From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Mathematics

2018

Diophantine Analysis on Moduli of Local Systems

Whang, Junho Peter

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp015q47rr43v

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

Diophantine Analysis on Moduli of Local Systems

Junho Peter Whang

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Mathematics
Adviser: Peter Clive Sarnak

June 2018

© Copyright by Junho Peter Whang, 2018.

All Rights Reserved

Abstract

We develop a Diophantine analysis on moduli of special linear rank two local systems over surfaces with prescribed boundary traces. We first show that such a moduli space is a log Calabi-Yau variety if the surface has nonempty boundary, and relate this property to a symmetry of generating series for combinatorial counts of essential multicurves on surfaces. We establish the finiteness of "class numbers" for integral orbits of mapping class group dynamics on the moduli space, generalizing a classical Diophantine work of Markoff. We also derive effective finiteness results for integral points of algebraic curves on the moduli space, as well as a structure theorem for morphisms from the affine line into the moduli space. As part of our work, we establish boundedness theorems for archimedean and nonarchimedean systoles of local systems. Finally, we give a complete classification of the finite orbits of the mapping class group on the moduli space, for surfaces of positive genus.

Acknowledgements

First and foremost, I thank my advisor Professor Peter Sarnak for providing insightful guidance, encouragement, and enthusiasm throughout my research, and for sharing his broad view and expertise in mathematics with generosity. Being his student has been a true pleasure and a transformative experience. I am also grateful to Professor Phillip Griffiths, whose generous guidance, support, and sharing of knowledge and deep insight were indispensable in the development of this thesis.

I thank Professor Manjul Bhargava for his warm support, encouragement, and reading of this thesis, as well as Professor Christopher Skinner and Professor Shou-Wu Zhang for examining my defense. I also thank Professor Sophie Morel, who provided important guidance during parts of my research. In addition, I thank the professors, postdoctoral fellows, staff, colleagues, and friends who enriched my mathematical and personal life while at Princeton. My research was partially supported by postgraduate scholarships from the Natural Sciences and Engineering Research Council of Canada.

Finally, I am greatly indebted to my family for their continuous love and support. I remember with special fondness my grandfather and physicist Byung Hyuk Lee (1926-2017), who inspired me through his life of constant learning and discovery. To him and all of my family, I dedicate this dissertation.

To my family

Contents

Abstract iii

Acknowledgements iv

1 Introduction **1**

1.1 Outline of main results 2

2 Background **8**

2.1 Surfaces 8

2.2 Character varieties 10

2.3 Markoff type cubic surfaces 15

3 Global geometry **18**

3.1 Word compactifications 21

3.2 Log Calabi-Yau property 29

3.3 Analysis of singularities 38

3.4 Curves on surfaces 57

3.5 Combinatorics of planar graphs 65

3.A Auxiliary results on matrices 72

4 Systoles of local systems **83**

4.1 Differential geometry 86

4.2 Harmonic maps 94

4.3 Systoles of local systems 104

5 Diophantine results **116**

5.1 Compactness criterion 120

5.2 Parabolic subvarieties 139

5.3 Arithmetic hyperbolic surfaces 154

5.4 Integral points on curves 160

6 Periodic points **173**

6.1 Background 176

6.2 Case of one-holed torus 182

6.3 Zariski dense representations 186

6.4 Sparse representations 195

Chapter 1

Introduction

The principle of descent in Diophantine analysis has been instrumental for several foundational discoveries in number theory, often expressed through finitude of orbits in integral solutions of Diophantine equations. Classical examples include Gauss's reduction theory of binary quadratic forms, Dirichlet's unit theorem, and Minkowski's geometry of numbers. Another instance of descent, different from the previous linear examples, concerns the cubic Diophantine equation

$$x^2 + y^2 + z^2 - xyz - 2 = k. \tag{1.1}$$

An equation of such form was considered by Markoff in his 1880 work [65] on indefinite quadratic forms. For each $k \in \mathbb{Z}$, the affine cubic surface defined by (1.1) admits a nonlinear group action, generated by maps such as $(x, y, z) \mapsto (x, z, xz - y)$, under which the integral points decompose into finitely many orbits when $k \neq 2$.

It has been understood that the Markoff cubic surfaces above fit among infinite families of algebraic varieties, which arise from topology and are endowed with natural nonlinear dynamics. Namely, given a compact oriented smooth surface of genus g with n boundary curves satisfying $3g+n-3 > 0$, we have the moduli space X_k of SL_2-local systems over the surface having prescribed boundary traces $k \in \mathbb{A}^n(\mathbb{Z})$. The mapping

class group of the surface acts naturally on each X_k via pullback of local systems, recovering the Markoff surfaces and their dynamics when $(g, n) = (1, 1)$.

In this dissertation, we establish, among others, the general theory of nonlinear descent for these moduli spaces X_k using tools of differential geometry. Inspired by classical Diophantine work on subvarieties of algebraic tori and abelian varieties, we also apply our philosophy of descent to obtain a complete and effective Diophantine analysis of integral points for nondegenerately embedded algebraic curves on X_k. Our results suggest that topology of surfaces has a fundamentally Diophantine nature, of which time seems ripe for further investigation.

1.1 Outline of main results

This thesis is based on the preprints [95], [96], [97], and [98] by the author. Together, these works form a coherent series of investigations into the moduli spaces X_k under the Diophantine theme mentioned above. Below, we give a motivated introduction of our main results. In Chapter 2, we shall collect necessary background on surfaces and the moduli of local systems.

Global geometry

A focal point of Diophantine geometry is the interplay between the behavior of integral points on an algebraic variety and the underlying geometry. Standard conjectures [91], [34] suggest that, given a normal irreducible projective variety Z and a reduced effective divisor $D \subset Z$, heuristically there should be *few* (e.g. not Zariski dense), resp. *many*, integral points on $Z \setminus D$ when the log canonical divisor $K_Z + D$ is *positive* (e.g. ample), resp. *negative*.

Definition 1.1. An irreducible algebraic variety V over $\mathbb{C}$ is *log Calabi-Yau* if there is a projective pair (Z, D) as above with $V = Z \setminus D$ and $K_Z + D$ linearly trivial.

It follows that log Calabi-Yau varieties form the boundary case in Diophantine analysis whose integral points are least well understood, even conjecturally. Moreover, any patterns in their integral points are more likely to reflect, and be governed by, a deeper underlying structure of Diophantine nature. For example, for algebraic tori and abelian varieties (which are indeed log Calabi-Yau), this deeper structure manifests itself in the group law. Dirichlet's unit theorem and the Mordell-Weil theorem show that the dynamics of these varieties are sufficiently rich to bring their Diophantine aspects to our understanding.

In Chapter 3, we analyze the global geometry of our moduli spaces X_k when the surface has nonempty boundary, and show that they are log Calabi-Yau (Theorem 3.1). This gives a first indication of the Diophantine richness of X_k, and motivates the Diophantine study of its dynamics taken up in this thesis. A key ingredient in our proof is the symmetry of a suitable Hilbert series associated to a compactification of X_k, which we interpret in Chapter 3 as a generating series for certain counts of multicurves on the given surface. In this way, we connect the global geometry of X_k to a new result (Theorem 3.2) on the combinatorial topology of surfaces.

Systoles of local systems

The crucial ingredient in our theory of descent on X_k is the moduli interpretation of the points of X_k. Indeed, it enables us to translate dynamical statements about the mapping class group action on X_k into intrinsic statements on local systems, which are amenable to differential geometric tools. This is in close parallel with the modern passage between Minkowski's work on the geometry of numbers and the dynamics of $\mathrm{SL}_m(\mathbb{Z})$ on the symmetric space $\mathrm{SL}_m(\mathbb{R})/\mathrm{SO}(m)$.

In Chapter 4, we introduce the notion of *systoles* for $\mathrm{SL}_2(\mathbb{C})$-local systems on the surface and study its properties. Using twisted harmonic maps to the hyperbolic three-space $\mathbb{H}^3$ and systolic geometry of Riemannian surfaces, we prove that systole

is bounded as a function on $X_k(\mathbb{C})$ (Theorem 4.1). This generalizes a classical result of Bers [4] on systoles of hyperbolic surfaces, and later provides a key step in our descent for integral points on X_k. Next, given a field F with a discrete valuation, we also prove a boundedness result (Theorem 4.2) for systoles of $\mathrm{SL}_2(F)$-local systems by a similar strategy, using the Bruhat-Tits tree of $\mathrm{SL}_2(F)$ in place of $\mathbb{H}^3$. This result is used later in our Diophantine analysis of algebraic curves on X_k in Chapter 5.

Finiteness of class numbers

Along with algebraic tori and abelian varieties mentioned above, a major source of analogy in our Diophantine work on the moduli spaces X_k is provided by linear arithmetic groups. Homogeneous varieties with linear arithmetic group actions have been studied extensively, and tools from the theory of arithmetic groups have been applied in their Diophantine study with great success; see for instance [78], [28]. The foundational result in the theory is the finiteness of *class numbers*, i.e. number of integral orbits, proved in general by Borel and Harish-Chandra [8] building on the works of Gauss, Hermite, Minkowski, Siegel, and others.

In Chapter 5, we prove the main Diophantine results of our thesis, one of which provides a nonlinear analogue of the finiteness of class numbers for the mapping class group action on $X_k(\mathbb{Z})$. Our proofs make essential uses of the boundedness results for systoles obtained in Chapter 4. In Section 5.1, we obtain an extension of Mumford's compactness theorem [70] to local systems (Theorem 5.1). One immediate corollary is an extension of McKean's theorem [67] on finiteness of length isotrivial Riemann surfaces. Another consequence of Theorem 5.1 is that the set of all integral points on X_k whose corresponding local systems have trace $\neq \pm 2$ along essential curves decomposes into finitely many mapping class group orbits.

In Section 5.2, we study the behavior of nonconstant morphisms $\mathbb{A}^1 \to X_k$ from the affine line. Our study motivates the following definition.

Definition 1.2. A (possibly reducible) algebraic variety V over $\mathbb{C}$ is *parabolic* if every closed point of V lies in the image of a nonconstant morphism $\mathbb{A}^1 \to V$.

A closely related notion is that of a *log uniruled* variety, introduced in [53]. We obtain a modular characterization of points lying on parabolic subvarieties of X_k, and in particular show (Theorem 5.2) that there is a closed parabolic subvariety $V \subsetneq X_k$ whose translates under the mapping class group contain the images of all nonconstant morphisms $\mathbb{A}^1 \to X_k$. This result is reminiscent of Sterk's finiteness theorem [85] for orbits of smooth rational curves on projective K3 surfaces. A consequence of our work is that certain polynomial deformations of Fuchsian surface groups are necessarily isotrivial. Theorem 5.2 mentioned above inspires the following definition.

Definition 1.3. A subvariety of a log Calabi-Yau variety V is *degenerate* if it lies on a parabolic subvariety of V. Otherwise, it is *nondegenerate*.

We remark that degenerate loci in algebraic tori and abelian varieties are empty, but for general log Calabi-Yau varieties they need not be. Informally, the log canonical divisor "drops in positivity" along parabolic subvarieties of a log Calabi-Yau variety. In discussing the Diophantine geometry of subvarieties on V, it is therefore sensible to focus on nondegenerate subvarieties. Combining results from Sections 5.1 and 5.2, we conclude that the set of nondegenerate integral points on X_k consists of finitely many mapping class group orbits. In Section 5.3, we discuss the context of our results in the setting of arithmetic hyperbolic surfaces.

Integral points on curves

Section 5.4 of this thesis is devoted to the analysis of integral points for nondegenerate algebraic curves on X_k. Our work is largely motivated by the corresponding theory for algebraic curves (and higher-dimensional varieties) on algebraic tori and abelian varieties, which we briefly recall below.

The use of log Calabi-Yau varieties as ambient spaces for Diophantine analysis of curves began in the 1929 work of Siegel (see [100] for an exposition), where he used the Abel-Jacobi maps of hyperbolic curves into Jacobians to prove the finiteness of integral points on the curves. Another instance is Skolem's 1935 approach (see for instance [9]) to solving the Thue equations by embedding them into algebraic tori defined by norm equations, which later enabled application of Baker's effective method using linear forms in logarithms. A powerful 1991 result of Faltings [30] (subsequently generalized to the semiabelian setting by Vojta [92]) states that any subvariety of an abelian variety over a number field contains only finitely many rational points, unless it contains a translate of an abelian subvariety.

Inspired by these results, we obtain a complete and effective analysis of integral points for nondegenerate algebraic curves on X_k. Our ingredients are the boundedness of nonarchimedean systoles obtained in Chapter 4 and known effective transcendental tools such as Runge's method and Baker's theory. The main result (Theorem 5.5) is that the set of integral points on a nondegenerate curve in X_k can be effectively determined, and is moreover finite unless the curve is invariant under a nontrivial Dehn multitwist action. Moreover, under additional hypotheses, the set of all imaginary quadratic integral points on the curve is shown to be finite. In this direction, our work extends the result of Long and Reid [60] on the finiteness of imaginary quadratic integral points for character curves of one cusped hyperbolic three-manifolds.

Finally, our work shows that, on any Markoff cubic surface defined by (1.1), the integral solutions to any Diophantine equation over $\mathbb{Z}$ can be effectively determined. By the work of Ghosh and Sarnak [36], almost all "admissible" such Markoff surfaces, in the sense of proportions, have a Zariski dense set of integral points.

Finite orbits

Our results on the moduli spaces X_k point to several possible directions of future research, especially concerning the refined Diophantine analysis of the integral points.

As a step in the study of strong approximation for reductions of integral points on X_k modulo congruences (which has been undertaken in the case $(g, n) = (1, 1)$ by Bourgain, Gamburd, and Sarnak [10]), in Chapter 6 we consider the finite orbits of the mapping class group on the moduli of local systems. Our main results (Theorem 6.A and 6.B) give a complete classification of the finite orbits when the surface has positive genus. Up to semisimplification, the finite orbits correspond precisely to the finite local systems when the surface has genus $g \geq 2$, and to the finite or special dihedral local systems when $g = 1$. Our work builds upon the results of Dubrovin and Mazzocco [27] on the case $(g, n) = (1, 1)$, which was motivated by the classification of algebraic solutions to a special family of Painlevé VI equations. As a byproduct of our work, we confirm a special case of a conjecture due to Kisin.

Chapter 2

Background

In this chapter, we collect relevant background on surfaces and their moduli of local systems. Aside from recording material common to the rest of this thesis, this chapter serves to set the main notations and conventions.

2.1 Surfaces

A *surface* is an oriented two dimensional smooth manifold, which we assume to be compact with at most finitely many boundary components unless otherwise indicated. A connected surface is said to have type (g, n) if it has genus g with n boundary components. A *curve* on a surface is an embedded copy of an unoriented circle, which we shall tacitly assume to be smooth in appropriate contexts. Given a surface Σ, we shall say that a curve $a \subset \Sigma$ is *nondegenerate* if it does not bound a disk, and *essential* if it is nondegenerate and is disjoint from, and not isotopic to, a boundary curve of Σ.

A *multicurve* on Σ is a finite union of disjoint curves on Σ. It is said to be *nondegenerate*, resp. *essential*, if each of its components is. A *pants decomposition* of Σ is an essential multicurve whose components are pairwise nonisotopic. If Σ is a surface of type (g, n) with $3g + n - 3 > 0$, then any pants decomposition of Σ consists

of $3g+n-3$ essential curves. Given a surface Σ and an essential multicurve $Q \subset \Sigma$, we denote by $\Sigma|Q$ the surface obtained by cutting Σ along the curves in Q. We shall say that an essential curve $a \subset \Sigma$ is *separating* if $\Sigma|a$ has a strictly greater number of connected components than Σ, and *nonseparating* otherwise.

2.1.1 Optimal generators

Let Σ be a surface of type (g,n), and choose a base point $x \in \Sigma$. We have the *standard presentation* of the fundamental group

$$\pi_1(\Sigma, x) = \langle \alpha_1, \beta_1', \cdots, \alpha_g, \beta_g', \gamma_1, \cdots, \gamma_n | [\alpha_1, \beta_1'] \cdots [\alpha_g, \beta_g'] \gamma_1 \cdots \gamma_n \rangle \tag{2.1}$$

where in particular $\gamma_1, \cdots, \gamma_n$ correspond to loops around the boundary curves of Σ. For $i = 1, \cdots, g$, let β_i be the based loop traversing β_i' in the opposite direction. We can choose the sequence of generating loops $(\alpha_1, \beta_1, \cdots, \alpha_g, \beta_g, \gamma_1, \cdots, \gamma_n)$ so that it satisfies the following:

(1) each loop in the sequence is simple,

(2) any two distinct loops in the sequence intersect exactly once (at x), and

(3) every product of distinct elements in the sequence preserving the cyclic ordering can be represented by a simple loop in Σ.

Some examples of products alluded to in (3) are $\alpha_1\beta_g$, $\alpha_1\alpha_2\beta_2\beta_g$, and $\beta_g\gamma_n\alpha_1$. We refer to $(\alpha_1, \beta_1, \cdots, \alpha_g, \beta_g, \gamma_1, \cdots, \gamma_n)$ as an *optimal sequence of generators* for $\pi_1\Sigma$. See Figure 2.1 for an illustration of optimal generators for $(g,n) = (2,1)$.

2.1.2 Mapping class group

Given a surface Σ, let $\Gamma = \Gamma(\Sigma) = \pi_0 \operatorname{Diff}^+(\Sigma, \partial\Sigma)$ denote its mapping class group. By definition, it is the group of isotopy classes of orientation preserving diffeomorphisms

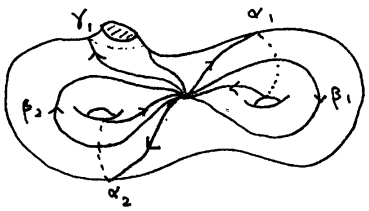


Figure 2.1: Optimal generators for $(g, n) = (2, 1)$

of Σ fixing the boundary of Σ pointwise. Given a (simple closed) curve $a \subset \Sigma$ disjoint from $\partial\Sigma$, the associated (left) *Dehn twist* $\tau_a \in \Gamma$ on Σ is defined as follows. Let $S^1 = \{z \in \mathbb{C} : |z| = 1\}$ be the unit circle. Let τ be the diffeomorphism from $S^1 \times [0, 1]$ to itself given by $(z, t) \mapsto (ze^{2\pi it}, t)$. Choose a closed tubular neighborhood N of a in Σ, and an orientation preserving diffeomorphism $f : N \to S^1 \times [0, 1]$. The Dehn twist τ_a is given by

$$\tau_a(x) = \begin{cases} f^{-1} \circ \tau \circ f(x) & \text{if } x \in N, \\ x & \text{otherwise.} \end{cases}$$

The class of τ_a in Γ is independent of the choices involved above, and depends only on the isotopy class of a. It is a standard fact that $\Gamma = \Gamma(\Sigma)$ is generated by Dehn twists along simple closed curves in Σ (see [31, Chapter 4]).

2.2 Character varieties

Throughout this thesis, an *algebraic variety* is a scheme of finite type over a field. Given an affine variety V over a given field k, we denote by $k[V]$ its coordinate ring over k. If moreover V is integral, then $k(V)$ denotes its function field over k. Given a commutative ring A with unity, the elements of A will be referred to as *regular functions* on the affine scheme $V = \operatorname{Spec} A$.

2.2.1 Character varieties of groups

Let π be a finitely generated group. Its (SL_2-)*representation variety* $\mathrm{Rep}(\pi)$ is the affine scheme determined by the functor

$$A \mapsto \mathrm{Hom}(\pi, \mathrm{SL}_2(A))$$

for every commutative ring A. Given a sequence of generators of π with m elements, we have a presentation of $\mathrm{Rep}(\pi)$ as a closed subscheme of SL_2^m defined by equations coming from relations among the generators. For each $a \in \pi$, let tr_a be the regular function on $\mathrm{Rep}(\pi)$ given by $\rho \mapsto \mathrm{tr}\,\rho(a)$.

The (SL_2-)*character variety* of π over $\mathbb{C}$ is the affine invariant theoretic quotient

$$X(\pi) = \mathrm{Rep}(\pi) /\!\!/ \mathrm{SL}_2 = \mathrm{Spec}\,\mathbb{C}[\mathrm{Rep}(\pi)]^{\mathrm{SL}_2(\mathbb{C})}$$

under the simultaneous conjugation action of SL_2. Note that the regular function tr_a for each $a \in \pi$ descends to a regular function on $X(\pi)$. Moreover, $X(\pi)$ has a natural model over $\mathbb{Z}$, defined as the spectrum of

$$R(\pi) = \mathbb{Z}[\mathrm{tr}_a : a \in \pi]/(\mathrm{tr}_1 - 2, \mathrm{tr}_a\,\mathrm{tr}_b - \mathrm{tr}_{ab} - \mathrm{tr}_{ab^{-1}}).$$

The relations in the above presentation arise from the fact that the trace of the 2×2 identity matrix is 2, and $\mathrm{tr}(A)\,\mathrm{tr}(B) = \mathrm{tr}(AB) + \mathrm{tr}(AB^{-1})$ for every $A, B \in \mathrm{SL}_2(\mathbb{C})$. Given an integral domain A with fraction field F of characteristic zero, the A-points of $X(\pi)$ parametrize the Jordan equivalence classes of $\mathrm{SL}_2(\overline{F})$-representations of π having character valued in A. We refer to [51], [74], [76] for details on the above statements and the following fact.

Fact 2.1. *If π is a group generated by $a_1, \cdots, a_m$, then $\mathbb{Q}[X(\pi)]$ is generated as a $\mathbb{Q}$-algebra by the collection $\{\mathrm{tr}_{a_{i_1} \cdots a_{i_k}} : 1 \le i_1 < \cdots < i_k \le m\}_{1 \le k \le 3}$.*

Example 2.2. We refer to Goldman [38] for details of examples below. Let F_m denote the free group on $m \geq 1$ generators $a_1, \cdots, a_m$.

(1) We have $\mathrm{tr}_{a_1} : X(F_1) \simeq \mathbb{A}^1$.

(2) We have $(\mathrm{tr}_{a_1}, \mathrm{tr}_{a_2}, \mathrm{tr}_{a_1a_2}) : X(F_2) \simeq \mathbb{A}^3$ by Fricke [38, Section 2.2].

(3) The coordinate ring $\mathbb{Q}[X(F_3)]$ is the quotient of the polynomial ring

$$\mathbb{Q}[\mathrm{tr}_{a_1}, \mathrm{tr}_{a_2}, \mathrm{tr}_{a_3}, \mathrm{tr}_{a_1a_2}, \mathrm{tr}_{a_2a_3}, \mathrm{tr}_{a_1a_3}, \mathrm{tr}_{a_1a_2a_3}, \mathrm{tr}_{a_1a_3a_2}]$$

by the ideal generated by two elements

$$\mathrm{tr}_{a_1a_2a_3} + \mathrm{tr}_{a_1a_3a_2} - (\mathrm{tr}_{a_1a_2}\,\mathrm{tr}_{a_3} + \mathrm{tr}_{a_1a_3}\,\mathrm{tr}_{a_2} + \mathrm{tr}_{a_2a_3}\,\mathrm{tr}_{a_1} - \mathrm{tr}_{a_1}\,\mathrm{tr}_{a_2}\,\mathrm{tr}_{a_3})$$

and

$$\begin{aligned}\mathrm{tr}_{a_1a_2a_3}\,\mathrm{tr}_{a_1a_3a_2} - \{&(\mathrm{tr}_{a_1}^2 + \mathrm{tr}_{a_2}^2 + \mathrm{tr}_{a_3}^2) + (\mathrm{tr}_{a_1a_2}^2 + \mathrm{tr}_{a_2a_3}^2 + \mathrm{tr}_{a_1a_3}^2)\\ &- (\mathrm{tr}_{a_1}\,\mathrm{tr}_{a_2}\,\mathrm{tr}_{a_1a_2} + \mathrm{tr}_{a_2}\,\mathrm{tr}_{a_3}\,\mathrm{tr}_{a_2a_3} + \mathrm{tr}_{a_1}\,\mathrm{tr}_{a_3}\,\mathrm{tr}_{a_1a_3})\\ &+ \mathrm{tr}_{a_1a_2}\,\mathrm{tr}_{a_2a_3}\,\mathrm{tr}_{a_1a_3} - 4\}.\end{aligned}$$

We record the following lemma, which is attributed by Goldman [38] to Vogt [90].

Lemma 2.3. *Given a finitely generated group π and $a_1, a_2, a_3, a_4 \in \pi$, we have*

$$\begin{aligned}2\mathrm{tr}_{a_1a_2a_3a_4} = &\,\mathrm{tr}_{a_1}\mathrm{tr}_{a_2}\mathrm{tr}_{a_3}\mathrm{tr}_{a_4} + \mathrm{tr}_{a_1}\mathrm{tr}_{a_2a_3a_4} + \mathrm{tr}_{a_2}\mathrm{tr}_{a_3a_4a_1} + \mathrm{tr}_{a_3}\mathrm{tr}_{a_4a_1a_2} + \mathrm{tr}_{a_4}\mathrm{tr}_{a_1a_2a_3}\\ &+ \mathrm{tr}_{a_1a_2}\mathrm{tr}_{a_3a_4} + \mathrm{tr}_{a_4a_1}\mathrm{tr}_{a_2a_3} - \mathrm{tr}_{a_1a_3}\mathrm{tr}_{a_2a_4}\\ &- \mathrm{tr}_{a_1}\mathrm{tr}_{a_2}\mathrm{tr}_{a_3a_4} - \mathrm{tr}_{a_3}\mathrm{tr}_{a_4}\mathrm{tr}_{a_1a_2} - \mathrm{tr}_{a_4}\mathrm{tr}_{a_1}\mathrm{tr}_{a_2a_3} - \mathrm{tr}_{a_2}\mathrm{tr}_{a_3}\mathrm{tr}_{a_4a_1}.\end{aligned}$$

2.2.2 Moduli of local systems on manifolds

Given a connected smooth (compact) manifold M, the moduli of local systems on M we shall study is the character variety $X(M) = X(\pi_1 M)$ of its fundamental group. More generally, given a smooth manifold $M = M_1 \sqcup \cdots \sqcup M_m$ with finitely many connected components M_i, we define

$$X(M) = X(M_1) \times \cdots \times X(M_m).$$

The construction of the moduli space $X(M)$ is functorial in the manifold M. Any smooth map $f : M \to N$ of manifolds induces a morphism $f^* : X(N) \to X(M)$, depending only on the homotopy class of f, given by pullback of local systems.

Let Σ be a surface. For each curve $a \subset \Sigma$, there is a well-defined regular function $\mathrm{tr}_a : X(\Sigma) \to X(a) \simeq \mathbb{A}^1$, which agrees with tr_α for any $\alpha \in \pi_1\Sigma$ freely homotopic to a parametrization of a. The boundary curves $\partial\Sigma$ of Σ induce a natural morphism

$$\mathrm{tr}_{\partial\Sigma} : X(\Sigma) \to X(\partial\Sigma) \simeq \mathbb{A}^n$$

where the latter isomorphism is given by a choice of ordering $\partial\Sigma = c_1 \sqcup \cdots \sqcup c_n$ of the boundary curves c_i. The fibers of $\mathrm{tr}_{\partial\Sigma}$ for $k \in \mathbb{A}^n$ will be denoted $X_k = X_k(\Sigma)$. Each X_k is often referred to as a *relative character variety* in the literature. If Σ is a surface of type (g, n) satisfying $3g + n - 3 > 0$, the relative character variety $X_k(\Sigma)$ is an irreducible algebraic variety of dimension $6g + 2n - 6$.

We shall often simplify our notation by combining parentheses where applicable, e.g. $X_k(\Sigma, \mathbb{Z}) = X_k(\Sigma)(\mathbb{Z})$. Given a fixed surface Σ, a subset $K \subseteq X(\partial\Sigma, \mathbb{C})$, and a subset $A \subseteq \mathbb{C}$, we shall denote by

$$X_K(A) = X_K(\Sigma, A)$$

the set of all $\rho \in X(\Sigma, \mathbb{C})$ such that $\mathrm{tr}_{\partial\Sigma}(\rho) \in K$ and $\mathrm{tr}_a(\rho) \in A$ for every essential curve $a \subset \Sigma$. The following lemma shows that there is no risk of ambiguity with this notation.

Lemma 2.4. *If A is a subring of $\mathbb{C}$ and $k \in A^n$, then X_k has a model over A and $X_k(A)$ recovers the set of A-valued points of X_k in the sense of algebraic geometry.*

Proof. Let A and $k \in A^n$ be as above. We have a model of X_k over A with coordinate ring $\mathrm{Spec}\, R(\pi_1\Sigma) \otimes_{\mathbb{Z}} A$. It is clear that an A-valued point in the sense of algebraic geometry corresponds to a point in $X_k(A)$. The converse follows from the observation, using the identity $\mathrm{tr}_a \mathrm{tr}_b = \mathrm{tr}_{ab} + \mathrm{tr}_{ab^{-1}}$, that tr_b for every $b \in \pi_1\Sigma$ can be written as a $\mathbb{Z}$-linear combination of products of traces tr_a for nondegenerate curves $a \subset \Sigma$. □

Similarly, given $k \in X(\partial\Sigma, \mathbb{C})$, a subvariety $V \subseteq X_k(\Sigma)$, and a subset $A \subseteq \mathbb{C}$, we shall denote

$$V(A) = \{\rho \in V(\mathbb{C}) : \mathrm{tr}\, \rho(a) \in A \text{ for every essential curve } a \subset \Sigma\}.$$

Given an immersion $\Sigma' \to \Sigma$ of surfaces, we have the associated restriction

$$(-)|_{\Sigma'} : X(\Sigma) \to X(\Sigma').$$

The mapping class group $\Gamma(\Sigma)$ acts naturally on $X(\Sigma)$ by pullback of local systems, preserving the integral structure as well as each relative character variety $X_k(\Sigma)$ and the sets $X_K(\Sigma, A)$ defined above. The dynamical aspects of this action on the complex points of $X(\Sigma)$ are not fully understood, but have been studied on certain special subloci. These include the locus of SU(2)-local systems (see [37]) on X, and the Teichmüller locus parametrizing *Fuchsian representations* associated to marked hyperbolic structures on Σ with geodesic boundary. This thesis is largely concerned with the descent properties of the dynamics on $X(\mathbb{C})$ beyond the classical setting.

2.3 Markoff type cubic surfaces

In this section, we give a description of the moduli spaces $X_k(\Sigma)$ and their mapping class group dynamics for $(g,n) = (1,1)$ and $(0,4)$. These cases are distinguished by the fact that each X_k is an affine cubic algebraic surface with an explicit equation.

2.3.1 Case $(g,n) = (1,1)$

Let Σ be a surface of type $(g,n) = (1,1)$, i.e. a one holed torus. Let (α, β, γ) be an optimal sequence of generators for $\pi_1\Sigma$. By Example 2.2.(2), we have an identification $(\mathrm{tr}_\alpha, \mathrm{tr}_\beta, \mathrm{tr}_{\alpha\beta}) : X(\Sigma) \simeq \mathbb{A}^3$. From the trace relations in Section 2.2, we obtain

$$\begin{aligned}
\mathrm{tr}_\gamma &= \mathrm{tr}_{\alpha\beta\alpha^{-1}\beta^{-1}} = \mathrm{tr}_{\alpha\beta\alpha^{-1}}\,\mathrm{tr}_{\beta^{-1}} - \mathrm{tr}_{\alpha\beta\alpha^{-1}\beta} \\
&= \mathrm{tr}_\beta^2 - \mathrm{tr}_{\alpha\beta}\,\mathrm{tr}_{\alpha^{-1}\beta} + \mathrm{tr}_{\alpha\alpha} = \mathrm{tr}_\beta^2 - \mathrm{tr}_{\alpha\beta}(\mathrm{tr}_{\alpha^{-1}}\,\mathrm{tr}_\beta - \mathrm{tr}_{\alpha\beta}) + \mathrm{tr}_\alpha^2 - \mathrm{tr}_1 \\
&= \mathrm{tr}_\alpha^2 + \mathrm{tr}_\beta^2 + \mathrm{tr}_{\alpha\beta}^2 - \mathrm{tr}_\alpha\,\mathrm{tr}_\beta\,\mathrm{tr}_{\alpha\beta} - 2.
\end{aligned}$$

Writing $(x,y,z) = (\mathrm{tr}_\alpha, \mathrm{tr}_\beta, \mathrm{tr}_{\alpha\beta})$ so that each of the variables x, y, and z corresponds to an essential curve on Σ as depicted in Figure 2.2, the moduli space $X_k \subset X$ has an explicit presentation as an affine cubic algebraic surface in $\mathbb{A}^3_{x,y,z}$ with equation

$$x^2 + y^2 + z^2 - xyz - 2 = k$$

as in (1.1). The mapping class group $\Gamma = \Gamma(\Sigma)$ acts on X_k via nonlinear polynomial transformations. For convenience, let a, b, and ab denote the essential curves lying in the free homotopy classes of loops α, β, and $\alpha\beta$, respectively. We have the following descriptions of the associated Dehn twist actions.

Lemma 2.5. *The Dehn twist actions τ_a, τ_b, and τ_{ab} on $X(\Sigma)$ are given by*

$$\tau_a^* : (x, y, z) \mapsto (x, z, xz - y),$$
$$\tau_b^* : (x, y, z) \mapsto (xy - z, y, x),$$
$$\tau_{ab}^* : (x, y, z) \mapsto (y, yz - x, z)$$

in terms of the above coordinates.

Proof. Note that $\tau_a(a)$ has the homotopy class of α, $\tau_a(b)$ has the homotopy class of $\alpha\beta$, and $\tau_a(ab)$ has the homotopy class of $\alpha\alpha\beta$. Noting that $\mathrm{tr}_{\alpha\alpha\beta} = \mathrm{tr}_\alpha \mathrm{tr}_{\alpha\beta} - \mathrm{tr}_\beta$, we obtain the desired expression for τ_a^*. The other Dehn twists are similar. □

Each of the morphisms τ_a^*, τ_b^*, and τ_{ab}^* is a composition of a Vieta involution, i.e. morphism of the form $(x, y, z) \mapsto (x, y, xy - z)$, with a transposition of two coordinates.

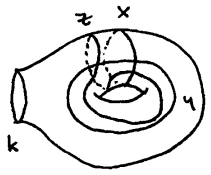


Figure 2.2: Curves on a surface of type $(1, 1)$ with corresponding functions

2.3.2 Case $(g, n) = (0, 4)$

Let Σ be a surface of type $(0, 4)$, i.e. a four holed sphere. Let $(\gamma_1, \cdots, \gamma_4)$ be an optimal sequence of generators for $\pi_1\Sigma$. Let us set $(x, y, z) = (\mathrm{tr}_{\gamma_1\gamma_2}, \mathrm{tr}_{\gamma_2\gamma_3}, \mathrm{tr}_{\gamma_1\gamma_3})$, so that each of the variables corresponds to an essential curve on Σ as depicted in Figure 2.3. By Example 2.2.(3), for $k = (k_1, k_2, k_3, k_4) \in \mathbb{A}^4(\mathbb{C})$ the relative character variety $X_k = X_k(\Sigma)$ is an affine cubic algebraic surface in $\mathbb{A}^3_{x,y,z}$ given by the equation

$$x^2 + y^2 + z^2 + xyz = Ax + By + Cz + D$$

with

$$\begin{cases} A = k_1k_2 + k_3k_4 \\ B = k_1k_4 + k_2k_3 \\ C = k_1k_3 + k_2k_4 \end{cases} \quad \text{and} \quad D = 4 - \sum_{i=1}^{4} k_i^2 - \prod_{i=1}^{4} k_i.$$

The mapping class group $\Gamma = \Gamma(\Sigma)$ acts on X_k via polynomial transformations. In particular, by an elementary argument as in Lemma 2.5, we see that Dehn twists give rise to morphisms

$$(x, y, z) \mapsto (x, B - xz - y, C - x(B - xz - y) - z),$$
$$(x, y, z) \mapsto (A - y(C - xy - z) - x, y, C - xy - z),$$
$$(x, y, z) \mapsto (A - yz - x, B - (A - yz - x)z - y, z).$$

Note that each transformation is a composition of two of the three Vieta involutions defined on X_k:

$$\tau_x : (x, y, z) \mapsto (A - yz - x, y, z),$$
$$\tau_y : (x, y, z) \mapsto (x, B - xz - y, z),$$
$$\tau_z : (x, y, z) \mapsto (x, y, C - xy - z).$$

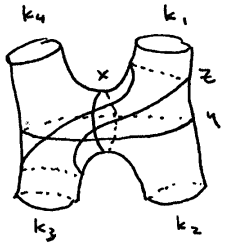


Figure 2.3: Curves on surfaces of type $(0, 4)$ with corresponding functions

Chapter 3

Global geometry

Let Σ be a surface of type (g, n) satisfying $3g + n - 3 > 0$. We assume in this chapter that Σ has nonempty boundary, i.e. $n \geq 1$. Throughout this chapter, we shall write $X_{g,n} = X(\Sigma)$ and $X_{g,n,k} = X_k(\Sigma)$ for the moduli spaces introduced in Chapter 2. The main result of this chapter is the following.

Theorem 3.1. *The variety $X_{g,n,k}$ is log Calabi-Yau for every $k \in \mathbb{A}^n(\mathbb{C})$.*

As in Chapter 1, we define an irreducible algebraic variety to be log Calabi-Yau if it has a normal projective compactification Z with reduced boundary divisor D such that the log canonical divisor is linearly trivial: $K_Z + D \sim 0$. We shall prove Theorem 3.1 by giving an algebraic construction of a compactification for $X_{g,n,k}$, which depends on the choice of free generators $\sigma = \{\alpha_1, \beta_1', \cdots, \alpha_g, \beta_g', \gamma_1, \cdots, \gamma_{n-1}\}$ under the standard presentation (2.1) of $\pi_1\Sigma$ given in Chapter 2. We also relate Theorem 3.1 to the following combinatorial result about curves on surfaces. Given a curve $a \subset \Sigma$, let $\text{length}_\sigma(a)$ be the minimum σ-word length of any $\alpha \in \pi_1\Sigma$ freely homotopic to a parametrization of a. This notion extends additively to multicurves on Σ. Consider the formal power series $Z_{g,n}(t) = \sum_{r=0}^{\infty} c_{g,n}(r)t^r$ where $c_{g,n}(r)$ denotes the number of isotopy classes of essential multicurves $a \subset \Sigma$ with $\text{length}_\sigma(a) = r$.

Theorem 3.2. *The series $Z_{g,n}(t)$ is rational, and satisfies $Z_{g,n}(1/t) = Z_{g,n}(t)$.*

We now outline the proofs of the above theorems and the contents of this chapter. Our proof of Theorem 3.1 proceeds in three steps: (i) construct a graded algebra whose Proj is a compactification of $X_{g,n,k}$, (ii) establish its algebraic properties, and (iii) convert these properties to geometric information about the projective variety. This graded algebra is constructed (with the choice σ above) using results from Section 3.1, where we introduce the notion of *word compactification* for the character variety of an arbitrary group with respect to a finite set of generators.

In Section 3.2, we show that the algebra constructed using Section 3.1 is a graded Gorenstein domain with an explicit canonical module, which (after applying works of Hochtster-Roberts [50] and Stanley [84]) amounts to a certain functional equation for the Hilbert series of the graded algebra. The latter is deduced from an elementary invariant theoretic work of Le Bruyn [57]. An analysis of singularities of the algebra (given separately in Section 3.3) shows that the algebra is moreover normal. We invoke results of Demazure [22] and Watanabe [94] to convert these algebraic properties to the geometry of our compactification, proving Theorem 3.1.

A result of Charles and Marché [15] states that the isotopy classes of nondegenerate multicurves on Σ provide a $\mathbb{C}$-linear basis of the coordinate ring of $X_{g,n}$. In Section 3.4, we use this result to establish an equivalence between the functional equation obtained in Section 3.2 and Theorem 3.2, proving the latter. In Section 3.5, we give another, purely combinatorial proof of Theorem 3.2 in the case Σ has genus zero, where the main input comes from properties of the associahedron. Appendix 3.A contains elementary computations in matrices relevant for the analysis in Section 3.3.

Given the explicit nature of our compactifications, the results of this chapter may be relevant to conjectures surrounding the boundary divisors of relative character varieties for surfaces, as formulated by Simpson [82] (verified therein for punctured spheres; see also [56]), or more generally of log Calabi-Yau varieties, cf. [55].

For Fock-Goncharov moduli spaces of local systems [33] which are related to, but

different from, our relative character varieties, log Calabi-Yau compactifications of their open subspaces can be constructed by general techniques of cluster varieties established by Gross-Hacking-Keel [44]; see also [45] for an elaboration of this point.

We mention another compactification of SL_2 (or more general) character varieties for free groups due to Manon [62], which uses Rees algebras associated to trivalent graphs, arising in the context of moduli spaces of principal bundles on closed Riemann surfaces and their degenerations. See also [63] for another compactification.

Conventions

Throughout this chapter, a ring or an algebra will always mean a commutative algebra with unity over the complex numbers $\mathbb{C}$, unless otherwise specified. All schemes will be defined over $\mathbb{C}$. Given an affine scheme X, we denote by $\mathbb{C}[X]$ its coordinate ring. Let $\mathbb{M}$ denote the affine scheme parametrizing 2×2 matrices. Let $\{x_{ij}\}_{i,j\in\{1,2\}}$ be the regular functions on $\mathbb{M}$ corresponding to the (i,j)th entries of a matrix, giving an identification $(x_{ij}) : \mathbb{M} \simeq \mathbb{A}^4$ with affine space. The standard matrix variable x and its adjugate x^* are the 2×2 matrices with coefficients in $\mathbb{C}[\mathbb{M}]$ given by

$$x = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}, \quad \text{and} \quad x^* = \begin{bmatrix} x_{22} & -x_{12} \\ -x_{21} & x_{11} \end{bmatrix}.$$

In particular, $\det(x)$ and $\mathrm{tr}(x)$ are regular functions on $\mathbb{M}$. Let $^* : \mathbb{M} \to \mathbb{M}$ denote the involution on $\mathbb{M}$ determined by $x \mapsto x^*$.

Given 2×2 matrices a and b, we shall denote $\langle a, b\rangle = aba^*b^*$ and $[a, b] = ab - ba$ in this chapter, to avoid confusion. Given complex numbers $t, k \in \mathbb{C}$, let $\mathbb{M}_t \subset \mathbb{M}$ be the subscheme of matrices with determinant t, and let $\mathbb{M}_{t,k} \subset \mathbb{M}_t$ be the subscheme of matrices with determinant t and trace k. We shall denote $\mathrm{SL}_2 = \mathbb{M}_1$ and $\mathrm{SL}_{2,k} = \mathbb{M}_{1,k}$. Let $\mathbf{1}$ be the identity matrix. We shall say that a matrix a is *scalar* if $a = \lambda\mathbf{1}$ for some $\lambda \in \mathbb{C}$, and *nonscalar* otherwise. As usual, let $\mathfrak{sl}_2(\mathbb{C})$ denote the tangent space

to SL_2 at $\mathbf{1}$. It is the space of traceless matrices in $\mathbb{M}_2(\mathbb{C})$.

Given an integer $m \geq 1$, consider $\mathfrak{C}_m = \mathbb{Z}/m\mathbb{Z}$ with its natural cyclic ordering. By a cyclic interval in $\mathfrak{C}_m$ we shall mean a sequence of the form $i_0, i_0+1, \cdots, i_0+k$ for some $i_0 \in \mathfrak{C}_m$ and for some $k \in \{0, \cdots, m-1\}$. If $|I| = k+1 < m$, then those elements of $\mathfrak{C}_m$ not contained in I together form a cyclic interval, denoted I^c.

3.1 Word compactifications

3.1.1 Graded rings

Given an affine scheme X with an action by the multiplicative group $\mathbb{G}_m$, its coordinate ring $A = \mathbb{C}[X]$ also carries a $\mathbb{G}_m$-action, or equivalently the structure of a graded ring via the $\mathbb{G}_m$-eigenspace decomposition:

$$A = \bigoplus_{r \in \mathbb{Z}} A_r, \quad \text{where} \quad A_r = \{a \in A : \lambda \cdot a = \lambda^{-r} a \text{ for all } \lambda \in \mathbb{G}_m(\mathbb{C})\}.$$

Conversely, the spectrum $\operatorname{Spec} A$ of a graded ring A carries a natural $\mathbb{G}_m$-action. Given a morphism of graded rings $B \to A$, we have an associated $\mathbb{G}_m$-equivariant morphism of schemes $\operatorname{Spec} A \to \operatorname{Spec} B$, and given $b \in \operatorname{Spec} B(\mathbb{C})$ and $\lambda \in \mathbb{G}_m(\mathbb{C})$ the fibers $(\operatorname{Spec} A)_b$ and $(\operatorname{Spec} A)_{\lambda \cdot b}$ are isomorphic via λ.

Example 3.3. For the grading of the polynomial ring $\mathbb{C}[t]$ by degree, the associated $\mathbb{G}_m$-action on the affine line $\mathbb{A}^1 = \operatorname{Spec} \mathbb{C}[t]$ is the usual one $\lambda \cdot a = \lambda a$ on points.

Let A be a graded algebra. An element $t \in A_1$ is non-nilpotent if and only if the morphism $\mathbb{C}[t] \to A$ is injective, or in other words the associated $\mathbb{G}_m$-equivariant morphism of schemes $\pi : X = \operatorname{Spec} A \to \mathbb{A}^1$ is dominant. Let $X_0 = \operatorname{Spec} A/(t)$ and $X_1 = \operatorname{Spec} A/(t-1)$ be the fibers of π over 0 and 1, respectively, and let $X|_{\mathbb{G}_m} = \pi^{-1}(\mathbb{G}_m)$ where $\mathbb{G}_m = \operatorname{Spec} \mathbb{C}[t, 1/t] \subset \mathbb{A}^1$ is the complement of 0.

Lemma 3.4. *Let A be a graded ring, and let $t \in A_1$ be non-nilpotent as above. We have a natural $\mathbb{G}_m$-equivariant isomorphism*

$$X_1 \times \mathbb{G}_m \simeq X|_{\mathbb{G}_m},$$

equipping X_1 with the trivial $\mathbb{G}_m$-action. In particular, $A/(t-1) \simeq A[1/t]_0$.

Proof. The morphisms $X_1 \times \mathbb{G}_m \to X|_{\mathbb{G}_m}$ and $X|_{\mathbb{G}_m} \to X_1 \times \mathbb{G}_m$ given at the level of points by $(a, \lambda) \mapsto \lambda \cdot a$ and $a \mapsto ((\pi(a))^{-1} \cdot a, \pi(a))$, respectively, are inverses to each other and are $\mathbb{G}_m$-equivariant. This proves the first part of the lemma, and gives us an isomorphism of graded rings $A/(t-1) \otimes \mathbb{C}[t, 1/t] \simeq A[1/t]$. Taking the degree 0 parts of both sides, the second assertion follows. □

Lemma 3.5. *Let A be a connected graded Cohen-Macaulay ring of pure dimension, and let $t \in A_1$ be non-nilpotent. Suppose that*

(1) *X_1 is a normal scheme, and*

(2) *X_0 is reduced of dimension* $\dim A - 1$.

Then A is a normal domain.

Proof. Since A is connected, if it is a normal ring then it is automatically a normal domain. Since A is Cohen-Macaulay, in order to show that A is normal it suffices by Serre's criterion for normality [66, Theorem 23.8, p.183] to show that A is regular in codimension 1.

Since t is not nilpotent, the $\mathbb{G}_m$-equivariant morphism $\pi : X \to \mathbb{A}^1$ is dominant, with isomorphic fibers of dimension $\dim A - 1$ over $\mathbb{G}_m$. By condition (2), in fact π has equidimensional fibers everywhere. By the miracle flatness theorem [66, Theorem 23.1, p.179], since A is Cohen-Macaulay and $\mathbb{C}[t]$ is regular it follows that A is flat over $\mathbb{C}[t]$. In particular, t is not a zero divisor in A.

By Lemma 3.4, we have $X|_{\mathbb{G}_m} \simeq X_1 \times \mathbb{G}_m$, which is regular in codimension 1 by condition (1). Thus, given any prime ideal $\mathfrak{p} \subset A$ of height ≤ 1 such that $t \notin \mathfrak{p}$, the localization $A_{\mathfrak{p}}$ is regular. Next, suppose that $\mathfrak{p} \subset A$ is a prime ideal of height 1 with $t \in \mathfrak{p}$. Since $t \in A$ is not a zero divisor, $A/(t)$ is Cohen-Macaulay of dimension $\dim A/(t) = \dim A - 1$. Hypothesis (2) implies that $(A/tA)_{\mathfrak{p}}$ is regular, and hence $A_{\mathfrak{p}}$ is regular. This proves the result. □

3.1.2 Rees algebras

By a *filtered algebra* (A, Fil) we mean an algebra A and an increasing filtration Fil of A by complex vector spaces, indexed by $\mathbb{Z}_{\geq 0}$, such that:

(1) $\mathbb{C} \subseteq \mathrm{Fil}_0 A$, and

(2) $\mathrm{Fil}_r \cdot \mathrm{Fil}_s \subseteq \mathrm{Fil}_{r+s}$ for every $r, s \geq 0$.

A morphism of filtered algebras $(B, \mathrm{Fil}) \to (A, \mathrm{Fil})$ is a morphism of algebras $B \to A$ mapping Fil_r into Fil_r for every $r \geq 0$. Given a $\mathbb{Z}_{\geq 0}$-graded algebra A, we have an associated filtered algebra (A, Fil) where Fil is given by $\mathrm{Fil}_r A = \bigoplus_{s=0}^{r} A_s$.

Example 3.6. For the grading of $\mathbb{C}[t]$ by degree, or more generally the grading of a polynomial ring $\mathbb{C}[x_1, \cdots, x_N]$ in several variables by (total) degree, we shall denote by $\mathrm{Fil}^{\mathrm{deg}}$ the associated filtration.

Given a filtered algebra (A, Fil), the Rees algebra A^{Fil} is the graded algebra

$$A^{\mathrm{Fil}} = \bigoplus_{r=0}^{\infty} \mathrm{Fil}_r \, .$$

A morphism of filtered algebras $(B, \mathrm{Fil}) \to (A, \mathrm{Fil})$ induces a morphism $B^{\mathrm{Fil}} \to A^{\mathrm{Fil}}$ of Rees algebras. Given a filtered algebra (A, Fil), we shall denote by t the element

of degree 1 in A^{Fil} corresponding to $1 \in \mathrm{Fil}_1 A = A_1^{\mathrm{Fil}}$. We have identifications

$$A^{\mathrm{Fil}}/(t-1) = A \quad \text{and} \quad A^{\mathrm{Fil}}/(t) = \mathrm{Gr}^{\mathrm{Fil}} A$$

where the latter preserves gradings. Note that $\operatorname{Spec} A^{\mathrm{Fil}}$ is connected if and only if $\operatorname{Spec} A_0^{\mathrm{Fil}}$ is connected. We have a natural identification $A = A^{\mathrm{Fil}}[1/t]_0$ by Lemma 3.4. Thus, if A^{Fil} is a finitely generated graded algebra, then $\operatorname{Proj} A^{\mathrm{Fil}}$ is a projective scheme containing $\operatorname{Spec} A$ as an open affine complement to the closed subscheme $\operatorname{Proj} \mathrm{Gr}^{\mathrm{Fil}} A$. For simplicity of notation, given a filtered algebra (A, Fil^x) where x is a symbol, we shall often write $A^x = A^{\mathrm{Fil}^x}$ for the Rees algebra.

Example 3.7. We have the following.

(1) Equipping $\mathbb{C}$ with the trivial filtration Fil given by $\mathrm{Fil}_r = \mathbb{C}$ for all $r \geq 0$, we have $\mathbb{C}^{\mathrm{Fil}} = \mathbb{C}[t]$ with grading by degree. For any filtered algebra (A, Fil), the obvious morphism $(\mathbb{C}, \mathrm{Fil}) \to (A, \mathrm{Fil})$ induces a $\mathbb{G}_m$-equivariant morphism of schemes $\pi : \operatorname{Spec} A^{\mathrm{Fil}} \to \mathbb{A}^1$.

(2) Given a filtered algebra (B, Fil) and a surjective algebra homomorphism $\varphi : B \to A$, we have an induced filtration Fil on A defined by $\mathrm{Fil}_r = \varphi(\mathrm{Fil}_r)$ for every $r \geq 0$ making φ a morphism of filtered algebras. We thus have an induced morphism $\varphi : B^{\mathrm{Fil}} \to A^{\mathrm{Fil}}$.

(3) For $(A = \mathbb{C}[x_1, \cdots, x_N], \mathrm{Fil}^{\deg})$ the polynomial ring with the filtration by degree, let $A^{\deg}$ be the Rees algebra. We then have an isomorphism $A^{\deg} \simeq \mathbb{C}[X_1, \cdots, X_N, T]$ given by

$$f(x_1, \cdots, x_N) \mapsto T^r f(X_1/T, \cdots, X_N/T)$$

for each $f \in \mathrm{Fil}_r^{\deg} A = A_r^{\deg}$. The projective space $\operatorname{Proj} A^{\deg} \simeq \mathbb{P}^N$ is the usual compactification of the affine space $\operatorname{Spec} A = \mathbb{A}^N$.

Lemma 3.8. *Given a filtered algebra* (A, Fil)*, we have the following.*

(1) A^{Fil} *is an integral domain if and only if* A *is.*

(2) *If* A^{Fil} *is a normal domain, then so is* A*.*

Proof. (1) If A^{Fil} is an integral domain, then so is $A = A^{\mathrm{Fil}}[1/t]_0$. Conversely, suppose that A is an integral domain. If $ab = 0$ in A^{Fil}, then replacing a and b by nonzero terms (if any) of highest degree we may assume that a and b are homogeneous of degree r and s, respectively. We then have $ab = 0 \in A^{\mathrm{Fil}}_{r+s} = \mathrm{Fil}_{r+s}$ within A, and hence $a = 0$ or $b = 0$. This proves that A^{Fil} is an integral domain.

(2) Given a normal $\mathbb{Z}$-graded domain B, its degree 0 component B_0 is also a normal domain. Indeed, it is obvious that B_0 is a domain; to see that B_0 is also normal, note that if $x \in B$ is integral over B_0 then we must have $x \in B_0$ by degree reasons. Applying this to $A = A^{\mathrm{Fil}}[1/t]_0$, we see that if A^{Fil} is a normal domain then so is A. □

3.1.3 Word compactifications

Let π be a finitely generated group. Throughout this chapter, we shall write $\mathrm{Rep}_\pi = \mathrm{Rep}(\pi)$ and $X_\pi = X(\pi)$ for the representation variety and the character variety of π, respectively. We shall also write R_π for the coordinate ring of X_π. Recall that a *length function* on π is a function $\ell : \pi \to \mathbb{R}_{\geq 0}$ satisfying the conditions:

(1) $\ell(1) = 0$,

(2) $\ell(g^{-1}) = \ell(g)$ for every $g \in \pi$, and

(3) $\ell(gh) \leq \ell(g) + \ell(h)$ for every $g, h \in \pi$.

A length function equips π with a pseudo-metric via the formula $d_\ell(g, h) = \ell(gh^{-1})$. Given a generating set $\sigma \subseteq \pi$, we have the σ-word length function $\mathrm{length}_\sigma : \pi \to \mathbb{Z}_{\geq 0}$

which is given by $\text{length}_\sigma(a) = \min\{r : a = w_1 \cdots w_r \text{ for some } w_i \in \sigma \cup \sigma^{-1}\}$. Let Fil^σ be the increasing filtration of the coordinate ring R_π of the character variety of π given by

$$\text{Fil}_r^\sigma R_\pi = \text{Span}\{\text{tr}_a : \text{length}_\sigma(a) \leq r\}.$$

We have $\mathbb{C} = \mathbb{C} \cdot \text{tr}_1 = \text{Fil}_0 R_\pi$, and the relations $\text{tr}_a \text{tr}_a = \text{tr}_{ab} + \text{tr}_{ab^{-1}}$ on R_π show that $\text{Fil}_r^\sigma \cdot \text{Fil}_s^\sigma \subseteq \text{Fil}_{r+s}^\sigma$ for every $r, s \in \mathbb{Z}_{\geq 0}$. Thus, $(R_\pi, \text{Fil}^\sigma)$ is a filtered algebra. Let R_π^σ denote the associated Rees algebra.

Definition 3.9. The *σ-word compactification* of X_π is the scheme $X_\pi^\sigma = \text{Proj}\, R_\pi^\sigma$.

Lemma 3.10. *Let π be a group generated by a finite set σ.*

(1) *We have $(R_\pi^\sigma)_0 = \mathbb{C}$. For each $r \geq 1$, we have $\dim_{\mathbb{C}}(R_\pi^\sigma)_r < \infty$.*

(2) *R_π^σ is finitely generated by homogeneous elements of degree ≤ 3.*

In particular, X_π^σ is projective and can be presented as a closed subscheme of some weighted projective space $\mathbb{P}(w_1, \cdots, w_s)$ with $w_i \in \{1, 2, 3\}$.

Proof. (1) We have $\text{length}_\sigma(g) = 1$ if and only if $g = 1$, and there are at most finitely many σ-words of length $\leq r$ in π for any given r.

(2) Applying Lemma 2.3 and induction, we see that the class $\text{tr}_w \in R_\pi$ of every word w of length $r \geq 4$ in π is a linear combination of terms of the form $\text{tr}_{w_1} \cdots \text{tr}_{w_s}$ with $\sum_{i=1}^s \text{length}_\sigma(w_i) \leq r$ and $\text{length}_\sigma(w_i) \leq 3$ for each $i = 1, \cdots, s$. This shows that R_π^σ is generated by $(R_\pi^\sigma)_1 \oplus (R_\pi^\sigma)_2 \oplus (R_\pi^\sigma)_3$. Applying part (1) of this lemma, we obtain the desired result. □

3.1.4 An alternate description

Following the notation from the beginning of this chapter, let $\mathbb{M}$ be the scheme parametrizing 2×2 matrices. Using the isomorphism $(x_{ij}) : \mathbb{M} \simeq \mathbb{A}^4$ via matrix

entries, we equip $\mathbb{C}[\mathbb{M}^m]$ for each $m \geq 1$ with grading by degree and associated filtration $\mathrm{Fil}^{\deg}$. We have SL_2 acting on $\mathbb{M}^m$ by simultaneous conjugation, preserving the grading (and hence the filtration) on $\mathbb{C}[\mathbb{M}^m]$. Therefore, SL_2 acts on the Rees algebra $\mathbb{C}[\mathbb{M}^m]^{\deg}$ of $(\mathbb{C}[\mathbb{M}^m], \mathrm{Fil}^{\deg})$ preserving the grading. By Example 3.7.(3), we have an identification

$$\mathbb{C}[\mathbb{M}^m]^{\deg} = \mathbb{C}[(X_1)_{ij}, \cdots, (X_m)_{ij}, T]$$

with the polynomial ring on $4m+1$ generators, so that $\operatorname{Spec} \mathbb{C}[\mathbb{M}^m]^{\deg} = \mathbb{M}^m \times \mathbb{A}^1$. The SL_2-action is the product of the conjugation action on $\mathbb{M}^m$ and the trivial action on $\mathbb{A}^1$. The morphism $\pi : \mathbb{M}^m \times \mathbb{A}^1 \to \mathbb{A}^1$ from Example 3.7.(1) is just the projection onto the second factor.

Let π be a group with finite generating set σ and $|\sigma| = m$. The choice of σ gives us a SL_2-equivariant closed immersion $\sigma^* : \mathrm{Rep}_\pi \hookrightarrow \mathrm{Rep}_m = \mathrm{SL}_2^m \subset \mathbb{M}^m$, where $\mathrm{Rep}_m = \mathrm{Rep}_{F_m}$ is the representation variety of the free group F_m on m generators. Via the surjective morphism $\mathbb{C}[\mathbb{M}^m] \to \mathbb{C}[\mathrm{Rep}_\pi]$, the filtration $\mathrm{Fil}^{\deg}$ on $\mathbb{C}[\mathbb{M}^m]$ induces a filtration on $\mathbb{C}[\mathrm{Rep}_\pi]$ which we shall denote Fil^σ.

Proposition 3.11. *We have* $\mathrm{Fil}^\sigma_r R_\pi = (\mathrm{Fil}^\sigma_r \mathbb{C}[\mathrm{Rep}_\pi])^{\mathrm{SL}_2}$ *for every* $r \geq 0$. *In particular, there is a surjective morphism of graded rings*

$$\Phi_\sigma : (\mathbb{C}[\mathbb{M}^m]^{\deg})^{\mathrm{SL}_2} \to R^\sigma_\pi.$$

Proof. We begin by considering the filtration $\mathrm{Fil}^{\deg}$ on $\mathbb{C}[\mathbb{M}^m]$. For each $r \geq 0$, a close inspection of the work of Procesi [73] (Theorems 1.1, 1.2, and 1.3 *loc.cit.*) shows that the space of invariants $(\mathbb{C}[\mathbb{M}^m]_r)^{\mathrm{SL}_2}$ is spanned by elements of the form

$$\mathrm{tr}(w_1)\,\mathrm{tr}(w_2)\cdots\mathrm{tr}(w_s)$$

where each w_i is a noncommutative monomial in the standard matrix variables $x_1, \cdots, x_m$ on $\mathbb{M}^m$ such that we have $\sum_{i=1}^{s} \deg(w_i) = r$. Here, the degree $\deg(w_i)$ refers to the total multiplicity of matrices $x_1, \cdots, x_m$ appearing in the monomial w_i. Since SL_2 is linearly reductive, we see that the map

$$(\mathrm{Fil}_r^{\deg} \mathbb{C}[\mathbb{M}^m])^{\mathrm{SL}_2} \to (\mathrm{Fil}_r^{\sigma} \mathbb{C}[\mathrm{Rep}_\pi])^{\mathrm{SL}_2},$$

obtained from the surjective morphism

$$\mathrm{Fil}_r^{\deg} \mathbb{C}[\mathbb{M}^m] \to \mathrm{Fil}_r^{\sigma} \mathbb{C}[\mathrm{Rep}_\pi]$$

of finite-dimensional complex representations of SL_2, remains surjective. In particular, $(\mathrm{Fil}_r^{\sigma} \mathbb{C}[\mathrm{Rep}_\pi])^{\mathrm{SL}_2}$ is spanned by elements of the form

$$\mathrm{tr}_{w_1} \mathrm{tr}_{w_2} \cdots \mathrm{tr}_{w_s}$$

where each w_i is a product of elements of σ (with multiplicities, but without inverses) such that the sum total number of elements appearing is $\leq r$. This shows that $(\mathrm{Fil}_r^{\sigma} \mathbb{C}[\mathrm{Rep}_\pi])^{\mathrm{SL}_2} \subseteq \mathrm{Fil}^{\sigma} R_\pi$ since each element of the above form lies in $\mathrm{Fil}^{\sigma} R_\pi$, recalling that the filtration Fil^{σ} is compatible with multiplication on R_π. The other containment $\mathrm{Fil}^{\sigma} R_\pi \subseteq (\mathrm{Fil}^{\sigma} \mathbb{C}[\mathrm{Rep}_\pi])^{\mathrm{SL}_2}$ follows from the observation that, using the relation $\mathrm{tr}_a \mathrm{tr}_a = \mathrm{tr}_{ab} + \mathrm{tr}_{ab^{-1}}$ in R_π, one can write any tr_a with $\mathrm{length}_\sigma(a) \leq r$ as a linear combination of elements of the above form. Finally, since the surjection $\mathbb{C}[\mathbb{M}^m] \to \mathbb{C}[\mathrm{Rep}_\pi]$ is compatible with the filtrations and the SL_2-actions, we obtain a surjection

$$(\mathbb{C}[\mathbb{M}^m]^{\deg})^{\mathrm{SL}_2} \to (\mathbb{C}[\mathrm{Rep}_\pi]^{\sigma})^{\mathrm{SL}_2} = (\mathbb{C}[\mathrm{Rep}_\pi]^{\mathrm{SL}_2})^{\sigma} = R_\pi^{\sigma}$$

which proves the second part of the proposition. □

3.2 Log Calabi-Yau property

3.2.1 Free groups

Let F_m be the free group on $m \geq 1$ generators σ. As before, let $\mathrm{Rep}_m = \mathrm{SL}_2^m$ be the representation variety of F_m, and let $X_m = \mathrm{Rep}_m /\!\!/ \mathrm{SL}_2$ be the SL_2 character variety of F_m with $R_m = \mathbb{C}[\mathrm{Rep}_m]^{\mathrm{SL}_2}$. Let Fil^σ be the filtration on the coordinate ring $\mathbb{C}[\mathrm{Rep}_m]$ of $\mathrm{Rep}_m = \mathrm{SL}_2^m$ defined in Section 3.1.4. In other words, it is the filtration obtained by the projection of $\mathrm{Fil}^{\mathrm{deg}}$ under the surjective ring homomorphism $\mathbb{C}[\mathbb{M}^m] \to \mathbb{C}[\mathrm{Rep}_m]$. For $i \in \{1, \cdots, m\}$, let us write $E_i = \det(X_i) - T^2$ for the homogeneous element in $\mathbb{C}[\mathbb{M}^m]^{\mathrm{deg}}$ of degree 2, with the identification $\mathbb{C}[\mathbb{M}^m]^{\mathrm{deg}} = \mathbb{C}[(X_1)_{ij}, \cdots, (X_m)_{ij}, T]$ from Section 3.1.4.

Lemma 3.12. *We have the following.*

(1) $\mathbb{C}[\mathbb{M}^m]^{\mathrm{deg}}/(E_1, \cdots, E_m)$ *is a normal graded Cohen-Macaulay domain.*

(2) $E_1, \cdots, E_m$ *is a regular sequence in* $\mathbb{C}[\mathbb{M}^m]^{\mathrm{deg}}$.

(3) *We have* $\mathbb{C}[\mathrm{Rep}_m]^\sigma \simeq \mathbb{C}[\mathbb{M}^m]/(E_1, \cdots, E_m)$.

Proof. Let us write $\mathrm{Rep}'_m = \mathrm{Spec}\, \mathbb{C}[\mathbb{M}^m]^{\mathrm{deg}}/(E_1, \cdots, E_m)$. Our proof of (1) and (2) will proceed by induction on m. Both statements are clear when $m = 1$ (we note by the Jacobian criterion that the singular locus of Rep'_1 consists of a single point). So let $m \geq 2$. For any $s \in \{1, \cdots, m\}$, note that we have

$$\mathbb{C}[\mathbb{M}^m]^{\mathrm{deg}}/(E_1, \cdots, E_s) = \mathbb{C}[\mathrm{Rep}'_s] \otimes \mathbb{C}[(X_{s+1})_{ij}, \cdots, (X_m)_{ij}].$$

In particular, each

$$\mathbb{C}[\mathbb{M}^m]^{\mathrm{deg}}/(E_1, \cdots, E_s)$$

is an integral domain, showing that $E_1, \cdots, E_m$ is a regular sequence in $\mathbb{C}[\mathbb{M}^m]^{\deg}$, and $\mathbb{C}[\mathrm{Rep}'_m]$ is a graded Cohen-Macaulay ring of pure dimension $3m+1$. Consider the $\mathbb{G}_m$-equivariant morphism

$$\pi : \mathrm{Rep}'_m \to \mathbb{A}^1$$

associated to the morphism of graded rings $\mathbb{C}[t] \to \mathbb{C}[\mathrm{Rep}'_m]$ given by $t \mapsto T$. The fiber $\pi^{-1}(1) = \mathrm{SL}_2^m$ is smooth, and $\pi^{-1}(0) = \mathbb{M}_0^m$ is reduced of dimension $3m$. We thus conclude by Lemma 3.5 that $\mathbb{C}[\mathrm{Rep}'_m]$ is a normal domain. This completes the induction, and we have proven (1) and (2). It remains to prove (3). Note that we have a natural surjective ring homomorphism $\mathbb{C}[\mathrm{Rep}'_m] \to \mathbb{C}[\mathrm{Rep}_m]^\sigma$. Note that $\mathbb{C}[\mathrm{Rep}_m]^\sigma$ is integral by Lemma 3.8, and has dimension at least $3m+1$. Thus, we see that $\mathbb{C}[\mathrm{Rep}'_m] \to \mathbb{C}[\mathrm{Rep}_m]^\sigma$ must be an isomorphism, as the former is also integral of dimension $3m+1$. This completes the proof of the lemma. □

Let Fil^σ be the σ-word filtration on R_m defined in Section 3.1, and let R_m^σ be the associated Rees algebra.

Proposition 3.13. *The ring R_m^σ is a normal graded Cohen-Macaulay domain.*

Proof. First, by a result of Hochster-Roberts [50, Main Theorem], we see that the ring $(\mathbb{C}[\mathbb{M}^m]^{\deg})^{\mathrm{SL}_2}$ is Cohen-Macaulay. We next claim that the surjective homomorphism $\Phi_\sigma : (\mathbb{C}[\mathbb{M}^m]^{\deg})^{\mathrm{SL}_2} \to R_m^\sigma$ constructed in Proposition 3.11 induces an isomorphism

$$(\mathbb{C}[\mathbb{M}^m]^{\deg})^{\mathrm{SL}_2}/(E_1, \cdots, E_m) \simeq R_m^\sigma.$$

Indeed, since $(\mathbb{C}[\mathbb{M}^m]^{\deg})^{\mathrm{SL}_2}$ is a pure subring of $\mathbb{C}[\mathbb{M}^m]^{\deg}$, the left hand side is isomorphic to $(\mathbb{C}[\mathbb{M}^m]^{\deg}/(E_1, \cdots, E_m))^{\mathrm{SL}_2} = (\mathbb{C}[\mathrm{Rep}_m]^{\deg})^{\mathrm{SL}_2} = R_m^\sigma$ by Lemma 3.12, which is the desired result. Since $\mathbb{C}[\mathrm{Rep}_m]^{\deg}$ is a normal graded domain by Lemma 3.12, so is R_m^σ. To show that R_m^σ is Cohen-Macaulay, it suffices to show that $E_1, \cdots, E_m$ is a regular sequence in $(\mathbb{C}[\mathbb{M}^m]^{\deg})^{\mathrm{SL}_2}$. But this follows from the fact that $E_1, \cdots, E_m$ is

a regular sequence in $\mathbb{C}[\mathbb{M}^m]^{\deg}$ and that $(\mathbb{C}[\mathbb{M}^m]^{\deg})^{\mathrm{SL}_2}$ is a pure subring of $\mathbb{C}[\mathbb{M}^m]^{\deg}$. The desired result follows. □

The Hilbert series $H_m(t)$ of the graded algebra R_m^σ is by definition the following formal power series. Note that $H_m(t)$ is well-defined by Lemma 3.10.

$$H_m(t) = \sum_{r=0}^{\infty} (\dim_{\mathbb{C}} (R_m^\sigma)_r) t^r = \sum_{r=0}^{\infty} (\dim_{\mathbb{C}} \mathrm{Fil}_r^\sigma R_m) t^r \in \mathbb{Z}[[t]].$$

Theorem 3.14. *Assume that $m \geq 2$. Then $H_m(t)$ is rational and satisfies*

$$H_m(1/t) = (-1)^{3m-2} t^{2m+1} H_m(t).$$

In particular, R_m^σ is Gorenstein with canonical module $R_m^\sigma(-2m-1)$.

Proof. For $m = 2$ and F_2 a free group on generators $\sigma = \{a_1, a_2\}$, we may verify the functional equation by direct computation as follows. Recall the isomorphism $(\mathrm{tr}_{a_1}, \mathrm{tr}_{a_2}, \mathrm{tr}_{a_1 a_2}) : X_2 \simeq \mathbb{A}^3$ from Example 2.2. It follows $R_2^\sigma \simeq \mathbb{C}[X_1, X_2, X_3, T]$ with each X_1, X_2, T given degree 1 and X_3 given degree 2, so that the Krull dimension of R_2^σ is 4 and

$$H_2(t) = \frac{1}{(1-t)^3(1-t^2)}$$

from which the desired symmetry follows. For $m \geq 3$, we have the following result of Le Bruyn [57] (based on rational expressions due to Weyl and Schur, cf. *loc.cit.*): the Hilbert series $h_m(t)$ of the graded algebra $\mathbb{C}[\mathbb{M}^m]^{\mathrm{SL}_2}$ has the functional equation $h_m(1/t) = -t^{4m} h_m(t)$. Our Hilbert series $H_m(t)$ is related to the series $h_m(t)$ by

$$H_m(t) = \frac{(1-t^2)^m}{1-t} h_m(t),$$

where the factor $(1-t^2)^m/(1-t)$ arises as we take the SL_2-invariants of

$$\mathbb{C}[\mathbb{M}^m]^{\deg}/(E_1, \cdots, E_m)$$

instead of $\mathbb{C}[\mathbb{M}^m]$, and the fact that $E_1, \cdots, E_m$ is a regular sequence of homogeneous elements of degree 2 in $\mathbb{C}[\mathbb{M}^m]^{\deg}$. Thus, the functional equation of $H_m(t)$ follows from that of $h_m(t)$. Now, the Krull dimension of R_m^σ is $(3m-3)+1$. Hence, our last claim follows from the functional equation and Proposition 3.13, by the result of Stanley [84, Theorem 4.4] (see also [8, Corollary 4.4.6, p.177]). □

3.2.2 Relative character varieties

Let $\Sigma_{g,n}$ be a compact oriented surface of genus $g \geq 0$ with $n \geq 1$ boundary components such that $\chi(\Sigma_{g,n}) = 2-2g-n < 0$. We fix a standard presentation of the fundamental group

$$\pi_1(\Sigma_{g,n}) = \langle a_1, \cdots, a_{2g+n} | [a_1, a_2] \cdots [a_{2g-1}, a_{2g}] a_{2g+1} \cdots a_{2g+n} \rangle.$$

By our assumption, $\pi_1(\Sigma_{g,n})$ is a free group of rank $m = 2g+n-1 \geq 2$. We shall refer to the set $\sigma = \{a_1, \cdots, a_{2g+n-1}\}$ of free generators as *standard*.

Let $\mathrm{Rep}_{g,n} = \mathrm{Rep}_{\pi_1(\Sigma_{g,n})}$ be the representation variety of $\Sigma_{g,n}$ (or of $\pi_1(\Sigma_{g,n})$), and let $X_{g,n} = \mathrm{Rep}_{g,n} /\!\!/ \mathrm{SL}_2$ be the character variety of $\Sigma_{g,n}$. Note that σ gives us isomorphisms

$$\sigma^* : \mathrm{Rep}_{g,n} \simeq \mathrm{Rep}_m = \mathrm{SL}_2^m,$$
$$\sigma^* : X_{g,n} \simeq X_m$$

where Rep_m and X_m are as in Section 3.2.1. For each $k = (k_1, \cdots, k_n) \in \mathbb{A}^n(\mathbb{C})$, the

corresponding fibers $\mathrm{Rep}_{g,n,k}$ and $X_{g,n,k}$ of the morphisms

$$(\mathrm{tr}(x_{2g+1}), \cdots, \mathrm{tr}(x_{2g+n-1}), \mathrm{tr}(x_{2g+n})) : \mathrm{Rep}_{g,n} \to \mathbb{A}^n,$$
$$(\mathrm{tr}_{a_{2g+1}}, \cdots, \mathrm{tr}_{a_{2g+n}}) : X_{g,n} \to \mathbb{A}^n$$

(where $x_1, \cdots, x_{2g+n-1}$ are the matrix variables of SL_2^m and we abbreviated here $x_{2g+n} = \langle x_1, x_2 \rangle \cdots \langle x_{2g-1}, x_{2g} \rangle x_{2g+1} \cdots x_{2g+n-1}$ for simplicity) are called the relative representation variety and relative character variety of $\Sigma_{g,n}$. Let us denote by $R_{g,n,k} = \mathbb{C}[X_{g,n,k}]$ the coordinate ring of the relative character variety.

Example 3.15. We review the presentations of some relative character varieties for surfaces of small Euler characteristic, following Goldman [38].

(1) $(g, n) = (0, 3)$. By Example 2.2.(2), we have $(\mathrm{tr}_{a_1}, \mathrm{tr}_{a_2}, \mathrm{tr}_{a_1 a_2}) : X_{0,3} \simeq \mathbb{A}^3$. Since the 3 boundary components of $\Sigma_{0,3}$ correspond to a_1, a_2, and $a_1 a_2$, we see that $X_{0,3,k}$ is simply a point for every $k \in \mathbb{A}^3(\mathbb{C})$.

(2) $(g, n) = (1, 1)$. By Example 2.2.(2), we have

$$(\mathrm{tr}_{a_1}, \mathrm{tr}_{a_2}, \mathrm{tr}_{a_1 a_2}) : X_{1,1} \simeq \mathbb{A}^3_{(x,y,z)}$$

where (x, y, z) denotes the sequence of standard coordinate functions on $\mathbb{A}^3$. The boundary component of $\Sigma_{1,1}$ corresponds to $\langle a_1, a_2 \rangle$, which defines the function given by the expression

$$[\langle a_1, a_2 \rangle] = f(x, y, z) = x^2 + y^2 + z^2 - xyz - 2$$

obtained using the identities $\mathrm{tr}_a \mathrm{tr}_b = \mathrm{tr}_{ab} + \mathrm{tr}_{ab^{-1}}$ in the coordinate ring R_2. Thus, $X_{1,1,k}$ for $k \in \mathbb{A}^1(\mathbb{C})$ is an affine cubic surface which is given by a level set $f^{-1}(k)$ in $\mathbb{A}^3$ of the above function, also called a (generalized) Markoff surface.

Looking at the partial derivatives of the above function, its critical locus is given by the conditions

$$2x = yz, \quad 2y = xz, \quad \text{and} \quad 2z = xy.$$

The crticial locus is thus $\{(0,0,0), (s_1 2, s_2 2, s_3 2) : s_1 s_2 s_3 = 1, s_i \in \{\pm 1\}\}$, and every relative character variety $X_{1,1,k}$ is normal.

(3) $(g,n) = (0,4)$. Using Example 2.2.(3), given $k = (k_1, \cdots, k_4) \in \mathbb{A}^4(\mathbb{C})$ we have a closed immersion

$$(\mathrm{tr}_{a_1 a_2}, \mathrm{tr}_{a_2 a_3}, \mathrm{tr}_{a_1 a_3}) : X_{0,4} \hookrightarrow \mathbb{A}^3_{(x,y,z)}$$

with the image given by the affine cubic surface

$$f_k(x,y,z) = x^2 + y^2 + z^2 + xyz - ax - by - cz - d = 0$$

where we define $a = k_1 k_2 + k_3 k_4$, $b = k_2 k_3 + k_1 k_4$, $c = k_1 k_3 + k_2 k_4$, and $d = 4 - k_1^2 - k_2^2 - k_3^2 - k_4^2 - k_1 k_2 k_3 k_4$. The critical locus of the function f is given by the conditions

$$2x + yz - a = 0, \quad 2y + xz - b = 0, \quad 2z + xy - c = 0.$$

We see that the locus $\subset \mathbb{A}^3$ defined by the above three equations is finite. Hence, every relative character variety $X_{0,4,k}$ is normal.

From Section 3.1, we have the σ-word filtration Fil^σ on the coordinate ring $R_{g,n}$ with Rees algebra $R^\sigma_{g,n} = (\mathbb{C}[\mathrm{Rep}_{g,n}]^\sigma)^{\mathrm{SL}_2}$. Let $F_1, \cdots, F_{n-1}$ and F_n respectively be

the homogeneous elements in $R^{\sigma}_{g,n}$ of degree $1, \cdots, 1$ and $4g+n-1$ given by

$$F_1 = \mathrm{tr}_{a_{2g+1}} - k_1 t, \quad \cdots, \quad F_{n-1} = \mathrm{tr}_{a_{2g+n-1}} - k_{n-1} t, \quad \text{and}$$
$$F_n = \mathrm{tr}_{a_{2g+n}} - k_n t^{4g+n-1}.$$

Here, F_n is homogeneous since $\mathrm{length}_{\sigma}(a_{2g+n}) = 4g+n-1$. Let us define the quotient $R^{\sigma}_{g,n,k} = R_{g,n}/(F_1, \cdots, F_n)$. Note that $R^{\sigma}_{g,n,k}/(t-1) = R_{g,n,k}$. The pair

$$(Z, D) = (\mathrm{Proj}\, R^{\sigma}_{g,n,k}, \mathrm{Proj}\, R^{\sigma}_{g,n,k}/(t))$$

is our projective compactification and boundary divisor of $X_{g,n,k}$ mentioned in Theorem 3.1. The goal of this section is to verify that (Z, D) satisfies the desired properties.

Theorem 3.16. *The ring $R^{\sigma}_{g,n,k}$ is a normal graded Gorenstein domain of dimension $6g+2n-5$, with canonical module $R^{\sigma}_{g,n,k}(-1)$.*

Proof. We have an identification $R^{\sigma}_{g,n} = R^{\sigma}_{m}$. We claim that $F_1, \cdots, F_n$ is a regular sequence in $R^{\sigma}_{m} = (\mathbb{C}[\mathrm{Rep}_m]^{\sigma})^{\mathrm{SL}_2}$. Indeed, using the fact that $(\mathbb{C}[\mathbb{M}^m]^{\deg})^{\mathrm{SL}_2}$ is a pure subring of $\mathbb{C}[\mathbb{M}^m]^{\deg}$, it suffices to note that $\mathbb{C}[\mathrm{Rep}_m]^{\sigma}/(F_1, \cdots, F_s)$ is an integral domain (of successively lower dimension) for $1 \leq s \leq n-1$. This follows from an easy application of Lemma 3.5 (as in the proof of Lemma 3.12). Now, by Theorem 3.14, R^{σ}_{m} is a graded Gorenstein ring of dimension $3m-2$ with canonical module $R^{\sigma}_{m}(-2m-1)$. Hence, it follows that

$$R^{\sigma}_{g,n,k} = R^{\sigma}_{g,n}/(F_1, \cdots, F_n)$$

is a graded Gorenstein ring of dimension $3m-2-n = 6g+2n-5$ with canonical module

$$R^{\sigma}_{g,n,k}(-2m-1+(n-1+4g+n-1)) = R^{\sigma}_{g,n,k}(-1).$$

It remains to show that $R^{\sigma}_{g,n,k}$ is a normal domain. By Lemma 3.5, it suffices to show the following:

(1) $R_{g,n,k}$ is a normal ring, and

(2) $R^{\sigma}_{g,n,k}/(t)$ is reduced of dimension $6g+2n-6$.

If $\chi(\Sigma_{g,n}) \leq -3$, then (1) and (2) follow from Propositions 3.25 and 3.27. Let us now consider the cases where $|\chi(\Sigma_{g,n})| \leq 2$. If $\chi(\Sigma_{g,n}) = -2$, then (1) and (2) follow from Propositions 3.26 and 3.27 for $(g,n) = (1,2)$, and they follow from Example 3.15.(3) and Proposition 3.27 for $(g,n) = (0,4)$. The case $\chi(\Sigma_{g,n}) = -1$ remains. The case $(g,n) = (0,3)$ is trivial, and for $(g,n) = (1,1)$ Example 3.15.(2) verifies (1). It thus remains to verify (2) for $(g,n) = (1,1)$. But we find that

$$R^{\sigma}_{1,1,k}/(t) \simeq \mathbb{C}[x,y,z]/(z(z-xy)),$$

whose spectrum is the union of the plane $\{z = 0\}$ with the surface $\{z = xy\}$ in $\mathbb{A}^3_{(x,y,z)}$, and hence is reduced and of dimension 2. This completes the proof. □

Let Z be the compactification of $X_{g,n,k}$ as above. Note that Z is a normal irreducible projective scheme, since $R^{\sigma}_{g,n,k}$ is a normal domain by Theorem 3.16. By the result of Demazure [22, Théorème (3.5) p.51], there is an ample $\mathbb{Q}$-divisor E on Z such that

$$R^{\sigma}_{g,n,k} = \bigoplus_{r=0}^{\infty} H^0(Z, \mathcal{O}_Z(rE))t^r$$

where $t \in (R^{\sigma}_{g,n,k})_1$ is the image of $t \in (R^{\sigma}_{g,n})_1$ under the projection $R^{\sigma}_{g,n} \to R^{\sigma}_{g,n,k}$. Following Demazure (*loc.cit.*), the $\mathbb{Q}$-divisor E is given explicitly as follows. Let us write $\mathrm{Div}(t) = \sum p_F F$ for the Weil divisor on $\mathrm{Spec}\, R^{\sigma}_{g,n,k}$ determined by t. Note that each F is given by $\mathrm{Spec}\, B_F$ where B_F is a graded integral quotient of $R^{\sigma}_{g,n,k}$. Defining $q_F \geq 0$ by

$$(B_F)_n \neq 0 \iff n \in q_F\mathbb{Z},$$

the $\mathbb{Q}$-divisor E on Z is given by

$$E = \sum_{q_F \neq 0} \frac{p_F}{q_F} \operatorname{Proj}(B_F).$$

Lemma 3.17. *We have $p_F \in \{0, 1\}$ and $q_F \in \{0, 1, 2\}$ for every prime Weil divisor F of* $\operatorname{Spec} R^{\sigma}_{g,n,k}$.

Proof. The claim $p_F \in \{0, 1\}$ follows from the fact that

$$\operatorname{Spec}(R^{\sigma}_{g,n,k}/t) = (\mathbb{C}[\operatorname{Rep}'_{g,n,k}]/(t))^{\mathrm{SL}_2}$$

is reduced by Proposition 3.27. To see that $q_F \in \{0, 1, 2\}$, note that B_F is generated in degrees 1, 2, and 3 by Lemma 3.10 so we must have $q_F \in \{0, 1, 2, 3\}$. Suppose toward contradiction that $q_F = 3$, so that $(B_F)_1 = (B_F)_2 = 0$. Given any $a_1, a_2, a_3 \in \pi_1(\Sigma_{g,n})$, on $R_{g,n,k}$ we have the relations

$$\operatorname{tr}_{a_1a_2a_3} + \operatorname{tr}_{a_1a_3a_2} = \operatorname{tr}_{a_1a_2} \operatorname{tr}_{a_3} + \operatorname{tr}_{a_1a_3} \operatorname{tr}_{a_2} + \operatorname{tr}_{a_2a_3} \operatorname{tr}_{a_1} - \operatorname{tr}_{a_1} \operatorname{tr}_{a_2} \operatorname{tr}_{a_3}$$

and

$$\begin{aligned} \operatorname{tr}_{a_1a_2a_3} \operatorname{tr}_{a_1a_3a_2} = {} & (\operatorname{tr}^2_{a_1} + \operatorname{tr}^2_{a_2} + \operatorname{tr}^2_{a_3}) + (\operatorname{tr}^2_{a_1a_2} + \operatorname{tr}^2_{a_2a_3} + \operatorname{tr}^2_{a_1a_3}) \\ & - (\operatorname{tr}_{a_1} \operatorname{tr}_{a_2} \operatorname{tr}_{a_1a_2} + \operatorname{tr}_{a_2} \operatorname{tr}_{a_3} \operatorname{tr}_{a_2a_3} + \operatorname{tr}_{a_1} \operatorname{tr}_{a_3} \operatorname{tr}_{a_1a_3}) \\ & + \operatorname{tr}_{a_1a_2} \operatorname{tr}_{a_2a_3} \operatorname{tr}_{a_1a_3} - 4. \end{aligned}$$

Since the right hand sides of both equations are zero on B_F which is an integral domain, this shows that $\operatorname{tr}_{a_1a_2a_3} = \operatorname{tr}_{a_1a_3a_2} = 0$ on B_F. This shows that we must have $(B_F)_3 = 0$ as well, a contradiction. Hence, we must have $q_F \in \{0, 1, 2\}$. □

Let $D = Z \setminus X_{g,n,k}$ be the reduced boundary divisor on Z. As a Weil divisor, D is a formal sum $\sum_V V$ of prime divisors V where $V = \operatorname{Proj} B_F$ runs over integral

quotients B_F of $R^{\sigma}_{g,n,k}/(t)$ by minimal homogeneous prime ideals. We are ready to prove the main part of Theorem 3.1, restated as follows.

Theorem 3.18. *We have $K_Z + D \sim 0$.*

Proof. Combining Corollary 3.16 with the result of Watanabe [94, Corollary 2.9], we find that $K + E' + E = \mathrm{Div}(f)$ for some $f \in \mathbb{C}(Z)$ (the function field of Z) and

$$E' = \sum_{q_F \neq 0} \frac{q'_F - 1}{q'_F} \operatorname{Proj} B_F,$$

where we have set $q'_F = q_F / \gcd(p_F, q_F)$ for $q_F \neq 0$, so that in particular $q'_F = 1$ if $p_F = 0$. But in light of Lemma 3.17, we have

$$\begin{aligned} 0 &\sim K + E' + E \\ &= K + \sum_{\substack{p_F=1 \\ q_F=2}} \frac{1}{2} \operatorname{Proj} B_F + \left\{ \sum_{\substack{p_F=1 \\ q_F=2}} \frac{1}{2} \operatorname{Proj} B_F + \sum_{\substack{p_F=1 \\ q_F=1}} \operatorname{Proj} B_F \right\} \\ &= K + D \end{aligned}$$

from which we conclude the result. □

3.3 Analysis of singularities

3.3.1 Introduction

Let $m \geq 1$ be an integer. Recall from Section 3.1.4 that we have an identification

$$\mathbb{C}[\mathbb{M}^m]^{\deg} = \mathbb{C}[(X_1)_{ij}, \cdots, (X_m)_{ij}, T] = \mathbb{C}[\mathbb{M}^m \times \mathbb{A}^1].$$

As before, for $i = 1, \cdots, m$, let $E_i = \det(X_i) - T^2$ be the element of degree 2 in $\mathbb{C}[\mathbb{M}^m]^{\deg}$. Recall that $\mathbb{C}[\mathrm{Rep}_m]^{\sigma} = \mathbb{C}[\mathbb{M}^m]^{\deg}/(E_1, \cdots, E_m)$, where Rep_m is the

representation variety of the free group on m generators σ.

Let $g \geq 0$ and $n \geq 0$ be integers with $2g+n-1 = m$. Fix $k = (k_1, \cdots, k_n) \in \mathbb{A}^n(\mathbb{C})$. Let $F_1, \cdots, F_{n-1}$ and F_n respectively be the homogeneous elements in $\mathbb{C}[\mathbb{M}^m]^{\deg}$ of degrees $1, \cdots, 1$ and $4g+n-1$ given by

$$F_1 = \mathrm{tr}(X_{2g+1}) - k_1 T, \quad \cdots, \quad F_{n-1} = \mathrm{tr}(X_{2g+n-1}) - k_{n-1} T, \quad \text{and}$$
$$F_n = \mathrm{tr}(\langle X_1, X_2\rangle \cdots \langle X_{2g-1}, X_{2g}\rangle X_{2g+1} \cdots X_{2g+n-1}) - k_n T^{4g+n-1}.$$

Let us define $\mathrm{Rep}'_m = \mathrm{Spec}\, \mathbb{C}[\mathrm{Rep}_m]^\sigma$ and $\mathrm{Rep}'_{g,n,k} = \mathrm{Spec}\, \mathbb{C}[\mathrm{Rep}'_m]/(F_1, \cdots, F_n)$. The singular locus of $\mathrm{Rep}'_{g,n,k}$ is the intersection of $\mathrm{Rep}'_{g,n,k}$ with the locus of critical points $\mathrm{Crit}(E, F)$ of the morphism $(E_1, \cdots, E_m, F_1, \cdots, F_n) : \mathbb{M}^m \times \mathbb{A}^1 \to \mathbb{A}^{m+n}$. Consider now the $\mathbb{G}_m$-equivariant morphism

$$\pi : \mathrm{Rep}'_{g,n,k} \to \mathbb{A}^1$$

of affine schemes induced by the the morphism of graded rings $\mathbb{C}[t] \to \mathbb{C}[\mathrm{Rep}'_{g,n,k}]$ sending $t \mapsto T$. Note that $\pi^{-1}(1) \simeq \mathrm{Rep}_{g,n,k}$ is the relative representation variety of the compact surface $\Sigma_{g,n}$ of genus g with n boundary components corresponding to k, as introduced in Section 3.2 (at least when $n \geq 1$). Let us define $H_{g,n} = \pi^{-1}(0)$. Note that it is independent of the choice of k.

3.3.2 Critical points

Fix integers $g \geq 0$ and $n \geq 1$ such that $m = 2g + n - 1 \geq 1$. Fix a complex number $t \in \mathbb{C}$. For each $s \in \mathbb{C}$ such that $s^2 = t$, the fiber of the morphism (considered in the proof of Lemma 3.12)

$$\pi : \mathrm{Rep}'_m \to \mathbb{A}^1$$

above s is the scheme $\mathbb{M}_t^m$. By definition, a point $a \in \mathbb{M}_t^m(\mathbb{C})$ is represented by an m-tuple of matrices $a = (a_1, \cdots, a_m)$, each a_i having determinant t. Given such a point, let us introduce the following supplementary notation.

(1) Let $b_i = \langle a_{2i-1}, a_{2i} \rangle$ for each $i \in \{1, \cdots, g\}$.

(2) Let $b_{g+j} = a_{2g+j}$ for each $j \in \{1, \cdots, n-1\}$.

(3) In generality, for a cyclic interval $I = (i_0, i_0+1, \cdots, i_0+k)$ in $\mathfrak{C}_{g+n-1}$ (cf. Section 1.4), let $b_I = b_{i_0} \cdots b_{i_0+k}$ where the indices are considered modulo $g+n-1$. Also, let $b^I = b_{I^c}$ when I^c is defined. Let us write $b^i = b^{(i)}$ for simplicity.

For example, we have

$$b_{(1,\cdots,g+n-1)} = b_1 \cdots b_{g+n-1} = \langle a_1, a_2 \rangle \cdots \langle a_{2g-1}, a_{2g} \rangle a_{2g+1} \cdots a_{2g+n-1}.$$

We shall attempt to describe, in terms of these parameters, the critical locus of the restriction to $\mathbb{M}_t^m \subset \mathrm{Rep}'_m$ of the morphism $F = (F_1, \cdots, F_n) : \mathrm{Rep}'_m \to \mathbb{A}^n$ defined in Section 3.3.1. We begin with a lemma on tangent spaces.

Lemma 3.19. *Let $a \in \mathbb{M}(\mathbb{C})$ be given.*

(1) $T_a\mathbb{M}_{\det a} = \{v \in \mathbb{M}(\mathbb{C}) : \mathrm{tr}(va^*) = 0\}$. *We have $\mathfrak{sl}_2(\mathbb{C}) \cdot a \subseteq T_a\mathbb{M}_{\det a}$ and $a \cdot \mathfrak{sl}_2(\mathbb{C}) \subseteq T_a\mathbb{M}_{\det a}$. The containments are equalities if $\det(a) \neq 0$.*

(2) $T_a\mathbb{M}_{\det a, \mathrm{tr}\, a} = \{v \in \mathbb{M}(\mathbb{C}) : \mathrm{tr}(v) = \mathrm{tr}(va^*) = 0\}$. *We have a containment $[\mathfrak{sl}_2(\mathbb{C}), a] \subseteq T_a\mathbb{M}_{\det a, \mathrm{tr}\, a}$ which is an equality if a is nonscalar.*

Proof. (1) For any $a, v \in \mathbb{M}(\mathbb{C})$, we have $\det(a + \varepsilon v) = \det(a) + \varepsilon \,\mathrm{tr}(va^*)$ within $\mathbb{C}[\varepsilon]/(\varepsilon^2)$, from which the first statement follows. In particular, $\dim T_a\mathbb{M}_{\det a} = 3$ provided that $a \neq 0$. Since $\mathrm{tr}(uaa^*) = \mathrm{tr}(aua^*) = \det(a)\,\mathrm{tr}(u) = 0$ for every $u \in \mathfrak{sl}_2(\mathbb{C})$, we have $\mathfrak{sl}_2(\mathbb{C}) \cdot a \subseteq T_a\mathbb{M}_{\det a}$ and $a \cdot \mathfrak{sl}_2(\mathbb{C}) \subseteq T_a\mathbb{M}_{\det a}$. Lastly, if $\det(a) \neq 0$

then $\dim \mathfrak{sl}_2(\mathbb{C}) \cdot a = \dim a \cdot \mathfrak{sl}_2(\mathbb{C}) = 3$ from which the equalities follow by dimension reasons.

(2) For any $a, v \in \mathbb{M}(\mathbb{C})$ we have $\operatorname{tr}(a + \varepsilon v) = \operatorname{tr}(a) + \varepsilon \operatorname{tr}(v)$ within $\mathbb{C}[\varepsilon]/(\varepsilon^2)$, from which the first statement follows. We have $\operatorname{tr}([v, a]a^*) = \operatorname{tr}(vaa^* - ava^*) = \det(a)\operatorname{tr}(v) - \det(a)\operatorname{tr}(v) = 0$ and $\operatorname{tr}([v, a]) = 0$ for any $v \in \mathfrak{sl}_2(\mathbb{C})$, and hence $[\mathfrak{sl}_2(\mathbb{C}), a] \subseteq T_{a,\mathrm{Char}}$. Finally, if a is nonscalar, then we have $\dim_{\mathbb{C}} T_a \mathbb{M}_{\det a, \operatorname{tr} a} = 2$ since $v \mapsto \operatorname{tr}(va^*)$ and $w \mapsto \operatorname{tr}(v)$ are linearly independent on $\mathbb{M}(\mathbb{C})$. On the other hand, $\dim_{\mathbb{C}}[\mathfrak{sl}_2(\mathbb{C}), a] = 2$ if a is nonscalar, and hence $T_a \mathbb{M}_{\det a, \operatorname{tr} a} = [\mathfrak{sl}_2(\mathbb{C}), a]$ by dimension reasons. □

Lemma 3.20. *A critical point $(a_1, \cdots, a_m)$ of $F : \mathbb{M}_t^m \to \mathbb{A}^n$ must satisfy at least one of the following two conditions.*

(1) *a_{2g+j} is scalar for some $j \in \{1, \cdots, n-1\}$.*

(2) *$[b_i, b^i] = 0$ for every $i \in \{1, \cdots, g+n-1\}$, and*

$$\begin{aligned}[a_{2i-1}, a_{2i}a_{2i-1}^*a_{2i}^*b^i] &= [a_{2i}, a_{2i-1}^*a_{2i}^*b^ia_{2i-1}] \\ &= [a_{2i-1}^*, a_{2i}^*b^ia_{2i-1}a_{2i}] = [a_{2i}^*, b^ia_{2i-1}a_{2i}a_{2i-1}^*] = 0\end{aligned}$$

for every $i \in \{1, \cdots, g\}$.

Proof. Assume that $a \in \mathbb{M}_t^m(\mathbb{C})$ satisfies neither of the conditions (1) and (2). We must show that $d_aF : T_a\mathbb{M}_t^m \to \mathbb{C}^n$ is surjective. Now, the first $n-1$ coordinates of d_aF are given by

$$(u_1, \cdots, u_m) \mapsto (\operatorname{tr}(u_{2g+1}), \cdots, \operatorname{tr}(u_{2g+n-1}))$$

for each $u = (u_1, \cdots, u_m) \in T_a\mathbb{M}_t^m$. Since a_{2g+j} is nonscalar for $j \in \{1, \cdots, n-1\}$ by hypothesis, each of the functionals $u \mapsto \operatorname{tr}(u_{2g+j})$ is nonzero. Thus, to show surjectivity of d_aF it remains to show that $(0, \cdots, 0, 1) \in \operatorname{Im} d_aF$.

Let $u \in \mathfrak{sl}_2(\mathbb{C})$ be given. We then have $[u, b_{g+j}] \in T_{b_{g+j}}\mathbb{M}_t$ for $j \in \{1, \cdots, n-1\}$, and the composition $T_{b_{g+j}}\mathbb{M}_t \hookrightarrow T_a\mathbb{M}_t^m \to \mathbb{C}^n$ with d_aF sends

$$[u, b_{g+j}] \mapsto (0, \cdots, 0, \operatorname{tr}([u, b_{g+j}]b^{g+j})) = (0, \cdots, 0, \operatorname{tr}(u[b_{g+j}, b^{g+j}])).$$

Similarly, after a short computation (using the fact $u^* = -u$ for $u \in \mathfrak{sl}_2(\mathbb{C})$) we also find that for $i \in \{1, \cdots, g\}$ the composition $T_{(a_{2i-1}, a_{2i})}\mathbb{M}_t^2 \hookrightarrow T_a\mathbb{M}_t^m \to \mathbb{C}^n$ with d_aF sends

$$\begin{aligned}
(ua_{2i-1} - a_{2i-1}u, -a_{2i}u) &\mapsto (0, \cdots, 0, \operatorname{tr}(u[a_{2i-1}, a_{2i}a_{2i-1}^*a_{2i}^*b^i])),\\
(a_{2i-1}u, 0) &\mapsto (0, \cdots, 0, \operatorname{tr}(u[a_{2i}, a_{2i-1}^*a_{2i}^*b^ia_{2i-1}])),\\
(0, a_{2i}u) &\mapsto (0, \cdots, 0, \operatorname{tr}(u[a_{2i-1}^*, a_{2i}^*b^ia_{2i-1}a_{2i}])),\\
(-a_{2i-1}u, ua_{2i} - a_{2i}u) &\mapsto (0, \cdots, 0, \operatorname{tr}(u[a_{2i}^*, b^ia_{2i-1}a_{2i}a_{2i-1}^*])).
\end{aligned}$$

In particular, summing up the above tangent vectors in $T_a\mathbb{M}_t^m$ we find that

$$(ua_{2i-1} - a_{2i-1}u, ua_{2i} - a_{2i}u) \mapsto (0, \cdots, 0, \operatorname{tr}(u[b_i, b^i])).$$

Thus, if a does not satisfy condition (2), at least one of the expressions above must be nonzero for some $u \in \mathfrak{sl}_2(\mathbb{C})$. This implies that $(0, \cdots, 0, 1) \in \operatorname{Im} d_aF$, and hence d_aF is surjective, as desired. □

3.3.3 Dimension estimates

We collect a number of estimates for the dimensions of certain schemes parametrizing sequences of matrices with prescribed conditions. We will later use them to analyze the singularities of $\operatorname{Rep}_{g,n,k}$ and $H_{g,n}$. Throughout Section 3.3.3, fix integers $g \geq 0$ and $n \geq 0$ such that $m = 2g + n - 1 \geq 1$.

Fix $k = (k_1, \cdots, k_n) \in \mathbb{A}^n(\mathbb{C})$, and let $\Phi : \operatorname{SL}_2^{2g} \times \prod_{i=1}^n \operatorname{SL}_{2,k_i} \to \operatorname{SL}_2$ be the

morphism given by

$$(a_1, \cdots, a_{2g+n}) \mapsto \langle a_1, a_2\rangle \cdots \langle a_{2g-1}, a_{2g}\rangle a_{2g+1} \cdots a_{2g+n}.$$

Lemma 3.21. *Φ is flat above $\mathrm{SL}_2 \setminus \{\pm\mathbf{1}\}$ with fibers of dimension $6g + 2n - 3$.*

Proof. First, Φ is dominant by Lemma 3.40. Note that $\mathrm{SL}_2^{2g} \times \prod_{i=1}^n \mathrm{SL}_{2,k_i}$ is a complete intersection scheme of dimension $6g + 2n$, and in particular Cohen-Macaulay. The target SL_2 of Φ is regular of dimension 3. By the miracle flatness theorem, to prove that Φ is flat it suffices to show that the fibers of Φ are equidimensional, of dimension $6g + 2n - 3$. Consider the composition $\mathrm{tr} \circ \Phi : \mathrm{SL}_2^{2g} \times \prod_{i=1}^n \mathrm{SL}_{2,k_i} \to \mathbb{A}^1$. Since $\mathrm{tr} \circ \Phi$ is dominant, it is flat. In particular, the fibers of $\mathrm{tr} \circ \Phi$ are all of pure dimension $6g + 2n - 1$.

Next, note that SL_2 acts by conjugation on the domain and target of Φ, and Φ is equivariant with respect to this action. Hence, given any $l \in \mathbb{C}$ the fibers of the restriction $\Phi : \Phi^{-1}(\mathrm{SL}_{2,l}) \to \mathrm{SL}_{2,l}$, away from the fibers over $\pm\mathbf{1}$, are isomorphic hence equidimensional of dimension $6g + 2n - 3$. □

Lemma 3.22. *If $2g + n - 1 \geq 2$, then $\Phi^{-1}(\pm\mathbf{1})$ has dimension at most $6g + 2n - 3$.*

Proof. Let us denote by F_+ (reps. F_-) the fiber of Φ above $\mathbf{1}$ (resp. $-\mathbf{1}$). We shall proceed by induction on (g, n). Consider first the case where $n = 0$. We must then have $g \geq 2$. Consider the projection

$$\pi : F_\pm \subset \mathrm{SL}_2^{2g} \to \mathrm{SL}_2^2$$

onto the last two copies of SL_2. Given $(b_1, b_2) \in \mathrm{SL}_2^2(\mathbb{C})$, the dimension of the fiber $\pi^{-1}(b_1, b_2)$ is estimated as follows.

(1) If $\langle b_1, b_2\rangle \neq \pm\mathbf{1}$, then $\dim \pi^{-1}(b_1, b_2) = 6(g-1) - 3$ by Lemma 3.21.

(2) If $\langle b_1, b_2\rangle = \pm\mathbf{1}$, then $\dim \pi^{-1}(b_1, b_2) \leq 6(g-1) - 2$.

Here, part (2) follows from the inductive hypothesis and the fact that, when $g = 2$ so that $g - 1 = 1$, the fiber of $\langle -, -\rangle : \mathrm{SL}_2^2 \to \mathrm{SL}_2$ above $\pm\mathbf{1}$ has dimension at most 4 by Lemma 3.41. Since the locus of $(b_1, b_2) \in \mathrm{SL}_2^2$ defined by part (1) is 6-dimensional, and the locus defined by condition (2) is at most 4-dimensional by Lemma 3.41, we see that $\dim F_\pm \leq 6g - 3$ as desired. Consider next the case where $n \geq 1$. If $g = 0$, then we must have $n \geq 3$. Consider the projection

$$\pi : F_\pm \subset \mathrm{SL}_2^{2g} \times \prod_{i=1}^{n} \mathrm{SL}_{2,k_i} \to \mathrm{SL}_{2,k_n}$$

onto the last factor of the product. Given $b \in \mathrm{SL}_{2,k_n}(\mathbb{C})$, the dimension of the fiber $\pi^{-1}(b)$ is estimated as follows.

(1) If $b \neq \pm\mathbf{1}$, then $\dim \pi^{-1}(b) = 6g + 2(n-1) - 3$ by the Lemma 3.21.

(2) If $b = \pm\mathbf{1}$, then $\dim \pi^{-1}(b) \leq 6g + 2(n-1) - 2$.

Here, part (2) follows from the inductive hypothesis, from Lemma 3.41 in the case $(g, n) = (1, 1)$, and from the observation that the fiber of $\mathrm{SL}_{2,l} \times \mathrm{SL}_{2,\pm l} \to \mathrm{SL}_2$, $(a_1, a_2) \mapsto a_1 a_2$ over $\pm\mathbf{1}$ has dimension 2 in the case $(g, n) = (0, 3)$. Since the locus of b satisfying (1) has dimension 2, we find that $\dim F_\pm \leq 6g + 2n - 3$ as desired. □

Let $\Psi : \mathbb{M}_0^{2g} \times \mathbb{M}_{0,0}^n \to \mathbb{M}_0$ be the morphism given by

$$(a_1, \cdots, a_{2g+n}) \mapsto \langle a_1, a_2\rangle \cdots \langle a_{2g-1}, a_{2g}\rangle a_{2g+1} \cdots a_{2g+n}.$$

Lemma 3.23. *Ψ is flat above $\mathbb{M}_0 \setminus \{0\}$ with fibers of dimension $6g + 2n - 3$.*

Proof. First, Ψ is dominant by Lemma 3.45. Note that $\mathbb{M}_0^{2g} \times \mathbb{M}_{0,0}^n$ is a complete intersection scheme of dimension $6g + 2n$, and in particular Cohen-Macaulay. The target $\mathbb{M}_0$ of Ψ is regular away from $\{0\}$, of dimension 3. By the miracle flatness

theorem, to prove that Ψ is flat over $\mathbb{M}_0 \setminus \{0\}$ it suffices to show that the fibers of Ψ are equidimensional, of dimension $6g+2n-3$. Consider the composition $\mathrm{tr} \circ \Psi_{g,n} : \mathbb{M}_0^{2g} \times \mathbb{M}_{0,0}^n \to \mathbb{A}^1$. Since $\mathrm{tr} \circ \Psi$ is dominant, it is flat. In particular, the fibers of $\mathrm{tr} \circ \Psi$ are all of pure dimension $6g+2n-1$.

Next, note that SL_2 acts by conjugation on the domain and target of Ψ, and Ψ is equivariant with respect to this action. Hence, given any $l \in \mathbb{C}$ the fibers of the restriction $\Psi : \Psi^{-1}(\mathbb{M}_{0,l}) \to \mathbb{M}_{0,l}$ are isomorphic, hence equidimensional of dimension $6g+2n-3$, away from the fiber over 0. □

Lemma 3.24. *$\Psi^{-1}(0)$ has dimension at most $6g+2n-1$.*

Proof. Since the domain $\mathbb{M}_0^{2g} \times \mathbb{M}_{0,0}^n$ of Ψ has dimension $6g+2n$ and is integral, if $\dim \Psi^{-1}(0) \geq 6g+2n$ then Ψ must be identically zero, contradicting the fact that Ψ is dominant by Lemma 3.45. □

3.3.4 Singularities

Let $\Sigma_{g,n}$ be a compact oriented surface of genus $g \geq 0$ with $n \geq 1$ boundary components and $\chi(\Sigma_{g,n}) = 2-2g-n < 0$. The fundamental group of $\Sigma_{g,n}$ is free of rank $m = 2g+n-1 \geq 2$. Let $k = (k_1, \cdots, k_n) \in \mathbb{A}^n(\mathbb{C})$, and let $F_1, \cdots, F_n \in \mathbb{C}[\mathbb{M}^m]^{\deg}$ be as in Section 3.3.1.

Proposition 3.25. *If $m \geq 4$, the scheme $\mathrm{Rep}_{g,n,k}$ is normal.*

Proof. Since $\mathrm{Rep}_{g,n,k}$ is a complete intersection of dimension $6g+2n-3$ in the regular scheme $\mathbb{M}^m$, it is Cohen-Macaulay. As a consequence of Serre's criterion for normality [66, Theorem 23.8, p.183], a Cohen-Macaulay ring is normal if and only if it is regular in codimension 1. It thus suffices to demonstrate that $\mathrm{Rep}_{g,n,k}$ is regular in codimension 1. Consider the restriction of $F = (F_1, \cdots, F_n)$ above to $\mathrm{SL}_2^m \subset \mathbb{M}^m \times \mathbb{A}^1$. Since SL_2^m is regular, the singular locus of $\mathrm{Rep}_{g,n,k} = F^{-1}(0)$ is the intersection of

$\mathrm{Rep}_{g,n,k}$ with the critical locus of F. It suffices to show that the locus in $\mathrm{Rep}_{g,n,k}$ defined by each of the conditions (1) and (2) of Proposition 3.20 has dimension at most $(6g+2n-3)-2$. We shall represent points of $\mathrm{Rep}_{g,n,k}$ by sequences $(a_1, \cdots, a_{2g+n-1})$ of matrices, and adopt the same notations b_I and b^I for the various products of matrices as used in Proposition 3.20.

First, consider the locus $W \subset \mathrm{Rep}_{g,n,k}$ defined by condition (1) of Proposition 3.20, which plays a role only when $n \geq 2$. For each $i = 1, \cdots, n-1$ and $s \in \{\pm 1\}$, the locus $\{b_{g+i} = s\mathbf{1}\} \subset \mathrm{Rep}_{g,n,k}$ is isomorphic to $\mathrm{Rep}_{g,n-1,k'}$ where k' is the $(n-1)$-tuple obtained from k by omitting k_i and replacing k_n by sk_n. By our assumption on (g,n), we have $2g + (n-1) - 1 \geq 2$, and we find by Lemma 3.22

$$\dim\{b_{g+i} = s\mathbf{1}\} = \dim \mathrm{Rep}_{g,n-1,k'} = 6g + 2(n-1) - 3 = (6g + 2n - 3) - 2.$$

As $W = \bigcup_{i=1}^{n-1} \{b_{g+i} = \pm\mathbf{1}\}$, it has codimension 2 in $\mathrm{Rep}_{g,n,k}$ as desired.

Next, let $Z \subset \mathrm{Rep}_{g,n,k}$ be the locus defined by condition (2) of Proposition 3.20. We stratify Z further into three subloci Z_1, Z_2, and Z_3, and estimate their dimensions as follows.

(1) Let $Z_1 \subset Z$ be the sublocus consisting of $(a_1, \cdots, a_m)$ such that $b_i = \pm\mathbf{1}$ for some $i \in \{1, \cdots, g\}$ or $b_{g+j}b_{g+j+1} = \pm\mathbf{1}$ for some $j \in \{1, \cdots, n-2\}$. We claim that $\dim Z_1 \leq (6g + 2n - 3) - 2$.

Given $i \in \{1, \cdots, g\}$ and $s \in \{\pm 1\}$, consider the locus where $b_i = s\mathbf{1}$. By Lemma 3.41, (a_{2i-1}, a_{2i}) must vary over a locus of dimension at most 4. For fixed (a_{2i-1}, a_{2i}) the remaining matrices in the sequence $(a_1, \cdots, a_m)$ vary over a locus isomorphic to $\mathrm{Rep}_{g-1,n,k'}$ for $k = (k_1, \cdots, k_{n-1}, sk_n)$, which has dimension $\dim \mathrm{Rep}_{g-1,n,k'} = 6(g-1) + 2n - 3$ by Lemma 3.22 since $2(g-1) + n - 1 \geq 2$ by our assumption. Thus, the locus in Z_1 where where $b_i = s\mathbf{1}$ for $i \in \{1, \cdots, g\}$

has dimension bounded by

$$4 + (6(g-1) + 2n - 3) = (6g + 2n - 3) - 2.$$

Given $j \in \{1, \cdots, n-2\}$ and $s \in \{\pm 1\}$, consider the locus where $b_{g+j}b_{g+j+1} = s\mathbf{1}$. The pair (b_{g+j}, b_{g+j+1}) then varies over a locus of dimension at most 2. Tor fixed (b_{g+j}, b_{g+j+1}) the remaining matrices in the sequence $(a_1, \cdots, a_m)$ vary over a locus isomorphic to $\mathrm{Rep}_{g,n-2,k'}$ where k' is obtained from k by removing k_j and k_{j+1}, and replacing k_n by sk_n. We have $\dim \mathrm{Rep}_{g,n-2,k'} = 6g+2(n-2)-3$ by Lemma 3.22 since $2g + (n-2) - 1 \geq 2$ by our assumption. Thus, the given locus has dimension bounded by

$$2 + (6g + 2(n-2) - 3) = (6g + 2n - 3) - 2.$$

This shows that $\dim Z_1 \leq (6g + 2n - 3) - 2$.

(2) Let $Z_2 \subset Z \setminus Z_1$ be the sublocus consisting of $(a_1, \cdots, a_m)$ with $\mathrm{tr}(b_i) = \pm 2$ for some $i \in \{1, \cdots, g\}$ or $\mathrm{tr}(b_{g+j}b_{g+j+1}) = \pm 2$ for some $j \in \{1, \cdots, n-2\}$. We claim that $\dim Z_2 \leq (6g + 2n - 3) - 2$.

Given $i \in \{1, \cdots, g\}$, consider the locus $\mathrm{tr}(b_i) = \pm 2$. By Lemma 3.21, the pair (a_{2i-1}, a_{2i}) varies over a locus of dimension at most 5. For fixed (a_{2i-1}, a_{2i}), by Lemma 3.39 the product b^i must vary over a locus of dimension at most 1. We have two possibilities.

(a) We have $b^i \in \{\pm\mathbf{1}\}$. We must have $2(g-1) + (n-1) - 1 \geq 2$ since Z_2 lies in the complement of Z_1. By Lemma 3.22, for fixed (a_{2i-1}, a_{2i}) the remaining matrices in $(a_1, \cdots, a_m)$ vary over a locus of dimension at most $6(g-1) + 2(n-1) - 3$.

(b) We have $b^i \neq \pm\mathbf{1}$. By Lemma 3.21, for fixed (a_{2i-1}, a_{2i}) and b^i the remaining matrices in $(a_1, \cdots, a_m)$ vary over a locus of dimension at most $6(g-1) - 2(n-1) - 3$.

Thus, we see that the locus of $(a_1, \cdots, a_m)$ in Z satisfying $\mathrm{tr}(b_i) = \pm 2$ for some $i \in \{1, \cdots, g\}$ has dimension bounded by

$$5 + 1 + (6(g-1) + 2(n-1) - 3) = (6g + 2n - 3) - 2.$$

Given $j \in \{1, \cdots, n-2\}$, consider the locus $\mathrm{tr}(b_{g+j}b_{g+j+1}) = \pm 2$. By Lemma 3.21, we see that the pair (b_{g+j}, b_{g+j+1}) varies over a locus of dimension 3. For fixed (b_{g+j}, b_{g+j+1}), by Lemma 3.39 the product $b^{\{g+j,g+j+1\}}$ must vary over a locus of dimension at most 1. We have two possibilities.

(a) We have $b^{\{g+j,g+j+1\}} \in \{\pm\mathbf{1}\}$. We must have $2g + (n-3) - 1 \geq 2$ since Z_2 lies in the complement of Z_1. By Lemma 3.22, for fixed (b_{g+j}, b_{g+j+1}) the remaining matrices in $(a_1, \cdots, a_m)$ vary over a locus of dimension at most $6g + 2(n-3) - 3$.

(b) We have $b^{\{g+j,g+j+1\}} \neq \pm\mathbf{1}$. By Lemma 3.22, for fixed (b_{g+j}, b_{g+j+1}) and $b^{\{g+j,g+j+1\}}$ the remaining matrices in $(a_1, \cdots, a_m)$ vary over a locus of dimension at most $6g - 2(n-3) - 3$.

Thus, we see that the locus of $(a_1, \cdots, a_m)$ in Z with $\mathrm{tr}(b_{g+j}b_{g+j+1}) = \pm 2$ for some $j \in \{1, \cdots, n-2\}$ has dimension bounded by

$$3 + 1 + (6g + 2(n-3) - 3) = (6g + 2n - 3) - 2.$$

This shows that $\dim Z_2 \leq (6g + 2n - 3) - 2$.

(3) We claim that $Z_3 = Z \setminus (Z_1 \cup Z_2)$ must have $\dim Z_3 \leq (6g+2n-3)-2$. Suppose

first that $g \geq 1$. The pair (a_1, a_2) varies over a locus of dimension 6. For fixed (a_1, a_2), since $\mathrm{tr}(b_1) \neq \pm 2$ we see by Lemma 3.39 that there are only finitely many possible values of b^i. For fixed (a_1, a_2) and value of b^i, the remaining matrices in $(a_1, \cdots, a_m)$ vary over a locus of dimension $6(g-1) + 2(n-1) - 3$ by Lemma 3.21 (noting that $2(g-1)+(n-1)-1 \geq 2$ if $b^i \in \{\pm \mathbf{1}\}$ by assumption that $Z_3 \cap Z_1 = \emptyset$). Hence, for $g \geq 1$, Z_3 has dimension bounded by

$$6 + (6(g-1) + 2(n-1) - 3) = (6g + 2n - 3) - 2.$$

Arguing similarly, if $g = 0$ so that $n \geq 5$, Z_3 has dimension bounded by

$$4 + (6g + 2(n-3) - 3) = (6g + 2n - 3) - 2.$$

Therefore, $\dim Z_3 \leq (6g + 2n - 3) - 2$.

From the above computations, we conclude that $\dim Z \leq (6g + 2n - 3) - 2$. Hence, $\mathrm{Rep}_{g,n,k}$ is normal, as desired. □

Proposition 3.26. $\mathrm{Rep}_{1,2,k}$ *is normal.*

Proof. Since $\mathrm{Rep}_{1,2,k}$ is a complete intersection of dimension 7 in the regular scheme $\mathbb{M}^m$, it is in particular Cohen-Macaulay. By Serre's criterion, it suffices to show that $\mathrm{Rep}_{1,2,k}$ is regular in dimension 1. It suffices to show that locus defined by each of conditions (1) and (2) of Proposition 3.20 is of dimension at most 5. We shall represent a point of $\mathrm{Rep}_{1,2,k}$ by a triple of matrices $(a_1, a_2, a_3) \in \mathrm{SL}_2^3$ satisfying $\mathrm{tr}(a_3) = k_1$ and $\mathrm{tr}(\langle a_1, a_2 \rangle a_3) = k_2$, and adopt the notations of Section 3.3.2.

Consider first the locus of $\mathrm{Rep}_{1,2,k}$ where $a_i = \pm \mathbf{1}$ for some $i \in \{1, 2, 3\}$. By Lemmas 3.21 and 3.41, we see that this locus has dimension at most $3 + 2 - 1 = 4$ when $i \in \{1, 2\}$, and at most $3+3-1 = 5$ when $i = 3$. The union $Y = \bigcup_{i=1}^{3} \{a_i = \pm \mathbf{1}\}$ includes the locus of $\mathrm{Rep}_{1,2,k}$ defined by condition (1) of Proposition 3.20, and is at

most 5-dimensional.

It remains to estimate the dimension of the locus $Z \subset \mathrm{Rep}_{1,2,k} \setminus Y$ defined by condition (2) of Proposition 3.20. In what follows, we stratify Z into subloci and estimate their respective dimensions.

(1) Let $Z_1 \subset Z$ be the locus where $\mathrm{tr}(a_1) \neq \pm 2$. From the condition

$$[a_1^{-1}, a_2^{-1} a_3 a_1 a_2] = 0,$$

and Lemma 3.43, we see that for given a_1 there are only finitely many possible values of $a_2^{-1} a_3 a_1 a_2$. The sublocus where $a_3 a_1 = \pm \mathbf{1}$ has dimension at most $2 + 3 = 5$, so we may restrict our attention to the sublocus of Z_1 where $a_3 a_1 \neq \pm \mathbf{1}$. For fixed a_1, there are only finitely many possible values of $\mathrm{tr}(a_3 a_1)$, and hence a_3 varies over a locus of dimension 1, and for fixed a_1 and a_3 since there are only finitely many possible values of $a_2^{-1} a_3 a_1 a_2$ showing that a_2 varies over a locus of dimension 1 (the centralizer of $a_1 a_3 \neq \pm \mathbf{1}$ being 1-dimensional). Thus, the sublocus of Z_1 where $a_3 a_1 \neq \pm \mathbf{1}$ has dimension at most $3 + 1 + 1 = 5$. Therefore, we conclude that $\dim Z_1 \leq 5$.

(2) Let $Z_2 \subset Z \setminus Z_1$ be the locus where $\mathrm{tr}(a_2) \neq \pm 2$. We then consider

$$[a_2, a_1^{-1} a_2^{-1} a_3 a_1] = 0,$$

and by the same argument as in part (1) we conclude that $\dim Z_2 \leq 5$.

(3) Let $Z_3 \subset Z \setminus (Z_1 \cup Z_2)$ be the locus where $\mathrm{tr}(a_1 a_2) \neq \pm 2$. By the condition

$$[a_1 a_2, a_1^{-1} a_2^{-1} a_3] = 0$$

and Lemma 3.43, for fixed a_1 and a_2 there are at most finitely many possible

values of $a_1^{-1}a_2^{-1}a_3$ and hence of a_3. Thus, $\dim Z_4 \leq 2+2=4$.

(4) Let $Z_4 = Z \setminus (Z_1 \cup Z_2 \cup Z_3)$. We must have $\mathrm{tr}(a_1), \mathrm{tr}(a_2), \mathrm{tr}(a_1a_2) \in \{\pm 2\}$, and hence (a_1, a_2) varies over a locus of dimension at most 3. Hence, $\dim Z_4 \leq 3+2=5$.

This completes the proof that $\dim Z \leq 5$, and hence $\mathrm{Rep}_{1,2,k}$ is normal. □

Proposition 3.27. *If $m \geq 3$, the intersection $\mathrm{Crit}(E,F) \cap H_{g,n}$ has codimension at least 1 in $H_{g,n}$, and in particular $H_{g,n}$ is reduced.*

Proof. First, the locus in $H_{g,n}$ determined by the condition $a_i = 0$ has dimension

$$\begin{aligned} &3+6(g-1)+2(n-1) = (6g+2n-3)-2 && \text{if } i \in \{1, \cdots, 2g\}, \\ &6g+2(n-2) = (6g+2n-3)-1 && \text{if } i \in \{2g+1, \cdots, 2g+n-1\}. \end{aligned}$$

Therefore, the union $Y = \bigcup_{i=1}^{m} \{a_i = 0\}$ has dimension at most $(6g+2n-3)-1$. Since $\mathrm{Crit}(E) \cap H_{g,n}$ and the locus defined by condition (1) of Lemma 3.20 both lie in Y, their dimensions also cannot be more than $(6g+2n-3)-1$.

Next, let $Z \subset H_{g,n} \setminus Y$ be the locus defined by condition (2) of Lemma 3.20. We must show that $\dim Z \leq (6g+2n-3)-1$. Let us first treat the cases $(g,n) = (0,4)$ and $(1,2)$ separately.

(1) Let $(g,n) = (0,4)$. The locus Z consists of $(a_1,a_2,a_3) \in \mathbb{M}_{0,0}^3(\mathbb{C})$ with a_i nonzero and $[a_1, a_2a_3] = [a_2, a_3a_1] = [a_3, a_1a_2] = 0$. Suppose that one of a_1a_2, a_2a_3, and a_3a_1 is nonzero; say $a_1a_2 \neq 0$. Then we must have $\mathrm{tr}(a_1a_2) \neq 0$ by Lemma 3.45.(1). But the condition $[a_3, a_1a_2] = 0$ implies that a_1a_2 is a scalar multiple of a_3 and hence $\mathrm{tr}(a_1a_2) = 0$, a contradiction. Arguing similarly for the other cases, we conclude that

$$a_1a_2 = a_2a_3 = a_3a_1 = 0.$$

The condition $\operatorname{tr} a_2 = \operatorname{tr} a_3 = 0$ then implies that a_2 and a_3 are both scalar multiples of a_1. Thus, Z is 4-dimensional.

(2) Let $(g, n) = (1, 2)$. The locus Z consists of $(a_1, a_2, a_3) \in \mathbb{M}_0^2 \times \mathbb{M}_{0,0}(\mathbb{C})$ with a_i nonzero satisfying, among other things, $[\langle a_1, a_2\rangle, a_3] = 0$. By Lemma 3.39.(1), this implies that $\langle a_1, a_2\rangle$ is a scalar multiple of a_3 and in particular $\operatorname{tr}\langle a_1, a_2\rangle = 0$. The condition $\operatorname{tr}(\langle a_1, a_2\rangle a_3) = 0$ then shows that in fact $\langle a_1, a_2\rangle a_3 = 0$ by Lemma 3.45(1). The sublocus of A where $\langle a_1, a_2\rangle \neq 0$ has dimension

$$2 + 1 + 3 = 6.$$

Indeed, a_3 varies over a locus of dimension 2, for fixed a_3 the value of $\langle a_1, a_2\rangle$ varies over a locus of dimension 1 by Lemma 3.39, and for fixed value of $\langle a_1, a_2\rangle \neq 0$ the pair (a_1, a_2) varies over a locus of dimension 3 by Lemma 3.23.

It remains to consider the locus $\langle a_1, a_2\rangle = 0$. This condition implies that at least one of $a_1a_2, a_2a_1^*, a_1^*a_2^*$ is zero, by Lemma 3.43.(2). The locus (a_1, a_2) where at least two of them are zero has dimension at most 4 by Lemma 3.44, and hence (a_1, a_2, a_3) would vary over a locus of dimension at most $4 + 2 = 6$, as desired. Thus, we may assume that exactly one of $a_1a_2, a_2a_1^*, a_1^*a_2^*$ is zero. We thus have the following possibilities.

(a) Consider $a_1a_2 = 0$ and $a_2a_1^*, a_1^*a_2^* \neq 0$. We must have $a_2a_1^*a_2^* \neq 0$ by Lemma 3.43.(1), and $a_1 \neq 0$. The condition

$$[a_2a_1^*a_2^*, a_3a_1] = 0$$

shows that $\operatorname{tr}(a_3a_1) = 0$ since $\operatorname{tr}(a_2a_1^*a_2^*) = 0$. For fixed a_3, the locus of $a_1 \in \mathbb{M}_0(\mathbb{C})$ with $\operatorname{tr}(a_3a_1) = 0$ has dimension 2 (as easily verified when

$a_3 = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$, and deduced from this in the other cases by conjugation). For fixed nonzero a_1 the locus of of $a_2 \in \mathbb{M}_0(\mathbb{C})$ with $a_1 a_2 = 0$ has dimension 2 by Lemma 3.42(1). Thus, the locus of $(a_1, a_2, a_3) \in Z(\mathbb{C})$ with $a_1 a_2 = 0$ has dimension at most $2 + 2 + 2 = 6$, as desired.

(b) Consider $a_2 a_1^* = 0$ and $a_1 a_2, a_1^* a_2^* \neq 0$, so that (a_1, a_2) varies over a locus of dimension at most 5. We have $a_1(\mathrm{tr}(a_2)\mathbf{1} - a_2) = (a_2 a_1^*)^* = 0$ which implies that $\mathrm{tr}\, a_2 \neq 0$ and similarly $\mathrm{tr}\, a_1 \neq 0$. Now, we have

$$a_1^* a_2^* a_3 a_1 a_2 = -[a_2, a_1^* a_2^* a_3 a_1] = 0.$$

Since $a_1^* a_2^*, a_1 a_2 \neq 0$ and $\mathrm{tr}(a_3) = 0$, the above condition implies that, for fixed (a_1, a_2), a_3 varies over a locus of dimension at most 1 by Lemma 3.42.(3). Hence, (a_1, a_2, a_3) varies over a locus of dimension $5 + 1 = 6$, as desired.

(c) Consider $a_1^* a_2^* = 0$. We have $a_2^* \neq 0$ and $a_1 a_2 a_1^* \neq 0$, and from the condition $[a_2^* a_3, a_1 a_2 a_1^*] = 0$ we argue as in (a) to see that this locus has dimension at most 6, as desired.

Thus, we have shown that Z has dimension 6.

In the remainder of the proof, we thus assume that $2g + n - 1 \geq 4$. Let us use the notations of Section 3.3.2. For a cyclic interval I in $\mathfrak{C}_{g+n-1}$, let $Z_I \subset Z$ be the sublocus defined by the condition $b_I \neq 0$. Note that if $I \subseteq J$ are two cyclic intervals (containment meaning that I is a subsequence of J) then $Z_J \subseteq Z_I$. Let

$$W_I = Z_I \setminus \bigcup_{I \subsetneq J} Z_J$$

where the union on the right hand side runs over all cyclic intervals J in $\mathfrak{C}_{g+n-1}$

containing I and distinct from I. Note that we have a stratification

$$Z = \coprod_I W_I$$

where the union runs over the finitely many cyclic intervals I of $\mathfrak{C}_{g+n-1}$. Hence, it suffices to show that each W_I has dimension at most $(6g+2n-3)-1$.

First, consider the case where $|I| = g+n-1$, and write $I = (i_0, \cdots, i_0+g+n-2)$. The condition $[b_{i_0}, b^{i_0}] = 0$ on Z with $b_{i_0}, b^{i_0} \neq 0$ then implies that, by Lemmas 3.39 and 3.23, the dimension of W_I is at most

$$\begin{aligned} &6+1+6(g-1)+2(n-1)-3 = (6g+2n-3)-1 && \text{if } g \geq 1, \text{ or} \\ &2+1+2(n-2)-3 = (6g+2n-3)-1 && \text{if } g = 0 \end{aligned}$$

as desired.

Now, suppose that $|I| < g+n-1$. Let us further stratify $W_I = W_I' \sqcup W_I''$ where W_I' (resp. W_I'') is the sublocus defined by $b^I \neq 0$ (resp. $b^I = 0$). The subloci W_I' can be handled as above and shown to have dimension at most $(6g+2n-3)-1$. Thus, it remains to consider the subloci W_I'' for cyclic intervals I with $|I| < g+n-1$. There are two possibilities to consider.

(1) We have $|I^c| = 1$. Since $Z \subset H_{g,n} \setminus Y$, we must then have $I^c = \{i\}$ for some $i \in \{1, \cdots, g\}$. Since $b_i = \langle a_{2i-1}, a_{2i} \rangle = 0$ on W_I'', by Lemma 3.43 at least one of $a_{2i-1}a_{2i}$, $a_{2i}a_{2i-1}^*$, and $a_{2i-1}^* a_{2i}^*$ must be zero. By Lemma 3.44, the sublocus of W_I'' where at least two of the products is zero has dimension

$$4 + 6(g-1) + 2(n-1) = (6g+2n-3)-1.$$

Thus, we are left to consider the sublocus W_I''' of W_I'' where exactly one of the products $a_{2i-1}a_{2i}$, $a_{2i}a_{2i-1}^*$, and $a_{2i-1}^* a_{2i}^*$ is be zero. We thus have the following

possibilities.

(a) Consider $a_{2i-1}a_{2i} = 0$ and $a_{2i}a_{2i-1}^*, a_{2i-1}^*a_{2i}^* \neq 0$. We must then have $a_{2i}a_{2i-1}^*a_{2i}^* \neq 0$ by Lemma 3.43, and $a_{2i-1} \neq 0$. The condition

$$[a_{2i}a_{2i-1}^*a_{2i}^*, b^i a_{2i-1}] = 0$$

shows that $\mathrm{tr}(b^i a_{2i-1}) = 0$ since $\mathrm{tr}(a_{2i}a_{2i-1}^*a_{2i}^*) = 0$. For fixed $b^i = b_I$ (which is nonzero on W_I), we see that the locus of $a_{2i-1} \in \mathbb{M}_0(\mathbb{C})$ with $\mathrm{tr}(b^i a_{2i-1}) = 0$ has dimension 2, and for fixed nonzero a_{2i-1} the locus of $a_{2i} \in \mathbb{M}_0(\mathbb{C})$ with $a_{2i-1}a_{2i} = 0$ has dimension at most 2. Thus, the locus of $(a_1, \cdots, a_m) \in W_I''(\mathbb{C})$ with $a_{2i-1}a_{2i} = 0$ has dimension at most

$$2 + 2 + 6(g-1) + 2(n-1) = (6g + 2n - 3) - 1$$

as desired.

(b) Consider $a_{2i}a_{2i-1}^* = 0$ and $a_{2i-1}a_{2i}, a_{2i-1}^*a_{2i}^* \neq 0$, so that (a_{2i-1}, a_{2i}) varies over a locus of dimension at most 5. Now, we have

$$a_{2i-1}^*a_{2i}^*b^i a_{2i-1}a_{2i} = -[a_{2i-1}^*, a_{2i}^*b^i a_{2i-1}a_{2i}] = 0.$$

Since $a_{2i-1}^*a_{2i}^*, a_{2i-1}a_{2i} \neq 0$, the above condition implies that, for fixed (a_1, a_2), the value of b^i varies over a locus of 2 by Lemma 3.42(3). Hence, by Lemma 3.23, the the sublocus of W_I'' with $a_{2i}a_{2i-1}^* = 0$ has dimension at most

$$5 + 2 + 6(g-1) + 2(n-1) - 3 = (6g + 2n - 3) - 1$$

as desired.

(c) Consider $a^*_{2i-1}a^*_{2i} = 0$ and $a_{2i-1}a_{2i}, a_{2i}a^*_{2i-1} \neq 0$. We must then have $a_{2i-1}a_{2i}a^*_{2i-1} \neq 0$ by Lemma 3.43, and $a^*_{2i} \neq 0$. From the condition $[a^*_{2i}b^i, a_{2i-1}a_{2i}a^*_{2i-1}] = 0$ we argue as in (a) to see that this locus has dimension at most $(6g + 2n - 3) - 1$, as desired.

(2) We have $|I^c| \geq 2$. Let us write $I^c = \{i_0, \cdots, i_0 + k\}$ for some $k \geq 1$. Let $J = \{i_0 + 1, \cdots, i_0 + k\}$, so that we have, on W''_I,

$$b_{i_0}b_J = 0, \quad b_I b_{i_0} = 0, \quad b_J b_I = 0.$$

Let us further stratify $W''_I = V_I \sqcup V'_I$ where V_I (resp. V'_I) is the sublocus defined by the condition $b_J \neq 0$ (resp. $b_J = 0$). It remains to show that V_I and V'_I each have dimension at most $(6g + 2n - 3) - 1$.

(a) Let us define

$$\begin{aligned} g_1 &= |I \cap \{1, \cdots, g\}|, \\ n_1 &= |I \cap \{g + 1, \cdots, g + n - 1\}|, \\ g_2 &= |J \cap \{1, \cdots, g\}|, \quad \text{and} \\ n_2 &= |J \cap \{g + 1, \cdots, g + n - 1\}|. \end{aligned}$$

Now, if $i_0 \in \{1, \cdots, g\}$, then by (a minor variant of) Lemmas 3.23 and 3.42 and condition $b_{i_0}b_J = b_I b_{i_0} = 0$, V_I has dimension bounded by

$$6 + (6g_1 + 2n_1 - 1) + (6g_2 + 2n_2 - 1) = 6g + 2(n - 1) - 2 = (6g + 2n - 3) - 1$$

where $g_1 + g_2 = g - 1$ and $n_1 + n_2 = n - 1$. On the other hand, if $i_0 \in \{g + 1, \cdots, g + n - 1\}$, then by a similar argument V_I has dimension

bounded by

$$2 + (6g_1 + 2n_1 - 1) + (6g_2 + 2n_2 - 1) = 6g + 2(n-2) = (6g + 2n - 3) - 1$$

where $g_1 + g_2 = g$ and $n_1 + n_2 = n - 2$. This proves that V_I has dimension at most $(6g + 2n - 3) - 1$, as desired.

(b) In addition to g_i and n_i as above, let us set

$$g_1' = |(I \cup \{i_0\}) \cap \{1, \cdots, g\}|, \quad \text{and}$$
$$n_1' = |(I \cup \{i_0\}) \cap \{g+1, \cdots, g+n-1\}|.$$

Since $g_1' + g_2 = g$ and $n_1' + n_2 = n - 1$, by Lemma 3.24 the dimension of V_I' is bounded by

$$(6g_1' + 2n_1' - 1) + (6g_2 + 2n_2 - 1) = 6g + 2(n-1) - 2 = (6g + 2n - 3) - 1$$

proving the desired result.

This completes the proof that $\dim W_I'' \leq (6g + 2n - 3) - 1$. Therefore, we have $\dim Z \leq (6g+2n-3)-1$, completing the proof that the intersection $\mathrm{Crit}(E, F) \cap H_{g,n}$ has codimension at least 1 in $H_{g,n}$. Finally, since $H_{g,n}$ is Cohen-Macaulay, the second statement of the proposition follows from the first. □

3.4 Curves on surfaces

Let $\Sigma_{g,n}$ be a compact oriented surface of type (g, n) such that $n \geq 1$ and $2g+n-2 > 0$. As in Chapter 2, let us fix a standard presentation of the fundamental group

$$\pi_1(\Sigma_{g,n}) = \langle a_1, \cdots, a_{2g+n} | \langle a_1, a_2 \rangle \cdots \langle a_{2g-1}, a_{2g} \rangle a_{2g+1} \cdots a_{2g+n} \rangle$$

and let $\sigma = \{a_1, \cdots, a_{2g+n-1}\}$ be the standard set of free generators. (We will later give a construction of $\Sigma_{g,n}$ making these standard choices concrete.)

Let us denote by $\chi_{g,n}$ the set of conjugacy classes of $\pi_1(\Sigma_{g,n})$, and let $\bar{\chi}_{g,n}$ be the quotient of $\chi_{g,n}$ obtained by setting the conjugacy class of $a \in \pi_1(\Sigma_{g,n})$ to be equivalent to that of a^{-1}. We shall refer to elements of $\bar{\chi}_{g,n}$ as reduced homotopy classes. We may identify $\chi_{g,n}$ with the set of free homotopy classes of loops $S^1 \to \Sigma_{g,n}$, and $\bar{\chi}_{g,n}$ with the set of free homotopy classes of loops considered up to (possibly orientation-reversing) reparametrizations of S^1. Let $\operatorname{Fin} \bar{\chi}_{g,n}$ denote the collection of all finite multisets of elements in $\bar{\chi}_{g,n}$.

Given a nondegenerate curve $a \subset \Sigma_{g,n}$, there is a well-defined element $a \in \bar{\chi}_{g,n}$ obtained by taking the reduced homotopy class of any choice of parametrization $S^1 \to a$. This extends to a well-defined assignment

$$\{\text{isotopy classes of nondegenerate multicurves in } \Sigma_{g,n}\} \to \operatorname{Fin} \bar{\chi}_{g,n} \tag{3.1}$$

sending $a = \coprod_i a_i \mapsto \{a_i\}$. Let $X_{g,n}$ be the character variety of $\Sigma_{g,n}$. In the coordinate ring $R_{g,n} = \mathbb{C}[X_{g,n}]$, we have $\operatorname{tr}_{a^{-1}} = \operatorname{tr}_a$ and $\operatorname{tr}_{aba^{-1}} = \operatorname{tr}_b$ for every $a, b \in \pi_1(\Sigma_{g,n})$. In particular, the function $\pi_1(\Sigma_{g,n}) \to R_{g,n}$ given by $a \mapsto \operatorname{tr}_a$ factors through the projection $\pi_1(\Sigma_{g,n}) \to \bar{\chi}_{g,n}$. We extend the function $\bar{\chi}_{g,n} \to R_{g,n}$ to an assignment

$$\operatorname{Fin} \bar{\chi}_{g,n} \to R_{g,n}$$

sending $a = \{a_1, \cdots, a_s\} \mapsto \operatorname{tr}_a = \operatorname{tr}_{a_1} \cdots \operatorname{tr}_{a_s} \in R_{g,n}$. In particular, for each isotopy class of a nondegenerate multicurve $a = \coprod_i a_i$ in $\Sigma_{g,n}$, there is a well-defined regular function $\operatorname{tr}_a = \prod_i \operatorname{tr}_{a_i} \in R_{g,n}$ (the empty multicurve $\emptyset$ corresponding to 1). We have the following remarkable result due to Charles and Marché [15, Theorem 1.1].

Theorem 3.28 (Charles-Marché [15])**.** *The regular functions* tr_a*, as* a *runs over the isotopy classes of nondegenerate multicurves in* $\Sigma_{g,n}$*, form a* $\mathbb{C}$*-linear basis of* $R_{g,n}$*.*

Remark. In particular, the assignment (3.1) considered above is injective.

We give an explicit presentation of $\Sigma_{g,n}$ as follows. For any compact oriented surface S, we shall denote by ∂S its boundary and $S^\circ = S \setminus \partial S$. Let

$$\mathbf{D} = \{z \in \mathbb{C} : |z| \leq 1\}$$

be the closed unit disk. Fix integers $g \geq 0$ and $n \geq 1$ with $m = 2g + n - 1 \geq 2$. Let $\mu_m = \{p_1, \cdots, p_m\}$, where $p_k = e^{2\pi ik/m}$, be the set of mth roots of unity. Let us fix $2m$ open intervals $\{\mathbf{I}'_k, \mathbf{I}''_k\}_{k=1,\cdots,m}$ on $\partial\mathbf{D}$ with disjoint closures in $\partial\mathbf{D} \setminus \{p_1, \cdots, p_m\}$, cyclically ordered counterclockwise on $\partial\mathbf{D}$ relative to the points $\{p_1, \cdots, p_m\}$ in the following way:

$$\begin{aligned}
&\mathbf{I}'_1 \to p_1 \to \mathbf{I}''_2 \to \mathbf{I}''_1 \to p_2 \to \mathbf{I}'_2 \to \cdots \\
&\to \mathbf{I}'_{2g-1} \to p_{2g-1} \to \mathbf{I}''_{2g} \to \mathbf{I}''_{2g-1} \to p_{2g} \to \mathbf{I}'_{2g} \\
&\to \mathbf{I}'_{2g+1} \to p_{2g+1} \to \mathbf{I}''_{2g+1} \to \cdots \to \mathbf{I}'_m \to p_m \to \mathbf{I}''_m.
\end{aligned}$$

Let us write $\mathbf{I}_k = \mathbf{I}'_k \sqcup \mathbf{I}''_k$ and set $\mathbf{I} = \coprod_{k=1}^m \mathbf{I}_k$. For $k = 1, \cdots, m$, we attach a compact rectangular strip $R_k = [0,1] \times [0,1]$ to $\mathbf{D}$, so that $\{0\} \times [0,1]$ is identified with $\mathbf{I}'_k$ and $\{1\} \times [0,1]$ is identified with $\mathbf{I}''_k$, in such a way that the resulting surface remains orientable. Note that the surface $\Sigma_{g,n}$ thus obtained has genus g with $n \geq 1$ boundary components. Let $0 \in \mathbf{D}$ be the fixed base point of $\Sigma_{g,n}$.

For $k = 1, \cdots, m$, let $a_k \in \pi_1(\Sigma_{g,n})$ be (the class of) a based loop which passes from 0 to a point in $\mathbf{I}'_k$ via a straight line segment in $\mathbf{D}$, then to a point of $\mathbf{I}''_k$ via a simple path in R_k, then back to 0 via a straight line segment in $\mathbf{D}$. Finally, let us fix a point $q \in \partial\mathbf{D} \setminus \bar{\mathbf{I}}$ situated so that we have a cyclic ordering counterclockwise

$$\overline{\mathbf{I}''_m} \to q \to \overline{\mathbf{I}'_1}$$

along $\partial\mathbf{D}$, and let $a_{2g+n} \in \pi_1(\Sigma_{g,n})$ be (the class of) the loop passing from 0 to q via a straight line segment, then traveling once clockwise along the boundary component of $\partial\Sigma_{g,n}$ containing q, and then returning to 0 via the same line segment. Under these choices, we have the standard presentation

$$\pi_1(\Sigma_{g,n}) = \langle a_1, \cdots, a_{2g+n} | [a_1, a_2] \cdots [a_{2g-1}, a_{2g}] a_{2g+1} \cdots a_{2g+n} \rangle$$

and we set $\sigma = \{a_1, \cdots, a_m\}$ to be the set of standard free generators. Note that each a_k for $k = 1, \cdots, m$ is (the class of) a simple loop in $\Sigma_{g,n}$. Given a collection of loops $a : \coprod_{i=1}^{s} S^1 \to \Sigma_{g,n}$, there exists an immersion $a_0 : \coprod_{i=1}^{s} S^1 \to \Sigma_{g,n}$ homotopic to a such that

- $\mathrm{Im}(a_0)$ intersects $\mathbf{I}$ only finitely many times,
- $\mathrm{Im}(a_0)$ contains no intersections outside of $\mathbf{D}^\circ$, and
- $|\mathrm{Im}(a_0) \cap \mathbf{I}'_k| = |\mathrm{Im}(a_0) \cap \mathbf{I}''_k|$ for every $k \in \{1, \cdots, m\}$.

This motivates the following definition. By a *chord* in $\mathbf{D}$ we mean a straight line segment in $\mathbf{D}$ joining two distinct points of $\partial\mathbf{D}$. Given two distinct points $v_1, v_2 \in \partial\mathbf{D}$, we shall write $\{v_1, v_2\}$ to denote the chord joining them.

Definition 3.29. Let $r \geq 0$ be an integer. We denote by $D_{g,n}(r)$ the collection of graphs $\Pi = (V_\Pi, E_\Pi)$ drawn (but not necessarily embedded) in $\mathbf{D}$, consisting of a collection $E_\Pi = \{e_1, \cdots, e_r\}$ of chords in $\mathbf{D}$ for edges, such that the set of vertices $V_\Pi = \bigcup_{i=1}^{r} \partial e_i$ consists of $2r$ distinct points lying in $\mathbf{I} \subset \partial\mathbf{D}$ and satisfying $|\mathbf{I}'_k \cap V_\Pi| = |\mathbf{I}''_k \cap V_\Pi|$ for every $k = 1, \cdots, m$. We set $D_{g,n} = \bigcup_{r \geq 0} D_{g,n}(r)$. Let $C_{g,n}(r)$ be the set of planar graphs in $D_{g,n}(r)$, and set $C_{g,n} = \bigcup_{r=0}^{\infty} C_{g,n}(r)$.

A graph $\Pi \in D_{g,n}$ is considered only up to isotopy of the vertex set V_Π within $\mathbf{I}$. By a generic choice of vertex positions, we may assume that each intersection

point among the chords lies on exactly two chords, and denote by $I(\Pi)$ the number of intersection points on Π under such configuration. Let us call $\Pi \in D_{g,n}$ *reduced* if no edge of Π joins two vertices lying on the same connected component of $\mathbf{I}$.

Given $\Pi \in D_{g,n}$, we obtain a finite collection $a(\Pi)$ of loops in $\Sigma_{g,n}$ (considered up to reparametrization of each loop) by joining vertices of Π in each $\mathbf{I}'_k$ to $\mathbf{I}''_k$ by pairwise non-intersecting segments in R_k. This gives rise to an assignment

$$D_{g,n} \to \operatorname{Fin} \bar{\chi}_{g,n}$$

which is surjective by the previous discussion. If $\Pi \in C_{g,n}$, the construction $a(\Pi)$ gives us a multicurve in $\Sigma_{g,n}$, and the assignment $D_{g,n} \to \operatorname{Fin} \bar{\chi}_{g,n}$ restricts to

$$C_{g,n} \to \{\text{isotopy classes of multicurves in } \Sigma_{g,n}\}$$

which is surjective (but not injective in general).

The standard generating set $\sigma = \{a_1, \cdots, a_m\} \subset \pi_1(\Sigma_{g,n})$ defines the word length function $\operatorname{length}_\sigma : \pi_1(\Sigma_{g,n}) \to \mathbb{Z}_{\geq 0}$, giving rise to $\operatorname{length}_\sigma : \overline{\chi}_{g,n} \to \mathbb{Z}_{\geq 0}$ defined by

$$\operatorname{length}_\sigma(a) = \min\{\operatorname{length}_\sigma(b) : b \in \pi_1(\Sigma_{g,n}) \text{ lies in class } a\}.$$

We extend this additively to a function $\operatorname{length}_\sigma : \operatorname{Fin} \bar{\chi}_{g,n} \to \mathbb{Z}_{\geq 0}$; namely, given $b = \{b_1, \cdots, b_s\} \in \operatorname{Fin} \bar{\chi}_{g,n}$ we define $\operatorname{length}_\sigma(b) = \sum_{i=1}^{s} \operatorname{length}_\sigma(b_i)$. In particular, the length of a nondegenerate multicurve is well-defined via the assignment (3.1). We remark that, by definition, the empty multicurve $\emptyset \subset \Sigma_{g,n}$ has length $\operatorname{length}_\sigma(\emptyset) = 0$, and it is the unique nondegenerate multicurve with this property.

Lemma 3.30. *For each isotopy class of a nondegenerate multicurve $a \subset \Sigma_{g,n}$, there is a unique reduced $\Pi \in C_{g,n}$ such that $a = a(\Pi)$. Furthermore, if $\operatorname{length}_\sigma(a) = r$, then $\Pi \in C_{g,n}(r)$.*

Proof. Observe that, if $\Pi, \Pi' \in D_{g,n}$ are reduced and there is a homotopy from $a(\Pi)$ to $a(\Pi')$ within $\Sigma_{g,n}$ (up to reparametrizations of the loops), then there is a homotopy $(a(t))_{t\in[0,1]}$ with $a(0) = a(\Pi)$ and $a(1) = a(\Pi')$ such that

$$|\operatorname{Im} a(t) \cap \mathbf{I}| \leq |\operatorname{Im} a(\Pi) \cap \mathbf{I}|$$

for all $t \in [0, 1]$. Similarly, given $\Pi, \Pi' \in C_{g,n}$ reduced such that there is an isotopy from $a(\Pi)$ to $a(\Pi')$ within $\Sigma_{g,n}$, there is an isotopy $(a(t))_{t\in[0,1]}$ from $a(\Pi)$ to $a(\Pi')$ satisfying the above inequality and in particular we must have $\Pi = \Pi'$. Further, given a nondegenerate multicurve a it is clear that there is a reduced $\Pi \in C_{g,n}$ such that $a(\Pi)$ is isotopic to a. This shows that the assignment

$$\{\Pi \in C_{g,n} : \Pi \text{ reduced}\} \to \{\text{isotopy classes of nondegenerate multicurves in } \Sigma_{g,n}\}$$

given by $\Pi \mapsto a(\Pi)$ is bijective, proving the first statement.

To prove the second statement, note first that if $\Pi \in D_{g,n}(r)$ for some $r \geq 0$ then $\operatorname{length}_\sigma(a(\Pi)) \leq r$. Thus, given a nondegenerate multicurve $a \subset \Sigma_{g,n}$ with $\operatorname{length}_\sigma(a) = r$, by our observation in the previous paragraph it suffices to show that there exists $\Pi \in D_{g,n}(r)$ such that $a(\Pi)$ is homotopic to a. In fact, we are reduced to the case where a is a curve. But for a curve, this statement is obvious. □

We prove the following refinement of Theorem 3.28. Recall that Fil^σ is the σ-word length filtration on $R_{g,n}$.

Lemma 3.31. *The regular functions* tr_a*, as* a *runs over the isotopy classes of nondegenerate multicurves in* $\Sigma_{g,n}$ *with* $\operatorname{length}_\sigma(a) \leq r$*, form a* $\mathbb{C}$*-linear basis of* $\mathrm{Fil}^\sigma_r R_{g,n}$*.*

Proof. By Theorem 3.28, the collection $B_r \subset \mathrm{Fil}^\sigma_r$ of functions tr_a, as a runs over the isotopy classes of nondegenerate multicurves in $\Sigma_{g,n}$ with $\operatorname{length}_\sigma(a) \leq r$, is linearly independent over $\mathbb{C}$. Therefore, it remains to show that B_r spans Fil^σ_r. Given any

$a \in \pi_1(\Sigma_{g,n})$ with $\mathrm{length}_\sigma(a) \leq r$, it is easy to see that there exists $\Pi \in D_{g,n}(s)$ with $s \leq r$ such that $a = a(\Pi)$. Thus, we are reduced to showing that $\mathrm{tr}_{a(\Pi)} \in \mathrm{Span}\, B_r$ for each $\Pi \in D_{g,n}(r)$.

So let $\Pi \in D_{g,n}(r)$ be given. We shall proceed by induction on the intersection number $I(\Pi)$. Let us choose the vertex positions for Π so that each of the $I(\Pi)$ intersection points lies on exactly two of the chords. If $I(\Pi) = 0$, then $\mathrm{tr}_{a(\Pi)} \in B_r$ and we are done. Suppose next that $I(\Pi) \geq 1$, and let $e = \{v_1, v_2\}$ and $e' = \{v_1', v_2'\}$ be two edges of Π intersecting at a point p. From this we can construct two graphs $\Pi', \Pi'' \in D_{g,n}(r)$ as follows.

- $V_{\Pi'} = V_{\Pi''} = V_\Pi$, and
- we have

$$E_{\Pi'} = (E_\Pi \setminus \{e, e'\}) \cup \{\{v_1, v_1'\}, \{v_1', v_2'\}\},$$
$$E_{\Pi''} = (E_\Pi \setminus \{e, e'\}) \cup \{\{v_1, v_2'\}, \{v_1', v_2\}\}.$$

The fact that $\Pi, \Pi'' \in D_{g,n}(r)$ implies that $\mathrm{tr}_{a(\Pi')}, \mathrm{tr}_{a(\Pi'')} \in \mathrm{Fil}_r^\sigma$. We verify easily that $I(\Pi'), I(\Pi'') < I(\Pi)$, so by the inductive hypothesis $\mathrm{tr}_{a(\Pi')}, \mathrm{tr}_{a(\Pi'')} \in \mathrm{Span}\, B_r$. But note that

$$\mathrm{tr}_{a(\Pi)} = \epsilon' \,\mathrm{tr}_{a(\Pi')} + \epsilon'' \,\mathrm{tr}_{a(\Pi'')}$$

for some $\epsilon', \epsilon'' \in \{\pm 1\}$, as easily seen by the condition that $\mathrm{tr}_a \,\mathrm{tr}_b = \mathrm{tr}_{ab} + \mathrm{tr}_{ab^{-1}}$ in $R_{g,n}$ (applied after changing our base point to the intersection point p). We thus have $\mathrm{tr}_{a(\Pi)} \in \mathrm{Span}\, B_r$, completing the induction. □

For $r \geq 0$, let $c'_{g,n}(r)$ be the number of isotopy classes of nondgenerate multicurves $a \subset \Sigma_{g,n}$ with $\mathrm{length}_\sigma(a) = r$, and let $c_{g,n}(r)$ be the number of isotopy classes of

essential multicurves $a \subset \Sigma_{g,n}$ with $\text{length}_\sigma(a) = r$. Note that we have

$$(1-t)^{n-1}(1-t^{4g+n-1})\sum_{r=0}^{\infty} c'_{g,n}(r)t^r = \sum_{r=0}^{\infty} c_{g,n}(r)t^r.$$

Indeed, the n boundary components of $\Sigma_{g,n}$ have lengths $1, \cdots, 1$, and $4g+n-1$, and any nondegenerate multicurve a' in $\Sigma_{g,n}$ can be written uniquely as a disjoint union $a' = a \coprod a''$ in $\Sigma_{g,n}$ where a is an essential multicurve and a'' is a finite disjoint union of curves each of which is isotopic to a boundary component of $\Sigma_{g,n}$. We are ready to prove Theorem 3.2, restated below.

Theorem 3.32. *The series $Z_{g,n}(t) = \sum_{r=0}^{\infty} c_{g,n}(r)t^r$ is a rational function, and*

$$Z_{g,n}(1/t) = Z_{g,n}(t).$$

Proof. Note that $R^\sigma_{g,n} \simeq R^\sigma_m$ where $m = 2g+n-1$ and R_m is the coordinate ring of the character variety X_m of the free group on m generators σ. By Lemma 3.31, its Hilbert series $H_m(t) = \sum_{r=0}^{\infty}(\dim \text{Fil}^\sigma_r)t^r$ is given by

$$H_m(t) = \frac{1}{1-t}\sum_{r=0}^{\infty} c'_{g,n}(r)t^r = \frac{Z_{g,n}(t)}{(1-t)^n(1-t^{4g+n-1})}.$$

By Theorem 3.14, we have $H_m(1/t) = (-1)^{3m-2}t^{2m+1}H_m(t)$. We then have

$$\begin{aligned} Z_{g,n}(1/t) &= (1-t^{-1})^n(1-t^{-(4g+n-1)})H_m(1/t) \\ &= (-1)^{3m-2}(1-t^{-1})^n(1-t^{-(4g+n-1)})t^{2m+1}H_m(t) \\ &= (-1)^{3m-2+(n+1)}(1-t)^n(1-t^{4g+n-1})H_m(t) = (-1)^{3m-2+(n+1)}Z_{g,n}(t) \end{aligned}$$

noting that $2m+1 = 2(2g+n-1)+1 = 4g+2n-1$. But lastly, we note that $3m-2+(n+1) = 3(2g+n-1)-2+(n+1)$ is even, and hence $Z_{g,n}(1/t) = Z_{g,n}(t)$, as desired. □

Remark. As the above proof shows, given $m \geq 2$ the functional equation in Theorem 3.14 is equivalent to Theorem 3.2 for any surface $\Sigma_{g,n}$ with $m = 2g + n - 1$. In the next section, we shall give an independent combinatorial proof of Theorem 3.2 for the surface $\Sigma_{0,m+1}$, which then implies Theorem 3.2 in the general case by this observation.

3.5 Combinatorics of planar graphs

Let $m \geq 2$ be a fixed integer. Let $\Sigma_m = \Sigma_{0,m+1}$ be a compact oriented surface of genus 0 with $m + 1$ boundary components. Under the standard presentation of the fundamental group, the standard set $\sigma = \{a_1, \cdots, a_m\}$ freely generates $\pi_1(\Sigma_m) \simeq F_m$. As in Section 3.4, let us consider the generating series

$$Z_m(t) = Z_{0,m+1}(t) = \sum_{r=0}^{\infty} c_m(r) t^r$$

where $c_m(r) = c_{0,m+1}(r)$ denotes the number of essential multicurves $a \subset \Sigma_m$ with $\text{length}_\sigma(a) = r$. In this section, we give a combinatorial proof of the following:

Theorem 3.33. *The series $Z_m(t)$ is rational, and satisfies $Z_m(1/t) = Z_m(t)$.*

We adopt the notations of Section 3.4. In particular, Σ_m is viewed as the closed unit disk $\mathbf{D}$ with $2m$ rectangular strips suitably attached to the boundary $\partial\mathbf{D}$ along the $2m$ intervals $\mathbf{I}$, situated in a particular way around the set $\mu_m = \{p_1, \cdots, p_m\}$, where $p_k = e^{2\pi ik/m}$, of mth roots of unity. Let us write $D_m(r) = D_{0,m+1}(r)$ and $D_m = D_{0,m+1}$, and similarly $C_m(r) = C_{0,m+1}(r)$ and $C_m = C_{0,m+1}$.

Given a reduced graph $\Pi \in C_m(r)$, note that one can identify the vertices in each $\mathbf{I}_k$ to a single vertex p_k, so as to obtain a planar multigraph in $\mathbf{D}$ with vertices $\mu_m = \{p_1, \cdots, p_m\} \subset \partial\mathbf{D}$ and r edges such that every vertex has even degree. Conversely, given such a planar multigraph, we can construct a unique reduced graph in $C_m(r)$

that gives rise to it. Let $B_m(r)$ denote the set of planar multigraphs Π without self-loops in $\mathbf{D}$ having vertex set μ_m and r edges such that every vertex in Π has even degree. In light of Lemma 3.30 and our discussion, we have

$$Z_m(t) = (1 - t^m) \sum_{r=0}^{\infty} |B_m(r)| t^r.$$

Here, the factor $(1 - t^m)$ is to cancel from our count the contribution of the $(m+1)$th boundary component of Σ_m, which has σ-length m.

Let $\mathbf{D}$ and $\partial\mathbf{D}$ be the closed unit disk and circle. Fix an integer $m \geq 3$, and let $\mu_m = \{p_1, \cdots, p_m : p_k = e^{2\pi ik/m}\} \subset \partial\mathbf{D}$ be the set of mth roots of unity. We shall say that two points $p_i, p_j \in \mu_m$ are *contiguous* if $i - j \in \{0, \pm 1\} \bmod m$.

For each integer $r \geq 0$, let $A_m(r)$ denote the collection of planar multigraphs in $\mathbf{D}$ with r edges on the m vertices μ_m, such that no edge joins contiguous vertices. Let $A_m^s(r)$ be the set of multigraphs in $A_m(r)$ which are simple (i.e. having no multiple edges). Note that we have $|A_m^s(r)| = 0$ for $r > m - 3$. Let $A_m = \bigcup_{r=0}^{\infty} A_m(r)$ and $A_m^s = \bigcup_{r=0}^{\infty} A_m^s(r)$. Define the generating series

$$F_m(t) = \sum_{r=0}^{\infty} |A_m(r)| t^r.$$

Lemma 3.34. *We have* $\sum_{r=1}^{m-3} (-1)^{r-1} |A_m^s(r)| = 1 + (-1)^{m-4}$.

Proof. The statement is obvious for $m \leq 4$, and hence we treat the case $m \geq 5$. We construct a simplicial complex $\mathbf{L}_m$ as follows. For each $r \geq 0$, the r-simplices of $\mathbf{L}_m$ are labeled by $A_m^s(r+1)$; we let $\Delta(\Pi)$ denote the r-simplex associated to $\Pi \in A_m^s(r+1)$. Whenever a graph $\Pi \in A_m(r+1)$ is obtained by deleting a number of edges from some $\Pi' \in A_m(r'+1)$ with $r < r'$, we glue the corresponding r-simplex $\Delta(\Pi)$ to the r'-simplex $\Delta(\Pi')$ compatibly. From this construction, the lemma is just

the statement

$$\chi(\mathbf{L}_m) = 1 + (-1)^{m-4}$$

where the left hand side is the Euler characteristic of $\mathbf{L}_m$. In fact, it is a relatively well known result (first published by Lee [58, Theorem 1], also found independently by Haiman (cf. Lee *loc.cit.*)) that $\mathbf{L}_m$ is isomorphic to the boundary complex of a convex polytope of dimension $m-3$, and therefore the following stronger result holds:

$$H_0(\mathbf{L}_m, \mathbb{Z}) = H_{m-4}(\mathbf{L}_m, \mathbb{Z}) = \mathbb{Z}, \quad \text{and} \quad H_r(\mathbf{L}_m, \mathbb{Z}) = 0 \quad \forall 1 \leq r < m-4.$$

This finishes the proof of the lemma. □

Let $r_0 \leq m-3$ be a nonnegative integer, and let $\Pi \in A_m^s(r_0)$. For $r_0 \leq r \leq m-3$, let us define $A_m^s(r, \Pi) = \{\Pi' \in A_m^s(r) : \Pi \subseteq \Pi'\}$. Here, the containment $\Pi \subseteq \Pi'$ means that the edge set of Π' contains the edge set of Π. Both graphs have the same vertex set, namely μ_m.

Corollary 3.35. *Let $r_0 \leq m-3$ be a nonnegative integer, and $\Pi \in A_m^s(r_0)$. Then*

$$\sum_{r=r_0}^{m-3} (-1)^{r-r_0} |A_m^s(r, \Pi)| = (-1)^{m-3-r_0}.$$

Proof. Note that Π divides $\mathbf{D}$ into a number of polygons with numbers of sides $m_1, \cdots, m_{r_0+1} \geq 3$ such that $\sum_{k=1}^{r_0+1}(m_k - 3) + r_0 = m-3$. Here, we are counting each arc of $\partial\mathbf{D}$ joining two contiguous vertices also as sides of a polygon. Using this and Lemma 3.34, we find

$$\begin{aligned}\sum_{r=r_0}^{m-3} (-1)^{r-r_0} |A_m^s(r, \Pi)| &= \prod_{k=1}^{r_0+1} \left\{ 1 - \sum_{r=1}^{m_k-3} (-1)^{r-1} |A_{m_k}(r)| \right\} \\ &= \prod_{k=1}^{r_0+1} (-1)^{m_k-3} = (-1)^{m-3-r_0},\end{aligned}$$

which is the desired result. □

Proposition 3.36. *The generating series* $F_m(t) = \sum_{r=0}^{\infty} |A_m(r)|t^r$ *satisfies*

$$F_m(t) = \frac{f_m(t)}{(1-t)^{m-3}}$$

where $f_m(t)$ *is a polynomial of degree* $m-3$ *such that* $t^{m-3}f_m(1/t) = f_m(t)$.

Proof. For each multigraph $\Pi \in A_m(r)$, there is a unique simple graph $\Pi^s \in A_m^s$ that has the same adjacent vertices as Π (i.e. vertices joined by an edge in Π are still joined in Π^s). Note that Π is obtained from Π^s by adding edges to adjacent vertices in Π^s. Thus, we have

$$F_m(t) = \sum_{\Pi \in A_m^s} \frac{t^{e(\Pi)}}{(1-t)^{e(\Pi)}}$$

where $e(\Pi)$ denotes the number of edges of Π. Since $|A_m^s(r)| = 0$ for $r > m-3$ and $e(\Pi) \leq m-3$ for all $\Pi \in A_m^s$, we have $F_m(t) = f_m(t)/(1-t)^{m-3}$ where $f_m(t) = \sum_{r=0}^{m-3} |A_m^s(r)|t^r(1-t)^{m-3-r}$ is a polynomial of degree at most $m-3$. As we have $f_m(0) = F_m(0) = 1$ from the interpretation of $F_m(t)$ as a generating function, to prove that $f_m(t)$ has degree exactly $m-3$ it suffices to prove the remaining assertion $f_m(t) = t^{m-3}f_m(1/t)$, or equivalently

$$f_m(t) = \sum_{r=0}^{m-3} |A_m^s(r)|(t-1)^{m-3-r}.$$

To prove this, we compute $F_m(t)$ in another way by a type of inclusion-exclusion principle. For each $\Pi \in A_m^s$, note that the generating function for the number of graphs $\Pi' \in A_N$ with $(\Pi')^s \subseteq \Pi$ is given by $1/(1-t)^{e(\Pi)}$. Thus, we may write $F_m(t)$ by first adding the contributions from $A_m^s(m-3)$, then adding contributions from $A_m^s(m-4)$ and subtracting away the corresponding "overcount" from the previous

step, and so on:

$$F_m(t) = \frac{|A_m^s(m-3)|}{(1-t)^{m-3}} + \frac{1}{(1-t)^{m-4}} \sum_{\Pi \in A_m^s(m-4)} (1 - |A_m^s(m-3,\Pi)|) + \cdots .$$

Therefore, we see upon reflection that

$$\begin{aligned} F_m(t) &= \sum_{r=0}^{m-3} \frac{1}{(1-t)^r} \sum_{\Pi \in A_m^s(r)} \left\{ 1 + \sum_{r'=r+1}^{m-3} \sum_{k=1}^{r'-r} (-1)^k \binom{r'-r}{k-1} |A_m^s(r',\Pi)| \right\} \\ &= \sum_{r=0}^{m-3} \frac{1}{(1-t)^r} \sum_{\Pi \in A_m^s(r)} \sum_{r'=r}^{m-3} (-1)^{r'-r} |A_m^s(r',\Pi)| \\ &= \sum_{r=0}^{m-3} \frac{(-1)^{m-3-r} |A_m^s(r)|}{(1-t)^r} = \frac{1}{(1-t)^{m-3}} \sum_{r=0}^{m-3} |A_m^s(r)| (t-1)^{m-3-r} \end{aligned}$$

(where the first equality on the last line follows by Corollary 3.35) from which we obtain $f_m(t) = \sum_{r=0}^{m-3} |A_m^s(r)|(t-1)^{m-3-r}$, showing that f_m has degree $m-3$ and that $t^{m-3} f(1/t) = f(t)$ as desired. □

Let $m \geq 1$ be an integer. As introduced in the beginning of Section 3.5, for each $r \geq 0$ let $B_m(r)$ denote the set of planar multigraphs Π without self-loops in the closed unit disk $\mathbf{D}$ having vertices $\mu_m = \{p_1, \cdots, p_m : p_k = e^{2\pi i k/m}\}$ and r edges such that every vertex in Π has even degree. Let $B_m^s(r)$ denote the set of those $\Pi \in B_m(r)$ that are simple graphs, i.e. between any two vertices there is at most one edge. Let $B_m = \bigcup_{r=0}^{\infty} B_m(r)$ and $B_m^s = \bigcup_{r=0}^{\infty} B_m^s(r)$. We are interested in the generating function

$$G_m(t) = \sum_{r=0}^{\infty} |B_m(r)| t^r.$$

By a *bigon* in a multigraph we shall mean a pair of edges whose endpoints coincide. Note that for every $\Pi \in B_m$ there exists a unique $\Pi^s \in B_m^s$ such that Π is obtained by adding bigons to Π^s without losing planarity. More precisely, Π^s is constructed from Π by repeatedly removing bigons from Π until none remain. Thus, $G_m(t)$ is the sum

of the contribution of each $\Pi \in B_m^s$ which is analyzed as follows. Let $\Pi \in B_m^s(r)$ be given. Let $e(\Pi)$ denote the number of edges of Π, and $e^\circ(\Pi)$ the number of edges of Π joining non-contiguous vertices. Note that the edges of Π joining non-contiguous vertices gives a decomposition of $\mathbf{D}$ into $e^\circ(\Pi)+1$ polygons (as in the proof of Corollary 3.35) with numbers of sides $m_1, \cdots, m_{e^\circ(\Pi)+1}$ such that

$$\sum_{k=1}^{e^\circ(\Pi)+1} (m_k - 3) + e^\circ(\Pi) = m - 3.$$

In this setting, the contribution from Π to the generating function $G_m(t)$ is

$$\begin{aligned}\frac{t^{e(\Pi)}}{(1-t^2)^{m+e^\circ(\Pi)}} \prod_{k=1}^{e^\circ(\Pi)+1} F_{m_k}(t^2) &= \frac{t^{e(\Pi)}}{(1-t^2)^{m+e^\circ(\Pi)}} \prod_{k=1}^{e^\circ(\Pi)+1} \frac{f_{m_k}(t^2)}{(1-t^2)^{m_k-3}} \\ &= \frac{t^{e(\Pi)} \prod_{k=1}^{e^\circ(\Pi)+1} f_{m_k}(t^2)}{(1-t^2)^{2m-3}}\end{aligned}$$

where each $f_{m_k}(t^2)$ is a polynomial of degree $2(m_k - 3)$ by Proposition 3.36. Thus, defining $d(\Pi) = 2\sum_{k=1}^{e^\circ(\Pi)+1}(m_k - 3)$ and $g_\Pi(t) = \prod_{k=1}^{e^\circ(\Pi)+1} f_{m_k}(t^2)$ for each $\Pi \in B_m^s$ with the m_k's as above, we see that $g_\Pi(t)$ is a polynomial of degree $d(\Pi)$ with the symmetry $t^{d(\Pi)} g_\Pi(1/t) = g_\Pi(t)$. We may thus write

$$G_m(t) = \frac{1}{(1-t^2)^{2m-3}} \sum_{\Pi \in B_m^s} t^{e(\Pi)} g_\Pi(t),$$

and we are ready to prove our main result of this section.

Theorem 3.37. *The series $G_m(t) = \sum_{r=0}^{\infty} |B_m(r)| t^r$ satisfies the symmetry*

$$G_m(1/t) = (-1)^{2m-3} t^m G_m(t).$$

Proof. Given $\Pi \in B_m^s$, we define its *dual* $\Pi^\vee$ as the graph obtained from Π by removing (resp. adding) one edge between contiguous vertices that were adjacent (resp. not

adjacent) in Π, and retaining any edges between non-contiguous vertices in Π. It is easy to see that in fact $\Pi^\vee \in B_m^s$, and $(\Pi^\vee)^\vee = \Pi$. Note that we have $g_\Pi(t) = g_{\Pi^\vee}(t)$ and $d(\Pi) = d(\Pi^\vee)$; these follow from the fact that the construction of $g_\Pi(t)$ only depended on the edges of Π joining non-contiguous vertices. Furthermore, note that we have

$$d(\Pi) + e(\Pi) + e(\Pi^\vee) = 3m - 6.$$

Hence, writing

$$G_m(t) = \frac{1}{2}\frac{1}{(1-t^2)^{2m-3}} \sum_{\Pi \in B_m^s} (t^{e(\Pi)} + t^{e(\Pi^\vee)}) g_\Pi(t),$$

we obtain

$$\begin{aligned} G_m(1/t) &= \frac{1}{2}\frac{1}{(1-t^{-2})^{2m-3}} \sum_{\Pi \in B_m^s} (t^{-e(\Pi)} + t^{-e(\Pi^\vee)}) g_\Pi(1/t) \\ &= \frac{1}{2}\frac{(-1)^{2m-3}t^{4m-6}}{(1-t^2)^{2m-3}} \sum_{\Pi \in B_m^s} \frac{(t^{e(\Pi^\vee)} + t^{e(\Pi)}) g_\Pi(t)}{t^{3m-6}} = (-1)^{2m-3} t^m G_m(t) \end{aligned}$$

which gives us the result. □

Returning to the beginning of Section 3.5, we have $Z_m(t) = (1-t^m)G_m(t)$. Hence, using Theorem 3.37, we find that

$$\begin{aligned} Z_m(1/t) &= (1-t^{-m})G_m(1/t) = (1-t^{-m})(-1)^{2m-3}t^m G_m(t) \\ &= (1-t^m)G_m(t) = Z_m(t), \end{aligned}$$

which concludes the proof of Theorem 6.1.

3.A Auxiliary results on matrices

3.A.1 Identities

Let $\mathbb{M}$ be the complex affine scheme parametrizing 2×2 matrices. Let x denote the standard matrix variable for $\mathbb{M}$. Viewing x and x^* as 2×2 matrices with coefficients in $\mathbb{C}[\mathbb{M}]$, we have

$$x + x^* = \operatorname{tr}(x)\mathbf{1}, \quad \text{and} \quad xx^* = x^*x = \det(x)\mathbf{1}. \tag{3.2}$$

Note that $\det(a^*) = \det(a)$ and $\operatorname{tr}(a^*) = \operatorname{tr}(a)$ for any 2×2 matrix a. Multiplying the matrix identity $b + b^* = \operatorname{tr}(b)\mathbf{1}$ by the matrix a and taking the trace, we also obtain the identity

$$\operatorname{tr}(a)\operatorname{tr}(b) = \operatorname{tr}(ab) + \operatorname{tr}(ab^*). \tag{3.3}$$

3.A.2 Invariant theory

We have an action of SL_2 on $\mathbb{M}$ by conjugation, and hence a diagonal conjugation action of SL_2 on $\mathbb{M}^2$. It is classical that the ring of invariants $\mathbb{C}[\mathbb{M}^2 /\!\!/ \mathrm{SL}_2] = \mathbb{C}[\mathbb{M}^2]^{\mathrm{SL}_2}$ is a polynomial ring on 5 generators

$$\mathbb{C}[\operatorname{tr}(x_1), \operatorname{tr}(x_2), \operatorname{tr}(x_1x_2), \det(x_1), \det(x_2)],$$

where x_1 and x_2 are the standard matrix variables on $\mathbb{M}^2$. (See for example [26, Theorem 5.3.1(ii), p.68]) In particular, any SL_2-invariant regular function in matrix variables x_1 and x_2 is a polynomial combination of the functions $\operatorname{tr}(x_1)$, $\operatorname{tr}(x_2)$, $\operatorname{tr}(x_1x_2)$, $\det(x_1)$, and $\det(x_2)$. For example, using the identities (1) and (2), we find

that

$$\begin{aligned}\operatorname{tr}(\langle a_1, a_2\rangle) = &\operatorname{tr}(a_1)^2 \det(a_2) + \operatorname{tr}(a_2)^2 \det(a_1) + \operatorname{tr}(a_1 a_2)^2 \\ &- \operatorname{tr}(a_1)\operatorname{tr}(a_2)\operatorname{tr}(a_1 a_2) - 2\det(a_1)\det(a_2).\end{aligned}$$

For each $t \in \mathbb{C}$, the action of SL_2 on $\mathbb{M}^2$ preserves the closed subscheme $\mathbb{M}^2_t \subset \mathbb{M}^2$, and since SL_2 is linearly reductive we have

$$\mathbb{C}[\mathbb{M}^2_t /\!\!/ \mathrm{SL}_2] = \mathbb{C}[\mathbb{M}^2]^{\mathrm{SL}_2}/(\det(x_1) - t, \det(x_2) - t) \simeq \mathbb{C}[\operatorname{tr}(x_1), \operatorname{tr}(x_2), \operatorname{tr}(x_1 x_2)].$$

In particular, $\mathrm{SL}_2^2 /\!\!/ \mathrm{SL}_2 \simeq \mathbb{A}^3$, which is due to Fricke (see Goldman [38, Section 2.2] for details).

Lemma 3.38. *The quotient morphism $\mathbb{M}^2 \to \mathbb{M}^2 /\!\!/ \mathrm{SL}_2 \simeq \mathbb{A}^5$ is surjective.*

Proof. By general theory, $\mathbb{C}[\mathbb{M}^2]^{\mathrm{SL}_2}$ is a pure subring of $\mathbb{C}[\mathbb{M}^2]$. In particular, for any ideal I of $\mathbb{C}[\mathbb{M}^2]^{\mathrm{SL}_2}$ we have $I\mathbb{C}[\mathbb{M}^2] \cap \mathbb{C}[\mathbb{M}^2]^{\mathrm{SL}_2} = I$. In particular, if I is a proper ideal of $\mathbb{C}[\mathbb{M}^2]^{\mathrm{SL}_2}$ then $I\mathbb{C}[\mathbb{M}^2]$ is a proper ideal of $\mathbb{C}[\mathbb{M}^2]$. This implies the lemma. □

3.A.3 Commutators

Lemma 3.39. *We have the following.*

(1) *For $b \in \mathbb{M}_0(\mathbb{C})$ nonscalar, $\{a \in \mathbb{M}_0(\mathbb{C}) : [a, b] = 0\}$ has dimension 1.*

(2) *For $b \in \mathrm{SL}_2(\mathbb{C})$ nonscalar and $k \in \mathbb{C}$, the locus of $a \in \mathrm{SL}_2(\mathbb{C})$ with $[a, b] = 0$ and $\operatorname{tr}(ab) = k$ has dimension 1 if $\operatorname{tr}(b), k \in \{\pm 2\}$, and is finite otherwise.*

Proof. Since b is nonscalar, any $a \in \mathbb{M}(\mathbb{C})$ with $[a, b] = 0$ must be of the form

$a = \lambda_1 \mathbf{1} + \lambda_2 b$ with $\lambda_i \in \mathbb{C}$. In such a case, we have

$$\det(a) = \det(\lambda_1 \mathbf{1} + \lambda_2 b) = \lambda_1^2 + \lambda_1 \lambda_2 \operatorname{tr}(b) + \lambda_2^2 \det(b), \tag{A}$$

$$\operatorname{tr}(ab) = \operatorname{tr}((\lambda_1 \mathbf{1} + \lambda_2 b)b) = \lambda_1 \operatorname{tr}(b) + \lambda_2 \operatorname{tr}(b^2). \tag{B}$$

We first prove (1). If $\det(a) = \det(b) = 0$, then we have $\lambda_1^2 + \lambda_1 \lambda_2 \operatorname{tr}(b) = 0$. It follows that the desired locus is the union $\{\lambda b : \lambda \in \mathbb{C}\} \cup \{\lambda b^* : \lambda \in \mathbb{C}\}$, and hence is 1-dimensional. It remains to prove (2). Let S be the locus of $a \in \mathrm{SL}_2(\mathbb{C})$ determined by the conditions of (2). Under the assumptions, by the Cayley-Hamilton theorem we have $\operatorname{tr}(b^2) - \operatorname{tr}(b)^2 + 2 = 0$ and hence $\operatorname{tr}(b^2) \neq 0$ or $\operatorname{tr}(b)^2 \neq 0$. Thus, equation (B) defines a line in the (λ_1, λ_2)-plane. In particular, S is at most 1-dimensional. Suppose it is not finite. The conic in the (λ_1, λ_2)-plane defined by equation (A) must then be degenerate, i.e. the discriminant

$$-\det(a) \det \begin{bmatrix} 1 & \operatorname{tr}(b)/2 \\ \operatorname{tr}(b)/2 & \det(b) \end{bmatrix} = -1(1 - \operatorname{tr}(b)^2/4)$$

is zero. Thus, we must have $4 = \operatorname{tr}(b)^2$, i.e. $\operatorname{tr}(b) = \pm 2$. Equations (A) and (B) then become

$$1 = (\lambda_1 + \lambda_2 \operatorname{tr}(b)/2)^2, \tag{A$'$}$$

$$k = \operatorname{tr}(b)(\lambda_1 + \lambda_2 \operatorname{tr}(b)/2). \tag{B$'$}$$

The degenerate conic defined by equation (A$'$) is a union of two disjoint lines. For S to be infinite, one of the two lines must coincide with the line defined by equation (B$'$). In other words, we must have $(k/\operatorname{tr}(b))^2 = 1$, or $k = \pm \operatorname{tr}(b) \in \{\pm 2\}$. If this happens, then S is one-dimensional. Thus, we have proved the lemma. □

3.A.4 Matrices of determinant one

Lemma 3.40. *Fix $(k_1, k_2, k_3) \in \mathbb{C}^3$. The morphism:*

(1) $\prod_{i=1}^{2} \mathrm{SL}_{2,k_i} \to \mathrm{SL}_2$ *given by* $(a_1, a_2) \mapsto a_1 a_2$ *is surjective over* $\mathrm{SL}_2 \setminus \{\pm \mathbf{1}\}$.

(2) $\prod_{i=1}^{3} \mathrm{SL}_{2,k_i} \to \mathrm{SL}_2$ *given by* $(a_1, a_2, a_3) \mapsto a_1 a_2 a_3$ *is surjective.*

(3) $\langle -, - \rangle : \mathrm{SL}_2^2 \to \mathrm{SL}_2$ *is surjective.*

Proof. (1) By Lemma 3.38, given any $b \in \mathrm{SL}_2(\mathbb{C})$ there exist $a_i \in \mathrm{SL}_{2,k_i}(\mathbb{C})$ such that $\mathrm{tr}(a_1 a_2) = \mathrm{tr}(b)$. If $\mathrm{tr}(b) \neq \pm 2$, then this implies $(g a_1 g^{-1})(g a_2 g^{-1}) = b$ for some $g \in \mathrm{SL}_2(\mathbb{C})$. Consider now the case $\mathrm{tr}(b) = 2s$ for some $s \in \{\pm 1\}$ but $b \neq s\mathbf{1}$. If $k_1 \neq s k_2$, then $a_1 a_2$ cannot be $s\mathbf{1}$ and hence $(g a_1 g^{-1})(g a_2 g^{-1}) = b$ for some $g \in \mathrm{SL}_2(\mathbb{C})$ as before. The case remains that $k_1 = s k_2 = k$. Let $\lambda \in \mathbb{C}^\times$ be a root of the polynomial $x^2 - kx + 1 = 0$. We then have

$$\begin{bmatrix} \lambda & 1 \\ 0 & \lambda^{-1} \end{bmatrix} \begin{bmatrix} s\lambda^{-1} & s \\ 0 & s\lambda \end{bmatrix} = \begin{bmatrix} s & 2s\lambda \\ 0 & s \end{bmatrix},$$

and letting a_1 and a_2 respectively be the two matrices on the left hand side, there exists $g \in \mathrm{SL}_2(\mathbb{C})$ such that $(g a_1 g^{-1})(g a_2 g^{-1}) = b$.

(2) By part (1), given any $b \in \mathrm{SL}_2(\mathbb{C})$ different from $\pm\mathbf{1}$ and any $k' \in \mathbb{C} \setminus \{\pm 2\}$, there exist $b' \in \mathrm{SL}_{2,k'}(\mathbb{C})$ and $a_3 \in \mathrm{SL}_{2,k_3}(\mathbb{C})$ such that $b' a_3 = b$. Since $b' \neq \pm\mathbf{1}$ due to the condition $k' \neq \pm 2$, again by part (1) there exist $a_1 \in \mathrm{SL}_{2,k_1}(\mathbb{C})$ and $a_2 \in \mathrm{SL}_{2,k_2}(\mathbb{C})$ such that $a_1 a_2 = b'$, and thus $a_1 a_2 a_3 = b$, as desired. If $b = s\mathbf{1}$ for some $s \in \{\pm 1\}$, then choosing $a_3 \in \mathrm{SL}_{2,k_3}(\mathbb{C})$ different from $\pm\mathbf{1}$, there exist $a_1 \in \mathrm{SL}_{2,k_1}(\mathbb{C})$ and $a_2 \in \mathrm{SL}_{2,k_2}(\mathbb{C})$ such that $a_1 a_2 = s a_3^{-1}$ and hence $a_1 a_2 a_3 = b$.

(3) From Lemma 3.38, given any $b \in \mathrm{SL}_2(\mathbb{C})$ there exist $a_1, a_2 \in \mathrm{SL}_2(\mathbb{C})$ such that $\mathrm{tr}\langle a_1, a_2 \rangle = \mathrm{tr}\, b$. If $b \neq \pm\mathbf{1}$, then $\langle g a_1 g^{-1}, g a_2 g^{-1} \rangle = g \langle a_1, a_2 \rangle g^{-1} = b$ for some

$g \in \mathrm{SL}_2(\mathbb{C})$, as desired. Finally, in the case where $b = \pm\mathbf{1}$, we have

$$\langle \mathbf{1}, \mathbf{1} \rangle = \mathbf{1} \quad \text{and} \quad \left\langle \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \right\rangle = -\mathbf{1}$$

which proves the desired result. □

Lemma 3.41. *The fiber of the morphism* $\langle -, - \rangle : \mathrm{SL}_2^2 \to \mathrm{SL}_2$ *above* $\mathbf{1}$ *has dimension* 4. *The fiber above* $-\mathbf{1}$ *has dimension* 3.

Proof. For $s \in \{+, -\}$, let F_s be the fiber of $\langle -, - \rangle$ over $s\mathbf{1}$. Consider the projection

$$\pi : F_\pm \subset \mathrm{SL}_2^2 \to \mathrm{SL}_2$$

onto the second factor. We first consider F_+. The projection $\pi : F_+ \to \mathrm{SL}_2$ is surjective, since for any $b \in \mathrm{SL}_2(\mathbb{C})$ we have $\langle \mathbf{1}, b \rangle = \mathbf{1}$. Given any $b \in \mathrm{SL}_2(\mathbb{C})$, we identify the fiber $\pi^{-1}(b)$ with the scheme of matrices $a \in \mathrm{SL}_2(\mathbb{C})$ such that $\langle a, b \rangle = \mathbf{1}$, or in other words $[a, b] = 0$. If $b = \pm\mathbf{1}$, then $\pi^{-1}(b) = \mathrm{SL}_2$ has dimension 3. If $b \neq \pm\mathbf{1}$, then any $a \in \pi^{-1}(b)(\mathbb{C})$ must be of the form $a = \lambda_1 \mathbf{1} + \lambda_2 b$ for some $\lambda_1, \lambda_2 \in \mathbb{C}$. The condition $\det(a) = 1$ then identifies $\pi^{-1}(b)$ with a curve in the (λ_1, λ_2)-plane. Hence, the fibers of $\pi : F_+ \to \mathrm{SL}_2$ above $\mathrm{SL}_2 \setminus \{\pm\mathbf{1}\}$ are 1-dimensional. Therefore F_+ is $3 + 1 = 4$-dimensional.

Consider next F_-. Given $b \in \mathrm{SL}_2(\mathbb{C})$, we identify the fiber $\pi^{-1}(b)$ with the scheme of matrices $a \in \mathrm{SL}_2(\mathbb{C})$ such that $\langle a, b \rangle = -\mathbf{1}$, or in other words $ab + ba = 0$. Writing $a = (a_{ij})$ and $b = (b_{ij})$, the condition $ab + ba = 0$ amounts to

$$\begin{bmatrix} 2b_{11} & b_{21} & b_{12} & \\ b_{12} & b_{11} + b_{22} & & b_{12} \\ b_{21} & & b_{11} + b_{22} & b_{21} \\ & b_{21} & b_{12} & 2b_{22} \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ a_{21} \\ a_{22} \end{bmatrix} = 0.$$

Hence, for the fiber $\pi^{-1}(b)$ to be nonempty, we need the determinant of the 4×4 matrix on the left hand side to be zero, or in other words $\mathrm{tr}(b) = b_{11} + b_{22} = 0$. Thus, the image of π is $\mathrm{SL}_{2,0}$. Given any $b \in \mathrm{SL}_{2,0}(\mathbb{C})$, the 4×4 matrix above has rank 2, and hence its kernel is 2-dimensional. Given the additional determinant condition $a_{11}a_{22} - a_{12}a_{21} = 1$, the fiber $\pi^{-1}(b)$ has dimension 1. Since $\mathrm{SL}_{2,0}$ is 2-dimensional, we see that F_- is $2 + 1 = 3$-dimensional. □

3.A.5 Matrices of determinant zero

Lemma 3.42. *Let $b, b' \in \mathbb{M}_0(\mathbb{C})$ both be nonzero.*

(1) *The locus of $a \in \mathbb{M}_0(\mathbb{C})$ with $ab = 0$ or $ba = 0$ is 2-dimensional.*

(2) *The locus of $a \in \mathbb{M}_0(\mathbb{C})$ with $ab = ba = 0$ is 1-dimensional.*

(3) *The locus of $a \in \mathbb{M}_0(\mathbb{C})$ with $bab' = 0$ is 2-dimensional. The intersection of this locus with $\mathbb{M}_{0,0}$ is at most 1-dimensional.*

Proof. We first prove (1) and (2). Let $a = (a_{ij}) \in \mathbb{M}(\mathbb{C})$. Without loss of generality, after conjugation we may assume that

$$b = \begin{bmatrix} \lambda & 0 \\ 0 & 0 \end{bmatrix} \text{ for some } \lambda \in \mathbb{C}^\times, \quad \text{or} \quad b = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

In the former case: $ab = 0$ if and only if $a_{11} = a_{21} = 0$, and $ba = 0$ if and only if $a_{11} = a_{12} = 0$. In the latter case: $ab = 0$ if and only if $a_{11} = a_{21} = 0$, and $ba = 0$ if and only if $a_{21} = a_{22} = 0$. In both cases, if $ab = 0$ or $ba = 0$ then we automatically have $\det(a) = 0$. Parts (1) and (2) follow immediately from these.

(3) Let us write $b = (b_{ij})$, and $b' = (b'_{ij})$. The linear map $a \mapsto bab'$ on $\mathbb{M}(\mathbb{C})$ is

given in terms of matrix coefficients $a = (a_{ij})$ by

$$\begin{bmatrix} a_{11} \\ a_{12} \\ a_{21} \\ a_{22} \end{bmatrix} \mapsto \begin{bmatrix} b_{11}b'_{11} & b_{11}b'_{21} & b_{12}b'_{11} & b_{12}b'_{21} \\ b_{11}b'_{12} & b_{11}b'_{22} & b_{12}b'_{12} & b_{12}b'_{22} \\ b_{21}b'_{11} & b_{21}b'_{21} & b_{22}b'_{11} & b_{22}b'_{21} \\ b_{21}b'_{12} & b_{21}b'_{22} & b_{22}b'_{12} & b_{22}b'_{22} \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ a_{21} \\ a_{22} \end{bmatrix}$$

and since $b, b' \in \mathbb{M}_0(\mathbb{C})$ are both nonzero, the 4×4 matrix above has rank 1. Thus, the kernel of the linear map $a \mapsto bab'$ on $\mathbb{M}(\mathbb{C})$ is a linear subspace of dimension 3. Since the hypersurface $\mathbb{M}_0(\mathbb{C}) \subset \mathbb{M}(\mathbb{C})$ is integral (and not linear), its intersection with the above kernel has dimension at most 2, proving the first assertion.

We now prove the last assertion. We claim that the linear subspace V consisting of $a \in \mathbb{M}(\mathbb{C})$ satisfying $\mathrm{tr}(a) = 0$ and $bab' = 0$ is 2-dimensional. Indeed, otherwise the second and third columns of the 4×4 matrix above must be identically zero, contradicting the hypothesis that b and b' are both nonzero matrices. Now, since V and $\mathbb{M}_{0,0}$ are both integral subschemes of dimension 2 in $\mathbb{M}$, they must coincide or have intersection with dimension at most 1. But $\mathbb{M}_{0,0}$ is not linear, and hence $V \cap \mathbb{M}_{0,0}$ has dimension at most 1, proving the last assertion. □

Lemma 3.43. *Given any $a, b \in \mathbb{M}_0(\mathbb{C})$, we have the following.*

(1) *$aba^* = 0$ if and only if at least one of ab, ba^* is zero.*

(2) *$\langle a, b \rangle = 0$ if and only if at least one of ab, ba^*, a^*b^* is zero.*

Proof. For both parts (1) and (2), one implication is clear: if one of ab, ba^*, a^*b^* is zero, then $\langle a, b \rangle = 0$, and if one of ab, ba^* is zero, then $aba^* = 0$. We shall now prove the converses. Since the lemma is trivial if a or b is zero, we may assume $a, b \in \mathbb{M}_0(\mathbb{C})$ are both nonzero. Without loss of generality, we shall assume after conjugation that

a is in Jordan normal form, so that

$$a = \begin{bmatrix} \lambda & 0 \\ 0 & 0 \end{bmatrix} \text{ for some } \lambda \in \mathbb{C}^\times, \quad \text{or} \quad a = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

In the former case, i.e. when $\mathrm{tr}(a) = \lambda \neq 0$, writing $b = (b_{ij})$ we have

$$ab = \begin{bmatrix} \lambda b_{11} & \lambda b_{12} \\ 0 & 0 \end{bmatrix}, \quad ba^* = \begin{bmatrix} 0 & \lambda b_{12} \\ 0 & \lambda b_{22} \end{bmatrix}, \quad a^*b^* = \begin{bmatrix} 0 & 0 \\ -\lambda b_{21} & \lambda b_{11} \end{bmatrix},$$

and

$$aba^* = \begin{bmatrix} 0 & \lambda^2 b_{12} \\ 0 & 0 \end{bmatrix}, \quad \langle a, b \rangle = \begin{bmatrix} -\lambda^2 b_{12} b_{21} & \lambda^2 b_{12} b_{11} \\ 0 & 0 \end{bmatrix}.$$

In the latter case, i.e. when $\mathrm{tr}(a) = 0$, writing $b = (b_{ij})$ we have

$$ab = \begin{bmatrix} b_{21} & b_{22} \\ 0 & 0 \end{bmatrix}, \quad ba^* = \begin{bmatrix} 0 & -b_{11} \\ 0 & -b_{21} \end{bmatrix}, \quad a^*b^* = \begin{bmatrix} b_{21} & -b_{11} \\ 0 & 0 \end{bmatrix},$$

and

$$aba^* = \begin{bmatrix} 0 & -b_{21} \\ 0 & 0 \end{bmatrix}, \quad \langle a, b \rangle = \begin{bmatrix} b_{21}^2 & -b_{21} b_{11} \\ 0 & 0 \end{bmatrix}.$$

We now proceed with our proof.

(1) Suppose that ab and ba^* are nonzero yet $aba^* = 0$. The condition $aba^* = 0$ would imply that $b_{12} = 0$ if $\mathrm{tr}(a) \neq 0$, and $b_{21} = 0$ if $\mathrm{tr}(a) = 0$. In both cases, since $ab, ba^* \neq 0$ we must have $b_{11}, b_{22} \neq 0$, contradicting the assumption that $\det(b) = 0$. Thus, we must have $aba^* = 0$.

(2) Assume toward contradiction that ab, ba^*, a^*b^* are all nonzero yet $\langle a, b \rangle = 0$. Consider the case $\mathrm{tr}(a) = \lambda \neq 0$. The condition $\langle a, b \rangle = 0$ would imply that $b_{12} = 0$ or $b_{21} = b_{11} = 0$. If $b_{12} = 0$, then since $ab, ba^* \neq 0$ we must have $b_{11}, b_{22} \neq 0$,

contradicting the assumption that $\det(b) = 0$. If $b_{21} = b_{11} = 0$, then $a^*b^* = 0$ contradicting our assumption. Thus, we must have $\langle a, b\rangle \neq 0$.

It remains to treat case $\mathrm{tr}(a) = 0$. The condition $\langle a, b\rangle = 0$ would imply that $b_{21} = 0$. But if $b_{21} = 0$, then since $ab, ba^* \neq 0$ we must have $b_{11}, b_{22} \neq 0$, contradicting the assumption that $\det(b) = 0$. Thus, we must have $\langle a, b\rangle \neq 0$, as desired. This finishes the proof. □

Lemma 3.44. *The locus of* $(a, b) \in \mathbb{M}_0^2(\mathbb{C})$ *such that at least two of* ab, ba^*, *and* a^*b^* *are zero has dimension at most* 4.

Proof. Since the locus where $a = 0$ or $b = 0$ has dimension at most 3, we may restrict our attention to the locus with $a, b \neq 0$.

(1) Consider the locus $ab = ba^* = 0$. We then have $ab^* = (ba^*)^* = 0$ and hence $\mathrm{tr}(b)a = ab + ab^* = 0$ and hence $\mathrm{tr}(b) = 0$, since $a \neq 0$ by assumption. Thus, b varies over a locus of dimension 2. For each fixed value of b, the condition $ab = 0$ implies that a varies over a locus of dimension 2 by Lemma 3.42. Hence, the locus where $ab = ba^* = 0$ has dimension at most $2 + 2 = 4$.

(2) Consider the locus $ab = a^*b^* = 0$. We then have $ba = (a^*b^*)^* = 0$ and hence $[a, b] = ab - ba = 0$. By Lemma 3.39, for fixed a (assumed nonzero) we see that b varies over a locus of dimension at most 1. Hence, the locus where $ab = a^*b^* = 0$ has dimension at most $3 + 1 = 4$.

(3) Consider the locus $ba^* = a^*b^* = 0$. We then have $ba = 0$ and $ab^* = 0$. Replacing the role of b and a, we reduce to case (1), showing that the said locus has dimension at most 4.

This concludes the proof of the lemma. □

Lemma 3.45. *The morphism:*

(1) $\mathbb{M}_{0,0}^2 \to \mathbb{M}_0$ *given by* $(a_1, a_2) \mapsto a_1 a_2$ *is surjective over* $\mathbb{M}_0 \setminus \mathbb{M}_{0,0}$. *The preimage of this morphism over* $\mathbb{M}_{0,0} \setminus \{0\}$ *is empty.*

(2) $(\mathbb{M}_0 \setminus \mathbb{M}_{0,0}) \times \mathbb{M}_{0,0} \to \mathbb{M}_0$ *given by* $(a_1, a_2) \mapsto a_1 a_2$ *is surjective.*

(3) $\langle -, - \rangle : \mathbb{M}_0^2 \to \mathbb{M}_0$ *is surjective.*

Proof. (1) By Lemma 3.38, Given any $b \in \mathbb{M}_0(\mathbb{C})$, there exist $a_1, a_2 \in \mathbb{M}_{0,0}(\mathbb{C})$ such that $\operatorname{tr}(a_1 a_2) = \operatorname{tr}(b)$. If $\operatorname{tr}(b) \neq 0$, then this implies that $(g a_1 g^{-1})(g a_2 g^{-1}) = b$ for some $g \in \mathrm{SL}_2(\mathbb{C})$, and $g a_1 g^{-1}, g a_2 g^{-1} \in \mathbb{M}_{0,0}(\mathbb{C})$. For the last statement, note first that

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x & y \\ z & w \end{bmatrix} = \begin{bmatrix} z & w \\ 0 & 0 \end{bmatrix}$$

and, provided that $\left[\begin{smallmatrix} x & y \\ z & w \end{smallmatrix}\right] \in \mathbb{M}_{0,0}(\mathbb{C})$, if the right hand side has trace zero then it must in fact be zero. Since any pair $(a_1, a_2) \in \mathbb{M}_{0,0}^2(\mathbb{C})$ with a_1 nonzero is conjugate to a pair of the form $(\left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} x & y \\ z & w \end{smallmatrix}\right])$, the last statement follows.

(2) The same argument as in the proof of part (1) goes through when $b \in \mathbb{M}_0(\mathbb{C})$ satisfies $\operatorname{tr} b \neq 0$ or $b = 0$. Suppose that $\operatorname{tr} b = 0$ and $b \neq 0$. Up to conjugation by SL_2, we may assume that $b = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$. Note then that we have

$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

This proves that the morphism $(\mathbb{M}_0 \setminus \mathbb{M}_{0,0}) \times \mathbb{M}_{0,0} \to \mathbb{M}_0$ is surjective, as desired.

(3) By Lemma 3.38, given any $b \in \mathbb{M}_0(\mathbb{C})$ there exist $a_1, a_2 \in \mathbb{M}_0(\mathbb{C})$ such that $\operatorname{tr}\langle a_1, a_2 \rangle = \operatorname{tr} b$. If $\operatorname{tr} b \neq 0$, then $\langle g a_1 g^{-1}, g a_2 g^{-1} \rangle = b$ for some $g \in \mathrm{SL}_2(\mathbb{C})$. Consider now $\operatorname{tr} b = 0$. We have $\langle 0, 0 \rangle = 0$. Furthermore, we have

$$\left\langle \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \right\rangle = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}$$

which shows that, letting a_1 and a_2 respectively be the two matrices on the left hand side, there exists $g \in \mathrm{SL}_2(\mathbb{C})$ such that $\langle g a_1 g^{-1}, g a_2 g^{-1} \rangle = b$, as desired. $\square$

Chapter 4

Systoles of local systems

Let Σ be a surface of type (g, n) satisfying $3g + n - 3 > 0$. In this chapter, we shall derive boundedness results for systoles of SL_2-local systems on surfaces. Our first main result is the following.

Theorem 4.1. *For each $B \geq 0$, there is $A \geq 0$ depending continuously on B such that, given any representation $\rho : \pi_1\Sigma \to \mathrm{SL}_2(\mathbb{C})$ whose boundary traces all have absolute value at most B, there is an essential curve $a \subset \Sigma$ such that $|\operatorname{tr} \rho(a)| \leq A$.*

Restricting to Fuchsian representations, this recovers the boundedness result of Bers [4] for systoles of hyperbolic surfaces with geodesic boundary. For closed surfaces, the above result also follows from the work of Deroin and Tholozan [23]. Our second main result is a nonarchimedean analogue of the first result.

Theorem 4.2. *Let $\mathcal{O}$ be a discrete valuation ring with fraction field F. Given any representation $\rho : \pi_1\Sigma \to \mathrm{SL}_2(F)$ whose boundary traces all take values in $\mathcal{O}$, there is an essential curve $a \subset \Sigma$ such that $\operatorname{tr} \rho(a) \in \mathcal{O}$.*

Our proof of Theorem 4.1 makes use of twisted harmonic maps into the hyperbolic three-space $\mathbb{H}^3 \simeq \mathrm{SL}_2(\mathbb{C})/\mathrm{SU}(2)$ and systolic inequalities for Riemannian surfaces. The proof of Theorem 4.2 follows similar strategy, the main differences being that we

replace $\mathbb{H}^3$ by the Bruhat-Tits tree of $\mathrm{SL}_2(F)$ and that the twisted maps we consider are not required to be harmonic (and are merely required to satisfy certain regularity properties). We remark that Bruhat-Tits trees also make an appearance [17] in the study of character curves for three-manifolds.

This chapter is organized as follows. In Section 4.1, we collect relevant background in differential geometry, including systolic inequalities for Riemannian surfaces. In Section 4.2, we introduce the notion of harmonic maps and recall their existence theory from the literature. In Section 4.3, we prove Theorems 4.1 and 4.2 using the tools and results introduced in the earlier sections.

Historical remarks

The use of systoles and harmonic maps in our Diophantine analysis is partly inspired by their appearances in the theory of linear arithmetic groups and related studies. We comment on these below, noting that our account is necessarily selective.

The origin of systolic geometry arguably lies in Minkowski's work on the geometry of numbers. The latter concerns the properties of convex bodies and lattice points in $\mathbb{R}^m$. Using the interpretation of the locally symmetric space $\mathrm{SL}_m(\mathbb{Z})\backslash \mathrm{SL}_m(\mathbb{R})/\mathrm{SO}(m)$ as the moduli of covolume one lattices in $\mathbb{R}^m$, Minkowski established the reduction theory of $\mathrm{SL}_m(\mathbb{Z})$ extending the classical work of Lagrange, Legendre, and Gauss on the modular group $\mathrm{SL}_2(\mathbb{Z})$. A basic result in the theory is the following.

Theorem. *There is a constant $c_m \geq 0$ such that every lattice $L \leq \mathbb{R}^m$ contains a nonzero vector of length at most $c_m \operatorname{Vol}(L\backslash\mathbb{R}^m)^{1/m}$.*

This follows immediately from a form of Dirichlet's pigeonhole principle and the observation that the volume of an m-disk in $\mathbb{R}^m$ is proportional to the mth power of its radius. The length of the shortest nonzero vector on a lattice $L \leq \mathbb{R}^m$ may be interpreted as the *systole*, defined here as the length of the shortest nontrivial closed

geodesic, on the flat torus $L\backslash\mathbb{R}^m$. The development of systolic geometry proper began with Loewner's extension of the above result (for $m = 2$) to arbitrary Riemannian metrics on the 2-torus. Later, Gromov [42] established general upper bounds for the systoles of m-tori and general closed *essential* manifolds (defined in *loc.cit.*) in terms of volume. Given the arithmetic origin of the theory, it is reasonable to expect that systolic inequalities should have applications in appropriate arithmetic contexts.

An especially rich part of systolic geometry is the study of systoles for hyperbolic surfaces. Since the volume of any finite-volume complete hyperbolic surface with cusps is determined by its topology, one may expect that systole is uniformly bounded as a function on the moduli of hyperbolic surfaces of each fixed type. This and its extension to compact hyperbolic surfaces with geodesic boundary were proved by Bers [4] using elementary hyperbolic geometry. For instance, on closed hyperbolic surfaces, which may be viewed as quotients of the hyperbolic plane $\mathbb{H}^2$ by lattices $L \leq \mathrm{PSL}_2(\mathbb{R})$, the boundedness of systoles follows from the same argument as the proof of Minkowski's theorem for Euclidean lattices, using the fact that the volume of a hyperbolic disk in $\mathbb{H}^2$ is exponential in its radius.

The question of uniformly bounding the systoles of hyperbolic surfaces naturally extends to that of bounding systoles (defined in Section 4.3) for general $\mathrm{SL}_2(\mathbb{C})$-local systems on surfaces. Contrary to the former situation, however, a general local system need not correspond to a metric on the surface. In view of the above discussion, a natural approach is to construct a Riemannian metric on the surface with uniformly controlled volume and boundary length, whose systole bounds the systole of the given local system from above so that Gromov's systolic inequality may be applied. We achieve this in Section 4.3, using a twisted map from the surface to $\mathbb{H}^3$ which is harmonic with respect to a suitable auxiliary metric on the surface. Here, we rely on foundational results on harmonic maps developed by Eells-Sampson [29], Hamilton [46], Donaldson [25], Corlette [18], and others, as shall be discussed in

Section 4.2. Our approach reflects a broader principle that geometric constructions, more specifically harmonic objects built using a choice of metric, can be used to derive purely topological or algebraic consequences.

In this spirit, we remark that twisted harmonic maps were also previously used to obtain strong rigidity results for lattices in Lie groups. For instance, building on earlier ideas of Siu [83] on certain rigidity of Kähler manifolds, Corlette [18] used the technique of twisted harmonic maps to derive archimedean superrigidity of lattices in certain real rank one Lie groups (of quaternionic and octonionic type), going beyond the classical work of Margulis [64] on superrigidity of lattices in Lie groups of higher rank. We refer to [19] for a survey of related topics.

The p-adic counterpart of Corlette's superrigidity result was obtained by Gromov and Schoen [43], who developed and used the theory of harmonic maps to buildings for this purpose. Their theory is relevant to the second main result of this chapter. Although our proof of the boundedness of nonarchimedean systoles for local systems does not strictly use harmonic maps, nevertheless the underlying ideas in [43] were important in the formulation of our approach.

4.1 Differential geometry

For the convenience of the reader and to set up our notation, we begin with a brief exposition of the relevant notions and results in differential geometry.

4.1.1 Curvature

Let M be a smooth manifold with Riemannian metric $\sigma = \langle -, - \rangle$. Let ∇ be the associated Levi-Civita connection, and $R = -\nabla^2$ the Riemannian curvature tensor. For $p \in M$ and two-dimensional subspace $V \subseteq T_pM$ spanned by tangent vectors x, y,

the *sectional curvature* of V is by definition

$$K(V) = K(x, y) = \frac{\langle R(x, y)x, y\rangle}{\langle x, x\rangle\langle y, y\rangle - \langle x, y\rangle^2}.$$

One can show that $K(V)$ is independent of the choice of basis $\{x, y\}$ for V. The manifold M is said to have nonpositive sectional curvature if $K(V) \leq 0$ for every two-dimensional subspace $V \subseteq T_pM$ for every $p \in M$. For any real $\epsilon > 0$, we shall denote by $M(\epsilon)$ the smooth manifold M with the scaled metric $\epsilon\sigma = \epsilon\langle -, -\rangle$.

When discussing a geometric construction in the presence of multiple manifolds, we shall use a subscript or superscript to indicate to which manifold it belongs. For instance, we shall write K^M for the sectional curvature on M. Given smooth manifolds M and N, let $\pi_M : M \times N \to M$ and $\pi_N : M \times N \to N$ be the projections.

Lemma 4.3. *Let M and N be Riemannian manifolds of dimension at least 2.*

(1) *If M and N have nonpositive curvature, then so does $M \times N$.*

(2) *For every $(p, q) \in M \times N$ and two-dimensional subspace $V \subseteq T_{(p,q)}(M \times N)$ such that $\dim_{\mathbb{R}} d\pi_N(V) = 2$,*

$$\lim_{\epsilon \to 0} K^{M(\epsilon)\times N}(V) = K^N(d\pi_N(V)).$$

Proof. A vector field on $T(M \times N) \simeq \pi_M^* TM \oplus \pi_N^* TN$ can be written as (X, Y) for some vector field X on M and Y on N. With this identification, the Levi-Civita connection on $M \times N$ is given by

$$\nabla^{M\times N}_{(X_1,Y_1)}(X_2, Y_2) = (\nabla^M_{X_1} X_2, \nabla^N_{Y_1} Y_2)$$

for vector fields X_i on M and Y_i on N. Therefore, by definition we have

$$\begin{aligned}
&R^{M\times N}((X_1,Y_1),(X_2,Y_2))(X_3,Y_3)\\
&=(-\nabla_{(X_1,Y_1)}\nabla_{(X_2,Y_2)}+\nabla_{(X_2,Y_2)}\nabla_{(X_1,Y_1)}+\nabla_{[(X_1,Y_1),(X_2,Y_2)]})(X_3,Y_3)\\
&=(R^M(X_1,X_2)X_3,R^M(Y_1,Y_2)Y_3).
\end{aligned}$$

It follows that, given a point $(p,q)\in M\times N$ and the two-dimensional subspace $V\subseteq T_{(p,q)}(M\times N)$ generated by two vectors (x_1,y_1) and (x_2,y_2), the sectional curvature $K^{M\times N}(V)$ of V is therefore given by

$$\frac{\langle R^M(x_1,x_2)x_1,x_2\rangle_M+\langle R^N(y_1,y_2)y_1,y_2\rangle_N}{(\langle x_1,x_1\rangle_M+\langle y_1,y_1\rangle_N)(\langle x_2,x_2\rangle_M+\langle y_2,y_2\rangle_N)-(\langle x_1,x_2\rangle_M+\langle y_1,y_2\rangle_N)^2}.$$

The two claims of the lemma follow immediately from this computation. □

Let $L>0$ be a real number. Let $\partial/\partial t$ be the standard vector field on the interval $[0,L]$. Let $\gamma:[0,L]\to M$ be a smooth path. Let us suppose that it is of unit speed, that is, the vector field

$$\frac{d\gamma}{dt}:=d\gamma\left(\frac{\partial}{\partial t}\right)\in C^\infty([0,1],\gamma^*TM)$$

along γ has norm 1 everywhere. Using the Levi-Civita connection on M, we can define the covariant derivative of this vector field:

$$\frac{D}{dt}\frac{d\gamma}{dt}:=\nabla_{\partial/\partial t}\frac{d\gamma}{dt}.$$

Here, ∇ is the connection on γ^*TM induced by the Levi-Civita connection on M. By definition, the *curvature* of γ is the norm of this covariant derivative:

$$k=\left\|\frac{D}{dt}\frac{d\gamma}{dt}\right\|.$$

It is a nonnegative real function on the interval $[0, L]$. We define γ to be a *geodesic* path (of unit speed) if its curvature vanishes identically, i.e.,

$$\frac{D}{dt}\frac{d\gamma}{dt} = 0.$$

We remark that, if M is a submanifold of a Riemannian manifold N with induced metric, and γ is a smooth path in M, then $k^M \leq k^N$ on γ.

For boundary curves of surfaces, the following refinement of curvature with sign is useful. Suppose S is a compact oriented Riemannian surface, and let c be a boundary curve. Let $\gamma : [0, L] \to c$ be a unit speed parametrization of c. Let ν be the inward facing unit normal vector field along c within S. The *signed curvature* of c in S is the function $k' : [0, L] \to \mathbb{R}$ given by

$$k' = \left\langle \nu, \frac{D}{dt}\frac{d\gamma}{dt} \right\rangle.$$

Note that we have $|k'| = k$. Below, we record the Gauss-Bonnet formula.

Theorem 4.4 (Gauss-Bonnet)**.** *If S is a compact oriented Riemannian surface with boundary curves $c_1, \cdots, c_n$, then*

$$\int_S K \, d\,\mathrm{vol}_S + \sum_{i=1}^{n} \int_{c_i} k' \, ds = 2\pi\chi(S)$$

where $d\,\mathrm{vol}_S$ is the volume element on S, ds is the line element on ∂S induced by the metric, and $\chi(S)$ is the Euler characteristic of S.

4.1.2 Second fundamental form

Let $f : M \to N$ be a smooth map of Riemannian manifolds with possible boundary. Let us view its differential df as a smooth section of $T^*M \otimes f^*TN$ over M. The

second fundamental form of f is defined by

$$\mathrm{II}^f = \nabla(df) \in C^\infty(M, T^*M \otimes T^*M \otimes f^*TN).$$

Here, by ∇ we mean the connection on $T^*M \otimes f^*TN$ induced from the Levi-Civita connection ∇^M on M and the pullback ∇^f of the Levi-Civita connection ∇^N on N. Explicitly, we have

$$\mathrm{II}^f(X, Y) = \nabla^f_X(df(Y)) - df(\nabla^M_X Y)$$

for vector fields X and Y on M. If $f : M \to N$ is an isometric embedding of Riemannian manifolds, then in fact $\mathrm{II}^f(X, Y)$ defines a section of the normal bundle of M in N. In such a case, we shall write $\mathrm{II}^M = \mathrm{II}^f$ if there is no risk of confusion. We record a theorem of Gauss (See [24, Chapter 6, Theorem 2.5]).

Theorem 4.5 (Gauss)**.** *If M is a submanifold of a Riemannian manifold N with the induced metric, and x, y are orthonormal vectors in T_pM for $p \in M$, then*

$$K^M(x, y) = K^N(x, y) + \langle \mathrm{II}^M(x, x), \mathrm{II}^M(y, y) \rangle - \| \mathrm{II}^M(x, y) \|^2.$$

4.1.3 Hyperbolic three-space

We recall the model of hyperbolic three-space $\mathbb{H}^3$ obtained using the quaternions. Let $\mathbb{H} = \mathbb{R} \oplus \mathbb{R}i \oplus \mathbb{R}j \oplus \mathbb{R}k$ be the algebra of Hamiltonian quaternions, with the usual embedding of $\mathbb{C} = \mathbb{R} \oplus \mathbb{R}i$ into $\mathbb{H}$. Let

$$\mathbb{H}^3 = \{(x_1, x_2, x_3) = x_1 + x_2 i + x_3 j + 0k \in \mathbb{H} : x_3 > 0\}$$

together with its standard hyperbolic metric $ds^2 = (dx_1^2 + dx_2^2 + dx_3^2)/x_3^2$. The action of $\mathrm{SL}_2(\mathbb{C})$ on $\mathbb{H}^3$ is given by

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot q = (aq + b)(cq + d)^{-1} \quad \text{for all} \quad \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}_2(\mathbb{C}) \text{ and } q \in \mathbb{H}.$$

This action is transitive, and the stabilizer of $j \in \mathbb{H}^3$ is the special unitary group $\mathrm{SU}(2)$, giving us the identification $\mathbb{H}^3 \simeq \mathrm{SL}_2(\mathbb{C})/\mathrm{SU}(2)$. For an element $g \in \mathrm{SL}_2(\mathbb{C})$, its *translation length* is defined to be

$$\mathrm{length}(g) = \inf\{d(x, g \cdot x) : x \in \mathbb{H}^3\}$$

where $d(\cdot, \cdot)$ denotes the hyperbolic distance on $\mathbb{H}^3$. Note that translation length descends to a function of the conjugacy classes in $\mathrm{SL}_2(\mathbb{C})$. If an element $g \in \mathrm{SL}_2(\mathbb{C})$ is elliptic or parabolic, i.e. $\mathrm{tr}(g) \in [-2, 2]$, then $\mathrm{length}(g) = 0$. Otherwise, up to conjugacy we may assume that

$$g = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{bmatrix}$$

for some unique $\lambda \in \mathbb{C}$ with $|\lambda| > 1$, and writing $\lambda = \exp((\ell + i\theta)/2)$ with $\ell, \theta \in \mathbb{R}$, we have $\mathrm{length}(g) = \ell$.

4.1.4 Bruhat-Tits trees

Let F be a field with a discrete valuation v, and let $\mathcal{O} = \{f \in F : v(f) \geq 0\}$ be the associated discrete valuation ring. Let $\pi \in \mathcal{O}$ be a uniformizer (so that $v(\pi) = 1$), with residue field $k = \mathcal{O}/\pi\mathcal{O}$. The main reference for our discussion below is [79].

The Bruhat-Tits building Λ associated to $\mathrm{SL}_2(F)$ is a tree, whose construction we outline as follows. Let $V = F^2$. A *lattice* $L \subset V$ is a free rank two $\mathcal{O}$-submodule of V

such that $F \cdot L = V$. Two lattices $L, L' \subset V$ are said to be equivalent if $L = x \cdot L'$ for some $x \in F^*$, and we shall denote by $[L]$ the equivalence class of a lattice L. Given two lattices L, L', their *distance* is defined to be

$$d(L, L') = |a - b|$$

where $(a, b) \in \mathbb{Z}^2$ are such that, given an $\mathcal{O}$-basis (e_1, e_2) of L, $(\pi^a e_1, \pi^b e_2)$ is a basis of L'. It can be shown that $d(L, L')$ depends only on the equivalence classes of L and L'. Moreover,

(1) $d([L], [L']) = 0$ if and only if $[L] = [L']$, and

(2) $d([L], [L']) = 1$ if and only if, up to replacing L and L' by equivalent lattices, we have $L' \subset L$ with $L/L' \simeq k$.

The *Bruhat-Tits tree* Λ of $\mathrm{SL}_2(F)$ is a tree whose vertex set is the set of equivalence classes of lattices in V, and where two vertices are joined by an edge if and only if their distance is 1. We refer to [79, Chapter 2] for a proof that Λ is indeed a tree. We shall work with the realization of Λ as a Riemannian simplicial complex, where each edge has the Euclidean metric of the unit interval $[0, 1]$. The distance function introduced above agrees with the Riemannian distance between vertices of Λ. We note that, unless the residue field k is finite, the tree Λ is not locally finite. The group $\mathrm{SL}_2(F)$ acts naturally on Λ as a group of isometries. Given an element $g \in \mathrm{SL}_2(F)$, its *translation length* is defined to be

$$\mathrm{length}(g) = \inf\{d(x, g \cdot x) : x \in \Lambda\} \in \mathbb{Z}_{\geq 0}.$$

Note that $\mathrm{length}(g)$ depends only on the conjugacy class of g, and that $\mathrm{length}(g) = 0$ if and only if $\mathrm{tr}(g) \in \mathcal{O}$.

4.1.5 Systoles of Riemannian surfaces

Given a surface S of type (g, n) satisfying $3g+n-3 > 0$ with an arbitrary Riemannian metric, we define its *systole* $\mathrm{sys}(S)$ to be the infimal length of an essential curve on S. This recovers the usual notion of systole for closed Riemannian surfaces. We begin by recording the following results in systolic geometry due to Gromov [42, Proposition 5.1.B] and Balacheff-Parlier-Sabourau [2, Theorem 6.10].

Theorem 4.6 (Gromov)**.** *For any closed surface S of genus $g \geq 1$ with an arbitrary Riemannian metric, we have* $\mathrm{sys}(S)^2 \leq 2\,\mathrm{Vol}(S)$.

Theorem 4.7 (Balacheff-Parlier-Sabourau)**.** *There is a constant C_g such that any complete Riemannian surface S of genus g with n ends, with volume normalized to $2\pi|\chi(S)|$, admits a pants decomposition whose total length does not exceed*

$$C_g n \log(n+1).$$

Using these results, we deduce the following.

Corollary 4.8. *There is a constant $C_{g,n}$ such that, for any Riemannian surface S of type (g, n) with $3g + n - 3 > 0$, we have*

$$\mathrm{sys}(S)^2 \leq C_{g,n}(\mathrm{Vol}(S) + \ell_1^2 + \cdots + \ell_n^2)$$

where $\ell_1, \cdots, \ell_n$ are the lengths of the boundary curves $c_1, \cdots, c_n$ of S, respectively.

Proof. Assume first that $g \geq 1$. For each $i = 1, \cdots, n$, let us attach a Riemannian disk D_i to the boundary curve c_i on S, satisfying the following two conditions:

(1) between any two points on the boundary of D_i, the shortest path between them lies on the boundary; and

(2) the area of D_i is bounded by a fixed positive constant times the square of the perimeter.

For instance, the disk D_i can be taken to be the union of the cylinder $c_i \times [0, 2\ell_i]$ with a spherical cap on one end. After attaching the disks $D_1, \cdots, D_n$ to S (and after a small perturbation to obtain a smooth Riemannian metric), the resulting closed Riemannian surface (M, σ_M) has genus $g \geq 1$ with volume

$$\mathrm{Vol}(M) \leq \mathrm{Vol}(S) + c \sum_i^n \ell_i^2$$

for some positive constant c. By Gromov's theorem (Theorem 4.6), there exists a shortest non-contractible curve a on M of length at most $\sqrt{2}\,\mathrm{Vol}(M)^{1/2}$. Note that a lies entirely within S by our assumption on the disks D_i, and in fact corresponds to an essential curve on S. Since the restriction of the metric M to S is the original metric (at least up to small perturbation near the boundary), we have

$$\mathrm{length}_S(a)^2 \leq C(\mathrm{Vol}(S) + \ell_1^2 + \cdots + \ell_n^2)$$

for some suitable constant $C > 0$ which is the desired result. The case $g = 0$ can be similarly deduced, by attaching suitable cylinders with cusps to S instead of disks, and using Theorem 4.7 in place of Theorem 4.6. □

4.2 Harmonic maps

Given a smooth map $f : M \to N$ of Riemannian manifolds, let us define the *tension field* of f to be the trace of the second fundamental form

$$\Delta f = \mathrm{tr}_M \, \mathrm{II}^f \in C^\infty(M, f^*TN).$$

Here, tr_M indicates the contraction of a section of $T^*M \otimes T^*M$ with the dual metric on T^*M, viewed as a section of $TM \otimes TM$. In other words, given a local orthonormal frame $\{e_1, \cdots, e_n\}$ of TM,

$$\Delta f = \mathrm{II}^f(e_1, e_1) + \cdots + \mathrm{II}^f(e_n, e_n).$$

The map f is said to be *harmonic* if its tension field vanishes identically, i.e.

$$\Delta f = 0.$$

For example, it follows from the discussion in Section 4.1.1 that harmonic maps from an interval to a Riemannian manifold are precisely the geodesic paths (of constant speed). To provide intuition in general, we shall recall the variational interpretation of harmonicity. For a smooth map $f : (M, \sigma_M) \to (N, \sigma_N)$ of Riemannian manifolds, the energy of f is defined to be the integral

$$E(f) = \frac{1}{2} \int_M \|df\|^2 \, d\,\mathrm{vol}_M$$

where the norm of $df \in C^\infty(M, T^*M \otimes f^*TN)$ is induced from σ_M and σ_N. Let us consider the variation of the energy functional E in the family of smooth maps $f : M \to N$ of finite energy, subject to one of the following conditions in the case where ∂M is nonempty:

(1) (Dirichlet) $f|_{\partial M} = h$ is fixed.

(2) (Neumann) $df(\nu) \equiv 0$ where ν is the normal vector field to ∂M in M.

The following result is well known.

Proposition 4.9. *The Euler-Lagrange equation for the energy functional E is*

$$\Delta f = 0.$$

In other words, harmonic maps of finite energy are the critical points of E.

Proof. Let $f_t : M \times (-\epsilon, \epsilon) \to N$ be a family of smooth maps with $f = f_0$ satisfying one of the following: (1) $(df_t/dt)|_{\partial M} \equiv 0$, or (2) $df_t(\nu) \equiv 0$ for any normal vector on ∂M. Our goal is to derive the variational formula

$$\left.\frac{d}{dt} E(f_t)\right|_{t=0} = -\int_M \left\langle \Delta f, \left.\frac{df}{dt}\right|_{t=0} \right\rangle \, d\,\mathrm{vol}_M \,. \tag{4.1}$$

where the inner product within the integral is with respect to the metric on f^*TN induced from the metric on N. As one varies the families f_t, this would show that f is a critical point of the energy functional if and only if $\Delta f = 0$. To obtain the variational formula, note first that

$$\begin{aligned}\left.\frac{d}{dt} E(f_t)\right|_{t=0} &= \frac{1}{2}\frac{d}{dt}\int_M \langle df, df\rangle \, d\,\mathrm{vol}_M = \int_M \langle \nabla_{\partial/\partial t}(df), df\rangle \, d\,\mathrm{vol}_M \\ &= \int_M \mathrm{tr}_M \left\langle \nabla_{\partial/\partial t}(df)(-), df(-)\right\rangle_{f^*TN} \, d\,\mathrm{vol}_M \,.\end{aligned}$$

Let us consider the 1-form ω on M given by

$$\omega(X) = \left\langle \left.\frac{df}{dt}\right|_{t=0}, df(X) \right\rangle .$$

Its divergence is given by

$$\begin{aligned}\operatorname{Div}\omega &= -\operatorname{tr}_M \nabla \left\langle \left.\frac{df}{dt}\right|_{t=0}, df(-)\right\rangle \\ &= -\operatorname{tr}_M \left\langle \nabla \left.\frac{df}{dt}\right|_{t=0}, df(-)\right\rangle_{f^*TN} - \operatorname{tr}_M \left\langle \left.\frac{df}{dt}\right|_{t=0}, \nabla df(-)\right\rangle_{f^*TN} \\ &= -\operatorname{tr}_M \left\langle (\nabla_{\partial/\partial t} df)(-), df(-)\right\rangle_{f^*TN} - \left\langle \left.\frac{df}{dt}\right|_{t=0}, \Delta f\right\rangle_{f^*TN}.\end{aligned}$$

Thus, we have

$$\left.\frac{d}{dt} E(f_t)\right|_{t=0} = \int_M \operatorname{Div}\omega \, d\operatorname{vol}_M - \int_M \left\langle \Delta f, \left.\frac{df}{dt}\right|_{t=0}\right\rangle \, d\operatorname{vol}_M .$$

On the other hand, by Green's theorem, we have

$$\int_M \operatorname{Div}\omega \, d\operatorname{vol}_M = \int_{\partial M} \omega(\nu) \, d\operatorname{vol}_{\partial M} = \int_{\partial M} \left\langle \left.\frac{df}{dt}\right|_{t=0}, df(\nu)\right\rangle \, d\operatorname{vol}_{\partial M} .$$

where ν is a unitary normal vector field along ∂M. By our hypothesis on f_t, we conclude that $\int_M \operatorname{Div}\omega \, d\operatorname{vol}_M = 0$, proving the desired variational formula. □

We shall consider a more general notion of "twisted" harmonic maps, or harmonic sections. Given a representation $\rho : \pi_1 M \to \operatorname{Isom}(N)$ of the fundamental group of M into the group of isometries of N, we consider the fibre bundle

$$N_\rho = \bar{M} \times_\rho N$$

over M, where $\bar{M}$ is the universal cover of M. By local triviality of N_ρ, we may equip N_ρ with the metric induced from those of M and N, and extend the notions of second fundamental forms and harmonicity to sections $s : M \to N_\rho$. In particular, a section s is *harmonic* if

$$\Delta s = \operatorname{tr}_M(\nabla(Ds)) = 0$$

where Ds is the vertical differential of s, and ∇ is the connection on $T^*M \otimes s^*VN_\rho$, for VN_ρ the vertical tangent bundle of N_ρ.

4.2.1 Constructing harmonic sections

The first general result for the existence of harmonic maps between Riemannian manifolds was proved by Eells-Sampson [29]. Namely, using the method of heat flow (discussed in the proof of Lemma 4.10 below), they showed that any smooth map between compact Riemannian manifolds is homotopic to a harmonic map when the target has nonpositive curvature. Their work was generalized to include manifolds with boundary by Hamilton [46]. In this subsection, we shall follow their method to construct twisted harmonic maps from surfaces into the hyperbolic three-space $\mathbb{H}^3$. We stress that the negative curvature of $\mathbb{H}^3$ is critical for the success of this approach.

Let Σ be a surface of type (g, n) with $3g + n - 3 > 0$. Let σ be a Riemannian metric on Σ such that $S = (\Sigma, \sigma)$ has constant sectional curvature -1 and each of its boundary curves is geodesic of length 1. Fix a representation $\rho : \pi_1 S \to \mathrm{SL}_2(\mathbb{C})$ with Zariski dense image in $\mathrm{SL}_2(\mathbb{C})$. The group $\mathrm{SL}_2(\mathbb{C})$ acts on $\mathbb{H}^3$ via isometries. We form the associated fibre bundle

$$\mathcal{H} = \bar{S} \times_\rho \mathbb{H}^3$$

where $\bar{S}$ is the universal cover of S, and construct harmonic sections of $\mathcal{H}$ satisfying certain conditions along the boundary curves of S. We have the following solution of Dirichlet problem for harmonic sections.

Lemma 4.10. *Given any smooth section $s_0 : S \to \mathcal{H}$, there is a smooth section s_∞ in the homotopy class of s_0 satisfying*

(1) $\Delta s_\infty = 0$ *and*

(2) $s_\infty|_{\partial S} = s_0|_{\partial S}$.

Proof. When the boundary of S is empty, this was proved by Donaldson [25] using an adaptation of the heat flow argument of Eells-Sampson [29] for harmonic maps to Riemannian manifolds of nonpositive curvature. The case where S has nonempty boundary was proved by Corlette [18, Proof of Theorem 2.1], adapting the work of Hamilton [46] for harmonic maps on manifolds with boundary. The strategy is to construct a 1-parameter family of sections

$$s(x,t) : S \times [0,\infty) \to \mathcal{H}$$

as the solution of the nonlinear heat equation

$$\frac{\partial s}{\partial t} = \Delta s$$

with initial boundary conditions:

(1) $s(-,0) = s_0$, and

(2) $s(-,t)|_{\partial S} = s_0|_{\partial S}$ for all $t \in [0,\infty)$.

Once such a family is constructed, one verifies that there is a sequences of sections $s_t = s(-,t)$ converging uniformly to a harmonic section of $\mathcal{H}$ satisfying the requisite boundary condition. We shall refer to Corlette [18] for details.

Here, we make some simple remarks to provide intuition for the heat flow method. The method is broadly divided into three steps. The first step is demonstrating the short time existence of solutions $s : S \times [0,\epsilon) \to \mathcal{H}$ to the boundary value problem for some $\epsilon > 0$. This is part of the standard theory of partial differential equations, and geometry plays no role here. The second and the third steps are to show that the solution to the boundary problem exists for all time, and that a sequence of sections s_t converges to a harmonic section. For these, one needs suitable uniform bounds on Ds_t/dt and s_t (and their derivatives) that are independent of t; here, the fact that

$\mathbb{H}^3$ has nonpositive sectional curvature will be critically used.

Let us assume for now that such bounds are given and assume that the solution s of the heat equation is defined for all time. Combining the heat equation with the variational formula (4.1) derived in the proof of Proposition 4.9, we have

$$\frac{d}{dt}E(s_t) = -\int_S \left\langle \frac{Ds_t}{dt}, \Delta s_t \right\rangle \, d\,\mathrm{vol}_S = -\int_S \left\| \frac{Ds_t}{dt} \right\|^2 \, d\,\mathrm{vol}_S$$

which shows that $E(s_t)$ is a nonincreasing function of t. Since the energy functional is nonnegative, together with suitable uniform bounds on Ds_t/dt and $s(-,t)$ (and their derivatives) this implies there is a sequence s_t of sections for which Ds_t/dt converges to zero uniformly as $t \to \infty$, and in turn s_t converges uniformly to a section s_∞ with $\Delta s_\infty = 0$, by following the heat equation.

Thus, it remains to obtain time-independent estimates of Ds_t/dt and s_t. (The proof that the image of s_t is bounded uniformly over all time is supplied in the argument of Corlette, *loc.cit.*) For this, we illustrate the importance of $\mathbb{H}^3$ having nonpositive curvature using the following more classical setup. Let $f_0 : M \to N$ be a smooth map between Riemannian manifolds, where M and N are compact, and suppose that $f : M \times [0,T] \to N$ is a solution of the heat equation

$$\frac{\partial f}{\partial t} = \Delta f$$

with boundary conditions $f(-,0) = f_0$ and $f|_{\partial M} = f_0|_{\partial M}$. Let

$$e(x,t) = \frac{1}{2}\|(df_t)_x\|^2$$

denote the *energy density* of $f_t = f(-,t)$ at $x \in M$. For a local orthonormal frame $\{e_i\}$ of TM, we have the following Bochner formula (see for example [46, p.128])

expressing the rate of change of energy density:

$$\frac{\partial e}{\partial t} = \Delta e - \|\nabla(df)\|^2 - \sum_i \langle \mathrm{Ric}_M\, df(e_i), df(e_i)\rangle + \sum_{i,j} \langle R_N(df(e_i), df(e_j))df(e_i), df(e_j)\rangle.$$

Here, the Ricci curvature tensor $\mathrm{Ric}_M = \mathrm{tr}_M\, R_M$ is defined as the trace of the Riemannian curvature tensor R_M on M, and R_N is the Riemannian curvature tensor on N. If N has nonpositive sectional curvature, then, since M is compact,

$$\frac{\partial}{\partial t} e \leq \Delta e + ce$$

for some positive constant c. This allows us to use a *maximum principle* to conclude that the energy density e, and hence $\|df\|$, is uniformly bounded on $M \times [0,T]$, as follows. We reproduce an argument of Hamilton [46, p. 102]. Let

$$h(x,t) = \exp(-(C+1)t) \cdot (e(x,t) - \max e(\cdot, 0)).$$

Then h satisfies $h_0 = h(\cdot, 0) \leq 0$ and $h_t|_{\partial M} \leq 0$ for all $t \in [0,T]$. Moreover,

$$\begin{aligned}\frac{\partial h}{\partial t} &= \exp(-(C+1)t)\frac{\partial e}{\partial t} - C\exp(-(C+1)t) \cdot e - h \\ &\leq \exp(-(C+1)t) \cdot \Delta e - h = \Delta h - h.\end{aligned}$$

Suppose that $(x,t) \in M \times [0,T]$ is a point at which $h(x,t)$ is maximal and positive. At (x,t) we therefore have, in local coordinates $\{x_i\}$:

(i) $\partial h/\partial x_i = 0$ for all i,

(ii) the Hessian $[\partial^2 h/\partial x_i \partial x_j]_{ij}$ is nonpositive as a matrix, and

(iii) $\partial h/\partial t \geq 0$.

The first two conditions imply that $\Delta h \leq 0$ at (x,t), as can be seen from the local expression for Δh. It follows that

$$0 \leq \frac{\partial h}{\partial t} \leq \Delta h - h \leq -h,$$

contradicting our initial hypothesis that $h(x,t) > 0$. This shows that $h(x,t) \leq 0$ for all $(x,t) \in M \times [0,T]$, and hence the energy density $e = \|df\|^2/2$ is uniformly bounded. $\square$

We describe the boundary conditions we shall impose on our harmonic sections. Let $c_1, \cdots, c_n$ be the geodesic boundary curves of S. For each $i = 1, \cdots, n$, let $\rho(c_i)$ denote the conjugacy class in $\mathrm{SL}_2(\mathbb{C})$ determined by the monodromy of ρ along c_i. Let us fix a geodesic parametrization $\gamma_i : [0,1] \to c_i$ of each boundary curve, and a trivialization

$$\gamma_i^* \mathcal{H} = \mathbb{H}^3 \times [0,1]$$

such that the parallel transport $\mathbb{H}^3 = (\gamma_i^* \mathcal{H})_0 \to (\gamma_i^* \mathcal{H})_1 = \mathbb{H}^3$ is given by the action of the unique element $a_i \in \rho(c_i)$ lying in

$$\left\{ \pm \mathbf{1}, \pm \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \right\} \cup \left\{ \begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix} : \theta \in (0,\pi) \right\} \cup \left\{ \begin{bmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{bmatrix} : \lambda \in \mathbb{C}, |\lambda| > 1 \right\}.$$

Let $h_i : c_i \to \mathcal{H}|_{c_i}$ be the unique section such that $h_i \circ \gamma_i : [0,1] \to \mathbb{H}^3 \times [0,1]$ is the graph of the path $f_i : [0,1] \to \mathbb{H}^3$ defined as follows.

(1) If $a_i = \pm \mathbf{1}$, then $f_i \equiv j \in \mathbb{H}^3$.

(2) If a_i is elliptic with unique fixed point $q \in \mathbb{H}^3$, then $f_i \equiv q$.

(3) If a_i is parabolic then

$$f_i(t) = \begin{bmatrix} 1 & t \\ & \\ 0 & 1 \end{bmatrix} j \text{ for all } t \in [0,1].$$

The curve underlying f_i has length 1.

(4) If $a_i = \left[\begin{smallmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{smallmatrix}\right]$ (with $\lambda \in \mathbb{C}$ and $|\lambda| > 1$) is hyperbolic, then

$$f_i(t) = \begin{bmatrix} \lambda^t & 0 \\ & \\ 0 & \lambda^{-t} \end{bmatrix} j \text{ for all } t \in [0,1]$$

This does not depend on the choice of branch for the logarithm of λ. Note that f_i is a geodesic path of length equal to length(a_i).

Let us consider the section

$$h = h_1 \sqcup \cdots \sqcup h_n : \partial S = c_1 \sqcup \cdots \sqcup c_n \to \mathcal{H}|_{\partial S}. \tag{4.2}$$

with the h_i defined as above.

Lemma 4.11. *There is a smooth section $s : S \to \mathcal{H}$ satisfying $s|_{\partial S} = h$.*

Proof. We shall keep the notations introduced above. For each $i = 1, \cdots, n$, let N_i be a closed half-collar neighborhood of the boundary curve c_i such that $N_i \cap N_j = \emptyset$ for $i \neq j$. Let $n_i : [0,1] \times [0,1] \to N_i$ be a parametrization of N_i with $n_i(s,1) = \gamma_i(s)$ and $n_i(0,t) = n_i(1,t)$ for all $s, t \in [0,1]$. We have a trivialization

$$n_i^* \mathcal{H} = \mathbb{H}^3 \times [0,1] \times [0,1]$$

that restricts to the trivialization $\gamma_i^* \mathcal{H} = \mathbb{H}^3 \times [0,1]$ chosen above. Thus, we have a bijective correspondence between the set of continuous sections $N_i \to \mathcal{H}|_{N_i}$ and the

set S_i of continuous maps $g : [0,1]^2 \to \mathbb{H}^3$ satisfying $g(1,t) = a_i \cdot g(0,t)$ for every $t \in [0,1]$. Given any $g_0 \in S_i$, there is a homotopy $g : [0,1]^2 \times [0,1] \to \mathbb{H}^3$ such that $g_u = g(-,-,u) \in S_i$ for any $u \in [0,1]$ and

(1) $g(s,t,0) = g_0(s,t)$,

(2) $g(s,0,u) = g_0(s,0)$, and

(3) $g(s,1,1) = f_i(s)$

for every $s,t,u \in [0,1]$. This can be seen, for instance, from the fact that between any two paths $m_0, m_1 : [0,1] \to \mathbb{H}^3$ satisfying $m_0(1) = a_i \cdot m_1(0)$ for $j = 0,1$ there is a homotopy $m : [0,1] \times [0,1] \to \mathbb{H}^3$ satisfying $m(-,j) = m_j$ for $j = 0,1$ and $m(1,t) = a_i \cdot m(0,t)$ for every $t \in [0,1]$. Therefore, given any section $s_0 : S \to \mathcal{H}$, considering the restrictions $s_0|_{N_i}$ for each $i = 1, \cdots, n$ and applying the above homotopy, we obtain a continuous section $s : S \to \mathcal{H}$ satisfying $s|_{\partial S} = h$. It is easily seen that we may arrange s to be smooth. □

Corollary 4.12. *There is a harmonic section $s : S \to \mathcal{H}$ satisfying $s|_{\partial S} = h$.*

Proof. This follows by combining Lemmas 4.10 and 4.11. □

4.3 Systoles of local systems

4.3.1 Archimedean systoles

Let Σ be a surface of type (g,n) satisfying $3g + n - 3 > 0$, with boundary curves $c_1, \cdots, c_n$. Given a representation $\rho : \pi_1\Sigma \to \mathrm{SL}_2(\mathbb{C})$ or a corresponding local system on Σ, we define its *systole* to be

$$\mathrm{sys}(\rho) = \inf\{\text{length}\, \rho(a) : a \subset \Sigma \text{ is an essential curve}\}$$

where length $\rho(a)$ denotes the translation length of the conjugacy class $\rho(a)$ in $\mathrm{SL}_2(\mathbb{C})$ acting on the hyperbolic three-space $\mathbb{H}^3$, as defined in Section 4.1.3. This recovers the usual notion of systole for hyperbolic surfaces (see Section 4.1.5) when $\rho : \pi_1\Sigma \to \mathrm{SL}_2(\mathbb{R})$ is a Fuchsian representation arising from a hyperbolic structure on Σ with geodesic boundary. By the relation between translation lengths and traces of elements in $\mathrm{SL}_2(\mathbb{C})$ given in Section 4.1.3, we see that Theorem 4.1 follows from the following claim.

Theorem 4.13. *There exists a real number $C > 0$ such that, for any representation $\rho : \pi_1\Sigma \to \mathrm{SL}_2(\mathbb{C})$, we have $C\,\mathrm{sys}(\rho)^2 \leq 2\pi|\chi(\Sigma)| + \sum_{i=1}^{n} \mathrm{length}\,\rho(c_i)^2$.*

Proof. Let us first choose a Riemannian metric σ on Σ such that $S = (\Sigma, \sigma)$ is a hyperbolic surface with geodesic boundary curves $c_1, \cdots, c_n$, each of length 1. Let us fix a representation $\rho : \pi_1\Sigma \to \mathrm{SL}_2(\mathbb{C})$, which we assume has Zariski dense image in $\mathrm{SL}_2(\mathbb{C})$ since otherwise the theorem is easy to deduce; this is easily seen from the classification of algebraic subgroups of $\mathrm{SL}_2(\mathbb{C})$, as found in Chapter 6. Consider the fibre bundle

$$\mathcal{H} = \bar{\Sigma} \times_\rho \mathbb{H}^3$$

where $\bar{\Sigma}$ is the universal cover of Σ. We may equip $\mathcal{H}$ with the Riemannian metric induced from the hyperbolic metric on $\mathbb{H}^3$ and the metric σ on S. Let us fix a section $h : \partial S \to \mathcal{H}$ as given by (4.2). By Corollary 4.12, there exists a harmonic section

$$s : S \to \mathcal{H}$$

satisfying $s|_{\partial S} = h$. For each $\epsilon > 0$, let $S(\epsilon)$ be the surface equipped with the scaled metric $\epsilon\sigma$, and let $\mathcal{H}[\epsilon]$ be the fibre bundle with metric induced from the hyperbolic metric on $\mathbb{H}^3$ and the metric on $S(\epsilon)$. Note that $s : S(\epsilon) \to \mathcal{H}[\epsilon]$ remains harmonic. Let

$$f : \bar{\Sigma} \to \mathbb{H}^3$$

be the ρ-equivariant map corresponding to s. Let us denote by σ_0 the standard hyperbolic metric on $\mathbb{H}^3$. The nonnegative symmetric tensor $f^*\sigma_0$ on $\bar{\Sigma}$ descends to Σ since f is ρ-equivariant. Let us consider the metric

$$\sigma(\epsilon) = f^*\sigma_0 + \epsilon\sigma$$

on the surface Σ. In fact, $\sigma(\epsilon)$ is precisely metric induced from $\mathcal{H}[\epsilon]$ on the image $M(\epsilon) = s(S(\epsilon))$ of the harmonic section $s : S(\epsilon) \to \mathcal{H}[\epsilon]$. Note that

$$\operatorname{length} \rho(a) \leq \operatorname{length}_{M(\epsilon)}(a)$$

for any essential closed geodesic a on $M(\epsilon)$, the left hand side being the translation length of $\rho(a)$ as it acts on $\mathbb{H}^3$. In particular, we see that $\operatorname{sys}(\rho) \leq \operatorname{sys}(M(\epsilon))$ for every $\epsilon > 0$. Now, note that the lengths of the boundary curves $c_1, \cdots, c_n$ of $M(\epsilon)$ satisfy

$$\lim_{\epsilon\to 0} \operatorname{length}_{M(\epsilon)}(c_i) = \begin{cases} 1 & \text{if } \rho(c_i) \text{ is parabolic, and} \\ \operatorname{length}(\rho(c_i)) & \text{otherwise.} \end{cases}$$

Moreover, since $\sigma(\epsilon)$ tends to the symmetric tensor $f^*\sigma_0$ uniformly as $\epsilon \to 0$, we have

$$\lim_{\epsilon\to 0} \operatorname{Vol}(M(\epsilon)) = \int_\Sigma d\operatorname{vol}_{f^*\sigma_0}\,.$$

(The "volume form" $d\operatorname{vol}_{f^*\sigma_0}$ associated to the nonnegative symmetric tensor $f^*\sigma_0$ may vanish at some places.) Therefore, by Corollary 4.8, for all sufficiently small $\epsilon > 0$, we have

$$C_1 \operatorname{sys}(M(\epsilon))^2 \leq \int_\Sigma d\operatorname{vol}_{f^*\sigma_0} + \sum_{i=1}^{n} \max\{1, \operatorname{length}(\rho(c_i))^2\}$$

for some absolute constant $C_1 > 0$. The second summand on the right hand side is bounded by our choice of bounded set $K \subset \mathbb{C}^n$. Therefore, it will suffice to show that

$$\int_\Sigma d\,\mathrm{vol}_{f^*\sigma_0} \leq 2\pi|\chi(\Sigma)| + c$$

for an absolute constant $c > 0$. The remainder of the proof is devoted to demonstrating this claim (due essentially to Reznikov), combining computations contained in the proof of Lemma 2.5 of Deroin-Tholozan [23] with an argument of Reznikov [75].

By a result of Sampson [77, Corollary of Theorem 3], the open subset $U \subseteq \Sigma$ corresponding to the locus in $\bar{\Sigma}$ where df has full rank is either empty or dense. Since the former case is trivial, we shall assume that df has full rank on a dense open subset of Σ. Locally, given a choice of orthonormal frame e_1, e_2 of $TM(\epsilon)$, by Gauss's Theorem (Theorem 4.5) we have

$$K^{M(\epsilon)} = K^{\mathcal{H}[\epsilon]}(TM(\epsilon)) + \langle \mathrm{II}^{M(\epsilon)}(e_1, e_1), \mathrm{II}^{M(\epsilon)}(e_2, e_2)\rangle - \| \mathrm{II}^{M(\epsilon)}(e_1, e_2)\|^2$$

where $\mathrm{II}^{M(\epsilon)}$ is the second fundamental form of $M(\epsilon)$ immersed in $\mathcal{H}[\epsilon]$, and the inner product is with respect to the metric on the normal bundle $NM(\epsilon)$ of $M(\epsilon)$. Choosing a local orthonormal frame $\{n_1, n_2, n_3\}$ of $NM(\epsilon)$, we have

$$\begin{aligned}
&\langle \mathrm{II}^{M(\epsilon)}(e_1, e_1), \mathrm{II}^{M(\epsilon)}(e_2, e_2)\rangle - \| \mathrm{II}^{M(\epsilon)}(e_1, e_2)\|^2 \\
&= \sum_{k=1}^{3} \det(\langle \mathrm{II}^{M(\epsilon)}(e_i, e_j), n_k\rangle) \\
&= \frac{1}{\mathrm{Jac}(s)^2} \sum_{k=1}^{3} \det_{\epsilon\sigma}(\langle \mathrm{II}^{M(\epsilon)}(ds(-), ds(-)), n_k\rangle) \\
&= \frac{3}{\mathrm{Jac}(s)^2} \mathbb{E}\left(\det_{\epsilon\sigma}\langle \mathrm{II}^{M(\epsilon)} \circ ds, n\rangle\right)
\end{aligned}$$

where the average $\mathbb{E}$ is taken over all unitary normal vectors n in $NM(\epsilon)$. Note that

we have by definition of the second fundamental form of s

$$\begin{aligned}\mathrm{II}^s(X,Y) &= \nabla^s_X Ds(Y) - Ds(\nabla^{\epsilon\sigma}_X Y) = \nabla^{s^*\sigma(\epsilon)}_X ds(Y) - ds(\nabla^{\epsilon\sigma}_X Y) \\ &= \mathrm{II}^{M(\epsilon)}(ds(X), ds(Y)) + ds\left(\nabla^{s^*\sigma(\epsilon)}_X Y - \nabla^{\epsilon\sigma}_X Y\right).\end{aligned}$$

But note that

$$\mathrm{II}^{M(\epsilon)}(ds(X), ds(Y)) \quad \text{and} \quad ds\left(\nabla^{s^*\sigma(\epsilon)}_X Y - \nabla^{\epsilon\sigma}_X Y\right)$$

are orthogonal. Therefore, taking the trace of both sides and using the harmonicity of s (i.e. $\mathrm{tr}_\sigma \mathrm{II}^s = 0$), we have $\mathrm{tr}_{\epsilon\sigma}(\mathrm{II}^{M(\epsilon)} \circ ds) = 0$, and *a fortiori*

$$\mathrm{tr}_{\epsilon\sigma}\langle \mathrm{II}^{M(\epsilon)}(ds(-), ds(-)), n\rangle = 0$$

for every normal vector $n \in NM(\epsilon)$. Since $\langle \mathrm{II}^{M(\epsilon)}(ds(-), ds(-)), n\rangle$ is a symmetric tensor, its eigenvalues (with respect to the metric $\epsilon\sigma$ on the surface) are real, and the vanishing of trace therefore implies $\det_{\epsilon\sigma}\langle \mathrm{II}^{M(\epsilon)}(ds(-), ds(-)), n\rangle \leq 0$. Thus, we conclude from Gauss's theorem that

$$-K^{\mathcal{H}[\epsilon]}(TM(\epsilon)) \leq -K^{M(\epsilon)}.$$

Applying the Gauss-Bonnet formula, we therefore have

$$\begin{aligned}-\int_{M(\epsilon)} K^{\mathcal{H}[\epsilon]}(TM(\epsilon))\, d\,\mathrm{vol}_{\sigma(\epsilon)} &\leq \int_{M(\epsilon)} (-K^{M(\epsilon)})\, d\,\mathrm{vol}_{\sigma(\epsilon)} \\ &\leq 2\pi|\chi(\Sigma)| + \int_{\partial M(\epsilon)} k^{M(\epsilon)}\, ds_{\sigma(\epsilon)}.\end{aligned}$$

Note that $-K^{\mathcal{H}[\epsilon]}$ is everywhere nonnegative by part (1) of Lemma 4.3, since $\mathbb{H}^3$ and

$S(\epsilon)$ have nonpositive sectional curvature. Furthermore, we have

$$\lim_{\epsilon\to 0} K^{\mathcal{H}[\epsilon]}(TM(\epsilon)) = K^{\mathbb{H}^3}(df(T\Sigma)) = -1$$

on the dense open set $U \subset \Sigma$ where df has full rank, by part (2) of Lemma 4.3. Recalling also that $\lim_{\epsilon\to 0}\sigma(\epsilon) = f^*\sigma_0$, we apply Fatou's lemma to deduce that

$$\begin{aligned}\int_\Sigma d\,\mathrm{vol}_{f^*\sigma_0} &\leq -\liminf_{\epsilon\to 0}\int_{M(\epsilon)} K^{\mathcal{H}[\epsilon]}(TM(\epsilon))\,d\,\mathrm{vol}_{\sigma(\epsilon)}\\ &\leq 2\pi|\chi(\Sigma)| + \liminf_{\epsilon\to 0}\int_{M(\epsilon)} k^{M(\epsilon)} ds_{\sigma(\epsilon)}.\end{aligned}$$

It therefore remains only to bound the total curvature of $\partial M(\epsilon)$ for all sufficiently small ϵ. But this was obtained in Lemma 4.14 below. This completes the proof. □

Lemma 4.14. *For $\epsilon > 0$, let $c_i(\epsilon)$ be the curve c_i equipped with the metric such that its length is ϵ. We have*

$$\lim_{\epsilon\to 0}\int_{c_i(\epsilon)} k^{\mathcal{H}[\epsilon]}\, ds_{\sigma(\epsilon)} = \begin{cases} 0 & \text{if } \rho(c_i) \text{ is central, elliptic, or hyperbolic,}\\ 1 & \text{if } \rho(c_i) \text{ is parabolic.}\end{cases}$$

Proof. We note that $h : c_i(\epsilon) \to \mathcal{H}[\epsilon]|_{c_i}$ is a geodesic path if $\rho(c_i)$ is central, elliptic, or hyperbolic. We shall therefore assume that $\rho(c_i)$ is parabolic. Recall that, in terms of the global coordinates (x_1, x_2, x_3) on $\mathbb{H}^3$ given above, the Christoffel symbols for the Levi-Civita connection on $\mathbb{H}^3$ are

$$\begin{aligned}\Gamma_{ij,k} &= \langle \nabla_{\partial/\partial x_i}(\partial/\partial x_j), \partial/\partial x_k\rangle\\ &= \frac{1}{2}\left(\frac{\partial}{\partial x_j}\left\langle\frac{\partial}{\partial x_i},\frac{\partial}{\partial x_k}\right\rangle + \frac{\partial}{\partial x_i}\left\langle\frac{\partial}{\partial x_j},\frac{\partial}{\partial x_k}\right\rangle - \frac{\partial}{\partial x_k}\left\langle\frac{\partial}{\partial x_i},\frac{\partial}{\partial x_j}\right\rangle\right).\end{aligned}$$

Thus, we have $\Gamma_{11,1} = \Gamma_{11,2} = 0$ and $\Gamma_{11,3} = 1/x_3$, and

$$\frac{D}{dt}\frac{df_i}{dt} = \Gamma_{11,3}x_3^2\frac{\partial}{\partial x_3} = \frac{1}{x_3}\frac{\partial}{\partial x_3}.$$

In particular, we have $k^{\mathbb{H}^3} \equiv 1$ on the image of f_i (in the notation of Section 4.2.1). It is easy to observe that $k^{\mathcal{H}[\epsilon]}$ tends to $k^{\mathbb{H}^3}$ uniformly as $\epsilon \to 0$. The desired result follows from this. □

4.3.2 Nonarchimedean systoles

Let Σ be a surface of type (g, n) satisfying $3g + n - 3 > 0$, with boundary curves $c_1, \cdots, c_n$. Let $\mathcal{O}$ be a discrete valuation ring, and let F be its fraction field. Given a representation $\rho : \pi_1\Sigma \to \mathrm{SL}_2(F)$ or a corresponding local system on Σ, we define its *systole* to be

$$\mathrm{sys}(\rho) = \inf\{\mathrm{length}\,\rho(a) : \text{essential curve } a \subset \Sigma\}$$

where $\mathrm{length}\,\rho(a)$ denotes the translation length of the conjugacy class $\rho(a)$ acting on the Bruhat-Tits tree Λ of $\mathrm{SL}_2(F)$, as defined in Section 4.1.4. Theorem 4.2 is a special case of the following result.

Theorem 4.15. *There exists a real number $C > 0$ such that, for any representation $\rho : \pi_1\Sigma \to \mathrm{SL}_2(F)$, we have $\mathrm{sys}(\rho)^2 \leq C \cdot \sum_{i=1}^{n}(\mathrm{length}\,\rho(c_i))^2$.*

Proof. Let us consider the fiber bundle

$$\Lambda_\rho = \Lambda \times_\rho \bar{\Sigma}$$

over Σ, where $\bar{\Sigma}$ is the universal cover of Σ. Our first step is to construct a section $s : \Sigma \to \Lambda_\rho$ satisfying the following two conditions.

(1) The subset $B \subset \Sigma$ where s fails to be smooth is finite.

(2) For each $i = 1, \cdots, n$, given some parametrization $\gamma_i : [0,1] \to c_i$ of the boundary curve c_i and trivialization

$$\gamma_i^* \Lambda_\rho \simeq \Lambda \times [0,1]$$

the restriction $\gamma_i^* s : [0,1] \to \Lambda \times [0,1]$ is the graph of a parametrization of a geodesic curve whose length equals $\text{length}(\rho(c_i))$.

In condition (1) above and our arguments, in consideration of the smoothness of s we shall take into account the various "apartments" of Λ, as in the work of Gromov and Schoen (see [43, p.178] for an elucidation). The construction of s can be given as follows. Let $x_0 \in \Sigma$ be a fixed base point lying on the interior of the surface. For each $j = 1, \cdots, n$, let x_j be a base point on c_j, and let γ_j be a simple loop on c_j based at x_j, so that the interior of Σ lies to the left as one traves along γ_j. Let $\alpha_1, \cdots, \alpha_{2g}$ be simple loops based on x_0, and let β_j for each $j = 1, \cdots, n$ be a simple path from x_0 to x_j not intersecting any of the aforementioned loops except (at the endpoints), such that the sequence

$$(\alpha_1, \alpha_2, \cdots, \alpha_{2g-1}, \alpha_{2g}, \beta_1 \gamma_1 \beta_1^{-1}, \cdots, \beta_n \gamma_n \beta_n^{-1})$$

of loops gives the standard generators of $\pi_1(\Sigma, x_0)$ pairwise intersecting only at the base point x_0. Fix a trivialization $\Psi_0 : (\Lambda_\rho)_{x_0} \simeq \Lambda$ of the fiber of Λ_ρ above x_0. By parallel transport this induces trivializations

- $\alpha_i^* \Lambda_\rho \simeq [0,1] \times \Lambda$ for $i = 1, \cdots, 2g$, and
- $\beta_j^* \Lambda_\rho \simeq [0,1] \times \Lambda$ for $j = 1, \cdots, n$,

agreeing with Ψ_0 at 0. For $j = 1, \cdots, n$, let $\Psi_j : (\Lambda_\rho)_{x_j} \simeq \Lambda$ be obtained from the second set of trivializations at 1. By parallel transport, we have a trivialization

- $\gamma_j^* \Lambda_\rho \simeq [0,1] \times \Lambda$ for $j = 1, \cdots, n$,

agreeing with Ψ_j at 0. Let a_i be the curve underlying α_i. Similarly, let b_j denote the 1-submanifold of Σ underlying β_i. By construction, the curve underlying the loop γ_j is the boundary curve c_j. Let

$$C = \bigcup_{i=1}^{2g} a_i \cup \bigcup_{j=1}^{n} (b_j \cup c_j).$$

Fix $v_0 \in \Lambda$. We have a section $t : C \to \Lambda_\rho|_C$ given by the following properties under the various trivializations above:

- $t(x_0) = v_0$ under the trivialization Ψ_0,
- the composition $t \circ \alpha_i$ corresponds to the graph of the geodesic path on Λ from v_0 to $\rho(\alpha_j) \cdot v_0$,
- the composition $t \circ \gamma_j$ corresponds to the graph of a geodesic path on Λ (from point v_j to v_j', say) whose length equals length $\rho(c_j)$, and
- the composition $t \circ \beta_j$ corresponds to the graph of the geodesic path on Λ from v_0 to the point v_j chosen above.

Our goal is to extend t to a section on the entire surface Σ as follows. Note that the latter corresponds to a ρ-equivariant map $\bar{\Sigma} \to \Lambda$. The complement of the preimage $\bar{C}$ of C in $\bar{\Sigma}$ divides $\bar{\Sigma}$ into a countable collection of components, each of which is an open fundamental domain for the covering $\bar{\Sigma} \to \Sigma$. Let $\mathcal{F}$ be one of them. Since t already specifies for us a ρ-equivariant map $\bar{C} \to \Lambda$ (and in particular a map $\partial\mathcal{F} \to \Lambda$ we shall call f), to construct a section $s : \Sigma \to \Lambda_\rho$ extending t it remains only to construct a map $F : \bar{\mathcal{F}} \to \Lambda$ such that $F|_{\partial\mathcal{F}} = f$. Note that the image of f is contained in a subtree $T \subset \Lambda$ with finitely many edges.

Let D be the closed unit disk. Let us fix a homeomorphism $\mathcal{F} \simeq D$ which is a diffeomorphism on the interior and restricts to a diffeomorphism on each of the

"edges" of $\partial\mathcal{F}$. Fix $0 < \epsilon < 1$. There is then a homotopy $g : [0, \epsilon] \times \partial D \to T$ such that $g_0 = f$ under the identification above, and $g|_{(0,\epsilon]}$ is smooth. (This serves to "round out the corners" of the map f.) Consideration of deformation retractions of T to a point $p \in T$ (or invoking a picture as in [43, p.178]) shows that we may extend g to a homotopy $[0, 1] \times \partial D \to T$ such that:

(i) $g|_{(0,1]}$ is smooth at all but finitely many points, and

(ii) $g_1 \equiv p$.

By construction, this factors through the quotient

$$F : \bar{\mathcal{F}} \simeq D \simeq ([0, 1] \times \partial D)/((x, 1) \sim (x', 1)) \to T$$

where the homeomorphism from $D \simeq ([0, 1] \times \partial D)/((x, 1) \sim (x', 1))$ appearing above is chosen in the obvious way (so that in particular it is a diffeomorphism away from the origin $o \in D$). This gives rise to a section $s : \Sigma \to \Lambda_\rho$ which is smooth away from C and finitely many points. By considering restrictions of Λ_ρ over open neighborhoods of a_i, b_j, and c_j, we see that s may be deformed to a section which is smooth away from a finite set $B \subset \Sigma$. This completes the construction of the section $s : \Sigma \to \Lambda_\rho$ satisfying the conditions (1) and (2) mentioned in the beginning.

Let σ be an arbitrary Riemannian metric on Σ. On the complement $\Sigma \setminus B$ of the finite set B given as in (1), we have a Riemannian metric given by

$$\sigma'(\epsilon) = s^*(\sigma_\Lambda + \epsilon\sigma)$$

for $\epsilon > 0$, where σ_Λ is the Euclidean piecewise metric on the Bruhat-Tits tree Λ. Moreover, for any curve $a \subset \Sigma$ its length with respect to $\sigma'(\epsilon)$ is well-defined. Setting

$s(\epsilon) = \inf\{\text{length}(a, \sigma'(\epsilon)) : a \subset \Sigma \text{ essential curve}\}$, we find that

$$\text{sys}(\rho) \leq s(\epsilon). \tag{a}$$

Our next step is to modify $\sigma'(\epsilon)$ to obtain a genuine Riemannian metric on Σ. More precisely, let $\{D_x\}_{x \in B}$ be a collection of pairwise disjoint closed disks lying on the interior of Σ, such that, for each $x \in B$,

(i) D_x contains x in its interior,

(ii) $\text{Vol}(D_x, \sigma'(\epsilon)) < \epsilon^2$, and

(iii) $\text{length}(\partial D_x, \sigma'(\epsilon)) < \epsilon$.

Let $\sigma(\epsilon)$ be a Riemannian metric on Σ obtained by extending the metric $\sigma'(\epsilon)$ on the complement of $\bigcup_{x \in B} D_x$ to Σ in such a way that $\sigma(\epsilon)|_{D_x}$ is the round metric on a half-sphere with base circumference equal to $\text{length}(\partial D_x, \sigma'(\epsilon))$, up to negligible smoothing. By our construction, we have

$$\text{Vol}(\Sigma, \sigma(\epsilon)) = \text{Vol}(\Sigma \setminus B, \sigma'(\epsilon)) + O(\epsilon^2) \tag{b}$$

and

$$s(\epsilon) \leq \text{sys}(\Sigma, \sigma(\epsilon)). \tag{c}$$

To observe the latter, we note that a shortest path between any two points on the boundary of the half-sphere $(D_x, \sigma(\epsilon))$ may be taken to be on the boundary ∂D_x. Using the fact that $\sigma'(\epsilon)$ tends to $s^*\sigma_\Lambda$ uniformly on Σ' as ϵ tends to 0 and the fact that Λ is a one-dimensional Riemannian simplicial complex, we conclude that

$$\lim_{\epsilon \to 0} \text{Vol}(\Sigma \setminus B, \sigma'(\epsilon)) = 0.$$

Therefore, taking $\epsilon > 0$ sufficiently small and combining the observations (a), (b), and (c) above with Corollary 4.8, we obtain the desired result. □

Let $X = X(\Sigma)$ be the moduli of local systems on Σ as defined in Chapter 2. Theorem 4.2 has the following corollary. A further special case, Lemma 4.17 below, will be used Section 5.2 of this thesis.

Corollary 4.16. *Let $\mathcal{O}$ and F be as above. If F has characteristic zero, then for any $\rho \in X_k(F)$ with $k \in \mathcal{O}^n$ there is an essential curve $a \subset \Sigma$ such that* $\operatorname{tr} \rho(a) \in \mathcal{O}$.

Proof. Given $\rho \in X_k(F)$, since F has characteristic zero there exists a finite field extension F'/F such that ρ is the class of a representation $\rho' : \pi_1\Sigma \to \mathrm{SL}_2(F')$. Choosing an extension of the valuation on F to F', let $\mathcal{O}' \subset F'$ be the associated valuation ring. It follows from Theorem 4.1 and our hypothesis $k \in \mathcal{O}^n \subseteq (\mathcal{O}')^n$ that there is an essential curve $a \subset \Sigma$ with $\operatorname{tr} \rho'(a) \in \mathcal{O}'$. This then implies that $\operatorname{tr} \rho(a) = \operatorname{tr} \rho'(a) \in \mathcal{O}' \cap F = \mathcal{O}$, which is the desired result. □

Lemma 4.17. *Given any morphism $f : \mathbb{A}^1 \to X_k$ over $\mathbb{C}$, there is an essential curve $a \subset \Sigma$ such that* $\operatorname{tr}_a \circ f : \mathbb{A}^1 \to \mathbb{A}^1$ *is constant.*

Proof. A morphism $f : \mathbb{A}^1 \to X_k$ corresponds to a $\mathbb{C}[t]$-valued point of X_k, giving rise to a $\mathbb{C}(t)$-valued point $\rho_f \in X_k(\mathbb{C}(t))$. Applying Corollary 4.16 with $F = \mathbb{C}(t)$, with discrete valuation given by the order of vanishing at ∞, we deduce that there is an essential curve $a \subset \Sigma$ such that $\operatorname{tr} \rho_f(a) = \operatorname{tr}_a \circ f : \mathbb{A}^1 \to \mathbb{A}^1$ has no pole at ∞, which implies that it must be constant, as desired. □

Chapter 5

Diophantine results

Let Σ be a surface of type (g, n) satisfying $3g + n - 3 > 0$. Let $X = X(\Sigma)$ and $X_k = X_k(\Sigma)$ be the moduli of local systems introduced in Chapter 2. In this chapter, we derive the main Diophantine results of our thesis on the moduli spaces.

In Section 5.1, we prove the following result, using as our main ingredient the boundedness of systoles for $\mathrm{SL}_2(\mathbb{C})$-local systems obtained in Chapter 4. We shall endow the sets of complex points of algebraic varieties with the analytic topology.

Theorem 5.1. *Let $K \subset \mathbb{A}^n(\mathbb{C})$ be a compact set. Suppose that*

(1) *$A \subset \mathbb{R}$ with $\mathrm{dist}_{\mathbb{R}}(A, \{\pm 2\}) > 0$, or*

(2) *$A \subset \mathbb{C}$ with $\mathrm{dist}_{\mathbb{C}}(A, [-2, 2]) > 0$.*

There is then a compact subset $L \subseteq X(\mathbb{C})$ satisfying $X_K(A) \subseteq \Gamma \cdot L$.

Upon restriction to the locus of Fuchsian representations, Theorem 5.1 recovers Mumford's compactness theorem [70] for Riemann surfaces. It also generalizes a related work of Bowditch, Maclachlan, and Reid [12, Theorem 2.1]. When the subset A is moreover assumed to be closed and discrete in $\mathbb{R}$ or $\mathbb{C}$, Theorem 5.1 implies that $X_K(A)$ consists of finitely many mapping class group orbits. This extends McKean's finiteness theorem [67] for length isospectral families of closed Riemann surfaces.

Further specializing to $A = \mathbb{Z} \setminus \{\pm 2\}$, we conclude that, for each $k \in \mathbb{A}^n(\mathbb{Z})$, the set of integral points on X_k having trace $\neq \pm 2$ along essential curves of Σ consists of finitely many mapping class group orbits. A similar conclusion can be drawn for the set of all imaginary quadratic integral points on X_k, if we consider $A = \bigcup_{d>0} O_d \setminus [-2, 2]$ where each O_d denotes the ring of integers of the imaginary quadratic field $\mathbb{Q}(\sqrt{-d})$.

In Section 5.2, we apply Theorem 4.2 to the field of rational functions to study the structure of morphisms from the affine line into the moduli space X_k for $k \in \mathbb{A}^n(\mathbb{C})$. Recall from Chapter 1 that we defined an algebraic variety V over $\mathbb{C}$ to be *parabolic* if every closed point of V lies in the image of a nonconstant morphism $\mathbb{A}^1 \to V$. We obtain the following.

Theorem 5.2. *There is a closed parabolic subvariety $V \subsetneq X_k$ such that every nonconstant morphism $\mathbb{A}^1 \to X_k$ is mapping class group equivalent to some $\mathbb{A}^1 \to V$.*

In fact, our analysis yields a more precise result (Theorem 5.16), giving a modular characterization of points lying on parabolic subvarieties of X_k. Away from the exceptional case $(g, n, k) = (1, 1, 2)$, these points correspond to the local systems which send an essential curve to a parabolic conjugacy class in $\mathrm{SL}_2(\mathbb{C})$ or have parabolic restriction to a nontrivial "pair of pants" in the surface. Theorem 5.2 is reminiscent of a result of Sterk [85] that the automorphism group of a projective K3 surface acts on the set of its smooth rational curve classes with finitely many orbits. Our work has an interesting consequence, Corollary 5.18, that certain polynomial deformations of Fuchsian surface groups must be isotrivial. Combining Theorems 5.1 and 5.16, we conclude the following result.

Theorem 5.3. *For each $k \in \mathbb{A}^n(\mathbb{Z})$, the set of nondegenerate integral points on X_k consists of finitely many mapping class group orbits. The degenerate integral points are generated from points on a parabolic subvariety of X_k.*

In Section 5.3, we discuss our work in the context of arithmetic hyperbolic surfaces,

whose theory can be used to derive nontrivial statements about the nondegenerate integral orbits on X_k in certain cases. In Section 5.4, we undertake the Diophantine analysis of integral points on nondegenerate algebraic curves in X_k. We define an algebraic subvariety $V \subseteq X_k$ to be *integrable* if there is a pants decomposition P of Σ such that tr_a is constant on V for every curve $a \subset P$. Otherwise, V is *nonintegrable*. We first prove the following in Section 5.4, by applying Theorem 4.2 to function fields of affine algebraic curves equipped with valuations given by points at infinity.

Theorem 5.4. *If $C \subset X_k$ is a geometrically irreducible nonintegrable algebraic curve, then $C(A)$ is finite for any closed discrete set $A \subset \mathbb{C}$.*

Moreover, our method will show that, given an embedding of C into affine space, the sizes of the coordinates of $C(A)$ from the theorem can be effectively determined. Applying Theorem 5.4 with $A = \bigcup_{d>0} O_d$, we conclude that a nonintegrable curve in X_k has at most finitely many imaginary quadratic integral points. As a special case, this recovers the finiteness result of Long and Reid [60] for imaginary quadratic integral points on character curves of one-cusped hyperbolic three-manifolds. Our approach to finiteness of integral points on nonintegrable curves shares its basis with Runge's method (see [99]).

By combining Theorem 5.4 with an analysis of integrable algebraic curves using Baker's theory on linear forms in logarithms, we also obtain the following result. We shall refer to an element in the mapping class group of Σ as a *multitwist* if it is given by a product of commuting Dehn twists (and their powers) along essential curves in a pants decomposition of Σ.

Theorem 5.5. *Let $C \subset X_k$ be a nondegenerate geometrically irreducible algebraic curve over $\mathbb{Z}$. Then $C(\mathbb{Z})$ can be effectively determined. Moreover,*

(1) *$C(\mathbb{Z})$ is finite, or*

(2) *$C(\mathbb{Z})$ is finitely generated under some nontrivial multitwist preserving C.*

If moreover C is not fixed pointwise by any nontrivial multitwist, the same result holds with $C(\mathbb{Z})$ replaced by the set of all imaginary quadratic integral points on C.

Theorem 5.5 gives a complete analysis of integral points on every nondegenerately embedded curve $C \subset X_k$ with no intrinsic restrictions, e.g. regarding the genus of the curve. The structure of Theorem 5.5 is strongly reminiscent of, and motivated by, classical Diophantine results on subvarieties in log Calabi-Yau varieties of linear type, such as algebraic tori and abelian varieties. Finally, for $(g, n) = (1, 1)$ or $(0, 4)$, the moduli space X_k has an explicit presentation as an affine cubic algebraic surface with equation of the form

$$x^2 + y^2 + z^2 + xyz = Ax + Dy + Cz + D \tag{5.1}$$

for some constants A, B, C, D depending on k. Affine algebraic surfaces of this type were first studied by Markoff [65], where he introduced a form of nonlinear descent which essentially coincides the mapping class group action. In this setting, one can give an elementary proof of Theorem 4.2 using Markoff's descent in a nonarchimedean setting; see also a related passage in [36, Section 5] discussing parametric solutions in Markoff type surfaces. Our work specializes to the following result, which may be proved elementarily (but still using the group action).

Corollary 5.6. *On any affine algebraic surface with an equation of the form* (5.1)*, the integral solutions to any Diophantine equation over $\mathbb{Z}$ can be effectively determined.*

The work of Ghosh and Sarnak [36] shows that, in the sense of proportions, almost all "admissible" Markoff type surfaces X_k for $(g, n) = (1, 1)$ have a Zariski dense set of integral points. Thus, Corollary 5.6 provides an infinite family of nontrivial ambient varieties of dimension two where every Diophantine equation over $\mathbb{Z}$ can be effectively solved.

5.1 Compactness criterion

Let Σ be a surface of type (g, n) satisfying $3g + n - 3 > 0$. The purpose of this section is to prove Theorem 5.1. The intuition behind our proof is the following geometric picture. By the work of Goldman (see for example [39]), the variety X carries an algebraic Poisson structure for which the subvarieties X_k, $k \in \mathbb{A}^n(\mathbb{C})$, form symplectic leaves in a suitable sense. In particular, given a pants decomposition P of Σ, the immersion $P \to \Sigma$ induces a morphism

$$\mathrm{tr}_P : X_k \to X(P) \simeq \mathbb{A}^{3g+n-3}$$

which should morally furnish the structure of an "integrable system" in the sense of Hamiltonian mechanics: a typical fibre of this morphism would be a "torus" of dimension $3g + n - 3$, preserved by the pairwise commuting actions of Dehn twists along the curves in P. (Moreover, these Dehn twists actions would arise from suitable Hamiltonian flows associated to the trace functions.) This setup was notably utilized by Goldman [37] to prove the ergodicity of mapping class group dynamics on the locus of SU(2)-local systems in $X_k(\mathbb{R})$. This suggests that the task of descent for a given point $\rho \in X_k(\mathbb{C})$ divides into two steps:

(1) ("action variables") Find a transformation $\gamma \in \Gamma$ such that $\mathrm{tr}_P(\gamma \cdot \rho) \in \mathbb{A}^n(\mathbb{C})$ lies in some predetermined compact set, and

(2) ("angle variables") Descend along fibers of tr_P using the Dehn twist actions.

Unraveling the definitions, we find that the boundedness of systoles for $\mathrm{SL}_2(\mathbb{C})$-local systems (Theorem 4.1) allows us to achieve the first step without difficulty. However, it turns out that the fibres of the morphisms tr_P as above can become complicated, and descent along the fibres by Dehn twist can fail. To avoid this issue, we shall instead give a proof of Theorem 5.1 using an inductive argument on (g, n); intuitively, our

argument takes care of the action/angle variables one at a time.

This section is organized as follows. In Section 5.1.1, we establish Theorem 5.1 for $(g, n) = (1, 1)$ and $(0, 4)$ using a classical elementary argument essentially due to Markoff. In Section 5.1.2, we collect some preliminary observations in preparation for the general case, including lifts of Dehn twist to representations considered in Golman and Xia [41]. In Section 5.1.3, we prove Theorem 5.1 in the general case.

5.1.1 Markoff descent

In this section, we establish Theorems 5.1 in the cases $(g, n) = (1, 1)$ and $(0, 4)$. Given $k = (k_1, \cdots, k_n) \in \mathbb{A}^n(\mathbb{C})$, we shall write $\text{Height}(k) = \max\{1, |k_1|, \cdots, |k_n|\}$.

Surfaces of type $(1, 1)$

Let Σ be a surface of type $(1, 1)$. By Section 2.3.1, we have an identification of the moduli space $X_k = X_k(\Sigma)$ with the affine cubic algebraic surface in $\mathbb{A}^3_{x,y,z}$ given by the equation

$$x^2 + y^2 + z^2 - xyz - 2 = k.$$

The mapping class group $\Gamma = \Gamma(\Sigma)$ acts on X_k via polynomial transformations. Up to finite index, it coincides with the group Γ' of automorphisms of X_k generated by the transpositions and even sign changes of coordinates as well as the Vieta involutions of the form $(x, y, z) \mapsto (x, y, xy - z)$.

The mapping class group dynamics on $X_k(\mathbb{R})$ was analyzed in detail by Goldman [40], and the work of Ghosh-Sarnak [36] establishes a remarkable exact fundamental set for the action of Γ' on the integral points $X_k(\mathbb{Z})$ for admissible k. We emphasize that our focus is on establishing descent for the complex points; cf. Silverman [80]. Lemmas 5.7 and 5.8 below establish Theorems 4.1 and 5.1 for $(g, n) = (1, 1)$.

Lemma 5.7. *There is a constant $C > 0$ independent of k such that, given any*

$\rho \in X_k(\mathbb{C})$, *there exists some* $\gamma \in \Gamma$ *such that* $\gamma \cdot \rho = (x, y, z)$ *satisfies*

$$\min\{|x|, |y|, |z|\} \leq C \cdot \mathrm{Height}(k)^{1/3}.$$

Proof. We notice that it suffices to prove the lemma with Γ replaced by Γ'. Suppose first that $\rho = (x, y, z)$ satisfies

$$|x| \leq |y| \leq |z| \leq |xy - z|. \qquad (*)$$

If $|x| \leq 8$ then we are done, so assume otherwise. Note that we have

$$|z|^2 \leq |z(xy - z)| = |x^2 + y^2 - 2 - k| \leq 2|y|^2 + 2 + |k|.$$

Suppose first that $|xy| \leq 2|z|$. We then have

$$\frac{|xy|^2}{4} \leq |z|^2 \leq |x^2 + y^2 - 2 - k| \leq 2|y|^2 + 2 + |k|.$$

Dividing both sides by $|y|^2$, we obtain

$$\frac{|x|^2}{4} \leq 2 + \frac{2 + |k|}{|y|^2} \leq 2 + \frac{2 + |k|}{|x|^2} \leq 2 \max\left\{2, \frac{2 + |k|}{|x|^2}\right\}$$

and therefore $|x|^4 \leq 8(2 + |k|)$, as desired. Suppose next that $|xy| \geq 2|z|$. We have

$$\frac{|z||xy|}{2} \leq |z(xy - z)| \leq 2|y|^2 + 2 + |k|.$$

Dividing both sides by $|yz|$ and arguing as above, we obtain $|x|^3 \leq 4(2 + |k|)$, as desired. It therefore remains to consider the case where no $\gamma \cdot \rho$ with $\gamma \in \Gamma'$ satisfies the analogue of $(*)$. Assuming $|x| \leq |y| \leq |z|$ without loss of generality, we must have $|xy - z| \leq |y|$, since otherwise we may replace (x, y, z) by (x, y, z'), $z' = xy - z$,

satisfying an analogue of $(*)$. If $|x| \leq 12$ then we are done, so assume otherwise. Suppose first that $2|y| \geq |z|$. We then have

$$|xz^2|/2 \leq |xyz| = |x^2 + y^2 + z^2 - 2 - k| \leq 3|z|^2 + 2 + |k|.$$

Dividing both sides by $|z|^2$ and arguing as before, we find that $|x|^3 \leq 4(2 + |k|)$. Thus, assume that $2|y| \leq |z|$. We then replace (x, y, z) by $(x, xy - z, y)$ and repeat the above argument. Since

$$2\max\{|x|, |y|, |xy - z|\} = 2|y| \leq |z| = \max\{|x|, |y|, |z|\},$$

this cannot continue indefinitely without reaching our desired result. □

Lemma 5.8. *Let $K \subset \mathbb{C}$ be a bounded set. Let A be a set such that*

(1) *$A \subset \mathbb{R}$ and $\operatorname{dist}(A, \{\pm 2\}) > 0$, or*

(2) *$A \subset \mathbb{C}$ and $\operatorname{dist}(A, [-2, 2]) > 0$.*

There is a constant $C = C(K, A) > 0$ such that, for every $\rho \in X_K(A)$, there exists $\gamma \in \Gamma$ such that $\gamma \cdot \rho = (x, y, z)$ satisfies $\max\{|x|, |y|, |z|\} \leq C$.

Proof. We notice that it suffices to prove the lemma with Γ replaced by Γ'. Fix $\delta > 0$ such that

$$|\lambda + \lambda^{-1} - x| > \delta$$

for every $x \in A$ and $\lambda \in \mathbb{K}$ with $1 \leq |\lambda| \leq \sqrt{1 + \delta}$, where $\mathbb{K} = \mathbb{R}$ if A satisfies the first condition in the lemma and $\mathbb{K} = \mathbb{C}$ otherwise. Such a number δ exists by our assumption on A. Let $\rho = (x, y, z) \in X_k(A)$ be any point with $k \in K$. By Lemma 5.7, we may assume without loss of generality that

$$m(\rho) = |x| \leq C_1(1 + |k|)^{1/3}$$

for a positive constant C_1 independent of ρ and k. Let $\epsilon = |x^2 - 2 - k|$. Writing

$$c = \inf\{|y'| : (x, y', z') \sim (x, y, z)\}$$

where $\sim$ indicates equivalence under Γ', we may further assume that

$$\min\{|y|, |z|\} \leq (c^2 + \epsilon)^{1/2}$$

and that $|y| \leq |z|$. We claim that $|y|^2 \leq \epsilon/\delta$. Indeed, suppose otherwise. We have the following three cases:

(1) We have $|xy - z| \leq |y| \leq |z|$. Replacing (x, y, z) by $(x, y', z') = (x, xy - z, y)$, we may thus assume $c \leq |y| \leq |z| \leq (c^2 + \epsilon)^{1/2}$ and thus

$$|z|^2 \leq |y|^2 + \epsilon \leq |y|^2(1 + \delta).$$

(2) We have $|y| \leq |xy - z| \leq |z|$. Replacing (x, y, z) by $(x, y', z') = (x, y, xy - z)$, we reduce to the third case below.

(3) We have $|y| \leq |z| \leq |xy - z|$. In particular, we have

$$|z|^2 \leq |z(xy - z)| = |x^2 + y^2 - 2 - k| \leq |y|^2 + \epsilon < |y|^2(1 + \delta).$$

Thus, in all cases, the ratio $\lambda = z/y$ satisfies $1 \leq |\lambda| \leq \sqrt{1 + \delta}$. Dividing the equation $y^2 + z^2 - xyz = 2 + k - x^2$ by yz, we have

$$\left|\lambda^{-1} + \lambda - x\right| = \frac{\epsilon}{|yz|} \leq \delta,$$

contradicting the definition of δ. Thus, we must have $|y| \leq \epsilon/\delta$. Finally, considering

the above three cases again, we have

$$\max\{|y|, |z|\} \leq \epsilon(1+\delta)/\delta,$$

up to replacing (x, y, z) by $(x, xy - z, y)$ in case $|xy - z| \leq |y| \leq |z|$. Combining this with the bound on $|x|$ in the beginning, we have proved our desired result. □

Surfaces of type $(0, 4)$

Let Σ be a surface of type $(0, 4)$. By Section 2.3.2, we have an identification of the moduli space $X_k = X_k(\Sigma)$ with the affine cubic algebraic surface in $\mathbb{A}^3_{x,y,z}$ given by the equation

$$x^2 + y^2 + z^2 + xyz = Ax + By + Cz + D$$

with A, B, C, D appropriately determined by k. The mapping class group $\Gamma = \Gamma(\Sigma)$ acts on X_k via polynomial transformations. Let Γ' be the group of automorphisms of X_k generated by the Vieta involutions

$$\tau_x : (x, y, z) \mapsto (A - yz - x, y, z),$$
$$\tau_y : (x, y, z) \mapsto (x, B - xz - y, z),$$
$$\tau_z : (x, y, z) \mapsto (x, y, C - xy - z).$$

Two points $\rho, \rho' \in X_k(\mathbb{C})$ are Γ'-equivalent if and only if they are Γ-equivalent or ρ is Γ-equivalent to all of $\tau_x \cdot \rho'$, $\tau_y \cdot \rho'$, and $\tau_z \cdot \rho'$. The following lemmas prove Theorems 4.1 and 5.1 for $(g, n) = (0, 4)$.

Lemma 5.9. *There is a constant $C > 0$ independent of k such that, given any $\rho \in X_k(\mathbb{C})$, there exists some $\gamma \in \Gamma$ such that $\gamma \cdot \rho = (x, y, z)$ satisfies one of:*

(1) $\min\{|x|, |y|, |z|\} \leq C$,

(2) $|yz| \leq C\,\mathrm{Height}(a)$,

(3) $|xz| \leq C\,\mathrm{Height}(b)$,

(4) $|xy| \leq C\,\mathrm{Height}(c)$, *or*

(5) $|xyz| \leq C\,\mathrm{Height}(d)$.

Proof. We proceed as in the proof of Lemma 5.7. Suppose first $\rho = (x, y, z)$ satisfies

$$|x| \leq |y| \leq |z| \leq |c - xy - z|. \qquad (*)$$

Note that we have

$$|z|^2 \leq |z(c - xy - z)| = |x^2 + y^2 - ax - by - d| \leq 4\max\{2|y|^2, |ax|, |by|, |d|\}.$$

Assume first that $|xy| \leq 2|z|$. We then have

$$\frac{|xyz|}{2} \leq |z|^2 \leq 4\max\{2|y|^2, |ax|, |by|, |d|\}$$

which gives us the desired result. Assume next that $|xy| \geq 2|z|$. If now $|xy| \leq 3|c|$, then we are done, we may assume $|xy| \geq 3|c|$. We then have

$$\frac{|xyz|}{6} \leq |z(c - xy - z)| \leq 4\max\{2|y|^2, |ax|, |by|, |d|\}$$

as desired. If $\rho = (x, y, z)$ satisfies an analogue of $(*)$ with a suitable change in the roles of the variables x, y, z, then the same argument applies. It therefore remains to consider the case where no $\gamma \cdot \rho$ with $\gamma \in \Gamma'$ satisfies an analogue of $(*)$. Let us assume that $|x| \leq |y| \leq |z|$; the other cases will follow similarly. We must have

$$|c - xy - z| \leq |y|,$$

since otherwise we may replace (x, y, z) by (x, y, z'), $z' = c - xy - z$, satisfying an analogue of $(*)$. Now, if $|x| \leq 8$ then we are done, so assume otherwise. Suppose first that $2|y| \geq |z|$. We then have

$$\begin{aligned} |xz^2|/2 \leq |xyz| &= |x^2 + y^2 + z^2 - ax - by - cz - d| \\ &\leq 5 \max\{3|z|^2, |ax|, |by|, |cz|, |d|\} \end{aligned}$$

which gives us the desired result. Thus, assume that $2|y| \leq |z|$. We then replace (x, y, z) by $(x, y, c - xy - z)$ and repeat the above argument. Since

$$2 \max\{|x|, |y|, |c - xy - z|\} = 2|y| \leq |z| = \max\{|x|, |y|, |z|\},$$

this cannot continue indefinitely without reaching our desired result. □

Lemma 5.10. *Let $K \subset \mathbb{C}^n$ be a compact set. Let A be a set such that:*

(1) *$A \subset \mathbb{R}$ and $\operatorname{dist}(A, \{\pm 2\}) > 0$, or*

(2) *$A \subset \mathbb{C}$ and $\operatorname{dist}(A, [-2, 2]) > 0$.*

There is a constant $C = C(K, A) > 0$ such that, for every $\rho \in X_K(A)$, there exists $\gamma \in \Gamma$ such that $\gamma \cdot \rho = (x, y, z)$ satisfies $\max\{|x|, |y|, |z|\} \leq C$.

Proof. We notice that it suffices to prove the lemma with Γ replaced by Γ'. We argue as in the proof of Lemma 5.8. Fix $\delta > 0$ such that

$$|\lambda + \lambda^{-1} - (-x)| > \delta$$

for every $x \in A$ and $\lambda \in \mathbb{K}$ with $1 \leq |\lambda| \leq \sqrt{1 + \delta}$, where $\mathbb{K} = \mathbb{R}$ if A satisfies the first condition in the lemma and $\mathbb{K} = \mathbb{C}$ otherwise. Such a number δ exists by our assumption on A. Let $\rho = (x, y, z) \in X_k(A)$ be any point with $k \in K$. By Lemma

5.9, we may assume without loss of generality that

$$\min\{|x|, |y|, |z|\} \leq C_1 \,\mathrm{Height}(k)^{4/3}$$

for a positive constant C_1 independent of ρ and k. Assume $|x| = \min\{|x|, |y|, |z|\}$; the other cases will follow similarly. Let $\epsilon = |x^2 - ax - d|$. Writing

$$c = \inf\{|y'|, |z'| : (x, y', z') \sim (x, y, z)\}$$

where $\sim$ indicates equivalence under Γ', we may further assume that

$$\min\{|y|, |z|\} \leq (c^2 + \epsilon)^{1/2}.$$

We assume $|y| \leq |z|$; the other cases will follow similarly. We shall first bound the size of $|y|$. If $|y| \leq 4|b|/\delta$ or $|y| \leq 4|c|/\delta$ or $|y|^2 \leq 2\epsilon/\delta$, then we are done, so suppose otherwise. We have the following three cases:

(1) We have $|c - xy - z| \leq |y| \leq |z|$. Replacing (x, y, z) by

$$(x, y, z') = (x, y, c - xy - z)$$

and switching the roles of y and z, we may assume $c \leq |y| \leq |z| \leq (c^2 + \epsilon)^{1/2}$ and thus

$$|z|^2 \leq |y|^2 + \epsilon \leq |y|^2(1 + \delta).$$

(2) We have $|y| \leq |c - xy - z| \leq |z|$. Replacing (x, y, z) by

$$(x, y', z') = (x, y, c - xy - z)$$

we reduce to the third case below.

(3) We have $|y| \leq |z| \leq |c - xy - z|$. In particular, we have

$$|z|^2 \leq |z(c - xy - z)| = |x^2 + y^2 - ax - by - d| < |y|^2(1 + \delta).$$

Thus, in all cases, the ratio $\lambda = z/y$ satisfies $1 \leq |\lambda| \leq \sqrt{1 + \delta}$. Dividing the equation $y^2 + z^2 + xyz = ax + by + cz + d - x^2$ by yz, we have

$$\left|\lambda^{-1} + \lambda + x\right| = \frac{|by| + |cz| + \epsilon}{|yz|} \leq \delta,$$

contradicting the definition of δ. Thus, we must have $|y| \leq 4\max\{|b|, |c|\}/\delta$ or $|y| \leq 2\epsilon/\delta$. Finally, considering the above three cases again, we also obtain a bound on $|z|$ in terms of $|x|$ and Height(k). Combining this with the bound on $|x|$ in the beginning, we have proved our desired result. □

5.1.2 Preliminary observations

Reconstruction

Let Σ be a surface of type (g, n) satisfying $3g + n - 3 > 0$, and let $a \subset \Sigma$ be an essential curve. Let $x \in \Sigma$ be a base point lying on a, and let α be a simple based loop parametrizing a. We shall summarize the reconstruction of a representation $\rho : \pi_1(\Sigma, x) \to \mathrm{SL}_2(\mathbb{C})$ from representations on connected components of $\Sigma|a$, as well as associated lifts of Dehn twists. Our main reference is Goldman-Xia [41]. There are two cases to consider, according to whether a is separating or nonseparating.

Nonseparating curves. Suppose that a is nonseparating, so $\Sigma|a$ is connected. Let a_1 and a_2 be the boundary curves of $\Sigma|a$ corresponding to a, and let (x_i, α_i) be the lifts of (x, α) to each a_i. We shall assume that we have chosen the numberings so that the interior of $\Sigma|a$ lies to the left as one travels along α_1. Let β be a simple loop on Σ based at x, intersecting the curve a once transversely at the base point,

such that β lifts to a path β' in $\Sigma|a$ from x_2 to x_1. Let us denote by α_2' the loop based at x_1 given by the path $\alpha_2' = (\beta')^{-1}\alpha_2\beta'$, where $(\beta')^{-1}$ refers to the path β' traversed in the opposite direction. The immersion $\Sigma|a \to \Sigma$ induces an embedding $\pi_1(\Sigma|a, x_1) \to \pi_1(\Sigma, x)$, giving us the isomorphism

$$\pi_1(\Sigma, x) = (\pi_1(\Sigma|a, x_1) \vee \langle\beta\rangle)/(\alpha_2' = \beta^{-1}\alpha_1\beta).$$

Thus, any representation $\rho : \pi_1(\Sigma, x) \to \mathrm{SL}_2(\mathbb{C})$ is determined uniquely by a pair (ρ', B), where $\rho' : \pi_1(\Sigma|a, x_1) \to \mathrm{SL}_2(\mathbb{C})$ is a representation and $B \in \mathrm{SL}_2(\mathbb{C})$ is an element such that $\rho'(\alpha_2') = B^{-1}\rho'(\alpha_1)B$, with the correspondence

$$\rho \mapsto (\rho', B) = (\rho|_{\pi_1(\Sigma|a,x_1)}, \rho(\beta)).$$

We define an automorphism τ_α of $\mathrm{Hom}(\pi_1(\Sigma, x), \mathrm{SL}_2)$ as follows. Given $\rho = (\rho_a, B)$, we set $\tau_\alpha(\rho_a, B) = (\rho_a, B')$ where $B' = \rho(\alpha)B$. This descends to the action τ_a of the left Dehn twist action along a on the moduli space $X(\Sigma)$.

Separating curves. Suppose that a is separating, so we have $\Sigma|a = \Sigma_1 \sqcup \Sigma_2$ with each Σ_i of type (g_i, n_i) satisfying $2g_i + n_i - 2 > 0$. Let a_i be the boundary curve of Σ_i corresponding to a. Let (x_i, α_i) be the lift of (x, α) to a_i. We shall assume that we have chosen the numberings so that the interior of Σ_1 lies to the left as one travels along α_1. The immersions $\Sigma_i \hookrightarrow \Sigma$ of the surfaces induce embeddings $\pi_1(\Sigma_i, x_i) \to \pi_1(\Sigma, x)$ of fundamental groups, and we have an isomorphism

$$\pi_1(\Sigma, x) \simeq (\pi_1(\Sigma_1, x_1) \vee \pi_1(\Sigma_2, x_2))/(\alpha_1 = \alpha_2).$$

Thus, any representation $\rho : \pi_1(\Sigma, x) \to \mathrm{SL}_2(\mathbb{C})$ is determined uniquely by a pair (ρ_1, ρ_2) of representations $\rho_i : \pi_1(\Sigma_i, x_i) \to \mathrm{SL}_2(\mathbb{C})$ such that $\rho_1(\alpha_1) = \rho_2(\alpha_2)$, with

the correspondence

$$\rho \mapsto (\rho_1, \rho_2) = (\rho|_{\pi_1(\Sigma_1, x_1)}, \rho|_{\pi_1(\Sigma_2, x_2)}).$$

We define an automorphism τ_α of $\mathrm{Hom}(\pi_1(\Sigma, x), \mathrm{SL}_2)$ as follows. For a representation $\rho = (\rho_1, \rho_2)$, we set $\tau_\alpha(\rho_1, \rho_2) = (\rho_1, \rho_2')$ where

$$\rho_2'(\gamma) = \rho(\alpha)\rho_2(\gamma)\rho(\alpha)^{-1}$$

for every $\gamma \in \pi_1(\Sigma_2, x_2)$. This descends to the action τ_a of the left Dehn twist along a on the moduli space $X(\Sigma)$.

Preparatory lemmas

We collect a number of elementary lemmas needed for our proof of Theorem 5.1. Let us say that a subset of a complex analytic variety V is *bounded* if its closure in V is compact. For example, a subset $U \subset \mathrm{SL}_2(\mathbb{C})$ is bounded if and only if the matrix entries of elements in U are uniformly bounded.

Lemma 5.11. *Let Σ be a surface of type (g, n) where $2g + n - 2 > 0$ and $n \geq 1$. Let $(\alpha_1, \cdots, \alpha_{2g+n})$ be an optimal sequence of generators for $\pi_1\Sigma$. Consider the set $K = K_1 \times K_2 \subset \mathbb{C}^n$ where*

(1) *$K_1 \subset \mathbb{C}^{n-1}$ is a bounded subset, and*

(2) *$K_2 \subset \mathbb{C}$ is a bounded subset with $\mathrm{dist}(K_2, \{\pm 2\}) > 0$.*

For each subset $L \subset X_K(\Sigma, \mathbb{C})$ which is bounded as a subset of $X(\Sigma, \mathbb{C})$, there exists a bounded set $M \subset \mathrm{Hom}(\pi_1\Sigma, \mathrm{SL}_2(\mathbb{C}))$ with $L = \pi(M)$ and

$$\rho(\alpha_{2g+n}) = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{bmatrix}$$

for some $\lambda \in \{z \in \mathbb{C} : |z| > 1\} \cup \{e^{\pi i t} : 0 < t < 1\}$ for every $\rho \in M$.

Proof. Let $\rho : \pi_1\Sigma \to \mathrm{SL}_2(\mathbb{C})$ be a representation whose class in $X(\mathbb{C})$ lies in L. This implies in particular that the traces $\operatorname{tr}\rho(\alpha_i)$, $\operatorname{tr}\rho(\alpha_i\alpha_j)$, and $\operatorname{tr}\rho(\alpha_i\alpha_j\alpha_k)$ are bounded by a constant depending only on L. Up to global conjugation, we have

$$\rho(\alpha_{2g+n}) = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{bmatrix}$$

with $\lambda \in \{z \in \mathbb{C} : |z| > 1\} \cup \{e^{\pi i t} : 0 < t < 1\}$, and there is a bounded set $\Lambda \subset \mathbb{C}$ with $\operatorname{dist}(\Lambda, \{\pm 1\}) > 0$, depending only on K_2, such that $\lambda \in \Lambda$. Let us write

$$\rho(\alpha_i) = \begin{bmatrix} a_i & b_i \\ c_i & d_i \end{bmatrix}$$

for $i = 1, \cdots, 2g+n-1$. Note that we have

$$\begin{bmatrix} a_i \\ d_i \end{bmatrix} = \frac{1}{\lambda^{-1} - \lambda} \begin{bmatrix} \lambda^{-1} & -1 \\ -\lambda & 1 \end{bmatrix} \begin{bmatrix} a_i + d_i \\ \lambda a_i + \lambda^{-1} d_i \end{bmatrix}.$$

Since $(\operatorname{tr}\rho(\alpha_i), \operatorname{tr}\rho(\alpha_i\alpha_{2g+n})) = (a_i + d_i, \lambda a_i + \lambda^{-1} d_i)$ is bounded in terms of L, the above expression shows that (a_i, d_i) lies in a bounded subset of $\mathbb{C}^2$ depending only on Λ and L. Now, suppose that ρ is irreducible, so that there exists at least one α_i such that $c_i \neq 0$. Up to global conjugation of ρ by a diagonal matrix in $\mathrm{SL}_2(\mathbb{C})$, noting that

$$\begin{bmatrix} \mu & 0 \\ 0 & \mu^{-1} \end{bmatrix} \begin{bmatrix} a_i & b_i \\ c_i & d_i \end{bmatrix} \begin{bmatrix} \mu^{-1} & 0 \\ 0 & \mu \end{bmatrix} = \begin{bmatrix} a_i & \mu^2 b_i \\ \mu^{-2} c_i & d_i \end{bmatrix}$$

we can assume that $\max\{|c_i| : 1 \leq i \leq 2g+n-1\} = 1$ and that $c_{i_0} = 1$ for some $1 \leq i_0 \leq 2g+n-1$. The equation $a_{i_0} d_{i_0} - b_{i_0} c_{i_0} = 1$ shows that b_{i_0} is bounded in

terms of Λ and L. Next, for $i \neq i_0$, we have

$$\operatorname{tr} \rho(\alpha_{i_0}\alpha_i) = a_{i_0}a_i + d_{i_0}d_i + b_{i_0}c_i + b_i c_{i_0}.$$

Since the left hand side as well as the first three terms on the right hand side are bounded in terms of Λ and L, it follows that so is b_i. This shows that ρ lies in a bounded subset of $\operatorname{Hom}(\pi_1\Sigma, \mathrm{SL}_2(\mathbb{C}))$. Lastly, if ρ is reducible, then up to replacing ρ by a diagonal representation it lies a bounded subset of $\operatorname{Hom}(\pi_1\Sigma, \mathrm{SL}_2(\mathbb{C}))$. We conclude that a bounded set M satisfying the conclusions of the lemma exists. □

Lemma 5.12. *Suppose K and M are sets given as follows:*

(1) *$K \subset \mathbb{C}$ is a bounded set with $\operatorname{dist}(K, \{\pm 2\}) > 0$.*

(2) *$M \subset \mathrm{SL}_2(\mathbb{C})$ is a bounded set.*

There is a bounded set $N \subset \mathrm{SL}_2(\mathbb{C})$ such that, given any pair (A_1, A_2) of conjugate matrices in $\mathrm{SL}_2(\mathbb{C})$ with $\operatorname{tr} A_1 \in K$ and $A_2 \in M$ and moreover

$$A_1 = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{bmatrix}, \quad \lambda \in \{z \in \mathbb{C} : |z| > 1\} \cup \{e^{\pi i t} : 0 < t < 1\},$$

there exists $B \in N$ with $A_2 = B^{-1}A_1B$.

Proof. Let A_1, A_2 be as in the hypothesis of the lemma. Given any matrix

$$B = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}_2(\mathbb{C}),$$

we have

$$B^{-1}A_1B = \begin{bmatrix} \lambda + bc(\lambda - \lambda^{-1}) & bd(\lambda - \lambda^{-1}) \\ -ac(\lambda - \lambda^{-1}) & \lambda^{-1} - bc(\lambda - \lambda^{-1}) \end{bmatrix}.$$

This shows that, if $A_2 = B^{-1}A_1B$, then the products bc, bd, and ac are all in a bounded subset of $\mathbb{C}$ depending only on K and M_1, and similarly for ad since $ad = bc + 1$. If $b = c = 0$ or $a = d = 0$, then A_2 is diagonal and we may take

$$B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},$$

respectively. So let us assume otherwise. Since B is invertible, note that we must have $ab \neq 0$ or $cd \neq 0$. Let us suppose that $ab \neq 0$; the case where $cd \neq 0$ will be similar. Multiplying B on the left by a diagonal matrix in $\mathrm{SL}_2(\mathbb{C})$, noting that

$$\begin{bmatrix} \mu & 0 \\ 0 & \mu^{-1} \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a\mu & b\mu \\ c\mu^{-1} & d\mu^{-1} \end{bmatrix},$$

we may assume that $\max\{|a|, |b|\} = 1$. In particular, c and d are bounded in terms of K and M_1 by the boundedness of the products ac, ad, bc, and bd remarked above. Hence, B may be taken to be in a bounded subset of $\mathrm{SL}_2(\mathbb{C})$ depending only on K and M_1, as desired. □

Lemma 5.13. *If $A, B, C \in \mathrm{SL}_2(\mathbb{C}) \setminus \{\pm \mathbf{1}\}$ are matrices such that each of the pairs (A, B), (A, C), and (A, BC) has an eigenvector in common in $\mathbb{C}^2$, then the triple (A, B, C) has a common eigenvector.*

Proof. Assume the contrary. If A is conjugate to a matrix of the form

$$\pm \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix},$$

then the lemma is clear. Thus, up to global conjugation we may assume without loss of generality that A is diagonal, B is strictly upper triangular (i.e. it is upper triangular and not diagonal), and C is strictly lower triangular. But then BC is

neither upper nor lower triangular, contradicting the assumption that (A, BC) has a common eigenvector. $\square$

5.1.3 General case

Let Σ be a surface of type (g, n) with $3g + n - 3 > 0$. Let $K \subset \mathbb{C}^n$ be a compact subset, and let A be a set satisfying one of the following:

(1) $A \subset \mathbb{R}$ and $\operatorname{dist}(A, \{\pm 2\}) > 0$, or

(2) $A \subset \mathbb{C}$ and $\operatorname{dist}(A, [-2, 2]) > 0$.

We shall prove Theorem 5.1 by induction on (g, n). We first state a lemma.

Lemma 5.14. *There is a constant $C = C(A, K)$ such that, for any irreducible local system $\rho \in X_K(A)$, there is an essential curve $a \subset \Sigma$ with $|\operatorname{tr}_a(\rho)| \leq C$ such that one of the following holds.*

(1) *We have $g \geq 1$, the curve a is nonseparating, and $\rho|(\Sigma|a)$ is irreducible.*

(2) *We have $g = 0$, the curve a is separating with $\Sigma|a = \Sigma_1 \sqcup \Sigma_2$, where $\rho|\Sigma_1$ is irreducible and Σ_2 is of type $(0, 3)$.*

Proof. Assume first that $g \geq 1$. Applying Theorem 4.1 a bounded number of times (with the bound only depending on (g, n)), there is a nonseparating essential curve a in Σ such that $|\operatorname{tr} \rho(a)| \leq C_1$ for some constant C_1 depending only on K and A. We are done if $\rho|(\Sigma|a)$ is irreducible, so let us assume otherwise. Let Σ' be an embedded one holed torus in Σ that contains a. Let c denote the boundary curve of Σ'. Since $\operatorname{tr} \rho(c) \neq 2$ by our hypothesis that $\pm 2 \notin A$ and that ρ is irreducible, we see $\rho|\Sigma'$ is irreducible. Writing $\operatorname{tr} \rho(a) = \lambda + \lambda^{-1}$ with $\lambda \in \mathbb{C}$ and $|\lambda| \geq 1$, we have

$$|\operatorname{tr} \rho(c)| = |\lambda^2 + \lambda^{-2}| \leq C_2(K, A).$$

Using Lemma 5.8 and the notation therein, we may assume that the restriction $\rho|\Sigma' = (x, y, z) \in X(\Sigma')$ satisfies

$$\max\{|x|, |y|, |z|\} \leq C_3(K, A).$$

Let $(\alpha_1, \alpha_2, \gamma)$ be an optimal sequence of generators for $\pi_1\Sigma'$, so that the essential curves $a = a_1$, a_2, and a_3 (lying in the free homotopy classes of α_1, α_2, and $\alpha_1\alpha_2$, respectively), correspond to the variables x, y, z. It suffices to show that $\rho|(\Sigma'|a_i)$ is irreducible for some i, since then $\rho|(\Sigma|a_i)$ is irreducible. The desired result follows by Lemma 5.13, taking our triple of matrices to be $(\rho(\alpha_1), \rho(\alpha_2), \rho(\gamma))$.

It remains to consider the case where $g = 0$. Applying Theorem 5.2 and Lemma 5.10 a bounded number of times (with the bound only depending on n), there is an essential curve $b \subset \Sigma$ with $\Sigma|b = \Sigma' \sqcup \Sigma''$, where Σ' is a four holed sphere and Σ'' is a $n - 2$ holed sphere, such that we have $\rho|\Sigma' = (x, y, z) \in X(\Sigma')$ satisfying

$$\max\{|x|, |y|, |z|\} \leq C_1(K, A).$$

Let $a_1, a_2, a_3 \subset \Sigma'$ be the essential curves corresponding to the variables x, y, z. If $\rho|\Sigma''$ is irreducible, then we may just take a to be one of the curves a_i and we are done, so assume otherwise. The desired result follows from similar considerations of Lemma 5.13 as before. □

Theorem 5.1. *There exists a bounded set $L \subseteq X_K(\mathbb{C})$ such that $X_K(A) \subseteq \Gamma \cdot L$.*

Proof. We shall first assume $\operatorname{dist}(A, [-2, 2]) > 0$, and later return to the case $A \subset \mathbb{R}$ with $\operatorname{dist}(A, \{\pm 2\}) > 0$ at the end of the proof. Let $\pi : \operatorname{Hom}(\pi_1\Sigma, \mathrm{SL}_2) \to X(\Sigma)$ be the usual projection. We shall prove the theorem by induction on (g, n).

We begin by treating the inductive step for the increase of genus g, to illustrate our approach. Suppose that Σ has genus $g \geq 1$, and the theorem has been proved

for all eligible surfaces of genus $g-1$. Let $a \subset \Sigma$ be a nonseparating curve. In our arguments below, we shall use the notations in the first part of Section 5.1.2. By Lemma 5.14.(1), there is a constant $C = C(K, A)$ such that, for any $\rho \in X_K(\Sigma, A)$ which is irreducible, there exists $\gamma \in \Gamma(\Sigma)$ such that

(1) $|\operatorname{tr}_a(\gamma \cdot \rho)| \leq C$ and

(2) $(\gamma \cdot \rho)|(\Sigma|a)$ is irreducible.

We remark that the irreducibility of $\rho \in X_K(\Sigma, A)$ is automatic unless we have $(g, n) = (1, 1)$, since we have $2 \notin A$ by assumption. Note that the theorem has already been proven for $(g, n) = (1, 1)$ in Lemma 5.8, so may restrict our attention to $(g, n) \neq (1, 1)$ below.

Let us set $K' = \{(k_1, \cdots, k_n, a_1, a_2) : (k_1, \cdots, k_n) \in K, a_i \in A, |a_i| \leq C\}$. By the inductive hypothesis, there is a bounded set $L' \subseteq X_{K'}(\Sigma|a, \mathbb{C})$ such that

$$X_{K'}(\Sigma|a, A) \subseteq \Gamma(\Sigma|a) \cdot L'.$$

By Lemma 5.11, there is a bounded set $M' \subseteq \pi^{-1}(X_{K'}(\Sigma', \mathbb{C}))$ satisfying $L' \subseteq \pi(M')$ and moreover every $\rho' \in M'$ satisfies

$$\rho'(\alpha_1) = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{bmatrix}$$

for some $\lambda \in \mathbb{C}^*$ with $|\lambda| > 1$. (Here, we use the notation in the first part of Section 5.1.2 and α_1 is a loop parametrizing one of the boundary curves of $\Sigma|a$ corresponding to a.) Using the description of the lift of Dehn twist in Section 5.1.2 and arguing as in the proof of Lemma 5.12, we conclude that there is a bounded set $N \subset \mathrm{SL}_2(\mathbb{C})$ depending only on K, A, and $C = C(K, A)$ such that every

$$\rho = (\rho', B) \in \pi^{-1}(X_K(\Sigma, A))$$

with $\rho' \in M'$ is $\langle \tau_\alpha \rangle$-equivalent to a point ρ_0 corresponding to $(\rho_0', B_0') \in M' \times N$. It follows that there is a bounded subset $M \subset \pi^{-1}(X_K(\Sigma, \mathbb{C}))$ such that every $\rho \in X_K(\Sigma, A)$ satisfying $\rho|(\Sigma|a) \in \pi(M')$ is $\langle \tau_a \rangle$-equivalent to a point in $\pi(M)$. But every $\rho \in X_K(\Sigma, A)$ is $\Gamma(\Sigma)$-equivalent to such a point by our hypothesis on M', and hence

$$X_K(\Sigma, A) \subseteq \Gamma(\Sigma) \cdot \pi(M).$$

Setting $L = \pi(M)$, this completes the inductive step in genus. It remains to treat the cases $g = 0$ with $n \geq 4$, which we establish by induction on n similar to the previous argument, with the following modifications:

- There are more than one (but at most finitely many) mapping class group equivalence classes of separating curves $a \subset \Sigma$, considered up to isotopy, such that $\Sigma|a = \Sigma_1 \sqcup \Sigma_2$ with Σ_2 of type $(0, 3)$.
- In appropriate places of the argument, we follow the lines of Section 5.1.2, Lemma 5.14.(2), and the proof of Lemma 5.11 instead of Section 5.1.1, Lemma 5.14.(1), and the proof of Lemma 5.12.

Note that the base case $(g, n) = (0, 4)$ was established by Lemma 5.10. This completes the induction, proving the theorem for $\operatorname{dist}(A, [-2, 2]) > 0$.

Finally, it remains to consider the case where $A \subset \mathbb{R}$ with $\operatorname{dist}(A, \{\pm 2\}) > 0$. We proceed by induction as before, with minor modifications. For instance, suppose that $g = 0$ and $a \subset \Sigma$ is a separating curve with $\Sigma|a = \Sigma_1 \sqcup \Sigma_2$ where Σ_2 is of type $(0, 3)$, and suppose that $\rho \in X_K(\Sigma, A)$ satisfies $\operatorname{tr} \rho(a) \in (-2, 2)$. Up to mapping class group action and induction, we may assume $\rho|\Sigma_1$ lies in a bounded set that only depends on K. Let $\alpha_1, \cdots, \alpha_n$ be an optimal generating sequence for $\pi_1\Sigma$ such that a lies in the free homotopy class of $\alpha_1\alpha_2$. Using Fact 2.1, the coordinate ring of $X(\Sigma)$ is generated by $\operatorname{tr}_{\alpha_{i_1} \cdots \alpha_{i_k}}$ where $i_1 < \cdots < i_k$ are integers in $\{2, \cdots, n\}$ and $k \leq 3$. The values of these functions at ρ are all bounded in terms of K and A, except possibly for those

of the form

$$\mathrm{tr}_{\alpha_2\alpha_r} \quad \text{and} \quad \mathrm{tr}_{\alpha_2\alpha_s\alpha_t}$$

for $3 \leq r \leq n$ and $3 \leq s < t \leq n$. By considering the restrictions of ρ to a finite number of subsurfaces of type $(0,4)$ in Σ containing Σ_2, the task of bounding the values of above functions is reduced to the following observation. Consider the moduli space $X_k(\Sigma)$ for Σ a surface of type $(0,4)$, with equation

$$x^2 + y^2 + z^2 + xyz = Ax + By + Cz + D$$

using the notation of Section 2.3.2, and suppose that $\rho = (x_0, y_0, z_0) \in X_k(\Sigma, \mathbb{R})$ is given with $z_0 \in (-2, 2)$. From the description of the action of Dehn twists given in Section 2.3.2, we see that we must have $A, B, C, D \in \mathbb{R}$ and thus the above equation with fixed $z = z_0$ determines a (possibly degenerate) ellipse in the (x, y)-plane. Moreover, the maximum possible size of (x, y) on this conic section is bounded in terms of A, B, C, D by elementary geometry. Applying this observation in our setting, we conclude that the values of the trace functions $\mathrm{tr}_{\alpha_2\alpha_r}$ and $\mathrm{tr}_{\alpha_2\alpha_s\alpha_t}$ are bounded in terms of the other functions in the collection $\{\mathrm{tr}_{\alpha_{i_1}\cdots\alpha_{i_k}}\}$ considered above, thus allowing us to continue the induction. The case $g \geq 1$ is also similar. □

5.2 Parabolic subvarieties

Let Σ be a surface of type (g, n) satisfying $3g + n - 3 > 0$, and let $X_k = X_k(\Sigma)$ for $k \in X(\partial\Sigma, \mathbb{C})$ be a relative character variety of Σ. Given a pants decomposition P of Σ, the immersion $P \to \Sigma$ induces a morphism

$$\mathrm{tr}_P : X_k \to X(P) \simeq \mathbb{A}^{3g+n-3}$$

whose fibers for $t \in X(P, \mathbb{C})$ will be denoted $X^P_{k,t} = \mathrm{tr}_P^{-1}(t)$. Let

$$(-)|_{\Sigma|P} : X_k \to X(\Sigma|P) = X(\Sigma_1) \times \cdots \times X(\Sigma_{2g+n-2})$$

be the morphism induced by the immersion $\Sigma|P \to \Sigma$, where the product on the right hand side is taken over the connected components Σ_i of $\Sigma|P$. It is clear that $(-)|_{\Sigma|P}$ is constant along each fiber $X^P_{k,t}$. We make the following definition.

Definition 5.15. Let (P, t) be as above. The fiber $X^P_{k,t}$ is *perfect* if

(1) $\mathrm{tr}_a(X^P_{k,t}) \neq \pm 2$ for every curve $a \subseteq P$, and

(2) $(g, n, k) = (1, 1, 2)$ or each factor of $(X^P_{k,t})|_{\Sigma|P}$ is irreducible.

The above definition is motivated by the following theorem, which is the main result of this section. Recall from Chapter 1 that an algebraic variety V over $\mathbb{C}$ is said to be *parabolic* if every closed point of V lies in the image of a nonconstant morphism $\mathbb{A}^1 \to V$. In the following, a pair (P, t) will denote a pants decomposition P of Σ and an element $t \in X(P, \mathbb{C})$.

Theorem 5.16. *We have the following.*

A. *For each (P, t), the fiber $X^P_{k,t}$ is either perfect or parabolic.*

B. *For any nonconstant morphism $f : \mathbb{A}^1 \to X_k$, there is a parabolic fiber $X^P_{k,t}$ for some (P, t) containing the image of f.*

The remainder of this section is organized as follows. In Section 5.2.1, we give a proof of Theorem 5.16 in the cases $(g, n) = (1, 1)$ and $(0, 4)$. The moduli space X_k in these cases is an explicitly defined algebraic surface, making the proof easier. We prove the general case of Theorem 5.16 in Section 5.2.2. Finally, in Section 5.2.3, we describe the conseuqences of Theorem 5.16, including Theorems 5.2 and 5.3 as well as a rigidity result (Corollary 5.18) for polynomial deformations of Fuchsian surface groups.

5.2.1 Base cases

In this subsection, we give a proof of Theorem 5.16 in the cases $(g, n) = (1, 1)$ and $(0, 4)$. We shall refer the reader to Section 2.3 for explicit presentations of X_k in these cases. First, it is useful to record an elementary lemma.

Lemma 5.17. *If Σ is a surface of type $(0, 3)$, then $k = (k_1, k_2, k_3) \in X(\Sigma) \simeq \mathbb{A}^3$ is reducible if and only if $k_1^2 + k_2^2 + k_3^2 - k_1k_2k_3 - 2 = 2$.*

Proof. This is elementary. For instance, it follows from Lemma 6.3 and the derivation in Section 2.3.1 of the trace of commutators $\mathrm{tr}([A, B])$ with $A, B \in \mathrm{SL}_2(\mathbb{C})$. □

Surfaces of type $(1, 1)$

Suppose Σ is of type $(1, 1)$, and let (α, β, γ) be optimal generators of $\pi_1\Sigma$. The moduli space X_k can be presented as an affine cubic algebraic surface, with equation

$$x^2 + y^2 + z^2 - xyz - 2 = k$$

where the variables (x, y, z) correspond to the monodromy traces $(\mathrm{tr}_\alpha, \mathrm{tr}_\beta, \mathrm{tr}_{\alpha\beta})$ and $\mathrm{tr}_\gamma \equiv k$. Now, let $a \subset \Sigma$ be the essential curve underlying α. Since any essential curve of Σ is, up to isotopy, mapping class group equivalent to a, it suffices to prove Theorem 5.16.A for $(g, n) = (1, 1)$ where $P = a$. Let $t \in X(a, \mathbb{C}) \simeq \mathbb{C}$. Since $X^a_{k,t}$ is a conic section

$$y^2 - tyz + z^2 + t^2 - 2 - k = 0$$

in the (y, z)-plane, elementary geometry shows that $X^a_{k,t}$ is parabolic if and only if (1) $t = \pm 2$, or (2) the conic section is degenerate. Let us assume $t \neq \pm 2$; the latter condition states that the equation for $X^a_{k,t}$ factors as

$$y^2 - tyz + z^2 + t^2 - 2 - k = (y - \lambda z + m_1)(y - \lambda^{-1}z + m_2) = 0$$

for some $\lambda \in \mathbb{C}^*$ and $m_i \in \mathbb{C}$. Expanding and comparing coefficients, we must have

- $\lambda + \lambda^{-1} = t$,
- we have
$$\begin{bmatrix} 1 & 1 \\ -\lambda & -\lambda^{-1} \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
and hence $m_1 = m_2 = 0$, and therefore
- $m_1 m_2 = t^2 - 2 - k = 0$, and in particular $k \neq 2$.

Thus, we see that $X^a_{k,t}$ is parabolic if and only if

(1) $t = \pm 2$, or

(2) $t \neq \pm 2$, $(g, n, k) \neq (1, 1, 2)$, and $(X^a_{k,t})|_{\Sigma|a} = (t, t, t^2 - 2)$.

Here, we made an identification $X(\Sigma|a) \simeq \mathbb{A}^3$. By Lemma 5.17, the last condition in (2) is equivalent to saying that $(X^a_{k,t})|_{\Sigma|a}$ is reducible. This concludes the proof of Theorem 5.16.A for $(g, n) = (1, 1)$. Combining this with Lemma 4.17, we immediately obtain Theorem 5.16.B in this case.

Surfaces of type $(0, 4)$

Suppose Σ is of type $(0, 4)$, and let $(\gamma_1, \gamma_2, \gamma_3, \gamma_4)$ be optimal generators of $\pi_1\Sigma$. The moduli space X_k is an affine cubic algebraic surface with equation

$$x^2 + y^2 + z^2 + xyz - Ax - By - Cz - D = 0$$

where the variables (x, y, z) correspond to the traces $(\mathrm{tr}_{\gamma_1\gamma_2}, \mathrm{tr}_{\gamma_2\gamma_3}, \mathrm{tr}_{\gamma_1\gamma_3})$ and

$$\begin{cases} A = k_1k_2 + k_3k_4 \\ B = k_2k_3 + k_1k_4 \\ C = k_1k_3 + k_2k_4 \end{cases} \quad \text{and} \quad D = 4 - k_1^2 - k_2^2 - k_3^2 - k_4^2 - k_1k_2k_3k_4$$

with $k_i \equiv \operatorname{tr}\gamma_i$. Now, let $a, b, c \subset \Sigma$ be essential curves lying in the free homotopy classes of $\gamma_1\gamma_2$, $\gamma_2\gamma_3$, and $\gamma_1\gamma_3$, respectively. Since any essential curve of Σ is, up to isotopy, mapping class group equivalent to one of the curves a, b, and c, it suffices to prove Theorem 5.16.A for $(g, n) = (0, 4)$ where P consists of one of the curves a, b, and c. We treat the case $P = a$ in what follows; the remaining cases will proceed similarly. Let $t \in X(a, \mathbb{C}) \simeq \mathbb{C}$. Again by elementary geometry, we see that $X^a_{k,t}$ is parabolic if and only if (1) $t = \pm 2$, or (2) $X^a_{k,t}$ is a degenerate conic in the (y, z)-plane. Let us assume $t \neq \pm 2$; the latter condition states that the equation for $X^a_{k,t}$ factors as

$$t^2 + y^2 + z^2 + tyz - At - By - Cz - D = (y + \lambda z + m_1)(y + \lambda^{-1}z + m_2) = 0$$

for some $\lambda \in \mathbb{C}^*$ and $m_i \in \mathbb{C}$. Expanding and comparing coefficients, we see that

- $\lambda + \lambda^{-1} = t$,
- we have
$$\begin{bmatrix} 1 & 1 \\ \lambda & \lambda^{-1} \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} = \begin{bmatrix} -B \\ -C \end{bmatrix},$$
and
- $m_1 m_2 = -At - D$.

This is equivalent to

$$-\frac{B^2 - tBC + C^2}{t^2 - 4} = \frac{1}{(\lambda^{-1} - \lambda)^2}(\lambda^{-1}B - C)(-\lambda B + C) = -At - D$$

or in other words $(t^2 - 4)(-Ax - D) + (B^2 - tBC + C^2) = 0$. Upon rearranging, this is seen to be equivalent to

$$(k_1^2 + k_2^2 + x^2 - tk_1k_2 - 4)(k_3^2 + k_4^2 + x^2 - tk_3k_4 - 4) = 0,$$

which in turn is equivalent to saying that at least one factor of $(X^a_{k,t})|_{\Sigma|a}$ is reducible, by Lemma 5.17. This proves Theorem 5.16.A for $(g,n) = (0,4)$, and Theorem 5.16.B follows by Lemma 4.17.

5.2.2 General case

Let Σ be a surface of type (g,n) with $3g+n-3 > 0$, and let $X_k = X_k(\Sigma)$ be a relative character variety of Σ. We shall first prove the following claim, by induction on (g,n). The claim will imply that a perfect fiber cannot be parabolic, proving one half of Theorem 5.16.A (as well as Theorem 5.16.B, conditional on Theorem 5.16.A).

Claim. *For any nonconstant morphism $f : \mathbb{A}^1 \to X_k$, there exists an imperfect fiber $X^P_{k,t}$ for some (P,t) containing the image of f.*

Proof. We have already proved Theorem 5.16 for $(g,n) = (1,1)$ and $(0,4)$, and these will provide the base cases for our induction. Let $f : \mathbb{A}^1 \to X_k$ be a nonconstant morphism. By Lemma 4.17, there is an essential curve $a \subset \Sigma$ such that $\mathrm{tr}_a \circ f \equiv t_0$ is constant. Suppose first that the restriction

$$f|_{\Sigma'} : \mathbb{A}^1 \to X_k(\Sigma) \xrightarrow{(-)|_{\Sigma'}} X(\Sigma')$$

of f to some connected component Σ' of $\Sigma|a$ is nonconstant. We observe that the image of $f|_{\Sigma'}$ lies in $X_{k'}(\Sigma')$ for some $k' \in X(\partial\Sigma', \mathbb{C})$ determined by k and t_0. Hence, by inductive hypothesis, there is a pants decomposition P' of Σ' and an element $t' \in X(P', \mathbb{C})$ such that $X^{P'}_{k',t'}(\Sigma')$ is an imperfect fiber containing the image of $f|_{\Sigma'}$. If a is nonseparating, then taking

$$(P,t) = (P' \sqcup a, (t', t_0))$$

we see that $X^P_{k,t}$ is an imperfect fiber containing the image of f, as desired. If f

is separating and $\Sigma|a = \Sigma' \sqcup \Sigma''$, then again $f|_{\Sigma''}$ has image lying in some $X_{k''}(\Sigma'')$ for some $k'' \in X(\partial\Sigma'', \mathbb{C})$ determined by k and t_0, and by a repeated application of Lemma 4.17 we see that there is a pants decomposition P'' of Σ'' and $t'' \in X(P'', \mathbb{C})$ such that $X^{P''}_{k'',t''}$ contains the image of $f|_{\Sigma''}$. Taking

$$(P, t) = (P' \sqcup a \sqcup P'', (t', t_0, t'')),$$

we again see that $X^P_{k,t}$ is an imperfect fiber containing the image of f.

To complete our proof of the claim, it remains only to consider the case where $f|_{\Sigma|a}$ is constant. Let $\alpha_1, \cdots, \alpha_{2g+n}$ be a sequence of optimal generators for $\pi_1\Sigma$. Below, we give a proof in the case where a is nonseparating; the case where a is separating will proceed in a similar manner, invoking the case of Theorem 5.16 for $(g, n) = (0, 4)$ appropriately.

Let us thus suppose that a is nonseparating. Up to mapping class group action, we may assume that a is the essential curve underlying α_1. By Fact 2.1, the coordinate ring of the fiber $X_k \to X(\Sigma|a)$ above $f|_{\Sigma|a}$ is generated by functions of the form

$$\mathrm{tr}_{\alpha_2}, \mathrm{tr}_{\alpha_2\alpha_i}, \mathrm{tr}_{\alpha_2\alpha_i\alpha_j}, \mathrm{tr}_{\alpha_1\alpha_2}, \mathrm{tr}_{\alpha_1\alpha_2\alpha_i}, \mathrm{tr}_{\alpha_1\alpha_2\alpha_i\alpha_j}$$

for $3 \leq i < j \leq 2g + n$. Thus, since f is nonconstant, the composition of f with at least one of the above coordinate functions must be nonconstant. Let us consider the case where $\mathrm{tr}_{\alpha_2} \circ f$ is nonconstant; the other cases will follow similarly. There is a surface $\Sigma' \subset \Sigma$ of type $(1, 1)$ containing the loops α_1 and α_2 in its interior. By our hypothesis, we see that $f|_{\Sigma'}$ is nonconstant, and the image of $f|_{\Sigma'}$ lies in $X_{k'}(\Sigma)$ for some $k' \in X(\partial\Sigma', \mathbb{C})$ determined by the (constant) value of $f|_{\Sigma|a}$; indeed, the boundary of Σ' lies in $\Sigma|a$. Thus, by the case of Theorem 5.16 for $(g, n) = (1, 1)$ proved above, there is an essential curve $a' \subset \Sigma'$ and $t' \in X(a', \mathbb{C})$ such that $X^{a'}_{k',t'}(\Sigma')$ is an imperfect fiber containing the image of $f|_{\Sigma'}$. Completing $a' \sqcup \partial\Sigma' \subset \Sigma$ to a pants

decomposition P of Σ, we see that there is some $t \in X(P, \mathbb{C})$ such that $X^P_{k,t}$ is an imperfect fiber containing the image of f, as desired. $\square$

We shall henceforth assume that $(g, n) \neq (1, 1), (0, 4)$, to simplify our remaining argument. To complete the proof of Theorem 5.16, it suffices to prove the following.

Claim. *Given a representation $\rho : \pi_1\Sigma \to \mathrm{SL}_2(\mathbb{C})$ with Jordan equivalence class lying in an imperfect fiber $X^P_{k,t}$, there is a one-parameter polynomial family*

$$\rho_T : \pi_1\Sigma \to \mathrm{SL}_2(\mathbb{C})$$

of representations with nonconstant Jordan equivalence classes all lying in $X^P_{k,t}$, so that we have $\rho = \rho_{T_0}$ for some $T_0 \in \mathbb{C}$.

Proof. Let ρ and $X^P_{k,t}$ be as above. Let us first consider the case where there is a curve $a \subseteq P$ with $\mathrm{tr}_a(X^P_{k,t}) = \pm 2$. We may assume to have fixed the base point of Σ to lie on a. Let α bs a smooth simple loop parametrizing a. Up to global conjugation, we may assume that

$$\rho(\alpha) = s \begin{bmatrix} 1 & u \\ 0 & 1 \end{bmatrix} \tag{$*$}$$

for $s \in \{\pm 1\}$ and $u \in \{0, 1\}$. There are several elementary cases to consider.

Suppose first that a is separating, and write $\Sigma | a = \Sigma_1 \sqcup \Sigma_2$. Up to conjugation, we may assume that on top of $(*)$ the following conditions hold:

(1) $\rho|\Sigma_1$ is irreducible or upper triangular, and

(2) $\rho|\Sigma_2$ is irreducible or lower triangular.

If $\rho_i = \rho|\Sigma_i$ is upper or lower triangular, then it is part of a 1-parameter family of upper or lower triangular representations ρ^t_i varying the off-diagonal entries. For

instance, if Σ_1 is a surface of type $(0,3)$, and $(\alpha_1, \alpha_2, \alpha_3)$ are a sequence of optimal generators for $\pi_1\Sigma_1$ with α_1 corresponding to α, and if

$$\rho_1(\alpha_2) = \begin{bmatrix} \lambda & x \\ 0 & \lambda^{-1} \end{bmatrix},$$

then ρ_1^t can be taken to be the representation determined by $\rho_1^t(\alpha_1) = \rho_1(\alpha_1)$ and

$$\rho_1^t(\alpha_2) = \begin{bmatrix} \lambda & x+t \\ 0 & \lambda^{-1} \end{bmatrix}.$$

We thus see that, if ρ_1 or ρ_2 is reducible, then at least one of the 1-parameter families (ρ_1^t, ρ_2^t), (ρ_1^t, ρ_2), and (ρ_1, ρ_2^t) (once defined) descends to a nontrivial family in $X_{k,t}^P$, giving us a nonconstant morphism $\mathbb{A}^1 \to X_{k,t}^P$ with the class of ρ in the image. It therefore remains only to consider the case where ρ_1 and ρ_2 are both irreducible. We consider the representation $\rho^t = (\rho_1, u_t \rho_2 u_t^{-1})$ for $t \in \mathbb{C}$ where

$$u_t = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}.$$

It is easy to see that the morphism $\mathbb{A}^1 \to X(\Sigma)$ given by $t \mapsto \rho_t$ is nonconstant, contains ρ in its image, and moreover lands in $X_{k,t}^P$, as desired.

Suppose next that a is nonseparating. Let us write $\rho = (\rho|(\Sigma|a), \rho(\beta)) = (\rho', B)$ using the notation of Section 5.1.2. If ρ is irreducible, then let us consider the representation

$$\rho^t = (\rho', B_t)$$

where

$$B_t = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} B.$$

It is easy to see that the morphism $\mathbb{A}^1 \to X^P_{k,t}$ defined by $t \mapsto \rho^t$ is nonconstant and $\rho^0 = \rho$, as desired. So suppose that ρ is reducible. Up to global conjugation, we may assume that ρ is diagonal. Since Σ is not of type $(1,1)$, there is a separating curve c with $\Sigma|c = \Sigma_1 \sqcup \Sigma_2$ where Σ_1 is a surface of type $(1,1)$ containing the curve b underlying β. Let us write $\rho|\Sigma_1 = ((\rho|\Sigma_1)', B)$ and consider the upper triangular deformation $(\rho|\Sigma_1)^t = ((\rho|\Sigma_1)', B_t)$ as above. This and a suitable lower triangular deformations of $\rho|\Sigma_2$ combine together to give a nonconstant morphism $\mathbb{A}^1 \to X^P_{k,t}$ whose image contains ρ, as desired.

It therefore remains to consider the case where the following conditions hold:

(1) $\mathrm{tr}_a(X^P_{k,t}) \neq \pm 2$ for every curve $a \subseteq P$,

(2) $(g,n,k) \neq (1,1,2)$, and

(3) $(X^P_{k,t})|_{\Sigma'}$ is reducible for some connected component Σ' of $\Sigma|P$.

We may assume for convenience that the base point $x \in \Sigma$ lies on Σ'. Let $\gamma_1, \gamma_2, \gamma_3$ be optimal generators for $\pi_1\Sigma'$ corresponding to the boundary curves c_1, c_2, c_3 of Σ'. We shall write $\rho|_{\Sigma'} = (t_1, t_2, t_3) \in X(\Sigma') \simeq \mathbb{A}^3$. By relabeling the boundary curves of Σ' if necessary, we may assume that c_1 corresponds to a curve of P. In particular, $t_1 \neq \pm 2$ by our hypothesis above. We further assume that, if the image of Σ' in Σ is a surface of type $(1,1)$, then c_1 and c_2 map to the same curve in P.

It will be convenient for us to introduce distinguished families of representations $\pi_1\Sigma' \to \mathrm{SL}_2(\mathbb{C})$ which are reducible. For each $T \in \mathbb{C}$, let $\rho^u_T, \rho^l_T : \pi_1\Sigma' \to \mathrm{SL}_2(\mathbb{C})$ be the representation determined by

$$\rho^u_T(\gamma_1) = \rho^l_T(\gamma_1) = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{bmatrix}, \quad \rho^u_T(\gamma_2) = \begin{bmatrix} \mu & T \\ 0 & \mu^{-1} \end{bmatrix}, \quad \rho^l_T(\gamma_2) = \begin{bmatrix} \mu & 0 \\ T & \mu^{-1} \end{bmatrix}$$

where $\lambda \in \{z \in \mathbb{C}^\times : \Im(z) \geq 0, z \notin [-1,1]\}$ and $\mu \in \mathbb{C}^\times$ are such that

$$\begin{cases} t_1 = \lambda + \lambda^{-1}, \\ t_2 = \mu + \mu^{-1}, \\ t_3 = \lambda\mu + \lambda^{-1}\mu^{-1}. \end{cases}$$

Note that the Jordan equivalence class of each ρ_T^u and ρ_T^l in $X(\Sigma)$ is equal to that of $\rho|_{\Sigma'}$. Up to global conjugation, we may assume that $\rho|_{\Sigma'} = \rho_{T_0}^u$ or $\rho_{T_0}^l$ for some $T_0 \in \mathbb{C}$. Below, we proceed under the assumption that $\rho|_{\Sigma'} = \rho_{T_0}^u$; the other case will proceed in a similar manner. We shall write also $\rho_T^u = \rho_T'$ for easier notation.

We must consider different cases, according to the relative position of Σ' in Σ. Suppose first that the image of Σ' in Σ is a surface of type $(1,1)$. By our hypothesis, the boundary curves c_1 and c_2 map to the same curve $a \subset P$, while c_3 maps to a separating curve $c \subset \Sigma$. Let us write $\Sigma|c = \Sigma'' \sqcup \Sigma'''$ with Σ'' being the image of Σ'. Without loss of generality, we may assume that the implicit base point $x \in \Sigma$ lies on c. Let (α, β, γ) be a sequence of otpimal generators for $\pi_1\Sigma''$, such that under the immersion $\Sigma' \to \Sigma''$ we have

$$\gamma_1 \mapsto \alpha, \quad \gamma_2 \mapsto \beta^{-1}\alpha^{-1}\beta, \quad \gamma_3 \mapsto \gamma.$$

Let us write

$$\rho(\beta) = \begin{bmatrix} B_1 & B_2 \\ B_3 & B_4 \end{bmatrix}.$$

The condition $\rho(\gamma_2) = \rho(\beta^{-1}\alpha^{-1}\beta)$ is then

$$\begin{aligned}\begin{bmatrix}\lambda & T_0\\ 0 & \lambda^{-1}\end{bmatrix} &= \begin{bmatrix}B_4 & -B_2\\ -B_3 & B_1\end{bmatrix}\begin{bmatrix}\lambda^{-1} & 0\\ 0 & \lambda\end{bmatrix}\begin{bmatrix}B_1 & B_2\\ B_3 & B_4\end{bmatrix}\\ &= \begin{bmatrix}\lambda + (\lambda^{-1}-\lambda)B_1B_4 & (\lambda^{-1}-\lambda)B_2B_4\\ (\lambda-\lambda^{-1})B_1B_3 & \lambda^{-1} + (\lambda-\lambda^{-1})B_1B_4\end{bmatrix}\end{aligned}$$

where we have $\lambda - \lambda^{-1} \neq 0$ by our hypothesis on ρ. This shows that we must have

$$\rho(\beta) = \begin{bmatrix}0 & B_2\\ -B_2^{-1} & \operatorname{tr}\rho(\beta)\end{bmatrix}$$

where furthermore $(\lambda^{-1} - \lambda)B_2 \operatorname{tr}\rho(\beta) = T_0$. Up to global conjugation of ρ by a diagonal matrix (which also results in a suitable adjustment of T_0), we may further assume that $B_2 = 1$, and hence $\operatorname{tr}\rho(\beta) = T_0/(\lambda^{-1} - \lambda)$. Let $\rho''_T : \pi_1\Sigma'' \to \mathrm{SL}_2(\mathbb{C})$ be the representation determined by

$$\rho''_T(\alpha) = \begin{bmatrix}\lambda & 0\\ 0 & \lambda^{-1}\end{bmatrix}, \quad \rho''_T(\beta) = \begin{bmatrix}0 & 1\\ -1 & \frac{T}{\lambda^{-1}-\lambda}\end{bmatrix}.$$

The preceding above observations show that $\rho''_{T_0} = \rho|_{\Sigma''}$ and $\rho''_T|_{\Sigma'} = \rho'_T$ for every $T \in \mathbb{C}$. Note that we have

$$\rho''_T(\gamma) = \rho'_T(\gamma_3) = \begin{bmatrix}\lambda^{-2} & -\lambda T\\ 0 & \lambda^2\end{bmatrix}.$$

For $T \in \mathbb{C}$, let $\rho_T : \pi_1\Sigma \to \mathrm{SL}_2(\mathbb{C})$ be the representation such that $\rho_T|\Sigma''' = \rho$ and

$$\rho_T|\Sigma'' = \begin{bmatrix}1 & \lambda\frac{T-T_0}{\lambda^2-\lambda^{-2}}\\ 0 & 1\end{bmatrix}\rho''_T\begin{bmatrix}1 & \lambda\frac{T-T_0}{\lambda^2-\lambda^{-2}}\\ 0 & 1\end{bmatrix}^{-1}.$$

Here, $\lambda^2 - \lambda^{-2} \neq 0$ since otherwise $\operatorname{tr}\rho(\gamma) = \pm 2$, which was precluded. It can be directly verified that

$$\rho(\gamma) = \begin{bmatrix} 1 & \lambda\frac{T-T_0}{\lambda^2-\lambda^{-2}} \\ 0 & 1 \end{bmatrix} \rho''_T(\gamma) \begin{bmatrix} 1 & \lambda\frac{T-T_0}{\lambda^2-\lambda^{-2}} \\ 0 & 1 \end{bmatrix}^{-1}$$

so ρ_T is well-defined by our discussion in Section 5.1.2. We have $\rho = \rho_{T_0}$ by construction, and we see that the morphism $\mathbb{A}^1 \to X_k$ given by $T \mapsto \rho_T$ lies in the fiber $X^P_{k,t}$. The fact that this morphism is nonconstant can be deduced from the observation that

$$\operatorname{tr}\rho_T(\beta) = \frac{T}{\lambda^{-1} - \lambda}$$

is a nonconstant function of T. Therefore, this proves the claim when the image of Σ' in Σ is a surface of type $(1, 1)$.

It remains to consider the case where the the boundary curves of Σ' map to three distinct curves in Σ under the immersion $\Sigma' \to \Sigma$, which we shall also denote c_1, c_2, c_3 for simplicity. Now, let us write

$$\begin{cases} \Sigma|c_3 &= \Sigma_3 \sqcup \Sigma_3^\circ \\ \Sigma_3|c_2 &= \Sigma_2 \sqcup \Sigma_2^\circ \\ \Sigma_2|c_1 &= \Sigma_1 \sqcup \Sigma_1^\circ \end{cases}$$

where Σ_3 is the connected component of $\Sigma|c_3$ containing Σ' (here, Σ_3° is empty if c_3 is nonseparating or is a boundary curve in Σ), Σ_2 is the connected component of $\Sigma_3|c_2$ containing Σ', and finally $\Sigma_1 = \Sigma'$.

Let us extend the polynomial family of representations $\rho'_T : \pi_1\Sigma' \to \mathrm{SL}_2(\mathbb{C})$ to the polynomial family $\rho''_T : \pi_1\Sigma_2 \to \mathrm{SL}_2(\mathbb{C})$ given by

$$\rho''_T|_{\Sigma'} = \rho'_T, \quad \rho''_T|_{\Sigma_1^\circ} = \rho|_{\Sigma_1^\circ}.$$

This is well-defined by our discussion in Section 5.1.2. Note that $\rho''_{T_0} = \rho|_{\Sigma_2}$. Our next step is to extend ρ''_T to a family $\rho'''_T : \pi_1\Sigma_3 \to \mathrm{SL}_2(\mathbb{C})$ such that $\rho'''_{T_0} = \rho|\Sigma_3$. We need to consider three cases.

(1) Suppose c_2 is a boundary curve on Σ_3, so that $\Sigma_2 = \Sigma_3$. Let $\rho'''_T = \rho''_T$.

(2) Suppose c_2 is a separating curve on Σ_3. We define ρ'''_T by requiring

$$\rho'''_T|_{\Sigma_2} = \rho''_T, \quad \rho'''_T|_{\Sigma_2^\circ} = \begin{bmatrix} 1 & \frac{T-T_0}{\mu^{-1}-\mu} \\ 0 & 1 \end{bmatrix} \rho|_{\Sigma_2^\circ} \begin{bmatrix} 1 & \frac{T-T_0}{\mu^{-1}-\mu} \\ 0 & 1 \end{bmatrix}^{-1}.$$

Note that we must have $\mu^{-1} - \mu \neq 0$ since otherwise $\operatorname{tr}\rho(c_2) = \pm 2$, which was precluded. By construction, we have

$$\rho''_T(\gamma_2) = \begin{bmatrix} 1 & \frac{T-T_0}{\mu^{-1}-\mu} \\ 0 & 1 \end{bmatrix} \rho(\gamma_2) \begin{bmatrix} 1 & \frac{T-T_0}{\mu^{-1}-\mu} \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} \mu & T \\ 0 & \mu^{-1} \end{bmatrix},$$

so that ρ'''_T is well defined by our discussion in Section 2.3.2.

(3) Suppose c_2 is a nonseparating curve on Σ_3. Let δ be a simple based loop on Σ_3 intersecting c_2 exactly once transversally, oriented as in the first part of Section 5.1.2 (with the pair (γ_2, δ) playing the role of (α, β) there). Let γ'_2 be the based loop on Σ_2 (like α'_2 in Section 5.1.2) whose image in Σ_3 lies in the homotopy class of $\delta^{-1}\gamma_2\delta$. We define the representation ρ'''_T by specifying the pair $(\rho'''_T|_{\Sigma_2}, \rho'''_T(\delta))$ as in the discussion of Section 5.1.2, where

$$\rho'''_T|_{\Sigma_2} = \rho''_T, \quad \rho'''_T(\delta) = \begin{bmatrix} 1 & \frac{T-T_0}{\mu^{-1}-\mu} \\ 0 & 1 \end{bmatrix} \rho(\delta).$$

We have $\rho'''_T(\delta^{-1})\rho'''_T(\gamma_2)\rho'''_T(\delta) = \rho(\delta^{-1})\rho(\gamma_2)\rho(\delta) = \rho(\gamma'_2) = \rho'''_T(\gamma'_2)$ by our construction, so that ρ'''_T is well defined.

Finally, by the same procedure, we define $\rho_T : \pi_1\Sigma \to \mathrm{SL}_2(\mathbb{C})$ extending ρ'''_T such that $\rho_{T_0} = \rho$. Moreover, the resulting morphism $\mathbb{A}^1 \to X_k$ given by $T \mapsto \rho_T$ must be nonconstant, unless $\rho = \rho_{T_0}$ is reducible.

It therefore remains to treat the case where ρ is reducible. This is elementary; for the benefit of the reader, we provide a sketch of the argument when P contains a separating curve $a \subset \Sigma$. Writing $\Sigma|a = \Sigma_1 \sqcup \Sigma_2$, up to Jordan equivalence we may assume that $\rho|\Sigma_1$ is lower triangular and *not* diagonal, while $\rho|\Sigma_2$ is diagonal. We may then define ρ_T to be any polynomial family of representations such that (i) $\rho_T|_{\Sigma_1} = \rho|_{\Sigma_1}$ for all $T \in \mathbb{C}$, and (ii) $\rho_T|_{\Sigma_2}$ is upper triangular and *not* diagonal for $T \neq 0$, with $\rho_0|_{\Sigma_2} = \rho|_{\Sigma_2}$. The existence of such a polynomial family ρ_T, with a nonconstant morphism $\mathbb{A}^1 \to X_k$ given by $T \mapsto \rho_T$, is clear. □

5.2.3 Applications

We prove Theorem 5.2 as a corollary of Theorem 5.16.

Theorem 2. *There is a closed parabolic subvariety $V \subsetneq X_k$ such that every nonconstant morphism $\mathbb{A}^1_{\mathbb{C}} \to X_k$ is mapping class group equivalent to some $\mathbb{A}^1_{\mathbb{C}} \to V$.*

Proof. For each pants decomposition P of Σ, the condition that $\rho \in X_k(\mathbb{C})$ lies in some parabolic fiber $X^P_{k,t}$ is evidently nontrivial and algebraic by Theorem 5.16 and the definition of perfect fibers. In particular, the union of all parabolic fibers of the form $X^P_{k,t}$ for fixed P is a proper closed parabolic algebraic subvariety of X_k. Since there are at most finitely many isotopy classes of pants decompositions of Σ up to mapping class group action, the desired result follows. □

Combining Theorems 5.1 and 5.16, we also immediately obtain Theorem 5.3. Finally, we record the following consequence of Theorem 5.16. By the uniformization theorem, given a marked hyperbolic structure σ on Σ with geodesic boundary

curves there is a Fuchsian representation $\rho_\sigma : \pi_1\Sigma \to \mathrm{SL}_2(\mathbb{R})$ such that the quotient $\mathbb{H}^2/\rho_\sigma(\pi_1\Sigma)$ of the upper half plane contains (Σ, σ) as a Nielsen core.

Corollary 5.18. *Every one-parameter polynomial deformation $\rho_t : \pi_1\Sigma \to \mathrm{SL}_2(\mathbb{R})$ of a Fuchsian representation ρ_0 preserving the boundary traces is isotrivial.*

Proof. This is immediate from Theorem 5.16 and the observation that a Fuchsian representation ρ_0 does not lie in an imperfect fiber $X^P_{k,t}$ for any (P, t). □

5.3 Arithmetic hyperbolic surfaces

In this section, we give a brief discussion of the relationship between the Diophantine geometry of the moduli spaces X_k and arithmetic hyperbolic surfaces. In Section 5.3.1, we shall recall some basic results on arithmetic hyperbolic surfaces, including a characterization of arithmetic lattices in $\mathrm{PSL}_2(\mathbb{R})$ due to Takeuchi [86] and a result on their finitude (for bounded covolume) due to Borel [7] and Takeuchi [87]. In the case of cocompact lattices, the latter result is strictly stronger than what immediately follows from our analysis. In Section 5.3.2, we mention specific consequences for the integral points $X_k(\mathbb{Z})$ of the moduli space for suitable values of k, and discuss the classical Markoff example where $(g, n) = (1, 1)$.

5.3.1 Arithmetic lattices and their finitude

We begin by recalling the characterization of arithmetic lattices in $\mathrm{PSL}_2(\mathbb{R})$. A *lattice* $L \leq \mathrm{PSL}_2(\mathbb{R})$ is a discrete subgroup of finite covolume. Let $k \subset \mathbb{R}$ be a totally real number field of degree d over $\mathbb{Q}$. Let $i_1, \cdots, i_d : k \hookrightarrow \mathbb{R}$ be the embeddings of k such that $i_1 = \mathrm{id}$. Let Q be a quaternion algebra over k which is split at i_1 and ramified at all the other archimedean places of k. The splitting induces an embedding $\rho : Q \to M_2(\mathbb{R})$. Let $\mathcal{O}$ be an order in Q, and let $\mathcal{O}^1$ be the group of norm one elements

in $\mathcal{O}$. The image of $\rho(\mathcal{O}^1) \leq \mathrm{SL}_2(\mathbb{R})$ under the projection $\mathrm{SL}_2(\mathbb{R}) \to \mathrm{PSL}_2(\mathbb{R})$, which we shall denote $L(Q, \mathcal{O})$, is a lattice in $\mathrm{PSL}_2(\mathbb{R})$.

Definition 5.19. A lattice $L \leq \mathrm{PSL}_2(\mathbb{R})$ is *arithmetic* if it is commensurable with some $L(Q, \mathcal{O})$ as above, and is *derived from a quaternion algebra* if it is moreover of finite index in some $L(Q, \mathcal{O})$.

Takeuchi [86] gave the following characterization of arithmetic lattices in $\mathrm{PSL}_2(\mathbb{R})$. Given a subset $M \subset \mathrm{PSL}_2(\mathbb{R})$, we shall denote $\mathrm{tr}(M) = \{\pm \mathrm{tr}(a) : a \in M\}$, where $\mathrm{tr}(a)$ is the trace of a which is well-defined up to sign.

Theorem 5.20 (Takeuchi)**.** *A lattice $L \leq \mathrm{PSL}_2(\mathbb{R})$ is arithmetic (up to conjugation) if and only if it satisfies the conditions* (I) *and* (II_1) *below, and is moreover derived from a quaternion algebra if and only if it satisfies* (I) *and* (II_2)*.*

(I) *The* trace field $K_L = \mathbb{Q}(\mathrm{tr}(L)) \subset \mathbb{R}$ *has finite degree over* $\mathbb{Q}$*, and* $\mathrm{tr}(L)$ *is contained in the ring of integers of* K_L*.*

(II_1) *For any field embedding $i : K_L \to \mathbb{C}$ which is not the identity on the subfield $K_L^{(2)} \subseteq K_L$ generated by $\{\mathrm{tr}(\gamma)^2 : \gamma \in L\}$, the set $i(\mathrm{tr}(L))$ is bounded in $\mathbb{C}$.*

(II_2) *For any field embedding $i : K_L \to \mathbb{C}$ which is not the identity, the set $i(\mathrm{tr}(L))$ is bounded in $\mathbb{C}$.*

One can show that $K_L^{(2)} = K_{L^{(2)}}$ where $L^{(2)}$ is the finite index subgroup of L generated by $\{\gamma^2 : \gamma \in L\}$ (see [86]). It follows that a lattice $L \leq \mathrm{PSL}_2(\mathbb{R})$ satisfying condition (I) above is arithmetic if and only if $L^{(2)}$ is derived from a quaternion algebra. We note that an arithmetic lattice $L \leq \mathrm{PSL}_2(\mathbb{R})$ is non-cocompact if and only if $L^{(2)}$ is conjugate to a subgroup of $\mathrm{PSL}_2(\mathbb{Z})$, or equivalently $K_L^{(2)} = \mathbb{Q}$.

Quotients of the hyperbolic plane $\mathbb{H}^2$ by arithmetic lattices in $\mathrm{PSL}_2(\mathbb{R})$ are referred to as *arithmetic hyperbolic surfaces*. In general, the quotient $S_L = L\backslash\mathbb{H}^2$ of $\mathbb{H}^2$ by a lattice $L \leq \mathrm{PSL}_2(\mathbb{R})$ is a complete hyperbolic surface of finite area, having

finitely many cusps and orbifold singularities. By classical works of Fricke and Klein, the lattice L is generated by hyperbolic elements $\alpha_1, \beta_1, \cdots, \alpha_g, \beta_g$ and elliptic or parabolic elements $\gamma_1, \cdots, \gamma_n$ with relations

$$[\alpha_1, \beta_1] \cdots [\alpha_g, \beta_g] \gamma_1 \cdots \gamma_n = 1, \quad \gamma_i^{e_i} = 1 \quad (i = 1, \cdots, n) \tag{5.2}$$

where $2 \leq e_1 \leq \cdots \leq e_n \leq \infty$ (with $e_i = \infty$ understood as saying that γ_i is parabolic). The sequence $(g; e_1, \cdots, e_n)$ is referred to as the *signature* of the lattice L or the hyperbolic surface S_L, and we have

$$\frac{\mathrm{Vol}(L\backslash\mathbb{H}^2)}{2\pi} = 2g - 2 + \sum_{i=1}^{n} \left(1 - \frac{1}{e_i}\right).$$

Let Σ be a surface of genus g with n punctures, and let $X = X(\Sigma)$ be the moduli space of SL_2-local systems on Σ. A diffeomorphism of Σ onto the complement of orbifold points on a hyperbolic surface S of signature $(g; e_1, \cdots, e_n)$ gives rise to a conjugacy class of a representation $\pi_1\Sigma \to \mathrm{PSL}_2(\mathbb{R})$, whose image is a lattice $L \leq \mathrm{PSL}_2(\mathbb{R})$ with $S \simeq S_L$. Such a representation admits finitely many lifts to $\mathrm{SL}_2(\mathbb{R})$ up to conjugacy, and we shall denote by $T(g; e_1, \cdots, e_n) \subset X(\mathbb{R})$ the set of representations so obtained over all hyperbolic surfaces S of signature $(g; e_1, \cdots, e_n)$. Each $T(g; e_1, \cdots, e_n)$ is contained in a finite union of relative moduli spaces $X_k(\mathbb{R})$, where the possible traces

$$k \in \{2\cos(2\pi r) : r \in \mathbb{Q}\}^n \subseteq \mathbb{A}^n(\mathbb{R})$$

are determined by $(g; e_1, \cdots, e_n)$. Each connected component of $T(g; e_1, \cdots, e_n)$ is isomorphic to the Teichmüller space of marked hyperbolic (orbifold) structures on Σ with signature $(g; e_1, \cdots, e_n)$, on which the mapping class group descent is classical. Below, we show how our arguments may be used to derive a finiteness statement for arithmetic lattices in $\mathrm{PSL}_2(\mathbb{R})$, which is superseded by a stronger result of Borel [7]

and Takeuchi [87] discussed further below.

Proposition 5.21. *For $M \geq 0$ and $D \geq 0$, there are only finitely many isomorphism classes of lattices $L \leq \mathrm{PSL}_2(\mathbb{R})$ derived from a quaternion algebra having covolume $\mathrm{Vol}(L\backslash\mathbb{H}^2) \leq M$ and trace field degree $[K_L : \mathbb{Q}] \leq D$.*

Proof. First, given $D \geq 0$ and $N \geq 0$, there are at most finitely many algebraic integers of degree at most D all of whose conjugates have absolute value at most N. It follows that there is a closed discrete subset $A_D \subset \mathbb{R}$ such that we have $\mathrm{tr}(L) \subseteq A_D$ for every arithmetic lattice $L \leq \mathrm{PSL}_2(\mathbb{R})$ derived from a quaternion algebra with trace field degree $[K_L : \mathbb{Q}] \leq D$. A bound on the volume $\mathrm{Vol}(L\backslash\mathbb{H}^2)$ shows that there are at most finitely many possibilities $(g; e_1, \cdots, e_n)$ for the signature. For each choice of signature, the set $T(g; e_1, \cdots, e_n) \cap X(A)$ consists of at most finitely mapping class group orbits. The desired result follows. □

Corollary 5.22. *For $M \geq 0$ and $D \geq 0$, there are only finitely many isomorphism classes of arithmetic lattices $L \leq \mathrm{PSL}_2(\mathbb{R})$ of covolume $\mathrm{Vol}(L\backslash\mathbb{H}^2) \leq M$ which are*

(1) *non-cocompact, or*

(2) *cocompact with trace field degree $[K_L : \mathbb{Q}] \leq D$.*

Proof. Given an upper bound on the covolume $\mathrm{Vol}(L\backslash\mathbb{H}^2)$, there are only finitely many possibilities for the signature of L. The index $[L : L^{(2)}]$ depends only on the signature of L. The previous proposition shows that there are at most finitely choices of $L^{(2)}$ up to conjugation, recalling that $K_L^{(2)} = \mathbb{Q}$ if L is non-cocompact. As noted in [87], for each choice of $L^{(2)}$ the normalizer $N_{L^{(2)}}$ of $L^{(2)}$ in $\mathrm{PSL}_2(\mathbb{R})$ contains $L^{(2)}$ with finite index. Since $L \leq N_{L^{(2)}}$, there are at most finitely choices of L. □

An analogous result is obtained in [12, Theorem 4.3] in the context of arithmetic hyperbolic surface bundles. For cocompact lattices, the finiteness result of Borel [7] and Takeuchi [87] below removes the assumption on the trace field degree.

Theorem ([7],[87]). *The collection of volumes of arithmetic hyperbolic surfaces is closed and discrete in* $\mathbb{R}$ *with finite multiplicities.*

We remark that, given a totally real number field $k \subset \mathbb{R}$ with ring of integers $\mathcal{O}_k$, it is not necessarily true that a point of $T(g; e_1, \cdots, e_n) \cap X(\mathcal{O}_k)$ is associated to an arithmetic hyperbolic surface. This amounts to the observation that there are lattices in $\mathrm{PSL}_2(\mathbb{R})$ satisfying condition (I) of Theorem 5.20 but failing (II_1); examples are given in [87]. Thus, the above theorem does not imply the finitude of mapping class group orbits on $T(g; e_1, \cdots, e_n) \cap X(\mathcal{O}_k)$, except in the case $k = \mathbb{Q}$ which we shall discuss in further detail.

5.3.2 Relation to integral points

We now restrict our discussion to the integral points $T(g; e_1, \cdots, e_n) \cap X(\mathbb{Z})$. We shall say that the signature $(g; e_1, \cdots, e_n)$ is of *compact type* if each e_i is finite, and is of *noncompact type* otherwise. The following is an immediate consequence of our discussion in Section 5.3.1.

Proposition 5.23. *If* $(g; e_1, \cdots, e_n)$ *is of compact type,* $T(g; e_1, \cdots, e_n) \cap X(\mathbb{Z}) = \emptyset$.

In particular, for $k \in \mathbb{A}^n(\mathbb{Z})$ such that $X_k(\mathbb{R}) \cap T(g; e_1, \cdots, e_n) \neq \emptyset$ where $(g; e_1, \cdots, e_n)$ is of compact type, this provides a partial obstruction to existence of integral points on X_k that is not *a priori* obvious without the moduli interpretation.

If $(g; e_1, \cdots, e_n)$ is of noncompact type, then each point of $T(g; e_1, \cdots, e_n) \cap X(\mathbb{Z})$ corresponds to a representation $\pi_1\Sigma \to \mathrm{SL}_2(\mathbb{R})$ whose image in $\mathrm{PSL}_2(\mathbb{R})$ is conjugate to a finite index subgroup of $\mathrm{PSL}_2(\mathbb{Z})$. As mentioned in [61], on combining results of Newman [71] and Moser-Wyman [69] one finds that the number a_N of index N subgroups in $\mathrm{PSL}_2(\mathbb{Z})$ is asymptotically

$$a_N \sim (12\pi\sqrt{e})^{-1/2} \exp\left(\frac{N \log N}{6} - \frac{N}{6} + N^{1/2} + N^{1/3} + \frac{\log N}{2}\right).$$

In particular, it follows that, on the moduli spaces $X_k(\Sigma)$ as Σ varies in type (g,n) with $n \geq 1$ and $k \in \{0, \pm 1, \pm 2\}^n$ with at least one component of k equal to ± 2, the average number of nondegenerate integral orbits grows exponentially in $|\chi(\Sigma)|$. This is also not *a priori* obvious without the moduli interpretation of the varieties $X_k(\Sigma)$.

We close this section by discussing the classical Markoff example $(g,n) = (1,1)$. As discussed in Chapter 2, the moduli space $X_k(\Sigma)$ in this case is an affine cubic algebraic surface with equation

$$x^2 + y^2 + z^2 - xyz - 2 = k, \quad k \in \mathbb{Z}.$$

The geometry and mapping class group dynamics on the set of real points $X_k(\mathbb{R})$ of the moduli space is described in [40]. To discuss the point of contact between arithmetic hyperbolic surfaces and integral points, we shall consider $k \in \{-2, -1, 0, 1\}$. For each such k, Goldman [40] observed that $X_k(\mathbb{R})$ consists of five connected components, each of which is mapping class group invariant: one component is compact and consists of unitary representations $\pi_1\Sigma \to \mathrm{SU}(2)$ having boundary trace k, while the remaining four components comprise the locus $T(1,e)$ for a suitable signature $(1,e)$ depending on k. If $k \in \{0, \pm 1\}$, then $e < \infty$ and Proposition 5.23 implies that $T(1,e)$ contains no integral point. Thus, any point on $X_k(\mathbb{Z})$ for $k \in \{0, \pm 1\}$ arises from a unitary representation, and we indeed find

$$\begin{aligned} X_{-1}(\mathbb{Z}) &= \Gamma \cdot \{(1,0,0)\}, \\ X_0(\mathbb{Z}) &= \Gamma \cdot \{(1,1,0)\}, \\ X_1(\mathbb{Z}) &= \emptyset. \end{aligned}$$

On the other hand, for $k = -2$, by Markoff descent we find that

$$X_k(\mathbb{Z}) = \Gamma \cdot \{(0,0,0), (3,3,3), (-3,-3,3), (-3,3,-3), (3,-3,-3)\}.$$

The compact component of $X_k(\mathbb{R})$ in this case is just the singleton $\{(0,0,0)\}$. From this, we find that there exists exactly one arithmetic hyperbolic surface of signature $(1,\infty)$ derived from a quaternion algebra, namely, the quotient of $\mathbb{H}^2$ by the lattice

$$L = \left\langle \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \right\rangle \leq \mathrm{PSL}_2(\mathbb{Z})$$

which is also called the *modular torus* in the literature. This is found by noting that

$$(3,3,3) = \left(\mathrm{tr} \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}, \mathrm{tr} \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}, \mathrm{tr} \begin{bmatrix} 0 & 1 \\ -1 & 3 \end{bmatrix} \right).$$

We remark that Takeuchi [87] gave a full classification of arithmetic hyperbolic surfaces of signature $(1,e)$. In a related context, Bowditch-Maclachlan-Reid [12] classified arithmetic once-punctured torus bundles by considering imaginary quadratic integral points on $X_k(\Sigma)$ for $(g,n,k) = (1,1,-2)$.

5.4 Integral points on curves

5.4.1 Nonintegrable curves

Let Σ be a surface of type (g,n) with $3g+n-3 > 0$, and let $X_k = X_k(\Sigma)$ be a relative character variety of Σ. We shall prove Theorem 5.4 using Corollary 4.16. We first repeat our definition of integrable subvarieties of X_k given in the beginning of the chapter.

Definition 5.24. An algebraic subvariety $V \subseteq X_k$ is *integrable* if there is a pants decomposition P of Σ with tr_P constant along V. Otherwise, V is *nonintegrable*.

Theorem 5.4. *Let $A \subset \mathbb{C}$ be a closed discrete subset. If $C \subset X_k$ is a nonintegrable geometrically irreducible algebraic curve, then $C(A)$ is finite, with effective bounds on*

sizes of the coordinates.

Proof. Let F be the function field of C over $\mathbb{C}$. Let $\pi : C_0 \to C$ be the normalization of C, and let $\overline{C}_0$ be a smooth compactification of C_0. Let

$$\{p_1, \cdots, p_m\} = \overline{C}_0(\mathbb{C}) \setminus C_0(\mathbb{C})$$

be the points at infinity. For each p_i, we have a discrete valuation v_i on F given by order of vanishing at p_i. By Corollary 4.16 and our assumption on C, we deduce that there is an essential curve $a_i \subset \Sigma$ such that $v_i(\pi^*(\mathrm{tr}_{a_i})) \geq 0$, meaning in particular that $\pi^*(\mathrm{tr}_{a_i})$ is bounded on C_0 near the point p_i. By our hypothesis that C is nonintegrable, we may further assume that each tr_{a_i} is nonconstant on C. In particular, for each $i = 1, \cdots, m$, setting $z_i = \mathrm{tr}_{a_i} \pi(p_i) \in \mathbb{C}$ the set

$$C(A) \cap \{\rho \in C(\mathbb{C}) : \mathrm{tr}\, \rho(a_i) = z_i\}$$

is finite. It therefore remains to consider

$$C(A) \setminus \bigcup_{i=1}^{m} \{\rho \in C(\mathbb{C}) : \mathrm{tr}\, \rho(a_i) = z_i\}.$$

But by the boundedness of each $\pi^* \mathrm{tr}_{a_i}$ near p_i and the discreteness of A, the above set lies in a compact subset. Again by the discreteness of A, this shows that $C(A)$ is finite, as desired.

To prove the last assertion, note that the desired curves a_i can in principle be found effectively by simply enumerating and going through the list of all isotopy classes of essential curves on Σ, with Corollary 4.16 guaranteeing the termination of this procedure. Once these functions are found, the above proof leads to the desired effective bounds on the sizes of points in $C(A)$. □

Theorem 5.4 yields a broad generalization and strengthening of the following result

of Long and Reid [60]. Let M be a finite volume hyperbolic three-manifold with a single cusp. By the work fo Thruston, its character variety $X(M)$ has as an irreducible component $C(M)$ containing the faithful discrete representation of $\pi_1 M$ which is an algebraic curve. As in Section 1.2, let O_d denote the ring of integers of the imaginary quadratic field $\mathbb{Q}(\sqrt{-d})$ for each squarefree integer $d > 0$.

Theorem (Long and Reid [60])**.** *For M as above, the set $\bigcup_{d>0} C(M)(O_d)$ is finite.*

The inclusion of the cuspidal torus $T \to M$ induces a morphism from $C(M)$ into the so-called Cayley cubic algebraic surface $X(T)$, which is isomorphic to $X_2(\Sigma)$ where Σ is a surface of type $(1, 1)$. The crucial ingredient in the proof of the above [60, Lemma 3.4] is that the image of $C(M)$ in $X_2(\Sigma)$ is a nonintegrable curve. Given this, Theorem 5.4 readily recovers the above theorem on the integral points of $C(M)$, bypassing a somewhat involved arithmetical argument in [60].

5.4.2 Application of Baker's theory

In this interlude, we demonstrate a result on certain lattice points lying on algebraic curves in algebraic tori. We shall obtain it as a straightforward consequence of Baker's theory on linear forms in logarithms. Let $\mathbb{G}_m$ denote be the multiplicative group, and let $d \geq 1$ be an integer. Let $\overline{\mathbb{Q}} \subset \mathbb{C}$ be the field of algebraic numbers. Fix a sequence of real algebraic numbers

$$(z_1, \cdots, z_d) \in \mathbb{G}_m^d(\overline{\mathbb{Q}} \cap \mathbb{R})$$

with each $z_i \neq \pm 1$, and let $\Gamma \leq \mathbb{G}_m^d(\overline{\mathbb{Q}} \cap \mathbb{R})$ be the subgroup consisting of elements of the form $(z_1^{l_1}, \cdots, z_d^{l_d})$ with $l_i \in \mathbb{Z}$. Let $C \subset \mathbb{G}_m^d$ be an irreducible algebraic curve defined over $\overline{\mathbb{Q}}$, and fix $p \in \mathbb{G}_m^d(\overline{\mathbb{Q}})$. Our goal is to prove the following.

Proposition 5.25. *One of the following holds:*

(1) $C(\overline{\mathbb{Q}}) \cap (\Gamma \cdot p)$ *is finite, or*

(2) C *is a translate of an algebraic subtorus* $T \leq \mathbb{G}_m^d$ *defined over* $\overline{\mathbb{Q}}$ *invariant under some nontrivial* $z \in \Gamma$.

Moreover, $C(\overline{\mathbb{Q}}) \cap (\Gamma \cdot p)$ *can be effectively determined.*

Proof. We may assume $p = (1, \cdots, 1)$ up to translation. We proceed by induction on $d \geq 1$. The claim is obvious if $d = 1$, so we may assume that $d \geq 2$. Assuming the result in the case $d = 2$, we shall first show below how the general case follows.

Suppose that $C(\overline{\mathbb{Q}}) \cap \Gamma$ is infinite. Up to rearranging the factors of $\mathbb{G}_m^d$, we may assume that the projection morphism $\pi : \mathbb{G}_m^d \to \mathbb{G}_m^{d-1}$ onto the first $d-1$ factors is nonconstant along C, since otherwise our claim is clear. Now, the Zariski closure C' of $\pi(C)$ in $\mathbb{G}_m^{d-1}$ contains infinitely many points lying in the group

$$\Gamma' = \{(z_1^{l_1}, \cdots, z_{d-1}^{l_{d-1}}) : l_i \in \mathbb{Z}\} \leq \mathbb{G}_m^{d-1}(\overline{\mathbb{Q}} \cap \mathbb{R}).$$

By inductive hypothesis, C' is a translate of an algebraic subtorus $T' \leq \mathbb{G}_m^{d-1}$, and C' is preserved by some nontrivial $z' \in \Gamma'$. Up to translation of C within $\mathbb{G}_m^d$, we may assume that $C' = T'$. Let us also denote by $z' \in \mathbb{G}_m(\overline{\mathbb{Q}} \cap \mathbb{R})$ the element corresponding to z' under the identification $T' \simeq \mathbb{G}_m$. Applying the case $d = 2$ of the proposition to the immersion $C \hookrightarrow T' \times \mathbb{G}_m \subset \mathbb{G}_m^d$, we see that C is a translate of an algebraic subtorus in $T' \times \mathbb{G}_m$ which is invariant under some nontrivial element of $\langle z' \rangle \times \langle z_d \rangle \leq \Gamma$. This gives the desired result.

It remains to prove the proposition in the case $d = 2$. We begin by making a number of simplifying remarks. First, it suffices to prove the proposition with Γ replaced by the monoid

$$\{(z_1^{l_1}, z_2^{l_2}) : l_i \in \mathbb{Z}_{\geq 0}\},$$

since Γ is a union of finitely many monoids of the above type (obtained by replacing

(z_1, z_2) with $(z_1^{\pm 1}, z_2^{\pm 1})$). Similarly, it suffices to treat the case where $|z_1|, |z_2| > 1$, since we reduce to this case by applying inversions to factors of $\mathbb{G}_m^2$ appropriately.

Let $f \in \overline{\mathbb{Q}}[X, Y]$ be an irreducible polynomial defining the Zariski closure of C in $\mathbb{A}^2$ under the obvious embedding $\mathbb{G}_m^2 \to \mathbb{A}^2$. In what follows, we shall write $(x, y) = (z_1, z_2)$ for convenience, so that in particular $|x|, |y| > 1$ by our assumption above. Assuming that the set

$$S = \{(m, n) \in \mathbb{Z}_{\geq 0}^2 : f(x^m, y^n) = 0\}$$

is infinite, we shall show that C is a translate of algebraic torus and is preserved by some nontrivial $z \in \Gamma$. Let us write

$$f(X, Y) = \sum_{i=1}^{r} a_i X^{d_i} Y^{e_i}$$

with nonzero $a_i \in \overline{\mathbb{Q}}$ and $d_i, e_i \in \mathbb{Z}_{\geq 0}$, such that $(d_i, e_i) \neq (d_j, e_j)$ whenever $i \neq j$. Note that the number r of terms in the above sum is at least 2 since $x, y \neq 0$. Upon relabeling the terms and passing to an infinite subset of S, we may assume that

$$(r-1)|a_2 x^{d_2 m} y^{e_2 n}| \geq |a_1 x^{d_1 m} y^{e_1 n}| \geq \cdots \geq |a_r x^{d_r m} y^{e_r n}|$$

for all $(m, n) \in S$. Let $q = \gcd(d_1 - d_2, e_1 - e_2) > 0$, and define $d, e \geq 0$ to be coprime integers such that $|d_1 - d_2| = qd$ and $|e_1 - e_2| = qe$. The above inequalities imply that there is a constant $K \geq 1$ with

$$K^{-1} \leq \left| \frac{(x^m)^d}{(y^n)^e} \right| \leq K \qquad (*)$$

for all $(m, n) \in S$. Assigning weights e and d to the variables X and Y respectively,

let us write

$$f(X,Y) = g(X,Y) + h(X,Y)$$

where g is the top degree weighted homogeneous part of f, and the remainder h is a polynomial of lower degree. We remark that g must include the term $a_1X^{d_1}Y^{e_1}$. Indeed, for any $i = 1, \cdots, r$ we must have

$$\begin{aligned}|a_i|K^{-d_i/d}|y^n|^{(d_ie+de_i)/d} &\leq |a_i(x^m)^{d_i}(y^n)^{e_i}| \\ &\leq |a_1(x^m)^{d_1}(y^n)^{e_1}| \leq |a_1|K^{d_1/d}|y^n|^{(d_1e+de_1)/d}.\end{aligned}$$

Thus, if $d_1e + de_1 < d_ie + de_i$, then $|y|^n$ must be bounded for every $(m,n) \in S$, and by a similar argument $|x|^m$ must be bounded for every $(m,n) \in S$, which is a contradiction. We also have

$$\begin{aligned}\deg(a_2X^{d_2}Y^{e_2}) = d_2e + de_2 &= (d_2 - d_1)e + d_1e + d(e_2 - e_1) + de_1 \\ &= (qde - qde) + (d_1e + de_1) = \deg(a_1X^{d_1}Y^{e_1})\end{aligned}$$

by definition, so g also includes the term $a_2X^{d_2}Y^{e_2}$. In fact, we see that $a_iX^{d_i}Y^{e_i}$ is a term in g if and only if $(d_i, e_i) = (d_1 + q'd, e_1 - q'e)$ for some $q' \in \mathbb{Z}$. This shows that we can factorize $g(X,Y)$ as

$$g(X,Y) = cX^{\lambda}Y^{\mu}\prod_j(\alpha_jX^d - Y^e)^{\nu_j}$$

for some nonzero $c \in \overline{\mathbb{Q}}$ and pairwise distinct nonzero $\alpha_j \in \overline{\mathbb{Q}}$. Up to relabeling and

passing to an infinite subset of S, there exists some $\kappa > 0$ with

$$\begin{aligned}\left|\alpha_1 \frac{(x^m)^d}{(y^n)^e} - 1\right|^{\nu_1} &= \frac{|g(x^m, y^n)|}{|c(x^m)^\lambda (y^n)^\mu \prod_{j\neq 1}(\alpha_j x^{md} - y^{ne})^{\nu_j}|} \frac{1}{|y^{ne}|^{\nu_1}} \\ &= \frac{1}{|c(x^m/y^n)^\lambda| \prod_{j\neq 1} |\alpha_j (x^{md}/y^{ne}) - 1|^{\nu_j}} \frac{|h(x^m, y^n)|}{|y^{ne}|^{(\lambda+\mu)/e + \sum_j \nu_j}} \\ &\ll \frac{1}{|y^{ne}|^\kappa} \ll \frac{1}{\max\{|x^d|^m, |y^e|^n\}^\kappa}.\end{aligned}$$

for every $(m, n) \in S$. Here, we have used the inequality $(*)$. We remark that we must have $\alpha_1 \in \overline{\mathbb{Q}} \cap \mathbb{R}$. In particular, there exists $M > 1$ such that

$$|\alpha_1| \frac{|x^d|^m}{|y^e|^n} = 1 + O\left(\frac{1}{M^{\max\{m,n\}}}\right).$$

for every $(m, n) \in S$. Taking logarithms and using the fact that $|\log(1 + \xi)| \leq 2|\xi|$ for every $|\xi| \leq 1/2$, upon passing to a suitable infinite subset of S we have

$$|\log|\alpha_1| + m \log|x^d| - n \log|y^e|| \ll \frac{1}{M^{\max\{m,n\}}}.$$

By Baker's theorem [1, Theorem 3.1], if $\log|\alpha_1|$, $\log|x^d|$, and $\log|y^e|$ were $\mathbb{Q}$-linearly independent, then we would have

$$|\log|\alpha_1| + m \log|x^d| - n \log|y^e|| \gg \max\{m, n\}^{-D}$$

for some effective constant $D > 0$. Therefore, what we have obtained shows that x and y must be multiplicatively dependent. Let $v > 1$ be the unique real algebraic number such that $(|x|, |y|) = (v^a, v^b)$ for positive coprime integers $a, b \in \mathbb{Z}$. We may rewrite $(*)$ as

$$K^{-1} \leq v^{adm - ben} \leq K.$$

This shows that $adm - ben$ takes at most finitely many values, and so there is some

$t \in \mathbb{Z}$ such that, up to passing to an infinite subset of S, we have $adm - ben = t$ for all $(m, n) \in S$. Fix $(m_0, n_0) \in S$. We then have

$$ad(m - m_0) = be(n - n_0)$$

for every $(m, n) \in S$. Since ad and be are coprime, for each $(m, n) \in S$ we have $N = N_{m,n} \in \mathbb{Z}$ with $(m, n) = (m_0, n_0) + N \cdot (be, ad)$. Consider now the morphism

$$\Phi : \mathbb{A}^1 \to \mathbb{A}^2$$

given by $\Phi(T) = (x^{m_0} T^e, y^{n_0} T^d)$. Note that Φ restricts to a morphism $\mathbb{G}_m \to \mathbb{G}_m^2$ so that the image of Φ is a translation of an algebraic torus. Furthermore, given $(m, n) \in S$ and $N = N_{m,n}$ as above, we have

$$\begin{aligned}\Phi(v^{Nab}) &= (x^{m_0} v^{Nabe}, y^{n_0} v^{Nabd}) = (x^{m_0} (v^{abe})^N, y^{m_0} (v^{abd})^N) \\ &= (x^{m_0} x^{beN}, y^{n_0} y^{adN}) = (x^m, y^n).\end{aligned}$$

This shows that the Zariski closure of C has infinitely many points in common with the image of Φ. This implies that $\Phi(\mathbb{G}_m) = C$, which is the desired result. □

5.4.3 Integrable curves

Fix a pants decomposition $P \subset \Sigma$. Let Γ_P be the free abelian subgroup of rank $3g + n - 3$ in the mapping class group $\Gamma(\Sigma)$ generated by Dehn twists along the curves in P. The action of Γ_P on X_k preserves the fiber $X_{k,t}^P$ for each $t \in X(P, \mathbb{C})$. For fixed $t = (t_1, \cdots, t_{3g+n-3})$, let us fix a sequence $(z_1, \cdots, z_{3g+n-3}) \in (\mathbb{C}^\times)^{3g+n-3}$ of complex numbers such that

$$z_i + z_i^{-1} = t_i \quad \text{for all} \quad i = 1, \cdots, 3g + n - 3.$$

Let Γ_z be the subgroup of $\mathbb{G}_m^{3g+n-3}(\mathbb{C})$ generated by translations by z_i (in the multiplicative sense) in the ith coordinate. Our discussion in Section 5.1.2 leads us to the following result.

Proposition 5.26. *If $X_{k,t}^P$ is perfect, then there is a morphism*

$$F : \mathbb{G}_m^{3g+n-3} \to X_{k,t}^P$$

of schemes (defined over $\overline{\mathbb{Q}}$ if k and t are algebraic) satisfying the following:

(1) *At the level of complex points, F is surjective with finite fibers,*

(2) *the action of Γ_z on $\mathbb{G}_m^{3g+n-3}$ lifts the Γ_P-action on $X_{k,t}^P$.*

Proof. We shall describe the map induced by F on the complex points. It will be conceptually clear from our construction, even if laborious to show, that the map is induced from a morphism of schemes, and is moreover defined over $\overline{\mathbb{Q}}$ provided that k and t are algebraic.

To construct F, we first fix an $\mathrm{SL}_2(\mathbb{C})$-local system ρ_0 on Σ whose class lies in the fiber $X_{k,t}^P$. Let us write $P = a_1 \sqcup \cdots \sqcup a_{3g+n-3}$ with each a_i a curve, on which we fix a base point $x_i \in a_i$. Let α_i be a choice of a simple loop based at x_i parametrizing a_i. We fix a trivialization of the fiber of ρ above each x_i so that the monodromy along α_i is given by a diagonal matrix of the form

$$\begin{bmatrix} z_i & 0 \\ 0 & z_i^{-1} \end{bmatrix} \in \mathrm{SL}_2(\mathbb{C}). \qquad (*)$$

This is possible since $t_i \neq \pm 2$ by our hypothesis that $X_{k,t}^P$ is perfect. Setting aside the case $(g, n, k) = (1, 1, 2)$ which is elementary, this hypothesis also implies that the restriction of ρ_0 to each connected component of $\Sigma | P$ is irreducible, and in fact

determines the isomorphism type of such restriction for every local system ρ whose class lies in $X^P_{k,t}$.

For each $i = 1, \cdots, 3g+n-3$, let us denote by a'_i and a''_i the two boundary curves on $\Sigma|P$ corresponding to a_i, and let (x'_i, α'_i) and (x''_i, α''_i) be the corresponding lifts of (x_i, α_i), respectively. We shall assume that we have chosen the labelings so that the interior of $\Sigma|a$ lies ot the left as one travels alons α'_i. The above observation shows that any local system with class in $X^P_{k,t}$ is determined by the isomorphisms of local systems $\rho_0|_{a'_i}$ and $\rho_0|_{a''_i}$ compatible with the gluing of a'_i and a''_i in Σ. For instance, ρ_0 itself is the local system determined by the identity isomorphisms

$$\mathrm{id} : \rho_0|_{a'_i} \simeq \rho_0|_{a_i} \simeq \rho_0|_{a''_i}.$$

More generally, an isomorphism of local systems $\rho_0|_{a'_i} \simeq \rho_0|_{a''_i}$ is specified by an element in the centralizer of the matrix $(*)$ above (namely, the group of diagonal matrices of determinant one). Thus, we have a map

$$F : \mathbb{G}_m^{3g+n-3}(\mathbb{C}) \to X^P_{k,t}(\mathbb{C})$$

sending $(v_1, \cdots, v_{3g+n-3}) \in \mathbb{G}_m^{3g+n-3}(\mathbb{C})$ to the class of the local system determined by the isomorphisms

$$\begin{bmatrix} v_i & 0 \\ 0 & v_i^{-1} \end{bmatrix} : \rho_0|_{a'_i} \simeq \rho_0|_{a''_i}.$$

The fact that this is surjective follows from the above discussion. To see that F has finite fibers, note that if we glue the local systems $\rho_0|_{\Sigma'}$ on the components of Σ' of $\Sigma|P$ along the curves $a_1, \cdots, a_{3g+n-3}$ in a fixed order, at each stage the resulting local system is uniquely determined, except possibly when a_i is a separating curve in which case there is a double ambiguity.

Finally, the compatibility of the action of Γ_z with the action of Γ_P under F follows

from the our description of lifts of Dehn twists in Section 5.1.2. □

Corollary 5.27. *Let $X^P_{k,t}$ be a perfect fiber. Then*

(1) $|\Gamma_P \backslash X^P_{k,t}(A)| < \infty$ *for any closed discrete* $A \subset \mathbb{R}$, *and*

(2) *if no coordinate of* $t \in X(P, \mathbb{C}) \simeq \mathbb{A}^{3g+n-3}$ *lies in* $[-2, 2]$, *then*

$$|\Gamma_P \backslash X^P_{k,t}(A)| < \infty$$

for any closed discrete $A \subset \mathbb{C}$.

Proof. Let us write $t = (t_1, \cdots, t_{3g+n-3})$, and let us first suppose $t_i \notin [-2, 2]$ for each $i = 1, \cdots, 3g + n - 3$. This shows in particular that each z_i (defined above so that $z_i + z_i^{-1} = t_i$) has absolute value different from 1. In particular, every point in $\mathbb{G}_m^{3g+n-3}(\mathbb{C})$ is Γ_z-equivalent to a point in a region $K \subset \mathbb{G}_m^{3g+n-3}(\mathbb{C})$ which is compact with respect to the Euclidean topology. Under the map

$$F : \mathbb{G}_m^{3g+n-3}(\mathbb{C}) \to X^P_{k,t}(\mathbb{C})$$

constructed in Proposition 5.26, the image of K in $X^P_{k,t}(\mathbb{C})$ is compact and hence has finite intersection with any closed discrete $A \subset \mathbb{C}$. The equivariance property of F proved in Proposition 5.26 then implies our claim, when $t_i \notin [-2, 2]$ for each $i = 1, \cdots, 3g + n - 3$.

Let us now assume that A is a closed discrete subset of $\mathbb{R}$. Note in particular that the components of t_i may be assumed to lie in A and hence are real, since otherwise $X^P_{k,t}(A)$ is empty. Now, let us consider the natural morphism

$$X^P_{k,t}(\Sigma) \to \prod_{i=1}^{3g+n-3} X^{a_i}_{k_i,t_i}(\Sigma_i)$$

where each Σ_i is the surface of type $(0, 4)$ or $(1, 1)$ obtained by gluing together the

two boundary curves on $\Sigma|P$ corresponding to a_i, and the boundary traces k_i are appropriately determined from k, P, t, and Σ_i. The fact that the map F constructed in Proposition 5.26 has finite fibers implies that the above morphism also has finite fibers (at the level of complex points). The claim to be proved thus reduces to the case where Σ is of type $(0,4)$ or $(1,1)$. But this is obvious by elementary geometric considerations (see also the last paragraph in the proof of Theorem 5.1). □

Proposition 5.28. *Let C be a geometrically irreducible algebraic curve over $\mathbb{Z}$ lying in a perfect fiber $X^P_{k,t}$. Then $C(\mathbb{Z})$ can be effectively determined, and*

(1) *$C(\mathbb{Z})$ is finite, or*

(2) *$C(\mathbb{Z})$ is finitely generated under some nontrivial $\gamma \in \Gamma_P$ preserving C.*

If moreover C is not fixed pointwise by any nontrivial $\gamma \in \Gamma_P$, the same result holds with $C(\mathbb{Z})$ replaced by the set of all imaginary quadratic integral points on C.

Proof. Note that $t \in X(P,\mathbb{Z}) \simeq \mathbb{Z}^{3g+n-3}$. We shall first consider the case where no coordinate of t lies in $[-2,2]$. By Corollary 5.27, the set $\bigcup_{d>0} X^P_{k,t}(O_d)$ consists of finitely many Γ_P-orbits. Thus, it suffices to consider the intersection of $C(\mathbb{C})$ with the orbit of a single point $\rho \in \bigcup_{d>0} X^P_{k,t}(O_d)$.

Let $F : \mathbb{G}_m^{3g+n-3} \to X^P_{k,t}$ be a Γ_P-equivariant morphism as constructed in the proof of Proposition 5.26. Let $z = (z_1, \cdots, z_{3g+n-3}) \in \mathbb{G}_m^{3g+n-3}$ be as defined earlier in this subsection. By our hypothesis on t, each z_i is a real quadratic integer. Let $C' = F^{-1}(C)$, and choose a point $p \in F^{-1}(\rho)$. Applying Proposition 5.25 to the curve $C' \subset \mathbb{G}_m^{3g+n-3}$ and projecting the result down to $X^P_{k,t}$, we obtain the result.

Let us write $P = a_1 \sqcup \cdots \sqcup a_{3g+n-3}$, and let us denote by $t_i \in \mathbb{Z}$ the component of t corresponding to a_i. Based on the above argument, it remains to consider the case where $t_i \in [-2,2]$ for some i. Since we must have $t_i \in \{0, \pm 1\}$, we see that the Dehn twist τ_{a_i} acts on the fiber $X^P_{k,t}$ with finite order (as seen from the discussion in

Section 5.1.2), so we need only to consider $C(\mathbb{Z})$. Let Σ_i be the surface of type $(0,4)$ or $(1,1)$ obtained by gluing together the two boundary curves on $\Sigma|P$ corresponding to a_i, and consider the composition of morphisms

$$C \to X^P_{k,t}(\Sigma) \to X^{a_i}_{k_i,t_i}(\Sigma_i)$$

where $k_i \in X(\partial\Sigma_i, \mathbb{C})$ is appropriately determined from k, t, and Σ_i. Note that the set of real points of $X^{a_i}_{k_i,t_i}(\Sigma_i)$ defines a ellipse in an appropriate coordinate plane, as seen from our discussion in Section 5.2.1. In particular, $X^{a_i}_{k_i,t_i}(\Sigma_i, \mathbb{Z})$ is finite, and if the above composition is nonconstant then we find that $C(\mathbb{Z})$ is finite, as desired. It thus remains to consider the case where the above composition is constant. This implies that the morphism

$$C \to X^P_{k,t}(\Sigma) \to X^{P'}_{k',t'}(\Sigma|a_i)$$

(where k', P', and t' are appropriately determined from k, P, t, and a_i) must be nonconstant, as seen from the consideration of the morphism $F : \mathbb{G}_m^{3g+n-3} \to X^P_{k,t}$ as constructed in the proof of Proposition 5.26. Thus, we may apply induction and a straightforward modification of the previous paragraph to conclude the result. $\square$

5.4.4 Proofs of Theorem 5.5 and Corollary 5.6

We obtain Theorem 5.5 by combining Theorem 5.4 for nonintegrable curves with Proposition 5.28 for integrable curves in X_k. Finally, Corollary 5.6 follows easily from Theorem 5.5 (and indeed from Theorem 5.4) and our understanding of the fibers $X^P_{k,t}$ from Section 5.2.1.

Chapter 6

Periodic points

Let Σ be a surface of type (g, n). The mapping class group of Σ acts on the moduli space $X = X(\Sigma)$ introduced in Chapter 2. By definition, the *periodic points* on $X(\mathbb{C})$ are the points with finite mapping class group orbits. In this chapter, we give a characterization of the periodic points on X when Σ has positive genus. We first state our main result for $g \geq 2$.

Theorem 6.A. *Let Σ be a surface of type (g, n) with $g \geq 2$. Up to semisimplification, the periodic points in X correspond precisely to the finite representations.*

In particular, this gives a dynamical characterization of the finite $\mathrm{SL}_2(\mathbb{C})$-local systems on surfaces of genus $g \geq 2$. Let us next define an irreducible representation $\pi_1\Sigma \to \mathrm{SL}_2(\mathbb{C})$, or the corresponding local system, to be *special dihedral* if its image lies in the infinite dihedral group (recalled in Section 6.2) and there is a nonseparating simple closed curve a in Σ such that the restriction of the representation to the complement $\Sigma\backslash a$ is diagonal. We state our main result for $g = 1$.

Theorem 6.B. *Let Σ be a surface of type $(1, n)$. Up to semisimplification, the periodic points in X correspond precisely to the finite or special dihedral representations.*

Theorems 6.A and 6.B show in particular that, for surfaces of genus $g > 0$, the representations of $\pi_1\Sigma$ with Zariski dense image in $\mathrm{SL}_2(\mathbb{C})$ have infinite mapping

class group orbits in the character variety. This suggests an analogy with the result of Previte-Xia [72] on the topological density of orbits for SU(2)-character varieties of positive genus surfaces. In contrast, for $(g, n) = (0, 4)$ there are periodic points on the character variety that correspond to representations with Zariski dense image in $\mathrm{SL}_2(\mathbb{C})$. This increase in complexity of periodic points may be attributed to the inverse change in complexity of mapping class groups.

We discuss the literature and motivation behind our results. Character varieties and their mapping class group actions appear in several different contexts, and the periodic points often hold special interest.

For $(g, n) = (0, 4)$ and $(1, 1)$, the periodic points are intimately connected to the algebraic solutions of Painlevé VI equations. These are a family of nonlinear second-order ordinary differential equations on the complex plane, whose solutions have three essential singularities at 0, 1, and ∞ and have no other singularities except movable poles. The nonlinear monodromy of the solutions is related to the mapping class group action on the character variety for $(g, n) = (0, 4)$. Using this connection, Lisovyy-Tyhkyy [59] completely classified the periodic points in this case as well as the algebraic solutions to Painlevé VI. This was preceded by the work of Dubrovin-Mazzocco [27], on a special subfamily of equations, which essentially settled the periodic points for $(g, n) = (1, 1)$. See also the work of Hitchin [49] on explicit constructions of algebraic Painlevé VI's arising from finite dihedral representations. Following Boalch [6], the classification of algebraic Painlevé VI may be viewed as a nonlinear analogue of *Schwarz's list* classifying the algebraic solutions to the hypergeometric equation of Riemann. The study of periodic points on character varieties for general (g, n) is a natural extension of this program.

Another source of interest for the finite mapping class group orbits lies in the Diophantine geometry of the relative character varieties $X_k \subset X$. For $(g, n) = (1, 1)$, the subvarieties X_k are (up to the coefficient 3) the level sets of the Markoff cubic

$x_1^2 + x_2^2 + x_3^2 - 3x_1x_2x_3$. Bourgain-Gamburd-Sarnak [10] examined the question of strong approximation for these affine cubic surfaces, which asks for the surjectivity of the reduction mod q map

$$X_k(\mathbb{Z}) \to X_k(\mathbb{Z}/q\mathbb{Z})$$

away from an exceptional locus. Their approach is based on showing the transitivity of the mapping class group action on the nonexceptional locus of $X_k(\mathbb{Z}/q\mathbb{Z})$. A finite mapping class group orbit over $\mathbb{C}$ appears in $X_k(\mathbb{Z}/q\mathbb{Z})$ for infinitely many q by reduction, and hence must be accounted for as part of the exceptional locus. Classification of the periodic points therefore arises naturally as part of our program to investigate strong approximation for general character varieties. Similarly, the classification of higher-dimensional mapping class group invariant subvarieties of character varieties is an interesting problem, which leave to future work.

Theorems 6.A and 6.B also verify the case $G = \mathrm{SL}_2$ of a conjecture of Kisin (See for instance Chapter 1 of [81]) that, for π the fundamental group of a closed surface or a free group of rank $r \geq 3$, the points with finite orbits for the action of the outer automorphism group $\mathrm{Out}(\pi)$ on the character variety $\mathrm{Hom}(\pi, G) /\!\!/ G$, for reductive algebraic groups G, correspond to representations $\pi \to G$ with virtually solvable image.

This chapter is organized as follows. We begin by recording background on subgroups of $\mathrm{SL}_2(\mathbb{C})$ in Section 6.1. We also introduce the notion of loop configurations as a tool to keep track of subsurfaces of a surface. In Section 6.2, we deduce the case $(g, n) = (1, 1)$ of Theorem 6.B from the results of Dubrovin-Mazzocco [27]. In Section 6.3, we show that Zariski dense representations, i.e. representations with Zariski dense image in $\mathrm{SL}_2(\mathbb{C})$, are not periodic in the character variety for surfaces of positive genus. We prove this using the case $(g, n) = (1, 1)$ and the following lemma,

which may be of independent interest.

Main Lemma. *Let Σ be a surface of type (g,n) with $g \geq 1$. A representation $\pi_1\Sigma \to \mathrm{SL}_2(\mathbb{C})$ is irreducible, resp. Zariski dense, if and only if its restriction to some embedded surface of type $(1,1)$ is irreducible, resp. Zariski dense.*

In Section 6.4, we verify Theorems 6.A and 6.B on the locus of representations that do not have Zariski dense image in $\mathrm{SL}_2(\mathbb{C})$, thus completing the proof.

6.1 Background

6.1.1 Subgroups of $\mathrm{SL}_2(\mathbb{C})$

Let G be a proper algebraic subgroup of $\mathrm{SL}_2(\mathbb{C})$. Up to conjugation, G satisfies one of the following [89, Theorem 4.29]:

(1) G is a subgroup of the standard Borel group

$$B = \left\{ \begin{bmatrix} a & b \\ 0 & a^{-1} \end{bmatrix} : a \in \mathbb{C}^*, b \in \mathbb{C} \right\}.$$

(2) G is a subgroup of the infinite dihedral group

$$D_\infty = \left\{ \begin{bmatrix} c & 0 \\ 0 & c^{-1} \end{bmatrix} : c \in \mathbb{C}^* \right\} \cup \left\{ \begin{bmatrix} 0 & c \\ -c^{-1} & 0 \end{bmatrix} : c \in \mathbb{C}^* \right\}.$$

(3) G is one of the finite groups BA_4, BS_4, and BA_5, which are the preimages in $\mathrm{SL}_2(\mathbb{C})$ of the finite subgroups A_4 (tetrahedral group), S_4 (octahedral group), and A_5 (icosahedral group) of $\mathrm{PGL}_2(\mathbb{C})$, respectively.

Remark. The finite subgroups BA_4, BS_4, BA_5 of $\mathrm{SL}_2(\mathbb{C})$ can be explicitly described

as follows. First, let us identify the group of unit quaternions

$$\mathrm{Sp}(1) = \{z = (a, b, c, d) = a + bi + cj + dk \in \mathbb{H} : |z| = a^2 + b^2 + c^2 + d^2 = 1\}$$

as a subgroup of $\mathrm{SL}_2(\mathbb{C})$ by the map

$$z = (a, b, c, d) \mapsto \begin{bmatrix} a + bi & c + di \\ -c + di & a - bi \end{bmatrix}.$$

Under the identification, the *binary tetrahedral group* BA_4 is given by

$$BA_4 = \{\pm 1, \pm i, \pm j, \pm k, (\pm 1 \pm i \pm j \pm k)/2\}$$

with all sign combinations taken in the above. The *binary octahedral group* BS_4 is the union of BA_4 with all quaternions obtained from $(\pm 1, \pm 1, 0, 0)/\sqrt{2}$ by all permutations of coordinates and all sign combinations. The *binary icosahedral group* BA_5 is the union of BA_4 with all quaternions obtained from $(0, \pm 1, \pm\varphi^{-1}, \pm\varphi)/2$ by an even permutation of coordinates and all possible sign combinations, where $\varphi = (1 + \sqrt{5})/2$ is the golden ratio.

The above classification has the following consequence. Given a subset $S \subset G(\mathbb{C})$ of an algebraic group G, let $\mathrm{Zcl}(S)$ denote the Zariski closure of S. If $\pi \leq G(\mathbb{C})$ is a group, then $\mathrm{Zcl}(\pi)$ is an algebraic subgroup of G. We say that π is Zariski dense in G if $\mathrm{Zcl}(\pi) = G$.

Lemma 6.1. *A torsion-free subgroup of* $\mathrm{SL}_2(\mathbb{C})$ *is Zariski dense in* $\mathrm{SL}_2(\mathbb{C})$ *if and only if it is not conjugate to a subgroup of the standard Borel group* B*.*

Proof. The *only if* direction is clear, and we prove the converse. Let π be a torsion-free subgroup of $\mathrm{SL}_2(\mathbb{C})$ not conjugate to a subgroup of B. This implies that π is nontrivial, and if π is not Zariski dense in $\mathrm{SL}_2(\mathbb{C})$ then up to conjugation it is a

subgroup of D_∞, BA_4, BS_4, or BA_5. The last three cases are immediately ruled out by the condition that π is torsion-free. If π is conjugate to a subgroup of D_∞ and is not conjugate to a subgroup of B, then there exists an element of π which is conjugate to a matrix of the form

$$\begin{bmatrix} 0 & c \\ -c^{-1} & 0 \end{bmatrix}$$

for some $c \in \mathbb{C}^*$. But this matrix has order 4, contradicting the hypothesis that π is torsion-free. Hence, π must be Zariski dense in $\mathrm{SL}_2(\mathbb{C})$. □

We record the following well-known observations.

Lemma 6.2. *Given subgroups $\pi' \leq \pi$ of an irreducible algebraic group G such that $[\pi : \pi'] < \infty$, we have $\mathrm{Zcl}(\pi) = G$ if and only if $\mathrm{Zcl}(\pi') = G$.*

Proof. The *if* direction is clear. For the converse, let $g_1, \cdots, g_N \in \pi$ be a complete set of coset representatives for π/π'. Note that $\mathrm{Zcl}(g_i\pi') = g_i \,\mathrm{Zcl}(\pi')$ for each $i = 1, \cdots, N$, and hence

$$\mathrm{Zcl}(\pi) = \mathrm{Zcl}\left(\bigcup_{i=1}^{N} g_i\pi'\right) = \bigcup_{i=1}^{N} \mathrm{Zcl}(g_i\pi') = \bigcup_{i=1}^{N} g_i \,\mathrm{Zcl}(\pi').$$

Suppose that $\mathrm{Zcl}(\pi) = G$. Since G is assumed to be irreducible, we see that $G \subseteq g_i \,\mathrm{Zcl}(\pi')$ for some g_i, which implies that $\mathrm{Zcl}(\pi') = G$, as desired. □

Lemma 6.3. *A pair (a, b) of elements in $\mathrm{SL}_2(\mathbb{C})$ has a common eigenvector in $\mathbb{C}^2$, or in other words lies in B up to simultaneous conjugation, if and only if $\mathrm{tr}([a, b]) = 2$, where $[a, b] = aba^{-1}b^{-1}$.*

Proof. After conjugating a and b by an element of $\mathrm{SL}_2(\mathbb{C})$, we reduce to one of the following three cases.

(1) We have $a = \pm\mathbf{1}$, where $\mathbf{1} = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$ is the identity matrix. In this case, (a, b) obviously has a common eigenvector and $\mathrm{tr}([a, b]) = \mathrm{tr}(\mathbf{1}) = 2$.

(2) We have $a = s\left[\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right]$ where $s \in \{\pm 1\}$. Writing $b = \left[\begin{smallmatrix} b_1 & b_2 \\ b_3 & b_4 \end{smallmatrix}\right]$, we have

$$\mathrm{tr}(aba^{-1}b^{-1}) = \mathrm{tr}\left(\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} b_4 & -b_2 \\ -b_3 & b_1 \end{bmatrix}\right) = 2 + b_3^2,$$

using the fact that $b_1b_4 - b_2b_3 = 1$. Hence, we have $\mathrm{tr}([a, b]) = 2$ if and only if $b_3 = 0$, which is equivalent to saying that (a, b) has a common eigenvector.

(3) We have $a = \left[\begin{smallmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{smallmatrix}\right]$ for some $\lambda \in \mathbb{C}^* \setminus \{\pm 1\}$. Writing $b = \left[\begin{smallmatrix} b_1 & b_2 \\ b_3 & b_4 \end{smallmatrix}\right]$,

$$\begin{aligned}\mathrm{tr}(aba^{-1}b^{-1}) &= \mathrm{tr}\left(\begin{bmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{bmatrix}\begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix}\begin{bmatrix} \lambda^{-1} & 0 \\ 0 & \lambda \end{bmatrix}\begin{bmatrix} b_4 & -b_2 \\ -b_3 & b_1 \end{bmatrix}\right) \\ &= 2b_1b_4 - (\lambda^2 + \lambda^{-2})b_2b_3.\end{aligned}$$

The condition $\mathrm{tr}([a, b]) = 2$ is then equivalent to $b_2b_3(\lambda^2 + \lambda^{-2} - 2) = 0$. Since $\lambda^2 + \lambda^{-2} - 2 = (\lambda - \lambda^{-1})^2 \neq 0$ by the hypothesis that $\lambda \neq \pm 1$, we conclude that this is equivalent to the statement that $b_2 = 0$ or $b_3 = 0$, or in other words, (a, b) has a common eigenvector.

Thus, (a, b) has a common eigenvector if and only if $\mathrm{tr}([a, b]) = 2$. □

Consider the morphism $i : \mathbb{C}^* \to \mathrm{SL}_2(\mathbb{C})$ given by $a \mapsto \left[\begin{smallmatrix} a & 0 \\ 0 & a^{-1} \end{smallmatrix}\right]$ and inclusion $j : D_\infty \to \mathrm{SL}_2(\mathbb{C})$. They induce $\mathrm{Out}(\pi)$-equivariant maps

$$i_* : \mathrm{Hom}(\pi, \mathbb{C}^*) \to X(\mathbb{C}) \quad \text{and} \quad j_* : \mathrm{Hom}(\pi, D_\infty)/D_\infty \to X(\mathbb{C}).$$

Lemma 6.4. *We have the following.*

(1) *The map $i_* : \mathrm{Hom}(\pi, \mathbb{C}^*) \to X(\mathbb{C})$ has finite fibers.*

(2) *The map* $j_* : \mathrm{Hom}(\pi, D_\infty)/D_\infty \to X(\mathbb{C})$ *has finite fibers.*

Proof. (1) Let $\rho_1, \rho_2 : \pi \to \mathbb{C}^*$ be characters such that $gi_*(\rho_1)g^{-1} = i_*(\rho_2)$ for some $g \in \mathrm{SL}_2(\mathbb{C})$. Without loss of generality, we may assume given some $a \in \pi$ such that $\rho_1(a) \notin \{\pm\mathbf{1}\}$, since otherwise the image of ρ_1 is finite and we are done. Writing $g = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ and $\rho_1(a) = \lambda$ with $\lambda \in \mathbb{C}^* \setminus \{\pm 1\}$, we have

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{bmatrix} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \begin{bmatrix} \lambda ad - \lambda^{-1}bc & (\lambda^{-1} - \lambda)ab \\ (\lambda - \lambda^{-1})cd & \lambda^{-1}ad - \lambda bc \end{bmatrix}.$$

For the matrix on the right hand side to be diagonal, we must thus have $a = d = 0$ or $b = c = 0$ since $\lambda \neq \pm 1$. If $a = d = 0$, then $\rho_2 = \rho_1^{-1}$. If $b = c = 0$, then $\rho_1 = \rho_2$. This proves (1).

(2) Let $\rho_1, \rho_2 : \pi \to D_\infty$ be representations such that $gj_*(\rho_1)g^{-1} = j_*(\rho_2)$ for some $g \in \mathrm{SL}_2(\mathbb{C})$. Without loss of generality, we may assume given $a \in \pi$ with $\mathrm{tr}\,\rho_1(a) \notin \{0, \pm 2\}$, since otherwise the image of ρ_1 is finite and we are done. Note that $\rho_1(a)$ must be diagonal. The equation $g\rho_1(a)g^{-1} = \rho_2(a)$ shows that, by the same computation as above, we have $g \in D_\infty$. This proves (2). □

We make the following definition.

Definition 6.5. A representation $\pi \to \mathrm{SL}_2(\mathbb{C})$ is said to be:

(1) *Zariski dense* if its image is Zariski dense in $\mathrm{SL}_2(\mathbb{C})$,

(2) *diagonal* if it factors through $i : \mathbb{C}^* \to \mathrm{SL}_2(\mathbb{C})$,

(3) *dihedral* if it factors through $j : D_\infty \to \mathrm{SL}_2(\mathbb{C})$, and

(4) *finite* if its image is finite.

6.1.2 Loop configurations

Let Σ be a surface of type (g, n). We fix a base point in Σ. For convenience, we shall say that a sequence $L = (\ell_1, \cdots, \ell_m)$ of based loops on Σ is *clean* if each loop is simple and the loops pairwise intersect only at the base point.

Definition 6.6. A *loop configuration* is a planar graph consisting of a single vertex v and a finite cyclically ordered sequence of directed rays, equipped with a bijection between the set of rays departing from v is and the set of rays arriving at v. We denote by $\mathcal{L}_{g,n}$ the loop configuration whose sequence of rays is of the form

$$(a_1, b_1, \bar{a}_1, \bar{b}_1, \cdots, a_g, b_g, \bar{a}_g, \bar{b}_g, c_1, \bar{c}_1, \cdots, c_n, \bar{c}_n),$$

where a_i, b_i, c_i are the rays directed away from v, respectively corresponding to the rays $\bar{a}_i, \bar{b}_i, \bar{c}_i$ directed towards v. See Figure 6.1 for an illustration of $\mathcal{L}_{2,1}$.

Given a clean sequence $L = (\ell_1, \cdots, \ell_m)$ of loops on Σ, we have an associated loop configuration $\mathcal{L}(L)$, obtained by taking a sufficiently small open neighborhood of the base point and setting the departing and arriving ends of the loops ℓ_i to correspond to each other. For example, if $(a_1, b_1, \cdots, a_g, b_g, c_1, \cdots, c_n)$ is a sequence of optimal generators for $\pi_1\Sigma$, then

$$\mathcal{L}(a_1, b_1, \cdots, a_g, b_g, c_1 \cdots, c_n) \simeq \mathcal{L}_{g,n}.$$

Definition 6.7. Let h and m be nonnegative integers. A sequence of based loops $L = (\ell_1, \cdots, \ell_{2h+m})$ on Σ is said to be *in* $(h, m+1)$*-position* if it is homotopic termwise to a clean sequence $L' = (\ell'_1, \cdots, \ell'_{2h+m})$ such that $\mathcal{L}(L') \simeq \mathcal{L}_{h,m}$. We denote by $\Sigma(L) \subset \Sigma$ the (isotopy class of a) subsurface of genus h with $m+1$ boundary curves obtained by taking a small closed tubular neighborhood of the union of the simple curves underlying $\ell'_1, \cdots, \ell'_{2h+m}$ in Σ.

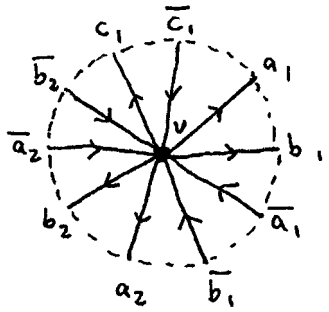


Figure 6.1: Loop configuration $\mathcal{L}_{2,1}$

Lemma 6.8. *Let (a, b, c) be a sequence of loops on a surface in $(1, 2)$-position.*

(1) *The following pairs are in $(1, 1)$-position:*

$$(a, b), (a, bc), (ca, b), (ab, bc), (ca, cb), (ac, bc), (ca, ab).$$

(2) *The triple $(c^{-1}b^{-1}a, b, c)$ is in $(1, 2)$-position.*

Proof. This is seen by drawing the corresponding loop configurations of homotopic clean sequences. See Figure 6.2 for the loop configurations occurring in part (1), and Figure 6.3 for the loop configuration of $(c^{-1}b^{-1}a, b, c)$. □

6.2 Case of one-holed torus

Let Σ be a surface of type $(1, 1)$. Let (α, β, γ) be a sequence of optimal generators for $\pi_1\Sigma$. Let $X = X(\Sigma)$ be the moduli of local systems defined in Chapter 2. As mentioned in Example 2.2, we have an isomorphism

$$(\mathrm{tr}_\alpha, \mathrm{tr}_\beta, \mathrm{tr}_{\alpha\beta}) : X(\Sigma) \xrightarrow{\sim} \mathbb{A}^3_{x_1,x_2,x_3}.$$

For simplicity, we shall denote the isotopy classes of simple curves lying in the free homotopy classes of α, β, and $\alpha\beta$ by a, b, ab. Let Π be the group of polynomial auto-

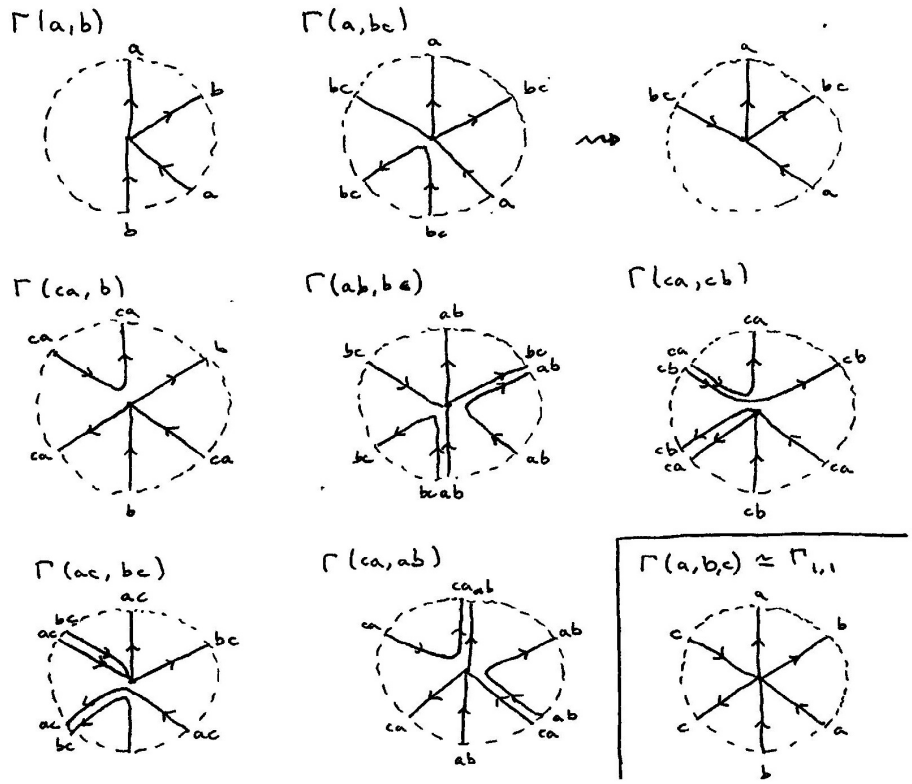


Figure 6.2: Loop configurations for part (1) of Lemma 6.8

morphisms of $\mathbb{A}^3$ generated by the Dehn transformations τ_a^*, τ_b^*, and τ_{ab}^* as described in Lemma 2.5 (with the variables x, y, z replaced by x_1, x_2, x_3). It is precisely the image of the mapping class group $\Gamma(\Sigma)$ in the group of polynomial automorphisms of $X(\Sigma) = \mathbb{A}^3$. Let Π' be the group generated by Π together with the following transformations:

$$\sigma_{12} : (x_1, x_2, x_3) \mapsto (-x_1, -x_2, x_3),$$

$$\sigma_{23} : (x_1, x_2, x_3) \mapsto (x_1, -x_2, -x_3),$$

$$\sigma_{13} : (x_1, x_2, x_3) \mapsto (-x_1, x_2, -x_3).$$

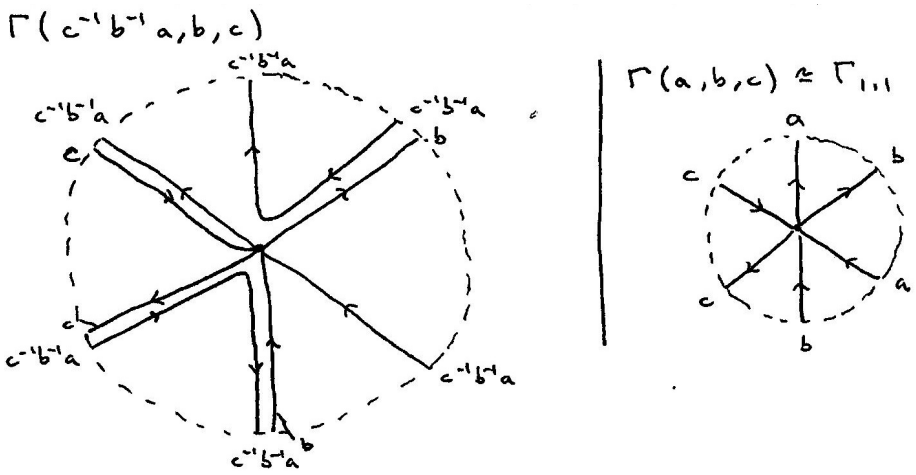


Figure 6.3: Loop configuration $\mathcal{L}(c^{-1}b^{-1}a, b, c)$

It is easy to see that $[\Pi' : \Pi] < \infty$. Hence, a point in $\mathbb{A}^3$ has finite Π-orbit if and only if it has finite Π'-orbit. Now, the group Π' contains the transformations

$$\beta_1 = \sigma_{12}\tau_{ab}^*(\tau_b^*\tau_a^*)^{-1} : (x_1, x_2, x_3) \mapsto (-x_1, x_3 - x_1x_2, x_2),$$
$$\beta_2 = \sigma_{23}\tau_a^*(\tau_b^*\tau_a^*)^{-1} : (x_1, x_2, x_3) \mapsto (x_3, -x_2, x_1 - x_2x_3).$$

whose joint orbits in $\mathbb{A}^3$ were studied in Dubrovin-Mazzocco [27, Theorem 1.6], in connection with algebraic solutions of special Painlevé VI equations. They defined a triple $(x_1, x_2, x_3) \in \mathbb{A}^3(\mathbb{C})$ to be *admissible* if it has at most one coordinate zero and $x_1^2 + x_2^2 + x_2^2 - x_1x_2x_3 - 2 \neq 2$. The result of [27] we shall use is the following.

Theorem 6.9 (Dubrovin-Mazzocco)**.** *The following is a complete set of representatives for the finite $\langle \beta_1, \beta_2 \rangle$-orbits of admissible triples in $\mathbb{A}^3$:*

$$(0, -1, -1), (0, -1, -\sqrt{2}), (0, -1, -\varphi), (0, -1, -\varphi^{-1}), (0, -\varphi, -\varphi^{-1})$$

where $\varphi = (1 + \sqrt{5})/2$ is the golden ratio.

Corollary 6.10. *If $(x_1, x_2, x_3) \in \mathbb{A}^3(\mathbb{C})$ is an admissible triple with finite $\Gamma(\Sigma)$-orbit, then it corresponds to a representation $\rho : \pi_1\Sigma \to \mathrm{SL}_2(\mathbb{C})$ with finite image.*

Proof of the corollary. Replacing (x_1, x_2, x_3) by another triple within its $\Gamma(\Sigma)$-orbit if necessary, we may assume that (x_1, x_2, x_3) is one of the triples in Theorem 6.9 or its image under one of the transformations σ_{12}, σ_{23}, or σ_{13}. We shall show that

$$(x_1, x_2, x_3) = (\operatorname{tr} a_1, \operatorname{tr} a_2, \operatorname{tr}(a_1 a_2))$$

where $a_1, a_2 \in \mathrm{SL}_2(\mathbb{C})$ are elements that together lie in one of the finite subgroups BA_4, BS_4, or BA_5 of $\mathrm{SL}_2(\mathbb{C})$. Since the matrix $-\mathbf{1}$ is contained in every one of these groups, it suffices to treat the case where (x_1, x_2, x_3) is one of the triples in Theorem 6.9. By explicit computation, we find that the triples in Theorem 6.9 respectively correspond to traces of the triples of matrices

$$\left(\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \begin{bmatrix} -\frac{1}{2}(1+i) & \frac{1}{2}(1-i) \\ -\frac{1}{2}(1+i) & -\frac{1}{2}(1-i) \end{bmatrix}, \begin{bmatrix} -\frac{1}{2}(1+i) & -\frac{1}{2}(1-i) \\ \frac{1}{2}(1+i) & -\frac{1}{2}(1-i) \end{bmatrix}\right),$$
$$\left(\begin{bmatrix} 0 & \frac{1}{\sqrt{2}}(1-i) \\ \frac{1}{\sqrt{2}}(1+i) & 0 \end{bmatrix}, \begin{bmatrix} -\frac{1}{2}(1+i) & \frac{1}{2}(1-i) \\ -\frac{1}{2}(1+i) & -\frac{1}{2}(1-i) \end{bmatrix}, \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}i \\ \frac{1}{\sqrt{2}}i & -\frac{1}{\sqrt{2}} \end{bmatrix}\right),$$
$$\left(\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \begin{bmatrix} -\frac{1}{2} & \frac{1}{2}(\varphi + \varphi^{-1}i) \\ -\frac{1}{2}(\varphi - \varphi^{-1}i) & -\frac{1}{2} \end{bmatrix}, \begin{bmatrix} -\frac{1}{2}(\varphi - \varphi^{-1}i) & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2}(\varphi + \varphi^{-1}i) \end{bmatrix}\right),$$
$$\left(\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{2}(1-\varphi i) & \frac{1}{2}\varphi^{-1} \\ -\frac{1}{2}\varphi^{-1} & -\frac{1}{2}(1+\varphi i) \end{bmatrix}, \begin{bmatrix} -\frac{1}{2}\varphi^{-1} & -\frac{1}{2}(1+\varphi i) \\ -\frac{1}{2}(1-\varphi i) & -\frac{1}{2}\varphi^{-1} \end{bmatrix}\right),$$
$$\left(\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \begin{bmatrix} -\frac{\varphi}{2} & \frac{\varphi^{-1}}{2} + \frac{1}{2}i \\ -\frac{\varphi^{-1}}{2} + \frac{1}{2}i & -\frac{\varphi}{2} \end{bmatrix}, \begin{bmatrix} -\frac{\varphi^{-1}}{2} + \frac{1}{2}i & -\frac{\varphi}{2} \\ \frac{\varphi}{2} & -\frac{\varphi^{-1}}{2} - \frac{1}{2}i \end{bmatrix}\right),$$

where $\varphi = (1 + \sqrt{5})/2$ is the golden ratio. The matrices for the first triple all lie in the binary tetrahedral group BA_4, the matrices for the second triple all lie in the binary octahedral group BS_4, and the matrices for the remaining three triples all lie in the binary icosahedral group BA_5. In each triple, the third matrix is the product

of the first two. Thus, each of the triples in Theorem 6.9 correspond to characters of representations $\pi_1\Sigma \to \mathrm{SL}_2(\mathbb{C})$ (Σ a one holed torus) with finite image, proving the corollary. □

Corollary 6.11. *Let Σ be a one holed torus. If $\rho : \pi_1\Sigma \to \mathrm{SL}_2(\mathbb{C})$ is a Zariski dense representation, then its mapping class group orbit in $X(\Sigma)$ is infinite.*

Proof. Let $x = (x_1, x_2, x_3) \in \mathbb{A}^3$ lie in a finite $\Gamma(\Sigma)$-orbit. By Corollary 6.10 above, if x is admissible then it corresponds to a representation with finite image. If x is not admissible, then by definition we have one of the following:

(1) at least two coordinates of x are zero, or

(2) $x_1^2 + x_2^2 + x_3^2 - x_1x_2x_3 - 2 = 2$.

In case (1), suppose that $x = (x_1, 0, 0)$. Let $\lambda \in \mathbb{C}^*$ be a complex number such that $\lambda + \lambda^{-1} = x_1$. We have

$$(x_1, 0, 0) = \left(\mathrm{tr} \begin{bmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{bmatrix}, \mathrm{tr} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \mathrm{tr} \begin{bmatrix} 0 & \lambda \\ -\lambda^{-1} & 0 \end{bmatrix} \right)$$

showing that x corresponds to the character of a representation with image in D_∞. The cases where $x = (0, x_2, 0)$ or $(0, 0, x_3)$ are treated similarly. Lastly, if case (2) occurs, then by Lemma 6.3 the point x corresponds to the character of a reducible representation. This gives the desired result. □

6.3 Zariski dense representations

Let Σ be a surface of type (g, n) with $g > 0$. Given a pair (a, b) of based loops in $(1, 1)$-position on Σ, there is an embedding $\Sigma(a, b) \subset \Sigma$ of a surface of type $(1, 1)$, i.e. a one holed torus. Up to isotopy, every embedding of a one holed torus is of the

form $\Sigma(a,b)$ for some choice of (a,b). For a representation $\rho : \pi_1\Sigma \to \mathrm{SL}_2(\mathbb{C})$, let $\rho|\Sigma(a,b)$ denote the representation $\pi_1\Sigma(a,b) \to \mathrm{SL}_2(\mathbb{C})$ obtained by restriction. We shall now restate and prove the irreducibility part of the Main Lemma.

Proposition 6.12. *Let Σ be a surface of type (g,n) with $g \geq 1$. A representation $\rho : \pi_1\Sigma \to \mathrm{SL}_2(\mathbb{C})$ is irreducible if and only if there is an embedded one holed torus $\Sigma(a,b) \subset \Sigma$ such that $\rho|\Sigma(a,b)$ is irreducible.*

Proof. The *if* direction is clear, and we now prove the converse. Let us fix a representation $\rho : \pi_1(S) \to \mathrm{SL}_2(\mathbb{C})$ whose restriction to every embedded one holed torus is reducible. In this proof, given $a \in \pi_1\Sigma$ we shall also denote by a the matrix $\rho(a) \in \mathrm{SL}_2(\mathbb{C})$ for simplicity. The statement that $\rho|\Sigma(a,b)$ is reducible for an embedding $\Sigma(a,b) \subset \Sigma$ associated to a pair (a,b) of loops in $(1,1)$-position is equivalent to saying that the pair $(\rho(a),\rho(b))$ of matrices in $\mathrm{SL}_2(\mathbb{C})$ has a common eigenvector in $\mathbb{C}^2$.

Throughout, we shall be using the following observation: if $a \in \mathrm{SL}_2(\mathbb{C}) \setminus \{\pm\mathbf{1}\}$, and if $x, y, z \in \mathbb{C}^2$ are eigenvectors of a, then at least two of them are proportional; in notation, $x \sim y$, $x \sim z$, or $y \sim z$. First, we prove the following claim.

Claim. *Any triple (a,b,c) of loops on Σ in $(1,2)$-position has a common eigenvector under the representation ρ.*

Let (a,b,c) be in $(1,2)$-position. Each of the following pairs is in $(1,1)$-position by part (1) of Lemma 6.8, and has a common eigenvector by our hypothesis on ρ:

$$(a,b),(a,bc),(ca,b),(ab,bc),(ca,cb),(ac,bc),(ca,ab).$$

If $ab = \pm\mathbf{1}$, then since (a,bc) has a common eigenvector we find that (a,b,c) has a common eigenvector, as desired. Henceforth, assume that $ab \neq \pm\mathbf{1}$. Now (a,b,ab) has a common eigenvector, say x; (ab,bc) has a common eigenvector, say y; and (ca,ab)

has a common eigenvector, say z. Since $ab \neq \pm\mathbf{1}$, we must have

$$x \sim y, \quad x \sim z, \quad \text{or} \quad y \sim z$$

where the relation $\sim$ indicates that the two vectors are proportional. If the first case occurs, then (a, b, bc) has a common eigenvector, and hence (a, b, c) does, as required. If the second case occurs, then (a, b, ca) has a common eigenvector, and hence (a, b, c) does, again as required. Henceforth, assume that the third case occurs, so that (bc, ca) has a common eigenvector.

Now, suppose first that $\mathrm{tr}(a) = \pm 2$. Up to conjugation, we have the following possibilities.

(1) We have $a = \pm\mathbf{1}$. Since (b, c) has a common eigenvector, this implies that (a, b, c) has a common eigenvector, as desired.

(2) We have

$$a = \pm \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Since (a, bc) and (a, b) each have a common eigenvector, bc and b must be upper triangular, and hence c is upper triangular. Thus (a, b, c) has a common eigenvector.

It remains to treat the case $\mathrm{tr}(a) \neq \pm 2$. As (a, bc) has a common eigenvector, by Lemma 6.3 we have

$$\mathrm{tr}(a)^2 + \mathrm{tr}(bc)^2 + \mathrm{tr}(abc)^2 - \mathrm{tr}(a)\,\mathrm{tr}(bc)\,\mathrm{tr}(abc) - 2 = 2$$

and this implies that we cannot have $\mathrm{tr}(bc) = \mathrm{tr}(abc) = 0$. After conjugation of ρ by an element of $\mathrm{SL}_2(\mathbb{C})$, we have one of the following three cases.

(1) We have $bc = \pm\mathbf{1}$. This implies that $b = \pm c^{-1}$. Since (a, b) has a common eigenvector, we are done.

(2) We have

$$bc = \pm \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Since (a, bc) and (bc, ca) each have a common eigenvector, a and ca are upper triangular, and hence c is upper triangular. This in turn implies that b is upper triangular, and we are done.

(3) We have

$$bc = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{bmatrix}, \quad \lambda \in \mathbb{C}^* \setminus \{\pm 1\}.$$

Since (a, bc), (bc, ca), and (ab, bc) each have a common eigenvector, at least two of a, ab, ca must be simultaneously upper or lower triangular. Let us assume that a is upper triangular; the case where a is lower triangular is dealt with similarly. If ab is also upper triangular, this implies that b is also upper triangular, in turn c is upper triangular, and we are done. Hence, we may assume ab is lower triangular. Similarly, we may assume that ca is lower triangular. Since we have

$$(ab)(ca) = a(bc)a$$

with left hand side lower triangular and right hand side upper triangular, $abca$ is diagonal. Writing

$$a = \begin{bmatrix} \mu & x \\ 0 & \mu^{-1} \end{bmatrix},$$

we compute

$$a(bc)a = \begin{bmatrix} \mu^2\lambda & x(\mu\lambda + \lambda^{-1}\mu^{-1}) \\ 0 & \mu^{-2}\lambda^{-1} \end{bmatrix}$$

which implies that $x = 0$ or $\mathrm{tr}(abc) = \mu\lambda + \lambda^{-1}\mu^{-1} = 0$. If $x = 0$, that is, a is diagonal, then we find that b and c are lower triangular, and we are done. The case $\mathrm{tr}(abc) = 0$ still remains; note the hypothesis that $\mathrm{tr}(bc) \neq \pm 2$ implies that $\mathrm{tr}(a) \neq 0$, and hence we have $\mathrm{tr}(a), \mathrm{tr}(bc) \neq 0$.

Hence, we have shown that if (a, b, c) is in $(1, 2)$-position then (a, b, c) has a common eigenvector, except possibly if $\mathrm{tr}_a, \mathrm{tr}(bc) \neq 0$ and $\mathrm{tr}(abc) = 0$. But in this exceptional case, since $(c^{-1}b^{-1}a, b, c)$ is in $(1, 2)$-position by part (2) of Lemma 6.8, and since

$$\begin{aligned}(\mathrm{tr}(c^{-1}b^{-1}a), \mathrm{tr}(bc), \mathrm{tr}(c^{-1}b^{-1}abc)) &= (\mathrm{tr}(a)\,\mathrm{tr}(bc) - \mathrm{tr}(abc), \mathrm{tr}(bc), \mathrm{tr}(a)) \\ &= (\mathrm{tr}(a)\,\mathrm{tr}(bc), \mathrm{tr}(bc), \mathrm{tr}(a))\end{aligned}$$

has none of the coordinates zero, running the same argument as above with (a, b, c) replaced by $(c^{-1}b^{-1}a, b, c)$ we find that $(c^{-1}b^{-1}a, b, c)$ has a common eigenvector. But then (a, b, c) also has a common eigenvector, as required.

Thus, we have proved our claim. To prove the proposition, we use the following inductive argument. Let $(a_1, a_2, \cdots, a_{2g-1}, a_{2g}, a_{2g+1}, \cdots, a_{2g+n})$ be an optimal sequence of generators of $\pi_1\Sigma$. We show that $(a_1, \cdots, a_{2g+n})$ has a common eigenvector. For simplicity of arguments we may assume that at least one element in each pair

$$(a_1, a_2), \cdots, (a_{2g-1}, a_{2g})$$

is not equal to $\pm\mathbf{1}$; for if some pair is of the form (a, b) with $a, b \in \{\pm\mathbf{1}\}$ we may

simply skip over that pair in the considerations below.

If $(g,n) = (1,0)$, then we are done since every representation $\pi_1\Sigma \to \mathrm{SL}_2(\mathbb{C})$ is abelian. We thus assume $(g,n) \neq (1,0)$. By our claim, (a_1, a_2, a_3) has a common eigenvector. Assume next that $4 \leq k \leq 2g+n$ and $(a_1, \cdots, a_{k-1})$ has a common eigenvector $x \in \mathbb{C}^2$. We show that $(a_1, \cdots, a_k)$ has a common eigenvector. We consider the following cases.

(1) $k = 5, 7, \cdots, 2g-1$, or $k = 2g+1, 2g+2, \cdots, 2g+n$. The triple (a_1, a_2, a_k) is in $(1,2)$-position, and hence has a common eigenvector y by our claim.

Assume now first that $a_1 \neq \pm\mathbf{1}$. Given any $3 \leq j < k$, note that (a_1, a_2a_j, a_k) is in $(1,2)$-position and hence has a common eigenvector by our claim, say y_j. Since $a_1 \neq \pm\mathbf{1}$, we must have

$$x \sim y, \quad x \sim y_j, \quad \text{or} \quad y_j \sim y.$$

If one of the first two occurs, then we conclude that $(a_1, \cdots, a_k)$ has a common eigenvector x, as required. Thus, we are left with the case $y_j \sim y$ for every $3 \leq j < k$. But this implies that (a_1, a_2, a_2a_j, a_k) has a common eigenvector y, which is thus shared also by a_j. Thus, y is a common eigenvector of $(a_1, \cdots, a_k)$, as desired.

Now, suppose that $a_1 = \pm\mathbf{1}$, and hence $a_2 \neq \pm\mathbf{1}$. Given any $3 \leq j < k$, we observe that $(a_2, a_1^{-1}a_2a_j, a_k)$ is in $(1,2)$-position and hence has a common eigenvector, say y_j. Since $a_2 \neq \pm\mathbf{1}$, we must have

$$x \sim y, \quad x \sim y_j, \quad \text{or} \quad y_j \sim y.$$

If one of the first two occurs, then we conclude that $(a_1, \cdots, a_k)$ has a common eigenvector x, as required. Thus, we are left with the case $y_j \sim y$ for every $3 \leq$

$j < k$. But this implies that $(a_1, a_2, a_1^{-1}a_2a_j, a_k)$ has a common eigenvector y, which is thus shared also by a_j. Thus, y is a common eigenvector of $(a_1, \cdots, a_k)$, as desired.

(2) $k = 4, 6, \cdots, 2g$. First, consider the case where $a_{k-1} = \pm\mathbf{1}$. It then suffices to show that the sequence $(a_1, \cdots, a_{k-2}, a_k)$ has a common eigenvector. This can be shown by repeating the argument the previous case (1) above. Thus, we may assume that $a_{k-1} \neq \pm\mathbf{1}$. Note that, for each integer m with $2m < k$, by our claim we have:

- a common eigenvector w_m of (a_{k-1}, a_k, a_{2m}), and
- a common eigenvector z_m of (a_{k-1}, a_k, a_{2m-1}).

Since $a_{k-1} \neq \pm\mathbf{1}$, we must have

$$x \sim w_m, \quad x \sim z_m, \quad \text{or} \quad w_m \sim z_m.$$

If one of the first two occurs, then conclude that $(a_1, \cdots, a_k)$ has a common eigenvector x, and we are done. Thus, we are left with the case where $w_m \sim z_m$ for every $2m < k$. Note in this case that w_m is a common eigenvector of $(a_{2m-1}, a_{2m}, a_{k-1}, a_k)$. Comparing the vectors x, w_m, and $w_{m'}$ for different m, m', we are in turn reduced to the case $w_m \sim w_{m'}$ for all m, m', in which case $(a_1, \cdots, a_k)$ has a common eigenvector w_1, as desired.

Thus, $(a_1, \cdots, a_k)$ has a common eigenvector. This completes the induction, and shows that $(a_1, \cdots, a_{2g+n})$ has a common eigenvector, proving the proposition. □

Remark. A corollary of the above proof is that, fixing a optimal set of generators of $\pi_1\Sigma$, we can (in principle) give an explicit finite set of equations for the locus of reducible characters on $X(\Sigma)$, in the case where the genus of Σ is at least 1.

We recall the following lemma of Selberg.

Lemma (Selberg)**.** *Let K be a field of characteristic zero. Given a finitely generated subgroup $\pi \leq \mathrm{GL}_n(K)$, there exists a torsion-free finite index subgroup $\pi' \leq \pi$.*

Corollary 6.13. *Let Σ be a surface, and let $\rho : \pi_1\Sigma \to \mathrm{SL}_2(\mathbb{C})$ be Zariski dense representation. There exists a finite covering $f : \Sigma' \to \Sigma$ such that the pullback*

$$\rho' = \rho \circ f_* : \pi_1 M' \to \mathrm{SL}_2(\mathbb{C})$$

has Zariski dense image which is moreover torsion-free.

Proof. Applying Selberg's lemma, there exists a finite-index subgroup π' of $\rho(\pi_1\Sigma)$ which is torsion-free. Since $\rho(\pi_1\Sigma)$ is Zariski dense in $\mathrm{SL}_2(\mathbb{C})$ by hypothesis, π' is also Zariski dense in $\mathrm{SL}_2(\mathbb{C})$ by Lemma 6.2. The subgroup $\rho^{-1}(\pi') \leq \pi_1\Sigma$ has finite index, and hence there is a corresponding finite covering $f : \Sigma' \to \Sigma$ with $\pi_1\Sigma' \simeq \rho^{-1}(\pi')$. The representation $\rho' = \rho \circ f_*$ satisfies the desired properties. □

Let Σ be a surface of type (g, n) with $g \geq 1$. We restate and prove the Zariski density part of the Main Lemma.

Proposition 6.14. *Let Σ be a surface of type (g, n) with $g \geq 1$. A representation $\rho : \pi_1\Sigma \to \mathrm{SL}_2(\mathbb{C})$ is Zariski dense if and only if there is an embedded one holed torus $\tau \subset \Sigma$ such that the restriction $\rho|\tau$ is Zariski dense.*

Proof. The *if* direction is clear. Let $\rho : \pi_1\Sigma \to \mathrm{SL}_2(\mathbb{C})$ be a representation with Zariski dense image. By Corollary 6.13, there exists a finite covering $f : \Sigma'' \to \Sigma$ such that $\rho'' = \rho \circ f_* : \pi_1\Sigma'' \to \mathrm{SL}_2(\mathbb{C})$ has Zariski dense image which is moreover torsion-free. By Proposition 6.12, there is an embedding $i : \Sigma' \to \Sigma''$ of a one holed torus such that $i^*(\rho'')$ is irreducible. Since $\mathrm{Im}\, i^*(\rho'') \leq \mathrm{Im}\, \rho''$ is torsion-free, we conclude by Corollary 6.1 that $i^*(\rho'')$ has Zariski dense image.

The composition $f \circ i : \Sigma' \to \Sigma$ presents Σ' as a finite covering of the image

$$\tau = f \circ i(\Sigma').$$

The induced map on the quotient topological spaces $\Sigma'/\partial\Sigma' \to \tau/\partial\tau$ is a branched covering of closed surfaces possibly ramified only at one point (corresponding to the boundary component of Σ'). By equipping the quotients suitably with complex structures and using the Riemann-Hurwitz formula, we conclude that $\tau/\partial\tau$ has genus one, and hence τ is a one holed torus. Writing $j : \tau \to \Sigma$ for the immersion, we have $f^*j^*(\rho) = i^*f^*(\rho)$ which is Zariski dense, and hence we conclude that $j^*(\rho)$ is Zariski dense, as desired. □

Combining Corollary 6.11 and the Main Lemma, we obtain the following. Let Σ be a compact oriented surface of genus $g > 0$ with n boundary curves.

Proposition 6.15. *If a representation $\rho : \pi_1\Sigma \to \mathrm{SL}_2(\mathbb{C})$ is Zariski dense, then its mapping class group orbit in $X(\Sigma)$ is infinite.*

Proof. Let $\rho : \pi_1\Sigma \to \mathrm{SL}_2(\mathbb{C})$ be Zariski dense. By the Main Lemma, there exists an embedded one-holed torus $\Sigma' \subset \Sigma$ such that the restriction $\rho|\Sigma'$ is Zariski dense. By Corollary 6.11, the $\Gamma(\Sigma')$-orbit of $\rho|\Sigma'$ in $X(\Sigma')$ must be infinite. *A fortiori*, the $\Gamma(\Sigma)$-orbit of ρ in $X(\Sigma)$ is infinite. □

We also record the following consequence of the above analysis.

Corollary 6.16. *Let Σ be a surface of type (g, n) with $g \geq 1$. The locus of Zariski dense representations in $X(\Sigma)$ is a Zariski open set.*

Proof. Let us write $X(\Sigma)_{\mathrm{Zar}} \subset X(\Sigma)$ for the locus of Zariski dense representations. Proposition 6.14 shows that $X(\Sigma)_{\mathrm{Zar}} = \bigcup_j (j^*)^{-1}(X(\Sigma')_{\mathrm{Zar}})$ where $j : \Sigma' \to \Sigma$ runs over all embeddings of one holed tori into Σ. Thus, we are reduced to proving the result for a one holed torus.

Suppose that Σ is a one holed torus. It suffices to show that, for $G = B$, D_∞, BA_4, BS_4, or BA_5, the locus of characters of representations ρ with $\rho(\pi_1\Sigma) \subseteq G$ (up to conjugation) is Zariski closed in $X(\Sigma)$. This is clear if G is one of the finite groups mentioned, as the locus consists of finite points.

We have $(\mathrm{tr}_a, \mathrm{tr}_b, \mathrm{tr}_{ab}) : X(\Sigma) \simeq \mathbb{A}^3$. Under this identification, the locus in question corresponding to $G = B$ is the closed subset of $\mathbb{A}^3$ defined by the equation

$$\mathrm{tr}_a^2 + \mathrm{tr}_{ab}^2 + \mathrm{tr}_{ab}^2 - \mathrm{tr}_a \,\mathrm{tr}_b \,\mathrm{tr}_{ab} - 2 = 2.$$

On the other hand, for $G = D_\infty$, the locus in question is the union of three lines $\{\mathrm{tr}_a = \mathrm{tr}_b = 0\}$, $\{\mathrm{tr}_a = \mathrm{tr}_{ab} = 0\}$, and $\{\mathrm{tr}_{ab} = \mathrm{tr}_a = 0\}$, and hence is closed. This proves the corollary. □

6.4 Sparse representations

Let Σ be a surface of type (g, n) with $g \geq 1$. Let X denote the character variety of Σ. In this section, we prove Theorems 6.A and 6.B on the locus of points on X corresponding to representations with non-Zariski-dense image. On the locus of finite representations, the claim is clear. In view of Lemma 6.4, it only remains to understand the finite mapping class group orbits on

$$\mathrm{Hom}(\pi_1\Sigma, \mathbb{C}^*) \quad \text{and} \quad \mathrm{Hom}(\pi_1\Sigma, D_\infty)/D_\infty.$$

Lemma 6.17. *Assume Σ is of type (g, n) with $g \geq 1$. A representation $\rho : \pi_1\Sigma \to \mathbb{C}^*$ has finite mapping class group orbit in $\mathrm{Hom}(\pi_1\Sigma, \mathbb{C}^*)$ if and only if it has finite image.*

Proof. The *if* direction is clear, and we now prove the converse. Let $\rho : \pi_1\Sigma \to \mathbb{C}^*$ be a representation with finite mapping class group orbit in $\mathrm{Hom}(\pi_1\Sigma, \mathbb{C}^*)$. Let $(a_1, b_1, \cdots, a_g, b_g, c_1, \cdots, c_n)$ be the optimal generators of $\pi_1\Sigma$. Assume first that

$n = 0$ or 1. We have a short exact sequence

$$0 \to \mathrm{Hom}(\pi_1\Sigma, \mathbb{Z}) \to \mathrm{Hom}(\pi_1\Sigma, \mathbb{C}) \xrightarrow{\exp(2\pi i \cdot)} \mathrm{Hom}(\pi_1\Sigma, \mathbb{C}^*) \to 0$$

which is $\Gamma(\Sigma)$-equivariant. The mapping class group action on $\mathrm{Hom}(\pi_1\Sigma, \mathbb{Z})$ and $\mathrm{Hom}(\pi_1\Sigma, \mathbb{C}) \simeq \mathrm{Hom}(\pi_1\Sigma, \mathbb{Z}) \otimes \mathbb{C}$ is linear, factoring through the quotient $\Gamma(\Sigma) \twoheadrightarrow \mathrm{Sp}(2g, \mathbb{Z})$ where $\mathrm{Sp}(2g, \mathbb{Z})$ is the integral symplectic group. Let us choose a lift $\tilde{\rho} \in \mathrm{Hom}(\pi_1\Sigma, \mathbb{C})$ of ρ. Let $\gamma \in \mathrm{Sp}(2g, \mathbb{Z})$ be an element such that none of its eigenvalues are roots of unity. By hypothesis, there exists $k \geq 1$ with $\gamma^k \cdot \rho = \rho$, or in other words

$$\gamma^k \cdot \tilde{\rho} \equiv \tilde{\rho} \bmod \mathrm{Hom}(\pi_1\Sigma, \mathbb{Z}).$$

Setting $\rho' = \gamma^k \cdot \tilde{\rho} - \tilde{\rho} \in \mathrm{Hom}(\pi_1\Sigma, \mathbb{Z})$, we see that $\tilde{\rho} = (\gamma^k - \mathbf{1})^{-1}\rho' \in \mathrm{Hom}(\pi_1\Sigma, \mathbb{Q})$ since the integral matrix $\gamma^k - \mathbf{1}$ has nonzero integral determinant and is therefore invertible over $\mathbb{Q}$. In particular, $\exp(2\pi i\tilde{\rho}) = \rho$ must be torsion, showing that the image of ρ is finite if $n \leq 1$, as desired. Thus, only the case $n \geq 2$ remains. Since $(a_1, b_1, \cdots, a_g, b_g)$ is in $(g, 1)$-position, the above analysis shows that $\rho(a_i)$ and $\rho(b_i)$ are roots of unity for $i = 1, \cdots, g$. Similarly, we see that the sequence

$$L_i = (a_1, b_1, \cdots, a_{g-1}, b_{g-1}, a_g, b_g c_i)$$

is in $(g, 1)$-position for every $i = 1, \cdots, n$. Since ρ restricted to the surface $\Sigma(L_i)$ must have finite $\Gamma(\Sigma(L_i))$-orbit, $\rho(c_i)$ is a root of unity for $i = 1, \cdots, n$ as well. This shows that ρ has finite image, as desired. □

Lemma 6.18. *Assume Σ has genus $g \geq 2$. A representation $\rho : \pi_1\Sigma \to D_\infty$ has finite mapping class group orbit in $\mathrm{Hom}(\pi_1\Sigma, D_\infty)/D_\infty$ if and only if it has finite image.*

Proof. The *if* direction is clear, and we now prove the converse. Let $\rho : \pi_1\Sigma \to D_\infty$ be a representation with finite mapping class group orbit in $\mathrm{Hom}(\pi_1\Sigma, D_\infty)/D_\infty$. We

have a short exact sequence $0 \to \mathbb{C}^* \to D_\infty \to \mathbb{Z}/2\mathbb{Z} \to 0$, where the morphism $D_\infty \to \mathbb{Z}/2\mathbb{Z}$ is given by

$$\begin{bmatrix} c & 0 \\ 0 & c^{-1} \end{bmatrix} \mapsto 0 \quad \text{and} \quad \begin{bmatrix} 0 & c \\ -c^{-1} & 0 \end{bmatrix} \mapsto 1 \quad \text{for all } c \in \mathbb{C}^*.$$

This give us a $\Gamma(\Sigma)$-equivariant map

$$\mathrm{Hom}(\pi_1\Sigma, D_\infty)/D_\infty \to \mathrm{Hom}(\pi_1\Sigma, \mathbb{Z}/2\mathbb{Z}).$$

The fibre of this map above the zero homomorphism consists of those points given by diagonal representations, to which Lemma 6.17 applies, noting that the map $\mathrm{Hom}(\pi_1\Sigma, \mathbb{C}^*) \to \mathrm{Hom}(\pi_1\Sigma, D_\infty)/D_\infty$ has finite fibers. It suffices to consider the case where $\rho : \pi_1\Sigma \to D_\infty$ is not in the fiber over the zero homomorphism. Let

$$(a_1, b_1, \cdots, a_g, b_g, c_1, \cdots, c_n)$$

be the optimal generators of $\pi_1\Sigma$. Note that $L = (a_1, b_1, \cdots, a_g, b_g)$ is in $(g, 1)$-position, and we have a $\Gamma(\Sigma(L))$-equivariant morphism

$$\mathrm{Hom}(\pi_1\Sigma, \mathbb{Z}/2\mathbb{Z}) \to \mathrm{Hom}(\pi_1\Sigma(L), \mathbb{Z}/2\mathbb{Z}).$$

The action of $\Gamma(\Sigma(L))$ on $\mathrm{Hom}(\pi_1\Sigma(L), \mathbb{Z}/2\mathbb{Z})$ factors through the projection

$$\Gamma(\Sigma(L)) \twoheadrightarrow \mathrm{Sp}(2g, \mathbb{Z}/2\mathbb{Z}).$$

From the transitivity of $\mathrm{Sp}(2g, \mathbb{Z}/2\mathbb{Z})$ on $(\mathbb{Z}/2\mathbb{Z})^{2g}$ away from the origin, it follows

that, up to $\Gamma(\Sigma)$-action, we may assume that

$$\bar{\rho}(a_1) = 1 \quad \text{and} \quad \bar{\rho}(b_1) = \bar{\rho}(a_2) = \cdots = \bar{\rho}(b_g) = 0$$

where $\bar{\rho}$ is the image of ρ in $\mathrm{Hom}(\pi_1\Sigma, \mathbb{Z}/2\mathbb{Z})$. Up to conjugation by an element in D_∞, we may moreover assume that

$$\rho(a_1) = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

It suffices to show that the entries of the matrices $\rho(b_1)$, $\rho(a_2)$, $\rho(b_2)$, $\cdots$, $\rho(a_g)$, $\rho(b_g)$, $\rho(c_1)$, $\cdots$, $\rho(c_n)$ are roots of unity. Let $i_1 < \cdots < i_q$ be precisely the indices in $\{1, \cdots, n\}$ such that $\bar{\rho}(c_{i_j}) = 0$. Since

$$L' = (a_2, b_2, \cdots, a_g, b_g, c_{i_1}, \cdots, c_{i_q}, b_1)$$

is in $(g-1, q+2)$-position and the restriction of ρ to $\Sigma(L')$ is diagonal, we see that $\rho(b_1)$, $\rho(a_2)$, $\rho(b_2)$, $\cdots$, $\rho(a_g)$, $\rho(b_g)$, $\cdots$, $\rho(c_{i_1})$, $\cdots$, $\rho(c_{i_q})$ are torsion by Lemma 6.17. Let us now take $i \in \{1, \cdots, n\} \setminus \{i_1, \cdots, i_q\}$. The restriction of ρ to the subsurface $\Sigma(c_i a_1, b_1)$ of genus 1 with 1 boundary curve is diagonal. Writing

$$\rho(c_i) = \begin{bmatrix} 0 & \lambda_i \\ -\lambda^{-1} & 0 \end{bmatrix}$$

with $\lambda_i \in \mathbb{C}^*$, we have

$$\rho(c_i a_i) = \begin{bmatrix} 0 & \lambda_i \\ -\lambda_i^{-1} & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} -\lambda_i & 0 \\ 0 & -\lambda_i^{-1} \end{bmatrix}.$$

It follows from Lemma 6.17 that λ_i is a root of unity. This completes the proof that

the entries of $\rho(b_1)$, $\rho(a_2)$, $\rho(b_2)$, $\cdots$, $\rho(a_g)$, $\rho(b_g)$, $\rho(c_1)$, $\cdots$, $\rho(c_n)$ are roots of unity, and hence the image of ρ is finite, as desired. □

Definition 6.19. A representation $\rho : \pi_1\Sigma \to D_\infty$ is *special dihedral* if:

(1) it is irreducible, and

(2) there is a nonseparating curve a in Σ such that $\rho|(\Sigma \setminus a)$ is diagonal.

Lemma 6.20. *Let Σ be a surface of genus 1 with $n \geq 0$ boundary curves. Let $\rho : \pi_1\Sigma \to D_\infty$ be special dihedral. Given a pair of loops (a, b) in $(1, 1)$-position on Σ, at least one of $\rho(a)$ and $\rho(b)$ is not diagonal.*

Proof. Let $\rho \in \pi_1\Sigma \to D_\infty$ be special dihedral. Assume toward contradiction that (a_1, b_1) is a pair of loops in $(1, 1)$-position on Σ with both $\rho(a_1)$ and $\rho(b_1)$ diagonal. We can complete (a_1, b_1) to a sequence of optimal generators $(a_1, b_1, c_1, \cdots, c_n)$ of $\pi_1\Sigma$. Since ρ is special dihedral, the matrices $\rho(c_1), \cdots, \rho(c_n)$ must be diagonal (noting that diagonality of a matrix is not changed under conjugation by an element in D_∞). This implies that ρ is in fact diagonal, contradicting the hypothesis that ρ is irreducible. □

Lemma 6.21. *Let Σ be a surface of genus 1 with $n \geq 0$ boundary curves. A representation $\rho : \pi_1\Sigma \to D_\infty$ has finite mapping class group orbit in $\mathrm{Hom}(\pi_1\Sigma, D_\infty)/D_\infty$ if and only if it is finite or special dihedral.*

Proof. The same argument as in Lemma 6.18 shows that if $\rho : \pi_1\Sigma \to D_\infty$ has finite $\Gamma(\Sigma)$-orbit in $\mathrm{Hom}(\pi_1\Sigma, D_\infty)/D_\infty$ then ρ is finite or special dihedral. Let now $\rho : \pi_1\Sigma \to D_\infty$ be a special dihedral representation. We shall show that ρ has finite mapping class group orbit in $\mathrm{Hom}(\pi_1\Sigma, D_\infty)/D_\infty$, or equivalently in the character variety $X(\Sigma)$.

The coordinate ring $R(\Sigma)$ of the character variety is finitely generated by the trace functions tr_a for a finite collection of essential curves a in Σ and the boundary curves,

in view of Fact 2.1 and the property of optimal sequence of generators. Therefore, it suffices to show that the set

$$\{\operatorname{tr}\rho(a) : a \subset \Sigma \text{ essential curve}\} \subseteq \mathbb{C}$$

is finite. Let $(a_1, b_1, \cdots, a_g, b_g, c_1, \cdots, c_n)$ be optimal generators of $\pi_1\Sigma$. Since ρ is special dihedral, it follows that $\rho(c_1)$, $\cdots$, $\rho(c_n)$ are diagonal matrices. Suppose a is a separating simple closed curve in Σ underlying a loop in the free homotopy class of $c_{i_1} \cdots c_{i_k}$ for some integers $i_1 < \cdots < i_k$ in $\{1, \cdots, n\}$. Since $\rho(c_i)$ are diagonal for $i = 1, \cdots, n$ it follows that $\operatorname{tr}\rho(a)$ lies in a finite set that only depends on the traces $\operatorname{tr}\rho(c_1)$, $\cdots$, $\operatorname{tr}\rho(c_n)$. But now, every separating simple closed curve in Σ is sent to one of the above form by some mapping class group element. Since the mapping class group action preserves the special dihedral representations, and since it fixes the traces $\operatorname{tr}\rho(c_1)$, $\cdots$, $\operatorname{tr}\rho(c_n)$, we conclude that $\{\operatorname{tr}\rho(a) : a \subset \Sigma \text{ separating curve}\}$ is finite.

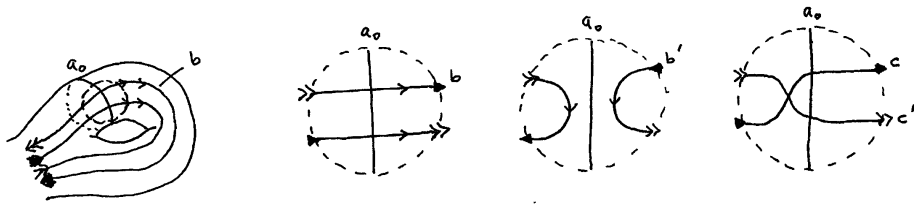


Figure 6.4: Building new loops out of old ones

It remains to show that $\{\operatorname{tr}\rho(a) : a \subset \Sigma \text{ nonseparating curve}\}$ is finite. Let a_0 be a nonseparating curve in Σ such that the restriction $\rho|(\Sigma \setminus a_0)$ is diagonal; such a curve exists since ρ is special dihedral. Let b be a nonseparating curve in Σ. Up to isotopy, we may assume that b intersects a only finitely many times. If b does not intersect a_0, then $b \subset \Sigma \setminus a_0$ and hence $\operatorname{tr}\rho(b)$ can take only finitely many values as $\rho|(\Sigma \setminus a_0)$ is diagonal.

If b intersects a exactly once, then we must have $\operatorname{tr}\rho(b) = 0$ by Lemma 6.20. Let

us now assume that b intersects a_0 more than once. Let us choose a parametrization of b. Since b is nonseparating, there must be two neighboring points of intersection of a_0 and b where the two segments of b have the same orientation, as in Figure 6.4. The operations as in Figure 6.4 produce for us a new simple closed curve b' which is also nonseparating, as well as a pair (c, c') of simple loops in $(1,1)$-position on Σ. We have the trace relation

$$\operatorname{tr}\rho(b) = \operatorname{tr}\rho(c)\operatorname{tr}\rho(c') - \operatorname{tr}\rho(b').$$

By Lemma 6.20, we have $\operatorname{tr}\rho(c)\operatorname{tr}\rho(c') = 0$, and therefore $\operatorname{tr}\rho(b) = -\operatorname{tr}\rho(b')$. Note furthermore that b' intersects a_0 in a smaller number of points than b does. Applying induction on the number of intersection points, we thus conclude that

$$\{\operatorname{tr}\rho(a) : a \subset \Sigma \text{ nonseparating curve}\}$$

is finite. This completes the proof that special dihedral representations have finite mapping class group orbits in X. □

The above lemmas combine together to give us the following result.

Proposition 6.22. *Let Σ be a surface of type (g,n) with $g \geq 1$. Let $\rho : \pi_1\Sigma \to \mathrm{SL}_2(\mathbb{C})$ be a semisimple representation whose image is not Zariski dense in $\mathrm{SL}_2(\mathbb{C})$. Then ρ has a finite mapping class group orbit in $X(\Sigma)$ if and only if one of the following holds:*

(1) *ρ is a finite representation, or*

(2) *$g = 1$ and ρ is special dihedral up to conjugation.*

6.4.1 Proof of Theorems 6.A and 6.B

Theorems 6.A and 6.B immediately follow by combining Propositions 6.15 and 6.22.

Bibliography

[1] Baker, Alan. *Transcendental number theory.* Second edition. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1990. x+165 pp. ISBN: 0-521-39791-X

[2] Balacheff, Florent; Parlier, Hugo; Sabourau, Stéphane *Short loop decompositions of surfaces and the geometry of Jacobians.* Geom. Funct. Anal. 22 (2012), no. 1, 37-73.

[3] Baragar, Arthur. *Integral solutions of Markoff-Hurwitz equations.* J. Number Theory 49 (1994), no. 1, 27-44.

[4] Bers, Lipman. *An inequality for Riemann surfaces.* Differential geometry and complex analysis, 87-93, Springer, Berlin, 1985.

[5] Biswas, Indranil; Koberda, Thomas; Mj, Mahan; Santharoubane, Ramanujan. *Representations of surface groups with finite mapping class group orbits.* preprint. arXiv:1702.03622

[6] Boalch, Phillip. *Geometry of moduli spaces of meromorphic connections on curves, Stokes data, wild nonabelian Hodge theory, hyperkahler manifolds, isomonodromic deformations, Painleve equations, and relations to Lie theory.* Habilitation memoir, Paris-Sud (Orsay) 12/12/12. arXiv:1305.6593

[7] Borel, Armand. *Commensurability classes and volumes of hyperbolic 3-manifolds.* Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) 8 (1981), no. 1, 1-33.

[8] Borel, Armand; Harish-Chandra. *Arithmetic subgroups of algebraic groups.* Ann. of Math. (2) 75 1962 485-535.

[9] Borevich, A. I.; Shafarevich, I. R. *Number theory.* Translated from the Russian by Newcomb Greenleaf. Pure and Applied Mathematics, Vol. 20 Academic Press, New York-London 1966 x+435 pp.

[10] Bourgain, Jean; Gamburd, Alexander; Sarnak, Peter *Markoff triples and strong approximation.* (English, French summary) C. R. Math. Acad. Sci. Paris 354 (2016), no. 2, 131-135.

[11] Bourgain, Jean; Gamburd, Alexander; Sarnak, Peter. *Markoff Surfaces and Strong Approximation: 1.* preprint. arXiv:1607.01530

[12] Bowditch, B. H.; Maclachlan, C.; Reid, A. W. *Arithmetic hyperbolic surface bundles.* Math. Ann. 302 (1995), no. 1, 31-60.

[13] Bruns, Winfried; Herzog, Jürgen. *Cohen-Macaulay rings.* Cambridge Studies in Advanced Mathematics, 39. Cambridge University Press, Cambridge, 1993. xii+403 pp. ISBN: 0-521-41068-1

[14] Chabauty, Claude. *Sur les points rationnels des courbes algébriques de genre supérieur à l'unité.* C. R. Acad. Sci. Paris 212, (1941). 882-885.

[15] Charles, Laurent; Marché, Julien *Multicurves and regular functions on the representation variety of a surface in SU(2).* Comment. Math. Helv. 87 (2012), no. 2, 409-431.

[16] Cohn, Harvey. *Approach to Markoff's minimal forms through modular functions.* Ann. of Math. (2) 61, (1955). 1-12.

[17] Cooper, D.; Culler, M.; Gillet, H.; Long, D. D.; Shalen, P. B. *Plane curves associated to character varieties of 3-manifolds.* Invent. Math. 118 (1994), no. 1, 47-84.

[18] Corlette, Kevin. *Archimedean superrigidity and hyperbolic geometry.* Ann. of Math. (2) 135 (1992), no. 1, 165-182.

[19] Corlette, Kevin. *Harmonic maps, rigidity, and Hodge theory.* Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994), 465-471, Birkhuser, Basel, 1995.

[20] Corti, Alessio; Kaloghiros, Anne-Sophie. *The Sarkisov program for Mori fibred Calabi-Yau pairs.* Algebr. Geom. 3 (2016), no. 3, 370-384.

[21] Corvaja, Pietro; Zannier, Umberto. *A subspace theorem approach to integral points on curves.* C. R. Math. Acad. Sci. Paris 334 (2002), no. 4, 267-271.

[22] Demazure, Michel *Anneaux gradués normaux.* Introduction la thorie des singularités, II, 35-68, Travaux en Cours, 37, Hermann, Paris, 1988.

[23] Deroin, Bertrand; Tholozan, Nicolas. *Dominating surface group representations by Fuchsian ones.* Int. Math. Res. Not. IMRN 2016, no. 13, 4145-4166.

[24] do Carmo, Manfredo Perdigão(BR-IMPA) *Riemannian geometry.* Translated from the second Portuguese edition by Francis Flaherty. Mathematics: Theory & Applications. Birkhuser Boston, Inc., Boston, MA, 1992. xiv+300 pp. ISBN: 0-8176-3490-8

[25] Donaldson, S. K. *Twisted harmonic maps and the self-duality equations.* Proc. London Math. Soc. (3) 55 (1987), no. 1, 127-131.

[26] Drensky, Vesselin; Formanek, Edward. *Polynomial identity rings.* Advanced Courses in Mathematics. CRM Barcelona. Birkhäuser Verlag, Basel, 2004. viii+200 pp. ISBN: 3-7643-7126-9

[27] Dubrovin, B.; Mazzocco, M. *Monodromy of certain Painlev-VI transcendents and reflection groups.* Invent. Math. 141 (2000), no. 1, 55-147.

[28] Duke, W.; Rudnick, Z.; Sarnak, P. *Density of integer points on affine homogeneous varieties.* Duke Math. J. 71 (1993), no. 1, 143-179.

[29] Eells, James, Jr.; Sampson, J. H. *Harmonic mappings of Riemannian manifolds.* Amer. J. Math. 86 1964 109-160.

[30] Faltings, Gerd. *Diophantine approximation on abelian varieties.* Ann. of Math. (2) 133 (1991), no. 3, 549-576.

[31] Farb, Benson; Margalit, Dan. *A primer on mapping class groups.* Princeton Mathematical Series, 49. Princeton University Press, Princeton, NJ, 2012.

[32] Fermat, Pierre. *Oeuvres.*

[33] Fock, Vladimir; Goncharov, Alexander. *Moduli spaces of local systems and higher Teichmüller theory.* Publ. Math. Inst. Hautes tudes Sci. No. 103 (2006), 1-211.

[34] Franke, Jens; Manin, Yuri I.; Tschinkel, Yuri. *Rational points of bounded height on Fano varieties.* Invent. Math. 95 (1989), no. 2, 421-435.

[35] Gamburd, A.; Magee, M; Ronan, R. *An asymptotic for Markoff-Hurwitz tuples.* Preprint. arXiv: 1603.06267

[36] Ghosh, A.; Sarnak, P. *Integral points on Markoff type cubic surfaces.* Preprint. arXiv: 1706.06712

[37] Goldman, William M. *Ergodic theory on moduli spaces.* Ann. of Math. (2) 146 (1997), no. 3, 475-507.

[38] Goldman, William M. *Trace coordinates on Fricke spaces of some simple hyperbolic surfaces.* Handbook of Teichmller theory. Vol. II, 611-684, IRMA Lect. Math. Theor. Phys., 13, Eur. Math. Soc., Zrich, 2009.

[39] Goldman, William M. *Mapping class group dynamics on surface group representations.* Problems on mapping class groups and related topics, 189-214, Proc. Sympos. Pure Math., 74, Amer. Math. Soc., Providence, RI, 2006.

[40] Goldman, William M. *The modular group action on real SL(2)-characters of a one-holed torus.* Geom. Topol. 7 (2003), 443-486.

[41] Goldman, William M.; Xia, Eugene Z. *Ergodicity of mapping class group actions on SU(2)-character varieties.* Geometry, rigidity, and group actions, 591-608, Chicago Lectures in Math., Univ. Chicago Press, Chicago, IL, 2011.

[42] Gromov, Mikhael *Filling Riemannian manifolds.* J. Differential Geom. 18 (1983), no. 1, 1-147.

[43] Gromov, Mikhail; Schoen, Richard. *Harmonic maps into singular spaces and p-adic superrigidity for lattices in groups of rank one.* Inst. Hautes Études Sci. Publ. Math. No. 76 (1992), 165-246.

[44] Gross, Mark; Hacking, Paul; Keel, Sean. *Birational geometry of cluster algebras.* Algebr. Geom. 2 (2015), no. 2, 137-175.

[45] Gross, Mark; Hacking, Paul; Keel, Sean; Kontsevich, Maxim. *Canonical bases for cluster algebras.* preprint. arXiv:1411.1394

[46] Hamilton, Richard S. *Harmonic maps of manifolds with boundary.* Lecture Notes in Mathematics, Vol. 471. Springer-Verlag, Berlin-New York, 1975. i+168 pp.

[47] Harpaz, Yonatan. *Geometry and arithmetic of certain log K3 surfaces.* preprint. arXiv: 1511.01285

[48] Hindry, Marc; Silverman, Joseph H. *Diophantine geometry. An introduction.* Graduate Texts in Mathematics, 201. Springer-Verlag, New York, 2000. xiv+558 pp.

[49] Hitchin, N. J. *Poncelet polygons and the Painlevé equations.* Geometry and analysis (Bombay, 1992), 151-185, Tata Inst. Fund. Res., Bombay, 1995.

[50] Hochster, Melvin; Roberts, Joel L. *Rings of invariants of reductive groups acting on regular rings are Cohen-Macaulay.* Advances in Math. 13 (1974), 115-175.

[51] Horowitz, Robert D. *Characters of free groups represented in the two-dimensional special linear group.* Comm. Pure Appl. Math. 25 (1972), 635-649.

[52] A. Hurwitz. *Über eine Aufgabe der unbestimmten Analysis* Archiv. Math. Phys. , 3 (1907) pp. 185-196

[53] Keel, Sean; McKernan, James *Rational curves on quasi-projective surfaces.* Mem. Amer. Math. Soc. 140 (1999), no. 669, viii+153 pp.

[54] Kollár, János. *Conic bundles that are not birational to numerical Calabi-Yau pairs.* preprint, 2016.

[55] Kollár, János; Xu, Chenyang. *The dual complex of Calabi-Yau pairs.* Invent. Math. 205 (2016), no. 3, 527-557.

[56] Komyo, Arata. *On compactifications of character varieties of n-punctured projective line.* Ann. Inst. Fourier (Grenoble) 65 (2015), no. 4, 1493-1523.

[57] Le Bruyn, Lieven. *The functional equation for Poincaré series of trace rings of generic* 2×2 *matrices.* Israel J. Math. 52 (1985), no. 4, 355-360.

[58] Lee, Carl W. *The associahedron and triangulations of the n-gon.* European J. Combin. 10 (1989), no. 6, 551-560.

[59] Lisovyy, Oleg; Tykhyy, Yuriy. *Algebraic solutions of the sixth Painlevé equation.* J. Geom. Phys. 85 (2014), 124-163.

[60] Long, D. D.; Reid, A. W. *Integral points on character varieties.* Math. Ann. 325 (2003), no. 2, 299-321.

[61] Lubotzky, Alexande. *Counting finite index subgroups.* Groups '93 Galway/St. Andrews, Vol. 2, 368-404, London Math. Soc. Lecture Note Ser., 212, Cambridge Univ. Press, Cambridge, 1995.

[62] Manon, Christopher. *Compactifications of character varieties and skein relations on conformal blocks.* Geom. Dedicata 179 (2015), 335-376.

[63] Manon, Christopher. *Toric geometry of* $\mathrm{SL}_2(\mathbb{C})$ *free group character varieties from outer space.* Canad. J. Math. 70 (2018), no. 2, 354-399.

[64] Margulis, G. A. *Discrete subgroups of semisimple Lie groups.* Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], 17. Springer-Verlag, Berlin, 1991.

[65] Markoff, A. *Sur les formes quadratiques binaires indéfinies.* Math. Ann. 17 (1880), no. 3, 379-399.

[66] Matsumura, Hideyuki. *Commutative ring theory.* Translated from the Japanese by M. Reid. Cambridge Studies in Advanced Mathematics, 8. Cambridge University Press, Cambridge, 1986. xiv+320 pp. ISBN: 0-521-25916-9

[67] McKean, H.P. *Selberg's trace formula as applied to a compact Riemann surface.* Comm. Pure Appl. Math. 25 (1972), 225-246.

[68] Mordell, L. J. *On the integer solutions of the equation* $x^2 + y^2 + z^2 + 2xyz = n$. J. London Math. Soc. 28, (1953). 500-510.

[69] Moser, Leo; Wyman, Max. *On solutions of* $x^d = 1$ *in symmetric groups.* Canad. J. Math. 7, (1955). 159-168.

[70] Mumford, David. *A remark on Mahler's compactness theorem.* Proc. Amer. Math. Soc. 28 (1971), 289-294.

[71] Newman, Morris. *Asymptotic formulas related to free products of cyclic groups.* Math. Comp. 30 (1976), no. 136, 838-846.

[72] Previte, Joseph P.; Xia, Eugene Z. *Topological dynamics on moduli spaces. II.* Trans. Amer. Math. Soc. 354 (2002), no. 6, 2475-2494.

[73] Procesi, C. *The invariant theory of* $n \times n$ *matrices.* Advances in Math. 19 (1976), no. 3, 306-381.

[74] Przytycki, Józef H.; Sikora, Adam S. *On skein algebras and* $\mathrm{SL}_2(\mathbf{C})$*-character varieties.* Topology 39 (2000), no. 1, 115-148.

[75] Reznikov, Alexander G. *Harmonic maps, hyperbolic cohomology and higher Milnor inequalities.* Topology 32 (1993), no. 4, 899-907.

[76] Saito, Kyoji. *Character variety of representations of a finitely generated group in SL2.* Topology and Teichmüller spaces (Katinkulta, 1995), 253-264, World Sci. Publ., River Edge, NJ, 1996.

[77] Sampson, J. H. *Some properties and applications of harmonic mappings.* Ann. Sci. École Norm. Sup. (4) 11 (1978), no. 2, 211-228.

[78] Sarnak, Peter C. *Diophantine problems and linear groups.* Proceedings of the International Congress of Mathematicians, Vol. I, II (Kyoto, 1990), 459-471, Math. Soc. Japan, Tokyo, 1991.

[79] Serre, Jean-Pierre. *Trees.* Translated from the French original by John Stillwell. Corrected 2nd printing of the 1980 English translation. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2003.

[80] Silverman, Joseph H. *The Markoff equation $X^2+Y^2+Z^2 = aXYZ$ over quadratic imaginary fields.* J. Number Theory 35 (1990), no. 1, 72-104.

[81] Sinz, Justin C. *Some results in representation dynamics.* Ph.D. Thesis. The University of Chicago, 2010.

[82] Simpson, Carlos. *The dual boundary complex of the SL2 character variety of a punctured sphere.* Ann. Fac. Sci. Toulouse Math. (6) 25 (2016), no. 2-3, 317-361.

[83] Siu, Yum Tong. *The complex-analyticity of harmonic maps and the strong rigidity of compact Khler manifolds.* Ann. of Math. (2) 112 (1980), no. 1, 73-111.

[84] Stanley, Richard P. *Hilbert functions of graded algebras.* Advances in Math. 28 (1978), no. 1, 57-83.

[85] Sterk, Hans. *Finiteness results for algebraic K3 surfaces.* Math. Z. 189 (1985), no. 4, 507-513.

[86] Takeuchi, Kisao. *A characterization of arithmetic Fuchsian groups.* J. Math. Soc. Japan 27 (1975), no. 4, 600-612.

[87] Takeuchi, Kisao. *Arithmetic Fuchsian groups with signature* $(1; e)$. J. Math. Soc. Japan 35 (1983), no. 3, 381-407.

[88] Thue, Axel. *Über Annäherungswerte algebraischer Zahlen.* (German) J. Reine Angew. Math. 135 (1909), 284-305.

[89] van der Put, Marius; Singer, Michael F. *Galois theory of linear differential equations.* Grundlehren der Mathematischen Wissenschaften [Fundamental Principles

of Mathematical Sciences], 328. Springer-Verlag, Berlin, 2003. xviii+438 pp. ISBN: 3-540-44228-6

[90] Vogt, H. *Sur les invariants fondamentaux des équations différentielles linéaires du second ordre.* (French) Ann. Sci. cole Norm. Sup. (3) 6 (1889), 3-71.

[91] Vojta, Paul. *Diophantine approximations and value distribution theory.* Lecture Notes in Mathematics, 1239. Springer-Verlag, Berlin, 1987. x+132 pp. ISBN: 3-540-17551-2

[92] Vojta, Paul. *Siegel's theorem in the compact case.* Ann. of Math. (2) 133 (1991), no. 3, 509-548.

[93] Vojta, Paul. *Integral points on subvarieties of semiabelian varieties. I.* Invent. Math. 126 (1996), no. 1, 133-181.

[94] Watanabe, Keiichi. *Some remarks concerning Demazure's construction of normal graded rings.* Nagoya Math. J. 83 (1981), 203-211.

[95] Whang, Junho Peter. *Simple loops on punctured surfaces and boundaries of character varieties.* preprint. arXiv:1612.02518

[96] Whang, Junho Peter. *Nonlinear descent on moduli of local systems.* Preprint. arXiv: 1710.01848

[97] Whang, Junho Peter. *Periodic points on character varieties.* Preprint. arXiv: 1707.00071

[98] Whang, Junho Peter. *Arithmetic of curve on moduli of local systems.* Preprint. arXiv: 1803.04583

[99] Zannier, Umberto. *Lecture notes on Diophantine analysis.* With an appendix by Francesco Amoroso. Lecture Notes. Scuola Normale Superiore di Pisa (New Series), 8. Edizioni della Normale, Pisa, 2009. xvi+237 pp.

[100] Zannier U. (eds) *On Some Applications of Diophantine Approximations*. Publications of the Scuola Normale Superiore, vol 2. Edizioni della Normale, Pisa.