Mobius-invariant Lorentz geometry of the space of circles on S^2

Ben Johnsrude
Advisor: David Barrett

Abstract

The space of circles and lines on the Riemann sphere has a natural topological 3-manifold structure. Coxeter observed that a geometry on this collection exists such that Mobius transformations are isometries. Properly defined, this geometry is double covered by (1+2)-dimensional de Sitter space. We discuss some of the properties circle space inherits from de Sitter space, as well as features original to the circle space.

Let $\mathscr{C}$ denote the orbit of the unit circle in the Riemann sphere $\widehat{\mathbb{C}}$ under action of the Mobius group $\mathscr{M}$; thus $\mathscr{C}$ is the "space of circles on $\widehat{\mathbb{C}}$," equivalently defined as the space of planar circles together with lines union the point at infinity.

We may parameterize $\mathscr{C}$ as pairs (ζ, ρ) of spherical center and spherical radius, respectively, with the caveat that, as $\mathscr{C}$ consists of *unoriented* circles, we identify

$$(\zeta, \rho) \sim (-\zeta, 1 - \rho) \tag{1}$$

Here we understand radius in the sense of the chordal metric

$$q(z, w) := \frac{|z - w|}{\sqrt{1 + |z|^2}\sqrt{1 + |w|^2}}$$

With this, $\mathscr{C}$ takes the topological structure of a 3-manifold

$$\mathscr{C} \cong (S^2 \times \mathbb{R})/\sim$$

where $\sim$ follows (1) above. An equivalent topology may be found by considering triples $(z_1, z_2, z_3) \in \widehat{\mathbb{C}}^3$; any three such distinct points determine a unique $c \in \mathscr{C}$, and for each triple of points determining a circle, each of the points may be varied along the circular arc to determine the same circle.

Classical inversive geometry [**1**] considers this collection with a group of rigid motions known as *circle inversions*: for a given circle $c \in \mathscr{C}$, this is the complex-analytic reflection across c, given in the case $|z| = 1$ by

$$w \mapsto \frac{w}{|w|^2}$$

Note that $\tilde{\mathscr{M}} := \langle \mathscr{M}, z \mapsto \bar{z} \rangle$ is exactly the group generated by circle inversions.

We may define a corresponding distance function δ^{inv} by noting the following (cf. [**2**]):

1. Two disjoint circles may be $\mathscr{M}$-mapped to the pair $|z| = 1, |z| = r > 1$
2. Two tangent circles may be $\mathscr{M}$-mapped to any other such pair
3. Two overlapping circles may be mapped to $\mathbb{R} \cup \{\infty\}, e^{i\theta}(\mathbb{R} \cup \{\infty\})$ with $\theta \in (0, \frac{\pi}{2}]$

The corresponding distance function is then defined

$$\delta^{\text{inv}}(c_1, c_2) := \begin{cases} \log r & \text{in case 1} \\ 0 & \text{in case 2} \\ i\theta & \text{in case 3} \end{cases} \tag{2}$$

By the previous note, $\delta^{\text{inv}}(c_1, c_2) = \delta^{\text{inv}}(c_3, c_4)$ precisely when $\phi(c_1) = c_3, \phi(c_2) = c_4$ for some $\phi \in \mathscr{M}$.

It may be noticed that δ^{inv} satisfies

$$\cosh(\delta^{\text{inv}}((\zeta_1, \rho_1), (\zeta_2, \rho_2))) = \frac{\rho_1^2 + \rho_2^2 - |\zeta_2 - \zeta_1|^2}{2\rho_1\rho_2} \tag{3}$$

when (ζ, ρ) are (center, radius) coordinates in *Euclidean* geometry, rather than spherical as above, and apply only for Euclidean circles. On occasion it is useful to consider an analytic version of δ^{inv} using (3); this version of δ^{inv} is multivalued and, for two disjoint circles which are not nested in the plane, has imaginary part $(n + 1)\pi$.

Any two $c_1, c_2 \in \mathscr{C}$ are joined by a geodesic in this metric. Coxeter [**2**] has demonstrated that the geodesic through c_1, c_2 is constructed in the following manner:

- There exists a 1-parameter family (*pencil*) of circles each of which is orthogonal to c_1, c_2; call this the orthogonal pencil to c_1, c_2.
- Every pair of circles from this orthogonal pencil form the same pencil orthogonal to themselves. Call this family the pencil spanned by c_1, c_2.

The pencil spanned by c_1, c_2 is the geodesic passing through them. They have the following properties:

- If c_1, c_2 overlap (i.e. intersect twice), the pencil they span is exactly the family of circles passing through the two points of intersection
- If c_1, c_2 are disjoint, the pencil they span limits to the two points common to every circle in the pencil orthogonal to c_1, c_2, and every two circles in the pencil spanned by c_1, c_2 are also disjoint
- If c_1, c_2 are tangent, the pencil they span is the family of circles all passing through the common point of c_1, c_2 with the same tangent direction at that point

As an example, take two circles c_1, c_2 concentric about 0. The pencil of circles orthogonal to both c_1, c_2 is the pencil of Euclidean lines passing through 0. Thus the geodesic passing through c_1, c_2 is the collection of all circles concentric about 0; clearly these limit to $0, \infty$ and are pairwise disjoint.

Any geodesic of this type, that is, any non-intersecting pencil spanned by disjoint c_1, c_2 is the $\mathscr{M}$-image of this example.

$\mathscr{C}$ is double-covered by the space of oriented circles $\mathscr{C}_{\circlearrowleft}$. On this space, topologically $S^2 \times \mathbb{R}$, we have that $(\delta^{\text{inv}})^2$ has infinitesimal form

$$\frac{d\rho^2 - |d\zeta|^2}{\rho^2} \tag{4}$$

which is a Lorenz metric. It was noticed as early as 1968 by Coxeter [**3**] that this Lorenz manifold is isometric to $(1+2)$ de Sitter space S^3_1, which may be regarded as the subspace of $(1+3)$ Minkowski space $\mathbb{R}^4_1$ by

$$S^3_1 = \{x \in \mathbb{R}^4_1 : \langle x, x \rangle = -1\}$$

where $\langle \cdot, \cdot \rangle$ is the Minkowski inner product

$$\langle (x_0, x_1, x_2, x_3), (y_0, y_1, y_2, y_3) \rangle = x_0 y_0 - x_1 y_1 - x_2 y_2 - x_3 y_3$$

Given a circle $c \in \mathscr{C}$ and a geodesic γ through c, choosing an orientation on c corresponds to choosing an orientation on every circle of γ. We note that, for γ non-intersecting or tangent, the two choices for orientations on c give two different and consistent orientations on the other circles of γ. However, intersecting pencils γ' form a loop, and following the loop around once reverses orientations. Thus intersecting pencils have only one lift to $\mathscr{C}_{\circlearrowleft}$.

We may now categorize the δ^{inv} trichotomy by the lift to S^3_1: disjoint circles $c_1, c_2 \in \mathscr{C}$ are called *timelike separated*, and the geodesic connecting them is called *timelike*. Similarly, tangent circles are said to be *lightlike separated*, and overlapping circles are *spacelike separated*.

1 Lifts of polygons

Consider polygons in the circle space $\mathscr{C}$. Every point of $\mathscr{C}$ lifts to two antipodal points of the de Sitter space S^3_1. If we restrict attention to polygons of $\mathscr{C}$ with disjoint vertices, every edge lifts to a pair of edges in S^3_1 connecting the appropriately-oriented circle vertices of $\mathscr{C}$. Hence every n-gon in $\mathscr{C}$ lifts in S^3_1 either to a $2n$-gon or to a pair of n-gons with distinct vertices.

To distinguish the two cases, one may begin by assigning a starting vertex v_1 in $\mathscr{C}$ with an orientation, then proceeding along edges and assigning the corresponding orientations. When we return to v_1, the pair v_{n-1}, v_1 will either have corresponding orientations or opposite ones. In the first, our polygon lifts to two n-gons. In the second, we obtain a lifted $2n$-gon such that vertices v_j and v_{j+n} are antipodal.

By a Mobius transformation we may take any finite collection of elements in $\mathscr{C}$ to the plane. In this setting, we may distinguish the two cases by counting the number of times an edge goes between non-nested circles, as each corresponds to a change in orientation. The parity of the number of these edges determines the two cases above.

Note that both possibilities are realizable for each $n > 2$: to produce two lifted n-gons, we start with any concentric collection of n circles and form a cycle; to produce a lifted $2n$-gon, we start with $n-1$ concentric circles, with edges drawn in order of descending radius, and add the nth to the ring domain bounded by circles $n-2$ and $n-1$ such that the nth circle does not separate those circles.

Further note that there exists a collection of circles in the plane for which different choices of edges determine different lift types: take three concentric circles with a fourth added between the two larger circles, not separating them. If the concentric circles are α, β, γ in descending radius order, and the fourth circle is δ, the following edges determine different lift types:

$$\{(\alpha, \beta), (\beta, \gamma), (\gamma, \delta), (\delta, \alpha)\} \text{ lifts to a } 2n\text{-gon}$$

$$\{(\alpha, \beta), (\beta, \delta), (\delta, \gamma), (\gamma, \alpha)\} \text{ lifts to two } n\text{-gons}$$

This construction generalizes to any $n > 3$.

2 Tempolateral triangles

Following the discussion above, we see that triangles in $\mathscr{C}$ which have nondegenerate timelike edges, which we call tempolateral triangles, fall into two categories: those that lift to two triangles in S^3_1, and those that lift to a hexagon in S^3_1. By considering the components of the mutual complement of the vertices of our triangle, we see that those triangles which lift to triangles in S^3_1 are topologically equivalent to the triangle generated by a concentric triple of circles. The vertices of a triangle lifting to a hexagon are topologically equivalent to three unnested planar circles. We distinguish the two cases by calling them *nested* and *unnested* triangles, respectively.

Theorem 1. *Let abc be a nested tempolateral triangle, such that the components F_1, F_2 of $\widehat{\mathbb{C}} - (a \cup b \cup c)$ adjacent to b (i.e. $b \subseteq \overline{F_1} \cap \overline{F_2}$) are topologically equivalent to annuli (i.e. are rings). Then the reverse triangle inequality holds:*

$$\delta^{inv}(a, c) \geq \delta^{inv}(a, b) + \delta^{inv}(b, c) \tag{5}$$

Additionally, if $a'b'c'$ is another nested tempolateral triangle with the same edge lengths, there is a Mobius transformation $\phi \in \mathscr{M}$ satisfying $\phi(a) = a', \phi(b) = b', \phi(c) = c'$.

Proof. The inequality (5) may be seen by either constructing a useful inequality in case 1 of (2), or by making an argument exploiting the hyperbolic geometry in one component of $\widehat{\mathbb{C}} - b$, or most directly by lifting to S^3_1 and using the inequality there (see e.g. [**4**, **5**]). Note that, in the latter consideration, the distinguished vertex b in $\mathscr{C}$ corresponds to the vertex in S^3_1 with one incident edge in the past light cone, the other in the future light cone.

The uniqueness condition also follows from the lifted antipodal de Sitter triangles. □

However, a qualitatively different situation occurs among unnested triangles:

Theorem 2. *Among unnested tempolateral triangles all triples $(r_1, r_2, r_3) \in \mathbb{R}^3_+$ are realized as side lengths. Further, any two unnested triangles with the same side lengths are Mobius equivalent.*

One immediate fact is very useful in proving the above result:

Lemma 1. *If abc is a unnested tempolateral triangle, there is some $\phi \in \mathscr{M}$ for which $\phi(a) = \{z : |z| = r > 1\}, \phi(b) = \{z : |z| = 1\}, \phi(c) = \{z : |z - \zeta| = \rho, \zeta \geq 0, \zeta + \rho < r, \zeta - \rho > 1\}$*

Proof. By a Mobius transformation, we may send a, c to a concentric pair about 0. Scale appropriately, then rotate. $\square$

By this normalization, all tempolateral triangles are parameterized by some $1 < a < b < r$, where a, b are the points of intersection of the normalized $\phi(c)$ with $\mathbb{R}$, and r is the radius of the encompassing circle. Clearly all such triples realize unnested tempolateral triangles. The missing facts - that all $1 < a < b < r$ generate all side length triples, and that any two such parameters generate the same side length triple precisely when the parameters coincide - can be seen by examining (3). On the question of uniqueness, we also present a topological proof, with the intermediate step:

Lemma 2. *Suppose $a = (0, r), b = (0, 1), c = (\zeta, \rho) \in \mathbb{C} \times \mathbb{R}_+$ with $\zeta \in \mathbb{R}, \zeta + \rho < r, \zeta - \rho > 1$ are three circles in the plane, parameterized by (center, radius) in Euclidean coordinates. Then there is a conformal automorphism of the component of $\widehat{\mathbb{C}} - a$ containing b, c sending b, c to circles centered on $\mathbb{R}_+$, b to the left of c, such that the Euclidean radius of the image of c is less than the Euclidean radius of the image of b.*

We omit the proof of this intuitively clear result. From this, we have the proof of uniqueness among unnested tempolateral triangles:

Proof. Let $abc, a'b'c'$ be two tempolateral triangles with the same side length. By a Mobius transformation we send them to the normalized form previously considered. Each of these new triples can be Mobius transformed to keep a, a' fixed, while moving b, c, b', c' to the right such that the Euclidean radius of b is larger than that of c, and the same for the other two. By performing both of these transformations, we may assume $a = a', b = b'$, along with the radius condition.

We now consider all possible differences between c and c', and show that in each case we get different side lengths. By moving the center of c to the left, the distance from a increases and the distance from b decreases; moving to the right does the opposite. By increasing the radius of c, we decrease the distance both to a and b; by decreasing the radius we increase this distance. Thus, by combining a shift and dilation of c to reach c', we find that one of the side lengths has either increased both times or decreased both times, giving a different collection of side lengths. As Mobius transformations preserve lengths, we see that the only way abc has the same side lengths of $a'b'c'$ is that their normalized forms above are identical. By combining the normalizing Mobius maps, we obtain a Mobius transformation from abc to $a'b'c'$. $\square$

3 Reflection in a plane

We restrict our attention to the Lorenzian plane

$$\mathscr{L} = \{c \in \mathscr{C} : c \text{ is centered on } \mathbb{R} \cup \{\infty\}\}$$

Consider timelike geodesic γ in $\mathscr{L}$. A Mobius transformation takes γ to the geodesic of circles centered at the origin with positive radii γ_C. Consider isometries ϕ of $\mathscr{C}$ such that $\phi(\mathscr{L}) = \mathscr{L}$ and $\phi(c) = c$ for all $c \in \gamma_C$. From the familiar fact that conformal automorphisms of annuli are Mobius transformations, the only such isometries are Id, $z \mapsto \bar{z}$, $z \mapsto -z$, and $z \mapsto -\bar{z}$. Only the latter two act nontrivially on $\mathscr{L}$, and their actions are identical there; we set reflection in

γ_C, denoted R_{γ_C}, to be that isometry. Reversing the Mobius normalization we obtain R_γ as inversion with respect to the unique element in $\mathscr{L}$ orthogonal to every element of γ. We do not attempt to extend R_γ to all of $\mathscr{C}$.

We consider the problem of reflecting triangles through their sides to produce a "tiling" of the ambient space. Suppose we wish to reflect unnested tempolateral triangle T in $\mathscr{L}$ through the geodesic connecting vertices c_1, c_2. A Mobius transformation takes T to the triangle in $\mathscr{L}$ with vertices $|z| = 1, |z| = r > 1, |z - \zeta| < \rho$ with $1 < \zeta - \rho < \zeta + \rho < r$; we relabel this triangle T, and the vertices c_1, c_2, c_3, respectively. Then reflection $R_{\{c_1,c_2\}}$ sends c_3 to $-c_3$. By generality, we have that reflection of a tempolateral unnested triangle through any of its edges produces a new tempolateral unnested triangle isometric to the first.

If c_3' is the new vertex produced by $R_{\{c_1,c_2\}}$, we may apply a Mobius transformation fixing the interior of c_2 sending c_3' to be centered at 0, and sending c_1, c_3 to the right of c_3'. Then we reflect with $R_{\{c_3',c_2\}}$ to produce c_2' left of c_3' and contained in c_1. By repeating this we "reflect around c_1"; rather than overlapping with our starting point in finitely many steps, as would happen in Euclidean geometry, we produce a sequence of circles of smaller and smaller radius which accumulate at $-r$ and are pairwise disjoint. In terms of de Sitter geometry, this may be called "accumulation at the light cone."

Consider now reflection across the other edges of T. The other edge of T containing c_2 produces reflections opposite to those of the first edge. $R_{\{c_1,c_3\}}$ reflects c_2 to a circle between c_1, c_3, not separating them.

Topological considerations give the following:

- Every circle other than c_2 that is generated by a sequence of reflections is contained in the finite disk bounded by c_2
- If any sequence of reflections across edges produces a circle containing another circle, then the larger circle is c_2
- Every circle generated is reached in a unique minimal sequence of reflections, where "minimal" means that no reflection is carried out twice in a row
- No two circles generated by any sequence of reflections intersect each other
- The closure of the points of intersection of these circles with the real axis is a Cantor set

Note also that unnested circles do not separate $\mathscr{L}$ into two regions; thus, the triangulations we consider do not 'cover the plane' in the usual sense.

Consider an arbitrary timelike geodesic γ. For any $c_1, c_2 \in |\gamma|$ distinct, we may map them to a concentric pair c_1', c_2' centered about 0 by a Mobius transformation ϕ. The image geodesic $\phi(\gamma)$ then is a timelike geodesic intersecting with the geodesic of circles centered about 0, γ_C, at the two points c_1', c_2'. By the inherited geometry of S_1^3, two disjoint circles have precisely one geodesic passing between them, so $\phi(\gamma) = \gamma_C$.

Now note that

$$\bigcup_{c \in \gamma_C} c = \widehat{\mathbb{C}} - \{0, \infty\}$$

As ϕ is a homeomorphism of $\widehat{\mathbb{C}}$,

$$\bigcup_{c\in\gamma} c = \widehat{\mathbb{C}} - \{\phi^{-1}(0), \phi^{-1}(\infty)\}$$

and $\phi^{-1}(0), \phi^{-1}(\infty)$ are distinct. Further, for any $a, b \in \widehat{\mathbb{C}}$ distinct, there is a Mobius transformation ψ with $\psi(a) = 0, \psi(b) = \infty$; then the timelike geodesic $\psi^{-1}(\gamma_C)$ satisfies

$$\bigcup_{c\in\psi^{-1}(\gamma_C)} c = \widehat{\mathbb{C}} - \{a, b\}$$

We thus may regard timelike geodesics of $\mathscr{C}$ as being parameterized by unordered distinct pairs of points in $\widehat{\mathbb{C}}$. For any geodesic γ, the pair $a, b \in \widehat{\mathbb{C}}$ is called the ideal boundary of γ.

Theorem 3. *Let γ_1, γ_2 be timelike geodesics with ideal boundaries $\{a_1, b_1\}, \{a_2, b_2\}$, respectively. Then γ_1, γ_2 intersect precisely when*

1. *a_1, b_1, a_2, b_2 are all distinct*
2. *a_1, b_1, a_2, b_2 lie in a common circle $w \in \mathscr{C}$ (equivalently, γ_1, γ_2 are coplanar)*
3. *The pair a_1, b_1 lie in the same component of $w - \{a_2, b_2\}$.*

Proof. Suppose first that γ_1, γ_2 intersect at circle c. Then we may find $c_1 \in \gamma_1, c_2 \in \gamma_2$ with c, c_1, c_2 pairwise disjoint; then c, c_1 uniquely determine γ_1, and c, c_2 uniquely determine γ_2, and we may apply a Mobius transformation taking c to $|z| = 1$,c_1 to $|z| = r > 1$, and c_2 to $|z-\alpha| = \rho$ disjoint with $\alpha \in \mathbb{R}_+$. As the images of c, c_1, c_2 lie on $\mathbb{R}$, we see that the images of γ_1, γ_2 consist entirely of circles centered on $\mathbb{R} \cup \{\infty\}$; thus the limit points of γ_1, γ_2 lie on the common circle $w = \phi^{-1}(\mathbb{R} \cup \{\infty\})$.

Now consider the images of a_1, b_1, a_2, b_2 under ϕ. By construction, $\phi(a_1) = 0, \phi(b_1) = \infty$. Suppose that a_1, b_1, a_2, b_2 are not all distinct; without loss of generality assume $\phi(b_2) = \infty$. Then $\phi(\gamma_2)$ is the geodesic of circles centered about $\phi(a_2)$. Since $\gamma_1 \neq \gamma_2$, $\phi(a_2) \neq 0$. Thus $\phi(\gamma_1)$ is disjoint from $\phi(\gamma_2)$, so γ_1 is disjoint from γ_2, contradicting our initial assumption. So we may conclude that a_1, b_1, a_2, b_2 are distinct points on the common circle w.

Lastly, we may note that, since $\phi(a_2), \phi(b_2)$ are distinct points on $\mathbb{R}$, the geodesic $\phi(\gamma_2)$ is parameterized by centers $(-\infty, a_2) \cup (b_2, +\infty) \cup \{\infty\}$ where we assume $\phi(a_2) < \phi(b_2)$. Since γ_1 intersects γ_2, one such center must be at 0, so $\phi(a_2), \phi(b_2)$ must be on the same side of $\mathbb{R} - \{0\}$. Thus, up to relabeling a_1, b_1 and a_2, b_2, we have that a_1, b_1 lie in the same component of $w - \{a_2, b_2\}$.

The reverse implication may be seen by reading the above arguments in reverse. □

Note that in hyperbolic 3-space H^3, two geodesics intersect exactly when they are coplanar, have distinct ideal boundary points, and satisfy the *opposite* condition as condition 3 above: a_1, b_1 ideal boundary points of γ_1 need to lie in distinct components of $w - \{a_2, b_2\}$, where w is the circle on the boundary determining the common plane of γ_1, γ_2.

4 4-point triangle inequality

Consider four disjoint unnested circles $c_1, \ldots, c_4$ in S^2. By Theorem 2 we may freely assign positive distances among c_1, c_2, c_3. We may further assign positive distances $\delta^{\text{inv}}(c_1, c_4)$ and $\delta^{\text{inv}}(c_2, c_4)$; by doing so, and putting c_1, c_2, c_3 in Mobius normalized form, we determine c_4 up to the argument of its center. By varying this argument we vary $\delta^{\text{inv}}(c_3, c_4)$, but not without bound! Depending on the previous distances, $\delta^{\text{inv}}(c_3, c_4)$ may have lower bound 0 or > 0, but $\delta^{\text{inv}}(c_3, c_4)$ maximizes when the center of c_3 and the center of c_4 are opposite each other with respect to 0. Hence

Theorem 4. *There exists a triangle inequality among unnested tempolateral tetrahedra.*

Bowers and Stephenson introduced in [**6**] a generalization of circle packings, called inversive distance circle packings, such that edges in the triangulation of S^2 represent inversive distance parameters $s \in i[0, 2\pi) \cup \mathbb{R}_+$, such that classical tangency relations are represented by inversive distance 0. Some work has been done on the uniqueness and rigidity of such packings (see [**7**]), but the theory of existence, especially among packings with positive inversive distance parameters, is not yet established. Theorem 4 gives an important limitation on the question of existence.

References

[1] H.S.M. Coxeter, *Inversive Geometry*. 1971.

[2] H.S.M. Coxeter, *Inversive distance*. 1966.

[3] H.S.M. Coxeter, *The Problem of Apollonius*. 1968.

[4] Immanuel Asmus, *Duality between hyperbolic and de Sitter geometry*. 2009.

[5] Dzan Jin Jee, *Trigonometric Laws on Lorentzian Sphere S^2_1*. 1985.

[6] Philip L. Bowers and Kenneth Stephenson, *Uniformizing Dessins and Belyi Mappings via Circle Packings*. 2004.

[7] Jiming Ma and Jean-Marc Schlenker, *Non-rigidity of Spherical Inversive Distance Circle Packings*. 2012.