ISSN 1364-0380 (on line) 1465-3060 (printed)

Geometry & Topology
Volume 9 (2005) 1539–1601
Published: 10 August 2005

A better proof of the Goldman–Parker conjecture

Richard Evan Schwartz

Department of Mathematics, University of Maryland
Collage Park, MD 20742, USA

Email: `res at math dot brown dot edu`

URL: `http://www.math.brown.edu/~res/`

Abstract

The Goldman–Parker Conjecture classifies the complex hyperbolic $\boldsymbol{C}$–reflection ideal triangle groups up to discreteness. We proved the Goldman–Parker Conjecture in [6] using a rigorous computer-assisted proof. In this paper we give a new and improved proof of the Goldman–Parker Conjecture. While the proof relies on the computer for extensive guidance, the proof itself is traditional.

AMS Classification numbers Primary: 20F67

Secondary: 20F65, 20F55

Keywords: Hyperbolic, complex reflection group, ideal triangle group, Goldman–Parker conjecture

Proposed: Benson Farb Received: 8 February 2005
Seconded: David Gabai, Martin Bridson Revised: 2 July 2005

1 Introduction

Let $\boldsymbol{H}^2$ be the hyperbolic plane. Let G denote the usual reflection ideal triangle group acting on $\boldsymbol{H}^2$. The standard generators of G are $\iota_0, \iota_1, \iota_2$.

$PU(2,1)$ is the holomorphic isometry group of $\boldsymbol{CH}^2$, the *complex hyperbolic plane*. See Section 2 for more details. A *$\boldsymbol{C}$-reflection* is an order 2 element of $PU(2,1)$ which is conjugate to the element which has the action $(z,w) \to (z,-w)$. A *complex hyperbolic ideal triangle group representation* is a representation of G which maps the generators to $\boldsymbol{C}$-reflections, and the products of pairs of generators to parabolic elements. Let $\mathrm{Rep}(G)$ denote the set of such representations, modulo conjugacy. It turns out that $\mathrm{Rep}(G)$ is a half-open interval, naturally parametrized by $s \in [0,\infty)$. See Section 2.

Define

$$\underline{s} = \sqrt{105/3}; \qquad \overline{s} = \sqrt{125/3}; \tag{1}$$

In [5], Goldman and Parker introduced $\mathrm{Rep}(G)$ (using different notation) and proved that ρ_s is a discrete embedding if $s \in [0,\underline{s}]$. They conjectured that ρ_s is a discrete embedding iff $\rho_s(\iota_0\iota_1\iota_2)$ is not an elliptic element of $PU(2,1)$. This corresponds to parameters $s \in [0,\overline{s}]$. We took care of the interval $(\underline{s},\overline{s}]$ in [6], using a rigorous computer-assisted proof, together with some new constructions in complex hyperbolic geometry. However, the proof in [6] is extremely complicated and requires massive computations.

The purpose of this paper is to give a new and improved proof of the Goldman–Parker Conjecture. Our new proof is based on an idea we worked out, to a limited extent, in [7, Sections 8–10]. To each of the three generators $I_{j,s} = \rho_s(\iota_j)$ we will associate an piecewise analytic sphere $\Sigma_{j,s}$. We call $\Sigma_{j,s}$ a *loxodromic $\boldsymbol{R}$-sphere*. Our construction is such that $I_{j,s}(\Sigma_{j,s}) = \Sigma_{j,s}$ and that $I_{j,s}$ interchanges the two components of $S^3 - \Sigma_{j,s}$. The key step in our argument is showing that $\Sigma_{i,s} \cap \Sigma_{j,s}$ is a contractible set—the union of 2 arcs arranged in a 'T' pattern—for $i \neq j$, and that $\Sigma_{j,s}$ is embedded. This sets up a version of the familiar ping-pong lemma, and it follows readily from this picture that ρ_s is a discrete embedding.

In [7, Sections 8–10] we established the intersection and embedding properties of our spheres for all $s \in [\overline{s}-\epsilon,\overline{s})$, using a perturbative argument. However, we couldn't get an effective estimate on ϵ back then. Here, in Section 3, we develop a theory for loxodromic $\boldsymbol{R}$-spheres and use it to establish the two desired properties for all $s \in [\underline{s},\overline{s})$. Pictures like Figure 4.2 indicate that our construction works for all $s \in [0,\overline{s})$. However, there are certain technical details we could not overcome when trying to deal with parameters outside the range $[\underline{s}, \overline{s})$.

We wrote a Java applet which illustrates this paper in great detail and, in particular, lets the reader plot pictures like Figure 4.2 for all parameter values. The paper is independent of the applet, but the applet greatly enhances the paper because it lets the reader see visually the objects we refer to here mainly with symbols. We encourage the reader to use the applet while reading the paper. One can access the applet from my website. The applet provides massive hands-on evidence that our construction works for all $s \in [0, \overline{s}]$. In fact, most of our proof works for all $s \in [0, \overline{s}]$ but there are certain technical estimates we rely on that do not hold over the whole range of parameters.

Since I wrote [6] 7 years ago, there has been considerable development of complex hyperbolic discrete groups. Some of us feel that all the new technology—eg, [1], [7], [3], [8]—should reprove the Goldman–Parker Conjecture without too much pain. Nonetheless, a new proof has never appeared and I thought that this paper would be of interest. Also, I never liked my proof in [6] and have wanted a better proof for a long time.

This paper divides into 2 halves. The first half is organized like this:

- Section 2: background;
- Section 3: theory of loxodromic $\boldsymbol{R}$–spheres;
- Section 4: the proof.

The proof requires a handful of technical estimates, which we make in Sections 5–7.

The technical estimates all concern the location in S^3 of a certain collection of arcs of circles. There is a 1–parameter family of these arcs and one can readily compute their positions numerically. You can see from my applet (or from your own experiments) that these estimates hold by a wide margin and are blatantly true for parameters in $[\underline{s}, \overline{s}]$. The original version of this paper had computer-aided estimates on the locations of these arcs. At the request of the referee of this paper, these computer-aided proofs have been replaced with analytic calculations.

The analytic calculations done in the paper are in part based on a brilliant algebraic idea due to the anonymous[1] referee. The idea can be summarized by

[1]Eventually I guessed that the referee was John Parker. You can tell the lion by his claw.

saying that one should introduce the parameter

$$x = \frac{e^2 + |e|^2 + \overline{e}^2}{1 - |e|^2}$$

and then write all relevant quantities in terms of x. (See subsections 5.1–5.2 for details.) Here e (which is not to be confused with the base of the natural log) is one of the coordinates of an eigenvector of the word $\rho(\iota_1\iota_0\iota_2)$.

I would like to thank Elisha Falbel, Bill Goldman and John Parker for many conversations, over the years, about complex hyperbolic geometry. Also, I would like to thank the University of Maryland, the Institute for Advanced Study, the National Science Foundation (Grant DMS-0305047) and the John Simon Guggenheim Memorial Foundation, for their generous support.

2 Background

2.1 Complex hyperbolic geometry

[2] and [4] are good references for complex hyperbolic geometry. [8] also has a good introduction.

2.1.1 The ball model

$\boldsymbol{C}^{2,1}$ is a copy of the vector space $\boldsymbol{C}^3$ equipped with the Hermitian form

$$\langle u, v\rangle = u_1\overline{v}_1 + u_2\overline{v}_2 - u_3\overline{v}_3 \tag{2}$$

$\boldsymbol{CH}^2$ and its ideal boundary are respectively the projective images, in the complex projective plane $\boldsymbol{CP}^2$, of

$$N_- = \{v \in \boldsymbol{C}^{2,1} |\ \langle v, v\rangle < 0\}; \qquad N_0 = \{v \in \boldsymbol{C}^{2,1} |\ \langle v, v\rangle = 0\} \tag{3}$$

(The set N_+ has a similar definition.) The projectivization map

$$(v_1, v_2, v_3) \to (v_1/v_3, v_2/v_3) \tag{4}$$

takes N_- and N_0 respectively to the open unit ball and unit sphere in $\boldsymbol{C}^2$. Henceforth we identify $\boldsymbol{CH}^2$ with the open unit ball. $\boldsymbol{CH}^2$ is called the *complex hyperbolic plane*. It is a symmetric space of negative curvature.

2.1.2 Slices

There are two kinds of totally geodesic 2–planes in $\boldsymbol{CH}^2$:

- The $\boldsymbol{R}$*–slices* are 2–planes, $PU(2,1)$–equivalent to $\boldsymbol{RH}^2 = \boldsymbol{R}^2 \cap \boldsymbol{CH}^2$.
- The $\boldsymbol{C}$*–slices* are 2–planes, $PU(2,1)$–equivalent to $\boldsymbol{CH}^1 = \boldsymbol{CH}^2 \cap \boldsymbol{C}^1$.

Let $\boldsymbol{F}$ stand either for $\boldsymbol{R}$ or $\boldsymbol{C}$. The accumulation set on S^3, of an $\boldsymbol{F}$–slice, is called an $\boldsymbol{F}$*–circle*. An $\boldsymbol{F}$–reflection is an involution in $\mathrm{Isom}(\boldsymbol{CH}^2)$ whose fixed point set is an $\boldsymbol{F}$–circle. The map $(z,w) \to (z,-w)$ is a prototypical $\boldsymbol{C}$–reflection and the map $(z,w) \to (\overline{z},\overline{w})$ is a prototypical $\boldsymbol{R}$–reflection. The $\boldsymbol{F}$–slice determines the $\boldsymbol{F}$–reflection and conversely.

2.1.3 Isometries

$SU(2,1)$ is the $\langle,\rangle$ preserving subgroup of $SL_3(\boldsymbol{C})$, the special complex linear group. $PU(2,1)$ is the projectivization of $SU(2,1)$. Elements of $PU(2,1)$ act isometrically on $\boldsymbol{CH}^2$ and are classified according to the usual scheme for groups acting on negatively curved spaces. Loxodromic elements move every point of $\boldsymbol{CH}^2$ greater than some $\epsilon > 0$; elliptic elements fix a point in $\boldsymbol{CH}^2$; and the remaining elements are parabolic.

We now discuss $\boldsymbol{C}$–reflections in more detail. Given a vector $C \subset N_+$ we define

$$I_C(U) = -U + \frac{2\langle U,C\rangle}{\langle C,C\rangle}C. \tag{5}$$

I_C is an involution fixing C and $I_C \in SU(2,1)$. See [4, page 70]. The element of $PU(2,1)$ corresponding to I_C is a $\boldsymbol{C}$–reflection. Every $\boldsymbol{C}$–reflection is conjugate to the map $(z,w) \to (z,-w)$ discussed above. $\boldsymbol{C}$–reflections are also called *complex reflections*.

2.2 Heisenberg space

2.2.1 Basic definitions

In the ball model, $\boldsymbol{CH}^2$ is a ball sitting inside complex projective space $\boldsymbol{CP}^2$. For this discussion we fix some $p \in S^3$, the ideal boundary of $\boldsymbol{CH}^2$. There exists a complex projective automorphism β of $\boldsymbol{CP}^2$ which maps p to a point in $\boldsymbol{CP}^2 - \boldsymbol{C}^2$ and which identifies $\boldsymbol{CH}^2$ with the *Siegel domain*:

$$Z = \{(z_1,z_2)|\ 2\mathrm{Re}(z_1) < -|z_2|^2\} \subset \boldsymbol{C}^2 \subset \boldsymbol{CP}^2 \tag{6}$$

We write $\infty = \beta(p)$ in this case. The isometries of $\boldsymbol{CH}^2$ which fix ∞ act as complex linear automorphisms of Z. The set ∂Z is characterized as the set of null vectors relative to the Hermitian form

$$\langle u, v\rangle' = u_1\overline{v}_3 + u_2\overline{v}_2 + u_3\overline{v}_1. \tag{7}$$

We call $\mathcal{H} = \boldsymbol{C} \times \boldsymbol{R}$ *Heisenberg space*. $\mathcal{H}$ is equipped with a group law:

$$(\zeta_1, t_1)\cdot(\zeta_2, t_2) = (\zeta_1 + \zeta_2, t_1 + t_2 + 2\mathrm{Im}(\overline{\zeta}_1\zeta_2)) \tag{8}$$

There is a natural map from ∂Z to $\mathcal{H}$, given by

$$\mu(z_1, z_2) = (\frac{z_2}{\sqrt{2}}, \mathrm{Im}(z_1)). \tag{9}$$

The inverse map is given by

$$(z, t) \to (-|z|^2 + it, z\sqrt{2}). \tag{10}$$

A *Heisenberg stereographic projection from* p is a map $\boldsymbol{B}$: $S^3 - \{p\} \to \mathcal{H}$ of the form $\mu \circ \beta$ where β is as above. We write $\infty = \boldsymbol{B}(p)$ in this case. We will somewhat abuse terminology and speak of elements of $PU(2,1)$ acting on $\mathcal{H}$. We mean that the conjugate of an element, by Heisenberg stereographic projection, acts on $\mathcal{H}$. If such a map stabilizes ∞, it acts as an affine map of $\mathcal{H}$.

- The $\boldsymbol{C}$–circles in $\mathcal{H}$ which contain ∞ all have the form $(\{z\} \times \boldsymbol{R}) \cup \infty$. The remaining $\boldsymbol{C}$–circles are ellipses which project to circles in $\boldsymbol{C}$. The plane containing the ellipse is the *contact plane* based at the center of mass of the ellipse. See below for more detail.
- The $\boldsymbol{R}$–circles which contain ∞ are straight lines. One of these $\boldsymbol{R}$–circles is $(\boldsymbol{R} \times \{0\}) \cup \infty$. The bounded $\boldsymbol{R}$–circles in $\mathcal{H}$ are such that their projections to $\boldsymbol{C}$ are lemniscates.

2.2.2 The contact distribution

The set of complex lines tangent to S^3 forms a $PU(2,1)$–invariant contact distribution on S^3. The $\boldsymbol{R}$–circles are tangent to this distribution and the $\boldsymbol{C}$–circles are transverse to it. The image of the contact distribution, under Heisenberg stereographic projection, is a contact distrubition on $\mathcal{H}$. It is defined as the kernel of the 1 form $dt + 2(xdy - ydx)$, when points in $\mathcal{H}$ are written as $(x + iy, t)$. Compare [4, page 124], Any element of $PU(2,1)$ acting on $\mathcal{H}$ respects this contact distribution. Each plane in the distribution is called a *contact plane*.

Area Principle Suppose α is a piecewise smooth curve in $\mathcal{H}$, tangent to the contact distribution, such that $\pi(\alpha)$ is a closed loop. Then the height difference—meaning the difference in the t–coordinates—between the two endpoints of α is 4 times the signed area of the compact region enclosed by $\pi(\alpha)$. This is basically Green's theorem. Compare [4, Section 4]. Call this principle the *area principle*.

2.3 Spinal spheres

2.3.1 Basic definitions

Basic information about bisectors and spinal spheres can be found in [4]. Here we recall some of the basics.

A *bisector* is a subset of $\boldsymbol{CH}^2$ of the form $\{x \in \boldsymbol{CH}^2 | \ d(x,p) = d(x,q)\}$. Here $p \neq q$ are two distinct points in $\boldsymbol{CH}^2$ and d is the complex hyperbolic metric. A *spinal sphere* is the ideal boundary of a bisector. Every two spinal spheres are equivalent under $PU(2,1)$, even though this is not immediately obvious. Equivalently, a spinal sphere is any set of the form

$$\boldsymbol{B}^{-1}((\boldsymbol{C} \times \{0\}) \cup \infty).$$

Here $\boldsymbol{B}$ is a Heisenberg stereographic projection. Thus,

$$S = (\boldsymbol{C} \times \{0\}) \cup \infty$$

is a model in $\mathcal{H}$ for a spinal sphere. From the second definition we see some of the structure of spinal spheres. Here are some objects associated to S:

- S has a singular foliation by $\boldsymbol{C}$–circles. The leaves are given by $C_r \times \{0\}$ where C_r is a circle of radius r centered at the origin. The singular points are 0 and ∞. We call this the *$\boldsymbol{C}$–foliation*.
- S has a singular foliation by $\boldsymbol{R}$–circles. The leaves are horizontal lines through the origin. The singular points are again 0 and ∞. We call this the *$\boldsymbol{R}$–foliation*.
- The singular points 0 and ∞ are called the *poles* of S.
- The *spine* of S is defined as the $\boldsymbol{C}$–circle containing the poles. In our case, the spine is $(\{0\} \times \boldsymbol{R}) \cup \infty$. Note that the spine of S only intersects S at the singular points.

Any other spinal sphere inherits this structure, by symmetry. The two foliations on a spinal sphere look like lines of lattitude and longitude on a globe. A spinal sphere is uniquely determined by its poles. Two spinal spheres are *cospinal* if they have the same spine.

2.3.2 Generic tangencies with spinal spheres

In this section we prove a useful technical result about how an $\boldsymbol{R}$–circle intersects a spinal sphere. Let $\pi\colon \mathcal{H} \to \boldsymbol{C}$ be projection. The next result is illustrated in Figure 2.1.

Lemma 2.1 *Let $S = (\boldsymbol{C} \times \{0\}) \cup \infty$ as above. Let γ be a finite $\boldsymbol{R}$–circle in $\mathcal{H}$. Suppose*

- *γ is tangent to S at $p \neq 0$.*
- *The line $L \subset \boldsymbol{C}$, containing 0 and $\pi(p)$, is not tangent to $\pi(\gamma)$ at the double point of $\pi(\gamma)$.*

Then γ has first but not second order contact with S. Moreover, a neighborhood of p in γ lies on one side of S.

Proof Since γ is tangent to S at p, and γ is also tangent to the contact distribution, and the contact distribution is not tangent to S at p, we see that γ is tangent to the $\boldsymbol{R}$–circle of S which contains p. This $\boldsymbol{R}$–circle is exactly $L \times \{0\}$. But then L is tangent to $\pi(\gamma)$ at $\pi(p)$. Figure 2.1 shows the situation when a lobe of $\pi(\gamma)$ surrounds 0. The other topological possibility has the same proof. The basic idea of the proof is that L does not have second order contact with $\pi(\gamma)$ at $\pi(p)$, by convexity.

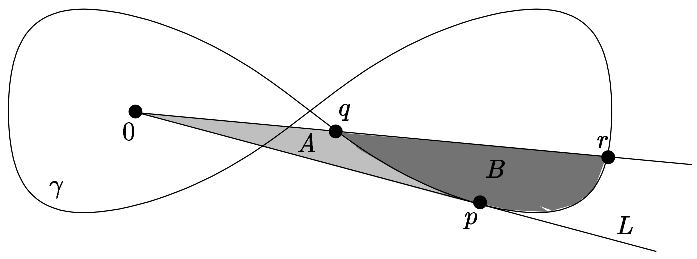

Figure 2.1

We first apply the area principle to the integral curve α made from two horizontal line segments and a portion of γ, so that $\pi(\alpha)$ bounds the lightly shaded region A shown in Figure 2.1. From the area principle we see that the height of q is positive, and also a quadratic function of the Euclidean distance from $\pi(q)$ to $\pi(p)$. The quadratic dependence comes from the strict convexity of $\pi(\gamma)$ in a neighborhood of $\pi(p)$.

A similar argument works when we take the relevant integral curve which projects to the region $A \cup B$. We see that the height of the point r is positive, and also a quadratic function of the Euclidean distance from $\pi(r)$ to $\pi(p)$. Now we know that points on γ, on either side of p, rise quadratically up and away from S. □

Corollary 2.2 *Suppose γ links the spine of a spinal sphere Σ and γ is tangent to Σ at some point p. Then γ has first but not second order contact with Σ at p and a neighborhood of p in γ lies to one side of Σ.*

Proof When we normalize so that $\Sigma = S$. then the spine of Σ projects to 0. One of the lobes of $\pi(\gamma)$ surrounds 0, and p projects to some nonzero point. In short, we have the picture in Figure 2.1, and the hypotheses of the previous result are forced. □

2.4 Equations for $\boldsymbol{C}$–circles

Suppose that C is a $\boldsymbol{C}$–circle in $\mathcal{H}$ which links $\{0\} \times \boldsymbol{R}$. Let $\pi\colon \mathcal{H} \to \boldsymbol{C}$ be projection as above. Then $\pi(C)$ is a circle in $\boldsymbol{C}$ which surrounds 0. As in [7, section 2] we study $\Psi_*(C)$, where Ψ_* is the map

$$\Psi_*(z,t) = (\arg z, t). \tag{11}$$

Define

$$r = \text{radius}(\pi(C)); \qquad d = |\text{center}(\pi(C))|; \qquad A = (r/d)^2. \tag{12}$$

We only define A when $d > 0$. We call A the *aspect* of C. Note that $A > 1$.

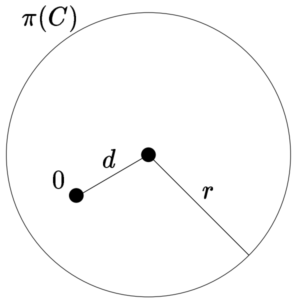

Figure 2.2

Lemma 2.3 *Let A be the aspect of C. Up to scaling and rotation $\Psi_*(C)$ is the graph of*

$$f_A(t) = \sin(t)(\cos(t) + \sqrt{A - \sin^2(t)}) \tag{13}$$

Proof We normalize so that $(1,0)$ is the center of mass of C. Then $d = 1$ and C is contained in the contact plane through $(1,0)$. This plane is spanned by $(1,0)$ and $(i,2)$. Let C_θ be the point on C such that the line through 0 and $\pi(C_\theta)$ makes an angle of θ with the x axis. Then $\Psi_*(C)$ is the graph of the function $\theta \to \text{height}(C_\theta) = 2y$, where $C_\theta = (x,y)$. Our formula comes from solving the equations $(x-1)^2 + y^2 = r^2$ and $x = y\cot(\theta)$ in terms of y. □

Lemma 2.4 *If $A \geq 9$ then f_A'' is negative on $(0,\pi)$ and positive on $(\pi, 2\pi)$.*

Proof We compute that

$$\frac{d}{dA} f_A''(t) = \frac{A(2 - A + \cos(2t))\sin(t)}{2(A - \sin^2(t))^{5/2}} = g_A(t)\sin(t), \tag{14}$$

where $g_A(t) < 0$. Hence $\frac{d}{dA} f_A''$ is negative on $(0,\pi)$ and positive on $(\pi, 2\pi)$. We just need to prove that f_9'' is negative on $(0,\pi)$ and positive on $(\pi,2\pi)$. We compute that $f_9''(\pi/2) = -7/\sqrt{8} < 0$. Thus, we just need to see that $f_9''(t) = 0$ only at $t = 0$ and $t = \pi$. Setting $u = \cos(t)$ we compute[2]

$$f_9'' = \frac{-4h(u)\sin(t)}{(8+u^2)^{3/2}}; \qquad h(u) = 14 + 12u^2 + u^4 + 8u\sqrt{8+u^2} + u^3\sqrt{8+u^2}. \tag{15}$$

For $u \in (-1,1]$ we have

$$h(u) \geq 14 + 12u^2 + u^4 + 24u + 3u^3 = (1+u)(14 + 10u + 2u^2 + u^3) > 0.$$

This shows that $f_9''(t) \neq 0$ if $t \notin \{0, \pi\}$. □

Remark The preceding lemma is essentially the same as Lemma 4.11 of [3], with the variable change $\phi = t + \pi/2$.

[2]We differentiate the function $\sin(t)(\cos(t) + \sqrt{(A-1) + \cos^2(t)})$, which is a re-writing of f_A, using Mathematica [10].

2.5 Ideal triangle groups

2.5.1 The basic definition

We will use the same set-up as in [6]. Given $s \in [0, \infty)$ we define

$$\beta_s = \frac{s+i}{\sqrt{2+2s^2}}. \tag{16}$$

Sometimes we write β instead of β_s, when the dependence is clear. As we showed in [6], every ideal triangle in S^3 is conjugate to a triangle with vertices

$$p_0 = (\beta, \overline{\beta}); \qquad p_1 = (\beta, \beta); \qquad p_2 = (\overline{\beta}, \overline{\beta}). \tag{17}$$

In brief, the idea is that we can apply an element of $PU(2,1)$ so that all three vertices of our ideal triangle lie on the Clifford torus

$$\{(z,w) | \ |z| = |w|\} \subset S^3$$

and then we can rotate the Clifford torus until the points are as above.

Let I_j be the $\boldsymbol{C}$–reflection which fixes p_{j-1} and p_{j+1}. We compute that the elements I_0, I_1, I_2 are given by

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}; \quad \begin{bmatrix} -1 & 0 & 0 \\ 0 & 3 & -4\overline{\beta} \\ 0 & 4\beta & -3 \end{bmatrix}; \quad \begin{bmatrix} 3 & 0 & -4\beta \\ 0 & -1 & 0 \\ 4\overline{\beta} & 0 & -3 \end{bmatrix}, \tag{18}$$

Letting $g_0 = I_1 I_0 I_2$ we compute

$$g_0 = \begin{bmatrix} 0 & -1 & 0 \\ -A_1 & 0 & A_2 \\ -A_2 & 0 & -\overline{A}_1 \end{bmatrix}; \quad A_1 = \frac{s+17i}{s+i}; \quad A_2 = \frac{12\sqrt{2}i}{\sqrt{1+s^2}}. \tag{19}$$

A direct computation, for example using the result on [4, page 201], shows that g is loxodromic for $s \in [0, \overline{s})$ and parabolic for $\overline{s}$. See [5].

As we mentioned in the introduction, we are mainly interested in the case when $s \in (\underline{s}, \overline{s}]$, though many of our constructions work for $s \in [0, \underline{s}]$ as well. Actually, we will carry out most of our constructions for $s \in (\underline{s}, \overline{s})$, because there are now several good discreteness proofs for the case $\overline{s}$. See [7] and [3].

2.5.2 Some associated objects

When $s < \overline{s}$, the element g_0 is loxodromic. In this case g_0 stabilizes a pair (E_0, Q_0), where E_0 is a $\boldsymbol{C}$–circle containing the fixed points of g_0 and Q_0 is an arc of E_0 bounded by the two fixed points. Of the two possible arcs, we choose Q_0 so that it varies continuously with the parameter and shrinks to a point as $s \to \overline{s}$. The curve

$$\rho_0 = \{(u\beta, \overline{u\beta})|\ u \in S^1\} \tag{20}$$

is an $\boldsymbol{R}$–circle fixed by the map $(z,w) = (\overline{w}, \overline{z})$. This map interchanges p_1 and p_2 and fixes p_0. In short ρ_0 is an $\boldsymbol{R}$–circle of symmetry for our configuration.

One can define Q_1, Q_2, etc. by cycling the indices mod 3. The objects

$$(C_j, E_j, p_j, Q_j); \qquad j = 0, 1, 2. \tag{21}$$

are the elementary objects of interest to us. Figure 2.3 shows those of the objects which lie on the Clifford torus, when the Clifford torus is drawn as a square torus (in "arg-arg coordinates"). The black dots are the points of $E_0 \cap \rho_0$.

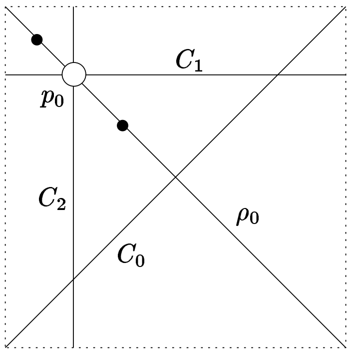

Figure 2.3

3 Loxodromic $\boldsymbol{R}$–spheres

3.1 The cospinal foliation

Our constructions are all based on the pair (E_0, Q_0) from subsection 2.5, though we could take any pair (E_0, Q_0) and make the same definitions. Again, E_0 is a $\boldsymbol{C}$–circle and Q_0 is a proper arc of E_0. Note that Q_0 determines E_0 uniquely. We include E_0 in our notation for emphasis.

Let $p, q \in E_0$ be two distinct points. The pair (p,q) is *harmonic* with respect to (E_0, Q_0) if the geodesic connecting p to q in $\boldsymbol{CH}^2$ is perpendicular to the geodesic connecting the endpoints of Q_0. (Both these geodesics lie in the $\boldsymbol{C}$–slice bounded by E_0.) A spinal sphere S is *harmonic* w.r.t. (E_0, Q_0) if the poles of S lie in E_0 and are harmonic w.r.t (E_0, Q_0). An $\boldsymbol{R}$–arc α is *harmonic* w.r.t. (E_0, Q_0) if the endpoints of α are harmonic w.r.t. (E_0, Q_0). Every harmonic $\boldsymbol{R}$–arc is contained in a harmonic spinal sphere.

- Let $S(E_0, Q_0)$ denote the union of all spinal spheres which are harmonic with respect to (E_0, Q_0). We call $S(E_0, Q_0)$ the *cospinal foliation*.
- Let $R(E_0, Q_0)$ denote the union of all $\boldsymbol{R}$–arcs which are harmonic with respect to (E_0, Q_0).
- Let $G(E_0, Q_0) \subset PU(2,1)$ denote the group which fixes the endpoints of Q_0. Then $G(E_0, Q_0)$ acts transitively the elements in $S(E_0, Q_0)$ and simply transitively on the elements of $R(E_0, Q_0)$.

To see a picture we work in $\mathcal{H}$ and normalize so that $E_0 = (\{0\} \times \boldsymbol{R}) \cup \infty$ and Q_0 is the unbounded arc whose endpoints are $(0, \pm 1)$. We call this *standard position*. In this case (p, q) is harmonic with respect to (E_0, Q_0) iff $p = (0, r)$ and $q = (0, r^{-1})$. We include the possibility that $r = 0$, so that $r^{-1} = \infty$. All the spinal spheres of interest to us are bounded in $\mathcal{H}$, except for $(\boldsymbol{C} \times \{0\}) \cup \infty$, which corresponds to the case $r = 0$.

Lemma 3.1 *Every bounded spinal sphere in $S(E_0, Q_0)$ is a convex surface of revolution.*

Proof All such spinal spheres are surfaces of revolution, by symmetry. Moreover, all such spinal spheres are affine images of the so-called *unit spinal sphere*, which has poles $(0, \pm 1)$. The unit spinal sphere satisfies the equation $|z|^4 + t^2 = 1$ and hence is convex. See [4, page 159]. Being affine images of a convex set, the other spinal spheres of interest to us are also convex. □

Lemma 3.2 *Every two distinct spinal spheres in $S(E_0, Q_0)$ are disjoint.*

Proof Let S_1 and S_2 be two distinct spinal spheres in $S(E_0, Q_0)$. Using the action of $G(E_0, Q_0)$ we can arrange that $S_1 = (\boldsymbol{C} \times \{0\}) \cup \infty$ and that the endpoints of S_2 are $(r, 0)$ and $(r^{-1}, 0)$, with $r \neq 0$. But then S_2 lies either entirely in the upper half space, or entirely the lower half space. In either case S_2 is disjoint from S_1. □

3.2 Loxodromic $\boldsymbol{R}$–cones and $\boldsymbol{R}$–spheres

Given $x \in S^3 - E_0$ there is a unique element $\alpha \in R(E_0, Q_0)$ such that $x \in \alpha$. We let $\Sigma(E_0, Q_0; x)$ denote the portion of α which connects x to a point in Q_0. Given a subset $S \subset S^3 - E_0$ we define

$$\Sigma(E_0, Q_0; S) = \bigcup_{x \in S} \Sigma(E_0, Q_0; x). \tag{22}$$

We call $\Sigma(E_0, Q_0; S)$ the *loxodromic cone* on S.

Let C_1 be a $\boldsymbol{C}$–circle which links E_0. Technical Lemma I (Lemma 4.1) establishes this linking property when (C_1, E_0) are as in subsection 2.5.) Let $(E_2, Q_2) = I_1(E_0, Q_0)$. Here I_1 is the $\boldsymbol{C}$–reflection fixing C_1. We say that a *loxodromic $\boldsymbol{R}$–sphere* is an object of the form

$$\Sigma_1 = \Sigma(E_0, Q_0; C_1) \cup \Sigma(E_2, Q_2; C_1) \tag{23}$$

If the $\boldsymbol{C}$–slice bounded by C_1 is perpendicular to the $\boldsymbol{C}$–slice bounded by E_0 then C_1 lies in one of the elements of $S(E_0, Q_0)$ and Σ_1 is a spinal sphere. In general Σ_1 is not a spinal sphere. In Section 4 we will show that the loxodromic $\boldsymbol{R}$–spheres of interest to us are embedded spheres but not spinal spheres.

We are interested in the case when Σ_1 is not a spinal sphere. Henceforth we assume that Σ_1 is not a spinal sphere. In this case we call Σ_1 *generic*.

Lemma 3.3 *Let Σ_1 be a generic $\boldsymbol{R}$–sphere. There exists a unique $\boldsymbol{R}$–circle R_1 such that R_1 intersects E_0, C_1, E_2 each in two points. Also, $R_1 \subset \Sigma_1$ and the $\boldsymbol{R}$–reflection in R_1 is a symmetry of Σ_1. We call R_1 the $\boldsymbol{R}$–axis of Σ_1.*

Proof We put (E_0, Q_0) in standard position. Recall that $G(E_0, Q_0)$ is the stabilizer subgroup of Q_0 which preserves the endpoints. Using the action of $G(E_0, Q_0)$ we can normalize so that the center of mass of C_1 is $(r, 0)$ for some $r > 0$. Consider the $\boldsymbol{R}$–circle $R_1 = (\boldsymbol{R} \times \{0\}) \cup \infty$. By symmetry R_1 intersects C_1 twice. Also R_1 intersects E_0 twice. Finally, we have $I_1(R_1) = R_1$ by symmetry. Hence R_1 intersects E_2 twice. The $\boldsymbol{R}$–reflection J_1 in R_1 is an anti-holomorphic element preserving (E_0, Q_0) and C_1, and hence a symmetry of Σ_1. If there was some other axis R_1' then the composition of the $\boldsymbol{R}$–reflection symmetries J_1 and J_1' would a non-trivial element of $PU(2,1)$ preserving both $(E_0.Q_0)$ and C_1. But no such element exists. Hence R_1 is unique. $\square$

3.3 The elevation map

The set $R(E_0, Q_0)$ is topologically a cylinder. Since $G(E_0, Q_0)$ acts transitively on $R(E_0, Q_0)$, this cylinder admits a natural family of flat metrics. Put another way, we can write $R(E_0, Q_0) = \boldsymbol{R}/2\pi\boldsymbol{Z} \times \boldsymbol{R}$. The identification is unique up to post-composition with a map of the form $(x, y) \to (x + a, by + c)$. That is, the identification is unique up to *affine maps*.

There is a tautological map Ψ_0: $S^3 - E_0 \to R(E_0, Q_0)$ defined as follows: $\Psi_0(x)$ is the element of $R(E_0, Q_0)$ which contains x. If we identify $R(E_0, Q_0)$ with $\boldsymbol{R}/2\pi \times \boldsymbol{R}$ then we have nice coordinates for this map. Given $x \in S^3 - E$ we let $\widehat{x}$ be a lift of x. We define

$$\Psi_0(x) = \left(\arg \frac{\langle \widehat{x}, \widehat{E}_0 \rangle}{\sqrt{\langle \widehat{x}, \widehat{Q}_1 \rangle \langle \widehat{x}, \widehat{Q}_2 \rangle}}, \log \left| \frac{\langle \widehat{x}, \widehat{Q}_1 \rangle}{\langle \widehat{x}, \widehat{Q}_2 \rangle} \right| \right) \tag{24}$$

Here $\widehat{Q}_1$ and $\widehat{Q}_2$ are lifts of the endpoints of Q_0 and $\widehat{E}_0$ is a polar vector of E_0. This is to say that $\langle \widehat{E}_0, \widehat{Y} \rangle = 0$ whenever $\widehat{Y}$ is a lift of a point $Y \in E_0$.

Remarks

(i) It is possible to choose a well-defined branch of the square root in Equation 24. This is basically a topological fact, deriving from the fact that the map

$$f(x) \to \langle \widehat{x}, \widehat{Q}_1 \rangle \langle \widehat{x}, \widehat{Q}_2 \rangle$$

induces the map on homology

$$f_*\colon \boldsymbol{Z} = H_1(S^3 - E_0) \to H_1(\boldsymbol{C} - \{0\}) = \boldsymbol{Z}$$

which is multiplication by 2.

(ii) Different choices of lifts lead to maps which differ by post-composition with affine maps.

(iii) To see that Equation 24 works as claimed we compute that Ψ_0 conjugates $G(E_0, Q_0)$ to isometries of $\boldsymbol{R}/2\pi \times \boldsymbol{R}$. The point here is that $(\widehat{Q}_1, \widehat{Q}_2, \widehat{E}_0)$ is an eigenbasis for the elements of $G(E_0, Q_0)$.

Henceforth we set $\Psi = \Psi_0$. Note that $\Psi(x) = \Psi(y)$ iff x and y belong to the same element of $R(E_0, Q_0)$. Ψ maps the orbit $G(E_0, Q_0)(x)$ diffeomorphically onto $\boldsymbol{R}/2\pi\boldsymbol{Z} \times \boldsymbol{R}$. hence $d\Psi$ has rank 2 everywhere.

For any $x \in S^3 - E_0$ let Π_x denote the contact plane at x. Let

$$L_x = d\Psi(\Pi_x) \tag{25}$$

As we just mentioned, $d\Psi$ has full rank at at x but $d\Psi$ maps the 3–dimensional vector space $T_x(S^3)$ onto a 2–dimensional tangent space. The kernel of $d\Psi$ is the vector tangent to the element of $R(E_0, Q_0)$ through x. This kernel is therefore contained in Π_x. Hence L_x is a line. The following result captures some of the basic features of this situation.

Lemma 3.4 *Let γ be an $\boldsymbol{F}$–circle and let $x \in \gamma - E_0$ be a point.*

- *If $\boldsymbol{F} = \boldsymbol{R}$ and γ is not tangent to $\Sigma(E_0, Q_0; x)$ then $\Psi(\gamma)$ is a nonsingular curve at $\Psi(x)$ and the tangent line is L_x.*
- *If $\boldsymbol{F} = \boldsymbol{C}$ then $\Psi(\gamma)$ is nonsingular at $\Psi(x)$ and transverse to L_x.*

Proof When γ is an $\boldsymbol{R}$–circle, the tangent vector v to γ at x lies in Π_x but is not contained in the kernel of $d\Psi_x$. Part 1 of our lemma follows from this fact. When γ is a $\boldsymbol{C}$–circle, $v \notin \Pi_x$ and hence $d\Psi_x(v) \notin L_x$. □

Remark At this point, the reader anxious to see our main construction should skip to Section 4.

3.4 More details on slopes

From Lemma 3.4 we see that L_x tells us a great deal about what Ψ does to $\boldsymbol{R}$–circles and $\boldsymbol{C}$–circles. We now investigate this further. Let σ_x denote the slope of L_x. Of course σ_x depends on our choice of normalization, but the general statements we make are independent of normalization. Let S_0 denote the spinal sphere whose poles are ∂Q_0.

Lemma 3.5 *If $x \in S_0$ then L_x is a vertical line and hence σ_x is infinite. Otherwise σ_x is finite and nonzero.*

Proof Let $H \subset G(E_0, Q_0)$ denote the 1–parameter subgroup consisting of the pure loxodromic elements. These elements do not twist at all in the direction normal to the slice bounded by E_0. By symmetry Ψ maps the orbit $H(x)$ to a vertical line in $\boldsymbol{R}/2\pi\boldsymbol{Z} \times \boldsymbol{R}$. On the other hand $H(x)$ is tangent to Π_x iff $x \in S_0$. From this we see that σ_x is infinite iff $x \in S_0$. Now Ψ maps the $\boldsymbol{C}$–circles foliating the spinal spheres in $S(E_0, Q_0)$ to horizontal lines. From this fact, and from Lemma 3.4, we see that $\sigma_x \neq 0$.

Lemma 3.6 *Let $x, y \in S^3 - E_0$. Then $\sigma_x = \sigma_y$ iff x and y lie in the same $G(E_0, Q_0)$ orbit.*

Proof By symmetry we have $\sigma_x = \sigma_y$ if x and y are $G(E_0, Q_0)$ equivalent. We just have to establish the converse. Each $x \in S^3 - E_0$ determines a 1–parameter subgroup $H_x \subset G(E_0, Q_0)$ which has the property that the orbit $H_x(x)$ is integral to the contact structure. Then Ψ maps $H_x(x)$ to a geodesic on $\boldsymbol{R}/2\pi\boldsymbol{Z} \times \boldsymbol{R}$ which is tangent to L_x. It suffices to show that $H_x \neq H_y$ if x and y lie in different $G(E_0, Q_0)$ orbits. Suppose, for the sake of contradiction, that there are $G(E_0, Q_0)$–inequivalent x, y for which $H_x = H_y$. Using the action of $G(E_0, Q_0)$ we can arrange that $\Psi(x) = \Psi(y)$. Let $h \in H_x = H_y$. By symmetry we have $\Psi(h(x)) = h(y)$. But then we can make a closed quadrilateral, $\mathcal{Q}$ as follows:

- One side of $\mathcal{Q}$ is the portion of $H_x(x)$ which connects x to $h(x)$.
- One side of $\mathcal{Q}$ is the portion of $H_y(y)$ which connects y to $h(y)$.
- One side of $\mathcal{Q}$ is $\Sigma(E_0, Q_0; y) - \Sigma(E_0, Q_0; x)$.
- One side of $\mathcal{Q}$ is $\Sigma(E_0, Q_0; h(y)) - \Sigma(E_0, Q_0; h(x))$.

Here we are choosing x, y so that $\Sigma(E_0, Q_0; x) \subset \Sigma(E_0, Q_0; y)$. The shaded region in Figure 3.1 is the projection of $\mathcal{Q}$ to $\boldsymbol{C}$.

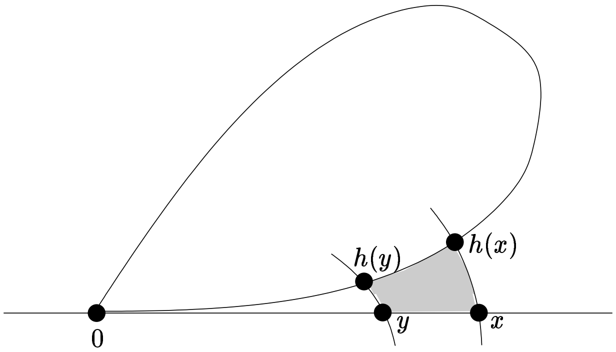

Figure 3.1

We normalize so that (E_0, Q_0) is the standard pair and $x = (r, 0)$ and $y = (s, 0)$. If we choose h close to the identity, then $\pi(\mathcal{Q})$ projects to an embedded quadrilateral in $\boldsymbol{C}$, as suggested by Figure 3.1. The point here is that the fibers of Ψ are lobes of lemniscates which have their double points at the origin. Since $\mathcal{Q}$ is integral to the contact structure, and yet a closed loop, we contradict the Area Principle of subsection 2.2.2. $\square$

Corollary 3.7 (Slope Principle) *Let $\gamma_1, \gamma_2 \in S^3 - E_0$ be two $\boldsymbol{R}$–arcs such that $\Psi(\gamma_1)$ and $\Psi(\gamma_2)$ are nonsingular at a point $x \in \boldsymbol{R}/2\pi\boldsymbol{Z} \times \boldsymbol{R}$, and tangent to each other at x. Then γ_1 and γ_2 intersect at some point $y \in \Psi^{-1}(x)$.*

Proof Each fiber of Ψ, including $\Psi^{-1}(x)$, intersects each orbit of $G(E_0, Q_0)$ in one point. Our result now follows from Lemma 3.4 and Lemma 3.6. □

We normalize Ψ so that σ_x is positive when S_0 separates Q_0 from x. In this case we call x *remote* from Q_0. Thus we can say that $\sigma_x \in (0, \infty)$ iff x is remote from S_0.

3.5 Images of $\boldsymbol{C}$–circles

Lemma 3.8 *Suppose C is a $\boldsymbol{C}$–circle which links E_0. Then $\Psi(C)$ is the graph of a function $\psi\colon \boldsymbol{R}/2\pi\boldsymbol{Z} \to \boldsymbol{R}$.*

Proof $\Psi(C)$ is a smooth loop by Lemma 3.4. We need to prove that $\Psi(C)$ is never vertical. Let $C = C_1 \cup C_2$, where C_1 is the closure of the remote points of C and C_2 is the complement. Let I be a $\boldsymbol{C}$–reflection in a $\boldsymbol{C}$–circle contained in S_0. Then I interchanges the two components of $S^3 - S_0$, and Ψ conjugates I to a reflection in a horizontal line of $\boldsymbol{R}/2\pi\boldsymbol{Z} \times \boldsymbol{R}$. (This is seen by choosing $\widehat{Q}_1$ and $\widehat{Q}_2$ in Equation 24 so that these vectors are swapped by I.) Hence $\Psi(I(C_2))$ is the image of $\Psi(C_2)$ reflected in a horizontal line. By symmetry, then, it suffices to show that $\Psi(C_1)$ is never vertical.

Let $x \in C_1$. We normalize so that (E_0, Q_0) is the standard pair and $x = (r, 0)$ for some $r > 0$. Since C links E, we can parameterize C as $C(\theta) = (z(\theta), t(\theta))$, where θ is the angle between the ray connecting 0 to $\pi(C(\theta))$ and $\boldsymbol{R}$. We are interested in $q = C(0)$.

If $t'(0) = 0$ then C is tangent to $S = (\boldsymbol{C} \times \{0\}) \cup \infty$ at q. But S is a member of $S(E_0, Q_0)$. Hence $\Psi(C)$ is horozontal at $\Psi(C(0))$. Conversely, if $\Psi(C)$ is horizontal at $\Psi(C(0))$ then C is tangent to S and hence $t'(0) = 0$.

If $t'(0) > 0$ then $w = d\Psi_q(C'(0))$ lies in the interior of the cone bounded by the x–axis and $L_{C(0)}$, a line whose slope is either positive or infinite. (The idea here is that the statement holds when the center of mass of C is near $(0, 0)$, and then Lemma 3.4 and the linking condition guarantee that the cone condition holds no matter how C varies.) Hence w is not vertical.

If $t'(0) < 0$ then the projection to $\boldsymbol{C}$ of the fiber of Ψ which contains $C(\epsilon)$ curves down and clockwise, as shown in Figure 3.2. But then the horizontal component of $w = d\Psi_q(\gamma'(0))$ exceeds 1 because the angle δ in Figure 3.2 exceeds the angle ϵ. Again w is not vertical. □

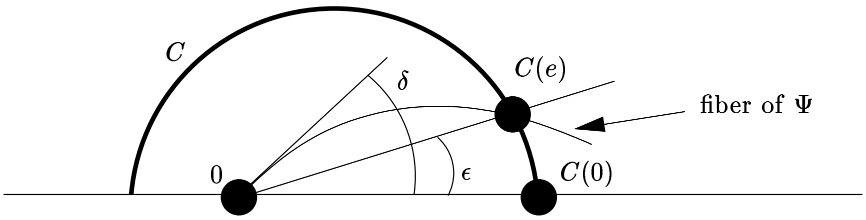

Figure 3.2

We say that C *generically links* E_0 if the $\boldsymbol{C}$–slice bounded by C is not perpendicular to the $\boldsymbol{C}$–slice bounded by E_0.

Lemma 3.9 *Suppose that C is a $\boldsymbol{C}$–circle which generically links E_0. Then $\Psi(C)$ is the graph of a function $\psi\colon S^1 \to \boldsymbol{R}$ which has one maximum and one mimimum.*

Proof Since C generically links E_0 the image $\Psi(C)$ is not contained in a horizontal line. The Ψ–preimages of horizontal lines are spinal spheres. Since C is not contained in any of these spinal spheres, C can intersect each of them at most twice. Hence $\Psi(C)$ intersects each horizontal line at most twice. Our result follows immediately. □

Corollary 3.10 *Suppose that C_1 generically links E_0. Then $\Sigma(E_0, Q_0; C_1)$ is an embedded topological disk, analytic away from $C_1 \cup Q_0$.*

Proof The set $\Sigma(E_0, Q_0; C_1) - Q_0$ is foliated by $\boldsymbol{R}$–arcs of the form $\gamma(\theta)$, where $\theta \in C_1$ is a point. Two arcs $\gamma(\theta_1)$ and $\gamma(\theta_2)$ are disjoint because $\Psi(\gamma(\theta_j)) = \Psi(\theta_j)$ and $\Psi(\theta_1) \neq \Psi(\theta_2)$. Moreover, the arcs vary analytically. Hence $\Sigma(E_0, Q_0; C_1) - Q_0$ is homeomorphic to an annulus (with one boundary component deleted) and analytic away from C_1.

There is a map $f\colon C_1 \to Q_0$ given as follows: $f(x)$ is defined to be the endpoint of $\Sigma(E_0, Q_0; x)$. If $f(C_1)$ is more than one point—as it is when C generically links E_0—then f is generically 2 to 1 and 1 to 1 at exactly two points. This follows from Lemma 3.9. Thus, f has the effect of folding C_1 in half over an arc of Q_0. Hence the $\boldsymbol{R}$–arcs foliating $\Sigma(E_0, Q_0; C_1)$ intersect Q in pairs, with two exceptions. Topologically, $\Sigma(E_0, Q_0; C_1)$ is obtained from an annulus by gluing the inner circle together by a folding map, as in Figure 3.3. From this description we see that $\Sigma(E_0, Q_0; C_1)$ is an embedded disk, analytic off of Q_0. □

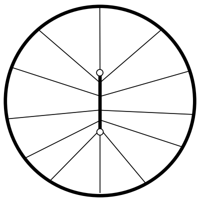

Figure 3.3

3.6 Images of linked $\boldsymbol{R}$–circles

Lemma 3.11 *Let γ be an $\boldsymbol{R}$–circle which links E_0. Then γ is tangent to exactly two spinal spheres in $S(E_0, Q_0)$.*

Proof We normalize so that (E_0, Q_0) is in standard position. Let Q_t denote the unbounded $\boldsymbol{C}$–arc whose endpoints are $(0, \pm(1+t))$. Then Q_t exits every compact subset of $\mathcal{H}$ as $t \to \infty$. Let $N(t)$ denote the number of elements of $S(E_0, Q_t)$ which are tangent to γ.

We define $S(E_0, Q_\infty)$ to be the collection of spinal spheres of the form $(\boldsymbol{C} \times \{s\}) \cup \infty$. Given this definition we can define $N(\infty)$ to be the number of spinal spheres in $S(E_0, Q_\infty)$ tangent to γ. Let's analyze $N(\infty)$ first. Let $\pi\colon \mathcal{H} \to \boldsymbol{C}$ be projection. Suppose γ is tangent to a horizontal spinal sphere $S_s = (\boldsymbol{C} \times \{s\}) \cup \infty$ at x. Since γ is also tangent to the contact plane Π_x, we see that γ is tangent to the line $\Pi_x \cap S_s$. But $\Pi_x \cap S_s$ is a horizontal line which intersects E_0. Hence $\pi(\gamma)$ is tangent at $\pi(x)$ to a line through the origin. But $\pi(\gamma)$ is a lemniscate, one of whose lobes surrounds the origin. See Figure 2.1. Hence there are only 2 lines through the origin which are tangent to $\pi(\gamma)$. Hence $N(\infty) = 2$.

Now fix some value of t. Since γ links E_0, Corollary 2.2 applies: If γ is tangent to a spinal sphere S of $\Sigma(E_0, Q_t)$ then γ locally lies to one side of S and has first but not second order contact with S. Moreover, as $t' \to t$, the spinal spheres of $S(E_0, Q_{t'})$ converge smoothly to the spinal spheres of $S(E_0, Q_t)$. These two properties imply that the tangency points vary continuously with t and cannot be created or destroyed as t changes. The two properties also hold at $t = \infty$.

From the discussion in the preceding paragraph we see that $N(t)$ is independent of t. Since $N(\infty) = 2$ we also have $N(1) = 2$. $\square$

Corollary 3.12 (Elevation Image) *Suppose that γ is an $\boldsymbol{R}$–circle which links E_0. Then $\Psi(\gamma)$ is the union of two non-singular arcs, each having nonzero slope at every point. The two non-singular arcs meet at two cusp points.*

Proof Everything but the statement about the cusps follows from Lemma 3.4, Lemma 3.5, and Lemma 3.11. The two cusps appear because γ locally lies on one side of each of the two spinal spheres to which it is tangent. $\square$

3.7 Linking of the poles

As above $I_1(E_0,Q_0)=(E_2,Q_2)$, where I_1 is the $\boldsymbol{C}$–reflection fixing the $\boldsymbol{C}$–circle C_1 which generically links E_0. Also R_1 is the $\boldsymbol{R}$–axis of the $\boldsymbol{R}$–sphere Σ_1 given in Equation 23.

Lemma 3.13 *E_0 and E_2 are linked.*

Proof We normalize so that (E_0,Q_0) is the standard pair and the center of mass of C_1 is $(r,0)$ for some $r>0$. Then $R_1=(\boldsymbol{R}\times\{0\})\cup\infty$. Let $A_1\subset R_1$ be the bounded interval whose endpoints are $C_1\cap R_1$. The center of A_1 is $(r,0)$. Since C_1 links E_0, we have $0\in A_1$. Since the center of A_1 is positive, 0 lies in the left half of A_1. Now $I_1|_{R_1}$ acts as a linear fractional transformation interchanging A_1 with R_1-A_1. But then $I_1(0,0)=(s,0)$ with $s<0$ and $I_1(\infty)=(r,0)$ with $r>0$. The two points of $E_2\cap R_1$ are $I_1(0,0)$ and $I_1(\infty)$. But these points separate $(0,0)$ from ∞ on R_1. Hence E_2 and E_0 are linked. $\square$

$\Psi(C_1)$ and $\Psi(E_2)$ have some symmetry: Let J_1 be the $\boldsymbol{R}$–reflection fixing R_1. Then Ψ conjugates J_1 to an isometric 180 degree rotation of $\boldsymbol{R}/2\pi\boldsymbol{Z}\times\boldsymbol{R}$. The fixed point set of this rotation is exactly $\Psi(R_1-E_0)$, which is a pair of points on the same horizontal level and π units around from each other. This rotation is a symmetry of $\Psi(C_1)$ and also of $\Psi(E_2)$. Figure 3.4 shows a picture of the three possibilities.

We say that Σ_1 is *interlaced* if the picture looks like the right hand side. That is, a vertical line separates the minimum of $\Psi(E_2)$ from the minimum of $\Psi(C_1)$. When we normalize the interlaced case as above, $(0,0)$ separates the center of mass of C_1 from the center of mass of E_2. From this we see that the set of interlaced $\boldsymbol{R}$–spheres is connected. In the interlaced case, the interlacing pattern of the extrema of $\Psi(C_1)$ and $\Psi(E_2)$ forces

$$\Psi(C_1)\cap\Psi(E_2)=\Psi(R_1-E_0). \tag{26}$$

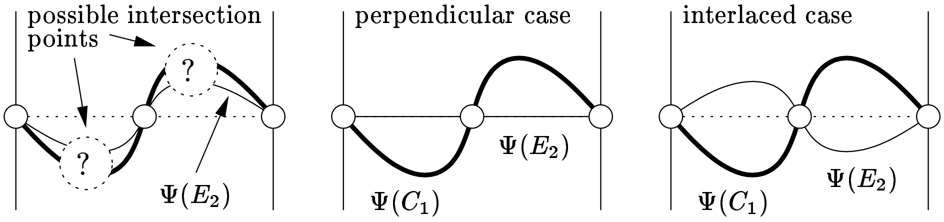

Figure 3.4

3.8 The Two Cusp Lemma

We say that an $\boldsymbol{R}$–circle $\widehat{\gamma}$ is *affiliated* with $\Sigma(E_0, Q_0; C_1)$ if $\widehat{\gamma}$ contains an $\boldsymbol{R}$–arc of the form $\Sigma(E_0, Q_0; x)$, for $x \in C_1$, but $\widehat{\gamma}$ is not the $\boldsymbol{R}$–axis R_1 of Σ_1. The purpose of this section is to prove the following result.

Lemma 3.14 (Two Cusp) *Suppose that Σ_1 is interlaced. Suppose that $\widehat{\gamma}$ is an $\boldsymbol{R}$–circle affiliated to $\Sigma(E_2, Q_2; C_1)$. Then $\Psi(\widehat{\gamma})$. is the union of two nonsingular arcs, each of which has everywhere nonzero slope. The two arcs are joined at two cusps.*

The Two Cusp Lemma is an immediate consequence of Lemma 3.15 below and the Elevation Image Lemma.

Lemma 3.15 *Suppose that Σ_1 is interlaced. Then every $\boldsymbol{R}$–circle affiliated to $\Sigma(E_0, q_0; C_1)$ links E_2.*

Proof By symmetry, every $\boldsymbol{R}$–circle affiliated to $\Sigma(E_0, q_0; C_1)$ links E_2 if and only if every $\boldsymbol{R}$–circle affiliated to $\Sigma(E_2, Q_2; C_1)$ links E_0.

Let $\widehat{\gamma}$ be an $\boldsymbol{R}$–circle affiliated to $\Sigma(E_2, Q_2; C_1)$. We claim that that $\widehat{\gamma} \cap E_0 = \emptyset$. Once we know this, we see that either all affiliates of $\Sigma(E_2, Q_2; C_1)$ link E_0 or all affiliates fail to link E_0. By continuity, the link/unlink option is independent of the choice of interlaced $\boldsymbol{R}$–sphere. We check explicitly, for one interlaced $\boldsymbol{R}$–sphere—eg, the one in Figure 4.5—that the link option holds for some of the affiliates. Hence the link option always holds.

It remains to establish our claim. By symmetry $\widehat{\gamma}' = I_{C_1}(\gamma)$ is affiliated with $\Sigma(E_0, Q_0; C_1)$. By construction, $\Psi(\gamma' - E_0)$ is a single point of $\Psi(C_1)$ and $\Psi(\gamma' - E_0) \in \Psi(R_1 - E_0)$ iff $R' = R_1$. The point is that Ψ is injective on C_1. Therefore

$$\Psi(\gamma' - E_0) \cap \Psi(R_1 - E_0) = \emptyset. \tag{27}$$

We have by hypotheses and Equation 26 that

$$\Psi(\gamma' - E_0) \cap \Psi(E_2) \subset \Psi(C_1) \cap \Psi(E_2) = \Psi(R_1 - E_0). \tag{28}$$

Combining Equations 27 and 28 we have

$$\Psi(\gamma' - E_0) \cap \Psi(E_2) = \emptyset. \tag{29}$$

Hence $(\gamma' - E_0) \cap E_2 = \emptyset$. Since $E_0 \cap E_2 = \emptyset$ we conclude that $\widehat{\gamma}' \cap E_2 = \emptyset$. Hence $\widehat{\gamma} \cap E_0 = \emptyset$. This establishes our claim. □

3.9 Asymmetry

Let Δ_0 denote the $\boldsymbol{C}$–slice which bounds E_0. Let S_0 denote the spinal sphere whose poles are Q_0. Let R_1 denote the $\boldsymbol{R}$–axis of the $\boldsymbol{R}$–sphere Σ_1 given in Equation 23. Again recall that $I_1(E_0, Q_0) = (E_2, Q_0)$. We we will assume explicitly that E_1 and E_2 are generically linked. Hence E_i and E_j are also generically linked.

Let η: $S^3 \to \Delta_0$ denote orthogonal projection. The generic linking condition implies that $\eta(E_j)$ is a circle (rather than a point) for $j = 1, 2$. Let $\Theta_1 = \eta(E_0)$. For $j = 0, 2$ let Θ_j be the circle which is perpendicular to $\eta(E_j)$ and contains the endpoints of $\eta(Q_j)$. Note that Θ_0 and Θ_1 intersect at right angles, since $\Theta_0 = \eta(S_0)$ and $\Theta_1 = \eta(E_0)$.Some of these objects are drawn in Figure 3.5. Say that Σ_1 is *asymmetric* if $\Theta_0 \cap \Theta_1 \cap \Theta_2 = \emptyset$. This is the generic case. The goal of this section is to prove:

Lemma 3.16 (Asymmetry Lemma) *Suppose Σ_1 is asymmetric and interlaced. Let $x, y \in E_2$ be two points which are harmonic with respect to Q_2. Then $\Psi(x)$ and $\Psi(y)$ lie on the same horizontal line in $\boldsymbol{R}/2\pi\boldsymbol{Z} \times \boldsymbol{R}$ iff $x, y \in R_1$.*

Proof We first list some basic properties of the map η.

- $\eta(E_0)$ is the boundary of Δ_0 and $\eta(Q_0)$ is an arc of $\eta(E_0)$. This follows from the fact that η is the identity on E_0.
- If C is a $\boldsymbol{C}$–circle which is disjoint from E_0 then $\eta(C)$ is a circle contained in the interior of E_0. The restriction of η to C is a linear fractional transformation. This property comes from the fact that η is holomorphic on complex lines.
- If S is a spinal sphere whose spine is E_0, then $\eta(S)$ is a geodesic in Δ_0. In particular, $\gamma_0 = \eta(S_0)$ is the geodesic whose endpoints are $\eta(Q_0)$. Indeed, an alternate definition of a spinal sphere is the preimage of such a geodesic under η. See [4].

- η maps each spinal sphere in the cospinal foliation to geodesics perpendicular to $\eta(S_0)$. This follows from symmetry: namely that η conjugates $G(E_0.Q_0)$ to isometries of Δ_0 which fix both endpoints of $\eta(Q_0)$.
- η maps the $\boldsymbol{R}$–axis of Σ_1 to a geodesic γ_1 which is simultaneously perpendicular to $\eta(S_0)$ and $\eta(C_1)$ and $\eta(Q_2)$. Again this follows from symmetry: The $\boldsymbol{R}$–reflection in the $\boldsymbol{R}$–axis of Σ_1 preserves both C_1 and Q_2. Indeed, the isometric reflection in γ_1 stabilizes $\eta(Q_0)$ and $\eta(Q_2)$ and $\eta(C_1)$ and Δ_0.

Remark The reader can see all these properties in action using my Applet.

Now we turn to the main argument in the proof of the Asymmetry Lemma. If $x, y \in R_1$ then $\Psi(x)$ and $\Psi(y)$ are precisely the two symmetry points of $\Psi(E_2)$ and $\Psi(C_1)$ discussed in Lemma 3.15, and then $\Psi(x)$ and $\Psi(y)$ lie on the same horizontal line.

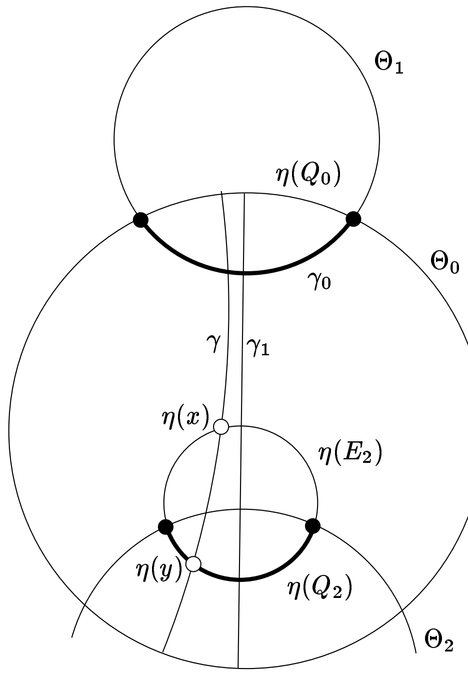

Figure 3.5

Suppose, conversely, that $\Psi(x)$ and $\Psi(y)$ lie on the same horizontal line. This means that $\eta(x)$ and $\eta(y)$ lie on the same geodesic γ of Δ_0, where γ is perpendicular to both Θ_0 and Θ_1. By assumption $x, y \in E_2$ are harmonic with respect to Q_2. Since the restriction of η to E_2 is a linear fractional transformation, the points $\eta(x)$ and $\eta(y)$ are in harmonic position with respect to $\eta(Q_2)$. So, we conclude that the geodesic γ has the following properties:

- γ is perpendicular to Θ_0 since γ is a geodesic.
- γ is perpendicular to γ_0, the geodesic connecting the endpoints of $\gamma(Q_0)$. Hence γ is perpendicular to Θ_1.
- γ intersects $\eta(E_2)$ in two points which are in harmonic position with respect to $\eta(Q_2)$. But then γ is perpendicular to Θ_2, the circle which is perpendicular to $\eta(E_2)$ and contains $\eta(Q_2)$ in its endpoints.

Figure 3.5 shows a picture. The lemma below says that there is only one geodesic which has this property, and this geodesic is $\gamma_1 = \eta(R_1)$. Hence $\Psi(x)$ and $\Psi(y)$ lie on the same horizontal line as $\Psi(R_1 - E_0)$. This forces $x, y \in R_1$. □

Lemma 3.17 *Let $\Theta_0, \Theta_1, \Theta_2$ be 3 circles in $\boldsymbol{C} \cup \infty$. Suppose that $\Theta_0 \cap \Theta_1$ is a pair of points and $\Theta_0 \cap \Theta_1 \cap \Theta_2 = \emptyset$. Then there is at most one circle which is simultaneously perpendicular to Θ_j for $j = 0, 1, 2$.*

Proof We normalize by a Moebius transformation so that Θ_0 and Θ_1 are lines through the origin. Then a circle in $\boldsymbol{C}$ is perpendicular to Θ_0 and Θ_1 iff this circle is centered at the origin. By assumption, Θ_2 is a finite circle in $\boldsymbol{C}$ which does not contain the origin. From here it is easy to see that at most one circle, centered at the origin, can be perpendicular to Θ_2. □

We end this chapter with a result which relates symmetry and asymmetry to the image of the arc $\Psi(Q_2)$.

Lemma 3.18 *Σ_1 is symmetric if and only if the maximum and minimum heights of $\Psi(E_2)$ occur at the endpoints of $\Psi(Q_2)$.*

Proof If Σ_1 is symmetric then the geodesics in Δ perpendicular to Θ_0 and Θ_1 and containing the endpoints of $\eta(Q_2)$ are tangent to $\eta(E_2)$. Recalling the 4th property of η mentioned above, the tangency property translates exactly into the statement that the height of $\Psi|_{E_2}$ takes on its maxima and minima at the endpoints of Q_2. The converse is proved simply by running the argument in reverse. □

3.10 Remoteness

We say that Σ_1 is *remote* if every point of $\Sigma(E_2, Q_2, C_1)$ is remote from Q_0. In this case $\alpha = \Psi(\gamma)$ has everywhere positive slope, when γ is an $\boldsymbol{R}$–arc of $\Sigma(E_2, Q_2; C_1)$. In this section we give a technical criterion for remoteness. We work in $\mathcal{H}$.

Lemma 3.19 *Σ_1 is remote provided:*

- *C_1 has aspect at least 9.*
- *The endpoints of Q_0 have the form $(0, \pm u)$, with $u \geq 4$.*
- *$\pi(C_1)$ has radius 1.*
- *C_1 is centered on the real axis.*

Proof Given our bound on the aspect, the center of C_1 is at most $1/3$ from 0. Hence every point of C_1 is at most $2/3$ from $\boldsymbol{C} \times \{0\}$. Also, there is a spinal sphere S_1, containing C_1 such that

$$S_1 \subset \pi(C_1) \times [-5/3, 5/3].$$

To obtain S_1 we simply take the spinal sphere with poles $(0, \pm 1)$ and left translate by less than $1/3$ along $\boldsymbol{R} \times \{0\}$.

The poles of S_0 are $(0, \pm u)$, with $u > 4$. It is easy to that the bounded portion of this huge (and convex) set contains S_1 in its interior. Let γ be an $\boldsymbol{R}$–arc of $\Sigma(E_0, Q_0; C_1)$. Then $I_1(\gamma)$ is an $\boldsymbol{R}$–arc of $\Sigma(E_2, Q_2; C_1)$. We claim that γ intersects S_1 only at its endpoint. In this case $I(\gamma)$ is contained in the bounded portion of $S_1 = I_1(S_1)$, which is in turn contained in the bounded portion of S_0. Hence $I(\gamma)$ is remote.

To finish our proof we need to establish our claim. Let $\widehat{\gamma}$ be the fiber of Ψ containing γ. Then $\widehat{\gamma}$ is harmonic with respect to (E_0, Q_0) and $\pi(\widehat{\gamma})$ is one lobe of a lemniscate. Let $(0, t_1)$ and $(0, t_2)$ be the two endpoints of $\widehat{\gamma}$, with $t_1 < t_2$. Without loss of generality assume that $t_1 > 0$. Then $\widehat{\gamma}$ rises up from its lower endpoint until it intersects C_1. Hence $t_1 \in [0, 2/3]$. Also $t_1 t_2 = u^2 > 14$. Hence $t_2 > 21$. Hence $\widehat{\gamma}$ rises up at least 20. The projection $\pi(\widehat{\gamma})$ is a huge lemniscate. From all this information we can see that γ only intersects S_1 at its endpoint: The only points $x \in \widehat{\gamma}$ with $\pi(x) \in \pi(C_1)$ have height greater than $5/3$. □

4 The proof

4.1 Main construction

Let $s \in [\underline{s}, \overline{s})$. Let $(C_j, E_j, Q_j, R_j, p_j)$ be as in subsection 2.5. (Actually, we defined R_j in Section 3, as the $\boldsymbol{R}$–axis of Σ_j.) Also define

$$Q_{21} = Q_2 \cap \Sigma_1; \qquad Q_{12} = Q_1 \cap \Sigma_2. \tag{30}$$

All our objects depend on a parameter s, though we typically suppress s from our notation.

Our proof includes several technical lemmas whose proofs will be given in Sections 5–7. It is to be understood that these results are only proved for parameters in $[\underline{s}, \overline{s}]$.

Lemma 4.1 (Technical Lemma I) *The following is true:*

(1) *E_0 and C_1 are linked.*

(2) *E_0 is normalized to be $(\{0 \times \boldsymbol{R}\}) \cup \infty$ in Heisenberg space then the aspect A of C_1 is at least 9.*

(3) *Suppose we normalize so that $C_1 \cap C_2 = (1, 0)$ in $\mathcal{H}$ and $E_0 = \{0\} \times \boldsymbol{R}$ and the map $(z, t) \to (\overline{z}, -t)$ swaps C_1 and C_2. Then 0 is closer to $1 = \pi(p_0)$ than it is to the other intersection point of $\pi(C_1)$ and $\pi(C_2)$.*

Lemma 4.2 (Technical Lemma II) *The following is true:*

(1) *Σ_1 satisfies the criteria of Lemma 3.19. Hence Σ_1 and Σ_2 are remote.*

(2) *The curve $\Psi(Q_2)$ has negative slope, even at the endpoints.*

Lemma 4.3 (Technical Lemma III) *The following is true:*

(1) *A horizontal line in $\boldsymbol{R}/2\pi\boldsymbol{Z} \times \boldsymbol{R}$ separates $\Psi(Q_{21})$ from $\Psi(C_2)$, with $\Psi(Q_{21})$ lying on top.*

(2) *If we normalize as in item 3 of Technical Lemma I then the center of C_1 lies above all points of C_2. Likewise the center of C_2 lies below all points of C_1.*

Now we are ready for our main construction. By symmetry C_i and E_j are linked for $i \neq j$. We define

$$\Sigma_{ij} = \Sigma(E_j, Q_j; C_i). \tag{31}$$

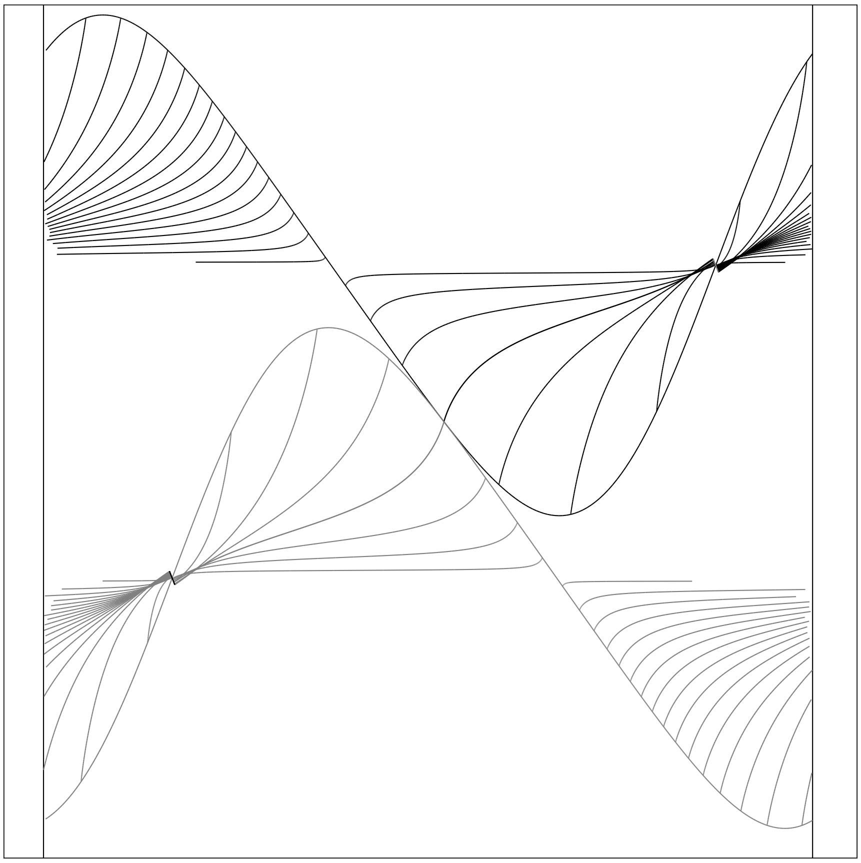

Figure 4.1

Then $\Sigma_i = \Sigma_{ij} \cup \Sigma_{ik}$ is an $\boldsymbol{R}$–sphere because I_i interchanges (E_j, Q_j) and (E_k, Q_k). Here i, j, k are distinct indices. Let $\Psi = \Psi_0$ be the map from Section 3.

Figure 4.1 shows $\Psi(\Sigma_1 - E_0)$ in black and $\Psi(\Sigma_2 - E_0)$ in grey for the parameter $\underline{s}$. All the black curves terminate on the tiny grey arc $\Psi(Q_{21})$ and all the grey curves terminate on the tiny black arc $\Psi(Q_{12})$. The Technical Lemma III says that the tiny grey arc lies above the grey curve $\Psi(C_2)$. This is obvious from the picture.

Figure 4.2 shows the same picture for $s = 1$. Even though the parameter value $s = 1$ is outside the interval of interest to us, we include the picture because the main features are more dramatic. Notice that the Technical Lemma III remains true even at this parameter. (However, our proof breaks down.)

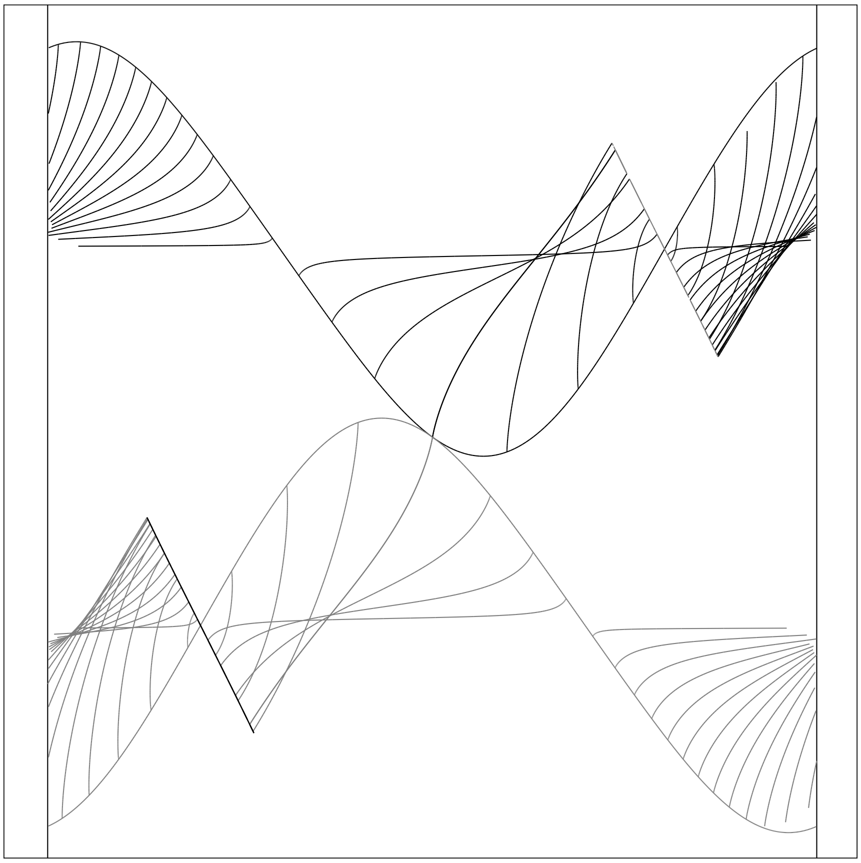

Figure 4.2

Recall that $\Sigma_1 = \Sigma_{10} \cup \Sigma_{12}$. The $\boldsymbol{R}$–arcs foliating Σ_{10} are mapped to individual points on $\Psi(C_1)$. The $\boldsymbol{R}$–arcs foliating Σ_{12} are mapped to black curves connecting $\Psi(C_1)$ to the black arc $\Psi(Q_{21})$. Here, in general

$$Q_{ij} = Q_i \cap \Sigma_{ji}; \qquad i \neq j. \tag{32}$$

The point here is that the $\boldsymbol{R}$–arcs foliating $\Sigma(E_2, Q_2; C_1)$ start on C_1 and end on a subset of Q_2. The point $\Psi(p_0)$ is the point of tangency between the black curve $\psi(C_1)$ and the grey curve $\psi(C_2)$.

Remark Using our applet, the reader can see the picture at any parameter, and can recolor each individual object, so as to see in a more direct way what part of the picture corresponds to what object.

We are going to draw two conclusions from the pictures. The main thrust of our proof is verifying that the pictures have the claimed property.

Contractible Intersections Figures 4.1 and 4.2 indicate that $\Psi(\Sigma_1 - E_0)$ and $\Psi(\Sigma_2 - E_0)$ only intersect in a single point at the parameters $s = \underline{s}$ and $s = 1$ respectively. Hence $A = (\Sigma_1 - E_0) \cap (\Sigma_2 - E_0)$ is a single arc. Note that $\Sigma_j \cap E_0$ is a connected arc $Q_j^* \subset Q_j$. The intersection $A^* = Q_1^* \cap Q_2^*$ is another arc, which intersects A in a single point. Hence $\Sigma_1 \cap \Sigma_2 = A \cup A^*$, the union of two arcs arranged in a T pattern—a contractible set.

Embedded Spheres Figures 4.1 and 4.2 also suggest that Σ_1 and Σ_2 are embedded. We work this out here. First, we have $\Psi(\Sigma_{10} - E_0) \subset \Psi(C_1)$ whereas the nontrivial arcs of $\Psi(\Sigma_{12} - E_0)$ only intersects $\Psi(C_1)$ at one point. Therefore

$$\Sigma_{10} \cap \Sigma_{12} \subset C_1 \cup E_0. \tag{33}$$

By symmetry

$$\Sigma_{10} \cap \Sigma_{12} \subset C_1 \cup E_2. \tag{34}$$

Lemma 4.5 below shows that $E_0 \cap E_2 = \emptyset$. Hence $\Sigma_{10} \cap \Sigma_{12} = C_1$. We already know that each hemisphere Σ_{1j} is an embedded disk, and we've just seen that these hemispheres just intersect along the equator. Hence Σ_1 is an embedded sphere. The same result holds for Σ_2 by symmetry.

Conclusions If the pictures are right—namely if $\Psi(\Sigma_1 - E_0) \cap \Psi(\Sigma_2 - E_0)$ is a single point—then $\Sigma_0, \Sigma_1, \Sigma_2$ are embedded spheres which have pairwise contractible intersections. But then these spheres bound balls B_0, B_1, B_2 with pairwise disjoint interiors. Moreover I_j interchanges the two components of $S^3 - \Sigma_j$. This picture, at the parameter s, easily implies that ρ_s is a discrete embedding. So, to prove the Goldman–Parker conjecture, we just need to show that $\Psi(\Sigma_1 - E_0) \cap \Psi(\Sigma_2 - E_0)$ is a single point for each $s \in [\underline{s}, \overline{s})$.

4.2 Containing the image

We define S_{12A} to be the region of $\boldsymbol{R}/2\pi\boldsymbol{Z} \times \boldsymbol{R}$ bounded by $\Psi(C_1)$, $\Psi(Q_{21})$ and the horizontal line through the highest point of $\Psi(Q_{21})$, as shown in Figure 4.3.

From Technical Lemma III, the horizontal edge through the lowest point of $\Psi(Q_{21})$ is disjoint from $\Psi(C_2)$. We define S_{12B} to be the region of $\boldsymbol{R}/2\pi\boldsymbol{Z} \times \boldsymbol{R}$ bounded by $\Psi(C_1)$, $\Psi(Q_{21})$ and the horizontal line through the lowest point of $\Psi(Q_{21})$.

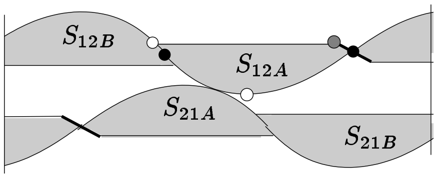

Figure 4.3

We define S_{21A} and S_{21B}, similarly, switching the roles of the indices 1 and 2.

The $\boldsymbol{R}$–axis R_1 divides $\Sigma(E_2, Q_2; C_1)$ into two halves, which we denote by Σ_{12A} and Σ_{12B}. Figure 4.4 shows schematically how $\Psi(\Sigma_{12A})$ sits inside S_{12A}. Compare Figure 4.2. Given Technical Lemma III, the Goldman–Parker Conjecture comes down to proving the following two items, which we establish in turn.

- Let $\alpha = \Psi(\gamma)$, where γ is an $\boldsymbol{R}$–arc of Σ_{12A}. Then α does not intersect ∂S_{12A} at an interior point.
- $\Psi(C_1)$ and $\Psi(C_2)$ intersect in one point

Item 1 implies that $\Psi(\Sigma_{12A}) \subset S_{12A}$ and also that $\Sigma_{12A} \cap \Sigma_{10} \subset C_1$. The same holds for Σ_{12B} and hence $\Sigma_{12} \cap \Sigma_{10} = C_1$. Hence Σ_1 is embedded. Item 2 combines with Item 1 and with the Technical Lemma III to show that $\Psi(\Sigma_1) \cap \Psi(\Sigma_2)$ is a single point.

Our proof of the Goldman–Parker Conjecture boils down to establishing Items 1 and 2.

4.3 The height property

Figure 4.4 shows an enlargement of part of Figure 4.3. As in Figure 4.4 we have distinguished 5 points on ∂S_{12A}: The 2 black points are $\Psi(R_1 - E_1)$, where R_1 is the $\boldsymbol{R}$–axis of Σ_1. The 2 white points are the points of maximum and minimum height on $\Psi(C_1)$. The grey point is the point of $\Psi(\Sigma_{12A})$ which lies on the same horizontal line as the top white point.

Lemma 4.4 *Two arcs of $\Psi(\Sigma_{12A})$, which start at the same horizontal level on $\Psi(C_1)$, end at the same point of $\Psi(Q_{21})$. In particular, there is an arc of $\Psi(\Sigma_{12A})$ which connects the bottom white point to the grey point.*

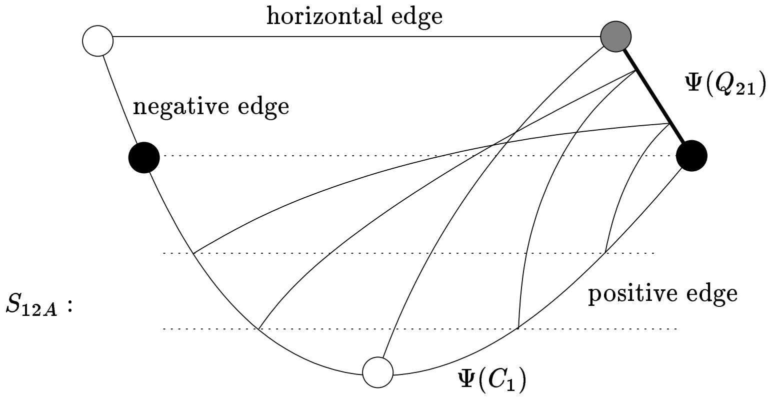

Figure 4.4

Proof The second statement follows as a limiting case of the first statement, so we will concentrate on the first statement. Let α_1 and α_2 be two $\boldsymbol{R}$–arcs of Σ_{12A}. Let $a_j = \Psi(\alpha_j)$ and let $\beta_j = I_1(\alpha_j)$.

(1) a_1 and a_2 have endpoints at the same horizontal level of $\Psi(C_1)$.
(2) Hence $\alpha_1 \cap C_1$ and $\alpha_2 \cap C_1$ lie in the same spinal sphere of the cospinal foliation $S(E_0, Q_0)$.
(3) Hence $\beta_1 \cap C_1$ and $\beta_2 \cap C_1$ lie in the same spinal sphere of the cospinal foliation $S(E_0, Q_0)$.
(4) Hence β_1 and β_2 lie in the same spinal sphere of the cospinal foliation $S(E_0, Q_0)$.
(5) Hence $\beta_1 \cap Q_0 = \beta_2 \cap Q_0$.
(6) Hence $\alpha_1 \cap Q_2 = \alpha_2 \cap Q_2$.
(7) Hence $\Psi(\alpha_1 \cap Q_2) = \Psi(\alpha_2 \cap Q_2)$.
(8) Hence a_1 and a_2 contain the same point of $\Psi(Q_2)$.

This is what we wanted to prove. □

Lemma 4.4 is borne out in Figures 4.1 and 4.2 and can be seen in more detail on our applet.

4.4 The interlacing property

Here we recall Figure 3.4 for convenience. Recall that Σ_1 is *interlaced* if the picture looks like the right hand side of Figure 3.4. That is, a vertical line separates the minimum of $\Psi(E_2)$ from the minimum of $\Psi(C_1)$. In this section we prove that Σ_1 is an interlaced $\boldsymbol{R}$–sphere for all $s \in [\underline{s}, \overline{s})$. This result lets us use all the machinery from Section 3.

Lemma 4.5 *The pairs* (C_1, E_0) *and* (E_0, E_2) *are each generically linked.*

Proof We already know that (E_0, C_1) are linked. Lemma 3.13 now says that (E_0, E_2) are linked. We just need the genericity condition. If C_1 and E_0 bound perpendicular $\boldsymbol{C}$–slices then I_1 stabilizes E_0. By symmetry I_2 stabilizes E_0. But then $I_1 I_2$ stabilizes a $\boldsymbol{C}$–slice. This does not happen. If E_0 and E_2 bound perpendicular $\boldsymbol{C}$–slices then E_0 and E_1 bound perpendicular slices, by symmetry. But then the slices bounded by E_1 and E_2 are disjoint. Hence E_1 and E_2 are unlinked. This is a contradiction. □

Comparing Figure 3.4 with Figure 4.5 we see that Σ_1 is interlaced for the parameter $\underline{s}$. By continuity, and Lemma 4.5, we get that Σ_1 is interlaced for all $s \in [\underline{s}, \overline{s})$. Lemma 4.5 prevents the picture from switching from the right hand side of Figure 3.4 to the left hand side of Figure 3.4 as s varies.

4.5 No interior cusps

Let γ be an $\boldsymbol{R}$–arc of $\Sigma(E_2, Q_2; C_1)$ and let $\widehat{\gamma}$ be the $\boldsymbol{R}$–circle which contains γ. By definition, $\widehat{\gamma}$ is an affiliate of $\Sigma(E_2, Q_2; C_1)$. Figure 4.5 shows a particular example at the parameter $\underline{s}$. Here $\Psi(\gamma)$ in black the rest of $\Psi(\widehat{\gamma})$ in grey. The loop $\Psi(C_1)$ is drawn in black and the loop $\Psi(E_2)$ is drawn in grey. The vertical direction in Figure 4.5 is scaled differently than the vertical direction in Figure 4.1, because otherwise $\Psi(E_2)$ would be quite a tall curve. (On my applet one can use many more colors.)

In general, let $s \in [\underline{s}, \overline{s})$. Suppose, for this parameter, that γ is an $\boldsymbol{R}$–arc of $\Sigma(E_2, Q_2; C_1)$, not contained in the $\boldsymbol{R}$–axis of Σ_1. Let $\widehat{\gamma}$ be the $\boldsymbol{R}$–circle which contains γ. Let $\widehat{\alpha} = \Psi(\widehat{\gamma})$ and let $\alpha = \Psi(\gamma)$. Let $\alpha_0 = \Psi(\gamma \cap C_1)$.

Lemma 4.6 α_0 *is a cusp of* $\widehat{\alpha}$.

Proof Let $\theta \in C_1$ be the endpoint of γ. Let C_θ be the contact plane at θ. Then I_1 rotates C_θ by 180 degrees. This means that γ_θ and $I_1(\gamma_\theta)$ are tangent at θ. But $I_1(\gamma_\theta) = \Sigma(E_0, q_0; \theta)$ is a fiber of Ψ. Hence γ_θ is tangent to a fiber of Ψ at θ. Hence $\alpha_0 = \Psi(\theta)$ is a cusp. □

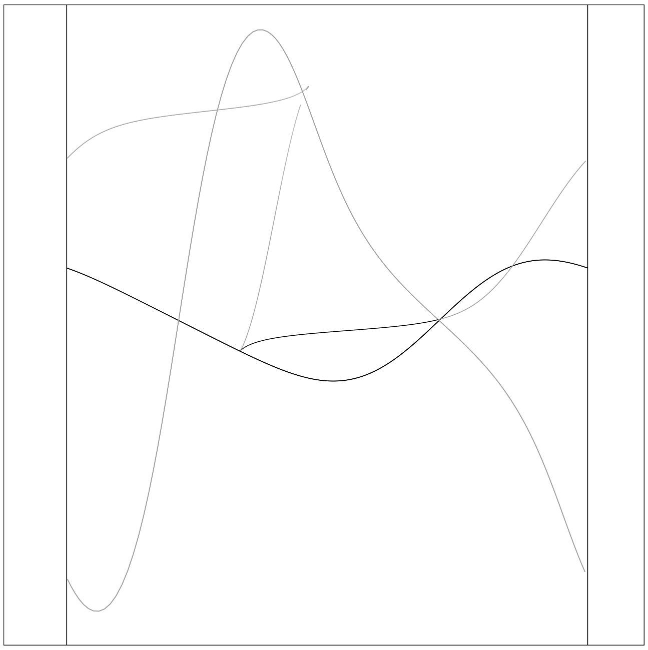

Figure 4.5

Lemma 4.7 *α has does not contain the second cusp of $\widehat{\alpha}$.*

Proof Write $\alpha = \Psi(\gamma)$ and let $\widehat{\gamma}$ be the $\boldsymbol{R}$–circle containing γ. Figure 4.5 shows an example where α does not contain the second cusp. If there are examples where α does contain the second cusp then there is an example where the two endpoints of α are both cusps. But then this second cusp is the point of maximum height of α. Note that $\widehat{\gamma}$ intersects E_2 in two points. For the example under consideration, we therefore have

$$\text{height}(\Psi(\widehat{\gamma} \cap Q_2)) > \text{height}(\Psi(\widehat{\gamma} \cap (E_2 - Q_2))).$$

However, for the example in Figure 4.5

$$\text{height}(\Psi(\widehat{\gamma} \cap Q_2)) < \text{height}(\Psi(\widehat{\gamma} \cap (E_2 - Q_2))).$$

But then, by continuity, we have an example where:

$$\text{height}(\Psi(\widehat{\gamma} \cap Q_2)) = \text{height}(\Psi(\widehat{\gamma} \cap (E_2 - Q_2))).$$

However $\gamma \cap Q_2$ and $\gamma \cap (E_2 - Q_2)$ are in harmonic position with respect to Q_2. Thus by Lemma 3.16, Σ_1 is symmetric. However, Technical Lemma II and Lemma 3.18 combine to say that Σ_1 is asymmetric. This is a contradiction. $\square$

Corollary 4.8 (Rising Property) *Let γ be an $\boldsymbol{R}$–arc of $\Sigma(E_2, Q_2; C_1)$. If γ is not contained in the axis then $\alpha = \Psi(\gamma)$ is nonsingular on its interior and has nowhere vanishing slope.*

Proof This follows immediately from Lemma 3.4 and Lemma 3.5, combined with the fact that α has no cusps in its interior. □

4.6 Proof of Item 1

At the end of subsection 4.2 we reduced the whole Goldman–Parker Conjecture to the verification of two items. Here we prove the first of these items.

Let γ be an $\boldsymbol{R}$–arc of Σ_{12A}. Let $\alpha = \Psi(\gamma)$. We want to show that α does not intersect S_{12A} at an interior point. We break ∂S_{12A} into three edges, as shown in Figure 4.4 From the Rising Property, the height of α attains its maximum at the endpoint of α which lies on $\Psi(Q_{21})$. But the horizontal edge only intersects $\Psi(Q_{21})$ at its point of maximum height by Technical Lemma III. Hence α does not hit the horizontal edge. Since Σ_1 is remote α has positive slope. Hence α cannot hit the negative edge. We just have to worry about the positive edge.

Before we deal with the problem of hitting the positive edge, we want to divide the arcs of $\Psi(\Sigma_{12A})$ into two categories. Say that a *positive arc* is one whose endpoint lies on the positive edge of ∂S_{12A}. Likewise define *negative arcs*. Say that the *middle arc* is the arc which contains the lowest point of $\Psi(C_1)$. This arc connects a white point to a grey point in Figure 4.4. Looking at Figure 4.2 we can see that two positive arcs appear never to cross each other whereas two negative arcs always cross each other.

Lemma 4.9 *Two positive arcs never cross each other at interior points.*

Proof Looking at the ordering of the endpoints of the positive arcs, on $\Psi(C_1)$ and $\Psi(Q_2)$ respectively, we see that two positive arcs α_1 and α_2 must cross in at least one pair of oppositely oriented crossings, if they cross at all. By varying α_2 away from α_1, as in Figure 4.6, we can produce a positive arc α_3 which is tangent to α_1 at some point x. We write $\alpha_j = \Psi(\gamma_j)$. By the Slope Principle of subsection 3.3, the $\boldsymbol{R}$–arcs γ_1 and γ_3 intersect at some point of $\Psi^{-1}(x)$. But this contradicts the fact that Σ_{12} is an embedded disk. □

Corollary 4.10 *A positive edge does not intersect the positive arc of C_1 in a point which is interior to the positive edge.*

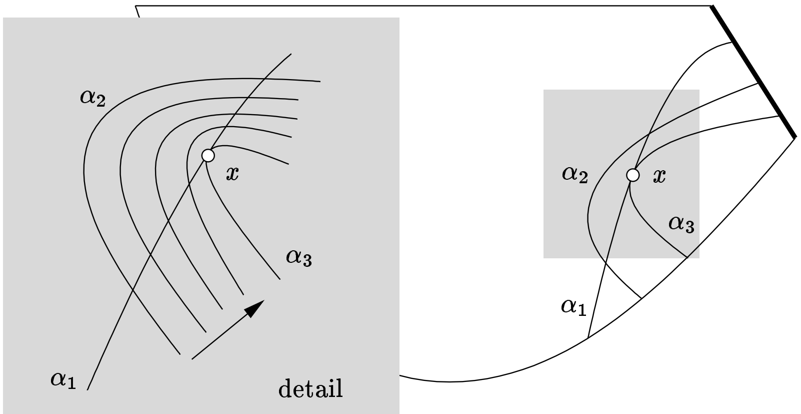

Figure 4.6

Proof Suppose α_0 is a positive arc which intersects the positive edge at a point $x_0 \in \Psi(C_1)$. By Lemma 3.4 the arc α_0 is transverse to $\Psi(C_1)$ at the endpoint. Hence an initial open segment of α_0 either is contained in the interior of S_{12A} or is contained in the complement of S_{12A}. Figure 4.7 shows the former option, which turns out to be the true option. Which option obtains is independent of the choice of α_0, and for α_0 very near the middle arc, the former option obtains by the Rising Property. Hence, the former option always obtains. In summary, some initial portion of α_0 is contained in the interior of S_{12A}.

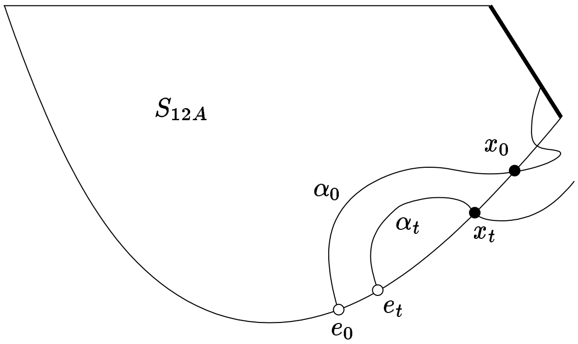

Figure 4.7

We can choose x_0 to be the first point where α_0 crosses the positive edge. Let

$$\{\alpha_t|\ t \in [0,1]\}$$

denote the family of positive arcs such that the endpoint e_t of α_t on $\Psi(C_1)$ interpolates between the endpoint e_0 of α_0 and the point $e_1 = x_0$. The arcs α_0

and α_t cannot cross, by the previous result. Also, an initial open segment of α_t is contained in the interior of S_{12A}. The only possibility is that α_t crosses the positive edge at some first point x_t and, in order along $\Psi(C)$ the points come as

$$e_0, e_t, x_t, x_0.$$

Since $\{e_t\}$ moves all the way from e_0 to x_0 and e_t comes before x_t (by the Rising Property) we must have some value s for which $e_s = x_s$, but this contradicts the fact that some initial open segment of α_s is contained in the interior of S_{12A}.

Lemma 4.11 *Suppose that α_1 is a negative arc and α_2 is a positive arc. Suppose the endpoint of α_1 on $\Psi(C_1)$ is lower than the endpoint of α_2 on $\Psi(C_2)$. Then α_1 and α_2 do not intersect.*

Proof The ordering of the endpoints for α_1 and α_2 is the same as in Lemma 4.9. The same argument as in Lemma 4.9 works here. □

Corollary 4.12 *A negative arc cannot intersect the positive edge of C_1 in a point which is interior to the negative arc.*

Proof Suppose α_1 is a negative arc which intersects the positive edge. Let x be the endpoint of α_1 on $\Psi(C_1)$. Then, from the Rising Property, α_1 intersects the positive edge at a point y which lies above x. But then α_1 must cross some positive arc α_2 whose endpoint lies above x and below y. This contradicts the previous result. □

Our lemmas cover all the cases. This establishes Item 1.

4.7 Proof of Item 2

Our goal is to show that $\Psi(C_1)$ and $\Psi(C_2)$ intersect in a single point, for all parameters $s \in [\underline{s}, \overline{s})$. We work in $\mathcal{H}$ and normalize so that (E_0, Q_0) is in standard position and $\rho_0 = (\boldsymbol{R} \times \{0\}) \cup \infty$. Then C_1 and C_2 are swapped by the $\boldsymbol{R}$–reflection $(z,t) \to (\overline{z}, -t)$ which fixes ρ_0. To fix the scale, we arrange that $C_1 \cap C_2 = (1,0)$. Both C_1 and C_2 have aspect greater than 9 by Technical Lemma I, part 3.

Lemma 4.13 *Define $\Psi_*(z,t) = (\arg z, t)$. Then $\Psi_*(C_1) \cap \Psi_*(C_2) = (0,0)$, with $\Psi_*(C_1)$ lying on top.*

Proof Let $s \in [\underline{s}, \overline{s})$ be some parameter. $\Psi_*(C_j)$ is the graph of a function ψ_j. Up to rotations and scaling, ψ satisfies the equation in Lemma 2.4, for some $A > 9$. Hence, by Lemma 2.4, the function ψ_j is convex on an interval of length π and concave on the complementary interval of length π. We have parametrized ψ so that $\psi(0) = \Psi(p_0)$. Here $p_0 = (1,0) = C_1 \cap C_2$.

Sub-Lemma 4.14 *$\psi''(0) > 0$ for all $s \in [\underline{s}, \overline{s})$.*

Proof We compute that $\psi_1''(0) > 0$ at the parameter $\overline{s}$. Suppose that $\psi_1''(0) = 0$ for some parameter s. Then $\Psi_*(C_1)$ and $\Psi_*(C_2)$ are tangent at their inflection point. Then, by symmetry, $\Psi_*(C_1) = \Psi_*(C_2)$. Here we are using Lemma 2.4 and the fact that $A(C_j) > 9$. Since C_1 and C_2 are generically linked, this forces $C_1 = C_2$, a contradiction. Hence $\psi''(0) > 0$ for all $s \in [\underline{s}, \overline{s})$. □

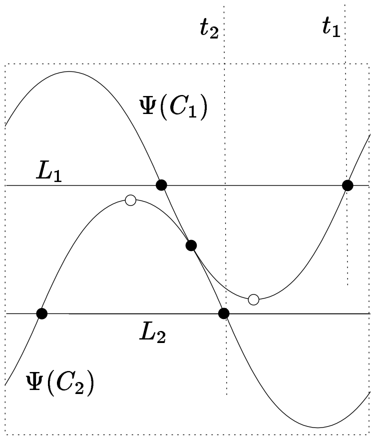

Figure 4.8

We have $\Psi_*(p_0) = (0,0)$. By symmetry we get

$$\psi_1(0) = \psi_2(0); \qquad \psi_1'(0) = \psi_2'(0); \qquad \psi_1''(0) = -\psi_2''(0). \tag{35}$$

The sublemma says that $\psi_1''(0) > 0$. Hence $\psi_2''(0) < 0$ by symmetry. To establish Item 2 it suffices to show that $\psi_1(t) > \psi_2(t)$ for $t \in (0,\pi]$. There are values $t_1, t_2 \in (0,\pi)$ such that $\psi_j''(t_j) = 0$. The point $(t_j, \psi_j''(t_j))$ is one of the points of $\Psi_*(C_j) \cap L_j$. Here L_j is the horizontal line through the inflection points of $\Psi_*(C_j)$. We suppose $t_2 \leq t_1$, as indicated in Figure 4.8. The other case is similiar. Then on $(0,t_2)$ we have $\psi_1 > \psi_2$ because of our initial conditions at 0, and the fact that $\psi_1'' > 0$ on $(0,t_1)$ and $\psi_2'' < 0$ on $(0,t_1)$. For $t \in [t_2,\pi)$ the curve $\Psi_*(C_2)$ lies below L_2, and $\Psi_*(C_1)$ lies completely above L_2 by Technical Lemma III (Section 7). (The point here is that L_j contains the point $\Psi_*(c_j)$, where c_j is the center of C_j.) This does it. □

To finish our proof we compare the loxodromic elevation map Ψ with Ψ_*. Let $A_2^* \subset C_2$ denote the arc such that $\Psi_*(A_2^*)$ ranges between 0 and $\pi/2$ in the S^1 direction on $S^1 \times \boldsymbol{R}$. See Figure 4.9. Let π be projection into $\boldsymbol{C}$.

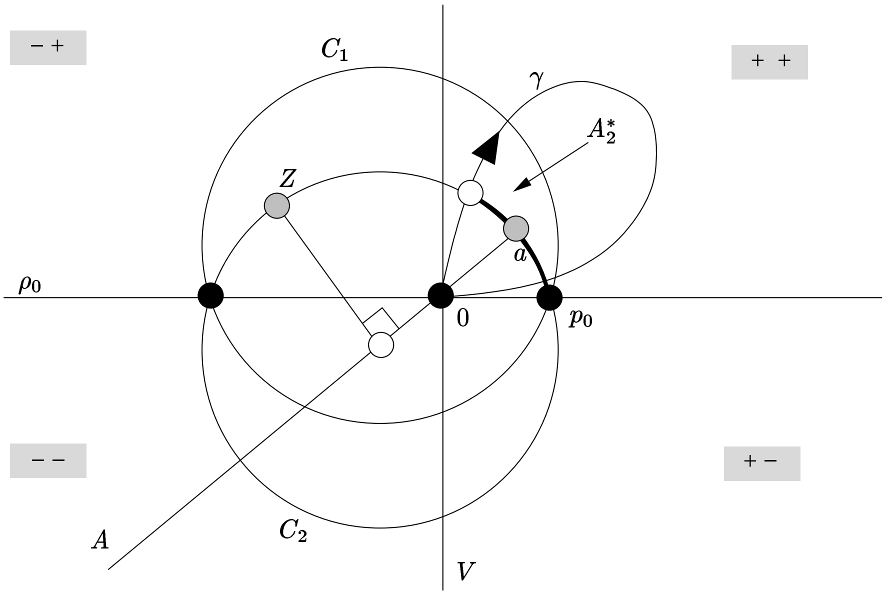

Figure 4.9

Lemma 4.15 $\Psi_*(A_2^*)$ *has negative slope.*

Proof Let m_2^* denote the point on $\Psi_*(C_2)$ having minimum height. We can locate $Z = \pi(\Psi_*^{-1}(m_2^*))$ as follows: Let A be the line through 0 which contains the point a on $\pi(C_2)$ closest to 0. Then Z is obtained from A by rotating 90 degrees about the center of $\pi(C_2)$, as shown. The key observations are that $\pi(C_2)$ lies more in the lower half plane than the upper half plane. This property is true for one parameter and cannot change as the parameter varies, because the $\pi(C_1)$ and $\pi(C_2)$ are symmetrically placed with respect to the real axis, and never coincide. Compare Figure 4.9. From this observation, and item 3 of Technical Lemma I (subsec 5.4) we conclude that Z lies in the $(-,+)$ quadrant, as shown.

Let A_2^{**} denote the arc of C_2 which projects to the $(+,+)$ quadrant. Since $\Psi^{-1}(m_2^*)$ projects to a point in the $(-,+)$ quadrant, we conclude that $\Psi_*(A_2^{**})$ has negative slope. Hence all points of A_2^{**} lie in the lower half-space of $\mathcal{H}$. But then the fibers of Ψ, which contain points on A_2^{**}, project to lemniscate lobes which curve around counter-clockwise. Figure 4.9 shows one such fiber γ. In particular, if γ contains the endpoint of A_2^* then $\pi(\gamma)$ is tangent to V at 0. (This is the $\pi/2$ condition, which defines A_2^*.) But then $\pi(\gamma)$ intersects $\pi(C_2)$ in a point which lies to the right of V. Hence $A_2^* \subset A_2^{**}$. □

Lemma 4.16 *Suppose $x_2 \in C_2$ is a point such that $\Psi(x_2) \in \Psi(C_1)$. Then x cannot be an interior point of A_2^*.*

Proof We will suppose this is false and derive a contradiction. Let $x_2 = (z_2, t_2)$. There is some $x_1 = (z_1, t_1) \in C_1$ and a fiber γ of Ψ such that $x_1, x_2 \in \gamma$. By the previous result $t_2 < 0$. Hence $\pi(\gamma)$ curves around clockwise, as shown in Figure 4.10.

We trace counterclockwise around γ as indicated by the arrow in Figure 4.10. Suppose for the moment that we encounter x_2 before we encounter x_1, as shown in Figure 4.10. The height of γ in $\mathcal{H}$ is monotone decreasing. Hence $t_2 > t_1$. The line L through 0 and $\pi(x_1)$ intersects $\pi(C_1)$ at some point between 1 and $\pi(x_2)$. Let $x_2^* = (z_2^*, t_2^*)$ be the corresponding point on C_2. Then x_2^* lies between $(1,0)$ and x_2. In particular $x_2^* \in A_2^*$. Since $\Psi_*(A_2^*)$ has negative slope, we have $t_2^* > t_2$. Hence $t_2^* > t_1$. But then $\Psi_*(x_2^*)$ lies above $\Psi_*(x_1)$, on the same vertical line. This is a contradiction.

To show that γ encounters x_2 before x_1 we first introduce some terminology. For $j = 1, 2$ we say that a *type j arc* is a portion of a fiber of Ψ which connects a point of $E_0 - Q_0$ to C_j. The type j subarc of γ projects to the portion of $\pi(\gamma)$ which connects 0 to x_j. Thus, we want to show that the type 1 sub-arc $\gamma_1 \subset \gamma$ contains the type 2 sub-arc $\gamma_2 \subset \gamma$. We are going to make an argument based on a computer plot, but we say in advance that we only use topological features of the plot. We simply need the plot to draw the reader's attention to the relevant details.

We use the projection map η from subsection 3.9. Figure 4.11 shows the relevant objects at the parameter $\underline{s}$. The picture looks similar at other parameters.

- The large black circle is $\eta(E_0)$. The small black circle is $\eta(C_1)$. The small grey circle is $\eta(C_2)$. The grey geodesic has $\eta(\partial Q_0)$ for endpoints.
- The black geodesic, which contains $\eta(C_1) \cap \eta(C_2)$, is $\eta(\rho_0)$. The black geodesic arcs are images of the type 1 arcs of interest to us. $\eta(\gamma)$ must

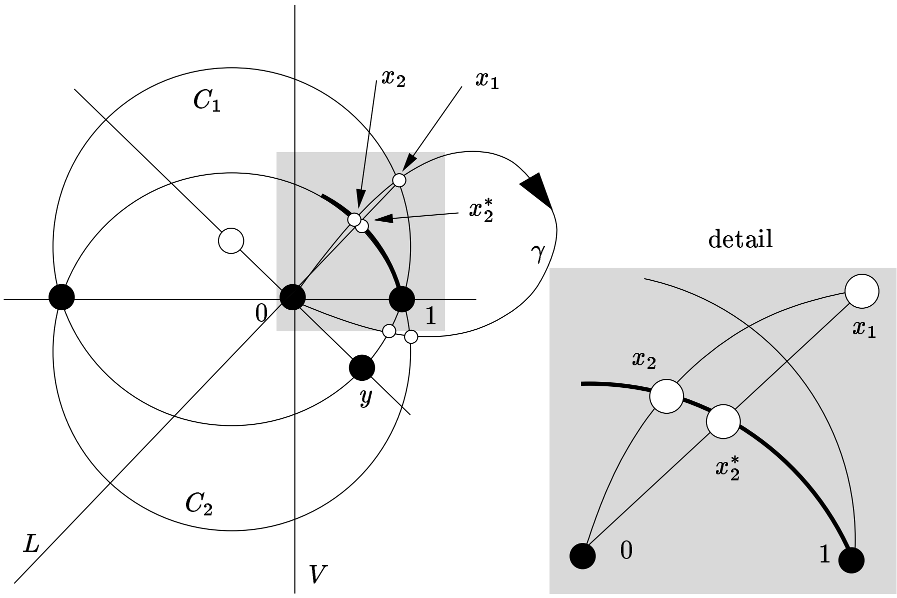

Figure 4.10

project onto the same side of $\eta(R_1)$ as these black arcs, and of course $\eta(\gamma)$ must intersect the black circle.

- The grey arcs are images of the type 2 arcs of interest to us. The last grey arc, the one tangent to the grey circle, is the type 2 subarc which contains the point $z \in C_2$ such that $\Psi(z)$ has minimum height on $\Psi(C_2)$. Thus $\eta(\gamma_2)$ must project into the region indicated by these grey arcs.

The type 1 and type 2 subarcs of R_0 coincide. As we move away from $\eta(R_0)$ the projections of the type 1 subarcs grow longer in comparison to the corresponding projections of the type 2 subarcs. In particular $\eta(\gamma_2)$ would be shorter than $\eta(\gamma_1)$. Hence $\gamma_2 \subset \gamma_1$. □

In summary, if $\Psi(C_1)$ and $\Psi(C_2)$ intersect in a second point (x,y) then $x \not\in (0,\pi/2]$. By symmetry $x \not\in [-\pi/2,0)$. Recall that L_j is the horizontal line in $S^1 \times \boldsymbol{R}$ which contains the points of symmetry of $\Psi(C_j)$. Then the two points of $L_1 \cap \Psi(C_1)$ are π apart. Let s_1 be this symmetry point. As we trace around $\Psi(C_1)$ from $(0,0)$ to s_1 we remain above $\Psi(C_2)$. Once we reach s_2 we remain above L_1 for another π radians of travel in the S^1 direction. It follows from Technical Lemma III that L_1 lies above every point of $\Psi(C_2)$. Depending on which option obtains, we have shown either that $x \not\in (0,3\pi/2]$

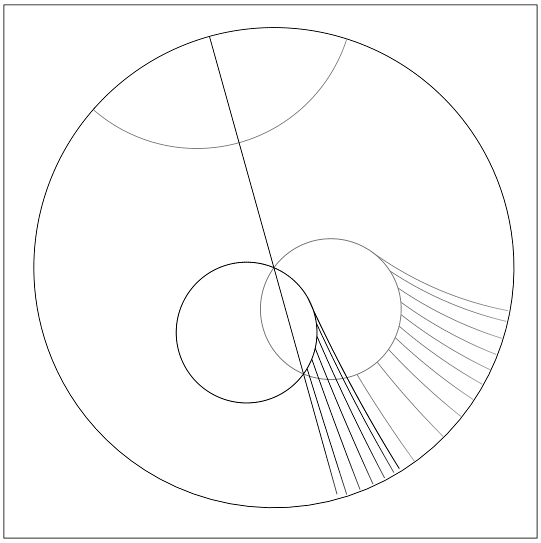

Figure 4.11

of $x \notin [-3\pi/2, 0)$. Either case implies the other by symmetry. But then $x = 0$ and we are done.

5 Technical Lemma I

5.1 A resume of formulas

Here we introduce the formulas which we use for our technical estimates. As a double-check, we verified all the numbered equations computationally, for the parameter $\overline{s}$. We will usually suppress the parameter s from our notation. One should view this section as a continuation of subsection 2.5. First, some quantities from subsection 2.5:

$$\beta = \frac{s+i}{\sqrt{2+2s^2}}; \qquad A_1 = \frac{s+17i}{s+i}; \qquad A_2 = \frac{12\sqrt{2}i}{\sqrt{1+s^2}}. \tag{36}$$

One verifies easily that

$$\overline{\beta} = \frac{3}{4} \times \frac{A_1 - 1}{A_2}. \tag{37}$$

The matrix $g_0 = I_1 I_0 I_2$ has a positive eigenvector $(e, \overline{e}, 1)$. The quantity e (not to be confused with the base of the natural log) figures heavily in our estimates.

As suggested by the referee, we introduce the quantity:

$$x = \frac{e^2 + |e|^2 + \overline{e}^2}{1 - |e|^2}. \tag{38}$$

Here are 4 equations, all due to the referee:

$$|e|^2 = \frac{x^2}{9(x-1)} \tag{39}$$

$$|e - \overline{e}|^2 = \frac{x(x-3)^2}{9(x-1)} \tag{40}$$

$$|2\overline{\beta}e - 1|^2 = \frac{(x-3)^2}{18(x-1)} \tag{41}$$

$$\frac{288}{1+s^2} = |A_2|^2 = \frac{9(x-1)(x-3)^2}{x} \tag{42}$$

We will give the derivations below.

Note that $x > 1$ by Equation 39. It follows readily from Equation 42 that

$$s \in [\underline{s}, \overline{s}] \qquad \Longrightarrow \qquad x \in [4, \frac{817}{200}]. \tag{43}$$

Recall that ρ_0 is the $\boldsymbol{R}$–circle $\{(z, \overline{z})\} \subset S^3$. Recall also that E_0 is the $\boldsymbol{C}$–circle stabilized by the element g_0. We have $E_0 \cap \rho_0 = \{q_0, q_0'\}$ where

$$q_0 = \begin{bmatrix} a \\ \overline{a} \end{bmatrix}; \quad q_0' = \begin{bmatrix} b \\ \overline{b} \end{bmatrix}; \qquad a = \frac{1 + i\sqrt{2|e|^2 - 1}}{2\overline{e}} \quad b = \frac{1 - i\sqrt{2|e|^2 - 1}}{2\overline{e}}. \tag{44}$$

This last equation follows from the fact that

$$|a|^2 = |b|^2 = 1/2; \qquad \langle (a, \overline{a}, 1), (e, \overline{e}, 1) \rangle = \langle (b, \overline{b}, 1), (e, \overline{e}, 1) \rangle = 0.$$

The endpoints of the arc $Q_0 \subset E_0$ are given by

$$\partial_1 Q_0 = (c, d); \qquad \partial_2 Q_0 = (\overline{d}, \overline{c}).$$

We label so that $|c|^2 \geq 1/2$ and $|d|^2 \leq 1/2$. We have

$$|c|^2 |d|^2 = \frac{1}{x(x-3)^2}; \qquad |c|^2 + |d|^2 = 1. \tag{45}$$

We will give the derivation below.

It follows readily from Equation 39 that:

$$\mathrm{Re}(a\overline{b}) = \frac{-9 + 9x - x^2}{2x^2} \tag{46}$$

Below we will derive:

$$\mathrm{Re}(a\overline{\beta}) = \frac{3(x-1)(9+x) - (x-3)\sqrt{(x-3)(2x-3)(9-x)}}{16x^2} \tag{47}$$

A similar derivation, which we omit, shows that:

$$\mathrm{Re}(\beta\overline{b}) = \frac{3(x-1)(9+x) + (x-3)\sqrt{(x-3)(2x-3)(9-x)}}{16x^2} \tag{48}$$

Finally:

$$\mathrm{Im}(a\overline{b} + b\overline{\beta} + \beta\overline{a}) = \frac{(x-3)\sqrt{(x-3)(2x-3)(x-1)}}{8x^2} \tag{49}$$

I'm proud to say that I found these last 3 equations myself, though of course the referee had a much better derivation for the last one, which he communicated to me in his final comments on the paper.

5.2 The derivations

The first 4 derivations, as well as the last one, are essentially due to the referee. We begin with a technical lemma:

Lemma 5.1

$$A_1 = \frac{e^2 - \overline{e}^4}{|e|^2 - |e|^4} \qquad A_2 = \frac{e^3 - \overline{e}^3}{|e|^2 - |e|^4} \qquad \frac{A_1 - 1}{A_2} = \frac{e}{x} + \overline{e} \tag{50}$$

Proof The vector $\widehat{E}_0$ is an eigenvalue for the matrix g_0 from Equation 19. That is:

$$\begin{bmatrix} 0 & -1 & 0 \\ -A_1 & 0 & A_2 \\ -A_2 & 0 & -\overline{A}_1 \end{bmatrix} \begin{bmatrix} e \\ \overline{e} \\ 1 \end{bmatrix} = \lambda \begin{bmatrix} e \\ \overline{e} \\ 1 \end{bmatrix}$$

Reading across the top row we see that $\lambda = -\overline{e}/e$. Reading across the second and third rows, using the relation $\overline{A}_2 = -A_2$, multiplying through by powers of e, and conjugating when necessary, we get:

$$-A_1 e^2 + A_2 e = -\overline{e}^2 \qquad - A_1\overline{e} + A_2\overline{e}^2 = -e$$

The first two equations in Equation 50 are now derived by the usual method of elimination. The third equation is verified by expanding out both sides, using the definition of x, and the identity $e^3 - \overline{e}^3 = (e-\overline{e})(e^2 + |e|^2 + \overline{e}^2)$. □

5.2.1 Derivation of equation 39

Inspecting Equation 36, we get the equation $9|A_1-1|^2=8|A_2|^2$. Therefore:

$$9|e^2-\overline{e}^4-|e|^2-|e|^4|^2=8|e^3-e^3|^2=8|e-\overline{e}|^2\ |e^2+|e|^2+\overline{e}^2|^2 \tag{51}$$

The left hand side equals

$$9|e-\overline{e}|^2|e(1-|e|^2)+\overline{e}(e^2+|e|^2+\overline{e}^2)^2.$$

Therefore

$$9|e(1-|e|^2)+\overline{e}(e^2+|e|^2+\overline{e}^2)|^2=8|e^2+|e|^2+\overline{e}^2|^2.$$

Dividing through by $(1-|e|^2)^2$ and cancelling the term $|e-\overline{e}|^2$ which occurs on both sides of Equation 51, we get $9|e+\overline{e}x|^2=8x^2$. Expanding:

$$9(|e|^2+|e|^2x^2+e^2x+\overline{e}^2x)=8x^2.$$

The left hand side can be re-written as

$$9(|e|^2+|e|^2x^2-|e|^2x+x(e^2+|e|^2+\overline{e}^2))=9|e|^2(1+x^2-x)+x^2(1-|e|^2).$$

Therefore

$$9|e|^2(1+x^2-x))+9x^2(1-|e|^2)=8x^2.$$

Solving this last equation for $|e|^2$ yields Equation 39.

5.2.2 Derivation of equation 40

$$|e-\overline{e}|^2=(e-\overline{e})(\overline{e}-e)=-e^2-\overline{e}^2+2|e|^2=-(e^2+|e|^2+\overline{e}^2)+3|e^2|=$$

$$-x(1-|e|^2)+\frac{x^2}{3(x-1)}=-x(1-\frac{x^2}{9(x-1)})+\frac{x^2}{3(x-1)}.$$

The last equation is equivalent to the right hand side of Equation 40.

5.2.3 Derivation of equation 41

From Equation 37 and the third part of Equation 50 we get

$$2\overline{\beta}e=\frac{3}{2x}\times(e^2+x|e|^2).$$

Substituting in Equation 39 and grouping terms we get

$$\frac{2x}{3}\times(2\overline{\beta}e-1)=e^2+P;\qquad P=\frac{6x-6x^2+x^3}{9x-9}. \tag{52}$$

Therefore
$$\frac{4x^2}{9} \times |2\overline{\beta}e - 1|^2 = |e|^4 + (e^2 + \overline{e}^2)P + P^2 =$$
$$|e|^4 + (e^2 + |e|^2 + \overline{e}^2)P + P^2 - P|e|^2 = |e|^4 + x(1 - |e|^2)P + P^2 - P|e|^2. \quad (53)$$
When we simplify Equation 53, using Equation 39, we get Equation 41.

5.2.4 Derivation of equation 42

The second equality in Equation 42 is the nontrivial one. From Equation 50 we get
$$A_2 = (e - \overline{e})\frac{x}{|e|^2(1 - |e|^2)}; \qquad |A_2|^2 = \frac{x^2|e - \overline{e}|^2}{|e|^2(1 - |e|^2)}.$$
Now we plug in our equations for $|e - \overline{e}|^2$ and $|e|^2$ and simplify.

5.2.5 Derivation of equation 45

The values $c, \overline{d}, e$ are all roots of
$$z^3 + \left[\frac{2\overline{A}_1}{A_2}\right] z^2 + \left[\frac{\overline{A}_1^2}{A_2^2} - \frac{A_1}{A_2^2}\right] z + \frac{1}{A_2}$$
This polynomial comes from solving the system of equations entailed by the equation $g_0(c, d, 1) = \lambda(c, d, 1)$. The last term is the product of the roots. Hence
$$|c|^2|d|^2 = \frac{1}{|A_2|^2|e|^2} = \frac{1}{x(x - 3)^2}.$$
Since $(c, d, 1)$ is a null vector we also have $|c|^2 + |d|^2 = 1$.

5.2.6 Derivation of equation 47

We will be a bit sketchy here, to avoid a huge mess which is best done symbolically. Note that
$$a\overline{\beta} = \frac{(2a\overline{e})(2\overline{\beta}e)}{4|e|^2} = \frac{(1 + i\sqrt{2|e|^2 - 1})(\frac{3}{2x} \times (e^2 + P) + 1)}{4|e|^2}. \quad (54)$$
Here P is as above. The only term on the right hand side of Equation 54 which is not readily expressible in terms of x is the e^2 term. However, when we take twice the real part of the right hand side of Equation 54, which amounts to

adding this formula to the conjugate of itself, the only terms not expressed in terms of x are the real and imaginary parts of e^2. We have

$$2\mathrm{Re}(e^2) = e^2+\overline{e}^2 = x(1-|e|^2)-|e|^2; \qquad (\mathrm{Re}(e^2))^2+(\mathrm{Im}(e^2))^2 = (|e|^2)^2. \quad (55)$$

Using these equations and Equation 39 we get expressions for the real and imaginary parts of e^2 in terms of x. When we simplify the massive expression we arrive at Equation 47.

5.2.7 Derivation of equation 49

Noting that $a\overline{e} = e\overline{b}$ we can write:

$$a\overline{b} + b\overline{\beta} + \beta\overline{a} = \frac{(2a\overline{e})^2 + 2\overline{a}e(2\beta\overline{e} + 2\overline{\beta}e)}{4|e|^2}$$

Using Equations 44 and 52, 55 we find that this last expression equals:

$$\frac{(1 + i\sqrt{2|e|^2 - 1})^2 + \frac{3}{2}(1 - i\sqrt{2|e|^2 - 1})(1 + |e|^2 - |e|^2/x)}{4|e|^2}$$

Using Equation 39 and expanding, we get Equation 49.

5.3 Items 1 and 2

We turn now to the proof of the Technical Lemma I. Here A denotes the aspect of C_1 relative to E_0. A *polar vector* to a $\boldsymbol{C}$–circle C is a vector X such that $\langle X, \widehat{y}\rangle = 0$ whenever $\widehat{y}$ is a lift of a point on C. The polar vector for C is unique up to scale.

Lemma 5.2 *Let $\widehat{E}_0$ and $\widehat{C}_1$ be polar vectors to E_0 and C_1 respectively. Let A be the aspect of C_0 when E_1 is normalized as above. Then:*

$$A = \frac{|\langle\widehat{E}_0, \widehat{E}_0\rangle||\langle\widehat{C}_1, \widehat{C}_1\rangle|}{|\langle\widehat{C}_1, \widehat{E}_0\rangle|^2} \qquad (56)$$

Moreover, the two $\boldsymbol{C}$–circles are linked provided that $A > 1$.

Proof The statement about the linking comes straight from [4, subsection 3.3.2]. Now for the second statement: If E_0 and C_1 are normalized as in Lemma 4.1 then the polar vectors to E_0 and C_1, in the Siegel model, are:

$$\widehat{E}_0 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \qquad \widehat{C}_1 = \begin{bmatrix} r^2 - d^2 \\ \sqrt{2}du \\ 1 \end{bmatrix} \qquad (57)$$

Here u is a unit complex number which is a real multiple of the projection of the center of C_1 to $\boldsymbol{C}$. For E_0 this is obvious and for C_1 it is proved by showing that any two distinct points on C_1 are $\langle,\rangle'$ orthogonal to $\widehat{C}_1$. From here an easy computation shows that the left hand side of Equation 56, computed with the Siegel Hermitian form, yields $r^2/d^2 = A$. □

In the ball model we have

$$\widehat{E}_0 = (e, \overline{e}, 1); \qquad \widehat{C}_1 = (0, 2\overline{\beta}, 1). \tag{58}$$

In fact $\widehat{E}_0$ is the positive eigenvector for g_0 and the eigenvalue is $-\overline{e}/e$. From Lemma 5.2 and Equations 39 and 41 we have

$$A = \frac{2|e|^2 - 1}{|2\overline{\beta}e - 1|^2} = \frac{2(2x-3)}{x-3}. \tag{59}$$

For x as in Equation 43 we compute that $A \in [9.5, 10]$. This establishes Items 1 and 2 of the Technical Lemma I.

5.4 Item 3

Here, as in Section 2,

$$C_1 \cap C_2 = p_0 = (\beta, \overline{\beta}, 1).$$

Let q_0 and q_0' be as in subsection 5.1. Let $\boldsymbol{B}\colon S^3 \to \mathcal{H}$ be the Heisenberg stereographic projection which normalizes as in item 3 of Technical Lemma I (subsec 5.4). Recall that $\pi(z,t) = z$.

Lemma 5.3 *Let $q_0'' = (-a, -\overline{a})$. Then $\pi \circ \boldsymbol{B}(q_0'')$ is the second intersection point of $\pi(C_1) \cap \pi(C_2)$.*

Proof Let V denote the $\boldsymbol{C}$–circle such that $\boldsymbol{B}(V)$ is vertical and $\pi(\boldsymbol{B}(v))$ is the second intersection point of $\pi(C_1) \cap \pi(C_2)$. By symmetry V lies on the Clifford torus and is stabilized by the map $(z, w) \to (\overline{w}, \overline{z})$. Also V contains $(a, \overline{a})$. Hence V also contains q_0''. □

We introduce the cross ratio

$$\chi(z_1, z_2, z_3, z_4) = \frac{(z_1 - z_3)(z_2 - z_4)}{(z_1 - z_2)(z_3 - z_4)}. \tag{60}$$

Let $\zeta \in (-\infty, 0)$. Note that $\chi(\infty, \zeta, 0, 1) > 2$ iff $\zeta < -1$. Note also that the restriction of Ψ_* to ρ_0 is a Mobius transformation. Hence $\zeta = \Psi_*(-a, -\overline{a})$ is closer $0 = \Psi_*(b, \overline{b})$ than is $1 = \Psi_*(\beta, \overline{\beta})$ iff

$$f(s) := \chi(a, -a, b, \beta) = \frac{(a-b)(-\beta-a)}{2a(b-\beta)} > 2. \tag{61}$$

Equivalently (since $|a|^2 = |b|^2 = |\beta|^2 = 1/2$) it suffices to show that

$$\frac{|a-b|^2|\beta+a|^2}{|\beta-b|^2} = \frac{(1-2\ \mathrm{Re}(a\overline{b}))(1+2\ \mathrm{Re}(a\overline{\beta}))}{1-2\ \mathrm{Re}(b\overline{\beta})} > 8 \tag{62}$$

It is an exercise in calculus to show that the the quantities $\mathrm{Re}(a\overline{b})$ and $\mathrm{Re}(a\overline{\beta})$ and $\mathrm{Re}(b\overline{\beta})$ are all monotone for x as in Equation 43. To sketch the idea, let f_1 be the function from Equation 47. We compute explicitly that $f_1'(4) = -1/16$ and we easily get the crude bound $|f_1''| < 1$ on $[4, 5]$. Hence $f_1' < 0$ for x in our range.

Computing at the endpoints of $[\underline{s}, \overline{s}]$ we have:

$$\mathrm{Re}(\mathrm{a}\overline{\mathrm{b}}) \in [.331, .344]; \qquad \mathrm{Re}(\mathrm{a}\overline{\beta}) \in [.432, .438]; \qquad \mathrm{Re}(\mathrm{b}\overline{\beta}) \in [.474, .477]. \tag{63}$$

Hence

$$\frac{(1-2\ \mathrm{Re}(a\overline{b}))(1+2\ \mathrm{Re}(a\overline{\beta}))}{1-2\ \mathrm{Re}(b\overline{\beta})} \geq \frac{(1-2(.334))(1+2(.438))}{1-2(.474)} = 11.9775. \tag{64}$$

This completes our proof.

Remark All our points in $\boldsymbol{C}$ are confined to a single quadrant. Hence, the imaginary parts also vary monotonically. We compute:

$$\mathrm{Im}(b\overline{a}) \in [-.374, -.363] \qquad \mathrm{Im}(a\overline{\beta}) \in [.242, .252] \qquad \mathrm{Im}(\beta\overline{b}) \in [.151, .171] \tag{65}$$

6 Technical Lemma II

6.1 Estimating Q_0

Recall that E_0 intersects the Clifford torus in points $(a, \overline{a})$ and $(b, \overline{b})$, where a and b are as in Equation 44. The vectors $(a, \overline{a}, 1)$ and $(b, \overline{b}, 1)$ have real Hermitian inner product. Therefore the point P_r, represented by the vector

$$r \begin{bmatrix} a \\ \overline{a} \\ 1 \end{bmatrix} + i \begin{bmatrix} b \\ \overline{b} \\ 1 \end{bmatrix} \tag{66}$$

lies in E_0. The idea here is that the chosen lift of P_r is null and lies in the span of $(a,\overline{a},1)$ and $(b,\overline{b},1)$. As $r\to\infty$ the point P_r converges to $(a,\overline{a})$ and $P_0=(b,\overline{b})$. The fixed points of g_0 have the form P_r and P_{-r} where $r=r(s)$. We define $r(\overline{s})=\infty$, in keeping with the analysis just made.

Lemma 6.1 *$r(s)>3$ for all $s\in[\underline{s},\overline{s}]$.*

Proof Equation 45 gives us $|d|^2(1-|d|^2)>.2079$. It follows easily that

$$|d|^2>.294 \tag{67}$$

when x is as in Equation 43. This also holds for the smaller range $s\in[\underline{s},\overline{s}]$.

We will suppose that there is some $s\in[\underline{s},\overline{s}]$ such that $r(s)=3$ and we will derive a contradiction. If $r(s)=3$ then we have

$$d=\frac{3\overline{a}+i\overline{b}}{3+i}.$$

Since $|3+i|^2=10$ we have

$$|3\overline{a}+i\overline{b}|^2>2.94$$

Using $|a|^2=|b|^2=1/2$ and expanding:

$$5+3i(a\overline{b}-\overline{a}b)=5-6\ \mathrm{Im}(a\overline{b})>2.94.$$

Hence

$$\mathrm{Im}(a\overline{b})<.343333. \tag{68}$$

This contradicts Equation 65, which says that $\mathrm{Im}(a\overline{b})>.36$ throughout $[\underline{s},\overline{s}]$. □

6.2 Item 1

There exists a (unique) Heisenberg stereographic projection $\boldsymbol{B}$ which maps E_0 to $(\{0\}\times\boldsymbol{R})\cup\infty$, and has the following properties:

- $\boldsymbol{B}(C_1)$ is centered at a point on the positive real axis.
- $\boldsymbol{B}(C_1)$ projects to a circle in $\boldsymbol{C}$ of radius 1.
- The endpoints of $\boldsymbol{B}(Q_0)$ are symmetrically located on $\boldsymbol{B}(E_0)$. Hence the $\boldsymbol{R}$–axis for Σ_1 is just the real line in $\boldsymbol{C}\times\{0\}$.

The Siegel model polar vector for $\boldsymbol{B}(C_1)$ is:

$$V_3 = \begin{bmatrix} D \\ \sqrt{2}d \\ 1 \end{bmatrix} \qquad D = 1 - d^2 \tag{69}$$

The Siegel model vectors representing the endpoints of $\boldsymbol{B}(Q_0)$ are given by:

$$V_1 = \begin{bmatrix} iu \\ 0 \\ 1 \end{bmatrix} \qquad V_2 \begin{bmatrix} -iu \\ 0 \\ 1 \end{bmatrix} \tag{70}$$

In this section we prove the following result, which implies item 1 of the Technical Lemma II.

Lemma 6.2 *$u > 4.2$ for all $s \in [\underline{s}, \overline{s}]$.*

Given 3 vectors $V_1, V_2, V_3 \in \boldsymbol{C}^{2,1}$ we define:

$$\delta(V_1, V_2, V_3) = \frac{|\mathrm{Im}(\tau)|}{|\mathrm{Re}(\tau)|} \qquad \tau = \langle V_1, V_2\rangle\langle V_2, V_3\rangle\langle V_3, V_1\rangle \tag{71}$$

Using the form in Equation 7 we compute readily that:

$$\delta(V_1, V_2, V_3) = \frac{|D^2 - u^2|}{2Du} \tag{72}$$

V_1, V_2, V_3 correspond to the following vectors in the ball model:

$$P_r \qquad -P_{-r} \qquad \widehat{C}_1 = \begin{bmatrix} 0 \\ 2\overline{\beta} \\ 1 \end{bmatrix} \tag{73}$$

(We prefer to use $-P_{-r}$ in place of P_{-r}.) It is convenient to define:

$$U_a = \langle \begin{bmatrix} a \\ \overline{a} \\ 1 \end{bmatrix}, \widehat{C}_1\rangle = 2\overline{a}\beta - 1 \qquad U_b = \langle \begin{bmatrix} b \\ \overline{b} \\ 1 \end{bmatrix}, \widehat{C}_1\rangle = 2\overline{b}\beta - 1 \tag{74}$$

Then

$$\langle P_r, \widehat{C}_1\rangle = rU_a + iU_b; \quad \langle \widehat{C}_1, -P_{-r}\rangle = r\overline{U}_a + i\overline{U}_b; \quad \langle P_r, -P_{-r}\rangle \in i\boldsymbol{R}.$$

From this we get:

$$\delta(V_1, V_2, V_3) = \left| \frac{r|U_a|^2 - r^{-1}|U_b|^2}{2\mathrm{Re}(U_a\overline{U}_b)} \right| \tag{75}$$

From Equation 63 we have:

$$|U_a|^2 = |1 - 2\overline{a}\beta|^2 = 2 - 4\ \mathrm{Re}(a\overline{\beta}) > .248$$

$$|U_a|^2 = |1 - 2\overline{b}\beta|^2 = 2 - 4\ \mathrm{Re}(b\overline{\beta}) < .104$$

Using Equations 46, 47, 54 and 74 we compute that:

$$2\mathrm{Re}(U_a\overline{U}_b) = \frac{-3(3-4x+x^2)}{2x^2} < .301$$

for x as in Equation 43. Therefore

$$\delta(V_1, V_2, V_3) = \delta(P_r, -P_{-r}, \widehat{C}_1) \geq \frac{3(.248) - (1/3)(.104)}{(.301)} > 2.35$$

Combining our last result with Equation 72 we get $|-D^2+u^2| \geq 4.7Du$. Since $u > 0$ for all parameters, the quantity $-D^2+u^2$ cannot change sign. For otherwise we would have $0 > Du$. Hence $u^2 > 4.7Du$. From Technical Lemma I we have:

$$D = 1 - \frac{1}{d^2} > 1 - \frac{1}{9.5} > .894$$

Hence $u > (4.7)(.894) > 4.2$. This completes our proof.

6.3 Item 2

We normalize by the map $\boldsymbol{B}$ so that all our calculations take place in $\mathcal{H}$. So, C_1 is a $\boldsymbol{C}$–circle, centered on a point $c_1 \in \boldsymbol{R}^+$ and C_1 projects to a circle of radius 1. Since the aspect of C_1 is at least 9 we conclude that $c \in (0, 1/3]$. Now I_1 is a $\boldsymbol{C}$–reflection in C_1. The x–axis is precisely R_1, the axis for Σ_1. Note that R_1 intersects C_1 twice, at points which are precisely 2 units away from each other. The restriction of I_1 to R_1 is a Mobius transformation.

Lemma 6.3 *$E_2 = I_1(E_0)$ intersects R_1 in two points. One of the points is $(c, 0)$ and the other one is $(-1/c + c, 0)$.*

Proof This follows from the fact that the restriction of I_1 to $\boldsymbol{R}^1$ is an inversion in the segment of radius 1 centered at 0. □

Thus we see that E_2 projects to a circle diameter $1/c \geq 3$. Moreover, the center of E_1 is $(-1/2c + c) < -7/6$. Now E_1 is contained in the contact plane based at its center. From this we see that E_2 is contained in a contact plane of slope at least $7/3$. (Here, and below, *slope* means vertical rise divided by horizontal run.) Compared to C_1, the $\boldsymbol{C}$–circle E_2 is a big and tall set.

Lemma 6.4 *Both endpoints of Q_2 are at most $1/4$ from the horizontal plane $\boldsymbol{C} \times \{0\}$.*

Proof Working in the Siegel model, the endpoints of Q_0 have lifts V_1 and V_2, as Equation 70 and the polar vector V_3 for C_1 is as in Equation 69. Using Equation 5 relative to the Siegel form we have

$$I_{C_1}(V_1) = -\begin{bmatrix} iu \\ 0 \\ 1 \end{bmatrix} + 2\frac{iu+D}{2D-2d^2}\begin{bmatrix} D \\ \sqrt{2}d \\ 1 \end{bmatrix} = \begin{bmatrix} D^2 + i(uD-u) \\ * \\ D-1+iu \end{bmatrix}.$$

(The starred entry is irrelevant to us.) Multiplying through by $D-1-iu$ we see that $I_{C_1}(V_1)$ is a scalar multiple of the matrix

$$\begin{bmatrix} P(u,D) + iu(1-2D) \\ * \\ u^2 + (D-1)^2 \end{bmatrix}. \tag{76}$$

Here $P(u,d)$ is a real polynomial in u and D whose form is not important to us. If (z,t) is the point represented by $I_{C_1}(V_1)$, then

$$|t| = \frac{u|1-2D|}{u^2+(D-1)^2} < \frac{|1-2D|}{u} \leq 1/u \leq 1/4.$$

But the point (z,t) is one of the endpoints of $Q_2 = I_1(Q_0)$. The same argment works for the other endpoint of Q_2. □

Here are three geometric facts:

(1) Q_2 is centered at 0 in $\mathcal{H}$ and extends upwards at most $1/4$ in either direction. Since E_2 is contained in a contact plane of slope at least $7/3$, the circle E_2 has slope $7/3$ at the origin by symmetry. Since Q_2 rises up only $1/4$ in either direction away from the origin, we see easily that every point of Q_2 has slope greater than 1.

(2) Given that every point of Q_2 has slope greater than 1, we see that the projection $\pi(Q_2)$ is contained in the dist $\Delta_{1/4}$ of radius $1/4$ about the origin.

(3) Any harmonic $\boldsymbol{R}$–arc which intersects Q_2 has slope less than $1/2$ at the intersection point, because this $\boldsymbol{R}$–arc is integral to the contact structure and the intersection point projects inside $\Delta_{1/4}$.

These properties together imply that $\Psi(Q_2)$ does not contain any extrema of the height function, even at the endpoints. This establishes Item 2 of Technical Lemma II.

Remark A more intuitive way to see the same result is that E_2 is extremely tall in comparison to Q_2, and the fibers of Ψ are fairly straight near Q_2 (because they have large diameter) and hence there is no way the height of $\Psi|_{E_2}$ takes on its extrema on Q_2.

7 Technical Lemma III

7.1 Estimating Q_{21}

We continue our work from the previous chapter. In this section, we continue to normalize using $\boldsymbol{B}$. However, our main estimate is independent of the normalization.

Recall that Ψ is the loxodromic elevation map, the map of interest to us. Given a subset X, which is disjoint from E_0, let $2\|X\|$ denote the "vertical diameter" of the set $\Psi(X)$. By this we mean that $2\|X\|$ denotes the maximum difference in heights between two points of the form $\Psi(p_1)$ and $\Psi(p_2)$, where $p_1, p_2 \in X$. We call $\|X\|$ the *vertical* Ψ*–radius* of X. This quantity is not quite canonical; it depends on us choosing a scaling factor for the image of Ψ. However, we shall always be interested in quantities of the form $\|X\|/\|Y\|$, and this ratio is independent of the way we scale Ψ. The main goal of this section is to prove

Lemma 7.1

$$\frac{\|Q_{12}\|}{\|C_1\|} < \frac{1}{15}. \tag{77}$$

for all $s \in [\underline{s}, \overline{s}]$.

So far we have estimated quantities in $\mathcal{H}$ and we need to translate the information we have gathered into terms related to Ψ. Here is an outline of how we will do this. For $X \subset \mathcal{H}$ let $\|X\|'$ denote half the width of the smallest vertical slab which contains X. At the parameter $s = \overline{s}$ we have $\|X\|' = \|X\|$. We want to compare $\|X\|'$ and $\|X\|$ in general. For this purpose we let $[X] \subset E_0 - Q_0$ denote the set of all (lower) endpoints of harmonic $\boldsymbol{R}$–arcs which contain points of X. These $\boldsymbol{R}$–arcs are meant to be harmonic with respect to (E_0, Q_0). They are the arcs used in the definition of Ψ. Here are the estimates we will prove:

(1) $\|Q_{12}\|'/\|C_1\|' < 1/17$.

(2) $\|[Q_{12}]\|' < \|Q_{12}\|'$ and $\|[C_1]\|' > (15/17)\|C_1\|'$.

(3) $\|Q_{12}\|/\|C_1\| < \|[Q_{12}]\|'/\|[C_1]\|'$.

These estimates combine, in a straightforward way, to establish Lemma 7.1.

7.1.1 First estimate

We have $Q_{21} = I_1(Q_{01})$. Here $Q_{01} \subset Q_0$ consists of points x such that x is the endpoint of an $\boldsymbol{R}$–arc which is harmonic with respect to (E_0, Q_0) and which intersects C_1. The endpoints of Q_{01} are represented by vectors of the form:

$$V_1' = \begin{bmatrix} iu' \\ 0 \\ 1 \end{bmatrix} \qquad V_2' = \begin{bmatrix} -iu' \\ 0 \\ 1 \end{bmatrix} \tag{78}$$

The two endpoints t_1 and t_2 of any $\boldsymbol{R}$–arc, harmonic with respect to (E_0, Q_0), satisfy

$$t_1 t_2 = u^2 > 17,$$

and also $|t_1| < \|C_1\|'$. Hence $u' > 17\|C_1\|'$. The same argument as in Lemma 6.4 now shows that both endpoints of Q_{21} are at most

$$1/(17/\|C_1\|') = \|C_1\|'/17$$

away from $\boldsymbol{C} \times \{0\}$. Hence $\|Q_{12}\|' < \|C_1\|'/17$.

7.1.2 The second estimate

Our harmonic arcs vary monotonically in height from their lower to their upper endpoints. Hence $\|[Q_{12}]\|' \leq \|Q_{12}'\|$. This estimate is true for any set X, actually. The other estimate is the interesting one.

Let η be the map described in subsection 3.9. To do our analysis efficiently we post-compose η with a Mobius transformation so that η is the identity on E_0 and $\eta(\mathcal{H})$ is the left half plane. (The HYP1 coordinate system in our applet is precisely this map.) η conjugates the Heisenberg automorphism $(z,t) \to (rz, r^2 t)$ to the hyperbolic isometry $z \to rz$. Hence, the restriction of η to the $\boldsymbol{R}$–circle $R_1 = R \times \{0\}$ is the map $r \to r^2$.

let $(c, 0)$ denote the center of C_1. Here $c \in (0, 1/3]$ as in subsection 6.3. Since C_1 projects to a circle of radius 1 we have $\|C_1\|' = 2c$. Now, C_1 intersects R_1 in the two points $c-1$ and $c+1$. Hence $\eta(C_1)$ is the circle of radius $2c$ which intersects the positive real axis in the points $(1-c)^2$ and $(1+c)^2$. Figure 7.1 shows the picture.

Let α be a harmonic arc which contains a point of C_1. Then $\eta(\alpha)$ is a semicircle which connects two points of the y–axis which are harmonic with respect to the endpoints of Q_0. The upper endpoint of Q_0 is at least 4 away from the origin. Also, the lower endpoint of $\eta(\alpha)$ is at most $2c$ away from 0. Hence the

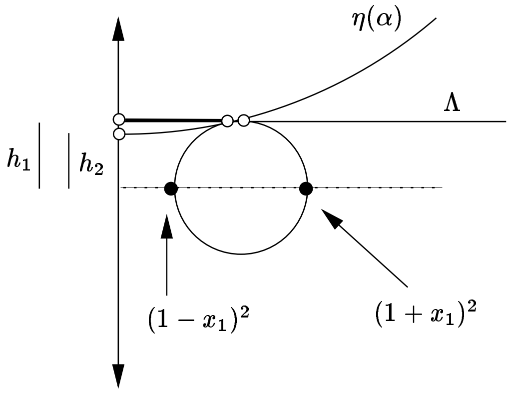

Figure 7.1

other endpoint of $\eta(\alpha)$ is at least $4^2/(2c) > 8/c$ away from the origin. Hence $\eta(\alpha)$ has radius at least

$$R(c) = \frac{4}{c} - c \tag{79}$$

Note that $R(c) > 11$ for $c \in [0, 1/3]$.

We want to compare the point h_2 where $\eta(\alpha)$ intersects the y–axis with the point h_1 where the horizontal line Λ tangent to $\eta(C_1)$ intersects the y–axis. These two points are shown in Figure 7.1. Our goal is to show that $h_2/h_1 > 15/17$. Let λ be the length of the portion of Λ contained inside $\eta(C_1)$. This segment is drawn thickly in Figure 7.1. Using a familiar fact from high school geometry we have

$$(h_1 - h_2)(2R(c) - (h_1 - h_2)) \leq \lambda^2 < (c + (1-c)^2)^2. \tag{80}$$

The right hand side of the equation comes from the fact that the intersection $\Lambda \cap \eta(C_1)$ occurs closer to the y–axis than does the center of $\eta(C_1)$.

To simplify this equation note that $2R(c) > 22$ and certainly $h_1 - h_2 < 1$. Hence

$$(2R(c) - (h_1 - h_2)) > \frac{21}{22} R(c).$$

Therefore (using the fact that $h_1 = 2c$) we have

$$1 - \frac{h_2}{h_1} = \frac{h_1 - h_2}{h_1} = \frac{h_1 - h_2}{2c} < \frac{22}{21} \times \frac{(100/81)}{4cR(c)} \leq \frac{1}{12} < \frac{2}{17}. \tag{81}$$

In particular $h_2/h_1 > 15/17$, as claimed. (We will use the stronger estimate in Equation 81 below.)

7.1.3 The third estimate

At this point we need to look at the loxodromic elevation map Ψ geometrically. We need to quantify the difference in heights between the two points $\Psi(p_1)$ and $\Psi(p_2)$. Let α_j be the harmonic arc which contains p_j. Let $x_j \in E_0 - Q_0$ be the relevant endpoint of α_j. Let $u_1 = u$ and $u_2 = -u$ be the two endpoints of Q_0. Then, up to a constant, the difference in heights of $\Psi(p_1)$ and $\Psi(p_2)$ is:

$$\log \frac{(u_1 - x_2)(u_2 - x_1)}{(u_1 - x_1)(u_2 - x_2)} \tag{82}$$

This quantity is none other than the hyperbolic radius of the segment $[x_1, x_2]$, when the segment $[-u, u] = E_0 - Q_0$ is identified with the hyperbolic line. This interpretation comes from the fact that the quantity in Equation 82 is invariant under Heisenberg automorphisms of (E_0, Q_0), and Ψ conjugates such maps to vertical translations.

After rotating the picture sideways (so as to work with the x–axis rather than the y–axis) our third estimate reduces to the following situation. We have points u and $-u$ as reference points. We have positive $0 < v < w < u$ and we consider the intervals $I_v = [-v.v]$ and $I_w = [-w, w]$. We let $\|I_v\|_u$ denote the hyperbolic radius of the segment I_v when the interval $[u, -u]$ is identified with the hyperbolic line. Likewise we define $\|I_w\|_u$.

To make our third estimate, we have in mind that

$$[Q_{12}] = I_v; \qquad [C_1] = I_w.$$

We have
$$\|[Q_{12}]\|' = \|I_v\|_\infty; \qquad \|[C_1]\|' = \|I_w\|_\infty;$$

$$\|Q_{12}\| = \|I_v\|_u; \qquad \|C_1\| = \|I_w\|_u.$$

Here our specific choice of u is as in Equation 70. Thus, to establish our third estimate, it suffices to show that:

$$\frac{v}{w} > \frac{\|I_v\|_u}{\|I_w\|_u} = \frac{\log(\frac{(u+v)^2}{(u-v)^2})}{\log(\frac{(u+w)^2}{(u-w)^2})} = \frac{\log(u+v) - \log(u-v)}{\log(u+w) - \log(u-w)}$$

This is an easy exercise in calculus. This completes our proof of the third estimate.

7.2 Estimating the gap

At this point we have done the hardest part of the estimate, which involved controlling $\Psi(Q_{21})$. Given Lemma 7.1 we are back to a problem involving the two curves $\Psi(C_1)$ and $\Psi(C_2)$. It is difficult to draw these curves well, so we will consider $\eta(C_1)$ and $\eta(C_2)$, where η is the map considered in the previous section. Given the work done in the previous section, we will see readily how to translate back and forth between η coordinates and Ψ coordinates.

Figure 7.2 shows a schematic (and fairly accurate) picture of $\eta(C_1)$ and $\eta(C_2)$. Here η is the map we used in the previous section.

Regarding this picture:

- The two white dots on the y–axis are the endpoints of $\eta(Q_0)$.
- The two black dots are the hyperbolic centers of $\eta(C_1)$ and $\eta(C_2)$.
- All the geodesics drawn are orthogonal to the geodesic joining the endpoints of $\eta(Q_0)$. Indeed, these geodesics are all certain images of harmonic $\boldsymbol{R}$–arcs.
- For $j = 1, 2$, we have $\gamma_j = \eta(R_j)$, where R_j is the $\boldsymbol{R}$–axis of C_j.
- $\|C_1\|$ is the hyperbolic distance between γ_1 and γ_2 and equally well the hyperbolic distance between γ_2 and γ_4.
- The geodesic γ_5 contains the lower endpoint of $\eta(Q_{21})$.

To finish the Technical Lemma III we just need to show that $\eta(Q_{21})$ lies above the geodesic γ_3. Let d_{ij} denote the hyperbolic distance between γ_i and γ_j. Lemma 7.1, interpreted in terms of η, says that:

$$\frac{d_{14}}{d_{15}} < \frac{1}{15} \tag{83}$$

To prove item 1 of Technical Lemma III it suffices to establish the inequality:

$$\frac{d_{13}}{d_{14}} > \frac{1}{15} \tag{84}$$

Using a bit of algebra, we see that Equation 84 is equivalent to the more symmetric:

$$\frac{d_{34}}{d_{12}} < \frac{7}{8} \tag{85}$$

Remark Our normalization here is slightly different than in the previous section. We are now normalizing so that reflection in the x axis interchanges $\eta(C_1)$ and $\eta(C_2)$. The reason we have changed normalizations is that previously we

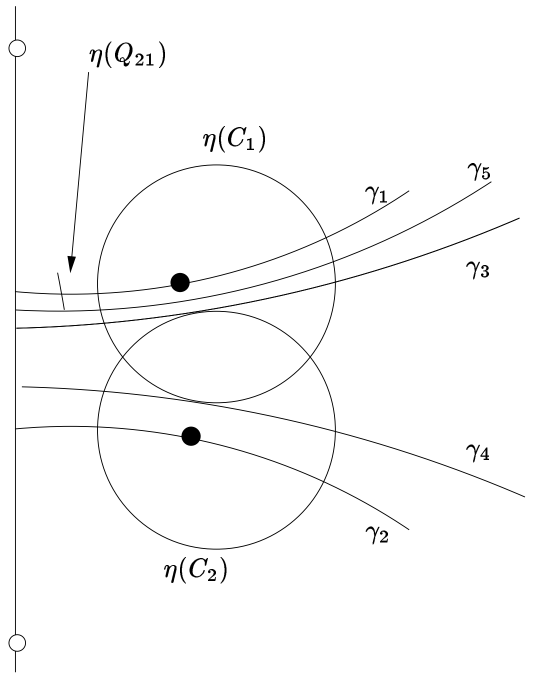

Figure 7.2

were just concentrating on C_1 and now we need to treat both C_1 and C_2. Note that Lemma 7.1 is independent of normalization, as we mentioned above.

Let d'_{12} denote the Euclidean distance between the centers of $\eta(C_1)$ and $\eta(C_2)$ and let d'_{34} denote the Euclidean distance between the point of maximum height on $\eta(C_2)$ and the point of minumum height of $\eta(C_1)$.

Lemma 7.2 *Equation 85 is true provided that*

$$\frac{d'_{34}}{d'_{12}} < \frac{4}{5}. \tag{86}$$

Proof Our proof here is very much like what we did in the previous section, for Estimates 2 and 3. Estimate 3 works exactly the same way, and Estimate 3 is replaced by the statements that

$$d_{34} < d'_{34}; \qquad d_{12} > \frac{11}{12} d'_{12}. \tag{87}$$

Thus, if Equation 86 holds then:

$$\frac{d_{34}}{d_{12}} < \frac{12}{11} \times \frac{4}{5} < \frac{7}{8}$$

The first part of Equation 87 is immediate, as above. The second part of Equation 87 is similar to what we have already done, though we have to think about our new normalization. In the old normalization we knew that $\eta(C_1)$ was centered at the real axis, and had a center which was at most $10/9$ from the y–axis. (In the old normalization $\eta(C_2)$ was some other circle below $\eta(C_1)$.) Our new normalization is obtained by the old one by applying a hyperbolic isometry I translating along the geodesic joining the endpoints of $\eta(Q_0)$. The hyperbolic isometry is chosen so that $\eta(C_1)$ and $\eta(C_2)$ are moved into symmetric position.

Let h_1 denote the Euclidean distance between the centers of $\eta(C_1)$ and $\eta(C_1)$. Let h_2 denote the Euclidean length of the segment on the y–axis between γ_1 and γ_2. We want to show that $h_2/h_1 < 11/12$. This is the same as showing that:

$$1 - \frac{h_2}{h_1} < 1/12$$

The argument given in subsection 7.1.2 goes through, once a few changes are noted.

By construction, the x–coordinate of the hyperbolic center of the new $\eta(C_1)$ is less than the x–coordinate of the hyperbolic center of the old $\eta(C_1)$, which is in turn less than the x coordinate of the Euclidean center of the old $\eta(C_1)$. We conclude that the Euclidean distance from the hyperbolic center of the new $\eta(C_1)$ to the y–axis is at most $10/9$. (This hyperbolic center is the intersection point of $\gamma_1 = \eta(\alpha_1)$ and Λ, the curves from subsection 7.1.2.) The radius of the circle containing γ_1 is still at least $R(c)$. Having made these definitions, we see that the rest of the argument in subsection 7.1.2 is the same. □

By the previous result, Equation 86 implies item 1 of Technical Lemma III. Since $4/5 < 1$, Equation 86 implies item 2 of Technical Lemma III. Thus, to finish the proof of Technical Lemma III we just have to verify Equation 86.

7.3 Back to Heisenberg space

Figure 7.3 shows the picture of $\pi(C_1)$ and $\pi(C_2)$ in $\boldsymbol{C}$. Referring to Figure 7.3, we will show that Technical Lemma III is true provided that $\sin(\theta) > 5/9$.

Let $c_1 = z + it$ be the center of C_1. The quantity d'_{12} is just the vertical distance between centers of the two circles. The quantity d'_{34} is the vertical distance between the minimum height point on C_1 and the maximum height point on C_2.

The outer edge of the rectangle in Figure 7.3 is tangent to $\pi(C_1)$ and parallel to axis 1, the $\boldsymbol{R}$–circle through the center of C_1 which intersects C_1 in two other points. The number R is the radius of C_1.

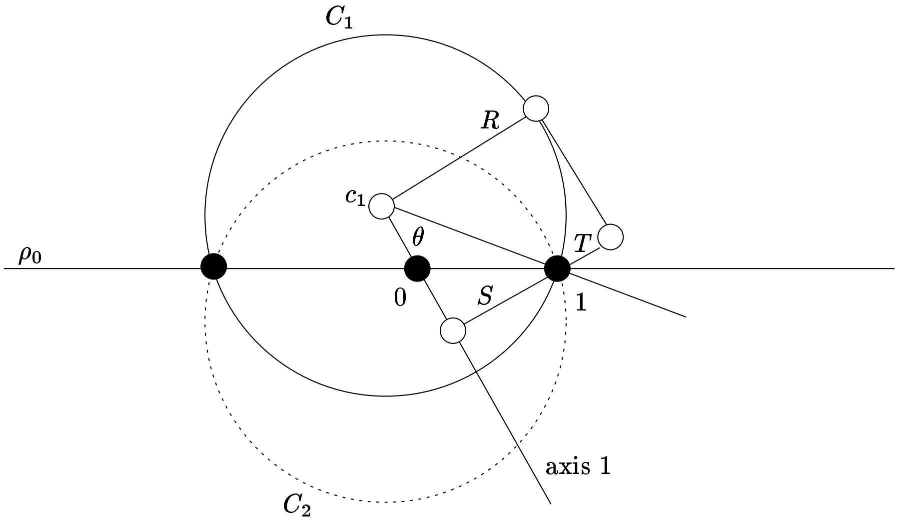

Figure 7.3

The vertical distance between the two centers is twice the distance from the center of C_1 to $\boldsymbol{C} \times \{0\}$. This latter distance is $2|z|S$ because S runs perpendicular to axis 1 and joins a point on C_1 to the point $(1,0)$. The vertical distance between the max of C_2 and the min of C_1 is twice the vertical distance from $\boldsymbol{C} \times \{0\}$ to the min of C_1. This distance is $2|z|T$. Therefore

$$\frac{d'_{34}}{d'_{12}} = \frac{T}{S} = \frac{R - R\sin(\theta)}{R\sin(\theta)} = \frac{1}{\sin(\theta)} - 1.$$

If $\sin(\theta) > 5/9$ then the quantity on the right is less than $4/5$.

7.4 The end of the proof

We will suppose that $\sin(\theta) < 5/9$ and derive a contradiction.

7.4.1 Siegel model computation

The center of C_1 is the point $c = z + it$. In Section 5 we saw that C_1 has aspect less than 10. Hence

$$\frac{R}{|z|} < \sqrt{10}. \tag{88}$$

In the Siegel model, the vectors representing $0, \infty, c$ are respectively:

$$V_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \qquad V_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad V_1 = \begin{bmatrix} -|z|^2 + it \\ \sqrt{2}z \\ 1 \end{bmatrix} \tag{89}$$

Letting δ be the invariant from Equation 71 we have

$$\delta(V_1, V_2, V_3) = \frac{t}{|z|^2} = \frac{2|z|R\sin(\theta)}{|z|^2} = \frac{R}{|z|} \times 2\sin(\theta) < \sqrt{10} \times \frac{10}{9} < 3.52. \tag{90}$$

7.4.2 Ball model computation

In the ball model the corresponding vectors are:

$$U_1 = \begin{bmatrix} b \\ \overline{b} \\ 1 \end{bmatrix} \qquad U_2 = \begin{bmatrix} a \\ \overline{a} \\ 1 \end{bmatrix} \qquad U_3 = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 3 & -4\overline{\beta} \\ 0 & 4\beta & -3 \end{bmatrix} \begin{bmatrix} a \\ \overline{a} \\ 1 \end{bmatrix} \tag{91}$$

Note that $\langle U_1, U_2 \rangle \in \boldsymbol{R}$. Also $\langle U_2, U_3 \rangle \in \boldsymbol{R}$ because these two vectors are exchanged by I_1. Therefore $\delta(V_1, V_2, V_3) = |\mathrm{Im}(\tau)|/|\mathrm{Re}(\tau)|$, where

$$\tau = \langle U_1, U_3 \rangle = 3 + 3(a\overline{b} + b\overline{a}) - 4a\overline{b} - 4b\overline{\beta} - 4\beta\overline{a}. \tag{92}$$

Using Equations 46, 47, 48, and 49 we get:

$$\delta(V_1, V_2, V_3) = \sqrt{\frac{(x-1)(2x-3)}{x-3}} \tag{93}$$

A bit of calculus now shows that

$$\delta(V_1, V_2, V_3) > 3.83 \tag{94}$$

in our range of parameters. This contradicts Equation 90 and we are done.

References

[1] **S Anan'in**, **H Grossi**, **N Gusevskii**, *Complex Hyperbolic Structures on Disc Bundles over Surfaces I*, preprint (2003)

[2] **D B A Epstein**, *Complex hyperbolic geometry*, from: "Analytical and geometric aspects of hyperbolic space (Coventry/Durham, 1984)", London Math. Soc. Lecture Note Ser. 111, Cambridge Univ. Press, Cambridge (1987) 93–111

[3] **E Falbel**, **J R Parker**, *The moduli space of the modular group in complex hyperbolic geometry*, Invent. Math. 152 (2003) 57–88

[4] **W M Goldman**, *Complex hyperbolic geometry*, Oxford Mathematical Monographs, The Clarendon Press Oxford University Press, New York (1999)

[5] **W M Goldman**, **J R Parker**, *Complex hyperbolic ideal triangle groups*, J. Reine Angew. Math. 425 (1992) 71–86

[6] **R E Schwartz**, *Ideal triangle groups, dented tori, and numerical analysis*, Ann. of Math. (2) 153 (2001) 533–598

[7] **R E Schwartz**, *Degenerating the complex hyperbolic ideal triangle groups*, Acta Math. 186 (2001) 105–154

[8] **R E Schwartz**, *Spherical CR Geometry and Dehn Surgery*, preprint of Research Monograph (2004)

[9] **R E Schwartz**, *An Interactive Proof of the G-P Conjecture*, Java applet (2004) `http://www.math.brown.edu/~res/Java/App45/test1.html`

[10] **S Wolfram**, *The Mathematica book*, Wolfram Media, Inc. Champaign, IL (1999)