J. DIFFERENTIAL GEOMETRY
64 (2003) 111-170

SURFACE GROUP REPRESENTATIONS AND U(p,q)-HIGGS BUNDLES

STEVEN B. BRADLOW[1,2], OSCAR GARCÍA-PRADA[1,3,5,6] & PETER B. GOTHEN[1,4,5]

Abstract

Using the L^2 norm of the Higgs field as a Morse function, we study the moduli spaces of $\mathrm{U}(p,q)$-Higgs bundles over a Riemann surface. We require that the genus of the surface be at least two, but place no constraints on (p,q). A key step is the identification of the function's local minima as moduli spaces of holomorphic triples. In a companion paper [7] we prove that these moduli spaces of triples are nonempty and irreducible.

Because of the relation between flat bundles and fundamental group representations, we can interpret our conclusions as results about the number of connected components in the moduli space of semisimple $\mathrm{PU}(p,q)$-representations. The topological invariants of the flat bundles are used to label subspaces. These invariants are bounded by a Milnor–Wood type inequality. For each allowed value of the invariants satisfying a certain coprimality condition, we prove that the corresponding subspace is nonempty and connected. If the coprimality condition does not hold, our results apply to the closure of the moduli space of irreducible representations.

[1] Members of VBAC (Vector Bundles on Algebraic Curves), which is partially supported by EAGER (EC FP5 Contract no. HPRN-CT-2000-00099) and by EDGE (EC FP5 Contract no. HPRN-CT-2000-00101).

[2] Partially supported by the National Science Foundation under grant DMS-0072073.

[3] Partially supported by the Ministerio de Ciencia y Tecnología (Spain) under grant BFM2000-0024.

[4] Partially supported by the Fundação para a Ciência e a Tecnologia (Portugal) through the Centro de Matemática da Universidade do Porto and through grant no. SFRH/BPD/1606/2000.

[5] Partially supported by the Portugal/Spain bilateral Programme Acciones Integradas, grant nos. HP2000-0015 and AI-01/24.

[6] Partially supported by a British EPSRC grant (October-December 2001).

Received 11/27/2002.

1. Introduction

The relation between Higgs bundles and fundamental group representations provides a vivid illustration of the interaction between geometry and topology. On the topological side we have a closed oriented surface X and the moduli space (or character variety) of representations of $\pi_1 X$ in a Lie group G. We cross over to complex geometry by fixing a complex structure on X, thereby turning it into a Riemann surface. The space of representations, or equivalently the space of flat G-bundles, then emerges as a complex analytic moduli space of G-Higgs bundles. In this guise, the moduli space carries a natural proper function whose restriction to the smooth locus is a Morse-Bott function. We can therefore use this function to determine topological properties of the moduli space of representations. Our goal in this paper is to pursue these ideas in the case where the group G is the real Lie group $\mathrm{PU}(p,q)$, the adjoint form of the noncompact group $\mathrm{U}(p,q)$.

The relevant Higgs bundles in our situation are $\mathrm{U}(p,q)$-Higgs bundles. These can be seen as a special case of the G-Higgs bundles defined by Hitchin in [22], where G is a real form of a complex reductive Lie group. Such objects provide a natural generalization of holomorphic vector bundles, which correspond to the case $G = \mathrm{U}(n)$ and zero Higgs field. In particular, they permit an extension to other groups of the Narasimhan and Seshadri theorem ([26]) on the relation between unitary representations of $\pi_1 X$ and stable vector bundles. By embedding $\mathrm{U}(p,q)$ in $\mathrm{GL}(p+q)$ we can give a concrete description of a $\mathrm{U}(p,q)$-Higgs bundle as a pair

$$\left(V \oplus W, \Phi = \left(\begin{smallmatrix} 0 & \beta \\ \gamma & 0 \end{smallmatrix}\right)\right) \tag{1.1}$$

where V and W are holomorphic vector bundles of rank p and q respectively, β is a section in $H^0(\mathrm{Hom}(W,V)\otimes K)$, and $\gamma \in H^0(\mathrm{Hom}(V,W)\otimes K)$, so that $\Phi \in H^0(\mathrm{End}(V\oplus W)\otimes K)$.

By the work of Hitchin [22, 23] Donaldson [12], Simpson [29, 30, 31, 32] and Corlette [10], we can define moduli spaces of polystable Higgs bundles, and these can be identified with moduli spaces of solutions to natural gauge theoretic equations. Moreover, since the gauge theory equations amount to a projective flatness condition, these moduli spaces correspond to moduli spaces of flat structures. In the case of $\mathrm{U}(p,q)$-Higgs bundles, the flat structures correspond to semi-simple representations of $\pi_1 X$ into the group $\mathrm{PU}(p,q)$. The Higgs bundle moduli spaces can thus be used, in a way which we make precise in Sections 2 and 3, to study the representation variety

$$\mathcal{R}(\mathrm{PU}(p,q)) = \mathrm{Hom}^+(\pi_1 X, \mathrm{PU}(p,q))/\mathrm{PU}(p,q),$$

where $\mathrm{Hom}^+(\pi_1 X, \mathrm{PU}(p,q))$ denotes the set of semi-simple representations of $\pi_1 X$ in $\mathrm{PU}(p,q)$, and the quotient is by the adjoint action.

Our main tool for studying the topology of the Higgs moduli space is the function which measures the L^2-norm of the Higgs field. When the moduli space is smooth, this turns out to provide a suitably non-degenerate Bott-Morse function which is, moreover, a proper map. In some cases (cf. [22, 18, 20]) the critical submanifolds are well enough understood to allow the extraction of topological information as detailed as the Poincaré polynomial. In our case our understanding is confined to the local minima of the function. This is sufficient to allow us to count the number of components of the Higgs moduli spaces, and thus of the representation varieties. A trivial but important observation is that the properness of the function allows us to draw conclusions about connected components also in the non-smooth case; we shall henceforth, somewhat imprecisely, refer to the function as the "Morse Function", whether or not the moduli space is smooth.

The criterion we use for finding the local minima can be applied more generally, for instance if $\mathrm{U}(p,q)$ is replaced by any real form of a complex reductive group. This should provide an important tool for future research. In the present case, this criterion allows us to identify the subspaces of local minima as moduli spaces in their own right, namely as moduli spaces of the holomorphic triples introduced in [4]. In a companion paper [7] we develop the theory of such objects and their moduli spaces. Using the results of [7] we are able to deduce several results about the Higgs moduli spaces and also about the corresponding representation spaces.

The relation between Higgs bundles and surface group representations has been successfully exploited by others, going back originally to the work of Hitchin and Simpson on complex reductive groups. The use of Higgs bundle methods to study $\mathcal{R}(G)$ for real G was pioneered by Hitchin in [23], and further developed in [18, 19]. It has also been used by Xia and Xia-Markman (in [34, 35, 36, 24]) to study various special cases of $G = \mathrm{PU}(p,q)$. None of these, though, address the general case of $\mathrm{PU}(p,q)$, as we do in this paper.

We now give a brief summary of the contents and main results of this paper.

In Sections 2 and 3 we give some background and describe the basic objects of our study. In Section 2 we describe the natural invariants associated with representations of $\pi_1 X$ into $\mathrm{PU}(p,q)$. We also discuss the invariants associated with representations of Γ, the universal central extensions of $\pi_1 X$, into $\mathrm{U}(p,q)$. The space of such representations is denoted by $\mathcal{R}_\Gamma(\mathrm{U}(p,q))$. In both cases, these involve a pair of integers (a,b) which can be interpreted respectively as degrees of rank p and rank q vector bundles over X. In the case of the $\mathrm{PU}(p,q)$ representations, the pair is well-defined only as a class in a quotient $\mathbb{Z} \oplus \mathbb{Z}/(p,q)\mathbb{Z}$. This leads us to define subspaces $\mathcal{R}[a,b] \subset \mathcal{R}(\mathrm{PU}(p,q))$ and $\mathcal{R}_\Gamma(a,b) \subset \mathcal{R}_\Gamma(\mathrm{U}(p,q))$. For fixed (a,b), the space $\mathcal{R}_\Gamma(a,b)$ fibers over $\mathcal{R}[a,b]$ with connected fibers.

In Section 3 we define $\mathrm{U}(p,q)$-Higgs bundles and their moduli spaces and establish their essential properties. Thinking of a $\mathrm{U}(p,q)$-Higgs bundle as a pair $(V \oplus W, \Phi)$, the parameters (a,b) appear here as the degrees of the bundles V and W. The moduli space of polystable $\mathrm{U}(p,q)$-Higgs bundles with $\deg(V) = a$ and $\deg W = b$, which we denote by $\mathcal{M}(a,b)$, is the space that can be identified with the component $\mathcal{R}_\Gamma(a,b)$ of $\mathcal{R}_\Gamma(\mathrm{U}(p,q))$. This, together with the fibration over $\mathcal{R}_\Gamma(\mathrm{U}(p,q))$ are the crucial links between the Higgs moduli and the surface group representation varieties.

Fixing p,q,a and b, we begin the Morse theoretic analysis of $\mathcal{M}(a,b)$ in Section 4. The basic results we need (cf. Proposition 4.3) are that the L^2-norm of the Higgs field has a minimum on each connected component of $\mathcal{M}(a,b)$, and hence if the subspace of local minima is connected then so is $\mathcal{M}(a,b)$. We identify the local minima, the loci of which we denote by $\mathcal{N}(a,b)$, and prove (cf. Theorem 4.6 and Proposition 4.8) that these correspond precisely to holomorphic triples in the sense of [4]. A full treatment of holomorphic triples and their moduli spaces is given in [7]. We summarize the salient features of these moduli spaces in Section 5.

In Section 6 we knit together all the strands. Using the properties of the moduli spaces of triples, we establish the key (for our purposes) topological properties of the strata $\mathcal{N}(a,b)$. These lead directly to our main results for the moduli spaces $\mathcal{M}(a,b)$. Some of the results depend on (a,b) only in the combination

$$\tau = \tau(a,b) = 2\frac{aq-bp}{p+q},$$

known as the Toledo invariant. Indeed, (a,b) is constrained by the bounds $0 \leqslant |\tau| \leqslant \tau_M$, where $\tau_M = 2\min\{p,q\}(g-1)$. Originally proved by Domic and Toledo in [11], these bounds emerge naturally from our point of view (cf. Corollary 3.27 and Remark 5.7). Bounds on invariants of this type, for representations of finitely generated groups in $\mathrm{U}(p,q)$, have also recently been studied using techniques from ergodic theory (see [9]). Summarizing our main results, we prove:

Theorem A (Theorems 6.1 and 6.5)**.** *Fix positive integers* (p,q). *Take* $(a,b) \in \mathbb{Z}\oplus\mathbb{Z}$ *and let* $\tau(a,b)$ *be the Toledo invariant. Let* $\mathcal{M}^s(a,b) \subseteq \mathcal{M}(a,b)$ *denote the moduli space of strictly stable* $\mathrm{U}(p,q)$*-Higgs bundles.*

(1) $\mathcal{M}(a,b)$ *is nonempty if and only if* $0 \leqslant |\tau(a,b)| \leqslant \tau_M$. *If* $\tau(a,b) = 0$, *or* $|\tau(a,b)| = \tau_M$ *and* $p \neq q$ *then* $\mathcal{M}^s(a,b)$ *is empty; otherwise it is nonempty whenever* $\mathcal{M}(a,b)$ *is nonempty.*

(2) *If* $|\tau(a,b)| = 0$ *or* $|\tau(a,b)| = \tau_M$ *and* $p \neq q$ *then* $\mathcal{M}(a,b)$ *is connected.*

(3) *Whenever nonempty, the moduli space* $\mathcal{M}^s(a,b)$ *is a smooth manifold of the expected dimension (i.e.,* $1+(p+q)^2(g-1)$*), with connected closure* $\overline{\mathcal{M}}^s(a,b) \subseteq \mathcal{M}(a,b)$. *In these cases, if* $\mathcal{M}(a,b)$ *has more than one connected component, then* $\mathrm{GCD}(p+q, a+b) \neq 1$ *and, if* $p = q$, $0 < |\tau| \leqslant (p-1)(2g-2)$.

Theorem B (Theorem 3.32)**.** *Suppose that* $p \neq q$ *and* $(a,b) \in \mathbb{Z}\oplus\mathbb{Z}$ *are such that* $|\tau(a,b)| = \tau_M$. *To be specific, suppose that* $p < q$ *and* $\tau(a,b) = p(2g-2)$. *Then every element in* $\mathcal{M}(a,b)$ *decomposes as the direct sum of a polystable* $\mathrm{U}(p,p)$*-Higgs bundle with maximal Toledo invariant and a polystable vector bundle of rank* $q-p$. *Thus*

$$\mathcal{M}(p,q,a,b) \cong \mathcal{M}(p,p,a,a-p(2g-2)) \times M(q-p, b-a+p(2g-2)). \tag{1.2}$$

In particular, the smooth locus in $\mathcal{M}(p,q,a,b)$ *has dimension* $2+(q^2+5p^2-2pq)(g-1)$. *This is strictly smaller than the expected dimension if* $g \geqslant 2$.

(A similar result holds if $p > q$, *as well as if* $\tau = -p(2g-2)$.*)*

Since we identify $\mathcal{M}(a,b) = \mathcal{R}_\Gamma(a,b)$, we can translate these results directly into statements about $\mathcal{R}_\Gamma(a,b)$ (given in Theorems 6.6 and 6.7).[7] The subspace in $\mathcal{R}_\Gamma(a,b)$ which corresponds to $\mathcal{M}^s(a,b) \subseteq \mathcal{M}(a,b)$ is

[7] Note added in proof: after this paper was submitted, M. Burger, A. Iozzi and A. Wienhard published the note *Surface group representations with maximal Toledo invariant*, Comptes Rendus **336** (2003) 387–390, in which they prove a result similar to our Theorem 6.7.

denoted by $\mathcal{R}^*_\Gamma(a,b)$. The representations it labels include all the simple representations. Defining $\mathcal{R}^*_\Gamma(\mathrm{U}(p,q)) \subset \mathcal{R}_\Gamma(\mathrm{U}(p,q))$ to be the union over all (a,b) of the components $\mathcal{R}^*_\Gamma(a,b)$ we thus obtain:

Theorem C (Corollary 6.16)**.** *The moduli space $\mathcal{R}^*_\Gamma(\mathrm{U}(p,q))$ has*

$$2(p+q)\min\{p,q\}(g-1) + \mathrm{GCD}(p,q)$$

connected components.

Since $\mathcal{R}_\Gamma(a,b)$ fibers over $\mathcal{R}[a,b]$ with connected fibers, we can apply our results to the latter. The results are given in Theorems 6.10 and 6.11.

The above results fall just short of saying that the full moduli spaces $\mathcal{M}(a,b)$ $(= \mathcal{R}(a,b))$ and $\mathcal{R}[a,b]$ are connected for all allowed choices of (a,b). They show however that if any one is not connected then it has one (nonempty) connected component which contains all the irreducible objects. Any other components must thus consist entirely of reducible (or strictly semisimple) elements. Theorem B and its analogs for $\mathcal{R}_\Gamma(a,b)$ and $\mathcal{R}[a,b]$ generalize rigidity results of Toledo [33] (when $p=1$) and Hernández [21] (when $p=2$).

This paper, together with its companion [7] form a substantially revised version of the preprint [6]. The main results proved in this paper were announced in the note [5]. In that note we claim (without proof) that the connectedness results for the moduli spaces $\mathcal{R}(a,b)$ and $\mathcal{R}[a,b]$ hold without the above qualifications. This is a reasonable conjecture, which we hope to come back to in a future publication.

We note, finally, that our methods surely apply more widely than to $\mathrm{U}(p,q)$-Higgs bundles and $\mathrm{PU}(p,q)$ representations (see, for example, Remark 4.16). Moreover, careful scrutiny of the Lie algebra properties used in the proofs suggests certain aspects can be generalized to representations in any real group G for which G/H is hermitian symmetric, where $H \subset G$ is a maximal compact subgroup. This will be addressed in a future publication.

Acknowledgements. We thank the mathematics departments of the University of Illinois at Urbana-Champaign, the University Autónoma of Madrid and the University of Aarhus, the Department of Pure Mathematics of the University of Porto, the Mathematical Sciences Research Institute of Berkeley and the Mathematical Institute of the University of Oxford, and the Erwin Schrödinger International Institute for Mathematical Physics in Vienna for their hospitality during various stages of this research. We thank Fran Burstall, Bill Goldman, Nigel Hitchin, Eyal Markman, S. Ramanan, Domingo Toledo, and Eugene Xia, for many insights and patient explanations.

2. Representations of surface groups

In this section we record some general facts about representations of a surface group in $\mathrm{U}(p,q)$ or $\mathrm{PU}(p,q)$ and set up our notation. A very useful reference for the general theory is Goldman's paper [16].

2.1 Moduli spaces of representations

Let X be a closed oriented surface of genus $g \geqslant 2$. By definition $\mathrm{U}(p,q)$ is the subgroup of $\mathrm{GL}(n,\mathbb{C})$ (with $n = p+q$) which leaves invariant a hermitian form of signature (p,q). It is a noncompact real form of $\mathrm{GL}(n,\mathbb{C})$ with center $\mathrm{U}(1)$ and maximal compact subgroup $\mathrm{U}(p)\times\mathrm{U}(q)$. The quotient $\mathrm{U}(p,q)/(\mathrm{U}(p)\times\mathrm{U}(q))$ is a hermitian symmetric space. The adjoint form $\mathrm{PU}(p,q)$ is given by the exact sequence of groups

$$1 \longrightarrow \mathrm{U}(1) \longrightarrow \mathrm{U}(p,q) \longrightarrow \mathrm{PU}(p,q) \longrightarrow 1,$$

and we have a standard inclusion $\mathrm{PU}(p,q) \subset \mathrm{PGL}(n,\mathbb{C})$.

Definition 2.1. By a *representation* of $\pi_1 X$ in $\mathrm{PU}(p,q)$ we mean a homomorphism $\rho\colon \pi_1 X \to \mathrm{PU}(p,q)$. We say that a representation of $\pi_1 X$ in $\mathrm{PU}(p,q)$ is *semi-simple* if the induced (adjoint) representation on the Lie algebra of $\mathrm{PU}(p,q)$ is semi-simple. The group $\mathrm{PU}(p,q)$ acts on the set of representations via conjugation. Restricting to the semi-simple representations, we get the *moduli space* of representations,

$$\mathcal{R}(\mathrm{PU}(p,q)) = \mathrm{Hom}^+(\pi_1 X, \mathrm{PU}(p,q))/\mathrm{PU}(p,q). \tag{2.1}$$

The moduli space of representations can be described more concretely as follows. From the standard presentation

$$\pi_1 X = \left\langle A_1, B_1, \ldots, A_g, B_g \;\;\middle|\;\; \prod_{i=1}^{g} [A_i, B_i] = 1 \right\rangle$$

we see that $\mathrm{Hom}^+(\pi_1 X, \mathrm{PU}(p,q))$ can be embedded in $\mathrm{PU}(p,q)^{2g}$ via

$$\begin{aligned} \mathrm{Hom}^+(\pi_1 X, \mathrm{PU}(p,q)) &\to \mathrm{PU}(p,q)^{2g} \\ \rho &\mapsto (\rho(A_1), \ldots \rho(B_g)). \end{aligned}$$

We give $\mathrm{Hom}^+(\pi_1 X, \mathrm{PU}(p,q))$ the subspace topology and $\mathcal{R}(\mathrm{PU}(p,q))$ the quotient topology. This topology is Hausdorff because we have restricted attention to semi-simple representations.

Clearly any representation of $\pi_1 X$ in $\mathrm{U}(p,q)$ gives rise to a representation in $\mathrm{PU}(p,q)$; however, not all representations in $\mathrm{PU}(p,q)$ lift to $\mathrm{U}(p,q)$. We are thus motivated to consider representations of the central extension

$$0 \longrightarrow \mathbb{Z} \longrightarrow \Gamma \longrightarrow \pi_1 X \longrightarrow 1. \tag{2.2}$$

Such extensions are defined (as in [1]) by the generators $A_1, B_1, \ldots, A_g$, B_g and a central element J subject to the relation $\prod_{i=1}^{g}[A_i, B_i] = J$. With Γ thus defined, any representation of $\pi_1 X$ in $\mathrm{PU}(p,q)$ can be lifted to a representation of Γ in $\mathrm{U}(p,q)$.

In analogy with Definition 2.1 we make the following definition.

Definition 2.2. We define the *moduli space* of semi-simple representations of Γ in $\mathrm{U}(p,q)$ by

$$\mathcal{R}_\Gamma(\mathrm{U}(p,q)) = \mathrm{Hom}^+(\Gamma, \mathrm{U}(p,q))/\mathrm{U}(p,q), \tag{2.3}$$

where semi-simplicity is defined with respect to the induced adjoint representation. This space is topologized in the same way as $\mathcal{R}(\mathrm{PU}(p,q))$.

2.2 Invariants

Our basic objective is to study the number of connected components of the spaces $\mathcal{R}(\mathrm{PU}(p,q))$ and $\mathcal{R}_\Gamma(\mathrm{U}(p,q))$. The first step in the study of topological properties of these spaces is to identify the appropriate topological invariant of a representation $\rho\colon \pi_1 X \to G$. For a general connected Lie group G the relevant invariant is an obstruction class in $H^2(X, \pi_1 G) \cong \pi_1 G$ (see Goldman [16, 17]). In the following we give an explicit description of this invariant in our case, using characteristic classes of the flat bundles associated to representations of the fundamental group. In fact we shall not need the more general description of the invariant.

We begin by considering the case $G = \mathrm{U}(p,q)$. By the same argument as in [1][8], $\mathcal{R}_\Gamma(\mathrm{U}(p,q))$ can be identified with the moduli space of connections with central curvature on a fixed $\mathrm{U}(p,q)$-bundle on X.

[8] While [1] gives the argument for $\mathrm{U}(n)$ and $\mathrm{PU}(n)$, there are no essential changes to be made in order to adapt for the case of $\mathrm{U}(p,q)$ and $\mathrm{PU}(p,q)$.

Taking a reduction to the maximal compact $\mathrm{U}(p)\times\mathrm{U}(q)$, we thus associate to each class $\widetilde{\rho}\in\mathcal{R}_\Gamma(\mathrm{U}(p,q))$ a vector bundle of the form $V\oplus W$, where V and W are rank p and q respectively, and thus a pair of integers $(a,b)=(\deg(V),\deg(W))$. There is thus a map

$$\widetilde{c}\colon \mathcal{R}_\Gamma(\mathrm{U}(p,q))\to\mathbb{Z}\oplus\mathbb{Z}$$

given by $\widetilde{c}(\widetilde{\rho})=(a,b)$. The corresponding map on $\mathrm{Hom}^+(\Gamma,\mathrm{U}(p,q))$ is clearly continuous and thus locally constant. Since $\mathrm{U}(p,q)$ is connected, the map $\widetilde{c}$ is likewise continuous and thus constant on connected components. We make the following definition.

Definition 2.3. The subspace of $\mathcal{R}_\Gamma(\mathrm{U}(p,q))$ corresponding to representations with invariants (a,b) is denoted by

$$\begin{aligned}\mathcal{R}_\Gamma(a,b)&=\widetilde{c}^{\,-1}(a,b)\\&=\{\widetilde{\rho}\in\mathcal{R}_\Gamma(\mathrm{U}(p,q))\ \mid\ \widetilde{c}(\widetilde{\rho})=(a,b)\in\mathbb{Z}\oplus\mathbb{Z}\}.\end{aligned}$$

Note that $\mathcal{R}_\Gamma(a,b)$ is a union of connected components, because $\widetilde{c}$ is constant on each connected component.

Next we consider the case $G=\mathrm{PU}(p,q)$. Any flat $\mathrm{PU}(p,q)$-bundle lifts to a $\mathrm{U}(p,q)$-bundle with a connection with constant central curvature. This lift is, however, not uniquely determined: in fact two such $\mathrm{U}(p,q)$-bundles give rise to the same flat $\mathrm{PU}(p,q)$-bundle if and only if one can be obtained from the other by twisting with a line bundle L with a unitary connection of constant curvature. If the invariant of the $\mathrm{U}(p,q)$-bundle is (a,b) and the degree of L is l, then the invariant associated to the twisted bundle is $(a+pl,b+ql)$. There is thus a well-defined map

$$c\colon \mathcal{R}(\mathrm{PU}(p,q))\longrightarrow(\mathbb{Z}\oplus\mathbb{Z})/(p,q)\mathbb{Z},\tag{2.4}$$

where $(\mathbb{Z}\oplus\mathbb{Z})/(p,q)\mathbb{Z}$ denotes the quotient of $\mathbb{Z}\oplus\mathbb{Z}$ by the $\mathbb{Z}$-action $l\cdot(a,b)=(a+pl,b+ql)$. Notice that $(\mathbb{Z}\oplus\mathbb{Z})/(p,q)\mathbb{Z}$ can be identified with $\pi_1(\mathrm{PU}(p,q))$. The invariant defined by c is the same as the obstruction class defined by Goldman [16, 17].

Definition 2.4. Denote the image of (a,b) in $(\mathbb{Z}\oplus\mathbb{Z})/(p,q)\mathbb{Z}$ by $[a,b]$. The subspace of $\mathcal{R}(\mathrm{PU}(p,q))$ corresponding to representations with invariant $[a,b]$ is denoted by

$$\begin{aligned}\mathcal{R}[a,b]&=c^{-1}[a,b]\\&=\{\rho\in\mathcal{R}(\mathrm{PU}(p,q))\ \mid\ c(\rho)=[a,b]\in(\mathbb{Z}\oplus\mathbb{Z})/(p,q)\mathbb{Z}\}.\end{aligned}$$

The space $\mathcal{R}[a,b]$ is a union of connected components in the same way as $\mathcal{R}_\Gamma(a,b)$. In order to compare the spaces $\mathcal{R}_\Gamma(a,b)$ and $\mathcal{R}[a,b]$ notice that we have surjective maps

$$\mathcal{R}_\Gamma(a,b) \to \mathcal{R}[a,b]. \tag{2.5}$$

Moreover, the preimage

$$\pi^{-1}(\mathcal{R}[a,b]) = \bigcup_{(a,b)} \mathcal{R}_\Gamma(a,b) \tag{2.6}$$

where the union is over all (a,b) in the class $[a,b] \in (\mathbb{Z}\oplus\mathbb{Z})/(p,q)\mathbb{Z}$. As mentioned above, tensoring by line bundles of degree l with constant curvature connections gives an isomorphism

$$\mathcal{R}_\Gamma(a,b) \xrightarrow{\cong} \mathcal{R}_\Gamma(a+pl, b+ql).$$

Notice that if $c(\rho) = [a,-a]$ for a representation $\rho \in \mathcal{R}(\mathrm{PU}(p,q))$, then the associated $\mathrm{U}(p,q)$-bundle can be taken to have degree zero and the projectively flat connection is actually flat. Then ρ defines a representation of $\pi_1 X$ in $\mathrm{U}(p,q)$. Under the correspondence between $\mathcal{R}(\mathrm{PU}(p,q))$ and $\mathcal{R}_\Gamma(\mathrm{U}(p,q))$, ρ corresponds to a Γ representation in which the central element J acts trivially. Furthermore, the subspaces $\mathcal{R}_\Gamma(a,-a) \subset \mathcal{R}_\Gamma(\mathrm{U}(p,q))$ can be identified with components of $\mathcal{R}(\mathrm{U}(p,q))$ (the moduli space for representations of $\pi_1 X$ in $\mathrm{U}(p,q)$). Indeed, defining

$$\mathcal{R}(a) = \mathcal{R}_\Gamma(a,-a), \tag{2.7}$$

we see that $\mathcal{R}(\mathrm{U}(p,q))$ is a union over $a \in \mathbb{Z}$ of the subspaces $\mathcal{R}(a)$.

Finally, we observe that the moduli space of flat degree zero line bundles acts by tensor product of bundles on $\mathcal{R}_\Gamma(a,b)$. Since this moduli space is isomorphic to the torus $\mathrm{U}(1)^{2g}$, we get the following relation between connected components.

Proposition 2.5. *The map $\mathcal{R}_\Gamma(a,b) \to \mathcal{R}[a,b]$ given in (2.5) defines a $\mathrm{U}(1)^{2g}$-fibration. Thus the subspace $\mathcal{R}[a,b] \subseteq \mathcal{R}(\mathrm{PU}(p,q))$ is connected if $\mathcal{R}_\Gamma(a,b)$ is connected.*

3. Higgs bundles and flat connections

We study the moduli spaces of representations by choosing a complex structure on X. This allows us to identify these spaces with certain moduli spaces of Higgs bundles. In this section we explain this correspondence and recall some general facts about Higgs bundles. Following this, we describe the special class of Higgs bundles relevant for the study of representations in $\mathrm{PU}(p,q)$ and $\mathrm{U}(p,q)$ and derive some basic results about these moduli spaces.

3.1 $\mathrm{GL}(n,\mathbb{C})$-Higgs bundles

Give X the structure of a Riemann surface. We recall (from [10, 12, 22, 29, 31, 32]) the following definition and basic facts about $\mathrm{GL}(n,\mathbb{C})$-Higgs bundles.

Definition 3.1.

(1) A $\mathrm{GL}(n,\mathbb{C})$-*Higgs bundle* on X is a pair (E,Φ), where E is a rank n holomorphic vector bundle over X and $\Phi \in H^0(\mathrm{End}(E)\otimes K)$ is a holomorphic endomorphism of E twisted by the canonical bundle K of X.

(2) The $\mathrm{GL}(n,\mathbb{C})$-Higgs bundle (E,Φ) is *stable* if the slope stability condition

$$\mu(E') < \mu(E) \tag{3.1}$$

holds for all proper Φ-invariant subbundles E' of E. Here the *slope* is defined by $\mu(E) = \deg(E)/\operatorname{rk}(E)$ and Φ-*invariance* means that $\Phi(E') \subset E' \otimes K$. *Semistability* is defined by replacing the above strict inequality with a weak inequality. A Higgs bundle is called *polystable* if it is the direct sum of stable Higgs bundles with the same slope.

(3) Given a hermitian metric on E, let A denote the unique unitary connection compatible with the holomorphic structure, and let F_A be its curvature. *Hitchin's equations* on (E,Φ) are

$$\begin{aligned} F_A + [\Phi,\Phi^*] &= -\sqrt{-1}\mu \mathrm{Id}_E\omega, \\ \overline{\partial}_A\Phi &= 0, \end{aligned} \tag{3.2}$$

where μ is a constant, Id_E is the identity on E, $\overline{\partial}_A$ is the anti-holomorphic part of the covariant derivative d_A and ω is the Kähler form on X. If we normalize ω so that $\int_X \omega = 2\pi$ then, taking the trace and integrating over X in the first equation, one sees that $\mu = \mu(E)$. A solution to Hitchin's equations is *irreducible* if there is no proper subbundle of E preserved by A and Φ.

Theorem 3.2.

(1) *Let (E, Φ) be a $\mathrm{GL}(n, \mathbb{C})$-Higgs bundle. Then (E, Φ) is polystable if and only if it admits a hermitian metric such that Hitchin's equations (3.2) are satisfied. Moreover, (E, Φ) is stable if and only if the corresponding solution is irreducible.*

(2) *Fix a hermitian metric in a smooth rank n complex vector bundle on X, then there is a gauge theoretic moduli space of pairs (A, Φ), consisting of a unitary connection A and an endomorphism valued $(1, 0)$-form Φ, which are solutions to Hitchin's equations (3.2), modulo $\mathrm{U}(n)$-gauge equivalence.*

(3) *The moduli space of rank n degree d polystable Higgs bundles is a quasi-projective variety of complex dimension $2(1 + n^2(g-1))$. There is a map from the gauge theoretic moduli space to this moduli space given by taking a solution (A, Φ) to Hitchin's equations to the Higgs bundle (E, Φ), where the holomorphic structure on E is given by $\overline{\partial}_A$. This map is a homeomorphism, and a diffeomorphism on the smooth locus.*

(4) *If we define a Higgs connection (as in [31]) by*

$$D = d_A + \theta \tag{3.3}$$

where $\theta = \Phi + \Phi^*$, then Hitchin's equations are equivalent to the conditions

$$\begin{aligned} F_D &= -\sqrt{-1}\mu Id_E \omega, \\ d_A \theta &= 0, \\ d_A^* \theta &= 0. \end{aligned} \tag{3.4}$$

In particular, D is a projectively flat connection. If $\deg(E) = 0$ then D is actually flat. It follows that in this case the pair (E, D) defines a representation of $\pi_1 X$ in $\mathrm{GL}(n, \mathbb{C})$. If $\deg(E) \neq 0$, then the pair (E, D) defines a representation of $\pi_1 X$ in $\mathrm{PGL}(n, \mathbb{C})$, or equivalently, a representation of Γ in $\mathrm{GL}(n, \mathbb{C})$. By the theorem of Corlette ([10]), every semisimple representation of Γ (and therefore every semisimple representation of $\pi_1 X$) arises in this way.

(5) *This correspondence gives rise to a homeomorphism between the moduli space of polystable Higgs bundles of rank n and the moduli*

space of semisimple representations of Γ *in* $\mathrm{GL}(n,\mathbb{C})$*. If the degree of the Higgs bundle is zero, then the moduli space is homeomorphic to the moduli space of representations of* $\pi_1 X$ *in* $\mathrm{GL}(n,\mathbb{C})$*.*

3.2 $\mathrm{U}(p,q)$-Higgs bundles

If we fix integers p and q such that $n = p + q$, then we can isolate a special class of $\mathrm{GL}(n,\mathbb{C})$-Higgs bundles by the requirements that

$$E = V \oplus W \tag{3.5}$$
$$\Phi = \begin{pmatrix} 0 & \beta \\ \gamma & 0 \end{pmatrix}$$

where V and W are holomorphic vector bundles of rank p and q respectively and the nonzero components in the Higgs field are

$$\beta \in H^0(\mathrm{Hom}(W,V)\otimes K), \quad \text{and} \quad \gamma \in H^0(\mathrm{Hom}(V,W)\otimes K).$$

The form of the Higgs field is determined by the Lie theory of the symmetric space $\mathrm{U}(p,q)/(\mathrm{U}(p)\times\mathrm{U}(q))$. Recall that for any real form G of a complex reductive group $G^{\mathbb{C}}$, with maximal compact subgroup H, there is an Ad-invariant decomposition

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

where $\mathfrak{g}$=Lie(G), $\mathfrak{h}$=Lie(H) is the $+1$ eigenspace of the Cartan involution and $\mathfrak{m}$ is the -1 eigenspace. This induces a decomposition

$$\mathfrak{g}^{\mathbb{C}} = \mathfrak{h}^{\mathbb{C}} + \mathfrak{m}^{\mathbb{C}} \tag{3.6}$$

of $\mathfrak{g}^{\mathbb{C}}$=Lie$(G^{\mathbb{C}})$. In the case of $G = \mathrm{U}(p,q)$, where $H = \mathrm{U}(p)\times\mathrm{U}(q)$ and thus $\mathfrak{h}^{\mathbb{C}} = \mathfrak{gl}(p,\mathbb{C})\oplus\mathfrak{gl}(q,\mathbb{C})$, the decomposition (3.6) becomes

$$\mathfrak{gl}(n,\mathbb{C}) = (\mathfrak{gl}(p,\mathbb{C})\oplus\mathfrak{gl}(q,\mathbb{C})) + \mathfrak{m}^{\mathbb{C}}. \tag{3.7}$$

If we identify $\mathfrak{gl}(p,\mathbb{C})\oplus\mathfrak{gl}(q,\mathbb{C})$ with the block diagonal elements in $\mathfrak{gl}(n,\mathbb{C})$, then $\mathfrak{m}^{\mathbb{C}}$ corresponds to the off diagonal matrices.

We can now describe the above Higgs bundles more intrinsically as follows. Let $P_{\mathrm{GL}(p,\mathbb{C})}$ and $P_{\mathrm{GL}(q,\mathbb{C})}$ be the principal frame bundles for V and W respectively. Let $P = P_{\mathrm{GL}(p,\mathbb{C})} \times P_{\mathrm{GL}(q,\mathbb{C})}$ be the fiber product, and let $\mathrm{Ad}\,P = P \times_{\mathrm{Ad}} \mathfrak{gl}(n,\mathbb{C})$ be the adjoint bundle, where $\mathrm{GL}(p,\mathbb{C})\times\mathrm{GL}(q,\mathbb{C}) \subset \mathrm{GL}(n,\mathbb{C})$ acts by the adjoint action on the Lie algebra of $\mathrm{GL}(n,\mathbb{C})$. This defines a subbundle

$$P_{\mathfrak{m}^{\mathbb{C}}} = P \times_{\mathrm{Ad}} \mathfrak{m}^{\mathbb{C}} \subset \mathrm{Ad}\,P. \tag{3.8}$$

We can then make the following definition.

Definition 3.3. A $\mathrm{U}(p,q)$-*Higgs bundle*[9] on X is a pair (P,Φ) where P is a holomorphic principal $\mathrm{GL}(p,\mathbb{C})\times\mathrm{GL}(q,\mathbb{C})$ bundle, and Φ is a holomorphic section of the vector bundle $P_{\mathfrak{m}^{\mathbb{C}}}\otimes K$ (where $P_{\mathfrak{m}^{\mathbb{C}}}$ is the bundle defined in (3.8)).

Remark 3.4. We can always write $P = P_{\mathrm{GL}(p,\mathbb{C})}\times P_{\mathrm{GL}(q,\mathbb{C})}$. If we let V and W be the standard vector bundles associated to $P_{\mathrm{GL}(p,\mathbb{C})}$ and $P_{\mathrm{GL}(q,\mathbb{C})}$ respectively, then any $\Phi\in H^0(P_{\mathfrak{m}^{\mathbb{C}}}\otimes K)$ can be written as in (3.5). We will usually adopt the vector bundle description of $\mathrm{U}(p,q)$-Higgs bundles.

Remark 3.5. Definition 3.3 is compatible with the definitions in [23] and [18], where G-Higgs bundles are defined for any real form G of a complex reductive Lie group $G^{\mathbb{C}}$. There, using the above notation, a G-Higgs bundle is a pair (P,Φ), where P is a principal $H^{\mathbb{C}}$-bundle and Φ is a holomorphic section of $(P\times_{\mathrm{Ad}}\mathfrak{m}^{\mathbb{C}})\otimes K$. From a different perspective, Definition 3.3 defines an example of a principal pair in the sense of [2] and [25]. Strictly speaking, since the canonical bundle K plays the role of a fixed 'twisting bundle', what we get is a principal pair in the sense of [8]. The defining data for the pair are then the principal $\mathrm{GL}(p,\mathbb{C})\times\mathrm{GL}(q,\mathbb{C})\times\mathrm{GL}(1)$-bundle $P_{\mathrm{GL}(p,\mathbb{C})}\times P_{\mathrm{GL}(q,\mathbb{C})}\times P_K$ (where P_K is the frame bundle for K), and the associated vector bundle $P_{\mathfrak{m}^{\mathbb{C}}}\otimes K$.

Lemma 3.6. *Let $(E=V\oplus W,\Phi)$ be a $\mathrm{U}(p,q)$-Higgs bundle with a hermitian metric such that $V\oplus W$ is a unitary orthogonal decomposition. Let A be a unitary connection and let $D=d_A+\theta$ be the corresponding Higgs connection, where $\theta=\Phi+\Phi^*$. Then D is a $\mathrm{U}(p,q)$-connection, i.e., in any unitary local frame the connection 1-form takes its values in the Lie algebra of $\mathrm{U}(p,q)$.*

Proof. Fix a local unitary frame. Then $D=d+A+\theta$, where A takes its values in $\mathfrak{u}(p)\oplus\mathfrak{u}(q)\subset\mathfrak{u}(p,q)$, while θ takes its values in $\mathfrak{m}$, where

$$\mathfrak{u}(p,q)=\mathfrak{u}(p)\oplus\mathfrak{u}(q)+\mathfrak{m}$$

is the eigenspace decomposition of the Cartan involution. q.e.d.

Definition 3.7. Let (E,Φ) be a $\mathrm{U}(p,q)$-Higgs bundle with $E=V\oplus W$ and $\Phi=\begin{pmatrix}0&\beta\\\gamma&0\end{pmatrix}$. We say (E,Φ) is a *stable* $\mathrm{U}(p,q)$-Higgs bundle if the slope stability condition $\mu(E')<\mu(E)$, is satisfied for all Φ-invariant subbundles of the form $E'=V'\oplus W'$, i.e., for all subbundles

[9] The reason for the name is explained by Remark 3.5 and Lemma 3.6

$V' \subset V$ and $W' \subset W$ such that

$$\beta : W' \longrightarrow V' \otimes K \tag{3.9}$$

$$\gamma : V' \longrightarrow W' \otimes K. \tag{3.10}$$

Semistability for $\mathrm{U}(p,q)$-Higgs bundles is defined by replacing the above strict inequality with a weak inequality, and (E,Φ) is *polystable* if it is a direct sum of stable $\mathrm{U}(p,q)$-Higgs bundles all of the same slope. We shall say that a polystable $\mathrm{U}(p,q)$-Higgs bundle which is not stable is *reducible*. A *morphism* between two $\mathrm{U}(p,q)$-Higgs bundles $(V \oplus W, \Phi)$ and $(V' \oplus W', \Phi')$ is given by maps $g_V : V \to V'$ and $g_W : W \to W'$ which intertwine Φ and Φ', i.e., such that $(g_V \oplus g_W) \otimes I_K \circ \Phi = \Phi' \circ (g_V \oplus g_W)$ where I_K is the identity on K. In particular we have a natural notion of isomorphism of $\mathrm{U}(p,q)$-Higgs bundles.

Remark 3.8. The stability condition for a $\mathrm{U}(p,q)$-Higgs bundle is *a priori* weaker than the stability condition given in Definition 3.1 for $\mathrm{GL}(n,\mathbb{C})$-Higgs bundles. However, it is shown in [19, Section 2.3] that the weaker condition is in fact equivalent to the ordinary stability of (E,Φ).

Proposition 3.9. *Let (E,Φ) be a $\mathrm{U}(p,q)$-Higgs bundle with $E = V \oplus W$ and $\Phi = \left(\begin{smallmatrix} 0 & \beta \\ \gamma & 0 \end{smallmatrix}\right)$. Then (E,Φ) is polystable if and only if it admits a hermitian metric such that $E = V \oplus W$ is an orthogonal decomposition and such that Hitchin's equations (3.2) are satisfied.*

Proof. This is a special case of the correspondence invoked in [23] for G-Higgs bundles where G is a real form of a reductive Lie group. By Remark 3.5 it can also be seen as a special case of the Hitchin–Kobayashi correspondence for principal pairs (cf. [2] and [25] and [8]). We note finally that in one direction the result follows immediately from Theorem 3.2 (1): if $(V \oplus W, \Phi)$ supports a compatible metric such that (3.2) is satisfied, then it is polystable as a $\mathrm{GL}(n,\mathbb{C})$-Higgs bundle, and hence it is $\mathrm{U}(p,q)$-polystable. q.e.d.

Definition 3.10. We define $\mathcal{M}(a,b)$ to be the moduli space of polystable $\mathrm{U}(p,q)$-Higgs bundles with $\deg(V) = a$ and $\deg W = b$. We denote by $\mathcal{M}^s(a,b)$ the subspace parameterizing the strictly stable $\mathrm{U}(p,q)$-Higgs bundles.

The construction of $\mathcal{M}(a,b)$ is essentially the same as in Section 9 of [32]. There the moduli space of G-Higgs bundles is constructed for any reductive group G. We take $G = \mathrm{GL}(p,\mathbb{C}) \times \mathrm{GL}(q,\mathbb{C})$. The difference between a $\mathrm{U}(p,q)$-Higgs bundle and a $\mathrm{GL}(p,\mathbb{C}) \times \mathrm{GL}(q,\mathbb{C})$-Higgs bundle is entirely in the nature of the Higgs fields. Taking the standard embedding of $\mathrm{GL}(p,\mathbb{C}) \times \mathrm{GL}(q,\mathbb{C})$ in $\mathrm{GL}(p+q,\mathbb{C})$ we see that in a $\mathrm{GL}(p,\mathbb{C}) \times \mathrm{GL}(q,\mathbb{C})$-Higgs bundle the Higgs field Φ takes its values in the subspace $(\mathfrak{gl}(p) \oplus \mathfrak{gl}(q)) \subset \mathfrak{gl}(p+q)$, while in a $\mathrm{U}(p,q)$-Higgs bundle the Higgs field Φ takes its values in the complementary subspace $\mathfrak{m}^{\mathbb{C}}$ (as in (3.7)). Since both subspaces are invariant under the adjoint action of $\mathrm{GL}(p,\mathbb{C}) \times \mathrm{GL}(q,\mathbb{C})$, the same method of construction works for the moduli spaces of both types of Higgs bundle.

We can describe the gauge theory version of the moduli space $\mathcal{M}(a,b)$ using standard methods; see Hitchin [22] for a construction in the case of ordinary rank 2 Higgs bundles. To adapt to our case we proceed as follows. Let $E = V \oplus W$ be a smooth complex vector bundle with a hermitian metric such that the direct sum decomposition is orthogonal. We let $\mathcal{A}$ denote the space of connections on E which are direct sums of unitary connections on V and W and we let $\mathbf{\Omega}$ denote the space of Higgs fields $\Phi \in \Omega^{1,0}(\mathrm{End}(E))$ of the form $\Phi = \left(\begin{smallmatrix} 0 & \beta \\ \gamma & 0 \end{smallmatrix}\right)$. The correspondence between unitary connections and holomorphic structures via $\overline{\partial}$-operators turns $\mathcal{A} \times \mathbf{\Omega}$ into a complex affine space which acquires a hermitian metric using the metric on E and integration over X. The group $\mathcal{G}$ of $\mathrm{U}(p) \times \mathrm{U}(q)$-gauge transformations acts on the configuration space $\mathcal{C} \subseteq \mathcal{A} \times \mathbf{\Omega}$ of solutions (A,Φ) to Hitchin's equations (3.2). The quotient $\mathcal{C}/\mathcal{G}$ is, by definition, the gauge theory moduli space. As in [22], the open subset of $\mathcal{C}/\mathcal{G}$ corresponding to irreducible solutions has a Kähler manifold structure.

To see that the gauge theory moduli space is homeomorphic to $\mathcal{M}(a,b)$ we can consider this latter space from the complex analytic point of view (cf. Remark 3.23 below): consider triples $(\overline{\partial}_V, \overline{\partial}_W, \Phi)$, where $\overline{\partial}_V$ and $\overline{\partial}_W$ are $\overline{\partial}$-operators on V and W, respectively, and $\Phi \in \mathbf{\Omega}$. Let $\mathcal{C}_{\mathbb{C}}$ be the set of such triples for which Φ is holomorphic and the associated $\mathrm{U}(p,q)$-Higgs bundle is polystable. We can then view $\mathcal{M}(a,b)$ as the quotient of $\mathcal{C}_{\mathbb{C}}$ by the complex gauge group. We clearly have an inclusion $\mathcal{C} \hookrightarrow \mathcal{C}_{\mathbb{C}}$ which descends to give a continuous map from the gauge theory moduli space to $\mathcal{M}(a,b)$. The Hitchin-Kobayashi correspondence of Proposition 3.9 now shows that this map is in fact a homeomorphism.

For a third perspective, we observe that provided that V and W are not isomorphic bundles, i.e., provided $p \neq q$ or $a \neq b$, we can view

$\mathcal{M}^s(a,b)$ as a subvariety of a moduli space of stable $\mathrm{GL}(p+q)$-Higgs bundle. If $V \simeq W$, then $\mathcal{M}^s(a,b)$ is a finite cover of a subvariety in the larger moduli space:

Proposition 3.11. *With $n = p+q$ and $d = a+b$, let $\mathcal{M}^s(d)$ denote the moduli space of stable $\mathrm{GL}(n,\mathbb{C})$-Higgs bundles of degree d. If $p \neq q$ or $a \neq b$ then $\mathcal{M}^s(a,b)$ embeds as a closed subvariety in $\mathcal{M}^s(d)$. If $p = q$ and $a = b$, then there is an involution on $\mathcal{M}^s(a,a)$ such that the quotient injects into $\mathcal{M}^s(d)$.*

Proof. Let $[V \oplus W, \Phi]_{p,q}$ denote the point in $\mathcal{M}^s(a,b)$ represented by the $\mathrm{U}(p,q)$-Higgs bundle $(V \oplus W, \Phi)$. Then $(E = V \oplus W, \Phi)$ is a stable $\mathrm{GL}(n,\mathbb{C})$-Higgs bundle and the map $\mathcal{M}^s(a,b) \to \mathcal{M}(d)$ is defined by

$$[V \oplus W, \Phi]_{p,q} \mapsto [E, \Phi]_n,$$

where $[,]_n$ denotes the isomorphism class in $\mathcal{M}(d)$. The only question is whether this map is injective. Suppose that $(E = V \oplus W, \Phi)$ and $(E' = V' \oplus W', \Phi')$ are isomorphic as $\mathrm{GL}(n,\mathbb{C})$-Higgs bundles. Let the isomorphism be given by a complex gauge transformation $g : E \to E'$. If g is not of the form $\left(\begin{smallmatrix} g_V & 0 \\ 0 & g_W \end{smallmatrix}\right)$ then the off diagonal components determine morphisms $\xi : V \to W'$ and $\sigma : W \to V'$. Let $N = \ker(\xi) \oplus \ker(\sigma)$ be the subbundle of $V \oplus W$ determined by the kernels of ξ and σ. If $p \neq q$ then N is a nontrivial proper subbundle. Moreover, using the fact that $g\Phi = \Phi' g$, we see that it is Φ-invariant. Since $(V \oplus W, \Phi)$ is stable, it follows that

$$\mu(N) < \mu(E). \tag{3.11}$$

Similarly, the images of ξ and σ determine a proper Φ'-invariant subbundle of E', say I, for which

$$\mu(I) < \mu(E'). \tag{3.12}$$

But if $\mu(E) = \mu(E')$ then (3.11) and (3.12) cannot both be satisfied. Thus ξ and σ must both vanish and hence $[V \oplus W, \Phi]_{p,q} = [V' \oplus W', \Phi']_{p,q}$.

If $p = q$, then this argument can fail, but only if ξ and σ are both isomorphisms. In that case, $N = 0$ and $I = E$. This also requires $a = b$. Under these conditions, if V and W are non-isomorphic, then $[V \oplus W, \left(\begin{smallmatrix} 0 & \beta \\ \gamma & 0 \end{smallmatrix}\right)]_n = [W \oplus V, \left(\begin{smallmatrix} 0 & \gamma \\ \beta & 0 \end{smallmatrix}\right)]_n$ but the Higgs bundles are not isomorphic as $\mathrm{U}(p,q)$-Higgs bundles. Hence the last statement of the Proposition follows taking the involution $[V \oplus W, \left(\begin{smallmatrix} 0 & \beta \\ \gamma & 0 \end{smallmatrix}\right)] \mapsto [W \oplus V, \left(\begin{smallmatrix} 0 & \gamma \\ \beta & 0 \end{smallmatrix}\right)]$ on $\mathcal{M}^s(a,a)$. q.e.d.

Proposition 3.12. *If* $\mathrm{GCD}(p+q, a+b) = 1$ *then* $\mathcal{M}^s(a,b) = \mathcal{M}(a,b)$*.*

Proof. If $\mathrm{GCD}(p+q, a+b) = 1$ then for purely numerical reasons there are no strictly semistable $\mathrm{U}(p,q)$-Higgs bundles in $\mathcal{M}(a,b)$. q.e.d.

The link to moduli spaces of representations is provided by the next result.

Proposition 3.13. *There is a homeomorphism* $\mathcal{M}(a,b) \cong \mathcal{R}_\Gamma(a,b)$*.*

Proof. Suppose that $(E = V \oplus W, \Phi)$ represents a point in $\mathcal{M}(a,b)$, i.e., suppose that it is a $\mathrm{U}(p,q)$-polystable Higgs bundle, and suppose that E has a hermitian metric such that the direct sum decomposition is orthogonal and Hitchin's equations (3.2) are satisfied. Rewriting the equations in terms of the Higgs connection $D = d_A + \theta$, where A is the metric connection and $\theta = \Phi + \Phi^*$, we see that D is projectively flat. By Lemma 3.6 it is a projectively flat $\mathrm{U}(p,q)$-connection, and thus defines a point in $\mathcal{R}_\Gamma(a,b)$. Conversely by Corlette's theorem [10], every representation in $\mathrm{Hom}^+(\pi_1 X, \mathrm{PU}(p,q))$, or equivalently every representation in $\mathrm{Hom}^+(\Gamma, \mathrm{U}(p,q))$, arises in this way. The fact that this correspondence gives a homeomorphism follows by the same argument as the one given in [32] for ordinary Higgs bundles. q.e.d.

Definition 3.14. Define the subspace $\mathcal{R}^*_\Gamma(a,b)$ to be the subspace corresponding to $\mathcal{M}^s(a,b)$ via the homeomorphism in Proposition 3.13. Using the fibration of $\mathcal{R}_\Gamma(a,b)$ over $\mathcal{R}[a,b]$, define $\mathcal{R}^*[a,b] \subset \mathcal{R}[a,b]$ to be the image of $\mathcal{R}^*_\Gamma(a,b)$.

Remark 3.15. Thus $\mathcal{R}^*_\Gamma(a,b)$ parameterizes the representations which give rise to stable $\mathrm{U}(p,q)$-Higgs bundles. Recall from Remark 3.8 that a $\mathrm{U}(p,q)$-Higgs bundle is stable (in the sense of Definition 3.7) if and only if its is stable as an ordinary $\mathrm{GL}(n,\mathbb{C})$-Higgs bundle. Now, a $\mathrm{GL}(n,\mathbb{C})$-Higgs bundle is stable if and only if the corresponding representation of Γ on $\mathbb{C}^n$ is irreducible (cf. Corlette [10]). Hence we see that the subspace $\mathcal{R}^*_\Gamma(a,b)$ corresponds to the representations of Γ in $\mathrm{U}(p,q)$ which are irreducible as $\mathrm{GL}(n,\mathbb{C})$ representations. Similarly, the subspace $\mathcal{R}^*[a,b]$ corresponds to the representations of $\pi_1 X$ which are irreducible as $\mathrm{PGL}(n,\mathbb{C})$ representations.

We point out, moreover, that the subspace $\mathcal{R}^*_\Gamma(a,b)$ includes as a dense open set the representations whose induced adjoint representations on the Lie algebra of $\mathrm{PU}(p,q)$ are irreducible. It may also contain some representations whose induced adjoint representation is reducible for the following reason. If $(E = V \oplus W, \Phi)$ is the $\mathrm{U}(p,q)$-Higgs bundle corresponding to a representation in $\mathcal{R}^*_\Gamma(a,b)$, then $(\mathrm{End}(E), \Phi)$ is a polystable Higgs bundle but it is not necessarily stable. The representations with reducible induced adjoint representation are the ones for which $(\mathrm{End}(E), \Phi)$ is strictly polystable.

3.3 Deformation theory

The results of Biswas and Ramanan [3] and Hitchin [23] readily adapt to describe the deformation theory of $\mathrm{U}(p,q)$-Higgs bundles.

Definition 3.16. Let $(E = V \oplus W, \Phi)$ be a $\mathrm{U}(p,q)$-Higgs bundle. We introduce the following notation:

$$\begin{aligned} U &= \mathrm{End}(E), \\ U^+ &= \mathrm{End}(V) \oplus \mathrm{End}(W), \\ U^- &= \mathrm{Hom}(W, V) \oplus \mathrm{Hom}(V, W). \end{aligned}$$

With this notation, $U = U^+ \oplus U^-$, $\Phi \in H^0(U^- \otimes K)$, and $\mathrm{ad}(\Phi)$ interchanges U^+ and U^-. We consider the complex of sheaves

$$C^\bullet : U^+ \xrightarrow{\mathrm{ad}(\Phi)} U^- \otimes K. \tag{3.13}$$

Lemma 3.17. *Let (E, Φ) be a stable $\mathrm{U}(p,q)$-Higgs bundle. Then*

$$\ker\big(\mathrm{ad}(\Phi) \colon H^0(U^+) \to H^0(U^- \otimes K)\big) = \mathbb{C}, \tag{3.14}$$

$$\ker\big(\mathrm{ad}(\Phi) \colon H^0(U^-) \to H^0(U^+ \otimes K)\big) = 0. \tag{3.15}$$

Proof. By Remark 3.8 (E, Φ) is stable as a $\mathrm{GL}(n, \mathbb{C})$-Higgs bundle. Hence it is simple, that is, its only endomorphisms are the nonzero scalars. Thus,

$$\ker\big(\mathrm{ad}(\Phi) \colon H^0(U) \to H^0(U \otimes K)\big) = \mathbb{C}.$$

Since $U = U^+ \oplus U^-$ and $\mathrm{ad}(\Phi)$ interchanges these two summands, the statements of the lemma follow. q.e.d.

Proposition 3.18 (Biswas-Ramanan [3])**.**

(1) *The space of endomorphisms of (E, Φ) is isomorphic to the zeroth hypercohomology group $\mathbb{H}^0(C^\bullet)$.*

(2) *The space of infinitesimal deformations of* (E, Φ) *is isomorphic to the first hypercohomology group* $\mathbb{H}^1(C^\bullet)$.

(3) *There is a long exact sequence*

$$(3.16)\quad 0 \longrightarrow \mathbb{H}^0(C^\bullet) \longrightarrow H^0(U^+) \longrightarrow H^0(U^- \otimes K) \longrightarrow \mathbb{H}^1(C^\bullet) \longrightarrow H^1(U^+) \longrightarrow H^1(U^- \otimes K) \longrightarrow \mathbb{H}^2(C^\bullet) \longrightarrow 0,$$

where the maps $H^i(U^+) \longrightarrow H^i(U^- \otimes K)$ *are induced by* $\mathrm{ad}(\Phi)$.

Proposition 3.19. *Let* (E, Φ) *be a stable* $\mathrm{U}(p,q)$*-Higgs bundle, then:*

(1) $\mathbb{H}^0(C^\bullet) = \mathbb{C}$ *(in other words* (E, Φ) *is simple) and*

(2) $\mathbb{H}^2(C^\bullet) = 0$.

Proof. (1) Follows immediately from Lemma 3.17 and Proposition 3.18 (3).

(2) We have natural ad-invariant isomorphisms $U^+ \cong (U^+)^*$ and $U^- \cong (U^-)^*$. Thus

$$\mathrm{ad}(\Phi)\colon H^1(U^+) \to H^1(U^- \otimes K)$$

is Serre dual to $\mathrm{ad}(\Phi)\colon H^0(U^-) \to H^0(U^+ \otimes K)$. Hence Lemma 3.17 and (3) of Proposition 3.18 show that $\mathbb{H}^2(C^\bullet) = 0$. q.e.d.

Proposition 3.20. *The moduli space of stable* $\mathrm{U}(p,q)$*-Higgs bundles is a smooth complex variety of dimension* $1 + (p+q)^2(g-1)$.

Proof. By Proposition 3.19 (2) $\mathbb{H}^2(C^\bullet) = 0$ at all points in the moduli space of stable $\mathrm{U}(p,q)$-Higgs bundles. Smoothness is thus a consequence of the results of [3], as follows. Let $e \in \mathcal{M}(a,b)$ be the point corresponding to a stable $\mathrm{U}(p,q)$-Higgs bundle (E, Φ) and let $\mathcal{F}$ be the infinitesimal deformation functor of (E, Φ) as in [3]. Then the completion of the local ring $\mathcal{O}_e$ pro-represents $\mathcal{F}$ (cf. Schlessinger [28]). Now Proposition 3.19 and Theorem 3.1 of [3] show that the completion of $\mathcal{O}_e$ is regular and hence $\mathcal{O}_e$ is itself regular. Thus $\mathcal{M}(a,b)$ is smooth at e.

Using (2) and (3) of Proposition 3.18, Proposition 3.19 and the Riemann-Roch Theorem, the dimension of the moduli space is given by

$$\begin{aligned}\dim \mathbb{H}^1(C^\bullet) &= 1 - \chi(U^+) + \chi(U^- \otimes K) \\ &= 1 + (p^2+q^2)(g-1) + 2pq(g-1) \\ &= 1 + (p+q)^2(g-1).\end{aligned}$$

q.e.d.

Remark 3.21. The dimension of the moduli space of stable $\mathrm{U}(p,q)$-Higgs bundles is half that of the moduli space of stable $\mathrm{GL}(p+q,\mathbb{C})$-Higgs bundles.

Remark 3.22. By Proposition 3.12 $\mathcal{M}(a,b)$ is smooth if $\mathrm{GCD}(p+q, a+b) = 1$.

Remark 3.23. As an alternative to the algebraic arguments of [3], the fact that the deformation theory of a $\mathrm{U}(p,q)$-Higgs bundle is controlled by the complex of sheaves (3.13) can be seen from the complex analytic point of view as follows. As in the gauge theory construction of $\mathcal{M}(a,b)$ (cf. Section 3.2) let $V \oplus W$ be a smooth complex vector bundle, and consider a $\mathrm{U}(p,q)$-Higgs bundle as being given by a triple $(\overline{\partial}_V, \overline{\partial}_W, \Phi)$. Now write down a Dolbeault resolution of the complex $C^\bullet$:

$$\begin{array}{ccc}\Omega^0(U^+) & \xrightarrow{\mathrm{ad}(\Phi)} & \Omega^{1,0}(U^-) \\ \downarrow{\scriptstyle\overline{\partial}} & & \downarrow{\scriptstyle -\overline{\partial}} \\ \Omega^{0,1}(U^+) & \xrightarrow{\mathrm{ad}(\Phi)} & \Omega^{1,1}(U^-) \\ \downarrow & & \downarrow \\ 0 & \longrightarrow & 0\end{array}.$$

Consider the associated total complex $\mathbf{C}^0 \xrightarrow{D^0} \mathbf{C}^1 \xrightarrow{D^1} \mathbf{C}^2$. Then $\mathbf{C}^0$ is the Lie algebra of the $\mathrm{GL}(p,\mathbb{C}) \times \mathrm{GL}(q,\mathbb{C})$-gauge group and $\mathbf{C}^1$ is the tangent space to the affine space of triples $(\overline{\partial}_V, \overline{\partial}_W, \Phi)$. Furthermore, D^0 is the infinitesimal action of the complex gauge group, while D^1 is the derivative of the holomorphicity condition: this gives the desired interpretation of the deformation complex $C^\bullet$ in complex analytic terms.

To conclude this line of thought we give an alternative argument for the smoothness of the moduli space of stable $\mathrm{U}(p,q)$-Higgs bundles: suppose that $(\overline{\partial}_V, \overline{\partial}_W, \Phi)$ corresponds to a stable $\mathrm{U}(p,q)$-Higgs bundle (E,Φ). Proposition 3.19 shows that $\mathbb{H}^0(C^\bullet) = \mathbb{C}$ and $\mathbb{H}^2(C^\bullet) = 0$. The differential of the holomorphicity condition is thus surjective and (E,Φ) has no nontrivial automorphisms. It follows by standard arguments that the moduli space can be constructed as a smooth complex manifold near (E,Φ).

3.4 Bounds on the topological invariants

In this section we show how the Higgs bundle point of view provides an easy proof of a result of Domic and Toledo [11] which allows us to bound the topological invariants $\deg(V)$ and $\deg(W)$ for which $\mathrm{U}(p,q)$-Higgs bundles may exist. The lemma is a slight variation on the results of [19, Section 3] (cf. also Lemma 3.6 of Markman and Xia [24]).

Lemma 3.24. *Let (E,Φ) be a semistable $\mathrm{U}(p,q)$-Higgs bundle. Then*

$$p(\mu(V)-\mu(E)) \leqslant \mathrm{rk}(\gamma)(g-1), \tag{3.17}$$

$$q(\mu(W)-\mu(E)) \leqslant \mathrm{rk}(\beta)(g-1). \tag{3.18}$$

If equality occurs in (3.17) then either (E,Φ) is strictly semistable or $p=q$ and γ is an isomorphism. If equality occurs in (3.18) then either (E,Φ) is strictly semistable or $p=q$ and β is an isomorphism.

Proof. If $\gamma=0$ then V is Φ-invariant. By stability, $\mu(V)\leqslant\mu(E)$ and equality can only occur if (E,Φ) is strictly semistable. This proves (3.17) in the case $\gamma=0$. We may therefore assume that $\gamma\neq 0$. Let $N=\ker(\gamma)\subseteq V$ and let $I=\mathrm{im}(\gamma)\otimes K^{-1}\subseteq W$. Then

$$\mathrm{rk}(N)+\mathrm{rk}(I)=p \tag{3.19}$$

and, since γ induces a nonzero section of $\det((V/N)^*\otimes I\otimes K)$,

$$\deg(N)+\deg(I)+\mathrm{rk}(I)(2g-2)\geqslant\deg(V). \tag{3.20}$$

The bundles N and $V\oplus I$ are Φ-invariant subbundles of E and hence we obtain by semistability that $\mu(N)\leqslant\mu(E)$ and $\mu(V\oplus I)\leqslant\mu(E)$ or, equivalently, that

$$\deg(N)\leqslant\mu(E)\,\mathrm{rk}(N), \tag{3.21}$$

$$\deg(I)\leqslant\mu(E)(p+\mathrm{rk}(I))-\deg(V). \tag{3.22}$$

Adding (3.21) and (3.22) and using (3.19) we obtain

$$\deg(N)+\deg(I)\leqslant 2\mu(E)p-\deg(V). \tag{3.23}$$

Finally, combining (3.20) and (3.23) we get

$$\deg(V) - \mathrm{rk}(I)(2g-2) \leqslant 2\mu(E)p - \deg(V),$$

which is equivalent to (3.17) since $\mathrm{rk}(\gamma) = \mathrm{rk}(I)$. Note that equality can only occur if we have equality in (3.21) and (3.22) and thus either (E, Φ) is strictly semistable or neither of the subbundles N and $V \oplus I$ is proper and nonzero. In the latter case, clearly $N = 0$ and $I = W$ and therefore $p = q$; furthermore we must also have equality in (3.20) implying that γ is an isomorphism. An analogous argument applied to β proves (3.18). q.e.d.

Remark 3.25. The proof also shows that if we have equality in, say, (3.17) then $\gamma\colon V/N \to I \otimes K$ is an isomorphism. In particular, if $p < q$ and $\mu(V) - \mu(E) = g - 1$ then $\gamma\colon V \xrightarrow{\cong} I \otimes K$.

We can re-formulate Lemma 3.24 to obtain the following corollary.

Corollary 3.26. *Let (E, Φ) be a semistable $\mathrm{U}(p,q)$-Higgs bundle. Then*

$$q(\mu(E) - \mu(W)) \leqslant \mathrm{rk}(\gamma)(g-1), \tag{3.24}$$

$$p(\mu(E) - \mu(V)) \leqslant \mathrm{rk}(\beta)(g-1). \tag{3.25}$$

Proof. Use $\mu(W) - \mu(E) = \frac{p}{q}\big(\mu(E) - \mu(V)\big)$ to see that (3.24) is equivalent to (3.17). Similarly (3.25) is equivalent to (3.18). q.e.d.

An important corollary of the lemma above is the following Milnor–Wood type inequality for $\mathrm{U}(p,q)$-Higgs bundles (due to Domic and Toledo [11], improving on a bound obtained by Dupont [13] in the case $G = \mathrm{SU}(p,q)$). This result gives bounds on the possible values of the topological invariants $\deg(V)$ and $\deg(W)$.

Corollary 3.27. *Let (E, Φ) be a semistable $\mathrm{U}(p,q)$-Higgs bundle. Then*

$$\frac{pq}{p+q}|\mu(V) - \mu(W)| \leqslant \min\{p,q\}(g-1). \tag{3.26}$$

Proof. Since $\mu(E) = \frac{p}{p+q}\mu(V) + \frac{q}{p+q}\mu(W)$ we have $\mu(V) - \mu(E) = \frac{q}{p+q}(\mu(V) - \mu(W)$ and therefore (3.17) gives

$$\frac{pq}{p+q}(\mu(V) - \mu(W)) \leqslant \mathrm{rk}(\gamma)(g-1).$$

A similar argument using (3.18) shows that

$$\frac{pq}{p+q}(\mu(W)-\mu(V)) \leqslant \mathrm{rk}(\beta)(g-1).$$

But, obviously, $\mathrm{rk}(\beta)$ and $\mathrm{rk}(\gamma)$ are both less than or equal to $\min\{p,q\}$.
q.e.d.

Definition 3.28. The *Toledo invariant* of the representation corresponding to $(E = V \oplus W, \Phi)$ is

$$\tau = \tau(a,b) = 2\frac{qa-pb}{p+q} \tag{3.27}$$

where $a = \deg(V)$ and $b = \deg(W)$.

Remark 3.29. Since

$$\tau = 2\frac{pq}{p+q}(\mu(V)-\mu(W)) = -2p(\mu(E)-\mu(V)) = 2q(\mu(E)-\mu(W)),$$

the inequalities in Lemma 3.24 and Corollary 3.26 can be written as

$$\frac{\tau}{2} \leqslant \mathrm{rk}(\gamma)(g-1), \tag{3.28}$$

$$-\frac{\tau}{2} \leqslant \mathrm{rk}(\beta)(g-1). \tag{3.29}$$

Similarly the inequality (3.26) can be written $|\tau| \leqslant \tau_M$, where

$$\tau_M = \min\{p,q\}(2g-2). \tag{3.30}$$

3.5 Rigidity and extreme values of the Toledo invariant

If $|\tau| = \tau_M$ then the moduli space $\mathcal{M}(a,b)$ has special features. These depend on whether $p = q$ or $p \neq q$.

Consider first the case $p = q$. Notice that if $p = q$ then $\tau(a,b) = a-b$ and $\tau_M = p(2g-2)$. We thus examine the moduli space $\mathcal{M}(a,b)$ when $|a-b| = p(2g-2)$. Before giving a description we review briefly the notion of L-twisted Higgs pairs. Let L be a line bundle. An L-twisted Higgs pair (V,θ) consists of a holomorphic vector bundle V and an L-twisted homomorphism $\theta : V \longrightarrow V \otimes L$. The notions of stability, semistability and polystability are defined as for Higgs bundles. The moduli space of semistable L-twisted Higgs pairs has been constructed by Nitsure using Geometric Invariant Theory [27]. Let $\mathcal{M}_L(n,d)$ be the moduli space of polystable L-twisted Higgs pairs of rank n and degree d.

Proposition 3.30. *Let $p = q$ and $|a - b| = p(2g - 2)$. Then*

$$\mathcal{M}(a, b) \cong \mathcal{M}_{K^2}(p, a) \cong \mathcal{M}_{K^2}(p, b).$$

Proof. Let $(E = V \oplus W, \Phi) \in \mathcal{M}(a, b)$. Suppose for definiteness that $b - a = p(2g - 2)$. From (3.18) it follows that $\gamma : V \longrightarrow W \otimes K$ is an isomorphism. We can then compose $\beta : W \longrightarrow V \otimes K$ with $\gamma \otimes \mathrm{Id}_K : V \otimes K \longrightarrow W \otimes K^2$ to obtain a K^2-twisted Higgs pair $\theta_W : W \longrightarrow W \otimes K^2$. Similarly, twisting $\beta : W \longrightarrow V \otimes K$ with K and composing with γ, we obtain a K^2-twisted Higgs pair $\theta_V : V \longrightarrow V \otimes K^2$. Conversely, given an isomorphism $\gamma : V \longrightarrow W \otimes K$, we can recover β from θ_V as well as from θ_W. It is clear that the (poly)stability of (E, Φ) is equivalent to the (poly)stability of (V, θ_V) and to the (poly)stability of (W, θ_W), proving the claim. q.e.d.

Remark 3.31. The moduli space $\mathcal{M}_{K^2}(p, a)$ contains an open (irreducible) subset consisting of a vector bundle over the moduli space of stable bundles of rank p and degree a. This is because the stability of V implies the stability of any K^2-twisted Higgs pair (V, θ_V), and $H^1(\mathrm{End}\, V \otimes K^2) = 0$. The rank of the bundle is determined by the Riemann–Roch Theorem.

Now consider the case $p \neq q$. For definiteness, we assume $p < q$. We use the more precise notation $\mathcal{M}(p, q, a, b)$ for the moduli space of $\mathrm{U}(p, q)$-Higgs bundles such that $\deg(V) = a$, and $\deg(W) = b$, and write the Toledo invariant as

$$\tau = \tau(p, q, a, b) = 2\frac{qa - pb}{p + q}. \tag{3.31}$$

Theorem 3.32. *Suppose (p, q, a, b) are such that $p < q$ and $|\tau(p, q, a, b)| = p(2g - 2)$. Then every element in $\mathcal{M}(p, q, a, b)$ is strictly semistable and decomposes as the direct sum of a polystable $\mathrm{U}(p, p)$-Higgs bundle with maximal Toledo invariant and a polystable vector bundle of rank $(q - p)$. If $\tau = p(2g - 2)$, then*

$$\begin{aligned} \mathcal{M}(p, q, a, b) &\cong \\ &\mathcal{M}(p, p, a, a - p(2g - 2)) \times M(q - p, b - a + p(2g - 2)), \end{aligned} \tag{3.32}$$

where $M(q - p, b - a + p(2g - 2))$ denotes the moduli space of polystable bundles of rank $q - p$ and degree $b - a + p(2g - 2)$. In particular, the dimension at a smooth point in $\mathcal{M}(p, q, a, b)$ is $2 + (p^2 + 5q^2 - 2pq)(g - 1)$, and it is hence strictly smaller than the expected dimension.

(A similar result holds if $\tau = -p(2g - 2)$ and also if $p > q$.)

Proof. Let $(E = V \oplus W, \Phi) \in \mathcal{M}(p,q,a,b)$ and suppose $\tau(p,q,a,b) = p(2g-2)$. Then $\mu(V) - \mu(E) = g-1$ and $\mu(E) - \mu(W) = \frac{p}{q}(g-1)$. Since $\mathrm{rk}(\beta)$ and $\mathrm{rk}(\gamma)$ are at most p, it follows from (3.17) and (3.25) that $\mathrm{rk}(\beta) = \mathrm{rk}(\gamma) = p$. Let $W_\gamma = \mathrm{im}(\gamma) \otimes K^{-1}$ and let $W_\beta = \ker(\beta)$. Then $V \oplus W_\gamma$ is a Φ-invariant subbundle of $V \oplus W$, and $\mu(V \oplus W_\gamma) = \mu(0 \oplus W_\beta) = \mu(E)$. We see that (E, Φ) is strictly semistable (as we already knew from Lemma 3.24). Since it is polystable it must split as

$$(V \oplus W_\gamma, \Phi) \oplus (0 \oplus W/W_\gamma, 0).$$

It is clear that $(V \oplus W_\gamma, \Phi) \in \mathcal{M}(p,p,a,a-p(2g-2))$ and that $(V \oplus W_\gamma, \Phi)$ has maximal Toledo invariant, that is, $\tau(p,p,a,a-p(2g-2)) = 2p(g-1)$. Also, using

$$0 \longrightarrow \ker(\Phi) \longrightarrow V \oplus W \longrightarrow (V \oplus W_\gamma) \otimes K \longrightarrow 0.$$

we see that $W/W_\gamma \in M(q-p, b-a+p(2g-2))$. To complete the proof we observe that

$$\begin{aligned} &\dim \mathcal{M}^s(p,p,a,a-p(2g-2)) + \dim M^s(q-p, b-a+p(2g-2)) \\ &= 1 + (2p)^2(g-1) + 1 + (q-p)^2(g-1) = 2 + (p^2 + 5q^2 - 2pq)(g-1). \end{aligned}$$

Since $q > 1$, this is smaller than $1 + (p+q)^2(g-1)$, the dimension of $\mathcal{M}(p,q,a,b)$ when the Toledo invariant is not maximal. q.e.d.

Remark 3.33. The fact the moduli space has smaller dimension than expected may be viewed as a certain kind of rigidity. This phenomenon (for large Toledo invariant) has been studied from the point of view of representations of the fundamental group by D. Toledo [33] when $p = 1$ and L. Hernández [21] when $p = 2$. We deal here with the general case which, as far as we know, has not appeared previously in the literature.

Corollary 3.34. *Fix (p,q,a,b) such that $p < q$ and $\tau(p,q,a,b) = p(2g-2)$. Then*

$$\mathcal{M}(p,q,a,b) \cong \mathcal{M}_{K^2}(p, a-p(2g-2)) \times M(q-p, b-a+p(2g-2)).$$

Proof. It follows from Theorem 3.32 and Proposition 3.30. q.e.d.

4. Morse theory

Morse theoretic techniques for studying the topology of moduli spaces of Higgs bundles were introduced by Hitchin [22, 23]. Though standard Morse theory cannot be applied to $\mathcal{M}(a,b)$ when it is not smooth, as we shall see in the following, we can still use Morse theory ideas to count connected components. Throughout this section we assume that p and q are any positive integers and that $(a,b) \in \mathbb{Z} \oplus \mathbb{Z}$ is such that $|\tau| \leqslant \tau_M$, where τ is as in Definition 3.28 and τ_M is given by (3.30).

4.1 The Morse function

Consider the moduli space $\mathcal{M}(a,b)$ from the gauge theory point of view (cf. Section 3.2). We can then define a real positive function

$$\begin{aligned} f\colon \mathcal{M}(a,b) &\to \mathbb{R} \\ [A,\Phi] &\mapsto \tfrac{1}{\pi}\|\Phi\|^2, \end{aligned} \tag{4.1}$$

where the L^2-norm of Φ is $\|\Phi\|^2 = \frac{\sqrt{-1}}{2}\int_X \operatorname{tr}(\Phi\Phi^*)$.

We have the following result due to Hitchin [22].

Proposition 4.1.

(1) *The function f is proper.*

(2) *The restriction of f to $\mathcal{M}^s(a,b)$ is a moment map (up to a constant) for the Hamiltonian circle action $[A,\Phi] \mapsto [A, e^{i\theta}\Phi]$.*

(3) *If $\mathcal{M}(a,b)$ is smooth, then f is a perfect Bott-Morse function.*

Thus, if the moduli space is smooth, then its number of connected components is bounded by the number of connected components of the subspace of local minima of f. However, even if $\mathcal{M}(a,b)$ is not smooth, f can be used to obtain information about the connected components of $\mathcal{M}(a,b)$ using the following elementary result.

Proposition 4.2. *Let Z be a Hausdorff space and let $f\colon Z \to \mathbb{R}$ be proper and bounded below. Then f attains a minimum on each connected component of Z and, furthermore, if the subspace of local minima of f is connected then so is Z.*

In particular this applies to our situation, giving:

Proposition 4.3. *The function $f\colon \mathcal{M}(a,b) \to \mathbb{R}$ defined in (4.1) has a minimum on each connected component of $\mathcal{M}(a,b)$. Moreover, if the subspace of local minima of f is connected then so is $\mathcal{M}(a,b)$.*

Definition 4.4. Let

$$\mathcal{N}(a,b) = \{(E,\Phi) \in \mathcal{M}(a,b) \mid \beta = 0 \text{ or } \gamma = 0\}. \tag{4.2}$$

Proposition 4.5. *For all $(E,\Phi) \in \mathcal{M}(a,b)$*

$$f(E,\Phi) \geqslant \frac{|\tau(a,b)|}{2} \tag{4.3}$$

with equality if and only if $(E,\Phi) \in \mathcal{N}(a,b)$.

Proof. Writing out the first of Hitchin's equations (3.2) for a $\mathrm{U}(p,q)$-Higgs bundle (E,Φ) in its componenents on V and W we get the pair of equations

$$\begin{aligned} F(A_V) + \beta\beta^* + \gamma^*\gamma &= -\sqrt{-1}\mu \operatorname{Id}_V \omega, \\ F(A_W) + \gamma\gamma^* + \beta^*\beta &= -\sqrt{-1}\mu \operatorname{Id}_W \omega, \end{aligned}$$

where A_V and A_W are the components on V and W, respectively, of the unitary connection A on $E = V \oplus W$. Taking the trace and integrating over X in the first of these equations we get from Chern-Weil theory

$$\deg(V) = \mu p - \tfrac{1}{\pi}\|\beta\|^2 + \tfrac{1}{\pi}\|\gamma\|^2,$$

where we have used $\int_X \omega = 2\pi$. Since $\mu = \mu(E)$, this is equivalent to

$$\tfrac{1}{\pi}\|\beta\|^2 - \tfrac{1}{\pi}\|\gamma\|^2 = p(\mu(E) - \mu(V)) = -\frac{\tau}{2}.$$

But $f(E,\Phi) = \frac{1}{\pi}\|\beta\|^2 + \frac{1}{\pi}\|\gamma\|^2$ and thus

$$\begin{aligned} f(E,\Phi) &= \tfrac{2}{\pi}\|\gamma\|^2 - \frac{\tau}{2} \\ &= \tfrac{2}{\pi}\|\beta\|^2 + \frac{\tau}{2}, \end{aligned} \tag{4.4}$$

from which the result is immediate. q.e.d.

The above proposition identifies $\mathcal{N}(a,b)$ as the set of global minima of f. The following theorem, which is of fundamental importance to our approach, shows that there are no other local minima.

Theorem 4.6. *Let (E, Φ) be a polystable* $\mathrm{U}(p,q)$*-Higgs bundle in* $\mathcal{M}(a,b)$*. Then (E,Φ) is a local minimum of $f\colon \mathcal{M}(a,b) \to \mathbb{R}$ if and only if (E,Φ) belongs to $\mathcal{N}(a,b)$.*

Proof. This follows directly from Proposition 4.5 above and Propositions 4.17 and 4.20, which are given in Sections 4.4 and 4.5, respectively. q.e.d.

Remark 4.7. This Theorem was already known to hold when $p, q \leqslant 2$ (by the results of [19], Hitchin [22], and Xia [36]), and also when $p = q$ and $(p-1)(2g-2) < |\tau| \leqslant p(2g-2)$ by Markman-Xia [24].

Which section actually vanishes for a minimum is given by the following.

Proposition 4.8. *Let $(E,\Phi) \in \mathcal{N}(a,b)$. Then:*

(1) *$\gamma = 0$ if and only if $a/p \leqslant b/q$ (i.e., $\tau \leqslant 0$). In this case,*

$$f(\mathcal{N}(a,b)) = b - q(a+b)/(p+q) = -\frac{\tau}{2}.$$

(2) *$\beta = 0$ if and only if $a/p \geqslant b/q$ (i.e., $\tau \geqslant 0$). In this case,*

$$f(\mathcal{N}(a,b)) = a - p(a+b)/(p+q) = \frac{\tau}{2}.$$

In particular, $\beta = \gamma = 0$ if and only if $a/p = b/q$ (i.e., $\tau = 0$) and, in this case, $f(E,\Phi) = 0$.

Proof. The relation between the conditions on τ and those on $a/p - b/q$ follows directly from the definition of τ (cf. (3.27)). The rest follows immediately from (4.4) and the fact that f is, by definition, nonnegative. Alternatively one can argue algebraically, using Lemma 3.24 and polystability. q.e.d.

Corollary 4.9. *If $a/p = b/q$ then $\mathcal{N}(a,b) \cong M(p,a) \times M(q,b)$.*

Proof. If $a/p = b/q$, then any $(E,\Phi) \in \mathcal{N}(a,b)$ has $E = V \oplus W$ and $\Phi = 0$. Polystability of (E,Φ) is thus equivalent to the polystability of V and W. q.e.d.

4.2 Critical points of the Morse function

In this section we recall Hitchin's method [22, 23] for determining the local minima of f and spell out how this works in the case of $\mathrm{U}(p,q)$-Higgs bundles.

Since f is a moment map, a smooth point of the moduli space is a critical point if and only if it is a fixed point of the circle action. To determine the fixed points, note that, if (A, Φ) represents a fixed point then there must be a 1-parameter family of gauge transformations $g(\theta)$ taking (A, Φ) to $(A, e^{i\theta}\Phi)$. This gives an infinitesimal $\mathrm{U}(p)\times\mathrm{U}(q)$-gauge transformation $\psi = \dot{g}$ which is covariantly constant (i.e., $d_A\psi = 0$) and such that $[\psi, \Phi] = i\Phi$. (Note that we can take ψ to be trace-free.) It follows that we can decompose E in holomorphic subbundles F_λ on which ψ acts as $i\lambda$ and furthermore that Φ maps F_λ to $F_{\lambda+1} \otimes K$. We thus have the following result.

Proposition 4.10. *A $\mathrm{U}(p, q)$-Higgs bundle (E, Φ) in $\mathcal{M}(a, b)$ represents a fixed point of the circle action if and only if it is a system of Hodge bundles, that is,*

$$E = F_1 \oplus \cdots \oplus F_m \tag{4.5}$$

for holomorphic vector bundles F_i such that the restriction

$$\Phi_i := \Phi_{|F_i} \in H^0(\mathrm{Hom}(F_i, F_{i+1}) \otimes K),$$

and the F_i are direct sums of bundles contained in V and W. Furthermore, each F_i is an eigenbundle for an infinitesimal trace-free gauge transformation ψ. If $\Phi_i \neq 0$, then the weight of ψ on F_{i+1} is one plus the weight of ψ on F_i. Moreover, if (E, Φ) is stable, then each restriction Φ_i is nonzero and the F_i are alternately contained in V and W.

Proof. Only the last statement requires a proof. But if some component of Φ vanished, or if some F_i had a nonzero component in both V and W, then (E, Φ) would be reducible and hence not stable. q.e.d.

When (E, Φ) is stable the decomposition $E = F_1 \oplus \cdots \oplus F_m$ gives a corresponding decomposition of the bundle $U = \mathrm{End}(E)$ into eigenbundles for the adjoint action of ψ:

$$U = \bigoplus_{k=-m+1}^{m-1} U_k,$$

where $U_k = \bigoplus_{i-j=k} \mathrm{Hom}(F_j, F_i)$ is the eigenbundle corresponding to the eigenvalue ik.

By Hitchin's calculations in [23, §8] (see also [18, Section 2.3.2]) the eigenvalues of the Hessian of f at a smooth critical point can be determined in the following way.

Proposition 4.11. *Let (E, Φ) be a stable $\mathrm{U}(p,q)$-Higgs bundle which represents a critical point of f. Then the eigenspace of the Hessian of f corresponding to the eigenvalue $-k$ is $\mathbb{H}^1$ of the following complex:*

$$C_k^\bullet : U_k^+ \xrightarrow{\mathrm{ad}(\Phi)} U_{k+1}^- \otimes K, \tag{4.6}$$

where we use the notation

$$U_k^+ = U_k \cap U^+,$$
$$U_k^- = U_k \cap U^-,$$

with U^+ and U^- as defined in Definition 3.16. In particular (E, Φ) corresponds to a local minimum of f if and only if

$$\mathbb{H}^1(C_k^\bullet) = 0$$

for all $k \geqslant 1$.

Remark 4.12. When (E, Φ) is a stable $\mathrm{U}(p,q)$-Higgs bundle, we know from Proposition 4.10 that the F_i are alternately contained in V and W. Thus we have

$$U^+ = \bigoplus_{k \text{ even}} U_k \; ; \quad U^- = \bigoplus_{k \text{ odd}} U_k. \tag{4.7}$$

In particular all the eigenvalues of the Hessian of f are even.

Remark 4.13. The description in Proposition 4.11 of the eigenspace of the Hessian of f gives rise to the long exact sequence

$$\begin{aligned} 0 \longrightarrow \mathbb{H}^0(C_k^\bullet) \longrightarrow H^0(U_k^+) \longrightarrow H^0(U_{k+1}^- \otimes K) \longrightarrow \mathbb{H}^1(C_k^\bullet) \\ \longrightarrow H^1(U_k^+) \longrightarrow H^1(U_{k+1}^- \otimes K) \longrightarrow \mathbb{H}^2(C_k^\bullet) \longrightarrow 0. \end{aligned}$$

Suppose that (E, Φ) is a stable $\mathrm{U}(p,q)$-Higgs bundle. The vanishing result of Proposition 3.19 shows that $\mathbb{H}^0(C_k^\bullet) = \mathbb{H}^2(C_k^\bullet) = 0$ for $k \neq 0$ (while $\mathbb{H}^0(C_0^\bullet) = \mathbb{C}$ and $\mathbb{H}^2(C_0^\bullet) = 0$). Hence one can use this exact sequence, Remark 4.12, and the Riemann–Roch formula to calculate the dimension of $\mathbb{H}^1(C_k^\bullet)$ for any k in terms of the ranks and the degrees of the F_i. This provides a method for calculating the Morse index of f at a critical point. However, we shall omit the formula since we have no need for it.

4.3 Local minima and the adjoint bundle

In this section we give a criterion for (E, Φ) to be a local minimum in terms of the adjoint bundle. This is the key step in the proof of Theorem 4.6. We use the notation introduced in Section 4.2.

Consider the complex $C_k^\bullet$ defined in (4.6) and let

$$\chi(C_k^\bullet) = \dim \mathbb{H}^0(C_k^\bullet) - \dim \mathbb{H}^1(C_k^\bullet) + \dim \mathbb{H}^2(C_k^\bullet).$$

Proposition 4.14. *Let (E, Φ) be a polystable $\mathrm{U}(p, q)$-Higgs bundle which is a fixed point of the S^1-action on $\mathcal{M}(a, b)$. Then $\chi(C_k^\bullet) \leqslant 0$ and equality holds if and only if*

$$\mathrm{ad}(\Phi) \colon U_k^+ \to U_{k+1}^- \otimes K$$

is an isomorphism.

Proof. For simplicity we shall adopt the notation

$$\Phi_k^\pm = \mathrm{ad}(\Phi)_{|U_k^\pm} \colon U_k^\pm \longrightarrow U_{k+1}^\mp \otimes K.$$

The key fact we need is that there is a natural ad-invariant isomorphism $U \cong U^*$ under which we have $U^+ \cong (U^+)^*$, $U^- \cong (U^-)^*$ and $U_k^\pm \cong (U_{-k}^\pm)^*$. Since $\mathrm{ad}(\Phi)^t = \mathrm{ad}(\Phi) \otimes 1_{K^{-1}}$ under this isomorphism we have

$$(\Phi_k^\pm)^t = \Phi_{-k-1}^\mp \otimes 1_{K^{-1}}. \tag{4.8}$$

We have the short exact sequence

$$0 \longrightarrow \ker(\Phi_k^+) \longrightarrow (U_{k+1}^- \otimes K)^* \longrightarrow \mathrm{im}(\Phi_k^+) \longrightarrow 0.$$

From (4.8) we have $\ker(\Phi_k^{+,t}) \cong \ker(\Phi_{-k-1}^-) \otimes K^{-1}$. Thus, tensoring the above sequence by K, we obtain the short exact sequence

$$0 \longrightarrow \ker(\Phi_{-k-1}^-) \longrightarrow (U_{k+1}^-)^* \longrightarrow \mathrm{im}(\Phi_k^+) \otimes K \longrightarrow 0.$$

It follows that

$$\deg(\mathrm{im}(\Phi_k^+)) \leqslant \deg(U_{k+1}^-) + (2g-2)\,\mathrm{rk}(\Phi_k^+) + \deg(\ker(\Phi_{-k-1}^-)).$$

Combining this inequality with the fact that

$$\deg(U_k^+) \leqslant \deg(\ker(\Phi_k^+)) + \deg(\mathrm{im}(\Phi_k^+)), \tag{4.9}$$

we obtain

$$\begin{aligned}(4.10)\qquad \deg(U_k^+) \leqslant \deg(U_{k+1}^-) &+ (2g-2)\operatorname{rk}(\Phi_k^+)\\ &+ \deg(\ker(\Phi_{-k-1}^-)) + \deg(\ker(\Phi_k^+)).\end{aligned}$$

Since (E, Φ) is semistable, so is the Higgs bundle $(\operatorname{End}(E), \operatorname{ad}(\Phi))$. Clearly the kernel $\ker(\Phi_k^\pm) \subseteq \operatorname{End}(E)$ is Φ-invariant and hence, from semistability,

$$\deg(\ker(\Phi_k^\pm)) \leqslant 0,$$

for all k. Substituting this inequality in (4.10), we obtain

$$(4.11)\qquad \deg(U_k^+) \leqslant \deg(U_{k+1}^-) + (2g-2)\operatorname{rk}(\Phi_k^+).$$

From the long exact sequence (4.6) and the Riemann–Roch formula we obtain

$$\begin{aligned}\chi(C_k^\bullet) &= \chi(U_k^+) - \chi(U_{k+1}^- \otimes K)\\ &= (1-g)\big(\operatorname{rk}(U_k^+) + \operatorname{rk}(U_{k+1}^-)\big) + \deg(U_k^+) - \deg(U_{k+1}^-).\end{aligned}$$

Using this identity and the inequality (4.11) we see that

$$\chi(C_k^\bullet) \leqslant (g-1)\big(2\operatorname{rk}(\Phi_k^+) - \operatorname{rk}(U_k^+) - \operatorname{rk}(U_{k+1}^-)\big).$$

Hence $\chi(C_k^\bullet) \leqslant 0$. Furthermore, if equality holds we have

$$\operatorname{rk}(\Phi_k^+) = \operatorname{rk}(U_k^+) = \operatorname{rk}(U_{k+1}^-)$$

and also equality must hold in (4.11) and so $\deg(\operatorname{im}(\Phi_k^+)) = \deg(U_{k+1}^- \otimes K)$, showing that Φ_k^+ is an isomorphism as claimed. q.e.d.

Corollary 4.15. *Let (E, Φ) be a stable $\mathrm{U}(p,q)$-Higgs bundle which represents a critical point of f. This critical point is a local minimum if and only if*

$$\operatorname{ad}(\Phi)\colon U_k^+ \to U_{k+1}^- \otimes K$$

is an isomorphism for all $k \geqslant 1$.

Proof. By Proposition 3.19 we have $\mathbb{H}^0(C_k^\bullet) = \mathbb{H}^2(C_k^\bullet) = 0$ for $k \geqslant 1$. Hence we have $-\chi(C_k^\bullet) = \mathbb{H}^1(C_k^\bullet)$ and the result follows from Propositions 4.11 and 4.14. q.e.d.

Remark 4.16. Let (P,Φ) be a G-Higgs bundle as defined in Remark 3.5 and define

$$\begin{aligned} U &= P \times_{\mathrm{Ad}} \mathfrak{g}^{\mathbb{C}}, \\ U^{+} &= P \times_{\mathrm{Ad}} \mathfrak{h}^{\mathbb{C}}, \\ U^{-} &= P \times_{\mathrm{Ad}} \mathfrak{m}^{\mathbb{C}}. \end{aligned}$$

Then $U = U^{+} \oplus U^{-}$ and if (P,Φ) is fixed under the circle action we can write $U = \bigoplus U_k$ as a direct sum of eigenbundles for an infinitesimal gauge transformation as before. Thus we can define a complex $C_k^{\bullet}$ as in (4.6). If (P,Φ) is a stable G-Higgs bundle, then the Higgs vector bundle $(U, \mathrm{ad}(\Phi))$ is semistable and so the proof of Proposition 4.14 goes through unchanged. Thus this key result is valid in the more general setting.

4.4 Stable Higgs bundles

In this section we prove Theorem 4.6 for stable Higgs bundles. The reducible (polystable) ones are dealt with in the next section. We continue to use the notation of Section 4.2.

Proposition 4.17. *Let $(E,\Phi) = (F_1 \oplus \cdots \oplus F_m, \Phi)$ be a stable $\mathrm{U}(p,q)$-Higgs bundle representing a critical point of f such that $m \geqslant 3$. Then (E,Φ) is not a local minimum of f.*

Proof. Note that $U_k = 0$ for $|k| \geqslant m$; in particular $U_m = 0$. We shall consider the cases when m is odd and even separately.

The case m odd. In this case $m-1$ is even and so, using Remark 4.12 we see that $U_{m-1}^{+} = U_{m-1} \neq 0$ while $U_m^{-} \subseteq U_m = 0$. Hence $\mathrm{ad}(\Phi)\colon U_{m-1}^{+} \to U_m^{-} \otimes K$ cannot be an isomorphism and we are done by Corollary 4.15.

The case m even. From Remark 4.12 we see that

$$\begin{aligned} U_{m-1}^{-} &= U_{m-1} = \mathrm{Hom}(F_1, F_m), \\ U_{m-2}^{+} &= U_{m-2} = \mathrm{Hom}(F_1, F_{m-1}) \oplus \mathrm{Hom}(F_2, F_m). \end{aligned}$$

Thus, by Corollary 4.15 it suffices to prove that

$$\mathrm{ad}(\Phi)\colon U_{m-2} \to U_{m-1} \otimes K$$

is not an isomorphism. In fact the restriction of $\mathrm{ad}(\Phi)$ to a fiber cannot even be injective. Indeed, if it were, then its restriction to

$\mathrm{Hom}(F_1, F_{m-1})$ would be injective and hence Φ_{m-1} would also be injective. Take a nonzero element $\eta \in \mathrm{Hom}(F_2, F_m)$ whose image is contained in the image of Φ_{m-1}. Define $\zeta = \Phi^{-1}\eta\Phi \in \mathrm{Hom}(F_1, F_{m-1})$. Then $\mathrm{ad}(\Phi)(\eta + \zeta) = 0$ which is a contradiction. q.e.d.

Remark 4.18. Let (E, Φ) be a stable $\mathrm{U}(p, q)$-Higgs bundle with $\beta = 0$ or $\gamma = 0$. Then, as pointed out above, Proposition 4.5 shows that (E, Φ) is a local minimum of f. This can also be seen from the Morse theory point of view, as follows. Such a Higgs bundle either has $\beta = \gamma = 0$ or it is a Hodge bundle of length 2. In the former case, clearly we have $\mathrm{End}(E) = U_0$. In the latter case, $E = F_1 \oplus F_2$ with $F_1 = V$ and $F_2 = W$ (if $\beta = 0$) or vice-versa (if $\gamma = 0$). Hence $\mathrm{End}(E) = U_{-1} \oplus U_0 \oplus U_1$. Hence, in both cases $U_k = 0$ for $|k| > 1$. It follows that the complex $C^\bullet_k$ is zero for any $k > 0$ and hence all eigenvalues of the Hessian of f are positive.

4.5 Reducible Higgs bundles

In this section we shall finally conclude the proof of Theorem 4.6 by showing that it also holds for reducible Higgs bundles. First we shall show that a reducible Higgs bundle which is not of the form given in Theorem 4.6 cannot be a local minimum of f; for this we use an argument similar to the one given by Hitchin [23, §8] for the case of $G = \mathrm{PSL}(n, \mathbb{R})$.

Let (E, Φ) be a strictly polystable $\mathrm{U}(p, q)$-Higgs bundle which is a local minimum of f. Since $f(E, \Phi)$ is the sum of the values of f on each of the stable direct summands (on the corresponding lower rank moduli space), it follows that each stable direct summand must be a local minimum in its moduli space and, therefore, a fixed point of the circle action. Hence (E, Φ) is itself fixed and thus (cf. Proposition 4.10)

$$E = \bigoplus F_\lambda,$$

where each F_λ is an $i\lambda$-eigenbundle for an infinitesimal trace-free $\mathrm{U}(p) \times \mathrm{U}(q)$-gauge transformation ψ. Moreover, if $\Phi_{|F_\lambda} \neq 0$, then its image is contained in $F_{\lambda+1} \otimes K$. In analogy with the case of stable $\mathrm{U}(p, q)$-Higgs bundles we write

$$\mathrm{End}\, E = \bigoplus U_\mu,$$

where U_μ is the $i\mu$-eigenbundle for the adjoint action of ψ. Let

$$U_\mu^+ = U_\mu \cap U^+,$$
$$U_\mu^- = U_\mu \cap U^-,$$

then we can define a complex of sheaves

$$C^\bullet_{>0} : \bigoplus_{\mu>0} U_\mu^+ \xrightarrow{\mathrm{ad}(\Phi)} \bigoplus_{\mu>1} U_\mu^- \otimes K. \tag{4.12}$$

In this language Hitchin's criterion [23, §8] for showing that a given fixed point is not a local minimum can be expressed as follows.

Lemma 4.19. *Let (E_t, Φ_t) be a 1-parameter family of polystable $\mathrm{U}(p,q)$-Higgs bundles such that (E_0, Φ_0) is a fixed point of the circle action. If the tangent vector $(\dot{E}, \dot{\Phi})$ at 0 is nontrivial and lies in the subspace*

$$\mathbb{H}^1(C^\bullet_{>0})$$

of the infinitesimal deformation space $\mathbb{H}^1(C^\bullet)$ of (E_0, Φ_0), then (E_0, Φ_0) is not a local minimum of f.

Proposition 4.20. *Let (E, Φ) be a reducible $\mathrm{U}(p,q)$-Higgs bundle. If $\beta \neq 0$ and $\gamma \neq 0$ then (E, Φ) is not a local minimum of f.*

Proof. As we noted above, each stable direct summand of (E, Φ) is a local minimum on its moduli space and therefore (by Proposition 4.17) it has $\beta = 0$ or $\gamma = 0$. Hence we can choose two stable direct summands $(E' = V' \oplus W', \Phi')$ and $(E'' = V'' \oplus W'', \Phi'')$ such that $\gamma' \neq 0$ and $\beta'' \neq 0$ and $\beta' = \gamma'' = 0$. It is clearly sufficient to show that $(E' \oplus E'', \Phi' \oplus \Phi'')$ is not a local minimum of f on the corresponding moduli space and we can therefore assume that $(E, \Phi) = (E' \oplus E'', \Phi' \oplus \Phi'')$ without loss of generality. We shall construct a family of deformations (E_t, Φ_t) of (E, Φ) satisfying the conditions of Lemma 4.19.

By Lemma 4.21 both $H^1(\mathrm{Hom}(W'', W'))$ and $H^1(\mathrm{Hom}(V', V''))$ are non-vanishing, so let $\eta \in H^1(\mathrm{Hom}(V', V''))$ and $\sigma \in H^1(\mathrm{Hom}(W'', W'))$ be nonzero. We can then define a deformation of (E, Φ) by using that η defines an extension

$$0 \longrightarrow V'' \longrightarrow V^\eta \longrightarrow V' \longrightarrow 0,$$

while σ defines an extension

$$0 \longrightarrow W' \longrightarrow W^\sigma \longrightarrow W'' \longrightarrow 0.$$

Let $E^{(\eta,\sigma)} = V^\eta \oplus W^\sigma$ and define $\Phi^{(\eta,\sigma)}$ by the compositions

$$b^{(\eta,\sigma)} \colon W^\sigma \longrightarrow W'' \xrightarrow{\beta''} V'' \to V^\eta,$$
$$c^{(\eta,\sigma)} \colon V^\eta \longrightarrow V' \xrightarrow{\gamma'} W' \longrightarrow W^\sigma.$$

Note that $(E^0, \Phi^0) = (E, \Phi)$ (the Higgs fields agree since $\beta' = \gamma'' = 0$). It is then easy to see that $(E^{\eta,\sigma}, \Phi^{\eta,\sigma})$ is stable: the essential point is that the destabilizing subbundles V' and W'' of (E, Φ) are not subbundles of the deformed Higgs bundle; we leave the details to the reader.

Now define the family $(E_t, \Phi_t) = (E^{(\eta t,\sigma t)}, \Phi^{(\eta t,\sigma t)})$. It is clear that the induced infinitesimal deformation of E is

$$\dot{E} = (\eta, \sigma) \in H^1(\mathrm{Hom}(V', V'')) \oplus H^1(\mathrm{Hom}(W'', W')) \subseteq H^1(\mathrm{End}(E)).$$

Considering the holomorphic structure as given by a $\overline{\partial}$-operator on the underlying smooth bundle, our definition of $(E^{(\eta,\sigma)}, \Phi^{(\eta,\sigma)})$ did not change the Higgs field but only the holomorphic structure on E. Thus, taking a Dolbeault representative (cf. Remark 3.23) for $(\dot{E}, \dot{\Phi}) \in \mathbb{H}^1(C^\bullet)$ we see that the weights of ψ on $(\dot{E}, \dot{\Phi})$ are given by its weights on $\dot{E}$. From Proposition 4.10 we have decompositions $E' = \bigoplus F'_k$ and $E'' = \bigoplus F''_k$ into eigenspaces of infinitesimal trace-free gauge transformations ψ' and ψ''. Note that the infinitesimal gauge transformation producing the decomposition of E is $\psi = \psi' + \psi''$. Clearly we have

$$\begin{aligned} F'_1 &= V', & F'_2 &= W', \\ F''_1 &= W'', & F''_2 &= V''. \end{aligned}$$

Let λ'_V and λ'_W be the weights of the action of ψ' on V' and W' respectively, and analogously for E''. We then have that

$$\lambda'_W = \lambda'_V + 1, \qquad \lambda''_V = \lambda''_W + 1.$$

and, since $\operatorname{tr}\psi' = \operatorname{tr}\psi'' = 0$,

$$\begin{aligned} \lambda'_V p' + \lambda'_W q' &= 0, \\ \lambda''_V p'' + \lambda''_W q'' &= 0, \end{aligned}$$

where $p' = \mathrm{rk}(V')$, $q' = \mathrm{rk}(W')$, $p'' = \mathrm{rk}(V'')$ and $q'' = \mathrm{rk}(W'')$. From these equations we conclude that

$$\begin{aligned} \lambda'_W - \lambda''_W &= \frac{p'}{p'+q'} + \frac{p''}{p''+q''} > 0, \\ \lambda''_V - \lambda'_V &= \frac{q''}{p''+q''} + \frac{q'}{p'+q'} > 0. \end{aligned}$$

It follows that the weights of ψ on $H^1(\mathrm{Hom}(W'', W'))$ and $H^1(\mathrm{Hom}(V', V''))$ are both positive and hence that $(\dot{E}, \dot{\Phi})$ lies in a direct sum of positive weight spaces of ψ. This concludes the proof of the proposition. q.e.d.

Lemma 4.21 *Let* $(E' = V' \oplus W', \Phi')$ *and* $(E'' = V'' \oplus W'', \Phi'')$ *be stable* $\mathrm{U}(p,q)$*-Higgs bundles of the same slope. Then the cohomology groups* $H^1(\mathrm{Hom}(V', V''))$ *and* $H^1(\mathrm{Hom}(W'', W'))$ *are both non-vanishing.*

Proof. Since $\gamma'' = 0$, V'' is a Φ-invariant subbundle of E''. Thus $\mu(V'') < \mu(E'')$. Using the Riemann–Roch formula and the equality $\mu(E'') = \mu(E')$ we obtain

$$\begin{aligned} h^0(\mathrm{Hom}(V', V'') - h^1(\mathrm{Hom}(V', V'') &= p'p''(1 - g + \mu(V'') - \mu(V')) \\ &< p'p''(1 - g + \mu(E') - \mu(V')). \end{aligned}$$

Since $\mathrm{rk}(\beta') \leqslant p'$ the inequality (3.25) of Corollary 3.26 shows that $\mu(E') - \mu(V') \leqslant g - 1$ and we therefore deduce that

$$h^0(\mathrm{Hom}(V', V'') - h^1(\mathrm{Hom}(V', V'') < 0,$$

from which it follows that $H^1(\mathrm{Hom}(V', V'') \neq 0$.

Similarly one sees that $H^1(\mathrm{Hom}(W'', W')) \neq 0$. q.e.d.

4.6 Local minima and connectedness

In this section we obtain connectedness results on $\mathcal{M}^s(a,b)$ and its closure $\overline{\mathcal{M}}^s(a,b)$. We denote by $\mathcal{N}^s(a,b) \subseteq \mathcal{N}(a,b)$ the subspace consisting of stable $\mathrm{U}(p,q)$-Higgs bundles, and denote its closure by $\overline{\mathcal{N}}^s(a,b)$.

The invariants (a,b) will be fixed in the following and we shall occasionally drop them from the notation and write $\mathcal{M} = \mathcal{M}(a,b)$, etc.

Proposition 4.22. *The closure of* $\mathcal{N}^s$ *in* $\mathcal{M}$ *coincides with* $\overline{\mathcal{N}}^s$ *and*

$$\overline{\mathcal{N}}^s = \overline{\mathcal{M}}^s \cap \mathcal{N}.$$

Proof. Clear. q.e.d.

Now consider the restriction of the Morse function to $\overline{\mathcal{M}}^s$,

$$f \colon \overline{\mathcal{M}}^s \to \mathbb{R}.$$

Proposition 4.23. *The restriction of f to $\overline{\mathcal{M}}^s$ is proper and the subspace of local minima of this function coincides with $\overline{\mathcal{N}}^s$.*

Proof. Properness of the restriction follows from properness of f and the fact that $\overline{\mathcal{M}}^s$ is closed in $\mathcal{M}$. By Proposition 4.5 f is constant on $\mathcal{N}$ and its value there is its global minimum on $\mathcal{M}$. Thus $\overline{\mathcal{N}}^s$ is contained in the subspace of local minima of f.

It remains to see that f has no other local minima on $\overline{\mathcal{M}}^s$. We already know that the subspace of local minima on $\mathcal{M}^s$ is $\mathcal{N}^s$. Thus, since $\mathcal{M}^s$ is open in $\overline{\mathcal{M}^s}$, there cannot be any additional local minima on $\mathcal{M}^s$. We need to prove therefore that there are no local minima in $(\overline{\mathcal{M}^s} \setminus \mathcal{M}^s) \setminus \overline{\mathcal{N}}^s$. So let (E, Φ) be a strictly poly-stable $\mathrm{U}(p, q)$-Higgs bundle representing a point in this space. From Proposition 4.22 we see that $\beta \neq 0$ and $\gamma \neq 0$. In the proof of Proposition 4.20 we constructed a family (E_t, Φ_t) of $\mathrm{U}(p, q)$-Higgs bundles such that $(E, \Phi) = (E_0, \Phi_0)$ and (E_t, Φ_t) is stable for $t \neq 0$. Furthermore we showed that the restriction of f to this family does not have a local minimum at (E_0, Φ_0). It follows that (E, Φ) is not a local minimum of f on $\overline{\mathcal{M}^s}$. q.e.d.

Proposition 4.24.

(1) *If $\mathcal{N}(a, b)$ is connected, then so is $\mathcal{M}(a, b)$.*

(2) *If $\mathcal{N}^s(a, b)$ is connected, then so is $\overline{\mathcal{M}}^s(a, b)$.*

Proof. (1) In view of Proposition 4.3, this follows from Theorem 4.6.

(2) If $\mathcal{N}^s(a, b)$ is connected, then so is its closure $\overline{\mathcal{N}}^s(a, b)$. But from Proposition 4.23, $\overline{\mathcal{N}}^s(a, b)$ is the subspace of local minima of the proper positive map $f \colon \overline{\mathcal{M}}^s(a, b) \to \mathbb{R}$. Hence the result follows from Proposition 4.2. q.e.d.

5. Local minima as holomorphic triples

The next step is to identify the spaces $\mathcal{N}(a, b)$ and $\mathcal{N}^s(a, b)$ as moduli spaces in their own right. By definition (cf. Definition 4.4), the Higgs bundles in $\mathcal{N}(a, b)$ all have $\beta = 0$ or $\gamma = 0$ in their Higgs fields. Suppose first that (E, Φ) is a $\mathrm{U}(p, q)$-Higgs bundle with $\gamma = 0$. Then (E, Φ) determines the triple $T = (E_1, E_2, \phi)$ where

$$\begin{aligned} E_1 &= V \otimes K, \\ E_2 &= W, \\ \phi &= \beta. \end{aligned}$$

Conversely, given two holomorphic bundles E_1, E_2 of rank p and q respectively, together with a bundle endomorphism $\Phi \in H^0(\mathrm{Hom}(E_2, E_1))$, we can use the above relations to define a $\mathrm{U}(p,q)$-Higgs bundle with $\gamma = 0$. Similarly, there is a bijective correspondence between $\mathrm{U}(p,q)$-Higgs bundles with $\beta = 0$ and holomorphic triples in which

$$\begin{aligned} E_1 &= W \otimes K, \\ E_2 &= V, \\ \phi &= \gamma. \end{aligned}$$

The triples (E_1, E_2, Φ) are examples of the holomorphic triples studied in [4] and [15].

5.1 Holomorphic triples

We briefly recall the relevant definitions, referring to [4] and [15] for details. A *holomorphic triple* on X, $T = (E_1, E_2, \phi)$, consists of two holomorphic vector bundles E_1 and E_2 on X and a holomorphic map $\phi\colon E_2 \to E_1$. Denoting the ranks E_1 and E_2 by n_1 and n_2, and their degrees by d_1 and d_2, we refer to $(\mathbf{n}, \mathbf{d}) = (n_1, n_2, d_1, d_2)$ as the *type* of the triple.

A homomorphism from $T' = (E_1', E_2', \phi')$ to $T = (E_1, E_2, \phi)$ is a commutative diagram

$$\begin{array}{ccc} E_2' & \xrightarrow{\phi'} & E_1' \\ \downarrow & & \downarrow \\ E_2 & \xrightarrow{\phi} & E_1. \end{array}$$

$T' = (E_1', E_2', \phi')$ is a subtriple of $T = (E_1, E_2, \phi)$ if the homomorphisms of sheaves $E_1' \to E_1$ and $E_2' \to E_2$ are injective.

For any $\alpha \in \mathbb{R}$ the α-*degree* and α-*slope* of T are defined to be

$$\begin{aligned} \deg_\alpha(T) &= \deg(E_1) + \deg(E_2) + \alpha\, \mathrm{rk}(E_2), \\ \mu_\alpha(T) &= \frac{\deg_\alpha(T)}{\mathrm{rk}(E_1) + \mathrm{rk}(E_2)} \\ &= \mu(E_1 \oplus E_2) + \alpha \frac{\mathrm{rk}(E_2)}{\mathrm{rk}(E_1) + \mathrm{rk}(E_2)}. \end{aligned}$$

The triple $T = (E_1, E_2, \phi)$ is α-*stable* if

$$\mu_\alpha(T') < \mu_\alpha(T) \tag{5.1}$$

for any proper sub-triple $T' = (E_1', E_2', \phi')$. Define *α-semistability* by replacing (5.1) with a weak inequality. A triple is called *α-polystable* if it is the direct sum of α-stable triples of the same α-slope. It is *strictly α-semistable (polystable)* if it is α-semistable (polystable) but not α-stable.

We denote the moduli space of isomorphism classes of *α-polystable* triples of type (n_1, n_2, d_1, d_2) by

$$\mathcal{N}_\alpha = \mathcal{N}_\alpha(\mathbf{n}, \mathbf{d}) = \mathcal{N}_\alpha(n_1, n_2, d_1, d_2). \tag{5.2}$$

Using Seshadri S-equivalence to define equivalence classes, this is the moduli space of equivalence classes of α-semistable triples. The isomorphism classes of *α-stable* triples form a subspace which we denoted by $\mathcal{N}_\alpha^s$.

Proposition 5.1 ([4, 15]). *The moduli space $\mathcal{N}_\alpha(n_1, n_2, d_1, d_2)$ is a complex analytic variety, which is projective when α is rational. A necessary condition for $\mathcal{N}_\alpha(n_1, n_2, d_1, d_2)$ to be nonempty is*

$$\begin{cases} 0 \leqslant \alpha_m \leqslant \alpha \leqslant \alpha_M & \text{if } n_1 \neq n_2 \\ 0 \leqslant \alpha_m \leqslant \alpha & \text{if } n_1 = n_2 \end{cases} \tag{5.3}$$

where

$$\alpha_m = \mu_1 - \mu_2, \tag{5.4}$$

$$\alpha_M = \left(1 + \frac{n_1 + n_2}{|n_1 - n_2|}\right)(\mu_1 - \mu_2) \tag{5.5}$$

and $\mu_1 = \frac{d_1}{n_1}$, $\mu_2 = \frac{d_2}{n_2}$.

Within the allowed range for α there is a discrete set of *critical values*. These are the values of α for which it is numerically possible to have a subtriple $T' = (E_1', E_2', \phi')$ such that $\mu(E_1' \oplus E_2') \neq \mu(E_1 \oplus E_2)$ but $\mu_\alpha(T') = \mu_\alpha(T')$. All other values of α are called *generic*. The critical values of α are precisely the values for α at which the stability properties of a triple can change, i.e., there can be triples which are strictly α-semistable, but either α'-stable or α'-unstable for $\alpha' \neq \alpha$.

Strict α-semistability can, in general, also occur at generic values for α, but only if there can be subtriples with $\mu(E_1' \oplus E_2') = \mu(E_1 \oplus E_2)$ and $\frac{n_2'}{n_1' + n_2'} = \frac{n_2}{n_1 + n_2}$. In this case the triple is strictly α-semistable for all values of α. We refer to this phenomenon as *α-independent semistability.* This cannot happen if $\mathrm{GCD}(n_2, n_1 + n_2, d_1 + d_2) = 1$.

5.2 Identification of $\mathcal{N}(a,b)$

The following result relates the stability conditions for holomorphic triples and that for $\mathrm{U}(p,q)$-Higgs bundles.

Proposition 5.2. *A $\mathrm{U}(p,q)$-Higgs bundle (E,Φ) with $\beta=0$ or $\gamma=0$ is (semi)stable if and only if the corresponding holomorphic triple is α-(semi)stable for $\alpha=2g-2$.*

Proof. Let $T=(E_1,E_2,\phi)$ be the triple corresponding to the Higgs bundle $(V\oplus W,\Phi)$. For definiteness we shall assume that $\gamma=0$ (of course, the same argument applies if $\beta=0$). Thus $E_1=V\otimes K$ and $E_2=W$ and, hence,

$$\deg(E_1)=\deg(V)+p(2g-2).$$

Since $p=\mathrm{rk}(E_1)$ and $q=\mathrm{rk}(E_2)$ it follows that

$$\mu_\alpha(T)=\mu(E)+\frac{p}{p+q}(2g-2)+\frac{q}{p+q}\alpha. \tag{5.6}$$

If we set $\alpha=2g-2$ we therefore have

$$\mu_\alpha(T)=\mu(E)+2g-2. \tag{5.7}$$

Clearly the correspondence between holomorphic triples and $\mathrm{U}(p,q)$-Higgs bundles gives a correspondence between sub-triples $T'=(E_1',E_2',\phi')$ and Φ-invariant subbundles of E which respect the decomposition $E=V\oplus W$ (i.e., subbundles $E'=V'\oplus W'$ with $V'\subseteq V$ and $W'\subseteq W$). Now, it follows from (5.7) that $\mu(E')<\mu(E)$ if and only if $\mu_\alpha(T')<\mu_\alpha(T)$ (and similarly for semistability), thus concluding the proof. q.e.d.

We thus have the following important characterization of the subspace of local minima of f on $\mathcal{M}(a,b)$.

Theorem 5.3. *Let $\mathcal{N}(a,b)$ be the subspace of local minima of f on $\mathcal{M}(a,b)$ and let τ be the Toledo invariant as defined in Definition 3.28.*

If $a/p\leqslant b/q$, or equivalently if $\tau\leqslant 0$, then $\mathcal{N}(a,b)$ can be identified with the moduli space of α-polystable triples of type $(p,q,a+p(2g-2),b)$, with $\alpha=2g-2$.

If $a/p\geqslant b/q$, or equivalently if $\tau\geqslant 0$, then $\mathcal{N}(a,b)$ can be identified with the moduli space of α-polystable triples of type $(q,p,b+q(2g-2),a)$, with $\alpha=2g-2$.

That is,

$$\mathcal{N}(a,b) \cong \begin{cases} \mathcal{N}_{2g-2}(p,q,a+p(2g-2),b) & \text{if } a/p \leqslant b/q \text{ (equivalently } \tau \leqslant 0) \\ \mathcal{N}_{2g-2}(q,p,b+q(2g-2),a) & \text{if } a/p \geqslant b/q \text{ (equivalently } \tau \geqslant 0). \end{cases}$$

Proof. This follows from Theorem 4.6, Proposition 4.8, and Proposition 5.2. q.e.d.

Thus, combining Proposition 4.24 and Theorem 5.3, we get:

Theorem 5.4.

(1) *Suppose $a/p \leqslant b/q$. If $\mathcal{N}_{2g-2}(p,q,a+p(2g-2),b)$ is connected then $\mathcal{M}(a,b)$ is connected. If $\mathcal{N}^s_{2g-2}(p,q,a+p(2g-2),b)$ is connected then $\overline{\mathcal{M}}^s(a,b)$ is connected.*

(2) *Suppose $a/p \geqslant b/q$. If $\mathcal{N}_{2g-2}(q,p,b+q(2g-2),a)$ is connected then $\mathcal{M}(a,b)$ is connected. If $\mathcal{N}^s_{2g-2}(q,p,b+q(2g-2),a)$ is connected then $\overline{\mathcal{M}}^s(a,b)$ is connected.*

5.3 The Toledo invariant, $2g-2$, and α-stability for triples

In view of Theorems 5.3 and 5.4, it is important to understand where $2g-2$ lies in relation to the range (given by Proposition 5.1) for the stability parameter α. Recall that for given (p,q,a,b), the Toledo invariant (Definition 3.28) is constrained by $0 \leqslant |\tau| \leqslant \tau_M$, where (see (3.30)) $\tau_M = \min\{p,q\}(2g-2)$. Recall also that α is constrained by the bounds given in Proposition 5.1. Whenever necessary we shal indicate the dependence of α_m and α_M on (p,q,a,b) by writing $\alpha_m = \alpha_m(p,q,a,b)$, and similarly for α_M.

Lemma 5.5. *Fix (p,q,a,b). Then*

$$\alpha_m(p,q,a,b) = (2g-2) - \frac{p+q}{2pq}|\tau| \tag{5.8}$$

where τ is the Toledo invariant. If $p \neq q$ then

$$\alpha_M(p,q,a,b) = \left(\frac{2\max\{p,q\}}{|p-q|} \right) \alpha_m(p,q,a,b). \tag{5.9}$$

If $p = q$ then $\alpha_M(p,q,a,b) = \infty$.

Proof. By Theorem 5.3 the type of the triple is determined by the sign of τ. The result thus follows by applying (5.3) and (5.4) to triples of type $(p, q, a + p(2g-2), b)$ (if $\tau \leqslant 0$) or type $(q, p, b + q(2g-2), a)$ (if $\tau \geqslant 0$). q.e.d.

Proposition 5.6. *Fix (p, q, a, b). Then*

$$0 \leqslant |\tau| \leqslant \tau_M \Leftrightarrow \begin{cases} 0 < \alpha_m(p,q,a,b) \leqslant 2g-2 \leqslant \alpha_M(p,q,a,b) & \text{if } p \neq q \\ 0 \leqslant \alpha_m(p,q,a,b) \leqslant 2g-2 & \text{if } p = q \end{cases} \tag{5.10}$$

Furthermore,

$$\tau = 0 \Leftrightarrow 2g-2 = \alpha_m \tag{5.11}$$

and

$$|\tau| = \tau_M \Leftrightarrow \begin{cases} 2g-2 = \alpha_M & \text{if } p \neq q \\ \alpha_m = 0 & \text{if } p = q. \end{cases} \tag{5.12}$$

Proof. Using (5.8) and (5.9) we see that $0 \leqslant |\tau| \leqslant \tau_M$ is equivalent to

$$2g-2 \geqslant \alpha_m \geqslant \left(\frac{|p-q|}{2\max\{p,q\}}\right)(2g-2), \tag{5.13}$$

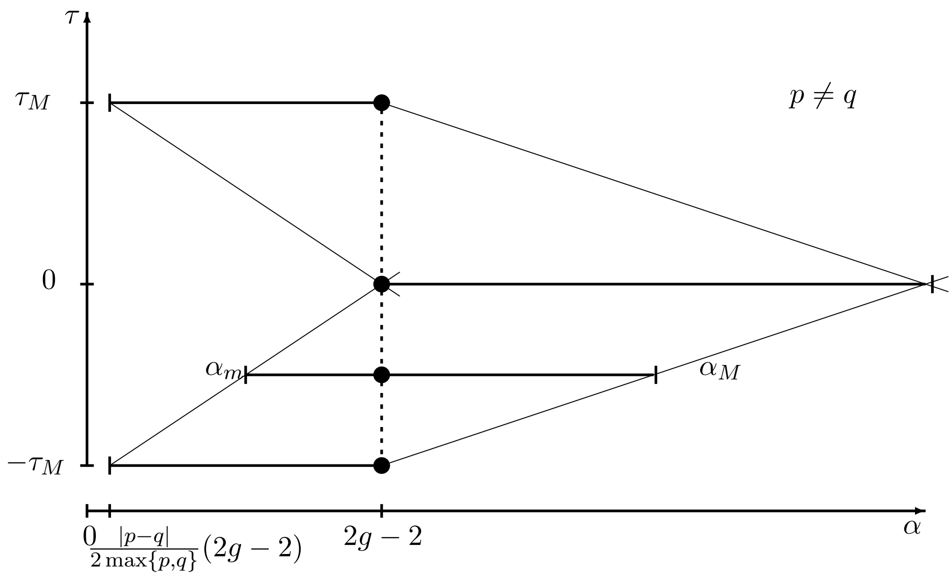
and hence also (assuming $p \neq q$) to

$$\left(\frac{2\max\{p,q\}}{|p-q|}\right)(2g-2) \geqslant \alpha_M \geqslant (2g-2). \tag{5.14}$$

In both (5.13) and (5.14), we get equality in the first place if and only if $\tau = 0$, and in the second place if and only if $|\tau| = \tau_M$. Notice that $\frac{|p-q|}{2\max\{p,q\}}$ is strictly positive if $p \neq q$ and is zero if $p = q$. The results follow. q.e.d.

These results are summarized in Figure 1, which can be used as follows. For any allowed value of τ, draw a horizontal line at height τ. The corresponding range for α and the relative location of $2g-2$ are then read off from the α-axis.

Remark 5.7. The above proposition gives another explanation for the Milnor–Wood inequality in Corollary 3.27. Using the fact that the



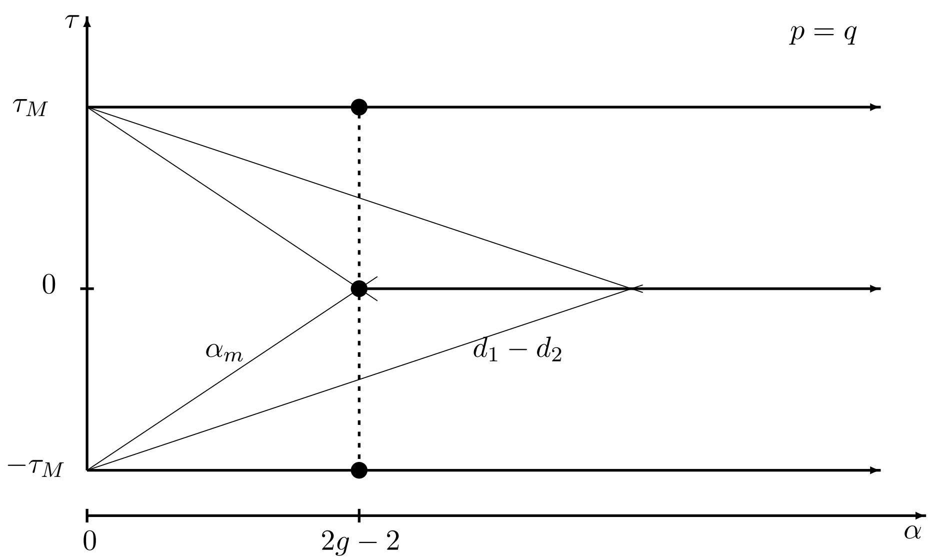


Figure 1: Range for the stability parameter α for triples in $\mathcal{N}(a, b)$, displayed as functions of $\tau = \frac{2pq}{p+q}(\frac{a}{p} - \frac{b}{q})$, and showing the relative location of $2g - 2$.

non-emptiness of $\mathcal{M}(a,b)$ is equivalent to the non-emptiness of $\mathcal{N}(a,b)$ and hence to that of either $\mathcal{N}_{2g-2}(p,q,a+p(2g-2),b)$ or $\mathcal{N}_{2g-2}(q,p,b+q(2g-2),a)$, we see that the Milnor–Wood inequality is equivalent to the condition that $2g-2$ lies within the range where α-polystable triples of the given kind exist.

5.4 Moduli spaces of triples

Proposition 5.6 shows that in order to study $\mathcal{N}(a,b)$ for different values of the Toledo invariant, we need to understand the moduli spaces of triples for values of α that may lie anywhere (including at the extremes α_m and α_M) in the α-range given in Proposition 5.1. The information we need can be found in [7]. From the results in [7] we get the following for triples of type (n_1,n_2,d_1,d_2).

Theorem 5.8 (Theorem A in [7])**.**

(1) *A triple $T=(E_1,E_2,\phi)$ of type (n_1,n_2,d_1,d_2) is α_m-polystable if and only if $\phi=0$ and E_1 and E_2 are polystable. We thus have*

$$\mathcal{N}_{\alpha_m}(n_1,n_2,d_1,d_2)\cong M(n_1,d_1)\times M(n_2,d_2).$$

where $M(n,d)$ denotes the moduli space of polystable bundles of rank n and degree d. In particular, $\mathcal{N}_{\alpha_m}(n_1,n_2,d_1,d_2)$ is nonempty and irreducible.

(2) *If $\alpha>\alpha_m$ is any value such that $2g-2\leqslant\alpha$ (and $\alpha<\alpha_M$ if $n_1\neq n_2$)) then $\mathcal{N}^s_\alpha(n_1,n_2,d_1,d_2)$ is nonempty and irreducible. Moreover:*

- *If $n_1=n_2=n$ then $\mathcal{N}^s_\alpha(n,n,d_1,d_2)$ is birationally equivalent to a $\mathbb{P}^N$-fibration over $M^s(n,d_2)\times\mathrm{Sym}^{d_1-d_2}(X)$, where $M^s(n,d_2)$ denotes the subspace of stable bundles of type (n,d_2), $\mathrm{Sym}^{d_1-d_2}(X)$ is the symmetric product, and the fiber dimension is $N=n(d_1-d_2)-1$.*

- *If $n_1>n_2$ then $\mathcal{N}^s_\alpha(n_1,n_2,d_1,d_2)$ is birationally equivalent to a $\mathbb{P}^N$-fibration over $M^s(n_1-n_2,d_1-d_2)\times M^s(n_2,d_2)$, where the fiber dimension is $N=n_2d_1-n_1d_2+n_2(n_1-n_2)(g-1)-1$. The birational equivalence is an isomorphism if $\mathrm{GCD}(n_1-n_2,d_1-d_2)=1$ and $\mathrm{GCD}(n_2,d_2)=1$.*

- *If $n_1<n_2$ then $\mathcal{N}^s_\alpha(n_1,n_2,d_1,d_2)$ is birationally equivalent to a $\mathbb{P}^N$-fibration over $M^s(n_2-n_1,d_2-d_1)\times M^s(n_1,d_1)$, where the*

fiber dimension is $N = n_2d_1 - n_1d_2 + n_1(n_2 - n_1)(g-1) - 1$*. The birational equivalence is an isomorphism if* $\mathrm{GCD}(n_2-n_1, d_2-d_1) = 1$ *and* $\mathrm{GCD}(n_1, d_1) = 1$*.*

In particular, if $n_1 \neq n_2$ *then* $\mathcal{N}^s_\alpha(n_1, n_2, d_1, d_2)$ *is a smooth manifold of dimension* $(g-1)(n_1^2 + n_2^2 - n_1n_2) - n_1d_2 + n_2d_1 + 1$*.*

(3) *If* $n_1 \neq n_2$ *then* $\mathcal{N}_{\alpha_M}(n_1, n_2, d_1, d_2)$ *is nonempty and irreducible. Moreover*

$$\mathcal{N}_{\alpha_M}(n_1, n_2, d_1, d_2) \cong \begin{cases} M(n_2, d_2) \times M(n_1 - n_2, d_1 - d_2) & \text{if } n_1 > n_2 \\ M(n_1, d_1) \times M(n_2 - n_1, d_2 - d_1) & \text{if } n_1 < n_2. \end{cases} \tag{5.15}$$

Theorem 5.9 (Corollary 8.2 and Theorem 8.10 in [7])**.** *If* $n_1 = n_2 = n$ *then:*

(1) *If* $\alpha_m = 0$*, i.e., if* $d_1 = d_2$ $(= d)$*, then* $\mathcal{N}_\alpha(n, n, d, d) \cong M(n, d)$ *for all* $\alpha > 0$*. In particular* $\mathcal{N}_\alpha(n, n, d, d)$ *is nonempty and irreducible.*

(2) *If* $0 < d_1 - d_2 < \alpha$*, then* $\mathcal{N}_\alpha(n, n, d_1, d_2)$ *is nonempty and irreducible.*

Remark 5.10. Notice that if $n_1 = n_2$ and $\alpha_m = 0$, then $\mathcal{N}_\alpha(n, n, d, d) \cong M(n, d)$ for all $\alpha > 0$, while $\mathcal{N}_0(n, n, d, d) \cong M(n, d) \times M(n, d)$. The picture is quite different if we restrict to the stable points in the moduli spaces. In fact there are no stable points in $\mathcal{N}_0(n, n, d, d)$, i.e., $\mathcal{N}^s_0(n, n, d, d)$ is empty, while $\mathcal{N}^s_\alpha(n, n, d, d) \cong M^s(n, d)$ for $\alpha > 0$.

Proposition 5.11 (Proposition 2.6 and Lemma 2.7 in [7])**.**

(1) *If* $\alpha \in [\alpha_m, \alpha_M]$ *is generic and* $\mathrm{GCD}(n_1, n_1 + n_2, d_1 + d_2) = 1$*, then*

$$\mathcal{N}_\alpha(n_1, n_2, d_1, d_2) = \mathcal{N}^s_\alpha(n_1, n_2, d_1, d_2).$$

In particular, the moduli space $\mathcal{N}_\alpha(n_1, n_2, d_1, d_2)$ *is nonempty and irreducible if in addition* $2g - 2 \leqslant \alpha$*.*

(2) *Let* $m \in \mathbb{Z}$ *be such that* $\mathrm{GCD}(n_1 + n_2, d_1 + d_2 - mn_1) = 1$*. Then* $\alpha = m$ *is not a critical value and there are no* α*-independent semistable triples.*

6. Main results

We now use the results of Section 5.4, applied to the case $\alpha = 2g-2$, to deduce our main results on the moduli spaces of $\mathrm{U}(p,q)$-Higgs bundles, and hence for the representation spaces $\mathcal{R}(\mathrm{PU}(p,q))$ and $\mathcal{R}_\Gamma(\mathrm{U}(p,q))$ (defined in Section 2). Recall that we identified components of $\mathcal{R}(\mathrm{PU}(p,q))$ labeled by $[a,b] \in \mathbb{Z}\oplus\mathbb{Z}/(p+q)\mathbb{Z}$, and similarly identified components of $\mathcal{R}_\Gamma(\mathrm{U}(p,q))$ labeled by $(a,b) \in \mathbb{Z}\oplus\mathbb{Z}$. Our arguments proceed along the following lines:

- By Proposition 2.5 $\mathcal{R}_\Gamma(a,b)$ is a $\mathrm{U}(1)^{2g}$-fibration over $\mathcal{R}[a,b]$. The number of connected components of $\mathcal{R}_\Gamma(a,b)$ is thus greater than or equal to that of $\mathcal{R}[a,b]$. In particular, $\mathcal{R}[a,b]$ is connected whenever $\mathcal{R}_\Gamma(a,b)$ is.
- By Proposition 3.13 there is a homeomorphism between $\mathcal{R}_\Gamma(a,b)$ and the moduli space $\mathcal{M}(a,b)$ of $\mathrm{U}(p,q)$-Higgs bundles. This restricts to give a homeomorphism between $\mathcal{R}^*_\Gamma(a,b)$ and $\mathcal{M}^s(a,b)$.
- By Proposition 4.3 the number of connected components of $\mathcal{M}(a,b)$ is bounded above by the number of connected components in the subspace of local minima for the Bott-Morse function defined in Section 4.1. By Proposition 4.24 the same conclusion holds for $\mathcal{M}^s(a,b)$.
- By Theorems 4.6 and 5.3 we can identify the subspace of local minima as a moduli space of α-stable triples, with $\alpha = 2g-2$.

Summarizing, we have:

$$\begin{aligned}|\pi_0(\mathcal{R}[a,b])| &\leqslant |\pi_0(\mathcal{R}_\Gamma(a,b))| = |\pi_0(\mathcal{M}(a,b))|\\ &\leqslant |\pi_0(\mathcal{N}(a,b))| = |\pi_0(\mathcal{N}_{2g-2}(n_1,n_2,d_1,d_2))|\end{aligned}$$

where $|\pi_0(\cdot)|$ denotes the number of components, and (in the notation of Section 5) the moduli space of triples which appears in the last line is either $\mathcal{N}_{2g-2}(p,q,a+p(2g-2),b)$ (if $a/p \leqslant b/q$) or $\mathcal{N}_{2g-2}(q,p,b+q(2g-2),a)$. Similarly, we get that

$$\begin{aligned}|\pi_0(\overline{\mathcal{R}}^*[a,b])| &\leqslant |\pi_0(\overline{\mathcal{R}}^*_\Gamma(a,b))| = |\pi_0(\overline{\mathcal{M}}^s(a,b))|\\ &\leqslant |\pi_0(\overline{\mathcal{N}}^s(a,b))| = |\pi_0(\overline{\mathcal{N}}^s_{2g-2}(n_1,n_2,d_1,d_2))|\end{aligned}$$

In particular, if the moduli spaces of triples are connected, then so are the Higgs moduli spaces and the moduli spaces of representations.

6.1 Moduli spaces of Higgs bundles

We begin with results for the $\mathrm{U}(p,q)$-Higgs moduli spaces. Recall from Proposition 3.20 that, whenever the moduli space $\mathcal{M}^s(a,b)$ of stable $\mathrm{U}(p,q)$-Higgs bundles with invariants (a,b) is nonempty, it is a smooth complex manifold of dimension $1+(p+q)^2(g-1)$. We shall refer to this dimension as the *expected dimension* in the following.

Theorem 6.1. *Let (p,q) be any pair of positive integers and let $(a,b)\in\mathbb{Z}\oplus\mathbb{Z}$ be such that $0\leqslant|\tau(a,b)|\leqslant\tau_M$.*

(1) *If either of the following sets of conditions apply, then the moduli space $\mathcal{M}^s(a,b)$ is a nonempty smooth manifold of the expected dimension, with connected closure $\overline{\mathcal{M}}^s(a,b)$:*

  (i) $0<|\tau(a,b)|<\tau_M$*,*

  (ii) $|\tau(a,b)|=\tau_M$ *and* $p=q$*.*

(2) *If any one of the following sets of conditions apply, then the moduli space $\mathcal{M}(a,b)$ is nonempty and connected:*

  (i) $\tau(a,b)=0$*,*

  (ii) $|\tau(a,b)|=\tau_M$ *and* $p\neq q$*,*

  (iii) $(p-1)(2g-2)<|\tau|\leqslant\tau_M=p(2g-2)$ *and* $p=q$*.*

Proof. (1) By Proposition 5.6 condition (i) implies that $\alpha_m<2g-2<\alpha_M$ for the triples corresponding to points in $\mathcal{N}(a,b)$. Thus Theorem 5.8(2) (together with Theorem 5.3) implies that $\mathcal{N}(a,b)$ is nonempty and connected. Similarly, condition (ii) implies that $\alpha_m=0$, and we can apply Theorem 5.9(1). The rest follows from Theorem 5.4.

(2) By Proposition 5.6, the conditions in (i) and (ii) are equivalent to $\alpha_m=2g-2$ and $\alpha_M=2g-2$ respectively. It follows by parts (1) and (3) of Theorem 5.8 (together with Theorem 5.3) that $\mathcal{N}(a,b)$ is nonempty and connected. The rest follows from Theorem 5.4.

For (iii), we use the fact that $|\tau|=|b-a|$ if $p=q$. The condition on $|\tau|$ is thus equivalent to $d_1-d_2<2g-2$ for the triples corresponding to points in $\mathcal{N}(a,b)$. The result thus follows by Theorem 5.9(2). q.e.d.

Remark 6.2. Combining (1) and (i)-(ii) of (2) in Theorem 6.1, we see that *the moduli space $\mathcal{M}(a,b)$ is nonempty for all (p,q,a,b) such that $0\leqslant|\tau|\leqslant\tau_M$.*

Remark 6.3. In Theorem 3.32 we gave a detailed description for $\mathcal{M}(a,b)$ in the case that $p \neq q$ and $|\tau(a,b)| = \tau_M$. The description was complete, provided that the space was nonempty. By the previous remark we can now remove this caveat.

In general, the stable locus $\mathcal{M}^s(a,b)$ is not the full moduli space and the full moduli space $\mathcal{M}(a,b)$ is not smooth. Singularities can occur at points representing strictly semistable objects, and these can also account for singularities in $\mathcal{N}(a,b)$, the space of local minima (as in Section 5). These types of singularities are prevented by a certain coprimality condition:

Proposition 6.4. *Suppose that* $\mathrm{GCD}(p+q, a+b) = 1$. *Then:*

(1) $\mathcal{M}(a,b)$ *is smooth.*

(2) $\alpha = 2g-2$ *is not a critical value for triples of type* $(p,q,a+p(2g-2),b)$ *or* $(q,p,b+q(2g-2),a)$.

(3) *The moduli spaces* $\mathcal{N}_{2g-2}(p,q,a+p(2g-2),b)$ *and* $\mathcal{N}_{2g-2}(q,p,b+q(2g-2),a)$ *are nonempty, smooth and irreducible.*

Proof.

(1) This is simply a re-statement of Proposition 3.12.

(2) Apply Proposition 5.11 (2) with (n_1,n_2,d_1,d_2) equal to $(p,q,a+p(2g-2),b)$ or $(q,p,b+q(2g-2),a)$ and $m = 2g-2$.

(3) Since $\mathrm{GCD}(p+q,a+b) = 1$ implies $\mathrm{GCD}(p,p+q,b+a+q(2g-2)) = 1$ (or $\mathrm{GCD}(q,p+q,b+a+p(2g-2)) = 1$), the result follows from (2) and Proposition 5.11 (1).

q.e.d.

Theorem 6.5. *Let* (p,q) *be any pair of positive integers and let* (a,b) *be such that* $0 \leqslant |\tau(a,b)| \leqslant \tau_M$. *Suppose also that* $\mathrm{GCD}(p+q, a+b) = 1$. *Then the moduli space* $\mathcal{M}(a,b)$ *is a (nonempty) smooth, connected manifold of the expected dimension.*

Proof. Combine Proposition 6.4 and Theorem 5.4. q.e.d.

Theorems 6.1 plus 6.5 are equivalent to Theorem A in the Introduction.

6.2 Moduli spaces of representations

Using Proposition 3.13 we can translate the results of Section 6.1 into results about the representation spaces $\mathcal{R}_\Gamma(a,b)$ and $\mathcal{R}^*_\Gamma(a,b)$ (for $\mathrm{U}(p,q)$ representations of the surface group Γ). We denote the closure of $\mathcal{R}^*_\Gamma(a,b)$ in $\mathcal{R}_\Gamma(a,b)$ by $\overline{\mathcal{R}}^*_\Gamma(a,b)$.

Theorem 6.6. *Let (p,q) be any pair of positive integers and let $(a,b) \in \mathbb{Z}\oplus\mathbb{Z}$ be such that $0 \leqslant |\tau(a,b)| \leqslant \tau_M$.*

(1) *The moduli space $\mathcal{R}_\Gamma(a,b)$ is nonempty.*

(2) *If either of the following sets of conditions apply, then the moduli space $\mathcal{R}^*_\Gamma(a,b)$ is a nonempty smooth manifold of the expected dimension, with connected closure $\overline{\mathcal{R}}^*_\Gamma(a,b)$ in $\mathcal{R}_\Gamma(a,b)$:*

   (i) $0 < |\tau(a,b)| < \tau_M$,

   (ii) $|\tau(a,b)| = \tau_M$ *and* $p = q$.

(3) *If any one of the following sets of conditions apply, then the moduli space $\mathcal{R}_\Gamma(a,b)$ is connected:*

   (i) $\tau(a,b) = 0$,

   (ii) $|\tau(a,b)| = \tau_M$ *and* $p \neq q$,

   (iii) $(p-1)(2g-2) < |\tau| \leqslant \tau_M = p(2g-2)$ *and* $p = q$,

   (iv) $\mathrm{GCD}(p+q, a+b) = 1$

(4) *If $\mathrm{GCD}(p+q, a+b) = 1$ then $\mathcal{R}_\Gamma(a,b)$ is a smooth manifold of the expected dimension.*

Proof. By Proposition 3.13, this follows from Theorem 6.1 and 6.5.
q.e.d.

Theorem 6.7. *Let (p,q) be any pair of positive integers such that $p \neq q$, and let (a,b) be such that $|\tau(a,b)| = \tau_M$. Then every representation in $\mathcal{R}_\Gamma(a,b)$ is reducible (i.e., $\mathcal{R}^*_\Gamma(a,b)$ is empty). If $p < q$, then every such representation decomposes as a direct sum of a semisimple representation of Γ in $\mathrm{U}(p,p)$ with maximal Toledo invariant and a semisimple representation in $\mathrm{U}(q-p)$. Thus, if $\tau = p(2g-2)$ then there is an isomorphism*

$$\mathcal{R}_\Gamma(p,q,a,b) \cong \mathcal{R}_\Gamma(p,p,a,a-p(2g-2)) \times \mathcal{R}_\Gamma(q-p, b-a+p(2g-2)),$$

where the notation $\mathcal{R}_\Gamma(p,q,a,b)$ indicates the moduli space of representations of Γ in $\mathrm{U}(p,q)$ with invariants (a,b), and $R_\Gamma(n,d)$ denotes the moduli space of degree d representations of Γ in $\mathrm{U}(n)$.

(A similar result holds if $p>q$, as well as if $\tau = -p(2g-2)$.)

Proof. Proposition 3.13 and Theorem 3.32. q.e.d.

As observed in Section 2.2 (cf. (2.7)), the spaces $\mathcal{R}(a) = \mathcal{R}_\Gamma(a,-a)$ can be identified with components of $\mathcal{R}(\mathrm{U}(p,q))$, i.e., with components of the moduli space for representations of $\pi_1 X$ in $\mathrm{U}(p,q)$. Applying Theorems 6.6 and 6.7, together with the observation that $\tau(a,-a) = 2a$ in the special case where $b=-a$, we thus obtain the following results for $\mathcal{R}(\mathrm{U}(p,q))$. Notice that the condition $\mathrm{GCD}(p+q,a+b)=1$ is never satisfied if $a+b=0$.

Theorem 6.8. *Let (p,q) be any pair of positive integers and let $a \in \mathbb{Z}\oplus\mathbb{Z}$ be such that $|a| \leqslant \min\{p,q\}(g-1)$.*

(1) *The moduli space $\mathcal{R}_\Gamma(a)$ is nonempty.*

(2) *If either of the following sets of conditions apply, then the moduli space $\mathcal{R}^*(a)$ is a nonempty, smooth manifold of the expected dimension, with connected closure $\overline{\mathcal{R}}^*(a)$ in $\mathcal{R}(a)$:*

   (i) $0 < |a| < \min\{p,q\}(g-1)$, *or*

   (ii) $|a| = p(g-1)$ *and* $p=q$.

(3) *If any one of the following sets of conditions apply, then the moduli space $\mathcal{R}(a)$ is connected:*

   (i) $a=0$,

   (ii) $|a| = \min\{p,q\}(g-1)$ *and* $p \neq q$,

   (iii) $(p-1)(g-1) < |a| \leqslant p(g-1)$ *and* $p=q$.

Theorem 6.9. *Let (p,q) be any pair of positive integers such that $p \neq q$. If $|a| = \min\{p,q\}(g-1)$ then $\mathcal{R}^*(a)$ is empty and every representation in $\mathcal{R}(a)$ is reducible. If $p<q$, then every such representation decomposes as a direct sum of a semisimple representation of Γ in $\mathrm{U}(p,p)$ with maximal Toledo invariant and a semisimple representation in $\mathrm{U}(q-p)$. Thus, if $a = p(g-1)$ then there is an isomorphism*

$$\mathcal{R}(a) \cong \mathcal{R}_\Gamma(p,p,a,a-p(2g-2)) \times \mathcal{R}_\Gamma(q-p,p(2g-2)),$$

where the notation $\mathcal{R}_\Gamma(p,q,a,b)$ *indicates the moduli space of representations of* Γ *in* $\mathrm{U}(p,q)$ *with invariants* (a,b)*, and* $R_\Gamma(n,d)$ *denotes the moduli space of degree* d *representations of* Γ *in* $\mathrm{U}(n)$*.*

(*A similar result holds if* $p>q$*, as well as if* $a=-p(g-1)$*.*)

From Theorem 6.6 and Proposition 2.5 we obtain the following theorem about the moduli spaces for $\mathrm{PU}(p,q)$ representations of $\pi_1 X$. Note that the closure $\overline{\mathcal{R}}^*[a,b]$ in $\mathcal{R}[a,b]$ is the image of $\overline{\mathcal{R}}^*_\Gamma(a,b)$ under the map of Proposition 2.5, hence these two spaces have the same number of connected components.

Theorem 6.10. *Let* (p,q) *be any pair of positive integers and let* $(a,b)\in\mathbb{Z}\oplus\mathbb{Z}$ *be such that* $0\leqslant|\tau(a,b)|\leqslant\tau_M$*.*

(1) *The moduli space* $\mathcal{R}[a,b]$ *is nonempty.*

(2) *If either of the following sets of conditions apply, then the moduli space* $\mathcal{R}^*[a,b]$ *is a nonempty smooth manifold of the expected dimension, with connected closure* $\overline{\mathcal{R}}^*[a,b]$ *in* $\mathcal{R}[a,b]$*:*

   (i) $0<|\tau(a,b)|<\tau_M$*, or*
   (ii) $|\tau(a,b)|=\tau_M$ *and* $p=q$*.*

(3) *If any one of the following sets of conditions apply, then the moduli space* $\mathcal{R}[a,b]$ *of all semi-simple representations is connected:*

   (i) $\tau(a,b)=0$*,*
   (ii) $|\tau(a,b)|=\tau_M$ *and* $p\neq q$*,*
   (iii) $(p-1)(2g-2)<|\tau|\leqslant\tau_M=p(2g-2)$ *and* $p=q$*,*
   (iv) $\mathrm{GCD}(p+q,a+b)=1$*.*

Theorem 6.11. *Let* (p,q) *be any pair of positive integers such that* $p\neq q$*, and let* (a,b) *be such that* $|\tau(a,b)|=\tau_M$*. Then* $\mathcal{R}^*[a,b]$ *is empty. If* $p<q$*, then every such representation reduces to a semisimple representation of* $\pi_1 X$ *in* $\mathrm{P}(\mathrm{U}(p,p)\times\mathrm{U}(q-p))$*, such that the* $\mathrm{PU}(p,p)$ *representation induced via projection on the first factor has maximal Toledo invariant.* (*A similar result holds if* $p>q$*.*)

Remark 6.12. As explained by Hitchin in [22, Section 5], the moduli space of irreducible representations in the adjoint form of a Lie group is liable to acquire singularities, because of the existence of stable vector bundles which are fixed under the action of tensoring by a finite order line bundle. For this reason we do not make any smoothness statements in Theorem 6.10.

6.3 Total number of components and coprimality conditions

We end with some elementary observations about the total number of components in the decomposition $\mathcal{R}(\mathrm{PU}(p,q)) = \bigcup_{(a,b)} \mathcal{R}[a,b]$, and about the number of such components for which the coprime condition $\mathrm{GCD}(p+q, a+b) = 1$ apply. We begin with the number of components.

By definition, $\tau(a,b)$ takes values in $\frac{2}{n}\mathbb{Z}$, where $n = p+q$.

Proposition 6.13. *Suppose that* $\mathrm{GCD}(p,q) = k$. *Then the map*

$$\begin{aligned} \tau\colon \mathbb{Z}\oplus\mathbb{Z}/(p,q)\mathbb{Z} &\longrightarrow \frac{2}{n}\mathbb{Z} \\ [a,b] &\longmapsto \frac{2}{n}(aq-bp) \end{aligned}$$

fits in an exact sequence

$$0 \longrightarrow \mathbb{Z}/k\mathbb{Z} \xrightarrow{\sigma} \mathbb{Z}\oplus\mathbb{Z}/(p,q)\mathbb{Z} \xrightarrow{\tau} \frac{2k}{n}\mathbb{Z} \longrightarrow 0$$

where the map σ *is* $[t] \mapsto [t\frac{p}{k}, t\frac{q}{k}]$. *In particular,* τ *is a* $k:1$ *map onto the subset* $\frac{2k}{n}\mathbb{Z} \subset \frac{2}{n}\mathbb{Z}$.

Proof. The map σ is clearly injective, and $\tau\circ\sigma = 0$. To see that $\ker(\tau) = \mathrm{im}(\sigma)$, observe that if $\tau[a,b] = 0$ then either $a = b = 0$ or $\frac{a}{b} = \frac{p}{q}$, i.e., $[a,b] = [t\frac{p}{k}, t\frac{q}{k}]$ for some $t \in \mathbb{Z}$. Finally, if $a_0q - b_0p = k$, then for any $l \in \mathbb{Z}$ we have $\tau[la_0, lb_0] = \frac{2k}{n}l$. Thus τ is surjective onto $\frac{2k}{n}\mathbb{Z}$. q.e.d.

Remark 6.14. Proposition 6.13 shows why[10] we must use $[a,b]$ rather than τ to label the components of $\mathcal{R}(\mathrm{PU}(p,q))$ or of $\mathcal{R}_\Gamma(\mathrm{U}(p,q))$.

Definition 6.15. Suppose that $\mathrm{GCD}(p,q) = k$. Define

$$\mathcal{C} = \tau^{-1}([-\tau_M, \tau_M] \cap \frac{2k}{n}\mathbb{Z}), \tag{6.1}$$

where τ is the map defined in Proposition 6.13.

The following is then an immediate corollary of Proposition 6.13.

[10] Unless p and q are coprime, in which case there is a bijective correspondence between $[a,b]$ and τ.

Corollary 6.16. *Suppose that* $\mathrm{GCD}(p,q) = k$ *and* $\mathcal{C}$ *is as above. Then* $\mathcal{C}$ *is precisely the set of all the points in* $\mathbb{Z} \oplus \mathbb{Z}/(p,q)\mathbb{Z}$ *which label components* $\mathcal{R}[a,b]$ *in* $\mathcal{R}(\mathrm{PU}(p,q))$*. The cardinality of* $\mathcal{C}$ *is*

$$\begin{aligned}|\mathcal{C}| &= 2n\min\{p,q\}(g-1)+k \\ &= |([-\tau_M,\tau_M] \cap \frac{2}{n}\mathbb{Z})| + \mathrm{GCD}(p,q) - 1.\end{aligned}$$

Proof. The first statement is a direct consequence of Proposition 6.13 and the bound on τ. Suppose for definiteness that $\min\{p,q\} = p$. Then since $\tau_M = 2\min\{p,q\}(g-1) = \frac{2k}{n}(n\frac{p}{k}(g-1)) \in \frac{2k}{n}\mathbb{Z}$, the number of points in $[-\tau_M,\tau_M] \cap \frac{2k}{n}\mathbb{Z}$ is $2n\frac{p}{k}(g-1)+1$. The second statement now follows from the fact that τ is a $k:1$ map. The proof is similar if $\min\{p,q\} = q$. q.e.d.

Finally, we examine the coprime condition $\mathrm{GCD}(p+q,a+b) = 1$. We regard p and q as fixed, but allow $[a,b]$ to vary. The coprime condition $\mathrm{GCD}(p+q,a+b) = 1$ can thus be satisfied on some components but not on others.

Definition 6.17. Fix p and q and let $\mathcal{C} \subset \mathbb{Z} \oplus \mathbb{Z}/(p+q)\mathbb{Z}$ be as in Definition 6.15. Define $\mathcal{C}_1$ to be the subset of classes $[a,b] \in \mathcal{C}$ for which the condition $\mathrm{GCD}(p+q,a+b) = 1$ is satisfied.

Proposition 6.18. *Fix* p *and* q *and let* $\mathcal{C}$ *and* $\mathcal{C}_1$ *be as above. Both* $\mathcal{C}_1$ *and its complement in* $\mathcal{C}$ *are nonempty.*

Proof. If $a = p$ and $b = q-1$ then $\mathrm{GCD}(p+q,a+b) = 1$. Also, $\tau(p,q-1) = \frac{2p}{p+q}$, which is in $[-\tau_M,\tau_M] \cap \frac{2k}{n}\mathbb{Z}$. Thus $[p,q-1]$ is in $\mathcal{C}_1$. It is similarly straightforward to see that $(a,b) = (0,0)$ defines an element in $\mathcal{C} - \mathcal{C}_1$, as does $(a,b) = (p,-p)$ if $p \leqslant q$ or $(a,b) = (q,-q)$ if $q \leqslant p$. q.e.d.

It seems somewhat complicated to go beyond this result and completely enumerate the elements in $\mathcal{C}_1$. The following result is, however, a step in that direction.

Definition 6.19. Let $\Omega \subset \mathbb{R} \oplus \mathbb{R}$ be the region depicted in Figure 2, i.e., the region bounded by (i) the ray $b = q$ and $a \leqslant p$, (ii) the ray $a = p$ and $b \leqslant q$, (iii) the ray $a = 0$ and $b \leqslant 0$, (iv) the ray $b = 0$ and $a \leqslant 0$, (v) the line $aq - bp = \frac{n}{2}\tau_M$, and (vi) the line $aq - bp = -\frac{n}{2}\tau_M$, and including all the boundary lines except the first two rays. Let $\Omega_{\mathbb{Z}}$ be the set of integer points in Ω, i.e. $\Omega_{\mathbb{Z}} = \Omega \bigcap \mathbb{Z} \oplus \mathbb{Z}$. We refer to Ω as the *fundamental region for* (p,q) (see Figure 2). Then $\Omega_{\mathbb{Z}}$ is the integer lattice inside the fundamental region.

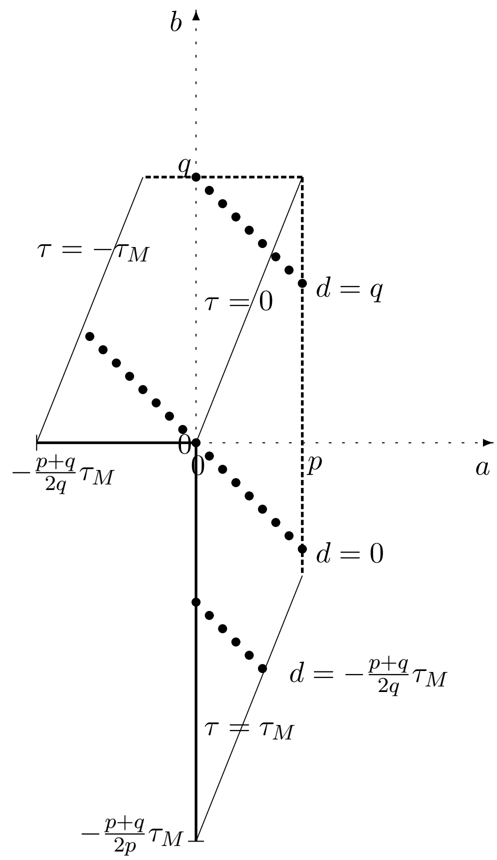


Figure 2: Fundamental region for (a, b). Components of $\mathcal{R}(\mathrm{PU}(p, q))$ correspond to the integer points in this region. Illustrative lines of constant τ (at $\tau = -\tau_M, 0, \tau_M$) and lines of constant d (at $d = -\frac{n}{2q}\tau_M, 0, q$) are shown.

Proposition 6.20. *Suppose that p and q are integers with* $\mathrm{GCD}(p,q) = k$.

(1) *There is a bijection between $\mathcal{C}$ and $\Omega_{\mathbb{Z}}$.*

(2) *If (a,b) lies in $\Omega_{\mathbb{Z}}$ then $d = a+b$ satisfies the bounds*

$$-n(g-1) \leqslant d < n. \tag{6.2}$$

All values of d in this range occur.

(3) *Let l_t denote the line $aq - bp = tk$. Then the points on $l_t \bigcap \Omega_{\mathbb{Z}}$ define the locus of points (a,b) for which $\tau(a,b) = t\frac{2k}{n}$.*

(4) *The line l_t intersects $\Omega_{\mathbb{Z}}$ for $-\frac{n}{2k}\tau_M \leqslant t \leqslant \frac{n}{2k}\tau_M$. For each integer t in this range, there are k points on $l_t \bigcap \Omega_{\mathbb{Z}}$.*

(5) *For a fixed t,* $\mathrm{GCD}(a+b, \frac{n}{k})$ *is the same for all integer points (a,b) on $l_t \bigcap \Omega_{\mathbb{Z}}$.*

(6) *Fix t and let (a,b) be any point in the set $l_t \bigcap \Omega_{\mathbb{Z}}$. If* $\mathrm{GCD}(a+b, \frac{n}{k}) \neq 1$ *then* $\mathrm{GCD}(a'+b', n) \neq 1$ *for all $(a',b') \in l_t \bigcap \Omega_{\mathbb{Z}}$.*

Proof. (1) Suppose first that $\frac{a}{p} \geqslant \frac{b}{q}$. Pick l such that $0 \leqslant a+lp \leqslant p$. Then $b+lq \leqslant q$, so that $(a+lp, b+lq)$ is in the fundamental region. Similarly, if $\frac{a}{p} \leqslant \frac{b}{q}$ then we pick l such that $0 \leqslant b+lq \leqslant q$ and see that $a+lp \leqslant p$. In this way we get a well-defined map from $\mathcal{C}$ to the fundamental region. The map is clearly injective. To see that it is surjective, notice that the boundary lines $aq - bp = \pm\frac{n}{2}\tau_M$ correspond to the conditions $\tau = \pm\tau_M$.

(2) This is clear from a sketch of the fundamental region (see Figure 2), in which the loci of points with constant value of $d = a+b$ are straight lines of slope -1. The maximal value for d corresponds to the line passing through the top right corner of the region, i.e., through (p,q). Thus $d_{max} = p+q = n$. The minimal value for d corresponds either to the line passing through $(-\frac{n}{2q}\tau_M, 0)$ or to the line through $(0, -\frac{n}{2p}\tau_M)$, depending on which yields the smaller value for d. Since $\tau_M = 2\min\{p,q\}(g-1)$, we find in all cases that $d_{min} = -n(g-1)$. It is straightforward to see that all intermediate values for d occur.

(3)-(4) This is simply a restatement of Proposition 6.13.

(5)-(6) Both follow from the fact that for any two points (a,b) and (a',b') on l_t, we get $d' = d + s\frac{n}{k}$ for some $s \in \mathbb{Z}$. q.e.d.

Remark 6.21. Part (6) says that for fixed t, if $\mathrm{GCD}(a+b,\frac{n}{k}) \neq 1$ for *any* point $(a,b) \in l_t \bigcap \Omega_{\mathbb{Z}}$, then $\mathrm{GCD}(a+b,n) \neq 1$ for *all* points $(a,b) \in l_t \bigcap \Omega_{\mathbb{Z}}$. That is, we can detect the non-coprimality of $(a+b,n)$ for all $(a,b) \in l_t \bigcap \Omega_{\mathbb{Z}}$ by checking the non-coprimality of $(a+b,\frac{n}{k})$ at any one $(a,b) \in l_t \bigcap \Omega_{\mathbb{Z}}$. We cannot however check for coprimality in the same way. If $\mathrm{GCD}(a+b,\frac{n}{k}) = 1$, it is possible that $\mathrm{GCD}(a'+b',n) \neq 1$ for some $(a',b') \in l_t \bigcap \Omega_{\mathbb{Z}}$. For example, take $p = 2, q = 4, a = -1, b = 0, a' = 0, b' = 2$, and $t = -2$. Then $\mathrm{GCD}(a'+b',n) = 2$ while $\mathrm{GCD}(a+b,\frac{n}{k}) = 1$.

References

[1] M.F. Atiyah & R. Bott, *The Yang-Mills equations over Riemann surfaces*, Philos. Trans. Roy. Soc. London Ser. A **308** (1983) 523–615, MR 85k:14006 Zbl 0509.14014.

[2] D. Banfield, *Stable pairs and principal bundles*, Q. J. Math. **5**1 (2000) 417–436, MR 2001k:53036, Zbl 0979.53028.

[3] I. Biswas & S. Ramanan, *An infinitesimal study of the moduli of Hitchin pairs*, J. London Math. Soc. (2) **49** (1994) 219–231, MR 94k:14006, Zbl 0819.58007.

[4] S.B. Bradlow & O. García-Prada, *Stable triples, equivariant bundles and dimensional reduction*, Math. Ann. **304** (1996) 225–252, MR 96m:32033, Zbl 0852.32016.

[5] S.B. Bradlow, O. García-Prada & P.B. Gothen, *Representations of the fundamental group of a surface in* $\mathrm{PU}(p,q)$ *and holomorphic triples*, C.R. Acad. Sci. Paris Sér. I Math. **333** (2001) 347–352, MR 2002q:14013, Zbl 0991.32012.

[6] S.B. Bradlow, O. García-Prada & P.B. Gothen, *Surface group representations, Higgs bundles and holomorphic triples*, preprint, math.AG/0206012.

[7] S.B. Bradlow, O. García-Prada & P.B. Gothen, *Moduli spaces of holomorphic triples over compact Riemann surfaces*, math.AG/0211428, to appear in Math. Ann.

[8] S.B. Bradlow, O. García-Prada & I. Mundet i Rierra, *Relative Hitchin–Kobayashi correspondences for principal pairs*, Quart. J. Math. **54** (2003) 171-208.

[9] M. Burger & A. Iozzi, *Bounded Kahler class rigidity of actions on Hermitian symmetric spaces*, preprint, 2002.

[10] K. Corlette, *Flat G-bundles with canonical metrics*, J. Differential Geom. **28** (1988) 361–382, MR 89k:58066, Zbl 0676.58007.

[11] A. Domic & D. Toledo, *The Gromov norm of the Kaehler class of symmetric domains*, Math. Ann. **276** (1987) 425–432, MR 88e:32057, Zbl 0595.53061.

[12] S.K. Donaldson, *Twisted harmonic maps and the self-duality equations*, Proc. London Math. Soc. (3) **55** (1987) 127–131, MR 88q:58040, Zbl 0634.53046.

[13] J.L. Dupont, *Bounds for characteristic numbers of flat bundles* in 'Algebraic topology', J.L. Dupont & I. Madsen (eds.), Aarhus 1978, Lecture Notes in Mathematics, **763**, 109–119, Berlin, Springer, 1978, MR 81f:57021, Zbl 0511.57018.

[14] T. Frankel, *Fixed points and torsion on Kähler manifolds*, Ann. of Math. (2) **70** (1959) 1–8, MR 24 #A1730, Zbl 0088.38002

[15] O. García-Prada, *Dimensional reduction of stable bundles, vortices and stable pairs*, Int. J. Math. **5** (1994) 1–52, MR 95d:32035, Zbl 0799.32022.

[16] W.M. Goldman, *Representations of fundamental groups of surfaces*, in 'Geometry and topology', Proceedings, J. Alexander & J. Harer (eds.), University of Maryland 1983–1984, Lecture Notes in Mathematics, **1167**, 95–117. Berlin-Heidelberg-New York, Springer, 1985, MR 87j:32068, Zbl 0575.57027.

[17] W.M. Goldman, *Topological components of spaces of representations*, Invent. Math. **93** (1988) 557–607, MR 89m:57001, Zbl 0655.57019.

[18] P.B. Gothen, *The topology of Higgs bundle moduli spaces*, Ph.D. Thesis, Mathematics Institute, University of Warwick, 1995.

[19] P.B. Gothen, *Components of spaces of representations and stable triples*, Topology **40** (2001) 823–850, MR 2002k:14017.

[20] P.B. Gothen, *Topology of $U(2,1)$ representation spaces*, Bull. London Math. Soc. **34** (2002) 729–738, MR 2003f:14036.

[21] L. Hernández, *Maximal representations of surface groups in bounded symmetric domains*, Transactions Amer. Math. Soc. **324** (1991) 405–420, MR 91f:32040, Zbl 0733.32024.

[22] N.J. Hitchin, *The self-duality equations on a Riemann surface*, Proc. London Math. Soc. **55** (1987) 59–126, MR 89a:32021, Zbl 0634.53045.

[23] N.J. Hitchin, *Lie groups and Teichmüller space*, Topology **31** (1992) 449–473, MR 93e:32023, Zbl 0769.32008.

[24] E. Markman & E.Z. Xia, *The moduli of flat* $\mathrm{PU}(p,p)$ *structures with large Toledo invariants*, Math. Z. **240** (2002) 95–109, MR 2003i:14009, Zbl 1008.32006.

[25] I. Mundet i Riera, *A Hitchin–Kobayashi correspondence for Kahler fibrations*, J. Reine Angew. Math. **528** (2000) 41–80, MR 2002b:53035, Zbl 1002.53057.

[26] M.S. Narasimhan & C.S. Seshadri, *Stable and unitary bundles on a compact Riemann surface*, Ann. of Math., **82**, 540–564 (1965), MR 30 #588, Zbl 0171.04803.

[27] N. Nitsure, *Moduli spaces of semistable pairs on a curve*, Proc. London Math. Soc. **62** (1991) 275–300, MR 92a:14010, Zbl 0733.14005.

[28] M. Schlessinger, *Functors on Artin rings*, Trans. Amer. Math. Soc. **130** (1968) 208–222, MR 36 #184, Zbl 0167.49503.

[29] C.T. Simpson, *Constructing variations of Hodge structure using Yang-Mills theory and applications to uniformization*, J. Amer. Math. Soc. **1** (1988) 867–918, MR 90e:58026, Zbl 0669.58008.

[30] C.T. Simpson, *Higgs bundles and local systems*, Inst. Hautes Études Sci. Publ. Math. **75** (1992) 5–95, MR 94d:32027, Zbl 0814.32003.

[31] C.T. Simpson, *Moduli of representations of the fundamental group of a smooth projective variety* I, Inst. Hautes Études Sci. Publ. Math. **79** (1994) 47–129 MR 96e:14012, Zbl 0891.14005.

[32] C.T. Simpson, *Moduli of representations of the fundamental group of a smooth projective variety* II, Inst. Hautes Études Sci. Publ. Math. **80** (1994) 5–79, MR 96e:14013, Zbl 0891.14006.

[33] D. Toledo, *Representations of surface groups in complex hyperbolic space*, J. Differential Geometry **29** (1989) 125–133, MR 90a:57016, Zbl 0676.57012.

[34] E.Z. Xia, *Components of* $\mathrm{Hom}(\pi_1, \mathrm{PGL}(2, \mathbf{R}))$, Topology **36** (1997) 481–499, MR 97j:57002, Zbl 0872.57004.

[35] E.Z. Xia, *The moduli of flat* $\mathrm{U}(p, 1)$ *structures on Riemann surfaces*, preprint, 1999, math.AG/9910037.

[36] E.Z. Xia, *The moduli of flat* $\mathrm{PU}(2, 1)$ *structures on Riemann surfaces*, Pacific Journal of Mathematics **195** (2000) 231–256, MR 2001q:32033.

Department of Mathematics
University of Illinois
Urbana, IL 61801

Instituto de Matemáticas y Física Fundamental
Consejo Superior de Investigaciones Científicas
28006 Madrid, Spain

Departamento de Matemática Pura
Universidade do Porto,
4169-007 Porto, Portugal