Ergod. Th. & Dynam. Sys., (2024), **44**, 1102–1122
doi:10.1017/etds.2023.32

Proper extensions of the 2-sphere's conformal group present entropy and are 4-transitive

ULISSES LAKATOS and FÁBIO TAL

Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão 1010, Cidade Universitária, 05508-090 São Paulo, SP, Brazil (e-mail: lakatos@ime.usp.br, fabiotal@ime.usp.br)

(*Received* 24 *June* 2022 *and accepted in revised form* 24 *March* 2023)

Abstract. In this paper, we prove using elementary techniques 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. As an important consequence, we derive that any such group must always contain an element of positive topological entropy. We also provide a self-contained characterization, in terms of transitivity, of the Möbius transformations within the full group of sphere diffeomorphisms.

Key words: transitivity, group actions on surfaces, topological groups, topological entropy
2020 Mathematics Subject Classification: 57M60 (Primary); 37B05, 57K20 (Secondary)

1. *Introduction*

Let M be a closed and oriented topological manifold, and consider the set $\mathrm{Homeo}(M)$ of all its *orientation-preserving homeomorphisms*. The usual *uniform metric* turns this set—endowed with the composition operation—into a topological group, the subgroups of which one may try to understand and classify.

In his early 2000s' essay [5], Ghys proposed such a classification for closed and transitive groups acting on the unit circle. 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 Giblin and Markovic, see [6].

A relevant part of understanding the closed subgroups of $\mathrm{Homeo}(M)$ 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, Le Roux proved in [10] that, in triangulable manifolds of dimension two or higher, the group of area-preserving homeomorphisms is always maximal.

Specializing in the 2-sphere $\mathbb{S}^2$, F. Kwakkel and the second author derived in [9] a number of results concerning subgroups of $\mathrm{Homeo}(\mathbb{S}^2)$ related to key symmetries, one of which is $\mathrm{Möb}(\mathbb{S}^2)$: the *Möbius group* of conformal diffeomorphisms. Among others, they left open the question of whether there is no proper intermediate group of diffeomorphisms between $\mathrm{Möb}(\mathbb{S}^2)$ and $\mathrm{Homeo}(\mathbb{S}^2)$, up to uniform closure. (Due to a difference of opinion between its authors, [9] remains as an unsubmitted preprint, of which we refer to version 2 in ArXiv. This note is self-contained, and makes no direct use of the results therein.)

This question is known to have a positive answer in the case of the circle, where the proof is related to higher orders of transitivity. In this paper, we provide the following insight into extensions of $\mathrm{Möb}(\mathbb{S}^2)$, from the transitivity viewpoint.

THEOREM A. *Let $G \subset \mathrm{Diff}^1(\mathbb{S}^2)$ be a group properly extending* $\mathrm{Möb}(\mathbb{S}^2)$. *Then, its identity component is arc 4-transitive. In particular, G is at least 4-transitive.*

Above, the stronger concept of *arc-transitivity* is introduced. It means that not only ordered lists $(p_1, \ldots, p_4)$ and $(q_1, \ldots, q_4)$ can be mapped one onto another by a transformation in G, each q_i is actually the endpoint of p_i's trajectory under a isotopy in G starting at the identity. Theorem A—or rather its proof—has an interesting dynamical implication: the constructions used to derive it also allow one to deduce the presence of an element having positive topological entropy in G.

THEOREM B. *Let $G \subset \mathrm{Diff}^1(\mathbb{S}^2)$ be a group properly extending* $\mathrm{Möb}(\mathbb{S}^2)$. *Then, G contains an element f fixing at least four points, and such that its restriction to the complement of these points is isotopic to a pseudo-Anosov map relative to them. In particular, f has strictly positive topological entropy.*

Theorem B could, in principle, be derived from Theorem A along with an abstract result of [9]. However, we favor the more explicit construction presented here, as both rely on Nielsen–Thurston classification theory. We also include an interesting result—albeit not new—stemming from the techniques used to show Theorems A and B. It is stated in terms of the key definition below.

Definition 1.1. Given $k \in \mathbb{N}$, the action of a subgroup $G \subset \mathrm{Homeo}(M)$ 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 consisting of mutually distinct points, there exists some transformation $g \in G$ such that $q_i = g(p_i)$ for each $i \in \{1, \ldots, k\}$. If, in addition, such g is unique, the action is said to be *sharply k-transitive*.

It is well known that $\mathrm{Möb}(\mathbb{S}^2)$ is sharply 3-transitive. As it turns out, this is a *defining property*: in [9], Kwakkel and the second author show that a sharply 3-transitive subgroup of $\mathrm{Diff}^1(\mathbb{S}^2)$ extending the group of rotations must be $\mathrm{Möb}(\mathbb{S}^2)$. Since the cited work has not been published, we present for completeness a new simple, independent and—in our opinion—rather amusing proof of this fact, as stated below.

THEOREM C. *Let $G \subset \mathrm{Diff}^1(\mathbb{S}^2)$ be a group extending the rotations group* $\mathrm{Rot}(\mathbb{S}^2)$. *Then, G sharply 3-transitive implies $G = \mathrm{Möb}(\mathbb{S}^2)$.*

https://doi.org/10.1017/etds.2023.32 Published online by Cambridge University Press

1.1. *Preliminaries.* Let us start by terminology. As aforementioned, $\mathrm{Homeo}(M)$ denotes the set of orientation-preserving homeomorphisms of a closed and oriented topological manifold M, on which the uniform metric is well defined. If M is smooth, we may also consider the subset $\mathrm{Diff}^1(M) \subset \mathrm{Homeo}(M)$ of orientation-preserving diffeomorphisms of class C^1 and its finer *C^1 topology*, which takes into account the local expressions of differentials as well.

Consider the circle $M = \mathbb{S}^1$, which can be thought of as the set of complex numbers of unit absolute value or as the compactification $\mathbb{R} \cup \{\infty\}$. Each such description is linked to a canonical group acting on it. In the first case, we have the *rotations group* $\mathrm{Rot}(\mathbb{S}^1)$, which is naturally identified with the circle itself, thus being a compact Lie group. In the second case, we have the *Möbius group* $\mathrm{Möb}(\mathbb{S}^1)$, obtained by the action of $\mathrm{PSL}(2; \mathbb{R})$ on the extended real line via *linear fractional transformations*.

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

THEOREM. (Giblin and Markovic [6]). *Let G be a closed and transitive subgroup of* $\mathrm{Homeo}(\mathbb{S}^1)$ *containing a non-trivial path connected component. Then, G is conjugate to one, and only one, of the following:*

(i) $\mathrm{Rot}(\mathbb{S}^1)$;
(ii) $\mathrm{Möb}_k(\mathbb{S}^1)$; *or*
(iii) $\mathrm{Homeo}_k(\mathbb{S}^1)$,

where the subscript $k \in \mathbb{N}$ indicates cyclic cover of order k. Furthermore, $\mathrm{Möb}(\mathbb{S}^1)$ *is maximal: there is no closed subgroup properly containing* $\mathrm{Möb}(\mathbb{S}^1)$ *and properly contained in* $\mathrm{Homeo}(\mathbb{S}^1)$.

Consider now the 2-sphere $M = \mathbb{S}^2$. It can be thought of either as the points of Euclidean 3-space 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 $(0, 0, 1)$. We denote points on the plane and their stereographic images by the same letters and confound them without notice. Any sphere map fixing ∞ thus defines, by stereographic conjugation, a planar map with the same degree of regularity, also denoted by the same letter.

The group of *sphere rotations* $\mathrm{Rot}(\mathbb{S}^2)$ consists of Euclidean isometries preserving both $\mathbb{S}^2$ and orientation. It is realized by $\mathrm{SO}(3; \mathbb{R})$, and is thus a compact Lie group. Each such transformation amounts to prescribing an axis and a rotation angle around that axis. From these facts, $\mathrm{Rot}(\mathbb{S}^2)$ is seen to be closed and transitive. Rotations are *minimal* in the following sense: a compact group $G \subset \mathrm{Homeo}(\mathbb{S}^2)$ must be topologically conjugate to a closed subgroup of $\mathrm{Rot}(\mathbb{S}^2)$—a result from the 1940s by Kerékjártó. A proof according to contemporary standards is given by Kolev in [8]. If G is further transitive, then it is actually conjugate to $\mathrm{Rot}(\mathbb{S}^2)$. The problem of classifying the groups contained between $\mathrm{Rot}(\mathbb{S}^2)$ and $\mathrm{Homeo}(\mathbb{S}^2)$ is thus called the *kernel subgroup problem*, and such intermediate groups containing the rotations are hereafter called *homogeneous*, following the terminology established in [9].

The *Möbius group* $\text{Möb}(\mathbb{S}^2)$ 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:

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

This procedure characterizes the *conformal diffeomorphisms* of $\mathbb{S}^2$ endowed with its canonical smooth structure. The Möbius group is thoroughly understood—see [4] and [12, Ch. 3]. In particular, spherical rotations are induced by the subgroup $\text{PSU}(2; \mathbb{C})$. Le Roux, Kwakkel, and the second author left open the question of whether the inclusion of $\text{Möb}(\mathbb{S}^2)$ in $\text{Homeo}(\mathbb{S}^2)$ is maximal. Our contribution to this problem is Theorem A above.

It should be stressed that, in the circle setting, 4-transitivity is sufficient to ensure k-transitivity for any $k > 4$. However, the argument used in [6] to do so presents no obvious generalization to higher dimensions, for it relies on the complement of a finite subset of $\mathbb{S}^1$ being composed of disjoint open intervals.

The constructions in this paper are mostly based upon *isotopies*, here understood as families $(f_t)_{t\in I}$ for which $t \mapsto f_t$ is continuous with respect to the uniform metric. When the interval in question is the *standard unit interval*, we denote it by $\mathbb{I} = [0, 1]$. Often, it will 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 the uniform metric. To avoid confusion, isotopies for which this association is actually continuous with respect to the C^1 topology will be explicitly referred to as *diffeotopies*.

Definition 1.2. Let $G \subset \text{Homeo}(M)$ be a subgroup. 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 *$\mathcal{I}G$-isotopy*. Given a point $x \in M$, we define its *trajectory* under f as

$$\gamma_f(x) = \{f_t(x) : t \in I\}.$$

It will often be thought of as a path, oriented according to its natural direction of travel along increasing values of t. If I is unbounded above, we further define the *ω-limit* as the following set of accumulation points:

$$\omega_f(x) = \{y \in M : \text{ there exists a sequence } t_n \nearrow +\infty \text{ such that } f_{t_n}(x) \to y\}.$$

There is an analogous notion of α-limit when I is unbounded below. Lastly, G is said to be *arc k-transitive* if for every pair of k-tuples, as in Definition 1.1, there exists an $\mathcal{I}G$-isotopy $(g_t)_{t\in\mathbb{I}}$ such that $g_1(p_i) = q_i$ for each $i \in \{1, \ldots, k\}$.

Despite analogies between isotopy trajectories and flows being very limited, such suggestive terminologies prove themselves pictorially useful in arguments to come. The reason is the concept we now develop, which plays vaguely the same role as that of the semigroup property for flows.

Definition 1.3. For a fixed subgroup $G \subset \text{Homeo}(M)$ and given $z, w \in M$, say that

$$z \sim_G w \iff \text{there exists an } \mathcal{I}G\text{-isotopy } (f_t)_{t\in\mathbb{I}} \text{ in } G \text{ such that } f_1(z) = w.$$

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

LEMMA 1.4. *Let $(f_t)_{t\in I}$ and $(h_t)_{t\in J}$ be $\mathcal{I}G$-isotopies, where I and J are intervals of any kind. Then, for any two $z, w \in M$, $\gamma_f(z) \cap \gamma_h(w) \neq \emptyset$ implies $\mathcal{A}_G(w) = \mathcal{A}_G(z)$.*

Proof. Assume $f_a(z) = h_b(w)$ for some $a \in I$ and $b \in J$. Given $y \in \mathcal{A}_G(w)$, there exists an $\mathcal{I}G$-isotopy $(g_t)_{t\in\mathbb{I}}$ such that $g_1(w) = y$. Thus,

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

is an $\mathcal{I}G$-isotopy satisfying $k_1(z) = y$. This shows that $\mathcal{A}_G(w) \subset \mathcal{A}_G(z)$, and the converse inclusion follows by symmetry. □

We finish this introduction fixing some references: $\boldsymbol{\infty}$ is the North Pole, $\mathbf{0}$ is the South Pole, corresponding to the plane's origin, and $\mathbf{1}$ is the point $(1, 0, 0)$, corresponding to its counterpart on the real axis. The meridian through $\mathbf{0}$, $\mathbf{1}$ and $\boldsymbol{\infty}$ is denoted by Γ. It defines on its left a *western hemisphere* $\mathcal{H}^+$, corresponding to the upper half-plane, and on its right an *eastern hemisphere* $\mathcal{H}^-$. For any group G, we also adopt the following notation for stabilizers:

$$G_1 := \text{Stab}_G\{\boldsymbol{\infty}\}, \quad G_2 := \text{Stab}_G\{\mathbf{0}, \boldsymbol{\infty}\} \quad \text{and} \quad G_3 := \text{Stab}_G\{\mathbf{0}, \mathbf{1}, \boldsymbol{\infty}\}.$$

1.2. *Paper outline.* Given a group G properly extending $\text{Möb}(\mathbb{S}^2)$, 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 §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.

From this starting point, we consider the reference meridian Γ and promote two parallel processes. In §3, we show that every point not in Γ admits a full time isotopy in G_3 under which its trajectory accumulates at the poles. In §4, we obtain a finite time isotopy in G_3 under which a point is moved from one side of the meridian to the other.

In §5, we show our main theorems. Theorem A is derived upon combining isotopies of the types just described and concluding that all but three points of the sphere are arc connected in the sense of Definition 1.2. 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.

In §6, we independently 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*

For a given $0 \leq \theta < 2\pi$, we let R_θ denote the counterclockwise rotation of angle θ around the axis from $\mathbf{0}$ to ∞. If it is thought of as a planar map, it may be written as $R_\theta(z) = \mathrm{e}^{\mathrm{i}\,\theta} z$. Also, for a given $\rho > 0$, we let $H_\rho \in \text{Möb}_2(\mathbb{S}^2)$ denote the transformation induced by the planar homothety $z \mapsto \rho z$. Since the tangent space $T_{\mathbf{0}}\mathbb{S}^2$ can be identified with the horizontal subspace $\mathbb{R}^2 \times \{0\} \simeq \mathbb{R}^2$, we may write $\mathrm{D}R_\theta(\mathbf{0}) = R_\theta$ and $\mathrm{D}H_\rho(\mathbf{0}) = \rho\,\text{id}$, in a slight abuse of notation.

LEMMA 2.1. *Let* $G \subset \text{Diff}^1(\mathbb{S}^2)$ *be a group properly extending* $\text{Möb}(\mathbb{S}^2)$. *Then, there exists* $\hat{g} \in G_2$ *for which* $\mathrm{D}\hat{g}(\mathbf{0}) = \text{diag}[\lambda, \lambda^{-1}]$ *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 map $\hat{h}$ which is non-conformal 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 $A = \mathrm{D}\hat{h}(\mathbf{0})$ is non-conformal.

This means that there exists a pair of unit vectors u, v such that $\text{ang}(Au, Av) \neq \text{ang}(u, v)$. Let $R_{(1)}$ be a planar rotation such that $R_{(1)}(Au)$ is a positive multiple of u. Then, $A_1 = R_{(1)}A$ has an eigenvalue $\lambda_1 > 0$ for which u is a unit eigenvector. Also, λ_1 cannot be of maximal geometric multiplicity, leaving two possibilities.

- If A_1 is defective, let $\{u, w\}$ form an orthogonal chain of generalized eigenvectors. That is, $A_1 w = u + \lambda_1 w$ and $\langle u, w\rangle = 0$, where $\langle\cdot,\cdot\rangle$ is the Euclidean inner product. For each $\phi \in [0, \pi/2]$, consider
$$x_\phi := \cos\phi\; u + \sin\phi\; \frac{w}{|w|} \quad \text{and} \quad y_\phi := -\sin\phi\; u + \cos\phi\; \frac{w}{|w|}.$$
Then, $A_1 x_{\phi_0}$ and $A_1 y_{\phi_0}$ are orthogonal for some $0 < \phi_0 < \pi/2$, since
$$0 < \langle A_1 x_0, A_1 y_0\rangle = -\langle A_1 x_{\pi/2}, A_1 y_{\pi/2}\rangle.$$
Define $\hat{x} := x_{\phi_0}$ and $\hat{y} := y_{\phi_0}$. Since A_1 preserves orientation, 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 planar rotation $R_{(2)}$. Therefore, $R_{(2)}(A_1\hat{x}) = \nu_1\,\hat{x}$ and $R_{(2)}(A_1\hat{y}) = \nu_2\,\hat{y}$, where (say) $0 < \nu_1 < \nu_2$.
Lastly, let R be a planar rotation such that $R^{-1}\,\hat{x} = \partial/\partial x$, $\rho := (\nu_1\nu_2)^{-1/2}$, $\lambda := (\nu_1/\nu_2)^{1/2} < 1$, and $\hat{A} := \rho\, R^{-1}R_{(2)}A_1R$. Then, $\hat{A}$ has the form described in the lemma's statement with respect to the canonical basis. Also, $\hat{A} = \mathrm{D}\hat{g}(\mathbf{0})$, where $\hat{g} = R^{-1} \circ H_\rho \circ R_{(2)} \circ R_{(1)} \circ \hat{h} \circ R \in G_2$. This completes the proof in this case.
- If A_1 has a second eigenvalue $\lambda_2 \neq \lambda_1$, fix a positive basis of unit eigenvectors $\{u, w\}$ and let $w^\perp$ be the orthogonal complement of w with respect to u. Letting
$$x_\phi := \cos\phi\; u + \sin\phi\; \frac{w^\perp}{|w^\perp|} \quad \text{and} \quad y_\phi := -\sin\phi\; u + \cos\phi\; \frac{w^\perp}{|w^\perp|},$$
the proof follows as before, but now $R_{(2)}$ may be the identity if $w = w^\perp$. □

LEMMA 2.2. (Extension Lemma) *Let* $G \subset \text{Diff}^1(\mathbb{S}^2)$ *be a group properly extending* $\text{Möb}(\mathbb{S}^2)$. *Then, there exists an* $\mathcal{I}G_2$*-diffeotopy* $(g_t)_{t\in\mathbb{I}}$ *such that:*

(1) *for every $t > 0$, the differential $\mathrm{D}g_t(\mathbf{0})$ is a hyperbolic saddle, having the tangent line $T_{\mathbf{0}}\Gamma$ as its stable direction;*
(2) $\mathrm{D}g_1(\mathbf{0}) = \mathrm{diag}[\lambda, \lambda^{-1}]$ *with respect to the canonical basis of $T_{\mathbf{0}}\mathbb{S}^2$, where $0 < \lambda < 1$.*

Proof. Lemma 2.1 yields $\hat{g} \in G_2$ such that $\hat{A} := \mathrm{D}\hat{g}(\mathbf{0}) = \mathrm{diag}[\mu, \mu^{-1}]$, $0 < \mu < 1$. For $s \in [0, \pi/2]$, we let $B_s := \hat{A}^{-1} R_s \hat{A}$, $v_s := B_s(\partial/\partial x)$, and consider the continuous function $\theta(s)$ given by the angle between v_s and $\partial/\partial x$. If $C_s := R_{-\theta(s)} B_s$, then $C_0 = \mathrm{id}$ and C_s has the x-axis as a contracting direction of rate $\lambda_s = |v_s| < 1$ for $s > 0$. Since C_s has unit determinant, it is actually a hyperbolic saddle. In general it may not be diagonal, but we explicitly compute $C_{\pi/2} = \mathrm{diag}[\lambda, \lambda^{-1}]$, where $\lambda = \mu^2$. Lastly, we define $(g_t)_{t\in\mathbb{I}}$ as

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

Since θ is continuous, equation (1) is continuous with respect to the C^1 topology. For each t, we have $\mathrm{D}g_t(\mathbf{0}) = C_{\pi t/2}$. Given that the x-axis corresponds to Γ—with the direction generated by $\partial/\partial x$ identified with $T_{\mathbf{0}}\Gamma$—the considerations previously made translate into the statements of the lemma, completing the proof. □

3. *A fundamental lemma*

For each angle $\alpha \in (0, \pi/2)$, consider the *closed cone*

$$\mathcal{C}_\alpha = \{r e^{\mathrm{i}\,\theta} : r \geq 0 \text{ and either } |\theta| \leq \alpha \text{ or } |\theta - \pi| \leq \alpha\}.$$

If $0 < \lambda < 1$, any such cone is 'broadened' under the action of the hyperbolic matrix $A = \mathrm{diag}[\lambda, \lambda^{-1}]$, $0 < \lambda < 1$. More precisely, $A(\mathcal{C}_\alpha) = \mathcal{C}_{\alpha+2\tau}$, for some $\tau > 0$. Approximating the polar angle of $g(z)$ by that of $\mathrm{D}g(\mathbf{0})z$ then implies the following lemma.

LEMMA 3.1. *Let g be a planar diffeomorphism for which $\mathrm{D}g(\mathbf{0}) = \mathrm{diag}[\lambda, \lambda^{-1}]$, $0 < \lambda < 1$. Then, for a given $0 < \alpha < \pi/2$, there exist $\tau > 0$ and $\delta > 0$ satisfying:*

$$0 < |z| < \delta \quad \text{and} \quad z \notin \mathcal{C}_\alpha \text{ imply } R_\omega(g(z)) \notin \mathcal{C}_\alpha, \quad \text{whenever } |\omega| < \tau.$$

In particular, while $\{z, g(z), \ldots, g^{k-1}(z)\}$ remains in $\mathbb{D}_\delta(\mathbf{0})$, $g^k(z) \notin \mathcal{C}_\alpha$.

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.

LEMMA 3.2. *Let $(g_t)_{t\in\mathbb{I}}$ be a planar diffeotopy such that the origin is a fixed point and $\mathrm{D}g_t(\mathbf{0})$ 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 $z \in \mathbb{D}_\rho(\mathbf{0}) \setminus \mathcal{C}_\alpha$ imply $g_t(z) \notin \mathcal{C}_\beta$ for every $t \in \mathbb{I}$.*

Proof. For each fixed $t \in \mathbb{I}$, let $A_t := \mathrm{D}g_t(\mathbf{0})$. Consider v_α and v_α^* unit vectors of polar angle α and $2\pi - \alpha$, respectively, whose spans delimit $\mathcal{C}_\alpha$. Upon defining $\beta_t = \min\{\alpha, \mathrm{ang}(A_t v_\alpha, {}^{\partial}\!/_{\partial x}), \mathrm{ang}(A_t v_\alpha^*, {}^{\partial}\!/_{\partial x})\}$, we have $0 < \beta_t \leq \alpha$ and $A_t v \notin \mathcal{C}_{\beta_t}$ whenever $v \notin \mathcal{C}_\alpha$. This yields a global solution to the associated linear problem.

Once this 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 $|A_t v| \geq c_t |v|$ for every v. One then obtains $\rho_t > 0$ such that:

$$|g_t(z) - A_t z| \leq \varepsilon_t |z|/2 \quad \text{whenever } 0 < |z| < \rho_t. \tag{2}$$

https://doi.org/10.1017/etds.2023.32 Published online by Cambridge University Press

With respect to the C^1 topology (see e.g. [1]), continuity of the isotopy yields $\delta_t > 0$ such that $s \in \mathbb{I}$ and $|s-t| < \delta_t$ imply:

$$\sup_{\overline{\mathbb{D}}_{\rho_t}(\mathbf{0})} \left\{ |g_s - g_t|, \|\mathrm{D}g_s - \mathrm{D}g_t\| \right\} \leq \varepsilon_t/2. \tag{3}$$

Thus, when $|s-t| < \delta_t$ and $|z| < \rho_t$ simultaneously, we have $|g_s(z) - A_t z| \leq \varepsilon_t |z|$, by the mean value inequality along with equations (2) and (3).

It follows that $|g_s(z)| \geq (c_t - \varepsilon_t)|z|$, and the right-hand side of the inequality is strictly positive. This allows us to conclude that $g_s(z)$ lies in a closed disk centered at $A_t z$ and of radius $\varepsilon_t |z|$, to which the origin is an external point. In particular, $\mathrm{ang}(g_s(z), A_t z) < \pi/2$. An elementary chord length formula then implies

$$\sin\left[\frac{\mathrm{ang}(g_s(z), A_t z)}{2}\right] \leq \frac{|g_s(z) - A_t z|}{(c_t - \varepsilon_t)|z|} \leq \frac{\varepsilon_t |z|}{(c_t - \varepsilon_t)|z|} = \sin\left(\frac{\beta_t}{4}\right).$$

We conclude that $\mathrm{ang}(g_s(z), A_t z) \leq \beta_t/2$, as long as $|s-t| < \delta_t$ and $0 < |z| < \rho_t$. If also $z \notin \mathcal{C}_\alpha$, the linear case implies $A_t z \notin \mathcal{C}_{\beta_t}$, and $g_s(z) \notin \mathcal{C}_{\beta_t/2}$ follows. Covering $\mathbb{I}$ by finitely many $(t - \delta_t, t + \delta_t)$ intervals finishes the proof. □

COROLLARY 3.3. *Let $(g_t)_{t\in\mathbb{I}}$ be a planar diffeotopy such that the origin is a fixed point and $\mathrm{D}g_t(\mathbf{0})$ has the x-axis as an invariant direction for every t. Then, given $0 < \alpha < \pi/2$, there exist $0 < \beta^- < \beta^+ < \alpha$ and $\rho > 0$ such that:*

(1) *$z \in \mathbb{D}_\rho(\mathbf{0})$ and $z \notin \mathcal{C}_\alpha$ imply $g_t(z) \notin \mathcal{C}_{\beta^+}$ for every $t \in \mathbb{I}$;*

(2) *$z \in \mathbb{D}_\rho(\mathbf{0})$ and $z \in \mathcal{C}_{\beta^-}$ imply $g_t(z) \in \mathcal{C}_{\beta^+}$ for every $t \in \mathbb{I}$.*

Proof. Given α, Lemma 3.2 yields $0 < \beta^+ < \alpha$ and $\rho^+ > 0$ such that $g_t(z) \notin \mathcal{C}_{\beta^+}$ for every $t \in \mathbb{I}$, whenever $z \in \mathbb{D}_{\rho^+}(\mathbf{0})$ and $z \notin \mathcal{C}_\alpha$. The diffeotopy $(g_t^{-1})_{t\in\mathbb{I}}$ also satisfies the hypotheses listed in the lemma. So, for this diffeotopy and the angle β^+ just encountered, we obtain $\rho^- > 0$ and $0 < \beta^- < \beta^+$ such that $w \in \mathbb{D}_{\rho^-}(\mathbf{0}) \setminus \mathcal{C}_{\beta^+}$ imply $g_t^{-1}(w) \notin \mathcal{C}_{\beta^-}$ for every $t \in \mathbb{I}$.

Let $\eta > 0$ be such that $|g_t(z)| < \rho^-$ for every $t \in \mathbb{I}$, whenever $|z| < \eta$. Then, by setting ρ as $\min\{\eta, \rho^+\}$, we have the proposed statements satisfied. Indeed, if that was not the case, $g_s(z) \notin \mathcal{C}_{\beta^+}$ for some $s \in \mathbb{I}$ and $z \in \mathbb{D}_\rho(\mathbf{0}) \cap \mathcal{C}_{\beta^-}$ would lead to a contradiction. □

Definition 3.4. Given finite and non-zero points z, w, we let $M[z, w] \in \mathrm{M\ddot{o}b}_2(\mathbb{S}^2)$ be the *unique* Möbius transformation fixing the poles and mapping z to w. Also, we denote $\hat{M}[z] := M[z, \mathbf{1}]$.

We can write down explicit formulas for these transformations and see that $(z, w) \mapsto M[z, w]$ is continuous. Also, if $\mathcal{K} \subset \mathbb{S}^2$ is a non-empty compact set bounded away from $\mathbf{0}$,

$$\text{the sets } \hat{M}[x](\mathcal{K}) \text{ converge to } \{\infty\} \text{ on the Hausdorff distance as } x \to \mathbf{0}. \tag{4}$$

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

LEMMA 3.5. (Fundamental Lemma) *Let $G \subset \mathrm{Diff}^1(\mathbb{S}^2)$ be a group properly extending $\mathrm{M\ddot{o}b}(\mathbb{S}^2)$. Then, for each given point $z_0 \notin \Gamma$, there exists an $\mathcal{I}G_3$-isotopy $(I_t^{z_0})_{t\geq 0}$, depending on z_0, such that:*

https://doi.org/10.1017/etds.2023.32 Published online by Cambridge University Press

(1) *the trajectory of z_0 under I^{z_0} does not intersect* Γ*; and*
(2) *the ω-limit of z_0 satisfies $\omega_{I^{z_0}}(z_0) = \{\infty\}$.*

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 $z_0 = R_0\, \mathrm{e}^{\mathrm{i}\,\theta_0}$, $0 < \theta_0 < \pi$.

Let $(g_t)_{t\in\mathbb{I}}$ be as in the Extension Lemma. Since ∞ is fixed throughout, it can be thought of as a planar $\mathcal{I}G_2$-diffeotopy such that $\mathrm{D}g_t(\mathbf{0})$ has the x-axis as an invariant direction for every t and $\mathrm{D}g_1(\mathbf{0}) = \mathrm{diag}[\lambda, \lambda^{-1}]$, $0 < \lambda < 1$. To ease notation, we write $g_1 := g$ and $\mathrm{D}g(\mathbf{0}) := A$.

Fix some $0 < \alpha < \pi/2$ such that the direction through θ_0 is external to $\mathcal{C}_{2\alpha}$. *With respect to* α, let $\delta > 0$ and $\tau > 0$ be as described in Lemma 3.1. Regarding this same α, and also the diffeotopy $(g_t)_{t\in\mathbb{I}}$, Corollary 3.3 yields a radius $\rho > 0$ and angles $0 < \beta^- < \beta^+ < \alpha$ such that, for $z \in \mathbb{D}_\rho(\mathbf{0})$ and every $t \in \mathbb{I}$, $z \notin \mathcal{C}_\alpha$ implies $g_t(z) \notin \mathcal{C}_{\beta^+}$, while $z \in \mathcal{C}_{\beta^-}$ implies $g_t(z) \in \mathcal{C}_{\beta^+}$.

Lastly, we characterize the stable manifold $\mathrm{W^s}$ of g at $\mathbf{0}$. Since the stable direction of A is the x-axis, $\mathrm{W^s}$ may be assumed to be a Lipschitz graph of the form $y = y(x)$ having horizontal tangent at the origin. So, for τ and β^- as obtained above and a sufficiently small radius $\sigma > 0$, it can be assumed that $\mathrm{W^s} \cap \mathbb{D}_\sigma(\mathbf{0}) \subset \mathcal{C}_{1/2\,\min\{\tau,\beta^-\}}$.

Picking $0 < \rho_0 < \min\{\delta, \rho, \sigma, 1\}$, all of the conditions described so far are satisfied in the disk $\mathbb{D}_0 := \mathbb{D}_{\rho_0}(\mathbf{0})$. Once these choices are made, fix a positive real number $0 < r_1 < \min\{\rho_0, \rho_0/R_0\}$ and let $\rho_1 := r_1 R_0$, $\mathcal{K}_0 := \overline{\partial\mathbb{D}_{\rho_1}(\mathbf{0}) \setminus \mathcal{C}_\alpha}$.

Then, $\mathcal{K}_0$ is a compact set contained in $\mathbb{D}_0$ while not intersecting the stable manifold, so there exists $n_0 \in \mathbb{N}$ such that:

$$\text{if } n_x = \min\{n \in \mathbb{N} : g^n(x) \notin \overline{\mathbb{D}_0}\} \quad \text{then } \; n_x \le n_0 \text{ for every } x \in \mathcal{K}_0. \tag{5}$$

Notice that $r_1 < \rho_0$. Since the stable manifold $\mathrm{W^s}$ is locally given as a Lipschitz graph $y = y(x)$, for some τ_1 with $|\tau_1| \le \frac{1}{2}\min\{\tau, \beta^-\}$, we have $r_1\mathrm{e}^{\mathrm{i}\,\tau_1} \in \mathrm{W^s}$.

CLAIM 3.5.1. *Define $M_1(z) := r_1\mathrm{e}^{\mathrm{i}\,\tau_1}z$. Then, M_1 has the following properties:*
(i) $M_1 \in \mathrm{M\ddot{o}b}_2(\mathbb{S}^2)$*;*
(ii) $v_1 := M_1(\mathbf{1}) \in \mathrm{W^s}$*;*
(iii) $w_1 := M_1(z_0) \in \mathcal{K}_0$.

Proof of claim. Item (i) follows from the form of M_1 and item (ii) from the choice of τ_1. As for item (iii), consider $w_1 = \mathrm{e}^{\mathrm{i}\,\tau_1}(r_1 z_0)$. Since $r_1 z_0$ also has polar angle θ_0, w_1 has polar angle $\theta_0 + \tau_1$. Then, $|\tau_1| < \beta^- < \alpha$ and $2\alpha < \theta_0 < \pi - 2\alpha$ together imply $\alpha < \theta_0 + \tau_1 < \pi - \alpha$, so $w_1 \notin \mathcal{C}_\alpha$. Furthermore, $|w_1| = \rho_1$, yielding $w_1 \in \mathcal{K}_0$. □

CLAIM 3.5.2. *Let $n_1 := n_{w_1}$ be as in equation (5). Consider $f_t := g_{t-\lfloor t\rfloor} \circ g^{\lfloor t\rfloor} \circ M_1$, where $t \in (0, n_1]$. Then, f is an isotopy such that $f_t(z_0) \notin \mathcal{C}_{\beta^+}$ and $f_t(\mathbf{1}) \in \mathcal{C}_{\beta^+}$ for any t.*

Proof of claim. On the one hand, $f_t(z_0) = g_{t-\lfloor t\rfloor}(g^{\lfloor t\rfloor}(w_1))$. Since $w_1 \notin \mathcal{C}_\alpha$, each $g^{\lfloor t\rfloor}(w_1)$ does not belong to $\mathcal{C}_\alpha$ either, as observed in Lemma 3.1. However, it does belong to $\mathbb{D}_0$ while $t < n_1$. Thus, $g_s(g^{\lfloor t\rfloor}(w_1)) \notin \mathcal{C}_{\beta^+}$ for every $s = t - \lfloor t\rfloor \in \mathbb{I}$. On the other hand, $f_t(\mathbf{1}) = g_{t-\lfloor t\rfloor}(g^{\lfloor t\rfloor}(v_1))$. Since $v_1 \in \mathrm{W^s}$, each $g^{\lfloor t\rfloor}(v_1)$ belongs to $\mathrm{W^s} \cap \mathbb{D}_0 \subset \mathcal{C}_{\beta^-} \cap \mathbb{D}_0$. Therefore, $g_s(g^{\lfloor t\rfloor}(v_1)) \in \mathcal{C}_{\beta^+}$ for every $s = t - \lfloor t\rfloor \in \mathbb{I}$, as claimed. □

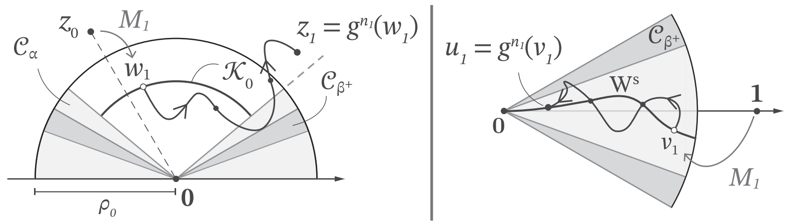

FIGURE 1. Promotion by f_t of two parallel processes: points in $\mathcal{K}_0$ are successively dragged out of $\mathbb{D}_0$ without entering the β^+-cone, while the images of **1** are dragged toward the origin over the stable manifold.

The above setting is pictured in Figure 1, where the points

$$z_1 := f_{n_1}(z_0) = g^{n_1}(w_1) \notin \mathcal{C}_\alpha \quad \text{and} \quad u_1 := f_{n_1}(\mathbf{1}) = g^{n_1}(v_1) \in \mathrm{W}^{\mathrm{s}}$$

are introduced.

By the choice of n_1, $|z_1| > \rho_0$. Let $r_2 = \rho_1/|z_1|$. Then, $r_2 z_1 \in \mathcal{K}_0$ and $r_2 u_1 \in \mathbb{D}_0 \cap \mathcal{C}_{\min\{\tau,\beta^-\}}$. In particular, $\mathrm{e}^{\mathrm{i}\,\tau_2}(r_2 u_1) \in \mathrm{W}^{\mathrm{s}}$ for some τ_2 such that $|\tau_2| \leq 2(\frac{1}{2}\min\{\tau, \beta^-\}) = \min\{\tau, \beta^-\}$.

CLAIM 3.5.3. *Define $M_2(z) := r_2 \mathrm{e}^{\mathrm{i}\,\tau_2} z$. Then, M_2 has the following properties:*

(i) $M_2 \in \mathrm{M\ddot{o}b}_2(\mathbb{S}^2)$;

(ii) $v_2 := M_2(u_1) \in \mathrm{W}^{\mathrm{s}}$;

(iii) $w_2 := M_2(z_1) \in \mathcal{K}_0$.

Proof of claim. Item (i) follows from the form of M_1 and item (ii) from the choice of τ_2. As for item (iii), notice that $r_2\, z_1$ does not belong to $\mathcal{C}_\alpha$—because it is obtained from z_1 through an homothety—but it does belong to $\mathbb{D}_0$. So Lemma 3.1 implies

$$\alpha < \theta(r_2 z_1) - |\tau_2| \leq \underbrace{\theta(r_2 z_1) + \tau_2}_{=\theta(w_2)} \leq \theta(r_2 z_1) + |\tau_2| < \pi - \alpha,$$

where it was used that $|\tau_2| \leq \tau$. This allows one to conclude that $w_2 \notin \mathcal{C}_\alpha$. Furthermore, $|w_2| = |r_2 z_1| = \rho_1$, establishing that $w_2 \in \mathcal{K}_0$. □

In particular, $n_2 := n_{w_2}$ is well defined. For $n_1 < t \leq n_1 + n_2$, consider now the expression $f_t = g_{t-\lfloor t \rfloor} \circ g^{\lfloor t \rfloor - n_1} \circ M_2 \circ f_{n_1}$. Arguments analogous to the previous ones imply $t \mapsto f_t$ to be continuous over the interval $(n_1, n_1 + n_2]$. Also, $f_t(z_0) \notin \mathcal{C}_{\beta^+}$ and $f_t(\mathbf{1}) \in \mathcal{C}_{\beta^+}$ for every t. In a similar fashion, we define inductively for $k \geq 0$:

$$f_t := \begin{cases} \mathrm{id} & \text{if } t = 0, \\ g_{t-\lfloor t \rfloor} \circ g^{\lfloor t \rfloor - N_k} \circ M_{k+1} \circ f_{N_k} & \text{on the interval } N_k < t \leq N_{k+1}, \end{cases} \tag{6}$$

where $N_0 = 0$, $N_k = \sum_{i=1}^{k} n_i$, and the numbers n_k and maps M_k satisfy:

- $M_{k+1} \in \mathrm{M\ddot{o}b}_2(\mathbb{S}^2)$ is a transformation of the form

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

mapping $z_k := f_{N_k}(z_0) \notin \overline{\mathbb{D}_0}$ to a point $w_{k+1} \in \mathcal{K}_0$ and $u_k := f_{N_k}(\mathbf{1}) \in \mathrm{W}^{\mathrm{s}}$ to a point $v_{k+1} \in \mathrm{W}^{\mathrm{s}}$, via an homothety of scaling factor $r_{k+1} = \rho_1/|z_k| < 1$ and a rotation of angle $|\tau_{k+1}| \leq \min\{\tau, \beta^-\}$;
- $n_k := n_{w_k} \leq n_0$ is given as in equation (5).

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) $f_t(z_0) \notin \mathcal{C}_{\beta^+}$ and $f_t(\mathbf{1}) \in \mathcal{C}_{\beta^+}$ for every $t \geq 0$.

CLAIM 3.5.4. *For $t \geq 0$, let*

$$I_t^{z_0} := \hat{M}[f_t(\mathbf{1})] \circ f_t,$$

where $\hat{M}[\cdot]$ is as in Definition 3.4. Then, $(I_t^{z_0})_{t\geq 0}$ is an $\mathcal{I}G_3$-isotopy.

Proof of claim. Since $f_t \in G_2$ for every $t \geq 0$, it is clear that $I_t^{z_0} \in G_3$ for every $t \geq 0$. One must see that $t \mapsto I_t^{z_0}$ defines a continuous curve of homeomorphisms. It 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 $0 < h < 1$,

$$I_{N_k+h}^{z_0} = \hat{M}[f_{N_k+h}(\mathbf{1})] \circ f_{N_k+h} = \hat{M}[g_h(v_{k+1})] \circ g_h \circ M_{k+1} \circ f_{N_k}.$$

However, notice that

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

is a Möbius transformation fixing the poles and mapping $g_h(v_{k+1})$ to $\mathbf{1}$. By sharp 3-transitivity, it must actually be $\hat{M}[g_h(v_{k+1})]$. However, since $g_h \to \mathrm{id}$ 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,

$$\begin{aligned} I_{N_k+h}^{z_0} \to \hat{M}[u_k] \circ M_{k+1}^{-1} \circ g_0 \circ M_{k+1} \circ \ f_{N_k} &= \hat{M}[u_k] \circ f_{N_k} \\ &= \hat{M}[f_{N_k}(\mathbf{1})] \circ f_{N_k} = I_{N_k}^{z_0}. \end{aligned}$$ □

CLAIM 3.5.5. $\gamma_{I^{z_0}}(z_0) \cap \Gamma = \emptyset$.

Proof of claim. Since $f_t(\mathbf{1}) \in \mathcal{C}_{\beta^+}$ for every $t \geq 0$, the transformation $\hat{M}[f_t(\mathbf{1})]$ may be explicitly written as $\hat{M}[f_t(\mathbf{1})](z) = |f_t(\mathbf{1})|^{-1}\, \mathrm{e}^{\mathrm{i}\,\psi} z$, where $|\psi| \leq \beta^+$. As we also know that $f_t(z_0) \notin \mathcal{C}_{\beta^+}$ for every $t \geq 0$, $\hat{M}[f_t(\mathbf{1})] \circ f_t(z_0)$ remains on the upper half-plane without ever touching the x-axis. This establishes **(1)**. □

CLAIM 3.5.6. *$f_t(\mathbf{1}) \to \mathbf{0}$ as $t \to +\infty$.*

Proof of claim. Further shrinking $\mathbb{D}_0$ if necessary, we may assume that $|g(z)| < |z|$ for every $z \in \mathrm{W}^{\mathrm{s}}$. Then, since $|u_k| = |g^{n_k}(v_k)| < |v_k|$ and $|v_{k+1}| = r_{k+1}|u_k|$, we have

$$|v_{k+1}| < (\rho_1/\rho_0)^k, \tag{7}$$

for $r_{k+1} = \rho_1/|z_k| \leq \rho_1/\rho_0$ and $v_1 \in \mathbb{D}_0$ has a norm not greater than one.

https://doi.org/10.1017/etds.2023.32 Published online by Cambridge University Press

Now, on each interval $(N_k, N_{k+1}]$, the expression $f_t(\mathbf{1}) = g_{t-\lfloor t\rfloor} \circ g^{\lfloor t\rfloor - N_k}(v_{k+1})$ holds. As t ranges through this interval, the quantity $t - \lfloor t\rfloor$ ranges over the interval [0, 1] and the quantity $\lfloor t\rfloor - N_k$ ranges through $\{0, \ldots, n_{k+1}\} \subset \{0, \ldots, n_0\}$. Thus,

$$\{f_t(\mathbf{1}) : N_k < t \leq N_{k+1}\} \subset \{g_s(g^i(v_{k+1})) : s \in \mathbb{I} \text{ and } 1 \leq i \leq n_0\}. \tag{8}$$

By the condition on the stable manifold, $|g^i(v_{k+1})| < |v_{k+1}|$ for each $1 \leq i \leq n_0$. The inclusion above—along with equation (7)—thus implies:

$$\{f_t(\mathbf{1}) : N_k < t \leq N_{k+1}\} \subset \{g_s(z) : s \in \mathbb{I} \text{ and } z \in \mathbb{D}_{(\rho_1/\rho_0)^k}(\mathbf{0})\}. \tag{9}$$

Since $(\rho_1/\rho_0) < 1$, the sets at the right-hand side of equation (9) are nested and decreasing with $k \in \mathbb{N}$. In particular,

$$\sup\{|f_t(\mathbf{1})| : t > N_k\} \leq \max\{|g_s(z)| : s \in \mathbb{I} \text{ and } z \in \mathbb{D}_{(\rho_1/\rho_0)^k}(\mathbf{0})\}. \tag{10}$$

Lastly, let $\varepsilon > 0$ be given and fix $\eta > 0$ such that $|z| < \eta$ implies $|g_s(z)| < \varepsilon$ for every $s \in \mathbb{I}$. Then, if $k_0 \in \mathbb{N}$ is so large that $(\rho_1/\rho_0)^{k_0} < \eta$, equation (10) yields $|f_t(\mathbf{1})| < \varepsilon$ whenever $t > N_{k_0}$. □

We may now finish the proof. We know that $f_t(z_0) = g_{t-\lfloor t\rfloor} \circ g^{\lfloor t\rfloor - N_k}(w_{k+1})$ on each interval $(N_k, N_{k+1}]$, where $w_k \in \mathcal{K}_0$ for every $k \in \mathbb{N}$. Observing the same ranges as in the proof of the previous claim, we see that, for every $t \geq 0$,

$$f_t(z_0) \in \{g_s \circ g^i(z) : s \in \mathbb{I},\ 1 \leq i \leq n_0 \text{ and } z \in \mathcal{K}_0\} := \mathcal{K}.$$

However, $\mathcal{K}$ is a compact set bounded away from $\mathbf{0}$. Since $f_t(\mathbf{1}) \to \mathbf{0}$ as $t \to +\infty$ and $I^{z_0}(z_0) \in \hat{M}[f_t(\mathbf{1})](\mathcal{K})$, the remark in equation (4) implies **(2)**. □

COROLLARY 3.6. *Let $G \subset \mathrm{Diff}^1(\mathbb{S}^2)$ be a group properly extending* $\mathrm{Möb}(\mathbb{S}^2)$. *Then, for each given point $z_0 \notin \Gamma$ and each pair of points $a, b \in \{\mathbf{0}, \mathbf{1}, \infty\}$, there exists a full-time $\mathcal{I}G_3$-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 Γ;*

(2) *the α and ω 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 $a, b \in \{\mathbf{0}, \mathbf{1}, \infty\}$ 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 Γ invariant, so neither $T_{\infty b}(z_0)$ nor $T_{\infty a}(z_0)$ lie on Γ. The Fundamental Lemma then yields $\mathcal{I}G_3$-isotopies $\left(\, I_t^{T_{\infty b}(z_0)}\,\right)_{t\geq 0}$ and $\left(\, I_t^{T_{\infty a}(z_0)}\,\right)_{t\geq 0}$ as described therein, from which we define $I^{z_0}_{ab} : \mathbb{R} \times \mathbb{S}^2 \to \mathbb{S}^2$ as

$$I^{z_0}_{ab}(t,\ z) = \begin{cases} T_{\infty b} \circ I_t^{T_{\infty b}(z_0)} \circ T_{\infty b}\ (z) & \text{if } t \geq 0, \\ T_{\infty a} \circ I_{-t}^{T_{\infty a}(z_0)} \circ T_{\infty a}\ (z) & \text{if } t \leq 0. \end{cases} \tag{11}$$

The statements then follow by noting that $t \mapsto -t$ turns the ω into the α-limit. □

4. *A crossing lemma*

Let us to now analyze the behavior of Γ under the action of G_0. Before doing so, we agree that, given three distinct points a, b, c, we denote by $\hat{M}[a, b, c]$ the unique Möbius transformation mapping a to $\mathbf{0}$, b to $\mathbf{1}$, and c to $\boldsymbol{\infty}$. If the points are all finite and non-zero, this map can be explicitly written as a certain cross ratio, from which it is seen that $(a, b, c) \mapsto \hat{M}[a, b, c]$ is continuous.

LEMMA 4.1. *Let $G \subset \mathrm{Diff}^1(\mathbb{S}^2)$ be a group properly extending* $\mathrm{M\ddot{o}b}(\mathbb{S}^2)$. *Then, there exists an $\mathcal{I}G_3$-isotopy $(h_t)_{t\in\mathbb{I}}$ such that $|\, h_1(\Gamma) \cap \Gamma \,| = 4$.*

Proof. The Extension Lemma yields an $\mathcal{I}G_2$-isotopy $(g_t)_{t\in\mathbb{I}}$ such that $g := g_1$ has a hyperbolic saddle fixed point at the origin. Fix any sufficiently small $0 < r_0 < 1$ such that $|\, S_{r_0} \cap g(S_{r_0}) \,| = 4$, and let $a, b, c, d \in S_{r_0}$ be four consecutive points in the usual anticlockwise cyclic order such that $S_{r_0} \cap g(S_{r_0}) = \{g(a), g(b), g(c), g(d)\}$. Then, we may consider the unique $M_0 \in \mathrm{M\ddot{o}b}(\mathbb{S}^2)$ mapping the ordered triple $(\mathbf{0}, \mathbf{1}, \infty)$ onto (a, b, c). Clearly, $M_0 = \hat{M}[a, b, c]^{-1}$ and $M_0(\Gamma) = S_{r_0}$. By defining

$$h_t := \hat{M}[g_t(a), g_t(b), g_t(c)] \circ g_t \circ M_0 , \quad 0 \le t \le 1 ,$$

we obtain an $\mathcal{I}G_3$-isotopy $(h_t)_{t\in\mathbb{I}}$. Since $\hat{M}[g(a), g(b), g(c)](S_{r_0}) = \Gamma$ and Möbius transformations are bijections of the sphere,

$$\begin{aligned} |h_1(\Gamma) \cap \Gamma| &= |\hat{M}[g(a), g(b), g(c)] \circ g \circ M_0\,(\Gamma) \ \cap\ \hat{M}[g(a), g(b), g(c)](S_{r_0})| \\ &= |\hat{M}[g(a), g(b), g(c)] \circ g(M_0(\Gamma)) \ \cap\ \hat{M}[g(a), g(b), g(c)](S_{r_0})| \\ &= |\hat{M}[g(a), g(b), g(c)] \circ g(S_{r_0}) \ \cap\ \hat{M}[g(a), g(b), g(c)](S_{r_0})| \\ &= |\, S_{r_0} \cap g(S_{r_0}) \,| = 4. \end{aligned}$$

□

COROLLARY 4.2. *Let $G \subset \mathrm{Diff}^1(\mathbb{S}^2)$ be a group properly extending* $\mathrm{M\ddot{o}b}(\mathbb{S}^2)$. *Then, there exist $z_0 \in \Gamma \setminus \{\mathbf{0}, \mathbf{1}, \infty\}$ and an $\mathcal{I}G_3$-isotopy $(h_t)_{t\in\mathbb{I}}$ such that $h_1(z_0) \notin \Gamma$.*

Proof. In the language of Lemma 4.1, let $h_1(\Gamma) \cap \Gamma = \{\mathbf{0}, \mathbf{1}, \infty, w_0\}$. It suffices to take any $z_0 \in \Gamma \setminus \{\mathbf{0}, \mathbf{1}, \infty, h_1^{-1}(w_0)\}$. □

Finite points a, b on the meridian Γ 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 $a < b$ with respect to this ordering:

- $[a, b]$ denotes the arc of the meridian Γ with endpoints a, b and not containing $\boldsymbol{\infty}$, corresponding to the compact interval of the x-axis with the associated endpoints;
- $[b, a]$ denotes the arc of the meridian Γ with endpoints a, b and containing $\boldsymbol{\infty}$, 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 Γ are defined accordingly by deletion of the suitable endpoints from the corresponding closed arcs.

LEMMA 4.3. (Crossing Lemma) *Let $G \subset \mathrm{Diff}^1(\mathbb{S}^2)$ be a group properly extending* $\mathrm{M\ddot{o}b}(\mathbb{S}^2)$. *Then, there exist a point $\hat{z}$ on the open segment $(\mathbf{0}, \mathbf{1})$ of Γ and an $\mathcal{I}G_3$-isotopy $(J_t)_{t\in[-1,1]}$ such that:*

(1) *the trajectory of $\hat{z}$ under J only intersects Γ on the arc $(\mathbf{0}, \mathbf{1})$;*

(2) *$J_{-1}(\hat{z}) \in \mathcal{H}^-$ and $J_1(\hat{z}) \in \mathcal{H}^+$.*

https://doi.org/10.1017/etds.2023.32 Published online by Cambridge University Press

Proof. By Corollary 4.2, we may fix an $\mathcal{I}G_3$-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 $h_1(z_0) \in \mathcal{H}^+$ and choose $u_0 \in \mathcal{H}^-$ near z_0 such that $h_1(u_0) \in \mathcal{H}^+$ as well. Consider the continuous path defined as $\gamma(t) = h_t(u_0)$. The sets $\gamma^{-1}(\mathcal{H}^-)$, $\gamma^{-1}(\Gamma)$, and $\gamma^{-1}(\mathcal{H}^+)$ form a partition of $\mathbb{I}$. Let

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

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

$$s \in \mathbb{I} \quad \text{and} \quad |s - t^\pm| < \delta \quad \text{imply } d(\gamma(s), \gamma(t^\pm)) < \rho. \tag{12}$$

Then, we may fix $t_1 \in \gamma^{-1}(\mathcal{H}^-) \cap (t^- - \delta, t^-)$ and $t_2 \in \gamma^{-1}(\mathcal{H}^+) \cap (t^+, t^+ + \delta)$. Let $\sigma : [-1, 1] \to [t_1, t_2]$ be any increasing bijection such that $\sigma(0)$ is the midpoint of the interval $[t^-, t^+]$. Defining $(\tilde{J}_t)_{t\in[-1,1]}$ by $\tilde{J}_t(z) = h_{\sigma(t)} \circ h^{-1}_{\sigma(0)}(z)$ yields an $\mathcal{I}G_3$-isotopy.

Letting $\tilde{z} = \gamma(\sigma(0))$, we have $\tilde{J}_1(\tilde{z}) \in \mathcal{H}^+$ and $\tilde{J}_{-1}(\tilde{z}) \in \mathcal{H}^-$. Furthermore, the trajectory of $\tilde{z}$ under $\tilde{J}$ is $\gamma_{\tilde{J}}(\tilde{z}) = \gamma([t^- - \delta^-, t^+ + \delta^+])$. In particular, it follows from the choice of ρ and from equation (12) that any point in $\gamma_{\tilde{J}}(\tilde{z}) \cap \Gamma$ lies on the same connected component of $\Gamma \setminus \{\mathbf{0}, \mathbf{1}, \infty\}$ as the segment $\gamma([t^-, t^+])$, which is precisely the component containing $\tilde{z}$.

Considering the cyclic order induced in Γ, let (a, b) be the only interval among $(\mathbf{0}, \mathbf{1})$, $(\mathbf{1}, \infty)$, and $(\infty, \mathbf{0})$ in which $\tilde{z}$ lies. Upon defining $\{c\} = \{\mathbf{0}, \mathbf{1}, \infty\} \setminus \{a, b\}$, the Möbius transformation $\hat{M}[a, b, c]$—which can be the identity—preserves this cyclic order, leaving the hemispheres invariant. Thus, we may take $\hat{z} = \hat{M}[a, b, c](\tilde{z})$ and $J_t = \hat{M}[a, b, c] \circ \tilde{J}_t \circ \hat{M}[a, b, c]^{-1}$. □

5. *Closing arguments*

5.1. *Proof of Theorem A.* From now on, a subgroup $G \subset \mathrm{Diff}^1(\mathbb{S}^2)$ properly extending $\mathrm{Möb}(\mathbb{S}^2)$ is fixed throughout, and we let the point $\hat{z} \in \Gamma$ and the $\mathcal{I}G_3$-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

$$\chi := \overline{\gamma_{I^{\hat{z}_-}}(\hat{z}_-) \cup \gamma_J(\hat{z}) \cup \gamma_{I^{\hat{z}_+}}(\hat{z}_+)}. \tag{13}$$

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, χ is a continuum, for it is the closure of the union of connected curves with points in common, and consists of adjoining $\{\infty\}$ to that union, as we suggest in Figure 2.

CLAIM 5.1.1. *The set χ separates $(\infty, \mathbf{0})$ and $(\mathbf{1}, \infty)$ in the following sense: for every path $\alpha : \mathbb{I} \to \mathbb{S}^2$ such that $\alpha(0) \in (\infty, \mathbf{0})$ and $\alpha(1) \in (\mathbf{1}, \infty)$, we have $\alpha(\mathbb{I}) \cap \chi \neq \emptyset$.*

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

https://doi.org/10.1017/etds.2023.32 Published online by Cambridge University Press

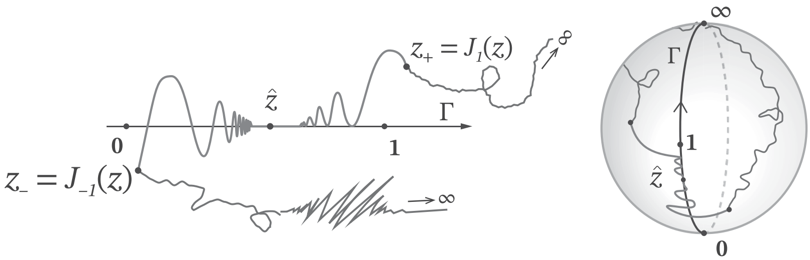

FIGURE 2. The set χ is constructed by gluing isotopy trajectories with points in common, one of which is bounded away from ∞ and two of which accumulate at ∞, and then taking their closure.

hypothesis made upon the endpoints of α, the following are well defined:

$$t^- := \sup\{t \in \mathbb{I} : \alpha(t) \in (.\infty, \mathbf{0}].\} \quad \text{and} \quad t^+ := \inf\{t \in (.t^-, 1]. : \alpha(t) \in [.\mathbf{1}, \infty).\}. \tag{14}$$

Continuity of α implies $\alpha(t^-) \in (\infty, \mathbf{0}]$ and $\alpha(t^+) \in [\mathbf{1}, \infty)$. In particular, $t^+ > t^-$, since these points are mapped to disjoint arcs. We further notice that any intersection between $\alpha((t^-, t^+))$ and Γ takes place in the open arc $(\mathbf{0}, \mathbf{1})$, by the very own definition of t^- and t^+. On the plane, we define the following continuous map, which is a line—that is, a simple and proper path:

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

Indeed, to see that ℓ is simple, notice first that since $\alpha(t^-) \in (\infty, \mathbf{0}]$, $\alpha(t^+) \in [\mathbf{1}, \infty)$ and $\hat{z} \in (\mathbf{0}, \mathbf{1})$ all lie in disjoint segments, both $(\alpha(t^-) - \hat{z})$ and $(\alpha(t^+) - \hat{z})$ are non-zero. Thus, the restrictions $\ell|_{t \le t^-}$ and $\ell|_{t \ge t^+}$ are injective. Furthermore, $\ell|_{(t^-, t^+)} = \alpha|_{(t^-, t^+)}$ is injective as well, for α was assumed simple.

Next, suppose for the sake of contradiction that $\ell(\bar{t}) = \ell(\bar{s})$ for some $\bar{s} \le t^-$ and $t^- < \bar{t} < t^+$. Thinking of Γ as the x-axis, we have $\alpha(t^-) \le \mathbf{0} < \hat{z}$, from which it follows that $\alpha(t^-) - \hat{z}$ is negative. However, since $(1 - \bar{s})/(1 - t^-) \ge 1$, it must be the case that $\ell(\bar{s}) \le \alpha(t^-) \le \mathbf{0}$. In particular, $\ell(\bar{s}) \in \Gamma$, so $\ell(\bar{t}) \in \Gamma$ as well. However, then

$$\ell(\bar{s}) = \ell(\bar{t}) = \alpha(\bar{t}) > \mathbf{0} \ge \ell(\bar{s}),$$

which is a contradiction. Analogous reasoning shows that $\ell(\bar{s}) = \ell(\bar{t})$ cannot happen for $\bar{s} \ge t^+$ and $t^- < \bar{t} < t^+$ both holding simultaneously either. Clearly, ℓ is proper, for $|\ell(t)| \to \infty$ as $t \to \pm\infty$.

Due to celebrated Jordan–Schoenflies theorem, the orientation inherited by ℓ from $\mathbb{R}$, which coincides with the intrinsic orientation of Γ, divides the plane into two open and connected components, its *right* $\mathsf{R}(\ell)$ and its *left* $\mathsf{L}(\ell)$, plus their common boundary ℓ [13].

Consider the compact set $\alpha([t^-, t^+])$, and fix some closed disk $D \subset \mathbb{R}^2$ containing it. Then, it must be the case that $[\mathbf{0}, \mathbf{1}] \subset D$. Indeed, since D is convex, it must contain

https://doi.org/10.1017/etds.2023.32 Published online by Cambridge University Press

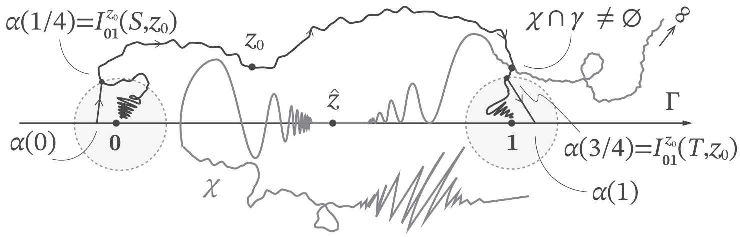

FIGURE 3. After leaving a compact neighborhood of $\mathbf{0}$ disjoint of χ and before entering a neighborhood of $\mathbf{1}$ disjoint from χ, the path of z_0 under $I_{\mathbf{01}}^{z_0}$ must cross the continuum χ.

the segment $[\alpha(t^-), \alpha(t^+)]$. However, from equation (14), we know that $\alpha(t^-) \leq \mathbf{0}$ and $\alpha(t^+) \geq \mathbf{1}$, which imply $[\mathbf{0}, \mathbf{1}] \subset [\alpha(t^-), \alpha(t^+)]$.

So, if we now consider the open set $\mathcal{O} := \mathbb{R}^2 \setminus D$, we see from the expression of ℓ that $\ell \cap \mathcal{O} = \Gamma \cap \mathcal{O}$. Also, ℓ traverses this intersection with the same orientation as Γ. These imply $\mathsf{L}(\ell) \cap \mathcal{O} = \mathsf{L}(\Gamma) \cap \mathcal{O} = \mathcal{H}^+ \cap \mathcal{O}$ and $\mathsf{R}(\ell) \cap \mathcal{O} = \mathsf{R}(\Gamma) \cap \mathcal{O} = \mathcal{H}^- \cap \mathcal{O}$. From the Fundamental Lemma, $\gamma_{I^{\hat{z}_+}}(\hat{z}_+)$ is fully contained within H^+ and accumulates at $\{\infty\}$. Since $\mathcal{O}$ is a neighborhood of ∞ on the sphere, it follows that $\gamma_{I^{\hat{z}_+}}(\hat{z}_+) \cap \mathsf{L}(\ell) \neq \emptyset$. Analogously, $\gamma_{I^{\hat{z}_-}}(\hat{z}_-) \cap \mathsf{R}(\ell) \neq \emptyset$. This translates to $\chi \cap \mathsf{L}(\ell) \neq \emptyset$ and $\chi \cap \mathsf{R}(\ell) \neq \emptyset$. Therefore, the line ℓ intersects the continuum χ.

Let $\bar{t} \in \mathbb{R}$ be such that $\ell(\bar{t}) \in \chi$. Then, it must be the case that $\bar{t} \in (t^-, t^+)$, for the Crossing Lemma and the Fundamental Lemma imply $(\chi \setminus \{\infty\}) \cap \Gamma \subset (\mathbf{0}, \mathbf{1})$. However, $t^- < \bar{t} < t^+$ means $\ell(\bar{t}) = \alpha(\bar{t})$, yielding an intersection between $\alpha(\mathbb{I})$ and χ, as claimed. □

Now, consider the equivalence relation $\sim_{G_3}$ described in Definition 1.3. Clearly, $\mathcal{A}_{G_3}(\mathbf{0}) = \{\mathbf{0}\}$, $\mathcal{A}_{G_3}(\mathbf{1}) = \{\mathbf{1}\}$ and $\mathcal{A}_{G_3}(\infty) = \{\infty\}$. Let us show that $z_0 \in \mathcal{A}_{G_3}(\hat{z})$ for any $z_0 \in \mathbb{S}^2 \setminus \{\mathbf{0}, \mathbf{1}, \infty\}$. By Corollary 4.2, it suffices to consider $z_0 \notin \Gamma$.

CLAIM 5.1.2. *If $z_0 \notin \Gamma$, then $z_0 \in \mathcal{A}_{G_3}(\hat{z})$.*

Proof. Let χ be as in equation (13). Fix $r > 0$ such that $\overline{\mathbb{D}}_r(\mathbf{0})$ and $\overline{\mathbb{D}}_r(\mathbf{1})$ are disjoint from χ, z_0, ∞, and also from each other. Consider the $\mathcal{I}G_3$-isotopy $I_{\mathbf{01}}^{z_0}$ yielded by Corollary 3.6. We encounter $S < 0$ *maximal* such that $I_{\mathbf{01}}^{z_0}(S, z_0) \in \partial\overline{\mathbb{D}}_r(\mathbf{0})$ and $T > 0$ *minimal* such that $I_{\mathbf{01}}^{z_0}(T, z_0) \in \partial\overline{\mathbb{D}}_r(\mathbf{1})$.

Let $\alpha : \mathbb{I} \to \mathbb{R}^2$ depart from a point $\alpha(0) \in (\infty, \mathbf{0}) \cap \overline{\mathbb{D}}_r(\mathbf{0})$; follow on a straight line until it reaches $\alpha(1/4) = I_{\mathbf{01}}^{z_0}(S, z_0)$ at the disk's boundary; follow z_0's isotopy path under $I_{\mathbf{01}}^{z_0}$ until it first reaches the boundary of the disk $\overline{\mathbb{D}}_r(\mathbf{1})$ at $\alpha(3/4) = I_{\mathbf{01}}^{z_0}(T, z_0)$; then move on a straight line until it reaches a point $\alpha(1) \in (\mathbf{1}, \infty) \cap \overline{\mathbb{D}}_r(\mathbf{1})$.

Claim 5.1.1 implies α to intersect $\chi \setminus \{\infty\}$. However, since the segments $\alpha([0, 1/4])$ and $\alpha([3/4, 1])$ are contained within disks disjoint from χ, we must have $\alpha(\,(1/4, 3/4)\,) \cap (\chi \setminus \{\infty\}) \neq \emptyset$. Since by construction $\alpha(\,(1/4, 3/4)\,) \subset \gamma_{I_{\mathbf{01}}^{z_0}}(z_0)$, Lemma 1.4 is readily seen to imply $z_0 \in \mathcal{A}_G(\hat{z})$. This process is conveyed in Figure 3. □

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 $z_0 := \hat{M}[a, b, c](d)$ and $w_0 := \hat{M}[p, q, r]^{-1}(s)$. Then, by Claim 5.1.2, both belong to $\mathcal{A}_{G_3}(\hat{z})$. This implies some

https://doi.org/10.1017/etds.2023.32 Published online by Cambridge University Press

$\mathcal{I}G_3$-isotopy $(f_t)_{t\in\mathbb{I}}$ to be such that $f_1(z_0) = w_0$. Since $\text{Möb}(\mathbb{S}^2)$ is path connected, in particular, $\hat{M}[p, q, r]^{-1} \circ f_1 \circ \hat{M}[a, b, c] \in G_0$ maps (a, b, c, d) onto (p, q, r, s). So, the arc 4-transitivity definition is satisfied.

5.2. *Discussion on Theorem B.* Given three distinct points p_0, p_1, p_2 on the plane, let us call the wedge of two circles based at p_0—each of them traversed once with contrary orientations, while leaving p_1 and p_2 in opposite components of their complements—a *prototypical figure-8*.

Given four distinguished points $P = \{p_0, \ldots, p_3\}$ on the sphere, a closed loop $\alpha : \mathbb{I} \to \mathbb{S}^2$ based at p_0 will be referred to as *topological figure-8* (relative to P) if, when p_3 is placed at infinity, α projects onto a path belonging to the same homotopy class in the fundamental group $\pi_1(\mathbb{R}^2 \setminus \{p_1, p_2\}; p_0)$ as a prototypical figure-8. If $G \subset \text{Diff}^1(\mathbb{S}^2)$ is a group properly extending $\text{Möb}(\mathbb{S}^2)$, our goal is to realize such an object as the trajectory of a point under an isotopy in G.

To do so, consider χ as defined in equation (13). We may construct an $\mathcal{I}G_3$ isotopy $(K_t)_{t\in\mathbb{R}}$ such that $\chi \setminus \{\infty\}$ is realized as the trajectory of $\hat{z}$ under K. From it, we may further obtain an $\mathcal{I}G_3$ isotopy $L_t := T_{\mathbf{0}\infty} \circ K_t \circ T_{\mathbf{0}\infty}$ and a point $\hat{y} := T_{\mathbf{0}\infty}(\hat{z}) \in (\mathbf{1}, \infty)$ with:

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

CLAIM 5.2.1. *There exist a point $\hat{w} \in (\mathbf{0}, \mathbf{1})$ and an $\mathcal{I}G_3$ isotopy $(\varphi_t)_{t\in\mathbb{I}}$ such that $\gamma_\varphi(\hat{w}) \cong \xi_1 \operatorname{rel}\{\mathbf{0}, \mathbf{1}, \infty\}$, where $\cong$ denotes fixed endpoints homotopy, and ξ_1 is a circle passing through $\hat{w}$—traversed once clockwise while leaving $\mathbf{1}$ on its right and both $\mathbf{0}, \infty$ on its left.*

Proof. Let $(L_t)_{t\in\mathbb{R}}$ and $\hat{y}$ be as in the previous discussion, and consider $\lambda(t) := L_t(\hat{y})$. We may fix

$$\tilde{t} := \min\{t \in \mathbb{R} : \lambda(t) \in \Gamma\} > -1.$$

Item (i) above yields $\lambda(t)$ in the same connected component of χ^{c} as $\mathbf{0}$ for $t \ll 0$, while item (iii) yields $\lambda(\tilde{t}) \in (\mathbf{1}, \infty)$. Claim 5.1.1 then allows us to consider $t^- := \max\{t \leq \tilde{t} : \lambda(t) \in \chi\}$ so that, by item (ii), $L_{t^-}(\hat{y}) \in \chi \cap \mathcal{H}^+$.

Proceeding analogously for times greater than $\max\{t \in \mathbb{R} : \lambda(t) \in \Gamma\}$, but using the parts of items (i) and (ii) concerning times ≥ 1, we obtain $t^+ > t^-$ such that $L_{t^+}(\hat{y}) \in \chi \cap \mathcal{H}^-$.

This means that $L_{t^-}(\hat{y}) = K_a(\hat{z})$ and $L_{t^+}(\hat{y}) = K_b(\hat{z})$, for some $a, b \in \mathbb{R}$. In particular, there exists an intermediate c for which $K_c(\hat{z}) \in \Gamma$. Upon defining $\hat{w} := K_c(\hat{z})$, we actually know from the Crossing Lemma that $\hat{w} \in (\mathbf{0}, \mathbf{1})$. Lastly, we define the $\mathcal{I}G_3$-isotopy $(\varphi_t)_{t\in\mathbb{I}}$ as:

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

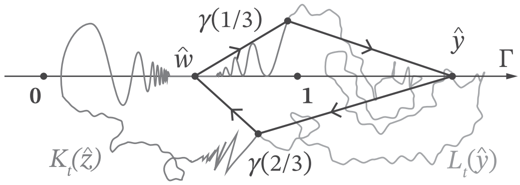

FIGURE 4. The trajectory of $\hat{w}$ under φ is homotopic to a clockwise loop turning around **1**, but not **0**, with fixed endpoint $\hat{w}$.

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 γ, we have four points distinguished during the construction:

- the starting and ending point $\hat{w} = \gamma(0) = \gamma(1) \in (\mathbf{0}, \mathbf{1})$;
- the point $\gamma(1/3) = K_b(\hat{z}) = L_{t^-}(\hat{y}) \in \mathcal{H}^+$;
- the point $\hat{y} \in (\mathbf{1}, \infty)$, which is of the form $\hat{y} = \gamma(\bar{s})$, for some $1/3 < \bar{s} < 2/3$; and
- the point $\gamma(2/3) = L_{t^+}(\hat{y}) = K_a(\hat{z}) \in \mathcal{H}^-$.

We also know that $\gamma|_{[0,1/3]}$ and $\gamma|_{[2/3,1]}$ only intersect Γ at $(\mathbf{0}, \mathbf{1})$, and that $\gamma|_{[1/3,\bar{s}]}$ and $\gamma|_{[\bar{s},2/3]}$ only intersect Γ at $(\mathbf{1}, \infty)$. These suffice to show that γ is homotopic on the plane—with fixed basepoint $\hat{w}$ and relative to $\{\mathbf{0}, \mathbf{1}\}$—to a simple closed curve (say, polygonal) turning once clockwise around **1** and leaving **0** outside of it, as suggested by Figure 4. This amounts to the claimed statement. □

We can now complete the construction. If $\hat{w}' := T_{\mathbf{01}}(\hat{w})$, Theorem A yields h in G_3 such that $\hat{w}' = h(\hat{w})$. Let us consider the $\mathcal{I}G_3$-isotopy

$$\psi_t := h^{-1} \circ T_{\mathbf{01}} \circ \varphi_t \circ T_{\mathbf{01}} \circ h.$$

The trajectory of $\hat{w}$ under ψ is described by the curve $\eta(s) = h^{-1}(T_{\mathbf{01}} \circ \gamma(s))$, where γ is as in the proof of Claim 5.2.1. Since $T_{\mathbf{01}}$ leaves $\{\mathbf{0}, \mathbf{1}, \infty\}$ invariant, $T_{\mathbf{01}} \circ \gamma \cong T_{\mathbf{01}} \circ \xi_1$ rel$\{\mathbf{0}, \mathbf{1}, \infty\}$. However, $T_{\mathbf{01}} \circ \xi_1$ is also a circle—traversed once clockwise, while leaving **0** on its right and both $\mathbf{1}, \infty$ on its left. Thus, since $h \in G_3$, $\xi_2 = h^{-1} \circ T_{\mathbf{01}} \circ \xi_1$ is a clockwise simple loop based at $\hat{w}$, leaving **0** on its right and both $\mathbf{1}, \infty$ on its left. Also, $\eta \cong \xi_2$ rel$\{\mathbf{0}, \mathbf{1}, \infty\}$. If we now define the $\mathcal{I}G_3$-isotopy as

$$F_t := \begin{cases} \varphi_{1-2t} \circ \varphi_1^{-1} & \text{if } 0 \le t \le 1/2, \\ \psi_{2t-1} \circ \varphi_1^{-1} & \text{if } 1/2 \le t \le 1, \end{cases}$$

then the trajectory of $\hat{w}$ under the isotopy $(F_t)_{t\in\mathbb{I}}$ is a topological figure-8 relative to the set $P = \{\hat{w}, \mathbf{0}, \mathbf{1}, \infty\}$, which is fixed by the terminal diffeomorphism $f = F_1$.

With respect to f, we now evoke the Nielsen–Thurston classification theorem, as presented in [11]. It takes as input the homeomorphism f and the set P, yielding:

- a homeomorphism $\Phi : \mathbb{S}^2 \to \mathbb{S}^2$ such that f and Φ are isotopic relative to P; and
- a (possibly empty) system of f-invariant closed simple loops $\alpha_1, \ldots, \alpha_r$.

Such *reducing curves* come equipped with disjoint tubular neighborhoods V_i, not intersecting P, such that the connected components of $\mathbb{S}^2 \setminus \bigcup_{i=1}^r V_i$ group into invariant cycles, restricted to which either Φ is of *finite order*—meaning some power of it equals the identity—or Φ is *pseudo-Anosov* relative to P.

Briefly, pseudo-Anosov means that Φ admits a pair of invariant and measured transverse foliations, one of which is expanded with ratio $\beta > 1$ and the other of which is contracted with ratio β^{-1}. *Relative to P* means that the points in P are kept fixed under Φ, and manifest as one-prong singularities of the foliations. Such maps are known to have strictly positive topological entropy $h_{\text{top}}(\Phi) = \log \beta > 0$.

Upon interpreting points in P as punctures (as implied by [11] and [3, §7.6]), each connected component of $\mathbb{S}^2 \setminus \bigcup_{i=1}^{r} V_i$ must have negative Euler characteristic. Thus, if the reducing system of curves is non-empty, $\hat{w}$ must be enclosed by some α_i along with some other reference point $a \in \{\mathbf{0}, \mathbf{1}, \infty\}$.

The key to our argument is that, if α is a simple loop enclosing $\hat{w}$ and a, then its iterates $f^m(\alpha)$ cannot be freely homotopic to α relative to P, allowing us to discard the finite-order cases. This is a property known to hold, and which can be intuitively seen by sliding any such loop along the figure-8. A reference discussion in terms of the action of Dehn half-twists may be found in of [2, Ch. 15].

In short, Φ must be a pseudo-Anosov map. However, the behavior of mappings isotopic to pseudo-Anosov homeomorphisms is described in [7]. More specifically, Theorem 2 therein implies the statement of Theorem B, and also $h_{\text{top}}(f) \geq h_{\text{top}}(\Phi)$. Thus, f must have positive topological entropy as well.

6. *A characterization of the conformal group*

Before proving Theorem C, we establish two auxiliary lemmas. The first is a property known *a priori* to be held by the actual $\text{Möb}(\mathbb{S}^2)$. The second establishes conformality at the poles in the subgroup G_2. In what follows, recall that by *homogeneous*, we mean a subgroup of $\text{Homeo}(\mathbb{S}^2)$ containing $\text{Rot}(\mathbb{S}^2)$.

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

Proof. A map $g \in G_2$ 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 γ, let $\lambda := g(\gamma)$. Then, λ is a Jordan loop, containing the origin in its interior. If λ is *not* a circle, it contains distinct points p_m and p_M of minimum and maximum norm, respectively. For each polar angle $0 \leq \theta < 2\pi$, the semiradius $\mathbf{r}_\theta := \{te^{\mathrm{i}\theta} : t \geq 0\}$ intersects λ in a compact set λ_θ. For some θ_0, it must be the case that $|p| < |p_M|$ for every $p \in \lambda_{\theta_0}$.

Now, if θ_M is such that $p_M \in \lambda_{\theta_M}$, let R be a planar rotation mapping the semiradius $\mathbf{r}_{\theta_M}$ onto the semiradius $\mathbf{r}_{\theta_0}$. Then, $R(p_M) \in R(\lambda)$, but $R(p_M) \in \text{ext}\lambda$, where int and ext are used in the Jordan curve theorem sense, as the bounded and unbounded open connected components of λ^{c}, sharing λ as their common boundary. 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 $R(\lambda) \cap \overline{\text{int}\lambda} \neq \emptyset$, and $R(\lambda) \cap \lambda \neq \emptyset$ follows.

Therefore, we may obtain $p, q \in \gamma$ such that $R(g(p)) = g(q)$. Since p, q lie on the same circle, there exists a planar rotation U such that $p = U(q)$. However 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

https://doi.org/10.1017/etds.2023.32 Published online by Cambridge University Press

identity. Since rotations leave γ invariant, this implies $R(\lambda) = \lambda$, which is a contradiction. Therefore, λ has to be a circle. $\square$

LEMMA 6.2. *Let $G \subset \mathrm{Diff}^1(\mathbb{S}^2)$ be a sharply 3-transitive homogeneous group of diffeomorphisms. Then, every $g \in G_2$ is conformal at the poles.*

Proof. By Lemma 6.1, it 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 $A = \mathrm{D}g(\mathbf{0})$ is a linear isomorphism, so we may fix z_m and z_M respectively minimizing and maximizing $|Az|$ over $\mathbb{S}^1$. Hypothesis then yields:

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

Therefore, $A(\mathbb{S}^1)$ is a circle. Since A preserves orientation, this is enough to infer that it is a conformal matrix. In other words, g is conformal at $\mathbf{0}$. $\square$

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 obtain $h_\delta \in G$ such that h_δ fixes $\mathbf{0}$, but not ∞, and $\mathrm{D}h_\delta(\mathbf{0})$ is δ-close to id.

Indeed, by 2-transitivity, we may fix $h \in G$ such that $h(\mathbf{0}) = \mathbf{0}$ and $h(\infty) = \mathbf{1}$. We then let $h_\delta := h^{-1} \circ R_t \circ h$, for a sufficiently small t. This might seem like an underuse of the 3-transitivity hypothesis. However, as it turns out, a result from [9] implies any 2-transitive homogeneous group to actually be 3-transitive.

6.1. *Proof of Theorem C.* Assume—for the sake of contradiction—that G contains a non-conformal map g. By precomposing and postcomposing with suitable rotations, it may be assumed that g fixes $\mathbf{0}$ and that $A = \mathrm{D}g(\mathbf{0})$ is a non-conformal matrix. This means that the angle α between some pair (u_0, v_0) of unit vectors is different from the angle β between their images. Given $\varepsilon = |\beta - \alpha|/2$, fix $\delta > 0$ such that $|u - w| < \delta$ and $|v - z| < \delta$ on $\mathbb{S}^1$ imply $|\mathrm{ang}(u, v) - \mathrm{ang}(w, z)| < \varepsilon$. By the prior discussion, we may fix $h_\delta \in G$ such that:

- $h_\delta(\mathbf{0}) = \mathbf{0}$;
- $h_\delta^{-1}(\infty) := p_\delta \neq \infty$;
- $\mathrm{D}h_\delta(\mathbf{0})$ is $\delta/2$-close to id.

Notice that, by Lemma 6.2, g cannot fix ∞, or it would be conformal at the South Pole. Therefore, the 3-transitivity of G yields $f_\delta \in G_2$ such that $f_\delta(g(\infty)) = p_\delta$. From Lemma 6.2, $\mathrm{D}f_\delta(\mathbf{0})$ is conformal. Letting $\tilde{g} := h_\delta \circ f_\delta \circ g$, also $\mathrm{D}\tilde{g}(\mathbf{0})$ is conformal. By the chain rule, $\mathrm{D}\tilde{g}(\mathbf{0}) = \mathrm{D}h_\delta(\mathbf{0})\mathrm{D}f_\delta(\mathbf{0})A$. However, the choices of ε and δ imply the right-hand side of this expression not to preserve the angle between u_0 and v_0, which is a contradiction. Thus, such non-conformal $g \in G$ cannot exist, and $G \subset \text{Möb}(\mathbb{S}^2)$. Equality then follows from the sharp 3-transitivity of $\text{Möb}(\mathbb{S}^2)$.

Acknowledgements. The authors are grateful to an anonymous reviewer for diligent reading and thoughtful comments that helped make the exposition clearer. U.L. would

https://doi.org/10.1017/etds.2023.32 Published online by Cambridge University Press

also like to thank Prof. S. Alvarez at Udelar for his genuine interest in Lemma 3.5. U.L. was partially supported by CNPq, Grant No. 159129/2015-0. F.T. was supported by Fapesp and CNPq-Brazil.

REFERENCES

[1] A. Banyaga. *The Structure of Classical Diffeomorphism Groups (Mathematics and Its Applications, 400)*, 1st edn. Springer, New York, NY, 1997.

[2] B. Farb and D. Margalit. *A Primer on Mapping Class Groups (Princeton Mathematical Series, 49)*. Princeton University Press, Princeton, NJ, 2011.

[3] J. Franks and M. Misiurewicz. Topological methods in dynamics. *Handbook of Dynamical Systems*. Eds. B. Hasselblat and A. Katok. Elsevier, Amsterdam, 2002. Ch. 7, pp. 547–598.

[4] D. Gabai. Convergence groups are Fuchsian groups. *Ann. of Math. (2)* **136**(3) (1992), 447–510.

[5] É. Ghys. Groups acting on the circle. *Enseign. Math. T* **47**(1–2) (2001), 329–407.

[6] J. Giblin and V. Markovic. Classification of continuously transitive circle groups. *Geom. Topol.* **10**(3) (2006), 1319–1346.

[7] M. Handel. Global shadowing of pseudo-Anosov homeomorphisms. *Ergod. Th. & Dynam. Sys.* **5**(3) (1985), 373–377.

[8] B. Kolev. Sous-groupes compacts d'homéomorphismes de la sphère. *Enseign. Math.* **52**(3–4) (2006), 193–214.

[9] F. Kwakkel and F. Tal. Homogeneous transformation groups of the sphere. *Preprint*, 2015, arXiv:1309.0179v2.

[10] F. Le Roux. On closed subgroups of the group of homeomorphisms of a manifold. *J. Éc. polytech. Math.* **1** (2014), 147–159.

[11] T. Matsuoka. Periodic points and braid theory. *Handbook of Topological Fixed Point Theory*. Eds. R. F. Brown, M. Furi, L. Górniewicz and B. Jiang. Springer-Verlag, Dordrecht, 2005, pp. 171–216.

[12] T. Needham. *Visual Complex Analysis*. Oxford University Press, Oxford, 1999.

[13] L. C. Siebenmann. The Osgood–Schoenflies theorem revisited. *Russian Math. Surveys* **60**(4) (2005), 645–672.