GEOMETRY ON SURFACES AND HIGGS BUNDLES

by **Peter B. Gothen***,**

ABSTRACT.—There are three complete plane geometries of constant curvature: spherical, Euclidean and hyperbolic geometry. We explain how a closed oriented surface can carry a geometry which locally looks like one of these. Focussing on the hyperbolic case we describe how to obtain all hyperbolic structures on a given topological surface, and how to parametrise them. Finally we introduce Higgs bundles and explain how they relate to hyperbolic surfaces.

1 INTRODUCTION

The idea of considering geometry on a surface is well known to inhabitants of Planet Earth. Indeed, as any explorer knows, spherical geometry is appropriate. In this geometry distance is measured along arcs of great circles. These are the *geodesics* of spherical geometry, just like the geodesics of plane Euclidean geometry are segments of straight lines.

The spherical surface and the Euclidean plane are both *complete*, meaning that any geodesic can be extended indefinitely. Moreover they both have constant curvature, positive in the case of the sphere, and zero in the case of the plane. There is also a complete 2-dimensional geometry of constant negative curvature, namely the hyperbolic plane (which we shall introduce below).

The sphere is an example of a *closed surface*, i.e., a surface which is compact and has no boundary (as opposed to a closed disk, for example). Topologically, closed orientable surfaces are classified by the genus g, a non-negative integer: a surface of genus g can be realised inside 3-space as a *g-holed torus* as illustrated in Figure 1. We have seen that the genus zero surface supports spherical geometry but what about the other surfaces? Our first main goal in this article is to explain how the torus (genus one) supports a complete geometry which locally looks like the Euclidean plane, while a surface of genus $g \geq 2$ can be given a complete locally hyperbolic geometry. This involves considering certain special subgroups of the matrix group $\mathrm{SL}(2,\mathbb{R})$. We shall then see how the algebra and geometry of the matrix group $\mathrm{SL}(2,\mathbb{R})$ interact in interesting ways, and how this sheds light on the question of which subgroups give rise to hyperbolic surfaces.

Our second main goal is to explain how considering the set of all possible hyperbolic structures on a fixed topological surface of genus $g \geq 2$ leads to interesting and beautiful mathematics. Thus we introduce moduli spaces and explore some of their properties.

Finally, we shall give an introduction to Higgs bundles. We shall show how they can be used to shed new light on the theme of hyperbolic structures on surfaces and indicate their role in recent generalisations of some of the results explained earlier in the article.

The paper is mostly expository, only the final Section 8.4 includes some results in which the author has been involved.

For reasons of space the references are by no means complete, but we hope the interested reader will be able to use them as a starting point for further exploration.

* **CMUP e Departamento de Matemática, Faculdade de Ciências da Universidade do Porto.**
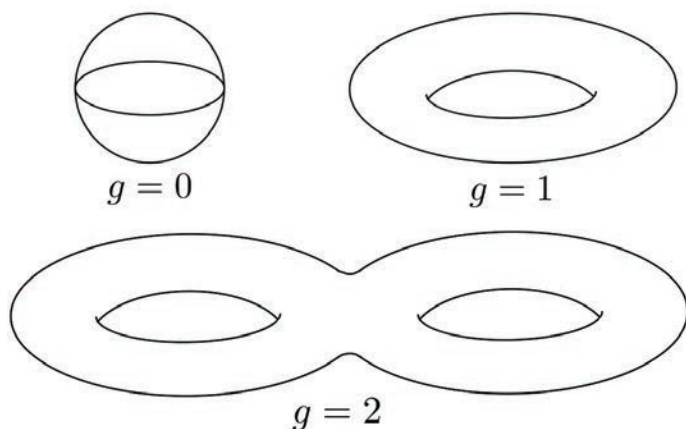   **Email: pbgothen@fc.up.pt**
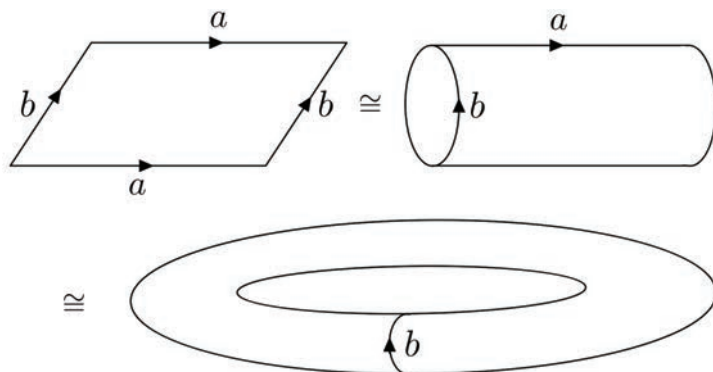
Figure 1.— Genus of a surface

Figure 2.—A Euclidean surface

2 EUCLIDEAN SURFACES

We want to explain how to do Euclidean geometry on a closed surface, in a way which makes the generalisation to the hyperbolic case natural.

2.1 THE EUCLIDEAN PLANE

Using Cartesian coordinates we identify the Euclidean plane $\mathbb{E}^2$ with the coordinate plane $\mathbb{R}^2$. Distance is determined by calculating the length of a parametrised curve $\alpha : [a, b] \to \mathbb{R}^2$ in the usual way: $l(\alpha) = \int_a^b |\alpha'(t)| dt$. This is usually expressed by saying that in Cartesian coordinates $(x, y) \in \mathbb{R}^2$ on $\mathbb{E}^2$ the Euclidean element of arc length ds is given by

$$ds^2 = dx^2 + dy^2. \tag{1}$$

Moreover, the distance preserving transformations form a group, $\mathrm{Isom}(\mathbb{E}^2)$, which is called the isometry group of $\mathbb{E}^2$. An example of isometries are *translations*. Using coordinates $\mathbb{E}^2 \cong \mathbb{R}^2$, the translation $A : \mathbb{E}^2 \to \mathbb{E}^2$ by the vector $\mathbf{a} \in \mathbb{R}^2$ can be written

$$A(P) = P + \mathbf{a}.$$

2.2 EUCLIDEAN SURFACES

The Euclidean plane $\mathbb{E}^2$ is obviously not a closed surface. However, we can build a closed surface by taking a parallellogram in the plane and gluing its oppo-site sides, as illustrated in Figure 2: sides labeled with the same letter are to be glued with the orientation indicated by the arrows. We can carry out this process in 3-space — in a non-distance preserving way! — to convince ourselves that the resulting surface is really a topological torus.

More formally, we take linearly independent vectors $\mathbf{a}$ and $\mathbf{b}$ generating the sides labeled a and b. Let A and B be the translations by the vectors $\mathbf{a}$ and $\mathbf{b}$, respectively, and consider the subgroup

$$\Gamma = \langle A, B \rangle \subseteq \mathrm{Isom}(\mathbb{E}^2)$$

generated by them inside the isometry group of $\mathbb{E}^2$. This is just the group of translations by vectors of the form $n\mathbf{a} + m\mathbf{b}$, where $m, n \in \mathbb{Z}$. Since translations commute, the generators of Γ satisfy the single relation

$$[A, B] = I,$$

where $[A, B] := ABA^{-1}B^{-1}$ is the *commutator* and I is the identity. The *orbit space*

$$\mathbb{E}^2/\Gamma$$

is obtained identifying points $P, Q \in \mathbb{E}^2$ if there is a $\gamma \in \Gamma$ such that $Q = \gamma(P)$. Its points correspond to *orbits* $\Gamma \cdot Q = \{\gamma(Q) \mid \gamma \in \Gamma\}$. Thus each point of the interior of the parallellogram generated by $\mathbf{a}$ and $\mathbf{b}$ corresponds to a unique point of $\mathbb{E}^2/\Gamma$, and pairs of points on opposite sides are identified via the corresponding translation, thus realising the desired gluing. Hence $\mathbb{E}^2/\Gamma$ is a locally Euclidean surface, which is topologically a torus.
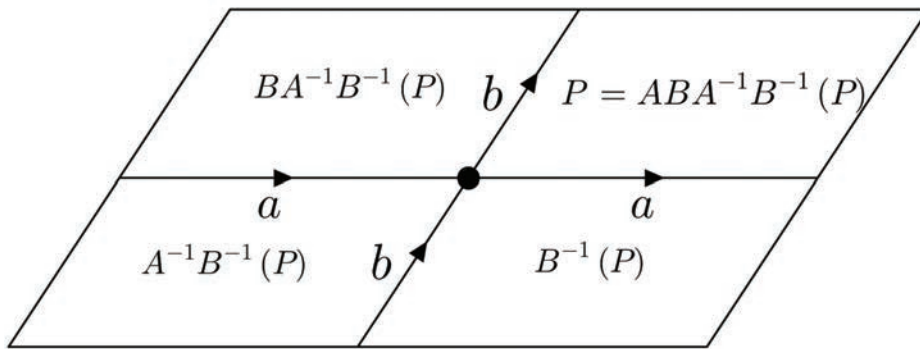
Figure 3.—Four copies of *P*

As illustrated in Figure 3 the four vertices of the parallelogram *P* get identified in $\mathbb{E}^2/\Gamma$. Moreover, there are four copies of the parallelogram meeting there, which fit together because of the relation $[A, B] = I$; the Euclidean metric is not distorted because the sum of the internal angles of the parallelogram is exactly 2π.

From a more abstract point of view, a key point is that the group Γ has the following property: for every point *P* in $\mathbb{E}^2$, there is an open neighbourhood *U* such that $\gamma(U) \cap U = \varnothing$ for all γ different from the identity. We say the action of Γ on $\mathbb{E}^2$ is *properly discontinuous*. This property ensures that for each $Q \in U$ its orbit $\Gamma \cdot Q = \{\gamma(Q) \mid \gamma \in \Gamma\}$ has a unique representative (namely *Q* itself) in *U*, so that *U* works as a *coordinate patch* for $\mathbb{E}/\Gamma$ around $\Gamma \cdot P$. This, together with the fact that the elements of Γ are isometries, means that $\mathbb{E}^2/\Gamma$ has a well defined distance function: indeed, the arc length of a parametrised curve in $\mathbb{E}^2/\Gamma$ can be calculated using the formula (1) which is invariant under isometries.

Note that we can also construct non-compact surfaces in this way. For example, if we take Γ to be the subgroup generated by a single translation, we obtain a cylinder. This is a locally Euclidean surface which, unlike the Euclidean torus, can be easily visualised in 3-space by rolling up a sheet of paper.

The so-called *Killing–Hopf Theorem* implies that any complete connected locally Euclidean surface can be represented as $\mathbb{E}^2/\Gamma$, where Γ acts freely and properly discontinuously on $\mathbb{E}^2$ (see, for example, Still-well [16]).

3 HYPERBOLIC SURFACES

3.1 THE HYPERBOLIC PLANE

We start by describing the *hyperbolic plane* $\mathbb{H}^2$. Hyperbolic geometry is different from spherical and Euclidean geometry in that it is not possible to embed (smoothly) $\mathbb{H}^2$ in Euclidean 3-space in a distance preserving way. Instead we consider the *upper half plane model*, defined by

$$\mathbb{H}^2 = \{z = x + iy \in \mathbb{C} \mid y > 0\}$$

with element of arc length

$$ds^2 = \frac{dx^2 + dy^2}{y^2}.$$

In the model $\mathbb{H}^2$ geodesics are open arcs of semi-circles orthogonal to the real axis $\mathbb{R} = \{y = 0\} \subset \mathbb{C}$ together with open half-lines orthogonal to $\mathbb{R}$. Note that the hyperbolic plane is complete, so these curves do in fact have infinite hyperbolic length. Moreover, orientation preserving isometries can be represented by *Möbius transformations*

$$z \mapsto A \cdot z = \frac{az + b}{cz + d},$$

where

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{R})$$

is a real 2×2-matrix of determinant one. As examples we can take

$$A = \begin{pmatrix} \rho & 0 \\ 0 & \rho \end{pmatrix}$$

which gives a hyperbolic translation whose axis is the imaginary axis in $\mathbb{H}^2$, and

$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

which gives a hyperbolic rotation about $i \in \mathbb{H}^2$ by the angle 2θ.

We note that A and $-A$ define the same Möbius transformation, so the group of orientation preserving isometries is really the quotient group $\mathrm{PSL}(2, \mathbb{R}) = \mathrm{SL}(2, \mathbb{R})/\{\pm I\}$. We shall mostly ignore this distinction in what follows but it will become relevant in Section 7 below.

We finish this section by commenting on the topology of $\mathrm{SL}(2, \mathbb{R})$. Identifying the set of all 2×2-matrices

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

with $\mathbb{R}^4$, the group $\mathrm{SL}(2, \mathbb{R})$ is the subset cut out by the equation $ac - bd = 1$. Thus it has a topology inherited from $\mathbb{R}^4$. In fact the Implicit Function Theorem applied to this equation shows that $\mathrm{SL}(2, \mathbb{R})$ is a 3-dimensional *Lie group*, meaning that it can be covered by local coordinate systems in 3-space and that the group operations are differentiable in these coordinates.

3.2 Hyperbolic surfaces

As we shall see, a closed orientable topological surface of genus g can be given a hyperbolic structure for any $g \geq 2$. In the case of $g = 2$, take an octagon with gluing instructions to create a surface as illustrated in Figure 4. If we cut the octagon along the diameter indicated, we see that indeed the resulting surface has genus 2, as desired.

In order to get a nice *hyperbolic* surface, the octagon should be taken in the hyperbolic plane (it will look very different from that of 4). And, in a manner

analogous to the Euclidean case, we require that pairs of sides which are to be glued have the same length. Moreover, the vertices of the octagon all get identified to one point in the surface, so the internal angles should add up to 2π. This condition sounds strange to our Euclidean wired brains but, it is a fact that such an octagon exists.[1]

In order to write the surface as $\mathbb{H}^2/\Gamma$ for a suitable subgroup $\Gamma \subset \mathrm{SL}(2, \mathbb{R})$ we take hyperbolic translations A_i and B_i giving the required identifications, and let Γ be the group generated by these translations. The octagon then becomes a fundamental domain for the action of Γ. The condition that the interior angles add up to 2π is equivalent to the identity

$$[A_1, B_1][A_2, B_2] = I$$

in $\mathrm{SL}(2, \mathbb{R})$. In general, we let Γ_g be the abstract group

$$\Gamma_g = \langle a_1, b_1, \ldots, a_g, b_g \mid \prod_{i=1}^{g} [a_i, b_i] = 1 \rangle.$$

This group is known as a *surface group*.[2] In view of the genus 2 example it is hopefully not a surprise that genus g surfaces can be obtained from subgroups of $\mathrm{SL}(2, \mathbb{R})$ isomorphic to Γ_g. In order to study all such subgroups we consider homomorphisms $\rho : \Gamma_g \to \mathrm{SL}(2, \mathbb{R})$ (often also called *representations*). We say that ρ is *Fuchsian* if it is injective and its image is discrete, i.e., consists of isolated points.[3] When ρ is Fuchsian it can be proved that the action of Γ_g on $\mathbb{H}^2$ is properly discontinuous. Hence the orbit space

$$S_\rho := \mathbb{H}^2/\rho(\Gamma_g),$$

is a nice hyperbolic surface of genus g with charts coming from $\mathbb{H}^2$. Conversely, the Killing–Hopf Theorem again tells us that any closed orientable hyperbolic surface is of this form.[4]

However, it is certainly not true that any homomorphism $\rho : \Gamma_g \to \mathrm{SL}(2, \mathbb{R})$ defines a closed hyperbolic surface: for example, the trivial homorphism clearly does not! This leaves us with the following

Question: Let $\rho : \Gamma_g \to \mathrm{SL}(2, \mathbb{R})$ be a representation. How can we tell if ρ defines a closed hyperbolic surface?

---

[1] In fact, in hyperbolic geometry the sum of the internal angles of a polygon depends on its area!

[2] The group Γ_g can be identified wth the fundamental group of a topological surface of genus g.

[3] Recall that topological notions make sense viewing $\mathrm{SL}(2, \mathbb{R}) \subseteq \mathbb{R}^4$.

[4] As already noted, we should really consider representations to $\mathrm{PSL}(2, \mathbb{R})$. However, it turns out that representations defining closed hyperbolic surfaces can always be lifted to $\mathrm{SL}(2, \mathbb{R})$.
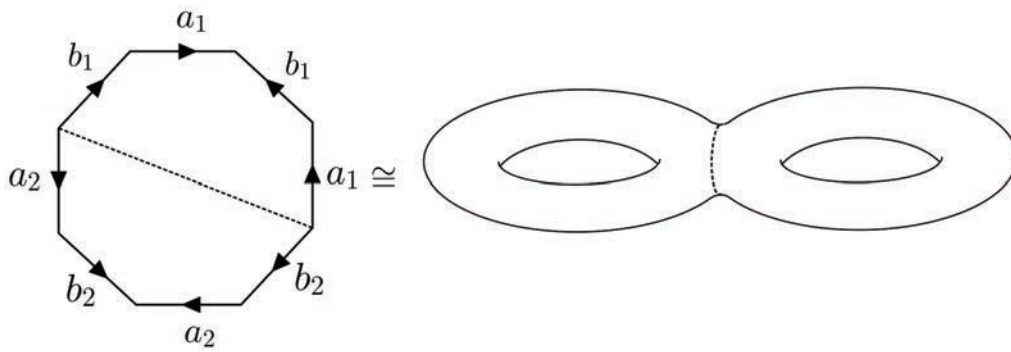
Figure 4.—Genus 2 surface from an octagon

4 Topology and algebra of $SL(2, \mathbb{R})$

In order to answer the question at the end of the last section we shall define an invariant of representations $\rho \colon \Gamma \to SL(2, \mathbb{R})$. For that we shall need to understand how the topology and algebra of $SL(2, \mathbb{R})$ interact.

The subgroup $SO(2) \subseteq SL(2, \mathbb{R})$ of rotation matrices

$$E(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

can be identified with a circle.

The map $E \colon \mathbb{R} \to SO(2)$, $\theta \mapsto E(\theta)$ wraps the real line around the circle, and it satisfies $E(0) = I$ and $E(\theta_1 + \theta_2) = E(\theta_1)E(\theta_2)$. In other words, E is a group homomorphism from the additive group $\mathbb{R}$ to $SO(2)$.

Now, thinking of $SO(2)$ inside $SL(2, \mathbb{R})$, we want to extend this picture and find a group $\widetilde{SL}(2, \mathbb{R})$ containing $\mathbb{R}$, with a surjective group homomorphism $p \colon \widetilde{SL}(2, \mathbb{R}) \to SL(2, \mathbb{R})$ which restricts to $E \colon \mathbb{R} \to SO(2)$, i.e., making the diagram

$$
\begin{array}{ccc}
\mathbb{R} & \longrightarrow & \widetilde{SL}(2, \mathbb{R}) \\
\downarrow{\scriptstyle E} & & \downarrow{\scriptstyle p} \\
SO(2) & \longrightarrow & SL(2, \mathbb{R})
\end{array}
$$

commutative (the horizontal maps are inclusions). In fact it follows from general theory that such a group exists and is essentially unique; it is known as the *universal covering group* of $SL(2, \mathbb{R})$. We shall explain how it can be constructed explicitly, following [13, §1.8], using the action of $SL(2, \mathbb{R})$ on the hyperbolic plane $\mathbb{H}^2$.

So let

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in SL(2, \mathbb{R}).$$

Write

$$j(A, z) = cz + d$$

for the denominator of $A \cdot z$. Note that

$$j(E(\theta), i) = i\sin\theta + \cos\theta = e^{i\theta},$$

which indicates that this function can used to keep track of the phase θ. For each fixed A, we can consider the holomorphic function

$$\mathbb{H}^2 \to \mathbb{C} \smallsetminus \{0\}$$
$$z \mapsto j(A, z) = cz + d.$$

Observe that $cz + d \neq 0$ for $z \in \mathbb{H}$. Therefore, since $\mathbb{H}^2$ is simply connected, there is a continuous determination of the logarithm of $j(A, z) = cz + d$, i.e., a continuous map $\phi \colon \mathbb{H}^2 \to \mathbb{C}$ such that

$$e^{\phi(z)} = cz + d.$$

We want to make the point that such a ϕ can be explicitly calculated: simply choose a value θ for the argument $\arg(ci + d)$, write $ci + d = re^{i\theta}$ and let $\phi(i) = \log(r) + i\theta$. Then

$$\phi(z) - \phi(i) = \int_\gamma \frac{dz}{z} = \int_0^1 \frac{c(z-i)dt}{c(i + t(z-i)) + d}$$

(here γ parametrises the segment joining $ci + d$ to $cz + d$). Note that ϕ is not unique, but it is uniquely determined by the choice of $\phi(i)$. Thus any two determinations ϕ differ by an integer multiple of $2\pi i$.

Now define $\widetilde{SL}(2, \mathbb{R})$ as the set of pairs (A, ϕ), where $A \in SL(2, \mathbb{R})$ and $\phi \colon \mathbb{H}^2 \to \mathbb{C}$ is any contin-

uous determination of the logarithm of $j(A, z)$. The product on $\widetilde{SL}(2, \mathbb{R})$ is defined by

$$(A_1, \phi_1) \cdot (A_2, \phi_2) = (A_1 A_2, \tilde{\phi}),$$

where

$$\tilde{\phi}(z) := \phi_1(A_2 \cdot z) + \phi_2(z).$$

It is an easy calculation to check that

$$j(A_1 A_2, z) = j(A_1, A_2 \cdot z) j(A_2, z)$$

which implies that indeed

$$e^{\tilde{\phi}(z)} = j(A_1 A_2, z)$$

as required. It is not hard to check that this defines a group structure on $\widetilde{SL}(2, \mathbb{R})$. For example, for $A = I$, the identity matrix, we can take $\phi(z) = 0$ and $(A, 0)$ is the neutral element. Moreover, $(A, \phi)^{-1} = (A^{-1}, \tilde{\phi})$, where

$$\tilde{\phi}(z) = -\phi(A^{-1} \cdot z). \tag{2}$$

The projection $p: \widetilde{SL}(2, \mathbb{R}) \to SL(2, \mathbb{R})$ is of course just $(A, \phi) \mapsto A$. The inclusion $\mathbb{R} \hookrightarrow \widetilde{SL}(2, \mathbb{R})$ is given by $\theta \mapsto (E(\theta), \phi_\theta)$, where ϕ_θ is the determination of $\log(j(E(\theta), z))$ which satisfies $\phi_\theta(i) = i\theta$ (recall that $j(E(\theta), i) = e^{i\theta}$).

PROPOSITION 1.— The kernel of $p: \widetilde{SL}(2, \mathbb{R}) \to SL(2, \mathbb{R})$ consists of pairs (I, ϕ), where I is the identity matrix and ϕ is a constant function taking values in $2\pi\mathbb{Z} \subset \mathbb{R}$.

PROOF.— Clearly $p(A, \phi) = I$ if and only if $A = I$. Moreover, $j(I, z) = 1$, so ϕ is a determination of the logarithm of the constant function $z \mapsto 1 \in C$, i.e., it is a constant $\phi \in 2\pi\mathbb{Z} \subset \mathbb{R}$. ∎

5   THE TOLEDO INVARIANT

Let $\rho: \Gamma \to SL(2, \mathbb{R})$ be a representation. We shall associate an integer invariant to ρ. This invariant is known as the Toledo invariant, even though it was actually introduced by Milnor [14], and sometimes is referred to as the Euler number. Write

$$A_i = \rho(a_i), \quad B_i = \rho(b_i)$$

for $i = 1, \dots, g$. Choose *lifts* $\tilde{A}_i$ and $\tilde{B}_i$ in $\widetilde{SL}(2, \mathbb{R})$ such that $p(\tilde{A}_i) = A_i$ and $p(\tilde{B}_i) = B_i$, and define the *Toledo invariant* of ρ to be

$$\tau(\rho) = \frac{1}{\pi} \prod_{i=1}^{g} [\tilde{A}_i, \tilde{B}_i].$$

In view of the relation defining Γ_g, the product $\prod_{i=1}^{g}[\tilde{A}_i, \tilde{B}_i]$ is in the kernel of p. Hence Proposition 1 shows that the Toledo invariant is an even integer.[5]

From the description of $\widetilde{SL}(2, \mathbb{R})$ of the preceding section, it is easy to check that the Toledo invariant is well defined, i.e., that it does not depend on the choice of lifts: the main point is that the ambiguity in the choice of ϕ is canceled by (2), because each lift occurs together with its inverse in the commutator. Moreover, the Toledo invariant of a representation defined by matrices A_i and B_i *can be explicitly calculated*.

6   GOLDMAN'S THEOREM

A celebrated inequality due to Milnor [14] states that

$$|\tau(\rho)| \leq 2g - 2$$

for every representation $\rho: \Gamma_g \to SL(2, \mathbb{R})$. The following beautiful result shows that representations with maximal Toledo invariant (known as *maximal representations*) have a special geometric significance.

THEOREM 2 (GOLDMAN [7]).—
A representation $\rho: \Gamma_g \to SL(2, \mathbb{R})$ is Fuchsian if and only if $|\tau(\rho)| = 2g - 2$.

REMARK 1.— One might wonder about the significance of the sign of the Toledo invariant. If we conjugate a representation ρ by the outer automorphism of $SL(2, \mathbb{R})$ given by conjugation by a reflection we obtain a representation $\bar{\rho}$ with $\tau(\bar{\rho}) = -\tau(\rho)$. In fact, the hyperbolic surface $S_{\bar{\rho}}$ is obtained from S_ρ by a change of orientation, i.e., by composing all charts with a reflection in $\mathbb{H}^2$.

7   THE MODULI SPACE OF REPRESENTATIONS

Let us now take a global view and consider all representations of Γ_g in $SL(2, \mathbb{R})$ simultaneously. The *representation space* for representations of Γ_g in $SL(2, \mathbb{R})$ is the set of homomorphisms $\mathrm{Hom}(\Gamma_g, SL(2, \mathbb{R}))$. It is natural to consider ρ_1 and ρ_2 equivalent if they differ by overall conjugation by an element of $SL(2, \mathbb{R})$, corresponding to a change of basis in $\mathbb{R}^2$. It also turns

[5]   Odd Toledo invariants arise from representations $\rho: \Gamma_g \to PSL(2, \mathbb{R})$ which do not lift to $SL(2, \mathbb{R})$.

out that two hyperbolic structures on the same topological surface are isometric by an isometry which can be continuously deformed to the identity if and only if the corresponding Fuchsian representations are equivalent in this sense. Thus the *moduli space* of representations is defined to be the orbit space

$$\mathscr{R}(\Gamma_g, \mathrm{SL}(2,\mathbb{R})) = \mathrm{Hom}(\Gamma_g, \mathrm{SL}(2,\mathbb{R}))/\mathrm{SL}(2,\mathbb{R})$$

under the conjugation action.[6]

A homomorphism $\rho \colon \Gamma_g \to \mathrm{SL}(2,\mathbb{R})$ is determined by $2g$ matrices

$$A_i = \rho(a_i), \ B_i = \rho(b_i), \quad i = 1, \dots, g$$

satisfying the single relation $\prod[A_i, B_i] = I$. Hence $\mathrm{Hom}(\Gamma_g, \mathrm{SL}(2,\mathbb{R}))$ can be identified with the subspace of $\mathbb{R}^{6g}$ cut out by the 3 scalar equations given by $\prod[A_i, B_i] = I$ (the equation takes values in the 3-dimensional group $\mathrm{SL}(2,\mathbb{R})$). It follows that it is a variety of dimension $6g - 3$. The conjugation action by $\mathrm{SL}(2,\mathbb{R})$ reduces the dimension by 3, and so the moduli space has dimension

$$\dim \mathscr{R}(\Gamma_g, \mathrm{SL}(2,\mathbb{R})) = 6g - 6.$$

The Toledo invariant separates the moduli space into subspaces

$$\mathscr{R}_d \subseteq \mathscr{R}(\Gamma_g, \mathrm{SL}(2,\mathbb{R}))$$

corresponding to representations with invariant d. Goldman [8] showed that the $\mathscr{R}_d$ are in fact connected components of the moduli space, except in the maximal case $|d| = 2g - 2$. It turns out that $\mathscr{R}_{2g-2}$ has 2^{2g} connected components. However, these components get identified after projecting onto

$$\mathscr{R}(\Gamma_g, \mathrm{PSL}(2,\mathbb{R})),$$

which thus has just one connected component with Toledo invariant $2g-2$. This is not surprising because, by Goldman's Theorem 2, the subspace $\mathscr{R}_{2g-2}$ is exactly the locus of Fuchsian representations and, moreover, any two Fuchsian representations into $\mathrm{SL}(2,\mathbb{R})$ define the same hyperbolic surface if and only if they coincide after projecting to $\mathrm{PSL}(2,\mathbb{R})$. Accordingly, the corresponding connected component $\mathscr{T} \subseteq \mathscr{R}(\Gamma_g, \mathrm{PSL}(2,\mathbb{R}))$ is known as the *Fuchsian locus*. As we have seen, it parametrises all hyperbolic structures on the topological surface S_g up to a natural equivalence. It is a classical result that the space of such hyperbolic structures can be identified with $\mathbb{R}^{6g-6}$. In the next section we shall explain how a parametrisation of this space can be obtained using Higgs bundles.

8 Higgs bundles

We now describe how the results of the previous section can be understood using non-abelian Hodge theory, a subject founded by Hitchin [11] and Simpson [15].

8.1 Riemann surfaces and holomorphic bundles

A *Riemann surface X* is a topological surface together with a family of local charts which together cover the surface, and are such that changes of coordinates are holomorphic functions between open sets in $\mathbb{C}$. An example of this is the *Riemann sphere* $\hat{\mathbb{C}} = \mathbb{C} \cup \infty$: we use the standard coordinate z in $\mathbb{C}$ and around $\infty \in \hat{\mathbb{C}}$ we use the coordinate $w = 1/z$. Thus the change of coordinates $T \colon \mathbb{C} \smallsetminus \{0\} \to \mathbb{C} \smallsetminus \{0\}$ given by $w = T(z) = 1/z$ is holomorphic in the domain where both z and w are defined.

In particular, if we have a hyperbolic surface $S_g \cong \mathbb{H}^2/\Gamma$ for a Fuchsian representation of Γ, then the local coordinates in $\mathbb{H}^2$ give S_g the structure of a Riemann surface: indeed the changes of coordinates are Möbius transformations of $\mathbb{H}^2$, which are certainly holomorphic. We write X_ρ for the Riemann surface constructed from a Fuchsian representation ρ in this way.

Note that not all holomorphic maps define isometries of $\mathbb{H}^2$, so the concept of Riemann surface is less rigid than that of hyperbolic surface. However, the famous Uniformisation Theorem, due to Köbe and Poincaré, asserts that any Riemann surface can be represented as a hyperbolic surface. This means, in particular, that the space of all Riemann surfaces with the same underlying topological surface of genus g (up to a suitable equivalence) can be identified with the Fuchsian locus $\mathscr{T}$. When thought of in this way, it is known as *Teichmüller space*.

A *holomorphic line bundle $L \to X$* on a Riemann surface X is a holomorphic family of 1-dimensional complex vector spaces parametrised by X. Thus, for each $p \in X$ we have a 1-dimensional complex vector space L_p, which varies holomorphically with p. The simplest example is the *trivial bundle $L = X \times \mathbb{C} \to X$*; here the map is projection onto X and the fibre

---

[6] In order to get a Hausdorff quotient, one should in fact exclude representations whose action on $\mathbb{R}^2$ is not semisimple.

$L_p = \{p\} \times \mathbb{C}$ for $p \in X$ with its vector space structure coming from $\mathbb{C}$. Locally on X, a holomorphic line bundle is required to look like the product $U \times \mathbb{C}$, where $U \subseteq X$ is open. We say that L is trivialised over U. This means that a holomorphic line bundle can be given by an open covering $\{U_\alpha\}$ of X and *trivialisations*

$$L_{|U_\alpha} \cong U_\alpha \times \mathbb{C}$$

for each α. This gives rise to *transition functions*

$$g_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathbb{C} \smallsetminus \{0\}$$

which compare the isomorphisms $L_p \cong \mathbb{C}$ given by the trivialisations over U_α and U_β, respectively.

More important than the line bundles themselves are their *sections*. These are holomorphic maps $s : X \to L$ such that $s(p) \in L_p$ for all $p \in X$. A section of the trivial bundle $U \times \mathbb{C}$ over U is of course nothing but a map $s : U \to \mathbb{C}$, and if we have local trivialisations of a line bundle L as above, a holomorphic section s corresponds to a collection of holomorphic maps $s_\alpha : U_\alpha \to \mathbb{C}$ which glue correctly, i.e., satisfy the condition

$$s_\alpha(p) = g_{\alpha\beta}(p)s_\beta(p)$$

for $p \in U_\alpha \cap U_\beta$. As an illustrative example, we take the *canonical bundle* $K \to X$. Its sections are *holomorphic differentials*. In a local coordinate z on X a holomorphic differential, say α, can be written

$$g(z)dz$$

for a holomorphic function $g(z)$ and if $h(w)dw$ is the representation of α in another holomorphic coordinate $w = T(z)$, then

$$g(z)dz = h(T(z))d(T(z)) = h(T(z))T'(z)dz.$$

Thus a holomorphic differential can be represented by a collection of holomorphic functions locally defined on coordinate charts which transform according to the preceding rule. It turns out that the vector space of holomorphic differentials on a closed Riemann surface X of genus g, usually denoted by $H^0(X, K)$, is finite dimensional, of dimension $2g - 2$. More generally, the vector space $H^0(X, L)$ of holomorphic sections of any holomorphic line bundle $L \to X$ is finite dimensional. Any holomorphic line bundle has a topological invariant called its *degree*; in case L has a non-zero holomorphic section, this is the number of zeroes of such a section, counted with multiplicity. For example, the degree of the canonical bundle is $2g - 2$. The fact that this is the same as the dimension of $H^0(X, K)$ is a consequence of a fundamental result known as the Riemann–Roch formula.

We can perform the usual operations of linear algebra, like taking duals and tensor products, fibrewise on line bundles. Thus, if L and M are line bundles with transition functions $g_{\alpha\beta}$ and $h_{\alpha\beta}$, respectively, the tensor product $L \otimes M$ has transition functions $g_{\alpha\beta}h_{\alpha\beta}$ (pointwise multiplication in $\mathbb{C} \smallsetminus \{0\}$) and the dual bundle L^* has transition functions $g_{\alpha\beta}^{-1}$.

8.2 Higgs bundles

A $\mathrm{PSL}(2, \mathbb{R})$-*Higgs bundle* on X consists of three pieces of data

$$(L, \beta, \gamma)$$

where, $L \to X$ is a holomorphic line bundle, and $\beta \in H^0(X; K \otimes L)$ and $\gamma \in H^0(X; K \otimes L^*)$ can be seen as holomorphic differentials which take values in the line bundles L and L^*, respectively.

In a manner analogous to the conjugation action on representations, there is a natural notion of isomorphism of Higgs bundles, and the set of isomorphism classes of $\mathrm{PSL}(2, \mathbb{R})$-Higgs bundles forms the *moduli space* $\mathscr{M}(X, \mathrm{PSL}(2, \mathbb{R}))$. It is a complex algebraic variety of complex dimension $3g - 3$. We note that in order to get a reasonable moduli space it is necessary to restrict to so-called semistable Higgs bundles. This is analogous to the way in which one restricts to semisimple representations in the moduli space of representations.

The *Non-abelian Hodge Theorem* (due to Corlette, Donaldson, Hitchin and Simpson) for this situation states the following.

THEOREM 3.— There is a real analytic isomorphism

$$\mathscr{R}(\Gamma_g, \mathrm{PSL}(2, \mathbb{R})) \cong \mathscr{M}(X, \mathrm{PSL}(2, \mathbb{R})).$$

This is a remarkable theorem for many reasons. Here we just point out that while the character variety $\mathscr{R}$ is **real** and depends only on the topological surface of genus g (through its fundamental group), the moduli space $\mathscr{M}$ depends on the Riemann surface structure X given to the topological surface and has a **complex** structure.

For fixed d we denote by $\mathscr{M}_d$ the subspace of $\mathrm{PSL}(2, \mathbb{R})$-Higgs bundles (L, β, γ) with $\deg(L) = d$. Then we have $\mathscr{R}_d \cong \mathscr{M}_d$ under the non-abelian Hodge Theorem. In particular, the Fuchsian locus $\mathscr{T}$ corresponds to $\mathscr{M}_{2g-2}$.

8.3 Hitchin's parametrisation of $\mathcal{T}$

A particular class of PSL$(2, \mathbb{R})$-Higgs bundles can be obtained by taking $L = K$. Then γ is a section of the line bundle $K \otimes K^*$ which is naturally isomorphic to the trivial line bundle on X. In other words, γ is simply a holomorphic function on X, so we can set $\gamma = 1$ (the constant function). Moreover, β is a section of $K^2 = K \otimes K$. In other words it is a *quadratic differential*, so it can locally be written as $\beta(z) = b(z)(dz)^2$, where $b(z)$ satisfies an appropriate transformation rule under changes of coordinates. The vector space $H^0(X, K^2)$ of quadratic differentials on X has complex dimension $3g - 3$ which equals the dimension of the moduli space $\mathcal{M}(X, \text{PSL}(2, \mathbb{R}))$. This construction defines a map

$$\Psi : H^0(X, K^2) \to \mathcal{M}(X, \text{PSL}(2, \mathbb{R})),$$
$$\beta \mapsto (K, \beta, 1).$$

The semistability condition alluded to earlier implies that all Higgs bundles in $\mathcal{M}_{2g-2}$ arise in this way. Hence Ψ is an isomorphism onto its image $\mathcal{M}_{2g-2}$.

From the non-abelian Hodge Theorem we already knew that $\mathcal{M}_{2g-2} \cong \mathcal{T}$ is a connected component. But the Higgs bundle construction gives an alternative proof. Using gauge theoretic methods Hitchin also shows that $\mathcal{M}_{2g-2}$ parametrises all hyperbolic metrics on the topological surface underlying X. Moreover, under this parametrisation the Higgs bundle $(K, 1, 0)$ corresponds to the hyperbolic metric which uniformises X. Thus Hitchin's approach gives alternative proofs of Goldman's theorems and the Uniformisation Theorem.

8.4 The general Cayley correspondence

Hitchin [12] generalised the construction of the map Ψ to a map

$$\Psi : \bigoplus_i H^0(X, K^{d_i}) \to \mathcal{M}(X, G)$$

whose image is again a connected component of the *moduli space $\mathcal{M}(X, G)$ of G-Higgs bundles* for any simple split real Lie group G, nowadays known as a *Hitchin component*.[7] The domain of Ψ is a direct sum of spaces of higher holomorphic differentials on X; the integers d_i are determined by the Lie group G (in fact they are the exponents of its Lie algebra).

Similar constructions of special connected components have later been given for Hermitian groups G of non-compact tube type, such as SU(p, p) (see, for example, [5, 6, 2]). In this case the domain of the map Ψ turns out to be a moduli space $\mathcal{M}_{K^2}(X, G')$ of so-called K^2-twisted G'-Higgs bundles, for a certain real Lie group G' associated to G (known as its *Cayley partner*).

Both Hitchin components and Cayley components are special because they are not (as all other known components of the moduli space) detected by standard topological invariants of the underlying bundles and the Higgs fields satisfy a certain non-degeneracy condition.

Recently (see [1, 4] and the recent survey [3]) both of these constructions have been unified and generalised. The class of Lie groups G covered are characterised by the fact that their Lie algebras admit a *magical $\mathfrak{sl}_2$-triple*. This new Lie theoretic notion builds on ideas of Hitchin [12] and generalises that of a principal $\mathfrak{sl}_2$-triple introduced by Kostant. Conjecturally the generalised Cayley components obtained by this construction account for all "special" (in the sense of the previous paragraph) connected components of the moduli space and thus opens the door to a complete determination of this important topological invariant.

One important piece of supporting evidence for this conjecture comes from the area of *Higher Teichmüller Theory*. Higher Teichmüller theory has developed in parallel with the Higgs bundle story just described, and there has been a rich cross-fertilisation of ideas between the two areas. We cannot do justice to this fast-growing, rich and important area of mathematics here but refer the interested reader to [10, 17] and references therein. Very briefly, a higher Teichmüller space is a connected component of the moduli space of representations, which consists exclusively of discrete and injective representations, like the Fuchsian locus in the PSL$(2, \mathbb{R})$-case. It turns out that the generalised Cayley components are indeed higher Teichmüller spaces [4, 9], and it is expected that all higher Teichmüller spaces are thus obtained.

---

[7] In the case of classical matrix groups this means that G is one of the groups SL$(n, \mathbb{R})$, Sp$(2n, \mathbb{R})$, SO(p, p) and SO$(p, p + 1)$.

References

[1] M. Aparicio-Arroyo, S. Bradlow, B. Collier, O. García-Prada, P. B. Gothen, and A. Oliveira, *SO(p, q)-Higgs bundles and higher Teichmüller components*, Invent. Math. **218** (2019), no. 1, 197–299.

[2] O. Biquard, O. García-Prada, and R. Rubio, *Higgs bundles, the Toledo invariant and the Cayley correspondence*, J. Topol. **10** (2017), no. 3, 795–826.

[3] S. Bradlow, *Global properties of Higgs bundle moduli spaces*, 2023, to appear.

[4] S. Bradlow, B. Collier, O. García-Prada, P. B. Gothen, and A. Oliveira, *A general Cayley correspondence and higher Teichmüller spaces*, arXiv:2101.09377 [math.AG], 2021.

[5] S. B. Bradlow, O. García-Prada, and P. B. Gothen, *Surface group representations and U(p, q)-Higgs bundles*, J. Differential Geom. **64** (2003), 111–170.

[6] ———, *Maximal surface group representations in isometry groups of classical hermitian symmetric spaces*, Geometriae Dedicata **122** (2006), 185–213.

[7] W. M. Goldman, *Representations of fundamental groups of surfaces*, Springer LNM 1167, 1985, pp. 95–117.

[8] ———, *Topological components of spaces of representations*, Invent. Math. **93** (1988), 557–607.

[9] O. Guichard, F. Labourie, and A. Wienhard, *Positivity and representations of surface groups*, arXiv:2106.14584 [math.DG], 2021.

[10] O. Guichard and A. Wienhard, *Positivity and higher Teichmüller theory*, European Congress of Mathematics, Eur. Math. Soc., Zürich, 2018, pp. 289–310.

[11] N. J. Hitchin, *The self-duality equations on a Riemann surface*, Proc. London Math. Soc. (3) **55** (1987), 59–126.

[12] ———, *Lie groups and Teichmüller space*, Topology **31** (1992), 449–473.

[13] G. Lion and M. Vergne, *The Weil representation, Maslov index and theta series*, Progress in Mathematics, vol. 6, Birkhäuser, Boston, Mass., 1980.

[14] J. W. Milnor, *On the existence of a connection with curvature zero*, Comment. Math. Helv. **32** (1958), 216–223.

[15] C. T. Simpson, *Higgs bundles and local systems*, Inst. Hautes Études Sci. Publ. Math. **75** (1992), 5–95.

[16] J. Stillwell, *Geometry of surfaces*, Universitext, Springer-Verlag, New York, 1992. MR 1171453 (94b:53001)

[17] A. Wienhard, *An invitation to higher Teichmüller theory*, Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. II. Invited lectures, World Sci. Publ., Hackensack, NJ, 2018, pp. 1013–1039.