A proof of McShane's identity via Markoff triples. †

B. H. Bowditch
Faculty of Mathematical Studies, University of Southampton,
Highfield, Southampton SO17 1BJ, Great Britain.
email: bhb@maths.soton.ac.uk

1. Introduction.

In [6], McShane described a remarkable identity concerning the lengths of simple closed geodesics on a once-punctured torus, $\mathbf{T}$, with a complete finite-area hyperbolic structure. Let $\mathcal{C}$ be the set of isotopy classes of simple closed curves on $\mathbf{T}$. Given a hyperbolic structure, we can associate to each $\gamma \in \mathcal{C}$ the length, $l(\gamma)$, of the unique closed geodesic in the isotopy class. McShane's identity tells us that

$$\sum_{\gamma \in \mathcal{C}} \frac{1}{1 + e^{l(\gamma)}} = \frac{1}{2}.$$

In this paper we offer an alternative proof of McShane's identity using "trees of Markoff triples". This proof is somewhat more direct in that it avoids using the Birman-Series result from ergodic theory, as quoted in [6]. However, it lacks some of the geometrical motivation of the original.

The idea of constructing trees of Markoff triples goes back essentially to the original work of Markoff. Their relationship with hyperbolic geometry has since been explored by several authors. Most of this work has focused on integral Markoff triples in connection with diophantine approximation. For more discussion, see, for example, [5] and the references therein.

The fundamental group of $\mathbf{T}$ is a free group, Γ, on two generators, a_0 and b_0, so that the commutator $[a_0, b_0]$ is peripheral. We may identify the group of isometries of the hyperbolic plane, $\mathbf{H}^2$, with the group $PSL(2, \mathbf{R})$. Thus, any complete finite-area hyperbolic structure on $\mathbf{T}$ is given by $\mathbf{H}^2/\rho(\Gamma)$ where $\rho : \Gamma \longrightarrow PSL(2, \mathbf{R})$ is a discrete faithful representation such that $\rho([a_0, b_0])$ is parabolic. Now each homotopy class of closed curves on $\mathbf{T}$ corresponds to an equivalence class in Γ, where two group elements, g and h, are equivalent if g is conjugate to h or to h^{-1}. If the homotopy class γ is represented by a group element $g \in \Gamma$, then we may verify that $\mathrm{tr}(\rho(g)) = 2\cosh(l(\gamma)/2)$. Thus McShane's identity can be interpreted in terms of the traces of such a representation.

This observation allows one to generalise to certain representations of Γ into the group $PSL(2, \mathbf{C})$. Of particular interest are the quasifuchsian representations (geometrically finite without accidental parabolics). Since quasifuchsian space is connected, McShane's identity extends unchanged to all quasifuchsian representations by analytic continuity (given the appropriate interpretation of "complex length"). One can also give a direct proof using complex Markoff triples, though the analysis becomes a little more complicated.

† *To appear in Bull. London Math. Soc.*

Markoff triples of complex numbers, and their relationship to geometry, are studied in [3]. These ideas are also useful in understanding some aspects of the geometry of hyperbolic once-punctured torus bundles fibring over the circle. For example one can give variations of McShane's identity applicable to such manifolds [2]. They are also used in the classification of arithmetic once-punctured torus bundles [4].

I would like to thank Greg McShane and Colin Maclachlan for helpful conversations. I am also indebted to Igor Rivin for suggesting a means of simplifying the analysis at the end of the paper.

2. The theorem.

We shall rephrase McShane's identity in terms of trees of Markoff triples.

By a "binary tree" we mean a simplicial tree all of whose vertices have degree 3. Let Σ be a binary tree properly embedded in the plane. We write $V(\Sigma)$ and $E(\Sigma)$ respectively for the set of vertices and the set of edges of Σ. By a *complementary region* we mean the closure of a connected component of the complement of Σ. We write Ω for the set of complementary regions.

Although we are only interested in the combinatorial structure of thus set-up, it is natural to imagine Σ as being dual to the regular tessellation of the hyperbolic plane by ideal triangles. In this way, the elements of Ω are in bijective correspondence with the "rational points" on the circle at infinity — the vertices of the tessellation. In particular, we see that Ω carries a natural dense cyclic order.

There is also a natural bijective correspondence between Ω and $\mathcal{C}$; "natural" at least once we have chosen to identify some element of Ω with some element of $\mathcal{C}$. There are several ways to describe this bijection, see for example [3] or [4]. It is characterised by the following property.

Suppose that $e \in E(\Sigma)$. Then e meets four regions $X, Y, Z, W \in \Omega$, in such a way that $e = X \cap Y$ and such that the singletons $e \cap Z$ and $e \cap W$ are the two endpoints of e. Now we may choose a free basis $\{a, b\}$ for the group $\Gamma = \pi_1(\mathbf{T})$ so that the simple closed curves in $\mathcal{C}$ corresponding respectively to X, Y, Z and W are represented by the group elements a, b, ab and ab^{-1}.

Now suppose that $\rho : \Gamma \longrightarrow PSL(2, \mathbf{R})$ is a discrete faithful representation such that $\rho([a_0, b_0])$ is parabolic. It follows that $\rho([a, b])$ is parabolic for any free generators a and b as above. Note that the sign of the trace of $\rho([a, b])$ is well-defined. In fact $tr(\rho([a, b])) = -2$ since it is well known that if two elements of $PSL(2, \mathbf{R})$ have the trace of their commutator equal to 2, then they generate an elementary group; (see [1, Theorem 4.3.5]).

Given such a representation we define a map

$$\phi : \Omega \longrightarrow (0, \infty)$$

by setting $\phi(X)$ to be the trace of $\rho(g)$ where $g \in \Gamma$ represents the simple closed curve corresponding to $X \in \Omega$. Note that we take the trace to be positive. In fact, since $\rho(g)$ is hyperbolic, we must have $\phi(X) > 2$. Now if $v \in V(\Sigma)$ is the intersection of $X, Y, Z \in \Omega$, we have the "vertex relation"

$$x^2 + y^2 + z^2 = xyz \qquad (*)$$

where $x = \phi(X)$, $y = \phi(Y)$ and $z = \phi(Z)$. A triple (x, y, z) satisfying $(*)$ is called a "Markoff triple". If the edge $e \in E(\Sigma)$ meets $X, Y, Z, W \in \Omega$ in the manner described earlier, then we have the edge relation

$$xy = z + w \tag{$**$}$$

where $w = \phi(W)$. These relations follow from the trace identities in $SL(2, \mathbf{R})$:

$$2 + \operatorname{tr}[A, B] = (\operatorname{tr} A)^2 + (\operatorname{tr} B)^2 + (\operatorname{tr} AB)^2 - \operatorname{tr} A \operatorname{tr} B \operatorname{tr} AB$$

$$\operatorname{tr} A \operatorname{tr} B = \operatorname{tr} AB + \operatorname{tr} AB^{-1}.$$

(We need to observe that the representation ρ can be lifted to a representation into $SL(2, \mathbf{R})$ in which the trace of every simple closed curve is positive.)

Definition : A map $\phi : \Omega \longrightarrow (0, \infty)$ is a *Markoff map* if it satisfies the vertex relation $(*)$ at each vertex of Σ and the edge relation $(**)$ at each edge of Σ.

Note that, in fact, it is sufficient that $(*)$ be satisfied at a single vertex of Σ — the relations at all other vertices following inductively using $(**)$. Note also that a Markoff map must satisfy $\phi(X) > 2$ for all $X \in \Omega$.

(In fact, it turns out that any Markoff map arises in this way from some representation ρ. This can be seen using Jørgensen's normalisation: see [4].)

Suppose that the Markoff map ϕ arises from the representation ρ. Suppose $X \in \Omega$ corresponds to $\gamma \in \mathcal{C}$, which is, in turn, represented by an element $g \in \Gamma$. We have $\phi(X) = \operatorname{tr} \rho(g) = 2\cosh(l(\gamma)/2)$, and so $\frac{1}{1+e^{l(\gamma)}} = h(\phi(X))$, where h is defined by

$$h(x) = \frac{1}{2}\left(1 - \sqrt{1 - \frac{4}{x^2}}\right)$$

for all $x > 2$.

We may now state McShane's identity as follows:

Theorem : *If $\phi : \Omega \longrightarrow (0, \infty)$ is a Markoff map, then*

$$\sum_{X \in \Omega} h(\phi(X)) = \frac{1}{2}.$$

The rest of this paper will be devoted to proving this result. No explicit use will be made of hyperbolic geometry.

3. The proof.

We have defined $E(\Sigma)$ as the set if (undirected) edges of Σ. We can think of a *directed* edge, $\vec{e}$, as an ordered pair of incident vertices, referred to as the *head* and *tail* of $\vec{e}$. Thus we can speak of $\vec{e}$ as directed from the tail to the head. We shall write $-\vec{e}$ for the same edge directed in the opposite direction. We shall always use the convention that e denotes the underlying edge in $E(\Sigma)$. We write $\vec{E}(\Sigma)$ for the set of directed edges of Σ.

Given a finite subtree, T, of Σ, we define the finite set $C(T) \subseteq \vec{E}(\Sigma)$ of directed edges of Σ by saying that $\vec{e} \in C(T)$ if and only if $e \cap T$ consists of a single point, namely the head of $\vec{e}$. We shall call a finite set of directed edges *circular* if it arises in this way.

Given a point $a \in V(\Sigma)$, and some $n \in \mathbf{N}$, set $T_n(a)$ to be the tree spanned by all those vertices a distance at most n from a, and let $\Omega_n(a)$ be the set of all complementary regions meeting $T_n(a)$. In other words $X \in \Omega_n(a)$ if it can be joined to a by a path in Σ with at most n edges. Let $C_n(a) = C(T_n(a))$. Note that if $X \in \Omega_n(a)$ then the boundary of X contains precisely two edges of $C_n(a)$. Conversely, the two regions either side of an edge of $C_n(a)$ both lie in $\Omega_n(a)$. Note also that $\Omega_0(a)$ consists of the three regions of Ω meeting a, and $C_0(a)$ conists of the three directed edges which have a as their head.

Convention : From now on we fix a Markoff map $\phi : \Omega \longrightarrow (0, \infty)$. Whenever $X, Y, Z, W \in \Omega$ we shall write $x = \phi(X)$, $y = \phi(Y)$, $z = \phi(Z)$ and $w = \phi(W)$.

Given a directed edge $\vec{e} \in \vec{E}(\Sigma)$, we define $\psi(\vec{e}) = z/xy$, where $e = X \cap Y$ and e meets Z in the head of $\vec{e}$. In these terms, the edge and vertex relations may be rephrased as

$$\psi(\vec{e}) + \psi(-\vec{e}) = 1$$

for any directed edge $\vec{e}$, and

$$\psi(\vec{e}_1) + \psi(\vec{e}_2) + \psi(\vec{e}_3) = 1,$$

where $\vec{e}_1$, $\vec{e}_2$ and $\vec{e}_3$ are the three directed edges with their heads at a vertex v, i.e. $\Omega_0(v) = \{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$.

By combining these relations, we find that:

Lemma 1 : *For any circular set $C \subseteq \vec{E}(\Sigma)$, we have*

$$\sum_{\vec{e} \in C} \psi(\vec{e}) = 1.$$

$\diamondsuit$

Recall the definition of $h : (2, \infty) \longrightarrow (0, \infty)$ by $h(x) = \frac{1}{2}(1 - \sqrt{1 - 4/x^2})$. Note that for all x we have $\frac{1}{x^2} \leq h(x) \leq \frac{2}{x^2}$. If $x, y \in (2, \infty)$ satisfy $\frac{1}{x^2} + \frac{1}{y^2} < \frac{1}{4}$, then we define

$$h(x, y) = h\left(\left(\frac{1}{x^2} + \frac{1}{y^2}\right)^{-\frac{1}{2}}\right) = \frac{1}{2}\left(1 - \sqrt{1 - 4\left(\frac{1}{x^2} + \frac{1}{y^2}\right)}\right).$$

One may easily verify that $h(x)+h(y) \leq h(x,y)$. Note also that for any fixed x, $h(x,y) \to h(x)$ as $y \to \infty$.

Now suppose that $\vec{e} \in \vec{E}(\Sigma)$ meets $X, Y, Z, W \in \Omega$ so that $X \cap Y = e$, and $e \cap Z$ and $e \cap W$ are the head and tail, respectively, of $\vec{e}$. If $z \leq w$, then we may verify from the vertex condition that $\psi(\vec{e}) = \frac{z}{xy} = h(x,y)$. If $z \geq w$, then $\psi(\vec{e}) = \frac{z}{xy} \geq \frac{w}{xy} = h(x,y)$. So, either way, we have $\psi(\vec{e}) \geq h(x,y) \geq h(x) + h(y)$.

Now, suppose $a \in V(\Sigma)$ and $n \in \mathbf{N}$. We see that

$$\sum_{X \in \Omega_n(a)} h(\phi(X)) \leq \frac{1}{2} \sum_{\vec{e} \in C_n(a)} \psi(\vec{e}) = \frac{1}{2}$$

by Lemma 1. Since this is true for all n, we conclude that:

Lemma 2 :

$$\sum_{X \in \Omega} h(\phi(X)) \leq \frac{1}{2}.$$

$\diamondsuit$

To obtain equality, we need a more careful analysis.

To this end, we shall choose a particular direction $\vec{e}$ for each edge $e \in E(\Sigma)$ as follows. Let $X, Y, Z, W \in \Omega$ be such that $X \cap Y = e$, and so that $e \cap Z$ and $e \cap W$ are the incident vertices. If $z < w$, we take the head of $\vec{e}$ to be $e \cap Z$, whereas if $w < z$, we take its head to be $e \cap W$. If it happens that $z = w$, then we direct e arbitrarily. Note that the statements $z \leq w$, $2z \leq xy$ and $2w \geq xy$ are all equivalent.

Lemma 3 : *There is at most one directed edge leaving any given vertex of Σ.*

Proof : Suppose $X, Y, Z \in \Omega$ meet at a single vertex v. Suppose (for contradiction) that the edges $X \cap Z$ and $Y \cap Z$ both have their tails at v. Then $xz \leq 2y$ and $yz \leq 2x$, so $z^2 \leq 4$. This contradicts the fact that $z > 2$. $\diamondsuit$

Lemma 4 : *There does not exist a path in Σ consisting of an infinite sequence of distinct edges, $e_1, e_2, e_3, \ldots$, such that $\vec{e}_i$ is directed towards $\vec{e}_{i+1}$ for all $i \in \mathbf{N}$.*

Proof : This is not hard to see directly — a detailed proof in a more general context is given in [3]. We shall not reproduce the argument here. Note that in the case that interests us, we can assume that the Markoff map ϕ is derived from a discrete representation ρ. Such a representation must have a discrete length spectrum, which shows there can be no such infinite path in this case. $\diamondsuit$

It follows from Lemmas 3 and 4 that there is a (unique) vertex $a \in V(\Sigma)$ such that every edge of Σ is directed towards a. In particular, we see that $\min\{\phi(X) \mid X \in \Omega\}$ is attained on $\Omega_0(a)$. Let $\mu = \min\{\phi(X) \mid X \in \Omega\} > 2$.

Given $\mu > 2$, it is easily seen that there some $M > 0$ such that $h(x,y) \leq h(x) + Mh(y)$ for all $x \geq \mu$ and y such that $x^{-2} + y^{-2} < 1/4$.

Now, given any $\epsilon > 0$, it follows from Lemma 2, that there is some $n \in \mathbf{N}$ such that

$$\sum_{X \in \Omega \setminus \Omega_n(a)} h(\phi(X)) \leq \epsilon / M.$$

Suppose $\vec{e} \in C_{n+1}(a)$. Then, $e = X \cap Y$ with $X \in \Omega_n(a)$ and $Y \in \Omega_{n+1}(a) \setminus \Omega_n(a)$. Since the direction on $\vec{e}$ agrees with that arising from ϕ in the manner we defined earlier, we have that $\psi(\vec{e}) = h(x, y)$, and so $\psi(\vec{e}) \leq h(x) + M h(y)$. Summing over all $\vec{e} \in C_{n+1}(a)$, using Lemma 1, we obtain:

$$1 = \sum_{\vec{e} \in C_{n+1}(a)} \psi(\vec{e}) \leq 2 \sum_{X \in \Omega_n(a)} h(\phi(X)) + 2M \sum_{X \in \Omega_{n+1}(a) \setminus \Omega_n(a)} h(\phi(X)).$$

Thus,

$$\sum_{X \in \Omega} h(\phi(X)) \geq \sum_{X \in \Omega_n(a)} h(\phi(X)) \geq \frac{1}{2} - M(\epsilon/M) = \frac{1}{2} - \epsilon.$$

Letting $\epsilon \to 0$, the theorem follows.

References.

[1] A.F.Beardon, *The geometry of discrete groups* : Graduate Texts in Mathematics No. 91, Springer-Verlag (1983).

[2] B.H.Bowditch, *A variation of McShane's identity for once-punctured torus bundles* : preprint, Southampton (1993).

[3] B.H.Bowditch, *Markoff triples and quasifuchsian groups* : preprint, Southampton (1995).

[4] B.H.Bowditch, C.Maclachlan, A.W.Reid, *Arithmetic hyperbolic surface bundles* : Math. Annalen **302** (1995) 31–62.

[5] T.W.Cusick, M.E.Flahive, *The Markoff and Lagrange Spectra* : Amer. Math. Soc. Mathematical Surveys and Monographs No. 30, Providence, Rhode Island (1989).

[6] G.McShane, *A remarkable identity for lengths of curves* : PhD Thesis, University of Warwick (1991).