On proper extensions of the conformal group

Ulisses Lakatos de Mello

Tese apresentada ao Instituto de Matemática
e Estatística da Universidade de São Paulo
para a obtenção do título de Doutor em Ciências

Programa: Matemática Aplicada
Orientador: Fábio Armando Tal

Durante parte do desenvolvimento deste trabalho,
o autor recebeu auxílio financeiro do CNPq

São Paulo, junho de 2022

On proper extensions of the conformal group

Ulisses Lakatos de Mello

Esta tese contém as correções e alterações sugeridas pela comissão julgadora durante a defesa, realizada em 15 de julho de 2022.

Comissão julgadora:

Prof. Dr. Alejandro Kocsard - IME / UFF
Prof. Dr. Alejandro Passeggi - Udelar
Prof. Dr. Bruno Santiago - IME / UFF
Prof. Dr. Fábio Armando Tal (orientador) - IME / USP
Prof. Dr. Sébastien Alvarez - Udelar

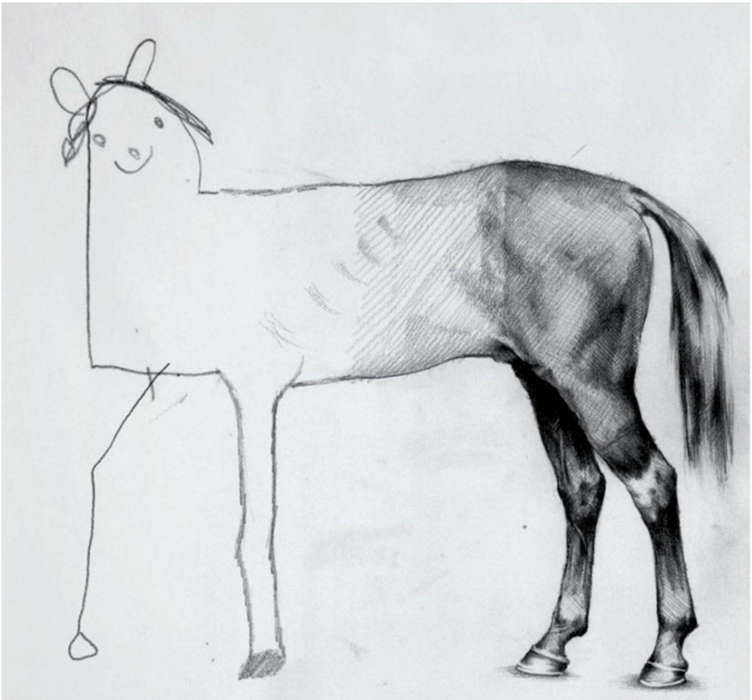
Ficha catalográfica elaborada com dados inseridos pelo(a) autor(a)
Biblioteca Carlos Benjamin de Lyra
Instituto de Matemática e Estatística
Universidade de São Paulo

Lakatos de Mello, Ulisses
On proper extensions of the conformal group / Ulisses Lakatos de Mello; orientador, Fábio Armando Tal. - São Paulo, 2022.
101 f.: il.

Tese (Doutorado) - Programa de Pós-Graduação em Matemática Aplicada / Instituto de Matemática e Estatística / Universidade de São Paulo.
Bibliografia
Versão corrigida

1. SISTEMAS DINÂMICOS. 2. VARIEDADES DE DIMENSÃO BAIXA. 3. GRUPOS TOPOLÓGICOS. 4. TOPOLOGIA GEOMÉTRICA. I. Tal, Fábio Armando. II. Título.

Bibliotecárias do Serviço de Informação e Biblioteca
Carlos Benjamin de Lyra do IME-USP, responsáveis pela
estrutura de catalogação da publicação de acordo com a AACR2:
Maria Lúcia Ribeiro CRB-8/2766; Stela do Nascimento Madruga CRB 8/7534.



Literally every human enterprise, by American artist Ali Bati

Agradecimentos

No início da pós-graduação, achava pretensioso incluir em uma dissertação a seção de agradecimentos. Para mim, tratava-se de demasiado peso dado ao cumprimento de mais uma dentre tantas “não mais do que obrigações”. Hoje —- inacreditáveis seis anos e dez meses depois — percebo que, fosse este texto mera obrigação, não teria chegado ao fim. E, independentemente da relevância acadêmica que venha a ter, é gigante seu valor quando medido pelo número de pessoas que contribuíram para o resultado — muitas vezes de maneira indireta, no melhor estilo “efeito borboleta”.

Antes de tudo, agradeço a meus pais Suzana e Fábio (*in memoriam*). Não só pelas razões biológico-existenciais óbvias, mas, também, por seu apoio incondicional ao longo dos últimos 30 anos — muitas vezes às custas de grandes sacrifícios financeiros e, pressuponho, emocionais. Sou da opinião que, quanto mais profundo e verdadeiro um sentimento, mais difícil é descrevê-lo com palavras: paradoxalmente, qualquer analogia que se faça será ao mesmo tempo hiperbólica e insuficiente. Assim, aposto na força dos atos e nos registros do tempo como expressão do meu carinho e retribuição a tamanha dedicação. Em especial, sei que a quantidade desproporcional de tempo gasto nas ilustrações desta tese são um sinal de que levo meu pai comigo, ainda que ele não esteja mais aqui.

Em segundo lugar, agradeço ao Tal, a quem algumas vezes já consigo chamar de Fábio, em vez de professor. Como orientador deste trabalho, também ele se enquadra na categoria dos homenageados por motivos existenciais sine qua non. Porém, este é um detalhe acessório: há no IME muitos professores capazes de orientar uma tese acadêmica. Mas, certamente, nenhum capaz de oferecer aos alunos tamanha empatia e comprometimento. Obrigado por não ter desistido de mim, apesar dos constantes atrasos!

Por fim, não posso deixar de agradecer a meu psiquiatra, Pedro Galvão Vianna Filho, cuja contribuição foi crucial para resolver alguns problemas de perspectiva que fugiam ao escopo da geometria projetiva.

As fórmulas para iniciar parágrafos de agradecimento tornam-se rapidamente escassas. Tendo por ora esgotado minha criatividade dando à tese seus retoques finais, “agradeço...”

- a minhas avós Sílvia (*in memoriam*) e Marli, por terem contado a tanta gente que seu neto seria doutor pela USP. Afinal, publicidade é a alma do negócio. Não só, agradeço também por toda a torcida e todo o carinho.
- ao *trio tánc* — meus amigos Pedrinho e Laís. Nosso retrato e minha arvorezinha são lembranças de que mesmo o caos apresenta alguma constância: alguém a quem podemos recorrer mesmo durante as reviravoltas mais imprevisíveis.
- ao Pedro Marques, um dos inusitados amigos a quem a vida por sorte me apresentou e a única pessoa do mundo com quem eu sou capaz de falar sobre absolutamente tudo, menos sobre a tese.

- à Panni, ao mesmo tempo analista e analisanda, o que nos coloca sobre uma inorientável faixa de Möbius que deixaria até Lacan intrigado. Obrigado por me ouvir falar tanto sobre este trabalho que é totalmente irrelevante para sua área de estudo e — ainda por cima — fazer perguntas pertinentes depois! Eu mesmo ainda estou aprendendo a fazer isso.
- ao Gabriel Cozzella. Embora já tenhamos passado muito mais tempo juntos no passado, nosso café anual é o ponto fixo mais estável a respeito do qual tenho notícia, e sem dúvida o mais divertido.
- à Marisa Cantarino, uma inspiração em muitos níveis, do pessoal ao profissional: Matemática Admirável, Realmente Incrível e Super Amiga.
- à Tereza, família Lacerda e, principalmente, Tia Auxiliadora, que pergunta sempre como andam as coisas, apesar de não estarmos mais tão próximos.
- aos amigos do *Pántlika* — no qual eu ingressei mais ou menos ao mesmo tempo que no doutorado — por pelo menos um dia na semana de diversão incondicional garantida. Também à Alinka, minha professora favorita, amiga e a curiosa conexão entre esses dois universos aparentemente disjuntos.
- à Sônia, cuja exposição clara e apaixonante da análise no $\mathbb{R}^n$ me levou ao IME e com quem, desde então, tive a sorte de trabalhar como monitor algumas vezes. Não poderia esquecer também do Mané e sua utilíssima caixa de ferramentas de três itens: teorema de Pitágoras, fórmula quadrática e soma de uma PG.
- à Luna, atualmente no IMPA, que talvez não se lembre de certa vez ter tomado comigo um café que foi crucial para minha insistência na vida acadêmica, mas que certamente se lembra de ter contribuído intensamente para meu apreço pelo lado docente da carreira.
- ao Bruno Santiago, da UFF, que incorpora todas as qualidades que eu sempre prezei na vida acadêmica — idoneidade, receptividade, entusiasmo não competitivo e humanidade — e que, talvez sem saber, foi imprescindível para dar um necessário impulso em minha autoestima acadêmica.
- ao Séba, da Udelar, por ter me recebido com entusiasmo em Montevidéu, demonstrado genuíno interesse por este trabalho e dado importantes sugestões para obter algumas cotas uniformes.
- ao Alejandro Kocsard, da UFF, por ter feito perguntas no TopDin que me motivaram a esboçar aqui a demonstração do teorema de Kerékjártó e Kolev.
- ao Salvador e ao André Salles, por ofertarem cursos daqueles que realmente mudam sua forma de pensar nas coisas.
- ao Nelsera e ao Dedé, por tornarem minha última passagem pelo IMPA um pouco menos sad e, também, pela parceria que se seguiu.
- ao Pips, que em determinado momento ajudou a reavivar meu entusiasmo pela matemática.
- ao Michel e ao Lucas Colucci, amigos de IME, por uma de suas muitas e sempre bem-vindas (re) aparições surpresa: dessa vez, na minha defesa.
- à dinda Solange, pela torcida e pelo entusiasmo.
- aos membros da comissão julgadora — Alejandros, Bruno e Séba — todos matemáticos que eu conheço e admiro. Foi uma grande honra, para mim, aceitarem o convite da banca.
- ao CNPq, pelo apoio financeiro e pelas constantes recordações das sanções legais e financeiras às quais eu estaria sujeito em caso de desistência — um inegável estímulo ao término deste trabalho.

Resumo

Neste ensaio, demonstra-se que qualquer grupo de difeomorfismos que preserve orientação e aja na 2-esfera, estendendo propriamente o grupo conforme das transformações de Möbius, precisa ser ao menos 4 transitivo ou, mais precisamente, 4-transitivo por arcos. Isso significa que quaisquer duas listas ordenadas de quatro pontos distintos podem ser aplicadas uma sobre a outra por alguma transformação do grupo, isotópica à identidade. Argumenta-se, também, que tais grupos apresentam sempre um elemento de entropia topológica positiva, para o qual é dada uma descrição como isotópico a um homeomorfismo pseudo-Anosov relativo da esfera 4-perfurada. Além disso, apresenta-se uma caracterização elementar — em termos de transitividade — das transformações de Möbius dentro do grupo total de difeomorfismos.

Palavras-chaves: transitividade; ação de grupos em superfícies; grupos topológicos; entropia topológica

Abstract

It is proven in this essay that any group of orientation preserving 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. This means that any two ordered lists of four distinct points can be mapped one onto the other via a transformation in the group, isotopic to the identity. In addition, it is shown that any such group must always contain an element of positive topological entropy, for which a description as isotopic to a relative pseudo-Anosov homeomorphism of the 4-punctured sphere is provided. Furthermore, an elementary characterisation of the Möbius transformations within the full group of sphere diffeomorphisms is given in terms of transitivity.

Keywords: transitivity; groups acting on surfaces; topological groups; topological entropy

Sumário

Chapter 1

Introduction

1.1 Overview

Let us approach the subjects of this thesis in two levels, one at a time. We here describe in general lines some of the historical background and motivations that led us to this work, as well as announce our contributions to the field. We then proceed to the more technical and prosy Preliminaries Section 1.2, which a more experienced reader may glance at diagonally or skip, but from which a reader less familiarised with the language may benefit from.

Let $\mathcal{M}$ be a closed and oriented topological manifold, which for all practical purposes may be thought of as either the circle or some surface of genus ≥ 0. If the set of all its orientation preserving homeomorphisms – denoted by $\mathrm{Homeo}_+(\mathcal{M})$ – is endowed with the composition operation $\circ$, then the usual uniform convergence metric turns it into a topological group, the subgroups of which one can try to understand and classify.

Placed in this degree of generality this may be a hopelessly difficult program, so specific approaches have been delimited. For instance, in an early 2000s' survey, Ghys proposed such a classification scheme for closed and transitive groups acting on the unit circle $\mathbb{S}^1$ (18, Problem 4.4). Here, closed refers to the uniform topology aforementioned, 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, discussed in Section 1.2.1, was later proven by Giblin and Markovic (19).

A relevant part of understanding closed subgroups of $\mathrm{Homeo}_+(\mathcal{M})$ 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 uniform closures.

For example, it was proven by Le Roux (33) that the (closed) subgroup of area preserving homeomorphisms is always maximal in triangulable manifolds of dimension ≥ 2 – see Theorem 1.16 for the exact statement. In that same paper, Le Roux pointed out how the rich one-dimensional theory contrasts with the yet to be developed higher dimensional setting, where mostly isolated results are known.

Concurrently, Kwakkel and Tal derived a number of specialised results concerning the unit sphere $\mathbb{S}^2$. Their findings are communicated in a preprint available at the ArXiv repository (30), not submitted for publication due to an irreconcilable difference of opinion between its authors on the presentation of the subject. While their paper is a strong source of technical inspiration, this thesis is self-contained. In particular, no result of ours makes direct use of results therein.

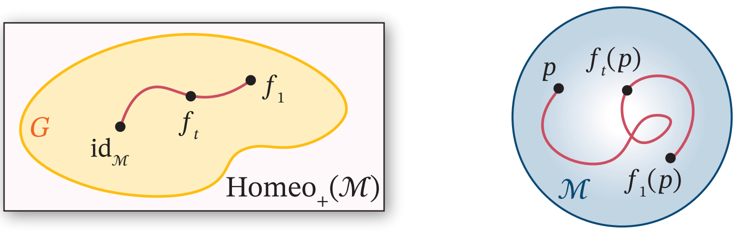
Among others, Kwakkel, Tal and Le Roux asked the question of whether the Möbius group $\text{Möb}(\mathbb{S}^2)$ – consisting of all orientation preserving conformal diffeomorphisms of the 2-sphere – is maximal within the full group of homeomorphisms. In other words, whether there are no uniformly closed groups of homeomorphisms properly contained between $\text{Möb}(\mathbb{S}^2)$ and $\text{Homeo}_+(\mathbb{S}^2)$.

This question is a direct parallel to a conjecture made by de la Harpe with respect to the unit circle $\mathbb{S}^1$ (*apud* (6), Q4.1), and which was proven to be true by Giblin and Markovic in their already mentioned 2006 paper. The milestone of the proof presented therein is that (some form of) 4-transitivity implies transitivity of any order. In this work, we provide the following insight – from the transitivity viewpoint – into extensions of the Möbius group.

Theorem A. *Let $G \subset \text{Diff}^1_+(\mathbb{S}^2)$ be a group properly extending $\text{Möb}(\mathbb{S}^2)$. 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 was introduced. Albeit more precisely defined in the following text, it essentially means that not only ordered lists $(p_1, \dots, p_4)$ and $(q_1, \dots, q_4)$ of distinct points can be mapped one onto another by 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, as pictured in Figure 1.1. Although the set of homeomorphisms is represented there as a flat and bounded piece of cardboard, it is actually to be thought of as a "huge" space – bigger than an infinite dimensional Lie group.

Figure 1.1 – A cartoon of arc transitivity: a continuous curve $t \in [0,1] \mapsto f_t \in \text{Homeo}_+(\mathcal{M})$ starting at the identity (that is, an isotopy) gives rise to curves on $\mathcal{M}$ connecting points to their images under the terminal map f_1.



It should be noticed that, if a subgroup G of $\text{Homeo}_+(\mathcal{M})$ happens to be k-transitive for every $k \in \mathbb{N}$, then an argument of separability implies its uniform closure to be the whole of $\text{Homeo}_+(\mathcal{M})$, yielding maximality. The natural step after Theorem A would thus be deriving higher orders of transitivity from 4-transitivity. Unfortunately, the argument used by Giblin and Markovic to do so presents no obvious generalization to higher dimensions, for it strongly relies on the complement of a finite set in the circle being composed of disjoint open intervals.

Theorem A would therefore be a mere curiosity, hadn't it – or rather its proof – had 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 any such proper extension $G \subset \text{Diff}^1_+(\mathcal{M})$ of the Möbius group, as described below.

Theorem B. *Let G be as in Theorem A. Then, G contains an element f fixing at least four distinct points and such that its restriction to their complement is isotopic to a pseudo-Anosov map relative to those 4 points. In particular, f has strictly positive topological entropy.*

In a certain sense, Theorems A and B when put together tell that the Möbius group is largely "enriched" by any extension of it, for:

- no individual element of $\text{Möb}(\mathbb{S}^2)$ presents interesting dynamical behaviour – their full description is summarised in Section 1.2;
- being 3-transitive is a defining property of the Möbius group in the transitive setting – a slightly vague statement that is made precise by Theorem C of Chapter 2.

These are clues that suggest an affirmative answer to the question of Kwakkel, Tal and Le Roux. However, as of the submission of this essay, the subject remains prone to further research.

1.2 Preliminaries

Let us now review some of the important concepts upon which this work is based. Our primary goal is to establish terminology, notation and a few instrumental results. Curiously enough, some of them are folklore for which explicit proofs — and sometimes even complete statements — are nowhere to be found in the (known to the author extent of the) standard literature. For this section not to become too clutched, some such results are collected in Appendix A, where they are presented in a "primer" fashion in hopes of providing a useful read on the subject. Consequently, some redundancy between contents here and there may be present. In particular, we start with a definition summing up the contents of Section A.1.

1.1 Definition. Let $\mathcal{M}$ be a closed and oriented topological manifold. We denote by $\text{Homeo}_+(\mathcal{M})$ the topological group of its *orientation preserving homeomorphisms* endowed with the uniform convergence topology and the composition operation.

In other words, if d is some (chosen and fixed) metric generating the manifold's topology, then the *uniform distance* between two maps f and g is given by $d_\infty(f,g) = \sup\{d(f(p), g(p)) : p \in \mathcal{M}\}$. In particular, f and g are said to be *ε-close* if $d_\infty(f,g) \leq \varepsilon$. A full discussion on what does it mean for a homeomorphism to preserve orientation is left to Appendix A, while specialised interpretations in dimensions 1 and 2 are postponed to subsequent sections, as they play no role for now.

Throughout this essay, $\mathcal{M}$ is typically a Riemannian manifold endowed with some canonical smooth structure, case in which d is the usual infimum of arc length metric induced accordingly. We then may also consider the subgroup $\text{Diff}^1_+(\mathcal{M}) \subset \text{Homeo}_+(M)$, consisting of all orientation preserving diffeomorphisms of class $C^1(\mathcal{M})$, and its finer C^1 *topology*, on occasion also referred to as the *Whitney weak topology* (24). This topology is generated by a class of subbasic neighbourhoods. Each such neighbourhood is specified by the following ingredients:

- a diffeomorphism $f \in \text{Diff}^1_+(M)$;
- a coordinate chart (U, Φ);
- a compact set $K \subset U$;
- a coordinate chart (V, Ψ) such that $f(K) \subset V$;
- and a real number $\varepsilon > 0$.

The corresponding subbasic neighbourhood is denoted by

[1.1] $$\mathcal{B}(f\,;\,(U,\Phi), K, (V,\Psi)\,;\,\varepsilon)\,,$$

and consists of all the orientation preserving diffeomorphisms g such that $g(K) \subset V$ and:

$$\max\Big\{ \big|\hat{f}(z) - \hat{g}(z)\big| \,,\, \|\mathrm{D}\hat{g}(z) - \mathrm{D}\hat{f}(z)\| \Big\} \leq \varepsilon \text{ for every } z \in \Phi(K)\,, \tag{1.2}$$

where $\hat{f} = \Psi \circ f \circ \Phi^{-1}$ and $\hat{g} = \Psi \circ g \circ \Phi^{-1}$ are the local expressions of f and g in coordinates and $\|\cdot\|$ is the usual operator norm for linear transformations of Euclidean space. Then, a set $\mathcal{W}$ is declared to be a neighbourhood of $f \in \mathrm{Diff}^{+}_{1}(\mathcal{M})$ if it contains a finite intersection of sets of the form [1.1].

The C^1 topology conveys the idea that two diffeomorphisms are close if both their local expressions and the respective differentials are uniformly close on compact sets. However, it is not always practical for computations, as they usually require chopping a compact set and its image in pieces subordinate to covers by coordinate domains. An alternative description is given in Section A.2, and used therein to establish the following key facts.

1.2 Proposition. Let $\mathcal{M}$ be a closed and oriented smooth manifold. Then, the C^1 topology in $\mathrm{Diff}^{+}_{1}(\mathcal{M})$ is metrisable and turns this set, endowed with the composition operation $\circ$, into a topological group such that the inclusion morphism $\big(\mathrm{Diff}^1_+(\mathcal{M}), C^1 \text{ topology}\big) \hookrightarrow (\mathrm{Homeo}_+(\mathcal{M}), d_\infty)$ is continuous.

In the differentiable context, the understanding of diffeomorphism groups is largely addressed by the so-called *Zimmer program*. Originally revolving around finitely generated subgroups of $\mathrm{Diff}^1(\mathcal{M})$, broader problems such as constraining possible actions of groups on $\mathcal{M}$, determining subgroups of $\mathrm{Diff}^1(\mathcal{M})$ beyond the "large" (in the sense of Labourie (31)) examples already known and further exploring analogies to Lie group theory may be understood as a part of it as well. The current *état de l'art* on the subject is often surveyed and communicated by Fisher (13, 14).

These aspects are addressed to some degree in this thesis for $\mathcal{M} = \mathbb{S}^2$, with its diffeomorphism groups being considered up to uniform closure. The underlying context is that of transitivity — that is, the possibility to map prescribed elements of $\mathcal{M}$ onto each other via transformations in the group, as precisely stated below.

1.3 Definition. Given $k \in \mathbb{N}$, the action of a subgroup $G \subset \mathrm{Homeo}_+(\mathcal{M})$ is said to be *k-transitive* if for every pair of k-tuples $(p_1, \dots, p_k)$ and $(q_1, \dots, q_k)$ — each of them consisting of mutually distinct points — there exists some transformation $g \in G$ such that

$$q_i = g(p_i) \text{ for each } i \in \{1, \dots, k\}\,.$$

When such a transformation is *unique*, the group is said to be *sharply k-transitive*.

Most constructions in this essay are based upon *isotopies*, here understood in a very broad sense as jointly continuous maps $f : I \times \mathcal{M} \to \mathcal{M}$, where:

- $I \subseteq \mathbb{R}$ is a (possibly unbouded) real interval, and
- for every $t \in I$, the function $f_t : \mathcal{M} \to \mathcal{M}$ defined by $f_t(p) = f(t, p)$ is an orientation preserving homeomorphism of the (not necessarily compact) oriented manifold $\mathcal{M}$.

As stressed in (37), whenever $\mathcal{M}$ is metrisable and compact we can either prescribe the jointly continuous rule f or a family $(f_t)_{t \in I}$ for which $t \in I \mapsto f_t \in \mathrm{Homeo}_+(\mathcal{M})$ is continuous with respect to d_∞. When the real interval I in question is the *standard unit interval* $[0, 1]$ — as it is often the case in the context of homotopies — we denote it by $\mathbb{I}$. In the following remark, we solve a small ambiguity that may arise at some point.

1.4 Remark. It will often be the case that $\mathcal{M}$ is a closed and smooth manifold such that $f_t \in \mathrm{Diff}^1_+(\mathcal{M})$ for every $t \in I$, yet the mapping $t \mapsto f_t$ can only be assured to be continuous with respect to d_∞. Those shall be referred to simply as isotopies, according to the previous terminology. When $t \in I \mapsto f_t \in \mathrm{Diff}^1(\mathcal{M})$ can actually be ensured to be continuous with respect to the C^1 topology, we shall name $(f_t)_{t\in I}$ a *diffeotopy*.

Let us now specialise in the relation between groups and isotopies. To do so, we borrow notations from (2) and from the theory of flows to define a handful of concepts.

1.5 Definition. Let $G \subset \mathrm{Homeo}_+(\mathcal{M})$ be a subgroup. An isotopy $(f_t)_{t\in I}$ such that $0 \in I$, $f_0 = \mathrm{id}$ and $f_t \in G$ for every $t \in I$ will be referred to as an *$\mathcal{I}G$-isotopy*. Given a point $p \in \mathcal{M}$, we define its *trajectory* under f as

$$\gamma_f(p) \stackrel{\text{def}}{=} \{f_t(p) : t \in I\}.$$

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, determined by increasing values of t. If I is unbounded above, we further define the *ω-limit* as the following (possibly empty) set of accumulation points:

$$\omega_f(p) \stackrel{\text{def}}{=} \left\{q \in \mathcal{M} : \text{there exists some sequence } t_n \nearrow +\infty \text{ such that } f_{t_n}(p) \to q\right\}.$$

There is an analogous notion of α-limit when I is unbounded below:

$$\alpha_f(p) \stackrel{\text{def}}{=} \left\{q \in \mathcal{M} : \text{there exists some sequence } t_n \searrow -\infty \text{ such that } f_{t_n}(p) \to q\right\}.$$

Lastly, G is said to be *arc k-transitive* if for every pair of k-tuples as in Definition 1.3 there exists an $\mathcal{I}G$-isotopy $(g_t)_{t\in\mathbb{I}}$ such that $g_1(p_i) = q_i$ for each $i \in \{1, \dots, k\}$.

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

1.6 Definition. For a fixed a subgroup $G \subset \mathrm{Homeo}_+(\mathcal{M})$ and given $p, q \in \mathcal{M}$, we agree that

[1.3] $$p \sim_G q \iff \text{there exists an } \mathcal{I}G\text{-isotopy } (f_t)_{t\in\mathbb{I}} \text{ such that } f_1(p) = q.$$

The set of all points q such that $p \sim_G q$ is referred to as the set of *points accessible from p* (in G), and denoted by $\mathcal{A}_G(p)$.

1.7 Lemma. The relation $\sim_G$ presented in [1.3] is an equivalence relation in $\mathcal{M}$.

Proof. For it to be an equivalence relation, $\sim_G$ must verify the subsequently listed properties.
Reflexive: for every $p \in \mathcal{M}$, $p \sim_G p$.

Indeed, it suffices to consider the constant $\mathcal{I}G$-isotopy $f_t = \mathrm{id}_\mathcal{M}$, $0 \le t \le 1$.

Symmetric: for every $p, q \in \mathcal{M}$, $p \sim_G q$ implies $q \sim_G p$.

If $p \sim_G q$, let $(f_t)_{t\in\mathbb{I}}$ be as in [1.3], and define $h_t = f_t^{-1}$ for each $t \in \mathbb{I}$. Then, $(h_t)_{t\in\mathbb{I}}$ is an $\mathcal{I}G$-isotopy satisfying $h_1(q) = f_1^{-1}(q) = f_1^{-1}(f_1(p)) = p$.

Transitive: for every $p, q, r \in \mathcal{M}$, $p \sim_G q$ and $q \sim_G r$ imply $p \sim_G r$.

The definition in [1.3] yields $\mathcal{I}G$-isotopies $(f_t)_{t\in\mathbb{I}}$ and $(h_t)_{t\in\mathbb{I}}$ such that $f_1(p) = q$ and $h_1(q) = r$. Consider

$$k_t \stackrel{\text{def}}{=} \begin{cases} f_{2t} & \text{if } 0 \le t \le \dfrac{1}{2}\,, \\ h_{2t-1} \circ f_1 & \text{if } \dfrac{1}{2} \le t \le 1\,. \end{cases}$$

By the topological group property, the two proposed expressions for k_t are continuous with respect to t and coincide when $t = 1/2$. The family $(k_t)_{t\in\mathbb{I}}$ thus defines a continuous curve in G (this kind of concatenation argument will often be employed from now on without further explicit mention). We have $k_0 = f_0 = \mathrm{id}_{\mathcal{M}}$ and $k_1(p) = h_1(f_1(p)) = h_1(q) = r$, so $p \sim_G r$ follows. ▨

The equivalence relation just defined partitions the manifold $\mathcal{M}$ into sets of points mutually accessible under trajectories of isotopies lying in the identity component of a given subgroup $G \subset \mathrm{Homeo}_+(\mathcal{M})$. We now present a practical accessibility criterion.

1.8 Lemma. 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 $p, q \in \mathcal{M}$ such that $\gamma_f(p) \cap \gamma_h(q) \neq \varnothing$ we have $\mathcal{A}_G(p) = \mathcal{A}_G(q)$.

Proof. Suppose that $f_a(p) = h_b(q)$, for some $a \in I$ and $b \in J$. Given $r \in \mathcal{A}_G(q)$, there exists an $\mathcal{I}G$-isotopy $(g_t)_{t\in\mathbb{I}}$ such that $g_1(q) = r$. Then,

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

is an $\mathcal{I}G$-isotopy, satisfying:

$$k_1(p) = \big(g_1 \circ h_b^{-1} \circ f_a\big)(p) = g_1\big(h_b^{-1}(f_a(p))\big) = g_1\big(h_b^{-1}(h_b(q))\big) = g_1(q) = r\,.$$

This shows that $\mathcal{A}_G(q) \subset \mathcal{A}_G(p)$, and the converse inclusion follows by the symmetry of $\sim_G$. ▨

Before pursuing other subjects, we finish this section stating an elementary yet instrumental result for further reference. On it, as in the rest of this essay, "planar" is a purposely broad term which may refer to $\mathbb{R}^2$ or $\mathbb{C}$, depending on the context, question which is immaterial when it comes to purely topological considerations.

1.9 Lemma. Let $(f_t)_{t\in\mathbb{I}}$ be a planar isotopy such that the origin $\mathbf{0}$ is a fixed point for every t. Then, given $\varepsilon > 0$ there exits $\eta > 0$ such that $|z| < \eta$ implies $|f_t(z)| < \varepsilon$ for every $t \in \mathbb{I}$.

Proof. Of course, one notices that the open manifold $\mathcal{M} = \mathbb{R}^2$ is not compact, so let us use the characterisation of f as a jointly continuous function of (t, z).

By hypothesis, $|f(t, \mathbf{0})| = \mathbf{0}$ for every $t \in \mathbb{I}$. Thus, there exist $\delta_t > 0$ and $\eta_t > 0$ such that $|s - t| < \delta_t$ and $|z| < \eta_t$ imply $|f(s, z)| < \varepsilon$. By compacity, $\mathbb{I} = \bigcup_{i=1}^n \big(t_j - \delta_j, t_j + \delta_j\big) \cap \mathbb{I}$, for some $t_1, \dots, t_n \in \mathbb{I}$ and their respective $\delta_j \stackrel{\text{def}}{=} \delta_{t_j}$. If we thus consider $\eta = \min_{1\le i\le n} \eta_{t_j}$, the proposed statement is readily seen to hold. ▨

1.2.1 The circle case

Consider the *unit circle* $\mathcal{M} = \mathbb{S}^1$, which is the only closed one-dimensional topological manifold. The most widespread ways of realising it are as the set T of complex numbers at unit distance from the origin — endowed with the topology induced by the ambient space — or as the abstract quotient space $\mathbb{R} / \mathbb{Z} = \{x + \mathbb{Z} : x \in \mathbb{R}\}$, obtained upon identifying each point on the real line with its integer translations.

The set T is more concrete geometrically, but is not practical for calculations, mostly due to the nuisances involving angle functions. The space $\mathbb{R} / \mathbb{Z}$, in turn, is more convenient algebraically, for it inherits an additive group structure, and also due to the usage of lifts of circle maps, a procedure soon to be described. Identification between the two viewpoints is provided by the homeomorphism $x + \mathbb{Z} \mapsto e^{2\pi i x}$, usually interpreted as (a factor of) "wrapping" the real line around the (geometric) circle.

If $\pi : \mathbb{R} \to \mathbb{R} / \mathbb{Z}$ denotes the quotient map, one sees that the interval $[0, 1)$ is a fundamental domain. It is thus usual to choose its element $x - \lfloor x \rfloor$ as the canonical representative of the class $x + \mathbb{Z}$, and also to think of the unit circle as the unit interval with identified endpoints. In particular, $d(x + \mathbb{Z}, y + \mathbb{Z}) \stackrel{\text{def}}{=} \{|x - y|, 1 - |x - y|\}$ for $x, y \in [0, 1)$ defines a metric in $\mathbb{R} / \mathbb{Z}$ which is equivalent to the quotient topology, for its open balls lift under π to countable unions of open intervals. This allows one to also consider the Borelian probability Leb, which is given, for $B \subset \mathbb{R} / \mathbb{Z}$, by the usual Lebesgue measure on the real line of the set $\pi^{-1}(B) \cap [0, 1)$.

It turns out that π is not only a quotient map, but rather a covering map. Any $f \in \mathrm{Homeo}\,(\mathbb{S}^1)$ can thus be lifted to a homeomorphism $F : \mathbb{R} \to \mathbb{R}$ of the real line satisfying $\pi \circ F = f \circ \pi$. Any two such lifts differ by a constant integer, and each of them satisfies $F(x + 1) = F(x) + d$, where $d = \pm 1$ is the *degree* of f. In particular, we may *define* a circle homeomorphism to be orientation preserving if it lifts to increasing homeomorphisms of the real line or, equivalently, if it has degree 1. Summing up, one may either prescribe $f \in \mathrm{Homeo}_+(\mathbb{S}^1)$ or $F \in \mathrm{Homeo}\,(\mathbb{R})$ such that $F(x + 1) = F(x) + 1$.

These facts, along with the full dynamical characterisation of each $f \in \mathrm{Homeo}_+(\mathbb{S}^1)$ in terms of the so-called Poincaré rotation number, are a fundamental component of the dynamical systems framework, being addressed by essentially any textbook on the subject (*cf. e.g.* (8), Sec. 7.1). Let us now present yet another form of seeing the unit circle, as the Alexandrov compactification $\mathbb{R} \cup \{\infty\}$ of the real line.

To do so, consider the set T sitting in the complex plane and fix a privileged point, say i. Then, stereographic projection $\Psi : \mathbb{S}^1 \setminus \{i\} \to \mathbb{R}$ from this point onto the real axis extends to the sought homeomorphism, the point at infinity being identified with the projection's basepoint i. For concreteness, the formulae are:

$$\Psi(z) = \frac{u}{1 - v} \text{ if } z = u + iv \quad \text{and} \quad \Psi^{-1}(t) = \frac{1}{t^2 + 1}\left[2t + i(t^2 - 1)\right].$$

Having settled the context, let us present two important subgroups of $\mathrm{Homeo}_+(\mathbb{S}^1)$ which act on the unit circle. The first is the group $\mathrm{Rot}(\mathbb{S}^1) = \{r_\alpha : 0 \leq \alpha < 1\}$ of *rotations*, given additively as $r_\alpha(x) = x + \alpha \bmod \mathbb{Z}$ for $x \in [0, 1)$. If one recalls the *special orthogonal group* $\mathrm{SO}_2(\mathbb{R})$ of real 2×2 orthogonal matrices of unit determinant, for each such matrix A there exists an unique $\alpha \in [0, 1)$ such that $Az = e^{2\pi i\alpha} z$. In view of the identification between $\mathbb{R}/\mathbb{Z}$ and T provided by the exponential, this yields an isomorphism of topological groups between $\mathrm{Rot}(\mathbb{S}^1)$ and $\mathrm{SO}_2(\mathbb{R})$.

Since the latter is a very well known compact Lie group, $\mathrm{Rot}(\mathbb{S}^1)$ is also compact. As it turns out, any compact subgroup G of $\mathrm{Homeo}_+(\mathbb{S}^1)$ is actually conjugate to a subgroup of the rotations. The argument is based on averaging its Haar measure — which always exist for compact topological groups — to conclude that G preserves in $\mathbb{S}^1$ a Borelian probability equivalent to Leb. Since rotations lift to translations, and translations are the only mappings that preserve both Lebesgue measure and orientation, the conclusion follows. The details of the proof, as well as a thorough discussion on the Lie groups that do act on $\mathbb{S}^1$ may be found in the first chapter of the book by Navas (41).

Recall now the real *projective special linear group* $\mathrm{PSL}_2(\mathbb{R})$, obtained from the special linear group $\mathrm{SL}_2(\mathbb{R})$ of 2×2 matrices of unit determinant upon declaring A and $-A$ as equivalent. This group acts on the extended real line by *linear fractional transformations*:

$$\varphi : \left(\pm \begin{bmatrix} a & b \\ c & d \end{bmatrix}, t \right) \in \mathrm{PSL}_2(\mathbb{R}) \times \mathbb{R} \cup \{\infty\} \mapsto \frac{at+b}{ct+d} \in \mathbb{R} \cup \{\infty\}.$$

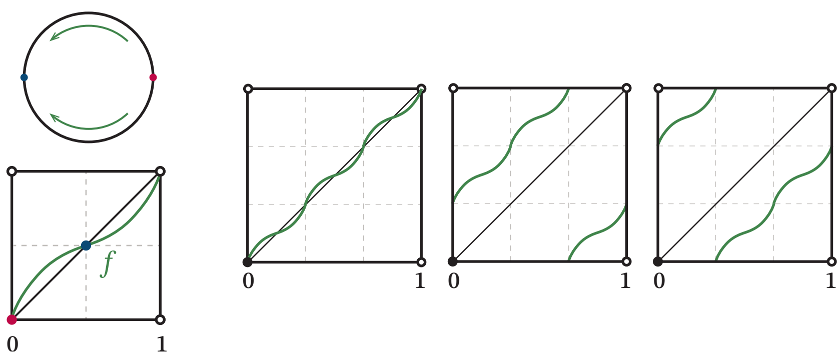
Via conjugation under Ψ, the above gives rise to an orientation preserving action on the unit circle, which we shall name the *Möbius action*, realised by a subgroup $\mathrm{M\ddot{o}b}(\mathbb{S}^1) \subset \mathrm{Homeo}_+(\mathbb{S}^1)$.

In particular, when A is an orthogonal matrix there exists an unique $0 < \beta \leq 1/2$ such that either A or $-A$ is represented by $e^{2\pi i \beta}$. Then, a lengthy yet elementary calculation shows that $\Psi^{-1} \circ \varphi(A, \Psi(z)) = e^{2\pi i(1-2\beta)} z$. In other words, $\mathrm{Rot}(\mathbb{S}^1)$ is realised by $\mathrm{PSO}_2(\mathbb{R})$ as a subgroup of $\mathrm{M\ddot{o}b}(\mathbb{S}^1)$, although its representation as such is *not* the same as by the orthogonal group itself.

The subgroups of $\mathrm{M\ddot{o}b}(\mathbb{S}^2)$ are fully characterised — up to conjugacy by an element of $\mathrm{Homeo}_+(\mathbb{S}^1)$ — as the so-called *convergence groups*. This is a pivotal milestone of the theory, established by the works of Hinkkanen (23) and Gabai (16).

Lastly, recollect that the unit circle admits a k-fold cover π_k by itself, which reads $x \mapsto kx \bmod \mathbb{Z}$ in additive notation. One verifies that the deck transformations consist in the cyclic group generated by the corresponding *rotation of finite order*, $r_k(x) \overset{\text{def}}{=} (x + 1/k) \bmod \mathbb{Z}$. Clearly, each point in $\mathbb{S}^1$ has k preimages on its fiber, and thus each continuous circle map admits k distinct lifts commuting with the deck transformations. For a given subgroup $G \subset \mathrm{Homeo}_+(f)$, the set of all such lifts thus defines a new group $G_{(k)}$, which is called a *cyclic cover* of G, for it projects onto G and has the same cyclic group of finite rotations as its deck transformations. Of course, $G_{(1)} = G$. A prototypical situation is shown in Figure 1.2.

Figure 1.2 – A hyperbolic Möbius transformation f and its three lifts in the cyclic cover $\mathrm{M\ddot{o}b}_{(3)}(\mathbb{S}^1)$. Notice that $G_{(k)}$ typically has no explicit representation whatsoever in terms of the base group G.



In his already mentioned essay on groups acting on the circle (18), Ghys posed the question of whether any closed and transitive subgroup G of $\mathrm{Homeo}_+(\mathbb{S}^1)$ would be conjugate to one of the following: $\mathrm{Rot}(\mathbb{S}^1)$, $\mathrm{Möb}_{(k)}(\mathbb{S}^1)$ or $\mathrm{Homeo}_{+,(k)}(\mathbb{S}^1)$. The conjecture was motivated by an extremely similar result available to the case of Lie groups, and roughly states that the circle supports only Euclidean and projective geometries as its intrinsic symmetries.

The corresponding result was proven a few years later by Giblin and Markovic (19). However, to do so, a new hypothesis was introduced: the group G in question must also contain a nonconstant isotopy. The following was then established.

1.10 First Giblin & Markovic Classification Theorem. Let $G \subset \mathrm{Homeo}_+(\mathbb{S}^1)$ be a transitive group containing a nonconstant isotopy. Then, one, and only one, of the following holds:

1) $G \simeq \mathrm{Rot}(\mathbb{S}^1)$;

2) $G \simeq \mathrm{Möb}(\mathbb{S}^1)$;

3) G is such that. for every $f \in \mathrm{Homeo}_+(\mathbb{S}^1)$ and every finite number of points $x_1, \dots, x_n \in \mathbb{S}^1$, there exists $g \in G$ satisfying $f(x_i) = g(x_i)$ for each $i \in \{1, \dots, n\}$;

4) $G \simeq \mathrm{Möb}_{(k)}(\mathbb{S}^1)$ for some $k > 1$;

5) $G \simeq H_{(k)}$ for some $k > 1$, where H is a subgroup of $\mathrm{Homeo}_+(\mathbb{S}^1)$ satisfying condition **3)**.

Here, $\simeq$ denotes conjugation in $\mathrm{Homeo}_+(\mathbb{S}^1)$.

It is interesting to notice that a full classification theorem is actually possible without the hypothesis of G being closed, which just further simplifies the description. Condition **3)** above is *a kind of n-transitivity* for every n: the homeomorphism f, which doesn't have anything to do with the group G, is there to ensure that this transitivity respects what the authors define as *matching orientations*. In other words, one is only allowed to map $(x_1, \dots, x_n)$ onto $(y_1, \dots, y_n)$ if both n-uples are cyclically ordered in the same fashion.

Most of the constructions in the paper are based upon a concept of continuous transitivity which is slightly stronger than what we called arc n-transitivity in Definition 1.5: G is said to be *continuously n-transitive* if for every pair of paths $(x_1(t), \dots x_n(t))$ and $(y_1(t), \dots y_n(t))$ such that their orientations match for every $t \in \mathbb{I}$, there exists an $\mathcal{I}G$-isotopy $(g_t)_{t\in\mathbb{I}}$ satisfying $g_t(x_i(t)) = y_i(t)$. The presence of a nonconstant isotopy is then used to establish the bridge between continuous transitivity and ordinary transitivity.

If G is further required to be closed, condition **3)** yields the whole group of homeomorphisms, since for a fixed enumeration $\{q_m : m \in \mathbb{N}\}$ of $\mathbb{Q} \cap [0, 1)$ and each given $f \in \mathrm{Homeo}_+(\mathbb{S}^1)$ one may obtain $g_n \in G$ such that $g_n(q_i) = f(q_i)$ for each $i \in \{1, \dots, n\}$, thus producing a sequence $(g_n)_{n\in\mathbb{N}}$ in G that converges uniformly to f. This establishes the following.

1.11 Second Giblin & Markovic Classification Theorem. Let $G \subset \mathrm{Homeo}_+(\mathbb{S}^1)$ be a closed and transitive group containing a nonconstant isotopy. Then, one, and only one, of the following holds:

1) $G \simeq \mathrm{Rot}(\mathbb{S}^1)$

2) $G \simeq \mathrm{Möb}_{(k)}(\mathbb{S}^1)$ for some $k \geq 1$.

3) $G \simeq \mathrm{Homeo}_{+,(k)}(\mathbb{S}^1)$

In particular, $\mathrm{Möb}(\mathbb{S}^2)$ is maximal in $\mathrm{Homeo}_+(\mathbb{S}^2)$.

The conclusion of maximality follows from the fact that elements in $\mathrm{Möb}(\mathbb{S}^1)$ can have at most two fixed points — as such points (in $\mathbb{R} \cup \{\infty\}$) must be the solutions to a second degree polynomial equation — while the cyclic covers in the statement of the theorem admit maps with more than two fixed points, as shown by Figure 1.2.

1.2.2 Specific context

Let us consider now the *unit sphere* $\mathcal{M} = \mathbb{S}^2$, which is the closed and orientable surface of genus 0. It can be thought of either as the set $\mathbb{S}^2 = \{P \in \mathbb{R}^3 : |P| = 1\}$ of points in Euclidean 3-space at unit distance from the origin or as the Alexandrov compactification $\mathbb{C} \cup \{\infty\}$ of the complex plane. In the latter case, identification is provided by *stereographic projection* from the North Pole $N \stackrel{\text{def}}{=} (0, 0, 1)$.

For the sake of completeness we remember that stereographic projection is a map Ψ_N from $\mathbb{S}^2 \setminus \{N\}$ onto the Euclidean plane, which sends P to the intersection point between the line passing through P and N with the xy-plane. If $P = (X, Y, Z)$, it is given explicitly as

$$\Psi_N(P) = \frac{X + iY}{1 - Z} .$$

Notice that $\mathbb{S}^2$ is an embedded Riemmanian submanifold of its ambient space $\mathbb{R}^3$ — the metric $\langle \cdot, \cdot \rangle_P$ at a point P being given by restriction of the usual Euclidean inner product to $T_P\mathbb{S}^2$. With respect to this structure, Ψ_N turns out to be a conformal diffeomorphism between the open submanifold $\mathbb{S}^2 \setminus \{N\}$ and the plane. Here, *conformal* means that angles between differentiable curves are preserved.

It is readily seen that Ψ_N is not defined at the North Pole itself. This pole is sent to the “point at infinity”, thus establishing a homeomorphism between the sphere and $\mathbb{C} \cup \{\infty\}$. Consequently, any sphere mapping fixing ∞ defines, by conjugation with stereographic projection, a planar mapping having the same degree of regularity, which is often denoted by the same letter. We stress that the converse procedure — from the plane to the sphere — is slightly more delicate.

Typically, we shall confound points in the plane and their stereographic images without notice, using the same letters to label them. Thus, $z = (x, y) \in \mathbb{R}^2$, which is naturally identified with the complex number $z = x + iy$, may also denote the point

$$\Psi_N^{-1}(z) = \left(\frac{2x}{x^2 + y^2 + 1} , \frac{2y}{x^2 + y^2 + 1} , \frac{x^2 + y^2 - 1}{x^2 + y^2 + 1} \right) \in \mathbb{S}^2 .$$

We also fix two other reference points that play a key role in the arguments to follow:

- $\mathbf{0}$ is the South Pole, and corresponds under Ψ_N to the plane’s origin;
- $\mathbf{1}$ is the point $(1, 0, 0)$, which corresponds under Ψ_N to its counterpart on the real axis.

There is also an analogously defined stereographic projection Ψ_S from the South Pole, so that $\big\{(\Psi_N, \mathbb{S}^2 \setminus \{\infty\}), (\Psi_S, \mathbb{S}^2 \setminus \{\mathbf{0}\})\big\}$ becomes a (smooth) conformal atlas for the sphere.

Some subsets of $\mathbb{S}^2$ are given special names: *parallels* are the circles obtained by intersection of the sphere with horizontal planes, whilst *meridians* are the circles obtained by intersection of the sphere with vertical planes containing the origin. More generally, an intersection of the sphere with a plane through the origin is called a *great circle*, and divides the sphere into two open connected components called *hemispheres*.

In particular, the meridian through $\mathbf{0}, \mathbf{1}$ and ∞ — which is the stereographic image of the (real)

x-axis — is denoted by $\varGamma$. To an observer external to the sphere's surface, this meridian defines on its right an *eastern hemisphere* $\mathcal{H}^+$, corresponding to the upper half-plane, and on its left an *eastern hemisphere* $\mathcal{H}^-$, corresponding to the lower half-plane.

It is widely known that, with respect to the Euclidean structure already discussed, maximal geodesics on the sphere have great circles as images. Therefore, one may explicitly compute the corresponding *round distance* $d(P, Q)$ between two points $P, Q \in \mathbb{S}^2$ as the length of the shortest arc determined by them on a great circle containing both P and Q. Such great circle can be obtained as the intersection of a plane containing P, Q and the origin of $\mathbb{R}^3$ with the sphere. It is thus unique unless P, Q are *antipodal*, that is $Q = -P$.

Meanwhile, for z, w in the plane we define their *chordal distance* $\hat{d}$ to be the Euclidean distance between their stereographic images, which is explicitly given by:

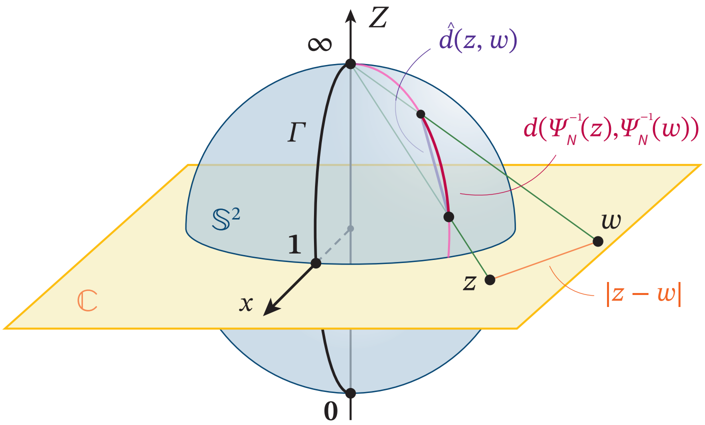
[1.4] $$\hat{d}(z, w) \overset{\text{def}}{=} \frac{2\,|z - w|}{\sqrt{1 + |z|^2}\,\sqrt{1 + |w|^2}} \quad .$$

Then, upon setting

$$\hat{d}(z, \infty) \overset{\text{def}}{=} \frac{2}{\sqrt{1 + |z|^2}}$$

one induces on the sphere a metric that is equivalent to the round metric, with $2^{-1}d \le \hat{d} \le 2d$. These two distances, whose relation is pictured in Figure 1.3, may thus be used interchangeably whenever convenient. One sees that stereographic projection preserves angles at the expense of heavily distorting distances and areas.

Figure 1.3 – The chordal distance $\hat{d}$ between planar points z, w is given by the Euclidean distance between their stereographic images, and is equivalent to their round distance d, measured along great circles.



In order to discuss some symmetries of the spherical setting, we first mention a theorem by Belliart (5), which states that a Lie group G admits a faithful and fixed-point free action by homeomorphisms of $\mathbb{S}^2$ if, and only if, it has $\mathrm{SO}_3(\mathbb{R})$, $\mathrm{PSL}_3(\mathbb{R})$ or $\mathrm{PSL}_2(\mathbb{C})$ as a quotient. Here, fixed-point free means that no point of $\mathbb{S}^2$ is kept fixed by all elements of G simultaneously. Although informative for the purposes of classification, this result does not reveal anything about the actual representation of G: it only tells that a surjective morphism does exist from G onto one such group.

It turns out that each of these groups is indeed canonically linked to a certain orientation preserving symmetry of the 2-sphere: $\mathrm{PSL}_2(\mathbb{C})$ relates to the preservation of circles and $\mathrm{PSL}_3(\mathbb{R})$

relates to the preservation of geodesics, while $\mathrm{SO}_3(\mathbb{R})$ relates to the previous two and also to isometries and preservation of area. The connection is established by their realisations as certain subgroups of $\mathrm{Diff}^1_+(\mathbb{S}^2)$ which we now introduce, along with some other canonical groups featuring in the paper by Kwakkel and Tal (30). We then briefly review some of their results in order to motivate and introduce ours.

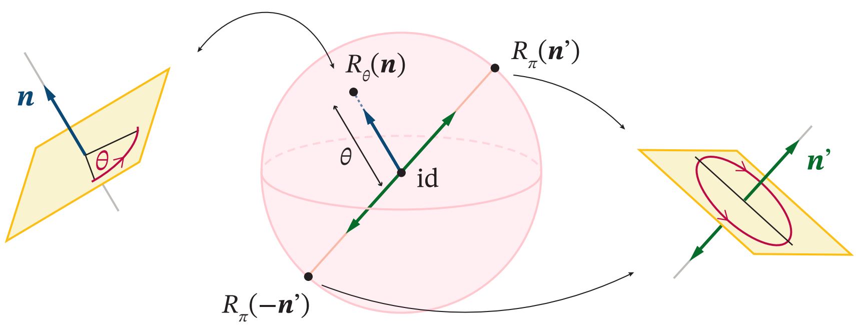
The rotation group

On the one hand, it is widely known that the isometries of Euclidean space $\mathbb{R}^{n+1}$ are affine transformations of the form $P \mapsto AP + b$, where $A \in \mathrm{O}_{n+1}(\mathbb{R})$ is an orthogonal matrix and b is a fixed vector. If one such transformation is required to preserve orientation, then actually $A \in \mathrm{SO}_{n+1}(\mathbb{R}) = \{A : A \text{ is orthogonal and of determinant } 1\}$. If it is further required to preserve $\mathbb{S}^n$ as a set, then $b = 0$. In particular, one thus obtains $\mathrm{SO}_3(\mathbb{R})$ as the compact Lie group of isometries of the round 2-sphere.

On the other hand, denote by $R_\theta(\boldsymbol{n})$ the anticlockwise rotation of angle θ around the axis positively oriented by the unit vector $\boldsymbol{n}$. Then, there exists a matrix $A \in \mathrm{SO}_3(\mathbb{R})$ such that $AV = R_\theta(\boldsymbol{n})(V)$ for every V. Conversely, given $A \in \mathrm{SO}_3(\mathbb{R})$ one may find an unit vector $\boldsymbol{n}$ and an angle $\theta \in [0, \pi]$ such that this identity holds. Namely, the solutions of $A\boldsymbol{n} = \boldsymbol{n}$ and $1 + 2\cos\theta = \operatorname{tr} A$ subject to $|\boldsymbol{n}| = 1$. Furthermore, $\boldsymbol{n}$ is unique when $0 < \theta < \pi$, while $R_\pi(\boldsymbol{n}) = R_\pi(-\boldsymbol{n})$ and $R_0(\cdot) = \mathrm{id}$ no matter which axis is chosen — *cf.* Vvedensky and Evans (48) for the actual calculations.

This allows one to define a continuous and surjective parameterisation of $\mathrm{SO}_3(\mathbb{R})$ on the closed ball $\overline{\mathbb{B}}_\pi(O) \subset \mathbb{R}^3$, by sending the origin O to id and each point of the form $\theta\boldsymbol{n}$ — where $0 < \theta \leq \theta$ and $\boldsymbol{n}$ is of unit length — to $R_\theta(\boldsymbol{n})$. This is the so-called *axis-angle representation*, and it is injective except for the ball's spherical surface, on which antipodal points become identified, as suggested by Figure 1.4.

Figure 1.4 – Rotations can be naturally identified with points in the ball $\overline{\mathbb{B}}_\pi(O)$ with antipodal surface points identified, which is the space $\mathbb{RP}^3$. The orange segment running across the ball is a noncontractible loop in this space.



A solid ball with its surface's antipodal points identified is a model for the *projective space* $\mathbb{RP}^3$, so the above parametrisation descends to a homeomorphism between $\mathbb{RP}^3$ and $\mathrm{SO}_3(\mathbb{R})$, implying the latter to have the same topology as that of three-dimensional projective space. In particular, it is orientable and connected, but not simply connected, since a loop running inside the ball while connecting antipodal points on the surface cannot be continuously shrunken to a point.

For the geometric reasons just discussed, we denote by $\mathrm{Rot}(\mathbb{S}^2)$ the subgroup of $\mathrm{Homeo}_+(\mathbb{S}^2)$

induced by the action of $SO_3(\mathbb{R})$, and name it the *rotation group*. A handful other ways exist to understand and represent rotations, the parameterisation by Euler angles and usage of the unit quaternions $\mathbb{S}^3$ being worth of mention (*cf. e.g.* (1)).

Clearly, $\mathrm{Rot}(\mathbb{S}^2)$ is transitive: given two distinct and non antipodal points $P, Q \in \mathbb{S}^2$, let Π be the unique plane through the origin containing them. Then, $\Pi \cap \mathbb{S}^2$ is a great circle upon which P, Q both lie and subtend an arc of length $\theta < \pi$. Fixed such arc, let θ be the angle of rotation, and choose as the axis $\boldsymbol{n}$ the vector $\overrightarrow{OP} \times \overrightarrow{OQ}$ normalised. Then, $Q = R_\theta(\boldsymbol{n})(P)$. When P, Q are antipodal, any rotation of angle π around an axis orthogonal to any of the infinitely many planes containing both points and the origin maps P to Q. Furthermore, $\mathrm{Rot}(\mathbb{S}^2)$ is closed, for it is compact. Our goal is now to clarify the following statement by Kwakkel and Tal (30):

> The rotation group is *minimal* among closed and transitive subgroups of $\mathrm{Homeo}_+(\mathbb{S}^2)$.

More precisely, we shall establish that any *compact and transitive group* of orientation preserving sphere homeomorphisms must be conjugate to the whole rotation group. To do so, we first recall a few terminologies.

Let G be a Lie group acting smoothly on a manifold $\mathcal{M}$. If the action is transitive, then $\mathcal{M}$ is said to be a *homogeneous G-space*. For a given set $S \subset \mathcal{M}$, the *stabiliser* of S in G is

$$\mathrm{Stab}_G\, S \stackrel{\text{def}}{=} \{g \in G \,:\, g \cdot p = p \text{ for every } p \in S\}\,,$$

where $\cdot$ denotes the action. It is always a (topologically) closed subgroup of G, and hence a Lie subgroup on its own. As it turns out, homogeneous G-spaces can be canonically described in terms of stabilisers, as conveyed below. Chapter 21 of the book by Lee (35) contains the relevant statements in full generality along with the pertinent arguments.

1.12 Proposition. Assume that $\mathcal{M}$ is a homogeneous G-space and let $p \in \mathcal{M}$ be any given point. Then, $G\,/\,\mathrm{Stab}_G\{p\}$ admits a unique smooth manifold structure such that it is also a homogeneous G-space under left multiplication. Furthermore, this manifold is diffeomorphic to $\mathcal{M}$ via an equivariant map explicitly given by $g \cdot \mathrm{Stab}_G\{p\} \mapsto g \cdot p$.

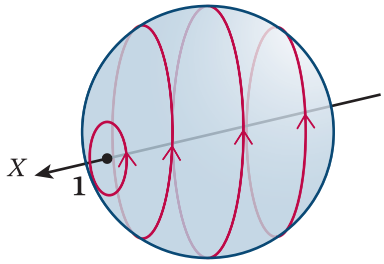
Clearly, $\mathbb{S}^2$ is a homogeneous $SO_3(\mathbb{R})$-space. If we wish to realise it in the fashion described by Proposition 1.12, we must characterize the stabiliser of one of its points, say $\mathbf{1}$ for concreteness. From the axis-angle parameterisation, these must be the matrices representing rotations around the X-axis. Explicitly,

$$[1.5] \qquad \mathrm{Stab}_{SO_3(\mathbb{R})}\{\mathbf{1}\} = \left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} : 0 \leq \theta < 2\pi \right\}.$$

All the relevant information is contained in the right bottom square block of such matrices, which consists of prescribing the angle of rotation on one (and hence all) of the invariant circles orthogonal to the X-axis that foliate $\mathbb{S}^2 \setminus \{\pm\mathbf{1}\}$ and are shown in Figure 1.5. These blocks, in turn, are naturally identified with $SO_2(\mathbb{R})$, so it is usual to write $\mathbb{S}^2 \cong SO_3(\mathbb{R})\,/\,SO_2(\mathbb{R})$.

Now, let $G \subset SO_3(\mathbb{R})$ be a closed subgroup which also acts transitively on $\mathbb{S}^2$. Then, $\mathrm{Stab}_G\{\mathbf{1}\}$ is a subgroup of [1.5], being thus identified with a subgroup of $SO_2(\mathbb{R})$. However, further recalling that

Figure 1.5 – A rotation around the X-axis (or any axis, for that matter) fixes an antipodal pair and induces planar rotations of the same angle θ in all of the invariant parallel circles obtained by intersection of the sphere with planes orthogonal to the axis.



matrices of $\mathrm{SO}_2(\mathbb{R})$ can be represented as multiplication by $e^{i\theta}$, one sees that $\mathrm{Stab}_G\{\mathbf{1}\}$ is isomorphic — as a topological group — to a multiplicative subgroup of the (complex) unit circle. However, such subgroups are widely known to be either finite cyclic or dense.

Since G was assumed to act transitively on the sphere, in particular its restriction to the invariant circles shown in Figure 1.5 is also transitive, so it cannot be cyclic of finite order. It must thus contain a dense set of possible rotation angles. However, as G was also assumed closed, it must actually contain all angles. In other words, $\mathrm{Stab}_G\{\mathbf{1}\} \simeq \mathrm{SO}_2(\mathbb{R})$ as well. Proposition 1.12 then implies G to be the whole of $\mathrm{SO}_3(\mathbb{R})$.

The sought conclusion can therefore be drawn from a result first announced by Hungarian mathematician Kerékjártó (26), according to which every compact subgroup H of $\mathrm{Homeo}(\mathbb{S}^2)$ is topologically conjugate to a closed subgroup of $\mathrm{Rot}(\mathbb{S}^2)$. A modern treatise on the proof was published by Kolev (27). Broadly, the idea is to build up from compact subgroups of $\mathrm{Homeo}(\mathbb{S}^1)$ to $\mathrm{Homeo}(\mathbb{S}^2)$, passing through $\mathrm{Homeo}(\mathbb{D}^2)$ halfway. The proof ends with an analysis of several possibilities for the group H, one of which is when it preserves orientation and acts transitively.

Curiously, this is the case that occupied most of Kerékjártó's work, but is now the one of shorter exposition, due to all the Riemannian geometry machinery developed ever since. Of course, this does not mean that the result has become simpler, only that many of the toilsome parts are now systematically conveyed in standard textbooks on the subject. For this reason, we now sum up the above discussion along with a specialised sketch of the argument.

1.13 Theorem [Kerékjártó, Kolev]. Let $G \subset \mathrm{Homeo}_+(\mathbb{S}^2)$ be a compact and transitive group. Then, G is conjugate to $\mathrm{Rot}(\mathbb{S}^2)$.

Sketch of proof. We resort to the fact that — by Gleason's solution to Hilbert's fifth problem (20, 47) and an estimate due to Newman ((27), Sec. 5) — any compact subgroup of $\mathrm{Homeo}(\mathbb{S}^2)$ (transitive or not) is actually a Lie group. That said, in particular the action of G turns $\mathbb{S}^2$ into a G-homogeneous space, diffeomorphic to the abstract surface $\mathcal{M} \stackrel{\text{def}}{=} G \,/\, \mathrm{Stab}_G\{\mathbf{1}\}$. Since this surface is compact, it admits a G-invariant Riemannian metric, say $\langle\!\langle \cdot, \cdot \rangle\!\rangle$, and this metric has constant scalar curvature (*cf. e.g.* (34), Corollary 3.18 and Exercise 8.24). However, in the particular case of a surface, the scalar curvature determines the Gaussian curvature as well. Since we know that $\chi(\mathcal{M}) = 2$, this curvature must be positive. Therefore, some positive constant c exists such that $c\langle\!\langle \cdot, \cdot \rangle\!\rangle$ has constant Gaussian curvature 1, while remaining G-invariant. The Killing-Hopf theorem thus yields a Riemannian isometry $\Phi : (\mathcal{M}, c\langle\!\langle \cdot, \cdot \rangle\!\rangle) \xrightarrow{\sim} (\mathbb{S}^2, \langle \cdot, \cdot \rangle)$, which in turn induces an conjugation between G and a closed and transitive subgroup of $SO_3(\mathbb{R})$. By the arguments precedent to this theorem's statement, the latter must actually be the full group $SO_3(\mathbb{R})$. ▨

For the reasons presented in this section, Kwakkel and Tal named $\mathrm{Rot}(\mathbb{S}^2)$ as the *kernel subgroup*, its extensions as *homogeneous groups* and posed the following.

The kernel subgroup problem. *Classify all homogeneous subgroups of* $\mathrm{Homeo}_+(\mathbb{S}^2)$ *and / or all homogeneous subgroups of* $\mathrm{Diff}^1_+(\mathbb{S}^2)$ *up to their uniform closures.*

We now turn our attention to some such groups.

Area preserving actions

Let $\mathcal{M}$ be a compact manifold, possibly with boundary, and call *good* a nonatomic Borelian probability μ of full support on $\mathcal{M}$. Then, we have the following.

1.14 Remark. If $\mathcal{M}$ is a compact manifold and a transitive subgroup $G \subset \mathrm{Homeo}(\mathcal{M})$ preserves a Borelian probability μ, then μ must be good.

Proof. Suppose first that μ has an atom $\{p\}$. Then, we may fix infinitely distinct points $\{p_n\}_{n\in\mathbb{N}}$ in $\mathcal{M}$ and, for each of them, obtain by transitivity $g_n \in G$ such that $p_n = g_n(p)$. Since μ is preserved by G, this implies $\mu(\{p_n\}_{n\in\mathbb{N}}) = \sum_{n\in\mathbb{N}} \mu\{p\} = +\infty$, a contradiction. On the other hand, suppose that there is a topological ball $B \subset \mathcal{M}$ of null measure. By transitivity of G and compacity of $\mathcal{M}$, one obtains finitely many maps $g_1, \dots, g_k \in G$ such that $\mathcal{M} = \bigcup_{j=1}^k g_j(B)$. Consequently, $1 = \mu(\mathcal{M}) \leq k\mu(B) = 0$, another contradiction. ▨

Consequently, good probabilites are the only ones that *may* be preserved in the transitive setting. The following result, which Fathi claims (on p. 53 of (12)) to be proven in Section II of (43), states that it is enough to study only one such measure as a model. We briefly remark that — in their more general contexts of origin — good measures may be defined in manifolds with boundary, but since such spaces cannot support transitive groups of homeomorphism, we do not bother to include the relevant statements.

1.15 The Oxtoby-Ulam Theorem. Let $\mathcal{M}$ be a closed manifold. If μ and ν are two good Borelian probabilities on $\mathcal{M}$, then there exists a homeomorphism $\Phi : \mathcal{M} \to \mathcal{M}$ such that $\mu = \Phi^*\nu$. Moreover (33), Φ can be assumed to be isotopic to the identity.

In the case of a Riemannian manifold, the natural choice of a good probability is the normalised volume form yielded by the metric. Albeit its explicit form is nowhere used, the area form corresponding to the round metric in $\mathbb{S}^2$ is

$$Z\, \mathrm{d}X \wedge \mathrm{d}Y - Y\, \mathrm{d}X \wedge \mathrm{d}Z + X\, \mathrm{d}Y \wedge \mathrm{d}Z\,.$$

The above 2-form pulls back under the usual parameterisation $(\cos\theta \sin\phi\,,\ \sin\theta\sin\phi\,,\ \cos\phi)$ of $\mathbb{S}^2$ by spherical coordinates $(\theta, \phi) \in [0, 2\pi] \times [0, \pi]$ to the familiar expression $\sin\phi\ \mathrm{d}\theta\mathrm{d}\phi$. We shall thus refer from now on to the Borelian probability λ, with density given by $\mathrm{d}\lambda = (4\pi)^{-1} \sin\phi\ \mathrm{d}\theta\mathrm{d}\phi$, as the *Lebesgue measure on* $\mathbb{S}^2$.

Lastly, for any given Borelian probability μ on a manifold $\mathcal{M}$ and any subgroup G of $\mathrm{Homeo}\,(\mathcal{M})$, let us denote by G_μ the subgroup of G consisting of those maps preserving μ. Since integration commutes with uniform convergence, this is *always* a closed subgroup. In the setting of good probabilities, it turns out to be maximal as well, a result proven by Le Roux (33) which we now quote.

1.16 Theorem [Le Roux]. Let $\mathcal{M}$ be a triangulable topological manifold of dimension ≥ 2, which is not assumed to be oriented and may or may not have boundary. Then, for every good Borelian probability μ on $\mathcal{M}$, the group $\mathrm{Homeo}_{\mu,0}(\mathcal{M})$ is maximal in $\mathrm{Homeo}_0(\mathcal{M})$.

Above, as in the rest of this essay, the subscript 0 indicates the identity component of a topological group, and may the replaced by + in the orientable case. In particular, when $\mathcal{M} = \mathbb{S}^2$ one readily sees that $\mathrm{Homeo}_{\lambda,+}(\mathbb{S}^2)$ is a maximal homogeneous group. We remark that an independent proof of this result is given by Kwakkel and Tal for the 2-sphere. This result strongly opposes to the one by Giblin and Markovic, because

- $\mathrm{Homeo}_{\mathrm{Leb},+}(\mathbb{S}^1)$ coincides with $\mathrm{Rot}(\mathbb{S}^1)$, while $\mathrm{Homeo}_{\lambda,+}(\mathbb{S}^2)$ extends $\mathrm{Rot}(\mathbb{S}^2)$ properly, a prototypical example being the time 1 flow of a conservative vector field;
- $\mathrm{Homeo}_{\mathrm{Leb},+}(\mathbb{S}^1)$ is properly contained in the larger $\mathrm{Möb}(\mathbb{S}^1)$, which in turn is maximal and possesses elements that do not preserve Lebesgue measure (for example, the one pictured in Figure 1.2).

Theorem 1.16 yields a machinery to produce subgroups of $\mathrm{Homeo}_+(\mathbb{S}^2)$ that are not closed. For instance, $\langle \mathrm{Homeo}_{+,\lambda}(\mathbb{S}^2) \cup \{h\} \rangle$ — the group generated upon adjunction of some $h \in \mathrm{Diff}^1_+(\mathcal{M})$ that does not preserve area to the area preserving maps.

Antipodal actions

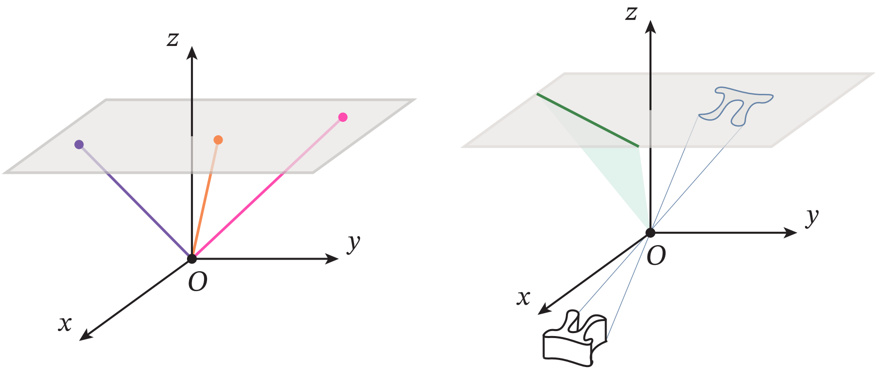
Projective geometry has its roots in the studies of perspective conduced by Dürer in the 16$^{\text{th}}$ century, and is related to the study of the properties of a figure that are preserved under projection on a screen, from the point of view of an observer seating at the origin of the space. For that reason, each line through the origin $O \in \mathbb{R}^3$ is called a *projective point*, and interpreted as a light beam emanating from it (7). Since each such line is uniquely determined by a nonzero direction vector $\boldsymbol{v}$, and any two direction vectors for the same line are scalar multiples of each other, one arrives at the definition of the *projective plane* as:

$$\mathbb{RP}^2 \overset{\text{def}}{=} \left\{ [\boldsymbol{v}] \,:\, \boldsymbol{v} \in \mathbb{R}^3 \setminus \{\mathbf{0}\} \right\},$$

where the equivalence classes $[\boldsymbol{v}]$ are determined by the relation $\boldsymbol{v} \sim \boldsymbol{w}$ if, and only if, $\boldsymbol{w} = \lambda \boldsymbol{v}$ for some $\lambda \neq 0$. Projective figures are then defined as subsets of $\mathbb{RP}^2$. In particular, *projective lines* are defined as planes of 3-space containing the origin — or rather, their images under the natural projection $\pi \,:\, \mathbb{R}^3 \twoheadrightarrow \mathbb{RP}^2$, as Figure 1.6 suggests. This projection is used to topologise $\mathbb{RP}^2$, thus yielding a Hausdorff and second countable space.

In Figure 1.6, the chosen "screen" was the plane $\{z = 1\}$. This is the so-called *standard embedding plane*, for it allows one to identify all of $\mathbb{RP}^2$ with a subset of the Euclidean plane, *except* for the projective line $\{z = 0\}$ parallel to the embedding plane, the so-called *ideal line*. This actually

Figure 1.6 – In projective geometry, Euclidean lines become points and figures are determined by beams of Euclidean lines. In particular, Euclidean planes containing the origin become projective lines.



amounts to prescribe a coordinate chart $\{U_z, \Phi_z\}$ in $\mathbb{RP}^2$, where

$$U_z = \{[x,y,z] \, : \, z \neq 0\} \text{ and } \Phi_z \, : \, [x,y,z] \mapsto \left(\frac{x}{z}, \frac{y}{z}\right).$$

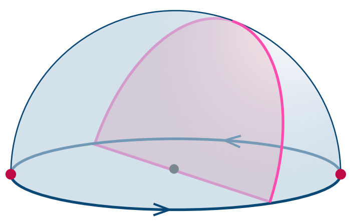
Analogous procedures for the x and y variables yield a smooth atlas, turning $\mathbb{RP}^2$ into a smooth manifold.

Consider now the group $\mathrm{PSL}_3(\mathbb{R})$, obtained from the general linear group $\mathrm{SL}_3(\mathbb{R})$ of 3×3 real matrices of unit determinant upon declaring each A and $-A$ as equivalent. Then, the following action via *projective transformations* is well defined:

$$\text{[1.6]} \qquad \big(\pm A, [\boldsymbol{v}]\big) \in \mathrm{PSL}_3(\mathbb{R}) \times \mathbb{RP}^2 \mapsto [A\boldsymbol{v}] \in \mathbb{RP}^2 \, .$$

This action gives rise to a group of homeomorphisms $\mathrm{Lin}(\mathbb{RP}^2) \subset \mathrm{Homeo}\,(\mathbb{RP}^2)$. To explain how it is related to the 2-sphere, we first observe that each Euclidean line through the origin of $\mathbb{R}^3$ meets $\mathbb{S}^2$ at a pair of antipodal points, thus defining a 2:1 smooth covering map $p \, : \, \mathbb{S}^2 \twoheadrightarrow \mathbb{RP}^2$ which identifies antipodal points. A fundamental domain for it is the sphere's northern hemisphere, along with the equator with antipodal points identified, as shown in Figure 1.7. From it, one sees that $\mathbb{RP}^2$ is not orientable, for a Möbius band may be found within it.

Figure 1.7 – A spherical model for the projective plane: on it, projective lines correspond to great circles, which are arcs of geodesics.



Clearly, the group of deck transformations for p is $\{\pm \,\mathrm{id}_{\mathbb{S}^2}\} \simeq \mathbb{Z}_2$. Each map in $\mathrm{Homeo}\,(\mathbb{RP}^2)$ therefore has two lifts to $\mathbb{S}^2$, only one of which is orientation preserving. We thus define $\mathrm{Ant}(\mathbb{S}^2)$ to be the subgroup of $\mathrm{Homeo}_+(\mathbb{S}^2)$ consisting of all orientation preserving lifts of projective homeomorphisms. The defining property for $f \in \mathrm{Homeo}_+(\mathbb{S}^2)$ to be in this group is, of course, that $f(-P) = -f(P)$ holds for every $P \in \mathbb{S}^2$. As a consequence, the group $\mathrm{Lin}(\mathbb{RP}^2)$ just defined lifts to a subgroup $\mathrm{Lin}(\mathbb{S}^2)$ of $\mathrm{Ant}(\mathbb{S}^2)$.

In the spherical model, projective lines lift to great circles, since those are precisely the intersections of $\mathbb{S}^2$ with planes containing the origin. Therefore, $\mathrm{Lin}(\mathbb{S}^2)$ must permute such circles, thus establishing the aforementioned connection between $\mathrm{PSL}_3(\mathbb{R})$ and preservation of geodesics. In particular, $\mathrm{Lin}(\mathbb{S}^2)$ is homogeneous, with the action of $\mathrm{Rot}(\mathbb{S}^2)$ descending to the action [1.6] of $\mathrm{PSO}_3(\mathbb{R})$ as the maximal compact subgroup of $\mathrm{PSL}_3(\mathbb{R})$.

These groups play in the spherical setting the same role as the cyclic covers did in $\mathbb{S}^1$. Some of the results announced by Kakkel and Tal concerning them ((30), Lemma 2.3, Theorems 5 and 6) are that $\mathrm{Lin}_\lambda(\mathbb{S}^2)$ coincides with $\mathrm{Rot}(\mathbb{S}^2)$ and that $\mathrm{Ant}_\lambda(\mathbb{S}^2)$ is maximal in $\mathrm{Ant}(\mathbb{S}^2)$.

The Möbius action

Consider the group $\mathrm{PSL}_2(\mathbb{C})$ obtained from $\mathrm{SL}_2(\mathbb{C})$ upon declaring A and $-A$ as equivalent. If $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in \mathrm{SL}_2(\mathbb{C})$, it acts on the extended complex plane by fractional linear transformations as follows:

$$(\pm A, z) \in \mathrm{PSL}_2(\mathbb{C}) \times \mathbb{C} \cup \{\infty\} \mapsto M_A(z) \stackrel{\text{def}}{=} \frac{az+b}{cz+d} \in \mathbb{C} \cup \{\infty\}\,.$$

Each M_A is a homemomorphism of the extended plane, and thus induces by conjugation with Ψ_N a map in $\mathrm{Homeo}\,(\mathbb{S}^2)$. This yields the subgroup $\mathrm{Möb}(\mathbb{S}^2)$ of *Möbius transformations*.

Möbius transformations are ubiquitous in several fields of Mathematics and have been thoroughly studied, so we provide here just a small glance at the aspects relevant to forthcoming arguments. The first important result we mention is their characterisation as conformal maps.

1.17 Theorem. $\mathrm{Möb}(\mathbb{S}^2)$ consists of the orientation preserving and conformal diffeomorphisms of the 2-sphere. More precisely, Möbius transformations are the conformal automorphisms of the Riemann sphere.

Above, *Riemann sphere* means that the underlying set $\mathbb{S}^2$ is being given the structure of a complex one-dimensional manifold, with the stereographic projections comprising the holomorphic atlas. In this setting, conformal automorphism becomes synonym to biholomorphic bijection, and this is how such result is usually stated and proven a regular Complex Analysis course. Very rigorously, to make sense out of the first sentence in Theorem 1.17 above, one would need to further evoke the notorious Uniformization Theorem, for it implies all Riemannian metrics on the (*a priori* only topological and real) oriented manifold $\mathbb{S}^2$ to be conformally equivalent — *cf.* (25), Sections 2.3, 3.11 and 4.4.

Möbius transformation are known for mapping circles into circles. If one restricts to the complex plane, some of these circles might actually be lines, which correspond stereographically to circles on $\mathbb{S}^2$ passing through ∞. This stems from the fact that each Möbius transformation can be decomposed into a sequence of simpler transformations — translations, homotheties and inversions. Enlightening proofs may be found in Chapter 3 of (42).

We further remark that the following implies $\mathrm{Möb}(\mathbb{S}^2)$ to be homogeneous.

1.18 A result by Gauss [1819]. A map $R : \mathbb{S}^2 \to \mathbb{S}^2$ is a rotation of the Riemann sphere if, and only if, R is a Möbius transformation induced by a matrix in

$$\mathrm{PSU}_2(\mathbb{C}) = \left\{ \pm \begin{bmatrix} a & b \\ -b^* & a^* \end{bmatrix} : a, b \in \mathbb{C} \text{ and } |a|^2 + |b|^2 = 1 \right\},$$

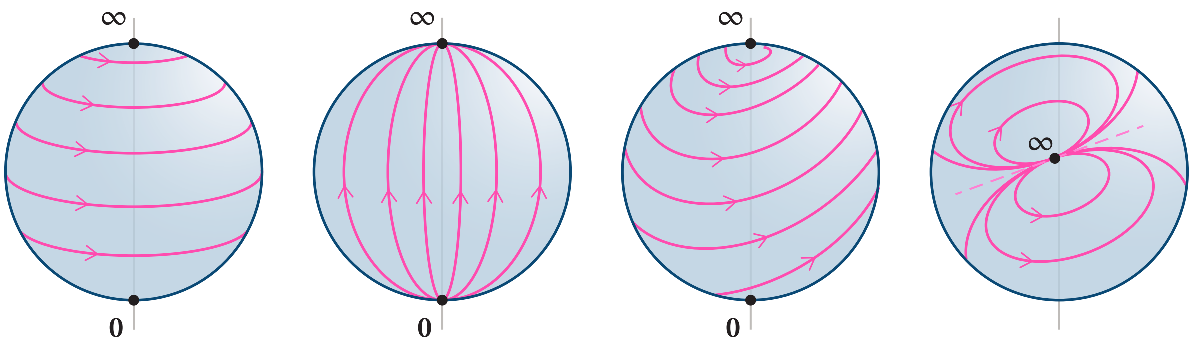
which is the maximal compact subgroup of $\mathrm{PSL}_2(\mathbb{C})$.

The fixed points of a Möbius transformations are solutions to the quadratic equation $M_A(z) = z$ over the field $\mathbb{C}$ (plus the point at infinity), so they always exist and either collapse to a single point or come in pairs. Actually, a full classification is possible in terms of the matrix A and a handful of prototypical transformations, of which we now remember:

- $z \mapsto e^{i\alpha} z$, where $\alpha \in \mathbb{R}$, is an *elliptical transformation*: on the sphere, it is a rotation around the Z-axis, thus leaving the poles fixed, parallels invariant and permuting meridians.
- $z \mapsto \rho z$, where $\rho \in \mathbb{R} \setminus \{0, 1\}$, is a *hyperbolic transformation*: on the sphere, it leaves the poles fixed, meridians invariant and permutes parallels. If $\rho > 1$, points are dragged monotonically from the South to the North Pole over meridians. If $0 < \rho < 1$, the opposite holds.
- $z \mapsto \rho e^{i\alpha} z$, where ρ and α are as above, is a *loxodromic transformation*. It is actually a combination of the two previous types, so both meridians and parallels form invariant families, but neither are fixed — instead, invariant curves spiral from one (fixed) pole to the other.
- $z \mapsto z + b$ is a *parabolic transformation*. It has ∞ as its single fixed point, and its invariant curves are circles passing through ∞ and sharing a common tangent parallel to b at that point, a configuration known as *horocyclic*.

These transformations are illustrated in Figure 1.8.

Figure 1.8 – The four prototypical Möbius transformations



$z \mapsto e^{i\alpha z}, \alpha > 0$ $\quad$ $z \mapsto \rho z, \rho > 1$ $\quad$ $z \mapsto \rho e^{i\alpha z}, \rho > 1, \alpha > 0$ $\quad$ $z \mapsto z + b, b \in \mathbb{C}$

We are now ready to provide the mentioned classification. An attempt to solve the equation yielding the fixed points shows that their multiplicity depends essentially on the absolute value of the trace of A, which is a well-defined quantity in $\mathrm{PSL}_2(\mathbb{C})$. In particular, unless $|\mathrm{tr}\, A| = 2$, M_A has two distinct fixed points, ξ_+ and ξ_-. Then, $H \in \text{Möb}(\mathbb{S}^2)$ exists such that $H(\xi_+) = \mathbf{0}$, $H(\xi_-) = \infty$ and $H \circ M_A \circ H^{-1}(z) = \mathfrak{m} z$. The complex number thus obtained is the *multiplier* of M_A, and can be computed in a number of ways. Namely,

- as $M'_A(\xi_+) = \mathfrak{m} = [M'_A(\xi_-)]^{-1}$ in terms of the complex function M_A, or
- as the solutions to $\mathfrak{m}^2 + (2 - \operatorname{tr}^2 A)\mathfrak{m} + 1 = 0$ in terms of the matrix A.

Lastly, if $|\operatorname{tr} A| = 2$, then M_A has a single fixed point ξ. This time, $H \in \text{Möb}(\mathbb{S}^2)$ exists such that $H(\xi) = \infty$ and $H \circ M_A \circ H^{-1}(z) = z + \tau$, for some $\tau \in \mathbb{C}$. In this case, the multiplier is agreed to be 1. Each Möbius transformation is then named after the prototypical map to which it is conjugate, as summarised in Table 1.1.

Table 1.1 – Classification of a Möbius transformation M_A represented by a non trivial matrix $A \in \mathrm{PSL}_2(\mathbb{C})$. According to its class, M_A is either conjugate to a translation (parabolic case) or to $z \mapsto \mathfrak{m}z$.

Trace of A	**Multipliers $\mathfrak{m}$**	**Class**
$\operatorname{tr} A \in \mathbb{R}$ and $\lvert\operatorname{tr} A\rvert < 2$	$e^{\pm i\alpha}$, $\alpha \in \mathbb{R} \setminus \{0\}$	Elliptic.
$\operatorname{tr} A = \pm 2$	1	Parabolic.
$\operatorname{tr} A \in \mathbb{R}$ and $\lvert\operatorname{tr} A\rvert > 2$	$\rho, \rho^{-1} \in \mathbb{R} \setminus \{0, 1\}$	Hyperbolic.
$\operatorname{tr} A \in \mathbb{C} \setminus \mathbb{R}$	$(\rho e^{i\alpha})^{\pm 1}$; α, ρ as above	Loxodromic.

A notorious and widely known property of $\text{Möb}(\mathbb{S}^2)$ is being a sharply 3-transitive group. As it turns out, this property completely determines the Möbius group among homogeneous subgroups of $\mathrm{Diff}^1_+(\mathbb{S}^2)$. This is established along Section 3.4 of the paper by Kwakkel and Tal. In this thesis, we offer a new — yet somewhat similar in spirit — proof of this fact in Chapter 2. Another relevant result derived by Kwakkel and Tal concerning Möbius transformations is that both $\text{Möb}_\lambda(\mathbb{S}^2)$ and $\text{Möb}(\mathbb{S}^2) \cap \text{Ant}(\mathbb{S}^2)$ reduce to $\text{Rot}(\mathbb{S}^2)$.

For they play a prominent role in subsequent chapters, we establish some further properties of Möbius transformations that shall be needed. In order to so, for any group $G \in \text{Homeo}(\mathbb{S}^2)$ we introduce the following notations for stabilizers:

[1.7] $$G_1 \overset{\text{def}}{=} \text{Stab}_G\{\infty\} \ , \ G_2 \overset{\text{def}}{=} \text{Stab}_G\{\mathbf{0}, \infty\} \ \text{ and } \ G_3 \overset{\text{def}}{=} \text{Stab}_G\{\mathbf{0}, \mathbf{1}, \infty\} \, .$$

1.19 Lemma. Given finite and nonzero points z, w, we let $M[z, w] \in \text{Möb}_2(\mathbb{S}^2)$ be the *unique* Möbius transformation fixing the poles and mapping z to w. Also, we denote $\hat{M}[z] \overset{\text{def}}{=} M[z, \mathbf{1}]$. Then,

1) The association $(x, y) \mapsto M[x, y]$ is continuous.

2) If $\mathcal{K} \subset \mathbb{S}^2$ is a nonempty compact set bounded away from $\mathbf{0}$ and $\hat{M}[x] \overset{\text{def}}{=} M[x, \mathbf{1}]$, the sets $\hat{M}[x](\mathcal{K})$ converge to $\{\infty\}$ in the Hausdorff distance as $x \to \mathbf{0}$.

Proof. For $z \in \mathbb{C}$, one has $M[x, y](z) = yz/x$.
Notice that $M[x, y]^{-1} = M[y, x]$, and also that the following relation to complex inversion holds:

[1.8] $$M[x, y](z) = (M[y, x](z^{-1}))^{-1} \, .$$

Let $((x_n, y_n))_{n \in \mathbb{N}}$ be a sequence in $\mathbb{C}^* \times \mathbb{C}^*$ converging to $(x_0, y_0) \in \mathbb{C}^* \times \mathbb{C}^*$, and let M_n, M_0 be the respective transformations. Since they all fix $\mathbf{0}$ and ∞, consider finite nonzero $z \in \mathbb{C}$.

On the one hand, if $0 < |z| \leq 1$, then [1.4] yields:

$$\hat{d}(M_n(z), M_0(z)) = \frac{2\,|M_n(z) - M_0(z)|}{\sqrt{1 + |M_n(z)|^2}\,\sqrt{1 + |M_0(z)|^2}}$$
$$< 2\,|M_n(z) - M_0(z)| = 2\left|\frac{y_n z}{x_n} - \frac{y_0 z}{x_0}\right| \leq 2\left|\frac{y_n}{x_n} - \frac{y_0}{x_0}\right| .$$

On the other hand, if $1 < |z| < +\infty$, by [1.8] and because complex inversion is an isometry of the Riemann sphere, the above calculation yields:

$$\hat{d}(M_n(z), M_0(z)) = \hat{d}\Big(\big[M_n^{-1}(z^{-1})\big]^{-1}, \big[M_0^{-1}(z^{-1})\big]^{-1}\Big) = \hat{d}(M_n^{-1}(z^{-1}), M_0^{-1}(z^{-1})) < 2\left|\frac{x_n}{y_n} - \frac{x_0}{y_0}\right| .$$

Putting together the possibilities, one obtains:

$$d_\infty(M_n, M_0) < 2\max\left\{\left|\frac{x_n}{y_n} - \frac{x_0}{y_0}\right|, \left|\frac{y_n}{x_n} - \frac{y_0}{x_0}\right|\right\} .$$

But the right-hand side of the above estimate tends to zero as $n \to +\infty$, so M_n tends to M_0 in $\mathrm{Homeo}\,(\mathbb{S}^2)$, establishing the desired continuity.

Turning to the second claim, let $d_{\mathcal{H}}(A, B)$ denote the Hausdorff distance between two nonempty compact sets of $\mathbb{S}^2$. When $B = \{\infty\}$ is a singleton, it reduces to $d_{\mathcal{H}}(A, \{\infty\}) = \max\{d(a, \infty) : a \in A\}$. Thus, given a compact set $\mathcal{K}$ such that $|z| \geq \rho > 0$ for every $z \in \mathcal{K}$ and some positive ρ, if one lets $A = \hat{M}[x](\mathcal{K})$, then each $a \in A$ takes the form $a = z/x$, where $z \in \mathcal{K}$. Therefore,

[1.9] $$d(a, \infty) \leq 2\,\hat{d}\Big(\frac{z}{x}, \infty\Big) = \frac{4}{\sqrt{1 + |z/x|^2}} \leq \frac{4}{|z/x|} = 4\left|\frac{x}{z}\right| \leq \frac{4\,|x|}{\rho} .$$

Since ρ depends only on $\mathcal{K}$ given, taking the limit as $x \to \mathbf{0}$ on the rightmost side of [1.9] yields the claimed convergence. ▨

Next,given $\mathbf{a}, \mathbf{b} \in \{\mathbf{0}, \mathbf{1}, \infty\}$, let $T_{\mathbf{ab}} \in \text{Möb}\,\mathbb{S}^2$ be the unique Möbius transformation permuting $\mathbf{a}$ and $\mathbf{b}$ and fixing the remaining special point. In symbols,

$$T_{\mathbf{ab}}(\mathbf{a}) = \mathbf{b} \quad , \quad T_{\mathbf{ab}}(\mathbf{b}) = \mathbf{a} \quad \text{and} \quad T_{\mathbf{ab}}\{\mathbf{0}, \mathbf{1}, \infty\} = \{\mathbf{0}, \mathbf{1}, \infty\} .$$

Note that the possibility $\mathbf{a} = \mathbf{b}$, and consequently $T_{\mathbf{ab}} = \mathrm{id}_{\mathbb{S}^2}$, is not being excluded, and that each such mapping is idempotent, satisfying $T_{\mathbf{ab}}^{-1} = T_{\mathbf{ba}} = T_{\mathbf{ab}}$. Albeit not explicitly used, a summary of these transformations is given in Table 1.2.

Table 1.2 – A brief description of the mappings $T_{\mathbf{ab}}$ for $a, b \in \{\mathbf{0}, \mathbf{1}, \infty\}$.

Mapping	**In** $\mathbb{C}$	**On the sphere**
$T_{\mathbf{0}\infty} = T_{\infty\mathbf{0}}$	$z \mapsto \dfrac{1}{z}$	Rotation of angle π around the X-axis.
$T_{\mathbf{01}} = T_{\mathbf{10}}$	$z \mapsto 1 - z$	Elliptic transformation with fixed points $\xi_- = \infty$ and $\xi_+ = (0.8, 0, -0.6)$ of multiplier $\mathfrak{m} = -1$.
$T_{\mathbf{1}\infty} = T_{\infty\mathbf{1}}$	$z \mapsto \dfrac{z}{z-1}$	Elliptic transformation with fixed points $\xi_- = (0.8, 0, 0.6)$ and $\xi_+ = \mathbf{0}$ of multiplier $\mathfrak{m} = 1$.

In particular, one readily sees from their real coefficients that all of them preserve Γ.

Lastly, given three finite, nonzero and distinct points a, b, c on the plane, let $\hat{M}[a,b,c]$ be the unique Möbius transformation mapping a to $\mathbf{0}$, b to $\mathbf{1}$ and c to ∞. It is widely known that this transformation is given by the cross-ratio

[1.10] $$\hat{M}[a,b,c](z) = \frac{(z-a)(b-c)}{(z-c)(b-a)} .$$

The association $(a,b,c) \mapsto \hat{M}[a,b,c]$ is continuous for finite, nonzero and mutually distinct points. This may be seen either from its matrix form or from the rather clutered yet clearly continuous expression

[1.11] $$\hat{M}[a,b,c] = T_{\mathbf{1}\infty} \circ \hat{M}[T_{\mathbf{1}\infty} \circ \hat{M}[T_{\mathbf{01}} \circ \hat{M}[a](b)] \circ T_{\mathbf{01}} \circ \hat{M}[a](c)] \circ T_{\mathbf{1}\infty} \circ \hat{M}[T_{\mathbf{01}} \circ \hat{M}[a](b)] \circ T_{\mathbf{01}} \circ \hat{M}[a] .$$

The structure of homogeneous groups

In this section we summarise some of the results announced by Kwakkel and Tal upon their work on the kernel subgroup problem, in order to arrive at our driving problem. To do so, let us first agree that a homogeneous group G has the property **P** according to whether all of its elements preserve **D**istance, **L**ebesgue measure, **G**reat circles, **A**ntipodal points or **C**ircles. For example, $\mathrm{Rot}(\mathbb{S}^2)$ has all the properties, whilst $\mathrm{Ant}_\lambda(\mathbb{S}^2)$ has properties **A** and **L**, but none of the others.

Then, given two homogeneous groups $H \subset G$, a *proper extension* edge of the form $H \overset{*}{—} G$ connecting them means that, for every property **P** that H has but G has not, the extension is *maximal* with respect to **P** *in the following sense*: for every group K *not* having property **P** and such that $H \subset K \subseteq G$, the uniform closure of both K and G coincide. For example, Theorem 1.16 by LeRoux may be restated as $\mathrm{Homeo}_{+,\lambda}(\mathbb{S}^2) \overset{*}{—} \mathrm{Homeo}_+(\mathbb{S}^2)$. With that in mind, one has the following ((30), Theorems A and C, Proposition 3.4).

1.20 The Kwakkel & Tal Classification Theorem. Let $G \subset \mathrm{Diff}^1_+(\mathbb{S}^2)$ be a homogeneous group. Then, the alternatives listed below are all mutually exclusive and exhaust the k-transitivity possibilities for $1 \leq k \leq 3$.

1) If G is 1-transitive, but not k-transitive for any other k, then $G \subset \mathrm{Ant}(\mathbb{S}^2)$.
Furthermore,
- **1.i** if it acts 1-transitively but not 2-transitively on $\mathbb{RP}^2$, then $G = \mathrm{Rot}(\mathbb{S}^2)$.
- **1.ii** If it acts 2-transitively on $\mathbb{RP}^2$, then
  - **1.ii.a** either $G = \mathrm{Lin}(\mathbb{S}^2)$, or
  - **1.ii.b** $(G_2)_0$ is nontrivial.

2) If G is 3-transitive, then
- **2.i** either $G = \mathrm{Möb}(\mathbb{S}^2)$, or
- **2.ii** $(G_3)_0$ is nontrivial.

Lastly, the following holds:

[1.12]

$$\begin{tikzcd}
 & & \mathrm{M\ddot{o}b}(\mathbb{S}^2) \arrow[ddll, dash, bend right, "*"'] \arrow[ddr, dash, bend left] & \\
 & & \mathrm{Homeo}_{\lambda,+}(\mathbb{S}^2) \arrow[dll, dash, bend right] \arrow[dl, dash] \arrow[dr, dash, "*"] & \\
\mathrm{Rot}(\mathbb{S}^2) \arrow[r, dash] \arrow[drr, dash, bend right] \arrow[ddrr, dash, bend right, "*"'] & \mathrm{Ant}_{\lambda}(\mathbb{S}^2) \arrow[rr, dash, "*"] \arrow[dr, dash, "*"] & & \mathrm{Homeo}_{+}(\mathbb{S}^2) \\
 & & \mathrm{Ant}(\mathbb{S}^2) \arrow[ur, dash, "*"'] \arrow[d, dash] & \\
 & & \mathrm{Lin}(\mathbb{S}^2) \arrow[uur, dash, bend right] &
\end{tikzcd}$$

Furthermore, the diagram contains all possible intersections between any two groups depicted on it.

It is unknown whether diagram [1.12] exhausts all homogeneous subgroups of $\mathrm{Diff}^1_+(\mathbb{S}^2)$ (or even $\mathrm{Diff}^\infty_+(\mathbb{S}^2)$, for that matter). Also, within it some extensions remained unresolved, thus resulting in the following.

Question [Kwakkel, Tal, Le Roux]. *Are the extensions*

$$\mathrm{Lin}(\mathbb{S}^2) — \mathrm{Ant}(\mathbb{S}^2)\,,\ \mathrm{Lin}(\mathbb{S}^2) — \mathrm{Homeo}_+(\mathbb{S}^2) \textit{ and } \mathrm{M\ddot{o}b}(\mathbb{S}^2) — \mathrm{Homeo}_+(\mathbb{S}^2)$$

maximal in the sense previously described?

In this essay we provide insight — from the transitivity viewpoint — into the last of the above extensions, as we shall now describe in general lines. Before doing so, we remark that Theorem B could in principle be derived from Theorem A along with an abstract and slightly foggy result of Kwakkel and Tal (namely, Theorem B in (30), along with its Corollary). However, we favour the more explicit construction presented here, since in the end of the day both rely on Nielsen-Thurston classification theory.

1.3 Outline

In Section 2.1, we state and derive Theorem C, which is a characterisation of the Möbius group in terms of transitivity. 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.

In the sequence, we focus on groups properly extending $\mathrm{M\ddot{o}b}(\mathbb{S}^2)$. Given one such subgroup G, we consider its subgroups G_k, $1 \le k \le 3$, as defined in [1.7]. Our final goal is to conclude that G_3 is (one) transitive, following the steps summarised in Figure 1.9.

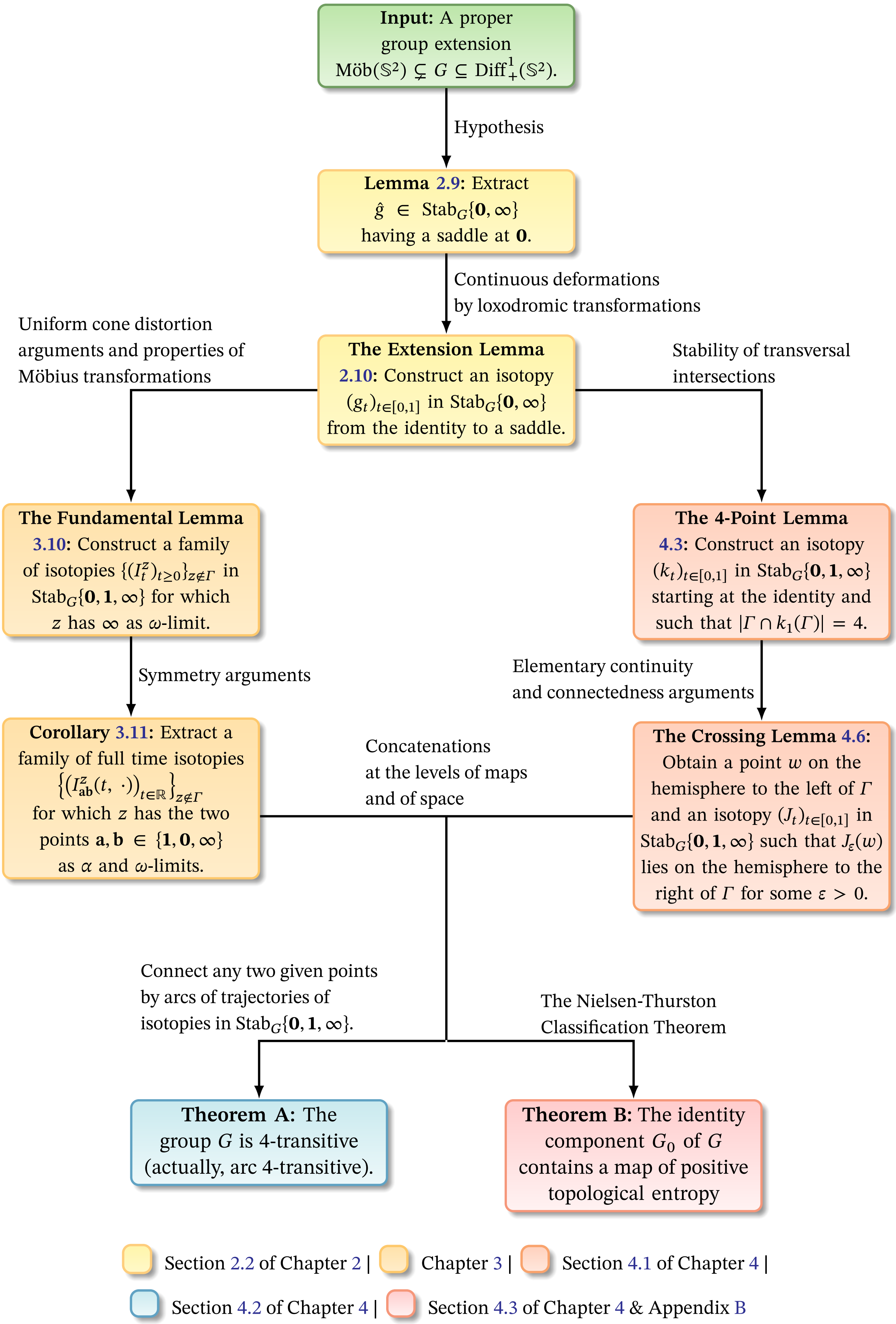
We begin with an Extension Lemma 2.10 at the end of Section 2.2, stating that G_2 must contain an isotopy between the identity and a map having a hyperbolic saddle point at the (fixed) South Pole. This is achieved by starting with a nonconformal map and continuously parameterising rotations and homotheties in such a way as to create eigendirections and modulate the corresponding eigenvalues.

From this starting point, we fix the privileged reference meridian Γ and promote two parallel processes. On the one hand, in Chapter 3 we prove our main result, the Fundamental Lemma 3.10. It states that any point outside of Γ admits a full time $\mathcal{I}G_3$-isotopy having any given pair of points in $\{\mathbf{0}, \mathbf{1}, \infty\}$ as α and ω-limits.

On the other hand, in Section 4.1 we prove that there exists a finite time $\mathcal{I}G_3$-isotopy for which some point on Γ starts at one side of it and ends at the other side. In particular, any point in $\Gamma \setminus \{\mathbf{0}, \mathbf{1}, \infty\}$ can then be moved out of Γ by $\mathcal{I}G_3$-isotopies

In Sections 4.2 and 4.3 at the end of Chapter 4, we show our main theorems. Theorem A is derived upon combining isotopies of the types previously described and concluding that all but three points of the sphere are actually arc connected in the sense of Definition 1.5, yielding the 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.

Figure 1.9 – Outline of the key steps towards theorems A and B, along with their interdependencies.



Chapter 2

Möbius and its extensions

2.1 A characterization of the conformal group

This section is devoted to understand how sharp 3-transitivity is a *defining* property of the Möbius group among *homogeneous* groups of diffeomorphisms. More precisely, we shall establish the following.

Theorem C. *Let $G \subset \mathrm{Diff}^1_+(\mathbb{S}^2)$ be a homogeneous group. If G is sharply 3-transitive, then $G =$* $\mathrm{Möb}(\mathbb{S}^2)$.

We reinforce that this is not a new result, and could be deduced from Section 3.4 in (30). However, 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, proof of this fact. Before proceeding to the actual argument, we quickly recall some elementary equivalences of conformality from linear algebra.

2.1 Lemma. Let A be a 2×2 real matrix of positive determinant. The following are equivalent:

(i) A is conformal.

(ii) $A = \lambda R$, where R is a rotation matrix and $\lambda > 0$.

(iii) $A(\mathbb{S}^1)$ is a circle.

Proof. This result relies on the fact that the invertible matrix A has a *polar decomposition* $A = UP$, where

- U is orthogonal and
- P is *positive* – meaning it admits an orthonormal frame of eigenvectors, all of which are associated to positive eigenvalues.

In particular, if $\{u, v\}$ is an orthonormal frame such that $Pu = \lambda u$ and $Pu = \mu v$, with $\lambda, \mu > 0$, then the vectors $u + v$ and $u - v$ are also mutually orthogonal, as well as $U(u)$ and $U(v)$. If we further assume (i), $A(u + v)$ and $A(u - v)$ must also be mutually orthogonal, yielding

$$0 = \langle A(u+v), A(u-v) \rangle = \lambda^2 - \mu^2 = (\lambda - \mu)(\lambda + \mu) .$$

This implies $\mu = \lambda > 0$ and $P = \lambda\,\mathrm{id}$. Since $\det A > 0$, U is an orthogonal matrix of positive determinant, meaning it is a rotation. This proves that (i) implies (ii).

If we now assume (ii), since $\mathbb{S}^1$ is invariant under rotations, $A(\mathbb{S}^1)$ is a circle of radius λ, proving that it implies (iii). Since (ii) straightforwardly implies (i), all that is left to verify is that (iii) implies (ii).

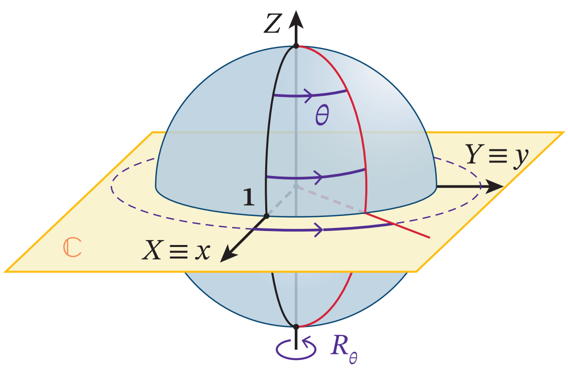
But, if $\lambda > 0$ is such that $A(\mathbb{S}^1) = \{w : |w - w_0| = \lambda\}$, it must be the case that $w_0 = 0$. Indeed, since $z \in \mathbb{S}^1$ if, and only if, $-z \in \mathbb{S}^1$, $|Az - w_0|^2 = |Az + w_0|^2$ must hold for every $z \in \mathbb{S}^1$, which, in turn, implies $\langle Az, w_0 \rangle = 0$ for every $z \in \mathbb{S}^1$ and, consequently, $w_0 = 0$. Thus, for every $z \in \mathbb{S}^1$, $|Az| = \lambda$. In particular, A/λ is an orientation preserving planar isometry fixing the origin. In other words, a rotation. This establishes (ii). ▨

It will be important – in this and all the constructions that follow – to consider rotations of a special kind: for $\theta \in (-\pi, \pi)$, let $R_\theta \in \mathrm{Rot}_2(\mathbb{S}^2)$ denote the rotation of angle θ around the Z axis. Its expression in coordinates, relative to the chart given by stereographic projection from the North Pole, coincides with the planar rotation of the same angle:

[2.1] $$\Psi_N \circ R_\theta \circ \Psi_N^{-1}(z) = e^{i\theta} z\,.$$

For this reason, both R_θ and its planar counterpart are denoted equally and referred to without distinction. Also, any given planar rotation may be though of as a mapping in $\mathrm{Rot}_2(\mathbb{S}^2)$, induced by the relation [2.1], as illustrated in Figure 2.1.

Figure 2.1 – Rotations around the Z axis are naturally identified with their planar counterparts, as well as $\mathrm{D}R_\theta(\mathbf{0})$.



Even more, if the (round) sphere $\mathbb{S}^2$ is considered as an embedded submanifold of $\mathbb{R}^3$, its tangent plane $T_{\mathbf{0}}\mathbb{S}^2$ at the South Pole is horizontal and generated by the ambient space tangent vectors $\{{}^{\partial}\!/\!_{\partial X}|_{(0,0,-1)}\,, {}^{\partial}\!/\!_{\partial Y}|_{(0,0,-1)}\}$, being naturally identified with the subspace $\mathbb{R}^2 \times \{0\} \simeq \mathbb{R}^2$. In stereographic coordinates, it is isomorphic to $\mathbb{R}^2$ via $\mathrm{D}\Psi_N(\mathbf{0})$, under which

[2.2] $$\mathrm{D}\Psi_N(\mathbf{0})\big({}^{\partial}\!/\!_{\partial X}|_{(0,0,-1)}\big) = \tfrac{1}{2}\,{}^{\partial}\!/\!_{\partial x} \quad \text{and} \quad \mathrm{D}\Psi_N(\mathbf{0})\big({}^{\partial}\!/\!_{\partial Y}|_{(0,0,-1)}\big) = \tfrac{1}{2}\,{}^{\partial}\!/\!_{\partial y}\,.$$

The differentials of a diffemorphism f fixing $\mathbf{0}$ thus act identically – up to a mere homothety – both on the plane and on $T_{\mathbf{0}}\mathbb{S}^2$. In particular, $\mathrm{D}R_\theta(\mathbf{0})$ acts as a rotation of the same angle θ, so we may write, in a slight abuse of notation, $\mathrm{D}R_\theta(\mathbf{0}) = R_\theta$.

2.2 Lemma. The mapping $t \in (-\pi, \pi) \mapsto R_t$ is continuous with respect to the C^1 topology of $\mathrm{Diff}^1_+(\mathbb{S}^2)$.

Proof. We will verify first that $R_t \to \mathrm{id}$ as $t \to 0$. To do so, notice that any compact subset $K \subset \mathbb{S}^2$ may be decomposed into two compact pieces, one of which is contained within $\mathbb{S}^2 \setminus \{\infty\}$ and the other of which is contained within $\mathbb{S}^2 \setminus \{\mathbf{0}\}$. Since R_t fixes the poles, we may thus prove convergence over compacts in each of the charts $\left(\mathbb{S}^2 \setminus \{\infty\}, \Psi_N\right)$ and $\left(\mathbb{S}^2 \setminus \{\mathbf{0}\}, \Psi_S\right)$ separately. By [2.1], these further reduce to verify ordinary C^1 convergence over compact sets of the plane.

But this is simple, since $e^{it} \to 1$ as $t \to 0$ and, also, the rotation matrix of R_t converges to the identity as $t \to 0$. Lastly, upon noticing that $R_{\theta+t} = R_t \circ R_\theta$ and recalling that composition with a fixed element is continuous in $\mathrm{Diff}^1_+(\mathbb{S}^2)$, it is seen that $R_{\theta+t} \to R_\theta$ as $t \to 0$, which amounts to the desired continuity at an arbitrary $\theta \in (-\pi, \pi)$. ▨

Now, the first step towards proving Theorem C is to show that the subgroup G_2 fixing the poles possesses a property known, *a priori*, to hold for the actual $\mathrm{Möb}_2(\mathbb{S}^2)$: it permutes parallels. This argument does not depend on differentiability, being actually purely topological, and will follow from the two remarks ahead.

To set the context up, we recall that a Jordan curve λ is a simple and closed path. The Jordan Curve Theorem states that, given one such path on the plane, it divides the space into two disjoint, open and connected components, its *interior* $\operatorname{int}\lambda$ and its *exterior* $\operatorname{ext}\lambda$, both of which share (the image of) λ as their common boundary. Also, the first of them is bounded, whilst the second is unbounded.

2.3 Remark. Let λ be a planar Jordan curve which is *not* a circle. Then, it cannot properly contain a circle either.

Proof. Let γ be a circle such that $\gamma \subset \lambda$, properly. Then, $\lambda^c \subset \gamma^c$, properly. More precisely,

$$\lambda^c = \operatorname{int}\lambda \sqcup \operatorname{ext}\lambda \subset \operatorname{int}\gamma \sqcup \operatorname{ext}\gamma = \gamma^c\ ,$$

implying $\operatorname{ext}\lambda \subset \operatorname{ext}\gamma$, for an unbounded connected component cannot be fully contained within a bounded one.

Now, if γ is centred at p_0, by compacity we may fix p_m and p_M in λ such that $|p_m| = \min_{p\in\lambda}|p - p_0|$ and $|p_M| = \max_{p\in\lambda}|p - p_0|$. Thus, at least one among p_m or p_M does not belong to γ, or λ itself would be a circle.

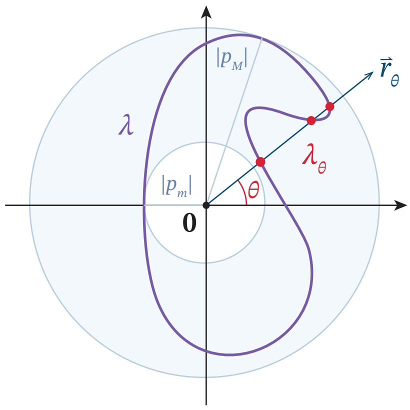
Suppose that $p_m \notin \gamma$. Then, $|p_m - p_0| < |p - p_0|$, for every $p \in \gamma$. In particular, $p_m \in \operatorname{int}\gamma$. But then, $\operatorname{int}\gamma$ would be an open set containing the boundary point $p_m \in \lambda$, yielding a forbidden intersection between $\operatorname{int}\gamma$ and $\operatorname{ext}\lambda$. The case $p_M \notin \gamma$ is handled similarly. ▨

2.4 Remark. Let λ be a planar Jordan curve which is not a circle and such that $\mathbf{0} \in \operatorname{int}\lambda$. Then, there exists a rotation R such that $R(\lambda) \cap \lambda \neq \varnothing$, but $R(\lambda) \neq \lambda$.

Proof. Let p_m and p_M in λ be such that $|p_m| = \min_{p\in\lambda}|p|$ and $|p_M| = \max_{p\in\lambda}|p|$. We notice that the hypotheses made upon λ imply $0 < |p_m| < |p_M|$. For each angle $0 \le \theta < 2\pi$, consider the *semiradius* $\vec{r}_\theta \overset{\text{def}}{=} \{te^{i\theta} : t \ge 0\}$. Then, each $\vec{r}_\theta$ intercepts λ in a compact set λ_θ, bounded away from the origin, in such a way that $\lambda = \bigsqcup\limits_{0\le\theta<2\pi} \lambda_\theta$, as conveyed in Figure 2.2.

It is claimed that, for some θ_0, it must be the case that $|p| < |p_M|$ for every $p \in \lambda_{\theta_0}$. Indeed, negating this would imply λ (properly) containing a circle of radius $|p_M|$, contradicting Remark 2.3.

Figure 2.2 – Each positive semiradius intercepts a Jordan curve leaving the origin in its interior on a compact set, which is contained within a closed annulus.



On the other hand, we may also fix θ_M such that $p_M \in \theta_M$. Let R be a planar rotation applying $\vec{r}_{\theta_M}$ onto $\vec{r}_{\theta_0}$. Then, $R(p_M) \in \vec{r}_{\theta_M}$ is a point in $R(\lambda)$ that cannot belong to λ, for that would imply a point of norm $|p_M|$ in λ_{θ_0}. Thus, $R(\lambda) \neq \lambda$. More than that, we can actually say that $R(p_M) \in R(\lambda) \cap \operatorname{ext}\lambda \neq \varnothing$. Now, since $\mathbf{0} \in \operatorname{int}\lambda$, we see that $\overline{\mathbb{D}}_{|p_m|}(\mathbf{0}) \subset \overline{\operatorname{int}\lambda}$. Equivalently,

$$\operatorname{ext}\lambda = \left(\overline{\operatorname{int}\lambda}\right)^c \subset \{z : |z| > |p_m|\}\,.$$

It thus cannot be the case that $R(\lambda)$ is fully contained within $\operatorname{ext}\lambda$, for it contains a point $R(p_m)$ of norm $|p_m|$, so $R(\lambda) \cap \overline{\operatorname{int}\lambda} \neq \varnothing$ as well. These yield an intersection $R(\lambda) \cap \lambda \neq \varnothing$. ▨

2.5 Proposition. Let $G \subset \operatorname{Homeo}(\mathbb{S}^2)$ be a sharply 3-transitive and homogeneous group. Then, G_2 permutes parallels.

Proof. Recalling that G_2 is the subgroup of G fixing $\mathbf{0}$ and ∞, $g \in G_2$ translates to an orientation preserving planar homeomorphism that fixes the origin, for which it must be proven that circles centred at the origin are mapped onto circles centred at the origin.

Let γ be such a circle and suppose, for the sake of contradiction, that the conclusion of the proposition does not hold. Then, $\lambda \overset{\text{def}}{=} g(\gamma)$ is a Jordan curve which is not a circle, having $\mathbf{0}$ in its interior. Remark 2.4 yields a rotation such that $R(\lambda) \cap \lambda \neq \varnothing$ but $R(\lambda) \neq \lambda$. Since G is homogeneous, $R \in G_2$.

The fact that $R(\lambda) \cap \lambda \neq \varnothing$ implies the existence of points p, q in γ such that $R(g(p)) = g(q)$ or, in other words, $\left(g^{-1} \circ R \circ g\right)(p) = q$. Since p,q lie at the same circle γ, 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. Since $q \notin \{\mathbf{0}, \infty\}$, this map fixes three distinct points of the sphere. By the sharp 3-transitivity of G, this amounts to $g^{-1} \circ R \circ g \circ U = \mathrm{id}$.

In particular, $\left(g^{-1} \circ R \circ g \circ U\right)(\gamma) = \gamma$. Since the planar rotation U leaves γ invariant, this implies $R\big(g(\gamma)\big) = g(\gamma)$, leading to $R(\lambda) = \lambda$, a contradiction. Therefore, the proposed statement must hold. ▨

When differentiability is added to the previous proposition, conformality at the poles follows, for the differentials there will have to preserve circles as well.

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

Proof. Given $g \in G_2$, it is known from the previous proposition that it maps parallels onto parallels. Consider first the chart $\left(\mathbb{S}^2 \setminus \{\infty\}, \Psi_N\right)$. In these coordinates, g may be thought of as a planar diffeomorphism fixing $\mathbf{0}$ and permuting circles centred at the origin. Given the linear isomorphism $A = \mathrm{D}g(\mathbf{0})$, there exist points z_m and z_M 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 maps circles into circles, we must have $|g(tz_m)| \,/\, |g(tz_M)| = 1$, for every $0 < t \leq 1$. But we may write $g(z) = Az + r(z)$, where the remainder term satisfies $r \in o(|z|)$, obtaining:

$$1 = \frac{|g(tz_m)|}{|g(tz_M)|} = \frac{|A(tz_m) + r(tz_m)|}{|A(tz_M) + r(tz_M)|} = \frac{\left|Az_m + \dfrac{r(tz_m)}{t}\right|}{\left|Az_M + \dfrac{r(tz_M)}{t}\right|}\,.$$

Taking the limit as $t \to 0^+$ in the above expression yields $|Az_m| \,/\, |Az_M| = 1$, which amounts to say that $A\left(\mathbb{S}^1\right)$ is a circle. From Lemma 2.1, conformality of A follows. If we consider in $\mathbb{S}^2 \setminus \{\mathbf{0}\}$ the chart given by stereographic projection from the South Pole, an analogous reasoning applies to $\mathrm{D}g(\infty)$, finishing the proof. ▨

To conclude, an auxiliary remark concerning 2-transitivity is needed. It might be seen, at first glance, as an underuse of the 3-transitivity hypothesis. However, Proposition 3.4 in (30) states that, if G is 2-transitive and homogeneous, then it is also 3-transitive, so we are not actually wasting any information.

2.7 Remark. Let $G \subset \mathrm{Diff}^1_+(\mathbb{S}^2)$ be a 2-transitive and homogeneous group. Then, given $0 < \delta < \pi$, there exists $h_\delta \in G$ such that

- h_δ fixes $\mathbf{0}$, but not ∞;
- h_δ and $\mathrm{D}h_\delta(\mathbf{0})$ are both δ-close to the identity.

Proof. By 2-transitivity, we may fix $h \in G$ such that $h(\mathbf{0}) = \mathbf{0}$ and $h(\infty) = \mathbf{1}$. Then, for each t, we consider $R_t \in \mathrm{Rot}_2(\mathbb{S}^2)$ and define $h_t = h^{-1} \circ R_t \circ h$. Each h_t fixes $\mathbf{0}$ but, as long as $0 < |t| < \pi$, $h_t(\infty) \neq \infty$. Indeed, if that was not the case, we would have $R_t(\mathbf{1}) = \mathbf{1}$, a contradiction, since $R_t(\mathbf{1})$ is some other point on the equator, distinct from $\mathbf{1}$.

By Lemma 2.2, the association $t \mapsto h_t$ is continuous with respect to the compact-open topology. Since $h_0 = \mathrm{id}_{\mathbb{S}^2}$, given $\delta > 0$ we may obtain $0 < \hat{t} < \delta$ sufficiently small so that

$$h_{\hat{t}} \in \mathcal{B}\big(\mathrm{id}_{\mathbb{S}^2}\,;\ (\mathbb{S}^2 \setminus \{\infty\}, \Psi_N)\,,\ \mathbb{S}^2 \setminus \mathbb{D}_\delta(\infty)\,,\ (\mathbb{S}^2 \setminus \{\infty\}, \Psi_N)\,;\ \delta\big)\,.$$

In particular, since $\mathbf{0} \notin \mathbb{D}_\delta(\infty)$, the proposed conclusions hold for $h_\delta \overset{\mathrm{def}}{=} h_{\hat{t}}$. ▨

It should be noticed that the mapping h_δ just obtained can be actually made (uniformly) C^1 δ-close to the identity with respect to any given cover of $\mathbb{S}^2$ by charts, although that would be slightly convoluted to express in terms of basic neighbourhoods.

The proof of Theorem C

Let G homogeneous and sharply 3-transitive be given. If G is not contained in $\text{Möb}(\mathbb{S}^2)$, there exist at least one $\tilde{g} \in G$ and a point $\tilde{z} \in \mathbb{S}^2$ such that $\mathrm{D}\tilde{g}(\tilde{z})$ is nonconformal. Let $U, R \in \mathrm{Rot}(\mathbb{S}^2)$ be such that $U(\mathbf{0}) = \tilde{z}$ and $R(\tilde{g}(\tilde{z})) = \mathbf{0}$. Then, $g \overset{\text{def}}{=} R \circ \tilde{g} \circ U \in G$ is such that g fixes $\mathbf{0}$, but $\mathrm{D}g(\mathbf{0})$ is nonconformal.

Let $A \overset{\text{def}}{=} \mathrm{D}g(\mathbf{0})$. Nonconformality implies the existence of unit vectors u_0, v_0 such that,

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

If we now consider the uniformly continuous function $(u, v) \in \mathbb{S}^1 \times \mathbb{S}^1 \mapsto \langle u, v \rangle$, given $\varepsilon = |\beta - \alpha|/2 > 0$ we may obtain $\delta > 0$ such that

[2.3] $$|u - w| \leq \delta \text{ and } |v - z| \leq \delta \text{ imply } |\langle u, v \rangle - \langle w, z \rangle| < \varepsilon\,.$$

With respect to δ as above, Remark 2.7 yields $h_\delta \in G$ such that $h_\delta(\mathbf{0}) = \mathbf{0}$, $h_\delta(\infty) \overset{\text{def}}{=} p_\delta \neq \infty$ and both h_δ and its differential $C_\delta \overset{\text{def}}{=} \mathrm{D}h_\delta(\mathbf{0})$ are $\delta/2$-close to the identity. Since G is 3-transitive, we may fix $f_\delta \in G_2$ such that $f_\delta(g(\infty)) = p_\delta$. By Corollary 2.6, $B \overset{\text{def}}{=} \mathrm{D}f_\delta(\mathbf{0})$ must be conformal. If we lastly define $\hat{g} \overset{\text{def}}{=} h_\delta \circ f_\delta \circ g \in G$, then $\hat{g}$ also fixes the poles and thus $D \overset{\text{def}}{=} \mathrm{D}\hat{g}(\mathbf{0})$ must be conformal. By the Chain Rule,

$$\mathrm{D}\hat{g}(\mathbf{0}) = \mathrm{D}h_\delta(\, f_\delta(g(\mathbf{0}))\,) \circ \mathrm{D}f_\delta(g(\mathbf{0})) \circ \mathrm{D}g(\mathbf{0}) = \mathrm{D}h_\delta(\mathbf{0}) \circ \mathrm{D}f_\delta(\mathbf{0}) \circ \mathrm{D}g(\mathbf{0})$$

or, in other words, $D = C_\delta BA$. Since C_δ is $\delta/2$-close to the identity, for every $w_0 \in \mathbb{S}^2$:

$$\begin{aligned}
\left| \frac{C_\delta w_0}{|C_\delta w_0|} - w_0 \right| &= \left| C_\delta w_0 - w_0 + \frac{1 - |C_\delta w_0|}{|C_\delta w_0|}\, C_\delta w_0 \right| \\
&\leq |C_\delta w_0 - w_0| + |1 - |C_\delta w_0|| \\
&= |C_\delta w_0 - w_0| + ||w_0| - |C_\delta w_0|| \\
&\leq 2|C_\delta w_0 - w_0| \\
&\leq 2\|C_\delta - \mathrm{id}\| \leq \delta\,.
\end{aligned}$$

In particular, for $w_0 = BAu_0/|BAu_0|$:

$$\left| \frac{C_\delta(BAu_0)}{|C_\delta(BAu_0)|} - \frac{BAu_0}{|BAu_0|} \right| = \left| \frac{C_\delta\big(BAu_0 \,/\, |BAu_0|\big)}{\big|C_\delta\big(BAu_0 \,/\, |BAu_0|\big)\big|} - \frac{BAu_0}{|BAu_0|} \right| \leq \delta\,.$$

Analogously, we obtain

$$\left| \frac{C_\delta(BAv_0)}{|C_\delta(BAv_0)|} - \frac{BAv_0}{|BAv_0|} \right| \leq \delta\,.$$

From [2.3], the pair of inequalities above imply

[2.4] $$\left| \left\langle \frac{C_\delta(BAu_0)}{|C_\delta(BAu_0)|}, \frac{C_\delta(BAv_0)}{|C_\delta(BAv_0)|} \right\rangle - \left\langle \frac{BAu_0}{|BAu_0|}, \frac{BAv_0}{|BAv_0|} \right\rangle \right| < \varepsilon\,.$$

But now, since B is conformal,

$$\left\langle \frac{BAu_0}{|BAu_0|}, \frac{BAv_0}{|BAv_0|} \right\rangle = \left\langle \frac{Au_0}{|Au_0|}, \frac{Av_0}{|Av_0|} \right\rangle = \beta \, .$$

Thus, recognising $D = C_\delta AB$ in [2.4], it yields

$$\left| \left\langle \frac{Du_0}{|Du_0|}, \frac{Dv_0}{|Dv_0|} \right\rangle - \beta \right| < \varepsilon \, .$$

From the choice of ε, $\langle Du_0 \,/\, |Du_0|, Dv_0 \,/\, |Dv_0| \rangle \neq \alpha$ follows. But this, in turn, implies D nonconformal, a contradiction.

We conclude that a nonconformal $\tilde{g} \in G$ could not exist in the first place, and therefore $G \subset \text{Möb}(\mathbb{S}^2)$. But now, given $M \in \text{Möb}(\mathbb{S}^2)$, by 3-transitivity there exists $h \in G$ such that $h(\mathbf{0}) = M(\mathbf{0})$, $h(\mathbf{1}) = M(\mathbf{1})$ and $h(\infty) = M(\infty)$. Since h is *a* Möbius transformation agreeing with M in three points, it must be *the* transformation M. Since M was arbitrary, $G = \text{Möb}(\mathbb{S}^2)$. This finishes the proof of Theorem C. ▨

2.2 An extension lemma

We shall now address the question of what kinds of transformations are expected to be found in proper extensions of the Möbius group. More precisely, we shall prove that any such extension must contain a transformation having an hyperbolic fixed point. For completeness, we include the relevant definition.

2.8 Definition. Let $\mathcal{M}$ be a Riemannian manifold and consider $f \in \text{Diff}(\mathcal{M})$. A fixed point x_0 of f is said to be *hyperbolic* if the tangent space $T_{x_0}\mathcal{M}$ admits a splitting into two subspaces, both of which are invariant under the self-map $\mathrm{D}f(x_0)$, say $T_{x_0}\mathcal{M} = E^{\mathrm{s}}_{x_0} \oplus E^{\mathrm{u}}_{x_0}$, and such that $\mathrm{D}f(x_0)\restriction_{E^{\mathrm{s}}_{x_0}}$ and $\mathrm{D}\big(f^{-1}\big)(x_0)\restriction_{E^{\mathrm{u}}_{x_0}}$ are contraction maps.

Above, we did not exclude the possibility that one of the subspaces is trivial. In our simple two-dimensional setting, we shall append the adjective *saddle* to the prototypical case in which the differential has one contracting eigendirection and one expanding eigendirection.

As in Section 2.1, we start with some elementary linear algebra preliminaries, followed by the description of yet another convenient identification between spherical and planar mappings. Then, we proceed to the actual development of the Lemma.

We start by considering a real 2×2 *defective* matrix A, meaning it has one single eigenvalue $\lambda \neq 0$ of geometrical multiplicity one and geometrical multiplicity two. It is widely known that it admits a chain $\{u, w\}$ of generalised eigenvectors, satisfying $Au = \lambda u$ and $Aw = u + \lambda w$. It may be assumed that $|u| = 1$. Let $w^\perp = w - \langle u, w \rangle\, u$. Then, $\langle w^\perp, u \rangle = 0$ and:

$$\begin{aligned} Aw^\perp = A(w - \langle u, w \rangle\, u) &= (u + \lambda w) - \langle u, w \rangle (\lambda u) \\ &= u + \lambda(w - \langle u, w \rangle\, u) = u + \lambda w^\perp \, . \end{aligned}$$

This means that, upon swapping w by $w^\perp$, a chain of generalised eigenvectors may always be supposed orthogonal (although not orthonormal).

Next, for a given $\rho > 0$, let H_ρ denote the planar homothety $H_\rho(z) = \rho z$ and, also, the hyperbolic mapping in $\text{Möb}_2(\mathbb{S}^2)$ induced accordingly. Upon considering [2.2], we write $\mathrm{D}H_\rho(\mathbf{0}) = \rho\,\mathrm{id}$, whether in $\mathbb{R}^2$ or in $T_{\mathbf{0}}\mathbb{S}^2$.

Lastly, let $G \subset \mathrm{Diff}^1_+(\mathbb{S}^2)$ be a group properly extending $\text{Möb}(\mathbb{S}^2)$. This means that there exist at least one transformation $\tilde{h} \in G$ and a point $\tilde{z} \in \mathbb{S}^2$ for which $\mathrm{D}\tilde{h}(\tilde{z}) : T_{\tilde{z}}\mathbb{S}^2 \to T_{\tilde{h}(\tilde{z})}\mathbb{S}^2$ is nonconformal.

Let $M_1 \in \text{Möb}(\mathbb{S}^2)$ and $M_2 \in \text{Möb}(\mathbb{S}^2)$ be such that $M_1(\mathbf{0}) = \tilde{z}$, $M_1(\infty) = \tilde{h}^{-1}(\infty)$, $M_2\big(\tilde{h}(\tilde{z})\big) = \mathbf{0}$ and $M_2(\infty) = \infty$. Then, $\hat{h} \stackrel{\text{def}}{=} M_2 \circ \tilde{h} \circ M_1$ is a mapping in G_2 such that $\mathrm{D}\hat{h}(\mathbf{0})$ is nonconformal. This will be the starting point of the constructions to follow.

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

Proof. Let $\hat{h} \in G_2$ be such that $A \stackrel{\text{def}}{=} \mathrm{D}\hat{h}(\mathbf{0})$ is nonconformal. Then, there exist a pair of unit vectors u and v such that $\alpha \stackrel{\text{def}}{=} \mathrm{ang}(u, v) \neq \mathrm{ang}(Au, Av) \stackrel{\text{def}}{=} \beta$. Now, if $\mathrm{ang}(u, Au) = \theta \in [0, \pi]$, it is uniquely determined by the relation $\cos\theta = \langle u, Au\rangle/|Au|$. Lagrange's identity implies, in turn, the expression $\sin\theta = |\det[u|Au]|/|Au|$ for its sine. But writing $u = \left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$ and $Au = \left[\begin{smallmatrix} a' \\ b' \end{smallmatrix}\right]$ in coordinates, one has, for a given angle α:

$$\begin{aligned} \frac{\langle R_\alpha(Au), u\rangle}{|R_\alpha(Au)|} &= \frac{(a'\cos\alpha - b'\sin\alpha)\,a + (a'\sin\alpha + b'\cos\alpha)\,b}{|Au|} \\ &= \frac{aa' + bb'}{|Au|}\cos\alpha - \frac{ab' - ba'}{|Au|}\sin\alpha \\ &= \cos\theta\cos\alpha - \frac{\det[u|Au]}{|Au|}\sin\alpha\,. \end{aligned} \tag{2.5}$$

Thus, upon defining

$$R \stackrel{\text{def}}{=} \begin{cases} R_{-\theta} & \text{if } \det[u|Au] > 0\,, \\ R_\theta & \text{if } \det[u|Au] \leq 0\,, \end{cases}$$

and $B \stackrel{\text{def}}{=} RA$, the parity of the trigonometric functions along with [2.5] imply

$$\frac{\langle Bu, u\rangle}{|Bu|} = \frac{\langle R(Au), u\rangle}{|R(Au)|} = 1\,.$$

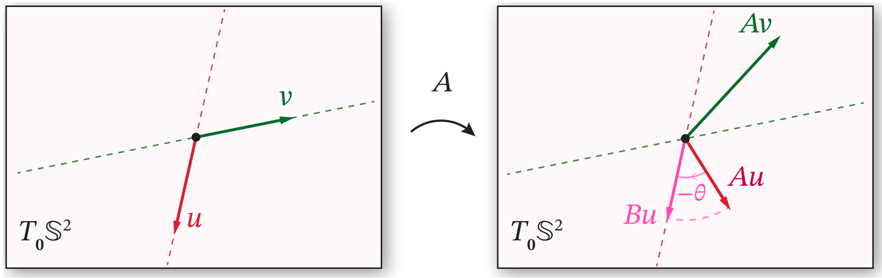
In other words, $\mathrm{ang}(Bu, u) = 0$, meaning that $Bu = \mu u$, where $\mu > 0$. Thus, B is a linear map having at least one positive eigenvalue, associated with the unit eigenvector u, as pictured in Figure 2.3. This leaves two possibilities for B.

i) B has a defective matrix.

As discussed prior to the Lemma's statement, we may fix an orthogonal chain of generalised eigenvectors $\{u, w\}$ satisfying $Bu = \mu u$, $Bw = u + \mu w$ and $\langle u, w\rangle = 0$. For each angle $\phi \in [0, \pi/2]$, we define

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

Figure 2.3 – Post-composing A with a convenient rotation yields a positive eigendirection.



This yields an orthonormal frame $\{x_\phi, y_\phi\}$ of $T_{\mathbf{0}}\mathbb{S}^2$, for which we consider the continuous function $\xi(\phi) = \langle Bx_\phi, B\,y_\phi\rangle \,/\, |Bx_\phi|\,|By_\phi|$. Then, on the one hand

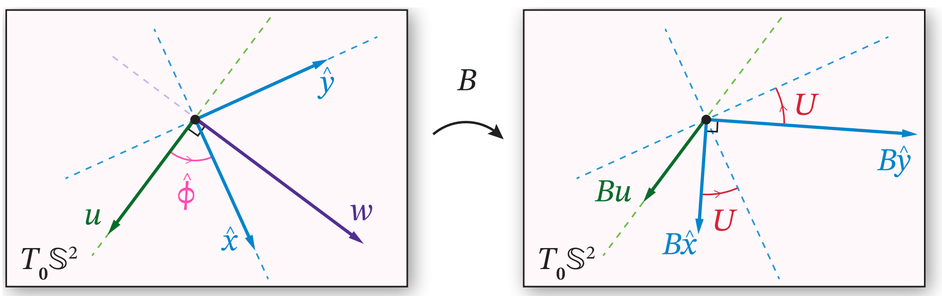
$$\xi(0) = \frac{\langle Bx_0, By_0\rangle}{|Bx_0|\,|By_0|} = \frac{\big\langle Bu, B\,(w/|w|)\big\rangle}{\big|Bu\big|\,\big|B\,(w/|w|)\big|} = \frac{\langle Bu, Bw\rangle}{|Bu|\,|Bw|} = \frac{\langle \mu u, u+\mu w\rangle}{\mu\,|u+\mu w|} = \frac{1}{|u+\mu w|}$$

and, on the other hand,

$$\xi\Big(\frac{\pi}{2}\Big) = \frac{\langle Bx_{\pi/2}, By_{\pi/2}\rangle}{|Bx_{\pi/2}|\,|By_{\pi/2}|} = \frac{\big\langle B\,(w/|w|), B\,(-u)\big\rangle}{\big|B\,(w/|w|)\big|\,\big|B\,(-u)\big|} = -\frac{\langle Bw, Bu\rangle}{|Bw|\,|Bu|} = -\frac{1}{|u+\mu w|} = -\xi(0)\,.$$

Therefore, it must be the case that $\xi(\hat\phi) = 0$ for some $0 < \hat\phi < \pi/2$. This means that, if $\hat x = x_{\hat\phi}$ and $\hat y = x_{\hat\phi}$, then $B\hat x$ and $B\hat y$ are orthogonal, as pictured in Figure 2.4. In particular, the orthonormal frame $\{B\hat x/|B\hat x|, B\hat y/|B\hat y|\}$ can be applied onto $\{\hat x, \hat y\}$ by a certain planar rotation U, since B preserves orientation. Thus, $U(B\hat x) = \nu_1 \hat x$ and $U(B\hat y) = \nu_2 \hat y$, where $\nu_1, \nu_2 > 0$ (explicitely, $\nu_1 = |B\hat x|$ and $\nu_2 = |B\hat y|$).

Figure 2.4 – The orthonormal frame $\{\hat x, \hat y\}$ — obtained from $\{u, w\}$ by a rotation and normalization — is applied by B onto a pair of orthogonal vectors.



It is claimed that $\nu_1 \neq \nu_2$. Indeed, let us assume for the sake of contradiction that $\nu_1 = \nu_2 = \nu$. Then, since $\{\hat x, \hat y\}$ is a basis, this would imply $UB = \nu\,\mathrm{id}$. Recalling that $B = RA$, $A = \nu R^{-1} U^{-1}$ would be conformal, contradicting the choice of $\hat h$.

Let $C \overset{\text{def}}{=} UB$ and suppose, without loss of generality, that $0 < \nu_1 < \nu_2$. If $\rho \overset{\text{def}}{=} 1 \,/\, \sqrt{\nu_1\nu_2}$ and $\lambda \overset{\text{def}}{=} \sqrt{\nu_1 \,/\, \nu_2}$, then $0 < \lambda < 1$,

$$(\rho C)\hat x = \rho\,(\nu_1 \hat x) = \frac{\nu_1}{\sqrt{\nu_1\nu_2}}\,\hat x = \sqrt{\frac{\nu_1}{\nu_2}}\,\hat x = \lambda \hat x \quad \text{and} \quad (\rho C)\hat y = \frac{\nu_2}{\sqrt{\nu_1\nu_2}}\,\hat y = \sqrt{\frac{\nu_2}{\nu_1}}\,\hat y = \lambda^{-1}\hat y\,.$$

Now, let P be a rotation such that $P^{-1}\hat{x} = {}^{\partial}\!/_{\partial x}$. Since $\{\hat{x}, \hat{y}\}$ and $\{{}^{\partial}\!/_{\partial x}, {}^{\partial}\!/_{\partial y}\}$ are both orthonormal frames, $P^{-1}\hat{y} = \varepsilon\, {}^{\partial}\!/_{\partial y}$, where $\varepsilon = \pm 1$, depending on the relative orientation of both bases.

Lastly, consider $\hat{A} \stackrel{\text{def}}{=} \rho\, P^{-1}CP$. We have

$$\hat{A}({}^{\partial}\!/_{\partial x}) = \rho\, P^{-1}CP\,({}^{\partial}\!/_{\partial x}) = P^{-1}\big((\rho\, C)\hat{x}\big) = P^{-1}(\lambda\hat{x}) = \lambda\, {}^{\partial}\!/_{\partial x} \quad \text{and}$$
$$\hat{A}({}^{\partial}\!/_{\partial y}) = \rho\, P^{-1}CP\,({}^{\partial}\!/_{\partial y}) = P^{-1}\big((\rho\, C)(\varepsilon\hat{y})\big) = \varepsilon\, P^{-1}\big(\lambda^{-1}\hat{y}\big) = \varepsilon^2\lambda^{-1}\, {}^{\partial}\!/_{\partial y}\,.$$

In short, with respect to the canonical basis, $\hat{A} = \operatorname{diag}[\lambda, \lambda^{-1}]$. But if

$$\hat{g} \stackrel{\text{def}}{=} H_\rho \circ P^{-1} \circ U \circ R \circ \hat{h} \circ P \in G_2\,,$$

the Chain Rule — along with the identifications made between rotations and homothethies with their differentials at the poles — yields

$$\mathrm{D}\hat{g}(\mathbf{0}) = (\rho\ \mathrm{id})\, P^{-1}\, U\, R\, \mathrm{D}\hat{h}(\mathbf{0})\, P = \rho\big(P^{-1}\, U\,(RA)\, P\big) = \rho\big(P^{-1}\,(U\, B)\, P\big) = \rho\big(P^{-1}\, C\, P\big) = \hat{A}\,.$$

This establishes the Lemma in this case.

ii) B has a second eigenvalue $\nu \in \mathbb{R}$.

We first notice that $\nu \neq \mu$. Indeed, if that was not the case, as previously argued $B = \mu\,\mathrm{id}$ would hold, what is incompatible with A nonconformal. Thus, we may fix a second unit eigenvector w associated with ν. In particular, $\{v, w\}$ is a linearly independent set, and the orthogonal complement $w^\perp = w - \langle w, u\rangle\, u$ is nonzero. Therefore,

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

define, for each $\phi \in [0, \pi/2]$, an orthonormal frame $\{x_\phi, y_\phi\}$ of $T_{\mathbf{0}}\mathbb{S}^2$. As made in case i), we consider the continuous function $\xi(\phi) = \langle Bx_\phi, B\,y_\phi\rangle \,/\, |Bx_\phi|\,|By_\phi|$.

If we let $c = \langle w, u\rangle$ to ease notation, then $Bx_0 = Bu = \mu u$,

$$\begin{aligned} B(y_0) = B\bigg(\frac{w^\perp}{|w^\perp|}\bigg) &= \frac{1}{|w^\perp|}B(w - cu) = \frac{1}{|w^\perp|}(\nu w - c\mu u) \\ &= \frac{1}{|w^\perp|}\big(\nu\big(w^\perp + cu\big) - c\mu u\big) = \frac{c(\nu-\mu)}{|w^\perp|}u + \nu\frac{w^\perp}{|w^\perp|} \end{aligned}$$

and, consequently,

$$\xi(0) = \frac{\langle Bx_\phi, B\,y_\phi\rangle}{|Bx_0|\,|By_0|} = \frac{\mu\,\dfrac{c\,(\nu-\mu)}{|w^\perp|}}{\mu\,\dfrac{\sqrt{c^2(\nu-\mu)^2 + \nu^2|w^\perp|^2}}{|w^\perp|}} = \frac{c\,(\nu-\mu)}{\sqrt{c^2(\nu-\mu)^2 + \nu^2|w^\perp|^2}}\,.$$

Analogously, $x_{\pi/2} = w^\perp/|w^\perp|$, $y_{\pi/2} = -u$. Thus, the calculations above yield

[2.6] $$\xi\Big(\frac{\pi}{2}\Big) = -\frac{c\,(\nu-\mu)}{\sqrt{c^2(\nu-\mu)^2 + \nu^2|w^\perp|^2}} = -\xi(0)\,.$$

We notice that, if $c = 0$, $\{u, w\}$ was already an orthonormal frame from the start. If, on the other hand, $c \neq 0$, [2.6] above implies the existence of some $0 < \hat{\phi} < \pi/2$ such that $\xi(\hat{\phi}) = 0$.

Either way, we may fix $\hat{\phi} \in [0, \pi/2)$ such that if $\hat{x} = x_{\hat{\phi}}$ and $\hat{y} = y_{\hat{\phi}}$, then $\operatorname{ang}(B\hat{x}, B\hat{y}) = \operatorname{ang}(\hat{x}, \hat{y}) = \pi / 2$. In particular, since B preserves orientation, the orthonormal frame $\{B\hat{x}/|B\hat{x}|, B\hat{y}/|B\hat{y}|\}$ can be applied onto $\{\hat{x}, \hat{y}\}$ by a certain planar rotation U — possibly the identity. From here, the proof follows exactly as in case i), by defining the ν_i, ρ, λ and P exactly as before. ▨

The basic idea of the above proof — post-composing a given nonconformal map with rotations to generate eigendirections and with homotheties to modulate the corresponding eigenvalues — can actually be implemented in a C^1 continuous fashion, yielding a diffeotopy connecting the identity to a transformation for which the South Pole is a hyperbolic fixed point, whilst keeping the poles fixed throughout the process. This is the content of our Extension Lemma to follow.

2.10 The Extension Lemma. Let $G \subset \mathrm{Diff}^1_+(\mathbb{S}^2)$ be a proper extension of $\text{Möb}(\mathbb{S}^2)$. Then, there exists an $\mathcal{J}G_2$ 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) $\mathrm{D}g_1(\mathbf{0}) = \operatorname{diag}[\lambda, \lambda^{-1}]$ with respect to the canonical basis of $T_{\mathbf{0}}\mathbb{S}^2$.

Proof. By Lemma 2.9 above, we may fix $\hat{g} \in G_2$ such that $\hat{A} \overset{\text{def}}{=} \mathrm{D}\hat{g}(\mathbf{0}) = \operatorname{diag}[\mu, \mu^{-1}]$ with respect to the canonical basis $\{\partial/\partial x, \partial/\partial y\}$ of $T_0\mathbb{S}^2$, where $0 < \mu < 1$. For $s \in [0, \pi/2]$, let $B_s = \hat{A}^{-1} R_s \hat{A}$ and $v_s \overset{\text{def}}{=} B_s(\partial/\partial x) = \left[\begin{smallmatrix}\cos s\\ \mu^2 \sin s\end{smallmatrix}\right]$. Then, $|v_s| < 1$ for $s > 0$.

We now consider the continuous function $\theta : [0, \pi/2] \to [0, \pi]$ given as the angle between v_s and $\partial/\partial x$:

$$\theta(s) = \operatorname{ang}(v_s, \partial/\partial x) = \arccos\left[\frac{\cos s}{|v_s|}\right].$$

If we then let $C_s \overset{\text{def}}{=} R_{-\theta(s)} B_s$, the canonical vector $\partial/\partial x$ defines a contracting direction for C_s, of rate $|v_s| < 1$. Indeed, in coordinates:

$$[2.7] \qquad C_s(\partial/\partial x) = R_{-\theta(s)}(B_s(\partial/\partial x)) = R_{-\theta(s)} v_s = \begin{bmatrix} \mu^2 \sin s\, \sin\theta(s) + \cos s \cos\theta(s) \\ \mu^2 \sin s\, \cos\theta(s) - \cos s \sin\theta(s) \end{bmatrix}.$$

But the identity $\sin^2\theta(s) = 1 - \cos^2\theta(s)$ yields

$$\sin^2\theta(s) = 1 - \frac{\cos^2 s}{|v_s|^2} = \frac{|v_s|^2 - \cos^2 s}{|v_s|^2} = \frac{\left(\cos^2 s + \mu^4 \sin^2 s\right) - \cos^2 s}{|v_s|^2} = \frac{\mu^4 \sin^2 s}{|v_s|^2}.$$

Since the sine function is positive over the considered range, $\sin\theta(s) = \mu^2 \sin s \,/\, |v_s|$ follows.
This implies

$$\mu^2 \sin s\, \sin\theta(s) + \cos s \cos\theta(s) = |v_s| \sin^2\theta(s) + |v_s| \cos^2\theta(s) \quad \text{and}$$
$$\mu^2 \sin s\, \cos\theta(s) - \cos s \sin\theta(s) = |v_s| \sin\theta(s)\, \cos\theta(s) - |v_s| \cos\theta(s)\, \sin\theta(s) = 0\,.$$

Comparison with [2.7] yields $C_s(\partial/\partial x) = \left[\begin{smallmatrix}|v_s|\\ 0\end{smallmatrix}\right] = |v_s|\, \partial/\partial x$, as claimed. In particular, when $s = \pi/2$, $\theta(\pi/2) = \pi/2$, and thus $|v_s| = \mu^2$. Or, in other words, $C_{\pi/2}(\partial/\partial x) = \mu^2\, \partial/\partial x$. Furthermore, for every $s > 0$, C_s is a saddle matrix, in the following sense:

$$\det C_s = \det\left[R_{-\theta(s)} \hat{A}^{-1} R_s \hat{A}\right] = \det\left[R_{-\theta(s)}\right] \det\left[\hat{A}^{-1}\right] \det R_s \det\hat{A} = \left(\det\hat{A}\right)^{-1} \det\hat{A} = 1\,,$$

and therefore C_s has $|\upsilon_s|^{-1} > 1$ as its other (real) eigenvalue.

In general, we cannot ensure that C_s is diagonal (with respect to the canonical basis). But, when $s = \pi/2$, we may explicitly compute $C_{\pi/2}(\partial/\partial y) = \left[\begin{smallmatrix} 0 \\ \mu^{-2} \end{smallmatrix}\right] = \mu^{-2}\,\partial/\partial y$. In other words, if we let $\lambda \stackrel{\text{def}}{=} \mu^2 < 1$, then $C_{\pi/2} = \operatorname{diag}[\lambda, \lambda^{-1}]$.

Lastly, since $\theta(\cdot)$ is a continuous function, the expression

$$g_t = R_{-\theta\left(\frac{\pi t}{2}\right)} \circ \hat{g}^{-1} \circ R_{\frac{\pi t}{2}} \circ \hat{g} \in G_2 \ , \ 0 \leq t \leq 1 \ ,$$

defines a diffeotopy $(g_t)_{t\in\mathbb{I}}$, by Lemma 2.2. Furthermore, the Chain Rule yields $\mathrm{D}g_t(\mathbf{0}) = C_{\pi t/2}$. Recalling that the x-axis corresponds to Γ on the sphere — the direction generated by $\partial/\partial x$ being identified with $T_{\mathbf{0}}\Gamma$ — the considerations previously made translate into the statements of the Lemma, completing the proof. ▨

Chapter 3

A fundamental lemma

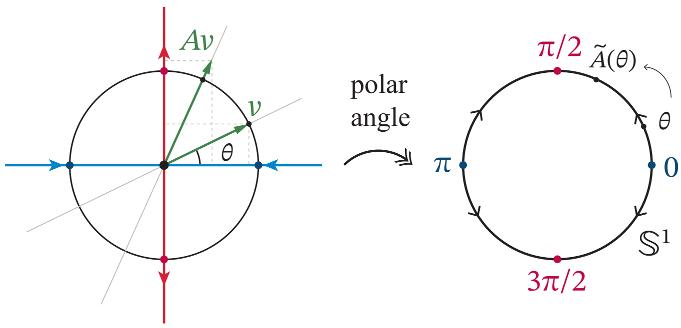
If we let $G \subset \mathrm{Diff}^1_+(\mathbb{S}^2)$ be a group properly extending $\mathrm{Möb}(\mathbb{S}^2)$, this section is devoted to characterise special kinds of isotopies in G, which form the basis of all our subsequent arguments. For this reason, the corresponding result is called – in the context of this work – the Fundamental Lemma. In the spirit of hyperbolic dynamics, we start by focusing on cones and how isotopies such as the ones defined in the Extension Lemma 2.10 act on them.

Given an angle $\alpha \in (0, \pi/2)$, by the *α-cone* we shall refer to the subset of the plane described in polar coordinates as follows:

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

We remark that, for $0 < \lambda < 1$, any cone is "broadened" under the action of the hyperbolic matrix $A = \mathrm{diag}[\lambda, \lambda^{-1}]$ or, yet, every vector "moves away" from the x-axis. More precisely, consider $\mathbb{S}^1$ parameterized by the counter-clockwise polar angle θ that each direction of space makes with the x-axis. Then, the action of A induces a monotone circle dynamics $\tilde{A}$ – by radial projection – such that $\tilde{A}$ has repelling fixed points at 0 and π and attracting fixed points at $\pi/2$ and $3\pi/2$, as pictured in Figure 3.1.

Figure 3.1 – The projective action of a saddle matrix on the lines through the origin.



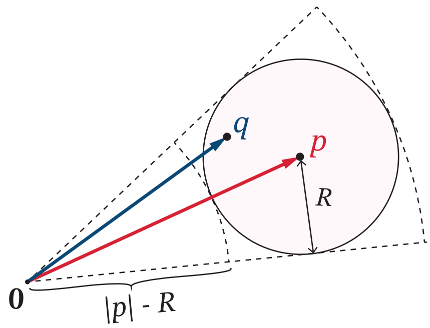
This behaviour has implications on how a diffeomorphism g having A as differential at a hyperbolic fixed point relates (locally) to compositions with rotations. To see how, we first record two elementary yet instrumental remarks steaming from plane geometry.

3.1 Remark. Let p, q be two points on the plane such that $q \in \overline{\mathbb{D}}_R(p)$ for some $R > 0$. If $|p| > R$ or, in other words, if the origin is external to the disk, then $\mathrm{ang}(p, q) < \pi/2$.

Proof. This geometrically intuitive fact, suggested by figure 3.2, is a straightforward consequence of the following polarisation formula:

$$\langle p, q\rangle = \frac{1}{2}\big(\underbrace{|q|^2}_{\geq 0} + \underbrace{|p|^2 - |p-q|^2}_{\text{strictly positive}}\big) > 0\,.$$ ▨

Figure 3.2 – Two points lying in a closed disk to which the origin is external make an angle $< \pi/2$.



3.2 Remark. Let p, q be two nonzero distinct points on the plane. Then, the following estimate holds:

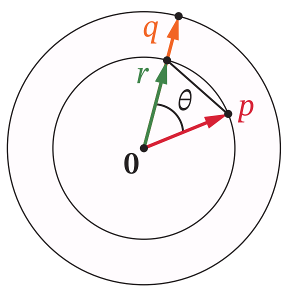
[3.1] $$|p-q| \geq \min\{|p|, |q|\}\,\sin\left[\frac{\operatorname{ang}(p,q)}{2}\right].$$

Proof. Assume without loss of generality that $|p| \leq |q|$, and let $r = (|p|/|q|)\,q$. The triangle inequality yields

$$|p-r| \leq |p-q| + |q-r| = |p-q| + (|q|-|p|) \leq 2|p-q|\,.$$

But $|p-r|$ is the length of the chord subtending $\theta = \operatorname{ang}(p,q)$ on the circle of radius $|p|$ centred at the origin, as suggested by figure 3.3. Therefore, it is known from elementary plane geometry that $|p-r| = 2\,|p|\,\sin(\theta\,/\,2)$. Comparison of the two formulae gives the desired result when the points are not collinear. In the latter case the angle between them – and thus its sine – is zero, and [3.1] holds trivially. ▨

Figure 3.3 – Relation between points on the plane and the chord subtending their angle.



These allow us to conclude that, for diffeomorphisms, the differential provides not only a numerical approximation to the action of g, but also angular, as stated below. This might not be true for maps that are only differentiable, as the differential may squish a cone onto a segment.

3.3 Lemma. Let g be a planar diffeomorphism fixing the origin. Then, given $0 < \varepsilon < \pi$ there exists $\delta > 0$ such that $\operatorname{ang}(g(z), \mathrm{D}g(\mathbf{0})z) \leq \varepsilon$ whenever $0 < |z| < \delta$.

Proof. Let $A \overset{\text{def}}{=} Dg(\mathbf{0})$. Given $\varepsilon > 0$, by the differentiability of g at the origin we obtain $\delta > 0$ satisfying:

$$0 < |z| < \delta \ \text{implies}\ |g(z) - Az| \leq cr\sin\left(\frac{\varepsilon}{2}\right)|z|\,,\ \text{where}\ r = \frac{1}{1+\sin(\varepsilon/2)}$$

and $c > 0$ has the property that $|Av| \geq c|v|$ for every v. It should be noticed that $1/2 < r < 1$, because $0 < \varepsilon/2 < \pi/2$, and that c indeed exists, because A is a linear isomorphism.

Now, given any such z, the triangle inequality along with the definition of c yield:

$$\begin{aligned} |g(z)| &\geq |Az| - |g(z) - Az| \\ &\geq |Az| - cr\sin\left(\frac{\varepsilon}{2}\right)|z| \\ &\geq c|z| - cr\sin\left(\frac{\varepsilon}{2}\right)|z| \\ &= c\left[1 - r\sin\left(\frac{\varepsilon}{2}\right)\right]|z|\,. \end{aligned}$$

Keeping in mind the bounds on r, the rightmost side of this inequality is seen to be strictly positive. This allows us to reach the following conclusions:

- By the aforementioned differentiability, $g(z)$ lies in a closed disk centred at Az and of radius $cr\sin(\varepsilon/2)\,|z|$.
- The origin is an external point to this disk, due to the second line's positiveness.
- In particular, by Remark 3.1, $\text{ang}(g(z), Az) < \pi/2$.
- Also, the following holds:

  $$\min\{|g(z)|, |Az|\} \geq c\left[1 - r\sin\left(\frac{\varepsilon}{2}\right)\right]|z| > 0\,,$$

  both because of the previous estimate and because subtracting a positive quantity from $c|z|$ only makes it smaller.

Considering these along with Remark 3.2 imply:

$$\sin\left[\frac{\text{ang}(g(z), Az,)}{2}\right] \leq \frac{|g(z) - Az|}{\min\{|g(z)|, |Az|\}} \leq \frac{cr\sin(\varepsilon/2)\,|z|}{c[1 - r\sin(\varepsilon/2)]\,|z|} = \frac{r\sin(\varepsilon/2)}{1 - r\sin(\varepsilon/2)} = \sin\left(\frac{\varepsilon}{2}\right).$$

Since none of the angles above exceed $\pi/2$, these calculations allow us to conclude that, as long as $0 < |z| < \delta$, $\text{ang}(g(z), Az) \leq \varepsilon$. ▨

Using the approximation described above, the cone-broadening property of a hyperbolic saddle can be seen to propagate locally to the diffeomorphism g itself, in a precise sense described by the following result.

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

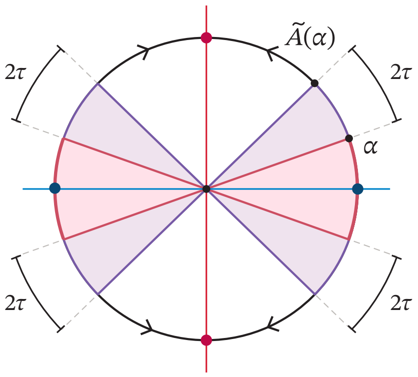
$$0 < |z| < \delta \ \text{and}\ z \notin \mathcal{C}_\alpha \ \text{imply}\ R_\omega(g(z)) \notin \mathcal{C}_\alpha \ \text{whenever}\ |\omega| < \tau\,.$$

Proof. First, notice that the set $\mathcal{C}_\alpha$ intercepts $\mathbb{S}^1$ at a pair of disjoint arcs, one with endpoints $2\pi - \alpha$ and α, the other with endpoints $\pi - \alpha$ and $\pi + \alpha$. Thus,if we let $A \overset{\text{def}}{=} Dg(\mathbf{0})$ and consider the circle dynamics $\tilde{A}$ induced accordingly, $\tau \stackrel{\text{def}}{=} \frac{1}{2}\left[\tilde{A}(\alpha) - \alpha\right]$ is such that $0 < 2\tau < \pi/2$. Furthermore, by the symmetry suggested in Figure 3.4,

$$2\tau = (\pi - \alpha) - \tilde{A}(\pi - \alpha) = \tilde{A}(\pi + \alpha) - (\pi + \alpha) = (2\pi - \alpha) - \tilde{A}(2\pi - \alpha)\,,$$

meaning that the action of $\tilde{A}$ broadens the arcs determined by $\mathcal{C}_\alpha$ by $\pm 2\tau$ at each endpoint.

Figure 3.4 – The action of the cyrcle dynamics $\tilde{A}$ induced by the saddle matrix $A = \operatorname{diag}[\lambda, \lambda^{-1}]$, broadening any cone to which the unstable direction is external.



Let $\theta \in \mathbb{S}^1 \setminus \mathcal{C}_\alpha$. Without loss of generality, we address the case $\alpha < \theta < \pi - \alpha$. Due to the induced dynamics monotonicity, $\tilde{A}(\alpha) < \tilde{A}(\theta) < \tilde{A}(\pi - \alpha)$, and thus:

$$\tilde{A}(\theta) + 2\tau < \tilde{A}(\pi - \alpha) + 2\tau = \tilde{A}(\pi - \alpha) + ((\pi - \alpha) - \tilde{A}(\pi - \alpha)) = \pi - \alpha\,,$$
$$\tilde{A}(\theta) - 2\tau > \tilde{A}(\alpha) - 2\tau = \tilde{A}(\alpha) - (\tilde{A}(\alpha) - \alpha) = \alpha\,.$$

Putting together the extremes of the above inequalities it is possible to conclude that

$$\alpha < \tilde{A}(\theta) - 2\tau < \tilde{A}(\theta) + 2\tau < \pi - \alpha\,.$$

In other words, if v is any nonzero vector of such a polar angle θ, then $R_{\pm 2\tau}(Av) \notin \mathcal{C}_\alpha$, due to the fact that the diagram below commutes:

$$\begin{tikzcd}
v \in \mathbb{R}^2 \setminus \{\mathbf{0}\} \arrow[r, mapsto, "\text{saddle}"] \arrow[d, mapsto, "\text{polar angle}"'] & Av \arrow[r, mapsto, "\text{rotation}"] & R_{\pm\tau}(Av) \arrow[d, mapsto, "\text{polar angle}"] \\
\theta \in \mathbb{S}^1 \arrow[r, mapsto, "\text{source-sink}"'] & \tilde{A}(\theta) \arrow[r, mapsto, "\text{translation}"'] & \tilde{A}(\theta) \pm \tau
\end{tikzcd}$$

Now, Lemma 3.3 applies, yielding $\delta > 0$ with the property that $\operatorname{ang}(g(z), Az) \leq \tau$ whenever $0 < |z| < \delta$. If z furthermore satisfies $\alpha < \theta(z) < \pi - \alpha$, then the following two are true simultaneously:

$$\theta(Az) - \tau \leq \theta(g(z)) \leq \theta(Az) + \tau \quad \text{and}$$
$$\alpha < \theta(Az) - 2\tau < \theta(Az) + 2\tau < \pi - \alpha\,.$$

Together they imply $\alpha < \theta(g(z)) - \tau < \theta(g(z)) + \tau < \pi - \alpha$. This is enough to conclude that $R_{\pm\tau}(g(z)) \notin \mathcal{C}_\alpha$ whenever $z \notin \mathcal{C}_\alpha$ also belongs to the upper half plane. An analogous reasoning applies to the case in which z lies on the lower half plane, thus yielding the desired result. ▨

3.5 Scholium. Under the hypotheses of Lemma 3.4, $g^k(z) \notin \mathcal{C}_\alpha$ for every $k \in \mathbb{N}$ such that the orbit $\{z, g(z), \dots, g^{k-1}(z)\}$ remains in $\mathbb{D}_\delta(\mathbf{0})$.

Moving on to a different approach, we now investigate how diffeotopies as a whole act on cones when a certain internal invariant direction is kept fixed throughout.

3.6 Lemma. Let $(g_t)_{t\in\mathbb{I}}$ be a planar diffeotopy such that the origin is a fixed point and $Dg_t(\mathbf{0})$ 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:

$$z \in \mathbb{D}_\rho(\mathbf{0}) \text{ and } z \notin \mathcal{C}_\alpha \text{ imply } g_t(z) \notin \mathcal{C}_\beta \text{ for every } t \in \mathbb{I}\,.$$

Proof. For a fixed $t \in \mathbb{I}$, let $A_t \overset{\text{def}}{=} Dg_t(\mathbf{0})$ and consider $v_\alpha = \left[\begin{smallmatrix}\cos\alpha\\ \sin\alpha\end{smallmatrix}\right]$ and $v_\alpha^* = \left[\begin{smallmatrix}\cos\alpha\\ -\sin\alpha\end{smallmatrix}\right]$, unit vectors of angle α and $2\pi - \alpha$, respectively, whose spans delimit $\mathcal{C}_\alpha$. Given that A_t preserves orientation and that the x-axis is invariant under its action, one has that $A_t v_\alpha$ lies to the left and $A_t v_\alpha^*$ lies to the right of the x-axis, since v_α and v_α^* do so.

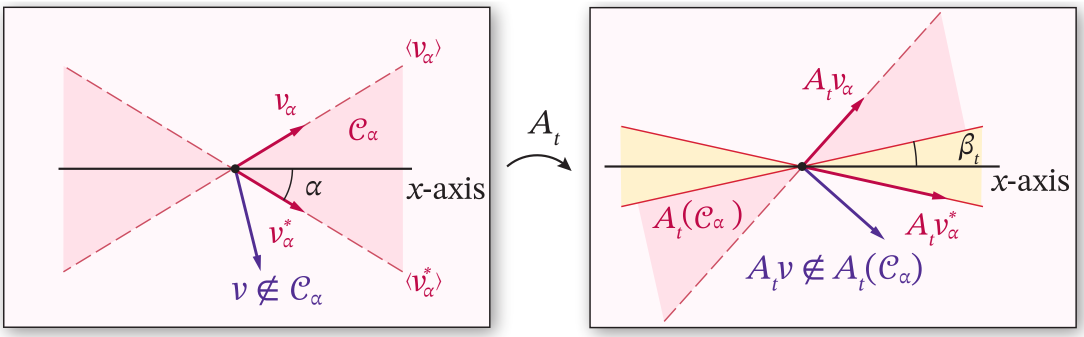
Notice that at least one among $\operatorname{ang}(A_t v_\alpha, {}^{\partial}\!/_{\partial x})$ and $\operatorname{ang}(A_t v_\alpha^*, {}^{\partial}\!/_{\partial x})$ has to be smaller than $\pi/2$. Indeed, if that was not the case, A_t would either revert orientation of the ordered basis $\{v_\alpha, v_\alpha^*\}$ or rupture its linear independence, contradicting the fact that it is an orientation-preserving linear isomorphism. It is claimed that, upon letting

$$\text{[3.2]} \qquad \beta_t \overset{\text{def}}{=} \min\{\alpha\,,\ \operatorname{ang}(A_t v_\alpha, {}^{\partial}\!/_{\partial x})\,,\ \operatorname{ang}(A_t v_\alpha^*, {}^{\partial}\!/_{\partial x})\}\,,$$

we have $0 < \beta_t \leq \alpha$, and also $A_t v \notin \mathcal{C}_{\beta_t}$ holds whenever $v \notin \mathcal{C}_\alpha$.

Indeed, by preservation of orientation, the exterior of $\mathcal{C}_\alpha$ is mapped onto the exterior of the set $A_t(\mathcal{C}_\alpha)$ which, in turn, is delimited by $\langle A_t v_\alpha\rangle \cup \langle A_t v_\alpha^*\rangle$ (because $\partial\mathcal{C}_\alpha = \langle v_\alpha\rangle \cup \langle v_\alpha^*\rangle$), and contains the x-axis. Thus, by the choice of β_t in [3.2], $\mathcal{C}_{\beta_t} \subseteq A_t(\mathcal{C}_\alpha)$. It follows that the complement of $A_t(\mathcal{C}_\alpha)$ is contained in the complement of $\mathcal{C}_{\beta_t}$, as suggested by Figure 3.5.

Figure 3.5 – Boundaries and exteriors of cones are preserved under the action of an orientation-preserving linear isomorphism.



This reasoning yields a global solution to the associated linear problem. Once it is done, let

$$\varepsilon_t = r_t c_t \sin\!\left(\frac{\beta_t}{4}\right)|z|\,, \text{ where } r_t = \frac{1}{1 + \sin(\beta_t/4)}\,,$$

and $c_t > 0$ has the property that $|A_t v| \geq c_t |v|$ for every v.

Having this choice in mind, given that g_t is differentiable at $\mathbf{0}$ it is possible to find $\rho_t > 0$ such that

$$[3.3] \qquad |g_t(z) - A_t z| \leq \frac{\varepsilon_t}{2}\,|z| \quad \text{whenever} \quad 0 < |z| < \rho_t\,.$$

With respect to the compact-open topology in global charts, we consider the following subbasic neighbourhood $\mathcal{B}_t$ of g_t:

$$\mathcal{B}_t \overset{\text{def}}{=} \mathcal{B}\Big(g_t\,;\,\overline{\mathbb{D}}_{\rho_t}(\mathbf{0})\,;\,\varepsilon_t/2\Big) = \left\{ f \in \mathrm{Diff}^1(\mathbb{R}^2)\ :\ \sup_{\overline{\mathbb{D}}_{\rho_t}(\mathbf{0})} |f - g_t| \leq \frac{\varepsilon_t}{2} \ \text{and} \ \sup_{\overline{\mathbb{D}}_{\rho_t}(\mathbf{0})} \|\mathrm{D}f - \mathrm{D}g_t\| \leq \frac{\varepsilon_t}{2} \right\}.$$

Since $t \mapsto g_t$ defines a continuous path of diffeomorphisms with respect to this topology, there exists $\delta_t > 0$ with the following property:

$$[3.4] \qquad s \in \mathbb{I} \text{ and } |s - t| < \delta_t \ \text{ imply } g_s \in \mathcal{B}_t\,.$$

Thus, when $|s - t| < \delta_t$ and $|z| < \rho_t$ simultaneously,

$$\begin{aligned} |g_s(z) - A_t z| \leq |g_s(z) - g_t(z)| + |g_t(z) - A_t z| &= |(g_s - g_t)(z)| + |g_t(z) - A_t z| \\ &\leq \Bigg(\sup_{\overline{\mathbb{D}}_{\rho_t}(\mathbf{0})} \|\mathrm{D}(g_s - g_t)\| \Bigg) |z| + \frac{\varepsilon_t}{2}\,|z| \leq \varepsilon_t\,|z|\,, \end{aligned}$$

where the Mean Value Inequality was applied in the compact convex set $\overline{\mathbb{D}}_{\rho_t}(\mathbf{0})$ to the continuously differentiable function $(g_s - g_t)$, along with [3.3] and [3.4]. So, by the triangle inequality,

$$|g_s(z)| \geq |A_t z| - \varepsilon_t |z| \geq (c_t - \varepsilon_t)\,|z|\,.$$

But, since $0 < \beta_t < \pi/2$, we have $1/2 < r_t < 1$, ensuring that $c_t - \varepsilon_t > 0$. This implies $g_s(z)$ lying in a closed disk of center $A_t z$ and radius $\varepsilon_t |z|$, to which the origin is an external point, as in Remark 3.1. Thus, $\mathrm{ang}(g_s(z), A_t z) < \pi/2$.

The estimate in Remark 3.2 – along with calculations identical to those made in the proof of Lemma 3.3 – yield:

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

These calculations allow us to 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 $z \notin \mathcal{C}_\alpha$, we know from the linear case that $A_t z \notin \mathcal{C}_{\beta_t}$. In other words,

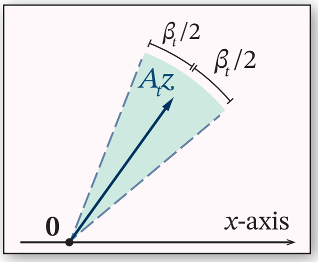
$$[3.5] \qquad \text{either} \ \ \beta_t < \theta(A_t z) < \pi - \beta_t \quad \text{or} \quad \pi + \beta_t < \theta(A_t z) < 2\pi - \beta_t\,,$$

where $\theta(\cdot)$ is the usual anti-clockwise polar angle. On the other hand, as suggested by Figure 3.6,

$$[3.6] \qquad \theta(A_t z) - \frac{\beta_t}{2} \leq \theta(g_s(z)) \leq \theta(A_t z) + \frac{\beta_t}{2}\,.$$

Putting [3.5] and [3.6] together, $g_s(z) \notin \mathcal{C}_{\beta_t/2}$ follows, as long as $z \notin \mathcal{C}_\alpha$, $0 < |z| < \rho_t$ and $|s - t| < \delta_t$.

Figure 3.6 – Angular interval around $A_t z$ on which $g_s(z)$ is allowed to lie.



Lastly, by compactness there exist $t_1, \dots, t_n$ such that $\mathbb{I} = \bigcup_{j=1}^{n}\left(t_j - \delta_{t_j}, t_j + \delta_{t_j}\right) \cap \mathbb{I}$. Thus, the desired conditions are satisfied if we consider

$$\rho \overset{\text{def}}{=} \min_{1\leq j\leq n} \rho_{t_j} \quad \text{and} \quad \beta \overset{\text{def}}{=} \min_{1\leq j\leq n} \frac{\beta_{t_j}}{2}\,.$$

3.7 Corollary (The Cone Lemma). 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 3.6 yields $0 < \beta^+ < \alpha$ and $\rho^+ > 0$ such that $z \in \mathbb{D}_{\rho^+}(\mathbf{0})$ and $z \notin \mathcal{C}_\alpha$ imply $g_t(z) \notin \mathcal{C}_{\beta^+}$ for every $t \in \mathbb{I}$. But, looking at the isotopy $(h_t)_{t\in\mathbb{I}}$ defined by the inverses $h_t \overset{\text{def}}{=} g_t^{-1}$, we notice that it satisfies exactly the same hypotheses as those listed in Lemma 3.6. So, for this isotopy and the angle β^+ just encountered, we obtain a radius $\rho^- > 0$ and a smaller angle $0 < \beta^- < \beta^+$ such that $w \in \mathbb{D}_{\rho^-}(\mathbf{0})$ and $w \notin \mathcal{C}_{\beta^+}$ imply $h_t(w) \notin \mathcal{C}_{\beta^-}$ for every $t \in \mathbb{I}$. Let $\eta > 0$ be such that $|g_t(z)| < \rho^-$ for every $t \in \mathbb{I}$, whenever $|z| < \eta$, as described in Lemma 1.9. Then, by setting ρ as $\min\{\eta, \rho^+\}$, we have the proposed statements satisfied.

Indeed, condition **1)** follows from the choice of β^+. Suppose for the sake of contradiction that **2)** does not hold. Then, $g_s(z) \notin \mathcal{C}_{\beta^+}$ for some $s \in \mathbb{I}$ and some $z \in \mathcal{C}_{\beta^-} \cap \mathbb{D}_\rho(0)$. But in this case, since $|z| < \eta$, the point $w \overset{\text{def}}{=} g_s(z)$ would satisfy $w \in \mathbb{D}_{\rho^-}(\mathbf{0})$ and $w \notin \mathcal{C}_{\beta^+}$. This, in turn, would imply $h_s(w) \notin \mathcal{C}_{\beta^-}$, whilst

$$h_s(w) = h_s(g_s(z)) = g_s^{-1}(g_s(z)) = z \in \mathcal{C}_{\beta^-}\,,$$

a contradiction. Thus, ρ, β^- and β^+ as obtained above must satisfy both of the listed properties.

Before moving on to our Fundamental Lemma, let us recall an important theorem about hyperbolic fixed points of diffeomorphisms, as presented in Chapter 6 of (3), and extract an elementary consequence in the form of a remark.

3.8 The Hadarmard-Perron Theorem. Let x_0 be an hyperbolic fixed point of $f \in \mathrm{Diff}^1(\mathbb{R}^m)$. Then there exists an open neighbourhood U of x_0 such that the *local stable set*:

$$\mathrm{W}^{\mathrm{s}}_{x_0} \overset{\text{def}}{=} \{x \in U : f^n(x) \to x_0 \text{ as } n \to +\infty\}$$

is a submanifold of class C^1, whose tangent space at x_0 is the differential's stable space $E^{\mathrm{s}}_{x_0}$.

In particular, $W^s_{x_0}$ is the graph of a Lipschitz function around x_0. Also, the following holds:

$$f^n(x) \in U \text{ for every } n \in \mathbb{N} \text{ if and only if } x \in W^s_{x_0} .$$

3.9 Remark. In the conditions the Hadamard-Perron Theorem 3.8 above, let B be an open ball such that $\overline{B} \subset U$ and $\mathcal{K} \subset B$ be a compact set not intercepting the stable set. Then, there exists $n_0 \in \mathbb{N}$ such that if $n_x \overset{\text{def}}{=} \min\{n \in \mathbb{N} : g^n(x) \notin \overline{B}\}$, then $n_x \leq n_0$ for every $x \in \mathcal{K}$.

Proof. First notice that, by the characterisation of the stable set as consisting of the only points which remain in U upon iteration, such n_x is well-defined and finite. Now, suppose for the sake of contradiction that the opposite holds: for every $N \in \mathbb{N}$ there exists $x_N \in \mathcal{K}$ such that $n_{x_N} > N$.

Since $\mathcal{K}$ is compact, switching to a subsequence if necessary it can be assumed that $x_N \to \bar{x} \in \mathcal{K}$. Let $N_0 = n_{\bar{x}}$. Then, $g^{N_0}(\bar{x}) \in (\overline{B})^c$, which is an open set. If V is a neighbourhood of $g^{N_0}(\bar{x})$ fully contained within $(\overline{B})^c$, $g^{-N_0}(V) \cap \mathcal{K}$ is an open neighbourhood of $\bar{x}$ in $\mathcal{K}$ and thus contains infinitely many x_N of the sequence.

Discarding the trivial case in which $\mathcal{K}$ consists of a finite number of isolated points, these x_N may be supposed all distinct. In particular, for some sufficiently large $M > N_0$ one has that x_M lies in this neighbourhood. Therefore, $g^{N_0}(x_M) \notin \overline{B}$, a contradiction, since $n_{x_M} > M > N_0$ was supposed to be minimal with the property that $g^{n_{x_M}}(x_M) \notin \overline{B}$. ▨

3.10 The Fundamental Lemma. Let $G \subset \mathrm{Diff}^1_+(\mathbb{S}^2)$ be a proper extension of $\mathrm{M\ddot{o}b}(\mathbb{S}^2)$. Then, for a given point z_0 *not* on the meridian Γ there exists an $\mathcal{I}G_3$ isotopy $\left(I^{z_0}_t\right)_{t\geq 0}$, *depending on* z_0, such that:

1) the trajectory of z_0 under I^{z_0} does not intercept 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 its planar counterpart 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 $(g_t)_{t\in\mathbb{I}}$ be as in the Extension Lemma 2.10. Since ∞ is fixed throughout, it can be thought of as a planar $\mathcal{I}G_1$-isotopy 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 $g \overset{\text{def}}{=} g_1$ and $A \overset{\text{def}}{=} \mathrm{D}g(\mathbf{0})$.

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 3.4:

$$[3.7] \qquad 0 < |z| < \delta \ \text{ and } \ z \notin \mathcal{C}_\alpha \ \text{ imply } \ R_\omega(g(z)) \notin \mathcal{C}_\alpha \text{ whenever } |\omega| < \tau .$$

Regarding this same α – and also the isotopy $(g_t)_{t\in\mathbb{I}}$ – the Cone Lemma 3.7 yields a radius $\rho > 0$ and angles $0 < \beta^- < \beta^+ < \alpha$ such that, for $z \in \mathbb{D}_\rho(\mathbf{0})$ and every $t \in \mathbb{I}$:

$$[3.8] \qquad \begin{aligned} z \notin \mathcal{C}_\alpha &\text{ implies } g_t(z) \notin \mathcal{C}_{\beta^+} , \\ z \in \mathcal{C}_{\beta^-} &\text{ implies } g_t(z) \in \mathcal{C}_{\beta^+} . \end{aligned}$$

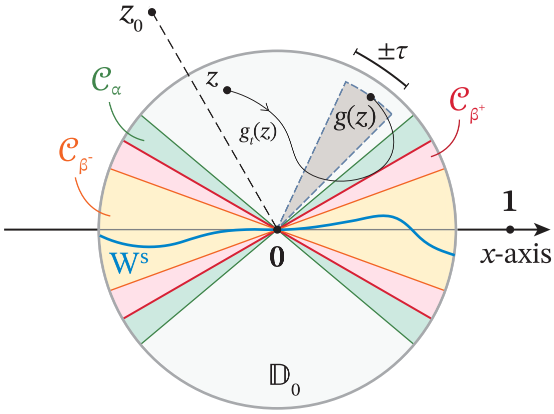
Lastly, the Hadamard-Perron Theorem 3.8 is used to characterise the stable manifold of g, denoted simply by W^s. Since $E^s_{\mathbf{0}}$ is the x-axis, by shrinking U if necessary W^s may be assumed to be a

Lipschitz graph of the form $y = y(x)$. So, for the angles τ and β^- as obtained above and a sufficiently small radius $\sigma > 0$, we may assume that

$$\mathrm{W}^s \cap \mathbb{D}_\sigma(\mathbf{0}) \subset \mathcal{C}_{\min\{\tau,\beta^-\}}\,. \tag{3.9}$$

Let $0 < \rho_0 < \min\{\delta, \rho, \sigma, 1\}$. Then, in the disk $\mathbb{D}_{\rho_0}(\mathbf{0})$, which shall be denoted simply by $\mathbb{D}_0$, conditions [3.7] through [3.9] are mutually satisfied, as pictorially suggested by Figure 3.7.

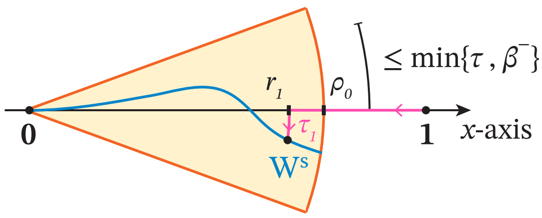
Figure 3.7 – Inside the disk $\mathbb{D}_0$ the isotopy path of points z outside of the α-cone never enter the β^+-cone, while the stable manifold of g is a Lipschitz graph fully contained within the $\min\{\tau, \beta^-\}$-cone.



Once these choices are made, fix a positive real number $0 < r_1 < \min\{\rho_0, \rho_0 / R_0\}$ and let $\rho_1 = r_1 R_0$ and $\mathcal{K}_0 = \partial\mathbb{D}_{\rho_1}(\mathbf{0}) \setminus \mathcal{C}_\alpha$. Then, $\mathcal{K}_0$ is a two-component compact set contained in $\mathbb{D}_0$, not containing $\mathbf{0}$ nor intercepting the stable manifold. In particular, by Remark 3.9 there exists $n_0 \in \mathbb{N}$ such that $n_x \leq n_0$ for every $x \in \mathcal{K}_0$, where $n_x = \min\{n \in \mathbb{N} : g^n(x) \notin \overline{\mathbb{D}_0}\}$.

Notice that $r_1 < \rho_0$. So, by the graph characterisation of W^s and by [3.9], for some τ_1 with $|\tau_1| \leq \min\{\tau, \beta^-\}$ we have $r_1 e^{i\tau_1} \in \mathrm{W}^s$, as suggested by figure 3.8.

Figure 3.8 – Since the stable manifold is locally given as a Lipschitz graph $y = y(x)$ it can be reached from the x-axis through an uniformly bounded rotation.



3.10.1 Claim. Define $M_1(z) = r_1 e^{i\tau_1}\, z$. Then, the mapping M_1 has the following properties:

(i) $M_1 \in \mathrm{Möb}_2(\mathbb{S}^2)$; (ii) $v_1 \stackrel{\text{def}}{=} M_1(\mathbf{1}) \in \mathrm{W}^s$; (iii) $w_1 \stackrel{\text{def}}{=} M_1(z_0) \in \mathcal{K}_0$.

Proof of Claim. Item (i) follows from the form of M_1 and (ii) follows from the choice of τ_1 depicted in Figure 3.8. With respect to (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 θ_0 as z_0, we have $\theta(w_1) = \theta_0 + \tau_1$. But α was chosen such that $2\alpha < \theta_0 < \pi - 2\alpha$, so the following holds:

$$|\tau_1| \leq \beta^- < \alpha < \min\left\{\frac{\theta_0}{2}\,,\, \frac{\pi - \theta_0}{2}\right\}.$$

Thus,

$$\theta_0 + |\tau_1| < \theta_0 + \frac{\pi - \theta_0}{2} = \frac{\theta_0 + \pi}{2} \leq \frac{(\pi - 2\alpha) + \pi}{2} = \pi - \alpha \quad \text{and}$$
$$\theta_0 - |\tau_1| > \theta_0 - \frac{\theta_0}{2} = \frac{\theta_0}{2} > \alpha\,.$$

Since $\theta_0 - |\tau_1| \leq \theta_0 + \tau_1 \leq \theta_0 + |\tau_1|$, we conclude that $\alpha < \theta_0 + \tau_1 < \pi - \alpha$ or, equivalently, $w_1 \notin \mathcal{C}_\alpha$. Furthermore, $|w_1| = |r_1 z_0| = r_1 R_0 = \rho_1$, yielding $w_1 \in \mathcal{K}_0$. □

In particular, $n_1 \stackrel{\text{def}}{=} n_{w_1}$ is well-defined.

3.10.2 Claim. Given $0 < t \leq n_1$, consider $f_t = \mathrm{g}_{t-\lfloor t \rfloor} \circ \mathrm{g}^{\lfloor t \rfloor} \circ M_1$. Then, $t \mapsto f_t$ is an isotopy satisfying

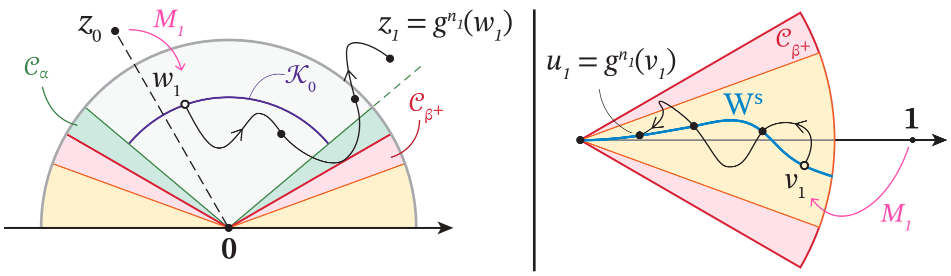
$$f_t(z_0) \notin \mathcal{C}_{\beta^+} \text{ and } f_t(\mathbf{1}) \in \mathcal{C}_{\beta^+} \text{ for every } t \in (0, n_1]\,.$$

Proof. On the one hand, $f_t(z_0) = \mathrm{g}_{t-\lfloor t \rfloor}\left(\mathrm{g}^{\lfloor t \rfloor}(w_1)\right)$. Since $w_1 \notin \mathcal{C}_\alpha$, as long as $t < n_1$ each $\mathrm{g}^{\lfloor t \rfloor}(w_1)$ does not belong to $\mathcal{C}_\alpha$ either, as observed in Scholium 3.5. On the other hand, they do belong to $\mathbb{D}_0$. Therefore, the Cone Lemma implies $\mathrm{g}_s\left(\mathrm{g}^{\lfloor t \rfloor}(w_1)\right) \notin \mathcal{C}_{\beta^+}$ for every $s = t - \lfloor t \rfloor \in \mathbb{I}$.
On the other hand, $f_t(\mathbf{1}) = \mathrm{g}_{t-\lfloor t \rfloor}(\mathrm{g}^{\lfloor t \rfloor}(v_1))$. But since $v_1 \in \mathrm{W}^{\mathrm{s}}$, each $\mathrm{g}^{\lfloor t \rfloor}(v_1)$ belongs to $\mathrm{W}^{\mathrm{s}} \cap \mathbb{D}_0$, which is contained in $\mathcal{C}_{\beta^-}$, by [3.9]. Therefore, $\mathrm{g}_s(\mathrm{g}^{\lfloor t \rfloor}(v_1)) \in \mathcal{C}_{\beta^+}$ for every $s = t - \lfloor t \rfloor \in [0, 1]$, also by the Cone Lemma. □

This setting is illustrated in figure 3.9, where the following points were introduced:

[3.10] $$z_1 \stackrel{\text{def}}{=} f_{n_1}(z_0) = \mathrm{g}^{n_1}(w_1) \notin \mathcal{C}_\alpha \quad \text{and} \quad u_1 \stackrel{\text{def}}{=} f_{n_1}(\mathbf{1}) = \mathrm{g}^{n_1}(v_1) \in \mathrm{W}^{\mathrm{s}}\,.$$

Figure 3.9 – f_t promotes two parallel processes: "macroscopically", points in $\mathcal{K}_0$ are successively dragged out of $\mathbb{D}_0$ without entering the β^+-cone, while "microscopically" the images of $\mathbf{1}$ are dragged towards the origin over the stable manifold.



By the choice of n_1, $|z_1| > \rho_0$. Let $r_2 \stackrel{\text{def}}{=} \rho_1/|z_1| < 1$ and notice that, since the product by r_2 does not change angles:

- $|r_2 z_1| = \rho_1$ and, given that $z_1 \notin \mathcal{C}_\alpha$, one has $r_2\, z_1 \in \mathcal{K}_0$.
- Also, $r_2 u_1 \in \mathbb{D}_0 \cap \mathcal{C}_{\min\{\tau,\beta^-\}}$, because $u_1 \in \mathrm{W}^{\mathrm{s}} \subset \mathbb{D}_0 \cap \mathcal{C}_{\min\{\tau,\beta^-\}}$.

In particular, by the second bullet above, the graph characterisation of W^{s} and [3.9], one has $e^{i\tau_2}(r_2 u_1) \in \mathrm{W}^{\mathrm{s}}$, for some τ_2 with $|\tau_2| \leq \min\{\tau, \beta^-\}$.

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

(i) $M_2 \in \mathrm{M\ddot{o}b}_2(\mathbb{S}^2)$, (ii) $v_2 \stackrel{\text{def}}{=} M_2(u_1) \in \mathrm{W}^{\mathrm{s}}$, (iii) $w_2 \stackrel{\text{def}}{=} M_2(z_1) \in \mathcal{K}_0$.

Proof. Item (i) is immediate from the form of M_2, while (ii) is a consequence of the very own choice of τ_2.

As for (iii), recall from [3.10] that $z_1 \notin \mathcal{C}_\alpha$. It follows that $r_2\, z_1$ does not belong to $\mathcal{C}_\alpha$ as well – because it is obtained from z_1 through an homothety – but it does belong to $\mathbb{D}_0$. So [3.7] 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 $w_2 \in K_0$. $\square$

In particular, $n_2 \stackrel{\text{def}}{=} n_{w_2}$ is well-defined. For $n_1 < t \leq n_1 + n_2$, consider now the expression $f_t = g_{t-\lfloor t \rfloor} \circ g^{\lfloor t \rfloor - n_1} \circ M_2 \circ f_{n_1}$. By arguments analogous to the ones developed previously, $t \mapsto f_t$ is seen to be continuous (with respect to d_∞) 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 $t \in (n_1, n_1 + n_2]$. In a similar fashion, we define inductively, for $k \geq 0$:

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

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:

- $M_{k+1} \in \text{Möb}_2(\mathbb{S}^2)$ is a transformation of the form
$$M_{k+1}(z) = r_{k+1} e^{i\tau_{k+1}} z\,,$$
which maps $z_k \stackrel{\text{def}}{=} f_{N_k}(z_0) \notin \overline{\mathbb{D}_0}$ to a point $w_{k+1} \in \mathcal{K}_0$ and $u_k \stackrel{\text{def}}{=} f_{N_k}(\mathbf{1}) \in \mathrm{W}^{\mathrm{s}}$ 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 $|\tau_{k+1}| \leq \min\{\tau, \beta^-\}$;
- $n_k \stackrel{\text{def}}{=} n_{w_k} = \min\big\{n \in \mathbb{N} \,:\, g^n(w_k) \notin \overline{\mathbb{D}_0}\big\} \leq n_0$.

The following properties hold, by construction:

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

(ii) $t \mapsto f_t$ is d_∞-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 $t \geq 0$.

3.10.4 Claim. 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 Lemma 1.19. Then, $\big(I_t^{z_0}\big)_{t\geq 0}$ is an $\mathcal{I}G_3$-isotopy.

Proof. Since $f_t \in G_2$ for every $t \geq 0$, it is clear that $I_t^{z_0} \in G_3$ for every $t \geq 0$. It is left to verify that $t \mapsto I_t^{z_0}$ defines a d_∞-continuous curve of homeomorphisms. By Lemma 1.19, the mappings $t \mapsto \hat{M}[f_t(\mathbf{1})]$ – and thus $t \mapsto I_t^{z_0}$ – are *a priori* as continuous as $t \mapsto f_t$.
In other words, over the intervals $(N_k, N_{k+1}]$. Thus, all that is needed to check is continuity from the right at their left endpoints. Now, for each $0 < h < 1$:

$$\begin{aligned} I_{N_k+h}^{z_0} &= \hat{M}[f_{N_k+h}(\mathbf{1})] \circ f_{N_k+h} \\ &= \hat{M}\big[g_h \circ g^0 \circ M_{k+1} \circ f_{N_k}(\mathbf{1})\big] \circ g_h \circ g^0 \circ M_{k+1} \circ f_{N_k} \\ &= \hat{M}\big[g_h(M_{k+1}(u_k))\big] \circ g_h \circ M_{k+1} \circ f_{N_k} \\ &= \hat{M}\big[g_h(v_{k+1})\big] \circ g_h \circ M_{k+1} \circ f_{N_k}\,. \end{aligned}$$

But, notice that

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

is *a* Möbius transformation fixing the poles and mapping $g_h(v_{k+1})$ to $\mathbf{1}$. By sharp 3-transitivity, it must be *the* transformation $\hat{M}\big[g_h(v_{k+1})\big]$. Since $g_h \xrightarrow{d_\infty}$ id as $h \to 0^+$, the continuity described in Lemma 1.19 applied to the expression [3.11] above yields:

$$\hat{M}\big[g_h(v_{k+1})\big] \xrightarrow{d_\infty} \hat{M}[u_k] \circ M_{k+1}^{-1} \circ \underbrace{M[v_{k+1}, v_{k+1}]}_{=\mathrm{id}} = \hat{M}[u_k] \circ M_{k+1}^{-1} \text{ as } h \to 0^+ \, .$$

Consequently,

$$I_{N_k+h}^{z_0} \xrightarrow{d_\infty} \hat{M}[u_k] \circ \underbrace{M_{k+1}^{-1} \circ g_0 \circ M_{k+1}}_{=\mathrm{id}} \circ f_{N_k} = \hat{M}[u_k] \circ f_{N_k} = \hat{M}\big[f_{N_k}(\mathbf{1})\big] \circ f_{N_k} = I_{N_k}^{z_0} \text{ as } t \to 0^+ \, ,$$

thus proving the desired continuity at the instants N_k. □

3.10.5 Claim. $\gamma_{I^{z_0}}(z_0) \cap \Gamma = \varnothing$.

Proof. Notice that, 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 $\hat{M}[f_t(\mathbf{1})](z) = |f_t(\mathbf{1})|^{-1} e^{i\psi} z$, for some $|\psi| \leq \beta^+$. Consequently,

$$I_t^{z_0}(z_0) = |f_t(\mathbf{1})|^{-1} e^{i\psi} f_t(z_0) \, .$$

But it is also known that $f_t(z_0) \notin \mathcal{C}_{\beta^+}$ for every $t \geq 0$. Thus, $f_t(z_0) = |f_t(z_0)| e^{i\theta}$, where either $\beta^+ < \theta < \pi - \beta^+$ or $\pi + \beta^+ < \theta < 2\pi - \beta^+$. Therefore, $-\beta^+ \leq \psi \leq \beta^+$ implies $\theta + \psi \notin \{0, \pi, 2\pi\}$. Since

$$I_t^{z_0}(z_0) = |f_t(z_0)| |f_t(\mathbf{1})|^{-1} e^{i(\theta+\psi)} \, .$$

we conclude that $\theta(I_t^{z_0}(z_0)) \notin \{0, \pi, 2\pi\}$ for every $t \geq 0$. In other words, the trajectory $\gamma_{I^{z_0}}(z_0)$ remains on the upper half-plane without ever touching the x-axis or – equivalently – remains on the left hemisphere of the sphere without ever touching the meridian Γ. This proves **1)**. □

3.10.6 Claim. $f_t(\mathbf{1}) \to \mathbf{0}$ as $t \to +\infty$.

Proof. 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 $|v_{k+1}| = r_{k+1}|u_k|$,

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

But on each interval $(N_k, N_{k+1}]$, we have $f_t(\mathbf{1}) = g_{t-\lfloor t \rfloor} \circ g^{\lfloor t \rfloor - N_k}(v_{k+1})$. 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,

[3.12] $$\sup\{|f_t(\mathbf{1})| \,:\, t > N_k\} \leq \max\Big\{|g_s^i(z)| \,:\, s \in \mathbb{I} \,, 1 \leq i \leq n_0 \text{ and } z \in \overline{\mathbb{D}}_{(\rho_1/\rho_0)^{k+1}}(\mathbf{0})\Big\} \, ,$$

where the right-hand side is seen to be a finite positive number, by compacity of all the sets involved along with the isotopy's joint continuity.

Let $\varepsilon > 0$ be given. Since each $(g^i_t)_{t\in\mathbb{I}}$ is an isotopy fixing the origin, Lemma 1.9 yields $\eta > 0$ such that $|g^i_s(z)| < \varepsilon$ for every $s \in \mathbb{I}$ and $1 \le i \le n_0$, whenever $|z| < \eta$. Consequently, if $k_0 \in \mathbb{N}$ is so large that $(\rho_1/\rho_0)^{k_0+1} < \eta$, [3.12] implies $|f_t(\mathbf{1})| < \varepsilon$ whenever $t > N_{k_0}$, establishing the claimed limit. □

We are, now, ready to finish the proof of the Fundamental Lemma. 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 Claim 3.10.6 above we see that, for every $t \ge 0$,

$$f_t(z_0) \in \mathcal{K} \stackrel{\text{def}}{=} \left\{ g^i_s(z) \,:\, s \in \mathbb{I}\,, 1 \le i \le n_0 \text{ and } z \in \mathcal{K}_0 \right\}.$$

Therefore, the segments of the trajectory $\gamma_{I^{z_0}}(z_0)$ corresponding to large instants t are contained within the corresponding $\hat{M}[f_t(\mathbf{1})](\mathcal{K})$ sets. However, $\mathcal{K}$ is a compact set bounded away from $\mathbf{0}$, by the injectivity of the diffeomorphisms g^i_s. Since $f_t(\mathbf{1}) \to \mathbf{0}$ as $t \to +\infty$, these sets are dragged towards infinity in the Hausdorff distance as $t \to +\infty$, as described by Lemma 1.19. Consequently, the only point in the trajectory's accumulation can be ∞, proving **2)**. ▨

As important as the Fundamental Lemma itself is the following Corollary, which states that isotopies can be build in the subgroup G_3 accumulating at any given pair $\mathbf{a}, \mathbf{b} \in \{\mathbf{0}, \mathbf{1}, \infty\}$ of reference points.

3.11 Corollary. Let $G \subset \mathrm{Diff}^1_+(\mathbb{S}^2)$ be a proper extension of $\mathrm{Möb}(\mathbb{S}^2)$. Then, for each z_0 not on the meridian Γ and each pair of distinct points $\mathbf{a}, \mathbf{b} \in \{\mathbf{0}, \mathbf{1}, \infty\}$ there exists a full-time $\mathcal{I}G_3$ isotopy $\left(I^{z_0}_{\mathbf{ab}}(t,\cdot)\right)_{t\in\mathbb{R}}$ such that:

1) the trajectory of z_0 under $I^{z_0}_{\mathbf{ab}}$ does not intercept Γ,
2) the α and ω limits of z_0 satisfy $\alpha_{I^{z_0}_{\mathbf{ab}}}(z_0) = \{\mathbf{a}\}$ and $\omega_{I^{z_0}_{\mathbf{ab}}}(z_0) = \{\mathbf{b}\}$.

Proof. Let $z_0 \notin \Gamma$ and $\mathbf{a}, \mathbf{b} \in \{\mathbf{0}, \mathbf{1}, \infty\}$ be given. Upon recalling Table 1.2, we see that neither $T_{\infty\mathbf{b}}(z_0)$ nor $T_{\infty\mathbf{a}}(z_0)$ lie on Γ, so the Fundamental Lemma 3.10 can be applied to both of them, yielding $\mathcal{I}G_3$ isotopies $\left(I^{T_{\infty\mathbf{b}}(z_0)}_t\right)_{t\ge 0}$ and $\left(I^{T_{\infty\mathbf{a}}(z_0)}_t\right)_{t\ge 0}$ as described therein, from which we define $I^{z_0}_{\mathbf{ab}} : \mathbb{R} \times \mathbb{S}^2 \to \mathbb{S}^2$ as:

$$I^{z_0}_{\mathbf{ab}}(t,\, z) = \begin{cases} T_{\infty\mathbf{b}} \circ I^{T_{\infty\mathbf{b}}(z_0)}_t \circ T_{\infty\mathbf{b}}(z) & \text{if } t \ge 0\,, \\ T_{\infty\mathbf{b}} \circ I^{T_{\infty\mathbf{a}}(z_0)}_{-t} \circ T_{\infty\mathbf{b}}(z) & \text{if } t \le 0\,. \end{cases}$$

Now, each mapping $I^{z_0}_{\mathbf{ab}}\restriction_{[0,+\infty)\times\mathbb{S}^2}$ and $I^{z_0}_{\mathbf{ab}}\restriction_{(-\infty,0]\times\mathbb{S}^2}$ is itself an isotopy, and they agree in the common slice $\{0\} \times \mathbb{S}^2$, both being equal to $\mathrm{id}_{\mathbb{S}^2}$ there. Consequently, $I^{z_0}_{ab}$ defines a global jointly continuous function of the variables $(t, z) \in \mathbb{R} \times \mathbb{S}^2$. The statements then follow from the Fundamental Lemma, recalling that each $T_{\mathbf{ab}}$ preserves Γ, and also that reversion of time turns the ω-limit into the α-limit. ▨

Chapter 4

Transitivity and entropy

4.1 A crossing lemma

The Fundamental Lemma 3.10 and its Corollary 3.11 gave us insight on how points in $\mathbb{S}^2 \setminus \Gamma$ can be moved under the action of the stabilising subgroup G_3 of a proper extension $G \subset \mathrm{Diff}^1_+(\mathbb{S}^2)$ of $\mathrm{Möb}(\mathbb{S}^2)$. We now wish to understand how this same subgroup acts on Γ. Our ultimate goal is to prove that it can take points "from one side of Γ to the other". To make this idea precise, we start by recalling some elementary facts about the stability of transverse intersections.

4.1 Lemma. Let C be an embedded planar curve of class C^1. If $\lambda : \mathbb{I} \to \mathbb{R}$ is a given $C^1(\mathbb{I}, \mathbb{R}^2)$ path that intersects C *transversally* at $\lambda(t_0)$, then there exist $\varepsilon > 0$ and $\delta > 0$ such that, for every path β that is C^1 ε-close to λ,

$$\Big|\beta\big((t_0 - \delta, t_0 + \delta)\big) \cap \mathrm{C}\Big| = 1 .$$

Furthermore, this unique intersection is transversal as well. Above:

- *transversal* means that the directions of the velocity vector $\lambda'(t_0)$ and of the (one-dimensional) tangent space $T_{\lambda(t_0)}\mathrm{C}$ are linearly independent;
- C^1 ε-close means that both β and β' are ε-close to λ and λ', in the following sense:

$$[4.1] \qquad |\beta(t) - \lambda(t)| \le \varepsilon \ \text{ and } \ |\beta'(t) - \lambda'(t)| \le \varepsilon \ \text{ for every } t \in \mathbb{I} .$$

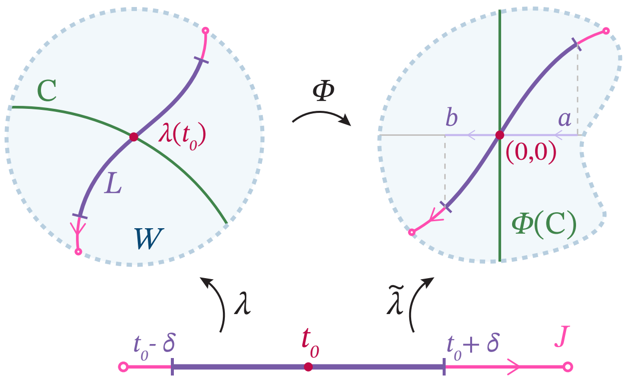
Proof. By the local characterisation of embedded submanifolds, C admits a slice-chart (W, Φ) around $\lambda(t_0)$. That is to say, a C^1 diffeomorphism $\Phi : W \to \Phi(W)$ from an open set onto its image such that $\Phi(\lambda(t_0)) = (0,0)$ and $\Phi(\mathrm{C} \cap W) = \{(0, y) : y \in \mathbb{R}\} \cap \Phi(W)$.

Let J be the connected component of $\lambda^{-1}(W)$ containing t_0. It may be assumed to be an open interval of $\mathbb{I}$, containing t_0 in its interior. Thus, by letting $\tilde{\lambda} \overset{\text{def}}{=} \Phi \circ \lambda : J \to \tilde{W}$ one obtains a C^1 path, of components $\tilde{\lambda} = (\tilde{\lambda}_1, \tilde{\lambda}_2)$.

Since $\mathrm{D}\Phi(\lambda(t_0))$ is a linear isomorphism, it preserves linear independence and, consequently, transversality. Given that C corresponds under Φ to a vertical segment, this translates to $\tilde{\lambda}_1'(t_0) \neq 0$. Say, without loss of generality, that $\tilde{\lambda}_1(t_0) = 0$ and $\tilde{\lambda}_1{}'(t_0) = c < 0$. Upon looking at the continuously differentiable real function $\tilde{\lambda}_1 : J \to \mathbb{R}$, it is then possible to find – as suggested by Figure 4.1 – some $\delta > 0$ such that $[t_0 - \delta, t_0 + \delta] \subset J$,

$$[4.2] \qquad \tilde{\lambda}_1(t_0 - \delta) = a > 0 \quad , \quad \tilde{\lambda}_1(t_0 + \delta) = b < 0 \quad \text{and} \quad \max_{[t_0-\delta, t_0+\delta]} \tilde{\lambda}_1' = \frac{c}{2} < 0 .$$

Figure 4.1 – A slice chart rectifies C and turns λ into a path whose horizontal component is monotone on a small neighbourhood of the instant t_0, yielding stability of the transversal intersection.



Given such δ, $L \overset{\text{def}}{=} \lambda([t_0-\delta, t_0+\delta])$ is a compact subset of the open set W, and thus admits a neighbourhood fully contained within W, say $L_\rho \overset{\text{def}}{=} \{q : \operatorname{dist}(q, L) \le \rho\} \subset W$. Let $\tilde\varepsilon = \min\{a, |b|, |c|\}/2 > 0$. Since $\mathrm{D}\Phi$ is uniformly continuous on the compact set L_ρ, we may fix $\eta > 0$ such that

$$[4.3] \quad \|\mathrm{D}\Phi(z) - \mathrm{D}\Phi(w)\| < \frac{\tilde\varepsilon}{2m} \text{ whenever } z, w \in L_\rho \text{ and } |z-w| < \eta\,, \text{ where } m = \max_{[t_0-\delta, t_0+\delta]} |\lambda'| > 0\,.$$

Also, we have $M \overset{\text{def}}{=} \max_{L_\rho} \|\mathrm{D}\Phi\| < +\infty$.

It is claimed that, upon defining $\varepsilon = \min\{\rho, \eta, \tilde\varepsilon/(2M)\} > 0$, if β is C^1 ε-close to λ, then $\tilde\beta \overset{\text{def}}{=} \Phi \circ \beta\restriction_J$ is C^1 $\tilde\varepsilon$-close to $\tilde\lambda$ in $[t_0-\delta, t_0+\delta]$. Indeed, on the one hand the Mean Value Inequality yields:

$$|\tilde\beta(t) - \tilde\lambda(t)| = |\Phi(\beta(t)) - \Phi(\lambda(t))| \le \Big(\max_{[\lambda(t),\beta(t)]} \|\mathrm{D}\Phi\|\Big)|\beta(t) - \lambda(t)|\,.$$

But, since $|\beta(t) - \lambda(t)| \le \varepsilon \le \rho$, the segment $[\lambda(t), \beta(t)]$ is contained in L_ρ. Therefore:

$$|\tilde\beta(t) - \tilde\lambda(t)| = |\Phi(\beta(t)) - \Phi(\lambda(t))| \le M\varepsilon \le M\Big(\frac{\tilde\varepsilon}{2M}\Big) = \frac{\tilde\varepsilon}{2}\,.$$

On the other hand, as long as $t \in [t_0-\delta, t_0+\delta]$:

$$\begin{aligned}
|\tilde\beta'(t) - \tilde\lambda'(t)| &= |\mathrm{D}\Phi(\beta(t))(\beta'(t)) - \mathrm{D}\Phi(\lambda(t))(\lambda'(t))| \\
&\le |\mathrm{D}\Phi(\beta(t))(\beta'(t)) - \mathrm{D}\Phi(\beta(t))(\lambda'(t))| + |\mathrm{D}\Phi(\beta(t))(\lambda'(t)) - \mathrm{D}\Phi(\lambda(t))(\lambda'(t))| \\
&\le \|\mathrm{D}\Phi(\beta(t))\|\,|\beta'(t) - \lambda'(t)| + \|\mathrm{D}\Phi(\beta(t)) - \mathrm{D}\Phi(\lambda(t))\|\,|\lambda'(t)| \\
&\le M\,|\beta'(t) - \lambda'(t)| + \|\mathrm{D}\Phi(\beta(t)) - \mathrm{D}\Phi(\lambda(t))\|\,m\,,
\end{aligned}$$

where the Mean Value Inequality was once again used, along with the fact that $\beta(t) \in L_\rho$. Since $|\beta(t) - \lambda(t)| \le \varepsilon \le \eta$, one can plug the estimate from [4.3] in the above inequality, along with the fact that $|\beta'(t) - \lambda'(t)| \le \varepsilon \le \tilde\varepsilon/(2M)$, to obtain:

$$|\tilde\beta'(t) - \tilde\lambda'(t)| \le M\Big(\frac{\tilde\varepsilon}{2M}\Big) + \Big(\frac{\tilde\varepsilon}{2m}\Big)m = \frac{\tilde\varepsilon}{2} + \frac{\tilde\varepsilon}{2} = \tilde\varepsilon\,.$$

Since $t \in [t_0-\delta, t_0+\delta]$ was arbitrary, this establishes the desired C^1 $\tilde\varepsilon$-closeness, from which follows

$$\tilde\beta_1(t_0+\delta) \le \tilde\lambda_1(t_0+\delta) + \tilde\varepsilon = b + \tilde\varepsilon \le a + \Big(-\frac{b}{2}\Big) = \frac{b}{2} < 0\,.$$

Analogously, $\tilde{\beta}_1(t_0 - \delta) > 0$. Therefore, there exists at least one $\bar{t} \in (t_0 - \delta, t_0 + \delta)$ such that $\tilde{\beta}_1(\bar{t}) = 0$. However, the last condition in [4.2] ensures that $\tilde{\lambda}_1'$ is never zero on this interval, from which it follows that such $\bar{t}$ is unique.

Since $\tilde{\lambda}_1(\bar{t}) = 0$ is equivalent to $\tilde{\lambda}(\bar{t})$ in the vertical segment $\Phi(\mathrm{C})$, and $\tilde{\lambda}_1'(\bar{t}) \neq 0$ is equivalent to say that this intersection is transversal, the result just obtained translates under Φ^{-1} to a (unique) transversal intersection between C and β, at $\bar{t} \in (t_0 - \delta, t_0 + \delta)$. ▨

Given $r > 0$, we let $\gamma_r : \mathbb{I} \to \mathbb{R}$ be the path $\gamma_r(\theta) = re^{2\pi i\theta}$ and denote its image – the circle of radius r centred at the origin – by $S_r = \gamma_r(\mathbb{I})$. In particular, when $r = 1$, S_1 is the unit circle $\mathbb{S}^1$. If $g \in \mathrm{Diff}^1(\mathbb{R})$ is a planar diffeomorphism fixing the origin, then

$$\frac{g \circ \gamma_r}{r} \xrightarrow{C^1(\mathbb{I},\mathbb{R}^2)} \mathrm{Dg}(\mathbf{0}) \circ \gamma_1 \ \text{ as } \ r \to 0^+ \ , \tag{4.4}$$

where the above $C^1\big(\mathbb{I}, \mathbb{R}^2\big)$ convergence refers to [4.1]. Indeed, for each $\theta \in \mathbb{I}$:

$$\Big(\frac{g \circ \gamma_r}{r}\Big)(\theta) = \frac{g\big(re^{2\pi i\theta}\big)}{r} \quad \text{and} \quad \big(\mathrm{Dg}(\mathbf{0}) \circ \gamma_1\big)(\theta) = \mathrm{Dg}(\mathbf{0})\big(e^{2\pi i\theta}\big) \, .$$

Thus, using basic linearity properties of the derivative:

$$\Big(\frac{g \circ \gamma_r}{r}\Big)'(\theta) = 2\pi \, \mathrm{Dg}(re^{2\pi i\theta})\big(ie^{2\pi i\theta}\big) \quad \text{and} \quad \big(\mathrm{Dg}(\mathbf{0}) \circ \gamma_1\big)'(\theta) = 2\pi \, \mathrm{Dg}(\mathbf{0})\big(ie^{2\pi i\theta}\big) \, .$$

If we let $\varepsilon > 0$ be given,

- by differentiability, we obtain $\eta_1 > 0$ such that $\dfrac{|g(z) - \mathrm{Dg}(\mathbf{0})(z)|}{|z|} < \varepsilon$ for $0 < |z| < \eta_1$;
- by C^1 regularity, we obtain $\eta_2 > 0$ such that $\|\mathrm{Dg}(z) - \mathrm{Dg}(\mathbf{0})\| < \dfrac{\varepsilon}{2\pi}$ when $|z| < \eta_2$.

Let us define $\eta \stackrel{\text{def}}{=} \min\{\eta_1, \eta_2\}$. When $0 < r < \eta$, any point of of the form $z = \gamma_r(\theta)$ satisfies the above conditions. Therefore, uniformly for $\theta \in \mathbb{I}$:

$$\Big|\Big(\frac{g \circ \gamma_r}{r}\Big)(\theta) - (\mathrm{Dg}(\mathbf{0}) \circ \gamma_1)(\theta)\Big| = \frac{\big|g\big(re^{2\pi i\theta}\big) - \mathrm{Dg}(\mathbf{0})\big(re^{2\pi i\theta}\big)\big|}{r} = \frac{|g(z) - \mathrm{Dg}(\mathbf{0})(z)|}{|z|} < \varepsilon$$

and

$$\begin{aligned} \Big|\Big(\frac{g \circ \gamma_r}{r}\Big)'(\theta) - (\mathrm{Dg}(\mathbf{0}) \circ \gamma_1)'(\theta)\Big| &= \big|2\pi \mathrm{Dg}(re^{2\pi i\theta})\big(ie^{2\pi i\theta}\big) - 2\pi \mathrm{Dg}(\mathbf{0})\big(ie^{2\pi i\theta}\big)\big| \\ &\leq 2\pi \Big\|\mathrm{Dg}(re^{2\pi i\theta}) - \mathrm{Dg}(\mathbf{0})\Big\| \, \big|ie^{2\pi i\theta}\big| \\ &= 2\pi \|\mathrm{Dg}(z) - \mathrm{Dg}(\mathbf{0})\| < 2\pi\Big(\frac{\varepsilon}{2\pi}\Big) = \varepsilon \, , \end{aligned}$$

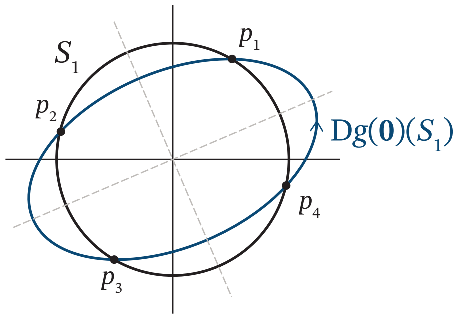
establishing [4.4]. If we further assume that the fixed point is a hyperbolic saddle, the routine calculations above yield a description of the curves $g(S_r)$ for small values of the parameter r, as summarised by the following Lemma.

4.2 Lemma. Let $g \in \mathrm{Diff}^1_+(\mathbb{R}^2)$ be a planar diffeomorphism fixing the origin such that $\mathrm{Dg}(\mathbf{0})$ is a hyperbolic saddle. Then, for any sufficiently small $r > 0$, $S_r \cap g(S_r)$ consists of exactly four points.

Proof. Given that $\mathrm{Dg}(\mathbf{0})$ is a hyperbolic saddle matrix, $\mathrm{Dg}(\mathbf{0})(S_1)$ is an ellipse having semi-major axis of length strictly smaller than one and semi-minor axis of length strictly greater than one. Thus,

$|S_1 \cap \mathrm{Dg}(\mathbf{0})(S_1)| = 4$, and all four intersections are non-tangential or, equivalently, transversal, as pictured in Figure 4.2.

Figure 4.2 – An ellipse centred at the origin for which none of the axis have unit length meets the unit circle transversely in four points.



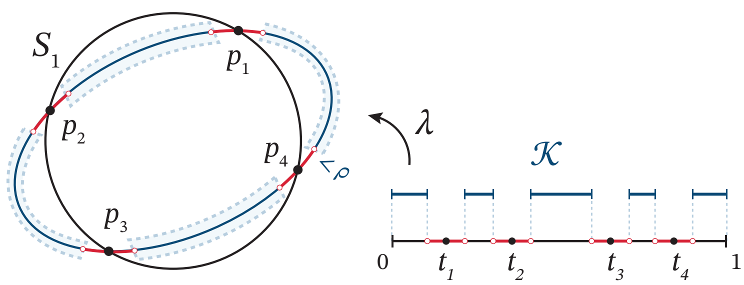
Under the usual parameterisation γ_1 of S_1, each such intersection point p_j, $1 \leq j \leq 4$, is correspondent to a transversal intersection of the C^∞ path $\lambda \overset{\text{def}}{=} \mathrm{Dg}(\mathbf{0}) \circ \gamma_1$ with the embedded curve $\mathrm{C} = S_1$ at $p_j = \lambda(t_j)$, for some $t_1, \dots, t_4 \in \mathbb{I}$. Thus, we may use Lemma 4.1 finitely many times to obtain $\varepsilon > 0$ and $\delta > 0$ such that, for every path β which is C^1 ε-close to λ, we have $\left|\beta\big((t_j - \delta, t_j + \delta)\big) \cap S_1\right| = 1$.

Let $\mathcal{K} \overset{\text{def}}{=} \mathbb{I} \setminus \left(\bigcup_{j=1}^{4} (t_j - \delta_j, t_j + \delta_j)\right)$. The set $\mathcal{K}$ is compact, and so is its image $\lambda(\mathcal{K})$. Also, since $\lambda(\mathcal{K}) \cap S_1 = \varnothing$, the distance between these two nonempty compact sets is positive, say $\rho > 0$.

In other words, $z \in \lambda(\mathcal{K})$ and $|w - z| < \rho$ simultaneously imply $w \notin S_1$. Let $\varepsilon_0 \overset{\text{def}}{=} \min\{\varepsilon, \rho\} > 0$. By [4.4] there exists r_0 such that $(g \circ \gamma_r) \,/\, r$ is C^1 ε_0-close to $\mathrm{Dg}(\mathbf{0}) \circ \gamma_1 = \lambda$ for every $0 < r < r_0$. Therefore:

- since $\varepsilon_0 \leq \varepsilon$, each path $g \circ \gamma_r$ has a single intersection point with S_1 at each one of the intervals $\big(t_j - \delta, t_j + \delta\big)$, for $1 \leq j \leq 4$;
- since $\varepsilon_0 \leq \rho$, the path $(g \circ \gamma_r) \,/\, r$ does not intercept S_1 for instants in $\mathcal{K}$, since its image is at distance at most ε_0 from the image of λ.

Figure 4.3 – Disjoint compact segments can be isolated from the circle by tubular neighbourhoods, while the intersection number remains stable around the intersection points.



These two facts are pictorially conveyed in Figure 4.3. Putting them together, we see that:

$$\left|\left(\frac{g \circ \gamma_r}{r}\right)(\mathbb{I}) \cap S_1\right| = 4 \text{ or, equivalently, } \left|(g \circ \gamma_r)(\mathbb{I}) \cap rS_1\right| = 4\,,$$

since $z \mapsto r_0 z$ is a bijection of the plane. But $rS_1 = S_r$ and $(g \circ \gamma_r)(\mathbb{I}) = g\big(\gamma_r(\mathbb{I})\big) = g(S_{r_0})$, so the desired result follows. ▨

When the previous Lemma 4.2 is combined with the Extension Lemma 2.10, it yields a first glance at the action of G_3 on Γ, as described below.

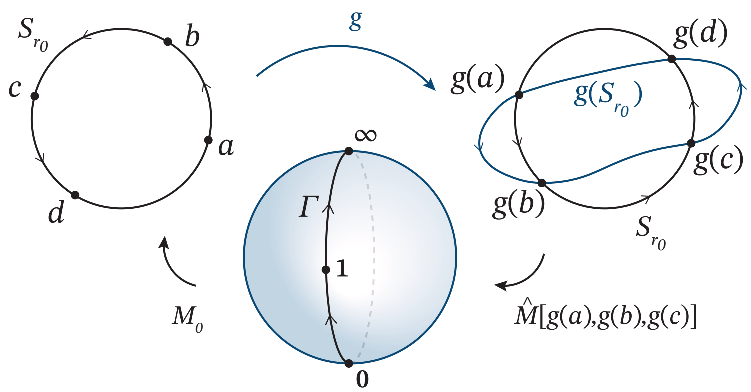
4.3 The 4-point Lemma. Let $G \subset \mathrm{Diff}^1_+(\mathbb{S}^2)$ be a proper extension of $\mathrm{Möb}(\mathbb{S}^2)$. Then, there exists an $\mathcal{I}G_3$ isotopy $(k_t)_{t\in\mathbb{I}}$ such that $|k_1(\Gamma) \cap \Gamma| = 4$.

Proof. By the Extension Lemma 2.10, it is known that there exists an $\mathcal{I}G_2$-isotopy $(g_t)_{t\in\mathbb{I}}$ such that $g \overset{\text{def}}{=} g_1$ has a hyperbolic saddle fixed point at $\mathbf{0}$. Moving to the plane $\mathbb{R}^2 \simeq \mathbb{S}^2 \setminus \{\infty\}$, we may fix a sufficiently small $0 < r_0 < 1$ such that $|S_{r_0} \cap g(S_{r_0})| = 4$, by Lemma 4.2.

Let $a, b, c, d \in S_{r_0}$ be four consecutive points in the usual anticlockwise cyclic order, and such that $S_{r_0} \cap g(S_{r_0}) = \{g(a), g(b), g(c), g(d)\}$. Then, we may consider the unique $M_0 \in \mathrm{Möb}(\mathbb{S}^2)$ mapping the ordered triple $(\mathbf{0}, \mathbf{1}, \infty)$ onto (a, b, c). Notice that $M_0 = \hat{M}[a, b, c]^{-1}$ and that $M_0(\Gamma) = S_{r_0}$, as pictured in Figure 4.4. Let

$$k_t \overset{\text{def}}{=} \hat{M}[g_t(a), g_t(b), g_t(c)] \circ g_t \circ M_0 \ , \ 0 \leq t \leq 1 \, .$$

Figure 4.4 – Given that the Möbius group is sharply 3-transitive and preserves both circles and orientation, M_0 takes the meridian Γ onto the circle S_{r_0}, in an order-preserving way.



By the continuity described in [1.11], we obtain an $\mathcal{I}G_3$-isotopy $(k_t)_{t\in\mathbb{I}}$. Since M_0 is a bijection of the sphere:

$$\begin{aligned} |k_1(\Gamma) \cap \Gamma| &= |M_0(k_1(\Gamma) \cap \Gamma)| \\ &= |M_0(k_1(\Gamma)) \cap M_0(\Gamma)| \\ &= |M_0 \circ \hat{M}[g(a), g(b), g(c)] \circ g \circ M_0(\Gamma) \cap M_0(\Gamma)| \\ &= |(M_0 \circ \hat{M}[g(a), g(b), g(c)]) \circ g(M_0(\Gamma)) \cap M_0(\Gamma)| \\ &= |g(S_{r_0}) \cap S_{r_0}| = 4 \, . \end{aligned}$$

▨

Even more important than the 4-point Lemma 4.3 is its Corollary 4.5, which we shall present now. Before doing so, however, we introduce some useful notations to easily describe the geometry of the meridian Γ in a familiar fashion.

4.4 Definition. Consider finite points a, b on the meridian Γ to be identified with their real counterparts on the x-axis, and the natural ordering induced accordingly. Whenever $a \leq b$ with respect to this order,

- $[a\!:\!b]$ denotes the arc of Γ with endpoints a, b and not containing ∞, which corresponds under stereographic projection Ψ_N to the compact segment of the x-axis with the associated endpoints;
- $[b\!:\!a]$ denotes the arc of Γ with endpoints a, b and containing ∞, which projects under Ψ_N onto $(-\infty, a] \cup [b, +\infty)$.

If $b = \infty$, then the corresponding arcs are defined via stereographic projection as

$$[\infty\!:\!a] = \Psi_N^{-1}((-\infty, a]) \cup \{\infty\} \quad \text{and} \quad [a\!:\!\infty] = \Psi_N^{-1}([a, +\infty)) \cup \{\infty\}\,.$$

Lastly, open and half-open arcs of Γ are defined accordingly by deletion of the suitable endpoints from the corresponding closed arcs.

4.5 Corollary. Let $G \subset \mathrm{Diff}^1_+(\mathbb{S}^2)$ be a proper extension of $\mathrm{Möb}(\mathbb{S}^2)$. Then, for a given $z_0 \in \Gamma \setminus \{\mathbf{0}, \mathbf{1}, \infty\}$ there exists an $\mathcal{I}G_3$-isotopy $(h_t)_{t\in\mathbb{I}}$ (depending on z_0) such that $h_1(z_0) \notin \Gamma$.

Proof. In the language of the previous 4-point Lemma 4.3, let $k_1(\Gamma) \cap \Gamma = \{\mathbf{0}, \mathbf{1}, \infty, w_0\}$. If $k_1(z_0) \neq w_0$, it suffices to consider $h_t = k_t$ for every $0 \leq t \leq 1$. Otherwise, assume for concreteness that $z_0 \in\ ]\mathbf{0}\!:\!\mathbf{1}[$ — analogous reasonings apply to the other cases. Let $M \in \mathrm{Möb}(\mathbb{S}^2)$ be such that $M(\mathbf{0}) = \mathbf{1}$, $M(\mathbf{1}) = \infty$ and $M(\infty) = \mathbf{0}$. Upon defining $h_t \stackrel{\text{def}}{=} M^{-1} \circ k_t \circ M$, where k_t is as given in the 4-point Lemma 4.3 for $0 \leq t \leq 1$, we obtain a new $\mathcal{I}G_3$ isotopy $(h_t)_{t\in\mathbb{I}}$.

Notice that M leaves Γ invariant, and $M\restriction_\Gamma$ acts as an interval exchange transformation free of fixed points. In particular, $\tilde{z}_0 \stackrel{\text{def}}{=} M(z_0) \neq z_0$, and thus $k_1(\tilde{z}_0) \notin \Gamma$. Consequently, $h_1(z_0) = M^{-1}(k_1(\tilde{z}_0)) \notin \Gamma$ as well. ▨

We are, now, ready to prove our Crossing Lemma, a key step towards establishing connectivity within the subgroup G_3. Once it is done, we can proceed to the proofs of this essay's main theorems.

4.6 The Crossing Lemma. Let $G \subset \mathrm{Diff}^1_+(\mathbb{S}^2)$ be a proper extension of $\mathrm{Möb}(\mathbb{S}^2)$. 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 intercepts Γ on the arc $]\mathbf{0}\!:\!\mathbf{1}[$,

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

Proof. By Corollary 4.5, we may fix an $\mathcal{I}G_3$-isotopy $(h_t)_{t\in\mathbb{I}}$ and a point $z_0 \in \Gamma$ such that $h_1(z_0) \notin \Gamma$. Suppose first that $h_1(z_0) \in \mathcal{H}^+$, and let D be a disk centred at $h_1(z_0)$, contained within $\mathcal{H}^+$. Then, $h_1^{-1}(D)$ is an open neighbourhood of $z_0 \in \Gamma = \partial\mathcal{H}^-$ and thus contains a point $u_0 \in h_1^{-1}(D) \cap \mathcal{H}^-$.

Consider the continuous path $\gamma : \mathbb{I} \to \mathbb{S}^2$ given by $\gamma(t) \stackrel{\text{def}}{=} h_t(u_0)$, that describes the trajectory of u_0 under the isotopy h. This path has the following properties: $\gamma(0) = u_0 \in \mathcal{H}^-$ and $\gamma(1) = h_1(u_0) \in \mathcal{H}^+$. It therefore intercepts the common boundary $\partial\mathcal{H}^- = \partial\mathcal{H}^+ = \Gamma$ at least once. In particular, the three sets $\gamma^{-1}(\mathcal{H}^-)$, $\gamma^{-1}(\Gamma)$ and $\gamma^{-1}(\mathcal{H}^+)$ are all nonempty and form a partition of $\mathbb{I}$. Let

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

4.6.1 Claim. $0 < t^- \leq t^+ < 1$, and $[t^-, t^+] \subset \gamma^{-1}(\Gamma)$ or, in other words, $\gamma\big([t^-, t^+]\big) \subset \Gamma$.

Proof. Since $\gamma^{-1}(\mathcal{H}^-)$ is open in $\mathbb{I}$ and $0 \in \gamma^{-1}(\mathcal{H}^-)$, certainly $t^- = \sup\{\gamma^{-1}(\mathcal{H}^-)\} > 0$. On the other hand, since $1 \in \gamma^{-1}(\mathcal{H}^+)$, which is an open set in $\mathbb{I}$ disjoint from $\gamma^{-1}(\mathcal{H}^-)$, we must have $t^- < 1$.

In particular, $\gamma^{-1}(\mathcal{H}^+) \cap [t^-, 1]$ is nonempty and bounded from below by t^-, thus admitting a certain infimum t^+, which must be $\geq t^-$ (by definition) and strictly smaller than 1 (by the openness of $\gamma^{-1}(\mathcal{H}^+)$).
Lastly, we notice that:

- $t > t^-$ implies $t \notin \gamma^{-1}(\mathcal{H}^-)$, which in turn implies $\gamma(t) \in \overline{\mathcal{H}^+}$;
- $t^- < t < t^+$ implies $t \notin \gamma^{-1}(\mathcal{H}^+)$, which in turn implies $\gamma(t) \in \overline{\mathcal{H}^-}$.

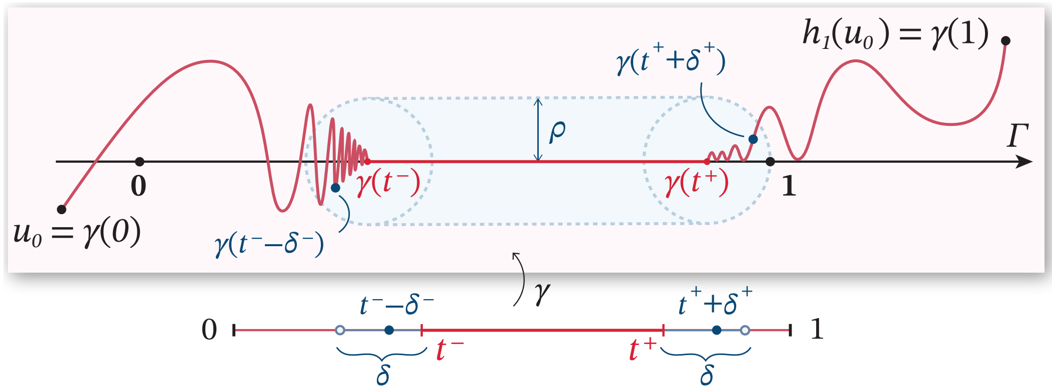
In other words, $\gamma(t) \in \overline{\mathcal{H}^+} \cap \overline{\mathcal{H}^-} = \Gamma$ for every $t \in (t^-, t^+)$, meaning that $(t^-, t^+) \subset \gamma^{-1}(\Gamma)$. But $\gamma^{-1}(\Gamma)$ is a closed subset of $\mathbb{I}$, so $[t^-, t^+] \subset \gamma^{-1}(\Gamma)$ follows. □

Now, since the image of γ cannot intercept $\{\mathbf{0}, \mathbf{1}, \infty\}$, the set $\gamma\big([t^-, t^+]\big)$ must be a compact arc contained within one of the three connected components of $\Gamma \setminus \{\mathbf{0}, \mathbf{1}, \infty\}$. In particular, it is at a positive distance $\rho > 0$ from those three points. Given such ρ, by continuity of γ we may obtain $\delta > 0$ with the following properties:

[4.5]
$$\begin{aligned} s \in \mathbb{I} \text{ and } |s - t^-| < \delta \text{ imply } d\big(\gamma(s), \gamma(t^-)\big) < \rho\,; \\ s \in \mathbb{I} \text{ and } |s - t^+| < \delta \text{ imply } d\big(\gamma(s), \gamma(t^+)\big) < \rho\,. \end{aligned}$$

But then, by the definitions of supremum and infimum we may find points of the form $t^- - \delta^- \in \gamma^{-1}(\mathcal{H}^-)$ and $t^+ + \delta^+ \in \gamma^{-1}(\mathcal{H}^+)$ satisfying $t^- - \delta < t^- - \delta^- < t^- \leq t^+ < t^+ + \delta^+ < t^+ + \delta$, as conveyed in Figure 4.5.

Figure 4.5 – After leaving the eastern hemisphere, the isotopy path of u_0 under h may remain trapped on a compact segment of $\Gamma \setminus \{\mathbf{0}, \mathbf{1}, \infty\}$ before entering the western hemisphere.



Let $\sigma : [-1, 1] \to [t^- - \delta^-, t^+ + \delta^+]$ be any increasing homeomorphism such that $\sigma(0)$ is the midpoint of the interval $[t^-, t^+]$, say piecewise linear for concreteness. Upon defining $\tilde{J} : [-1, 1] \times \mathbb{S}^2 \to \mathbb{S}^2$ as

$$\tilde{J}(t, z) = \tilde{J}_t(z) \overset{\text{def}}{=} h_{\sigma(t)} \circ h^{-1}_{\sigma(0)}(z)\,,$$

we readily see that $(\tilde{J})_{t \in [-1,1]}$ is an $\mathcal{I}G_3$-isotopy. Furthermore, letting $\tilde{z} = \gamma\big(\sigma(0)\big) = h_{\sigma(0)}(u_0) \in \Gamma$, we have $\tilde{J}_1(\tilde{z}) = \gamma(t^+ + \delta^+) \in \mathcal{H}^+$ and $\tilde{J}_{-1}(\tilde{z}) = \gamma(t^- - \delta^-) \in \mathcal{H}^-$.
Lastly, observing the range of σ we may describe the trajectory of $\tilde{z}$ under $\tilde{J}$ as:

$$\begin{aligned} \gamma_{\tilde{J}}(\tilde{z}) = \{\tilde{J}_t(\tilde{z}) : -1 \leq t \leq 1\} &= \{h_{\sigma(t)}(u_0) : -1 \leq t \leq 1\} \\ &= \{h_s(u_0) : t^- - \delta^- \leq s \leq t^+ + \delta^+\} \\ &= \{\gamma(s) : t^- - \delta^- \leq s \leq t^+ + \delta^+\} = \gamma\big([t^- - \delta^-, t^+ + \delta^+]\big)\,. \end{aligned}$$

In particular, the choice of ρ along with [4.5] implies that any point in $\gamma_{\tilde{J}}(\tilde{z}) \cap \Gamma$ must lie in the same connected component of $\Gamma \setminus \{\mathbf{0}, \mathbf{1}, \infty\}$ as the one containing the segment $\gamma([t^-, t^+])$, which is, in turn, precisely the one containing $\tilde{z}$.

If this component happens to be $]\mathbf{0} : \mathbf{1}[$, as already pictured in Figure 4.5, we may let $J = \tilde{J}$ and $\hat{z} = \tilde{z}$ to accomplish the Lemma's statement. Otherwise, consider Γ endowed with its cyclic order induced by the real line. If $\tilde{z} \in]\mathbf{a} : \mathbf{b}[$ and $\{\mathbf{c}\} = \{\mathbf{0}, \mathbf{1}, \infty\} \setminus \{\mathbf{a}, \mathbf{b}\}$,let $\hat{M}[\mathbf{a}, \mathbf{b}, \mathbf{c}]$ be the unique Möbius transformation performing the associations described in [1.10], where now $\mathbf{a}, \mathbf{b}, \mathbf{c}$ were allowed to be $\mathbf{0}, \mathbf{1}$ or ∞ as well. Then, since $\hat{M}[\mathbf{a}, \mathbf{b}, \mathbf{c}]$ applies $]\mathbf{a} : \mathbf{b}[$ onto $]\mathbf{0} : \mathbf{1}[$ in an orientation-preserving way, we may let

$$\hat{z} \overset{\text{def}}{=} \hat{M}[\mathbf{a}, \mathbf{b}, \mathbf{c}](\tilde{z}) \qquad \text{and} \qquad J \overset{\text{def}}{=} \hat{M}[\mathbf{a}, \mathbf{b}, \mathbf{c}] \circ \tilde{J}_t \circ \hat{M}[\mathbf{a}, \mathbf{b}, \mathbf{c}]^{-1}$$

to obtain the sought point and isotopy. Lastly, if we had $h_1(z_0) \in \mathcal{H}^-$ instead, the same reasoning would apply, with the roles of $\gamma^{-1}(\mathcal{H}^-)$ and $\gamma^{-1}(\mathcal{H}^+)$ reversed on the construction of $\tilde{J}$. ▨

4.2 The proof of Theorem A

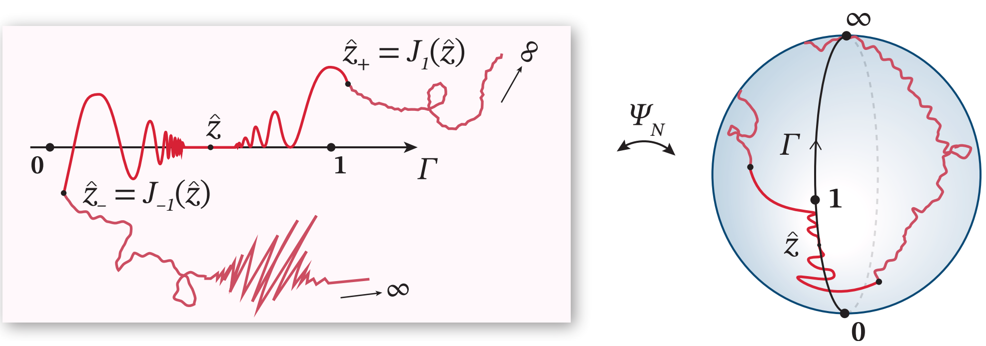
From now on, a proper extension $G \subset \mathrm{Diff}^1_+(\mathbb{S}^2)$ of $\mathrm{Möb}(\mathbb{S}^2)$ is fixed throughout, and we let the point $\hat{z} \in \Gamma$ and the $\mathcal{J}G_3$-isotopy $(J_t)_{t\in[-1,1]}$ be as in the Crossing Lemma 4.6. Upon denoting $\hat{z}_- \overset{\text{def}}{=} J_{-1}(\hat{z})$ and $\hat{z}_+ \overset{\text{def}}{=} J_1(\hat{z})$, we define the set

$$\chi \overset{\text{def}}{=} \overline{\gamma_{I^{\hat{z}_-}}(\hat{z}_-) \cup \gamma_J(\hat{z}) \cup \gamma_{I^{\hat{z}_+}}(\hat{z}_+)}\,, \tag{4.6}$$

where $I^{\hat{z}_-}$ and $I^{\hat{z}_+}$ are the isotopies yielded by the Fundamental Lemma 3.10 when considering the points $\hat{z}_- \in \mathcal{H}^-$ and $\hat{z}_+ \in \mathcal{H}^+$.

On the sphere, χ is a continuum. Indeed, $\gamma_{I^{\hat{z}_-}}(\hat{z}_-) \cup \gamma_J(\hat{z}) \cup \gamma_{I^{\hat{z}_+}}(\hat{z}_+)$ is connected, as the union of (connected) curves with points in common, while its closure is automatically compact on the compact space $\mathbb{S}^2$, and consists of adjoining $\{\infty\}$ to this union, as suggested in Figure 4.6.

Figure 4.6 – The continuum χ is constructed by gluing together isotopy trajectories with points in common, one of which is bounded away from ∞ and two of which are known to accumulate at ∞, and then taking their closure.



We will now derive an important separating property for χ. The argument relies on the folklore fact that, given a path $\alpha : \mathbb{I} \to \mathbb{R}^n$ connecting two distinct points, one may obtain another path $\beta : \mathbb{I} \to \mathbb{R}^n$ which is simple, connects the same pair of points and such that $\beta(\mathbb{I}) \subseteq \alpha(\mathbb{I})$[1].

[1]This is stated as Lemma 3.1 in the book (10), on fractals. A patch up of the proof presented therein is provided by Prof. Lee Mosher as an answer to `Question 857066` at the `Mathematics Stack Exchange` portal.

4.7 Lemma. The set χ defined in [4.6] is a continuum separating the open arcs $]\infty:\mathbf{0}[$ and $]\mathbf{1}:\infty[$ in the following sense: whenever $\alpha : \mathbb{I} \to \mathbb{S}^2$ is a path such that $\alpha(0) \in]\infty:\mathbf{0}[$ and $\alpha(1) \in]\mathbf{1}:\infty[$, we have $\alpha(\mathbb{I}) \cap \chi \neq \varnothing$.

Proof. If $\infty \in \alpha(\mathbb{I})$ there is nothing to prove. Otherwise, suppose that α never passes through ∞. Since we shall look for intersections, the argument in the previous paragraph allows us to assume that α is simple. By the hypothesis made upon its endpoints, the following are well-defined:

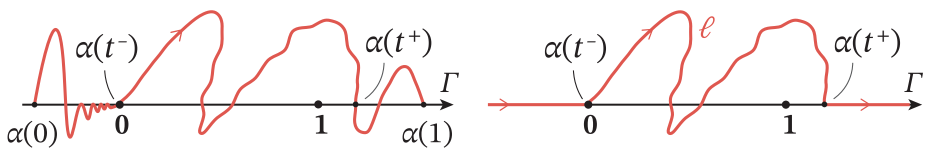
[4.7]
$$t^- \overset{\text{def}}{=} \sup\, \alpha^{-1}\big(]\infty:\mathbf{0}]\big) = \sup\big\{t \in \mathbb{I} : \alpha(t) \in]\infty:\mathbf{0}]\big\} \text{ and}$$
$$t^+ \overset{\text{def}}{=} \inf\big\{t \in (t^-, 1] : \alpha(t) \in [\mathbf{1}:\infty[\big\}\,.$$

Recalling that the supremum and infimum of a set are limit points of the set, continuity of α implies $\alpha(t^-) \in]\infty:\mathbf{0}]$ and $\alpha(t^+) \in [\mathbf{1}:\infty[$. In particular, $t^+ > t^-$, for their images belong to disjoint arcs. We further notice that any intersection between $\alpha\big((t^-, t^+)\big)$ and Γ takes place in the open arc $]\mathbf{0}:\mathbf{1}[$, by the very own definition of t^- and t^+.

Equipped with this data, we define on the plane a continuous mapping $\ell : \mathbb{R} \to \mathbb{R}^2$ as pictured in Figure 4.7, and explicitly given by:

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

Figure 4.7 – The segment of α comprehended between the instant it leaves $]\infty:\mathbf{0}]$ and the instant it enters $[\mathbf{1}:\infty[$ can be glued to Γ – traversed the usual way – to generate a line ℓ.



4.7.1 Claim. The mapping ℓ is a line or, in other words, a simple and proper path.

Proof. 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 $\big(\alpha(t^-) - \hat{z}\big)$ and $\big(\alpha(t^+) - \hat{z}\big)$ are nonzero. Thus, the restrictions $\ell\restriction_{t \leq t^-}$ and $\ell\restriction_{t \geq t^+}$ are injective. Furthermore, $\ell\restriction_{(t^-,t^+)} = \alpha\restriction_{(t^-,t^+)}$ 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 $\bar{s} \leq t^-$ and $t^- < \bar{t} < t^+$. Thinking of Γ as the x-axis, as described in Definition 4.4, we have $\alpha(t^-) \leq \mathbf{0} < \hat{z}$, from which follows $\alpha(t^-) - \hat{z}$ negative. But, since $(1-\bar{s})/(1-t^-) \geq 1$, it must be the case that $\ell(\bar{s}) \leq \alpha(t^-) \leq \mathbf{0}$. In particular, $\ell(\bar{s}) \in \Gamma$, so $\ell(\bar{t}) \in \Gamma$ as well. But then:

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

a contradiction. Analogous reasonings show that $\ell(\bar{s}) = \ell(\bar{t})$ cannot happen for $\bar{s} \geq t^+$ and $t^- < \bar{t} < t^+$ both holding simultaneously either.

Lastly, $\ell\big((-\infty, t^-]\big) \subset\ ]\infty\!:\!\mathbf{0}]$ and $\ell\big([t^+, +\infty)\big) \subset [\mathbf{1}\!:\!\infty[$, which are disjoint subsets of the plane. It must therefore be the case that ℓ is simple.
To see that ℓ is proper, notice first that

$$\lim_{t\to-\infty} \frac{1-t}{1-t^-} = \lim_{t\to+\infty} \frac{1+t}{1+t^+} = +\infty\,,$$

so that, for $t \geq t^+$,

$$|\ell(t)| \geq \underbrace{\frac{1+t}{1+t^+}}_{\to+\infty} \underbrace{|\alpha(t^+) - \hat{z}|}_{\text{nonzero, fixed}} - |\hat{z}| \to +\infty \text{ as } t \to +\infty$$

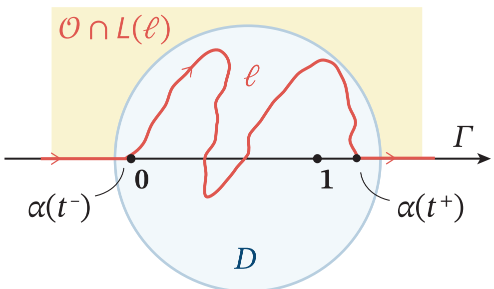
and, similarly, $|\ell(t)| \to +\infty$ as $t \to -\infty$. □

As it is usually done for lines, the mapping ℓ and its trace will be confounded without notice, and endowed with the natural orientation inherited from $\mathbb{R}$, which in this case is compatible with the intrinsic orientation of Γ. Due to the usually misquoted and often misunderstood Jordan-Schoenflies Theorem[2], this automatically divides $\mathbb{S}^2$ — and thus the plane — into two open and connected components, the right $R(\ell)$ and the left $L(\ell)$ of ℓ, plus their common boundary ℓ.

Consider the compact set $\alpha\big([t^-, t^+]\big)$, and fix some closed disk $D \subset \mathbb{R}^2$ fully containing it. Then, it must be the case that $[\mathbf{0}\!:\!\mathbf{1}] \subset D$. Indeed, since D is convex it must contain the segment $[\alpha(t^-), \alpha(t^+)]$. But from **[4.7]** we know that $\alpha(t^-) \leq \mathbf{0}$ and $\alpha(t^+) \geq \mathbf{1}$, which imply $[\mathbf{0}\!:\!\mathbf{1}] \subset [\alpha(t^-), \alpha(t^+)]$.

Thus, 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. Consequently, $L(\ell) \cap \mathcal{O} = L(\Gamma) \cap \mathcal{O} = \mathcal{H}^+ \cap \mathcal{O}$ and $R(\ell) \cap \mathcal{O} = R(\Gamma) \cap \mathcal{O} = \mathcal{H}^- \cap \mathcal{O}$, as suggested by Figure 4.8.

Figure 4.8 – Sufficiently close to ∞, the line ℓ follows the meridian Γ, and its left is fully contained within the upper half-plane (western hemisphere).



From the Fundamental Lemma 3.10, $\gamma_{I\hat{z}_+}(\hat{z}_+)$ is fully contained within $\mathcal{H}^+$ and accumulates at $\{\infty\}$. Since $\mathcal{O}$ defines a neighbourhood of ∞ in the sphere, $\gamma_{I\hat{z}_+}(\hat{z}_+) \cap L(\ell) \neq \varnothing$ follows. Analogously, $\gamma_{I\hat{z}_-}(\hat{z}_-) \cap R(\ell) \neq \varnothing$. This translates to $\chi \cap L(\ell) \neq \varnothing$ and $\chi \cap R(\ell) \neq \varnothing$. Since χ is connected, these imply $\chi \cap \partial L(\ell) = \chi \cap \partial L(\ell) \neq \varnothing$. In other words, ℓ intercepts the continuum χ.

Let $\bar{t} \in \mathbb{R}$ be such that $\ell(\bar{t}) \in \chi$. Then, it must be the case that $\bar{t} \in (t^-, t^+)$. Indeed, on the one hand, $\ell(\bar{t}) \in [\mathbf{1}\!:\!\infty[$ if $\bar{t} \geq t^+$ and $\ell(\bar{t}) \in\ ]\infty\!:\!\mathbf{0}]$ if $\bar{t} \leq t^-$. On the other hand, any intersection between χ and Γ must take place on the open segment $]\mathbf{0}\!:\!\mathbf{1}[$, by the Fundamental Lemma 3.10 and the Crossing Lemma 4.6. But $t^- < \bar{t} < t^+$ means that $\ell(\bar{t}) = \alpha(\bar{t})$, yielding an intersection between $\alpha(\mathbb{I})$ and χ, as claimed. ▨

[2]As stated in (45, 44), see also (46)

Consider the accessibility relation $\sim_{G_3}$ as described in Definition 1.6. Clearly, $\mathcal{A}_{G_3}(\mathbf{0}) = \{\mathbf{0}\}$, $\mathcal{A}_{G_3}(\mathbf{1}) = \{\mathbf{1}\}$ and $\mathcal{A}_{G_3}(\infty) = \{\infty\}$. Our goal is now to show that $z_0 \in \mathcal{A}_{G_3}(\hat{z})$ for any given $z_0 \in \mathbb{S}^2 \setminus \{\mathbf{0}, \mathbf{1}, \infty\}$. Notice that by Corollary 4.5 it suffices to consider only the case $z_0 \notin \Gamma$, as done next.

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

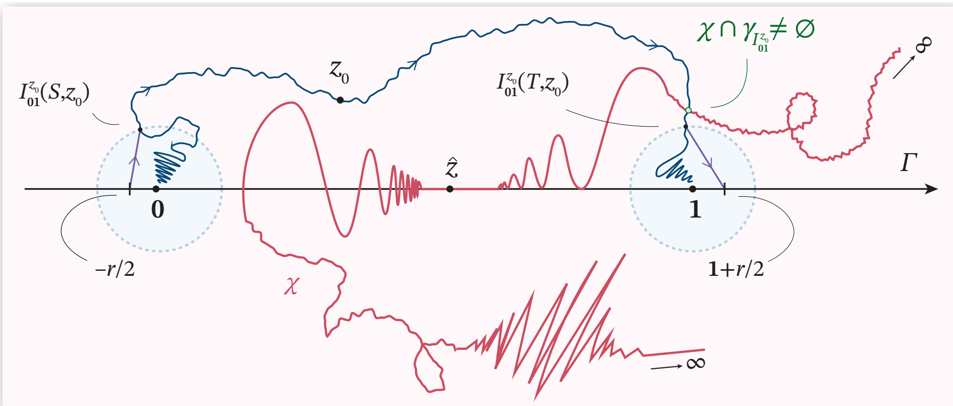
Proof. Let χ be as in [4.6]. Given that $\mathbf{0}, \mathbf{1} \notin \chi$, we may fix $r > 0$ such that $\overline{\mathbb{D}}_r(\mathbf{0}) \cap \chi$, $\overline{\mathbb{D}}_r(\mathbf{1}) \cap \chi$ and $\overline{\mathbb{D}}_r(\mathbf{0}) \cap \overline{\mathbb{D}}_r(\mathbf{1})$ are all empty, and also such that both closed disks are simultaneously disjoint from $\{z_0\}$ and $\{\infty\}$.

Given z_0, consider the $\mathcal{J}G_3$-isotopy $I^{z_0}_{\mathbf{01}}$ yielded by Corollary 3.11. Then, the continuous path $t \mapsto I^{z_0}_{\mathbf{01}}(t, z_0)$ accumulates at $\{\mathbf{0}\}$ for arbitrarily negative times, and at $\{\mathbf{1}\}$ for arbitrarily positive times. Thus, using supremum and infimum arguments analogous to those previously encountered in the proofs of Lemmas 4.6 and 4.7, we encounter $S < 0$ *maximal* such that $I^{z_0}_{\mathbf{01}}(S, z_0) \in \partial\overline{\mathbb{D}}_r(\mathbf{0})$ and $T > 0$ *minimal* such that $I^{z_0}_{\mathbf{01}}(T, z_0) \in \partial\overline{\mathbb{D}}_r(\mathbf{1})$. We define a new continuous path $\alpha : \mathbb{I} \to \mathbb{R}^2$ as:

$$\alpha(t) \overset{\text{def}}{=} \begin{cases} 4t\, I^{z_0}_{\mathbf{01}}(S, z_0) - (1-4t)\,\dfrac{r}{2} & \text{if } 0 \le t \le \dfrac{1}{4}\,, \\ I^{z_0}_{\mathbf{01}}\Big(2t(T-S) + \dfrac{3S-T}{2},\, z_0\Big) & \text{if } \dfrac{1}{4} \le t \le \dfrac{3}{4}\,, \\ (4-4t)\, I^{z_0}_{\mathbf{01}}(T, z_0) + (4t-3)\Big(\mathbf{1} + \dfrac{r}{2}\Big) & \text{if } \dfrac{3}{4} \le t \le 1\,. \end{cases} \tag{4.8}$$

Geometrically, α departs from the point $\alpha(0) \in\,]\infty\,{:}\,\mathbf{0}[\, \cap \overline{\mathbb{D}}_r(\mathbf{0})$ 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 α follows z_0's isotopy trajectory until it first reaches the boundary of the disk $\overline{\mathbb{D}}_r(\mathbf{1})$. Then, α moves on a straight line until it reaches the point $\alpha(1) \in\,]\mathbf{1}\,{:}\,\infty[\, \cap \overline{\mathbb{D}}_r(\mathbf{1})$. This process is conveyed in Figure 4.9.

Figure 4.9 – After leaving a compact neighbourhood of $\mathbf{0}$ disjoint of χ and before entering a neighbourhood of $\mathbf{1}$ disjoint from χ, the path of z_0 under $I^{z_0}_{\mathbf{01}}$ must cross the continuum χ.



In particular, $\alpha(0) \in\,]\infty\,{:}\,\mathbf{0}[$, $\alpha(1) \in\,]\mathbf{1}\,{:}\,\infty[$ and $\alpha(\mathbb{I}) \subset \mathbb{S}^2 \setminus \{\infty\}$. Thus, by Lemma 4.7, α must intercept $\chi \setminus \{\infty\}$. But, since the segments $\alpha\big([0, 1/4]\big)$ and $\alpha\big([3/4, 1]\big)$ are contained within disks disjoint from χ, we must have $\alpha\big((1/4, 3/4)\big) \cap \big(\chi \setminus \{\infty\}\big) \neq \varnothing$.

However, it is seen from [4.8] that $\alpha\big((1/4, 3/4)\big) \subset \gamma_{I_{\mathbf{01}}^{z_0}}(z_0)$. By Lemma 1.8, this is readily seen to imply $z_0 \in \mathcal{A}_{G_3}(\hat{z})$.

This is enough to derive the (arc) 4-transitivity of G_0 for, if (p_0, p_1, p_2, p_3) and (q_0, q_1, q_2, q_3) are two given lists of mutually distinct points on the sphere, let $z_0 = \hat{M}[p_1, p_2, p_3](p_0)$ and $w_0 = \hat{M}[q_1, q_2, q_3]^{-1}(q_0)$. Then, neither z_0 nor w_0 belong to $\{\mathbf{0}, \mathbf{1}, \infty\}$ and thus, by Lemma 4.8 above, both z_0 and w_0 belong to $\mathcal{A}_{G_3}(\hat{z})$. This implies that there is some $\mathcal{J}G_3$-isotopy $(f_t)_{t\in\mathbb{I}}$ such that $f_1(z_0) = w_0$. Since Möb$(\mathbb{S}^2)$ is a path connected group, in particular $\hat{M}[q_1, q_2, q_3]^{-1} \circ f_1 \circ \hat{M}[p_1, p_2, p_3]$ lies in G_0 and maps (p_0, p_1, p_2, p_3) onto (q_0, q_1, q_2, q_3). Therefore, the arc 4-transitivity definition is seen to be satisfied, and Theorem A is proven.

4.3 The proof of Theorem B

Before proceeding to the actual proof, we introduce a handful of auxiliary terminology in order to make some pictorial arguments slightly more precise.

4.9 Definition. Let P be a fixed subset of a manifold $\mathcal{M}$. Given two paths $\alpha, \beta : \mathbb{I} \to \mathcal{M}$, we shall use the notation $\alpha \simeq \beta \operatorname{rel} P$ to indicate that α, β are homotopic – with fixed endpoints $\alpha(0) = \beta(0)$ and $\alpha(1) = \beta(1)$ – in the set $\mathcal{M} \setminus P$.

In other words, $\alpha \simeq \beta \operatorname{rel} P$ whenever α and β are two paths joining the same pair of points, and can be continuously deformed one onto another in such a way that none of the intermediate paths meet the distinguished set P, while the endpoints are kept fixed throughout. This, just like ordinary homotopy with fixed endpoints, is an equivalence relation.

In particular, we may consider loops based at a certain point and the associated (relative) fundamental group equipped with its usual operations, which are well-defined in homotopy classes:

- $\beta * \alpha$ denotes *concatenation*: first α and then β are traversed, each twice as fast as their original parameterisations;
- $\bar{\alpha}$ denotes *inversion*: the image of α is traversed in the opposite direction, via the reparameterisation $t \mapsto 1 - t$.

These may be used to define the prototype of what we shall call a topological figure 8.

4.10 Definition. Let $\{p_0, \dots, p_3\}$ be four distinguished points of $\mathbb{S}^2$. A loop $\alpha : \mathbb{I} \to \mathbb{S}^2$ based at p_0 shall be named a *topological figure 8* (relative to the $\{p_i\}$) if

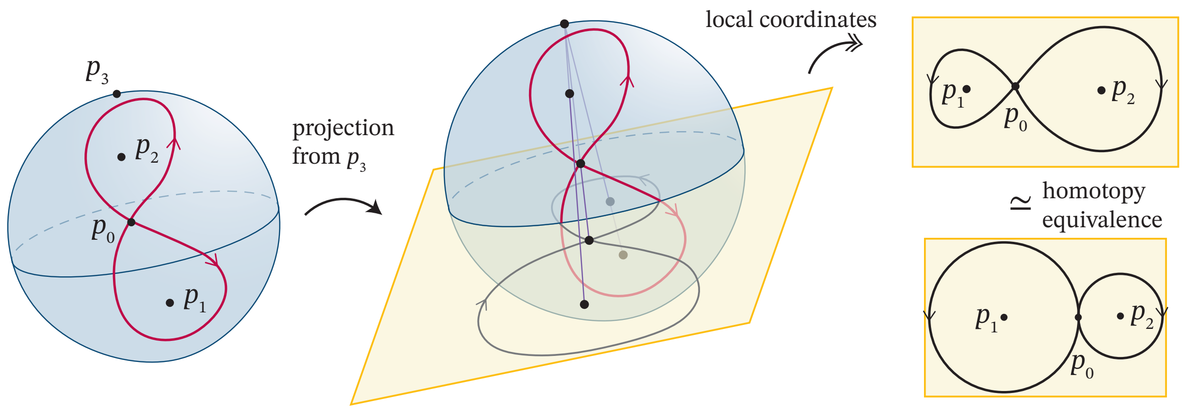
$$\alpha \simeq \zeta_2 * \bar{\zeta}_1 \operatorname{rel}\{p_1, p_2, p_3\},$$

where each $\zeta_i \in \pi_1\big(\mathbb{S}^2 \setminus \{p_1, p_2, p_3\}, p_0\big)$ is a Jordan curve that leaves p_i on its left and separates it from the remaining two points,

The importance of a topological figure 8 lies in the folklore fact that, if such an object is realised as the trajectory of p_0 under an $\mathcal{J}$-isotopy fixing the remaining distinguished points, then the terminal homeomorphism must possess positive topological entropy. This can be argued to be a consequence of the classification theory due to Nielsen and Thurston. Such connection is outlined in Appendix B, where the relevant concepts are introduced and developed.

Notice that whenever we consider p_3 as the point at infinity and stereographically project from it, a topological figure 8 translates into a curve (relatively) homotopic to the prototypical *planar figure 8*, consisting of the wedge of two circles based at (the image of) p_0, each of them traversed once with contrary orientations, whilst leaving (the images of) p_1 and p_2 in opposite components of their complements. This setting is pictured in Figure 4.10, and is the configuration that we shall be aiming at in the constructions to follow.

Figure 4.10 – A topological figure 8 on the surface of the sphere and its planar projection. Notice that stereographic projection may reverse the orientation with which loops are traversed to an external observer.



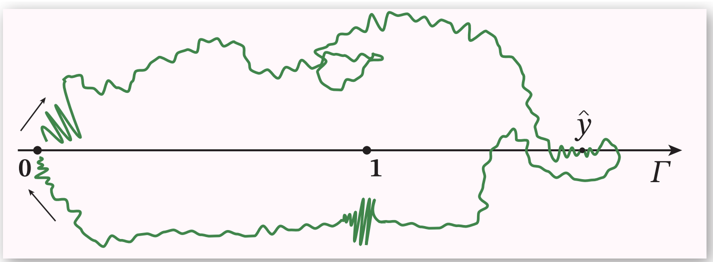
As in Section 4.2, fix a proper group extension $G \subset \mathrm{Diff}^1_+(\mathbb{S}^2)$ of $\mathrm{Möb}(\mathbb{S}^2)$ and consider the continuum χ as defined in [4.6]. We start by observing that $\chi \setminus \{\infty\}$ can be realised as the trajectory of the point $\hat{z}$ under a certain $\mathcal{I}G_3$-isotopy $(K_t)_{t\in\mathbb{R}}$, which is given explicitly as:

[4.9]
$$K_t \overset{\text{def}}{=} \begin{cases} I^{\hat{z}_-}_{-1-t} \circ J_{-1} & \text{if } t \le -1\,, \\ J_t & \text{if } -1 \le t \le 1\,, \\ I^{\hat{z}_+}_{-1+t} \circ J_1 & \text{if } t \ge 1\,, \end{cases}$$

where the isotopy J and the family of points $\hat{z}_\pm = J_{\pm 1}(\hat{z})$ are described in details in the Crossing Lemma 4.6, while the isotopies I are as in the Fundamental Lemma.

We can also use such descriptions, along with the property of preserving circles possessed by Möbius transformations, to derive yet another full-time isotopy having a trajectory which plays the part of a “symmetric” to χ, in the sense described by the following Lemma 4.11 and suggested by Figure 4.11.

Figure 4.11 – The trajectory of $\hat{y}$ under L has properties analogous to those of χ, but relative to the arc $]\mathbf{1}:\infty[$ and the accumulation point $\mathbf{0}$.

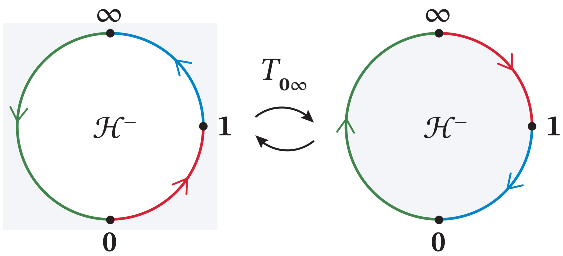


4.11 Lemma. There exist a point $\hat{y}$ in the open segment $]\mathbf{1}:\infty[$ and a full-time $\mathcal{J}G_3$-isotopy $(L_t)_{t\in\mathbb{R}}$ such that:

1) $\{L_t(\hat{y}) : t \geq 1\} \subset \mathcal{H}^-$ and $\{L_t(\hat{y}) : t \leq -1\} \subset \mathcal{H}^+$,
2) the trajectory $\gamma_L(\hat{y})$ only intersects the meridian Γ at points on the open arc $]\mathbf{1}:\infty[$,
3) $\omega_L(\hat{y}) = \alpha_L(\hat{y}) = \{\mathbf{0}\}$.

Proof. Let $(K_t)_{t\in\mathbb{R}}$ be as in [4.9], and consider $L_t \stackrel{\text{def}}{=} T_{\mathbf{0}\infty} \circ K_t \circ T_{\mathbf{0}\infty}$ and $\hat{y} = T_{\mathbf{0}\infty}(\hat{z})$. Notice first that $\hat{y}$ indeed lies in $]\mathbf{1}:\infty[$, since $T_{\mathbf{0}\infty}$ keeps the meridian Γ invariant, acting on it as an interval exchange transformation. Also, it switches $\mathcal{H}^+$ and $\mathcal{H}^-$, as conveyed in Figure 4.12

Figure 4.12 – The action of $T_{\mathbf{0}\infty}$, which on the sphere amounts to a rotation of π around the X-axis.



In particular, when $t \geq 1$ we have $s = -1 + t \geq 0$, and thus $L_t(\hat{y}) = T_{\mathbf{0}\infty}\big(I_s^{\hat{z}+}(\hat{z}_+)\big) \in T_{\mathbf{0}\infty}\big(\mathcal{H}^+\big) = \mathcal{H}^-$, for $\hat{z}_+ \in \mathcal{H}^+$ by the Crossing Lemma 4.6 and the trajectory $\{I_s^{\hat{z}+} : s \geq 0\}$ remains in $\mathcal{H}^+$ by the Fundamental Lemma 3.10. Analogously, $L_t(\hat{y}) \in \mathcal{H}^+$ for every $t \leq -1$, establishing **1)**.

Due to **1)**, if $L_t(\hat{y}) \in \Gamma$, then necessarily $|t| < 1$. Under this circumstance, $L_t(\hat{y}) \in \Gamma$ or, explicitly, $T_{\mathbf{0}\infty} \circ J_t \circ T_{\mathbf{0}\infty}(\hat{y}) \in \Gamma$. But this is equivalent to $J_t(\hat{z}) \in T_{\mathbf{0}\infty}\big(\Gamma\big) = \Gamma$. By the first item in the Crossing Lemma 4.6, this implies $J_t(\hat{z}) \in]\mathbf{0}:\mathbf{1}[$ and thus – equivalently – $L_t(\hat{y}) \in T_{\mathbf{0}\infty}\big(]\mathbf{0}:\mathbf{1}[\big) =]\mathbf{1}:\infty[$. This proves **2)**.

Lastly, **3)** is implied by the fact that, for arbitrarily large values of t, $K_t(\hat{z})$ accumulates at $\{\infty\}$ and thus $L_t(\hat{y})$ accumulates at $T_{\mathbf{0}\infty}\big(\{\infty\}\big) = \{\mathbf{0}\}$. A similar reasoning applies to arbitrarily negative times. ▨

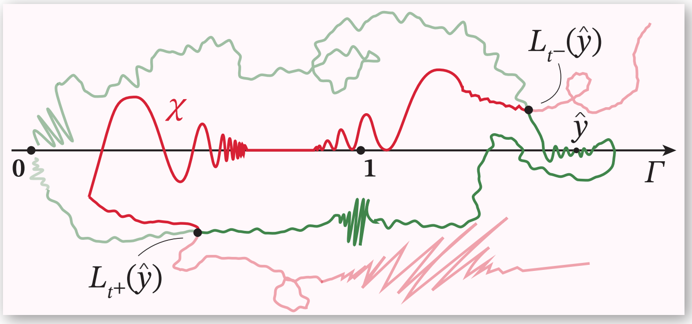
4.12 Lemma. There exist a point $\hat{w} \in]\mathbf{0}:\mathbf{1}[$ and an $\mathcal{J}G_3$-isotopy $(\varphi_t)_{t\in\mathbb{I}}$ such that $\gamma_\varphi(\hat{w}) \simeq \xi_1$ rel $\{\mathbf{0},\mathbf{1},\infty\}$, where $\xi_1(s) = \mathbf{1} + |\hat{w} - \mathbf{1}|\, e^{-2\pi i(s-1/2)}$, $0 \leq s \leq 1$, describes a circle passing through $\hat{w}$ traversed once clockwise while leaving $\mathbf{1}$ on its right and both $\mathbf{0}, \infty$ on its left. In particular, $\hat{w}$ is fixed by φ_1.

Proof. Let $(L_t)_{t\in\mathbb{R}}$ and $\hat{y} \in]\mathbf{1}:\infty[$ be as in Lemma 4.11 above, and let us denote by $\lambda(t) \stackrel{\text{def}}{=} L_t(\hat{y})$ the curve describing the trajectory of the point $\hat{y}$ under the isotopy L. By itens **1)** and **2)** of such Lemma, $\lambda^{-1}(\Gamma)$ is a compact subset of the open interval $(-1,1)$, so we may consider its minimum $\tilde{t} \stackrel{\text{def}}{=} \min \lambda^{-1}(\Gamma) > -1$ and look at the restricted path $\tilde{\lambda} = \lambda\restriction_{(-\infty,\tilde{t}]}$.

Recall the continuum χ from [4.6]. Since $\mathbf{0} \notin \chi$, item **3)** of Lemma 4.11 implies that $\tilde{\lambda}(t_n)$ lies in the same (open) path connected component of χ^c as $\mathbf{0}$ for some very large negative values of $t_n < 0$. In particular, a construction identical to the one described described in [4.8] may be used to conclude – from the fact that χ separates $]\infty:\mathbf{0}[$ from $]\mathbf{1}:\infty[$ – that $\tilde{\lambda}$ must intercept χ. More precisely, $\{t < \tilde{t} : L_t(\hat{y}) \in \chi\} \neq \varnothing$, so its maximum $t^- < \tilde{t}$ is well-defined. Also, it satisfies $L_{t^-}(\hat{y}) \in \mathcal{H}^+$.

In an analogous fashion, but looking at $\max \lambda^{-1}(\Gamma) < 1$ and to the part of item **3)** in Lemma 4.11 concerning arbitrarily large values of time, we may also obtain $t^+ > 0$ such that $L_{t^+}(\hat{y}) \in \chi \cap \mathcal{H}^-$, as suggested by Figure 4.13.

Figure 4.13 – The instants $t^- < 0 < t^+$ are obtained in such a way that $L_{t^-}(\hat{y}) \in \chi \cap \mathcal{H}^+$ and $L_{t^+}(\hat{y}) \in \chi \cap \mathcal{H}^-$.



Lastly, let $(K_t)_{t\in\mathbb{R}}$ be the isotopy in [4.9]. Then, there are real parameters $a \leq -1$ and $b \geq 1$ such that $L_{t^-}(\hat{y}) = K_b(\hat{z})$ and $L_{t^+}(\hat{y}) = K_a(\hat{z})$, since $\hat{z}$ describes χ under K. In particular, $K_c(\hat{z}) \in \Gamma$ for some intermediate $a < c < b$. Even more, the Crossing Lemma 4.6 implies $K_c(\hat{z}) \in \,]\mathbf{0}\,{:}\,\mathbf{1}[$. We set $\hat{w} \stackrel{\text{def}}{=} K_c(\hat{z})$, and finally:

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

The family $(\varphi_t)_{t\in\mathbb{I}}$ is readily seen to form an $\mathcal{IG}_3$-isotopy. Also,

$$\begin{aligned} \varphi_1(\hat{w}) &= K_c \circ K_a^{-1} \circ L_{t^+} \circ L_{t^-}^{-1} \circ K_b \circ K_c^{-1}(\hat{w}) \\ &= K_c \circ K_a^{-1} \circ L_{t^+} \circ L_{t^-}^{-1} \circ K_b\,(\hat{z}) && (\text{for } \hat{w} = K_c(\hat{z})) \\ &= K_c \circ K_a^{-1} \circ L_{t^+}\,(\hat{y}) && (\text{for } K_b(\hat{z}) = L_{t^+}(\hat{y})) \\ &= K_c\,(\hat{z}) = \hat{w} && (\text{for } L_{t^+}(\hat{y}) = K_a(\hat{z})) \end{aligned}$$

Thus, the path $\gamma(s) \stackrel{\text{def}}{=} \varphi_s(\hat{w})$ describing the trajectory of the point $\hat{w}$ is indeed a closed loop based at $\hat{w}$. We now must prove that $\gamma \simeq \xi_1 \text{ rel}\,\{\mathbf{0},\mathbf{1},\infty\}$. To do so, we distinguish four special points along γ:

- the starting and terminal 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 $\gamma(2/3) = L_{t^+}(\hat{y}) = K_a(\hat{z}) \in \mathcal{H}^-$.

Consider first the subpath $\gamma\restriction_{[0,1/3]}$. Since $\gamma(0) \in \,]\mathbf{0}\,{:}\,\mathbf{1}[$ and $\gamma(1/3) \in \mathcal{H}^+$, we may obtain $s_1 = \max\{s \leq 1/3 : \gamma(s) \in \,]\mathbf{0}\,{:}\,\mathbf{1}[\}$. Let $\ell_1(s) = \hat{w} + (\gamma(s_1) - \hat{w})\,s/s_1$ be the standard parameterisaton of the (oriented) arc $[\hat{w}\,{:}\,\gamma(s_1)] \subset \,]\mathbf{0}\,{:}\,\mathbf{1}[$ over the real interval $[0, s_1]$. Then,

$$\kappa : (t,s) \in \mathbb{I} \times [0,s_1] \mapsto (1-t)\,\gamma(s) + t\,\ell_1(s)$$

defines a (straight line) planar homotopy between $\gamma\restriction_{[0,s_1]}$ and ℓ_1, with fixed endpoints $\hat{w}$ and $\gamma(1/3)$.

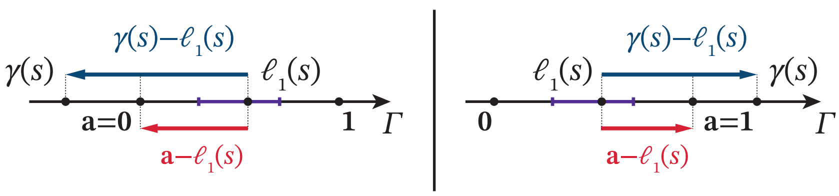
4.12.1 Claim. When thought of as a sphere-valued function, $\varkappa$ is a homotopy relative to $\{\mathbf{0},\mathbf{1},\infty\}$.

Proof. Consider first $\mathbf{a} \in \{\mathbf{0},\mathbf{1}\}$. If $\varkappa(t,s) = \mathbf{a}$ for some $0 \le t \le 1$ and $0 \le s \le s_1$, then

$$(1-t)\,\gamma(s) + t\,\ell_1(s) = \mathbf{a} \text{ or, equivalently,}$$
$$\gamma(s) - \ell_1(s) = \underbrace{(1-t)^{-1}}_{\text{scalar } \ge 1}\big(\mathbf{a} - \ell_1(s)\big)\,.$$

Consequently, $\gamma(s) \in \Gamma$ and also $d\big(\gamma(s), \ell_1(s)\big) > d(\mathbf{a}, \ell_1(s))$. In particular, $\gamma(s) \notin\,]\mathbf{0}:\mathbf{1}[$, as represented in Figure 4.14.

Figure 4.14 – The ruling out of $\varkappa(t,s) = \mathbf{a}$ for $\mathbf{a} = \mathbf{0}$ (left) and $\mathbf{a} = \mathbf{1}$ (right).

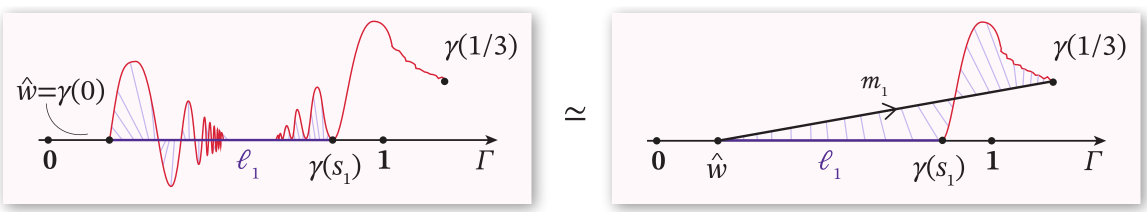


But this is a contradiction, since $\gamma\big([0,1/3]\big) = \{\varphi_t(\hat{w}) : 0 \le t \le 1/3\} = \{K_u(\hat{z}) : c \le u \le b\} \subset \chi$ and $\chi \cap \Gamma \subset\,]\mathbf{0}:\mathbf{1}[$. This reasoning shows that $\varkappa$ is a homotopy with fixed endpoints relative to $\{\mathbf{0},\mathbf{1}\}$. However, since all isotopies under consideration are in G_3, for each t the path $\varkappa(t,\cdot)$ is defined entirely in terms of points known to be finite. Thus, $\varkappa$ is actually a homotopy with fixed endpoints relative to $\{\mathbf{0},\mathbf{1},\infty\}$. $\square$

Now, since by choice of s_1 one has $\gamma(s) \in \mathcal{H}^+$ for every $s_1 < s < 1/3$, and also $\ell_1([0,s_1]) = [\gamma(0):\gamma(s_1)] \subset\,]\mathbf{0}:\mathbf{1}[$, the path $\gamma\restriction_{[s_1,1/3]} * \ell_1$ can be straight line homotoped into the path $m_1 : \mathbb{I} \to \mathbb{S}^2 \setminus \{\mathbf{0},\mathbf{1},\infty\} \cong \mathbb{R}^2 \setminus \{\mathbf{0},\mathbf{1}\}$ describing the line segment from $\gamma(\mathbf{0}) = \hat{w}$ to $\gamma(1/3)$, with fixed endpoints and relative to $\{\mathbf{0},\mathbf{1},\infty\}$, as described in Figure 4.15. Thus:

$$\gamma\restriction_{[0,1/3]} \simeq \gamma\restriction_{[s_1,1/3]} * \gamma\restriction_{[1/3,s_1]} \simeq \gamma\restriction_{[s_1,1/3]} * \ell_1 \simeq m_1 \ \mathrm{rel}\,\{\mathbf{0},\mathbf{1},\infty\}\,.$$

Figure 4.15 – The subpath $\gamma\restriction_{[0,1/3]}$ can be deformed with fixed endpoints and relative to $\{\mathbf{0},\mathbf{1},\infty\}$ into a straight line by juxtaposition of at most two straight line homotopies.



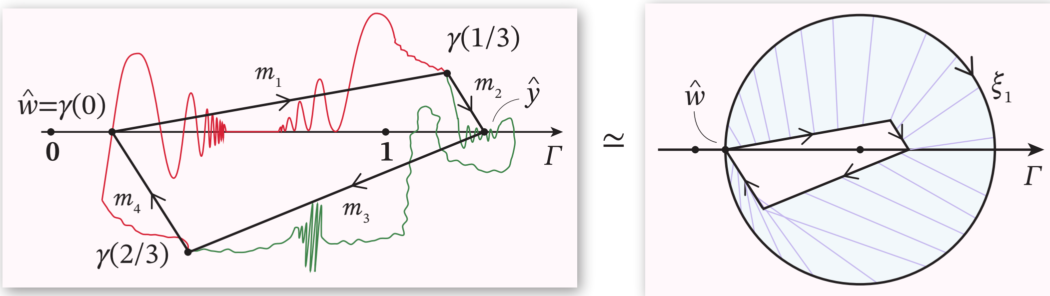
Analogous reasonings involving the juxtaposition of two straight line homotopies at a time further yield the following equivalences:

$$\gamma\restriction_{[1/3,\bar{s}]} \simeq m_2\,, \gamma\restriction_{[\bar{s},2/3]} \simeq m_3 \text{ and } \gamma\restriction_{[2/3,1]} \simeq m_4 \ \mathrm{rel}\,\{\mathbf{0},\mathbf{1},\infty\}\,,$$

where each m_j describes the line segment connecting the endpoints of the corresponding subpath of γ, in the same direction.

Thus, $\gamma \simeq m_4 * m_3 * m_2 * m_1 \ \mathrm{rel}\, \{\mathbf{0}, \mathbf{1}, \infty\}$. But the oriented polygonal $m_4 * \ldots * m_1$ is clearly homotopic to ξ_1 with fixed endpoints relative to $\{\mathbf{0}, \mathbf{1}, \infty\}$, as suggested by Figure 4.16. By transitivity, the Lemma follows. ▨

Figure 4.16 – The path γ describing the trajectory of $\hat{w}$ under φ is homotopic to an oriented polygonal which, in turn, is homotopic to the circle ξ_1.



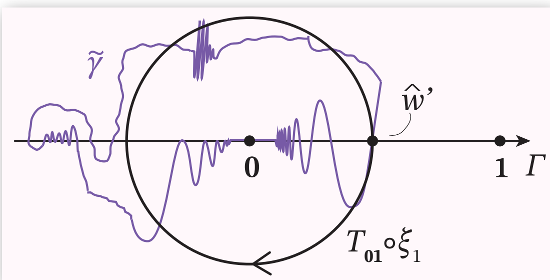
4.13 Corollary. Let $\hat{w}$ be as in Lemma 4.12. Then, there exists an $\mathcal{J}G_3$-isotopy $(\psi_t)_{t\in\mathbb{I}}$ such that $\gamma_\psi(\hat{w}) \simeq \xi_2 \ \mathrm{rel}\, \{\mathbf{0}, \mathbf{1}, \infty\}$, where $\xi_2(s) = |\hat{w}|\, e^{2\pi i s}$, $0 \le s \le 1$, describes a circle passing through $\hat{w}$ traversed once anticlockwise while leaving $\mathbf{0}$ on its left and both $\mathbf{1}, \infty$ on its right.

Proof. Consider $\hat{w}$ and $(\varphi_t)_{t\in\mathbb{I}}$ as yielded by Lemma 4.12, and let $\hat{w}' \overset{\text{def}}{=} T_{\mathbf{01}}(\hat{w})$, which is also a point in the open arc $]\mathbf{0}:\mathbf{1}[$. By Theorem A, there exists h in $\left(G_3\right)_0$ such that $h(\hat{w}') = \hat{w}$. For each $0 \le t \le 1$, we set $\varrho_t = h \circ T_{\mathbf{01}} \circ \varphi_t$.

Then, $(\varrho_t)_{t\in\mathbb{I}}$ is an $\mathcal{J}G_3$-isotopy satisfying $\varrho_1(\hat{w}) = \hat{w}$. Thus, the path describing the trajectory of $\hat{w}$ under ϱ is indeed a closed loop based at $\hat{w}$, explicitly given by $s \in \mathbb{I} \mapsto h\big(\tilde{\gamma}(s)\big)$, where $\tilde{\gamma} = T_{\mathbf{01}} \circ \gamma$ and γ is the path describing the trajectory of $\hat{w}$ under φ, as defined in Lemma 4.12. From that Lemma, we know that $\gamma \simeq \xi_1 \ \mathrm{rel}\, \{\mathbf{0}, \mathbf{1}, \infty\}$. Since $T_{\mathbf{01}}$ leaves the set $\{\mathbf{0}, \mathbf{1}, \infty\}$ invariant, $\tilde{\gamma} \simeq T_{\mathbf{01}} \circ \xi_1 \ \mathrm{rel}\, \{\mathbf{0}, \mathbf{1}, \infty\}$ follows.

Now, since Möbius transformations preserve circles and orientation, $T_{\mathbf{01}} \circ \xi_1$ is a circle, traversed once, clockwise, while leaving $\mathbf{0}$ on its right and both $\mathbf{1}, \infty$ on its left, as pictured in Figure 4.17.

Figure 4.17 – The path $\tilde{\gamma}$ is homotopic with fixed endpoint $\hat{w}'$ to a clockwise traversed circle leaving $\mathbf{0}$ on its interior, relative to $\{\mathbf{0}, \mathbf{1}, \infty\}$.



Let us consider $h \circ \left(T_{\mathbf{01}} \circ \xi_1\right)$. Since h is an orientation preserving homeomorphism fixing $\{\mathbf{0}, \mathbf{1}, \infty\}$, the former is a Jordan curve, also traversed clockwise while leaving $\mathbf{0}$ on its right and both $\mathbf{1}, \infty$ on its left. However, it is now a loop based at $\hat{w}$, for which we have $h \circ \left(T_{\mathbf{01}} \circ \xi_1\right) \simeq h \circ \tilde{\gamma} \ \mathrm{rel}\, \{\mathbf{0}, \mathbf{1}, \infty\}$.

Lastly, for each $t \in \mathbb{I}$ we let $\psi_t \overset{\text{def}}{=} \varrho_{1-t}$. Then, the trajectories $\gamma_\varrho(\hat{w})$ and $\gamma_\psi(\hat{w})$ coincide as sets, but the latter is described by $\overline{h \circ \tilde{\gamma}}$, since the direction of travel is reversed. Consequently,

$$\overline{h \circ \left(T_{\mathbf{01}} \circ \xi_1\right)} \simeq \overline{h \circ \tilde{\gamma}} \ \mathrm{rel}\, \{\mathbf{0}, \mathbf{1}, \infty\}\,.$$

But $\overline{h \circ \left(T_{\mathbf{01}} \circ \xi_1\right)} \simeq \xi_2 \ \mathrm{rel}\, \{\mathbf{0}, \mathbf{1}, \infty\}$ (say via a radial retraction, for concreteness), and the Corollary is thus established. ▨

Equipped with Lemma 4.12 and its Corollary 4.13, we are ready to finish our argument, for setting

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

yields an $\mathcal{J}G_3$-isotopy$(F_t)_{t \in \mathbb{I}}$ under which

$$\gamma_F(\hat{w}) = \gamma_\psi(\hat{w}) * \gamma_\varphi(\hat{w}) \simeq \xi_2 * \xi_1 \ \mathrm{rel}\, \{\mathbf{0}, \mathbf{1}, \infty\}\,,$$

whilst $\xi_2 * \xi_1$ is a prototypical planar figure 8, translating to a topological figure 8 on the surface of the sphere, based at $\hat{w}$ and relative to $\{\mathbf{0}, \mathbf{1}, \infty\}$. This is enough to guarantee that the terminal homeomorphism $F_1 \in \left(G_3\right)_0$ has positive topological entropy as claimed in Theorem B, finishing its proof.

Appendix A

Review on groups of transformations

A.1 Groups of homeomorphisms

Let (X, d) be a compact metric space, and consider the set of all its self homeomorphisms, Homeo (X). On the one hand, it can be endowed with the following *uniform convergence metric*:

[A.1]
$$d_\infty(f, g) \stackrel{\text{def}}{=} \max\Big\{d(f(p), g(p)) \,:\, p \in X\Big\}\,,$$

which is well-defined due to the compacity of X. Its name is justified by the elementary fact that a sequence $(f_n)_{n\in\mathbb{N}}$ converges uniformly to a map f if, and only if, $d_\infty(f_n, f) \to 0$ as $n \to +\infty$.

On the other hand, Homeo (X) can also be endowed with a group structure under the composition operation $(f, g) \mapsto f \circ g$. Then, the identity map id_X plays the role of the identity element and the inverse function f^{-1} plays the role of the inverse element of f. As it turns out, the topological structure defined by [A.1] is *compatible* with this group structure, in the precise sense described below.

A.1 Proposition. Let X be a compact metric space. Then, Homeo (X) is a topological group. More precisely, the maps

$$(f, g) \in \text{Homeo}\,(X) \times \text{Homeo}\,(X) \mapsto f \circ g \in \text{Homeo}\,(X) \quad \text{and}$$
$$f \in \text{Homeo}\,(X) \mapsto f^{-1} \in \text{Homeo}\,(X)$$

are continuous with respect to d_∞ and the corresponding product topology.

Proof. It suffices to show that $(f, g) \mapsto f \circ g^{-1}$ is continuous (p. 143 of (40)). To do so, let $(f_n)_{n\in\mathbb{N}}$ and $(g_n)_{n\in\mathbb{N}}$ be two sequences in Homeo (X) such that $d_\infty(f_n, f) \to 0$ and $d_\infty(g_n, g) \to 0$ as $n \to +\infty$ for some fixed pair $f, g \in$ Homeo (X). We first establish an intermediate result: given $\delta > 0$, there exists $n_1 \in \mathbb{N}$ such that $d_\infty(g_n^{-1}, g^{-1}) \le \delta$ for every $n \ge n_1$.

Indeed, since X is compact, the self map g^{-1} is uniformly continuous, so we may fix $\eta > 0$ such that $d\big(g^{-1}(p), g^{-1}(q)\big) < \delta$ whenever $d(p, q) < \eta$. But then, uniform convergence yields $n_1 \in \mathbb{N}$ such that $d_\infty(g_n, g) < \eta$ for every $n \ge n_1$. In particular, for a fixed $q \in X$ and any $n \ge n_1$:

$$d\Big(g^{-1} \circ g_n\,(q), q\Big) = d\Big(g^{-1}\big(g_n(q)\big)\,,\, g^{-1}\big(g(q)\big)\Big) < \delta\,.$$

The above implies $d_\infty\big(g^{-1} \circ g_n, \mathrm{id}_X\big) \le \delta$ for every $n \ge n_1$, whilst

$$\begin{aligned} d_\infty\big(g^{-1} \circ g_n, \mathrm{id}_X\big) &= \max\Big\{d\Big(g^{-1}(g_n(p)), g_n^{-1}(g_n(p))\Big) \,:\, p \in X\Big\} \\ &= \max\Big\{d\Big(g^{-1}(q), g_n^{-1}(q)\Big) \,:\, q \in X\Big\} = d_\infty\big(g^{-1}, g_n^{-1}\big)\,, \end{aligned}$$

since each $g_n : X \to X$ is a bijection. This proves the claimed intermediate result.

Lastly, given $\varepsilon > 0$, since f : is uniformly continuous as well, there exists $\delta > 0$ such that $d(p,q) < \delta$ implies $d\big(f(p), f(q)\big) < \varepsilon/2$. For this $\delta > 0$, we let $n_1 \in \mathbb{N}$ be as above and $n_0 \geq n_1$ be such that $d_\infty(f_n, f) < \varepsilon/2$ for every $n \geq n_0$. Then, for each $p \in X$ and $n \geq n_0$:

$$\begin{aligned} d\Big(f_n\big(g_n^{-1}(p)\big), f\big(g^{-1}(p)\big)\Big) &\leq d\Big(f_n\big(g_n^{-1}(p)\big), f\big(g_n^{-1}(p)\big)\Big) + d\Big(f\big(g_n^{-1}(p)\big), f\big(g^{-1}(p)\big)\Big) \\ &< d_\infty(f_n, f) + \varepsilon/2 < \varepsilon\,. \end{aligned}$$

In other words, given $\varepsilon > 0$ we obtained $n_0 \in \mathbb{N}$ such that $d_\infty\big(f_n \circ g_n^{-1}, f \circ g^{-1}\big) \leq \varepsilon$ for every $n \geq n_0$, which amounts to the sought convergence. ▨

Let now $\mathcal{M}$ be a *topological manifold* $\mathcal{M}$. This means — by definition — that $\mathcal{M}$ is a topological space which is Hausdorff, second countable and locally Euclidean of dimension m. In particular, it is completely metrisable (17), so we may fix some complete metric $d(\cdot,\cdot)$ generating its topology. If $\mathcal{M}$ further happens to be compact[1] and connected, then it is called a *closed manifold* and, in light of Proposition A.1 above, we may consider the topological group Homeo$(\mathcal{M})$ corresponding to the metric space $(\mathcal{M}, d)$.

Orientation preserving actions

Often, a closed manifold $\mathcal{M}$ can also be oriented, and one wishes to further specialise in the elements of Homeo$(\mathcal{M})$ which are orientation preserving. However, in a purely topological setting it may not be entirely clear at first glance what does it mean for a manifold to be orientable, let alone for a nondifferentiable homeomorphism to preserve orientation. A precise definition is given in terms of *relative homology groups*. Let us quickly review the general theory, and then particularise it.

Let X be a topological space. Given $k \in \mathbb{N}_0$, a *singular k-simplex* in X is a continuous function $\sigma : \triangle_k \to X$ defined on the *standard k-simplex* $\triangle_k$, which is the convex hull of the origin $\boldsymbol{e}_0$ along with the vectors $\boldsymbol{e}_1, \dots, \boldsymbol{e}_k$ comprising the canonical basis of $\mathbb{R}^k$. This fact is summarised by the alternative notation $\triangle_k = [\boldsymbol{e}_0, \boldsymbol{e}_1, \dots, \boldsymbol{e}_k]$, which highlights the natural orientation provided to the simplex by the canonical basis. A *singular k-chain* in X is a (finite) formal sum of the form $\sum_\alpha n_\alpha \sigma_\alpha$, where each σ_α is a singular k-simplex and the coefficients n_α are integers. The free abelian group consisting of all such sums is denoted by $C_k(X)$.

For $k \geq 1$ and $0 \leq i \leq k$, we let $[\boldsymbol{e}_0, \dots, \widehat{\boldsymbol{e}}_i, \dots, \boldsymbol{e}_k] : \triangle_{k-1} \to \triangle_k$ denote the so-called *i-th face mapping*: it is the *unique* affine transformation orderly mapping the $k-1$ vertices of $\triangle_{k-1}$ onto the k vertices of $\triangle_k$ but the i-th. As a set, $\bigcup_{i=0}^k [\boldsymbol{e}_0, \dots, \widehat{\boldsymbol{e}}_i, \dots, \boldsymbol{e}_k](\triangle_{k-1})$ describes the topological boundary of $\triangle_k$. A finer description is provided algebraically by the corresponding *boundary operators*, which are the morphisms $\partial_k : C_k(X) \to C_{k-1}(X)$ defined on generators as:

$$\partial_k \sigma \overset{\text{def}}{=} \sum_{i=1}^{k} (-1)^i \, \sigma \circ [\boldsymbol{e}_0, \dots, \widehat{\boldsymbol{e}}_i, \dots, \boldsymbol{e}_k]\,. \tag{A.2}$$

If the coefficients of a chain are interpreted as a prescription of the net number of times that the corresponding simplex is to be traversed — negative signs meaning "reverse orientation" — then the alternating signs in [A.2] are seen to be chosen so that, in the particular case $\sigma = \mathrm{id}_{\triangle_k}$, the singular $(k-1)$-simplexes that add up to the boundary are oriented coherently with the higher dimensional k-simplex $\triangle_k$ that their images bound topologically.

In general, $C_k(X)$ is a very large object both as a set and as a group: each generator σ is required to be nothing more than a continuous mapping, the term "singular" being attached to it as a remainder that its image must not resemble the corresponding standard simplex at all. The group of k-chains itself is thus not directly studied, but rather a certain quotient of it.

[1] Actually, any compact, Hausdorff and locally Euclidean space is automatically metrisable, but noncompact manifolds shall appear from time to time, so we also embed second countability in the definition.

The starting point is the well-known and purely algebraic observation that $\partial_k \circ \partial_{k+1} \equiv 0$ for every $k \in \mathbb{N}_0$, where ∂_0 is agreed to be trivial by definition. Thus, the set $\operatorname{img} \partial_{k+1}$ of *boundaries* is always a (normal) subgroup of the group $\ker \partial_k$ of the so-called *k-cycles*. This setting, called a *chain complex*, allows for the definition of the *k-th (singular) homology group* of X as the quotient $H_k(X) \stackrel{\text{def}}{=} \ker \partial_k \,/\, \operatorname{img} \partial_{k+1}$, which trivialises cycles that are boundaries of higher dimensional chains. We recall some basic properties:

1) Homology groups are topological invariants or, more precisely, homotopy invariants. In particular, if the space X can be deformation retracted onto a subspace Y via r, then the induced map over singular simplices $\sigma \mapsto r \circ \sigma$ descends to a homology isomorphism $r_* : H_k(X) \xrightarrow{\sim} H_k(Y)$.

2) Of all homology groups, there is one that can always be readily computed: if $X = \bigsqcup_\alpha X_\alpha$ is decomposed into its path-connected components, $H_0(X) \simeq \bigoplus_\alpha \mathbb{Z}$, where each copy of $\mathbb{Z}$ is generated by a chosen point $p_\alpha \in X_\alpha$ or, in other words, a singular 0-simplex.

3) If X consists of a single point, then $H_0(X) \simeq \mathbb{Z}$ and $H_k(X)$ is trivial for every other $k \in \mathbb{N}$. As a consequence of **1)**, the same holds for any contractible space.

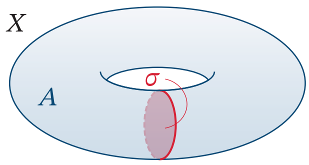
Having recollected some facts about absolute homology, let us consider relative homology, which trivialises a given and fixed subspace A of X. A prototypical situation is pictured in Figure A.1.

A.2 Definition. For $A \subseteq X$, the *k-th relative (singular) homology group* $H_k(X, A)$ consists of all the equivalence classes $[\sigma]$ such that:

1) σ is a *relative k-cycle*, meaning that $\sigma \in C_k(X)$ is a k-chain in X whose boundary $\partial_k \sigma \in C_{k-1}(A)$ is a $(k-1)$-cycle in A,

2) $[\sigma]$ is declared to be trivial if, and only if, $\sigma \in C_k(A) + \operatorname{img} \partial_{k+1}$, meaning that σ consists of chains already in the subspace which is being trivialised plus some absolute boundary from the ambient space.

Figure A.1 – If $X = \mathbb{R}^3$ and A is (the usual embedding of) the 2-torus, σ is a relative 2-cycle which is trivial in $H_2(X)$, but not in $H_2(X, A)$.



There are two important results concerning relative homology which, when put together, lead to a topological definition of orientation. The first relates the absolute and relative groups. To state it, we first recall that a sequence $(\phi_{k+1} : G_{k+1} \to G_k)_{k \in \mathbb{N}_0}$ of groups G_k and homomorphisms ϕ_k is said to be *exact* if $\ker \phi_k = \operatorname{img} \phi_{k+1}$. Then, purely algebraic considerations lead to the following.

A.3 Proposition. For $k \in \mathbb{N}$ there exist morphisms $\partial'_k : H_k(X, A) \to H_{k-1}(A)$ such that

$$\dots \xrightarrow{\partial'_{k+1}} H_k(A) \xrightarrow{(i_k)_*} H_k(X) \xrightarrow{(j_k)_*} H_k(X, A) \xrightarrow{\partial'_k} H_{k-1}(A) \xrightarrow{(i_{k-1})_*} \dots \xrightarrow{(j_0)_*} H_0(X, A) \xrightarrow{\partial'_0} \{0\}$$

is an exact sequence, where:

- each $(i_k)_*$ is induced by the inclusion map $i_k : C_k(A) \hookrightarrow C_k(X)$,
- each $(j_k)_*$ is induced by the quotient map $j_k : c \in C_k(X) \mapsto c + C_k(A)$,
- each ∂'_k maps $[\sigma] \in H_k(X, A)$ to the homology class of $\partial_k \sigma$ in $H_{k-1}(A)$.

The second relevant result, on the other hand, is related to the possibility of deleting or — in the jargon — *excising* a smaller subspace in order to enable computations. Its proof is involved: loosely, it relies on the possibility of subdividing a chain in a way that is subordinate to a given open cover, via iterated barycentric subdivisions. After the theorem, we present two classical and elementary consequences of it.

A.4 The Excision Theorem. Let Z and A be subspaces of X such that $\overline{Z} \subset A^{\circ}$. Then, the inclusion map $\iota : X \setminus Z \hookrightarrow X$ induces an isomorphism $\iota_* : H_k(X \setminus Z, A \setminus Z) \xrightarrow{\sim} H_k(X, A)$ for every $k \in \mathbb{N}_0$.

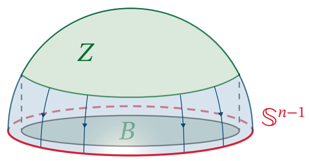
A.5 Lemma. Given $n \in \mathbb{N}_0$, let $\mathbb{S}^n$ denote the *n-sphere* in $\mathbb{R}^{n+1}$. Then, for $n \in \mathbb{N}$:

$$H_k(\mathbb{S}^n) \simeq \begin{cases} \mathbb{Z} & \text{if } k \in \{0, n\}\,, \\ \{0\} & \text{otherwise.} \end{cases}$$

If $n = 0$, then $H_0(\mathbb{S}^0) \simeq \mathbb{Z} \oplus \mathbb{Z}$ and $H_k(\mathbb{S}^0)$ is trivial otherwise.

Proof. Notice first that, for $n \in \mathbb{N}$, $\mathbb{S}^{n-1}$ includes homeomorphically into $\mathbb{S}^n$ as the *equator* $\mathbb{S}^n \cap \{x_{n+1} = 0\} = \mathbb{S}^{n-1} \times \{0\}$. Consider now $H^n_+ \overset{\text{def}}{=} \mathbb{S}^n \cap \{x_{n+1} \geq 0\}$ and $H^n_- \overset{\text{def}}{=} \mathbb{S}^n \cap \{x_{n+1} \leq 0\}$. Each of these hemispheres is homeomorphic to the closed unit ball $\mathbb{B}^n \subset \mathbb{R}^n$ via the projection $\mathrm{prj} : \mathbb{R}^{n+1} \to \mathbb{R}^n$ onto the n first coordinates. We then consider $A = H^n_+$ and $Z = \mathrm{prj}^{-1}(B)$, where B is an open ball centred at the origin of radius slightly smaller than 1. When $n \in \{1, 2\}$, Z is just a cap slightly smaller than the northern hemisphere, as pictured in Figure A.2. The Excision Theorem A.4 then yields an isomorphism $H_k(\mathbb{S}^n \setminus Z, H^n_+ \setminus Z) \simeq H_k(\mathbb{S}^n, H^n_+)$.

Figure A.2 – A sphere from which an (open) spherical cap is deleted may be deformation retracted onto its equator, which is homeomorphic to the one lesser-dimensional sphere.



However, as suggested by Figure A.2, the set $\mathbb{S}^n \setminus Z$ can be deformation retracted onto H^n_- in such a way that $H^n_+ \setminus Z$ is retracted onto the equator $\mathbb{S}^{n-1}$. This retraction thus descends to an isomorphism $H_k(\mathbb{S}^n \setminus Z, H^n_+ \setminus Z) \simeq H_k(H^n_-, \mathbb{S}^{n-1})$ between relative homologies. Collecting the available isomorphisms:

[A.3] $$H_k(\mathbb{S}^n, H^n_+) \simeq H_k(H^n_-, \mathbb{S}^{n-1})\,.$$

Let us analyze each of these groups separately, using the long exact sequence property from Proposition A.3. First, for any $k \in \mathbb{N}$ we consider the segment

$$\dots \xrightarrow{\partial'_{k+1}} H_k(H^n_+) \xrightarrow{(i_k)_*} H_k(\mathbb{S}^n) \xrightarrow{(j_k)_*} H_k(\mathbb{S}^n, H^n_+) \xrightarrow{\partial'_k} H_{k-1}(H^n_+) \xrightarrow{(i_{k-1})_*} H_{k-1}(\mathbb{S}^n) \xrightarrow{(j_{k-1})_*} \dots\,.$$

Fix $p_0 \in \cap_{m \geq 0} \mathbb{S}^m = \mathbb{S}^0$. Since p_0 is a point in the equator of every sphere — and hence in the contractible space H^n_+ — for every $n \in \mathbb{N}$ there exists a (deformation) retraction $r_n : H^n_+ \to \{p_0\}$. Consider the constant map $\phi : \mathbb{S}^n \to \{p_0\}$, which is automatically continuous. Then, if $i : H^n_+ \hookrightarrow \mathbb{S}^n$ denotes inclusion:

$$\begin{tikzcd} H^n_+ \arrow[r, hook, "i"] \arrow[dr, "r_n"'] & \mathbb{S}^n \arrow[d, "\phi"] \\ & \{p_0\} \end{tikzcd} \quad \text{at the level of spaces induces} \quad \begin{tikzcd} H_k(H^n_+) \arrow[r, "(i_k)_*"] \arrow[dr, "\simeq", "(r_n)_*"'] & H_k(\mathbb{S}^n) \arrow[d, "\phi_*"] \\ & H_k(\{p_0\}) \end{tikzcd} \quad \text{in homology.}$$

In particular, $(i_k)_*$ admits a left inverse and is thus injective for every $k \in \mathbb{N}_0$. Exactness then implies $\operatorname{img} \partial'_k = \ker (i_{k-1})_* = \{0\}$ for every $k \in \mathbb{N}$. Since H^n_+ is contractible, all of its homology groups are trivial for $k \in \mathbb{N}$, so we can extract the short exact sequence

$$\{0\} \xrightarrow{(i_k)_*} H_k(\mathbb{S}^n) \xrightarrow{(j_k)_*} H_k(\mathbb{S}^n, H^n_+) \xrightarrow{\partial'_k \equiv 0} H_{k-1}(H^n_+)\,,$$

which allows one to conclude:

[A.4] $$H_k(\mathbb{S}^n) \simeq H_k(\mathbb{S}^n, H^n_+) \text{ for every } k \in \mathbb{N}\,.$$

We are now left to analyze

$$\dots \xrightarrow{\partial'_{k+1}} H_k\big(\mathbb{S}^{n-1}\big) \xrightarrow{(i_k)_*} H_k(H^n_-) \xrightarrow{(j_k)_*} H_k\big(H^n_-, \mathbb{S}^{n-1}\big) \xrightarrow{\partial'_k} H_{k-1}\big(\mathbb{S}^{n-1}\big) \xrightarrow{(i_{k-1})_*} H_{k-1}(H^n_-) \xrightarrow{(j_{k-1})_*} \dots \; .$$

To do so, reset notation so that i now denotes the inclusion $i : \mathbb{S}^{n-1} \hookrightarrow H^n_-$ and r_n denotes the deformation retraction $r_n : H^n_- \to \{p_0\}$. We also let j be the inclusion $\{p_0\} \hookrightarrow \mathbb{S}^{n-1}$. Then, at the level of spaces, $r_n \circ i \circ j = \mathrm{id}_{\{p_0\}}$ holds. Thus, at any homology level $(i_k)_* \circ j_* = (r_n)_*^{-1}$ is an isomorphism, implying $(i_k)_*$ surjective. Consequently, $(j_k)_*$ is trivial for every $k \in \mathbb{N}_0$. In particular, since H^n_- is contractible, for each $k \in \mathbb{N}$ we may extract the following short exact sequence:

$$\{0\} \xrightarrow{(j_k)_*} H_k\big(H^n_-, \mathbb{S}^{n-1}\big) \xrightarrow{\partial'_k} H_{k-1}\big(\mathbb{S}^{n-1}\big) \xrightarrow{(i_{k-1})_*} H_{k-1}(H^n_-) \xrightarrow{(j_{k-1})_* \equiv 0} H_{k-1}\big(H^n_-, \mathbb{S}^{n-1}\big),$$

which, in turn, implies $H_{k-1}(H^n_-) \simeq H_{k-1}\big(\mathbb{S}^{n-1}\big) \,/\, H_k\big(H^n_-, \mathbb{S}^{n-1}\big)$. Considering [A.3] and [A.4],

$$\frac{H_{k-1}\big(\mathbb{S}^{n-1}\big)}{H_k(\mathbb{S}^n)} \simeq H_{k-1}(\mathbb{B}^n) \qquad \text{[A.5]}$$

follows. Since $\mathbb{B}^n$ is contractible for every $n \in \mathbb{N}$, all its homology groups but $H_0(\mathbb{B}^n) \simeq \mathbb{Z}$ are trivial. Also, upon writing $\mathbb{S}^0 = \{-1\} \sqcup \{1\}$ we see that the proposed result holds for it. Then, backward substitution in [A.5] yields the remaining sought results. ▨

A.6 Corollary. Let $\mathcal{M}$ be a topological manifold of dimension $m \geq 1$ and $p \in \mathcal{M}$ be a given point. Then, $H_k(\mathcal{M}, \mathcal{M} \setminus \{p\}) \simeq \mathbb{Z}$ is a free abelian group in one generator if $k = m$, or trivial otherwise.

Proof. Let (U, Φ) be a coordinate chart around p with $\Phi(U) = \mathbb{R}^m$ and $\Phi(p) = \mathbf{0}$, where $\mathbf{0}$ is the origin of Euclidean space. The Excision Theorem A.4 applied to $X = \mathcal{M}$, $A = \mathcal{M} \setminus \{p\}$ and $Z = \mathcal{M} \setminus U$ yields $H_k(\mathcal{M}, \mathcal{M} \setminus \{p\}) \simeq H_k(U, U \setminus \{p\})$ for every $k \in \mathbb{N}_0$, whilst the isomorphism induced by Φ yields $H_k(U, U \setminus \{p\}) \simeq H_k(\mathbb{R}^m, \mathbb{R}^m \setminus \{\mathbf{0}\})$, so it suffices to characterise the latter group.

Refer to the long exact sequence in Proposition A.3. We notice first that $\mathbb{R}^m$ may always be deformation retracted onto some point in $\mathbb{R}^m \setminus \{\mathbf{0}\}$, implying $(i_k)_*$ surjective for every $k \in \mathbb{N}_0$, as already argued during the proof of the preceding Lemma A.5. Exactness then implies $\ker (j_k)_* = \operatorname{img} (i_k)_* = H_k(\mathbb{R}^m)$ or, in other words, $(j_k)_*$ trivial. Thus $\ker \partial'_k = \operatorname{img} (j_k)_* = \{0\}$, meaning that each ∂'_k is injective, for every $k \in \mathbb{N}_0$.

In particular, since ∂'_0 is the trivial morphism, $H_0(\mathbb{R}^m, \mathbb{R}^m \setminus \{\mathbf{0}\}) \simeq \{0\}$ must be trivial as well. Also, since $\mathbb{R}^m$ is contractible, we obtain the following short exact sequences for $k \geq 2$:

$$\{0\} \xrightarrow{(j_k)_*} H_k(\mathbb{R}^m, \mathbb{R}^m \setminus \{\mathbf{0}\}) \xrightarrow{\partial'_k} H_{k-1}(\mathbb{R}^m \setminus \{\mathbf{0}\}) \xrightarrow{(i_{k-1})_*} \{0\},$$

which, in turn, imply ∂'_k surjective and thus an isomorphism. Since $\mathbb{R}^m \setminus \{\mathbf{0}\}$ can be (radially) deformation retracted onto $\mathbb{S}^{m-1}$, we obtain:

$$H_k(\mathbb{R}^m, \mathbb{R}^m \setminus \{\mathbf{0}\}) \simeq H_{k-1}\big(\mathbb{S}^{m-1}\big) \text{ for every } k \geq 2 \, . \qquad \text{[A.6]}$$

Now, since $\mathbb{R}^m$ is path-connected, $H_0(\mathbb{R}^m) \simeq \mathbb{Z}$, whilst surjectivity of $(i_0)_*$ along with the isomorphism theorem yield

$$\mathbb{Z} \simeq \frac{H_0\big(\mathbb{S}^{m-1}\big)}{H_1(\mathbb{R}^m, \mathbb{R}^m \setminus \{\mathbf{0}\})} \simeq \begin{cases} \mathbb{Z} \,/\, H_1(\mathbb{R}^m, \mathbb{R}^m \setminus \{\mathbf{0}\}) & \text{if } m \geq 2 \, , \\ \mathbb{Z} \oplus \mathbb{Z} \,/\, H_1(\mathbb{R}^m, \mathbb{R}^m \setminus \{\mathbf{0}\}) & \text{if } m = 1 \, , \end{cases}$$

in a slight abuse of notation due to the injectivity of ∂'_1. But since all groups involved are free abelian, comparison of ranks implies $H_1(\mathbb{R}^m, \mathbb{R}^m \setminus \{\mathbf{0}\})$ trivial if $m \geq 2$ and $H_1(\mathbb{R}, \mathbb{R} \setminus \{0\}) \simeq \mathbb{Z}$. This, along with [A.6] and the previous Lemma A.5 yield the sought result. ▨

Each of the two possible choices of a generator for $H_m(\mathcal{M}, \mathcal{M} \setminus \{p\})$ is called a *local orientation* at p. The proof of Corollary A.6 above conveys the geometrical information making it a reasonable nomenclature.

Indeed, recall that $H_m(\mathbb{R}^m \setminus \{\mathbf{0}\})$ is identified with $H_m(\mathbb{S}^{m-1})$ via radial projection, and consider first the slightly more complicated case $m = 1$. Then, the isomorphism $H_0(\mathbb{S}^0) \simeq \mathbb{Z} \oplus \mathbb{Z}$ consists of choosing the two points $\{-1, 1\}$ as generators: $(a, b) \leftrightarrow a\,[1] + b\,[-1]$. Since both are also points in the real line, in which any two 0-simplices are equivalent, the inclusion induced homomorphism reads $(i_0)_* : (a, b) \mapsto a + b$. Therefore, $\operatorname{img} \partial_1' = \ker(i_0)_* = \langle(-1, 1)\rangle$. Since ∂_1' was seen to be injective, $H_1(\mathbb{R}, \mathbb{R} \setminus \{0\}) \simeq \langle(-1, 1)\rangle$, which admits two generators, $(+1, -1)$ and $(-1, +1)$. Each such choice basically amounts to prescribe a positive sign to the right side of 0 and a negative sign to the left side or vice-versa.

When $m = 2$, the aforementioned proof implied ∂_2' to be an isomorphism, so we may look directly at the generators of $H_1(\mathbb{S}^1)$. These are loops traversing the circle once, as may be seen from the Hurewicz homomorphism[2], which connects homology to the fundamental group. This same homomorphism shows that changing the direction of travel amounts to a change of sign in homology, so a choice of generator for $H_2(\mathbb{R}^2, \mathbb{R}^2 \setminus \{\mathbf{0}\})$ boils down to a choice between clockwise and anticlockwise.

More generally, it can be shown using degree theory that any choice of generator for $H_{m-1}(\mathbb{S}^{m-1})$ is preserved under rotations and has its sign reversed under reflections — properties which may be taken as reasonable axioms for orientation. Clearly, a local orientation at $\mathbf{0}$ can be transported to any other point of $\mathbb{R}^m$, and thus defines a global orientation. In the case of a general manifold, a global orientation consists of local orientations chosen in such a way that a certain local consistency property holds, as we describe below and picture in Figure A.3.

A.7 Definition. Let $\mathcal{M}$ be a topological manifold without boundary of dimension $m \geq 1$. A choice of local orientations $p \mapsto \alpha_p$ is said to be *continuous* if for every $p \in \mathcal{M}$ there exist

- a coordinate chart (U_p, Φ) around p with $\Phi(U_p) = \mathbb{R}^m$ and $\Phi(p) = \mathbf{0}$;
- a ball $B \subset \mathbb{R}^m$ centred at $\mathbf{0}$;
- and a generator α_B of $H_m(\mathcal{M}, \mathcal{M} \setminus \Phi^{-1}(B))$

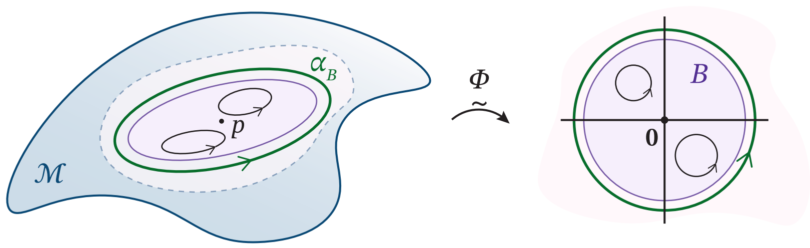
such that, for every $q \in \Phi^{-1}(B)$,

$$\alpha_q = (i_q)_*(\alpha_B)\,,$$

where $(i_q)_* : H_m(\mathcal{M}, \mathcal{M} \setminus \Phi^{-1}(B)) \to H_m(\mathcal{M}, \mathcal{M} \setminus \{p\})$ is the isomorphism induced by the natural inclusion $i_q : \mathcal{M} \setminus \Phi^{-1}(B) \hookrightarrow \mathcal{M} \setminus \{q\}$. Lastly, $\mathcal{M}$ is *orientable* if it admits a continuous choice of local orientations.

A.8 Remark. For the above definition to make sense, we quickly remark that i_q leaves the subspace $\mathcal{M} \setminus \{q\}$ invariant and also that $H_m(\mathcal{M}, \mathcal{M} \setminus \Phi^{-1}(B)) \simeq H_m(\mathbb{R}^m, \mathbb{R}^m \setminus B)$ is isomorphic to $\mathbb{Z}$ for every $m \geq 1$, by radial retraction arguments analogous to those in the proof of Corollary A.6. In particular, it is naturally isomorphic to $H_{m-1}(\mathbb{S}^{m-1})$ when $m \geq 2$.

Figure A.3 – A manifold is orientable if a continuous choice of local orientations is possible — in other words, if small $(m-1)$-chains can be consistently oriented within a neighbourhood of every point via a single chart.



[2] In this case, isomorphism, *cf.* (36), pp. 351–355

Once a continuous choice of local orientations $p \mapsto \alpha_p$ is made, $\mathcal{M}$ is said to be *oriented*. Then, $f \in \mathrm{Homeo}\,(\mathcal{M})$ is *orientation-preserving* if the induced isomorphism f_* in homology satisfies $f_*(\alpha_p) = \alpha_{f(p)}$ for every $p \in \mathcal{M}$. Since induced isomorphisms respect composition, we readily see that $\mathrm{Homeo}_+(\mathcal{M})$ forms a group. We summarise this conclusion below and stretch it a little further without going into details.

A.9 Proposition. Let $\mathcal{M}$ be a closed and oriented topological manifold. Then, the set $\mathrm{Homeo}_+(\mathcal{M})$ of all its orientation-preserving homeomorphisms is a closed subgroup of $\mathrm{Homeo}\,(M)$.

Sketch of proof. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of orientation-preserving maps, converging uniformly to f. We argue that f must be orientation-preserving as well. To do so, we first notice that if σ is a singular k-simplex, then $(f_n \circ \sigma)_{n\in\mathbb{N}}$ converges uniformly to $f \circ \sigma$ over the compact set Δ_k for every $k \in \mathbb{N}_0$.

We fix p in $\mathcal{M}$, and confound a coordinate neighbourhood of $f(p)$ with Euclidean space, thus letting B be a ball centred at $\mathbf{0} = f(p)$ such that $\alpha_y = (i_y)_*(\alpha_B)$ for every $y \in B$, as in Definition A.7. Then, we may choose for α_p a representative σ fully contained in $f^{-1}(B)$. This means that $f_*\alpha_p$ is represented by the m-chain $\beta \overset{\text{def}}{=} f \circ \sigma$ in B, whose boundary is a (compact) $(m-1)$-cycle not meeting $\mathbf{0}$.

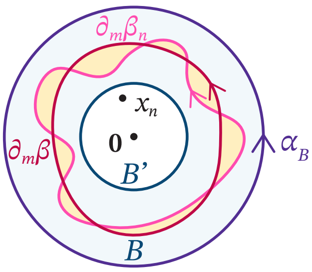
If $\beta_n \overset{\text{def}}{=} f_n \circ \sigma$, each β_n is a representative of $(f_n)_*\alpha_p$. By the uniform convergence property described earlier, we may fix a smaller ball B' centred at $\mathbf{0}$ and $n' \in \mathbb{N}$ such that $x_n \overset{\text{def}}{=} f_n(p) \in B'$ and both $\partial_m\beta, \partial_m\beta_n$ are $(m-1)$-cycles in $\mathbb{R}^m \setminus B'$, for every $n \geq n'$.

Let us denote classes in $H_m(\mathbb{R}^m, \mathbb{R}^m \setminus \{x\})$ by $[\cdot]_x$, and classes in $H_m(\mathbb{R}^m, \mathbb{R}^m \setminus B')$ by $[\cdot]'$. Whenever $n \geq n'$, the inclusion $\mathbb{R}^m \setminus B' \subset \mathbb{R}^m \setminus \{x_n\}$ holds. Since each f_n is orientation-preserving, we thus have the following diagram:

$$\begin{tikzcd} \alpha_p \arrow[r, mapsto] & {[\beta_n]'} \arrow[r, hook, "\text{injection}"] & {[\beta_n]_{x_n} = (f_n)_*\alpha_p} & \alpha_B \arrow[l, hook', "(i_{x_n})_*"'] \end{tikzcd}.$$

In particular, $[\beta_n]' = [\beta_s]'$ whenever $n, s \geq n'$. Recalling Definition A.2, this means that $\beta_n - \beta_s \in C_m(\mathbb{R}^m \setminus B')$. Since $\mathbb{R}^m \setminus B'$ is a closed set, taking the pointwise limit as $s \to \infty$ implies $\beta_n - \beta \in C_m(\mathbb{R}^m \setminus B')$. Hence, $[\beta]' = [\beta_n]'$ for every $n \geq n'$, as suggested in Figure A.4.

Figure A.4 – The important information in relative homology is actually encoded by the boundaries, which become equivalent modulo a bounded neighbourhood of $f(p)$ for large n.



Let now r_x denote radial projection from $x \in B'$ onto $\partial B'$. Then, it determines a deformation retraction from $\mathbb{R}^m\setminus\{x\}$ onto $\mathbb{R}^m\setminus B'$, thus descending to an isomorphism in relative homology groups. Since $\mathbb{R}^m\setminus B \subset \mathbb{R}^m\setminus B'$, the following diagram commutes for every $n \geq n'$:

$$\begin{tikzcd} & H_m(\mathbb{R}^m, \mathbb{R}^m \setminus \{x_n\}) \arrow[dr, "(r_{x_n})_*", "\simeq"'] & \\ H_m(\mathbb{R}^m, \mathbb{R}^m \setminus B) \arrow[ur, hook, "(i_{x_n})_*"'] \arrow[dr, hook, "(i_{\mathbf{0}})_*"'] & & H_m(\mathbb{R}^m, \mathbb{R}^m \setminus B') \\ & H_m(\mathbb{R}^m, \mathbb{R}^m \setminus \{\mathbf{0}\}) \arrow[ur, "\simeq", "(r_{\mathbf{0}})_*"'] & \end{tikzcd}$$

Lastly, $[\beta]' = [\beta_n]'$ and $[\beta_n]_{x_n} = [\alpha_B]_{x_n}$ imply $[\beta]_{\mathbf{0}} = [\alpha_B]_{\mathbf{0}}$, which translates to $f_*\alpha_p = \big(i_{f(p)}\big)_*(\alpha_B)$. ▨

Although the considerations made so far settle the meaning of the group $\mathrm{Homeo}_+(\mathcal{M})$, no concrete instances of it are available yet. Indeed, very few examples — if any — can be given of orientable topological manifolds from Definition A.7 alone. In a regular Algebraic Topology course, the subject is further developed by the introduction of a new manifold $\widetilde{\mathcal{M}}$, whose underlying set consists of all local orientations of $\mathcal{M}$, topologised in such a way that $\alpha_p \mapsto p$ is a double sheeted covering map. Then, global orientations — if they exist — arise as continuous sections in $\Gamma(\mathcal{M}, \widetilde{\mathcal{M}})$. This is a framework well-suited to prove Proposition A.9, for then it is simply the statement that convergent sequences assuming discrete values are eventually constant.

Nevertheless, even the classical textbook theorems on the subject yield as orientable topological manifolds essentially those which are simply-connected or products of manifolds known to be orientable, apart from a plethora of necessary conditions or specialised results given as exercises — *cf. e.g.* Sec. 3.3 of (22) and Ch. VII §2 of (9).

Although topological manifolds not admitting a differentiable structure do exist, much more familiar orientability criteria are available for smooth manifolds (29): the existence of an atlas for which the transition functions have positive Jacobian determinant; the existence of a continuous volume form or even the existence of a continuous normal field for a realisation of the manifold as a hypersurface of Euclidean space, among others.

A.2 Groups of diffeomorphisms

Let $\mathcal{M}$ be a smooth manifold. In other words, a topological manifold for which an atlas $\{(U_\alpha, \Phi_\alpha)\}_\alpha$ presenting smooth transition functions exists. Then, the Whitney weak topology — introduced via subbasic neighbourhoods for diffeomorphisms — can actually be used to topologise the entire set $C^1(\mathcal{M})$ of functions $\mathcal{M} \to \mathcal{M}$ of class C^1. To do so, the subbasic neighbourhoods $\mathcal{B}(f\,;\,(U,\Phi), K, (V,\Psi)\,;\,\varepsilon)$ are defined in the exact same way, only now its elements (f included) are taken from the whole of $C^1(\mathcal{M})$. As mentioned earlier, this topology is not very convenient for computations. An alternative description is given by jets, a generalisation of the Taylor polynomial that we now recall.

A.10 Definition. Let $\mathcal{M}$ be a smooth manifold, and consider the following equivalence relation in $\mathcal{M} \times C^1(\mathcal{M})$:

$$(p, f) \sim (q, g) \iff q = p\,,\ f(p) = g(p) \text{ and } \mathrm{D}f(p) = \mathrm{D}g(p)\,.$$

The *jet of f at p* is the equivalence class $j_p f$ of (p, f) under this relation, while the *space of jets* is the quotient $J\mathcal{M} \overset{\text{def}}{=} \mathcal{M} \times C^1(\mathcal{M})/\sim\, = \{j_p f \,:\, p \in \mathcal{M} \text{ and } f \in C^1(\mathcal{M})\}$.

A.11 Remark. Clearly, two jets $j_p f$ and $j_q g$ agree if, and only if, $q = p$ and for every pair of coordinate charts (U, Φ) around p and (V, Ψ) around $f(p)$ such that both $f(U)$ and $g(U)$ are cointaned in V, the local representations $\hat{f}$ and $\hat{g}$ have the same first order Taylor polynomial at $\Phi(p) \in \mathbb{R}^m$.

The set $J\mathcal{M}$ can be topologised as to become itself a manifold, a process that can be canonically carried out as long as reasonable candidates for local charts are available — cf. e.g. Lee (35, Lemma 1.35). In this particular case, if $\{(U_\alpha, \Phi_\alpha)\}_\alpha$ is an atlas of $\mathcal{M}$, let $\mathcal{V}_{\alpha\beta}$ denote the set of jets $j_p f$ such that $p \in U_\alpha$ and $f(p) \in U_\beta$. Then, consider

$$\begin{aligned} \Theta_{\alpha\beta} \,:\, \mathcal{V}_{\alpha\beta} &\to \Phi(U_\alpha) \times \Phi\big(U_\beta\big) \times \mathrm{GL}_m(\mathbb{R}) \\ j_p f &\mapsto \Big(\Phi_\alpha(p),\ f_{\alpha\beta}(\Phi_\alpha(p))\,,\ \mathrm{D}f_{\alpha\beta}(\Phi_\alpha(p))\Big) \end{aligned}\,,$$

where $f_{\alpha\beta} = \Phi_\beta \circ f \circ \Phi_\alpha^{-1}$ is the local expression of f between suitable open sets of Euclidean m-space. Due to Remark A.11, the above yields a well-defined bijection between $\mathcal{V}_{\alpha\beta}$ and an open subset of $\mathbb{R}^m \times \mathbb{R}^m \times$

$\mathrm{GL}_m(\mathbb{R}) \simeq \mathbb{R}^{2m+m^2}$. Upon checking that the collection $\{\mathcal{V}_{\alpha\beta}\}_{\alpha,\beta}$ inherits from $\mathcal{M}$ Hausdorffness and second countability, $J\mathcal{M}$ is made into a $(2m+m^2)$-dimensional topological manifold by the charts $\{\mathcal{V}_{\alpha\beta}, \Theta_{\alpha\beta}\}_{\alpha,\beta}$.

For each $f \in C^1(\mathcal{M})$, one may thus consider the so-called *prolongation* $jf : \mathcal{M} \to J\mathcal{M}$, mapping $p \in \mathcal{M}$ to $(jf)(p) \overset{\text{def}}{=} j_p f$. These are continuous as functions of p, for a local expression is $z \in \Phi_\alpha(U_\alpha) \subset \mathbb{R}^m \mapsto \big(z,\, f_{\alpha\beta}(z),\, \mathrm{D}f_{\alpha\beta}(z)\big)$. Therefore, each jf lies in the set $\Gamma(\mathcal{M}, J\mathcal{M})$ of continuous sections of the bundle $J\mathcal{M} \longrightarrow \mathcal{M}$. Even more, j is actually an injection of $C^1(\mathcal{M})$ into $\Gamma(\mathcal{M}, J\mathcal{M})$, for $jf = jg$ implies $f(p) = g(p)$ for every $p \in \mathcal{M}$.

The set $\Gamma(\mathcal{M}, J\mathcal{M})$, in turn, carries the *compact-open topology*, which is also specified by a subbasis. Namely, the one comprising neighbourhoods of the form $\mathcal{N}(K, \mathcal{V}) = \{\sigma : \sigma(K) \subset \mathcal{V}\}$, where K is compact and $\mathcal{V}$ is open. In particular, this subbasis pulls back under j to a subbasis of some topology in $C^1(\mathcal{M})$, say a "jets topology". With respect to it, we have the following.

A.12 Lemma. The "jets topology" and the Whitney weak topology in $C^1(\mathcal{M})$ actually coincide.

Proof. Let $f \in C^1(\mathcal{M})$ be such that $jf \in \mathcal{N}(K, \mathcal{V})$. Then, one may fix finitely many open sets $B_1, \dots, B_n$ such that $K \subset \bigcup_{i=1}^n B_i$ and each $\overline{B_i}$ is contained in some coordinate neighbourhood U_{α_i} satisfying $f(U_{\alpha_i}) \subset U_{\beta_i}$, where U_{β_i} is some other coordinate neighbourhood. Clearly, each $\Theta_{\alpha_i\beta_i}\big(\mathcal{V}_{\alpha_i\beta_i} \cap \mathcal{V}\big)$ is an open set containing the compact set $K_i \overset{\text{def}}{=} \Theta_{\alpha_i\beta_i}\big[jf(\overline{B_i})\big]$. Thus, one may fix $\varepsilon > 0$ such that $\{w : \operatorname{dist}(w, K_i) \le \varepsilon\} \subset \Theta_{\alpha_i\beta_i}\big(\mathcal{V}_{\alpha_i\beta_i} \cap \mathcal{V}\big)$ for every $i \in \{1, \dots, n\}$. We claim the following:

[A.7] $$jg(K) \subset \mathcal{V} \text{ whenever } g \in \bigcap_{i=1}^{n} \mathcal{B}\Big(f\,;\, (U_{\alpha_i}, \Phi_{\alpha_i}), \overline{B_i}, (U_{\beta_i}, \Phi_{\beta_i})\,;\, \varepsilon\Big)\,.$$

Indeed, given such g, it is known by definition that $g(\overline{B_i}) \subset U_{\beta_i}$ for each i, so for a point $p \in K \cap B_i$ it makes sense to compute

$$\Theta_{\alpha_i\beta_i}(j_p g) = \Big(\,\Phi_{\alpha_i}(p)\,,\, g_{\alpha_i\beta_i}(\Phi_{\alpha_i}(p))\,,\, \mathrm{D}g_{\alpha_i\beta_i}(\Phi_{\alpha_i}(p))\Big) \in \Phi_{\alpha_i}(U_{\alpha_i}) \times \Phi_{\beta_i}(U_{\beta_i}) \times \mathbb{R}^{m^2}\,.$$

Since $\Theta_{\alpha_i\beta_i}(j_p f) \in K_i$ for $p \in B_i \cap K$, condition [1.2] yields $\operatorname{dist}\big(\Theta_{\alpha_i\beta_i}(j_p g), K_i\big) \le \varepsilon$, say with respect to the maximum norm. The choice of ε then implies $\Theta_{\alpha_i\beta_i}(j_p g) \in \Theta_{\alpha_i\beta_i}\big(\mathcal{V}_{\alpha_i\beta_i} \cap \mathcal{V}\big)$ for every $p \in B_i \cap K$, or yet $jg(B_i \cap K) \subset \mathcal{V}_{\alpha_i\beta_i} \cap \mathcal{V}$ for each $i \in \{1, \dots, n\}$, which is enough to establish the claim.
Claim [A.7] above states that the "jets topology" is coarser than the Whitney weak topology. The converse inclusion is simpler: if $g \in \mathcal{B}\big(f\,;\,(U_\alpha, \Phi_\alpha), K, (U_\beta, \Phi_\beta)\,;\,\varepsilon\big)$ then $jg \in \mathcal{N}(K, \mathcal{V})$, where

$$\mathcal{V} = \Theta_{\alpha\beta}^{-1}\Big(\,\Phi_\alpha(U_\alpha) \times \Big(K_\varepsilon \cap [\,\Phi_\beta(U_\beta) \times \mathbb{R}^{m^2}\,]\Big)\Big)$$

and K_ε is an open 2ε-neighbourhood of the compact set $(f_{\alpha\beta} \times \mathrm{D}f_{\alpha\beta})\big(\Phi_\alpha(K)\big)$. ▨

Lemma A.12 above has key implications to the C^1 topology, as the compact-open topology is textbook material (§46 of (40)):

- First, when the target space is metric, the compact-open topology is equivalent to that of uniform convergence over compact sets. This implies the injection of $C^1(\mathcal{M})$ to be a closed subspace of $\Gamma(\mathcal{M}, J\mathcal{M})$, as may be seen upon resorting to convex balls compactly contained in coordinate domains and standard Real Analysis results on the uniform convergence of a sequence of functions and their derivatives.
- Second, when the source space $\mathcal{M}$ is compact, the compact-open topology further simplifies to the (metrisable) topology of uniform convergence, thus confirming that convergence in the C^1 topology implies convergence in the metric d_∞ defined in [A.1]. This also attaches a meaning to expressions such as "C^1 close". Once a finite atlas is fixed, one may even consider the "C^1 norm" of a function, an expression often found in the literature.

- Lastly, since $\mathcal{M}$ and $J\mathcal{M}$ are both manifolds, the first is locally compact with countable base, whilst the second is completely metrisable. The compact-open topology in $\Gamma(\mathcal{M}, J\mathcal{M})$ is thus completely metrisable as well (Theorem 4.1 in (24)). This allows one to use sequential criteria for continuity, as in Proposition A.13 ahead.

A.13 Proposition. Let $\mathcal{M}$ be a smooth manifold, not necessarily compact. Then, $\mathrm{Diff}^1(\mathcal{M})$ endowed with the C^1-topology and the composition operation is a topological group.

Proof. Let $(f_n)_{n\in\mathbb{N}}$ and $(g_n)_{n\in\mathbb{N}}$ be two sequences in $\mathrm{Diff}^1(\mathcal{M})$ such that $jf_n \to jf$ and $jg_n \to jg$ with respect to the compact-open topology. The goal is to prove that $j(f_n \circ g_n)$ converges uniformly to $j(f\circ g)$ on compact sets of $\mathcal{M}$. Let thus $K \subset \mathcal{M}$ be a given compact set.

For every $p \in K$, there exist coordinate neighbourhoods U_α and U_β — both depending on p — such that $p \in U_\alpha$ and $(f\circ g)(U_\alpha) \subset U_\beta$. Also, there exists a coordinate neighbourhood U_γ such that $g(p) \in U_\gamma$. Thus, $g(p)$ belongs to the open set $U_\gamma \cap f^{-1}(U_\beta)$. Hence, one may fix an open set B_p around p such that $\overline{B_p}$ is a compact subset of U_α satisfying $g(\overline{B_p}) \subset U_\gamma \cap f^{-1}(U_\beta)$. In particular, $jg(\overline{B_p}) \subset \mathcal{V}_{\alpha\gamma}$, $j(f\circ g)(\overline{B_p}) \subset \mathcal{V}_{\alpha\beta}$ and $jf\big[g(\overline{B_p})\big] \subset \mathcal{V}_{\gamma\beta}$. Given this setting, we establish an intermediate local result before concluding.

A.13.1 Claim. Let $B \subset \mathbb{R}^m$ be a bounded open set and let $(\hat{g}_n)_{n\in\mathbb{N}}$, $(\hat{f}_n)_{n\in\mathbb{N}}$ be sequences of functions of class C^1 such that:

- $(\hat{g}_n)_{n\in\mathbb{N}}$ and their differentials converge uniformly on $\overline{B}$ to a function g;
- $(\hat{f}_n)_{n\in\mathbb{N}}$ and their differentials converge uniformly to a function f on compact sets.

Then, $(\hat{f}_n \circ \hat{g}_n)_{n\in\mathbb{N}}$ and their differentials converge uniformly to $\hat{f}\circ\hat{g}$ on $\overline{B}$.

Proof. Since the uniform convergence $\hat{f}_n \circ \hat{g}_n \to \hat{f}\circ\hat{g}$ is standard Real Analysis, let us only verify uniform convergence of the differentials. To do so, we switch to the prime notation $\mathrm{D}f(x) = f'(x)$. For $x \in B$:

$$\begin{aligned}\Big\| \hat{f}_n'(\hat{g}_n(x)) \circ \hat{g}_n'(x) - \hat{f}'(\hat{g}(x)) \circ \hat{g}'(x) \Big\| \le{}& \Big\| \hat{f}_n'(\hat{g}_n(x)) - \hat{f}'(\hat{g}_n(x)) \Big\| \, \| \hat{g}_n'(x) \| \\ &+ \Big\| \hat{f}'(\hat{g}_n(x)) - \hat{f}'(\hat{g}(x)) \Big\| \, \| \hat{g}_n'(x) \| \\ &+ \| \hat{f}'(\hat{g}(x)) \| \, \| \hat{g}_n'(x) - \hat{g}'(x) \| \,.\end{aligned}$$

The sought convergence then follows from the uniform boundedness of $(g_n')_{n\in\mathbb{N}}$ and the fact that $\hat{f}'\circ\hat{g}_n$ converges uniformly to $\hat{f}'\circ\hat{g}$ on $\overline{B}$ as well. □

Now, there are finitely many points $p_1, \dots, p_k$ such that the sets $B_i \overset{\text{def}}{=} B_{p_i}$ cover K. Thus, for every sufficiently large n one is allowed to compute both $\Theta_{\alpha_i\gamma_i} \circ jg_n$ and $\Theta_{\alpha_i\beta_i} \circ j(f_n\circ g_n)$ on each set $\overline{B_i}$, as well as $\Theta_{\gamma_i\beta_i} \circ jf_n \circ g_n$. Since jg_n converges uniformly to jg on $\overline{B_i}$ and jf_n converges uniformly to jf on $g(\overline{B_i})$, Claim A.13.1 above applies to $B = B_i$, $\hat{g}_n = \Phi_{\gamma_i} \circ g_n \circ \Phi_{\alpha_i}^{-1}$, $\hat{f}_n = \Phi_{\beta_i} \circ f_n \circ \Phi_{\gamma_i}^{-1}$ and $\hat{g}$, $\hat{f}$ defined accordingly. But

$$\hat{f}_n \circ \hat{g}_n = \big(\Phi_{\beta_i} \circ f_n \circ \Phi_{\gamma_i}^{-1}\big) \circ \big(\Phi_{\gamma_i} \circ g_n \circ \Phi_{\alpha_i}^{-1}\big) = \Phi_{\beta_i} \circ \big(f_n \circ g_n\big) \circ \Phi_{\alpha_i}^{-1} \,.$$

Therefore, the uniform convergence of $(\hat{f}_n\circ\hat{g}_n)_{n\in\mathbb{N}}$ and its derivatives on $\overline{B_i}$ amounts to that of $\Theta_{\alpha_i\beta_i}\circ j(f_n\circ g_n)$. This implies $j(f_n\circ g_n)$ to converge uniformly to $j(f\circ g)$ on each B_i, and thus on K. Since K was arbitrary, continuity of composition is proven. It now remains to prove continuity of inversion.

Given that continuity of composition in $\mathrm{Diff}^1(\mathcal{M})$ is established, it suffices to show that $f\circ f_n^{-1} \to \mathrm{id}_{\mathcal{M}}$ under the assumption that $f_n \to f$, all convergences being with respect to the C^1-topology. But, if $K \subset \mathcal{M}$ is a compact set contained in the domains of two charts (U,Φ) and (V,Ψ), in coordinates one has:

$$|(\hat{f}\circ\hat{f}_n^{-1})(x) - x| = |\hat{f}(\hat{f}_n^{-1}(x)) - \hat{f}_n(\hat{f}_n^{-1}(x))| \le \sup_{y\in\Phi(K)} |\hat{f}(y) - \hat{f}_n(y)| \,,$$

which tends to zero as $n\to\infty$. Convergence of the differentials then follows from the identity $\mathrm{D}\big(\hat{f}_n^{-1}\big)(x) = \big[\,\mathrm{D}\hat{f}_n\,(\hat{f}_n^{-1}(x))\,\big]^{-1}$, along with the continuity of matrix inversion. ▨

Appendix B

Thurston classification and the figure 8

This appendix is devoted to shed some light into the reasons why the presence of a topological figure 8, as defined during the proof of Theorem B and depicted in Figure 4.10, implies the presence of positive topological entropy in G_0. This is a generally accepted fact in the Dynamical Systems lore, and we by no means intend to give here a mathematically correct proof of it. Instead, we give a general idea of why such must hold, and point out some of the subtleties involved.

In order to do so, we must first remember the concept of a pseudo-Anosov homeomorphism — or rather, a relative pseudo-Anosov. For that, we follow (15).

B.1 Definition. Let $\mathcal{S}$ be a closed smooth surface and P be a finite set. Then, a homeomorphism $f : \mathcal{S} \to \mathcal{S}$ is said to be *pseudo-Anosov relative to P* if it leaves invariant two mutually transverse measured singular foliations $(\mathcal{F}^{\mathrm{s}}, \mu^{\mathrm{s}})$ and $(\mathcal{F}^{\mathrm{u}}, \mu^{\mathrm{u}})$ such that:

1) $f_* \mu^{\mathrm{s}} = \beta^{-1} \mu^{\mathrm{s}}$ and $f_* \mu^{\mathrm{u}} = \beta \mu^{\mathrm{u}}$ for some $\beta > 1$,

2) a singularity in $\mathrm{Sing}(\mathcal{F}^{\mathrm{s}}) = \mathrm{Sing}(\mathcal{F}^{\mathrm{u}})$ is a 1-prong if, and only if, it lies in P. In particular, P is permuted under f.

When $P = \varnothing$ or — in other words — the foliations present k-prongs only for $k \geq 3$, f is called simply *pseudo-Anosov*.

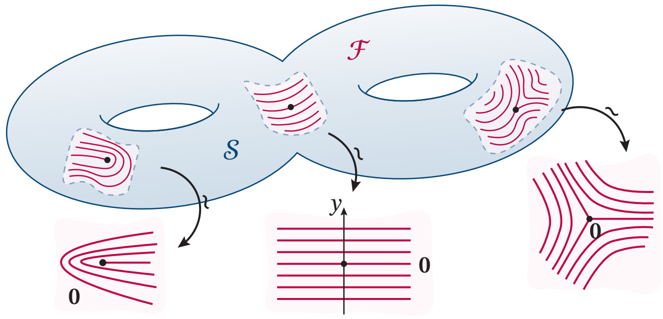
Let us briefly clarify the terminologies introduced above, for the sake of completeness. First, recall that a *singular foliation* $\mathcal{F}$ is a partition of a closed surface $\mathcal{S}$ into a set $\{\phi_x\}_{x \in \mathcal{S}}$ of lesser dimensional submanifolds called *leaves*, along with the prescription of a *finite* set of *singularities* $\mathrm{Sing}(\mathcal{F})$ such that:

- for every nonsingular point $p \in \mathcal{S} \setminus \mathrm{Sing}\,\mathcal{F}$ there exists a coordinate chart $\left(U_p, \Psi\right)$ around p such that $\Psi(p) = \mathbf{0}$ and, for every leaf $\phi \in \mathcal{F}$ intercepting U_p, the set $\phi \cap U_p$ corresponds under Ψ to some horizontal line segment;
- for every singular point $p \in \mathrm{Sing}(\mathcal{F})$ there exists exists a coordinate chart $\left(U_p, \Psi\right)$ around p such that $\Psi(p) = \mathbf{0}$ and, for every leaf $\phi \in \mathcal{F}$ intercepting U_p, the set $\phi \cap U_p$ corresponds under Ψ to some level set of $\mathrm{Im}\left(z^{k/2}\right)$, where $k \in \mathbb{N} \setminus \{2\}$.

This extensive definition is conveyed in Figure B.1. A singular point is named a *k-prong* after the number k figuring in the second bullet above, while the charts around nonsingular points are called *flow boxes*.

We can now assign meaning to transversality: an arc $\alpha : \mathbb{I} \to \mathcal{S}$ is *$\mathcal{F}$-transverse* if its image does not intercept $\mathrm{Sing}(\mathcal{F})$ and every point in $\alpha(\mathbb{I}^\circ)$ admits a flow box in which α corresponds to a curve transverse to the foliation by horizontal lines in the usual sense of the complex plane.

Figure B.1 – Charts around points of a closed surface $\mathcal{S}$ endowed with a singular foliation $\mathcal{F}$: a 1-prong, a flow box and a 3-prong, respectively.

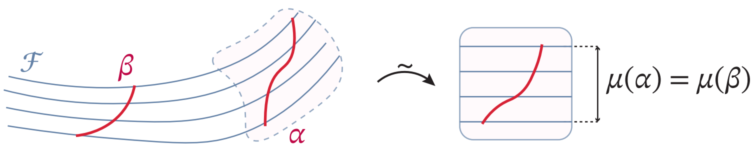


Now, a pair $(\mathcal{F},\mu)$ is said to be a *measured foliation* if μ is a real valued function defined on the set of all $\mathcal{F}$-transverse arcs and such that:

- if β is homotopic to α along $\mathcal{F}$, then $\mu(\alpha)=\mu(\beta)$;
- every nonsingular point of $\mathcal{F}$ admits a flow box in which the measure of transverse arcs pushes forward to the vertical Lebesgue measure $|\mathrm{d}y|$.

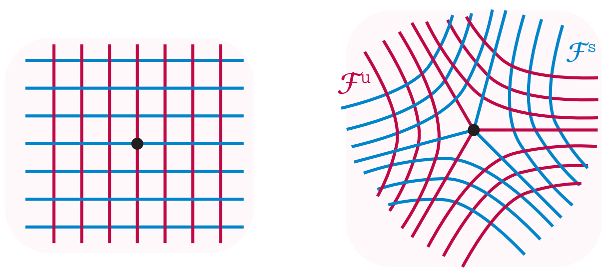
Once more, a handful of terms asking for a precise definition were introduced, but we rather just illustrate them in Figure B.2.

Figure B.2 – The conditions on a measured foliation: pushing-forward to the vertical Lebesgue measure in flow boxes and assigning the same measure to transverse arcs that can be slided onto each other over the same leaves.



In light of the above, we can now make sense of Definition B.1. First, to say that the two measured foliations $(\mathcal{F}^{\mathrm{s}},\mu^{\mathrm{s}})$ and $(\mathcal{F}^{\mathrm{u}},\mu^{\mathrm{u}})$ are transverse means that they share the same set of singularities and, furthermore, around every point of $\mathcal{S}$ there exists a chart under which these foliations correspond to mutually orthogonal families of curves in the usual sense of the complex plane. The prototypical situations for a nonsingular point and a 3-prong are shown in Figure B.3.

Figure B.3 – Prototypical models for a flow box and a 3-prong of a pair of transverse foliations $\mathcal{F}^{\mathrm{u}}$ and $\mathcal{F}^{\mathrm{s}}$.



Now, it may not be clear at first glance what condition **1)** implies, for the measures are *a priori* defined in arcs transverse to the given foliations: measured foliations are to be thought of as providing an "intrinsic" arc length function to transverse arcs. Thus, μ^{u} provides a way to measure arcs of the stable foliation and vice-versa.

In this case, **1)** reads for an arc α of stable foliation:

$$\mu^{\mathrm{u}}(f(\alpha)) = \beta^{-1}\,(f_*\mu^{\mathrm{u}})(f(\alpha)) = \beta^{-1}\,\mu^{\mathrm{u}}(\alpha)\,.$$

Since $\beta > 1$, we see at once that the leaves of the stable foliation are contracted under the action of f, whilst the leaves of the unstable foliation are stretched. For this reason, the number β is called the *stretching factor* of f. This is to be seen as a generalisation of the concept of Anosov diffeomorphisms — that is, diffeomorphisms which are uniformly hyperbolic. In that case, actual stable and unstable manifolds are available, and their lengths are indeed contracted and stretched accordingly.

Let us not bother with condition **2)** for now, and consider instead ordinary pseudo-Anosov maps, for which a very broad and well developed theory is available from the Dynamical Systems viewpoint. For example, it is known that

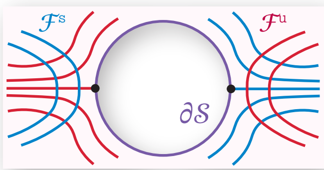
- the periodic orbits for f are dense in $\mathcal{S}$, and actually form a residual set;
- f has strictly positive topological entropy;

and, furthermore, some form of these properties carry over to a map g that is only isotopic to a pseudo-Anosov. More specifically, Theorem 2 in (21) by Handel establishes the existence of a closed and g-invariant subset $Y \subset \mathcal{S}$ on which the dynamics of g has the dynamics of f as a factor. In particular, this implies g to have strictly positive topological entropy as well.

As it turns out, all of the above holds for relative pseudo-Anosov homeomorphisms as well. The idea is a little roundabout, for it consists in blowing up the points in P into boundary components of a new surface $\mathcal{S}'$, in such a way that f induces on this new surface an actual pseudo-Anosov map. Then, this new map factors over the original one via collapsing these components into points. How this may be carried out in general is hinted at in p. 559 of (28), while (38) does it a little more explicitly for the case in which P is a periodic orbit of a surface diffeomorphism, fitting us rather well.

Of course, one notices that we first defined a pseudo-Anosov map on a closed surface. Fortunately, surfaces with boundary are comprised by the available theory as well, but some further restrictions must be made. Namely, one asks the usual Definition B.1 to hold in $\mathcal{S}\setminus\partial\mathcal{S}$ and also for $\partial\mathcal{S}$ to decompose into components that are both stable and unstable leaves at the same time, in such a way that their singularities alternate and are all 3-prongs. Once again, we don't bother making these precise, referring instead to Figure B.4 and to Chapters 4 through 7 of (4) for an exposition aimed at graduate students.

Figure B.4 – On a surface with boundary, each segment of a boundary component is a stable and unstable leaf.

$p \in \mathrm{Sing}(\mathcal{F}^{\mathrm{u}})$  $q \in \mathrm{Sing}(\mathcal{F}^{\mathrm{s}})$

We are now ready to state the (presently classical) Nielsen-Thurston Classification Theorem. Several forms of it became available in the literature since its consolidation in 1978. We present the one given in Section 5 of (39), for it allows one to locate the periodic points of a given map at once.

B.2 The Nielsen-Thurston Classification Theorem. Let $\mathcal{S}$ be a compact surface and $g : \mathcal{S} \to \mathcal{S}$ be a homeomorphism for which a (possibly empty) finite set $P \subset \mathcal{S} \setminus \partial\mathcal{S}$ of distinguished fixed points is given. Then, another homeomorphism $f : \mathcal{S} \to \mathcal{S}$ exists such that $g \simeq f \ \mathrm{rel}\, P$ and:

1) either f is *periodic*, meaning that $f^m = \mathrm{id}_{\mathcal{S}}$ for some $m \in \mathbb{N}$;

2) or f is pseudo-Anosov relative to P;

3) or there exists a f-invariant system of simple loops $\{\gamma_1, \dots, \gamma_r\}$ — called *reducing curves* — with the following properties:

- each connected component of the punctured surface obtained upon deleting P from $\mathcal{S} \setminus \bigcup_{i=1}^r \gamma_i$ has negative Euler characteristic, and
- each reducing curve comes equipped with a f-invariant tubular neighbourhood $\mathcal{U}(\gamma_i)$, disjoint from P, such that if $\mathcal{N}$ is a connected component of $\mathcal{S} \setminus \bigcup_{i=1}^r \mathcal{U}(\gamma_i)$, then $f\restriction_{\mathcal{N}}$ is either periodic or pseudo-Anosov relative to $P \cap \mathcal{N}$.

A few comments are in order concerning the above theorem. First, the requirement of P being composed of fixed points is artificial: in principle it could be any finite g-invariant set, to the effect that powers of f would figure in **3)** instead. Second, the notation $g \simeq f \ \mathrm{rel}\, P$ means that g and f are isotopic in the usual sense, say via $(g_t)_{t\in\mathbb{I}}$ such that $g_0 = g$ and $g_1 = f$, but with the further requirement that $g_t\restriction_P = g\restriction_P$ for every t.

Lastly, the classification theorem is to be understood as a sort of recursive algorithm. Once it finishes running and spits out f and $\{\gamma_1, \dots, \gamma_r\}$, the original surface $\mathcal{S}$ is separated by the reducing curves into invariant components — restricted to which f presents regions of either extremely regular (periodic) or extremely chaotic (pseudo-Anosov) behaviour. The tubular neighbourhoods function as “transition” areas, about which nothing can be said in principle. Running the isotopy backwards then yields information about the homeomorphism g which was given as input, thanks to results such as the aforementioned one due to Handel.

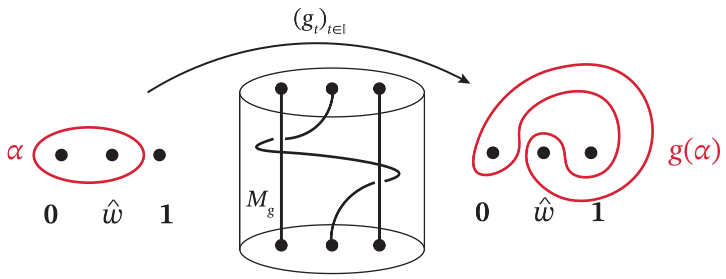
Having that in mind, let $G \subset \mathrm{Homeo}(\mathbb{S}^2)$ be a given subgroup, and assume that $(g_t)_{t\in\mathbb{I}}$ is an $\mathcal{I}G_3$-isotopy under which a certain point $\hat{w}$ has as its trajectory a topological figure 8 relative to $\{\mathbf{0}, \mathbf{1}, \infty\}$. Let us argue informally that this implies G to have an element of positive topological entropy.

To do so, we feed into the classification theorem the homeomorphism $g \overset{\text{def}}{=} g_1$, along with the set of fixed points $P = \{\mathbf{0}, \mathbf{1}, \infty, \hat{w}\}$, thus obtaining f as described there. The first thing we want to do is discard possibility **1)**.

Since we are working within a group, we may take powers to suppose at once that $f = \mathrm{id}_{\mathbb{S}^2}$. Then, concatenation with the isotopy between g and f implies any simple loop α avoiding P to be freely homotopic to its image $g(\alpha)$ relative to P. In particular, this must hold for a simple loop α separating $\{\mathbf{0}, \hat{w}\}$ from $\{\mathbf{1}, \infty\}$, such as the one depicted in the left of Figure B.5. We would like to prove that this cannot hold, and that is exactly the delicate part of the argument.

Intuitively, this can be seen upon unfolding the isotopy $(g_t)_{t\in\mathbb{I}}$ in the mapping torus M_g, as pictured in the centre of Figure B.5. When one tries to slide α onto its image $g(\alpha)$ — which is naively sketched on the right of Figure B.5 — without breaking α open nor crossing the strings, this seems impossible.

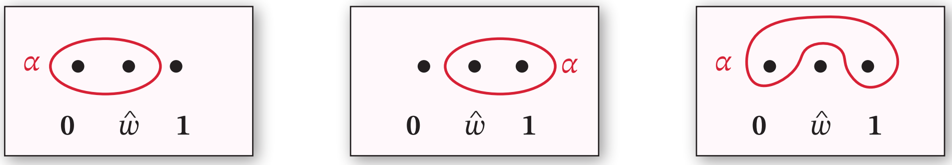
Figure B.5 – An isotopy determines a braid in the mapping torus $M_g \overset{\text{def}}{=} (\mathbb{S}^2 \times \mathbb{I}) \,/\, \{(p,0) \sim (g(p),1)\}$.



Since such arguments are usually misleading and prone to ingenious counterexamples, a rigorous proof is due. One possible approach would indeed be to look at braid types (39). Another one would be to unwrap the curves in the universal cover of the four-punctured sphere and lift the isotopies, along the lines of Bers' proof of the classification theorem. Be as it may, let us move on.

If **2)** holds, we are done. Otherwise, we assume that no simple loop α such as the ones depicted in Figure B.6 — separating $\{\hat{w}, \mathbf{a}\}$ from the remaining two points in $\{\mathbf{0}, \mathbf{1}, \infty\} \setminus \{\mathbf{a}\}$ — is freely homotopic to $g(\alpha)$ relative to P.

Figure B.6 – Possible ways for a simple loop α to separate $\hat{w}$ and another point $\mathbf{a} \in \{\mathbf{0}, \mathbf{1}, \infty\}$ from $\{\mathbf{0}, \mathbf{1}, \infty\} \setminus \{\mathbf{a}\}$ — recall that there is a point at infinity.



If **3)** holds, there can be only one reducing curve γ determining two topological disks within which the points of P lie in pairs, due to the condition on the Euler characteristics. If we thus let $\mathcal{N}$ be the connected component of $\mathbb{S}^2 \setminus \mathcal{U}(\gamma)$ containing $\hat{w}$ and one point $\mathbf{a} \in \{\mathbf{0}, \mathbf{1}, \infty\}$, then $f\restriction_{\mathcal{N}}$ must be a pseudo-Anosov relative to $\{\hat{w}, \mathbf{a}\}$, by the same argument that we used to dismiss case **1)**. Also, the same must hold in the other component, by symmetry. The classification theorem then implies Theorem B as it is stated in the Outline, bringing this essay to its tombstone.

The care that must be taken here is that a simple loop α enclosing $\{\hat{w}, \mathbf{a}\}$ in a component of $\mathbb{S}^2 \setminus \mathcal{U}(\gamma)$ definitely must *not* look at all like any of the ones in Figure B.6! Indeed, the reducing curve γ — and consequently the interior of the disk it bounds — typically may wind quite badly around the surface: we refer to Figure 1 in (11).

We close by remarking that an alternative and more direct proof of the positive entropy part of Theorem B is very likely possible using the forcing techniques recently introduced by Le Calvez and Tal (32), a gain in clarity at the expense of a more explicit (yet not particularly useful) description of the involved maps' "anatomy".

Referências bibliográficas

1 ALTMANN, Simon L. **Rotations, Quaternions, and Double Groups**. Republication of the Oxford, 1986 edition. Dover Publications, 2005. ISBN 9780486445182.

2 BANYAGA, Augustin. **The Structure of Classical Diffeomorphism Groups**. 1. ed.: Springer, 1997. (Mathematics and Its Applications). ISBN 978-1-4419-4774-1. DOI: https://doi.org/10.1007/978-1-4757-6800-8.

3 BARREIRA, Luis; VALLS, Claudia. **Dynamical Systems**: An Introduction. Springer-Verlag London, 2013. cap. 5-6. DOI: 10.1007/978-1-4471-4835-7.

4 BÉGUIN, François; LE ROUX, Frédéric. **DYNAMIQUE TOPOLOGIQUE SUR LES SURFACES**. 2007. Disponível em: <https://www.math.univ-paris13.fr/~beguin/Enseignement_files/Cours_1.pdf>. Acesso em: 5 jun. 2022.

5 BELLIART, Michel. Actions sans points fixes sur les surfaces compactes. **Mathematische Zeitschrift**, Springer Science e Business Media LLC, v. 225, n. 3, p. 453–465, jul. 1997. DOI: 10.1007/pl00004317.

6 BESTVINA, Mladen. **Questions in Geometric Group Theory**. Jul. 2004. Disponível em: <https://www.math.utah.edu/~bestvina/eprints/questions-updated.pdf>. Acesso em: 24 mai. 2022.

7 BRANNAN, David A.; ESPLEN, Matthew F.; GRAY, Jeremy J. **Geometry**. 2. ed.: Cambridge University Press, 2011. ISBN 9781107647831.

8 BRIN, Michael; STUCK, Garrett J. **Introduction to Dynamical Systems**. Cambridge University Press, 2002. ISBN 0-521-80841-3. DOI: https://doi.org/10.1017/CBO9780511755316.

9 DOLD, Albrecht. **Lecture on Algebraic Topology**. Springer-Verlag Berlin Heidelberg, 1995. (Classics in Mathematics). ISBN 978-3-540-58660-9. DOI: 10.1007/978-3-642-67821-9.

10 FALCONER, Kenneth John. **The Geometry of Fractal Sets**. Cambridge University Press, 1985. cap. 3. (Cambridge Tracts in Mathematics, 85). DOI: https://doi.org/10.1017/CBO9780511623738.

11 FARB, Benson; MARGALIT, Dan. **A Primer on Mapping Class Groups**. Princeton Univeristy Press, 2011. (Princeton Mathematical Series). ISBN 978-0-691-14794-9.

12 FATHI, Albert. Structure of the group of homeomorphisms preserving a good measure on a compact manifold. **Annales scientifiques de l'École normale supérieure**, Societe Mathematique de France, T. 13, n. 1, p. 45–93, 1980. DOI: 10.24033/asens.1377.

13 FISHER, David. **Groups acting on manifolds: around the Zimmer program**. ArXiv [math.DS], 5 dez. 2008. DOI: https://doi.org/10.48550/arXiv.0809.4849. Disponível em: <https://arxiv.org/abs/0809.4849v2>. Acesso em: 19 mai. 2022.

14 ________. Recent Developments in the Zimmer Program. **Notices of the American Mathematical Society**, American Mathematical Society (AMS), v. 67, n. 04, p. 492–499, abr. 2020. ISSN 1088-9477. DOI: 10.1090/noti2058.

15 FRANKS, John; MISIUREWICZ, Michal. Topological methods in dynamics. In: HANDBOOK of Dynamical Systems. Elsevier, 2002. cap. 7, p. 547–598. DOI: 10.1016/s1874-575x(02)80009-1.

16 GABAI, David. Convergence Groups are Fuchsian Groups. **The Annals of Mathematics**, Mathematics Department, Princeton University, v. 136, n. 3, p. 447–510, 1992. DOI: https://doi.org/10.2307/2946597.

17 GAULD, David. **Metrisability of Manifolds**. ArXiv [math.GN], 5 out. 2009. DOI: 10.48550/arXiv.0910.0885. Disponível em: <https://arxiv.org/abs/0910.0885>. Acesso em: 25 mai. 2022.

18 GHYS, Étienne. Groups acting on the circle. **L'Enseignement Mathématique**, T. 47, Fascicule 1-2, p. 329–407, 2001. DOI: 10.5169/seals-65441.

19 GIBLIN, James; MARKOVIC, Vladimir. Classification of continuously transitive circle groups. **Geometry and Topology**, v. 10, n. 3, p. 1319–1346, 18 set. 2006. DOI: 10.2140/gt.2006.10.1319.

20 GLEASON, Andrew M. Groups Without Small Subgroups. **Annals of Mathematics**, Mathematics Department, Princeton University, v. 56, n. 2, p. 193–212, 1952. DOI: https://doi.org/10.2307/1969795.

21 HANDEL, Michael. Global shadowing of pseudo-Anosov homeomorphisms. **Ergodic Theory and Dynamical Systems**, Cambridge University Press, v. 5, n. 3, p. 373–377, 1985. DOI: 10.1017/s0143385700003011.

22 HATCHER, Allen. **Algebraic Topology**. Cambridge University Press, 2002. ISBN 978-0-521-79540-1. Disponível em: <http://pi.math.cornell.edu/~hatcher/AT/AT.pdf>. Acesso em: 21 abr. 2022.

23 HINKKANEN, Aimo. Abelian and nondiscrete convergence groups on the circle. **Transactions of the American Mathematical Society**, American Mathematical Society (AMS), v. 318, n. 1, p. 87–121, 1990. DOI: https://doi.org/10.1090/S0002-9947-1990-1000145-X.

24 HIRSCH, Morris W. **Differential Topology**. 1. ed. New York: Springer, 1976. DOI: https://doi.org/10.1007/978-1-4684-9449-5.

25 JOST, Jürgen. **Compact Riemann Surfaces**: An Introduction to Contemporary Mathematics. 3. ed.: Springer Berlin, Heidelberg, 2006. (Universitext). ISBN 978-3-540-33065-3. DOI: https://doi.org/10.1007/978-3-540-33067-7.

26 KERÉKJÁRTÓ, Béla. Sur les groupes compacts de transformations topologiques des surfaces. **Acta Mathematica**, International Press of Boston, v. 74, n. 0, p. 129–173, 1941. DOI: 10.1007/bf02392252.

27 KOLEV, Boris. Sous-groupes compacts d'homéomorphismes de la sphère. **L'Enseignement Mathématique**, Fascicule 3-4, p. 193–214, 2006. DOI: 10.5169/SEALS-2231.

28 KOROPECKI, Andres. Realizing rotation numbers on annular continua. **Mathematische Zeitschrift**, Springer Science e Business Media LLC, v. 285, n. 1-2, p. 549–564, 2016. DOI: 10.1007/s00209-016-1720-z.

29 KRECK, Matthias. **Orientation of manifolds**. Bulletin of the Manifold Atlas, 2013. Disponível em: <http://www.boma.mpim-bonn.mpg.de/data/47screen.pdf>. Acesso em: 25 mai. 2022.

30 KWAKKEL, Ferry; TAL, Fábio. **Homogeneous Transformation Groups of the Sphere**. ArXiv [math.GT], 14 set. 2014. DOI: 10.48550/ARXIV.1309.0179. Disponível em: <https://arxiv.org/abs/1309.0179v2>. Acesso em: 24 mai. 2022.

31 LABOURIE, François. Large groups actions on manifolds. In: PROCEEDINGS OF THE INTERNATIONAL CONGRESS OF MATHEMATICIANS, 1998. v. 2.

32 LE CALVEZ, Patrice; TAL, Fábio A. Forcing theory for transverse trajectories of surface homeomorphisms. **Inventiones mathematicae**, Springer Science e Business Media LLC, v. 212, n. 2, p. 619–729, 2017. DOI: 10.1007/s00222-017-0773-x.

33 LE ROUX, Frédéric. On closed subgroups of the group of homeomorphisms of a manifold. **Journal de l'École Polytechnique – Mathématiques**, Cellule MathDoc/CEDRAM, t. I, p. 147–159, 2014. DOI: 10.5802/jep.7.

34 LEE, John M. **Introduction to Riemannian Manifolds**. 2. ed.: Springer International Publishing, 2018. (Graduate Texts in Mathematics). ISBN 978-3-319-91755-9. DOI: 10.1007/978-3-319-91755-9.

35 ________. **Introduction to Smooth Manifolds**. 2. ed.: Springer New York, 2012. (Graduate Texts in Mathematics). ISBN 978-1-4419-9981-8. DOI: 10.1007/978-1-4419-9982-5.

36 ________. **Introduction to Topological Manifolds**. 2. ed.: Springer, 2011. ISBN 978-1-4419-7940-7. DOI: 10.1007/978-1-4419-7940-7.

37 LIMA, Elon Lages. **Grupo fundamental e espaços de recobrimento**. 4. ed.: IMPA, 2012. P. 4. (Projeto Euclides). ISBN 978-8-5244-0086-5.

38 LLIBRE, Jaume; MACKAY, Robert S. A Classification of Braid Types for Diffeomorphisms of Surfaces of Genus Zero with Topological Entropy Zero. **Journal of the London Mathematical Society**, Wiley, s2-42, Issue 3, p. 562–576, 1990. DOI: 10.1112/jlms/s2-42.3.562.

39 MATSUOKA, Takashi. Periodic Points and Braid Theory. In: HANDBOOK of Topological Fixed Point Theory. Springer-Verlag, 2005. P. 171–216. DOI: 10.1007/1-4020-3222-6_5.

40 MUNKRES, James. **Topology**: Pearson New International Edition. 2. ed.: Pearson, 2013. ISBN 978-1-292-02362-5.

41 NAVAS, Andrés. **Groups of circle diffeomorphisms**. The University of Chicago Press, 2011. (Chicago Lectures in Mathematics). ISBN 978-0-226-56951-2.

42 NEEDHAM, Tristan. **Visual Complex Analysis**. Oxford University Press, 1999. ISBN 978-0198534464.

43 OXTOBY, John C.; ULAM, Stanisław M. Measure-Preserving Homeomorphisms and Metrical Transitivity. **Annals of Mathematics**, Mathematics Department, Princeton University, v. 42, n. 4, p. 874–920, 1941. DOI: 10.2307/1968772.

44 SIEBENMANN, Laurent Carl. Errata to "The Osgood-Schoenflies theorem revisited". **Russian Mathematical Surveys**, IOP Publishing, v. 60, n. 5, p. 1001–1001, 2005. DOI: 10.1070/rm2005v060n05abeh004295.

45 ________. The Osgood-Schoenflies theorem revisited. **Russian Mathematical Surveys**, IOP Publishing, v. 60, n. 4, p. 645–672, 2005. DOI: 10.1070/rm2005v060n04abeh003672.

46 THOMASSEN, Carsten. The Jordan-Schönflies Theorem and the Classification of Surfaces. **The American Mathematical Monthly**, Taylor & Francis, v. 99, n. 2, p. 116–131, 1992. DOI: https://doi.org/10.2307/2324180.

47 VAN DEN DRIES, Lou; GOLDBRING, Isaac. Hilbert's 5th Problem. **L'Enseignement Mathématique**, T. 61, Fascicule 1-2, p. 3–43, 2015. DOI: https://10.4171/LEM/61-1/2-2.

48 VVENDENSKY, Dimitri; EVANS, Timothy. **Symmetry, Groups, and Representations in Physics**. World Scientific, 2011. (Imperial College Press Advanced Physics Texts). ISBN 9781848163713.