Abstract. In this paper, we prove that any group of diffeomorphisms acting on the 2-sphere and properly extending the conformal group of Möbius transformations must be at least 4-transitive or, more precisely, arc 4-transitive. In addition, we show that any such group must always contain an element of positive topological entropy. We also provide an elementary characterization, in terms of transitivity, of the Möbius transformations within the full group of sphere diffeomorphisms.

Contents

1. Introduction 1
1.1. Preliminaries 2
1.2. Paper outline 5
2. Extensions of Möbius 5
3. A Fundamental Lemma 7
4. A Crossing Lemma 12
5. Conclusions 14
5.1. Proof of Theorem A 14
5.2. Discussion on Theorem B 16
6. Characterization of the conformal group in terms of transitivity 19
6.1. Proof of Theorem C 20
Acknowledgements 21
References 21

1. Introduction

Let $M$ be a closed and oriented topological manifold, and consider the set of all its orientation preserving homeomorphisms, denoted simply by $\text{Homeo}(M)$. The usual uniform convergence metric turns this set – endowed with the composition operation $\circ$ – into a topological group, the subgroups of which one can try to understand and classify.

In his early 2000s’ essay [1], É. Ghys proposed such a classification scheme for closed and transitive groups acting on the unit circle $S^1$. Here, closed refers to the uniform topology mentioned above, while transitive means that any given point $p$ can be mapped onto another given point $q$ via some transformation in the group. The corresponding result, which we quote later on this paper, was proven in 2006 by J. Giblin and V. Markovic, see [2].

A relevant part of understanding the closed subgroups of $\text{Homeo}(M)$ is to deal not only with their inclusions, but also with questions of maximality. In other words, determining whether or not between a given subgroup and the full group of homeomorphisms one may find proper intermediate subgroups, up to their closures. For example, F. Le Roux proved in [3] that, in triangulable manifolds of dimension two or higher, the group of area preserving homeomorphisms is always maximal.
Specializing in the unit sphere $S^2$, F. Kwakkel and the second author derived in [4] a number of results concerning important subgroups of $\text{Homeo}(S^2)$, one of which is $\text{Möb}(S^2)$, the Möbius group of conformal 2-sphere diffeomorphisms. Among others, they left the question of whether the Möbius group is maximal within the full group of homeomorphisms, meaning that there would be no proper intermediate group of diffeomorphisms between $\text{Möb}(S^2)$ and $\text{Homeo}(S^2)$, up to uniform closure. \footnote{[4] is an unpublished work. Due to a difference of opinion between its authors, it remains as an unsubmitted preprint. Note that we reference here the v2 in arXiv, not the v3. In any case, this note, being mostly self-contained, makes no direct use of any result in that work.}

This question is known to have a positive answer in the case of the circle, and the proof there is related to higher orders of transitivity. In this paper we provide insight, from the transitivity viewpoint, into extensions of the Möbius group acting on the sphere, as stated below.

**Theorem A.** Let $G \subset \text{Diff}^1(S^2)$ be a group that is a proper extension of $\text{Möb}(S^2)$. Then, its identity component $G_0$ is arc 4-transitive. In particular, $G$ is at least 4-transitive.

Above, the stronger concept of arc-transitivity is introduced. Although more precisely defined in the text to follow, it essentially means that not only ordered lists $(p_1, \ldots, p_4)$ and $(q_1, \ldots, q_4)$ of four points can be mapped one onto another via a single transformation in the group $G$, each $q_i$ is actually the endpoint of $p_i$'s trajectory under an isotopy in $G$ starting at the identity map.

A straightforward application of Theorem A, using both the techniques developed in its proof and some basics of Nielsen-Thurston classification theory, is the following result. It could also be derived with Theorem A and an abstract result from [4]).

**Theorem B.** Let $G \subset \text{Diff}^1(S^2)$ be a proper extension of $\text{Möb}(S^2)$. Then, $G$ contains an element $f$ fixing at least 4 points, and such that the restriction of $f$ to the complement of these points is isotopic to a pseudo-Anosov map relative to these 4 points. In particular, $f$ has strictly positive topological entropy.

Finally, we also include an interesting result – albeit not new – steaming from the same sort of techniques developed in the proofs of Theorems A and B. It is stated in terms of the general definition given below, which plays a key role in this paper.

**Definition 1.1.** Given $k \in \mathbb{N}$, the action of a subgroup $G \subset \text{Homeo}(M)$ is said to be $k$-transitive if, for every pair of $k$-tuples $(p_1, \ldots, p_k)$ and $(q_1, \ldots, q_k)$, each of them composed by mutually distinct points, there exists some transformation $g \in G$ such that $q_i = g(p_i)$ for each $i \in \{1, \ldots, k\}$. When it holds, in addition, that for such a given $k$-tuple the only transformation fixing its every point is the identity map, the group is said to be sharply $k$-transitive.

It is a widely known fact of complex analysis that $\text{Möb}(S^2)$ is sharply 3-transitive. As it turns out, this is a defining property of such group: in their 2014 work, Kwakkel and the second author show that any sharply 3-transitive subgroup of $\text{Diff}^1(S^2)$ extending the rotations group $\text{Rot}(S^2)$ must be $\text{Möb}(S^2)$. For completeness, and given that the cited work has not been published, we also present herein a new simple – and in our opinion rather enjoyable – direct proof of this fact, formally stated below.

**Theorem C.** Let $G \subset \text{Diff}^1(S^2)$ be a homogeneous group of diffeomorphisms. That is, $G$ is an extension of the rotations group $\text{Rot}(S^2)$. Then, $G$ sharply 3-transitive implies $G = \text{Möb}(S^2)$.

### 1.1. Preliminaries.

Let us start by agreeing upon terminology. As aforementioned, $\text{Homeo}(M)$ denotes the set of orientation preserving homeomorphisms of a closed and oriented topological manifold $M$, on which the uniform metric

$$d_\infty(f, g) = \max_{p \in M} \{d(f(p), g(p))\}$$

is well defined. Above, $d$ is some distance function generating the manifold’s topology. If $d$ comes from a Riemannian structure on $M$, we may also consider the subset $\text{Diff}^1(M) \subset \text{Homeo}(M)$ of...
orientation preserving diffeomorphisms of class $C^1$ and its finer $C^1$ topology – or compact-open – which takes into account the local expressions of differentials as well.

Consider the unit circle, $M = S^1$. It can be thought of as the set of complex points at unit distance from the origin or as the compactification $\mathbb{R} \cup \{\infty\}$. In the latter case, identification is provided by stereographic projection from the point $(0,1)$. Each such description is linked to a canonical group acting on $S^1$.

In the first case, we have the rotations group, $\text{Rot}(S^1)$. This group is naturally identified with the circle itself, the rotation of angle $\alpha$ being given as $z \mapsto e^{2\pi i \alpha} z$. It is thus a compact Lie group. In the second case, we have the Möbius group, $\text{Möb}(S^1)$, obtained by transposing to the circle the action of $\text{PSL}(2; \mathbb{R})$ on the extended real line by linear fractional transformations.

The rotations group can be realized as a proper subgroup of the Möbius group. Interestingly enough, these two groups yield a full description of the closed and transitive subgroups of $\text{Homeo}(S^1)$, which is the content of the theorem by Giblin and Markovic.

**Theorem** (Giblin and Markovic, [2]). Let $G$ be a closed and transitive subgroup of $\text{Homeo}(S^2)$ containing a nontrivial path connected component. Then, it is conjugate to one, and only one, of the following:

(i) $\text{Rot}(S^1)$,
(ii) $\text{Möb}_k(S^1)$,
(iii) $\text{Homeo}_k(S^1)$.

Above, the subscript $k \in \mathbb{N}$ indicates cyclic covering of order $k$. Furthermore, $\text{Möb}(S^1)$ is maximal among the closed and transitive groups of $\text{Homeo}(S^1)$. This means that there is no such a subgroup properly containing $\text{Möb}(S^1)$ and properly contained in $\text{Homeo}(S^1)$.

Consider now the unit 2-sphere, $M = S^2$. It can be thought of either as the points of Euclidean 3-space $\mathbb{R}^3$ at unit distance from the origin or as the compactification $\mathbb{C} \cup \{\infty\}$. In the latter case, identification is provided by stereographic projection from the North Pole $(0,0,1)$. Points on the plane and their stereographic images will be denoted by the same letters and confounded without notice. Accordingly, any sphere mapping that fixes $\infty$ defines, by stereographic conjugation, a mapping of the plane with the same degree of regularity, also denoted by the same letter.

The group of **sphere rotations**, $\text{Rot}(S^2)$, consists of Euclidean isometries preserving $S^2$. It is realized by the orthogonal matrices $\text{SO}(3; \mathbb{R})$, and is thus also a compact Lie group. Each such transformation amounts to prescribing an axis and a rotation angle around that axis. From these two facts, it can be seen that $\text{Rot}(S^2)$ is closed and transitive.

Rotations are **minimal** among transitive groups in the following sense: any compact subgroup $G \subset \text{Homeo}(S^2)$ is conjugate to a closed subgroup of $\text{Rot}(S^2)$ – a result known to B. Kerékjártó since the 40s, and for which a proof according to contemporary standards is given by B. Kolev in [5]. If $G$ is also transitive, then it is conjugate to $\text{Rot}(S^2)$. The problem of classifying the groups contained between $\text{Rot}(S^2)$ and $\text{Homeo}(S^2)$ is thus called the kernel subgroup problem, and such groups are hereafter called homogeneous, following the terminology established in [4].

The **Möbius group** $\text{Möb}(S^2)$ is an example of a homogeneous group. It is defined similarly to its circle counterpart, by the action of $\text{PSL}(2; \mathbb{C})$ on the extended plane via linear fractional transformations. More precisely, one associates to (the class of) the matrix $A$ a mapping $M_A$ as follows:

$$\text{if } A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{PSL}(2; \mathbb{C}), \text{ then } M_A(z) = \frac{az + b}{cz + d} \text{ for every } z \in \mathbb{C} \cup \{\infty\}. $$

This procedure not only induces a group of orientation preserving homeomorphisms of the sphere, it characterizes the conformal diffeomorphisms of $S^2$ endowed with its canonical smooth structure. The Möbius group is thoroughly understood: each transformation has either one or two fixed points, and is conjugated to one out of four prototypical mappings – called parabolic, elliptical, hyperbolic or loxodromic – depending on the relation between the coefficients $a$ and $d$. In particular, rotations of the sphere are induced by the subgroup $\text{PSU}(2; \mathbb{C})$. 


These are classical results, some of them already known to Gauss, that can be found, for instance, in Chapter 3 of [6]. Kwakkel and the second author left open the question of whether the inclusion of $\text{Mob}(S^2)$ in $\text{Homeo}(S^2)$ is maximal. Our contribution to this problem was stated in Theorem A above. It should be stressed that, in the circle setting, 4-transitivity is sufficient to ensure $k$-transitivity for any $k \geq 4$ (Theorem 6.4 in [2]).

The constructions in this paper are mostly based upon isotopies, here understood as jointly continuous maps $f : I \times M \to M$, where $I$ is a (possibly unbounded) real interval, such that the functions $f_t = f(t, \cdot)$ are homeomorphisms of $M$. We can either prescribe $f$ or a family $(f_t)_{t \in I}$ for which $t \mapsto f_t$ is continuous with respect to $d_\infty$. When the interval in question is the standard unit interval, we denote it by $I = [0, 1]$.

It will often be the case that $f_t \in \text{Diff}^1(M)$ for every $t \in I$, but $t \mapsto f_t$ can only be assured to be continuous with respect to $d_\infty$. To avoid confusions, isotopies for which this association is actually continuous with respect to the $C^1$-topology will be explicitly referred to as $C^1$-isotopies.

**Definition 1.2.** Let $G \subset \text{Homeo}(M)$. An isotopy $(f_t)_{t \in I}$ such that $0 \in I$, $f_0 = \text{id}$ and $f_t \in G$ for every $t \in I$ will be referred to as an $\mathcal{IG}$-isotopy. By letting $\sup I = +\infty$ and $\inf I = -\infty$ whenever suitable, we then define:

1. The trajectory of a point $x \in M$ under $f$ as $\gamma_f(x) = \{f_t(x) : t \in I\}$. Although this is a priori just a set, it will often be thought of as a curve oriented according to its natural direction of travel.

2. Its $\omega$-limit as the following set of accumulation points:

   $$\omega_f = \{y \in M : \text{there exists some sequence } t_n \not\to \sup I \text{ such that } f_{t_n}(x) \to y\}.$$ 

   A similar notion of $\alpha$-limit is defined for sequences such that $t_n \not\to \inf I$.

We can now introduce the concept of arc transitivity, mentioned in Theorem A.

**Definition 1.3.** Given $k \in \mathbb{N}$, a subgroup $G \subset \text{Homeo}(M)$ is said to be arc $k$-transitive if, for every pair of $k$-tuples as in Definition 1.1, there exists an $\mathcal{IG}$-isotopy $(g_t)_{t \in I}$ such that $g_t(p_i) = q_i$ for each $i \in \{1, \ldots, k\}$.

The suggestive terminologies above are useful in the developments that follow. Although analogies with flows are limited, the following establishes some parallels.

**Definition 1.4.** For a fixed a subgroup $G \subset \text{Homeo}(M)$ and given $z, w \in M$, say that $z \sim_G w$ if there exists an $\mathcal{IG}$-isotopy $(f_t)_{t \in I}$ in $G$ such that $f_1(z) = w$.

This is an equivalence relation, under which the class of a point $z \in M$ is denoted by $A_z$, and referred to as the points accessible from $z$ (in $G$).

**Lemma 1.5.** Let $(f_t)_{t \in I}$ and $(h_t)_{t \in J}$ be $\mathcal{IG}$-isotopies, where $I$ and $J$ are intervals of any kind. Then, for any two $z, w \in M$ such that $\gamma_f(z) \cap \gamma_h(w) \neq \emptyset$ we have that $A_w = A_z$.

**Proof.** Say that $f_a(z) = h_b(w)$, for some $a \in I$, $b \in J$. Given $y \in A_w$, there exists an $\mathcal{IG}$-isotopy $(g_t)_{t \in I}$ such that $g_1(w) = y$. Then,

$$k_t = \begin{cases} 
   f_{3at} & \text{if } 0 \leq t \leq \frac{1}{3}, \\
   h_{(3t-2)b} \circ h_b^{-1} \circ f_a & \text{if } \frac{1}{3} \leq t \leq \frac{2}{3}, \\
   g_{3t-2} \circ h_b^{-1} \circ f_a & \text{if } \frac{2}{3} \leq t \leq 1,
\end{cases}$$

is an $\mathcal{IG}$-isotopy satisfying $k_1(z) = y$. This shows that $A_w \subset A_z$, and the converse inclusion follows by symmetry. \qed

We finish this introduction by fixing some reference points that play a key role in the arguments that follow:

- $\infty$ is the North Pole,
Lemma 2.2. Let \( G \) be a proper extension of the conformal group of sphere diffeomorphisms, and \( \rho > 0 \) is a linear mapping having at least one eigenvalue \( \lambda > 0 \), for which \( \rho \) is an unit eigenvector.

The meridian through 0, 1 and \( \infty \) – which is the stereographic image of the \( x \)-axis (or real axis) – is denoted by \( \Gamma \). This meridian defines on its left a western hemisphere \( \mathcal{H}^+ \), corresponding to the upper half-plane, and on its right an eastern hemisphere \( \mathcal{H}^- \), corresponding to the lower half-plane. For any group \( G \) we also adopt the following notations for stabilizers:

\[
G_1 = \text{Stab}_G(\infty), \quad G_2 = \text{Stab}_G(0, \infty) \quad \text{and} \quad G_3 = \text{Stab}_G(0, 1, \infty).
\]

1.2. Paper outline. Given a group \( G \) extending Möbius\( (S^2) \), we consider the subgroups \( G_k, 1 \leq k \leq 3 \), as above. Our objective is to conclude that \( G_3 \) is (one) transitive.

We begin by establishing an Extension Lemma at the end of Section 2, which states that \( G_2 \) must contain an isotopy between the identity and a map having a hyperbolic saddle point at the (fixed) South Pole. This is achieved by starting with a nonconformal map and continuously parametrizing rotations and homotheties, both of which are Möbius transformations, in such a way as to create eigendirections and modulate the corresponding eigenvalues.

From this starting point, we fix the privileged reference meridian \( \Gamma \) and promote two parallel processes. In Section 3, we prove that points outside of it admit full time isotopies in \( G_3 \) accumulating at the poles. In Section 4, we prove that there is a finite time isotopy in \( G_3 \) for which some point on this meridian starts at one side of it and ends at the other side.

In Section 5 we show our main theorems. Theorem A is derived upon combining isotopies of the types described and concluding that all but three points of the sphere are actually arc connected in the sense of Definition 1.3, which yields arc transitivity of \( G_3 \). Theorem B is also derived by convenient combinations of segments of such isotopies, but to produce a “topological figure 8”, a device that implies positive entropy due to the Nielsen-Thurston classification theory.

Lastly, in Section 6 we derive Theorem C. First, a purely topological argument shows that, if \( G \) is a sharply 3-transitive and homogeneous group of homeomorphisms, then \( G_2 \) must permute parallels. This fact, when combined with differentiability, yields conformality – first at the poles, and then at every point.

2. Extensions of Möbius

Definition 2.1. For a given \( 0 \leq \theta < 2\pi \), \( R_\theta \) denotes the counterclockwise rotation of angle \( \theta \)

- around the \( z \)-axis, if it is thought of as a mapping of the 2-sphere,
- around the origin, if it is thought of as a planar mapping. In this case, it may be written as \( R_\theta(z) = e^{2\pi i \theta} z \).

Also, for a given \( \rho > 0 \), we let \( H_\rho \in \text{Möb}_2(S^2) \) denote the transformation induced by the plane homothety \( z \mapsto \rho z \). Since the tangent space \( T_0 S^2 \) can be identified with the horizontal subspace \( \mathbb{R}^2 \times \{0\} \cong \mathbb{R}^2 \), generated by the canonical basis \( (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}) \), we may write, in a slight abuse of notation:

\[
DR_\theta(0) = R_\theta \quad \text{and} \quad DH_\rho(0) = \rho \text{id}.
\]

We remark that, if \( u, v \) are nonzero vectors, the angle \( 0 \leq \angle(z, w) \leq \pi \) between them is defined in the usual way, as \( \arccos \frac{\langle u, v \rangle}{||u|| \cdot ||v||} \).

Lemma 2.2. Let \( G \subset \text{Diff}^1(S^2) \) be a proper extension of Mőb\( (S^2) \). Then, there exists \( \hat{g} \in G_2 \) for which 0 is a hyperbolic saddle point. More precisely, \( D\hat{g}(0) = \text{diag} [\lambda, \lambda^{-1}] \) with respect to the canonical basis, where \( 0 < \lambda < 1 \).

Proof. Since \( G \) is a proper extension of the conformal group, it contains at least one mapping \( h \) which is nonconformal at some point of the sphere. By precomposing and postcomposing with suitable Möbius transformations, it can be assumed that such map lies in \( G_2 \) (that is, it fixes the poles) and that \( A = D\hat{h}(0) \) is nonconformal.

This means that there exists a pair of unit vectors \( u, v \) such that \( \angle(Au, Av) \neq \angle(u, v) \). Let \( R_{(1)} \) be a planar rotation such that \( R_{(1)}(Au) \) is a positive multiple of \( u \). Then, \( A_1 = R_{(1)}A \) is a linear mapping having at least one eigenvalue \( \lambda_1 > 0 \), for which \( u \) is an unit eigenvector.
It cannot be the case that $\lambda_1$ is of maximal geometric multiplicity, for that would imply $A$ conformal. This leaves two possibilities.

**Case 1.** $A_1$ is a defective matrix.

Then, we may fix an orthogonal chain of generalized eigenvectors $\{u, w\}$. That is, $A_1w = u + \lambda_1w$ and $\langle u, w \rangle = 0$. For each $\phi \in [0, \pi/2]$ we define

$$x_\phi = \cos \phi \, u + \sin \phi \, w/|w| \quad \text{and} \quad y_\phi = -\sin \phi \, u + \cos \phi \, w/|w|,$$

obtaining an orthonormal frame $\{x_\phi, y_\phi\}$ of $T_0S^2$. Consider now the continuous function

$$\xi(\phi) = \langle A_1x_\phi, A_1y_\phi \rangle / |A_1x_\phi| |A_1y_\phi|.$$

It satisfies $\xi(0) \xi(\pi/2) < 0$. We thus have $\xi(\phi_0) = 0$ for some $0 < \phi_0 < \pi/2$. This means that, if $\hat{x} = x_{\phi_0}$ and $\hat{y} = x_{\phi_0}$, then the orthonormal frame $\{A_1\hat{x}/|A_1\hat{x}|, A_1\hat{y}/|A_1\hat{y}|\}$ can be applied onto $\{\hat{x}, \hat{y}\}$ by a certain planar rotation $R_2$, because $A_1$ preserves orientation. Thus, $R_2(A_1\hat{x}) = v_1 \hat{x}$ and $R_2(A_1\hat{y}) = v_2 \hat{y}$, where $v_1, v_2 > 0$.

Once again, we must have $v_1 \neq v_2$, or $A$ would be conformal. Let us assume, without loss of generality, that $0 < v_1 < v_2$. By defining $A_2 = R_2(A_1)$, $\rho = 1/\sqrt{v_1v_2}$ and $\lambda = \sqrt{v_1/v_2} < 1$, we have $(\rho A_2)\hat{x} = \lambda \hat{x}$ and $(\rho A_2)\hat{y} = \lambda^{-1} \hat{y}$. Lastly, let $R$ be a planar rotation such that $R^{-1} \hat{x} = \partial/\partial x$ and $R^{-1} \hat{y} = \varepsilon \partial/\partial y$, where $\varepsilon = 1$ if $\{\hat{x}, \hat{y}\}$ is a positive basis, or $\varepsilon = -1$ otherwise. Defining $\hat{A} = \rho R^{-1} A_2 R:

$$\hat{A} \left( \frac{\partial}{\partial x} \right) = \lambda \frac{\partial}{\partial x} \quad \text{and} \quad \hat{A} \left( \frac{\partial}{\partial y} \right) = \varepsilon^2 \lambda^{-1} \frac{\partial}{\partial y} = \lambda^{-1} \frac{\partial}{\partial y}.$$

Therefore, $\hat{A}$ has the form described in the Lemma’s statement with respect to the canonical basis. But, if $\hat{g} = R^{-1} \circ H_\rho \circ R_2 \circ R_1 \circ \hat{h} \circ R \in G_2$, then $D\hat{g}(0) = \hat{A}$, completing the proof in this case.

**Case 2.** $A_1$ has a second eigenvalue $\lambda_2 \neq \lambda_1$

Fix a positive basis of unit eigenvectors $\{u, w\}$. Let $w^\perp$ be the orthogonal complement of $w$ with respect to $u$. For each $\phi \in [0, \pi/2]$, we define an orthonormal frame $\{x_\phi, y_\phi\}$ as

$$x_\phi = \cos \phi \, u + \sin \phi \, w^\perp/|w^\perp| \quad \text{and} \quad y_\phi = -\sin \phi \, u + \cos \phi \, w^\perp/|w^\perp|,$$

and consider the same continuous function $\xi$ as before. We now have $\xi(0) \xi(\pi/2) \leq 0$, so there exists $0 < \phi_0 < \pi/2$ such that $A_1\hat{x}$ and $A_2 \hat{y}$ are orthogonal, where $\hat{x} = x_{\phi_0}$ and $\hat{y} = y_{\phi_0}$. We may thus fix a planar rotation $R_2$ – possibly the identity – applying the orthonormal frame $\{A_1\hat{x}/|A_1\hat{x}|, A_1\hat{y}/|A_1\hat{y}|\}$ onto $\{\hat{x}, \hat{y}\}$. From here, the proof follows as in Case 1, by defining $v_1, \rho, \lambda$ and $R$ exactly as before.

**Lemma 2.3 (The Extension Lemma).** Let $G \subset \text{Diff}^1(S^2)$ be a proper extension of $\text{M"ob}(S^2)$. Then, there exists an $\mathcal{IG}_2 C^1$-isotopy $(g_t)_{t \in \mathbb{R}}$ such that:

1. for every $t > 0$, the differential $Dg_t(0)$ is a hyperbolic saddle, having the tangent line $T_0\Gamma$ as its stable direction,
2. $Dg_t(0) = \text{diag} [\lambda, \lambda^{-1}]$ with respect to the canonical basis of $T_0S^2$.

**Proof.** By Lemma 2.2, we may fix $\hat{g} \in G_2$ such that $\hat{A} = D\hat{g}(0) = \text{diag} [\mu, \mu^{-1}]$, where $0 < \mu < 1$. For $s \in [0, \pi/2]$, let $B_s = \hat{A}^{-1} R_s \hat{A}$ and $v_s = B_s \left( \frac{\partial}{\partial x} \right)$. Then, $|v_s| < 1$ for $0 < s \leq \pi/2$. Next, we define the continuous function $\theta : [0, \pi/2] \to \mathbb{R}$ to be the angle between $v_s$ and $\frac{\partial}{\partial x}$, and let $C_s = R_{-\theta(s)} B_s$. Then, $C_0 = \text{id}$ and, for $0 < s \leq \pi/2$, the $x$-axis is a contracting direction for $C_s$, of rate $\lambda_s = |v_s|$. But, since $\det C_s = 1$, it must also have an invariant direction of expansion rate $\lambda_s^{-1}$. In other words, $C_s$ is a hyperbolic saddle, although not necessarily diagonal. When $s = \pi/2$, however, it can be computed directly that $C_{\pi/2} = \text{diag} [\lambda, \lambda^{-1}]$, where $\lambda = \mu^2$. Lastly, we define $(g_t)_{t \in \mathbb{R}}$ as:

1. $g_t = R_{-\theta (\frac{\pi}{2})} \circ \hat{g}^{-1} \circ R_{\frac{\pi}{2}} \circ \hat{g} \in G_2$.

Since $\theta (\cdot)$ is a continuous function, (1) defines an $\mathcal{IG}_2$-isotopy which is continuous only with respect to the uniform topology, but also to the $C^1$ topology. For each $t$, we have $Dg_t(0) = C_{\pi t/2}$.
Thus, given that the x-axis corresponds to $\Gamma$ on the sphere – with the direction generated by $\frac{\partial}{\partial x}$ identified with $T_0\Gamma$ – the considerations previously made translate into the statements of the lemma, completing the proof. □

3. A Fundamental Lemma

For each angle $\alpha \in (0, \pi/2)$, the subset of the plane describe in polar coordinates as follows shall be referred to as the $\alpha$-cone:

$$C_\alpha = \left\{ re^{i\theta} : r \geq 0 \text{ and either } |\theta| \leq \alpha \text{ or } |\theta - \pi| \leq \alpha \right\}.$$ 

We remark that, for $0 < \lambda < 1$, any cone is "broadened" under the action of the hyperbolic matrix $A = \text{diag}[\lambda, \lambda^{-1}]$. More precisely, consider $S^1$ parametrized by the counterclockwise polar angle $\theta$ that each direction of space makes with the x-axis. Then, the action of $A$ induces a monotone circle dynamics $\hat{A}$ – by radial projection – such that $\hat{A}$ has repelling fixed points at 0 and $\pi$ and attracting fixed points at $\pi/2$ and $3\pi/2$.

Now, if we let $g$ be a planar diffeomorphism fixing the origin, then, for a given $0 < \varepsilon < \pi$, differentiability yields $\delta > 0$ such that $\text{ang}(g(z), Dg(0)z) \leq \varepsilon$ whenever $0 < |z| < \delta$. The next lemma is a consequence of these two remarks, along with the fact that the following diagram commutes:

$$v \in \mathbb{R}^2 \setminus \{0\} \xrightarrow{\text{polar angle}} Av \xrightarrow{\text{rotation}} R_{\pm\tau}(Av)$$

$$\theta \in S^1 \xrightarrow{\text{source-sink}} \hat{A}(\theta) \xrightarrow{\text{translation}} \hat{A}(\theta) \pm \tau$$

Lemma 3.1. Let $g$ be a planar diffeomorphism for which the origin is a hyperbolic fixed point satisfying $Dg(0) = \text{diag}[\lambda, \lambda^{-1}]$, $0 < \lambda < 1$. Then, for a given $0 < \alpha < \pi/2$, there exist $\tau > 0$ and $\delta > 0$ such that:

$$0 < |z| < \delta \text{ and } z \notin C_\alpha \implies R_{\tau}(g(z)) \notin C_\alpha, \text{ whenever } |\omega| < \tau.$$ 

In particular, $g^k(z) \notin C_\alpha$ for every $k \in \mathbb{N}$ such that the orbit $\{z, g(z), \ldots, g^{k-1}(z)\}$ remains in $D_\delta(0)$.

We keep this result aside for now and move on to understand how isotopies of the kind defined in the Extension Lemma act on cones. For that, we record the following consequence of the chord length formula from elementary geometry:

$$|z - w| \geq \min\{|z|, |w|\} \sin\left(\frac{\text{ang}(z, w)}{2}\right). \tag{2}$$

Lemma 3.2. Let $(g_t)_{t \in \mathbb{R}}$ be a planar $C^1$-isotopy such that the origin is a fixed point and $Dg_t(0)$ has the x-axis as an invariant direction for every $t$. Then, given $0 < \alpha < \pi/2$, there exist $0 < \beta < \alpha$ and $\rho > 0$ such that:

$$z \in D_\rho(0) \text{ and } z \notin C_\alpha \implies g_t(z) \notin C_\beta \text{ for every } t \in \mathbb{R}.$$ 

Proof. For a fixed $t \in \mathbb{R}$, let $A_t = Dg_t(0)$. Consider $v_\alpha$ and $v_\alpha^*$ unit vectors of angle $\alpha$ and $2\pi - \alpha$, respectively, whose spans delimit $C_\alpha$. Given that the x-axis is $A_t$-invariant, $A_tv_\alpha$ lies to the left and $A_tv_\alpha^*$ lies to the right of it. Also, at least one among $\text{ang}(A_tv_\alpha, \frac{\partial}{\partial x})$ and $\text{ang}(A_tv_\alpha^*, \frac{\partial}{\partial x})$ has to be smaller than $\pi/2$, because $A_t$ is an orientation preserving linear isomorphism. Thus, by taking

$$\beta_t = \min\left\{\alpha, \text{ang}\left(A_tv_\alpha, \frac{\partial}{\partial x}\right), \text{ang}\left(A_tv_\alpha^*, \frac{\partial}{\partial x}\right)\right\}$$

we have $0 < \beta_t \leq \alpha$ and $A_tv \notin C_{\beta_t}$ whenever $v \notin C_\alpha$. This reasoning yields a global solution to the associated linear problem. Once it is done, let $\varepsilon_t = r_t c_t \sin(\beta_t/4)$, where $r_t = (1 + \sin(\beta_t/4))^{-1}$ and $c_t > 0$ is such that $|A_tv| \geq c_t |v|$ for every vector $v$. Having these choices in mind, we obtain $\rho_t > 0$ such that:

$$|g_t(z) - A_tz| \leq \frac{\varepsilon_t}{2} |z| \text{ whenever } 0 < |z| < \rho_t.$$ 

Thus, when $s$ is an external point. In particular, $(4)$

$C_0 < t$

Given finite and nonzero points $z, w$, we let $M[z, w] \in \text{Möb}(S^2)$ be the unique Möbius transformation fixing the poles and mapping $z$ to $w$. Also, we denote $M[z] = M[z, 1]$.

We can write down explicit formulae for these transformations and see that $(z, w) \mapsto M[z, w]$ is continuous. Also, if $K \subset S^2$ is a nonempty compact set bounded away from $0$, the sets $M[x](K)$ converge to $\{\infty\}$ on the Hausdorff distance as $x \to 0$.

Having settled these notations and technical results, we are now able to prove our Fundamental Lemma.

**Lemma 3.5** (The Fundamental Lemma). Let $G \subset \text{Diff}^1(S^2)$ be a proper extension of $\text{Möb}(S^2)$. Then, for a given point $z_0$ not on the meridian $\Gamma$ there exists an $\mathcal{I}G_3$-isotopy $(I^t_0)_{t \geq 0}$, depending on $z_0$, such that:

1. the trajectory of $z_0$ under $I^{z_0}$ does not intersect $\Gamma$, and
(2) the ω-limit of \( z_0 \) satisfies \( \omega_{t \to 0} (z_0) = \{ \infty \} \).

**Proof.** Given \( z_0 \notin \Gamma \), we assume for concreteness that it lies on the upper half-plane, and is thus given in polar coordinates as \( z_0 = R_0 e^{i \theta_0} \), \( 0 < \theta_0 < \pi \).

Let \((g_t)_{t \in I}\) be as in the Extension Lemma. Since \( \infty \) is fixed throughout, it can be thought of as a planar \( TG_1 \)-isotopy such that \( Dg_t(0) \) has the x-axis as an invariant direction for every \( t \) and \( Dg_t(0) = \text{diag} \left[ \lambda, \lambda^{-1} \right] \), \( 0 < \lambda < 1 \). To ease notation, we write \( g := g_1 \) and \( A := Dg(0) \).

Fix some \( 0 < \alpha < \pi/2 \) such that the direction through \( \theta_0 \) is external to \( C_{2\alpha} \). With respect to \( \alpha \), let \( \delta > 0 \) and \( \tau > 0 \) be as described in Lemma 3.1. Regarding this same \( \alpha \), and also the isotopy \((g_t)_{t \in I}\), Corollary 3.3 yields a radius \( \rho > 0 \) and angles \( 0 < \beta^- < \beta^+ < \alpha \) such that, for \( z \in D_\rho (0) \) and every \( t \in I \), \( z \notin C_\alpha \) implies \( g_t(z) \notin C_{\beta^+} \), whilst \( z \in C_{\beta^-} \) implies \( g_t(z) \in C_{\beta^-} \). Lastly, we characterize the stable manifold \( W^s \) of \( g \) at \( 0 \). Since the stable direction of \( A \) is the x-axis, \( W^s \) may be assumed to be a Lipschitz graph of the form \( y = y(x) \) having horizontal tangent at the origin. So, for \( \tau \) and \( \beta^- \) as obtained above and a sufficiently small radius \( \sigma > 0 \), it can be assumed that \( W^s \cap D_\sigma (0) \subset C_{\min \{ \tau, \beta^- \}} \).

Let \( 0 < \rho_0 < \min \{ \delta, \rho, \sigma, 1 \} \). Then, in the disk \( D_0 := D_{\rho_0} (0) \) all of the conditions described above are mutually satisfied, as conveyed in Figure 1.

![Figure 1](image1)

**Figure 1.** Inside the disk \( D_0 \), the isotopy path under \( g_t \) of points \( z \) outside of the \( \alpha \)-cone never enters the \( \beta^- \)-cone, while the stable manifold of \( g \) is a Lipschitz graph fully contained within the \( \min \{ \tau, \beta^- \} \)-cone.

Once these choices are made, fix a positive real number \( 0 < r_1 < \min \{ \rho_0, \rho_0/R_0 \} \) and let \( \rho_1 = r_1R_0 \) and \( K_0 = \partial D_{\rho_1} (0) \setminus C_\alpha \). Thus, \( K_0 \) is a compact set – contained in \( D_0 \) and not intercepting the stable manifold – composed of two closed arcs of a circumference. In particular, there exists \( n_0 \in \mathbb{N} \) such that:

\[
\text{if } n_x = \min \{ n \in \mathbb{N} : g^n(x) \notin D_0 \}, \text{ then } n_x \leq n_0 \text{ for every } x \in K_0.
\]

Note that \( r_1 < \rho_0 \). So, by the graph characterization of \( W^s \), for some \( \tau_1 \) with \( |\tau_1| \leq \min \{ \tau, \beta^- \} \) we have that \( r_1 e^{i \tau_1} \in W^s \), as suggested by Figure 2.

![Figure 2](image2)

**Figure 2.** Since the stable manifold \( W^s \) is locally given as a Lipschitz graph \( y = y(x) \), it can be reached from the x-axis by an uniformly bounded rotation.
Claim 3.5.2. Define $M_1(z) = r_1 e^{i \tau_1} z$. Then, the mapping $M_1$ has the following properties:

(i) $M_1 \in \text{Möb}_2(S^2)$,
(ii) $v_1 := M_1(1) \in W^s$,
(iii) $w_1 := M_1(z_0) \in K_0$.

Proof of claim. Item (i) follows from the form of $M_1$ and (ii) from the choice of $\tau_1$. As for (iii), consider $w_1 = M_1(z_0) = e^{i \tau_1} r_1 z_0$. Since $r_1 z_0$ has the same polar angle $\theta_0$ as $z_0$, $\theta(w_1) = \theta_0 + \tau_1$.

But $|\tau_1| < \beta^{-} - \alpha$ and $2 \alpha < \theta_0 < \pi - 2 \alpha$, which together imply $\theta_0 + |\tau_1| < \pi - \alpha$ and $\theta_0 - |\tau_1| > \alpha$. It follows that $\alpha < \theta_0 + \tau_1 < \pi - \alpha$, allowing us to conclude that $w_1 \notin C_\alpha$. Furthermore, $|w_1| = \rho_1$, yielding $w_1 \in K_0$.

In particular, $n_1 := n_{w_1}$ as in (6) is well-defined.

Claim 3.5.3. Consider $f_t = g_{t-[t]} \circ g^{[t]} \circ M_1$, for $0 < t \leq n_1$. Then, $f$ is an isotopy satisfying $f_t(z_0) \notin C_{\beta^+}$ and $f_t(1) \in C_{\beta^+}$, for every $t \in (0, n_1]$.

Proof of claim. On the one hand, $f_t(z_0) = g_{t-[t]} \left( g^{[t]} \left( w_1 \right) \right)$. Since $w_1 \notin C_\alpha$, each $g^{[t]}(w_1)$ does not belong to $C_\alpha$ either, as observed in Lemma 3.1. However, it does belong to $D_0$ while $t < n_1$. Thus, $g_s \left( g^{[t]} \left( w_1 \right) \right) \notin C_{\beta^+}$ for every $s = t - [t] \in I$. On the other hand, $f_t(1) = g_{t-[t]} \left( g^{[t]} \left( v_1 \right) \right)$. Since $v_1 \in W^u$, each $g^{[t]}(v_1)$ belongs to $W^u \cap D_0 \subset C_{\beta^-} \cap D_0$. Therefore, $g_s \left( g^{[t]} \left( v_1 \right) \right) \in C_{\beta^+}$ for every $s = t - [t] \in [0, 1]$, as claimed.

This setting is pictorially represented in Figure 3, where the points $z_1 := f_{n_1}(z_0) = g^{n_1}(w_1) \notin C_\alpha$ and $u_1 := f_{n_1}(1) = g^{n_1}(v_1) \in W^s$ were introduced.

![Figure 3](image_url)

By the choice of $n_1$, $|z_1| > \rho_0$. Let $r_2 = \rho_1/|z_1|$. Since product by $r_2$ does not change angles, $r_2 z_1 \in K_0$ and $r_2 u_1 \in D_0 \cap C_{\min(\tau, \beta^-)}$. In particular, $e^{i \tau_2} \left( r_2 u_1 \right) \in W^s$ for some $\tau_2$ with $|\tau_2| \leq \min \{ \tau, \beta^- \}$.

Claim 3.5.4. Define $M_2(z) = r_2 e^{i \tau_2} z$. Then, the mapping $M_2$ has the following properties:

(i) $M_2 \in \text{Möb}_2(S^2)$,
(ii) $v_2 := M_2(u_1) \in W^s$,
(iii) $w_2 := M_2(z_1) \in K_0$.

Proof of claim. Item (i) follows from the form of $M_1$ and (ii) from the choice of $\tau_2$. As for (iii), recall that $r_2 z_1$ belongs to $D_0 \setminus C_\alpha$, which yields $\alpha < \theta(r_2 z_1) - \tau < \theta(r_2 z_1) + \tau < \pi - \alpha$. This implies, along with $|\tau_2| \leq \tau$, that $w_2 \notin C_\alpha$. Furthermore, $|w_2| = |r_2 z_1| = \rho_1$, establishing that $w_2 \in K_0$.

In particular, $n_2 := n_{w_2}$ is well-defined. Consider now the expression $f_t = g_{t-[t]} \circ g^{[t]-n_1} \circ M_2 \circ f_{n_1}$, for $n_1 < t \leq n_1 + n_2$. By arguments analogous to the ones developed previously, we have
that $t \mapsto f_t$ is continuous over the interval $(n_1, n_1 + n_2)$. Also, $f_t(z_0) \notin C_{\beta^+}$ and $f_t(1) \in C_{\beta^+}$, for every $t \in (n_1, n_1 + n_2)$. In a similar fashion, we define inductively, for $k \geq 0$:

$$f_t = \begin{cases} 
\text{id} & \text{if } t = 0, \\
g_{t-[|t|]} \circ g^{[|t|]-N_k} \circ M_{k+1} \circ f_{N_k} & \text{on the interval } N_k < t \leq N_{k+1}, 
\end{cases}$$

where $N_0 = 0$, $N_k = \sum_{i=1}^{k} n_i$ and the numbers $n_k$ and the mappings $M_k$ are determined as follows:

- $M_{k+1} \in \text{Möb}_2(S^2)$ is a transformation of the form

$$M_{k+1}(z) = r_{k+1} e^{i\gamma_{k+1}} z,$$

mapping $z_k := f_{N_k}(z_0) \notin \overline{D_0}$ to a point $w_{k+1} \in K_0$ and $u_k := f_{N_k}(1) \in W^+$ to a point $v_{k+1} \in W^+$, via an homothety of scaling factor $r_{k+1} = \rho_1/|z_k| < 1$ and a rotation of angle $|\gamma_{k+1}| \leq \min \{\tau, \beta^\ast\}$;

- $n_k := n_{w_k} \leq n_0$ is given as in (6).

The following properties hold, by construction:

(i) $f_t \in G_2$ for every $t \geq 0$,

(ii) $t \mapsto f_t$ is continuous over each interval of the form $(N_k, N_{k+1}]$,

(iii) $f_t(z_0) \notin C_{\beta^+}$ and $f_t(1) \in C_{\beta^+}$ for every $t \geq 0$.

Claim 3.5.4. For $t \geq 0$, let

$$I^t_{\beta^+} = \hat{M}[f_t(1)] \circ f_t,$$

where $\hat{M}[\cdot]$ is as in Definition 3.4. Then, $(I^t_{\beta^+})_{t \geq 0}$ is an $IG_\beta^\ast$-isotopy.

Proof of claim. Since $f_t \in G_2$ for every $t \geq 0$, it is clear that $I^t_{\beta^+} \in G_3$ for every $t \geq 0$. It is left to verify that $t \mapsto I^t_{\beta^+}$ defines a continuous curve of homeomorphisms. This mapping is, a priori, as continuous as $t \mapsto f_t$. Thus, all that is needed to check is continuity from the right at the left endpoints of each interval $(N_k, N_{k+1}]$. For $0 < h \leq 1$,

$$I^{h}_{N_k+h} = \hat{M}[f_{N_k+h}(1)] \circ f_{N_k+h} = \hat{M}[g_h(v_{k+1})] \circ g_h \circ M_{k+1} \circ f_{N_k}.$$

But notice that

$$\hat{M}[u_k] \circ M_{k+1}^{-1} \circ M[g_h(v_{k+1})]$$

is a Möbius transformation fixing the poles and mapping $g_h(v_{k+1})$ to 1. By sharp 3-transitivity, it must be the transformation $\hat{M}[g_h(v_{k+1})]$. But, since $g_h \to \text{id}$ as $h \to 0^+$, the above expression implies $\hat{M}[g_h(v_{k+1})] \to \hat{M}[u_k] \circ M_{k+1}^{-1}$ as $h \to 0^+$. Consequently,

$$I^{h}_{N_k+h} = \hat{M}[u_k] \circ M_{k+1}^{-1} \circ g_0 \circ M_{k+1} \circ f_{N_k} = \hat{M}[u_k] \circ f_{N_k} = \hat{M}[f_{N_k}(1)] \circ f_{N_k} = I^{t}_{\beta^+}.$$

Claim 3.5.5. $\gamma_{t=0}(z_0) \cap \Gamma = \emptyset$.

Proof of claim. Since $f_t(1) \in C_{\beta^+}$ for every $t \geq 0$, the transformation $\hat{M}[f_t(1)]$ may be explicitly written as $\hat{M}[f_t(1)](z) = |f_t(1)|^{-1} e^{i\psi} z$, where $|\psi| \leq \beta^\ast$. We also known that $f_t(z_0) \notin C_{\beta^+}$ for every $t \geq 0$. Thus, $f_t(1) = |f_t(z_0)| e^{i\theta}$, where $\beta^+ < \theta < \pi - \beta^\ast$. Consequently, $0 < \theta + \psi < \pi$. But $\theta(I^t_{\beta^+}(z_0)) = \theta + \psi$. Therefore, the trajectory $\gamma_{t=0}(z_0)$ remains on the upper half-plane without ever touching the x-axis. This establishes (1).

Claim 3.5.6. $f_t(1) \to 0$ as $t \to +\infty$.

Proof of claim. We may assume that $|g(z)| < |z|$ for every $z \in W^+$. Then, since $|u_k| = |g^{n_k}(v_k)| < |v_k|$ and $|v_{k+1}| = r_{k+1} |u_k|$, we have

$$|v_{k+1}| \leq (\rho_1/\rho_0)^{k+1}.$$

But, on each interval $(N_k, N_{k+1}]$, we have $f_t(1) = g_{t-[|t|]} \circ g^{[|t|]-N_k}(v_{k+1})$. As $t$ ranges through this interval, the quantity $t-[|t|]$ ranges over the interval $[0, 1]$ and the quantity $|t| - N_k$ ranges
through \( \{0,\ldots,n_{k+1}\} \subset \{0,\ldots,n_0\} \). Thus,
\[
(9) \quad \sup \{ |f_t(1)| : t > N_k \} \leq \max \left\{ |g^i(z)| : s \in I, 1 \leq i \leq n_0 \text{ and } z \in \overline{D}_{(\rho_1/\rho_0)^{k+1}}(0) \right\}.
\]

Let \( \varepsilon > 0 \) be given. Since each \( (g^i)' \) is an isotopy fixing the origin, we may find \( \eta > 0 \) such that \( |g^i(z)| < \varepsilon \) for every \( s \in I \) and \( 1 \leq i \leq n_0 \), whenever \( |z| < \eta \). Consequently, if \( k_0 \in \mathbb{N} \) is so large that \((\rho_1/\rho_0)^{k_0+1} < \eta, (9) \) implies \( |f_t(1)| < \varepsilon \) whenever \( t > N_{k_0} \).

We are, now, ready to finish the proof. We know that \( f_t(z_0) = g_{t-|t|} \circ g_{|t|}^{-N_k} (w_{k+1}) \) on each interval \((N_k, N_{k+1}]\), where \( w_k \in \mathcal{K}_0 \) for every \( k \in \mathbb{N} \). By observing the same ranges as in the proof of the previous claim, we see that, for every \( t \geq 0 \),
\[
f_t(z_0) \in \mathcal{K} := \left\{ g^i_s(z) : s \in I, 1 \leq i \leq n_0 \text{ and } z \in \mathcal{K}_0 \right\}.
\]

But \( \mathcal{K} \) is a compact set bounded away from \( 0 \). Since \( f_t(1) \to 0 \) as \( t \to +\infty \) and \( I^{z_0}(z_0) \in M[f_t(1)](\mathcal{K}) \), the remark in (5) implies (2).

**Corollary 3.6.** Let \( G \subset \text{Diff}^1(S^2) \) be a proper extension of \( \text{Möb}(S^2) \). Then, for each \( z_0 \) not on the meridian \( \Gamma \) and each pair of distinct points \( a, b \in \{0, 1, \infty\} \) there exists a full-time \( I\Gamma_3 \)-isotopy \((I^{z_0}_{ab}(t, \cdot))_{t \in \mathbb{R}} \) such that:

1. the trajectory of \( z_0 \) under \( I^{z_0}_{ab} \) does not intersect \( \Gamma \),
2. the \( \alpha \) and \( \omega \) limits of \( z_0 \) satisfy \( \alpha_{I^{z_0}_{ab}}(z_0) = \{a\} \) and \( \omega_{I^{z_0}_{ab}}(z_0) = \{b\} \).

**Proof.** Let \( z_0 \notin \Gamma \) and \( a, b \in \{0, 1, \infty\} \) be given. We denote by \( T_{ab} \) the unique idempotent Möbius transformation permuting \( a \) and \( b \) and fixing the remaining reference point. Each \( T_{ab} \) leaves \( \Gamma \) invariant, so neither \( T_{\infty b}(z_0) \) nor \( T_{\infty a}(z_0) \) lie on \( \Gamma \). The Fundamental Lemma then yields \( I\Gamma_3 \)-isotopies \((I^z_{\infty b}(z_0))_{t \geq 0} \) and \((I^z_{T_{\infty a}(z_0)})_{t \geq 0} \) as described therein, from which we define \( I^{z_0}_{ab} : \mathbb{R} \times S^2 \to S^2 \) as:
\[
(10) \quad I^{z_0}_{ab}(t, z) = \begin{cases} T_{\infty b} \circ I^t_{\infty b}(z_0) \circ T_{\infty b}(z) & \text{if } t \geq 0, \\ T_{\infty a} \circ I^t_{T_{\infty a}(z_0)} \circ T_{\infty a}(z) & \text{if } t \leq 0. \end{cases}
\]

Each mapping \( I^{z_0}_{ab}|_{[0, +\infty) \times S^2} \) and \( I^{z_0}_{ab}|_{(-\infty, 0) \times S^2} \) is itself an isotopy, and they agree in the common slice \( \{0\} \times S^2 \), both being equal to id there. Consequently, \( I^{z_0}_{ab} \) defines a global jointly continuous function of the variables \((t, z) \in \mathbb{R} \times S^2 \). The statements then follow from the Fundamental Lemma, noticing also that reversion of time turns the \( \omega \)-limit into the \( \alpha \)-limit.

**4. A Crossing Lemma**

Given \( r > 0 \), we let \( \gamma_r : I \to \mathbb{R} \) be the path \( \gamma_r(\theta) = re^{2\pi i \theta} \) and \( S_r \) be its image. Then, if \( g \in \text{Diff}^1(\mathbb{R}^2) \) is a planar diffeomorphism fixing the origin, we can check by a calculation that:
\[
(11) \quad g \circ \gamma_r \xrightarrow{C^1(I)} Dg(0) \circ \gamma_1 \quad \text{as} \quad r \to 0^+.
\]

Assume further that \( Dg(0) \) is a hyperbolic saddle matrix. Then, it maps \( S_1 \) onto an ellipse having semiminor axis of length strictly smaller than one and semimajor axis of length strictly greater than one. Thus, \( S_1 \cap Dg(0)(S_1) \) consists of four nontangential intersections. It then follows from (11), along with the stability of nontangential intersections, that for a sufficiently small \( r_0 > 0 \) (indeed, for infinitely many), \( S_{r_0} \cap g(S_{r_0}) \) contains exactly four points.

This remark allows us to now analyze the behaviour of \( \Gamma \) under the action of \( G \). Before doing so, we agree that given three distinct points \( a, b, c \), we denote by \( M[a, b, c] \) the unique Möbius transformation mapping \( a \) to \( 0 \), \( b \) to \( 1 \) and \( c \) to \( \infty \). If the points are all finite and nonzero, this transformation can be explicitly written as a certain cross ratio, from which it is seen that the association \( (a, b, c) \mapsto M[a, b, c] \) is continuous.

**Lemma 4.1.** Let \( G \subset \text{Diff}^1(S^2) \) be a proper extension of \( \text{Möb}(S^2) \). Then, there exists an \( I\Gamma_3 \)-isotopy \((k_t)_{t \in I}\) such that \( |k_1(\Gamma) \cap \Gamma| = 4 \).
Proof. By the Extension Lemma, it is known that there exists an $IG_2$-isotopy $(g_t)_{t \in \mathbb{R}}$ such that $g := g_t$ has a hyperbolic saddle fixed point at the origin. So, we may fix any sufficiently small $0 < r_0 < 1$ such that $|S_{t_0} \cap g(S_{t_0})| = 4$.

Let $a, b, c, d \in S_{t_0}$ be four consecutive points in the usual anticlockwise cyclic order, and such that $S_{t_0} \cap g(S_{t_0}) = \{g(a), g(b), g(c), g(d)\}$. Then, we may consider the unique $M_0 \in \text{M"{o}b}(S^2)$ mapping the ordered triple $(0, 1, \infty)$ onto $(a, b, c)$. Notice that $M_0 = M[a, b, c]^{-1}$ and that $M_0(\Gamma) = S_{t_0}$. By defining

$$k_t = \tilde{M}[g_t(a), g_t(b), g_t(c)] \circ g_t \circ M_0, \quad 0 \leq t \leq 1,$$

we obtain an $IG_3$-isotopy $(k_t)_{t \in \mathbb{R}}$. Since $M_0$ is a bijection of the sphere:

$$|k_1(\Gamma) \cap \Gamma| = \left| (M_0 \circ \tilde{M}[g(a), g(b), g(c)]) \circ g(M_0(\Gamma)) \cap M_0(\Gamma) \right| = |g(S_{t_0}) \cap S_{t_0}| = 4.$$

Finite points $a, b$ on the meridian $\Gamma$ are identified with their real counterparts on the $x$-axis. This induces a natural ordering, for which one may speak of the oriented interval with endpoints $a, b$. Whenever $a \leq b$ with respect to this ordering,

- $[a, b]$ denotes the arc of the meridian $\Gamma$ with endpoints $a, b$ and not containing $\infty$, which corresponds to the compact interval of the $x$-axis with the associated endpoints;

- $[b, a]$ denotes the arc of the meridian $\Gamma$ with endpoints $a, b$ and containing $\infty$, which projects onto $(-\infty, a] \cup [b, +\infty)$.

If $b = \infty$, then the corresponding arcs are defined via stereographic projection as $[\infty, a] = (-\infty, a] \cup \{\infty\}$ and $[a, \infty) = [a, +\infty) \cup \{\infty\}$. Lastly, open and half-open arcs of $\Gamma$ are defined accordingly by deletion of the suitable endpoints from the corresponding closed arcs.

Corollary 4.2. Let $G \subset \text{Diff}^1(S^2)$ be a proper extension of $\text{M"{o}b}(S^2)$. Then, given $z_0 \in \Gamma \setminus \{0, 1, \infty\}$, there exists an $IG_3$-isotopy $(h_t)_{t \in \mathbb{I}}$ such that $h_1(z_0) \notin \Gamma$.

Proof. In the language of Lemma 4.1, let $k_1(\Gamma) \cap \Gamma = \{0, 1, \infty, w_0\}$. If $k_1(z_0) \neq w_0$, it suffices to take $h_t = k_t$ for every $t$. Otherwise, assume for concreteness that $z_0 \notin (0, 1)$ – analogous reasonings apply to the other cases. By defining $h_t = \tilde{M}[1, \infty, 0] \circ k_t \circ \tilde{M}[1, \infty, 0]^{-1}$, we obtain a new $IG_3$-isotopy. Since $\tilde{M}[1, \infty, 0] \mid r$ is an interval exchange transformation free of fixed points, $\tilde{z}_0 := \tilde{M}[1, \infty, 0]^{-1}(z_0) \neq z_0$. Thus, $k_1(\tilde{z}_0) \notin \Gamma$. But $\tilde{M}[1, \infty, 0]$ leaves $\Gamma$ invariant, so $h_1(z_0) = \tilde{M}[1, \infty, 0](k_1(\tilde{z}_0)) \notin \Gamma$ as well.

Lemma 4.3 (The Crossing Lemma). Let $G \subset \text{Diff}^1(S^2)$ be a proper extension of $\text{M"{o}b}(S^2)$. Then, there exist a point $\hat{z}$ on the open segment $(0, 1)$ of $\Gamma$ and an $IG_3$-isotopy $(J_t)_{t \in [-1, 1]}$ such that:

1. The trajectory of $\hat{z}$ under $J$ only intercepts $\Gamma$ on the arc $(0, 1)$,
2. $J_{-1}(\hat{z}) \in \mathcal{H}^-$ and $J_1(\hat{z}) \in \mathcal{H}^+$.

Proof. By Corollary 4.2, we may fix an $IG_3$-isotopy $(h_t)_{t \in \mathbb{I}}$ and a point $z_0 \in \Gamma$ such that $h_1(z_0) \notin \Gamma$. Without loss of generality, assume $h_1(z_0) \in \mathcal{H}^+$. By continuity, there is some $w_0 \in \mathcal{H}^-$ near $z_0$ such that $h_1(w_0) \in \mathcal{H}^+$ as well. Consider the continuous path $\gamma : I \to S^2$ defined as $\gamma(t) = h_t(w_0)$. The sets $\gamma^{-1}(\mathcal{H}^-), \gamma^{-1}(\Gamma)$ and $\gamma^{-1}(\mathcal{H}^+)$ form a partition of $I$. Let

$$t^- = \sup \gamma^{-1}(\mathcal{H}^-) \quad \text{and} \quad t^+ = \inf \gamma^{-1}(\mathcal{H}^+) \cap [t^-, 1].$$

We have that $0 < t^- \leq t^+ < 1$, and that $[t^-, t^+] \subset \gamma^{-1}(\Gamma)$. Since $\gamma$ cannot intercept the points $(0, 1, \infty)$, it follows that $\gamma([t^-, t^+])$ is a compact arc contained within one of the three connected components of $\Gamma \setminus \{0, 1, \infty\}$. It is, thus, at a positive distance $\rho > 0$ from those three points. Let $\delta > 0$ be such that

1. $s \in I$ and $|s - t^-| < \delta$ imply $d(\gamma(s), \gamma(t^-)) < \rho$;
2. $s \in I$ and $|s - t^+| < \delta$ imply $d(\gamma(s), \gamma(t^+)) < \rho$.
Then, we may find points of the form \( t^- - \delta^- \in \gamma^{-1}(\mathcal{H}^-) \) and \( t^+ + \delta^+ \in \gamma^{-1}(\mathcal{H}^+) \) satisfying \( t^- - \delta^- < t^- - \delta^- < t^- < t^+ + \delta^+ < t^+ + \delta^- \) as conveyed in Figure 4 — and let \( \sigma : [-1, 1] \to [t^- - \delta^- , t^+ + \delta^+] \) be any increasing bijection such that \( \sigma(0) \) is the midpoint of the interval \([t^-, t^+]\).

Defining \( \tilde{J} : [-1, 1] \times S^2 \to S^2 \) as \( \tilde{J}(t, z) = \hat{J}_t(z) = h_{\sigma(t)} \circ h_{\sigma(0)}^{-1}(z) \), we readily see that it is an \( I\mathcal{G}_3 \)-isotopy. By letting \( \tilde{z} = \gamma(\sigma(0)) = h_{\sigma(0)}(u_0) \in \Gamma \), we have that \( \tilde{J}_1(\tilde{z}) = \gamma(t^+ + \delta^+) \in \mathcal{H}^+ \) and \( \tilde{J}_{-1}(\tilde{z}) = \gamma(t^- - \delta^-) \in \mathcal{H}^- \). Also, by observing the range of \( \sigma \), the trajectory of \( \tilde{z} \) under \( \hat{J} \) is \( \gamma_{\hat{J}}(\tilde{z}) = \gamma([t^- - \delta^- , t^+ + \delta^+]) \). In particular, it follows from the choice of \( \rho \) and from (12) that any point in \( \gamma_{\hat{J}}(\tilde{z}) \cap \Gamma \) lies on the same connected component of \( \Gamma \setminus \{0, 1, \infty\} \) as the segment \( \gamma([t^-, t^+]) \), which is precisely the component containing \( \tilde{z} \).

If this component happens to be \( (0,1) \), as pictured in Figure 4, we simply let \( J = \hat{J} \) and \( \tilde{z} = \hat{z} \). Otherwise, consider \( \Gamma \) endowed with the cyclic order induced by the real line. If \( \hat{z} \in (a, b) \), let \( \{c\} = \{0, 1, \infty\} \setminus \{a, b\} \). Since \( M[a, b, c] \) applies \( (a, b) \) onto \( (0, 1) \) in an orientation-preserving way, one may take \( \hat{z} = M[a, b, c](\hat{z}) \) and \( J = M[a, b, c] \circ \tilde{J} \circ M[a, b, c]^{-1} \) to obtain the desired point and isotopy.

5. Conclusions

5.1. Proof of Theorem A. From now on, a proper extension \( G \subseteq Diff^1(S^2) \) of \( \text{Mob}(S^2) \) is fixed throughout, and we let the point \( \hat{z} \in \Gamma \) and the \( I\mathcal{G}_3 \)-isotopy \( (J_t)_{t \in [-1, 1]} \) be as in the Crossing Lemma. Upon denoting \( \hat{z}_- = J_{-1}(\hat{z}) \) and \( \hat{z}_+ = J_1(\hat{z}) \), we define the set

\[
\chi = \gamma_{I^\hat{z}_-} \circ \gamma_{I^\hat{z}_+} \cup \gamma_{I^\hat{z}_-} \circ \gamma_{I^\hat{z}_+} \cup \gamma_{I^{\hat{z}_+}} \circ \gamma_{I^{\hat{z}_-}}.
\]

Above, \( I^{\hat{z}_-} \) and \( I^{\hat{z}_+} \) are the isotopies yielded by the Fundamental Lemma when considering the points \( \hat{z}_- \in \mathcal{H}^- \) and \( \hat{z}_+ \in \mathcal{H}^+ \). On the sphere, \( \chi \) is a continuum. Indeed, \( \gamma_{I^\hat{z}_-} \circ \gamma_{I^{\hat{z}_-}} \cup \gamma_{I^\hat{z}_+} \circ \gamma_{I^{\hat{z}_-}} \) is connected, as it is the union of (connected) curves with points in common, while its closure is automatically compact on the compact space \( S^2 \), and consists of adjoining \( \{\infty\} \) to this union, as we suggest in Figure 5.

Claim 5.1.1. The set \( \chi \) is a continuum separating \( (\infty, 0) \) and \( (1, \infty) \) in the following sense: whenever \( \alpha : \mathbb{I} \to S^2 \) is a path such that \( \alpha(0) \in (\infty, 0) \) and \( \alpha(1) \in (1, \infty) \), we have that \( \alpha(\mathbb{I}) \cap \chi \neq \emptyset \).

Proof. If \( \infty \in \alpha(\mathbb{I}) \) there is nothing to prove. Otherwise, suppose that \( \alpha \) never passes through \( \infty \). By a standard Zorn Lemma argument, it can also be assumed simple. We let

\[
t^- = \sup \alpha^{-1}(\infty, 0) \quad \text{and} \quad t^+ = \inf \left( \alpha^{-1}[1, \infty) \right) \cap [t^-, 1].
\]

Then, \( 0 \leq t^- < t^+ \leq 1 \), and any intersection between \( \alpha(t) \) and \( \Gamma \) for \( t^- < t < t^+ \) happens on the open segment \((0, 1)\).
Figure 5. The continuum $\chi$ is constructed by gluing together isotopy trajectories with points in common, one of which is bounded away from $\infty$ and two of which are known to accumulate at $\infty$, and then taking their closure.

On the plane, we define a continuous mapping $\ell : \mathbb{R} \to \mathbb{R}^2$ as pictured in Figure 6, and explicitly given by:

$$\ell(t) = \begin{cases} 
\frac{1-t}{1-t^-} (\alpha(t^-) - \hat{z}) + \hat{z} & \text{if } t \leq t^-, \\
\alpha(t) & \text{if } t^- < t < t^+, \\
\frac{1+t}{1+t^+} (\alpha(t^+) - \hat{z}) + \hat{z} & \text{if } t \geq t^+.
\end{cases}$$

The mapping $\ell$ is a line or, in other words, a simple and proper path. The orientation it inherits from the real line automatically divides the plane into two open and connected components, the right $R(\ell)$ and the left $L(\ell)$ of $\ell$, plus their common boundary $\ell$.

![Figure 6](image)

Figure 6. The segment of $\alpha$ comprehended between the instant it leaves $(\infty, 0]$ and the instant it enters $[1, \infty)$ can be glued to $\Gamma$ – traversed the usual way – to generate a line $\ell$.

We now consider the compact set $\alpha([t^-, t^+])$, and fix someclosed disk $D \subset \mathbb{R}^2$ fully containing it. Then, $[0, 1] \subset D$. Thus, if we now consider the open set $\mathcal{O} = \mathbb{R}^2 \setminus D$, we see from the expression of $\ell$ that $\ell \cap \mathcal{O} = \Gamma \cap \mathcal{O}$. Also, $\ell$ traverses this intersection with the same orientation as $\Gamma$. It follows that $L(\ell) \cap \mathcal{O} = L(\Gamma) \cap \mathcal{O} = \mathcal{H}^+ \cap \mathcal{O}$ and $R(\ell) \cap \mathcal{O} = R(\Gamma) \cap \mathcal{O} = \mathcal{H}^- \cap \mathcal{O}$.

From the Fundamental Lemma, $\gamma_{I^+} (\hat{z}_+) \cap \mathcal{H}^+$ and accumulates at $\{\infty\}$. Since $\mathcal{O}$ is a neighbourhood of $\infty$ on the sphere, it follows that $\gamma_{I^+} (\hat{z}_+) \cap \mathcal{H}^+ = \emptyset$. Analogously, $\gamma_{I^-} (\hat{z}_-) \cap \mathcal{H}^- = \emptyset$. This translates to $\chi \cap \mathcal{H}^+ = \emptyset$ and $\chi \cap \mathcal{H}^- = \emptyset$. Therefore, the line $\ell$ intercepts the continuum $\chi$.

Let $\tilde{t} \in \mathbb{R}$ be such that $\ell(\tilde{t}) \in \chi$. Then, it must be the case that $\tilde{t} \in (t^-, t^+)$. Indeed, on the one hand, $\ell(\tilde{t}) \in [1, \infty)$ if $\tilde{t} \geq t^+$ and $\ell(\tilde{t}) \in (-\infty, 0]$ if $\tilde{t} \leq t^-$. On the other hand, any intersection between $\chi$ and $\Gamma$ must take place on the open segment $(0, 1)$, by the Crossing Lemma and the Fundamental Lemma. However, $t^- < \tilde{t} < t^+$ means that $\ell(\tilde{t}) = \alpha(\tilde{t})$, yielding an intersection between $\alpha(I)$ and $\chi$, as claimed.

Next, consider the equivalence relation $\sim_{G_3}$, as described in Definition 1.4. Clearly, $\mathcal{A}_0 = \{0\}$, $\mathcal{A}_1 = \{1\}$ and $\mathcal{A}_\infty = \{\infty\}$. Our goal is to show that $z_0 \in \mathcal{A}_z$ for any $z_0 \in S^2 \setminus \{0, 1, \infty\}$. By Corollary 4.2, it suffices to consider $z_0 \notin \Gamma$.

Claim 5.1.2. If $z_0 \notin \Gamma$, then $z_0 \in \mathcal{A}_z$. 


Proof. Let \( \chi \) be as in (13). We fix \( r > 0 \) such that \( \overline{D_r}(\mathbf{0}) \cap \chi = \overline{D_r}(\mathbf{1}) \cap \chi = \overline{D_r}(\mathbf{0}) \cap \overline{D_r}(\mathbf{1}) = \emptyset \), and also such that both closed disks are disjoint from \( \{ z_0 \} \) and \( \{ \infty \} \). Consider the \( \mathcal{IG}_3 \)-isotopy \( I_{01}^{\infty} \) yielded by Corollary 3.6. We encounter \( S < 0 \) maximal such that \( I_{01}^S(S, z_0) \in \partial \overline{D_r}(\mathbf{0}) \) and \( T > 0 \) minimal such that \( I_{01}^T(T, z_0) \in \partial \overline{D_r}(\mathbf{1}) \). Let \( \alpha : \mathbb{I} \to \mathbb{R}^2 \) be given by:

\[
\alpha(t) = \begin{cases} 
4t I_{01}^S(S, z_0) - (1 - 4t) \frac{r}{2} & \text{if } 0 \leq t \leq \frac{1}{4}, \\
I_{01}^S \left( 2t(T - S) + \frac{3S - T}{2}, z_0 \right) & \text{if } \frac{1}{4} \leq t \leq \frac{3}{4}, \\
(4 - 4t) I_{01}^T(T, z_0) + (4t - 3) \left( 1 + \frac{r}{2} \right) & \text{if } \frac{3}{4} \leq t \leq 1.
\end{cases}
\]

Geometrically, \( \alpha \) departs from a point \( \alpha(0) \in (\infty, \mathbf{0}) \cap \overline{D_r}(\mathbf{0}) \), and follows on a straight line until it reaches a certain point of \( z_0 \)'s trajectory in the disk's boundary. This intersection point is such that \( z_0 \)'s trajectory returns to this first disk at most finitely many times in the future. From it, the path \( \alpha \) follows \( z_0 \)'s path until it first reaches the boundary of the disk \( \overline{D_r}(\mathbf{1}) \). Then, \( \alpha \) moves on a straight line until it reaches a point \( \alpha(1) \in (1, \infty) \cap \overline{D_r}(\mathbf{1}) \). This process is conveyed in Figure 7.

![Figure 7](image-url)

**Figure 7.** After leaving a compact neighbourhood of \( \mathbf{0} \) disjoint of \( \chi \) and before entering a neighbourhood of \( \mathbf{1} \) disjoint from \( \chi \), the path of \( z_0 \) under \( I_{01}^{\infty} \) must cross the continuum \( \chi \).

In particular, \( \alpha(0) \in (\infty, \mathbf{0}) \), \( \alpha(1) \in (1, \infty) \) and \( \alpha(\mathbb{I}) \subset S^2 \setminus \{ \infty \} \). By Claim 5.1.1, we have that \( \alpha \) intercepts \( \chi \setminus \{ \infty \} \). But, since the segments \( \alpha([0, 1/4]) \) and \( \alpha([3/4, 1]) \) are contained within disks disjoint from \( \chi \), we must have \( \alpha((1/4, 3/4)) \cap (\chi \setminus \{ \infty \}) \neq \emptyset \). However, it is seen from (14) that \( \alpha((1/4, 3/4)) \subset \gamma_{I_{01}^0}(z_0) \). By Lemma 1.5, this is readily seen to imply \( z_0 \in \mathcal{A}_z \). \( \square \)

This is enough to derive the (arc) 4-transitivity of \( G_0 \) for, if \( (a, b, c, d) \) and \( (p, q, r, s) \) are two given lists of distinct points on the sphere, let \( z_0 = M[a, b, c] (d) \) and \( w_0 = M[p, q, r]^{-1} (s) \). Then, neither \( z_0 \) nor \( w_0 \) belong to \( \{ \mathbf{0}, \mathbf{1}, \infty \} \) and thus, by Claim 5.1.2, both \( z_0 \) and \( w_0 \) belong to \( \mathcal{A}_z \). This implies that there is some \( \mathcal{IG}_3 \)-isotopy \( (f_i)_{i \in \mathbb{I}} \) such that \( f_i(z_0) = w_0 \). Since \( \text{Möb}(S^2) \) is a path connected group, in particular \( M[p, q, r]^{-1} \circ f_i \circ M[a, b, c] \) lies in \( G_0 \) and maps \( (a, b, c, d) \) onto \( (p, q, r, s) \). Therefore, the arc 4-transitivity definition is seen to be satisfied.

5.2. Discussion on Theorem B. Let us start by making our terminology precise: relative to four distinguished points \( P = \{ p_0, \ldots, p_3 \} \) on the sphere, a closed loop \( \alpha : \mathbb{I} \to S^2 \) based at \( p_0 \) will be referred to as topological figure 8 if it defines in the fundamental group \( \pi_1(S^2 \setminus \{ p_1, p_2, p_3 \}; p_0) \) an element in the same homotopy class of a path of the form \( \tilde{\zeta} \ast \tilde{\zeta} \). Here, each \( \zeta \) is a Jordan curve separating \( p_i \) from the remaining two points, while leaving \( p_i \) in the connected component of its complement locally at its left, and \( \ast, \tilde{\ast} \) denote the usual operations of concatenation and inversion. When \( p_3 \) is placed at infinity, we may think of the prototypical planar figure 8, consisting of the wedge of two circles based at \( p_0 \), each of them traversed once with contrary orientations, and leaving \( p_1 \) and \( p_2 \) in opposite components of their complements, as pictured in Figure 8. That is the situation we shall be aiming at.
As before, a proper extension $G \subset \text{Diff}^1(S^2)$ of $\text{M"{o}b}(S^2)$ is fixed throughout. Thus, if $\chi$ is defined as in (13), we may construct an $IG_3$ isotopy $(K_t)_{t \in \mathbb{R}}$ such that $\chi \setminus \{\infty\}$ is realized as the trajectory of $\hat{z}$ under $K$. It is given explicitly by

$$K_t = \begin{cases} I_{-1-t} \circ J_{-1} & \text{if } t \leq -1, \\ J_t & \text{if } -1 \leq t \leq 1, \\ I_{1+t} \circ J_1 & \text{if } t \geq 1, \end{cases}$$

where the isotopy $J$ and the family of points $\hat{z}$ are described in details in the Crossing Lemma, while the isotopies $I$ are as in the Fundamental Lemma. In particular, based on the descriptions given by such lemmas, we obtain a new $IG_3$ isotopy $L_t = T_{0\infty} \circ K_t \circ T_{0\infty}$ and a point $\hat{y} = T_{0\infty}(\hat{z}) \in (1,\infty)$ such that:

(i) $\omega_L(\hat{y}) = \alpha_L(\hat{y}) = \{0\}$,
(ii) $\{L_t(\hat{y}) : t \geq 1\} \subset H^-$ and $\{L_t(\hat{y}) : t \leq -1\} \subset H^+$,
(iii) the trajectory $\gamma_L(\hat{y})$ only intersects the meridian $\Gamma$ on the open arc $(1,\infty)$.

Claim 5.2.1. There exist a point $\hat{w} \in (0,1)$ and an $IG_3$ isotopy $(\varphi_t)_{t \in \mathbb{R}}$ such that $\gamma_{\varphi}(\hat{w}) \cong \xi_1 \cap \{0,1,\infty\}$, where $\cong$ denotes fixed endpoints homotopy, and $\xi_1$ is a circle passing through $\hat{w}$ traversed once clockwise while leaving $1$ on its right and both $0, \infty$ on its left.

Proof. Let $(L_t)_{t \in \mathbb{R}}$ and $\hat{y}$ be as in the previous discussion, and consider the path $\lambda(t) = L_t(\hat{y})$ describing the trajectory of $\hat{y}$ under $L$. Since $\lambda(0) = \hat{y} \in \Gamma$ and, by item (ii) above, $\lambda(-1) \in H^+$, we can fix $t = \min \lambda^{-1}(\Gamma) > -1$ and look at the restricted curve $\tilde{\lambda} = \lambda|_{(-\infty,\tilde{t})}$.

On the one hand, item (i) above implies $\lambda(t_N)$ in the same connected component of $\chi^c$ as $0$, for sufficiently negative $t_N < 0$. Thus, it can be joined to $(0,1)$ by a path not intersecting $\chi$. On the other hand, $\tilde{\lambda}(\tilde{t}) \in (1,\infty)$, by item (iii). It follows from Claim 5.1.1 that $\tilde{\lambda}$ must intersect $\chi \setminus \{\infty\}$ at least once. Therefore, we define $t^- = \max \lambda^{-1}(\chi) \cap (-\infty,\tilde{t}) < \tilde{t}$. Then, $\lambda(t^-) = L_{t^-}(\hat{y}) \in \chi \cap H^+$. Proceeding analogously for times greater than $\max \lambda^{-1}(\Gamma)$, but using the parts of items (i) and (ii) concerning positive times, we obtain $t^+ > 0$ such that $\lambda(t^+) = L_{t^+}(\hat{y}) \in \chi \cap H^-$. Also, $\tilde{\lambda}(t) = L_t(\hat{y})$ does not intersect $\chi$ for $t^- < t < t^+$.

Now, let $(K_t)_{t \in \mathbb{R}}$ be the isotopy under which the trajectory of $\hat{z}$ is $\chi \setminus \{\infty\}$. This means that $\lambda(t^-) = L_{t^-}(\hat{y}) = K_a(\hat{z})$ and that $\lambda(t^+) = L_{t^+}(\hat{y}) = K_b(\hat{z})$, for some $a, b \in \mathbb{R}$. In particular, $K_a(\hat{z})$ and $K_b(\hat{z})$ lie in opposite hemispheres, so there must be an intermediate parameter $c$ for which $K_c(\hat{z}) \in \Gamma$. Upon defining $\tilde{w} = K_c(\hat{z})$, we actually know from the Crossing Lemma that $\tilde{w} \in (0,1)$. We finally define:

$$\varphi_t = \begin{cases} K_{c+3t(b-c)} \circ K_c^{-1} & \text{if } 0 \leq t \leq 1 - \frac{1}{3}, \\ L_{2t-t^++3t(t^+-t^-)} \circ L_t^{-1} \circ K_b \circ K_c^{-1} & \text{if } 1 - \frac{2}{3} \leq t \leq 2 - \frac{2}{3}, \\ K_{3a-2c+3t(a-c)} \circ K_a^{-1} \circ L_t \circ L_t^{-1} \circ K_b \circ K_c^{-1} & \text{if } \frac{1}{3} \leq t \leq 1. \end{cases}$$

The family $(\varphi_t)_{t \in \mathbb{R}}$ is readily seen to form an $IG_3$ isotopy. Also, $\varphi_1(\hat{w}) = \hat{w}$, so the path $\gamma(s) = \varphi_s(\hat{w})$ describing the trajectory of the point $\hat{w}$ is indeed a closed loop based at $\hat{w}$. Along $\gamma$, we have four special points, distinguished during the previous constructions:

- the starting and terminal point $\hat{w} = \gamma(0) = \gamma(1) \in (0,1)$,
We also know that \( \gamma_{[0,1/3]} \) and \( \gamma_{[2/3,1]} \) only intercept \( \Gamma \) at \( (0,1) \), and that \( \gamma_{[1/3,3]} \) and \( \gamma_{[2/3,2]} \) only intercept \( \Gamma \) at \( (1,\infty) \). These facts are enough to show, by combining at most two consecutive straight line homotopies, that \( \gamma \) is homotopic on the plane – with fixed basepoint \( \hat{w} \) and relative to \( \{0,1\} \) – to a simple closed curved (say, polygonal) turning once clockwise around \( 1 \) and leaving \( 0 \) outside of it, as suggested by Figure 9. Moving back to the sphere by adjoining \( \infty \), this amounts to the claimed statement. 

\[ \begin{figure} 
\centering 
\includegraphics{figure9.png} 
\caption{The geometrical information gathered about the paths \( t \mapsto K_t(\hat{z}) \) and \( t \mapsto L_t(\hat{y}) \) is enough to prove that the trajectory of the point \( \hat{w} \) under \( \varphi \) is homotopic to a closed loop turning around 1, but not 0, with fixed endpoint \( \hat{w} \).}
\end{figure} 
\]

We can now complete the construction, for if \( \hat{w}' = T_{01}(\hat{w}) \), Theorem A yields \( h \) in the identity component of \( G_3 \) such that \( \hat{w}' = f(\hat{w}) \). We then consider the following \( IG_3 \) isotopy:

\[ \psi_t = h^{-1} \circ T_{01} \circ \varphi_t \circ T_{01} \circ h. \]

The trajectory of \( \hat{w} \) under \( \psi \) is described by the curve \( \eta(s) = h^{-1} (T_{01} \circ \gamma(s)) \), where \( \gamma \) describes \( \gamma_{\varphi}(\hat{w}) \), as in the proof of Claim 5.2.1. Since \( T_{01} \) leaves \( \{0,1,\infty\} \) invariant, this implies \( T_{01} \circ \gamma \cong T_{01} \circ \xi_1 \) rel \( \{0,1,\infty\} \). But we know that \( T_{01} \circ \xi_1 \) is also a circle – traversed once clockwise, while leaving \( 0 \) on its right and both \( 1,\infty \) on its left. Thus, since \( h \in G_3 \), \( \xi_2 = h^{-1} \circ T_{01} \circ \xi_1 \) is a clockwise oriented Jordan curve based at \( \hat{w} \), leaving \( 0 \) on its right and both \( 1,\infty \) on its left. Also, \( \eta \cong \xi_2 \) rel \( \{0,1,\infty\} \). If we now define the \( IG_3 \) isotopy

\[ F_t = \begin{cases} 
\varphi_{1-2t} \circ \varphi_1^{-1} & \text{if } 0 \leq t \leq \frac{1}{2}, \\
\psi_{2t-1} \circ \varphi_1^{-1} & \text{if } \frac{1}{2} \leq t \leq 1,
\end{cases} \]

then, clearly

\[ \gamma_F(\hat{w}) = \gamma_{\varphi}(\hat{w}) * \gamma_{\varphi}(\hat{w}) = \eta * \gamma = \xi_2 * \xi_1 \text{ rel } \{0,1,\infty\}, \]

meaning that (upon considering \( \zeta = \xi_1 \)) the diffeomorphism \( f := F_1 \) fixes \( P = \{\hat{w},0,1,\infty\} \), whilst the trajectory \( \gamma_F(\hat{w}) \) of \( \hat{w} \) under the isotopy \( (F_t)_{t \in \mathbb{E}} \) is a topological figure 8 relative to this set.

With respect to \( f \), we now evoke the Nielsen-Thurston Classification Theorem, as presented in Section 7.5 of [8]. It takes as input the homeomorphism \( f : S^2 \to S^2 \) and the set \( P \), for which the four punctured sphere \( S^2 \setminus P \) has negative Euler characteristic. Then, it yields

- a homeomorphism \( \Phi : S^2 \to S^2 \) such that \( f \) and \( \Phi \) are isotopic relative to \( P \) or, in other words, such that every point of \( P \) is fixed throughout the isotopy between \( f \) and \( \Phi \),
- a (possibly empty) system of closed simple loops \( \alpha_1, \ldots, \alpha_r \), called reducing curves.

Such curves come equipped with pairwise disjoint tubular neighbourhoods \( V_i \), disjoint from \( P \), such that the connected components of \( S^2 \setminus \bigcup_{i=1}^r V_i \) group into invariant cycles, restricted to which either \( \Phi \) is of finite order – meaning some power of it equals the identity – or \( \Phi \) is pseudo-Anosov relative to \( P \).
Briefly, pseudo-Anosov means that $\Phi$ (or its restriction to the appropriate component) admits a pair of invariant transverse foliations – one of which is expanded with ratio $\beta > 1$ and the other of which is contracted with ratio $\beta^{-1}$ (in a measure-theoretical precise sense). Relative to $P$ means that the points in $P$ are kept fixed under $\Phi$, and manifest as one-prong singularities of the foliations.

If the reducing system of curves is empty, we notice that no power $f^m$ can be isotopic to the identity relative to $P$. Indeed, let $\alpha$ be any simple closed loop separating two pairs of points of $P$. Then, $f^m(\alpha)$ would have to be (freely) homotopic to $\alpha$ relative to $P$. But, considering the $\mathbb{Z}\mathbb{O}_3$ isotopy $(F_t^m)_{t \in \mathbb{I}}$, the family $F_t^m(\alpha)$ would have to enclose the figure 8 in a fashion incompatible with such fact.

Upon interpreting points in $P$ as punctures (as implied by [9] and Section 7.6 of [8]), each connected component of $\mathbb{S}^2 \setminus \bigcup_{i=1}^r V_i$ must also have negative Euler characteristic. Thus, if the reducing system of curves is nonempty, $\hat{w}$ must be enclosed by $\alpha_i$ along with some other reference point $a \in \{0, 1, \infty\}$. Then, the argument from the previous paragraph also shows that $\Phi$ cannot be of finite order in any component.

In short, $\Phi$ must be a pseudo-Anosov map, which is classically known to have strictly positive topological entropy $h_{\text{top}}(\Phi) = \log \beta > 0$. But the behaviour of mappings isotopic to pseudo-Anosov homeomorphisms is described in [10]. More specifically, Theorem 2 therein implies the statement of Theorem B, and also $h_{\text{top}}(f) \geq h_{\text{top}}(\Phi)$. Thus, $f \in G_0$ must have positive topological entropy as well.

6. Characterization of the Conformal Group in Terms of Transitivity

In this section, we show that sharp 3-transitivity is a defining property of $\text{M"{o}b}(\mathbb{S}^2)$ among the homogeneous groups of diffeomorphisms, as more precisely stated in Theorem C. Before proving it, we establish two auxiliary lemmas. The first shows that the subgroup $G_2$ fixing the poles possesses a property that we already know a priori to be held by the actual $\text{M"{o}b}(\mathbb{S}^2)$. The second establishes conformality at the poles in this subgroup.

Lemma 6.1. Let $G \subset \text{Homeo}(\mathbb{S}^2)$ be a sharply 3-transitive homogeneous group. Then, the subgroup $G_2$ permutes $\mathbb{S}^2$ paralllely.

Proof. Recalling that $G_2$ is the subgroup fixing $0$ and $\infty$, $g \in G_2$ translates to a planar homeomorphism fixing the origin, for which we must prove that circles centered at the origin are mapped onto circles centered at the origin. Given one such circle $\gamma$, let $\lambda = g(\gamma)$ be its image. Then, $\lambda$ is a Jordan curve, containing the origin in its interior. If $\lambda$ is not a circle, it contains points $p_m$ and $p_M$ such that

$$|p_m| \leq \min_{p \in \lambda} |p| < \max_{p \in \lambda} |p| = |p_M|.$$

For each polar angle $0 \leq \theta < 2\pi$, the semiradius $\vec{r}_{\theta} = \{te^{i\theta} : t \geq 0\}$ intercepts $\lambda$ in a compact set $\lambda_{\theta}$, in such a way that $\lambda = \bigcup_{0 \leq \theta < 2\pi} \lambda_{\theta}$. For some $\theta_0$, it must be the case that $|p| < |p_M|$ for every $p \in \lambda_{\theta_0}$.

Now, if $\theta_M$ is such that $p_M \in \lambda_{\theta_M}$, let $R$ be a planar rotation mapping the semiradius $\vec{r}_{\theta_M}$ onto the semiradius $\vec{r}_{\theta_0}$. Then, $R(p_M) \in R(\lambda)$, but $R(p_M) \in \text{ext} \lambda = \left(\text{int} \lambda\right)^c$, where int and ext are used in the Jordan Curve Theorem sense, as the bounded and unbounded open connected components of $\lambda^c$, sharing $\lambda$ as their common boundary. We remark that $\left(\text{int} \lambda\right)^c \subset \{z : |z| > |p_m|\}$. So, it cannot be the case that $R(\lambda)$ is fully contained within $\text{ext} \lambda$, since $|R(p_m)| = |p_m|$. Therefore, we also must have $R(\lambda) \cap \text{int} \lambda \neq \emptyset$. It follows that $R(\lambda) \cap \lambda \neq \emptyset$.

That said, we may obtain $p, q \in \gamma$ such that $R(g(p)) = g(q)$, implying $(g^{-1} \circ R \circ g)(p) = q$. Since $p, q$ lie on the same circle, there exists a planar rotation $U$ such that $p = U(q)$. But then, $g^{-1} \circ R \circ g \circ U$ defines an element of $G_2$ fixing $q$. By sharp 3-transitivity, it must be the identity. In particular, $g^{-1} \circ R \circ g \circ U(\gamma) = \gamma$. Since rotations leave $\gamma$ invariant, the above implies $R(\lambda) = \lambda$, which is a contradiction. Therefore, $\lambda$ has to be a circle. \qed
**Lemma 6.2.** Let $G \subset \text{Homeo}(\mathbb{S}^2)$ be a sharply $3$-transitive homogeneous group of diffeomorphisms. Then, every $g \in G_2$ is conformal at the poles.

**Proof.** Given such $g$, we know from **Lemma 6.1** that it permutes parallels. It thus suffices to consider the case of a planar diffeomorphism fixing the origin and mapping circles centered at the origin onto circles centered at the origin. We know that $A = Dg(0)$ is a linear isomorphism, so we may fix $z_m, z \in \mathbb{S}^1$ such that

$$0 < |Az_m| = \min_{z \in \mathbb{S}^1} |Az| \leq \max_{z \in \mathbb{S}^1} |Az| = |Az_M|.$$ 

Since $g$ preserves circles, for each $t \in (0,1)$:

$$1 = \frac{|g(tz_m)|}{|g(tz_M)|} = \frac{|A(tz_m) + o(|tz_m|)|}{|A(tz_M) + o(|tz_M|)|} = \frac{|A(z_m) + o(|tz_m|)|}{|A(z_M) + o(|tz_M|)|} \to |A(z_m)| \text{ as } t \to 0^+.$$ 

It follows that $A(\mathbb{S}^1)$ is a circle. Since $A$ is orientation preserving, this is enough to conclude that it is a conformal matrix. In other words, $g$ is conformal at $0$. □

Before proceeding to the proof of Theorem C, let us make a small remark: if $G$ is a $2$-transitive homogeneous group of diffeomorphisms and $\delta > 0$ is given, we may find $h_\delta \in G$ such that $h_\delta$ fixes $0$, but not $\infty$, and $Dh_\delta(0)$ is $\delta$-close to id.

Indeed, this can be verified as follows: by $2$-transitivity, we may fix $h \in G$ such that $h(0) = 0$ and $h(\infty) = 1$. Then, as long as $0 \leq t < 2\pi$, $h_t = h^{-1} \circ R_t \circ h$ fixes $0$, but not $\infty$, since $R_t(1)$ is a point on the equator distinct from $1$. It thus suffices to take $h_\delta = h_{\frac{\delta}{t}}$, for sufficiently small $t$. This might seem like an underuse of the $3$-transitivity hypothesis. But, as it turns out, a result from [4] implies that any $2$-transitive homogeneous group must actually be $3$-transitive. Either way, we are now ready to establish Theorem C.

**6.1. Proof of Theorem C.** Suppose, for the sake of contradiction, that $G$ contains a non-conformal mapping $g$. By precomposing and postcomposing with suitable rotations, it may be assumed that $g$ fixes $0$ and that $A = Dg(0)$ is a nonconformal matrix. This means that, for some pair of unit vectors $u_0$ and $v_0$,

$$\text{(15)} \quad \text{if } \alpha = \langle u_0, v_0 \rangle \text{ and } \beta = \left( \frac{Au_0}{|Au_0|}, \frac{Av_0}{|Av_0|} \right), \text{ then } \alpha \neq \beta.$$ 

Given $\varepsilon = |\beta - \alpha|/2$ there exists $\delta > 0$ such that, for every $u, v, w, z$ unit vectors, $|u - z| < \delta$ and $|v - w| < \delta$ imply $|\langle u, v \rangle - \langle z, w \rangle| < \varepsilon$. We then fix $h_\delta \in G$ such that

- $h_\delta(0) = 0$,
- $h_\delta^{-1}(\infty) = p_\delta \neq \infty$, and
- $C_{\delta} = Dh_{\delta}(0)$ is $\delta/2$-close to id.

Since $G$ is $3$-transitive, we may fix $f_\delta \in G_2$ such that $f_\delta(g(\infty)) = p_\delta$. From **Lemma 6.2**, we know that $B = Df_\delta(0)$ is conformal. If we let $g = h_\delta \circ f_\delta \circ g$, then also $g \in G_2$, and thus $D = Dg(0)$ is conformal. But, by the Chain Rule, $D = C_{\delta}BA$. We remark that, if $u_0$ is any unit vector, then

$$\left| \frac{C_{\delta}w_0}{C_{\delta}u_0} - w_0 \right| = \left| C_{\delta}w_0 - w_0 - \frac{|C_{\delta}w_0| - 1}{|C_{\delta}w_0|} C_{\delta}w_0 \right| \leq 2 \left| C_{\delta}w_0 - w_0 \right| < \delta.$$ 

Since $B$ is conformal,

$$\left| \frac{Du_0}{|Du_0|} - \frac{Au_0}{|Au_0|} \right| < \delta \text{ and } \left| \frac{Dv_0}{|Dv_0|} - \frac{Av_0}{|Av_0|} \right| < \delta.$$ 

Accordingly,

$$\left| \frac{Du_0}{|Du_0|} - \frac{Dv_0}{|Dv_0|} \right| = \left| \left( \frac{Du_0}{|Du_0|}, \frac{Dv_0}{|Dv_0|} \right) - \beta \right| < \varepsilon.$$ 

By the choice of $\varepsilon$, this ensures that the angle between $Du_0$ and $Dv_0$ is different from the angle between $u_0$ and $v_0$, contradicting $D$ conformal. Thus, such nonconformal $g \in G$ cannot exist, and $G$ is a subgroup of Möb($\mathbb{S}^2$).
Lastly, let $M \in \text{M"ob}(S^2)$. Then, there exists $h \in G \subset \text{M"ob}(S^2)$ such that $h(0) = M(0)$, $h(1) = M(1)$ and $h(\infty) = M(\infty)$. By the sharp 3-transitivity of $\text{M"ob}(S^2)$, this implies $M = h \in G$. Since $M$ was arbitrary, $G = \text{M"ob}(S^2)$ follows.

**Acknowledgements.** U. Lakatos would like to thank Prof. S. Alvarez at Udelar for his genuine interest and fruitful discussions on the obtainment of uniform bounds on the proof of Lemma 3.5.

**References**

[1] Étienne Ghys, *Groups acting on the circle*, L’Enseignement Mathématique t. 47 (2001), 329–407, DOI https://dx.doi.org/10.5169/seals-65441.

[2] James Giblin and Vladimir Markovic, *Classification of continuously transitive circle groups*, Geometry and Topology Volume 10, issue 3 (2006), 1319–1346, DOI https://dx.doi.org/10.2140/gt.2006.10.1319.

[3] Frédéric Le Roux, *On closed subgroups of the group of homeomorphisms of a manifold*, Journal de l’École Polytechnique – Mathématiques t. 1 (2014), 147–159, DOI https://10.5802/jep.7.

[4] Ferry Kwakkel and Fábio Armando Tal, *Homogeneous transformation groups of the sphere v2* (2014), arXiv:1309.0179 [math.GT]

[5] Boris Kolev, *Sous-groupes compacts d’homéomorphismes de la sphère*, L’Enseignement Mathématique t. 52 (2006), 193–214, DOI https://dx.doi.org/10.5169/seals-2231.

[6] Tristan Needham, *Visual Complex Analysis*, Oxford University Press, 2000.

[7] Augustin Banyaga, *The Structure of Classical Diffeomorphism Groups*, Springer Science, 1997.

[8] John Franks and Michal Misiurewicz, *Chapter 7 – Topological methods in dynamics*, in: Handbook of Dynamical Systems, 2002, pp. 547–598, DOI https://doi.org/10.1016/S1874-575X(02)80009-1.

[9] Takashi Matsuoka, *5. Periodic points and braid theory*, in: Handbook of Topological Fixed Point Theory, 2005, pp. 129–171.

[10] Michael Handel, *Global shadowing of pseudo-Anosov homeomorphisms*, Ergodic Theory and Dynamical Systems 5 (1985), no. 3, 373–377, DOI https://doi.org/10.1017/S0143385700003011.

**INSTITUTO DE MATEMÁTICA E ESTATÍSTICA, UNIVERSIDADE DE SÃO PAULO, RUA DO MATÃO 1010, CIDADE UNIVERSITÁRIA, 05508-090 SÃO PAULO, SP, BRAZIL**

Email address: lakatos@ime.usp.br

URL: www.ime.usp.br/~lakatos

**INSTITUTO DE MATEMÁTICA E ESTATÍSTICA, UNIVERSIDADE DE SÃO PAULO, RUA DO MATÃO 1010, CIDADE UNIVERSITÁRIA, 05508-090 SÃO PAULO, SP, BRAZIL**

Email address: fabiotal@ime.usp.br