

Nonabelian Hodge theory and vector valued modular forms

Cameron Franc and Steven Rayan

ABSTRACT. We examine the relationship between nonabelian Hodge theory for Riemann surfaces and the theory of vector valued modular forms. In particular, we explain how one might use this relationship to prove a conjectural three-term inequality on the weights of free bases of vector valued modular forms associated to complex, finite dimensional, irreducible representations of the modular group. This conjecture is known for irreducible unitary representations and for all irreducible representations of dimension at most 12. We prove new instances of the three-term inequality for certain nonunitary representations, corresponding to a class of maximally-decomposed variations of Hodge structure, by considering the same inequality with respect to a new type of modular form, called a “Higgs form”, that arises naturally on the Dolbeault side of nonabelian Hodge theory. The paper concludes with a discussion of a strategy for reducing the general case of nilpotent Higgs bundles to the case under consideration in our main theorem.

CONTENTS

1. Introduction	1
2. Filtered objects	4
3. Vector valued modular forms	9
4. Nonabelian Hodge theory	10
5. New cases of the three-term inequality for $\mathrm{PSL}_2(\mathbf{Z})$	18
References	21

1. Introduction

Suppose that Y is a hyperbolic orbifold curve uniformized by the complex upper half plane \mathcal{H} , so that Y can be identified with a quotient $Y = \Gamma \backslash \mathcal{H}$ of \mathcal{H} under the action of a discrete group of linear fractional transformations Γ . Assume that Γ is a Fuchsian group of the first kind, so that Y has finite volume and a finite number of cusps. Vector bundles \mathcal{V} on Y can be pulled back to \mathcal{H} and trivialized there — this trivialization can even be achieved holomorphically for holomorphic bundles, by the Stein property of \mathcal{H} . Thus, by uniformization, all vector bundles on Y can be described

Key words and phrases. Nonabelian Hodge correspondence; modular form; filtered flat bundle; Higgs bundle; Higgs form; tame harmonic bundle; three-term inequality; stability.

as trivial bundles $\mathcal{H} \times \mathbf{C}^n$ endowed with an action of the fundamental group Γ :

$$\gamma(\tau, v) = (\gamma\tau, c(\gamma, \tau)v).$$

The function $c: \Gamma \times \mathcal{H} \rightarrow \mathrm{GL}_n(\mathbf{C})$ appearing above satisfies the cocycle identity $c(\alpha\beta, \tau) = c(\alpha, \beta\tau)c(\beta, \tau)$. The bundle-sheaf dichotomy says that giving the data of the bundle \mathcal{V} is equivalent to giving the data of its sheaf of local sections, which are vector valued functions on open subsets of \mathcal{H} satisfying the modular transformation law

$$F(\gamma\tau) = c(\gamma, \tau)F(\tau).$$

Thus, in this sense, the study of vector bundles on Y is equivalent to the study of generalized automorphic functions for the fundamental group Γ .

In the last several decades geometers have been exploring the geometry of Higgs bundles and their relationship with Hodge theory, beginning with [29, 47, 48]. Higgs bundles originated in mathematical investigations into gauge theory, specifically the self-dual Yang-Mills equations on a Riemann surface [29], now known as the *Hitchin equations*. The rich geometry inherited by moduli spaces of Higgs bundles has been exploited in a variety of fields: surface group theory (see [38] for a survey); integrable systems (beginning with [30]); mirror symmetry ([26, 31, 11] for instance); and topological recursion and free probability ([16]), to name a few areas. Most spectacularly, Higgs bundles lie at the heart of Ngô's proof of the Fundamental Lemma [44] from the Langlands Program, which is perhaps the first glimpse into a truly deep relationship between Higgs bundles and number theory.

The goal of the present paper is to discuss applications of ideas of nonabelian Hodge theory to a question concerning the weights of basis elements of spaces of (weakly holomorphic) vector valued modular forms for the modular group $\Gamma(1) := \mathrm{PSL}_2(\mathbf{Z})$. To be more precise, let $Y = \Gamma(1) \backslash \mathcal{H}$ and let X stand for its compactification by cusps. Then X can be identified with the weighted projective line $\mathbf{P}(2, 3)$, and the splitting principle holds for holomorphic vector bundles on X . If $\rho: \Gamma(1) \rightarrow \mathrm{GL}_n(\mathbf{C})$ denotes a representation of the modular group, then holomorphic modular functions for ρ are global sections of a logarithmic extension $\mathcal{V}(\rho)$ to the cusp of the natural flat connection on Y associated to ρ . If $\mathcal{V}_{2l}(\rho) := \mathcal{V}(\rho) \otimes \mathcal{O}(l)$, then global sections of $\mathcal{V}_{2l}(\rho)$ are modular forms of weight $k = 2l$. Let $M_k(\rho)$ denote the space of global sections of $\mathcal{V}_k(\rho)$, and define the space of *holomorphic vector valued modular forms* for ρ to be

$$M(\rho) := \bigoplus_{k \in 2\mathbf{Z}} M_k(\rho).$$

The splitting principle for holomorphic bundles on X implies that $M(\rho)$ is a free module over the ring of classical scalar valued modular forms of level one, which is a polynomial ring in two generators E_4 and E_6 in degrees 2 and 3, respectively (corresponding to Eisenstein series of weights 4 and 6) isomorphic with the projective coordinate ring of $\mathbf{P}(2, 3)$. Moreover, if $\mathcal{V}(\rho)$ decomposes as

$$\mathcal{V}(\rho) \cong \bigoplus_{j=1}^n \mathcal{O}(r_j),$$

then $M(\rho)$ has a free-basis of modular forms of weights $k_j = -2r_j$.

Computing the roots r_j above, or equivalently, the weights k_j , is a difficult problem in general. When ρ is unitarizable, one can use positivity and Serre duality to give a formula for the roots, but the nonunitarizable case is much more difficult. Note that, if one works with $\mathrm{SL}_2(\mathbf{Z})$ instead of $\mathrm{PSL}_2(\mathbf{Z})$, one encounters difficulties even

for representations of finite image: in that case, forms of weight one are mapped to themselves under Serre duality, and the geometric argument collapses. Due to this, even to this day the dimensions of spaces of classical holomorphic congruence modular forms of weight one are poorly understood. See [15] and [41] for some asymptotic results.

In [18] it was observed that for irreducible ρ , the weights of a free basis for $M(\rho)$ tend to satisfy a three-term inequality. To state this precisely, let m_k denote the multiplicity of forms of (even) weight k among a free basis for $M(\rho)$ over the ring of scalar valued modular forms. Equivalently, it is the multiplicity of $\mathcal{O}(-\frac{1}{2}k)$ appearing in the decomposition of $\mathcal{V}(\rho)$.

CONJECTURE 1.1. *For irreducible ρ , the three-term inequality*

$$m_k \leq m_{k+2} + m_{k-2}$$

is satisfied for all $k \in \mathbf{Z}$.

Conjecture 1.1 was proved in [18] for irreducible unitary representations of $\mathrm{SL}_2(\mathbf{Z})$ (not just $\mathrm{PSL}_2(\mathbf{Z})$) using positivity, and it was further established for *all* irreducible representations satisfying $\dim \rho \leq 12$ via computation. While Conjecture 1.1 sheds no new light on the question of modular forms of weight one (in that case, for unitary representations of $\mathrm{SL}_2(\mathbf{Z})$, Conjecture 1.1 says simply that $\dim M_1(\rho) \leq \dim M_3(\rho)$, since $\dim M_{-1}(\rho) = 0$ for unitary ρ), it is a surprising restriction on the sorts of bases that can occur for spaces of modular forms.

It has been known since the work of Narasimhan-Sheshadri [43] in the case of compact curves, and Mehta-Seshadri [40] for punctured curves, that irreducible unitary representations of a Fuchsian group Γ correspond to holomorphic bundles on the compactified quotient X with parabolic structures at cusps. A few decades later, Hitchin [29], Donaldson [14] and Simpson [47, 48] discovered that if one wishes to consider all irreducible representations — not just the unitary ones — then the right category to look at is the category of *filtered* or *parabolic Higgs bundles* on X . Such an object consists of a filtered, or parabolic, bundle E on X paired with an \mathcal{O}_X -linear endomorphism

$$\theta: E \rightarrow E \otimes \Omega_X^1(S)$$

where $S = X \setminus Y$ denotes the (possibly empty) set of cusps, and θ respects the filtered structure on E . For higher dimensional spaces there is an additional flatness condition, but for curves this condition is automatic.

The goal of this paper is to explain, in automorphic terms, the constructions of nonabelian Hodge theory that allow one to pass between representations and filtered Higgs bundles, and to apply them to the study of Conjecture 1.1. In Section 2 we describe the various filtered objects that arise in the statement of nonabelian Hodge theory, keeping in mind the automorphic description of these objects. Section 3 recalls some facts on vector valued modular forms and gives a precise discussion of Conjecture 1.1. Section 4 next describes the nonabelian Hodge correspondence for noncompact curves, following [48]¹, and we introduce the natural space of *Higgs modular forms* $\mathrm{Higgs}(\rho)$ that is the output of nonabelian Hodge theory. When $\Gamma = \mathrm{PSL}_2(\mathbf{Z})$ we use the splitting principle to show that $\mathrm{Higgs}(\rho)$ is a free module of rank $\dim \rho$ over the classical ring of scalar valued modular forms for Γ , analogously to the module $M(\rho)$

¹Although [48] only treats the case of curves, and not orbifolds, Simpson's arguments all take place on the universal cover, and so they generalize readily to the case of orbifolds required in this paper. See also [?], particularly Section 9

discussed above. In the final Section 5 we prove Conjecture 1.1 for stable filtered Higgs bundles corresponding to certain variations of Hodge structure:

THEOREM 1.2. *Let (E, θ) be a stable filtered Higgs bundle corresponding to a variation of Hodge structure, such that θ is nilpotent and decomposes E into a sequence of line bundles L_i for which θ shifts the sequence in the following way:*

$$\theta : L_i \rightarrow L_{i-1} \otimes \Omega_X^1(S).$$

Then, Conjecture 1.1 holds for E .

Note that the representations that correspond to (E, θ) as in Theorem 1.2 are not unitary, and so Theorem 1.2 represents a new class of cases for which the three-term inequality is known. The rest of Section 5 discusses ideas for using the geometry of the moduli space of Higgs bundles to extend this argument to the general case.

In the final section of the paper, we speculate on how Morse-theoretic techniques associated with the topology of the Higgs bundle moduli space, namely a natural \mathbf{C}^* -action and its flows, can be used to extend Theorem 1.2 to nilpotent filtered Higgs bundles that do not arise from complex variations of Hodge structure. This program rests on the observation that certain directions of the downward Morse-Bott flow are reorganizations of the line bundles L_i in the statement of the theorem.

This note reports on a talk given at a conference in honour of Geoff Mason that took place at Sacramento State University in 2018, and we thank the organizers for the opportunity to speak on this work. This is fitting, given that Knopp and Mason initiated the general study of vector valued modular forms for any representation of $\mathrm{SL}_2(\mathbf{Z})$ in the sequence of papers [32, 33, 34, 35, 36]. The original approach of Knopp-Mason was rooted in classical techniques of analytic number theory and motivated by questions arising in the theory of vertex operator algebras. This article aims to show that new approaches to these problems can arise by appealing to the correspondence between vector valued modular forms and Higgs modular forms. It is our general hope that this article will bring together geometers, both differential and algebraic, and automorphic representation theorists in the pursuit of new techniques and tools for solving open problems coming from either community.

Acknowledgements: Both authors are grateful for the support provided by NSERC Discovery Grants during the preparation of this manuscript.

2. Filtered objects

Let Y be an orbifold curve uniformized by the complex upper half plane \mathcal{H} via a map $j: \mathcal{H} \rightarrow Y$. Assume that the fundamental group $\Gamma \subseteq \mathrm{PSL}_2(\mathbf{R})$ of this cover is a Fuchsian group of the first kind, so that the compactification $X = Y \cup S$ is obtained by adding a finite, possibly empty, set of cusps S to Y .

2.1. Filtered representations. For each cusp $s \in S$, we let $\Gamma_s \subseteq \Gamma$ denote the stabilizer subgroup of s .

DEFINITION 2.1. A *filtered representation* of Γ consists of a finite dimensional complex representation $\rho: \Gamma \rightarrow \mathrm{GL}(V)$ and filtrations $F_{\beta,s}V$ for each $s \in S$ indexed by real numbers β such that the following axioms are satisfied:

- (1) (decreasing) $F_{\alpha,s}V \subseteq F_{\beta,s}V$ for all $\alpha \geq \beta$;
- (2) (exhaustive) $\cup_{\beta} F_{\beta,s}V = V$ and $\cap_{\beta} F_{\beta,s}V = 0$ for all $s \in S$;
- (3) (left continuous) for all $\beta \in \mathbf{R}$, there exists $\varepsilon > 0$ such that $F_{\beta-\varepsilon,s}V = F_{\beta,s}V$;
- (4) (monodromy invariance) $\Gamma_s \cdot F_{\beta,s}V \subseteq F_{\beta,s}V$ for all $\beta \in \mathbf{R}$.

REMARK 2.2. In the literature on moduli of parabolic bundles, both decreasing and increasing filtrations appear. This is a matter of convention and we have opted for decreasing filtrations in order to match Simpson's choice in [47].

Note that since V is finite dimensional, condition (2) implies that there exists $A \leq B$ such that $F_{A,s}V = V$ and $F_{B,s}V = 0$. If (ρ, F) is a filtered representation of Γ on a vector space V , then for each $s \in S$ define

$$\begin{aligned} \text{Gr}_{\beta,s}(\rho, F) &:= \frac{F_{\beta,s}V}{\bigcup_{\alpha > \beta} F_{\alpha,s}V}, \\ \text{Gr}_s(\rho, F) &:= \bigoplus_{\beta \in \mathbf{R}} \text{Gr}_{\beta,s}(\rho, F). \end{aligned}$$

We call a real number β a *jump* of (ρ, F) provided that $\text{Gr}_{\beta,s}(\rho, F) \neq 0$. A jump is characterized by the property that $F_{\beta,s}V \neq F_{\beta+\varepsilon,s}V$ for all $\varepsilon > 0$.

DEFINITION 2.3. The *degree* of a filtered representation (ρ, F) is defined as

$$\deg(\rho, F) := \sum_{s \in S} \sum_{\beta \in \mathbf{R}} \beta \dim_{\mathbf{C}}(\text{Gr}_{\beta,s}(\rho, F)).$$

The *slope* of a filtered representation (ρ, F) is the quantity $\mu(\rho, F) = \deg(\rho, F) / \dim_{\mathbf{C}} V$.

DEFINITION 2.4. A filtered representation (ρ, F) is said to be *semistable* provided that for every subobject $(\rho', F') \subseteq (\rho, F)$, the inequality $\mu(\rho', F') \leq \mu(\rho, F)$ holds. If the inequality holds strictly whenever (ρ', F') is nontrivial, then (ρ, F) is said to be *stable*.

EXAMPLE 2.5. Let (ρ, F) be a filtered representation with $\dim_{\mathbf{C}} \rho = 1$. For each $s \in S$, there is one jump $\beta \in \mathbf{R}$ and $\dim \text{Gr}_{\beta,s}(\rho, F) = 1$. Hence $\deg(\rho, F) = \mu(\rho, F)$ is the sum of the jumps. In this case (ρ, F) is stable since there are no nontrivial sub-objects to test against the stability condition.

DEFINITION 2.6. If $s \in S$ is a cusp, and if Γ_s denotes the stabilizer subgroup, let $\gamma_s \in \Gamma_s$ denote the unique generator that is conjugate to a matrix $(\begin{smallmatrix} 1 & w \\ 0 & 1 \end{smallmatrix})$ where w is a positive real number. The *residue* $\text{res}_s(\rho, F)$ of (ρ, F) at s is the endomorphism $\rho(\gamma_s)$.

2.2. Filtered regular connections. Closely related to the notion of a filtered representation is that of a filtered connection or filtered flat bundle. To make precise this notion, let (ρ, F) denote a filtered representation and define an action of Γ on the trivial bundle $\mathcal{H} \times V$ by setting $\gamma(\tau, v) = (\gamma\tau, \rho(\gamma)v)$. This descends to a vector bundle on Y that we denote $\mathcal{V}(\rho)$. Sections of \mathcal{V} over $U \subseteq Y$ are identified with functions $f: j^{-1}(U) \rightarrow \mathbf{C}^d$ satisfying $f(\gamma\tau) = \rho(\gamma)f(\tau)$ for all $\gamma \in \Gamma$ and $\tau \in j^{-1}(U)$. In particular, global holomorphic sections are holomorphic modular forms of weight zero for ρ that have no conditions imposed at the cusp.

Weakly holomorphic modular forms associated to Γ are global sections of the meromorphic extension of $\mathcal{V}(\rho)$ to the cusp. Since the dagger symbol is used to denote spaces of weakly holomorphic modular forms in the literature, we will write $\mathcal{V}_\dagger(\rho)$ for the meromorphic extension. It can be described concretely as follows: each stabilizer Γ_s of a cusp s has two possible generators to choose from. Let γ_s denote the unique generator such that

$$g^{-1}\gamma_s g = \begin{pmatrix} 1 & w \\ 0 & 1 \end{pmatrix}$$

for some $g \in \mathrm{PSL}_2(\mathbf{R})$, where w is a *positive* real number. Define a new function $f_g(\tau) = f(g\tau)$ and observe that $f_g(\tau + w) = \rho(\gamma_s)f_g(\tau)$. Since the exponential is surjective, we can choose an endomorphism $L_s \in \mathrm{End}_{\mathbf{C}}(V)$ such that $\rho(\gamma_s) = e^{2\pi i L_s}$. We call L_s a choice of exponents for $\rho(\gamma_s)$, or a choice of exponents at s . For every section of $\mathbf{C} \rightarrow \mathbf{C}/\mathbf{Z}$, there is a unique choice of L_s with eigenvalues taking values in the section. If $\tilde{f}_g(\tau) := e^{-2\pi i L_s \tau} f_g(\tau)$, then it follows that $\tilde{f}_g(\tau + w) = \tilde{f}_g(\tau)$. Since \tilde{f}_g is holomorphic, it follows that \tilde{f}_g admits a Fourier expansion

$$(1) \quad \tilde{f}_g(q_w) = \sum_{n \in \mathbf{Z}} a_n q_w^n$$

where $a_n \in \mathbf{C}^d$ and $q_w = e^{2\pi i \tau/w}$. Whether this q_w -expansion is meromorphic is independent of the choice of L_s . One sees this by describing all possible choices of L_s , typically by using the Jordan canonical form of $\rho(\gamma_s)$, and comparing the results for the different choices.

A section of $\mathcal{V}(\rho)$ in a punctured neighbourhood of the cusp s extends to a section of $\mathcal{V}_\dagger(\rho)$ at s if and only if the q_w -expansion in (1) is meromorphic. If $L = (L_s)_{s \in S}$ denotes a choice of exponents for each cusp, then let $\mathcal{V}(\rho, L) \subseteq \mathcal{V}_\dagger(\rho)$ denote the vector bundle on X isomorphic to $\mathcal{V}(\rho)$ on Y , and such that a section of $\mathcal{V}_\dagger(\rho)$ at s is contained in $\mathcal{V}(\rho, L)$ if and only if the q_w -expansion (1) obtained using the exponents L is in fact holomorphic. Thus, for every choice of exponents L , we obtain inclusions of sheaves $\mathcal{V}(\rho, L) \subseteq \mathcal{V}_\dagger(\rho)$. These bundles are often called *lattices* in the meromorphic extension $\mathcal{V}_\dagger(\rho)$.

The meromorphic extension is naturally endowed with a connection

$$\nabla: \mathcal{V}_\dagger(\rho) \rightarrow \mathcal{V}_\dagger(\rho) \otimes \Omega_X^1(S) \cong \mathcal{V}_\dagger(\rho) \otimes \Omega_X^1,$$

where $\Omega_X^1(S)$ denotes the sheaf of regular differentials on Y with at worst simple poles at the cusps. Viewing sections of $\mathcal{V}_\dagger(\rho)$ as functions $f: \mathcal{H} \rightarrow V$, the holomorphic connection is described simply by $\nabla f = \frac{df}{d\tau} \otimes d\tau$. This connection keeps stable every lattice $\mathcal{V}(\rho, L)$ and restricts to a regular connection

$$\nabla: \mathcal{V}(\rho, L) \rightarrow \mathcal{V}(\rho, L) \otimes \Omega_X^1(S)$$

for every choice of exponents L .

The connection matrix at a cusp can be described explicitly as follows: for simplicity assume that the cusp is ∞ and it is of width $w > 0$ as above. For large enough N , the open set

$$U = \{x + iy \in \mathcal{H} \mid y > N, 0 < x < w\}$$

satisfies $\gamma U \cap U = \emptyset$ for all $\gamma \in \Gamma$. Hence each vector $v \in V$ extends uniquely to a locally constant multivalued section c_v of $\mathcal{V}(\rho, L)$ over $j(U)$. By varying v over a basis for V , we obtain a single valued frame

$$\tilde{c}_v(\tau) = e^{-2\pi i L_s \tau} c_v(\tau)$$

for the fiber of $\mathcal{V}(\rho, L)$ over the cusp s . Since c_v is locally constant, it follows that

$$\nabla(\tilde{c}_v(\tau)) = -2\pi i L_s \tilde{c}_v(\tau) \otimes d\tau.$$

Thus, the connection matrix at the cusp s is $-2\pi i L_s$. The *residue* of a regular connection at a cusp s is the connection matrix divided by $2\pi i$. Therefore,

$$\mathrm{res}_s(\mathcal{V}(\rho, L), \nabla) = -L_s.$$

So far we have not incorporated the filtration F of the filtered representation (ρ, F) into our discussion of regular connections. Our next goal is to introduce a

corresponding filtered object $(\mathcal{V}(\rho, F), G)$, where for each $s \in S$ and real number α , the bundle $G_{\alpha,s} \mathcal{V}(\rho, F) \subseteq \mathcal{V}_\dagger(\rho)$ is an extension of the bundle $\mathcal{V}(\rho)$ to X . The stalk of sections of $G_{\alpha,s} \mathcal{V}(\rho, F)$ over s is generated by all sections of the following type: for every $v \in F_{\beta,s} V$, and for every choice of exponents L_s satisfying $\rho(\gamma_s) = e^{2\pi i L_s}$ and such that the real parts r of the eigenvalues of L_s all satisfy $r < \beta - \alpha$, then $e^{-2\pi i L_s \tau} c_v$ defines a local section of $G_{\alpha,s} \mathcal{V}(\rho, F)$ over s .

This gives an example of the following type of filtered object.

DEFINITION 2.7. A *filtered bundle* on X consists of a holomorphic bundle E on Y , along with filtered families of extensions $E_{\beta,s}$ of E to $s \in S$ for each cusp such that

- (1) (decreasing) $E_{\alpha,s} \subseteq E_{\beta,s}$ for all $\alpha \geq \beta$;
- (2) (left continuous) for all $\beta \in \mathbf{R}$, there exists $\varepsilon > 0$ such that $E_{\beta-\varepsilon,s} = E_{\beta,s}$;
- (3) (periodicity) if z is a local coordinate vanishing to order 1 at s , then $E_{\beta+1,s} = zE_\beta$.

If E is a filtered bundle on X , then let \bar{E} denote the vector bundle on X obtained by using the extensions $E_{0,s}$ at each cusp. The fiber $\bar{E}(s)$ over s is a finite dimensional vector space that inherits a decreasing filtration $\bar{E}_\alpha(s)$ from the extensions $E_{\alpha,s}$ for $\alpha \in [0, 1)$. In more down to earth terms, when E arises as $\mathcal{V}(\rho)$ above, $\bar{E}_\alpha(s)$ is spanned by constant terms at s of modular forms that are sections of $E_{\alpha,s}$. The left continuous and periodicity conditions imply that $\bar{E}_{1-\varepsilon}(s) = 0$ for some $\varepsilon > 0$.

DEFINITION 2.8. The *filtered degree* of a filtered bundle E is the quantity

$$\deg(E) := \deg(\bar{E}) + \sum_{s \in S} \sum_{0 \leq \alpha < 1} \alpha \dim(\text{Gr}_\alpha(\bar{E}(s)))$$

where $\deg(\bar{E})$ is the standard notion of degree for the holomorphic bundle \bar{E} on the compact curve X . The (filtered) *slope* of E is the quantity $\mu(E) = \deg(E) / \text{rank}(E)$.

DEFINITION 2.9. A *filtered regular connection* on X consists of a holomorphic connection (V, ∇) on Y endowed with the structure of a filtered bundle $V_{\beta,s}$ such that for every $s \in S$ and $\beta \in \mathbf{R}$, such that the connection extends to a regular connection

$$\nabla: V_{\beta,s} \rightarrow V_{\beta,s} \otimes \Omega_X^1(s).$$

The *residue* $\text{res}_s(V, \nabla)$ of (V, ∇) at s is the residue of $(V_{0,s}, \nabla)$ at s .

DEFINITION 2.10. A filtered regular connection (V, ∇) is said to be *semistable* provided that for every filtered regular subconnection $(U, \nabla) \subseteq (V, \nabla)$, the inequality $\mu(U, \nabla) \leq \mu(V, \nabla)$ holds. If the inequality holds without equality whenever (U, ∇) is nontrivial, then (V, ∇) is said to be *stable*.

REMARK 2.11. Note that in the stability condition of a filtered regular condition, one only tests against filtered sub-objects, so that the filtered pieces must be preserved by the connection.

We now state, using the precise language and conventions adopted here, a classical theorem that forms one half of nonabelian Hodge theory.

THEOREM 2.12 (Riemann-Hilbert correspondence). *The association*

$$(\rho, F) \mapsto (\mathcal{V}(\rho, F), G, \nabla)$$

described above defines a degree-preserving equivalence between the category of filtered representations of Γ and the category of filtered regular connections on X . This equivalence is compatible with direct sums, determinants, duals and tensor products. Semistable (resp. stable) objects correspond to one another under this equivalence.

See Lemma 3.2 of [47] for a proof.

REMARK 2.13. To lighten notation we will often write $\mathcal{V}(\rho, F)$ for the filtered regular connection associated to a filtered representation (ρ, F) . Subscripts $\mathcal{V}_{s,\beta}(\rho, F)$ indicate that this is a filtered collection of bundles that extend $\mathcal{V}(\rho, F)$ to the cusp s . An overline $\overline{\mathcal{V}}(\rho, F)$ denotes the extension of $\mathcal{V}(\rho, F)$ to X using $\mathcal{V}_{s,0}(\rho, F)$ at each cusp. The fiber of $\overline{\mathcal{V}}(\rho, F)$ over a cusp s will be denoted $\overline{\mathcal{V}}(\rho, F, s)$. Its filtered pieces will be denoted as $\overline{\mathcal{V}}(\rho, F, s)_\beta$ to distinguish this filtration of a finite dimensional vector space from the filtration $\mathcal{V}_{s,\beta}(\rho, F)$ by sheaves.

EXAMPLE 2.14. Let $\Gamma = \mathrm{PSL}_2(\mathbf{Z})$, for which the group of characters is cyclic of order 6 and generated by $\chi: \Gamma \rightarrow \mathbf{C}^\times$ satisfying $\chi(T) = e^{2\pi i/6}$, $\chi(S) = -1$, $\chi(R) = e^{4\pi i/3}$. For $a = 0, \dots, 5$, the choices of logarithm for χ^a are the rational numbers $\frac{a}{6} + n$ for $n \in \mathbf{Z}$. A filtered character consists of a pair (χ^a, b) where $b \in \mathbf{R}$ is the jump of the filtration of χ^a . The filtered degree $\deg(\chi^a, b)$ is equal to b . The residue is $\mathrm{res}_\infty(\chi^a, b) = e^{2\pi i a/6}$.

Next we describe the corresponding filtered regular connection. The bundle on Y underlying $\mathcal{V}(\chi^a, b)$ is the usual bundle associated to χ^a . For $\alpha \in \mathbf{R}$, for every $\beta \leq b$ and for every choice of logarithm $\frac{a}{6} + n < \beta - \alpha$, if c_v denotes the locally constant section associated to a basis $v \in V$ for the one-dimensional complex vector space V , then

$$e^{-2\pi i(\frac{a}{6}+n)\tau} c_v(\tau) = q^{-n} e^{-2\pi i\tau/6} c_v(\tau)$$

defines a local section of $\mathcal{V}_{\alpha,\infty}(\chi^a, b)$. Therefore, if $\mathcal{V}(\chi^a)$ denotes the canonical extension of χ^a defined by the logarithm $a/6$, then sections of $\mathcal{V}_{\alpha,\infty}(\chi^a, b)$ are sections of $\mathcal{V}(\chi^a)$ with a pole of order $n < b - \frac{a}{6} - \alpha$ at the cusp. It follows that

$$\mathcal{V}_{\alpha,\infty}(\chi^a, b) \cong \mathcal{V}(\chi^a, a/6) \otimes \mathcal{O}_X \left(\left\lfloor b - \frac{a}{6} - \alpha \right\rfloor \infty \right)$$

where $\mathcal{O}_X(m\infty)$ denotes the sheaf of holomorphic functions on Y with a pole of order at most m at the cusp (meaning that the functions vanish at the cusp if $m < 0$), and $\mathcal{V}(\chi^a, a/6)$ denotes the canonical extension of $\mathcal{V}(\chi^a)$ to the cusp using the exponent $a/6$. The jumps occur at the values of $\alpha \in b - \frac{a}{6} + \mathbf{Z}$. Therefore, the jump in $[0, 1)$ occurs at $\{b - (a/6)\}$, the fractional part of $b - (a/6)$. To compute the filtered degree of $\mathcal{V}(\chi^a, b)$ we now need to compute the *unfiltered* degree of $\deg \mathcal{V}_{0,\infty}(\chi^a, b)$:

$$\begin{aligned} \deg \mathcal{V}_{0,\infty}(\chi^a, b) &= \deg \mathcal{V}(\chi^a, a/6) \otimes \mathcal{O}_X \left(\left\lfloor b - \frac{a}{6} \right\rfloor \infty \right) \\ &= (a/6) + \lfloor b - (a/6) \rfloor \\ &= (a/6) + (b - (a/6) - \{b - (a/6)\}) \\ &= b - \{b - (a/6)\}. \end{aligned}$$

This shows that the filtered degree of $\mathcal{V}(\chi^a, b)$ is also b .

2.3. Filtered Higgs bundles. Filtered Higgs bundles provide a third collection of filtered objects, whose relationship to filtered connections brings about the other half of nonabelian Hodge theory. While the connection between filtered local systems and filtered regular connections is relatively classical and has long been a part of the study of modular forms, the study of modular forms through Higgs bundles is a more recent development.

DEFINITION 2.15. A *filtered regular Higgs bundle* on X consists of a pair (E, θ) where E is a filtered bundle on X , and θ is a collection of \mathcal{O}_X -linear maps

$$\theta: E_{\beta,s} \rightarrow E_{\beta,s} \otimes \Omega_X^1(s).$$

The degree of a filtered regular Higgs bundles is the degree of the underlying filtered bundle on X . The *residue* of (E, θ) at $s \in S$ is the operator $\theta \left(q \frac{d}{dq} \right)$ acting on the fiber of $E_{0,s}$ over s , where q is a local parameter at s .

DEFINITION 2.16. A filtered regular Higgs bundle (E, θ) is said to be *semistable* provided that for every filtered subbundle $E' \subseteq E$ satisfying $\theta(E') \subseteq E'$, one has $\mu(E', \theta) \leq \mu(E, \theta)$. If the inequality is strict whenever (E', θ) is a nontrivial subobject, then (E, θ) is said to be *stable*.

As will be discussed below, nonabelian Hodge theory extends the Riemann-Hilbert correspondence to a triangle of equivalences including the category of filtered Higgs bundles.

3. Vector valued modular forms

3.1. Definition. For all even integers k , let \mathcal{L}_k denote the line bundle on X that pulls back to \mathcal{H} as the trivial bundle $j^* \mathcal{L}_k = \mathcal{H} \times \mathbf{C}$ endowed with the action $\gamma(\tau, v) = (\tau, (c\tau + d)^k v)$ for $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$. The extension of this bundle to the cusps is the canonical one, so that global sections of \mathcal{L}_k are the usual scalar valued holomorphic modular forms for Γ of weight k .

Let $\rho: \Gamma \rightarrow \mathrm{GL}(V)$ denote a complex finite dimensional representation of a Fuchsian group $\Gamma \subseteq \mathrm{PSL}_2(\mathbf{Z})$. A weakly holomorphic modular form for ρ of weight k is a global holomorphic section of the bundle $\mathcal{V}_\dagger(\rho) \otimes \mathcal{L}_k$. That is, such an object is a holomorphic function $f: \mathcal{H} \rightarrow V$ such that

$$f(\gamma\tau) = (c\tau + d)^k \rho(\gamma) f(\tau)$$

for all $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$, and which is *meromorphic at the cusps* in the following sense: for each cusp s , choose $\gamma_s \in \Gamma$ stabilizing s such that $g\gamma_s g^{-1} = \begin{pmatrix} 1 & w \\ 0 & 1 \end{pmatrix}$ for some $w > 0$ and $g \in \mathrm{PSL}_2(\mathbf{R})$. Then for any choice of logarithm L_s matrix satisfying $\rho(\gamma_s) = e^{2\pi i L_s}$, the function $e^{-2\pi i L_s \tau} f(\tau)$ has a meromorphic $q_r = e^{2\pi i \tau/r}$ expansion.

One often wants to impose bounds on the poles at cusps (or even holomorphy or vanishing), and this is achieved through the mechanism of fixing exponent matrices L_s at each cusp. This defines a holomorphic bundle $\mathcal{V}(\rho, L)$ on the compact curve X as discussed in the previous section. A global section f of $\mathcal{V}_k(\rho, L) := \mathcal{V}(\rho, L) \otimes \mathcal{L}_k$ is a section of $\mathcal{V}_\dagger(\rho) \otimes \mathcal{L}_k$ such that the function $e^{-2\pi i L_s \tau} f(\tau)$ has a *holomorphic* $q_r = e^{2\pi i \tau/r}$ expansion at each cusp. We write

$$M_k(\rho, L) := H^0(X, \mathcal{V}_k(\rho, L))$$

for the corresponding (finite dimensional) space of modular forms. The usual holomorphic modular forms correspond to choosing logarithms L_s with real parts of their eigenvalues inside $[0, 1]$, the *canonical choice* of exponents. Cusp forms correspond to the interval $(0, 1]$, and in general it is advantageous to allow arbitrary exponents. We will see some justification for this below, but one might also mention functorial linear algebraic constructions such as tensor products, symmetric powers, induction, et cetera. These functorial constructions lead naturally to non-canonical exponents — see [19] for examples.

3.2. The three-term inequality. The authors began to think about possible applications of nonabelian Hodge theory to vector valued modular forms when discussing an open question concerning the case of $\Gamma = \mathrm{PSL}_2(\mathbf{Z})$. In this case X can be identified with the weighted projective line $\mathbf{P}(2, 3)$. Every line bundle on $\mathbf{P}(2, 3)$ is isomorphic with $\mathcal{O}(k)$ for an even integer k^2 . For example, one has $\mathcal{L}_k \cong \mathcal{O}(k)$. Moreover every holomorphic bundle on $\mathbf{P}(2, 3)$ decomposes into a direct sum of line bundles, and this decomposition is unique up to isomorphism. For details on these facts one can consult [7].

In this simplified case a choice of logarithms for a representation ρ is a single matrix L satisfying $\rho \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = e^{2\pi i L}$, and the corresponding module

$$M(\rho, L) = \bigoplus_{k \in \mathbf{Z}} M_k(\rho, L)$$

of modular forms is known to be free of rank $d = \dim \rho$ over the ring $\mathbf{C}[E_4, E_6] \cong \bigoplus_{k \in \mathbf{Z}} H^0(X, \mathcal{L}_k)$ of holomorphic scalar valued modular forms of level one. If $d = \dim \rho$, one can always choose a basis of forms of weights k_1, k_2, \dots, k_d for $M(\rho, L)$ over $\mathbf{C}[E_4, E_6]$, where

$$\mathcal{V}_0(\rho, L) \cong \bigoplus_{j=1}^d \mathcal{L}_{-k_j}.$$

Let m_k denote the multiplicity of \mathcal{L}_{-k} occurring in the decomposition above. In [18] it was observed that these weight multiplicities satisfy the inequalities

$$(2) \quad m_k \leq m_{k+2} + m_{k-2}$$

for irreducible representations ρ of dimension at most 12. Moreover in the case that ρ is irreducible and *unitary* of arbitrarily large dimension, [18] proved that this three-term inequality holds. We expect this inequality to hold for all irreducible representations regardless of unitarity hypotheses. Since the nonabelian Hodge correspondence arose from considerations of extending classical results of Narasimhan-Seshadri [43] (in the compact case) and Mehta-Seshadri [40] (in the noncompact case) from unitary bundles to all holomorphic bundles, it is natural to suggest that the study of modular forms associated to nonunitary representations could profit from contact with nonabelian Hodge theory. In particular, can the three-term inequality (2) be verified for all irreducible representations of $\mathrm{PSL}_2(\mathbf{Z})$ by exploiting the rich geometry of the moduli space of filtered Higgs bundles by, say, deforming to the unitary case or to complex variations of Hodge structure? We begin a discussion of this question here and provide some partial answers.

4. Nonabelian Hodge theory

4.1. The nonabelian Hodge correspondence. Here, we review the parts of the full nonabelian Hodge theorem that we need for the current investigation. For a broader survey on the nonabelian Hodge correspondence, we refer the reader to either of [13, 20]. To begin, we let $\rho: \Gamma \rightarrow \mathrm{GL}(V)$ denote a representation of a Fuchsian group Γ . The nonabelian Hodge correspondence is mediated by the existence of a Hermitian metric on $\mathcal{V}(\rho)$ satisfying a nonlinear differential equation, and with prescribed growth conditions at cusps. In order to describe this in concrete automorphic terms, let Herm_n

²Since we are working with $\mathrm{PSL}_2(\mathbf{Z})$ as opposed to $\mathrm{SL}_2(\mathbf{Z})$, all weights should be even. Hence $\mathcal{O}(1)$ is not defined on X , but rather, it is a line bundle on a μ_2 -gerbe over X . We will only work with even weights in this paper.

denote the space of $n \times n$ positive definite Hermitian matrices, and let $K: \mathcal{H} \rightarrow \text{Herm}_n$ denote a smooth map. Define a Hermitian metric on $\mathcal{V}(\rho)$ by setting

$$\langle u, v \rangle_\tau := u^T \cdot K(\tau) \cdot \bar{v}.$$

In order for this to descend to a well-defined metric on $\mathcal{V}(\rho)$ over the open curve $Y = \Gamma \backslash \mathcal{H}$, it is necessary and sufficient that $\langle u, v \rangle_\tau = \langle \rho(\gamma)u, \rho(\gamma)v \rangle_{\gamma\tau}$ for all $\gamma \in \Gamma$. That is, K must satisfy the transformation law

$$K(\tau) = \rho(\gamma)^T \cdot K(\gamma\tau) \cdot \overline{\rho(\gamma)}$$

for all $\tau \in \mathcal{H}$ and $\gamma \in \Gamma$. This means that K is a matrix valued smooth modular form of weight zero taking values in Herm_n . We have not yet imposed any conditions of slow growth at the cusps.

Let \mathcal{A}_Y^1 denote the sheaf of smooth differential one forms on Y . Writing also $\mathcal{V}(\rho)$ for the smooth sheaf associated to the vector bundle $\mathcal{V}(\rho)$, then coordinatewise exterior differentiation defines a map of sheaves

$$d: \mathcal{V}(\rho) \rightarrow \mathcal{V}(\rho) \otimes \mathcal{A}_Y^1.$$

In fact, d is a flat connection since the exterior derivative is flat. The decomposition $\mathcal{A}_Y^1 = \mathcal{A}_Y^{1,0} \oplus \mathcal{A}_Y^{0,1}$ yields a decomposition $d = \partial + \bar{\partial}$ into holomorphic and antiholomorphic components

$$\partial: \mathcal{V}(\rho) \rightarrow \mathcal{V}(\rho) \otimes \mathcal{A}_Y^{1,0}, \quad \bar{\partial}: \mathcal{V}(\rho) \rightarrow \mathcal{V}(\rho) \otimes \mathcal{A}_Y^{0,1}.$$

To be explicit, if a section f of $\mathcal{V}(\rho)$ has coordinate functions f_j , then

$$\partial \begin{pmatrix} f_1(\tau) \\ f_2(\tau) \\ \vdots \\ f_n(\tau) \end{pmatrix} = \begin{pmatrix} (\partial f_1)(\tau) \\ (\partial f_2)(\tau) \\ \vdots \\ (\partial f_n)(\tau) \end{pmatrix} d\tau, \quad \bar{\partial} \begin{pmatrix} f_1(\tau) \\ f_2(\tau) \\ \vdots \\ f_n(\tau) \end{pmatrix} = \begin{pmatrix} (\bar{\partial} f_1)(\tau) \\ (\bar{\partial} f_2)(\tau) \\ \vdots \\ (\bar{\partial} f_n)(\tau) \end{pmatrix} d\bar{\tau}.$$

The bundle $\mathcal{V}(\rho)$ has a natural holomorphic structure given by $\bar{\partial}$ such that the holomorphic sections for this structure are holomorphic functions $f: \mathcal{H} \rightarrow \mathbf{C}^n$ in the usual sense.

If K is a Hermitian metric on $\mathcal{V}(\rho)$, then a smooth connection D on $\mathcal{V}(\rho)$ is said to be a *metric connection* provided that

$$d(K(u, v)) = K(Du, v) + K(u, Dv)$$

for all sections u, v of $\mathcal{V}(\rho)$. The constructions of nonabelian Hodge theory rely on choosing a $(1, 0)$ -connection δ'_K and a $(0, 1)$ -connection δ''_K such that $\partial + \delta''_K$ and $\bar{\partial} + \delta'_K$ are both metric connections relative to K . On a curve these operators can be expressed in terms of matrices as $\delta'_K = \partial + M d\tau$ and $\delta''_K = \bar{\partial} + N d\bar{\tau}$, where M and N are two matrices of smooth functions on Y . By examining the $(1, 0)$ -component, or equivalently, the $(0, 1)$ -component, of the metric connection condition for each of these derivations, one finds the identities:

$$\begin{aligned} \delta'_K &= \partial + \partial \log \bar{K} d\tau, \\ \delta''_K &= \bar{\partial} + \bar{\partial} \log \bar{K} d\bar{\tau}. \end{aligned}$$

Above, if G is an invertible matrix of functions and D is a derivation, we write $D \log G := G^{-1} D(G)$ where $D(G)$ denotes the matrix obtained from G by applying D entrywise.

Using these operators, we can describe the rest of the constructions of nonabelian Hodge theory. They involve the following list of operators:

$$\begin{aligned}\partial_K &= \partial + \frac{1}{2}\partial \log \bar{K}d\tau, & \bar{\partial}_K &= \bar{\partial} + \frac{1}{2}\bar{\partial} \log \bar{K}d\bar{\tau}, \\ \theta_K &= -\frac{1}{2}\partial \log \bar{K}d\tau, & \bar{\theta}_K &= -\frac{1}{2}\bar{\partial} \log \bar{K}d\bar{\tau}.\end{aligned}$$

For a general choice of K , these operators do not need to have desirable properties. In particular, the operators $D'_K = \partial_K + \bar{\theta}_K$ and $D''_K = \bar{\partial}_K + \theta_K$ will not be flat in general. But it turns out that if at least one of them is flat, then both are, and they allow one to pass back and forth between flat regular connections and regular Higgs bundles.

Recall that a Higgs bundle on Y consists of a triple (E, δ, θ) where E is a smooth bundle, $\delta: E \rightarrow E \otimes \mathcal{A}_Y^{0,1}$ is a holomorphic structure on E , and $\theta: E \rightarrow E \otimes \Omega_Y^1$ is holomorphic for the complex structure defined by δ . If E is the smooth bundle underlying $\mathcal{V}(\rho)$, whenever $(D''_K)^2 = 0$, then K is called a *harmonic metric* for ρ , and the triple $(E, \partial_K, \theta_K)$ is the Higgs bundle associated to ρ via nonabelian Hodge theory. While the two bundles associated to one another via nonabelian Hodge theory have the same underlying smooth bundle, in general their complex structures will be different. They will be equal only when $\partial \log \bar{K} = 0$, as is clear from the equations above.

The succinct condition $(D''_K)^2 = 0$ can be made more explicit. Since Y is a curve, a straightforward computation shows that this condition simplifies to

$$\bar{\partial} \partial \log \bar{K} = \frac{1}{2}[\partial \log \bar{K}, \bar{\partial} \log \bar{K}]$$

or equivalently $\partial \bar{\partial} \log K = \frac{1}{2}[\bar{\partial} \log K, \partial \log K]$. Observe that if K is a metric on a line bundle, then the commutator vanishes and this simply says that K is log harmonic in the sense of Definition 4.1.7 of [21]. We thus obtain the following automorphic description of a harmonic metric on a uniformized flat regular connection.

DEFINITION 4.1. Let $\rho: \Gamma \rightarrow \mathrm{GL}_d(\mathbf{C})$ denote a representation of a Fuchsian group Γ . Then a *harmonic metric* for ρ is a smooth function $K: \mathcal{H} \rightarrow \mathrm{Herm}_d$ valued in positive definite Hermitian matrices, and such that the following two conditions are satisfied:

(1) for all $\gamma \in \Gamma$ one has

$$K(\gamma\tau) = \rho(\gamma)^{-T} \cdot K(\tau) \cdot \overline{\rho(\gamma)}^{-1};$$

(2) the harmonic equation is satisfied:

$$\partial \bar{\partial} \log(K) = \frac{1}{2}[\bar{\partial} \log(K), \partial \log(K)].$$

Note that we have not yet imposed any conditions at the cusps.

EXAMPLE 4.2. The simplest examples are unitary representations ρ . In this case the constant identity map $K(\tau) = I$ is a harmonic metric, the complex structure is the standard one under a uniformization, and the corresponding Higgs field is the zero endomorphism. This case gives rise to the usual Petersson scalar product in the theory of modular forms. Thus, one can regard harmonic metrics as generalizations of the Petersson product to nonunitary representations.

More generally, if Γ has cusps and ρ is endowed with filtrations at the cusps, then growth conditions must be imposed at the cusps. Even for unitary representations ρ ,

the harmonic metric with appropriate growth conditions will differ from the Petersson scalar product if the filtration is nontrivial. We will discuss this below.

EXAMPLE 4.3. If K is a harmonic metric for ρ , and if χ is a unitary one-dimensional representation of Γ , then K is also a harmonic metric for $\rho \otimes \chi$.

EXAMPLE 4.4. A more interesting class of examples is provided by the *totally geodesic* metrics. Such examples arise as follows: a metric $K: \mathcal{H} \rightarrow \text{Herm}_n$ is harmonic if and only if the trace of the second fundamental form of K vanishes (see Theorem 3.3.3 of [1] for details). If the second fundamental form itself vanishes, then K is a totally geodesic map. This means that any geodesic in Herm_n tangent to the image of K at some point is in fact entirely contained in the image of K . By Proposition I.4.12 of [6], all such (irreducible) totally geodesic metrics K arise from injective irreducible representations $\alpha: \text{SL}_2(\mathbf{R}) \rightarrow \text{GL}_n(\mathbf{C})$ making the following diagram commute:

$$\begin{array}{ccc} \text{SL}_2(\mathbf{R}) & \xrightarrow{\alpha} & \text{GL}_n(\mathbf{C}) \\ a \mapsto ai \downarrow & & \downarrow g \mapsto gg^T \\ \mathcal{H} & \xrightarrow{K} & \text{Herm}_n \end{array}$$

This construction can be used to obtain the harmonic metric for the inclusion representation $\rho: \Gamma \rightarrow \text{GL}_2(\mathbf{C})$ of a Fuchsian subgroup Γ of $\text{SL}_2(\mathbf{R})$. Let $\rho^*: \text{SL}_2(\mathbf{R}) \rightarrow \text{GL}_2(\mathbf{C})$ denote the dual of the inclusion, so that $\rho^*(a) = a^{-T}$. The resulting totally geodesic metric K can be described, as usual, by lifting $\tau = x + iy \in \mathcal{H}$ to the matrix $a_\tau = \begin{pmatrix} y^{1/2} & xy^{-1/2} \\ 0 & y^{-1/2} \end{pmatrix} \in \text{SL}_2(\mathbf{R})$ so that

$$K(\tau) = \rho^*(a_\tau) \overline{\rho^*(a_\tau)}^T = \frac{1}{y} \begin{pmatrix} 1 & -x \\ -x & x^2 + y^2 \end{pmatrix}.$$

This is the harmonic metric for the inclusion representation of a Fuchsian subgroup of $\text{SL}_2(\mathbf{R})$. The Higgs bundle corresponding to the inclusion representation under the non-abelian Hodge correspondence thus has a complex structure and corresponding Higgs field defined by the operators:

$$\bar{\partial}_K = \bar{\partial} + \frac{1}{(\tau - \bar{\tau})^2} \begin{pmatrix} \tau & -\tau^2 \\ 1 & -\tau \end{pmatrix} d\bar{\tau}, \quad \theta_K = \frac{1}{(\tau - \bar{\tau})^2} \begin{pmatrix} -\bar{\tau} & \bar{\tau}^2 \\ -1 & \bar{\tau} \end{pmatrix} d\tau.$$

Observe that the inclusion representation is not unitarizable, $\bar{\partial}_K$ does not yield the standard complex structure on the uniformized bundle, and the Higgs field θ_K is nonzero.

Totally geodesic metrics have appeared in the literature on automorphic forms – see for example [22], [23], [24], [17].

REMARK 4.5. Since we have insisted that Fuchsian groups are contained in $\text{PSL}_2(\mathbf{R})$, in the previous Example 4.4 we should have twisted the inclusion of the subgroup $\Gamma \subseteq \text{SL}_2(\mathbf{R})$ by a unitary character to be consistent with this hypothesis. However, the harmonic metric is the same in both cases, and so in that example we worked inside $\text{SL}_2(\mathbf{R})$ to avoid introducing an unnecessary twist that would have complicated matters without adding clarity. More generally, the results of Simpson and this paper extend in a straightforward way to subgroups of $\text{SL}_2(\mathbf{R})$ (or higher degree covers) and their associated gerbes over \mathcal{H} by twisting with a unitary character to descend to $\text{PSL}_2(\mathbf{R})$.

REMARK 4.6. In general it is a difficult problem to write down exact formulas for harmonic metrics other than in the rank one case, or in the higher rank unitary

or totally geodesic cases discussed above. The existence proofs discussed in [29], [14], [9], [47], [48], [3], and [4] are not constructive in nature. A natural question is to give a representation-theoretic proof of the existence of harmonic metrics that utilizes the representation theory of (possibly infinite-dimensional) Lie groups and algebras, or even more exotic structures, such as vertex operator algebras (VOAs). Indeed, connections between Higgs bundles and VOAs have already emerged in a number of different contexts, especially in regards to the quantization of the Hitchin system, that is, of the natural integrable system structure on the moduli space of ordinary Higgs bundles on a smooth compact curve. See [12, 2] for instance.

Such a proof concerning harmonic metrics would make the relationship between automorphic forms and nonabelian Hodge theory more transparent. For example, what is the form of the Selberg trace formula for the L^2 -space associated to a tame harmonic metric for a nonunitary representation? In the unitary case this goes back to Selberg and Hejhal (see [27], [28]), and more recently Muller [42] has established a version of the Selberg trace formula for a non-selfadjoint flat Laplacian associated to nonunitary representations, in the case of cocompact groups Γ . (See also [10].) What more can one say if one works systematically with a harmonic metric? Hopefully, these questions will motivate further work and interesting results down the line.

DEFINITION 4.7. Let (ρ, F) denote a filtered representation of Γ . Then a *tame harmonic metric* for (ρ, F) is a harmonic metric K for ρ satisfying the following tameness condition. For every cusp $s \in S$, for every $\beta \in \mathbf{R}$ and $\varepsilon > 0$, the following holds: there exists $N > 0$ and constants c and C such that for all $v \in F_{\beta,s}V$ but $v \notin \bigcup_{\alpha > \beta} F_{\alpha,s}V$, if $\text{im}(\tau) > N$ then

$$c |q_w|^{\beta+\varepsilon} \leq |v|_K \leq C |q_w|^{\beta-\varepsilon}.$$

A harmonic metric for ρ is said to be *tame* if it is tame with respect to some filtered representation (ρ, F) .

Recall that in Definition 4.7, $q_w = e^{2\pi i \tau/w}$ is a local coordinate at the cusp s . The width $w > 0$ of the cusp s is characterized by the fact that the stabilizer $\gamma_s \in \Gamma$ for s is conjugate to $(\begin{smallmatrix} 1 & w \\ 0 & 1 \end{smallmatrix})$. The length $|v|_K$ is a smooth function of $\tau \in \mathcal{H}$ equal to $\sqrt{v^T K(\tau) \bar{v}}$.

EXAMPLE 4.8. Suppose that the filtration F is *trivial*, meaning that for each cusp s ,

$$F_{\beta,s}V = \begin{cases} V & \beta \leq 0, \\ 0 & \beta > 0. \end{cases}$$

In this case we can apply the condition of Definition 4.7 to all basis vectors, with $\beta = 0$. If $K_{ij}(\tau)$ is the (i, j) -entry of K , then tameness is seen to be equivalent to the inequality

$$c |q_w|^\varepsilon \leq K_{ij}(g\tau) \leq C |q_w|^{-\varepsilon}$$

for all τ with $\text{im}(g\tau)$ large enough. So for example, if the inclusion representation is endowed with the trivial filtration, then the totally geodesic metric from Example 4.4 is seen to be tame for this filtration.

In [47], Simpson usually expresses tameness of a harmonic metric in terms of the associated filtered regular Higgs bundle. Above we have expressed it in terms of the filtered local system, as this is the object that is most closely tied to the classical theory of modular forms. One of the key results in [47] is the equivalence of these definitions of tameness. Tameness of a harmonic metric can be described intrinsically without

reference to a filtered local system, a filtered regular connection, or a filtered regular Higgs bundle. Simpson introduced the notion of a *harmonic bundle* to denote such an object.

DEFINITION 4.9. A *harmonic bundle* consists of a representation $\rho: \Gamma \rightarrow \mathrm{GL}(V)$, and an associated harmonic metric K that is *tame* in the following sense: the metric behaves rationally polynomially in the Euclidean distance to each puncture, where the Euclidean distance is measured as usual in terms of a local coordinate q_w at each cusp.

Tame harmonic bundles form a category, and they determine regular filtrations on the associated representations, connections and Higgs bundles. A harmonic bundle (ρ, K) is *irreducible* if it is impossible to decompose ρ and K compatibly into a nontrivial direct sum. A filtered object on X will be said to be *polystable* if it is a direct sum of stable filtered objects. The following Theorem is the main result of [47].

THEOREM 4.10. *Let $\Gamma \subseteq \mathrm{PSL}_2(\mathbf{R})$ be a Fuchsian group, $Y = \Gamma \backslash \mathcal{H}$ and let $X = Y \cup S$ be the compactification by cusps. The constructions discussed above define equivalences of categories between all four of the following categories:*

- (1) *tame harmonic bundles on X ;*
- (2) *polystable filtered representations of Γ ;*
- (3) *polystable filtered regular connections on X ;*
- (4) *polystable filtered regular Higgs bundles on X .*

Irreducible harmonic bundles correspond to irreducible filtered representations, stable filtered regular connections and stable filtered regular Higgs bundles.

PROOF. See page 755 of [47]. Full-faithfulness and essential surjectivity of the various functors are established in Sections 4 and 6, respectively, of [47]. Since the proofs take place locally on the universal cover \mathcal{H} and at the cusps, they extend essentially without change to the case where Γ contains elliptic points. \square

Suppose that (ρ, F) is a filtered representation with corresponding tame harmonic metric K . Let (\mathcal{V}, ∇) and (E, θ) denote the corresponding filtered regular flat connection and filtered regular Higgs bundle, respectively. For each $s \in S$ we have defined residues for each of these filtered objects. In [47], Simpson describes how these residues are related to one another. The description involves breaking up the operator according to Jordan blocks. Then the nonabelian Hodge correspondence respects this Jordan decomposition, but it shifts both eigenvalues and jumps in a prescribed manner. The correspondence is as in the following table.

	(ρ, F)	(\mathcal{V}, ∇)	(E, θ)
Jump	β	$\beta + u$	$-u$
Eigenvalue	$e^{2\pi i(u+vi)}$	$-(u+vi)$	$-\frac{1}{2}(\beta+vi)$

Note that we have a slightly different labelling of variables than that found in [47]. Ours prefers the filtered representation, while that of [47] prefers the Higgs bundle. One surprising feature worth noting is that the jumps in the filtration of the representation shows up in the eigenvalues of the Higgs field. Thus, even if ρ is a unitary representation, the Higgs field will be nonzero if the jumps of (ρ, F) are nonzero.

4.2. Higgs forms. Fix now a representation ρ of a Fuchsian group Γ and a tame harmonic metric K for ρ . The Higgs or Dolbeault side of nonabelian Hodge theory has the following automorphic interpretation: for an integer k , let $\mathcal{A}_k(\rho)$ denote the space

of smooth automorphic forms for ρ of weight k , where we do not impose, initially, any conditions at the cusps. Such an automorphic form is a smooth function $f: \mathcal{H} \rightarrow \mathbf{C}^n$ satisfying

$$f(\gamma\tau) = (c\tau + d)^k \rho(\gamma) f(\tau)$$

for all $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$. Define the subset $\text{Higgs}_{\dagger, k}(\rho, K) \subseteq \mathcal{A}_k(\rho)$ to be those forms $f \in \mathcal{A}_k(\rho)$ satisfying the linear differential equation $\bar{\partial}_K f = 0$ and which have slow growth (relative to K) at the cusps. We call such an automorphic form a *Higgs form* of weight k for ρ .

Observe that $\text{Higgs}_{\dagger, 0}(\rho, K)$ is nothing but the space of global holomorphic sections of $(E, \theta) = (\mathcal{V}, \bar{\partial}_K, \theta_K)$. By abuse of notation let θ_K denote the matrix of functions appearing in the Higgs field. Then this matrix defines a linear map

$$\theta_K: \text{Higgs}_{\dagger, k}(\rho, K) \rightarrow \text{Higgs}_{\dagger, k+2}(\rho, K).$$

If we write

$$\text{Higgs}_{\dagger}(\rho, K) = \bigoplus_{k \in \mathbf{Z}} \text{Higgs}_{\dagger, k}(\rho, K)$$

then this becomes a module over the ring of holomorphic scalar valued modular forms $M(\Gamma)$, and θ_K defines a graded $M(\Gamma)$ -linear endomorphism of $\text{Higgs}_{\dagger}(\rho, K)$. We stress that θ_K is *linear*, and not a derivation.

REMARK 4.11. In fact, since we have only imposed slow growth at cusps, $\text{Higgs}_{\dagger}(\rho, K)$ even has the structure of an $M(\Gamma)[1/\Delta]$ -module, where Δ is the Ramanujan Δ function, and θ_K also commutes with Δ^{-1} .

REMARK 4.12. Note that if (ρ, K) corresponds to a unitary representation with trivial filtration, then $\bar{\partial}_K = \bar{\partial}$, and Higgs forms are nothing but classical modular forms for the standard holomorphic structure.

EXAMPLE 4.13. Consider again Example 4.4. The harmonic metric described there is the tame harmonic metric for the inclusion representation endowed with the trivial filtration. The differential equation satisfied by Higgs forms $f \in \text{Higgs}_k(\rho, K)$ is

$$\bar{\partial}(f) = \frac{1}{(\tau - \bar{\tau})^2} \begin{pmatrix} -\tau & \tau^2 \\ -1 & \tau \end{pmatrix} f.$$

If we write $f = (f_j)$ in coordinates, then this is the pair of equations

$$\begin{aligned} \bar{\partial}(f_1) &= \frac{1}{(\tau - \bar{\tau})^2} (-\tau f_1 + \tau^2 f_2), \\ \bar{\partial}(f_2) &= \frac{1}{(\tau - \bar{\tau})^2} (-f_1 + \tau f_2). \end{aligned}$$

Observe that then $\bar{\partial}(f_1 - \tau f_2) = 0$, so that $g = f_1 - \tau f_2$ is holomorphic in the usual sense, and $\bar{\partial}(f_2) = -\frac{1}{(\tau - \bar{\tau})^2} g$. It follows that there is another holomorphic function h so that $f_2(\tau) = h(\tau) - \frac{1}{\tau - \bar{\tau}} g(\tau)$ and

$$f(\tau) = \begin{pmatrix} -\frac{\bar{\tau}}{\tau - \bar{\tau}} & \tau \\ -\frac{1}{\tau - \bar{\tau}} & 1 \end{pmatrix} \begin{pmatrix} g(\tau) \\ h(\tau) \end{pmatrix}.$$

The automorphy of f implies that g and h are classical holomorphic scalar valued modular forms for Γ of weights $k - 1$ and $k + 1$, respectively. Hence left multiplication

by $M(\tau) = \begin{pmatrix} -\frac{\bar{\tau}}{\tau-\bar{\tau}} & \tau \\ -\frac{1}{\tau-\bar{\tau}} & 1 \end{pmatrix}$ defines an isomorphism

$$(3) \quad M(\tau): M_{\dagger, k-1}(\Gamma) \oplus M_{\dagger, k+1}(\Gamma) \xrightarrow{\cong} \text{Higgs}_{\dagger, k}(\rho, K).$$

The matrix of θ_K in the basis given by g and h is the nilpotent matrix

$$M(\tau)^{-1} \theta_K M(\tau) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

For each $\lambda \in \mathbf{C}^\times$ there is an isomorphism $a_\lambda: (\mathcal{V}, \bar{\partial}_K, \theta_K) \cong (\mathcal{V}, \bar{\partial}_K, \lambda \theta_K)$, and hence this Higgs bundle is fixed by the natural \mathbf{C}^\times -action on Higgs bundles, which rescales the Higgs field. Indeed, the isomorphism is

$$a_\lambda = M(\tau) \begin{pmatrix} \lambda & 0 \\ 0 & 1 \end{pmatrix} M(\tau)^{-1} = \frac{1}{\tau - \bar{\tau}} \begin{pmatrix} \tau - \lambda \bar{\tau} & (\lambda - 1)\tau \bar{\tau} \\ 1 - \lambda & \lambda \tau - \bar{\tau} \end{pmatrix}.$$

The analogue of the isomorphism (3) for the usual complex structure on $\mathcal{V}(\rho)$ for the inclusion representation ρ goes back at least to [37], and the isomorphism (3) says that the complex structure arising from the harmonic metric yields an isomorphic bundle. It is not clear in what generality the bundles underlying a stable flat connection and a stable Higgs bundle that are related under nonabelian Hodge theory are isomorphic. For the inclusion representation, though, one can use this observation to give an interpretation of the complex structure coming from the harmonic metric in terms of the Maass lowering operator $L = (\tau - \bar{\tau})^2 \bar{\partial}$. In this case the partial differential equation $\bar{\partial}_K(f) = 0$ can be expressed as saying that $\text{Higgs}_{\dagger, k}(\rho, K)$ is an inverse image of a space of holomorphic modular forms under Maass's lowering operator:

$$\text{Higgs}_{\dagger, k}(\rho, K) = L^{-1}(M_{\dagger, k-1} \cdot M_{-1}(\rho)).$$

Here $M_{\dagger, k-1}$ is the usual space of scalar valued weakly holomorphic modular forms of level one and weight $k-1$, and $M_{-1}(\rho)$ is the usual space of classical holomorphic modular forms for ρ . Note that $M_{-1}(\rho)$ is described concretely as the one-dimensional vector space spanned by $(\tau, 1)^T$. Modular forms that are inverse images of forms under the Maass lowering operator occur frequently in the theory of mock modular forms and related subjects. It is not clear how or if this relationship between $\bar{\partial}_K$ and Maass's lowering operator generalizes to other filtered representations.

We return now to the general discussion. Since the metric K is tame, it determines filtrations on all of the objects involved in the nonabelian Hodge correspondence. If (E, θ_K) is the filtered regular Higgs bundle corresponding to (ρ, K) , then recall that we write \bar{E} for the bundle on X obtained by extending E to each cusp using the zero-th filtered piece of E . Define

$$\text{Higgs}_k(\rho, K) = H^0(X, \bar{E} \otimes \mathcal{L}_k).$$

Since \bar{E} is a holomorphic bundle on a compact space, each space $\text{Higgs}_k(\rho, K)$ is a finite dimensional vector space. The graded vector space

$$\text{Higgs}(\rho, K) = \bigoplus_{k \in \mathbf{Z}} \text{Higgs}_k(\rho, K)$$

has the structure of a graded $M(\Gamma)$ -module. As above, the Higgs field defines an $M(\Gamma)$ -linear graded endomorphism of this module that increases weights by two. The reader should beware, however, that the Higgs field is the zero map in most cases of classical interest (for example, when ρ is unitary and the filtration induced on ρ by K is trivial).

THEOREM 4.14. *Let $\Gamma = \mathrm{PSL}_2(\mathbf{Z})$, and let Y and X denote the corresponding open and compact curves, respectively. If (ρ, K) describes an irreducible tame harmonic bundle on X , then the module $\mathrm{Higgs}(\rho, K)$ of Higgs forms is free of rank $\dim \rho$ over the ring $M(\Gamma) = \mathbf{C}[E_4, E_6]$ of scalar valued holomorphic modular forms for ρ .*

PROOF. The proof is the same as the geometric proof of the free module theorem for vector valued modular forms from [7]. It uses the splitting principle for vector bundles on $X = \mathbf{P}(2, 3)$, and the fact that the spaces $\mathrm{Higgs}_k(\rho, K)$ are the global sections of the twisted bundles $\overline{E} \otimes \mathcal{L}_k$ on X . Since $\mathrm{Pic}(X) \cong \mathbf{Z}$ with $\mathcal{L}_1 \in \mathrm{Pic}(X)$ a generator, and $M(\Gamma) = \bigoplus_{k \in \mathbf{Z}} H^0(X, \mathcal{L}_k)$, the proof follows. \square

Now, given a tame harmonic bundle (ρ, K) , we obtain two spaces of modular forms: the more classical space $M(\rho, F)$ associated to the filtered representation (ρ, F) obtained from (ρ, K) , and the space $\mathrm{Higgs}(\rho, K)$ of Higgs forms introduced above. Both spaces are finitely generated graded modules over the ring $M(\Gamma)$ of holomorphic modular forms for Γ . In the special case that $\Gamma = \mathrm{PSL}_2(\mathbf{Z})$, both modules are free of rank $\dim \rho$ over $M(\Gamma)$. If ρ is unitary and the filtration that K induces on ρ is trivial, then $M(\rho, F)$ and $\mathrm{Higgs}(\rho, K)$ are in fact *equal*, not just isomorphic. There is no reason why these spaces must be isomorphic in general. Since the three-term inequality for modular forms on $\mathrm{PSL}_2(\mathbf{Z})$ is a statement about the weights of a graded basis for $M(\rho, F)$, it is not a priori clear how to use nonabelian Hodge theory to study $M(\rho, F)$. In principle the module $M(\rho, F)$ of interest could look quite different from the corresponding module $\mathrm{Higgs}(\rho, K)$ on the Dolbeault side of the nonabelian Hodge correspondence. The solution is that while nonabelian Hodge theory need not preserve the isomorphism type of the holomorphic bundles involved in its formulation, it does preserve the isomorphism type of their associated cohomology groups (see Section 12 of [3] for the noncompact case, or the earlier paper [48] for the compact case). It turns out that this is sufficient to show that the three-term inequality is true for all filtered regular connections if and only if it is true for all filtered regular Higgs bundles. Thus, when proving the three-term inequality, we may as well work within the category of filtered regular Higgs bundles. In the next section we describe a class of regular Higgs bundles that satisfy the three-term inequality, but which do not correspond to unitary representations of $\mathrm{PSL}_2(\mathbf{Z})$.

5. New cases of the three-term inequality for $\mathrm{PSL}_2(\mathbf{Z})$

5.1. Maximally-decomposed variations of Hodge structure: a proof. In [18], the three-term inequality for irreducible unitary representations of $\mathrm{PSL}_2(\mathbf{Z})$ was established using the positivity of the corresponding regular connections. Unitary representations endowed with trivial filtrations correspond to Higgs bundles with Higgs field equal to zero. In the next Proposition we establish a new case of the conjecture for a simple type of variation of Hodge structure on X , in which the filtered Higgs bundle has a distinguished decomposition into an ordered list of sub-line bundles and with the Higgs field shifting the list by -1 .

Stability of the Higgs bundle will play a key role, and it suggests that the three-term inequality should be interpreted as a natural manifestation of stability. Results possessing a similar flavour already exist in the Higgs bundle literature. For instance, the so-called “co-Higgs bundles” on the ordinary projective line \mathbf{P}^1 , whose Higgs fields take values in the anticanonical line bundle $\mathcal{O}(2)$ — which, incidentally, bear a formal similarity to the objects below — have Grothendieck splitting numbers that are forced

by slope stability to satisfy a certain inequality [45]. In general, one can view constructing a stack of Higgs bundles with fixed topological data (rank, degree, etc.) as an unbounded partition problem in which degrees of sub-sheaves are (possibly negative) integer parts of the total degree. Passing to the moduli space via stability intervenes to make the problem bounded and, hence, well-posed. For more on the combinatorics of Higgs bundles and the stability condition, see [46, §5].

THEOREM 5.1. *Let (E, θ) denote a stable filtered regular Higgs bundle on X . Assume that it is a complex variation of Hodge structure, such that \bar{E} , the zero-th filtered piece of E , decomposes into line bundles $\bar{E} \cong \bigoplus_{j=1}^n \mathcal{O}(r_j)$ such that for each $1 \leq j \leq n - 1$, the Higgs field θ restricts to a map*

$$\theta: \mathcal{O}(r_j) \rightarrow \mathcal{O}(r_{j+1}) \otimes \Omega_X^1(\infty) \cong \mathcal{O}(r_{j+1} + 2).$$

If m_r denotes the multiplicity of r among the roots r_j of \bar{E} , then

$$m_r \leq m_{r-2} + m_{r+2}$$

for all r .

PROOF. Stability forces all of the maps θ among the line bundles to be nonzero, save for $\theta|_{\mathcal{O}(r_n)} = 0$, and all of the roots r_j must have the same parity. Since $H^0(X, \mathcal{O}(r)) = 0$ if $r < 0$ or $r = 2$, it follows that since $\theta: \mathcal{O}(r_j) \rightarrow \mathcal{O}(r_{j+1} + 2)$ is nonzero, we must have $r_{j+1} + 2 - r_j \geq 0$ (there are no holomorphic modular forms of negative weight) and $r_{j+1} + 2 - r_j \neq 2$ (there are no holomorphic modular forms of weight 2 and level one). Therefore, if we consider the graph G obtained by joining the points (j, r_j) by straight lines, then G has the following properties:

- (1) it has no horizontal segments;
- (2) it can drop at most by steps of 2;
- (3) it rises by even integers.

We wish to show that if we intersect G with the horizontal lines at height r , $r - 2$ and $r + 2$, then the number of points with integer coordinates on the line at height r is bounded by the sums of the number of points on the lines at heights $r + 2$ and $r - 2$. To see that this is so, label the regions between intersection points with the line at height r with either an A , B , or C . The A and B indicate that G is either entirely above or entirely below the line at height r in the given region. Since G drops by at most two at each step, the only other possibility is that G starts below the line at height r , but then rises up by an even amount without hitting the line at height r at an integer lattice point. Label such regions by C . Notice that after a region labeled by C , the next time G intersects the line at height r , it must do so at an integer lattice point from above, as G decreases at each step only by 2.

Consider walking along G from left to right. At a point of intersection with the line at height r , if the region to the right is labelled B , we may pair the intersection point at height r with the following intersection point that must be at height $r - 2$, since the graph decreases exactly by 2. If instead the region is labelled A , then we may travel along the region up until the next intersection point at height r . This next point is approached from above. We thus pair the point under consideration with the point at height $r + 2$ to the left of the *next* intersection point at height r . If the region is labelled C , we can in fact pair our point with the point to the right of it at height $r - 2$, or the point to the left of the *next* intersection point, necessarily at height $r + 2$.

It remains to handle the rightmost intersection point. Suppose first that $r < r_1$. Then to the left of the leftmost intersection point at height r , the graph must drop down

to meet that intersection point. That means that the line at height $r + 2$ contributes one more intersection point to m_{r+2} to the left of the leftmost intersection point at height r , and hence $m_r \leq m_{r-2} + m_{r+2}$.

Finally, suppose instead that $r \geq r_1$. Stability implies that $r_n < r_1$. Thus, G must eventually drop below r_1 after the rightmost intersection point at height $r \geq r_1$. Hence the line at height $r - 2$ will contribute an extra point of intersection, contributing $+1$ to m_{r-2} , and this again proves that $m_r \leq m_{r-2} + m_{r+2}$. \square

5.2. Nilpotent Higgs bundles: a sketch of an idea. As noted previously, one of the nice features of working with Higgs bundles is the confluence of geometric structures on their moduli space that can be used to study properties of Higgs bundles that remain invariant in a family. With this in mind, one strategy for proving the three-term inequality for a larger class of filtered regular Higgs bundles is to use the differential and symplectic topology of the moduli space to reduce the problem to the complex variations of Hodge structure in Proposition 5.1. The key fact here is that the complex variations of Hodge structure are fixed points of a \mathbf{C}^* -action on the moduli space of stable filtered regular Higgs bundles. This action, whose compact analogue was first exploited by Hitchin in [29], simply rescales the Higgs field:

$$\theta \xrightarrow{\lambda} \lambda\theta.$$

This action interacts with the structure of the moduli space via the *Hitchin map* that sends each filtered Higgs bundle to the tuple of coefficients of the characteristic polynomial of θ . In this sense, the Hitchin map is a generalized Chevalley morphism for the Lie algebra of the structure group of the filtered bundles. The target space of this map is an affine space and the fibres are called *Hitchin fibres*.³ The complex variations of Hodge structure have nilpotent θ (meaning as per usual that $\theta^k = 0$ for some $k \geq 1$), and so they lie in a special fibre of the Hitchin map, called the *nilpotent cone* or *core*, which we signify by \mathcal{N} . Now, the particular variations considered in Proposition 5.1 correspond in the Higgs bundle literature to fixed points of type $(1, \dots, 1)$, where the 1's are simply recording the fact that each summand in the Hodge structure is of rank 1. In general, the type of a fixed point is some ordered partition of the total rank, n . The various types of fixed points are organized as a poset within the cone according to the value of a real-valued functional defined on the moduli space. Typically, this functional is a nondegenerate function $f \geq 0$, called the *Morse function*, which in this case is some scalar multiple of the L^2 -norm of θ . This norm descends to the moduli space from an infinite-dimensional affine space of smooth connections via the Kähler quotient construction [5] and is a moment map for the compact group $U(1) \hookrightarrow \mathbf{C}^*$, which has the same fixed points as the noncompact action. In general, \mathcal{N} is a highly singular variety with multiple irreducible components. One expects that the type $(1, \dots, 1)$ fixed points are local maxima of $f|_{\mathcal{N}}$. In other words, to each irreducible component, there should be an ordered n -tuple of integers m_j , and vice-versa. (Different orderings may be possible and correspond to distinct components.) The downward gradient flow of f from the type $(1, \dots, 1)$ fixed points eventually captures all of \mathcal{N} . The flow trajectories are determined by the positive weight space of the normal bundle to \mathcal{N} , which in turn

³As a side note, it is now known from the work of Laumon, Ngô, Chaudouard, and others (for example, [8, 39, 44]) that the Hitchin fibration is closely related to the affine Springer fibration in geometric representation theory, which is yet another avenue along which Higgs bundles enter automorphic representation theory.

is determined by the positive eigenvalues of $\text{Hess}(f)$. Some of these directions correspond to deformations of θ that leave the bundle \bar{E} invariant, while others transform the holomorphic structure on \bar{E} alone. This former type of deformation reorganizes the line bundles in \bar{E} but preserves the numbers m_j .

The challenge here is to show that:

- (1) for any $(E, \theta) \in \mathcal{N}$, its splitting into line bundles is a permutation of those appearing in a connected component of the $(1, \dots, 1)$ fixed point locus and that (E, θ) is connected to that locus via the Morse flow;
- (2) all reorganizations of the line bundles coming from the θ -deforming flow trajectories either continue to satisfy the three-term inequality (by virtue of having satisfied it for the type $(1, \dots, 1)$ fixed points) or become unstable;
- (3) a version of Morse-Bott or Bialynicki-Birula theory, which have both been developed for classical moduli of Higgs bundles (for example, [49, 25]) can be correctly adapted to the moduli space in this setting, in order to ensure that the gradient flow proceeds normally.

Regarding the last point, these theories, which are usually deployed to detect the topology of a space, are sensitive to singularities and other defects that can interrupt flow lines. Although the cone itself is singular, the smoothness of the total space of the moduli space is usually sufficient to guarantee sensible results. Despite the technical wrinkles, these ideas suggest a potentially fruitful avenue to proving the three-term inequality for a large number of cases. With \mathcal{N} under control, one could then examine flow lines into the cone from the rest of the moduli space, potentially establishing the three-term inequality for less restrictive Higgs bundles and hence for further types of representations.

References

- [1] Paul Baird and John C. Wood. *Harmonic morphisms between Riemannian manifolds*, volume 29 of *London Mathematical Society Monographs. New Series*. The Clarendon Press, Oxford University Press, Oxford, 2003.
- [2] David Ben-Zvi and Edward Frenkel. Geometric realization of the Segal-Sugawara construction. In *Topology, geometry and quantum field theory*, volume 308 of *London Math. Soc. Lecture Note Ser.*, pages 46–97. Cambridge Univ. Press, Cambridge, 2004.
- [3] Olivier Biquard. Fibrés de Higgs et connexions intégrables: le cas logarithmique (diviseur lisse). *Ann. Sci. École Norm. Sup. (4)*, 30(1):41–96, 1997.
- [4] Olivier Biquard and Philip Boalch. Wild non-abelian Hodge theory on curves. *Compos. Math.*, 140(1):179–204, 2004.
- [5] Indranil Biswas, Peter B. Gothen, and Marina Logares. On moduli spaces of Hitchin pairs. *Math. Proc. Cambridge Philos. Soc.*, 151(3):441–457, 2011.
- [6] Armand Borel and Lizhen Ji. *Compactifications of symmetric and locally symmetric spaces*. Mathematics: Theory & Applications. Birkhäuser Boston, Inc., Boston, MA, 2006.
- [7] Luca Candelori and Cameron Franc. Vector-valued modular forms and the modular orbifold of elliptic curves. *Int. J. Number Theory*, 13(1):39–63, 2017.
- [8] Pierre-Henri Chaudouard and Gérard Laumon. Sur l’homologie des fibres de Springer affines tronquées. *Duke Math. J.*, 145(3):443–535, 2008.
- [9] Kevin Corlette. Flat G -bundles with canonical metrics. *J. Differential Geom.*, 28(3):361–382, 1988.
- [10] Anton Deitmar. A non-unitary trace formula for arbitrary groups. arXiv:1703.03709.
- [11] R. Donagi and T. Pantev. Langlands duality for Hitchin systems. *Invent. Math.*, 189(3):653–735, 2012.
- [12] Ron Donagi, Sheldon Katz, and Eric Sharpe. Spectra of D-branes with Higgs vevs. *Adv. Theor. Math. Phys.*, 8(5):815–860 (2005), 2004.
- [13] Ron Donagi and Tony Pantev. Geometric Langlands and non-abelian Hodge theory. In *Surveys in differential geometry. Vol. XIII. Geometry, analysis, and algebraic geometry: forty years of*

the Journal of Differential Geometry, volume 13 of *Surv. Differ. Geom.*, pages 85–116. Int. Press, Somerville, MA, 2009.

- [14] S. K. Donaldson. Twisted harmonic maps and the self-duality equations. *Proc. London Math. Soc.* (3), 55(1):127–131, 1987.
- [15] W. Duke. The dimension of the space of cusp forms of weight one. *Internat. Math. Res. Notices*, (2):99–109, 1995.
- [16] Olivia Dumitrescu and Motohico Mulase. Quantum curves for Hitchin fibrations and the Eynard-Orantin theory. *Lett. Math. Phys.*, 104(6):635–671, 2014.
- [17] Hélène Esnault and Michael Harris. Chern classes of automorphic vector bundles. *Pure Appl. Math. Q.*, 13(2):193–213, 2017.
- [18] Cameron Franc and Geoffrey Mason. On the structure of modules of vector valued modular forms. *Ram. J.*, 2017.
- [19] Cameron Franc and Geoffrey Mason. Constructions of vector-valued modular forms of rank four and level one. (*submitted*), 2018.
- [20] Alberto García-Raboso and Steven Rayan. Introduction to nonabelian Hodge theory: flat connections, Higgs bundles and complex variations of Hodge structure. In *Calabi-Yau varieties: arithmetic, geometry and physics*, volume 34 of *Fields Inst. Monogr.*, pages 131–171. Fields Inst. Res. Math. Sci., Toronto, ON, 2015.
- [21] William M. Goldman and Eugene Z. Xia. Rank one Higgs bundles and representations of fundamental groups of Riemann surfaces. *Mem. Amer. Math. Soc.*, 193(904):viii+69, 2008.
- [22] Michael Harris. Arithmetic vector bundles and automorphic forms on Shimura varieties. I. *Invent. Math.*, 82(1):151–189, 1985.
- [23] Michael Harris. Arithmetic vector bundles and automorphic forms on Shimura varieties. II. *Compositio Math.*, 60(3):323–378, 1986.
- [24] Michael Harris. Automorphic forms and the cohomology of vector bundles on Shimura varieties. In *Automorphic forms, Shimura varieties, and L-functions, Vol. II (Ann Arbor, MI, 1988)*, volume 11 of *Perspect. Math.*, pages 41–91. Academic Press, Boston, MA, 1990.
- [25] Tamás Hausel and Fernando Rodriguez Villegas. Cohomology of large semiprojective hyperkähler varieties. *Astérisque*, (370):113–156, 2015.
- [26] Tamás Hausel and Michael Thaddeus. Mirror symmetry, Langlands duality, and the Hitchin system. *Invent. Math.*, 153(1):197–229, 2003.
- [27] Dennis A. Hejhal. *The Selberg trace formula for $\mathrm{PSL}(2, \mathbf{R})$. Vol. I*. Lecture Notes in Mathematics, Vol. 548. Springer-Verlag, Berlin-New York, 1976.
- [28] Dennis A. Hejhal. *The Selberg trace formula for $\mathrm{PSL}(2, \mathbf{R})$. Vol. 2*, volume 1001 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1983.
- [29] N. J. Hitchin. The self-duality equations on a Riemann surface. *Proc. London Math. Soc.* (3), 55(1):59–126, 1987.
- [30] Nigel Hitchin. Stable bundles and integrable systems. *Duke Math. J.*, 54(1):91–114, 1987.
- [31] Anton Kapustin and Edward Witten. Electric-magnetic duality and the geometric Langlands program. *Commun. Number Theory Phys.*, 1(1):1–236, 2007.
- [32] Marvin Knopp and Geoffrey Mason. On vector-valued modular forms and their Fourier coefficients. *Acta Arith.*, 110(2):117–124, 2003.
- [33] Marvin Knopp and Geoffrey Mason. Vector-valued modular forms and Poincaré series. *Illinois J. Math.*, 48(4):1345–1366, 2004.
- [34] Marvin Knopp and Geoffrey Mason. Logarithmic vector-valued modular forms. *Acta Arith.*, 147(3):261–262, 2011.
- [35] Marvin Knopp and Geoffrey Mason. Logarithmic vector-valued modular forms and polynomial-growth estimates of their Fourier coefficients. *Ramanujan J.*, 29(1-3):213–223, 2012.
- [36] Marvin Knopp and Geoffrey Mason. Vector-valued modular forms with an unnatural boundary. In *From Fourier analysis and number theory to Radon transforms and geometry*, volume 28 of *Dev. Math.*, pages 341–352. Springer, New York, 2013.
- [37] Michio Kuga and Goro Shimura. On vector differential forms attached to automorphic forms. *J. Math. Soc. Japan*, 12:258–270, 1960.
- [38] François Labourie. *Lectures on representations of surface groups*. Zurich Lectures in Advanced Mathematics. European Mathematical Society (EMS), Zürich, 2013.

- [39] Gérard Laumon and Bao Châu Ngô. Le lemme fondamental pour les groupes unitaires. *Ann. of Math.* (2), 168(2):477–573, 2008.
- [40] V. B. Mehta and C. S. Seshadri. Moduli of vector bundles on curves with parabolic structures. *Math. Ann.*, 248(3):205–239, 1980.
- [41] Philippe Michel and Akshay Venkatesh. On the dimension of the space of cusp forms associated to 2-dimensional complex Galois representations. *Int. Math. Res. Not.*, (38):2021–2027, 2002.
- [42] Werner Müller. A Selberg trace formula for non-unitary twists. *Int. Math. Res. Not. IMRN*, (9):2068–2109, 2011.
- [43] M. S. Narasimhan and C. S. Seshadri. Holomorphic vector bundles on a compact Riemann surface. *Math. Ann.*, 155:69–80, 1964.
- [44] Bao Châu Ngô. Le lemme fondamental pour les algèbres de Lie. *Publ. Math. Inst. Hautes Études Sci.*, (111):1–169, 2010.
- [45] Steven Rayan. Co-Higgs bundles on \mathbb{P}^1 . *New York J. Math.*, 19:925–945, 2013.
- [46] Steven Rayan. Aspects of the topology and combinatorics of Higgs bundle moduli spaces. *SIGMA*, 14:129, 18 pages, 2018.
- [47] Carlos T. Simpson. Harmonic bundles on noncompact curves. *J. Amer. Math. Soc.*, 3(3):713–770, 1990.
- [48] Carlos T. Simpson. Higgs bundles and local systems. *Inst. Hautes Études Sci. Publ. Math.*, (75):5–95, 1992.
- [49] Graeme Wilkin. Morse theory for the space of Higgs bundles. *Comm. Anal. Geom.*, 16(2):283–332, 2008.

E-mail address: franc@math.usask.ca; rayan@math.usask.ca

DEPARTMENT OF MATHEMATICS & STATISTICS, UNIVERSITY OF SASKATCHEWAN, MCLEAN HALL, 106 WIGGINS ROAD, SASKATOON, SK, CANADA S7N 5E6