Geodesic and orthogeodesic identities on hyperbolic surfaces

Hugo Parlier[*]

Abstract. The lengths of geodesics on hyperbolic surfaces satisfy intriguing equations, known as identities, relating these lengths to geometric quantities of the surface. This paper is about a large family of identities that relate lengths of closed geodesics and orthogeodesics to boundary lengths or number of cusps. These include, as particular cases, identities due to Basmajian, to McShane and to Mirzakhani and Tan-Wong-Zhang. In stark contrast to previous identities, the identities presented here include the lengths taken among *all* closed geodesics.

1. Introduction

The relationship between lengths of closed geodesics, hyperbolic surfaces, and their underlying moduli spaces is intriguing and has been studied from many different perspectives. One exciting feature appeared in the work of Basmajian [1] and McShane [20] who discovered two different *identities* which relate infinite sums, where the terms depend on lengths of geodesics, to geometric quantities. In the case of Basmajian, the geometric quantity is the boundary length of the surface and in the original McShane identity, it is the length of a horocycle surrounding a cusp. Both of these identities have been generalized in multiple contexts, most famously perhaps being the generalization of the McShane identity by Mirzakhani who then showed how to use the identity to compute volumes of moduli spaces [21].

The exact statements of these original identities will be given below (and indeed scrutinized carefully), and although they seem at first glance to be unrelated, they share an aesthetic feature of simplicity. The Basmajian identity is a sum over all orthogeodesics of terms $\log(\coth(\ell/2))$ that depend naturally on the length ℓ of the orthogeodesic. The sum gives the boundary length. The original McShane identity is on a one cusped torus, and is a sum over all simple closed geodesics of terms $\frac{1}{e^\ell+1}$ depending only on the length ℓ of the simple closed geodesic. The sum is equal to 1/2. Their many successors have plenty of redeeming features, but are no match in terms of simplicity of the statement. One of the goals of the present paper is to show how these identities are related, and are in fact part of a family of identities that involve both lengths of closed geodesics and orthogeodesics.

The identities will be for orientable hyperbolic surfaces (of finite type) with non-empty boundary. Points in the moduli space $\mathcal{M}(\Sigma)$ of hyperbolic surfaces X homeomorphic to fixed given topological surface Σ with be denoted $\mathcal{M}(X)$. A hyperbolic surface X will be required to have either simple closed geodesics or cusps as boundary. Non-empty boundary means that $\pi_1(\Sigma)$ is a free group, and conjugacy classes of elements in $\pi_1(\Sigma)$ correspond to (free) homotopy classes of closed curves $\mathcal{C}(\Sigma)$, which it turn corresponds to $\mathcal{G}(X)$, the set of oriented closed geodesics of X. $\mathcal{G}(X)$ will be the set of oriented closed geodesics of X, and can thus be viewed from an algebraic, topological or geometric viewpoint. In particular, if you move X in its moduli space, then $\mathcal{G}(X)$ remains the same from the algebraic or topological viewpoint. Similarly, we use $\mathcal{O} = \mathcal{O}(X)$ for the set of all oriented orthogeodesics of X: these are geodesic segments with endpoints on and orthogonal to ∂X. If one of the endpoints is a cusp c, an orthogeodesic that leaves or returns to or from c is of infinite length so the term orthogonal refers to it being orthogonal to a (or any) horocyclic boundary of c.

The identities are all given by a choice of subset $M \subset \mathcal{C}(\Sigma)$ which is *coherent*, by which it is meant that M must satisfy a property which we now outline. An element in M is a closed curve, and if it is not simple and primitive, it might contain proper subloops. Being coherent means that if $\gamma \in M$, then none of the subloops of γ belong to M (see Section 2.1 for a more detailed definition). Note that orthogeodesics can also contain subloops, and if an orthogeodesic has a subloop freely homotopic to an element $\gamma \in \mathcal{C}(\Sigma)$, we say that it *supports* γ.

Some immediate examples of coherent markings are when M is empty or M is any single curve. Any coherent marking (finite or infinite) can be obtained as follows: take any curve γ and add it to M, now add any other curve that is not homotopic to a subloop of a curve in M and that does not contain an element of M as a subloop, and so forth. Note that there is no requirement that if $\gamma \in M$, then γ^{-1} also lie in M. A perhaps less obvious example: M is the set of all simple primitive closed geodesics (with both orientations). Or, alternatively, choose exactly one orientation for each simple closed geodesic, and let M be the set of those. More generally, any subset of a coherent marking is coherent, and using this fact it is not hard to convince oneself that for fixed Σ, there are uncountably many coherent markings. Some of them are mapping class group invariant, but in general they are not. Other examples of a coherent markings can be obtained by fixing a filling closed curve γ, and taking the full mapping class group orbit of γ (thus $M = \mathrm{Mod}_\Sigma(\gamma)$).

This allows, given a coherent marking M, for a separation of $\mathcal{O}$ into two sets: those that are supportive of elements of M, and those that are not. Note that, by definition, all simple orthogeodesics are unsupportive. Among the unsupportive elements of $\mathcal{O}$ (denoted $\mathcal{O}_M$), certain are *peripheral*, meaning that together with a boundary arc they form a homotopy class

2

in M. A more detailed definition is given in Section 2, but for now we denote the set of orthogeodesics peripheral to M by $\mathcal{O}_M$ and write $\partial\eta \ni \alpha$ if η is peripheral to α.

The abstract identity is as follows:

Theorem 1.1. *Let X be a hyperbolic surface with $\partial X \neq \emptyset$ consisting of cusps and/or simple closed geodesics. Let β be a boundary element, M a coherent marking of $\mathcal{C}(\Sigma)$ and let $\mathcal{O}\beta$ be the set of orthogeodesics leaving from β.*

If β is a closed geodesic, then:

$$\ell_X(\beta) = \sum_{\eta \in \mathcal{O}\beta_M} \phi(\eta) + \sum_{\alpha \in M} \left(\sum_{\eta \in \mathcal{O}\beta_M^\alpha} \psi(\eta) \right)$$

where $\phi(\eta)$ is a measure that depends only on the length of η, and $\psi(\eta)$ is a measure depending only the length of η and α (the curve η is peripheral to).

When β is a cusp:

$$2 = \sum_{\eta \in \mathcal{O}\beta_M} \phi_\circ(\eta) + \sum_{\alpha \in M} \left(\sum_{\eta \in \mathcal{O}\beta_M^\alpha} \psi_\circ(\eta) \right)$$

where $\phi_\circ(\eta)$ is a measure that depends only on the truncated length of η, and $\psi_\circ(\eta)$ depends only the length of doubly truncated length of η and α (the curve η is peripheral to).

The truncated length of an orthogeodesic leaving from a cusp is the restriction of the orthogeodesic to the subarc that leaves from the boundary of a standard horocyclic neighborhood of the cusp. For an orthogeodesic that leaves and returns to the same cusp, the doubly truncated length is the length of the subarc that leaves from the cusp neighborhood until it returns to the cusp neighborhood for the last time. (Note that truncated orthogeodesics might return multiple times to the the horocyclic neighborhood.) The functions (or measures) are often called gap functions or just gaps because they are measures of segments of the boundary associated to index terms. They are not only abstract, and can be quantified. Here is a first quantification.

Theorem 1.2. *Let $X \in \mathcal{M}(\Sigma)$ be a hyperbolic surface of finite type with $\partial X \neq \emptyset$ containing at least one simple closed geodesic, and M a coherent marking of $\mathcal{C}(\Sigma)$. The lengths of curves and orthogeodesics of X satisfy*

$$\ell(\partial X) = \sum_{\eta \in \mathcal{O}_M} \log\left(\coth^2(\eta/2)\right)$$

$$+ \sum_{\alpha \in M} \sum_{\eta \in \mathcal{O}_M^\alpha} \log\left(\frac{\cosh(\alpha/2) + \sqrt{\cosh^2(\alpha/2) + \cosh^2(\eta/2) - 1}}{\sinh(\alpha/2) + \sqrt{\cosh^2(\alpha/2) + \cosh^2(\eta/2) - 1}}\right).$$

3

Note that any $\eta \in \mathcal{O}_M$ appears twice: once in the first sum and once in the second. It would of course be possible to regroup these terms, but geometrically these terms describe different phenomena.

The above equation can be expressed in terms of half-traces (that is the hyperbolic cosine of half-lengths: $a = \cosh(\alpha/2)$, $b = \cosh(\beta/2)$, and $t = \cosh(\eta/2)$). By taking the exponential, the following product formula appears:

$$b + \sqrt{b^2 - 1} = \left(\prod_{\eta} \frac{t^2}{t^2 - 1} \right) \prod_{\alpha \in M} \left(\prod_{\partial\eta \ni \alpha} \left(\frac{a + \sqrt{a^2 + t^2 - 1}}{\sqrt{a^2 - 1} + \sqrt{a^2 + t^2 - 1}} \right) \right) \tag{1}$$

where the first product is taken over all orthogeodesics unsupportive of M, and the last product is over all orthogeodesics unsupportive of M *and* peripheral to α.

The first sum (or product in the half-trace formulation) is exactly the same term as the term in the Basmajian identity [1], and is the projection of a copy of β onto itself (following the reverse direction of the orthogeodesic). And in fact, when $M = \varnothing$, the above identity is the Basmajian identity: all orthogeodesics are unsupportive, the first sum is empty, and the second sum sums over all orthogeodesics leaving from β.

When M however is the set of all (primitive) simple closed geodesics, taken with both possible orientations and including boundary curves, the above identity is an expression of the McShane identity for surfaces with geodesic boundary due to Mirzakhani [21] and Tan-Wong-Zhang [22]. This is another extremal case of the identity, because it is a sum over simple orthogeodesics. And simple orthogeodesics are always unsupportive. Hence, the first sum has the "smallest" possible index set, in contrast with the Basmajian case where the first sum has the "largest" possible set. The expression of the identity is clearly different from previous formulations however, and this is due to several factors. The gap functions in the work of Mirzakhani and Tan-Wong-Zhang are all computed using the boundary curves of the associated pair of pants. In particular, the length of β plays a part in the function and the functions depend on two factors. Here the gaps are divided into parts: one part corresponding the Basmajian type term, which only depends on the associated ortholength, and two different gaps on each side of the orthogeodesic, which have another dynamical interpretation to be be discussed later, and which depend on two geometric quantities, the ortholength and the length of α. This difference in interpretation of the gaps shows how to express the length of β via a sum of gaps which depend on geometric quantities which do not involve β, similarly to the original Basmajian identity.

By taking an appropriate geometric limit, the identity can be quantified as follows on a surface with cusps. Note that one of the sum disappears.

Theorem 1.3. *Let X be a hyperbolic surface with $\partial X \neq \varnothing$ consisting of $n > 0$ cusps, and M a*

coherent marking.

$$n = \frac{1}{2} \sum_{\alpha \in M} e^{-\frac{\alpha}{2}} \left(\sum_{\eta \in \mathcal{O}_M^\alpha} e^{-\frac{\eta_\odot}{2}} \right)$$

When M is the set of all primitive simple closed geodesics, the above expression is a reformulation of the original McShane identity [20]. This will be explained in Section 4.2.

Both quantifications of the identity (Theorems 1.2 and 1.3) are obtained as sums of a quantification of the abstract identity, which takes into account one boundary component at a time:

Theorem 1.4. *If β is a boundary simple closed geodesic of X, it satisfies:*

$$\ell(\beta) = \sum_{\eta \in \mathcal{O}\beta_M} \log\left(\coth^2(\eta/2) \right)$$

$$+ \sum_{\alpha \in M} \sum_{\eta \in \mathcal{O}\beta_M^\alpha} \log\left(\frac{\cosh(\alpha/2) + \sqrt{\cosh^2(\alpha/2) + \cosh^2(\eta/2) - 1}}{\sinh(\alpha/2) + \sqrt{\cosh^2(\alpha/2) + \cosh^2(\eta/2) - 1}} \right).$$

If β is a cusp:

$$2 = 2 \sum_{\eta \in \mathcal{O}\beta_M} e^{-\eta_\circ} + \sum_{\alpha \in M} \left(\sum_{\eta \in \mathcal{O}\beta_M^\alpha} e^{-\frac{\alpha + \eta_\odot}{2}} \right)$$

where $\eta_\circ$ is the truncated length of η and, for orthogeodesics that leave and return to β, $\eta_\odot$ is the doubly truncated length.

Organization.

The article is organized as follows. Section 2 contains a lot of the groundwork: after having introduced a bit of notation, coherent markings are properly introduced and studied, followed by a detailed computation of the gaps. In Section 3, the abstract identity is proved. In the final section, the quantified identities are formulated by inserting the previously computed gaps, and the relationship to existing identities is then discussed.

Acknowledgments.

2. Background, notation and preliminaries

Let Σ be a topological orientable surface with a non-empty set of marked points and negative Euler characteristic. Take X to be a hyperbolic surface, homeomorphic to Σ and with either cusp or geodesic boundary. The moduli space $\mathcal{M}(\Sigma)$ is the set of such X up to isometry and $\mathrm{Mod}(\Sigma)$ is the (full) mapping class group of Σ, that is homeomorphisms up to isotopy.

2.1. Curves, arcs and geodesics

In a bit of somewhat non-standard notation, $\mathcal{C}(\Sigma)$ will designate the set of *all* curves on Σ, by which we mean the set of free homotopy classes of oriented closed curves, including curves peripheral to boundary. Elements of $\mathcal{C}(\Sigma)$ are not parametrized, but are oriented, and include all non-primitive elements. In particular, there are infinitely many homotopy classes peripheral to boundary. Because of our condition on X, via the unicity of geodesics in free homotopy classes, elements of $\mathcal{C}(\Sigma)$ correspond to the set of all closed geodesics $\mathcal{G}(X)$, including those that are not primitive and boundary elements. There is a slight matter of convention here: if X has cusp boundary, it might not make much sense to consider a cusp as a boundary geodesic, much less an oriented boundary geodesic or a power of one, but this slight discrepancy between $\mathcal{C}(\Sigma)$ and $\mathcal{G}(X)$ as sets of free homotopy classes is not a fundamental issue and will be discussed further below.

Similarly, $\mathcal{A}(\Sigma)$ will be the set of topological arcs on Σ with endpoints on the marked points and up to homotopy preserving the endpoints. Note there is no notion of primitive for elements of $\mathcal{A}(\Sigma)$ (said otherwise, all elements of $\mathcal{A}(\Sigma)$ are primitive). Given X, the geometric realization of $\mathcal{A}(\Sigma)$ is the set $\mathcal{O}(X)$ consisting of orthogeodesics of X: these are geodesic arcs with endpoints orthogonal to the boundary of X. Again, if X has cusps, the term orthogeodesic might seem inappropriate for the infinite length geodesics travelling to a cusp, but in fact these geodesics are, in the a region sufficiently deep in the cusp, orthogonal to the boundary of any horocylic neighborhood of the cusp.

Before introducing markings, we begin by making some observations about $\mathcal{C}(\Sigma)$ and $\mathcal{A}(\Sigma)$.

Closed curves, arcs and their subloops

Let $\gamma \in \mathcal{C}(\Sigma)$ which we always think of being in minimal position, meaning that minimizes self-intersection number. For the most part, we will think of closed curves as being realized topologically in minimal position, and to avoid further confusion, in a way such that all intersection points are double. Hence in an intersection point, because curves are oriented, there are exactly two outgoing directions, each associated to an incoming direction. If γ is primitive, meaning not the power of another curve, then its realization as a closed geodesic in $\mathcal{G}(X)$ (for any X) is in minimal position. There is a subtlety here when γ is not primitive however: its unique closed geodesic representative will be some number of iterates of the

associated primitive closed geodesic. While this geometric representative minimizes *transversal* self-intersection, of course there are infinitely many double points, but of course there are different ways of getting around this difficulty. In any event, we denote by $i(\alpha, \alpha)$ the minimal number of double points among topological representations (higher order multiple points are counted with multiplicities so that minimality is reached by certain representatives with double points).

Associated to a curve in $\mathcal{C}(\Sigma)$ are (oriented and proper) subloops. Given any curve with self-intersection points, in each intersection point (which we suppose double) there are exactly two (not necessarily) distinct subloops, obtained by taking one of the two outgoing paths until it returns. In any point that is not a self-intersection point, there is a unique loop beginning and ending given by the curve itself, and so there are no proper subloops in that point. Hence, any closed curve α, has at most $2i(\alpha, \alpha)$ proper subloops, each corresponding to (not necessarily distinct) free homotopy classes of oriented closed curves. Further observe that this set of elements of $\mathcal{C}(\Sigma)$ associated to a given α does *not* depend on the choice of representation of α. Indeed, given an intersection point of a curve, following the associated loops as the point moves under a free homotopy, it is easy to see that the two free homotopy classes do not change. Similarly, arcs can have proper subloops, corresponding to homotopy classes of closed curves. This brings us to the following:

Definition 2.1. Let α and β be curves in $\mathcal{C}(\Sigma)$. We say that β is supported by α if β is freely homotopic to a proper subloop of α and we write $\beta \hookrightarrow \alpha$.

Similarly, for $\eta \in \mathcal{A}(\Sigma)$, we say β is supported by η if it is freely homotopic to a proper subloop of η and we write $\beta \hookrightarrow \eta$.

We now make a few observations about curves, arcs and their proper subloops.

Proposition 2.2. *Let $\beta \hookrightarrow \alpha$ and suppose β is not prime, meaning $\beta = (\beta')^k$ for some natural number $k > 1$. Then $\beta' \hookrightarrow \alpha$. More generally $(\beta')^l \hookrightarrow \alpha$ for all $l \leq k$.*

Proof. As will be detailed in the sequel, for a given hyperbolic structure this is certainly true for any closed geodesic representative. This comes from the fact that a geodesic only loops k times around a given geodesic β' if it belongs to a small enough collar around β'. And the width of the collar depends only on the length of β', and is decreasing. In particular, if you enter the collar for β, you've entered the collar for each $(\beta')^k$. See Remark 2.4 for more details about the collar.

Although we make no use of this in the sequel, it is interesting to point out that this is really a topological fact, and does not even depend on having β in minimal position. To see this, the easiest is probably to look at β in the annular cover corresponding to β' (this is a finite cover

where β' has a lift freely homotopic to a simple primitive loop which generates the annulus). The curve β has a lift $\tilde{\beta}$ that is freely homotopic to multiple of generator of the annulus. Now it is not too difficult to convince oneself that $\tilde{\beta}$ contains a subloop that is homotopic to the generator, which doesn't contain the basepoint (this last point is key, otherwise it won't project to the desired loop). $\qquad\square$

Also observe that $\alpha \in \mathcal{C}(\Sigma)$ is simple (and primitive) if and only if it does not support *any* curves, and similarly for $\eta \in \mathcal{A}(\Sigma)$. However, if α is a power of a simple curve, then it supports all lesser powers. More generally, any curve that is a proper power supports all lesser powers.

More generally, one might hope that its a transitive property (if $\beta \hookrightarrow \alpha$ and $\gamma \hookrightarrow \beta$, then $\gamma \hookrightarrow \alpha$) but this is not always the case, see Figure 1.

Figure 1: The left most curve is not a subloop of the right most curve

We now pass to the geometric counterparts of these notions.

Closed geodesics and orthogeodesics

Now consider $X \in \mathcal{M}(\Sigma)$. To any element of $\mathcal{C}(\Sigma)$ that is not peripheral to boundary, there is a unique geodesic representative. Likewise, any interior element of $\mathcal{G}(X)$ corresponds to a unique element of $\mathcal{C}(\Sigma)$. This is also true for boundary geodesics and their corresponding peripheral homotopy classes, but not quite to cusps. The cusp is the geometric realization of all homotopy classes peripheral to it. There is however a complete one to one correspondence between elements of $\mathcal{A}(\Sigma)$ and those of $\mathcal{O}(X)$. Via this correspondence, closed geodesics and orthogeodesics of X inherit the notion of *support*.

Convention. Throughout the article, geometric quantities, namely lengths realized on an underlying hyperbolic surface X, will be denoted by the same symbol as the corresponding homotopy classes in $\mathcal{C}(\Sigma)$ and $\mathcal{A}(\Sigma)$. For example, with this convention the quantity $\cosh(\alpha/2)$ for $\alpha \in \mathcal{C}(\Sigma)$, can be thought of as a function of X in $\mathcal{M}(\Sigma)$ which associates to X the hyperbolic cosine of half of the length of the unique geodesic in the free homotopy class of α.

Oriented geodesics on X naturally correspond to subset of vectors in $T^1(X)$, the unit tangent bundle of X. To each $\gamma \in \mathcal{G}(X)$, we can associate a so-called stable neighborhood [2], but here it is more relevant to associate another subset of $T^1(X)$ to γ.

To define it, introduce a bit of notation: given $v \in T^1(X)$, we let geod$_v$ be the complete oriented geodesic on X obtained by exponentiating v forwards and backwards, but only taking the forwards orientation (given by v).

Definition 2.3. Let $\gamma \in \mathcal{C}(\Sigma)$. The loop set of γ, denoted $L_X(\gamma)$, is the subset of $T^1(X)$ consisting of all vectors belonging to oriented geodesics that form a loop freely homotopic to γ. Stated differently:

$$L_X(\gamma) = \{v \in T^1(X) \mid \text{geod}_v \text{ contains a geodesic loop freely homotopic to } \gamma\}.$$

This is well-defined, even if γ is a homotopy class peripheral to boundary, and even if this boundary is geometrically realized as a cusp.

Remark 2.4. To γ we can associated a (generally immersed) collar (or stable neighborhood) which is related to the loop set of γ. If you look at an annular lift of γ (a finite lift where γ is simple), by an adaptation of the classical collar lemma, see for instance [3], one obtains an embedded collar of half-width exactly

$$w(\gamma) = \text{arcsinh}\left(\frac{1}{\sinh(\gamma/2)}\right)$$

where to simply notation, γ also denotes the length of the geodesic in $\mathcal{G}(X)$ corresponding to γ. The stable neighborhood is the projection of this collar to X. By construction, it includes all points of X distance at most $w(\gamma)$ from γ. Now if a geodesic comes at least $w(\gamma)$ close to γ, it either crosses γ immediately, by which it is meant that the geodesic follows γ until it intersects γ transversally, or it belongs to $L_X(\gamma)$.

2.2. Coherent markings and peripheral orthogeodesics

We can now introduce markings which are essential to the index sets of the identities.

A *marking* of a set of curves is a choice of a subset M of $\mathcal{C}(\Sigma)$, which is allowed to be empty. (Equivalently, one could think of a marking as a map from $\mathcal{C}(\Sigma)$ to $\mathbb{Z}_2$, where the image is 1 for all marked curves, and 0 for the others.)

We say that a marking M is *coherent* if it satisfies the property that if $\alpha \in M$, then all curves supported by α are not. In other words, if $\alpha \in M$ and if $\beta \hookrightarrow \alpha$, then $\beta \notin M$.

Here are some examples of coherent markings:

1. $M = \emptyset$. This example will give rise the index set of the Basmajian identity in the sequel.

2. $M = \{\alpha\}$ where $\alpha \in \mathcal{C}(\Sigma)$ is any choice of curve. Similarly, you can construct finite sets of curves by taking any finite set, and remove curves one by one that are supported by other curves. Note that there may be more than one way to do this.

3. $M = \{\alpha \mid \alpha \in \mathcal{C}(\Sigma)$ is simple and primitive.$\}$. Note that this includes curves peripheral to boundary, and if $\alpha \in M$, then $\alpha^{-1} \in M$. This example will give rise the index set of McShane type identities as will be explained in Section 4.

4. $M = \{\phi(\alpha) \mid \phi \in \mathrm{Mod}(\Sigma)\}$ for a choice of $\alpha \in \mathcal{C}(\Sigma)$. These are the curves obtained in the mapping class group orbit of a curve. For certain choices of α, it might happen that $\phi(\alpha) = \phi'(\alpha)$ for different $\phi, \phi' \in \mathrm{Mod}(\Sigma)$, but in this case, the curve is not counted with multiplicity. An example where this *doesn't* happen is when α is a filling curve (its complementary region is a collection of topological disks with at most one marked point).

Here are some immediate properties of coherent markings.

Property 2.5. *Coherent markings satisfy the following:*

- *Any coherent marking can be obtained by taking curves in succession $\{\gamma_1, \gamma_2, \ldots\}$ and verifying that γ_k does not support and is not supported by an element of $\{\gamma_1, \ldots, \gamma_{k-1}\}$.*

- *Any subset $M' \subset M$ of a coherent marking M is a coherent marking.*

- *There are uncountably many coherent markings.*

- *If $\alpha \in M$ then for all integers $k > 1$, $\alpha^k \notin M$.*

Proof. The first property follows from the fact that the set of elements of $\mathcal{C}(\Sigma)$ is countable.

The second property is immediate, because if a subset of M violates coherency, then so does M.

For the third property, it suffices to find a marking of infinite cardinality, and then to consider all subsets of this marking. The examples 3 and 4 above provide such markings.

The last property follows from the definition of a coherent marking and that roots of a non-primitive curve are subloops. $\square$

Given a coherent marking M, $\mathcal{A}(\Sigma)$ naturally splits into two sets: those that support at least one element of M, and those that don't. An arc $\eta \in \mathcal{A}(\Sigma)$ is said to be *unsupportive* if it does not support any element of M.

Certain arcs have the same starting and ending points. For an associated orthogeodesic, these are those that leave and return to the same boundary curve. Naturally associated to an arc η, whose endpoints are both the same marked point p, are two elements of $\mathcal{C}(\Sigma)$. These the two curves obtained by looking at η on $\Sigma \setminus \{p\}$ as a closed curve, and pushing it slightly away from p (see Figure 2).
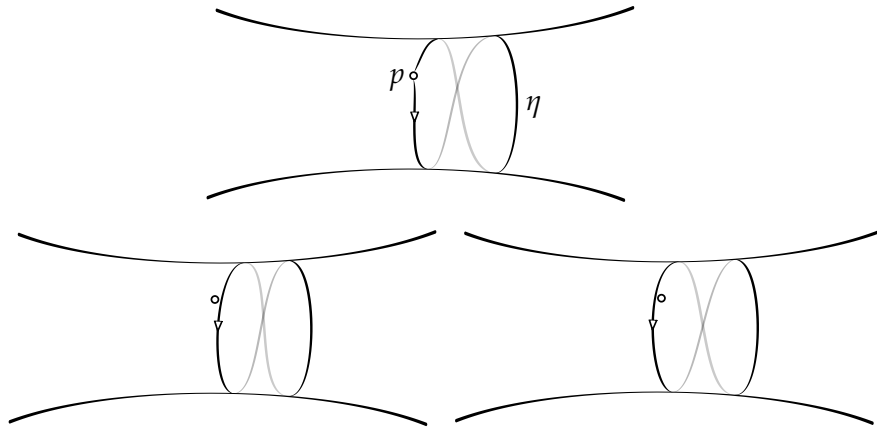
Figure 2: The two curves obtained by pushing η slightly away from p

Often these two curves are distinct, but there is one case where they are not. Take Σ to be a torus with a single marked point, and take η to be a simple arc. Then the two (oriented) curves are freely homotopic (see Figure 3).
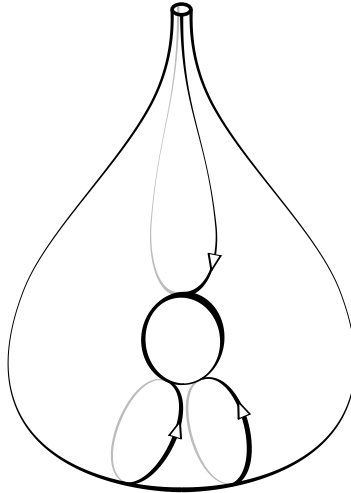

Figure 3: An η with identical curves in $\partial\eta$

Given an arc $\eta \in \mathcal{A}(\Sigma)$, we denote the set of these two (not necessarily distinct) curves $\partial\eta$, and if $\alpha \in \partial\eta$, we say that η is *peripheral* to α.

With this in hand, for a given coherent marking M, we see that certain unsupportive arcs are special: we say that an unsupportive arc η is *peripheral* to M if there exists $\alpha \in M$ such that $\alpha \in \partial\eta$. If an arc η is peripheral to α, note that via the homotopy on $\Sigma \cup \{p\}$, η and α bound a

(possibly) immersed cylinder H on Σ, with p lying on one of its boundaries. Similarly, p and the two curves in $\partial\eta$ form the boundary of an immersed thrice punctured sphere (a pair of pants). The cylinder H we call an immersed *half-pants*. See Figure 4 for an example.
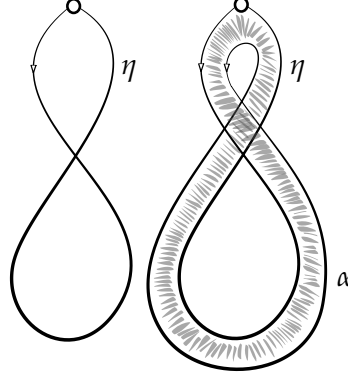

Figure 4: An arc η and an associated half-pants

2.3. Gaps and their inclusion

Associated to an orthogeodesic are two types of gaps. One of them is the gap that appears in the Basmajian identity, and the other is a generalization of the type of gap that appears in McShane type identities.

Type 1: Basmajian type gaps

For completion we describe, and compute, the gap associated to an orthogeodesic following [1].

Let η be an orthogeodesic between boundary elements β and β'. We begin with the case when β and β' are both simple closed geodesics. If we take a lift of η in the universal cover $\mathbb{H}$, it lies between lifts of β and β', say $\tilde{\beta}$ and $\tilde{\beta}'$. If you take the shortest point projection to $\tilde{\beta}$ of $\tilde{\beta}'$, you obtain a segment b on $\tilde{\beta}$ of length that only depends on η. The shortest point projection geodesics, together with the segment b and $\tilde{\beta}'$, form a quadrilateral in $\mathbb{H}$ with two right angles and two ideal points, as portrayed in Figure 5. By splitting the quadrilateral in two along the lift of η, we get a Lambert quadrilateral (called a trirectangle in [10]), which satisfies

$$\sinh(b/2)\sinh(\eta) = 1.$$

From this:

$$b = 2\operatorname{arcsinh}\left(\frac{1}{\sinh(\eta)}\right) = 2\log\left(\coth\left(\frac{\eta}{2}\right)\right).$$

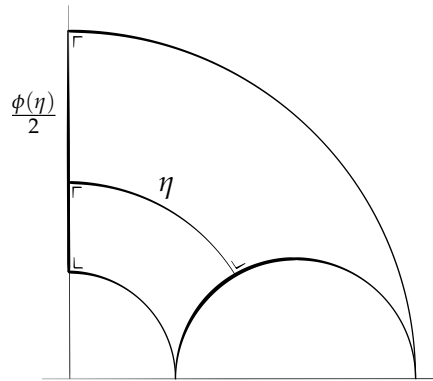Now b projects to X to a segment along β of the same length, say $\phi(\eta)$.

Figure 5: How to compute the Basmajian gap in $\mathbb{H}$

Remark 2.6. If you double the surface X along its boundary, the two copies of η become a closed geodesic of length 2η. On this doubled surface, this closed geodesic has a collar, as observed in Remark 2.4, which is exactly of width $\phi(\eta)/2$. This is not, of course, a complete coincidence.

If the orthogeodesic goes from a cusp and a closed geodesic, the closed geodesic can be projected onto a horocyclic boundary of the cusp. The gap this time will depend on the truncated length, by which we mean the length of the orthogeodesic that leaves from the cusp neighborhood and arrives on the geodesic. Note that the orthogeodesic is allowed to pass through the cusp neighborhood (as many times as it likes) and that length is still accounted for in the truncated length. In other words, it is only the initial infinite length segment from the cusp to the boundary of the neighborhood that is truncated.
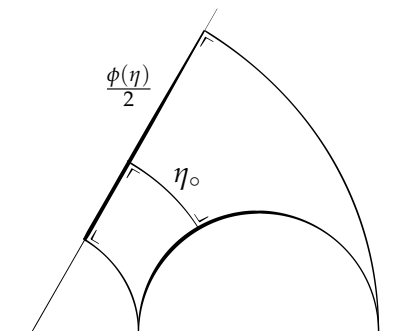

Figure 6: Computing the truncated gap in $\mathbb{H}$

If we denote by $\eta_\circ$ its truncated length, we obtain a gap of size

$$\phi(\eta) = \frac{2}{e^{\eta_\circ}}$$

associated to the orthogeodesic η. One way to compute this is by seeing this gap as a limit of the gap along a collar of β as the length of β goes to 0. Alternatively, one can do a direct computation using the geometry portrayed in Figure 6.

Note that if we have an orthogeodesic ending in a cusp, we do *not* consider a type of doubly truncated orthogeodesic and the gap associated to this orthogeodesic is always 0. As we will see now, for orthogeodesics that return to the same cusp, the second type of gap *will* use the doubly truncated length.

Type 2: McShane type gaps

A slight word of warning: although these are similar to the gaps described in [21] and [22], they are fundamentally different in several ways. First of all, they are associated to an orthogeodesic, and not an embedded pair of pants. (Of course an embedded pair of pants is essentially equivalent to a simple *unoriented* orthogeodesic.) Here the gaps involve *any* oriented orthogeodesic, not necessarily simple and they are related to the half-pants alluded to earlier. We also point out that these only concern orthogeodesics that leave and return to a same boundary element.

Although most of the discussion is mostly topological, it is sometime convenient to rely on the underlying geometry. In what follows, β is implicitly a simple closed geodesic. The case where β is a cusp is of course topologically identical, and can be done geometrically in a near identical fashion by replacing β by a horocyclic neighborhood around β.

As mentioned before, associated to such an orthogeodesic η, are two curves, the collection of which is denoted $\partial\eta$. These both are homotopic to the concatenation of η with an arc of β between the endpoints of η (see Figure 7).
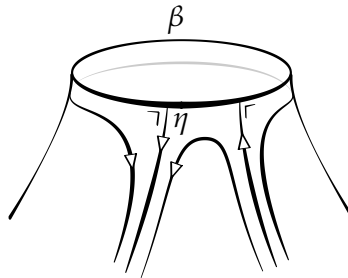

Figure 7: An orthogeodesic η and the curves $\partial\eta$ near β

Take one of the geodesic realizations of the curve, say α, and observe that it bounds an

immersed geodesic half-pants H, bounded on one end by α, and on the other by η and the segment b of β such that $b * \eta$ is freely homotopic to α. The half-pants is embedded if and only if η is simple.

Another way of seeing this is purely topological: if η (as an arc) is peripheral to α, then it is homotopic to a concatenation $c * \alpha * c^{-1}$ where c is a path from β to α (see Figure 8).
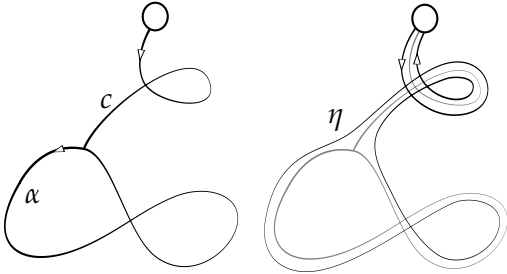

Figure 8: An arc η peripheral to α

The pre-immersed version of the cylinder is obtained by lifting c and α to simple curves $\tilde{c}$ and $\tilde{\alpha}$. From this we obtain a lift of η to $\tilde{\eta}$ (see Figure 9).
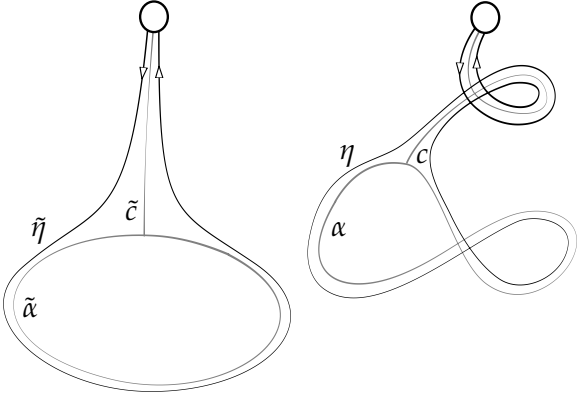

Figure 9: Lifting η, c and α to $\tilde{H}$ (on the left)

The half-pants H are exactly the image via the immersion of the half-pants $\tilde{H}$ with boundary curves $\tilde{\alpha}$ on one side, and a piecewise geodesic boundary made of $\tilde{\eta}$ and a segment of $\tilde{\beta}$, the lift of β (see Figure 10).
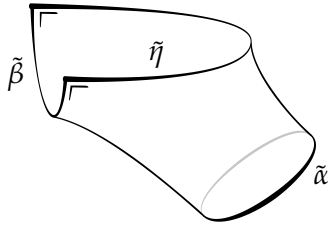
Figure 10: A lift of the geodesic half-pants

The lifts of the immersed half-pants preserve the lengths of curves, and so it suffices to compute in $\tilde{H}$. For this we consider the simple geodesic orthoray r leaving from $\tilde{\beta}$ and wrapping infinitely many times around $\tilde{\alpha}$ (in the same direction as $\tilde{\alpha}$). For those familiar with the gaps in McShane identities, this allows us to picture a gap in between $\tilde{\eta}$ and r (see Figure 11).
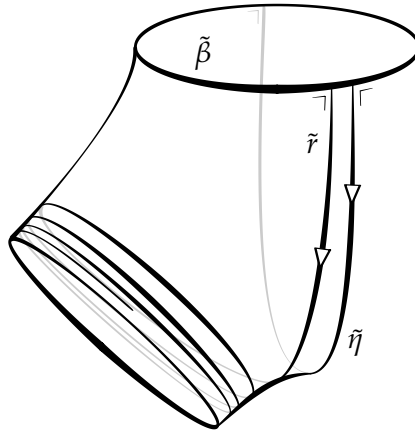

Figure 11: The distance between the basepoints of $\tilde{r}$ and $\tilde{\eta}$ is the sum of $\psi(\eta)$ and $\phi(\eta)/2$

The gap we consider here is *not* this whole gap however. From this gap, we remove the part of the gap coming from the Basmajian gap associated to η (or $\tilde{\eta}$) to obtain a gap $\psi(\eta)$. This is best illustrated in the universal cover (see Figure 12 where order not to introduce too much notation, the lifts of η, α, β and r are also denoted $\tilde{\eta}$, $\tilde{\alpha}$, $\tilde{\beta}$ and $\tilde{r}$). The gap $\psi(\eta)$ associated to η is the length of the segment as indicated.

There is a dynamic interpretation of the gap: the gap is the set of (basepoints of) orthorays that lie entirely in H until they make a loop around α. The associated base vectors in $T^1(X)$ all belong to $L(\alpha)$. They might belong to other $L(\alpha')$ for $\alpha' \in M$, but they will form a loop around α before forming a loop around any other curve in M. This will be discussed in detail
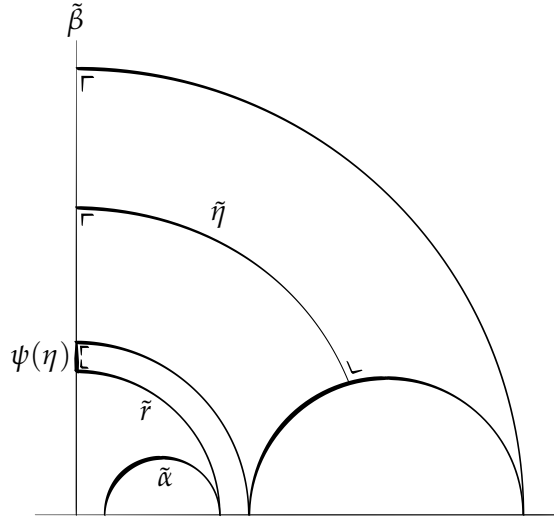
Figure 12: The gap $\psi(\eta)$ viewed in $\mathbb{H}$

in the proofs of the identities.

The gap is bounded by two extremal cases: an orthoray that stays inside H and wraps infinitely many times around α, and an orthoray that corresponds to, on the completion of the surface, an ideal geodesic loop that is homotopic to α. Note that, if H is immersed and not embedded, there might be many (possibly infinitely many) such orthorays. The extremal ones we are interested in are the projections of the simple ones with this behavior in $\tilde{H}$.

We can now compute $\psi(\eta)$. Up until now, the discussion has been essentially topological, and there are really two geometric situations to consider: when β is a closed geodesic or when β is a cusp. We begin with the first case, and the do the computation in the second case via a limiting argument.

Computing the gap when β is a closed geodesic

By cutting and unfolding the half-pants, we get a symmetric right-angled hexagon as in Figure 13. By putting it in $\mathbb{H}$ and unwrapping r, we can compute $\psi(\eta)$.
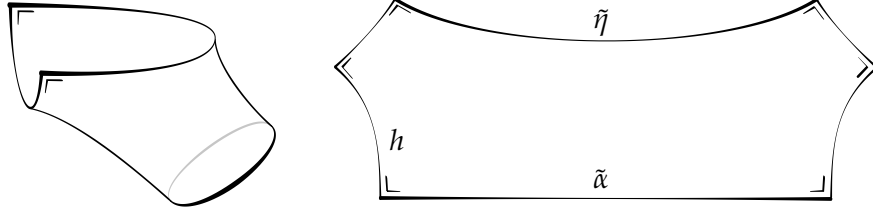
Figure 13: Cutting and unfolding the (lifted) half-pants

In the following formulas, we use α for the length of α (which is the same as the length of $\tilde{\eta}$, and similarly for η. We also abbreviate $\psi(\eta)$ to ψ and the Basmajian measure $\phi(\eta)$ to ϕ. We denote by x the quantity $\psi + \phi/2$ which is the length of the segment in between the base point of η and the base point of r. And we denote by y the complementary region to x on the side of the hexagon on which x lies. The quantity h is one of the sides of the hexagon as indicated on Figures 13 and 14.

$$\sinh(h)\sinh(\psi) = 1.$$

Figure 14: The pre-immersed pentagon

Now by looking in the right-angled pentagon that forms the left side of the hexagon (see Figure 14) and using the pentagon relation, we get

$$\sinh(x+y)\sinh(\eta/2) = \cosh(\alpha/2).$$

Putting this all together by elementary computations:

$$x = \operatorname{arcsinh}\left(\frac{\cosh(\alpha/2)}{\sinh(\eta/2)}\right) - y = \operatorname{arcsinh}\left(\frac{\cosh(\alpha/2)}{\sinh(\eta/2)}\right) - \operatorname{arcsinh}\left(\frac{1}{\sinh(h)}\right).$$

18

Now using the formula pentagon formula, $\sinh(h) = \cosh(\eta/2)/\sinh(\alpha/2)$ and so

$$
\begin{aligned}
x &= \operatorname{arcsinh}\left(\frac{\cosh(\alpha/2)}{\sinh(\eta/2)}\right) - \operatorname{arcsinh}\left(\frac{\sinh(\alpha/2)}{\cosh(\eta/2)}\right) \\
&= \log\left(\coth\left(\frac{\eta}{2}\right)\right) + \log\left(\frac{\cosh(\alpha/2) + \sqrt{\cosh^2(\alpha/2) + \cosh^2(\eta/2) - 1}}{\sinh(\alpha/2) + \sqrt{\cosh^2(\alpha/2) + \cosh^2(\eta/2) - 1}}\right) \\
&= \frac{\phi}{2} + \log\left(\frac{\cosh(\alpha/2) + \sqrt{\cosh^2(\alpha/2) + \cosh^2(\eta/2) - 1}}{\sinh(\alpha/2) + \sqrt{\cosh^2(\alpha/2) + \cosh^2(\eta/2) - 1}}\right)
\end{aligned}
$$

and because $\psi = x - \phi/2$, we get

$$
\psi(\eta) = \log\left(\frac{\cosh(\alpha/2) + \sqrt{\cosh^2(\alpha/2) + \cosh^2(\eta/2) - 1}}{\sinh(\alpha/2) + \sqrt{\cosh^2(\alpha/2) + \cosh^2(\eta/2) - 1}}\right).
$$

Half-trace notation

It is sometimes convenient to use *traces* or *half-traces* of lengths. These are the traces and half-traces (or absolute value of) the corresponding matrices in $\mathrm{PSL}_2(\mathbb{R})$. Note that to the gaps themselves, there is no corresponding surface group element, but there is a trace corresponding to the corresponding translation element in $\mathrm{PSL}_2(\mathbb{R})$.

We denote the half-traces by

$$
a := \cosh(\alpha/2),\ b := \cosh(\beta/2)\ \text{and}\ t := \cosh(\eta/2)
$$

and then we can express $\psi(\eta)$ and $\phi(\eta)$ as follows:

$$
\phi(\eta) = \log\left(\frac{t^2}{t^2 - 1}\right) \tag{2}
$$

$$
\psi(\eta) = \log\left(\frac{a + \sqrt{a^2 - 1 + t^2}}{\sqrt{a^2 - 1} + \sqrt{a^2 - 1 + t^2}}\right) \tag{3}
$$

$$
\beta = \log(b + \sqrt{b^2 - 1}) \tag{4}
$$

Limiting to a cusp

To compute the gap in the cusp case (this only concerns gaps of type 2), we use a standard limiting trick which involves multiplying the measure by a factor that varies in terms of the length of β. The resulting measure has the additional advantage of varying continuously from the boundary geodesic case to the cusp case.

To do this, instead of viewing the measure on the boundary geodesic, we project it to the boundary of the collar of β, where the collar has been described previously in Remark 2.4.

The collar boundary is not geodesic of course, but is a curve of constant curvature. By a standard computation, the boundary length of the collar of β is $\beta \coth(\beta/2)$. The collar limits (continuously) to the standard horocyclic neighborhood of a cusp of boundary length 2 as β goes to 0.

A segment of length ℓ on β projects to segment of length $\ell \coth(\beta/2)$ on the collar boundary. We get a new projected gaps given by $\coth(\beta/2)\psi(\eta)$ and $\coth(\beta/2)\phi(\eta)$, but the latter limits to 0 as β goes to 0. The former however, even though $\psi(\eta)$ tends to 0 as β goes to 0, doesn't disappear in the limit, and in fact, turns into a very nice expression.

Before computing the limit, we need to know what we are computing the limit of. The length of the actual orthogeodesic η goes to ∞ as β goes to 0, so we consider the twice truncated orthogeodesic $\eta_\circ$: this is the orthogeodesic to and from the horocyclic neighborhood of the cusp where the initial and final geodesic segment which goes respectively from and to the cusp has been removed. This truncated orthogeodesic is the limit of the truncated orthogeodesic to and from the boundary of the collar of β as the length of β goes to 0 (which we also denote $\eta_\circ$). Note that, as described previously, the doubly truncated orthogeodesic may pass through the collar again, and in fact only does so if it winds around β, but we won't make use of that fact here.

So we have
$$\eta = \eta_\circ + 2\operatorname{arcsinh}\left(\frac{1}{\sinh(\beta/2)}\right)$$
and thus by a standard manipulation and limit argument:
$$\lim_{\beta \to 0}\left(e^{\eta_\circ/2}\frac{\cosh(\eta/2)}{\sinh(\beta/2)}\right) = 1.$$

With this in hand, and using the fact that
$$\lim_{\beta \to 0}\left(\coth(\beta/2)\psi(\eta)\right) = \lim_{\beta \to 0}\left(\coth(\beta/2)\sinh(\psi(\eta))\right),$$

one has

$$
\begin{aligned}
\lim_{\beta \to 0}\left(\coth(\beta/2)\psi(\eta)\right) &= \lim_{\beta \to 0}\left(\coth(\beta/2)\frac{e^{-\alpha/2}}{\cosh(\eta/2)}\right) \\
&= \lim_{\beta \to 0}\left(\frac{e^{-\alpha/2}}{\sinh(\beta/2)\cosh(\eta/2)}\right) \\
&= e^{-\alpha/2} \cdot e^{-\eta_\circ/2}.
\end{aligned}
$$

For a surface with a cusp, to an orthogeodesic η of truncated length $\eta_\circ$ we thus associate the gap
$$\psi_\circ(\eta) = e^{-\frac{\alpha+\eta_\circ}{2}}.$$

20

3. Proof of the identity

We now fix a marking M of $\mathcal{C}(\Sigma)$, and take $X \in \mathcal{M}(\Sigma)$. The identity is a way of breaking up a boundary element of X into gaps that depend on the behavior of the associated orthoray.

We begin with the case when the boundary curve to be partitioned into gaps is a simple closed geodesic β. To each $p \in \beta$ corresponds an element v_p of $T^1(X)$, the unique vector orthogonal to β in p and aiming inwards towards X. There is also the associated *orthoray* $\vec{r}_p$ given by forward exponentiating v_p.

We follow $\vec{r}_p$ from its basepoint until one of three things happen:

- It forms a loop around an element in M, in which case we stop and denote the associated geodesic arc r_p.

- It hits ∂X, the boundary of X, and is freely homotopic to an orthogeodesic which we denote η_p.

- It wanders forever in X without ever forming a loop around an element of M or hitting ∂X.

We will show in the sequel that the last case only happens for a measure 0 set of points on β.

First case: a loop is made

Let us begin with the first case, where we have an arc r_p, which ends exactly after having formed a loop around an element of M, say α_0, for the first time. We can construct a homotopy class of arc and corresponding orthogeodesic by concatenating r_p with the path that follows r_p with the opposite orientation from the endpoint. We denote the corresponding arc and orthogeodesic by η_0.

Key observation: By construction: η_0 is peripheral to α_0 and p (or $\vec{r}_p$) belongs to the gap associated to η_0.

The latter follows from the dynamics of the behavior of all orthorays in the gap associated to η_0.

However, η_0 might not be unsupportive, that is η_0 might have a subloop belonging to M. If this is the case, we consider, following η_0 from its basepoint, the first time it forms a loop around an element of M, say α_1, and denote by r_1 be the corresponding oriented subpath. We then construct a new homotopy class of orthogeodesic η_1 by following, from the endpoint of r_1, η_0^{-1} back to β. Similarly to above, the key observation is that the entire gap of η_0 is contained in the gap of η_1. And by construction, η_1 is peripheral to α_1.

We can now repeat the above process: if η_1 is *not* unsupportive, then it forms at least one loop

around an element of M. We can then construct η_2, containing the gap associated to η_1, hence the gap associated to η_0, hence p, and so forth.

It is relatively straightforward to see that his process finishes after a finite number of steps, which only depend on the number of self-intersections of η_0 (but the only thing we really need is that it finishes). The point p belongs to the gap of the resulting orthogeodesic, say η_p, which is both peripheral to an element of M, and unsupportive.

Second case: $\vec{r}_p$ returns to ∂X

In this case, we stop $\vec{r}_p$ at ∂X to obtain a finite geodesic arc, which in turn corresponds to an orthogeodesic η_p. We associate p to the gap associated to η_p. Note that if it ends in a cusp, then the gap is of length 0.

To prove the identity we need to show that the gaps are disjoint, and that the set of for ever wandering orthorays is a measure 0 subset of β. This latter fact - in the case where β is a simple closed geodesic - follows from the same limit set argument as in [1], namely because the set of for ever wandering orthorays is a subset of the for ever wandering rays in the Basmajian identity. If however, β is (the horocyclic boundary of) a cusp, this requires an additional argument.

Lemma 3.1. *Let X be a surface with a cusp, and let β be a horocyclic neighborhood of a cusp. Suppose that $M \neq \varnothing$ and/or ∂X contains a boundary geodesic. Then the set of points on β which are the basepoints of orthorays of infinite length that are not supported on M or that do not hit ∂X in finite time are a subset of measure 0 of β. In other words, the complementary regions of the gaps are of measure 0.*

Proof. If ∂X has boundary curves, then this follows from the same limit set argument as in [1]. Suppose then that ∂X consists only of cusps. In this case, the geodesic flow is ergodic and by hypothesis $M \neq \varnothing$.

Consider p with $\vec{r}_p$ which is infinitely long and never forms a loop around an element of M. By ergodicity, arbitrarily close to any orthoray $\vec{r}_p$ leaving from $p \in \beta$ is a geodesic that passes through $L(\alpha)$ for any $\alpha \in M$. This is because the latter forms a set of positive measure in $T^1(X)$. That means that arbitrarily close to the orthogonal vector v_p corresponding to $\vec{r}_p$ a vector v with angle arbitrarily close to $\frac{\pi}{2}$ that exponentiates to reach $L(\alpha)$. By continuity of geodesic behavior, arbitrarily close to p is an orthoray leaving from β belonging to a gap. Hence p is isolated in β, proving the lemma. $\qquad\square$

For the following lemma, we denote the gaps β_i, indexed by $i \in I$, and think of each gap as an open subset of β (hence without its endpoints).

Lemma 3.2. *Let β_i and β_j be two different gaps. Then $\beta_i \cap \beta_j = \varnothing$.*

Proof. Suppose that β_i and β_j are of different type, say type 1 and type 2. Then if $p \in \beta_i$, the associated orthoray returns to β without forming a loop in M, hence it cannot be of type 2.

Note that two gaps of type 1 are disjoint, as there is a unique orthogeodesic associated to each homotopy class of arc between boundary elements.

Now suppose that β_i and β_j are both of type 2.

It suffices to show that for all points of β_i, the first loop in M that the associated orthoray forms is freely homotopic to α_i, and that the associated arc constructed above is always homotopic to η_i.

Consider the boundary point p_i of β_i corresponding to an ideal orthoray freely homotopic to α_i. By continuity, for all points in β_i sufficiently close to p_i, the associated orthorays first form a loop around α_i creating an associated arc homotopic to η_i. Note that throughout the gap, the orthoray will continue to form this loop. We just need to analyse what happens for in each self-intersection point of the orthoray that occurs before this loop is formed.

Now suppose p' is the first point in the gap such that the associated orthoray $\vec{r}_{p'}$ forms another loop , say α', before looping around α_i. By continuity, as this is the first point, it occurs exactly at a previous intersection point where the associated loop is homotopic to α_i. Hence, it is easy to see that α' is a subloop of α_i. So either it is homotopic to α_i, in which we are happy, or its a proper subloop, in which case its homotopy class cannot belong to M by definition of a coherent marking. Note that this is exactly the crucial moment where coherent markings come into play. And further note that this *never* happens if η_i is simple.
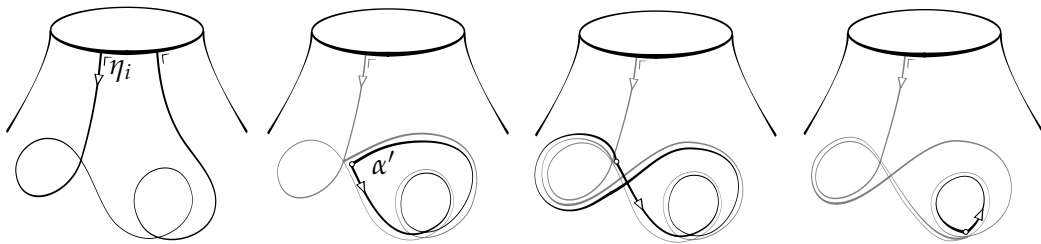

Figure 15: An example orthogeodesic η_i and subloops that appear along orthorays in the gaps

This was only the first critical point (meaning the first time an extra intersection point appeared). As we continue to push orthorays along the gap, the same process occurs again, and again, by the same argument, the new loop that appears is freely homotopic to α_i or to one of its subloops. This process is illustrated in Figure 15.

23

Hence the only loops that appear that belong to M are all freely homotopic to α_i. Thus there is no first point of the gap with a different behavior, which proves that two gaps of type 2 cannot overlap. □

This proves the (abstract) identities (Theorem 1.1), and in the next section, we restate and discuss the quantified identities, obtained by replacing the measures of the gaps with their computed values.

4. The full identities and final remarks

Now that the index set has been properly introduced, that gaps have been quantified, and that the proof that they form a full measure partition of the boundary length is done, we can state the identities.

4.1. The general identities

We begin with a full identity, which decomposes the actual boundary of X. In particular if ∂X contains a cusp, it is considered a geodesic of length 0, all orthogeodesics that have at least one endpoint in it are of infinite length, and contribute nothing to the sum.

We recall the notation: for a given coherent marking M, $\mathcal{O}_M$ denotes the orthogeodesics unsupportive of M (for implicit X). And $\mathcal{O}_M^\alpha$ is the subset of $\mathcal{O}_M$ which is peripheral to α. Note that by definition of being peripheral, orthogeodesics in $\mathcal{O}_M^\alpha$ leave and return to the same connected component of ∂X.

Theorem 4.1. *Let $X \in \mathcal{M}(\Sigma)$ be a hyperbolic surface of finite type with $\partial X \neq \varnothing$ containing at least one simple closed geodesic, and M a coherent marking of $\mathcal{C}(\Sigma)$. The lengths of curves and orthogeodesics of X satisfy*

$$\ell(\partial X) = \sum_{\eta \in \mathcal{O}_M} \log\left(\coth^2(\eta/2)\right)$$

$$+ \sum_{\alpha \in M} \sum_{\eta \in \mathcal{O}_M^\alpha} \log\left(\frac{\cosh(\alpha/2) + \sqrt{\cosh^2(\alpha/2) + \cosh^2(\eta/2) - 1}}{\sinh(\alpha/2) + \sqrt{\cosh^2(\alpha/2) + \cosh^2(\eta/2) - 1}}\right).$$

The condition on the boundary containing at least one simple closed geodesic is to avoid the above identity be a sum of 0s equal to 0.

The above identity is really a sum of identities taken over the connected components of ∂X that are simple closed geodesics. If we fix a boundary element β, the identity associated to β takes two forms, which only depends on whether β is realized on X or as a cusp or as a geodesic.

24

For M is a coherent marking, $\mathcal{O}\beta_M$ will denote the subset of $\mathcal{O}_M$ consisting of orthogeodesics leaving from β. And $\mathcal{O}\beta_M^\alpha$ is the subset of $\mathcal{O}\beta_M$ which is peripheral to α. As before, by definition, elements of $\mathcal{O}\beta_M^\alpha$ leave and return to β.

If β is a boundary simple closed geodesic of X, it satisfies:

$$\ell(\beta) = \sum_{\eta \in \mathcal{O}\beta_M} \log\left(\coth^2(\eta/2)\right)$$

$$+ \sum_{\alpha \in M} \sum_{\eta \in \mathcal{O}\beta_M^\alpha} \log\left(\frac{\cosh(\alpha/2) + \sqrt{\cosh^2(\alpha/2) + \cosh^2(\eta/2) - 1}}{\sinh(\alpha/2) + \sqrt{\cosh^2(\alpha/2) + \cosh^2(\eta/2) - 1}}\right).$$

Note that using the half-trace convention and computations (Equations 2, 3 and 3), the above becomes Equation 1.

If β is a cusp:

$$2 = 2\sum_{\eta \in \mathcal{O}\beta_M} e^{-\eta_\circ} + \sum_{\alpha \in M}\left(\sum_{\eta \in \mathcal{O}\beta_M^\alpha} e^{-\frac{\alpha + \eta_\circ}{2}}\right) \tag{5}$$

where $\eta_\circ$ is the truncated length of η and, for orthogeodesics that leave and return to β, $\eta_\circledcirc$ is the doubly truncated length. Note that in this case, every element of $\mathcal{O}\beta_M$ appears exactly once. Indeed, an orthogeodesic in $\mathcal{O}\beta_M$ returning to β has infinite truncated length, and thus contributes 0 to the lefthand sum, but does contribute to the right hand sum with its finite doubly truncated length. If however an orthogeodesic does not return to β, it cannot be peripheral to an element of M. Rewriting this sum taking this into account, and reorganizing the last terms, gives us:

$$1 = \sum_{\eta \in \mathcal{O}\beta_M^{np}} e^{-\eta_\circ} + \frac{1}{2}\sum_{\alpha \in M} e^{-\frac{\alpha}{2}}\left(\sum_{\eta \in \mathcal{O}\beta_M^\alpha} e^{-\frac{\eta_\circledcirc}{2}}\right) \tag{6}$$

where $\mathcal{O}\beta_M^{np}$ is the set of non-peripheral elements of $\mathcal{O}\beta_M$. In particular

$$\mathcal{O}\beta_M = \mathcal{O}\beta_M^{np}\dot\cup\left(\dot\cup_{\alpha \in M}\mathcal{O}\beta_M^\alpha\right).$$

With this viewpoint, Equation 6 gives a nice probability measure on orthogeodesics in $\mathcal{O}\beta_M$. The measure associated to an orthogeodesic η in $\mathcal{O}\beta_M^{np}$ is $e^{-\eta_\circ}$ and to an orthogeodesic η in $\mathcal{O}\beta_M^\alpha$ is $e^{-\frac{\alpha + \eta_\circledcirc}{2}}/2$.

Remark 4.2. There is something to check in as we have inverted (possibly) infinite sums. To show the sums are indeed the same, it suffices to show that the sum of $e^{-\eta_\circledcirc/2}$ for all η peripheral to α and unsupportive of M converges. Depending on M, the index set for this

sum is a subset of the set $\mathcal{O}\beta_\alpha^\alpha$ of orthogeodesics peripheral to α and unsupportive of α. Now, for *any* $\alpha \in \mathcal{C}(\Sigma)$, from 5 with $M = \{\alpha\}$:

$$2 = 2 \sum_{\eta \in \mathcal{O}\beta_\alpha^{\mathrm{np}}} e^{-\eta_\circ} + e^{-\frac{\alpha}{2}} \left(\sum_{\eta \in \mathcal{O}\beta_\alpha^\alpha} e^{-\frac{\eta_\circ}{2}} \right).$$

In particular, the sum $\sum_{\eta \in \mathcal{O}\beta_\alpha^\alpha} e^{-\frac{\eta_\circ}{2}}$ converges.

If all boundary elements of X are cusps, then the first sum of Equation 6 disappears. In this case we have

$$1 = \frac{1}{2} \sum_{\alpha \in M} e^{-\frac{\alpha}{2}} \left(\sum_{\eta \in \mathcal{O}\beta_M^\alpha} e^{-\frac{\eta_\circ}{2}} \right) \tag{7}$$

This time we can also view Equation 7 as a type of probability measure on elements in M, where the measure on α is given by

$$\frac{1}{2} e^{-\frac{\alpha}{2}} \left(\sum_{\eta \in \mathcal{O}\beta_M^\alpha} e^{-\frac{\eta_\circ}{2}} \right).$$

If X has n cusps, by adding up the terms of Equation 7, one obtains

$$n = \frac{1}{2} \sum_{\alpha \in M} e^{-\frac{\alpha}{2}} \left(\sum_{\eta \in \mathcal{O}_M^\alpha} e^{-\frac{\eta_\circ}{2}} \right) \tag{8}$$

4.2. Relationship to previous identities

The Basmajian identity

When $M = \varnothing$, there are no marked curves, hence no peripheral orthogeodesics, and $\mathcal{O}_M = \mathcal{O}$. The identity becomes

$$\ell(\partial X) = \sum_{\eta \in \mathcal{O}} 2 \log \left(\coth(\eta/2) \right)$$

which is exactly the identity proved by Basmajian [1]. Note it only has content when ∂X contains at least one simple closed geodesic, and has no content when ∂X consists solely of cusps. Note there is a way of making sense of the identity when X only has cusps, or even cone-points, obtained by regrouping terms: see the results from [3].

Note this is not the same time the Basmajian identity has been linked to another identity. Bridgeman and Tan [8] showed how the Basmajian identity is linked to moments of the

Liouville measure. In fact it is the first term of a family of identities that you can derive from this point of view, where the second term is the Bridgeman identity [6] (and Bridgeman-Kahn [7] in higher dimensions). Also see [11].

McShane type identities

When M is the set of all simple closed curves (primitive, with both orientations, including those peripheral to boundary), $\mathcal{O}_M$ becomes the set of simple orthogeodesics $\mathcal{O}_0$. This is because any non-simple orthogeodesic supports a closed curve, and as such, supports a simple closed curve. The last implication follows from the observation that any non-simple closed curve has a simple (non-trivial) subloop, and hence it supports a simple closed curve.

In McShane type identities, the sums are over embedded pants with either one or two of its cuffs peripheral to boundary.

Consider an embedded pair of pants P with a single peripheral cuff β and two other cuffs α_1 and α_2. P There are two simple orthogeodesics η and η^{-1} leaving from β and contained in P: they have the same traces but opposite orientations (see Figure 16). The measure associated to P is the sum of the measures associated to η and η^{-1}. If β is a geodesic, this constitutes 6 terms: the measures $\phi(\eta)$ and $\phi(\eta^{-1})$ (the Basmajian type terms), the terms associated to η and α_1, resp. α_2, and the terms associated to η^{-1} and α_1, resp. α_2. If β is a cusp, then the Basmajian type terms disappear.
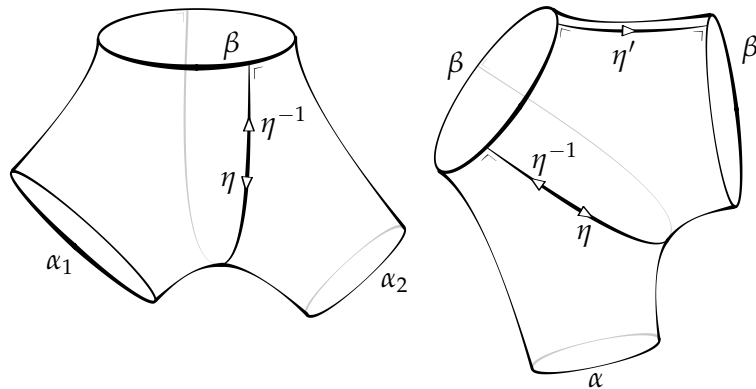

Figure 16: The pairs of pants P on the left and Q on the right.

To an embedded pair of pants Q with two peripheral cuffs, β and β', and an additional cuff α, in the McShane setup there is a gap associated to β and one associated to β'. There are three simple orthogeodesics in Q that leave from β_1 (see Figure 16). Two of them, say η and η^{-1} return to β, and like before have the same traces and different orientations. The 6 gaps associated to them are as before, where α plays the part of α_1 and β' plays the part of α_2.

27

There is another simple orthogeodesic, say η', that goes from β to β', and to it we associate an additional Basmajian type gap. Again, the sum of the usual McShane measure is the sum of these 7 gaps this time. As before, if β is a cusp, the Basmajian type terms disappear.

Final remarks

The Basmajian and McShane identities have been studied in many contexts. For instance, the Basmajian identity has been generalized to inequalities on maximal representations [13], to higher Teichmüller theory [23], and to situations with limit sets of different Hausdorff dimension [14]. The McShane identity found its first generalizations in the work of Bowditch [5], Mirzakhani [21] and Tan-Wong-Zhang [22], the latter having generalized it to cone surfaces. Bowditch also provided another proof of the original torus identity [4], recently generalized by Labourie-Tan [18]. Further generalizations include [17], [16], and [12]. A very useful survey is [9], although of course it doesn't take into account the most recent work.

It might be interesting to see to what extent the identities investigated in this paper generalize to other contexts. As mentioned previously, there is direct link between the Bridgeman identity and the Basmajian identity [8], and the two identities share the same index set. In the current identities, the measures are always sums of Basmajian terms (the McShane type terms being infinite sums of Basmajian type terms that share the same initial dynamic behavior). Hence there is an immediate abstract identity related to the Bridgeman identity, obtained by summing the terms corresponding to the same index set. In a related direction, the Luo-Tan identity [19] decomposes the volume of the unit tangent bundle (also see [15]). The index set in these identities is the set of embedded pairs of pants. It seems reasonable to hope to be able to generalize the techniques of the current paper to this context to obtain identities with certain types of immersed pairs of pants.

References

[1] Basmajian, Ara. The orthogonal spectrum of a hyperbolic manifold. Amer. J. Math. 115 (1993), no. 5, 1139–1159.

[2] Basmajian, Ara. The stable neighborhood theorem and lengths of closed geodesics. Proc. Amer. Math. Soc. 119 (1993), no. 1, 217–224.

[3] Basmajian, Ara, Parlier, Hugo and Tan, Ser Peow. Prime orthogeodesics, concave cores and families of identities on hyperbolic surfaces. Preprint, 2020.

[4] Bowditch, Brian H. A proof of McShane's identity via Markoff triples. Bull. London Math. Soc. 28 (1996), no. 1, 73–78.

[5] Bowditch, Brian H. A variation of McShane's identity for once-punctured torus bundles. Topology 36 (1997), no. 2, 325–334.

[6] Bridgeman, Martin. Orthospectra of geodesic laminations and dilogarithm identities on moduli space. Geom. Topol. 15 (2011), no. 2, 707–733.

[7] Bridgeman, Martin and Kahn, Jeremy. Hyperbolic volume of manifolds with geodesic boundary and orthospectra. Geom. Funct. Anal. 20 (2010), no. 5, 1210–1230.

[8] Bridgeman, Martin and Tan, Ser Peow. Moments of the boundary hitting function for the geodesic flow on a hyperbolic manifold. Geom. Topol. 18 (2014), no. 1, 491–520.

[9] Bridgeman, Martin and Tan, Ser Peow. Identities on hyperbolic manifolds. Handbook of Teichmüller theory. Vol. V, 19–53, IRMA Lect. Math. Theor. Phys., 26, Eur. Math. Soc., Zürich, 2016.

[10] Buser, Peter. Geometry and spectra of compact Riemann surfaces. Progress in Mathematics, 106. Birkhäuser Boston, Inc., Boston, MA, 1992.

[11] Calegari, Danny. Chimneys, leopard spots and the identities of Basmajian and Bridgeman. Algebr. Geom. Topol. 10 (2010), no. 3, 1857–1863.

[12] Charette, Virginie and Goldman, William M. McShane-type identities for affine deformations. Ann. Inst. Fourier (Grenoble) 67 (2017), no. 5, 2029–2041.

[13] Fanoni, Federica and Pozzetti, Maria Beatrice. Basmajian-type inequalities for maximal representations. J. Differential Geom., to appear.

[14] He, Yan Mary. Basmajian-type identities and Hausdorff dimension of limit sets. Ergodic Theory Dynam. Systems 38 (2018), no. 6, 2224–2244.

[15] Hu, Hengnan and Tan, Ser Peow. New identities for small hyperbolic surfaces. Bulletin of the London Mathematical Society 46.5 (2014): 1021–1031.

[16] Huang, Yi and Sun, Zhe. McShane identities for Higher Teichmüller theory and the Goncharov-Shen potential. Preprint, 2019.

[17] Labourie, François and McShane, Greg. Cross ratios and identities for higher Teichmüller-Thurston theory. Duke Math. J. 149 (2009), no. 2, 279–345.

[18] Labourie, François and Tan, Ser Peow. The probabilistic nature of McShane's identity: planar tree coding of simple loops. Geom. Dedicata 192 (2018), 245–266.

[19] Luo, Feng and Tan, Ser Peow. A dilogarithm identity on moduli spaces of curves. J. Differential Geom. 97 (2014), no. 2, 255–274.

[20] McShane, Greg. Simple geodesics and a series constant over Teichmüller space. Invent. Math. 132 (1998), no. 3, 607–632.

[21] Mirzakhani, Maryam. Simple geodesics and Weil-Petersson volumes of moduli spaces of bordered Riemann surfaces. Invent. Math. 167 (2007), no. 1, 179–222.

[22] Tan, Ser Peow, Wong, Yan Loi and Zhang, Ying. Generalizations of McShane's identity to hyperbolic cone-surfaces. J. Differential Geom. 72 (2006), no. 1, 73–112.

[23] Vlamis, Nicholas G. and Yarmola, Andrew. Basmajian's identity in higher Teichmüller-Thurston theory. J. Topol. 10 (2017), no. 3, 744–764.

Address:

Department of Mathematics, University of Luxembourg, Luxembourg

Email: hugo.parlier@uni.lu