Dubins' problem in the hyperbolic plane using the open ball model

Dirk Mittenhuber

November 18, 1996

1 The motion group

The group G of isometries is isomorphic to $\mathrm{PSU}(1,1)$. In fact the group action can be described as the action of $\mathrm{SU}(1,1) \subseteq \mathrm{SL}(2,\mathbb{C})$ on $\mathbb{C}$ by Moebius transformations.

$$\mathrm{SU}(1,1) = \left\{ \begin{pmatrix} a & b \\ \bar{b} & \bar{a} \end{pmatrix} \mid a, b \in \mathbb{C},\ |a|^2 - |b|^2 = 1 \right\}, \quad g.z = \frac{az+b}{\bar{b}z+\bar{a}}.$$

This action leaves the unit circle invaraint, hence its interior and exterior too. Using this description has the advantage that we can look at the ambient space as well and exploit our knowledge about Moebius transformations. The key property is that a Moebius transformation on $\mathbb{C}$ maps circles and straight lines onto circels and straight lines. We also observe that rotations around the origin $0 \in \mathbb{C}$ are isometries. Indeed we have

$$\begin{pmatrix} e^{it} & 0 \\ 0 & e^{-it} \end{pmatrix}.z = \frac{e^{it}z}{e^{-it}} = e^{2it}z.$$

In this model the curves in $\mathbb{H}^2$ with constant curvature κ are the intersection of a euclidean circle $C(z_0, \rho) = \{z_0 + \rho e^{it} \mid t \in [-\pi, \pi)\} \subseteq \mathbb{C} \cong \mathbb{R}^2$ with the open unit disc $\mathbb{D}$. This intersection is nonempty if and only if the point on $C(z_0,\rho)$ closest to the origin lies in the unit disc. This means that

$$-1 < |z_0| - \rho < 1, \quad \text{i.e.,} \quad |z_0| - 1 < \rho < |z_0| + 1.$$

The circles that are entirely contained in $\mathbb{D}$ have curvature > 1. They are actually balls with respect to the hyperbolic metric d_H, cf. [3]. A circle which is contained in the closed unit disc $\overline{D}$ and is not the unit circle has precisely one point on the boundary. These circles are called *horocycles*. They have curvature 1.

Now we want to compute the hyperbolic curvature of $C(z_0, \rho)$. We will do this first in the case $\rho = |z_0|$.

In order to translate from $\mathrm{SO}_0(1,2)$ to $\mathrm{SU}(1,1)$ we use a special Lie algebra isomorphism $\phi\colon \mathfrak{so}(1,2) \to \mathfrak{su}(1,1)$.

Lemma 1.1 *The linear map $\phi\colon \mathfrak{so}(1,2) \to \mathfrak{su}(1,1)$ defined by*

$$\begin{array}{ccc} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \\ \downarrow & \downarrow & \downarrow \\ \frac{1}{2}\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} & \frac{1}{2}\begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix} & \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{array}$$

is a Lie algebra isomorphism.

In Monroy's setting we passed from a curve $g(t) \in \mathrm{SO}_0(1,2)$ to a curve $x(t) \in \mathbb{H}^2$ by sending $g(t)$ to its first column vector $x(t) = g(t)e_1$. Let $K \subseteq \mathrm{SO}_0(1,2)$ denote the subgroup that fixes e_1, i.e.,

$$K = G_{e_1} = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos t & -\sin t \\ 0 & \sin t & \cos t \end{pmatrix} \mid t \in \mathbb{R} \right\} = \exp \mathbb{R} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \cong \mathrm{SO}(2).$$

Then $\mathbb{H}^2 = G/G_{e_1} = \mathrm{SO}_0(1,2)/K$. Now we can translate from Monroy's $\mathrm{SO}_0(1,2)$ setting to our $\mathrm{SU}(1,1)$ setting. For with $\mathfrak{k} = \mathbf{L}(K)$ the Lie algebra of K, we have $\phi(\mathfrak{k}) = \mathbb{R}\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$. The stabilizer G_{x_0} of the point $x_0 = 0 \in \mathbb{D}$ has Lie algebra $\mathfrak{g}_{x_0} = \phi(\mathfrak{k})$. Thus $G_{x_0} = \exp \phi(\mathfrak{k}) =: \widetilde{K}$. Since $-\mathbf{1} \in \mathrm{SU}(1,1)$ acts trivially on $\mathbb{D}$, the homogeneous spaces

$$\mathbb{D} = \mathrm{SU}(1,1)/\widetilde{K}, \quad \mathbb{H}^2 = SO_0(1,2)/K$$

are isomorphic. Moreover let $\pi\colon \mathrm{SO}_0(1,2) \to \mathrm{SO}_0(1,2)/K = \mathbb{H}^2$ be the canonical projection, i.e., $g \mapsto g.e_1$. Then Monroy showed that $x(t) = \pi(\exp tX_\kappa)$ is a curve in $\mathbb{H}^2$ with constant curvature $|\kappa|$ and it is parametrized by arclength. Thus the same statement is true for the canonical map $\widetilde{\pi}\colon \mathrm{SU}(1,1) \to \mathrm{SU}(1,1)/\widetilde{K} = \mathbb{D}$ which is given by $\widetilde{\pi}(g) = g.0$.

Lemma 1.2 *For $g \in \mathrm{SU}(1,1)$ fixed let $x' = g.0$. Let $\kappa \in \mathbb{R}$, then the curve*

$$x(t;\kappa) = g \exp\left(\frac{it}{2} \begin{pmatrix} \kappa & 1 \\ -1 & -\kappa \end{pmatrix} \right).0$$

is parametrized by arclength, has constant hyperbolic curvature $|\kappa|$, and starts at $x(0;\kappa) = g.0 = x'$.

Proof. This follows from $x(t;\kappa) = \widetilde{\pi}\Big(g \exp(t\phi(X_\kappa))\Big)$ and the previous discussion. $\square$

Corollary 1.3 *For every $z_0 \in \mathbb{D}$ let $g_0 \in \mathrm{SU}(1,1)$ be such that $g_0.0 = z_0$. Then the curve*

$$x(t) = g_0 \exp \frac{it}{2}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.0$$

is a geodesic parametrized by arclength starting at z_0. For every $c \in \mathbb{C}$ the curve

$$x(t;c) = g_0 \exp \frac{t}{2}\begin{pmatrix} 0 & c \\ \bar{c} & 0 \end{pmatrix}.0$$

is a geodesic with arclength $\ell(t) = |c| \cdot t$.

Lemma 1.4 (The First Circle Lemma) *The circle $C(\rho,\rho)$ has hyperbolic curvature $\kappa = \frac{1}{2\rho}$.*

Proof. We must distinguish the three cases $\kappa^2 > 1,\ = 1,\ < 1$. For $\kappa^2 > 1$ we let $w = \sqrt{\kappa^2 - 1}$ and obtain

$$\exp\left(\frac{it}{2}\begin{pmatrix} \kappa & 1 \\ -1 & -\kappa \end{pmatrix}\right) = \begin{pmatrix} \cos(wt/2) + i\kappa\frac{\sin(wt/2)}{w} & i\frac{\sin(wt/2)}{w} \\ -i\frac{\sin(wt/2)}{w} & \cos(wt/2) - i\kappa\frac{\sin(wt/2)}{w} \end{pmatrix}$$

Thus

$$x(t) = \exp(\,\cdot\,).0 = \frac{i\sin(wt/2)}{w\cos(wt/2) - i\kappa\sin(wt/2)}.$$

The arc $x(t)$ hits the horizontal axis if and only if $\cos(wt/2) = 0$. This happens first when $wt/2 = \pi/2$, thus $\sin(wt/2) = 1$ and $x = -1/\kappa$. So the center has to be at $-1/(2\kappa)$. We simply verify

$$\begin{aligned} x(t) - \left(-\frac{1}{2\kappa}\right) &= \frac{2\kappa i\sin(wt/2) + w\cos(wt/2) - i\kappa\sin(wt/2)}{2\kappa(w\cos(wt/2) - i\kappa\sin(wt/2))} \\ &= \frac{1}{2\kappa}\frac{w\cos(wt/2) + i\kappa\sin(wt/2)}{w\cos(wt/2) - i\kappa\sin(wt/2)}. \\ \left|x(t) + \frac{1}{2\kappa}\right| &= \frac{1}{2|\kappa|}\frac{|z|}{|\bar{z}|} = \frac{1}{2|\kappa|}. \end{aligned}$$

For $\kappa^2 = 1$ we get the so-called horocycles,

$$\exp\frac{it}{2}\begin{pmatrix} \kappa & 1 \\ -1 & -\kappa \end{pmatrix} = \begin{pmatrix} 1 + \frac{it}{2}\kappa & \frac{it}{2} \\ -\frac{it}{2} & 1 - \frac{it}{2}\kappa \end{pmatrix}$$

and

$$x(t) = \exp(\,\cdot\,).0 = \frac{it}{2 - it\kappa} = \frac{-t^2\kappa + 2it}{4 + \kappa^2 t^2} = \frac{-t^2\kappa + 2it}{4 + t^2}$$

because of $\kappa^2 = 1$. This is a circle of radius $1/2$ with center at $\frac{1}{2}\kappa = \frac{1}{2\kappa}$.

Finally, if $\kappa^2 < 1$, then with $w = \sqrt{1 - \kappa^2}$ we obtain

$$\exp\frac{it}{2}\begin{pmatrix} \kappa & 1 \\ -1 & -\kappa \end{pmatrix} = \begin{pmatrix} \cosh(wt/2) + i\kappa\frac{\sinh(wt/2)}{w} & i\frac{\sinh(wt/2)}{w} \\ -i\frac{\sinh(wt/2)}{w} & \cosh(wt/2) - i\kappa\frac{\sinh(wt/2)}{w} \end{pmatrix}.$$

So
$$x(t) = \exp(\,\cdot\,).0 = \frac{i\sinh(wt/2)}{w\cosh(wt/2) - i\kappa\sinh(wt/2)}$$
and
$$x(t) - \left(-\frac{1}{2\kappa}\right) = \frac{1}{2\kappa}\frac{w\cosh(wt/2) + i\kappa\sinh(wt/2)}{w\cosh(wt/2) - i\kappa\sinh(wt/2)}.$$
Thus $\left|x(t) + \frac{1}{2\kappa}\right| = \frac{1}{2|\kappa|}$ is constant. This finishes the proof. $\square$

Now we start computing the hyperbolic curvature for $C(z_0, \rho)$. Since rotations around the origin are symmetries, we can assume w.l.o.g. that $z_0 = iy_0$ with $y_0 \in \mathbb{R}^+$. By the First Circle Lemma we know the curvature of every euclidean circle passing through the origin. Now we look what happens when we shift along the vertical axis. For $z \in \mathbb{D}$ let
$$g(z) = \frac{1}{\sqrt{1-|z|^2}}\begin{pmatrix} 1 & z \\ \bar{z} & 1 \end{pmatrix}.$$
Then $g(z) \in \mathrm{SU}(1,1)$ and
$$g(z).z = \frac{2z}{1+|z|^2}, \quad g(z).0 = z, \quad g(z).(-z) = 0.$$
Moreover $g(z)$ leaves the line $\mathbb{R}z$ invariant:
$$g(z).tz = \frac{tz+z}{t|z|^2+1} = \frac{1+t}{1+t|z|^2}z.$$
If $|z| > 0$, then $\pm z/|z|$ are fixed ponts of $g(z)$ because
$$\frac{1 \pm |z|^{-1}}{1 \pm |z|} = \frac{1}{|z|}\frac{|z| \pm 1}{1 \pm |z|} = \pm\frac{1}{|z|}.$$
In particular for $z = ir \in i\mathbb{R}$ we obtain
$$g(ir).iy = \frac{iy+ir}{-ir\,iy+1} = i\frac{y+r}{1+ry}.$$
Moreover we have
$$g(ir).0 = ir, \quad g(ir).-ir = 0, \quad g(ir).i = i, \quad g(ir).-i = -i.$$
Since $g(ir) \in \mathrm{SU}(1,1)$ is an isometry, it leaves the curvature unchanged. Now we obtain

Lemma 1.5 *For $r \in (0,1)$ we have*
$$g(r).C(id/2,\, d/2) = C\left(y_0(r,d),\, \rho(r,d)\right)$$
with
$$y_0(r,d) = \frac{1}{2}\left(\frac{r+d}{1+rd} + r\right), \quad \rho(r,d) = \frac{1}{2}\left(\frac{r+d}{1+rd} - r\right).$$

Proof. since Moebius transformations map circles and straight lines onto circles and straight lines, the set $g(r).C(id/2, d/2)$ is a circle through $g(r).0 = ir$ and $g(r).id = i\frac{r+d}{1+rd}$. Moreover the diameter line segment from 0 to id is mapped onto the line segment from ir to $i\frac{r+d}{1+rd}$. Since Moebius transformations are conformal mappings, this image again has to be a diameter. Thus the center of this circle is

$$y_0(r,d) = \frac{1}{2}(g(r).id + g(r).0) = \frac{1}{2}\left(\frac{r+d}{1+rd} + r\right)$$

and its radius is

$$\rho(r,d) = \frac{1}{2}(g(r).id - g(r).0) = \frac{1}{2}\left(\frac{r+d}{1+rd} - r\right)$$

□

Lemma 1.6 (The Hyperbolic Curvature Formula) *For $z_0 \in \mathbb{C}$ and $\rho > 0$ The hyperbolic curvature of $C(z_0, \rho)$ is given by*

$$\kappa(z_0, \rho) = \left|\frac{1 + \rho^2 - |z_0|^2}{2\rho}\right|.$$

Proof. Since rotations around the origin are isometries, we may assume w.l.o.g. that $z_0 = iy_0$ with $y_0 \geq 0$. By the previous Lemma we know that for $r \in (0,1)$ and $d > 0$ the circles $C(y_0(r,d), \rho(r,d))$ and $C(id/2, d/2)$ have the same hyperbolic curvature. By the First Circle Lemma the hyperbolic curvature of $C(id/2, d/2)$ is $\kappa = d^{-1}$. Thus we must compute the inverse of the map $(r,d) \mapsto (y_0, \rho)$. We obtain $r = y_0 - \rho$ and

$$\begin{aligned}
y_0 + \rho &= \frac{r+d}{1+rd} = \frac{y_0 - \rho + d}{1 + r(y_0 - \rho)} \\
y_0 + \rho + (y_0 + \rho)(y_0 - \rho)d &= y_0 - \rho + d, \\
2\rho &= d(1 - y_0^2 + \rho^2) \\
d &= \frac{2\rho}{1 - y_0^2 + \rho^2} \\
\kappa &= \frac{1}{d} = \frac{1 - y_0^2 + \rho^2}{2\rho}.
\end{aligned}$$

□

Remark 1.7 We observe that for $|z_0| > 1$ and $\rho = \sqrt{|z_0|^2 - 1} > 0$ we obtain $\kappa(z_0, \rho) = 0$, i.e., $C(z_0, \sqrt{|z_0|^2 - 1})$ is a geodesic.

We also observe that for $z_0 \in \mathbb{D}$ we always have $1 - |z_0|^2 + \rho^2 > 0$, hence we can forget about taking the absolute value. Since we will be interested in Dubins' problem with $\kappa > 1$, all the circles we will need to determine will have curvature either $\kappa > 1$ or $\kappa = 0$. In both cases the hyperbolic curvature formula applies in the form

$$\kappa(z, r) = \frac{1 + r^2 - |z|^2}{2r}.$$

Sometimes it turns out to be useful to write the hyperbolic curvature formula in implicit form

$$1 + \rho^2 - |z|^2 - 2\kappa\rho = 0.$$

We can also solve this equation for $\rho = \rho(z, \kappa)$ and $|z| = |z|(\kappa, \rho)$ which will be needed. We obtain

$$\begin{aligned} \rho(z,\kappa) &= \kappa - \sqrt{\kappa^2 + |z|^2 - 1}, \\ |z|(\kappa,\rho) &= \sqrt{1 - 2\kappa\rho + \rho^2}. \end{aligned}$$

We conclude with the formula for the hyperbolic curvature of the euclidean line $L = z_0 + \mathbb{R}v$ where $z_0 \in \mathbb{C}$ and $v \in \mathbb{C} \setminus \{0\}$. We obtain

Lemma 1.8 *The hyperbolic curvature of the euclidean line $L = z_0 + \mathbb{R}v$ where $z_0 \in \mathbb{C}$ is given as*

$$\kappa = \min_{t \in \mathbb{R}} |z_0 + tv| = \left| z_0 - \frac{\Re(z_0 \bar{v})}{|v|^2} v \right| = \frac{1}{|v|^2} |z_0|v|^2 - (z_0, v)v|.$$

Proof. All we need to do is compute the hyperbolic curvature of the line $x_0 + \mathbb{R}i$ with $x_0 \in (-1, 1)$. Now we observe that $g(-x_0) \in \mathrm{SU}(1,1)$ and $g(-x_0).x_0 = 0$. Since $g(-x_0)$ is a Möbius transformation, the line $L \cup \{\infty\}$ is mapped onto a line or a circle. since $1 - x_0(x_0 + it) = 1 - x_0^2 - ix_0t \neq 0$ for all $t \in \mathbb{R}$, the image must be a circle through the origin. The line $\mathbb{R}$ is invariant under $g(-x_0)$ and perpendicular to $i\mathbb{R}$, hence its image is a diameter of this circle. Now we have $g(-x_0).\infty = -\frac{1}{x_0}$, hence the diameter of this circle is $d = \frac{1}{|x_0|}$. Now the First Circle Lemma already yields $\kappa = |x_0|$. □

2 Hyperbolic angles and distance

For $z_0, z_1 \in \mathbb{H}^2$ we want to compute the hyperbolic distance. Using the $\mathrm{SU}(1,1)$-invariance we obtain:

Lemma 2.1 *For z_0, z_1 the hyperbolic distance is given by*

$$d_H(z_0, z_1) = 2 \operatorname{artanh}\left(\frac{|z_1 - z_0|}{|1 - \overline{z_0} z_1|} \right).$$

Proof. Let

$$g = \frac{1}{\sqrt{1-|z_0|^2}} \begin{pmatrix} 1 & -z_0 \\ -\overline{z_0} & 1 \end{pmatrix} \in \mathrm{SU}(1,1),$$

then $d_H(z_0, z_1) = d_H(g.z_0, g.z_1)$. Now $g.z_0 = 0$ and rotations around the origin are isometries, thus $d_H(z_0, z_1) = d_H(0, |g.z_1|)$. Thus we only need to compute $d_H(0,t)$ for $0 \leq t < 1$. Since

$$z(s) := \exp \frac{s}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} .0 = \tanh(\frac{s}{2})$$

is a geodesic parametrized by arclength, we simply have to solve $\tanh(\frac{s}{2}) = t$ for s. Thus we obtain $d_H(0,t) = 2\,\mathrm{artanh}(t)$, hence

$$d_H(z_0, z_1) = d_H\left(0, \frac{|z_1 - z_0|}{|1 - \overline{z_0}z_1|}\right) = 2\,\mathrm{artanh}\left(\frac{|z_1 - z_0|}{|1 - \overline{z_0}z_1|}\right).$$

□

Remark 2.2 We observe that for $z \in \mathbb{D}$ we have $d_H(-z, z) = 2\,\mathrm{artanh}\left(\frac{2|z|}{1-|z|^2}\right)$.

Next we consider hyperbolic angles. Let three distinct points z_0, z_1, z_2 be given. Then there exist unique geodesic segments through two of these. Thus we obtain a hyperbolic triangle. Figure 2 illustrates this construction. We define the hyperbolic angle $\theta(z_0, z_1, z_2)$ to be the angle between the segments $\overline{z_1 z_0}$ and $\overline{z_1 z_2}$. Since Moebius transformations are conformal mappings, we may shift z_1 to the origin using

$$g = g(z_1) = \frac{1}{\sqrt{1-|z_1|^2}} \begin{pmatrix} 1 & -z_1 \\ -\overline{z_1} & 1 \end{pmatrix} \in \mathrm{SU}(1,1),$$

and $\theta(z_0, z_1, z_2) = \theta(g.z_0, 0, g.z_2) = \angle(g.z_0, g.z_2)$. Using $\arg(z)\colon \mathbb{C} \setminus \{0\} to [0, 2\pi)$ for oriented angles we obtain

$$\theta(z_0, z_1, z_2) = \arg\left(\frac{g.z_2}{g.z_0}\right) = \arg\left(\frac{(z_2 - z_1)(1 - \overline{z_1}z_0)}{(z_0 - z_1)(1 - \overline{z_1}z_2)}\right).$$

Of course we can express the arclength of a circular arc in terms of the corresponding hyperbolic angle and curvature. We obtain

Lemma 2.3 *The arclength of the circular arc of hyperbolic angle θ and constant hyperbolic curvature $\kappa > 1$ is given by*

$$\ell = \frac{\theta}{\sqrt{\kappa^2 - 1}}.$$

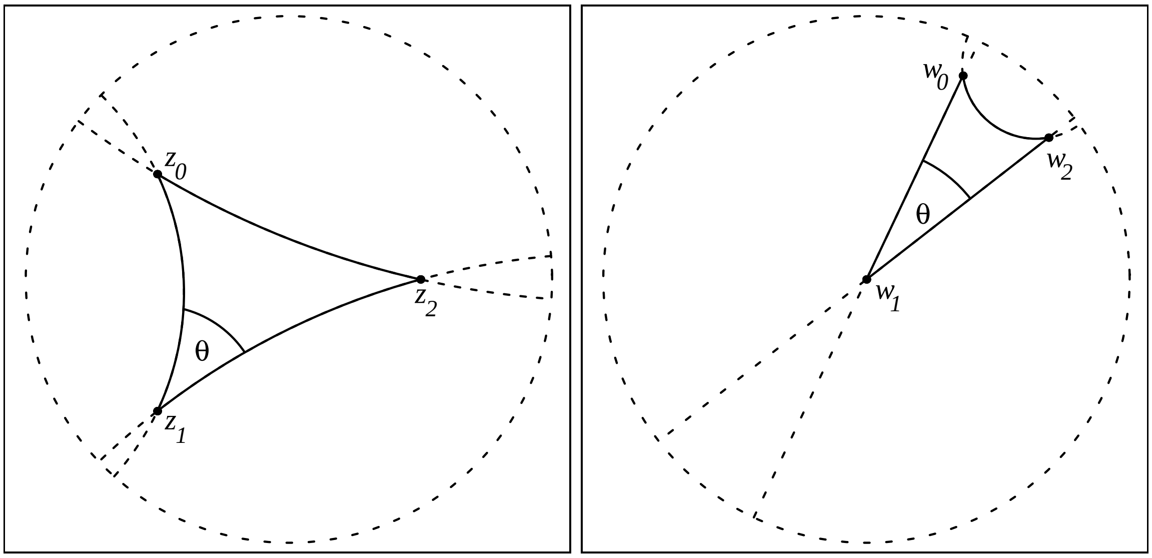

Figure 1: Measuring hyperbolic angles

Proof. Since rotations around the origin are symmetries, it is clear that ℓ is of the form $\frac{\theta}{2\pi}L(\kappa)$ where $L(\kappa)$ is the hyperbolic arclength of the circle with hyperbolic curvature κ. Since the smallest $t > 0$ for which

$$\exp \frac{it}{2}\begin{pmatrix} \kappa & 1 \\ -1 & -\kappa \end{pmatrix} \in \{-\mathbf{1}, \mathbf{1}\} = Z(\mathrm{SU}(1,1))$$

is $\frac{2\pi}{\sqrt{\kappa^2-1}}$, we have $L(\kappa) = \frac{2\pi}{\sqrt{\kappa^2-1}}$, proving our claim. □

We can also define the *hyperbolic center* of a circle $C(z_0, \rho)$ with hyperbolic curvature $\kappa(z_0, \rho) > 1$. This is simply the unique point ζ that has the same hyperbolic distance to all points of $C(z_0, \rho)$. Such a point always exists because we can find a $g \in \mathrm{SU}(1,1)$ such that $g.C(z_0, \rho) = C(0, r)$ with a suitable $r \in (0, 1)$. And for a circle around 0 the hyperbolic center and the euclidean center coincide. One actually obtains $r = \kappa - \sqrt{\kappa^2 - 1}$ where $\kappa = \kappa(z_0, \rho) > 1$. And the hyperbolic center of $C(z_0, \rho)$ is located at $g^{-1}.0$.

Lemma 2.4 *If $0 < |z_0| < 1$ and $\kappa(z_0, \rho) > 1$, then the hyperbolic center of $C(z_0, \rho)$ is located at*

$$\zeta = \frac{1 - \sqrt{1 - 4\rho|z_0|}}{2|z_0|^2} z_0 = \frac{2\rho}{1 + \sqrt{1 - 4\rho|z_0|}} \frac{z_0}{|z_0|}.$$

Proof. Since rotations around the origin are symmetries, we may assume w.l.o.g. that $z_0 \in \mathbb{R}^+$. Let us write $z_0 = x_0 \in (0, 1)$. Now we must solve the equation

$d_H(\zeta, x_0 - \rho) = d_H(\zeta, x_0 + \rho)$. It is clear that $\zeta \in (x_0 - \rho, x_0 + \rho)$. Using the hyperbolic distance formula we obtain

$$d_H(\zeta, x_0 - \rho) = 2\,\mathrm{artanh}\left(\frac{\zeta - (x_0 - \rho)}{1 - \zeta(x_0 - \rho)}\right),\ d_H(\zeta, x_0 + \rho) = 2\,\mathrm{artanh}\left(\frac{(x_0 + \rho) - \zeta}{1 - \zeta(x_0 + \rho)}\right).$$

Since artanh: $\mathbb{R} \to (-1, 1)$ is bijective, this yields

$$\frac{\zeta - (x_0 - \rho)}{1 - \zeta(x_0 - \rho)} = \frac{x_0 + \rho - \zeta}{1 - \zeta(x_0 + \rho)}.$$

This leads to a quadratic equation for ζ with the solutions

$$\zeta_1 = \frac{1 - \sqrt{1 - 4\rho x_0}}{2x_0}, \quad \zeta_2 = \frac{1 + \sqrt{1 - 4\rho x_0}}{2x_0}.$$

First we observe that $x_0 + \rho < 1$, hence $\rho < (1 - x_0)$ and therefore $4x_0\rho < 4x_0(1 - x_0) < 1$ for all $x_0 \in (0, 1)$. Thus solutions for ζ exist. Of course we must have $\zeta < 1$. We claim that $\zeta_2 \geq 1$. In fact, since $\rho < (1 - x)$ we obtain the following estimate:

$$\begin{aligned} \zeta_2 &= \frac{1 + \sqrt{1 - 4\rho x_0}}{2x_0} > \frac{1 + \sqrt{1 - 4(1 - x_0)x_0}}{2x_0} = \frac{1 + \sqrt{1 - 4x_0 + 4x_0^2}}{2x_0} \\ &= \frac{1 + \sqrt{(1 - 2x_0)^2}}{2x_0} = \frac{1 + |2x_0 - 1|}{2x_0} \geq 1 \end{aligned}$$

because for $x_0 \in (0, 1/2]$ we already have $\frac{1}{2x_0} \geq 1$ and for $x \in (1/2, 1)$ we have $1 + |2x_0 - 1| = 2x_0$, so $\frac{1 + |2x_0 - 1|}{2x_0} = 1$. □

We can also define *hyperbolic diameters, semicircles*, and *antipodal points*. A hyperbolic diameter of $C(z, r)$ is a geodesic that intersects $C(z, r)$ perpendicular. Two points on $C(z, r)$ are hyperbolic antipodal points if the unique geodesic through these points is a diameter. And a circular arc is a hyperbolic semicircle if its endpoints are antipodal points. Since all these definitions only use arclengths and angles, it is clear that the familiar properties from euclidean geometry are still valid, i.e., all hyperbolic diameters meet in the hyperbolic center etc.

3 Adhoc results: developping geometric intuition

In the euclidean plane $\mathbb{R}^2$ we could prove a lot of adhoc results that certain arcs cannot be optimal by finding a shorter arc. For example, abnormal extremals were not needed to describe the optimal arcs because all CC-arcs could be obtained from regular extremals and abnormal CCC-arcs could be replaced by a shorter CLC-arc, see [2, 4], for example. One of the key results is Dubins' CCC-Lemma stating that a concatenation of four circular arcs cannot be optimal in the euclidean plane. Now we want to prove similar results in $\mathbb{H}^2$. Thus we want to find different paths for the same Dubins' data. The difficulty is that in principle we have to solve equations of the form

$$\exp(t_1X_1)\exp(t_2X_2)\exp(t_3X_3) = \exp(s_1Y_1)\exp(s_2Y_2)\exp(s_3Y_3)$$

for $t_i, s_i \in \mathbb{R}^+$ and X_i, Y_i corresponding to C-and L-arcs. If we try the brute force approach using matrices in G we will soon get lost in lengthy formulas. The main reason why we could prove those adhoc results was the geometric intuiton. Given two circles in $\mathbb{R}^2$, for example, it is not difficult to *see* the lines that are tangent to both circles. In the hyperbolic plane this is more delicate. Of course we only have to look at circles and straight lines, but we do not *see* the hyperbolic curvature, more precisely, its actual magnitude. Nevertheless if we want to find different paths for the same Dubins' data, eventually everything boils down to quadratic equations since we only have to deal with circles and straight lines. But the formulas are much more complicated than in the euclidean case. And nice formulas for the length of certain arcs are almost impossible to obtain because the computation of hyperbolic angles and distances are required.

4 How to make a U-turn

Let $\kappa > 1$ be given. We want to find a CCC-path for the Dubins' data $x_0 = x_1 = (0,0)$, $v_0 = (1,0)$, $v_1 = (-1,0)$. Let $r = \frac{1}{2\kappa}$. The two C-arcs starting at x_0 in the v_0-direction are the circles of radius r with center at $(0,r)$, resp., $(0,-r)$. Their hyperbolic curvature is

$$\kappa(\pm ir, r) = \frac{1 + r^2 - |\pm ir|^2}{2r} = \frac{1}{2r}.$$

Let $x_0 > 0$ be given, then there exists a unique euclidean circle that is tangent to both C-arcs. Let ρ denote its radius. Then $(r+\rho)^2 = r^2 + x_0^2$, i.e., $\rho^2 + 2r\rho = x_0^2$. In order for this curve to be an extremal, the hyperbolic curvature of $C(x_0,\rho)$ must be equal to κ. Thus we obtain

$$\begin{aligned}
\kappa &= \frac{1+\rho^2 - x_0^2}{2\rho} = \frac{1-2r\rho}{2\rho}, \\
2\rho\kappa &= 1 - 2r\rho, \\
\rho &= \frac{1}{2(\kappa + r)} = \frac{1}{2(r+\frac{1}{2r})} = \frac{r}{1+2r^2}, \\
x_0 &= \sqrt{\rho^2 + 2r\rho} = \frac{r\sqrt{3+4r^2}}{1+2r^2}.
\end{aligned}$$

Alternatively we can try a CLC-arc. In this case the geodesic segment is part of a circle $C(-x_0,\rho)$. Again we have $x_0^2 + r^2 = (r+\rho)^2$, hence $x_0^2 = \rho^2 + 2r\rho$. But this

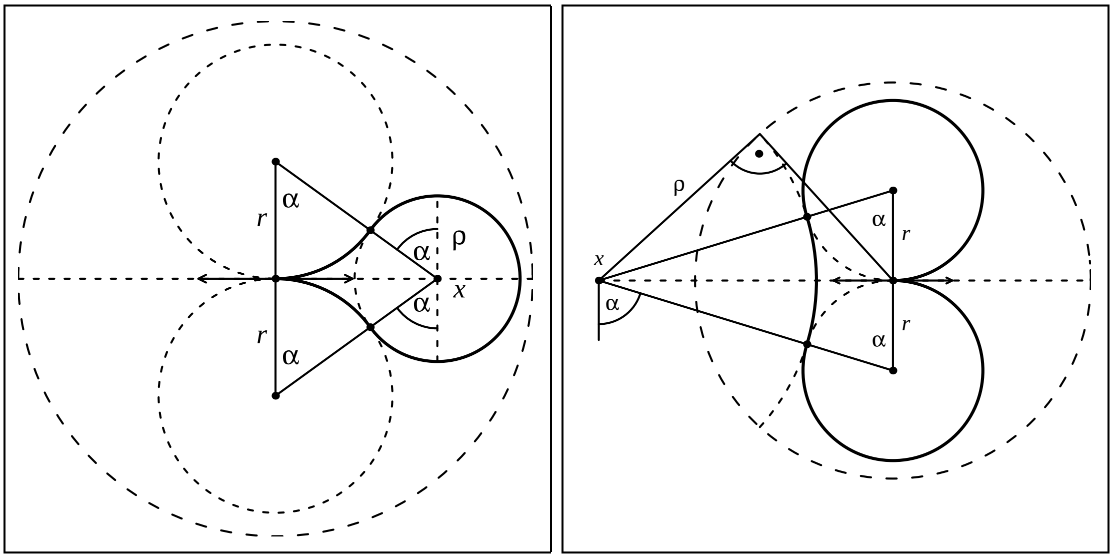

Figure 2: The U-turns of type CCC and CLC

time we must have $\kappa(-x_0, \rho) = 0$. Thus we obtain $0 = 1 + \rho^2 - x_0^2 = 1 - 2r\rho$, hence

$$\rho = \frac{1}{2r} = \kappa > 1, \quad \text{and} \quad x_0 = \sqrt{\rho^2 + 2r\rho} = \sqrt{\kappa^2 + 1} = \frac{1}{2r}\sqrt{1 + 4r^2}.$$

Figure 2 Shows these paths.

5 How to replace certain CCC-arcs by a geodesic segment L

In this section we will prove that an arc with five C's cannot be optimal. Our result will be better and eventually our goal is to prove that as in the euclidean case, an optimal abnormal extremal has at most two arcs. But this will require much more computational work. As an immediate consequence we have reduced the solution of Dubins' problem in $\mathbb{H}^2$ with $\kappa > 1$ to a finite problem. Given a set of Dubuns' data, determine all arcs CLC and $CCCC$ satisfying the given boundary conditions. These are only finitely many. Computation of their length yields the shortest paths.

Now let us first consider a CCC-arc coming from an abnormal extremal. The middle C of such an arc is a *hyperbolic semicircle*. In general, hyperbolic semicircles *need not be euclidean semicircles!* But they are semicircles in the sense that the geodesic joining the endpoints is a diameter, i.e., perpendicular to the circle. Actually one has

Lemma 5.1 *A hyperbolic semicircle is a euclidean semicircle if and only if the geodesic joining the endpoints is a line segment.*

For sake of a symmetric picture we want to have the center of the middle circle at the origin. Then all diameters are straight line segments. Let $r \in (0,1)$ be fixed. The circle $C(0,r)$ has hyperbolic curvature $\kappa(0,r) = \frac{1+r^2}{2r} =: \kappa_0 > 1$. Of course we could also start with $\kappa_0 > 1$, then the equation $\kappa(0,r) = \kappa_0$ has two solutions for r, namely $r = \kappa_0 - \sqrt{\kappa_0^2 - 1}$ and $r = \kappa_0 + \sqrt{\kappa_0^2 - 1}$. The first solution is in $(0,1)$ and the second solution is in $(1,\infty)$. Thus $r = r(\kappa_0) = \kappa_0^2 - \sqrt{\kappa_0^2 - 1}$.

The adjacent circles are of the form $C(r+\tau,\tau)$ and $C(-r-\tau,\tau)$. We obtain for their curvature

$$\kappa(r+\tau,\tau) = \frac{1-(r^2+2r\tau+\tau^2)+\tau^2}{2\tau} = \frac{1-2r\tau-r^2}{2\tau}.$$

Thus we must solve the equation

$$\frac{1-2r\tau-r^2}{2\tau} = \frac{1+r^2}{2r}$$

for τ. We obtain

$$\tau = \frac{r(1-r^2)}{1+3r^2}.$$

The geodesic that is tangent to the adjacent circles is of the form $C(-iy,\rho)$. Now this circle is a geodesic if and only $\kappa(-iy,\rho) = 0$. This is equivalent to $1-y^2+\rho^2 = 0$, i.e., $y^2 = 1+\rho^2$. On the other hand Pythagoras' Theorem yields $(r+\tau)^2 + y^2 = (\rho+\tau)^2$, hence

$$\rho^2 + 2\rho\tau + \tau^2 = \rho^2 + 1 + (r+\tau)^2, \quad \Rightarrow \quad \rho = \frac{1+(r+\tau)^2-\tau^2}{2\tau} = \frac{1+2r\tau+r^2}{2\tau}.$$

For the critical angle α we obtain

$$\alpha = \arctan\left(\frac{y}{r+\tau}\right).$$

Figure 5 shows the geodesic shortcut Eventually we would like to prove that an optimal abnormal extremal has at most two segments, i.e., is CC. This would also imply that every optimal arc could be obtained from a regular extremal. But at least we have proved:

Corollary 5.2 *If an arc coming from an abnormal extremal is optimal, then it has at most four segments. In other words, an arc of the form $CCCCC$ cannot be optimal.*

Proof. An arc coming from an abnormal extremal with more than four segments has a subpath of three consecutive hyperbolic semicircles. This subpath is not optimal because it can be replaced by a shorter CLC-arc. Since every subarc of an optimal arc must be optimal, our claim follows. □

We do not have to require that the middle circle is a hyperbolic semicircle. The same type of argument works in a more general setting. We are primarily interested in the case where the middle circle has hyperbolic angle $\theta \in (\pi, 2\pi)$. Figure 4 shows the underlying construction. Thus we have

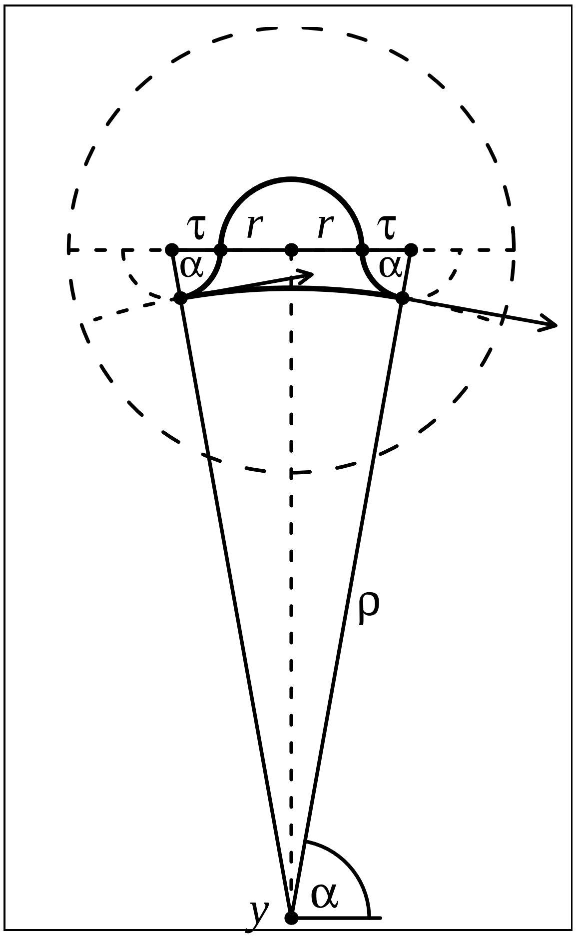

Figure 3: The L-shortcut replacing certain abnormal CCC-arcs

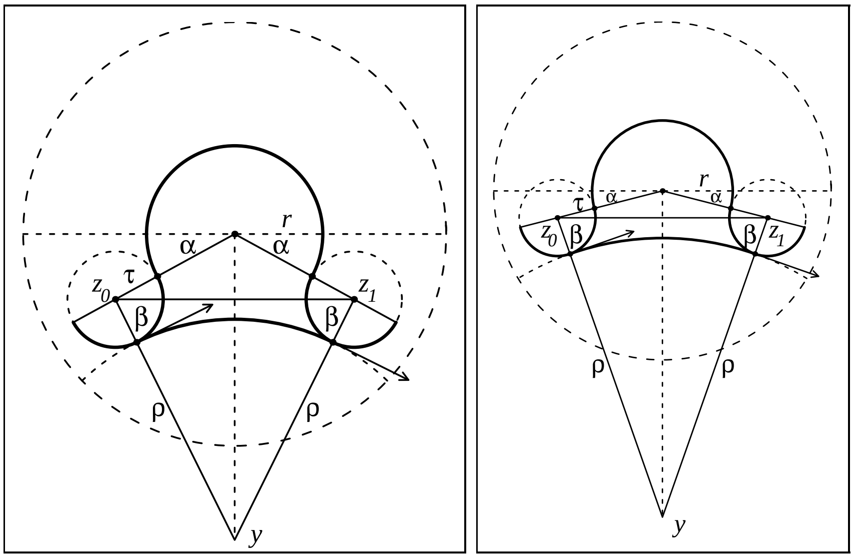

Figure 4: The L-shortcut replacing a CCC-arc

Corollary 5.3 (The 5C-Lemma) *An optimal arc has at most five segments.*

Proof. Let κ_0 be given, again we have $r = \kappa_0 - \sqrt{\kappa_0^2 - 1}$ and τ is given by the same formula as before. With $z_0 = (r+\tau)e^{\pi+\alpha}$ and $z_1 = (r+\tau)e^{-i\alpha}$ we have $z_0 = -\xi - i\eta$ with $\xi = (r+\tau)\cos\alpha \geq 0$ and $\eta = (r+\tau)\sin\alpha \geq 0$. Since the circle $C(z_0, \tau)$ has hyperbolic curvature $\kappa_0 > 1$, it is entirely contained in $\mathbb{D}$, thus $|z_0 - (-i)| > \tau$. Let $\rho(y) = |z_0 + iy| - \tau$, then $\rho(y) > 0$ is obviously true for $y \geq 1$. The circle $C(-iy, \rho(y))$ touches the circles $C(z_0, \tau)$, $C(z_1, \tau)$ from outside. Now we obtain for

$$\hat{\kappa}(y) := \frac{1 - y^2 + \rho(y)^2}{2\rho(y)}$$

that $\hat{\kappa}(1) = \frac{1}{2\rho(1)} > 0$ and

$$\lim_{y\to\infty} \hat{\kappa}(y) = -(\eta + \tau) < 0.$$

Since $\hat{\kappa}$ is continuous in $[1, \infty)$, the existence of a solution $y^* > 1$ of the equality $\hat{\kappa}(y^*) = 0$ follows. The circle $C(-iy^*, \rho(y^*))$ is the geodesic we are looking for. □

Remark 5.4 We can actually give an explicit formula for y^* since it is the solution of a quadratic equation. From $\rho = \sqrt{\xi^2 + (\eta - y)^2} - \tau$ (touching condition) and $1 + \rho^2 - y^2 = 0$ (curvature condition) we obtain

$$0 \;=\; 1 + \rho^2 - y^2$$

$$\begin{aligned}
&= 1+\xi^2+y^2-2y\eta+\eta^2-2\tau\sqrt{\xi^2+(\eta-y)^2}+\tau^2-y^2 \\
2\tau\sqrt{\xi^2+(\eta-y)^2} &= -2y\eta+(1+\xi^2+\eta^2+\tau^2) \\
4\tau^2(\xi^2+\eta^2-2y\eta+y^2) &= y^2\cdot 4\eta^2-y\cdot 4\eta(1+\xi^2+\eta^2+\tau^2)+(1+\xi^2+\eta^2+\tau^2)^2 \\
0 &= y^2\cdot\underbrace{4(\eta^2-\tau)^2}_{=:a_2}+y\cdot\underbrace{(-4\eta)(1+\xi^2+\eta^2-\tau^2)}_{=:a_1} \\
&\quad +\underbrace{(1+\xi^2+\eta^2+\tau^2)-4\tau^2(\xi^2+\eta^2)}_{=:a_0}.
\end{aligned}$$

Now we have to distinguish the three cases $a_2>0,\ =0,\ <0$.

If $a_2<0$, i.e., $\eta<\tau$ the circles $C(z_i,\tau)$ have points in the upper as well as the lower half plane. Of course there are two geodesics touching both circles, one in the upper halfplane corresponding to a negative solution of the quadratic equation and the other one in the lower halfplane corresponding to a positive y. Since $a_2<0$ this positive solution is given by

$$y^*=\frac{-a_1-\sqrt{a_1^2-4a_2a0}}{2a_2}.$$

If $a_2=0$, the quadratic equation is a linear equation. The other geodesic tangent to both circles is the horizontal axis, of course. We get $y^*=-a_0/a_1$. We observe that $a_1<0$ and $a_0>0$ because of $1-\tau^2>0$, and for a_0 one can prove that $a_0=1+2(\xi^2+\eta^2+\tau^2)+(\xi^2+\eta^2-\tau^2)^2$.

Finally if $a_2>0$, then the two solutions of the quadratic equation correspond to a smaller circle touching the circles $C(z_i,\tau)$ from outside and a larger circle containing the circles $C(z_i,\tau)$ in its interior. Thus we have to pick the smaller solution of the quadratic equation which is again given by the formula

$$y^*=\frac{-a_1-\sqrt{a_1^2-4a_2a_0}}{2a_2}.$$

6 CC versus CL

Let $\kappa>1$ and $\alpha\in[0,\pi/2)$ be given. We want to construct a CC- and a CL-arc for the same Dubins data. Let $r=\kappa-\sqrt{\kappa^2-1}$, then $C(0,r)$ has hyperbolic curvature κ. Let

$$\tau=\frac{r(1-r^2)}{1+3r^2},$$

then $C((r+\tau)e^{i\phi},\tau)$ has hyperbolic curvature κ, too. Now fix $\alpha\in[0,\pi/2)$ and let $z_0=-(r+\tau)e^{i\alpha}$, $z_1=(r+\tau)e^{-i\alpha}$. Then the two geodesics simultaneously tangent to $C(z_0,\tau)$ and $C(z_1,\tau)$ are of the form $C(-iy,\rho)$ where $\rho^2+1=y^2$ (vanishing curvature) and y must solve the following equation:

$$(y-(r+\tau)\sin\alpha)^2+(r+\tau)^2\cos^2\alpha=(\rho+\tau)^2=y^2-1-2\tau\sqrt{y^2-1}+\tau^2.$$

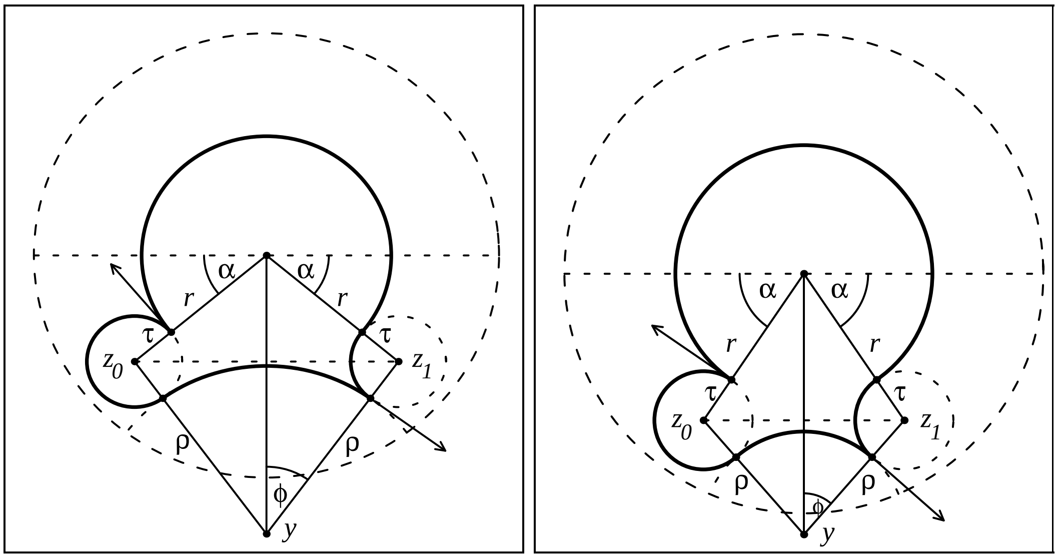

Figure 5: Arcs of type CC and CL for the same Dubins data

Thus we obtain

$$
\begin{aligned}
& y^2 - 1 + 2\tau\sqrt{y^2-1} + \tau^2 = y^2 - 2y(r+\tau)\sin\alpha + r^2 + 2r\tau + \tau^2 \\
2\tau\sqrt{y^2-1} &= 1 + 2r\tau + r^2 - 2y(r+\tau)\sin\alpha \\
4\tau^2(y^2-1) &= 4y^2(r+\tau)^2\sin^2\alpha - 4y(r+\tau)(1+2r\tau+r^2)\sin\alpha + (1+2r\tau+r^2)^2 \\
0 &= y^2\cdot\underbrace{\Big(4(r+\tau)^2\sin^\alpha - 4\tau^2\Big)}_{=:a_2} + y\cdot\underbrace{\Big(-4(r+\tau)(1+2r\tau+r^2)\sin\alpha\Big)}_{=:a_1} \\
&\quad + \underbrace{(1+2r\tau+r^2)^2 + 4\tau^2}_{=:a_0}.
\end{aligned}
$$

This equation will have two solutions. We must pick the smaller one. The bigger solution corresponds to the other geodesic touching $C(z_0,\tau)$ and $C(z_1,\tau)$ from above.

$$
y = \frac{-a_1 - \sqrt{a_1^2 - 4a_2a_0}}{2a_2}.
$$

Figure 5 shows the the corresponding arcs. For α small the CC-arc will be shorter. This is the case in the picture on the lefthand side. On the second picture the CL-arc is slightly shorter than the CC-arc. This happens in general for α large, i.e., sufficiently close to $\frac{\pi}{2}$.

7 Finding CCC and CLC arcs for the same Dubins' data

Now we try to replace a CCC-arc where the middle C is a hyperbolic semicircle and the initial and terminal C have the same length > 0 by a CLC-arc.

Let $z_0 = -(r+\tau)+(\tau+\rho)e^{-i\beta}$. Then we must pick $\rho > 0$ such that $C(z_0,\rho)$ has hyperbolic curvature $\kappa(z_0,\rho) = \kappa_0$. Thus we must solve

$$\frac{1+\rho^2-|z_0|^2}{2\rho} = \kappa_0, \quad \text{i.e.,} \quad 0 = 1-2\rho\kappa_0+\rho^2-\left|-(r+\tau)+(\tau+\rho)e^{-i\beta}\right|^2$$

for ρ. We compute

$$\begin{aligned}
|z_0|^2 &= |-(r+\tau)+(\tau+\rho)(\cos\beta - i\sin\beta)|^2 \\
&= (\rho\cos\beta+(\tau\cos\beta-r-\tau))^2+(\rho+\tau)^2\sin^{\beta} \\
&= \rho^2(\cos^2\beta+\sin^2\beta)+2\rho\,\cos\beta(\tau\cos\beta-r-\tau) \\
&\quad +2\rho\tau\sin^2\beta+(\tau\cos\beta-r-\tau)^2+\tau^2\sin^2\beta \\
&= \rho^2+2\rho\cdot\Big(\cos\beta(\tau\cos\beta-r-\tau)+\tau\sin^2\beta\Big)+(\tau\cos\beta-r-\tau)^2+\tau^2\sin^2\beta \\
&= \rho^2+2\rho\cdot\Big(\tau-r\cos\beta-\tau\cos\beta\Big)+\Big(\tau^2-2\tau\cos\beta(r+\tau)+(r+\tau)^2\Big)
\end{aligned}$$

Thus we obtain

$$\begin{aligned}
0 &= |z_0|^2-\rho^2+2\kappa_0\rho-1 \\
&= 2\rho\Big(\kappa_0+\tau-(r+\tau)\cos\beta\Big)+\Big(-1+\tau^2-2\tau(r+\tau)\cos\beta+(r+\tau)^2\Big) \\
\rho &= \frac{1-\tau^2+2\tau(r+\tau)\cos\beta-(r+\tau)^2}{2(\kappa_0+\tau-(r+\tau)\cos\beta)}.
\end{aligned}$$

Now the next big task is to find a geodesic that is tangent to both circles. Let $\xi = \Re(z_0)$ and $\eta = -\Im(z_0)$, so $z_0 = \xi - i\eta$. By construction $\eta > 0$ because z_0 is below the horizontal axis. We have to distinguish three cases.

(1) If $\rho = \eta$, then the horizontal axis is a geodesic and tangent to both circles. Thus the line segment from $(-\xi, 0)$ to $(\xi, 0)$ is the L of our CLC-arc.

(2) If $\rho < \eta$, then both circles $C(z_0,\rho)$, $C(z_1,\rho)$ are below the horizontal axis. The geodesic tangent to these circles is a euclidean circle. Its center is on the imaginary axis and $C(z_i,\rho)$ are inside this circle. Therefore this circle has the form $C(-iy, |z_0-(-iy)|+\rho)$ with $y \in \mathbb{R}$. The requirement that this is a geodesic means $\kappa(-iy, |z_0+iy|+\rho) = 0$, i.e.,

$$0 = 1-y^2+\left(\sqrt{\xi^2+(y-\eta)^2}+\rho\right)^2.$$

Eventually this will lead us to a quadratic equation for y. Geometrically it is clear that there will be two solutions $y_1 < \eta < y_2$. From these we have to pick the bigger one because the center of the geodesic circle must be below the centers of the small circles $C(z_0,\rho)$, $C(z_1,\rho)$. We obtain

$$\begin{aligned}
0 &= -y^2+1+\rho^2+2\rho\sqrt{\xi^2+(y-\eta)^2}+\xi^2+y^2-2\eta y+\eta^2\\
2\rho\sqrt{\xi^2+(y-\eta)^2} &= 2\eta\cdot y-(1+\rho^2+\xi^2+\eta^2)\\
4\rho^2(\xi^2+y^2-2\eta y+\eta^2) &= 4\eta^2y^2-4\eta(1+\rho^2+\xi^2+\eta^2)\cdot y+(1+\rho^2+\xi^2+\eta^2)^2\\
0 &= y^2\cdot(4\eta^2-4\rho^2)+y\cdot\Big(8\eta\rho^2-4\eta(1+\rho^2+\xi^2+\eta^2)\Big)\\
&\quad+\Big((1+\rho^2+\xi^2+\eta^2)^2-4\rho^2(\xi^2+\eta^2)\Big)\\
&= y^2\cdot\underbrace{4(\eta^2-\rho^2)}_{=:a_2}+y\cdot\underbrace{4\eta(\rho^2-1-\xi^2-\eta^2)}_{=:a_1}\\
&\quad+\underbrace{\Big((1+\rho^2+\xi^2+\eta^2)^2-4\rho^2(\xi^2+\eta^2)\Big)}_{=:a_0}.
\end{aligned}$$

There are two solutions for y. We must take the bigger solution we obtain

$$y=\frac{-a_1+\sqrt{a_1^2-4a_2a_0}}{2a_2}.$$

The second solution also has a meaning. In fact it corresponds to the other geodesic in the lower halfplane simultaneously tangent to $C(z_0,\rho)$, $C(z_1,\rho)$. But we have to change the radius. The lower geodesic is $C(-iy,-\rho+|z_0-iy|)$.

(3) Finally if $\rho>\eta$, then the geodesic tangent to the circles $C(z_0,\rho)$, $C(z_1,\rho)$ is a circle in the upper half plane and touches from outside. For $y\in\mathbb{R}$ the circle $C(iy,|z_0-iy|-\rho)$ touches these circles from outside. In order to be a geodesic, its hyperbolic curvature must vanish, i.e.,

$$0=1-y^2+(|z_0-iy|-\rho)^2=1-y^2+\xi^2+(y+\eta)^2-2\rho\sqrt{\xi^2+(y+\eta)^2}+\rho^2$$

Thus

$$\begin{aligned}
2\rho\sqrt{\xi^2+(y+\eta)^2} &= 2\eta y+\Big(1+\xi^2+\eta^2+\rho^2\Big)\\
4\rho^2(\xi^2+y^2+2\eta y+\eta^2) &= 4\eta^2y^2+4\eta\Big(1+\xi^2+\eta^2+\rho^2\Big)\cdot y+\Big(1+\xi^2+\eta^2+\rho^2\Big)^2\\
0 &= y^2\cdot\underbrace{4(\eta^2-\rho^2)}_{=:a_2}+y\cdot\underbrace{4\eta(1+\xi^2+\eta^2-\rho^2)}_{=:a_1}\\
&\quad+\underbrace{\Big(1+\xi^2+\eta^2+\rho^2\Big)^2-4\rho^2(\xi^2+\eta^2)}_{=:a_0}.
\end{aligned}$$

This quadratic equation has two solutions $y_1<0<y_2$, and we have to pick the positive one. Since $a_2=4(\eta^2-\rho^2)<0$ this means

$$y=\frac{-a_1-\sqrt{a_1^2-4a_2a_0}}{2a_2}.$$

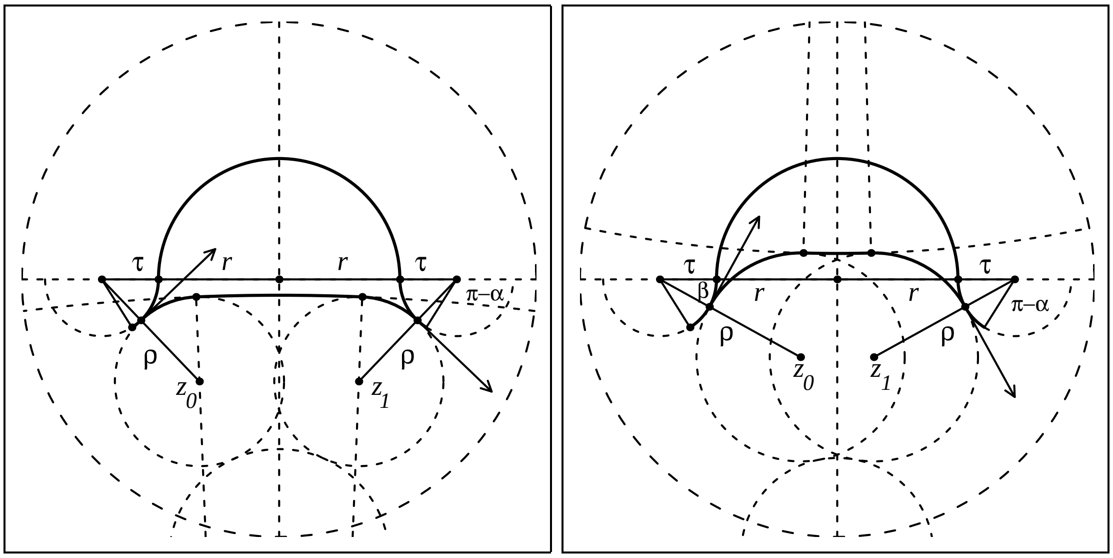

Figure 6: Replacing CCC by CLC

The second solution also has a meaning. In fact it corresponds to the geodesic in the lower halfplane simultaneously tangent to $C(z_0, \rho)$, $C(z_1, \rho)$, but we have to change the radius. The lower geodesic is $C(iy, -\rho + |z_0 - iy|)$.

Figure 6 illustrates cases (2) and (3).

Now the question is, of course, whether the CLC arc found this way is shorter than the CCC arc. Unfortunately the euclidean angles do not convert simply into hyperbolic arclength.

This will be our next task. Thus for q_0, q_1 on $C(z_0, \rho)$, i.e., $q_i = z_0 + \rho e^{i\phi_i}$, $i = 1, 2$, we must find the length of the circular arc from q_0 to q_1.

8 Optimality

Our conjecture is that the CLC-arc is always shorter than the CCC-arc. But it seems to be impossible to prove this directly because the formulas are too complicated. Nevertheless we will prove that the CCC-arc is *not* optimal. For $t \geq 0$ let γ_t denote the arc $C_t^+ C_{\frac{\pi}{w}}^- C_t^+$. From the previous discussion it is clear that for $t > 0$ sufficiently small we can always find an arc $\tilde{\gamma}_t$ of the form $C_r^- L_{2s} C_r^-$ for the same Dubins' data, i.e., initial and terminal position and tangent. We can translate this into an equation on the group $\mathrm{SU}(1,1)$. Let

$$X = \frac{i}{2}\begin{pmatrix} \kappa & 1 \\ -1 & -\kappa \end{pmatrix}, \quad Y = \frac{i}{2}\begin{pmatrix} -\kappa & 1 \\ -1 & \kappa \end{pmatrix}, \quad Z = \frac{i}{2}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Then for $t > 0$ sufficiently small, we can find $r(t), s(t)$ such that

$$\exp(tX)\exp\left(\frac{\pi}{w}Y\right)\exp(tX) = \exp(rY)\exp(2sZ)\exp(rY).$$

For $t = 0$ we take $r(0) = \frac{\pi}{2w}$ and $s(0) = 0$, of course. The paths $\gamma, \tilde{\gamma}$ have length $\ell(\gamma) = 2t + \frac{\pi}{w}$, $\ell(\tilde{\gamma}) = 2r + 2s$. Now let

$$\delta(t) = \ell(\gamma) - \ell(\tilde{\gamma}) = 2t + \frac{\pi}{w} - (2r + 2s).$$

Obviously $\delta(0) = 0$, and $\delta(t) > 0$ means that the CLC-arc is shorter. We will show that $\delta'(0) > 0$. This implies that for small $t > 0$ the CLC-arc $\tilde{\gamma}$ is shorter than γ. Since every subarc of an optimal arc must be optimal, γ cannot be optimal for any $t > 0$.

The idea of proof is to apply the implicit function theorem, compute $r'(0)$ and $s'(0)$ and verify that $\delta'(0) = 2(1 - r'(0) - s'(0)) > 0$. These are lengthy computations but one can handle them.

(1) Since we already *know* that for $t > 0$ small, the equation

$$\exp(tX)\exp\left(\frac{\pi}{w}Y\right)\exp(tX) = \exp(r_1Y)\exp(2sZ)\exp(r_2Y)$$

has a solution with $r_2 = r_1$, we may exploit this knowledge and reduce the complexity of the problem. It is sufficient to consider only the $(1,1)$-entries of the matrices in this equation. Since this is a complex number, we have two equations and three indeterminates.

(2) Let

$$z(r,s,t) = e_1^T\left(\exp(tX)\exp\left(\frac{\pi}{w}Y\right)\exp(tX) - \exp(rY)\exp(2sZ)\exp(rY)\right)e_1$$

and let $x(r,s,t) = \Re(z)$, $y(r,s,t) = \Im(z)$. At $p_0 = (r_0, s_0, t_0) = \left(\frac{\pi}{2w}, 0, 0\right)$ we have $x(p_0) = 0$, $y(p_0) = 0$. We will show that

$$\mathcal{J} := \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial s} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial s} \end{pmatrix}_{(p_0)}$$

is invertible. Then the Implicit Function Theorem (IFT) applies. We find differentiable functions $r(t), s(t)$ defined on an interval $I = (-t_0, t_0)$ such that $(r(0), s(0)) = (\frac{\pi}{2w}, 0)$ and $z(r(t), s(t), t) \equiv 0$. Moreover we obtain $(r'(0), s'(0))$ from the equation

$$\mathcal{J}\begin{pmatrix} r'(0) \\ s'(0) \end{pmatrix} + \begin{pmatrix} \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial t} \end{pmatrix} = 0, \quad \text{i.e.,} \quad \begin{pmatrix} r'(0) \\ s'(0) \end{pmatrix} = -\mathcal{J}^{-1}\begin{pmatrix} \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial t} \end{pmatrix}.$$

(3) Now we compute $z(r,s,t)$. Let $\mu = \frac{w}{2}$ then we obtain the following formulas:

$$X^2 = -\mu^2 I, \quad Y^2 = -\mu^2 I, \quad Z^2 = \frac{1}{4}I, \quad XY + YX = \frac{\kappa^2+1}{2}I,$$

and

$$XYX = \frac{i}{8}\begin{pmatrix} \kappa(3+\kappa^2) & 1+3\kappa^2 \\ -(1+3\kappa^2) & \kappa(3+\kappa^2) \end{pmatrix}.$$

For the exponentials we obtain

$$\begin{aligned}
\exp(tX) &= \cos(\mu t)I + \frac{\sin(\mu t)}{\mu}X, \\
\exp(tY) &= \cos(\mu t)I + \frac{\sin(\mu t)}{\mu}Y, \\
\exp\left(\frac{\pi}{w}Y\right) &= \frac{1}{\mu}Y, \\
\exp(tZ) &= \cosh(t/2)I + 2\sinh(t/2)Z = \begin{pmatrix} \cosh(t/2) & i\sinh(t/2) \\ -i\sinh(t/2) & \cosh(t/2) \end{pmatrix}.
\end{aligned}$$

Next let $g(t) = \exp(tX)\exp\left(\frac{\pi}{w}Y\right)\exp(tX)$. We obtain

$$\begin{aligned}
g(t) &= (\cos(\mu t) + \mu^{-1}\sin(\mu t)X)\cdot \mu^{-1}Y \cdot (\cos(\mu t) + \mu^{-1}\sin(\mu t)X) \\
&= \mu^{-1}(\cos(\mu t)Y + \mu^{-1}\sin(\mu t)XY)(\cos(\mu t)I + \mu^{-1}\sin(\mu t)X) \\
&= \mu^{-1}(\cos^2(\mu t)Y + \mu^{-1}\sin(\mu t)\cos(\mu t)(XY+YX) + \mu^{-2}\sin^2(\mu t)XYX) \\
&= \frac{\kappa^2+1}{2\mu^2}\sin(\mu t)\cos(\mu t)I + \frac{\cos^2(\mu t)}{\mu}Y + \frac{\sin^2(\mu t)}{\mu^3}XYX.
\end{aligned}$$

For $h(r,s) = \exp(rY)\exp(2sZ)\exp(rY)$ we obtain

$$\begin{aligned}
h(r,s) &= \left(\cos(\mu r)I + \frac{\sin(\mu r)}{\mu}Y\right)(\cosh(s)I + \sinh(s)Z)\left(\cos(\mu r)I + \frac{\sin(\mu r)}{\mu}Y\right) \\
&= \cos^2\cosh I + \frac{\sin^2\cosh}{\mu^2}Y^2 + \frac{2\cos\sin\cosh}{\mu}Y + \cos^2\sinh Z \\
&\quad + \frac{\cos\sin\sinh}{\mu}(ZY+YZ) + \frac{\sin^2\sinh}{\mu^2}YZY \\
&= \Big(\cosh(s)\cos(wr) + \frac{\sin(wr)}{w}\sinh(s)\Big)\cdot I \\
&\quad + \cosh(s)\frac{\sin(wr)}{2w}\begin{pmatrix} -i\kappa & i \\ -i & i\kappa \end{pmatrix} + \frac{1+\cos(wr)}{4}\sinh(s)\begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix} \\
&\quad + \frac{(1-\cos(wr))}{2w^3}\sinh(s)\begin{pmatrix} -2i\kappa & i(1+\kappa^2) \\ -i(1+\kappa^2) & 2i\kappa \end{pmatrix}
\end{aligned}$$

Thus $z(r,s,t) = z_1(t) - z_2(r,s)$ with

$$\begin{aligned}
z_1(t) &= e_1^T g(t) e_1 = \frac{\kappa^2+1}{2\mu^2} sc + \frac{c^2}{\mu}\frac{-i\kappa}{2} + \frac{s^2}{\mu^3}\frac{i\kappa(3+\kappa^2)}{8} \\
&= \frac{\kappa^2+1}{\kappa^2-1}\sin(wt) + i\kappa\frac{(3+\kappa^2)s^2 - w^2c^2}{w^3} \\
&= \frac{w^2+2}{w^2}\sin(wt) + i\kappa\frac{(4+w^2)s^2 - w^2c^2}{w^3} \\
&= \frac{w^2+2}{w^2}\sin(wt) + i\kappa\frac{2-(w^2+2)\cos(wt)}{w^3} \\
\tilde{a}(r,s) &= \cosh(s)\cos(wr) + \frac{\sin(wr)}{w}\sinh(s) \\
&\quad -\frac{i\kappa}{2}\left(\cosh(s)\frac{\sin(wr)}{w} + 2\sinh(s)\frac{1-\cos(wr)}{w^3}\right).
\end{aligned}$$

Thus

$$\begin{aligned}
x(r,s,t) &= \frac{2+w^2}{w^2}\sin(wt) - \cosh(s)\cos(wr) - \sinh(s)\frac{\sin(wr)}{w}, \\
y(r,s,t) &= \kappa\,\frac{2-(w^2+2)\cos(wt)}{w^3} + \frac{\kappa}{2}\left(\cosh(s)\frac{\sin(wr)}{w} + 2\sinh(s)\frac{1-\cos(wr)}{w^3}\right).
\end{aligned}$$

For the partial derivatives we find

$$\begin{aligned}
\frac{\partial x}{\partial r} &= -w\cosh(s)\sin(wr) + \sinh(s)\cos(wr), \\
\frac{\partial x}{\partial s} &= \sinh(s)\cos(wr) + \cosh(s)\frac{\sin(wr)}{w}, \\
\frac{\partial x}{\partial t} &= -\frac{2+w^2}{w}\cos(wt), \\
\frac{\partial y}{\partial r} &= -\frac{\kappa}{2}\cosh(s)\cos(wr) - \kappa\sinh(s)\frac{\sin(wr)}{w^2}, \\
\frac{\partial y}{\partial s} &= -\frac{\kappa}{2}\sinh(s)\frac{\sin(wr)}{w} - \kappa\cosh(s)\frac{1-\cos(wr)}{w^3}, \\
\frac{\partial y}{\partial t} &= -\kappa(w^2+2)\frac{\sin(wt)}{w^2}.
\end{aligned}$$

Thus we obtain

$$\frac{\partial(x,y)}{\partial(r,s,t)}\left(\frac{\pi}{2w}, 0, 0\right) = \begin{pmatrix} -w & 1 & -\frac{w^2+2}{w} \\ 0 & -\frac{\kappa}{w^3} & 0 \end{pmatrix}.$$

The first two columns are linear independent, i.e., $\mathcal{J}$ is indeed invertible. We compute

$$\begin{pmatrix} r'(0) \\ s'(0) \end{pmatrix} = -\begin{pmatrix} -w & 1 \\ 0 & -\kappa w^{-3} \end{pmatrix}^{-1} \begin{pmatrix} -w^2+2w \\ 0 \end{pmatrix}$$

$$
\begin{aligned}
&= \frac{w^2+2}{w}\begin{pmatrix} -w^{-1} & -w^2\kappa^{-1} \\ 0 & -w^3\kappa^{-1} \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} \\
&= -\frac{w^2+2}{w^2}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = -\frac{\kappa^2+1}{\kappa^2-1}\begin{pmatrix} 1 \\ 0 \end{pmatrix}.
\end{aligned}
$$

(4) Thus we obtain

$$\delta'(0) = 2(1 - r'(0) - s'(0)) = 2\left(1 + \frac{\kappa^2+1}{\kappa^2-1}\right) > 0.$$

Thus we have proved:

Lemma 8.1 (The Hyperbolic Semicircle Lemma) *For $t > 0$ the arc $C_t^+ C_{\frac{\pi}{w}}^- C_t^+$ is not optimal. In other words, a CCC-arc with a semicircle in the middle is not optimal.*

Remark 8.2 Once again, since we already *know* that nonegative solutions $r(t), s(t)$ exists, we do not need to prove neither their existence nor their being nonnegative. The (IFT) yields existence of solutions, and $r(t) > 0$ for small $t > 0$ simply follows from $r(0) = \frac{\pi}{2w} > 0$. But $s(t) \geq 0$ does *not* follow from the (IFT). If we wanted to use only the IFT we would have to show that the first nonzero higher order derivative of $s(t)$ at 0 is nonnegative. Hopefully this will be the second derivative, $s''(0) > 0$. Of course the (IFT) allows us to compute all higher order derivatives at 0, but it would be a matter of tedious computations to determine $s''(0)$.

As a consequence of the Hyperbolic Semicircle Lemma we obtain that abnormal extremals are not needed in order to find the optimal arcs.

Corollary 8.3 *Every optimal arc comes from a regular extremal.*

9 Hyperbolic Rhombic Configurations

Our goal is to prove Monroy's Lemma in the hyperbolic plane, i.e., for $t > 0$ and $\alpha \in \left(\frac{\pi}{w}, \frac{2\pi}{w}\right)$ the path $C_t^+ C_\alpha^- C_\alpha^+$ is not optimal. In order to get some geometric intuition, we generalize his idea of rhombic configurations to the hyperbolic plane.

We look for a configuration of four circles C_0, C_1, C_2, C_3 such that $C_i C_{i+1}$ touch. The centers of these circles are the vertices of a rhombus. Now we look for the same configuration in hyperbolic space. Let $z_0, \ldots, z_3$ denote these center points, then we can assume w.l.o.g. that for suitable $\xi, \eta > 0$ we have $z_0 = -\xi$, $z_1 = i\eta$, $z_2 = \xi$, $z_3 = -i\eta$. Thus our four circles are

$$C(z_0, r),\ C(z_1, \rho),\ C(z_2, r),\ C(z_3, \rho).$$

Given $\kappa > 0$ we have to find the suitable values for ξ, η, r, ρ. Since all circles must have hyperbolic curvature κ, we obtain

$$\begin{aligned} \xi &= \xi(r,\kappa) = \sqrt{1-2r\kappa+r^2}, \\ \eta &= \eta(\rho,\kappa) = \sqrt{1-2\rho\kappa+\rho^2}. \end{aligned}$$

If $\kappa \leq 1$, then ξ, η are defined for all r, ρ. But if $\kappa > 1$, then we have the additional restriction that $1-2r\kappa+r^2 \geq 0$, $1-2\rho\kappa+\rho^2 \geq 0$. This yields either $r, \rho \leq \kappa - \sqrt{\kappa^2-1}$ or $r, \rho \geq \kappa + \sqrt{\kappa^2-1}$. But since for $\kappa > 1$ the radii r, ρ must be strictly less than 1, the second possibility is not admissible. Hence we must require

$$\max\{r, \rho\} \leq \kappa - \sqrt{\kappa^2-1} =: r_{\max}(\kappa).$$

The requirement that the circles $C(\xi, r)$ and $C(i\eta, \rho)$ touch yields the equation

$$\xi^2 + \eta^2 = (r+\rho)^2 = r^2 + 2r\rho + \rho^2.$$

Using our formulas for ξ and η we obtain

$$\begin{aligned} r^2 + 2r\rho + \rho^2 &= 1 - 2r\kappa + r^2 + 1 - 2\rho\kappa + \rho^2 \\ 2\rho(r+\kappa) &= 2 - 2r\kappa \\ \rho &= \frac{1-r\kappa}{r+\kappa}. \end{aligned}$$

Since *rho* must be nonnegative, we get the additional restriction

$$1 - \kappa r \geq 0.$$

If $\kappa = 0$, then there is 'no' restriction on r. We simply obtain

$$\rho = \frac{1}{r}, \ \xi = \sqrt{1+r^2}, \ \eta = \frac{1}{r}\sqrt{1+r^2}, \quad \text{if } \kappa = 0.$$

This makes sense even for $r = 0$, in that case $z_1 = z_3 = \infty$ and $C(\mathfrak{z}_1, \rho)$ is the horizontal axis. If $\kappa \in (0, 1]$, then we must require $\rho \geq 0$, hence $r \leq \frac{1}{\kappa}$. Finally if $\kappa > 1$, then we must have $r \leq \kappa - \sqrt{\kappa^2-1}$. This implies

$$1 - r\kappa \geq 1 - \kappa(\kappa - \sqrt{\kappa^2-1}) = 1 - \kappa^2 + \kappa\sqrt{\kappa^2-1} = \sqrt{\kappa^2-1}(-\sqrt{\kappa^2-1} + \kappa) > 0.$$

So $\rho \geq 0$ is always satisfied. But the second requirement $\rho \leq \kappa - \sqrt{\kappa^2-1}$ yields

$$\frac{1-r\kappa}{r+\kappa} \leq \kappa - \sqrt{\kappa^2-1}, \qquad \Longleftrightarrow \qquad r \geq \frac{\sqrt{\kappa^2-1}(\kappa - \sqrt{\kappa^2-1})}{2\kappa - \sqrt{\kappa^2-1}} =: r_{\min}(\kappa).$$

The boundary cases $r = r_{\min}$ and $r = r_{\max}$ correspond to degenerate rhombus.

$$\begin{aligned} r = r_{\min} \qquad & \rho = r_{\max}, \ \xi = r_{\min} + r_{\max}, \ \eta = 0, \\ r = r_{\max} \qquad & \rho = r_{\min}, \ \xi = 0, \ \eta = r_{\min} + r_{\max}. \end{aligned}$$

In the first case we have three circles on the horizontal axis, in the second case three circles on the vertical axis.

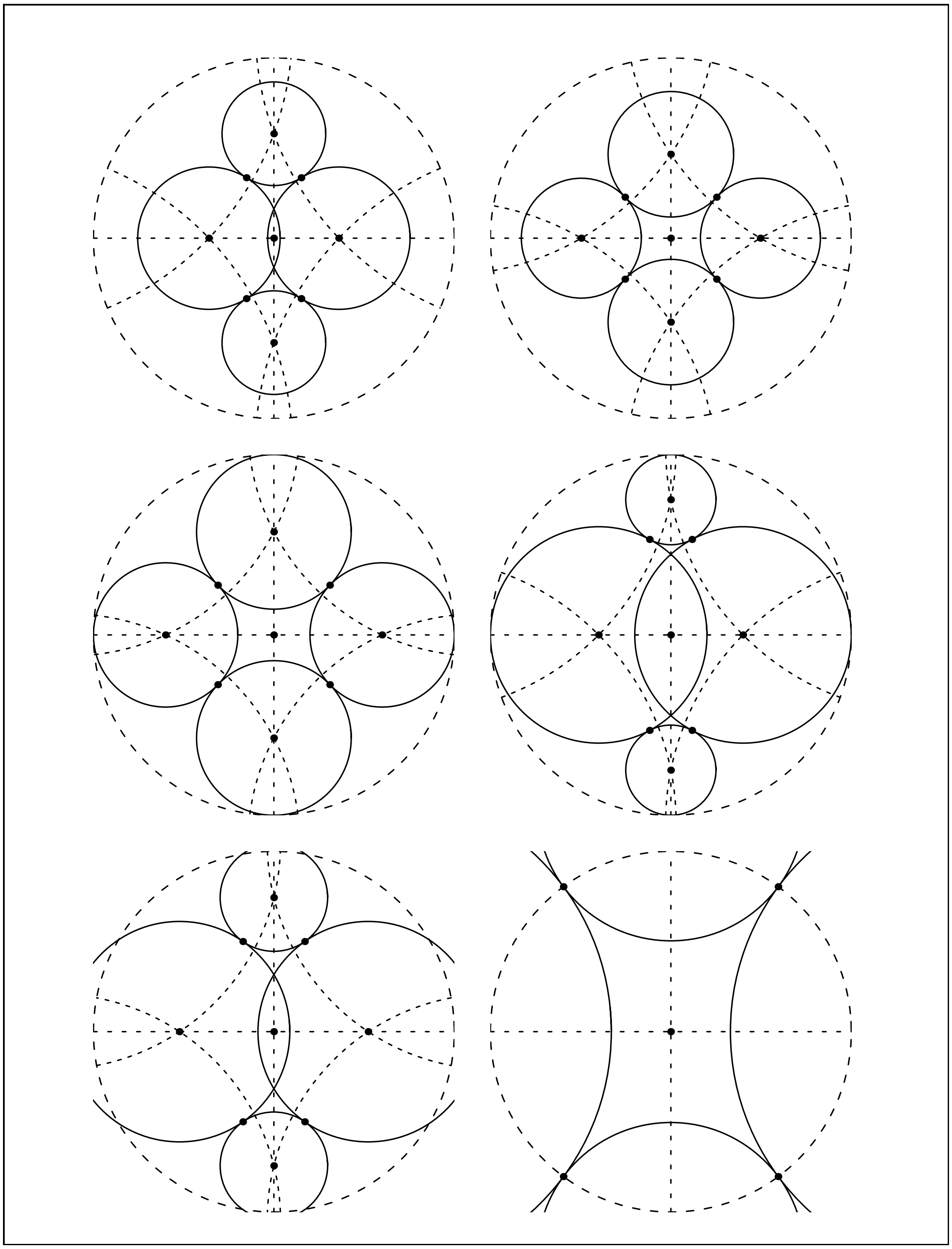

Figure 7: Hyperbolic rhombus, examples with $\kappa > 1$, $\kappa = 1$, $\kappa \in (0,1)$ and $\kappa = 0$

10 Finding a general rhombus

In this section we consider the following problem: suppose we are given two circles $C(w,\rho)$ and $C(z,\tau)$ both of hyperbolic curvature κ_0. Then we want to find ζ and μ such that $C(\zeta,\mu)$ touches both circles and has hyperbolic curvature κ_0. The questions are of course: do solutions exist, and if so, how do we find them?

Lemma 10.1 (The Outer Circle Lemma) *Let $w \in \mathbb{D}$ and $\rho > 0$ be given such that $\kappa(w,\rho) = \kappa_0 > 1$. For $\varphi \in [0,2\pi)$ let*

$$\lambda(\varphi) = \frac{2\kappa_0\rho}{\rho + \kappa_0 + (w, e^{i\varphi})}, \quad \textit{and} \quad \zeta(\varphi) = w + \lambda(\varphi)e^{i\varphi}.$$

Then $\lambda > \rho$, the circle $C(\zeta, \lambda - \rho)$ touches $C(w,\rho)$ from outside and has hyperbolic curvature $\kappa(\zeta, \lambda - \mu) = \kappa_0$.

Proof. First of all we observe that $C(w,\rho)$ must be contained in the open unit disc because of $\kappa_0 > 1$. Thus $|w| + \rho < 1$ follows, hence $|w| < 1 - \rho$ and $\rho + \kappa_0 + (w, e^{i\varphi}) > 2\rho + \kappa_0 - 1 > 0$. Therefore λ is always positive. Now $\lambda > \rho$ is equivalent to $\lambda/\rho > 1$. The latter is equivalent to $2\kappa_0 > \rho + \kappa_0 + (w, e^{i\varphi})$. This is equivalent to

$$\kappa_0 - \rho - (w, e^{i\varphi}) > 0.$$

But this is true because $|w| < 1 - \rho$ yields

$$\kappa_0 - \rho - (w, e^{i\varphi}) \geq \kappa_0 - \rho - |w| \geq \kappa_0 - \rho - (1 - \rho) = \kappa_0 - 1 > 0.$$

This proves $\lambda > \rho$. In order to prove $\kappa(\zeta, \lambda - \rho) = \kappa_0$ we compute $|\zeta - w|^2 = \lambda^2$, and $C(\zeta, \lambda - \rho)$ automatically touches $C(w,\rho)$. Thus we only have to make sure that $\kappa(\zeta, \lambda - \rho) = \kappa_0$. this means we must solve the equation

$$1 + (\lambda - \rho)^2 - |\zeta|^2 = 2\kappa_0(\lambda - \rho).$$

We compute

$$\begin{aligned} |\zeta|^2 &= |w|^2 + 2\lambda(w, e^{i\varphi}) + \lambda^2, \\ 0 &= 1 + \lambda^2 - 2\lambda\rho + \rho^2 - |w|^2 - 2\lambda(w, e^{i\varphi}) - \lambda^2 - 2\kappa_0\lambda + 2\kappa_0\rho, \\ &= \underbrace{1 + \rho^2 - |w|^2}_{=2\kappa_0\rho} + 2\kappa_0\rho - 2\lambda(\rho + \kappa_0 + (w, e^{i\varphi})). \end{aligned}$$

Thus

$$\lambda = \frac{4\kappa_0\rho}{2(\rho + \kappa_0 + (w, e^{i\varphi}))} = \frac{2\kappa_0\rho}{\rho + \kappa_0 + (w, e^{i\varphi})} = \lambda(\varphi)$$

is the unique solution. $\square$

Remark 10.2 In particular, if $w = 0$, then λ does not depend on φ at all which is no surprise since rotations around the origin are symmetries. In that case we have

$$\rho = \kappa_0 - \sqrt{\kappa_0^2 - 1}, \quad \text{and} \quad \lambda = \frac{2\kappa_0\rho}{\rho + \kappa_0}.$$

Now suppose we are given two circles $C(w, \rho)$, $C(z, \tau)$ with the same hyperbolic curvature κ_0. Can we find a third circle $C(\zeta, \mu)$ touching the given circles and with hyperbolic curvature κ? The answer is:

Lemma 10.3 (The Rhombus Condition) *Such a circle $C(\zeta, \mu)$ exists if and only if*

$$|w - z| \leq \rho + \tau + \frac{2\kappa_0\rho|z - w|}{|z - w|(\rho + \kappa_0) + (w, z - w)}.$$

Proof. Let $e^{i\varphi} = \frac{z-w}{|z-w|}$. With $\lambda = \lambda(\varphi)$ and $\zeta = \zeta(\lambda, \varphi)$ as given in the Outer Circle Lemma it is obvious that a solution exists if and only if $C(\zeta, \lambda - \rho)$ intersects $C(z, \tau)$. This happens if and only if $|z - w| \leq \lambda + \rho + \tau$. For λ we have

$$\lambda = \frac{2\kappa_0\rho}{\rho + \kappa_0 + (w, e^{i\varphi})} == \frac{2\kappa_0\rho}{\rho + \kappa_0 + (w, \frac{z-w}{|z-w|})} = \frac{2\kappa_0\rho|z - w|}{|z - w|(\rho + \kappa_0) + (w, z - w)}.$$

Thus

$$|z - w| \leq \rho + \tau + \frac{2\kappa_0\rho|z - w|}{|z - w|(\rho + \kappa_0) + (w, z - w)}$$

follows. □

Now we look for arbitrary circles that touch the given ones.

Lemma 10.4 *Suppose that $z_1, z_2 \in \mathbb{C}$ and $r_1, r_2 > 0$ are such that the two circles $C(z_1, r_1)$, $C(z_2, r_2)$ touch at $z^* \in \mathbb{D}$ and have the same hyperbolic curvature $\kappa(z_1, r_1) = \kappa(z_2, r_2)$. Then they either coincide, $z_1 = z_2$, $r_1 = r_2$, or they touch from outside, i.e., $|z_1 - z_2| = r_1 + r_2$.*

Proof. Let $z^* \in \mathbb{D}$ denote the point where the circles touch. Using $g(z^*) \in \mathrm{SU}(1, 1)$ we can shift z^* to the origin 0. Let C'_k denote the image of the circle C_k under $g(z^*)$. Let $\kappa = \kappa(z_k, r_k)$, then the First Circle Lemma implies that C'_k is of the form $C(\frac{1}{2\kappa}e^{i\phi_k}, \frac{1}{2\kappa})$. Since both circles touch, we have either $\phi_1 = \phi_2$, so the circles coincide, or $\phi_1 = \phi_2 + \pi$, hence they touch from outside. □

Lemma 10.5 (The Touching Circle Lemma) *Let $w, z \in \mathbb{D}$ and $\rho, \tau > 0$ be given such that $\kappa(w, \rho) = \kappa(z, \tau) = \kappa_0 > 1$. For $\varphi \in [0, 2\pi)$ and $\lambda \geq \rho$ let*

$$\zeta(\lambda, \varphi) = w + \lambda(\varphi)e^{i\varphi}.$$

If the circle $C(\zeta, \lambda - \rho)$ touches $C(z, \tau)$ from outside, then

$$\lambda = \frac{(\rho - \tau)^2 - |w - z|^2}{2(\rho - \tau + (e^{i\varphi}, w - z))}.$$

Proof. Suppose that $C(\zeta, \lambda-\rho)$ touches $C(z,\tau)$ from outside, then $|\zeta - z| = \lambda - \rho + \tau$ follows. Now we compute

$$\begin{aligned}
|\zeta - z|^2 &= |\lambda e^{i\varphi} + w - z|^2 = \lambda^2 + 2\lambda(e^{i\varphi}, w-z) + |w-z|^2. \\
0 &= |\zeta - z|^2 - (\lambda - (\rho - \tau))^2 \\
&= \lambda^2 + 2\lambda(e^{i\varphi}, w-z) + |w-z|^2 - \lambda^2 + 2\lambda(\rho-\tau) - (\rho-\tau)^2, \\
\lambda &= \frac{|z-w|^2 - (\rho-\tau)^2}{2((e^{i\varphi}, z-w) - (\rho-\tau))} = \frac{(\rho-\tau)^2 - |z-w|^2}{2((\rho-\tau) + (e^{i\varphi}, w-z))}.
\end{aligned}$$

□

If the Rhombus Condition is satisfied as an equality, then

$$\zeta = w + \left(\rho + \frac{|z-w| - \rho - \tau}{2}\right)\frac{(z-w)}{|z-w|}, \quad \text{and} \quad \mu = \frac{|z-w| - \rho - \tau}{2}$$

is the unique circle of curvature κ_0 touching $C(w,\rho)$ and $C(z,\tau)$ simultaneously. If the Rhombus Condition is satisfied as a strict inequality, then there do exist two circles $C(\zeta_i, \mu_i)$. We want to determine these circles.

Lemma 10.6 *Let $\zeta = \zeta(\varphi) = w + \lambda(\varphi)e^{i\varphi}$ with $\lambda(\varphi)$ as in the Outer Circle Lemma. So $C(\zeta, \lambda - \rho)$ has hyperblic curvature κ_0 and touches $C(w,\rho)$ from outside. Let $\xi = \cos\varphi$ and define $a, b, c \in \mathbb{R}$ through*

$$a + ib = 4\kappa_0\rho(z-w) + ((\rho-\tau)^2 - |z-w|^2)w, \; c = (|z-w|^2 - (\rho-\tau)^2)(\rho+\kappa_0) + 4\kappa_0\rho(\rho-\tau).$$

Suppose that $C(\zeta, \lambda-\rho)$ touches $C(z,\tau)$ from outside. Then ξ solves the quadratic equation

$$(a^2+b^2)\xi^2 - 2ac\xi + c^2 - b^2 = 0, \quad i.e., \quad \xi = \frac{ac \pm b\sqrt{a^2+b^2-c^2)}}{a^2+b^2}.$$

Thus we have up to two solutions ξ_1, ξ_2. For φ we obtain the following solutions:

$$\begin{array}{ll}
b = 0 & \xi_1 = \xi_2 = \frac{c}{a}, \; \varphi_1 = \arccos(\frac{c}{a}), \; \varphi_2 = -\arccos(\frac{c}{a}), \\
b \neq 0 & \varphi_i = \left\{\begin{array}{ll} \arccos\xi_i, & \text{if } \frac{c - a\xi_i}{b} \geq 0, \\ -\arccos\xi_i, & \text{otherwise} \end{array}\right., i = 1, 2.
\end{array}$$

Proof. We have two expressions for $\lambda = \lambda(\varphi)$, one from the Outer Circle Lemma, the other one from the Touching Circle Lemma. Thus we obtain the equation

$$\frac{(\rho-\tau)^2 - |w-z|^2}{2(\rho - \tau + (e^{i\varphi}, w-z))} = \frac{2\kappa_0\rho}{\rho + \kappa_0 + (w, e^{i\varphi})}.$$

Thus

$$((\rho-\tau)^2 - |w-z|^2)(\rho + \kappa_0 + (w, e^{i\varphi})) = 4\kappa_0\rho(\rho - \tau + (e^{i\varphi}, w-z)),$$

and therefore

$$\left(e^{i\varphi}, \underbrace{((\rho-\tau)^2 - |w-z|^2)w + 4\kappa_0\rho(z-w)}_{=:a+ib}\right) =$$
$$= \underbrace{4\kappa_0\rho(\rho-\tau) - ((\rho-\tau)^2 - |w-z|^2)(\rho+\kappa_0)}_{=:c}.$$

Thus we must solve the equation $a\cos\varphi + b\sin\varphi = c$. Geometrically this means we must find the intercepts of the line $ax + by = c$ with the unit circle $x^2 + y^2 = 1$. With $\xi = \cos\varphi$ we obtain

$$(c - a\xi)^2 = b^2(1-\xi^2), \quad \text{hence} \quad (a^2+b^2)\xi^2 - 2a\xi + c^2 - b^2 = 0.$$

Thus we obtain

$$\xi = \frac{ac \pm \sqrt{a^2c^2 - (a^2+b^2)(c^2-b^2)}}{a^2+b^2} = \frac{ac \pm b\sqrt{a^2+b^2-c^2}}{a^2+b^2}.$$

Let ξ_1, ξ_2 denote the solutions of this equation. We must determine the corresponding angles φ_i. Of course $\cos\varphi_i = \xi_i$, so we have at most two possibilities for φ_i. Now we must distinguish two cases.

If $b = 0$, then $\xi_1 = \xi_2 = \frac{c}{a}$ and $\varphi_1 = \arccos(c/a)$, $\varphi_2 = -\arccos(c/a)$ are the solutions for φ.

If $b \neq 0$, then we observe that $a\cos\varphi_i + b\sin\varphi_i = c$ implies $\sin\varphi_i = \frac{c-a\xi_i}{b}$. Since the sine is positive in $(0,\pi)$ and negative in $(-\pi,0)$, we deduce

$$\varphi = \begin{cases} \arccos\xi_i, & \text{if } \frac{c-a\xi_i}{b} > 0, \\ -\arccos\xi_i, & \text{if } \frac{c-a\xi_i}{b} < 0. \end{cases}$$

Finally, if $a\xi_i - c = 0$, then $0 = (c-a\xi_i)^2 = b^2(1-\xi_i^2)$ and $b \neq 0$ implies $\xi_i^2 = 1$. Thus $\xi_i \in \{-1, 1\}$ and $\varphi_i = \arccos\xi_i$ is the only solution of $\cos\varphi = \xi_i$ in $(-\pi,\pi]$. □

Figure 8 shows the alternative CCC-path.

11 The proof of the CCC-Lemma

In this section we prove

Lemma 11.1 (The Hyperbolic Monroy-Lemma) *The path $C_t^+C_\alpha^-C_\alpha^+$ is not optimal for any $t > 0$ and $\alpha \in \left(\frac{\pi}{w}, \frac{2\pi}{w}\right)$.*

In connection with the Hyperbolic Semicircle Lemma we immediately obtain:

Lemma 11.2 (The Hyperbolic CCC-Lemma) *An optimal arc has at most three segments, i.e., it is either CLC or CCC or a subarc of one of these.*

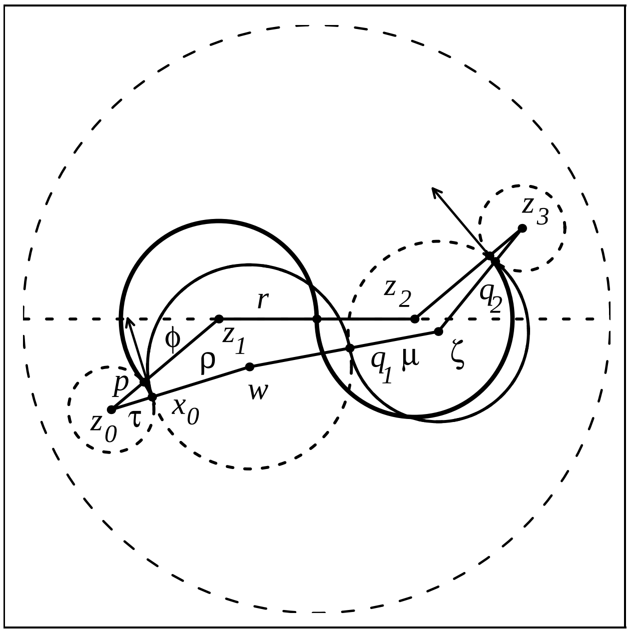

Figure 8: Replacing $C_t^+C_\alpha^-C_\alpha^+$ with $C_{s_1}^-C_{s_2}^+C_{s_3}^-$

Proof of the Hyperbolic Monroy Lemma. We consider the following function $F\colon \mathbb{R}^4 \to \mathrm{SU}(1,1)$:

$$F(t, s_1, s_2, s_3) = \exp(-tX)\exp(s_1 Y)\exp(s_2 X)\exp(s_3 Y).$$

For $t \geq 0$ let $\gamma = C_t^+C_\alpha^-C_\alpha^+$. We have already seen that for $t \geq 0$ sufficiently small we can find $s(t) = (s_1(t), s_2(t), s_3(t)) \in \mathbb{R}_+^3$ such that

$$\exp(tX)\exp(\alpha Y)\exp(\alpha X) = \exp(s_1 Y)\exp(s_2 X)\exp(s_3 Y).$$

In other words, if we let $g_0 = F(0, \alpha, \alpha, 0)$, then we have

$$F(t, s_1(t), s_2(t), s_3(t)) \equiv F(0, \alpha, \alpha, 0) = g_0.$$

We follow the same strategy as before, i.e., we want to apply the IFT. This will prove, once again, the existence of $s(t)$, but it does not necessarily yield the nonnegativity of $s(t)$. Nevertheless we do not have to worry about that because we already have a different existence proof which also yields nonnegativity. It does not necessarily yield smoothness, though.

(2) For an application of the IFT we must show that

$$\frac{\partial F}{\partial s_1}, \frac{\partial F}{\partial s_2}, \frac{\partial F}{\partial s_3} \in T_{g_0}\,\mathrm{SU}(1,1) \cong \mathfrak{su}(1,1)$$

are linearly independent. We obtain

$$\begin{aligned}
\frac{\partial F}{\partial s_1} &= \exp(-tX)Y\exp(s_1Y)\exp(s_2X)\exp(s_3Y)\\
&= \left(e^{-t\,\mathrm{ad}(X)}Y\right)F(t,s_1,s_2,s_3) = M_1(t)\,F(t,s),\\
\frac{\partial F}{\partial s_2} &= \exp(-tX)\exp(s_1Y)X\exp(s_2X)\exp(s_3Y)\\
&= \left(e^{-t\,\mathrm{ad}(X)}e^{s_1\,\mathrm{ad}(Y)}X\right)F(t,s_1,s_2,s_3) = M_2(t,s_1)\,F(t,s),\\
\frac{\partial F}{\partial s_3} &= \exp(-tX)\exp(s_1Y)\exp(s_2X)Y\exp(s_3Y)\\
&= \left(e^{-t\,\mathrm{ad}(X)}e^{s_1\,\mathrm{ad}(Y)}e^{s_2\,\mathrm{ad}(X)}Y\right)F(t,s_1,s_2,s_3) = M_3(t,s_1,s_2),\\
\frac{\partial F}{\partial t} &= -X\,F(t,s_1,s_2,s_3).
\end{aligned}$$

At $a=(0,\alpha,\alpha,0)$ we have

$$\begin{aligned}
\frac{\partial F}{\partial s_1}(a) &= Y\,g_0, & \frac{\partial F}{\partial s_2}(a) &= \left(e^{\alpha\,\mathrm{ad}(Y)}X\right)g_0,\\
\frac{\partial F}{\partial s_3}(a) &= \left(e^{\alpha\,\mathrm{ad}(Y)}e^{\alpha\,\mathrm{ad}(X)}Y\right)g_0, & \frac{\partial F}{\partial t}(a) &= -X\,g_0.
\end{aligned}$$

Let

$$M_1^0 = Y, \quad M_2^0 = e^{\alpha\,\mathrm{ad}(Y)}X, \quad \text{and} \quad M_3^0 = e^{\alpha\,\mathrm{ad}(Y)}e^{\alpha\,\mathrm{ad}(X)}Y.$$

We must show that $M_1^0, M_2^0, M_3^0 \in \mathfrak{su}(1,1)$ are linearly independent, i.e., a basis of $\mathfrak{su}(1,1)$. Once we have done that, the IFT applies, proving the existence of differentiable—in fact even analytic—functions $s_i(t)$. And the derivative $s_i'(0)$ are obtained from the linear equation

$$M_1^0 s_1'(0) + M_2^0 s_2'(0) + M_3^0 s_3'(0) = X.$$

In order to compute the M_i's and verify their being linear independent we must pick a basis of $\mathfrak{su}(1,1)$ and compute the adjoint action.

(3) Let

$$v_1 = \frac{i}{2}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}, \quad v_2 = \frac{1}{2}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}, \quad v_3 = \frac{i}{2}\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix},$$

then the brackets are $[v_1,v_2]=v_3$, $[v_1,v_3]=-v_2$ and $[v_2,v_3]=-v_1$. With respect to this basis we have

$$X = \begin{pmatrix}\kappa\\ 0\\ 1\end{pmatrix}, \quad Y = \begin{pmatrix}-\kappa\\ 0\\ 1\end{pmatrix}, \quad \mathrm{ad}(X) = \begin{pmatrix}0 & 1 & 0\\ 1 & 0 & -\kappa\\ 0 & \kappa & 0\end{pmatrix}, \mathrm{ad}(Y) = \begin{pmatrix}0 & 1 & 0\\ 1 & 0 & \kappa\\ 0 & -\kappa & 0\end{pmatrix}.$$

Since $\mathrm{ad}(X)^3 = -w^2\,\mathrm{ad}(X)$ we obtain

$$\begin{aligned}
e^{\alpha\,\mathrm{ad}(X)} &= I + \frac{\sin(w\alpha)}{w}\,\mathrm{ad}(X) + \frac{1-\cos(w\alpha)}{w^2}\,\mathrm{ad}(X)^2 \\
&= \begin{pmatrix} \frac{\kappa^2-\cos(w\alpha)}{w^2} & \frac{\sin(w\alpha)}{w} & -\frac{\kappa(1-\cos(w\alpha))}{w^2} \\ \frac{\sin(w\alpha)}{w} & \cos(w\alpha) & -\frac{\kappa\sin(w\alpha)}{w} \\ \frac{\kappa(1-\cos(w\alpha))}{w^2} & \frac{\kappa\sin(w\alpha)}{w} & \frac{-1+\kappa^2\cos(w\alpha)}{w^2} \end{pmatrix} \\
e^{\alpha\,\mathrm{ad}(Y)} &= e^{\alpha\,\mathrm{ad}(X)^T} = \left(e^{\alpha\,\mathrm{ad}(X)}\right)^T \\
&= \begin{pmatrix} \frac{\kappa^2-\cos(w\alpha)}{w^2} & \frac{\sin(w\alpha)}{w} & \frac{\kappa(1-\cos(w\alpha))}{w^2} \\ \frac{\sin(w\alpha)}{w} & \cos(w\alpha) & \frac{\kappa\sin(w\alpha)}{w} \\ -\frac{\kappa(1-\cos(w\alpha))}{w^2} & -\frac{\kappa\sin(w\alpha)}{w} & \frac{-1+\kappa^2\cos(w\alpha)}{w^2} \end{pmatrix}.
\end{aligned}$$

We compute

$$\begin{aligned}
e^{\alpha\,\mathrm{ad}(Y)}X &= \begin{pmatrix} \frac{\kappa^2-\cos(w\alpha)}{w^2} & * & \frac{\kappa(1-\cos(w\alpha))}{w^2} \\ \frac{\sin(w\alpha)}{w} & * & \frac{\kappa\sin(w\alpha)}{w} \\ -\frac{\kappa(1-\cos(w\alpha))}{w^2} & * & \frac{-1+\kappa^2\cos(w\alpha)}{w^2} \end{pmatrix} \begin{pmatrix} \kappa \\ 0 \\ 1 \end{pmatrix} \\
&= \begin{pmatrix} \dfrac{\kappa^3+\kappa-2\kappa\cos(w\alpha)}{w^2} \\ 2\dfrac{\kappa}{w}\sin(w\alpha) \\ \dfrac{2\kappa^2\cos(w\alpha)-\kappa^2-1}{w^2} \end{pmatrix} \\
e^{\alpha\,\mathrm{ad}(X)}Y &= \begin{pmatrix} \dfrac{-\kappa^3-\kappa+2\kappa\cos(w\alpha)}{w^2} \\ -\dfrac{2\kappa\sin(w\alpha)}{w} \\ \dfrac{2\kappa^2\cos(w\alpha)-\kappa^2-1}{w^2} \end{pmatrix}
\end{aligned}$$

The following computations were only done on a computer (under MATHEMATICA):

$$e^{\alpha\,\mathrm{ad}(Y)}e^{\alpha\,\mathrm{ad}(X)}Y = \begin{pmatrix} \dfrac{\kappa(1-4\kappa^2-\kappa^4+2(1+3\kappa^2)\cos(w\alpha)-4\cos^2(w\alpha))}{w^4} \\ \dfrac{2\kappa(2\cos(w\alpha)\sin(w\alpha)-(\kappa^2+1)\sin(w\alpha))}{w^3} \\ \dfrac{1+3\kappa^4-2\kappa^2(3+\kappa^2)\cos(w\alpha)+4\kappa^2\cos^2(w\alpha)}{w^4} \end{pmatrix}.$$

Let $M = (M_1^0, M_2^0, M_3^0) = \left(Y,\ e^{\alpha\,\mathrm{ad}(Y)}X,\ e^{\alpha\,\mathrm{ad}(Y)}e^{\alpha\,\mathrm{ad}(X)}Y\right)$. We obtain

$$\det(M) = -4\frac{\kappa^2\sin(w\alpha)}{w}.$$

In other words: $M = M(\alpha)$ is invertible if and only if $\sin(w\alpha) \neq 0$, i.e., $\alpha \notin \frac{\pi}{w}\mathbb{Z}$. We obtain $\dot{s}(0) = (\dot{s}_1(0), \dot{s}_2(0), \dot{s}_3(0))$ through

$$M\dot{s}(0) = X, \quad \text{i.e.,} \quad \dot{s}(0) = M^{-1}\begin{pmatrix} \kappa \\ 0 \\ 1 \end{pmatrix}.$$

We obtain

$$s_1'(0) = \frac{2\cos(w\alpha) - \kappa^2 - 1}{w^2}, \quad s_2'(0) = \frac{\kappa^2 + 1 - 2\cos(w\alpha)}{w^2}, \quad s_3'(0) = 1.$$

The path $\gamma = C_t^+ C_\alpha^- C_\alpha^+$ has length $\ell(\gamma) = t + 2\alpha$. The path $\tilde{\gamma} = C_{s_1}^- C_{s_2}^+ C_{s_3}^-$ fitting the same Dubins' data exists for all sufficiently small t if $w\alpha \notin \pi\mathbb{Z}$. For us only the case $w\alpha \in (\pi, 2\pi)$ is of interest. The path $\tilde{\gamma}$ has length $s_1(t) + s_2(t) + s_3(t)$. Let $\delta(t) = \ell(\gamma) - \ell(\tilde{\gamma}) = t + 2\alpha - s_1 - s_2 - s_3$. Then $\delta(0) = 0$ because of $s_1(0) = s_2(0) = \alpha$, $s_3(0) = 0$. With the above values for $s_i'(0)$ we obtain $\delta'(0) = 1 - 1 = 0$. In other words, we must compute the second derivative $s''(0)$. Hopefully this will yield that $\delta''(0) = -(s_1''(0) + s_2''(0) + s_3''(0)) > 0$. With the matrices $M_i(t,s)$ we obtain that

$$s_1'(t)e^{-t\,\mathrm{ad}(X)}Y + s_2'(t)e^{-t\,\mathrm{ad}(X)}e^{s_1\,\mathrm{ad}(Y)}X + s_3'(t)e^{-t\,\mathrm{ad}(X)}e^{s_1\,\mathrm{ad}(Y)}e^{s_2\,\mathrm{ad}(X)}Y \equiv X.$$

This is equivalent to

$$s_1'(t)Y + s_2'(t)e^{s_1\,\mathrm{ad}(Y)}X + s_3'(t)e^{s_1\,\mathrm{ad}(Y)}e^{s_2\,\mathrm{ad}(X)}Y \equiv X.$$

Of course $s(t)$ will be analytic because $F(t,s)$ is analytic. Thus we can differentiate the last equality. This yields

$$\begin{aligned} 0 \;\equiv\; & s_1''(t)Y + s_2''(t)e^{s_1\,\mathrm{ad}(Y)}X + s_3''(t)e^{s_1\,\mathrm{ad}(Y)}e^{s_2\,\mathrm{ad}(X)}Y + s_2'(t)s_1'(t)\,\mathrm{ad}(Y)e^{s_1\,\mathrm{ad}(Y)}X \\ & + s_3'(t)\Big(s_1'(t)\,\mathrm{ad}(Y)e^{s_1\,\mathrm{ad}(Y)}e^{s_2\,\mathrm{ad}(X)} + s_2'(t)e^{s_1\,\mathrm{ad}(Y)}\,\mathrm{ad}(X)e^{s_2\,\mathrm{ad}(X)}\Big)Y. \end{aligned}$$

In particular for $t = 0$ we obtain (recalling $s_3'(0) = 1$ and $s_2'(0) = -s_1'(0)$):

$$\begin{aligned} 0 \;=\; & s_1''(0)M_1^0 + s_2''(0)M_2^0 + s_3''(0)M_3^0 - (s_2'(0))^2[Y,X] \\ & - s_2'(0)[Y, e^{\alpha\,\mathrm{ad}(Y)}e^{\alpha\,\mathrm{ad}(X)}Y] + s_2'(0)e^{\alpha\,\mathrm{ad}(Y)}e^{\alpha\,\mathrm{ad}(X)}[X,Y], \\ =\; & M\,s''(0) - (s_2'(0))^2[Y,X] - s_2'(0)[Y, M_3^0] + s_2'(0)[e^{\alpha\,\mathrm{ad}(Y)}X, e^{\alpha\,\mathrm{ad}(Y)}e^{\alpha\,\mathrm{ad}(X)}Y], \\ =\; & Ms''(0) - s_2'(0)\Big(s_2'(0)[Y,X] + [Y, M_3^0] - [M_2^0, M_3^0]\Big). \end{aligned}$$

Thus we obtain

$$s''(0) = s_2'(0)M^{-1}\left(s_2'(0)[Y,X] + [Y, M_3^0] - [M_2^0, M_3^0]\right).$$

Note that

$$s_2'(0) = \frac{\kappa^2 + 1 - 2\cos(w\alpha)}{w^2} = \frac{\kappa^2 - 1 + 2(1 - \cos(w\alpha)}{w^2} > 0,$$

thus we only need to show that $(1,1,1)M^{-1}\left(s_2'(0)[Y,X]+[Y,M_3^0]-[M_2^0,M_3^0]\right)<0$. Again we use MATHEMATICA in order to compute this expression. We obtain

$$s_1''(0)+s_2''(0)+s_3''(0)=s_2'(0)\,\frac{8(1-\cos(w\alpha))(\kappa^2-\cos(w\alpha))}{w^3\sin(w\alpha)}.$$

Now for $w\alpha\in(\pi,2\pi)$, the numerator is strictly positive because of $1-\cos(w\alpha)>0$ and $\kappa^2-\cos(w\alpha)>0$. The denominator will be strictly negative because $\sin(w\alpha)<0$. Hence $\delta''(0)=-(s_1''(0)+s_2''(0)+s_3''(0))$ is strictly positive. This proves that the arc $\tilde{\gamma}$ is strictly shorter than the arc γ, at least for $t>0$ sufficiently small. Hence the Hyperbolic Monroy Lemma follows.

References

[1] L. E. Dubins. On curves of minimal length with a constraint on average curvature and with presribed initial and terminal positions and tangents. *Am. J. Math.*, 79:497–516, 1957.

[2] R. Felipe Monroy Pérez. Non-Euclidean Dubins' problem: A control theoretic approach. PhD thesis, University of Toronto, 1995.

[3] John G. Ratcliffe. *Foundations of Hyperbolic Manifolds*, volume 149 of *Graduate Texts in Mathematics*. Springer, Berlin, New York, 1994.

[4] J.A. Reeds and L.A. Shepp. Optimal paths for a car that goes both forwards and backwards. *Pacific Journal of Mathematics*, 145:367–393, 1990.

[5] Héctor J. Sussmann. Shortest 3-dimensional paths with a prescribed curvature bound. In *Proceedings of the 1995 IEEE Conference on Decision and Control*, pages 3306–3312, 1995.