Expanding Thurston Maps

December 30, 2016

Mario Bonk

Daniel Meyer

Department of Mathematics, University of California, Los Angeles, CA 90095, USA
E-mail address: mbonk@math.ucla.edu

Department of Mathematics and Statistics, P.O.Box 35, FI-40014 University of Jyväskylä, Finland
E-mail address: dmeyermail@gmail.com

Contents

List of Figures

Preface

This book is the result of an intended research paper that grew out of control. A preprint containing a substantial part of our investigations was already published on arXiv in 2010. To make the content more accessible, we decided to include additional material. This more than doubled the size of this work compared to the 2010 version and caused a long delay until its completion.

More than fifteen years ago we became independently interested in some basic problems on quasisymmetric parametrization of 2-spheres. This problem is related to the dynamics of rational maps; it seems that Rick Kenyon was the first to make this observation. During our common time at the University of Michigan we decided to join forces and to investigate this connection systematically.

We realized that for the relevant rational maps an explicit analytic expression is not so important, but rather a geometric-combinatorial description. As this became our preferred way of looking at these objects, it was a natural step to consider a more general class of maps that are not necessarily holomorphic. The relevant properties can be condensed into the notion of an *expanding Thurston map* which is the topic of this book. We will discuss the underlying ideas more thoroughly in the Introduction (Chapter 1).

Part of this work has overlap with studies by other researchers, notably Haïssinsky-Pilgrim [**HP09**], and Cannon-Floyd-Parry [**CFP07**]. We would like to clarify some of the interrelations of our investigations with these works. Theorem 15.1 (in the body of the text) was announced by the first author during an Invited Address at the AMS Meeting at Athens, Ohio, in March 2004, where he gave a short outline of the proof. After the talk he was informed by Bill Floyd and Walter Parry that related results had been independently obtained by Cannon-Floyd-Parry (which later appeared as [**CFP07**]).

Theorem 18.1 was already published by Haïssinsky-Pilgrim as part of a more general statement [**HP09**, Theorem 4.2.11]. Special cases go back to work by the second author [**Me02**] and unpublished joint work by Bruce Kleiner and the first author. The current more general

version seems to have emerged after a visit of the first author at the University of Indiana at Bloomington in February 2003.

During this visit the first author explained concepts of quasiconformal geometry to Kevin Pilgrim and his joint work with Bruce Kleiner on Cannon's conjecture in geometric group theory. Kevin Pilgrim in turn pointed out Theorem 11.1 and the ideas for its proof to the first author. After this visit versions of Theorem 18.1 with an outline for the proof were found independently by Kevin Pilgrim and the first author. A proof of Theorem 18.1 was discovered soon afterwards by the authors using ideas from [**Me02**] (see [**Me10**] for an argument along similar lines) in combination with Theorem 15.1.

We are indebted to many people. Conversations with Bruce Kleiner, Peter Haïssinsky, and Kevin Pilgrim have been especially fruitful. We would also like to thank Jim Cannon, Bill Floyd, Lukas Geyer, Misha Hlushchanka, Zhiqiang Li, Dimitrios Ntalampekos, Walter Parry, Juan Souto, Dennis Sullivan, and Mike Zieve for various useful comments. We are indebted to the two anonymous referees who provided us with valuable feedback on a draft of our work.

Qian Yin was so kind to let us incorporate parts of her thesis. We are grateful to Jana Kleineberg for careful proofreading and helping us with some of the pictures. We are also happy to acknowledge the patient support of our editors from the American Mathematical Society, Ed Dunne and Ina Mette.

Notation

Here we summarize some of the most important notation used in this book for easy reference.

When an object A is defined to be another object B, we write $A := B$ for emphasis.

We denote by $\mathbb{N} = \{1, 2, \dots\}$ the set of natural numbers and by $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ the set of natural numbers including 0. We write $\mathbb{Z}$ for the set of integers, and $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ for the set of rational, real, and complex numbers, respectively. For $k \in \mathbb{N}$, we let $\mathbb{Z}_k = \mathbb{Z}/k\mathbb{Z}$ be the cyclic group of order k.

We also consider $\widehat{\mathbb{N}} := \mathbb{N} \cup \{\infty\}$. Given $a, b \in \widehat{\mathbb{N}}$ we write $a | b$ if a divides b. This notation is extended to $\widehat{\mathbb{N}}$-valued functions (see Section 2.5).

The *floor* of a real number x, denoted by $\lfloor x \rfloor$, is the largest integer $m \in \mathbb{Z}$ with $m \leq x$. The *ceiling* of a real number x, denoted by $\lceil x \rceil$, is the smallest integer $m \in \mathbb{Z}$ with $x \leq m$.

The symbol $\boldsymbol{i}$ stands for the imaginary unit in the complex plane $\mathbb{C}$. The real and imaginary part of a complex number z are indicated by $\mathrm{Re}(z)$ and $\mathrm{Im}(z)$, respectively, and its complex conjugate by $\overline{z}$. We use the notation $\mathbb{D} := \{z \in \mathbb{C} : |z| < 1\}$ for the open unit disk in $\mathbb{C}$, and $\mathbb{H} := \{z \in \mathbb{C} : \mathrm{Re}(z) > 0\}$ for the open upper half-plane.

We let $\widehat{\mathbb{C}} := \mathbb{C} \cup \{\infty\}$ be the Riemann sphere. It carries the *chordal metric* σ given by (A.3). Similarly, we let $\widehat{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$. Here we consider $\widehat{\mathbb{R}}$ as a subset of $\widehat{\mathbb{C}}$, and so $\widehat{\mathbb{R}} \subset \widehat{\mathbb{C}}$.

The *Lebesgue measure* on $\mathbb{R}^2$, $\mathbb{C}$, $\widehat{\mathbb{C}}$, or $\mathbb{D}$ is denoted by $\mathcal{L}$. If necessary, we add a subscript here to avoid ambiguities. More precisely, $\mathcal{L} = \mathcal{L}_{\mathbb{R}^2}$ and $\mathcal{L} = \mathcal{L}_{\mathbb{C}}$ are the Euclidean area measures on $\mathbb{R}^2$ and $\mathbb{C}$, $\mathcal{L} = \mathcal{L}_{\widehat{\mathbb{C}}}$ is the spherical area measure on $\widehat{\mathbb{C}}$, and $\mathcal{L} = \mathcal{L}_{\mathbb{D}}$ the hyperbolic area measure on $\mathbb{D}$ considered as the hyperbolic plane.

When we consider two objects A and B and there is a natural identification between them that is clear from the context, we write $A \cong B$. For example, $\mathbb{R}^2 \cong \mathbb{C}$ if we identify a point $(x, y) \in \mathbb{R}^2$ with $x + \boldsymbol{i}y \in \mathbb{C}$.

The derivative of a holomorphic function f is denoted by f' as usual. If $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is a rational map, then $f^{\sharp}$ stands for its *spherical derivative* (see (A.4)). For a differentiable (not necessarily holomorphic) map, we use Df to denote its derivative considered as a linear map between suitable tangent spaces. If these tangent spaces are equipped with norms, then we let $\|Df\|$ be the operator norm of Df. Sometimes we use subscripts here to indicate the norms.

Two non-negative quantities a and b are said to be *comparable* if there is a constant $C \geq 1$ depending on some obvious ambient parameters such that

$$\frac{1}{C}a \leq b \leq Ca.$$

We then write $a \asymp b$. The constant C is referred to as $C(\asymp)$. Similarly, we write $a \lesssim b$ or $b \gtrsim a$, if there is a constant $C > 0$ such that $a \leq Cb$, and refer to the constant C as $C(\lesssim)$ or $C(\gtrsim)$.

The cardinality of a set X is denoted by $\#X$ and the identity map on X by id_X. If $x_n \in X$ for $n \in \mathbb{N}$ are points in X, we denote the sequence of these points by $\{x_n\}_{n \in \mathbb{N}}$, or just by $\{x_n\}$ if the index set $\mathbb{N}$ is understood.

If $f \colon X \to X$ is a map and $n \in \mathbb{N}$, then

$$f^n := \underbrace{f \circ \cdots \circ f}_{n\text{-times}}$$

is the n-th iterate of f. We set $f^0 := \mathrm{id}_X$.

Let $f \colon X \to Y$ be a map between sets X and Y. If $U \subset X$, then $f|U$ stands for the *restriction* of f to U. If $A \subset Y$, then $f^{-1}(A) := \{x \in X : f(x) \in A\}$ is the preimage of A in X. Similarly, $f^{-1}(y) := \{x \in X : f(x) = y\}$ is the preimage of a point $y \in Y$.

If $f \colon X \to X$ is a map, then preimages of a set $A \subset X$ or a point $p \in X$ under the n-th iterate f^n are denoted by $f^{-n}(A) := \{x \in X : f^n(x) \in A\}$ and $f^{-n}(p) := \{x \in X : f^n(x) = p\}$, respectively.

Let (X, d) be a metric space, $a \in X$, and $r > 0$. We denote by $B_d(a, r) = \{x \in X : d(a, x) < r\}$ the open and by $\overline{B}_d(a, r) = \{x \in X : d(a, x) \leq r\}$ the closed ball of radius r centered at a. If $A, B \subset X$, we let $\mathrm{diam}_d(A)$ be the diameter, $\overline{A}$ be the closure of A in X, and

$$\mathrm{dist}_d(A, B) := \inf\{d(x, y) : x \in A, y \in B\}$$

be the distance of A and B. If $p \in X$, we let $\mathrm{dist}_d(p, A) := \mathrm{dist}_d(\{p\}, A)$. For $\epsilon > 0$,

$$\mathcal{N}_{d,\epsilon}(A) := \{x \in X : \mathrm{dist}_d(x, A) < \epsilon\}$$

is the *open ϵ-neighborhood* of A with respect to d. If $\gamma \colon [0, 1] \to X$ is a path, we denote by $\mathrm{length}_d(\gamma)$ the length of γ. For $Q \geq 0$, we denote by $\mathcal{H}_d^Q$ the *Q-dimensional Hausdorff measure* on X with respect to d. We drop the subscript d in our notation for $B_d(a, r)$, etc., if the metric d is clear from the context. For the Euclidean metric on $\mathbb{C}$ we sometimes use the subscript $\mathbb{C}$ for emphasis. So, for example,

$$B_{\mathbb{C}}(a, r) := \{z \in \mathbb{C} : |z - a| < r\}$$

denotes the Euclidean (metric) ball of radius $r > 0$ centered at $a \in \mathbb{C}$.

The *Gromov product* of two points $x, y \in X$ with respect to a basepoint $p \in X$ in a metric space X is denoted by $(x \cdot y)_p$ or by $(x \cdot y)$ if the basepoint p is understood (see Section 4.2). The *boundary at infinity* of a Gromov hyperbolic space X is represented by $\partial_\infty X$. If a group G acts on a space X, then we write $G \curvearrowright X$ to indicate this action.

Often we use the notation $I := [0, 1]$. If X and Y are topological spaces, then a *homotopy* is a continuous map $H \colon X \times I \to Y$. For $t \in I$, we let $H_t(\cdot) := H(\cdot, t)$ be the *time-t map* of the homotopy.

The symbol S^2 indicates a 2-sphere, which we think of as a topological object. Similarly, T^2 is a topological 2-torus. For a 2-torus with a Riemann surface structure we write $\mathbb{T}$ (see Section A.8).

Often S^2 (or the Riemann sphere $\widehat{\mathbb{C}}$) is equipped with certain metrics that induce its topology. The *visual metric* induced by an expanding Thurston map f is usually denoted by ϱ (see Chapter 8). The *canonical orbifold metric* for a rational Thurston map is indicated by ω (see Section A.10).

The (topological) *degree* of a branched covering map f between surfaces is denoted by $\deg(f)$ and the *local degree* of f at a point z by $\deg_f(x)$ or $\deg(f, x)$ (see Section 2.1). We write $\operatorname{crit}(f)$ for the *set of critical points* of a branched covering map (see Section 2.1), and $\operatorname{post}(f)$ for the set of *postcritical points* of a Thurston map f (see Section 2.2).

The *ramification function* of a Thurston map f is denoted by α_f (see Definition 2.7), and the *orbifold* associated with f by $\mathcal{O}_f$ (see Definition 2.10).

For a given Thurston map $f \colon S^2 \to S^2$ we usually use the symbol $\mathcal{C}$ to indicate a Jordan curve $\mathcal{C} \subset S^2$ that satisfies $\operatorname{post}(f) \subset \mathcal{C}$.

When we consider objects that are defined in terms of the n-th iterate of a given Thurston map, then we often use the upper index "n" to emphasize this.

For a topological cell c in a topological space $\mathcal{X}$ we denote by ∂c the *boundary* of c, and by $\operatorname{int}(c)$ the *interior* of c (see Section 5.1). Note that ∂c and $\operatorname{int}(c)$ usually do not agree with the boundary or interior of c as a subset of $\mathcal{X}$.

Cell decompositions of a space $\mathcal{X}$ are usually denoted by $\mathcal{D}$ (see Chapter 5). Let $n \in \mathbb{N}_0$, $f \colon S^2 \to S^2$ be a Thurston map, and $\mathcal{C} \subset S^2$ a Jordan curve with $\operatorname{post}(f) \subset \mathcal{C}$. We then write $\mathcal{D}^n(f, \mathcal{C})$ for the cell decomposition of S^2 consisting of the *cells of level n* or *n-cells* defined in terms of f and $\mathcal{C}$ (see Definition 5.14). The set of corresponding *n-tiles* is denoted by $\mathbf{X}^n$, the set of *n-edges* by $\mathbf{E}^n$, and the set of *n-vertices* by $\mathbf{V}^n$ (see Section 5.3).

In this context we often "color" tiles "black" or "white". We then use the subscripts $\mathtt{b}$ and $\mathtt{w}$ to indicate the color (see the end of Section 5.3). For example, the black and white 0-tiles are denoted by $X_{\mathtt{b}}^0$ and $X_{\mathtt{w}}^0$, respectively.

The n-*flower* of an n-vertex v is denoted by $W^n(v)$ (see Section 5.6). The number $D_n = D_n(f, \mathcal{C})$ is the minimal number of n-tiles required to join opposite sides (see (5.14)).

The number $m(x, y) = m_{f, \mathcal{C}}(x, y)$ is defined in Definition 8.1. The *expansion factor* of a visual metric is usually denoted by Λ (see Definition 8.2).

We write $\Lambda_0(f)$ for the *combinatorial expansion factor* of a Thurston map f (see Proposition 16.1).

The *topological entropy* of a map f is denoted by $h_{top}(f)$, and the *measure-theoretic entropy* of f with respect to a measure μ by $h_\mu(f)$. The *measure of maximal entropy* of an expanding Thurston map f is indicated by ν_f. See Chapter 17 for these concepts.

For a rational Thurston map $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ we write Ω_f for its *canonical orbifold measure* (see Section A.10) and λ_f for the unique probability measure on $\widehat{\mathbb{C}}$ that is absolutely continuous with respect to Lebesgue measure (see Chapter 19).

CHAPTER 1

Introduction

In this work we study the dynamics of Thurston maps under iteration. A Thurston map is a branched covering map on a 2-sphere S^2 such that each of its critical points has a finite orbit. The most important examples are given by postcritically-finite rational maps on the Riemann sphere $\widehat{\mathbb{C}}$. Most of the time we will also assume that a Thurston map is expanding in a suitable sense. For postcritically-finite rational maps f on $\widehat{\mathbb{C}}$ expansion is equivalent to the requirement that f does not have periodic critical points or that its Julia set is equal to $\widehat{\mathbb{C}}$.

These objects were first considered by Thurston as topological model maps in the context of his celebrated characterization of rational maps (see Theorem 2.17). The terminology was introduced by Douady and Hubbard in their proof of this theorem.

Every expanding Thurston map $f\colon S^2 \to S^2$ gives rise to a type of fractal geometry on the underlying sphere S^2. This geometry is represented by a class of *visual metrics* ϱ that are associated with the map. Many dynamical properties of the map are encoded in the geometry of the corresponding *visual sphere*, meaning S^2 equipped with a visual metric ϱ. For example, we will see that an expanding Thurston map is topologically conjugate to a rational map if and only if (S^2, ϱ) is quasisymmetrically equivalent to $\widehat{\mathbb{C}}$ (see Section 4.1 for the terminology). For us this relation between dynamics and fractal geometry is one of the main motivations for studying expanding Thurston maps.

In order to define a visual metric for a given Thurston map $f\colon S^2 \to S^2$, we will extract some combinatorial data from f. For this we consider a cell decomposition of S^2 and its pull-backs by the iterates f^n. When f is expanding, the diameters of the cells in these decompositions shrink to 0; so we get discrete approximations of S^2 that get finer with larger level n. Given two distinct points in S^2, one can ask at which level the cell decompositions will allow us to distinguish them. Our definition of a visual metric is based on this information.

The visual sphere (S^2, ϱ) of an expanding Thurston map is fractal in the sense that its Hausdorff dimension is typically larger than 2. With

a suitable choice of ϱ, the local behavior of f becomes very simple though. Namely, there is a number $\Lambda > 1$ (the *expansion factor* of ϱ) such that f expands ϱ locally by the factor Λ in a sense that will be made precise. So the local behavior of f on (S^2, ϱ) is simplified at the expense of a more complicated geometry of (S^2, ϱ). This point of view is in contrast to the usual setting for complex dynamics, where one studies the action of a rational map on a *smooth* underlying space, namely the Riemann sphere $\widehat{\mathbb{C}}$, considered as a Riemann surface.

It is possible to construct a graph $\mathcal{G}$ that combines the combinatorial data of the cell decompositions on all levels generated by an expanding Thurston map and its iterates. This graph $\mathcal{G}$ is Gromov hyperbolic and its boundary at infinity can naturally be identified with the underlying sphere S^2. Under this identification a metric is a visual metric for the given map f according to our definition if and only if it is a visual metric in the sense of Gromov hyperbolic spaces. This fact relates the study of expanding Thurston maps and of Gromov hyperbolic spaces.

There is an intriguing connection of these ideas to *Cannon's conjecture* in geometric group theory. Roughly speaking, this conjecture predicts that a group G that shares the topological properties of the fundamental group of a closed hyperbolic 3-manifold "is" such a fundamental group (see Section 4.3 for precise statements). In this context one assumes that the group G is Gromov hyperbolic and that its boundary at infinity $\partial_\infty G$ is a 2-sphere. Here $\partial_\infty G$ is naturally equipped with a visual metric that provides $\partial_\infty G$ with a fractal geometry. Then Cannon's conjecture is equivalent to showing that the fractal sphere $\partial_\infty G$ is quasisymmetrically equivalent to $\widehat{\mathbb{C}}$.

So for both types of dynamical systems, namely expanding Thurston maps and Gromov hyperbolic groups G with 2-sphere boundary $\partial_\infty G$, we are led to the investigation of a fractal geometry on the underlying 2-sphere. This analogy can be viewed as an example of *Sullivan's dictionary* which exhibits similarities in complex dynamics and the theory of Kleinian groups. Common to both areas is the desire to characterize conformal dynamical systems in a wider class of dynamical systems characterized by suitable metric-topological conditions. One should not push the analogies too far though: while Cannon's conjecture is generally believed to be true and, accordingly, one expects that the fractal 2-spheres arising from Gromov hyperbolic groups are always quasisymmetrically equivalent to $\widehat{\mathbb{C}}$, this is not always the case for Thurston maps, because not every Thurston map is equivalent to a rational map.

After these remarks about some of the motivations for our investigation, we now state some basic definitions more precisely (more details can be found in the body of the text). Let $f\colon S^2 \to S^2$ be an (orientation-preserving) branched covering map. As usual, we call a point $c \in S^2$ a *critical point* of f if near c the map f is not a local homeomorphism. A *postcritical point* is any point obtained as an image of a critical point under forward iteration of f. So if we denote by $\mathrm{crit}(f)$ the set of critical points of f and by f^n the n-th iterate of f, then the set of postcritical points of f is given by

$$\mathrm{post}(f) := \bigcup_{n \geq 1} \{f^n(c) : c \in \mathrm{crit}(f)\}.$$

It is a fundamental fact in complex dynamics that much information on the dynamics can be deduced from the structure of the orbits of critical points. A very strong assumption in this respect is that each such orbit is finite, i.e., that $\mathrm{post}(f)$ is a finite set. In this case the map f is called *postcritically-finite*. A *Thurston map* is a (non-homeomorphic) branched covering map $f\colon S^2 \to S^2$ that is postcritically-finite.

Thurston maps are abundant and include specific *rational Thurston maps* (i.e., postcritically-finite rational maps on $\widehat{\mathbb{C}}$) such as $f(z) = 1 - 2/z^2$ or $f(z) = 1 + (\boldsymbol{i} - 1)/z^4$. More examples can be found in Section 12.3, and a complete list of all examples considered in this book is given in Section 1.9. We will later provide a general method for producing Thurston maps (see Proposition 12.3 in combination with Theorem 14.2); it follows from one of our main results (Theorem 15.1) that at least some iterate of every expanding Thurston map can be obtained from this construction.

We now turn to the discussion of more specific topics in this introductory chapter. Our main purpose is to give some guidance for the intuition of the reader. We will present some examples and discuss the main concepts and results of this work. Full details can be found in subsequent chapters.

1.1. A Lattès map as a first example

A large class of well-understood Thurston maps are Lattès maps. These are rational maps obtained as quotients of holomorphic torus endomorphisms. We will discuss these maps in more detail in Chapter 3; results concerning them will be outlined in Section 1.7. Note that the terminology is not uniform and some authors use the term Lattès map with a slightly different meaning. They were the first known examples of rational maps whose Julia set is the whole sphere.
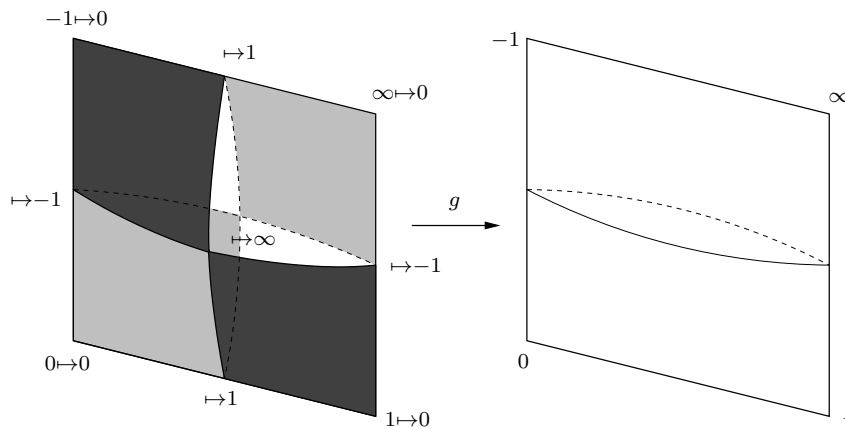
FIGURE 1.1. The Lattès map g.

We will encounter Lattès maps quite often in this book. On the one hand, they are easy to visualize and construct, and thus often serve as convenient examples to illustrate various phenomena. On the other hand, these maps are quite special and arise in many situations as exceptional cases. In order to introduce some of the main themes of this work, we will now consider a specific Lattès map.

The map is essentially given by Figure 1.1. We will explain this picture in detail momentarily, but we will first define the map by a more standard approach. This may be helpful for readers that are already familiar with Lattès maps.

The square $[0, \frac{1}{2}]^2 \subset \mathbb{R}^2 \cong \mathbb{C}$ can be mapped conformally to the upper half-plane in $\widehat{\mathbb{C}}$ such that the vertices $0, \frac{1}{2}, \frac{1}{2} + \frac{i}{2}, \frac{i}{2}$ of the square are mapped to the points $0, 1, \infty, -1$, respectively. Note that here and in the following a "conformal map" is always bijective. By Schwarz reflection we can extend this to a holomorphic map $\Theta \colon \mathbb{C} \to \widehat{\mathbb{C}}$. Up to postcomposition by a Möbius transformation, this map is a classical *Weierstraß $\wp$-function*; it is doubly-periodic with respect to the lattice $\Gamma := \mathbb{Z} \oplus \mathbb{Z}i$ and induces a double branched covering map of the torus $\mathbb{T} := \mathbb{C}/\Gamma$ to the sphere $\widehat{\mathbb{C}}$.

Consider the map

$$A \colon \mathbb{C} \to \mathbb{C}, \quad u \mapsto A(u) := 2u.$$

From the properties of the $\wp$-function or directly from the definition of Θ by the reflection process, one can see that $\Theta(v) = \Theta(u)$ for $u, v \in \mathbb{C}$ if and only if $v = \pm u + \gamma$ with $\gamma \in \Gamma$. In this case, $\Theta(2v) = \Theta(2u)$. This implies that there is a well-defined and unique holomorphic map

$g \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ such that the diagram

(1.1)
$$
\begin{array}{ccc}
\mathbb{C} & \xrightarrow{\;A\;} & \mathbb{C} \\
{\scriptstyle\Theta}\big\downarrow & & \big\downarrow{\scriptstyle\Theta} \\
\widehat{\mathbb{C}} & \xrightarrow{\;g\;} & \widehat{\mathbb{C}}
\end{array}
$$

commutes. The map g obtained in this way is a *Lattès map*. It is a rational map given by

$$
g(z) = 4\frac{z(1 - z^2)}{(1 + z^2)^2} \quad \text{for } z \in \widehat{\mathbb{C}}.
$$

One can show that the Julia set of g is the whole sphere.

More relevant for us than this explicit formula for g is that one can describe g geometrically as indicated in Figure 1.1. To explain this, note that there is an essentially unique path metric on $\widehat{\mathbb{C}}$ obtained as a "push-forward" of the Euclidean metric on $\mathbb{C}$ by the map Θ. This metric is in fact the *canonical orbifold metric* of g (see Section A.10 and Section 2.5).

Geometrically, the sphere equipped with this metric looks like a pillow. In general, a *pillow* (see Section A.10) is a metric space P obtained from gluing two identical copies $X_\mathtt{w}$ and $X_\mathtt{b}$ of a (simple and compact) Euclidean polygon $X \subset \mathbb{C}$ together along their boundaries. The pillow is equipped with the induced path metric. Under the given identification, $\partial X_\mathtt{w} \cong \partial X_\mathtt{b}$ is a Jordan curve in the pillow P called its *equator*.

In our case, the upper and lower half-planes in $\widehat{\mathbb{C}}$ equipped with the canonical orbifold metric are isometric to copies of the square $S = [0, \frac{1}{2}]^2$. If we glue two copies of S together along their boundaries, then we obtain the pillow P. We color one of these squares, say the one corresponding to the upper half-plane, white, and the other square black.

The square $S = [0, 1/2]^2 \subset \mathbb{R}^2 \cong \mathbb{C}$ (and each of its translates by $\frac{1}{2}(m + n\boldsymbol{i})$ where $m, n \in \mathbb{Z}$) can be subdivided into four squares of side length $1/4$. If S' is such a square, then $A(S')$ is a square of side length $1/2$ that is mapped by Θ to either the upper or the lower half-plane, meaning to either the black or the white face of the pillow P. It follows from (1.1) that g has a very similar mapping behavior on P.

More precisely, we divide each of the two sides of P (each of the two isometric copies of S contained in P) into four smaller squares of half the side length, and color the eight small squares in a checkerboard fashion black and white. If we map one such small white square to the large white square by a Euclidean similarity (that scales by the factor

2), then this map extends by reflection to the whole pillow. There are obviously many different ways to color and map the small squares. If we do this in an appropriate way as indicated in Figure 1.1, then we obtain the map g.

The vertices where four small squares intersect are the critical points of g. They are mapped by g to the set $\{1, \infty, -1\}$, which in turn is mapped to $\{0\}$. The point 0 is a fixed point of g. So g is a postcritically-finite branched covering map on the 2-sphere P with $\mathrm{post}(g) = \{0, 1, \infty, -1\}$, and hence a Thurston map. The postcritical points of g are the vertices of the pillow, which are the conical singularities of our canonical orbifold metric. The extended real line $\mathcal{C} := \widehat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ (corresponding to the equator of the pillow) is a Jordan curve that is invariant under g in the sense that $g(\mathcal{C}) \subset \mathcal{C}$ and contains the set $\{0, 1, \infty, -1\}$ of postcritical points of g. The set $g^{-1}(\mathcal{C})$ is an embedded graph in the pillow consisting of all sides of the small squares on the left hand side of Figure 1.1 as edges and the points in $g^{-1}(\mathrm{post}(g))$, i.e., the corners of these squares, as vertices. This graph $g^{-1}(\mathcal{C})$ determines the tiling in this picture.

The set $g^{-2}(\mathcal{C})$ is obtained by pulling $g^{-1}(\mathcal{C})$ back by the map g. Since g restricted to any small square S' is a homeomorphism onto one of the two large squares S forming the pillow, in this process S' is subdivided in the same way as S was subdivided by the small squares of side length $1/4$ (i.e., S' is subdivided into 4 squares). It follows that $g^{-2}(\mathcal{C})$ subdivides the pillow into $4 \times 8 = 32$ squares of side length $1/8$. Proceeding in this way inductively, we see that the preimage $g^{-n}(\mathcal{C})$ of $\mathcal{C}$ under the iterate g^n subdivides the pillow into $2 \cdot 4^n$ squares of side length 2^{-n-1} for $n \in \mathbb{N}$.

The complementary components of $g^{-n}(\mathcal{C})$ are the interiors of these squares. In particular, the diameters of these components tend to 0 uniformly as $n \to \infty$. This fact will be the basis of our definition of an *expanding* Thurston map. Accordingly, g is such a map.

For each $n \in \mathbb{N}_0$ the set $g^{-n}(\mathcal{C})$ forms an embedded graph in the pillow P with the points in $g^{-n}(\mathrm{post}(g))$ as vertices. This is also meaningful for $n = 0$, if we interpret g^0 as the identity map on the pillow P. Then this graph is just the Jordan curve $\mathcal{C}$ with the points in $\mathrm{post}(g)$ as vertices.

The graph $g^{-n}(\mathcal{C})$ is the 1-dimensional skeleton or 1-skeleton of a *cell decomposition* $\mathcal{D}^n = \mathcal{D}^n(g, \mathcal{C})$ of the pillow P generated by g and $\mathcal{C}$ (see Chapter 5 for the terminology that we use here and below). The 2-dimensional cells or *tiles* of the cell decomposition $\mathcal{D}^n$ are squares of side length 2^{-n-1} and are given by the closures of the complementary

components of $g^{-n}(\mathcal{C})$ in P. The map g sends each cell in $\mathcal{D}^{n+1}$ home-omorphically to a cell in $\mathcal{D}^n$ (for all $n \in \mathbb{N}_0$); so g is *cellular* for each pair $(\mathcal{D}^{n+1}, \mathcal{D}^n)$ of cell decompositions.

Since $\mathcal{C}$ is g-invariant in the sense that $g(\mathcal{C}) \subset \mathcal{C}$, we have $g^{-n}(\mathcal{C}) \subset g^{-(n+1)}(\mathcal{C})$ for each $n \in \mathbb{N}_0$. This inclusion for 1-skeleta implies that the cell decomposition $\mathcal{D}^{n+1}$ is a *refinement* of $\mathcal{D}^n$. On a more intuitive level, this means that the tiles in $\mathcal{D}^n$ are subdivided by the tiles in $\mathcal{D}^{n+1}$.

The tiles in $\mathcal{D}^0$ are the two initial squares of side length $1/2$ forming the pillow, and $\mathcal{D}^1$ is formed by squares of side length $1/4$ subdividing these squares. Since we repeat the same subdivision procedure in the passage from $\mathcal{D}^n$ to $\mathcal{D}^{n+1}$, this whole sequence of cell decompositions $\mathcal{D}^n$ is essentially generated by the initial pair $(\mathcal{D}^1, \mathcal{D}^0)$. This pair $(\mathcal{D}^1, \mathcal{D}^0)$ is a *cellular Markov partition* for g (see Definition 5.8). The map g sends each cell in $\mathcal{D}^1$ to a cell in $\mathcal{D}^0$. The cellular Markov partition $(\mathcal{D}^1, \mathcal{D}^0)$ together with this information completely determines the map g (up to conjugation). In this sense the dynamics of g is described by finite combinatorial data.

In fact, one can turn this process around and construct g from this combinatorial data, meaning essentially from the information encoded in Figure 1.1. A related discussion can be found in Section 1.5.

1.2. Cell decompositions

The previous example motivates several concepts for a general Thurston map $f\colon S^2 \to S^2$, in particular the combinatorial description of f that we will employ.

We choose a Jordan curve $\mathcal{C} \subset S^2$ with $\operatorname{post}(f) \subset \mathcal{C}$, and consider the preimages $f^{-n}(\mathcal{C})$. Then for each $n \in \mathbb{N}_0$ one obtains an associated cell decomposition $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$ of S^2. Its vertices are the points in $f^{-n}(\operatorname{post}(f))$, and its 1-skeleton the set $f^{-n}(\mathcal{C})$. The condition $\operatorname{post}(f) \subset \mathcal{C}$ ensures that the closure of each complementary component of $f^{-n}(\mathcal{C})$ is a closed Jordan region. These sets are the 2-dimensional cells in $\mathcal{D}^n$. We call each such set a *tile of level n* or an *n-tile* (of the cell decomposition). Similarly, we call any point $v \in f^{-n}(\operatorname{post}(f))$ a *vertex* of level n or an *n-vertex*; then $\{v\}$ is a 0-dimensional cell in $\mathcal{D}^n$. Finally, the closure e of a component of $f^{-n}(\mathcal{C}) \setminus f^{-n}(\operatorname{post}(f))$ is called an *edge* of level n or an *n-edge*; then e is a 1-dimensional cell of $\mathcal{D}^n$. The cells in $\mathcal{D}^n(f, \mathcal{C})$ of any dimension are called the *n-cells* for given f and $\mathcal{C}$. Note that here n always refers to the level of the cell and not to its dimension.

The cell decomposition $\mathcal{D}^0$ contains two tiles (the two closed Jordan regions in S^2 bounded by $\mathcal{C}$), $k = \#\operatorname{post}(f)$ vertices (the points $p \in \operatorname{post}(f)$), and k edges (the closed arcs into which the points in $\operatorname{post}(f)$ divide $\mathcal{C}$). We will study cell decompositions and their relation to Thurston maps in more detail in Chapter 5. Various examples for the cell decompositions $\mathcal{D}^n$ generated in this way can be found in Figures 8.1, 12.1, 12.7, and 15.1.

We say that a Thurston map $f \colon S^2 \to S^2$ is *expanding* if there exists a Jordan curve $\mathcal{C} \subset S^2$ with $\operatorname{post}(f) \subset \mathcal{C}$ such that the complementary components of $f^{-n}(\mathcal{C})$ become uniformly small in diameter as $n \to \infty$. Here S^2 is equipped with any metric inducing the topology on S^2. It is easy to see that this condition is independent of the choice of this base metric. Later we will show that it is also independent of the choice of $\mathcal{C}$ and will give other characterizations of expansion (see Chapter 6, in particular Proposition 6.3). A rational Thurston map $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is expanding precisely if it does not have periodic critical points or if its Julia set is equal to $\widehat{\mathbb{C}}$ (see Proposition 2.3). Note that in general an expanding Thurston map (that is not necessarily rational) may have periodic critical points (see Example 12.21).

Put differently, a Thurston map f is expanding if and only if the tiles in $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$ shrink to 0 in diameter uniformly as $n \to \infty$. This allows us to describe points in S^2 by suitable sequences of tiles. So we can think of $\mathcal{D}^n$ as a discrete approximation of S^2 that becomes finer with larger n.

We have seen that for the example g discussed in the previous section the n-tiles become uniformly small in diameter as $n \to \infty$; so we conclude that g is an expanding Thurston map.

Often the precise choice of the Jordan curve $\mathcal{C}$ with $\operatorname{post}(f) \subset \mathcal{C}$ will play no essential role, meaning that we may chose any such curve for our considerations. If the curve $\mathcal{C}$ is not f-invariant (as in the example discussed in the previous section), then in general the cell decompositions $\mathcal{D}^n$ will not be compatible for different levels n. Without precise knowledge of the behavior of the map we will then not have any information on how $(n+1)$-cells and n-cells intersect. The situation changes when the Jordan curve $\mathcal{C} \subset S^2$ with $\operatorname{post}(f) \subset \mathcal{C}$ is f-invariant. We will discuss existence of such invariant Jordan curves and the resulting combinatorial description of Thurston maps in Section 1.5.

1.3. Fractal spheres

We want to motivate other important concepts of our investigation, in particular the concept of visual metrics. To do this, we will discuss
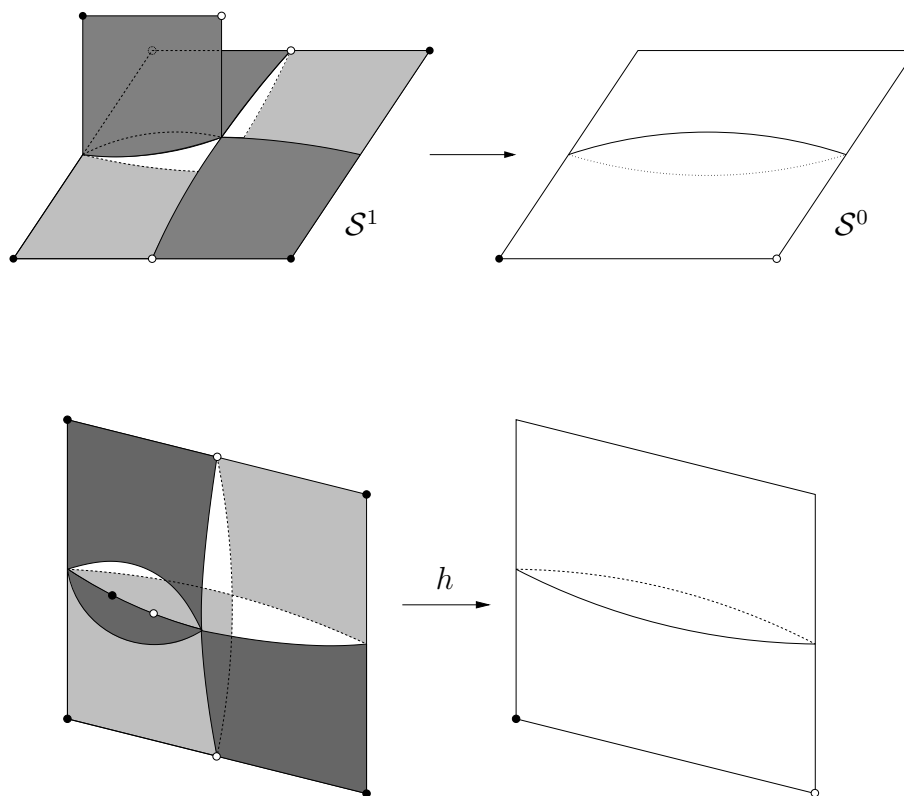
FIGURE 1.2. The map h.

another Thurston map and an associated fractal 2-sphere. As our main purpose here is to provide the reader with some intuition on the definition of a visual metric ϱ and on the fractal nature of the sphere (S^2, ϱ), we will omit the justification of some details.

The map arises from a geometric construction that is similar to the one used to describe the Lattès map in Section 1.1. Again we start with a pillow obtained by gluing together two squares along their boundaries; see the top right of Figure 1.2. This pillow is a polyhedral surface $\mathcal{S}^0$ homeomorphic to the 2-sphere. It carries a natural cell decomposition $\mathcal{D}^0$ with the two squares as tiles, the four sides of the common boundary of the squares as edges, and the four common corners of the squares as vertices. To distinguish them from other topological cells that we will introduce momentarily, we consider them as cells of level 0 and accordingly call them 0-tiles, 0-edges, and 0-vertices. As in the example of Section 1.1, we assign colors to the tiles; say the square on top of $\mathcal{S}^0$ as shown in Figure 1.2 is white, and the bottom square is black.

To obtain cells on the next level 1, each of the two squares, or more precisely 0-tiles, is divided into four squares of half the side length. We

call these eight smaller squares tiles of level 1, or simply 1-tiles. The edges of these squares are 1-edges. We slit the sphere along one such 1-edge in the white 0-tile and glue in two small squares at the slit, as indicated on the upper left of Figure 1.2. This gives two additional 1-tiles and we obtain a polyhedral surface $\mathcal{S}^1$ homeomorphic to the 2-sphere. The surface $\mathcal{S}^1$ carries a cell decomposition given by the topological cells of level 1 as described. We color the 1-tiles black and white in a checkerboard fashion so that 1-tiles sharing an edge have different color, as indicated in Figure 1.2.

To define a Thurston map based on our construction, we choose an identification of the polyhedral surface $\mathcal{S}^1$ with $\mathcal{S}^0$. To do this, we represent the six 1-tiles that replaced the white 0-tile topologically as subsets of this tile, and similarly the other four 1-tiles as subsets of the black 0-tile. So the 0-tiles are "subdivided" by 1-tiles. This is indicated in the lower left of Figure 1.2. Under this identification the cell decomposition of $\mathcal{S}^1$ gives a cell decomposition $\mathcal{D}^1$ on $\mathcal{S}^0$ that is a refinement of the cell decomposition $\mathcal{D}^0$.

Now we can define a Thurston map as follows. We map each white 1-tile on the polyhedral surface $\mathcal{S}^1$ to the white 0-tile in $\mathcal{S}^0$, and each black 1-tile in $\mathcal{S}^1$ to the black 0-tile in $\mathcal{S}^0$ by a similarity map (preserving orientation). This is a well-defined and uniquely determined map on $\mathcal{S}^1$ if we do this so that the 1-vertices marked by black and by white dots in the upper left of Figure 1.2 are sent to 0-vertices in the upper right of the picture with the same markings.

If we identify $\mathcal{S}^1$ with $\mathcal{S}^0$ as discussed, we get a map $h\colon S^2 \to S^2$ on the 2-sphere $S^2 \coloneqq \mathcal{S}^0$. Since h restricted to each 1-tile is a homeomorphism onto a 0-tile, h is a branched covering map. The critical points of h are the 1-vertices where at least four 1-tiles intersect. These critical points are all mapped to vertices of the pillow, i.e., to 0-vertices. All 0-vertices in turn are mapped to the 0-vertex marked black. Thus h is a postcritically-finite branched covering map on S^2, i.e., a Thurston map. Note that the equator $\mathcal{C}$ of the pillow is an h-invariant Jordan curve (i.e., $h(\mathcal{C}) \subset \mathcal{C}$) and that the cell decomposition $\mathcal{D}^1$ on $\mathcal{S}^1 \cong \mathcal{S}^0$ is determined by $h^{-1}(\mathcal{C})$. Namely, $h^{-1}(\mathcal{C})$ is a topological graph that gives the 1-skeleton of this cell decomposition, and each 1-tile is the closure of a complementary component of $h^{-1}(\mathcal{C})$.

The relevant information on the map h is contained in the combinatorics of the cell decompositions $\mathcal{D}^0$ and $\mathcal{D}^1$ and a map $L\colon \mathcal{D}^1 \to \mathcal{D}^0$ that records how h associates the cells in $\mathcal{D}^1$ with cells in $\mathcal{D}^0$. This triple $(\mathcal{D}^1, \mathcal{D}^0, L)$ forms a *two-tile subdivision rule* that is *realized* by h. We will give precise definitions of these concepts in Chapter 12. The

map h depends on choices and is not uniquely determined, but another map realizing the same subdivision rule $(\mathcal{D}^1, \mathcal{D}^0, L)$ is *Thurston equivalent* to h (see Definition 2.4 for the terminology).

There is no rational map that realizes this combinatorial picture as our map h. More precisely, h is not Thurston equivalent to a rational map, because h has a *Thurston obstruction*. This, together with the terminology, will be explained in Section 2.6.

We will now describe a fractal sphere $\mathcal{S}$ that is associated with our construction and gives an alternative way to view our map h. The sphere $\mathcal{S}$ is obtained in a similar way as the well-known snowflake curve. We will also define a metric ϱ on $\mathcal{S}$.

To construct the space $\mathcal{S}$, we do not identify the surfaces $\mathcal{S}^0$ and $\mathcal{S}^1$. Instead, we consider the passage from $\mathcal{S}^0$ to $\mathcal{S}^1$ as a replacement procedure. The white 0-tile is replaced by the top part of $\mathcal{S}^1$, consisting of six 1-tiles, i.e., squares of side length $1/2$; we call this top part of $\mathcal{S}^1$ the *white generator*. Similarly, the four 1-tiles subdividing the black 0-tile form the *black generator*. The polyhedral surface $\mathcal{S}^1$ consisting of ten squares is the first approximation of the fractal space $\mathcal{S}$ that we are about to construct by iterating this procedure.

Namely, the 1-tiles of the black and white generators are also colored as indicated in Figure 1.2. So if we replace each black or white 1-tile by a suitably scaled copy of the black or white generators, then we obtain a polyhedral surface $\mathcal{S}^2$ glued together from squares of side length $1/4$ as 2-tiles. Here we have to be careful about how precisely a tile is replaced by an appropriate generator, because the generators with their given colorings of tiles are not symmetric with respect to rotations. To specify the replacement rule uniquely, we use the additional markings of some points. Each generator carries two points corresponding to the points on $\mathcal{S}^0$ marked black and white. In the replacement process we require that these points match the corresponding points with the same markings on 1-tiles.

If we iterate the replacement procedure in this way, we obtain polyhedral surfaces $\mathcal{S}^n$ for all levels $n \in \mathbb{N}_0$ glued together from squares of side length $1/2^n$. Each surface $\mathcal{S}^n$ carries a natural cell decomposition $\mathcal{D}^n$ given by these squares as tiles. Some iterates of this construction are shown in Figure 1.3. The pictures essentially indicate the gluing pattern of the squares which give the surfaces. One should view them as abstract polyhedral surfaces, and not confuse them with the underlying subsets of $\mathbb{R}^3$ in these pictures. Each surface $\mathcal{S}^n$ is a topological 2-sphere and carries a piecewise Euclidean path metric ϱ_n with conical singularities.
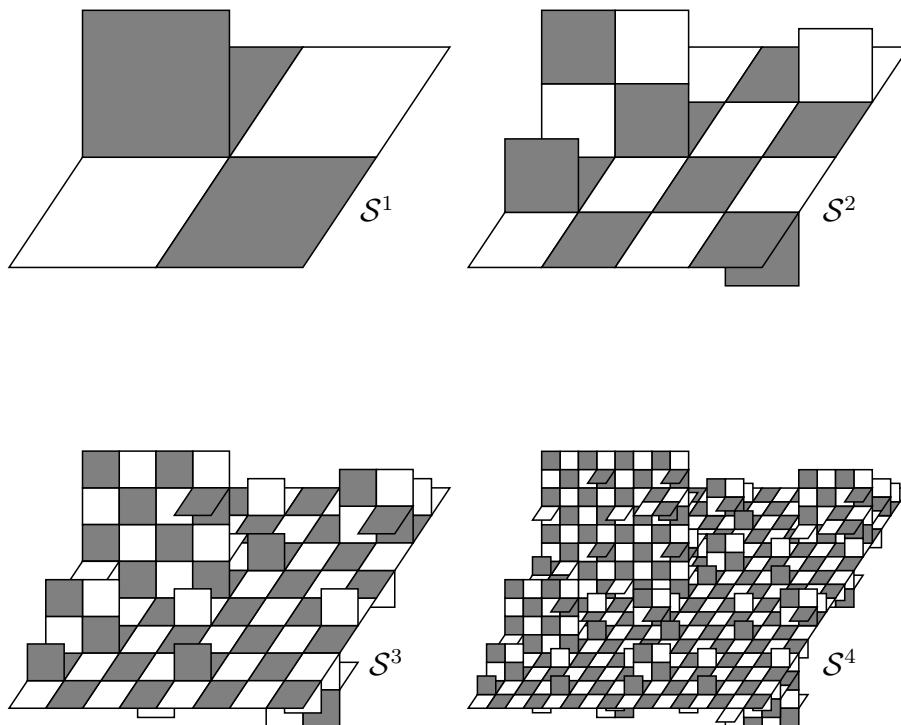
FIGURE 1.3. Polyhedral surfaces obtained from the replacement rule.

One can now extract a self-similar "fractal" space $\mathcal{S}$ as a limit $\mathcal{S}^n \to \mathcal{S}$ for $n \to \infty$ in several ways. One possibility is to pass to a Gromov-Hausdorff limit of the sequence $(\mathcal{S}^n, \varrho_n)$ of metric spaces. We will discuss a different method that is closer in spirit to our general definition of a visual metric (see Chapter 8 and Chapter 10; similar considerations appear in Chapter 14).

Namely, given an n-tile $\mathcal{X}^n \subset \mathcal{S}^n$ (which is a square of side length 2^{-n}), and an $(n+1)$-tile $\mathcal{X}^{n+1} \subset \mathcal{S}^{n+1}$, we write $\mathcal{X}^n \sqsupset \mathcal{X}^{n+1}$ if $\mathcal{X}^{n+1}$ is contained in the scaled copy of a generator that replaced $\mathcal{X}^n$ in the construction of $\mathcal{S}^{n+1}$ from $\mathcal{S}^n$. We now consider descending sequences $\mathcal{X}^0 \sqsupset \mathcal{X}^1 \sqsupset \mathcal{X}^2 \sqsupset \ldots$. On an intuitive level the squares in such a sequence should shrink to a point in our desired limit space $\mathcal{S}$ represented by the sequence. Here we consider two sequences $\{\mathcal{X}^n\}$ and $\{\mathcal{Y}^n\}$ as equivalent and representing the same point if $\mathcal{X}^n \cap \mathcal{Y}^n \neq \emptyset$ for all $n \in \mathbb{N}_0$. It is not hard to see that this indeed defines an equivalence relation for descending sequences. By definition our limit space $\mathcal{S}$ is now the set of all equivalence classes.

For $x, y \in \mathcal{S}$ we set

$$(1.2) \qquad \varrho(x, y) := \limsup_{n \to \infty} \mathrm{dist}_{\varrho_n}(\mathcal{X}^n, \mathcal{Y}^n),$$

where $\{\mathcal{X}^n\}$ and $\{\mathcal{Y}^n\}$ are sequences representing x and y, respectively. Then ϱ is well-defined and one can show that this is a metric on $\mathcal{S}$.

For $x, y \in \mathcal{S}$, $x \neq y$, we define

$$(1.3) \qquad m(x, y) := \inf\{n \in \mathbb{N} : \mathcal{X}^n \cap \mathcal{Y}^n = \emptyset\},$$

where the infimum is taken over all sequences $\{\mathcal{X}^n\}$ and $\{\mathcal{Y}^n\}$ representing x and y, respectively. Then

$$(1.4) \qquad \varrho(x, y) \asymp 2^{-m(x,y)}$$

for $x, y \in \mathcal{S}$, $x \neq y$. This notation (which will be used frequently) means that there is a constant $C > 0$ such that

$$\frac{1}{C}\varrho(x, y) \le 2^{-m(x,y)} \le C\varrho(x, y).$$

We refer to the constant as $C(\asymp) = C$. Here $C(\asymp)$ does not depend on x, y, or n. So roughly speaking, the distance of two distinct points in $\mathcal{S}$ is given in terms of the minimal level on which two descending sequences representing the points can distinguish them.

It is intuitively clear that $(\mathcal{S}, \varrho)$ is a topological 2-sphere. To outline a rigorous proof for this fact, we return to the Thurston map h defined above. Recall that $\mathcal{C} \subset S^2$ is the h-invariant Jordan curve given by the common boundary of the 0-tiles. The curve $\mathcal{C}$ contains the vertices of the pillow, which are the postcritical points of h. We consider the cell decompositions $\mathcal{D}^n(h, \mathcal{C})$ as discussed in the previous section. Note that the 1-tiles (i.e., the tiles in $\mathcal{D}^1(h, \mathcal{C})$) are exactly the 1-tiles in $\mathcal{S}^1$ under the identification $\mathcal{S}^1 \cong \mathcal{S}^0 = S^2$ (see bottom left in Figure 1.2).

The map h^n sends each n-tile to a 0-tile homeomorphically, and we can assign colors to n-tiles so that h^n sends an n-tile to the 0-tile of the same color. In the passage from $\mathcal{D}^n(h, \mathcal{C})$ to $\mathcal{D}^{n+1}(h, \mathcal{C})$ each n-tile is subdivided by $(n+1)$-tiles in the same way as the 0-tile of the same color is subdivided by 1-tiles. From this it is clear that there is a one-to-one correspondence between n-tiles in $\mathcal{S}^n$ and n-tiles for the pair $(h, \mathcal{C})$. Moreover, these tiles realize identical combinatorics. More precisely, we have

$$\mathcal{X}^{n+1} \sqsupset \mathcal{X}^n \quad \Leftrightarrow \quad X^{n+1} \supset X^n, \quad \text{and}$$
$$\mathcal{X}^n \cap \mathcal{Y}^n \neq \emptyset \quad \Leftrightarrow \quad X^n \cap Y^n \neq \emptyset,$$

where the n-tiles $\mathcal{X}^n$ and $\mathcal{Y}^n$ in $\mathcal{S}^n$ and the $(n+1)$-tile $\mathcal{X}^{n+1}$ in $\mathcal{S}^{n+1}$ correspond to the n-tiles X^n and Y^n and the $(n+1)$-tile X^{n+1} for $(h, \mathcal{C})$, respectively.

Recall from Section 1.2 that h is expanding if the diameters of n-tiles for $(h, \mathcal{C})$ tend to 0 uniformly as $n \to \infty$ (with respect to some fixed base metric on S^2 representing the topology). In this case one obtains a well-defined map $\varphi \colon \mathcal{S} \to S^2$ by sending a point in $\mathcal{S}$ represented by a descending sequence $\{\mathcal{X}^n\}$ to the unique point in the intersection $\bigcap_{n \in \mathbb{N}_0} X^n$ of the corresponding n-tiles for $(h, \mathcal{C})$. In general, our map h need not be expanding, but we may assume this if we choose the identification of $\mathcal{S}^1$ with $\mathcal{S}^0$ carefully (this hinges on the fact that h is "combinatorially expanding" and so the map can be corrected if necessary to make it expanding; see Theorem 14.2 for details). It is then not hard to see that φ is a homeomorphism, and so $\mathcal{S}$ is a 2-sphere.

Though $(\mathcal{S}, \varrho)$ is a topological 2-sphere, it is not a *quasisphere*. This means that this space is not quasisymmetrically equivalent to the standard 2-sphere (i.e., the unit sphere in $\mathbb{R}^3$, or equivalently the Riemann sphere $\widehat{\mathbb{C}}$ equipped with the chordal metric; see Section 4.1 for the definition of a quasisymmetry). This is closely related to the fact that h is not (equivalent to) a rational map. One can deduce that $(\mathcal{S}, \varrho)$ is not a quasisphere from a general result (see Theorem 18.1 (ii) mentioned in the next section), but one can also show this directly (we will outline an argument in Section 4.4).

The fractal sphere $(\mathcal{S}, \varrho)$ is in a sense the natural domain for our map h; namely, we can conjugate our original map $h \colon S^2 \to S^2$ by the homeomorphism $\varphi \colon \mathcal{S} \to S^2$ to obtain a map on $\mathcal{S}$, also denoted by h. We also obtain n-tiles in $\mathcal{S}$ corresponding to the n-tiles in S^2 under the homeomorphism φ. Roughly speaking, an n-tile in $\mathcal{S}$ is the part of $\mathcal{S}$ that "sits on top" of an n-tile in $\mathcal{S}^n$. The new map $h \colon \mathcal{S} \to \mathcal{S}$ then behaves locally like a similarity map: it scales each $(n+1)$-tile in $\mathcal{S}$ by a factor 2 and matches it with the corresponding n-tile.

1.4. Visual metrics and the visual sphere

After this example we return to the general setting. Let $f \colon S^2 \to S^2$ be an expanding Thurston map. We fix a Jordan curve $\mathcal{C} \subset S^2$ with $\operatorname{post}(f) \subset S^2$. Then for $n \in \mathbb{N}_0$ we have cell compositions $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$ with 1-skeleton $f^{-n}(\mathcal{C})$ as defined in Section 1.2.

Since f is expanding, the diameters of n-tiles (i.e., tiles in $\mathcal{D}^n$) shrink to 0 uniformly as $n \to \infty$. So if $x, y \in S^2$ are distinct points and X^n and Y^n are tiles of level n with $x \in X$ and $y \in Y$, then $X^n \cap Y^n = \emptyset$ for sufficiently large n. This implies that the number

$$(1.5) \qquad m(x, y) := \max\{n \in \mathbb{N}_0 : \text{ there exist non-disjoint } n\text{-tiles}$$
$$X^n \text{ and } Y^n \text{ with } x \in X^n, y \in Y^n\}$$

is finite. Similar to (1.3), it records the level at which x and y can be separated by tiles. In Figure 8.1 we have illustrated this separation by tiles in an example.

Generalizing (1.4), we consider metrics ϱ on S^2 satisfying

$$\varrho(x, y) \asymp \Lambda^{-m(x,y)},$$

for some $\Lambda > 1$. We call such a metric a *visual metric* for f, and Λ its *expansion factor*. We will start investigating visual metrics in earnest in Chapter 8.

Visual metrics for a given Thurston map f are not unique, but two different visual metrics with the same expansion factor are bi-Lipschitz equivalent. They are snowflake equivalent if they have different expansion factors (see Section 4.1 for this terminology). Whether a metric is visual does not depend on the choice of the Jordan curve $\mathcal{C}$ that was used to define the quantity (1.5) via the cell decompositions $\mathcal{D}^n(f, \mathcal{C})$. Moreover, if $F = f^k$ is an iterate of f (where $k \in \mathbb{N}$), then a metric is visual for f if and only if it is visual for F. These (and more) basic properties of visual metrics can be found in Proposition 8.3.

If σ is a tile or an edge in the cell decomposition $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$, then $\operatorname{diam}_\varrho(\sigma) \asymp \Lambda^{-n}$. In addition, any two disjoint cells $\sigma, \tau \in \mathcal{D}^n$ satisfy $\operatorname{dist}_\varrho(\sigma, \tau) \gtrsim \Lambda^{-n}$. This notation means that there is a constant $C > 0$ such that $C \operatorname{dist}_\varrho(\sigma, \tau) \geq \Lambda^{-n}$. We refer to the constant as $C = C(\gtrsim)$. Equivalently, we write $\Lambda^{-n} \lesssim \operatorname{dist}_\varrho(\sigma, \tau)$ and refer to the constant as $C = C(\lesssim)$. Here the constants $C(\asymp)$ and $C(\gtrsim)$ do not depend on n or the cells involved. In fact, these two geometric properties characterize visual metrics (see Proposition 8.4).

For the map g from Section 1.1 the length metric induced by the Euclidean metric on the pillow P is a visual metric with expansion factor $\Lambda = 2$. Similarly, the particular metric ϱ defined in Section 1.3 is a visual metric for h with expansion factor $\Lambda = 2$ (here we identify $\mathcal{S}$ with S^2 by the homeomorphism φ). In this case, we obtain visual metrics with arbitrary expansion factor $1 < \Lambda \leq 2$ if we consider a "snowflaked" metric ϱ^α with suitable $\alpha \in (0, 1)$, but there is no visual metric for h with $\Lambda > 2$. Indeed, if ϱ is a visual metric with expansion factor $\Lambda > 1$ and X^n is an n-tile, then $\operatorname{diam}_\varrho(X^n) \lesssim \Lambda^{-n}$. Now it is easy to see that one can form a connected chain of n-tiles with 2^n elements that joins two non-adjacent 0-edges (as Figure 1.3 suggests, one obtains such a chain by running along the bottom 0-edge). Then by the triangle inequality $2^n \cdot \Lambda^{-n} \gtrsim 1$ for all $n \in \mathbb{N}$, and so $\Lambda \leq 2$.

If $f \colon S^2 \to S^2$ is an arbitrary Thurston map, then the supremum of all $\Lambda > 1$ for which there exist visual metrics with expansion factor Λ agrees with the *combinatorial expansion factor* of f, denoted by

$\Lambda_0(f)$. It is computed from data associated with the cell decomposi-
tions $\mathcal{D}^n(f,\mathcal{C})$ determined by the map f and a Jordan curve $\mathcal{C} \subset S^2$
with $\mathrm{post}(f) \subset \mathcal{C}$. For this we consider the combinatorial quantity
$D_n(f,\mathcal{C})$ defined to be the minimal number of tiles in $\mathcal{D}^n(f,\mathcal{C})$ that
are needed to form a connected set joining opposite sides of $\mathcal{C}$, i.e.,
two non-adjacent 0-edges (the definition is slightly different in the case
$\#\mathrm{post}(f) = 3$; see Section 5.7). For the examples discussed in Sec-
tions 1.1 and 1.3 we have $D_n(g,\mathcal{C}) = 2^n$ and $D_n(h,\mathcal{C}) = 2^n$ for $n \in \mathbb{N}_0$.

In general, the number $D_n(f,\mathcal{C})$ depends on $\mathcal{C}$. For an expanding
Thurston map it grows at an exponential rate as $n \to \infty$. This growth
rate is independent of $\mathcal{C}$, and determined only by f. Moreover, the
limit

$$(1.6) \qquad\qquad \Lambda_0(f) := \lim_{n\to\infty} D_n(f,\mathcal{C})^{1/n}$$

exists, satisfies $1 < \Lambda_0(f) < \infty$, and is defined to be the combinatorial
expansion factor of f (see Proposition 16.1). It is invariant under topo-
logical conjugacy and well-behaved under iteration (see Chapter 16,
and in particular Theorem 16.3 and Proposition 16.2). For our two
examples we have $\Lambda_0(g) = 2$ and $\Lambda_0(h) = 2$.

As already mentioned above, $\Lambda_0(f)$ gives the range of possible ex-
pansion factors of visual metrics for an expanding Thurston map f.
This is made precise in the following theorem.

THEOREM 16.3 (Visual metrics and their expansion factors). *Let*
$f\colon S^2 \to S^2$ *be an expanding Thurston map, and* $\Lambda_0(f) \in (1,\infty)$ *be
its combinatorial expansion factor. Then the following statements are
true:*

(i) *If* Λ *is the expansion factor of a visual metric for* f*, then*
$1 < \Lambda \le \Lambda_0(f)$.

(ii) *Conversely, if* $1 < \Lambda < \Lambda_0(f)$*, then there exists a visual metric*
ϱ *for* f *with expansion factor* Λ*. Moreover, the visual metric*
ϱ *can be chosen to have the following additional property: For
every* $x \in S^2$ *there exists a neighborhood* U_x *of* x *such that*

$$\varrho(f(x), f(y)) = \Lambda\varrho(x,y) \text{ for all } y \in U_x.$$

(Note that in this introduction we label the results as they appear
in later chapters.)

In general, one cannot guarantee the existence of a visual metric
with expansion factor $\Lambda = \Lambda_0(f)$ (see Example 16.8).

The combinatorial expansion factor always satisfies the inequality
$\Lambda_0(f) \le \deg(f)^{1/2}$, where $\deg(f)$ is the (topological) degree of f (see
Proposition 20.1). For our examples g and h from the previous sections

we have $\Lambda_0(g) = 2 = \deg(g)^{1/2}$ and $\Lambda_0(h) = 2 < \deg(h)^{1/2} = \sqrt{5}$. The equality for the Lattès map g is not a coincidence. Closely related results will be discussed in Section 1.7.

According to Theorem 16.3 (ii), for each expanding Thurston map f we can find a visual metric ϱ so that f scales the metric ϱ by a constant factor at each point. The Lattès map $g \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ discussed in Section 1.1 illustrates this statement: if we equip $\widehat{\mathbb{C}}$ with a suitable visual metric for g (the path metric on the pillow in Figure 1.1), then g behaves like a piecewise similarity map, where distances are scaled by the factor $\Lambda = 2$.

The space $\mathcal{S}$ from Section 1.3 equipped with the visual metric ϱ in (1.2) is a fractal sphere. It is self-similar in the sense that the part of the surface that is "built on top" of some n-tile $\mathcal{X}^n$ is similar (i.e., is isometric up to scaling by the factor 2^n) to the part of the surface that is "built on top" of the white or black 0-tile. Similarly, we can find visual metrics for any Thurston map f such that f scales tiles by a constant factor. Then the metric behavior of the dynamics on tiles becomes very simple, while the space on which f acts is a fractal sphere and geometrically more complicated.

Our choice of the term "visual metric" is motivated by the close relation of this concept to the notion of a visual metric on the boundary of a Gromov hyperbolic space (see Section 4.2 for general background; very similar ideas can be found in [**HP09**]). Namely, if $f \colon S^2 \to S^2$ is an expanding Thurston map and $\mathcal{C} \subset S^2$ a Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}$, then one can define an associated *tile graph* $\mathcal{G}(f, \mathcal{C})$ as follows. Its vertices are given by the tiles in the cell decompositions $\mathcal{D}^n(f, \mathcal{C})$ on all levels $n \in \mathbb{N}_0$, and one joins two vertices by an edge if the corresponding tiles intersect and have levels differing by at most 1 (see Chapter 10). The graph $\mathcal{G}(f, \mathcal{C})$ depends on the choice of $\mathcal{C}$, but if $\mathcal{C}' \subset S^2$ is another Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}'$, then $\mathcal{G}(f, \mathcal{C})$ and $\mathcal{G}(f, \mathcal{C}')$ are rough-isometric (Theorem 10.4); note that this is much stronger than being quasi-isometric (see Section 4.2 for the terminology).

The graph $\mathcal{G}(f, \mathcal{C})$ is Gromov hyperbolic (Theorem 10.1). Its boundary at infinity $\partial_\infty \mathcal{G}(f, \mathcal{C})$ can be identified with S^2. Under this identification the class of visual metrics in the sense of Gromov hyperbolic spaces coincides with the class of visual metrics for f in our sense (Theorem 10.2). The number $m(x, y)$ defined in (1.5) is the Gromov product of the points $x, y \in S^2 \cong \partial_\infty \mathcal{G}(f, \mathcal{C})$ up to a uniformly bounded additive constant (Lemma 10.3).

If $f \colon S^2 \to S^2$ is an expanding Thurston map and ϱ a visual metric for f, then we call the metric space (S^2, ϱ) the *visual sphere* of f.

For fixed f different visual metrics ϱ_1 and ϱ_2 give snowflake equivalent spaces (S^2, ϱ_1) and (S^2, ϱ_1). So an expanding Thurston map determines its visual sphere uniquely up to snowflake equivalence.

Many dynamical properties of f are encoded in the geometry of its visual sphere. The following statement is one of the main results of this work.

THEOREM 18.1 (Properties of f and its associated visual sphere). *Let* $f \colon S^2 \to S^2$ *be an expanding Thurston map and* ϱ *be a visual metric for* f. *Then the following statements are true:*

(i) (S^2, ϱ) *is doubling if and only if* f *has no periodic critical points.*

(ii) (S^2, ϱ) *is quasisymmetrically equivalent to* $\widehat{\mathbb{C}}$ *if and only if* f *is topologically conjugate to a rational map.*

(iii) (S^2, ϱ) *is snowflake equivalent to* $\widehat{\mathbb{C}}$ *if and only if* f *is topologically conjugate to a Lattès map.*

Here it is understood that $\widehat{\mathbb{C}}$ is equipped with the chordal metric. For the terminology used in the statements see Section 4.1.

As we already discussed, part (ii) of the previous theorem provides an analog of Cannon's conjecture in geometric group theory (see Section 4.3 for a more detailed discussion). According to this conjecture every Gromov hyperbolic group G whose boundary at infinity $\partial_\infty G$ is a 2-sphere should arise from some standard situation in hyperbolic geometry. The conjecture is equivalent to showing that $\partial_\infty G$ equipped with a visual metric (in the sense of Gromov hyperbolic spaces) is quasisymmetrically equivalent to $\widehat{\mathbb{C}}$. One of the reasons why Cannon's conjecture is still open may be the lack of non-trivial examples that guide the intuition (see the paper [**BK11**] though, which in a sense addresses this issue). All examples come from fundamental groups G of compact hyperbolic manifolds where one already has a natural identification of $\partial_\infty G$ with $\widehat{\mathbb{C}}$; according to Cannon's conjecture there are no other examples. In contrast, the visual spheres of expanding Thurston maps provide a rich supply of metric 2-spheres that sometimes are and sometimes are not quasisymmetrically equivalent to $\widehat{\mathbb{C}}$ (see Section 4.4).

The proof of one of the implications in Theorem 18.1 (ii) (the "if" part) uses some well-known ingredients. Namely, if (S^2, ϱ) is quasisymmetrically equivalent to the standard sphere $\widehat{\mathbb{C}}$, then one can conjugate f to a map g on $\widehat{\mathbb{C}}$. Since the map f dilates distances with respect to a suitable visual metric by a fixed factor (see Theorem 16.3 (ii) mentioned above), the map g is uniformly quasiregular (see Section 4.1 for

the terminology). Hence g, and therefore also f, are conjugate to a rational map by a standard theorem.

The converse direction (the "only if" part) is harder to establish. If f is conjugate to a rational map, then we may assume without loss of generality that f is an expanding rational Thurston map on $\widehat{\mathbb{C}}$ to begin with. If ϱ is a visual metric for f, then one shows that the identity map from $(\widehat{\mathbb{C}}, \varrho)$ to $(\widehat{\mathbb{C}}, \sigma)$ is a quasisymmetry, where σ is the chordal metric. This follows from a careful analysis of the geometry of the tiles in the cell decompositions $\mathcal{D}^n(f, \mathcal{C})$ with respect to the metric σ (see Proposition 18.8). For example, while it is fairly obvious from the definitions that adjacent tiles in $\mathcal{D}^n(f, \mathcal{C})$ have comparable diameter with respect to a visual metric ϱ (with uniform constants independent of the level n), the same assertion is also true for the chordal metric σ. Our proof of this and related statements is based on Koebe's distortion theorem and the fact that if f has no periodic critical points, then in the cell decompositions $\mathcal{D}^n(f, \mathcal{C})$ we see locally only finitely many different combinatorial types.

Thurston studied the question when a given Thurston map is represented by a conformal dynamical system from a point of view different from the one suggested by Theorem 18.1 (ii) (see Section 2.6 for a short overview). He asked when a Thurston map $f \colon S^2 \to S^2$ is in a suitable sense *(Thurston) equivalent* (see Definition 2.4) to a rational map and obtained a necessary and sufficient condition (see [**DH93**]). For expanding Thurston maps his notion of equivalence actually means the same as topological conjugacy of the maps (Theorem 11.1).

The proof of part (ii) of Theorem 18.1 does not use Thurston's theorem. Indeed, none of our statements relies on this, and so our methods possibly provide a different approach for its proof.

It is not clear how useful Theorem 18.1 (ii) is for deciding whether an explicitly given expanding Thurston map is topologically conjugate to a rational map. It is likely that our techniques can be used to formulate a more efficient criterion, but we will not pursue this further here.

1.5. Invariant curves

The Jordan curve $\mathcal{C}$ chosen in Section 1.1 is invariant for the map g in the sense that $g(\mathcal{C}) \subset \mathcal{C}$. In this case the cell decomposition $\mathcal{D}^{n+1}(g, \mathcal{C})$ is a refinement of $\mathcal{D}^n(g, \mathcal{C})$ for each $n \in \mathbb{N}_0$. We have a similar situation for the Jordan curve $\mathcal{C}$ and the map h in Section 1.3.

Some of our main results are about the existence and uniqueness of such invariant Jordan curves $\mathcal{C}$. In particular, we will show that they exist for sufficiently high iterates of *every* expanding Thurston map.

THEOREM 15.1 (Sufficiently high iterates have invariant curves). *Let* $f\colon S^2 \to S^2$ *be an expanding Thurston map, and* $\mathcal{C} \subset S^2$ *be a Jordan curve with* $\operatorname{post}(f) \subset \mathcal{C}$. *Then for each sufficiently large* $n \in \mathbb{N}$ *there exists a Jordan curve* $\widetilde{\mathcal{C}} \subset S^2$ *that is invariant for* f^n *and isotopic to* $\mathcal{C}$ *rel.* $\operatorname{post}(f)$.

A discussion of isotopies and related terminology can be found in Section 2.4. Since $\widetilde{\mathcal{C}}$ is isotopic to $\mathcal{C}$ rel. $\operatorname{post}(f)$, it will also contain the set $\operatorname{post}(f)$.

In Example 15.11 we exhibit an expanding Thurston map $f\colon S^2 \to S^2$ that has no f-invariant Jordan curve $\widetilde{\mathcal{C}} \subset S^2$ with $\operatorname{post}(f) \subset \widetilde{\mathcal{C}}$. This shows that in general it is necessary to pass to an iterate in Theorem 15.1.

If a curve $\widetilde{\mathcal{C}}$ is invariant for some iterate f^n, then one cannot expect it to be invariant for some other iterate f^k unless k is a multiple of n (see Remark 15.16). So typically the curve $\widetilde{\mathcal{C}}$ in the previous theorem will depend on n.

The proof of Theorem 15.1 is based on a necessary and sufficient criterion on the existence of f-invariant curves given in Theorem 15.4. An outline of the proof of this latter theorem is presented in Example 15.6 (see Figure 15.1 for an illustration).

One can actually formulate a related criterion for the existence of an invariant curve in a given isotopy class rel. $\operatorname{post}(f)$ or rel. $f^{-1}(\operatorname{post}(f))$ (see Remark 15.13 (iii)). Moreover, if an f-invariant Jordan curve $\widetilde{\mathcal{C}}$ exists, then it is the Hausdorff limit of a sequence of Jordan curves $\mathcal{C}^n$ that can be obtained from a simple iterative procedure (Proposition 15.20).

Our existence results are complemented by the following uniqueness statement for invariant Jordan curves.

THEOREM 15.5 (Uniqueness of invariant curves). *Let* $f\colon S^2 \to S^2$ *be an expanding Thurston map, and* $\mathcal{C}, \mathcal{C}' \subset S^2$ *be* f-invariant Jordan *curves that both contain the set* $\operatorname{post}(f)$. *Then* $\mathcal{C} = \mathcal{C}'$ *if and only if* $\mathcal{C}$ *and* $\mathcal{C}'$ *are isotopic rel.* $f^{-1}(\operatorname{post}(f))$.

As a consequence one can prove that if $\#\operatorname{post}(f) = 3$, then there are at most finitely many f-invariant Jordan curves $\mathcal{C} \subset S^2$ with $\operatorname{post}(f) \subset \mathcal{C}$ (Corollary 15.8). This is also true if f is rational and not a *flexible Lattès map*. In general, a Thurston map f can have infinitely many

such invariant curves (Example 15.9), but there are at most finitely many in a given isotopy class rel. $\mathrm{post}(f)$ (Corollary 15.7).

Let $f\colon S^2 \to S^2$ be a Thurston map and suppose $\mathcal{C} \subset S^2$ is an f-invariant Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}$. We consider the cell decompositions $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$ as discussed in Section 1.2. The f-invariance of $\mathcal{C}$ implies that $f^{n+1}(f^{-n}(\mathcal{C})) \subset \mathcal{C}$, or equivalently that $f^{-n}(\mathcal{C}) \subset f^{-(n+1)}(\mathcal{C})$ for all $n \in \mathbb{N}_0$. Since n-tiles and $(n+1)$-tiles were defined to be the closures of the complementary components of $f^{-n}(\mathcal{C})$ and $f^{-(n+1)}(\mathcal{C})$, respectively, each $(n+1)$-tile is contained in an n-tile. Similarly, every cell in $\mathcal{D}^{n+1}$ is contained in a cell in $\mathcal{D}^n$. More precisely, $\mathcal{D}^{n+1}$ is a refinement of $\mathcal{D}^n$ (see Definition 5.6 and Proposition 12.5). In particular, $\mathcal{D}^1$ refines $\mathcal{D}^0$.

One can essentially recover the Thurston map f from the cell decomposition $\mathcal{D}^0$ and its refinement $\mathcal{D}^1$ if one specifies some additional data. In the example in Figure 1.1 we labeled the vertices in domain and range to indicate their correspondence under the map. Similarly, in the general case this additional information is provided by a *labeling*, which is a map $L\colon \mathcal{D}^1 \to \mathcal{D}^0$ (see Section 5.4).

The triple $(\mathcal{D}^1, \mathcal{D}^0, L)$ records how 0-cells are subdivided by 1-cells and how 1-cells are mapped to 0-cells. We call such triples $(\mathcal{D}^1, \mathcal{D}^0, L)$ *two-tile subdivision rules* (see Definition 12.1), because $\mathcal{D}^0$ contains two tiles (namely the two closed Jordan regions bounded by $\mathcal{C}$). Every Thurston map f with an f-invariant Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}$ gives rise to a two-tile subdivision rule (see Proposition 12.2).

Note that L is a map between finite sets. This means that the information encoded in $(\mathcal{D}^1, \mathcal{D}^0, L)$ is given in terms of finite data. So a two-tile subdivision rule can be considered as a combinatorial object.

Conversely, every two-tile subdivision rule can be *realized* by a Thurston map (see Proposition 12.3). It is unique up to Thurston equivalence. In this way, two-tile subdivision rules give simple combinatorial models for Thurston maps. There is no obvious difference between the two-tile subdivisions realized by rational maps and the ones that are not. This is another motivation for investigating general Thurston maps.

The map g in Section 1.1 and the map h in Section 1.3 were constructed from Figure 1.1 and Figure 1.2, respectively. This really means that the pictures represent two-tile subdivision rules, and the maps realize these subdivision rules according to Proposition 12.3. This is our preferred way to define Thurston maps. To be able to discuss examples,

we will use this way of constructing Thurston maps in an informal way even before we provide the theoretical foundations in Chapter 12.

Our concept of a two-tile subdivision rule is closely related to the general *subdivision rules* that have been studied extensively by Cannon, Floyd, and Parry (see for example [**CFP01**]).

The main consequence of Theorem 15.1 is that we get a combinatorial description by a two-tile subdivision rule for sufficiently high iterates $F = f^n$ of every expanding Thurston map f.

COROLLARY 15.2 (Thurston maps as two-tile subdivision rules). *Let* $f\colon S^2 \to S^2$ *be an expanding Thurston map. Then for each sufficiently large* $n \in \mathbb{N}$ *there exists a two-tile subdivision rule that is realized by* $F = f^n$.

In particular, we obtain a cellular Markov partition for F. There are several other approaches to providing combinatorial models for certain classes of maps. For example, a postcritically-finite polynomial can be described by its Hubbard tree (see [**DH84**]) or a rational map with three critical values by a dessin d'enfant (see [**Gro97**] and [**LZ04**]). A very general setting that allows one to address similar questions is the recently developed theory of self-similar group actions. In this context one investigates algebraic objects such as the iterated monodromy group and the biset (or bimodule) defined for a Thurston map (see [**Ne05**], in particular Chapter 6).

Our approach is more geometric and adapted to Thurston maps. One of its main features is that we have good geometric control for the cells in the decompositions $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$ if $\mathcal{C}$ is f-invariant. In particular, with respect to any visual metric the curve $\mathcal{C}$ is actually a quasicircle (Theorem 15.3) and the boundaries of the tiles in $\mathcal{D}^n$ are quasicircles with uniform parameters independent of the level n (Proposition 15.26). For a rational expanding Thurston map $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ the tiles in $\mathcal{D}^n$ are in fact uniform quasidisks with respect to the chordal metric σ on $\widehat{\mathbb{C}}$ (Theorem 18.4 (iii)).

1.6. Miscellaneous results

In this section we collect various noteworthy results that may be useful for the orientation of the reader.

The concept of *Thurston equivalence* for Thurston maps was already mentioned before. We record its slightly technical definition in Section 2.4. At first sight the concept does not seem to be adapted to the dynamics under iteration. However, in Theorem 11.1 we will prove the important fact that two expanding Thurston maps are Thurston equivalent if and only if they are topologically conjugate.

In Chapter 9 we make a brief excursion to symbolic dynamics. The properties of visual metrics are essential for proving the following statement (see Chapter 9 for the relevant definitions).

THEOREM 9.1. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map. Then f is a factor of the left-shift $\Sigma\colon J^\omega \to J^\omega$ on the space J^ω of all sequences in a finite set J of cardinality $\#J = \deg(f)$.*

The proof of this theorem does not use invariant Jordan curves for f or its iterates; so in a sense it is independent of Theorem 15.1 mentioned above. It can be used to obtain another Markov partition for f, but we have very little control of the geometric shape of the "tiles".

An immediate consequence of this theorem and its proof is the fact that the periodic points of an expanding Thurston map $f\colon S^2 \to S^2$ form a dense subset of S^2 (Corollary 9.2).

In Chapter 13 we investigate equivalence relations $\sim$ on the sphere S^2, and the question when a Thurston map $f\colon S^2 \to S^2$ descends to a Thurston map on the quotient space $S^2/\sim$. Here we assume that $\sim$ is of *Moore-type* (see Definition 13.7) which implies that $S^2/\sim$ is a 2-sphere. Under this assumption the relevant condition is that $\sim$ is *strongly invariant* for f in the sense that f maps each equivalence class onto another equivalence class (see Definition 13.1 and Lemma 13.19). We will prove that for a given equivalence relation $\sim$ of Moore-type on S^2 a Thurston map $f\colon S^2 \to S^2$ descends to a Thurston map if and only if $\sim$ is strongly invariant for f (see Theorem 13.2 and Corollary 13.3).

Often it is desirable to promote a given Thurston map f that is not expanding to an expanding one. More precisely, we want to find an expanding Thurston map $\widetilde{f}$ that is Thurston equivalent to f. In general, the existence of $\widetilde{f}$ is not guaranteed. However, if f is *combinatorially expanding* (see Definition 12.4) such a map $\widetilde{f}$ does exist. Roughly speaking, it is constructed by defining an equivalence relation that collapses the sets where f fails to be expanding to points (see Chapter 14).

We will also investigate some measure-theoretic aspects of expanding Thurston maps. Each such map has a natural measure adapted to its dynamics.

THEOREM 17.1. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map. Then there exists a unique measure ν_f of maximal entropy for f. The map f is mixing for ν_f.*

This theorem follows from results due to Haïssinsky-Pilgrim [**HP09**, Theorem 3.4.1]. We will present a different proof and give an explicit

description of ν_f in terms of the cell decompositions $\mathcal{D}^n(F, \mathcal{C})$, where $F = f^n$ is a suitable iterate and $\mathcal{C}$ is an invariant curve as in Theorem 15.1. In particular, $\nu_f = \nu_F$ assigns equal mass to all tiles in the cell decompositions $\mathcal{D}^n(F, \mathcal{C})$ of a given "color" (see Proposition 17.11 and Theorem 17.12).

The measure ν_f can be used to study the topological and measure-theoretic dynamics of f under iteration. For example, we will see that $h_{top}(f) = \log(\deg(f))$, where $h_{top}(f)$ is the topological entropy and $\deg(f)$ the topological degree of f (Corollary 17.2).

If μ is a Borel measure on a metric space (X, d), then we call the metric measure space (X, d, μ) *Ahlfors Q-regular* for $Q > 0$ if

$$\mu(\overline{B}_d(x, r)) \asymp r^Q$$

for each closed ball $\overline{B}_d(x, r)$ in X whose radius r does not exceed the diameter of the space. If an expanding Thurston map has no periodic critical points, then its visual sphere together with its measure of maximal entropy has this property.

PROPOSITION 18.2. *Let $f \colon S^2 \to S^2$ be an expanding Thurston map without periodic critical points, ϱ be a visual metric for f with expansion factor $\Lambda > 1$, and ν_f be the measure of maximal entropy of f. Then the metric measure space (S^2, ϱ, ν_f) is Ahlfors Q-regular with*

$$Q := \frac{\log(\deg(f))}{\log(\Lambda)}.$$

In particular, (S^2, ϱ) has Hausdorff dimension Q and

$$0 < \mathcal{H}_\varrho^Q(S^2) < \infty.$$

Here $\mathcal{H}_\varrho^Q$ denotes Hausdorff Q-measure on the metric space (S^2, ϱ).

In Chapter 19 we delve into a deeper analysis of measure-theoretic properties of rational expanding Thurston maps. If we denote by $\mathcal{L}_{\widehat{\mathbb{C}}}$ Lebesgue measure on $\widehat{\mathbb{C}}$ normalized such that $\mathcal{L}_{\widehat{\mathbb{C}}}(\widehat{\mathbb{C}}) = 1$, we will establish the following (well-known) statement.

THEOREM 19.1. *Let $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a rational expanding Thurston map. Then Lebesgue measure $\mathcal{L}_{\widehat{\mathbb{C}}}$ is ergodic for f.*

Note that $\mathcal{L}_{\widehat{\mathbb{C}}}$ is essentially never f-invariant, but ergodicity is interpreted as for f-invariant measures (see the discussion in Section 17.1): if $A \subset \widehat{\mathbb{C}}$ is a Borel set with $f^{-1}(A) = A$ and $\mathcal{L}_{\widehat{\mathbb{C}}}(A) > 0$, then $\mathcal{L}_{\widehat{\mathbb{C}}}(A) = 1$.

In our context one can actually find an f-invariant measure that is absolutely continuous with respect to Lebesgue measure.

THEOREM 19.2. *Let $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a rational expanding Thurston map. Then there exists a unique f-invariant Borel probability measure λ_f on $\widehat{\mathbb{C}}$ that is absolutely continuous with respect to Lebesgue measure $\mathcal{L}_{\widehat{\mathbb{C}}}$. This measure has the form $d\lambda_f = \rho\, d\mathcal{L}_{\widehat{\mathbb{C}}}$, where ρ is a positive continuous function on $\widehat{\mathbb{C}} \backslash \mathrm{post}(f)$. Moreover, the measure λ_f is ergodic for f.*

Again this statement is essentially well known. We will prove it by interpreting the existence of λ_f as a fixed point problem for a suitable *Ruelle operator*. This is a standard technique in ergodic theory reviewed in Chapter 19.

1.7. Characterizations of Lattès maps

Lattès maps form another major theme in this book. One may define such a map as a rational expanding Thurston map with a parabolic orbifold (see Section 2.5 for the terminology). Equivalently, they are characterized as quotients of holomorphic torus endomorphisms, or as quotients of holomorphic automorphisms on the complex plane $\mathbb{C}$ by a crystallographic group (see Theorem 3.1). While these latter descriptions are more technical to state, they contain more information and allow us to construct all Lattès maps explicitly.

In Section 1.1 the Lattès map g was constructed from maps $A\colon \mathbb{C} \to \mathbb{C}$ and $\Theta\colon \mathbb{C} \to \widehat{\mathbb{C}}$. For these maps we have $g \circ \Theta = \Theta \circ A$, and so we obtain a commutative diagram as in (1.1). The push-forward of the Euclidean metric on $\mathbb{C}$ by Θ is the *canonical orbifold metric* ω of g. In this example, it is the path metric on the pillow in Figure 1.1. Moreover, ω is a visual metric for g. This is characteristic for Lattès maps.

To make this precise, we will introduce some terminology in an informal way (the details can be found in the body of the text). Each rational Thurston map $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ has an associated orbifold $\mathcal{O}_f$ (see Section 2.5), for which there is in turn a universal orbifold covering map $\Theta\colon X \to \widehat{\mathbb{C}}$ (see Section A.9). Here $X = \mathbb{C}$ or $X = \mathbb{D}$ depending on whether f has a parabolic or hyperbolic orbifold. The map Θ is holomorphic. The *canonical orbifold metric* ω of f is the push-forward of the Euclidean metric (in the case that $X = \mathbb{C}$) or hyperbolic metric (in the case that $X = \mathbb{D}$) by Θ (see Section A.10). The metric ω is a conformal metric on $\widehat{\mathbb{C}}$ (see Section A.1 for the terminology), and closely related to the chordal metric σ on $\widehat{\mathbb{C}}$. In fact, if f does not have periodic critical points, the metric space $(\widehat{\mathbb{C}}, \omega)$ is bi-Lipschitz equivalent to $(\widehat{\mathbb{C}}, \sigma)$ (see Lemma A.32). Note that all this is tied to a

holomorphic setting, and so we cannot define such a canonical metric ω unless the Thurston map f is rational.

PROPOSITION 8.5 (Canonical orbifold metric as visual metric). *Let* $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ *be a rational Thurston map without periodic critical points, and* ω *be the canonical orbifold metric for* f. *Then* ω *is a visual metric for* f *if and only if* f *is a Lattès map.*

In Theorem 18.1 (iii) we have already encountered another (much deeper) characterization of Lattès maps in terms of visual metrics.

It is not hard to see that for a Lattès map f the space $(\widehat{\mathbb{C}}, \omega)$ is Ahlfors 2-regular. The existence of a visual metric with this property again characterizes Lattès maps.

THEOREM 20.4. *Let* $f\colon S^2 \to S^2$ *be an expanding Thurston map. Then* f *is topologically conjugate to a Lattès map if and only if there is a visual metric* ϱ *for* f *such that* (S^2, ϱ) *is Ahlfors 2-regular.*

Together with Proposition 18.2 (mentioned above) the previous theorem implies that an expanding Thurston map without periodic critical points is topologically conjugate to a Lattès map if and only if there is a visual metric with expansion factor $\Lambda = \deg(f)^{1/2}$ (see Corollary 20.5).

Recall from Theorem 16.3 that for an expanding Thurston map f the supremum of all expansion factors of visual metrics is given by the combinatorial expansion factor $\Lambda_0(f)$. This number was defined in (1.6) as the limit of $D_n(f, \mathcal{C})^{1/n}$ as $n \to \infty$. Here $D_n(f, \mathcal{C})$ is the minimal number of n-tiles that are needed to form a connected set joining opposite sides of $\mathcal{C}$. We will show in Proposition 20.1 that $D_n(f, \mathcal{C}) \lesssim \deg(f)^{n/2}$. Lattès maps are precisely those expanding Thurston maps for which this maximal growth rate for $D_n(f, \mathcal{C})$ is attained.

THEOREM 20.2. *Let* $f\colon S^2 \to S^2$ *be an expanding Thurston map. Then* f *is topologically conjugate to a Lattès map if and only if the following conditions are true:*

   (i) f *has no periodic critical points.*
   (ii) *There exists* $c > 0$, *and a Jordan curve* $\mathcal{C} \subset S^2$ *with* $\mathrm{post}(f) \subset \mathcal{C}$ *such that*
$$D_n(f, \mathcal{C}) \geq c \deg(f)^{n/2}$$
   for all $n \in \mathbb{N}_0$.

This theorem is due to Qian Yin [**Yi15a**]. It is remarkable that one can characterize the conformal dynamical systems given by iteration of Lattès maps in terms of essentially combinatorial data.

An immediate consequence of the maximal growth rate of $D_n(f, \mathcal{C})$ is the inequality $\Lambda_0(f) \leq \deg(f)^{1/2}$ for each expanding Thurston map. Here equality is attained for Lattès maps, and so one might expect that this property again characterizes Lattès maps similar to Theorem 20.2. However, there are Thurston maps f satisfying $\Lambda_0(f) = \deg(f)^{1/2}$ that are not topologically conjugate to a Lattès map (see Example 16.8).

The proof of Theorem 20.2 relies on Theorem 20.4, which in turn depends on yet another characterization of Lattès maps. For this we compare the Lebesgue measure $\mathcal{L}_{\widehat{\mathbb{C}}}$ on $\widehat{\mathbb{C}}$ with the measure of maximal entropy ν_f for a given rational Thurston map $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$. As the following result shows, ν_f and $\mathcal{L}_{\widehat{\mathbb{C}}}$ lie in different measure classes unless f is a Lattès map.

THEOREM 19.4. *Let* $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ *be a rational expanding Thurston map. Then its measure of maximal entropy* ν_f *is absolutely continuous with respect to Lebesgue measure* $\mathcal{L}_{\widehat{\mathbb{C}}}$ *if and only if* f *is a Lattès map.*

This is a special case of a more general theorem due to Zdunik [**Zd90**] which gives a similar characterization of Lattès maps among all, not only postcritically-finite, rational maps. Crucial in the proof of Theorem 19.4 is the existence of the f-invariant measure λ_f on $\widehat{\mathbb{C}}$ that is absolutely continuous with respect to Lebesgue measure $\mathcal{L}_{\widehat{\mathbb{C}}}$ (see Theorem 19.2 mentioned above). In fact, for a Lattès map f the measures λ_f, ν_f, and the *canonical orbifold measure* Ω_f (see Section A.10) all agree (Theorem 19.3). In the example g from Section 1.1 these measures are given by the Euclidean area measure on the pillow in Figure 1.1 (normalized to be a probability measure).

1.8. Outline of the presentation

Our work is an introduction to the subject. We hope that it will stimulate more research in the area and will serve as a foundation for future investigations. Therefore, we kept our presentation elementary, as self-contained as possible, and rather detailed.

For the most part, the prerequisites for the reader are modest and include some basic knowledge of complex analysis and topology, in particular plane topology and the topology of surfaces. A background in complex dynamics is helpful, but not absolutely necessary. In later chapters our demands on the reader are more substantial. In particular, in Chapter 17 and Chapter 19 we require some concepts and results from topological and measure-theoretic dynamics, but we will state and review the necessary facts.

When writing this book, we were faced with conflicting objectives. On the one hand, it was desirable to present the material in linear order, meaning that we should only use results in an argument that have been discussed or established before. On the other hand, a too rigid implementation of this idea would have resulted in long detours that might have distracted the reader from the subject at hand. Moreover, we often had to invoke results that are "well known", but difficult to find in the required form in the literature. For this reason we have included an appendix, where such results are collected. We use and refer to the appendix throughout the text.

We will now give an outline of how of this book is organized and will briefly discuss some main concepts and ideas.

In the last section of this introduction the reader can find a list of all examples of Thurston maps that we consider in this work (Section 1.9).

In Chapter 2 we turn to Thurston maps, the main object of our investigation. We first review branched covering maps in Sections 2.1 and then define Thurston maps in Section 2.2. Our notion of *expansion* is introduced in Section 2.3 (see Definition 2.2). In this section we also give a characterization when a rational Thurston map is expanding. While the concept of expansion is discussed more systematically later (in Chapter 6), readers familiar with complex dynamics may find it helpful to get some perspective early on. The notion of *(Thurston) equivalence* for Thurston maps is discussed in Section 2.4.

Every Thurston map f has an associated *orbifold* $\mathcal{O}_f$, defined in terms of the *ramification function* of f. This orbifold can be parabolic in some exceptional cases (including Lattès maps) and is hyperbolic otherwise (see Section 2.5). For rational Thurston maps the *universal orbifold covering map* induces a natural metric (the *canonical orbifold metric*) and a natural measure (the *canonical orbifold measure*). As we did not want to overburden the reader with technicalities at this early stage, we delegated a detailed discussion of these topics to the appendix (see Section A.9 and Section A.10).

As already mentioned, Thurston gave a characterization when a Thurston map is equivalent to a rational map. We will review this without proofs in Section 2.6. This material is not really essential for the rest of the book, but we included this discussion for general background.

In Chapter 3 we discuss a large class of Thurston maps, namely Lattès maps and a related class that we call Lattès-type maps. Lattès-type maps are quotients of torus endomorphisms, but in contrast to the Lattès case we do not require the endomorphism to be holomorphic.

We will classify Lattès maps, which is surprisingly involved, and will discuss many examples.

In Chapter 4 we collect facts from quasiconformal geometry (Section 4.1) and from the theory of Gromov hyperbolic spaces (Section 4.2) that will be relevant later on. We then turn to Cannon's conjecture in geometric group theory (Section 4.3). As we already remarked earlier, this conjecture gives an intriguing analog to some of the main themes of our study of expanding Thurston maps. We illustrate this with an explicit description of some examples of fractal 2-spheres that arise as visual spheres of expanding Thurston maps (Section 4.4). We included Sections 4.3 and 4.4 mainly to give some motivating background for our investigation.

This starts in earnest in Chapter 5, the technical core of our combinatorial approach. Here we discuss cell decompositions and their relation to Thurston maps. In Section 5.1 we collect some general (well-known) facts about cell decompositions, including the definition of a cell decomposition and related concepts such as refinements and cellular maps. In Section 5.2 we specialize to cell decompositions on 2-spheres.

In Section 5.3 we consider cell decompositions induced by a Thurston map f. Here we define cell decompositions $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$ for each level $n \in \mathbb{N}_0$ from a Jordan curve $\mathcal{C} \subset S^2$ with $\operatorname{post}(f) \subset \mathcal{C}$ as outlined in Section 1.2. These cell decompositions are our most important technical tool for studying Thurston maps. Their properties are summarized in Proposition 5.17.

Given such a sequence $\mathcal{D}^n$ of cell decompositions induced by (the iterates of) a Thurston map f, we may *label* the cells in $\mathcal{D}^n$ to record to which cells in $\mathcal{D}^0$ they are mapped by f^n. This is explained in Section 5.4. By using two cell decompositions $\mathcal{D}^0$ and $\mathcal{D}^1$ and a labeling (satisfying some additional assumptions), it is possible to construct a Thurston map f that *realizes* this data in a suitable way (see Proposition 5.26). Roughly speaking, this means that we may construct Thurston maps in a geometric fashion, very similar to the example indicated in Figure 1.1.

In Section 5.6 we introduce the concept of an *n-flower* $W^n(p)$ of a vertex p in the cell decomposition $\mathcal{D}^n$. The set $W^n(p)$ is formed by the interiors of all cells in $\mathcal{D}^n$ that meet p (see Definition 5.27 and Lemma 5.28). An important fact is that while in general a component of the preimage $f^{-n}(K)$ of a small connected set K will not be contained in an n-tile (i.e., a 2-dimensional cell in $\mathcal{D}^n$), it is always contained in an n-flower (Lemma 5.34).

In Section 5.7 we give a precise definition for a connected set *joining opposite sides* of $\mathcal{C}$. In addition, we define the quantity $D_n = D_n(f, \mathcal{C})$ that measures the combinatorial expansion rate of a Thurston map. It is given as the minimal number of n-tiles needed to form a connected set joining opposite sides of $\mathcal{C}$ (see (5.14)).

In Chapter 6 we revisit our notion of expansion. The main result here is Proposition 6.3, which gives several equivalent conditions for a Thurston map to be expanding. In particular, it follows that expansion is a topological condition, and does not depend on the choice of metric on S^2 used in the definition. Section 6.2 collects some additional results about expansion, and Section 6.3 provides a simple criterion when a Lattès-type map is expanding.

In Chapter 7 we consider Thurston maps with two or three post-critical points. Every such map is equivalent to a rational map (see Theorem 7.3). In fact, every Thurston map f with $\#\operatorname{post}(f) = 2$ is equivalent to $z \mapsto z^n$ (for $n \in \mathbb{Z} \setminus \{-1, 0, 1\}$) (see Proposition 7.1). In Section 7.2 we consider Thurston maps with an associated parabolic orbifold of signature (∞, ∞) or $(2, 2, \infty)$. Such a map is equivalent to $z \mapsto z^n$ in the case (∞, ∞), or to a *Chebyshev polynomial* in the case $(2, 2, \infty)$ (up to a sign). This completes the classification of Thurston maps with parabolic orbifold begun in Chapter 3.

In Chapter 8 we introduce *visual metrics* for expanding Thurston maps, one of our central concepts, as outlined in Section 1.4. Basic properties of visual metrics are listed in Proposition 8.3. A characterization (already mentioned in Section 1.4) is given in Proposition 8.4.

If an expanding Thurston map f is a rational map on the Riemann sphere $\widehat{\mathbb{C}}$, then one would like to know whether some natural metrics on $\widehat{\mathbb{C}}$ are visual metrics for f. The chordal metric on $\widehat{\mathbb{C}}$ is never a visual metric. The canonical orbifold metric of f is visual if and only if f is a Lattès map. This is discussed in Section 8.3.

In Chapter 9 we briefly turn to symbolic dynamics and show that every expanding Thurston map is a factor of a shift operator (see Theorem 9.1).

In Chapter 10 we connect our development of expanding Thurston maps to the theory of *Gromov hyperbolic spaces*. We define the *tile graph* $\mathcal{G} = \mathcal{G}(f, \mathcal{C})$ associated with an expanding Thurston map $f \colon S^2 \to S^2$ and a Jordan curve $\mathcal{C} \subset S^2$ with $\operatorname{post}(f) \subset S^2$. The graph $\mathcal{G}(f, \mathcal{C})$ only depends on $\mathcal{C}$ up to rough-isometry (Theorem 10.4). We show that it is Gromov hyperbolic (Theorem 10.1) and that its boundary at infinity $\partial_\infty \mathcal{G}$ can be identified with S^2 (Theorem 10.2). Under this identification a metric ϱ on $S^2 \cong \partial_\infty \mathcal{G}$ is visual in the sense of

Gromov hyperbolic spaces if and only if it is visual for f as defined in Chapter 8 (see Theorem 10.2). This is the reason why we chose the term "visual" for the metrics associated with a Thurston map.

In Chapter 11 we consider *isotopies* on S^2 and their lifts by Thurston maps. We show that if two expanding Thurston maps are Thurston equivalent, then they are in fact topologically conjugate (Theorem 11.1). We also prove some results on isotopies of Jordan curves (Section 11.2). The subsequent Section 11.3 contains some auxiliary statements on graphs. The main result is the important, but rather technical Lemma 11.17, which gives a sufficient criterion when a Jordan curve can be isotoped into the 1-skeleton of a given cell decomposition of a 2-sphere.

In Chapter 12 we study the cell decompositions $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$ under the additional assumption that $\mathcal{C}$ is f-invariant. Then $\mathcal{D}^{n+k}$ is a refinement of $\mathcal{D}^n$ for all $n, k \in \mathbb{N}_0$; so each cell in any of the cell decompositions $\mathcal{D}^n$ is "subdivided" by cells of a higher level. Moreover, the pair $(\mathcal{D}^{n+k}, \mathcal{D}^n)$ is a cellular Markov partition for f^k (Proposition 12.5). In this case the Thurston map f can be described by a *two-tile subdivision rule* (Definition 12.1) as discussed in Section 12.2. Conversely, we may construct a Thurston map from a two-tile subdivision rule by Proposition 12.3. This is the main result in this chapter. This way to construct Thurston maps from a combinatorial viewpoint is illustrated in Section 12.3, where we consider many examples of Thurston maps given in this form.

In Chapter 13 we study the question when a Thurston map $f\colon S^2 \to S^2$ descends to a Thurston map on the quotient space $S^2/\!\sim$ obtained from an equivalence relation $\sim$ on S^2. For this, we first review closed equivalence relations and Moore's theorem in Section 13.1. We also require some auxiliary results on the mapping behavior of branched covering maps discussed in Section 13.2. We prove the main result of this chapter in Section 13.3: a Thurston map $f\colon S^2 \to S^2$ descends to a Thurston map on the quotient $S^2/\!\sim$ obtained from an equivalence relation $\sim$ on S^2 of Moore-type (see Definition 13.7) if and only if $\sim$ is strongly f-invariant (see Definition 13.1, Theorem 13.2, and Corollary 13.3).

Can a given two-tile subdivision rule be realized by an *expanding* Thurston map? This question is addressed in Chapter 14. A necessary condition is that the subdivision rule is *combinatorially expanding* (see Definition 12.18 and Definition 12.4). We show that every combinatorially expanding Thurston map is equivalent to an expanding Thurston map (Proposition 14.3 and Theorem 14.2).

To prove this statement, we "correct" a Thurston map $f\colon S^2 \to S^2$ that realizes a combinatorially expanding two-tile subdivision rule so that the map becomes expanding. On an intuitive level, it is very plausible that this should be possible (see the discussion at the beginning of Chapter 14), but a rigorous implementation is somewhat cumbersome. For this we define an equivalence relation $\sim$ on S^2 that essentially collapses components where the map fails to be expanding to single points. We show that $\sim$ is of Moore-type and will obtain a suitable Thurston map on the quotient $S^2/\!\sim$.

Existence and uniqueness results for *invariant Jordan curves* are proved in Chapter 15. It is one of the central chapters of the present work. Here we establish Theorems 15.1, 15.4, and 15.5, and Corollary 15.2 about existence and uniqueness of invariant curves mentioned in Section 1.5. One can obtain invariant curves from an iterative procedure discussed in detail in Section 15.2.

In Section 15.3 we prove that a Jordan curve $\mathcal{C}$ is a *quasicircle* if it is invariant for an expanding Thurston map f (Theorem 15.3). If the cell decompositions $\mathcal{D}^n(f,\mathcal{C})$, $n \in \mathbb{N}_0$, are obtained from such an f-invariant Jordan curve $\mathcal{C}$, then the edges in these cell decompositions are uniform quasiarcs and the boundaries of tiles are uniform quasicircles (Proposition 15.26). The underlying metric in all these statements is any visual metric for f.

In Chapter 16 we revisit visual metrics. We introduce the combinatorial expansion factor $\Lambda_0(f)$ and prove Theorem 16.3 (see the outline in Section 1.4). In its proof we use the invariant curves constructed in Chapter 15 to obtain particularly nice visual metrics for a given expanding Thurston map f. They have the property that the map f expands distances locally by the constant factor Λ (see (16.1)).

Chapter 17 is devoted to the measure-theoretic dynamics of an expanding Thurston map f. The main result is Theorem 17.1 about existence and uniqueness of a measure of maximal entropy ν_f for f. For the convenience of the reader we review (mostly standard) material from measure-theoretic dynamics in Section 17.1.

The geometry of the visual sphere (S^2, ϱ) of an expanding Thurston map $f\colon S^2 \to S^2$ is explored in Chapter 18, another central part of our work. Here we show the important fact—already discussed in Section 1.4—that f is conjugate to a rational map if and only if (S^2, ϱ) is quasisymmetrically equivalent to the standard 2-sphere (see Theorem 18.1 (ii)). In addition, we prove linear local connectivity of the visual sphere (Proposition 18.5), as well as its Ahlfors regularity in the absence of periodic critical points of the map (Proposition 18.2).

In Chapter 19 we study measure-theoretic properties of rational expanding Thurston maps $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$. We first construct a measure λ_f on $\widehat{\mathbb{C}}$ that is f-invariant and absolutely continuous with respect to Lebesgue measure on $\widehat{\mathbb{C}}$ (Theorem 19.2). This allows us to apply methods from ergodic theory. As a consequence we recover (a weak form of) Zdunik's result that the measure of maximal entropy of f is absolutely continuous with respect to Lebesgue measure if and only if f is a Lattès map (Theorem 19.4).

This in turn allows us to finish the discussion of the visual sphere of an expanding Thurston map f begun in Chapter 18. Specifically, in Section 19.4 we prove that the visual sphere of such a map f is snowflake equivalent to the standard 2-sphere if and only if f is topologically conjugate to a Lattès map (Theorem 18.1 (iii)).

Chapter 20 gives another application of (Zdunik's) Theorem 19.4. By using this theorem we show that Lattès maps can be characterized in terms of their combinatorial expansion behavior (see Theorem 20.2 mentioned in Section 1.6).

Some further developments and future perspectives are presented in Chapter 21. Here we discuss some recent related work and open problems.

The appendix is devoted to several subjects whose inclusion in the main text would have been too distracting. For example, in one of its sections we establish a useful variant of Janiszewski's theorem in plane topology, whose proof is a bit technical. We also review some fairly standard material about conformal metrics, Koebe's distortion theorem, orientation on surfaces, covering maps, lattices and tori, and quotient spaces. We discuss these topics so that we can refer to them in the main text and make this work as self-contained as possible.

The appendix also contains fairly lengthy sections on branched covering maps, orbifolds, and the canonical orbifold metric. Though the expert will find no surprises here, it is hard to track down this material in an accessible form in the literature.

1.9. List of examples for Thurston maps

Throughout the book we consider many examples of Thurston maps in order to illustrate various phenomena. We list them here with a short description for the reader's convenience. The relevant terms used in these descriptions are defined in later chapters. Often the maps in our examples are Lattès maps. While Lattès maps sometimes have special properties compared to general Thurston maps, they often provide convenient examples with generic behavior.

A Lattès map is discussed in Section 1.1. In the terminology of Chapter 3 it is a flexible Lattès map.

In Section 1.3 we consider an expanding Thurston map h that "generates" a fractal sphere $\mathcal{S}$. The map h is not topologically conjugate (or Thurston equivalent) to a rational map. This is examined in Example 2.18. Closely related is the fact that the fractal sphere $\mathcal{S}$ is not quasisymmetrically equivalent to the standard sphere (see Example 4.7).

The examples $f(z) = 1 - 2/z^2$ and $g(z) = \frac{i}{2}(z + 1/z)$ are used in Section 2.2 to explain the concept of a ramification portrait. Both are in fact Lattès maps.

In Example 2.6 our main purpose is to familiarize the reader with the concept of Thurston equivalence. To this end we consider two Thurston maps f and g and show that they are Thurston equivalent. While the map g is given in a combinatorial fashion and realizes a certain two-tile subdivision rule, the map f is rational and given by an explicit formula.

In Section 3.5 we look at several examples of Lattès maps. In Example 3.23 the map is $f(z) = 1 - 2/z^2$, which is a Lattès map with orbifold signature $(2, 4, 4)$. Similarly, in Example 3.24 and Example 3.25 we have Lattès maps with orbifold signatures $(3, 3, 3)$ and $(2, 3, 6)$, respectively.

Certain types of Lattès maps are called *flexible*; see Definition 3.26 and (3.38) for an example. They have orbifold signature $(2, 2, 2, 2)$, but not all Lattès maps with this signature are flexible. Such a non-flexible Lattès map with signature $(2, 2, 2, 2)$ is given in Example 3.27.

The Thurston map in Example 6.10 has a Levy cycle.

In Example 6.14 the Thurston map is eventually onto, but not expanding.

In Section 7.2 we consider Thurston maps with signatures (∞, ∞) and $(2, 2, \infty)$. They are equivalent to $z \mapsto z^n$ (where $n \in \mathbb{Z} \setminus \{-1, 0, 1\}$) and to *Chebyshev polynomials* (up to sign), respectively.

The example $f(z) = i(z^4 - i)/(z^4 + i)$ is used in Figure 8.1 to illustrate the definition of a visual metric.

In Example 12.6 we show some cell decompositions generated by an f-invariant curve for the map $f(z) = 1 - 2/z^4$.

In Example 12.11 we consider two maps that realize two-tile subdivision rules that only differ by the labeling.

Several examples of two-tile subdivision rules and the Thurston maps realizing them are discussed in Section 12.3. The map in Example 12.20 is $f_1(z) = z^2 - 1$; it realizes a two-tile subdivision rule that is not combinatorially expanding.

The two maps f_2 and $\widetilde{f_2}$ in Example 12.21 both realize the barycentric subdivision rule. The map f_2 is a rational map, but it is not expanding (i.e., its Julia set is not the whole Riemann sphere $\widehat{\mathbb{C}}$). However, the map $\widetilde{f_2}$ is expanding. It is an example of an expanding Thurston map with periodic critical points.

The map f_3 in Example 12.22 is the map h considered in Section 1.3. It realizes a certain two-tile subdivision rule and is an obstructed map. This means that f_3 is not Thurston equivalent to a rational map.

The map f_4 in Example 12.23 is a Lattès-type map that is again not Thurston equivalent to a rational map. While it is somewhat easier to define than the map f_3 in Example 12.22, it is less generic, since f_4 has a parabolic orbifold, while f_3 has a hyperbolic orbifold. The map f_4 realizes the 2-by-3 subdivision rule. If we equip the underlying sphere S^2 with a suitable visual metric for f_4, then S^2 consists of two copies of Rickman's rug.

Example 12.24 provides a whole class of Thurston maps. One of them is $f_5(z) = 1 - 2/z^2$ which realizes a simple two-tile subdivision rule. By "adding flaps" we obtain the other maps. All these maps are rational; in fact, they are given by an explicit formula, which makes them easy to understand and visualize.

In Example 13.17 we consider an equivalence relation of Moore-type on $\widehat{\mathbb{C}}$ such that the map $z \mapsto z^2$ does not descend to a branched covering map. In contrast, the equivalence relation on $\widehat{\mathbb{C}}$ in Example 13.18 is not of Moore-type, but $z \mapsto z^2$ descends to a Thurston map on the quotient.

Example 14.22 shows why the additional condition $\mathrm{post}(f) = \mathbf{V}^0$ in Theorem 14.1 is necessary.

The Thurston map f in Example 14.23 is not combinatorially expanding, yet Thurston equivalent to an expanding Thurston map g. This shows that the sufficient condition in Proposition 14.3 is not necessary.

The map f in Example 15.6 is the same as in Example 2.6. We use it to outline the main ideas of Chapter 15. In particular, we show how to construct an f-invariant curve $\widetilde{\mathcal{C}}$ with $\mathrm{post}(f) \subset \widetilde{\mathcal{C}}$ (see Figure 15.1).

In Example 15.9 we return to the Lattès map g from Section 1.1 and prove that it has infinitely many distinct g-invariant curves $\mathcal{C}$ with $\mathrm{post}(g) \subset \mathcal{C}$.

In Example 15.11 we exhibit an expanding Thurston map f for which no f-invariant Jordan curve $\mathcal{C}$ with $\mathrm{post}(f) \subset \mathcal{C}$ exists.

Remark 15.16 justifies why the f^n-invariant curve $\widetilde{\mathcal{C}}$ given by Theorem 15.1 will in general depend on n.

In Example 15.17 we use another Lattès map to illustrate an iterative construction of invariant curves (see Figure 15.4). The invariant curve obtained is quite "fractal" (its Hausdorff dimension is > 1).

Example 15.23 shows what can happen if one of the necessary conditions in the iterative procedure for producing invariant curve is violated. Namely, the limiting object $\widetilde{\mathcal{C}}$ is not a Jordan curve anymore. The map here is again a Lattès map.

The map in Example 15.24 (yet another Lattès map) has a nontrivial (in particular non-smooth) invariant curve that is rectifiable.

In Example 16.8 we exhibit an expanding Thurston map f that has no visual metric with an expansion factor Λ equal to its combinatorial expansion factor $\Lambda_0(f)$. Therefore, statement (ii) in Theorem 16.3 cannot be improved in general.

In Example 18.11 we revisit the two maps from Example 12.21 that realize the barycentric subdivision rule; these maps show that in Theorem 18.1 (ii) we cannot replace "topologically conjugate to a rational map" by "Thurston equivalent to a rational map".

CHAPTER 2

Thurston maps

Here we set the stage for the subsequent developments. We define
Thurston maps and what it means for such a map to be expanding.
We also explain several closely related notions, namely Thurston equiv-
alence and the orbifold associated with a Thurston map. In the last
section of this chapter we briefly discuss Thurston's characterization of
rational maps among Thurston maps.

2.1. Branched covering maps

Branched covering maps are modeled on (non-constant) holomor-
phic maps between Riemann surfaces. As our immediate purpose in
this section is to prepare the definition of a Thurston map, we will
discuss only some basic facts on this topic and relegate more technical
aspects to the appendix (see Section A.6 in particular).

We call a 2-dimensional connected manifold (without boundary) a
surface. All surfaces that we consider will be orientable, and we will
assume that some fixed orientation has been chosen on such a surface
(for a review of orientation and related concepts see Section A.4). A
surface is called a *topological disk* if it is homeomorphic to the open
unit disk $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ in the complex plane.

In the following, X and Y are two compact and connected oriented
surfaces. If $f \colon X \to Y$ is a continuous and surjective map, then f is
called a *branched covering map* if we can write it locally as the map $z \mapsto$
z^d for some $d \in \mathbb{N}$ in orientation-preserving homeomorphic coordinates
in domain and target. More precisely, we require that for each point
$p \in X$ there exist $d \in \mathbb{N}$, topological disks $U \subset X$ and $V \subset Y$ with
$p \in U$ and $q := f(p) \in V$, and orientation-preserving homeomorphisms
$\varphi \colon U \to \mathbb{D}$ and $\psi \colon V \to \mathbb{D}$ with $\varphi(p) = 0$ and $\psi(q) = 0$ such that

$$(2.1) \qquad\qquad (\psi \circ f \circ \varphi^{-1})(z) = z^d$$

for all $z \in \mathbb{D}$. This means that the following diagram commutes:

$$
\begin{array}{ccc}
p \in U \subset X & \xrightarrow{\ \ f\ \ } & q \in V \subset Y \\
{\scriptstyle \varphi} \downarrow & & \downarrow {\scriptstyle \psi} \\
0 \in \mathbb{D} & \xrightarrow{\ z \mapsto z^d\ } & 0 \in \mathbb{D}.
\end{array}
$$

For the concept of a branched covering map between non-compact surfaces see Definition A.7. If $f \colon X \to Y$ is a branched covering map, then one can find conformal structures on the surfaces X and Y so that f becomes a holomorphic map (see Lemma A.11). In this way, one can often derive statements for branched covering maps (as in the ensuing discussion) from analogous statements for holomorphic maps.

The integer $d \geq 1$ in (2.1) is uniquely determined by f and p, and called the *local degree* of the map f at p, interchangeably denoted by $\deg_f(p)$ or $\deg(f, p)$ (depending on whether our emphasis is on the point p or the map f). A point $c \in X$ with $\deg_f(c) \geq 2$ is called a *critical point* of f, and a point in Y that has a critical point as a preimage a *critical value*. The set of all critical points of f is denoted by $\operatorname{crit}(f)$. Obviously, if f is a branched covering map on X, then $\operatorname{crit}(f)$ is *discrete in X*, i.e., it has no limit points in X. Hence $\operatorname{crit}(f)$ is a finite set, because X is assumed to be compact. Moreover, f is *open* (images of open sets are open), and *finite-to-one* (every point in Y has finitely many preimages under f). Actually, if $d = \deg(f)$ is the topological degree of f (see Section A.4), then

$$
(2.2) \qquad\qquad \sum_{p \in f^{-1}(q)} \deg_f(p) = d
$$

for every $q \in Y$ (see [**Ha02**, Section 2.2]). In particular, if q is not a critical value of f, then q has precisely d preimages.

We denote by $\chi(X)$ the *Euler characteristic* of a compact surface X (see, for example, [**Ha02**, Theorem 2.44]). If $X = S^2$ is a 2-sphere, then $\chi(S^2) = 2$. The only other case relevant for us is if $X = T^2$ is a 2-dimensional torus, in which case $\chi(T^2) = 0$. The degrees of a branched covering map $f \colon X \to Y$ at critical points are related to the Euler characteristics of the surfaces by the *Riemann-Hurwitz formula* (see, for example, [**Hu06**, Theorem A3.4]); namely,

$$
(2.3) \qquad \chi(X) + \sum_{c \in \operatorname{crit}(f)} (\deg_f(c) - 1) = \deg(f) \cdot \chi(Y).
$$

Suppose Z is another compact and connected oriented surface, and $f \colon X \to Y$ and $g \colon Y \to Z$ are branched covering maps. Then $g \circ f$ is

also a branched covering map (see Lemma A.15 (i)) and

$$(2.4) \qquad \deg(g \circ f, x) = \deg(g, f(x)) \cdot \deg(f, x)$$

for all $x \in X$ (see Lemma A.16). Moreover, by counting the number of preimages of a point that is not a critical value of $g \circ f$ we see that

$$\deg(g \circ f) = \deg(g) \cdot \deg(f).$$

2.2. Definition of Thurston maps

We will be mostly interested in the case of branched covering maps on a 2-sphere S^2. We use notation S^2 for such a sphere, because we think of it as purely topological object that should be distinguished from the Riemann sphere $\widehat{\mathbb{C}}$ or other 2-spheres with additional structure (such as a Riemann surface structure); in other words, S^2 indicates a topological space homeomorphic to $\widehat{\mathbb{C}}$. We call the topology on S^2 the *given* topology. Sometimes it is convenient to express topological notation on S^2 in metric terms. Then we choose a *base metric* on S^2 that induces the given topology on S^2. Such a base metric can be obtained, for example, by pulling back the chordal metric on $\widehat{\mathbb{C}}$ by a homeomorphism from S^2 onto $\widehat{\mathbb{C}}$.

Branched covering maps on S^2 are topologically not very different from rational maps on the Riemann sphere $\widehat{\mathbb{C}}$, because one can show that if $f \colon S^2 \to S^2$ is a branched covering map, then there exist homeomorphisms $\psi \colon S^2 \to \widehat{\mathbb{C}}$ and $\varphi \colon S^2 \to \widehat{\mathbb{C}}$ such that $R = \psi \circ f \circ \varphi^{-1}$ is a rational map on $\widehat{\mathbb{C}}$ (see Corollary A.12 for a slightly stronger statement). In other words, these maps are not only locally modeled on holomorphic maps as in (2.1), but even globally. In general though, a branched covering map $f \colon S^2 \to S^2$ is not topologically conjugate (see Section 2.4) to a rational map, i.e., the homeomorphisms φ and ψ above will be different. So these maps may exhibit behavior under iteration different from rational maps.

We denote by f^n the n-th iterate of f for any $n \in \mathbb{N}_0$ (where $f^0 := \mathrm{id}_{S^2}$). If f is a branched covering map on S^2, then the same is true for f^n and we have $\deg(f^n) = \deg(f)^n$.

A *postcritical point* of f is a point $p \in S^2$ of the form $p = f^n(c)$ with $n \in \mathbb{N}$ and $c \in \mathrm{crit}(f)$. So the set of postcritical points of f is given by

$$\mathrm{post}(f) := \bigcup_{n \geq 1} \{f^n(c) : c \in \mathrm{crit}(f)\}.$$

If the cardinality $\# \operatorname{post}(f)$ is finite, then f is called *postcritically-finite.* This is equivalent to the requirement that the *orbit* $\{f^n(c) : n \in \mathbb{N}_0\}$ of each critical point c of f is finite.

For $n \in \mathbb{N}$ we have

$$(2.5) \qquad \operatorname{crit}(f^n) = \operatorname{crit}(f) \cup f^{-1}(\operatorname{crit}(f)) \cup \cdots \cup f^{-(n-1)}(\operatorname{crit}(f)).$$

This implies that $\operatorname{post}(f^n) = \operatorname{post}(f)$ and

$$(2.6) \qquad\qquad \operatorname{post}(f) = \bigcup_{n \in \mathbb{N}} f^n(\operatorname{crit}(f^n)).$$

So $\operatorname{post}(f)$ is equal to the union of the critical values of all iterates f^n, $n \in \mathbb{N}$. It follows that $f^n \colon S^2 \setminus f^{-n}(\operatorname{post}(f)) \to S^2 \setminus \operatorname{post}(f)$ is a covering map (see Section A.5 for a review of this topic) and away from $\operatorname{post}(f)$ all "branches of the inverse of f^n" are defined; more precisely, if $U \subset S^2 \setminus \operatorname{post}(f)$ is a path-connected and simply connected set, $q \in U$, and $p \in S^2$ a point with $f^n(p) = q$, then there exists a unique continuous map $g \colon U \to S^2$ with $g(q) = p$ and $f^n \circ g = \operatorname{id}_U$ (this easily follows from Lemma A.6). We refer to such a right inverse of f^n informally as a "branch of f^{-n}".

We can now record the definition of the main object of investigation in this work.

DEFINITION 2.1 (Thurston maps). A *Thurston map* is a branched covering map $f \colon S^2 \to S^2$ of a 2-sphere S^2 with $\deg(f) \geq 2$ and a finite set of postcritical points.

Note that for a branched covering map $f \colon S^2 \to S^2$ the condition $\deg(f) \geq 2$ is equivalent to the requirement that f is not a homeomorphism.

Away from its finitely many critical points, a Thurston map is an orientation-preserving local homeomorphism on the oriented sphere S^2 (see the end of Section A.4 for the relevant terminology here). Since $\operatorname{post}(f^n) = \operatorname{post}(f)$, each iterate f^n, $n \in \mathbb{N}$, of a Thurston map is also a Thurston map.

If $S^2 = \widehat{\mathbb{C}}$ is the Riemann sphere and $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is a non-constant holomorphic map, then f is a rational map (and can be represented as the quotient of two polynomials). If, in addition, f is a Thurston map (i.e., it is postcritically-finite and satisfies $\deg(f) \geq 2$), then we call f a *rational Thurston map.* Note that if $f = P/Q$, where $P, Q \neq 0$ are polynomials without common zero, then $\deg(f) = \max\{\deg(P), \deg(Q)\}$. Here the degree $\deg(P)$ of a polynomial $P \neq 0$ is equal to $n \in \mathbb{N}_0$ if $P(z) = a_n z^n + \cdots + a_0$ with $a_0, \ldots, a_n \in \mathbb{C}$ and $a_n \neq 0$. This agrees with the topological degree of P considered as a map $P \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$.

There are no Thurston maps with $\#\operatorname{post}(f) \in \{0,1\}$ (see Remark 5.16), and all Thurston maps with $\#\operatorname{post}(f) = 2$ are Thurston equivalent (see Section 2.4) to a map $z \mapsto z^k$, $k \in \mathbb{Z} \setminus \{-1,0,1\}$, on the Riemann sphere (see Proposition 7.1). Postcritically-finite rational maps give a large class of Thurston maps. Examples include $P(z) = z^2 - 1$ (commonly known as the *Basilica map*, since its Julia set supposedly resembles St Mark's Basilica reflected in the water), $f(z) = 1 - 2/z^2$, or $g(z) = i/2(z + 1/z)$. Many other examples can be found throughout this work (see also [**B–P00**]).

Since the orbit of each critical point of a Thurston map f is finite, it is often convenient to represent such orbits by the *ramification portrait*. This is a directed graph, where the vertex set V is the union of the orbits of all critical points. For $p, q \in V$ there is a directed edge from p to q if and only if $f(p) = q$. Moreover, if p is a critical point with $\deg_f(p) = d$ we label this edge by "$d : 1$".

For example, the map $f(z) = 1 - 2/z^2$ has the ramification portrait

$$0 \xrightarrow{\;2:1\;} \infty \xrightarrow{\;2:1\;} 1 \longrightarrow -1 \;\circlearrowright$$

and for $g(z) = \frac{i}{2}(z + 1/z)$ we obtain:

$$
\begin{array}{c}
1 \xrightarrow{\;2:1\;} i \\
\qquad\qquad\qquad \searrow \\
\qquad\qquad\qquad\qquad 0 \longrightarrow \infty. \circlearrowright \\
\qquad\qquad\qquad \nearrow \\
-1 \xrightarrow{\;2:1\;} -i
\end{array}
$$

2.3. Definition of expansion

Let $f\colon S^2 \to S^2$ be a Thurston map and $\mathcal{C}$ be a *Jordan curve* in S^2 (i.e., a set homeomorphic to the unit circle in $\mathbb{R}^2$) with $\operatorname{post}(f) \subset \mathcal{C}$. We fix a base metric d on S^2 that induces the given topology on S^2. For $n \in \mathbb{N}$ we denote by $\operatorname{mesh}(f, n, \mathcal{C})$ the supremum of the diameters of all connected components of the set $f^{-n}(S^2 \setminus \mathcal{C})$.

DEFINITION 2.2 (Expansion). A Thurston map $f\colon S^2 \to S^2$ is called *expanding* if there exists a Jordan curve $\mathcal{C}$ in S^2 with $\operatorname{post}(f) \subset \mathcal{C}$ and

$$(2.7) \qquad\qquad \lim_{n \to \infty} \operatorname{mesh}(f, n, \mathcal{C}) = 0.$$

We will study the concept of expansion in more detail in Chapter 6 after we have built up some methods for a systematic investigation. For the moment we summarize some main facts related to this concept.

The set $f^{-n}(S^2 \setminus \mathcal{C})$ has actually only finitely many components; so the supremum in the definition of $\operatorname{mesh}(f, n, \mathcal{C})$ is a maximum (see

Proposition 5.17). We will see in Lemma 6.1 that if condition (2.7) is satisfied for one Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}$, then it actually holds for every such curve. So expansion is a property of the map f alone. Moreover, it is really a topological property, since it is independent of the choice of the base metric d on S^2 (as long as d induces the topology on S^2). Our notion of expansion for a Thurston map is equivalent to a similar concept of expansion introduced by Haïssinky-Pilgrim (see [**HP09**, Section 2.2] and Proposition 6.3).

It is immediate that expansion is preserved under topological conjugacy, i.e., if f and g are topologically conjugate Thurston maps, then f is expanding if and only if g is expanding. On the other hand, expansion is not preserved under Thurston equivalence (see the next section for the terminology). A related fact is that if two expanding Thurston maps are Thurston equivalent, then they are actually topologically conjugate (see Theorem 11.1).

Expansion is compatible with iteration of the map. Namely, if $f \colon S^2 \to S^2$ is a Thurston map, and $F = f^n$, $n \in \mathbb{N}$, is an iterate, then f is expanding if and only if F is expanding (Lemma 6.4).

A map $f \colon S^2 \to S^2$ is called *eventually onto*, if for any non-empty open set $U \subset S^2$ there is an iterate f^n such that $f^n(U) = S^2$. Every expanding Thurston map is eventually onto (Lemma 6.5). The converse does not hold: there are Thurston maps that are eventually onto, but not expanding (see Example 6.14).

If $f \colon S^2 \to S^2$ is a branched covering map, then a point $p \in S^2$ is called *periodic* if $f^n(p) = p$ for some $n \in \mathbb{N}$. The smallest n for which this is true is called the *period* of the periodic point. The point p is called *preperiodic* if there exists $k \in \mathbb{N}_0$ such that $q = f^k(p)$ is periodic. Finally, a *periodic critical point* is a periodic point $c \in \mathrm{crit}(f)$.

The following statement gives a criterion when a rational Thurston map is expanding.

PROPOSITION 2.3. *Let $R \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a rational Thurston map. Then the following conditions are equivalent:*

  (i) *R is expanding.*

 (ii) *The Julia set of R is equal to $\widehat{\mathbb{C}}$.*

(iii) *R has no periodic critical points.*

An immediate consequence of this proposition is that no postcritically-finite polynomial $P \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ (with $\deg(P) \geq 2$) can be expanding. Indeed, in this case $\infty \in \widehat{\mathbb{C}}$ is both a critical point and a fixed point of P. So condition (iii) is violated. For a related fact see Lemma 6.7.

In general, expanding Thurston maps may have periodic critical points (see Example 12.21). On the other hand, there are Thurston maps that do not have periodic critical points, yet are not Thurston equivalent to any expanding map (see Example 6.10).

Our proof of Proposition 2.3 relies on facts (in particular, on Lemma 6.5, Lemma 6.6, and Proposition A.34) that we will establish later. We will also use some basic concepts from complex dynamics, which can be found in [**CG93**] and [**Mi06a**], for example.

PROOF OF PROPOSITION 2.3. It suffices to show the chain of implications (i)$\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (i).

(i) $\Rightarrow$ (ii) Let $R\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be an expanding rational Thurston map. Then its Julia set $\mathcal{J}$ is non-empty. Let $\mathcal{F} = \widehat{\mathbb{C}}\setminus\mathcal{J}$ be the Fatou set of R. By Lemma 6.5 the map R is eventually onto. So if we assume $\mathcal{F} \neq \emptyset$, then, since $\mathcal{F}$ is open, this implies that $R^n(\mathcal{F}) = \widehat{\mathbb{C}}$ for a sufficiently high iterate R^n. Now $\mathcal{F}$ is invariant for R; so this means that $\mathcal{F} = \widehat{\mathbb{C}}$ and $\mathcal{J} = \emptyset$. This is a contradiction.

(ii) $\Rightarrow$ (iii) If the Julia set of R is equal to $\widehat{\mathbb{C}}$, then its Fatou set is empty. This implies that R cannot have periodic critical points, because a periodic critical point of a rational map is part of a super-attracting cycle and belongs to the Fatou set.

(iii) $\Rightarrow$ (i) Suppose R has no periodic critical points. Then there exists a geodesic metric ω on $\widehat{\mathbb{C}}$ (the canonical orbifold metric of f; see Sections 2.5 and A.10) such that R expands the ω-length of each path in $\widehat{\mathbb{C}}$ by a fixed factor $\rho > 1$ (Proposition A.34). This implies that R is expanding (Lemma 6.6). $\qquad\square$

2.4. Thurston equivalence

Suppose $f\colon S^2 \to S^2$ and $g\colon \widehat{S}^2 \to \widehat{S}^2$ are two Thurston maps. Here $\widehat{S}^2$ is another 2-sphere. Often $S^2 = \widehat{S}^2$, but sometimes it is important to distinguish the spheres on which the Thurston maps are defined. We call the maps f and g *topologically conjugate* if there exists a homeomorphism $h\colon S^2 \to \widehat{S}^2$ such that $h \circ f = g \circ h$. This defines a notion of equivalence for Thurston maps. Topologically conjugate maps have essentially the same dynamics under iteration up to "change of coordinates".

It is often convenient to consider a weaker notion of equivalence for Thurston maps. To define it, we first recall the definition of homotopies and isotopies between spaces. Let $I = [0, 1]$, and X, Y be topological spaces. A *homotopy between X and Y* is a continuous map $H\colon X\times I \to$

Y. We define $H_t := H(\cdot, t)\colon X \to Y$ for $t \in I$. The map H_t is called the *time-t map* of the homotopy. A homotopy $H\colon X \times I \to Y$ is called an *isotopy* if H_t is a homeomorphism of X onto Y for each $t \in I$. If $X = Y$, then H is called a homotopy (or isotopy) *on X*.

Let $H\colon X \times I \to Y$ be a homotopy and $A \subset X$. Then we say H is a homotopy *relative* to A (abbreviated "H is a homotopy rel. A") if $H_t(a) = H_0(a)$ for all $a \in A$ and $t \in I$. So this means that the image of each point in A remains fixed during the homotopy. Similarly, we speak of *isotopies rel. A*.

Two homeomorphisms $h_0, h_1\colon X \to Y$ are called *isotopic rel. A* if there exists an isotopy $H\colon X \times I \to Y$ rel. A with $H_0 = h_0$ and $H_1 = h_1$.

DEFINITION 2.4 (Thurston equivalence). Suppose $f\colon S^2 \to S^2$ and $g\colon \widehat{S}^2 \to \widehat{S}^2$ are Thurston maps. Then they are called *(Thurston) equivalent* if there exist homeomorphisms $h_0, h_1\colon S^2 \to \widehat{S}^2$ that are isotopic rel. $\mathrm{post}(f)$ and satisfy $h_0 \circ f = g \circ h_1$.

In this case, the following diagram commutes:

$$(2.8) \qquad \begin{array}{ccc} S^2 & \xrightarrow{\ h_1\ } & \widehat{S}^2 \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle g} \\ S^2 & \xrightarrow{\ h_0\ } & \widehat{S}^2. \end{array}$$

Our 2-spheres S^2 and $\widehat{S}^2$ are assumed to be oriented. Since the homeomorphisms h_0 and h_1 are isotopic, they are either both orientation-preserving or orientation-reversing. One defines Thurston equivalence sometimes differently by insisting on the homeomorphisms h_0 and h_1 in Definition 2.4 being orientation-preserving. In this case, we say that f and g are *orientation-preserving Thurston equivalent*.

This is a stronger notion of Thurston equivalence. For example, according to our definition, a rational Thurston map R on $\widehat{\mathbb{C}}$ is Thurston equivalent to the map $z \mapsto \overline{R(\overline{z})}$ (we use $h_0(z) = h_1(z) = \overline{z}$ in Definition 2.4), but these maps are in general not orientation-preserving Thurston equivalent.

If two Thurston maps are topologically conjugate, then they are Thurston equivalent. However, they will not be orientation-preserving Thurston equivalent in general. This is the main reason why we use our more general concept of Thurston equivalence.

LEMMA 2.5. *Let $f\colon S^2 \to S^2$ be a Thurston map, $g\colon \widehat{S}^2 \to \widehat{S}^2$ be a continuous map, and $h_0, h_1\colon S^2 \to \widehat{S}^2$ be homeomorphisms that are isotopic rel. $\mathrm{post}(f)$ and satisfy $h_0 \circ f = g \circ h_1$. Then g is also a*

Thurston map and we have

$$\operatorname{crit}(g) = h_1(\operatorname{crit}(f)),$$
$$\operatorname{post}(g) = h_0(\operatorname{post}(f)) = h_1(\operatorname{post}(f)).$$

Note that we again have the diagram (2.8). So under the given assumptions, f and g are Thurston equivalent.

If we already know that g is a Thurston map (which was part of the conclusion of the lemma), then, as the proof will show, the statements about $\operatorname{crit}(g)$ and $\operatorname{post}(g)$ are valid under the weaker assumptions that h_0 and h_1 are homeomorphisms that satisfy $h_0 \circ f = g \circ h_1$ and $h_0| \operatorname{post}(f) = h_1| \operatorname{post}(f)$.

PROOF. Since h_0 and h_1 are isotopic, they both preserve or both reverse orientation. So it is clear that $g = h_0 \circ f \circ h_1^{-1}$ is a branched covering map with $\operatorname{crit}(g) = h_1(\operatorname{crit}(f))$. Thus

$$g(\operatorname{crit}(g)) = (g \circ h_1)(\operatorname{crit}(f)) = (h_0 \circ f)(\operatorname{crit}(f)).$$

Since $h_0| \operatorname{post}(f) = h_1| \operatorname{post}(f)$ and $f^n(\operatorname{crit}(f)) \subset \operatorname{post}(f)$ for all $n \in \mathbb{N}$, we inductively derive

$$g^n(\operatorname{crit}(g)) = h_0(f^n(\operatorname{crit}(f))) = h_1(f^n(\operatorname{crit}(f)))$$

for all $n \in \mathbb{N}$. Hence

$$\operatorname{post}(g) = \bigcup_{n \in \mathbb{N}} g^n(\operatorname{crit}(g)) = \bigcup_{n \in \mathbb{N}} h_0(f^n(\operatorname{crit}(f)))$$
$$= h_0(\operatorname{post}(f)) = h_1(\operatorname{post}(f)).$$

In particular, this shows that $\operatorname{post}(g)$ is finite and so g is a Thurston map. $\qquad\square$

The previous lemma implies that the roles of the maps f and g in Definition 2.4 are symmetric. Indeed, suppose $H\colon S^2 \times I \to \widehat{S}^2$ is an isotopy rel. $\operatorname{post}(f)$ with $H_0 = h_0$ and $H_1 = h_1$. Define $K\colon \widehat{S}^2 \times I \to S^2$ as $K(\hat{x}, t) = (H_t)^{-1}(\hat{x})$ for $\hat{x} \in \widehat{S}^2$, $t \in I$. The map $(x, t) \mapsto (H(x, t), t)$ is a continuous bijection between the compact Hausdorff spaces $S^2 \times I$ and $\widehat{S}^2 \times I$, and hence a homeomorphism. Its inverse is given by $(\hat{x}, t) \mapsto (K(\hat{x}, t), t)$. This implies that K is continuous, and so obviously an isotopy rel. $h_0(\operatorname{post}(f)) = h_1(\operatorname{post}(f)) = \operatorname{post}(g)$. Since $k_0 := h_0^{-1} = K_0$ and $k_1 := h_1^{-1} = K_1$, the homeomorphisms k_0 and k_1 are isotopic rel. $\operatorname{post}(g)$. We also have

$$k_0 \circ g = h_0^{-1} \circ g = f \circ h_1^{-1} = f \circ k_1,$$

and so the conditions in Definition 2.4 are also satisfied if we interchange f and g.

It is also clear that if f, g, h are Thurston maps, f is (Thurston) equivalent to g, and g is equivalent to h, then f is equivalent to h. Thus, Thurston equivalence leads to a notion of equivalence for Thurston maps.

If two Thurston maps f and g are equivalent, then for each $n \in \mathbb{N}$ the iterates f^n and g^n are also equivalent. To show this statement, one needs to lift the isotopy rel. $\mathrm{post}(f)$ between the homeomorphisms h_0 and h_1 as in Definition 2.4. We postpone the proof to Chapter 11, where such lifts are considered (see Corollary 11.6).

EXAMPLE 2.6. Up to Thurston equivalence, one can often construct combinatorial models of maps that are given in some other specific way, for example by an analytic formula. To illustrate this, we consider the map $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ given by

$$(2.9) \qquad\qquad f(z) = 1 + (\omega - 1)/z^3,$$

where $\omega = e^{4\pi i/3}$. Then $\mathrm{crit}(f) = \{0, \infty\}$, and f has the ramification portrait

$$(2.10) \qquad\qquad 0 \xrightarrow{\ 3:1\ } \infty \xrightarrow{\ 3:1\ } 1 \longrightarrow \omega. \circlearrowright$$

So $\mathrm{post}(f) = \{1, \omega, \infty\}$, and f is a Thurston map.

We now construct a related Thurston map $g \colon S^2 \to S^2$ in a similar fashion as the map h in Section 1.3. For this let T be a right-angled isosceles Euclidean triangle whose hypotenuse has length 1. The angles of T are then $\pi/2, \pi/4, \pi/4$. We also consider a triangle T' that is similar to T by the scaling factor $\sqrt{2}/2$. We glue two copies of T together along their boundaries to form a pillow $\mathcal{S}^0$ (see Section A.10), which is a topological 2-sphere. These two triangles are called the 0-tiles. As before, we color one of them white, and the other black.

We divide each of the two 0-tiles by the perpendicular bisector of the hypotenuse into two triangles similar to T and isometric to T'. We slit the pillow open along one such bisector and glue two copies of T' into the slit as indicated on the left of Figure 2.1. This results in a polyhedral surface $\mathcal{S}^1$ consisting of six triangles, each isometric to T'. These six small triangles are called the 1-tiles. They are colored in a checkerboard fashion black and white so that triangles sharing an edge have different colors, as indicated in the picture. Note that there are two points (labeled "$0 \mapsto \infty$" and "$\infty \mapsto 1$") in which all 1-tiles intersect.

Each small triangle in $\mathcal{S}^1$ is now mapped by a similarity to the triangle in $\mathcal{S}^0$ of the same color. This determines a unique map $\mathcal{S}^1 \to$

$\mathcal{S}^0$ if the vertices of the 1-tiles are mapped as indicated on the top of Figure 2.1.

We identify $\mathcal{S}^1$ with $\mathcal{S}^0 = S^2$ such that the four triangles shown on the top left in Figure 2.1 are identified with the white triangle in $\mathcal{S}^0$, and the other two small triangles in $\mathcal{S}^1$ are identified with the black triangle in $\mathcal{S}^0$. Here we require that this identification respects the natural identification of the three common corners of the two copies of T (labeled $\omega, 1, \infty$ in the picture) which are contained in both $\mathcal{S}^0$ and $\mathcal{S}^1$. This yields a Thurston map $g\colon S^2 \to S^2$, indicated at the bottom of Figure 2.1. Here we have cut the pillow $\mathcal{S}^0$ along the two legs of the two triangles. The two pairs of legs marked with the same symbol have to be identified, i.e., glued together, to form the pillow.

Note that g depends on how precisely $\mathcal{S}^1$ is identified with $\mathcal{S}^0$. A different identification yields another Thurston map $\widetilde{g}$, but it is easy to see that $\widetilde{g}$ is always Thurston equivalent to g. Thus the notion of Thurston equivalence allows us some latitude in specifying the precise identification of $\mathcal{S}^1$ with $\mathcal{S}^0 = S^2$.

Our main point here is the following statement: *The map f as defined in* (2.9) *and the map g constructed above are Thurston equivalent.* As our purpose is to give the reader some intuition for the concept of Thurston equivalence, we only outline a proof omitting some details.

Note that $f(z) = \tau(z^3)$, where $\tau(\zeta) = 1 + (\omega - 1)/\zeta$ is a Möbius transformation that maps the upper half-plane to the half-plane above the line through the points ω and 1 (indeed, τ maps $0, 1, \infty$ to $\infty, \omega, 1$, respectively).

Let $\mathcal{C} \subset \widehat{\mathbb{C}}$ be the circle through $\omega, 1, \infty$ (i.e., the extended line through 1 and ω). Then $\mathcal{C}$ contains all postcritical points of f. The closures of the two components of $\widehat{\mathbb{C}} \setminus \mathcal{C}$ are called the 0-tiles. The 0-tile containing $0 \in \widehat{\mathbb{C}}$ (i.e., the half-plane above the line through 1 and ω) is colored white, the other 0-tile is colored black.

Since $f(z) = \tau(z^3)$, we have $f^{-1}(\mathcal{C}) = \bigcup_{k \in \{1,\dots,6\}} R_k$, where

$$R_k = \{re^{ik\pi/3} : 0 \leq r \leq \infty\}$$

for $k \in \mathbb{Z}$ is the ray from 0 through the sixth root of unity $e^{ik\pi/3}$ (note that $R_0 = R_6$). These rays divide $\widehat{\mathbb{C}}$ into open sectors that are the complementary components of $\widehat{\mathbb{C}} \setminus f^{-1}(\mathcal{C})$. For $k = 1, \dots, 6$ let

$$X_k := \{re^{it} : 0 \leq r \leq \infty,\ (k-1)\pi/3 \leq t \leq k\pi/3\}.$$

be the closure of the sector bounded by R_{k-1} and R_k. We call these sets the 1-tiles. Note that f maps each 1-tile X_k homeomorphically to a 0-tile $X^0 \subset \widehat{\mathbb{C}}$. We color X_k white if X^0 is white, and black otherwise.
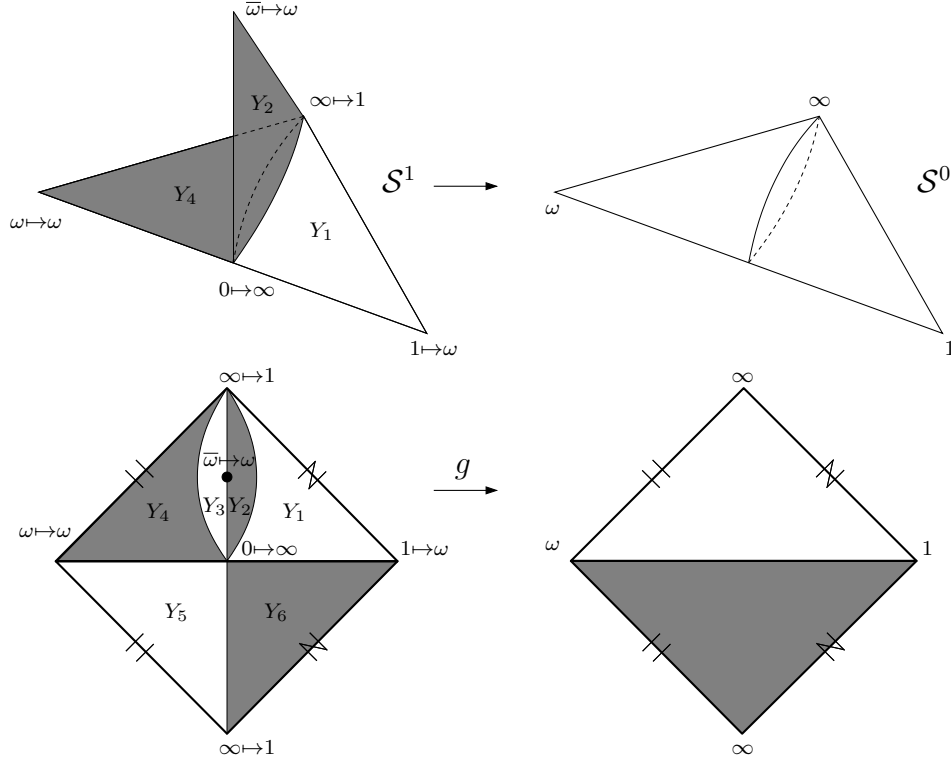
FIGURE 2.1. The map g.

Then the 1-tile X_k, $k = 1, \ldots, 6$, is colored black or white depending on whether k is even or odd.

In order to show that f and g are Thurston equivalent, we need two homeomorphisms h_0 and h_1 as in Definition 2.4. The homeomorphism $h_0 \colon \widehat{\mathbb{C}} \to S^2 = \mathcal{S}^0$ is defined as follows. We map the white and black 0-tiles in $\widehat{\mathbb{C}}$ homeomorphically to the white and black 0-tile in $\mathcal{S}^0$, respectively. We can do this so that the points $1, \omega, \infty \in \widehat{\mathbb{C}}$ are mapped to the points labeled $1, \omega, \infty \in S^2$ on the top right of Figure 2.1, and so that the homeomorphisms on these two 0-tiles match along the common boundary.

The homeomorphism $h_1 \colon \widehat{\mathbb{C}} \to \mathcal{S}^1$ is constructed similarly by mapping 1-tiles in $\widehat{\mathbb{C}}$ homeomorphically to corresponding 1-tiles in $\mathcal{S}^1$. We will then view h_1 as a map to the domain $S^2 = \mathcal{S}^0$ of g by using the given identification of $\mathcal{S}^1$ with $\mathcal{S}^0$

For the precise definition of h_1, we denote the 1-tiles in $\mathcal{S}^1$ by $Y_1, \ldots, Y_6$ so that they follow in positive cyclic order around the point labeled "$0 \mapsto \infty$" on the left in Figure 2.1 and such that Y_1 is the white 1-tile containing the point labeled "$1 \mapsto \omega$". With our labeling the 1-tile X_k in $\widehat{\mathbb{C}}$ has the same color as the 1-tile Y_k in $\mathcal{S}^1$. So for

fixed $k = 1, \ldots, 6$, the map $f|X_k \colon X_k \to X^0$ is a homeomorphism of X_k onto a 0-tile X^0 in $\widehat{\mathbb{C}}$, and $g|Y_k \colon Y_k \to Y^0$ is a homeomorphism of Y_k onto a 0-tile Y^0 in $\mathcal{S}^0$, where the tiles X_k, Y_k, X^0, Y^0 all have the same color. In particular, h_0 maps X^0 homeomorphically onto Y^0, and we can define
$$h_1|X_k := (g|Y_k)^{-1} \circ h_0 \circ (f|X_k).$$
This is a homeomorphism from X_k onto Y_k. It is then straightforward to check that these partial homeomorphisms on the 1-tiles X_k of $\widehat{\mathbb{C}}$ paste together to a well-defined homeomorphism $h_1 \colon \widehat{\mathbb{C}} \to \mathcal{S}^1 \cong S^2$. Moreover, it follows directly from the definition of h_1 that $h_0 \circ f = g \circ h_1$, and so we get a commutative diagram as in (2.8).

It remains to argue that the homeomorphisms h_0 and h_1 are isotopic rel. $\mathrm{post}(f)$. Note that h_0 and h_1 are orientation-preserving and agree on the set $\mathrm{post}(f) = \{1, \omega, \infty\}$. Moreover, h_0 maps $\mathcal{C}$ (the extended line through ω and 1) and h_1 maps $R_0 \cup R_4$ to the equator of the pillow. So basically we need to deform $\mathcal{C}$ to $R_0 \cup R_4$ while keeping $\mathrm{post}(f) = \{\omega, 1, \infty\} = \mathcal{C} \cap (R_0 \cup R_4)$ fixed to obtain the desired isotopy between h_0 and h_1. It is intuitively clear that this is possible. Since $\mathrm{post}(f) = 3$ the existence of the desired isotopy actually follows from a general fact (see Lemma 11.11).

2.5. The orbifold associated with a Thurston map

An orbifold is a space that is locally represented as a quotient of a model space by a group action (see [**Th**, Chapter 13]). In our present context we are only interested in the 2-dimensional case where the group actions are given by cyclic groups near points on a surface. Then all the relevant information is given by a pair (S, α), where S is a surface and $\alpha \colon S \to \widehat{\mathbb{N}} := \mathbb{N} \cup \{\infty\}$ is a map such that the set of points $p \in S$ with $\alpha(p) \neq 1$ is a closed and discrete set in S. We call such a function α a *ramification function* on S, and the pair (S, α) an *orbifold*. Each number $\alpha(p)$ should be thought of as the order of an associated cyclic group (see Section A.9 and in particular Proposition A.29 (ii)). The set $\mathrm{supp}(\alpha) := \{p \in S : \alpha(p) \geq 2\}$ is the *support* of α. It is closed and discrete in S; so if S is compact, then $\mathrm{supp}(\alpha)$ is a finite set.

Orbifolds are useful for the study of branched covering maps, in particular Thurston maps. In this case, the underlying surface of the orbifold is a topological 2-sphere. Every Thurston map $f \colon S^2 \to S^2$ has an associated orbifold $\mathcal{O}_f = (S^2, \alpha_f)$, where the orbifold data are determined by the *ramification function* $\alpha_f \colon S^2 \to \widehat{\mathbb{N}}$ of f. We will first define α_f and discuss the main properties of this function before we turn our attention to $\mathcal{O}_f$.

For a general Thurston map f we consider its orbifold $\mathcal{O}_f$ as a purely topological object that encodes some ramification data of f. In particular, the orbifold $\mathcal{O}_f$ does not carry any canonical geometric structure associated with f. This is different for *rational* Thurston maps f defined on the Riemann sphere $\widehat{\mathbb{C}}$. Here the additional conformal structure on $\widehat{\mathbb{C}}$ can be used to define the *canonical orbifold metric* of f. We will discuss this briefly at the end of this section, and in more detail in Section A.10.

We use the notation $\widehat{\mathbb{N}} := \mathbb{N} \cup \{\infty\}$, and extend the usual order relations $<, \leq, >, \geq$ on $\mathbb{N}$ in the obvious way to $\widehat{\mathbb{N}}$. So $a < \infty$ for $a \in \mathbb{N}$, $a \leq \infty$ for $a \in \widehat{\mathbb{N}}$, etc. We also extend multiplication of natural numbers to $\widehat{\mathbb{N}}$ by setting $a \cdot \infty = \infty \cdot a = \infty$ for $a \in \widehat{\mathbb{N}}$. For $a, b \in \widehat{\mathbb{N}}$ we say that a *divides* b, written $a|b$, if there exists $k \in \widehat{\mathbb{N}}$ such that $b = ak$. So the relation $a|b$ is an extension of the usual divisor relation in $\mathbb{N}$ under the additional convention that every value in $\mathbb{N} \cup \{\infty\}$ divides ∞.

Suppose $A \subset \widehat{\mathbb{N}}$ is arbitrary. Then there exists a unique $L \in \widehat{\mathbb{N}}$, called the *least common multiple* of the elements of A and denoted by $\mathrm{lcm}(A)$, with the following properties: $a|L$ for all $a \in A$, and if $L' \in \widehat{\mathbb{N}}$ is such that $a|L'$ for all $a \in A$, then $L|L'$. It is easy to see that if $A \subset \mathbb{N}$ is a finite set of natural numbers, then $\mathrm{lcm}(A) \in \mathbb{N}$ is the least common multiple of the numbers in A in the usual sense, and $\mathrm{lcm}(A) = \infty$ otherwise.

If $\alpha, \beta \colon S^2 \to \widehat{\mathbb{N}}$ are functions on a 2-sphere S^2, then we write $\alpha \leq \beta$ and $\alpha|\beta$ if $\alpha(p) \leq \beta(p)$ and $\alpha(p)|\beta(p)$ for all $p \in S^2$, respectively.

DEFINITION 2.7 (Ramification function of a Thurston map). Let $f \colon S^2 \to S^2$ be a Thurston map. Then its *ramification function* is the map $\alpha_f \colon S^2 \to \widehat{\mathbb{N}}$ defined as

$$\alpha_f(p) = \mathrm{lcm}\{\deg(f^n, q) : q \in S^2,\ n \in \mathbb{N},\ \text{and}\ f^n(q) = p\}$$

for $p \in S^2$.

We will see momentarily (see the remark after Proposition 2.9) that α_f is indeed a ramification function on the surface S^2. It admits the following characterization.

PROPOSITION 2.8. *Let $f \colon S^2 \to S^2$ be a Thurston map with its associated ramification function $\alpha_f \colon S^2 \to \widehat{\mathbb{N}}$. Then we have:*

(i) $\deg(f, q)\alpha_f(q) \,|\, \alpha_f(p)$ *whenever $p, q \in S^2$ and $f(q) = p$,*

(ii) *if $\beta \colon S^2 \to \widehat{\mathbb{N}}$ is any function such that $\deg(f, q)\,\beta(q)\,|\,\beta(p)$ whenever $p, q \in S^2$ and $f(q) = p$, then $\alpha_f|\beta$.*

Moreover, α_f is the unique function with the properties (i) *and* (ii).

Before we turn to the proof of this proposition, we point out a fact that follows from repeated application of (i). Namely, suppose that $p, q \in S^2$, $n \in \mathbb{N}$, and $f^n(q) = p$. Let $q_k := f^k(q)$ for $k = 0, \ldots, n$. Then
$$\deg(f^n, q) = \deg(f, q_0) \deg(f, q_1) \cdots \deg(f, q_{n-1})$$
as follows from (2.4). Moreover,
$$\alpha_f(q_k) \deg(f, q_k) \,|\, \alpha_f(q_{k+1})$$
for $k = 0, \ldots, n-1$. Since $q_0 = q$ and $q_n = p$, this implies that $\deg(f^n, q)\alpha_f(q) \,|\, \alpha_f(p)$.

In the following, we will often use relation (2.4) without mentioning it explicitly.

PROOF OF PROPOSITION 2.8. Let $f\colon S^2 \to S^2$ be a Thurston map and α_f its ramification function as given in Definition 2.7.

To establish (i) for α_f, suppose $p, q \in S^2$ and $f(q) = p$. If $\alpha_f(p) = \infty$, then $\deg(f, q)\alpha_f(q) \,|\, \alpha_f(p)$, and there is nothing to prove. So we may assume that $\alpha_f(p) \in \mathbb{N}$.

Suppose $q' \in S^2$ is a point with $f^n(q') = q$ for some $n \in \mathbb{N}$. Then $f^{n+1}(q') = p$, and
$$\deg(f^{n+1}, q') = \deg(f^n, q') \deg(f, q);$$
so $\deg(f^n, q') \deg(f, q) \,|\, \alpha_f(p)$ by definition of $\alpha_f(p)$. Then the quotient $\alpha_f(p)/\deg(f, q)$ must be a natural number that has $\deg(f^n, q')$ as a divisor. This implies that the least common multiple $\alpha_f(q)$ of all such numbers $\deg(f^n, q')$ divides $\alpha_f(p)/\deg(f, q)$. We conclude that
$$\deg(f, q)\alpha_f(q) \,|\, \alpha_f(p),$$
and (i) follows.

Let $\beta\colon S^2 \to \widehat{\mathbb{N}}$ be another function with the property (i). Then by repeated use of this property we see that $\deg(f^n, q)\beta_f(q) \,|\, \beta(p)$ whenever $p, q \in S^2$, $n \in \mathbb{N}$, and $f^n(q) = p$. In particular, $\deg(f^n, q)|\beta(p)$. As this is true for all $q \in S^2$ with $f^n(q) = p$ for some $n \in \mathbb{N}$, we conclude $\alpha_f(p)|\beta(p)$ for all $p \in S^2$. This establishes the desired property (ii) of α_f.

The uniqueness of a function with properties (i) and (ii) is clear. □

To state other important properties of the ramification function, we first need a definition. A *critical cycle* C of a Thurston map $f\colon S^2 \to S^2$ is an orbit of a periodic critical point of f; so then there exists $c \in \mathrm{crit}(f)$, and $n \in \mathbb{N}$ with $f^n(c) = c$ such that $C = \{f^k(c) : k = 0, \ldots, n-1\}$.

PROPOSITION 2.9. *Let $f \colon S^2 \to S^2$ be a Thurston map with its associated ramification function $\alpha_f \colon S^2 \to \widehat{\mathbb{N}}$. Then for $p \in S^2$ we have:*

(i) *$\alpha_f(p) \geq 2$ if and only if $p \in \mathrm{post}(f)$,*

(ii) *$\alpha_f(p) = \infty$ if and only if p is contained in a critical cycle of f.*

In particular, this implies that $\mathrm{supp}(\alpha_f) = \{p \in S^2 : \alpha_f(p) \geq 2\} = \mathrm{post}(f)$ is a finite set and so α_f is really a ramification function on the surface S^2 (as defined in the beginning of this section).

PROOF. (i) If $p \in S^2 \setminus \mathrm{post}(f)$, then $\deg(f^n, q) = 1$ whenever $q \in S^2$, $n \in \mathbb{N}$, and $f^n(q) = p$, because none of the points $f^k(q)$, $k = 0, \ldots, n-1$, can be a critical point of f. Hence $\alpha_f(p) = 1$ by definition of α_f.

If $p \in \mathrm{post}(f)$, then there exist $c \in \mathrm{crit}(f)$ and $n \in \mathbb{N}$ such that $f^n(c) = p$. Then $\deg(f, c) \geq 2$ which implies that $\deg(f^n, c) \geq 2$. By definition of α_f we have $\deg(f^n, c) | \alpha_f(p)$; so $\alpha_f(p) \geq 2$.

(ii) If p is contained in a critical cycle, then there exists a periodic critical point c of f such that $f^n(c) = p$ for some $n \in \mathbb{N}_0$. There exists $k \in \mathbb{N}$ such that $f^k(c) = c$, and so $f^{n+km}(c) = p$ for all $m \in \mathbb{N}$. Then the numbers $\deg(f^{n+km}, c) \geq \deg(f, c)^m \geq 2^m$ divide $\alpha_f(p)$ for all $m \in \mathbb{N}$. This is only possible if $\alpha_f(p) = \infty$.

Conversely, suppose $p \in S^2$ and $\alpha_f(p) = \infty$. Then by definition of α_f the set $\{\deg(f^n, q) : q \in S^2, n \in \mathbb{N}, f^n(q) = p\}$ is unbounded. In particular, there exist $q \in \mathbb{N}$ and $n \in \mathbb{N}$ with $f^n(q) = p$ such that $\deg(f^n, q) > M$, where

$$ M := \prod_{c \in \mathrm{crit}(f)} \deg(f, c) \in \mathbb{N}. $$

Let $q_k = f^k(q)$ for $k = 0, \ldots, n-1$. Then

$$ M < \deg(f^n, q) = \prod_{k=0}^{n-1} \deg(f, q_k). $$

Since $\deg(f, q_k) > 1$ only if q_k is a critical point and since $\deg(f^n, q) > M$, there exists a critical point c of f that appears at least twice in the list $q_0, \ldots, q_{n-1}$. Then c is periodic and p belongs to the orbit of c. Hence p is an element of a critical cycle of f. $\square$

We can now define the orbifold of a Thurston map from the point of view explained in the beginning of this section.

DEFINITION 2.10 (Orbifold of a Thurston map). Let $f\colon S^2 \to S^2$ be a Thurston map. The *orbifold* associated with f is the pair $\mathcal{O}_f \coloneqq (S^2, \alpha_f)$,

where $\alpha_f\colon S^2 \to \widehat{\mathbb{N}}$ is the ramification function of f.

If $\mathcal{O} = (S^2, \alpha)$ is an orbifold, then the *Euler characteristic* of $\mathcal{O}$ is defined as

$$(2.11) \qquad \chi(\mathcal{O}) = 2 - \sum_{p \in S^2} \left(1 - \frac{1}{\alpha(p)}\right).$$

Here and elsewhere we use the convention that $a/\infty = 0$ for $a \in \mathbb{N}$. The sum in (2.11) is really a finite sum, where only the points in $\operatorname{supp}(\alpha)$ give a non-zero contribution, but it is convenient to write it (and similar sums below) in this form. A geometric interpretation of $\chi(\mathcal{O})$ is given in Section A.9.

We call $\mathcal{O}$ *parabolic* if $\chi(\mathcal{O}) = 0$, and *hyperbolic* if $\chi(\mathcal{O}) < 0$. The orbifold $\mathcal{O}_f$ of a Thurston map f is always parabolic or hyperbolic. In order to show this, we first need a lemma.

LEMMA 2.11. *Let $\alpha, \alpha'\colon S^2 \to \widehat{\mathbb{N}}$ be ramification functions on a 2-sphere S^2, and let $f\colon S^2 \to S^2$ be a branched covering map satisfying $\deg_f(p) \cdot \alpha'(p) = \alpha(f(p))$ for all $p \in S^2$. Then for the orbifolds $\mathcal{O}' = (S^2, \alpha')$ and $\mathcal{O} = (S^2, \alpha)$ we have $\chi(\mathcal{O}') = \deg(f) \cdot \chi(\mathcal{O})$.*

This statement is an orbifold version of the well-known identity $\chi(X) = \deg(f) \cdot \chi(Y)$ for covering maps $f\colon X \to Y$ between surfaces (or more general spaces) X and Y. For compact surfaces this is a special case of the Riemann-Hurwitz formula (2.3).

PROOF. Let $d \coloneqq \deg(f)$. Then for each point $q \in S^2$ we have

$$\sum_{p \in f^{-1}(q)} \deg_f(p) = d.$$

So the Riemann-Hurwitz formula (2.3) implies that

$$2 - \chi(\mathcal{O}') = \sum_{p \in S^2} \left(1 - \frac{1}{\alpha'(p)} \right) = \sum_{p \in S^2} \left(1 - \frac{\deg_f(p)}{\alpha(f(p))} \right)$$

$$= \sum_{p \in S^2} (1 - \deg_f(p)) + \sum_{p \in S^2} \left(\deg_f(p) - \frac{\deg_f(p)}{\alpha(f(p))} \right)$$

$$= 2 - 2d + \sum_{q \in S^2} \sum_{p \in f^{-1}(q)} \left(\deg_f(p) - \frac{\deg_f(p)}{\alpha(f(p))} \right)$$

$$= 2 - 2d + d \sum_{q \in S^2} \left(1 - \frac{1}{\alpha(q)} \right)$$

$$= 2 - 2d + d(2 - \chi(\mathcal{O})) = 2 - d\chi(\mathcal{O}).$$

Note that each sum in this computation, where the summation range is over all points in S^2, actually has only finitely many non-zero terms. The claim follows.   □

PROPOSITION 2.12. *Let* $f \colon S^2 \to S^2$ *be a Thurston map, and* α_f *be its associated ramification function. Then*

$$(2.12) \qquad \chi(\mathcal{O}_f) = 2 - \sum_{p \in S^2} \left(1 - \frac{1}{\alpha_f(p)} \right) \le 0.$$

Here we have equality if and only if $\deg_f(p) \cdot \alpha_f(p) = \alpha_f(f(p))$ *for all* $p \in S^2$.

PROOF. For $p \in S^2$ we define

$$\alpha'(p) = \begin{cases} \alpha_f(f(p))/\deg_f(p) & \text{if } \alpha_f(f(p)) < \infty, \\ \infty & \text{if } \alpha_f(f(p)) = \infty. \end{cases}$$

By Proposition 2.8 the function α' takes values in $\widehat{\mathbb{N}}$ and we have $\alpha' \ge \alpha_f$. Moreover, $\{p \in S^2 : \alpha'(p) \ge 2\} \subset f^{-1}(\mathrm{post}(f))$ is a finite set, and so α' is a ramification function on S^2. It satisfies $\deg_f(p) \cdot \alpha'(p) = \alpha_f(f(p))$ for all $p \in S^2$. Hence by Lemma 2.11 we have $\chi(\mathcal{O}') = d\chi(\mathcal{O}_f)$, where $d = \deg(f)$ and $\mathcal{O}' = (S^2, \alpha')$.

On the other hand, the fact that $\alpha' \ge \alpha_f$ and the definition of the Euler characteristic of an orbifold imply that $\chi(\mathcal{O}') \le \chi(\mathcal{O}_f)$. Hence

$$(d - 1)\chi(\mathcal{O}_f) = \chi(\mathcal{O}') - \chi(\mathcal{O}_f) \le 0.$$

We conclude that $\chi(\mathcal{O}_f) \le 0$. Here we have equality if and only if $\chi(\mathcal{O}') = \chi(\mathcal{O}_f)$ which in view of $\alpha' \ge \alpha_f$ is in turn equivalent to $\alpha' = \alpha_f$. This last condition is the same as the requirement that $\deg_f(p) \cdot \alpha_f(p) = \alpha_f(f(p))$ for all $p \in S^2$. The statement follows.   □

By the previous proposition the orbifold of a Thurston map $f\colon S^2 \to S^2$ is parabolic or hyperbolic.

Let $\mathcal{O} = (S^2, \alpha)$ be an orbifold. If we label the finitely many points $p_1, \ldots, p_k$ in $\mathrm{supp}(\alpha)$ so that $2 \le \alpha(p_1) \le \cdots \le \alpha(p_k)$, then the k-tuple

$$(\alpha(p_1), \ldots, \alpha(p_k))$$

is called the *signature* of $\mathcal{O}$. The *signature of a Thurston map* $f\colon S^2 \to S^2$ is the signature of its orbifold $\mathcal{O}_f = (S^2, \alpha_f)$. Note that in this case the support of the ramification function α_f consists precisely of the points in $\mathrm{post}(f)$ (see Proposition 2.9), and so the signature of f is determined by the restriction of α_f to $\mathrm{post}(f)$.

The Lattès map g from Section 1.1, for example, has signature $(2, 2, 2, 2)$, and accordingly its associated orbifold $\mathcal{O}_g$ is parabolic.

PROPOSITION 2.13 (Thurston maps with parabolic orbifold). *Let $f\colon S^2 \to S^2$ be a Thurston map. Then the following conditions are equivalent:*

(i) *$\mathcal{O}_f$ is parabolic,*

(ii) *the signature of $\mathcal{O}_f$ is*

$$(\infty, \infty),\ (2, 2, \infty),\ (2, 2, 2, 2),\ (2, 4, 4),\ (3, 3, 3),\ or\ (2, 3, 6),$$

(iii) *the ramification function $\alpha_f\colon S^2 \to \widehat{\mathbb{N}}$ has the following property: for all $p \in S^2$ and $q \in f^{-1}(p)$, we have*

$$\alpha_f(p) = \deg_f(q)\alpha_f(q).$$

Rational Thurston maps with parabolic orbifolds are investigated in Chapter 3 and Section 7.2. Another characterization of Thurston maps with parabolic orbifolds is given in Lemma 19.12.

PROOF. (i) $\Leftrightarrow$ (ii) If $\mathcal{O}_f$ has one of the signatures listed in (ii), then $\chi(\mathcal{O}_f) = 0$, and so $\mathcal{O}_f$ is parabolic. Conversely, if $\chi(\mathcal{O}_f) = 0$ then one first notes that f can have at most four postcritical points. Exhausting all combinatorial possibilities, we are led to the signatures in (ii).

(i) $\Leftrightarrow$ (iii) This immediately follows from the second part of Proposition 2.12. $\square$

It is an elementary fact that the signatures of equivalent Thurston maps are the same.

PROPOSITION 2.14. *Two Thurston maps that are Thurston equivalent have the same signatures.*

Proof. Let $f: S^2 \to S^2$ and $g: \widehat{S}^2 \to \widehat{S}^2$ be two Thurston maps on 2-spheres S^2 and $\widehat{S}^2$, and suppose they are Thurston equivalent. Then there exist homeomorphisms $h_0, h_1: S^2 \to \widehat{S}^2$ as in Definition 2.4.

Let $\alpha_f: S^2 \to \widehat{\mathbb{N}}$ and $\alpha_g: \widehat{S}^2 \to \widehat{\mathbb{N}}$ be the ramification functions of f and g, respectively. The claim will follow if we can show that $\alpha_f = \alpha_g \circ h_0$.

To establish this identity, define $\nu := \alpha_g \circ h_0$. Since α_g is supported on $\mathrm{post}(g)$, $h_0|\, \mathrm{post}(f) = h_1|\, \mathrm{post}(f)$, and $h_0(\mathrm{post}(f)) = h_1(\mathrm{post}(f)) = \mathrm{post}(g)$ (see Lemma 2.5), we have $\nu = \alpha_g \circ h_0 = \alpha_g \circ h_1$.

Now let $p \in S^2$ be arbitrary, and define $\widehat{p} := h_1(p)$. By what we have seen, $\nu(p) = \alpha_g(h_1(p)) = \alpha_g(\widehat{p})$; moreover, the relation $h_0 \circ f = g \circ h_1$ implies that $\deg(f, p) = \deg(g, \widehat{p})$ and $h_0(f(p)) = g(\widehat{p})$. Hence

$$\nu(p) \cdot \deg(f, p) = \alpha_g(\widehat{p}) \cdot \deg(g, \widehat{p})$$

divides

$$\alpha_g(g(\widehat{p})) = \alpha_g(h_0(f(p))) = \nu(f(p)).$$

Proposition 2.8 (ii) implies that α_f divides $\nu = \alpha_g \circ h_0$. If we reverse the roles of f and g, then a similar argument shows that $\alpha_g | \alpha_f \circ h_0^{-1}$, or equivalently, $\alpha_g \circ h_0 | \alpha_f$. So $\alpha_f = \alpha_g \circ h_0$ as desired. $\square$

The ramification function and hence the signature of a Thurston map do not change if we pass to any of its iterates.

Proposition 2.15. *Let $f: S^2 \to S^2$ be a Thurston map. Then $\alpha_f = \alpha_{f^n}$ for each $n \in \mathbb{N}$.*

Proof. Fix $n \in \mathbb{N}$, and let $F = f^n$. If $p \in S^2$, $k \in \mathbb{N}$, and $q \in F^{-k}(p)$, then $p = F^k(q) = f^{nk}(q)$, and so

$$\deg(F^k, q) = \deg(f^{nk}, q) | \alpha_f(p)$$

by definition of α_f (see Definition 2.7). Since this is true for all $k \in \mathbb{N}$ and $q \in F^{-k}(p)$, this in turn implies $\alpha_F(p) | \alpha_f(p)$ by definition of α_F; so $\alpha_F | \alpha_f$.

On the other hand, suppose $p \in S^2$, and let $k \in \mathbb{N}$ and $q \in f^{-k}(p)$ be arbitrary. Then there exist $l, m \in \mathbb{N}$ such that $F^m = f^k \circ f^l$. We can find a point $q' \in S^2$ such that $f^l(q') = q$. Then

$$\deg(F^m, q') = \deg(f^k, q) \cdot \deg(f^l, q'),$$

and so $\deg(f^k, q) | \deg(F^m, q')$. Since $F^m(q') = f^k(q) = p$, we have $\deg(F^m, q') | \alpha_F(p)$ by definition of α_F, which implies $\deg(f^k, q) | \alpha_F(p)$. Since $k \in \mathbb{N}$ and $q \in f^{-k}(p)$ were arbitrary, we have $\alpha_f(p) | \alpha_F(p)$ by definition of α_f. We conclude that $\alpha_f | \alpha_F$; but we have seen above that $\alpha_F | \alpha_f$, and so $\alpha_f = \alpha_F$ as desired. $\square$

We finish this section with a brief discussion of the *canonical orbi-fold metric* associated with an orbifold $\mathcal{O} = (\widehat{\mathbb{C}}, \alpha)$ whose underlying surface is the Riemann sphere $\widehat{\mathbb{C}}$. Here we assume that $\mathcal{O}$ is parabolic or hyperbolic, the only cases relevant for Thurston maps. With the ramification function α understood, we use the notation

$$\widehat{\mathbb{C}}_0 := \widehat{\mathbb{C}} \setminus \{z \in \widehat{\mathbb{C}} : \alpha(z) = \infty\}.$$

So $\widehat{\mathbb{C}}_0$ is the Riemann sphere with each *puncture* of the orbifold $\mathcal{O} = (\widehat{\mathbb{C}}, \alpha)$ (i.e., a point $p \in \widehat{\mathbb{C}}$ with $\alpha(p) = \infty$) removed.

We set $X = \mathbb{C}$ or $X = \mathbb{D}$ depending on whether $\mathcal{O}$ is parabolic or hyperbolic. Then there exists a holomorphic branched covering map $\Theta \colon X \to \widehat{\mathbb{C}}_0$ such that

$$\deg(\Theta, z) = \alpha(\Theta(z))$$

for each $z \in X$. The map Θ is unique up to a precomposition by a holomorphic biholomorphism of X and called the *universal covering map* of the orbifold $\mathcal{O}$. These facts are discussed in detail in Section A.9.

We equip X with its natural metric d_0, namely the Euclidean metric if $X = \mathbb{C}$ and the hyperbolic metric if $X = \mathbb{D}$. Then one can show that there exists a metric ω on $\widehat{\mathbb{C}}$, called the *canonical orbifold metric* of $\mathcal{O}$, which is given by

$$\omega(p, q) = \inf\{d_0(z, w) : z \in \Theta^{-1}(p),\ w \in \Theta^{-1}(q)\}$$

for $p, q \in \widehat{\mathbb{C}}_0$. Since the universal covering map Θ of $\mathcal{O}$ is essentially unique, the metric ω is uniquely determined if $\mathcal{O}$ is hyperbolic and uniquely determined up to a scaling factor if $\mathcal{O}$ is parabolic. Note that this conclusion strongly relies on holomorphicity of the maps involved; namely, it follows from the fact that a biholomorphism of $\mathbb{D}$ preserves the hyperbolic metric and biholomorphism of $\mathbb{C}$ is a Euclidean similarity and so scales Euclidean distances by a fixed factor.

If we equip X with the metric d_0 and $\widehat{\mathbb{C}}_0$ with the metric ω, then Θ is a path isometry in the sense that

$$\text{length}_\omega(\Theta \circ \beta) = \text{length}_{d_0}(\beta)$$

for all paths β in X. This property characterizes the metric ω; so roughly speaking, one can say that the canonical orbifold metric ω of $\mathcal{O}$ is obtained by pushing forward the natural metric d_0 on X by the universal covering map Θ to $\widehat{\mathbb{C}}_0$.

If $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is a rational Thurston map, then its orbifold $\mathcal{O}_f = (\widehat{\mathbb{C}}, \alpha_f)$ is parabolic or hyperbolic and so the preceding discussion applies. We call the metric $\omega = \omega_f$ for the orbifold $\mathcal{O}_f$, the *canonical orbifold metric* of f. Note that in the parabolic case, it is only unique

up to scaling, but often this ambiguity does not matter. Since the essential uniqueness of ω strongly relies on the holomorphicity assumption, one cannot define a similar canonical metric for a general Thurston map $f\colon S^2 \to S^2$ defined on a topological 2-sphere S^2 with no holomorphic structure (note though that sometimes it is useful to identify S^2 with $\widehat{\mathbb{C}}$ and pick a suitable orbifold metric on $\widehat{\mathbb{C}}$; see the proof of Proposition 6.11, for example).

For a detailed discussion of the universal orbifold metric (and also the definition of a natural associated measure, the *canonical orbifold measure*), see Section A.10.

2.6. Thurston's characterization of rational maps

Thurston's criterion when a Thurston map is equivalent to a rational map is an important theorem in complex dynamics. To formulate this statement, we need some definitions.

DEFINITION 2.16 (Invariant multicurves). Let $f\colon S^2 \to S^2$ be a Thurston map.

(i) A Jordan curve $\gamma \subset S^2 \setminus \mathrm{post}(f)$ is called *non-peripheral* if each of the two components of $S^2 \setminus \gamma$ contains at least two points from $\mathrm{post}(f)$, and is called *peripheral* otherwise.

(ii) A *multicurve* is a non-empty finite set of non-peripheral Jordan curves in $S^2 \setminus \mathrm{post}(f)$ that are pairwise disjoint and pairwise non-homotopic rel. $\mathrm{post}(f)$.

(iii) A multicurve Γ is called f-*invariant* (or simply *invariant* if f is understood) if each non-peripheral component of the preimage $f^{-1}(\gamma)$ of a curve $\gamma \in \Gamma$ is homotopic rel. $\mathrm{post}(f)$ to a curve $\gamma' \in \Gamma$.

Note that if $\gamma \subset S^2$ is a Jordan curve, then by the Schönflies theorem $S^2 \setminus \gamma$ has precisely two components, each of which is a topological disk.

If $\#\,\mathrm{post}(f) = 3$ every Jordan curve in $S^2 \setminus \mathrm{post}(f)$ is peripheral; so there are no multicurves in this case. If $\#\,\mathrm{post}(f) = 4$ then every multicurve consists of a single Jordan curve.

In (iii) we call two Jordan curves $\gamma, \gamma' \subset S^2 \setminus \mathrm{post}(f)$ *homotopic rel.* $\mathrm{post}(f)$ if there exists a homotopy $H\colon S^2 \times [0,1] \to S^2$ rel. $\mathrm{post}(f)$ such that $H_0 = \mathrm{id}_{S^2}$ and $H_1(\gamma) = \gamma'$. Here we also implicitly used that if $\gamma \subset S^2 \setminus \mathrm{post}(f)$ is a Jordan curve, then each component σ of $f^{-1}(\gamma)$ is also a Jordan curve in $S^2 \setminus \mathrm{post}(f)$. Essentially, this follows from the fact that a suitable branch of f^{-1} gives a local homeomorphism of γ onto σ.

Suppose $\Gamma = \{\gamma_1, \ldots, \gamma_n\}$ is an invariant multicurve for a given Thurston map $f\colon S^2 \to S^2$. Then one can associate an $(n \times n)$-matrix with f and Γ as follows. Fix $i, j \in \{1, \ldots, n\}$, and let $\sigma_1, \ldots, \sigma_k$ be the components of $f^{-1}(\gamma_j)$ that are homotopic to γ_i rel. $\mathrm{post}(f)$ (here $k = k(i, j) \in \mathbb{N}_0$). Then each set σ_l is a Jordan curve in $S^2 \setminus \mathrm{post}(f)$, and $f|\sigma_l$ is a covering map of σ_l onto γ_j. Let

$$d_{i,j,l} := \deg(f|\sigma_l)$$

be the (unsigned) topological degree of this map (in this case, this is just the number of preimages of each point $p \in \gamma_j$ under the map $f|\sigma_l$). Then the *Thurston matrix* $A = A(f, \Gamma) = (a_{ij})$ is the matrix with non-negative entries

$$a_{ij} = \sum_{l=1}^{k(i,j)} \frac{1}{d_{i,j,l}}$$

for $i, j \in \{1, \ldots, n\}$; if $k(i, j) = 0$, then the sum is interpreted as the empty sum, in which case $a_{ij} = 0$.

A *Thurston obstruction* for a Thurston map f is an invariant multicurve Γ such that the spectral radius (which is the largest eigenvalue by the Perron-Frobenius theorem) of the Thurston matrix $A(f, \Gamma)$ is ≥ 1.

With these definitions Thurston's criterion can be formulated as follows.

THEOREM 2.17 (Thurston's characterization of rational maps). *Let $f\colon S^2 \to S^2$ be a Thurston map with a hyperbolic orbifold. Then f is Thurston equivalent to a rational map if and only if there exists no Thurston obstruction for f.*

The proof can be found in [**DH93**], see also [**Hu16**, Theorem 10.1.14]. We will not use this theorem in any essential way, and included its statement for general background and to put our work into context.

A Thurston map f with a parabolic orbifold is not covered by Theorem 2.17. In this case, the map has at most four postcritical points. If $\# \mathrm{post}(f) \leq 3$, then f is always equivalent to a rational map (see Proposition 7.1 and Theorem 7.3 (i)). If $\# \mathrm{post}(f) = 4$ and f has a parabolic orbifold, then the signature of f is $(2, 2, 2, 2)$. A criterion when such a map is equivalent to a rational map can be derived from Proposition 3.6 in combination with Theorem 3.22.

In Theorem 18.1 (ii) we give a criterion for an expanding Thurston map to be topologically conjugate to a rational map. The proof does not use Thurston's theorem.
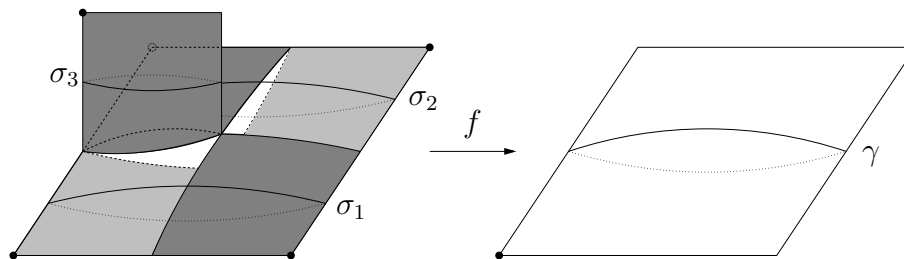
FIGURE 2.2. An obstructed map.

EXAMPLE 2.18. To illustrate Theorem 2.17 and the concepts we introduced for its formulation, we consider the Thurston map $f = h$ constructed in Section 1.3.

Recall that $\#\operatorname{post}(f) = 4$, and that the postcritical points of f are given by the vertices of the pillow on the right in Figure 2.2. The signature of the orbifold $\mathcal{O}_f$ of f is $(2, 6, 6, 6)$, and so $\mathcal{O}_f$ is hyperbolic.

Let γ be the Jordan curve indicated on the right in Figure 2.2. The preimage $f^{-1}(\gamma)$ of γ has three components σ_1, σ_2, σ_3 indicated on the left in Figure 2.2. The Jordan curves σ_1 and σ_2 are non-peripheral, while σ_3 is peripheral. Both curves σ_1 and σ_2 are homotopic to γ rel. $\operatorname{post}(f)$. Thus $\Gamma = \{\gamma\}$ is an invariant multicurve.

The degree of $f|\sigma_l \colon \sigma_l \to \gamma$ is 2 for $l = 1, 2$; so the Thurston matrix $A(f, \Gamma)$, which is a 1×1-matrix, has the single entry $1/2 + 1/2 = 1$.

It follows that the spectral radius of $A(f, \Gamma)$ is equal to 1. Hence Γ is a Thurston obstruction for f, and f is not Thurston equivalent to a rational map by Thurston's criterion.

Theorem 2.17 can be interpreted as a condition for the existence of a Thurston equivalence. It is complemented by the following statement which is essentially a uniqueness statement for Thurston equivalences.

THEOREM 2.19 (Thurston's uniqueness theorem). *Let $f, g \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be two rational Thurston maps with hyperbolic orbifolds, and suppose that there are two orientation-preserving homeomorphisms $h_0, h_1 \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ that are isotopic rel. $\operatorname{post}(f)$ and satisfy $h_0 \circ f = g \circ h_1$. Then there exists a conformal homeomorphism $\varphi \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ that is isotopic to h_0 and h_1 rel. $\operatorname{post}(f)$ and satisfies $\varphi \circ f = g \circ \varphi$.*

A conformal homeomorphism on $\widehat{\mathbb{C}}$ is of course a Möbius transformation. So in particular, if two rational Thurston maps with hyperbolic orbifolds are orientation-preserving Thurston equivalent (see the discussion after Definition 2.4), then they are conjugate by a Möbius transformation.

Theorem 2.19 is contained in [**DH93**]; there it is not formulated explicitly, but it can be easily derived from the considerations in this paper.

CHAPTER 3

Lattès maps

A *Lattès map* is a rational Thurston map that is expanding and has a parabolic orbifold. There are other equivalent ways to characterize these maps. For example, a Lattès map is a quotient of a holomorphic automorphism of the complex plane by the action of a crystallographic group or a quotient of a holomorphic endomorphism on a complex torus. We will explain this more precisely below.

These maps play a special role in the theory. On the one hand, they are very easy to construct and visualize, and provide a convenient class of examples (one was given in Section 1.1). On the other hand, they often show exceptional behavior compared to generic rational Thurston maps (that are expanding). This is already apparent in Thurston's characterization of rational maps (Theorem 2.17). In general, Lattès maps are distinguished among typical rational Thurston maps in terms of metric geometry (Theorem 18.1 (iii)), by their measure-theoretic properties (Theorem 19.4), or by their "combinatorial expansion rate" (Theorem 20.2). Some of these statements are among the main results of this work and so we will take a closer look at these maps. We will also define the related class of *Lattès-type* maps. These are Thurston maps with a parabolic orbifold and no periodic critical points, but not necessarily (equivalent to) rational maps.

Some aspects of a thorough treatment of the underlying theory are rather technical. As we do not want to overburden the reader with details at this point, we will rely on various results that are more fully developed in the appendix.

To motivate our definition of Lattès maps in terms of three equivalent conditions, we will now consider a specific example. For precise definitions of the terminology in the ensuing discussion we refer to the beginning of Section 3.1.

Let f be the map from Section 1.1 (there denoted by g). Then f is a rational Thurston map that is expanding, or equivalently, has no periodic critical points. Its orbifold has signature $(2, 2, 2, 2)$, and is hence parabolic.

The map f is a quotient of the automorphism $A\colon \mathbb{C} \to \mathbb{C}$, $z \mapsto A(z) = 2z$, on $\mathbb{C}$ by a holomorphic map $\Theta\colon \mathbb{C} \to \widehat{\mathbb{C}}$ in the sense that the following diagram commutes:

(3.1)
$$
\begin{array}{ccc}
\mathbb{C} & \xrightarrow{\;A\;} & \mathbb{C} \\
{\scriptstyle \Theta}\big\downarrow & & \big\downarrow{\scriptstyle \Theta} \\
\widehat{\mathbb{C}} & \xrightarrow{\;f\;} & \widehat{\mathbb{C}}.
\end{array}
$$

Here Θ is essentially a Weierstrass $\wp$-function for the lattice $\Gamma = \mathbb{Z} \oplus \mathbb{Z}\boldsymbol{i}$ (see Section 3.5 for the definition of $\wp$ and a related discussion). Note that $\Theta(z) = \Theta(w)$ for $z, w \in \mathbb{C}$ if and only if $w = \pm z + m + n\boldsymbol{i}$ with $m, n \in \mathbb{Z}$. This last condition can most conveniently be expressed in terms of an action of a *crystallographic group* of orientation-preserving isometries on $\mathbb{C}$ (see Section 3.1).

Indeed, let G be the group of all maps $g\colon \mathbb{C} \to \mathbb{C}$ of the form

$$
g(z) = \pm z + m + n\boldsymbol{i},
$$

where $m, n \in \mathbb{Z}$. Then G is a crystallographic group and the map Θ is *induced* by G in the sense that $\Theta(z) = \Theta(w)$ for $z, w \in \mathbb{C}$ if and only if there exists $g \in G$ such that $w = g(z)$ (related concepts and facts are discussed in more detail in Section A.7). This implies that the quotient $\mathbb{C}/G$ can be identified with $\widehat{\mathbb{C}}$, and the map Θ with the quotient map $\mathbb{C} \to \mathbb{C}/G$ (see Corollary A.21).

The property that allows us to pass to a quotient f in (3.1) is that the map $z \mapsto A(z) = 2z$ is *G-equivariant* (see Lemma A.22). This means that A maps points that are in the same G-orbit to points that are also in the same G-orbit, or equivalently that

(3.2) $$A \circ g \circ A^{-1} \in G \text{ for all } g \in G.$$

The translations in G form a subgroup G_{tr} isomorphic (as a group) to $\mathbb{Z}^2 = \mathbb{Z} \oplus \mathbb{Z}\boldsymbol{i}$. The quotient $\mathbb{C}/G_{\mathrm{tr}}$ is naturally a *complex torus* $\mathbb{T}$, i.e., a Riemann surface whose underlying 2-manifold is a 2-dimensional torus (for more on tori see Section A.8). The maps $A\colon \mathbb{C} \to \mathbb{C}$ and $\Theta\colon \mathbb{C} \to \widehat{\mathbb{C}}$ descend to $\mathbb{T}$, and we obtain holomorphic maps $\overline{A}\colon \mathbb{T} \to \mathbb{T}$ and $\overline{\Theta}\colon \mathbb{T} \to \widehat{\mathbb{C}}$ such that $f \circ \overline{\Theta} = \overline{\Theta} \circ \overline{A}$. So we have the following commutative diagram:

(3.3)
$$
\begin{array}{ccc}
\mathbb{T} & \xrightarrow{\;\overline{A}\;} & \mathbb{T} \\
{\scriptstyle \overline{\Theta}}\big\downarrow & & \big\downarrow{\scriptstyle \overline{\Theta}} \\
\widehat{\mathbb{C}} & \xrightarrow{\;f\;} & \widehat{\mathbb{C}}.
\end{array}
$$

We call a non-constant holomorphic map $\overline{A}\colon \mathbb{T} \to \mathbb{T}$ on a complex torus $\mathbb{T}$ a *holomorphic torus endomorphism*. The Riemann-Hurwitz formula (2.3) implies that such a map $\overline{A}$ has no critical points and is hence a covering map (in the usual topological sense; see Section A.5).

The relations of our example f to crystallographic groups or to holomorphic torus endomorphisms hold for a more general class of rational maps, called *Lattès maps*, as the following statement shows.

THEOREM 3.1 (Characterization of Lattès maps). *Let $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a map. Then the following conditions are equivalent:*

(i) *f is a rational Thurston map that has a parabolic orbifold and no periodic critical points.*

(ii) *There exists a crystallographic group G, a G-equivariant holomorphic map $A\colon \mathbb{C} \to \mathbb{C}$ of the form $A(z) = \alpha z + \beta$, where $\alpha, \beta \in \mathbb{C}$, $|\alpha| > 1$, and a holomorphic map $\Theta\colon \mathbb{C} \to \widehat{\mathbb{C}}$ induced by G such that $f \circ \Theta = \Theta \circ A$.*

(iii) *There exists a complex torus $\mathbb{T}$, a holomorphic torus endomorphism $\overline{A}\colon \mathbb{T} \to \mathbb{T}$ with $\deg(\overline{A}) > 1$, and a non-constant holomorphic map $\overline{\Theta}\colon \mathbb{T} \to \widehat{\mathbb{C}}$ such that $f \circ \overline{\Theta} = \overline{\Theta} \circ \overline{A}$.*

So in (ii) the map f is given as in (3.1), and in (iii) as in (3.3). We will see that we have $\deg(f) = \deg(\overline{A}) = |\alpha|^2 > 1$ (Lemma 3.16).

As we already indicated, the previous theorem motivates the following definition.

DEFINITION 3.2 (Lattès maps). A map $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is called a *Lattès map* if it satisfies one of the conditions (and hence every condition) in Theorem 3.1.

The terminology is not uniform in the literature and some authors use the term "Lattès map" with a slightly different meaning (see the discussion in Section 3.5). Lattès maps became more widely known through Lattès paper [**La18**], but they had been studied about half a century earlier by Schroeder, for example. See [**Mi06a**] for more on the history of these maps.

Theorem 3.1 is well known (see, for example, [**Mi06a**]). We will prove it in Sections 3.1 and 3.2. The map A in (ii) is subject to strong further restrictions. See Proposition 3.14 (or [**Mi06a**] and [**DH84**, Appendix]) for more details.

By Proposition 2.3, condition (i) can equivalently be expressed as:

(i') *f is a rational Thurston map that has a parabolic orbifold and is expanding.*

This is how we introduced Lattès maps in the beginning of the chapter.

The most convenient way to construct Lattès maps is based on condition (ii) in Theorem 3.1. One starts with a crystallographic group G and an G-equivariant map A as in this statement. Then there exists a holomorphic branched covering map $\Theta\colon \mathbb{C} \to \widehat{\mathbb{C}}$ induced by G; it is unique up to postcomposition by a Möbius transformation (see Proposition 3.9). The existence of a Lattès map f as in (3.1) then follows from the G-equivariance of A (see Lemma A.22).

If $\alpha_f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{N}}$ is the ramification function of f (see Definition 2.7), then $\Theta\colon \mathbb{C} \to \widehat{\mathbb{C}}$ is a holomorphic branched covering map such that $\deg(\Theta, z) = \alpha_f(\Theta(z))$ for all $z \in \mathbb{C}$ (this follows from Remark 3.17). Therefore, Θ is in fact the *universal orbifold covering map* of $\mathcal{O}_f = (\widehat{\mathbb{C}}, \alpha_f)$ (see Theorem 3.10 and Section A.9).

It is quite natural to consider more general maps f as in (3.1) or as in (3.3), where the maps involved are branched covering maps, but not necessarily holomorphic. To state this more precisely, we first recall some terminology.

As usual, we call a map $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ *affine*, if it has the form

$$(3.4) \qquad\qquad A(u) = L_A(u) + u_0, \quad u \in \mathbb{R}^2,$$

where $L_A\colon \mathbb{R}^2 \to \mathbb{R}^2$ is $\mathbb{R}$-linear and $u_0 \in \mathbb{R}^2$. We call L_A the *linear part* of A.

Let G be a crystallographic group not isomorphic to $\mathbb{Z}^2$. Then the quotient $\mathbb{R}^2/G$ is homeomorphic to a 2-sphere S^2 and the quotient map $\Theta\colon \mathbb{R}^2 \to S^2 \cong \mathbb{R}^2/G$ is a branched covering map induced by G. Suppose $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ is an affine map that is G-equivariant and whose linear part L_A satisfies $\det(L_A) > 1$. In this case, there is a Thurston map $f\colon S^2 \to S^2$ such that the diagram

$$(3.5) \qquad\qquad \begin{array}{ccc} \mathbb{R}^2 & \xrightarrow{\;A\;} & \mathbb{R}^2 \\ {\scriptstyle \Theta}\downarrow & & \downarrow{\scriptstyle \Theta} \\ S^2 & \xrightarrow{\;f\;} & S^2 \end{array}$$

commutes (see the beginning of Section 3.4). This is the basis of the following definition.

DEFINITION 3.3 (Lattès-type maps). Suppose $f\colon S^2 \to S^2$ is a map such that there exists a crystallographic group G, an affine map $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ with $\det(L_A) > 1$ that is G-equivariant, and a branched covering map $\Theta\colon \mathbb{R}^2 \to S^2$ induced by G such that $f \circ \Theta = \Theta \circ A$. Then f is called a *Lattès-type map*.

So Lattès-type maps are given as in (3.5), where A is affine. It is clear that every Lattès map belongs to this class. For the degree of such a map we have $\deg(f) = \det(L_A)$ (see Lemma 3.16). The requirement $\det(L_A) > 1$ guarantees the condition $\deg(f) \geq 2$ which is part of our definition of a Thurston map. One can also consider maps as in Definition 3.3 with $\det(L_A) = 1$ or $\det(L_A) < 0$. This gives homeomorphisms and orientation-reversing maps, respectively. According to our definition, no such map is a Thurston map.

Before we discuss some other properties of Lattès-type maps, we will first define another generalization of Lattès maps based on (3.3). We denote by T^2 a 2-dimensional torus, now considered as a purely topological object with no complex structure. If $\overline{A} \colon T^2 \to T^2$ is a branched covering map, then again by the Riemann-Hurwitz formula (2.3) the map $\overline{A}$ cannot have any critical points and must be an orientation-preserving covering map. We call such a map $\overline{A}$ a *(topological) torus endomorphism*. So a continuous map $\overline{A} \colon T^2 \to T^2$ is a torus endomorphism precisely if it is an orientation-preserving local homeomorphism.

DEFINITION 3.4 (Quotients of torus endomorphisms). Let $f \colon S^2 \to S^2$ be a map on a 2-sphere S^2 such that there exists a torus endomorphism $\overline{A} \colon T^2 \to T^2$ with $\deg(\overline{A}) \geq 2$, and a branched covering map $\overline{\Theta} \colon T^2 \to S^2$ such that $f \circ \overline{\Theta} = \overline{\Theta} \circ \overline{A}$. Then f is called a *quotient of a torus endomorphism*.

In this case, we have a commutative diagram of the form

$$
(3.6) \qquad\qquad
\begin{array}{ccc}
T^2 & \xrightarrow{\ \overline{A}\ } & T^2 \\
{\scriptstyle \overline{\Theta}}\downarrow & & \downarrow{\scriptstyle \overline{\Theta}} \\
S^2 & \xrightarrow{\ f\ } & S^2.
\end{array}
$$

Properties of quotients of torus endomorphisms are recorded in Lemma 3.12. In particular, every such map is a Thurston map without periodic critical points. Lattès-type maps are in this class.

PROPOSITION 3.5. *Every Lattès-type map $f \colon S^2 \to S^2$ is a quotient of a torus endomorphism and hence a Thurston map. It has a parabolic orbifold and no periodic critical points.*

This implies that the orbifold of every Lattès-type map has one of the signatures $(2, 2, 2, 2)$, $(2, 4, 4)$, $(3, 3, 3)$, or $(2, 3, 6)$. The last three signatures do not lead to genuinely new maps, as each Lattès-type map whose orbifold has such a signature is topologically conjugate to a Lattès map (Proposition 3.18). The most interesting case is signature $(2, 2, 2, 2)$. More details on these maps can be found in Example 3.20

(see also Theorem 3.22). These maps include *flexible Lattès maps* (see Definition 3.26 and the discussion that follows). The last statement in Proposition 3.5 essentially characterizes Lattès-type maps.

PROPOSITION 3.6. *Let $f\colon S^2 \to S^2$ be a Thurston map. Then f is Thurston equivalent to a Lattès-type map if and only if f has a parabolic orbifold and no periodic critical points.*

If f has a parabolic orbifold $\mathcal{O}_f$, but also periodic critical points, then the signature of $\mathcal{O}_f$ is (∞, ∞) or $(2, \infty, \infty)$. It is easy to classify these maps up to Thurston equivalence as well (see Theorem 7.4).

Each Lattès map is expanding, but this is not always true for a Lattès-type map (see Example 6.14). One can state a simple criterion though when this is the case. Namely, a Lattès-type map is expanding if and only if the two (possibly complex) eigenvalues λ_1 and λ_2 of the linear part L_A of the affine map A in (3.5) satisfy $|\lambda_1|, |\lambda_2| > 1$ (Proposition 6.11).

Every Lattès map is a Lattès-type map, and every Lattès-type map is a quotient of a torus endomorphism. On the other hand, not every Lattès-type map is (conjugate to) a Lattès map. It is a natural question whether every quotient of a torus endomorphism f is Thurston equivalent to a Lattès-type map. One can show that this is true if f is expanding (in this case, f is even conjugate to a Lattès-type map), but we have been unable to answer this question in full generality.

3.1. Crystallographic groups and Lattès maps

In this section we focus on maps as in statement (ii) of Theorem 3.1. We first review some facts related to crystallographic groups. For a more detailed discussion related to group actions and quotient spaces see Section A.7.

We use the notation

$$\mathrm{Aut}(\mathbb{C}) = \{z \in \mathbb{C} \mapsto \alpha z + \beta : \alpha, \beta \in \mathbb{C},\ \alpha \neq 0\}$$

for the group of all holomorphic automorphisms of $\mathbb{C}$ and

$$\mathrm{Isom}(\mathbb{C}) = \{z \in \mathbb{C} \mapsto \alpha z + \beta : \alpha, \beta \in \mathbb{C},\ |\alpha| = 1\} \subset \mathrm{Aut}(\mathbb{C})$$

for the group of all orientation-preserving isometries of $\mathbb{C}$ (equipped with the Euclidean metric).

Let G be a group of homeomorphisms acting on $\mathbb{C}$. If $z \in \mathbb{C}$, then we denote by $G_z := \{g \in G : g(z) = z\}$ its *stabilizer subgroup* and by $Gz = \{g(z) : g \in G\}$ its *orbit under G* or *G-orbit*. The group G induces a natural equivalence relation on $\mathbb{C}$ whose equivalence classes are given by the G-orbits. The corresponding quotient space is denoted by $\mathbb{C}/G$.

The group G acts *properly discontinuously* on $\mathbb{C}$ if for each compact set $K \subset \mathbb{C}$ there are only finitely many maps $g \in G$ with $g(K) \cap K \neq \emptyset$. Then the stabilizer G_z is finite for each $z \in \mathbb{C}$. The group G acts *cocompactly* on $\mathbb{C}$ if there exists a compact set $K \subset \mathbb{C}$ such that the sets $g(K)$, $g \in G$, cover $\mathbb{C}$. In this case, $\mathbb{C}/G$ is compact.

We call G a *(planar) crystallographic group* if each element $g \in G$ is an orientation-preserving isometry on $\mathbb{C}$ and if the action of G on $\mathbb{C}$ is properly discontinuous and cocompact. Note that this definition of a crystallographic group is more restrictive than usual, since we require that the isometries in G are orientation-preserving.

We say that two crystallographic groups G and $\widetilde{G}$ are *conjugate* if there exists $h \in \mathrm{Aut}(\mathbb{C})$ such that

$$\widetilde{G} = h \circ G \circ h^{-1} := \{h \circ g \circ h^{-1} : g \in G\}.$$

The following statement gives a classification of crystallographic groups up to conjugacy.

THEOREM 3.7 (Classification of crystallographic groups). *Let $G \subset \mathrm{Isom}(\mathbb{C})$ be a planar crystallographic group. Then G is conjugate to one of the following groups $\widetilde{G}$ consisting of all $g \in \mathrm{Isom}(\mathbb{C})$ of the form*

(torus) $z \mapsto g(z) = z + m + n\tau, \quad m, n \in \mathbb{Z};$

(2222) $z \mapsto g(z) = \pm z + m + n\tau, \quad m, n \in \mathbb{Z};$

(244) $z \mapsto g(z) = \boldsymbol{i}^k z + m + n\boldsymbol{i}, \quad m, n \in \mathbb{Z},\ k = 0, 1, 2, 3;$

(333) $z \mapsto g(z) = \omega^{2k} z + m + n\omega, \quad m, n \in \mathbb{Z},\ k = 0, 1, 2;$

(236) $z \mapsto g(z) = \omega^k z + m + n\omega, \quad m, n \in \mathbb{Z},\ k = 0, \ldots, 5.$

Here $\tau \in \mathbb{C}$ is a fixed number with $\mathrm{Im}(\tau) > 0$ in the first two cases and $\omega = e^{i\pi/3}$ in the last two cases.

This classification of (planar) crystallographic groups is classical. Proofs can be found in [**Be87**] and [**Ar91**]; see also [**We52**].

We used Conway's orbifold notation for planar crystallographic groups, see [**Co92**] (with one difference: Conway denotes the (torus) case by $(\circ)$). In Theorem 3.7 the group G determines the type of the conjugate group $\widetilde{G}$ uniquely. Accordingly, we speak of a crystallographic group G of type (2222), etc., if $\widetilde{G}$ belongs to the corresponding class. This terminology is explained by the fact that the quotient space $\mathbb{C}/G$ is a torus in the first case of the theorem and homeomorphic to $\widehat{\mathbb{C}}$ in the other cases (see Section 3.5 for more details). In addition, G induces a natural ramification function α on $\widehat{\mathbb{C}} = \mathbb{C}/G$ so that the orbifold $(\widehat{\mathbb{C}}, \alpha)$ has a signature as indicated by its type (see the discussion below).

REMARK 3.8. A crystallographic group G is isomorphic as a group to its conjugate $\widetilde{G}$. This implies that if G is of (torus) type, then G is isomorphic to $\mathbb{Z}^2$. In the other cases, G is isomorphic to a semidirect product $\mathbb{Z}^2 \rtimes \mathbb{Z}_k$ of $\mathbb{Z}^2$ and a cyclic group $\mathbb{Z}_k = \mathbb{Z}/k\mathbb{Z}$. Here $k = 2, 4, 3, 6$ if G is of type (2222), (244), (333), or (236), respectively. To see this, one considers the isomorphic group $\widetilde{G}$ and identifies $\mathbb{Z}^2$ with the lattice $\Gamma = \mathbb{Z} \oplus \mathbb{Z}\tau$, where $\tau \in \mathbb{C}$ satisfies $\operatorname{Im}(\tau) > 0$ in case (2222), $\tau = \boldsymbol{i}$ in case (244), and $\tau = \omega$ in cases (333) and (236). In addition, we identify $\mathbb{Z}_k$ with the multiplicative group consisting of the k-th roots of unity. Then each element in $\mathbb{Z}_k$ acts by multiplication as an automorphism on $\Gamma \cong \mathbb{Z}^2$ and one derives the isomorphism $G \cong \widetilde{G} \cong \mathbb{Z}^2 \rtimes \mathbb{Z}_k$. All this is standard and well known, and we skip the details.

If G is a crystallographic group, we denote by G_{tr} the subgroup consisting of all translations in G, i.e., G_{tr} consists of all maps $g \in G$ of the form $z \mapsto g(z) = z + \gamma$ with $\gamma \in \mathbb{C}$. Theorem 3.7 implies that G_{tr} is a normal subgroup of finite index in G, and that there is a lattice $\Gamma \subset \mathbb{C}$ such that $G_{\mathrm{tr}} = \{z \mapsto z + \gamma : \gamma \in \Gamma\}$. This lattice Γ has *rank* 2 in the sense that it spans $\mathbb{R}^2 \cong \mathbb{C}$ (see Section A.8 for more discussion). We call it the *underlying lattice* of the crystallographic group G. The underlying lattice Γ of $\widetilde{G}$ as in Theorem 3.7 is equal to $\mathbb{Z} \oplus \mathbb{Z}\tau$ in cases (torus) and (2222), to $\mathbb{Z} \oplus \mathbb{Z}\boldsymbol{i}$ in case (244), and to $\mathbb{Z} \oplus \mathbb{Z}\omega$ in cases (333) and (236). Note that a crystallographic group G is of (torus) type if and only if $G = G_{\mathrm{tr}}$, or, equivalently, if and only if G is isomorphic to $\mathbb{Z}^2$.

Crystallographic groups G of type (244), (333), and (236) are represented in Figure 3.1, Figure 3.2, and Figure 3.3, respectively. For each type (abc) we see (part of) a tiling of $\mathbb{C}$ given by isometric copies of Euclidean triangles with angles $\pi/a, \pi/b, \pi/c$. The corresponding group G consists of all orientation-preserving isometries of the plane $\mathbb{C}$ that keep the tiling invariant. This means that each group element $g \in G$ maps each triangle in the tiling to another one of the same color. Moreover, g maps each point marked by a dot to another such point. If we assume that $0 \in \mathbb{C}$ is one of these points, then all points marked by a dot form the underlying lattice Γ of G, i.e., the orbit of $0 \in \mathbb{C}$ by the group of translations $G_{\mathrm{tr}} \subset G$. Actually, the full G-orbit of 0 is equal to its G_{tr}-orbit Γ.

Suppose G is a crystallographic group and $z, w \in \mathbb{C}$ are contained in the same G-orbit. Then their stabilizers G_z and G_w have the same order $\#G_z = \#G_w$, since they are conjugate subgroups of G. Each non-trivial stabilizer G_z is cyclic and of order 2, 3, 4, or 6, as can be seen from Theorem 3.7 (actually, an independent proof of this fact is
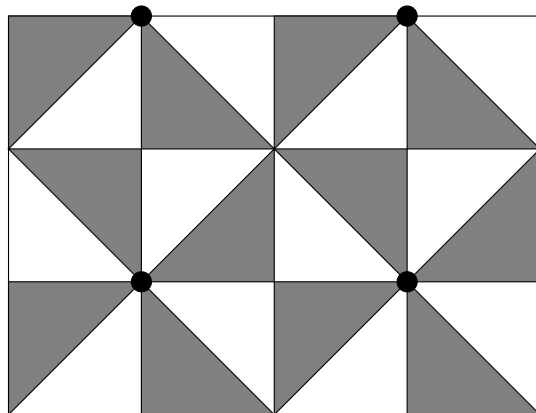
FIGURE 3.1. Invariant tiling for type (244).
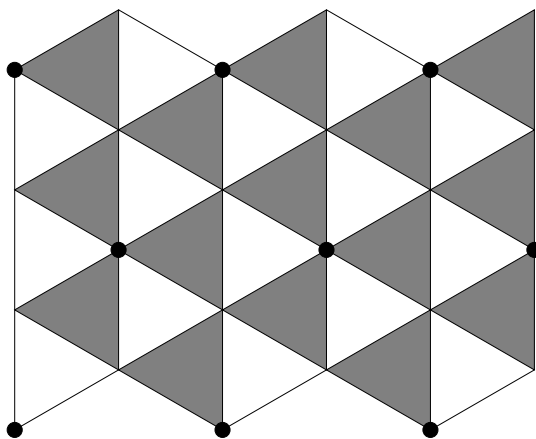
FIGURE 3.2. Invariant tiling for type (333).
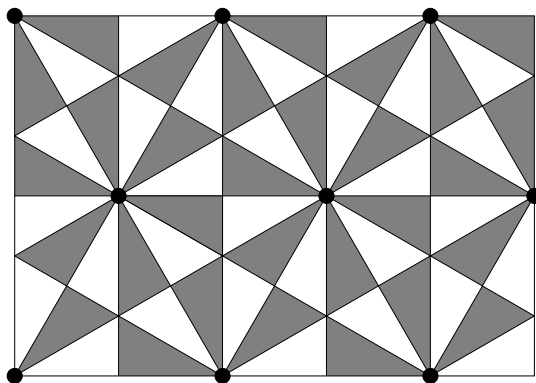
FIGURE 3.3. Invariant tiling for type (236).

one of the main steps in the proof of Theorem 3.7). The numbers in the label for the type of a group indicate the orders of non-trivial stabilizers in distinct orbits. For example, a crystallographic group of type (236) has three distinguished orbits consisting of points with non-trivial stabilizers of order 2, 3, or 6. Note that every $g \in G_z$ is a rotation around z, since it is an orientation-preserving isometry that fixes z.

Recall that a map $\Theta \colon \mathbb{C} \to \widehat{\mathbb{C}}$ is *induced* by a crystallographic group G if $\Theta(z) = \Theta(w)$ for $z, w \in \mathbb{C}$ if and only if there is a $g \in G$ such that $w = g(z)$.

PROPOSITION 3.9. *Let G be a crystallographic group not isomorphic to $\mathbb{Z}^2$. Then there exists a holomorphic map $\Theta \colon \mathbb{C} \to \widehat{\mathbb{C}}$ that is induced by G. Moreover, Θ is a branched covering map, satisfies*

$$(3.7) \qquad\qquad \deg(\Theta, z) = \#G_z$$

for all $z \in \mathbb{C}$, and is unique up to postcomposition by a Möbius transformation.

For the notion of a branched covering map that applies here see Definition A.7. The uniqueness statement means more explicitly that if $\Theta \colon \mathbb{C} \to \widehat{\mathbb{C}}$ and $\widetilde{\Theta} \colon \mathbb{C} \to \widehat{\mathbb{C}}$ are holomorphic maps that are both induced by G, then there is a Möbius transformation $\varphi \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ such that $\widetilde{\Theta} = \varphi \circ \Theta$.

PROOF. Suppose that we have two crystallographic groups G and $\widetilde{G}$ that are conjugate, meaning that $\widetilde{G} = h \circ G \circ h^{-1}$ for some $h \in \mathrm{Aut}(\mathbb{C})$. Then a map $\widetilde{\Theta} \colon \mathbb{C} \to \widehat{\mathbb{C}}$ is induced by $\widetilde{G}$ if and only if $\Theta := \widetilde{\Theta} \circ h$ is induced by G. To see this, suppose $\widetilde{\Theta} \colon \mathbb{C} \to \widehat{\mathbb{C}}$ is induced by $\widetilde{G}$ and let $z, w \in \mathbb{C}$ be arbitrary. Then $\Theta(z) = \Theta(w)$ if and only if $\widetilde{\Theta}(h(z)) = \widetilde{\Theta}(h(w))$. This in turn is equivalent to the existence of $\widetilde{g} \in \widetilde{G}$ with $h(w) = \widetilde{g}(h(z))$ or $w = (h^{-1} \circ \widetilde{g} \circ h)(z)$. Since $g := h^{-1} \circ \widetilde{g} \circ h \in G$, and each $g \in G$ is of this form, Θ is induced by G as desired. Conversely, if Θ is induced by G, then by reversing the roles of $\widetilde{\Theta}$ and Θ in the previous argument, we see that $\widetilde{\Theta}$ is induced by $\widetilde{G}$.

This shows that if G is a crystallographic group as in the statement, then we may assume that $G = \widetilde{G}$ for one of the groups $\widetilde{G}$ as in Theorem 3.7 not of (torus) type. If G is of type (2222) with corresponding lattice $\Gamma = \mathbb{Z} \oplus \mathbb{Z}\tau$ (where $\mathrm{Im}(\tau) > 0$), we can choose Θ as the Weierstraß $\wp$-function for the lattice Γ. In the other cases, the map Θ can be constructed from the tiling in Figure 3.1, Figure 3.2, or Figure 3.3. Roughly speaking, we obtain Θ by first mapping a triangle in the tiling conformally to the upper half-plane and then extending

the map by Schwarz reflection to all of $\mathbb{C}$. In all cases we obtain a holomorphic map $\Theta\colon \mathbb{C} \to \widehat{\mathbb{C}}$ induced by G. We postpone the discussion of the details to Section 3.5.

If Γ is the underlying lattice of G, then each $\gamma \in \Gamma$ is a period of Γ. So Θ is an elliptic function, i.e., a (non-constant) doubly-periodic holomorphic map $\Theta\colon \mathbb{C} \to \widehat{\mathbb{C}}$. This implies that Θ is surjective, because the image $\Theta(\mathbb{C})$ is compact and open in $\widehat{\mathbb{C}}$, and so $\Theta(\mathbb{C}) = \widehat{\mathbb{C}}$.

To see (3.7), let $z_0 \in \mathbb{C}$ be arbitrary and $d \coloneqq \#G_{z_0}$. We have to show that Θ is locally d-to-1 near z_0. Since G acts properly discontinuously on $\mathbb{C}$, the orbit Gz_0 is a discrete set in $\mathbb{C}$. This implies that

$$\epsilon_0 \coloneqq \inf\{|g(z_0) - z_0| : g \in G \setminus G_{z_0}\} > 0.$$

We pick $\epsilon \in (0, \epsilon_0/2]$ and consider the Euclidean disk $B \coloneqq B_{\mathbb{C}}(z_0, \epsilon)$. Then

(3.8) $g(B) \cap h(B) \neq \emptyset$ for $g, h \in G$ if and only if $h^{-1} \circ g \in G_{z_0}$.

Indeed, suppose $g, h \in G$ and $g(u) = h(v)$ for some $u, v \in B$. Since g and h are isometries, we then have

$$|(h^{-1} \circ g)(z_0) - z_0| \leq |(h^{-1} \circ g)(z_0) - (h^{-1} \circ g)(u)| + |(h^{-1} \circ g)(u) - z_0|$$
$$= |g(z_0) - g(u)| + |v - z_0|$$
$$= |z_0 - u| + |v - z_0| < 2\epsilon \leq \epsilon_0,$$

and so $h^{-1} \circ g \in G_{z_0}$ by definition of ϵ_0. Conversely, if $h^{-1} \circ g \in G_{z_0}$, then $h^{-1} \circ g$ is a rotation around z_0. Hence $(h^{-1} \circ g)(B) = B$ which actually implies $g(B) = h(B)$.

Now if $u, v \in B$ and $\Theta(u) = \Theta(v)$, then there exists $g \in G$ with $v = g(u)$, because Θ is induced by G. In particular, $v \in B \cap g(B)$, and so $g \in G_{z_0}$ by (3.8). It follows that two points in B have the same image under Θ if and only if they belong to the same G_{z_0}-orbit. Now G_{z_0} is a cyclic group of order d consisting of rotations around z_0. Hence the orbit of each point $z \in B \setminus \{z_0\}$ consists of precisely d points in B. This shows that Θ is indeed locally d-to-1 near z_0 (for a similar argument in greater generality see the proof of Proposition A.29 (ii)).

From the previous considerations one can also easily see that Θ is a branched covering map. Indeed, since $z_0 \in \mathbb{C}$ was arbitrary and Θ is surjective, it is enough to show that $w_0 \coloneqq \Theta(z_0)$ has a neighborhood that is evenly covered by Θ as in Definition A.7.

The invariance property of Θ with respect to the cyclic rotation group G_{z_0} on B implies that

$$\Theta(z) = f((z - z_0)^d/\epsilon^d) \quad \text{for } z \in B,$$

where $f \colon \mathbb{D} \to \widehat{\mathbb{C}}$ is an injective holomorphic map with $f(0) = w_0$.

In particular, $V := \Theta(B) = f(\mathbb{D})$ is a topological disk containing w_0; moreover, if we define $\varphi \colon B \to \mathbb{D}$ by $\varphi(z) = (z - z_0)/\epsilon$ for $z \in B$ and set $\psi := f^{-1} \colon V \to \mathbb{D}$, then φ and ψ are orientation-preserving homeomorphisms with $\varphi(z_0) = 0$, $\psi(w_0) = 0$, and

(3.9) $$(\psi \circ \Theta \circ \varphi^{-1})(u) = u^d$$

for all $u \in \mathbb{D}$.

Since Θ is induced by G, we have by (3.8) that

$$\Theta^{-1}(V) = \bigcup_{g \in G} g(B) = \bigcup_{g \in I} g(B),$$

where $I \subset G$ is a set that contains precisely one element from each left coset of G_{z_0} in G. Note that by (3.8) this makes the sets in the latter union pairwise disjoint. For each set $g(B)$ with $g \in I$, we have a relation as in (3.9) if we replace $\varphi \colon B \to \mathbb{D}$ by $\varphi \circ g^{-1} \colon g(B) \to \mathbb{D}$. This shows that V is evenly covered by Θ. It follows that Θ is a branched covering map.

It remains to show the uniqueness statement. To this end, suppose $\Theta \colon \mathbb{C} \to \widehat{\mathbb{C}}$ and $\widetilde{\Theta} \colon \mathbb{C} \to \widehat{\mathbb{C}}$ are two holomorphic maps that are induced by G. We pick three points $z_0, z_1, z_\infty \in \mathbb{C}$ with distinct G-orbits. By postcomposing Θ and $\widetilde{\Theta}$ with suitable Möbius transformations, we may assume that these maps send z_0, z_1, z_∞ to $0, 1, \infty$, respectively. Then the zero sets of Θ and $\widetilde{\Theta}$ are the same and given by the G-orbit of z_0. Moreover, by (3.7) the order of each zero is the same for Θ and $\widetilde{\Theta}$. Similarly, Θ and $\widetilde{\Theta}$ have poles at the same points of the same order.

This implies that the quotient $h = \widetilde{\Theta}/\Theta$ has removable singularities and has an extension to a holomorphic function on $\mathbb{C}$. In addition, h is invariant under precomposition by elements in G and in particular under the translations induced by the associated lattice Γ of G. In other words, h is a doubly-periodic holomorphic function without poles and hence constant (by Liouville's first theorem for elliptic functions). We conclude that there exists a constant $c \in \mathbb{C}$ such that $\widetilde{\Theta} = c\Theta$. Since $\widetilde{\Theta}(z_1) = \Theta(z_1) = 1$, it follows that $c = 1$, and so $\widetilde{\Theta} = \Theta$ as desired. $\square$

Crystallographic groups are closely related to parabolic orbifolds $(\widehat{\mathbb{C}}, \alpha)$ satisfying $\alpha(p) < \infty$ for $p \in \widehat{\mathbb{C}}$. These orbifolds have one of the signatures $(2, 2, 2, 2)$, $(2, 4, 4)$, $(3, 3, 3)$, or $(2, 3, 6)$. This corresponds precisely to the types of crystallographic groups that are not of (torus) type, i.e., not isomorphic to $\mathbb{Z}^2$. This relation is based on the *orbifold uniformization theorem* which is discussed in detail in the appendix

(see Section A.9). For the present purpose we will formulate a relevant special case explicitly. First, we recall some terminology.

If $\Theta\colon \mathbb{C} \to \widehat{\mathbb{C}}$ is a branched covering map, then a *deck transformation* of Θ is a homeomorphism $g\colon \mathbb{C} \to \mathbb{C}$ such that $\Theta \circ g = \Theta$. If Θ is holomorphic, then this is also true for each deck transformation g of Θ (see the last part of Lemma A.15) and so $g \in \mathrm{Aut}(\mathbb{C})$. The deck transformations of Θ form a group G.

We say that $\Theta\colon \mathbb{C} \to \widehat{\mathbb{C}}$ is a *regular* branched covering map if its deck transformations act transitively on the fibers of Θ; this means that if $z, w \in \mathbb{C}$ and $\Theta(z) = \Theta(w)$, then there exists $g \in G$ such that $w = g(z)$. Note that Θ is regular if and only if it is induced by its group of deck transformations G.

THEOREM 3.10. *Let* $(\widehat{\mathbb{C}}, \alpha)$ *be a parabolic orbifold with* $\alpha(p) < \infty$ *for all* $p \in \widehat{\mathbb{C}}$. *Then there exists a holomorphic branched covering map* $\Theta\colon \mathbb{C} \to \widehat{\mathbb{C}}$ *such that*

$$(3.10) \qquad \deg(\Theta, z) = \alpha(\Theta(z)) \ \textit{for each } z \in \mathbb{C}.$$

The branched covering map Θ *is regular and its deck transformation group* G *is a crystallographic group.*

Moreover, if $\widetilde{\Theta}\colon \mathbb{C} \to \widehat{\mathbb{C}}$ *is another holomorphic map satisfying* (3.10), *then there exists* $h \in \mathrm{Aut}(\mathbb{C})$ *such that* $\widetilde{\Theta} = \Theta \circ h$.

The essentially unique map Θ is the *universal orbifold covering map* of the orbifold $(\widehat{\mathbb{C}}, \alpha)$ (see Section A.9).

We will not prove Theorem 3.10 here, but refer the reader to the appendix, where more general facts are discussed from which this theorem can be derived. More specifically, the existence of Θ in Theorem 3.10 is a special case of Theorem A.24 (see also the discussion below). The uniqueness statement for Θ follows from Corollary A.27 and Remark A.28. Finally, the statement about the deck transformation group of Θ follows from Proposition A.29.

Since Θ is regular, it is induced by the crystallographic group G given by its deck transformations. So we can naturally identify $\mathbb{C}/G$ with $\widehat{\mathbb{C}}$ by sending each orbit Gz of G (considered as a point in $\mathbb{C}/G$) to $\Theta(z) \in \widehat{\mathbb{C}}$ (see Corollary A.21). Under this identification, the quotient map $\mathbb{C} \to \mathbb{C}/G$ then corresponds to Θ. In particular, $\mathbb{C}/G$ is a topological 2-sphere and so G is not isomorphic to $\mathbb{Z}^2$ (in which case $\mathbb{C}/G$ is a torus).

If $\Theta\colon \mathbb{C} \to S^2$ is any (not necessarily holomorphic) branched covering map induced by a crystallographic group G (as the universal

orbifold covering map just discussed), then

$$(3.11) \qquad\qquad \deg(\Theta, z) = \#G_z$$

for $z \in \mathbb{C}$. This follows from an argument as in the proof of Proposition 3.9. Indeed, for each $z \in \mathbb{C}$ the stabilizer G_z is a cyclic rotation group fixing z. So the G-orbit of a point $w \neq z$ sufficiently close to z contains precisely $d = \#G_z$ distinct points near z. Since Θ is induced by G, this means that near z the map Θ is locally d-to-1 and (3.11) follows.

The relation (3.11) implies that the crystallographic group G arising in Theorem 3.10 has a type corresponding to the signature of the orbifold $(\widehat{\mathbb{C}}, \alpha)$; so for example, if the signature of $(\widehat{\mathbb{C}}, \alpha)$ is $(2, 2, 2, 2)$, then G is of type (2222).

One can use this remark to actually show the existence of the orbifold covering map Θ for a given parabolic orbifold $(\widehat{\mathbb{C}}, \alpha)$ as in Theorem 3.10. Indeed, if the orbifold has three singularities and so one of the signatures $(2, 4, 4)$, $(3, 3, 3)$, or $(2, 3, 6)$, one picks a crystallographic group G of a corresponding type and considers the map Θ induced by G as provided by Proposition 3.9. As follows from (3.7), the map Θ has three critical values. In the fibers over these values the local degree of Θ is constant and corresponds to the different values a, b, or c giving the type (abc) of G. One can postcompose this map by a Möbius transformation so that these three critical values match the three singularities of the orbifold. In this way, one obtains a new holomorphic map $\Theta \colon \mathbb{C} \to \widehat{\mathbb{C}}$ that satisfies (3.10).

If $(\widehat{\mathbb{C}}, \alpha)$ has signature $(2, 2, 2, 2)$, the argument is more complicated. Here one has to choose the crystallographic G of type (2222) carefully so that the map Θ from Proposition 3.9 (which is essentially a Weierstraß $\wp$-function with respect to the underlying lattice Γ of G) has four critical values that can be matched with the four singularities of the orbifold by a Möbius transformation. We will discuss all this in more detail in Section 3.5.

We are now ready to prove the first implication of Theorem 3.1.

PROOF OF (ii) $\Rightarrow$ (iii) IN THEOREM 3.1. Assume $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is a map as in (ii). Then there exists a crystallographic group G, a G-equivariant map $A \colon \mathbb{C} \to \mathbb{C}$ of the form $A(z) = \alpha z + \beta$ where $\alpha, \beta \in \mathbb{C}$, $|\alpha| > 1$, and a holomorphic map $\Theta \colon \mathbb{C} \to \widehat{\mathbb{C}}$ induced by G such that $f \circ \Theta = \Theta \circ A$. Note that then $\mathbb{C}/G$ is homeomorphic to $\widehat{\mathbb{C}}$ and so G is not isomorphic to $\mathbb{Z}^2$.

Let $G_{\mathrm{tr}} \subset G$ be the normal subgroup of translations in G. Consider the quotient space $\mathbb{T} \coloneqq \mathbb{C}/G_{\mathrm{tr}}$ and the quotient map $\pi \colon \mathbb{C} \to \mathbb{T} =$

$\mathbb{C}/G_{\mathrm{tr}}$. Then π is a covering map and $\mathbb{T}$ a topological torus. Moreover, there is a natural conformal structure on $\mathbb{T}$ so that π is holomorphic and $\mathbb{T}$ a complex torus (see Section A.8 for more details).

Let $g \in G_{\mathrm{tr}}$ with $g \neq \mathrm{id}_{\mathbb{C}}$ be arbitrary. Since A is G-equivariant, we have $\widetilde{g} := A \circ g \circ A^{-1} \in G$. So the map $\widetilde{g}$ is an orientation-preserving isometry on $\mathbb{C}$. Moreover, this map has no fixed points, because this is true for its conjugate map g, which is a translation. Hence $\widetilde{g} \in G$ is also a translation, i.e., $\widetilde{g} \in G_{\mathrm{tr}}$. This shows that $A \circ g \circ A^{-1} \in G_{\mathrm{tr}}$ whenever $g \in G_{\mathrm{tr}}$.

We conclude that A is G_{tr}-equivariant and so descends to the quotient $\mathbb{T} = \mathbb{C}/G_{\mathrm{tr}}$ (see Lemma A.22). More explicitly, if we set

$$\overline{A}(\pi(z)) := \pi(A(z))$$

for $z \in \mathbb{C}$, then $\overline{A} \colon \mathbb{T} \to \mathbb{T}$ is a well-defined continuous map satisfying $\overline{A} \circ \pi = \pi \circ A$. Since π and $\pi \circ A$ are holomorphic, the map $\overline{A}$ is holomorphic as well (see Lemma A.15). It follows that $\overline{A}$ is a holomorphic torus endomorphism.

Similarly, the map Θ descends to $\mathbb{T}$. Indeed, suppose $z, w \in \mathbb{C}$ and $\pi(z) = \pi(w)$. Then there exists $g \in G_{\mathrm{tr}}$ such that $w = g(z)$. Since Θ is induced by $G \supset G_{\mathrm{tr}}$, we then have $\Theta(z) = \Theta(w)$. So if we set

$$\overline{\Theta}(\pi(z)) := \Theta(z)$$

for $z \in \mathbb{C}$, then we get a well defined map $\overline{\Theta} \colon \mathbb{T} \to \widehat{\mathbb{C}}$ such that $\overline{\Theta} \circ \pi = \Theta$. This last relation implies that $\overline{\Theta}$ is non-constant and holomorphic, and a branched covering map (see Lemma A.15 (ii)).

Note that

$$f \circ \overline{\Theta} \circ \pi = f \circ \Theta = \Theta \circ A = \overline{\Theta} \circ \pi \circ A = \overline{\Theta} \circ \overline{A} \circ \pi.$$

Since $\pi \colon \mathbb{C} \to \mathbb{T}$ is surjective, it follows that $f \circ \overline{\Theta} = \overline{\Theta} \circ \overline{A}$.

The holomorphic maps considered, and their relations, can be summarized in the following commutative diagram:

(3.12)

$$
\Theta \left(
\begin{array}{ccc}
\mathbb{C} & \xrightarrow{\;A\;} & \mathbb{C} \\
\pi \downarrow & & \downarrow \pi \\
\mathbb{T} & \xrightarrow{\;\overline{A}\;} & \mathbb{T} \\
\overline{\Theta} \downarrow & & \downarrow \overline{\Theta} \\
\widehat{\mathbb{C}} & \xrightarrow{\;f\;} & \widehat{\mathbb{C}}.
\end{array}
\right) \Theta
$$

This shows that f is a quotient of a holomorphic torus endomorphism. If we have maps as in (3.12), then

(3.13)
$$\deg(f) = \deg(\overline{A}) = |\alpha|^2$$

(recall that $A(z) = \alpha z + \beta$). We postpone the justification of this to Lemma 3.16, where we will establish a more general fact (not relying on Theorem 3.1, of course). In particular, $\deg(\overline{A}) = |\alpha|^2 > 1$, and so $\deg(\overline{A}) \geq 2$. It follows that f is indeed a map as in (iii). $\square$

REMARK 3.11. The action of the crystallographic group G descends to the complex torus $\mathbb{T} = \mathbb{C}/G_{\mathrm{tr}}$; more precisely, the group $\overline{G} = G/G_{\mathrm{tr}}$ acts naturally on $\mathbb{T}$. It easily follows from Theorem 3.7 that $\overline{G}$ is a cyclic group. Accordingly, condition (iii) in Theorem 3.1 can be formulated similar to condition (ii) in terms of a cyclic group action on $\mathbb{T}$. This is discussed in [**Mi06a**] in more detail. We do not pursue this point of view here, since the underlying geometry is not as easy to visualize as for crystallographic groups.

PROOF OF (i) $\Rightarrow$ (ii) IN THEOREM 3.1. Let $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a map as in (i), i.e., a rational Thurston map with a parabolic orbifold $\mathcal{O}_f = (\widehat{\mathbb{C}}, \alpha_f)$ and no periodic critical points. By Proposition 2.9 (ii) we then have $\alpha_f(p) < \infty$ for $p \in \widehat{\mathbb{C}}$.

Let $\Theta \colon \mathbb{C} \to \widehat{\mathbb{C}}$ be the (holomorphic) universal orbifold covering map of the orbifold $\mathcal{O}_f$ (see Theorem 3.10) and G be the group of deck transformations of Θ. Then G is a crystallographic group and Θ is induced by G.

We have to find an automorphism $A \colon \mathbb{C} \to \mathbb{C}$ such that $f \circ \Theta = \Theta \circ A$. For this we consider the holomorphic map $f \circ \Theta \colon \mathbb{C} \to \widehat{\mathbb{C}}$. Since $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ and $\Theta \colon \mathbb{C} \to \widehat{\mathbb{C}}$ are branched covering maps, $f \circ \Theta \colon \mathbb{C} \to \widehat{\mathbb{C}}$ is a branched covering map as well (see Lemma A.15 (i)).

Since $\mathcal{O}_f = (\widehat{\mathbb{C}}, \alpha_f)$ is parabolic, by Proposition 2.13 we have

$$\deg(f, p) \cdot \alpha_f(p) = \alpha_f(f(p))$$

for all $p \in \widehat{\mathbb{C}}$. For each $z \in \mathbb{C}$ we have $\deg(\Theta, z) = \alpha_f(\Theta(z))$, and so

$$\deg(f \circ \Theta, z) = \deg(f, \Theta(z)) \cdot \deg(\Theta, z)$$
$$= \deg(f, \Theta(z)) \cdot \alpha_f(\Theta(z)) = \alpha_f\big(f(\Theta(z))\big).$$

This shows that $f \circ \Theta$ is another universal orbifold covering map of $\mathcal{O}_f$. The essential uniqueness of the universal orbifold cover (see Theorem 3.10) implies that there is an automorphism $A \colon \mathbb{C} \to \mathbb{C}$ satisfying $f \circ \Theta = \Theta \circ A$. In other words, we have a commutative diagram as in (3.1). Since Θ is induced by G, it follows that the map A is G-equivariant (see Lemma A.22).

The map $A \in \mathrm{Aut}(\mathbb{C})$ is necessarily of the form $z \mapsto A(z) = \alpha z + \beta$, where $\alpha, \beta \in \mathbb{C}$, $\alpha \neq 0$. As in (3.13), we have $\deg(f) = |\alpha|^2 \geq 2$, and so $|\alpha| > 1$. It follows that f is as in (ii). $\square$

3.2. Quotients of torus endomorphisms and parabolicity

We now prepare the proof of the implication (iii) $\Rightarrow$ (i) in Theorem 3.1. The main difficulty here is to show that a map as in Theorem 3.1 (iii) has a parabolic orbifold. To address this, we will first establish some general statements for quotients of endomorphisms on a torus T^2 (see Definition 3.4).

LEMMA 3.12 (Properties of quotients of torus endomorphisms). *Let $f\colon S^2 \to S^2$ be a quotient of a torus endomorphism, and $\overline{\Theta}\colon T^2 \to S^2$ and $\overline{A}\colon T^2 \to T^2$ with $\deg(\overline{A}) \geq 2$ be as in Definition 3.4. Then the following statements are true:*

   (i) *The map f is a Thurston map without periodic critical points, and $\deg(f) = \deg(\overline{A}) \geq 2$.*

   (ii) *The set $\mathrm{post}(f)$ is equal to the set of critical values of $\overline{\Theta}$, i.e.,*

$$\mathrm{post}(f) = \overline{\Theta}(\mathrm{crit}(\overline{\Theta})).$$

  (iii) *The ramification function of f is given by*

$$\alpha_f(p) = \mathrm{lcm}\{\deg(\overline{\Theta}, x) : x \in \overline{\Theta}^{-1}(p)\}$$

   for $p \in S^2$.

PROOF. Let f, $\overline{\Theta}$, and $\overline{A}$ be as in the statement. Then $\overline{A}$ and $\overline{\Theta}$ are branched covering maps. Since $f \circ \overline{\Theta} = \overline{\Theta} \circ \overline{A}$, it follows from Lemma A.15 (i) and (ii) that f is a branched covering map as well.

Note that

$$\deg(f) \cdot \deg(\overline{\Theta}) = \deg(f \circ \overline{\Theta}) = \deg(\overline{\Theta} \circ \overline{A}) = \deg(\overline{\Theta}) \cdot \deg(\overline{A}),$$

and so

$$\deg(f) = \deg(\overline{A}) \geq 2,$$

as claimed. To show that f is a Thurston map without periodic critical points, we first establish (ii) and (iii).

(ii) Let $V_{\overline{\Theta}} := \overline{\Theta}(\mathrm{crit}(\overline{\Theta}))$ denote the set of critical values of $\overline{\Theta}$. Since T^2 is compact and the set of critical points of $\overline{\Theta}$ has no accumulation point in T^2, there are only finitely many critical points of $\overline{\Theta}$. Hence the set $V_{\overline{\Theta}}$ is also finite. We will first prove that $\mathrm{post}(f) \subset V_{\overline{\Theta}}$.

Let $p \in \mathrm{post}(f)$ be arbitrary. Then by (2.6) the point p is a critical value of some iterate of f. So there exist $n \in \mathbb{N}$ and $q \in S^2$ with $\deg(f^n, q) \geq 2$ and $f^n(q) = p$. As a branched covering map, $\overline{\Theta}$ is surjective, and so we can find $x \in T^2$ with $\overline{\Theta}(x) = q$.

Recall that by the Riemann-Hurwitz formula (2.3) the map $\overline{A}$ cannot have critical points, and hence is locally injective. In particular,

$\deg(\overline{A}^n, x) = 1$. Since $f \circ \overline{\Theta} = \overline{\Theta} \circ \overline{A}$, we have $f^n \circ \overline{\Theta} = \overline{\Theta} \circ \overline{A}^n$. It follows that

$$
\begin{aligned}
\deg(\overline{\Theta}, \overline{A}^n(x)) &= \deg(\overline{\Theta}, \overline{A}^n(x)) \cdot \deg(\overline{A}^n, x) \\
&= \deg(\overline{\Theta} \circ \overline{A}^n, x) = \deg(f^n \circ \overline{\Theta}, x) \\
&= \deg(f^n, q) \cdot \deg(\overline{\Theta}, x) \geq 2.
\end{aligned}
$$

Thus $\overline{A}^n(x)$ is a critical point of $\overline{\Theta}$. So we have

$$
p = f^n(q) = (f^n \circ \overline{\Theta})(x) = (\overline{\Theta} \circ \overline{A}^n)(x) \in V_{\overline{\Theta}}.
$$

The desired inclusion $\mathrm{post}(f) \subset V_{\overline{\Theta}}$ follows.

To show that actually $\mathrm{post}(f) = V_{\overline{\Theta}}$, we argue by contradiction and assume that there exists a point $p \in V_{\overline{\Theta}} \setminus \mathrm{post}(f)$. Then by (2.6) the point p is not a critical value of any iterate f^n of f. Since p is a critical value of $\overline{\Theta}$, the set $\overline{\Theta}^{-1}(p)$ contains a critical point c of $\overline{\Theta}$. This implies that for each $n \in \mathbb{N}$, the set $\overline{A}^{-n}(c)$ consists of critical points of $\overline{\Theta}$. Indeed, if $a \in \overline{A}^{-n}(c)$, then

$$
f^n(\overline{\Theta}(a)) = \overline{\Theta}(\overline{A}^n(a)) = \overline{\Theta}(c) = p,
$$

and so $\deg(f^n, \overline{\Theta}(a)) = 1$; moreover, $\overline{A}^n(a) = c$, and so

$$
\begin{aligned}
\deg(\overline{\Theta}, a) &= \deg(f^n, \overline{\Theta}(a)) \cdot \deg(\overline{\Theta}, a) \\
&= \deg(f^n \circ \overline{\Theta}, a) = \deg(\overline{\Theta} \circ \overline{A}^n, a) \\
&= \deg(\overline{\Theta}, \overline{A}^n(a)) \cdot \deg(\overline{A}^n, a) = \deg(\overline{\Theta}, c) \geq 2.
\end{aligned}
$$

Since $\overline{A}$ is a covering map with $\deg(\overline{A}) \geq 2$, we have

$$
\#\overline{A}^{-n}(c) = \deg(\overline{A})^n \geq 2^n,
$$

and so there must be at least 2^n distinct critical points of $\overline{\Theta}$. Since $n \in \mathbb{N}$ was arbitrary, and the number of critical points of $\overline{\Theta}$ is finite, this is a contradiction showing $\mathrm{post}(f) = V_{\overline{\Theta}}$.

Since $\mathrm{post}(f) = V_{\overline{\Theta}}$ is a finite set, we conclude that f is a Thurston map.

(iii) We define

$$
\nu(p) := \mathrm{lcm}\{\deg(\overline{\Theta}, x) : x \in \overline{\Theta}^{-1}(p)\}
$$

for $p \in S^2$. We claim that the function ν is equal to the ramification function α_f of f. To prove this claim, we will show that ν satisfies the characterization of α_f in Proposition 2.8.

To see this, let $p \in S^2$ be arbitrary, and $x \in \overline{\Theta}^{-1}(p)$. If $y := \overline{A}(x)$, then

$$
\overline{\Theta}(y) = \overline{\Theta}(\overline{A}(x)) = f(\overline{\Theta}(x)) = f(p).
$$

Hence $y \in \overline{\Theta}^{-1}(f(p))$, and

$$\deg(f, p) \cdot \deg(\overline{\Theta}, x) = \deg(f \circ \overline{\Theta}, x) = \deg(\overline{\Theta} \circ \overline{A}, x)$$
$$= \deg(\overline{\Theta}, \overline{A}(x)) \cdot \deg(\overline{A}, x) = \deg(\overline{\Theta}, y).$$

This shows that $\deg(f, p) \cdot \deg(\overline{\Theta}, x)$ divides $\nu(f(p))$. Since this is true for all $x \in \overline{\Theta}^{-1}(p)$ we conclude that $\deg(f, p) \cdot \nu(p)$ divides $\nu(f(p))$. So the function ν satisfies the condition in Proposition 2.8 (i).

Now suppose $\beta \colon S^2 \to \widehat{\mathbb{N}}$ is another function such that $\deg(f, p) \cdot \beta(p)$ divides $\beta(f(p))$ for each $p \in S^2$. Then $\deg(f^n, q) \cdot \beta(q)$ divides $\beta(f^n(q))$ for all $q \in S^2$ and $n \in \mathbb{N}$ (see the remarks after Proposition 2.8).

Let $p \in S^2$ and $x \in \overline{\Theta}^{-1}(p)$ be arbitrary. Then $\# \overline{A}^{-n}(x) = \deg(\overline{A})^n \geq 2^n$. Since there are only finitely many critical points of $\overline{\Theta}$, there exist $n \in \mathbb{N}$ and $y \in \overline{A}^{-n}(x)$ such that $y \notin \operatorname{crit}(\overline{\Theta})$. Let $q := \overline{\Theta}(y)$. Then

$$f^n(q) = f^n(\overline{\Theta}(y)) = \overline{\Theta}(\overline{A}^n(y)) = \overline{\Theta}(x) = p,$$

and

$$\deg(\overline{\Theta}, x) = \deg(\overline{\Theta}, x) \cdot \deg(\overline{A}^n, y) = \deg(\overline{\Theta} \circ \overline{A}^n, y)$$
$$= \deg(f^n \circ \overline{\Theta}, y) = \deg(f^n, q) \cdot \deg(\overline{\Theta}, y)$$
$$= \deg(f^n, q).$$

Clearly, $\deg(\overline{\Theta}, x) = \deg(f^n, q)$ divides $\deg(f^n, q) \cdot \beta(q)$, which in turn divides $\beta(p) = \beta(f^n(q))$ by the remark above. Hence $\deg(\overline{\Theta}, x) | \beta(p)$ for all $x \in \overline{\Theta}^{-1}(p)$. By definition of ν this implies that $\nu(p) | \beta(p)$ for $p \in S^2$. This means that ν satisfies condition (ii) in Proposition 2.8.

From the uniqueness property of α_f given by Proposition 2.8 we conclude $\nu = \alpha_f$ as desired.

(i) We have already seen that f is a Thurston map with $\deg(f) = \deg(\overline{A})$. From (iii) it follows that $\alpha_f(p) < \infty$ for all $p \in S^2$. Thus f has no periodic critical points (see Proposition 2.9 (ii)). $\square$

LEMMA 3.13 (Criterion for orbifold parabolicity). *Let $f \colon S^2 \to S^2$ be a quotient of a torus endomorphism and $\overline{\Theta} \colon T^2 \to S^2$ be as in Definition 3.4. Then f has a parabolic orbifold if and only if*

$$(3.14) \qquad\qquad \deg(\overline{\Theta}, x) = \deg(\overline{\Theta}, y)$$

for all $x, y \in T^2$ with $\overline{\Theta}(x) = \overline{\Theta}(y)$.

We do not know whether condition (3.14) is always true, or equivalently, whether every quotient of a torus endomorphism has a parabolic

orbifold. One can show this under the additional assumption that the
map is expanding. The proof is rather involved, and so we will not
discuss it.

PROOF. As in Definition 3.4, let $\overline{A}\colon T^2 \to T^2$ be a torus endomor-
phism with $\deg(\overline{A}) \geq 2$ for our given maps f and $\overline{\Theta}$.

Suppose first that $\deg(\overline{\Theta}, x) = \deg(\overline{\Theta}, y)$, whenever $x, y \in T^2$ and
$\Theta(x) = \Theta(y)$. This is equivalent to the statement that the local degree
of $\overline{\Theta}$ is constant in each fiber $\overline{\Theta}^{-1}(p)$, $p \in S^2$. Then Lemma 3.12 (iii)
implies that $\alpha_f(p) = \deg(\overline{\Theta}, x)$ whenever $x \in \overline{\Theta}^{-1}(p)$. If $y \coloneqq \overline{A}(x)$,
then $\overline{\Theta}(y) = f(p)$, and so

$$
\begin{aligned}
\alpha_f(f(p)) = \deg(\overline{\Theta}, y) &= \deg(\overline{\Theta}, y) \cdot \deg(\overline{A}, x) \\
&= \deg(\overline{\Theta} \circ \overline{A}, x) = \deg(f \circ \overline{\Theta}, x) \\
&= \deg(f, p) \cdot \deg(\overline{\Theta}, x) = \deg(f, p) \cdot \alpha_f(p)
\end{aligned}
$$

for $p \in S^2$. Hence $\mathcal{O}_f = (S^2, \alpha_f)$ is parabolic by the condition in
Proposition 2.13 (iii).

Conversely, suppose that f has a parabolic orbifold. We claim
that the local degree of $\overline{\Theta}$ is constant in each fiber over a point in
S^2. For this it suffices to show that if $p \in S^2$ and $x \in \overline{\Theta}^{-1}(p)$, then
$\deg(\overline{\Theta}, x) = \alpha_f(p)$.

Note that the set $\overline{\Theta}^{-1}(\mathrm{post}(f))$ is finite. So by picking $n \in \mathbb{N}$ large
enough, we can find a point $y \in \overline{A}^{-n}(x)$ with $q \coloneqq \overline{\Theta}(y) \notin \mathrm{post}(f)$.
Then $\alpha_f(q) = 1$ and $\deg(\overline{\Theta}, y) = 1$ by Lemma 3.12 (ii). We also have

$$
f^n(q) = f^n(\overline{\Theta}(y)) = \overline{\Theta}(\overline{A}^n(y)) = \overline{\Theta}(x) = p,
$$

and

$$
\begin{aligned}
\deg(\overline{\Theta}, x) &= \deg(\overline{\Theta}, x) \cdot \deg(\overline{A}^n, y) \\
&= \deg(\overline{\Theta} \circ \overline{A}^n, y) = \deg(f^n \circ \overline{\Theta}, y) \\
&= \deg(f^n, q) \cdot \deg(\overline{\Theta}, y) = \deg(f^n, q).
\end{aligned}
$$

The parabolicity of $\mathcal{O}_f$ implies that

$$
\alpha_f(p) = \alpha_f(f^n(q)) = \alpha_f(q) \cdot \deg(f^n, q) = \deg(f^n, q).
$$

We conclude that

$$
\deg(\overline{\Theta}, x) = \deg(f^n, q) = \alpha_f(p)
$$

as desired.   □

To complete the proof of Theorem 3.1, and to establish the remaining implication (iii) $\Rightarrow$ (i), let $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be given as in (3.3), where the corresponding maps $\overline{A}\colon \mathbb{T} \to \mathbb{T}$ and $\overline{\Theta}\colon \mathbb{T} \to \widehat{\mathbb{C}}$ are holomorphic and defined on a complex torus $\mathbb{T}$. This implies that f is a holomorphic map as well and hence a rational map on $\widehat{\mathbb{C}}$ (see Lemma A.15).

The universal cover of $\mathbb{T}$ (as a Riemann surface) is $\mathbb{C}$ and so there exists a holomorphic covering map $\pi\colon \mathbb{C} \to \mathbb{T}$. Actually, we can identify $\mathbb{T}$ with a quotient space $\mathbb{C}/\Gamma$, where Γ is a suitable rank-2 lattice. Under such an identification $\mathbb{T} \cong \mathbb{C}/\Gamma$, the map $\pi\colon \mathbb{C} \to \mathbb{C}/\Gamma \cong \mathbb{T}$ is just the quotient map.

The map $\overline{A}$ can be lifted by π to a holomorphic homeomorphism $A\colon \mathbb{C} \to \mathbb{C}$ such that $\overline{A} \circ \pi = \pi \circ A$ (see Section A.8 and in particular Lemma A.23 (ii)). The map A has to be of the form $A(z) = \alpha z + \beta$ with $\alpha, \beta \in \mathbb{C}$, $\alpha \neq 0$. Then we have the following commutative diagram:

(3.15)

$$
\begin{array}{ccc}
\mathbb{C} & \xrightarrow{\ A\ } & \mathbb{C} \\
\downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \pi} \\
\mathbb{T} & \xrightarrow{\ \overline{A}\ } & \mathbb{T} \\
\downarrow{\scriptstyle \overline{\Theta}} & & \downarrow{\scriptstyle \overline{\Theta}} \\
\widehat{\mathbb{C}} & \xrightarrow{\ f\ } & \widehat{\mathbb{C}}.
\end{array}
$$

We are now ready to show the implication (iii) $\Rightarrow$ (i) of Theorem 3.1. As we will see, the proof strongly relies on the fact that the involved maps are holomorphic.

PROOF OF (iii) $\Rightarrow$ (i) IN THEOREM 3.1. Suppose $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is a map as in statement (iii) of Theorem 3.1. Then there is a complex torus $\mathbb{T}$, a holomorphic torus endomorphism $\overline{A}\colon \mathbb{T} \to \mathbb{T}$, and a holomorphic map $\overline{\Theta}\colon \mathbb{T} \to \widehat{\mathbb{C}}$ such that $f \circ \overline{\Theta} = \overline{\Theta} \circ \overline{A}$. As we discussed, f is also a holomorphic map and hence a rational map on $\widehat{\mathbb{C}}$. Moreover, f is a quotient of a torus endomorphism and so by Lemma 3.12 (i) a Thurston map without periodic critical points.

It remains to show that the orbifold of f is parabolic. By Lemma 3.13 it is enough to prove that the local degree of $\overline{\Theta}$ is constant in each fiber $\overline{\Theta}^{-1}(p)$, $p \in \widehat{\mathbb{C}}$. We argue by contradiction and assume that there exist $p \in \widehat{\mathbb{C}}$, and $x, y \in \overline{\Theta}^{-1}(p)$ with

(3.16) $$\deg(\overline{\Theta}, x) \neq \deg(\overline{\Theta}, y).$$

In particular, one of these degrees must be ≥ 2; so p is a critical value of $\overline{\Theta}$ and hence belongs to $\mathrm{post}(f)$ by Lemma 3.12 (ii).

For all $n \in \mathbb{N}$ we have $\overline{A}^n(x), \overline{A}^n(y) \in \overline{\Theta}^{-1}(f^n(p))$, and, since $\overline{A}$ does not have critical points,

$$\begin{aligned}
\deg(\overline{\Theta}, \overline{A}^n(x)) &= \deg(\overline{\Theta}, \overline{A}^n(x)) \cdot \deg(\overline{A}^n, x) = \deg(\overline{\Theta} \circ \overline{A}^n, x) \\
&= \deg(f^n \circ \overline{\Theta}, x) = \deg(f^n, p) \cdot \deg(\overline{\Theta}, x) \\
&\neq \deg(f^n, p) \cdot \deg(\overline{\Theta}, y) = \deg(\overline{\Theta}, \overline{A}^n(y)).
\end{aligned}$$

In other words, the iterates $\overline{A}^n(x)$ and $\overline{A}^n(y)$ lie in the fiber over the point $f^n(p) \in \mathrm{post}(f)$ and $\overline{\Theta}$ has different local degrees at these points. Since there are only finitely many points in $\mathrm{post}(f)$, and each fiber $\overline{\Theta}^{-1}(p)$ contains only finitely many points, the points x and y must be preperiodic under iteration of $\overline{A}$, and p must be preperiodic under iteration of f. This implies that in (3.16) we may in addition assume that x and y are periodic points for $\overline{A}$, and that p is a periodic point for f. Moreover, by replacing the maps $\overline{A}$ and f by suitable iterates, we are further reduced to the case where x and y are fixed points of $\overline{A}$, and p is a fixed point of f.

If we introduce a suitable holomorphic coordinate w near p such that p corresponds to $w = 0$, then f has a local power series representation of the form

$$f(w) = \lambda w^d + \dots,$$

where $\lambda \neq 0$, and $d \in \mathbb{N}$. Since f has no periodic critical points by Lemma 3.12 (i), we actually have $d = 1$, and so

$$f(w) = \lambda w + \dots$$

As we discussed before the proof, the map $\overline{A}$ lifts to a map $A \colon \mathbb{C} \to \mathbb{C}$ of the form $A(z) = \alpha z + \beta$ for $z \in \mathbb{C}$, where $\alpha, \beta \in \mathbb{C}$, $\alpha \neq 0$, such that we have a commutative diagram as in (3.15). This implies that by introducing a suitable holomorphic coordinate u near x such that $u = 0$ corresponds to x, the maps $\overline{A}$ and $\overline{\Theta}$ can be given the forms $\overline{A}(u) = \alpha u$ and

$$w = \overline{\Theta}(u) = bu^k + \dots$$

near $u = 0$, where $b \neq 0$ and $k = \deg(\overline{\Theta}, x)$. Similarly, by using a suitable holomorphic coordinate v near y such that $v = 0$ corresponds to y, we can write $\overline{A}(v) = \alpha v$ and

$$w = \overline{\Theta}(v) = cv^n + \dots$$

near $v = 0$, where $c \neq 0$ and $n = \deg(\overline{\Theta}, y)$. Since $f \circ \overline{\Theta} = \overline{\Theta} \circ \overline{A}$ near $u = 0$, we obtain

$$f(\overline{\Theta}(u)) = \lambda bu^k + \dots = \overline{\Theta}(\overline{A}(u)) = b\alpha^k u^k + \dots$$

In particular, $\lambda = \alpha^k$. Similarly, by considering the relation $f(\overline{\Theta}(v)) = \overline{\Theta}(\overline{A}(v))$ near $v = 0$, we obtain $\lambda = \alpha^n$. We conclude that $\alpha^k = \lambda = \alpha^n$. Now $2 \leq \deg(f) = \deg(\overline{A}) = |\alpha|^2$ by (3.13), and so $|\alpha| > 1$; but then $\alpha^k = \alpha^n$ implies $k = n$. This contradicts our assumption that $k = \deg(\overline{\Theta}, x) \neq \deg(\overline{\Theta}, y) = n$. $\qquad\square$

This finishes the proof of Theorem 3.1.

3.3. Classifying Lattès maps

Theorem 3.1 allows us to explicitly construct each Lattès map as a quotient of a holomorphic automorphism $A \colon \mathbb{C} \to \mathbb{C}$ by a crystallographic group G. For such a map A to pass to the quotient $\mathbb{C}/G$ it has to be G-equivariant. In this section we study this condition on A and some related questions in more detail. Our results essentially provide a complete classification of all Lattès maps.

Let G be a crystallographic group not isomorphic to $\mathbb{Z}^2$, and let $\Theta \colon \mathbb{C} \to \widehat{\mathbb{C}}$ be a holomorphic map induced by G as provided by Proposition 3.9. Let $A \colon \mathbb{C} \to \mathbb{C}$ be a map of the form $A(z) = \alpha z + \beta$, where $\alpha, \beta \in \mathbb{C}$ with $|\alpha| > 1$. Then by Lemma A.22 there is a (unique) map $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ that satisfies $f \circ \Theta = \Theta \circ A$ if and only if A is G-equivariant. In this case, f is a Lattès map by condition (ii) in Theorem 3.1.

Recall from Proposition 3.9 that for a given crystallographic group G, the map Θ is unique up to postcomposition by a Möbius transformation. So suppose $\widetilde{\Theta} = \varphi \circ \Theta$ is another (holomorphic) map induced by G, where $\varphi \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is a Möbius transformation. Then if $f \circ \Theta = \Theta \circ A$ it is immediate to check that $\widetilde{f} := \varphi \circ f \circ \varphi^{-1}$ is the unique map that satisfies $\widetilde{f} \circ \widetilde{\Theta} = \widetilde{\Theta} \circ A$. Thus the Lattès map induced by a map A that is equivariant for a given crystallographic group G is unique up to conjugation by a Möbius transformation.

For a given Lattès map $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$, the map $A \colon \mathbb{C} \to \mathbb{C}$ and the group G provided by Theorem 3.1 (ii) are not unique. Indeed, we can conjugate G and A by an arbitrary map $h \in \mathrm{Aut}(\mathbb{C})$. Then $\widetilde{G} = \{h^{-1} \circ g \circ h : g \in G\}$ is also a crystallographic group. If $\widetilde{A} = h^{-1} \circ A \circ h$ and $\widetilde{\Theta} = \Theta \circ h$, then we obtain the same map f in Theorem 3.1 (ii), if we replace G, Θ, A by $\widetilde{G}, \widetilde{\Theta}, \widetilde{A}$, respectively. The situation is illustrated

in the following commutative diagram:

(3.17)

$$
\widetilde{\Theta}\left(
\begin{array}{ccc}
\mathbb{C} & \xrightarrow{\widetilde{A}} & \mathbb{C} \\
h\downarrow & & \downarrow h \\
\mathbb{C} & \xrightarrow{A} & \mathbb{C} \\
\Theta\downarrow & & \downarrow \Theta \\
\widehat{\mathbb{C}} & \xrightarrow{f} & \widehat{\mathbb{C}}.
\end{array}
\right)\widetilde{\Theta}
$$

By using such a conjugation, in Theorem 3.1 (ii) we can always assume that the group G is one of the groups $\widetilde{G}$ listed in Theorem 3.7.

The requirement that the map $A(z) = \alpha z + \beta$ is $\widetilde{G}$-equivariant can then be explicitly analyzed and puts strong restrictions on α and β as the following proposition shows.

PROPOSITION 3.14. *Let $G = \widetilde{G}$ be a crystallographic group as in Theorem 3.7 not isomorphic to $\mathbb{Z}^2$. Let $\Gamma = \mathbb{Z} \oplus \mathbb{Z}\tau$ be the underlying lattice, where $\tau \in \mathbb{C}$ with $\mathrm{Im}(\tau) > 0$ in the case (2222), $\tau = \boldsymbol{i}$ in the case (244), and $\tau = \omega = e^{\pi i/3}$ in the cases (333) and (236). Then $A\colon \mathbb{C} \to \mathbb{C}$ given by $A(z) = \alpha z + \beta$ (where $\alpha, \beta \in \mathbb{C}$, $\alpha \neq 0$) is G-equivariant if and only if*

$$
\left.
\begin{array}{l}
\alpha, \alpha\tau, 2\beta \\
\quad \alpha, (1+\boldsymbol{i})\beta \\
\quad \alpha, (1+\omega)\beta \\
\quad \alpha, \beta
\end{array}
\right\}
\textit{ are elements of } \Gamma \textit{ when } G \textit{ is of type }
\left\{
\begin{array}{l}
(2222), \\
(244), \\
(333), \\
(236).
\end{array}
\right.
$$

As we will see in the proof, the condition on α is equivalent to the requirement that $\alpha\Gamma \subset \Gamma$. According to this proposition, we can always choose $\alpha \in \mathbb{Z} \setminus \{0\}$ and $\beta = 0$ for any lattice Γ. This and Theorem 3.1 (ii) imply that Lattès maps exist for all signatures $(2, 2, 2, 2)$, $(2, 4, 4)$, $(3, 3, 3)$, and $(2, 3, 6)$. We will discuss more explicit examples for each of these signatures later in Section 3.5.

PROOF. Recall from (3.2) that a map $z \mapsto A(z) = \alpha z + \beta$ as in the statement is G-equivariant if and only if $A \circ g \circ A^{-1} \in G$ for all $g \in G$.

The maps $g \in G$ have the form $g(z) = \lambda z + \gamma$, where $\gamma \in \Gamma$ and λ is a root of unity depending on the type of G. An elementary computation shows that

$$
(A \circ g \circ A^{-1})(z) = \lambda z + \alpha\gamma + (1 - \lambda)\beta.
$$

Using this first for $\gamma = 0 \in \Gamma$, we see that (3.2) can only be valid if

(3.18) $(1 - \lambda)\beta \in \Gamma$,

for the appropriate roots of unity λ; in addition, it is necessary that

$$(3.19) \qquad\qquad \alpha\gamma \in \Gamma \text{ for all } \gamma \in \Gamma.$$

Conversely, if (3.18) and (3.19) are true, A satisfies (3.2). Thus A is G-equivariant if and only if α and β satisfy (3.18) and (3.19).

Note that (3.19) is equivalent to the condition that $\alpha\Gamma \subset \Gamma$. Since 1 and τ generate the lattice Γ, this in turn is the same as the requirement that

$$(3.20) \qquad\qquad \alpha \in \Gamma \text{ and } \alpha\tau \in \Gamma.$$

If G is of type (244), (333), or (236), then we can omit the second condition here. Indeed, in these cases $\tau\Gamma = \Gamma$, and so $\alpha \in \Gamma$ if and only if $\alpha\tau \in \Gamma$.

This discussion shows that A is G-equivariant if and only if α satisfies the conditions as in the statement of the proposition, and β satisfies (3.18). To analyze this latter condition further, we consider several cases depending on the type of G.

If G is of type (2222), then we have $\lambda = \pm 1$. Thus (3.18) is true if and only if $2\beta \in \Gamma$.

If G is of type (244), then we have $\Gamma = \mathbb{Z} \oplus \mathbb{Z}\boldsymbol{i}$ and $\lambda = 1, \boldsymbol{i}, -1, -\boldsymbol{i}$. Thus (3.18) implies that $(1 + \boldsymbol{i})\beta \in \Gamma$. Conversely, suppose this last condition is true. Note that $-\boldsymbol{i}\Gamma = \Gamma$ and $(1-\boldsymbol{i})\Gamma \subset \Gamma$. So we conclude that $-\boldsymbol{i}(1+\boldsymbol{i})\beta = (1-\boldsymbol{i})\beta \in \Gamma$ and $(1-\boldsymbol{i})(1+\boldsymbol{i})\beta = 2\beta \in \Gamma$. Therefore, (3.18) holds for $\lambda = 1, \boldsymbol{i}, -1, -\boldsymbol{i}$.

If G is of type (333), then we have $\Gamma = \mathbb{Z} \oplus \mathbb{Z}\omega$ and $\lambda = 1, \omega^2, \omega^4$ (recall that $\omega = e^{\pi \boldsymbol{i}/3}$). So (3.18) implies that $(1-\omega^4)\beta = (1+\omega)\beta \in \Gamma$. Conversely, if this last condition is true, then using $\omega\Gamma = \Gamma$ we obtain $\omega^5(1+\omega)\beta = (1-\omega^2)\beta \in \Gamma$; so (3.18) holds for $\lambda = \omega^j$ where $j = 0, 2, 4$.

Finally, if G is of type (236), then $\Gamma = \mathbb{Z} \oplus \mathbb{Z}\omega$ and $\lambda = \omega^j$ where $j = 0, \ldots, 5$. Thus (3.18) implies $(1 - \omega^5)\beta = (1+\omega^2)\beta \in \Gamma$. Note that $1 + \omega^2 = \omega$, and so we obtain $\omega\beta \in \Gamma$. Since $\omega\Gamma = \Gamma$, this shows that $\beta \in \Gamma$.

Conversely, assume that $\beta \in \Gamma$. Using $\omega\Gamma = \Gamma$ again, we conclude that $\omega^j\beta \in \Gamma$ for $j = 0, \ldots, 5$. This implies $\beta - \omega^j\beta = (1 - \omega^j)\beta \in \Gamma$. Thus (3.18) is satisfied for all $\lambda = \omega^j$ with $j = 0, \ldots, 5$.

The claim follows. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \square$

Suppose a Lattès map $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is given as in Theorem 3.1 (ii). Then $f \circ \Theta = \Theta \circ A$, where $\Theta \colon \mathbb{C} \to \widehat{\mathbb{C}}$ is induced by a crystallographic group G (not isomorphic to $\mathbb{Z}^2$) and $A(z) = \alpha z + \beta$ (with $\alpha, \beta \in \mathbb{C}$ and $|\alpha| > 1$) is G-equivariant. As we discussed, here we can always assume that $G = \widetilde{G}$ is as in Theorem 3.7.

One can make another reduction. Namely, we can replace A with any map $\widetilde{A} = g \circ A$, where $g \in G$. Indeed, we then have $\Theta \circ \widetilde{A} = \Theta \circ A = f \circ \Theta$, because Θ is induced by G and so $\Theta \circ g = \Theta$. In particular, $\widetilde{A}$ is also G-equivariant and induces the same Lattès map f.

One can use this remark to substantially restrict the values of the coefficient β of the map A.

PROPOSITION 3.15. *Let $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a Lattès map as above obtained from a map $A \colon \mathbb{C} \to \mathbb{C}$ given by $A(z) = \alpha z + \beta$, a crystallographic group $G = \widetilde{G}$ as in Theorem 3.7, and a holomorphic map $\Theta \colon \mathbb{C} \to \widehat{\mathbb{C}}$ induced by G. Then by postcomposing A with a suitable translation $g \in G_{\mathrm{tr}} \subset G$ we can always assume that β has one of the following forms:*

$$
\left.
\begin{array}{l}
\beta \in \{0, \tfrac{1}{2}, \tfrac{1}{2}\tau, \tfrac{1}{2}(1+\tau)\} \\
\beta \in \{0, \tfrac{1}{2}(1+\boldsymbol{i})\} \\
\beta \in \{0, \tfrac{1}{3} + \tfrac{1}{3}\omega, \tfrac{2}{3} + \tfrac{2}{3}\omega\} \\
\beta = 0
\end{array}
\right\}
\quad \text{when } G \text{ is of type}
\left\{
\begin{array}{l}
(2222), \\
(244), \\
(333), \\
(236).
\end{array}
\right.
$$

PROOF. As before, we denote by Γ the underlying lattice of G. Let $\sim$ be the equivalence relation on $\mathbb{C}$ defined by $z \sim w$ if and only if $w - z \in \Gamma$ for $z, w \in \mathbb{C}$. This is the equivalence relation induced by the action of the subgroup of all translations $G_{\mathrm{tr}} \subset G$. As we have seen, if $g(z) = z + \gamma$ with $\gamma \in \Gamma$, then $g \in G_{\mathrm{tr}}$ and we can replace the G-equivariant map $A(z) = \alpha z + \beta$ with

$$
\widetilde{A}(z) = (g \circ A)(z) = \alpha z + (\beta + \gamma).
$$

This means that we can change β to any element β' with $\beta' \sim \beta$ without affecting the Lattès map f. We now analyze this in combination with the condition on β in Proposition 3.14 for the different types of G.

If G is of type (2222), then $2\beta \in \Gamma$ by Proposition 3.14 or equivalently, $\beta \in \tfrac{1}{2}\Gamma$. As we can replace β by any β' with $\beta \sim \beta'$, we may assume that

$$
(3.21) \qquad \beta = \tfrac{1}{2}(k + \ell\tau), \quad \text{where } k, \ell \in \{0, 1\}.
$$

So β has the desired form.

If G is of type (244), then by Proposition 3.14 the relevant condition is $(1 + \boldsymbol{i})\beta \in \Gamma$. This implies that

$$
\beta \in (1 + \boldsymbol{i})^{-1}\Gamma = \tfrac{1}{2}(1 - \boldsymbol{i})\Gamma \subset \tfrac{1}{2}\Gamma.
$$

So again we may assume that β is as in (3.21) with $\tau = \boldsymbol{i}$. For such β we have

$$
(1 + \boldsymbol{i})\beta = \tfrac{1}{2}((k - \ell) + (k + \ell)\boldsymbol{i}) \in \Gamma
$$

precisely if $k = \ell \in \{0, 1\}$. The statement follows in this case.

If G is of type (236), then by Proposition 3.14 the condition on β is $\beta \in \Gamma$, or equivalently $\beta \sim 0$. This means that we can always take $\beta = 0$ in this case.

Finally, if G is of type (333), then by Proposition 3.14 the relevant condition is $(1 + \omega)\beta \in \Gamma$, where $\omega = e^{\pi i/3}$ and $\Gamma = \mathbb{Z} \oplus \mathbb{Z}\omega$. Since $\omega\Gamma = \Gamma$ and $(1 + \omega)(1 + \omega^5) = 3$, this implies

$$\beta \in (1 + \omega)^{-1}\Gamma = \tfrac{1}{3}(1 + \omega^5)\Gamma \subset \tfrac{1}{3}\Gamma.$$

Hence β can be assumed to be of the form

$$\beta = \tfrac{1}{3}(k + \ell\omega), \quad \text{where } k, \ell \in \{0, 1, 2\}.$$

If we use $\omega^2 = \omega - 1$, we see that for such β we have

$$(1 + \omega)\beta = \tfrac{1}{3}((k - \ell) + (k + 2\ell)\omega) \in \Gamma$$

precisely if $k = \ell \in \{0, 1, 2\}$. Again β can be given the desired form. $\square$

3.4. Lattès-type maps

We now consider Lattès-type maps $f\colon S^2 \to S^2$ as in Definition 3.3. If f is such a map, then there exists a crystallographic group G acting on $\mathbb{R}^2 \cong \mathbb{C}$, a G-equivariant affine map $A\colon \mathbb{R}^2 \to \mathbb{R}^2$, and a branched covering map $\Theta\colon \mathbb{R}^2 \to S^2$ induced by G such that $f \circ A = \Theta \circ A$. Since Θ is induced by G, the quotient space $\mathbb{R}^2/G$ is homeomorphic to S^2 which implies that G is not isomorphic to $\mathbb{Z}^2$.

In explicit constructions of Lattès-type maps one usually turns this around and starts with a crystallographic group G not isomorphic to $\mathbb{Z}^2$ and an G-equivariant affine map $A\colon \mathbb{R}^2 \to \mathbb{R}^2$. Then $S^2 = \mathbb{R}^2/G$ is a topological 2-sphere and the quotient map $\Theta\colon \mathbb{R}^2 \to \mathbb{R}^2/G \cong S^2$ is a branched covering map induced by G. The G-equivariance of A ensures that this map descends to the quotient $\mathbb{R}^2/G \cong S^2$ and so there exists a map $f\colon S^2 \to S^2$ such that $f \circ A = \Theta \circ A$. As the considerations below will show, the additional condition that the linear part L_A of A (see (3.4)) satisfies $\det(L_A) > 1$ ensures that f is a Thurston map.

Indeed, let G_{tr} be the subgroup of translations in G. We know that then $T^2 = \mathbb{R}^2/G_{\mathrm{tr}}$ is a (topological) 2-torus. We denote by $\pi\colon \mathbb{R}^2 \to T^2 = \mathbb{R}^2/G_{\mathrm{tr}}$ the quotient map.

The argument in the proof of the implication (ii) $\Rightarrow$ (iii) in Theorem 3.1 (see Section 3.1) shows that A and Θ descend to maps $\overline{A}$ and $\overline{\Theta}$ on T^2. In this proof the maps were assumed to be holomorphic, but this played no role in the existence proof for $\overline{A}$ and $\overline{\Theta}$. So we obtain maps $\overline{A}\colon T^2 \to T^2$ and $\overline{\Theta}\colon T^2 \to S^2$ such that $\overline{A} \circ \pi = \pi \circ A$ and $\Theta = \overline{\Theta} \circ \pi$. These last relations imply that $\overline{A}$ and $\overline{\Theta}$ are branched

covering maps (see Lemma A.15 (ii)). As we discussed, this implies that $\overline{A}$ is a (topological) torus endomorphism.

Similar to (3.12), one can summarize the relations of these map in the following commutative diagram:

$$(3.22)$$

$$
\Theta\left(
\begin{array}{ccc}
\mathbb{R}^2 & \xrightarrow{\ A\ } & \mathbb{R}^2 \\
\pi\downarrow & & \downarrow\pi \\
T^2 & \xrightarrow{\ \overline{A}\ } & T^2 \\
\overline{\Theta}\downarrow & & \downarrow\overline{\Theta} \\
S^2 & \xrightarrow{\ f\ } & S^2.
\end{array}
\right)\Theta
$$

Since $\overline{\Theta}$ and $\overline{\Theta}\circ\overline{A}$ are branched covering maps, and $f\circ\overline{\Theta} = \overline{\Theta}\circ\overline{A}$, it follows from Lemma A.15 (ii) that f is also a branched covering map.

It is easy to see that $\deg(f) = \deg(\overline{A})$ (see the beginning of the proof of Lemma 3.12), but the degree of f can also be computed from A.

LEMMA 3.16. *Let $f\colon S^2 \to S^2$ be a Lattès-type map, $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ be an affine map as in Definition 3.3, L_A be the linear part of A, and $\overline{A}$ be a torus endomorphism as in (3.22). Then $\deg(f) = \deg(\overline{A}) = \det(L_A)$.*

In particular, if $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is a Lattès map and $A(z) = \alpha z + \beta$ is as in Theorem 3.1 (ii), then $\deg(f) = |\alpha|^2$.

PROOF. Let $f\colon S^2 \to S^2$ be a Lattès-type map, and suppose G is a crystallographic group and A an affine map as in Definition 3.3. Then we have a commutative diagram as in (3.22) and we know that $\deg(f) = \deg(\overline{A})$. So we have to verify that $\deg(\overline{A}) = \det(L_A)$. Essentially, this follows from standard facts about degrees of torus endomorphisms as discussed in more detail in Section A.8.

Indeed, let $\Gamma \subset \mathbb{R}^2 \cong \mathbb{C}$ be the underlying lattice of G. Then G_{tr} consists of all translations of the form $u \in \mathbb{R}^2 \mapsto \tau_\gamma(u) := u + \gamma$, where $\gamma \in \Gamma$. Accordingly, we can identify the torus $T^2 = \mathbb{R}^2/G_{\mathrm{tr}}$ with the quotient $\mathbb{R}^2/\Gamma$ and can think of the lattice Γ as representing the fundamental group of T^2 (see the discussion after Lemma A.23).

Now an elementary computation shows that

$$A \circ \tau_\gamma \circ A^{-1} = \tau_{L_A(\gamma)}$$

for each $\gamma \in \mathbb{R}^2$, and in particular for each $\gamma \in \Gamma$. Since A is a lift of $\overline{A}$ to $\mathbb{R}^2$, it follows that L_A is the unique map induced by $\overline{A}$ on the fundamental group Γ of T^2 (see Lemma A.23 (iii)). Now Lemma A.23 (iv) implies $\deg(\overline{A}) = \det(L_A)$ as desired.

If f is a Lattès map and $A(z) = \alpha z + \beta$ as in Theorem 3.1 (ii), then in complex notation $L_A(z) = \alpha z$ for $z \in \mathbb{C}$. For the determinant of L_A considered as a $\mathbb{R}$-linear map, we have $\det(L_A) = |\alpha|^2$. So it follows from the first part of the proof that $\deg(f) = \det(L_A) = |\alpha|^2$ as claimed. $\qquad\square$

PROOF OF PROPOSITION 3.5. Let $f\colon S^2 \to S^2$ be a Lattès-type map, and G, A, and Θ be as in Definition 3.3. Then we have a diagram as in (3.22). Here $\det(L_A) > 1$ by assumption which by Lemma 3.16 translates into $\deg(f) = \deg(\overline{A}) = \det(L_A) \geq 2$. We conclude that f is a quotient of a torus endomorphism (see Definition 3.4).

So we can apply Lemma 3.12 and it follows that f is a Thurston map without periodic critical points. It remains to show that f has a parabolic orbifold.

For this we verify the criterion in Lemma 3.13 with the branched covering map $\overline{\Theta}\colon T^2 \to S^2$ as provided by (3.22). So suppose $x, y \in T^2$ and $\overline{\Theta}(x) = \overline{\Theta}(y)$. Since $\pi\colon \mathbb{R}^2 \to T^2$ is surjective, there exist $u, v \in \mathbb{R}^2$ with $\pi(u) = x$ and $\pi(v) = y$. Then

$$\Theta(u) = (\overline{\Theta} \circ \pi)(u) = \overline{\Theta}(x) = \overline{\Theta}(y) = (\overline{\Theta} \circ \pi)(v) = \Theta(v).$$

Since Θ is induced by the crystallographic group G, there exists $g \in G$ with $v = g(u)$. Now $\Theta = \Theta \circ g$ and

$$\deg(\pi, u) = \deg(\pi, v) = \deg(g, u) = 1.$$

We conclude that

$$\begin{aligned}
\deg(\overline{\Theta}, y) &= \deg(\overline{\Theta}, y) \cdot \deg(\pi, v) = \deg(\Theta, v) \\
&= \deg(\Theta, v) \cdot \deg(g, u) = \deg(\Theta, u) \\
&= \deg(\overline{\Theta}, x) \cdot \deg(\pi, u) = \deg(\overline{\Theta}, x)
\end{aligned}$$

as desired. $\qquad\square$

REMARK 3.17. The considerations in the previous proof actually allow us to determine the signature of the orbifold $\mathcal{O}_f = (S^2, \alpha_f)$ associated with a Lattès-type map $f\colon S^2 \to S^2$. Indeed, if $p \in S^2$, then the argument given shows that the degree of $\overline{\Theta}$ in each point of the fiber $\overline{\Theta}^{-1}(p)$ is the same and is equal to $\deg(\Theta, u)$ for any point $u \in \Theta^{-1}(p)$. If we combine this with (3.11) and Lemma 3.12 (iii), we see that

$$\alpha_f(p) = \deg(\Theta, u) = \#G_u$$

for each $u \in \Theta^{-1}(p)$. As we discussed, the type of a crystallographic group G is determined by the orders of the point stabilizers $\#G_u$,

$u \in \mathbb{R}^2$. So the signature of $\mathcal{O}_f$ corresponds to the type of the crystallographic group G. For example, if G is of type (2222), then the signature of $\mathcal{O}_f$ is $(2, 2, 2, 2)$.

Since a Lattès-type map has parabolic orbifold, we know by Proposition 2.13 that the orbifold of each such map has one of the signatures $(2, 2, 2, 2)$, $(2, 4, 4)$, $(3, 3, 3)$, or $(2, 3, 6)$ (this also follows from Remark 3.17). The last three cases lead to nothing new and essentially give Lattès maps.

PROPOSITION 3.18. *Let $f \colon S^2 \to S^2$ be a Lattès-type map with orbifold signature $(2, 4, 4)$, $(3, 3, 3)$, or $(2, 3, 6)$. Then f is topologically conjugate to a Lattès map.*

To prove this statement we need a lemma that gives a criterion when an $\mathbb{R}$-linear map $L \colon \mathbb{C} \to \mathbb{C}$ is $\mathbb{C}$-linear. Here the $\mathbb{R}$-linearity or $\mathbb{C}$-linearity for L of course mean that $L(\lambda z) = \lambda L(z)$ for all $z \in \mathbb{C}$ and all $\lambda \in \mathbb{R}$ or all $\lambda \in \mathbb{C}$, respectively.

LEMMA 3.19. *Let $L \colon \mathbb{C} \to \mathbb{C}$ be an $\mathbb{R}$-linear map with $\det(L) > 0$. Suppose there exist $\zeta \in \mathbb{C} \setminus \mathbb{R}$ and $\eta \in \mathbb{C}$ with $L(\zeta z) = \eta L(z)$ for all $z \in \mathbb{C}$. Then L is $\mathbb{C}$-linear.*

PROOF. Since L is $\mathbb{R}$-linear, there exist unique numbers $a, b \in \mathbb{C}$ such that $L(z) = az + b\overline{z}$ for $z \in \mathbb{C}$. Then the determinant of L (as an $\mathbb{R}$-linear map) is given by $\det(L) = |a|^2 - |b|^2 > 0$. It follows that $a \neq 0$.

Now for all $z \in \mathbb{C}$ we have

$$L(\zeta z) = \zeta a z + \overline{\zeta}\, b\overline{z} = \eta L(z) = \eta a z + \eta b\overline{z},$$

and so

$$\zeta a = \eta a \quad \text{and} \quad \overline{\zeta} b = \eta b.$$

Since $a \neq 0$, the first equation implies $\zeta = \eta$. Then the second equation combined with the fact that $\zeta \notin \mathbb{R}$ gives $b = 0$. Hence $L(z) = az$ for $z \in \mathbb{C}$. This shows that L is $\mathbb{C}$-linear. $\square$

PROOF OF PROPOSITION 3.18. We know that there exists a crystallographic group G as well as a G-equivariant affine homeomorphism $A \colon \mathbb{R}^2 \to \mathbb{R}^2$ with $\det(L_A) > 1$ such that f arises as in (3.5). By conjugation with a suitable map in $\mathrm{Aut}(\mathbb{C})$, we may assume that G is one of the groups $\widetilde{G}$ listed in Theorem 3.7 (see the discussion preceding (3.17)). The quotient $S^2 = \mathbb{R}^2/G \cong \mathbb{C}/G$ has a natural complex structure, and so we can identify it with the Riemann sphere $\widehat{\mathbb{C}}$ by the uniformization theorem. We will show that under this identification, f is a Lattès map.

By assumption the signature of the orbifold of f is $(2, 4, 4)$, $(3, 3, 3)$, or $(2, 3, 6)$. By Remark 3.17 this means that G is a crystallographic group $\widetilde{G}$ of type (244), (333), or (236) in Theorem 3.7. In these cases, G contains a rotation g_0 of the form $z \in \mathbb{C} \mapsto g_0(z) = \zeta z$, where $\zeta = e^{2\pi i/n}$ is a primitive n-th root of unity with $n \in \{3, 4, 6\}$. In particular, $\zeta \in \mathbb{C} \setminus \mathbb{R}$. Since the homeomorphism A passes to the quotient $\widehat{\mathbb{C}} \cong \mathbb{C}/G$, this map is G-equivariant (Lemma A.22) and so there exists $g_1 \in G$ such that

$$(3.23) \qquad\qquad A \circ g_0 = g_1 \circ A.$$

Let $L = L_A$ be the linear part of A. This is an $\mathbb{R}$-linear map satisfying $\det(L) > 0$. Since G consists of orientation-preserving isometries, the linear part of g_1 is $\mathbb{C}$-linear, as for every map in G. Comparing linear parts of the maps in (3.23), we see that there exists $\eta \in \mathbb{C}$, $|\eta| = 1$, such that

$$L(\zeta z) = \eta L(z)$$

for all $z \in \mathbb{C}$. This shows that L satisfies the hypotheses of Lemma 3.19 and we conclude that L is $\mathbb{C}$-linear. Hence A can be written in the form $A(z) = \alpha z + \beta$ for $z \in \mathbb{C}$, where $\alpha, \beta \in \mathbb{C}$, $\alpha \neq 0$. In particular, A is holomorphic, and it follows that f is indeed a Lattès map. $\square$

By Proposition 3.18 only Lattès-type maps whose orbifolds have signature $(2, 2, 2, 2)$ give genuinely new maps beyond Lattès maps. We summarize some facts about these maps in the following discussion. For specific maps see Examples 6.14 and 16.8.

EXAMPLE 3.20 (Lattès-type maps with signature $(2, 2, 2, 2)$). We know that each such Lattès-type map arises from a crystallographic group G of type (2222) (see Remark 3.17). In this case, it is elementary to check that the isometries in G remain isometries on $\mathbb{R}^2$ not only if we conjugate them by an isometry on $\mathbb{R}^2$, but even if we conjugate them by an affine homeomorphism $h \colon \mathbb{R}^2 \to \mathbb{R}^2$. It follows that the class of crystallographic groups G of type (2222) is preserved under conjugation by such a homeomorphism h. Similarly, the class of orientation-preserving affine homeomorphisms $A \colon \mathbb{R}^2 \to \mathbb{R}^2$ is preserved under conjugation by h. As in the discussion preceding (3.17), we may therefore assume in the construction of Lattès-type maps with orbifold signature $(2, 2, 2, 2)$ that the underlying lattice Γ of G is equal to the integer lattice $\Gamma = \mathbb{Z}^2$ and that the group G consists of all isometries $g \colon \mathbb{R}^2 \to \mathbb{R}^2$ of the form

$$(3.24) \qquad u \in \mathbb{R}^2 \mapsto g(u) = \pm u + \gamma, \quad \text{where } \gamma \in \Gamma = \mathbb{Z}^2.$$

The quotient space $\mathbb{R}^2/G$ is a 2-sphere S^2. One can identify it with a pillow (see Section A.10) as follows. Let

$$R := [0,1] \times [0,1/2],$$

$$S_1 := [0,1/2] \times [0,1/2], \quad \text{and} \quad S_2 := [1/2,1] \times [0,1/2].$$

Then $R = S_1 \cup S_2$ is a fundamental domain (see Section A.7) for the action of G on $\mathbb{R}^2$, i.e., R contains a representative from every orbit, and this representative is unique in R if it lies in the interior of R. So $\mathbb{R}^2/G$ is obtained from R by identifying certain points on the boundary of R. For this one folds the rectangle R along the line $\ell = \{(x,y) \in \mathbb{R}^2 : x = 1/2\}$ and identifies corresponding points on the boundaries of the two squares S_1 and S_2 under this folding operation; so for example the point $(0,t) \in \mathbb{R}^2$ is identified with $(1,t) \in \mathbb{R}^2$ for $t \in [0,1/2]$. The pillow obtained in this way is our quotient space $\mathbb{R}^2/G$.

Let $A(u) = L_A(u) + u_0$, $u \in \mathbb{R}^2$, be an orientation-preserving affine homeomorphism. Then A induces a map on the quotient $\mathbb{R}^2/G$ if and only if A is G-equivariant, or equivalently if $A \circ g \circ A^{-1} \in G$ for each $g \in G$ (see Lemma A.22). Exactly as in Proposition 3.14, this is the case if and only if $2u_0 \in \Gamma$ and $L_A(\Gamma) \subset \Gamma$.

Since $\Gamma = \mathbb{Z}^2$ the latter condition is true precisely if the matrix representing L_A with respect to the standard basis in $\mathbb{R}^2$ has integer coefficients. So we conclude that the homeomorphism A induces a map on $S^2 = \mathbb{R}^2/G$ precisely if A has the form

$$(3.25) \quad A(u) = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \frac{1}{2} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \text{ for } u = \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2,$$

where $a, b, c, d, x_0, y_0 \in \mathbb{Z}$ and $\det(L_A) = ad - bc \geq 1$.

Here the last inequality follows from our assumption that A is orientation-preserving.

Let $f \colon S^2 \to S^2$ be the branched covering map induced by A as in (3.25). If $\det(L_A) = 1$, then A^{-1} also is of the form (3.25), and so f has a continuous inverse induced by A^{-1}. In this case $f \colon S^2 \to S^2$ is a homeomorphism.

If $\det(L_A) \geq 2$, then $\deg(f) = \det(L_A) \geq 2$ by Lemma 3.16. Then f is a Lattès-type map, and we know that in this case the orbifold of f has signature $(2,2,2,2)$ (see Remark 3.17).

By the previous discussion we established the following statement.

PROPOSITION 3.21. *Let G be the group consisting of all isometries $g \colon \mathbb{R}^2 \to \mathbb{R}^2$ of the form* (3.24) *and $A \colon \mathbb{R}^2 \to \mathbb{R}^2$ be an affine orientation-preserving homeomorphism as in* (3.25).

Then A descends to a map $f\colon S^2 \to S^2$ on the quotient $\mathbb{R}^2/G \cong S^2$, i.e., if $\Theta\colon \mathbb{R}^2 \to \mathbb{R}^2/G \cong S^2$ is the quotient map, then $\Theta \circ A = f \circ \Theta$.

If $ad - bc = 1$, then f is an orientation-preserving homeomorphism. If $ad - bc \geq 2$, then f is a Lattès-type map whose orbifold has signature $(2, 2, 2, 2)$.

Moreover, every Lattès-type map with orbifold signature $(2, 2, 2, 2)$ is conjugate to a map f obtained in this way.

An obvious question is when a Lattès-type map f is Thurston equivalent to a rational map. This is always true if the signature of $\mathcal{O}_f$ is equal to $(2, 4, 4)$, $(3, 3, 3)$, or $(2, 3, 6)$, because then f is even conjugate to a Lattès map (Proposition 3.18).

If $\mathcal{O}_f$ has signature $(2, 2, 2, 2)$, the question is answered by the following statement which can be seen as complementary to Thurston's characterization of rational Thurston maps with hyperbolic orbifold (Theorem 2.17).

THEOREM 3.22 (Rational maps with orbifold signature $(2, 2, 2, 2)$). *Let $f\colon S^2 \to S^2$ be a Lattès-type map with orbifold signature $(2, 2, 2, 2)$ and $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ be an affine map as in Definition 3.3 with linear part L_A. Then f is Thurston equivalent to a rational map if and only if L_A is a real multiple of the identity map on $\mathbb{R}^2$ or the eigenvalues of L_A belong to $\mathbb{C} \setminus \mathbb{R}$.*

We will not prove this here, but refer to [**DH93**, Proposition 9.7] for an essentially equivalent statement.

In the final part of this section we provide a proof of Proposition 3.6 that characterizes Lattès-type maps up to Thurston equivalence.

PROOF OF PROPOSITION 3.6. Suppose first that f is a Thurston map that is Thurston equivalent to a Lattès-type map g. Then f and g have the same orbifold signature (Proposition 2.14). Since g has a parabolic orbifold and no periodic critical points by Proposition 3.5, the same is true for the map f as follows from Propositions 2.13 and 2.9 (ii).

For the converse direction, suppose that f is a Thurston map with parabolic orbifold $\mathcal{O}_f$ and no periodic critical points. We know that then the signature of $\mathcal{O}_f$ is $(2, 2, 2, 2)$, $(2, 4, 4)$, $(2, 3, 6)$, or $(3, 3, 3)$.

In the last three cases the map has precisely three postcritical points. As we will see later (Theorem 7.3), every Thurston map f with three postcritical points is Thurston equivalent to a rational map R. Then the signatures of the orbifolds of f and R are the same (Proposition 2.14), and so R is a Thurston map with parabolic orbifold $\mathcal{O}_f$ and no periodic critical points. By Definition 3.2 the map R is a Lattès map, and the statement follows in this case.

So we are left with the case where $\mathcal{O}_f$ has signature $(2, 2, 2, 2)$. We may assume that f is defined on $\widehat{\mathbb{C}}$. Let $\alpha = \alpha_f$ be the ramification function and $\Theta\colon \mathbb{C} \cong \mathbb{R}^2 \to (\widehat{\mathbb{C}}, \alpha)$ be the universal orbifold covering map as provided by Theorem 3.10. The group G of deck transformations of Θ is a crystallographic group of type (2222). Let Γ be the underlying lattice of G. Then G consists precisely of the isometries on $\mathbb{R}^2$ of the form $u \mapsto \pm u + \gamma$, $\gamma \in \Gamma$. We also consider the torus $T^2 = \mathbb{R}^2/G_{\mathrm{tr}} = \mathbb{R}^2/\Gamma$ and the quotient map $\pi\colon \mathbb{R}^2 \to \mathbb{R}^2/\Gamma$.

We now repeat part of the arguments for the proof of the implications (i) $\Rightarrow$ (ii) and (ii) $\Rightarrow$ (iii) in Theorem 3.1. The difference is that we do not have holomorphicity of the maps involved, but this is mostly inessential. First, the parabolicity of $\mathcal{O}_f$ in combination with the uniqueness part of Theorem 3.10 implies that there exists an orientation-preserving homeomorphism $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ such that $f \circ \Theta = \Theta \circ A$. Again A is G and G_{tr}-equivariant. This allows us to push the map A to T^2, and we obtain a diagram as in (3.15) (with $S^2 = \widehat{\mathbb{C}}$), where $\overline{\Theta}\colon T^2 \to \widehat{\mathbb{C}}$ is a branched covering map and $\overline{A}\colon T^2 \to T^2$ is a torus endomorphism.

In particular, f is a quotient of the torus endomorphism $\overline{A}$ and so by Lemma 3.12 (ii) the set $\mathrm{post}(f)$ is equal to the set of critical values of $\overline{\Theta}$. Since $\Theta = \overline{\Theta} \circ \pi$ and π is a covering map, the set of critical values of Θ and $\overline{\Theta}$ are the same. Now

$$\mathrm{crit}(\Theta) = \{u \in \mathbb{R}^2 : \#G_u = 2\} = \tfrac{1}{2}\Gamma$$

and it follows that

$$(3.26) \qquad\qquad \mathrm{post}(f) = \Theta\big(\tfrac{1}{2}\Gamma\big).$$

By Lemma A.23 (iii) there exists a linear map $L\colon \mathbb{R}^2 \to \mathbb{R}^2$ (essentially the map induced by $\overline{A}$ on the fundamental group Γ of T^2) with $L(\Gamma) \subset \Gamma$ such that

$$(3.27) \qquad\qquad A \circ \tau_\gamma \circ A^{-1} = \tau_{L(\gamma)}$$

for $\gamma \in \Gamma$, where τ_γ denotes the translation $u \in \mathbb{R}^2 \mapsto \tau_\gamma(u) := u + \gamma$. This relation can be rewritten as

$$(3.28) \qquad A(u + \gamma) = A(u) + L(\gamma) \quad \text{for each } u \in \mathbb{R}^2, \gamma \in \Gamma.$$

By Lemma A.23 (iv) we have $\det(L) = \deg(\overline{A}) = \deg(f) \geq 2$, and so L is an orientation-preserving linear homeomorphism on $\mathbb{R}^2$.

The G-equivariance of A also implies that for suitable sign and $\gamma_0 \in \Gamma$ we have $A(-u) = \pm A(u) + \gamma_0$ for each $u \in \mathbb{R}^2$. Here we necessarily have the negative sign on the right hand side; otherwise, by

setting $u = 0$ we obtain $\gamma_0 = 0$ and $A(u) = A(-u)$ for $u \in \mathbb{R}^2$. This is impossible, because A is a homeomorphism and hence injective.

It follows that $A(-u) = -A(u) + \gamma_0$ for $u \in \mathbb{R}^2$; setting $u = 0$ we see that $\gamma_0 = 2A(0)$. We conclude that

$$A(0) \in \tfrac{1}{2}\Gamma \tag{3.29}$$

and

$$A(-u) = -A(u) + 2A(0) \quad \text{for each } u \in \mathbb{R}^2. \tag{3.30}$$

Now we define
$$\widetilde{A}(u) = L(u) + A(0)$$
for $u \in \mathbb{R}^2$. Then $\widetilde{A}$ is an orientation-preserving affine homeomorphism whose linear part L satisfies $\det(L) \geq 2$. Moreover, the map $\widetilde{A}$ satisfies relations as in (3.28) and (3.30). This together with the fact $\widetilde{A}(0) = A(0) \in \tfrac{1}{2}\Gamma$ implies that $\widetilde{A}$ is G-equivariant. It follows that there exists a Lattès-type map $g \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ with $g \circ \Theta = \Theta \circ \widetilde{A}$.

We claim that f and g are Thurston equivalent. To see this, consider the orientation-preserving homeomorphism $B := \widetilde{A}^{-1} \circ A$. Note that $\widetilde{A}^{-1}(u) = L^{-1}(u - A(0))$ for $u \in \mathbb{R}^2$. This and the relations (3.28) and (3.30) imply that

$$B(\pm u + \gamma) = \pm B(u) + \gamma \quad \text{for each } u \in \mathbb{R}^2, \gamma \in \Gamma. \tag{3.31}$$

In particular, B is G-equivariant and so there exists a continuous map $h \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ such that

$$h \circ \Theta = \Theta \circ B. \tag{3.32}$$

Note that

$$g \circ h \circ \Theta = g \circ \Theta \circ B = \Theta \circ \widetilde{A} \circ B = \Theta \circ A = f \circ \Theta.$$

Since $\Theta \colon \mathbb{R}^2 \to \widehat{\mathbb{C}}$ is surjective, we conclude that $g \circ h = f$. So the Thurston equivalence of f and g will follow if we can show that h is a homeomorphism that is isotopic to $\mathrm{id}_{\widehat{\mathbb{C}}}$ rel. $\mathrm{post}(f)$.

It is easy to see that the inverse map B^{-1} satisfies a similar relation as in (3.31). Hence B^{-1} is also G-equivariant and descends to a continuous map on $\widehat{\mathbb{C}}$. This map is an inverse map for h, and it follows that h is a homeomorphism. Since B is orientation-preserving, the same is true for h.

Using (3.31) with the negative sign and $u = \tfrac{1}{2}\gamma$, one sees that

$$B(\tfrac{1}{2}\gamma) = \tfrac{1}{2}\gamma \quad \text{for each } \gamma \in \Gamma.$$

If we combine this with (3.26) and (3.32), we conclude that h fixes each of the four points in $\mathrm{post}(f)$.

It remains to show that h is isotopic to $\mathrm{id}_{\widehat{\mathbb{C}}}$ rel. $\mathrm{post}(f)$. There is no easy self-contained argument for this and we have to invoke some facts about the mapping class group of a sphere with four punctures (as given by the points in $\mathrm{post}(f)$) and its relation to the mapping class group of a torus (for the definition of the mapping class group and the facts needed see [**FM11**, Sections 2.1 and 2.2]).

We have a well-defined torus involution $\overline{I}\colon T^2 \to T^2$ induced by the map $u \in \mathbb{R}^2 \mapsto I(u) = -u$. Then $\pi \circ I = \overline{I} \circ \pi$ and $\overline{I}$ is the generator of a cyclic group $\overline{G}$ of order 2 acting on T^2. The map $\overline{\Theta}\colon T^2 \to \widehat{\mathbb{C}}$ corresponds to the quotient map $T^2 \to T^2/\overline{G} \cong \widehat{\mathbb{C}}$. Moreover, $\overline{I}$ has four fixed points given by the images of the points in $\frac{1}{2}\Gamma$ under the quotient map $\pi\colon \mathbb{R}^2 \to T^2 \cong \mathbb{R}^2/\Gamma$. These fixed points of $\overline{I}$ in turn are mapped by $\overline{\Theta}$ to the points in the set $P \coloneqq \mathrm{post}(f)$ by $\overline{\Theta}$ as follows from (3.26).

It is now a general fact that an orientation-preserving homeomorphism on $\widehat{\mathbb{C}}$ fixing the points in P (i.e., the images of the fixed points of $\overline{I}$ under the projection map $\overline{\Theta}$) is isotopic to the identity rel. P if it has a lift to T^2 that induces the identity map on the fundamental group of T^2 (this is essentially shown in [**FM11**, Proof of Proposition 2.7, pp. 59–60]).

In our situation, such a lift to T^2 of the homeomorphism h can easily be obtained. Namely, it follows from (3.31) there exists a homeomorphism $\overline{B}\colon T^2 \to T^2$ such that $\pi \circ B = \overline{B} \circ \pi$. Then by definition B is a lift of $\overline{B}$ under π. Now the translations $u \mapsto u + \gamma$, $\gamma \in \Gamma$, form the group of deck transformations of π which represents the fundamental group of T^2. Then it follows from (3.31) that the induced map of $\overline{B}$ on the fundamental group is the identity (see the remarks after Lemma A.23 for a related discussion). Finally, $\overline{B}$ is a lift of h under $\overline{\Theta}$, because we have

$$\overline{\Theta} \circ \overline{B} \circ \pi = \overline{\Theta} \circ \pi \circ B = \Theta \circ B = h \circ \Theta = h \circ \overline{\Theta} \circ \pi,$$

and so

$$\overline{\Theta} \circ \overline{B} = h \circ \overline{\Theta}.$$

This shows that h is indeed isotopic to $\mathrm{id}_{\widehat{\mathbb{C}}}$ rel. $P = \mathrm{post}(f)$. $\square$

3.5. Examples of Lattès maps

In this section we present some examples of Lattès maps based on Theorem 3.1 (ii).

The map Θ in Theorem 3.1 (ii) is the universal orbifold covering map of $\mathcal{O}_f = (\widehat{\mathbb{C}}, \alpha_f)$, the orbifold associated with f. We can use Θ

to push the Euclidean metric on $\mathbb{C}$ forward to a metric ω on $\widehat{\mathbb{C}}$. This metric is the *canonical orbifold metric* of f introduced at the end of Section 2.5 and discussed in detail in Section A.10. In the case of Lattès maps, $\mathcal{O}_f$ is parabolic and the canonical orbifold metric ω is unique up to a scaling factor. The metric space $(\widehat{\mathbb{C}}, \omega)$ is locally isometric to $\mathbb{C}$, except at the postcritical points of f (i.e., the points with $\alpha_f(p) \geq 2$) where $(\widehat{\mathbb{C}}, \omega)$ is locally isometric to a Euclidean cone of angle $2\pi/\alpha_f(p)$.

One can reverse this procedure, and construct Θ in a simple geometric way. This leads to a very explicit geometric description of Lattès maps. To explain this, we start with a given crystallographic group G not isomorphic to $\mathbb{Z}^2$ as in Theorem 3.7. We first consider the types (244), (333), and (236). The type (2222) is different and will be treated later.

So let G be the group of type (244) as in Theorem 3.7. We explain the construction in detail in this case. For the types (333) and (236) the considerations are completely analogous and we will not repeat the construction in these cases.

We start with the invariant tiling for G shown in Figure 3.1. The union of a white and a black triangle with a common edge forms a fundamental domain for the action of G on $\mathbb{C}$.

Let T be one of the white triangles in this tiling. Then T is a right-angled Euclidean triangle with angles $\pi/2$, $\pi/4$, $\pi/4$. We glue an isometric copy $T_{\mathtt{w}}$ of T, colored white, with another isometric copy $T_{\mathtt{b}}$ of T, colored black, along their boundaries to form a pillow Δ (see Section A.10) which represents the quotient space $\mathbb{C}/G$. Then Δ is a topological 2-sphere. We fix an orientation on Δ, and equip Δ with the path metric ω induced by the Euclidean metric on the two copies of T. One can now define a map $\widetilde{\Theta}\colon \mathbb{C} \to \Delta$ as follows. The map $\widetilde{\Theta}$ sends each white triangle $T \subset \mathbb{C}$ from the tiling in Figure 3.1 to (the white triangle) $T_{\mathtt{w}} \subset \Delta$ and each black triangle $T \subset \Delta$ to (the black triangle) $T_{\mathtt{b}} \subset \Delta$ by an orientation-preserving isometry. Then $\widetilde{\Theta}\colon \mathbb{C} \to \Delta$ is a well-defined branched covering map. Note that if G is of type (333), then a similar construction does not lead to a well-defined map due to a rotational ambiguity which is not present for type (244) and (236). In this case, one has to single out one of the common vertices v of $T_{\mathtt{w}}$ and $T_{\mathtt{b}}$ and impose the additional requirement that the piecewise isometry $\widetilde{\Theta}$ sends the points marked by a black dot in Figure 3.2 to v.

It is clear from the definition of $\widetilde{\Theta}$ that $\widetilde{\Theta} = \widetilde{\Theta} \circ g$ for each $g \in G$. On the other hand, if $\widetilde{\Theta}(w) = \widetilde{\Theta}(z)$ for $z, w \in \mathbb{C}$, then we may pick two triangles T and T' of the same color in the tiling as represented by Figure 3.1 such that $z \in T$ and $w \in T'$. There is a unique element

$g \in G$ with $g(T) = T'$. Then $g(z), w \in T'$ and

$$\widetilde{\Theta}(g(z)) = \widetilde{\Theta}(z) = \widetilde{\Theta}(w).$$

Since $\widetilde{\Theta}$ is an isometry on T' and hence injective on T', it follows that $w = g(z)$. This shows that $\widetilde{\Theta}(w) = \widetilde{\Theta}(z)$ for $z, w \in \mathbb{C}$ if and only if there exists $g \in G$ such that $w = g(z)$. Hence $\widetilde{\Theta}$ is induced by G.

The topological 2-sphere Δ is a polyhedral surface equipped with a locally Euclidean metric with three conical singularities. In particular, Δ carries a natural conformal structure, with respect to which it is conformally equivalent to $\widehat{\mathbb{C}}$. Let $a, b, c \in \widehat{\mathbb{C}}$ be three distinct points. Then by the uniformization theorem there exists a conformal map $\varphi \colon \Delta \to \widehat{\mathbb{C}}$ that sends the vertices of the pillow to the points a, b, c. In fact, one can construct φ quite explicitly by first mapping $T_{\mathtt{w}}$ conformally to the upper half-plane, and $T_{\mathtt{b}}$ to the lower half-plane such that the vertices of the triangles are sent to $0, 1, \infty$, followed by a suitable Möbius transformation. If we define $\Theta = \varphi \circ \widetilde{\Theta} \colon \mathbb{C} \to \widehat{\mathbb{C}}$, then this map is holomorphic and induced by G.

We may assume that a is the image of the common vertex of $T_{\mathtt{w}}$ and $T_{\mathtt{b}}$ with angle $\pi/2$. We define the map $\alpha \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{N}}$ by $\alpha(a) = 2$, $\alpha(b) = \alpha(c) = 4$, and $\alpha(p) = 1$ for all $p \in \widehat{\mathbb{C}} \setminus \{a, b, c\}$. Then Θ is the universal orbifold covering map of the orbifold $\mathcal{O} = (\widehat{\mathbb{C}}, \alpha)$. If we equip $\widehat{\mathbb{C}}$ with the canonical orbifold metric (scaled appropriately) given by Θ, then the map φ is in fact an isometry.

EXAMPLE 3.23 (A Lattès map with orbifold signature $(2, 4, 4)$). Based on the previous considerations, it is now easy to give an example of a Lattès map with orbifold signature $(2, 4, 4)$. Let Δ be the pillow as above obtained from gluing together the triangles $T_{\mathtt{w}}$ and $T_{\mathtt{b}}$. We call $T_{\mathtt{w}} \subset \Delta$ the white 0-tile, and $T_{\mathtt{b}} \subset \Delta$ the black 0-tile.

We divide $T_{\mathtt{w}}$ along the perpendicular bisector of its hypotenuse into two triangles T_1 and T_2 that are similar to $T_{\mathtt{w}}$ by the scaling factor $\sqrt{2}$. In the same way we divide $T_{\mathtt{b}}$ into two similar triangles. These four small triangles are called 1-tiles; we color them black and white in a checkerboard fashion so that two 1-tiles with a common edge have distinct colors as indicated on the left in Figure 3.4. Note that the triangle with the lighter gray shading is on the back of the triangular pillow.

A map $g \colon \Delta \to \Delta$ is now given as follows. We send each white 1-tile to $T_{\mathtt{w}}$ and each black 1-tile to $T_{\mathtt{b}}$ by a similarity as indicated in Figure 3.4. Then g is a branched covering map. The points labeled 0 and ∞ are its critical points and we have the following ramification
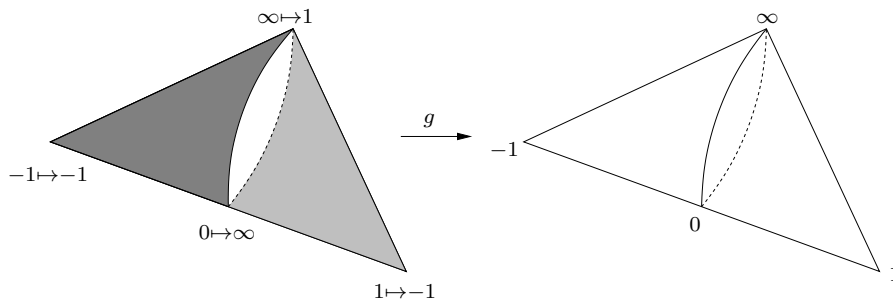
FIGURE 3.4. A Lattès map with orbifold signature $(2, 4, 4)$.
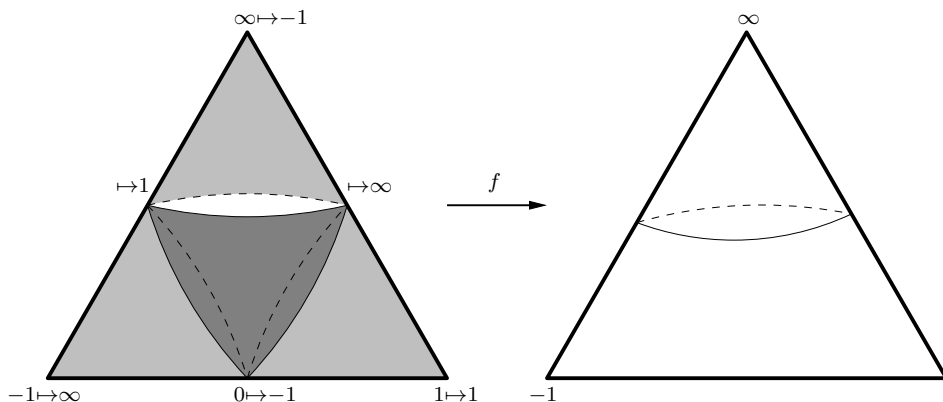
FIGURE 3.5. A Lattès map with orbifold signature $(3, 3, 3)$.

portrait:

$$(3.33) \qquad\qquad 0 \xrightarrow{\;2:1\;} \infty \xrightarrow{\;2:1\;} 1 \longrightarrow -1.$$

Thus g is a Thurston map with $\operatorname{post}(g) = \{-1, 1, \infty\}$ and orbifold signature $(2, 4, 4)$. Note that g is given as $z \mapsto (1 + \boldsymbol{i})z$ in suitable Euclidean coordinates (i.e., if we identify $T_{\mathtt{w}}$ and $T_{\mathtt{b}}$ with a triangle $T \subset \mathbb{C}$ in the obvious way, so that the fixed point $-1 \in \Delta$ is identified with $0 \in \mathbb{C}$). In particular, $g\colon \Delta \to \Delta$ is holomorphic with respect to the given conformal structure on Δ.

Let $\varphi\colon \Delta \to \widehat{\mathbb{C}}$ be the conformal map considered earlier. Then the map $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ given by $f := \varphi \circ g \circ \varphi^{-1}$ is a rational Thurston map. This map is actually a Lattès map by Theorem 3.1, because its orbifold signature is $(2, 4, 4)$. If the conformal conjugacy φ is chosen such that it sends the points labeled $-1, 1, \infty \in \Delta$ to $-1, 1, \infty \in \widehat{\mathbb{C}}$, respectively, then it is not hard to check that $f(z) = 1 - 2/z^2$.

The map f can also be represented as in Theorem 3.1 (ii). Namely, f is the quotient of the map $A\colon z \mapsto (1 + \boldsymbol{i})z$ by the crystallographic group $\widetilde{G}$ of type (244) in Theorem 3.7.
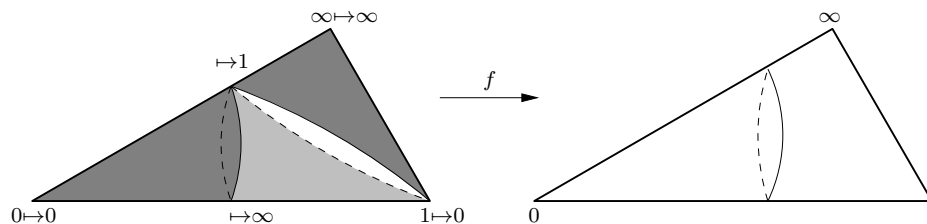
FIGURE 3.6. A Lattès map with orbifold signature $(2, 3, 6)$.

EXAMPLE 3.24. In Figure 3.5 we illustrate a Lattès map f with orbifold signature $(3, 3, 3)$. The map given here is obtained as a quotient of the map $A(z) = 2z$ by the crystallographic group $\widetilde{G}$ of type (333) in Theorem 3.7. The Riemann sphere $\widehat{\mathbb{C}}$ equipped with the canonical orbifold metric is isometric to a pillow obtained by gluing together two equilateral triangles along their boundaries. The map f has in fact the form

$$f(z) = 1 - 2\frac{(z-1)(z+3)^3}{(z+1)(z-3)^3}.$$

EXAMPLE 3.25. A Lattès map f with orbifold signature $(2, 3, 6)$ is illustrated in Figure 3.6. It is obtained as the quotient of the map $A(z) = \frac{1}{2}(3 + i\sqrt{3})z$ by a crystallographic group $\widetilde{G}$ of type (236) in Theorem 3.7. The map f is given by

$$f(z) = 1 - \frac{(3z+1)^3}{(9z-1)^2}.$$

Note that the rational maps f discussed in the previous three examples all have the property that the extended real line $\mathcal{C} := \widehat{\mathbb{R}} = \mathbb{R} \cup \{\infty\} \subset \widehat{\mathbb{C}}$ contains $\mathrm{post}(f)$ and is f-invariant in the sense that $f(\mathcal{C}) \subset \mathcal{C}$. This leads to a simple description of the maps in geometric terms.

In Chapter 12 we will consider more general Thurston maps with such invariant Jordan curves, and in Chapter 15 we will see that such an invariant Jordan curve exists for each sufficiently high iterate of an expanding Thurston map. This means that we obtain similar geometric descriptions in much greater generality.

We now turn to crystallographic groups of type (2222) and the corresponding Lattès maps. For the moment we allow groups G of this type with an arbitrary underlying (rank-2) lattice $\Gamma \subset \mathbb{C}$. Then G consists of all isometries on $\mathbb{C}$ of the form $z \mapsto \pm z + \gamma$, where $\gamma \in \Gamma$.

Holomorphic maps $\Theta\colon \mathbb{C} \to \widehat{\mathbb{C}}$ induced by G can be obtained from the Weierstraß $\wp$-function (see [**Ah79**, Chapter 7.3] for general background). Recall that the Weierstraß $\wp$-function for a given lattice

$\Gamma \subset \mathbb{C}$ is defined as

$$(3.34) \qquad \wp(z; \Gamma) = \frac{1}{z^2} + \sum_{\gamma \in \Gamma \setminus \{0\}} \left(\frac{1}{(z - \gamma)^2} - \frac{1}{\gamma^2} \right) \quad \text{for } z \in \mathbb{C}.$$

Then $\wp = \wp(\,\cdot\,; \Gamma)$ is an even meromorphic function on $\mathbb{C}$ with the period lattice Γ. It descends to a holomorphic map $\overline{\wp} \colon \mathbb{T} \to \widehat{\mathbb{C}}$ on the torus $\mathbb{T} = \mathbb{C}/\Gamma$ with $\deg(\overline{\wp}) = 2$. The function $\wp$ satisfies the differential equation

$$(3.35) \qquad (\wp')^2 = 4(\wp - e_1)(\wp - e_2)(\wp - e_3).$$

The numbers $e_1, e_3, e_3 \in \mathbb{C}$ here are three distinct values (depending on Γ) with

$$(3.36) \qquad e_1 + e_2 + e_3 = 0.$$

The critical values of $\wp \colon \mathbb{C} \to \widehat{\mathbb{C}}$ are e_1, e_2, e_3 and ∞. Actually, $\wp \colon \mathbb{C} \to \widehat{\mathbb{C}}$ is a holomorphic branched covering map satisfying

$$(3.37) \qquad \deg(\wp, z) = \begin{cases} 2 & \text{if } \wp(z) \in \{e_1, e_2, e_3, \infty\}, \\ 1 & \text{otherwise.} \end{cases}$$

This shows that $\wp$ is the universal orbifold covering map of the orbifold $(\widehat{\mathbb{C}}, \alpha)$, where $\alpha(p) = 2$ if $p \in \{e_1, e_2, e_3, \infty\}$ and $\alpha(p) = 1$ for $p \in \widehat{\mathbb{C}} \setminus \{e_1, e_2, e_3, \infty\}$. If G is the crystallographic group corresponding to Γ as above, then $\wp$ is induced by G. Indeed, $\wp(w) = \wp(z)$ if and only if $w = \pm z + \gamma$ for some $\gamma \in \Gamma$.

As is well known and classical, one can reverse this procedure and start with the differential equation (3.35): if three distinct values e_1, e_2, $e_3 \in \mathbb{C}$ with (3.36) are given, then there exists a (unique) lattice Γ such that the corresponding function $\wp = \wp(\,\cdot\,; \Gamma)$ satisfies (3.35). See [**Ah79**, Chapters 7.3.3 and 7.3.4].

This implies that the universal orbifold covering map Θ of an orbifold $(\widehat{\mathbb{C}}, \alpha)$ with signature $(2, 2, 2, 2)$ can always be obtained from a Weierstraß $\wp$-function followed by a suitable Möbius transformation. Indeed, if $p_1, \ldots, p_4 \in \widehat{\mathbb{C}}$ are the four distinct points with $\alpha(p_k) = 2$ for $k = 1, \ldots, 4$, then there exists a Möbius transformation ψ on $\widehat{\mathbb{C}}$ such that $\psi(\{p_1, \ldots, p_4\}) = \{e_1, e_2, e_3, \infty\}$, where $e_1, e_2, e_3 \in \mathbb{C}$ are three distinct points satisfying (3.36) (first map p_4 to ∞ by a Möbius transformation and then apply a suitable translation). Then we can find a lattice Γ such that the corresponding function $\wp = \wp(\,\cdot\,; \Gamma)$ satisfies (3.35) with these values e_1, e_2, e_3. If we define $\Theta = \psi^{-1} \circ \wp(\,\cdot\,; \Gamma)$, then Θ is the universal orbifold covering map of the given orbifold $(\widehat{\mathbb{C}}, \alpha)$.

Actually, one can make one more reduction here. Namely, in the last statement we may assume that the lattice has the form $\Gamma = \mathbb{Z} \oplus \mathbb{Z}\tau$ with $\tau \in \mathbb{C}$ and $\mathrm{Im}(\tau) > 0$. This immediately follows from the homogeneity property

$$\wp(\lambda z; \lambda \Gamma) = \frac{1}{\lambda^2} \wp(z; \Gamma), \quad z \in \mathbb{C}, \, \lambda \in \mathbb{C} \setminus \{0\},$$

of the $\wp$-function.

We will now describe a more geometric construction for the map Θ and the orbifold $\mathbb{C}/G$ similar in spirit to the one given for the crystallographic groups of type (244) as discussed above.

For this we fix $\tau \in \mathbb{C}$ with $\mathrm{Im}(\tau) > 0$ and consider the crystallographic group G of type (2222) given by all isometries $z \mapsto \pm z + \gamma$, where $\gamma \in \Gamma := \mathbb{Z} \oplus \mathbb{Z}\tau$. A fundamental domain for G is the Euclidean triangle $T \subset \mathbb{C}$ with vertices $0, 1, \tau$. To obtain the quotient space $\mathbb{C}/G$ from T, we divide T into four similar triangles by connecting the midpoints of each side and fold T along the edges of these smaller triangles. In this way, we can build a tetrahedron Δ in Euclidean 3-space as indicated in Figure 3.7 (the two halves of each side are identified). Here we allow the degenerate case that Δ is a pillow. This happens precisely when T has a right angle.

We equip Δ with the path metric ω induced by the Euclidean metric on T, and the orientation induced by the orientation on $T \subset \mathbb{C}$. Then Δ is an oriented topological 2-sphere carrying a locally Euclidean metric with four conical singularities. The cone angles at these singularities are π. If we assign to each of these singularities the value 2 for a ramification function on Δ and the value 1 to all other points, then we can consider Δ as an orbifold with signature $(2, 2, 2, 2)$.

The parallelogram $P \subset \mathbb{C}$ spanned by 1 and τ is a fundamental domain for the group of translations $G_{\mathrm{tr}} \subset G$ given by all maps of the form $z \mapsto z + \gamma$, where $\gamma \in \Gamma = \mathbb{Z} \oplus \mathbb{Z}\tau$. We can divide P into two triangles that are isometric to T, and into 8 smaller triangles similar to T by the scaling factor 2 (see Figure 3.8).

As indicated in Figure 3.8, there is a map $\widetilde{\Theta} \colon P \to \Delta$ such that each small triangle in P is mapped isometrically to a face of Δ. This map extends naturally to a map $\widetilde{\Theta} \colon \mathbb{C} \to \Delta$ that respects the lattice translations in the sense that $\widetilde{\Theta} \circ g = \widetilde{\Theta}$ for all $g \in G_{\mathrm{tr}}$. Note that then $\widetilde{\Theta}(z) = \widetilde{\Theta}(-z)$ for $z \in \mathbb{C}$, and it easily follows that $\widetilde{\Theta}$ is induced by G. So we can identify $\mathbb{C}/G$ with the tetrahedron Δ. In fact, $\mathbb{C}/G$ equipped with the canonical orbifold metric (see Section A.10) is isometric to (Δ, ω) (up to scaling).
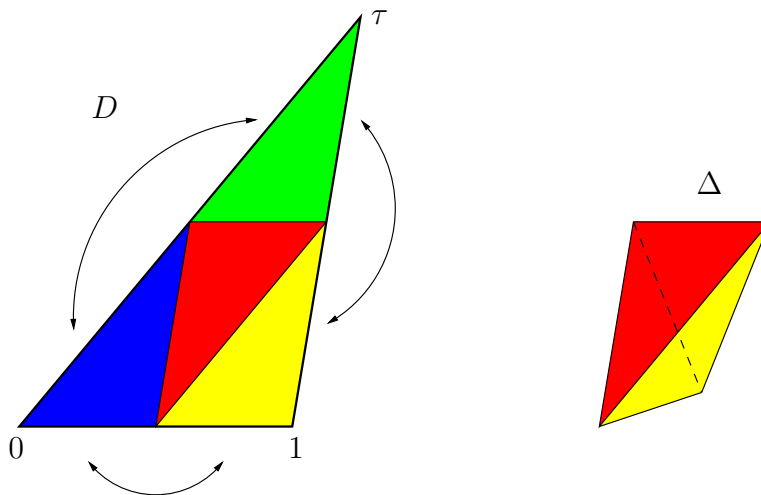
FIGURE 3.7. Folding a tetrahedron from a triangle.
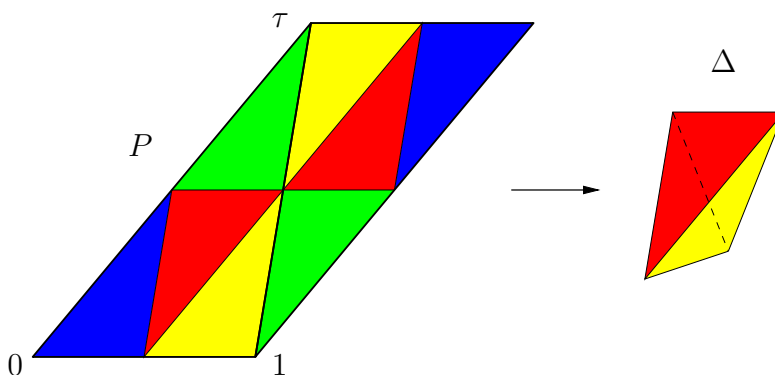
FIGURE 3.8. Construction of $\Theta = \wp$.

As a polyhedral surface Δ carries a natural conformal structure. By the uniformization theorem there is a conformal map $\varphi \colon \Delta \to \widehat{\mathbb{C}}$. It is uniquely determined up to postcomposition by a Möbius transformation. Then $\Theta = \varphi \circ \widetilde{\Theta} \colon \mathbb{C} \to \widehat{\mathbb{C}}$ is a holomorphic map induced by G. If we choose a suitably normalized map $\varphi = \varphi_0$ here, then $\Theta = \varphi_0 \circ \widetilde{\Theta} = \wp(\,\cdot\,; \Gamma)$. Moreover, it follows from our analytic discussion above that if we set $\varphi = \psi \circ \varphi_0$ for a suitable Möbius transformation ψ and choose $\tau \in \mathbb{C}$ with $\operatorname{Im}(\tau) > 0$ appropriately, then a universal orbifold covering map Θ of any orbifold on $\widehat{\mathbb{C}}$ with signature $(2, 2, 2, 2)$ can be written as $\Theta = \varphi \circ \widetilde{\Theta}$.

Lattès maps with orbifold signature $(2, 2, 2, 2)$ are obtained from G-equivariant maps A as in Theorem 3.1 (ii). The conditions on A specified in Proposition 3.14 are quite restrictive. For generic $\tau \in \mathbb{C}$

with $\mathrm{Im}(\tau) > 0$ there are no $\alpha \in \mathbb{C} \setminus \mathbb{Z}$ with $\alpha\Gamma \subset \Gamma$. More precisely, such $\alpha \in \mathbb{C} \setminus \mathbb{Z}$ exists if and only if the lattice allows so-called complex multiplication and α is an integer in an imaginary quadratic field. On the other hand, if $\alpha \in \mathbb{Z} \setminus \{0\}$, then $\alpha\Gamma \subset \Gamma$ for each lattice Γ. This leads to a class of Lattès maps given by the following definition.

DEFINITION 3.26 (Flexible Lattès map). A Lattès map $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is called *flexible* if its orbifold has signature $(2, 2, 2, 2)$ and can be represented as in Theorem 3.1 (ii) with a crystallographic group $G = \widetilde{G}$ of type (2222) as in Theorem 3.7 and a map $A \colon \mathbb{C} \to \mathbb{C}$ of the form $A(z) = \alpha z + \beta$ with $\alpha \in \mathbb{Z} \setminus \{-1, 0, 1\}$ and $2\beta \in \Gamma$, where Γ is the underlying lattice of G.

Note that we have to rule out $\alpha = \pm 1$ due to the requirement that $\deg(f) \geq 2$ for a Thurston map f.

The term "flexible" derives from the fact that by deforming the underlying lattice Γ (and possibly the parameter β of A) of a flexible Lattès map, one obtains a family of such maps depending on one complex parameter (for example, the parameter τ in the representation of the lattice $\Gamma = \mathbb{Z} \oplus \mathbb{Z}\tau$). It is not hard to see that this leads to maps that are topologically conjugate, but in general not conjugate by a Möbius transformation on $\widehat{\mathbb{C}}$ (see the discussion below). An explicit example of such a family (with $A(z) = 2z$ and $\beta = 0$) is given by

$$(3.38) \qquad\qquad f(z) = \frac{4z(1-z)(1-k^2 z)}{(1-k^2 z^2)^2},$$

where $k \in \mathbb{C} \setminus \{0, 1, -1\}$. The map g discussed in Section 1.1 corresponds to $k = i$.

Flexible Lattès maps are exceptional in many respects. For example, Thurston's uniqueness theorem (see Theorem 2.19) fails for these maps. They also carry invariant line fields (see [**McM94a**, Chapter 3.5]). According to a well-known conjecture they are the only rational maps with this property. Its validity would imply the "density of hyperbolic rational maps", which is possibly the most famous open problem in complex dynamics (see [**MSS83**] and [**McM94b**] for an overview).

Flexible Lattès maps can be considered as generic Lattès maps, because they are the only type of Lattès maps that can be defined for arbitrary lattices $\Gamma = \mathbb{Z} \oplus \mathbb{Z}\tau$ or, equivalently, can be obtained as quotients of maps on arbitrary complex tori $\mathbb{T}$. This is the reason why some authors use the term "Lattès map" only for maps with orbifold signature $(2, 2, 2, 2)$. For example, Milnor uses this more restrictive
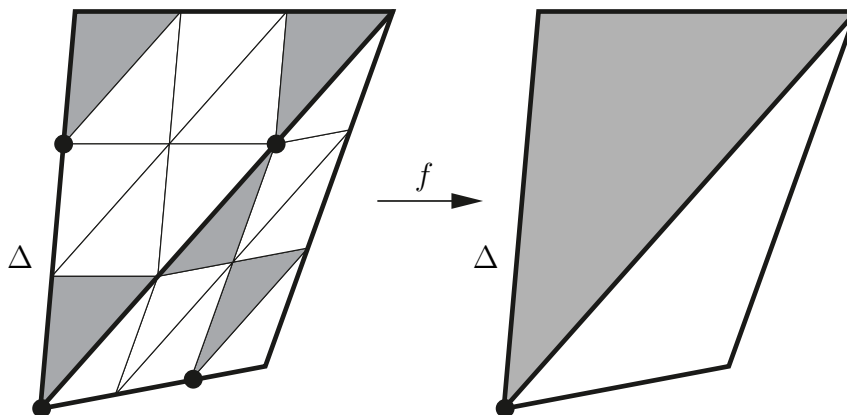
FIGURE 3.9. A Euclidean model for a flexible Lattès map.

definition in the second edition (and earlier editions), while he uses the definition used here in the third edition of [**Mi06a**].

To discuss a geometric description of flexible Lattès maps, suppose f is such a map as in Definition 3.26 obtained from a map $A \colon \mathbb{C} \to \mathbb{C}$ of the form $A(z) = nz + \beta$ with $n \in \mathbb{Z} \setminus \{-1, 0, 1\}$ and $2\beta \in \Gamma = \mathbb{Z} \oplus \mathbb{Z}\tau$. We get the same map f if we replace A by $g \circ A$, where $g \in G$. This allows us to assume that A has the form $A(z) = nz + \beta$, where $n \in \mathbb{N}$, $n \geq 2$, and $\beta \in \{0, 1/2, \tau/2, (\tau+1)/2\}$ (see Proposition 3.15). Note that the map $\widetilde{\Theta}$ considered above sends the four points $0, 1/2, \tau/2, (\tau+1)/2$ to the four vertices of the tetrahedron Δ.

The triangle T from Figure 3.7 can be divided into n^2 triangles that are similar to T by the scaling factor n. For this we divide each side of Δ into n edges of the same length and draw the line segments parallel to the sides of T through the endpoints of these edges. The tetrahedron Δ has four faces, denoted by T_1, T_2, T_3, T_4. Each of them is similar to T, and so we can divide each face of Δ in the same way into n^2 triangles similar to T. In the following, we will call them the *small* triangles in Δ.

We now construct a map $\widetilde{f} \colon \Delta \to \Delta$ as follows. We consider a small triangle S that contains one of the four vertices of Δ and send S to one of the faces T' of Δ by an orientation-preserving similarity that scales by the factor n. For given S and T' there is only one such map unless the base triangle T, and hence also S and T', are equilateral. In this case, there are three such maps, only one of which will lead to a flexible Lattès map. We will ignore this special case in order to keep our discussion simple.
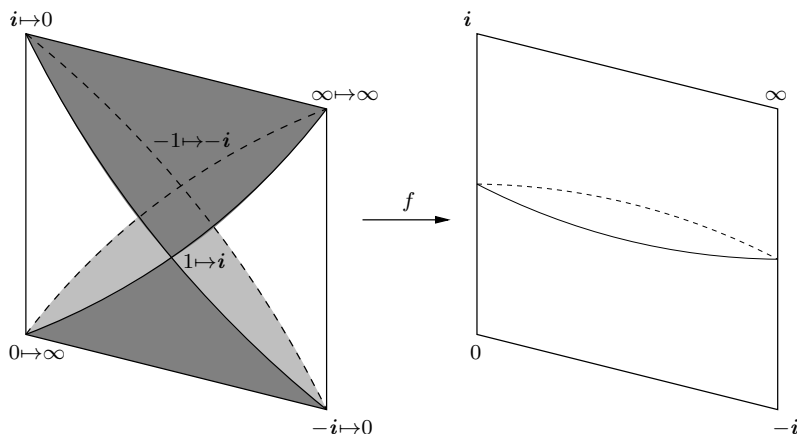
FIGURE 3.10. The map f in Example 3.27.

The similarity between S and T' can be uniquely extended to a continuous map $\widetilde{f}$ on Δ that sends each small triangle to one of the triangles T_j, $j \in \{1, 2, 3, 4\}$, by an orientation-preserving similarity with scaling factor n. It is clear that $\widetilde{f}$ is a branched covering map. Its critical points are the vertices of the small triangles with the exception of the four vertices of Δ. The map $\widetilde{f}$ sends each vertex of a small triangle to one of the vertices of Δ. It follows that the postcritical points of $\widetilde{f}$ are the four vertices of Δ and $\widetilde{f}$ is postcritically-finite. Since the local degree at each critical point is 2, it is easy to see that the orbifold signature of $\widetilde{f}$ is $(2, 2, 2, 2)$.

If we conjugate $\widetilde{f}$ by a uniformizing map $\varphi \colon \Delta \to \widehat{\mathbb{C}}$, then we obtain a flexible Lattès map $f \coloneqq \varphi \circ \widetilde{f} \circ \varphi^{-1} \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$. Conversely, each flexible Lattès map can be obtained in this form. One such map (with $n = 3$) is illustrated in Figure 3.9 (note that here the tetrahedron Δ is embedded in $\mathbb{R}^3$).

If we use exactly the same construction as just described with another parameter $\tau' \in \mathbb{C}$, $\operatorname{Im}(\tau') > 0$, then we obtain a different tetrahedron Δ'. This leads to a different flexible Lattès map $g \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$. It is easy to see that the f and g are always topologically conjugate, but in general not conjugate by a Möbius transformation. So we obtain a 1-parameter family of Lattès maps with similar dynamics.

EXAMPLE 3.27. Not all Lattès maps whose orbifolds have signature $(2, 2, 2, 2)$ are flexible Lattès maps. For example, suppose G is the crystallographic group consisting of the maps $g(z) = \pm z + \gamma$, where $\gamma \in \Gamma \coloneqq \mathbb{Z} \oplus \mathbb{Z}i$. Then $A(z) = (1 - i)z$ is G-equivariant and so there is a Lattès map f that arises as the quotient of A by G as in Theorem 3.1 (ii). Then the map f has orbifold signature $(2, 2, 2, 2)$.

In fact, f is (up to conjugation by a Möbius transformation) given by $f(z) = \frac{i}{2}(z + 1/z)$. Since $\deg(f) = 2$, this map is not a flexible Lattès map, because the degree of each such map is the square of an integer. A geometric model for f is shown in Figure 3.10. A detailed discussion of this map can be found in [**Mi04**].

CHAPTER 4

Quasiconformal and rough geometry

The immediate purpose of this chapter is to provide a framework for discussing Cannon's conjecture in geometric group theory and related questions. This will give some motivating background for our study of expanding Thurston maps. The chapter will also serve as a reference for some geometric terminology and facts that we will use throughout this work.

4.1. Quasiconformal geometry

In this section we record some material related to quasiconformal geometry and the analysis on metric spaces (see [**He01**] for an exposition of this subject).

Let (X, d_X) and (Y, d_Y) be metric spaces, and $f \colon X \to Y$ be a homeomorphism. Then f is called *bi-Lipschitz* if there exists a constant $L \geq 1$ such that

$$\frac{1}{L} d_X(u, v) \leq d_Y(f(u), f(v)) \leq L d_X(u, v)$$

for all $u, v \in X$. If there exists $\alpha > 0$ and $L \geq 1$ such that

(4.1) $$\frac{1}{L} d_X(u, v)^\alpha \leq d_Y(f(u), f(v)) \leq L d_X(u, v)^\alpha$$

for all $u, v \in X$, then f is called a *snowflake homeomorphism*. The map f is called a *quasisymmetric homeomorphism* or a *quasisymmetry* if there exists a homeomorphism $\eta \colon [0, \infty) \to [0, \infty)$ such that

(4.2) $$\frac{d_Y(f(u), f(v))}{d_Y(f(u), f(w))} \leq \eta \left(\frac{d_X(u, v)}{d_X(u, w)} \right)$$

for all $u, v, w \in X$ with $u \neq w$. If we want to emphasize η here, then we speak of an *η-quasisymmetry*.

The inverse of a bi-Lipschitz homeomorphism and the composition of two such maps (when it is defined) is again a bi-Lipschitz homeomorphism. Similarly, the classes of snowflake homeomorphisms and quasisymmetries are closed under taking inverses or compositions. This implies that all these classes of maps lead to a corresponding notion of equivalence for metric spaces. So X and Y are called *bi-Lipschitz*,

snowflake, or *quasisymmetrically equivalent* if there exists a homeo-
morphism between X and Y that is of the corresponding type. In
the same way, we call two metrics d and d' on a space X *bi-Lipschitz*,
snowflake, or *quasisymmetrically equivalent* if the identity map from
(X, d) to (X, d') is a homeomorphism with the corresponding property.
Since every bi-Lipschitz homeomorphism is also a snowflake homeo-
morphism and every snowflake homeomorphism is a quasisymmetry,
this leads to correspondingly weaker notions of equivalence.

A collection of metrics on a given space X that are mutually qua-
sisymmetrically equivalent is called a *quasisymmetric gauge* on X. One
defines a *bi-Lipschitz* or *snowflake gauge* on X similarly (see [**He01**,
Chapter 15] for related terminology and further discussion).

As we will later see in Chapter 8, each expanding Thurston map
$f \colon S^2 \to S^2$ induces a natural snowflake gauge on its underlying 2-
sphere S^2. This is related to the fact that if X is a Gromov hyperbolic
space, then there is a natural snowflake gauge on its boundary at in-
finity $\partial_\infty X$ (see Section 4.2 below). In both cases we will later use
the term *visual metric* for a metric in the natural snowflake gauge.
A visual metric ϱ gives the underlying 2-sphere S^2 of an expanding
Thurston map $f \colon S^2 \to S^2$ some geometric structure related to the dy-
namics of the map. An important question in this context is whether
this sphere (S^2, ϱ) is quasisymmetrically equivalent to the standard 2-
sphere, i.e., the Riemann sphere equipped with the chordal metric (see
Theorem 18.1). This is an instance of a more general problem that can
be formulated as follows.

THE QUASISYMMETRIC UNIFORMIZATION PROBLEM. *Suppose X
is a metric space homeomorphic to some "standard" metric space Y.
When is X quasisymmetrically equivalent to Y?*

This problem is very similar to questions in classical uniformization
theory where one asks whether a given region in $\widehat{\mathbb{C}}$ or a given Riemann
surface is conformally equivalent to a "standard" region or Riemann
surface. The quasisymmetric uniformization problem can be seen as
a metric space version of this where the class of conformal maps is
replaced by the class of quasisymmetric homeomorphisms.

It depends on the context how the term "standard space Y" is
precisely interpreted. A satisfactory answer to the quasisymmetric
uniformization problem for given Y is essentially equivalent to cha-
racterizing Y up to quasisymmetric equivalence. We will see that this
is related to Cannon's conjecture in geometric group theory, for exam-
ple. One can also pose a similar problem for other classes of maps such
as bi-Lipschitz or snowflake maps.

The prime instance for a quasisymmetric uniformization result is a theorem due to Tukia and Väisälä that characterizes *quasicircles* and *quasiarcs*. In order to state this result, we first have to discuss some terminology.

The *standard unit circle* $\mathbb{S}^1$ is the unit circle in $\mathbb{R}^2$ equipped with the restriction of the Euclidean metric. A *metric circle*, i.e., a metric space homeomorphic to $\mathbb{S}^1$, is called a *quasicircle* if it is quasisymmetrically equivalent to $\mathbb{S}^1$. Similarly, a *metric arc* (a metric space homeomorphic to the unit interval $[0, 1]$) is called a *quasiarc* if it is quasisymmetrically equivalent to $[0, 1]$.

A *continuum* is a compact and connected topological space. If a continuum contains two distinct points, then it is called *non-degenerate*. If x and y are points in a continuum X, then we say that X *connects* or *joins* x and y.

Let (X, d) be a metric space. Then X is called *doubling* if there exists a number $N \in \mathbb{N}$ such that every ball in X of radius $R > 0$ can be covered by N balls of radius $R/2$. We say that X is of *bounded turning* if there is a constant $K \geq 1$ such that for all points $x, y \in X$ there exists a continuum $\gamma \subset X$ with $x, y \in \gamma$ and

$$(4.3) \qquad \operatorname{diam}_d(\gamma) \leq K d(x, y).$$

If X is a metric circle, then it is of bounded turning precisely when the last inequality is always true for one of the subarcs $\gamma \subset X$ with endpoints x and y. Similarly, if X is a metric arc, then it is of bounded turning precisely when (4.3) is always true for the unique subarc γ of X with endpoints x and y.

The Tukia-Väisälä theorem [**TV80**, Theorem 4.9, p. 113] can now be formulated as follows.

THEOREM 4.1. *Let X be a metric circle or a metric arc. Then X is a quasicircle or a quasiarc, respectively, if and only if X is doubling and of bounded turning.*

A metric space X is called *linearly locally connected* if there exists a constant $\lambda \geq 1$ satisfying the following conditions: If $B(a, r)$ is an open ball in X and $x, y \in B(a, r)$, then there exists a continuum in $B(a, \lambda r)$ connecting x and y. Moreover, if $x, y \in X \setminus B(a, r)$, then there exists a continuum in $X \setminus B(a, r/\lambda)$ connecting x and y.

It is easy to see that a metric circle is of bounded turning if and only if it is linearly locally connected. This gives an alternative version of part of the Tukia-Väisälä theorem: *A metric circle is a quasicircle if and only if it is doubling and linearly locally connected.*

Every subset of $\widehat{\mathbb{C}}$ (equipped with the restriction of the chordal metric) is doubling; this implies that a Jordan curve $J \subset \widehat{\mathbb{C}}$ is a quasicircle if and only if it is of bounded turning. A similar remark applies to arcs $\alpha \subset \widehat{\mathbb{C}}$.

A quasisymmetric characterization of the standard $1/3$-Cantor set can be found in [**DS97**]. Work by Semmes [**Se96a, Se96b**] shows that the quasisymmetric characterization of $\mathbb{R}^n$ or the standard sphere $\mathbb{S}^n$ for $n \geq 3$ is a problem that seems to be beyond reach at the moment. The intermediate case $n = 2$ is particularly interesting. To formulate a specific result in this direction we need one more definition.

Let (X, d) be a locally compact metric space, and μ a Borel measure on X. Then the metric measure space (X, d, μ) is called *Ahlfors Q-regular*, where $Q > 0$, if

$$(4.4) \qquad \frac{1}{C} R^Q \leq \mu(\overline{B}(x, R)) \leq C R^Q$$

for all closed balls $\overline{B}(x, R)$ with $x \in X$ and $0 < R \leq \mathrm{diam}(X)$, where $C \geq 1$ is independent of the ball. If (X, d) is understood, we also say that μ is Ahlfors Q-regular.

If this condition is satisfied, then the Q-dimensional Hausdorff measure $\mathcal{H}^Q$ actually satisfies an inequality as in (4.4), and μ and $\mathcal{H}^Q$ are comparable and in particular mutually absolutely continuous with respect to each other (for the definition of Hausdorff Q-measure $\mathcal{H}^Q$ see [**He01**], for example). If we want to emphasize the underlying metric d, we use the notation $\mathcal{H}^Q_d$ for the Hausdorff Q-measure. Note that every Ahlfors regular measure μ on (X, d) is *doubling*, i.e., there exists a constant $C \geq 1$ such that

$$\mu(\overline{B}(x, 2R)) \leq C\mu(\overline{B}(x, R))$$

whenever $x \in X$ and $R > 0$.

A locally compact metric space (X, d) is called *Ahlfors Q-regular* for $Q > 0$ if the Q-dimensional Hausdorff measure $\mathcal{H}^Q$ has this property. Since every Ahlfors regular space admits a doubling measure, it is doubling (as a metric space) (see [**Se99**, B.3.4 Lemma, p. 412] for this last implication).

For a metric 2-sphere X to be quasisymmetrically equivalent to the standard 2-sphere $(\widehat{\mathbb{C}}, \sigma)$ it is necessary that X is linearly locally connected. This alone is not sufficient, but will be if we add Ahlfors 2-regularity as an assumption [**BK02**, Theorem 1.1].

THEOREM 4.2. *Suppose X is a metric space homeomorphic to $\widehat{\mathbb{C}}$. If X is linearly locally connected and Ahlfors 2-regular, then X is quasisymmetrically equivalent to $(\widehat{\mathbb{C}}, \sigma)$.*

A similar result for other simply connected surfaces has been obtained by K. Wildrick [**Wi08**].

The assumption of Ahlfors regularity for some exponent $Q \geq 2$ is quite natural, because it is satisfied in many interesting cases: for example, for boundaries of Gromov hyperbolic groups (see [**Co93**]) or for 2-spheres equipped with visual metrics of an expanding Thurston map (under some mild extra conditions; see Proposition 18.2). There are metric 2-spheres X though that are linearly locally connected and Q-regular with $Q > 2$, but are not quasisymmetrically equivalent to $(\widehat{\mathbb{C}}, \sigma)$ [**Vä88**].

We will now prove a fact that will be useful in Chapter 20, where Theorem 4.2 is applied to give a characterization of Lattès maps.

PROPOSITION 4.3. *Let d be a metric on $\widehat{\mathbb{C}}$ that is quasisymmetrically equivalent to the chordal metric σ. Let μ be a Borel measure on $\widehat{\mathbb{C}}$ such that $(\widehat{\mathbb{C}}, d, \mu)$ is Ahlfors 2-regular. Then μ and the Lebesgue measure $\mathcal{L}$ on $\widehat{\mathbb{C}}$ are absolutely continuous with respect to each other.*

PROOF. There is no simple justification of this fact. Accordingly, we have to rely on concepts and statements that we will not explain in detail, but rather refer to the literature.

The quasisymmetric equivalence of σ and d implies that a metric ball with respect to σ roughly looks like a metric ball with respect to d (typically with quite different radius); more precisely, there exists a constant $\lambda \geq 1$ such that for all $p \in \widehat{\mathbb{C}}$ and all $R > 0$ there exists $R' > 0$ such that

$$B_d(p, R'/\lambda) \subset B_\sigma(p, R) \subset B_d(p, \lambda R').$$

One can deduce from this and the Ahlfors 2-regularity of $(\widehat{\mathbb{C}}, d, \mu)$ that μ is a doubling measure on $(\widehat{\mathbb{C}}, \sigma)$. Actually, μ is a *metric doubling measure* on $(\widehat{\mathbb{C}}, \sigma)$: if $x, y \in \widehat{\mathbb{C}}$ are arbitrary, $R := \sigma(x, y)$, and we set

$$d'(x, y) := \mu(B_\sigma(x, R) \cup B_\sigma(y, R))^{1/2},$$

then d' is comparable to a metric (in fact to d, meaning that $d' \asymp d$). See [**He01**] for the terminology employed here.

It is known that a metric doubling measure μ on $(\widehat{\mathbb{C}}, \sigma)$ is absolutely continuous with respect to Lebesgue measure $\mathcal{L}$ on $\widehat{\mathbb{C}}$ with a Radon-Nikodym derivative w that is positive almost everywhere (in fact w is a so-called A_∞-weight; see [**Se93**]). Then $d\mu = w \, d\mathcal{L}$. Since w does

not vanish on a set of positive Lebesgue measure, this implies that $\mathcal{L}$ is also absolutely continuous with respect to μ. $\square$

A homeomorphism $f\colon X \to Y$ between metric spaces (X, d_X) and (Y, d_Y) is called *weakly quasisymmetric* if there exists a constant $H \geq 1$ such that for all $u, v, w \in X$ the following implication holds:

$$d_X(u, v) \leq d_X(u, w) \Rightarrow d_Y(f(u), f(v)) \leq H d_Y(f(u), f(w)).$$

Under mild extra assumptions on the spaces, a weak quasisymmetry is in fact a quasisymmetry ([**He01**, Theorem 10.19]).

PROPOSITION 4.4. *Let (X, d_X) and (Y, d_Y) be metric spaces, and $f\colon X \to Y$ be a weakly quasisymmetric homeomorphism. If X and Y are connected and doubling, then f is a quasisymmetry.*

This proposition is very useful if one wants to establish that a given map is a quasisymmetry.

Let U and V be open regions in $\widehat{\mathbb{C}}$. A continuous map $f\colon U \to V$ is called *quasiregular* if f belongs to the Sobolev space $W^{1,2}_{loc}$ (i.e., f has weak partial derivatives in L^2_{loc} on U) and if there exists $K \geq 1$ such that

$$\|Df(p)\|^2_\sigma \leq K \det(Df(p))$$

for almost every $p \in \widehat{\mathbb{C}}$. Here $Df(p)$ denotes the (formal) derivative of f at p considered as a linear map between tangent spaces and $\|Df(p)\|_\sigma$ denotes the operator norm of $Df(p)$ with respect to the spherical metric (see Section A.1). If, in addition, f is a homeomorphism between U and V, then f is called *quasiconformal*. If we want to emphasize the parameter K, then we call such a map K-*quasiconformal* or K-*quasiregular*. For more background on quasiconformal and quasiregular maps see [**Vä71**] and [**Ri93**].

If one considers a family $\mathcal{F}$ of maps in one of the classes that we discussed, then it is often important to know whether the distortion properties of the maps are controlled by the same parameters. If this is the case, we say that $\mathcal{F}$ is a *uniform* family of maps in the given class. For example, a family $\mathcal{F}$ of homeomorphisms is said to consist of *uniform quasisymmetries* if there exists a homeomorphism $\eta\colon [0, \infty) \to [0, \infty)$ such that each map in $\mathcal{F}$ is an η-quasisymmetry. Similarly, we say that the maps in $\mathcal{F}$ are *uniformly quasiregular* if there exists $K \geq 1$ such that each $f \in \mathcal{F}$ is K-quasiregular, etc.

We conclude this section with a discussion of Hausdorff distance and Hausdorff convergence. Let (X, d) be a metric space. If $A, B \subset X$

are subsets of X, then their *Hausdorff distance* is defined as

(4.5) $\mathrm{dist}_d^H(A, B) =$
$$\inf\{\delta > 0 : A \subset \mathcal{N}_\delta(B) \text{ and } B \subset \mathcal{N}_\delta(A)\} \in [0, \infty].$$

Here for $\delta > 0$,
$$\mathcal{N}_\delta(M) = \mathcal{N}_{d,\delta}(M) = \{x \in X : \mathrm{dist}_d(x, M) < \delta\}$$

denotes the δ-neighborhood of a set $M \subset X$. If A and A_n for $n \in \mathbb{N}$ are non-empty closed subsets of X, then we say that $A_n \to A$ *in the sense of Hausdorff convergence* or A_n *Hausdorff converges* to A as $n \to \infty$ if
$$\lim_{n \to \infty} \mathrm{dist}_d^H(A_n, A) = 0.$$

Note that in this case a point $x \in X$ lies in A if and only if there exists a sequence $\{x_n\}$ of points in X such that $x_n \in A_n$ for $n \in \mathbb{N}$ and $x_n \to x$ as $n \to \infty$.

It is known that if X is compact, then the space of all non-empty closed subsets of X equipped with the Hausdorff distance is a complete metric space (see [**BBI01**, Section 7.3.1]).

4.2. Gromov hyperbolicity

In this section we review some standard material on Gromov hyperbolic spaces. For general background on this topic see [**BS07, GH90, Gr87**].

Let (X, d) be a metric space. Then for $p, x, y \in X$ the quantity

(4.6) $$(x \cdot y)_p := \frac{1}{2}\big(d(x, p) + d(y, p) - d(x, y)\big)$$

is called the *Gromov product* of x and y with respect to the basepoint p. The space X is called δ-*hyperbolic* for $\delta \geq 0$, if the inequality

(4.7) $$(x \cdot z)_p \geq \min\{(x \cdot y)_p, (y \cdot z)_p\} - \delta$$

holds for all $x, y, z, p \in X$. If this condition is true for some basepoint $p \in X$ (and all $x, y, z \in X$), then it is actually true for all basepoints if one changes the constant δ to 2δ. We say that X is *Gromov hyperbolic* if X is δ-hyperbolic for some $\delta \geq 0$.

The space X is called *geodesic* if any two points in X can be joined by a path whose length is equal to the distance of the points. For a geodesic metric space Gromov hyperbolicity is equivalent to a thinness condition for geodesic triangles [**GH90**].

Roughly speaking, the Gromov hyperbolicity of a space requires it to be "negatively curved" on large scales. Examples for such spaces are

simplicial trees, or Cartan-Hadamard manifolds with a negative upper curvature bound such as the real hyperbolic spaces $\mathbb{H}^n$, $n \geq 2$.

Let (X, d_X) and (Y, d_Y) be metric spaces. A map $f \colon X \to Y$ is called a *quasi-isometry* if there exist constants $\lambda \geq 1$ and $k \geq 0$ such that

$$(4.8) \qquad \frac{1}{\lambda} d_X(u, v) - k \leq d_Y(f(u), f(v)) \leq \lambda d_X(u, v) + k$$

for all $u, v \in X$ and if

$$(4.9) \qquad \inf_{x \in X} d_Y(f(x), y) \leq k$$

for all $y \in Y$. If $\lambda = 1$, then we call f a *rough-isometry*. The spaces X and Y are called *quasi-isometric* or *rough-isometric* if there exists a map $f \colon X \to Y$ that is a quasi-isometry or a rough-isometry, respectively. In coarse geometry one often considers two metric spaces the same if they are quasi-isometric or rough-isometric.

Quasi-isometries form a natural class of maps in the theory of Gromov hyperbolic spaces. For example, Gromov hyperbolicity of geodesic metric spaces is invariant under quasi-isometries [**GH90**].

A subset A of a metric space (X, d) is called *cobounded* if there exists a constant $k \geq 0$ such that for every $x \in X$ there exists $a \in A$ with $d(a, x) \leq k$. Then every point in X lies within uniformly bounded distance of the set A. With this terminology, the condition (4.9) can be reformulated by saying that the map f has cobounded image in Y. If A is cobounded in X, then X is Gromov hyperbolic if and only if A (equipped with the restriction of the ambient metric) is Gromov hyperbolic.

With each Gromov hyperbolic space X one can associate a *boundary at infinity* $\partial_\infty X$ as follows. We fix a basepoint $p \in X$, and consider sequences of points $\{x_i\}$ in X *converging to infinity* in the sense that

$$(4.10) \qquad \lim_{i,j \to \infty} (x_i \cdot x_j)_p = \infty.$$

We declare two such sequences $\{x_i\}$ and $\{y_i\}$ in X as *equivalent* if

$$(4.11) \qquad \lim_{i \to \infty} (x_i \cdot y_i)_p = \infty.$$

Then $\partial_\infty X$ is defined as the set of equivalence classes of sequences converging to infinity. It is easy to see that the choice of the basepoint p does not matter here. Moreover, if A is cobounded in X, then one can represent each equivalence class by a sequence in A, and so we have a natural identification $\partial_\infty X \cong \partial_\infty A$.

A metric space X is called *proper* if every closed ball in X is compact. If a Gromov hyperbolic space X is proper and geodesic, then

there is an equivalent definition of $\partial_\infty X$ as the set of equivalence classes of geodesic rays emanating from the basepoint p. One declares two such rays as equivalent if they stay within bounded Hausdorff distance. Intuitively, a ray represents its "endpoint" in $\partial_\infty X$ (see [**BS07**] for the details).

The Gromov product on a Gromov hyperbolic space X has a natural extension to the boundary $\partial_\infty X$. Namely, if $p \in X$ and $a, b \in \partial_\infty X$, we set

$$(4.12) \qquad (a \cdot b)_p := \inf \big\{ \liminf_{i \to \infty} (x_i \cdot y_i)_p : \{x_i\} \in a, \ \{y_i\} \in b \big\}.$$

Note that by definition a point in $\partial_\infty X$ is an equivalence class (i.e., a set) of sequences in X. So in (4.12) it makes sense to take the infimum over all sequences $\{x_i\}$ and $\{y_i\}$ that represent (i.e., are contained in) a and b, respectively. We have $(a \cdot b)_p \in [0, \infty]$, where $(a \cdot b)_p = \infty$ if and only if $a = b$.

In (4.12) one can actually use any sequences representing the points a and b to determine $(a \cdot b)_p$ (up to an irrelevant additive constant). Namely, there exists a constant $k \geq 0$ independent of a and b such that for all $\{x_i\} \in a$ and $\{y_i\} \in b$ we have

$$(4.13) \qquad \liminf_{i \to \infty} (x_i \cdot y_i)_p - k \leq (a \cdot b)_p \leq \liminf_{i \to \infty} (x_i \cdot y_i)_p.$$

The boundary $\partial_\infty X$ is equipped with a natural class of visual metrics. By definition a metric ϱ on $\partial_\infty X$ is called *visual* if there exists a basepoint $p \in X$, and constants $C \geq 1$ and $\Lambda > 1$ such that

$$(4.14) \qquad \frac{1}{C} \Lambda^{-(a \cdot b)_p} \leq \varrho(a, b) \leq C \Lambda^{-(a \cdot b)_p}$$

for all $a, b \in \partial_\infty X$ (here we use the convention that $\Lambda^{-\infty} = 0$). We call Λ the *visual parameter* of ϱ. If X is δ-hyperbolic, then there exists a visual metric ϱ for each visual parameter $\Lambda > 1$ sufficiently close to 1.

Later we will also define a notion of a visual metric for an expanding Thurston map. So then we have two notions of visual metrics—one for expanding Thurston maps and one for boundaries of Gromov hyperbolic spaces. For clarity we will sometimes use the phrases *visual metric in the sense of Thurston maps* and *visual metric in the sense of Gromov hyperbolic spaces* to distinguish between these two notions. As we will see in Chapter 10, with each expanding Thurston map $f \colon S^2 \to S^2$ one can associate a Gromov hyperbolic space $\mathcal{G}$ such that $\partial_\infty \mathcal{G}$ can be identified with S^2 and such that the class for visual metric of f (in the sense of Thurston maps) is exactly the same as the class for visual metric on $\partial_\infty \mathcal{G} \cong S^2$ (in the sense of Gromov hyperbolic spaces), see

Theorem 10.2. This fact was actually the reason for our choice of the term "visual metric" for expanding Thurston maps.

We can think of the boundary at infinity of a Gromov hyperbolic space $\partial_\infty X$ as a metric space if we equip it with a fixed visual metric. If ϱ_1 and ϱ_2 are two visual metrics on $\partial_\infty X$, then the identity map is a snowflake equivalence between $(\partial_\infty X, \varrho_1)$ and $(\partial_\infty X, \varrho_2)$. So the visual metrics form a snowflake gauge on $\partial_\infty X$. In particular, $\partial_\infty X$ carries a well-defined topology induced by any visual metric, and the ambiguity of the visual metric is irrelevant if one wants to speak of snowflake or quasisymmetric maps on $\partial_\infty X$. One should consider the space $\partial_\infty X$ equipped with such a visual metric ϱ as being very "fractal". For example, assume there exists a visual metric ϱ_0 with visual parameter $\Lambda_0 > 1$. Then for every visual metric ϱ with a visual parameter Λ that satisfies $1 < \Lambda < \Lambda_0$, the space $(\partial_\infty X, \varrho)$ does not contain any non-constant rectifiable curves.

The following fact links the theory of Gromov hyperbolic spaces to quasisymmetric maps (see [**BS00**] for more on this subject).

PROPOSITION 4.5. *Let X and Y be proper and geodesic Gromov hyperbolic spaces. Then every quasi-isometry $f\colon X \to Y$ induces a natural quasisymmetric boundary map $\tilde{f}\colon \partial_\infty X \to \partial_\infty Y$.*

The boundary map $\tilde{f}$ is defined by assigning to a point $a \in \partial_\infty X$ represented by the sequence $\{x_i\}$ the point $b \in \partial_\infty Y$ represented by the sequence $\{f(x_i)\}$.

This statement lies at the heart of Mostow's proof for rigidity of rank-one symmetric spaces [**Mo78**]. The point is that a quasi-isometry may locally exhibit very irregular behavior, but gives rise to a quasisymmetric boundary map that can be analyzed by analytic tools.

4.3. Gromov hyperbolic groups and Cannon's conjecture

The theory of Gromov hyperbolic spaces can be used to define a class of discrete groups. Here one adopts a geometric point of view by studying the Cayley graph of the group. We review some standard definitions related to this, but will not attempt an in-depth treatment of the subject (for more details, see [**GH90**]). We will just develop the necessary background to state and discuss Cannon's conjecture that served as one of our motivations for studying expanding Thurston maps.

Let G be a finitely generated group, and S a finite set of generators of G that is *symmetric*, i.e., if s is in S, then its inverse s^{-1} is also in S. The *Cayley graph* $\mathcal{G}(G, S)$ of G with respect to S is now defined

as follows: the group elements are the vertices of G, and one joins two vertices given by $g, h \in G$ by an edge, if there exists $s \in S$ such that $g = hs$ (here we use the common convention that juxtaposition of group elements means their composition in the group). Since S is symmetric, this "edge relation" for vertices is also symmetric, and so we consider edges as undirected. If we identify each edge in $\mathcal{G}(G, S)$ with a closed interval of length 1, then $\mathcal{G}(G, S)$ becomes a cell complex, where singleton sets consisting of group elements are the cells of dimension 0 and the edges are the cells of dimension 1. The graph $\mathcal{G}(G, S)$ is connected and carries a unique path metric so that each edge is isometric to the unit interval $[0, 1]$. In the following, we always consider $\mathcal{G}(G, S)$ as a metric space equipped with this path metric. Then $\mathcal{G}(G, S)$ is proper and geodesic.

The group G is called *Gromov hyperbolic* if the metric space $\mathcal{G}(G, S)$ is Gromov hyperbolic for some (finite and symmetric) set S of generators of G. If this is the case, then $\mathcal{G}(G, S')$ is Gromov hyperbolic for all generating sets S'. This essentially follows from the fact that $\mathcal{G}(G, S)$ and $\mathcal{G}(G, S')$ are quasi-isometric.

Examples of Gromov hyperbolic groups are free groups, fundamental groups of compact negatively curved manifolds, or small cancellation groups.

If G is a Gromov hyperbolic group, then one defines its boundary at infinity as $\partial_\infty G = \partial_\infty \mathcal{G}(G, S)$. A priori this depends on the choice of the generating set S, but if S' is another generating set, then there is a natural identification $\partial_\infty \mathcal{G}(G, S') \cong \partial_\infty \mathcal{G}(G, S)$; namely, since G is cobounded in $\mathcal{G}(G, S)$ and $\mathcal{G}(G, S')$, one can represent points in the boundaries of both spaces by equivalence classes of sequences *in G* converging to infinity where the equivalence relation is independent of the generating set. So $\partial_\infty G$ is well-defined.

One has to be careful though when one considers visual metrics. If ϱ is a visual metric on $\partial_\infty \mathcal{G}(G, S)$, then in general ϱ will not be a visual metric on $\partial_\infty \mathcal{G}(G, S')$; but if ϱ' is a visual metric on $\partial_\infty \mathcal{G}(G, S')$, then ϱ and ϱ' are quasisymmetrically equivalent. In other words, the identity map between $(\partial_\infty \mathcal{G}(G, S), \varrho)$ and $(\partial_\infty \mathcal{G}(G, S'), \varrho')$ (given by the natural identification of these spaces as discussed) is a quasisymmetry. This follows from Proposition 4.5 and the fact that $\mathcal{G}(G, S)$ and $\mathcal{G}(G, S')$ are quasi-isometric. So $\partial_\infty G$ carries a natural quasisymmetric gauge.

If we equip $\partial_\infty G$ with any of these visual metrics ϱ, then we can unambiguously speak of quasisymmetric maps on $\partial_\infty G$. Another consequence of this is that $\partial_\infty G$ carries a unique topology induced by any visual metric ϱ on $\partial_\infty G \cong \partial_\infty \mathcal{G}(G, S)$.

Letting a group element $g \in G$ act on the vertices of $\mathcal{G}(G, S)$ by left-translation, we get a natural action $G \curvearrowright \mathcal{G}(G, S)$. This action is *geometric*, i.e., it is isometric, properly discontinuous, and cocompact. To get a better understanding of the properties of a group, one often wants to find a "better" space than $\mathcal{G}(G, S)$ on which G admits a geometric action (for a systematic exploration of this point of view see [**Kl06**]).

A related question is how the topological structure of the boundary $\partial_\infty G$ of a Gromov hyperbolic group determines its algebraic structure. Since the Cayley graph of G and any of its subgroups of finite index are quasi-isometric and hence indistinguishable from the perspective of coarse geometry, one is mostly interested in *virtual properties* of G, i.e., algebraic properties that are true for some subgroup of finite index.

The spaces $\partial_\infty G$ form a very restricted class. For example, if G is *non-elementary* (meaning that $\#\partial_\infty G \geq 3$) and has topological dimension 0, then $\partial_\infty G$ is homeomorphic to a Cantor set. Moreover, in this case G is *virtually isomorphic* to a free group (i.e., some finite-index subgroup of G is free).

If $\partial_\infty G$ is homeomorphic to a circle, then G is *virtually Fuchsian*; so G is virtually isomorphic to a fundamental group of a compact hyperbolic surface, or equivalently, there is a geometric action of G on $\mathbb{H}^2$ (see [**KB02**] for an overview on this subject).

If $\partial_\infty G$ has no local cut points and has topological dimension one, then $\partial_\infty G$ is a Menger curve or a Sierpiński carpet. Moreover, a conjecture due to Kapovich and Kleiner [**KK00**] predicts that in the latter case, G admits a geometric action on a convex subset of $\mathbb{H}^3$ with non-empty totally geodesic boundary.

This conjecture is related to (and implied by) another conjecture, due to Cannon (see [**Ca94**, p. 232]).

CONJECTURE (Cannon's conjecture. Version I). *Let G be a Gromov hyperbolic group and suppose $\partial_\infty G$ is homeomorphic to $\widehat{\mathbb{C}}$. Then there exists a geometric action of G on hyperbolic 3-space $\mathbb{H}^3$.*

If this were true, then G would be virtually isomorphic to the fundamental group of a compact hyperbolic 3-manifold.

In higher dimensions a corresponding statement is false; there are Gromov hyperbolic groups G with $\partial_\infty G$ homeomorphic to an n-sphere, $n \geq 3$, that do not admit geometric actions on $\mathbb{H}^{n+1}$. One can obtain such examples as fundamental groups of *Gromov-Thurston manifolds* [**GT87**]. These are closed negatively-curved manifolds that do not carry a hyperbolic metric, and exist in dimension $n + 1 \geq 4$.

Cannon's conjecture can be reformulated in equivalent form as a quasisymmetric uniformization problem (see [**Bo06**] for more discussion).

CONJECTURE (Cannon's conjecture. Version II). *Let G be a Gromov hyperbolic group and suppose $\partial_\infty G$ is homeomorphic to $\widehat{\mathbb{C}}$. Then $\partial_\infty G$ equipped with a visual metric is quasisymmetrically equivalent to $(\widehat{\mathbb{C}}, \sigma)$.*

In view of this formulation of the conjecture it is very interesting to study the quasisymmetric uniformization problem for metric 2-spheres in general and ask for general conditions under which such a sphere is quasisymmetrically equivalent to the standard 2-sphere $(\widehat{\mathbb{C}}, \sigma)$. Here the conditions should be similar to those that one can establish for boundaries of Gromov hyperbolic groups.

In all known examples where $\partial_\infty G$ is a 2-sphere, G is (essentially) the fundamental group of a hyperbolic manifold and $\partial_\infty G$ can naturally be identified with $\partial_\infty \mathbb{H}^3$ which is the standard 2-sphere. So in these cases, no uniformization problem arises. Cannon's conjecture predicts that there are no other examples. One of the difficulties in making progress on Cannon's conjecture is this lack of non-trivial examples that may guide the intuition.

In contrast, the theory of Thurston maps provides a large class of self-similar fractal 2-spheres that sometimes are, and sometimes are not, quasisymmetrically equivalent to the standard 2-sphere. By analyzing these examples, one may hope to discover some general features that could also be relevant in the solution of Cannon's conjecture.

4.4. Quasispheres

A metric space that is quasisymmetrically equivalent to $\widehat{\mathbb{C}}$ equipped with the chordal metric is called a *quasisphere*. In view of the previous discussion of Cannon's conjecture and the characterization of rational Thurston maps as given by Theorem 18.1 (ii), we now want to discuss two examples that may guide the reader's intuition. As this is our main purpose here, we will skip the justification of most details.

EXAMPLE 4.6. A *snowball* is a compact set in $\mathbb{R}^3$ that is obtained in a way analogous to the set in the plane bounded by the classical von Koch snowflake curve. The boundary of a snowball is a *snowsphere* $\mathcal{S}$. In many cases this is a quasisphere.

The simplest example is obtained as follows. We start with the unit cube $[0,1]^3 \subset \mathbb{R}^3$ as the 0-th approximation $\mathcal{B}^0$ of the snowball. The boundary of $\mathcal{B}^0$ is a polyhedral surface $\mathcal{S}^0$ consisting of six copies of
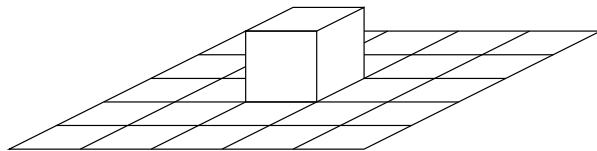
FIGURE 4.1. The generator of the snowsphere $\mathcal{S}$.

the unit square $[0,1]^2 \subset \mathbb{R}^2$ as faces. We divide each of these six faces into 5×5 squares of side length $1/5$ (a $1/5$-*square*). On the $1/5$-square in the middle of each face we place a cube that has side length $1/5$ and sticks out of $\mathcal{B}^0$. This results in a set $\mathcal{B}^1 \supset \mathcal{B}^0$. The boundary of $\mathcal{B}^1$ is a polyhedral surface $\mathcal{S}^1$ consisting of 6×29 $1/5$-squares. The procedure is now iterated; namely, each $1/5$-square is divided into 5×5 squares of side length $1/25$, on each middle square we put a cube of side length $1/25$, and so on. We obtain an increasing sequence $\mathcal{B}^0 \subset \mathcal{B}^1 \subset \ldots$ of compact sets in $\mathbb{R}^3$. Their union is the snowball $\mathcal{B}$ with the snowsphere $\mathcal{S} := \partial\mathcal{B}$ as its boundary. One can show that $\mathcal{S}$ is indeed a 2-sphere. For each $n \in \mathbb{N}_0$ the boundary of $\mathcal{B}^n$ is a polyhedral surface $\mathcal{S}^n$ that consists of $1/5^n$-squares. The surfaces $\mathcal{S}^n$ give better and better approximations of the snowsphere $\mathcal{S}$ for larger and larger n.

One can also give another construction of $\mathcal{S}$ by a replacement procedure very similar to the one in Section 1.3. For this we let the *generator* of the snowball be the polyhedral surface shown in Figure 4.1. The approximation $\mathcal{S}^{n+1}$ is then obtained from $\mathcal{S}^n$ by replacing each 5^{-n}-square of $\mathcal{S}^n$ by a scaled copy of the generator. If X^n is one of the 5^{-n}-squares from which $\mathcal{S}^n$ is built, and X^{n+1} is a $5^{-(n+1)}$-square in the scaled copy of the generator that replaces X^n, we write $X^n \sqsupset X^{n+1}$.

The snowsphere $\mathcal{S}$ inherits the Euclidean metric from $\mathbb{R}^3$. It is not hard to see that if $x, y \in \mathcal{S}$, then there exists a rectifiable path $\gamma \subset \mathcal{S}$ joining x and y whose length is comparable to $|x - y|$. If we define $d(x, y)$ for $x, y \in \mathcal{S}$ as the infimum of the lengths of such paths, then we get a length metric d on $\mathcal{S}$ that is bi-Lipschitz equivalent to the Euclidean metric (see [**Me02**] and [**Me10**]).

One can also obtain the Euclidean metric on $\mathcal{S}$ in an intrinsic way similar to the example in Section 1.3. For this we note that for each point $x \in \mathcal{S}$ there exist sequences $X^0 \sqsupset X^1 \sqsupset \ldots$ of $1/5^n$-squares X^n such that $\{X^n\}$ Hausdorff converges to $\{x\}$ in $\mathbb{R}^3$. Now for $x, y \in \mathcal{S}$, $x \neq y$, we define (compare with (1.3))

(4.15) $m(x, y) := \inf \min\{n \in \mathbb{N}_0 : X^n \cap Y^n = \emptyset\}$,

where the infimum is taken over all such sequences $\{X^n\}$ for x and $\{Y^n\}$ for y. Then Euclidean distance of x and y can essentially be

expressed in terms of $m(x, y)$; namely,

$$(4.16) \qquad\qquad |x - y| \asymp 5^{-m(x,y)},$$

where $C = C(\asymp)$ is independent of x and y. So up to a multiplicative constant, the Euclidean metric on $\mathcal{S}$ can be recovered from the combinatorics of the sets X^n.

It is a small step from here to the theory of Gromov hyperbolic spaces. We construct a graph $\mathcal{G}$ as follows (see Chapter 10 for very similar considerations). The set of vertices of $\mathcal{G}$ is the set of $1/5^n$-squares X^n for all $n \in \mathbb{N}_0$. It is convenient to add another vertex X^{-1}. We declare that $X^{-1} \sqsupset X^0$ for any $1/5^0$-square X^0. Then each vertex of $\mathcal{G}$ as represented by X^n has an attached level $n \in \mathbb{N}_0 \cup \{-1\}$.

The set of (undirected) edges of $\mathcal{G}$ is now given as follows. We connect two distinct vertices by an edge if they have the same level n and are represented by two $1/5^n$-squares X^n and Y^n with $X^n \cap Y^n \neq \emptyset$. Moreover, we join two vertices represented by X^n and X^{n+1} if $X^n \sqsupset X^{n+1}$. There are no other edges in $\mathcal{G}$.

If we identify each edge with a copy of the unit interval $[0, 1]$, then $\mathcal{G}$ carries a natural path metric (corresponding to combinatorial distance in $\mathcal{G}$ on the set of vertices). It can be shown that with this path metric $\mathcal{G}$ is a Gromov hyperbolic metric space.

There is a natural identification of $\mathcal{S}$ with the boundary at infinity $\partial_\infty \mathcal{G}$. Namely, if $x \in \mathcal{S}$, then we choose a sequence $X^0 \sqsupset X^1 \sqsupset \ldots$ of $1/5^n$-squares such that $\{X^n\}$ Hausdorff converges to $\{x\}$ in $\mathbb{R}^3$. Then $\{X^n\}$, now considered as a sequence of vertices in $\mathcal{G}$, converges at infinity (see Section 4.2). Sending a point x to the equivalence class of $\{X^n\}$ (considered as a point on $\partial_\infty \mathcal{G}$), we get a bijection between $\mathcal{S}$ and $\partial_\infty \mathcal{G}$ that we use to identify these two sets.

We choose the basepoint $p = X^{-1}$ in $\mathcal{G}$. Then for the Gromov product $(x \cdot y)_p$ of two points $x, y \in \mathcal{S} = \partial_\infty \mathcal{G}$ we have

$$m(x, y) - c \leq (x \cdot y)_p \leq m(x, y) + c,$$

where $c \geq 0$ is a constant independent of x and y. From this and (4.16) it follows that the Euclidean metric on $\mathcal{S} = \partial_\infty \mathcal{G}$ is a visual metric in the sense of Gromov hyperbolic spaces; indeed, it satisfies (4.14) with $\Lambda = 5$. It is easy to see that there are no visual metrics with $\Lambda > 5$.

The snowsphere $\mathcal{S}$ (equipped with the metric inherited from $\mathbb{R}^3$) is a quasisphere. This was shown in [**Me02**] (see also [**Me10**] and [**Me09a**]). For the proof one constructs a rational Thurston map (in a non-obvious way) leading to sets that mirror the combinatorics of the sets X^n related to $\mathcal{S}$. One can use this to show directly that $\mathcal{S}$
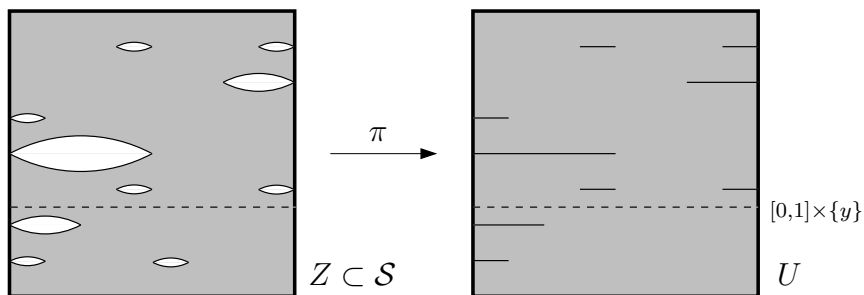
FIGURE 4.2. The set Z.

is a quasisphere, or one invokes the general criterion given by Theorem 18.1 (ii).

EXAMPLE 4.7 (A non-quasisphere). Our second example is the sphere $\mathcal{S}$ equipped with the metric ϱ as discussed in Section 1.3. As we already remarked, it follows from the fact that the associated Thurston map h has a Thurston obstruction in combination with Theorem 18.1 (ii) that $\mathcal{S}$ is not a quasiphere. Here we want to outline a direct argument for this statement (it emerged in discussions with B. Kleiner). We use the notation from Section 1.3.

Consider the top part of $\mathcal{S}$ given by all equivalence classes of sequences $X^0 \sqsupset X^1 \sqsupset \ldots$ where X^0 is the top white 0-tile of $\mathcal{S}^0$. From this top part we remove all the "flaps" that were successively added to X^0 in the construction of $\mathcal{S}$. What remains is a subset $Z \subset \mathcal{S}$ that looks like the unit square $U = [0,1]^2$ with countably many slits (related to Z are the "slit carpets" considered in [**Mer10**]). These slits are all parallel to one of the sides of U, say to $[0,1] \times \{0\}$. Each point $z \in Z$ corresponds to a unique point in U. This gives a surjective map $\pi \colon Z \to U$. If a point $p \in U$ is an interior point of one of the slits, then there are two points in Z (one for each side of the slit) that map to p. For all other points $p \in U$ we have $\#\pi^{-1}(p) = 1$.

We equip Z with the metric ϱ and U with the Euclidean metric. Then the map $\pi \colon Z \to U$ is Lipschitz. Actually, π is *David-Semmes regular* (see [**DS97**, Chapter 12]). For π this means that in addition to being Lipschitz, there exists a number $N \in \mathbb{N}$ such that the preimage $\pi^{-1}(B(p,r))$ of each Euclidean ball $B(p,r)$ in U can be covered by N balls in Z of the same radius r.

At least on an intuitive level, one can see that π has this last property as follows. If no slit cuts through $B(p,r)$, then $\pi^{-1}(B(p,r))$ is contained in a ball in Z whose radius is not much larger, and hence comparable to r. If a slit cuts through $B(p,r)$, then $\pi^{-1}(B(p,r))$ is

split into two parts each of which is contained in a ball in Z with radius comparable to r. This implies that in any case, $\pi^{-1}(B(p,r))$ can be covered by a controlled number of balls in Z with radius r.

Since U is Ahlfors 2-regular, and π is David-Semmes regular, Z is Ahlfors 2-regular as well ([**DS97**, Lemma 12.5]).

Let Γ be the set of all paths in Z that project under π to a line segment that has the form $[0,1] \times \{y\}$, $y \in [0,1]$, and does not contain a slit. Restricted to a path $\gamma \in \Gamma$, the map π is bi-Lipschitz with a uniform constant independent of γ. This implies that the 2-modulus of Γ in Z (see [**He01**, Section 7.3] for the definition of the modulus of a path family) cannot be much smaller than the 2-modulus of $\pi(\Gamma)$ and is hence positive. Together with the Ahlfors 2-regularity of Z this implies that an image of Z under any quasisymmetric homeomorphism has Hausdorff dimension ≥ 2 ([**He01**, Theorem 15.10]).

One other property of Z will be important. Namely, Z is a *porous* subset of $\mathcal{S}$. This means that there exists a constant $c \in (0,1)$ with the following property: if $a \in Z$ and $r > 0$ with $r \leq \operatorname{diam}(Z)$ are arbitrary, then there exists $x \in B(a,r)$ with $B(x,cr) \cap Z = \emptyset$. So the set $B(a,r) \cap Z$ has the "hole" $B(x,cr)$ of comparable size.

The porosity of Z follows from the fact that each ball $B(a,r)$ in $\mathcal{S}$ with $r \leq \operatorname{diam}(\mathcal{S})$ contains a flap of size comparable to r. Since in the construction of Z we removed all flaps from the top side of $\mathcal{S}$, this means that all sufficiently small balls centered in Z contain a hole of about the same size.

Now we can see that $\mathcal{S}$ is not a quasisphere as follows. We argue by contradiction and assume that there exists a quasisymmetry φ of $\mathcal{S}$ onto $\widehat{\mathbb{C}}$ (equipped with the chordal metric). From what we have discussed above, it then follows that the Hausdorff dimension of $\varphi(Z)$ is ≥ 2. On the other hand, images of porous sets under quasisymmetries are porous (see [**Vä87**, Theorem 4.2] for a proof in $\mathbb{R}^n$; it easily generalizes to a metric space setting). Hence $\varphi(Z)$ is a porous subset of $\widehat{\mathbb{C}}$. A porous subset of an Ahlfors Q-regular space, $Q > 0$, has Hausdorff dimension $< Q$ ([**DS97**, Lemma 5.8]). It follows that the Hausdorff dimension of $\varphi(Z)$ is < 2. This is a contradiction, and so $\mathcal{S}$ cannot be a quasisphere.

CHAPTER 5

Cell decompositions

In this chapter we discuss some technical, but very crucial aspects of our work, namely cell decompositions and their relation to Thurston maps. Since this is the basis of our approach to the investigation of Thurston maps, we collected all the relevant facts in one place with a detailed presentation.

Accordingly, this chapter is quite long and covers a mix of concepts and results that are fairly standard (Sections 5.1 and 5.2), are important right away for our further developments (Sections 5.3, 5.6, and 5.7) or are not needed until much later (Sections 5.4 and 5.5 are not used before Chapter 12). For this reason, the reader may not want to peruse this chapter in linear order and only skim through some of its parts, in particular at first reading.

We start with a general review of cell decompositions for arbitrary spaces (Section 5.1). The reader already acquainted with these concepts may want to look at Definition 5.1 to get familiar with our terminology and notation. We also define refinements of cell decompositions and cellular Markov partitions for a map (see Definition 5.6 and Definition 5.8). For the most part this in preparation of Chapter 12.

In Section 5.2 we specialize to cell decompositions of 2-spheres. The most important facts about them are recorded in Lemma 5.9. We also consider isomorphisms of cell complexes (Definition 5.10) and the homeomorphisms that are induced by them (Lemma 5.11).

In Section 5.3 we consider cell decompositions of the 2-sphere S^2 induced by a Thurston map $f \colon S^2 \to S^2$. For this we consider a Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}$ and an associated cell decomposition $\mathcal{D}^0(f, \mathcal{C})$ of S^2. Its cells are given by the points in $\mathrm{post}(f)$ as vertices, the closed arcs into which the points in $\mathrm{post}(f)$ divide $\mathcal{C}$, and the closures of the two components of $S^2 \setminus \mathcal{C}$. Pulling this decomposition back by f^n, we obtain a cell decomposition $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$ of S^2 for each $n \in \mathbb{N}_0$ (see Corollary 5.13 and Definition 5.14). These cell decompositions $\mathcal{D}^n$ are our most important tool for studying Thurston maps.

The properties of $\mathcal{D}^n$ are collected in Proposition 5.17. This is an elementary, but central result of this chapter and will be used throughout this work.

The 2-dimensional cells or tiles in $\mathcal{D}^0(f, \mathcal{C})$ are the two closed Jordan regions in S^2 bounded by $\mathcal{C}$. If we assign the colors black and white to them, then we can pull this coloring back by f^n and obtain colors for the tiles in $\mathcal{D}^n$ (see Lemma 5.21). This is closely related to the more general notion of a *labeling* considered in Section 5.4. For the cell decompositions $\mathcal{D}^0 = \mathcal{D}^0(f, \mathcal{C})$ and $\mathcal{D}^1 = \mathcal{D}^1(f, \mathcal{C})$ as in Definition 5.14, this is simply the map $L\colon \mathcal{D}^1 \to \mathcal{D}^0$ given by $L(c) = f(c)$ for $c \in \mathcal{D}^1$. We will turn this around in Section 5.5, and construct Thurston maps from cell decompositions $\mathcal{D}^1$ and $\mathcal{D}^0$ and such a labeling L (see Proposition 5.26). These notions will be revisited later in Chapter 12, where we will assume in addition that $\mathcal{D}^1$ is a refinement of $\mathcal{D}^0$. This will allow us to define and describe Thurston maps by finite combinatorial data. The reader may safely skip Sections 5.4 and 5.5 until this material is needed in Chapter 12.

In Section 5.6 we consider *flowers*. These are simply connected neighborhoods of vertices in $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$ (see Definition 5.27). Their properties are summarized in Lemma 5.28. They behave well under the maps f^k (see Lemma 5.29). We also introduce *edge-flowers* as neighborhoods of edges in $\mathcal{D}^n$ (see Definition 5.30 and Lemma 5.31).

In the last section (Section 5.7) we introduce the notion of *joining opposite sides* of the Jordan curve $\mathcal{C}$. The most important result here is Lemma 5.35, which says that if a connected set K joins two disjoint cells in $\mathcal{D}^n(f, \mathcal{C})$, then $f^n(K)$ joins opposite sides of $\mathcal{C}$. This fact will be of major importance when we construct visual metrics in Chapter 8 (see also Chapter 16).

5.1. Cell decompositions in general

Here we review some facts about cell decompositions of arbitrary spaces. Most of this material is well known (see, for example, [**CF67**, Chapter 1]). For the purpose of the present work we could have restricted ourselves to cell decompositions of subsets of a 2-sphere, but it is more transparent to discuss the topic in greater generality.

In this section $\mathcal{X}$ will always be a locally compact Hausdorff space. A *(closed topological) cell c of dimension* $n = \dim(c) \in \mathbb{N}$ in $\mathcal{X}$ is a set $c \subset \mathcal{X}$ that is homeomorphic to the closed unit ball $\overline{\mathbb{B}}^n$ in $\mathbb{R}^n$. We denote by ∂c the set of points corresponding to $\partial \overline{\mathbb{B}}^n$ under such a homeomorphism between c and $\overline{\mathbb{B}}^n$. This is independent of the homeomorphism chosen, and the set ∂c is well-defined. We call ∂c the

boundary and $\mathrm{int}(c) = c \setminus \partial c$ the *interior* of c. Note that boundary and interior of c in this sense will in general not agree with the boundary and interior of c regarded as a subset of the topological space $\mathcal{X}$. A *cell of dimension* 0 in $\mathcal{X}$ is a set $c \subset \mathcal{X}$ consisting of a single point. We set $\partial c = \emptyset$ and $\mathrm{int}(c) = c$ in this case.

DEFINITION 5.1 (Cell decompositions). Suppose that $\mathcal{D}$ is a collection of cells in a locally compact Hausdorff space $\mathcal{X}$. We say that $\mathcal{D}$ is a *cell decomposition* of $\mathcal{X}$ provided the following conditions are satisfied:

(i) the union of all cells in $\mathcal{D}$ is equal to $\mathcal{X}$,

(ii) we have $\mathrm{int}(\sigma) \cap \mathrm{int}(\tau) = \emptyset$, whenever $\sigma, \tau \in \mathcal{D}$, $\sigma \neq \tau$,

(iii) if $\tau \in \mathcal{D}$, then $\partial \tau$ is a union of cells in $\mathcal{D}$,

(iv) every point in $\mathcal{X}$ has a neighborhood that meets only finitely many cells in $\mathcal{D}$.

Note that the literature one often uses the term *regular* for cell decompositions as in Definition 5.1 in order to distinguish them from more general notions of cell decompositions (as in the theory of CW-complexes, for example).

If $\mathcal{D}$ is a collection of cells in some ambient space $\mathcal{X}$, then we call $\mathcal{D}$ a *cell complex* if $\mathcal{D}$ is a cell decomposition of the underlying set

$$|\mathcal{D}| := \bigcup \{c : c \in \mathcal{D}\}.$$

Suppose $\mathcal{D}$ is a cell decomposition of $\mathcal{X}$. By (iv), every compact subset of $\mathcal{X}$ can only meet finitely many cells in $\mathcal{D}$. In particular, if $\mathcal{X}$ is compact, then $\mathcal{D}$ consists of only finitely many cells. Moreover, for each $\tau \in \mathcal{D}$, the set $\partial \tau$ is compact and hence equal to a finite union of cells in $\mathcal{D}$. It follows from basic dimension theory that if $\dim(\tau) = n$, then $\partial \tau$ is equal to a union of cells in $\mathcal{D}$ that have dimension $n - 1$.

The union $\mathcal{X}^n$ of all cells in $\mathcal{D}$ of dimension $\leq n$ is called the *n-skeleton* of the cell decomposition. It is useful to set $\mathcal{X}^{-1} = \emptyset$. It follows from the local compactness of $\mathcal{X}$ and property (iv) of a cell decomposition that $\mathcal{X}^n$ is a closed subset of $\mathcal{X}$ for each $n \in \mathbb{N}_0$. By the last remark in the previous paragraph, we have $\partial \tau \subset \mathcal{X}^{n-1}$ for each $\tau \in \mathcal{D}$ with $\dim(\tau) = n$.

LEMMA 5.2. *Let $\mathcal{D}$ be a cell decomposition of $\mathcal{X}$. Then for each $n \in \mathbb{N}_0$ the n-skeleton $\mathcal{X}^n$ is equal to the disjoint union of the sets $\mathrm{int}(c)$, $c \in \mathcal{D}$, $\dim(c) \leq n$. The space $\mathcal{X}$ is equal to the disjoint union of the sets $\mathrm{int}(c)$, $c \in \mathcal{D}$. Similarly, every cell $\tau \in \mathcal{D}$ is the disjoint union of the sets $\mathrm{int}(c)$, where $c \in \mathcal{D}$ and $c \subset \tau$.*

So in particular, the interiors of the cells in a cell decomposition partition the space $\mathcal{X}$. This is of prime importance and will be used frequently throughout this work.

PROOF. We show the first statement by induction on $n \in \mathbb{N}_0$. Since $\mathrm{int}(c) = c$ for each cell c in $\mathcal{D}$ of dimension 0, it is clear that $\mathcal{X}^0$ is the disjoint union of the interiors of all cells $c \in \mathcal{D}$ with $\dim(c) = 0$.

Suppose that the first statement is true for $\mathcal{X}^n$, and let $p \in \mathcal{X}^{n+1}$ be arbitrary. If $p \in \mathcal{X}^n$, then p is contained in the interior of a cell $c \in \mathcal{D}$ with $\dim(c) \leq n$ by induction hypothesis. In the other case, $p \in \mathcal{X}^{n+1} \setminus \mathcal{X}^n$, and so there exists $c \in \mathcal{D}$ with $\dim(c) = n + 1$ and $p \in c$. Since $\partial c \subset \mathcal{X}^n$, it follows that $p \in c \setminus \partial c = \mathrm{int}(c)$. So $\mathcal{X}^{n+1}$ is the union of the interiors of all cells c in $\mathcal{D}$ with $\dim(c) \leq n+1$. This union is disjoint, because distinct cells in a cell decomposition have disjoint interiors.

The second statement follows from the first, and the obvious fact that $\mathcal{X} = \bigcup_{n \in \mathbb{N}_0} \mathcal{X}^n$.

To see the last statement, let $\mathcal{D}_\tau := \{c \in \mathcal{D} : c \subset \tau\}$. Then it is clear that $\mathcal{D}_\tau$ is a cell decomposition of (the compact Hausdorff space) τ. So the claim follows from the previous statement. $\square$

The lemma immediately implies that if $\tau \in \mathcal{D}$ and $\dim(\tau) = n$, then each point $p \in \mathrm{int}(\tau)$ is an interior point of τ regarded as a subset of the topological space $\mathcal{X}^n$.

LEMMA 5.3. *Let $\mathcal{D}$ be a cell decomposition of $\mathcal{X}$.*

(i) *If σ and τ are two distinct cells in $\mathcal{D}$ with $\sigma \cap \tau \neq \emptyset$, then one of the following statements holds: $\sigma \subset \partial\tau$, $\tau \subset \partial\sigma$, or $\sigma \cap \tau = \partial\sigma \cap \partial\tau$ and this intersection consists of cells in $\mathcal{D}$ of dimension strictly less than $\min\{\dim(\sigma), \dim(\tau)\}$.*

(ii) *If $\sigma, \tau_1, \ldots, \tau_n$ are cells in $\mathcal{D}$ and $\mathrm{int}(\sigma) \cap (\tau_1 \cup \cdots \cup \tau_n) \neq \emptyset$, then $\sigma \subset \tau_i$ for some $i \in \{1, \ldots, n\}$.*

PROOF. (i) We may assume that $l := \dim(\sigma) \leq m := \dim(\tau)$, and prove the statement by induction on m. The case $m = 0$ is vacuous and hence trivial. Assume that the statement is true whenever both cells have dimension $< m$. If $l = m$ then by definition of a cell decomposition $\mathrm{int}(\sigma)$ is disjoint from $\tau \subset \mathrm{int}(\tau) \cup \mathcal{X}^{m-1}$, and similarly $\mathrm{int}(\tau) \cap \sigma = \emptyset$. Hence $\sigma \cap \tau = \partial\tau \cap \partial\sigma$. Moreover, both sets $\partial\sigma$ and $\partial\tau$ consist of finitely many cells in $\mathcal{D}$ of dimension $\leq m - 1$. Applying the induction hypothesis to pairs of these cells, we see that $\partial\tau \cap \partial\sigma$ consists of cells of dimension $< m$ as desired.

If $l < m$, then $\sigma \subset \mathcal{X}^{m-1}$ and so $\sigma \cap \mathrm{int}(\tau) = \emptyset$. This shows that $\sigma \cap \tau = \sigma \cap \partial\tau$. Moreover, we have $\partial\tau = c_1 \cup \cdots \cup c_s$, where $c_1, \ldots, c_s$

are cells of dimension $m - 1$. So we can apply the induction hypothesis to the pairs (σ, c_i). If $\sigma = c_i$ or $\sigma \subset \partial c_i$ for some i, then $\sigma \subset \partial \tau$; we cannot have $c_i \subset \partial \sigma$, because c_i has dimension $m - 1$, and $\partial \sigma$ is a set of topological dimension $< m - 1$. So if none of the first possibilities occurs, then $\sigma \cap c_i = \emptyset$ or $\sigma \cap c_i = \partial \sigma \cap \partial c_i$ and this set consists of cells of dimension $< l$ (by induction hypothesis) contained in $\partial c_i \subset c_i \subset \partial \tau$ for all i. In this case $\sigma \cap \tau = \partial \sigma \cap \partial \tau$, and this set consists of cells of dimension $< l$ as desired. The claim follows.

(ii) There exists $i \in \{1, \ldots, n\}$ with $\mathrm{int}(\sigma) \cap \tau_i \neq \emptyset$. By the alternatives in (i) we then must have $\sigma = \tau_i$ or $\sigma \subset \partial \tau_i$. Hence $\sigma \subset \tau_i$. $\quad\square$

LEMMA 5.4. *Let $A \subset \mathcal{X}$ be a closed set, and $U \subset \mathcal{X} \setminus A$ be a non-empty open and connected set. If $\partial U \subset A$, then U is a connected component of $\mathcal{X} \setminus A$.*

PROOF. Since U is a non-empty connected set in the complement of A, this set is contained in a unique connected component V of $\mathcal{X} \setminus A$. Since $\partial U \subset A \subset \mathcal{X} \setminus V$, we have $V \cap \overline{U} = V \cap U = U$ showing that U is relatively open and closed in V. Since $U \neq \emptyset$ and V is connected, it follows that $U = V$ as desired. $\quad\square$

LEMMA 5.5. *Let $\mathcal{D}$ be a cell decomposition of $\mathcal{X}$ with n-skeleton $\mathcal{X}^n$, $n \in \{-1\} \cup \mathbb{N}_0$. Then for each $n \in \mathbb{N}_0$ the non-empty connected components of $\mathcal{X}^n \setminus \mathcal{X}^{n-1}$ are precisely the sets $\mathrm{int}(\tau)$, where $\tau \in \mathcal{D}$ and $\dim(\tau) = n$.*

PROOF. Let τ be a cell in $\mathcal{D}$ with $\dim(\tau) = n$. Then $\mathrm{int}(\tau)$ is a connected set contained in $\mathcal{X}^n \setminus \mathcal{X}^{n-1}$ that is relatively open with respect to $\mathcal{X}^n$. Its relative boundary is a subset of $\partial \tau$ and hence contained in the closed set $\mathcal{X}^{n-1}$. It follows by Lemma 5.4 that $\mathrm{int}(\tau)$ is equal to a component V of $\mathcal{X}^n \setminus \mathcal{X}^{n-1}$.

Conversely, suppose that V is a non-empty connected component of $\mathcal{X}^n \setminus \mathcal{X}^{n-1}$. Pick a point $p \in V$. Then p lies in the interior of a unique cell $\tau \in \mathcal{D}$ with $\dim(\tau) = n$. It follows from the first part of the proof that $V = \mathrm{int}(\tau)$. $\quad\square$

DEFINITION 5.6 (Refinements). Let $\mathcal{D}'$ and $\mathcal{D}$ be two cell decompositions of the space $\mathcal{X}$. We say that $\mathcal{D}'$ is a *refinement* of $\mathcal{D}$ if the following two conditions are satisfied:

(i) For every cell $\sigma \in \mathcal{D}'$ there exists a cell $\tau \in \mathcal{D}$ with $\sigma \subset \tau$.

(ii) Every cell $\tau \in \mathcal{D}$ is the union of all cells $\sigma \in \mathcal{D}'$ with $\sigma \subset \tau$.

It is easy to see that if $\mathcal{D}'$ is a refinement of $\mathcal{D}$ and $\tau \in \mathcal{D}$, then the cells $\sigma \in \mathcal{D}'$ with $\sigma \subset \tau$ form a cell decomposition of τ. Moreover,

every cell in $\mathcal{D}'$ arises in this way. So roughly speaking, the refinement $\mathcal{D}'$ of the cell decomposition $\mathcal{D}$ is obtained by decomposing each cell in $\mathcal{D}$ into smaller cells. We informally refer to this process as *subdividing* the cells in $\mathcal{D}$ by the smaller cells in $\mathcal{D}'$.

LEMMA 5.7. *Let $\mathcal{D}'$ and $\mathcal{D}$ be two cell decompositions of $\mathcal{X}$, and $\mathcal{D}'$ be a refinement of $\mathcal{D}$. Then for every cell $\sigma \in \mathcal{D}'$ there exists a minimal cell $\tau \in \mathcal{D}$ with $\sigma \subset \tau$, i.e., if $\widetilde{\tau} \in \mathcal{D}$ is another cell with $\sigma \subset \widetilde{\tau}$, then $\tau \subset \widetilde{\tau}$. Moreover, τ is the unique cell in $\mathcal{D}$ with $\mathrm{int}(\sigma) \subset \mathrm{int}(\tau)$.*

PROOF. First note that if $\sigma \in \mathcal{D}'$, $\tau_1, \ldots, \tau_n \in \mathcal{D}$ and

$$\mathrm{int}(\sigma) \cap (\tau_1 \cup \cdots \cup \tau_n) \neq \emptyset,$$

then $\sigma \subset \tau_i$ for some $i \in \{1, \ldots, n\}$. Indeed, by definition of a refinement the union of all cells in $\mathcal{D}'$ contained in some τ_i covers $\tau_1 \cup \cdots \cup \tau_n$. Hence this union meets $\mathrm{int}(\sigma)$. It follows from Lemma 5.3 (ii) that σ is contained in one of these cells from $\mathcal{D}'$ and hence in one of the cells τ_i.

Now if $\sigma \in \mathcal{D}'$ is arbitrary, then σ is contained in some cell of $\mathcal{D}$ by definition of a refinement, and hence in a cell $\tau \in \mathcal{D}$ of minimal dimension. Then τ is minimal among all cells in $\mathcal{D}$ containing σ. Indeed, let $\widetilde{\tau} \neq \tau$ be another cell in $\mathcal{D}$ containing σ. We want to show that $\tau \subset \widetilde{\tau}$.

One of the alternatives in Lemma 5.3 (i) occurs. If $\tau \subset \partial\widetilde{\tau} \subset \widetilde{\tau}$ we are done. The second alternative, $\widetilde{\tau} \subset \partial\tau$, is impossible, since τ has minimal dimension among all cells containing σ. The third alternative leads to $\sigma \subset \tau \cap \widetilde{\tau} = \partial\tau \cap \partial\widetilde{\tau}$, where the latter intersection consists of cells in $\mathcal{D}$ of dimension $< \dim(\tau)$. By the first part of the proof σ is contained in one of these cells, again contradicting the definition of τ. Hence τ is minimal.

We have $\mathrm{int}(\sigma) \subset \mathrm{int}(\tau)$; for otherwise $\mathrm{int}(\sigma)$ meets $\partial\tau$ which is a union of cells in $\mathcal{D}$. Then σ would be contained in one of these cells by the first part of the proof. This contradicts the minimality of τ.

Finally, it is clear that $\tau \in \mathcal{D}$ is the unique cell with $\mathrm{int}(\sigma) \subset \mathrm{int}(\tau)$, because distinct cells in a cell decomposition have disjoint interiors. $\square$

DEFINITION 5.8 (Cellular maps and cellular Markov partitions). Let $\mathcal{D}'$ and $\mathcal{D}$ be two cell decompositions of $\mathcal{X}$, and $f\colon \mathcal{X} \to \mathcal{X}$ be a continuous map. We say that f is *cellular* for $(\mathcal{D}', \mathcal{D})$ if the following condition is satisfied:

> If $\sigma \in \mathcal{D}'$ is arbitrary, then $f(\sigma)$ is a cell in $\mathcal{D}$ and $f|\sigma$ is a homeomorphism of σ onto $f(\sigma)$.

If f is cellular with respect to $(\mathcal{D}', \mathcal{D})$ and $\mathcal{D}'$ is a refinement of $\mathcal{D}$, then the pair $(\mathcal{D}', \mathcal{D})$ is called a *cellular Markov partition* for f.

For the most part we will be considering cellular Markov partitions only starting in Chapter 12. Since almost all examples of Thurston maps we consider are in fact constructed from cellular Markov partitions, we chose to introduce this notion already here.

5.2. Cell decompositions of 2-spheres

We now turn to cell decompositions of 2-spheres. We first review some standard concepts and results from plane topology (see [**Mo77**] for general background and more details).

Let S^2 be a 2-sphere. An *arc* α in S^2 is a homeomorphic image of the unit interval $[0, 1]$. The points corresponding to 0 and 1 under such a homeomorphism are called the *endpoints* of α. They are the unique points $p \in \alpha$ such that $\alpha \setminus \{p\}$ is connected. If p is an *interior point* of α, i.e., a point in α distinct from the endpoints, then there exist arbitrarily small open neighborhoods $W \subset S^2$ of p such that $W \setminus \alpha$ has precisely two open connected components U and V.

A *closed Jordan region* X in S^2 is a homeomorphic image of the closed unit disk $\overline{\mathbb{D}}$. The boundary ∂X of a closed Jordan region $X \subset S^2$ is a *Jordan curve*, i.e., the homeomorphic image of the unit circle $\partial \mathbb{D}$. If $J \subset S^2$ is a Jordan curve, then by the Schönflies theorem there exists a homeomorphism $\varphi \colon S^2 \to \widehat{\mathbb{C}}$ such that $\varphi(J) = \partial \mathbb{D}$. In particular, the set $S^2 \setminus J$ has two connected components, both homeomorphic to $\mathbb{D}$. Note that arcs and closed Jordan regions are cells of dimension 1 and 2, respectively.

Let $\mathcal{D}$ be a cell decomposition of S^2. Since the topological dimension of S^2 is equal to 2, no cell in $\mathcal{D}$ can have dimension > 2. We call the 2-dimensional cells in $\mathcal{D}$ the *tiles*, and the 1-dimensional cells in $\mathcal{D}$ the *edges* of $\mathcal{D}$. The *vertices* of $\mathcal{D}$ are the points $v \in S^2$ such that $\{v\}$ is a cell in $\mathcal{D}$ of dimension 0. So there is a somewhat subtle distinction between vertices and cells of dimension 0: a vertex is an element of S^2, while a cell of dimension 0 is a subset of S^2 with one element.

If c is a cell in $\mathcal{D}$, we denote by ∂c the boundary and by $\mathrm{int}(c)$ the interior of c as introduced in the beginning of Section 5.1. Note that for edges and 0-cells c this is different from the boundary and the interior of c as a subset of the topological space S^2.

We now summarize some facts related to orientation. See Section A.4 for a more detailed discussion.

We always assume that the sphere S^2 is *oriented*, i.e., one of the two generators of the singular homology group $H_2(S^2) \cong \mathbb{Z}$ (with coefficients in $\mathbb{Z}$) has been chosen as the *fundamental class* of S^2.

The orientation on S^2 induces an orientation on every Jordan region $X \subset S^2$ which in turn induces an orientation on ∂X and on every arc $\alpha \subset \partial X$. Here an orientation of an arc is just a selection of one of the endpoints as the *initial point* and the other endpoint as the *terminal point*. Let $X \subset S^2$ be a Jordan region in the oriented 2-sphere S^2 equipped with the induced orientation. If $\alpha \subset \partial X$ is an arc with a given orientation, then we say that X lies *to the left* or *to the right* of α depending on whether the orientation on α induced by the orientation of X agrees with the given orientation on α or not. Similarly, we say that with a given orientation of ∂X the Jordan region X lies to the left or right of ∂X.

To describe orientations, one can also use the notion of a flag. By definition a *flag* in S^2 is a triple (c_0, c_1, c_2), where c_i is an i-dimensional cell for $i = 0, 1, 2$, $c_0 \subset \partial c_1$, and $c_1 \subset \partial c_2$. So a flag in S^2 is a closed Jordan region c_2 with an arc c_1 contained in its boundary, where the point in c_0 is a distinguished endpoint of c_1. We orient the arc c_1 so that the point in c_0 is the initial point in c_1. The flag is called *positively-* or *negatively-oriented* (for the given orientation on S^2) depending on whether c_2 lies to the left or to the right of the oriented arc c_1.

A positively-oriented flag determines the orientation on S^2 uniquely. The standard orientation on $\widehat{\mathbb{C}}$ is the one for which the *standard flag* (c_0, c_1, c_2) is positively-oriented, where $c_0 = \{0\}$, $c_1 = [0, 1] \subset \mathbb{R}$, and

$$c_2 = \{z \in \mathbb{C} : 0 \leq \mathrm{Re}(z) \leq 1, \ 0 \leq \mathrm{Im}(z) \leq \mathrm{Re}(z)\}.$$

Since edges and tiles in a cell decomposition $\mathcal{D}$ of S^2 are arcs and closed Jordan regions, respectively, it makes sense to speak of oriented edges and tiles in $\mathcal{D}$. A *flag in* $\mathcal{D}$ is a flag (c_0, c_1, c_2), where c_0, c_1, c_2 are cells in $\mathcal{D}$. If c_i are i-dimensional cells in $\mathcal{D}$ for $i = 0, 1, 2$, then (c_0, c_1, c_2) is a flag in $\mathcal{D}$ if and only if $c_0 \subset c_1 \subset c_2$.

Cell decompositions of a 2-sphere S^2 have special properties that we summarize in the next lemma.

LEMMA 5.9. *Let $\mathcal{D}$ be a cell decomposition of S^2. Then it has the following properties:*

(i) *There are only finitely many cells in $\mathcal{D}$.*

(ii) *The tiles in $\mathcal{D}$ cover S^2.*

(iii) *Let X be a tile in $\mathcal{D}$. Then there exists a number $k \in \mathbb{N}$, $k \geq 2$, such that X contains precisely k edges $e_1, \ldots, e_k$ and k vertices $v_1, \ldots, v_k$ in $\mathcal{D}$. Moreover, these edges and vertices lie on the boundary ∂X of X, and we have*

$$\partial X = e_1 \cup \cdots \cup e_k.$$

> *The indexing of these vertices and edges can be chosen such that $v_j \in \partial e_j \cap \partial e_{j+1}$ for $j = 1, \ldots, k$ (where $e_{k+1} := e_1$).*

(iv) *Every edge $e \in \mathcal{D}$ is contained in the boundary of precisely two tiles $\mathcal{D}$. If X and Y are these tiles, then $\mathrm{int}(X) \cup \mathrm{int}(e) \cup \mathrm{int}(Y)$ is a simply connected region.*

(v) *Let v be a vertex of $\mathcal{D}$. Then there exists a number $d \in \mathbb{N}$, $d \geq 2$, such that v is contained in precisely d tiles $X_1, \ldots, X_d$, and d edges $e_1, \ldots, e_d$ in $\mathcal{D}$. We have $v \in \partial X_j$ and $v \in \partial e_j$ for each $j = 1, \ldots, d$. Moreover, the indexing of these tiles and edges can be chosen such that $e_j \subset \partial X_j \cap \partial X_{j+1}$ for $j = 1, \ldots, d$ (where $X_{d+1} := X_1$).*

(vi) *The 1-skeleton of $\mathcal{D}$ is connected and equal to the union of all edges in $\mathcal{D}$.*

Statement (iii) actually holds for all tiles (i.e., 2-dimensional cells) in each cell decomposition of a locally compact space. If the boundary of a tile X is subdivided into vertices and edges as in (iii), we say that X is a *(topological) k-gon.*

If the edge e and the tiles X and Y are as in (iv), then there exists a unique orientation of e such that X lies to the left and Y to the right of e.

We say that the cells $\{v\}, e_1, \ldots, e_d, X_1, \ldots, X_d$ as in (v) form the *cycle* of the vertex v and call d the *length* of the cycle. We refer to $X_1, \ldots, X_d$ as the tiles and to $e_1, \ldots, e_d$ as the edges of the cycle.

PROOF. (i) By Definition 5.1 (iv) this follows from the compactness of S^2 and the fact that every point in S^2 has a neighborhood that meets only finitely many cells in $\mathcal{D}$.

(ii) The set consisting of all vertices and the union of all edges has empty interior (in the topological sense) by (i) and Baire's theorem. Hence the union of all tiles is a dense set in S^2. Since this union is also closed by (i), it is all of S^2.

(iii) Let X be a tile in $\mathcal{D}$. Then $\mathrm{int}(X)$ does not meet any edge or vertex, and ∂X is a union of edges and vertices. Since there are only finitely many vertices, ∂X must contain an edge, and hence at least two vertices. Suppose $v_1, \ldots, v_k$, $k \geq 2$, are all the vertices on ∂X. Since ∂X is a Jordan curve, we can choose the indexing of these vertices so that ∂X is a union of arcs α_j with pairwise disjoint interior such that α_j has the endpoints v_j and v_{j+1} for $j = 1, \ldots, k$, where $v_{k+1} = v_1$. Then for each $j = 1, \ldots, k$ the set $\mathrm{int}(\alpha_j)$ is connected and lies in the 1-skeleton of the cell decomposition $\mathcal{D}$, it is disjoint from the 0-skeleton and has boundary contained in the 0-skeleton. It follows
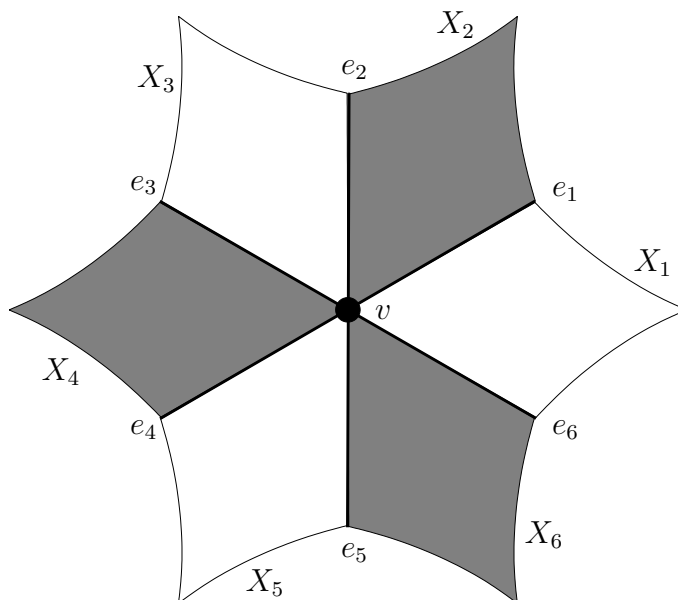
FIGURE 5.1. The cycle of a vertex v.

from Lemma 5.4 and Lemma 5.5 that there exists an edge e_j in $\mathcal{D}$ with $\text{int}(e_j) = \text{int}(\alpha_j)$. Hence $\alpha_j = e_j$, and so α_j is an edge in $\mathcal{D}$. It is clear that ∂X does not contain other edges in $\mathcal{D}$. The statement follows.

(iv) Let e be an edge in $\mathcal{D}$. Pick $p \in \text{int}(e)$. By (ii) the point p is contained in some tile X in $\mathcal{D}$. By Lemma 5.3 (ii) we have $e \subset X$. On the other hand, $\text{int}(X)$ is disjoint from each edge and so $e \subset \partial X$. It follows from the Schönflies theorem that the set X does not contain a neighborhood of p. Hence every neighborhood of p must meet tiles distinct from X. Since there are only finitely many tiles, it follows that there exists a tile Y distinct from X with $p \in Y$. By the same reasoning as before, we have $e \subset \partial Y$.

Let $q \in \text{int}(e)$ be arbitrary. Then there exists a small open and connected neighborhood W of q such that $W \setminus \text{int}(e)$ consists of two connected components U and V. If W is small enough, then U and V do not meet ∂X. Since $q \in \overline{\text{int}(X)}$, one of the sets, say U, meets $\text{int}(X)$, and so $U \subset \text{int}(X)$. We can also assume that the set W is small enough so that it does not meet ∂Y either. By the same reasoning, U or V must be contained in $\text{int}(Y)$, and, since $\text{int}(X) \cap \text{int}(Y) = \emptyset$, we have $V \subset \text{int}(Y)$. Hence $M := \text{int}(X) \cup \text{int}(e) \cup \text{int}(Y)$ is a neighborhood of each point in $\text{int}(e)$ which implies that this set is open. Since the sets $\text{int}(X)$, $\text{int}(e)$, $\text{int}(Y)$ are connected, and each of them meets the connected set W, it follows that their union M is also connected. So M is a region. To see that M is simply connected, first note that the

Schönflies theorem implies there exists a homotopy on $\operatorname{int}(X) \cup \operatorname{int}(e)$ that deforms this set into $\operatorname{int}(e)$ and keeps the points in $\operatorname{int}(e)$ fixed during the homotopy. If we combine this homotopy with a similar homotopy on $\operatorname{int}(Y) \cup \operatorname{int}(e)$, then we see that M is homotopic to $\operatorname{int}(e)$, and hence also to a point. So M is simply connected.

Suppose that Z is another tile in $\mathcal{D}$ with $e \subset \partial Z$. Since $X \cup Y$ contains an open neighborhood for p, there exists a point $x \in \operatorname{int}(Z)$ near p with $x \in X \cup Y$, say $x \in X$. Since the interior of a tile is disjoint from all other cells, we conclude $X = Z$. This shows the uniqueness of X and Y.

(v) Let v be a vertex of $\mathcal{D}$. If an edge e in $\mathcal{D}$ contains v, then v is an endpoint of e and we orient e so that v is the initial point of e. By (ii) there exists a tile X_1 in $\mathcal{D}$ with $v \in X_1$. Then $v \in \partial X_1$, and so by (iii) there exist two edges in ∂X_1 that contain v. For one of these oriented edges, denote it by e_1, the tile X_1 will lie on the right of e_1. Then $v \in e_1 \subset \partial X_1$ and X_1 will lie on the left of the other oriented edge. By (iv) there exists a unique tile $X_2 \neq X_1$ with $e_1 \subset \partial X_2$. Then X_2 will lie on the left of e_1. By (iii) there exists a unique edge $e_2 \subset \partial X_2$ distinct from e_1 with $v \in e_2$. The tile X_2 will lie on the right of e_2. We can continue in this manner to obtain tiles $X_1, X_2, \ldots$ and edges $e_1, e_2, \ldots$ that contain v and satisfy $X_j \neq X_{j+1}$, $e_j \neq e_{j+1}$, and $e_j \subset \partial X_j \cap \partial X_{j+1}$ for all $j \in \mathbb{N}$. Moreover, X_j will lie on the right and X_{j+1} on the left of the oriented edge e_j. Since there are only finitely many tiles, there exists a smallest number $d \in \mathbb{N}$ such that the tiles $X_1, \ldots, X_d$ are all distinct and X_{d+1} is equal to one of the tiles $X_1, \ldots, X_d$. Since $X_1 \neq X_2$, we have $d \geq 2$.

Moreover, $X_{d+1} = X_1$. To see this, we argue by contradiction and assume that X_{d+1} is equal to one of the tiles $X_2, \ldots, X_d$ say $X_{d+1} = X_j$. Note that $X_d \neq X_{d+1}$, so $2 \leq j \leq d-1$. Then $e = e_d$ is an edge with $v \in e$ that is contained in ∂X_d and in $\partial X_{d+1} = \partial X_j$. Hence $e = e_{j-1}$ or $e = e_j$. Since $X_{d+1} = X_j$ lies on the left of $e = e_d$, we must have $e = e_{j-1}$. Then e is contained in the boundary of the three distinct tiles X_{j-1}, X_j, X_d which is impossible by (iv). So indeed $X_{d+1} = X_1$.

By a similar reasoning we can show that the edges $e_1, \ldots, e_d$ are all distinct. Indeed, suppose $e = e_j = e_k$, where $1 \leq j < k \leq d$. Then $k > j+1$ and e is contained in the boundary of the three distinct tiles X_j, X_{j+1}, X_k which is again absurd.

To show that there are no other edges and tiles containing v note that by (iii) the set

$$U = \operatorname{int}(X_1) \cup \operatorname{int}(e_1) \cup \operatorname{int}(X_2) \cup \cdots \cup \operatorname{int}(e_d) \cup \operatorname{int}(X_{d+1})$$

is open. Moreover, its boundary ∂U consists of the point v and a closed set

$$A \subset \bigcup_{j=1}^{d} \partial X_j$$

disjoint from $\{v\}$. Hence v is an isolated boundary point of U which implies that $W = U \cup \{v\}$ is an open neighborhood of v.

If c is an arbitrary cell in $\mathcal{D}$ with $v \in c$ and $c \neq \{v\}$, then $v \in \overline{\mathrm{int}(c)}$. This implies that $\mathrm{int}(c)$ meets U. Since interiors of distinct cells in $\mathcal{D}$ are disjoint, this is only possible if c is equal to one of the edges $e_1, \ldots, e_d$ or one of the tiles $X_1, \ldots, X_d$. The statement follows.

(vi) By (v) every vertex is contained in an edge. Hence the 1-skeleton E of $\mathcal{D}$ is equal to the union of all edges in $\mathcal{D}$. To show that E is connected, let $x, y \in E$ be arbitrary. Since the tiles in $\mathcal{D}$ cover S^2, there exist tiles X and Y with $x \in X$ and $y \in Y$. The interior of each tile is disjoint from the 1-skeleton E, and so $x \in \partial X$ and $y \in \partial Y$. Since S^2 is connected, there exist tiles $X_1, \ldots, X_N$ in $\mathcal{D}$ such that $X_1 = X$, $X_N = Y$, and $X_i \cap X_{i+1} \neq \emptyset$ for $i = 1, \ldots, N-1$. The interior of a tile meets no other tile. Hence $\partial X_i \cap \partial X_{i+1} \neq \emptyset$ for $i = 1, \ldots, N-1$. Since each set ∂X_i is connected, it follows that

$$K = \partial X_1 \cup \cdots \cup \partial X_N$$

is a connected subset of E containing x and y. Hence E is connected. $\square$

Let $d \in \mathbb{N}$, $d \geq 2$, and the tiles X_j and edges e_j for $j \in \mathbb{N}$ be as defined in the proof of statement (v) of the previous lemma. Then we showed that $X_{d+1} = X_1$, but it is useful to point out that actually $X_j = X_{d+j}$ and $e_j = e_{d+j}$ for all $j \in \mathbb{N}$.

Indeed we have seen that $X_{d+1} = X_1$. Moreover, e_1, e_d, e_{d+1} are edges in $\mathcal{D}$ that contain v and are contained in the boundary of the tile $X_1 = X_{d+1}$. Since there are only two such edges, $e_1 \neq e_d$, and $e_d \neq e_{d+1}$, we conclude that $e_{d+1} = e_1$. Then $e_1 = e_{d+1}$ is an edge contained in the boundary of the tiles X_1, X_2, X_{d+2}. Since there are precisely two tiles containing an edge in its boundary, $X_1 \neq X_2$ and $X_1 = X_{d+1} \neq X_{d+2}$ it follows that $X_{d+2} = X_2$.

If we continue in this manner, shifting all indices by 1 in each step, we see that $e_{d+2} = e_2$, $X_{d+3} = X_3$, etc., as claimed.

Note that if we choose the indexing of the edges e_j and X_j as in the proof of statement (v) of the previous lemma, then for each $j \in \mathbb{N}$ the flag $(\{v\}, e_j, X_{j+1})$ in $\mathcal{D}$ is positively-oriented, and flag $(\{v\}, e_j, X_j)$ is negatively-oriented. In other words, if e_j is oriented so that v is the

initial point of e_j, then X_{j+1} lies to the left and X_j lies to the right of e_j.

We will now discuss how cell decompositions can be used to define homeomorphisms and isotopies of the underlying spaces. We start with a general definition.

DEFINITION 5.10 (Isomorphisms of cell complexes). Let $\mathcal{D}$ and $\widetilde{\mathcal{D}}$ be cell complexes. A bijection $\phi\colon \mathcal{D} \to \widetilde{\mathcal{D}}$ is called an *isomorphism (of cell complexes)* if the following conditions are satisfied:

(i) $\dim(\phi(\tau)) = \dim(\tau)$ for all $\tau \in \mathcal{D}$.

(ii) if $\sigma, \tau \in \mathcal{D}$, then $\sigma \subset \tau$ if and only if $\phi(\sigma) \subset \phi(\tau)$.

Let $h\colon \mathcal{X} \to \widetilde{\mathcal{X}}$ be a homeomorphism between two locally compact Hausdorff spaces $\mathcal{X}$ and $\widetilde{\mathcal{X}}$, and suppose $\mathcal{D}$ is a cell decomposition of $\mathcal{X}$. Then it is easy to see that $\widetilde{\mathcal{D}} := \{h(c) \subset \widetilde{\mathcal{X}} : c \in \mathcal{D}\}$ is a cell decomposition of $\widetilde{\mathcal{X}}$ and $\phi\colon \mathcal{D} \to \widetilde{\mathcal{D}}$ given by $\phi(c) = h(c)$ for $c \in \mathcal{D}$ is an isomorphism. We will see that this procedure can be reversed and one can construct a homeomorphism from a given cell complex isomorphism. For simplicity we will restrict ourselves to the case of 2-spheres. As a preparation for the proof of the corresponding Lemma 5.11, we first record some facts about homeomorphisms and isotopies on subsets of 2-spheres.

If α is an arc, then every homeomorphism $\varphi\colon \alpha \to \alpha$ that fixes the endpoints of α is isotopic to the identity rel. $\partial\alpha$. Indeed, we may assume that α is equal to the unit interval $I = [0, 1]$. Then $\varphi(0) = 0$, $\varphi(1) = 1$, and φ is strictly increasing on $[0, 1]$. Define $H\colon I \times I \to I$ by

$$H(s, t) = (1 - t)\varphi(s) + ts$$

for $s, t \in I$. Then for each $t \in I$, the map $H_t = H(\cdot, t)$ is strictly increasing on I. It follows that H is an isotopy. We have $H_t(0) = 0$ and $H_t(1) = 1$ for all $t \in I$, and $H_0 = \varphi$ and $H_1 = \mathrm{id}_I$. Hence φ and id_I are isotopic rel. $\partial I = \{0, 1\}$ by the isotopy H.

Let $X \subset S^2$ be a closed Jordan region. If $h\colon \partial X \times I \to \partial X$ is an isotopy with $h(\cdot, 0) = \mathrm{id}_{\partial X}$, then there exists an isotopy $H\colon X \times I \to X$ such that $H(\cdot, 0) = \mathrm{id}_X$ and $H(p, t) = h(p, t)$ for all $p \in \partial X$ and $t \in I$. So an isotopy h on the boundary of X with $h_0 = \mathrm{id}_{\partial X}$ can be extended to an isotopy H on X with $H_0 = \mathrm{id}_X$. To see this, we may assume that $X = \overline{\mathbb{D}}$. Then H is obtained from h by radial extension; more precisely, we define

$$H(re^{is}, t) = rh(e^{is}, t)$$

for all $r \in [0, 1]$ and $s \in [0, 2\pi]$. Then H is well-defined and it is easy to see that H is an isotopy with the desired properties. By using the

Schönflies theorem and a similar radial extension one can also show that if X and X' are closed Jordan regions in S^2, then every homeomorphism $\varphi \colon \partial X \to \partial X'$ extends to a homeomorphism $\Phi \colon X \to X'$.

If $\varphi \colon X \to X$ is a homeomorphism with $\varphi|\partial X = \mathrm{id}_{\partial X}$, then φ is isotopic to id_X rel. ∂X. Indeed, again we may assume that $X = \overline{\mathbb{D}}$. Then we obtain the desired isotopy by the "Alexander trick": for $z \in \overline{\mathbb{D}}$ and $t \in I$ we define $H(z,t) = t\varphi(z/t)$ if $|z| < t$, and $H(z,t) = z$ if $|z| \geq t$. It is easy to see that H is an isotopy rel. $\partial \mathbb{D}$ with $H_0 = \mathrm{id}_X$ and $H_1 = \varphi$.

If $\varphi, \widetilde{\varphi} \colon X \to X$ are two homeomorphisms with $\varphi|\partial X = \widetilde{\varphi}|\partial X$, then we can apply the previous remark to $\psi = \widetilde{\varphi}^{-1} \circ \varphi$ and conclude that φ and $\widetilde{\varphi}$ are isotopic rel. ∂X.

We are now ready to state and prove a fact that allows us to construct homeomorphisms from cell complex isomorphisms.

LEMMA 5.11. *Let $\mathcal{D}$ and $\widetilde{\mathcal{D}}$ be isomorphic cell decompositions of 2-spheres S^2 and $\widetilde{S}^2$, respectively, and let $\phi \colon \mathcal{D} \to \widetilde{\mathcal{D}}$ be an isomorphism. Then the following statements are true:*

(i) *If $h \colon S^2 \to \widetilde{S}^2$ is a map such that $h|\tau$ is a homeomorphism of τ onto $\phi(\tau)$ for each $\tau \in \mathcal{D}$, then h is a homeomorphism of S^2 onto $\widetilde{S}^2$.*

(ii) *There exists a homeomorphism $h \colon S^2 \to \widetilde{S}^2$ such that $h(\tau) = \phi(\tau)$ for all $\tau \in \mathcal{D}$.*

(iii) *Let $\mathbf{V}$ be the set of vertices of $\mathcal{D}$. If $h_0, h_1 \colon S^2 \to \widetilde{S}^2$ are two homeomorphisms with $h_0(\tau) = \phi(\tau) = h_1(\tau)$ for all $\tau \in \mathcal{D}$, then h_0 and h_1 are isotopic rel. $\mathbf{V}$.*

Note that the map h in (i) is cellular for $(\mathcal{D}, \widetilde{\mathcal{D}})$. If h is as in (ii), then we say that h *realizes* the cell complex isomorphism ϕ. So an isomorphism between cell decompositions of 2-spheres can always be realized by a homeomorphism h and by (iii) this homeomorphism is unique up to isotopy. A similar fact is actually true in greater generality, but Lemma 5.11 will be enough for our purposes.

PROOF. In the following we write $\widetilde{\tau} := \phi(\tau)$ for $\tau \in \mathcal{D}$.

(i) Let $h \colon S^2 \to \widetilde{S}^2$ be a map such that $h|\tau$ is a homeomorphism of τ onto $\widetilde{\tau}$ for each $\tau \in \mathcal{D}$. Then h is continuous, because the restriction $h|\tau$ is continuous for each $\tau \in \mathcal{D}$ and the cells $\tau \in \mathcal{D}$ form a finite cover of S^2 by closed sets. The image cells $\widetilde{\tau} = f(\tau)$ form the cell decomposition $\widetilde{\mathcal{D}}$ of $\widetilde{S}^2$ and hence cover $\widetilde{S}^2$. So h is also surjective.

In order to conclude that $h \colon S^2 \to \widetilde{S}^2$ is a homeomorphism, it suffices to show that h is injective. To see this, let $x_1, x_2 \in S^2$ and assume

that $y := h(x_1) = h(x_2)$. Then there exist unique cells $\tau_1, \tau_2 \in \mathcal{D}$ such that $x_1 \in \operatorname{int}(\tau_1)$ and $x_2 \in \operatorname{int}(\tau_2)$. Since $f|\tau_i$ is a homeomorphism of τ_i onto $\widetilde{\tau}_i$ for $i = 1, 2$, we have $y \in \operatorname{int}(\widetilde{\tau}_1) \cap \operatorname{int}(\widetilde{\tau}_2)$. This implies that $\widetilde{\tau}_1 = \widetilde{\tau}_2$. Since the map $\tau \in \mathcal{D} \mapsto \widetilde{\tau} \in \widetilde{\mathcal{D}}$ is an isomorphism, it follows that $\tau_1 = \tau_2$. So x_1 and x_2 are contained in the same cell $\tau := \tau_1 = \tau_2 \in \mathcal{D}$. Since $h|\tau$ is a homeomorphism onto $\widetilde{\tau}$ and hence injective, we conclude that $x_1 = x_2$ as desired.

(ii) By (i) it suffices to find a map $h: S^2 \to \widetilde{S}^2$ such that $h|\tau$ is a homeomorphism of τ onto $\widetilde{\tau}$ for each $\tau \in \mathcal{D}$. The existence of h follows from the well-known procedure of successive extensions to the skeleta of the cell decomposition $\mathcal{D}$.

Indeed, if v is vertex in $\mathcal{D}$, then there exists a unique vertex $\widetilde{v}$ in $\widetilde{\mathcal{D}}$ such that $\phi(\{v\}) = \{\widetilde{v}\}$. We define $h(v) = \widetilde{v}$. Then h is a bijection of the set of vertices of $\mathcal{D}$ onto the set of vertices in $\widetilde{\mathcal{D}}$. To extend h from the 0-skeleton of $\mathcal{D}$ to the 1-skeleton, let e be an arbitrary edge in $\mathcal{D}$ and u and v be the vertices in $\mathcal{D}$ that are the endpoints of e. Then $\widetilde{u}$ and $\widetilde{v}$ are the endpoints of $\widetilde{e}$. So we can extend h to e by choosing a homeomorphism of e onto $\widetilde{e}$ that agrees with h on the endpoints of e. In this way we can continuously extend f to the 1-skeleton of $\mathcal{D}$ so that $f|\tau$ is a homeomorphism of τ onto $\widetilde{\tau}$, whenever τ is a cell in $\mathcal{D}$ with $\dim(\tau) \leq 1$. An argument as in the proof of (i) shows that h is a homeomorphism of the 1-skeleton of $\mathcal{D}$ onto the 1-skeleton of $\widetilde{\mathcal{D}}$.

If X is an arbitrary tile in $\mathcal{D}$, then ∂X is a subset of the 1-skeleton of $\mathcal{D}$ and hence h is already defined on ∂X. Then $h|\partial X$ is an injective and continuous mapping of ∂X into the boundary $\partial \widetilde{X}$ of the tile $\widetilde{X} \in \widetilde{\mathcal{D}}$. Since an injective and continuous map of a Jordan curve into another Jordan curve is necessarily surjective, $h|\partial X$ is a homeomorphism of ∂X onto $\partial \widetilde{X}$. Hence h can be extended to a homeomorphism of X onto $\widetilde{X}$. These extensions on different tiles paste together to a map $h: S^2 \to S^2$ with the desired property that $h|\tau$ is a homeomorphism of τ onto $\widetilde{\tau}$ for each $\tau \in \mathcal{D}$. By (i) the map h is a homeomorphism of S^2 onto $\widetilde{S}^2$ with $h(\tau) = \widetilde{\tau} = \phi(\tau)$ for $\tau \in \mathcal{D}$.

(iii) Suppose $h_0, h_1: S^2 \to \widetilde{S}^2$ are as in the statement. Then $\varphi := h_1^{-1} \circ h_0$ is a homeomorphism on S^2 that maps each cell $\tau \in \mathcal{D}$ onto itself. In particular, φ is the identity on the set $\mathbf{V}$ of vertices of $\mathcal{D}$.

By successive extensions to the 1- and the 2-skeleton of $\mathcal{D}$ we will show that φ is actually isotopic to id_{S^2} rel. $\mathbf{V}$. For this we denote the set of edges of $\mathcal{D}$ by $\mathbf{E}$ and by $E = \bigcup\{e : e \in \mathbf{E}\}$ the 1-skeleton of $\mathcal{D}$. Let $e \in \mathbf{E}$ be an arbitrary edge in $\mathcal{D}$. Since $\varphi(e) = e$ and φ is the identity on $\mathbf{V}$, the map $\varphi|e$ is isotopic to id_e rel. ∂e. These isotopies

on edges paste together to an isotopy of $\varphi|E$ to id_E rel. $\mathbf{V}$. If X is a tile in $\mathcal{D}$, then this isotopy is defined on $\partial X \subset E$, and we can extend it to an isotopy of a homeomorphism on X that agrees with $\varphi|\partial X$ on ∂X to id_X. These extensions on tiles X paste together to an isotopy $\Phi\colon S^2 \times [0,1] \to S^2$ rel. $\mathbf{V}$ such that $\psi|E = \varphi|E$, where $\psi \coloneqq \Phi(\cdot, 0)$, and $\Phi(\cdot, 1) = \mathrm{id}_{S^2}$.

For each tile $X \in \mathcal{D}$ the maps $\varphi|X$ and $\psi|X$ are homeomorphisms of X onto itself that agree on $\partial X \subset E$. As we have seen in the discussion before the proof of the lemma, this implies that $\varphi|X$ and $\psi|X$ are isotopic rel. ∂X. Again by pasting these isotopies on tiles together, we can find an isotopy $\Psi\colon S^2 \times [0,1] \to S^2$ rel. E with $\Psi(\cdot, 0) = \varphi$ and $\Psi(\cdot, 1) = \psi$. The concatenation of the isotopies Ψ and Φ gives an isotopy rel. $\mathbf{V}$ between $\varphi = h_1^{-1} \circ h_0$ and id_{S^2}. If we postcompose this isotopy by h_1, we get an isotopy between h_0 and h_1 rel. $\mathbf{V}$ as desired. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

5.3. Cell decompositions induced by Thurston maps

Let $f\colon S^2 \to S^2$ be a Thurston map, and $\mathcal{C} \subset S^2$ be a Jordan curve such that $\mathrm{post}(f) \subset \mathcal{C}$. In this section we will discuss how the pair $(f, \mathcal{C})$ induces natural cell decompositions of S^2.

By the Schönflies theorem there are two closed Jordan regions $X_{\mathsf{b}}^0, X_{\mathsf{w}}^0 \subset S^2$ whose boundary is $\mathcal{C}$. Our notation for these regions is suggested by the fact that we often think of X_{b}^0 as being assigned or carrying the color "black", represented by the symbol b, and X_{w}^0 as being colored "white" represented by w. We will discuss this more precisely later in this section (see Lemma 5.21).

The sets X_{b}^0 and X_{w}^0 are topological cells of dimension 2. We call them *tiles of level* 0 or 0-*tiles*. The postcritical points of f are on the boundary of X_{w}^0 and X_{b}^0. We consider them as *vertices* of X_{w}^0 and X_{b}^0, and the closed arcs of $\mathcal{C}$ between vertices as the *edges* of the 0-tiles. In this way, we think of X_{w}^0 and X_{b}^0 as topological m-gons where $m = \#\mathrm{post}(f) \geq 2$ (see Remark 5.16). To emphasize that these edges and vertices belong to 0-tiles, we call them 0-*edges* and 0-*vertices*. A 0-*cell* is a 0-tile, a 0-edge, or a set consisting of a 0-vertex. Obviously, the 0-cells form a cell decomposition of S^2 that we denote by $\mathcal{D}^0 = \mathcal{D}^0(f, \mathcal{C})$. Roughly speaking, we can now obtain cell decompositions $\mathcal{D}^n(f, \mathcal{C})$ of S^2 for each $n \in \mathbb{N}_0$ by taking preimages of $\mathcal{D}^0(f, \mathcal{C})$ under f^n. This is based on the following lemma.

LEMMA 5.12. *Let $f\colon S^2 \to S^2$ be a branched covering map and $\mathcal{D}$ a cell decomposition of S^2 such that every point in $f(\mathrm{crit}(f))$ is a vertex*

in $\mathcal{D}$. Then there exists a unique cell decomposition $\mathcal{D}'$ of S^2 such that f is cellular with respect to $(\mathcal{D}', \mathcal{D})$.

In general, $\mathcal{D}'$ will not be a refinement of $\mathcal{D}$. As we will see in the proof, $\mathcal{D}'$ consists precisely of all cells $c \subset S^2$ such that $f(c)$ is a cell in $\mathcal{D}$ and $f|c$ is a homeomorphism of c onto $f(c)$. In particular, if $\mathbf{V}$ and $\mathbf{V}'$ denote the set of vertices of $\mathcal{D}$ and $\mathcal{D}'$, respectively, then $\mathbf{V}' = f^{-1}(\mathbf{V})$.

If we make the stronger assumption $\mathrm{post}(f) \subset \mathbf{V}$ in Lemma 5.12, then $f^n(\mathrm{crit}(f^n)) \subset \mathbf{V}$ for all $n \in \mathbb{N}$ and we can apply the lemma to all iterates of f.

Before we prove Lemma 5.12, we record some immediate consequences.

COROLLARY 5.13. *Let $f \colon S^2 \to S^2$ be a Thurston map, $\mathcal{C} \subset S^2$ be a Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}$, and $\mathcal{D}^0(f, \mathcal{C})$ be defined as above. Then there exists a unique sequence of cell decompositions $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$, $n \in \mathbb{N}_0$, such that f is cellular with respect to $(\mathcal{D}^{n+1}, \mathcal{D}^n)$ for each $n \in \mathbb{N}_0$.*

PROOF. Since the points in $\mathrm{post}(f) \supset f(\mathrm{crit}(f))$ form the vertices in $\mathcal{D}^0 = \mathcal{D}^0(f, \mathcal{C})$, we can apply Lemma 5.12 to obtain a cell decomposition $\mathcal{D}^1 = \mathcal{D}^1(f, \mathcal{C})$ such that f is cellular with respect to $(\mathcal{D}^1, \mathcal{D}^0)$. By the remark following Lemma 5.12 a point is a vertex in $\mathcal{D}^1$ precisely if its image under f is a vertex of $\mathcal{D}^0$. So the set of vertices of $\mathcal{D}^1$ is given by $f^{-1}(\mathrm{post}(f)) \supset \mathrm{post}(f) \supset f(\mathrm{crit}(f))$. Hence we can apply Lemma 5.12 again and obtain a cell decomposition $\mathcal{D}^2 = \mathcal{D}^2(f, \mathcal{C})$ such that f is cellular with respect to $(\mathcal{D}^2, \mathcal{D}^1)$. Continuing in this manner, we obtain cell decompositions $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$ of S^2 for $n \in \mathbb{N}_0$ such that f is cellular for $(\mathcal{D}^{n+1}, \mathcal{D}^n)$ for all $n \in \mathbb{N}_0$.

The last property uniquely determines the cell decompositions $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$ for all $n \in \mathbb{N}_0$ as follows from the uniqueness statement in Lemma 5.12. □

The cell decompositions $\mathcal{D}^n(f, \mathcal{C})$ will be used throughout this work.

DEFINITION 5.14 (Cell decompositions for f and $\mathcal{C}$). Given a Thurston map $f \colon S^2 \to S^2$ and a Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}$, the cell decompositions $\mathcal{D}^n(f, \mathcal{C})$ for $n \in \mathbb{N}_0$ are the ones provided by Corollary 5.13.

We call the elements in $\mathcal{D}^n(f, \mathcal{C})$ the *n-cells* for $(f, \mathcal{C})$, or simply *n*-cells if f and $\mathcal{C}$ are understood. We call n the *level* of an n-cell. When we speak of n-cells, then n always refers to this level and not to the dimension of the cell. An n-cell of dimension 2 is called an *n-tile*,

and an n-cell of dimension 1 an *n-edge*. An *n-vertex* is a point $p \in S^2$ such that $\{p\}$ is an n-cell of dimension 0. We denote the set of all n-tiles, n-edges, and n-vertices for $(f, \mathcal{C})$ by $\mathbf{X}^n(f, \mathcal{C})$, $\mathbf{E}^n(f, \mathcal{C})$, and $\mathbf{V}^n(f, \mathcal{C})$, respectively. If f and $\mathcal{C}$ are understood, we simply write $\mathbf{X}^n$ for $\mathbf{X}^n(f, \mathcal{C})$, etc.

In Proposition 5.17 we will record the properties of the cell decompositions $\mathcal{D}^n(f, \mathcal{C})$ and also give a more explicit description of their cells. We first turn to the proof of Lemma 5.12. We require a lemma.

LEMMA 5.15. *Let X and Y be metric spaces, and $f\colon X \to Y$ be a continuous map. Suppose X is compact and $A \subset Y$ is closed. Then for each $\epsilon > 0$ there exists $\delta > 0$ such that*

$$f^{-1}(\mathcal{N}_\delta(A)) \subset \mathcal{N}_\epsilon(f^{-1}(A)).$$

Here $\mathcal{N}_r(M)$ for $r > 0$ denotes the open r-neighborhood of a set M in a metric space.

PROOF. We argue by contradiction and assume that for some $\epsilon > 0$ the statement is not true. Then for each $n \in \mathbb{N}$ there exists a point $x_n \in f^{-1}(\mathcal{N}_{1/n}(A))$ with $x_n \notin \mathcal{N}_\epsilon(f^{-1}(A))$. Since X is compact, by passing to a subsequence if necessary, we may assume that $\{x_n\}$ converges, say $x_n \to x \in X$ as $n \to \infty$. Then

$$\operatorname{dist}(x, f^{-1}(A)) = \lim_{n \to \infty} \operatorname{dist}(x_n, f^{-1}(A)) \geq \epsilon > 0.$$

On the other hand, $f(x_n) \in \mathcal{N}_{1/n}(A)$ for $n \in \mathbb{N}$ and $f(x_n) \to f(x)$ as $n \to \infty$. This implies that $f(x) \in \overline{A} = A$, and so $x \in f^{-1}(A)$. This is a contradiction. $\qquad\square$

PROOF OF LEMMA 5.12. To show existence, we define $\mathcal{D}'$ to be the set of all cells $c \subset S^2$ such that $f(c)$ is a cell in $\mathcal{D}$ and $f|c$ is a homeomorphism of c onto $f(c)$. It is clear that $\mathcal{D}'$ does not contain cells of dimension > 2. As usual, we call the cells c in $\mathcal{D}'$ edges or tiles depending on whether c has dimension 1 or 2, respectively. The vertices p of $\mathcal{D}'$ are the points in S^2 such that $\{p\}$ is a cell in $\mathcal{D}'$ of dimension 0. It is clear that the set of vertices of $\mathcal{D}'$ is equal to $f^{-1}(\mathbf{V})$, where $\mathbf{V}$ is the set of vertices of $\mathcal{D}$.

In order to show that $\mathcal{D}'$ is a cell decomposition of S^2, we first establish two claims.

Claim 1. If $p \in S^2$ and $q = f(p) \in \operatorname{int}(X)$ for some tile $X \in \mathcal{D}$, then there exists a unique tile $X' \in \mathcal{D}'$ with $p \in X'$.

In this case, let $U = \operatorname{int}(X)$. Then U is an open and simply connected set in the complement of $\mathbf{V} \supset f(\operatorname{crit}(f))$. Hence there exists a unique continuous map $g\colon U \to U' := g(U)$ (a "branch of the inverse of

f^{-1}") with $f \circ g = \mathrm{id}_U$ and $g(q) = p$. The map g is a homeomorphism onto its image U'. Hence $U' \subset S^2$ is open and simply connected.

We equip S^2 with some base metric inducing the given topology. In the following, metric terms will refer to this metric. We write $\mathcal{N}_\epsilon(A)$ for the open ϵ-neighborhood of a set $A \subset S^2$. Then it follows from Lemma 5.15 that f has the following property: for all $w \in S^2$ and all $\epsilon > 0$, there exists $\delta > 0$ such that

$$(5.1) \qquad f^{-1}(B(w, \delta)) \subset \mathcal{N}_\epsilon(f^{-1}(w)).$$

We want to prove that g has a continuous extension to $\overline{U} = X$. For this it suffices to show that $\{g(w_i)\}$ converges whenever $\{w_i\}$ is a sequence in U converging to a point $w \in \partial U$. Since g is a right inverse of f, it follows that the limit points of $\{g(w_n)\}$ are contained in $f^{-1}(w)$. Since f is finite-to-one, the point w has finitely many preimages $z_1, \ldots, z_m$ under f.

We can choose $\epsilon > 0$ so small that the sets $B(z_i, \epsilon)$, $i = 1, \ldots, m$, are pairwise disjoint. By (5.1) we can find $\delta > 0$ such that

$$(5.2) \qquad f^{-1}(B(w, \delta)) \subset \bigcup_{i=1}^{m} B(z_i, \epsilon).$$

The set $\overline{U} = X$ is a closed Jordan region, and hence locally connected. So there exists an open connected set $V \subset U$ such that $\overline{V}$ is a neighborhood of w in $\overline{U}$ and $\overline{V} \subset B(w, \delta)$. Then $g(V)$ is a connected subset of $f^{-1}(B(w, \delta))$. Since the union on the right hand side of (5.2) is disjoint, the set $g(V)$ must be contained in one of the sets of this union, say $g(V) \subset B(z_k, \epsilon)$. Now $w_i \in V$ for sufficiently large i, and so all limit points of $\{g(w_i)\}$ are contained in $\overline{g(V)} \subset \overline{B}(z_k, \epsilon)$. On the other hand, the only possible limit points of $\{g(w_i)\}$ are $z_1, \ldots, z_m$, and z_k is the only one contained in $\overline{B}(z_k, \epsilon)$. This implies $\{g(w_i)\} \to z_k$. So g has indeed a continuous extension to $\overline{U}$, which is again denoted by g. It is clear that

$$(5.3) \qquad f \circ g = \mathrm{id}_{\overline{U}}.$$

This implies that g is a homeomorphism of $\overline{U} = X$ onto its image $X' := g(\overline{U}) = \overline{g(U)}$. Then X' is a closed Jordan region, and by (5.3) the map $f|X'$ is a homeomorphism of X' onto $\overline{U} = X$. Hence X' is a tile in $\mathcal{D}'$ with $p \in g(U) \subset X'$.

So a tile $X' \in \mathcal{D}'$ containing p exists. We want to show that it is the only tile in $\mathcal{D}'$ containing p. Indeed, suppose $Y' \in \mathcal{D}'$ is another tile with $p \in Y'$. Then $f(Y')$ is a tile in $\mathcal{D}$ containing the point $q = f(p) \in \mathrm{int}(X)$. Hence $f(Y') = X$, and so $f|Y'$ is a homeomorphism of Y' onto

X. Let $h = (f|Y')^{-1}$. Then g and h are both inverse branches of f defined on the simply connected region U with $g(q) = p = h(q)$. Hence h and g agree on U, and so by continuity also on $\overline{U}$. We conclude that $X' = g(X) = h(X) = Y'$ as desired, and so Claim 1 follows.

Claim 2. If $p \in S^2$ and $q = f(p) \in \mathrm{int}(e)$ for some edge $e \in \mathcal{D}$, then there exists a unique edge $e' \in \mathcal{D}'$, and precisely two distinct tiles X' and Y' in $\mathcal{D}'$ that contain p. Moreover, $e' \subset \partial X' \cup \partial Y'$.

By Lemma 5.9 (iv) we know that there are precisely two distinct tiles $X, Y \in \mathcal{D}$ that contain e in their boundary, and that $U = \mathrm{int}(X) \cup \mathrm{int}(e) \cup \mathrm{int}(Y)$ is an open and simply connected region in the complement of the set $\mathbf{V} \supset f(\mathrm{crit}(f))$. Hence there exists a unique continuous map $g \colon U \to S^2$ with $g(q) = p$ and $f \circ g = \mathrm{id}_U$. As before one can show that the maps $g_1 := g|\mathrm{int}(X)$ and $g_2 := g|\mathrm{int}(Y)$ have continuous extensions to X and Y, respectively. We use the same notation g_1 and g_2 for these extensions. It is clear that $g_1|e = g_2|e$. Moreover, g_1 is a homeomorphism of X onto a closed Jordan region $X' = g_1(X)$ with inverse map $f|X'$. In particular, X' is a tile in $\mathcal{D}'$. Similarly, $Y' = g_2(Y)$ is a tile in $\mathcal{D}'$. The tiles X' and Y' are distinct, because f maps them to different tiles in $\mathcal{D}$. Moreover, $e' := g_1(e) = g_2(e)$ is an edge in $\mathcal{D}'$ with $p \in e' \subset \partial X' \cap \partial Y'$.

It remains to prove the uniqueness part. If $\widetilde{e}$ is another edge in $\mathcal{D}'$ with $p \in \widetilde{e}$, then $f(\widetilde{e})$ is an edge in $\mathcal{D}$ and $f|\widetilde{e}$ is a homeomorphism of $\widetilde{e}$ onto $f(\widetilde{e})$. Hence $q = f(p) \in \mathrm{int}(e) \cap f(\widetilde{e})$ which implies that $f(\widetilde{e}) = e$. So $f|\widetilde{e}$ is actually a homeomorphism of $\widetilde{e}$ onto e. Then $(f|\mathrm{int}(e'))^{-1}$ and $(f|\mathrm{int}(\widetilde{e}))^{-1}$ are right inverses of f defined on the open arc $\mathrm{int}(e)$ that both map q to p. Hence these right inverses must agree on $\mathrm{int}(e)$. By continuity this implies $(f|e')^{-1} = (f|\widetilde{e})^{-1}$ on e, and so $e' = (f|e')^{-1}(e) = (f|\widetilde{e})^{-1}(e) = \widetilde{e}$.

If Z' is another tile in $\mathcal{D}'$ with $p \in Z'$, then f maps $\partial Z'$ homeomorphically to the boundary $\partial f(Z')$ of the tile $f(Z') \in \mathcal{D}$. Moreover, $p \in \partial Z'$; for otherwise $f(p)$ would lie in the set $\mathrm{int}(f(Z'))$ which is disjoint of e. It follows that there is an edge in $\mathcal{D}'$ that contains p and is contained in the boundary of $\partial Z'$. Since this edge in $\mathcal{D}'$ is unique, as we have just seen, we know that $e' \subset \partial Z'$. Note that $X' \cup Y' \supset g(U) \supset \mathrm{int}(e')$, and so $X' \cup Y'$ contains an open neighborhood for each point in $\mathrm{int}(e')$. Since $e' \subset \partial Z'$ there exists a point $x \in \mathrm{int}(Z')$ near p with $x \in X' \cup Y'$, say $x \in X'$. Then $f(x)$ is contained in the interior of the tile $f(Z') \in \mathcal{D}$. Since $x \in X' \cap Z'$ and X' and Z' are both tiles in $\mathcal{D}'$, we conclude $X' = Z'$ by Claim 1. This completes the proof of Claim 2.

Now that we have established the claims, we can show that $\mathcal{D}$ is a cell decomposition of S^2 by verifying conditions (i)–(iv) of Definition 5.1.

Condition (i): If $p \in S^2$ is arbitrary, then $f(p)$ is a vertex of $\mathcal{D}$ or $f(p)$ lies in the interior of an edge or in the interior of a tile in $\mathcal{D}$. In the first case p is a vertex of $\mathcal{D}'$, and in the other two cases p lies in cells in $\mathcal{D}'$ by Claim 1 and Claim 2. It follows that the cells in $\mathcal{D}'$ cover S^2.

Condition (ii): Let σ, τ be cells in $\mathcal{D}'$ with $\mathrm{int}(\sigma) \cap \mathrm{int}(\tau) \neq \emptyset$. Then $f(\sigma)$ and $f(\tau)$ are cells in $\mathcal{D}$ with $\mathrm{int}(f(\sigma)) \cap \mathrm{int}(f(\tau)) \neq \emptyset$. Hence $\lambda := f(\sigma) = f(\tau)$ and $\lambda \in \mathcal{D}$. In particular, σ and τ have the same dimension.

If σ and τ are both tiles, then $\sigma = \tau$ by Claim 1, because every point in $\mathrm{int}(\sigma) \cap \mathrm{int}(\tau) \neq \emptyset$ has an image under f in $\mathrm{int}(\lambda)$. Similarly, if σ and τ are edges, then $\sigma = \tau$ by Claim 2.

If σ and τ consist of vertices in $\mathcal{D}'$, then the relation $\mathrm{int}(\sigma) \cap \mathrm{int}(\tau) \neq \emptyset$ trivially implies $\sigma = \tau$.

Condition (iii): Let $\tau' \in \mathcal{D}'$ be arbitrary. Then $f|\tau'$ is a homeomorphism of τ' onto the cell $\tau = f(\tau') \in \mathcal{D}$. Note that $(f|\tau')^{-1}(\sigma) \in \mathcal{D}'$ whenever $\sigma \in \mathcal{D}$ and $\sigma \subset \tau$. Since $\partial\tau' = (f|\tau')^{-1}(\partial\tau)$ and $\partial\tau$ is a union of cells in $\mathcal{D}$, it follows that $\partial\tau'$ is a union of cells in $\mathcal{D}'$.

Condition (iv): To establish the final property of a cell decomposition for $\mathcal{D}'$, we will show that $\mathcal{D}'$ consists of only finitely many cells. Indeed, let $N_i \in \mathbb{N}$ be the number of cells of dimension i in $\mathcal{D}$ for $i = 0, 1, 2$. Since the vertices in $\mathcal{D}'$ are the preimages of the vertices of $\mathcal{D}$, we have at most $\deg(f)N_0$ vertices in $\mathcal{D}'$.

Pick one point in the interior of each edge in $\mathcal{D}$. The set M of these points consists of N_1 elements. If $q \in M$, then $q \notin \mathbf{V} \supset f(\mathrm{crit}(f))$, and so q is not a critical value of f. Hence $\#f^{-1}(M) = N_1 \deg(f)$. It follows from Claim 2 that each element of $f^{-1}(M)$ is contained in a unique edge in $\mathcal{D}'$, and it follows from the definition of $\mathcal{D}'$ that each edge in $\mathcal{D}'$ contains a unique point in $f^{-1}(M)$. Hence the number of edges in $\mathcal{D}'$ is equal to $\#f^{-1}(M) = N_1 \deg(f)$.

Similarly, pick a point in the interior of each tile in $\mathcal{D}$ and let M be the set of these points. Then $\#f^{-1}(M) = N_2 \deg(f)$ and by a similar reasoning as above based on Claim 1, we see that the number of tiles in $\mathcal{D}'$ is equal to $\#f^{-1}(M) = N_2 \deg(f)$.

We have shown that $\mathcal{D}'$ is a cell decomposition of S^2. It follows immediately from the definition of $\mathcal{D}'$ that f is cellular with respect to $(\mathcal{D}', \mathcal{D})$.

To show uniqueness of $\mathcal{D}'$, suppose that $\widetilde{\mathcal{D}}$ is another cell decomposition such that f is cellular with respect to $(\widetilde{\mathcal{D}}, \mathcal{D})$. Then by definition of $\mathcal{D}'$ every cell in $\widetilde{\mathcal{D}}$ also lies in $\mathcal{D}'$. So we have $\widetilde{\mathcal{D}} \subset \mathcal{D}'$. If this inclusion were strict, then there would be a cell $\tau \in \mathcal{D}'$ whose interior $\mathrm{int}(\tau) \neq \emptyset$ is disjoint from the interior of all cells in $\widetilde{\mathcal{D}}$. This is impossible, since these interiors form a cover of S^2. Hence $\widetilde{\mathcal{D}} = \mathcal{D}'$. $\hspace{1cm}$ $\square$

REMARK 5.16. No Thurston maps $f\colon S^2 \to S^2$ with $\#\mathrm{post}(f) \in \{0,1\}$ exist. Indeed, suppose that f is such a map. Then $U := S^2 \setminus \mathrm{post}(f)$ is simply connected, and so there exists a continuous map $g\colon U \to S^2$ with $f \circ g = \mathrm{id}_U$.

If $\#\mathrm{post}(f) = 0$, we have $U = S^2$ and so we conclude that g is a homeomorphism onto its image. This image must be all of S^2. Hence g and f are homeomorphisms, contradicting our assumption $\deg(f) \geq 2$ (see Definition 2.1).

If $\#\mathrm{post}(f) = 1$, we have $U = S^2 \setminus \{p\}$ for some $p \in S^2$. Then by an argument as in the proof of Claim 1 in Lemma 5.12, one can show that g has a continuous extension to the point p, and hence to S^2. If we denote this extension to S^2 also by g, then $f \circ g = \mathrm{id}_{S^2}$, and again we conclude that f is a homeomorphism and obtain a contradiction.

We now return to the cell decompositions $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$ from Definition 5.14.

PROPOSITION 5.17. *Let* $k, n \in \mathbb{N}_0$, *let* $f\colon S^2 \to S^2$ *be a Thurston map,* $\mathcal{C} \subset S^2$ *be a Jordan curve with* $\mathrm{post}(f) \subset \mathcal{C}$, $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$, *and* $m = \#\mathrm{post}(f)$. *Then the following statements are true:*

(i) *The map* f^k *is cellular with respect to* $(\mathcal{D}^{n+k}, \mathcal{D}^n)$. *In particular, if* τ *is any* $(n+k)$-*cell, then* $f^k(\tau)$ *is an* n-*cell, and* $f^k|\tau$ *is a homeomorphism of* τ *onto* $f^k(\tau)$.

(ii) *Let* σ *be an* n-*cell. Then* $f^{-k}(\sigma)$ *is equal to the union of all* $(n+k)$-*cells* τ *with* $f^k(\tau) = \sigma$.

(iii) *The* 0-*skeleton (i.e., the set of vertices) of* $\mathcal{D}^n$ *is given by* $\mathbf{V}^n = f^{-n}(\mathrm{post}(f))$, *and we have* $\mathbf{V}^n \subset \mathbf{V}^{n+k}$. *The* 1-*skeleton of* $\mathcal{D}^n$ *is equal to* $f^{-n}(\mathcal{C})$.

(iv) *We have* $\#\mathbf{V}^n \leq m \deg(f)^n$, $\#\mathbf{E}^n = m \deg(f)^n$, *and* $\#\mathbf{X}^n = 2 \deg(f)^n$ *for* $n \in \mathbb{N}_0$.

(v) *The* n-*edges are precisely the closures of the connected components of* $f^{-n}(\mathcal{C}) \setminus f^{-n}(\mathrm{post}(f))$. *The* n-*tiles are precisely the closures of the connected components of* $S^2 \setminus f^{-n}(\mathcal{C})$.

(vi) *Every* n-*tile is an* m-*gon, i.e., the number of* n-*edges and* n-*vertices contained in its boundary is equal to* m.

(vii) *Let $F = f^k$ be an iterate of f with $k \geq 1$. Then $\mathcal{D}^n(F, \mathcal{C}) = \mathcal{D}^{nk}$.*

In the proof we will use the following fact about open maps $g \colon S^2 \to S^2$ (such as iterates of Thurston maps): if $A \subset S^2$ is arbitrary, then

$$(5.4) \qquad\qquad g^{-1}(\overline{A}) \subset \overline{g^{-1}(A)}.$$

Indeed, if U is an open neighborhood of a point $p \in g^{-1}(\overline{A})$, then $g(U)$ is an open neighborhood of $g(p) \in \overline{A}$. Hence there exists a point $p' \in U$ with $g(p') \in A$, and so $U \cap g^{-1}(A) \neq \emptyset$. The inclusion (5.4) follows.

PROOF. We know that $\mathbf{V}^0 = \operatorname{post}(f) \supset f(\operatorname{crit}(f))$ is the set of vertices of $\mathcal{D}^0$ and that f is cellular with respect to $(\mathcal{D}^{n+1}, \mathcal{D}^n)$ for each $n \in \mathbb{N}_0$. As we have seen in the proof of Lemma 5.12, this implies $\mathbf{V}^{n+1} = f^{-1}(\mathbf{V}^n)$. It follows by induction that $\mathbf{V}^n = f^{-n}(\operatorname{post}(f))$ for $n \in \mathbb{N}_0$. After this preliminary remark, we now turn to the proofs of the statements.

(i) This immediately follows from the facts that f is cellular with respect to $(\mathcal{D}^{n+1}, \mathcal{D}^n)$ for each n, and that compositions of cellular maps are cellular (if, as in our case, the obvious compatibility requirement for the cell decompositions involved is satisfied).

(ii) Note that the set $\mathbf{V}^n = f^{-n}(\operatorname{post}(f)) \supset \operatorname{post}(f)$ of vertices of $\mathcal{D}^n$ contains the critical values $f^k(\operatorname{crit}(f^k)) \subset \operatorname{post}(f)$ of f^k. So we can apply Lemma 5.12 and conclude from from (i) that $\mathcal{D}^{n+k}$ is the unique cell decomposition of S^2 such that f^k is cellular with respect to $(\mathcal{D}^{n+k}, \mathcal{D}^n)$. Moreover, recall from the proof of Lemma 5.12 that a topological cell $c \subset S^2$ is an $(n+k)$-cell if and only if $f^k(c)$ is an n-cell and $f^k|c$ is a homeomorphism of c onto $f^k(c)$.

This immediately implies the statement if $\sigma = \{q\}$, where q is an n-vertex.

Suppose σ is equal to an n-edge e. Let M be the union of all $(n+k)$-edges e' with $f^k(e') = e$. It is clear that $M \subset f^{-k}(e)$.

To see the reverse inclusion, first note that because there are only finitely many $(n+k)$-edges, the set M is closed.

Let $p \in f^{-k}(\operatorname{int}(e))$ be arbitrary. Then from Claim 2 in the proof of Lemma 5.12 it follows that there exists an $(n+k)$-edge e' with $p \in e'$. Then $f^k(e')$ is an n-edge that contains $q = f^k(p) \in \operatorname{int}(e)$. Hence $e = f^k(e')$, and so $f^{-k}(\operatorname{int}(e)) \subset M$. Since f^k is an open and continuous map and M is closed, it follows from (5.4) that

$$f^{-k}(e) = f^{-k}(\overline{\operatorname{int}(e)}) \subset \overline{f^{-k}(\operatorname{int}(e))} \subset \overline{M} = M.$$

Hence $M = f^{-k}(e)$ as desired.

If σ is equal to an n-tile X, let M be the union of all $(n+k)$-tiles X' with $f^k(X') = X$. Then $M \subset f^{-k}(X)$ and M is closed.

If $p \in f^{-k}(\mathrm{int}(X))$, then by Claim 1 in the proof of Lemma 5.12 there exists an $(n+k)$-tile with $p \in X'$. As above, we conclude $f^k(X') = X$, and so $p \in M$. Hence $f^{-k}(\mathrm{int}(X)) \subset M$. Now again by (5.4) we have
$$f^{-k}(X) = f^{-k}(\overline{\mathrm{int}(X)}) \subset \overline{f^{-k}(\mathrm{int}(X))} \subset \overline{M} = M.$$
We conclude that $M = f^{-k}(X)$ as desired.

(iii) The 0-skeleton of $\mathcal{D}^n$ is the set $\mathbf{V}^n$ of all vertices of $\mathcal{D}^n$. We have already seen that $\mathbf{V}^n = f^{-n}(\mathrm{post}(f))$. Moreover,
$$f^{n+k}(\mathbf{V}^n) = f^{n+k}(f^{-n}(\mathrm{post}(f))) \subset f^k(\mathrm{post}(f)) \subset \mathrm{post}(f),$$
and so $\mathbf{V}^n \subset f^{-n-k}(\mathrm{post}(f)) = \mathbf{V}^{n+k}$.

The 1-skeleton of $\mathcal{D}^n$ is equal to the set consisting of all n-vertices and the union of all n-edges. As follows from (ii) this set is equal to the preimage of the 1-skeleton of $\mathcal{D}^0$ under the map f^n. Since the 1-skeleton of $\mathcal{D}^0$ is equal to $\mathcal{C}$, it follows that the 1-skeleton of $\mathcal{D}^n$ is equal to $f^{-n}(\mathcal{C})$.

(iv) Note that f^n is cellular for $(\mathcal{D}^n, \mathcal{D}^0)$. Moreover, $\deg(f^n) = \deg(f)^n$, $\#\mathbf{V}^0 = m$, $\#\mathbf{E}^0 = m$, and $\#\mathbf{X}^0 = 2$. The statements about $\mathbf{V}^n, \mathbf{E}^n$, and $\mathbf{X}^n$, then follow from the corresponding statement established in the last part of the proof of Lemma 5.12.

(v) This immediately follows from (iii) and Lemma 5.5.

(vi) If X is an n-tile, then $f^n|X$ is a homeomorphism of X onto the 0-tile $f^n(X)$. The n-vertices contained in X are precisely the preimages of the 0-vertices contained in $f^n(X)$; hence X contains exactly $m = \#\mathrm{post}(f)$ n-vertices, and hence also the same number of n-edges (Lemma 5.9 (iii)). So every n-tile is an m-gon.

(vii) We know that $F = f^k$ is a Thurston map with $\mathrm{post}(F) = \mathrm{post}(f)$ (see Section 2.2). It follows that $\mathcal{D}^0(F, \mathcal{C}) = \mathcal{D}^0$ and that every point in $F^n(\mathrm{crit}(F^n)) \subset \mathrm{post}(F) = \mathrm{post}(f)$ is a vertex of $\mathcal{D}^0(F, \mathcal{C})$. By (i) the map $F^n = f^{nk}$ is cellular for $(\mathcal{D}^n(F, \mathcal{C}), \mathcal{D}^0(F, \mathcal{C}))$ and also cellular for $(\mathcal{D}^{nk}, \mathcal{D}^0(F, \mathcal{C}))$. Hence $\mathcal{D}^n(F, \mathcal{C}) = \mathcal{D}^{nk}$ by the uniqueness statement in Lemma 5.12. $\quad\square$

Instead of an inequality for $\#\mathbf{V}^n$ as in (iv) one can easily give a precise formula for this number; namely, if we set $d = \deg(f)$, and $m = \#\mathrm{post}(f)$, then $\#\mathbf{X}^n = 2d^n$ and $\mathbf{E}^n = md^n$. Moreover, by Euler's polyhedral formula we have
$$\#\mathbf{X}^n - \#\mathbf{E}^n + \#\mathbf{V}^n = 2,$$

and so
$$\#\mathbf{V}^n = (m-2)d^n + 2.$$

We record another lemma that relates cells with the mapping properties of a given Thurston map.

LEMMA 5.18. *Let $k, n \in \mathbb{N}_0$, $f \colon S^2 \to S^2$ be a Thurston map, and $\mathcal{C} \subset S^2$ be a Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}$.*

    (i) *If $c \subset S^2$ is a topological cell such that $f^n|c$ is a homeomorphism onto its image and $f^n(c)$ is a k-cell, then c is an $(n+k)$-cell.*

    (ii) *If X' is a k-tile and $p \in S^2$ is a point with $f^n(p) \in \mathrm{int}(X')$, then there exists a unique $(n+k)$-tile X with $p \in X$ and $f^n(X) = X'$.*

Here it is understood that all m-cells, $m \in \mathbb{N}_0$, are for $(f, \mathcal{C})$. A priori this is not true for the cell c; the point of (i) is to give a criterion when c is a cell for $(f, \mathcal{C})$ (of appropriate level).

PROOF. Let $\mathcal{D}^m = \mathcal{D}^m(f, \mathcal{C})$ for all $m \in \mathbb{N}_0$ according to Definition 5.14. Note that f^n is cellular for $(\mathcal{D}^{n+k}, \mathcal{D}^k)$, and the set $f^{-k}(\mathrm{post}(f)) \supset \mathrm{post}(f)$ of vertices of $\mathcal{D}^k$ contains the set $f^n(\mathrm{crit}(f^n)) \subset \mathrm{post}(f)$ (the last inclusion follows from (2.6)). Hence we are in the situation of Lemma 5.12 with $\mathcal{D} = \mathcal{D}^k$ and $\mathcal{D}' = \mathcal{D}^{n+k}$.

Then (i) follows from the uniqueness statement of Lemma 5.12 and the definition of $\mathcal{D}'$ in the first paragraph of the proof of this lemma.

Moreover, under the assumptions of (ii) it follows from Claim 1 in the proof of Lemma 5.12 that there exists a unique $(n+k)$-tile X with $p \in X$. Then $f^n(X)$ is a k-tile containing $f^n(p) \in \mathrm{int}(X')$, and so $f^n(X) = X'$. □

Tiles can be used to create connections between sets and points. To make this precise, we will now introduce various notions of *chains*. For an illustration of the following definitions see Figure 5.2.

DEFINITION 5.19 (Chains). A *chain* P in S^2 is a finite sequence $A_1, \ldots, A_N$ of sets in S^2 such that $A_i \cap A_{i+1} \neq \emptyset$ for $i = 1, \ldots, N-1$. We call $N = \mathrm{length}(P)$ the *length* of the chain. It *joins* two sets A and B in S^2, if $A \cap A_1 \neq \emptyset$ and $B \cap A_N \neq \emptyset$. Similarly, P *joins* the points $x, y \in S^2$ if $x \in A_1$ and $y \in A_N$ (so P joins $\{x\}$ and $\{y\}$).

A *subchain* P' of P is a chain $A_{j_1}, \ldots, A_{j_M}$ with $1 \leq j_1 < j_2 < \cdots < j_M \leq N$. We say that the chain P joining the sets A and B is *simple* if there is no proper subchain P' of P that joins A and B. Clearly, if a chain P joins the sets A and B, then there is a simple subchain P'

of P that joins A and B. The same terminology and a similar remark apply to chains joining two points $x, y \in S^2$.

We will often consider chains $X_1, \ldots, X_N$, where the sets X_j are tiles in a given cell decomposition of S^2. In this case, P is called a *chain of tiles* or a *tile chain*.

If each set X_j of the chain is an n-tile for a given Thurston map $f \colon S^2 \to S^2$ and a Jordan curve $\mathcal{C} \subset S^2$ with $\operatorname{post}(f) \subset \mathcal{C}$, then we call the chain (with $(f, \mathcal{C})$ understood) a *chain of n-tiles* or an *n-chain*.

DEFINITION 5.20 (*e*-chains). Let $\mathcal{D}$ be a cell decomposition of S^2 and P be a tile chain consisting of the tiles $X_1, \ldots, X_N$. If P has the additional property that for each $j = 1, \ldots, N - 1$ we have $X_j \neq X_{j+1}$ and there is an edge e_j in $\mathcal{D}$ with $e_j \subset \partial X_j \cap \partial X_{j+1}$, then P is called an *e-chain*.

With the given vertices and edges the 1-skeleton of $\mathcal{D}$ can be considered as a graph embedded in S^2. Then the *dual graph* has the set of tiles as vertices, and two vertices as represented by tiles are joined by an edge if the tiles both contain an edge $e \in \mathcal{D}$ in their boundaries. Then an *e*-chain in $\mathcal{D}$ is essentially a path in this dual graph. Having this interpretation in mind, we say that the *e*-chain P given by $X_1, \ldots, X_N$ *joins* the tiles $X = X_1$ and $Y = X_N$. Note that tiles $X' \neq X$ and $Y' \neq Y$ are not joined by this *e*-chain according to our definition. If $1 \leq i \leq j \leq N$, then $X_i, \ldots, X_j$ is a subchain of P; it is an *e*-chain that joins X_i and X_j.

An *e*-chain for a given Thurston map $f \colon S^2 \to S^2$ and a Jordan curve $\mathcal{C} \subset S^2$ with $\operatorname{post}(f) \subset \mathcal{C}$, is an *e*-chain in one of the cell decompositions $\mathcal{D} = \mathcal{D}^n(f, \mathcal{C})$, $n \in \mathbb{N}_0$. In particular, the tiles in such an *e*-chain are of the same level n.

If X is an arbitrary tile in a cell decomposition $\mathcal{D}$ of S^2, then every tile Y in $\mathcal{D}$ can be joined to X by an *e*-chain. This follows from the fact that the union of the tiles Y that can be joined to X is equal to S^2; indeed, this union is a non-empty closed set; it is also open, as follows from Lemma 5.9 (vi) and (v). Hence the union is all of S^2. In general, we call a set M of tiles *e-connected* if every two tiles in M can be joined by an *e*-chain consisting of tiles in M.

Let $f \colon S^2 \to S^2$ be a Thurston map, and $\mathcal{C} \subset S^2$ be a Jordan curve with $\operatorname{post}(f) \subset \mathcal{C}$. It is often useful, in particular in graphical representations, to assign to each tile in $\mathcal{D}^n(f, \mathcal{C})$ one of the two colors "black" and "white" represented by the symbols $\mathtt{b}$ and $\mathtt{w}$, respectively. To formulate this, we denote by $\mathbf{X}^\infty$ the disjoint union of the sets $\mathbf{X}^n = \mathbf{X}^n(f, \mathcal{C})$, $n \in \mathbb{N}_0$. More informally, $\mathbf{X}^\infty$ is the set of all tiles for $(f, \mathcal{C})$. Note that in general a set can be a tile for different levels n, so the
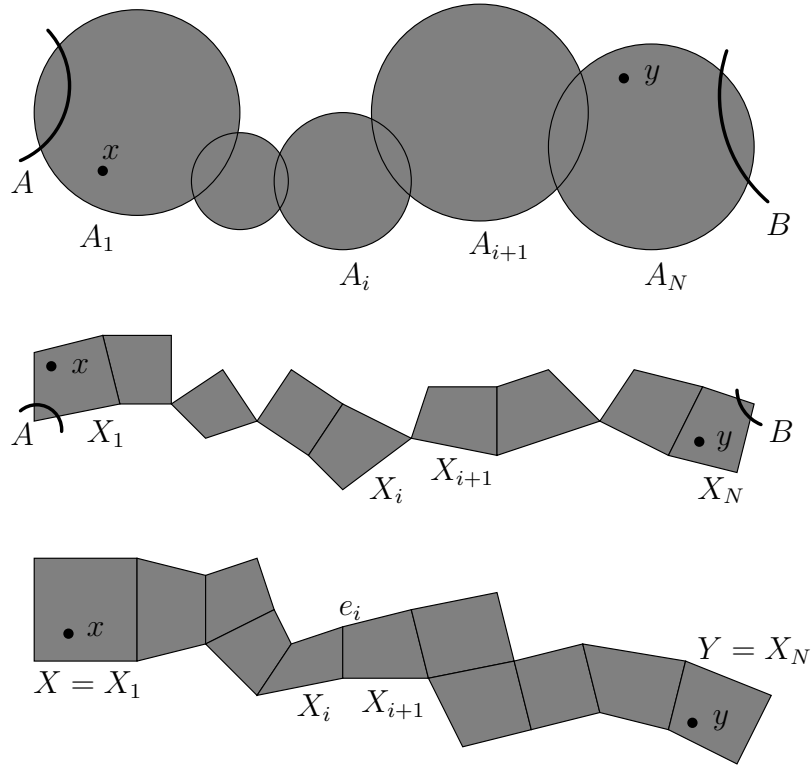
FIGURE 5.2. A chain, an n-chain, and an e-chain.

same tile may be represented by multiple copies in $\mathbf{X}^\infty$ distinguished by their levels.

LEMMA 5.21 (Colors of tiles). *There exists a map $L\colon \mathbf{X}^\infty \to \{\mathtt{b}, \mathtt{w}\}$ with the following properties:*

(i) $L(X_{\mathtt{b}}^0) = \mathtt{b}$ *and* $L(X_{\mathtt{w}}^0) = \mathtt{w}$.

(ii) *If* $n, k \in \mathbb{N}_0$, $X^{n+k} \in \mathbf{X}^{n+k}$, *and* $X^n = f^k(X^{n+k}) \in \mathbf{X}^n$, *then* $L(X^n) = L(X^{n+k})$.

(iii) *If* $n \in \mathbb{N}_0$, *and* X^n *and* Y^n *are two distinct n-tiles that have an n-edge in common, then* $L(X^n) \neq L(Y^n)$.

Moreover, L is uniquely determined by properties (i) *and* (ii).

So with the normalization (i) one can uniquely assign colors "black" or "white" to the tiles so that all iterates of f are color-preserving as in (ii). By (iii) colors of distinct n-tiles are different if they share an n-edge.

To find the number of white n-tiles for $n \in \mathbb{N}_0$, pick a point $p \in \operatorname{int}(X_{\mathtt{w}}^0) \subset S^2 \setminus \operatorname{post}(f)$. Then p is not a critical value of f^n,

and so $\#f^{-n}(p) = \deg(f)^n$. On the other hand, it follows from Proposition 5.17 (i) that each white n-tile X^n contains a unique point $q \in f^{-n}(p)$. It lies in the interior of X^n, and so for different n-tiles these points are distinct; moreover, by Lemma 5.18 (ii) each $q \in \#f^{-n}(p)$ is contained in a unique white n-tile. So we have a bijection between the set of white n-tiles and $f^{-n}(p)$. Hence the number of white n-tiles is equal to $\#f^{-n}(p) = \deg(f)^n$. A similar argument shows that the number of black n-tiles is also equal to $\deg(f)^n$.

Our notion of colorings of tiles is related to the more general concept of a *labeling* of cells in a cell decomposition (see the next Section 5.4, in particular Lemma 5.23).

PROOF OF LEMMA 5.21. To define L we assign colors to the two 0-tiles $X_{\mathtt{b}}^0$ and $X_{\mathtt{w}}^0$ as in (i). If Z^n is an n-tile for some arbitrary level $n \geq 0$, then $f^n(Z^n)$ is a 0-tile (Proposition 5.17 (i)), and so it already has a color assigned. We set $L(Z^n) := L(f^n(Z^n))$.

This defines a map $L\colon \mathbf{X}^\infty \to \{\mathtt{b}, \mathtt{w}\}$. By definition, L has property (i). To show (ii), assume that $n, k \in \mathbb{N}_0$ and $X^{n+k} \in \mathbf{X}^{n+k}$. Then by Proposition 5.17 (i), we have $X^n := f^k(X^{n+k}) \in \mathbf{X}^n$, and $f^{n+k}(X^{n+k}), f^n(X^n) \in \mathbf{X}^0$. So by definition of L we have

$$L(X^n) = L(f^n(X^n)) = L(f^n(f^k(X^{n+k})))$$
$$= L(f^{n+k}(X^{n+k})) = L(X^{n+k})$$

as desired.

Let X^n and Y^n be as in (iii). Then again by Proposition 5.17 (i), we have $f^n(X^n), f^n(Y^n) \in \mathbf{X}^0$. There exists an n-edge e such that $e \subset \partial X^n \cap \partial Y^n$. We orient e so that X^n lies to the left of e. The set of n-vertices is equal to $f^{-n}(\text{post}(f))$ and disjoint from $\text{int}(e)$. It follows that no point in $\text{int}(e)$ is a critical point of f^n. In particular, f^n is a local homeomorphism near each point in $\text{int}(e)$ which implies that $f^n(X^n)$ and $f^n(Y^n)$ are distinct. We conclude that $L(f^n(X^n)) \neq L(f^n(Y^n))$, and so by definition of L we have

$$L(X^n) = L(f^n(X^n)) \neq L(f^n(Y^n)) = L(Y^n)$$

as desired. It follows that L has the properties (i)–(iii).

It is clear that L is uniquely determined by (i) and (ii).   □

If the tiles in a cell decomposition $\mathcal{D}$ of a 2-sphere S^2 are assigned colors "black" and "white" so that two distinct tiles sharing an edge have different colors, then we say the cell decomposition is a *checkerboard tiling* of S^2. It is clear that for the existence of such a coloring the length of the cycle of each vertex in $\mathcal{D}$ has to be even. In Lemma 5.23 we will see that this necessary condition is also sufficient.

If there exists $m \in \mathbb{N}$, $m \geq 2$, such that each tile in $\mathcal{D}$ is an m-gon, then we say that $\mathcal{D}$ is a *tiling by m-gons*. With this terminology we can summarize some of the main results of this section by saying that the cell decompositions $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$ (with the colorings given by the previous lemma) are checkerboard tilings by m-gons, where $m = \# \operatorname{post}(f)$.

5.4. Labelings

Suppose $\mathcal{D}^0$ and $\mathcal{D}^1$ are cell decompositions of a 2-sphere S^2. In Section 5.5 we will see that under suitable conditions one can construct a Thurston map that is cellular for $(\mathcal{D}^1, \mathcal{D}^0)$. If one wants to obtain a unique map up to Thurston equivalence, one needs additional data; namely, for each cell in $\mathcal{D}^1$ we have to assign an image in $\mathcal{D}^0$. The necessary properties of such assignments can be abstracted in the notion of a *labeling*.

DEFINITION 5.22 (Labelings). Let $\mathcal{D}^1$ and $\mathcal{D}^0$ be cell complexes. Then a *labeling* of $(\mathcal{D}^1, \mathcal{D}^0)$ is a map $L \colon \mathcal{D}^1 \to \mathcal{D}^0$ satisfying the following conditions:

(i) $\dim(L(\tau)) = \dim(\tau)$ for all $\tau \in \mathcal{D}^1$.

(ii) if $\sigma, \tau \in \mathcal{D}^1$ and $\sigma \subset \tau$, then $L(\sigma) \subset L(\tau)$.

(iii) if $\sigma, \tau, c \in \mathcal{D}^1$, $\sigma, \tau \subset c$, and $L(\sigma) = L(\tau)$, then $\sigma = \tau$.

So a labeling is a map $L \colon \mathcal{D}^1 \to \mathcal{D}^0$ that preserves inclusions and dimensions of cells, and is "injective on cells" $c \in \mathcal{D}^1$ in the sense of (iii). In particular, every cell of dimension 0 in $\mathcal{D}^1$ is mapped to a cell of dimension 0 in $\mathcal{D}^0$. If v is a *vertex* in $\mathcal{D}^1$, i.e., if $\{v\}$ is a cell of dimension 0 in $\mathcal{D}^1$, then we can write $L(\{v\}) = \{w\}$, where w is a vertex in $\mathcal{D}^0$. We define $L(v) = w$. In the following, we always assume that a labeling $L \colon \mathcal{D}^1 \to \mathcal{D}^0$ has been extended to the set of vertices of $\mathcal{D}^1$ in this way; this will allow us to ignore the subtle distinction between vertices and cells of dimension 0, i.e., sets consisting of one vertex.

Let S^2 be an oriented 2-sphere, and $\mathcal{D}$ be a cell decomposition of S^2. Recall (see Section 5.2) that a flag in $\mathcal{D}$ is a triple (c_0, c_1, c_2), where c_i is a cell in $\mathcal{D}$ of dimension i for $i = 0, 1, 2$ and $c_0 \subset c_1 \subset c_2$. If $L \colon \mathcal{D}^1 \to \mathcal{D}^0$ is a labeling of a pair $(\mathcal{D}^1, \mathcal{D}^0)$ of cell decompositions of S^2 and (c_0, c_1, c_2) is a flag in $\mathcal{D}^1$, then $(L(c_0), L(c_1), L(c_2))$ is a flag in $\mathcal{D}^0$. This follows from the definition of a labeling. So a labeling maps "flags to flags". We say that the labeling is *orientation-preserving* if it maps flags in $\mathcal{D}^1$ to flags in $\mathcal{D}^0$ of the same (positive or negative)

orientation. Here the orientation of a flag is determined by the given orientation of the underlying 2-sphere S^2.

If $f\colon S^2 \to S^2$ is cellular for $(\mathcal{D}^1, \mathcal{D}^0)$, then the *labeling induced by* f is the labeling $L\colon \mathcal{D}^1 \to \mathcal{D}^0$ given by $L(\tau) = f(\tau)$ for $\tau \in \mathcal{D}^1$. Moreover, suppose addition that in $f|X$ is orientation-preserving for each tile X in $\mathcal{D}^1$ in the sense that f preserves the orientation of flags contained in X (which is always true if f is a branched covering map). Then the labeling L induced by f is orientation-preserving.

If a labeling $L\colon \mathcal{D}^1 \to \mathcal{D}^0$ is given, then we say that a map $f\colon S^2 \to S^2$ that is cellular for $(\mathcal{D}^1, \mathcal{D}^0)$ is *compatible* with the labeling L if $L(\tau) = f(\tau)$ for each $\tau \in \mathcal{D}^1$, i.e., if the labeling induced by f is equal to the given labeling.

Let X be a closed Jordan region on the oriented 2-sphere S^2 with $k \geq 3$ distinct points $v_0, \dots, v_{k-1}, v_k = v_0$. Here the indices are elements of $\mathbb{Z}_k = \{0, 1, \dots, k-1\} = \mathbb{Z}/k\mathbb{Z}$, the cyclic group with k elements. Suppose further that the points $v_0, \dots, v_{k-1}$ are indexed such that if we start at v_0 and run through ∂X with suitable orientation, then the points $v_0, \dots, v_{k-1}$ are traversed in successive order. If this is true and if with this orientation of ∂X the region X lies on the left, then we call the points $v_1, \dots, v_k$ in *cyclic* order on ∂X, and otherwise, if X lies on the right, in *anti-cyclic* order on ∂X. If the points $v_0, \dots, v_{k-1}$ are in cyclic or anti-cyclic order on ∂X, then ∂X is decomposed into unique arcs $e_0, \dots, e_{k-1}$; here e_l for $l \in \mathbb{Z}_k$ is the unique subarc of ∂X that has the endpoints v_l and v_{l+1}, but does not contain any other of the points v_i, $i \in \mathbb{Z}_k \setminus \{l, l+1\}$. We say that the arcs $e_0, \dots, e_{k-1}$ are in *cyclic* or *anti-cyclic* order on ∂X, if this is true for the points $v_0, \dots, v_{k-1}$, respectively.

If we have a labeling $L\colon \mathcal{D}^1 \to \mathcal{D}^0$, we should think of each element $\tau \in \mathcal{D}^1$ as "carrying" the label $L(\tau) \in \mathcal{D}^0$. In applications it is often more intuitive and convenient to allow more general index sets $\mathcal{L}$ of the same cardinality as $\mathcal{D}^0$ as labeling sets for the elements in $\mathcal{D}^1$. In such situations we fix a bijection $\psi\colon \mathcal{D}^0 \to \mathcal{L}$ and call a map $L'\colon \mathcal{D}^1 \to \mathcal{L}$ a labeling if $\psi^{-1} \circ L'\colon \mathcal{D}^1 \to \mathcal{D}^0$ is a labeling in the sense of Definition 5.22.

We will discuss this in a case that will be relevant for us later. Namely, suppose that the cell decomposition $\mathcal{D}^0$ of the oriented sphere S^2 has two tiles $X_{\mathtt{b}}^0$ and $X_{\mathtt{w}}^0$ with common boundary $\mathcal{C} \coloneqq \partial X_{\mathtt{b}}^0 = \partial X_{\mathtt{w}}^0$. We represent $X_{\mathtt{b}}^0$ by the symbol $\mathtt{b}$ for "black" and $X_{\mathtt{w}}^0$ by $\mathtt{w}$ for "white". By Lemma 5.9 (iii) the set $\mathcal{C}$ is a Jordan curve containing $k \geq 2$ vertices and edges, and there are no other edges and vertices in $\mathcal{D}^0$. Let us assume that $k \geq 3$ and that we have indexed the vertices $v_0, \dots, v_{k-1}$ so that they are cyclically ordered on $\partial X_{\mathtt{w}}^0$. Then they are anti-cyclically

ordered on $\partial X_{\mathtt{b}}^0$. As above, we index the edges such that e_l is the unique subarc on $\mathcal{C} = \partial X_{\mathtt{w}}^0$ with endpoints v_l and v_{l+1}.

There exists a bijection of $\mathcal{D}^0$ with the set $\mathcal{L}$ consisting of the symbols $\mathtt{b}$ and $\mathtt{w}$ (for the two tiles in $\mathcal{D}^0$) and two copies of $\mathbb{Z}_k$, one for the edges and the other one for the vertices in $\mathcal{D}^0$. More explicitly, such a bijection $\psi\colon \mathcal{D}^0 \to \mathcal{L}$ is given by

$$(5.5) \quad \psi(X_{\mathtt{w}}^0) = \mathtt{w}, \ \psi(X_{\mathtt{b}}^0) = \mathtt{b}, \ \psi(v_l) = l \text{ and } \psi(e_l) = l \text{ for } l \in \mathbb{Z}_k.$$

Now suppose that in this situation $\mathcal{D}^0$ is equal to the cell decomposition $\mathcal{D}^0(f, \mathcal{C})$ (as defined in Section 5.3) for a Thurston map $f\colon S^2 \to S^2$. In other words, $\mathrm{post}(f) \subset \mathcal{C}$ and the vertex set of $\mathcal{D}^0$ is equal to $\mathrm{post}(f)$. Let $\mathcal{D}^1 = \mathcal{D}^1(f, \mathcal{C})$ be the cell decomposition given by cells of level 1, and $\mathbf{X} = \mathbf{X}^1$, $\mathbf{E} = \mathbf{E}^1$, and $\mathbf{V} = \mathbf{V}^1$ be the sets of 1-tiles, 1-edges, and 1-vertices, respectively. One can then define three maps $L_{\mathbf{X}}\colon \mathbf{X} \to \{\mathtt{w}, \mathtt{b}\}$, $L_{\mathbf{E}}\colon \mathbf{E} \to \mathbb{Z}_k$ and $L_{\mathbf{V}}\colon \mathbf{V} \to \mathbb{Z}_k$ as

$$L_{\mathbf{X}}(X) = \psi(f(X)), \quad L_{\mathbf{E}}(e) = \psi(f(e)), \quad L_{\mathbf{V}}(v) = \psi(f(v))$$

for $X \in \mathbf{X}$, $e \in \mathbf{E}$, and $v \in \mathbf{V}$. Accordingly, a 1-tile X is called "white" if $L_{\mathbf{X}}(X) = \mathtt{w}$, and called "black" if $L_{\mathbf{X}}(X) = \mathtt{b}$. Recall from Proposition 5.17 (i) that for each $X \in \mathbf{X}$ the map $f|X$ is an orientation-preserving homeomorphism onto either $X_{\mathtt{w}}^0$ or $X_{\mathtt{b}}^0$. This implies that if we use the map $L_{\mathbf{V}}\colon \mathbf{V} \to \mathbb{Z}_k$ to index 1-vertices, then they are in cyclic order on the boundary ∂X of a white 1-tile X and in anti-cyclic order on a boundary of a black 1-tile. Similarly, the 1-edges are in cyclic order on the boundary of white 1-tiles, and in anti-cyclical order on the boundary of black 1-tiles. The maps $L_{\mathbf{X}}$, $L_{\mathbf{E}}$ and $L_{\mathbf{V}}$ can be combined in the obvious way to a map $L\colon \mathcal{D}^1 \to \mathcal{L} \cong \mathcal{D}^0$ (so that $L|\mathbf{X} = L_{\mathbf{X}}$, etc.), and it follows easily from the previous discussion that L is a labeling.

In the following lemma, we will turn this construction around and ask when a labeling with similar properties exists on a given cell decomposition $\mathcal{D} = \mathcal{D}^1$ of a 2-sphere that is *a prori* not related to a Thurston map. This will later be useful when we want to construct Thurston maps.

LEMMA 5.23. *Let $\mathcal{D}$ be a cell decomposition of S^2, and denote by $\mathbf{V}$ the set of vertices, by $\mathbf{E}$ the set of edges, and by $\mathbf{X}$ the set of tiles in $\mathcal{D}$. Suppose that the length of the cycle of every vertex in $\mathcal{D}$ is even and that there exists $k \geq 3$ such that every tile in $\mathbf{X}$ is a k-gon.*

Then for each positively-oriented flag (c_0, c_1, c_2) in $\mathcal{D}$ there exist maps $L_{\mathbf{V}}\colon \mathbf{V} \to \mathbb{Z}_k$, $L_{\mathbf{E}}\colon \mathbf{E} \to \mathbb{Z}_k$, and $L_{\mathbf{X}}\colon \mathbf{X} \to \{\mathtt{b}, \mathtt{w}\}$ with the following properties:

(i) $L_{\mathbf{V}}(c_0) = 0$, $L_{\mathbf{E}}(c_1) = 0$, *and* $L_{\mathbf{X}}(c_2) = \mathtt{w}$,

(ii) *if $X, Y \in \mathbf{X}$ are two distinct tiles with a common edge on their boundaries, then $L_{\mathbf{X}}(X) \neq L_{\mathbf{X}}(Y)$,*

(iii) *if X is an arbitrary tile in $\mathbf{X}$, then $L_{\mathbf{V}}$ induces a bijection of the set of vertices in ∂X with $\mathbb{Z}_k$ so that the order of these vertices is cyclic if $L_{\mathbf{X}}(X) = \mathtt{w}$ and anti-cyclic if $L_{\mathbf{X}}(X) = \mathtt{b}$,*

(iv) *if $e \in \mathbf{E}$ and $l = L_{\mathbf{E}}(e)$, then $L_{\mathbf{V}}(\partial e) = \{l, l + 1\}$,*

(v) *if X is an arbitrary tile in $\mathbf{X}$, then $L_{\mathbf{E}}$ induces a bijection of the set of edges contained in ∂X so that the order of these edges is cyclic if $L_{\mathbf{X}}(X) = \mathtt{w}$ and anti-cyclic if $L_{\mathbf{X}}(X) = \mathtt{b}$,*

(vi) *a flag (τ_0, τ_1, τ_2) in $\mathcal{D}$ is positively-oriented if and only if there exists $l \in \mathbb{Z}_k$ such that $(L_{\mathbf{V}}(\tau_0), L_{\mathbf{E}}(\tau_1), L_{\mathbf{X}}(\tau_2))$ is equal to $(l, l, \mathtt{w})$ or $(l, l - 1, \mathtt{b})$.*

The maps $L_{\mathbf{V}}$, $L_{\mathbf{E}}$, and $L_{\mathbf{X}}$ are uniquely determined by the properties (i)–(iv).

Here in (i) and (vi) we again ignored the distinction between vertices in $\mathcal{D}$ and 0-dimensional cells in $\mathcal{D}$ by setting $L_{\mathbf{V}}(c) = L_{\mathbf{V}}(v)$ if $c = \{v\}$ is a 0-dimensional cell consisting of the vertex v.

Recall (see the discussion after Lemma 5.9) that the length of the cycle of a vertex v is the number of edges as well as the number of tiles that contain v. So instead of saying that the length of each cycle of every vertex is even, we could have said that every vertex is contained in an even number of tiles (equivalently contained in an even number of edges).

Condition (ii) says that one of the two tiles containing an edge is "black" and the other is "white". So if the tiles have been labeled in this way, then $\mathcal{D}$ becomes a checkerboard tiling by k-gons. Condition (ii) is equivalent to the statement that the dual graph of the 1-skeleton of $\mathcal{D}$ is bipartite.

Condition (iii) says that for any white tile X, the vertices on ∂X are labeled cyclically; for any black tile Y, the vertices on ∂Y are labeled anti-cyclically. A more precise formulation is as follows: for each $i \in \mathbb{Z}_k$ there is exactly one vertex $v \in \partial X$ with $L_{\mathbf{V}}(v) = i$, and if we write $v = v_i$ if $L_{\mathbf{V}}(v) = i$, then the vertices $v_0, \dots, v_{k-1} \in \partial X$ are in cyclic order on ∂X if X is white and in anti-cyclic order if X is black. Condition (v) has to be interpreted in a similar way.

By (iv) the label $L_{\mathbf{E}}(e)$ of an edge $e \in \mathbf{E}$ is determined by the labels $L_{\mathbf{V}}(u)$ and $L_{\mathbf{V}}(v)$ of the two endpoints u and v of e (here it is important that $k \geq 3$).

PROOF OF LEMMA 5.23. We first establish the following fact.

Claim. Suppose that $J \subset S^2$ is a Jordan curve that does not contain any vertex (in $\mathcal{D}$) and has the property that for every edge e the intersection $e \cap J$ is either empty, or e meets both components of $S^2 \setminus J$ and $e \cap J$ consists of a single point. Then J meets an even number of edges.

To see this, pick one of the complementary components U of $S^2 \setminus J$, and let $v_1, \ldots, v_n$ be the vertices contained in U, where $n \in \mathbb{N}_0$ (for $n = 0$ we consider this as an empty list). For $i = 1, \ldots, n$ let d_i be the length of the cycle of v_i, i.e., the number of edges containing v_i. We denote by $\mathbf{E}_J$ the set of all edges that meet J and by $\mathbf{E}_U$ the set of all edges contained in U. From our assumption on the intersection property of J with edges it follows that an edge is contained in U if and only if its two endpoints are in U, and it meets J if and only if one endpoint is in U and the other in $S^2 \setminus \overline{U}$. Hence

$$d_1 + \cdots + d_n = \#\mathbf{E}_J + 2\#\mathbf{E}_U,$$

because the sum on the left hand side counts every edge in $\mathbf{E}_J$ once, and every edge in $\mathbf{E}_U$ twice. Since all the numbers $d_1, \ldots, d_n$ are even by our assumptions, we conclude that the number $\#\mathbf{E}_J$ of edges that J meets is also even. The claim follows.

We now proceed to show existence and uniqueness of the map $L_{\mathbf{X}}$. For every tile Y there exists an e-chain $Y_0 = c_2, \ldots, Y_N = Y$ of tiles joining the "base tile" c_2 (from the given flag (c_0, c_1, c_2)) to Y. Recall from Definition 5.20 that such an e-chain is a finite sequence $Y_0, \ldots, Y_N$ of tiles such that $Y_i \neq Y_{i+1}$ and there is an edge $e_i \subset \partial Y_i \cap Y_{i+1}$ for $i = 0, \ldots, N-1$. We put $L_{\mathbf{X}}(Y) = \mathtt{w}$ or $L_{\mathbf{X}}(Y) = \mathtt{b}$ depending on whether N is even or odd. It is clear that if this is well-defined, then it is the unique choice for $L_{\mathbf{X}}(Y)$. This follows from the normalization (i) and the fact that by (ii) the labels of tiles have to alternate along an e-chain.

To see that $L_{\mathbf{X}}$ is well-defined, it is enough to show that if an e-chain $X_0, X_1, \ldots, X_N$ forms a cycle, i.e., if $X_0 = X_N$, then N is even. To prove this, we may make the additional assumption that $N \geq 3$ and that the chain is *simple*, i.e., that the tiles $X_1, \ldots X_N$ are all distinct.

Let e_i be an edge with $e_i \subset \partial X_{i-1} \cap \partial X_i$ for $i = 1, \ldots, N$. Then the edges $e_1, \ldots, e_N$ are all distinct. For otherwise, $e_i = e_j$ for some $1 \leq i < j \leq N$. Then $e_i = e_j$ is contained in the boundary of the tiles $X_{i-1}, X_i, X_{j-1}, X_j$ which is impossible, because three of these tiles must be distinct (note that $N \geq 3$).

We now construct a Jordan curve J that "follows" our closed e-chain. Formally, for each edge e_i pick a point $x_i \in \mathrm{int}(e_i)$. Moreover, for

$i = 1, \ldots, N$, we can choose an arc $\alpha_i \subset X_i$ with endpoints x_i and x_{i+1} such that $\mathrm{int}(\alpha_i) \subset \mathrm{int}(X_i)$. Here $x_{N+1} \coloneqq x_1$. Then $J = \alpha_1 \cup \cdots \cup \alpha_N$ is a Jordan curve that has properties as in the claim above. The curve J meets the edges $e_1, \ldots, e_N$ and no others. Hence N is even. Thus $L_{\mathbf{X}}$ is well-defined, and it has property (ii) and is normalized as in (i).

To show the existence of $L_{\mathbf{V}}$, it is useful to quickly recall some basic definitions from the homology and cohomology of chain complexes. Denote by $\mathbf{E}_o$ the set of oriented edges in $\mathcal{D}$. Let $C(\mathbf{X})$ and $C(\mathbf{E}_o)$ be the free modules over $\mathbb{Z}_k$ generated by the sets $\mathbf{X}$ and $\mathbf{E}_o$, respectively. So $C(\mathbf{E}_o)$, for example, is just the set of formal finite sums $\sum a_i e_i$, where $a_i \in \mathbb{Z}_k$ and $e_i \in \mathbf{E}_o$. Note that in contrast to other commonly used definitions of chain complexes we have $e + \widetilde{e} \neq 0$ if e and $\widetilde{e}$ are oriented edges with the same underlying set, but opposite orientations.

There is a unique boundary operator $b \colon C(\mathbf{X}) \to C(\mathbf{E}_o)$ that is a module homomorphism and satisfies

$$bX \coloneqq b(X) = \sum_{e \subset \partial X} e$$

for each tile X, where the sum is extended over all oriented edges $e \subset \partial X$ so that X lies on the left of e.

Let e be an oriented edge and X be the unique tile with $e \subset \partial X$ that is on the left of e. We put $\alpha(e) = 1 \in \mathbb{Z}_k$ or $\alpha(e) = -1 \in \mathbb{Z}_k$ depending on whether $L_{\mathbf{X}}(X) = \mathtt{w}$ (X is a white tile) or $L_{\mathbf{X}}(X) = \mathtt{b}$ (X is a black tile). If e and $\widetilde{e}$ are oriented edges with the same underlying set, but opposite orientations, then $\alpha(e) + \alpha(\widetilde{e}) = 0$ as follows from property (ii) of $L_{\mathbf{X}}$.

The map α extends uniquely to a homomorphism $\alpha \colon C(\mathbf{E}_o) \to \mathbb{Z}_k$. In the language of cohomology it is a "cochain". This cochain α is a cocycle, i.e.,

$$(5.6) \qquad \alpha(bX) = \sum_{e \subset \partial X} \alpha(e) = \pm k = 0 \in \mathbb{Z}_k$$

for every tile X, considered as one of the generators of $C(\mathbf{X})$. Indeed, by our convention on the orientation of edges $e \subset \partial X$ in the above sum, for each such edge we get the same contribution $\alpha(e)$, and so, since X has k edges, the sum is equal to $\pm k = 0 \in \mathbb{Z}_k$.

Consider an arbitrary closed edge path consisting of the oriented edges $e_1, \ldots, e_n$; so the terminal point of e_i is the initial point of e_{i+1} for $i = 1, \ldots, n$, where $e_{n+1} \coloneqq e_1$. We claim that

$$(5.7) \qquad \sum_{i=1}^{n} \alpha(e_i) = 0.$$

Essentially, this is a consequence of the fact that we have $H^1(S^2, \mathbb{Z}_k) = 0$ for the first cohomology group of S^2 with coefficients in $\mathbb{Z}_k$. This implies that the cocycle α is a coboundary and gives 5.7.

We will present a simple direct argument. To show (5.7), it is clearly enough to establish this for simple closed edge paths, i.e., for closed edge paths where the underlying sets of all edges are distinct and have a Jordan curve $J \subset S^2$ as a union. In this case, let U be the complementary component of $S^2 \setminus J$ so that U lies on the left if we traverse J according to the orientation given by the edges e_i. If $X_1, \ldots, X_M$ are all the tiles contained in $\overline{U}$, then

$$b(X_1 + \cdots + X_M) = \sum_{e \subset \overline{U}} e,$$

where the sum is extended over oriented edges contained in $\overline{U}$. Each edge on J is equal to one of the edges e_i and it appears in the above sum exactly once and with the same orientation as e_i. All other edges in $\overline{U}$ appear twice and with opposite orientations. Hence by (5.6),

$$\sum_{i=1}^{n} \alpha(e_i) = \sum_{e \subset \overline{U}} \alpha(e) = \sum_{i=1}^{M} \alpha(bX_i) = 0.$$

We now define $L_{\mathbf{V}} \colon \mathbf{V} \to \mathbb{Z}_k$ as follows. Suppose $c_0 = \{p_0\}$ consists of the vertex p_0. For $v \in \mathbf{V}$ we can pick an edge path consisting of the oriented edges $e_1, \ldots, e_n$ that joins the base point p_0 to v (this list of edges may be empty if $v = p_0$). The existence of such an edge path follows from the connectedness of the 1-skeleton of $\mathcal{D}$ (see Lemma 5.9 (vi)). Put

(5.8) $$L_{\mathbf{V}}(v) = \sum_{i=1}^{n} \alpha(e_i).$$

This is well-defined, because we have (5.7) for every closed edge path; we also have the normalization $L_{\mathbf{V}}(c_0) = L_{\mathbf{V}}(p_0) = 0$.

The definition of $L_{\mathbf{V}}$ implies that if e is an oriented edge, and u is the initial and v the terminal point of e, then

(5.9) $$L_{\mathbf{V}}(v) = L_{\mathbf{V}}(u) + \alpha(e).$$

This means that if we go from the initial point u of e to the terminal point v, then the value of $L_{\mathbf{V}}$ is increased by 1 or decreased by -1 depending on whether the tile on the left of e is white or black. The desired property (iii) of $L_{\mathbf{V}}$ immediately follows from this.

This shows existence of $L_{\mathbf{V}}$. Conversely, every function $L_{\mathbf{V}}$ with property (iii) must satisfy (5.9). Together with the normalization

$L_{\mathbf{V}}(p_0) = 0$ this implies that $L_{\mathbf{V}}$ is given by the formula (5.8), and so we have uniqueness.

To define $L_{\mathbf{E}}$ note that if $e \in \mathbf{E}$, then by (ii) we can choose a unique orientation for e such that the tile on the left is white, and the one on the right is black. If u is the initial and v the terminal point of e according to this orientation, and $L_{\mathbf{V}}(u) = l \in \mathbb{Z}_k$, then $L_{\mathbf{V}}(v) = l + 1$. Now set $L_{\mathbf{E}}(e) := l$. Then $L_{\mathbf{E}}$ has property (iv). Moreover, we also have the normalization (i) for $L_{\mathbf{E}}$; indeed, is c_1 is oriented so that p_0 is the initial point of c_1, then c_2 lies on the left of c_1, because the flag (c_0, c_1, c_2) is positively-oriented. Since $L_{\mathbf{X}}(c_2) = \mathtt{w}$, the tile c_2 is white and so $L_{\mathbf{E}}(e) = L_{\mathbf{V}}(p_0) = 0$. Uniqueness of $L_{\mathbf{E}}$ follows from (iii) and the uniqueness of $L_{\mathbf{V}}$.

We have proved (i)–(iv) and the uniqueness statement. It remains to establish (v) and (vi).

To show (v) let $X \in \mathbf{X}$ be arbitrary. Then by (iii) we can assume that the indexing of the k vertices $v_0, \ldots, v_{k-1}$ on ∂X is such that $L_{\mathbf{V}}(v_i) = i$ for all $i \in \mathbb{Z}_k$, and that $v_0, \ldots, v_{k-1}$ are met in successive order if we traverse ∂X. This implies that for each $i \in \mathbb{Z}_k$ there exists a unique edge $e_i \subset \partial X$ in $\mathcal{D}$ with endpoints v_i and v_{i+1}. Hence by (iv) we have $L_{\mathbf{E}}(e_i) = i$. Moreover, by (iii) the edges $e_0, \ldots, e_{k-1}$ are in cyclic or anti-cyclic order on ∂X depending on whether $L_{\mathbf{X}}(X) = \mathtt{w}$ or $L_{\mathbf{X}}(X) = \mathtt{b}$. So (v) holds.

Finally, to see that (vi) is true, let (τ_0, τ_1, τ_2) be a flag in $\mathcal{D}$. Then $\tau_0 = \{u\}$ for some $u \in \mathbf{V}$. The vertex u is the initial point of the oriented edge τ_1. Let $v \in \mathbf{V}$ be the terminal point of τ_1, and define $l = L_{\mathbf{V}}(u)$.

Depending on whether the flag is positively- or negatively-oriented, the vertex v follows u in cyclic or anti-cyclic order on $\partial \tau_2$. So if the flag is positively-oriented, then by property (iii) we have $L_{\mathbf{V}}(v) = l + 1$ if $L_{\mathbf{X}}(\tau_2) = \mathtt{w}$ and $L_{\mathbf{V}}(v) = l - 1$ if $L_{\mathbf{X}}(\tau_2) = \mathtt{b}$. Property (iv) implies that $L_{\mathbf{E}}(\tau_1) = l$ if $L_{\mathbf{X}}(\tau_2) = \mathtt{w}$ and $L_{\mathbf{E}}(\tau_1) = l - 1$ if $L_{\mathbf{X}}(\tau_2) = \mathtt{b}$.

So if (τ_0, τ_1, τ_2) is positively-oriented, then the cells in this flag carry the labels l, l, $\mathtt{w}$, or l, $l - 1$, $\mathtt{b}$, respectively.

Similarly, if (τ_0, τ_1, τ_2) is negatively-oriented, then we get the labels l, $l - 1$, $\mathtt{w}$, or l, l, $\mathtt{b}$ for the cells in the flag. Statement (vi) follows from this. □

5.5. Thurston maps from cell decompositions

In Section 5.3 we have seen how to obtain cell decompositions from Thurston maps. In this section we reverse this procedure and ask when a pair $(\mathcal{D}^1, \mathcal{D}^0)$ of cell decompositions gives rise to a Thurston

map f that is cellular for $(\mathcal{D}^1, \mathcal{D}^0)$. In general, one cannot expect f to be determined just by the cell decompositions alone, but one needs additional information on how f is supposed to map the cells in $\mathcal{D}^1$ to cells in $\mathcal{D}^0$. This is given by an orientation-preserving labeling as discussed in the previous section (see Definition 5.22 and the discussion following this definition).

We start with a lemma that allows us to recognize branched covering maps.

LEMMA 5.24. *Let $\mathcal{D}'$ and $\mathcal{D}$ be two cell decompositions of S^2, and $f\colon S^2 \to S^2$ be a cellular map for $(\mathcal{D}', \mathcal{D})$ such that $f|X$ is orientation-preserving for each tile X in $\mathcal{D}'$.*

(i) *Then f is a branched covering map on S^2. Each critical point of f is a vertex of $\mathcal{D}'$.*

(ii) *If in addition each vertex in $\mathcal{D}$ is also a vertex in $\mathcal{D}'$, then every point in $\mathrm{post}(f)$ is a vertex of $\mathcal{D}$. In particular, f is postcritically-finite, and hence a Thurston map if f is not a homeomorphism.*

The assumption that $f|X$ is orientation-preserving means that f preserves the orientation of flags contained in X.

PROOF. (i) We will show that for each point $p \in S^2$, there exists $k \in \mathbb{N}$, an orientation-preserving homeomorphism φ of the open unit disk $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ onto a neighborhood W' of p, and an orientation-preserving homeomorphism ψ of a neighborhood $W \supset f(W')$ of $q = f(p)$ onto $\mathbb{D}$ such that $\varphi(0) = p$, $\psi(q) = 0$, and

$$(\psi \circ f \circ \varphi)(z) = z^k$$

for all $z \in \mathbb{D}$. The desired relation between the points and maps can be represented by the commutative diagram

$$(5.10) \qquad \begin{array}{ccc} p \in W' & \xrightarrow{\ f\ } & q \in W \\ {\scriptstyle \varphi}\big\uparrow & & \big\downarrow{\scriptstyle \psi} \\ 0 \in \mathbb{D} & \xrightarrow{\ z \mapsto z^k\ } & 0 \in \mathbb{D}. \end{array}$$

Note that the homeomorphism φ here runs in the opposite direction compared to (2.1), but this is slightly more convenient in Case 3 of the ensuing proof.

We will use the fact that if f is an orientation-preserving local homeomorphism near p, then we can take $k = 1$ in (5.10) and always find suitable homeomorphisms φ and ψ.

Let $p \in S^2$ be arbitrary. Since S^2 is the disjoint union of the interiors of the cells in $\mathcal{D}'$, the point p is contained in the interior of a tile or an edge in $\mathcal{D}'$, or is a vertex of $\mathcal{D}'$. Accordingly, we consider three cases.

Case 1: There exists a tile $X' \in \mathcal{D}'$ with $p \in \operatorname{int}(X')$. Then $W' := \operatorname{int}(X')$ is an open neighborhood of p, and $f|W'$ is an orientation-preserving homeomorphism of $W' = \operatorname{int}(X')$ onto $W := \operatorname{int}(X)$, where $X = f(X') \in \mathcal{D}$. Hence f is an orientation-preserving local homeomorphism near p.

Case 2: There exists an edge $e' \in \mathcal{D}'$ with $p \in \operatorname{int}(e')$. By Lemma 5.9 (iv) there exist distinct tiles $X', Y' \in \mathcal{D}'$ such that $e' \subset \partial X' \cap \partial Y'$. Then $W' = \operatorname{int}(X') \cup \operatorname{int}(e') \cup \operatorname{int}(Y')$ is an open neighborhood of p. Since f is cellular, $X = f(X')$ and $Y = f(Y')$ are tiles in $\mathcal{D}$ and $e = f(e')$ is an edge in $\mathcal{D}$. Moreover, $e \subset \partial X \cap \partial Y$.

We orient e' so that X' lies to the left and Y' to the right of e'. Since f is orientation-preserving if restricted to cells in $\mathcal{D}'$, the tile X lies to the left, and Y to the right of the image e of e'. In particular, $X \neq Y$, and so the sets $\operatorname{int}(X), \operatorname{int}(e), \operatorname{int}(Y)$ are pairwise disjoint, and their union is open. Since f is cellular and hence a homeomorphism if restricted to cells (and interior of cells), it follows that $f|W'$ is a homeomorphism of W' onto the open set $W = \operatorname{int}(X) \cup \operatorname{int}(e) \cup \operatorname{int}(Y)$. Moreover, it is clear that $f|W'$ is orientation-preserving. Since W' is open and contains p, the map f is an orientation-preserving local homeomorphism near p.

Case 3: The point p is a vertex of $\mathcal{D}'$. Since we already know that in the complement of the vertex set of $\mathcal{D}'$ our map f is an orientation-preserving local homeomorphism, one can deduce the desired local epresentation (5.10) of f near p from a general fact (see Lemma A.9). We will provide a direct argument for this that will also give us additional insight how the cells in the cycle of p are mapped (this is summarized in Remark 5.25 after the proof).

As in the proof of Lemma 5.9 (v), we can choose tiles $X'_j \in \mathcal{D}'$ and edges $e'_j \in \mathcal{D}'$ for $j \in \mathbb{N}$ that contain p and satisfy $X'_j \neq X'_{j+1}$, $e'_j \neq e'_{j+1}$, and $e'_j \subset \partial X'_j \cap \partial X'_{j+1}$ for all $j \in \mathbb{N}$. There exists $d' \in \mathbb{N}$ such that $X'_{d'+1} = X'_1$, the tiles $X'_1, \dots, X'_{d'}$ as well as the edges $e'_1, \dots, e'_{d'}$ are all distinct, and

$$W' = \{p\} \cup \operatorname{int}(X'_1) \cup \operatorname{int}(e'_1) \cup \operatorname{int}(X'_2) \cup \cdots \cup \operatorname{int}(e'_{d'})$$

is an open neighborhood of p. Moreover, by the remark following Lemma 5.9, we know that $X'_j = X'_{d'+j}$ and $e'_j = e'_{d'+j}$ for all $j \in \mathbb{N}$.

Define $X_j = f(X'_j)$ and $e_j = f(e'_j)$ for $j \in \mathbb{N}$. Since f is cellular for $(\mathcal{D}', \mathcal{D})$, the set X_j is a tile and e_j an edge in $\mathcal{D}$. Note that $e_j \subset \partial X_j \cap \partial X_{j+1}$ for $j \in \mathbb{N}$. Since X'_j and X'_{j+1} are distinct tiles containing the edge e'_{j+1} in their boundaries, it follows by an argument as in Case 2 above that $X_j \neq X_{j+1}$ for $j \in \mathbb{N}$. Similarly, e'_j and e'_{j+1} are distinct edges in $\mathcal{D}'$ contained in X'_{j+1}, and $f|X'_{j+1}$ is a homeomorphism, and so again $e_j \neq e_{j+1}$ for $j \in \mathbb{N}$.

As in the proof of Lemma 5.9 (v) we see that there exists a number $d \in \mathbb{N}$, $d \geq 2$, such that $X_{d+1} = X_1$, and such that the tiles $X_1, \ldots, X_d$ and the edges $e_1, \ldots, e_d$ are all distinct. Moreover,

$$W = \{q\} \cup \operatorname{int}(X_1) \cup \operatorname{int}(e_1) \cup \operatorname{int}(X_2) \cup \cdots \cup \operatorname{int}(e_d)$$

is an open neighborhood of $q = f(p)$, and $X_j = X_{d+j}$ and $e_j = e_{d+j}$ for all $j \in \mathbb{N}$.

The periodicity properties of the indexing of the tiles X'_j and X_j imply that $d \leq d'$ and that d is a divisor of d'. Hence there exists $k \in \mathbb{N}$ such that $d' = kd$.

We now claim that after suitable coordinate changes near p and q, the map f can be given the form $z \mapsto z^k$.

For $N \in \mathbb{N}$, $N \geq 2$, and $j \in \mathbb{N}$ define half-open line segments

$$L_j^N = \{re^{2\pi i j/N} : 0 \leq r < 1\} \subset \mathbb{D}$$

and sectors

$$\Sigma_j^N = \{re^{it} : 2\pi(j-1)/N \leq t \leq 2\pi j/N \text{ and } 0 \leq r < 1\} \subset \mathbb{D}.$$

We then construct a homeomorphism $\psi\colon W \to \mathbb{D}$ with $\psi(q) = 0$ as follows. For each $j = 1, \ldots, d$ we first map the half-open arc $\{q\} \cup \operatorname{int}(e_j)$ homeomorphically to the half-open line segment L_j^d. Then q is mapped to 0, so these maps are consistently defined for q. Since X_j is a Jordan region, we can extend the homeomorphisms on $\{q\} \cup \operatorname{int}(e_{j-1}) \subset \partial X_j$ and on $\{q\} \cup \operatorname{int}(e_j) \subset \partial X_j$ to a homeomorphism of

$$\{q\} \cup \operatorname{int}(e_{j-1}) \cup \operatorname{int}(e_j) \cup \operatorname{int}(X_j)$$

onto the sector Σ_j^d for each $j = 2, \ldots, d+1$. Since the sets

$$\{q\}, \operatorname{int}(e_1), \ldots, \operatorname{int}(e_d), \operatorname{int}(X_2), \ldots, \operatorname{int}(X_{d+1}) = \operatorname{int}(X_1)$$

are pairwise disjoint and have W as a union, these homeomorphisms paste together to a well-defined homeomorphism ψ of W onto $\mathbb{D}$. Note that $\psi(q) = 0$ and $\psi(X_j \cap W) = \Sigma_j^d$ for each $j = 1, \ldots, d$.

A homeomorphism $\varphi\colon \mathbb{D} \to W'$ is defined as follows. If $z \in \mathbb{D}$ is arbitrary, then $z \in \Sigma_j^{d'}$ for some $j = 1, \ldots, d'$. Hence $z^k \in \Sigma_j^d$, and

so $\psi^{-1}(z^k) \in X_j \cap W$. Since f is a homeomorphism of $X'_j \cap W'$ onto $X_j \cap W$, it follows that $(f|X'_j)^{-1}(\psi^{-1}(z^k))$ is defined and lies in $X'_j \cap W'$.

We set

$$\varphi(z) = (f|X'_j)^{-1}(\psi^{-1}(z^k)).$$

It is straightforward to verify that φ is well-defined and a homeomorphism of $\mathbb{D}$ onto W' with $\varphi(0) = p$. It follows from the definition of φ that $(\psi \circ f \circ \varphi)(z) = z^k$ for $z \in \mathbb{D}$, and so we have the diagram (5.10).

We assume that the tiles X'_j and the edges e'_j are indexed by the procedure in the proof of Lemma 5.9 (v). Then each flag $(\{p\}, e'_j, X'_{j+1})$ is positively-oriented (see the remark after the proof of Lemma 5.9). Since $f|X'_j$ is orientation-preserving, this implies that the flag $(\{q\}, e_j, X_{j+1})$ is also positively-oriented. We conclude that ψ is orientation-preserving, since ψ maps the positively-oriented flag $(\{q\}, e_j, X_{j+1})$ in S^2 to the positively-oriented flag $(\{0\}, L^d_j, \Sigma^d_{j+1})$ in $\widehat{\mathbb{C}}$. As follows from its definition, the map φ is then also orientation-preserving. Hence ψ and φ are local homeomorphisms as desired.

We have shown that in all cases the map f has a local behavior as claimed. It follows that f is a branched covering map. Moreover, we have seen that f near each point is a local homeomorphism unless p is a vertex of $\mathcal{D}'$. It follows that each critical point of f is a vertex of $\mathcal{D}'$.

(ii) Suppose in addition that every vertex of $\mathcal{D}$ is also a vertex of $\mathcal{D}'$. Let p be a critical point of f. Then by (i) the point p is a vertex of $\mathcal{D}'$. Since f is cellular for $(\mathcal{D}', \mathcal{D})$, the point $f(p)$ is a vertex of $\mathcal{D}$. Hence $f(p)$ is also a vertex of $\mathcal{D}'$, and we can apply the argument again, to conclude that $f^2(p)$ is a vertex of $\mathcal{D}$, etc. It follows that $\mathrm{post}(f)$ is a subset of the set of vertices of $\mathcal{D}$. In particular, $\mathrm{post}(f)$ is finite, and so f is postcritically-finite. $\qquad\square$

REMARK 5.25. Let the map $f\colon S^2 \to S^2$ and the cell decompositions $\mathcal{D}'$ and $\mathcal{D}$ be as in the previous lemma, and let p be a vertex in $\mathcal{D}'$. Then $q = f(p)$ is a vertex in $\mathcal{D}$. If d' and d are the lengths of the cycles of p in $\mathcal{D}'$ and q in $\mathcal{D}$, respectively, then $d' = dk$, where $k = \deg_f(p)$. Moreover, if X'_j for $j \in \mathbb{N}$ are the tiles in the cycle of p in $\mathcal{D}'$ labeled so that $X'_{d'+j} = X'_j$ for $j \in \mathbb{N}$, then $X_j = f(X'_j)$ are the tiles in the cycle of q in $\mathcal{D}$. In addition, $X_{d+j} = f(X'_{d+j}) = f(X'_j) = X_j$ for $j \in \mathbb{N}$. This was established in Case 3 of the proof of Lemma 5.24.

So on a more intuitive level, if we follow the tiles in the cycle of p in a cyclic order modulo d', then under the map we follow the tiles in the cycle of $q = f(p)$ also in a cyclic order modulo $d = d'/k$. Here each tile $X_j = f(X'_j)$ in the image cycle has precisely $k = \deg_f(p)$ distinct

preimage tiles, namely $X'_j, X'_{d+j}, \ldots, X'_{(k-1)d+j}$. A similar statement is true for edges.

We can now prove the following fact which allows us the construction of many Thurston maps (see Section 12.3 for specific examples).

PROPOSITION 5.26. *Let $\mathcal{D}^0$ and $\mathcal{D}^1$ be cell decompositions of an oriented 2-sphere S^2, and $L\colon \mathcal{D}^1 \to \mathcal{D}^0$ be an orientation-preserving labeling. Suppose that every vertex of $\mathcal{D}^0$ is also a vertex of $\mathcal{D}^1$. Then there exists a branched covering map $f\colon S^2 \to S^2$ that is cellular for $(\mathcal{D}^1, \mathcal{D}^0)$ and is compatible with the given labeling L. The map f is a homeomorphism or a Thurston map. In the latter case, f is unique up to Thurston equivalence, and $\mathrm{post}(f)$ is contained in the set of vertices of $\mathcal{D}^0$.*

Later we will consider triples $(\mathcal{D}^1, \mathcal{D}^0, L)$ as in the previous proposition related to Thurston maps f with invariant curves $\mathcal{C}$. These triples satisfy additional and somewhat technical conditions and lead to the notion of a *two-tile subdivision rule* (see Definition 12.1).

PROOF OF PROPOSITION 5.26. As in the proof of Lemma 5.11 (ii), a map f as desired is obtained from successive extensions to the skeleta of the cell decomposition $\mathcal{D}^1$. Indeed, let $L\colon \mathcal{D}^1 \to \mathcal{D}^0$ be an orientation-preserving labeling. If $v \in S^2$ is a 1-vertex (i.e., a vertex in $\mathcal{D}^1$), then $L(v)$ is a 0-vertex (i.e., a vertex in $\mathcal{D}^0$). Set $f(v) = L(v)$. This defines f on the 0-skeleton of $\mathcal{D}^1$. To extend this to the 1-skeleton of $\mathcal{D}^1$, let e be an arbitrary 1-edge. Then $e' = L(e)$ is a 0-edge. Moreover, if u and v are the 1-vertices that are the endpoints of e, then $u' = f(u) = L(u)$ and $v' = f(v) = L(v)$ are distinct 0-vertices contained in e'. Hence they are the endpoints of e'. So we can extend f to e by choosing a homeomorphism of e onto e' that agrees with f on the endpoints of e. In this way we can continuously extend f to the 1-skeleton of $\mathcal{D}^1$ so that $f|\tau$ is a homeomorphism of τ onto $L(\tau)$ whenever $\tau \in \mathcal{D}^1$ is a cell of dimension ≤ 1.

If X is an arbitrary 1-tile, then ∂X is a subset of the 1-skeleton of $\mathcal{D}^1$ and hence f is already defined on ∂X. Then $f|\partial X$ is a continuous mapping of ∂X into the boundary $\partial X'$ of the 0-tile $X' = L(X)$. The map $f|\partial X$ is injective. Indeed, suppose that $u, v \in \partial X$ and $f(u) = f(v)$. Then there exist unique 1-cells $\sigma, \tau \subset \partial X$ of dimension ≤ 1 such that $u \in \mathrm{int}(\sigma)$ and $v \in \mathrm{int}(\tau)$. Then

$$f(u) = f(v) \in \mathrm{int}(f(\sigma)) \cap \mathrm{int}(f(\tau)) = \mathrm{int}(L(\sigma)) \cap \mathrm{int}(L(\tau))$$

and so the 1-cells $L(\sigma)$ and $L(\tau)$ must be the same. Since L is a labeling and $\sigma, \tau \subset X \in \mathcal{D}^1$, it follows that $\sigma = \tau$. As the map f restricted to the 1-cell $\sigma = \tau$ is injective, we conclude $u = v$ as desired.

So $f|\partial X$ is a continuous and injective map of ∂X into $\partial X'$, and hence a homeomorphism between these sets. We conclude that f can be extended to a homeomorphism of X onto X'. These extensions on different 1-tiles paste together to a continuous map $f\colon S^2 \to S^2$ that is cellular and compatible with the given labeling. Moreover, $f|X$ is orientation-preserving for each 1-tile X as follows from the fact that the labeling is orientation-preserving. By Lemma 5.24 (i) and (ii) the map f is a postcritically-finite branched covering map. In particular, f is a homeomorphism or a Thurston map. This shows that a map with the stated properties exists.

Suppose f is a Thurston map. To show uniqueness up to Thurston equivalence, let $g\colon S^2 \to S^2$ be another continuous map that is cellular for $(\mathcal{D}^1, \mathcal{D}^0)$ and compatible with L. We will prove that g is a Thurston map that is equivalent to f.

First note that for each cell $\tau \in \mathcal{D}^1$, the maps $f|\tau$ and $g|\tau$ are homeomorphisms of τ onto $L(\tau) \in \mathcal{D}^0$. Hence $\varphi_\tau = (g|\tau)^{-1} \circ (f|\tau)$ is a homeomorphism of τ onto itself. The family φ_τ, $\tau \in \mathcal{D}^1$, of these homeomorphism is obviously compatible under inclusions: if $\sigma, \tau \in \mathcal{D}^1$ and $\sigma \subset \tau$, then $\varphi_\tau(p) = \varphi_\sigma(p)$ for all $p \in \sigma$.

Using this, we can define a map $\varphi\colon S^2 \to S^2$ as follows. For $p \in S^2$ pick $\tau \in \mathcal{D}^1$ with $p \in \tau$. Then set $\varphi(p) = \varphi_\tau(p)$. The compatibility properties of the homeomorphisms φ_τ imply that φ is well-defined. Indeed, suppose that τ, τ' are cells in $\mathcal{D}^1$ with $p \in \tau \cap \tau'$. There exists a unique cell $\sigma \in \mathcal{D}^1$ with $p \in \mathrm{int}(\sigma)$. It follows from Lemma 5.3 (ii) that $\sigma \subset \tau \cap \tau'$. Hence

$$\varphi_\tau(p) = \varphi_\sigma(p) = \varphi_{\tau'}(p).$$

It is clear that $g \circ \varphi = f$. Moreover, $\varphi|\tau = \varphi_\tau$ is a homeomorphism of τ onto itself whenever $\tau \in \mathcal{D}^1$. By Lemma 5.11 (i) and (iii) this implies that φ is a homeomorphism of S^2 onto itself that is isotopic to id_{S^2} rel. $\mathbf{V}^1$, where $\mathbf{V}^1$ is the set of 1-vertices.

The sets of postcritical points of f is contained in the set of 0-vertices and hence in $\mathbf{V}^1$. So if we use the facts that $g \circ \varphi = f$ and that φ is isotopic to id_{S^2} rel. $\mathbf{V}^1$, then Lemma 2.5 (with $h_1 = \varphi$ and $h_0 = \mathrm{id}_{S^2}$) shows that g is a Thurston map that is Thurston equivalent to f. $\qquad\square$

5.6. Flowers

Throughout this section $f\colon S^2 \to S^2$ is a given Thurston map with $\#\operatorname{post}(f) \geq 3$ and $\mathcal{C} \subset S^2$ a Jordan curve with $\operatorname{post}(f) \subset \mathcal{C}$. We consider the cell decompositions $\mathcal{D}^n = \mathcal{D}^n(f,\mathcal{C})$ and use the related terminology and notation as discussed in Section 5.3

The results in this section are based on the following concept.

DEFINITION 5.27 (n-Flowers). Let $n \in \mathbb{N}_0$, and $p \in S^2$ be an n-vertex. Then the n-*flower* of p is defined as

$$W^n(p) := \bigcup \{\operatorname{int}(c) : c \in \mathcal{D}^n, \ p \in c\}.$$

So the n-flower $W^n(p)$ of the n-vertex p is the union of the interiors of all cells in the cycle of p in $\mathcal{D}^n$ (see Figure 5.1 as well as Lemma 5.9 (v) and the discussion after this lemma).

The main reason why we introduce flowers is the following. Consider a simply connected domain $U \subset S^2$ not containing a postcritical point of f and branches g_n of f^{-n} defined on U. Then it may happen that the number of n-tiles intersecting $g_n(U)$ is unbounded as $n \to \infty$, even if the diameter of U (with respect to some base metric on S^2) is small. For example, this happens when f has a periodic critical point p (see Section 18.2), and U spirals around one of the points in the cycle generated by p. However if $\operatorname{diam}(U)$ is sufficiently small, then $g_n(U)$ is always contained in one n-flower as we shall see. Similar issues that are resolved by the use of flowers are addressed in Lemma 5.37 and Lemma 5.38.

We first prove some basic properties of flowers.

LEMMA 5.28. *Let $n \in \mathbb{N}_0$, and $p \in S^2$ be an n-vertex. As in Lemma 5.9 let $e_1, \ldots, e_d$ be the n-edges and $X_1, \ldots, X_d$ be the n-tiles of the cycle of p, where $d \in \mathbb{N}$, $d \geq 2$, is the length of the cycle.*

(i) *Then $d = 2\deg(f^n, p)$ and the set $W^n(p)$ is homeomorphic to $\mathbb{D}$, i.e., it is an open, connected, and simply connected neighborhood of p. It contains no other n-vertex, and we have*

(5.11) $$W^n(p) = \{p\} \cup \bigcup_{i=1}^d \operatorname{int}(X_i) \cup \bigcup_{i=1}^d \operatorname{int}(e_i)$$
$$= S^2 \setminus \bigcup \{c \in \mathcal{D}^n : c \in \mathcal{D}^n, \ p \notin c\}.$$

(ii) *We have $\overline{W^n(p)} = X_1 \cup \cdots \cup X_d$, and the set $\partial W^n(p)$ is the union of all n-edges e with $p \notin e$ and $e \subset \partial X_i$ for some $i \in \{1, \ldots, d\}$.*

(iii) *If c is an arbitrary n-cell, then either $p \in c$ and $c \subset \overline{W^n(p)}$, or $c \subset S^2 \setminus W^n(p)$.*

Note that by (i) each n-vertex p is contained in precisely $d = 2 \deg(f^n, p)$ distinct n-edges and in precisely d distinct n-tiles.

PROOF. (i) By Remark 5.25 the length d of the cycle of the vertex p (in the cell decomposition $\mathcal{D}^n$) is a multiple $d = kd'$ of the length d' of the cycle of the image point $q = f^n(p)$ (in the cell decomposition $\mathcal{D}^0$), where k is the degree of f^n at p. Since $d' = 2$, we have $d = 2 \deg(f^n, p)$ as claimed.

The first equality in (5.11) follows from Lemma 5.9 (v). Based on this, the argument in Case 3 of the proof of Lemma 5.24 shows that there is a homeomorphism of the set $W^n(p)$ onto $\mathbb{D}$. Hence $W^n(p)$ is open and simply connected, and it follows from the first equality in (5.11) that $W^n(p)$ contains no other n-vertex than p.

Let $M = S^2 \setminus \bigcup \{c \in \mathcal{D}^n : c \in \mathcal{D}^n, \ p \notin c\}$. If $x \in W^n(p)$, then x is an interior point in one of the cells τ forming the cycle of p. So if c is any n-cell with $x \in c$, then $\tau \subset c$ by Lemma 5.3 (ii). This implies $p \in c$, and so $x \in M$ by definition of M. Hence $W^n(p) \subset M$.

Conversely, if $x \in M$, let τ be an n-cell of smallest dimension that contains x. Obviously, $x \in \mathrm{int}(\tau)$. On the other hand, the definition of M implies that $p \in \tau$. Hence τ is a cell in the cycle of p, and so $x \in W^n(p)$. We conclude $M \subset W^n(p)$, and so $M = W^n(p)$ as desired.

(ii) Equation (5.11) implies $\overline{W^n(p)} = X_1 \cup \cdots \cup X_n$.

Every point $x \in \partial W^n(p)$ is contained in one of the sets ∂X_i. Since $W^n(p)$ is open, the point x is not contained in $\{p\} \cup \mathrm{int}(e_{i-1}) \cup \mathrm{int}(e_i) \subset W^n(p)$ and hence belongs to one of the n-edges e in the boundary of X_i distinct from e_{i-1} and e_i; note that there exists such an edge, because each n-tile is an m-gon, where $m = \#\mathrm{post}(f) \geq 3$, and so contains more than two n-edges in its boundary. Then $p \notin e$, and x is contained in an n-edge with the desired properties.

Conversely, if e is an n-edge with $p \notin e$ and $e \subset \partial X_i$, then $e \subset S^2 \setminus W^n(p)$ by (5.11), and $e \subset X_i \subset \overline{W^n(p)}$. Hence $e \subset \partial W^n(p)$.

(iii) This follows from (i) and (5.11).   □

Note that if we color tiles as in Lemma 5.21, then the colors of the tiles $X_1, \ldots, X_d$ associated with an n-flower as in the previous lemma will alternate.

LEMMA 5.29. *Let $k, n \in \mathbb{N}_0$. Then the following statements are true:*

(i) *If $p \in S^2$ is an $(n + k)$-vertex, then f^k maps the edges and tiles in the cycle of p to the edges and tiles in the cycle of the n-vertex $f^k(p)$ in cyclic order in an m-to-1 fashion, where $m = \deg(f^k, p)$. Moreover, $f^k(W^{n+k}(p)) = W^n(f^k(p))$.*

(ii) *If $q \in S^2$ is an n-vertex, then the connected components of $f^{-k}(W^n(q))$ are the $(n + k)$-flowers $W^{n+k}(p)$, $p \in f^{-k}(q)$.*

(iii) *A connected set $K \subset S^2$ is contained in an $(n + k)$-flower if and only if $f^k(K)$ is contained in an n-flower.*

(iv) *The set of all n-flowers $W^n(p)$, $p \in \mathbf{V}^n$, forms an open cover of S^2.*

In the proof we will explain the precise meaning of the first statement in (i).

PROOF. (i) It is clear that $q = f^k(p)$ is an n-vertex. Let e_i' and X_i' for $i \in \mathbb{N}$ be the $(n + k)$-edges and $(n + k)$-tiles in the cycle of p, respectively. Here we can choose the indexing so that it has precise period $d' = 2\deg(f^{n+k}, p)$, i.e., $e_{d'+i}' = e_i'$ and $X_{d'+i}' = X_i'$ for $i \in \mathbb{N}$ and d' is the smallest possible number here, because it is the length of the cycle of p (see Lemma 5.28 (i)).

Define $e_i = f^k(e_i')$ and $X_i = f^k(X_i')$ for $i \in \mathbb{N}$. Since the map f^k is cellular for $(\mathcal{D}^{n+k}, \mathcal{D}^n)$, it follows from Remark 5.25 that e_i and X_i for $i \in \mathbb{N}$ are the n-edges and n-tiles in the cycle of q. Here $e_{d+i} = e_i$ and $X_{d+i} = X_i$ for $i \in \mathbb{N}$ with $d = 2\deg(f^n, q)$ and again d is the smallest number with this property. In this sense, f^k maps the edges and tiles in the cycle of p to the edges and tiles in the cycle of q in cyclic order.

The map f^k between these cycles is m-to-1 with $m = \deg(f^k, p)$, because each edge or tile in the cycle of q has precisely

$$m = d'/d = \deg(f^{n+k}, p)/\deg(f^n, q) = \deg(f^k, p)$$

distinct preimages in the cycle of p.

Note that we also have $f^k(\mathrm{int}(e_i')) = \mathrm{int}(e_i)$ and $f^k(\mathrm{int}(X_i')) = \mathrm{int}(X_i)$ for $i \in \mathbb{N}$. This and (5.11) imply the last statement in (i).

(ii) If $p \in f^{-k}(q)$, then p is an $(n+k)$-vertex. By (i) the $(n+k)$-flower $W^{n+k}(p)$ is an open and connected subset of $f^{-k}(W^n(q))$. Suppose that $x \in \partial W^{n+k}(p)$. Then by Lemma 5.28 (ii) there exists an $(n + k)$-tile X', and an $(n + k)$-edge e' with $p \in X'$, $p \notin e'$, and $x \in e' \subset \partial X'$. Then $X = f^k(X')$ is an n-tile, $e = f^k(e')$ is an n-edge, and $q \in X$, $f(x) \in e \subset \partial X$. Since $f^k|X'$ is a homeomorphism of X' onto X, we also have $q \notin e$. Lemma 5.28 (ii) implies that $f^k(x) \in \partial W^n(q)$, and so $f^k(x) \notin W^n(q)$, because flowers are open sets.

We conclude that $x \in S^2 \setminus f^{-k}(W^n(q))$, and so $\partial W^{n+k}(p) \subset S^2 \setminus f^{-k}(W^n(q))$. It now follows from Lemma 5.4 that $W^{n+k}(p)$ is a connected component of $f^{-k}(W^n(q))$.

Conversely, suppose that U is a connected component of the set $f^{-k}(W^n(q))$. Then U is an open set and so it meets the interior $\mathrm{int}(X')$ of some $(n+k)$-tile X'. Then $X = f^k(X')$ is an n-tile that meets $W^n(q)$. Hence $q \in X$, and so there exists an $(n+k)$-vertex $p \in X'$ with $f^k(p) = q$.

Then by the first part of the proof, the set $W^{n+k}(p)$ is a connected component of $f^{-k}(W^n(q))$. Since $W^{n+k}(p)$ contains the set $\mathrm{int}(X')$ and so meets U, we must have $W^{n+k}(p) = U$.

(iii) Suppose K is contained in the $(n+k)$-flower $W^{n+k}(p)$. Then by (i) the set $f^k(W^{n+k}(p)) = W^n(f^k(p))$ is an n-flower and it contains $f^k(K)$.

Conversely, if $f^k(K)$ is contained in the n-flower $W^n(q)$, then K is a connected set in $f^{-k}(W^n(q))$. Hence K lies in a connected component of $f^{-k}(W^n(q))$, and hence in an $(n+k)$-flower by (ii).

(iv) We know by Lemma 5.28 (i) that flowers are open sets. If $x \in S^2$ is arbitrary, then there exists an n-cell c such that $x \in \mathrm{int}(c)$. We can find an n-vertex p such that $p \in c$. Then $x \in \mathrm{int}(c) \subset W^n(p)$. So the n-flowers form indeed an open cover of S^2. $\qquad\square$

Similar to the definition of an n-flower for an n-vertex, one can also define an *edge flower* for an n-edge. These sets provide "canonical" neighborhoods for n-vertices and n-edges defined in terms of n-cells.

DEFINITION 5.30 (Edge flowers). Let $n \in \mathbb{N}_0$, and e be an n-edge. Then the *n-edge flower* of e is defined as

$$W^n(e) := \bigcup \{\mathrm{int}(c) : c \in \mathcal{D}^n, \ c \cap e \neq \emptyset\}.$$

We list some properties of edge flowers. They correspond to similar properties of n-flowers as in Lemma 5.28. Note that in contrast to an n-flower, an edge flower $W^n(e)$ will not be simply connected in general (for example, if there is another n-edge e' with the same endpoints as e).

LEMMA 5.31. *Let e be an n-edge whose endpoints are the n-vertices u and v.*

(i) *Then $W^n(e)$ is an open set containing e, and*

$$(5.12) \quad W^n(e) = W^n(u) \cup W^n(v) = S^2 \setminus \bigcup \{c : c \in \mathcal{D}^n, \ c \cap e = \emptyset\}.$$

(ii) *We have* $\overline{W^n(e)} = \bigcup\{X \in \mathbf{X}^n : X \cap e \neq \emptyset\}$ *and*

$$\partial W^n(e) = \bigcup\{c \in \mathcal{D}^n : c \cap e = \emptyset \text{ and}$$

$$\text{there exists } X \in \mathbf{X}^n \text{ with } X \cap e \neq \emptyset \text{ and } c \subset \partial X\},$$

where each n-cell c in the last union either consists of one n-vertex or is an n-edge.

(iii) *If c is an arbitrary n-cell, then either $c \cap e \neq \emptyset$ and $c \subset \overline{W^n(e)}$, or $c \subset S^2 \setminus W^n(e)$.*

PROOF. (i) It follows from Lemma 5.3 (i) that an n-cell c meets e if and only if it contains one of the endpoints u and v of e. Hence $W^n(e) = W^n(u) \cup W^n(v)$ by the definition of flowers. By Lemma 5.28 (i) this implies that $W^n(e)$ is open, and, since e is an edge in the cycles of u and v, we also have

$$e = \{u\} \cup \text{int}(e) \cup \{v\} \subset W^n(u) \cup W^n(v) = W^n(e).$$

Let $M = S^2 \setminus \bigcup\{c : c \in \mathcal{D}^n, \ c \cap e = \emptyset\}$. If an n-cell c does not meet e, then it contains neither u nor v. Hence by (5.11) we have $S^2 \setminus M \subset (S^2 \setminus W^n(u)) \cap (S^2 \setminus W^n(v)) = S^2 \setminus W^n(e)$, and so $W^n(e) \subset M$.

Conversely, let $x \in M$ be arbitrary, and c be the unique n-cell c such that $x \in \text{int}(c)$. Then $c \cap e \neq \emptyset$ and so $u \in c$ or $v \in c$. It follows that $x \in W^n(u) \cup W^n(v) = W^n(e)$. We conclude that $M \subset W^n(e)$, and so $M = W^n(e)$ as claimed.

(ii) By Lemma 5.3 (i) an n-tile X meets e if and only if X contains u or v. Hence by (i) and Lemma 5.28 (ii) we have

$$\overline{W^n(e)} = \overline{W^n(u)} \cup \overline{W^n(v)} = \bigcup\{X \in \mathbf{X}^n : X \cap e \neq \emptyset\}$$

as desired.

For the second claim suppose that c is an n-cell and X an n-tile with $c \cap e = \emptyset$, $X \cap e \neq \emptyset$, and $c \subset \partial X$. Then $c \subset S^2 \setminus W^n(e)$ and c must be an n-edge or consist of an n-vertex. Moreover, $c \subset X \subset \overline{W^n(e)}$. It follows that $c \subset \partial W^n(e)$.

Conversely, let x be a point in $\partial W^n(e)$. Then by (i) the point x is also a boundary point of $W^n(u)$ or $W^n(v)$, say $x \in \partial W^n(u)$.

By Lemma 5.28 (ii) there exists an n-edge e' and an n-tile X with $x \in e'$, $u \in X$, $u \notin e'$ and $e' \subset \partial X$. If x is an n-vertex, we let $c = \{x\}$. Then c is an n-cell and we have $c \cap e = \emptyset$, because $W^n(e)$ is an open neighborhood of e and c lies in $\partial W^n(e) \subset S^2 \setminus W^n(e)$. Moreover, $X \cap e \neq \emptyset$ and $c \subset e' \subset \partial X$. So c is an n-cell with the desired properties containing x.

If x is not a vertex we put $c = e'$. Again if $c \cap e = e' \cap e = \emptyset$, then c is an n-cell with the desired properties containing x.

The other case, where $e' \cap e \neq \emptyset$, leads to a contradiction. Indeed, then we have $v \in e'$. Moreover, since x is not a vertex, it follows that $x \in \text{int}(e')$; but then $x \in \text{int}(e') \subset W^n(v) \subset W^n(e)$ which is impossible, because $x \in \partial W^n(e) \subset S^2 \setminus W^n(e)$.

(iii) If c is an n-cell and $c \cap e = \emptyset$, then $c \subset S^2 \setminus W^n(e)$. If $c \cap e \neq \emptyset$, then c contains u or v, and so $c \subset \overline{W^n(u)} \cup \overline{W^n(v)} = \overline{W^n(e)}$. $\square$

5.7. Joining opposite sides

In this section $f \colon S^2 \to S^2$ will again be a Thurston map with $\# \operatorname{post}(f) \geq 3$ and $\mathcal{C} \subset S^2$ be a Jordan curve with $\operatorname{post}(f) \subset \mathcal{C}$. We fix a base metric d on S^2 that induces the given topology and consider the cell decompositions $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$ as discussed in Section 5.3.

We will define a constant $\delta_0 > 0$ such that any connected set of diameter $< \delta_0$ (with respect to the base metric d) is contained in a single 0-flower (as introduced in Section 5.6). There is a slight difference for the cases $\# \operatorname{post}(f) = 3$ and $\# \operatorname{post}(f) \geq 4$. In order to treat these two cases simultaneously, the following definition is useful.

DEFINITION 5.32 (Joining opposite sides). A set $K \subset S^2$ *joins opposite sides* of $\mathcal{C}$ if $\# \operatorname{post}(f) \geq 4$ and K meets two disjoint 0-edges, or if $\# \operatorname{post}(f) = 3$ and K meets all three 0-edges.

We will mostly use this for connected sets K (when the phrase "joining" really makes sense), but it is convenient to allow arbitary sets here. We now define

$$(5.13) \quad \delta_0 = \delta_0(f, \mathcal{C}) = \inf\{\operatorname{diam}(K) : K \subset S^2 \text{ is a set}$$
$$\text{joining opposite sides of } \mathcal{C}\}.$$

Then $\delta_0 > 0$. Indeed, if $\# \operatorname{post}(f) = 4$, then

$$\delta_0 = \min\{\operatorname{dist}(e, e') : e \text{ and } e' \text{ are disjoint 0-edges}\} > 0.$$

If $\# \operatorname{post}(f) = 3$ and we had $\delta_0 = 0$, then it would follow from a simple limiting argument that the three 0-edges had a common point. This is absurd.

LEMMA 5.33. *A connected set $K \subset S^2$ joins opposite sides of $\mathcal{C}$ if and only if K is not contained in a single 0-flower (of a 0-vertex).*

PROOF. If K is contained in a 0-flower $W^0(p)$, where $p \in \mathcal{C}$ is a 0-vertex, then K meets at most two 0-edges, namely the ones that have the common endpoint p. So K does not join opposite sides of $\mathcal{C}$.

Conversely, suppose K does not join opposite sides of $\mathcal{C}$. We have to show that K is contained in some 0-flower. Note that K cannot meet three distinct 0-edges.

If K does not meet any 0-edge, then K does not meet $\mathcal{C}$ and is hence contained in the interior of one of the two 0-tiles. This implies that K is actually contained in every 0-flower.

If K meets only one 0-edge e, then K is contained in the 0-flowers $W^0(u)$ and $W^0(v)$, where u and v are the endpoints of e.

If K meets two edges, then these edges share a common endpoint $v \in \mathbf{V}^0 = \mathrm{post}(f)$. This is always true if $\#\mathrm{post}(f) = 3$ and follows from the fact that K does not join opposite sides of $\mathcal{C}$ if $\#\mathrm{post}(f) \geq 4$. Moreover, K cannot meet a third 0-edge which implies that $K \subset W^0(v)$. $\qquad\square$

By the previous lemma every connected set $K \subset S^2$ satisfying $\mathrm{diam}(K) < \delta_0$ is contained in a 0-flower.

LEMMA 5.34. *Let $n \in \mathbb{N}_0$, and $\delta_0 > 0$ be as in (5.13).*

(i) *If $K \subset S^2$ is a connected set with $\mathrm{diam}(K) < \delta_0$, then each component of $f^{-n}(K)$ is contained in some n-flower.*

(ii) *If $\gamma\colon [0,1] \to S^2$ is a path such that $\mathrm{diam}(\gamma) < \delta_0$, then each lift $\widetilde{\gamma}$ of γ by f^n has an image that is contained in some n-flower.*

Here by definition a *lift* of γ by f^n is any path $\widetilde{\gamma}\colon [0,1] \to S^2$ with $\gamma = f^n \circ \widetilde{\gamma}$.

PROOF. (i) The set K is contained in some 0-flower $W^0(p)$, $p \in \mathbf{V}^0$, by Lemma 5.33 and the definition of δ_0. So if K' is a component of $f^{-n}(K)$, then K' is contained in a component of $f^{-n}(W^0(p))$, and hence in an n-flower by Lemma 5.29 (ii).

(ii) The reasoning is exactly the same as in (i). The image of γ is contained in some 0-flower; by Lemma 5.29 (ii) this implies that the image of $\widetilde{\gamma}$ is contained in an n-flower. $\qquad\square$

We will often have to estimate how many tiles are needed to connect certain points. If we have a condition that is formulated "at the top level", i.e., for connecting points in $\mathcal{C}$, then the map f^n can be used to translate this to n-tiles.

LEMMA 5.35. *Let $n \in \mathbb{N}_0$, and $K \subset S^2$ be a connected set. If there exist two disjoint n-cells σ and τ with $K \cap \sigma \neq \emptyset$ and $K \cap \tau \neq \emptyset$, then $f^n(K)$ joins opposite sides of $\mathcal{C}$.*

PROOF. It suffices to show that K is not contained in any n-flower, because then $f^n(K)$ is not contained in any 0-flower (Lemma 5.29 (iii)) and so $f^n(K)$ joins opposite sides of $\mathcal{C}$ (Lemma 5.33). We consider several cases.

Case 1: One of the cells is an n-vertex, say $\sigma = \{v\}$, where $v \in \mathbf{V}^n$. Then $v \in K$, so the only n-flower that K could possibly be contained in is $W^n(v)$, because no other n-flower contains the n-vertex v. But since σ and τ are disjoint, we have $v \notin \tau$, and so $\tau \subset S^2 \setminus W^n(v)$. Hence $K \cap (S^2 \setminus W^n(v)) \neq \emptyset$, and so $W^n(v)$ does not contain K.

Case 2: Suppose one of the cells is an n-edge, say $\sigma = e \in \mathbf{E}^n$. Then e has two endpoints $u, v \in \mathbf{V}^n$. The only n-flowers that meet e are $W^n(u)$ and $W^n(v)$; so these n-flowers are the only ones that could possibly contain K. But the set $W^n(e) = W^n(u) \cup W^n(v)$ does not contain K, because K meets the set τ which lies in the complement of $W^n(e)$.

Case 3: One of the cells in an n-tile, say $\sigma \in \mathbf{X}^n$. Then K meets ∂X. Since ∂X consists of n-edges, the set K meets an n-edge disjoint from τ. So we are reduced to Case 2. $\qquad\square$

For $n \in \mathbb{N}_0$ we denote by $D_n(f, \mathcal{C})$ the minimal number of n-tiles required to form a connected set joining opposite sides of $\mathcal{C}$; more precisely,

$$(5.14) \quad D_n(f, \mathcal{C}) = \min \Big\{ N \in \mathbb{N} : \text{there exist } X_1, \ldots, X_N \in \mathbf{X}^n$$

$$\text{such that } K = \bigcup_{j=1}^{N} X_j \text{ is connected and joins opposite sides of } \mathcal{C} \Big\}.$$

We simply write D_n for $D_n(f, \mathcal{C})$ if f and $\mathcal{C}$ are clear from the context (as in this section).

From Lemma 5.35 we can immediately derive the following consequence.

LEMMA 5.36. *Let $n, k \in \mathbb{N}_0$. Every set of $(n+k)$-tiles whose union is connected and meets two disjoint n-cells contains at least D_k elements.*

PROOF. Suppose K is a union of $(n+k)$-tiles with the stated properties. Then the images of these tiles under f^n are k-tiles and $f^n(K)$ joins opposite sides of $\mathcal{C}$ by Lemma 5.35. Hence there exist at least D_k distinct k-tiles in the union forming $f^n(K)$ and hence at least D_k distinct $(n + k)$-tiles in K. $\qquad\square$

The following two lemmas give some motivation for introducing flowers. Namely, the number of $(n - 1)$-tiles or the number of $(n +$

1)-tiles required to cover some n-tile X may not be bounded by a constant independent of X and n. Similarly, in general there will be no universal bound on the number of n-tiles defined with respect to a different Jordan curve $\widetilde{\mathcal{C}}$ needed to cover X. Both issues are resolved by considering flowers instead of tiles.

LEMMA 5.37. *There exists $M \in \mathbb{N}$ with the following properties:*

(i) *Each n-tile, $n \in \mathbb{N}$, can be covered by M $(n-1)$-flowers.*

(ii) *Each n-tile, $n \in \mathbb{N}_0$, can be covered by M $(n+1)$-flowers.*

For easier formulation of this lemma and the subsequent proof, we assume for simplicity that a cover by *at most M* elements contains precisely M elements. This can always be achieved by repetition of elements in the cover.

PROOF. It suffices to consider the statements (i) and (ii) separately and find a corresponding number M for each of them.

(i) Let $\delta_0 > 0$ be as in (5.13). Then there exists $M \in \mathbb{N}$ such that each of the finitely many 1-tiles X is a union of M connected sets $U \subset X$ with $\operatorname{diam}(U) < \delta_0$. If Y is an arbitrary n-tile, $n \geq 1$, then $Z = f^{n-1}(Y)$ is a 1-tile and $f^{n-1}|Y$ a homeomorphism of Y onto Z. Hence Y is a union of M sets of the form $(f^{n-1}|Y)^{-1}(U)$, where $U \subset Z$ is connected and $\operatorname{diam}(U) < \delta_0$. Each set $(f^{n-1}|Y)^{-1}(U)$ is connected and so by Lemma 5.34 (i) it lies in an $(n-1)$-flower. Hence Y can be covered by M $(n-1)$-flowers.

(ii) There exists $M \in \mathbb{N}$ such that each of the two 0-tiles X can be covered by M connected sets $U \subset X$ with $\operatorname{diam}(f(U)) < \delta_0$. If Y is an arbitrary n-tile, then $Z = f^n(Y)$ is a 0-tile. By the same reasoning as above, the set Y is a union of M sets of the form $(f^n|Y)^{-1}(U)$, where $U \subset Z$ is connected and $\operatorname{diam}(f(U)) < \delta_0$.

Then $U' = (f^n|Y)^{-1}(U)$ is connected, and $f^{n+1}(U') = f(U)$ which implies $\operatorname{diam}(f^{n+1}(U')) < \delta_0$. Hence by Lemma 5.34 (i) the set U' is contained in some $(n+1)$-flower. Since M of the sets U' cover Y, it follows that each n-tile can be covered by M $(n+1)$-flowers. $\square$

LEMMA 5.38. *Let $\mathcal{C}$ and $\widetilde{\mathcal{C}}$ be two Jordan curves in S^2 that both contain $\operatorname{post}(f)$. Then there exists a number M such that each n-tile for $(f, \widetilde{\mathcal{C}})$ is covered by M n-flowers for $(f, \mathcal{C})$.*

PROOF. The proof is very similar to the proof of Lemma 5.37.

Let $\delta_0 = \delta_0(f, \mathcal{C}) > 0$ be the number as defined in (5.13). There exists a number M such that each of the two 0-tiles X for $(f, \widetilde{\mathcal{C}})$ is a union of M connected sets $U \subset X$ with $\operatorname{diam}(U) < \delta_0$. If Y is an

arbitrary n-tile for $(f, \widetilde{\mathcal{C}})$, then $Z = f^n(Y)$ is a 0-tile for $(f, \widetilde{\mathcal{C}})$ and $f^n|Y$ is a homeomorphism of Y onto Z. Hence Y is a union of M sets of the form $(f^n|Y)^{-1}(U)$, where $U \subset Z$ is connected and $\mathrm{diam}(U) < \delta_0$. Each set $(f^n|Y)^{-1}(U)$ is connected and so by Lemma 5.34 (i) it lies in an n-flower for $(f, \mathcal{C})$. Hence Y can be covered by M such n-flowers. $\square$

Expansion

In this chapter we revisit the notion of expansion for Thurston maps (see Definition 2.2) and study it in greater depth. We will establish basic properties of this concept.

In Section 6.1 the main result is Proposition 6.3 which gives several conditions that are equivalent to our notion of expansion. In particular, one of these conditions (namely, condition (iv) in Proposition 6.3) can be formulated in terms of open covers without reference to a metric. This shows (as we remarked after Definition 2.2) that expansion is an entirely topological property of a given Thurston map.

In Section 6.2 we prove various other results about expansion. For example, in Lemma 6.6 we show that a Thurston map is expanding if it uniformly expands the length of paths with respect to an underlying length metric. This result was already used in our characterization of expanding rational Thurston maps (see the proof of Proposition 2.3). It is not known if for every expanding Thurston map there is a length metric with respect to which it is expanding.

In Section 6.3 we return to Lattès-type maps. We show that such a map is expanding if and only if each eigenvalue of the linear part L_A of the affine map A in Definition 3.3 has absolute value > 1 (see Proposition 6.11).

6.1. Definition of expansion revisited

Let S^2 be a 2-sphere. In the following, it is often convenient to formulate some essentially topological properties in metric terms. For this we fix a base metric on S^2 that induces the given topology. In this and the next section notation for metric terms will refer to this base metric unless otherwise indicated.

Let $f \colon S^2 \to S^2$ be a Thurston map and $\mathcal{C} \subset S^2$ be a Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}$. For $n \in \mathbb{N}_0$ we consider the cell decompositions $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$ as given by Definition 5.14 with corresponding set $\mathbf{X}^n = \mathbf{X}^n(f, \mathcal{C})$ of n-tiles. Recall (from the beginning of Section 2.3) that $\mathrm{mesh}(f, n, \mathcal{C})$ is defined as the supremum of the diameters of the connected components of $f^{-n}(S^2 \setminus \mathcal{C}) = S^2 \setminus f^{-n}(\mathcal{C})$. We know that the

n-tiles for $(f, \mathcal{C})$ are precisely the closures of the connected components of $S^2 \setminus f^{-n}(\mathcal{C})$ (see Proposition 5.17 (v)), and so

$$\mathrm{mesh}(f, n, \mathcal{C}) = \max_{X \in \mathbf{X}^n} \mathrm{diam}(X).$$

Thus a Thurston map f is expanding (see Definition 2.2) if and only if there is a Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}$ such that

$$(6.1) \qquad\qquad \max_{X \in \mathbf{X}^n} \mathrm{diam}(X) \to 0 \text{ as } n \to \infty,$$

where the tiles are defined with respect to $(f, \mathcal{C})$. We first want to convince ourselves that this condition is independent of the choice of the curve $\mathcal{C}$.

LEMMA 6.1. *Let* $f \colon S^2 \to S^2$ *be a Thurston map and* $\mathcal{C}, \widetilde{\mathcal{C}} \subset S^2$ *be Jordan curves with* $\mathrm{post}(f) \subset \mathcal{C}, \widetilde{\mathcal{C}}$. *Then*

$$\lim_{n \to \infty} \mathrm{mesh}(f, n, \mathcal{C}) = 0 \text{ if and only if } \lim_{n \to \infty} \mathrm{mesh}(f, n, \widetilde{\mathcal{C}}) = 0.$$

PROOF. Let $\mathcal{C}, \widetilde{\mathcal{C}} \subset S^2$ be as in the statement of the lemma. Assume that $\lim_{n \to \infty} \mathrm{mesh}(f, n, \mathcal{C}) = 0$. This means that f is expanding. Then

$$\max_{X \in \mathbf{X}^n} \mathrm{diam}(X) = \mathrm{mesh}(f, n, \mathcal{C}) \to 0$$

as $n \to \infty$, where $\mathbf{X}^n$ is the set of n-tiles for $(f, \mathcal{C})$. Lemma 5.28 (ii) implies that

$$(6.2) \qquad\qquad \mathrm{diam}(W^n(v)) \le 2 \max_{X \in \mathbf{X}^n} \mathrm{diam}(X)$$

for each n-flower for $(f, \mathcal{C})$.

Now we consider tiles for $(f, \widetilde{\mathcal{C}})$. By Lemma 5.38 that there exists a number $M \in \mathbb{N}$ such that each n-tile for $(f, \widetilde{\mathcal{C}})$ can be covered by M n-flowers for $(f, \mathcal{C})$. If a connected set is covered by a finite union of connected sets, then its diameter is bounded by the sum of the diameters of the sets in the union. Combining this with (6.2), we conclude that

$$
\begin{aligned}
\mathrm{mesh}(f, n, \widetilde{\mathcal{C}}) \;&=\; \max\{\mathrm{diam}(\widetilde{X}) : \widetilde{X} \text{ is an } n\text{-tile for } (f, \widetilde{\mathcal{C}})\} \\
&\le\; M \max_{p \in \mathbf{V}^n} \mathrm{diam}(W^n(p)) \\
&\le\; 2M \max_{X \in \mathbf{X}^n} \mathrm{diam}(X) \\
&=\; 2M \,\mathrm{mesh}(f, n, \mathcal{C}).
\end{aligned}
$$

Here $\mathbf{V}^n$ denotes the set of n-vertices for $(f, \mathcal{C})$. Hence $\mathrm{mesh}(f, n, \widetilde{\mathcal{C}}) \to 0$ as $n \to \infty$ as desired.

The other implication is obtained by reversing the roles of $\mathcal{C}$ and $\widetilde{\mathcal{C}}$. $\qquad\square$

The lemma shows that a Thurston map $f\colon S^2 \to S^2$ is expanding if and only if $\mathrm{mesh}(f, n, \mathcal{C}) \to 0$ as $n \to \infty$ whenever $\mathcal{C} \subset S^2$ is an *arbitrary* Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}$. In particular, expansion is a property of the map f alone and independent of the choice of the Jordan curve $\mathcal{C}$.

Lemma 5.38, which was used in the previous proof, admits an improvement for expanding Thurston maps.

LEMMA 6.2. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map. Suppose that $\mathcal{C}$ and $\widetilde{\mathcal{C}}$ are two Jordan curves in S^2 that both contain $\mathrm{post}(f)$. Then there exists a number $M \in \mathbb{N}$ with the following property: if $n, k \in \mathbb{N}_0$, then every $(n + k)$-tile for $(f, \widetilde{\mathcal{C}})$ can be covered by M n-flowers for $(f, \mathcal{C})$.*

PROOF. The argument is a small variation of the one that we used to establish Lemma 5.38.

Let $\delta_0 = \delta_0(f, \mathcal{C}) > 0$ be the number as defined in (5.13). Since f is expanding, there exists a number $M \in \mathbb{N}$ such that *each* tile X for $(f, \widetilde{\mathcal{C}})$ is a union of M connected sets $U \subset X$ with $\mathrm{diam}(U) < \delta_0$ (in the proof of Lemma 5.38 we could guarantee this only for the two 0-tiles for $(f, \widetilde{\mathcal{C}})$). Indeed, since f is expanding this is trivially true for all tiles X of sufficiently high level, because then $\mathrm{diam}(X) < \delta_0$. There are only finitely many tiles X for $(f, \widetilde{\mathcal{C}})$ with $\mathrm{diam}(X) \geq \delta_0$. The existence of a suitable constant M easily follows.

Now let $n, k \in \mathbb{N}_0$ and suppose Y is an arbitrary $(n + k)$-tile for $(f, \widetilde{\mathcal{C}})$. Then $Z = f^n(Y)$ is a k-tile for $(f, \widetilde{\mathcal{C}})$ and $f^n|Y$ is a homeomorphism of Y onto Z. Hence Y is a union of M sets of the form $(f^n|Y)^{-1}(U)$, where $U \subset Z$ is connected and $\mathrm{diam}(U) < \delta_0$. Each set $(f^n|Y)^{-1}(U)$ is connected and so by Lemma 5.34 (i) it lies in an n-flower for $(f, \mathcal{C})$. Hence Y can be covered by M such n-flowers. $\square$

Our definition of expansion is somewhat *ad hoc*, but it has the advantage that it relates to the geometry of tiles. As we will see, equivalent and maybe more conceptual descriptions can be given in terms of the behavior of open covers of S^2 under pull-backs by the iterates of the map. This shows that expansion is a topological property of the map. Our definition was based on a metric concept (namely the mesh size), but this was just for convenience.

We start with some definitions. Let $\mathcal{U}$ be an open cover of S^2. We define $\mathrm{mesh}(\mathcal{U})$ to be the supremum of all diameters of connected components of sets in $\mathcal{U}$. If $g\colon S^2 \to S^2$ is a continuous map, then the *pull-back of $\mathcal{U}$ by g* is defined as

$$g^{-1}(\mathcal{U}) = \{V : V \text{ is a connected component of } g^{-1}(U), \text{ where } U \in \mathcal{U}\}.$$

Obviously, $g^{-1}(\mathcal{U})$ is also an open cover of S^2. Similarly, we denote by $g^{-n}(\mathcal{U})$ the pull-back of $\mathcal{U}$ by g^n.

PROPOSITION 6.3. *Let* $f\colon S^2 \to S^2$ *be a Thurston map. Then the following conditions are equivalent:*

(i) *The map* f *is expanding.*

(ii) *There exists* $\delta_0 > 0$ *with the following property: if* $\mathcal{U}$ *is a cover of* S^2 *by open and connected sets that satisfies* $\mathrm{mesh}(\mathcal{U}) < \delta_0$, *then*
$$\lim_{n\to\infty} \mathrm{mesh}(f^{-n}(\mathcal{U})) = 0.$$

(iii) *There exists an open cover* $\mathcal{U}$ *of* S^2 *with*
$$\lim_{n\to\infty} \mathrm{mesh}(f^{-n}(\mathcal{U})) = 0.$$

(iv) *There exists an open cover* $\mathcal{U}$ *of* S^2 *with the following property: for every open cover* $\mathcal{V}$ *of* S^2 *there exists* $N \in \mathbb{N}$ *such that* $f^{-n}(\mathcal{U})$ *is finer than* $\mathcal{V}$ *for every* $n \in \mathbb{N}$ *with* $n > N$; *i.e., for every set* $U' \in f^{-n}(\mathcal{U})$ *there exists a set* $V \in \mathcal{V}$ *such that* $U' \subset V$.

Condition (iii) is the notion of expansion as defined by Haïssinsky-Pilgrim (see [**HP09**, Section 2.2]). So our notion of expansion agrees with the one in [**HP09**]. Condition (iv) is essentially a reformulation of (iii) in purely topological terms without reference to the base metric on S^2 (which enters in the definition of the mesh of an open cover). One can reformulate (ii) in a similar spirit. We will see in the proof below that the constant δ_0 in (ii) can be chosen to be the number from (5.13). If there exists a Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}$ and $f(\mathcal{C}) \subset \mathcal{C}$, then expansion of the map f can be characterized in yet another way (see Lemma 12.7).

PROOF. We will show (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (i) and (iii) $\Rightarrow$ (iv) $\Rightarrow$ (iii).

(i) $\Rightarrow$ (ii) Suppose f is expanding. Pick a Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}$, and let $\delta_0 > 0$ be as in (5.13) (note that $\#\mathrm{post}(f) \geq 3$ as we will later prove in Corollary 7.2). Suppose $\mathcal{U}$ is a cover of S^2 by open and connected sets that satisfies $\mathrm{mesh}(\mathcal{U}) < \delta_0$. If $U \in \mathcal{U}$, then U is connected and $\mathrm{diam}(U) < \delta_0$. So if V is an arbitrary connected component of $f^{-n}(U)$, then by Lemma 5.34 the set V is contained in an n-flower for $(f, \mathcal{C})$. Hence
$$\mathrm{diam}(V) \leq 2\,\mathrm{mesh}(f, n, \mathcal{C}),$$

which implies

$$\text{mesh}(f^{-n}(\mathcal{U})) \leq 2\,\text{mesh}(f, n, \mathcal{C}).$$

Since f is an expanding Thurston map, we have $\text{mesh}(f, n, \mathcal{C}) \to 0$, and so $\text{mesh}(f^{-n}(\mathcal{U})) \to 0$ as $n \to \infty$.

(ii) $\Rightarrow$ (iii) Obvious.

(iii) $\Rightarrow$ (i) Suppose $\mathcal{U}$ is an open cover of S^2 as in (iii). Pick a Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) \subset \mathcal{C}$, and let $\delta > 0$ be a *Lebesgue number* for the cover $\mathcal{U}$, i.e., every set $K \subset S^2$ with $\text{diam}(K) < \delta$ is contained in a set $U \in \mathcal{U}$. We can find a number $M \in \mathbb{N}$ such that each of the two 0-tiles for $(f, \mathcal{C})$ can be written as a union of M connected sets V with $\text{diam}(V) < \delta$. Then each such set V is contained in a set $U \in \mathcal{U}$.

Now if X is an arbitrary n-tile for $(f, \mathcal{C})$, then $Y = f^n(X)$ is a 0-tile for $(f, \mathcal{C})$ and $f^n|X$ is a homeomorphism of X onto Y. Hence X is a union of M connected sets of the form $(f^n|Y)^{-1}(V)$, where $V \subset Y$ is connected and lies in a set $U \in \mathcal{U}$. Then $(f^n|X)^{-1}(V)$ lies in a component of $f^{-n}(U)$, and so

$$\text{diam}((f^n|X)^{-1}(V)) \leq \text{mesh}(f^{-n}(\mathcal{U})).$$

This implies

$$\text{diam}(X) \leq M\,\text{mesh}(f^{-n}(\mathcal{U})).$$

Hence

$$\text{mesh}(f, n, \mathcal{C}) \leq M\,\text{mesh}(f^{-n}(\mathcal{U})).$$

Since $\text{mesh}(f^{-n}(\mathcal{U})) \to 0$, we also have $\text{mesh}(f, n, \mathcal{C}) \to 0$ as $n \to \infty$. It follows that f is expanding.

(iii) $\Rightarrow$ (iv) Suppose $\mathcal{U}$ is an open cover of S^2 as in (iii), and $\mathcal{V}$ is an arbitrary open cover of S^2. Let $\delta > 0$ be a Lebesgue number for the cover $\mathcal{V}$, i.e., every set $K \subset S^2$ with $\text{diam}(K) < \delta$ is contained in a set $V \in \mathcal{V}$. By (iii) we can find $N \in \mathbb{N}$ such that $\text{mesh}(f^{-n}(\mathcal{U})) < \delta$ for $n > N$. If $n > N$ and U' is a set in $f^{-n}(\mathcal{U})$, then $\text{diam}(U') < \delta$ by definition of $\text{mesh}(f^{-n}(\mathcal{U}))$. Hence there exists $V \in \mathcal{V}$ such that $U' \subset V$.

(iv) $\Rightarrow$ (iii) Suppose $\mathcal{U}$ is an open cover of S^2 as in (iv). Then $\mathcal{U}$ also satisfies condition (iii); indeed, let $\epsilon > 0$ be arbitrary, and let $\mathcal{V}$ be the open cover of S^2 consisting of all open balls of radius $\epsilon/2$. Then $\text{diam}(V) \leq \epsilon$ for all $V \in \mathcal{V}$. Moreover, by (iv) there exists $N \in \mathbb{N}$ such that for $n > N$ every set in $f^{-n}(\mathcal{U})$ is contained in a set in $\mathcal{V}$. In particular, $\text{mesh}(f^{-n}(\mathcal{U})) \leq \epsilon$ for $n > N$. This shows that $\mathcal{U}$ satisfies condition (iii). $\qquad\square$

6.2. Further results on expansion

In this section we collect various other useful results related to expansion.

LEMMA 6.4. *Let* $f\colon S^2 \to S^2$ *be a Thurston map,* $n \in \mathbb{N}$, *and* $F = f^n$. *Then* F *is a Thurston map with* $\operatorname{post}(F) = \operatorname{post}(f)$. *The map* f *is expanding if and only if* F *is expanding.*

PROOF. Since f is a Thurston map, the map F is a branched covering map on S^2 with $\operatorname{post}(F) = \operatorname{post}(f)$ (see Section 2.2) and $\deg(F) = \deg(f)^n \geq 2$. Hence F is also a Thurston map.

Fix a Jordan curve $\mathcal{C} \subset S^2$ with $\operatorname{post}(f) = \operatorname{post}(F) \subset \mathcal{C}$. It follows from the definitions that

$$\operatorname{mesh}(F, k, \mathcal{C}) = \operatorname{mesh}(f, nk, \mathcal{C})$$

for all $k \in \mathbb{N}_0$. If f is expanding, then by Lemma 6.1 we have

$$\lim_{k \to \infty} \operatorname{mesh}(f, k, \mathcal{C}) = 0$$

which implies that

$$\operatorname{mesh}(F, k, \mathcal{C}) = \operatorname{mesh}(f, nk, \mathcal{C}) \to 0$$

as $k \to \infty$. Hence F is expanding.

Conversely, suppose that F is expanding. Then we know that

$$(6.3) \qquad \lim_{k \to \infty} \operatorname{mesh}(F, k, \mathcal{C}) = \lim_{k \to \infty} \operatorname{mesh}(f, nk, \mathcal{C}) = 0.$$

Let the constant $M \geq 1$ be as in Lemma 5.37 for the map f and the Jordan curve $\mathcal{C}$. By an argument similar to the proof of Lemma 6.1 one can show that

$$\operatorname{mesh}(f, l + 1, \mathcal{C}) \leq 2M \operatorname{mesh}(f, l, \mathcal{C})$$

for all $l \in \mathbb{N}_0$. This implies

$$\operatorname{mesh}(f, l, \mathcal{C}) \leq (2M)^n \operatorname{mesh}(f, n\lfloor l/n \rfloor, \mathcal{C})$$

for all $l \in \mathbb{N}_0$ and so by (6.3) we have $\operatorname{mesh}(f, l, \mathcal{C}) \to 0$ as $l \to \infty$. This shows that f is expanding. $\qquad\qquad\square$

A map $f\colon S^2 \to S^2$ is called *eventually onto*, if for each non-empty open set $U \subset S^2$ there is an iterate f^n such that $f^n(U) = S^2$.

LEMMA 6.5. *Let* $f\colon S^2 \to S^2$ *be an expanding Thurston map. Then* f *is eventually onto.*

As we will see (Example 6.14), there are Thurston maps that are eventually onto, but not expanding.

PROOF. Let $f\colon S^2 \to S^2$ be an expanding Thurston map. Pick a Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}$ as in Definition 2.2. As before, we denote the black and white 0-tiles for $(f, \mathcal{C})$ by X_{b}^0 and X_{w}^0, respectively.

Let $U \subset S^2$ be an arbitrary non-empty open set, and $B(a, \epsilon)$ with $a \in U$ and $\epsilon > 0$ be an open ball contained in U. Since f is expanding, there is a number $n \in \mathbb{N}$ such that $\mathrm{mesh}(f, n, \mathcal{C}) < \epsilon/4$. Then each n-tile has diameter $< \epsilon/4$. Let X be an n-tile containing the center $a \in U$ of $B(a, \epsilon)$, and Y an n-tile that shares an n-edge with X. Then $X \cup Y \subset B(a, \epsilon) \subset U$, and so $f^n(U) \supset f^n(X \cup Y) = X_{\mathsf{w}}^0 \cup X_{\mathsf{b}}^0 = S^2$. The claim follows. $\square$

A metric d on a space S is called a *length metric* or *path metric* (see Section A.1) if for any two points $x, y \in S$ we have $d(x, y) = \inf_\gamma \mathrm{length}(\gamma)$, where the infimum is taken over all paths γ in S joining x to y. Using this concept, one can formulate a simple criterion when a Thurston map is expanding.

LEMMA 6.6. *Let d be a length metric on S^2 that induces the given topology on S^2, and let $f\colon S^2 \to S^2$ be a Thurston map. If f uniformly expands the d-length of paths, i.e., if there is a number $\rho > 1$ such that for every path γ in S^2 we have*

$$\mathrm{length}_d(f \circ \gamma) \geq \rho \, \mathrm{length}_d(\gamma),$$

then f is expanding.

PROOF. Let d be a length metric on S^2 such that the Thurston map $f\colon S^2 \to S^2$ expands the d-length of paths as in the statement of the lemma. In the following, all metric notions refer to this metric d. To prove that f is expanding, we will show that condition (iii) in Proposition 6.3 is satisfied for a suitable cover $\mathcal{U}$ of S^2.

We pick a Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}$, and consider cells for $(f, \mathcal{C})$. Then the corresponding 0-flowers $W^0(p)$, $p \in \mathrm{post}(f)$, form a cover of S^2 (see Lemma 5.29 (iv)). In order to obtain a cover $\mathcal{U}$ as in Proposition 6.3 (iii), we want to shrink each 0-flower $W^0(p)$ slightly to a new set U so that we have good control for the length of paths joining points in U to p inside $W^0(p)$. Note that since $W^0(p)$ is a region and d is a length metric, every point in $W^0(p)$ can be joined to p by a path in $W^0(p)$ of finite length, but in general there will be no uniform upper bound for the length of these paths.

In order to obtain such a bound, let $r > 0$ and define $W_r^0(p)$ for $p \in \mathrm{post}(f)$ to be the set of all points $u \in W^0(p)$ such that u and p can be joined by a path γ in $W^0(p)$ with $\mathrm{length}(\gamma) < r$.

Claim. There exists $r > 0$ such that each point in S^2 is contained in one of the sets $W_r^0(p)$, $p \in \mathrm{post}(f)$.

To prove this, let $u \in S^2$ be arbitrary. Since the 0-flowers cover S^2, there exists $p \in \mathrm{post}(f)$ such that $u \in W^0(p)$. Then we can find a path γ in $W^0(p)$ joining u and p with $r_u := \mathrm{length}(\gamma) < \infty$.

We can choose $\delta_u > 0$ such that $B_u := B(u, \delta_u) \subset W^0(p)$. Since d is a length metric, every point v in B_u can be joined with u by a path in B_u of length $< \delta_u$. If we concatenate such a path with γ, then we obtain a path in $W^0(p)$ that has length $< r_u + \delta_u$ and stays inside $W^0(p)$. In particular, we have uniform control for the length of such paths for all points in B_u. Since finitely many of the balls B_u, $u \in S^2$, cover S^2, the claim follows.

Now pick $r > 0$ as in the claim, and consider the open cover $\mathcal{U}$ of S^2 given by the sets $W_r^0(p)$, $p \in \mathrm{post}(f)$. Let $n \in \mathbb{N}_0$ and $p \in \mathrm{post}(f)$ be arbitrary, and consider a component V of $f^{-n}(W_r^0(p))$. Then V is contained in a component of $f^{-n}(W^0(p))$, and so there exists an n-flower $W^n(q)$ such that $V \subset W^n(q)$ (see Lemma 5.29 (iii)). Here $q \in S^2$ is an n-vertex. Then $f^n(q)$ is a 0-vertex contained in $W^0(p)$ which implies that $f^n(q) = p$.

Now let $v \in V$ be arbitrary, and $u := f^n(v) \in W_r^0(p)$. Then there exists a path γ in $W_r^0(p) \subset W^0(p)$ with $\mathrm{length}(\gamma) < r$ that joins u and p. By Lemma A.17 there exists a lift α of γ under f^n that starts at v. Then $f^n \circ \alpha = \gamma$ and $\alpha \subset V \subset W^n(q)$. One endpoint of α is v, while the other endpoint of α is a preimage of p under f^n and hence an n-vertex. Since q is the only n-vertex in $V \subset W^n(q)$, it follows that α joins v and q.

By using the fact that f expands the d-length paths by the factor ρ, we see that

$$\mathrm{length}(\alpha) < \frac{1}{\rho^n} \mathrm{length}(f^n \circ \alpha) = \frac{1}{\rho^n} \mathrm{length}(\gamma) < \frac{r}{\rho^n}.$$

So every point in V can be joined to q by a path of length $< r/\rho^n$. This implies that $\mathrm{diam}(V) < 2r/\rho^n$, and it follows that $\mathrm{mesh}(f^{-n}(\mathcal{U})) < 2r/\rho^n$. Since $\rho > 1$, we conclude that $\mathrm{mesh}(f^{-n}(\mathcal{U})) \to 0$ as $n \to \infty$. By Proposition 6.3, the map f is expanding. $\square$

A Thurston map $f\colon S^2 \to S^2$ is called a *Thurston polynomial* if there exists a point in S^2, denoted by ∞, that is completely invariant, i.e., $f^{-1}(\infty) = \{\infty\}$.

LEMMA 6.7. *No Thurston polynomial f is expanding.*

PROOF. Let f be a Thurston polynomial. We can choose a point $\infty \in S^2$ that is completely invariant. Then $d := \mathrm{deg}_f(\infty) = \mathrm{deg}(f) \geq 2$

by (2.2), and so ∞ is a critical point of f. Since ∞ is a fixed point as well, it follows that $\infty \in \operatorname{post}(f)$.

We pick a Jordan curve $\mathcal{C} \subset S^2$ with $\operatorname{post}(f) \subset \mathcal{C}$, and consider tiles for $(f, \mathcal{C})$. Each n-tile X^n is mapped by f^n homeomorphically to a 0-tile X^0 (see Proposition 5.17 (i)). Since

$$\infty \in \operatorname{post}(f) \subset \mathcal{C} = \partial X^0 \subset X^0,$$

it follows that X^n contains a preimage of ∞ by f^n. Since ∞ is completely invariant, the only such preimage is ∞ itself, and so $\infty \in X^n$. Therefore, each n-tile contains ∞.

We pick a point $p \in S^2$ distinct from ∞. Since for each $n \in \mathbb{N}$ the set of all n-tiles forms a cover of S^2, there exists an n-tile X^n containing p. Then

$$\operatorname{mesh}(f, n, \mathcal{C}) \geq \operatorname{diam}(X^n) \geq \operatorname{dist}(p, \infty) > 0,$$

and so $\operatorname{mesh}(f, n, \mathcal{C}) \not\to 0$ as $n \to \infty$. By (6.1) this means that f is not expanding. $\qquad\square$

Let $f\colon S^2 \to S^2$ a Thurston map. A *Levy cycle* for f is a multicurve $\Gamma = \{\gamma_1, \dots, \gamma_n\}$ (see Definition 2.16 (ii)) with the following property: for each $j = 1, \dots, n$ the set $f^{-1}(\gamma_{j+1})$ contains a component $\widetilde{\gamma}_j$ that is isotopic to γ_j rel. $\operatorname{post}(f)$ such that the map $f|\widetilde{\gamma}_j\colon \widetilde{\gamma}_j \to \gamma_{j+1}$ is a homeomorphism (here we set $\gamma_{n+1} = \gamma_1$).

Since $f|\widetilde{\gamma}_j\colon \widetilde{\gamma}_j \to \gamma_{j+1}$ is a covering map, the last condition is equivalent to the requirement that the degree of this map is equal to 1.

LEMMA 6.8. *Let $f\colon S^2 \to S^2$ be a Thurston map and suppose that γ and σ are Jordan curves in $S^2 \setminus \operatorname{post}(f)$ that are isotopic rel. $\operatorname{post}(f)$. Let $\gamma_1, \dots, \gamma_k$ with $k \in \mathbb{N}$ be the components of $f^{-1}(\gamma)$. Then $f^{-1}(\sigma)$ has also k components. Moreover, we can label them as $\sigma_1, \dots, \sigma_k$ such that for $j = 1, \dots, k$,*

  (i) *the curves γ_j and σ_j are isotopic rel. $\operatorname{post}(f)$,*

  (ii) *the degrees of $f|\gamma_j\colon \gamma_j \to \gamma$ and $f|\sigma_j\colon \sigma_j \to \sigma$ agree.*

PROOF. This follows immediately if we lift a suitable isotopy by f. Indeed, for every isotopy $H\colon S^2 \times [0,1] \to S^2$ rel. $\operatorname{post}(f)$ with $H_0 = H(\cdot, 0) = \operatorname{id}_{S^2}$ there is an isotopy $\widetilde{H}\colon S^2 \times [0,1] \to S^2$ rel. $\operatorname{post}(f)$ with $\widetilde{H}_0 = \operatorname{id}_{S^2}$ such that

(6.4) $$(H_t \circ f)(p) = (f \circ \widetilde{H}_t)(p)$$

for all $p \in S^2$, $t \in [0,1]$ (see Proposition 11.3).

Let H be an isotopy rel. $\operatorname{post}(f)$ that deforms γ to σ rel. $\operatorname{post}(f)$, i.e., $H_0 = \operatorname{id}_{S^2}$ and $H_1(\gamma) = \sigma$, and let $\widetilde{H}$ be the lift of H by f as above. Define $\sigma_j := \widetilde{H}_1(\gamma_j)$. Then by definition γ_j and σ_j are isotopic

rel. $\mathrm{post}(f)$ for $j = 1, \ldots, n$. Moreover, (6.4) implies that $\sigma_1, \ldots, \sigma_k$ are the components of $f^{-1}(\sigma)$ and that the degrees of $f|\gamma_j \colon \gamma_j \to \gamma$ and $f|\sigma_j \colon \sigma_j \to \sigma$ agree. $\square$

Now suppose that f is a Thurston map that has a Levy cycle $\Gamma = \{\gamma_1, \ldots, \gamma_n\}$. We consider the iterate $F = f^n$ and denote $\gamma = \gamma_1$ for convenience. If we use the previous lemma repeatedly, then we see that there is a component $\widetilde{\gamma}$ of $F^{-1}(\gamma)$ that is isotopic to γ rel. $\mathrm{post}(f)$ such that the degree of $F|\widetilde{\gamma} \colon \widetilde{\gamma} \to \gamma$ is 1. We remark the existence of such an iterate $F = f^n$ and such a (non-peripheral) Jordan curve $\gamma \subset S^2 \setminus \mathrm{post}(f)$ is in fact equivalent to the existence of a Levy cycle, but we will not prove this here.

By using a lifting argument as in the proof of the previous Lemma (based on Proposition 11.3) one can easily show Levy cycles persist under Thurston equivalence; so if the maps $f \colon S^2 \to S^2$ and $g \colon S^2 \to S^2$ are Thurston equivalent, then f has a Levy cycle if and only if g has a Levy cycle.

If a Levy cycle Γ is an *invariant* multicurve, then it is clearly a Thurston obstruction, since the spectral radius of the corresponding Thurston matrix $A(f, \Gamma)$ is ≥ 1. If the Levy cycle Γ is not invariant, then it is not hard to show that there is an invariant multicurve $\Gamma' \supset \Gamma$ (see [**Ta92**, Lemma 2.2]). Then the spectral radius of $A(f, \Gamma')$ is ≥ 1, and so every Levy cycle is contained in a Thurston obstruction. Hence according to Thurston's theorem (Theorem 2.17) if a Thurston map has a Levy cycle (and a hyperbolic orbifold), then it cannot be equivalent to a rational map.

Our next lemma shows that a Levy cycle is also an obstruction for a Thurston map to be expanding.

LEMMA 6.9. *Let $f \colon S^2 \to S^2$ be a Thurston map that has a Levy cycle. Then f is not expanding.*

Of course, Levy cycles are not the only obstructions for expansion of a Thurston map. For example, by Lemma 6.7 no holomorphic Thurston polynomial is expanding, but, as follows from Thurston's theorem, such a polynomial cannot have Levy cycles either.

PROOF. Assume f has a Levy cycle. Then there exists an iterate $F = f^n$ and a non-peripheral Jordan curve $\gamma_1 \subset S^2 \setminus \mathrm{post}(f)$ such that a component γ_2 of $F^{-1}(\gamma_1)$ is isotopic to γ_1 rel. $\mathrm{post}(f)$ and the degree of $F|\gamma_1 \colon \gamma_2 \to \gamma_1$ is equal to 1. From Lemma 6.8 it follows by induction that there is a sequence $\{\gamma_k\}_{k \in \mathbb{N}}$ of Jordan curves in $S^2 \setminus \mathrm{post}(f)$ all of which are isotopic to γ_1 rel. $\mathrm{post}(f)$ such that γ_{k+1} is a component of $F^{-1}(\gamma_k)$ and the degree of $F|\gamma_{k+1} \colon \gamma_{k+1} \to \gamma_k$ is equal to 1 for

all $k \in \mathbb{N}$. Hence γ_k is a component of $F^{-k}(\gamma_1)$ and the degree of $F^k|\gamma_{k+1}\colon \gamma_{k+1} \to \gamma_1$ is equal to 1. Moreover, each curve γ_k is non-peripheral.

We fix a Jordan curve $\mathcal{C} \subset S^2$ with $\operatorname{post}(f) = \operatorname{post}(F) \subset \mathcal{C}$, and consider tiles and flowers for $(F, \mathcal{C})$. Let δ_0 be as in (5.13) for the map F. Then we can decompose γ_1 into finitely many arcs $\alpha_1, \ldots, \alpha_l$ such that $\operatorname{diam}(\alpha_j) < \delta_0$ for $j = 1, \ldots, l$. By Lemma 5.34 (ii) every component of $F^{-k}(\alpha_j)$ is contained in a k-flower W_j^k for $j = 1, \ldots, l$. It follows that for every $k \in \mathbb{N}$ the curve γ_k is contained in the union of l k-flowers $W_1^k, \ldots, W_l^k$.

To reach a contradiction, assume now that f, and hence also F, is expanding. Then we have

$$\operatorname{diam}(\gamma_k) \leq \sum_{j=1}^{l} \operatorname{diam}(W_j^k) \leq 2l \max_{X \in \mathbf{X}^k} \operatorname{diam}(X) \to 0$$

as $k \to \infty$.

On the other hand, we can find a finite open cover $\mathcal{U}$ of S^2 consisting of simply connected regions U each of which contains at most one postcritical point of f (for example, the 0-flowers form such a cover). By what we have just seen, we can find $k \in \mathbb{N}$ such that $\operatorname{diam}(\gamma_k)$ is smaller than the Lebesgue number of $\mathcal{U}$. Then γ_k is contained in a set $U \in \mathcal{U}$. Since U is simply connected and contains at most one postcritical point of f, the curve γ_k is peripheral. This is a contradiction showing that f is not expanding. $\qquad\square$

EXAMPLE 6.10. We now present an example of a Thurston map f with a Levy cycle. The map f will have no periodic critical points, and so it provides an example of a Thurston map without periodic critical points that is not expanding, contrasting Proposition 2.3. Since Levy cycles persist under Thurston equivalence, f is also not equivalent to any expanding Thurston map. The same is actually true of the map from Example 6.14 below, but the map f considered here will have a hyperbolic orbifold.

For the construction we start with a topological sphere S^2 that is a pillow (see Section A.10). Similar to Section 1.1, the pillow is obtained by gluing two unit squares together along their boundaries. As before, we color one side (i.e., one square) of the pillow white, and the other black.

The black side is divided horizontally into two rectangles, one of which is colored white and the other colored black. The white side of the pillow is subdivided into four quadrilaterals, two white and two black ones as indicated on the left in Figure 6.1. Here we have cut the
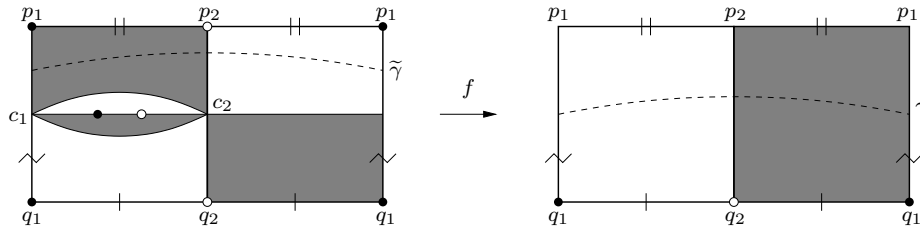
FIGURE 6.1. A map with a Levy cycle.

pillow along three sides so that we obtain a rectangle as shown in the picture. The symbols in the picture indicate which sides have to be identified to recover the pillow.

The map f is now constructed by mapping each white quadrilateral homeomorphically to the white face, and each black quadrilateral to the black side of the pillow. Here f maps vertices to vertices. In Figure 6.1 we have marked two vertices of each quadrilateral (on the left), as well as two vertices of the pillow (on the right) by a black or white dot to indicate the correspondence of vertices under f. Finally, we require that f agrees on sides shared by two vertices. The map f thus defined is indeed a Thurston map (see Chapter 12 and in particular Proposition 12.3). The postcritical points correspond to the vertices of the pillow. The map f has two critical points c_1 and c_2 and the following ramification portrait:

$$ c_1 \xrightarrow{3:1} p_1 \longrightarrow q_1 \circlearrowright \qquad c_2 \xrightarrow{3:1} p_2 \longrightarrow q_2. \circlearrowright $$

Thus f has no periodic critical points, its signature is $(3,3,3,3)$, and f has a hyperbolic orbifold.

We consider the Jordan curve $\gamma \subset S^2 \setminus \mathrm{post}(f)$ as indicated on the right of Figure 6.1. One of the components $\widetilde{\gamma}$ of $f^{-1}(\gamma)$ is shown on the left (the other components of $f^{-1}(\gamma)$ are not shown). Clearly, $\widetilde{\gamma}$ is isotopic rel. $\mathrm{post}(f)$ to γ. Furthermore, the degree of $f\colon \widetilde{\gamma} \to \gamma$ is equal to 1, and so $\Gamma = \{\gamma\}$ is a Levy cycle.

Lemma 6.9 implies that f is not expanding, and, by our earlier discussion, no Thurston map equivalent to f is expanding.

6.3. Lattès-type maps and expansion

We know (see Theorem 3.1 (i) and (i')) that every Lattès map is expanding. This is not always true for Lattès-type maps, but it is easy to decide when this is the case. The relevant condition is based on the following definition.

Let $L\colon \mathbb{R}^2 \to \mathbb{R}^2$ be a linear map. We call it *expanding* if $|\lambda| > 1$ for each of the two (possibly complex) eigenvalues λ of L.

PROPOSITION 6.11. *Let $f\colon S^2 \to S^2$ be a Lattès-type map and $L = L_A$ be the linear part of an affine map A as in Definition 3.3. Then f is expanding (as a Thurston map) if and only if L is expanding (as a linear map).*

For the proof we require two lemmas.

LEMMA 6.12. *Suppose $L\colon \mathbb{R}^2 \to \mathbb{R}^2$ is an expanding linear map. Then there exist constants $\rho > 1$ and $n_0 \in \mathbb{N}$ such that*

$$(6.5) \qquad\qquad |L^n(v)| \geq \rho^n |v|$$

for all $v \in \mathbb{R}^2$ and all $n \in \mathbb{N}$ with $n \geq n_0$.

Here $|v| = \sqrt{x^2 + y^2}$ denotes the usual Euclidean norm of $v = (x, y) \in \mathbb{R}^2$.

PROOF. The eigenvalues of L are the two (possibly identical) roots $\lambda_1, \lambda_2 \in \mathbb{C}$ of the characteristic polynomial $P(\lambda) = \det(L - \lambda \, \mathrm{id}_{\mathbb{R}^2})$ of L. We may assume $|\lambda_1| \leq |\lambda_2|$. Since L is expanding we have $|\lambda_1| > 1$. The polynomial P has real coefficients, and so $\lambda_2 = \overline{\lambda_1}$ if λ_1 is not real.

There exists a basis of $\mathbb{R}^2$ consisting of two linearly independent vectors $v_1, v_2 \in \mathbb{R}^2$ such that the linear map L has a matrix representation with respect to this basis of one of the following forms:

$$M_1 = |\lambda_1| \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \text{where } \theta \in \mathbb{R},$$

$$M_2 = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}, \text{ or } M_3 = \begin{pmatrix} \lambda_1 & 1 \\ 0 & \lambda_1 \end{pmatrix}.$$

We can find an inner product on $\mathbb{R}^2$ such that v_1 and v_2 form an orthonormal basis of $\mathbb{R}^2$ with respect to this inner product. If $\| \cdot \|$ is the norm induced by this product, then

$$\|av_1 + bv_2\| = \sqrt{a^2 + b^2}$$

for $a, b \in \mathbb{R}$. If L has a matrix representation given by M_1 or M_2, it is clear that $\|L(v)\| \geq |\lambda_1| \cdot \|v\|$ and so

$$(6.6) \qquad\qquad \|L^n(v)\| \geq |\lambda_1|^n \cdot \|v\|$$

for all $v \in \mathbb{R}^2$ and $n \in \mathbb{N}$. If L is represented by the matrix M_3, then a similar estimate is harder to obtain, but one can show that

$$\|L^n(v)\| \geq \frac{|\lambda_1|^{n+1}}{\sqrt{2\lambda_1^2 + n^2}} \|v\|.$$

To see this, one bounds the operator norm of the matrix M_3^{-n} by its Hilbert-Schmidt norm. We leave the details to the reader.

Since all norms on $\mathbb{R}^2$ are comparable, it follows that

$$|L^n(v)| \geq \rho^n |v|$$

for all sufficiently large n independent of $v \in \mathbb{R}^2$ with $\rho = |\lambda_1|^{1/2} > 1$. $\qquad\square$

Recall that if G is a planar crystallographic group, then we say that a continuous map $\Theta\colon \mathbb{R}^2 \to S^2$ is induced by G if $\Theta(u) = \Theta(v)$ for $u, v \in \mathbb{R}^2$ if and only if there exists $g \in G$ such that $v = g(u)$ (see Section A.7). In this case, G is necessarily of non-torus type (see Theorem 3.7) and Θ is essentially the quotient map $\Theta\colon \mathbb{R}^2 \to \mathbb{R}^2/G \cong S^2$ (see Corollary A.21). Under a suitable identification $S^2 \cong \widehat{\mathbb{C}}$ by a homeomorphism, the map Θ is equal to the universal orbifold covering map of a parabolic orbifold $(\widehat{\mathbb{C}}, \alpha)$, whose signature corresponds to the type of G (see Theorem 3.10).

LEMMA 6.13. *Let G be a planar crystallographic group and $\Theta\colon \mathbb{R}^2 \to S^2$ be a continuous map induced by G. Suppose $K_n \subset \mathbb{R}^2$ is a connected set for $n \in \mathbb{N}$. Then*

$$\lim_{n\to\infty} \operatorname{diam}(K_n) = 0 \text{ if and only if } \lim_{n\to\infty} \operatorname{diam}(\Theta(K_n)) = 0.$$

Here $\operatorname{diam}(K_n)$ is the Euclidean diameter of K_n, and $\operatorname{diam}(\Theta(K_n))$ the diameter of $\Theta(K_n)$ with respect to some base metric d on S^2.

PROOF. "$\Rightarrow$" For this implication it is enough to show that Θ is uniformly continuous on $\mathbb{R}^2$. This follows from the fact that Θ is induced by G and that G acts isometrically and cocompactly on $\mathbb{R}^2$; indeed, we can find a compact fundamental domain $F \subset \mathbb{R}^2$ for the action of G on $\mathbb{R}^2$. Now suppose $x, y \in \mathbb{R}^2$ and $\delta := |x - y|$ is small. Then there exists $g \in G$ such that $g(x) \in F$. If δ is small enough, then $g(x), g(y) \in U$, where U is a compact neighborhood of F. Since Θ is uniformly continuous on U, and $|g(x) - g(y)| = |x - y| = \delta$, it follows that

$$d(\Theta(x), \Theta(y)) = d(\Theta(g(x)), \Theta(g(y)))$$

is small only depending on δ. The uniform continuity of Θ follows.

"$\Leftarrow$" We argue by contradiction and assume that the statement is false. Then there exist connected sets $K_n \subset \mathbb{R}^2$ such $\operatorname{diam}(\Theta(K_n)) \to 0$ as $n \to \infty$, but $\operatorname{diam}(K_n) \geq \epsilon_0$ for $n \in \mathbb{N}$, where $\epsilon_0 > 0$.

We pick a point $x_n \in K_n$ for $n \in \mathbb{N}$. If we replace each set K_n with its image $K'_n = g_n(K_n)$ for suitable $g_n \in G$ (note that $\operatorname{diam}(K'_n) = \operatorname{diam}(K_n)$ and $\Theta(K'_n) = \Theta(K_n)$), and pass to a subsequence if necessary, then we may assume that the sequence $\{x_n\}$ converges, say $x_n \to x \in \mathbb{R}^2$ as $n \to \infty$.

Let $p := \Theta(x)$. Then the set $\Theta^{-1}(p)$ is equal to the orbit Gx of x under G. Since the action of G on $\mathbb{R}^2$ is properly discontinuous, the set $\Theta^{-1}(p) = Gx$ has no limit point in $\mathbb{R}^2$. Since G also acts cocompactly on $\mathbb{R}^2$, this implies that the distance of distinct points in $\Theta^{-1}(p)$ is bounded away from 0; so there exists a constant $m > 0$ such that $|u - v| \geq m$ whenever $u, v \in \Theta^{-1}(p)$ and $u \neq v$.

Pick a constant c with $0 < c < \min\{\epsilon_0/2, m\}$. The set K_n is connected, and has diameter $\mathrm{diam}(K_n) \geq \epsilon_0 > 2c$. Hence K_n cannot be contained in the disk $\{z \in \mathbb{R}^2 : |z - x_n| < c\}$, and so it meets the circle $\{z \in \mathbb{R}^2 : |z - x_n| = c\}$. It follows that there exists a point $y_n \in K_n$ with $|x_n - y_n| = c$. By passing to another subsequence if necessary, we may assume that the sequence $\{y_n\}$ converges, say $y_n \to y \in \mathbb{R}^2$ as $n \to \infty$. Then $|x - y| = c < m$. Note that $\Theta(x_n), \Theta(y_n) \in \Theta(K_n)$ for $n \in \mathbb{N}$, and $\mathrm{diam}(\Theta(K_n)) \to 0$ as $n \to \infty$. So

$$p = \Theta(x) = \lim_{n \to \infty} \Theta(x_n) = \lim_{n \to \infty} \Theta(y_n) = \Theta(y),$$

and $x, y \in \Theta^{-1}(p)$. Since $|x - y| = c > 0$, we have $x \neq y$. So x and y are two distinct points in $\Theta^{-1}(p)$ with $|x - y| = c < m$. This contradicts the choice of m, and the claim follows. $\qquad\square$

PROOF OF PROPOSITION 6.11. In the proof metric notions on $\mathbb{R}^2$ will refer to the Euclidean metric.

Let $f\colon S^2 \to S^2$ be the given Lattès-type map, and A, Θ, G be as in Definition 3.3. Then $\Theta\colon \mathbb{R}^2 \to S^2$ is a branched covering map induced by the crystallographic group G. We know that here G is not isomorphic to $\mathbb{Z}^2$, because the quotient space $\mathbb{R}^2/G \cong S^2$ is not a torus. So by Proposition 3.9 there exists a holomorphic branched covering map $\widetilde{\Theta}\colon \mathbb{R}^2 \cong \mathbb{C} \to \widehat{\mathbb{C}}$ induced by G. Since both Θ and $\widetilde{\Theta}$ are induced by G, these maps essentially correspond to the quotient map $\mathbb{R}^2 \to \mathbb{R}^2/G$ and so we have natural identifications $S^2 \cong \mathbb{R}^2/G \cong \widehat{\mathbb{C}}$. More precisely, Corollary A.21 implies that there exists a unique homeomorphism $\varphi\colon S^2 \to \widehat{\mathbb{C}}$ such that $\widetilde{\Theta} = \varphi \circ \Theta$.

We now conjugate f by φ to obtain a Thurston map $\widetilde{f} := \varphi \circ f \circ \varphi^{-1}$ defined on $\widehat{\mathbb{C}}$. Then $\widetilde{f}$ is a Lattès-type map with the triple A, $\widetilde{\Theta}$, G as in Definition 3.3. Note that the affine map A has not changed here and that f is expanding if and only if $\widetilde{f}$ is expanding. In other words, in order to prove the statement, we can make the additional assumptions that the Lattès-type map f is defined on $\widehat{\mathbb{C}}$ and the map $\Theta\colon \mathbb{C} \to \widehat{\mathbb{C}}$ is holomorphic.

Then $\Theta\colon \mathbb{C} \to \widehat{\mathbb{C}}$ is the universal orbifold covering map of the parabolic orbifold $\mathcal{O}_f = (\widehat{\mathbb{C}}, \alpha_f)$ of f (see Proposition 3.6, Remark 3.17, and Theorem 3.10).

Let ω be the canonical orbifold metric of $\mathcal{O}_f$. Since $\mathcal{O}_f$ is parabolic, ω is essentially the push-forward of the Euclidean metric on $\mathbb{R}^2$ by Θ (see Section A.10 and Section 2.5). The metric ω is a length metric that induces the standard topology on $\widehat{\mathbb{C}}$ (so it can be used as a base metric on $\widehat{\mathbb{C}}$ as in Section 6.1 or Lemma 6.13) and the map $\Theta\colon \mathbb{R}^2 \to \widehat{\mathbb{C}}$ is a path isometry in the sense that

$$\mathrm{length}(\alpha) = \mathrm{length}_\omega(\Theta \circ \alpha)$$

for each path α in $\mathbb{R}^2$.

If $L = L_A$ is the linear part of A, then the map L^n is the linear part of A^n for each $n \in \mathbb{Z}$. So L^n and A^n only differ by a translation. Since translations are isometries, it follows that

$$(6.7) \qquad\qquad \mathrm{length}(L^n \circ \alpha) = \mathrm{length}(A^n \circ \alpha)$$

for all $n \in \mathbb{Z}$ whenever α is a path in $\mathbb{R}^2$. If $\gamma := \Theta \circ \alpha$, then

$$f^n \circ \gamma = f^n \circ \Theta \circ \alpha = \Theta \circ A^n \circ \alpha$$

for $n \in \mathbb{N}$. Since Θ is a path isometry, we conclude that

$$\mathrm{length}_\omega(f^n \circ \gamma) = \mathrm{length}(A^n \circ \alpha) = \mathrm{length}(L^n \circ \alpha)$$

for $n \in \mathbb{N}$.

Now suppose that L is expanding. Then Lemma 6.12 implies that there exist $N \in \mathbb{N}$ and a constant $\rho > 1$ such that

$$\mathrm{length}(L^N \circ \alpha) \geq \rho\, \mathrm{length}(\alpha)$$

for all paths α in $\mathbb{R}^2$. If γ is an arbitrary path in $\widehat{\mathbb{C}}$, then it has a lift under the branched covering map Θ (see Lemma A.17); so it can be written in the form $\gamma = \Theta \circ \alpha$, where α is a path in $\mathbb{R}^2$. Hence

$$\mathrm{length}_\omega(f^N \circ \gamma) = \mathrm{length}(L^N \circ \alpha) \geq \rho\, \mathrm{length}(\alpha)$$
$$= \rho\, \mathrm{length}_\omega(\Theta \circ \alpha) = \rho\, \mathrm{length}_\omega(\gamma).$$

Lemma 6.6 implies that f^N is an expanding Thurston map. Hence f is expanding (see Lemma 6.4).

To prove the converse, we assume that f is expanding, but L is not. Then one of the eigenvalues of L has absolute value ≤ 1. Since the product of these eigenvalues is equal to $\det(L) = \deg(f) \geq 2$ (see Lemma 3.16), it follows from the considerations in the beginning of the proof of Lemma 6.12 that both eigenvalues of L are real. So L has a real eigenvalue λ with $|\lambda| \leq 1$. Note that $\lambda \neq 0$, because L is invertible.

Then there exists $u \in \mathbb{R}^2$ with $|u| = 1$ such that $L(u) = \lambda u$. Let α be the parametrized line segment joining 0 and u, and define $\alpha_n := A^{-n} \circ \alpha$ for $n \in \mathbb{N}$. Then

$$(6.8) \qquad \operatorname{diam}(\alpha_n) = \operatorname{diam}(A^{-n} \circ \alpha) = \operatorname{length}(A^{-n} \circ \alpha)$$

$$= \operatorname{length}(L^{-n} \circ \alpha) = \frac{1}{|\lambda|^n} \operatorname{length}(\alpha)$$

$$\geq \operatorname{length}(\alpha) = 1$$

for all $n \in \mathbb{N}$.

Let $\gamma := \Theta \circ \alpha$, and $\gamma_n := \Theta \circ \alpha_n$ for $n \in \mathbb{N}$. Then γ_n is a lift of γ under f^n, because

$$f^n \circ \gamma_n = f^n \circ \Theta \circ A^{-n} \circ \alpha = \Theta \circ A^n \circ A^{-n} \circ \alpha = \Theta \circ \alpha = \gamma.$$

Since f is expanding and the paths γ_n are lifts of some path γ under f^n, we have $\operatorname{diam}_\omega(\gamma_n) \to 0$ as $n \to \infty$. Indeed, if γ is a path whose diameter is less than the Lebesgue number $l_0 > 0$ of an open cover $\mathcal{U}$ as in Proposition 6.3 (iii), then there exists $U \in \mathcal{U}$ such that $\gamma \subset U$. Then γ_n lies in a connected component of $f^{-n}(U)$ and so

$$\operatorname{diam}_\omega(\gamma_n) \leq \operatorname{mesh}(f^{-n}(\mathcal{U})) \to 0$$

as $n \to \infty$. The statement $\operatorname{diam}_\omega(\gamma_n) \to 0$ as $n \to \infty$ remains true for arbitrary paths γ, because we can break γ up into finitely many paths of diameter $< l_0$. (We will later see that with respect to a *visual metric* for f (see Chapter 8) the diameters of lifts of any path under f^n actually shrink to 0 exponentially fast as $n \to \infty$ (see Lemma 8.9)).

Now we obtain a contradiction from Lemma 6.13, because the sets α_n are connected and

$$\operatorname{diam}_\omega(\gamma_n) = \operatorname{diam}_\omega(\Theta(\alpha_n)) \to 0$$

as $n \to \infty$, but $\operatorname{diam}(\alpha_n) \geq 1$ for all $n \in \mathbb{N}$ by (6.8).

It follows that L is expanding if f is. Together with the first part of the proof, we conclude that f is expanding as a Thurston map if and only if $L = L_A$ is expanding as a linear map. $\square$

We finish this chapter by giving an example of a Thurston map that is eventually onto, but not expanding. The example is due to K. Pilgrim.

EXAMPLE 6.14. Let G be the crystallographic group consisting of all maps g of the form

$$u \in \mathbb{R}^2 \mapsto g(u) = \pm u + \gamma,$$

where $\gamma \in \mathbb{Z}^2$. So G is of type (2222) (see Theorem 3.7). Let $\Theta \colon \mathbb{R}^2 \to \mathbb{R}^2/G \cong S^2$ be the quotient map.

We consider the matrix

$$A = \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix}.$$

and the map $A\colon \mathbb{R}^2 \to \mathbb{R}^2$, $u \in \mathbb{R}^2 \mapsto Au$ given by left-multiplication of $u \in \mathbb{R}^2$ (written as a column vector) by the matrix A. For simplicity, here (and also below) we not not distinguish in our notation between a matrix and the linear map it induces on $\mathbb{R}^2$ by left-multiplication.

The map A has the form (3.25). The linear part $L = L_A$ of A agrees with A. So L is also represented by the matrix A. Since $\det(L_A) = \det(A) \geq 2$, the map A induces a Lattès-type map $f\colon S^2 \to S^2$ on the quotient $S^2 = \mathbb{R}^2/G$ (see Proposition 3.21). We claim that f *is not expanding*, but *eventually onto*.

Recall that the latter property means that for any non-empty open set $U \subset S^2$ there is an iterate f^n such that $f^n(U) = S^2$ (see also Lemma 6.5).

The map $L = A$ has the eigenvalues $\lambda_1 = 3 - \sqrt{5}$ and $\lambda_2 = 3 + \sqrt{5}$. Since $|\lambda_1| < 1$, the map f is not expanding by Proposition 6.11.

Now consider the linear maps B and C given by left-multiplication of $u \in \mathbb{R}^2$ with the matrices

$$B = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \text{ and } C = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix},$$

respectively. Then $A = B \circ C = C \circ B$. The maps B and C again have the form (3.25) and so descend to maps $g\colon S^2 \to S^2$ and $h\colon S^2 \to S^2$ respectively. Note that $\det(B) = 1$, and so g is a homeomorphism with an inverse induced by B^{-1} (see Proposition 3.21).

These maps satisfy $f = g \circ h = h \circ g$. Indeed, note that $f \circ \Theta = \Theta \circ A$, $g \circ \Theta = \Theta \circ B$, and $h \circ \Theta = \Theta \circ C$. Thus

$$f \circ \Theta = \Theta \circ A = \Theta \circ B \circ C = g \circ \Theta \circ C = g \circ h \circ \Theta.$$

Since Θ is surjective, it follows that $f = g \circ h$. A similar argument shows that $f = h \circ g$. Since $f = g \circ h = h \circ g$, we have $f^n = g^n \circ h^n$ for all $n \in \mathbb{N}$.

Now let $U \subset S^2$ be an arbitrary non-empty open set. Then $V := \Theta^{-1}(U)$ is a non-empty open set in $\mathbb{R}^2$. Since $C^n(V) = 2^n V$ this set $C^n(V)$ will contain arbitrarily large disks if n is sufficiently large. In particular, there exists $n \in \mathbb{N}$ such that $C^n(V)$ contains a translate $\gamma + R$ with $\gamma \in \mathbb{Z}^2$ of the rectangle $R = [0,1] \times [0,1/2]$, which is a fundamental domain (see Section A.7) for the action of G. For such n we have

$$S^2 = \Theta(R) = \Theta(\gamma + R) = \Theta(C^n(V)) = h^n(\Theta(V)) = h^n(U),$$

and so, since g is a homeomorphism,
$$f^n(U) = (g^n \circ h^n)(U) = g^n(S^2) = S^2.$$
This shows that f is eventually onto.

CHAPTER 7

Thurston maps with two or three postcritical points

In this chapter we investigate Thurston maps $f\colon S^2 \to S^2$ with a postcritical set consisting of two or three elements (note that we always have $\#\operatorname{post}(f) \geq 2$ by Remark 5.16). The considerations here are not essential for our main story and may safely be skipped by the impatient reader.

For starters it is easy to classify all Thurston maps with two postcritical points up to Thurston equivalence.

PROPOSITION 7.1. *A Thurston map $f\colon S^2 \to S^2$ with $\#\operatorname{post}(f) = 2$ is Thurston equivalent to a power map $z \mapsto z^n$ on $\widehat{\mathbb{C}}$, where $n \in \mathbb{Z} \setminus \{-1, 0, 1\}$.*

This will be proved in Section 7.1. We record the following immediate consequence.

COROLLARY 7.2. *If $f\colon S^2 \to S^2$ is an expanding Thurston map, then $\#\operatorname{post}(f) \geq 3$.*

PROOF. By Remark 5.16 we know that $\#\operatorname{post}(f) \geq 2$. When $\#\operatorname{post}(f) = 2$ it follows from Proposition 7.1 that the second iterate f^2 is a Thurston polynomial. Then Lemmas 6.7 and 6.4 imply that f is not expanding. $\square$

According to this corollary, we can restrict ourselves to the case $\#\operatorname{post}(f) \geq 3$ when we consider expanding Thurston maps.

In case $\#\operatorname{post}(f) = 3$, the relation to rational Thurston maps is clarified by the following statement.

THEOREM 7.3. *Let $f\colon S^2 \to S^2$ be a Thurston map such that $\#\operatorname{post}(f) = 3$. Then the following statements are true:*

(i) *f is Thurston equivalent to a rational Thurston map;*

(ii) *if the map f is expanding, then f is topologically conjugate to a rational Thurston map if and only if f has no periodic critical points.*

Part (i) of this statement is essentially a trivial case of Thurston's characterization of rational maps given in Theorem 2.17. We will present a proof in Section 7.1. Part (ii) easily follows if this is combined with some of our other results.

In Chapter 3 we considered Lattès and Lattès-type maps. These are Thurston maps $f\colon S^2 \to S^2$ with a parabolic orbifold and no periodic critical points. The case when f has a parabolic orbifold, but also periodic critical points, is very special. Then the signature of f must be (∞, ∞) or $(2, 2, \infty)$ as follows from Propositions 2.13 and 2.9. These maps can easily be classified up to Thurston equivalence.

THEOREM 7.4. *Let* $f\colon S^2 \to S^2$ *be a Thurston map. Then* f *has signature*

(i) (∞, ∞) *if and only if* f *is Thurston equivalent to a power map* $z \mapsto z^n$ *on* $\widehat{\mathbb{C}}$, *where* $n \in \mathbb{Z} \setminus \{-1, 0, 1\}$;

(ii) $(2, 2, \infty)$ *if and only if* f *is Thurston equivalent to* χ *or* $-\chi$, *where* $\chi = \chi_n$ *is a* Chebyshev polynomial *of degree* $n \in \mathbb{N} \setminus \{1\}$.

The case of signature (∞, ∞) is essentially already covered by Proposition 7.1. The proof of Theorem 7.4 is given in Section 7.2, where we will also review the definition of Chebyshev polynomials (see also [**Ri90**]).

7.1. Thurston maps with $\# \operatorname{post}(f) \in \{2, 3\}$ are equivalent to rational maps

We begin by looking at Thurston maps f with $\# \operatorname{post}(f) = 3$.

PROOF OF THEOREM 7.3. (i) Let $f\colon S^2 \to S^2$ be a Thurston with three postcritical points, which we denote by p_0, p_1, p_∞. Let $h_0\colon S^2 \to \widehat{\mathbb{C}}$ be an orientation-preserving homeomorphism. By postcomposing with a suitable Möbius transformation, we may assume that $h_0(p_0) = 0$, $h_0(p_1) = 1$, and $h_0(p_\infty) = \infty$. We can think of the map h_0 as a global chart on S^2; by this chart S^2 carries a conformal structure.

If we pull-back this conformal structure by the map f, then we obtain another conformal structure on S^2. The map f is then holomorphic with respect to these two conformal structures on S^2. To be more precise, Corollary A.12 gives the existence of an orientation-preserving homeomorphism $h_1\colon S^2 \to \widehat{\mathbb{C}}$ and a rational map $R\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ such that $h_0 \circ f = R \circ h_1$. On a more intuitive level, R is the representation of the holomorphic map f if we use suitable global charts.

Again by postcomposing with a suitable Möbius transformation, we may assume that $h_1(p_0) = 0$, $h_1(p_1) = 1$, and $h_1(p_\infty) = \infty$.

Two orientation-preserving homeomorphisms on S^2 that agree on a set $P \subset S^2$ containing at most three points are isotopic rel. P (see Lemma 11.11). Thus h_0 and h_1 are isotopic rel. $\operatorname{post}(f) = \{p_0, p_1, p_\infty\}$. It now follows from Lemma 2.5 that R is a Thurston map, and it is clear that f and R are Thurston equivalent.

(ii) Now suppose in addition that f is expanding and has no periodic critical points. Since the latter condition is invariant under Thurston equivalence, the rational Thurston map R constructed above will then not have periodic critical points either, and is hence expanding by Proposition 2.3. Therefore, the maps f and R are topologically conjugate by a general result that will be proved later (see Theorem 11.1).

Conversely, if f is expanding and topologically conjugate to a rational map R, then R is an expanding Thurston map. Hence R has no periodic critical points by Proposition 2.3, which implies that f cannot have periodic critical points either. $\square$

To deal with the case $\# \operatorname{post}(f) = 2$, we first prove an auxiliary statement.

LEMMA 7.5. *Let $f \colon S^2 \to S^2$ be a Thurston map with $\# \operatorname{post}(f) = 2$. Then*

(i) *the signature of f is (∞, ∞);*

(ii) *the set $\operatorname{post}(f)$ is completely invariant, i.e.,*

$$f^{-1}(\operatorname{post}(f)) = \operatorname{post}(f).$$

PROOF. Let $\alpha_f \colon S^2 \to \widehat{\mathbb{N}}$ be the ramification function and $\mathcal{O}_f = (S^2, \alpha_f)$ be the orbifold associated with f.

(i) If this were not true, then the Euler characteristic of $\mathcal{O}_f$ would be strictly positive, contradicting Proposition 2.12.

(ii) Let $p \in f^{-1}(\operatorname{post}(f))$. Then (i) and condition (iii) in Proposition 2.13 imply that $\alpha_f(p) = \infty$. Thus $p \in \operatorname{post}(f)$, showing that $f^{-1}(\operatorname{post}(f)) \subset \operatorname{post}(f)$. The reverse inclusion $f^{-1}(\operatorname{post}(f)) \supset \operatorname{post}(f)$ is true for every Thurston map, and so the claim follows. $\square$

Let us now prove Proposition 7.1.

PROOF OF PROPOSITION 7.1. Let $f \colon S^2 \to S^2$ be a Thurston map with $\# \operatorname{post}(f) = 2$. We want to show that f is Thurston equivalent to the map $z \mapsto z^n$ on $\widehat{\mathbb{C}}$, where $n \in \mathbb{Z} \setminus \{-1, 0, 1\}$.

The proof that f is Thurston equivalent to a rational map $R \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is identical to the proof of Theorem 7.3 (i). The only adjustment is that the maps $h_0 \colon S^2 \to \widehat{\mathbb{C}}$ and $h_1 \colon S^2 \to \widehat{\mathbb{C}}$ have to agree on the set

$\text{post}(f)$, which now contains just two points, instead of three. Again this is achieved by postcomposing with a suitable Möbius transformation.

The rational map R has also two postcritical points, and so we may assume that $\text{post}(R) = \{0, \infty\}$ (by conjugating R with a suitable Möbius transformation). Let $A = R^{-1}(0)$ be the set of all zeros and $B = R^{-1}(\infty)$ be the set of poles of R. From Lemma 7.5 it follows that $A \cup B = \text{post}(R) = \{0, \infty\}$. This implies that $R(z) = cz^n$ for $z \in \widehat{\mathbb{C}}$, where $c \in \mathbb{C} \setminus \{0\}$ and $n \in \mathbb{Z} \setminus \{0\}$. Here actually $n \in \mathbb{Z} \setminus \{-1, 0, 1\}$, because R is a Thurston map and hence not a homeomorphism. So in particular, $n \neq 1$, which implies that by conjugating R with an auxiliary map of the form $z \mapsto \alpha z$, $\alpha \in \mathbb{C} \setminus \{0\}$, if necessary, we may assume that $c = 1$. The claim follows. $\qquad\square$

7.2. Thurston maps with signature (∞, ∞) or $(2, 2, \infty)$

In this section we consider Thurston maps f with a parabolic orbifold and periodic critical points. Equivalently, the associated orbifold $\mathcal{O}_f$ has signature (∞, ∞) or $(2, 2, \infty)$. These maps together with Lattès and Lattés-type maps considered in Chapter 3 cover all cases of Thurston maps with a parabolic orbifold (see Proposition 3.6).

The case when the signature is (∞, ∞) has already been treated in Proposition 7.1 (see also Lemma 7.5). It remains to consider the case of signature $(2, 2, \infty)$. Our presentation follows [**Mi06b**].

LEMMA 7.6. *Let* $f\colon S^2 \to S^2$ *be a Thurston map with signature* $(2, 2, \infty)$. *Then* f *is a Thurston polynomial.*

PROOF. Suppose the signature of f is $(2, 2, \infty)$ and let $p \in S^2$ be the unique point with $\alpha_f(p) = \infty$. Lemma 2.13 (iii) implies that $\alpha_f(q) = \infty$ for each point $q \in f^{-1}(p)$. Thus, $q = p$ and so p is completely invariant. Therefore, f is a Thurston polynomial. $\qquad\square$

We call a Thurston polynomial f with a parabolic orbifold a *parabolic Thurston polynomial*. Then f has signature (∞, ∞) or $(2, 2, \infty)$. Conversely, if f is a Thurston map with signature $(2, 2, \infty)$, then f is a parabolic Thurston polynomial by Lemma 7.6. If f has signature (∞, ∞), then f or f^2 is a parabolic Thurston polynomial as follows from Lemma 7.5 (ii). A classification of parabolic Thurston polynomials is obtained from Theorem 7.4 which we will prove below.

Lemma 6.7 implies a parabolic Thurston polynomial f cannot be expanding. Moreover, if f is rational and suitably normalized, then f is a polynomial.

The polynomials that appear here are very special, namely power maps $z \mapsto z^n$ or Chebyshev polynomials. By definition the *Chebyshev polynomial* χ_n for $n \in \mathbb{N}_0$ is the unique polynomial such that

$$(7.1) \qquad \cos(nu) = \chi_n(\cos u), \quad u \in \mathbb{C}.$$

Thus, the first Chebyshev polynomials are $\chi_0(z) = 1$, $\chi_1(z) = z$, $\chi_2(z) = 2z^2 - 1$, and $\chi_3(z) = 4z^3 - 3z$. They satisfy the recurrence relation $\chi_{n+1}(z) = 2z\chi_n(z) - \chi_{n-1}(z)$ for $n \in \mathbb{N}$. This implies that $\deg(\chi_n) = n$. Since

$$\chi_n(-\cos u) = \chi_n(\cos(u + \pi)) = \cos(nu + n\pi)$$
$$= (-1)^n \cos(nu) = (-1)^n \chi_n(\cos u),$$

it follows that χ_n is an even function when n is even, and χ_n is an odd function when n is odd.

To find the critical points of χ_n, we differentiate (7.1) and obtain

$$-n \sin(nu) = -\chi_n'(\cos u) \sin u, \quad u \in \mathbb{C}.$$

For $n \geq 1$ the left hand side has a simple zero whenever $nu \in \mathbb{Z}\pi$. In particular, if we consider $u_k = k\pi/n$ and the corresponding points $z_k = \cos(k\pi/n)$ for $k = 0, \ldots, n$, we see that χ_n' has a simple zero at z_k for $k = 1, \ldots, n-1$. Moreover, $\chi_n'(z_0) = \chi_n'(1) \neq 0$ and $\chi_n'(z_n) = \chi_n'(-1) \neq 0$, because $\sin(u_0) = \sin(u_n) = 0$. Since χ_n has at most $n-1$ critical points and $u \mapsto \cos(u)$ is injective on $[0, \pi]$, the points z_k for $k = 1, \ldots, n-1$ are the distinct critical points of χ_n and $\deg(\chi_n, z_k) = 2$. By (7.1) we have $\chi_n(z_k) = (-1)^k$ for $k = 1, \ldots, n-1$, $\chi_n(-1) = \chi_n(1) = 1$ for n even, and $\chi_n(\pm 1) = \pm 1$ for n odd. This implies that $\chi = \chi_n$ for $n \geq 2$ is a postcritically-finite polynomial with $\mathrm{post}(\chi) = \{-1, 1, \infty\}$. For the ramification function α_χ of χ we conclude from this analysis that $\alpha_\chi(-1) = \alpha_\chi(1) = 2$ and $\alpha_\chi(\infty) = \infty$.

We will also have to consider the polynomial $\chi = -\chi_n$ for $n \geq 2$. The same considerations show that again $\mathrm{post}(\chi) = \{-1, 1, \infty\}$, $\alpha_\chi(-1) = \alpha_\chi(1) = 2$, and $\alpha_\chi(\infty) = \infty$. It follows that $\chi = \pm\chi_n$ for $n \geq 2$ is a postcritically-finite polynomial whose orbifold has signature $(2, 2, \infty)$.

The postcritical points -1 and 1 are mapped by $\chi = \pm\chi_n$ as indicated in the following diagrams: if $\chi = \chi_n$,

$$(7.2) \qquad -1 \longrightarrow 1 \circlearrowleft \quad \text{for } n \text{ even and} \quad -1 \circlearrowleft \quad 1 \circlearrowleft \quad \text{for } n \text{ odd};$$

and if $\chi = -\chi_n$,

$$(7.3) \qquad 1 \longrightarrow -1 \circlearrowleft \quad \text{for } n \text{ even and} \quad -1 \rightleftarrows 1 \quad \text{for } n \text{ odd}.$$

The diagrams when n is even are similar in both cases, because the maps involved are topologically conjugate. Indeed, suppose $n \in \mathbb{N}$ is even, and let $\tau(z) = -z$ for $z \in \mathbb{C}$. Since χ_n is an even function, it follows that

$$(\tau \circ \chi_n \circ \tau^{-1})(z) = -\chi_n(-z) = -\chi_n(z)$$

for $z \in \mathbb{C}$, which implies that $-\chi_n = \tau \circ \chi_n \circ \tau^{-1}$ for $n \in \mathbb{N}$ even.

When $n \in \mathbb{N}$ is odd, the maps χ_n and $-\chi_n$ are not topologically conjugate (or Thurston equivalent) as the above diagrams show: all postcritical points are fixed points for χ_n, but not for $-\chi_n$.

After this preliminary discussion, we can now prove Theorem 7.4.

PROOF OF THEOREM 7.4. Statement (i) immediately follows from Proposition 7.1, Lemma 7.5, and Proposition 2.14.

To prove (ii), let $f \colon S^2 \to S^2$ be a Thurston map with signature $(2, 2, \infty)$. Then $\# \mathrm{post}(f) = 3$, and so by Theorem 7.3 (i) the map f is Thurston equivalent to a rational map, necessarily with the same signature. So we may assume that f is a rational map on $S^2 = \widehat{\mathbb{C}}$ to begin with. Moreover, by conjugating the map with a suitable Möbius transformation, we may assume that $\mathrm{post}(f) = \{-1, 1, \infty\}$ and that for the ramification function $\alpha_f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{N}}$ of f we have $\alpha_f(-1) = \alpha_f(1) = 2$ and $\alpha_f(\infty) = \infty$. By the argument in the proof of Lemma 7.6 we see that then f is a polynomial.

The map f has a parabolic orbifold, and so Proposition 2.13 implies that

(7.4) $$\alpha_f(z) \cdot \deg_f(z) = \alpha_f(f(z))$$

for all $z \in \widehat{\mathbb{C}}$. Equation (7.4) shows that we have $z \in f^{-1}(\{-1, 1\})$ if and only if $z \neq \infty$ and one of the factors on the left-hand side of (7.4) is different from 1. Then one factor is equal to 1 and the other equal to 2. We conclude that

$$f^{-1}(\{-1, 1\}) = \{-1, 1\} \cup \mathrm{crit}(f) \setminus \{\infty\}$$

with $\deg_f(-1) = \deg_f(1) = 1$ and $\deg_f(z) = 2$ for $z \in \mathrm{crit}(f) \setminus \{\infty\}$. This implies that the polynomials $1 - f(z)^2$ and $(1 - z^2)f'(z)^2$ have the same zeros of exactly the same order. If we define $n = \deg(f) \geq 2$, then comparison of the highest-order coefficient gives

(7.5) $$n^2(1 - f(z)^2) = (1 - z^2)f'(z)^2$$

for $z \in \mathbb{C}$. It is well known and easy to prove that then $f = \pm \chi_n$.

Indeed, to see this, consider the even entire function g defined as $g(u) = f(\cos u)$ for $u \in \mathbb{C}$. Then (7.5) leads to

$$g'(u)^2 = n^2(1 - g(u)^2)$$

for $u \in \mathbb{C}$. If we differentiate this equation, then we obtain the linear ordinary differential equation

$$g''(u) + n^2 g(u) = 0, \quad u \in \mathbb{C}.$$

It has the general solution $g(u) = c_1 \cos(nu) + c_2 \sin(nu)$, $c_1, c_2 \in \mathbb{C}$. Since g is even, we must have $c_2 = 0$; moreover, $c_1 = g(0) = f(1) \in \{-1, 1\}$. Hence $g(u) = \pm \cos(nu) = f(\cos u)$ for $u \in \mathbb{C}$. This implies $f = \pm \chi_n$.

For the converse direction suppose that the Thurston map $f \colon S^2 \to S^2$ is Thurston equivalent to $\chi = \pm \chi_n$ with $n \in \mathbb{N} \setminus \{1\}$. We have seen earlier in this section that χ has signature $(2, 2, \infty)$. Hence f has the same signature by Proposition 2.14. $\qquad\square$

As we have seen in Chapter 3, Lattès maps are related to crystallographic groups G acting on $\mathbb{C}$. Here G contains a subgroup G_{tr} of translations isomorphic to a rank-2 lattice. We will now discuss how the maps $z \mapsto z^n$ for $n \in \mathbb{Z} \setminus \{-1, 0, 1\}$ and $z \mapsto \pm \chi_n(z)$ for $n \in \mathbb{N} \setminus \{1\}$ can be described in a similar fashion. According to Theorem 7.4, every Thurston map with signature (∞, ∞) or $(2, 2, \infty)$ is Thurston equivalent to such a map. Since the following considerations are fairly elementary, we will skip some details.

Recall from Section 3.1 that $\mathrm{Isom}(\mathbb{C})$ denotes the group of all orientation-preserving isometries of $\mathbb{C}$ (equipped with the Euclidean metric) and $\mathrm{Aut}(\mathbb{C})$ the group of holomorphic automorphisms of $\mathbb{C}$. For a group $G \subset \mathrm{Isom}(\mathbb{C})$, we denote by G_{tr} the subgroup of G consisting of translations in G, i.e., G_{tr} consists of all maps $g \in G$ of the form $z \mapsto g(z) = z + \gamma$ with a $\gamma \in \mathbb{C}$. If we denote by $\Gamma \subset \mathbb{C}$ the set of all such $\gamma \in \mathbb{C}$, then $G_{\mathrm{tr}} = \{z \mapsto z + \gamma : \gamma \in \Gamma\}$. If the action of G on $\mathbb{C}$ is properly discontinuous, then Γ is necessarily a discrete set in $\mathbb{C}$ and so a lattice.

We now focus on the case that Γ is a rank-1 lattice, meaning that Γ spans a 1-dimensional subspace of $\mathbb{R}^2 \cong \mathbb{C}$. The following lemma is closely related to Theorem 3.7.

LEMMA 7.7. *Let $G \subset \mathrm{Isom}(\mathbb{C})$ be a group such that the action of G on $\mathbb{C}$ is properly discontinuous, and $\Gamma \subset \mathbb{C}$ as defined above is of rank 1. Then G is conjugate to one of the following groups $\widetilde{G}$ consisting of all $g \in \mathrm{Isom}(\mathbb{C})$ of the form*

$(\infty\infty)$ $z \mapsto g(z) = z + k$, *where $k \in \mathbb{Z}$;*

(22∞) $\qquad$ $z \mapsto g(z) = \pm z + k,$ *where* $k \in \mathbb{Z}.$

As in Chapter 3, we say that G and $\widetilde{G}$ are *conjugate* if there exists $h \in \mathrm{Aut}(\mathbb{C})$ such that $\widetilde{G} = h \circ G \circ h^{-1}$. We do not provide the (well-known) proof of Lemma 7.7 here; it can be found in [**Ar91**].

If for a group G as in Lemma 7.7 the conjugate group $\widetilde{G}$ has the form $(\infty\infty)$ or (22∞), then we say that G is of type $(\infty\infty)$ or type (22∞), respectively. Again we are using Conway's orbifold notation. Clearly a group of type $(\infty\infty)$ is isomorphic (as a group) to $\mathbb{Z}$, and a group of type (22∞) is isomorphic to the *infinite dihedral group* $D_\infty = \mathbb{Z} \rtimes \mathbb{Z}_2$.

If $G = \widetilde{G}$ is one of the groups in Lemma 7.7, then its subgroup of translations G_{tr} consists of the maps of the form $z \in \mathbb{C} \mapsto g(z) = z + k$, where $k \in \mathbb{Z}$. The quotient $\mathbb{C}/G_{\mathrm{tr}}$ is an *(infinite) cylinder* (see below for a geometric justification of this terminology).

Clearly, $\exp(2\pi i z) = \exp(2\pi i w)$ for $z, w \in \mathbb{C}$ if and only if $w = z + k$ for some $k \in \mathbb{Z}$. This means that the map $\pi \colon \mathbb{C} \to \mathbb{C}^* := \mathbb{C} \setminus \{0\}$ given by $\pi(z) = \exp(2\pi i z)$ for $z \in \mathbb{C}$ is induced by G_{tr}; so we can identify the cylinder $\mathbb{C}/G_{\mathrm{tr}}$ with $\mathbb{C}^*$ and consider $\pi \colon \mathbb{C} \to \mathbb{C}^* \cong \mathbb{C}/G_{\mathrm{tr}}$ as the quotient map (see Lemma A.19). The cylinder $\mathbb{C}^*$ plays a similar role for the groups in Lemma 7.7 as tori obtained as quotients $\mathbb{C}/G_{\mathrm{tr}}$ of crystallographic groups G.

Now let $f(z) = z^n$ with $n \in \mathbb{Z} \setminus \{-1, 0, 1\}$. We know that the orbifold $(\widehat{\mathbb{C}}, \alpha_f)$ of f has signature (∞, ∞) and two *punctures* (i.e., points p with $\alpha_f(p) = \infty$) at 0 and ∞. So if we remove these punctures from $\widehat{\mathbb{C}}$ we obtain the set $\widehat{\mathbb{C}}_0 = \widehat{\mathbb{C}} \setminus \{0\} = \mathbb{C}^*$ (see Section A.9 for a related discussion). The (holomorphic) universal orbifold covering map $\Theta \colon \mathbb{C} \to \widehat{\mathbb{C}}_0 = \mathbb{C}^*$ is given by $\Theta(z) = \exp(2\pi i z)$ for $z \in \mathbb{C}$. It is induced by the group $G = \widetilde{G}$ of the form $(\infty\infty)$ in Lemma 7.7. The map $A \colon \mathbb{C} \to \mathbb{C}$ given by $A(z) = nz$ for $z \in \mathbb{C}$ is G-equivariant. If we define $\overline{A}(z) = z^n$ for $z \in \mathbb{C}^*$ and $\overline{\Theta} = \mathrm{id}_{\mathbb{C}^*}$, then we obtain the following commutative diagram:

(7.6)

$$
\begin{array}{ccc}
\mathbb{C} & \xrightarrow{A(z)=nz} & \mathbb{C} \\
{\scriptstyle \pi}\downarrow & & \downarrow{\scriptstyle \pi} \\
\mathbb{C}^* & \xrightarrow{\overline{A}} & \mathbb{C}^* \\
{\scriptstyle \overline{\Theta}}\downarrow & & \downarrow{\scriptstyle \overline{\Theta}} \\
\mathbb{C}^* & \xrightarrow{z \mapsto z^n} & \mathbb{C}^*.
\end{array}
$$

with $\Theta(z)=\exp(2\pi i z)$ on the left and $\Theta(z)=\exp(2\pi i z)$ on the right.

It is the analog of the diagram in (3.12) that we obtained for Lattès maps based on Theorem 3.1.

It is elementary to check that a map $A \in \mathrm{Aut}(\mathbb{C})$ is G-equivariant (see (A.27)) if and only if it is of the form $A(z) = nz + \beta$ with $n \in \mathbb{Z} \backslash \{0\}$ and $\beta \in \mathbb{C}$; so the map A in (7.6) corresponds to the case $\beta = 0$. It is easy to see that for general $\beta \in \mathbb{C}$ the map obtained as a quotient of A on $\widehat{\mathbb{C}}_0 = \mathbb{C}^*$ as in (7.6) is conjugate to $f(z) = z^n$ (for $n \in \mathbb{Z} \backslash \{-1, 0, 1\}$).

The canonical orbifold metric ω of $f(z) = z^n$ (see Section A.10 and in particular the discussion before Lemma A.31) is the conformal metric on $\mathbb{C}^*$ with length element $2\pi |dz|/|z|$ (see Section A.1 for the terminology). Equipped with this metric, $\widehat{\mathbb{C}}_0 = \mathbb{C}^* \cong \mathbb{C}/G$ is isometric to an infinite cylinder.

We can describe the maps $\chi = \pm\chi_n$ with $n \in \mathbb{N} \setminus \{1\}$ obtained from Chebyshev polynomials in a similar vein. For these maps we have $\alpha_f(-1) = \alpha(1) = 2$ and $\alpha_f(\infty) = \infty$. So the orbifold $(\widehat{\mathbb{C}}, \alpha_f)$ of f has a puncture at ∞ and so $\widehat{\mathbb{C}}_0 = \mathbb{C}$. The universal orbifold covering map $\Theta \colon \mathbb{C} \to \widehat{\mathbb{C}}_0 = \mathbb{C}$ is given by $\Theta(z) = \cos(2\pi z)$ for $z \in \mathbb{C}$. To see this, note that $\Theta \colon \mathbb{C} \to \mathbb{C}$ is a branched covering map with the critical values -1 and 1 and that $\deg(\Theta, z) = 2$ whenever $z \in \Theta^{-1}(\{-1, 1\})$.

Clearly, for $z, w \in \mathbb{C}$ we have

(7.7) $$\cos(2\pi z) = \cos(2\pi w) \text{ if and only if}$$
$$w = \pm z + k \text{ for some } k \in \mathbb{Z}.$$

So Θ is induced by the group G of isometries of the form $z \mapsto \pm z + k$ with $k \in \mathbb{Z}$, i.e., $G = \widetilde{G}$, where $\widetilde{G}$ as in case (22∞) of Lemma 7.7. As before, we identify $\mathbb{C}/G_{\mathrm{tr}}$ with $\mathbb{C}^*$ and consider $\pi(z) = \exp(2\pi i z)$ for $z \in \mathbb{C}$ as the quotient map $\pi \colon \mathbb{C} \to \mathbb{C}^*$. Let $\overline{\Theta}(z) = \frac{1}{2}(z + 1/z)$ for $z \in \mathbb{C}^*$. Then $\Theta = \overline{\Theta} \circ \pi$.

A map $A \in \mathrm{Aut}(\mathbb{C})$ is G-equivariant if and only if it is of the form $A(z) = nz + l/2$ where $n \in \mathbb{Z} \setminus \{0\}$ and $l \in \mathbb{Z}$ (we omit the elementary proof for this fact). For $n \in \mathbb{N} \setminus \{1\}$ and $l \in \{0, 1\}$ the map A descends to the map $\chi = \pm\chi_n$ on $\widehat{\mathbb{C}}_0 = \mathbb{C}$ under the map $\Theta \colon \mathbb{C} \to \mathbb{C}$. Actually, if we define $\overline{A}(z) = (-1)^l z^n$ for $z \in \mathbb{C}^*$, then we have the commutative diagram

(7.8)

$$
\begin{array}{ccc}
\mathbb{C} & \xrightarrow{A} & \mathbb{C} \\
\pi \downarrow & & \downarrow \pi \\
\mathbb{C}^* & \xrightarrow{\overline{A}} & \mathbb{C}^* \\
\overline{\Theta} \downarrow & & \downarrow \overline{\Theta} \\
\mathbb{C} & \xrightarrow{\chi} & \mathbb{C}.
\end{array}
$$

$\Theta(z) = \cos(2\pi z)$ on the left, $\Theta(z) = \cos(2\pi z)$ on the right.

This is again analogous to (3.12) obtained for Lattès maps.

Based on the considerations in Section A.10 it is not hard to see that the canonical orbifold metric ω of χ is given by the length element

$$\frac{|dz|}{2\pi|z-1|^{1/2}|z+1|^{1/2}}$$

on $\mathbb{C}$. We will describe a more geometric picture for $(\mathbb{C}, \omega)$ that will lead to models for the maps $\pm\chi_n$ similar to the models for Lattès maps as described in Section 1.1 and Section 3.5.

Our universal orbifold covering map $z \mapsto \Theta(z) = \cos(2\pi z)$ is induced by the group $G = \widetilde{G}$ as in case (22∞) of Lemma 7.7. So we can identify the quotient space $\mathbb{C}/G$ with the target $\mathbb{C}$ of $\Theta\colon \mathbb{C} \to \mathbb{C}$ and consider $\Theta\colon \mathbb{C} \to \mathbb{C} \cong \mathbb{C}/G$ as the quotient map. If we denote by $[u] \in \mathbb{C}/G$ the equivalence class of a point $u \in \mathbb{C}$ under the equivalence relation on $\mathbb{C}$ induced by G, then this identification is more explicitly given by the well-defined map $[z] \in \mathbb{C}/G \mapsto \Theta(z) \in \mathbb{C}$.

The canonical orbifold metric ω is essentially the push-forward of the Euclidean metric under the quotient map. Under our identification $\mathbb{C} \cong \mathbb{C}/G$ we have

(7.9) $$\omega([x], [y]) := \inf\{|z - w| : z \in [x], w \in [y]\}$$

for $[x], [y] \in \mathbb{C}/G$ (see Section A.10 and in particular (A.38)).

The metric space $(\mathbb{C}/G, \omega)$ is isometric to the space Δ obtained by gluing two copies of the half-infinite strip $S = [0, 1/2] \times [0, \infty) \subset \mathbb{R}^2 \simeq \mathbb{C}$ together along their boundaries. Here the strips carry the Euclidean metric and Δ the induced path metric.

To see this, note that the infinite strip $F = [0, 1/2] \times \mathbb{R} \subset \mathbb{R}^2 \cong \mathbb{C}$ is a fundamental domain of G (see Section A.7). Under the action of G two points in the boundary of F are identified if they are complex conjugates of each other. So the quotient $\mathbb{C}/G$ is obtained by folding the strip F along the real axis and gluing together corresponding boundary parts of the half-infinite strips $S_{\mathtt{w}} := [0, 1/2] \times [0, \infty)$ and $S_{\mathtt{b}} := [0, 1/2] \times (-\infty, 0]$. The metric ω is a path metric and corresponds to the Euclidean metric on $S_{\mathtt{w}}$ and $S_{\mathtt{b}}$. So $(\mathbb{C}/G, \omega)$ and Δ are indeed isometric.

In the following we will identify these spaces. We will also consider the half-infinite strips $S_{\mathtt{w}}$ and $S_{\mathtt{b}}$ as subsets and sides of Δ. We color $S_{\mathtt{w}}$ white, and $S_{\mathtt{b}}$ black. Then Δ is a locally Euclidean surface with two conical singularities as indicated on the right in Figure 7.1; the conical singularities are labeled by -1 and 1, because they correspond to these points under the identification $\mathbb{C} \cong \mathbb{C}/G$. Indeed, $[0] \cong \Theta(0) = 1$ and $[-1/2] \cong \Theta(-1/2) = -1$.
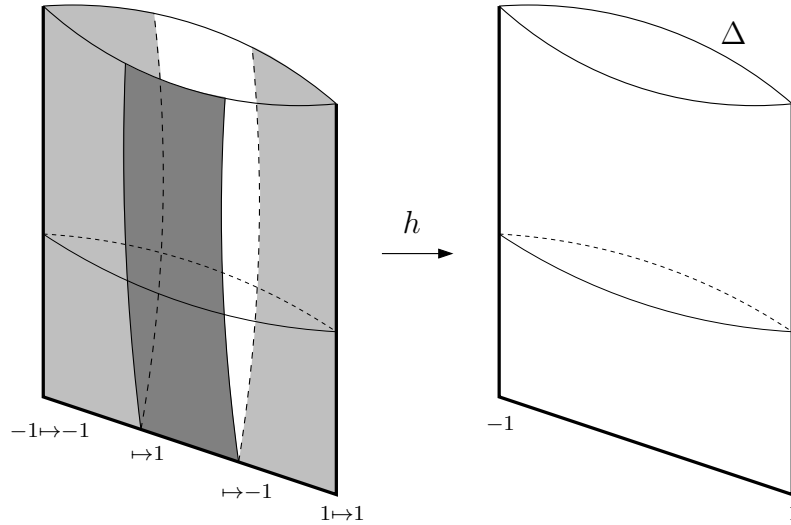
FIGURE 7.1. Model for a Chebyshev polynomial.

We now fix $n \in \mathbb{N}$ and divide each side of Δ into n strips S' of equal size. Then each strip S' is isometric to $[0, \frac{1}{2n}] \times [0, \infty)$, and hence similar to S by the scaling factor n. We color the strips S' in a checkerboard fashion black and white so that strips sharing an edge have different colors.

One can now define a map $h \colon \Delta \to \Delta$ as follows. We map each white strip S' to $S_{\mathtt{w}}$ and each black strip S' to $S_{\mathtt{b}}$ by an orientation-preserving Euclidean similarity. Then h is well-defined, because the definitions for h match on the edges where two strips intersect. An example of such a map is indicated in Figure 7.1.

These maps h give a Euclidean model for the maps $\pm\chi_n$ due to the following fact.

PROPOSITION 7.8. *Let $n \in \mathbb{N}$. Then every map $h \colon \Delta \to \Delta$ obtained from the above construction is topologically conjugate to χ_n or $-\chi_n$. Conversely, every polynomial χ_n and $-\chi_n$ is topologically conjugate to such a map h.*

PROOF. Fix $n \in \mathbb{N}$. Then the half-strips S' of the form

$$[\tfrac{k}{2n}, \tfrac{k+1}{2n}] \times [0, \infty) \subset S_{\mathtt{w}} \text{ and } [\tfrac{k}{2n}, \tfrac{k+1}{2n}] \times (-\infty, 0] \subset S_{\mathtt{b}}$$

for $k = 0, \dots, n-1$ divide the sides $S_{\mathtt{w}}$ and $S_{\mathtt{b}}$ of Δ, respectively. The checkerboard coloring of these strips S' is uniquely determined if we specify the coloring of $S'_0 := [\tfrac{0}{2n}, \tfrac{1}{2n}] \times [0, \infty)$.

If S'_0 is colored white, then the map $A(z) = nz$ passes to the quotient $\Delta = \mathbb{C}/G$ and sends white strips S' to $S_{\mathtt{w}}$ and black strips S' to $S_{\mathtt{b}}$ by a Euclidean similarity. In other words, A induces the map

$h\colon \Delta \to \Delta$ discussed above. On the other hand, by (7.8) this map A passes to the quotient χ_n under the map $z \mapsto \Theta(z) = \cos(2\pi z)$. This implies that the induced homeomorphism $\widetilde{\Theta} : \Delta = \mathbb{C}/G \to \mathbb{C}$ defined as $\widetilde{\Theta}([z]) = \Theta(z)$ for $[z] \in \mathbb{C}/G$ gives the conjugacy $h = \widetilde{\Theta}^{-1} \circ \chi_n \circ \widetilde{\Theta}$.

If S_0' is colored black, then $A(z) = nz + 1/2$ passes to the quotient $\Delta = \mathbb{C}/G$ and a strip S' is sent to a strip $S_{\mathtt{w}}$ or $S_{\mathtt{b}}$ of the same color by a Euclidean similarity. So again A induces the map h and by (7.8) we get a conjugacy $h = \widetilde{\Theta}^{-1} \circ \chi \circ \widetilde{\Theta}$, where $\chi = -\chi_n$. $\qquad\square$

CHAPTER 8

Visual Metrics

In this chapter we construct a natural class of metrics for an expanding Thurston map that we call *visual metrics*. We haven chosen this name, because there is a close relation between these metrics and visual metrics on the boundary at infinity of a Gromov hyperbolic space. Indeed, for an expanding Thurston map $f\colon S^2 \to S^2$ one can define a Gromov hyperbolic *tile graph* whose boundary at infinity can naturally be identified with S^2. By this identification, a metric ϱ on S^2 is visual in the sense of Gromov hyperbolic spaces if and only if it is visual as it will be defined in this chapter (see Chapter 10 and in particular Theorem 10.2). In general, a visual metric ϱ is not a length metric on S^2. In Chapter 16 we will investigate the resulting metric space (S^2, ϱ) in more detail.

We first give a quick overview of the definition and the basic properties o l metrics. In Sections 8.1 and 8.2 we will then provide the technical details. In the final section (Section 8.3) we consider rational expanding Thurston maps. In particular, we show that the canonical orbifold metric (see Section A.10 and Section 2.5) for such a map f is a visual metric precisely if f is a Lattès map (see Proposition 8.5).

Let $f\colon S^2 \to S^2$ be an expanding Thurston map, and $\mathcal{C} \subset S^2$ be a Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}$. We consider the cell decompositions of S^2 for $(f, \mathcal{C})$ as defined in Section 5.3. One can think of the set of n-tiles as a discrete approximation of the sphere S^2 and measure distances of points by a quantity related to combinatorics of n-tiles.

Indeed, let $x, y \in S^2$ be two distinct points, and X and Y be n-tiles with $x \in X$, $y \in Y$. Since f is expanding, X and Y must be disjoint if n is sufficiently large (see (6.1) and Lemma 6.1). This leads to the following definition.

DEFINITION 8.1. Let $f\colon S^2 \to S^2$ be an expanding Thurston map, $\mathcal{C} \subset S^2$ be a Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}$, and $x, y \in S^2$. For $x \neq y$ we define

$$m_{f,\mathcal{C}}(x,y) := \max\{n \in \mathbb{N}_0 : \text{there exist non-disjoint } n\text{-tiles}$$
$$X \text{ and } Y \text{ for } (f,\mathcal{C}) \text{ with } x \in X, y \in Y\}.$$
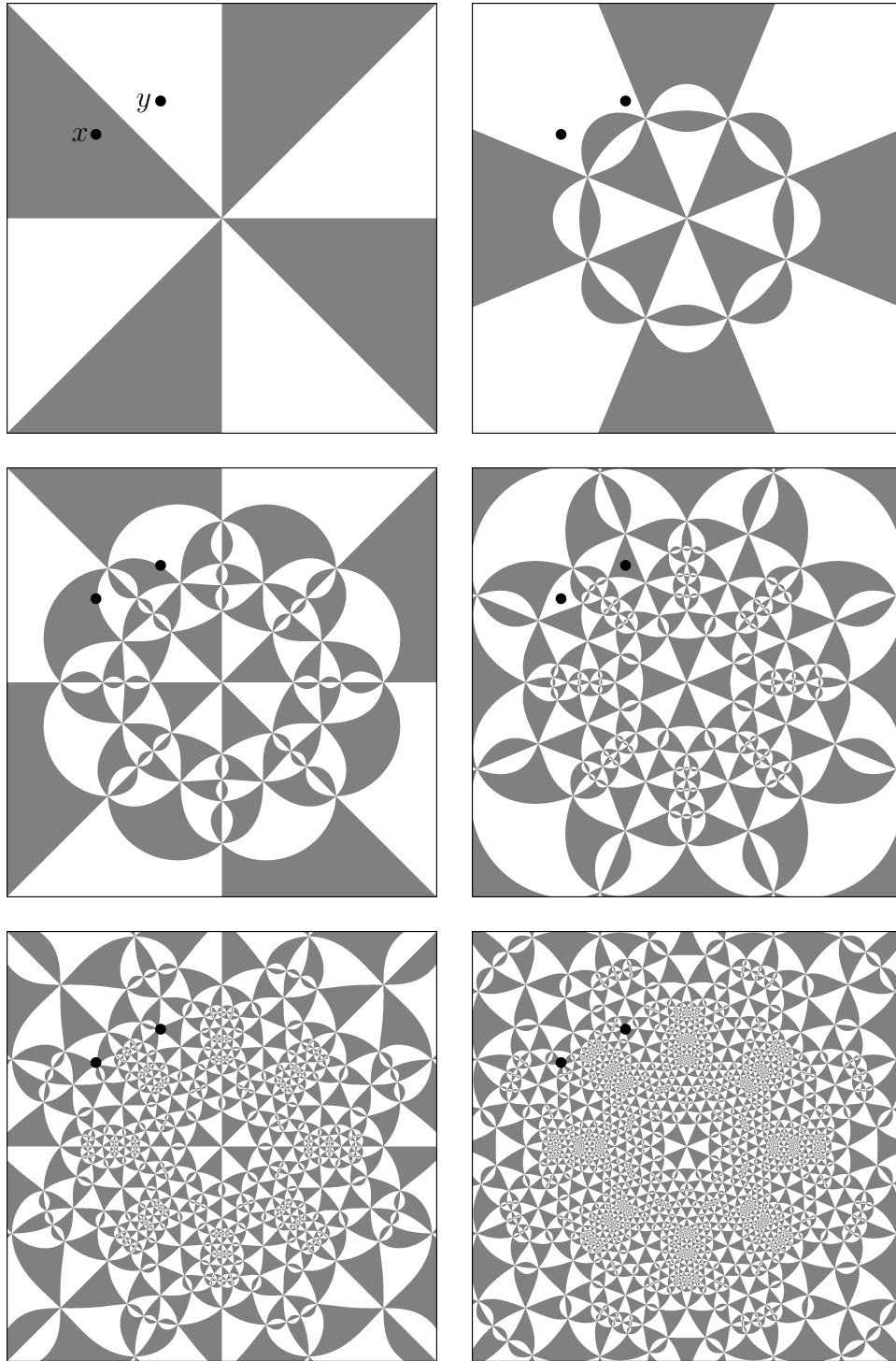
FIGURE 8.1. Separating points by tiles.

If $x = y$ we define $m_{f,\mathcal{C}}(x, x) := \infty$.

Note that $m_{f,\mathcal{C}}(x, y) \in \mathbb{N}$ if $x \neq y$. We usually drop both subscripts in $m_{f,\mathcal{C}}(x, y)$ if f and $\mathcal{C}$ are clear from the context. A similar combinatorial quantity that is essentially equivalent to $m_{f,\mathcal{C}}(x, y)$ (see Lemma 8.7 (v)) is

(8.1) $m'_{f,\mathcal{C}}(x, y) := \min\{n \in \mathbb{N}_0 :$ there exist disjoint n-tiles

$$X \text{ and } Y \text{ for } (f, \mathcal{C}) \text{ with } x \in X, \, y \in Y\},$$

for $x \neq y$ and $m'_{f,\mathcal{C}}(x, x) := \infty$.

These quantities are illustrated in Figure 8.1. Here we consider the map $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ given by

$$f(z) = i\frac{z^4 - i}{z^4 + i}$$

for $z \in \widehat{\mathbb{C}}$ (we will consider this map again in Example 15.11). We have $\mathrm{post}(f) = \{1, i, -i\}$, and so we can choose the unit circle $\mathcal{C} = \partial \mathbb{D}$ as a Jordan curve containing the postcritical set of f. In Figure 8.1 the n-tiles for $(f, \mathcal{C})$ are shown for $n = 1, \dots, 6$. For the points x and y as indicated in the figure, we have $m_{f,\mathcal{C}}(x, y) = 3$ and $m'_{f,\mathcal{C}}(x, y) = 4$.

The number $m_{f,\mathcal{C}}(x, y)$ is large if x and y are close together, i.e., if n-tiles of high level are needed to separate the points. This is the basis of the following definition.

DEFINITION 8.2 (Visual metrics). Let $f \colon S^2 \to S^2$ be an expanding Thurston map. A metric ϱ on S^2 is called a *visual metric* (for f) if there exists a Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}$, and a constant $\Lambda > 1$ such that

(8.2) $\varrho(x, y) \asymp \Lambda^{-m(x,y)}$

for all $x, y \in S^2$, where $m(x, y) = m_{f,\mathcal{C}}(x, y)$ and where the constant $C(\asymp)$ is independent of x and y.

Here we use the convention $\Lambda^{-\infty} = 0$. The number Λ is called the *expansion factor* of the metric ϱ. It is easy to see that the expansion factor of each visual metric is uniquely determined. Different visual metrics may have different expansion factors.

It is possible to identify the sphere S^2 with the boundary at infinity of a certain Gromov hyperbolic graph constructed from tiles. Under this identification, the numbers $m_{f,\mathcal{C}}(x, y)$ and $m'_{f,\mathcal{C}}(x, y)$ are the Gromov product of x and y, up to some additive constants (see Section 4.2, Chapter 10, and Lemma 10.3). A metric is then visual as defined above if and only if it is visual in the sense of Gromov hyperbolic spaces (see Theorem 10.2). This explains our terminology.

Obvious questions are whether visual metrics exist, and how they depend on the chosen Jordan curve $\mathcal{C}$ and the expansion factor Λ. This is answered by the following proposition.

PROPOSITION 8.3. *For an expanding Thurston map* $f\colon S^2 \to S^2$ *the following statements are true:*

(i) *There exist visual metrics for* f.

(ii) *Every visual metric induces the given topology on* S^2.

(iii) *Let* ϱ *be a visual metric for* f *with expansion factor* Λ, $\widetilde{\mathcal{C}} \subset S^2$ *be a Jordan curve with* $\mathrm{post}(f) \subset \widetilde{\mathcal{C}}$, *and* $m = m_{f,\widetilde{c}}$ *be defined as in Definition 8.1. Then a relation as in* (8.2) *is true with the same expansion factor* Λ, *where the constant* $C(\asymp)$ *depends on* $\widetilde{\mathcal{C}}$.

(iv) *Any two visual metrics are snowflake equivalent, and bi-Lipschitz equivalent if they have the same expansion factor* Λ.

(v) *A metric* ϱ *is a visual metric for some iterate* $F = f^n$ *with* $n \in \mathbb{N}$ *if and only if it is a visual metric for* f. *If* $\Lambda > 1$ *is the expansion factor of* ϱ *for* f, *then* $\Lambda_F = \Lambda^n$ *is the expansion constant of* ϱ *for* $F = f^n$.

(vi) *If* ϱ *is a visual metric for* f, *then* $f\colon (S^2, \varrho) \to (S^2, \varrho)$ *is a Lipschitz map.*

The notions of snowflake and bi-Lipschitz equivalence were defined in Section 4.1. The proof of Proposition 8.3 will be provided in Section 8.2. Theorem 16.3 gives a stronger result on the existence of visual metrics.

In Sections 1.3 and 4.4 we introduced visual metrics on a more intuitive level, where we considered certain self-similar fractal spheres constructed as limits of polyhedral surfaces $\mathcal{S}^n$. Each surface $\mathcal{S}^n$ was built from tiles whose size was about Λ^{-n} for some constant $\Lambda > 1$. A similar statement is true in general for visual metrics and gives in fact a characterization of these metrics.

PROPOSITION 8.4 (Characterization of visual metric). *Let* $f\colon S^2 \to S^2$ *be an expanding Thurston map and* ϱ *be a metric on* S^2. *Then* ϱ *is a visual metric for* f *with expansion factor* $\Lambda > 1$ *if and only if the following two conditions hold:*

(i) $\mathrm{dist}_\varrho(\sigma, \tau) \gtrsim \Lambda^{-n}$, *whenever* σ *and* τ *are disjoint n-cells,*

(ii) $\mathrm{diam}_\varrho(\tau) \asymp \Lambda^{-n}$ *for all n-edges and all n-tiles* τ.

Here cells are defined in terms of some Jordan curve $\mathcal{C} \subset S^2$ with post$(f) \subset \mathcal{C}$, and the implicit constants $C(\gtrsim)$ and $C(\asymp)$ are independent of the cells.

We will prove this lemma in Section 8.2.

For a rational expanding Thurston map $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ there are some other natural metrics on the Riemann sphere $\widehat{\mathbb{C}}$ besides the visual metrics for f, in particular the chordal metric σ and the canonical orbifold metric ω of f as introduced in Section 2.5. The chordal metric σ is never visual for f (see Lemma 8.12). For the canonical orbifold metric we will prove the following statement in Section 8.3.

PROPOSITION 8.5 (Canonical orbifold metric as visual metric). *Let $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a rational Thurston map without periodic critical points, and ω be the canonical orbifold metric for f. Then ω is a visual metric for f if and only if f is a Lattès map.*

8.1. The number $m(x, y)$

We now turn to a more detailed exposition of the basic properties of the quantity $m(x, y) = m_{f, \mathcal{C}}(x, y)$ as in Definition 8.1. In the following, $f \colon S^2 \to S^2$ will be an expanding Thurston map and $\mathcal{C} \subset S^2$ be a Jordan curve with post$(f) \subset \mathcal{C}$.

Recall the quantity $D_n = D_n(f, \mathcal{C})$ as defined in (5.14) that measures distances in terms of lengths of tile chains. We consider a slight variant here.

We define $\widetilde{D}_n = \widetilde{D}_n(f, \mathcal{C})$ as the minimal number of tiles of level $k \geq n$ for $(f, \mathcal{C})$ required to join opposite sides of $\mathcal{C}$, i.e., the smallest number $N \in \mathbb{N}$ for which there are tiles $X_i \in \bigcup_{k \geq n} \mathbf{X}^k$, $i = 1, \ldots, N$, such that $K = \bigcup_{i=1}^{N} X_i$ is connected and joins opposite sides of $\mathcal{C}$ (see Definition 5.32).

While the sets K used to define D_n are unions of tiles of level n, the sets K in the definition of $\widetilde{D}_n$ are unions of tiles of level $k \geq n$; in particular, $D_k \geq \widetilde{D}_n$ for $k \geq n$.

LEMMA 8.6. *Let $f \colon S^2 \to S^2$ be an expanding Thurston map, and $\mathcal{C} \subset S^2$ be a Jordan curve with post$(f) \subset \mathcal{C}$. Let $D_n = D_n(f, \mathcal{C})$ and $\widetilde{D}_n = \widetilde{D}_n(f, \mathcal{C})$ for $n \in \mathbb{N}_0$.*

Then $D_n \to \infty$ and $\widetilde{D}_n \to \infty$ as $n \to \infty$.

PROOF. We know that $D_k \geq \widetilde{D}_n$ whenever $k \geq n$. So it suffices to show $\widetilde{D}_n \to \infty$ as $n \to \infty$.

Let $\delta_0 > 0$ be defined as in (5.13) (for some base metric on S^2) and suppose $K = X_1 \cup \cdots \cup X_N$ is a connected union of tiles of level $\geq n$

that joins opposite sides of $\mathcal{C}$. Then

$$
\begin{aligned}
\delta_0 \;\; &\leq \;\; \operatorname{diam}(K) \leq \sum_{i=1}^{N} \operatorname{diam}(X_i) \\
&\leq \;\; N \max_{i=1,\ldots,N} \operatorname{diam}(X_i) \\
&\leq \;\; N \sup_{k \geq n} \operatorname{mesh}(f, k, \mathcal{C}).
\end{aligned}
$$

Putting $c_n := \sup_{k \geq n} \operatorname{mesh}(f, k, \mathcal{C})$, we conclude that $N \geq \delta_0/c_n$, and so $\widetilde{D}_n \geq \delta_0/c_n$.

Since f is expanding we have $\operatorname{mesh}(f, n, \mathcal{C}) \to 0$ and so also $c_n \to 0$ as $n \to \infty$. This implies that $\widetilde{D}_n \to \infty$ as desired. $\qquad\square$

If f is expanding and $\mathcal{C}$ is given, then in view of the last lemma, we can find a number $k_0 = k_0(f, \mathcal{C}) \in \mathbb{N}$ such that

(8.3) $$ \widetilde{D}_{k_0} = \widetilde{D}_{k_0}(f, \mathcal{C}) \geq 10. $$

This inequality will be useful in the following.

In the next lemma we collect some of the properties of the function $m_{f,\mathcal{C}}$.

LEMMA 8.7. *Let* $f \colon S^2 \to S^2$ *be an expanding Thurston map,* $\mathcal{C} \subset S^2$ *be a Jordan curve with* $\operatorname{post}(f) \subset \mathcal{C}$, *and* $m = m_{f,\mathcal{C}}$. *Then the following statements are true:*

(i) *There exists a number* $k_1 > 0$ *such that*

$$ \min\{m(x, z), m(y, z)\} \leq m(x, y) + k_1 $$

for all $x, y, z \in S^2$.

(ii) *We have*

$$ m(f(x), f(y)) \geq m(x, y) - 1 $$

for all $x, y \in S^2$.

(iii) *Let* $\widetilde{\mathcal{C}} \subset S^2$ *be another Jordan curve with* $\operatorname{post}(f) \subset \widetilde{\mathcal{C}}$. *Then there exists a constant* $k_2 > 0$ *such that*

$$ m(x, y) - k_2 \leq m_{f,\widetilde{\mathcal{C}}}(x, y) \leq m(x, y) + k_2 $$

for all $x, y \in S^2$.

(iv) *Let* $F = f^n$ *for* $n \in \mathbb{N}$ *be an iterate of* f. *Then there exists a constant* $k_3 > 0$ *such that*

$$ m(x, y) - k_3 \leq n \cdot m_{F,\mathcal{C}}(x, y) \leq m(x, y) $$

for all $x, y \in S^2$.

(v) *The quantities m and $m' = m'_{f,\mathcal{C}}$ as defined in (8.1) are comparable in the following sense: there exists a constant $k_4 > 0$ such that*

$$m(x,y) - k_4 \leq m'_{f,\mathcal{C}}(x,y) \leq m(x,y) + 1$$

for all $x, y \in S^2$.

In Chapter 10 we will prove that $m = m_{f,\mathcal{C}}$ can essentially be interpreted as a Gromov product in a suitable metric space (see Lemma 10.3). Property (i) is then related to the Gromov hyperbolicity of this space (compare with (4.7)).

PROOF. We fix $k_0 = k_0(f,\mathcal{C}) \in \mathbb{N}$ as in (8.3). Let $x, y \in S^2$ be arbitrary. In order to establish the desired inequalities we may always assume $x \neq y$. Unless otherwise stated, tiles will be for $(f,\mathcal{C})$.

(i) Let $m := m(x,y) \in \mathbb{N}_0$ be as in Definition 8.1. We can pick $(m+1)$-tiles X_0 and Y_0 containing x and y, respectively. Then $X_0 \cap Y_0 = \emptyset$ by definition of m.

Define $n := m + k_0$, and let $z \in S^2$ be arbitrary. We claim that $m(x,z) \leq n$ or $m(y,z) \leq n$.

Otherwise, $m(x,z) \geq n+1$ and $m(y,z) \geq n+1$, and so by Definition 8.1 there exist numbers $m_1, m_2 \geq n+1$ and m_1-tiles X and Z with $x \in X$, $z \in Z$ and $X \cap Z \neq \emptyset$, and m_2-tiles Y and Z' with $y \in Y$, $z \in Z'$ and $X \cap Z' \neq \emptyset$.

Then the set $K = X \cup Z \cup Z' \cup Y$ is connected and meets the disjoint $(m+1)$-tiles X_0 and Y_0. Thus $f^{m+1}(K)$ joins opposite sides of $\mathcal{C}$ by Lemma 5.35, and consists of four tiles of level $\geq n - m = k_0$. This contradicts (8.3), proving the claim.

So we have $m(x,z) \leq m + k_0$ or $m(y,z) \leq m + k_0$. This implies (i) with the constant $k_1 = k_0$ which is independent of x and y.

(ii) We may assume that $m := m(x,y) \geq 1$. There are non-disjoint m-tiles X and Y with $x \in X$ and $y \in Y$. It follows that $f(X)$ and $f(Y)$ are non-disjoint $(m-1)$-tiles with $f(x) \in f(X)$ and $f(y) \in f(Y)$. Hence $m(f(x), f(y)) \geq m - 1$ as desired.

(iii) Let $\widetilde{m} := m_{f,\widetilde{\mathcal{C}}}(x,y) \in \mathbb{N}_0$. Then there exist $\widetilde{m}$-tiles $\widetilde{X}$ and $\widetilde{Y}$ for $(f,\widetilde{\mathcal{C}})$ with $x \in \widetilde{X}$, $y \in \widetilde{Y}$, and $\widetilde{X} \cap \widetilde{Y} \neq \emptyset$. By Lemma 5.38 the sets $\widetilde{X}$ and $\widetilde{Y}$ are each contained in M $\widetilde{m}$-flowers for $(f,\mathcal{C})$, where M is independent of $\widetilde{X}$ and $\widetilde{Y}$. In particular, this implies that we can find a chain of at most $2M$ such $\widetilde{m}$-flowers joining x and y (recall that chains were introduced in Definition 5.19). Since any two tiles in the closure $\overline{W^n(v)}$ of an n-flower have the point v in common, it follows that there exists a chain $X_1, \ldots, X_N$ of $\widetilde{m}$-tiles for $(f,\mathcal{C})$ joining x and

y with $N \leq 4M$. Let $x_1 := x$, $x_N := y$, and for $i = 2, \ldots, N-1$, pick a point $x_i \in X_i$. Then $m(x_i, x_{i+1}) \geq \widetilde{m}$ for $i = 1, \ldots, N-1$. Hence by repeated application of (i) we obtain

$$
\begin{aligned}
\widetilde{m} &\leq \min\{m(x_i, x_{i+1}) : i = 1, \ldots, N-1\} \\
&\leq m(x_1, x_N) + Nk_1 \leq m(x, y) + 4Mk_1.
\end{aligned}
$$

Since $4Mk_1$ is independent of x and y, we get an upper bound as in (iii). A lower bound is obtained by the same argument if we reverse the roles of $\mathcal{C}$ and $\widetilde{\mathcal{C}}$.

(iv) The map F is also an expanding Thurston map, and we have $\mathrm{post}(f) = \mathrm{post}(F)$ (see Lemma 6.4); so the Jordan curve $\mathcal{C}$ contains the set of postcritical points of F and $m_{F,\mathcal{C}}$ is defined. The m-tiles for $(F, \mathcal{C})$ are precisely the (nm)-tiles for $(f, \mathcal{C})$ (see Proposition 5.17 (vii)). In the ensuing proof we will only consider tiles for $(f, \mathcal{C})$.

Let $m_F := m_{F,\mathcal{C}}(x, y)$ and $m := m(x, y)$; then there are non-disjoint (nm_F)-tiles X and Y with $x \in X$ and $y \in Y$. So $m \geq nm_F$ which gives the desired upper bound.

We claim that on the other hand, we have $m \leq nm_F + k_3$, where $k_3 = n + k_0 - 1$. To see this, assume that

$$
m \geq nm_F + k_3 + 1 = n(m_F + 1) + k_0.
$$

Then we can find non-disjoint m-tiles X and Y with $x \in X$, $y \in Y$. Moreover, we can pick $n(m_F + 1)$-tiles X' and Y' with $x \in X'$ and $y \in Y'$. By definition of m_F we know that $X' \cap Y' = \emptyset$; so X' and Y' are disjoint $n(m_F + 1)$-tiles joined by the connected set $K = X \cup Y$. Hence by Lemma 5.36 K must consist of at least

$$
D_{m-n(m_F+1)} \geq \widetilde{D}_{k_0} \geq 10
$$

m-tiles; but K consists of only two such m-tiles. This is a contradiction showing the desired claim.

(v) Let $m' := m'_{f,\mathcal{C}}(x, y)$ be defined as in (8.1). Then $m' \geq 1$, because the two 0-tiles have non-empty intersection. So $m' - 1 \geq 0$, and there exist $(m' - 1)$-tiles X and Y with $x \in X$ and $y \in Y$. Then $X \cap Y \neq \emptyset$ by definition of m', and so $m(x, y) \geq m' - 1$.

Conversely, let $m := m(x, y)$. Suppose $m' < m - k_0$. Then there exist m'-tiles X' and Y' with $X' \cap Y' = \emptyset$, m-tiles X and Y with $X \cap Y \neq \emptyset$, and $x \in X \cap X'$, $y \in Y \cap Y'$. Hence $K = X \cup Y$ is a union of two m-tiles joining the disjoint m'-tiles X' and Y'; but such a union must consist of at least

$$
D_{m-m'} \geq \widetilde{D}_{k_0} \geq 10
$$

m-tiles by Lemma 5.36. This is a contradiction showing that $m - k_0 \leq m'$. So the claim is true with $k_4 = k_0$. $\qquad\qquad\qquad\qquad\qquad\square$

8.2. Existence and basic properties of visual metrics

We are now ready to prove Propositions 8.3 and 8.4, and in particular the existence of visual metrics. In this section we will also collect various other and somewhat more technical statements related to visual metrics that will be useful later on.

PROOF OF PROPOSITION 8.3. (i) Fix a Jordan curve $\mathcal{C} \subset S^2$ with post$(f) \subset \mathcal{C}$.

Recall that a function $q\colon S^2 \times S^2 \to [0, \infty)$ is called a *quasimetric* if it has the symmetry property $q(x, y) = q(y, x)$, satisfies the condition $q(x, y) = 0 \Leftrightarrow x = y$, and the inequality

$$(8.4) \qquad\qquad q(x, y) \leq K(q(x, z) + q(z, y)),$$

holds for a constant $K \geq 1$ and all $x, y, z \in S^2$.

We now define a quasimetric q on S^2. For this purpose, we fix $\Lambda > 1$ and set

$$(8.5) \qquad\qquad\qquad q(x, y) := \Lambda^{-m(x,y)},$$

for $x, y \in S^2$, where $m(x, y) = m_{f,\mathcal{C}}(x, y) \in \mathbb{N}_0 \cup \{\infty\}$ is as in Definition 8.1.

Symmetry and the property $q(x, y) = 0 \Leftrightarrow x = y$ are clear. The quasi-triangle inequality (8.4) follows from Lemma 8.7 (i).

It is well known (see [**He01**, Proposition 14.5]) that a sufficient "snowflaking" of a quasimetric leads to a distance function that is comparable to a metric. This means there is $0 < \epsilon < 1$ and a metric ϱ on S^2 such that $\varrho \asymp q^\epsilon$. Then ϱ is a visual metric for f (with expansion factor Λ^ϵ).

(ii) Let ϱ be a visual metric for f satisfying (8.2), and d a fixed base metric on S^2 that induces the given topology of S^2. We have to show that if $x \in S^2$ and $\{x_i\}$ is a sequence in S^2, then $\varrho(x_i, x) \to 0$ if and only if $d(x_i, x) \to 0$ as $i \to \infty$.

Assume first that $\varrho(x_i, x) \to 0$ as $i \to \infty$. By (8.2) this is obviously equivalent to $m_i := m_{f,\mathcal{C}}(x_i, x) \to \infty$. For each i there are non-disjoint m_i-tiles X_i and Y_i with $x \in X, x_i \in Y$. Thus

$$d(x_i, x) \leq \operatorname{diam}_d(X_i) + \operatorname{diam}_d(Y_i) \leq 2\operatorname{mesh}(f, m_i, \mathcal{C}).$$

Since f is expanding and $m_i \to \infty$, the latter expression approaches 0 as $i \to \infty$. Hence $d(x_i, x) \to 0$ as $i \to \infty$.

Conversely, suppose that $d(x_i, x) \to 0$ as $i \to \infty$. Let $n \in \mathbb{N}_0$ be arbitrary. Then x lies in some n-flower $W^n(p)$ (see Lemma 5.29 (iv)).

Since flowers are open sets, we have $x_i \in W^n(p)$ for sufficiently large i. For each of these i we can find n-tiles X and Y with $x \in X$, $x_i \in Y$, and $p \in X \cap Y$. This implies $m_i \geq n$. Therefore $m_i \to \infty$, and hence $\varrho(x_i, x) \to \infty$ as desired.

(iii) This follows from Lemma 8.7 (iii).

(iv) This follows from (iii) and the definition of a visual metric.

(v) This follows from (iii) and Lemma 8.7 (iv).

(vi) This follows from Lemma 8.7 (ii). $\square$

If two expanding Thurston maps are topologically conjugate, then their visual metrics are closely related.

PROPOSITION 8.8. *Let $f\colon S^2 \to S^2$ and $g\colon \widehat{S}^2 \to \widehat{S}^2$ be expanding Thurston maps that are topologically conjugate. Then S^2 equipped with any visual metric with respect to f is snowflake equivalent to $\widehat{S}^2$ equipped with any visual metric with respect to g. Every homeomorphism $h\colon S^2 \to \widehat{S}^2$ satisfying $h \circ f = g \circ h$ is a snowflake equivalence.*

PROOF OF PROPOSITION 8.8. By our assumptions there exists a topological conjugacy between f and g, i.e., a homeomorphism $h\colon S^2 \to \widehat{S}^2$ such that $h \circ f = g \circ h$. Let ϱ be a visual metric on S^2 with respect to f, and $\widehat{\varrho}$ be a visual metric on $\widehat{S}^2$ with respect to g. Let $\Lambda > 1$ and $\widehat{\Lambda} > 1$ be the expansion factors of ϱ and $\widehat{\varrho}$, respectively. It suffices to show that $h\colon (S^2, \varrho) \to (\widehat{S}^2, \widehat{\varrho})$ is a snowflake equivalence.

To see this, pick a Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}$. Then $\widehat{\mathcal{C}} = h(\mathcal{C})$ is a Jordan curve in $\widehat{S}^2$ with $\mathrm{post}(g) = h(\mathrm{post}(f)) \subset \widehat{\mathcal{C}}$. Since h conjugates f and g, it follows from Proposition 5.17 (iii) and (v) or, alternatively, from the uniqueness statement in Lemma 5.12 that for each $n \in \mathbb{N}_0$ the images of the cells in the cell decomposition $\mathcal{D}^n \coloneqq \mathcal{D}^n(f, \mathcal{C})$ of S^2 under the map h are precisely the cells in the cell decomposition $\widehat{\mathcal{D}}^n = \mathcal{D}^n(g, \widehat{\mathcal{C}})$ of $\widehat{S}^2$; so we have

(8.6) $$\widehat{\mathcal{D}}^n = \{h(c) : c \in \mathcal{D}^n\}$$

for all $n \in \mathbb{N}_0$. This implies that

$$\widehat{m}(h(x), h(x')) = m(x, x')$$

for all $x, x' \in S^2$, where $\widehat{m} = m_{g,\widehat{\mathcal{C}}}$ and $m = m_{f,\mathcal{C}}$ (recall Definition 8.1). Combining this with Proposition 8.3 (iii) we see that

$$\widehat{\varrho}(h(x), h(x')) \asymp \widehat{\Lambda}^{-\widehat{m}(h(x), h(x'))} = \widehat{\Lambda}^{-m(x,x')} = \Lambda^{-\alpha m(x,x')} \asymp \varrho(x, x')^\alpha$$

for all $x, x' \in S^2$, where $\alpha = \log(\widehat{\Lambda})/\log(\Lambda)$ and the implicit multiplicative constants do not depend on x and x'. It follows that h is a snowflake equivalence. $\square$

We now prove the geometric characterization of visual metrics.

PROOF OF PROPOSITION 8.4. Let $f\colon S^2 \to S^2$ be an expanding Thurston map and $\mathcal{C} \subset S^2$ be a Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}$.

We first show that a visual metric ϱ for f has the properties (i) and (ii). By Proposition 8.3 (iii) we may assume that ϱ satisfies (8.2) for $m = m_{f,\mathcal{C}}$ and a constant $C = C(\asymp)$.

(i) Let $k_0 \in \mathbb{N}$ be defined as in (8.3), and let σ and τ be disjoint n-cells. If $x \in \sigma$ and $y \in \tau$ are arbitrary, then $m(x, y) < n + k_0$. Indeed, if this were not the case, then we could find $(n+k)$-tiles X and Y with $x \in X$, $y \in Y$, $X \cap Y \neq \emptyset$, and $k \geq k_0$. Then $K = X \cup Y$ is a connected set meeting disjoint n-cells. Hence by Lemma 5.35 the set $f^n(K)$ joins opposite sides of $\mathcal{C}$. On the other hand, $f^n(K)$ consists of two k-tiles $X' = f^n(X)$ and $Y' = f^n(Y)$, where $k \geq k_0$. This is impossible by definition of k_0.

Therefore, $\varrho(x, y) \geq (1/C)\Lambda^{-n-k_0}$, and so we get the desired bound $\mathrm{dist}_\varrho(\sigma, \tau) \geq (1/C')\Lambda^{-n}$ with the constant $C' = C\Lambda^{k_0}$ that is independent of n, σ, and τ.

(ii) If x, y are points in some n-tile X, then $m(x, y) \geq n$. Since every n-edge is contained in an n-tile, this inequality is still true if x and y are contained in an n-edge. Hence $\varrho(x, y) \leq C\Lambda^{-m(x,y)} \leq C\Lambda^{-n}$, and so $\mathrm{diam}_\varrho(\tau) \leq C\Lambda^{-n}$ whenever τ is an n-tile or n-edge, where $C = C(\asymp)$ is the constant from (8.2).

A similar lower bound for the diameter of an n-edge or n-tile τ follows from (i) and the fact that every n-edge or n-tile contains two distinct n-vertices.

To prove the converse implication, suppose that ϱ is a metric on S^2 with the properties (i) and (ii) as in the statement. We want to show that ϱ is visual for f. Let $x, y \in S^2$, $x \neq y$, be arbitrary, and $m = m_{f,\mathcal{C}}(x, y)$.

Then we can find m-tiles X and Y with $x \in X$, $y \in Y$ and $X \cap Y \neq \emptyset$. By (ii) we have

$$\varrho(x, y) \leq \mathrm{diam}_\varrho(X) + \mathrm{diam}_\varrho(Y) \lesssim \Lambda^{-m}.$$

We can also find $(m+1)$-tiles X' and Y' with $x \in X'$, $y \in Y'$. By definition of m we then have $X' \cap Y' = \emptyset$. Hence by (i)

$$\varrho(x, y) \geq \mathrm{dist}_\varrho(X', Y') \gtrsim \Lambda^{-m}.$$

Since the implicit multiplicative constants in the previous inequalities are independent of x and y, it follows that ϱ is a visual metric for f. $\square$

It is possible to establish the phenomenon of "exponential shrinking" as in Proposition 8.4 (ii) also for other types of sets. For example, we have

$$(8.7) \qquad \operatorname{diam}_\varrho(W^n(p)) \le C\Lambda^{-n}$$

for every n-flower for $(f, \mathcal{C})$ where the constant C is independent of n and p. Of particular importance will be exponential shrinking for lifts of paths.

LEMMA 8.9. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map, and ϱ be a visual metric for f with expansion factor $\Lambda > 1$. Then for every path $\gamma\colon [0, 1] \to S^2$ there exists a constant $A > 0$ with the following property: if $n \in \mathbb{N}$ and $\widetilde{\gamma}$ is any lift of γ under f^n, then*

$$\operatorname{diam}_\varrho(\widetilde{\gamma}) \le A\Lambda^{-n}.$$

Note that lifts of γ under f^n exist by Lemma A.17.

PROOF. Pick a Jordan curve $\mathcal{C} \subset S^2$ with $\operatorname{post}(f) \subset \mathcal{C}$, and let $\delta_0 > 0$ be as in (5.13) with ϱ as the base metric on S^2. Then we can break up γ into a finite number of paths γ_i, $i = 1, \dots, N$, traversed in successive order such that $\operatorname{diam}_\varrho(\gamma_i) < \delta_0$ for $i = 1, \dots, N$. By Lemma 5.34 (ii) each lift of a pieces γ_i is contained in one n-flower, and so the whole lift $\widetilde{\gamma}$ in N n-flowers. Hence by (8.7) we have $\operatorname{diam}_\varrho(\widetilde{\gamma}) \le CN\Lambda^{-n}$, where $C > 0$ is independent of n and γ. The statement follows with $A = CN$. $\square$

In general, the constant A in the last lemma will depend on γ, but the proof shows that we can take the same constant A for a family of paths if there exists $N \in \mathbb{N}$ such that each path can be broken up into at most N subpaths of diameter $< \delta_0$.

Let f be an expanding Thurston map, and $\mathcal{C} \subset S^2$ be a Jordan curve with $\operatorname{post}(f) \subset \mathcal{C}$. It is useful to define neighborhoods of points by using the cells in our decompositions $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$. To do this, let $x \in S^2$, $n \in \mathbb{N}_0$, and set

$$(8.8) \qquad U^n(x) = \bigcup \{Y \in \mathbf{X}^n : \text{there exists an } n\text{-tile } X \text{ with}$$

$$x \in X \text{ and } X \cap Y \ne \emptyset\}.$$

It is convenient to define $U^n(x)$ also for negative integers n. We set $U^n(x) = U^0(x) = S^2$ for $n < 0$.

As we will see momentarily, the sets $U^n(x)$ resemble metric balls defined in terms of a visual metric very closely.

LEMMA 8.10. *Let ϱ be a visual metric for f with expansion factor $\Lambda > 1$. Then there are constants $K \geq 1$ and $n_0 \in \mathbb{N}_0$ with the following properties.*

(i) *For all $x \in S^2$ and all $n \in \mathbb{Z}$,*

$$B_\varrho(x, r/K) \subset U^n(x) \subset B_\varrho(x, Kr),$$

where $r = \Lambda^{-n}$.

(ii) *For all $x \in S^2$ and all $r > 0$,*

$$U^{n+n_0}(x) \subset B_\varrho(x, r) \subset U^{n-n_0}(x),$$

where $n = \lceil -\log r / \log \Lambda \rceil$.

PROOF. (i) Let $m = m_{f,\mathcal{C}}$. If $y \in U^n(x)$, then $m(x, y) \geq n$, and so $\varrho(x, y) \lesssim \Lambda^{-n} = r$. This gives the inclusion $U^n(x) \subset B_\varrho(x, Kr)$ for a suitable constant K independent of x and n.

Conversely, suppose that $y \notin U^n(x)$. Then $n \geq 1$. If we pick any n-tiles X and Y with $x \in X$ and $y \in Y$, then $X \cap Y = \emptyset$ by definition of $U^n(x)$. So by Proposition 8.4 (i) we have

$$\varrho(x, y) \geq \text{dist}_\varrho(X, Y) \gtrsim \Lambda^{-n} = r.$$

Hence $B_\varrho(x, r/K) \subset U^n(x)$ if K is suitably large independent of x and r.

(ii) Choose $n_0 = \lceil \log K / \log \Lambda \rceil + 1$, where K is as in (i). Then $\Lambda^{-n_0} \leq 1/(\Lambda K)$. Moreover, $\Lambda^{-n} \leq r \leq \Lambda \Lambda^{-n}$, and so

$$K\Lambda^{-n-n_0} \leq r \leq (1/K)\Lambda^{-n+n_0}.$$

The desired inclusion then follows from (i).   □

We next show that when S^2 is equipped with a visual metric, the tiles are "quasi-round". Furthermore, every tile contains points that are "deep inside" the tile.

LEMMA 8.11. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map, $\mathcal{C} \subset S^2$ be a Jordan curve with $\text{post}(f) \subset \mathcal{C}$, and ϱ be a visual metric for f with expansion factor $\Lambda > 1$. Then there exists a constant $C \geq 1$ with the following property: for every n-tile X for $(f, \mathcal{C})$ there exists a point $p \in X$ such that*

$$B_\varrho(p, (1/C)\Lambda^{-n}) \subset X \subset B_\varrho(p, C\Lambda^{-n}).$$

PROOF. With a suitable constant C independent of n, an inclusion of the form

$$X \subset B_\varrho(p, C\Lambda^{-n})$$

holds for every n-tile X and every point $p \in X$ as follows from Proposition 8.4 (ii).

The main difficulty for an inclusion in the opposite direction is to find an appropriate point p. For this purpose, let $k_0 \in \mathbb{N}$ be the number defined in (8.3), and X be an arbitrary n-tile. Since f is an expanding Thurston map, we have $\mathrm{post}(f) \geq 3$ (see Corollary 7.2), and so ∂X contains at least three distinct n-vertices v_1, v_2, v_3. Using these vertices, we can find three arcs $\alpha_1, \alpha_2, \alpha_3 \subset \partial X$ with pairwise disjoint interior such that $\partial X = \alpha_1 \cup \alpha_2 \cup \alpha_3$ and such that α_i has the endpoints v_i and v_{i+1} for $i = 1, 2, 3$, where $v_4 = v_1$. In general, α_i will not be an n-edge, but since it lies on ∂X, and its endpoints are n-vertices, it is the union of all the n-edges that it contains.

We now define

$$A_i = \bigcup_{x \in \alpha_i} U^{n+k_0}(x)$$

for $i = 1, 2, 3$, where $U^{n+k_0}(x)$ is given as in (8.8). Then the set A_i is the union of all $(n + k_0)$-tiles that meet an $(n + k_0)$-tile that has non-empty intersection with α_i. In particular, A_i is a closed set that contains α_i.

We claim that the sets A_1, A_2, A_3 do not form a cover of X. To reach a contradiction, suppose that $X \subset A_1 \cup A_2 \cup A_3$. We can regard X as a topological simplex with the sides α_i, $i = 1, 2, 3$. Then the closed sets A_1, A_2, A_3 form a cover of X such that each set A_i contains the side α_i of the simplex for $i = 1, 2, 3$. A well-known result due to Sperner [**AH35**, p. 378] then implies that $A_1 \cap A_2 \cap A_3 \neq \emptyset$.

Pick a point $x \in A_1 \cap A_2 \cap A_3$. Then by definition of A_i, there exist $(n + k_0)$-tiles X_i and Y_i with $X_i \cap \alpha_i \neq \emptyset$, $x \in Y_i$, and $X_i \cap Y_i \neq \emptyset$, where $i = 1, 2, 3$. Then the set

$$K = \bigcup_{i=1}^{3} (X_i \cup Y_i)$$

consists of at most six $(n+k_0)$-tiles, is connected, and meets each of the arcs $\alpha_1, \alpha_2, \alpha_3$. Hence $K' = f^n(K)$ is a connected set that consists of at most six k_0-tiles, and meets each of the arcs $\beta_i = f^n(\alpha_i)$, $i = 1, 2, 3$. Note that each arc β_i is the union of all 0-edges that it contains. Hence for $i = 1, 2, 3$ there exists a 0-edge $e_i \subset \beta_i$ with $e_i \cap K' \neq \emptyset$. Since the arcs $\beta_1, \beta_2, \beta_3$ have pairwise disjoint interior, it follows that the 0-edges e_1, e_2, e_3 are all distinct. So K' is a connected set that meets three distinct 0-edges. Hence it joins opposite sides of $\mathcal{C}$. So K' should contain at least $D_{k_0} \geq \widetilde{D}_{k_0} \geq 10$ tiles of level k_0. This is a contradiction, because K' is a union of at most six k_0-tiles.

This proves the claim that the sets A_1, A_2, A_3 do not cover X, and we conclude that we can find a point

$$p \in X \setminus (A_1 \cup A_2 \cup A_3).$$

We claim that $U^{n+k_0}(p) \subset X$. Otherwise, we could find a point $y \in U^{n+k_0}(p) \setminus X$, and $(n + k_0)$-tiles U and V with $p \in U$, $y \in V$, and $U \cap V \neq \emptyset$. Then the connected set $U \cup V$ must meet ∂X, and hence one of the arcs α_i; but then $p \in A_i$ by definition of A_i. This is a contradiction showing the desired inclusion $U^{n+k_0}(p) \subset X$.

From Lemma 8.10 (i) it now follows that $B_\varrho(p, (1/C)\Lambda^{-n}) \subset X$, where $C \geq 1$ is a constant independent of n and X. $\qquad \square$

8.3. The canonical orbifold metric as a visual metric

If $f\colon S^2 \to S^2$ is an expanding Thurston map f, one can ask whether other standard metrics on S^2 are visual metrics. To have some natural metrics available, we restrict ourselves here to rational expanding Thurston maps f defined on $\widehat{\mathbb{C}}$. Recall that a rational Thurston map is expanding if and only if it does not have any periodic critical point (see Proposition 2.3).

LEMMA 8.12. *Let $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a rational expanding Thurston map. Then the chordal metric σ on $\widehat{\mathbb{C}}$ is not a visual metric for f.*

PROOF. We argue by contradiction and assume that σ is a visual metric for f with expansion factor $\Lambda > 1$. We pick a critical point $c \in \widehat{\mathbb{C}}$ of f and set $d := \deg(f, c) \geq 2$.

We consider tiles for $(f, \mathcal{C})$, where $\mathcal{C} \subset \widehat{\mathbb{C}}$ is a fixed Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}$ as usual. For each $n \in \mathbb{N}$ let X^n be an n-tile that contains c. Then $\mathrm{diam}_\sigma(X^n) \asymp \Lambda^{-n}$ by Proposition 8.4 (ii). Since in suitable local conformal coordinates the map f near c behaves like $z \mapsto z^d$ near 0, for $Y^n := f(X^{n+1})$ we have

$$(8.9) \qquad \mathrm{diam}_\sigma(Y^n) = \mathrm{diam}_\sigma(f(X^{n+1})) \asymp \mathrm{diam}_\sigma(X^{n+1})^d$$

$$\asymp \Lambda^{-(n+1)d} \asymp \Lambda^{-nd}$$

for $n \in \mathbb{N}_0$, where $C(\asymp)$ is independent of n. On the other hand, Y^n is an n-tile and so $\mathrm{diam}_\sigma(Y^n) \asymp \Lambda^{-n}$ by Proposition 8.4 (ii), where $C(\asymp)$ is independent of n. Since $d \geq 2$ this is irreconcilable with (8.9) for large n and so we reach a contradiction. $\qquad \square$

We now turn to the canonical orbifold metric $\omega = \omega_f$ of a given rational expanding Thurston map f (see Section 2.5) and its relation to visual metrics. We will reformulate and prove the two implications in Proposition 8.5 separately. First we prove the "only if" statement.

LEMMA 8.13. *Let $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a rational expanding Thurston map such that its canonical orbifold metric $\omega = \omega_f$ is a visual metric for f. Then f is a Lattès map.*

PROOF. The metric ω is the canonical orbifold metric of the associated orbifold $\mathcal{O}_f = (\widehat{\mathbb{C}}, \alpha_f)$. Note that $\alpha_f(u) < \infty$ for $u \in \widehat{\mathbb{C}}$ by Proposition 2.9 (ii), because f is an expanding rational Thurston map and so it does not have periodic critical points.

Suppose ω is a visual metric for f with expansion factor $\Lambda > 1$. Let $p \in \mathbb{C}$ and $q \in f^{-1}(p)$ be arbitrary, and set $d := \deg_f(q)$.

We consider tiles for $(f, \mathcal{C})$, where $\mathcal{C} \subset \widehat{\mathbb{C}}$ is a fixed Jordan curve with $\operatorname{post}(f) \subset \mathcal{C}$. For each $n \in \mathbb{N}_0$ we pick an n-tile X^n with $q \in X^n$. Then $Y^n := f(X^{n+1})$ is an n-tile containing $p = f(q)$.

As in the proof of Lemma 8.12, we have

$$\operatorname{diam}_\sigma(Y^n) \asymp \operatorname{diam}_\sigma(X^{n+1})^d.$$

On the other hand, one can relate ω and the chordal metric σ. Namely, one can show (see (A.41)) that if U is a sufficiently small neighborhood of p, then

$$\omega(p, u) \asymp \sigma(p, u)^{1/\alpha_f(p)}$$

for all $u \in U$. This implies that

$$\operatorname{diam}_\omega(Y^n) \asymp \operatorname{diam}_\sigma(Y^n)^{1/\alpha_f(p)}$$

for large n. Similarly,

$$\operatorname{diam}_\omega(X^n) \asymp \operatorname{diam}_\sigma(X^n)^{1/\alpha_f(q)}$$

for large n.

Since ω is a visual metric for f with expansion factor Λ, we also have

$$\operatorname{diam}_\omega(X^n) \asymp \operatorname{diam}_\omega(Y^n) \asymp \Lambda^{-n}$$

for $n \in \mathbb{N}_0$ by Proposition 8.4 (ii).

If we combine all these estimates, we arrive at

$$\Lambda^{-\alpha_f(p)n} \asymp \operatorname{diam}_\omega(Y^n)^{\alpha_f(p)} \asymp \operatorname{diam}_\sigma(Y^n)$$
$$\asymp \operatorname{diam}_\sigma(X^{n+1})^d \asymp \operatorname{diam}_\omega(X^{n+1})^{d\alpha_f(q)}$$
$$\asymp \Lambda^{-d\alpha_f(q)(n+1)} \asymp \Lambda^{-d\alpha_f(q)n}$$

for all large n, where all the implicit constants $C(\asymp)$ are independent of n. This is only possible if $\alpha_f(p) = d\alpha_f(q) = \deg_f(q)\alpha_f(q)$.

We conclude that α_f satisfies condition (iii) in Proposition 2.13, and so $\mathcal{O}_f$ is parabolic. Hence f is a Lattès map by Theorem 3.1. $\square$

We now prove the "if" implication of Proposition 8.5.

PROPOSITION 8.14. *Let $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a Lattès map, and $\omega = \omega_f$ be the canonical orbifold metric of f on $\widehat{\mathbb{C}}$. Then ω is a visual metric for f with expansion factor $\Lambda = \deg(f)^{1/2} > 1$. Moreover, ω is a geodesic metric on $\widehat{\mathbb{C}}$ and for all paths γ in $\widehat{\mathbb{C}}$ we have*

$$(8.10) \qquad \operatorname{length}_\omega(f \circ \gamma) = \Lambda \operatorname{length}_\omega(\gamma).$$

We will see later (in Proposition 20.1 and Theorem 16.3) that for given degree of an expanding Thurston map f the number $\Lambda = \deg(f)^{1/2}$ is the largest possible expansion factor of a visual metric. So Lattès maps are special as they realize this maximal factor. This is closely related to the characterization of Lattès maps among expanding Thurston maps given in Theorem 20.2.

PROOF. Let $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a Lattès map. Then ω is the canonical orbifold metric of the parabolic orbifold $\mathcal{O} = (\widehat{\mathbb{C}}, \alpha_f)$ associated with f. Since f has no periodic critical points, we have $\alpha_f(p) < \infty$ for all $p \in \widehat{\mathbb{C}}$ (Proposition 2.9 (ii)) and so $\mathcal{O}$ has no punctures. In particular, ω is a geodesic metric defined on the whole Riemann sphere $\widehat{\mathbb{C}}$ (see the discussion after (A.40) in Section A.10).

Let $\Theta \colon \mathbb{C} \to \widehat{\mathbb{C}}$ and $A \colon \mathbb{C} \to \mathbb{C}$ be holomorphic maps for f as in Theorem 3.1 (ii). Then $f \circ \Theta = \Theta \circ A$, and A has the form $A(z) = \alpha z + \beta$, where $\alpha, \beta \in \mathbb{C}$ with $\deg(f) = |\alpha|^2 \geq 2$ (see Lemma 3.16). We know that Θ is the universal orbifold covering map of $\mathcal{O}$ and that the metric ω is essentially the push-forward of the Euclidean metric on $\mathbb{C}$ by the map Θ. More precisely, Θ is a path isometry in the sense that

$$(8.11) \qquad \operatorname{length}(\gamma) = \operatorname{length}_\omega(\Theta \circ \gamma),$$

whenever γ is a path in $\mathbb{C}$ (see (A.40)). Here and below metric notions on $\mathbb{C}$ (such as $\operatorname{length}(\beta)$) refer to the Euclidean metric, while on $\widehat{\mathbb{C}}$ we use the metric ω indicated by a subscript.

Define $\Lambda = |\alpha| = \deg(f)^{1/2} > 1$ and let γ be an arbitrary path in $\widehat{\mathbb{C}}$. Then γ has a lift under the branched covering map Θ, and so there exists a path $\widetilde{\gamma}$ in $\mathbb{C}$ such that $\gamma = \Theta \circ \widetilde{\gamma}$ (see Lemma A.17). Then $f \circ \Theta = \Theta \circ A$ implies

$$\Theta \circ A^n \circ \widetilde{\gamma} = f^n \circ \Theta \circ \widetilde{\gamma} = f^n \circ \gamma$$

for all $n \in \mathbb{N}$.

Note that A^n for $n \in \mathbb{N}$ is a Euclidean similarity on $\mathbb{C}$ scaling distances by the factor $|\alpha|^n = \Lambda^n$. Since Θ is a path isometry, we

conclude that

$$(8.12) \qquad \text{length}_\omega(f^n \circ \gamma) = \text{length}(A^n \circ \widetilde{\gamma})$$
$$= \Lambda^n \, \text{length}(\widetilde{\gamma}) = \Lambda^n \, \text{length}_\omega(\gamma).$$

If we take $n = 1$ here, then (8.10) follows.

In order to verify that ω is a visual metric for f with expansion factor Λ, we will show that ω satisfies the conditions in Proposition 8.4. For this, we pick a Jordan curve $\mathcal{C} \subset \widehat{\mathbb{C}}$ with $\text{post}(f) \subset \mathcal{C}$ and consider n-cells for $(f, \mathcal{C})$.

Suppose σ and τ are two disjoint n-cells, where $n \in \mathbb{N}_0$. Since ω is a geodesic metric, we can find a path γ in $\widehat{\mathbb{C}}$ joining σ and τ with

$$(8.13) \qquad \text{length}_\omega(\gamma) = \text{dist}_\omega(\sigma, \tau).$$

By Lemma 5.35 the path $f^n \circ \gamma$ joins opposite sides of $\mathcal{C}$, and so

$$\text{length}_\omega(f^n \circ \gamma) \geq \text{diam}_\omega(f^n \circ \gamma) \geq \delta_0,$$

where $\delta_0 > 0$ is defined as in (5.13) for the base metric ω on $\widehat{\mathbb{C}}$.

Combining the last inequality with (8.13) and (8.12), we arrive at the inequality

$$(8.14) \qquad \text{dist}_\omega(\sigma, \tau) \gtrsim \Lambda^{-n}.$$

Here and in the following, the implicit multiplicative constant $C(\gtrsim)$ is independent of the cells and their level n. So ω satisfies condition (i) in Proposition 8.4.

Since every n-edge or n-tile τ contains two distinct n-vertices, (8.14) also implies that

$$(8.15) \qquad \text{diam}_\omega(\tau) \gtrsim \Lambda^{-n}.$$

To complete the proof, we have to establish an inequality in the opposite direction. First note that by Lemma 6.13 we can find a number $\delta_1 > 0$ with the following property: if $K \subset \mathbb{C}$ is a connected set with $\text{diam}_\omega(\Theta(K)) < \delta_1$, then we have $\text{diam}(K) < 1$. Since f is expanding, we can choose $n_0 \in \mathbb{N}$ such that $\text{diam}_\omega(X) < \delta_1$, whenever X is an n-tile with level $n \geq n_0$.

Now suppose X is an arbitrary n-tile with $n \geq n_0$. Define $k = n - n_0 \in \mathbb{N}_0$ and $U = \text{int}(X)$. Then $U \subset \widehat{\mathbb{C}}$ is a simply connected region that does not contain any n-vertex and so $U \subset \widehat{\mathbb{C}} \backslash \text{post}(f)$. Since $\Theta \colon \mathbb{C} \setminus \Theta^{-1}(\text{post}(f)) \to \widehat{\mathbb{C}} \setminus \text{post}(f)$ is a covering map, the inclusion map $U \to \widehat{\mathbb{C}} \setminus \text{post}(f)$ lifts under Θ (see Lemma A.6). This implies that there exists a region $V \subset \mathbb{C}$ such that $\Theta(V) = U$. Note that $\text{diam}_\omega(U) \leq \text{diam}(V)$, because distances do not increase under the map Θ as follows from (8.11).

Moreover,

$$\Theta(A^k(V)) = f^k(\Theta(V)) = f^k(U) \subset f^k(X). \tag{8.16}$$

On the other hand, $f^k(X)$ is a tile of level $n - k = n_0$, and so

$$\delta_1 > \operatorname{diam}_\omega(f^k(X)) \geq \operatorname{diam}_\omega(f^k(U)).$$

Since $A^k(V)$ is connected, this inequality, the relation (8.16), and the definition of δ_1 imply that

$$\operatorname{diam}(A^k(V)) \leq 1.$$

It follows that

$$\operatorname{diam}_\omega(X) = \operatorname{diam}_\omega(\overline{U}) = \operatorname{diam}_\omega(U) \leq \operatorname{diam}(V)$$
$$= \Lambda^{-k} \operatorname{diam}(A^k(V)) \leq \Lambda^{-k} = \Lambda^{n_0}\Lambda^{-n} \lesssim \Lambda^{-n}.$$

For n-tiles X with $n \leq n_0$ we trivially have $\operatorname{diam}_\omega(X) \asymp 1 \asymp \Lambda^{-n}$. Hence $\operatorname{diam}_\omega(X) \lesssim \Lambda^{-n}$ for tiles X on any level $n \in \mathbb{N}_0$ if we choose a suitable constant $C(\lesssim)$. Since every n-edge is contained in an n-tile, we have $\operatorname{diam}_\omega(\tau) \lesssim \Lambda^{-n}$ whenever τ is an n-tile or n-edge, $n \in \mathbb{N}_0$. Here we actually have $\operatorname{diam}_\omega(\tau) \asymp \Lambda^{-n}$ by (8.15).

This shows that ω also satisfies condition (ii) in Proposition 8.4. It follows that ω is indeed a visual metric for f with expansion factor Λ. $\square$

Note that Proposition 8.5 clearly follows from Proposition 8.14 and Lemma 8.12.

CHAPTER 9

Symbolic dynamics

If one wants to understand a dynamical system (X, f) given by the iteration of a map f on a space X, then often one tries to find a link to symbolic dynamics and the theory of shift operators, in particular to shifts of finite type. These operators serve as an important paradigm in dynamics. In this chapter we study this for expanding Thurston maps, but we will be exclusively concerned with topological aspects. One could also investigate measure-theoretic properties of expanding Thurston maps and their relation to (Bernoulli) shift operators, but we will not pursue this here (see [**HH02**] for a relevant paper in this context).

We will prove the following statement.

THEOREM 9.1. *Let* $f\colon S^2 \to S^2$ *be an expanding Thurston map. Then* f *is a factor of the left-shift* $\Sigma\colon J^\omega \to J^\omega$ *on the space* J^ω *of all sequences in a finite set* J *of cardinality* $\#J = \deg(f)$.

The notation and terminology will be explained below.

Theorem 9.1 is essentially due to Kameyama (see [**Ka03**, Theorem 3.4]). The basic idea seems to go back to [**Jo98**] (see also [**Pr85**]). Kameyama's notion of an expanding Thurston map is different from ours, but his proof carries over to our setting with only minor modifications.

It is a standard fact in complex dynamics that the repelling periodic points of a rational map on $\widehat{\mathbb{C}}$ are dense in its Julia set. The following statement is an analog of this for expanding Thurston maps. As we will see, it easily follows from the proof of Theorem 9.1.

COROLLARY 9.2. *Let* $f\colon S^2 \to S^2$ *be an expanding Thurston map. Then the periodic points of* f *are dense in* S^2.

Before we supply the proofs of these results, we will first review some basic definitions related to shift operators.

Let J be a finite non-empty set. We consider J as an *alphabet* and its elements as *letters* in this alphabet. A *word* is a finite sequence $w = i_1 i_2 \dots i_n$, where $n \in \mathbb{N}_0$ and $i_1, i_2, \dots, i_n \in J$. For $n = 0$ we interpret this as the *empty word* $\emptyset$. The number n is called the *length* of

the word $w = i_1 i_2 \ldots i_n$. The words of length n can be identified with n-tuples in J and are elements in the Cartesian power J^n. The letters, i.e., the elements in J, are precisely the words of length 1. If $w = i_1 i_2 \ldots i_n$ and $w' = j_1 j_2 \ldots j_m$, then we denote by $ww' = i_1 i_2 \ldots i_n j_1 j_2 \ldots j_m$ the word obtained by concatenating w and w'.

Let J^* be the set of all words (including the empty word) in the alphabet J. The *(left-)shift* $\Sigma \colon J^* \setminus \{\emptyset\} \to J^*$ is defined by setting $\Sigma(i_1 i_2 \ldots i_n) = i_2 \ldots i_n$ for a word $w = i_1 i_2 \ldots i_n \in J^* \setminus \{\emptyset\}$. We denote by J^ω the set of all sequences $\{i_k\} = \{i_k\}_{k \in \mathbb{N}}$ with $i_k \in J$ for $k \in \mathbb{N}$. More informally, we consider a sequence $s = \{i_k\} \in J^\omega$ as a word of infinite length and write $s = i_1 i_2 \ldots$.

If $s = \{i_k\} \in J^\omega$ and $n \in \mathbb{N}_0$, then we denote by $[s]_n \in J^*$ the word $s_n = i_1 i_2 \ldots i_n$ consisting of the first n elements of the sequence s. The *(left-)shift* $\Sigma \colon J^\omega \to J^\omega$ is the map that assigns to each sequence $\{i_k\} \in J^\omega$ the sequence $\{j_k\} \in J^\omega$ with $j_k = i_{k+1}$ for all $k \in \mathbb{N}$. In our notation we do not distinguish the shifts on $J^* \setminus \{\emptyset\}$ and J^ω and denote both maps by Σ. Note that $[\Sigma(s)]_n = \Sigma([s]_{n+1})$ for all $s \in J^\omega$; indeed, if $s = i_1 i_2 \ldots$, then we have

$$[\Sigma(s)]_n = [i_2 i_3 \ldots]_n = i_2 \ldots i_{n+1} = \Sigma(i_1 i_2 \ldots i_{n+1}) = \Sigma([s]_{n+1}).$$

If we equip J with the discrete topology, then J^ω carries a natural metrizable product topology. This topology is induced by the ultra-metric d given by $d(s, s') = 2^{-N}$ for $s = \{i_k\} \in J^\omega$ and $s' = \{j_k\} \in J^\omega$, $s \neq s'$, where $N = \min\{k \in \mathbb{N} : i_k \neq j_k\}$. In particular, two elements $s, s' \in J^\omega$ are close if and only if $s_k = s'_k$ for all $k = 1, \ldots, n$, where n is large. Equipped with this topology, the space J^ω is compact.

Suppose that X and $\widetilde{X}$ are topological spaces, and $f \colon X \to X$ and $\widetilde{f} \colon \widetilde{X} \to \widetilde{X}$ are continuous maps. We say that the dynamical system (X, f) is a *factor* of the dynamical system $(\widetilde{X}, \widetilde{f})$ if there exists a surjective continuous map $\varphi \colon \widetilde{X} \to X$ such that $\varphi \circ \widetilde{f} = f \circ \varphi$.

We are now ready for the proofs.

PROOF OF THEOREM 9.1. Let $f \colon S^2 \to S^2$ be an expanding Thurston map, and $k := \deg(f) \geq 2$. Fix a visual metric ϱ for f, and let $\Lambda > 1$ be its expansion factor. In the following, metric concepts refer to ϱ.

We fix a Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}$ and consider tiles for $(f, \mathcal{C})$. We color them black and white as in Lemma 5.21. Let $p \in S^2 \setminus \mathrm{post}(f)$ be a basepoint in the interior of the white 0-tile $X_{\mathtt{w}}^0$.

Claim 1. For all $n \in \mathbb{N}_0$ the estimate

$$\sup_{x \in S^2} \mathrm{dist}(x, f^{-n}(p)) \lesssim \Lambda^{-n}$$

is true, where $C(\lesssim)$ is independent of n.

In other words, the set $f^{-n}(p)$ forms a very dense net in S^2 if n is large. To see this, let $x \in S^2$ be arbitrary. Then x lies in some n-tile X^n. If X^n is white, then X^n contains a point in $f^{-n}(p)$ and so $\mathrm{dist}(x, f^{-n}(p)) \leq \mathrm{diam}(X^n)$. If X^n is black, then X^n shares an edge with a white n-tile Y^n. Then Y^n contains a point in $f^{-n}(p)$, and so $\mathrm{dist}(x, f^{-n}(p)) \leq \mathrm{diam}(X^n) + \mathrm{diam}(Y^n)$.

From the inequalities in both cases and Proposition 8.4 we conclude

$$\mathrm{dist}(x, f^{-n}(p)) \lesssim \Lambda^{-n},$$

where $C(\lesssim)$ is independent of x and n. Claim 1 follows.

None of the points in $S^2 \setminus \mathrm{post}(f)$ is a critical value for any of the iterates f^n of f. Moreover, each iterate f^n is a covering map $f^n \colon S^2 \setminus f^{-n}(\mathrm{post}(f)) \to S^2 \setminus \mathrm{post}(f)$. Since $p \in S^2 \setminus \mathrm{post}(f)$, we have $f^{-n}(p) \subset S^2 \setminus f^{-n}(\mathrm{post}(f))$ and

(9.1) $$\# f^{-n}(p) = \deg(f^n) = \deg(f)^n = k^n$$

for $n \in \mathbb{N}$. In particular,

$$f^{-1}(p) \subset S^2 \setminus f^{-1}(\mathrm{post}(f)) \subset S^2 \setminus \mathrm{post}(f),$$

and $\# f^{-1}(p) = k$. Let $q_1, \ldots, q_k \in S^2 \setminus \mathrm{post}(f)$ be the points in $f^{-1}(p)$. For each $i = 1, \ldots, k$ we pick a path $\alpha_i \colon [0, 1] \to S^2 \setminus \mathrm{post}(f)$ with $\alpha_i(0) = p$ and $\alpha_i(1) = q_i$.

Let $J := \{1, \ldots, k\}$, and consider the shift $\Sigma \colon J^\omega \to J^\omega$. We want to show that f is a factor of Σ, i.e., that there exists a continuous and surjective map $\varphi \colon J^\omega \to S^2$ with $f \circ \varphi = \varphi \circ \Sigma$. In order to define φ, we first construct a suitable map ψ that assigns to each word in J^* a point in S^2.

Definition of ψ: The map $\psi \colon J^* \to S^2$ will be defined inductively such that

$$\psi(w) \in f^{-n}(p),$$

whenever $n \in \mathbb{N}_0$ and $w \in J^n \subset J^*$ is a word of length n. For the empty word $\emptyset$ we set $\psi(\emptyset) = p$, and for the word consisting of the single letter $i \in J$ we set $\psi(i) := q_i \in f^{-1}(p)$.

Now suppose that ψ has been defined for all words of length $\leq n$, where $n \in \mathbb{N}$. Let w be an arbitrary word of length $n+1$. Then $w = w'i$, where $w' \in J^*$ is a word of length n and $i \in J$. So $\psi(w') \in f^{-n}(p)$ is already defined. Since $f^n(\psi(w')) = p$ and $f^n \colon S^2 \setminus f^{-n}(\mathrm{post}(f)) \to S^2 \setminus \mathrm{post}(f)$ is a covering map, the path α_i has a unique lift with initial point $\psi(w')$, i.e., there exists a unique path $\widetilde{\alpha}_i \colon [0, 1] \to S^2$ with

$\widetilde{\alpha}_i(0) = \psi(w')$ and $f^n \circ \widetilde{\alpha}_i = \alpha_i$ (see Lemma A.6). We now define $\psi(w) := \widetilde{\alpha}_i(1)$. Note that then

$$f^{n+1}(\psi(w)) = f^{n+1}(\widetilde{\alpha}_i(1)) = f(\alpha_i(1)) = f(q_i) = p.$$

Hence $\psi(w) \in f^{-(n+1)}(p)$. This shows that a map $\psi: J^* \to S^2$ with the desired properties exists.

Claim 2. $f(\psi(w)) = \psi(\Sigma(w))$ for all non-empty words $w \in J^*$.

We prove this by induction on the length of the word w. If $w = i \in J$, then

$$f(\psi(w)) = f(\psi(i)) = f(q_i) = p = \psi(\emptyset) = \psi(\Sigma(i)) = \psi(\Sigma(w)).$$

So the claim is true for words of length 1.

Suppose the claim is true for words of length $\leq n$, where $n \in \mathbb{N}$. Let w be a word of length $n+1$. Then $w = w'i$, where w' is a word of length n and $i \in J$. Let $\widetilde{\alpha}_i$ be the path as above, used in the definition of $\psi(w)$. Define $\widetilde{\beta}_i := f \circ \widetilde{\alpha}_i$. Then $\widetilde{\beta}_i$ is a lift of α_i by f^{n-1}. By induction hypothesis its initial point is

$$\widetilde{\beta}_i(0) = f(\widetilde{\alpha}_i(0)) = f(\psi(w')) = \psi(\Sigma(w')).$$

In other words, $\widetilde{\beta}_i$ is the unique path as in the definition of ψ used to determine $\psi(\Sigma(w')i)$ from $\psi(\Sigma(w'))$, and so $\psi(\Sigma(w')i) = \widetilde{\beta}_i(1)$. Hence

$$\psi(\Sigma(w)) = \psi(\Sigma(w')i) = \widetilde{\beta}_i(1) = f(\widetilde{\alpha}_i(1)) = f(\psi(w'i)) = f(\psi(w))$$

as desired, and Claim 2 follows.

Claim 3. For each $n \in \mathbb{N}$ the map $\psi|J^n: J^n \to f^{-n}(p)$ is a bijection.

In other words, the map ψ provides a coding of the points in $f^{-n}(p)$ by words of length n. Again we prove this by induction on n. By definition of ψ it is true for $n = 1$.

Suppose it is true for some $n \in \mathbb{N}$. It suffices to show that the map $\psi|J^{n+1}: J^{n+1} \to f^{-(n+1)}(p)$ is surjective, since both sets J^{n+1} and $f^{-(n+1)}(p)$ have the same cardinality k^{n+1}. So let $x \in f^{-(n+1)}(p)$ be arbitrary. Then $f^n(x) \in f^{-1}(p)$, and so there exists $i \in J$ with $f^n(x) = q_i$. Since

$$x \in f^{-(n+1)}(p) \subset S^2 \setminus f^{-(n+1)}(\mathrm{post}(f)) \subset S^2 \setminus f^{-n}(\mathrm{post}(f)),$$

and $f^n: S^2 \setminus f^{-n}(\mathrm{post}(f)) \to S^2 \setminus \mathrm{post}(f)$ is a covering map, we can lift the path α_i by f^n to a path $\widetilde{\alpha}_i: [0,1] \to S^2$ whose terminal point is x (to see this, lift α_i, traversed in opposite direction, so that the initial point of the lift is x). Then $f^n(\widetilde{\alpha}_i(0)) = \alpha_i(0) = p$, and so $\widetilde{\alpha}_i(0) \in f^{-n}(p)$. By induction hypothesis there exists a word $w' \in J^n$

with $\psi(w') = \widetilde{\alpha}_i(0)$. Then $\widetilde{\alpha}_i$ is a path as used to determine $\psi(w'i)$ from $\psi(w')$. So if we set $w := w'i \in J^{n+1}$, then

$$\psi(w) = \psi(w'i) = \widetilde{\alpha}_i(1) = x.$$

This shows that $\psi|J^{n+1}\colon J^{n+1} \to f^{-(n+1)}(p)$ is surjective. Claim 3 follows.

Claim 4. If $s \in J^\omega$, then the points $\psi([s]_n)$, $n \in \mathbb{N}$, form a Cauchy sequence in S^2 (recall that $[s]_n$ is the word consisting of the first n elements of the sequence s).

Indeed, by definition of ψ the points $\psi([s]_n)$ and $\psi([s]_{n+1})$ are joined by a lift of one of the paths $\alpha_1, \dots, \alpha_k$ by f^n. So by Lemma 8.9 we have

$$(9.2) \qquad \varrho(\psi([s]_n), \psi([s]_{n+1})) \lesssim \Lambda^{-n},$$

where $C(\lesssim)$ is independent of n and s. Hence $\{\psi([s]_n)\}$ is a Cauchy sequence, proving Claim 4.

Definition of φ: If $s \in J^\omega$, then by Claim 4 the limit

$$\varphi(s) := \lim_{n\to\infty} \psi([s]_n)$$

exists. This defines a map $\varphi\colon J^\omega \to S^2$.

Claim 5. $f \circ \varphi = \varphi \circ \Sigma$.

To see this, let $s \in J^\omega$ be arbitrary. Note that $\Sigma([s]_n) = [\Sigma(s)]_{n-1}$ for $n \in \mathbb{N}$. Hence by Claim 2 and the continuity of f we have

$$f(\varphi(s)) = \lim_{n\to\infty} f(\psi([s]_n)) = \lim_{n\to\infty} \psi(\Sigma([s]_n))$$
$$= \lim_{n\to\infty} \psi([\Sigma(s)]_{n-1}) = \varphi(\Sigma(s)).$$

Claim 5 follows.

Claim 6. The map $\varphi\colon J^\omega \to S^2$ is continuous and surjective.

Let $s \in J^\omega$ and $n \in \mathbb{N}$. Then (9.2) shows that

$$(9.3) \qquad \varrho(\varphi(s), \psi([s]_n)) \lesssim \sum_{l=n}^{\infty} \Lambda^{-l} \lesssim \Lambda^{-n},$$

where $C(\lesssim)$ is independent of n and s. Hence if $s, s' \in \Sigma$ and $[s]_n = [s']_n$, then

$$\varrho(\varphi(s), \varphi(s')) \lesssim \Lambda^{-n},$$

where $C(\lesssim)$ is independent of n, s, and s'. The continuity of φ follows from this; indeed, if s and s' are close in J^ω, then $[s]_n = [s']_n$ for some large n, and so the image points $\varphi(s)$ and $\varphi(s')$ are close in S^2.

Since J^ω is compact, the continuity of φ implies that the image $\varphi(J^\omega)$ is also compact and hence closed in S^2. The surjectivity of φ will follow, if we can show that φ has a dense image in S^2.

To see this, let $x \in S^2$ and $n \in \mathbb{N}$ be arbitrary. Then by Claim 1 we can find a point $y \in f^{-n}(p)$ with $\varrho(x,y) \lesssim \Lambda^{-n}$, where $C(\lesssim)$ is independent of x and n. Moreover, by Claim 3 there exists a word $w \in J^n$ with $\psi(w) = y$. Pick $s \in J^\omega$ such that $[s]_n = w$. Then by (9.3) we have

$$\varrho(x, \varphi(s)) \le \varrho(x,y) + \varrho(y, \varphi(s)) = \varrho(x,y) + \varrho(\psi([s]_n), \varphi(s)) \lesssim \Lambda^{-n},$$

where $C(\lesssim)$ is independent of the choices. Hence

$$\sup_{x \in S^2} \operatorname{dist}(x, \varphi(J^\omega)) \lesssim \Lambda^{-n}$$

for all n, where $C(\lesssim)$ is independent of n. This shows that $\varphi(J^\omega)$ is dense in S^2. Claim 6 follows.

The theorem now follows from Claim 5 and Claim 6.   □

The procedure that we employed to code the elements in $f^{-n}(p)$ by words of length n and the points in S^2 by infinite words is well known (see [**Ne05**, Section 5.2], for example). Note that an expanding Thurston map may have periodic critical points. Then there are points in S^2 that are coded by an uncountable number of sequences in J^ω.

PROOF OF COROLLARY 9.2. We use the notation and setup of the proof of Theorem 9.1.

It suffices to show that if $x \in S^2$ and $n \in \mathbb{N}$ are arbitrary, then there exists a point $z \in S^2$ with $f^n(z) = z$ and $\varrho(x,z) \lesssim \Lambda^{-n}$. Here and in the following, $C(\lesssim)$ is independent of x and n.

To find such a point z, we apply Claim 1 in the proof of Theorem 9.1 and conclude that there exists $y \in f^{-n}(p)$ with $\varphi(x,y) \lesssim \Lambda^{-n}$. By Claim 3 in this proof there exists a word $w \in J^*$ of length n such that $\psi(w) = y$. Let s be the unique sequence obtained by periodic repetition of the letters in w, i.e., $s \in J^\omega$ is the unique sequence with $[s]_n = w$ and $\Sigma^n(s) = s$. Put $z := \varphi(s)$. Then Claim 5 in the proof of Theorem 9.1 implies

$$f^n(z) = f^n(\varphi(s)) = \varphi(\Sigma^n(s)) = \varphi(s) = z.$$

Moreover, by (9.3) we have

$$\varrho(y,z) = \varrho(\psi(w), \varphi(s)) = \varrho(\psi([s]_n), \varphi(s)) \lesssim \Lambda^{-n},$$

and so

$$\varrho(x,z) \le \varrho(x,y) + \varrho(y,z) \lesssim \Lambda^{-n}.$$

The statement follows.   □

If in the previous argument we choose a constant sequence $s \in J^\omega$ and set $z = \varphi(s)$, then $\Sigma(s) = s$, and so

$$f(z) = f(\varphi(s)) = \varphi(\Sigma(s)) = \varphi(s) = z.$$

This shows that every expanding Thurston map has a fixed point. A more systematic investigation of fixed points and periodic points of expanding Thurston maps can be found in [**Li16a**].

CHAPTER 10

Tile graphs

An interesting feature of expanding Thurston maps is that they are linked to negatively curved spaces. Namely, if $f\colon S^2 \to S^2$ is such a map and $\mathcal{C} \subset S^2$ is a Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}$, then one can use the associated cell decompositions to define an infinite graph $\mathcal{G} = \mathcal{G}(f, \mathcal{C})$. The set of vertices of this graph is given by the collection of tiles on all levels, where it is convenient to add $X^{-1} := S^2$ as a tile of level -1 and basepoint of the graph. One connects two vertices by an edge if the corresponding tiles have non-empty intersection and their levels differ by at most 1. We will study the properties of this *tile graph* in the present chapter. The main results are based on work by Q. Yin (see [**Yi15b**]).

THEOREM 10.1. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map, and $\mathcal{C} \subset S^2$ be a Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}$. Then the associated tile graph $\mathcal{G}(f, \mathcal{C})$ is Gromov hyperbolic.*

For the boundary at infinity $\partial_\infty \mathcal{G}$ we have a natural identification $\partial_\infty \mathcal{G} = S^2$. By this identification the class of visual metrics in the sense of Thurston maps (see Chapter 8) and in the sense of Gromov hyperbolic spaces (see Section 4.2) are the same.

THEOREM 10.2. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map, $\mathcal{C} \subset S^2$ a Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}$, and $\mathcal{G} = \mathcal{G}(f, \mathcal{C})$ be the associated tile graph. Then $\partial_\infty \mathcal{G}$ can naturally be identified with S^2. Under this identification, a metric on $\partial_\infty \mathcal{G} \cong S^2$ is visual in the sense of Gromov hyperbolic spaces if and only if it is visual in the sense of expanding Thurston maps.*

Under the identification of S^2 with $\partial_\infty \mathcal{G}$ as in the previous theorem, the number $m_{f,\mathcal{C}}$ (see Definition 8.1) is the Gromov product $(x \cdot y)$ (with basepoint X^{-1}) up to some additive constant.

LEMMA 10.3. *In the setting of Theorem 10.2 there is a constant $c \geq 0$ such that*
$$m_{f,\mathcal{C}}(x, y) - c \leq (x \cdot y) \leq m_{f,\mathcal{C}}(x, y) + c,$$
for all $x, y \in S^2$.

In [**HP09**] Haïssinsky and Pilgrim also considered the sphere S^2 as the boundary at infinity of a suitable Gromov hyperbolic space very similar to the setting of Theorem 10.2.

An obvious question is how the graphs $\mathcal{G}(f,\mathcal{C})$ and $\mathcal{G}(f,\widetilde{\mathcal{C}})$ are related for different Jordan curves $\mathcal{C},\widetilde{\mathcal{C}} \subset S^2$ containing $\mathrm{post}(f)$. For a Cayley graph of a group a change of the generating set leads to quasi-isometric Cayley graphs. So in the context of expanding Thurston maps one may expect a similar result. Actually, a stronger statement is true: the graphs $\mathcal{G}(f,\mathcal{C})$ and $\mathcal{G}(f,\widetilde{\mathcal{C}})$ are even rough-isometric (see Section 4.2).

THEOREM 10.4. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map, and $\mathcal{C},\mathcal{C}' \subset S^2$ be Jordan curves with $\mathrm{post}(f) \subset \mathcal{C},\mathcal{C}'$. Then the graphs $\mathcal{G}(f,\mathcal{C})$ and $\mathcal{G}(f,\mathcal{C}')$ are rough-isometric.*

Throughout this chapter, $f\colon S^2 \to S^2$ will be an expanding Thurston map, and $\mathcal{C} \subset S^2$ a Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}$. We consider tiles in the cell decompositions $\mathcal{D}^n(f,\mathcal{C})$, $n \in \mathbb{N}_0$, and add $X^{-1} := S^2$ as a tile of level -1. Let $\mathbf{X}'$ be the collection of tiles on all levels $n \in \mathbb{N}_0 \cup \{-1\}$. In $\mathbf{X}'$ we consider tiles as different if their levels are different even if the underlying sets of the tiles are the same. If $X \in \mathbf{X}'$, we denote by

$$(10.1) \qquad\qquad \ell(X) \in \mathbb{N}_0 \cup \{-1\}$$

the level of the tile X; so X is an $\ell(X)$-tile.

As discussed above, we define the *tile graph* $\mathcal{G} = \mathcal{G}(f,\mathcal{C})$ of f with respect to $\mathcal{C}$ as follows. The set of vertices of $\mathcal{G}$ is equal to the set $\mathbf{X}'$ of all tiles. Moreover, two distinct vertices given by a k-tile X^k and an n-tile X^n are joined by an edge precisely if

$$(10.2) \qquad\qquad |k - n| \le 1 \text{ and } X^k \cap X^n \ne \emptyset.$$

So we join two vertices if the corresponding tiles intersect and their levels differ by at most 1. The graph $\mathcal{G}$ is a 1-dimensional cell complex where each edge is identified with an interval of length 1. Then the graph $\mathcal{G}$ is connected.

Indeed, each point in an edge in $\mathcal{G}$ can be joined to a vertex, and each vertex in $\mathcal{G}$, given by an n-tile X^n, can be joined to $X^{-1} = S^2 \in \mathcal{G}$ as follows. Pick a point $p \in X^n$, and for $i = 0,\dots,n-1$ let X^i be an i-tile with $p \in X^i$. Then in the vertex sequence $X^n, X^{n-1},\dots,X^{-1}$ two consecutive elements are joined by an edge, because the levels of the tiles differ by 1 and all tiles contain p and hence have non-empty intersection. So there exists a path joining X^n and X^{-1} in $\mathcal{G}$ as desired.

Since $\mathcal{G}$ is connected, this graph carries a unique path metric so that each edge is isometric to the unit interval. If X and Y are vertices in $\mathcal{G}$, i.e., tiles in $\mathbf{X}'$, we denote by $|X - Y|$ the distance of X and Y in $\mathcal{G}$, and call this quantity the *combinatorial distance* of the tiles. By definition of the metric in $\mathcal{G}$ it is clear that $|X - Y|$ is equal to the minimal number $n \in \mathbb{N}_0$ such that there exist tiles $X_0 = X, X_1, \ldots, X_n = Y$ in $\mathbf{X}'$ satisfying

$$(10.3) \qquad |\ell(X_{i-1}) - \ell(X_i)| \leq 1 \text{ and } X_{i-1} \cap X_i \neq \emptyset$$

for $i = 1, \ldots, n$.

Note that

$$|X - Y| \geq |\ell(X) - \ell(Y)|.$$

Moreover, if $X \cap Y \neq \emptyset$, then a simple argument similar to the one used to show connectedness of $\mathcal{G}$ gives that

$$|X - Y| \leq |\ell(X) - \ell(Y)| + 1.$$

We pick $X^{-1} = S^2$ as the basepoint in $\mathcal{G}$. Note that the combinatorial distance of a vertex $X \in \mathcal{G}$ to X^{-1} is

$$|X - X^{-1}| = \ell(X) + 1.$$

We denote by $(X \cdot Y)$ the Gromov product (see (4.6)) of two vertices $X, Y \in \mathcal{G}$ with respect to the basepoint X^{-1}; so

$$(10.4) \qquad (X \cdot Y) = \frac{1}{2}(|X - X^{-1}| + |Y - X^{-1}| - |X - Y|)$$

$$= 1 + \frac{1}{2}(\ell(X) + \ell(Y) - |X - Y|).$$

We want to show that the graph $\mathcal{G}$ equipped with its path metric is a Gromov hyperbolic space. Since every point in $\mathcal{G}$ has distance $\leq 1/2$ to a vertex, the set $\mathbf{X}'$ of vertices in $\mathcal{G}$ is cobounded in $\mathcal{G}$. Hence it is enough to consider the space of vertices $\mathbf{X}'$ equipped with the metric given by the combinatorial distance of vertices. The key for proving Gromov hyperbolicity of $\mathcal{G}$ is to relate the Gromov product to visual metrics for f as discussed in Chapter 8. The basic idea goes back to a similar argument in [**BP03**]. Our presentation mostly follows [**Yi15b**].

For the rest of this chapter we pick a fixed visual metric ϱ for f on S^2. Metric notions on S^2 will usually refer to ϱ unless otherwise stated. If $\Lambda > 1$ is the expansion factor of ϱ, then by Proposition 8.4

$$(10.5) \qquad \operatorname{diam}(X) \asymp \Lambda^{-\ell(X)}$$

for all $X \in \mathbf{X}'$, and

$$(10.6) \qquad \operatorname{dist}(X, Y) \gtrsim \Lambda^{-\ell(X)}$$

for all $X, Y \in \mathbf{X}'$ with $X \cap Y = \emptyset$ and $\ell(X) = \ell(Y)$. In both inequalities the implicit constants are independent of the tiles involved.

We may view tiles in two distinct ways: as vertices in the graph $\mathcal{G}$, or as subsets of the sphere S^2. The following lemma provides the key for relating these two viewpoints.

LEMMA 10.5. *For all tiles $X, Y \in \mathbf{X}' \subset \mathcal{G}$ we have*

$$\Lambda^{-(X \cdot Y)} \asymp \operatorname{diam}(X \cup Y),$$

where $C(\asymp)$ is independent of X and Y.

See [**BP03**, Lemma 2.2] for a similar statement in a different (but related) context.

PROOF. Let $X, Y \in \mathbf{X}'$ be arbitrary, and set $n = |X - Y|$. To prove the upper bound, we pick a tile chain (see Definition 5.19)

$$X_0 = X, X_1, \ldots, X_n = Y$$

satisfying (10.3) and realizing the combinatorial distance n between X and Y. Note that for the levels of the tiles in this chain we have

$$\ell(X_i) \geq \max\{\ell(X) - i, \ell(Y) - (n - i)\}$$

for $i = 0, \ldots, n$. The minimum of the right hand side occurs for

$$i = l := \lfloor (\ell(X) - \ell(Y) + n)/2 \rfloor \in [0, n].$$

Using (10.4) we have the estimates

$$l - \ell(X) \leq \frac{1}{2}(n - \ell(X) - \ell(Y)) = -(X \cdot Y) + 1 \text{ and}$$
$$n - l - 1 - \ell(Y) \leq -(X \cdot Y) + 1.$$

So by (10.5) we have

$$
\begin{aligned}
\operatorname{diam}(X \cup Y) \ &\leq \ \sum_{i=0}^{n} \operatorname{diam}(X_i) \lesssim \sum_{i=0}^{n} \Lambda^{-\ell(X_i)} \\
&\lesssim \ \sum_{i=0}^{l} \Lambda^{i-\ell(X)} + \sum_{i=l+1}^{n} \Lambda^{(n-i)-\ell(Y)} \\
&\lesssim \ \Lambda^{-(X \cdot Y)}
\end{aligned}
$$

with implicit constants independent of X and Y.

For the lower bound, let

$$-1 \leq m \leq \min\{\ell(X), \ell(Y)\}$$

be the maximal integer for which there exist m-tiles X^m and Y^m with $X \cap X^m \neq \emptyset$, $Y \cap Y^m \neq \emptyset$, and $X^m \cap Y^m \neq \emptyset$. Then

$$|X - X^m| \leq \ell(X) - m + 1,$$
$$|Y - Y^m| \leq \ell(Y) - m + 1,$$
$$|X^m - Y^m| \leq 1.$$

This implies

$$
\begin{aligned}
|X - Y| &\leq |X - X^m| + |X^m - Y^m| + |Y^m - Y| \\
&\leq \ell(X) + \ell(Y) - 2m + 3,
\end{aligned}
$$

and so

$$(X \cdot Y) = \frac{1}{2}(\ell(X) + \ell(Y) - |X - Y|) + 1 \geq m - 1/2.$$

Now if $m = \min\{\ell(X), \ell(Y)\}$, then by (10.5) we have

$$
\begin{aligned}
\operatorname{diam}(X \cup Y) &\geq \max\{\operatorname{diam}(X), \operatorname{diam}(Y)\} \\
&\gtrsim \max\{\Lambda^{-\ell(X)}, \Lambda^{-\ell(Y)}\} \\
&\asymp \Lambda^{-m} \gtrsim \Lambda^{-(X \cdot Y)}
\end{aligned}
$$

with implicit constants independent of X and Y. This gives the desired lower bound in this case.

In the other case, where $m < \min\{\ell(X), \ell(Y)\}$, we pick $(m+1)$-tiles X^{m+1} and Y^{m+1} with $X \cap X^{m+1} \neq \emptyset$ and $Y \cap Y^{m+1} \neq \emptyset$. Then we can find points $x \in X \cap X^{m+1}$ and $y \in Y \cap Y^{m+1}$. Moreover, by definition of m we have $X^{m+1} \cap Y^{m+1} = \emptyset$. Hence by (10.6) we have

$$
\begin{aligned}
\operatorname{diam}(X \cup Y) &\geq \varrho(x, y) \geq \operatorname{dist}(X^{m+1}, Y^{m+1}) \\
&\gtrsim \Lambda^{-(m+1)} \gtrsim \Lambda^{-(X \cdot Y)}
\end{aligned}
$$

with implicit constants independent of X and Y. So we get the desired lower bound also in this case. $\square$

The following corollary to the previous lemma relates sequences converging to infinity in $\mathcal{G}$ with points in the sphere S^2.

LEMMA 10.6. *Let $\{X_i\}$ be a sequence of points (i.e., tiles) in $\mathbf{X}'$. Then the following statements are true:*

(i) *$\{X_i\}$ converges to infinity in $\mathcal{G}$ if and only if there is a unique point $p \in S^2$ such that $X_i \to \{p\}$ as $i \to \infty$ in the sense of Hausdorff convergence on S^2.*

(ii) *Another sequence $\{Y_i\}$ in $\mathbf{X}'$ that converges to infinity in $\mathcal{G}$ is equivalent to $\{X_i\}$ if and only if the sequences Hausdorff converge to the same singleton set $\{p\} \subset S^2$.*

For the definition of Hausdorff convergence see the end of Section 4.1. Recall from (4.10) that a sequence $\{X_i\}$ in $\mathbf{X}'$ converges to infinity if and only if

$$(10.7) \qquad\qquad \lim_{i,j\to\infty} (X_i \cdot X_j) = \infty$$

and from (4.11) that a sequence $\{Y_i\}$ in $\mathbf{X}'$ that converges to infinity is equivalent to $\{X_i\}$ if (and only if)

$$\lim_{i\to\infty} (X_i \cdot Y_i) = \infty.$$

PROOF. Note that (10.7) is equivalent to

$$(10.8) \qquad\qquad \lim_{i,j\to\infty} \mathrm{diam}(X_i \cup X_j) = 0$$

by Lemma 10.5. This is equivalent to $\mathrm{diam}(X_i) \to 0$ as $i \to \infty$ and that $\{X_i\}$ is a Cauchy sequence with respect to Hausdorff distance on S^2. This in turn happens if and only if there exists a unique point $p \in S^2$ such that $X_i \to \{p\}$ as $i \to \infty$ in the sense of Hausdorff convergence. Thus (i) holds.

Let $\{Y_i\}$ be another sequence in $\mathbf{X}'$ that converges to infinity. From Lemma 10.5 we see that $\{Y_i\}$ is equivalent to $\{X_i\}$ if and only if $\lim_{i\to\infty} \mathrm{diam}(X_i \cup Y_i) = 0$. This happens if and only if X_i and Y_i Hausdorff converge to the same singleton $\{p\}$ as $i \to \infty$. Thus (ii) also holds. $\qquad\square$

PROOF OF THEOREM 10.1. We use notation as before. Since the set of vertices $\mathbf{X}'$ is cobounded in $\mathcal{G}$, it suffices to show that $\mathbf{X}'$ equipped with the combinatorial distance is Gromov hyperbolic.

Now if $X, Y, Z \in \mathbf{X}'$ are arbitrary, then

$$\begin{aligned} \mathrm{diam}(X \cup Y) &\leq \mathrm{diam}(X \cup Z) + \mathrm{diam}(Z \cup Y) \\ &\leq 2\max\{\mathrm{diam}(X \cup Z), \mathrm{diam}(Z \cup Y)\}. \end{aligned}$$

Invoking Lemma 10.5 and taking logarithms with base Λ in the last inequality, we obtain

$$(X \cdot Y) \geq \min\{(X \cdot Z), (Z \cdot Y)\} - \delta,$$

where $\delta \geq 0$ is a suitable constant independent of X, Y, Z. Thus $\mathcal{G}$ is Gromov hyperbolic (see (4.7)). $\qquad\square$

We are now ready to prove that our notion of visual metric on S^2 agrees with the standard one on $\partial_\infty \mathcal{G}$ under a suitable identification. Recall from Section 4.2 that $\partial_\infty \mathcal{G}$ is defined to be the set of all equivalence classes of sequences in $\mathcal{G}$ converging to infinity.

PROOF OF THEOREM 10.2. The identification will be given by a bijection between $\partial_\infty \mathcal{G}$ and S^2. Since $\mathbf{X}'$ is cobounded in $\mathcal{G}$, every point $x \in \partial_\infty \mathcal{G}$ can be represented by a sequence of tiles $\{X_i\}$ in $\mathbf{X}'$ converging to infinity. By Lemma 10.6 (i) there is a unique point $p \in S^2$ such that $X_i \to \{p\}$ as $i \to \infty$ in the sense of Hausdorff convergence. Any two sequences $\{X_i\}, \{Y_i\}$ representing x converge to the same singleton $\{p\}$ by Lemma 10.6 (ii). Thus the map

$$\varphi \colon \partial_\infty \mathcal{G} \to S^2 \ \text{ given by } \ \varphi(x) \coloneqq p$$

is well-defined.

The map φ is surjective. Indeed, let $p \in S^2$ be arbitrary. For each $i \in \mathbb{N}$ we pick an i-tile $X_i \in \mathbf{X}'$ such that $p \in X_i$. Since f is expanding, we have $\operatorname{diam}(X_i) \to 0$ as $i \to \infty$. Thus $X_i \to \{p\}$ as $i \to \infty$ (in the sense of Hausdorff convergence on S^2). By Lemma 10.6 (i) the sequence $\{X_i\}$ converges to infinity and the point $x \in \partial_\infty \mathcal{G}$ represented by $\{X_i\}$ is mapped to p by φ.

To show that φ is injective, consider two points $x, y \in \partial_\infty \mathcal{G}$ that are represented by sequences $\{X_i\}$ and $\{Y_i\}$ in $\mathbf{X}'$ converging to infinity. By Lemma 10.6 (ii) they converge to the same singleton set if and only if they are equivalent. Thus $\varphi(x) = \varphi(y)$ if and only if $x = y$. Thus φ is injective.

Having proved that φ is bijective, we ignore the original distinction between points in $\partial_\infty \mathcal{G}$ and in S^2, and identify $\partial_\infty \mathcal{G} \cong S^2$ by the map φ.

To prove the second part of the statement, we first consider the visual metric ϱ (in the sense of Thurston maps) for f fixed earlier. Let x and y be arbitrary points in S^2 and let $\{X_i\}$ and $\{Y_i\}$ be two sequences in $\mathbf{X}'$ representing them in $\partial_\infty \mathcal{G}$, respectively. As we have seen, this means $X_i \to \{x\}$ and $Y_i \to \{y\}$ as $i \to \infty$ in the sense of Hausdorff convergence on S^2. Thus

$$\operatorname{diam}(X_i \cup Y_i) \to \varrho(x, y)$$

as $i \to \infty$.

Recall the definition of the Gromov product $(x \cdot y)$ for $x, y \in \partial_\infty \mathcal{G} \cong S^2$ as given in (4.12). Here we choose $X^{-1} = S^2$ as the basepoint in $\mathcal{G}$. By (4.13) there exists a constant $k \geq 0$ independent of x and y, and of the choice of the sequences $\{X_i\}$ and $\{Y_i\}$ such that

$$\liminf_{i \to \infty} (X_i \cdot Y_i) - k \leq (x \cdot y) \leq \liminf_{i \to \infty} (X_i \cdot Y_i).$$

If we combine the previous two estimates with Lemma 10.5, then we conclude that $\varrho(x, y) \asymp \Lambda^{-(x \cdot y)}$.

Since ϱ is a visual metric for f, by Proposition 8.3 (iii) we have $\varrho(x, y) \asymp \Lambda^{-m(x,y)}$, where $m = m_{f,\mathcal{C}}$ is as in Definition 8.1. It follows that

$$(10.9) \qquad m(x, y) - c \leq (x \cdot y) \leq m(x, y) + c,$$

where $c \geq 0$ is independent of x and y.

In the definition of visual metrics in the sense of Gromov hyperbolic spaces (4.14) we may choose any basepoint for the Gromov product (up to an adjustment of the multiplicative constant). Similarly, by Proposition 8.3 (iii) it is not a restriction to use our given curve $\mathcal{C}$ in Definition 8.2 of visual metrics in the sense of expanding Thurston maps.

Hence (10.9) shows that any metric $\widetilde{\varrho}$ on $\partial_\infty \mathcal{G} \cong S^2$ is a visual metric in the sense of Gromov hyperbolic spaces if and only if $\widetilde{\varrho}$ is a visual metric in the sense of expanding Thurston maps. $\qquad\square$

Note that in (10.9) we have proved Lemma 10.3.

PROOF OF THEOREM 10.4. It suffices to find a rough-isometry between the set of vertices in $\mathcal{G} = \mathcal{G}(f, \mathcal{C})$ and $\widetilde{\mathcal{G}} = \mathcal{G}(f, \widetilde{\mathcal{C}})$. As before, we denote the set of tiles for $(f, \mathcal{C})$ by $\mathbf{X}'$, and use the notation $\widetilde{\mathbf{X}}'$ for the set of tiles for $(f, \widetilde{\mathcal{C}})$ (including $\widetilde{X}^{-1} = X^{-1} = S^2$).

For each tile $X \in \mathbf{X}'$ we pick a tile $\widetilde{X} \in \widetilde{\mathcal{C}}$ of the same level (i.e., $\ell(X) = \ell(\widetilde{X})$) with $X \cap \widetilde{X} \neq \emptyset$. This assignment $X \mapsto \widetilde{X}$ gives a level-preserving map $\psi \colon \mathbf{X}' \to \widetilde{\mathbf{X}}'$. We claim that ψ is a rough-isometry between $\mathbf{X}'$ and $\widetilde{\mathbf{X}}'$, where the spaces are equipped with their respective combinatorial distances.

To see this, let $X, Y \in \mathbf{X}'$ be arbitrary, and consider $\widetilde{X} := \psi(X)$ and $\widetilde{Y} := \psi(Y)$. Since ψ preserves levels of tiles, we have

$$\mathrm{diam}(X) \asymp \Lambda^{-\ell(X)} = \Lambda^{-\ell(\widetilde{X})} \asymp \mathrm{diam}(\widetilde{X}),$$

and similarly $\mathrm{diam}(Y) \asymp \mathrm{diam}(\widetilde{Y})$. Hence

$$\begin{aligned} \mathrm{diam}(X \cup Y) &\leq \mathrm{diam}(X) + \mathrm{diam}(\widetilde{X} \cup \widetilde{Y}) + \mathrm{diam}(Y) \\ &\lesssim \mathrm{diam}(\widetilde{X} \cup \widetilde{Y}), \end{aligned}$$

and the same argument gives $\mathrm{diam}(\widetilde{X} \cup \widetilde{Y}) \lesssim \mathrm{diam}(X \cup Y)$. In all the previous relations the implicit multiplicative constants are independent of X and Y. Lemma 10.5 implies that there exists a constant $c \geq 0$ independent of X and Y such that

$$(X \cdot Y) - c \leq (\widetilde{X} \cdot \widetilde{Y}) \leq (X \cdot Y) + c.$$

Here Gromov products are in $\mathbf{X}'$ and $\widetilde{\mathbf{X}}'$, respectively, with respect to the basepoint $X^{-1} = \widetilde{X}^{-1} = S^2$. Since $\ell(X) = \ell(\widetilde{X})$ and $\ell(Y) = \ell(\widetilde{Y})$, based on (10.4) we deduce the inequality

$$|X - Y| - k \leq |\widetilde{X} - \widetilde{Y}| \leq |X - Y| + k$$

for combinatorial distances. Here $k := 2c$ is independent of X and Y.

This is the first condition (4.8) (with $\lambda = 1$) for ψ to be a rough-isometry. It remains to show that $\psi(\mathbf{X}')$ is cobounded in $\widetilde{\mathbf{X}}'$. To verify this, let $\widetilde{Y} \in \widetilde{\mathbf{X}}'$ be arbitrary. Pick a tile $X \in \mathbf{X}'$ with $\ell(X) = \ell(\widetilde{Y})$ and $X \cap \widetilde{Y} \neq \emptyset$. Define $\widetilde{X} := \psi(X)$. It suffices to produce a uniform upper bound for the combinatorial distance $|\widetilde{X} - \widetilde{Y}|$ of $\widetilde{X}$ and $\widetilde{Y}$ in $\widetilde{\mathbf{X}}'$ independent of $\widetilde{Y}$. Now

$$\mathrm{diam}(X) \asymp \mathrm{diam}(\widetilde{X}) \asymp \mathrm{diam}(\widetilde{Y}) \asymp \Lambda^{-\ell(\widetilde{Y})},$$

and so

$$\mathrm{diam}(\widetilde{X} \cup \widetilde{Y}) \leq \mathrm{diam}(\widetilde{X}) + \mathrm{diam}(X) + \mathrm{diam}(\widetilde{Y}) \lesssim \Lambda^{-\ell(\widetilde{Y})},$$

where again all implicit multiplicative constants are independent of the choice of the tiles. So by Lemma 10.5 we have

$$(\widetilde{X} \cdot \widetilde{Y}) \geq \ell(\widetilde{Y}) - c',$$

where $c' \geq 0$ is independent of the choices. Since $\ell(\widetilde{X}) = \ell(\widetilde{Y})$, we conclude

$$|\widetilde{X} - \widetilde{Y}| = 2 + 2\ell(\tilde{Y}) - 2(\widetilde{X} \cdot \widetilde{Y}) \leq k' := 2 + 2c',$$

which gives the desired uniform bound. $\square$

REMARK 10.7. The rough-isometry ψ between the graphs $\mathcal{G}$ and $\widetilde{\mathcal{G}}$ constructed in the previous proof is compatible with the identifications $\partial_\infty \mathcal{G} \cong S^2$ and $\partial_\infty \widetilde{\mathcal{G}} \cong S^2$. Indeed, let $p \in S^2$ be arbitrary. Viewed as an element in $\partial_\infty \mathcal{G}$, the point p is represented by a sequence $\{X_i\}$ in $\mathbf{X}'$ such that $X_i \to \{p\}$ as $i \to \infty$ in the sense of Hausdorff convergence (see Lemma 10.6). If $\widetilde{X}_i := \psi(X_i)$ for $i \in \mathbb{N}$, then $\{\widetilde{X}_i\}$ is a sequence of tiles in $\widetilde{\mathbf{X}}'$. By definition of ψ the levels of X_i and $\widetilde{X}_i$ are the same, and so $\mathrm{diam}(\widetilde{X}_i) \to 0$ as $i \to \infty$. In addition, $X_i \cap \widetilde{X}_i \neq \emptyset$ for $i \in \mathbb{N}$. This implies that $\widetilde{X}_i \to \{p\}$ as $i \to \infty$. So if $\{X_i\}$ represents the point $p \in S^2$ under the identification $\partial_\infty \mathcal{G} \cong S^2$, then the image sequence $\{\widetilde{X}_i\}$ under ψ also represents the point p under the identification $\partial_\infty \widetilde{\mathcal{G}} \cong S^2$.

CHAPTER 11

Isotopies

In this chapter we consider various questions related to isotopies. We first revisit the notion of Thurston equivalence, which is defined in terms of certain isotopies. Then we investigate when two Jordan curves in S^2 are isotopic relative to a finite set of points. This is in preparation for results about the existence of invariant Jordan curves for expanding Thurston maps (see Chapter 15).

Recall that two Thurston maps $f\colon S^2 \to S^2$ and $g\colon \widehat{S}^2 \to \widehat{S}^2$ on 2-spheres S^2 and $\widehat{S}^2$ are (Thurston) equivalent (see Definition 2.4) if there exist homeomorphisms $h_0, h_1\colon S^2 \to \widehat{S}^2$ that are isotopic rel. $\mathrm{post}(f)$ and satisfy that $h_0 \circ f = g \circ h_1$. We then have the commutative diagram:

$$(11.1) \qquad\qquad \begin{array}{ccc} S^2 & \xrightarrow{\ h_1\ } & \widehat{S}^2 \\ {\scriptstyle f}\big\downarrow & & \big\downarrow{\scriptstyle g} \\ S^2 & \xrightarrow{\ h_0\ } & \widehat{S}^2. \end{array}$$

The maps f and g are topologically conjugate if there exists a homeomorphism $h\colon S^2 \to \widehat{S}^2$ such that $h \circ f = g \circ h$.

Obviously, Thurston equivalence is a weaker notion than topological conjugacy. However, two expanding Thurston maps f and g are Thurston equivalent if and only if they are topologically conjugate.

THEOREM 11.1 (Thurston equivalence and topological conjugacy). *Let $f\colon S^2 \to S^2$ and $g\colon \widehat{S}^2 \to \widehat{S}^2$ be expanding Thurston maps that are Thurston equivalent. Then they are topologically conjugate.*

More precisely, if we have a Thurston equivalence between f and g as in (11.1), then there exists a homeomorphism $h\colon S^2 \to \widehat{S}^2$ such that h is isotopic to h_1 rel. $f^{-1}(\mathrm{post}(f))$ and satisfies $h \circ f = g \circ h$.

Since $\mathrm{post}(f) \subset f^{-1}(\mathrm{post}(f))$ and h_0 and h_1 are isotopic rel. $\mathrm{post}(f)$ this implies that h is also isotopic to h_0 rel. $\mathrm{post}(f)$.

A statement very similar to the theorem above was proved by Kameyama [**Ka03**]. Since his notion of "expanding" is different from ours, we will present the details of the proof.

Theorem 11.1 will be shown in Section 11.1. To this end, we will repeatedly lift the isotopy between h_0 and h_1 in (11.1). The relevant result about the existence of such lifts is established in Proposition 11.3. Since the maps f and g in the above statement are expanding, the "tracks" of the lifted isotopies under the n-th iterates will shrink exponentially with respect to a visual metric as $n \to \infty$ (see Lemma 11.4). We will obtain the desired conjugacy between f and g essentially by concatenating these lifts. The basic idea of this argument is well known in dynamics (see [**Sh69**], for example).

In Section 11.2 and Section 11.3 we present some technical results on isotopies of Jordan curves. The most important result here is Lemma 11.17, which gives a criterion when a Jordan curve can be isotoped into the 1-skeleton of a given cell decomposition of S^2. This will be a crucial ingredient in the proof of Theorem 15.1.

Before we go into the details, we first fix some notation and terminology related to homotopies and isotopies that will be used throughout this chapter (for the basic definitions see Section 2.4). We denote by $I := [0, 1]$ the unit interval. If X and Y are topological spaces, and $H \colon X \times I \to Y$ is a homotopy between X and Y, then, as usual, $H_t := H(\cdot, t) \colon X \to Y$ for $t \in I$ denotes the time-t map of the homotopy.

Conversely, when say that a family H_t (where it is understood that $t \in I$) of continuous maps from X into Y is a homotopy between X and Y, we consider t as a variable in I and mean that the map $(x, t) \in X \times I \mapsto H_t(x)$ is a homotopy. This is a slightly imprecise, but convenient way of expression. Such a family H_t is an isotopy between X and Y if each map H_t is a homeomorphism between X and Y.

11.1. Equivalent expanding Thurston maps are conjugate

In preparation for the proof of Theorem 11.1, we first record a simple lemma about preimages of sets.

LEMMA 11.2. *Let $f \colon X \to X$ and $g \colon Y \to Y$ be maps defined on some sets X and Y, and $h, \widetilde{h} \colon X \to Y$ be bijections with $g \circ \widetilde{h} = h \circ f$. Then for every set $A \subset X$ we have*

$$g^{-1}(h(A)) = \widetilde{h}(f^{-1}(A)).$$

PROOF. Since $h, \widetilde{h}$ are bijections, $g \circ \widetilde{h} = h \circ f$ implies $h^{-1} \circ g = f \circ \widetilde{h}^{-1}$. Thus

$$g^{-1}(h(A)) = \left(h^{-1} \circ g\right)^{-1}(A) = \left(f \circ \widetilde{h}^{-1}\right)^{-1}(A) = \widetilde{h}(f^{-1}(A)),$$

as desired.   □

We now turn to lifts of isotopies by Thurston maps (see [**Ka03**, Lemma 4.3] for a similar statement).

PROPOSITION 11.3 (Lifts of isotopies by Thurston maps). *Suppose* $f\colon S^2 \to S^2$ *and* $g\colon \widehat{S}^2 \to \widehat{S}^2$ *are Thurston maps, and* $h_0, \widetilde{h}_0\colon S^2 \to \widehat{S}^2$ *are homeomorphisms such that* $h_0|\operatorname{post}(f) = \widetilde{h}_0|\operatorname{post}(f)$ *and* $h_0 \circ f = g \circ \widetilde{h}_0$. *Let* $H\colon S^2 \times I \to \widehat{S}^2$ *be an isotopy rel.* $\operatorname{post}(f)$ *with* $H_0 = h_0$.

Then the isotopy H *uniquely lifts to an isotopy* $\widetilde{H}\colon S^2 \times I \to \widehat{S}^2$ *rel.* $f^{-1}(\operatorname{post}(f))$ *such that* $\widetilde{H}_0 = \widetilde{h}_0$ *and* $g \circ \widetilde{H}_t = H_t \circ f$ *for all* $t \in I$.

So if we set $h_1 := H_1$ and $\widetilde{h}_1 := \widetilde{H}_1$, then we obtain the following commutative diagram:

$$
\begin{array}{ccc}
S^2 & \xrightarrow{\ \widetilde{H}\colon \widetilde{h}_0 \simeq \widetilde{h}_1\ } & \widehat{S}^2 \\
{\scriptstyle f}\Big\downarrow & & \Big\downarrow{\scriptstyle g} \\
S^2 & \xrightarrow{\ H\colon h_0 \simeq h_1\ } & \widehat{S}^2.
\end{array}
$$

PROOF. We have

(11.2) $$h_0(\operatorname{post}(f)) = \widetilde{h}_0(\operatorname{post}(f)) = \operatorname{post}(g)$$

as follows from the remark after Lemma 2.5. This implies that

$$H_t(\operatorname{post}(f)) = \operatorname{post}(g)$$

for all $t \in I$. Hence $H_t|S^2 \setminus \operatorname{post}(f)$ is an isotopy between $S^2 \setminus \operatorname{post}(f)$ and $\widehat{S}^2 \setminus \operatorname{post}(g)$.

Moreover, it follows from Lemma 11.2 and (11.2) that

$$\widetilde{h}_0(f^{-1}(\operatorname{post}(f))) = g^{-1}(h_0(\operatorname{post}(f))) = g^{-1}(\operatorname{post}(g)).$$

So the map $\widetilde{h}_0|S^2 \setminus f^{-1}(\operatorname{post}(f))$ can be considered as a lift of

$$(H_0 \circ f)|S^2 \setminus f^{-1}(\operatorname{post}(f)) = (h_0 \circ f)|S^2 \setminus f^{-1}(\operatorname{post}(f))$$

by the (non-branched) covering map

$$g\colon \widehat{S}^2 \setminus g^{-1}(\operatorname{post}(g)) \to \widehat{S}^2 \setminus \operatorname{post}(g).$$

By the usual homotopy lifting theorem for covering maps (see [**Ha02**, Proposition 1.30, p. 60]) the homotopy $(H_t \circ f)|S^2 \setminus f^{-1}(\operatorname{post}(f))$ lifts to a unique homotopy $\widetilde{H}\colon (S^2 \setminus f^{-1}(\operatorname{post}(f))) \times I \to \widehat{S}^2 \setminus g^{-1}(\operatorname{post}(g))$ such that

$$\widetilde{H}_0 = \widetilde{h}_0|S^2 \setminus f^{-1}(\operatorname{post}(f))$$

and $g \circ \widetilde{H}_t = H_t \circ f$ on $S^2 \setminus f^{-1}(\operatorname{post}(f))$ for all $t \in I$.

We claim that $\widetilde{H}$ has a unique extension to a homotopy between S^2 and $\widehat{S}^2$. To see this, let $q \in f^{-1}(\text{post}(f))$ be arbitrary, and set $p \coloneqq f(q) \in \text{post}(f)$. Then there exists $\widehat{p} \in \text{post}(g)$ such that $H_t(p) = \widehat{p}$ for all $t \in I$. Since g is a branched covering map, we can find a small topological disk $\widehat{V} \subset \widehat{S}^2$ containing $\widehat{p}$ such that each of the components of $g^{-1}(\widehat{V})$ contains precisely one point in $g^{-1}(\widehat{p})$. Since $H(\{p\} \times I) = \{\widehat{p}\}$ and H is uniformly continuous on $S^2 \times I$, we can choose a small neighborhood $V \subset S^2$ of p such that $H(V \times I) \subset \widehat{V}$. Finally, we can find a small topological disk $U \subset S^2$ containing q such that $f(U) \subset V$ and

$$U' \coloneqq U \setminus \{q\} \subset S^2 \setminus f^{-1}(\text{post}(f)).$$

Then the set $U' \times I \subset (S^2 \setminus f^{-1}(\text{post}(f))) \times I$ is connected; so $\widetilde{H}(U' \times I)$ is also connected. Moreover,

$$g(\widetilde{H}(U' \times I)) = \{g(\widetilde{H}_t(u)) : u \in U', \, t \in I\}$$
$$= \{H_t(f(u)) : u \in U', \, t \in I\} \subset H(V \times I) \subset \widehat{V}.$$

Hence the connected set $\widetilde{H}(U' \times I)$ is contained in a unique component $\widehat{U}$ of $g^{-1}(\widehat{V})$. By choice of $\widehat{V}$, this component contains a unique point $\widehat{q} \in g^{-1}(\widehat{p})$. By making $\widehat{V}$ smaller if necessary, we can guarantee that the corresponding component $\widehat{U}$ of $g^{-1}(\widehat{V})$ containing $\widehat{q}$ lies in an arbitrarily small neighborhood of $\widehat{q}$ (this easily follows from the fact that g is a branched covering map). Since $\widetilde{H}(U' \times I) \subset \widehat{U}$ as we have just seen, this implies that we can continuously extend $\widetilde{H}$ to $\{q\} \times I$ by setting $\widetilde{H}(q, t) = \widehat{q}$ for $t \in I$. Since q was an arbitrary element of the finite set $f^{-1}(\text{post}(f))$, we see that $\widetilde{H}$ has indeed an extension to a homotopy between S^2 and $\widehat{S}^2$, also called $\widetilde{H}$. This extension is unique, because $S^2 \setminus f^{-1}(\text{post}(f))$ is dense in S^2. The previous argument also shows that the extension $\widetilde{H}$ is a homotopy rel. $f^{-1}(\text{post}(f))$. Moreover, again by density of $S^2 \setminus f^{-1}(\text{post}(f))$ in S^2 it is clear that on S^2 we have $\widetilde{H}_0 = \widetilde{h}_0$ and $g \circ \widetilde{H}_t = H_t \circ f$ for $t \in I$. We conclude that the isotopy H can be lifted to a unique *homotopy* $\widetilde{H}$ with the desired properties.

To show that $\widetilde{H}$ is actually an *isotopy* between S^2 and $\widehat{S}^2$, we first note that the roles of f and g in the previous argument can be reversed. So by lifting the isotopy H_t^{-1}, we can find a unique homotopy $\widetilde{K}_t$ between $\widehat{S}^2$ and S^2 such that $\widetilde{K}_0 = \widetilde{h}_0^{-1}$ and $f \circ \widetilde{K}_t = H_t^{-1} \circ g$ for $t \in I$. Then $\widetilde{K}_0 \circ \widetilde{H}_0 = \widetilde{h}_0^{-1} \circ \widetilde{h}_0 = \text{id}_{S^2}$ and

$$f \circ \widetilde{K}_t \circ \widetilde{H}_t = H_t^{-1} \circ g \circ \widetilde{H}_t = H_t^{-1} \circ H_t \circ f = f.$$

This implies that for each $p \in S^2$ the (continuous) path $t \in I \mapsto \widetilde{K}_t(\widetilde{H}_t(p))$ starts at p for $t = 0$ and is contained in the finite set $f^{-1}(f(p))$. Hence $\widetilde{K}_t(\widetilde{H}_t(p)) = p$ for all $t \in I$ and $p \in S^2$, or equivalently, $\widetilde{K}_t \circ \widetilde{H}_t = \mathrm{id}_{S^2}$ for $t \in I$. A similar argument shows that $\widetilde{H}_t \circ \widetilde{K}_t = \mathrm{id}_{\widehat{S}^2}$ for $t \in I$. It follows that for each $t \in I$, the map $\widetilde{H}_t$ is a homeomorphism from S^2 onto $\widehat{S}^2$ with the inverse $\widetilde{K}_t$. So $\widetilde{H}_t$ is indeed the unique isotopy with the desired properties. $\qquad\square$

Note that if in the previous proposition H is an isotopy relative to a set $M \subset S^2$ with $\mathrm{post}(f) \subset M$, then the lift $\widetilde{H}$ is an isotopy rel. $f^{-1}(M)$. Indeed, if $p \in f^{-1}(M)$, then $f(p) \in M$ and so

$$g(\widetilde{H}_t(p)) = H_t(f(p)) = h_0(f(p)) =: \widehat{q}$$

for all $t \in I$. Thus $t \mapsto \widetilde{H}_t(p)$ is a path contained in the finite set $g^{-1}(\widehat{q})$ and hence a constant path.

The expansiveness of the Thurston maps implies the shrinking of repeated lifts. This will be made precise in the following lemma, which will be of crucial importance in the proof of Theorem 11.1.

LEMMA 11.4 (Exponential shrinking of tracks of isotopies). *Suppose* $f \colon S^2 \to S^2$ *and* $g \colon \widehat{S}^2 \to \widehat{S}^2$ *are Thurston maps, and* $H^n \colon S^2 \times I \to \widehat{S}^2$ *are isotopies rel.* $\mathrm{post}(f)$ *satisfying* $g \circ H_t^{n+1} = H_t^n \circ f$ *for* $n \in \mathbb{N}_0$ *and* $t \in I$.

If g *is expanding and* $\widehat{S}^2$ *is equipped with a visual metric for* g, *then the tracks of the isotopies* H^n *shrink exponentially as* $n \to \infty$. *More precisely, if* ϱ *is a visual metric for* g *with expansion factor* $\Lambda > 1$, *then there exists a constant* $C \geq 1$ *such that*

$$\text{(11.3)} \qquad \sup_{x \in S^2} \mathrm{diam}_\varrho(\{H_t^n(x) : t \in I\}) \leq C\Lambda^{-n}$$

for all $n \in \mathbb{N}_0$.

PROOF. For all $n \in \mathbb{N}_0$ and $t \in I$ we have $g^n \circ H_t^n = H_t^0 \circ f^n$; so for fixed $x \in S^2$ and $n \in \mathbb{N}_0$ the path $t \mapsto H_t^n(x)$ in $\widehat{S}^2$ is a lift of the path $t \mapsto H_t^0(f^n(x))$ by the map g^n. Recall that in the proof of Lemma 8.9 we had to break up the path γ into N pieces γ_j so that $\mathrm{diam}(\gamma_j) < \delta_0$ (see also (5.13)). Since H^0 is uniformly continuous, we can choose the number N uniformly for all the paths $t \mapsto H_t^0(y)$, $y \in S^2$. Since g is expanding, Lemma 8.9 then implies that

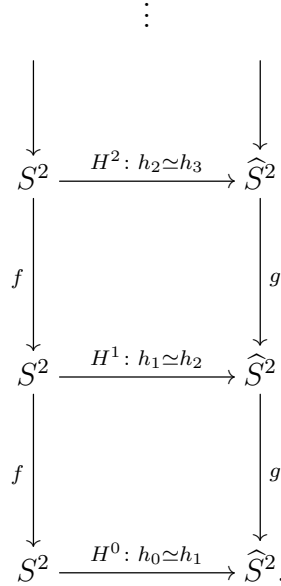$$\sup_{x \in S^2} \mathrm{diam}_\varrho(\{H_t^n(x) : t \in I\}) \lesssim \Lambda^{-n}$$

for all $n \in \mathbb{N}$, where $C(\lesssim)$ is independent of n. $\qquad\square$

After these preparations, we are ready for the proof of the main result in this section.

PROOF OF THEOREM 11.1. Let $f\colon S^2 \to S^2$ and $g\colon \widehat{S}^2 \to \widehat{S}^2$ be two expanding Thurston maps that are equivalent. We want to prove that they are in fact topologically conjugate. The main idea of the proof is to lift a suitable initial isotopy repeatedly and use the fact that by Lemma 11.4 the tracks of the isotopies shrink exponentially fast. The desired conjugacy is then obtained as a limit.

By assumption there exists an isotopy H_t^0 between S^2 and $\widehat{S}^2$ rel. $\mathrm{post}(f)$ such that $h_0 \circ f = g \circ h_1$, where $h_0 = H_0^0$ and $h_1 = H_1^0$. By Proposition 11.3 we can lift the isotopy H_t^0 between h_0 and h_1 to an isotopy H_t^1 rel. $f^{-1}(\mathrm{post}(f)) \supset \mathrm{post}(f)$ between h_1 and $h_2 := H_1^1$. Note that the map h_1 plays two roles here: it is the endpoint H_1^0 of the initial isotopy H_t^0, and also a lift of h_0.

Repeating this argument, we get homeomorphisms h_n and isotopies H_t^n between S^2 and $\widehat{S}^2$ rel. $f^{-n}(\mathrm{post}(f)) \supset \mathrm{post}(f)$ for all $n \in \mathbb{N}_0$ such $H_t^n \circ f = g \circ H_t^{n+1}$, $H_0^n = h_n$ and $H_1^n = h_{n+1}$ for all $n \in \mathbb{N}_0$ and $t \in I$. So we have the following "infinite tower" of isotopies:

$$
\vdots
$$

$$
\begin{array}{ccc}
S^2 & \xrightarrow{\;H^2\colon\, h_2 \simeq h_3\;} & \widehat{S}^2 \\[2pt]
{\scriptstyle f}\big\downarrow & & \big\downarrow{\scriptstyle g} \\[2pt]
S^2 & \xrightarrow{\;H^1\colon\, h_1 \simeq h_2\;} & \widehat{S}^2 \\[2pt]
{\scriptstyle f}\big\downarrow & & \big\downarrow{\scriptstyle g} \\[2pt]
S^2 & \xrightarrow{\;H^0\colon\, h_0 \simeq h_1\;} & \widehat{S}^2.
\end{array}
$$

We want to show that for $n \to \infty$ the maps h_n converge to a homeomorphism h_∞ that gives the desired topological conjugacy between f and g.

To see this, fix a visual metric ϱ on $\widehat{S}^2$, and assume that it has the expansion factor $\Lambda > 1$. Metric concepts on $\widehat{S}^2$ will refer to this metric

in the following. Since g is expanding, Lemma 11.4 implies that

$$(11.4) \qquad \sup_{x \in S^2} \operatorname{diam}(\{H_t^n(x) : t \in I\}) \lesssim \Lambda^{-n}$$

for all $n \in \mathbb{N}$, where $C(\lesssim)$ is independent of n. In particular,

$$\operatorname{dist}(h_{n+1}, h_n) := \sup_{x \in S^2} \operatorname{dist}(h_n(x), h_{n+1}(x)) \lesssim \Lambda^{-n}$$

for all $n \in \mathbb{N}_0$, and so there is a continuous map $h_\infty \colon S^2 \to \widehat{S}^2$ such that $h_n \to h_\infty$ uniformly on S^2 as $n \to \infty$. Since $h_{n-1} \circ f = g \circ h_n$, we have $h_\infty \circ f = g \circ h_\infty$.

The map h_∞ is a homeomorphism. To prove this, we repeat the argument where we interchange the roles of f and g. More precisely, we consider the isotopy $(H_t^0)^{-1}$ between h_0^{-1} and h_1^{-1}. The corresponding tower of repeated lifts of this initial isotopy is given by the isotopies $(H_t^n)^{-1}$ between h_n^{-1} and h_{n+1}^{-1}. By the argument in the first part of the proof we see that the maps h_n^{-1} converge to a continuous map $k_\infty \colon \widehat{S}^2 \to S^2$ uniformly on $\widehat{S}^2$ as $n \to \infty$. By uniform convergence we have $k_\infty \circ h_\infty(x) = \lim_{n \to \infty} h_n^{-1} \circ h_n(x) = x$ for all $x \in S^2$. Hence $k_\infty \circ h_\infty = \operatorname{id}_{S^2}$. Similarly, $h_\infty \circ k_\infty = \operatorname{id}_{\widehat{S}^2}$, and so k_∞ is a continuous inverse of h_∞. Hence h_∞ is a homeomorphism.

The conjugating map $h = h_\infty$ is isotopic to h_1 rel. $f^{-1}(\operatorname{post}(f))$. To see this, we will define an isotopy rel. $f^{-1}(\operatorname{post}(f))$ that is obtained by concatenating (with suitable time change) the isotopies $H^1, H^2, \ldots$ and take $h = h_\infty$ as the endpoint at time $t = 1$. The precise definition is as follows. We break up the unit interval into intervals

$$I = [0, 1] = \left[0, \tfrac{1}{2}\right] \cup \left[\tfrac{1}{2}, \tfrac{3}{4}\right] \cup \cdots \cup \left[1 - 2^{-n}, 1 - 2^{-n-1}\right] \cup \cdots \cup \{1\}.$$

The n-th interval in this union is denoted by $I^n = [1 - 2^{-n}, 1 - 2^{-n-1}]$. Let $s_n \colon I^n \to I$, $s_n(t) = 2^{n+1}(t - (1 - 2^{-n}))$, for $n \in \mathbb{N}_0$. We define $H \colon S^2 \times I \to \widehat{S}^2$ by

$$H(x, t) := H^{n+1}(x, s_n(t))$$

if $x \in S^2$ and $t \in I^n$ for some $n \in \mathbb{N}_0$, and $H(x, t) = h(x)$ for $x \in S^2$ and $t = 1$. We claim that H is indeed an isotopy between h_1 and h rel. $f^{-1}(\operatorname{post}(f))$.

Note that H is well-defined, $H_1 = h$, and $H_{1-1/2^n} = h_{n+1}$ for $n \in \mathbb{N}_0$. Moreover, H_t is a homeomorphism for each $t \in I$, and $H_t|f^{-1}(\operatorname{post}(f))$ does not depend on t. To establish our claim, it remains to verify that H is continuous. It is clear that H is continuous at each point $(x, t) \in S^2 \times [0, 1)$.

Moreover, as follows from the uniform convergence $h_n \to h$ as $n \to \infty$ and inequality (11.4), we have $H_t \to H_1$ uniformly on S^2 as $t \to 1$. This together with the continuity of $h = H_1$ implies the continuity of H at points $(x, t) \in S^2 \times I$ with $t = 1$. □

REMARK 11.5. The previous proof gives a procedure for approximating the conjugating map $h = h_\infty$. Indeed, as follows from the remark after the proof of Proposition 11.3, the map H_t^n is constant in t on $f^{-n}(\text{post}(f))$ for all $n \in \mathbb{N}_0$. This implies that $h_n = h_{n+1} = \cdots = h_\infty$ on the set $f^{-n}(\text{post}(f))$, and so the map h_n sends the points in $f^{-n}(\text{post}(f))$ to the "right" points in $g^{-n}(\text{post}(g))$. The isotopy H_t^n then deforms h_n to a map h_{n+1} such that the points in $f^{-(n+1)}(\text{post}(f))$ have the correct images in $g^{-(n+1)}(\text{post}(g))$ as well, etc. Since by expansion the union of the sets

$$\text{post}(f) \subset f^{-1}(\text{post}(f)) \subset f^{-2}(\text{post}(f)) \subset \ldots$$

is dense in S^2, this gives better and better approximations of the limit map h_∞.

The following fact, already mentioned in Section 2.4, is an immediate consequence of the considerations in the proof of Theorem 11.1.

COROLLARY 11.6. *Let* $f\colon S^2 \to S^2$ *and* $g\colon \widehat{S}^2 \to \widehat{S}^2$ *be Thurston maps. If* f *and* g *are (Thurston) equivalent, then* f^n *and* g^n *are equivalent for each* $n \in \mathbb{N}$.

In general, it is not true that f and g are equivalent if f^n and g^n are equivalent for some $n \geq 2$.

PROOF. We use the same notation as in the proof of Theorem 11.1. For the construction of the infinite tower of isotopies the assumption that f and g are expanding was not needed; so we also obtain such a tower under our given assumption that f and g are equivalent Thurston maps.

Let $n \in \mathbb{N}$ be arbitrary. Then $h_n = H_1^{n-1}$ is a homeomorphism such that

$$g^n \circ h_n = g^{n-1} \circ h_{n-1} \circ f = \cdots = h_0 \circ f^n.$$

Moreover, the homeomorphisms h_0 and h_n are isotopic rel. $\text{post}(f)$, because a suitable isotopy can be obtained by concatenating the isotopies $H^0, \ldots, H^{n-1}$. Hence f^n and g^n are equivalent as desired. □

11.2. Isotopies of Jordan curves

Let X be a topological space, and $A, B, C \subset X$. We say that B is *isotopic to C rel. A*, or *B can be isotoped (or deformed) into C rel. A*,
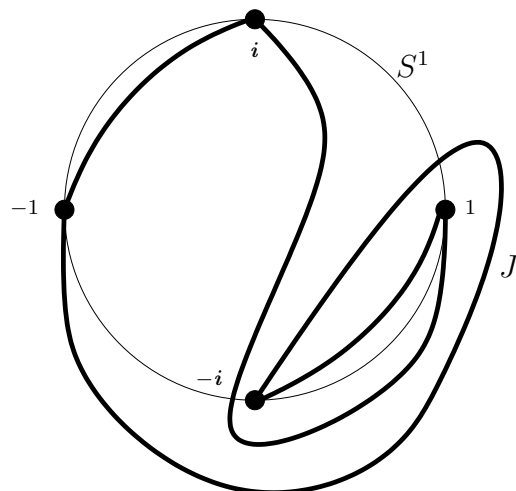
FIGURE 11.1. J is not isotopic to S^1 rel. $\{1, \boldsymbol{i}, -1, -\boldsymbol{i}\}$.

if there exists an isotopy $H\colon X \times I \to X$ rel. A with $H_0 = \mathrm{id}_X$ and $H_1(B) = C$. Note that this notion depends on the ambient space X containing the sets A, B, C.

In the following, the ambient space for all isotopies will be a fixed 2-sphere S^2 equipped with a base metric. We will study the problem when two Jordan curves J and K on S^2 passing through a given finite set P of points in the same order can be deformed into each other by an isotopy of S^2 rel. P. If $\#P \leq 3$ this is always the case (see Lemma 11.10 below).

For $\#P \geq 4$ this is not always true as the example in Figure 11.1 shows. Here $K = S^1$ is the unit circle and $P = \{1, \boldsymbol{i}, -1, -\boldsymbol{i}\} \subset S^1$. The Jordan curve J (which contains P) is drawn with a thick line. The curves $K = S^1$ and J are not isotopic rel. P. In fact, J can be obtained from S^1 by a "Dehn twist" around the points $-\boldsymbol{i}$ and 1. Note that in this example we can make the Hausdorff distance (see (4.5)) between J and S^1 arbitrarily small.

We will need the following statement.

PROPOSITION 11.7. *Suppose J is a Jordan curve in S^2 and $P \subset J$ a set consisting of $n \geq 3$ distinct points $p_1, \ldots, p_n, p_{n+1} = p_1$ in cyclic order on J. For $i = 1, \ldots, n$ let α_i be the unique arc on J with endpoints p_i and p_{i+1} such that $\mathrm{int}(\alpha_i) \subset J \setminus P$. Then there exists $\delta > 0$ with the following property:*

Let K be another Jordan curve in S^2 passing through the points $p_1, \ldots, p_n$ in cyclic order, and let β_i for $i = 1, \ldots, n$ be the arc with

endpoints p_i and p_{i+1} such that $\text{int}(\beta_i) \subset J \setminus P$. If

$$\beta_i \subset \mathcal{N}_\delta(\alpha_i)$$

for all $i = 1, \ldots, n$, then there exists an isotopy H_t on S^2 rel. P such that $H_0 = \text{id}_{S^2}$ and $H_1(J) = K$.

In other words, if the arcs β_i of the Jordan curve K are contained in sufficiently small neighborhoods of the corresponding arcs α_i of J, then one can deform J into K by an isotopy of S^2 that keeps the points in P fixed. Even though this statement seems "obvious", a complete proof is surprisingly difficult and involved. We will derive it from two lemmas in [**Bu92**].

LEMMA 11.8. *Let $\Omega \subset S^2$ be a simply connected region, $p, q \in \Omega$ distinct points, and α and β arcs in Ω with endpoints p and q. Then α is isotopic to β rel. $\{p, q\} \cup S^2 \setminus \Omega$.*

So arcs in a simply connected region with the same endpoints can be deformed into each other so that the endpoints and the complement of the region stay fixed. The lemma follows from [**Bu92**, A.6 Theorem (ii), p. 413].

LEMMA 11.9. *Suppose we have two Jordan curves J and K as in Proposition 11.7 such that for each $i = 1, \ldots, n$ the arc α_i is isotopic to β_i rel. P. Then J is isotopic to K rel. P.*

This is essentially [**Bu92**, A.5 Theorem, p. 411].

PROOF OF PROPOSITION 11.7. For each arc α_i there exists a simply connected region Ω_i that contains α_i but does not contain any element of P different from the endpoints of α_i. There exists $\delta > 0$ such that $\mathcal{N}_\delta(\alpha_i) \subset \Omega_i$ for all $i = 1, \ldots, n$. Then by Lemma 11.8 every arc β_i in $\mathcal{N}_\delta(\alpha_i)$ with the same endpoints as α_i can be isotoped to α_i rel. P. The proposition now follows from Lemma 11.9.   □

If $\#P \leq 3$ in Proposition 11.7, then J can always be isotoped to K rel. P.

LEMMA 11.10. *Suppose J and K are Jordan curves in S^2 and $P \subset J \cap K$ is a set with $\#P \leq 3$. Then J is isotopic to K rel. P.*

PROOF. Suppose first that P consists of three distinct points p_1, p_2, p_3. Define the arcs α_i and β_i as in Proposition 11.7. Then for each $i = 1, 2, 3$ the arcs α_i and β_i have the same endpoints p_i and p_{i+1} and are contained in the simply connected region $\Omega_i = S^2 \setminus \{p_{i+2}\}$, where indices are understood modulo 3. Hence by Lemma 11.8 each arc α_i is

isotopic to β_i rel. P. Again Lemma 11.9 implies that J is isotopic to K rel. P.

If $\#P \leq 2$, we may assume that $S^2 = \widehat{\mathbb{C}}$. Then by applying the first part of the proof (by adding auxiliary points to P) one sees that both J and K are isotopic to circles in $\widehat{\mathbb{C}}$ rel. P. Hence J is isotopic to K rel. P. $\hfill\square$

LEMMA 11.11. *Let S^2 and $\widehat{S}^2$ be oriented 2-spheres, and $P \subset S^2$ be a set with $\#P \leq 3$. If $h_0 \colon S^2 \to \widehat{S}^2$ and $h_1 \colon S^2 \to \widehat{S}^2$ are orientation-preserving homeomorphisms with $h_0|P = h_1|P$, then h_0 and h_1 are isotopic rel. P.*

PROOF. The statement is essentially well known. For the sake of completeness we will give a proof, but will leave some of the details to the reader. These details can easily be filled in along the lines of the proof of Proposition 5.26. In particular, the proof of the uniqueness part of this proposition used very similar arguments.

By considering $h := h_0 \circ h_1^{-1}$ one can reduce the lemma to the case where $S^2 = \widehat{S}^2$ and $h_1 = \mathrm{id}_{S^2}$. Then h fixes the points in P, and we have to show that h is isotopic to id_{S^2} rel. P. We first assume that $\#P = 3$.

Pick a Jordan curve $K \subset S^2$ with $P \subset K$, and let $J = h(K)$. Then $P \subset J \cap K$, and so by Lemma 11.10 the Jordan curve J can be isotoped into K rel. P. This implies that h is isotopic rel. P to a homeomorphism φ_1 on S^2 with $\varphi_1(K) = K$. Let e be one of the three subarcs of K determined by P. Since φ_1 fixes the three points in P, this map restricts to a homeomorphism of e that does not move the endpoints of e. Hence on e the map φ_1 is isotopic to the identity on e rel. ∂e. By pasting the isotopies on these arcs together, one can find an isotopy $H \colon K \times I \to K$ rel. P such that $H_0 = \varphi_1|K$ and $H_1 = \mathrm{id}_K$. One can extend H to each of the complementary components of K to obtain an isotopy $\overline{H} \colon S^2 \times I \to S^2$ rel. P such that $\overline{H}_1 = \mathrm{id}_{S^2}$ and $\overline{H}(p,t) = H(p,t)$ for all $p \in K$ and $t \in I$. Then $\varphi_2 := \overline{H}_0$ is a homeomorphism on S^2 that is isotopic to id_{S^2} rel. P such that $\varphi_1|K = \varphi_2|K$. This implies that φ_1 and φ_2 are isotopic rel. $K \supset P$. If $\sim$ indicates that two homeomorphisms on S^2 are isotopic rel. P, then we have $h \sim \varphi_1 \sim \varphi_2 \sim \mathrm{id}_{S^2}$, and so $h \sim \mathrm{id}_{S^2}$ as desired.

If $\#P \leq 2$, then we pick a set $P' \subset S^2$ with $\#P' = 3$ and $P' \supset P$. By the first part of the proof it suffices to find an isotopy rel. P of the given map h to a homeomorphism h' that fixes the points in P'. It is clear that such an isotopy can always be found; for an explicit

construction one can assume that $S^2 = \widehat{\mathbb{C}}$ and obtain the desired iso-
topy by postcomposing h with a suitable continuous family of Möbius
transformations, for example. $\square$

The following lemma will be crucial for the proof of the uniqueness
statement on invariant Jordan curves. In its proof we will use the
following topological fact: if D is a 2-dimensional cell and $\varphi\colon D \to S^2$
is a continuous map such that $\varphi|\partial D$ is injective, then the set $\varphi(\text{int}(D))$
contains one of the two complementary components of the Jordan curve
$\varphi(\partial D)$. Indeed, by applying the Schönflies theorem and using auxiliary
homeomorphisms we can reduce to the case where $D = \overline{\mathbb{D}}$, $S^2 = \widehat{\mathbb{C}}$,
$\varphi|\partial\mathbb{D} = \text{id}_{\partial\mathbb{D}}$, and $\infty \notin \varphi(D)$. Then $\mathbb{D} \subset \varphi(\mathbb{D})$. This follows from a
simple degree argument and the statement can be generalized to higher
dimensions; for an elementary exposition of this and related facts in
dimension 2 see [**Bu81**], in particular [**Bu81**, Corollary 3.5].

LEMMA 11.12. *Let $\mathcal{D}$ be a cell decomposition of S^2 with 1-skeleton
E and vertex set $\mathbf{V}$, and suppose that every tile in $\mathcal{D}$ contains at least
three vertices in its boundary. If J and K are Jordan curves that are
both contained in E and are isotopic rel. $\mathbf{V}$, then $J = K$.*

PROOF. Let $H\colon S^2 \times I \to S^2$ be an isotopy rel. $\mathbf{V}$ such that $H_0 =
\text{id}_{S^2}$ and $H_1(J) = K$.

Note that if $M \subset S^2$ is a set disjoint from $\mathbf{V}$, then it remains disjoint
from $\mathbf{V}$ during the isotopy, i.e., if $M \cap \mathbf{V} = \emptyset$, then $H_t(M) \cap \mathbf{V} = \emptyset$ for
all $t \in I$. This follows from the fact that each map H_t, $t \in [0, 1]$, is a
homeomorphism on S^2 with $H_t|\mathbf{V} = \text{id}_{\mathbf{V}}$.

Let e be an edge in $\mathcal{D}$. We claim that if $H_1(e) \subset E$, then $H_1(e) = e$.
First note that $H_1(e)$ is an edge in $\mathcal{D}$. Indeed, since $\partial e \subset \mathbf{V}$ and the
isotopy H does not move vertices, the arc $H_1(e)$ has the same endpoints
as e. Moreover, $\text{int}(e) \cap \mathbf{V} = \emptyset$, and so $H_1(\text{int}(e)) \cap \mathbf{V} = \emptyset$ by what we
have just seen. So $H_1(\text{int}(e))$ is a connected set in the 1-skeleton E of
$\mathcal{D}$ disjoint from the 0-skeleton $\mathbf{V}$. By Lemma 5.5 there exists an edge
e' in $\mathcal{D}$ with $H_1(\text{int}(e)) \subset \text{int}(e')$. Since the endpoints of $H_1(e)$ lie in
$\mathbf{V}$, this implies that $e' = H_1(e)$.

To show that $e' = e$ we argue by contradiction and assume that
$e \neq e'$. Then e and e' have the same endpoints, but no other points
in common. Therefore $\alpha = e \cup e'$ is a Jordan curve that contains two
vertices, namely the endpoints of e and e', but no other vertices. Let Ω_1
and Ω_2 be the two open Jordan regions that form the complementary
components of α. Then both regions Ω_1 and Ω_2 contain vertices.

To see this, note that the interior of every tile X is a connected
set disjoint from the 1-skeleton E, and hence also disjoint from α.

Therefore $\mathrm{int}(X)$ is contained in Ω_1 or Ω_2. Moreover, since the union of the interiors of tiles is dense in S^2, both regions Ω_1 and Ω_2 must contain the interior of at least one tile.

Now consider Ω_1, for example, and pick a tile X with $\mathrm{int}(X) \subset \Omega_1$. Then by our hypotheses the set $X \subset \overline{\Omega}_1 = \Omega_1 \cup \alpha$ contains at least three vertices. Since only two of them can lie on α, the set Ω_1 must contain a vertex. Similarly, Ω_2 must contain at last one vertex.

A contradiction can now be obtained from the fact that during the isotopy H the set $\mathrm{int}(e)$ remains disjoint from the set of vertices, but on the other hand it has to sweep out one of the domains Ω_1 or Ω_2 and hence it meets a vertex.

To make this rigorous, we apply the topological fact mentioned before the statement of the lemma. Let u and v be the endpoints of e. We collapse $\{u\} \times I$ and $\{v\} \times I$ in $e \times I$ to obtain a set D. Formally D is the quotient of $e \times I$ obtained by identifying all points (u,t), $t \in I$, and by identifying all points (v,t), $t \in I$. Then D is a 2-dimensional cell. Since the isotopy H does not move the points u and v, the map $(p,t) \mapsto H_t(p)$ on $e \times I$ induces a continuous map $\varphi \colon D \to S^2$. Moreover, $\varphi|\partial D$ is a homeomorphism of ∂D onto α. Hence Ω_1 or Ω_2 is contained in the set

$$\varphi(\mathrm{int}(D)) = \bigcup_{t \in (0,1)} H_t(\mathrm{int}(e)).$$

In particular, the set $\varphi(\mathrm{int}(D))$ contains a vertex. This is a contradiction, because we know that no set $H_t(\mathrm{int}(e))$, $t \in I$, meets $\mathbf{V}$. Thus $H_1(e) = e$ as desired.

Having verified the statement about edges, it is now easy to see that $J = K$. Indeed, J is a union of edges in $\mathcal{D}$; to see this, consider the components of the set $J \setminus \mathbf{V}$. If γ is such a component, then $\overline{\gamma} \setminus \gamma \subset \mathbf{V}$. Moreover, γ is contained in the 1-skeleton E, and does not meet the 0-skeleton $\mathbf{V}$. Again by Lemma 5.5 the set γ must be contained in the interior $\mathrm{int}(e)$ of some edge e. This is only possible if $\gamma = \mathrm{int}(e)$. Hence $\overline{\gamma} = e$. Since J is the union of the closures of these components γ, it follows that J is the union of edges e. For each such edge e we have $H_1(e) \subset K \subset E$ and so $H_1(e) = e$ by the first part of the proof. This implies $J \subset K$. Since J and K are Jordan curves, the desired identity $J = K$ follows. $\square$

11.3. Isotopies and cell decompositions

The main result in this section is Lemma 11.17 which gives a criterion when a Jordan curve $\mathcal{C}$ in a 2-sphere S^2 can be isotoped relative to a finite set $P \subset \mathcal{C}$ into the 1-skeleton of a given cell decomposition $\mathcal{D}$

of S^2. We first discuss some facts about graphs that are needed in the proof. Since all the graphs we consider will be embedded in a 2-sphere, we base the concept of a graph on a topological definition rather than a combinatorial one as usual (in Chapter 20 it will be more convenient to adopt the combinatorial viewpoint).

A *(finite) graph* is a compact Hausdorff space G equipped with a fixed cell decomposition $\mathcal{D}$ such that $\dim(c) \leq 1$ for all $c \in \mathcal{D}$. The cells c in $\mathcal{D}$ of dimension 1 are called the *edges* of the graph, and the points $v \in G$ such that $\{v\}$ is a 0-dimensional cell in $\mathcal{D}$ the *vertices* of the graph. Note that we do allow multiple edges, i.e., two or more edges with the same endpoints v, w. Loops however, meaning edges where the two endpoints agree, are not allowed according to our definition.

An *oriented edge e* in a graph is an edge, where one of the vertices in ∂e has been chosen as the *initial point* and the other vertex as the *terminal point* of e. An *edge path* in G is a finite sequence α of oriented edges $e_1, \ldots, e_N$ such that the terminal point of e_i is the initial point of e_{i+1} for $i = 1, \ldots, N-1$. We denote by $|\alpha| = e_1 \cup \cdots \cup e_N$ the underlying set of the edge path. The edge path α *joins* the vertices $a, b \in G$ if the initial point of e_1 is a and the terminal point of e_N is b. The number N is called the *length* of the edge path. The edge path is called *simple* if e_i and e_j are disjoint for $1 \leq i < j \leq N$ and $j - i \geq 2$, and $e_i \cap e_j$ consists of precisely one point (the terminal point of e_i and initial point of e_j) when $j = i + 1$. If the edge path α is simple, then $|\alpha|$ is an arc. The edge path is called a *loop* if the terminal point of e_N is the initial point of e_1.

A graph is connected (as a topological space) if and only if any two vertices $a, b \in G$, $a \neq b$, can be joined by an edge path. The *combinatorial distance* of two vertices a and b in a connected graph G is defined as the minimal length of all edge paths joining the points (interpreted as 0 if $a = b$). The vertices $a, b \in G$ are called *neighbors* if their combinatorial distance is equal to 1, i.e., if there exists an edge e in G whose endpoints are a and b. A vertex $q \in G$ is called a *cut point* of G if $G \setminus \{q\}$ is not connected. A vertex $q \in G$ is not a cut point if and only if all vertices $a, b \in G \setminus \{q\}$, $a \neq b$, can be joined by an edge path α with $q \notin |\alpha|$.

LEMMA 11.13. *Let G be a connected graph without cut points. Then for all vertices $a, b, p \in G$ with $a \neq b$ there exists a simple edge path γ in G with $p \in |\gamma|$ that joins a and b.*

PROOF. Since G is connected, there exist edge paths in G joining a and b. By removing loops from such a path if necessary, we can also obtain such an edge path in G that is simple. Among all such simple
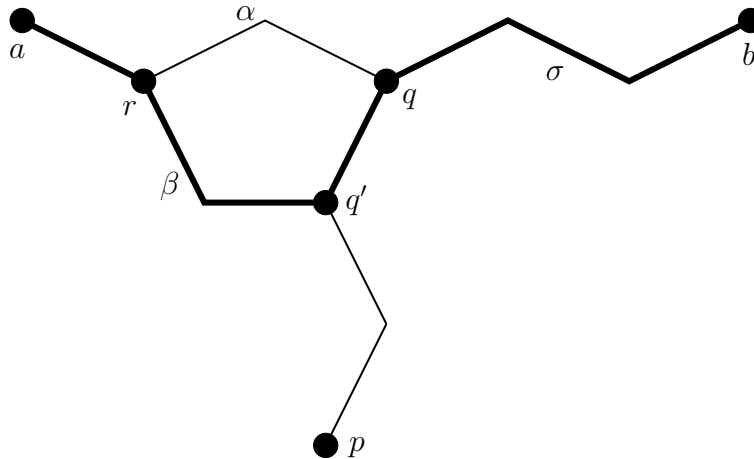
FIGURE 11.2. Constructing a path through a, b, p.

paths, there is one that contains a vertex with minimal combinatorial distance to p. More precisely, there exists a simple edge path α in G with endpoints a and b, and a vertex $q \in |\alpha|$ such that the combinatorial distance $k \in \mathbb{N}_0$ of q and p is minimal among all combinatorial distances between p and vertices on simple paths joining a and b. If $k = 0$ then $q = p$ and we can take $\gamma = \alpha$.

We will show that the alternative case $k \geq 1$ leads to a contradiction. By definition of combinatorial distance, there exists an edge path joining q to p consisting of $k \geq 1$ edges. The second vertex q' on this path as traveling from q to p is a neighbor of q whose combinatorial distance to p is $k - 1$ and hence strictly smaller than the combinatorial distance of q to p. In particular, $q' \notin |\alpha|$ by choice of q and α. We will obtain the desired contradiction if we can show that there exists a simple edge path σ in G that joins a and b and passes through q'.

For the construction of σ we apply our assumption that G has no cut points; in particular, q is no cut point and hence there exists an edge path β with $q \notin |\beta|$ that joins q' to a vertex in the (non-empty) set $A = |\alpha| \setminus \{q\}$. We may assume that β is simple and that the endpoint $r \neq q'$ of β is the only point in $|\beta| \cap A$.

Moreover, we may assume that r lies between a and q on the path α (the argument in the other case where r lies between q and b is similar). Now let σ be the edge path obtained by traveling from a to r along α, then from r to q' along β, then from q' to q along an edge (this is possible since q and q' are neighbors), and finally from q to b along α. See the illustration in Figure 11.2. Then σ is a simple edge path in G that passes through q' and has the endpoints a and b. This gives the desired contradiction. $\qquad\square$

Now let S^2 be a 2-sphere, and $\mathcal{D}$ be a cell decomposition of S^2. We denote the set of tiles, edges, and vertices in $\mathcal{D}$ by $\mathbf{X}$, $\mathbf{E}$, and $\mathbf{V}$, respectively. In the following the terms cell, tile, etc., refer to elements of these sets.

Let $M \subset \mathbf{X}$ be a set of tiles. We denote by $|M|$ its underlying set; so

$$|M| = \bigcup_{X \in M} X.$$

The set

(11.5) $$G_M := \bigcup_{X \in M} \partial X$$

admits a natural cell decomposition consisting of all cells contained in G_M. Obviously, no such cell can be a tile, so with this cell decomposition G_M is a graph.

Recall from Definition 5.20 that a sequence $X = X_1, \ldots, X_N = Y$ of tiles is an *e-chain* if $X_i \neq X_{i+1}$ and there exists an edge e_i with $e_i \subset \partial X_i \cap \partial X_{i+1}$ for $i = 1, \ldots, N-1$. It *joins* the tiles X and Y. A set M of tiles is *e-connected* if every two tiles in M can be joined by an e-chain consisting of tiles in M.

LEMMA 11.14. *Let $M \subset \mathbf{X}$ be a set of tiles that is e-connected. Then the graph G_M is connected and has no cut points.*

PROOF. Let $a, b \in G_M$ be arbitrary vertices with $a \neq b$. We can pick tiles X and Y in M such that a is a vertex in X and b is a vertex in Y. By assumption there exists an e-chain $X_1, X_1, \ldots, X_N$ in M with $X_1 = X$ and $X_N = Y$. The vertices of a tile X_i lie in G_M; they subdivide the Jordan curve ∂X_i such that successive vertices on ∂X_i are connected by an edge and are hence neighbors in G_M. An edge path α in G_M joining a and b can now be obtained as follows: starting from $a \in \partial X_0$, use edges on the boundary of X_0 to find an edge path in G_M that joins $p_0 = a$ to a vertex p_1 of X_1. This is possible, since X_0 and X_1 have a common edge and hence at least two common vertices. Then run from p_1 along edges on ∂X_1 to a vertex p_2 of X_2, and so on. Once we arrived at a vertex p_N of X_N, we can reach b by running from p_N to $p_{N+1} := b$ along edges on X_N. In this way we obtain an edge path α in G_M that joins a and b.

A slight refinement of this argument also shows that we can construct the path α so that it avoids any given vertex q in G_M distinct from a and b. Indeed, choose $p_0 = a$ as before. Since X_0 and X_1 have at least two vertices in common, we can pick a common vertex p_1 of

X_0 and X_1 that is distinct from q. There exists an arc on ∂X_0 (possibly degenerate) that does not contain q and joins p_0 and p_1. This arc (if non-degenerate) consists of edges and if we follow these edges, we obtain an edge path in G_M that does not contain q and joins p_0 and p_1. In the same way we can find an edge path in G_M that avoids q and joins p_1 to a vertex $p_2 \in \partial X_1 \cap \partial X_2$, and so on. Concatenating all these edge paths we get a path α as desired.

This shows that G_M is connected and has no cut points. □

LEMMA 11.15. *Let $M \subset \mathbf{X}$ be a set of tiles that is e-connected, and let $a, b, p \in |M|$ be distinct vertices. Then there exists a simple edge path α in G_M with $p \in |\alpha|$ that joins a and b.*

In particular, this applies if M consists of a single e-chain.

PROOF. This follows from Lemma 11.14 and Lemma 11.13. □

LEMMA 11.16. *Let $\gamma\colon J \to S^2$ be a path in S^2 defined on an interval $J \subset \mathbb{R}$ and $M = M(\gamma)$ be the set of tiles having non-empty intersection with γ. Then M is e-connected.*

PROOF. We first prove the following claim. If $[a, b] \subset \mathbb{R}$, $\alpha\colon [a, b] \to S^2$ is a path, and X and Y are tiles with $\alpha(a) \in X$ and $\alpha(b) \in Y$, then there exists an e-chain $X_1 = X, X_2, \ldots, X_N = Y$ such that $X_i \cap \alpha \neq \emptyset$ for all $i = 1, \ldots, N$.

In the proof of this claim, we call an e-chain $X_1, \ldots, X_N$ *admissible* if $X_1 = X$ and $X_i \cap \alpha \neq \emptyset$ for all $i = 1, \ldots, N$. So we want to find an admissible e-chain whose last tile is Y.

Let $T \subset [a, b]$ be the set of all points $t \in [a, b]$ for which there exists an admissible e-chain $X_1, \ldots, X_N$ with $\alpha(t) \in X_N$. We first want to show that $b \in T$.

First note that the set T is closed. Indeed, suppose that $\{t_k\}$ is a sequence in T with $t_k \to t_\infty \in [a, b]$ as $k \to \infty$. Then for each $k \in \mathbb{N}$ there exists an admissible e-chain $X_1^k, \ldots, X_{N_k}^k$ with $\alpha(t_k) \in X_{N_k}^k$. Define $Z_k = X_{N_k}^k$ to be the last tile in this chain. Since there are only finitely many tiles, there exists one tile, say Z, among the tiles $Z_1, Z_2, Z_3, \ldots$ that appears infinitely often in this sequence. Since $\alpha(t_k) \to \alpha(t_\infty)$, we have $\alpha(t_k) \in Z$ for infinitely many k. Since tiles are closed, we conclude $\alpha(t_\infty) \in Z$. By definition of Z there exists an admissible e-chain $X_1, \ldots, X_N$ with $X_N = Z$. Then $\alpha(t_\infty) \in Z = X_N$, and so $t_\infty \in T$.

Obviously, $a \in T$ and so T is non-empty. Since T is also closed, the set T has a maximum, say $m \in [a, b]$. We have to show that $m = b$; we will see that the assumption $m < b$ leads to a contradiction.

We consider $p := \alpha(m)$. Then there exists an admissible e-chain $X_1, \ldots, X_N$ with $p \in Z := X_N$.

If $p \in \operatorname{int}(Z)$, then $\alpha(t) \in Z$ and so $t \in T$ for $t \in (m, b]$ close to m. This is impossible by definition of m.

If p does not belong to $\operatorname{int}(Z)$, then p must be a boundary point of Z. Suppose first that p is in the interior of an edge $e \subset \partial Z$. By Lemma 5.9 (iv) there exists precisely one tile Z' distinct from Z such that $e \subset \partial Z'$. Moreover, $Z \cup Z'$ is a neighborhood of p, and so points $\alpha(t)$ with $t \in (m, b]$ close to m belong to Z or Z'. Since Z' contains p and hence meets α, and Z and Z' share an edge, $X_1, \ldots, X_N, Z'$ is an admissible e-chain. It follows that $t \in T$ for $t \in (m, b]$ close to m. Again this is impossible by definition of m.

If p is a boundary point of Z, but not in the interior of an edge, then p is a vertex. The tiles in the cycle of p form a neighborhood of p, and so a point $\alpha(t)$ for some $t \in (m, b]$ close to m will belong to a tile Z' in the cycle of p. It follows from Lemma 5.9 (v) that any two tiles in the cycle of a vertex can be connected by an e-chain consisting of tiles in the cycle. Hence there exists an e-chain $Z = Z_1, \ldots, Z_K = Z'$ such that $p \in Z_j$ for $j = 1, \ldots, K$. In particular, $\alpha \cap Z_j \neq \emptyset$ for $j = 1, \ldots, K$, and so $X_1, \ldots, X_N = Z = Z_1, \ldots, Z_K = Y'$ is an admissible e-chain. Since $\alpha(t) \in Z' = Z_K$, we have $t \in T$, again a contradiction.

We have exhausted all possibilities proving that $b \in T$ as desired. This implies that there exists an admissible e-chain $X_1 = X, \ldots, X_N$ with $\alpha(b) \in X_N$. If $X_N = Y$, then we are done. If $X_N \neq Y$, then $\alpha(b) \in \partial X_N \cap \partial Y$, and so $\alpha(b)$ is an interior point of an edge e with $e \subset \partial X_N \cap \partial Y$, or $\alpha(b)$ is a vertex. As in the first part of the proof, one can then extend the admissible e-chain $X_1 = X, \ldots, X_N$ to obtain an admissible e-chain whose last tile is Y. The claim made in the beginning of the proof follows.

This claim now easily implies the statement of the lemma. Indeed, let $X, Y \in M = M(\gamma)$ be arbitrary. Then there exist $a, b \in J$ with $\gamma(a) \in X$ and $\gamma(b) \in Y$. If $a \leq b$, then we apply the claim to the path $\alpha = \gamma|[a, b]$, and if $b \leq a$ to the path $\alpha = \gamma|[b, a]$. This shows that we can find an e-chain in M that joins X and Y. $\qquad\square$

For the formulation of the next statement, we need a slight extension of Definition 5.32. Let $\mathcal{C} \subset S^2$ be a Jordan curve, and $P \subset \mathcal{C}$ be a finite set with $\#P \geq 3$. The points in P divide $\mathcal{C}$ into subarcs that have endpoints in P, but whose interiors are disjoint from P. We say that a (not necessarily connected) set $K \subset S^2$ *joins opposite sides of* $(\mathcal{C}, P)$ if $k = \#\operatorname{post}(f) \geq 4$ and K meets two of these arcs that are

non-adjacent (i.e., disjoint), or if $k = \#\operatorname{post}(f) = 3$ and K meets all of these arcs (in this case there are three arcs).

In the following lemma and its proof, metric notions refer to some fixed base metric on S^2.

LEMMA 11.17. *Let* $\mathcal{C} \subset S^2$ *be a Jordan curve, and* $P \subset \mathcal{C}$ *a finite set with* $k := \#P \geq 3$. *Then there exists* $\epsilon_0 > 0$ *satisfying the following condition:*

Suppose that $\mathcal{D}$ *is a cell decomposition of* S^2 *with vertex set* $\mathbf{V}$ *and 1-skeleton E. If* $P \subset \mathbf{V}$ *and*

$$\max_{c \in \mathcal{D}} \operatorname{diam}(c) < \epsilon_0,$$

then there exists a Jordan curve $\mathcal{C}' \subset E$ *that is isotopic to* $\mathcal{C}$ *rel. P, and has the property that no tile in* $\mathcal{D}$ *joins opposite sides of* $(\mathcal{C}', P)$.

PROOF. We fix an orientation of $\mathcal{C}$ and let $p_1, \dots, p_k$ be the points in P in cyclic order on $\mathcal{C}$. The points in P divide $\mathcal{C}$ into subarcs $\mathcal{C}_1, \dots, \mathcal{C}_k$ such that for $i = 1, \dots, k$ the arc $\mathcal{C}_i$ has the endpoints p_i and p_{i+1} and has interior disjoint from P. Here and in the following the index i is understood modulo k, i.e., $p_{k+1} = p_1$, etc. Note that $\mathcal{C}_i \cap \mathcal{C}_{i+1} = \{p_{i+1}\}$ for $i = 1, \dots, k$. There exists a number $\delta_0 > 0$ such that no set $K \subset S^2$ with $\operatorname{diam}(K) < \delta_0$ joins opposite sides of $(\mathcal{C}, P)$ (this can be seen as in the discussion after (5.13)).

Now choose $\delta > 0$ as in Proposition 11.7 for $J = \mathcal{C}$ (and $n = k$). We may assume that $3\delta < \delta_0$.

We break up $\mathcal{C}$ into subarcs

(11.6) $$\alpha_1, \gamma_1, \alpha_2, \gamma_2, \dots, \alpha_k, \gamma_k, \alpha_1,$$

arranged in cyclic order on $\mathcal{C}$, such that p_i is an interior point of α_i and we have $\alpha_i \subset B(p_i, \delta/2)$ for each $i = 1, \dots, k$. The arcs in (11.6) have disjoint interiors, and two arcs have an endpoint in common if and only if they are adjacent in this cyclic order in which case they share one endpoint. So each "middle piece" γ_i does not contain any point from P and is contained in the interior of $\mathcal{C}_i$.

We choose $0 < \epsilon_0 < \delta/4$ so small that the distance between non-adjacent arcs in (11.6) is $\geq 10\epsilon_0$ and so that

$$\operatorname{dist}(p_i, \gamma_{i-1} \cup \gamma_i) \geq 10\epsilon_0$$

for $i = 1, \dots, k$.

Now suppose we have a cell decomposition $\mathcal{D}$ of S^2 such that P is contained in the vertex set $\mathbf{V}$ of $\mathcal{D}$ and

$$\max_{c \in \mathcal{D}} \operatorname{diam}(c) < \epsilon_0.$$
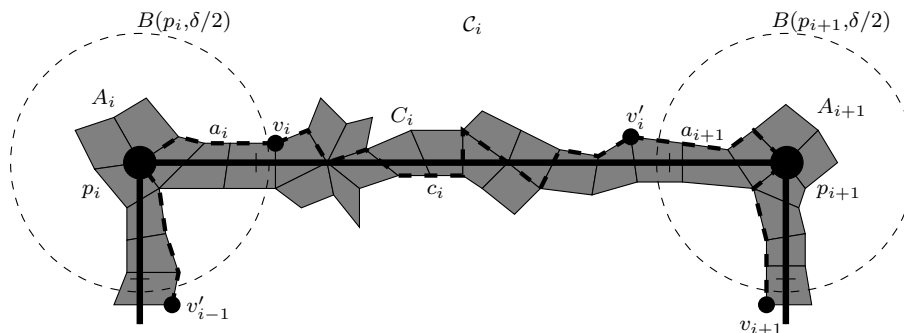
FIGURE 11.3. Construction of the curve $\mathcal{C}'$.

Our goal is to find a Jordan curve $\mathcal{C}' \subset S^2$ consisting of arcs $\mathcal{C}'_i$ that are unions of edges, have endpoints p_i and p_{i+1}, and satisfy

$$\mathcal{C}'_i \subset \mathcal{N}_\delta(\mathcal{C}_i)$$

for $i = 1, \ldots, k$.

Let $\mathbf{A}_i$ be the set of all tiles intersecting α_i and $\mathbf{C}_i$ be the set of all tiles intersecting γ_i for $i = 1, \ldots, k$. Recall that for a given set of tiles M, we denote by $|M|$ the union of tiles in M. Let $A_i := |\mathbf{A}_i|$ and $C_i := |\mathbf{C}_i|$.

Note that

$$A_i \subset \mathcal{N}_{\epsilon_0}(\alpha_i) \quad \text{and} \quad C_i \subset \mathcal{N}_{\epsilon_0}(\gamma_i).$$

Moreover,

$$
\begin{aligned}
A_i \cup C_i \cup A_{i+1} \quad &\subset \quad \mathcal{N}_{\epsilon_0}(\alpha_i) \cup \mathcal{N}_{\epsilon_0}(\gamma_i) \cup \mathcal{N}_{\epsilon_0}(\alpha_{i+1}) \\
&\subset \quad B(p_i, \delta) \cup \mathcal{N}_{\epsilon_0}(\gamma_i) \cup B(p_{i+1}, \delta) \\
&\subset \quad \mathcal{N}_\delta(\mathcal{C}_i),
\end{aligned}
$$
(11.7)

and the natural cyclic order of these sets is

(11.8) $A_1, C_1, A_2, C_2, \ldots, A_k, C_k, A_1.$

By choice of ϵ_0 we know that if two of the sets in (11.8) are not adjacent in the cyclic order, then their distance is $\geq 8\epsilon_0$ and so their intersection is empty. Moreover, for $i = 1, \ldots, k$ the only one of these sets that contains p_i is A_i.

The construction that now follows is illustrated in Figure 11.3. Here the two large dots represent two points p_i, p_{i+1} and the thick line the curve $\mathcal{C}$.

For $i = 1, \ldots, k$ we consider the graphs $G_{\mathbf{A}_i}$ and $G_{\mathbf{C}_i}$ associated with the tile sets $\mathbf{A}_i$ and $\mathbf{C}_i$, respectively, as in (11.5). Each of these graphs is a union of edges in $\mathcal{D}$. Moreover, $G_{\mathbf{A}_i} \subset A_i$ and $G_{\mathbf{C}_i} \subset C_i$. Note that there is at least one tile contained in both $\mathbf{A}_i$ and $\mathbf{C}_i$, namely

any tile containing the common endpoint of α_i and γ_i. Hence $G_{\mathbf{C}_i}$ and $G_{\mathbf{A}_i}$ have a common vertex contained in A_i; similarly $G_{\mathbf{C}_i}$ and $G_{\mathbf{A}_{i+1}}$ have a common vertex contained in A_{i+1}.

It follows from Lemmas 11.16 and 11.15 that $G_{\mathbf{C}_i}$ is connected. Hence we can find a simple edge path c_i' in $G_{\mathbf{C}_i}$ joining a vertex $v_i \in A_i$ as the initial point to a vertex $v_i' \in A_{i+1}$ as the terminal point. Let $c_i := |c_i'| \subset C_i$ be the underlying arc. By deleting edges from c_i' if necessary, we may assume that v_i is the only vertex in $c_i \cap A_i$ and v_i' is the only vertex in $c_i \cap A_{i+1}$. Then c_i has no other points in common with A_i or A_{i+1}.

To see this, suppose that there exists a point $x \neq v_i, v_i'$ with $x \in c_i \cap (A_i \cup A_{i+1})$, say $x \in c_i \cap A_i$. Then x is contained in an edge e of the edge path c_i'. The point $x \neq v_i$ cannot be a vertex, because v_i is the only vertex in $c_i \cap A_i$. So $x \in \mathrm{int}(e) \cap A_i$ which implies that $e \subset A_i$; but then both endpoints of e are vertices in $c_i \cap A_i$, which is impossible by our choice of c_i'.

Note that $v_{i-1}' \in C_{i-1}$ and $v_i \in C_i$ are distinct vertices in A_i, and recall that $p_i \in A_i$. Then Lemmas 11.16 and 11.15 imply that there exists an arc $a_i \subset A_i$ with $p_i \in a_i$ that consists of edges and has the endpoints v_{i-1}' and v_i. Since $p_i \notin C_{i-1} \cup C_i$, we have $v_{i-1}', v_i \neq p_i$, and so $p_i \in \mathrm{int}(a_i)$.

If we arrange the arcs a_i and c_i in cyclic order

$$a_1, c_1, a_2, c_2 \ldots, a_k, c_k, a_1,$$

then two of these arcs have non-empty intersection if and only if they are adjacent in this order. If two arcs are adjacent, then their intersection consists of a common endpoint. Therefore, the set

$$\mathcal{C}' := a_1 \cup c_1 \cup a_2 \cup c_2 \cup \cdots \cup a_k \cup c_k$$

is a Jordan curve that passes through the points $p_1, \ldots, p_k$. Moreover, $\mathcal{C}'$ consists of edges and is hence contained in the 1-skeleton E of $\mathcal{D}$.

By construction each vertex p_i is an interior point of the arc a_i. Thus it divides a_i into two subarcs a_i^- and a_i^+ consisting of edges such that p_i is a common endpoint of a_i^- and a_i^+, and such that a_i^- shares an endpoint with c_{i-1} and a_i^+ one with c_i. Then

$$\mathcal{C}_i' := a_i^+ \cup c_i \cup a_{i+1}^-$$

for $i = 1, \ldots, k$ is an arc that consists of edges and has endpoints p_i and p_{i+1}. The arcs $\mathcal{C}_1', \ldots, \mathcal{C}_k'$ have pairwise disjoint interior. Moreover,

$$\mathcal{C}' = \mathcal{C}_1' \cup \cdots \cup \mathcal{C}_k',$$

and by (11.7) we have

$$\mathcal{C}'_i \subset A_i \cup C_i \cup A_{i+1} \subset \mathcal{N}_\delta(\mathcal{C}_i).$$

Hence by Proposition 11.7 and choice of δ, the curve $\mathcal{C}'$ is isotopic to $\mathcal{C}$ rel. P.

It remains to show that no tile in $\mathcal{D}$ joins opposite sides of $(\mathcal{C}', P)$. To see this, we argue by contradiction. Suppose that there exists a tile X in $\mathcal{D}$ that joins opposite sides of $(\mathcal{C}', P)$. Then $K := \mathcal{N}_\delta(X)$ joins opposite sides of $(\mathcal{C}, P)$, since $\mathcal{C}'_i \subset \mathcal{N}_\delta(\mathcal{C}_i)$ for all $i = 1, \ldots, k$. By choice of δ_0 we then have

$$\delta_0 \leq \operatorname{diam}(K) \leq 2\delta + \operatorname{diam}(X) \leq 2\delta + \epsilon_0 < 3\delta < \delta_0,$$

which is impossible. $\square$

CHAPTER 12

Subdivisions

In complex dynamics the iteration of polynomials is much better understood than the iteration of general rational maps. One of the reasons is that for polynomials powerful combinatorial methods are available such as external rays, Hubbard trees, or Yoccoz puzzles (see [**DH84**]). It is desirable to develop similar concepts for other classes of maps as well. For Thurston maps we will introduce the notion of a *two-tile subdivision rule* in this chapter. It provides a useful combinatorial tool for their investigation.

This concept can be extracted from various previous examples (see Sections 1.1 and 1.3, or Examples 2.6 and 6.10), where we have described Thurston maps by a subdivision procedure. In these examples we consider a topological 2-sphere obtained as a pillow (see Section A.10) by gluing two k-gons together along their boundaries. Then the two faces of the pillow (the 0-tiles) are subdivided into k-gons (the 1-tiles) and it is specified how the map sends a 1-tile to one of the 0-tiles. The equator of the pillow is a Jordan curve that is invariant under the map and contains its postcritical points.

More generally, suppose that $f: S^2 \to S^2$ is a Thurston map with $\# \operatorname{post}(f) \geq 3$, and $\mathcal{C} \subset S^2$ is a Jordan curve with $\operatorname{post}(f) \subset \mathcal{C}$. If $\mathcal{C}$ is f-invariant (i.e., $f(\mathcal{C}) \subset \mathcal{C}$), then the cell decompositions $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$ (see Definition 5.14) have nice compatibility properties given by Proposition 12.5. In particular, $\mathcal{D}^{n+k}$ is a refinement of $\mathcal{D}^n$, whenever $n, k \in \mathbb{N}_0$. Intuitively, this means that each cell $\mathcal{D}^n$ is "subdivided" by the cells in $\mathcal{D}^{n+k}$. A cell $c \in \mathcal{D}^n$ is actually subdivided by the cells in $\mathcal{D}^{n+k}$ "in the same way" as the cell $f^n(c) \in \mathcal{D}^0$ by the cells in $\mathcal{D}^k$ (see Proposition 12.5 (v) for a precise statement). This implies that the "combinatorics" of the sequence $\mathcal{D}^0, \mathcal{D}^1, \mathcal{D}^2 \ldots \ldots$ is uniquely determined by the pair $(\mathcal{D}^1, \mathcal{D}^0)$ and the map $\tau \in \mathcal{D}^1 \to f(\tau) \in \mathcal{D}^0$, i.e., the labeling $L: \mathcal{D}^1 \to \mathcal{D}^0$ induced by f (see Section 5.4). For more discussion see Remark 12.12 (ii) and the related Proposition 12.19.

The triples $(\mathcal{D}^1, \mathcal{D}^0, L)$ arising in this way lead to the following definition (see the beginning of Section 12.2 for more motivation).

DEFINITION 12.1 (Two-tile subdivision rules). Let S^2 be a 2-sphere. A *two-tile subdivision rule* for S^2 is a triple $(\mathcal{D}^1, \mathcal{D}^0, L)$ of cell decompositions $\mathcal{D}^0$ and $\mathcal{D}^1$ of S^2 and an orientation-preserving labeling $L\colon \mathcal{D}^1 \to \mathcal{D}^0$. We assume that the cell decompositions satisfy the following conditions:

(i) $\mathcal{D}^0$ contains precisely two tiles.

(ii) $\mathcal{D}^1$ is a refinement of $\mathcal{D}^0$, and $\mathcal{D}^1$ contains more than two tiles.

(iii) If k is the number of vertices in $\mathcal{D}^0$, then $k \geq 3$ and every tile in $\mathcal{D}^1$ is a k-gon.

(iv) Every vertex in $\mathcal{D}^1$ is contained in an even number of tiles in $\mathcal{D}^1$.

If $\mathcal{D}^0$ is a cell decomposition of S^2 with precisely two tiles X and Y, then necessarily $\partial X = \partial Y$. The set $\mathcal{C} \coloneqq \partial X = \partial Y$ is a Jordan curve which we call *the Jordan curve of $\mathcal{D}^0$*. Then $\mathcal{C}$ is the 1-skeleton of $\mathcal{D}^0$ and all vertices and edges of $\mathcal{D}^0$ lie on $\mathcal{C}$. If k is the number of these vertices on $\mathcal{C}$ and $\mathcal{D}^1$ is another cell decomposition of S^2, then a Thurston map f that is cellular for $(\mathcal{D}^1, \mathcal{D}^0)$ can only exist if each tile in $\mathcal{D}^1$ is a k-gon, i.e., it contains exactly k vertices and edges in its boundary. Since f is not a homeomorphism, $\mathcal{D}^1$ contains more than two tiles. The number of tiles in $\mathcal{D}^1$ that contain a given vertex v in $\mathcal{D}^1$ is equal to the length of the cycle of v in $\mathcal{D}^1$. This number has to be even, because it must be an integer multiple of the length of a vertex cycle in $\mathcal{D}^0$ which is always equal to 2. This motivated the requirements (ii)–(iv) in Definition 12.1.

We say that a continuous map $f\colon S^2 \to S^2$ *realizes* the two-tile subdivision rule $(\mathcal{D}^1, \mathcal{D}^0, L)$ if f is cellular for $(\mathcal{D}^1, \mathcal{D}^0)$ and $f(\tau) = L(\tau)$ for each $\tau \in \mathcal{D}^1$. Note that in this case $(\mathcal{D}^1, \mathcal{D}^0)$ is a cellular Markov partition for f (see Definition 5.8).

The following proposition shows that two-tile subdivision rules arise from Thurston maps with invariant curves.

PROPOSITION 12.2 (Two-tile subdivision rules via Thurston maps). *Suppose $f\colon S^2 \to S^2$ is a Thurston map with $\operatorname{post}(f) \geq 3$, and $\mathcal{C} \subset S^2$ is an f-invariant Jordan curve with $\operatorname{post}(f) \subset \mathcal{C}$. If we define $\mathcal{D}^0 = \mathcal{D}^0(f, \mathcal{C})$, $\mathcal{D}^1 = \mathcal{D}^1(f, \mathcal{C})$, and $L\colon \mathcal{D}^1 \to \mathcal{D}^0$ by setting $L(\tau) = f(\tau)$ for $\tau \in \mathcal{D}^1$, then $(\mathcal{D}^1, \mathcal{D}^0, L)$ is a two-tile subdivision rule realized by f.*

Theorem 15.1 implies that every expanding Thurston map f has an iterate $F = f^n$ that realizes a two-tile subdivision rule.

Conversely, a two-tile subdivision rule gives rise to a Thurston map with an invariant curve and this map is unique up to Thurston equivalence.

PROPOSITION 12.3 (Thurston maps via two-tile subdivision rules). *Let $(\mathcal{D}^1, \mathcal{D}^0, L)$ be a two-tile subdivision rule on S^2. Then there exists a Thurston map $f\colon S^2 \to S^2$ that realizes $(\mathcal{D}^1, \mathcal{D}^0, L)$. The map f is unique up to Thurston equivalence.*

Moreover, the Jordan curve $\mathcal{C}$ of $\mathcal{D}^0$ is f-invariant and contains the set $\mathrm{post}(f)$.

When we constructed or described certain Thurston maps, we used these propositions informally several times before (see, for example, Figures 1.1, 1.2, 2.1, 2.2, 3.4, 3.5, 3.6, 6.1, and 7.1). In these examples we used a geometric picture to represent the cell decompositions and the labeling of a two-tile subdivision rule. Formally, we obtained the corresponding map from Proposition 12.3.

Our concept of a two-tile subdivision rule is inspired by the more general concept of a *subdivision rule* as introduced by Cannon, Floyd, and Parry (see [**CFP01, CFP06a, C–P03**], and also [**BS97, Me02**]). In their definition an explicit map from the 1-cells to 0-cells is specified (corresponding to the map f in our case); in contrast, our definition is purely combinatorial.

The reason for the name *two-tile* subdivision rule is that the data given by $(\mathcal{D}^1, \mathcal{D}^0)$ determines how the *two* 0-*tiles* are subdivided by the cells in $\mathcal{D}^1$, and this together with the labeling L can be used to create a sequence of cell decompositions $\mathcal{D}^n$ where each cell $\tau \in \mathcal{D}^1$ is subdivided by the cells in $\mathcal{D}^2$ in the same way as the cell $L(\tau) \in \mathcal{D}^0$ is subdivided by the 1-cells, etc. Our definition is tailored to generate Thurston maps, so a more accurate term would have been a "two-tile subdivision rule generating a Thurston map", but we chose the shorter term for brevity.

Since we are mostly interested in expanding Thurston maps, we want to find a combinatorial condition on a two-tile subdivision rule that ensures that it can be realized by an expanding Thurston map. To motivate a relevant definition, suppose that $f\colon S^2 \to S^2$ is a Thurston map and $\mathcal{C} \subset S^2$ is a Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}$. We consider the quantities $D_n = D_n(f, \mathcal{C})$ defined in (5.14). If the Jordan curve $\mathcal{C}$ here is f-invariant, then it is easy to see that the numbers D_n are non-decreasing as $n \to \infty$. We will show that we actually have an exponential increase under the additional assumption that there exists $n_0 \in \mathbb{N}$ with $D_{n_0} \geq 2$ (see Lemma 12.8). This will turn out to be a key condition related to expansion of Thurston maps realizing two-tile subdivision rules.

DEFINITION 12.4 (Combinatorial expansion). Let $f\colon S^2 \to S^2$ be a Thurston map. We call f *combinatorially expanding* if $\mathrm{post}(f) \geq 3$,

and if there exists a Jordan curve $\mathcal{C} \subset S^2$ that is f-invariant, satisfies $\mathrm{post}(f) \subset \mathcal{C}$, and for which there is a number $n_0 \in \mathbb{N}$ such that $D_{n_0}(f, \mathcal{C}) \geq 2$.

The condition $D_{n_0}(f, \mathcal{C}) \geq 2$ means that no single n_0-tile for $(f, \mathcal{C})$ joins opposite sides of $\mathcal{C}$.

If f and $\mathcal{C}$ are as in the previous definition, then we say that f is *combinatorially expanding for $\mathcal{C}$*. This condition is indeed combinatorial in nature, because it can be verified just by knowing the combinatorics of the cell decompositions $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$, $n \in \mathbb{N}_0$. This in turn is determined by the combinatorics of the pair $(\mathcal{D}^1, \mathcal{D}^0)$ and the labeling $\tau \in \mathcal{D}^1 \mapsto f(\tau) \in \mathcal{D}^0$ induced by f (see Remark 12.12 (ii) and Proposition 12.19).

If a Thurston map $f \colon S^2 \to S^2$ is expanding and $\mathcal{C} \subset S^2$ is an f-invariant Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}$, then f is also combinatorially expanding for $\mathcal{C}$, because in this case $D_n(f, \mathcal{C}) \to \infty$ as $n \to \infty$ (see Lemma 8.6). The converse is not true in general, as a combinatorially expanding Thurston map need not be expanding (see Example 12.21). However, in Chapter 14 we will see that each combinatorially expanding Thurston map is equivalent to an expanding Thurston map with an invariant curve (Theorem 14.2).

Let $(\mathcal{D}^1, \mathcal{D}^0, L)$ be a two-tile subdivision rule and $\mathcal{C}$ be the Jordan curve of $\mathcal{D}^0$. We will show that if a Thurston map realizing this subdivision rule is combinatorially expanding for $\mathcal{C}$, then this is true for every Thurston map realizing the subdivision rule (Lemma 12.17). In this case, we say that the subdivision rule is *combinatorially expanding* (see Definition 12.18). We will later prove that under an additional hypothesis a two-tile subdivision rule is combinatorially expanding if and only if it can be realized by an expanding Thurston map (Theorem 14.1).

Our definition of combinatorial expansion is set up to be compatible with the description of a Thurston map by a two-tile subdivision rule. It is clearly not invariant under Thurston equivalence, because we require the existence of an invariant Jordan curve. For Thurston maps f with an invariant curve, combinatorial expansion is sufficient for f to be equivalent to an expanding Thurston map (Theorem 14.2). However, this condition not necessary (Example 14.23). We are not aware of a necessary and sufficient condition that is easy to check in practice (see [**HP12b**, Theorem 1.4] for an algebraic condition for Thurston maps without periodic critical points).

This chapter is organized as follows. In Section 12.1 we summarize facts related to Thurston maps with invariant curves and the associated cell decompositions (Proposition 12.5). Then we give yet another

characterization when f is expanding (Lemma 12.7). We show that the quantities $D_n(f, \mathcal{C})$ are supermultiplicative (Lemma 12.8), which is done in preparation of Chapter 16. In Section 12.2 we discuss two-tile subdivision rules and prove Propositions 12.2 and 12.3. For two-tile subdivision rules the information given by an orientation-preserving labeling $L\colon \mathcal{D}^1 \to \mathcal{D}^0$ can be further compressed: for example, it is uniquely determined if one knows the image of one flag in $\mathcal{D}^1$ (see Lemma 12.15 which is based on Lemma 5.23). We will also discuss facts related to combinatorial expansion (such as Lemma 12.17) and conclude the section with a precise version of the statement that the combinatorics of the sequence of cell decompositions $\mathcal{D}^n$ of a Thurston map realizing a subdivision rule is determined by the subdivision rule alone (Proposition 12.19).

Our results pave the way for a convenient construction of Thurston maps from a combinatorial perspective. Section 12.3 is devoted to this. We will exhibit several Thurston maps arising from two-tile subdivision rules.

12.1. Thurston maps with invariant curves

In the following, $f\colon S^2 \to S^2$ is a Thurston map, $\mathcal{C} \subset S^2$ is a Jordan curve with $\operatorname{post}(f) \subset \mathcal{C}$, and $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$ for $n \in \mathbb{N}_0$ is the cell decomposition of S^2 given by the n-cells for $(f, \mathcal{C})$ according to Definition 5.14.

As usual, a set $M \subset S^2$ is called f-*invariant* (or simply *invariant* if f is understood) if

$$(12.1) \qquad f(M) \subset M \quad \text{or equivalently} \quad M \subset f^{-1}(M).$$

We will mostly be interested in the case when $M = \mathcal{C}$ is a Jordan curve with $\operatorname{post}(f) \subset \mathcal{C}$. The reason for this is that then the cell decomposition $\mathcal{D}^n(f, \mathcal{C})$ induced by such an f-invariant Jordan curve $\mathcal{C}$ is refined by each cell decomposition $\mathcal{D}^m(f, \mathcal{C})$ of higher level $m \geq n$ (see Proposition 12.5 below).

Since the set $\operatorname{post}(f)$ is f-invariant, we have

$$(12.2) \qquad \operatorname{post}(f) \subset f^{-1}(\operatorname{post}(f)) \subset f^{-2}(\operatorname{post}(f)) \subset \ldots$$

We know by Proposition 5.17 (iii) that

$$\mathbf{V}^n = \mathbf{V}^n(f, \mathcal{C}) = f^{-n}(\operatorname{post}(f))$$

for $n \in \mathbb{N}_0$, and so (12.2) is equivalent to the inclusions

$$\mathbf{V}^0 \subset \mathbf{V}^1 \subset \mathbf{V}^2 \subset \ldots$$

for the vertex sets of the cell decompositions $\mathcal{D}^n$.

In general, a similar inclusion chain will not hold for the 1-skeleta $E^n := f^{-n}(\mathcal{C})$ of $\mathcal{D}^n$, but if $\mathcal{C}$ is f-invariant, then it follows by induction that

$$\mathcal{C} = E^0 \subset E^1 \subset E^2 \subset \dots$$

The following proposition summarizes the properties of the cell decomposition $\mathcal{D}^n(f, \mathcal{C})$ if $\mathcal{C}$ is f-invariant.

PROPOSITION 12.5. *Let $k, n \in \mathbb{N}_0$, $f\colon S^2 \to S^2$ be a Thurston map, and $\mathcal{C} \subset S^2$ be an f-invariant Jordan curve with* $\mathrm{post}(f) \subset \mathcal{C}$. *Then we have:*

  (i) *$(\mathcal{D}^{n+k}, \mathcal{D}^k)$ is a cellular Markov partition for f^n.*

  (ii) *Every $(n+k)$-tile X^{n+k} is contained in a unique k-tile X^k.*

  (iii) *Every k-tile X^k is equal to the union of all $(n+k)$-tiles X^{n+k} with $X^{n+k} \subset X^k$.*

  (iv) *Every k-edge e^k is equal to the union of all $(n+k)$-edges e^{n+k} with $e^{n+k} \subset e^k$.*

  (v) *Let $c' \subset S^2$ be an n-cell and $c := f^n(c')$. Define*

$$\mathcal{M}' := \{\tau' : \tau' \text{ is an } (n+k)\text{-cell with } \tau' \subset c'\} \text{ and}$$

$$\mathcal{M} := \{\tau : \tau \text{ is a } k\text{-cell with } \tau \subset c\}.$$

  Then $\mathcal{M}'$ and $\mathcal{M}$ are cell decompositions of c' and c, respectively, and the map $\tau' \in \mathcal{M}' \mapsto f^n(\tau')$ is an isomorphism of the cell complexes $\mathcal{M}'$ and $\mathcal{M}$.

If f and $\mathcal{C}$ are as in this proposition, then, in particular, the pair $(\mathcal{D}^1, \mathcal{D}^0)$ is a cellular Markov partition for f by statement (i). If X^n is any n-tile, then by (ii) there exist unique i-tiles X^i for $i = 0, \dots, n-1$ such that

$$X^n \subset X^{n-1} \subset \dots \subset X^0.$$

We refer to the statements (iii) and (iv) informally by saying that tiles and edges are "subdivided" by tiles and edges of higher level. By statement (v) each n-cell c' is subdivided by the $(n+k)$-cells contained in c' "in the same way" as the corresponding 0-cell $c = f^n(c')$ is subdivided by the k-cells contained in c.

PROOF. (i) We know that the map f^n is cellular for $(\mathcal{D}^{n+k}, \mathcal{D}^n)$ (Proposition 5.17 (i)); so we have to show that $\mathcal{D}^{n+k}$ is a refinement of $\mathcal{D}^n$ (see Definition 5.6). By the invariance of $\mathcal{C}$ we have $E^{n+k} = f^{-(n+k)}(\mathcal{C}) \supset E^k = f^{-k}(\mathcal{C})$, and so $S^2 \setminus E^{n+k} \subset S^2 \setminus E^k$.

To establish the first property of a refinement, we will show that every $(n+k)$-cell is contained in some k-tile.

Let σ be an arbitrary $(n + k)$-cell. If σ is an $(n + k)$-tile, then int(σ) is a connected set in $S^2 \setminus E^{n+k} \subset S^2 \setminus E^k$ and hence contained in the interior of a k-tile τ (see Proposition 5.17 (v)). It follows that $\sigma = \overline{\text{int}(\sigma)} \subset \tau$.

If σ is an $(n + k)$-edge or an $(n + k)$-vertex, then it is contained in an $(n + k)$-tile (Lemma 5.9 (iv) and (v)), and hence in some k-tile by what we have just seen.

To establish the second property of a refinement, let τ be an arbitrary k-cell. We have to show that the $(n + k)$-cells σ contained in τ cover τ.

If τ consists of a k-vertex p, then p is also an $(n + k)$-vertex, and the statement is trivial.

If τ is a k-edge, consider the points in $\mathbf{V}^{n+k}$ that lie on τ. Note that this includes the elements of $\partial\tau \subset \mathbf{V}^k \subset \mathbf{V}^{n+k}$. By using these points to partition τ, we can find finitely many arcs $\alpha_1, \ldots, \alpha_N$ such that $\tau = \alpha_1 \cup \cdots \cup \alpha_N$ and such that each arc α_i has endpoints in $\mathbf{V}^{n+k} \supset \mathbf{V}^k$ and interior int(α_i) disjoint from $\mathbf{V}^{n+k}$. Then for each $i = 1, \ldots, N$ the set int(α_i) is a connected set in $E^k \setminus \mathbf{V}^{n+k} \subset E^{n+k} \setminus \mathbf{V}^{n+k}$. It follows that int$(\alpha_i)$ and hence also α_i is contained in some $(n + k)$-edge σ_i (Proposition 5.17 (v)). Since the endpoints of α_i lie in $\mathbf{V}^{n+k}$, they cannot lie in int(σ_i), and so they are also endpoints of σ_i. This implies that $\alpha_i = \sigma_i$. In particular, the $(n + k)$-edges $\sigma_1, \ldots, \sigma_N$ are contained in τ and form a cover of τ. The statement follows in this case.

Finally, let τ be a k-tile. If $p \in \text{int}(\tau)$ is arbitrary, then p is contained in an $(n+k)$-tile σ. By the first part of the proof, σ is contained in a k-cell. Since τ is the only k-cell that contains p, we must have $\sigma \subset \tau$. This implies that the union of the $(n + k)$-tiles contained in τ covers int(τ). On the other hand, this union consists of finitely many tiles and is hence a closed set. It follows that the union also contains $\overline{\text{int}(\tau)} = \tau$.

(ii) We have just seen that every $(n+k)$-tile X^{n+k} is contained in a k-tile X^k. This tile is unique. To see this, suppose that $\widetilde{X}^k$ is another k-tile with $X^{n+k} \subset \widetilde{X}^k$. Then

$$\emptyset \neq \text{int}(X^{n+k}) \subset \text{int}(X^k) \cap \text{int}(\widetilde{X}^k),$$

and so X^k and $\widetilde{X}^k$ have common interior points. This implies $X^k = \widetilde{X}^k$.

(iii)–(iv) Both statements were established in the proof of (i).

(v) Note that under the given assumptions it follows from (i) that $c = f^n(c')$ is a 0-cell. Moreover, again by (i) the cell decomposition $\mathcal{D}^{n+k}$ is a refinement of $\mathcal{D}^n$, and $\mathcal{D}^k$ is a refinement of $\mathcal{D}^0$. This implies that $\mathcal{M}'$ and $\mathcal{M}$ are cell decompositions of c' and c, respectively.

It also follows from (i) that $f^n(\tau') \in \mathcal{M}$ whenever $\tau' \in \mathcal{M}'$. So we can define a map $\varphi \colon \mathcal{M}' \to \mathcal{M}$ by setting $\varphi(\tau') = f^n(\tau)$ for $\tau' \in \mathcal{M}'$. We have to show that this map φ is an isomorphism of cell complexes (see Definition 5.10).

It follows from (i) that $f^n|c'$ is a homeomorphism of the n-cell c' onto the 0-cell c. This implies that φ is injective and that it satisfies the conditions (i) and (ii) in Definition 5.10.

It remains to show that φ is also surjective. To see this, let $\tau \in \mathcal{M}$ be an arbitrary k-cell with $\tau \subset c$. Since $f^n|c'$ is a homeomorphism of c' onto c, the set $\tau' := (f^n|c')^{-1}(\tau) \subset c'$ is a topological cell. Moreover, $f^n|\tau'$ is a homeomorphism of τ' onto the k-cell τ. Lemma 5.18 (i) now implies that τ' is an $(n+k)$-cell, and so $\tau' \in \mathcal{M}'$. Then $\varphi(\tau') = f^n(\tau') = \tau$, and so φ is indeed surjective. □

We illustrate the previous proposition by an example that shows how cells are subdivided.

EXAMPLE 12.6. In Figure 12.1 we indicate the cell decompositions generated by a Thurston map with an invariant Jordan curve and the corresponding subdivision of cells. Here we use the map $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ given by $f(z) = 1 - 2/z^4$. The ramification portrait of f is

$$0 \xrightarrow{\ 4:1\ } \infty \xrightarrow{\ 4:1\ } 1 \longrightarrow -1. \circlearrowleft$$

Thus $\mathrm{post}(f) = \{-1, 1, \infty\}$. Let $\mathcal{C} := \widehat{\mathbb{R}}$ be the extended real line. Clearly, $\mathrm{post}(f) \subset \mathcal{C}$ and $f(\mathcal{C}) \subset \mathcal{C}$, since the coefficients of f are real (indeed $f(\mathcal{C}) = [-\infty, 1]$).

We let the closure of the upper half-plane be the white 0-tile, and the closure of the lower half-plane be the black 0-tile. The intersections of the resulting tiles of levels 0 to 5 with the square $[-2, 2]^2 \subset \mathbb{R}^2 \cong \mathbb{C}$ are shown in Figure 12.1.

The upper and lower half-planes (i.e., the two 0-tiles X_{w}^0 and X_{b}^0 shown on the top left) are both subdivided into four 1-tiles (shown on the top right). Similarly, each n-tile is subdivided into four $(n+1)$-tiles. For emphasis we have marked the boundary of one white 3-tile (shown in the middle right) and the four 4-tiles into which it is subdivided (shown on the bottom left).

Similarly, the 0-edge $[-1, 1] \subset \mathcal{C} = \widehat{\mathbb{R}}$ is subdivided into the two 1-edges $[-1, 0]$ and $[0, 1]$. Note that the 0-edges $[-\infty, -1]$ and $[1, \infty]$ are also 1-edges. So these 0-edges are each replaced by one 1-edge. In the same way, each n-edge is replaced by one or by two $(n+1)$-edges.

Note that it is possible to obtain the $(n+1)$-tiles from the n-tiles in the following way. Given an n-tile X, let X^0 be the 0-tile of the same
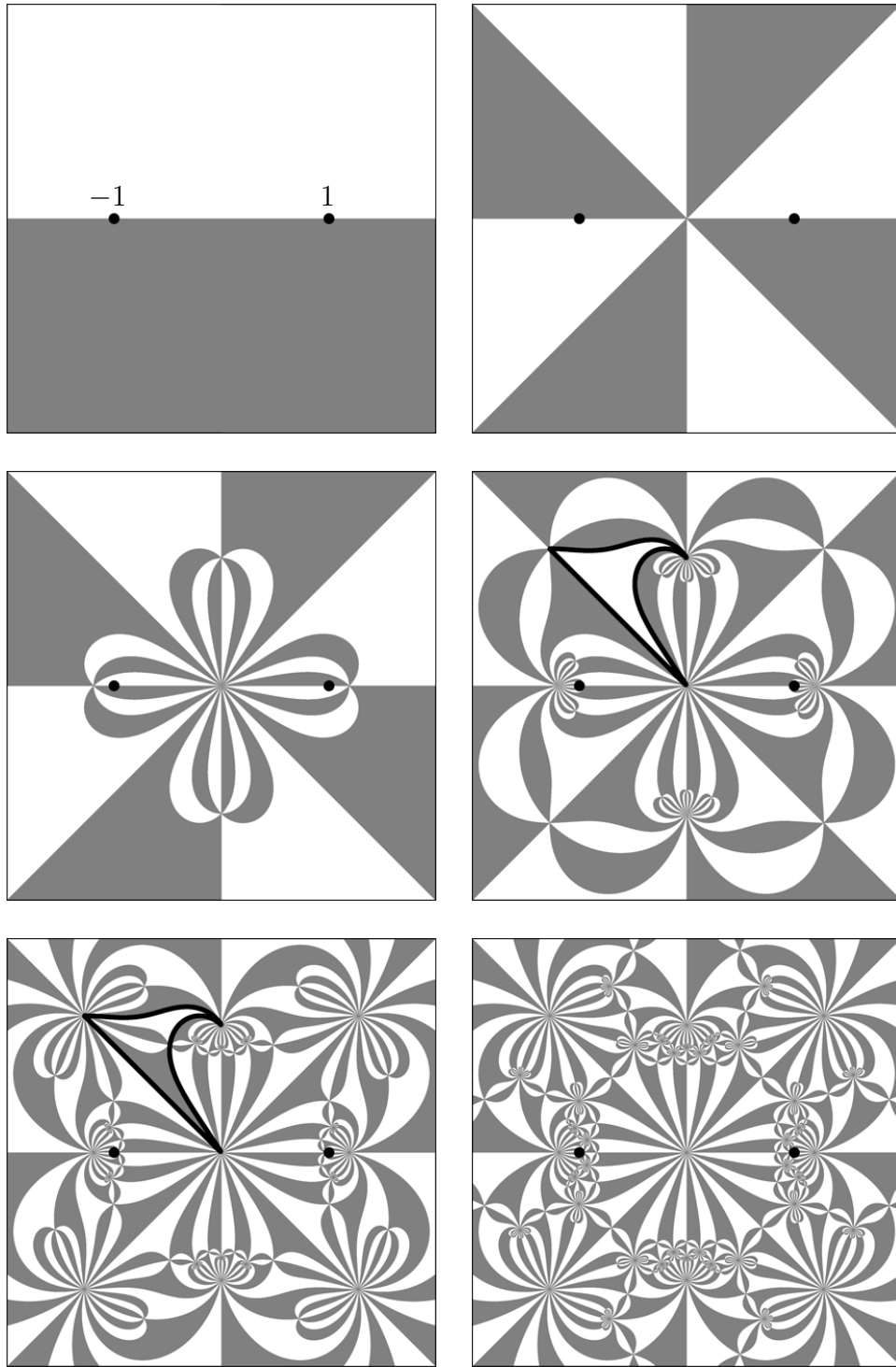
FIGURE 12.1. Subdividing tiles.

color. Then there exists a unique homeomorphism φ from X^0 (the upper or lower half-plane) onto X that is a conformal map between the interiors of these tiles and sends the points $-1, 1, \infty$ (the 0-vertices) to the vertices of X. The map φ sends the four 1-tiles that subdivide X^0 to X. The images of these 1-tiles are the $(n+1)$-tiles into which X is subdivided. Similarly, φ gives a bijection between the 1-edges and 1-vertices contained in X and the $(n+1)$-edges and $(n+1)$-vertices contained in X. Note that φ is determined uniquely by X, once we know the color of X, and the correspondence between the vertices of X and the 0-vertices. We introduced the concept of a labeling (see Section 5.4) to keep track of such information.

As before, let $f\colon S^2 \to S^2$ be a Thurston map and $\mathcal{C} \subset S^2$ be an f-invariant Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}$. We consider cells for $(f, \mathcal{C})$. By $\mathcal{S} = \mathcal{S}(f, \mathcal{C})$ we denote the set of all sequences $\{X^n\}$, where X^n is an n-tile for $n \in \mathbb{N}_0$ and

(12.3) $X^0 \supset X^1 \supset X^2 \supset \ldots$

Since tiles are subdivided by tiles of higher level, for each point $p \in S^2$ we can find a sequence $\{X^n\} \in \mathcal{S}$ such that $p \in \bigcap_n X^n$. Here it is understood that the intersection is taken over all $n \in \mathbb{N}_0$. In the following, we use a similar convention for intersections of sets labeled by some index n, k, etc., if the range of the indices is clear from the context.

In general, a sequence $\{X^n\} \in \mathcal{S}$ that contains a given point $p \in S^2$ is not unique. Moreover, the intersection $\bigcap_n X^n$ may contain more than one point. It turns out that this gives a criterion when f is expanding.

LEMMA 12.7. *Let $f\colon S^2 \to S^2$ be a Thurston map, and $\mathcal{C} \subset S^2$ be an f-invariant Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}$. Then the map f is expanding if and only if for each sequence $\{X^n\} \in \mathcal{S}(f, \mathcal{C})$ the intersection $\bigcap_n X^n$ consists of precisely one point.*

PROOF. Fix a metric on S^2 that induces the given topology on S^2. If f is expanding and $\{X^n\} \in \mathcal{S} = \mathcal{S}(f, \mathcal{C})$, then $\mathrm{diam}(X^n) \to 0$ as $n \to \infty$. Hence $\bigcap_n X^n$ cannot contain more than one point. On the other hand, this set is an intersection of a nested sequence of non-empty compact sets and hence non-empty. So the set $\bigcap_n X^n$ contains precisely one point.

For the converse direction suppose that $\bigcap_n X^n$ is a singleton set for each sequence $\{X^n\} \in \mathcal{S}$. To establish that f is expanding we have to show that

$$\lim_{n \to \infty} \max \{\mathrm{diam}(X) : X \text{ is an } n\text{-tile}\} = 0.$$

We argue by contradiction and assume that this is not the case. Then there exists $\delta > 0$ such that $\operatorname{diam}(X) \geq \delta$ for some tiles X of arbitrarily high level.

We now define a descending sequence of tiles $X^0 \supset X^1 \supset X^2 \supset \dots$ as follows. Let X^0 be a 0-tile such that X^0 contains tiles X of arbitrarily high level with $\operatorname{diam}(X) \geq \delta$. Since the tiles are subdivided by tiles of higher level, and so every tile is contained in one of the finitely many 0-tiles (in our case there are actually two 0-tiles), there exists such a 0-tile. Note that then $\operatorname{diam}(X^0) \geq \delta$. Moreover, among the finitely many 1-tiles into which X^0 is subdivided there must be a 1-tile $X^1 \subset X^0$ such that X^1 contains tiles X of arbitrarily high level with $\operatorname{diam}(X) \geq \delta$. Again this implies that $\operatorname{diam}(X^1) \geq \delta$. Repeating this procedure, we obtain a sequence $\{X^n\} \in \mathcal{S}$ such that $\operatorname{diam}(X^n) \geq \delta$ for all $n \in \mathbb{N}_0$. It is easy to see that this implies that the set $\bigcap_n X^n$ also has diameter $\geq \delta > 0$, and so it contains at least two points. This is a contradiction showing that f is expanding. $\qquad\square$

The main idea in the previous proof is essentially Kőnig's infinity lemma from graph theory (see for example [**Di10**, Lemma 9.1.3]); it says that a (locally finite) simplicial tree with arbitrarily long branches has an infinite branch.

Recall the definition of the numbers $D_n = D_n(f, \mathcal{C})$ in (5.14). We know that $D_n \to \infty$ if f is expanding (see Lemma 8.6). If f is not necessarily expanding, but the Jordan curve $\mathcal{C}$ used in the definition of D_n is invariant, then it follows from the previous discussion that the numbers D_n are non-decreasing, i.e., $D_{n+1} \geq D_n$ for all $n \in \mathbb{N}_0$. The following lemma establishes the much deeper fact that these quantities are actually supermultiplicative (in an appropriate sense). This implies that the numbers D_n increase exponentially fast under the additional assumption that there exists $n_0 \in \mathbb{N}$ with $D_{n_0} \geq 2$ (see Lemma 12.9).

LEMMA 12.8. *Let* $f\colon S^2 \to S^2$ *be a Thurston map with* $\#\operatorname{post}(f) \geq 3$, $\mathcal{C} \subset S^2$ *be a Jordan curve with* $\operatorname{post}(f) \subset \mathcal{C}$, *and* $D_n = D_n(f, \mathcal{C})$ *for* $n \in \mathbb{N}_0$. *Suppose that* $\mathcal{C}$ *is* f-*invariant. Then for all* $n, k \in \mathbb{N}_0$ *we have*

$$D_{n+k} \geq D_n D_k$$

if $\#\operatorname{post}(f) \geq 4$, *and*

(12.4) $$D_{n+k} \geq D_n(D_k - 1) + 1$$

if $\#\operatorname{post}(f) = 3$.

In the proof of this lemma we will use n-chains. Recall from Definition 5.19 that such an n-chain is a finite sequence of n-tiles $X_1, \dots, X_N$ with $X_i \cap X_{i+1} \neq \emptyset$ for $i = 1, \dots, N-1$.
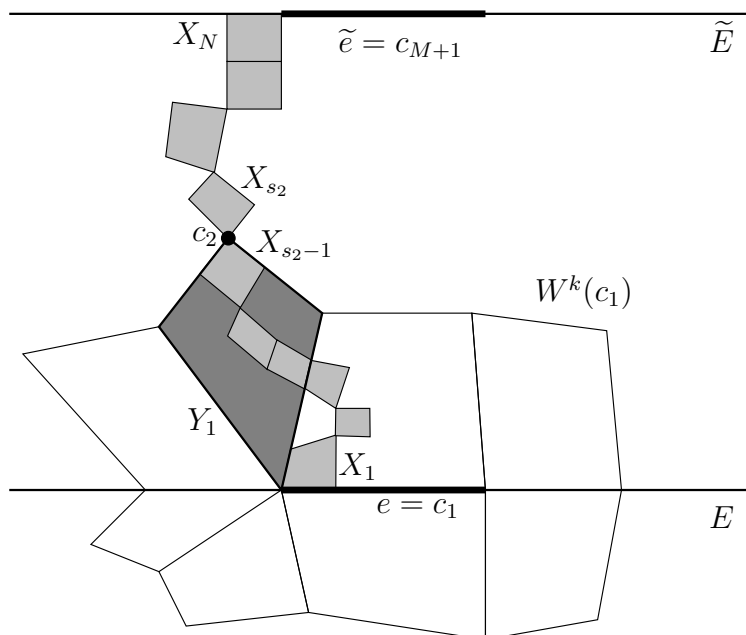
FIGURE 12.2. The proof of Lemma 12.8.

PROOF OF LEMMA 12.8. *Case 1:* $\# \operatorname{post}(f) \geq 4$. Let $X_1, \ldots, X_N$ be a set of $(n+k)$-tiles whose union is connected and joins opposite sides of $\mathcal{C}$. We may assume that these tiles form a chain joining disjoint 0-edges E and $\widetilde{E}$. To prove the desired inequality, we will break this chain into M subchains $X_{s_i}, \ldots, X_{s_{i+1}-1}$, where $M \in \mathbb{N}$, $i = 1, \ldots, M$, and $s_1 = 1 < s_2 < \cdots < s_{M+1} = N + 1$. The length of each subchain (i.e., the number $s_{i+1} - s_i$) will be at least D_n. The number M of subchains will be at least D_k. Thus $N \geq D_n D_k$, and since the minimum over all N is equal to D_{n+k}, the desired inequality follows.

To guarantee the desired lower bound on the length, we will ensure that each subchain $X_{s_i}, \ldots, X_{s_{i+1}-1}$ joins disjoint k-cells. Then the length of such a subchain is at least D_n by Lemma 5.36.

To control the number of subchains, we will associate with each one a k-tile Y_i. These k-tiles $Y_1, \ldots, Y_M$ will form a k-chain joining E and $\widetilde{E}$, and hence opposite sides of $\mathcal{C}$. Thus M, which is the number of k-tiles in this chain, as well as the number of subchains, is at least D_k (by definition of this quantity; see (5.14)).

We now provide the details of the construction, which is illustrated in Figure 12.2. We will use auxiliary k-cells $c_1, c_2, \ldots$ of dimension ≤ 1. If c_i is 0-dimensional, then c_i consists of a k-vertex p_i, and we let

$W^k(c_i) := W^k(p_i)$ (see Definition 5.27). If c_i is 1-dimensional, then c_i is a k-edge and $W^k(c_i)$ is the edge flower of c_i as in Definition 5.30.

Since $\mathcal{C}$ is f-invariant, the cell decomposition $\mathcal{D}^k$ is a refinement of $\mathcal{D}^0$. Hence there exist disjoint k-edges $e \subset E$ and $\widetilde{e} \subset \widetilde{E}$ with $X_1 \cap e \neq \emptyset$ and $X_N \cap \widetilde{e} \neq \emptyset$.

For some number $M \in \mathbb{N}$ we will now inductively define k-cells $c_1, \ldots, c_{M+1}$ of dimension ≤ 1, k-tiles $Y_1, \ldots, Y_M$, and indices $s_1 = 1 < s_2 < \cdots < s_{M+1} = N + 1$ with the following properties:

(i) $c_1 = e$, $c_{M+1} = \widetilde{e}$, and $c_i \cap c_{i+1} = \emptyset$ for $i = 1, \ldots, M$.

(ii) $c_i \cap Y_i \neq \emptyset$ for $i = 1, \ldots, M$, $c_{i+1} \subset \partial Y_i$ for $i = 1, \ldots, M-1$, and $\widetilde{e} \cap Y_M \neq \emptyset$.

(iii) $X_{s_i}, \ldots, X_{s_{i+1}-1}$ is an $(n + k)$-chain joining c_i and c_{i+1} for $i = 1, \ldots, M$.

Note that (i) and (ii) imply that $E \cap Y_1 \supset e \cap Y_1 \neq \emptyset$, $\widetilde{E} \cap Y_M \supset \widetilde{e} \cap Y_M \neq \emptyset$, and $Y_i \cap Y_{i+1} \supset c_{i+1} \cap Y_{i+1} \neq \emptyset$ for $i = 1, \ldots, M-1$. Hence $Y_1, \ldots, Y_M$ will be a k-chain joining the 0-edges E and $\widetilde{E}$ as desired.

Let $s_1 = 1$ and $c_1 = e$. Suppose first that $\widetilde{e}$ meets $\overline{W^k(c_1)}$. Since $\widetilde{e}$ is disjoint from $e = c_1$ and hence from $W^k(c_1)$, the points in $\widetilde{e} \cap \overline{W^k(c_1)}$ lie in $\partial W^k(c_1)$. By Lemma 5.31 (ii) there exists a k-tile Y_1 that meets both c_1 and $\widetilde{e} \supset \widetilde{e} \cap \overline{W^k(c_1)}$. We let $M = 1$, set $c_2 = \widetilde{e}$, and stop the construction. We have all the desired properties (i)–(iii).

In the other case where $\widetilde{e} \cap \overline{W^k(c_1)} = \emptyset$ not all the $(n + k)$-tiles $X_1, \ldots, X_N$ are contained in $\overline{W^k(c_1)}$. So there exists a smallest index $s_2 \geq 1$ such that X_{s_2} meets $S^2 \setminus \overline{W^k(c_1)}$. Then $s_2 > s_1 = 1$, because X_1 meets $e = c_1$ and is hence contained in $\overline{W^k(c_1)}$. To see this, we use Lemma 5.31 (iii) and the fact that X_1 is contained in some k-tile. Moreover, for a similar reason we have $X_{s_2} \subset S^2 \setminus W^k(c_1)$. By definition of s_2 the set X_{s_2-1} is contained in $\overline{W^k(c_1)}$. Hence every point in the non-empty intersection $X_{s_2-1} \cap X_{s_2}$ lies in $\partial W^k(c_1)$. Note that the $(n + k)$-tiles X_{s_2-1} and X_{s_2} do not necessarily meet in a k-vertex; but by Lemma 5.31 (ii) there exists a k-cell $c_2 \subset \partial W^k(c_1)$ of dimension ≤ 1 that has common points with both X_{s_2-1} and X_{s_2}, and a k-tile $Y_1 \subset \overline{W^k(c_1)}$ with $c_1 \cap Y_1 \neq \emptyset$ and $c_2 \subset \partial Y_1$. Then $c_1 \cap c_2 = \emptyset = c_2 \cap \widetilde{e}$, and the chain $X_{s_1} = X_1, \ldots, X_{s_2-1}$ joins c_1 and c_2.

We can now repeat the construction as in the first step by using the chain $X_{s_2}, \ldots, X_N$ that joins the disjoint k-cells c_2 and $\widetilde{e}$, etc. If in the process one of the cells c_i has dimension 0, we invoke Lemma 5.28 (ii) and (iii) instead of Lemma 5.31 (ii) and (iii) in the above construction. The construction eventually stops, and it is clear that we obtain cells

and indices with the desired properties. The statement follows in this
case.

Case 2: $\# \operatorname{post}(f) = 3$. Let E_1, E_2, E_3 be the three 0-edges. Consider a connected union K of $(n+k)$-tiles joining opposite sides of $\mathcal{C}$
with $N = D_{n+k}$ elements. Then K meets k-edges contained in the
0-edges, say k-edges $e_i \subset E_i$ for $i = 1, 2, 3$. From K we can extract a
simple $(n+k)$-chain joining e_1 and e_2 as well as another simple chain
that joins e_3 to one tile X in the chain joining e_1 and e_2. Starting
from this "center tile" X, we can find three simple $(n+k)$-chains that
join X to the edges e_1, e_2, e_3, respectively, and have only the tile X in
common.

More precisely, for $i = 1, 2, 3$ we can find $N_i \in \mathbb{N}_0$ and $(n+k)$-chains
$X, X_1^i, \ldots, X_{N_i}^i$ that join X and e_i. Here the first tile X is the same
in all chains and it is understood that the chain consists only of X if
$N_i = 0$. Moreover, all the $(n+k)$-tiles

$$X, X_1^1, \ldots, X_{N_1}^1, X_1^2, \ldots, X_{N_2}^2, X_1^3, \ldots, X_{N_3}^3$$

are pairwise distinct tiles from K. Thus their number is bounded by
the number of $(n+k)$-tiles in K. Since they still form a connected set
joining opposite sides of $\mathcal{C}$, we have $N_1 + N_2 + N_3 + 1 = N = D_{n+k}$.

Let Y be the unique k-tile with $X \subset Y$, and consider the chain
$X, X_1^1, \ldots, X_{N_1}^1$. Suppose that $Y \cap e_1 = \emptyset$. Since $X \subset Y$, we have
$N_1 \geq 1$, and the chain $X_1, \ldots, X_{N_1}$ joins Y and e_1. Hence this chain
or a subchain must also join a k-edge $e \subset \partial Y$ and e_1. Then $e \cap e_1 = \emptyset$.
As in the first part of the proof, we can find k-tiles $Y_1, \ldots, Y_{M_1}$ joining
e and e_1, where $M_1 \in \mathbb{N}$ and $N_1 \geq M_1 D_n$.

If $Y \cap e_1 \neq \emptyset$, we set $M_1 = 0$ and do not define new k-tiles. In any
case we have that $Y, Y_1^1, \ldots, Y_{M_1}^1$ is a chain joining Y and e_1 (again we
use the convention that this chain consists only of Y if $M_1 = 0$). We
also have $N_1 \geq M_1 D_n$ (which is trivial if $M_1 = 0$).

Using a similar construction for the other indices $i = 2, 3$, we obtain
numbers $M_i \in \mathbb{N}_0$ for each $i = 1, 2, 3$ that satisfy $N_i \geq M_i D_n$, and
chains $Y, Y_1^i, \ldots, Y_{M_i}^i$ of k-tiles that join Y and e_i. The union of these
k-tiles is a connected set joining opposite sides of $\mathcal{C}$. Therefore, it
contains at least D_k distinct elements. On the other hand, the number
of distinct k-tiles in the union is at most $M_1 + M_2 + M_3 + 1$ (note
that the three chains may have other k-tiles in common apart from Y).
Hence $D_k \leq M_1 + M_2 + M_3 + 1$, and it follows that

$$D_n(D_k - 1) + 1 \leq D_n(M_1 + M_2 + M_3) + 1$$
$$\leq N_1 + N_2 + N_3 + 1 \leq D_{n+k},$$

which is the desired inequality (12.4). $\qquad\square$

LEMMA 12.9. *Let $f\colon S^2 \to S^2$ be a Thurston map with $\#\operatorname{post}(f) \geq 3$, $\mathcal{C} \subset S^2$ be a Jordan curve with $\operatorname{post}(f) \subset \mathcal{C}$, and $D_n = D_n(f, \mathcal{C})$ for $n \in \mathbb{N}_0$. Suppose that $\mathcal{C}$ is f-invariant. Then the limit*

$$\Lambda_0 = \lim_{n\to\infty} D_n^{1/n}$$

exists and $\Lambda_0 = \sup_{n\in\mathbb{N}} D_n^{1/n} \leq \deg(f)$.

If in addition there exists $n_0 \in \mathbb{N}$ with $D_{n_0} \geq 2$, then $\Lambda_0 > 1$ and $D_n \to \infty$ as $n \to \infty$.

The lemma shows that if f is combinatorially expanding for $\mathcal{C}$, then $\Lambda_0 > 1$. So if $\Lambda \in (1, \Lambda_0)$ is arbitrary, then $D_n \gtrsim \Lambda^n$ for all n.

We will see later (Proposition 16.1) that if f is expanding, but $\mathcal{C}$ is not necessarily f-invariant, then the limit $\Lambda_0 = \lim_{n\to\infty} D_n(f, \mathcal{C})^{1/n}$ still exists and is independent of $\mathcal{C}$. Moreover, the improved estimate $\Lambda_0 \leq \deg(f)^{1/2}$ holds (see Proposition 20.1).

PROOF. We use Lemma 12.8. First note that inequality (12.4) is also true if $\operatorname{post}(f) = 4$. The statements now essentially follow from Fekete's lemma (see [**KH95**, Proposition 9.6.4]). We provide the details for the convenience of the reader.

A simple induction argument using (12.4) shows that if $D_N \geq 2$ for some $N \in \mathbb{N}$, then

$$(12.5) \qquad\qquad D_{kN} \geq D_N^{k-1} + 1$$

for all $k \in \mathbb{N}$. For such N let $k(n) = \lfloor n/N \rfloor$. Noting that the sequence $\{D_n\}$ is non-decreasing and using (12.5), we obtain

$$
\begin{aligned}
\liminf_{n\to\infty} \frac{1}{n} \log(D_n) &\geq \liminf_{n\to\infty} \frac{1}{n} \log(D_{k(n)N}) \\
&\geq \liminf_{n\to\infty} \frac{k(n)-1}{n} \log(D_N) = \frac{1}{N} \log(D_N).
\end{aligned}
$$

This inequality is trivially true if $D_N = 1$, and so

$$\liminf_{n\to\infty} \frac{1}{n} \log(D_n) \geq \sup_{n\in\mathbb{N}} \frac{1}{n} \log(D_n).$$

On the other hand,

$$\limsup_{n\to\infty} \frac{1}{n} \log(D_n) \leq \sup_{n\in\mathbb{N}} \frac{1}{n} \log(D_n),$$

and so the limit

$$\alpha := \lim_{n\to\infty} \frac{1}{n} \log(D_n) \quad \text{exists and} \quad \alpha = \sup_{n\in\mathbb{N}} \frac{1}{n} \log(D_n).$$

Note that $D_n \leq \#\mathbf{X}^n(f, \mathcal{C}) \leq 2\deg(f)^n$ which implies $\alpha \leq \log(\deg(f))$. The first part of the statement now follows by taking exponentials.

For the last part suppose that there exists $n_0 \in \mathbb{N}$ with $D_{n_0} \geq 2$. Then

$$\Lambda_0 = \sup_{n \in \mathbb{N}} D_n^{1/n} \geq D_{n_0}^{1/n_0} > 1,$$

and it is clear from the definition of Λ_0 as a limit that $D_n \to \infty$ as $n \to \infty$. $\qquad\square$

We conclude this section with a lemma that shows that if a Thurston map is combinatorially expanding, then there are points that lie "deep inside" a given edge or tile. It is related to Lemma 8.11 which implies a similar statement for expanding Thurston maps.

LEMMA 12.10. *Let* $f\colon S^2 \to S^2$ *be a Thurston map, and* $\mathcal{C} \subset S^2$ *be an* f-*invariant Jordan curve with* $\mathrm{post}(f) \subset \mathcal{C}$. *Suppose that* $D_{n_0}(f, \mathcal{C}) \geq 2$ *for some* $n_0 \in \mathbb{N}$.

  (i) *If* $n \in \mathbb{N}_0$ *and* e *is an* n-*edge, then there exists an* $(n + n_0)$-*vertex* p *with* $p \in \mathrm{int}(e)$.

  (ii) *If* $n \in \mathbb{N}_0$ *and* X *is an* n-*tile, then there exists an* $(n + n_0)$-*edge with* $\mathrm{int}(e) \subset \mathrm{int}(X)$, *and an* $(n + 2n_0)$-*vertex* p *with* $p \in \mathrm{int}(X)$.

PROOF. In the previous statements and the ensuing proof it is understood that the term k-cell for $k \in \mathbb{N}_0$ refers to a cell in $\mathcal{D}^k = \mathcal{D}^k(f, \mathcal{C})$.

(i) Suppose e is an n-edge that does not contain $(n + n_0)$-vertices in its interior. By Proposition 12.5 (iv) the n-edge e is equal to the union of all $(n + n_0)$-edges contained in e. Thus e must be an $(n + n_0)$-edge itself. Let u and v be the endpoints of e, and X be an $(n + n_0)$-tile containing e in its boundary. Then $K = X$ meets the two disjoint n-cells $\{u\}$ and $\{v\}$. Hence by Lemma 5.36 the set K should consist of at least $D_{n_0} = D_{n_0}(f, \mathcal{C}) \geq 2$ $(n + n_0)$-tiles. This is a contradiction proving the statement.

(ii) Let X be an n-tile. By Proposition 12.5 (iii) we know that X is the union of all $(n + n_0)$-tiles contained in X. In particular, there exists an $(n + n_0)$-tile Y with $Y \subset X$. We claim that there exists an $(n + n_0)$-edge in the boundary of Y that meets $\mathrm{int}(X)$. Otherwise, $\partial Y \cap \mathrm{int}(X) = \emptyset$, and as $Y \subset X$, we must have $\partial Y \subset \partial X$. Since both sets ∂Y and ∂X are Jordan curves, this is only possible if $\partial Y = \partial X$. Then Y meets all n-vertices contained in ∂X, and two distinct n-vertices in particular. As in the proof of (i), this leads to a contradiction.
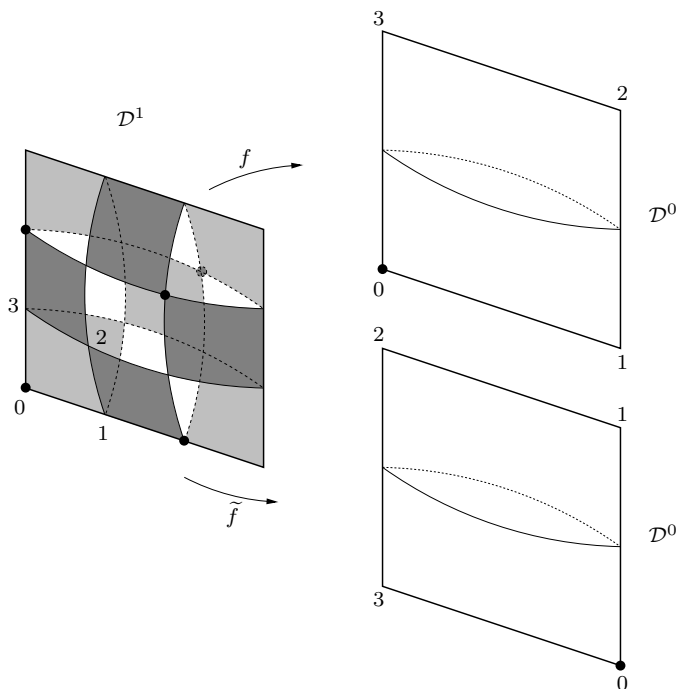
FIGURE 12.3. Two two-tile subdivision rules.

Hence there exists an $(n + n_0)$-edge e with $e \cap \operatorname{int}(X) \neq \emptyset$. Since $\operatorname{int}(X)$ is an open subset of S^2, we then also have $\operatorname{int}(e) \cap \operatorname{int}(X) \neq \emptyset$. Since $\mathcal{D}^{n+n_0}$ is a refinement of $\mathcal{D}^n$, by Lemma 5.7 we know that there is a unique cell τ in $\mathcal{D}^n$ with $\operatorname{int}(e) \subset \operatorname{int}(\tau)$. Then $\operatorname{int}(\tau) \cap \operatorname{int}(X) \neq \emptyset$, and so $X = \tau$ by Lemma 5.2. Hence $\operatorname{int}(e) \subset \operatorname{int}(X)$ as desired.

By (i) there exists an $(n + 2n_0)$-vertex p with $p \in \operatorname{int}(e)$. Then we also have $p \in \operatorname{int}(X)$ as desired. $\qquad \square$

12.2. Two-tile subdivision rules

In this section we consider a 2-sphere S^2 and two given cell decompositions $\mathcal{D}^0$ and $\mathcal{D}^1$ of S^2. We call the cells in $\mathcal{D}^0$ the 0-cells and the cells in $\mathcal{D}^1$ the 1-cells. Similarly, we refer to the tiles in $\mathcal{D}^0$ as the 0-tiles, the edges in $\mathcal{D}^1$ as the 1-edges, etc. We know by Proposition 5.26 that under suitable additional assumptions for such a pair $(\mathcal{D}^1, \mathcal{D}^0)$ there exists a postcritically-finite branched covering map $f \colon S^2 \to S^2$ that is cellular for $(\mathcal{D}^1, \mathcal{D}^0)$. The map f is unique up to Thurston equivalence if additional data is provided, namely a labeling $L \colon \mathcal{D}^1 \to \mathcal{D}^0$. The following simple example illustrates that the pair $(\mathcal{D}^1, \mathcal{D}^0)$ alone without the labeling is in general not enough to determine the map f.

EXAMPLE 12.11. In Figure 12.3 we indicated two-tile subdivision rules $(\mathcal{D}^1, \mathcal{D}^0, L)$ and $(\mathcal{D}^1, \mathcal{D}^0, \widetilde{L})$ on a sphere S^2 given as a pillow obtained by gluing two copies of a square together along their boundaries. The cell decomposition $\mathcal{D}^0$ is shown twice on the right of the figure with the two squares as the 0-tiles and their common sides and corners as 0-edges and 0-vertices, respectively. Each of the two 0-tiles is subdivided into nine squares of equal size. From this one obtains a refinement $\mathcal{D}^1$ of $\mathcal{D}^0$ as indicated on the left of the figure. We color the tiles black and white so that $\mathcal{D}^1$ and $\mathcal{D}^0$ are checkerboard tilings. Here the top square of the pillow is white and the colors of the 1-tiles are as indicated on the left of Figure 12.3.

There are unique orientation-preserving labelings L and $\widetilde{L}$ that send 1-tiles to 0-tiles of the same colors and the 1-vertices marked as a black dot on the left to the one 0-vertex marked in the same way in the two representations of $\mathcal{D}^0$ on the right (one can easily see this directly or derive it as a special case of Lemma 12.15 (ii) below). In the figure we also show additional markings of some corresponding vertices for better illustration.

In this way we get two subdivision rules $(\mathcal{D}^1, \mathcal{D}^0, L)$ and $(\mathcal{D}^1, \mathcal{D}^0, \widetilde{L})$ realized by Thurston maps f and $\widetilde{f}$ as indicated. The maps are uniquely determined if we require in addition that they are piecewise scaling maps on 1-tiles. The map $\widetilde{f}$ assigns the same 0-tile to each 1-tile as f, followed by an additional rotation.

The maps f and $\widetilde{f}$ are not Thurston equivalent. In fact, every postcritical point for f is a fixed point, whereas no postcritical point for $\widetilde{f}$ is a fixed point (each postcritical point for $\widetilde{f}$ is periodic with period 4). Note that both maps are (Thurston equivalent to) Lattès maps.

As we know from Section 5.3, every Thurston map $f\colon S^2 \to S^2$ arises as a cellular map for a pair of cell decompositions $\mathcal{D}^1$ and $\mathcal{D}^0$ of S^2. This gives a useful description of a Thurston map in combinatorial terms. If one wants to study the dynamics of f, one is interested in the cell decompositions $\mathcal{D}^n$ obtained from pulling back $\mathcal{D}^0$ by f^n as in Lemma 5.12. In general, in order to determine the combinatorics of the whole sequence $\mathcal{D}^n$, $n \in \mathbb{N}_0$ (i.e., the inclusion and intersection patterns of cells of possibly distinct levels), is not enough to just know the pair $(\mathcal{D}^1, \mathcal{D}^0)$ and the labeling $\tau \in \mathcal{D}^1 \mapsto f(\tau) \in \mathcal{D}^0$, but one also needs specific information on the *pointwise* mapping behavior of f on the cells in $\mathcal{D}^1$. Indeed, suppose g is another map that is cellular for $(\mathcal{D}^1, \mathcal{D}^0)$ and induces the same labeling as f, i.e., $f(\tau) = g(\tau)$ for all $\tau \in \mathcal{D}^1$. Let $\widetilde{\mathcal{D}}^n$ be the cell decomposition of S^2 obtained from $\mathcal{D}^0$ by

pulling back by g^n. Then one can show (by an argument very similar to the considerations in the proof of Lemma 12.16 below) that $\mathcal{D}^n$ and $\widetilde{\mathcal{D}}^n$ are isomorphic cell complexes (see Definition 5.10) for *fixed* $n \in \mathbb{N}_0$. In contrast, the intersection patterns of corresponding cells $\sigma \in \mathcal{D}^n$, $\tau \in \mathcal{D}^m$ and $\widetilde{\sigma} \in \widetilde{\mathcal{D}}^n$, $\widetilde{\tau} \in \widetilde{\mathcal{D}}^m$, for different levels $n, m \in \mathbb{N}_0$, may not be the same.

The situation changes if $(\mathcal{D}^1, \mathcal{D}^0)$ is a cellular Markov partition for f, because then the combinatorics of the sequence $\mathcal{D}^0, \mathcal{D}^1, \mathcal{D}^2, \ldots$ is completely determined by $(\mathcal{D}^1, \mathcal{D}^0)$ and the combinatorial data given by the labeling $\tau \in \mathcal{D}^1 \mapsto f(\tau) \in \mathcal{D}^0$ (see Remark 12.12 (ii) and Proposition 12.19). This suggests that if one wants to study Thurston maps as given by Proposition 5.26 from a purely combinatorial point of view, then one should add the additional assumption that $\mathcal{D}^1$ is a refinement of $\mathcal{D}^0$. If we restrict ourselves to the case where $\mathcal{D}^0$ contains only two tiles, then we are led to the concept of a two-tile subdivision rule as defined in the introduction of this chapter.

The proofs of Propositions 12.2 and 12.3 are easy consequences of our previous considerations.

PROOF OF PROPOSITION 12.2. Suppose f, $\mathcal{C}$, $\mathcal{D}^0$, $\mathcal{D}^1$, and L are as in the statement. Then it follows from Proposition 5.17 (i), Proposition 12.5 and the discussion after Definition 12.1 that $(\mathcal{D}^0, \mathcal{D}^1, L)$ is a two-tile subdivision rule realized by f. $\qquad\square$

PROOF OF PROPOSITION 12.3. The first part is just a special case of Proposition 5.26. Note that f is a Thurston map; indeed, the number of 1-tiles is equal to $2\deg(f)$, and also > 2 by Definition 12.1 (ii). So $\deg(f) \geq 2$.

Since $\mathcal{D}^1$ is a refinement of $\mathcal{D}^0$, the 1-skeleton $\mathcal{C}$ of $\mathcal{D}^0$ is contained in the 1-skeleton of $\mathcal{D}^1$. Moreover, since f is cellular for $(\mathcal{D}^1, \mathcal{D}^0)$, this map sends the 1-skeleton of $\mathcal{D}^1$ into the 1-skeleton of $\mathcal{D}^0$. Hence $f(\mathcal{C}) \subset \mathcal{C}$, and so $\mathcal{C}$ is f-invariant. Each postcritical point of f is a vertex of $\mathcal{D}^0$ and hence contained in $\mathcal{C}$. $\qquad\square$

REMARK 12.12. Let $(\mathcal{D}^1, \mathcal{D}^0, L)$ be a two-tile subdivision rule on S^2 and $f\colon S^2 \to S^2$ be a Thurston map that realizes $(\mathcal{D}^1, \mathcal{D}^0, L)$ according to Proposition 12.3.

(i) If $\mathbf{V}^0$ denotes the set of vertices of $\mathcal{D}^0$ and $\mathbf{V}^1$ the set of vertices $\mathcal{D}^1$, we then have $\operatorname{crit}(f) \subset \mathbf{V}^1$ and $\operatorname{post}(f) \subset \mathbf{V}^0$ (see Lemma 5.24). In general, $\operatorname{post}(f) \neq \mathbf{V}^0$ (see Example 14.22).

Since the length of each cycle in $\mathcal{D}^0$ is 2, a vertex v in $\mathbf{V}^1$ is a critical point of f if and only if the length of the cycle of v in $\mathcal{D}^1$ is ≥ 4 (see Remark 5.25). Hence if f and g both realize the subdivision

rule, then $\mathrm{crit}(f) = \mathrm{crit}(g) \subset \mathbf{V}^1$. Moreover, since the orbit of any point in $\mathbf{V}^1$ is completely determined by the labeling, we then also have $\mathrm{post}(f) = \mathrm{post}(g) \subset \mathbf{V}^0$.

Let $\mathcal{C}$ be the Jordan curve of $\mathcal{D}^0$, (i.e., the 1-skeleton of $\mathcal{D}^0$), and $\mathcal{D}^n(f,\mathcal{C})$ for $n \in \mathbb{N}_0$ be the cell decomposition defined according to Definition 5.14. Then $\mathcal{D}^0 = \mathcal{D}^0(f,\mathcal{C})$ if and only if $\mathbf{V}^0 = \mathrm{post}(f)$. If this is not true, then the points in $\mathbf{V}^0 \setminus \mathrm{post}(f)$ are additional points that are not distinguished by the dynamics of the map f. In view of this, it seems to make sense to include a requirement that forces $\mathbf{V}^0 = \mathrm{post}(f)$ in the definition of a two-tile subdivision, but we chose not to do so in order to keep the definition a little simpler.

(ii) Let us now assume that $\mathbf{V}^0 = \mathrm{post}(f)$. In this case, $\mathcal{D}^0 = \mathcal{D}^0(f,\mathcal{C})$, and also $\mathcal{D}^1 = \mathcal{D}^1(f,\mathcal{C})$ as follows from the uniqueness statement in Lemma 5.12.

By Proposition 12.3 we know that $\mathcal{C}$ is f-invariant and $\mathrm{post}(f) \subset \mathcal{C}$. This means that the cell decompositions $\mathcal{D}^n := \mathcal{D}^n(f,\mathcal{C})$, $n \in \mathbb{N}_0$, satisfy the properties listed in Proposition 12.5. In particular, $\mathcal{D}^{n+k}$ is a refinement of $\mathcal{D}^n$ for $k,n \in \mathbb{N}_0$.

By Proposition 12.5 (v) each 1-cell c' is subdivided into $(n+1)$-cells "in the same way" as the corresponding 0-cell $f(c') = L(c')$ is subdivided into n-cells. From this it is intuitively clear that the "combinatorics" of the cells in the sequence $\mathcal{D}^n$, $n \in \mathbb{N}_0$, that is, their inclusion and intersection patterns, can be determined inductively from the subdivision rule $(\mathcal{D}^1, \mathcal{D}^0, L)$ independently of the map realizing it.

To make this more precise, suppose $g \colon S^2 \to S^2$ is another Thurston map realizing the subdivision rule. We denote by $\mathcal{D}^\infty$ the disjoint union of the cell decompositions $\mathcal{D}^n$, $n \in \mathbb{N}_0$, i.e., $\mathcal{D}^\infty$ is the set of all cell for $(f,\mathcal{C})$, where cells with the same underlying sets, but of different level, are considered distinct. Similarly, let $\widetilde{\mathcal{D}}^\infty$ be the set of all cells for $(g,\mathcal{C})$. Then there exists a bijection $\varphi \colon \mathcal{D}^\infty \to \widetilde{\mathcal{D}}^\infty$ that preserves the level and the dimension of each cell, and all inclusion patterns (i.e., $\sigma \subset \tau$ for $\sigma, \tau \in \mathcal{D}^\infty$ if and only if $\varphi(\sigma) \subset \varphi(\tau)$). This last property of φ implies that also all intersection patterns are preserved (i.e., $\sigma \cap \tau \neq \emptyset$ for $\sigma, \tau \in \mathcal{D}^\infty$ if and only if $\varphi(\sigma) \cap \varphi(\sigma) \neq \emptyset$). In this sense, $\mathcal{D}^\infty$ and $\widetilde{\mathcal{D}}^\infty$ have exactly the same combinatorics.

This will not be proved here, but we refer to Proposition 12.19 for a related statement.

We will later need a more general version of the uniqueness part of Proposition 12.3. To formulate this, we first introduce a suitable notion of an *isomorphism* between two-tile subdivision rules.

Let $(\mathcal{D}^1, \mathcal{D}^0, L)$ and $(\widetilde{\mathcal{D}}^1, \widetilde{\mathcal{D}}^0, \widetilde{L})$ be two-tile subdivision rules on 2-spheres S^2 and $\widetilde{S}^2$, respectively. We say that these subdivision rules are *isomorphic* if there exist cell complex isomorphisms $\phi_i \colon \mathcal{D}^i \to \widetilde{\mathcal{D}}^i$ for $i = 0, 1$ such that $\widetilde{L}(\phi_1(\tau)) = \phi_0(L(\tau))$ for $\tau \in \mathcal{D}^1$ and such that $\sigma \subset \tau$ for $\sigma \in \mathcal{D}^1$, $\tau \in \mathcal{D}^0$ if and only if $\phi_1(\sigma) \subset \phi_0(\tau)$.

If, by abuse of notation, we denote the image of a cell $\tau \in \mathcal{D}^i$ under ϕ_i by $\widetilde{\tau}$ for $i = 0, 1$, then the last two conditions require that $\widetilde{L}(\widetilde{\tau}) = \widetilde{L(\tau)}$ for $\tau \in \mathcal{D}^1$ and that $\sigma \subset \tau$ for $\sigma \in \mathcal{D}^1$, $\tau \in \mathcal{D}^0$ if and only if $\widetilde{\sigma} \subset \widetilde{\tau}$. So $(\mathcal{D}^1, \mathcal{D}^0, L)$ and $(\widetilde{\mathcal{D}}^1, \widetilde{\mathcal{D}}^0, \widetilde{L})$ are isomorphic if the combinatorics of cells under the correspondence $\tau \leftrightarrow \widetilde{\tau}$ is the same and if this correspondence is also compatible with the labelings.

LEMMA 12.13. *Let $(\mathcal{D}^1, \mathcal{D}^0, L)$ and $(\widetilde{\mathcal{D}}^1, \widetilde{\mathcal{D}}^0, \widetilde{L})$ be two-tile subdivision rules on 2-spheres S^2 and $\widetilde{S}^2$ that are isomorphic. If the Thurston map $f \colon S \to S^2$ realizes $(\mathcal{D}^1, \mathcal{D}^0, L)$ and the Thurston map $\widetilde{f} \colon \widetilde{S}^2 \to \widetilde{S}^2$ realizes $(\widetilde{\mathcal{D}}^1, \widetilde{\mathcal{D}}^0, \widetilde{L})$, then f and $\widetilde{f}$ are Thurston equivalent.*

PROOF. The proof uses very similar ideas as the proof of the uniqueness part of Proposition 5.26.

By assumption, there exist cell complex isomorphisms $\phi_i \colon \mathcal{D}^i \to \widetilde{\mathcal{D}}^i$ for $i = 0, 1$ with the following properties: $\widetilde{L}(\widetilde{\tau}) = \widetilde{L(\tau)}$ for each $\tau \in \mathcal{D}^1$, and for $\sigma \in \mathcal{D}^1$, $\tau \in \mathcal{D}^0$ we have the inclusion $\sigma \subset \tau$ if and only if $\widetilde{\sigma} \subset \widetilde{\tau}$. Here we denote the image of a cell $\tau \in \mathcal{D}^i$ under ϕ_i by $\widetilde{\tau}$ for $i = 0, 1$.

By Lemma 5.11 (ii) there exists a homeomorphism $h_0 \colon S^2 \to \widetilde{S}^2$ such that $h_0(\tau) = \widetilde{\tau}$ for each $\tau \in \mathcal{D}^0$.

The map f realizes the subdivision rule $(\mathcal{D}^1, \mathcal{D}^0, L)$. So if $\tau \in \mathcal{D}^1$, then $f(\tau) = L(\tau) \in \mathcal{D}^0$. Since $\widetilde{f}$ realizes $(\widetilde{\mathcal{D}}^1, \widetilde{\mathcal{D}}^0, \widetilde{L})$, we have

$$\widetilde{f}(\widetilde{\tau}) = \widetilde{L}(\widetilde{\tau}) = \widetilde{L(\tau)} = \widetilde{f(\tau)} = h_0(f(\tau)) \in \widetilde{\mathcal{D}}^0$$

for $\tau \in \mathcal{D}^1$. The map f is cellular for $(\mathcal{D}^1, \mathcal{D}^0)$ and so for each $\tau \in \mathcal{D}^1$ the map $f|\tau$ is a homeomorphism of τ onto $f(\tau)$. Similarly, the map $\widetilde{f}|\widetilde{\tau}$ is a homeomorphism of $\widetilde{\tau}$ onto $\widetilde{f}(\widetilde{\tau}) = h_0(f(\tau))$. Hence for each $\tau \in \mathcal{D}^1$ the map

$$\varphi_\tau := (\widetilde{f}|\widetilde{\tau})^{-1} \circ h_0 \circ (f|\tau)$$

is well-defined and a homeomorphism from τ onto $\widetilde{\tau}$. If $x \in \tau$, then $y = \varphi_\tau(x)$ is the unique point $y \in \widetilde{\tau}$ with $\widetilde{f}(y) = h_0(f(x))$.

If $\sigma, \tau \in \mathcal{D}^1$ and $\sigma \subset \tau$, then

$$\varphi_\tau|\sigma = \varphi_\sigma.$$

Indeed, if $x \in \sigma$, then $y = \varphi_\sigma(x) \in \widetilde{\sigma} \subset \widetilde{\tau}$ and $\widetilde{f}(y) = h_0(f(x))$. Hence $\varphi_\sigma(x) = y = \varphi_\tau(x)$ by the uniqueness property of $\varphi_\tau(x)$.

If a point $x \in S^2$ lies in two cells $\tau, \tau' \in \mathcal{D}^1$, then $\varphi_\tau(x) = \varphi_{\tau'}(x)$. Indeed, there exists a unique cell $\sigma \in \mathcal{D}^1$ with $x \in \operatorname{int}(\sigma)$. Then $\sigma \subset \tau \cap \tau'$ by Lemma 5.3 (ii), and so, by what we have just seen, we conclude

$$\varphi_\tau(x) = \varphi_\sigma(x) = \varphi_{\tau'}(x).$$

This allows us to define a map $h_1 \colon S^2 \to \widetilde{S}^2$ as follows. If $x \in S^2$, pick a cell $\tau \in \mathcal{D}^1$ with $x \in \tau$, and set

$$h_1(x) = \varphi_\tau(x).$$

Then h_1 is well-defined.

The definitions of h_1 and φ_τ imply that $h_0 \circ f = \widetilde{f} \circ h_1$. Moreover, $h_1|\tau = \varphi_\tau$ is a homeomorphism of τ onto $\widetilde{\tau}$ for each $\tau \in \mathcal{D}^1$. So by Lemma 5.11 (i) the map h_1 is a homeomorphism of S^2 onto $\widetilde{S}^2$.

We claim that $h_1(\tau) = \widetilde{\tau}$ not only for $\tau \in \mathcal{D}^1$, but also for $\tau \in \mathcal{D}^0$. Indeed, let $\tau \in \mathcal{D}^0$ and $x \in \tau$ be arbitrary. Since $\mathcal{D}^1$ refines $\mathcal{D}^0$, there exists $\sigma \in \mathcal{D}^1$ such that $x \in \sigma \subset \tau$. Then $\widetilde{\sigma} \subset \widetilde{\tau}$ and $h_1(x) \in h_1(\sigma) = \widetilde{\sigma} \subset \widetilde{\tau}$. We conclude that $h_1(\tau) \subset \widetilde{\tau}$. Conversely, if $y \in \widetilde{\tau}$, then there exists a cell in $\widetilde{\mathcal{D}}^1$ that contains y and is contained in $\widetilde{\tau}$. This cell has the form $\widetilde{\sigma}$ with $\sigma \subset \tau$. Hence $y \in \widetilde{\sigma} = h_1(\sigma) \subset h_1(\tau)$. This shows that $h_1(\tau) = \widetilde{\tau}$ as desired.

This implies that for the homeomorphism $h_1^{-1} \circ h_0 \colon S^2 \to S^2$ we have $(h_1^{-1} \circ h_0)(\tau) = h_1^{-1}(\widetilde{\tau}) = \tau$ for each $\tau \in \mathcal{D}^0$. So by Lemma 5.11 (iii) the homeomorphism $h_1^{-1} \circ h_0$ is isotopic to id_{S^2} rel. $\mathbf{V}^0$, where $\mathbf{V}^0$ is the set of vertices of $\mathcal{D}^0$. If we postcompose an isotopy rel. $\mathbf{V}^0$ between $h_1^{-1} \circ h_0$ and id_{S^2} by h_1, we see that h_0 and h_1 are isotopic rel. $\mathbf{V}^0$, and hence also isotopic rel. $\operatorname{post}(f)$, because $\operatorname{post}(f) \subset \mathbf{V}^0$. Since $h_0 \circ f = \widetilde{f} \circ h_1$, the maps f and $\widetilde{f}$ are Thurston equivalent. $\qquad\square$

REMARK 12.14. Suppose the setup is as in the previous lemma and its proof. We record some observations that will be important later in the proof of Theorem 15.10.

(i) Let $\mathcal{C}$ be the Jordan curve of $\mathcal{D}^0$ and $\widetilde{\mathcal{C}}$ be the Jordan curve of $\widetilde{\mathcal{D}}^0$. Then $\mathcal{C}$ and $\widetilde{\mathcal{C}}$ are the 1-skeletons of $\mathcal{D}^0$ and $\widetilde{\mathcal{D}}^0$, respectively. The homeomorphisms h_0 and h_1 constructed in the previous proof have the property that they send each cell $\tau \in \mathcal{D}^0$ to the corresponding cell $\widetilde{\tau} \in \widetilde{\mathcal{D}}^0$. Since the bijection $\tau \mapsto \widetilde{\tau}$ is an isomorphism of the cell complexes $\mathcal{D}^0$ and $\widetilde{\mathcal{D}}^0$, it preserves the dimension of a cell. This implies that the maps h_0 and h_1 send the 1-skeleton of $\mathcal{D}^0$ to the 1-skeleton of $\widetilde{\mathcal{D}}^0$, and so $h_0(\mathcal{C}) = \widetilde{\mathcal{C}} = h_1(\mathcal{C})$.

(ii) The cell complex isomorphisms $\phi_i \colon \mathcal{D}^i \to \widetilde{\mathcal{D}}^i$, $i = 0, 1$, as in the definition of an isomorphism between $(\mathcal{D}^1, \mathcal{D}^0, L)$ and $(\widetilde{\mathcal{D}}^1, \widetilde{\mathcal{D}}^0, \widetilde{L})$ send flags in $\mathcal{D}^i$ to flags in $\widetilde{\mathcal{D}}^i$. In the definition of an isomorphism of two-tile subdivision rules one can make the stronger additional requirement that positively-oriented flags are sent to positively-oriented flags. If one assumes this stronger notion of isomorphism in the previous lemma, then the maps h_0 and h_1 will be orientation-preserving.

(iii) For a fixed number of cells in $\mathcal{D}^1$, there is only a finite number of two-tile subdivision rules $(\mathcal{D}^1, \mathcal{D}^0, L)$ up to isomorphism. This implies that if we consider Thurston maps $f \colon S^2 \to S^2$ of fixed degree and a fixed number of postcritical points, then up to isomorphism there is only a finite number of two-tile subdivision rules given by such a map f and an f-invariant Jordan curve $\mathcal{C} \subset S^2$ according to Proposition 12.2. These statements remain true if one uses the strong notion of isomorphism between two-tile subdivision rules as in (ii).

Based on these observations one can show that a rational expanding Thurston map $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ with hyperbolic orbifold has at most finitely many f-invariant Jordan curves $\mathcal{C} \subset \widehat{\mathbb{C}}$ with $\mathrm{post}(f) \subset \mathcal{C}$ (see Theorem 15.10).

If one wants to discuss specific examples of Thurston maps that realize a given two-tile subdivision rule $(\mathcal{D}^1, \mathcal{D}^0, L)$, then it is convenient to represent the relevant data in a compressed form. The information on the labeling L is completely determined by a pair of corresponding positively-oriented flags in $\mathcal{D}^1$ and $\mathcal{D}^0$.

LEMMA 12.15. *Let $(\mathcal{D}^1, \mathcal{D}^0)$ be a pair of cell decompositions of S^2 satisfying conditions* (i)–(iv) *in Definition 12.1.*

(i) *If (c_0', c_1', c_2') and (c_0, c_1, c_2) are positively-oriented flags in $\mathcal{D}^1$ and $\mathcal{D}^0$, respectively, then there exists a unique orientation-preserving labeling $L \colon \mathcal{D}^1 \to \mathcal{D}^0$ with $(L(c_0'), L(c_1'), L(c_2')) = (c_0, c_1, c_2)$.*

(ii) *Let v' be a vertex and X' be a tile in $\mathcal{D}^1$, and v be a vertex and X be a tile in $\mathcal{D}^0$. If $v' \in X'$, then there exists a unique orientation-preserving labeling $L \colon \mathcal{D}^1 \to \mathcal{D}^0$ such that $L(v') = v$ and $L(X') = X$.*

So in both cases, $(\mathcal{D}^1, \mathcal{D}^0, L)$ is a two-tile subdivision rule. In (ii) we automatically have $v \in X$, because every vertex in $\mathcal{D}^0$ is contained in each of the two tiles in $\mathcal{D}^0$.

Note that (i) easily follows from (ii). We formulated (i) explicitly, because this version puts the statement in a more conceptual setting

and because in the proof we use Lemma 5.23 to first establish (i) and then derive (ii).

PROOF. For $i = 0, 1$ denote by $\mathbf{V}^i, \mathbf{E}^i, \mathbf{X}^i$ the set of vertices, edges, and tiles of $\mathcal{D}^i$, respectively.

(i) To describe the labeling for $(\mathcal{D}^1, \mathcal{D}^0)$, we proceed in the manner discussed after Definition 5.22 and choose a particular index set $\mathcal{L}$ for the labeling of the elements in $\mathcal{D}^0$ and $\mathcal{D}^1$.

We let $\mathcal{L}$ be the set that consists of two disjoint copies of $\mathbb{Z}_k$ (one will be for the vertices, and one for the edges), and the set $\{\mathtt{b}, \mathtt{w}\}$, where again we think of $\mathtt{w}$ representing "white" and $\mathtt{b}$ representing "black".

We assign to $c_2 \in \mathbf{X}^0$ the color "white", and "black" to the other tile in $\mathbf{X}^0$. We assign $0 \in \mathbb{Z}_k$ to the 0-vertex $v_0 \in c_0$. Then there is a unique way to assign labels in $\mathbb{Z}_k$ to the other vertices on $\mathcal{C} := \partial c_2$ (and the corresponding cells of dimension 0) such that if $v_0, v_1, \ldots, v_{k-1}$ are the vertices indexed by their label, then they are in cyclic order on $\mathcal{C}$ as considered as the boundary of the white 0-tile and in anti-cyclic order for the black 0-tile. Each 0-edge e is an arc on $\mathcal{C}$ with endpoints v_l and v_{l+1} for a unique $l \in \mathbb{Z}_k$. We label e by l (where we think of l as belonging to the second copy of $\mathbb{Z}_k$). Since (c_0, c_1, c_2) is a positively-oriented flag, and v_0 is the initial point of c_1, the edge c_1 has the label 0. All this is just a special case of Lemma 5.23. If in this way we assign to each element in $\mathcal{D}^0$ a label in $\mathcal{L}$, we get a bijection $\psi \colon \mathcal{D}^0 \to \mathcal{L}$. Note that if (τ_0, τ_1, τ_2) is any positively-oriented flag in $\mathcal{D}^0$, then its image under ψ has the form $(l, l, \mathtt{w})$ or $(l, l-1, \mathtt{b})$ for some $l \in \mathbb{Z}_k$ (see Lemma 5.23 (vi)).

For $\mathcal{D}^1$ we invoke Lemma 5.23 directly to set up a suitable map $\varphi \colon \mathcal{D}^1 \to \mathcal{L}$. Since $\mathcal{D}^1$ satisfies the conditions of Lemma 5.23, we can find maps $L_{\mathbf{V}} \colon \mathbf{V}^1 \to \mathbb{Z}_k$, $L_{\mathbf{E}} \colon \mathbf{E}^1 \to \mathbb{Z}_k$, and $L_{\mathbf{X}} \colon \mathbf{X}^1 \to \{\mathtt{b}, \mathtt{w}\}$ with the properties (ii)–(iv) stated in the lemma and the normalizations $L_{\mathbf{V}}(c_0') = 0$ (recall our convention that here we do not distinguish between vertices and 0-dimensional cells) , $L_{\mathbf{E}}(c_1') = 0$, and $L_{\mathbf{X}}(c_2') = \mathtt{w}$. The maps $L_{\mathbf{V}}$, $L_{\mathbf{E}}$, $L_{\mathbf{X}}$ induce a unique map $\varphi \colon \mathcal{D}^1 \to \mathcal{L}$ such that $\varphi(c) = L_{\mathbf{X}}(c)$ if c is a 1-tile, $\varphi(c) = L_{\mathbf{E}}(c)$ if c is a 1-edge, and $\varphi(c) = L_{\mathbf{V}}(v)$ if $c = \{v\}$ consists of a 1-vertex v.

Now define $L := \psi^{-1} \circ \varphi \colon \mathcal{D}^1 \to \mathcal{D}^0$. The map L assigns to each 1-cell c the unique 0-cell that has the same dimension as c and carries the same label in $\mathcal{L}$ as c.

It follows immediately from the properties of the maps ψ and φ that L preserves dimensions, respects inclusions, and is injective on cells. Hence L is a labeling according to Definition 5.22. By our normalizations the map L sends the flag (c_0', c_1', c_2') to (c_0, c_1, c_2).

Moreover, L is orientation-preserving. Indeed, φ maps the cells τ_0, τ_1, τ_2 in a positively-oriented flag in $\mathcal{D}^1$ to l, l, $\mathtt{w}$, or to l, $l-1$, $\mathtt{b}$, respectively, where $l \in \mathbb{Z}_k$. These triples correspond to positively-oriented flags in $\mathcal{D}^0$. It follows that L has the desired properties.

To show uniqueness, we reverse the process. Given a labeling $L\colon \mathcal{D}^1 \to \mathcal{D}^0$ with the stated properties, we use the same map $\psi\colon \mathcal{D}^0 \to \mathcal{L}$ as above and define maps $L_{\mathbf{V}}\colon \mathbf{V}^1 \to \mathbb{Z}_k$, $L_{\mathbf{E}}\colon \mathbf{E}^1 \to \mathbb{Z}_k$, $L_{\mathbf{X}}\colon \mathbf{X}^1 \to \{\mathtt{b}, \mathtt{w}\}$ such that $L_{\mathbf{X}}(c) = (\psi \circ L)(c)$ if c is a 1-tile, $L_{\mathbf{E}}(c) = (\psi \circ L)(c)$ if c is a 1-edge, and $L_{\mathbf{V}}(v) = (\psi \circ L)(c)$ if $c = \{v\}$ consists of a 1-vertex v.

Then we have normalizations $L_{\mathbf{V}}(c_0') = 0$, $L_{\mathbf{E}}(c_1') = 0$, and $L_{\mathbf{X}}(c_2') = \mathtt{w}$ as in Lemma 5.23 (i). If we can show that $L_{\mathbf{V}}$, $L_{\mathbf{E}}$, $L_{\mathbf{X}}$ have the properties (ii)–(iv) in Lemma 5.23, then the uniqueness of L will follow from the corresponding uniqueness statement in this lemma.

To see this, let $e \in \mathcal{D}^1$ be arbitrary, and $X, Y \in \mathcal{D}^1$ be the two tiles that contain e in its boundary. Let $u, v \in \mathbf{V}^1$ be the two endpoints of e. We may assume that notation is chosen so that the flag $(\{u\}, e, X)$ is positively-oriented. Then $(\{v\}, e, Y)$ is also positively-oriented. It follows that the images of these flags under L are positively-oriented. Since L is injective on cells, and so $L(u) \neq L(v)$, this implies that $L(X) \neq L(Y)$. So $L(X)$ and $L(Y)$ carry different colors (given by ψ) which implies that X and Y also carry different colors by definition of $L_{\mathbf{X}}$. Hence $L_{\mathbf{X}}$ has property (ii) in Lemma 5.23. By switching the notation for u and v and X and Y if necessary, we may assume that X and $L(X)$ are white tiles. Since the flag $(\{L(u)\}, L(e), L(X))$ is positively-oriented, and $L(X)$ is white, it follows that for some $l \in \mathbb{Z}_k$ we have $\psi(L(u)) = l$ and $\psi(L(e)) = l$. Hence $L_{\mathbf{V}}(u) = l$ and $L_{\mathbf{E}}(e) = l$. Similarly, using that $L(Y)$ is black and that $(\{L(v)\}, L(e), L(Y))$ is positively-oriented, we see that $L_{\mathbf{V}}(v) = l + 1$.

In other words, if we run along an oriented edge e in $\mathcal{D}^1$ so that a white tile lies on the left of e, then the label of the endpoint of e (given by $L_{\mathbf{V}}$) is increased by one, and decreased by one if a black tile lies on the left. Hence $L_{\mathbf{V}}$ has the property (iii) in Lemma 5.23. Moreover, we also see that the label $L_{\mathbf{E}}(e)$ is related to the labels of its endpoints as in statement (iv) of Lemma 5.23. The uniqueness of L follows.

(ii) Since $v' \in X'$ and $v' \in \mathbf{V}^1$, we have $v' \in \partial X'$. There are precisely two edges in $\mathbf{E}^1$ that are contained in $\partial X'$ and have v' as one of their endpoints. For precisely one of these edges e', the triple $(\{v'\}, e', X')$ is the unique positively-oriented flag in $\mathcal{D}'$ that includes $\{v'\}$ and X'.

Similarly, there exists a unique edge $e \in \mathbf{E^0}$ such that $(\{v\}, e, X)$ is a positively-oriented flag in $\mathcal{D}^0$. By (i) there exists an orientation-preserving labeling $L\colon \mathcal{D}^1 \to \mathcal{D}^0$ that sends $(\{v'\}, e', X')$ to $(\{v\}, e, X)$. In particular, $L(v') = v$ and $L(X') = X$. This shows existence of a labeling as desired.

To prove uniqueness, suppose that $L\colon \mathcal{D}^1 \to \mathcal{D}^0$ is an orientation-preserving labeling with $L(v') = v$ and $L(X') = X$. Then the image of $(\{v'\}, e', X')$ under L is a positively-oriented flag of the form $(v, L(e'), X)$. Since $(\{v\}, e, X)$ is the unique positively-oriented flag in $\mathcal{D}'$ that includes $\{v\}$ and X, we have $L(e') = e$. Uniqueness of L now follows from (i). $\qquad\square$

If we are given cell decompositions $\mathcal{D}^1$ and $\mathcal{D}^0$ as in the last lemma, then by part (ii) we can specify a unique labeling so that $(\mathcal{D}^1, \mathcal{D}^0, L)$ becomes a two-tile subdivision rule in a very condensed form: all we need to know is the image 0-tile X of one 1-tile X', and the image 0-vertex $v \in X$ of one 1-vertex $v' \in X'$. In specific examples (see Section 12.3), one usually wants to include more information on the labeling to get a better understanding of the mapping properties of the Thurston map that realizes the subdivision rule.

Let f be a map realizing a two-tile subdivision rule $(\mathcal{D}^1, \mathcal{D}^0, L)$. We want to show next that the property of f being combinatorially expanding for the Jordan curve $\mathcal{C}$ of $\mathcal{D}^0$ is independent of the realization. In contrast, this is not true for expansion of the map (see Example 12.21). We require a lemma.

LEMMA 12.16. *Let $f\colon S^2 \to S^2$ and $g\colon \widetilde{S}^2 \to \widetilde{S}^2$ be Thurston maps. Suppose $\#\operatorname{post}(f) \geq 3$, $\mathcal{C} \subset S^2$ is an f-invariant Jordan curve with $\operatorname{post}(f) \subset \mathcal{C}$, and $h_0, h_1\colon S^2 \to \widetilde{S}^2$ are orientation-preserving homeomorphisms satisfying $h_0|\operatorname{post}(f) = h_1|\operatorname{post}(f)$, $h_0 \circ f = g \circ h_1$, and $h_0(\mathcal{C}) = h_1(\mathcal{C})$.*

Then f is combinatorially expanding for $\mathcal{C}$ if and only if g is combinatorially expanding for $\widetilde{\mathcal{C}} := h_0(\mathcal{C}) = h_1(\mathcal{C})$.

As we will see momentarily, the conditions of the lemma imply that $\widetilde{\mathcal{C}}$ is g-invariant.

PROOF. We have $\operatorname{post}(g) = h_0(\operatorname{post}(f)) = h_1(\operatorname{post}(f))$ (see the remark after Lemma 2.5). Hence $\#\operatorname{post}(g) = \#\operatorname{post}(f) \geq 3$. Moreover, $\widetilde{\mathcal{C}} \subset \widetilde{S}^2$ is a Jordan curve with $\operatorname{post}(g) \subset \widetilde{\mathcal{C}}$. This curve is g-invariant, since

$$g(\widetilde{\mathcal{C}}) = g(h_1(\mathcal{C})) = h_0(f(\mathcal{C})) \subset h_0(\mathcal{C}) = \widetilde{\mathcal{C}}.$$

So the statement that g is combinatorially expanding for $\widetilde{\mathcal{C}}$ is meaningful (see Definition 12.4).

Pick an orientation of $\mathcal{C}$. By our assumptions the map $\varphi := h_1^{-1} \circ h_0$ fixes the elements of $\operatorname{post}(f)$ pointwise and the Jordan curve $\mathcal{C}$ setwise. Since $\#\operatorname{post}(f) \geq 3$ and $\operatorname{post}(f) \subset \mathcal{C}$, this implies that φ preserves the orientation of $\mathcal{C}$. Since φ is an orientation-preserving homeomorphism on S^2, the map φ sends each of the complementary components of $\mathcal{C}$ to itself. Thus, φ is cellular for $(\mathcal{D}^0, \mathcal{D}^0)$, where $\mathcal{D}^0 = \mathcal{D}^0(f, \mathcal{C})$, and we have $\varphi(c) = c$ for each cell $c \in \mathcal{D}^0$. As in the proof of Lemma 5.11 (iii), this implies that φ is isotopic to id_{S^2} rel. $\operatorname{post}(f)$. Hence $h_0 = h_1 \circ \varphi$ is isotopic to $h_1 = h_1 \circ \operatorname{id}_{S^2}$ rel. $\operatorname{post}(f)$, and so there exists an isotopy $H^0 \colon S^2 \times I \to \widetilde{S}^2$ rel. $\operatorname{post}(f)$ with $H_0^0 = h_0$ and $H_1^0 = h_1$.

As in the proof of Theorem 11.1, we can repeatedly lift the initial isotopy H^0 based on Proposition 11.3. In this way we can find isotopies $H^n \colon S^2 \times I \to \widetilde{S}^2$ rel. $\operatorname{post}(f)$ such that $H_t^n \circ f = g \circ H_t^{n+1}$ and $H_0^{n+1} = H_1^n$ for all $n \in \mathbb{N}_0$ and $t \in I$. Note that H^n for $n \geq 1$ is actually an isotopy rel. $f^{-1}(\operatorname{post}(f)) \supset \operatorname{post}(f)$.

Define homeomorphisms $h_n := H_0^n$ for $n \in \mathbb{N}_0$ (note that for $n = 0$ and $n = 1$ these maps agree with our given maps h_0 and h_1). Then $h_n \circ f = g \circ h_{n+1}$, and so

$$(12.6) \qquad\qquad h_0 \circ f^n = g^n \circ h_n$$

for all $n \in \mathbb{N}_0$.

We have $h_n | \operatorname{post}(f) = h_0 | \operatorname{post}(f)$ which implies

$$(12.7) \qquad\qquad h_n(\operatorname{post}(f)) = \operatorname{post}(g)$$

for all $n \in \mathbb{N}_0$. Moreover, $h_n | f^{-1}(\operatorname{post}(f)) = h_1 | f^{-1}(\operatorname{post}(f))$ and so

$$(12.8) \qquad\qquad h_n(f^{-1}(\operatorname{post}(f))) = g^{-1}(\operatorname{post}(g))$$

for $n \in \mathbb{N}$ as follows from Lemma 11.2.

Our hypotheses imply that if c is a cell in $\mathcal{D}^0(f, \mathcal{C})$, then $h_0(c)$ is a cell in $\mathcal{D}^0(g, \widetilde{\mathcal{C}})$. Since the set

$$\widetilde{\mathcal{D}}^n := \{h_n(c) : c \in \mathcal{D}^n(f, \mathcal{C})\}$$

is a cell decomposition of $\widetilde{S}^2$, it follows from this and (12.6) that g^n is cellular for $(\widetilde{\mathcal{D}}^n, \mathcal{D}^0(g, \widetilde{\mathcal{C}}))$. Since g^n is also cellular for the pair $(\mathcal{D}^n(g, \widetilde{\mathcal{C}}), \mathcal{D}^0(g, \widetilde{\mathcal{C}}))$, the uniqueness statement in Lemma 5.12 implies that $\widetilde{\mathcal{D}}^n = \mathcal{D}^n(g, \mathcal{C})$ for all $n \in \mathbb{N}_0$. In other words, the n-cells for $(g, \widetilde{\mathcal{C}})$ are precisely the images of the n-cells for $(f, \mathcal{C})$ under the homeomorphism h_n.

We also have

$$(12.9) \qquad\qquad h_n(\mathcal{C}) = \widetilde{\mathcal{C}}$$

for each $n \in \mathbb{N}_0$. This can be seen by induction on n as follows. The statement is true for $n = 0$ and $n = 1$ by our assumption and by the definition of $\widetilde{\mathcal{C}}$. Assume that $h_n(\mathcal{C}) = \widetilde{\mathcal{C}}$ for some $n \in \mathbb{N}$. Then by Lemma 11.2 and the induction hypotheses we have

$$J := h_{n+1}(\mathcal{C}) \subset h_{n+1}(f^{-1}(\mathcal{C})) = g^{-1}(h_n(\mathcal{C})) = g^{-1}(\widetilde{\mathcal{C}}).$$

The identity (12.8) implies that $(H_t^n) \circ h_n^{-1}$ is an isotopy on $\widetilde{S}^2$ rel. $g^{-1}(\mathrm{post}(g))$. It isotopes $\widetilde{\mathcal{C}} = h_n(\mathcal{C}) \subset g^{-1}(\widetilde{\mathcal{C}})$ into $J = h_{n+1}(\mathcal{C})$ rel. $g^{-1}(\mathrm{post}(g))$. So $\widetilde{\mathcal{C}}$ and J are Jordan curves contained in the 1-skeleton $g^{-1}(\widetilde{\mathcal{C}})$ of $\mathcal{D}^1(g, \widetilde{\mathcal{C}})$ that are isotopic relative to the set $g^{-1}(\mathrm{post}(g))$ of vertices of $\mathcal{D}^1(g, \widetilde{\mathcal{C}})$. Lemma 11.12 implies that $J = \widetilde{\mathcal{C}}$, and (12.9) follows.

Now (12.9) and (12.7) imply that a chain of n-tiles for $(f, \mathcal{C})$ joins opposite sides of $\mathcal{C}$ if and only if their images under h_n form a chain joining opposite sides of $\widetilde{\mathcal{C}}$. Since the images of the n-tiles for $(f, \mathcal{C})$ under h_n are precisely the n-tiles for $(g, \widetilde{\mathcal{C}})$, we have $D_n(f, \mathcal{C}) = D_n(g, \widetilde{\mathcal{C}})$ for each $n \in \mathbb{N}_0$. The statement follows. $\qquad\square$

Now we can show the desired independence of combinatorial expansion from the realization of a two-tile subdivision rule.

LEMMA 12.17. *Let $(\mathcal{D}^1, \mathcal{D}^0, L)$ be a two-tile subdivision rule on S^2 and $\mathcal{C}$ be the Jordan curve of $\mathcal{D}^0$. Suppose that the maps $f\colon S^2 \to S^2$ and $g\colon S^2 \to S^2$ both realize the subdivision rule and that $\#\,\mathrm{post}(f) = \#\,\mathrm{post}(g) \geq 3$. Then f is combinatorially expanding for $\mathcal{C}$ if and only if g is combinatorially expanding for $\mathcal{C}$.*

PROOF. Let $\mathbf{V}^0$ and $\mathbf{V}^1$ be the set of vertices of $\mathcal{D}^0$ and $\mathcal{D}^1$, respectively. Then $P := \mathrm{post}(f) = \mathrm{post}(g) \subset \mathbf{V}^0 \subset \mathbf{V}^1$.

It follows from the proof of the uniqueness part of Proposition 5.26 that there exists a homeomorphism $h_1\colon S^2 \to S^2$ isotopic to id_{S^2} rel. $\mathbf{V}^1 \supset \mathrm{post}(f) = \mathrm{post}(g)$ that satisfies $f = g \circ h_1$. Moreover, $h_1(e) = e$ for each edge e in $\mathcal{D}^1$. Since $\mathcal{D}^1$ is a refinement of $\mathcal{D}^0$ and so the 1-skeleton $\mathcal{C}$ of $\mathcal{D}^0$ is contained in the 1-skeleton of $\mathcal{D}^1$, this implies $h_1(\mathcal{C}) = \mathcal{C}$. Define $h_0 = \mathrm{id}_{S^2}$. Since h_1 is isotopic to id_{S^2} rel. P we have $h_1|P = \mathrm{id}_{S^2}|P = h_0|P$. Moreover, $h_0 \circ f = g \circ h_1$, $h_1(\mathcal{C}) = \mathcal{C} = h_0(\mathcal{C})$, and both h_0 and h_1 are orientation-preserving homeomorphisms on S^2. This shows that the hypotheses of Lemma 12.16 are satisfied (with $\widetilde{S}^2 = S^2$), and so f is combinatorially expanding for $\mathcal{C}$ if and only if g is combinatorially expanding for $\widetilde{\mathcal{C}} = h_0(\mathcal{C}) = h_1(\mathcal{C}) = \mathcal{C}$. $\qquad\square$

The previous lemma motivates the following definition.

DEFINITION 12.18 (Combinatorially expanding two-tile subdivision rules). Let $(\mathcal{D}^0, \mathcal{D}^1, L)$ be a two-tile subdivision rule, and $\mathcal{C}$ be the Jordan curve of $\mathcal{D}^0$. We call $(\mathcal{D}^0, \mathcal{D}^1, L)$ *combinatorially expanding* if every Thurston map f realizing $(\mathcal{D}^0, \mathcal{D}^1, L)$ is combinatorially expanding for $\mathcal{C}$.

We know that if this condition is true for *one* Thurston map realizing the subdivision rule, then it is true for *all* such maps by Lemma 12.17. We will see that under an additional mild technical assumption a two-tile subdivision rule can be realized by an expanding Thurston map if and only if the subdivision rule is combinatorially expanding (see Theorem 14.1).

We conclude this section by proving a statement related to Remark 12.12 (ii).

PROPOSITION 12.19. *Let $(\mathcal{D}^1, \mathcal{D}^0, L)$ be a two-tile subdivision rule on S^2 and $\mathcal{C}$ be the Jordan curve of $\mathcal{D}^0$. Suppose that the maps $f\colon S^2 \to S^2$ and $g\colon S^2 \to S^2$ both realize the subdivision rule. Let $\mathbf{V}^0$ be the vertex set of $\mathcal{D}^0$, and assume that $\mathrm{post}(f) = \mathrm{post}(g) = \mathbf{V}^0$ and $\#\mathbf{V}^0 \geq 3$.*

Then there exist homeomorphisms $h_n\colon S^2 \to S^2$ for $n \in \mathbb{N}_0$ with the following properties: for all $k, n \in \mathbb{N}_0$ with $n \geq k$ we have $h_n(c) \in \widetilde{\mathcal{D}}^k := \mathcal{D}^k(g, \mathcal{C})$ whenever $c \in \mathcal{D}^k := \mathcal{D}^k(f, \mathcal{C})$. Moreover, the map

$$c \in \mathcal{D}^k \mapsto h_n(c) \in \widetilde{\mathcal{D}}^k$$

is an isomorphism between the cell decompositions $\mathcal{D}^k$ and $\widetilde{\mathcal{D}}^k$ that does not depend on $n \geq k$.

We added the assumption $\mathrm{post}(f) = \mathrm{post}(g) = \mathbf{V}^0$ for convenience, because then one does not have to worry about vertices in $\mathcal{D}^0$ without dynamical relevance (see Remark 12.12 (i)).

We know that under the given assumptions $\mathcal{C}$ is invariant under f and g. So in the sequence $\mathcal{D}^0, \mathcal{D}^1, \mathcal{D}^2, \ldots$ each cell decomposition is a refinement of the previous one. Similarly, $\widetilde{\mathcal{D}}^0, \widetilde{\mathcal{D}}^1, \widetilde{\mathcal{D}}^2, \ldots$ forms a sequence of finer and finer cell decompositions. By the proposition there exists a homeomorphism h_n that induces isomorphisms of the cell decompositions in the first sequence with the corresponding cell decompositions in the second one up to level n. Moreover, the isomorphism between $\mathcal{D}^k$ and $\widetilde{\mathcal{D}}^k$ given by h_n actually does not depend on $n \geq k$. Based on this, one can easily show the combinatorics of the sequences $\mathcal{D}^n$ and $\widetilde{\mathcal{D}}^n$, $n \in \mathbb{N}_0$, are exactly the same (as formulated more precisely in Remark 12.12 (ii)).

Note that in general there is no single homeomorphism h on S^2 that induces an isomorphism between $\mathcal{D}^k$ and $\widetilde{\mathcal{D}}^k$ for *all* levels $k \in \mathbb{N}_0$. For example, no such h can exist if one of the Thurston maps is expanding, but the other one is not (for a specific case, see Example 12.21).

PROOF OF PROPOSITION 12.19. As in the statement, we use the notation $\mathcal{D}^i := \mathcal{D}^i(f, \mathcal{C})$ and $\widetilde{\mathcal{D}}^i := \mathcal{D}^i(g, \mathcal{C})$ for $i \in \mathbb{N}_0$. Note that for $i = 0, 1$ the first definition is consistent with our notation $\mathcal{D}^0$ and $\mathcal{D}^1$ for the cell decompositions of the subdivision rule, because under our assumptions we have $\mathcal{D}^0 = \mathcal{D}^0(f, \mathcal{C}) = \mathcal{D}^0(g, \mathcal{C}) = \widetilde{\mathcal{D}}^0$ and $\mathcal{D}^1 = \mathcal{D}^1(f, \mathcal{C}) = \mathcal{D}^1(g, \mathcal{C}) = \widetilde{\mathcal{D}}^1$.

By the proof of Lemma 12.17 there exist orientation-preserving homeomorphisms $h_0 = \mathrm{id}_{S^2}$ and $h_1 \colon S^2 \to S^2$ with $\mathcal{C} = h_0(\mathcal{C}) = h_1(\mathcal{C})$ satisfying the hypotheses of Lemma 12.16 (with $\widetilde{S}^2 = S^2$). Moreover, both h_0 and h_1 fix the points in $\mathrm{post}(f) = \mathrm{post}(g) = \mathbf{V^0}$. It follows from the considerations in the proof of Lemma 12.16 that we obtain orientation-preserving homeomorphisms $h_n \colon S^2 \to S^2$ for all $n \in \mathbb{N}_0$ that fix the curve $\mathcal{C}$ as a set and each point in $\mathbf{V^0}$. This implies that each h_n fixes all cells in $\mathcal{D}^0$ as sets.

As we have seen in the proof of Lemma 12.16, these homeomorphisms also satisfy

(12.10) $$h_n \circ f = g \circ h_{n+1}$$

for $n \in \mathbb{N}_0$.

Now let $n, k \in \mathbb{N}_0$ with $k \leq n$ be arbitrary, and consider the cell decomposition

$$h_n(\mathcal{D}^k) := \{h_n(c') : c' \in \mathcal{D}^k\}.$$

It follows from repeated application of (12.10) that $g^k = h_{n-k} \circ f^k \circ h_n^{-1}$. This implies that if $c' \in \mathcal{D}^k$, then g^k is a homeomorphism of $h_n(c') \in h_n(\mathcal{D}^k)$ onto $h_{n-k}(f^k(c'))$. Now $c := f^k(c')$ is a cell in $\mathcal{D}^0 = \widetilde{\mathcal{D}}^0$ and so $h_{n-k}(c) = c$. In other words, g^k is cellular for $(h_n(\mathcal{D}^k), \widetilde{\mathcal{D}}^0)$. Since g^k is also cellular for $(\widetilde{\mathcal{D}}^k, \widetilde{\mathcal{D}}^0)$, the uniqueness statement in Lemma 5.12 implies that $\widetilde{\mathcal{D}}^k = h_n(\mathcal{D}^k)$. The first part of statement follows.

It remains to show that the isomorphism between $\mathcal{D}^k$ and $\widetilde{\mathcal{D}}^k$ induced by h_n does not depend on $n \geq k$. Since we will not use this statement in the following, we provide only a sketch of the proof leaving some details to the reader.

It is enough to show that

(12.11) $$(h_{n+1}^{-1} \circ h_n)(c') = c',$$

whenever $c' \in \mathcal{D}^k$ and $n \geq k$. Since both h_n and h_{n+1} induce iso-morphisms of $\mathcal{D}^k$ and $\widetilde{\mathcal{D}}^k$, we know that $(h_{n+1}^{-1} \circ h_n)(c') \in \mathcal{D}^k$ for each $c' \in \mathcal{D}^k$.

Now (12.11) is true if c' consists of a vertex in $\mathcal{D}^k$, i.e., a point in $f^{-k}(\text{post}(f))$. This follows from the fact that h_{n+1} and h_h are isotopic rel. $f^{-n}(\text{post}(f))$, and so $h_{n+1}^{-1} \circ h_n$ fixes each point in $f^{-n}(\text{post}(f)) \supset f^{-k}(\text{post}(f))$.

Since $h_{n+1}^{-1} \circ h_n$ is isotopic to id_{S^2} rel. $f^{-n}(\text{post}(f)) \supset f^{-k}(\text{post}(f))$, the argument in the proof of Lemma 11.12 shows that (12.11) is also valid for each edge c' in $\mathcal{D}^k$. This in turn implies that if $X \in \mathcal{D}^k$ is a tile, then $(h_{n+1}^{-1} \circ h_n)(X)$ is a tile in $\mathcal{D}^k$ with the same boundary as X. We conclude that these tiles are the same. Equation (12.11) follows. $\square$

12.3. Examples of two-tile subdivision rules

In this section we present some examples of two-tile subdivision rules $(\mathcal{D}^1, \mathcal{D}^0, L)$ and maps that realize them. This is based on Proposition 12.3. We have already used this method for constructing and describing Thurston maps before (see the remark after Proposition 12.3).

In most of our examples we will represent the underlying sphere S^2 as a pillow P (see Section A.10) obtained by gluing together two isometric copies X_{w}^0 and X_{b}^0 of a (simple) Euclidean polygon $X \subset \mathbb{C}$. This gives us a natural cell decomposition $\mathcal{D}^0$ of S^2, where X_{w}^0 and X_{b}^0 are the 0-tiles, and the sides and corners on the common boundaries of the polygons the 0-edges and 0-vertices. We assign the color "white" to X_{w}^0, and "black" to X_{b}^0. In our figures the top polygon of the pillow will be the white 0-tile. The pillow P carries a natural orientation as discussed in Section A.10. This in turn determines an orientation of the equator $\mathcal{C} := \partial X_{\mathsf{w}}^0 = \partial X_{\mathsf{b}}^0$ of P so that X_{w}^0 lies on the left and X_{b}^0 on the right of $\mathcal{C}$ (see Section A.4).

The description of the cell decomposition $\mathcal{D}^1$ is usually more complicated and depends on the specific case.

We know that in order to uniquely specify the labeling $L\colon \mathcal{D}^1 \to \mathcal{D}^0$ it is enough to know the image of a pair (v', X'), where X' is a 1-tile and $v' \in X'$ a 1-vertex (see Lemma 12.15 (ii)). In general, we will include more information on the labeling L for a better illustration of the behavior of the map f realizing the subdivision rule. We will assign the colors "black" and "white" to the 1-tiles indicating to which of the 0-tiles they are sent by L (and f). With these labels the cell decomposition $\mathcal{D}^1$ will be a checkerboard tiling of k-gons, where k is the number of vertices in $\mathcal{D}^0$.
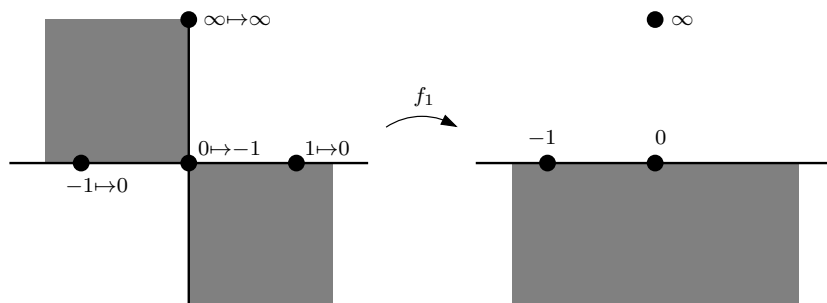
FIGURE 12.4. The two-tile subdivision rule for $z^2 - 1$.
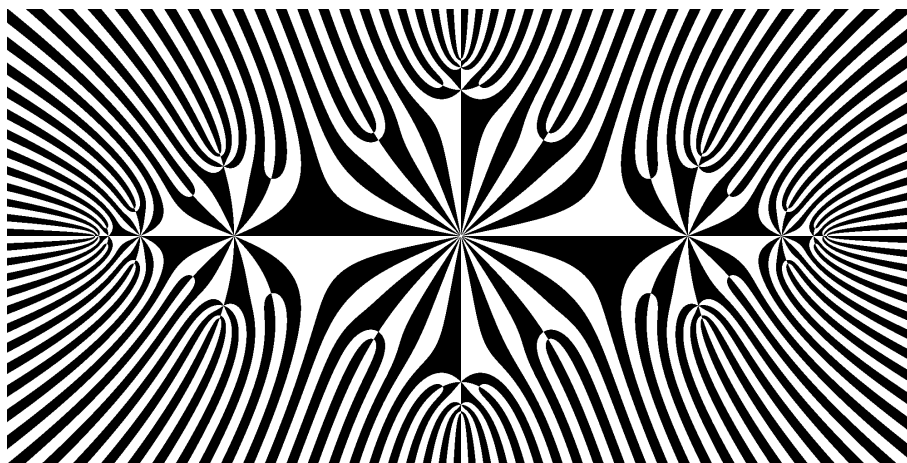
FIGURE 12.5. Tiles of level 7 for Example 12.20.

Sometimes we will introduce markings for the 0-vertices and some 1-vertices, often suggested by a natural identification of the underlying sphere S^2 with the Riemann sphere $\widehat{\mathbb{C}}$. A 1-vertex that is also a 0 vertex marked a is then marked "$a \mapsto b$" if the labeling L sends it to the 0-vertex marked b. Similarly, "$\mapsto b$" marks a 1-vertex that is not a 0-vertex and is sent to the 0-vertex b by L.

After these preliminaries we now proceed to discussing the examples.

EXAMPLE 12.20. Here the white 0-tile is the (closure of the) upper half-plane, and the black 0-tile is the (closure of the) lower half-plane in $\widehat{\mathbb{C}}$. The 0-vertices are the points $-1, 0, \infty$. Thus the 0-edges are $[-\infty, -1], [-1, 0], [0, \infty] \subset \widehat{\mathbb{R}}$. The cell decomposition $\mathcal{D}^0$ is indicated on the right of Figure 12.4.

The white 1-tiles are the first and third quadrants, and the black 1-tiles are the second and forth quadrants. The 1-vertices and their
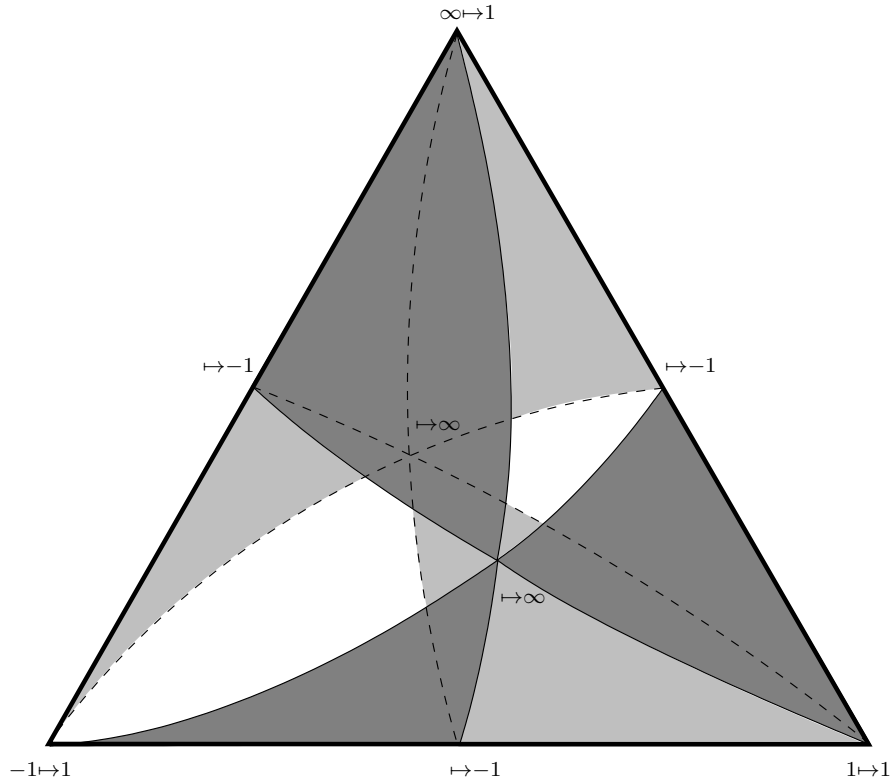
FIGURE 12.6. The barycentric subdivision rule.

labelings are as follows: the point ∞ is the only 1-vertex labeled ∞, the 1-vertices -1 and 1 are labeled 0, the 1-vertex 0 is labeled -1. The cell decomposition is indicated on the left in Figure 12.4.

The pair $(\mathcal{D}^1, \mathcal{D}^0)$ together with this orientation-preserving labeling L is a two-tile subdivision rule. It is straightforward to check that the map $f_1(z) = z^2 - 1$ realizes the subdivision rule $(\mathcal{D}^1, \mathcal{D}^0, L)$.

Since f_1 is a Thurston polynomial, it cannot be expanding by Lemma 6.7. In fact, for all n there exist two n-tiles that contain all postcritical points $-1, 0, \infty$. Figure 12.5 shows the tiles of level 7.

EXAMPLE 12.21 (The barycentric subdivision rule). We glue two equilateral triangles together along their boundaries to form a pillow S^2. It is a polyhedral surface and so conformally equivalent to $\widehat{\mathbb{C}}$. The two triangles are the 0-tiles. We can find a conformal equivalence of S^2 with $\widehat{\mathbb{C}}$ such that the triangles correspond to the upper and lower half-planes, and the vertices to the points $-1, 1, \infty$. For convenience we identify the vertices with $-1, 1, \infty$; they are the 0-vertices. The 0-edges are the three edges of the triangles. The bisectors divide each triangle (each 0-tile) into six smaller triangles. These 12 small triangles are the
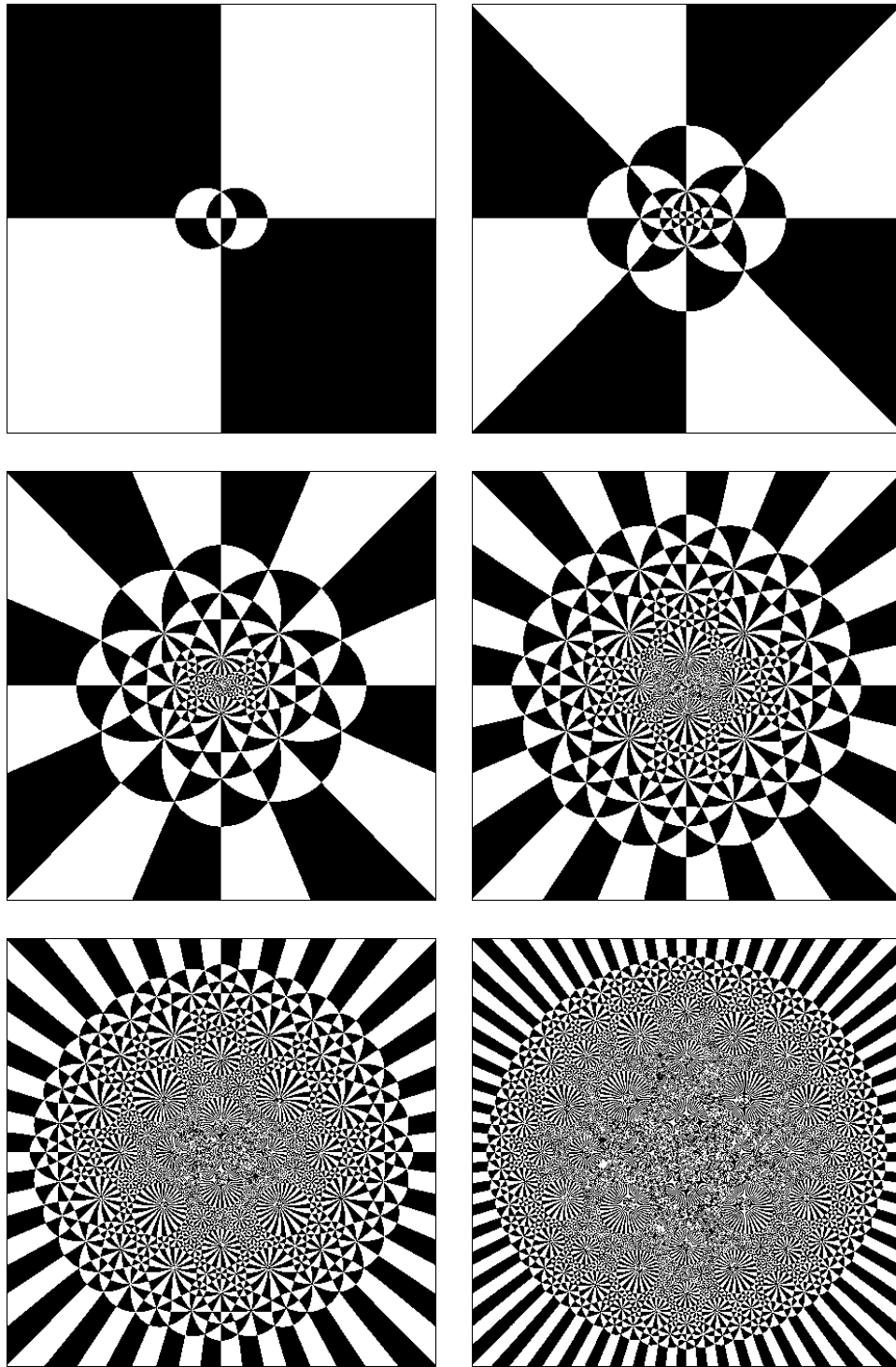
FIGURE 12.7. Tiles of levels 1–6 for the barycentric subdivision map.

1-tiles. The labeling of the 1-vertices is indicated in Figure 12.6. Again
we obtain a two-tile subdivision rule. We can realize this subdivision
rule by a map f_2 that *conformally* maps 1-tiles to the 0-tiles. Under
the indicated identification of S^2 with $\widehat{\mathbb{C}}$, the map is given by

$$f_2(z) = 1 - \frac{54(z^2 - 1)^2}{(z^2 + 3)^3}$$

(see [**C–P03**, Example 4.6]). The subdivision rule is combinatorially
expanding, but the map f_2 is not expanding. This follows from Propo-
sition 2.3, because the point 1 is both a critical and a fixed point of
f_2. In Figure 12.7 the tiles of levels 1–6 are shown. The tiles inter-
secting the borders of a picture frame are actually unbounded; these
tiles form the (closures of the) flowers at ∞. One can show that for a
fixed n-vertex v, the intersection of the m-flowers $W^m(v)$, for $m \geq n$,
is not a single point, but in fact the closure of the Fatou component
of f_2 containing v. The Julia set of f_2 is a Sierpiński carpet, i.e., a set
homeomorphic to the standard Sierpiński carpet fractal.

It is possible to choose a different realization of the two-tile subdi-
vision indicated in Figure 12.6 by a map $\widetilde{f_2}$ that is expanding. Namely,
we can use *affine* maps to map the 1-tiles (the small triangles in the
barycentric subdivision of the equilateral triangles) to the 0-tiles. In
this case, the n-tiles are Euclidean triangles for each $n \in \mathbb{N}_0$. The
collection of all n-tiles is obtained from the $(n-1)$-tiles as the 1-tiles
were constructed from the 0-tiles: one subdivides each Euclidean tri-
angle representing an $(n-1)$-tile by its bisectors. It is not difficult
to see that the diameters of n-tiles tend to 0 as $n \to \infty$. Hence $\widetilde{f_2}$ is
expanding, and so this map is an example of an expanding Thurston
map with periodic critical points.

In Chapter 14 we will present a general procedure how to obtain
an expanding Thurston map from a combinatorially expanding one.
Roughly speaking, we define an equivalence relation on the underlying
sphere that collapses sets where the map fails to be expanding to a
point. For this example f_2, these equivalence classes are given by the
sets of the form $\bigcap_{m \geq n} W^m(v)$, where v is an n-vertex.

EXAMPLE 12.22. The map $f_3 = h$ constructed in Section 1.3 real-
izes the two-tile subdivision rule shown in Figure 1.2. We revisited this
map in Example 2.18, where we saw that f_3 is not Thurston equivalent
to a rational map. The map f_3 is combinatorially expanding as follows
from the considerations in Section 1.4.

This map is related to the Lattès map g considered in Section 1.1.
Namely, we cut the sphere along one 1-edge of g, and glued in two small
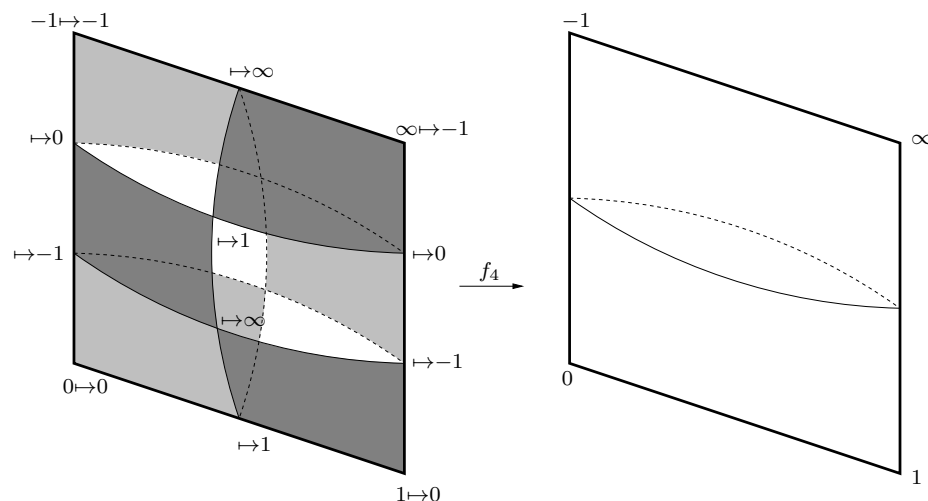
FIGURE 12.8. The 2-by-3 subdivision rule.

squares that are mapped to the 0-tiles. This increased the degree of the map by 1. A similar construction is possible in greater generality. This was introduced by Pilgrim and Tan Lei, who called this operation "blowing up an arc" (see [**PT98**]).

EXAMPLE 12.23 (The 2-by-3 subdivision rule). We present another example of an expanding Thurston map f_4 that is not (Thurston) equivalent to a rational map. In a sense, this is the easiest example, but it has a parabolic orbifold in contrast to the previous one.

The map f_4 is a Lattès-type map (see Definition 3.3) with signature $(2, 2, 2, 2)$ as provided by Proposition 3.21. Here the map $A\colon \mathbb{C} \to \mathbb{C}$ in (3.25) is given by $A(x + y\boldsymbol{i}) = 2x + 3y\boldsymbol{i}$ for $x, y \in \mathbb{R}$ and the associated crystallographic group G consists of all isometries on $\mathbb{R}^2$ of the form $g(u) = \pm u + \gamma$, where $\gamma \in \mathbb{Z}^2$. Then A descends to the map $f_4\colon S^2 \to S^2$ on the quotient $S^2 = \mathbb{R}^2/G$. As discussed in Example 3.20, we can represent the quotient space by a pillow P obtained by gluing two squares of side length $1/2$ together along their boundaries. We have seen in Section 1.1, that this pillow may be naturally identified with $\widehat{\mathbb{C}}$ via the Weierstraß $\wp$-function (which is discussed in more detail in Section 3.5). This explains the markings of the four 0-vertices in Figure 12.8 which represents the two-tile subdivision rule realized by f_4.

Each of the two faces (i.e., squares) of the pillow is divided into six rectangles as shown in Figure 12.8. These 12 rectangles are the 1-tiles. Their sides and vertices are the 1-edges and 1-vertices. The coloring of 1-tiles and the labeling of the 1-vertices is indicated on the left of Figure 12.8. The map f_4 sends each of the 12 rectangles affinely to

one of the two squares forming the faces of the pillow. This implies that each n-tile is a rectangle with side lengths $\frac{1}{2}2^{-n}$ and $\frac{1}{2}3^{-n}$. In particular, f_4 is an expanding Thurston map.

The fact that f_4 is not equivalent to a rational map can be derived from Theorem 3.22. In the following, we will sketch a different argument for this which is more in line with our general framework. In our outline, we will rely on some results and concepts that will be discussed later on.

To reach a contradiction, suppose that f_4 is equivalent to a rational map $R\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$. Then R is a Thurston map with no periodic critical points and hence is expanding (Proposition 2.3). So by Theorem 11.1 the maps f_4 and R are topologically conjugate. This implies by Theorem 18.1 (ii) that if our pillow P is equipped with a visual metric ϱ for f_4, then (P, ϱ) is quasisymmetrically equivalent to the standard 2-sphere, i.e., $\widehat{\mathbb{C}}$ equipped with the chordal metric. In particular, if X^0 is a 0-tile (i.e., one of the faces of the pillow P) equipped with a visual metric ϱ, then it can be embedded into $\widehat{\mathbb{C}}$ by a quasisymmetry.

Now there are visual metrics for f_4 with expansion factor $\Lambda = 2$. It is not hard to see this directly; it also follows from the general argument in the proof of Theorem 16.3 based on (16.5). Indeed, if $\mathcal{C}$ is the equator of the pillow (which is f_4-invariant), then we have $D_1 = D_1(f_4, \mathcal{C}) = 2$ in (16.5), which guarantees the existence of the desired visual metric.

If ϱ is such a metric, then (X^0, ϱ) is bi-Lipschitz equivalent to a Rickman's rug R_α. Here by definition the *Rickman's rug* R_α for $0 < \alpha < 1$ is the unit square $[0, 1]^2 \subset \mathbb{R}^2$ equipped with the metric d_α given by

$$d_\alpha((x_1, y_1), (x_2, y_2)) = |x_1 - x_2| + |y_1 - y_2|^\alpha$$

for $(x_1, y_1), (x_2, y_2) \in [0, 1]^2$. In our case, (X^0, ϱ) is bi-Lipschitz equivalent to R_α with $\alpha = \log 2 / \log 3$. It is well known that no quasisymmetry can lower the Hausdorff dimension

$$\dim_H(R_\alpha) = 1 + \log 3 / \log 2 > 2$$

of R_α (see [**He01**, Theorem 15.10]); in particular, R_α and hence also (X^0, ϱ), cannot be embedded into $\widehat{\mathbb{C}}$ by a quasisymmetry. This is a contradiction showing that f_4 is not Thurston equivalent to a rational map.

EXAMPLE 12.24. We now present a whole class of examples. In fact, the Lattès map $z \mapsto 1 - 2/z^2$ from Example 3.23, the map from Example 2.6, and the one from Example 12.6 are all members of this family. The construction of these maps is illustrated in Figure 12.9.

The starting point is the Lattès map $f_5(z) := 1 - 2/z^2$, which is
the map in our family of lowest degree. We briefly recall the geometric
description of this map as indicated in Figure 3.4. For this let T be the
right-angled isosceles Euclidean triangle whose hypotenuse has length
1; its angles are $\pi/2, \pi/4, \pi/4$. We also consider a triangle T' similar to
T and scaled by the factor $\sqrt{2}/2$. We obtain a pillow Δ by gluing two
isometric copies $T_{\mathtt{w}}$ and $T_{\mathtt{b}}$ of T together along their boundaries. The
pillow carries a natural cell decomposition $\mathcal{D}^0$ whose 0-tiles are $T_{\mathtt{w}}$ and
$T_{\mathtt{b}}$, with the common corners and sides of these triangles as 0-vertices
and 0-edges. A cell decomposition $\mathcal{D}^1$ of Δ is now obtained by subdi-
viding $T_{\mathtt{w}}$ and $T_{\mathtt{b}}$ by the bisectors perpendicular to their hypotenuses
into two triangles each. Then $\mathcal{D}^1$ contains four 1-tiles isometric to T'.
If we choose a labeling as indicated on the top of Figure 12.9 (cor-
responding to Figure 3.4), then we obtain a two-tile subdivision rule
$(\mathcal{D}^1, \mathcal{D}^0, L)$. It can be realized by a Thurston map $g = g_5 \colon \Delta \to \Delta$ that
sends each of the four small triangles isometric to T' to $T_{\mathtt{w}}$ or $T_{\mathtt{b}}$ by a
suitable similarity.

Since Δ is a polyhedral surface, it may naturally be viewed as a
Riemann surface. By the uniformization theorem there is a conformal
map $\varphi \colon \Delta \to \widehat{\mathbb{C}}$. It can be chosen so that its sends the 0-vertices of
Δ (i.e., the common corners of $T_{\mathtt{w}}$ and $T_{\mathtt{b}}$) to the points -1, 1, ∞. If
we conjugate g_5 by this map φ, then we obtain the Lattès map $f_5 =
\varphi \circ g \circ \varphi^{-1}$ (see Example 3.23 for more details). The homeomorphism φ
can be used in an obvious way to transfer $(\mathcal{D}^0, \mathcal{D}^1, L)$ to an isomorphic
two-tile subdivision rule $(\widetilde{\mathcal{D}}^1, \widetilde{\mathcal{D}}^0, \widetilde{L})$ on $\widehat{\mathbb{C}}$. It is realized by f_5.

Similarly to Example 12.22, we can modify the map g_5 as follows.
Namely, we take the pillow Δ as above, but now label the vertices of Δ
by $\omega, 1, \infty$ as shown in the middle right of Figure 12.9. Each side of Δ is
subdivided into two triangles isometric to T' as before. We cut $T_{\mathtt{w}}$ along
the perpendicular bisector of its hypotenuse and glue in two isometric
copies of T'. Informally, we refer to this procedure as "adding a flap".
This results in a surface Δ' homeomorphic to Δ that is built from six
isometric copies $T_1, \ldots, T_6$ of T'. There is a map $\widetilde{g}_6 \colon \Delta' \to \Delta$ that
sends each T_j by a similarity to $T_{\mathtt{w}}$ or $T_{\mathtt{b}}$ as indicated in the picture.
We can identify Δ' and Δ by a homeomorphism $\psi \colon \Delta' \to \Delta$ that
respects the correspondence of the common three corners of $T_{\mathtt{w}}$ and $T_{\mathtt{b}}$
(labeled $\omega, 1, \infty$ in both Δ' and Δ), sends the top of Δ' (consisting of
four small triangles) to $T_{\mathtt{w}}$, and the bottom of Δ' (consisting of two
small triangles) to $T_{\mathtt{b}}$. Then $g_6 := \widetilde{g}_6 \circ \psi^{-1}$ is a Thurston map with the
three postcritical points $\omega, 1, \infty$.
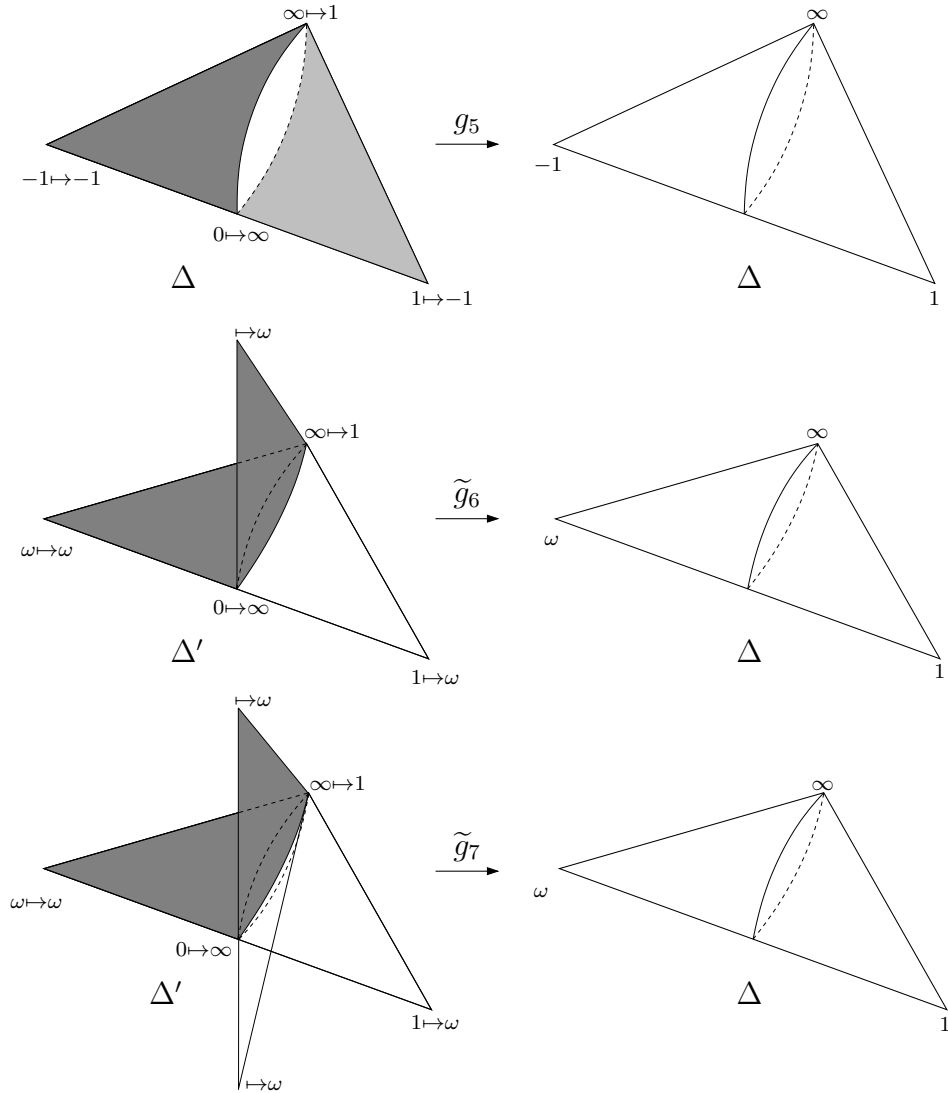
FIGURE 12.9. Adding flaps to a Lattès map.

This map was already considered in Example 2.6. There we saw that it is equivalent to the rational Thurston map $f_6(z) = 1 + (\omega - 1)/z^3$ with $\omega = e^{4\pi i/3}$.

It is possible to generalize this construction. For example, instead of adding just one flap on the top face $T_{\mathtt{w}}$ of Δ, we may add one flap on $T_{\mathtt{w}}$ and $T_{\mathtt{b}}$ each. This is illustrated at the bottom of Figure 12.9.

Moreover, instead of adding just one flap to the 1-edge bisecting $T_{\mathtt{w}}$, we can add $n \in \mathbb{N}_0$ flaps. Similarly, we can glue in $m \in \mathbb{N}_0$ flaps at the 1-edge bisecting $T_{\mathtt{b}}$. This again results in a polyhedral surface Δ' consisting of $2n + 2$ triangles isometric to T' on the top, and $2m + 2$
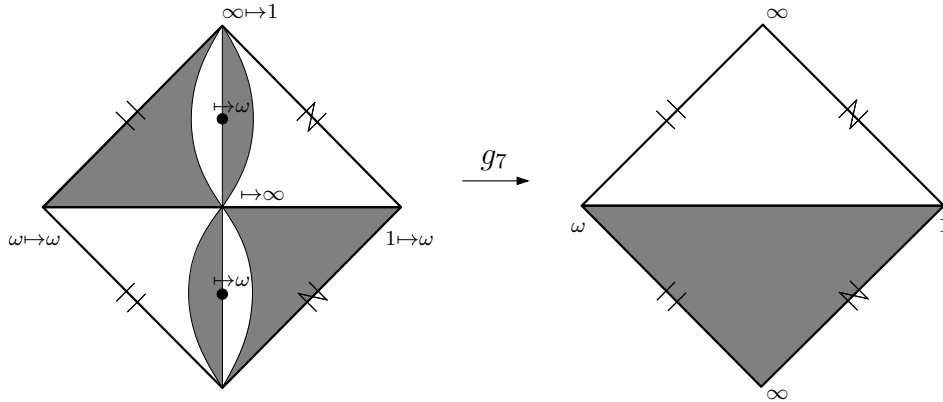
FIGURE 12.10. Two-tile subdivision rule realized by g_7.

small triangles isometric to T' on the bottom of Δ'. We label the
vertices of Δ by $\omega, 1, \infty$ as in the middle of Figure 12.9, and consider
these as vertices of Δ' as well. There is a unique small triangle $\widetilde{T}$ with
$\omega \in \widetilde{T}$ that is contained in the top part of Δ'. We color the tiles of Δ'
in checkerboard fashion so that $\widetilde{T}$ is black (as indicated in Figure 12.9).
Then with a proper choice of an orientation on Δ' there is a unique
branched covering map

$$\widetilde{g} \colon \Delta' \to \Delta$$

that sends each of the small triangles in Δ' to either $T_{\mathtt{w}}$ or $T_{\mathtt{b}}$ by a
similarity, fixes the vertex ω, and respects the coloring of tiles. Note
that $\widetilde{g}$ is not a Thurston map, because the domain Δ' and the range
Δ of $\widetilde{g}$ are different sets. To obtain a Thurston map, we consider,
as before, an identification of Δ' and Δ by an orientation-preserving
homeomorphism $\psi \colon \Delta' \to \Delta$ that preserves the correspondence of the
points labeled $\omega, 1, \infty$, sends the top of Δ' to $T_{\mathtt{w}}$, and the bottom of Δ'
to $T_{\mathtt{b}}$. Then

$$g \coloneqq \widetilde{g} \circ \psi^{-1} \colon \Delta \to \Delta$$

is a Thurston map with $\mathrm{post}(g) = \{\omega, 1, \infty\}$. Moreover, if $\mathcal{C}$ is the
common boundary of $T_{\mathtt{w}}$ and $T_{\mathtt{b}}$, then $\mathcal{C}$ is a g-invariant Jordan curve
with $\mathrm{post}(g) \subset \mathcal{C}$.

   Let g_7 be the Thurston map obtained in this way by adding $n = 1$
on the top and $m = 1$ flaps on the bottom. The two-tile subdivision
rule given by the map g_7 and its invariant curve $\mathcal{C}$ according to Propo-
sition 12.2 is shown in Figure 12.10. Here the edges on the top and
bottom have to be identified as indicated to obtain a 2-sphere.

   Since the Thurston map g has three postcritical points, it follows
from Theorem 7.3 that g is Thurston equivalent to a rational map

$f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$. In fact, we may choose f as

$$(12.12) \qquad\qquad f(z) = 1 + \frac{\omega - 1}{z^d},$$

where $d = n + m + 2$ and $\omega = e^{2\pi i(n+1)/d}$. Note that for $n = m = 1$ we obtain the rational map given by $f_7(z) = 1 - 2/z^4$, which is the map considered in Example 12.6. Thus the maps $f_5, f_6,$ and f_7 can be considered as special cases of our construction. The proof that g is Thurston equivalent to f is similar to the one given in Example 2.9; we omit the details.

The map f can be obtained more explicitly from $\widetilde{g}$ as follows. Since Δ' and Δ are polyhedral surfaces, they are Riemann surfaces. By the uniformization theorem there are conformal maps $\varphi\colon \Delta \to \widehat{\mathbb{C}}$ and $\widetilde{\varphi}\colon \Delta' \to \widehat{\mathbb{C}}$. We can normalize them so that $\widetilde{\varphi}(\omega) = \varphi(\omega) = \omega$, $\widetilde{\varphi}(1) = \varphi(1) = 1$, $\widetilde{\varphi}(\infty) = \varphi(\infty) = \infty$. Then one can show that $f = \varphi \circ \widetilde{g} \circ \widetilde{\varphi}^{-1}$ is exactly the map given in (12.12).

Let $\mathcal{C} \subset S^2$ be the g-invariant Jordan curve as before. Then g is combinatorially expanding for $\mathcal{C}$. We will see in Theorem 14.2 that by possibly choosing a different identification $\psi\colon \Delta' \to \Delta$ in the definition of g, we may assume that g is expanding. Since f is also expanding as follows from Proposition 2.3, the maps g and f are topologically conjugate by Theorem 11.1. So there is a homeomorphism $h\colon \Delta \to \widehat{\mathbb{C}}$ such that $f = h \circ g \circ h^{-1}$. Then $\widetilde{\mathcal{C}} := h(\mathcal{C}) \subset \widehat{\mathbb{C}}$ is an f-invariant Jordan curve with $\mathrm{post}(f) \subset \widetilde{\mathcal{C}}$. The previous argument is closely related to the general construction of invariant Jordan curves in Chapter 15. In Figure 15.1 the invariant Jordan curve $\widetilde{\mathcal{C}}$ is shown for the map f_6.

Let $(\mathcal{D}^1, \mathcal{D}^0, L)$ be the two-tile subdivision rule given by g and the g-invariant Jordan curve $\mathcal{C}$ according to Proposition 12.2, and $(\widetilde{\mathcal{D}}_1, \widetilde{\mathcal{D}}_0, \widetilde{L})$ be the one given by f and the f-invariant Jordan curve $\widetilde{\mathcal{C}}$. Then these two-tile subdivision rules are isomorphic (the isomorphism is naturally induced by h). In this sense, g and f realize the same two-tile subdivision rule.

EXAMPLE 12.25 (Subdivision rules from tilings). We now describe a general method for obtaining subdivisions rules from tilings of the Euclidean or hyperbolic plane. This can be used to find Thurston maps with an arbitrarily large set of postcritical points.

First, we consider the unit square $[0, 1]^2$ in $\mathbb{R}^2$ and its translates under the lattice $\mathbb{Z}^2 \subset \mathbb{R}^2$. These squares form a tiling of $\mathbb{R}^2$. More precisely, they are 2-dimensional cells or tiles of a cell decomposition $\mathcal{D}$ of $\mathbb{R}^2$ whose vertex set is $\mathbb{Z}^2$ and whose 1-skeleton is the "square grid" $S = (\mathbb{Z} \times \mathbb{R}) \cup (\mathbb{R} \times \mathbb{Z})$. Let $\mathcal{C}$ be a Jordan curve contained in S, and

$X \subset \mathbb{R}^2$ be the closed Jordan region with boundary $\mathcal{C}$. Then X is a union of tiles in $\mathcal{D}$. We assume X consists of at least two such tiles. We take two identical copies $X_{\mathtt{w}}$ and $X_{\mathtt{b}}$ of X (which we call 0-tiles), and glue them along their boundaries to form a pillow Δ. Note that this common boundary can be identified with $\mathcal{C}$. Among the vertices of $\mathcal{D}$ (i.e., the lattice points $\mathbb{Z}^2$) contained in $\mathcal{C} \subset \Delta$ we fix four distinct ones and consider them as the 0-vertices of Δ. The four arcs into which they divide $\mathcal{C}$ are the 0-edges. These 0-tiles, 0-edges, and 0-vertices form cell decomposition $\mathcal{D}^0$ of $\Delta \cong S^2$.

Our tiling of $\mathbb{R}^2$ by copies of $[0,1]^2$ gives a natural subdivision of X, and hence also of $X_{\mathtt{w}}$ and $X_{\mathtt{b}}$, into unit squares. Let $\mathcal{D}^1$ be the cell decomposition of $\Delta \cong S^2$ that is given by these squares as 1-tiles, their sides as 1-edges, and their corners as 1-vertices. Clearly, $\mathcal{D}^1$ is a refinement of $\mathcal{D}^0$ and every tile in $\mathcal{D}^1$ is a 4-gon. Every 1-vertex $v \notin \mathcal{C}$ is contained in four 1-tiles, and every 1-vertex $v \in \mathcal{C}$ is contained in the same number of tiles 1-tiles in $X_{\mathtt{w}}$ and in $X_{\mathtt{b}}$. It follows that every 1-vertex is contained in an even number of 1-tiles. This shows that the pair $(\mathcal{D}^1, \mathcal{D}^0)$ satisfies the conditions (i)–(iv) in Definition 12.1.

To define a corresponding labeling, we fix a 1-tile X', a 1-vertex $v' \in X'$, and a 0-vertex v. Since $v \in X_{\mathtt{w}}$ there is a unique labeling $L \colon \mathcal{D}^1 \to \mathcal{D}^0$ such that $L(X') = X_{\mathtt{w}}$ and $L(v') = v$ by Lemma 12.15 (ii). Then $(\mathcal{D}^1, \mathcal{D}^0, L)$ is a two-tile subdivision rule. By Proposition 12.3 it can be realized by Thurston map $f \colon \Delta \to \Delta$. Roughly speaking, f is constructed by mapping X' to $X_{\mathtt{w}}$, normalized such that $f(v') = v$, and extended to all of Δ "by reflection".

Note that the Lattès map from Section 1.1 and the Lattès-type map from Example 12.23 may be viewed as examples of this procedure. In [**HP12a**] Haïssinsky and Pilgrim constructed certain rational maps with Sierpiński carpet Julia sets in this way.

Instead of the square tilings, one can also use other tilings of $\mathbb{R}^2$ for this construction. The Lattès maps in Example 3.23, 3.24, 3.25, and the map in Example 12.21 are of this form. See also [**Me02**] and [**HM**].

Finally, we may use tilings of the hyperbolic plane instead. For example, if $n \geq 5$ is fixed, then one can tile the hyperbolic plane $\mathbb{H}^2$ with right-angled n-gons. This gives a cell decomposition $\mathcal{D}$ of $\mathbb{H}^2$ so that four n-gons intersect at each vertex. Again we consider a pillow Δ obtained by gluing together two copies of a closed Jordan region $X \subset \mathbb{H}^2$ whose boundary is contained in the 1-skeleton of $\mathcal{D}$ and encloses at least two n-gons of the tiling. If we define a Thurston map $f \colon \Delta \to \Delta$ by the analog of the above construction, then (under some mild additional assumptions) f will have n postcritical points.

More examples of maps constructed from subdivision rules can be found in [**C–P03**], and more examples of subdivisions in [**CFP06b**].

Figures 12.5, 12.7, and 15.1 show *symmetric conformal tilings* of $\widehat{\mathbb{C}}$. This means that if two tiles share an edge, they are conformal reflections of each other along this edge. The tiling can be produced by successive reflections, and so each individual tile encodes the information for the whole tiling.

CHAPTER 13

Quotients of Thurston maps

In this chapter we study the general problem when a Thurston map $f\colon S^2 \to S^2$ passes to another Thurston map on a quotient of S^2 induced by an equivalence relation $\sim$ (see Section A.7 for some basic facts about equivalence relations and quotient spaces).

For this, the quotient space $\widetilde{S}^2 := S^2/\!\sim$ equipped with the quotient topology has to be a 2-sphere itself. Well-known sufficient conditions for this to be the case are due to Moore (see Theorem 13.8). We call an equivalence relation $\sim$ on S^2 that satisfies these conditions to be of *Moore-type* (see Definition 13.7). So if $\sim$ is of Moore-type, then the quotient space $\widetilde{S}^2$ is a 2-sphere.

We denote by $[x] := \{y \in S^2 : y \sim x\}$ the equivalence class of a point and by $\pi\colon S^2 \to \widetilde{S}^2 = S^2/\!\sim$ the quotient map defined as $\pi(x) = [x] \in S^2/\!\sim$ for $x \in S^2$. The general question when $f\colon S^2 \to S^2$ *descends* to the quotient $\widetilde{S}^2$, i.e, when there exists a map $\widetilde{f}\colon \widetilde{S}^2 \to \widetilde{S}^2$ such that $\pi \circ f = \widetilde{f} \circ \pi$, is very easy to answer. Namely, $\sim$ needs to be *f-invariant* in the sense that we have the implication

(13.1) $$x \sim y \Rightarrow f(x) \sim f(y)$$

for all $x, y \in S^2$ (see Lemma A.20). This is equivalent to the requirement that

(13.2) $$f([x]) \subset [f(x)]$$

for each $x \in S^2$.

If $\sim$ is f-invariant and the map $f\colon S^2 \to S^2$ descends to a map $\widetilde{f}\colon \widetilde{S}^2 \to \widetilde{S}^2$, then we have the following commutative diagram:

(13.3)
$$
\begin{array}{ccc}
S^2 & \xrightarrow{\ f\ } & S^2 \\
\downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \pi} \\
\widetilde{S}^2 & \xrightarrow{\ \widetilde{f}\ } & \widetilde{S}^2.
\end{array}
$$

We call $\widetilde{f}$ the *quotient map* of f on $\widetilde{S}^2$. It is uniquely determined (by $\sim$ and f) and continuous.

317

Even if f is a Thurston map and $\sim$ is an f-invariant equivalence relation of Moore-type, it is not guaranteed that the map $\widetilde{f}$ as in (13.3) defined on the 2-sphere $\widetilde{S}^2$ is a Thurston map or even a branched covering map (see Example 13.17). For this we need a stronger condition on $\sim$.

DEFINITION 13.1 (Strongly invariant equivalence relations). Let $f\colon S^2 \to S^2$ be a branched covering map. Then an equivalence relation $\sim$ on S^2 is called *strongly f-invariant* if the image of each equivalence class is an equivalence class, or equivalently, if

$$f([x]) = [f(x)]$$

for each $x \in S^2$.

Clearly, each strongly f-invariant equivalence relation is f-invariant. Based on this concept of a strongly invariant equivalence relation, we can state the main result of this chapter.

THEOREM 13.2 (Quotients of branched covering maps). *Suppose $f\colon S^2 \to S^2$ is a branched covering map, $\sim$ is an f-invariant equivalence relation of Moore-type on S^2, $\pi\colon S^2 \to \widetilde{S}^2 \coloneqq S^2/\!\sim$ is the quotient map, and $\widetilde{f}\colon \widetilde{S}^2 \to \widetilde{S}^2$ is the induced map as in (13.3). Then $\widetilde{f}$ is a branched covering map if and only if $\sim$ is strongly f-invariant.*

Moreover, in this case the following statements are true:

(i) $\deg(\widetilde{f}) = \deg(f)$.

(ii) $\operatorname{crit}(\widetilde{f}) = \pi(\operatorname{crit}(f))$ *and* $\operatorname{post}(\widetilde{f}) = \pi(\operatorname{post}(f))$.

(iii) *If $x \in S^2$ and the equivalence class $[x]$ contains the (distinct) critical points $c_1, \ldots, c_n \in S^2$ of f, then the local degree of $\widetilde{f}$ at $[x] \in \widetilde{S}^2$ is given by*

$$(13.4) \qquad \deg(\widetilde{f}, [x]) = 1 + \sum_{i=1}^{n} (\deg(f, c_i) - 1).$$

We obtain the following immediate consequence.

COROLLARY 13.3 (Quotients of Thurston maps). *Suppose $f\colon S^2 \to S^2$ is a Thurston map, $\sim$ is a strongly f-invariant equivalence relation of Moore-type on S^2, and $\widetilde{f}\colon \widetilde{S}^2 \to \widetilde{S}^2$ the induced map on the quotient $\widetilde{S}^2 = S^2/\!\sim$ as in (13.3). Then $\widetilde{f}$ is a Thurston map.*

Moreover, $\deg(\widetilde{f}) = \deg(f)$, and, if $\pi\colon S^2 \to \widetilde{S}^2 = S^2/\!\sim$ is the quotient map, $\operatorname{crit}(\widetilde{f}) = \pi(\operatorname{crit}(f))$, and $\operatorname{post}(\widetilde{f}) = \pi(\operatorname{post}(f))$.

PROOF. It follows from Theorem 13.2 that $\widetilde{f}$ is a branched covering map on the 2-sphere $\widetilde{S}^2$ with the properties specified in the second part

of the statement. The map $\widetilde{f}$ is a actually a Thurston map because f is, and so

$$\# \operatorname{post}(\widetilde{f}) = \#\pi(\operatorname{post}(f)) \leq \# \operatorname{post}(f) < \infty$$

and $\deg(\widetilde{f}) = \deg(f) \geq 2$. $\qquad\qquad\square$

Quotients of rational maps (not necessarily postcritically-finite rational maps) were considered by McMullen in a somewhat different setting (see [**McM94a**, Appendix B]).

This chapter is organized as follows. In Section 13.1 we review some facts about equivalence relations relevant for the statement of Moore's theorem. In particular, we discuss the important concept of a *closed equivalence relation* (see Definition 13.4) and state various conditions that characterize closed equivalence relations.

In Section 13.2 we prove two facts about branched covering maps that are relevant for the proof of Theorem 13.2 (see Lemma 13.13 and Lemma 13.16).

Finally, in Section 13.3 we discuss some properties of strongly invariant equivalence relations (see Lemma 13.19) and establish a fact needed in the proof of Theorem 13.2 (see Lemma 13.20). The proof of this theorem concludes the section and the chapter.

13.1. Closed equivalence relations and Moore's theorem

We first take a closer look at the situation when a topological space $\widetilde{X}$ is obtained as the quotient of some other topological space X by an equivalence relation. Often $\widetilde{X}$ is then called a *decomposition space* (for a standard reference see [**Da86**]).

Let $\sim$ be an equivalence relation on a set X. As before, we denote by $[x]$ the equivalence class of any $x \in X$, by $X/\!\!\sim \;:= \{[x] : x \in X\}$ the quotient space, and by $\pi\colon X \to X/\!\!\sim$ the natural projection given by $\pi(x) = [x]$ for $x \in X$. If X is a topological space, then we equip $X/\!\!\sim$ with the quotient topology. See Section A.7 for more details.

A set $A \subset X$ is called *saturated* if $x \in A$ and $x \sim y$ implies $y \in A$ for all $x, y \in X$, or equivalently, if A is a union of equivalence classes.

We define the *saturated interior* of a set $U \subset X$ as

$$(13.5) \qquad U_s := \bigcup\{[x] : x \in X,\, [x] \subset U\}$$

$$= \bigcup\{A \subset X : A \text{ is saturated and } A \subset U\}.$$

The saturated interior of U is the largest saturated set contained in U.

LEMMA 13.4 (Closed equivalence relations). *Let X be a compact metric space and $\sim$ be an equivalence relation on X. Then the following conditions are equivalent:*

(i) *The set $\{(x, y) : x, y \in X, \ x \sim y\} \subset X \times X$ is closed.*

(ii) *Let $\{x_n\}_{n \in \mathbb{N}}$ and $\{y_n\}_{n \in \mathbb{N}}$ be convergent sequences in X. Then*

$$x_n \sim y_n \text{ for all } n \in \mathbb{N} \text{ implies } \lim_{n \to \infty} x_n \sim \lim_{n \to \infty} y_n.$$

(iii) *For each $x \in X$ and each neighborhood $U \subset X$ of $[x]$ there is a neighborhood $V \subset U$ of $[x]$ such that*

$$[y] \cap V \neq \emptyset \Rightarrow [y] \subset U \text{ for all } y \in X.$$

(iv) *For each open set $U \subset X$ the saturated interior U_s is open.*

In the first condition $X \times X$ is equipped with the product topology. Each of these four equivalent conditions implies that equivalence classes are closed, and hence compact subsets of X.

DEFINITION 13.5 (Closed equivalence relations). An equivalence relation $\sim$ on a compact metric space X is called *closed* if it satisfies one, and hence every, condition in Lemma 13.4.

In the literature closed equivalence relations are often called "upper semi-continuous" instead. One may define upper semi-continuous equivalence relations in any topological space (usually by condition (iii) from Lemma 13.4, together with the requirement that each equivalence class is compact). To simplify the discussion, we chose to restrict ourselves to compact metric spaces.

In [**MP12**] several other characterizations of closed equivalence relations can be found.

PROOF OF LEMMA 13.4. The equivalence of (i) and (ii) immediately follows from the definition of the product topology.

(ii) $\Rightarrow$ (iii) We first note that (ii) implies that each equivalence class $[x]$, $x \in X$, is closed, and hence compact. Indeed, if we choose $x_n = x \in X$ to be a constant sequence, then (ii) shows that the limit of any convergent sequence $\{y_n\}$ in $[x]$ is contained in $[x]$.

We now argue by contradiction and assume that (ii) is satisfied, but (iii) is not. Let $[x]$ be an equivalence class and U be a neighborhood of $[x]$ that violates (iii). Since $[x]$ is compact, it follows that for each sufficiently large $n \in \mathbb{N}$ the $1/n$-neighborhood of $[x]$ with respect to the underlying metric on X satisfies $\mathcal{N}_{1/n}([x]) \subset U$. By assumption, for each large enough $n \in \mathbb{N}$ there exists $x_n \in \mathcal{N}_{1/n}([x])$ such that $[x_n] \not\subset U$. This means there exists $y_n \in [x_n]$ with $y_n \notin U$. Taking a

subsequence, we may assume that $\{x_n\}$ and $\{y_n\}$ are convergent. Then $y := \lim_{n\to\infty} y_n$ is contained in the closure of $X \setminus U$, and so $y \notin [x]$. On the other hand, $x' := \lim_{n\to\infty} x_n \in [x]$, and so $x' \not\sim y$. This is a contradiction to (ii), proving the claim.

(iii) $\Rightarrow$ (iv) Let $U \subset X$ be an open set, and $x \in U_s$ be arbitrary. Then $[x] \subset U$, and so we can find a neighborhood $V \subset U$ of $[x]$ that satisfies the condition in (iii). This condition implies that $V \subset U_s$, and so U_s is a neighborhood of x. Since $x \subset U_s$ was arbitrary, it follows that U_s is open.

(iv) $\Rightarrow$ (ii) We first note that (iv) implies that every equivalence class is closed. Indeed, let $x \in X$ be arbitrary and consider the open set $U = X \setminus \{x\}$. Then $U_s = U \setminus [x]$. By (iv) the set U_s is open, and so $[x] = X \setminus U_s$ is closed.

Now let $\{x_n\}$ and $\{y_n\}$ be convergent sequences in X with $x_n \sim y_n$ for all $n \in \mathbb{N}$. Define $x = \lim_{n\to\infty} x_n$ and $y = \lim_{n\to\infty} y_n$. For $\epsilon > 0$ let $U = \mathcal{N}_\epsilon([x])$. Then by (iv) the set U_s is open, and it contains $x \in [x] \subset U_s$. Hence $x_n \in U_s$ and so $y_n \in U_s$ for sufficiently large n. This implies that $y \in \overline{U}_s \subset \overline{U}$, and so $\text{dist}(y, [x]) \le \epsilon$. Since $\epsilon > 0$ was arbitrary and $[x]$ is closed, we have $y \in [x]$ and so $x \sim y$ as desired. $\square$

An equivalence relation $\sim$ on a compact metric space X is called *monotone*, if every equivalence class of $\sim$ is connected. The following statement is well-known (see [**Da86**, Proposition 1.4.1, p. 18]). We provide a proof for the convenience of the reader.

LEMMA 13.6. *Let $\sim$ be a closed and monotone equivalence relation on a compact metric space X, and $\widetilde{K} \subset \widetilde{X} := X/\sim$ be a connected set. Then $K := \pi^{-1}(\widetilde{K}) \subset X$ is connected.*

Recall that the quotient space $\widetilde{X} = X/\sim$ is equipped with the quotient topology.

PROOF. Assume the statement is not true. Then there are disjoint open sets $U, V \subset X$ such that $K_U := K \cap U$ and $K_V := K \cap V$ are non-empty and $K \subset U \cup V$. Each equivalence class $[x] \subset K = \pi^{-1}(\widetilde{K})$ is either contained in K_U or in K_V, because $[x]$ is connected by our hypotheses. This implies that K_U and K_V are saturated sets.

Let U_s and V_s be the saturated interiors of U and V, respectively. Then U_s and V_s are disjoint, $K_U \subset U_s$ and $K_V \subset V_s$. Since $\sim$ is closed, U_s and V_s are open sets by condition (iv) in Lemma 13.4. Thus $\widetilde{U}_s := \pi(U_s)$ and $\widetilde{V}_s := \pi(V_s)$ are disjoint open sets in $\widetilde{X}$. Moreover, $\widetilde{U}_s \cap \widetilde{K} = \pi(K_U) \neq \emptyset$, $\widetilde{V}_s \cap \widetilde{K} = \pi(K_V) \neq \emptyset$, and $\widetilde{K} \subset \widetilde{U}_s \cup \widetilde{V}_s$. This contradicts our assumption that $\widetilde{K}$ is connected. $\square$

After these general considerations, we now turn to equivalence re-lations on a 2-sphere S^2.

DEFINITION 13.7 (Moore-type equivalence relations). An equiva-lence relation $\sim$ on S^2 is said to be of *Moore-type*, if the following conditions are satisfied:

(i) the equivalence relation $\sim$ is closed;

(ii) the equivalence relation $\sim$ is monotone;

(iii) no equivalence class of $\sim$ *separates* S^2, i.e., $S^2 \setminus [x]$ is connected for each $x \in S^2$;

(iv) the equivalence relation $\sim$ is *non-trivial*, i.e., there are at least two distinct equivalence classes.

The reason for our terminology is the following important theorem due to Moore. See [**Mo25**] for the original proof, [**Da86**, Theorem 25.1, p. 187] for a stronger statement, and [**Ca78**, Supplement 1] for a general discussion on the 2-sphere recognition problem.

THEOREM 13.8 (Moore). *Let $\sim$ be an equivalence relation on S^2 of Moore-type. Then the quotient space $S^2/\sim$ is homeomorphic to S^2.*

As before, it is understood that $S^2/\sim$ is equipped with the quotient topology. There is an equivalent description of Moore-type equivalence relations. To discuss this, we first require a definition.

DEFINITION 13.9 (Pseudo-isotopies). Let X and Y be topological spaces. A homotopy $H\colon X \times [0,1] \to Y$ is a *pseudo-isotopy* of X to Y if for each $t \in [0,1)$ the map $H(\cdot, t)\colon X \to Y$ is a homeomorphism.

So a pseudo-isotopy can fail to be a homeomorphism only at time $t = 1$. An equivalence relation $\sim$ on a topological space X is *realized* or *induced by a pseudo-isotopy* if there is a pseudo-isotopy $H\colon X \times [0,1] \to X$ with $H_0 = H(\cdot, 0) = \mathrm{id}_X$ such that

$$(13.6) \qquad\qquad x \sim y \Leftrightarrow H(x, 1) = H(y, 1)$$

for all $x, y \in X$.

Conversely, if a pseudo-isotopy $H\colon X \times [0,1] \to X$ with $H_0 = \mathrm{id}_X$ is given, then (13.6) defines an equivalence relation $\sim$ on X induced by H.

THEOREM 13.10. *Let $\sim$ be an equivalence relation on S^2. Then $\sim$ is of Moore-type if and only if $\sim$ is induced by a pseudo-isotopy $H\colon S^2 \times [0,1] \to S^2$.*

The "if"-direction is the easy implication in this statement. Its proof can be found in [**Me14**, Lemma 2.4]; the proof of the "only-if"-implication is quite involved and can be found in [**Da86**, Theorem 25.1 and Theorem 13.4].

COROLLARY 13.11. *Let $\sim$ an equivalence relation of Moore-type on S^2 and $\pi\colon S^2 \to \widetilde{S}^2 \coloneqq S^2/\!\sim$ be the quotient map. Then the induced map $\pi_*\colon H_2(S^2) \to H_2(\widetilde{S}^2)$ on singular homology groups is an isomorphism.*

Recall that we always assume that S^2 is oriented. The given orientation on S^2 can be represented by a generator $[S^2]$ of $H_2(S^2) \cong \mathbb{Z}$, called the fundamental class of S^2 (see Section A.4). By the corollary we may choose an orientation on the 2-sphere $\widetilde{S}^2$, such that $\pi_*([S^2])$ is the fundamental class on $\widetilde{S}^2$. With these choices we then have $\deg(\pi) = 1$ for the degree of π (as defined in Section A.4 in terms of the induced map on homology).

PROOF. We know by Theorem 13.10 that $\sim$ is induced by a pseudo-isotopy $H\colon S^2 \times [0,1] \to S^2$. Let $h \coloneqq H_1$ be the time-1 map. Then h is a map homotopic to $H_0 = \mathrm{id}_{S^2}$. Since the degree of a map is a homotopy invariant (see [**Ha02**, p. 134]), we conclude that $\deg(h) = \deg(\mathrm{id}_{S^2}) = 1$. In particular, h is surjective (it is a standard fact that a non-surjective continuous map on S^2 is null-homotopic and so has vanishing degree; see [**Ha02**, p. 134]).

Since $\sim$ is induced by H, we know that $x \sim y$ if and only if $h(x) = h(y)$ for all $x, y \in X$. This allows us to define a map $\varphi\colon \widetilde{S}^2 \to S^2$ as follows. If $x \in X$ we set $\varphi([x]) \coloneqq h(x)$. This map is well-defined, and a homeomorphism of $\widetilde{S}^2 = S^2/\!\sim$ onto S^2 (see Lemma A.19 (ii)).

Note that $h = \varphi \circ \pi$. Since φ is a homeomorphism, we can choose a fundamental class $[\widetilde{S}^2]$ on the 2-sphere $\widetilde{S}^2$, i.e., a generator $[\widetilde{S}^2]$ of $H_2(\widetilde{S}^2) \cong \mathbb{Z}$, such that $\varphi_*([\widetilde{S}^2]) = [S^2]$. With this choice $\deg(\varphi) = 1$ and so

$$\deg(\pi) = \deg(\varphi) \cdot \deg(\pi) = \deg(\varphi \circ \pi) = \deg(h) = 1.$$

This implies that $\pi_*([S^2]) = [\widetilde{S}^2]$ and so π_* is an isomorphism. $\square$

In the previous proof we saw explicitly how to find a homeomorphism φ between the quotient space $\widetilde{S}^2 = S^2/\!\sim$ and the 2-sphere S^2 if the equivalence relation $\sim$ is induced by a pseudo-isotopy. This shows that Theorem 13.10 implies Theorem 13.8, and so Theorem 13.10 can be regarded as a stronger version of Moore's theorem.

In Chapter 14 we will also require a 1-dimensional version of Moore's theorem. It can easily be derived from the topological characterization of arcs and topological circles (in equivalent form this is stated in [**Wh42**, Section 9.1, (1.1), p. 165] or as two exercises in [**Da86**, Exercise 4.2 and 4.3, p. 21]).

PROPOSITION 13.12. *Let J be an arc or a topological circle, and $\sim$ be an equivalence relation on J. Suppose that*

(i) *each equivalence class of $\sim$ is a compact and connected subset of J,*

(ii) *there are at least two distinct equivalence classes.*

Then the quotient space $\widetilde{J} = J/\sim$ is an arc or a topological circle, respectively.

13.2. Branched covering maps and continua

In this section we establish two topological properties of branched covering maps formulated in Lemma 13.13 and Lemma 13.16. They are needed for the proof of Theorem 13.2.

LEMMA 13.13. *Let X and Y be a compact metric spaces, $f\colon X \to Y$ be an open and continuous map, and $K \subset Y$ be a compact connected set. Then each component C of $f^{-1}(K)$ satisfies $f(C) = K$.*

In particular, this applies to the situation where $X = Y$ is a 2-sphere S^2 and f is a branched covering map f on S^2. In this context, a similar statement is also true for open connected sets: if $V \subset S^2$ is a region and $U \subset S^2$ a component of $f^{-1}(V)$, then $f(U) = V$ (see Lemma A.8 (ii)).

Lemma 13.13 was proved in [**Wh42**, Theorem 7.5, p. 148]. Since it is a bit hard to see the main ideas of the argument in this reference, we decided to include a proof. We need the following fact.

THEOREM 13.14 (Šura-Bura). *Let X be a compact metric space. Then every component C of X is the intersection of all clopen subsets of X that contain C.*

Here a *clopen* subset of a topological space is a set that is both open and closed. Proofs of Theorem 13.14 can be found in [**Bu79**, Corollary 1.34] and [**Re98**, Appendix to Chapter 14]. The theorem is in fact still true for locally compact Hausdorff spaces. We will use a slight variant of the Šura-Bura theorem.

COROLLARY 13.15. *Let X be a compact metric space and C be a component of X. Then there is a nested sequence $A_1 \supset A_2 \supset \ldots$ of clopen subsets of X such that $C = \bigcap_n A_n$.*

PROOF. Let C be a component of X. We define

$$\mathcal{A} := \{A \subset X : C \subset A, \ A \text{ clopen in } X\}, \text{ and}$$
$$\mathcal{B} := \{B \subset X : B \cap C = \emptyset, \ B \text{ clopen in } X\}$$
$$= \{X \setminus A : A \in \mathcal{A}\}.$$

Then $\bigcap_{A \in \mathcal{A}} A = C$ by the Šura-Bura theorem, which is equivalent to $\bigcup_{B \in \mathcal{B}} B = X \setminus C$.

Now for each $n \in \mathbb{N}$ we consider the set

$$K_n := \{x \in X : \text{dist}(x, C) \geq 1/n\}.$$

This is a compact subset of X that is disjoint from C. Thus it is covered by finitely many sets in $\mathcal{B}$. Since $\mathcal{B}$ is stable under taking finite unions of sets in $\mathcal{B}$, there exists *one* set $B_n \in \mathcal{B}$ with $K_n \subset B_n$.

Now define $A_n = X \setminus (B_1 \cup \cdots \cup B_n)$ for $n \in \mathbb{N}$. Then A_n is clopen, $C \subset A_n$, and $A_n \supset A_{n+1}$ for $n \in \mathbb{N}$. Moreover, we have

$$C \subset \bigcap_n A_n \subset \bigcap_n (X \setminus B_n) \subset \bigcap_n (X \setminus K_n) = C.$$

Here the last equality follows from the fact that C is closed. We conclude that $\bigcap_n A_n = C$ as desired. □

PROOF OF LEMMA 13.13. Under the given assumptions, consider the set $Z := f^{-1}(K)$, and let C be a component of Z. Then Z equipped with the restriction of the metric on X is a compact metric space itself, and we can apply Corollary 13.15 to Z.

Let $A \subset Z$ be clopen in Z. Then $f(A) \subset K$ is clopen in K. Indeed, since A is open in Z, there exists an open set $U \subset X$ such that $A = U \cap Z$. Then $f(A) = f(U) \cap K$, because $y \in f(U) \cap K$ if and only if there exists $x \in U \cap f^{-1}(K) = A$ with $f(x) = y$. Since f is an open map, $f(U)$ is open. This means that $f(A) = f(U) \cap K$ is open in K.

Similarly, A is closed in Z, and hence a compact subset of X. Therefore, $f(A) \subset K$ is compact, and hence closed in K.

In particular, if $A \subset Z$ is non-empty and clopen in Z, then $f(A)$ is non-empty and clopen in K. This implies that $f(A) = K$, since K is connected.

Now let $A_1 \supset A_2 \ldots$ be a decreasing sequence of clopen sets in Z with $\bigcap_n A_n = C$, as in Corollary 13.15. Then $f(A_n) = K$ for each $n \in \mathbb{N}$ by what we have just seen.

Now let $p \in K$ be arbitrary. Then for each $n \in \mathbb{N}$ there exists $q_n \in A_n$ such that $f(q_n) = p$. Since X is compact, by passing to a subsequence if necessary, we may assume that the sequence $\{q_n\}$

converges, say $q_n \to q \in X$ as $n \to \infty$. Then $f(q) = \lim_{n\to\infty} f(q_n) = p$ by continuity of f.

Since the sets A_n are decreasing, we have $q_k \in A_n$, whenever $k \geq n$. Since A_n is closed in Z and hence also in X, this implies that $q \in A_n$ for each $n \in \mathbb{N}$. Then $q \in \bigcap_n A_n = C$, and so $p = f(q) \in f(C)$. Since $p \in K$ was arbitrary, we conclude $f(C) = K$ as desired. $\square$

Before we formulate the next lemma, we discuss some simple facts about branched covering maps between regions in S^2. Suppose $V \subset S^2$ is a *finitely connected region*, i.e., a region with finitely many complementary components. Then the *Euler characteristic $\chi(V)$* of V is given by $\chi(V) = 2 - k_V$, where $k_V \in \mathbb{N}_0$ is the number of complementary components of V. Note that k_V is also equal to the number of components of ∂V.

Let $f\colon S^2 \to S^2$ be a branched covering map and U be a connected component of $f^{-1}(V)$. Then $f(U) = V$ and the map $f|U\colon U \to V$ is proper (see Lemma A.8 (ii)). Moreover, each point $q \in V$ has the same number of preimages under f in U if we count multiplicity given by the local degree of f at a preimage point. This number is called the *degree of f on U* and denoted by $\deg(f|U)$. So

$$\deg(f|U) = \sum_{p\in U\cap f^{-1}(q)} \deg_f(p)$$

for each $q \in V$. By a variant of the Riemann-Hurwitz formula we have

$$(13.7) \qquad \deg(f|U) \cdot \chi(V) = \chi(U) + \sum_{c\in U\cap\mathrm{crit}(f)} (\deg_f(c) - 1).$$

Implicitly this includes the statement that U is also a finitely-connected region.

All of this is well known if f is a rational map on the Riemann sphere (see, for example, [**Be87**, Section 5.4]); these facts are also true for a general branched covering map f, since we can reduce this case to the rational one by Corollary A.12.

In particular, if $f|U\colon U \to V$ is a covering map, then

$$\deg(f|U) \cdot \chi(V) = \chi(U).$$

If $k_V = 2$, then V is called a *ring domain*. In this case, $\chi(V) = 0$, which implies that $\chi(U) = 0$. In other words, a finite cover of a ring domain under f is also a ring domain.

LEMMA 13.16. *Let $f\colon S^2 \to S^2$ be a branched covering map, $K \subset S^2$ be a compact connected set, $C \subset S^2$ be a component of $f^{-1}(K)$, and*

$V \subset S^2$ be a Jordan region with $K \subset V$ such that $\overline{V} \setminus K$ contains no critical value of f. Suppose that neither K nor C separates S^2.

Then the unique component U of $f^{-1}(V)$ that contains C is a Jordan region that contains no other component of $f^{-1}(K)$. Moreover, for the degree of the proper map $f\colon U \to V$ we have

$$(13.8) \qquad \deg(f|U) = 1 + \sum_{c \in C \cap \mathrm{crit}(f)} (\deg_f(c) - 1).$$

PROOF. Note that C is a connected set in $f^{-1}(K) \subset f^{-1}(V)$, and so is contained in a unique component U of $f^{-1}(V)$. In the ensuing argument, we will actually define U in a different, less direct way. This will make it easier to establish our claims.

We start by considering the open set $R := V \setminus K$. It has two complementary components, namely K and the closed Jordan region $S^2 \setminus V$. Since neither K nor $S^2 \setminus V$ separate S^2, their union $S^2 \setminus R$ does not separate S^2 either (this follows from Janiszewski's lemma; see Lemma A.3). So R is connected and hence a ring domain.

We can find a connected component R' of $f^{-1}(R)$ such that $C \cap \partial R' \neq \emptyset$. To see this, we run along some path in V from a point in $R \subset V$ towards K until we first hit K. In this way, we can find a path γ in V whose endpoint y lies in K, but that has no other points with K in common. By Lemma 13.13 we have $f(C) = K$ and so we can find a point $x \in C$ with $f(x) = y$. We can lift the path γ under f to a path α that ends in x (see Lemma A.17). Then

$$f(\alpha \setminus \{x\}) = \gamma \setminus \{y\} \subset V \setminus K = R,$$

and so the connected set $\alpha \setminus \{x\}$ must lie in a component R' of $f^{-1}(R)$. Then $x \in C \cap \partial R'$.

By a similar path lifting argument one can also see that there exists a component J of $f^{-1}(\partial V)$ such that $J \cap \partial R' \neq \emptyset$.

Since ∂V is a Jordan curve that does not contain any critical values of f, all components of $f^{-1}(\partial V)$, and in particular J, are Jordan curves. The sets J, R', and C, are all disjoint, because f maps them to the disjoint sets ∂V, R, and K, respectively. In particular, the connected sets R' and C must each be contained in one of the two complementary components of J. Let U be the complementary component of J that contains R'. Then U is a Jordan region and we also have $C \subset U$, as follows from $C \cap \partial R' \neq \emptyset$.

We know that $f(R') = R$, because R' is a component of $f^{-1}(R)$ (see Lemma A.8 (ii)). Moreover, R' contains no critical points of f, since $R = f(R') = V \setminus K$ contains no critical value of f. This implies that $f|R'$ is a covering map of R' onto R.

To see this, let $q \in R$ be arbitrary. We have to find a neighborhood of q that is evenly covered by the map $f|R'$. For this we choose a small topological disk $D \subset S^2$ with $q \in D \subset R$ that is evenly covered by the branched covering map f as in Definition A.7 (see Lemma A.10). If D' is a component of $f^{-1}(D)$ and p' the unique point in D' with $f(p') = q$, then either $D' \cap R' = \emptyset$ or $D' \subset R'$ and $\deg_f(p') = 1$. So in the latter case f is a homeomorphism of D' onto D. It easily follows that D is evenly covered by the map $f|R \colon R' \to R$ in the sense of (unbranched) covering maps.

Since $f|R' \colon R' \to R$ is a covering map and R is a ring domain, we conclude that R' is a ring domain as well (see the discussion before the statement of the lemma). This implies that the boundary $\partial R'$ of R' has precisely two connected components B_1 and B_2. It follows from Lemma A.8 (ii) that

$$f(B_1) \cup f(B_2) = f(B_1 \cup B_2) = f(\partial R') \subset \partial R \subset K \cup \partial V.$$

The sets K and ∂V are compact and disjoint, and the sets $f(B_1)$ and $f(B_2)$ are connected. Hence each of the sets $f(B_1)$ and $f(B_2)$ is completely contained in one of the sets K or ∂V.

Since $C \cap \partial R' \neq \emptyset$, one of the sets B_1 or B_2 must meet C, say $C \cap B_1 \neq \emptyset$. Since $f(C) = K$, this forces $f(B_1) \subset K$ by what we have just seen. Then $C \cup B_1$ is a connected subset of $f^{-1}(K)$. Since C is a component of $f^{-1}(K)$, it follows that $B_1 \subset C$.

We also know that $J \cap \partial R' \neq \emptyset$, and so one of the sets B_1 or B_2 must meet J. Since $B_1 \subset C$, we necessarily have $B_2 \cap J \neq \emptyset$. This forces $f(B_2) \subset \partial V$, and by a similar argument as for B_1 we see that $B_2 \subset J$.

We now consider the set $U \setminus C$. This is a ring domain, because its complement has the two connected components C and $S^2 \setminus U$, whose union does not separate S^2 by Janiszewski's lemma. We know that R' is open and that $R' \subset U \setminus C$. Moreover, R' is relatively closed in $U \setminus C$, because

$$\partial R' = B_1 \cup B_2 \subset C \cup J = C \cup \partial U,$$

and so R' has no boundary points in $U \setminus C$. Since $U \setminus C$ is connected, we conclude that $R' = U \setminus C$.

In particular, $U = R' \cup C$ and so

$$f(U) \subset f(R') \cup f(C) \subset V.$$

This implies $U \subset f^{-1}(V)$. Now U is a Jordan region and hence connected. Moreover, it is a maximal connected set in $f^{-1}(V)$, because any point in $f^{-1}(V)$ not in U is separated from U by the Jordan curve $J \subset f^{-1}(\partial V)$ that lies in the complement of $f^{-1}(V)$. Hence U is a

component of $f^{-1}(V)$. Moreover, C is the only component of $f^{-1}(K)$ contained in U, because $U = R' \cup C$ and $f(R') = R = V \setminus K$, which implies that R' is disjoint from $f^{-1}(K)$.

Finally, (13.8) follows from the Riemann-Hurwitz formula (13.7). Indeed, U and V are Jordan regions and so $\chi(U) = \chi(V) = 1$; moreover, the only critical points of f in $U = R' \cup C$ are those contained in C, because R' does not contain any. $\qquad\square$

13.3. Strongly invariant equivalence relations

We now consider the question when a given Thurston map descends to a quotient map that is itself a Thurston map. More precisely, the setting is as follows. Let $\sim$ be an equivalence relation on a 2-sphere S^2. As before, we denote by $[x]$ the equivalence class of a point $x \in S^2$. Let $\widetilde{S}^2 := S^2/\!\sim$ be the quotient space equipped with the quotient topology and $\pi\colon S^2 \to \widetilde{S}^2$ be the projection given by $x \mapsto \pi(x) = [x] \in \widetilde{S}^2$ for $x \in S^2$. In this section, $\sim$ will often be of Moore-type, in which case the quotient space $\widetilde{S}^2$ is also a 2-sphere.

Condition (13.1) is necessary for f to descend to a Thurston map $\widetilde{f}$. However, this condition is not sufficient even when $\sim$ is of Moore-type, as the following example shows.

EXAMPLE 13.17. Let $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be the Thurston map given by $f(z) = z^2$. Let $\sim$ be the equivalence relation on $\widehat{\mathbb{C}}$ that is obtained by collapsing the positive real line $[0, \infty] \subset \widehat{\mathbb{C}}$ to a point, meaning that

$$x \sim y \;:\Leftrightarrow\; x, y \in [0, \infty] \text{ or } x = y$$

for $x, y \in \widehat{\mathbb{C}}$. Clearly, this is an equivalence relation that is f-invariant and of Moore-type. Thus by Theorem 13.8 (Moore's theorem) the quotient $\widetilde{S}^2 := \widehat{\mathbb{C}}/\!\sim$ is a 2-sphere, and by Lemma A.20 there is a continuous map $\widetilde{f}\colon \widetilde{S}^2 \to \widetilde{S}^2$ as in (13.3). However, the map $\widetilde{f}$ is not a branched covering map, and hence not a Thurston map. Indeed, note that in $\widetilde{S}^2$ all points $[x]$ with $x \in (-\infty, 0]$ are distinct, but $\widetilde{f}$ maps each such point to $[x^2] = [0] \in \widetilde{S}^2$. Thus the point $[0] \in \widetilde{S}^2$ has infinitely many preimages under $\widetilde{f}$, which is impossible for a branched covering map on the 2-sphere $\widetilde{S}^2$.

The map $\widetilde{f}$ can be described as follows. First $\widetilde{f}$ collapses an equator of the sphere $\widetilde{S}^2$ to a point. This results in two topological 2-spheres that are connected at one point. Then $\widetilde{f}$ maps each of these two spheres (by orientation-preserving homeomorphisms) to the sphere $\widetilde{S}^2$.

In contrast, it can happen that a Thurston map descends to a Thurston map on a 2-sphere quotient $S^2/\!\sim$ even though $\sim$ is not of Moore-type.

EXAMPLE 13.18. We again consider the rational Thurston map $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ given by $f(z) = z^2$. Let $\sim$ be the equivalence relation on $\widehat{\mathbb{C}}$ defined as $z \sim w$ if and only if $w = \pm z$ for $z, w \in \widehat{\mathbb{C}}$. Clearly, all equivalence classes except $[0]$ and $[\infty]$ are disconnected and so $\sim$ is not of Moore-type.

Since the equivalence relation $\sim$ is f-invariant, f descends a to map $\widetilde{f}\colon \widehat{\mathbb{C}}/\!\sim \; \to \widehat{\mathbb{C}}/\!\sim$ as in (13.3). We claim that $\widehat{\mathbb{C}}/\!\sim$ is a 2-sphere and $\widetilde{f}$ is topologically conjugate to f. In particular, $\widetilde{f}$ is also a Thurston map.

Indeed, if $\overline{\overline{\mathbb{H}}} = \{z \in \widehat{\mathbb{C}} : \operatorname{Im}(z) \geq 0\}$ denotes the closed upper half-plane, then $\widehat{\mathbb{C}}/\!\sim \; = \overline{\overline{\mathbb{H}}}/\!\sim$ and $\sim$ identifies the points x and $-x$ for $x \in (0, \infty) \subset \partial\mathbb{H}$ and no other points in $\overline{\overline{\mathbb{H}}}$. This implies that $\widehat{\mathbb{C}}/\!\sim$ is a 2-sphere. An explicit homeomorphism $h\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}/\!\sim$ is given by $h(z) := [\sqrt{z}] \in \widehat{\mathbb{C}}/\!\sim$ for $z \in \widehat{\mathbb{C}}$, as can easily be verified.

Let $\pi\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}/\!\sim$ be the quotient map. Then $\widetilde{f}\colon \widehat{\mathbb{C}}/\!\sim \; \to \widehat{\mathbb{C}}/\!\sim$ is given by $\widetilde{f}([z]) = \widetilde{f}(\pi(z)) = \pi(f(z)) = [z^2]$ for $z \in \widehat{\mathbb{C}}$. Thus $(h \circ f)(z) = [\sqrt{z^2}] = [(\sqrt{z})^2] = (\widetilde{f} \circ h)(z)$ for $z \in \widehat{\mathbb{C}}$. This means that f and $\widetilde{f}$ are topologically conjugate by the homeomorphism h.

Our general criterion for obtaining branched covering maps $\widetilde{f}$ on quotients $S^2/\!\sim$ as formulated in Theorem 13.2 uses the notion of a strongly f-invariant equivalence relation as in Definition 13.1. The following statement puts this condition into perspective.

LEMMA 13.19 (Strongly invariant equivalence relations). *Suppose* $f\colon S^2 \to S^2$ *is a branched covering map, and $\sim$ is an f-invariant equivalence relation of Moore-type on S^2. Then the following conditions are equivalent:*

(i) *The equivalence relation $\sim$ is strongly f-invariant.*

(ii) *If $y \in S^2$, then each component of $f^{-1}([y])$ is a single equivalence class.*

(iii) *If $y \in S^2$, then $f^{-1}([y])$ is a union of finitely many equivalence classes.*

(iv) *The induced map $\widetilde{f}$ in (13.3) is discrete.*

PROOF. (i) $\Rightarrow$ (ii) Let $y \in S^2$ be arbitrary and C be a component of $f^{-1}([y])$. If $x \in C$, then $[x]$ is connected, since $\sim$ is monotone, and $f([x]) \subset [y]$, since $\sim$ is f-invariant. So $[x] \subset C$, showing that C is a union of equivalence classes. Each such equivalence class $[x] \subset C$ is

mapped by assumption (i) to $[y]$, i.e., $f([x]) = [y]$. Thus $[x]$ contains a point from the finite set $f^{-1}(y)$, and so C consists of finitely many equivalence classes. Since $\sim$ is of Moore-type, each equivalence class is a compact connected set. A finite union of such pairwise disjoint sets is disconnected, unless it consists of a single such set. This implies that C consists of a single equivalence class as desired.

(ii) $\Rightarrow$ (iii) Let $y \in S^2$ be arbitrary. By our assumption (ii), each component of $f^{-1}([y])$ is an equivalence class $[x]$. Then $f([x]) = [y]$ by Lemma 13.13. Thus $[x]$ contains a point from the finite set $f^{-1}(y)$. Hence there are only finitely many such equivalence classes, or components of $f^{-1}([y])$.

(iii) $\Rightarrow$ (iv) Since $\sim$ is f-invariant, there exists a well-defined continuous map $\widetilde{f} \colon \widetilde{S}^2 \to \widetilde{S}^2$ as in (13.3) (see Lemma A.20).

Now consider an arbitrary point in $\widetilde{S}^2$ as given by an equivalence class $[y] \in \widetilde{S}^2$, $y \in S^2$. Then by (13.3) we have $[x] \in \widetilde{f}^{-1}([y])$ for $x \in S^2$ if and only if $f([x]) \subset [y]$ or $[x] \subset f^{-1}([y])$. By assumption (iii), there are only finitely many such equivalence classes $[x]$. Thus, $[y]$ has only finitely many preimages under $\widetilde{f}$, and so $\widetilde{f}$ is discrete.

(iv) $\Rightarrow$ (i) Let $x \in S^2$ be arbitrary, $y = f(x)$, and C be the component of $f^{-1}([y])$ that contains x. Then $f(C) = [y]$ by Lemma 13.13.

If $x' \in C$ is arbitrary, then $f(x') \in f(C) = [y]$, and so $f(x') \sim y$. The f-invariance of $\sim$ implies $f([x']) \subset [f(x')] = [y]$, and so $[x'] \subset f^{-1}([y])$. Now $\sim$ is monotone, and so $[x']$ is connected. We conclude that $[x'] \subset C$. So C is saturated, i.e., a union of equivalence classes. Each of these equivalence classes is mapped into $[y]$, and hence a preimage of $[y] \in \widetilde{S}^2 = S^2/\!\sim$ under $\widetilde{f}$.

Since $\sim$ is of Moore-type, the quotient space $\widetilde{S}^2 = S^2/\!\sim$ is a 2-sphere. Moreover, since $\widetilde{f}$ is discrete by our assumption (iv), it must be finite-to-one. So $[y]$ can have only finitely many preimages under $\widetilde{f}$. By what we have seen, this implies that C consists of finitely many equivalence classes. Since $\sim$ is closed and monotone, each of these equivalence classes is compact and connected. So C can only be connected if it consists of a single equivalence class, i.e., $C = [x]$. Hence $f([x]) = f(C) = [y] = [f(x)]$, and (i) follows. $\qquad\square$

To prove Theorem 13.2, we first consider the mapping behavior of f near an individual equivalence class.

LEMMA 13.20. *Suppose $f \colon S^2 \to S^2$ is a branched covering map, and $\sim$ is an equivalence relation of Moore-type on S^2 that is strongly*

f-invariant. Let $x \in S^2$ and $U' \subset S^2$ be a neighborhood of $[x]$. Then there exists a neighborhood $U \subset U'$ of $[x]$ with the following properties:

- (i) *U is a Jordan region.*
- (ii) *$U \setminus [x]$ does not contain any critical point of f.*
- (iii) *The restriction $f|U \colon U \to f(U)$ is a proper map.*
- (iv) *If $c_1, \ldots, c_n$ are the (distinct) critical points of f contained in $[x]$, then the degree of f on U is given by*

$$\deg(f|U) = 1 + \sum_{i=1}^{n} (\deg(f, c_i) - 1).$$

In particular, if $[x]$ does not contain a critical point, the map $f|U \colon U \to f(U)$ is a homeomorphism.

Recall that $\deg(f|U)$ was defined as the (constant) number of preimages of a point $q \in f(U)$ counting multiplicities (see the discussion before Lemma 13.16).

Note that under our given assumptions on $\sim$ and f, we immediately obtain the following implication from Lemma 13.20 (iv):

(13.9) $[x]$ does not contain a critical point $\Rightarrow$

f is a homeomorphism of $[x]$ onto $f([x])$.

PROOF. Let $\sim$, f, $x \in S^2$, and $U' \subset S^2$ be as in the statement of the lemma. Moreover, let $c_1, \ldots, c_n$ be the critical points contained in $[x]$, and $y \coloneqq f(x)$.

Since $\sim$ is strongly f-invariant, we know that $f([x]) = [y]$. By statements (ii) and (iii) in Lemma 13.19 the components of $f^{-1}([y])$ are given by finitely many distinct equivalences classes, say $[x_1], \ldots, [x_k]$. Here $[x]$ is one of these classes, and so we may assume $x_1 = x$.

If we equip S^2 with a base metric that induces the given topology, then we can choose $\epsilon > 0$ so small that the neighborhoods

$$\mathcal{N}_\epsilon([x_1]), \ldots, \mathcal{N}_\epsilon([x_k])$$

are all disjoint and $\mathcal{N}_\epsilon([x_1]) = \mathcal{N}_\epsilon([x]) \subset U'$. We can find a corresponding $\delta > 0$ such that $f^{-1}(\mathcal{N}_\delta([y])) \subset \mathcal{N}_\epsilon(f^{-1}([y]))$ (see Lemma 5.15).

We now choose a Jordan region $V \subset S^2$ such that $[y] \subset V \subset \overline{V} \subset \mathcal{N}_\delta([y])$ and $\overline{V} \setminus [y]$ does not contain any critical value of f. It is clear that such a region V exists if $[y]$ is the singleton set $\{y\}$. If $[y]$ contains at least two points, then $[y]$ is a non-degenerate continuum. The existence of V is then most easily established by identifying S^2 with $\widehat{\mathbb{C}}$ under some homeomorphism. Then by the Riemann mapping theorem there exists a conformal map $\varphi \colon \mathbb{D} \to \widehat{\mathbb{C}} \setminus [y]$. If we now define

$V = \widehat{\mathbb{C}} \setminus \varphi(\overline{B}_{\mathbb{C}}(0, r))$ with $r \in (0, 1)$ sufficiently close to 1, then V has the desired properties.

Since $\sim$ is of Moore-type, neither $[y]$ nor any of the components $[x_1], \ldots, [x_n]$ of $f^{-1}([y])$ separates S^2. If U is the unique component of $f^{-1}(V)$ that contains the component $[x] = [x_1]$ of $f^{-1}([y])$, then U is a Jordan region by Lemma 13.16. Moreover, by choice of V, we have $U \subset \mathcal{N}_\epsilon(f^{-1}([y]))$. Since the connected set U can only meet one of the disjoint open sets $\mathcal{N}_\epsilon([x_1]), \ldots, \mathcal{N}_\epsilon([x_k])$, whose union is equal to $\mathcal{N}_\epsilon(f^{-1}([y]))$, it follows that $U \subset \mathcal{N}_\epsilon([x]) \subset U'$.

Now U has clearly properties (i) and (ii) as in the statement. By Lemma A.8 (ii) the map $f|U \colon U \to V$ is proper and $f(U) = V$. Property (iii) follows.

Since U and V are Jordan regions, we have $\chi(U) = \chi(V) = 1$ for their Euler characteristics. Since the only critical points of f in U are those contained in $[x]$, the identity for $\deg(f|U)$ is implied by (13.8) in Lemma 13.16. If $[x]$ does not contain critical points, then $\deg(f|U) = 1$ and so f is a homeomorphism of U onto $V = f(U)$. Statement (iv) follows. $\qquad\square$

We can now prove the main result of this chapter.

PROOF OF THEOREM 13.2. Let $f \colon S^2 \to S^2$ be a branched covering map, and $\sim$ be an f-invariant equivalence relation on S^2 of Moore-type.

Assume first that $\widetilde{f}$ is a branched covering map. Then $\widetilde{f}$ is discrete, and thus $\sim$ is strongly f-invariant by Lemma 13.19.

Conversely, suppose that $\sim$ is strongly f-invariant. In order to see that f is a branched covering map, we want to apply the criterion provided by Corollary A.13.

First, the map $\widetilde{f}$ is continuous, and discrete by condition (iv) in Lemma 13.19. To see that $\widetilde{f}$ is also an open map, consider an arbitrary open set $\widetilde{U} \subset \widetilde{S}^2$. Then $U \coloneqq \pi^{-1}(\widetilde{U}) \subset S^2$ is open and saturated. Since $\sim$ is strongly f-invariant, it follows that f maps any saturated set to a saturated set. Since f is open, it follows that $V \coloneqq f(U) \subset S^2$ is open and saturated. Now by (13.3) we have that

$$\widetilde{f}(\widetilde{U}) = \widetilde{f}(\pi(U)) = \pi(f(U)) = \pi(V),$$

which implies that $\widetilde{f}(\widetilde{U})$ is open. Thus $\widetilde{f}$ is an open map.

With a suitable choice of a fundamental class on the 2-sphere $\widetilde{S}^2$ we have $\deg(\pi) = 1$ (see Corollary 13.11 and the subsequent discussion).

Then

$$\deg(\widetilde{f}) = \deg(\widetilde{f}) \cdot \deg(\pi) = \deg(\widetilde{f} \circ \pi)$$
$$= \deg(\pi \circ f) = \deg(\pi) \cdot \deg(f) = \deg(f) > 0.$$

In particular, (i) is true and $\widetilde{f}$ has positive degree.

In order to apply Corollary A.13 and to conclude that $\widetilde{f}$ is indeed a branched covering map, it remains to show that $\widetilde{f}$ is a local homeomorphism in the complement of some finite subset of $\widetilde{S}^2$. We will do this by an argument that will also establish formula (13.4).

Let $x \in S^2$ be arbitrary and $[x]$ the corresponding equivalence class. Let $U \subset S^2$ be a neighborhood of $[x]$ as in Lemma 13.20. We set $y := f(x)$. Then $[y] = f([x])$, since $\sim$ is strongly f-invariant. Let U_s be the saturated interior of U. Then $[x] \subset U_s$ and by condition (iv) in Lemma 13.4 the set U_s is open. Moreover, since f is strongly invariant, the set $V' := f(U_s)$ is open and saturated. Let $\widetilde{U} := \pi(U_s)$ and $\widetilde{V} := \pi(V')$. Then $[x] \in \widetilde{U}$, the sets $\widetilde{U}$ and $\widetilde{V}$ are open, and $\widetilde{f}|\widetilde{U} \colon \widetilde{U} \to \widetilde{V}$ is a continuous, open, and surjective map.

Let us consider a point in $\widetilde{V}$ distinct from $[y]$. It is represented by an equivalence class $[y'] \neq [y]$, where $y' \in V' \subset f(U)$. Suppose that $x_1, \dots, x_k \in U$ are the distinct preimage points of y' under f that lie in U; so $\{x_1, \dots, x_k\} = U \cap f^{-1}(y')$. Then none of these points is a critical point of f by choice of U and so it follows from Lemma 13.20 (iv) that

$$k = d_x := 1 + \sum_{i=1}^{n} (\deg(f, c_i) - 1),$$

where $c_1, \dots, c_n$ are the critical points of f contained in $[x]$.

Consider $i \in \{1, \dots, k\}$. Since $\sim$ is strongly f-invariant, we have $f([x_i]) = [y']$. This implies that $[x_i]$ is contained in U. Indeed, otherwise the connected set $[x_i]$ must meet the boundary of U and so $U \cap [x_i]$ is not relatively compact in U. On the other hand, $U \cap [x_i]$ is contained in the subset $U \cap f^{-1}([y']) = (f|U)^{-1}([y'])$ of U which is compact, because $f|U \colon U \to f(U)$ is a proper map. This is a contradiction showing that indeed $[x_i] \subset U$. This implies that actually $[x_i] \subset U_s$ and so $[x_i] \in \widetilde{U}$. Note that

$$\widetilde{f}([x_i]) = (\widetilde{f} \circ \pi)(x_i) = (\pi \circ f)(x_i) = [y'].$$

So the points $[x_1], \dots, [x_k] \in \widetilde{U}$ are preimages of $[y']$ under $\widetilde{f}$. It is clear that $[y']$ cannot have other preimages in $\widetilde{U}$; indeed, suppose such a preimage is represented by an equivalence class $[x'] \subset U_s$ distinct from $[x_1], \dots, [x_k]$. Then $f([x']) = [y']$ by strong f-invariance of $\sim$ and so

there would be another preimage $x'' \in [x'] \subset U_s \subset U$ of y' in U distinct from the points $x_1, \ldots, x_k$.

The points $[x_1], \ldots, [x_k] \in \widetilde{U}$ are distinct; indeed, these equivalence classes are distinct from $[x]$ and so do not contain any critical points of f. So by (13.9) the map f is a homeomorphism of each of the equivalence classes $[x_1], \ldots, [x_k]$ onto $[y]$. In particular, in each of these equivalences classes the point y' has exactly one preimage which implies that the equivalence classes $[x_1], \ldots, [x_k]$ are distinct, because the points $x_1, \ldots, x_k$ are. We conclude that $[y']$ has precisely $k = d_x$ preimages under $\widetilde{f}$ that lie in $\widetilde{U}$.

Now suppose in addition that $[x]$ does not contain critical points of f. Then $d_x = 1$. Actually, then f is a homeomorphism of U onto $f(U)$ and our argument shows that each point in $\widetilde{V}$ (including $[y]$) has precisely one preimage under $\widetilde{f}$ in $\widetilde{U}$. In this case, the map $\widetilde{f}|\widetilde{U} \colon \widetilde{U} \to \widetilde{V}$ is a continuous open bijection and hence a homeomorphism. In particular, $\widetilde{f}$ is a local homeomorphism near each point in $S^2 \setminus C$, where $C := \pi(\mathrm{crit}(f))$ is a finite set. Corollary A.13 now implies that $\widetilde{f}$ is indeed a branched covering map on the 2-sphere $\widetilde{S}^2$.

Now that we know that $\widetilde{S}^2$ is a branched covering map, we return to the general case where we allow critical points of f in $[x]$. By what we have seen, each point $[y'] \neq [y] = \widetilde{f}([x])$ in $\widetilde{V}$ has precisely $k = d_x$ preimages under $\widetilde{f}$ in $\widetilde{U}$. Here $\widetilde{U}$ can be chosen to be contained in any given neighborhood of $[x]$, because the Jordan region U that led to the definition of $\widetilde{U}$ can be chosen to lie in an arbitrary neighborhood of $[x]$. In other words, each point $[y'] \neq [y]$ close to $[y]$ has precisely $k = d_x$ preimages under $\widetilde{f}$ close to $[x]$. The formula (13.4) for the local degree of $\widetilde{f}$ at $[x]$ follows.

Now (13.4) immediately implies that $\pi(\mathrm{crit}(f)) = \mathrm{crit}(\widetilde{f})$. This, in combination with the identity $\pi \circ f^n = \widetilde{f}^n \circ \pi$ for all $n \in \mathbb{N}$, gives that $\pi(\mathrm{post}(f)) = \mathrm{post}(\widetilde{f})$. $\qquad\square$

CHAPTER 14

Combinatorially expanding Thurston maps

In Chapter 12 we have constructed Thurston maps in a geometric way from two-tile subdivision rules. We want to know when the Thurston map realizing a subdivision rule can be chosen to be expanding. The key concept for an answer is the notion of combinatorial expansion.

THEOREM 14.1 (Subdivision rules and expansion). *Let $(\mathcal{D}^1, \mathcal{D}^0, L)$ be a two-tile subdivision rule on a 2-sphere S^2 that can be realized by a Thurston map $f\colon S^2 \to S^2$ with $\operatorname{post}(f) = \mathbf{V}^0$, where $\mathbf{V}^0$ is the vertex set of $\mathcal{D}^0$. Then $(\mathcal{D}^1, \mathcal{D}^0, L)$ can be realized by an expanding Thurston map if and only if $(\mathcal{D}^1, \mathcal{D}^0, L)$ is combinatorially expanding.*

Recall that combinatorial expansion (see Definition 12.18) for a two-tile subdivision rule means that every Thurston map $f\colon S^2 \to S^2$ realizing the subdivision rule is combinatorially expanding for the Jordan curve $\mathcal{C}$ of $\mathcal{D}^0$. In this case, $\mathcal{C}$ is f-invariant, $\#\operatorname{post}(f) \geq 3$, $\operatorname{post}(f) \subset \mathcal{C}$, and there exists $n_0 \in \mathbb{N}$ such that no n_0-tile for $(f, \mathcal{C})$ joins opposite sides of $\mathcal{C}$ (see Definition 12.4).

In general, one only has $\operatorname{post}(f) \subset \mathbf{V}^0$ for a Thurston map f realizing a subdivision rule as in Theorem 14.1. The stronger condition $\operatorname{post}(f) = \mathbf{V}^0$ prevents the existence of additional vertices in $\mathbf{V}^0$ that have no dynamical relevance and force an additional normalization on the Thurston map. Without the condition $\operatorname{post}(f) = \mathbf{V}^0$, Theorem 14.1 is not true in general, as we will show in Example 14.22.

If a Thurston map $f\colon S^2 \to S^2$ with $\#\operatorname{post}(f) \geq 3$ has an invariant Jordan curve $\mathcal{C}$ with $\operatorname{post}(f) \subset \mathcal{C}$, then for f to be expanding it is necessary that f is combinatorially expanding for $\mathcal{C}$ (see Lemma 8.6). The converse is not true in general: a combinatorially expanding Thurston map need not be expanding. One still obtains a converse if one allows a change of the map by a suitable isotopy.

THEOREM 14.2 (Expansion and combinatorial expansion). *Suppose $f\colon S^2 \to S^2$ is a Thurston map that has an invariant Jordan curve $\mathcal{C} \subset S^2$ with $\operatorname{post}(f) \subset \mathcal{C}$. If f is combinatorially expanding for $\mathcal{C}$, then there exists an orientation-preserving homeomorphism $\phi\colon S^2 \to$*

S^2 with $\phi(\mathcal{C}) = \mathcal{C}$ that is isotopic to the identity on S^2 rel. $\mathrm{post}(f)$ such that $g = \phi \circ f$ and $\widetilde{g} = f \circ \phi$ are expanding Thurston maps.

Clearly, g and $\widetilde{g}$ are both Thurston equivalent to f. Note that $\widetilde{g} = \phi^{-1} \circ g \circ \phi$; so g and $\widetilde{g}$ are topologically conjugate. Moreover, $\mathrm{post}(g) = \mathrm{post}(\widetilde{g}) = \mathrm{post}(f)$ as follows from Lemma 2.5, and $g(\mathcal{C}) = \widetilde{g}(\mathcal{C}) = \mathcal{C}$. So the theorem says that if a Thurston map f is combinatorially expanding for an invariant Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}$, then by "correcting" the map by post- or precomposing with a suitable homeomorphism, we can obtain an expanding Thurston map with the same invariant curve and the same set of postcritical points.

The previous two theorems are easy consequences of the following slightly more technical result.

PROPOSITION 14.3. *Let $f\colon S^2 \to S^2$ be a Thurston map that has an invariant Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}$. If f is combinatorially expanding for $\mathcal{C}$, then there exists an expanding Thurston map $\widetilde{f}\colon \widetilde{S}^2 \to \widetilde{S}^2$ that is Thurston equivalent to f and has an $\widetilde{f}$-invariant Jordan curve $\widetilde{\mathcal{C}} \subset \widetilde{S}^2$ with $\mathrm{post}(\widetilde{f}) \subset \widetilde{\mathcal{C}}$. Moreover, there exist orientation-preserving homeomorphisms $h_0, h_1\colon S^2 \to \widetilde{S}^2$ that are isotopic rel. $\mathrm{post}(f)$ and satisfy $h_0 \circ f = \widetilde{f} \circ h_1$ as well as $h_0(\mathcal{C}) = \widetilde{\mathcal{C}} = h_1(\mathcal{C})$.*

So up to Thurston equivalence every *combinatorially expanding* Thurston map with an invariant Jordan curve can be promoted to an *expanding* Thurston map with an invariant curve.

Proposition 14.3 shows that for a Thurston map with an invariant Jordan curve combinatorial expansion is *sufficient* for the existence of an equivalent map that is expanding. One may ask whether combinatorial expansion is *necessary* for this as well. The answer is negative, as we will see in Example 14.23. The Thurston map f in this example has an invariant Jordan curve $\mathcal{C}$ containing all its postcritical points. It is not combinatorially expanding for $\mathcal{C}$ (and hence not expanding), yet is equivalent to an expanding Thurston map g.

On an intuitive level the assertion of Proposition 14.3 seems quite plausible. Namely, based on Proposition 5.26 a map $\widetilde{f}$ as in Proposition 14.3 can easily be constructed if one can find cell decompositions with the same combinatorics as $\mathcal{D}^n(f, \mathcal{C})$ where the cells are small in diameter (with respect to a given background metric) when n is large. Since f is combinatorially expanding, Lemma 12.8 implies that $D_n(f, \mathcal{C}) \to \infty$ as $n \to \infty$. So if the level n increases, one needs more and more tiles to form a connected set joining opposite sides of $\mathcal{C}$, or more generally, to join any two disjoint k-cells. Therefore, it

seems evident that one should be able to make the cells small while keeping their combinatorics the same. If one wants to implement this idea, then one faces serious difficulties that make it hard to convert this into a valid proof (in [**CFP01**, Theorem 2.3] the authors claim a more general statement with an argument along these lines).

For this reason our approach to the proof of Proposition 14.3 is different. For the map $f\colon S^2 \to S^2$ in this proposition to be expanding, the intersection $\bigcap_n X^n$ of any nested sequence $\{X^n\}$ of n-tiles should consist of only one point (see Lemma 12.7). In order to enforce this condition, we introduce a suitable equivalence relation $\sim$ on the sphere S^2 that collapses these intersections $\bigcap_n X_n$ to points. We then use Moore's theorem (Theorem 13.8) to show that the quotient space $S^2/\!\!\sim$ is also a 2-sphere. The map $\widetilde{f}$ will be the induced map on $S^2/\!\!\sim$. It encodes the same combinatorial information as the original map f, because f and $\widetilde{f}$ realize isomorphic two-tile subdivision rules (see Corollary 14.16 and Remark 14.21). In particular, these maps are Thurston equivalent as can be deduced from Lemma 12.13 (we will actually give a different and more direct argument for this). While our approach is quite natural, it is somewhat lengthy to carry out and will occupy the whole chapter.

In the following, $f\colon S^2 \to S^2$ is a Thurston map and $\mathcal{C} \subset S^2$ an f-invariant Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}$ for which f is combinatorially expanding. We consider the cell decompositions $\mathcal{D}^n = \mathcal{D}^n(f,\mathcal{C})$ for $n \in \mathbb{N}_0$ as given by Definition 5.14. As before, we denote by $\mathbf{X}^n$, $\mathbf{E}^n$ and $\mathbf{V}^n$ the set of n-tiles, n-edges, and n-vertices for $(f,\mathcal{C})$, respectively. A subset $\tau \subset S^2$ is called a *tile* if it is an n-tile for some $n \in \mathbb{N}_0$. We use the terms *edge*, *vertex*, and *cell* in a similar way. In particular, in this chapter the term "cell" will always be used with this specific meaning. We will use the term *topological cell* to refer to the more general notion of cells as defined in Section 5.1.

Since $\mathcal{C}$ is f-invariant, $\mathcal{D}^{n+k}$ is a refinement of $\mathcal{D}^n$ for $n, k \in \mathbb{N}_0$. For each $X \in \mathbf{X}^{n+k}$ there exists a unique $Y \in \mathbf{X}^n$ with $X \subset Y$. Conversely, each n-tile Y is equal to the union of all $(n+k)$-tiles contained in Y, and similarly each n-edge e is equal to the union of all $(n+k)$-edges contained in e (all this was proved in Proposition 12.5). We will use this fact that cells are subdivided by cells of the same dimension and higher level repeatedly in the following.

The equivalence relation on S^2**.** As in (12.3), we denote by $\mathcal{S} = \mathcal{S}(f,\mathcal{C})$ the set of all sequences $\{X^n\}$ with $X^n \in \mathbf{X}^n$ for $n \in \mathbb{N}_0$ and

$$X^0 \supset X^1 \supset X^2 \supset \dots$$

We know (see Lemma 12.7) that expansion of a Thurston map with an invariant curve is characterized by the condition that $\bigcap_n X^n$ is always a singleton set if $\{X^n\} \in \mathcal{S}$. This may not be the case for our given map f, and so we want to identify all points in such an intersection $\bigcap_n X^n$. This will not lead to an equivalence relation, since transitivity may fail. As we will see, this issue is resolved if we define the relation as follows.

DEFINITION 14.4. Let $x, y \in S^2$ be arbitrary. We write $x \sim y$ if and only if for all $\{X^n\}, \{Y^n\} \in \mathcal{S}$ with $x \in \bigcap_n X^n$ and $y \in \bigcap_n Y^n$ we have $X^n \cap Y^n \neq \emptyset$ for all $n \in \mathbb{N}_0$.

Recall from (5.14) that $D_n = D_n(f, \mathcal{C})$ denotes the minimal number of n-tiles forming a connected set K^n joining opposite sides of $\mathcal{C}$. Since f is combinatorially expanding for $\mathcal{C}$ (see Definition 12.4), we have $\#\operatorname{post}(f) \geq 3$ and so the term "joining opposite sides" is meaningful (see Definition 5.32). Moreover, there exists $n_0 \in \mathbb{N}$ such that $D_{n_0}(f, \mathcal{C}) \geq 2$, and so by Lemma 12.8 we have $D_n = D_n(f, \mathcal{C}) \to \infty$ as $n \to \infty$. In combination with Lemma 5.36 this implies that if τ, σ are disjoint k-cells and K^n is a connected set of n-tiles with $\sigma \cap K^n \neq \emptyset$ and $\tau \cap K^n \neq \emptyset$, then the number of tiles in K^n tends to infinity and thus cannot stay bounded as $n \to \infty$. We will use this fact in the proof of the following lemma.

LEMMA 14.5. *The relation $\sim$ is an equivalence relation on S^2.*

PROOF. Reflexivity and symmetry of the relation $\sim$ are clear. To show transitivity, let $x, y, z \in S^2$ be arbitrary and assume that $x \sim y$ and $y \sim z$. Let $\{X^n\}, \{Z^n\} \in \mathcal{S}$ with $x \in \bigcap_n X^n$ and $z \in \bigcap_n Z^n$ be arbitrary. We have to show that $X^n \cap Z^n \neq \emptyset$ for all $n \in \mathbb{N}_0$.

If this is not the case, then there exists $n_0 \in \mathbb{N}_0$ such that $X^{n_0} \cap Z^{n_0} = \emptyset$. To reach a contradiction, pick a sequence $\{Y^n\} \in \mathcal{S}$ with $y \in \bigcap_n Y^n$. Since $x \sim y$ and $y \sim z$, we have $X^n \cap Y^n \neq \emptyset$ and $Y^n \cap Z^n \neq \emptyset$ for all $n \in \mathbb{N}_0$. Then $X^{n_0} \cap Y^n \supset X^n \cap Y^n \neq \emptyset$ and $Z^{n_0} \cap Y^n \supset Z^n \cap Y^n \neq \emptyset$ for all $n \geq n_0$. So the n-tile Y^n connects the disjoint n_0-tiles X^{n_0} and Y^{n_0} for all $n \geq n_0$. As we discussed, this is impossible by Lemma 5.36. $\square$

The following lemma gives convenient characterizations when two points are equivalent.

LEMMA 14.6. *Let $x, y \in S^2$ be arbitrary. Then the following conditions are equivalent:*

(i) $x \sim y$,

(ii) *there exist $\{X^n\}, \{Y^n\} \in \mathcal{S}$ with $x \in \bigcap_n X^n$, $y \in \bigcap_n Y^n$, and*
    $X^n \cap Y^n \neq \emptyset$ *for all $n \in \mathbb{N}_0$,*

(iii) *for all cells $\sigma, \tau \subset S^2$ with $x \in \sigma$, $y \in \tau$, we have $\sigma \cap \tau \neq \emptyset$.*

PROOF. The implication (i) $\Rightarrow$ (ii) is clear.

To show the reverse implication (ii) $\Rightarrow$ (i), we assume that there exist $\{X^n\}, \{Y^n\} \in \mathcal{S}$ with $x \in \bigcap_n X^n$, $y \in \bigcap_n Y^n$, and $X^n \cap Y^n \neq \emptyset$ for all $n \in \mathbb{N}_0$. We claim that if $\{U^n\}, \{V^n\} \in \mathcal{S}$ are two other sequences with $x \in \bigcap_n U^n$ and $y \in \bigcap_n V^n$, then $U^n \cap V^n \neq \emptyset$ for all $n \in \mathbb{N}_0$. To reach a contradiction, assume that $U^{n_0} \cap V^{n_0} = \emptyset$ for some $n_0 \in \mathbb{N}_0$. We then have

$$U^{n_0} \cap X^n \supset \{x\} \neq \emptyset \quad \text{and} \quad V^{n_0} \cap Y^n \supset \{y\} \neq \emptyset$$

for all $n \in \mathbb{N}$. Moreover, $X^n \cap Y^n \neq \emptyset$, and so for each $n \in \mathbb{N}_0$, the set $K^n := X^n \cup Y^n$ is connected, consists of two n-tiles, and meets the disjoint n_0-tiles U^{n_0} and V^{n_0}. As before, this contradicts Lemma 5.36. Hence $x \sim y$ as desired.

The implication (iii) $\Rightarrow$ (i) is again clear. To prove (i) $\Rightarrow$ (iii), suppose that $x \sim y$. We argue by contradiction and assume that there exist cells σ, τ with $x \in \sigma$, $y \in \tau$, and $\sigma \cap \tau = \emptyset$. By subdividing the cells if necessary, we may assume that σ and τ are cells on the same level n_0.

There are sequences $\{X^n\}, \{Y^n\} \in \mathcal{S}$ with $x \in \bigcap_n X^n$ and $y \in \bigcap_n Y^n$. Since $x \sim y$, we have $X^n \cap Y^n \neq \emptyset$ for all n.

This implies that for $n \in \mathbb{N}_0$ the set $K^n = X^n \cup Y^n$ is connected and consists of at most two n-tiles. Moreover,

$$K^n \cap \sigma \supset X^n \cap \sigma \supset \{x\} \neq \emptyset,$$

and similarly, $K^n \cap \tau \neq \emptyset$. Hence K^n connects the disjoint n_0-cells σ and τ. Since f is combinatorially expanding, this is impossible by Lemma 5.36 for large n. This gives the desired contradiction. $\square$

The previous lemma implies that all points in an intersection $\bigcap_n X^n$ with $\{X^n\} \in \mathcal{S}$ are equivalent. It is clear that $\sim$ is the "smallest" equivalence relation with this property.

If $x \in S^2$ we denote by $[x] \subset S^2$ the equivalence class of x with respect to the equivalence relation $\sim$, and by

$$\widetilde{S}^2 = S^2/\!\sim \, = \{[x] : x \in S^2\}$$

the quotient space of S^2 under $\sim$. So $\widetilde{S}^2$ consists of all equivalence classes of $\sim$. Such an equivalence class is both a point in $\widetilde{S}^2$ and a subset of S^2. We equip $\widetilde{S}^2$ with the quotient topology. Then the quotient map $\pi \colon S^2 \to \widetilde{S}^2$, $x \in S^2 \mapsto [x]$, is continuous.

In order to prove that $\widetilde{S}^2$ is in fact a topological 2-sphere, we want to show that $\sim$ is of Moore-type (see Definition 13.7) and apply Theorem 13.8 (Moore's theorem). To verify the relevant conditions for $\sim$, we need a good geometric description of the equivalence classes. To set this up, consider a point $x \in S^2$ and let $n \in \mathbb{N}_0$ be arbitrary. We define

$$(14.1) \qquad \Omega^n = \Omega^n(x) = \bigcup_{x \in c^n} \text{int}(c^n)$$

where the union is taken over all n-cells c^n that contain x. Recall that $\text{int}(c^n) = c^n$ if c^n has dimension 0.

Note that

$$(14.2) \qquad \overline{\Omega}^n = \bigcup_{x \in X^n} X^n,$$

where the union is taken over all n-tiles X^n that contain x. Indeed, every cell c^n in the union in (14.1) is contained in an n-tile X^n that contains x. Therefore, $\overline{\Omega}^n \subset \bigcup_{x \in X^n} X^n$, since the right hand side is closed. On the other hand, for each n-tile X^n containing x we have $\text{int}(X^n) \subset \Omega^n$. Thus $\bigcup_{x \in X^n} X^n \subset \overline{\Omega}^n$, and (14.2) follows.

LEMMA 14.7. *The set $\Omega^n \subset S^2$ is a simply connected region.*

PROOF. We have to consider three cases. When $x = v$ is an n-vertex then $\Omega^n(v) = W^n(v)$ is the n-flower of v by Definition 5.27. Recall from Lemma 5.28 (i) that such a vertex flower is a simply connected region.

Suppose that x is not an n-vertex, but x is contained in an n-edge e^n. Then x is necessarily contained in the interior $\text{int}(e^n)$ of e^n. There are precisely two distinct n-tiles X^n and Y^n that contain e^n in their boundaries. These are all the n-tiles that contain x. Thus

$$\Omega^n = \text{int}(X^n) \cup \text{int}(e^n) \cup \text{int}(Y^n).$$

Then Ω^n is a simply connected region (see Lemma 5.9 (iv)).

Finally, suppose that x is not contained in any n-edge. Then there is a unique n-tile X^n that contains x. Then $\Omega^n = \text{int}(X^n)$ is an open Jordan region, and so simply connected. $\qquad\square$

Let $M \subset S^2$ be an equivalence class with respect to $\sim$. We select a point $x \in M$ as follows.

Case 1: If M contains a vertex v, then $x := v$.

Case 2: If M contains no vertex, but intersects an edge e, then we choose a point in $M \cap e$ for x.

Case 3: If M contains no vertex and does not intersect any edge, then we choose an arbitrary point in M for x.
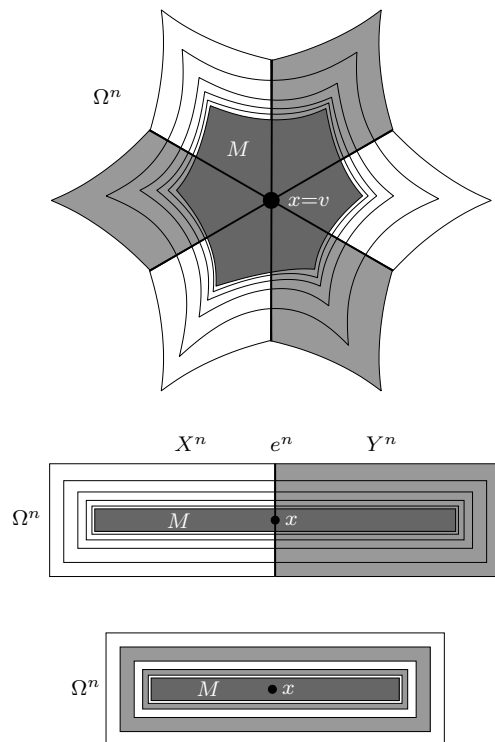
FIGURE 14.1. Equivalence classes of vertex-, edge-, and tile-type.

We say that M is of *vertex-type* in the first, of *edge-type* in the second, and of *tile-type* in the last case. We call the point x a *center* of the equivalence class M. By Lemma 14.6 an equivalence class cannot contain two distinct vertices. So if M is of vertex-type, then its center is unique, but this may not be true in the other two cases.

With such a choice of a center x for given M, we define $\Omega^n = \Omega^n(x)$ as in (14.1) for $n \in \mathbb{N}_0$. Note that every $(n+1)$-cell c^{n+1} with $x \in c^{n+1}$ is contained in an n-cell c^n with $x \in c^n$ and $\mathrm{int}(c^{n+1}) \subset \mathrm{int}(c^n)$ (see Lemma 5.7). Thus $\{\Omega^n\}$ is a decreasing sequence of sets, i.e.,

$$(14.3) \qquad \Omega^0 \supset \Omega^1 \supset \Omega^2 \supset \dots$$

The different types of equivalence classes are illustrated in Figure 14.1.

LEMMA 14.8. *Let $M \subset S^2$ be an arbitrary equivalence class with respect to $\sim$, and Ω^n be defined as above for $n \in \mathbb{N}_0$. Then*

$$(14.4) \qquad M = \bigcap_n \Omega^n = \bigcap_n \overline{\Omega}^n$$

PROOF. Let $x \in M$ be a center of M. In order to establish the inclusion

$$\text{(14.5)} \qquad \bigcap_n \overline{\Omega}^n \subset M,$$

let $y \in \bigcap_n \overline{\Omega}^n$ be arbitrary. We have to show that $x \sim y$ (which implies $y \in M$). If this is not the case, then $x \nsim y$, and so there exist sequences $\{X^n\}, \{Y^n\} \in \mathcal{S}$ with $x \in \bigcap_n X^n$ and $y \in \bigcap_n Y^n$, and $n_0 \in \mathbb{N}_0$ such that $X^{n_0} \cap Y^{n_0} = \emptyset$. On the other hand, for each $n \in \mathbb{N}_0$ we have $y \in \overline{\Omega}^n$, and so by (14.2) we can find an n-tile Z^n with $x, y \in Z^n$. Then $X^{n_0} \cap Z^n \supset \{x\} \neq \emptyset$ and $Y^{n_0} \cap Z^n \supset \{y\} \neq \emptyset$, i.e., the tile Z^n intersects the disjoint tiles X^{n_0} and Y^{n_0}, for all $n \in \mathbb{N}_0$. This is impossible, since f is combinatorially expanding for $\mathcal{C}$ (see Lemma 5.36). We obtain a contradiction and (14.5) follows.

To finish the proof, it is enough to show that $M \subset \bigcap_n \Omega^n$, or equivalently, that if

$$y \in S^2 \setminus \bigcap_n \Omega^n = \bigcup_n (S^2 \setminus \Omega^n)$$

is arbitrary, then $y \nsim x$ (and so $y \notin M$). Note that the sets $S^2 \setminus \Omega^n$ for $n \in \mathbb{N}_0$ form an increasing sequence, and so $y \notin \Omega^n$ for all sufficiently large n. In order to show $y \nsim x$, we now consider three cases according to the type of M.

Case 1: M is of vertex-type. In this case, x is a vertex, say an n_0-vertex, where $n_0 \in \mathbb{N}_0$. Then x is also an n-vertex for all $n \geq n_0$. Fix $n \geq n_0$ such that $y \notin \Omega^n$. Then there exists a unique n-cell τ with $y \in \text{int}(\tau)$. Since $y \notin \Omega^n$, we have $x \notin \tau$. Now x is an n-vertex and so $\{x\}$ is an n-cell, and the n-cells $\tau \supset \{y\}$ and $\{x\}$ are disjoint. By Lemma 14.6 this implies $y \nsim x$ as desired.

Case 2: M is of edge-type. Then x is contained in an edge, say an n_0-edge e^{n_0}, where $n_0 \in \mathbb{N}_0$. By successive subdivisions (see Proposition 12.5 (iv)) we can find n-edges e^n for $n \geq n_0$ that contain x and that satisfy

$$e^{n_0} \supset e^{n_0+1} \supset \dots.$$

Fix $n \geq n_0$ such that $y \notin \Omega^n$. Then there exists a unique n-cell τ with $y \in \text{int}(\tau)$. Since $y \notin \Omega^n$, we have $x \notin \tau$ and so $e^n \not\subset \tau$. Hence $\tau \cap \text{int}(e^n) = \emptyset$ by Lemma 5.3 (ii). Let u and v be the endpoints of e^n. Since these points are vertices, they do not belong to M by assumption. So the set

$$\bigcap_{n \geq n_0} e^n \subset \bigcap_n \overline{\Omega}^n \subset M$$

does not contain u or v either. It follows that there exists $m \geq n$ such that $u, v \notin e^m$. Then $y \in \tau$, $x \in e^m$, and $\tau \cap e^m = \emptyset$, because

$$\tau \cap e^m \subset \tau \cap (e^n \setminus \{u, v\}) = \tau \cap \operatorname{int}(e^n) = \emptyset.$$

By Lemma 14.6 this implies $y \not\sim x$ as desired.

Case 3: M is of tile-type. We pick sequences $\{X^n\}, \{Y^n\} \in \mathcal{S}$ with $x \in \bigcap_n X^n$ and $y \in \bigcap_n Y^n$. Since $x \in M$ and M does not meet any edges, for each $n \in \mathbb{N}_0$ the tile X^n is the unique n-tile with $x \in X^n$. Then $x \in \operatorname{int}(X^n)$ and $\Omega^n = \operatorname{int}(X^n)$. In order to show that $y \not\sim x$, we argue by contradiction and assume $y \sim x$. Then $K^n := X^n \cap Y^n \neq \emptyset$ for each $n \in \mathbb{N}_0$ (see Definition 14.4). The sets K^n, $n \in \mathbb{N}_0$, are non-empty nested compact sets. Hence there exists a point $z \in \bigcap_n K^n$. Then $x \sim z$ and so $z \in M$.

On the other hand, if $n_0 \in \mathbb{N}_0$ is large enough, then $y \notin \Omega^{n_0} = \operatorname{int}(X^{n_0})$. Since $\partial X^{n_0} = X^{n_0} \setminus \operatorname{int}(X^{n_0})$ consists of n_0-edges, $y \sim x$ and so $y \in M$, and M does not meet edges, we then actually have $y \notin X^{n_0}$. Thus $X^{n_0} \neq Y^{n_0}$. This means that the intersection $K^{n_0} = X^{n_0} \cap Y^{n_0}$ consists of n_0-cells on the boundary of X^{n_0} and of Y^{n_0}, and is hence contained in a union of n_0-edges. Since $z \in M \cap K_{n_0}$, this implies that M meets an edge, contradicting our assumption in this case. So indeed $y \not\sim x$ as desired. $\qquad\square$

The following consequence of the previous lemma will be one of the essential ingredients in the proof that $\widetilde{S}^2$ is a topological 2-sphere.

COROLLARY 14.9. *Each equivalence class M of $\sim$ is a compact connected set with connected complement $S^2 \setminus M$.*

PROOF. Let M be an arbitrary equivalence class of $\sim$. Then Lemma 14.8 and (14.3) imply that the set M is the intersection of the nested sequence of the compact sets $\overline{\Omega}^n$, $n \in \mathbb{N}_0$. It follows from (14.2) that each set $\overline{\Omega}^n$ is connected. Hence M is also compact and connected.

The complement of an open simply connected set in S^2 is connected. So Lemma 14.8 also shows that the complement $S^2 \setminus M$ of M is the union of the increasing sequence of the connected sets $S^2 \setminus \Omega^n$. Hence $S^2 \setminus M$ is connected. $\qquad\square$

The quotient space $\widetilde{S}^2$ is a topological 2-sphere. After these preparations we are ready to show that $\widetilde{S}^2$ is a topological 2-sphere.

LEMMA 14.10. *Let $\sim$ be the equivalence relation on S^2 as in Definition 14.4. Then $\sim$ is of Moore-type and the quotient space $\widetilde{S}^2 = S^2/\sim$ is homeomorphic to S^2.*

PROOF. By Lemma 14.5 our relation $\sim$ is indeed an equivalence relation. It remains to verify the conditions (i)–(iv) in Definition 13.7. Then $\widetilde{S}^2$ is a 2-sphere by Theorem 13.8 (Moore's theorem).

Conditions (ii) and (iii) were already proved in Corollary 14.9.

Condition (iv): There are at least two equivalence classes, because no two distinct vertices are equivalent by Lemma 14.6, and each post-critical point of f is a vertex (there are at least three such points).

Condition (i): Let $\{x_n\}$ and $\{y_n\}$ be convergent sequences in S^2 with $x_n \to x$ and $y_n \to y$ as $n \to \infty$, and suppose that $x_n \sim y_n$ for all $n \in \mathbb{N}$. We have to show that $x \sim y$. Suppose this is not the case. Then the equivalence classes $[x]$ and $[y]$ are disjoint. By Lemma 14.7 and Lemma 14.8 there exist simply connected nested regions Ω_x^n and Ω_y^n for $n \in \mathbb{N}_0$ such that

$$[x] = \bigcap_n \Omega_x^n = \bigcap_n \overline{\Omega}_x^n \text{ and } [y] = \bigcap_n \Omega_y^n = \bigcap_n \overline{\Omega}_y^n.$$

Since $[x]$ and $[y]$ are disjoint, the sets $\overline{\Omega}_x^n$ and $\overline{\Omega}_y^n$ will also be disjoint for sufficiently large n, say $\overline{\Omega}_x^{n_0} \cap \overline{\Omega}_y^{n_0} = \emptyset$. On the other hand, since $\Omega_x^{n_0} \supset [x]$ and $\Omega_y^{n_0} \supset [y]$ are open, there exists $n_1 \in \mathbb{N}$ such that $x_{n_1} \in \Omega_x^{n_0}$ and $y_{n_1} \in \Omega_y^{n_0}$. Since $\overline{\Omega}_x^{n_0}$ and $\overline{\Omega}_y^{n_0}$ consist of n_0-tiles and are disjoint, this means that there exist n_0-tiles σ and τ with $x_{n_1} \in \sigma$, $y_{n_1} \in \tau$ and $\sigma \cap \tau = \emptyset$. Hence $x_{n_1} \not\sim y_{n_1}$ by Lemma 14.6. This is a contradiction. It follows that $\sim$ is closed. $\square$

Quotients of cells and the induced cell decompositions on $\widetilde{S}^2$. We now study what happens to our cells under the quotient map $\pi \colon S^2 \to \widetilde{S}^2$. If $A \subset S^2$ is an arbitrary set, we denote by $\widetilde{A}$ its image under π. So $\widetilde{A} = \pi(A) = \{[x] : x \in A\} \subset \widetilde{S}^2$. We will see that if σ is an arbitrary cell (i.e., an element of $\mathcal{D}^n$ for some $n \in \mathbb{N}_0$), then $\widetilde{\sigma}$ is a topological cell of the same dimension (Lemma 14.14). Moreover, the images $\widetilde{\sigma}$ of the n-cells $\sigma \in \mathcal{D}^n$ form a cell decomposition of $\widetilde{S}^2$ (Lemma 14.15).

LEMMA 14.11. *Let M be an arbitrary equivalence class with center $x \in M$. If τ is an arbitrary cell, then*

$$\tau \cap M \neq \emptyset \quad \text{if and only if} \quad x \in \tau.$$

PROOF. The "if"-implication is obvious.

To show the other implication, assume that τ is a cell of level n and $x \notin \tau$. Consider an n-cell $\sigma \subset \tau$. Then $x \notin \sigma$ and so $\text{int}(\sigma)$ is disjoint from $\Omega^n = \Omega^n(x)$ by (14.1), because distinct n-cells have disjoint interiors. Recall from Lemma 5.2 that τ is the disjoint union

of the interiors of all n-cells $\sigma \subset \tau$. Thus $\tau \cap \Omega^n = \emptyset$, and so $\tau \cap M = \emptyset$ by Lemma 14.8. $\qquad\square$

The following lemma states that if we pass to the quotient space $\widetilde{S}^2 = S^2/\!\sim$, then intersection and inclusion relations of cells are preserved. In particular, we do not create "new" intersections or inclusions between cells.

LEMMA 14.12. *If σ and τ are cells, then $\widetilde{\sigma} \cap \widetilde{\tau} = \widetilde{\sigma \cap \tau}$. Moreover, we have $\widetilde{\sigma} \subset \widetilde{\tau}$ if and only if $\sigma \subset \tau$.*

PROOF. The inclusion $\widetilde{\sigma \cap \tau} \subset \widetilde{\sigma} \cap \widetilde{\tau}$ is trivial.

For the other inclusion consider an arbitrary point $[x] \in \widetilde{\sigma} \cap \widetilde{\tau} \subset \widetilde{S}^2$. We can assume that x is a center of $M = [x] \subset S^2$. Then M meets both cells σ and τ, and so $x \in \sigma \cap \tau$ by Lemma 14.11. Thus $[x] \in \widetilde{\sigma \cap \tau}$. We have proved $\widetilde{\sigma} \cap \widetilde{\tau} \subset \widetilde{\sigma \cap \tau}$ as desired.

In the second statement the implication $\sigma \subset \tau \Rightarrow \widetilde{\sigma} \subset \widetilde{\tau}$ is trivial. For the other implication assume that $\widetilde{\sigma} \subset \widetilde{\tau}$. Let k and n be the levels of σ and τ, respectively. For the moment, we make the additional assumption that $k \geq n$.

By Lemma 12.10 there exists a vertex v such that $v \in \operatorname{int}(\sigma)$ (note that this is trivial if σ is a 0-dimensional cell). Then $[v] \in \widetilde{\sigma} \subset \widetilde{\tau}$. This means that there exists a point $x \in \tau$ such that $[v] = [x]$, or $v \sim x$. Condition (iii) in Lemma 14.6 implies that $\{v\} \cap \tau \neq \emptyset$ or $v \in \tau$. Thus

$$(14.6) \qquad\qquad \operatorname{int}(\sigma) \cap \tau \neq \emptyset.$$

Since $k \geq n$, the cell decomposition $\mathcal{D}^k$ containing σ is a refinement of the cell decomposition $\mathcal{D}^n$ containing τ. Therefore, as we have seen in the first part of the proof of Lemma 5.7, the relation (14.6) forces the inclusion $\sigma \subset \tau$.

If $k < n$, we subdivide σ into cells of level n. By the previous argument, τ will contain each of these cells, and so we always have $\sigma \subset \tau$ as desired. $\qquad\square$

LEMMA 14.13. *Let $M \subset S^2$ be an equivalence class and $E \subset S^2$ be a finite union of edges. Then $E \cap M$ is connected.*

PROOF. Let x be a center of M. By subdividing the edges in E, we can assume that E consists of n_0-edges, where $n_0 \in \mathbb{N}_0$ is large enough. For $n \geq n_0$ let

$$E^n := \bigcup \{e \in \mathbf{E}^n : e \subset E, x \in e\}.$$

Clearly, each set E^n is compact and connected. We claim that these sets form a decreasing sequence, i.e., $E^{n+1} \subset E^n$ for $n \geq n_0$. To see

this, suppose e is one of the $(n + 1)$-edges forming the union E^{n+1}, where $n \geq n_0$. Then $x \in e \subset E$. In particular, e is contained in the union of the n_0-edges forming E. By subdividing these edges into n-edges, we see that e is covered by n-edges contained in E. This implies that e is contained in one of these n-edges e' (this again follows from considerations as in the first part of the proof of Lemma 5.7). Hence $x \in e \subset e' \subset E$, and so $e \subset e' \subset E^n$. Since the $(n + 1)$-edge $e \subset E^{n+1}$ was arbitrary, we conclude $E^{n+1} \subset E^n$ as desired.

The set $C := \bigcap_{n \geq n_0} E^n \subset E$ is an intersection of a decreasing sequence of compact and connected set, and so it is also compact and connected.

We claim that $C = E \cap M$. Indeed, $E^n \subset \overline{\Omega}^n$, where $\Omega^n = \Omega^n(x)$ and $n \geq n_0$, as follows from (14.2). Hence $C \subset E \cap \bigcap_{n \geq n_0} \overline{\Omega}^n = E \cap M$ by Lemma 14.8.

For the other inclusion, let $y \in E \cap M$ be arbitrary. Then for each $n \geq n_0$ the point y is contained in an n-edge $e \subset E$. Since $y \sim x$, we have $e \cap M \neq \emptyset$, and so $x \in e$ by Lemma 14.11. Hence $e \subset E^n$ and so $y \in e \subset E^n$. It follows that $y \in \bigcap_{n \geq n_0} E^n = C$. This shows the other inclusion $E \cap M \subset C$. We conclude that $E \cap M = C$ is connected. $\square$

For the proof of the next lemma we need the 1-dimensional version of Moore's Theorem as provided by Proposition 13.12.

LEMMA 14.14. *Let τ be an edge or a tile. Then $\widetilde{\tau}$ is an arc or a closed Jordan region, respectively. Moreover, $\partial \widetilde{\tau} = \widetilde{\partial \tau}$.*

Here $\partial \tau$ (and similarly $\partial \widetilde{\tau}$) refers as usual to the boundary of a cell τ as defined in Section 5.1. So $\partial \tau$ is the topological boundary of τ in S^2 if τ is a tile, and equal to the set consisting of the two endpoints of τ if τ is an edge. If τ is a 0-dimensional cell, i.e., a singleton set consisting of a vertex, then $\partial \tau = \emptyset$, and the statement in the lemma is trivially also true. So the lemma can be formulated in an equivalent form by saying that if $\tau \subset S^2$ is a cell (in one of the cell decompositions $\mathcal{D}$), then $\widetilde{\tau} \subset \widetilde{S^2}$ is a cell (in the general topological sense) of the same dimension, and the boundary of $\widetilde{\tau}$ is the image of the boundary of τ under the quotient map.

PROOF. Suppose first that τ is an edge. Then our equivalence relation $\sim$ on S^2 restricts to an equivalence relation on τ whose quotient space can be identified with the subset $\widetilde{\tau}$ of $\widetilde{S^2}$. The equivalence classes on τ have the form $\tau \cap M$, where $M \subset S^2$ is an equivalence class with respect to $\sim$.

Each set $\tau \cap M$ is compact, as $\sim$ is closed, and connected by Lemma 14.13. Moreover, τ meets at least two distinct equivalence classes,

as its endpoints are distinct vertices and hence not equivalent. Proposition 13.12 implies that $\widetilde{\tau} \subset \widetilde{S}^2$ is indeed an arc.

Let u and v be the two endpoints of τ. Then $[u] \cap \tau$ is a compact connected subset of τ containing u. Hence this set is a subarc of τ with one endpoint equal to u. This implies that the set $\tau \setminus [u]$ is connected, and so the set $\pi(\tau \setminus [u]) = \widetilde{\tau} \setminus \{\pi(u)\}$ is also connected. Therefore, $\pi(u)$ is an endpoint of $\widetilde{\tau}$. By the same reasoning we see that $\pi(v)$ is also an endpoint of $\widetilde{\tau}$. Since u and v are distinct vertices, we have $u \not\sim v$ and so $\pi(u) \neq \pi(v)$. Hence $\partial\widetilde{\tau} = \{\pi(u), \pi(v)\} = \pi(\{u, v\}) = \widetilde{\partial\tau}$.

If τ is a tile, say an n-tile, then τ is a closed Jordan region whose boundary $J = \partial\tau$ is a topological circle consisting of finitely many edges. By the Schönflies theorem we can write S^2 as a disjoint union $S^2 = U_1 \cup J \cup U_2$, where U_1 and U_2 are open Jordan regions bounded by J. Then τ coincides with one of the sets $\overline{U}_1$ or $\overline{U}_2$, say $\tau = \overline{U}_1$.

The set $\widetilde{J} \subset \widetilde{S}^2$ is also a topological circle as follows from the fact that $\sim$ is closed, Lemma 14.13, and Proposition 13.12. So we can also write $\widetilde{S}^2$ as a disjoint union $\widetilde{S}^2 = D_1 \cup \widetilde{J} \cup D_2$, where D_1 and D_2 are open Jordan regions in $\widetilde{S}^2$ bounded by $\widetilde{J}$. If we take preimages under the quotient map $\pi \colon S^2 \to \widetilde{S}^2$, we get the disjoint union $S^2 = \pi^{-1}(D_1) \cup \pi^{-1}(\widetilde{J}) \cup \pi^{-1}(D_2)$. Recall from Lemma 14.10 that $\sim$ is monotone, meaning that each equivalence class is connected. Therefore, preimages of connected sets under π are connected (see Lemma 13.6). So the sets $\pi^{-1}(D_1)$ and $\pi^{-1}(D_2)$ are connected open sets disjoint from $\pi^{-1}(\widetilde{J}) \supset J$. It follows that each of the sets $\pi^{-1}(D_1)$ and $\pi^{-1}(D_2)$ is contained in one of the regions U_1 and U_2.

These sets cannot be contained in the same region U_i. Indeed, if $\pi^{-1}(D_1) \cup \pi^{-1}(D_2) \subset U_1$, for example, then $U_2 \subset \pi^{-1}(\widetilde{J})$, and so $\pi(U_2) \subset \widetilde{J}$. This means that every point in U_2 is equivalent to a point in J. This is impossible, because U_2 contains the interior of an n-tile, and hence a k-vertex for some $k > n$ (see Lemma 12.10 (ii)). Such a vertex is not equivalent to any point in J by condition (iii) in Lemma 14.6.

By what we have just seen, we may assume that indices are chosen such that $\pi^{-1}(D_1) \subset U_1$ and $\pi^{-1}(D_2) \subset U_2$. Since π is surjective, it follows that

$$\overline{D}_1 = D_1 \cup \widetilde{J} = \pi(\pi^{-1}(D_1)) \cup \pi(J) \subset \pi(U_1 \cup J) = \pi(\overline{U}_1).$$

On the other hand, $\overline{U}_1 \cap \pi^{-1}(D_2) \subset \overline{U}_1 \cap U_2 = \emptyset$, and so no point in $\overline{U}_1$ is sent to $D_2 = \widetilde{S}^2 \setminus \overline{D}_1$ by π. Hence $\pi(\overline{U}_1) \subset \overline{D}_1$. It follows that $\overline{D}_1 = \pi(\overline{U}_1) = \widetilde{\tau}$, and so $\widetilde{\tau}$ is indeed a closed Jordan region. Moreover, $\partial\widetilde{\tau} = \partial D_1 = \widetilde{J} = \widetilde{\partial\tau}$. $\square$

We now come to the main result of this subsection, which says that the quotient map π maps the cell decompositions $\mathcal{D}^n$ to cell decompositions $\widetilde{\mathcal{D}}^n$ such that "all combinatorial properties are preserved".

LEMMA 14.15. *Let $n, k \in \mathbb{N}_0$. Then the following statements are true:*

(i) *For each $\tau \in \mathcal{D}^n$ the set $\widetilde{\tau}$ is a topological cell in $\widetilde{S}^2$ of the same dimension as τ.*

(ii) *For each cell $\tau \in \mathcal{D}^n$ we have $\widetilde{\partial \tau} = \partial \widetilde{\tau}$.*

(iii) *For $\sigma, \tau \in \mathcal{D}^n$, we have $\widetilde{\sigma} = \widetilde{\tau}$ if and only if $\sigma = \tau$.*

(iv) *$\widetilde{\mathcal{D}}^n := \{\widetilde{\tau} : \tau \in \mathcal{D}^n\}$ is a cell decomposition of $\widetilde{S}^2$.*

(v) *The map $\tau \in \mathcal{D}^n \mapsto \widetilde{\tau} \in \widetilde{\mathcal{D}}^n$ is an isomorphism between the cell complexes $\mathcal{D}^n$ and $\widetilde{\mathcal{D}}^n$.*

(vi) *$\widetilde{\mathcal{D}}^{n+k}$ is a refinement of $\widetilde{\mathcal{D}}^n$. Moreover, for all $\sigma \in \mathcal{D}^{n+k}$ and $\tau \in \mathcal{D}^n$ we have*

$$\sigma \subset \tau \quad \text{if and only if} \quad \widetilde{\sigma} \subset \widetilde{\tau}.$$

PROOF. (i) and (ii) follow from Lemma 14.14.

(iii) Let σ and τ be arbitrary n-cells, and suppose that $\operatorname{int}(\widetilde{\sigma}) \cap \operatorname{int}(\widetilde{\tau}) \neq \emptyset$. Pick a point $p \in \operatorname{int}(\widetilde{\sigma}) \cap \operatorname{int}(\widetilde{\tau})$. Then $p \in \widetilde{\sigma} \cap \widetilde{\tau} = \widetilde{\sigma \cap \tau}$ (see Lemma 14.12), and so there exists $x \in \sigma \cap \tau$ with $\pi(x) = p$. Then $x \in \operatorname{int}(\sigma)$, for otherwise $x \in \partial \sigma$ and so $p = \pi(x) \in \partial \widetilde{\sigma}$ by (ii), contradicting the choice of p. Similarly, $x \in \operatorname{int}(\tau)$. So $x \in \operatorname{int}(\sigma) \cap \operatorname{int}(\tau)$ which implies that $\sigma = \tau$. Statement (iii) follows.

(iv) From what we have seen, it follows that the topological cells $\widetilde{\tau}$ for $\tau \in \mathcal{D}^n$ are all distinct, and no two have a common interior point. Moreover, there are finitely many of these cells, and they cover $\widetilde{S}^2$, because the cells in $\mathcal{D}^n$ cover S^2. Finally, for a cell $\widetilde{\tau}$ we have $\partial \widetilde{\tau} = \widetilde{\partial \tau}$ by (ii). Since $\partial \tau$ is a union of cells in $\mathcal{D}^n$, the set $\partial \widetilde{\tau}$ is a union of cells in $\widetilde{\mathcal{D}}^n$. This shows that $\widetilde{\mathcal{D}}^n$ is a cell decomposition of $\widetilde{S}^2$.

(v) By (i) and (iii) the map $\tau \in \mathcal{D}^n \mapsto \widetilde{\tau} \in \widetilde{\mathcal{D}}^n$ is a bijection between $\mathcal{D}^n$ and $\widetilde{\mathcal{D}}^n$ that preserves dimensions of cells. By Lemma 14.12 the map also satisfies condition (ii) in Definition 5.10. Hence it is an isomorphism between the cell complexes $\mathcal{D}^n$ and $\widetilde{\mathcal{D}}^n$.

(vi) It follows immediately from the definitions and the fact that $\mathcal{D}^{n+k}$ is a refinement of $\mathcal{D}^n$ that $\widetilde{\mathcal{D}}^{n+k}$ is a refinement of $\widetilde{\mathcal{D}}^n$. The second statement was proved in Lemma 14.12. $\qquad\square$

Let $L\colon \mathcal{D}^1 \to \mathcal{D}^0$ be the labeling induced by f. It is given by $L(\tau) = f(\tau) \in \mathcal{D}^0$ for $\tau \in \mathcal{D}^1$ (see Section 5.4). Recall that the Jordan curve

$\mathcal{C} \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}$ that was used to define $\mathcal{D}^0 = \mathcal{D}^0(f, \mathcal{C})$ and $\mathcal{D}^1 = \mathcal{D}^1(f, \mathcal{C})$ is f-invariant. Thus $(\mathcal{D}^1, \mathcal{D}^0, L)$ is a two-tile subdivision rule realized by f (see Proposition 12.2).

We now consider the associated cell decompositions $\widetilde{\mathcal{D}}^0$ and $\widetilde{\mathcal{D}}^1$ of the 2-sphere $\widetilde{S}^2$ as given by Lemma 14.15 (iv). By Lemma 14.15 (v) each cell in $\widetilde{\mathcal{D}}^1$ can be represented as $\widetilde{\tau}$ for a unique $\tau \in \mathcal{D}^1$. This implies that if we set $\widetilde{L}(\widetilde{\tau}) := \widetilde{L(\tau)} = \widetilde{f(\tau)} \in \widetilde{\mathcal{D}}^0$ for $\tau \in \mathcal{D}^1$, then we obtain a well-defined map $\widetilde{L} \colon \widetilde{\mathcal{D}}^1 \to \widetilde{\mathcal{D}}^0$.

COROLLARY 14.16. *The map $\widetilde{L} \colon \widetilde{\mathcal{D}}^1 \to \widetilde{\mathcal{D}}^0$ is a labeling. Moreover, $(\widetilde{\mathcal{D}}^1, \widetilde{\mathcal{D}}^0, \widetilde{L})$ is a two-tile subdivision rule isomorphic to $(\mathcal{D}^1, \mathcal{D}^0, L)$.*

Essentially, this corollary says that all the combinatorial information encoded in f and its associated two-tile subdivision rule $(\mathcal{D}^1, \mathcal{D}^0, L)$ is preserved if we pass to the quotient space $\widetilde{S}^2$.

PROOF. The statement easily follows from the definitions in combination with Lemma 14.15. Since we will not use the corollary in any essential way in the following, we leave the straightforward (and somewhat tedious) details to the reader.   □

The induced map $\widetilde{f}$ on $\widetilde{S}^2$. We will now show that f induces a map $\widetilde{f}$ on the sphere $\widetilde{S}^2$.

LEMMA 14.17. *The equivalence relation $\sim$ is f-invariant.*

PROOF. Let $x, y \in S^2$ with $x \sim y$ be arbitrary. We have to show that then $f(x) \sim f(y)$ (see (13.1)).

Pick $\{X^n\}, \{Y^n\} \in \mathcal{S}$ with $x \in \bigcap X^n$ and $y \in \bigcap_n Y^n$. Define $U^n = f(X^{n+1})$ and $V^n = f(Y^{n+1})$ for $n \in \mathbb{N}_0$. Then U^n and V^n are n-tiles, and so $\{U^n\}, \{V^n\} \in \mathcal{S}$. Moreover, $f(x) \in \bigcap_n U^n$ and $f(y) \in \bigcap_n V^n$. Since $x \sim y$ we have $X^n \cap Y^n \neq \emptyset$ for all $n \in \mathbb{N}$. Hence

$$U^n \cap V^n = f(X^{n+1}) \cap f(Y^{n+1}) \supset f(X^{n+1} \cap Y^{n+1}) \neq \emptyset$$

for all $n \in \mathbb{N}_0$. Lemma 14.6 now implies that $f(x) \sim f(y)$ as desired.   □

By the previous lemma the map $\widetilde{f} \colon \widetilde{S}^2 \to \widetilde{S}^2$ given by

$$\widetilde{f}([x]) = [f(x)] \quad \text{for} \quad x \in S^2$$

is well-defined. Then $\widetilde{f} \circ \pi = \pi \circ f$, and it follows from the properties of the quotient topology that $\widetilde{f}$ is continuous (see Lemma A.20).

REMARK 14.18. With some additional effort, one can actually establish that $\sim$ is strongly f-invariant (see Definition 13.1). For this one first shows (using Lemmas 14.6 and 14.11) that each equivalence class M of $\sim$ can be represented in the form

$$(14.7) \qquad M = \bigcup \left\{ \bigcap_n X^n : \{X^n\} \in \mathcal{S}, c \in \bigcap_n X^n \right\},$$

where c is a center of M. For given $x \in S^2$ one then chooses a center c of $M = [f(x)]$ and shows that it has a preimage $c' \in [x]$ under f. By analyzing the different types for M and invoking (14.7) in combination with Lemma 14.8, one can then prove that $f([x]) = [f(x)]$. This implies that $\sim$ is indeed strongly f-invariant.

From this one can conclude that $\widetilde{f}$ is a Thurston map based on Corollary 13.3. We will actually provide a direct simple argument for this that will also show that f and $\widetilde{f}$ are Thurston equivalent.

In the following, $\widetilde{\mathcal{D}}^n = \{\widetilde{\tau} : \tau \in \mathcal{D}^n\}$ for $n \in \mathbb{N}_0$ will denote the cell decomposition of $\widetilde{S}^2$ as provided by Lemma 14.15 (iv). As the next lemma shows, the map $\widetilde{f}^n$ has injectivity properties similar to f^n.

LEMMA 14.19. *Let τ be an n-cell, $n \in \mathbb{N}$. Then $\widetilde{f}^n$ is a homeomorphism of $\widetilde{\tau}$ onto $\widetilde{\sigma}$, where $\sigma = f^n(\tau)$. In particular, $\widetilde{f}^n$ is cellular for $(\widetilde{\mathcal{D}}^n, \widetilde{\mathcal{D}}^0)$.*

PROOF. Since $\widetilde{f}$ is continuous, $\widetilde{f}^n$ is also continuous. Note that f^n is a homeomorphism of τ onto σ. Hence

$$\widetilde{f}^n(\widetilde{\tau}) = (\widetilde{f}^n \circ \pi)(\tau) = (\pi \circ f^n)(\tau) = \widetilde{\sigma}$$

showing that $\widetilde{f}^n$ maps $\widetilde{\tau}$ onto $\widetilde{\sigma}$.

So it remains to show the injectivity of $\widetilde{f}^n$ on $\widetilde{\tau}$, or equivalently, that if $x, y \in \tau$ and $f^n(x) \sim f^n(y)$, then $x \sim y$. Since every vertex, as well as every edge, is contained in a tile, we may also assume that τ is an n-tile.

If $x, y \in \tau$, then we can pick sequences $\{X^k\}$ and $\{Y^k\}$ in $\mathcal{S}$ such that $X^n = Y^n = \tau$ and $x \in \bigcap_k X^k$, $y \in \bigcap_k Y^k$. Then $f^n(X^{k+n})$ and $f^n(Y^{k+n})$ are k-tiles for $k \in \mathbb{N}_0$. Moreover, the sequences $\{f^n(X^{k+n})\}$ and $\{f^n(Y^{k+n})\}$ are in $\mathcal{S}$, and $f^n(x) \in \bigcap_k f^n(X^{k+n})$ and $f^n(y) \in \bigcap_k f^n(Y^{k+n})$. Since $f^n(x) \sim f^n(y)$, this implies that $f^n(X^{k+n}) \cap f^n(Y^{k+n}) \neq \emptyset$ for all $k \in \mathbb{N}_0$. Since $X^{k+n}, Y^{k+n} \subset \tau$ for $k \geq 0$ and $f^n|\tau$ is injective, we conclude that $X^{k+n} \cap Y^{k+n} \neq \emptyset$ for $k \geq 0$. Since $X^n = Y^n = \tau$, we also have $X^k = Y^k$ for $k = 0, \ldots, n-1$. Hence

$X^k \cap Y^k \neq \emptyset$ for all $k \geq 0$. Lemma 14.6 then shows that $x \sim y$ as desired.

The fact that $\widetilde{f}^n$ is cellular for $(\widetilde{\mathcal{D}}^n, \widetilde{\mathcal{D}}^0)$ follows from the first part of the proof and the fact that f^n is cellular for $(\mathcal{D}^n, \mathcal{D}^0)$. $\qquad\square$

The auxiliary homeomorphisms h_0 and h_1. To prove that $\widetilde{f}$ is a Thurston map equivalent to f, we need to define homeomorphisms $h_0, h_1 \colon S^2 \to \widetilde{S}^2$ that make the diagram

(14.8)
$$\begin{array}{ccc} S^2 & \xrightarrow{h_1} & \widetilde{S}^2 \\ f\downarrow & & \downarrow\widetilde{f} \\ S^2 & \xrightarrow{h_0} & \widetilde{S}^2 \end{array}$$

commutative and are isotopic rel. $\mathbf{V}^0 = \mathrm{post}(f)$. The construction of these maps follows ideas in the proof of Lemma 12.13.

For the definition of h_0 recall that S^2 is the union of two 0-tiles $X_{\mathtt{b}}^0$ and $X_{\mathtt{w}}^0$ with common boundary $\mathcal{C}$. The Jordan curve $\mathcal{C}$ is further decomposed into $k = \#\mathbf{V}^0 \geq 3$ 0-edges and 0-vertices. The cell decomposition $\widetilde{\mathcal{D}}^0$ of $\widetilde{S}^2$ contains two tiles $\widetilde{X}_{\mathtt{b}}^0$ and $\widetilde{X}_{\mathtt{w}}^0$. Lemma 14.15 (ii) and Lemma 14.14 show that the common boundary of $\widetilde{X}_{\mathtt{b}}^0$ and $\widetilde{X}_{\mathtt{w}}^0$ is $\widetilde{\mathcal{C}} = \pi(\mathcal{C})$, which is a Jordan curve. There are k distinct vertices and edges on $\widetilde{\mathcal{C}}$. There are no other cells in $\widetilde{\mathcal{D}}^0$.

We know by Lemma 14.15 (v) that the map $\tau \in \mathcal{D}^0 \mapsto \widetilde{\tau} \in \widetilde{\mathcal{D}}^0$ is an isomorphism between the cell complexes $\mathcal{D}^0$ and $\widetilde{\mathcal{D}}^0$. So Lemma 5.11 (ii) implies that there exists a homeomorphism $h_0 \colon S^2 \to \widetilde{S}^2$ such that $h_0(\tau) = \widetilde{\tau}$ for all cells $\tau \in \mathcal{D}^0$.

Now let $\tau \in \mathcal{D}^1$ be arbitrary. Then $f(\tau) \in \mathcal{D}^0$, and by Lemma 14.19 the map $\widetilde{f}|\widetilde{\tau}$ is a homeomorphism of $\widetilde{\tau}$ onto $\widetilde{f(\tau)} = h_0(f(\tau))$. Hence the map

$$\varphi_\tau := (\widetilde{f}|\widetilde{\tau})^{-1} \circ h_0 \circ (f|\tau)$$

is well-defined and a homeomorphism from τ onto $\widetilde{\tau}$. If $x \in \tau$, then $y = \varphi_\tau(x)$ is the unique point $y \in \widetilde{\tau}$ with $\widetilde{f}(y) = h_0(f(x))$. As in the proof of Lemma 12.13, this uniqueness property implies that if $\sigma, \tau \in \mathcal{D}^1$ and $\sigma \subset \tau$, then $\varphi_\tau|\sigma = \varphi_\sigma$. From this in turn one can deduce that if a point $x \in S^2$ lies in two cells $\tau, \tau' \in \mathcal{D}^1$, then $\varphi_\tau(x) = \varphi_{\tau'}(x)$. This allows us to define a map $h_1 \colon S^2 \to \widetilde{S}^2$ as follows. If $x \in S^2$, we pick $\tau \in \mathcal{D}^1$ with $x \in \tau$ and set $h_1(x) := \varphi_\tau(x)$. Then $h_1 \colon S^2 \to \widetilde{S}^2$ is well-defined.

LEMMA 14.20. *The map* $h_1 \colon S^2 \to \widetilde{S}^2$ *is a homeomorphism of* S^2 *onto* $\widetilde{S}^2$ *satisfying* $h_0 \circ f = \widetilde{f} \circ h_1$. *Moreover, we have* $h_0(\mathcal{C}) = \widetilde{\mathcal{C}} = h_1(\mathcal{C})$ *and the homeomorphisms* h_0 *and* h_1 *are isotopic rel.* $\mathbf{V}^0 = \mathrm{post}(f)$.

PROOF. We have $h_1|\tau = \varphi_\tau$ for each cell $\tau \in \mathcal{D}^1$. So the definitions of h_1 and φ_τ show that $h_0 \circ f = \widetilde{f} \circ h_1$ and that $h_1(\tau) = \varphi_\tau(\tau) = \widetilde{\tau}$ for each $\tau \in \mathcal{D}^1$. Since $\tau \in \mathcal{D}^1 \mapsto \widetilde{\mathcal{D}}^1$ is an isomorphism of cell complexes by Lemma 14.15 (v), the last statement implies that h_1 is a homeomorphism of S^2 onto $\widetilde{S}^2$ (Lemma 5.11 (i)).

Note that $h_1(\tau) = \widetilde{\tau}$ also for each $\tau \in \mathcal{D}^0$. Indeed, suppose $\tau \in \mathcal{D}^0$ is arbitrary. Since $\mathcal{D}^1$ is a refinement of $\mathcal{D}^0$, for each $x \in \tau$ there exists $\sigma \in \mathcal{D}^1$ such that $x \in \sigma \subset \tau$. Then $h_1(x) \in h_1(\sigma) = \widetilde{\sigma} \subset \widetilde{\tau}$. So $h_1(\tau) \subset \widetilde{\tau}$. Conversely, let $y \in \widetilde{\tau}$ be arbitrary. Since $\widetilde{\mathcal{D}}^1$ is a refinement of $\widetilde{\mathcal{D}}^0$ (Lemma 14.15 (vi)), there exists a cell $\sigma \in \mathcal{D}^1$ such that $y \in \widetilde{\sigma} \subset \widetilde{\tau}$. By Lemma 14.12 we then have $\sigma \subset \tau$, and so $y \in \widetilde{\sigma} = h_1(\sigma) \subset h_1(\tau)$. We conclude that $h_1(\tau) = \widetilde{\tau}$ for each $\tau \in \mathcal{D}^0$ as claimed.

The Jordan curve $\mathcal{C}$ is the 1-skeleton of $\mathcal{D}^0$ and thus equal to the union of all edges $e \in \mathcal{D}^0$. We know that $h_0(e) = h_1(e) = \widetilde{e} = \pi(e)$ for each such edge e. Hence $h_0(\mathcal{C}) = h_1(\mathcal{C}) = \pi(\mathcal{C}) = \widetilde{\mathcal{C}}$.

If $\tau \in \mathcal{D}^0$ is arbitrary, then $(h_1^{-1} \circ h_0)(\tau) = h_1^{-1}(\widetilde{\tau}) = \tau$. So by Lemma 5.11 (iii) the homeomorphism $h_1^{-1} \circ h_0$ is isotopic to id_{S^2} rel. $\mathbf{V}^0$. Hence h_0 and h_1 are isotopic rel. $\mathbf{V}^0 = \mathrm{post}(f)$. $\qquad\square$

We are now ready to prove the main results of this chapter.

PROOF OF PROPOSITION 14.3. Let f be a Thurston map that is combinatorially expanding for a Jordan curve $\mathcal{C}$ as in the statement. Then all of our previous considerations apply.

We consider the 2-sphere $\widetilde{S}^2 = S^2/\!\sim$, the quotient map $\pi \colon S^2 \to \widetilde{S}^2$, and the map $\widetilde{f} \colon \widetilde{S}^2 \to \widetilde{S}^2$ defined earlier. As we have seen in Lemma 14.20, there exist homeomorphisms $h_0, h_1 \colon S^2 \to \widetilde{S}^2$ that are isotopic rel. $\mathrm{post}(f)$ and satisfy $h_0 \circ f = \widetilde{f} \circ h_1$. Then $\widetilde{f}$ is a Thurston map with $\mathrm{post}(\widetilde{f}) = h_0(\mathrm{post}(f)) = \pi(\mathrm{post}(f))$ by Lemma 2.5, and it is clear that f and $\widetilde{f}$ are Thurston equivalent.

The Jordan curve $\widetilde{\mathcal{C}} = \pi(\mathcal{C})$ satisfies

$$\mathrm{post}(\widetilde{f}) = \pi(\mathrm{post}(f)) \subset \pi(\mathcal{C}) = \widetilde{\mathcal{C}}.$$

Since $f(\mathcal{C}) \subset \mathcal{C}$, we also have

$$\widetilde{f}(\widetilde{\mathcal{C}}) = (\widetilde{f} \circ \pi)(\mathcal{C}) = (\pi \circ f)(\mathcal{C}) \subset \pi(\mathcal{C}) = \widetilde{\mathcal{C}}.$$

This shows that $\widetilde{\mathcal{C}}$ is invariant with respect to $\widetilde{f}$ and contains the set of postcritical points of $\widetilde{f}$. In the proof of the previous lemma we have also seen that $h_0(\mathcal{C}) = h_1(\mathcal{C}) = \widetilde{\mathcal{C}}$.

It remains to show that $\widetilde{f}$ is expanding. Since $\widetilde{\mathcal{C}} \subset \widetilde{S}^2$ is an $\widetilde{f}$-invariant Jordan curve with $\mathrm{post}(\widetilde{f}) \subset \widetilde{\mathcal{C}}$, we can do this by verifying the condition in Lemma 12.7 for $\widetilde{\mathcal{C}}$.

Note that $\mathcal{D}^0(\widetilde{f}, \widetilde{\mathcal{C}}) = \widetilde{\mathcal{D}}^0$. Moreover, since the map $\widetilde{f}^n$ is cellular for $(\widetilde{\mathcal{D}}^n, \widetilde{\mathcal{D}}^0)$, it follows from Lemma 5.12 that $\mathcal{D}^n(\widetilde{f}, \widetilde{\mathcal{C}}) = \widetilde{\mathcal{D}}^n$ for all $n \in \mathbb{N}_0$. So the n-tiles for $(\widetilde{f}, \widetilde{\mathcal{C}})$ are precisely the sets $\widetilde{X} = \pi(X)$, where X is an n-tile on S^2 for $(f, \mathcal{C})$.

Let $\widetilde{X}^0 \supset \widetilde{X}^1 \supset \widetilde{X}^2 \supset \ldots$ be a nested sequence of n-tiles for $(\widetilde{f}, \widetilde{\mathcal{C}})$. Clearly, $\bigcap_n \widetilde{X}^n$ is non-empty. We have to show that this intersection does not contain more than one point. From Lemma 14.15 (vi) it follows that the corresponding sequence $\{X^n\}$ of n-tiles for $(f, \mathcal{C})$ is nested, and so $\{X^n\} \in \mathcal{S} = \mathcal{S}(f, \mathcal{C})$. To see that $\bigcap_n \widetilde{X}^n$ consists of precisely one point, we argue by contradiction and assume that $\bigcap_n \widetilde{X}^n$ contains more than one point, or equivalently, there exist two distinct (and hence disjoint) equivalence classes M and N with respect to $\sim$ such that $M^n := M \cap X^n \neq \emptyset$ and $N^n := N \cap X^n \neq \emptyset$ for all $n \in \mathbb{N}_0$. Since equivalence classes and tiles are compact, in this way we get descending sequences $M^0 \supset M^1 \supset \ldots$ and $N^0 \supset N^1 \supset \ldots$ of non-empty and compact sets. Hence the sets $\bigcap_n M^n = M \cap \bigcap_n X^n$ and $\bigcap_n N^n = N \cap \bigcap_n X^n$ are non-empty. So there exist points $x \in M \cap \bigcap_n X^n$ and $y \in N \cap \bigcap_n X^n$. Since x and y lie in different equivalence classes, they are not equivalent. On the other hand, we have $x, y \in \bigcap_n X^n$ and $\{X^n\} \in \mathcal{S}$. Hence $x \sim y$ by Lemma 14.6. This is a contradiction and we conclude that $\widetilde{f}$ is indeed expanding. $\square$

REMARK 14.21. The Thurston map $\widetilde{f}$ realizes the subdivision rule $(\widetilde{\mathcal{D}}^1, \widetilde{\mathcal{D}}^0, \widetilde{L})$ given by Corollary 14.16. Indeed, we know that $\widetilde{f}$ is cellular for $(\widetilde{\mathcal{D}}^1, \widetilde{\mathcal{D}}^0)$ (see Lemma 14.19). To show compatibility of $\widetilde{f}$ with the labeling $\widetilde{L}$, first note that each cell in $\widetilde{\mathcal{D}}^1$ has a representation in the form $\widetilde{\tau}$ with a unique $\tau \in \mathcal{D}^1$. As we have seen in the proof of Lemma 14.20, we have $h_1(\tau) = \widetilde{\tau}$. Moreover, $f(\tau) \in \mathcal{D}^0$ and so $h_0(f(\tau)) = \widetilde{f(\tau)} = \widetilde{L}(\widetilde{\tau})$ by the definitions of h_0 and $\widetilde{L}$. This leads to the desired relation

$$\widetilde{f}(\widetilde{\tau}) = (\widetilde{f} \circ h_1)(\tau) = (h_0 \circ f)(\tau) = h_0(f(\tau)) = \widetilde{L}(\widetilde{\tau}).$$

Our previous considerations immediately give the proofs of Theorems 14.2 and 14.1.

PROOF OF THEOREM 14.2. We use the same notation as in Proposition 14.3 and define $\phi = h_1^{-1} \circ h_0$. Then ϕ is isotopic to the identity on S^2 rel. post(f) and $\phi(\mathcal{C}) = \mathcal{C}$. Moreover,

$$g = \phi \circ f = h_1^{-1} \circ h_0 \circ f = h_1^{-1} \circ \widetilde{f} \circ h_1,$$

and so g is topologically conjugate to the expanding Thurston map $\widetilde{f}$, and hence itself an expanding Thurston map.

Similarly, the map

$$\widetilde{g} = f \circ \phi = f \circ h_1^{-1} \circ h_0 = h_0^{-1} \circ \widetilde{f} \circ h_0.$$

is topologically conjugate to $\widetilde{f}$, and hence an expanding Thurston map. $\square$

PROOF OF THEOREM 14.1. Let $(\mathcal{D}^1, \mathcal{D}^0, L)$ be a two-tile subdivision rule on S^2 as in the statement, $\mathcal{C}$ be the Jordan curve and $\mathbf{V}^0$ be the vertex set of $\mathcal{D}^0$.

If the subdivision rule $(\mathcal{D}^1, \mathcal{D}^0, L)$ can be realized by an expanding Thurston map $f\colon S^2 \to S^2$, then $\mathcal{C}$ is f-invariant, post(f) $\subset \mathcal{C}$, and $\#\,\mathrm{post}(f) \geq 3$. So Lemma 8.6 implies that f is combinatorially expanding for $\mathcal{C}$, and so $(\mathcal{D}^1, \mathcal{D}^0, L)$ is combinatorially expanding according to Definition 12.18 and the discussion following this definition.

Conversely, suppose $(\mathcal{D}^1, \mathcal{D}^0, L)$ is combinatorially expanding. Then this subdivision rule can be realized by a Thurston map $f\colon S^2 \to S^2$ that is combinatorially expanding for $\mathcal{C}$. From our assumption we obtain that post(f) $= \mathbf{V}^0$ (see Remark 12.12 (i)). Note that then $\mathcal{D}^0 = \mathcal{D}^0(f, \mathcal{C})$. Moreover, f is cellular for $(\mathcal{D}^1, \mathcal{D}^0)$ and so necessarily $\mathcal{D}^1 = \mathcal{D}^1(f, \mathcal{C})$ (see Lemma 5.12).

We again use the notation of Proposition 14.3. We define $\phi = h_1^{-1} \circ h_0$. Then ϕ is a orientation-preserving homeomorphism on S^2 that is isotopic to id_{S^2} rel. $\mathbf{V}^0 = \mathrm{post}(f)$. As in the proof of Theorem 14.2, let $g = \phi \circ f = h_1^{-1} \circ \widetilde{f} \circ h_1$. Then $g\colon S^2 \to S^2$ is an expanding Thurston map.

Since $\phi(\mathcal{C}) = \mathcal{C}$ and ϕ is orientation-preserving and the identity on $\mathbf{V}^0$, we have $\phi(c) = c$ for each $c \in \mathcal{D}^0$. Since f is cellular for $(\mathcal{D}^1, \mathcal{D}^0)$, the map $g = \phi \circ f$ is cellular for $(\mathcal{D}^1, \mathcal{D}^0)$ and we have $g(c) = f(c)$ for each cell $c \in \mathcal{D}^1$. Since f realizes the subdivision rule with the given labeling, this shows that g is also a realization. Now g is expanding, and so the claim follows. $\square$

We end this chapter with two examples. The first one illustrates why we need the condition post(f) $= \mathbf{V}^0$ in Theorem 14.1.
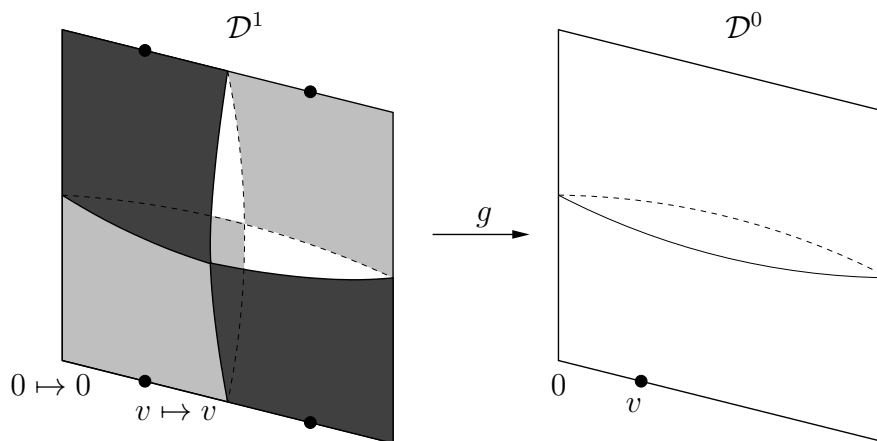
FIGURE 14.2. A two-tile subdivision rule realized by a map g with $\mathrm{post}(g) \neq \mathbf{V}^0$.

EXAMPLE 14.22. Consider the two-tile subdivision rule $(\mathcal{D}^1, \mathcal{D}^0, L)$ indicated in Figure 14.2. It is almost the same as the one in Figure 1.1 realized by the Lattès map discussed in Section 1.1. However, there is one difference: $\mathcal{D}^0$ contains an additional vertex v; so $\mathcal{D}^0$ contains five vertices (and five edges). Compared to Figure 1.1, the cell decomposition $\mathcal{D}^1$ of the subdivision rule $(\mathcal{D}^1, \mathcal{D}^0, L)$ contains four additional vertices (represented by black dots in Figure 14.2), one of which agrees with v.

Let $g \colon S^2 \to S^2$ be an arbitrary Thurston map that realizes this two-tile subdivision rule. Then the set $\mathrm{post}(g)$ consists of the four corners of the squares forming the pillow, and so $\mathrm{post}(g) \neq \mathbf{V}^0$. Moreover, g is combinatorially expanding for the Jordan curve $\mathcal{C}$ of $\mathcal{D}^0$ (which is the equator of the pillow), because no tile in $\mathcal{D}^1$ joins opposite sides of $\mathcal{C}$. Note that for combinatorial expansion of g for $\mathcal{C}$ only the points in $\mathrm{post}(g)$ are relevant, and so the extra point v plays no role here.

Let e be the edge in $\mathcal{D}^0$ whose endpoints are the vertices labeled by 0 and v. Then e is also an edge in $\mathcal{D}^1$ and $g(e) = e$. This implies that e is an n-edge for $(g, \mathcal{C})$ for each $n \in \mathbb{N}_0$ and so g cannot be expanding.

Recall from Theorem 14.2 that combinatorial expansion is a sufficient condition for a Thurston map to be equivalent to an expanding Thurston map. Our last example in this chapter shows that this condition is not necessary.

EXAMPLE 14.23. We consider the map $f \colon S^2 \to S^2$ represented by the top part of Figure 14.3. Here we identify S^2 with a pillow that is obtained by gluing two squares together along their boundaries. The map f has four postcritical points, which are the vertices of the pillow
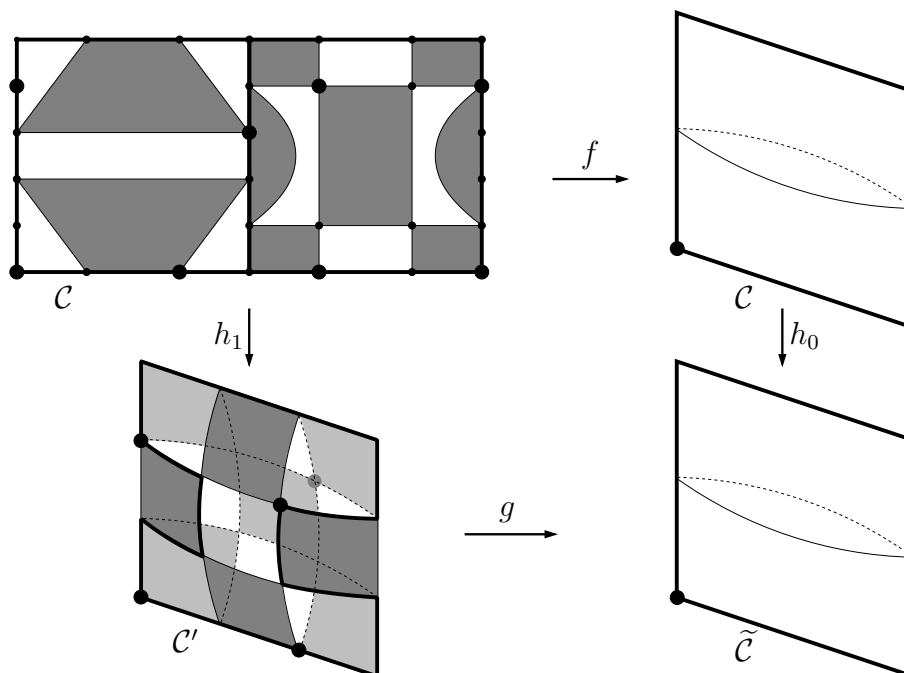
FIGURE 14.3. The map f is not combinatorially expanding, but equivalent to the expanding map g.

shown on the top right. Its front is the white 0-tile, and its back the black 0-tile. The subdivision of the 0-tiles is indicated on the top left of the figure. Here we have cut the pillow along three 0-edges and folded the back of the pillow up so that we see two adjacent squares. The left one shows the subdivision of the white 0-tile, and the right one the subdivision of the black 0-tile. One postcritical point (which is a vertex of the pillow) is marked by a large black dot. On the left we indicated its preimages, meaning the 1-vertices that are labeled by this 0-vertex. The other 1-vertices are shown as small black dots.

The equator $\mathcal{C}$ of the pillow S^2 is an f-invariant Jordan curve containing $\mathrm{post}(f)$. The map f is not combinatorially expanding for $\mathcal{C}$: for each $n \in \mathbb{N}_0$ there is a white n-tile (contained in the white 0-tile) that joins the 0-edges given by the left and the right side of the white 0-tile.

We want to show that f is equivalent to an expanding Thurston map $g\colon S^2 \to S^2$ defined on the same pillow as f. The map g is indicated at the bottom of Figure 14.3. If we identify the pillow S^2 with $\widehat{\mathbb{C}}$ in the same way as in Section 1.1, then g is a Lattès map. It is obtained according to Theorem 3.1 (ii) as a quotient of $A\colon \mathbb{C} \to \mathbb{C}$,

$z \mapsto A(z) := 3z$, by the crystallographic group of type (2222) in (3.24). Since g is a Lattès map, it is expanding.

Let $h_0 := \mathrm{id}_{S^2}$. We consider $\mathcal{D}^1 = \mathcal{D}^1(f, \mathcal{C})$ and $\widetilde{\mathcal{D}}^1 = \mathcal{D}^1(g, \mathcal{C})$ as given by Definition 5.14. These are the cell decompositions of S^2 shown on the left of Figure 14.3. Note that $\mathcal{D}^1$ and $\widetilde{\mathcal{D}}^1$ are in fact isomorphic. More precisely, there is a bijection $\phi \colon \mathcal{D}^1 \to \widetilde{\mathcal{D}}^1$ as in Definition 5.10 that preserves the color of tiles and sends the 1-vertices on the top left marked by a large black dot to the 1-vertices on the bottom left marked in the same way. From this one can deduce that $g(\phi(\tau)) = f(\tau) = h_0(f(\tau))$ for each $\tau \in \mathcal{D}^1$ (this is closely related to Lemma 12.15 (ii)).

This in turn allows us to define a map $h_1 \colon S^2 \to S^2$ by setting

$$h_1(x) := \big((g|\phi(\tau))^{-1} \circ h_0 \circ (f|\tau)\big)(x),$$

whenever $x \in S^2$ and $x \in \tau \in \mathcal{D}^1$. As in the proof of Lemma 12.13, one can show that h_1 is well-defined. Lemma 5.11 (i) implies that h_1 is a homeomorphism on S^2. From the definition of h_1 it is also clear that $g \circ h_1 = h_0 \circ f$.

So in order to conclude that f and g are Thurston equivalent, it remains to show that h_1 and $h_0 = \mathrm{id}_{S^2}$ are isotopic rel. $\mathrm{post}(f)$. First note that the definition of h_1 implies that this map fixes the points in $\mathrm{post}(f) = \mathrm{post}(g)$, i.e., the 0-vertices.

Let $\mathcal{C}' := h_1(\mathcal{C}) \subset S^2$. This is the Jordan curve drawn on the bottom left of Figure 14.3 with a thick line. It is intuitively clear and not hard to prove that $\mathcal{C}'$ can be deformed into $\mathcal{C}$ by an isotopy rel. $\mathrm{post}(f)$. By using such an isotopy one can show that h_1 is isotopic rel. $\mathrm{post}(f)$ to a homeomomorphism $\widetilde{h}_0$ on S^2 that preserves all 0-cells as sets (i.e., all cells in $\mathcal{D}^0(f, \mathcal{C}) = \mathcal{D}^0(g, \mathcal{C})$). Lemma 5.11 (iii) then implies that $\widetilde{h}_0$, and hence also h_1, is isotopic to $h_0 = \mathrm{id}_{S^2}$ rel. $\mathrm{post}(f)$. The Thurston equivalence of f and g follows.

CHAPTER 15

Invariant curves

This chapter is central for this work. We will prove existence and uniqueness results for invariant curves $\mathcal{C}$ of an expanding Thurston map f. We will also show that if an invariant curve exists, then it can be obtained from an iterative procedure, and that it is a quasicircle. We always require that $\mathcal{C}$ is a Jordan curve and that $\operatorname{post}(f) \subset \mathcal{C}$, but in the following discussion we will often refer to such a curve $\mathcal{C}$ simply as an invariant curve for brevity.

One of our main results can be formulated as follows.

THEOREM 15.1 (Sufficiently high iterates have invariant curves). *Let* $f\colon S^2 \to S^2$ *be an expanding Thurston map, and* $\mathcal{C} \subset S^2$ *be a Jordan curve with* $\operatorname{post}(f) \subset \mathcal{C}$. *Then for each sufficiently large* $n \in \mathbb{N}$ *there exists a Jordan curve* $\widetilde{\mathcal{C}} \subset S^2$ *that is invariant for* f^n *and isotopic to* $\mathcal{C}$ *rel.* $\operatorname{post}(f)$.

This existence result has the following important implication.

COROLLARY 15.2 (Thurston maps as two-tile subdivision rules). *Let* $f\colon S^2 \to S^2$ *be an expanding Thurston map. Then for each sufficiently large* $n \in \mathbb{N}$ *there exists a two-tile subdivision rule that is realized by* $F = f^n$.

This justifies our approach of studying expanding Thurston maps from a combinatorial perspective based on cellular Markov partitions.

Invariant curves are quasicircles if the underlying metric is visual.

THEOREM 15.3 (Invariant curves are quasicircles). *Let* $f\colon S^2 \to S^2$ *be an expanding Thurston map, and* $\mathcal{C} \subset S^2$ *be a Jordan curve with* $\operatorname{post}(f) \subset \mathcal{C}$. *If* $\mathcal{C}$ *is* f*-invariant, then* $\mathcal{C}$ *equipped with (the restriction of) a visual metric for* f *is a quasicircle.*

This also applies to invariant curves of iterates, because if $f\colon S^2 \to S^2$ is an expanding Thurston map, then the same is true for each iterate $F = f^n$, $n \in \mathbb{N}$.

If one studies rational Thurston maps f that are expanding, then the underlying 2-sphere is the Riemann sphere $\widehat{\mathbb{C}}$, and it is natural to equip it with the chordal metric σ. Then an invariant $\mathcal{C}$ curve as in

the previous theorem is also a quasicircle with respect to σ. This can be deduced from Theorem 15.3 once we know that for such maps the chordal metric is quasisymmetrically equivalent to each visual metric. This will be proved in Chapter 18; see in particular Corollary 18.10.

As we discussed in Section 12.1, if $\mathcal{C}$ an f-invariant Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}$, then we get a sequence of cell decompositions $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$, $n \in \mathbb{N}_0$, so that each cell decomposition is refined by the cell decompositions of higher level. We will see that Theorem 15.3 implies that we get good control for the geometry of edges and tiles in these cell decompositions. Namely, the family of edges consists of *uniform quasiarcs* and the boundary of tiles are *uniform quasicircles*. See Section 15.3 for an explanation of this terminology and Proposition 15.26 for a precise statement.

In Theorem 15.1 it is necessary to pass to an iterate of the map to guarantee existence of an invariant Jordan curve, because there are examples of maps for which an invariant curve does not exist (see Example 15.11). One can formulate a necessary and sufficient criterion for the existence of invariant curves.

THEOREM 15.4 (Existence of invariant curves). *Let $f\colon S^2 \to S^2$ be an expanding Thurston map. Then the following conditions are equivalent:*

(i) *There exists a Jordan curve $\widetilde{\mathcal{C}} \subset S^2$ with $\mathrm{post}(f) \subset \widetilde{\mathcal{C}}$ that is f-invariant.*

(ii) *There exist Jordan curves $\mathcal{C}, \mathcal{C}' \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}, \mathcal{C}'$ and $\mathcal{C}' \subset f^{-1}(\mathcal{C})$, and an isotopy $H\colon S^2 \times I \to S^2$ rel. $\mathrm{post}(f)$ with $H_0 = \mathrm{id}_{S^2}$ and $H_1(\mathcal{C}) = \mathcal{C}'$ such that the map*

$$\widehat{f} := H_1 \circ f \text{ is combinatorially expanding for } \mathcal{C}'.$$

Moreover, if (ii) *is true, then there exists an f-invariant Jordan curve $\widetilde{\mathcal{C}} \subset S^2$ with $\mathrm{post}(f) \subset \widetilde{\mathcal{C}}$ that is isotopic to $\mathcal{C}$ rel. $\mathrm{post}(f)$ and isotopic to $\mathcal{C}'$ rel. $f^{-1}(\mathrm{post}(f))$.*

The first condition in (ii) says that there exists a Jordan curve $\mathcal{C}$ with $\mathrm{post}(f) \subset \mathcal{C}$ that can be isotoped rel. $\mathrm{post}(f)$ into its preimage under f. This condition alone ensures that an associated f-invariant set $\widetilde{\mathcal{C}}$ with $\mathrm{post}(f) \subset \widetilde{\mathcal{C}}$ exists, but in general $\widetilde{\mathcal{C}}$ will not be a Jordan curve (see Lemma 15.18 (viii) and Example 15.23). If, in addition, the map $\widehat{f}$ is combinatorially expanding as stipulated in (ii), then one obtains a Jordan curve $\widetilde{\mathcal{C}}$.

Invariant curves can be constructed by an iterative procedure that will be described in Section 15.2. In the situation of Theorem 15.4

one lifts the isotopy H by the map f repeatedly to obtain a sequence of isotopies $H^n \colon S^2 \times I \to S^2$, $n \in \mathbb{N}_0$, with $H^0 \coloneqq H$ such that $H_0^n = H^n(\cdot, 0) = \mathrm{id}_{S^2}$. One sets $\mathcal{C}^0 \coloneqq \mathcal{C}$ and defines inductively $\mathcal{C}^{n+1} \coloneqq H_1^n(\mathcal{C}^n) = H^n(\mathcal{C}^n, 1)$ for $n \in \mathbb{N}_0$. It can then be shown that the sequence $\{\mathcal{C}_n\}$ Hausdorff converges to the desired invariant curve $\widetilde{\mathcal{C}}$ (see Proposition 15.20). An explicit knowledge of the isotopies is not really necessary, because one can interpret this as an edge replacement procedure (see Remark 15.22). In Section 15.2 we will discuss this and several examples that illustrate various phenomena in this context.

Theorem 15.1, which is our basic existence result for invariant curves, is complemented by the following uniqueness statement.

THEOREM 15.5 (Uniqueness of invariant curves). *Let $f \colon S^2 \to S^2$ be an expanding Thurston map, and $\mathcal{C}, \mathcal{C}' \subset S^2$ be f-invariant Jordan curves that both contain the set $\mathrm{post}(f)$. Then $\mathcal{C} = \mathcal{C}'$ if and only if $\mathcal{C}$ and $\mathcal{C}'$ are isotopic rel. $f^{-1}(\mathrm{post}(f))$.*

This implies that in a given isotopy class rel. $\mathrm{post}(f)$ there are only finitely many invariant curves $\mathcal{C}$ (Corollary 15.8). It follows that an expanding Thurston map f with $\# \mathrm{post}(f) = 3$ can have only finitely many invariant curves $\mathcal{C}$ (Corollary 15.8).

The situation changes if one does not restrict the isotopy class of $\mathcal{C}$. An expanding Thurston map f may have infinitely many invariant curves $\mathcal{C}$ in general (see Example 15.9). If, in addition, the map is rational and has a hyperbolic orbifold, then this cannot happen and f can have only finitely many invariant curves $\mathcal{C}$ (see Theorem 15.10; we are grateful to K. Pilgrim for pointing this out to us).

The chapter is organized as follows. Section 15.1 is devoted to existence and uniqueness results, where we provide proofs for the statements discussed above. The iterative procedure for the construction of invariant curves is explained in Section 15.2. In Section 15.3 we discuss the quasiconformal geometry of invariant curves. Here we prove Theorem 15.3 and related results.

Much of our discussion in this chapter is quite technical. Before we go into the details, we look at a specific example that will illustrate some of the main ideas.

EXAMPLE 15.6. Let $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be the map defined by

$$f(z) = 1 + (\omega - 1)/z^3$$

for $z \in \widehat{\mathbb{C}}$, where $\omega = e^{4\pi i/3}$. This map was already considered in Examples 2.6 and 12.24. It realizes the two-tile subdivision rule shown in Figure 2.1.
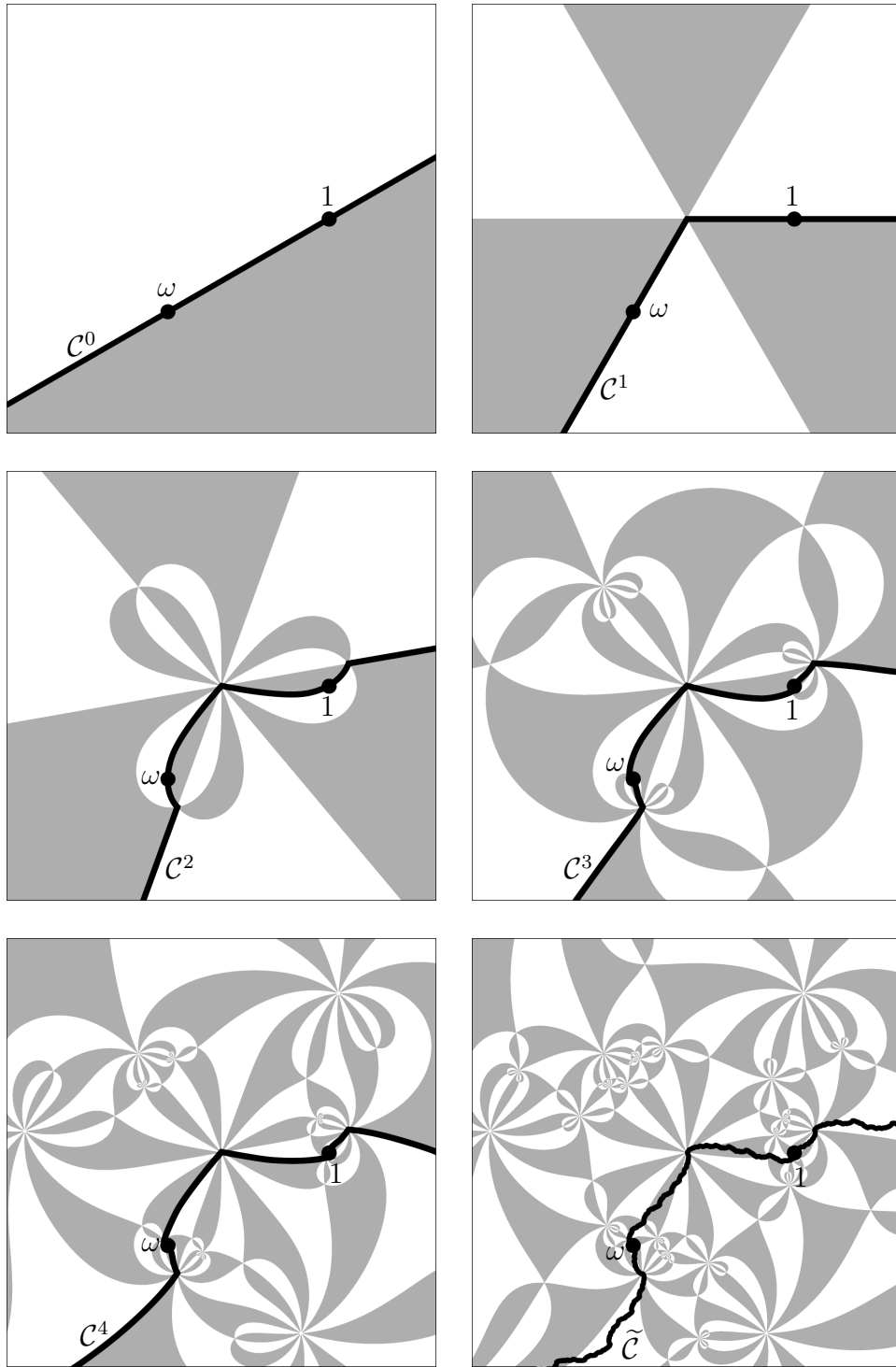
FIGURE 15.1. The invariant curve for Example 15.6.

Note that $f(z) = \tau(z^3)$, where $\tau(w) = 1 + (\omega - 1)/w$ is a Möbius transformation that maps the upper half-plane to the half-plane above the line through the points ω and 1 (indeed, τ maps $0, 1, \infty$ to $\infty, \omega, 1$, respectively). We have $\text{crit}(f) = \{0, \infty\}$ and $\text{post}(f) = \{\omega, 1, \infty\}$.

One can obtain an f-invariant Jordan curve $\widetilde{\mathcal{C}} \subset \widehat{\mathbb{C}}$ with $\text{post}(f) \subset \widetilde{\mathcal{C}}$ as follows. We first pick a Jordan curve $\mathcal{C}^0 \subset \widehat{\mathbb{C}}$ containing all postcritical points of f. More specifically, let $\mathcal{C}^0$ be the (extended) line through ω and 1 (i.e., the circle on $\widehat{\mathbb{C}}$ through $\omega, 1, \infty$).

Now consider $f^{-1}(\mathcal{C}^0) = \bigcup_{k=0,\dots,5} R_k$, where

$$R_k = \{re^{ik\pi/3} : 0 \le r \le \infty\}$$

is the ray from 0 through the sixth root of unity $e^{ik\pi/3}$; see the top right in Figure 15.1. We choose a Jordan curve $\mathcal{C}^1 \subset \widehat{\mathbb{C}}$ such that

$$\mathcal{C}^1 \subset f^{-1}(\mathcal{C}^0), \ \text{post}(f) \subset \mathcal{C}^1, \ \text{and} \ \mathcal{C}^1 \text{ is isotopic to } \mathcal{C}^0 \text{ rel. post}(f).$$

For general Thurston maps a similar choice is not always possible, but in our specific case there is a unique Jordan curve $\mathcal{C}^1 \subset f^{-1}(\mathcal{C}^0)$ with $\text{post}(f) \subset \mathcal{C}^1$, namely $\mathcal{C}^1 = R_0 \cup R_4$, the union of the two rays through ω and through 1. Since $\#\text{post}(f) = 3$, the requirement that $\mathcal{C}^1$ is isotopic to $\mathcal{C}^0$ rel. $\text{post}(f)$ is automatic for our specific map f by Lemma 11.10. Let $H \colon \widehat{\mathbb{C}} \times I \to \widehat{\mathbb{C}}$ be an isotopy rel. $\text{post}(f)$ that deforms $\mathcal{C}^0$ to $\mathcal{C}^1$, i.e., $H_0 = \text{id}_{\widehat{\mathbb{C}}}$ and $H_1(\mathcal{C}^0) = \mathcal{C}^1$.

Given the data $\mathcal{C}^0$, $\mathcal{C}^1$, and H, there are two (essentially equivalent) ways to obtain an f-invariant Jordan curve isotopic to $\mathcal{C}^1$ rel. $f^{-1}(\text{post}(f))$ and hence also isotopic to $\mathcal{C}^0$ rel. $\text{post}(f)$.

For the first approach we consider the Thurston map $\widehat{f} := H_1 \circ f$. Since $\mathcal{C}^1 \subset f^{-1}(\mathcal{C}^0)$ we have $f(\mathcal{C}^1) \subset \mathcal{C}^0$, and so

$$\widehat{f}(\mathcal{C}^1) = (H_1 \circ f)(\mathcal{C}^1) \subset H_1(\mathcal{C}^0) = \mathcal{C}^1.$$

Thus $\mathcal{C}^1$ is $\widehat{f}$-invariant. The two-tile subdivision rule given by $\mathcal{D}^1 = \mathcal{D}^1(\widehat{f}, \mathcal{C}^1)$, $\mathcal{D}^0 = \mathcal{D}^0(\widehat{f}, \mathcal{C}^1)$, and the labeling induced by $\widehat{f}$ is as in Figure 2.1. The map $\widehat{f}$ is combinatorially expanding for $\mathcal{C}^1$; indeed, no 2-tile for $(\widehat{f}, \mathcal{C}^1)$ joins opposite sides of $\mathcal{C}^1$. Thus by Theorem 14.2 there is a homeomorphism $\phi \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ isotopic to the identity on $\widehat{\mathbb{C}}$ rel. $\text{post}(\widehat{f}) = \text{post}(f)$ such that $\phi(\mathcal{C}^1) = \mathcal{C}^1$ and $g = \phi \circ \widehat{f}$ is expanding. Since f is also expanding (as follows from Proposition 2.3) and g is Thurston equivalent to f, there is a homeomorphism $h \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ such that $h \circ f = g \circ h$ (Theorem 11.1). Then $\widetilde{\mathcal{C}} := h^{-1}(\mathcal{C}^1)$ is an f-invariant Jordan curve containing $\text{post}(f)$. The general existence result for invariant curves given by Theorem 15.4 is proved in the same way.

For the second approach we use Proposition 11.3 to lift $H = H^0$ by the map f to an isotopy H^1 with $H^1_0 = \mathrm{id}_{\widehat{\mathbb{C}}}$. Then we lift H^1 to an isotopy H^2 with $H^2_0 = \mathrm{id}_{\widehat{\mathbb{C}}}$, etc. In this way, we find a sequence of isotopies H^n and inductively define $\mathcal{C}^{n+1} := H^n_1(\mathcal{C}^n)$. We will see in Proposition 15.20 that the sequence $\{\mathcal{C}^n\}$ of Jordan curves Hausdorff converges to an f-invariant Jordan curve $\widetilde{\mathcal{C}}$ containing all postcritical points of f as desired. This is illustrated in Figure 15.1; indeed, the invariant curve $\widetilde{\mathcal{C}}$ in this picture was obtained by approximating it by the curves $\mathcal{C}_n$ (as were the invariant curves in Figures 15.4, 15.6, and 15.7).

In our example the f-invariant Jordan curve $\widetilde{\mathcal{C}} \subset \widehat{\mathbb{C}}$ with $\mathrm{post}(f) \subset \widetilde{\mathcal{C}}$ is in fact *unique*. To see this, note that since $\#\,\mathrm{post}(f) = 3$, every such curve $\widetilde{\mathcal{C}}$ is isotopic rel. $\mathrm{post}(f)$ to the curve $\mathcal{C}^0$ chosen above. Thus we can find an isotopy $K \colon \widehat{\mathbb{C}} \times I \to \widehat{\mathbb{C}}$ rel. $\mathrm{post}(f)$ with $K_0 = \mathrm{id}_{\widehat{\mathbb{C}}}$ and $K_1(\widetilde{\mathcal{C}}) = \mathcal{C}^0$. By Proposition 11.3 we can lift K to an isotopy $\widetilde{K} \colon \widehat{\mathbb{C}} \times I \to \widehat{\mathbb{C}}$ rel. $f^{-1}(\mathrm{post}(f))$ with $\widetilde{K}_0 = \mathrm{id}_{\widehat{\mathbb{C}}}$ and $K_t \circ f = f \circ \widetilde{K}_t$ for $t \in I$. Then by Lemma 11.2 we have

$$\mathcal{C}' := \widetilde{K}_1(\widetilde{\mathcal{C}}) \subset \widetilde{K}_1(f^{-1}(\widetilde{\mathcal{C}})) = f^{-1}(K_1(\widetilde{\mathcal{C}})) = f^{-1}(\mathcal{C}^0).$$

So $\mathcal{C}'$ is a Jordan curve in $\widehat{\mathbb{C}}$ with $\mathcal{C}' \subset f^{-1}(\mathcal{C}^0)$ and $\mathrm{post}(f) \subset \mathcal{C}'$. Since in this particular example $\mathcal{C}^1$ is the unique such curve, we conclude $\mathcal{C}' = \widetilde{K}_1(\widetilde{\mathcal{C}}) = \mathcal{C}^1$. In particular, $\widetilde{\mathcal{C}}$ is isotopic to $\mathcal{C}^1$ rel. $f^{-1}(\mathrm{post}(f))$ by the isotopy $\widetilde{K}$. So every f-invariant Jordan curve $\widetilde{\mathcal{C}}$ with $\mathrm{post}(f) \subset \widetilde{\mathcal{C}}$ lies in the same isotopy class rel. $f^{-1}(\mathrm{post}(f))$ as $\mathcal{C}^1$. Hence by Theorem 15.5 (which we will prove momentarily) there is at most one such Jordan curve $\widetilde{\mathcal{C}}$. The uniqueness of $\widetilde{\mathcal{C}}$ follows.

In Example 15.17 the reader can find another illustration for the construction of an invariant curve (see Figure 15.4).

15.1. Existence and uniqueness of invariant curves

We now turn to establishing existence and uniqueness results for general Thurston maps. We start with uniqueness results.

PROOF OF THEOREM 15.5. Suppose $f \colon S^2 \to S^2$ is an expanding Thurston map, and $\mathcal{C}$ and $\mathcal{C}'$ are f-invariant Jordan curves in S^2 that both contain the set $\mathrm{post}(f)$ and are isotopic rel. $f^{-1}(\mathrm{post}(f))$. We have to show that $\mathcal{C} = \mathcal{C}'$.

Under the given assumptions, there exists an isotopy $H^0 \colon S^2 \times I \to S^2$ rel. $f^{-1}(\mathrm{post}(f))$ with $H^0_0 = \mathrm{id}_{S^2}$ and $H^0_1(\mathcal{C}) = \mathcal{C}'$. Since $\mathrm{post}(f) \subset f^{-1}(\mathrm{post}(f))$, the map H^0 is also an isotopy rel. $\mathrm{post}(f)$.

Hence by Proposition 11.3 we can find an isotopy $H^1\colon S^2\times I\to S^2$ rel. $f^{-1}(\mathrm{post}(f))$ with $H_0^1=\mathrm{id}_{S^2}$ and $f\circ H_t^1=H_t^0\circ f$ for $t\in I$. Repeating this argument, we obtain isotopies $H^n\colon S^2\times I\to S^2$ rel. $f^{-1}(\mathrm{post}(f))$ with $H_0^n=\mathrm{id}_{S^2}$ and $f\circ H_t^{n+1}=H_t^n\circ f$ for $t\in I$ and $n\in\mathbb{N}_0$.

Claim. $H_1^n(\mathcal{C})=\mathcal{C}'$ for $n\in\mathbb{N}_0$.

To see this, we use induction on n. For $n=0$ the claim is true by choice of H^0.

Suppose that $H_1^n(\mathcal{C})=\mathcal{C}'$ for some $n\in\mathbb{N}_0$. Then Lemma 11.2 and the identity $f\circ H_1^{n+1}=H_1^n\circ f$ imply that

$$H_1^{n+1}(f^{-1}(\mathcal{C}))=f^{-1}(H_1^n(\mathcal{C}))=f^{-1}(\mathcal{C}').$$

Since $\mathcal{C}$ and $\mathcal{C}'$ are f-invariant, we have the inclusions $\mathcal{C}\subset f^{-1}(\mathcal{C})$ and $\mathcal{C}'\subset f^{-1}(\mathcal{C}')$. In particular,

$$\widetilde{\mathcal{C}}:=H_1^{n+1}(\mathcal{C})\subset H_1^{n+1}(f^{-1}(\mathcal{C}))=f^{-1}(\mathcal{C}')$$

is a Jordan curve contained in $f^{-1}(\mathcal{C}')$. Moreover, $\mathcal{C}$ and $\widetilde{\mathcal{C}}$ are isotopic rel. $f^{-1}(\mathrm{post}(f))$ (by the isotopy H^{n+1}). Since $\mathcal{C}$ and $\mathcal{C}'$ are isotopic rel. $f^{-1}(\mathrm{post}(f))$ by our hypotheses, it follows that $\mathcal{C}'$ and $\widetilde{\mathcal{C}}$ are also isotopic rel. $f^{-1}(\mathrm{post}(f))$. Both sets are contained in $f^{-1}(\mathcal{C}')$.

Now $f^{-1}(\mathcal{C}')$ is the 1-skeleton of the cell decomposition $\mathcal{D}^1(f,\mathcal{C}')$. This cell decomposition has the vertex set $f^{-1}(\mathrm{post}(f))$. Moreover, since f is expanding, $\#\,\mathrm{post}(f)\geq 3$, and so every tile in $\mathcal{D}^1(f,\mathcal{C}')$ has at least three vertices. So the hypotheses of Lemma 11.12 are satisfied and we conclude that $\mathcal{C}'=\widetilde{\mathcal{C}}=H_1^{n+1}(\mathcal{C})$. The claim above follows.

Now fix a visual metric on S^2. Then the tracks of the isotopies H^n shrink at an exponential rate as $n\to\infty$ (Lemma 11.4). Since $H_0^n=\mathrm{id}_{S^2}$, it follows that $H_1^n\to\mathrm{id}_{S^2}$ uniformly as $n\to\infty$. Since $H_1^n(\mathcal{C})=\mathcal{C}'$ for all $n\in\mathbb{N}_0$ by the claim, we conclude $\mathcal{C}=\mathcal{C}'$ as desired. $\qquad\square$

COROLLARY 15.7 (Invariant curves rel. $\mathrm{post}(f)$). *Let $f\colon S^2\to S^2$ be an expanding Thurston map and $\mathcal{C}\subset S^2$ be a Jordan curve with $\mathrm{post}(f)\subset\mathcal{C}$. Then there are at most finitely many f-invariant Jordan curves $\widetilde{\mathcal{C}}\subset S^2$ with $\mathrm{post}(f)\subset\widetilde{\mathcal{C}}$ that are isotopic to $\mathcal{C}$ rel. $\mathrm{post}(f)$.*

PROOF. Let $\widetilde{\mathcal{C}}$ be such an f-invariant Jordan curve. Then there exists an isotopy $H\colon S^2\times I\to S^2$ rel. $\mathrm{post}(f)$ with $H_0=\mathrm{id}_{S^2}$ and $H_1(\widetilde{\mathcal{C}})=\mathcal{C}$. Lifting H we get an isotopy $\widetilde{H}\colon S^2\times I\to S^2$ rel. $f^{-1}(\mathrm{post}(f))$ such that $\widetilde{H}_0=\mathrm{id}_{S^2}$ and $f\circ\widetilde{H}_t=H_t\circ f$ for $t\in I$. Since $\widetilde{\mathcal{C}}$ is f-invariant, we have $\widetilde{\mathcal{C}}\subset f^{-1}(\widetilde{\mathcal{C}})$. So Lemma 11.2 implies that

$$\widetilde{H}_1(\widetilde{\mathcal{C}})\subset\widetilde{H}_1(f^{-1}(\widetilde{\mathcal{C}}))=f^{-1}(H_1(\widetilde{\mathcal{C}}))=f^{-1}(\mathcal{C}).$$
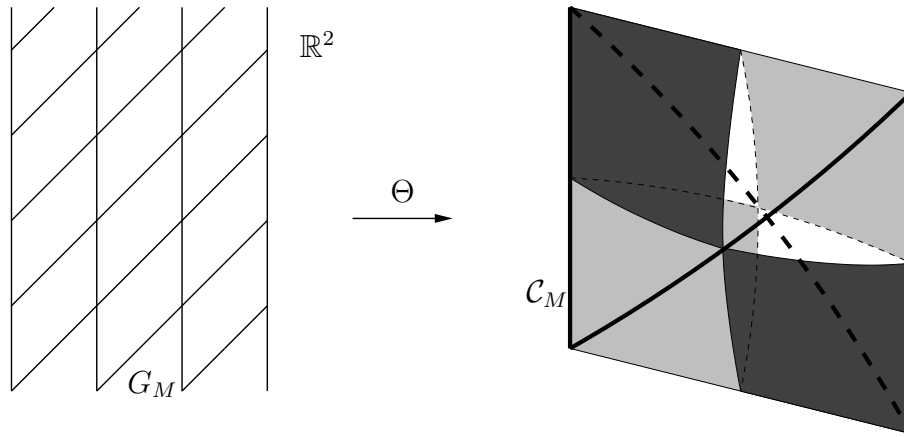
FIGURE 15.2. Invariant curves for the Lattès map g.

Hence $\widetilde{\mathcal{C}}$ is isotopic rel. $f^{-1}(\text{post}(f))$ to the Jordan curve $\widetilde{H}_1(\widetilde{\mathcal{C}})$ that is contained in $f^{-1}(\mathcal{C})$. Any such Jordan curve is a union of edges in the cell decomposition $\mathcal{D}^1(f, \mathcal{C})$ (see the last part of the proof of Lemma 11.12). In particular, there are only finitely many distinct Jordan curves contained in $f^{-1}(\mathcal{C})$. This implies that there are only finitely many isotopy classes rel. $f^{-1}(\text{post}(f))$ represented by curves $\widetilde{\mathcal{C}}$ satisfying the assumptions of the corollary. Since an f-invariant Jordan curve $\widetilde{\mathcal{C}} \subset S^2$ with $\text{post}(f) \subset \widetilde{\mathcal{C}}$ is unique in its isotopy class rel. $f^{-1}(\text{post}(f))$ by Theorem 15.5, the statement follows. $\qquad\square$

COROLLARY 15.8. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map with $\text{post}(f) = 3$. Then there are at most finitely many f-invariant Jordan curves $\widetilde{\mathcal{C}} \subset S^2$ with $\text{post}(f) \subset \widetilde{\mathcal{C}}$.*

PROOF. Pick a Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) \subset \mathcal{C}$. Since we have $\#\text{post}(f) = 3$, by Lemma 11.10 every Jordan curve $\widetilde{\mathcal{C}} \subset S^2$ with $\text{post}(f) \subset \widetilde{\mathcal{C}}$ is isotopic to $\mathcal{C}$ rel. $\text{post}(f)$. The statement now follows from Corollary 15.7. $\qquad\square$

In contrast to the case $\#\text{post}(f) = 3$, expanding Thurston maps f with $\#\text{post}(f) \geq 4$ can have infinitely many distinct invariant curves.

EXAMPLE 15.9 (Infinitely many invariant curves). Let g be the Lattès map from Section 1.1. In the following, it is advantageous to use real notation and consider the maps A and Θ used in the definition of g as in (1.1) as maps on $\mathbb{R}^2$. Then $A(u) = 2u$ for $u \in \mathbb{R}^2$. Moreover, for $u_1, u_2 \in \mathbb{R}^2$ we have $\Theta(u_1) = \Theta(u_2)$ if and only if $u_2 = \pm u_1 + \gamma$ for $\gamma \in \Gamma := \mathbb{Z}^2$. Recall that the extended real line $\widehat{\mathbb{R}}$ (which is the equator of the pillow) is g-invariant and contains $\{-1, 0, 1, \infty\} = \text{post}(g) =$

$\Theta(\Gamma)$. Let $S = \partial Q$ be the boundary of the square $Q = [0, 1/2]^2 \subset \mathbb{R}^2$. We denote by G the square grid given as the union of the horizontal and vertical lines in $\mathbb{R}^2$ that pass through a point in $\frac{1}{2}\Gamma = \frac{1}{2}\mathbb{Z}^2$. Then $\Theta|S$ is injective and $\Theta(S) = \Theta(G) = \widehat{\mathbb{R}}$. So the g-invariant curve $\widehat{\mathbb{R}}$ is the image of G under Θ. One can obtain other g-invariant Jordan curves by mapping other grids by Θ.

To explain this, we consider a (2×2)-matrix $M \in \mathrm{SL}_2(\mathbb{Z})$ (here $\mathrm{SL}_2(\mathbb{Z})$ denotes as usual the set of (2×2)-matrices with integer entries and determinant 1). We identify M with the linear map $u \mapsto Mu$ on $\mathbb{R}^2$ induced by left-multiplication of $u \in \mathbb{R}^2$ (considered as a column vector) by the matrix M. Define $Q_M = M(Q)$, $S_M = \partial Q_M = M(S)$, and the corresponding grid $G_M = M(G)$. Since M is a linear map and $M(\Gamma) = \Gamma$, it follows that for $u_1, u_2 \in \mathbb{R}^2$ we have $\Theta(M(u_1)) = \Theta(M(u_2))$ if and only if $u_2 = \pm u_1 + \gamma$, where $\gamma \in \Gamma$. This implies that $\Theta|S_M$ is injective, and so $\mathcal{C}_M := \Theta(S_M) \subset \widehat{\mathbb{C}}$ is a Jordan curve. Moreover, $\Theta(G_M) = \Theta(S_M)$. Since $A \circ M = M \circ A$, we have

$$A(G_M) = A(M(G)) = M(A(G)) \subset M(G) = G_M,$$

and so

$$g(\mathcal{C}_M) = g(\Theta(G_M)) = \Theta(A(G_M)) \subset \Theta(G_M) = \mathcal{C}_M.$$

Hence $\mathcal{C}_M$ is g-invariant. Since $\Gamma = M(\Gamma) \subset G_M$, we also have $\mathrm{post}(g) = \Theta(\Gamma) \subset \Theta(G_M) = \mathcal{C}_M$. So $\mathcal{C}_M$ is an g-invariant Jordan curve that contains the set $\mathrm{post}(g)$. An example of this construction is indicated in Figure 15.2. The curve $\mathcal{C}_M$ is drawn in thick on the right.

The curve $\mathcal{C}_M$ determines the grid G_M uniquely; indeed, one obtains generating vectors of the two lines in G_M through 0 by locally lifting $\mathcal{C}_M$ near $\Theta(0) = 0 \in \mathrm{post}(g) \subset \mathcal{C}_M$ to 0 by the map Θ. The whole grid G_M is obtained by translating these two lines by vectors in Γ.

This implies that the map $M \in \mathrm{SL}_2(\mathbb{Z}) \mapsto \mathcal{C}_M$ is four-to-one; indeed, if $M, N \in \mathrm{SL}_2(\mathbb{Z})$, then, as we have seen, $\mathcal{C}_M = \mathcal{C}_N$ if and only if $G_M = G_N$. On the other hand, $G_M = G_N$ if and only if $M^{-1} \circ N \in \mathrm{SL}_2(\mathbb{Z})$ is one of the four rotations around 0 (by integer multiples of $\pi/2$) that preserve the grid G. In particular, there exist infinitely many g-invariant Jordan curves $\widetilde{\mathcal{C}} \subset \widehat{\mathbb{C}}$ with $\mathrm{post}(g) \subset \widetilde{\mathcal{C}}$.

Note that by Proposition 3.21 the map M descends to a orientation-preserving homeomorphism $h \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ such that $h \circ \Theta = \Theta \circ M$. It easily follows from the above considerations that h is an *automorphism* of g in the sense that $g \circ h = h \circ g$. So our map g (as each flexible Lattès map) has a large associated group formed by these automorphisms. This is the deeper underlying reason why infinitely many g-invariant Jordan curves exist.

The map g in the previous example is very special, since it is a flexible Lattès map. In contrast, we have the following result (pointed out to us by K. Pilgrim).

THEOREM 15.10. *Let $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a rational Thurston map. Suppose that f is expanding and has a hyperbolic orbifold. Then there are at most finitely many f-invariant Jordan curves $\mathcal{C} \subset \widehat{\mathbb{C}}$ with $\operatorname{post}(f) \subset \mathcal{C}$.*

PROOF. If $\mathcal{C} \subset \widehat{\mathbb{C}}$ is an f-invariant Jordan curve with $\operatorname{post}(f) \subset \mathcal{C}$, then we have an associated two-tile subdivision rule $(\mathcal{D}^1, \mathcal{D}^0, L)$ accoding to Proposition 12.2. Recall that this means that $\mathcal{D}^1 = \mathcal{D}^1(f, \mathcal{C})$, $\mathcal{D}^0 = \mathcal{D}^0(f, \mathcal{C})$ and $L\colon \mathcal{D}^1 \to \mathcal{D}^0$ is the labeling induced by f (i.e., $L(\tau) = f(\tau) \in \mathcal{D}^0$ for $\tau \in \mathcal{D}^1$).

The number of cells in $\mathcal{D}^1$ is bounded by a constant only depending on $\deg(f)$ and $\#\operatorname{post}(f)$. In particular, we have a uniform bound independent of $\mathcal{C}$. This implies that among the two-tile subdivision rules obtained in such a way from f-invariant curves $\mathcal{C}$, there are only finitely many up to isomorphism (defined in the discussion before Lemma 12.13; see also Remark 12.14 (iii)). Here we use the strong notion of isomorphism where we require that the cell complex isomorphisms as in the definition of an isomorphism between two-tile subdivision rules send positively-oriented flags to positively-oriented flags (see Remark 12.14 (ii)).

This allows us to pick a finite family $\mathcal{F}$ of such curves $\mathcal{C}$ such that the associated two-tile subdivision rule of any f-invariant Jordan curve $\widetilde{\mathcal{C}} \subset \widehat{\mathbb{C}}$ with $\operatorname{post}(f) \subset \widetilde{\mathcal{C}}$ is isomorphic to one associated with a curve in $\mathcal{F}$.

Let $\mathcal{G}$ be the family of all Möbius transformations $\varphi\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ with $\varphi \circ f = f \circ \varphi$. If $\varphi \in \mathcal{G}$, then $\varphi(\operatorname{post}(f)) = \operatorname{post}(f)$. Now f is expanding and so $\#\operatorname{post}(f) \geq 3$. Since Möbius transformation are uniquely determined by images of three distinct points in $\widehat{\mathbb{C}}$, this implies that each $\varphi \in \mathcal{G}$ is uniquely determined by the bijection it induces on $\operatorname{post}(f)$. Since there are only finitely many such bijections, $\mathcal{G}$ consists of finitely many elements.

Now let $\widetilde{\mathcal{C}} \subset \widehat{\mathbb{C}}$ be an arbitrary f-invariant Jordan curve with $\operatorname{post}(f) \subset \widetilde{\mathcal{C}}$. We claim that $\widetilde{\mathcal{C}}$ is isotopic rel. $\operatorname{post}(f)$ to one of the finitely many Jordan curves $\varphi(\mathcal{C})$, where $\mathcal{C} \in \mathcal{F}$ and $\varphi \in \mathcal{G}$. Since each isotopy class rel. $\operatorname{post}(f)$ contains only finitely many f-invariant Jordan curves $\mathcal{C} \subset \widehat{\mathbb{C}}$ with $\operatorname{post}(f) \subset \mathcal{C}$ (Corollary 15.7), this claim implies the theorem.

To prove the claim, we use the fact that the two-tile subdivision rule associated with $\widetilde{\mathcal{C}}$ is isomorphic to one of the two-tile subdivision rules

associated with a curve $\mathcal{C} \in \mathcal{F}$. So by Lemma 12.13 and Remark 12.14 there exist orientation-preserving homeomorphisms $h_0, h_1 \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ that are isotopic rel. $\mathrm{post}(f)$ such $h_0 \circ f = f \circ h_1$ and $h_0(\mathcal{C}) = h_1(\mathcal{C}) = \widetilde{\mathcal{C}}$. By Thurston's uniqueness theorem (Theorem 2.19) there exists a Möbius transformation $\varphi \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ that is isotopic to h_0 rel. $\mathrm{post}(f)$ such that $\varphi \circ f = f \circ \varphi$. Then $\varphi \in \mathcal{G}$, and $\varphi(\mathcal{C})$ is isotopic to $h_0(\mathcal{C}) = \widetilde{\mathcal{C}}$ rel. $\mathrm{post}(f)$ as desired. $\qquad\square$

Using similar arguments as in the previous proof together with Theorem 11.1, one can show that for an expanding Thurston map $f \colon S^2 \to S^2$ with infinitely many invariant curves there are infinitely many homeomorphisms $h \colon S^2 \to S^2$ with $h \circ f = f \circ h$.

We now turn our attention to existence results. As the following example shows, for a Thurston map $f \colon S^2 \to S^2$ an f-invariant Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}$ need not exist.

EXAMPLE 15.11. Consider the map $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ defined by

$$f(z) = i \frac{z^4 - i}{z^4 + i}$$

for $z \in \widehat{\mathbb{C}}$. The critical points of f are 0 and ∞, and the map has the following ramification portrait:

(15.1)
$$
\begin{array}{c}
0 \xrightarrow{4:1} -i \searrow \\
\phantom{0 \xrightarrow{4:1}} \quad \circlearrowright 1. \\
\infty \xrightarrow{4:1} i \nearrow
\end{array}
$$

So the set of postcritical points of f is given by $\mathrm{post}(f) = \{-i, 1, i\}$, and f is a Thurston map. This map is also expanding as follows from Proposition 2.3.

LEMMA 15.12. *Let f be the map from Example 15.11. Then there is no f-invariant Jordan curve $\widetilde{\mathcal{C}} \subset \widehat{\mathbb{C}}$ with $\mathrm{post}(f) \subset \widetilde{\mathcal{C}}$.*

PROOF. We have $f(z) = \varphi(z^4)$ for $z \in \widehat{\mathbb{C}}$, where

(15.2)
$$\varphi(w) = i \frac{w - i}{w + i}, \quad w \in \widehat{\mathbb{C}},$$

is a Möbius transformation that maps the upper half-plane to the unit disk (note that φ maps $0, 1, \infty$ to $-i, 1, i$, respectively). Let $\mathcal{C} := \partial \mathbb{D}$ be the unit circle. Then

$$f^{-1}(\mathcal{C}) = \bigcup_{k=0,\dots,7} R_k, \text{ where } R_k = \{re^{ik\pi/4} : 0 \le r \le \infty\}.$$
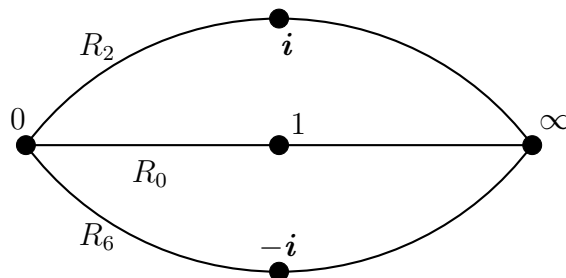
FIGURE 15.3. No invariant Jordan curve $\widetilde{\mathcal{C}} \supset \mathrm{post}(f)$.

The postcritical points $-\boldsymbol{i}, 1, \boldsymbol{i}$ lie on distinct rays R_k. Two such rays have the points 0 and ∞ in common and no other points. Thus there is no Jordan curve in $f^{-1}(\mathcal{C})$ containing all postcritical points (see Figure 15.3). As we will see in Theorem 15.4, the existence of such a Jordan curve is a necessary condition for the existence of an f-invariant Jordan curve $\widetilde{\mathcal{C}} \subset \widehat{\mathbb{C}}$ with $\mathrm{post}(f) \subset \widetilde{\mathcal{C}}$ (in our specific case, where $\#\,\mathrm{post}(f) = 3$, the choice of $\mathcal{C}$ does not matter, since all Jordan curves that contain $\mathrm{post}(f)$ are isotopic rel. $\mathrm{post}(f)$)). Hence there is no f-invariant Jordan curve $\widetilde{\mathcal{C}} \subset \widehat{\mathbb{C}}$ with $\mathrm{post}(f) \subset \widetilde{\mathcal{C}}$. One can also see this by a simple argument directly.

Indeed, suppose that $\widetilde{\mathcal{C}} \subset \widehat{\mathbb{C}}$ is a Jordan curve with $\mathrm{post}(f) \subset \widetilde{\mathcal{C}}$ and $f(\widetilde{\mathcal{C}}) \subset \widetilde{\mathcal{C}}$. The unit circle $\mathcal{C} = \partial \mathbb{D}$ is also a Jordan curve containing the set $\mathrm{post}(f)$. Hence by Lemma 11.10 there exists an isotopy $H\colon \widehat{\mathbb{C}} \times I \to \widehat{\mathbb{C}}$ rel. $\mathrm{post}(f)$ such that $H_0 = \mathrm{id}_{\widehat{\mathbb{C}}}$ and $H_1(\widetilde{\mathcal{C}}) = \mathcal{C}$.

By Proposition 11.3 the isotopy H can be lifted to an isotopy $\widetilde{H}\colon \widehat{\mathbb{C}} \times I \to \widehat{\mathbb{C}}$ rel. $\mathrm{post}(f)$ such that $\widetilde{H}_0 = \mathrm{id}_{\widehat{\mathbb{C}}}$ and $H_t \circ f = f \circ \widetilde{H}_t$ for $t \in I$.

Since $\widetilde{\mathcal{C}} \subset f^{-1}(\widetilde{\mathcal{C}})$, it follows from Lemma 11.2 that

$$\widetilde{H}_1(\widetilde{\mathcal{C}}) \subset \widetilde{H}_1(f^{-1}(\widetilde{\mathcal{C}})) = f^{-1}(H_1(\widetilde{\mathcal{C}})) = f^{-1}(\mathcal{C}).$$

This means that the Jordan curve $\mathcal{C}' := \widetilde{H}_1(\widetilde{\mathcal{C}})$ is contained in $f^{-1}(\mathcal{C})$. Moreover, it contains all postcritical points, since $\widetilde{\mathcal{C}}$ does, and the points in $\mathrm{post}(f)$ stay fixed under the isotopy $\widetilde{H}$. As we have seen above, no such Jordan curve exists and we get a contradiction as desired. $\square$

By a similar (though somewhat lengthier) argument one can show that the Lattès map $f(z) = \frac{i}{2}(z + 1/z)$ does not have an f-invariant Jordan curve $\mathcal{C}$ with $\mathrm{post}(f) \subset \mathcal{C}$ (this map was considered in Example 3.27). Another such example can be found in [**CFP10**, Section 4].

We now turn to the proof of the necessary and sufficient criterion for the existence of an invariant Jordan curve as formulated in Theorem 15.4. Note that in condition (ii) of this theorem the requirement on $\widehat{f}$ is meaningful. Indeed, $\widehat{f} = H_1 \circ f$ is a branched covering map of S^2 with $\mathrm{crit}(\widehat{f}) = \mathrm{crit}(f)$. Since H_1 is isotopic to id_{S^2} rel. $\mathrm{post}(f)$ we have $\mathrm{post}(\widehat{f}) = \mathrm{post}(f)$, and so $\widehat{f}$ is a Thurston map.

Furthermore, $\mathcal{C}'$ is a Jordan curve with $\mathrm{post}(\widehat{f}) = \mathrm{post}(f) \subset \mathcal{C}'$. Since $\mathcal{C}' = H_1(\mathcal{C}) \subset f^{-1}(\mathcal{C})$, we have that $\widehat{f}(\mathcal{C}') = (H_1 \circ f)(\mathcal{C}') \subset H_1(\mathcal{C}) = \mathcal{C}'$. Hence $\mathcal{C}'$ is invariant with respect to $\widehat{f}$, and it makes sense to require that $\widehat{f}$ is combinatorially expanding for $\mathcal{C}'$ (note that $\#\mathrm{post}(\widehat{f}) = \#\mathrm{post}(f) \geq 3$, because f is expanding).

PROOF OF THEOREM 15.4. (i) $\Rightarrow$ (ii) This implication is trivial. Indeed, suppose $\widetilde{\mathcal{C}}$ is as in (i). Then in (ii) we let $\mathcal{C} = \mathcal{C}' = \widetilde{\mathcal{C}}$, and the isotopy H be such that $H_t = \mathrm{id}_{S^2}$ for all $t \in I$. Then $\mathcal{C}' = \widetilde{\mathcal{C}} \subset f^{-1}(\widetilde{\mathcal{C}}) = f^{-1}(\mathcal{C})$, and $\widehat{f} = f$ is combinatorially expanding for the invariant curve $\mathcal{C}' = \widetilde{\mathcal{C}}$, because f is expanding.

(ii) $\Rightarrow$ (i) Let $\mathcal{C}$, $\mathcal{C}'$, H, $\widehat{f}$ be as in (ii), and define $\chi = H_1$. As we have seen in the discussion before the proof, $\widehat{f}$ is a Thurston map with $\mathrm{post}(\widehat{f}) = \mathrm{post}(f)$, and $\mathcal{C}'$ is an $\widehat{f}$-invariant Jordan curve containing the set $\mathrm{post}(\widehat{f}) = \mathrm{post}(f)$.

Since $\widehat{f}$ is combinatorially expanding for $\mathcal{C}'$, Theorem 14.2 implies that there exists a homeomorphism $\phi\colon S^2 \to S^2$ that is isotopic to the identity rel. $\mathrm{post}(\widehat{f}) = \mathrm{post}(f)$ such that $\phi(\mathcal{C}') = \mathcal{C}'$ and $g = \phi \circ \widehat{f}$ is an expanding Thurston map. Since $g = (\phi \circ \chi) \circ f$, and $\phi \circ \chi$ is isotopic to the identity on S^2 rel. $\mathrm{post}(f)$, the expanding Thurston maps f and g are Thurston equivalent: if notation is as in (2.8) (with $\widehat{S}^2 = S^2$), then we can take $h_0 = \phi \circ \chi$ and $h_1 = \mathrm{id}_{S^2}$. By Theorem 11.1 we can find a homeomorphism $h\colon S^2 \to S^2$ that is isotopic to $h_1 = \mathrm{id}_{S^2}$ rel. $f^{-1}(\mathrm{post}(f))$ with $h \circ f = g \circ h$. Note that then the homeomorphism h^{-1} is also isotopic to id_{S^2} rel. $f^{-1}(\mathrm{post}(f))$ and we have $f \circ h^{-1} = h^{-1} \circ g$.

Let $\widetilde{\mathcal{C}} = h^{-1}(\mathcal{C}')$. Then $\widetilde{\mathcal{C}}$ is a Jordan curve in S^2 that is isotopic to $\mathcal{C}'$ rel. $f^{-1}(\mathrm{post}(f))$, and hence isotopic to $\mathcal{C}$ rel. $\mathrm{post}(f)$; in particular, $\widetilde{\mathcal{C}}$ contains the set $\mathrm{post}(f)$. Moreover, $\widetilde{\mathcal{C}}$ is f-invariant, because we have

$$
\begin{aligned}
f(\widetilde{\mathcal{C}}) &= f(h^{-1}(\mathcal{C}')) = h^{-1}(g(\mathcal{C}')) = h^{-1}(\phi(\widehat{f}(\mathcal{C}'))) \\
&\subset h^{-1}(\phi(\mathcal{C}')) = h^{-1}(\mathcal{C}') = \widetilde{\mathcal{C}}.
\end{aligned}
$$

Here we used Lemma 11.2. The proof is complete. $\square$

REMARK 15.13. (i) The condition of combinatorial expansion in condition (ii) of Theorem 15.4 is combinatorial in nature and can easily be checked in principle. A simple sufficient criterion for this can be formulated as follows: if no 1-tile for $(f, \mathcal{C})$ joins opposite sides of $\mathcal{C}'$, then $\widehat{f}$ is combinatorially expanding for $\mathcal{C}'$.

To see this, note that

$$\widehat{f}^{-1}(\mathcal{C}') = f^{-1}(H_1^{-1}(\mathcal{C}')) = f^{-1}(\mathcal{C}).$$

By Proposition 5.17 (v) this implies that the 1-tiles for $(\widehat{f}, \mathcal{C}')$ are precisely the 1-tiles for $(f, \mathcal{C})$. Hence if no 1-tile for $(f, \mathcal{C})$ joins opposite sides of $\mathcal{C}'$, then $D_1(\widehat{f}, \mathcal{C}') \geq 2$ and so $\widehat{f}$ is combinatorially expanding for $\mathcal{C}'$. We will later formulate a necessary and sufficient condition for combinatorial expansion of $\widehat{f}$ (see Proposition 15.19; Example 15.23 illustrates the situation when this condition is not satisfied).

(ii) The condition of combinatorial expansion in condition (ii) of Theorem 15.4 is independent of the chosen isotopy H. More precisely, suppose $H^1, H^2 \colon S^2 \times I \to S^2$ are two isotopies rel. $\mathrm{post}(f)$ with $H_0^1 = H_0^2 = \mathrm{id}_{S^2}$ and $H_1^1(\mathcal{C}) = H_1^2(\mathcal{C}) = \mathcal{C}'$. Then $\widehat{f}_1 = H_1^1 \circ f$ is combinatorially expanding for $\mathcal{C}'$ if and only if $\widehat{f}_2 = H_1^2 \circ f$ is combinatorially expanding for $\mathcal{C}'$. This follows immediately from Lemma 12.16 (with $f = \widehat{f}_1$, $g = \widehat{f}_2$, $h_0 = H_1^2 \circ (H_1^1)^{-1}$, $h_1 = \mathrm{id}_{S^2}$, and $\mathcal{C} = \mathcal{C}'$).

(iii) Theorem 15.4 can be slightly modified to give necessary and sufficient conditions for the existence of an invariant curve in a given isotopy class rel. $\mathrm{post}(f)$ or rel. $f^{-1}(\mathrm{post}(f))$. An existence statement for a given isotopy class rel. $f^{-1}(\mathrm{post}(f))$ is especially relevant in view of the complementary uniqueness statement given by Theorem 15.5.

To formulate this precisely, let $\widehat{\mathcal{C}} \subset S^2$ be a given Jordan curve with $\mathrm{post}(f) \subset \widehat{\mathcal{C}}$. Then an f-invariant Jordan curve $\widetilde{\mathcal{C}} \subset S^2$ *isotopic to $\widehat{\mathcal{C}}$ rel.* $\mathrm{post}(f)$ exists if and only if condition (ii) in Theorem 15.4 is true for a Jordan curve $\mathcal{C}$ *isotopic to $\widehat{\mathcal{C}}$ rel.* $\mathrm{post}(f)$. This immediately follows from the proof of this theorem.

Similarly, an f-invariant Jordan curve $\widetilde{\mathcal{C}} \subset S^2$ *isotopic to $\widehat{\mathcal{C}}$ rel.* $f^{-1}(\mathrm{post}(f))$ exists if and only if condition (ii) in Theorem 15.4 is true with the extra assumption that $\mathcal{C}'$ *is isotopic to $\widehat{\mathcal{C}}$ rel.* $f^{-1}(\mathrm{post}(f))$.

The proof of the implication (ii) $\Rightarrow$ (i) in Theorem 15.4 does not only give the existence of an f-invariant Jordan curve $\widetilde{\mathcal{C}}$, but $\widetilde{\mathcal{C}}$ is constructed quite explicitly from the given Jordan curves $\mathcal{C}$ and $\mathcal{C}'$. So one can actually say more about the combinatorial description of f in terms of $\widetilde{\mathcal{C}}$, or, more precisely, about the two-tile subdivision rule

that is given by $\widetilde{\mathcal{C}}$ according to Proposition 12.2. Namely, the 1-tiles for $(f, \mathcal{C})$ subdivide the two 0-tiles defined by $(f, \mathcal{C}')$ in the same way as the 1-tiles for $(f, \widetilde{\mathcal{C}})$ subdivide the 0-tiles for $(f, \widetilde{\mathcal{C}})$. This is made precise in the following statement.

COROLLARY 15.14. *Let $\widetilde{\mathcal{C}}$ be the f-invariant Jordan curve obtained in the proof of Theorem 15.4 from the Jordan curves $\mathcal{C}, \mathcal{C}' \subset S^2$ as in condition* (ii) *in Theorem 15.4. Then there is a homeomorphism $h \colon S^2 \to S^2$ that is isotopic to id_{S^2} rel. $f^{-1}(\mathrm{post}(f))$ with the following properties: $h(\mathrm{post}(f)) = \mathrm{post}(f)$, $h(\widetilde{\mathcal{C}}) = \mathcal{C}'$, and h maps the 1-cells for $(f, \widetilde{\mathcal{C}})$ to the 1-cells for $(f, \mathcal{C})$.*

Here the map h is in fact the map that appears in the proof of Theorem 15.4. Recall that a 1-cell is a 1-tile, 1-edge, or a singleton set $\{v\}$ where v is a 1-vertex.

To illustrate the statement, let us consider the map f from Example 15.6. In the top right image of Figure 15.1 we can see that one of the two 0-tiles for $\mathcal{C}^1 \subset f^{-1}(\mathcal{C})$ is subdivided into four 1-tiles for $(f, \mathcal{C})$, and the other 0-tile into two 1-tiles for $(f, \mathcal{C})$. Thus the corollary above shows that the two 0-tiles for the (unique) f-invariant curve $\widetilde{\mathcal{C}}$ are subdivided into four or into two 1-tiles for $(f, \widetilde{\mathcal{C}})$.

PROOF. We will use the notation from the proof of the implication (ii) $\Rightarrow$ (i) in Theorem 15.4. So we have homeomorphisms $\chi, \phi, h \colon S^2 \to S^2$ that satisfy

$$\chi(\mathcal{C}) = \mathcal{C}', \quad \phi(\mathcal{C}') = \mathcal{C}', \quad h(\widetilde{\mathcal{C}}) = \mathcal{C}'.$$

The map h is isotopic to id_{S^2} rel. $f^{-1}(\mathrm{post}(f))$, and so it fixes the points in $f^{-1}(\mathrm{post}(f)) \supset \mathrm{post}(f)$.

The map $g = \phi \circ \chi \circ f$ is a Thurston map conjugate to f so that $h \circ f = g \circ h$. Lemma 11.2 implies that $h(f^{-1}(A)) = g^{-1}(h(A))$ for each set $A \subset S^2$. Hence

$$h(f^{-1}(\widetilde{\mathcal{C}})) = g^{-1}(h(\widetilde{\mathcal{C}})) = g^{-1}(\mathcal{C}') = f^{-1}(\chi^{-1}(\phi^{-1}(\mathcal{C}')))$$
$$= f^{-1}(\chi^{-1}(\mathcal{C}')) = f^{-1}(\mathcal{C}).$$

So by Proposition 5.17 (iii) the homeomorphism h maps the 1-skeletons of the cell decompositions $\mathcal{D}^1(f, \widetilde{\mathcal{C}})$ and $\mathcal{D}^1(f, \mathcal{C})$ (see Definition 5.14) onto each other. The vertices of both cell decompositions are the points in $f^{-1}(\mathrm{post}(f))$ which are fixed by h. Since these cell decompositions are uniquely determined by their 1-skeletons and vertices (see Proposition 5.17 (v)), h map the cells in $\mathcal{D}^1(f, \widetilde{\mathcal{C}})$ to the cells in $\mathcal{D}^1(f, \mathcal{C})$. Hence h maps 1-cells for $(f, \widetilde{\mathcal{C}})$ to 1-cells for $(f, \mathcal{C})$. $\qquad\square$

For the proof of Theorem 15.1 we require the following auxiliary result.

LEMMA 15.15. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map, and $\mathcal{C} \subset S^2$ be a Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}$. Then for all sufficiently large n there exists a Jordan curve $\mathcal{C}' \subset f^{-n}(\mathcal{C})$ that is isotopic to $\mathcal{C}$ rel. $\mathrm{post}(f)$. Moreover, $\mathcal{C}'$ can be chosen so that no n-tile for $(f,\mathcal{C})$ joins opposite sides of $\mathcal{C}'$.*

PROOF. We fix some base metric on S^2. Let $P \coloneqq \mathrm{post}(f)$. Since f is expanding, we have $k \coloneqq \#P = \#\mathrm{post}(f) \geq 3$ by Corollary 7.2. Pick $\epsilon_0 > 0$ as in Lemma 11.17. Since f is expanding, for large enough n we have

$$\mathrm{mesh}(f,n,\mathcal{C}) = \max_{c \in \mathcal{D}^n(f,\mathcal{C})} \mathrm{diam}(c) < \epsilon_0.$$

For such n consider the cell decomposition $\mathcal{D} = \mathcal{D}^n(f,\mathcal{C})$ of S^2. Its vertex set is the set $f^{-n}(\mathrm{post}(f)) \supset \mathrm{post}(f) = P$ of n-vertices and its 1-skeleton is the set $f^{-n}(\mathcal{C})$. Hence by Lemma 11.17 there exists a Jordan curve $\mathcal{C}' \subset f^{-n}(\mathcal{C})$ that is isotopic to $\mathcal{C}$ rel. $P = \mathrm{post}(f)$ and so that no tile in $\mathcal{D}$, i.e., no n-tile for $(f,\mathcal{C})$, joins opposite side of $\mathcal{C}'$. □

PROOF OF THEOREM 15.1. Let f and $\mathcal{C}$ be as in the statement of the theorem. By Lemma 15.15 for sufficiently large $n \in \mathbb{N}$ there exists an isotopy $H\colon S^2 \times I \to S^2$ rel. $\mathrm{post}(f)$ such that $H_0 = \mathrm{id}_{S^2}$ and $\mathcal{C}' \coloneqq H_1(\mathcal{C}) \subset f^{-n}(\mathcal{C})$ and such that no n-tile for $(f,\mathcal{C})$ joins opposite sides of $\mathcal{C}'$.

If we define $F = f^n$ for such n, then the map F is an expanding Thurston map with $\mathrm{post}(F) = \mathrm{post}(f)$. The sets $\mathcal{C}$ and $\mathcal{C}'$ are Jordan curves with $\mathrm{post}(F) \subset \mathcal{C}, \mathcal{C}'$, and H is an isotopy rel. $\mathrm{post}(F)$ that deforms $\mathcal{C}$ into $\mathcal{C}' \subset f^{-n}(\mathcal{C}) = F^{-1}(\mathcal{C})$. By Proposition 5.17 (vii) the 1-cells for $(F,\mathcal{C})$ are precisely the n-cells for $(f,\mathcal{C})$. So no 1-tile for $(F,\mathcal{C})$ joins opposite side of $\mathcal{C}'$ and by Remark 15.13 (i) the map $H_1 \circ F$ is combinatorially expanding for $\mathcal{C}'$. This shows that condition (ii) in Theorem 15.4 is satisfied. Hence there exists a Jordan curve $\widetilde{\mathcal{C}} \subset S^2$ that is F-invariant and isotopic to $\mathcal{C}$ rel. $\mathrm{post}(F) = \mathrm{post}(f)$ as desired. □

REMARK 15.16. In general, the f^n-invariant Jordan curve $\widetilde{\mathcal{C}}$ as in Theorem 15.1 will depend on n, and one cannot expect that $\widetilde{\mathcal{C}}$ is invariant for *all* sufficiently high iterates of f. To illustrate this, consider the map f from Example 15.11 (see also Lemma 15.12). Recall that $f(z) = \varphi(z^4)$ for $z \in \widehat{\mathbb{C}}$, where φ is as in (15.2).

The Möbius transformation φ maps the extended real line $\widehat{\mathbb{R}}$ to the unit circle $\partial \mathbb{D}$, and $\partial \mathbb{D}$ to $\widehat{\mathbb{R}}$. This implies that the unit circle $\widetilde{\mathcal{C}} \coloneqq \partial \mathbb{D}$

satisfies $f^{2n}(\widetilde{\mathcal{C}}) \subset \widetilde{\mathcal{C}}$ for every $n \in \mathbb{N}$. Note that $\mathrm{post}(f) = \{-\boldsymbol{i}, 1, \boldsymbol{i}\} \subset \widetilde{\mathcal{C}}$. Thus $\widetilde{\mathcal{C}}$ is a Jordan curve with $\mathrm{post}(f) \subset \widetilde{\mathcal{C}}$ that is invariant for every *even* iterate f^{2n}.

On the other hand, for $n \in \mathbb{N}_0$ we have $f^{2n+1}(\partial \mathbb{D}) \subset \widehat{\mathbb{R}}$, and so we cannot have $f^{2n+1}(\partial \mathbb{D}) \subset \partial \mathbb{D}$ (for otherwise, $f^{2n+1}(\partial \mathbb{D}) \subset \partial \mathbb{D} \cap \widehat{\mathbb{R}} = \{-1, 1\}$). Thus the unit circle $\partial \mathbb{D} = \widetilde{\mathcal{C}}$ is not invariant for any *odd* iterate of f.

PROOF OF COROLLARY 15.2. Let $f\colon S^2 \to S^2$ be an expanding Thurston map. It follows from Theorem 15.1 that for each sufficiently large $n \in \mathbb{N}$ there exists an f^n-invariant Jordan curve $\widetilde{\mathcal{C}} \subset S^2$ with $\mathrm{post}(f) = \mathrm{post}(f^n) \subset \widetilde{\mathcal{C}}$. For such n let $F = f^n$. By Proposition 12.2 there exists a two-tile subdivision rule that is realized by F. $\square$

15.2. Iterative construction of invariant curves

Given data as in Theorem 15.4 (ii), the f-invariant curve $\widetilde{\mathcal{C}}$ can be obtained by an iterative procedure. To explain this, let $f\colon S^2 \to S^2$ be an arbitrary Thurston map, and assume as in Theorem 15.4 (ii) that $\mathcal{C}, \mathcal{C}' \subset S^2$ are Jordan curves with $\mathrm{post}(f) \subset \mathcal{C}, \mathcal{C}'$ and $\mathcal{C}' \subset f^{-1}(\mathcal{C})$, and that $H\colon S^2 \times I \to S^2$ is an isotopy rel. $\mathrm{post}(f)$ that deforms $\mathcal{C}$ to $\mathcal{C}'$, i.e., $H_0 = \mathrm{id}_{S^2}$ and $H_1(\mathcal{C}) = \mathcal{C}'$. For the moment, we do *not* assume that the map f is expanding or that $\widehat{f} = H_1 \circ f$ is combinatorially expanding for $\mathcal{C}'$.

Let $H^0 := H$. Using Proposition 11.3 repeatedly, we can find isotopies $H^n\colon S^2 \times I \to S^2$ rel. $f^{-1}(\mathrm{post}(f))$ such that $H_0^n = \mathrm{id}_{S^2}$ and $f \circ H_t^n = H_t^{n-1} \circ f$ for all $n \in \mathbb{N}$, $t \in I$. Now define Jordan curves inductively by setting $\mathcal{C}^0 := \mathcal{C}$, and $\mathcal{C}^{n+1} := H_1^n(\mathcal{C}^n)$ for $n \in \mathbb{N}_0$. Note that then $\mathcal{C}^1 = \mathcal{C}'$.

To summarize, we start with the following data for our given Thurston map f:

(i) a Jordan curve $\mathcal{C}^0 = \mathcal{C} \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}^0$,

(ii) a Jordan curve $\mathcal{C}^1 = \mathcal{C}' \subset S^2$ isotopic to $\mathcal{C}^0 \subset S^2$ rel. $\mathrm{post}(f)$ with $\mathcal{C}^1 \subset f^{-1}(\mathcal{C}^0)$,

(iii) an isotopy $H^0\colon S^2 \times I \to S^2$ rel. $\mathrm{post}(f)$ such that $H_0^0 = \mathrm{id}_{S^2}$ and $H_1^0(\mathcal{C}^0) = \mathcal{C}^1$.

We then define inductively:

(i) isotopies $H^n\colon S^2 \times I \to S^2$ such that $H_0^n = \mathrm{id}_{S^2}$ and $f \circ H_t^n = H_t^{n-1} \circ f$ for all $n \in \mathbb{N}$, $t \in I$,

(ii) Jordan curves $\mathcal{C}^{n+1} := H_1^n(\mathcal{C}^n)$ for $n \in \mathbb{N}_0$.

Figure 15.1 illustrates this procedure for Example 15.6. Since this example is rather complicated and it is hard to grasp the involved isotopies, we present a simpler example for the construction.

EXAMPLE 15.17. Let $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a Lattès map constructed as the example g in Section 1.1, but with the map

$$A\colon \mathbb{C} \to \mathbb{C}, \quad u \mapsto A(u) \coloneqq 5u.$$

More precisely, f is obtained according to Theorem 3.1 (ii) as the quotient of A by a crystallographic group of type (2222) as in (3.24). It is straightforward to check that the extended real line $\mathcal{C} \coloneqq \widehat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ is f-invariant and contains all postcritical points $0, 1, \infty, -1$ of f.

As in Figure 1.1, we represent the sphere $\widehat{\mathbb{C}}$ as a pillow, i.e., two squares glued together along their boundaries. The equator of the pillow represents the curve $\mathcal{C}$, and the two squares represent the 0-tiles, one of which is colored white, the other black.

The map f can then be described as follows. Each of the two sides of the pillow is divided into 5×5 squares, which are colored in a checkerboard fashion. The map f acts by mapping each small white square to the white side of the pillow, and each small black square to the black side. The two sides of the pillow are the 0-tiles with respect to $\mathcal{C}$; the 4 vertices of the pillow are the postcritical points in this model. The small squares are the 1-tiles (for $(f, \mathcal{C})$). The coloring of the 0- and 1-tiles corresponds to a labeling map $L_{\mathbf{X}}$ as in Lemma 5.23.

There exist f-invariant Jordan curves that are isotopic to $\mathcal{C}$ rel. $\mathrm{post}(f)$, but distinct from $\mathcal{C}$. The construction of one such curve is illustrated in Figure 15.4. Namely, we set $\mathcal{C}^0 \coloneqq \mathcal{C}$. The Jordan curve $\mathcal{C}^1$ is shown on the top right, as well as in the middle left picture. In the latter picture, we see that $\mathcal{C}^1$ consists of 1-edges, i.e., $\mathcal{C}^1 \subset f^{-1}(\mathcal{C}^0)$. Moreover, there exists an isotopy $H^0\colon \widehat{\mathbb{C}} \times I \to \widehat{\mathbb{C}}$ rel. $\mathrm{post}(f)$ that deforms $\mathcal{C}^0$ to $\mathcal{C}^1$ (i.e., $H_0^0 = \mathrm{id}_{\widehat{\mathbb{C}}}$ and $H_1^0(\mathcal{C}^0) = \mathcal{C}^1$). We also see here how the black and the white 0-tile are deformed by H_1^0; namely, the four small black squares on the top right of Figure 15.4 are part of the image of the black 0-tile (which is at the back of the pillow) under H_1^0.

The Jordan curve $\mathcal{C}^2 \coloneqq H_1^1(\mathcal{C}^1)$ consists of 2-edges, i.e., $\mathcal{C}^2 \subset f^{-2}(\mathcal{C}^0)$ (see the bottom left).

The two pictures in the middle of Figure 15.4 indicate how H^1 deforms 1-tiles. Roughly speaking, H^1 deforms each black/white 1-tile "in the same way" as H^0 deforms the black/white 0-tiles.

The curves $\mathcal{C}^n$ Hausdorff converge to $\widetilde{\mathcal{C}}$, which is an f-invariant Jordan curve with $\mathrm{post}(f) \subset \widetilde{\mathcal{C}}$ (see Lemma 15.18 (viii) and Proposition 15.20).

$$\mathcal{C}^0 \qquad \xrightarrow{\;H_1^0\;} \qquad \mathcal{C}^1$$

$$\mathcal{C}^1 \qquad \xrightarrow{\;H_1^1\;} \qquad \mathcal{C}^2$$

$$\mathcal{C}^2 \qquad \cdots \qquad \widetilde{\mathcal{C}}$$
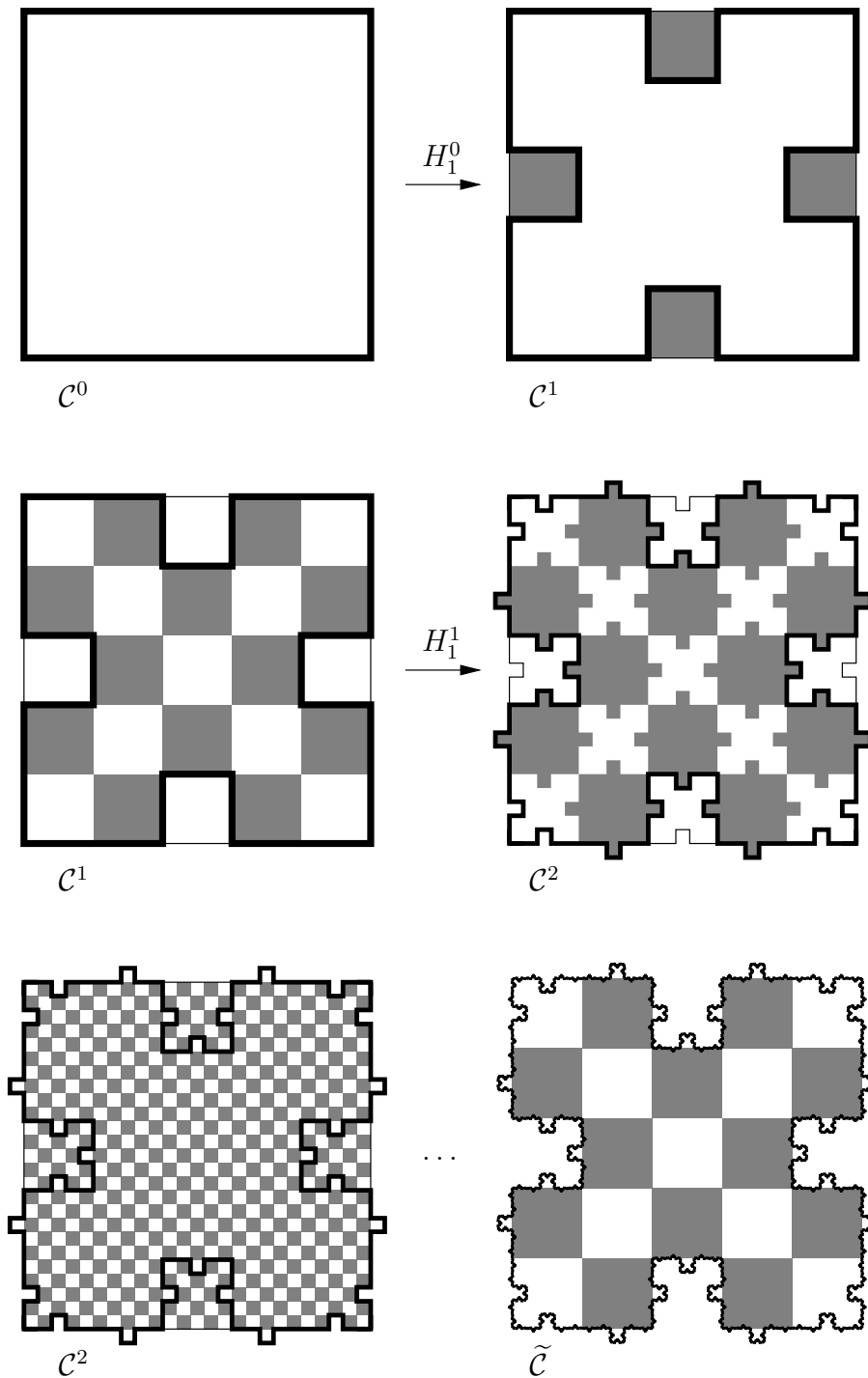
FIGURE 15.4. Iterative construction of an invariant curve.

There is a conceptually different way to obtain $\mathcal{C}^{n+1}$ from $\mathcal{C}^n$, which will be explained in detail in Remark 15.22. Namely, we replace each n-edge $\alpha^n \subset \mathcal{C}^n$ by $(n + 1)$-edges "in the same way" as the 0-edge $\alpha^0 := f^n(\alpha^n) \subset \mathcal{C}^0$ is replaced by the arc $\beta^1 \subset \mathcal{C}^1$ with the same endpoints as α^0 (which are postcritical points). Note that $\beta^1 = H_1^0(\alpha^0)$, and that β^1 consists of 1-edges.

To prepare the proof that under suitable conditions our iteration process has an invariant Jordan curve as a limit in the sense of Hausdorff convergence, we summarize some properties of the Jordan curves $\mathcal{C}^n$.

LEMMA 15.18. *Let $f\colon S^2 \to S^2$ be a Thurston map that satisfies $\#\operatorname{post}(f) \geq 3$, and let the Jordan curves $\mathcal{C}^n$ for $n \in \mathbb{N}_0$ be defined as above. Then the following statements are true:*

(i) $\mathcal{C}^{n+k} \subset f^{-k}(\mathcal{C}^n)$ *for $n, k \in \mathbb{N}_0$.*

(ii) $\mathcal{C}^{n+k}$ *is isotopic to $\mathcal{C}^n$ rel. $f^{-n}(\operatorname{post}(f))$ for $n, k \in \mathbb{N}_0$.*

(iii) $\mathcal{C}^{n+k} \cap f^{-n}(\operatorname{post}(f)) = \mathcal{C}^n \cap f^{-n}(\operatorname{post}(f))$ *for $n, k \in \mathbb{N}_0$.*

(iv) $\operatorname{post}(f) \subset \mathcal{C}^n$ *for $n \in \mathbb{N}_0$.*

(v) *For $n, k \in \mathbb{N}_0$ the curve $\mathcal{C}^{n+k}$ consists of n-edges for $(f, \mathcal{C}^k)$.*

(vi) *For $n \in \mathbb{N}$ the curve $\mathcal{C}^n$ is the unique Jordan curve in S^2 with $\mathcal{C}^n \subset f^{-1}(\mathcal{C}^{n-1})$ that is isotopic to $\mathcal{C}^1$ rel. $f^{-1}(\operatorname{post}(f))$.*

(vii) *The sequence $\mathcal{C}^n$, $n \in \mathbb{N}_0$, only depends on $\mathcal{C}^0$ and $\mathcal{C}^1$ and not of the choice of the initial isotopy $H = H^0$ used in the definition of the sequence.*

(viii) *Suppose in addition that f is expanding. Then, as $n \to \infty$, the sets $\mathcal{C}^n$ Hausdorff converge to a closed f-invariant set $\widetilde{\mathcal{C}} \subset S^2$ with $\operatorname{post}(f) \subset \widetilde{\mathcal{C}}$.*

Recall that Hausdorff convergence was discussed at the end of Section 4.1.

PROOF. In the following, we use the isotopies H^n as in the definition of the sequence $\mathcal{C}^n$, and set $h_n = H_1^n$ for $n \in \mathbb{N}_0$.

(i) It suffices to show that $\mathcal{C}^n \subset f^{-1}(\mathcal{C}^{n-1})$ for $n \in \mathbb{N}$. We prove this by induction on n; this is clear for $n = 1$. Assume that the statement holds for some $n \in \mathbb{N}$; so $\mathcal{C}^n \subset f^{-1}(\mathcal{C}^{n-1})$. Since $h_n = H_1^n$ and $h_{n-1} = H_1^{n-1}$ are homeomorphisms with $f \circ h_n = h_{n-1} \circ f$, we have $h_n(f^{-1}(\mathcal{C}^{n-1})) = f^{-1}(h_{n-1}(\mathcal{C}^{n-1}))$ by Lemma 11.2. Thus

$$\mathcal{C}^{n+1} = h_n(\mathcal{C}^n) \subset h_n(f^{-1}(\mathcal{C}^{n-1})) = f^{-1}(h_{n-1}(\mathcal{C}^{n-1})) = f^{-1}(\mathcal{C}^n),$$

and (i) follows.

(ii)–(iv) From the definition of H^n, the remark after the proof of Proposition 11.3, and induction on n we conclude that H^n is an isotopy rel. $f^{-n}(\mathrm{post}(f))$. Since $H_0^n = \mathrm{id}_{S^2}$ and

$$f^{-n}(\mathrm{post}(f)) \subset f^{-(n+k)}(\mathrm{post}(f))$$

for $n, k \in \mathbb{N}_0$, statements (ii) and (iii) immediately follow from this by induction on k for fixed n. Statement (iv) follows from (iii) (with $n = 0$ and $k \in \mathbb{N}_0$ arbitrary) and the fact that $\mathrm{post}(f) \subset \mathcal{C}^0$.

(v) By (iv) we have $\#(f^{-n}(\mathrm{post}(f)) \cap \mathcal{C}^{n+k}) \geq \#\,\mathrm{post}(f) \geq 3$. In particular, the points in $f^{-n}(\mathrm{post}(f))$ that lie on $\mathcal{C}^{n+k}$ subdivide this curve into arcs whose endpoints lie in $f^{-n}(\mathrm{post}(f))$ and whose interiors are disjoint from $f^{-n}(\mathrm{post}(f))$. Let $\alpha \subset \mathcal{C}^{n+k}$ be one of these arcs. Then we have $\mathrm{int}(\alpha) \subset f^{-n}(\mathcal{C}^k) \setminus f^{-n}(\mathrm{post}(f))$ by (i), and $\partial\alpha \subset f^{-n}(\mathrm{post}(f))$. Since by Proposition 5.17 (iii) the set $f^{-n}(\mathcal{C}^k)$ is the 1-skeleton and the set $f^{-n}(\mathrm{post}(f))$ the 0-skeleton of the cell decomposition $\mathcal{D}^n(f, \mathcal{C}^k)$, we conclude from Lemmas 5.4 and 5.5 that α is an edge in $\mathcal{D}^n(f, \mathcal{C}^k)$, i.e., an n-edge for $(f, \mathcal{C}^k)$. Hence $\mathcal{C}^{n+k}$ consists of n-edges for $(f, \mathcal{C}^k)$.

(vi) By (i) and (ii) we know that $\mathcal{C}^n$ for $n \in \mathbb{N}$ is a Jordan curve with $\mathcal{C}^n \subset f^{-1}(\mathcal{C}^{n-1})$ that is isotopic to $\mathcal{C}^1$ rel. $f^{-1}(\mathrm{post}(f))$. Let $\widehat{\mathcal{C}} \subset f^{-1}(\mathcal{C}^{n-1})$ be another Jordan curve isotopic to $\mathcal{C}^1$ rel. $f^{-1}(\mathrm{post}(f))$. Then $\mathcal{C}^n$ and $\widehat{\mathcal{C}}$ are isotopic to each other rel. $f^{-1}(\mathrm{post}(f))$. Note that $f^{-1}(\mathcal{C}^{n-1})$ is the 1-skeleton of the cell decomposition $\mathcal{D}^1(f, \mathcal{C}^{n-1})$ and $f^{-1}(\mathrm{post}(f))$ its set of vertices. Since each tile in $\mathcal{D}^1(f, \mathcal{C}^{n-1})$ has at least $\#\,\mathrm{post}(f) \geq 3$ vertices, we can apply Lemma 11.12 and conclude that $\widehat{\mathcal{C}} = \mathcal{C}^n$. The uniqueness statement for $\mathcal{C}^n$ follows.

(vii) It follows from (vi) and induction on n that $\mathcal{C}^n$ is uniquely determined by $\mathcal{C}^0$ and $\mathcal{C}^1$.

(viii) Since f is expanding, we can pick a visual metric ϱ for f. Let $\Lambda > 1$ be the expansion factor of ϱ. By Lemma 11.4 the diameters of the tracks of the isotopy H^n are bounded by $C\Lambda^{-n}$, where C is a fixed constant. Since $H_0^n = \mathrm{id}_{S^2}$ and $\mathcal{C}^{n+1} = H_1^n(\mathcal{C}^n)$ for $n \in \mathbb{N}_0$ this implies that $\mathrm{dist}_\varrho^H(\mathcal{C}^n, \mathcal{C}^{n+1}) \leq C\Lambda^{-n}$ for $n \in \mathbb{N}_0$. It follows that the sequence $\{\mathcal{C}^n\}$ is a Cauchy sequence with respect to Hausdorff distance. Recall that the space of all non-empty closed subsets of a compact metric space is complete if it is equipped with the Hausdorff distance. Thus there exists a non-empty closed set $\widetilde{\mathcal{C}} \subset S^2$ such that $\mathcal{C}^n \to \widetilde{\mathcal{C}}$ as $n \to \infty$ in the sense of Hausdorff convergence. Since $\mathrm{post}(f) \subset \mathcal{C}^n$ for all $n \in \mathbb{N}_0$ by (iv), we have $\mathrm{post}(f) \subset \widetilde{\mathcal{C}}$.

It remains to show that $\widetilde{\mathcal{C}}$ is f-invariant. To see this, let $p \in \widetilde{\mathcal{C}}$ be arbitrary. Then there exists a sequence $\{p_n\}$ of points in S^2 such that $p_n \in \mathcal{C}^n$ for $n \in \mathbb{N}_0$ and $p_n \to p$ as $n \to \infty$. By continuity of f we have $f(p_n) \to f(p)$ as $n \to \infty$. Moreover, (i) implies that $f(p_n) \in \mathcal{C}^{n-1}$ for $n \in \mathbb{N}$. Hence $f(p) \in \widetilde{\mathcal{C}}$, and so the set $\widetilde{\mathcal{C}}$ is indeed f-invariant. $\qquad\square$

As an application of the preceding setup we prove a statement that gives a necessary and sufficient condition for the map $\widehat{f}$ in Theorem 15.4 to be combinatorially expanding.

PROPOSITION 15.19. *Suppose $f\colon S^2 \to S^2$ is a Thurston map with $\#\operatorname{post}(f) \geq 3$, and let the isotopy $H^0\colon S^2 \times I \to S^2$ and Jordan curves $\mathcal{C}^n$ for $n \in \mathbb{N}_0$ be defined as above.*

Then $\widehat{f} = H_1^0 \circ f$ is combinatorially expanding for $\mathcal{C}^1 = \mathcal{C}'$ if and only if there exists $n \in \mathbb{N}$ such that no n-tile for $(f, \mathcal{C}^0)$ joins opposite sides of $\mathcal{C}^n$.

PROOF. Let H^n for $n \in \mathbb{N}_0$ be the isotopies used in the definition of the curves $\mathcal{C}^n$. Set $h_n := H_1^n$. Then $\widehat{f} = h_0 \circ f$, $\mathcal{C}^{n+1} = h_n(\mathcal{C}^n)$, and $h_n \circ f = f \circ h_{n+1}$ for $n \in \mathbb{N}_0$. It follows by induction that for $n \in \mathbb{N}$ we have

$$\widehat{f}^n = h_0 \circ f \circ \cdots \circ h_0 \circ f = h_0 \circ f^n \circ h_{n-1} \circ \cdots \circ h_1,$$

and so

$$h_0 \circ f^n = \widehat{f}^n \circ h_1^{-1} \circ \ldots h_{n-1}^{-1}.$$

Hence

$$f^{-n}(\mathcal{C}^0) = f^{-n}(h_0^{-1}(\mathcal{C}^1)) = (h_{n-1} \circ \cdots \circ h_1)(\widehat{f}^{-n}(\mathcal{C}^1)).$$

Recall that the n-tiles for $(f, \mathcal{C}^0)$ are the closures of the complementary components of $f^{-n}(\mathcal{C}^0)$, and the n-tiles for $(\widehat{f}, \mathcal{C}^1)$ the closures of the complementary components of $\widehat{f}^{-n}(\mathcal{C}^1)$ (Proposition 5.17 (v)). So from the previous identity we conclude that the n-tiles for $(f, \mathcal{C}^0)$ are precisely the images of the n-tiles for $(\widehat{f}, \mathcal{C}^1)$ under the homeomorphism $h_{n-1} \circ \cdots \circ h_1$. Note that this homeomorphism is isotopic to id_{S^2} rel. $\operatorname{post}(f) = \operatorname{post}(\widehat{f})$ and maps $\mathcal{C}^1$ to $\mathcal{C}^n$. Thus no n-tile for $(\widehat{f}, \mathcal{C}^1)$ joins opposite sides of $\mathcal{C}^1$ if and only if no n-tile for $(f, \mathcal{C}^0)$ joins opposite sides of $\mathcal{C}^n$.

Now $\widehat{f}$ is combinatorially expanding for $\mathcal{C}^1$ if and only if there exists $n \in \mathbb{N}$ such that no n-tile for $(\widehat{f}, \mathcal{C}^1)$ joins opposite sides of $\mathcal{C}^1$. By what we have seen, this is the case if and only if there exists $n \in \mathbb{N}$ such that no n-tile for $(f, \mathcal{C}^0)$ joins opposite sides of $\mathcal{C}^n$. $\qquad\square$

Let us now assume that our Thurston map f is expanding. Then the curves $\mathcal{C}^n$ Hausdorff converge to an f-invariant closed set $\widetilde{\mathcal{C}}$ by

Lemma 15.18 (viii). In general, $\widetilde{\mathcal{C}}$ will not be a Jordan curve (see Example 15.23). The following proposition shows that $\widetilde{\mathcal{C}}$ is a Jordan curve if the map $\widehat{f} = H_1^0 \circ f$ is combinatorially expanding for $\mathcal{C}^1$. Actually, one can also show that this condition is also necessary for $\widetilde{\mathcal{C}}$ to be a Jordan curve, but we will not present the proof for this statement as it is somewhat involved.

PROPOSITION 15.20 (Iterative procedure for an invariant curve). *Let* $f\colon S^2 \to S^2$ *be an expanding Thurston map, and suppose the isotopy* $H^0\colon S^2 \times I \to S^2$ *and Jordan curves* $\mathcal{C}^n$ *for* $n \in \mathbb{N}_0$ *are defined as above.*

If $\widehat{f} = H_1^0 \circ f$ *is combinatorially expanding for* $\mathcal{C}^1 = \mathcal{C}'$, *then* $\mathcal{C}^n$ *Hausdorff converges to a Jordan curve* $\widetilde{\mathcal{C}} \subset S^2$ *as* $n \to \infty$. *In this case, the curve* $\widetilde{\mathcal{C}}$ *is* f-*invariant and* $\mathrm{post}(f) \subset \widetilde{\mathcal{C}}$. *Moreover,* $\widetilde{\mathcal{C}}$ *is isotopic to* $\mathcal{C}^1$ *rel.* $f^{-1}(\mathrm{post}(f))$.

By Theorem 15.5 the curve $\widetilde{\mathcal{C}}$ will be the unique Jordan curve with the given properties.

PROOF. Suppose that $\widehat{f} = H_1^0 \circ f$ is combinatorially expanding for $\mathcal{C}^1$. From Theorem 15.4 it follows that there exists an f-invariant Jordan curve $\widetilde{\mathcal{C}} \subset S^2$ with $\mathrm{post}(f) \subset \widetilde{\mathcal{C}}$ that is isotopic to $\mathcal{C}^0$ rel. $\mathrm{post}(f)$ and isotopic to $\mathcal{C}^1$ rel. $f^{-1}(\mathrm{post}(f))$. Let $K^0\colon S^2 \times I \to S^2$ be an isotopy rel. $\mathrm{post}(f)$ that deforms $\widetilde{\mathcal{C}}$ to $\mathcal{C}^0$; so $K_0^0 = \mathrm{id}_{S^2}$ and $K_1^0(\widetilde{\mathcal{C}}) = \mathcal{C}^0$. Using Proposition 11.3 repeatedly, we can find isotopies $K^n\colon S^2 \times I \to S^2$ rel. $f^{-1}(\mathrm{post}(f))$ with $K_0^n = \mathrm{id}_{S^2}$ such that $f \circ K_1^n = K_1^{n-1} \circ f$ for $n \in \mathbb{N}$.

Claim. $\widetilde{\mathcal{C}}^n := K_1^n(\widetilde{\mathcal{C}}) = \mathcal{C}^n$ for all $n \in \mathbb{N}_0$.

We prove this claim by induction on n; it follows from the choice of K^0 for $n = 0$. Assume that the statement is true for some $n \in \mathbb{N}_0$. Then $K_1^n(\widetilde{\mathcal{C}}) = \mathcal{C}^n$, and so by Lemma 11.2 we have

$$\widetilde{\mathcal{C}}^{n+1} = K_1^{n+1}(\widetilde{\mathcal{C}}) \subset K_1^{n+1}(f^{-1}(\widetilde{\mathcal{C}})) = f^{-1}(K_1^n(\widetilde{\mathcal{C}})) = f^{-1}(\mathcal{C}^n).$$

Since K^{n+1} is an isotopy rel. $f^{-1}(\mathrm{post}(f))$, the curve $\widetilde{\mathcal{C}}^{n+1}$ is isotopic to $\widetilde{\mathcal{C}}$ and hence to $\mathcal{C}^1$ rel. $f^{-1}(\mathrm{post}(f))$. So Lemma 15.18 (vi) implies that $\widetilde{\mathcal{C}}^{n+1} = \mathcal{C}^{n+1}$. This proves the claim.

By Lemma 11.4 the maps K_1^n converge uniformly to the identity on S^2. Hence $\mathcal{C}^n = K_1^n(\widetilde{\mathcal{C}})$ Hausdorff converges to the Jordan curve $\widetilde{\mathcal{C}}$ as $n \to \infty$. The statement follows.   □

REMARK 15.21. If $f\colon S^2 \to S^2$ is an expanding Thurston map, then every f-invariant Jordan curve $\widetilde{\mathcal{C}}$ with $\mathrm{post}(f) \subset \widetilde{\mathcal{C}}$ can be obtained by our iterative procedure. Indeed, suppose that $\widetilde{\mathcal{C}}$ is such a curve.

Trivially, we can then take $\mathcal{C} = \mathcal{C}^0 = \widetilde{\mathcal{C}}$, $\mathcal{C}' = \mathcal{C}^1 = \widetilde{\mathcal{C}}$, and $H_t^0 = \mathrm{id}_{S^2}$ for $t \in I$. Then $\mathcal{C}^n = \widetilde{\mathcal{C}}$ for all $n \in \mathbb{N}_0$ and so $\mathcal{C}^n \to \widetilde{\mathcal{C}}$ as $n \to \infty$.

Actually, a much stronger statement is true. Namely, we can start with *any* Jordan curve $\mathcal{C}$ in the same isotopy class rel. $\mathrm{post}(f)$ as $\widetilde{\mathcal{C}}$. Suppose that $\mathcal{C}$ is such a curve. First, we claim that then there exists a unique Jordan curve $\mathcal{C}' \subset f^{-1}(\mathcal{C})$ that is isotopic to $\widetilde{\mathcal{C}}$ rel. $f^{-1}(\mathrm{post}(f))$. To see this, let $K^0 \colon S^2 \times I \to S^2$ be an isotopy rel. $\mathrm{post}(f)$ with $K_0^0 = \mathrm{id}_{S^2}$ and $K_1^0(\widetilde{\mathcal{C}}) = \mathcal{C}$. By Proposition 11.3 we can lift K^0 by f to an isotopy K^1 rel. $f^{-1}(\mathrm{post}(f))$ with $K_0^1 = \mathrm{id}_{S^2}$ and $K_t^0 \circ f = f \circ K_t^1$ for $t \in I$. Then the Jordan curve $\mathcal{C}' := K_1^1(\widetilde{\mathcal{C}})$ satisfies

$$\mathcal{C}' = K_1^1(\widetilde{\mathcal{C}}) \subset K_1^1(f^{-1}(\widetilde{\mathcal{C}})) = f^{-1}(K_1^0(\widetilde{\mathcal{C}})) = f^{-1}(\mathcal{C}).$$

Here we used $\widetilde{\mathcal{C}} \subset f^{-1}(\widetilde{\mathcal{C}})$ and Lemma 11.2. This shows existence of a curve $\mathcal{C}'$ with the desired properties. Uniqueness of $\mathcal{C}'$ follows from Lemma 11.12.

Define $H \colon S^2 \times I \to S^2$ by setting $H_t = K_t^1 \circ (K_t^0)^{-1}$ for $t \in I$. Then H is an isotopy rel. $\mathrm{post}(f)$ that deforms $\mathcal{C}^0 := \mathcal{C}$ into $\mathcal{C}^1 := \mathcal{C}'$. Indeed, we have $H_0 = \mathrm{id}_{S^2}$ and

$$H_1(\mathcal{C}^0) = K_1^1((K_1^0)^{-1}(\mathcal{C})) = K_1^1(\widetilde{\mathcal{C}}) = \mathcal{C}' = \mathcal{C}^1.$$

Moreover,

$$\widehat{f} := H_1 \circ f = K_1^1 \circ (K_1^0)^{-1} \circ f = K_1^1 \circ f \circ (K_1^1)^{-1}.$$

Thus it follows from Lemma 12.16 that $\widehat{f}$ is combinatorially expanding for $\mathcal{C}^1 = \mathcal{C}' = K_1^1(\widetilde{\mathcal{C}})$.

Define the sequence $\{\mathcal{C}^n\}$ starting from $\mathcal{C}^0$ and $\mathcal{C}^1$ as before. From Proposition 15.20 it follows that as $n \to \infty$ the curves $\mathcal{C}^n$ Hausdorff converge to an f-invariant Jordan curve that is isotopic to $\mathcal{C}^1$, and hence isotopic to $\widetilde{\mathcal{C}}$, rel. $f^{-1}(\mathrm{post}(f))$. From Theorem 15.5 it follows that the unique such curve is $\widetilde{\mathcal{C}}$. Thus $\mathcal{C}^n \to \widetilde{\mathcal{C}}$ in the Hausdorff sense as $n \to \infty$.

REMARK 15.22. In the inductive definition of $\mathcal{C}^{n+1} = H_1^n(\mathcal{C}^n)$ one can construct $\mathcal{C}^{n+1}$ from $\mathcal{C}^n$ by an *edge replacement procedure* without explicitly knowing the isotopy H^n. To explain this, suppose that $n \in \mathbb{N}$, and that $\mathcal{C}^n$ has already been constructed (starting from given curves $\mathcal{C}^0$ and $\mathcal{C}^1$). We know by Lemma 15.18 (v) that $\mathcal{C}^n$ consists of n-edges α^n for $(f, \mathcal{C}^0)$. Then $\mathcal{C}^{n+1}$ is obtained from $\mathcal{C}^n$ by replacing each n-edge $\alpha^n \subset \mathcal{C}^n$ by a certain arc β^{n+1} with the same endpoints as α^n.

Indeed, we can set $\beta^{n+1} := H_1^n(\alpha^n) \subset \mathcal{C}^{n+1}$. Then the union of these arcs β^{n+1} is equal to $\mathcal{C}^{n+1}$. Moreover, since H^n is an isotopy relative to the set $f^{-n}(\mathrm{post}(f))$ of n-vertices, and α^n is an n-edge for
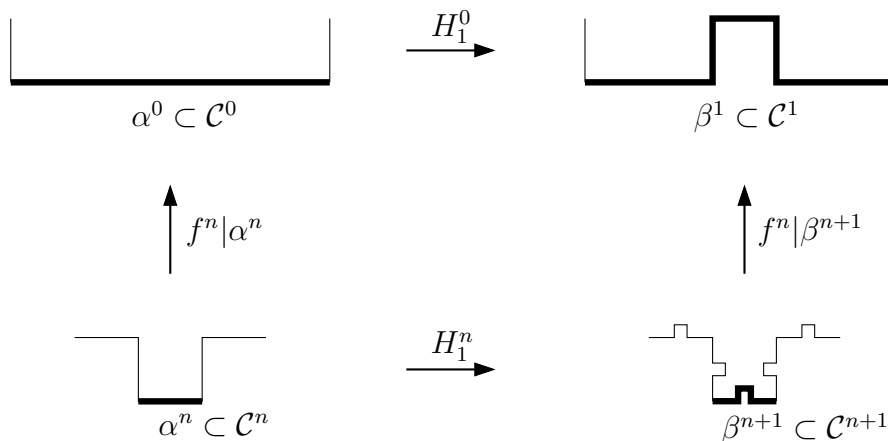
FIGURE 15.5. Iterative construction by replacing edges.

$(f, \mathcal{C}^0)$ and so has n-vertices as endpoints, the arcs α^n and β^{n+1} have the same endpoints.

Now the arc β^{n+1} is the unique arc in $f^{-n}(\mathcal{C}^1)$ that is isotopic to α^n rel. $f^{-n}(\operatorname{post}(f))$. This property often allows one to determine β^{n+1} directly from α^n.

To see that this characterization of β^{n+1} holds, note that by Lemma 15.18 (i) we have $\beta^{n+1} \subset \mathcal{C}^{n+1} \subset f^{-n}(\mathcal{C}^1)$. Moreover, $\beta^{n+1} = H_1^n(\alpha^n)$ is isotopic to α^n rel. $f^{-n}(\operatorname{post}(f))$.

Suppose $\widetilde{\beta}^{n+1} \subset f^{-n}(\mathcal{C}^1)$ is another arc that is isotopic to α^n rel. $f^{-n}(\operatorname{post}(f))$. Then the arcs β^{n+1} and $\widetilde{\beta}^{n+1}$ have endpoints in $f^{-n}(\operatorname{post}(f))$, but contain no other points in this set, since this is true for α^n. This and the inclusions $\beta^{n+1}, \widetilde{\beta}^{n+1} \subset f^{-n}(\mathcal{C}^1)$ imply that β^{n+1} and $\widetilde{\beta}^{n+1}$ are n-edges for $(f, \mathcal{C}^1)$ (see the argument in the proof of Lemma 15.18 (v)). Since β^{n+1} and $\widetilde{\beta}^{n+1}$ are isotopic relative to the set $f^{-n}(\operatorname{post}(f))$, which is the 0-skeleton of $\mathcal{D}^n(f, \mathcal{C}^1)$, it follows from the first part of the proof of Lemma 11.12 that $\beta^{n+1} = \widetilde{\beta}^{n+1}$ as desired.

As we have just seen, β^{n+1} is an n-edge for $(f, \mathcal{C}^1)$. Since β^{n+1} has endpoints in the set $f^{-n}(\operatorname{post}(f)) \subset f^{-(n+1)}(\operatorname{post}(f))$ and $\beta^{n+1} \subset \mathcal{C}^{n+1} \subset f^{-(n+1)}(\mathcal{C}^0)$, a similar argument also shows that β^{n+1} consists of $(n+1)$-edges for $(f, \mathcal{C}^0)$.

One can look at the arc replacement procedure $\alpha^n \to \beta^{n+1}$ from yet another point of view. Since α^n is an n-edge for $(f, \mathcal{C}^0)$, the map $f^n|\alpha^n$ is a homeomorphism of α^n onto the 0-edge $\alpha^0 := f^n(\alpha^n) \subset \mathcal{C}^0$ for $(f, \mathcal{C}^0)$ (Proposition 5.17 (i)). The endpoints of α^0 lie in $\operatorname{post}(f)$. Then $\beta^1 := H_1^0(\alpha^0)$ is the unique subarc of $\mathcal{C}^1$ that has the same endpoints as α^0, but contains no other points in $\operatorname{post}(f)$ (here it is important that $\#(\mathcal{C}^1 \cap \operatorname{post}(f)) = \# \operatorname{post}(f) \geq 3$). Since $f^n \circ H_1^n = H_1^0 \circ f^n$

and $\beta^{n+1} = H_1^n(\alpha^n)$, the map $f^n|\beta^{n+1}$ is thus a homeomorphism of β^{n+1} onto β^1. Often, this information (together with the fact that α^n and β^{n+1} share endpoints) is enough to determine β^{n+1} uniquely. We illustrate this procedure in Figure 15.5. Here the map f (as well as the curves $\mathcal{C}^0, \mathcal{C}^1, \dots$ and the isotopies $H^0, H^1, \dots$) are as in Example 15.17, see also Figure 15.4.

For example, suppose that β^1 lies in a *single* 0-tile X^0 for $(f, \mathcal{C}^0)$, i.e., in one of the Jordan regions bounded by $\mathcal{C}^0$. This is not always true, but in Example 15.17 as well as the Examples 15.23 and 15.24 discussed below this is the case. Then there exists a unique n-tile X^n for $(f, \mathcal{C}^0)$ with $\alpha^n \subset \partial X^n$ and $f^n(X^n) = X^0$; if we assign colors to tiles for $(f, \mathcal{C}^0)$ as in Lemma 5.21, then X^n is the unique n-tile for $(f, \mathcal{C}^0)$ that contains α^n in its boundary and has the same color as X^0.

Consider the arc $\widetilde{\beta}^{n+1} := (f^n|X^n)^{-1}(\beta^1) \subset X^n$. Then $\widetilde{\beta}^{n+1}$ has the same endpoints as $(f^n|X^n)^{-1}(\alpha^0) = \alpha^n$ and is contained in $f^{-n}(\mathcal{C}^1)$. Moreover, $\widetilde{\beta}^{n+1}$ is isotopic to α^n rel. $f^{-n}(\mathrm{post}(f))$; this easily follows from Lemma 11.8, since our assumptions imply that one can find a suitable simply connected domain $\Omega \subset S^2$ that contains $\widetilde{\beta}^{n+1}$ and α^n and no point in $f^{-n}(\mathrm{post}(f))$ except the endpoints of $\widetilde{\beta}^{n+1}$ and α^n. By what we have seen above, we conclude $\beta^{n+1} = \widetilde{\beta}^{n+1}$, and so

$$(15.3) \qquad\qquad \beta^{n+1} = (f^n|X^n)^{-1}(\beta^1).$$

In the special case under consideration, this leads to a very convenient edge replacement procedure that can be summarized as follows: Suppose the arc $\beta^1 \subset \mathcal{C}^1$ corresponding to $\alpha^0 = f^n(\alpha^n) \subset \mathcal{C}^0$ lies in a single 0-tile X^0, and let X^n be the n-tile that contains α^n in its boundary and has the same color as X^0 (so that $f^n(X^n) = X^0$). Then α^n is replaced by the arc β^{n+1} in X^n that corresponds to $\beta^1 \subset X^0$ under the homeomorphism $f^n|X^n$ of X^n onto X^0.

The next example illustrates what happens if the map $\widehat{f}$ in Proposition 15.20 is not combinatorially expanding.

EXAMPLE 15.23. Let $g \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be the Lattès map obtained according to Theorem 3.1 (ii) as a quotient of the map

$$A \colon \mathbb{C} \to \mathbb{C}, \quad u \mapsto A(u) := 3u,$$

by a crystallographic group of type (2222) as in (3.24). This map was the second map in Example 14.23.

As in Section 1.1 we identify $\widehat{\mathbb{C}}$ with a pillow that is obtained by gluing two squares together, see also the bottom of Figure 14.3. Let $\mathcal{C}^0$ be the equator of the pillow. The curve $\mathcal{C}^1 \subset g^{-1}(\mathcal{C}^0)$ is drawn with a
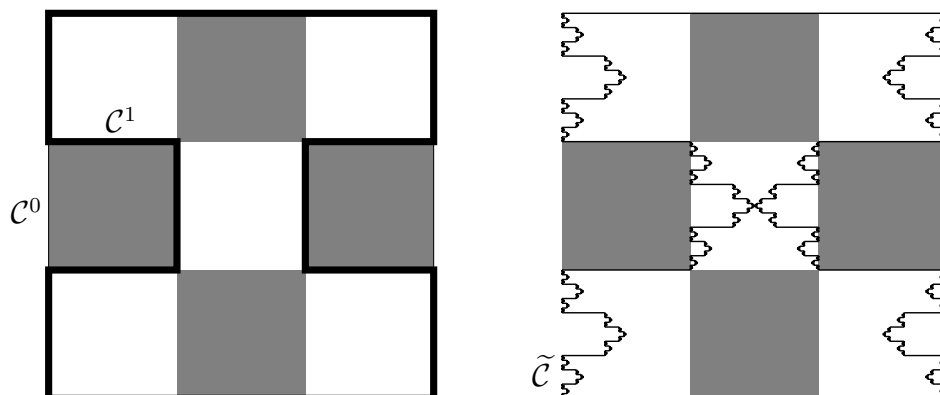
FIGURE 15.6. Since $\widehat{g}$ is not combinatorially expanding, $\widetilde{\mathcal{C}}$ is not a Jordan curve.

thick line in the top left of Figure 15.6. Clearly, there is an isotopy H^0 rel. $\mathrm{post}(g)$ (the four vertices of the pillow) that deforms $\mathcal{C}^0$ to $\mathcal{C}^1$. Note that $\widehat{g} = H_1^0 \circ g$ is not combinatorially expanding for $\mathcal{C}^1$ (see Figure 14.3). Starting with the data $\mathcal{C}^0$, $\mathcal{C}^1$, H^0, we inductively define Jordan curves $\mathcal{C}^n$ as described before.

It is slightly more difficult than in Example 15.17 to see here how $\mathcal{C}^{n+1}$ evolves from $\mathcal{C}^n$, since the 0-edges are replaced by arcs (consisting of 1-edges) in different ways. Each 0-edge that is drawn horizontally in Figure 15.6 is replaced by itself. Since every horizontal 1-edge is mapped by g to a horizontal 0-edge, it is also replaced by itself in the construction of $\mathcal{C}^2$ from $\mathcal{C}^1$; in general, if in one step we obtain a horizontal edge, then it remains unchanged in subsequent steps (see Remark 15.22).

Then $\mathcal{C}^n \to \widetilde{\mathcal{C}}$ as $n \to \infty$ in the sense of Hausdorff convergence, where the set $\widetilde{\mathcal{C}}$ is indicated on the right of Figure 15.6. In this case, $\widetilde{\mathcal{C}}$ is not a Jordan curve and $\widehat{\mathbb{C}} \setminus \widetilde{\mathcal{C}}$ has three components. Of course, the "self-intersections" of the limit set $\widetilde{\mathcal{C}}$ can be more complicated in general.

We conclude this section with one more example. It shows a non-trivial invariant curve that is rectifiable.

EXAMPLE 15.24. Let f be the map from Example 15.17, i.e., the Lattès map obtained as in (1.1), where we choose $A \colon \mathbb{C} \to \mathbb{C}$, $u \mapsto A(u) := 5u$. The curve $\mathcal{C} = \mathcal{C}^0$ is the equator of the pillow as before, and we consider cells for $(f, \mathcal{C})$. On the pillow the map f sends the lower left 1-tile to the white 0-tile by the map $u \mapsto 5u$ and extends to other 1-tiles by reflection.
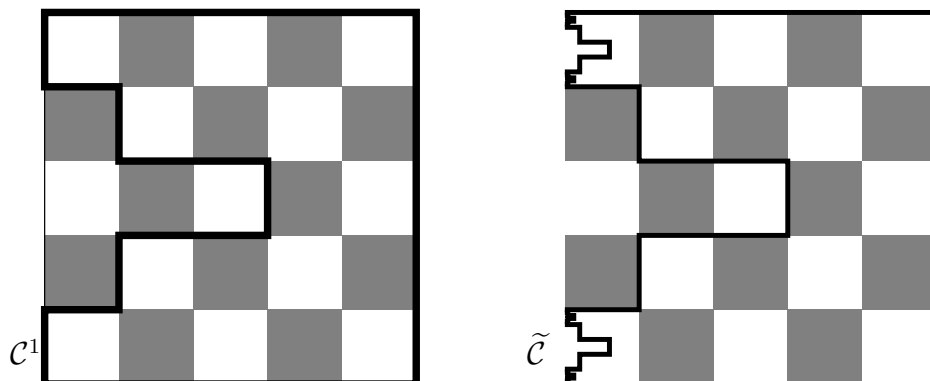
FIGURE 15.7. A non-trivial rectifiable invariant Jordan curve.

The curve $\mathcal{C}^1$ (which is isotopic to $\mathcal{C}^0$ rel. post(f) by an isotopy H^0) is the thick curve indicated on the left of Figure 15.7. Note that no 1-tile for $(f, \mathcal{C}^0)$ joins opposite sides of $\mathcal{C}^1$. Thus the sequence of curves $\{\mathcal{C}^n\}$, defined as before, Hausdorff converges to an f-invariant Jordan curve $\widetilde{\mathcal{C}}$ by Proposition 15.19 and Proposition 15.20.

Note that the three 0-edges on the top, bottom, and right side of the pillow are deformed by H^0 to themselves. This means that each n-edge in $\mathcal{C}^n$ that is sent to one of these 0-edges by f^n remains unchanged in the passage from $\mathcal{C}^n$ to $\mathcal{C}^{n+1}$.

The resulting f-invariant Jordan curve $\widetilde{\mathcal{C}}$ is shown on the right. It is not hard to see that $\widetilde{\mathcal{C}}$ is a rectifiable curve on the pillow. Indeed, if as before we identify the top square of the pillow with $[0, 1/2]^2$, then the n-edges have length $5^{-n}/2$. The curve $\mathcal{C}^n$ contains 2^n "alive" n-edges that will not remain unchanged in subsequent steps. In $\mathcal{C}^{n+1}$ each of them is replaced by 11 edges of level $n+1$. A simple computation gives

$$\text{length}(\mathcal{C}^{n+1}) = \text{length}(\mathcal{C}^n) + \tfrac{3}{5}(2/5)^n,$$

which implies that indeed length($\widetilde{\mathcal{C}}$) $< \infty$.

15.3. Invariant curves are quasicircles

Recall from Section 4.1 that a metric circle (S, d) is called a quasicircle if it is quasisymmetrically equivalent to the unit circle in $\mathbb{R}^2$ (equipped with the Euclidean metric). This is the case if and only if (S, d) is doubling and of bounded turning (see Theorem 4.1). We will now verify that this is true for an invariant curve as in Theorem 15.3.

PROOF OF THEOREM 15.3. Suppose $\mathcal{C}$ is an f-invariant Jordan curve as in the statement, and let ϱ be a visual metric on S^2 with

expansion factor $\Lambda > 1$. Metric notions will be with respect to this metric in the following.

In the ensuing proof, we will consider edges for $(f, \mathcal{C})$. Since $\mathcal{C}$ is f-invariant, edges are subdivided by edges of higher level (see Proposition 12.5 (iv)). The Jordan curve $\mathcal{C}$ is the union of all 0-edges, so this implies that $\mathcal{C}$ is a union of n-edges for all $n \in \mathbb{N}_0$. If $n, k \in \mathbb{N}_0$ and $\widetilde{e}$ is an arbitrary $(n + k)$-edge with $\widetilde{e} \subset \mathcal{C}$, then there exists a unique n-edge e' with $\widetilde{e} \subset e' \subset \mathcal{C}$.

If e' is an n-edge, then the number of $(n + k)$-edges $\widetilde{e}$ contained in e' is bounded by $\# \operatorname{post}(f) \deg(f)^k$. Indeed, the map $f^n|e'$ is injective; so the images of these $(n + k)$-edges $\widetilde{e}$ under the map f^n are distinct k-edges, and the number of k-edges is equal to $\# \operatorname{post}(f) \deg(f)^k$ (see Proposition 5.17 (iv)).

After these preliminaries, we are ready to show that $\mathcal{C}$ equipped with (the restriction of) ϱ is a quasicircle. We first establish that $\mathcal{C}$ is doubling. Note that in contrast (S^2, ϱ) is not doubling in general (see Theorem 18.1 (i)).

Let $x \in \mathcal{C}$, and $0 < r \leq 2 \operatorname{diam}(\mathcal{C})$. In order to show that $\mathcal{C}$ is doubling, it suffices to cover $B(x, r) \cap \mathcal{C}$ by a controlled number of sets of diameter $< r/4$.

It follows from Proposition 8.4 that we can find $n \in \mathbb{N}_0$ depending on r, as well as constants $C(\asymp) > 0$ and $k_0 \in \mathbb{N}_0$ that are independent of x and r with the following properties:

  (i) $r \asymp \Lambda^{-n}$,

 (ii) $\operatorname{diam}(e) < r/4$ whenever e is an $(k_0 + n)$-edge,

(iii) $\operatorname{dist}(e, e') \geq r$ if $n - k_0 \geq 0$ and e and e' are disjoint $(n - k_0)$-edges.

Let E be the set of all $(n + k_0)$-edges contained in $\mathcal{C}$ that meet $B(x, r)$. Then the collection E forms a cover of $\mathcal{C} \cap B(x, r)$ and consists of sets of diameter $< r/4$ by (ii). Hence it suffices to find a uniform upper bound on $\#E$. If $n < k_0$, then $\#E \leq \# \operatorname{post}(f) \deg(f)^{2k_0}$.

Otherwise, $n - k_0 \geq 0$. Then we can find an $(n - k_0)$-edge $e \subset \mathcal{C}$ with $x \in e$. Let $\widetilde{e}$ be an arbitrary $(n + k_0)$-edge in E. Again we can find an $(n - k_0)$-edge $e' \subset \mathcal{C}$ that contains $\widetilde{e}$.

There exists a point $y \in \widetilde{e} \cap B(x, r)$. Hence $\operatorname{dist}(e, e') \leq \varrho(x, y) < r$. This implies $e \cap e' \neq \emptyset$ by (iii). So whatever $\widetilde{e} \in E$ is, the corresponding $(n - k_0)$-edge $e' \subset \mathcal{C}$ meets the fixed $(n - k_0)$-edge e. This leaves at most three possibilities for e', namely e, and the two "neighbors" of e on $\mathcal{C}$. So there are three or less $(n - k_0)$-edges that contain all the edges in E. Since each $(n - k_0)$-edge contains at most $\# \operatorname{post}(f) \deg(f)^{2k_0}$

edges of level $(n + k_0)$, it follows that $\#E \leq 3\# \operatorname{post}(f) \deg(f)^{2k_0}$. In both cases, we get an upper bound for $\#E$ as desired.

It remains to show that $\mathcal{C}$ is of bounded turning. Let $x, y \in \mathcal{C}$ with $x \neq y$ be arbitrary, and let $n_0 \geq 0$ be the smallest integer for which there exist n_0-edges $e_x \subset \mathcal{C}$ and $e_y \subset \mathcal{C}$ with $x \in e_x$, $y \in e_y$ and $e_x \cap e_y = \emptyset$. Note that n_0 is well-defined, because f is expanding and so the diameter of n-edges approaches 0 uniformly as $n \to \infty$.

Then by Proposition 8.4 (i),

$$\varrho(x, y) \gtrsim \Lambda^{-n_0}.$$

If $n_0 = 0$, then

$$\operatorname{diam}(\mathcal{C}) \lesssim \varrho(x, y)$$

and there is nothing to prove. If $n_0 \geq 1$, we can find $(n_0 - 1)$-edges $e'_x \subset \mathcal{C}$ and $e'_y \subset \mathcal{C}$ with $x \in e'_x$, $y \in e'_y$, and $e'_x \cap e'_y \neq \emptyset$. Then $e'_x \cup e'_y$ must contain one of the subarcs γ of $\mathcal{C}$ with endpoints x and y. Hence

$$\operatorname{diam}(\gamma) \leq \operatorname{diam}(e'_x) + \operatorname{diam}(e'_y) \lesssim \Lambda^{-n_0} \lesssim \varrho(x, y).$$

Since the implicit multiplicative constants in the previous inequalities do not depend on x and y, we get a bound as desired. $\qquad\square$

Recall that a metric if visual for f if and only only if it is visual for a given iterate of f (see Proposition 8.3 (v)). Hence we may apply Theorem 15.3 to any Jordan curve $\mathcal{C} \subset S^2$, with $\operatorname{post}(f) \subset \mathcal{C}$, that is invariant for an iterate of f. In particular, the invariant Jordan curve in Theorem 15.1 is a quasicircle if equipped with a visual metric for f.

A family of quasisymmetries (possibly defined on different spaces) is called *uniformly quasisymmetric* if there exists a homeomorphism $\eta \colon [0, \infty) \to [0, \infty)$ such that each map in the family is an η-quasi-symmetry. Obviously, each finite family of quasisymmetries is uniformly quasisymmetric. If h is an η-quasisymmetry, then h^{-1} is an $\widetilde{\eta}$-quasisymmetry, where $\widetilde{\eta}$ only depends on η; actually, one can take $\widetilde{\eta} \colon [0, \infty) \to [0, \infty)$ defined by $\widetilde{\eta}(0) = 0$ and $\widetilde{\eta}(t) = 1/\eta^{-1}(1/t)$ for $t > 0$. This implies that if a family of maps is uniformly quasisymmetric, then the family of inverse maps is also uniformly quasisymmetric.

If X, Y, Z are metric spaces, and $h_1 \colon X \to Y$ is η_1-quasisymmetric and $h_2 \colon Y \to Z$ is η_2-quasisymmetric, then $h_2 \circ h_1$ is η-quasisymmetric, where $\eta = \eta_2 \circ \eta_1$. Hence the family of all compatible compositions of maps in two uniformly quasisymmetric families is again uniformly quasisymmetric.

Recall (see Section 4.1) that an arc α equipped with some metric d is called a *quasiarc* if there exists a quasisymmetry of the unit interval $[0, 1]$ onto (α, d). This is true if and only if (α, d) is doubling and there

exists a constant $K \geq 1$ such that $\operatorname{diam}_d(\gamma) \leq Kd(x,y)$, whenever $x, y \in \alpha$, $x \neq y$, and γ is the subarc of α with endpoints x and y (see Theorem 4.1).

A family of arcs is said to consist of *uniform quasiarcs* if there exists a homeomorphism $\eta \colon [0, \infty) \to [0, \infty)$ such that for each arc α in the family there exists an η-quasisymmetry $h \colon [0, 1] \to \alpha$. Similarly, a family of quasicircles is said to consist of *uniform quasicircles* if there exists a homeomorphism $\eta \colon [0, \infty) \to [0, \infty)$ such that for each quasicircle S in the family there exists an η-quasisymmetry $h \colon \partial \mathbb{D} \to S$. A family of quasicircles consists of uniform quasicircles if and only if the geometric conditions characterizing quasicircles, i.e., the doubling condition and the bounded turning condition, holds with uniform parameters. A similar statement is true for families of quasiarcs (see [**TV80**]).

We want to show that if the assumptions are as in Theorem 15.3, then all boundaries of tiles for $(f, \mathcal{C})$ are quasicircles and all edges for $(f, \mathcal{C})$ are quasiarcs. Actually, the family of all boundaries of tiles consists of uniform quasicircles and the family of all edges consists of uniform quasiarcs. One way to do this is to repeat the proof Theorem 15.3 and show that the geometric conditions characterizing quasiarcs and quasicircles are true for the edges and boundaries of tiles with uniform constants. We choose a different approach that is based on the following lemma which is of independent interest.

LEMMA 15.25. *Let $f \colon S^2 \to S^2$ be an expanding Thurston map, and $\mathcal{C} \subset S^2$ be an f-invariant Jordan curve with $\operatorname{post}(f) \subset \mathcal{C}$. Suppose that S^2 is equipped with a visual metric ϱ for f with expansion factor $\Lambda > 1$, and denote by $\mathbf{X}^n$ for $n \in \mathbb{N}_0$ the set of n-tiles for $(f, \mathcal{C})$. Then there exists a constant $C \geq 1$ with the following property:*

If $k, n \in \mathbb{N}_0$, $X^{n+k} \in \mathbf{X}^{n+k}$, and $x, y \in X^{n+k}$, then

$$(15.4) \qquad \frac{1}{C} \varrho(x, y) \leq \frac{\varrho(f^n(x), f^n(y))}{\Lambda^n} \leq C \varrho(x, y).$$

In particular, the family

$$\mathcal{F} = \{f^n | X^{n+k} : k, n \in \mathbb{N}_0, \ X^{n+k} \in \mathbf{X}^{n+k}\}$$

is uniformly quasisymmetric.

The distortion estimate (15.4) is closely related to the concept of a *conformal elevator* as introduced by Haïssinsky and Pilgrim [**HP09**, Theorem 2.2]. See also (16.1) in Theorem 16.3 for a related statement.

PROOF. In the following, all cells will be for $(f, \mathcal{C})$. Let $m = m_{f, \mathcal{C}}$ be as in Definition 8.1. We know by Definition 8.2 and by Lemma 8.7 (iii) that $\varrho(x, y) \asymp \Lambda^{-m(x,y)}$, whenever $x, y \in S^2$. If $n \in \mathbb{N}_0$,

then Lemma 8.7 (ii) implies that

$$m(f^n(x), f^n(y)) \geq m(x, y) - n,$$

and so

$$\varrho(f^n(x), f^n(y)) \lesssim \Lambda^n \varrho(x, y).$$

Here the implicit multiplicative constant is independent of x, y, and n.

To obtain an inequality in the other direction, let $x, y \in X^{n+k} \in \mathbf{X}^{n+k}$, where $n, k \in \mathbb{N}_0$. We may assume that $x \neq y$. Then by definition of $m(x, y)$ we have $n + k \leq m(x, y) < \infty$. Let $l := m(x, y) + 1 \in \mathbb{N}$. Since $l > n + k$, the $(n + k)$-tile X^{n+k} is subdivided by tiles of level l (Proposition 12.5 (iii)). Hence there exist l-tiles $X, Y \subset X^{n+k}$ with $x \in X$ and $y \in Y$. Then $X \cap Y = \emptyset$ by definition of $m(x, y)$. Let $X' := f^n(X)$ and $Y' := f^n(Y)$. Then by Proposition 5.17 (i) the sets X' and Y' are $(l - n)$-tiles. Since $f^n | X^{n+k}$ is injective, these tiles are disjoint, and we have $f^n(x) \in X'$ and $f^n(y) \in Y'$. So from Proposition 8.4 (i) we conclude that

$$\varrho(f^n(x), f^n(y)) \geq \mathrm{dist}_\varrho(X', Y') \gtrsim \Lambda^{-(l-n)} \asymp \Lambda^n \Lambda^{-m(x,y)} \asymp \Lambda^n \varrho(x, y).$$

Here the implicit multiplicative constants are again independent of x, y, and n. The other desired inequality follows.

Inequality (15.4) immediately implies that the family $\mathcal{F}$ is uniformly quasisymmetric. To see this, let $k, n \in \mathbb{N}_0$ and $X^{n+k} \in \mathbf{X}^{n+k}$. Then $f^n | X^{n+k}$ is a homeomorphism onto its image (see Proposition 5.17 (i)). Moreover, if $u, v, w \in X^{n+k}$, $u \neq w$, then by (15.4) we have

$$\frac{\varrho(f^n(u), f^n(v))}{\varrho(f^n(u), f^n(w))} \leq C^2 \frac{\varrho(u, v)}{\varrho(u, w)}.$$

Hence $f^n | X^{n+k}$ is η-quasisymmetric, where $\eta(t) = C^2 t$ for $t \geq 0$. Since η is independent of the chosen map, the family $\mathcal{F}$ is uniformly quasisymmetric. $\square$

PROPOSITION 15.26. *Let $f \colon S^2 \to S^2$ be an expanding Thurston map and $\mathcal{C} \subset S^2$ be an f-invariant Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}$. Suppose that S^2 is equipped with a visual metric for f, and for $n \in \mathbb{N}_0$ denote by $\mathbf{X}^n$ the set of n-tiles and by $\mathbf{E}^n$ the set of n-edges for $(f, \mathcal{C})$.*

Then the family $\{\partial X : n \in \mathbb{N}_0 \text{ and } X \in \mathbf{X}^n\}$ consists of uniform quasicircles and the family $\{e : n \in \mathbb{N}_0 \text{ and } e \in \mathbf{E}^n\}$ of uniform quasiarcs.

In particular, edges for $(f, \mathcal{C})$ are quasiarcs and the boundaries of all tiles are quasicircles.

PROOF. By Theorem 15.3 there exists a quasisymmetry $h \colon \partial \mathbb{D} \to \mathcal{C}$. Let X be an arbitrary tile for $(f, \mathcal{C})$, say an n-tile, where $n \in \mathbb{N}_0$.

Then $f^n|X$ is a homeomorphism of X onto the 0-tile $f^n(X)$ (Proposition 5.17 (i)), and so

$$f^n(\partial X) = \partial f^n(X) = \mathcal{C}.$$

By Lemma 15.25 the map $f^n|X$, and hence also the map $(f^n|X)^{-1}$, is a quasisymmetry. It follows that $(f^n|X)^{-1} \circ h$ is a quasisymmetric map from $\partial \mathbb{D}$ onto ∂X. Hence ∂X is a quasicircle. Actually, the family of these quasicircles ∂X is uniform, since the family of all relevant maps $(f^n|X)^{-1} \circ h$ is uniformly quasisymmetric as follows from Lemma 15.25.

The proof that the family $\{e : n \in \mathbb{N}_0 \text{ and } e \in \mathbf{E}^n\}$ consists of uniform quasiarcs runs along the same lines. First note that each 0-edge is a subarc of $\mathcal{C}$, and hence corresponds to a subarc of $\partial \mathbb{D}$ under the quasisymmetry h. Since this subarc can be mapped to the unit interval $[0, 1]$ by a bi-Lipschitz homeomorphism, each 0-edge is quasisymmetrically equivalent to $[0, 1]$ and hence a quasiarc.

Now let e be an arbitrary edge for $(f, \mathcal{C})$, say an n-edge, where $n \in \mathbb{N}_0$. Then $f^n|e$ is a homeomorphism of e onto the 0-edge $f^n(e)$ (Proposition 5.17 (i)). Moreover, there exists an n-tile X with $e \subset X$ (Lemma 5.9). Then $f^n|e$ is the restriction of the map $f^n|X$ to e, and it follows from Lemma 15.25 that $f^n|e$ is a quasisymmetry. Hence e is quasisymmetrically equivalent to a 0-edge and hence a quasiarc.

Lemma 15.25 actually implies that the family consisting all maps $f^n|e$ with $n \in \mathbb{N}_0$ and $e \in \mathbf{E}^n$ is uniformly quasisymmetric. So each edge is quasisymmetrically equivalent to a 0-edge by a quasisymmetry in a uniformly quasisymmetric family. Since there are only finitely many 0-edges, this implies that the family of all edges for $(f, \mathcal{C})$ consists of uniform quasiarcs. $\qquad\square$

A *quasidisk* is a closed topological disk (i.e., a topological cell of dimension 2) that is quasisymmetrically equivalent to the closed unit disk $\overline{\mathbb{D}}$. A family of closed topological disks is said to consist of *uniform quasidisks* if each disk X in the family can be mapped to $\overline{\mathbb{D}}$ by an η-quasisymmetry, where η is independent of X. It is a natural question whether the family $\{X : n \in \mathbb{N}_0, X \in \mathbf{X}^n\}$ of tiles obtained from an invariant curve as in the previous theorem actually consists of uniform quasidisks. This is true if and only if the expanding Thurston map f is topologically conjugate to a rational map without periodic critical points. One direction easily follows from Theorem 18.4 (iii) and Corollary 18.10 proved later.

For the other direction suppose that f and $\mathcal{C}$ are as in Proposition 15.26 and that the two 0-tiles equipped with a visual metric ϱ are quasidisks. Then one can show that (S^2, ϱ) is a quasisphere (this

requires the solution of a so-called *welding problem*). So by Theorem 18.1 (ii) the map f is topologically conjugate to a rational map without periodic critical points. We skip the details for this implication as we will not use the result.

CHAPTER 16

The combinatorial expansion factor

Suppose $f\colon S^2 \to S^2$ is a Thurston map with $\mathrm{post}(f) \geq 3$, and $\mathcal{C} \subset S^2$ is a Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}$. In Section 5.7 we introduced the quantity $D_n(f, \mathcal{C})$ as the minimal number of n-tiles for $(f, \mathcal{C})$ to form a connected set that joins opposite sides of $\mathcal{C}$ (see Definition 5.32 and (5.14)). In this chapter we study the asymptotic behavior of $D_n(f, \mathcal{C})$ as $n \to \infty$. We will see that for an expanding Thurston map, $D_n(f, \mathcal{C})$ grows at an exponential rate independent of $\mathcal{C}$.

PROPOSITION 16.1. *Let* $f\colon S^2 \to S^2$ *be an expanding Thurston map, and* $\mathcal{C} \subset S^2$ *be a Jordan curve with* $\mathrm{post}(f) \subset \mathcal{C}$. *Then the limit*

$$\Lambda_0(f) := \lim_{n \to \infty} D_n(f, \mathcal{C})^{1/n}$$

exists. Moreover, this limit is independent of $\mathcal{C}$ *and we have* $1 < \Lambda_0(f) < \infty$.

We call $\Lambda_0(f)$ the *combinatorial expansion factor* of f. Later we will see that $\Lambda_0(f) \leq \deg(f)^{1/2}$ (Proposition 20.1).

The combinatorial expansion factor is well-behaved under taking iterates and invariant under topological conjugacy.

PROPOSITION 16.2. *Let* $f\colon S^2 \to S^2$ *be an expanding Thurston map. Then the following statements are true:*

(i) $\Lambda_0(f^n) = \Lambda_0(f)^n$ *for* $n \in \mathbb{N}$.

(ii) *Suppose* $g\colon \widehat{S}^2 \to \widehat{S}^2$ *is an expanding Thurston map that is topologically conjugate to* f. *Then* $\Lambda_0(g) = \Lambda_0(f)$.

The main result of this chapter relates the combinatorial expansion factor to expansion factors of visual metrics.

THEOREM 16.3 (Visual metrics and their expansion factors). *Let* $f\colon S^2 \to S^2$ *be an expanding Thurston map, and* $\Lambda_0(f) \in (1, \infty)$ *be its combinatorial expansion factor. Then the following statements are true:*

(i) *If* Λ *is the expansion factor of a visual metric for* f, *then* $1 < \Lambda \leq \Lambda_0(f)$.

(ii) *Conversely, if $1 < \Lambda < \Lambda_0(f)$, then there exists a visual metric ϱ for f with expansion factor Λ. Moreover, the visual metric ϱ can be chosen to have the following additional property:*

For every $x \in S^2$ there exists a neighborhood U_x of x such that

$$(16.1) \qquad \varrho(f(x), f(y)) = \Lambda \varrho(x, y) \text{ for all } y \in U_x.$$

This statement shows that if Λ is the expansion factor of a visual metric for f, then $1 < \Lambda \leq \Lambda_0(f)$, but conversely, the existence of a visual metric with expansion factor Λ is only guaranteed for $1 < \Lambda < \Lambda_0(f)$. This statement is optimal, since a visual metric with expansion factor $\Lambda = \Lambda_0(f)$ need not exist in general. We will discuss an example at the end of this chapter (see Example 16.8). However, in the proof of Theorem 16.3 (ii) we will establish an existence statement that is somewhat stronger: if $\mathcal{C} \subset S^2$ is an f-invariant Jordan curve with $\operatorname{post}(f) \subset \mathcal{C}$ and $1 < \Lambda \leq D_1(f, \mathcal{C})$, then there exists a visual metric ϱ for f with expansion factor Λ (see (16.5)).

We now proceed to supply the proofs. We require some preparation and start with some lemmas.

LEMMA 16.4. *Let $n \in \mathbb{N}_0$, $f\colon S^2 \to S^2$ be a Thurston map with $\#\operatorname{post}(f) \geq 3$, and $\mathcal{C} \subset S^2$ be a Jordan curve with $\operatorname{post}(f) \subset \mathcal{C}$. If there exists a connected set $K \subset S^2$ that joins opposite sides of $\mathcal{C}$ and that can be covered by $M \in \mathbb{N}$ n-flowers for $(f, \mathcal{C})$, then $D_n(f, \mathcal{C}) \leq 4M$.*

PROOF. We first assume that $\#\operatorname{post}(f) = 3$. Let K be as in the statement. By picking a point from the intersection of K with each of the three 0-edges, we can find a set $\{x, y, z\} \subset K$ such that $\{x, y, z\}$ joins opposite sides of $\mathcal{C}$. Since K is connected and can be covered by M n-flowers, we can find n-vertices $v_1, \ldots, v_M \in S^2$ such that $x \in W^n(v_1)$, $y \in W^n(v_M)$ and $W^n(v_i) \cap W^n(v_{i+1}) \neq \emptyset$ for $i = 1, \ldots, M - 1$. Then it follows from Lemma 5.28 (ii) that there exists a chain of n-tiles $X_1, \ldots, X_{2M}$ joining x and y (recall the terminology from Definition 5.19). Similarly, there exists a chain $X'_1, \ldots, X'_{2M}$ of n-tiles joining x and z. The union K' of the n-tiles in these two chains is a connected set consisting of at most $4M$ n-tiles. It contains the set $\{x, y, z\}$ and hence joins opposite sides of $\mathcal{C}$. Thus $D_n(f, \mathcal{C}) \leq 4M$.

If $\#\operatorname{post}(f) \geq 4$, the proof is similar and easier. In this case we can find a set $\{x, y\} \subset K$ that joins opposite sides of $\mathcal{C}$. By the same argument as before, we get the bound $D_n(f, \mathcal{C}) \leq 2M$. $\square$

LEMMA 16.5. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map, and $\mathcal{C}, \widetilde{\mathcal{C}} \subset S^2$ Jordan curves with $\operatorname{post}(f) \subset \mathcal{C}, \widetilde{\mathcal{C}}$. Then*

$$(16.2) \qquad D_n(f, \mathcal{C}) \asymp D_{n+1}(f, \mathcal{C})$$

and

$$D_n(f, \mathcal{C}) \asymp D_n(f, \widetilde{\mathcal{C}}) \tag{16.3}$$

for all $n \in \mathbb{N}_0$, where $C(\asymp)$ is independent of n.

PROOF. Set $D_n = D_n(f, \mathcal{C})$ and $\widetilde{D}_n = D_n(f, \widetilde{\mathcal{C}})$ for $n \in \mathbb{N}_0$.

To show (16.2), we fix $n \in \mathbb{N}_0$ and pick a connected set K joining opposite sides of $\mathcal{C}$ that consists of D_n n-tiles for $(f, \mathcal{C})$. According to Lemma 5.37 (ii) we can cover K by MD_n $(n+1)$-flowers, where $M \in \mathbb{N}$ is independent of n. Hence by Lemma 16.4 we have $D_{n+1} \leq CD_n$, where $C = 4M$. An inequality in the opposite direction follows from a similar argument based on Lemma 5.37 (i) and Lemma 16.4.

To establish (16.3), we consider $\tilde{\delta}_0 = \delta_0(f, \widetilde{\mathcal{C}}) > 0$ defined as in (5.13) for f, $\widetilde{\mathcal{C}}$, and a base metric d on S^2. Since f is expanding, there exists $n_0 \in \mathbb{N}_0$ such that $\operatorname{diam}_d(X) < \tilde{\delta}_0/2$, whenever X is an n_0-tile for $(f, \mathcal{C})$.

We can find a compact connected set $\widetilde{K}$ joining opposite sides of $\widetilde{\mathcal{C}}$ that consists of $\widetilde{D}_n$ n-tiles for $(f, \widetilde{\mathcal{C}})$. Then $\operatorname{diam}_d(\widetilde{K}) \geq \tilde{\delta}_0$ and so $\widetilde{K}$ contains two points with $d(x, y) \geq \tilde{\delta}_0$. There exist n_0-tiles X and Y for $(f, \mathcal{C})$ such that $x \in X$ and $y \in Y$. By choice of n_0 we have $X \cap Y = \emptyset$, and so $\widetilde{K}$ joins n_0-tiles for $(f, \mathcal{C})$ that are disjoint. Hence $f^{n_0}(\widetilde{K})$ joins opposite sides of $\mathcal{C}$ by Lemma 5.35. Every n-tile for $(f, \widetilde{\mathcal{C}})$ can be covered by M n-flowers for $(f, \mathcal{C})$, where M only depends on $\mathcal{C}$ and $\widetilde{\mathcal{C}}$ (Lemma 5.38). This and Lemma 5.29 imply that if $n \geq n_0$, then we can cover $f^{n_0}(\widetilde{K})$ by $M\widetilde{D}_n$ $(n - n_0)$-flowers for $(f, \mathcal{C})$.

So by Lemma 16.4 we have

$$D_{n-n_0} \leq 4M\widetilde{D}_n,$$

and by the first part of the proof there is a constant $C_1 < \infty$ such that

$$D_n \leq C_1^{n_0} D_{n-n_0} \leq 2MC_1^{n_0}\widetilde{D}_n.$$

If $n < n_0$ we get a similar bound from the inequalities $D_n \leq 2\deg(f)^{n_0}$ and $\widetilde{D}_n \geq 1$. It follows that there exists a constant C independent of n such that

$$D_n \leq C\widetilde{D}_n$$

for all $n \in \mathbb{N}_0$. An inequality in the opposite direction is obtained by reversing the roles of $\mathcal{C}$ and $\widetilde{\mathcal{C}}$ and using an estimate analogous to (16.2) for $\widetilde{D}_n$. $\square$

PROOF OF PROPOSITION 16.1. A consequence of (16.3) is that if $\mathcal{C}, \widetilde{\mathcal{C}} \subset S^2$ are Jordan curves with $\operatorname{post}(f) \subset \mathcal{C}, \widetilde{\mathcal{C}}$ and the sequence

$\{D_n(f,\mathcal{C})^{1/n}\}$ converges as $n \to \infty$, then $\{D_n(f,\widetilde{\mathcal{C}})^{1/n}\}$ also converges and has the same limit. So if the limit exists, it does not depend on $\mathcal{C}$.

To show existence, we may impose additional assumptions on $\mathcal{C}$; namely by Theorem 15.1, we may assume that $\mathcal{C}$ is invariant for some iterate $F = f^N$ of f. Since F is also an expanding Thurston map (Lemma 6.4), it follows from Lemma 8.6 and Lemma 12.9 that the limit

$$\Lambda_0(F,\mathcal{C}) := \lim_{n\to\infty} D_n(F,\mathcal{C})^{1/n}$$

exists and that $\Lambda_0(F,\mathcal{C}) \in (1,\infty)$.

Since the n-tiles for $(F,\mathcal{C})$ are precisely the (nN)-tiles for $(f,\mathcal{C})$ (see Proposition 5.17 (vii)), we have $D_{nN}(f,\mathcal{C}) = D_n(F,\mathcal{C})$ for all $n \in \mathbb{N}_0$, and so

$$D_{nN}(f,\mathcal{C})^{1/(nN)} = D_n(F,\mathcal{C})^{1/(nN)} \to \Lambda_0(f) := \Lambda_0(F,\mathcal{C})^{1/N} \in (1,\infty)$$

as $n \to \infty$. If we combine this with (16.2), it easily follows that that $D_n(f,\mathcal{C})^{1/n} \to \Lambda_0(f)$ as $n \to \infty$. The statement follows. $\square$

PROOF OF PROPOSITION 16.2. (i) If $F = f^n$ is an iterate of f, then, as was pointed out in the previous proof, we have

$$D_k(F,\mathcal{C}) = D_{nk}(f,\mathcal{C})$$

whenever $k \in \mathbb{N}_0$ and $\mathcal{C}$ is a Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}$. This implies

$$\Lambda_0(f^n) = \Lambda_0(F) = \lim_{k\to\infty} D_k(F,\mathcal{C})^{1/k} = \lim_{k\to\infty} D_{nk}(f,\mathcal{C})^{1/k} = \Lambda_0(f)^n.$$

(ii) By assumption there exists a homeomorphism $h\colon S^2 \to \widehat{S}^2$ such that $h \circ f = g \circ h$. Pick a Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}$ and let $\widehat{\mathcal{C}} = h(\mathcal{C})$. Then $\widehat{\mathcal{C}}$ is a Jordan curve that contains $\mathrm{post}(g)$, and, as in the proof of Proposition 8.8, we have

$$\mathcal{D}^n(g,\widehat{\mathcal{C}}) = \{h(c) : c \in \mathcal{D}^n(f,\mathcal{C})\}$$

for $n \in \mathbb{N}_0$. This implies that $D_n(f,\mathcal{C}) = D_n(g,\widehat{\mathcal{C}})$ for all $n \in \mathbb{N}_0$ and so

$$\Lambda_0(g) = \lim_{k\to\infty} D_n(g,\widehat{\mathcal{C}})^{1/n} = \lim_{n\to\infty} D_n(f,\mathcal{C})^{1/n} = \Lambda_0(f)$$

as desired. $\square$

We now proceed to prove Theorem 16.3, the main result of this chapter. So let $f\colon S^2 \to S^2$ be an expanding Thurston map. We fix a Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}$, and let $D_k = D_k(f,\mathcal{C})$ for $k \in \mathbb{N}_0$. In the following, cells will be for $(f,\mathcal{C})$. In Proposition 16.1 the combinatorial expansion factor $\Lambda_0(f)$ was defined, and we proved that $1 < \Lambda_0(f) < \infty$.

The proof of the first part of Theorem 16.3 is easy.

PROOF OF THEOREM 16.3 (i). Suppose ϱ is a visual metric for f with expansion factor Λ. Then there exists a constant $C \geq 1$ such that

$$\operatorname{diam}(X) \leq C\Lambda^{-k}$$

for all k-tiles (Proposition 8.4 (ii)). Let $\delta_0 = \delta_0(f, \mathcal{C}) > 0$ be defined as in (5.13) for f, $\mathcal{C}$, and the metric ϱ.

For each $k \in \mathbb{N}_0$ there exists a connected set joining opposite sides of $\mathcal{C}$ that consists of D_k k-tiles. Hence

(16.4) $$\delta_0 \leq \operatorname{diam}(K) \leq CD_k\Lambda^{-k}.$$

Taking the k-th root here and letting $k \to \infty$, we conclude that $\Lambda \leq \Lambda_0(f)$ as desired. $\qquad\square$

It remains to prove part (ii). For a given expansion factor $\Lambda \in (1, \Lambda_0(f))$ we have to construct a visual metric that satisfies (16.1). We have already encountered visual metrics with this local expansion property; indeed, one can show that for the map h in Section 1.3 the metric given by (1.2) has the property (16.1) with $\Lambda = \Lambda_0(h) = 2$.

The construction in the general case is much more difficult and involved than the general construction of visual metrics in Section 8.2. We will first do this under additional assumptions on the map and then for the general case.

Construction of the metric under additional assumptions. Let $\mathcal{C} \subset S^2$ be the Jordan curve with $\operatorname{post}(f) \subset \mathcal{C}$ used to define our cell decompositions (and $D_n(f, \mathcal{C})$). We now assume in addition that $\mathcal{C}$ is f-invariant and that $\Lambda \in (1, \Lambda_0(f)]$ satisfies

(16.5) $$\Lambda \leq D_1 = D_1(f, \mathcal{C}).$$

In this case, we will now construct a visual metric ϱ with expansion factor Λ that satisfies (16.1). Note that we do allow $\Lambda = \Lambda_0(f)$ here. We first introduce some terminology.

Recall from Definition 5.19 that a *tile chain* P is a finite sequence of tiles $X_1, \ldots, X_N$, where $X_j \cap X_{j+1} \neq \emptyset$ for $j = 1, \ldots, N - 1$. Here we do not require the tiles to be of the same level.

Define the *weight* of a k-tile X^k to be

(16.6) $$w(X^k) := \Lambda^{-k},$$

and the *w-length* of a tile chain P consisting of the tiles $X_1, \ldots, X_N$ as

$$\operatorname{length}_w(P) := \sum_{j=1}^{N} w(X_j).$$

Now for $x, y \in S^2$ we define

$$(16.7) \qquad \varrho(x, y) := \inf_{P} \operatorname{length}_w(P),$$

where the infimum is taken over all tile chains P joining x and y. Obviously, such tile chains exist and the infimum can be taken over simple tile chains P.

LEMMA 16.6. *The distance function ϱ defined in* (16.7) *is a visual metric with expansion factor* Λ.

PROOF. Symmetry and the triangle inequality immediately follow from the definition of ϱ. Since f is expanding, we also have $\varrho(x, x) = 0$ for $x \in S^2$.

Let $x, y \in S^2$ with $x \neq y$ be arbitrary, and define $m = m(x, y) = m_{f, \mathcal{C}}(x, y)$ (see Definition 8.1). Then there exist m-tiles X and Y with $x \in X$, $y \in Y$ and $X \cap Y \neq \emptyset$. So X, Y is a tile chain joining x and y, and thus

$$\varrho(x, y) \leq w(X) + w(Y) = 2\Lambda^{-m}.$$

In order to prove that ϱ is a metric and is visual for f, it remains to establish a lower bound $\varrho(x, y) \geq (1/C)\Lambda^{-m}$ for a suitable constant C independent of x and y.

Pick $(m + 1)$-tiles X' and Y' with $x \in X'$ and $y \in Y'$. Then $X' \cap Y' = \emptyset$ by definition of m. Every tile chain joining x and y contains a simple tile chain joining X' and Y'.

So let P be a simple tile chain joining X' and Y', and suppose it consists of the tiles $X_1, \ldots, X_N$. Let $k \in \mathbb{N}_0$ be the largest level of any tile in P. If $k \leq m + 1$, then we get the favorable estimate

$$(16.8) \qquad \operatorname{length}_w(P) \geq \Lambda^{-k} \geq \Lambda^{-m-1}.$$

Otherwise, $k > m + 1$. We want to show that then we can replace the k-tiles in P with $(k - 1)$-tiles without increasing the w-length of the tile chain (the construction is illustrated in Figure 16.1).

To see this, set $X_0 = X'$, $X_{N+1} = Y'$, and let X_i, where $1 \leq i \leq N$, be the first k-tile in P. Since P is a simple tile chain joining X' and Y', the tile X_i is not contained in X_{i-1} and so it has to meet ∂X_{i-1}. Since the level of X_{i-1} is $< k$, we can find a $(k - 1)$-edge $e \subset \partial X_{i-1}$ with $e \cap X_i \neq \emptyset$. Here and below we use the fact that $\mathcal{C}$ is f-invariant, and so cells of any level are subdivided by cells of higher level. Every $(k - 1)$-tile meets e or is contained in the complement of the edge flower $W^{k-1}(e)$ (see Lemma 5.31 (iii)). Since tiles of level $\leq k - 1$ are subdivided into tiles of level $k - 1$, this implies also that every tile of level $\leq k - 1$ meets e or is contained in the complement of $W^{k-1}(e)$.
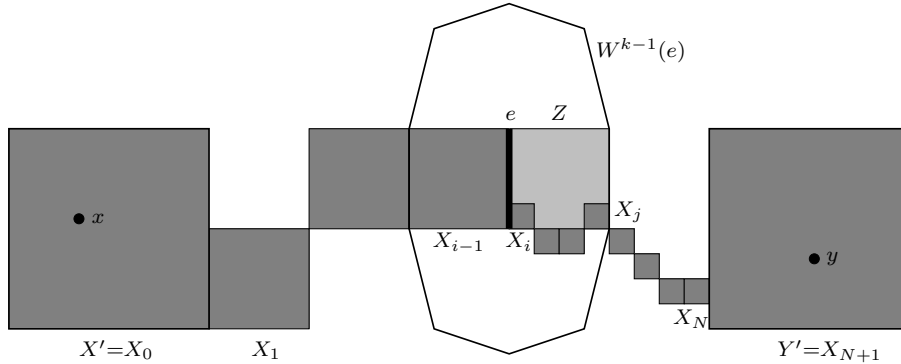
FIGURE 16.1. Replacing k-tiles with $(k-1)$-tiles.

Now P is simple and so no tile in the "tail" $X_{i+1}, \ldots, X_N, X_{N+1}$ meets e. Let $j' \in \mathbb{N}$ be the largest number such that $i \leq j' \leq N$ and all tiles $X_i, \ldots, X_{j'}$ are k-tiles. Then $X_{j'+1}$ has level $\leq k-1$. Since this tile does not meet e, it is contained in $S^2 \setminus W^{k-1}(e)$, and so the tiles $X_i, \ldots, X_{j'}$ form a chain of k-tiles joining e and $S^2 \setminus W^{k-1}(e)$. Let $j \in \mathbb{N}$ be the smallest number with $i \leq j \leq j'$ such that X_j meets the complement of $W^{k-1}(e)$. Then $X_i, \ldots, X_j$ is a chain P^k of k-tiles joining e and $S^2 \setminus W^{k-1}(e)$. In particular, P^k joins two disjoint $(k-1)$-cells as follows from the definition of an edge flower (see Definition 5.30). Moreover, X_j is the only tile in the chain P^k that meets the complement of $W^{k-1}(e)$.

Since P^k joins disjoint $(k-1)$-cells, it follows from Lemma 5.36 that P^k has at least D_1 elements, and so by (16.5),

$$\operatorname{length}_w(P^k) \geq D_1 \Lambda^{-k} \geq \Lambda^{-k+1}.$$

Let Z be the unique $(k-1)$-tile with $Z \supset X_j$. Then $Z \cap X_{j+1} \neq \emptyset$. We also have $Z \cap e \neq \emptyset$; for otherwise $X_j \subset Z \subset S^2 \setminus W^{k-1}(e)$. Then $j > i$ and X_{j-1} meets X_j and so the complement of $W^{k-1}(e)$ contradicting the definition of j. So $Z \cap X_{i-1} \supset Z \cap e \neq \emptyset$. Thus we can replace the subchain P^k of P by the single $(k-1)$-tile Z to obtain a chain P' joining X' and Y'. It satisfies

$$\operatorname{length}_w(P') = \operatorname{length}_w(P) - \operatorname{length}_w(P^k) + w(Z) \leq \operatorname{length}_w(P).$$

By passing to a subchain of P' we can find a simple tile chain P'' joining X' and Y' that contains fewer k-tiles than P and satisfies $\operatorname{length}_w(P'') \leq \operatorname{length}_w(P)$.

Continuing this process, we can remove all k-tiles from the tile chain joining X' and Y' without increasing its w-length. If $k-1 > m+1$, we can repeat the process and remove the $(k-1)$-tiles without increasing

the w-length, etc. In the end we obtain a tile chain $\widetilde{P}$ joining X' and Y' that contains no tiles of level $> m + 1$ and satisfies $\text{length}_w(\widetilde{P}) \leq \text{length}_w(P)$. Thus

$$\text{length}_w(P) \geq \text{length}_w(\widetilde{P}) \geq \Lambda^{-m-1}.$$

This together with the previous estimate (16.8) implies

$$\varrho(x, y) \geq \Lambda^{-m-1}.$$

This is an inequality as desired, and so ϱ is indeed a visual metric with expansion factor Λ. In particular, this implies that ϱ induces the given topology on S^2 (see Proposition 8.3 (ii)).   □

LEMMA 16.7. *The visual metric ϱ as defined in (16.7) has the expansion property (16.1).*

PROOF. We first show that

(16.9) $\varrho(f(x), f(y)) \leq \Lambda \varrho(x, y),$

for all $x, y \in S^2$ with $\varrho(x, y) < 1$.

Indeed, suppose $x, y \in S^2$ are arbitrary points with $\varrho(x, y) < 1$. Let P be an arbitrary tile chain that joins x and y and suppose that it consists of the tiles $X_1, \ldots, X_N$. We may assume in addition that P satisfies $\text{length}_w(P) < 1$. Then P does not contain 0-tiles and hence $f(X_1), \ldots, f(X_N)$ is a tile chain joining $f(x)$ and $f(y)$. Denoting the latter chain by $f(P)$, we have

$$\text{length}_w(f(P)) = \Lambda \, \text{length}_w(P).$$

Taking the infimum over all such tile chains P, we obtain the desired inequality (16.9).

For an inequality in the other direction we now consider two cases for $x \in S^2$.

Case 1: $x \notin \text{crit}(f)$. Then we can find an open neighborhood U of x on which f is a homeomorphism. Then $U' = f(U)$ is an open set containing $f(x)$. We can choose $\epsilon > 0$ and $\delta \in (0, 1)$ such that $B_\varrho(x, \delta) \subset U$, $B_\varrho(f(x), \epsilon) \subset U'$, and $f(B_\varrho(x, \delta)) \subset B_\varrho(f(x), \epsilon)$.

Define $U_x = B_\varrho(x, \delta)$, and let $y \in U_x$ be arbitrary. Then we have $\varrho(f(x), f(y)) < \epsilon$. Consider a tile chain P' joining $f(x)$ and $f(y)$ whose w-length is close enough to $\varrho(f(x), f(y))$ so that $\text{length}_w(P') < \epsilon$. By definition of the metric ϱ, for every point z that belongs to a tile in P', we have $\varrho(f(x), z) \leq \text{length}_w(P') < \epsilon$. Hence P' lies in $B_\varrho(f(x), \epsilon) \subset U'$.

It follows that $(f|U)^{-1}$ is defined on every tile X' in P', and so by Lemma 5.18 (i) the Jordan region $X = (f|U)^{-1}(X')$ is a tile contained

in U. If k is the level of X', then $k+1$ is the level of X. By considering these images of tiles in P' under $(f|U)^{-1}$, we get a tile chain P joining x and y with $\text{length}_w(P) = (1/\Lambda)\,\text{length}_w(P')$. Taking the infimum over such P', we obtain

$$(16.10) \qquad\qquad \varrho(x,y) \le (1/\Lambda)\varrho(f(x),f(y)).$$

Case 2: $x \in \text{crit}(f)$. Then $x \in f^{-1}(\text{post}(f))$, and so x is a 1-vertex. Consider the flower $U = W^1(x)$, and its image $U' = f(W^1(x)) = W^0(f(x))$. These are open neighborhoods of x and $f(x)$, respectively, and the map $f|U \setminus \{x\}$ is a (non-branched) covering map of $U \setminus \{x\}$ onto $U' \setminus \{f(x)\}$ (all this follows from the considerations in the proof of Lemma 5.24). Again we can choose $\epsilon > 0$ and $\delta \in (0,1)$ such that $B_\varrho(x,\delta) \subset U$, $B_\varrho(f(x),\epsilon) \subset U'$, and $f(B_\varrho(x,\delta)) \subset B_\varrho(f(x),\epsilon)$.

Define $U_x = B_\varrho(x,\delta)$, and let $y \in U_x$ be arbitrary. In order to show (16.10), we may assume $x \ne y$. Then $\varrho(f(x),f(y)) < \epsilon$ and $f(x) \ne f(y)$. Consider a tile chain P' joining $f(x)$ and $f(y)$ consisting of tiles $X'_1, \ldots, X'_N$. We can make the further assumptions that X'_1 is the only tile in this chain that contains $f(x)$ and that $\text{length}_w(P')$ is close enough to $\varrho(f(x),f(y))$ such that $\text{length}_w(P') < \epsilon$. As before this implies that P' lies in U'. We now choose a path $\gamma\colon [0,N] \to U'$ with the following properties:

(i) $\gamma(0) = f(x)$, $\gamma(N) = f(y)$ and $\gamma(t) \ne f(x)$ for $t \ne 0$,

(ii) $\gamma([i-1,i]) \subset X'_i$ for $i = 1, \ldots, N$,

(iii) $\gamma(i - 1/2) \in \text{int}(X'_i)$ for $i = 1, \ldots, N$.

Since the tiles X'_i are Jordan regions, such a path γ can easily be obtained by first running from $f(x)$ in X'_1 to an interior point of X'_1, then to a point in $X'_1 \cap X'_2$, then to an interior point of X'_2, etc., and finally to $f(y) \ne f(x)$ in X'_N. Since X'_1 is the only tile in P' containing $f(x)$, this can be done so that the path never meets $f(x)$ except in its initial point.

There exists a lift α of this path (under f) with endpoints x and y, i.e., a path $\alpha\colon [0,N] \to U$ with $\alpha(0) = x$, $\alpha(N) = y$, and $f \circ \alpha = \gamma$. To obtain α, lift $\gamma|(0,N]$ under the covering map $f|U \setminus \{x\}$ such that the lift ends at y (see Lemma A.6 and the subsequent discussion), and note that the lift has a unique continuous extension to $[0,N]$ by choosing x to be its initial point.

Using this lift α, we can construct a lift of our tile chain P' as follows. Consider a tile X'_i in P' and let k_i be its level. Set $p_i := \alpha(i - 1/2)$ and $p'_i := \gamma(i - 1/2)$. Then $f(p_i) = p'_i \in \text{int}(X'_i)$. By Lemma 5.18 (ii) there exists a unique $(k_i + 1)$-tile X_i with $p_i \in X_i$ and $f(X_i) = X'_i$.

Note that

$$\gamma((0, N]) \subset U' \setminus \{f(x)\} = W^0(f(x)) \setminus \{f(x)\} \subset S^2 \setminus \text{post}(f)$$

and that the map

$$f \colon S^2 \setminus f^{-1}(\text{post}(f)) \to S^2 \setminus \text{post}(f)$$

is a covering map. This implies that $\alpha|[i-1, i]$ is the unique lift of $\gamma|[i-1, i]$ with $\alpha(i-1/2) = p_i$ (see Lemma A.6 (i)). On the other hand, the path $\beta_i = (f|X_i)^{-1} \circ (\gamma|[i-1, i])$ is also a lift of $\gamma|[i-1, i]$ under f with $\beta_i(i-1/2) = p_i$ by definition of X_i. Hence $\beta_i = \alpha|[i-1, i]$ and so $\alpha([i-1, i]) \subset X_i$. It follows that $x = \alpha(0) \in X_1$, $y = \alpha(N) \in X_N$, and $X_i \cap X_{i+1} \supset \{\alpha(i)\} \neq \emptyset$ for $i = 1, \ldots, N-1$.

Therefore, the tiles $X_1, \ldots, X_N$ form a tile chain P joining x and y. The level of each tile in P exceeds the level of the corresponding tile in P' by exactly 1. Hence $\text{length}_w(P) = (1/\Lambda) \text{length}_w(P')$. Taking the infimum over such P', we again obtain inequality (16.10)

Combining (16.9) and (16.10), we see that every point $x \in S^2$ has a neighborhood U_x such that (16.1) holds.   □

This concludes the proof for the existence of the visual metric ϱ with the desired properties as in Theorem 16.3 (ii) under the additional assumptions that $\mathcal{C}$ is f-invariant and that (16.5) holds. We now consider the general case.

PROOF OF THEOREM 16.3 (ii). Suppose that $1 < \Lambda < \Lambda_0(f)$. We can choose an iterate $F = f^n$ of f such that F has an F-invariant Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) = \text{post}(F) \subset \mathcal{C}$ (Theorem 15.1). Note that $D_k(f, \mathcal{C})^{1/k} \to \Lambda_0(f) > \Lambda$ by Proposition 16.1. Hence if n is sufficiently large, which we may assume by passing to an iterate of F, we also have

$$D_1(F, \mathcal{C}) = D_n(f, \mathcal{C}) \geq \Lambda^n.$$

This means that F is an expanding Thurston map that satisfies the condition (16.5) for the F-invariant Jordan curve $\mathcal{C}$. As we have seen above, this guarantees the existence of a metric with the desired properties. We call this metric d in order to distinguish it from the metric ϱ that we are trying to construct. Then d is a visual metric for F with expansion factor Λ^n, and for each $x \in S^2$ there exists an open neighborhood U_x of x such that

(16.11) $d(F(x), F(y)) = d(f^n(x), f^n(y)) = \Lambda^n d(x, y)$

for all $y \in U_x$.

We now define ϱ as

$$(16.12) \qquad \varrho(x, y) = \frac{1}{n} \sum_{i=0}^{n-1} \Lambda^{-i} d(f^i(x), f^i(y))$$

for $x, y \in S^2$. It is clear that ϱ is a metric on S^2.

Property (16.1) for the metric ϱ follows from the corresponding property (16.11) for d with the same sets U_x, $x \in S^2$; indeed, if $x \in S^2$ and $y \in U_x$ then by (16.11) we have

$$\varrho(f(x), f(y)) = \frac{1}{n} \sum_{i=0}^{n-1} \Lambda^{-i} d(f^{i+1}(x), f^{i+1}(y))$$

$$= \frac{1}{n} \left(\Lambda \sum_{i=0}^{n-2} \Lambda^{-(i+1)} d(f^{i+1}(x), f^{i+1}(y)) + \Lambda d(x, y) \right)$$

$$= \Lambda \frac{1}{n} \sum_{i=0}^{n-1} \Lambda^{-i} d(f^i(x), f^i(y)) = \Lambda \varrho(x, y).$$

It remains to show that ϱ is a visual metric for f with expansion factor Λ. Let $m = m_{f,\mathcal{C}}$ and $m_F = m_{F,\mathcal{C}}$ be as in Definition 8.1. Since d is a visual metric for F with expansion factor Λ^n, we have

$$d(x, y) \asymp \Lambda^{-nm_F(x,y)} \asymp \Lambda^{-m(x,y)}$$

for all $x, y \in S^2$ by Lemma 8.7 (iv). Hence

$$\varrho(x, y) \geq \frac{1}{n} d(x, y) \gtrsim \Lambda^{-m(x,y)}.$$

Moreover, by Lemma 8.7 (ii) we have

$$m(f^i(x), f^i(y)) \geq m(x, y) - i$$

and so

$$d(f^i(x), f^i(y)) \asymp \Lambda^{-m(f^i(x), f^i(y))} \leq \Lambda^i \Lambda^{-m(x,y)}$$

for all $i \in \mathbb{N}_0$. Hence

$$\varrho(x, y) \lesssim \frac{1}{n} \sum_{i=0}^{n-1} \Lambda^{-m(x,y)} = \Lambda^{-m(x,y)}.$$

It follows that $\varrho(x, y) \asymp \Lambda^{-m(x,y)}$ for all $x, y \in S^2$, where $C(\asymp)$ is independent of x and y. This shows that ϱ is a visual metric for f with expansion factor Λ. $\qquad\qquad\qquad\square$

We conclude the chapter with an example showing that for an expanding Thurston map f one can in general not expect the existence of a visual metric with expansion factor $\Lambda = \Lambda_0(f)$.

EXAMPLE 16.8. The example is a Lattès-type map as in Example 3.20. We consider the crystallographic group G consisting of all isometries g on $\mathbb{R}^2$ of the form $u \mapsto g(u) = \pm u + \gamma$, where $\gamma \in \mathbb{Z}^2$. Then the quotient space $S^2 := \mathbb{R}^2/G$ is a 2-sphere. Let $\Theta \colon \mathbb{R}^2 \to S^2 = \mathbb{R}^2/G$ be the quotient map. We know that Θ is induced by G and so $\Theta(u_1) = \Theta(u_2)$ for $u_1, u_2 \in \mathbb{R}^2$ if and only if there exists $g \in G$ such that $u_2 = g(u_1)$.

As described in Example 3.20, one may view $S^2 = \mathbb{R}^2/G$ as a pillow obtained by folding the rectangle $R = [0,1] \times [0,1/2]$ along the line $\ell = \{(x,y) \in \mathbb{R}^2 : x = 1/2\}$ and identifying the boundaries of the squares $Q = [0,1/2] \times [0,1/2]$ and $Q' = [1/2,1] \times [0,1/2]$ under this operation. In particular, $\mathcal{C} := \Theta(\partial Q) = \Theta(\partial Q')$ is a Jordan curve containing the four critical values of Θ which are the four vertices of the pillow.

Let

$$A = \begin{pmatrix} 2 & 2 \\ 0 & 2 \end{pmatrix}.$$

Note that then for $n \in \mathbb{N}$ we have

(16.13) $A^n = \begin{pmatrix} 2^n & n2^n \\ 0 & 2^n \end{pmatrix}$ and $A^{-n} = \begin{pmatrix} 2^{-n} & -n2^{-n} \\ 0 & 2^{-n} \end{pmatrix}.$

We consider the map $A \colon \mathbb{R}^2 \to \mathbb{R}^2$, $u \in \mathbb{R}^2 \mapsto Au$, given by left-multiplication of $u \in \mathbb{R}^2$ (considered as a column vector) by the matrix A. For simplicity we use the same notation for the matrix A and this linear map on $\mathbb{R}^2$.

Then there exists a unique Lattès-type map $f \colon S^2 \to S^2$ such that the diagram

$$\begin{array}{ccc} \mathbb{R}^2 & \xrightarrow{\ A\ } & \mathbb{R}^2 \\ {\scriptstyle\Theta}\downarrow & & \downarrow{\scriptstyle\Theta} \\ S^2 & \xrightarrow{\ f\ } & S^2 \end{array}$$

commutes (see Proposition 3.21). The map f has signature $(2,2,2,2)$ and its four postcritical points are the critical values of Θ. In particular, $\mathrm{post}(f) \subset \mathcal{C}$. Proposition 6.11 implies that the Thurston map f is expanding.

Since A induces a map on the quotient $\mathbb{R}^2/G$, it is G-equivariant and so we have $A \circ g \circ A^{-1} \in G$ whenever $g \in G$ (Lemma A.22). This implies that if $n \in \mathbb{N}_0$ and $\alpha \in \frac{1}{2}\mathbb{Z}^2$, then Θ is injective on the parallelogram $A^{-n}(\alpha + Q)$. Indeed, if $u_1, u_2 \in \alpha + Q$ and $\Theta(A^{-n}(u_1)) = \Theta(A^{-n}(u_2))$, then there exists $g \in G$ such that $A^{-n}(u_2) = g(A^{-n}(u_1))$. Then $u_2 = h(u_1)$ with $h := A^n \circ g \circ A^{-n} \in G$, and so $\Theta(u_2) = \Theta(u_1)$;

since Θ is injective on the square $\alpha + Q$, we conclude that $u_2 = u_1$. The injectivity of Θ on $A^{-n}(\alpha + Q)$ follows.

This implies that the set $X^n = \Theta(A^{-n}(\alpha + Q))$ is a Jordan region whenever $n \in \mathbb{N}_0$ and $\alpha \in \frac{1}{2}\mathbb{Z}^2$. The n-tiles for $(f, \mathcal{C})$ are precisely the sets X^n of this form. This is clear for $n = 0$. If $n \in \mathbb{N}_0$ is arbitrary, then $f^n \circ \Theta = \Theta \circ A^n$. Since $\Theta|(\alpha + Q)$ is a homeomorphism of $\alpha + Q$ onto the 0-tile $\Theta(\alpha + Q)$, it follows that $f^n|X^n$ is a homeomorphism of the Jordan region X^n onto a 0-tile. So by Lemma 5.18 (i), the set X^n is indeed an n-tile. Since these sets X^n cover $\Theta(\mathbb{R}^2) = S^2$, there are no other n-tiles.

It follows from (16.13) that each set $A^{-n}(\alpha + Q) \subset \mathbb{R}^2$ is a parallelogram congruent to the parallelogram $P_n = \{xu_n + yv_n : 0 \le x, y \le 1\}$ spanned by the vectors $u_n = \frac{1}{2^{n+1}}(1, 0)$ and $v_n = \frac{1}{2^{n+1}}(-n, 1)$. Thus $\operatorname{diam}(A^{-n}(\alpha + Q)) \asymp n2^{-n}$, where $C(\asymp)$ is independent of n. This implies that if we equip the pillow S^2 with the locally Euclidean metric (obtained by pushing the Euclidean metric forward by Θ), then for each n-tile X^n we have $\operatorname{diam}(X^n) \asymp n2^{-n}$ (so by Proposition 8.4 (ii)) this metric on S^2 is *not* a visual metric for f). We conclude that $D_n := D_n(f, \mathcal{C}) \gtrsim 2^n/n$ for $n \in \mathbb{N}$.

If $X_i := \Theta(A^{-n}(\alpha_i + Q))$ with $\alpha_i = (0, -i/2)$ for $i = 1, \ldots, N := \lceil 2^n/n \rceil$, then $X_1, \ldots, X_N$ is a chain of n-tiles joining opposite sides of $\mathcal{C}$. Hence $D_n \le N \lesssim 2^{-n}/n$ for $n \in \mathbb{N}$. It follows that $D_n \asymp 2^n/n$, and so

$$\Lambda_0(f) = \lim_{n \to \infty} D_n^{1/n} = 2.$$

If Λ is the expansion factor of a visual metric, then for all $n \in \mathbb{N}$ we must have $1 \lesssim D_n \Lambda^{-n} \asymp 2^n \Lambda^{-n}/n$ (see (16.4) in the proof Theorem 16.3 (i)). It follows that a visual metric for f with expansion factor $\Lambda = \Lambda_0(f) = 2$ does not exist.

CHAPTER 17

The measure of maximal entropy

In this chapter we investigate the measure of maximal entropy of an expanding Thurston map. We will first review some definitions and the necessary background from measure-theoretic dynamics in Section 17.1. Our goal in Section 17.2 is then to prove the following statement.

THEOREM 17.1. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map. Then there exists a unique measure ν_f of maximal entropy for f. The map f is mixing for ν_f.*

Since mixing implies ergodicity, it follows that f is ergodic with respect to ν_f.

On our way to prove the previous theorem, we will be able to compute the topological entropy $h_{top}(f)$ of f.

COROLLARY 17.2. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map. Then $h_{top}(f) = \log(\deg(f))$.*

The corresponding statements of Theorem 17.1 and Corollary 17.2 for rational maps (not necessarily postcritically finite) where proved by Lyubich in [**Ly83**]. For expanding Thurston maps without periodic critical points, Theorem 17.1 can be derived from general results due to Haïssinsky and Pilgrim [**HP09**]. We will present a different approach. We consider an iterate $F = f^n$ of our given expanding Thurston map f that has an F-invariant Jordan curve $\mathcal{C}$ with $\mathrm{post}(F) \subset \mathcal{C}$. Then in the cell decomposition $\mathcal{D}^k(F, \mathcal{C})$, $k \in \mathbb{N}_0$, generated by F and $\mathcal{C}$ each cell is subdivided by cells of higher level. This will allow us to construct a specific F-invariant probability measure ν_F that assigns to each k-tile mass proportional to $\deg(F)^{-k}$, where the proportionality factor only depends on the color of the tile (see Proposition 17.11).

By using coverings by tiles, it is also easy to obtain the estimate $h_{top}(F) \leq \log(\deg(F))$ (see the proof of Lemma 17.8). On the other hand, the 1-tiles in $\mathcal{D}^1(F, \mathcal{C})$ form a measurable partition of S^2 that generates (in a suitable sense) the Borel σ-algebra on S^2 (see Lemma 17.6). This allows us to explicitly compute the measure-theoretic entropy $h_{\nu_F}(F)$ of F with respect to ν_F as $h_{\nu_F}(F) = \log(\deg(F))$ (see

the proof of Proposition 17.11). It follows that ν_F is a measure of maximal entropy for F. One then shows that this measure ν_F is actually f-invariant and the unique measure of maximal entropy for f; this is formulated in Theorem 17.12 which immediately implies Theorem 17.1.

Our main point here is to give a rather concrete and elementary description of the measure of maximal entropy of an expanding Thurston map and to establish some of its basic properties. We will not touch upon many interesting related questions such as equidistribution of preimages and periodic points or more general measures such as equilibrium measures for suitable potentials. These subjects are thoroughly investigated in [**Li16a, Li15b**].

17.1. Review of measure-theoretic dynamics

In this section we briefly discuss the necessary concepts related to topological and measure-theoretic entropy. For more background on these topics see [**KH95, Wa82**].

In the following, (X, d) is a compact metric space, and $g\colon X \to X$ is a continuous map. For $n \in \mathbb{N}$ and $x, y \in X$ we define

$$(17.1) \qquad d_g^n(x, y) = \max\{d(g^k(x), g^k(y)) : k = 0, \dots, n-1\}.$$

Then d_g^n is a metric on X. Let $D(g, \epsilon, n)$ be the minimal number of sets whose d_g^n-diameter is at most $\epsilon > 0$ and whose union is equal to X. One can show that the limit

$$h(g, \epsilon) := \lim_{n \to \infty} \frac{1}{n} \log(D(g, \epsilon, n))$$

exists [**KH95**, Lemma 3.1.5]. Obviously, the quantity $h(g, \epsilon)$ is non-increasing in ϵ. One defines the *topological entropy* of g (see [**KH95**, Section 3.1.b]) as

$$h_{top}(g) := \lim_{\epsilon \to 0} h(g, \epsilon) \in [0, \infty].$$

If one uses another metric d' on X, then one obtains the same quantity for $h_{top}(g)$ if d' induces the same topology on X as d [**KH95**, Proposition 3.1.2]. The topological entropy is also well-behaved under iteration. Indeed, if $n \in \mathbb{N}$, then $h_{top}(g^n) = nh_{top}(g)$ [**KH95**, Proposition 3.1.7 (3)].

We denote by $\mathcal{B}$ the σ-algebra of all Borel sets on X. A measure on X is understood to be a Borel measure, i.e., one defined on $\mathcal{B}$. We call a measure μ on X *g-invariant* if

$$(17.2) \qquad\qquad \mu(g^{-1}(A)) = \mu(A)$$

for all $A \in \mathcal{B}$. Note that by continuity of g, we have $g^{-1}(A) \in \mathcal{B}$ whenever $A \in \mathcal{B}$. We denote by $\mathcal{M}(X, g)$ the set of all g-invariant Borel probability measures on X.

If μ is a probability measure on a compact metric space X, then it is *regular*. This means that for every $\epsilon > 0$ and every Borel set $A \subset X$ there exists a compact set $K \subset A$ with $\mu(A \setminus K) < \epsilon$ (*inner regularity*) and an open set $U \subset X$ with $A \subset U$ and $\mu(U \setminus A) < \epsilon$ (*outer regularity*). See [**Ru87**, Theorem 2.18] for a more general result that contains this statement as a special case.

A *semi-algebra* $\mathcal{S}$ is a family of subsets of X satisfying the following conditions: (i) $\emptyset \in \mathcal{S}$, (ii) $A \cap B \in \mathcal{S}$, whenever $A, B \in \mathcal{S}$, and (iii) $X \setminus A$ is a finite union of disjoint sets in $\mathcal{S}$, whenever $A \in \mathcal{S}$. A semi-algebra $\mathcal{S}$ *generates* a σ-algebra $\mathcal{A}$ on X if $\mathcal{A}$ is the smallest σ-algebra containing $\mathcal{S}$.

Let $\mathcal{S}$ be a semi-algebra generating $\mathcal{B}$. If μ and ν are two probability measures on X and $\mu(A) = \nu(A)$ for all $A \in \mathcal{S}$, then $\mu = \nu$. Similarly, in order to show that a probability measure μ is g-invariant it is enough to verify (17.2) for all sets A in $\mathcal{S}$ (see [**Wa82**, proof of Theorem 1.1, p. 20] for the simple argument on how to verify these statements).

Let $\mu \in \mathcal{M}(X, g)$. Then we say that g is *ergodic* for μ (or μ is *ergodic* for g) if for each set $A \in \mathcal{B}$ with $g^{-1}(A) = A$ we have $\mu(A) = 0$ or $\mu(A) = 1$. The map g is called *mixing* for μ if

$$(17.3) \qquad \lim_{n \to \infty} \mu(g^{-n}(A) \cap B) = \mu(A)\mu(B)$$

for all $A, B \in \mathcal{B}$. It is easy to see that if g is mixing for μ, then g is also ergodic.

To establish mixing, one only has to verify (17.3) for sets A and B in a semi-algebra generating $\mathcal{B}$ ([**Wa82**, Theorem 1.17 (iii)]; note that the terminology in [**Wa82**] slightly differs from ours). If $\mu, \nu \in \mathcal{M}(X, g)$, g is ergodic for μ, and ν is absolutely continuous with respect to μ, then $\nu = \mu$ [**Wa82**, Remark (1), p. 153].

Our next goal is to define the measure-theoretic entropy of g for a measure μ. We will follow [**KH95**, Section 4.3] with slight differences in notation and terminology (see also [**Wa82**, Chapter 4]).

Let $\mu \in \mathcal{M}(X, g)$. A *measurable partition* ξ for (X, μ) is a countable collection $\xi = \{A_i : i \in I\}$ of sets in $\mathcal{B}$ such that $\mu(A_i \cap A_j) = 0$ for $i, j \in I$, $i \neq j$ and

$$\mu\left(X \setminus \bigcup_{i \in I} A_i\right) = 0.$$

Here I is a countable (i.e., finite or countably infinite) index set. The *symmetric difference* of two sets $A, B \subset X$ is defined as

$$A \Delta B = (A \setminus B) \cup (B \setminus A).$$

Two measurable partitions ξ and η for (X, μ) are called *equivalent* if there exists a bijection between the sets of positive measure in ξ and the sets of positive measure in η such that corresponding sets have a symmetric difference of vanishing μ measure. Roughly speaking, this means that the partitions are the same up to sets of measure zero.

Let $\xi = \{A_i : i \in I\}$ and $\eta = \{B_j : j \in J\}$ be measurable partitions of (X, μ). Then

$$\xi \vee \eta := \{A_i \cap B_j : i \in I, \, j \in J\}$$

is also a measurable partition, called the *join* of ξ and η. The join of finitely many measurable partitions is defined similarly.

Let

$$g^{-1}(\xi) := \{g^{-1}(A_i) : i \in I\}$$

and for $n \in \mathbb{N}$ define

(17.4) $$\xi_g^n := \xi \vee g^{-1}(\xi) \vee \cdots \vee g^{-(n-1)}(\xi).$$

The *entropy* of ξ is

$$H_\mu(\xi) := \sum_{i \in I} \mu(A_i) \log(1/\mu(A_i)) \in [0, \infty].$$

Here it is understood that the function $\phi(x) = x \log(1/x)$ is continuously extended to 0 by setting $\phi(0) = 0$.

One can show that if $H_\mu(\xi) < \infty$, then for a given map g the quantities $H_\mu(\xi_g^n)$, $n \in \mathbb{N}_0$, are *subadditive* in the sense that

$$H_\mu(\xi_g^{n+k}) \le H_\mu(\xi_g^n) + H_\mu(\xi_g^k)$$

for all $k, n \in \mathbb{N}_0$ [**KH95**, Proposition 4.3.6]. This implies that

$$h_\mu(g, \xi) := \lim_{n \to \infty} \frac{1}{n} H_\mu(\xi_g^n) \in [0, \infty)$$

exists and we have

$$h_\mu(g, \xi) = \inf_{n \in \mathbb{N}} \frac{1}{n} H_\mu(\xi_g^n)$$

([**Wa82**, Theorem 4.9]; see also the proof of Lemma 12.9). The quantity $h_\mu(g, \xi)$ is called the *(measure-theoretic) entropy of g relative to ξ*. The *(measure-theoretic) entropy* of g for μ is defined as

(17.5) $$h_\mu(g) = \sup\{h_\mu(g, \xi) : \; \xi \text{ is a measurable partition}$$
$$\text{of } (X, \mu) \text{ with } H_\mu(\xi) < \infty\}.$$

In this definition it is actually enough to take the supremum over all finite measurable partitions ξ (this easily follows from "Rokhlin's inequality" [**KH95**, Proposition 4.3.10 (4)]).

We call a finite measurable partition ξ a *generator* for (g, μ) if the following condition is true: Let $\mathcal{A}$ be the smallest σ-algebra containing all sets in the partitions ξ_g^n, $n \in \mathbb{N}$. Then we require that for each Borel set $B \in \mathcal{B}$ there exists a set $A \in \mathcal{A}$ such that $\mu(A \Delta B) = 0$.

If for every set $B \in \mathcal{B}$ and for every $\epsilon > 0$, there exists $n \in \mathbb{N}$ and a union A of sets in ξ_g^n with $\mu(A \Delta B) < \epsilon$, then ξ is a generator for (g, μ). If ξ is a generator, then $h_\mu(g) = h_\mu(g, \xi)$ by the Kolmogorov-Sinai theorem [**Wa82**, Theorem. 4.17].

If $\mu \in \mathcal{M}(g, X)$ and $n \in \mathbb{N}$, then [**KH95**, Proposition 4.3.16 (4)]

$$(17.6) \qquad h_\mu(g^n) = n h_\mu(g).$$

If $\alpha \in [0, 1]$ and $\nu \in \mathcal{M}(g, X)$ is another measure, then [**Wa82**, Theorem 8.1]

$$h_{\alpha\mu + (1-\alpha)\nu}(g) = \alpha h_\mu(g) + (1 - \alpha) h_\nu(g).$$

The topological entropy is related to the measure-theoretic entropy by the so-called variational principle. It states that [**Wa82**, Theorem 8.6]

$$(17.7) \qquad h_{top}(g) = \sup\{h_\mu(g) : \mu \in \mathcal{M}(g, X)\}.$$

A measure $\mu \in \mathcal{M}(g, X)$ for which $h_{top}(g) = h_\mu(g)$ is called a *measure of maximal entropy.*

Let $\widetilde{X}$ be another compact metric space. If μ is a measure on X and $\varphi \colon X \to \widetilde{X}$ is continuous, then the *push-forward* $\varphi_* \mu$ of μ by φ is the measure given by $\varphi_* \mu(A) = \mu(\varphi^{-1}(A))$ for all Borel sets $A \subset \widetilde{X}$. Note that if $\widetilde{X} = X$, then μ is φ-invariant is and only if $\varphi_* \mu = \mu$.

Suppose $\widetilde{g} \colon \widetilde{X} \to \widetilde{X}$ is a continuous map, and $\mu \in \mathcal{M}(X, g)$ and $\widetilde{\mu} \in \mathcal{M}(\widetilde{X}, \widetilde{g})$. Then the dynamical system $(\widetilde{X}, \widetilde{g}, \widetilde{\mu})$ is called a *(topological) factor* of (X, g, μ) if there exists a continuous and surjective map $\varphi \colon X \to \widetilde{X}$ such that $\varphi_* \mu = \widetilde{\mu}$ and $\widetilde{g} \circ \varphi = \varphi \circ g$. Then we have the following commutative diagram:

$$
\begin{array}{ccc}
(X, \mu) & \xrightarrow{\;g\;} & (X, \mu) \\
\downarrow{\scriptstyle\varphi} & & \downarrow{\scriptstyle\varphi} \\
(\widetilde{X}, \widetilde{\mu}) & \xrightarrow{\;\widetilde{g}\;} & (\widetilde{X}, \widetilde{\mu}).
\end{array}
$$

In this case, $h_{\widetilde{\mu}}(\widetilde{g}) \leq h_\mu(g)$ [**KH95**, Proposition 4.3.16].

If μ and ν are Borel probability measures on a compact metric space X, then μ has a unique Lebesgue decomposition with respect to ν. More precisely, μ can uniquely be written as $\mu = \mu_a + \mu_s$, where μ_a and μ_s are finite Borel measures on X such that μ_a is absolutely continuous and μ_s is singular with respect to ν (see [**Ru87**, Theorem 6.10]).

We require the following fact.

LEMMA 17.3. *Let X be a compact metric space, and $g\colon X \to X$ be a continuous map. Suppose μ and ν are Borel probability measures on X, and $\mu = \mu_a + \mu_s$ is the Lebesgue decomposition of μ with respect to ν. If μ and ν are g-invariant, then μ_a and μ_s are also g-invariant.*

PROOF. We claim that the measure $g_*\mu_a$ is absolutely continuous with respect to ν. For this is suffices to show that if $A \subset X$ is a Borel set with $\nu(A) = 0$, then $g_*\mu_a(A) = 0$. Now ν is f-invariant and so $\nu(f^{-1}(A)) = \nu(A) = 0$ for such a set A. Since μ_a is absolutely continuous with respect to ν, this implies that $g_*\mu_a(A) = \mu_a(g^{-1}(A)) = 0$ as desired.

Similarly, we claim that $g_*\mu_s$ is singular with respect to ν. Since μ_s is singular with respect to ν, there exists a Borel set $B \subset X$ with $\nu(B) = 0$ and $\mu_s(X \setminus B) = 0$. Then $\mu_a(B) = 0$ and $\mu_a(g^{-1}(B)) = g_*\mu_a(B) = 0$. This combined with the g-invariance of μ implies that

$$\mu_s(g^{-1}(B)) = \mu(g^{-1}(B)) = \mu(B) = \mu_s(B).$$

It follows that

$$g_*\mu_s(X \setminus B) = \mu_s(X \setminus g^{-1}(B)) = \mu_s(X) - \mu_s(g^{-1}(B))$$
$$= \mu_s(X) - \mu_s(B) = \mu_s(X \setminus B) = 0.$$

This shows that B is a set of full $g_*\mu_s$-measure and has ν-measure zero. We conclude that $g_*\mu_s$ is indeed singular with respect to ν.

Now $\mu = g_*\mu = g_*\mu_a + g_*\mu_s$. By what we have seen, here $g_*\mu_a$ is absolutely continuous and $g_*\mu_s$ singular with respect to ν. The uniqueness of the Lebesgue decomposition of μ implies that $g_*\mu_a = \mu_a$ and $g_*\mu_s = \mu_s$. The statement follows. $\square$

17.2. Construction of the measure of maximal entropy

In this section we fix an expanding Thurston map $f\colon S^2 \to S^2$. Our goal is to describe a measure of maximal entropy for f and show its uniqueness. We will freely use the notation and the results discussed in the previous section.

By Theorem 15.1 we can find a sufficiently high iterate $F = f^n$ of f that has an F-invariant Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) =$

$\mathrm{post}(F) \subset \mathcal{C}$. Then F is also an expanding Thurston map (Lemma 6.4). In the following, we consider the cell decompositions $\mathcal{D}^k = \mathcal{D}^k(F, \mathcal{C})$ for $k \in \mathbb{N}_0$. A *cell* is a cell in any of the cell decompositions $\mathcal{D}^k$, $k \in \mathbb{N}_0$, and the terms *tiles*, *edges*, and *vertices* are used in a similar way. As usual, we denote by $\mathbf{X}^k$ and $\mathbf{E}^k$ the set of k-tiles and k-edges for $(F, \mathcal{C})$, respectively. By Proposition 12.5 the cell decomposition $\mathcal{D}^{m+k}$ is a refinement of $\mathcal{D}^k$ for $m, k \in \mathbb{N}_0$, and so cells are subdivided by cells of higher level.

We denote by X_{w}^0 and X_{b}^0 the two 0-tiles, and color the tiles for $(F, \mathcal{C})$ as in Lemma 5.21. In particular, X_{w}^0 is colored white and X_{b}^0 is colored black.

For $k \in \mathbb{N}_0$ let w_k be the number of white and b_k be the number of black k-tiles contained in X_{w}^0, and similarly let w_k' and b_k' be the number of white and black k-tiles contained in X_{b}^0. Then it follows from the discussion after Lemma 5.21 that

$$(17.8) \qquad w_k + w_k' = b_k + b_k' = \deg(F)^k.$$

Note that $b_1, w_1' \neq 0$. Indeed, suppose that $b_1 = 0$, for example. Then X_{w}^0 contains only white 1-tiles. Let $X \subset X_{\mathsf{w}}^0$ be such a 1-tile, $e \subset X$ be a 1-edge with $e \subset \partial X$, and Y be the other 1-tile containing e. Then Y is black and so $Y \subset X_{\mathsf{b}}^0$. Hence

$$e \subset X \cap Y \subset X_{\mathsf{w}}^0 \cap X_{\mathsf{b}}^0 = \partial X_{\mathsf{w}}^0.$$

Since ∂X is a union of 1-edges, it follows that $\partial X \subset \partial X_{\mathsf{w}}^0$. As X_{w}^0 and X are Jordan regions and $X \subset X_{\mathsf{w}}^0$, this is only possible if $X = X_{\mathsf{w}}^0$. Hence X_{w}^0 is a 1-tile and $F|X_{\mathsf{w}}^0$ is a homeomorphism of X_{w}^0 onto itself. Applying Lemma 5.18 (i) repeatedly, we see that X_{w}^0 is a k-tile for each $k \in \mathbb{N}_0$. This is impossible, because F is expanding and so the diameters of k-tiles approach 0 as $k \to \infty$.

Define

$$(17.9) \qquad w := \frac{b_1}{b_1 + w_1'}, \quad b := \frac{w_1'}{b_1 + w_1'}.$$

Then $w, b > 0$ and $w + b = 1$. It follows from (17.8) for $k = 1$ that the matrix

$$(17.10) \qquad A = \begin{pmatrix} w_1 & b_1 \\ w_1' & b_1' \end{pmatrix}$$

has the eigenvalues $\lambda_1 = \deg(F)$ and $\lambda_2 = w_1 - b_1$ with respective eigenvectors

$$v_1 = \begin{pmatrix} w \\ b \end{pmatrix} \text{ and } v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Here $|\lambda_2| = |w_1 - b_1| < \lambda_1 = \deg(F)$. Indeed, since $1 \leq b_1 \leq \deg(F)$ and $0 \leq w_1 \leq \deg(F)$, we otherwise have $w_1 = 0$ and $b_1 = \deg(F) \geq 2$. Then the white 0-tile contains only black 1-tiles. Arguing as in the discussion above, we see that then there can be only one such tile, and so $b_1 = 1$. This is a contradiction.

The existence of a largest positive eigenvalue λ_1 for A with a corresponding eigenvector with all positive coordinates is an instance of the Perron-Frobenius theorem ([**KH95**, Theorem 1.9.11]).

Let $k, l, m \in \mathbb{N}_0$ with $m \geq l \geq k$ be arbitrary. The map F^k preserves colors of tiles, i.e., if X^m is an m-tile, then $F^k(X^m)$ is an $(m-k)$-tile with the same color as X^m. Moreover, if Y^l is an l-tile, then it follows from Lemma 5.18 (i) that the map $F^k|Y^l$ induces a bijection $X^m \mapsto F^k(X^m)$ between the m-tiles contained in Y^l and the $(m-k)$-tiles contained in the $(l-k)$-tile $Y^{l-k} := F^k(Y^l)$.

If we use this for $m = k+1$ and $l = k$, then we see that a white k-tile contains w_1 white and b_1 black $(k+1)$-tiles, and similarly each black k-tile contains w_1' white and b_1' black $(k+1)$-tiles. This leads to the identity

$$(17.11) \qquad \begin{pmatrix} w_{k+1} & b_{k+1} \\ w_{k+1}' & b_{k+1}' \end{pmatrix} = \begin{pmatrix} w_k & b_k \\ w_k' & b_k' \end{pmatrix} \begin{pmatrix} w_1 & b_1 \\ w_1' & b_1' \end{pmatrix},$$

for $k \in \mathbb{N}_0$. This implies

$$A^k = \begin{pmatrix} w_k & b_k \\ w_k' & b_k' \end{pmatrix}$$

for $k \in \mathbb{N}_0$. The following lemma is another consequence of (17.11).

LEMMA 17.4. *For all $k \in \mathbb{N}_0$ we have*

$$w_k = w \deg(F)^k + b(w_1 - b_1)^k, \quad b_k = w \deg(F)^k - w(w_1 - b_1)^k,$$
$$w_k' = b \deg(F)^k - b(w_1 - b_1)^k, \quad b_k' = b \deg(F)^k + w(w_1 - b_1)^k.$$

Since $|w_1 - b_1| < \deg(F)$, the terms with $\deg(F)^k$ in these identities are the main terms for large k.

PROOF. This follows from (17.8), (17.9), and (17.11) by induction. $\square$

The next lemma provides an important connection between cells for $(F, \mathcal{C})$ and general Borel sets.

LEMMA 17.5. *Let $\mathcal{S}$ be the set collection consisting of the empty set and the interiors of all cells for $(F, \mathcal{C})$. Then $\mathcal{S}$ is a semi-algebra generating the Borel σ-algebra $\mathcal{B}$ on S^2.*

PROOF. We first show that $\mathcal{S}$ is a semi-algebra by verifying conditions (i)–(iii) of a semi-algebra.

Condition (i): By definition of $\mathcal{S}$ we have $\emptyset \in \mathcal{S}$.

Condition (ii): Let $A, B \in \mathcal{S}$. In order to show that $A \cap B \in \mathcal{S}$, we may assume that $A = \text{int}(\sigma)$ and $B = \text{int}(\tau)$, where σ is a k-cell and τ is an l-cell, and $k \geq l$. Let $p \in \text{int}(\tau)$ be arbitrary. Then by Lemma 5.2 there exists a unique k-cell c with $p \in \text{int}(c)$. Since $\mathcal{D}^k$ is a refinement of $\mathcal{D}^l$, there exists a unique l-cell τ' with $\text{int}(c) \subset \text{int}(\tau')$ (see Lemma 5.7). Then τ and τ' are both l-cells containing the point p in their interiors. This implies that $\tau' = \tau$, and so $\text{int}(c) \subset \text{int}(\tau)$.

It follows that $\text{int}(\tau)$ can be written as a disjoint union of interiors of k-cells. This implies that either $A \cap B = \text{int}(\sigma)$ or $A \cap B = \emptyset$. In both cases, $A \cap B \in \mathcal{S}$.

Condition (iii): Let $A \in \mathcal{S}$ be arbitrary. If $A = \emptyset$, then $S^2 \setminus A = S^2$, and so $S^2 \setminus A$ is equal to the disjoint union of the interiors of the 0-cells, and so a disjoint union of elements in $\mathcal{S}$.

If $A = \text{int}(\tau)$ where τ is a k-cell, then $S^2 \setminus A$ is the disjoint union of the interiors of all k-cells distinct from τ. Again $S^2 \setminus A$ is a disjoint union of sets in $\mathcal{S}$.

So $\mathcal{S}$ is indeed a semi-algebra.

$\mathcal{S}$ *generates* $\mathcal{B}$: Let $\mathcal{A}$ be the smallest σ-algebra containing $\mathcal{S}$. Since $\mathcal{S}$ consists of Borel sets, we have $\mathcal{A} \subset \mathcal{B}$. So in order to show that $\mathcal{A} = \mathcal{B}$ is suffices to establish that $U \in \mathcal{A}$ for each open subset U of S^2.

Let $p \in U$ be arbitrary. Then for each $k \in \mathbb{N}_0$ the point p is contained in the interior of some k-cell. Since F is expanding, the diameters of k-cells approach 0 as $k \to \infty$. Hence there exists a cell c with $p \in \text{int}(c) \subset U$. This implies that U is a union of elements in $\mathcal{S}$. Since for each $k \in \mathbb{N}_0$ there are only finitely many k-cells, the set $\mathcal{S}$ is countable, and so U is a countable union of elements in $\mathcal{S}$. Hence $U \in \mathcal{A}$ as desired. $\square$

In the following, we set

$$E^\infty = \bigcup_{k \in \mathbb{N}_0} F^{-k}(\mathcal{C}).$$

Then E^∞ is a Borel set. Proposition 5.17 (iii) (applied to the map F) implies that E^∞ is equal to the union of all edges. Since every vertex is contained in an edge, the set E^∞ also contains all vertices. Moreover, we have

(17.12) $$F^{-1}(E^\infty) = E^\infty.$$

Indeed, note that $F^{-1}(\mathcal{C}) \supset \mathcal{C}$ and so

$$F^{-1}(E^\infty) = F^{-1}\left(\bigcup_{k \in \mathbb{N}_0} F^{-k}(\mathcal{C})\right) = \bigcup_{k \in \mathbb{N}_0} F^{-(k+1)}(\mathcal{C})$$

$$= \bigcup_{k \in \mathbb{N}} F^{-k}(\mathcal{C}) = \mathcal{C} \cup \bigcup_{k \in \mathbb{N}} F^{-k}(\mathcal{C}) = E^\infty.$$

LEMMA 17.6. *Let μ be an F-invariant probability measure on S^2 with $\mu(E^\infty) = 0$. Then for each $k \in \mathbb{N}$ the set $\mathbf{X}^k$ of k-tiles forms a measurable partition of S^2. It is equivalent to the partition ξ_F^k where $\xi = \mathbf{X}^1$. Moreover, $\xi = \mathbf{X}^1$ is a generator for (F, μ).*

PROOF. Note that $\mu(E^\infty) = 0$ implies that all edges are sets of μ-measure zero. Since every vertex is contained in an edge, we also have $\mu(\{v\}) = 0$ for all vertices v. The k-tiles cover S^2, and two distinct k-tiles have only edges or vertices, i.e., a set of μ-measure zero, in common. Hence $\mathbf{X}^k$ is a measurable partition of S^2.

Let $X \in \mathbf{X}^k$ be arbitrary. Then for $i = 1, \ldots, k$ there exist unique i-tiles X^i with $X = X^k \subset X^{k-1} \subset \ldots \subset X^1$. Put $Y_i = F^{i-1}(X^i)$ for $i = 1, \ldots, k$. Then $Y_1, \ldots, Y_k$ are 1-tiles. We claim that

(17.13) $\qquad X = Y_1 \cap F^{-1}(Y_2) \cap \cdots \cap F^{-(k-1)}(Y_k).$

To verify this, denote the right hand side in this equation by $\widetilde{X}$. Then it is clear that $X \subset \widetilde{X}$. We verify $X = \widetilde{X}$ by inductively showing that for any point $x \in \widetilde{X}$ we have $x \in X^i$ for $i = 1, \ldots, k$, and so $x \in X^k = X$.

Indeed, since $\widetilde{X} \subset Y_1 = X^1$ this is clear for $i = 1$. Suppose $x \in X^{i-1}$ for some i with $2 \leq i \leq k$. To complete the inductive step, we have to show $x \in X^i$. Note that $x \in \widetilde{X} \subset F^{-(i-1)}(Y_i)$ and so $F^{i-1}(x) \in Y_i$. The map $F^{i-1}|X^{i-1}$ is a homeomorphism of X^{i-1} onto the 0-tile $F^{i-1}(X^{i-1})$. Moreover, $x \in X^{i-1}$, $X^i \subset X^{i-1}$, and $F^{i-1}(x) \in Y_i = F^{i-1}(X^i)$. Hence by injectivity of F^{i-1} on X^{i-1} we have $x \in X^i$ as desired.

Equation (17.13) shows that every element in $\mathbf{X}^k$ belongs to ξ_F^k. This implies that the measurable partitions $\mathbf{X}^k$ and ξ_F^k are equivalent (ξ_F^k may contain additional sets, but they have to be of measure zero).

To establish that $\xi = \mathbf{X}^1$ is a generator, let $B \subset S^2$ be an arbitrary Borel set and $\epsilon > 0$. By what we have seen, it is enough to show that there exists $k \in \mathbb{N}$ and a union A of k-tiles such that $\mu(A \Delta B) < \epsilon$.

By regularity of μ there exist a compact set $K \subset A$ and an open set $U \subset S^2$ with $K \subset A \subset U$ and $\mu(U \setminus K) < \epsilon$. Since the diameters of tiles goes to 0 uniformly with the levels of the tiles, we can choose $k \in \mathbb{N}$ so large that every k-tile that meets K is contained in the open

neighborhood U of K. Define

$$A = \bigcup \{X \in \mathbf{X}^k : X \cap K \neq \emptyset\}.$$

Then $K \subset A \subset U$. This implies $A \Delta B \subset U \setminus K$, and so

$$\mu(A \Delta B) \leq \mu(U \setminus K) < \epsilon$$

as desired. The proof is complete. $\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

The last lemma allows us to easily compute the entropy of F-invariant measures μ once we know that $\mu(E^\infty) = 0$. The following fact will be useful for verifying this.

LEMMA 17.7. *There exists $1 \leq L < \deg(F)$ such that for all $k, m \in \mathbb{N}_0$ and each m-edge e there exists a collection M of $(m+k)$-tiles with $\#M \leq CL^k$ such that e is contained in the interior of the set $\bigcup_{X \in M} X$. Here C is independent of k.*

The total number of $(m+k)$-tiles is $2 \deg(F)^{m+k}$. So the lemma says that for large k, the m-edge e can be covered by a substantially smaller number of $(m+k)$-tiles.

PROOF. It follows from Lemma 8.11 and Proposition 8.4 (ii) that one can find $k_0 \in \mathbb{N}$ such that for every s-tile X, $s \in \mathbb{N}_0$, there exist two $(s+k_0)$-tiles Y and Z, one white and one black, with $Y \subset \operatorname{int}(X)$ and $Z \subset \operatorname{int}(X)$.

Every white s-tile contains w_{k_0} white and b_{k_0} black $(s+k_0)$-tiles, and every black s-tile contains w'_{k_0} white and b'_{k_0} black $(s+k_0)$-tiles. By (17.8) we also know that $w_{k_0} + w'_{k_0} = b_{k_0} + b'_{k_0} = \deg(F)^{k_0}$.

Now let e be an arbitrary m-edge. For each $l \in \mathbb{N}_0$ we will define certain collections T_l of $(m+lk_0)$-tiles whose union contains e in its interior. We will denote the number of white tiles in T_l by N_l^{w}, the number of black tiles in T_l by N_l^{b}, and define $N_l = \max\{N_l^{\mathrm{w}}, N_l^{\mathrm{b}}\}$. Then the number of tiles in T_l is bounded by $2N_l$.

Let T_0 be the set of all m-tiles that meet e. Then the union of the tiles in T_0 is the closure of the edge flower of e and so it contains e in its interior.

Suppose the collection T_l has been constructed. Then we subdivide each of the tiles U in T_l into $(m+(l+1)k_0)$-tiles and remove one white and one black $(m+(l+1)k_0)$-tile from the interior of U. We define T_{l+1} as the collection of all tiles obtained in this way from tiles in T_l. Since $\operatorname{int}(U) \cap e = \emptyset$ for each $U \in T_l$, the union of the tiles in T_{l+1} still contains e in its interior. Then for the number of white tiles in T_{l+1} we

have the estimate

$$
\begin{aligned}
N_{l+1}^{\mathtt{w}} &= N_l^{\mathtt{w}}(w_{k_0} - 1) + N_l^{\mathtt{b}}(w_{k_0}' - 1) \\
&\leq N_l(w_{k_0} + w_{k_0}' - 2) = N_l(\deg(F)^{k_0} - 2).
\end{aligned}
$$

Similarly,

$$
N_{l+1}^{\mathtt{b}} \leq N_l(\deg(F)^{k_0} - 2),
$$

and so

$$
N_{l+1} \leq N_l(\deg(F)^{k_0} - 2).
$$

Let

$$
L := (\deg(F)^{k_0} - 2)^{1/k_0} < \deg(F).
$$

Then

$$
\#T_l \leq 2N_l \leq 2N_0 L^{k_0 l}
$$

is a bound for the total number of tiles in T_l.

Now let $k \in \mathbb{N}_0$ be arbitrary. Then we can choose $l \in \mathbb{N}_0$ such that $k \leq lk_0 < k + k_0$. For each $(m + lk_0)$-tile U in T_l we can pick an $(m + k)$-tile that contains U. Let M be that collection of all $(m + k)$-tiles obtained in this way. Then the union of all tiles in M contains e in its interior and we have

$$
\#M \leq \#T_l \leq 2N_0 L^{k_0 l} \leq 2N_0 L^{k_0} L^k = C L^k,
$$

where $C = 2N_0 L^{k_0}$. The claim follows. $\qquad\square$

The constant C in the previous lemma depends on e. If we require the weaker property that the collection M of $(m+k)$-tiles only covers e, then we can choose the collection so that $\#M \leq C L^k$ with a constant C independent of e. Indeed, in this case, we can choose T_0 to consist of the two m-tiles X and Y, one white and one black, that contain e in their boundary. Then $N_0 = 1$ and this leads to an inequality of the desired type with a constant C independent of e.

In the next lemma we obtain an upper bound for the topological entropy of f.

LEMMA 17.8. $h_{top}(f) \leq \log(\deg(f))$.

We will verify later that actually $h_{top}(f) = \log(\deg(f))$ (see the proof of Corollary 17.2).

PROOF. Since $h_{top}(F) = nh_{top}(f)$ and $\deg(F) = \deg(f)^n$, it suffices to show that $h_{top}(F) \leq \log(\deg(F))$.

To show that $h_{top}(F) \leq \log(\deg(F))$, we fix a base metric d on S^2 and let $\epsilon > 0$ be arbitrary. Since F is expanding, we can find $k_0 \in \mathbb{N}_0$ such that $\operatorname{diam}(X) \leq \epsilon$ whenever $X \in \mathbf{X}^k$ for $k > k_0$.

Now if $k \in \mathbb{N}$ and $X \in \mathbf{X}^{k+k_0}$ is arbitrary, then $F^i(X)$ is a tile of level $k - i + k_0 > k_0$ for $i = 0, 1 \ldots, k-1$, and so $\text{diam}(F^i(X)) \leq \epsilon$. This implies that the diameter of X with respect to the metric d_F^k (see (17.1)) is $\leq \epsilon$. Since the number of $(k + k_0)$-tiles is equal to $2 \deg(F)^{k+k_0}$ and these tiles form a cover of S^2, it follows that $D(F, \epsilon, k) \leq 2 \deg(F)^{k+k_0}$, and so $h(\epsilon, F) \leq \log(\deg(F))$. Letting $\epsilon \to 0$ we conclude $h_{top}(F) \leq \log(\deg(F))$ as desired. $\qquad\square$

Since the curve $\mathcal{C}$ is F-invariant, we can restrict F to $\mathcal{C}$ to obtain a map $F|\mathcal{C}\colon \mathcal{C} \to \mathcal{C}$. The following lemma shows that the topological entropy of this restriction is strictly smaller than $\log(\deg(F))$.

LEMMA 17.9. $h_{top}(F|\mathcal{C}) < \log(\deg(F))$.

PROOF. The proof is very similar to the first part of the proof of Lemma 17.8. Again let d be a base metric on S^2.

Since $\mathcal{C}$ consists of $\# \text{post}(f)$ 0-edges, by Lemma 17.7 we can cover $\mathcal{C}$ by a collection M_k of k-tiles, where $\# M_k \leq CL^k$. Here $1 \leq L < \deg(F)$ and C is independent of k. The k-edges in the boundaries of the k-tiles in M_k then form a cover of $\mathcal{C}$. It is clear that each k-edge contained in $\mathcal{C}$ must belong to this collection. Hence if E_k is the set of all k-edges contained in $\mathcal{C}$, we have $\# E_k \leq C'L^k$ with a constant C' independent of k.

Now let $\epsilon > 0$ be arbitrary. Since F is expanding, we can find $k_0 \in \mathbb{N}_0$ such that $\text{diam}(X) \leq \epsilon$ whenever $X \in \mathbf{X}^k$ for $k > k_0$. Since every k-edge is contained in a k-tile, we also have $\text{diam}(e) \leq \epsilon$ whenever $e \in \mathbf{E}^k$ for $k > k_0$.

If $k \in \mathbb{N}$ and $e \in E_{k+k_0}$ is arbitrary, then $F^i(e)$ is an edge of level $k - i + k_0 > k_0$ for $i = 0, 1 \ldots, k - 1$, and so $\text{diam}(F^i(e)) \leq \epsilon$. This implies that the diameter of e with respect to the metric d_F^k is $\leq \epsilon$.

It follows that $D(F|\mathcal{C}, \epsilon, k) \leq \# E_{k+k_0} \leq C'L^{k_0+k}$, and so $h(\epsilon, F) \leq \log(L)$. Letting $\epsilon \to 0$ we conclude $h_{top}(F|\mathcal{C}) \leq \log(L) < \log(\deg(F))$ as desired. $\qquad\square$

REMARK 17.10. The topological entropy of $F|\mathcal{C}$ can in fact be computed explicitly. Since we do not need this in the following, we will only give an outline of the procedure without proof.

As described in Section 5.4, we label the 0-edges $e_1, \ldots, e_m$ so that they are in cyclic order on the boundary of the white 0-tile X_{w}^0. Here $m = \# \text{post}(F)$. If e^n is an n-edge, then F^n maps it homeomorphically to a 0-edge. We say that e^n is of *type i* if $F^n(e^n) = e_i$.

Each 0-edge is subdivided into 1-edges. Let a_{ij} be the number of 1-edges of type j into which e_i is subdivided, and $A = (a_{ij})_{i,j=1,\ldots,m}$ be the $(m \times m)$-matrix with entries a_{ij}. It is easy to see that for $n \in \mathbb{N}$

the entry a_{ij}^n of the n-th power $A^n = (a_{ij}^n)$ of A gives the number of n-edges of type j into which the 0-edge e_i is subdivided. Based on this one can show that for the topological entropy of $F|\mathcal{C}$ we have $h_{top}(F|\mathcal{C}) = \log(\rho(A))$, where $\rho(A)$ is the spectral radius of A.

The spectral radius $\rho(A)$ in turn can easily be computed from any matrix norm. If for an $(m \times m)$-matrix $B = (b_{ij})$ we set

$$\|B\| := \sum_{i,j=1}^{m} |b_{ij}|$$

for example, then $h_{top}(F|\mathcal{C}) = \rho(A) = \lim_{n\to\infty} \|A^n\|^{1/n}$. Note that for this matrix norm, $\|A^n\|$ is equal to the number of n-edges contained in $\mathcal{C}$.

Recall from Chapter 12 that the F-invariant Jordan curve $\mathcal{C}$ gives a two-tile subdivision rule that is realized by F (see Proposition 12.2). The quantity $h_{top}(F|\mathcal{C})$ can be viewed as a measure of the complexity of this two-tile subdivision rule (see also [**MT88**]). This is similar to the "core-entropy" of the Hubbard tree of a postcritically-finite polynomial (see [**Ti16**]).

After these preparations we are now ready to construct a measure ν_F that will turn out to be the unique measure of maximal entropy for F and f.

PROPOSITION 17.11. *There exists a unique probability measure ν_F on S^2 such that for each $X \in \mathbf{X}^k$, $k \in \mathbb{N}_0$, we have*

(17.14) $$\nu_F(X) = \begin{cases} w \deg(F)^{-k} & \text{if } X \text{ is white,} \\ b \deg(F)^{-k} & X \text{ is black.} \end{cases}$$

Then $\nu_F(E^\infty) = 0$. Moreover, the measure ν_F is F-invariant, F is mixing for ν_F, and $h_{\nu_F}(F) = \log(\deg(F))$.

Here w and b are as in (17.9). The proposition implies in particular that edges and vertices are sets of ν_F-measure zero, and that F is ergodic for ν_F.

PROOF. The proof proceeds in several steps.

Construction of ν_F and $\nu_F(E^\infty) = 0$: For each k-tile X, $k \in \mathbb{N}_0$, set $w(X) = w(\deg(F))^{-k}$ if X is white and $w(X) = b(\deg(F))^{-k}$ if X is black.

If $X \in \mathbf{X}^k$ is white, then w_1 is the number of white $(k+1)$-tiles contained in X, and b_1 the number of black $(k+1)$-tiles contained in

X. Since $w_1 + w_1' = \deg(F)$, we have

$$\sum_{Y \in \mathbf{X}^{k+1}, Y \subset X} w(Y) = \frac{w_1 w + b_1 b}{\deg(F)^{k+1}} = \frac{w_1 b_1 + b_1 w_1'}{(b_1 + w_1') \deg(F)^{k+1}}$$

$$= \frac{b_1}{(b_1 + w_1') \deg(F)^k} = w(X).$$

A similar equation is also true for black k-tiles. If we iterate these identities, we get

$$(17.15) \qquad \sum_{Y \in \mathbf{X}^{k+m}, Y \subset X} w(Y) = w(X)$$

for all $k, m \in \mathbb{N}_0$ and all $X \in \mathbf{X}^k$.

For $A \subset S^2$ we now define

$$(17.16) \qquad \nu^*(A) = \inf_{\mathcal{U}} \sum_{X \in \mathcal{U}} w(X),$$

where the infimum is taken over all covers $\mathcal{U}$ of A by tiles (not necessarily of the same level). By subdividing tiles into tiles of high level and using (17.15), one sees that in the infimum in the definition of $\nu^*(A)$ it is enough to only consider covers by tiles whose levels exceed a given number k. Based on this and the fact that $\max_{x \in \mathbf{X}^k} \operatorname{diam}(X) \to 0$ as $k \to \infty$, it is clear that ν^* is a *metric outer measure*, i.e., ν^* is an outer measure and if $A, B \subset S^2$ are sets with $\operatorname{dist}(A, B) > 0$ (for some base metric on S^2), then

$$\nu^*(A \cup B) = \nu^*(A) + \nu^*(B).$$

It is a known fact that the restriction of a metric outer measure to the σ-algebra of Borel sets is a measure (see [**StS05**, Theorem 1.2, p. 267]). We denote this restriction of ν^* by ν_F.

If $A \subset S^2$ is compact, then it is enough to consider only finite covers by tiles in (17.16). Indeed, suppose $\mathcal{U} = \{X_i : i \in \mathbb{N}\}$ is an infinite cover of the compact set $A \subset S^2$ by tiles, and let $\epsilon > 0$ be arbitrary. It follows from Lemma 17.7 that for each $i \in \mathbb{N}$ we can find a finite collection $\mathcal{U}_i$ of tiles such that $X_i \subset \operatorname{int}(X_i')$, where

$$X_i' = \bigcup_{X \in \mathcal{U}_i} X$$

and

$$\sum_{X \in \mathcal{U}_i} w(X) \leq w(X_i) + \epsilon/2^i.$$

Finitely many sets $\operatorname{int}(X_{i_1}'), \ldots, \operatorname{int}(X_{i_m}')$ will cover A. Then

$$\mathcal{U}' = \mathcal{U}_{i_1} \cup \cdots \cup \mathcal{U}_{i_m}$$

is a finite collection of tiles that covers A, and we have

$$\sum_{X \in \mathcal{U}'} w(X) \le \sum_{X \in \mathcal{U}} w(X) + \epsilon.$$

Since $\epsilon > 0$ was arbitrary, we conclude that for compact sets A we get the same infimum in (17.16) if we restrict ourselves to finite covers by tiles.

A consequence of this is that $\nu_F(X) = w(X)$ for each $X \in \mathbf{X}^k$, $k \in \mathbb{N}_0$. Indeed, by definition of ν_F we obviously have $\nu_F(X) \le w(X)$.

For an inequality in the opposite direction, it is enough to consider an arbitrary finite cover $\mathcal{U}$ of X by tiles. By subdividing the tiles in $\mathcal{U}$ if necessary, we may also assume that they all have the same level l and that $l \ge k$. Since $\mathcal{U}$ is a cover of X and l-tiles have pairwise disjoint interiors, this implies that $Y \in \mathcal{U}$ whenever $Y \in \mathbf{X}^l$ and $Y \subset X$. Hence

$$w(X) = \sum_{Y \in \mathbf{X}^l, Y \subset X} w(Y) \le \sum_{Y \in \mathcal{U}} w(Y).$$

Taking the infimum over all $\mathcal{U}$ we get $w(X) \le \nu_F(X)$ as desired.

Since $\nu_F(X) = w(X)$ for all tiles X, we have (17.14).

It follows from Lemma 17.7 and the definition of ν_F, that if e is an edge, then $\nu_F(e) = 0$. Since E^∞ is the (countable) union of all edges, we have $\nu_F(E^\infty) = 0$. This also shows that

$$\nu_F(S^2) = \sum_{X \in \mathbf{X}^0} \nu_F(X) = \sum_{X \in \mathbf{X}^0} w(X) = w + b = 1,$$

and so ν_F is a probability measure.

Uniqueness of ν_F: Suppose that ν is another probability measure on S^2 satisfying the analog of (17.14). Then from Lemma 17.7 it follows that each edge is a set of ν-measure zero. Hence $\nu(\text{int}(c)) = \nu_F(\text{int}(c))$ whenever c is a cell for $(F, \mathcal{C})$. Since the empty set together with the interiors of all cells for $(F, \mathcal{C})$ form a semi-algebra $\mathcal{S}$ generating the Borel σ-algebra on S^2 (see Lemma 17.5), we conclude that $\nu = \nu_F$.

ν_F *is F-invariant*: To show that ν_F is F-invariant, it is enough to verify that

(17.17) $$\nu_F(F^{-1}(A)) = \nu_F(A)$$

for all sets A in the semi-algebra $\mathcal{S}$. This is true if $A = \emptyset$.

Edges are sets of measure zero, and the preimage of an edge is a finite union of edges (see Proposition 5.17 (ii)). This implies that (17.17) holds if $A = \text{int}(e)$ for an edge e, or if $A = \{v\}$ for a vertex v (since every vertex is contained in an edge).

Moreover, if X is a tile, then $X \setminus \operatorname{int}(X)$ is a union of edges, and so we have $\nu_F(F^{-1}(\operatorname{int}(X))) = \nu_F(F^{-1}(X))$ and $\nu_F(\operatorname{int}(X)) = \nu_F(X)$. So in order to establish (17.17), it remains to show that

$$\nu_F(F^{-1}(X)) = \nu_F(X)$$

for all tiles X. To see this, note that if X is a k-tile, then $F^{-1}(X)$ is a union of $\deg(F)$ $(k+1)$-tiles that have the same color as X. Since the intersection of any two distinct $(k+1)$-tiles is contained in a union of edges, and hence a set of measure zero, it follows from (17.14) that

$$\nu_F(F^{-1}(X)) = \deg(F)\frac{\nu_F(X)}{\deg(F)} = \nu_F(X).$$

F *is mixing for* ν_F: It suffices to show that for all sets A and B in the semi-algebra $\mathcal{S}$ we have

$$\nu_F(F^{-m}(A) \cap B) \to \nu_F(A)\nu_F(B)$$

as $m \to \infty$. Based on the fact that edges are sets of measure zero and that the preimage of each edge under F is a finite union of edges, for this it suffices to show that for all tiles X and Y we have

$$\nu_F(F^{-m}(X) \cap Y) \to \nu_F(X)\nu_F(Y)$$

as $m \to \infty$.

So let $k, l, m \in \mathbb{N}_0$, $X = X^k \in \mathbf{X}^k$ and $Y = Y^l \in \mathbf{X}^l$ be arbitrary. We may assume that $m \geq l$. Then $F^{-m}(X^k)$ is a union of $(m+k)$-tiles that have the same color as X^k. Since edges are sets of measure zero and the $(m+k)$-tiles subdivide the l-tile Y^l, it follows that

$$(17.18) \qquad \nu_F(F^{-m}(X^k) \cap Y^l) = \frac{\nu_F(X^k)}{\deg(F)^m} \cdot \#M,$$

where

$$(17.19) \qquad M := \{Z^{m+k} \in \mathbf{X}^{m+k} : Z^{m+k} \subset Y^l, \, F^m(Z^{m+k}) = X^k\}.$$

Let $X^0 \in \mathbf{X}^0$ be the unique 0-tile with $X^k \subset X^0$, and $Y^0 := F^l(Y^l) \in \mathbf{X}^0$. We assume that X^0 and Y^0 are both white; the other cases are similar. Then $Y = Y^l$ is also white.

Every $(m+k)$-tile Z^{m+k} lies in a unique "parent" m-tile Z^m. Since Y^l is an l-tile and $m \geq l$, we have $Z^{m+k} \subset Y^l$ if and only if $Z^m \subset Y^l$. If $F^m(Z^{m+k}) = X^k$, then $F^m(Z^m)$ is a 0-tile containing X^k, and so $F^m(Z^m) = X^0$. Conversely, if Z^m is an m-tile and $F^m(Z^m) = X^0$, then it follows from Lemma 5.18 (i) that $Z^{m+k} := (F^m|Z^m)^{-1}(X^k)$ is the unique $(m+k)$-tile with $Z^{m+k} \subset Z^m$ and $F^m(Z^{m+k}) = X^k$. The situation is illustrated in Figure 17.1. These statements imply that the map $Z^{m+k} \mapsto Z^m$ that assigns to each $(m+k)$-tile Z^{m+k} its unique
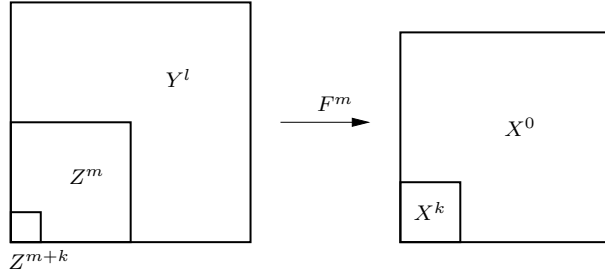
FIGURE 17.1. Bijection of tiles.

parent m-tile Z^m induces a bijection between the set M defined in (17.19) and

$$N := \{Z^m \in \mathbf{X}^m : Z^m \subset Y^l,\ F^m(Z^m) = X^0\}.$$

Hence $\#M = \#N$.

Since X^0 is white, $\#N$ is equal to the number of white m-tiles contained in Y^l. Applying the homeomorphism $F^l|Y^l$, we see that this number is equal to w_{m-l}, the number of white $(m-l)$-tiles contained in the white 0-tile $Y^0 = F^l(Y^l)$. It follows that $\#M = \#N = w_{m-l}$. So from (17.18) and Lemma 17.4 we conclude that

$$\nu_F(F^{-m}(X) \cap Y) = \frac{\nu_F(X)}{\deg(F)^m} \cdot w_{m-l} \to \frac{\nu_F(X)}{\deg(F)^l} \cdot w = \nu_F(X)\nu_F(Y)$$

as $m \to \infty$.

The identity $h_{\nu_F}(F) = \log(\deg(F))$: According to Lemma 17.6 the measurable partition $\xi = \mathbf{X}^1$ is a generator for (F, ν_F), and so $h_{\nu_F}(F) = h_{\nu_F}(F, \xi)$. Moreover, for each $k \in \mathbb{N}$, the measurable partition ξ_F^k is equivalent to the measurable partition $\mathbf{X}^k$ given by the k-tiles. Since the number of black and the number of white k-tiles are both equal to $\deg(F)^k$, it follows that

$$
\begin{aligned}
H_{\nu_F}(\xi_F^k) &= H_{\nu_F}(\mathbf{X}^k) = \sum_{X \in \mathbf{X}^k} \nu_F(X) \log(1/\nu_F(X)) \\
&= w \log(\deg(F)^k/w) + b \log(\deg(F)^k/b) \\
&= k \log(\deg(F)) + w \log(1/w) + b \log(1/b).
\end{aligned}
$$

This implies

$$h_{\nu_F}(F) = h_{\nu_F}(F, \xi) = \lim_{k \to \infty} \frac{1}{k} H_{\nu_F}(\xi_F^k) = \log(\deg(F)).$$

The proof is complete. $\square$

We can now identify the topological entropy of f.

PROOF OF COROLLARY 17.2. Lemma 17.8 (applied to F) implies that $h_{top}(F) \le \log(\deg(F))$. We also have $h_{\nu_F}(F) = \log(\deg(F))$ by Proposition 17.11, and so $h_{top}(F) \ge \log(\deg(F))$ by the variational principle (17.7). It follows that $h_{top}(F) = \log(\deg(F))$. Since $F = f^n$ and so $\deg(F) = \deg(f)^n$ and $h_{top}(F) = n h_{top}(f)$, the claim follows. $\square$

We know that $h_{\nu_F}(F) = \log(\deg(F)) = h_{top}(F)$. So ν_F is a measure of maximal entropy for F. As we will see momentarily, it is the unique measure of maximal entropy for F and also for f.

THEOREM 17.12. *The measure ν_F is the unique measure of maximal entropy for f, i.e., the unique f-invariant probability measure ν_F with $h_{\nu_F}(f) = h_{top}(f)$. Moreover, f is mixing for ν_F.*

PROOF. We first show uniqueness. So let μ be a probability measure that is f-invariant and satisfies $h_\mu(f) = h_{top}(f)$. Then μ is F-invariant and satisfies

$$(17.20) \qquad h_\mu(F) = n h_\mu(f) = n h_{top}(f) = h_{top}(F) = \log(\deg(F)).$$

We will see that this implies $\mu = \nu_F$. In particular, this will show that ν_F is the unique measure of maximal entropy for F. The proof proceeds in several steps.

By using the Lebesgue decomposition of μ with respect to ν_F, we can represent μ as a convex combination $\mu = \beta \mu_a + (1 - \beta)\mu_s$, where $\beta \in [0, 1]$, μ_a is a probability measure that is absolutely continuous and μ_s is a probability measure that is singular with respect to ν_F (if $\beta = 0$ or $\beta = 1$, the decomposition is trivial and only one of the measures μ_a or μ_s exists). Since ν_F and μ are F-invariant, Lemma 17.3 implies that the measures μ_a and μ_s are also F-invariant. Since F is ergodic for ν_F, and μ_a is F-invariant and absolutely continuous with respect to μ_F, it follows that $\mu_a = \nu_F$.

If $\beta = 1$, then $\mu = \nu_F$ and we are done. If $\beta \in [0, 1)$, then we can use the equation

$$\begin{aligned} \log(\deg(F)) &= h_\mu(F) = \beta h_{\nu_F}(F) + (1 - \beta) h_{\mu_s}(F) \\ &= \beta \log(\deg(F)) + (1 - \beta) h_{\mu_s}(F) \end{aligned}$$

to conclude that

$$h_{\mu_s}(F) = \log(\deg(F)).$$

We will show that this is impossible by proving that for every F-invariant probability measure μ that is singular with respect to ν_F we must have

$$h_\mu(F) < \log(\deg(F)).$$

The uniqueness of ν_F will then follow.

So let μ be such a measure and consider the union E^∞ of all edges. Assume first that $\mu(E^\infty) > 0$. By (17.12) we can then write μ as a convex combination $\mu = \alpha\mu_1 + (1-\alpha)\mu_2$ of two F-invariant probability measures μ_1 and μ_2, where $\alpha = \mu(E^\infty)$, μ_1 is concentrated on E^∞, and μ_2 on $S^2 \setminus E^\infty$.

Since μ_1 is F-invariant, we have $\mu_1(F^{-k}(\mathcal{C})) = \mu_1(\mathcal{C})$ for all $k \in \mathbb{N}_0$. On the other hand, $\mathcal{C} \subset F^{-k}(\mathcal{C})$, and so $\mu_1(F^{-k}(\mathcal{C}) \setminus \mathcal{C}) = 0$. This implies that $\mu_1(E^\infty \setminus \mathcal{C}) = 0$. So μ_1 is actually concentrated on $\mathcal{C}$. Therefore, by the variational principle (17.7) and by Lemma 17.9 we have

$$h_{\mu_1}(F) = h_{\mu_1}(F|\mathcal{C}) \le h_{top}(F|\mathcal{C}) < \log(\deg(F)).$$

We also have $h_{\mu_2}(F) \le h_{top}(F) = \log(\deg(F))$, and so

$$h_\mu(F) = \alpha h_{\mu_1}(F) + (1-\alpha)h_{\mu_2}(F) < \log(\deg(F)).$$

In this case we are done.

In the remaining case we have $\mu(E^\infty) = 0$. Then by Lemma 17.6 $\xi = \mathbf{X}^1$ is a generator for (F, μ), and so $h_\mu(F) = h_\mu(F, \mu)$. In particular,

$$h_\mu(F) = \lim_{k \to \infty} \frac{1}{k} \sum_{X \in \mathbf{X}^k} \mu(X) \log(1/\mu(X)),$$

and the limit is given by the infimum of the sequence elements.

Since μ and ν_F are mutually singular, we can find a Borel set $A \subset S^2$ with $\mu(A) = 1$ and $\nu_F(A) = 0$. Using inner regularity of μ and outer regularity of ν_F, for each $\epsilon > 0$ we can find a compact set $K \subset S^2$ and an open set $U \subset S^2$ with $K \subset A \subset U$, $\mu(K) > 1 - \epsilon$, and $\nu_F(U) < \epsilon$. If k is sufficiently large, then we can cover the set K by k-tiles contained in U.

Using this for smaller and smaller $\epsilon > 0$, we conclude that for each $k \in \mathbb{N}$ we can find a set $M_k \subset \mathbf{X}^k$ such that for $A_k := \bigcup\{X \in M_k\}$ we have $\mu(A_k) \to 1$ and $\nu_F(A_k) \to 0$ as $k \to \infty$. Note that $\nu_F(X) \ge c\deg(F)^{-k}$ for each $X \in \mathbf{X}^k$, where $c > 0$ is independent of k and X. Hence

$$\#M_k \le \nu_F(A_k)\deg(F)^k/c.$$

We also have $\#\mathbf{X}^k = 2\deg(F)^k$.

The function $x \mapsto \phi(x) = x\log(1/x)$ is concave and has a maximum equal to $1/e$ on $[0, 1]$. This implies that if $M \subset \mathbf{X}^k$ is arbitrary and $A = \bigcup_{X \in M} X$, then

$$\sum_{X \in M} \mu(X) \log(1/\mu(X)) \le \#M \cdot \phi\left(\frac{1}{\#M} \sum_{X \in M} \mu(X)\right)$$

$$= \mu(A) \log(\#M/\mu(A)) \le \mu(A) \log(\#M) + 1/e.$$

To reach a contradiction, let us assume that $h_\mu(F) = \log(\deg(F))$. Then for each $k \in \mathbb{N}$ we have

$$k \log(\deg(F)) = k h_\mu(F)$$

$$\leq \sum_{X \in \mathbf{X}^k} \mu(X) \log(1/\mu(X))$$

$$= \sum_{X \in M_k} \mu(X) \log(1/\mu(X)) + \sum_{X \in \mathbf{X}^k \setminus M_k} \mu(X) \log(1/\mu(X))$$

$$\leq \mu(A_k) \log(\#M_k) + \mu(S^2 \setminus A_k) \log(\#\mathbf{X}^k) + C_1$$

$$\leq \mu(A_k) \log(\nu_F(A_k)) + (\mu(A_k) + \mu(S^2 \setminus A_k)) \log(\deg(F)^k) + C_2$$

$$= \mu(A_k) \log(\nu_F(A_k)) + k \log(\deg(F)) + C_2.$$

Here the constants C_1 and C_2 do not depend on k. An inequality of this type is impossible as

$$\mu(A_k) \log(\nu_F(A_k)) \to -\infty$$

for $k \to \infty$.

This shows that if there is a measure of maximal entropy for f, then it has to agree with ν_F. We have also proved that ν_F is the unique measure of maximal entropy for F.

We now show that ν_F is f-invariant and a measure of maximal entropy for f. Indeed, the measure $f_*\nu_F$ is F-invariant and the triple $(S^2, F, f_*\nu_F)$ is a factor of (S^2, F, ν_F) by the map f. It follows that $h_{f_*\nu_F}(F) \leq h_{\nu_F}(F)$. Iterating this and noting that $f_*^n \nu_F = F_*\nu_F = \nu_F$ by F-invariance of ν_F, we obtain

$$h_{\nu_F}(F) = h_{f_*^n \nu_F}(F) \leq h_{f_*^{n-1} \nu_F}(F) \leq \cdots \leq h_{f_*\nu_F}(F) \leq h_{\nu_F}(F).$$

Hence $h_{f_*\nu_F}(F) = h_{\nu_F}(F)$, and so $f_*\nu_F$ is a measure of maximal entropy for F. By uniqueness of ν_F we have $f_*\nu_F = \nu_F$ showing that ν_F is f-invariant. Moreover,

$$h_{\nu_F}(f) = h_{\nu_F}(F)/n = \log(\deg(F))/n = \log(\deg(f)) = h_{top}(f),$$

and so ν_F is a measure of maximal entropy for f. By the first part of the proof we know that it is the unique such measure.

It remains to show that f is mixing for ν_F. Indeed, since F is mixing for ν_F (Proposition 17.11) and ν_F is f-invariant, we have that for all $m \in \{0, \ldots, n-1\}$, and all Borel sets $A, B \subset S^2$,

$$\nu_F(f^{-(nl+m)}(A) \cap B) = \nu_F(F^{-l}(f^{-m}(A)) \cap B) \to$$

$$\nu_F(f^{-m}(A))\mu_F(B) = \nu_F(A)\nu_F(B)$$

as $l \to \infty$. This implies the desired relation

$$\nu_F(f^{-k}(A) \cap B) \to \nu_F(A)\nu_F(B)$$

as $k \to \infty$. The proof is complete.   □

PROOF OF THEOREM 17.1. Obviously, the statement follows from Theorem 17.12.   □

CHAPTER 18

The geometry of the visual sphere

When $f\colon S^2 \to S^2$ is an expanding Thurston map and ϱ a visual metric for f, we call the metric space (S^2, ϱ) the *visual sphere* for f. Of course, (S^2, ϱ) depends on the choice of the visual metric ϱ, but by Proposition 8.3 (iv) any two such choices for a given Thurston map f produce snowflake equivalent metric spaces. Accordingly, we think of the visual sphere of a Thurston map as uniquely determined up to snowflake equivalence. In this chapter we investigate geometric features of the visual sphere that are invariant under such equivalences and relate them to the dynamics of f.

The following statement is one of the main results here.

THEOREM 18.1 (Properties of f and its associated visual sphere). *Let $f\colon S^2 \to S^2$ be an expanding Thurston map and ϱ be a visual metric for f. Then the following statements are true:*

(i) *(S^2, ϱ) is doubling if and only if f has no periodic critical points.*

(ii) *(S^2, ϱ) is quasisymmetrically equivalent to $\widehat{\mathbb{C}}$ if and only if f is topologically conjugate to a rational map.*

(iii) *(S^2, ϱ) is snowflake equivalent to $\widehat{\mathbb{C}}$ if and only if f is topologically conjugate to a Lattès map.*

Here it is understood that $\widehat{\mathbb{C}}$ is equipped with the chordal metric. Statement (ii) characterizes the visual spheres that are quasispheres. As we have already discussed, it provides an interesting analog of Cannon's conjecture (see Section 4.3).

We know that two expanding Thurston maps are Thurston equivalent if and only if they are topologically conjugate (Theorem 11.1). Thus (ii) and (iii) can be reformulated as follows:

(ii') *(S^2, ϱ) is quasisymmetrically equivalent to $\widehat{\mathbb{C}}$ if and only if f is Thurston equivalent to a rational Thurston map with no periodic critical points.*

(iii') *(S^2, ϱ) is snowflake equivalent to $\widehat{\mathbb{C}}$ if and only if f is Thurston equivalent to a Lattès map.*

For the "if" implication in (ii') one has to assume that the rational map has no periodic critical points (or impose an equivalent condition); see Example 18.11.

Much more can be said in case (i) of the previous theorem.

PROPOSITION 18.2. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map without periodic critical points, ϱ be a visual metric for f with expansion factor $\Lambda > 1$, and ν_f be the measure of maximal entropy of f. Then the metric measure space (S^2, ϱ, ν_f) is Ahlfors Q-regular with*

$$Q := \frac{\log(\deg(f))}{\log(\Lambda)}.$$

In particular, (S^2, ϱ) has Hausdorff dimension Q and

$$0 < \mathcal{H}^Q_\varrho(S^2) < \infty.$$

Here $\mathcal{H}^Q_\varrho$ is Q-dimensional Hausdorff measure on (S^2, ϱ). For the definition of the measure of maximal entropy ν_f see Chapter 17. The statement implies that under the given assumptions we have $\mathcal{H}^Q_\varrho(M) \asymp \nu_f(M)$ for each Borel set $M \subset S^2$ with $C(\asymp)$ independent of M.

Since the Hausdorff dimension of (S^2, ϱ) must be ≥ 2, it also follows that $\Lambda \leq \deg(f)^{1/2}$. Combining this with Theorem 16.3 (ii) we obtain the upper bound $\Lambda_0(f) \leq \deg(f)^{1/2}$ for the combinatorial expansion factor of f. We will later see that this is true for every expanding Thurston map without the additional assumption that f has no periodic critical points (see Proposition 20.1).

Recall from Proposition 8.3 (v) that a metric ϱ on S^2 is a visual metric for an expanding Thurston map $f\colon S^2 \to S^2$ if and only if it is a visual metric for an iterate $F = f^n$ (which is also an expanding Thurston map by Lemma 6.4). Hence Theorem 18.1 immediately gives the following corollary.

COROLLARY 18.3. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map, and $F = f^n$ with $n \in \mathbb{N}$ be an iterate of f. Then the following statements are true.*

  (i) *The map f is topologically conjugate to a rational map if and only if F is topologically conjugate to a rational map.*

  (ii) *The map f is topologically conjugate to a Lattès map if and only if F is topologically conjugate to a Lattès map.*

We now record some of the consequences of our results for rational Thurston maps explicitly.

THEOREM 18.4. *Let $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a rational Thurston map with no periodic critical points. Then the following statements are true:*

(i) *For each sufficiently large $n \in \mathbb{N}$ there exists a quasicircle $\mathcal{C} \subset \widehat{\mathbb{C}}$ with $\mathrm{post}(f) \subset \mathcal{C}$ that is f^n-invariant (i.e., $f^n(\mathcal{C}) \subset \mathcal{C}$).*

(ii) *Each f^n-invariant Jordan curve $\mathcal{C} \subset \widehat{\mathbb{C}}$ with $\mathrm{post}(f) \subset \mathcal{C}$ is a quasicircle.*

(iii) *Let $\mathcal{C}$ be an f-invariant Jordan curve $\mathcal{C} \subset \widehat{\mathbb{C}}$ with $\mathrm{post}(f) \subset \mathcal{C}$, $\mathbf{E}^n$ be the set of n-edges, and $\mathbf{X}^n$ be the set of n-tiles for $(f, \mathcal{C})$. Then the family $\{e : n \in \mathbb{N}_0, e \in \mathbf{E}^n\}$ consists of uniform quasiarcs, and the family $\{X : n \in \mathbb{N}_0, X \in \mathbf{X}^n\}$ of uniform quasidisks.*

Here the underlying metric is again the chordal metric on $\widehat{\mathbb{C}}$. For the concepts of uniformity used here see Section 15.3.

An immediate consequence of (iii) is that the family $\{\partial X : n \in \mathbb{N}_0, X \in \mathbf{X}^n\}$ consists of uniform quasicircles. Note that this and the statement about the arcs in (iii) do not *a priori* follow from Proposition 15.26, because we use different underlying metrics. We will prove though that for a rational Thurston map f without periodic critical points the chordal metric is quasisymmetrically equivalent to each visual metric for f (see Lemma 18.10). Once we know this, Theorem 18.4 (iii) can easily be deduced from Proposition 15.26.

A consequence of Theorem 18.4 is that each sufficiently high iterate of a rational Thurston map $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ that is expanding has a particularly nice Markov partition, where the tiles are quasidisks.

Another important property of visual spheres from the viewpoint of quasiconformal geometry is that they are *linearly locally connected*. This and related properties will be discussed in Section 18.1.

Theorem 18.1 (i) and Proposition 18.2 will be proved in Section 18.2. In Section 18.3 we will establish Theorem 18.1 (ii). We postpone the proof of Theorem 18.1 (iii) to the end of Section 19.4 (this part was essentially proved in [**Me09a**]).

18.1. Linear local connectedness

Recall (see Section 4.1) that a metric space (X, d) is said to be *linearly locally connected* (often abbreviated as LLC) if there exists a constant $C \geq 1$ such that the following two conditions are satisfied:

(LLC1) If $p \in X$, $r > 0$, and $x, y \in B_d(p, r)$, then there exists a continuum $E \subset X$ with $x, y \in E$ and $E \subset B_d(p, Cr)$.

(LLC2) If $p \in X$, $r > 0$, and $x, y \in X \setminus B_d(p, r)$, then there exists a continuum $E \subset X$ with $x, y \in E$ and $E \subset X \setminus B_d(p, r/C)$.

It is easy to see that LLC1 is satisfied if and only if X is of bounded turning (as defined in (4.3)).

The space (X, d) is called *annularly linearly locally connected* (abbreviated as ALLC) if there exists a constant $C \geq 1$ with the following property: If $p \in X$, $r > 0$, and $x, y \in \overline{B}_d(p, 2r) \setminus B_d(p, r)$, then there exists a path γ in X joining x and y with

$$\gamma \subset \overline{B}_d(p, Cr) \setminus B_d(p, r/C).$$

The following proposition shows that the visual sphere of an expanding Thurston map is linearly locally connected and annularly linearly locally connected.

PROPOSITION 18.5. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map, and suppose that S^2 is equipped with a visual metric ϱ for f. Then the following statements are true:*

(i) *(S^2, ϱ) is of bounded turning.*

(ii) *(S^2, ϱ) is annularly linearly locally connected.*

(iii) *(S^2, ϱ) is linearly locally connected.*

In this statement and in its proof below all metric concepts refer to the visual metric ϱ.

The statements (i)–(iii) are not logically independent, but one can show the implications (ii) $\Rightarrow$ (iii) $\Rightarrow$ (i) for quite general spaces. The ensuing proof will not rely on this directly.

PROOF. Let $\Lambda > 1$ be the expansion factor of ϱ. Then for some Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}$ we have $\varrho(u, v) \asymp \Lambda^{-m(u,v)}$ for $u, v \in S^2$, where $m(u, v) = m_{f,\mathcal{C}}(u, v)$ (see Definitions 8.1 and 8.2). In the following, all cells will be for $(f, \mathcal{C})$.

(i) Let $x, y \in S^2$ be arbitrary. If $x = y$ there nothing to prove. So we may assume that $x \neq y$. Let $n = m(x, y) \in \mathbb{N}_0$. Then there exist n-tiles X and Y with $x \in X$, $y \in Y$, and $X \cap Y \neq \emptyset$. Since X and Y are Jordan regions, we can find a path α in $X \cup Y$ that joins x and y. Then by Proposition 8.4 (ii) we have

$$\mathrm{diam}(\alpha) \leq \mathrm{diam}(X) + \mathrm{diam}(Y) \lesssim \Lambda^{-n} \asymp \varrho(x, y).$$

In particular,

(18.1) $\mathrm{diam}(\alpha) \leq K\varrho(x, y)$

with a constant $K \geq 1$ independent of x and y. Statement (i) follows.

(ii) Let $p \in S^2$, $r > 0$, and $x, y \in \overline{B}(p, 2r) \setminus B(p, r)$. In the following, all implicit multiplicative constants will be independent of these initial
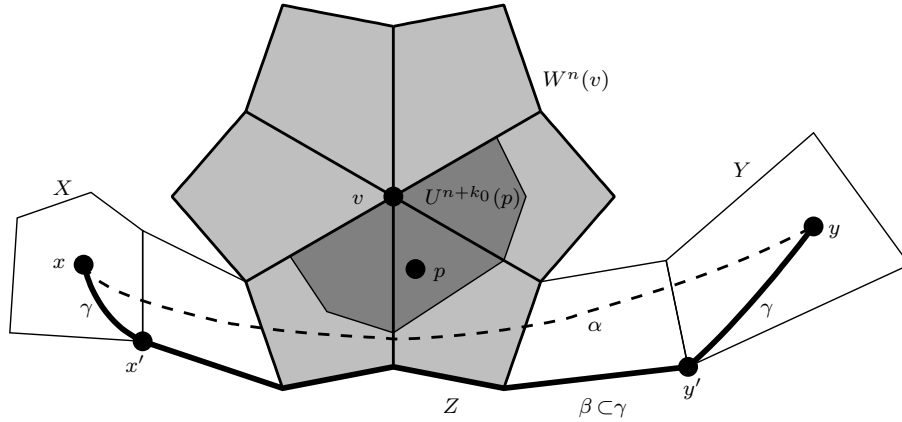
FIGURE 18.1. Construction of γ.

choices of p, r, x, and y. We have to find a path γ joining x and y such that

$$(18.2) \qquad \gamma \subset \overline{B}(p, Cr) \setminus B(p, r/C),$$

where $C \geq 1$ is a suitable constant. The construction of γ is illustrated in Figure 18.1.

Define

$$n := \max\{m(p, x), m(p, y)\} + 1.$$

Then

$$\Lambda^{-n} \asymp \min\{\varrho(p, x), \varrho(p, y)\} \asymp r.$$

Let X, Y, Z be n-tiles with $x \in X$, $y \in Y$, and $p \in Z$. Then by definition of n we have $X \cap Z = \emptyset$ and $Y \cap Z = \emptyset$.

Since f is expanding, we can choose $k_0 \in \mathbb{N}_0$ as in (8.3). In particular, every connected set of k_0-tiles joining opposite sides of $\mathcal{C}$ must contain at least ten k_0-tiles.

Consider the set $U^{n+k_0}(p)$ as defined in (8.8). This is the set of all $(n + k_0)$-tiles that intersect an $(n + k_0)$-tile containing p. Then $f^n(U^{n+k_0}(p))$ is connected, and consists of k_0-tiles. This set cannot join opposite sides of $\mathcal{C}$; for otherwise, we could find a connected set of six k_0-tiles with this property (see the proof Lemma 8.11 for a similar reasoning). This is impossible by definition of k_0. Hence $f^n(U^{n+k_0}(p))$ is contained in a 0-flower (Lemma 5.33) which implies that $U^{n+k_0}(p)$ is contained in an n-flower (Lemma 5.29 (iii)). So there exists an n-vertex v with $p \in U^{n+k_0}(p) \subset W^n(v)$. Since Z contains p, this tile must be one of the n-tiles forming the cycle of v. So $v \in Z$, and $v \notin X, Y$. This in turn implies that X and Y do not meet $W^n(v)$ (see Lemma 5.28 (iii)).

Pick a path α in S^2 that joins x and y and satisfies (18.1). By Lemma 11.16 we can find a set M of n-tiles that forms an e-chain

joining X and Y (see Definition 5.20) so that each tile in M has non-empty intersection with α. Pick n-vertices $x' \in \partial X$, $y' \in \partial Y$. Since X and Y do not contain v, we have $x', y' \neq v$. Consider the graph $G_M = \{\partial U : U \in M\}$. It consists of n-edges, is connected, has no cut points (Lemma 11.14), and contains x' and y' as vertices. Hence there exists an edge path in G_M joining x' and y' whose underlying set β does not contain v. Then this edge path does not contain any edge in the cycle of v and so $\beta \cap W^n(v) = \emptyset$. Let γ be the path in S^2 that is obtained by running from x to x' along some path in X, then from x' to y' along β, and then from y' to y along some path in Y. Then γ joins x and y.

Since the sets X, Y, β have empty intersection with $W^n(v)$ and hence with $U^{n+k_0}(p)$, it follows that $\gamma \cap U^{n+k_0}(p) = \emptyset$. Thus we have by Lemma 8.10 (i)

$$\operatorname{dist}(p, \gamma) \gtrsim \Lambda^{-(n+k_0)} \asymp \Lambda^{-n} \asymp r.$$

Hence there exists a constant $C_1 \geq 1$ independent of the initial choices such that

$$\gamma \cap B(p, r/C_1) = \emptyset.$$

The set γ can be covered by n-tiles that meet α. Since

$$\operatorname{diam}(\alpha) \leq K\varrho(x, y) \leq 4Kr \lesssim r,$$

and

$$\max\{\operatorname{diam}(U) : U \text{ is an } n\text{-tile}\} \lesssim \Lambda^{-n} \asymp r,$$

we conclude that

$$\operatorname{diam}(\gamma) \leq \operatorname{diam}(\alpha) + 2\max\{\operatorname{diam}(U) : U \text{ is an } n\text{-tile}\} \lesssim r.$$

Since the initial point x of γ has distance $\leq 2r$ from p, it follows that there exists a constant $C_2 \geq 1$ independent of the initial choices such that $\gamma \subset B(p, C_2 r)$. If we set $C = \max\{C_1, C_2\}$, then (18.2) follows.

(iii) To show that (S^2, ϱ) is linearly locally connected, we verify the two relevant conditions LLC1 and LLC2; here we can use possibly different constants C in each of the conditions.

Let $p \in S^2$, $r > 0$, and $x, y \in B(p, r)$ be arbitrary. We choose a path α joining x and y that satisfies (18.1). Define $E := \alpha$ and $C = 2K + 1$. Then $x, y \in E$, and, since $\operatorname{diam}(\alpha) \leq K\varrho(x, y) \leq 2Kr$, we have

$$E \subset B(p, r + \operatorname{diam}(\alpha)) \subset B(p, Cr).$$

This shows that (S^2, ϱ) satisfies LLC1.

In order to prove LLC2, let $p \in S^2$, $r > 0$, and $x, y \in S^2 \setminus B(p, r)$ be arbitrary. Let α be a path in S^2 joining x and y. If $\alpha \cap B(p, r) = \emptyset$,

define $E := \alpha$. Then E is a continuum with $x, y \in E$ and $E \subset S^2 \setminus B(p, r)$.

If α meets $B(p, r)$, then, as we travel from x to y along α, there exists a first point with $x' \in \overline{B}(p, 2r)$. Note that if $x \in \overline{B}(p, 2r)$, then $x' = x$, and $d(p, x') = 2r$ otherwise. In any case, $x' \in \overline{B}(p, 2r) \setminus B(p, r)$. Let α_x be the subpath of α obtained by traveling along α starting from x until we reach x'. Then $\alpha_x \subset S^2 \setminus B(p, r)$.

Traveling along α in the opposite direction starting from y, we similarly define a point $y' \in \overline{B}(p, 2r) \setminus B(p, r)$ and a subpath $\alpha_y \subset S^2 \setminus B(p, r)$ of α joining y and y'. Then $x', y' \in \overline{B}(p, 2r) \setminus B(p, r)$. Hence by (ii) there exists a path γ in S^2 that joins x' and y' and satisfies $\gamma \subset S^2 \setminus B(p, r/C)$. Here $C \geq 1$ is a constant independent of the initial choices. Now define $E = \alpha_x \cup \gamma \cup \alpha_y$. Then E is a continuum with $x, y \in E$ and $E \subset S^2 \setminus B(p, r/C)$. It follows that LLC2 is satisfied as well. $\square$

18.2. Doubling and Ahlfors regularity

Here part (i) of Theorem 18.1 and Proposition 18.2 are proved. We first need some preparation.

Let $f \colon S^2 \to S^2$ be a branched covering map on a 2-sphere S^2. Recall that a point $p \in S^2$ is called periodic if there exists $n \in \mathbb{N}$ such that $f^n(p) = p$ and that the smallest n for which this is true is called the period of the periodic point.

The following lemma is essentially well known.

LEMMA 18.6. *Let $f \colon S^2 \to S^2$ be a branched covering map. Then f has no periodic critical points if and only if there exists $N \in \mathbb{N}$ such that*

$$\deg(f^n, p) \leq N$$

for all $p \in S^2$ and all $n \in \mathbb{N}$.

PROOF. Note that for $p \in S^2$ and $n \in \mathbb{N}$ we have

$$(18.3) \qquad \deg(f^n, p) = \prod_{k=0}^{n-1} \deg(f, f^k(p)).$$

So if p is a periodic critical point of period l, say, and $d = \deg(f, p) \geq 2$, then

$$\deg(f^n, p) \geq d^{\lceil n/l \rceil} \geq 2^{n/l} \to \infty$$

as $n \to \infty$. Hence $\deg(f^n, p)$ is not uniformly bounded.

If f has no periodic critical point, then the orbit $p, f(p), f^2(p), \ldots$ of a point $p \in S^2$ can contain each critical point at most once. Hence

by (18.3) we have

$$\deg(f^n, p) \leq N := \prod_{c \in \mathrm{crit}(f)} \deg(f, c).$$

Note that the last product is finite, because f has only finitely many critical points. $\square$

COROLLARY 18.7. *Let $f\colon S^2 \to S^2$ be a Thurston map with no periodic critical points. Then there is a constant $N \in \mathbb{N}$ with the following property: if $\mathcal{C} \subset S^2$ is a Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}$, then for each $n \in \mathbb{N}_0$ and each vertex v in $\mathcal{D}^n(f, \mathcal{C})$ the cycle of v has length at most N.*

In other words, the closure of each n-flower $W^n(v)$ contains at most N tiles of level n, where N only depends on f.

PROOF. By Lemma 18.6 there exists $N' \in \mathbb{N}$ such that the inequality $\deg(f^n, p) \leq N'$ is valid for all $p \in S^2$ and $n \in \mathbb{N}_0$. This implies that the cycle of each n-vertex (defined with respect to any Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}$) has length at most $N := 2N'$ (see Lemma 5.28 (i)). $\square$

We are now ready to prove the first part of Theorem 18.1.

PROOF OF THEOREM 18.1 (i). Assume first that f has no periodic critical points. By Theorem 15.1 there exists an iterate $F = f^n$ and an F-invariant Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) = \mathrm{post}(F) \subset \mathcal{C}$. Then F is also an expanding Thurston map (Lemma 6.4) and it has no periodic critical points as easily follows from (2.5). So by Corollary 18.7 there exists a number $N \in \mathbb{N}$ such that the cycle for each vertex in $\mathcal{D}^n(F, \mathcal{C})$, $n \in \mathbb{N}_0$, has length $\leq N$.

It suffices to show that S^2 equipped with a visual metric for F is doubling, since the class of visual metrics for f and F agree (Proposition 8.3 (v)).

Fix such a visual metric ϱ for F, and denote by $\Lambda > 1$ its expansion factor. In the following, all cells are for $(F, \mathcal{C})$. To establish that S^2 is doubling, we now proceed as in the proof of Theorem 15.3.

Let $x \in S^2$ and $0 < r \leq 2\,\mathrm{diam}(S^2)$ be arbitrary. We have to cover $B(x, r)$ by a controlled number of sets of diameter $< r/4$. Using Proposition 8.4, we can find $n \in \mathbb{N}_0$ depending on r, as well as constants $C(\asymp) > 0$ and $k_0 \in \mathbb{N}_0$ independent of x and r with the following properties:

  (i) $r \asymp \Lambda^{-n}$,
  (ii) $\mathrm{diam}(X) < r/4$ whenever X is an $(n + k_0)$-tile,

(iii) $\operatorname{dist}(X, Y) \geq r$ whenever $n - k_0 \geq 0$ and X and Y are disjoint $(n - k_0)$-tiles.

Let T be the set of all $(n + k_0)$-tiles that meet $B(x, r)$. Then the collection T forms a cover of $B(x, r)$ and consists of sets of diameter $< r/4$ by (ii). Hence it suffices to find a uniform upper bound on $\#T$, independent of x and r. If $n < k_0$, then $\#T \leq 2 \deg(F)^{2k_0}$ (see Proposition 5.17 (iv)) and we have such a bound.

Otherwise, $n - k_0 \geq 0$. Pick an $(n - k_0)$-tile X with $x \in X$. If Z is an arbitrary $(n + k_0)$-tile in T, then we can find a unique $(n - k_0)$-tile Y that contains Z (here we use that $\mathcal{C}$ is F-invariant and so each tile is subdivided by tiles of higher level).

There exists a point $y \in Z \cap B(x, r)$. Hence $\operatorname{dist}(X, Y) \leq \varrho(x, y) < r$. This implies $X \cap Y \neq \emptyset$ by (iii). So whatever $Z \in T$ is, the corresponding $(n - k_0)$-tile $Y \supset Z$ meets the fixed $(n - k_0)$-tile X. Hence Y must share an $(n - k_0)$-vertex v with X which implies $Y \subset \overline{W^{n-k_0}(v)}$. Since by choice of N the set $\overline{W^{n-k_0}(v)}$ contains at most N tiles of level $(n - k_0)$, and the number of $(n - k_0)$-vertices in X is equal to $\# \operatorname{post}(F)$, this leaves at most $N \# \operatorname{post}(f)$ possibilities for Y.

Since every $(n - k_0)$-tile contains at most $2 \deg(f)^{2k_0}$ tiles of level $(n + k_0)$, it follows that $\#T \leq 2N \# \operatorname{post}(f) \deg(f)^{2k_0}$. So we get a uniform bound as desired, which shows that (S^2, ϱ) is doubling.

To show the reverse implication, we use the following fact about doubling spaces, which is easy to show: in every ball there cannot be too many pairwise disjoint smaller balls that all have the same radius. More precisely, for every $\eta \in (0, 1)$ there is a number K such that every open ball of radius r contains at most K pairwise disjoint open balls of radius ηr.

Now suppose $f \colon S^2 \to S^2$ is an expanding Thurston map such that S^2 equipped with some visual metric is doubling. Pick a Jordan curve $\mathcal{C} \subset S^2$ with $\operatorname{post}(f) \subset \mathcal{C}$. In the following, cells will be for $(f, \mathcal{C})$. Let $p \in S^2$ and $n \in \mathbb{N}$. In order to show that f has no periodic critical points, it suffices to give a uniform bound on $d = \deg(f^n, p)$ (see Lemma 18.6). For this we may assume that $\deg(f^n, p) \geq 2$. Then p is an n-vertex and the closure of the n-flower $W^n(p)$ consists of precisely $2 \deg(f^n, p)$ tiles of level n. These n-tiles have pairwise disjoint interior and each interior contains a ball of radius $r \asymp \Lambda^{-n}$ (see Lemma 8.11). On the other hand, $\operatorname{diam}(\overline{W^n(p)}) \lesssim \Lambda^{-n}$. Since S^2 is doubling, it follows that the number of these tiles and hence $\deg(f^n, p)$ is uniformly bounded from above by a constant independent of n and p. Hence f has no periodic critical points. $\qquad \square$

We now prove the Ahlfors regularity of (S^2, ϱ) when f has no periodic critical points.

PROOF OF PROPOSITION 18.2. As in the statement, let $f\colon S^2 \to S^2$ be an expanding Thurston map without periodic critical points, ϱ be a visual metric, and $\nu = \nu_f$ be the measure of maximal entropy for f. Assume that $\Lambda > 1$ is the expansion factor of ϱ.

By Theorem 15.1 we can fix an iterate $F = f^n$ and an F-invariant Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}$. The map F is expanding by Lemma 6.4 and ϱ is a visual metric for F with expansion factor $\Lambda_F :=$ Λ^n by Proposition 8.3 (v). In the following, cells are defined for $(F, \mathcal{C})$ and metric notions refer to ϱ. The measure ν is also the measure of maximal entropy ν_F for F (see Proposition 17.11 and Theorem 17.12).

Let $\overline{B}(x, R) \subset S^2$ be an arbitrary closed ball, where $x \in S^2$ and $0 < R \le \mathrm{diam}(S^2)$. We use the sets $U^m(x)$ as defined in (8.8) for the map F. Since these sets are closed, Lemma 8.10 (ii) (applied to F) gives the inclusions

$$U^{m+n_0}(x) \subset \overline{B}(x, R) \subset U^{m-n_0}(x),$$

where $m = \lceil -\log(R)/\log(\Lambda_F) \rceil$ and $n_0 \in \mathbb{N}_0$ is a constant independent of the ball.

Since F does not have periodic critical points, the length of the cycle of each vertex is uniformly bounded (Corollary 18.7). This implies that each set $U^m(x)$ consists of a uniformly bounded number of tiles.

Noting that

$$Q := \frac{\log(\deg(f))}{\log(\Lambda)} = \frac{\log(\deg(F))}{\log(\Lambda_F)}$$

and using Proposition 17.11, we obtain

$$\begin{aligned} \nu_F(U^{m+n_0}(x)) \ &\asymp\ \nu_F(U^{m-n_0}(x)) \asymp \deg(F)^{-m} \\ &\asymp\ \exp\big(\log(R)\log(\deg(F))/\log(\Lambda_F)\big) = R^Q, \end{aligned}$$

and so $\nu(\overline{B}(x, R)) = \nu_F(\overline{B}(x, R)) \asymp R^Q$. Here the constants $C(\asymp)$ are independent of the ball. The Ahlfors Q-regularity of (S^2, ϱ, ν) follows.

This in turn implies that $\nu(M) \asymp \mathcal{H}^Q_\varrho(M)$ for every Borel set $M \subset S^2$, where $C(\asymp)$ is independent of M. Since ν is a probability measure, we conclude that $0 < \mathcal{H}^Q_\varrho(S^2) < \infty$. It also follows that the Hausdorff dimension of (S^2, ϱ) is equal to Q. $\qquad\square$

18.3. Quasisymmetry and rational Thurston maps

In this section we prove Theorem 18.1 (ii). We will also derive Theorem 18.4 as a consequence of this and other previous results. The

proof of Theorem 18.1 (ii) mostly follows [**Me02**] and [**Me10**]. It was independently established in [**HP09**] by a different method.

The more difficult implication in Theorem 18.1 (ii) amounts to proving that if a rational Thurston map $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is expanding, then its visual sphere is a quasisphere. For this one wants to show that the chordal metric σ on $\widehat{\mathbb{C}}$ is quasisymmetrically equivalent to each visual metric ϱ for f. The key for this is to analyze the metric properties of the cell decompositions $\mathcal{D}^n(f, \mathcal{C})$ with respect to the chordal metric σ. This is of independent interest and we record the relevant facts in a separate statement. A similar approach to prove quasisymmetric equivalence can be found in [**Ki14**, Theorem 3.4].

PROPOSITION 18.8 (Tiles in the chordal metric). *Let* $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ *be a rational Thurston map without periodic critical points and* $\mathcal{C} \subset \widehat{\mathbb{C}}$ *be a Jordan curve with* $\mathrm{post}(f) \subset \mathcal{C}$. *We equip* $\widehat{\mathbb{C}}$ *with the chordal metric* σ *and denote by* $\mathbf{X}^n$ *for* $n \in \mathbb{N}_0$ *the collection of n-tiles for* $(f, \mathcal{C})$. *Then the following statements are true:*

(i) *If* $X, Y \in \mathbf{X}^n$, $n \in \mathbb{N}_0$, *and* $X \cap Y \neq \emptyset$, *then*
$$\mathrm{diam}(X) \asymp \mathrm{diam}(Y).$$

(ii) *If* $X, Y \in \mathbf{X}^n$, $n \in \mathbb{N}_0$, *and* $X \cap Y = \emptyset$, *then*
$$\mathrm{dist}(X, Y) \gtrsim \mathrm{diam}(X).$$

(iii) *If* $k, n \in \mathbb{N}_0$, $X \in \mathbf{X}^n$, $Y \in \mathbf{X}^{n+k}$, *and* $X \cap Y \neq \emptyset$, *then*
$$\mathrm{diam}(X) \gtrsim \mathrm{diam}(Y).$$

(iv) *If* $k, n \in \mathbb{N}_0$, $X \in \mathbf{X}^n$, $Y \in \mathbf{X}^{n+k}$, *and* $X \cap Y \neq \emptyset$, *then*
$$\mathrm{diam}(X) \lesssim \mathrm{diam}(Y),$$
where $C(\lesssim) = C(k)$.

(v) *Let* $\widetilde{\mathcal{C}} \subset \widehat{\mathbb{C}}$ *be another Jordan curve with* $\mathrm{post}(f) \subset \widetilde{\mathcal{C}}$. *If* $n \in \mathbb{N}_0$, X *is an n-tile for* $(f, \mathcal{C})$, $\widetilde{X}$ *is an n-tile for* $(f, \widetilde{\mathcal{C}})$, *and* $X \cap \widetilde{X} \neq \emptyset$, *then*
$$\mathrm{diam}(X) \asymp \mathrm{diam}(\widetilde{X}).$$

(vi) *If* $x, y \in \widehat{\mathbb{C}}$ *and* $x \neq y$, *then*
$$\sigma(x, y) \asymp \mathrm{diam}(X),$$
whenever $X \in \mathbf{X}^n$ *contains* x, *where* $n = m_{f,\mathcal{C}}(x, y)$.

The implicit multiplicative constant in (i), (ii), (iii), (v) *is independent of the tiles involved in the inequalities and their levels, in* (vi) *independent of* x, y, X, *and in* (iv) *only depends on* k.

So in the previous inequalities we can choose all implicit multiplicative constants only depending on f and $\mathcal{C}$ with the exception of (iv), where we also have dependence on the level difference k of the tiles (but not on the specific tiles). Statements (iii) and (iv) essentially say that if two tiles for $(f, \mathcal{C})$ have non-empty intersection, then their diameters are comparable in the following way: the diameter of the lower-level tile bounds the diameter of the higher-level tile up to a uniform constant, while in the converse inequality the constant depends only on the level difference.

The proof of the previous proposition will be based on two observations. First, the unions of two sets of tiles with the same combinatorics are conformally equivalent. Second, there are only finitely many different combinatorial types, because the local degrees of all iterates of the map are uniformly bounded. The statements can then be derived from Koebe distortion estimates for conformal maps.

The distortion estimates we will need are formulated in Lemma A.2. There they are stated in terms of the chordal metric and spherical derivatives of the conformal map. A subtlety here is that in order to get uniform estimates one has to assume that the image of the map is not too large. In Lemma A.2 we stipulate that the image of the map is contained in a hemisphere, i.e., a chordal disk of radius $\sqrt{2}$. So when we apply Lemma A.2 we have to make sure that this hypothesis is true (see also the discussion after the proof of Theorem A.1).

We now fill in the details of this outline. Let $\mathcal{D}$ be a cell complex. A subset $\mathcal{D}' \subset \mathcal{D}$ is called a *subcomplex* of $\mathcal{D}$ if the following condition is true: if $\tau \in \mathcal{D}'$, $\sigma \in \mathcal{D}$, and $\sigma \subset \tau$, then $\sigma \in \mathcal{D}'$. If $\mathcal{D}'$ is a subcomplex of $\mathcal{D}$, then the cells in $\mathcal{D}'$ form a cell decomposition of the underlying set

$$|\mathcal{D}'| \coloneqq \bigcup \{c : c \in \mathcal{D}'\}.$$

Let $\mathcal{D}$, $\mathcal{D}'$, $\widetilde{\mathcal{D}}$ be cell complexes and suppose we have labelings $L \colon \mathcal{D} \to \widetilde{\mathcal{D}}$ and $L' \colon \mathcal{D}' \to \widetilde{\mathcal{D}}$ (see Definition 5.22). Then we say that an isomorphism $\phi \colon \mathcal{D} \to \mathcal{D}'$ of cell complexes (see Definition 5.10) is *label-preserving* if $L(\tau) = L'(\phi(\tau))$ for each $\tau \in \mathcal{D}$.

Now suppose that $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is a rational Thurston map, and $\mathcal{C} \subset \widehat{\mathbb{C}}$ is a Jordan curve with $\operatorname{post}(f) \subset \mathcal{C}$. We consider the cell decompositions $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$ of $\widehat{\mathbb{C}}$ for $n \in \mathbb{N}_0$. If $\tau \in \mathcal{D}^n$, then $f^n|\tau$ is a homeomorphism of the n-cell τ onto the 0-cell $f^n(\tau)$. So the map $\tau \mapsto f^n(\tau)$ induces a labeling $\mathcal{D}^n \to \mathcal{D}^0$. We call this the *natural labeling* on $\mathcal{D}^n$. Similarly, the map $\tau \mapsto f^n(\tau)$ induces a natural labeling on every subcomplex of $\mathcal{D}^n$.

LEMMA 18.9. *Let $n, m \in \mathbb{N}_0$, $\mathcal{D}$ be a subcomplex of $\mathcal{D}^n$, and $\mathcal{D}'$ be a subcomplex of $\mathcal{D}^m$ both equipped with their natural labelings. If $\phi \colon \mathcal{D} \to \mathcal{D}'$ is a label-preserving isomorphism, then there exists a homeomorphism $h \colon |\mathcal{D}| \to |\mathcal{D}'|$ such that*

(i) *$h(\tau) = \phi(\tau)$ for each $\tau \in \mathcal{D}$,*

(ii) *h maps $\text{int}(|\mathcal{D}|)$ conformally onto $\text{int}(|\mathcal{D}'|)$.*

Here $\text{int}(|\mathcal{D}|)$ and $\text{int}(|\mathcal{D}'|)$ denote the interiors of $|\mathcal{D}|$ and $|\mathcal{D}'|$, respectively, as subsets of $\widehat{\mathbb{C}}$. Roughly speaking, the lemma says that combinatorial equivalence of two subcomplexes $\mathcal{D}$ and $\mathcal{D}'$ gives conformal equivalence of their underlying sets.

Note that since f and all of its iterates are orientation-preserving, f^n maps positively-oriented flags in $\mathcal{D}^n$ to positively-oriented flags in $\mathcal{D}^0$. A similar statement is true for f^m and $\mathcal{D}^m$. Hence ϕ maps positively-oriented flags in $\mathcal{D}$ to positively-oriented flags in $\mathcal{D}'$. This implies that the homeomorphism h in the lemma is necessarily orientation-preserving. Of course, this also follows from the conformality statement (ii) if $\text{int}(|\mathcal{D}|)$ is dense in $|\mathcal{D}|$.

PROOF. If $\tau \in \mathcal{D}$, then $f^n(\tau) = f^m(\phi(\tau))$, because ϕ is label-preserving. Hence we can define a homeomorphism $h_\tau := (f^m|\phi(\tau))^{-1} \circ (f^n|\tau)$ of τ onto $\phi(\tau)$. It is clear that the maps h_τ are compatible under inclusions of cells: if $\sigma, \tau \in \mathcal{D}$ and $\sigma \subset \tau$, then $h_\tau|\sigma = h_\sigma$. We now define a map $h \colon |\mathcal{D}| \to |\mathcal{D}'|$ as follows. For $p \in |\mathcal{D}|$ pick $\tau \in \mathcal{D}$ with $p \in \tau$. Set $h(p) := h_\tau(p)$. As in the proof of Proposition 5.26 one sees that h is well-defined and as in the proof of Lemma 5.11 (i) that h is a homeomorphism of $|\mathcal{D}|$ onto $|\mathcal{D}'|$. Obviously, h has property (i).

To establish property (ii), first note that h maps $\text{int}(|\mathcal{D}|)$ homeomorphically onto $\text{int}(|\mathcal{D}'|)$ (this follows from the "invariance of domain"; see [**Ha02**, Theorem 2B.3, p. 172]). So it suffices to show that h is holomorphic on $U := \text{int}(|\mathcal{D}|) \subset \widehat{\mathbb{C}}$.

The definition of h implies that $f^n = f^m \circ h$ on U. So if $p \in U$ and $q = h(q)$ is not a critical point of f^m, then there exists on open neighborhood V of q where $f^m|V$ has a holomorphic inverse $(f^m|V)^{-1}$. Then $h = (f^m|V)^{-1} \circ f^n$ near p, which shows that that h is holomorphic near p. This implies that h is holomorphic on $U \setminus h^{-1}(\text{crit}(f^m))$. The finitely many points in $h^{-1}(\text{crit}(f^m)) \cap U$ are removable singularities for h, because h is continuous. It follows that h is holomorphic on U as desired. $\square$

PROOF OF PROPOSITION 18.8. Unless otherwise stated, in the following all cells are for $(f, \mathcal{C})$.

(i) For $n \in \mathbb{N}_0$ and $X, Y \in \mathbf{X}^n$ with $X \cap Y \neq \emptyset$, we consider the complex $\mathcal{D}(X, Y)$, equipped with the natural labeling, consisting of all n-cells c for which there exists an n-tile Z with $c \subset Z$ and $Z \cap (X \cup Y) \neq \emptyset$. Obviously,

$$(18.4) \qquad |\mathcal{D}(X, Y)| = \bigcup \{Z \in \mathbf{X}^n : Z \cap (X \cup Y) \neq \emptyset\}.$$

Let $\Omega(X, Y)$ be the interior of $|\mathcal{D}(X, Y)|$. Then $\Omega(X, Y)$ is a region containing X and Y.

Suppose that X', Y' is a pair of non-disjoint m-tiles, $m \in \mathbb{N}_0$. We call $\mathcal{D}(X, Y)$ and $\mathcal{D}(X', Y')$ equivalent if there exists a label-preserving isomorphism $\phi \colon \mathcal{D}(X, Y) \to \mathcal{D}(X', Y')$ with $\phi(X) = X'$ and $\phi(Y) = Y'$. If $\mathcal{D}(X, Y)$ and $\mathcal{D}(X', Y')$ are equivalent, then by Lemma 18.9 there exists a conformal map $h \colon \Omega(X, Y) \to \Omega(X', Y')$ with $h(X) = X'$ and $h(Y) = Y$.

Since f has no periodic critical points, the length of the cycle of each vertex is uniformly bounded (Corollary 18.7). So each n-vertex is contained in a uniformly bounded number of n-tiles independent of n. This implies that the number of n-tiles, and hence the number of n-cells, in $\mathcal{D}(X, Y)$ is uniformly bounded by a number independent of X, Y, and n. Therefore, among the complexes $\mathcal{D}(X, Y)$ there are only finitely many equivalence classes. Since f is expanding, there are also only finitely many complexes $\mathcal{D}(X, Y)$ such that $\Omega(X, Y)$ is not contained in a hemisphere. Hence we can find finitely many complexes $\mathcal{D}(X_1, Y_1), \ldots, \mathcal{D}(X_N, Y_N)$ such that each complex $\mathcal{D}(X, Y)$ not in this list is equivalent to one complex $\mathcal{D}(X_i, Y_i)$ and such that $\Omega(X, Y)$ is contained in a hemisphere. It follows from (A.16) in Lemma A.2 (applied to $A = X_i$, $B = Y_i$, $\Omega = \Omega(X_i, Y_i)$, and the conformal map $h \colon \Omega(X_i, Y_i) \to \Omega(X, Y)$ produced by Lemma 18.9) that $\operatorname{diam}(X) \asymp \operatorname{diam}(Y)$ with $C(\asymp)$ independent of X, Y, and n.

(ii) The argument is very similar to the previous one. For $n \in \mathbb{N}_0$ and $X \in \mathbf{X}^n$, we consider the cell complex $\mathcal{D}(X)$, equipped with the natural labeling, consisting of all n-cells c for which there exists an n-tile Z with $c \subset Z$ and $Z \cap X \neq \emptyset$. Then

$$(18.5) \qquad |\mathcal{D}(X)| = \bigcup \{Z \in \mathbf{X}^n : X \cap Z \neq \emptyset\}.$$

If we define $\Omega(X)$ to be the interior of $|\mathcal{D}(X)|$, then $\Omega(X)$ is a region that contains X. Moreover, if Y is an n-tile with $X \cap Y = \emptyset$, then $\Omega(X)$ is disjoint from Y.

If $m \in \mathbb{N}_0$ and X' is an m-tile, then we call the complexes $\mathcal{D}(X)$ and $\mathcal{D}(X')$ equivalent if there exists a label-preserving isomorphism $\phi \colon \mathcal{D}(X) \to \mathcal{D}(X')$ with $\phi(X) = X'$.

Again there are only finitely many equivalence classes of the complexes $\mathcal{D}(X)$. Based on Lemma 18.9 and (A.14), (A.15) in Lemma A.2, we conclude that for each n-tile Y with $X \cap Y = \emptyset$, we have

$$\operatorname{dist}(X, Y) \geq \operatorname{dist}(X, \partial\Omega(X)) \gtrsim \operatorname{diam}(X),$$

where $C(\gtrsim)$ does not depend on X and Y.

(iii)–(iv) Let $n, k \in \mathbb{N}_0$, $X \in \mathbf{X}^n$, $Y \in \mathbf{X}^{n+k}$, and $X \cap Y \neq \emptyset$.

If W is any n-flower, then any two n-tiles contained in $\overline{W}$ have an n-vertex in common, and hence have comparable diameter by (i). This implies that $\operatorname{diam}(Z) \asymp \operatorname{diam}(W)$ whenever Z is an n-tile with $Z \cap \overline{W} \neq \emptyset$. We also see that $\operatorname{diam}(W) \asymp \operatorname{diam}(W')$, whenever W and W' are n-flowers with $W \cap W' \neq \emptyset$.

Now by Lemma 6.2 (applied for $\widetilde{\mathcal{C}} = \mathcal{C}$) we can cover the $(n+k)$-tile Y with M n-flowers $W_1, \ldots, W_M$, where M is independent of Y. Since X and Y have a point in common, we may assume that $X \cap W_1 \neq \emptyset$. Moreover, since Y is connected, we may assume that each flower in the list meets one of the previous ones. Then

$$\operatorname{diam}(X) \asymp \operatorname{diam}(W_1) \asymp \operatorname{diam}(W_i)$$

for $i = 1, \ldots, M$. This implies

$$\operatorname{diam}(Y) \leq \sum_{i=1}^{M} \operatorname{diam}(W_i) \asymp \operatorname{diam}(W_1) \asymp \operatorname{diam}(X).$$

Since in the previous inequalities all implicit multiplicative constants only depended on f and $\mathcal{C}$, claim (iii) follows.

For (iv) note that by choosing tiles that contain a point in $X \cap Y$, we can find tiles X^i of level $i = n, \ldots, n+k$ such that $X = X^n$, $Y = X^{n+k}$, and $X^i \cap X^{i+1} \neq \emptyset$ for $i = n, \ldots, n+k-1$. If we can show that $\operatorname{diam}(X^i) \lesssim \operatorname{diam}(X^{i+1})$ with a uniform constant $C(\lesssim)$ only depending on f and $\mathcal{C}$, then we conclude that $\operatorname{diam}(X) \lesssim \operatorname{diam}(Y)$ with a constant $C(\lesssim) = C(k)$ as desired.

In other words, we are reduced to proving the inequality $\operatorname{diam}(X) \lesssim \operatorname{diam}(Y)$ under the additional assumption that $Y \in \mathbf{X}^{n+1}$, where the implicit multiplicative constant is supposed to depend only on f and $\mathcal{C}$.

By (i) and Lemma 5.37 (ii) the n-tile X can be covered by a controlled number of $(n+1)$-flowers whose diameters are comparable to $\operatorname{diam}(Y)$. If one combines this with similar estimates as in the previous argument, then (iv) immediately follows.

(v) If X and $\widetilde{X}$ are as in the statement, then by (i) and Lemma 5.38 we can cover $\widetilde{X}$ by M n-flowers $W_1, \ldots, W_M$ for $(f, \mathcal{C})$ with $\operatorname{diam}(W_i) \asymp$

$\operatorname{diam}(X)$ for $i = 1, \dots, M$. Here $C(\asymp)$ and the number M are independent of n, X, and $\widetilde{X}$. By an estimate as in the proof of (iii), we conclude $\operatorname{diam}(\widetilde{X}) \lesssim \operatorname{diam}(X)$. For the other inequality we reverse the roles of X and $\widetilde{X}$ and use an analog of (i) for the curve $\widetilde{\mathcal{C}}$.

(vi) Let $x, y \in \widehat{\mathbb{C}}$ with $x \neq y$ be arbitrary, $n := m_{f,\mathcal{C}}(x, y)$, and X be an n-tile with $x \in X$. By Definition 8.1 there are n-tiles X^n and Y^n with $x \in X^n$, $y \in Y^n$ and $X^n \cap Y^n \neq \emptyset$. Then by (i) we have

$$\operatorname{diam}(X) \asymp \operatorname{diam}(X^n) \asymp \operatorname{diam}(Y^n),$$

and so

$$\sigma(x, y) \leq \operatorname{diam}(X^n) + \operatorname{diam}(Y^n) \asymp \operatorname{diam}(X).$$

On the other hand, pick $(n+1)$-tiles X^{n+1} and Y^{n+1} with $x \in X^{n+1}$ and $y \in Y^{n+1}$. Then $X^{n+1} \cap Y^{n+1} = \emptyset$ by definition of n. We have $x \in X \cap X^{n+1}$, and so $\operatorname{diam}(X^{n+1}) \asymp \operatorname{diam}(X)$ by (iii) and (iv). Then (ii) implies that

$$\sigma(x, y) \geq \operatorname{dist}(X^{n+1}, Y^{n+1}) \gtrsim \operatorname{diam}(X^{n+1}) \asymp \operatorname{diam}(X).$$

Since in the previous inequalities all the implicit multiplicative constants can be chosen independent of x, y, and X, the statement follows. $\qquad\square$

We can now prove the following crucial result.

LEMMA 18.10. *Let $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a rational Thurston map with no periodic critical points. Then each visual metric ϱ for f is quasisymmetrically equivalent to the chordal metric σ.*

PROOF. By Proposition 2.3 the map f is expanding. Let ϱ be a visual metric for f. We have to show that the identity map $\operatorname{id}_{\widehat{\mathbb{C}}} \colon (\widehat{\mathbb{C}}, \varrho) \to (\widehat{\mathbb{C}}, \sigma)$ is a quasisymmetry. We will actually prove that this map is weakly quasisymmetric: there exists a constant $H \geq 1$ such that we have the implication

(18.6) $$\varrho(x, z) \leq \varrho(x, y) \Rightarrow \sigma(x, z) \leq H\sigma(x, y)$$

for all $x, y, z \in \widehat{\mathbb{C}}$. A weak quasisymmetry from a connected doubling space to a doubling space is quasisymmetric (see Proposition 4.4). We may apply this statement, because $(\widehat{\mathbb{C}}, \sigma)$ and $(\widehat{\mathbb{C}}, \varrho)$ are doubling. Since f has no periodic critical points, the last fact follows from Theorem 18.1 (i).

We pick a Jordan curve $\mathcal{C} \subset \widehat{\mathbb{C}}$ with $\operatorname{post}(f) \subset \mathcal{C}$ and denote by $\Lambda > 1$ the expansion factor of ϱ. In the following, we will consider tiles for $(f, \mathcal{C})$.

Now suppose that $x, y, z \in \widehat{\mathbb{C}}$ are points with $\varrho(x, z) \leq \varrho(x, y)$. In the following estimates, all implicit multiplicative constants can be chosen independent of x, y, z.

Define $n = m_{f,\mathcal{C}}(x, y)$ and $l = m_{f,\mathcal{C}}(x, z)$, where $m_{f,\mathcal{C}}$ is as in Definition 8.1. Then

$$\Lambda^{-l} \asymp \varrho(x, z) \leq \varrho(x, y) \asymp \Lambda^{-n}.$$

Hence there exists a constant $k_0 \in \mathbb{N}_0$ independent of x, y, z such $n \leq l + k_0$. Now we pick an n-tile X^n, an l-tile X^l, and an $(l + k_0)$-tile X^{n+k_0} that all contain x. Then

$$\sigma(x, z) \asymp \operatorname{diam}_\sigma(X^l) \qquad \text{by Proposition 18.8 (vi),}$$
$$\asymp \operatorname{diam}_\sigma(X^{l+k_0}) \qquad \text{by Proposition 18.8 (iii) and (iv),}$$
$$\lesssim \operatorname{diam}_\sigma(X^n) \asymp \sigma(x, y) \quad \text{by Proposition 18.8 (iii) and (vi).}$$

Since the implicit constants here are independent of x, y, z, the map $\operatorname{id}_{\widehat{\mathbb{C}}} \colon (\widehat{\mathbb{C}}, \varrho) \to (\widehat{\mathbb{C}}, \sigma)$ is indeed weakly quasisymmetric. The statement follows. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

We are now ready to prove the second part of Theorem 18.1.

PROOF OF THEOREM 18.1 (ii). Let $f \colon S^2 \to S^2$ be an expanding Thurston map, and ϱ be a visual metric for f.

Suppose first that f is topologically conjugate to a rational map. Obviously, this rational map is then itself an expanding Thurston map. Moreover, by Proposition 8.8 the conjugating homeomorphism is a snowflake equivalence with respect to visual metrics, and in particular a quasisymmetry. This implies that in order to show that the visual sphere of f is a quasisphere, we may actually assume that f itself is a rational Thurston map that is expanding.

Then $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ has no periodic critical points by Proposition 2.3. So by Lemma 18.10 the visual metric ϱ is quasisymmetrically equivalent to the chordal metric σ. Hence the identity map $\operatorname{id}_{\widehat{\mathbb{C}}} \colon (\widehat{\mathbb{C}}, \varrho) \to (\widehat{\mathbb{C}}, \sigma)$ is a quasisymmetry, and so $(\widehat{\mathbb{C}}, \varrho)$ is a quasisphere. This proves the first implication of the theorem.

For the converse direction suppose that $f \colon S^2 \to S^2$ is an expanding Thurston map, ϱ is a visual metric for f on S^2, and that there exists a quasisymmetry $h \colon (S^2, \varrho) \to (\widehat{\mathbb{C}}, \sigma)$. Since all visual metrics are snowflake and hence also quasisymmetrically equivalent, we may also assume that ϱ is a visual metric for f satisfying (16.1) in Theorem 16.3.

The map h^{-1} is also a quasisymmetry; so h and h^{-1} are η-quasi-symmetric for some distortion function η. We consider the conjugate $g = h \circ f \circ h^{-1} \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ of f by h.

We claim that the family of iterates $\{g^n \colon n \in \mathbb{N}\}$ is *uniformly quasiregular*, i.e., each map g^n is K-quasiregular with K independent of n (the definition of a quasiregular map was given near the end of Section 4.1). This is true, because with the metric ϱ satisfying (16.1), the map f is locally "conformal", and so the dilatation of $g^n = h \circ f^n \circ h^{-1}$ can be bounded by the dilatations of h and h^{-1}, and hence by a constant independent of n.

To be more precise, let $n \in \mathbb{N}$, $u \in \widehat{\mathbb{C}}$, and for small $\epsilon > 0$ consider points $v, w \in \widehat{\mathbb{C}}$ with $\sigma(u, v) = \sigma(u, w) = \epsilon$. Define $x = h^{-1}(u)$, $y = h^{-1}(v)$, $z = h^{-1}(w)$. By Theorem 16.3 (ii) we have that if $\epsilon > 0$ is sufficiently small (depending on u and n), then

$$
\begin{aligned}
\frac{\sigma(g^n(u), g^n(v))}{\sigma(g^n(u), g^n(w))} &= \frac{\sigma(h(f^n(x)), h(f^n(y)))}{\sigma(h(f^n(x)), h(f^n(z)))} \\
&\leq \eta\left(\frac{\varrho(f^n(x), f^n(y))}{\varrho(f^n(x), f^n(z))}\right) = \eta\left(\frac{\varrho(x, y)}{\varrho(x, z)}\right) \\
&= \eta\left(\frac{\varrho(h^{-1}(u), h^{-1}(v))}{\varrho(h^{-1}(u), h^{-1}(w))}\right) \leq H := \eta(\eta(1)).
\end{aligned}
$$

Hence

$$
H(g^n, u) :=
$$

$$
\limsup_{\epsilon \to 0} \max\left\{ \frac{\sigma(g^n(u), g^n(v))}{\sigma(g^n(u), g^n(w))} : v, w \in \widehat{\mathbb{C}}, \, \sigma(u, v) = \sigma(u, w) = \epsilon \right\}
$$

$$
\leq H
$$

for all $u \in \widehat{\mathbb{C}}$ and $n \in \mathbb{N}$. This inequality implies that g^n is locally H-quasiconformal on the set $\widehat{\mathbb{C}} \setminus \operatorname{crit}(g^n)$ (according to the so-called "metric" definition of quasiconformality; see [**Vä71**, Section 34]).

In particular, this implies that $g^n | \widehat{\mathbb{C}} \setminus \operatorname{crit}(g^n)$ is K-quasiregular with $K = K(H)$ independent of n. Since the finite set $\operatorname{crit}(g^n)$ is removable for quasiregularity (see [**Ri93**, Section 7.1]), we conclude that g^n is K-quasiregular with K independent of n.

So the family of iterates $\{g^n \colon n \in \mathbb{N}\}$ of g is uniformly quasiregular. This implies that g is topologically conjugate to a rational map (see [**IM01**, Theorem 21.5.2]). Hence f is also topologically conjugate to a rational map. $\qquad\square$

PROOF OF THEOREM 18.4. By Proposition 2.3 the map f is an expanding Thurston map (and hence also every iterate of f). So by

Theorem 15.1 for each large $n \in \mathbb{N}$ there exists a Jordan curve $\mathcal{C} \subset \widehat{\mathbb{C}}$ with $\mathrm{post}(f) \subset \mathcal{C}$ that is f^n-invariant. Equipped with the chordal metric, every such curve is a quasicircle as follows from Theorem 15.3 and Lemma 18.10 (note that the class of visual metrics for f and for any iterate of f are the same). Statements (i) and (ii) follow.

Suppose the curve $\mathcal{C}$ is actually f-invariant, and consider cells for $(f, \mathcal{C})$. Then by Proposition 15.26 the family of all edges consists of uniform quasiarcs if the underlying metric on $\widehat{\mathbb{C}}$ is a visual metric for f. Again by Lemma 18.10 we can switch to the chordal metric σ in this statement.

Similarly, the family of the boundaries of all tiles consists of uniform quasicircles for σ. Now it is a standard fact that a closed Jordan region $X \subset \widehat{\mathbb{C}}$ bounded by a quasicircle ∂X is a quasidisk. More precisely, if $h\colon \partial \mathbb{D} \to \partial X$ is an η-quasisymmetry, then it can be extended to an $\tilde{\eta}$-quasisymmetry $H\colon \overline{\mathbb{D}} \to X$. Here $\tilde{\eta}$ depends not only on η, but also on a lower bound for the diameter of $\widehat{\mathbb{C}} \setminus X$ (see [**Bo11**, Proposition 5.3 (ii)]). In particular, if we have a family of such Jordan regions whose boundaries form a family of uniform quasicircles, then the family will consist of uniform quasidisks, if there is a positive uniform lower bound for the diameters of the complements of the regions. Since f is expanding, there are only finitely many tiles not contained in hemispheres, and so we have such a uniform lower bound for the family of all tiles for $(f, \mathcal{C})$. Statement (iii) follows. $\square$

EXAMPLE 18.11. The "if" part of Theorem 18.1 (ii) is not true if one only requires that f is Thurston equivalent to a rational map. Indeed, in Example 12.21 we considered two Thurston maps f_2 and $\widetilde{f}_2$ on a triangular pillow identified with the Riemann sphere $\widehat{\mathbb{C}}$. Here $f_2\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is rational and not expanding (it has a critical fixed point), while $\widetilde{f}_2\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is an expanding (non-rational) Thurston map. Both maps realize the barycentric subdivision rule and are hence Thurston equivalent (Proposition 12.3). So $\widetilde{f}_2$ is Thurston equivalent to a rational map.

Now $\widetilde{f}_2$ has a critical fixed point. So if ϱ is a visual metric for $\widetilde{f}_2$, then $(\widehat{\mathbb{C}}, \varrho)$ is not doubling by Theorem 18.1 (i). On the other hand, each quasisphere is doubling, because $(\widehat{\mathbb{C}}, \sigma)$ is doubling and this condition is invariant under quasisymmetries. Hence $(\widehat{\mathbb{C}}, \varrho)$ cannot be a quasisphere.

CHAPTER 19

Measure-theoretic properties of rational expanding Thurston maps

In this chapter we consider *rational* expanding Thurston maps f on the Riemann sphere $\widehat{\mathbb{C}}$. We will investigate various measures that are associated with f. This will allow us to complete the proof of Theorem 18.1 by providing the missing justification for part (iii) of this theorem (see the end of Section 19.4). Theorem 19.4 below will be crucial for the characterization of Lattès maps given in Chapter 20. None of this material is new. We include it to make our presentation more self-contained.

We equip the Riemann sphere $\widehat{\mathbb{C}}$ with (normalized) Lebesgue measure $\mathcal{L}_{\widehat{\mathbb{C}}}$ given by

$$d\mathcal{L}_{\widehat{\mathbb{C}}}(z) = \frac{1}{\pi(1+|z|^2)^2} \, d\mathcal{L}_{\mathbb{C}}(z).$$

The normalization means that $\mathcal{L}_{\widehat{\mathbb{C}}}(\widehat{\mathbb{C}}) = 1$. Mostly, we will drop the subscript and simply write $\mathcal{L} = \mathcal{L}_{\widehat{\mathbb{C}}}$ if no ambiguity can arise. Throughout this chapter we will use "Polish notation" and denote by $|z - w|$ the chordal distance $\sigma(z, w)$ of two points $z, w \in \widehat{\mathbb{C}}$. All metric notions will refer to the chordal metric unless otherwise mentioned.

A first link to measure-theoretic dynamics and ergodic theory is provided by the following statement.

THEOREM 19.1. *Let* $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ *be a rational expanding Thurston map. Then Lebesgue measure* $\mathcal{L}_{\widehat{\mathbb{C}}}$ *is ergodic for* f.

Note that $\mathcal{L}_{\widehat{\mathbb{C}}}$ is essentially never f-invariant, but ergodicity is interpreted as for f-invariant measures (see the discussion in Section 17.1): if $A \subset \widehat{\mathbb{C}}$ is a Borel set with $f^{-1}(A) = A$ and $\mathcal{L}_{\widehat{\mathbb{C}}}(A) > 0$, then $\mathcal{L}_{\widehat{\mathbb{C}}}(A) = 1$.

Theorem 19.1 is well known. We will present a proof in Section 19.2, where we follow the argument from [**McM94a**, Theorem 3.9] closely.

In our context one can actually find an f-invariant measure that is absolutely continuous with respect to Lebesgue measure.

THEOREM 19.2. *Let $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a rational expanding Thurston map. Then there exists a unique f-invariant probability measure λ_f on $\widehat{\mathbb{C}}$ that is absolutely continuous with respect to Lebesgue measure $\mathcal{L}_{\widehat{\mathbb{C}}}$. This measure has the form $d\lambda_f = \rho\, d\mathcal{L}_{\widehat{\mathbb{C}}}$, where ρ is a positive continuous function on $\widehat{\mathbb{C}}\backslash\mathrm{post}(f)$. Moreover, the measure λ_f is ergodic for f.*

It immediately follows from the second part of this statement that λ_f and $\mathcal{L}_{\widehat{\mathbb{C}}}$ are mutually absolutely continuous and are hence in the same *measure class*, i.e., the two measures have precisely the same (Borel) null-sets. The points in $\mathrm{post}(f)$ are singularities for the Radon-Nikodym derivative $\rho = d\lambda_f/d\mathcal{L}_{\widehat{\mathbb{C}}}$. One can describe the asymptotic behavior near these points explicitly (see Proposition 19.13).

Theorem 19.2 is again a well-known statement. The first part is actually true for more general rational maps (see, for example, [**GPS90**, Theorem 3]). We will present the proof of Theorem 19.2 in Section 19.3. There we will reinterpret the existence problem for the measure λ_f as a fixed point problem for a certain operator, the *Ruelle* or *transfer operator* acting on the space of continuous functions on $\widehat{\mathbb{C}} \setminus \mathrm{post}(f)$.

There is another natural measure for f that is in the same measure class as $\mathcal{L}_{\widehat{\mathbb{C}}}$, namely the *canonical orbifold measure* $\Omega = \Omega_f$ that is associated with the orbifold $\mathcal{O}_f = (\widehat{\mathbb{C}}, \alpha_f)$ of f. To quickly review its definition (see (A.56) in Section A.10 for more details), let $\Theta\colon X \to \widehat{\mathbb{C}}$ be the universal orbifold covering map of $\mathcal{O}_f$. Here $X = \mathbb{D}$ or $X = \mathbb{C}$ depending on whether $\mathcal{O}_f$ is hyperbolic or parabolic. Roughly speaking, Ω is then the "local" push-forward of the natural measure $\mathcal{L}_X$ on X, namely hyperbolic area in the hyperbolic and Euclidean area in the parabolic case. More precisely, Ω is the unique measure on $\widehat{\mathbb{C}}$ such that for the Jacobian $J_{\Theta,\Omega,\mathcal{L}_X}$ of Θ with respect to $\mathcal{L}_X$ and Ω we have $J_{\Theta,\mathcal{L}_X,\Omega} \equiv 1$ on X (see Section 19.1 for a general discussion of Jacobians). In the hyperbolic case Ω is independent of the choice of Θ and hence unique, but in the parabolic case Ω is only unique up to a positive multiplicative constant. One can enforce uniqueness in the parabolic case by the normalization $\Omega(\widehat{\mathbb{C}}) = 1$.

A Lattès map f has a parabolic orbifold $\mathcal{O}_f$. In this case the normalized measure $\Omega = \Omega_f$ is equal to the measure of maximal entropy and the measure in Theorem 19.2.

THEOREM 19.3. *Let $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a Lattès map. Suppose ν_f is its measure of maximal entropy, λ_f the unique f-invariant probability measure that is absolutely continuous with respect to $\mathcal{L}_{\widehat{\mathbb{C}}}$, and Ω_f the canonical orbifold measure of $\mathcal{O}_f$ normalized such that $\Omega_f(\widehat{\mathbb{C}}) = 1$.*

Then

$$\nu_f = \lambda_f = \Omega_f.$$

An immediate consequence of the previous theorem is that for a Lattès map the measure of maximal entropy ν_f is absolutely continuous with respect to $\mathcal{L}_{\widehat{\mathbb{C}}}$. This property actually characterizes these maps among rational expanding Thurston maps.

THEOREM 19.4. *Let $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a rational expanding Thurston map. Then its measure of maximal entropy ν_f is absolutely continuous with respect to Lebesgue measure $\mathcal{L}_{\widehat{\mathbb{C}}}$ if and only if f is a Lattès map.*

A much stronger version of this theorem is actually true. Namely, according to a result by A. Zdunik [**Zd90**] Lattès maps are characterized by this property among *all* rational maps, and not only among rational expanding Thurston maps.

This chapter is organized as follows. In Section 19.1 we review some general facts about Jacobians. Section 19.2 is devoted to the proof of Theorem 19.1 while Theorems 19.2 and 19.3 are established in Section 19.3. In Section 19.4 we give a characterization of Lattès maps (see Theorem 19.11) that will lead to the proofs of Theorem 19.4 and of Theorem 18.1 (iii).

19.1. The Jacobian of a measurable map

In this section we discuss some general facts about Jacobians. We are mostly interested in Jacobians of holomorphic maps on the Riemann sphere, but we will discuss the subject in greater generality. For more background see [**PU10**, Section 2.9].

Recall that a *measure space* is a triple $(X, \mathcal{F}, \mu)$ of consisting a set X, a σ-algebra $\mathcal{F}$ on X, and a measure μ defined on the sets in $\mathcal{F}$. Let $(X', \mathcal{F}', \mu')$ be another measure space, and $T\colon X \to X'$ be a measurable map, i.e., T satisfies $T^{-1}(A) \in \mathcal{F}$ whenever $A \in \mathcal{F}'$. In this context we call a set $A \subset X$ *admissible* if $A \in \mathcal{F}$, $T(A) \in \mathcal{F}'$, and T is injective on A. Then a measurable function $J_T\colon X \to [0, \infty]$ is called a *Jacobian* of T if for each admissible set $A \subset X$, we have

$$\mu'(T(A)) = \int_A J_T \, d\mu.$$

We will sometimes write $J_T = J_{T,\mu,\mu'}$ if we also want to mention the measures involved. Often our measure spaces will be identical, i.e., $(X, \mathcal{F}, \mu) = (X', \mathcal{F}', \mu')$. Then we write $J_f = J_\mu = J_{f,\mu}$ based on which dependence we want to emphasize.

If $A \subset X$ is an admissible set, we can define a measure ν_A on A given as

$$\nu_A(M) = \mu'(T(M)), \text{ whenever } M \subset A \text{ is admissible.}$$

A sufficient condition for the existence of a Jacobian J_T is that there exists a countable partition of X into admissible sets such that for each set $A \subset X$ in the partition the measure ν_A is absolutely continuous with respect to μ. In this case, ν_A is absolutely continuous with respect to μ for each admissible set $A \subset X$. Moreover, J_T is uniquely determined μ-almost everywhere on X, because on each admissible set A it is equal to the Radon-Nikodym derivative $d\nu_A/d\mu$.

If J_T is a Jacobian of T, $A \subset X$ is an admissible set, and $\rho' \colon X' \to [0,\infty]$ a non-negative measurable function, then

$$(19.1) \qquad \int_{T(A)} \rho' \, d\mu' = \int_A (\rho' \circ T) \cdot J_T \, d\mu.$$

A similar relation holds for all integrable functions $\rho' \in L^1(\mu')$.

Let $(X'', \mathcal{F}'', \mu'')$ be a third measure space, and $S \colon X' \to X''$ be a measurable map. Then a chain rule for Jacobians is valid: If the Jacobians J_T and J_S exist, then a Jacobian of $S \circ T$ is given as

$$J_{S \circ T} = (J_S \circ T) \cdot J_T.$$

We will only be interested in the cases where the spaces are $\mathbb{D}, \mathbb{C}$, or $\widehat{\mathbb{C}}$ equipped with their corresponding Borel σ-algebras, and the measures are absolutely continuous with respect to Lebesgue measures. In addition, $T = f$ will be a holomorphic map. Then the existence of Jacobians follows from the transformation formula for integrals.

For example, let $X = X' = \widehat{\mathbb{C}}$ and $\mathcal{L} = \mathcal{L}_{\widehat{\mathbb{C}}}$ be normalized Lebesgue measure on $\widehat{\mathbb{C}}$. Suppose μ and μ' are Borel measures on $\widehat{\mathbb{C}}$ that are absolutely continuous with respect to $\mathcal{L}$. Then $d\mu = \kappa \, d\mathcal{L}$ and $d\mu' = \kappa' \, d\mathcal{L}$ for some non-negative measurable functions κ and κ' on $\widehat{\mathbb{C}}$. Let us assume in addition that κ is positive $\mathcal{L}$-almost everywhere on $\widehat{\mathbb{C}}$. Then if $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is a rational function, we have

$$(19.2) \qquad J_{f,\mu,\mu'}(z) = \frac{\kappa'(f(z))}{\kappa(z)} f^\sharp(z)^2$$

for $\mathcal{L}$-almost every $z \in \mathbb{C}$, where

$$f^\sharp(z) = |f'(z)| \frac{1 + |z|^2}{1 + |f(z)|^2}$$

is the spherical derivative of f. In particular, if $\mu = \mu' = \mathcal{L}$, then

$$(19.3) \qquad J_f = J_{f,\mathcal{L}} = (f^\sharp)^2.$$

Suppose now that $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is a postcritically-finite rational map (i.e., a rational Thurston map), and that $\mu = \mu'$ is an f-invariant probability measure on $\widehat{\mathbb{C}}$. We pick a Jordan curve $\mathcal{C} \subset \widehat{\mathbb{C}}$ with $\mathrm{post}(f) \subset \mathcal{C}$ and $\mathcal{L}(\mathcal{C}) = \mu(\mathcal{C}) = 0$ and consider tiles for $(f, \mathcal{C})$. The map $f|X$ is a homeomorphism onto its image for each 1-tile $X \in \mathbf{X}^1$. This shows that 1-tiles are admissible and it follows that

$$(19.4) \qquad \int_{\widehat{\mathbb{C}}} J_{f,\mu}\, d\mu = \sum_{X \in \mathbf{X}^1} \int_X J_{f,\mu}\, d\mu = \sum_{X \in \mathbf{X}^1} \mu(f(X)) = \deg(f).$$

The measure-theoretic entropy h_μ of μ for the given map f can be expressed using the Jacobian as

$$(19.5) \qquad h_\mu = \int_{\widehat{\mathbb{C}}} \log(J_{f,\mu})\, d\mu.$$

This is known as *Rokhlin's formula* (see, for example, [**PU10**, Theorem 2.9.7]).

If we combine (19.4) and (19.5) with Jensen's inequality, we recover the inequality

$$h_\mu = \int_{\widehat{\mathbb{C}}} \log(J_{f,\mu})\, d\mu \leq \log\left(\int_{\widehat{\mathbb{C}}} J_{f,\mu}\, d\mu \right) = \log(\deg(f))$$

that we derived in Chapter 17 for an arbitrary expanding Thurston map f.

19.2. Ergodicity of Lebesgue measure

In this section we will prove Theorem 19.1. To prepare the proof, we fix a rational expanding Thurston map $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ and define $\mathbf{V}^\infty = \bigcup_{n \in \mathbb{N}_0} f^{-n}(\mathrm{post}(f))$. We require a lemma.

LEMMA 19.5. *There exists a neighborhood $U \subset \widehat{\mathbb{C}}$ of $\mathrm{post}(f)$ with the following property: if $z_0 \in \widehat{\mathbb{C}} \setminus \mathbf{V}^\infty$ is arbitrary and if we define $z_n = f^n(z_0)$ for $n \in \mathbb{N}_0$, then there exists a subsequence $\{z_{n_k}\}$ of $\{z_n\}$ such that $z_{n_k} \in \widehat{\mathbb{C}} \setminus U$ for all $k \in \mathbb{N}$.*

PROOF. For the proof it is convenient to use a visual metric ϱ for f as provided by Theorem 16.3 (ii). Let $\Lambda > 1$ be the expansion factor of ϱ. If we pick $k_0 \in \mathbb{N}$ sufficiently large, then for each $p \in \mathrm{post}(f)$ the ball $U_p := B_\varrho(p, \Lambda^{-k_0})$ satisfies (16.1), i.e.,

$$\varrho(f(x), f(p)) = \Lambda \varrho(x, p) \text{ for all } x \in U_p,$$

and we also have

$$\varrho(p, q) \geq 2\Lambda^{-k_0+1} \text{ for distinct points } p, q \in \mathrm{post}(f).$$

In particular, the balls U_p and U_q are disjoint for distinct $p, q \in \text{post}(f)$. Now define $U = \bigcup_{p \in \text{post}(f)} U_p$. This is an open neighborhood of the set $\text{post}(f)$.

To see that U has the property as in the statement of the lemma, let $z_0 \in \widehat{\mathbb{C}} \setminus \mathbf{V}^\infty$ be arbitrary and consider $z_n = f^n(z_0)$ for some $n \in \mathbb{N}$. It suffices to show that there exists $m \geq n$ such that $z_m = f^m(z_0) \in \widehat{\mathbb{C}} \setminus U$.

If $z_n \in \widehat{\mathbb{C}} \setminus U$ we are done. So assume $z_n \in U$; then $z_n \in U_p$ for a point $p \in \text{post}(f)$. Since $z_0 \notin \mathbf{V}^\infty$, we have $z_n \notin \text{post}(f)$ and so $z_n \neq p$. On the other hand, $z_n \in U_p = B_\varrho(p, \Lambda^{-k_0})$, and so there exists a number $k \in \mathbb{N}$, $k \geq k_0$ such that $\Lambda^{-k-1} \leq \varrho(z_n, p) < \Lambda^{-k}$. We now consider two cases.

Case 1: We have $k = k_0$ and so $\Lambda^{-k_0-1} \leq \varrho(z_n, p) < \Lambda^{-k_0}$. Then $\Lambda^{-k_0} \leq \varrho(f(z_n), f(p)) < \Lambda^{-k_0+1}$. So if $q = f(p)$, then $q \in \text{post}(f)$ and $z_{n+1} = f(z_n) \notin U_q$. Moreover, for each $p' \in \text{post}(f) \setminus \{q\}$ we have

$$\varrho(z_{n+1}, p') \geq \varrho(p', q) - \varrho(z_{n+1}, q) \geq 2\Lambda^{-k_0+1} - \Lambda^{-k_0+1} \geq \Lambda^{-k_0},$$

and so $z_{n+1} \notin U_{p'}$. Thus $z_{n+1} \notin U$ as desired.

Case 2: $\Lambda^{-k-1} \leq \varrho(z_n, p) < \Lambda^{-k}$ for some $k > k_0$. In this case, $z_{n+k-k_0} = f^{k-k_0}(z_n)$ satisfies the assumptions of Case 1 for the point $p' = f^{k-k_0}(p) \in \text{post}(f)$, and so $z_{n+k-k_0+1} \in \widehat{\mathbb{C}} \setminus U$ as desired. $\square$

In the proof of Theorem 19.1 we need the *Lebesgue density theorem.* For Lebesgue measure $\mathcal{L}$ on $\widehat{\mathbb{C}}$ this theorem says that if $A \subset \widehat{\mathbb{C}}$ is a Borel set, the $\mathcal{L}$-almost every point $z_0 \in A$ is a *(Lebesgue) density point* of A, meaning that

$$(19.6) \qquad \lim_{\epsilon \to 0^+} \frac{\mathcal{L}(B(z_0, \epsilon) \cap A)}{\mathcal{L}(B(z_0, \epsilon))} = 1.$$

Here and in the following balls are defined with respect to the chordal metric on $\widehat{\mathbb{C}}$. We need a slight modification of (19.6) where we allow "roundish" sets with controlled "eccentricity" instead of metric balls.

LEMMA 19.6. *Let $A \subset \widehat{\mathbb{C}}$ be a Borel set and $z_0 \in A$ be a density point of A. Suppose that $V_n \subset \widehat{\mathbb{C}}$ is a Borel set for $n \in \mathbb{N}$ such that*

$$B(z_0, r_n/K) \subset V_n \subset B(z_0, r_n),$$

where $K \geq 1$, $r_n > 0$, and $r_n \to 0$ as $n \to \infty$. Then

$$\lim_{n \to \infty} \frac{\mathcal{L}(V_n \cap A)}{\mathcal{L}(V_n)} = 1.$$

PROOF. Note that (19.6) is equivalent to

$$\frac{\mathcal{L}(B(z_0, \epsilon) \setminus A)}{\mathcal{L}(B(z_0, \epsilon))} \to 0 \text{ as } \epsilon \to 0^+.$$

Suppose V_n, $n \in \mathbb{N}$, are sets as in the statement. Define $B_n = B(z_0, r_n)$ and $B_n' = B(z_n, r_n/K)$ for $n \in \mathbb{N}$. Then

$$\mathcal{L}(V_n \setminus A) \leq \mathcal{L}(B_n \setminus A)$$

and

$$\mathcal{L}(V_n) \geq \mathcal{L}(B_n') \gtrsim \mathcal{L}(B_n),$$

where $C(\gtrsim)$ is independent of n. Hence

$$\frac{\mathcal{L}(V_n \setminus A)}{\mathcal{L}(V_n)} \lesssim \frac{\mathcal{L}(B_n \setminus A)}{\mathcal{L}(B_n)} \to 0$$

as $n \to \infty$, and the statement follows. $\square$

After these preparations we are ready to prove the main result of this section.

PROOF OF THEOREM 19.1. Let $A \subset \widehat{\mathbb{C}}$ be a Borel set with $\mathcal{L}(A) > 0$ and $f^{-1}(A) = A$. We have to show that A has full measure, i.e., $\mathcal{L}(A) = 1$. Since f is surjective, we also have $f(A) = A$, and so A is *fully invariant* under f in the sense that $f^{-1}(A) = A = f(A)$.

Before we delve into the details, let us give an outline of the argument. By the Lebesgue density theorem, we can find a sequence of small balls where A has density approaching 1. We then use the dynamics and blow up these balls by iterates of f to find a sequence of balls of fixed size with this property. Passing to a subsequential limit, we obtain a ball B where A has density 1. Since f is eventually onto and A is fully invariant, we conclude that A must have full measure. We now present the details.

Claim 1. There exists $\delta > 0$, and balls $B_k = B(w_k, \delta) \subset \widehat{\mathbb{C}}$ for $k \in \mathbb{N}$ such that $\mathcal{L}(B_k \setminus A) \to 0$ as $k \to \infty$.

The main point here is that the balls B_k have a fixed radius. To prove the claim, first note that $\mathbf{V}^\infty = \bigcup_{n \in \mathbb{N}_0} f^{-n}(\mathrm{post}(f))$ is a countable set, A has positive measure, and so we can find a Lebesgue density point $z_0 \in A \setminus \mathbf{V}^\infty$ of A. Now let $U \subset \widehat{\mathbb{C}}$ be a neighborhood of $\mathrm{post}(f)$ as provided by Lemma 19.5. Then there is a subsequence $\{z_{n_k}\}$ of the sequence $\{z_n\}$ given by $z_n = f^n(z_0)$ for $n \in \mathbb{N}_0$ such that $z_{n_k} \in \widehat{\mathbb{C}} \setminus U$ for all $k \in \mathbb{N}$.

Let $\delta_0 := \mathrm{dist}(\mathrm{post}(f), \widehat{\mathbb{C}} \setminus U) > 0$. Then the disk $B_k' := B(z_{n_k}, \delta_0)$ is a simply connected region contained in $\widehat{\mathbb{C}} \setminus \mathrm{post}(f)$. In particular,

each iterate is a covering map over B'_k and so there exists a conformal map g_k on B'_k that is an inverse branch of f^{-n_k} and sends z_{n_k} to z_0.

Let $B_k := B(z_{n_k}, \delta_0/2)$ and $V_k := g_k(B_k)$. We note that the diameters of the sets V_k tend to 0 as $k \to \infty$. The quickest way to see this is to use the canonical orbifold metric ω of f. Namely, it follows from Proposition A.34 that there is a constant $\rho > 1$ such that

$$\operatorname{diam}_\omega(V_k) \lesssim \rho^{-n_k}$$

for $k \in \mathbb{N}$ with $C(\lesssim)$ independent of k. Since ω induces the standard topology on $\widehat{\mathbb{C}}$, it follows that $\operatorname{diam}_\sigma(V_k) \to 0$ as $k \to \infty$ for the chordal metric σ.

Koebe's distortion theorem (see (A.10) in Theorem A.1) implies that the sets V_n satisfy the assumption of Lemma 19.6 for the Lebesgue point $z_0 \in V_k$. Thus

$$\frac{\mathcal{L}(V_k \cap A)}{\mathcal{L}(V_k)} \to 1$$

as $k \to \infty$. Since A is f-invariant, it again follows from Koebe's distortion theorem (more precisely, we apply (A.8) in Theorem A.1) that

$$\mathcal{L}(V_k \setminus A) \asymp \mathcal{L}(B_k \setminus A) J_{g_k}(z_{n_k}) \asymp \mathcal{L}(B_k \setminus A)\frac{\mathcal{L}(V_k)}{\mathcal{L}(B_k)}$$

$$\asymp \mathcal{L}(B_k \setminus A)\mathcal{L}(V_k).$$

Here $J_{g_k} = (g_k^\sharp)^2$ is the Jacobian of g_k with respect to $\mathcal{L}$ (see (19.3)) and the constants $C(\asymp)$ are independent of k. Hence

$$\mathcal{L}(B_k \setminus A) \asymp \frac{\mathcal{L}(V_k \setminus A)}{\mathcal{L}(V_k)} \to 0$$

as $k \to \infty$. Claim 1 follows (with $w_k = z_{n_k}$ and $\delta = \delta_0/2$).

Claim 2. There exists an open ball $B \subset \widehat{\mathbb{C}}$ with $\mathcal{L}(B \setminus A) = 0$.

Indeed, if $B_k = B(w_k, \delta)$, $k \in \mathbb{N}$, is a sequence of balls as in Claim 1, then by compactness $\{w_k\}$ has a convergent subsequence which we still denote by $\{w_k\}$ for convenience. Let $w_\infty := \lim_{k \to \infty} w_k$ and $B := B(w_\infty, \delta/2)$.

Then $B \subset B_k$ for sufficiently large k and so $\mathcal{L}(B \setminus A) \leq \mathcal{L}(B_k \setminus A)$. Since the right hand side tends to 0 as $k \to \infty$, we conclude that $\mathcal{L}(B \setminus A) = 0$. Claim 2 is proved.

We know (see Lemma 6.5) that f is eventually onto; this implies that there exists $n \in \mathbb{N}$ such that $f^n(B) = \widehat{\mathbb{C}}$ if B is a ball as in the previous claim. Since A is completely f-invariant, we then have $\widehat{\mathbb{C}} \setminus A = f^n(B \setminus A)$. Now f^n, as a rational map, preserves sets of

measure zero, and so $\mathcal{L}(\widehat{\mathbb{C}} \setminus A) = 0$. This implies that $\mathcal{L}(A) = 1$. Thus $\mathcal{L}$ is ergodic for f. $\qquad\square$

We record a quick consequence of Theorem 19.1.

COROLLARY 19.7. *Let $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a rational expanding Thurston map and μ an f-invariant probability measure that is absolutely continuous with respect to $\mathcal{L}_{\mathbb{C}}$. Then μ is ergodic for f.*

PROOF. Let $A \subset \widehat{\mathbb{C}}$ be a Borel set with $f^{-1}(A) = A$ and $\mu(A) > 0$. We have to show that $\mu(A) = 1$.

Now $\mu(A) > 0$ and so $\mathcal{L}(A) > 0$ since μ is absolutely continuous with respect to $\mathcal{L}$. Hence $\mathcal{L}(A) = 1$ as follows from Theorem 19.1, or equivalently $\mathcal{L}(\widehat{\mathbb{C}} \setminus A) = 0$. This implies $\mu(\widehat{\mathbb{C}} \setminus A) = 0$ which gives $\mu(A) = 1$ as desired. $\qquad\square$

19.3. The absolutely continuous invariant measure

In this section $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is again a rational Thurston map. We assume that f is expanding, or equivalently, that f has no periodic critical points (see Proposition 2.3). Our goal is to construct an f-invariant probability measure $\lambda = \lambda_f$ on $\widehat{\mathbb{C}}$ that is absolutely continuous with respect to (normalized) Lebesgue measure $\mathcal{L} = \mathcal{L}_{\widehat{\mathbb{C}}}$ on $\widehat{\mathbb{C}}$.

In order to motivate our approach, let us consider a finite (Borel) measure μ on $\widehat{\mathbb{C}}$ that is absolutely continuous with respect to $\mathcal{L}$. Then $d\mu = \rho \, d\mathcal{L}$ for a suitable integrable function ρ on $\widehat{\mathbb{C}}$. As we will see momentarily, then $f_*\mu$ is also absolutely continuous with respect to $\mathcal{L}$ and hence can be written in the form $d(f_*\mu) = \mathcal{R}(\rho) \, d\mathcal{L}$, where $\mathcal{R}(\rho)$ is integrable. The requirement that μ is f-invariant, i.e., the relation $f_*\mu = \mu$, then translates into the condition $\mathcal{R}(\rho) = \rho$.

Let us assume that ρ has some additional regularity, namely, that it is a non-negative continuous function outside the finite subset $\mathrm{post}(f)$ of $\widehat{\mathbb{C}}$. If $p \in \widehat{\mathbb{C}} \setminus \mathrm{post}(f)$, then p is not a critical value of f and so it has precisely $d = \deg(f)$ preimage points $q_1, \ldots, q_d$. Near each of these points q_j the map f is a conformal map. In particular, there exists an open neighborhood V of p, and pairwise disjoint open neighborhoods U_j of q_j such that $f|U_j$ is a conformal map of U_j onto V for $j = 1, \ldots, d$.

Let $g_j := (f|U_j)^{-1}$. We will now consider Jacobians of these maps with $\mathcal{L}$ as the underlying measure. Recall that if h is a holomorphic map defined on a subset of $\widehat{\mathbb{C}}$, then $J_h = J_{h,\mathcal{L}} = (h^\sharp)^2$, where $h^\sharp$ denotes the spherical derivative (see (19.3)). For the Jacobians of g_j and f we have the relation

$$(19.7) \qquad\qquad J_{g_j}(w) = J_f(g_j(w))^{-1}$$

for $w \in V$ which follows from the chain rule.

Now consider a Borel set $B \subset V$. Then its preimage under f decomposes into the Borel sets $A_j = g_j(B)$, $j = 1, \ldots, d$. It follows that

$$(19.8) \qquad f_*\mu(B) = \mu(f^{-1}(B)) = \sum_{j=1}^{d} \mu(A_j) = \sum_{j=1}^{d} \int_{A_j} \rho \, d\mathcal{L}$$

$$= \sum_{j=1}^{d} \int_B (\rho \circ g_j) \cdot J_{g_j} \, d\mathcal{L}$$

$$= \int_B \sum_{j=1}^{d} \rho(g_j(w)) \cdot J_f(g_j(w))^{-1} \, d\mathcal{L}(w)$$

$$= \int_B \sum_{z \in f^{-1}(w)} \rho(z) J_f(z)^{-1} \, d\mathcal{L}(w).$$

The equality between the first and the last terms in this identity actually remains valid for arbitrary Borel sets $B \subset \widehat{\mathbb{C}} \setminus \mathrm{post}(f)$, because we can split each such set B into countably many disjoint pieces such that each of these pieces lies in a suitable set V as chosen above.

It follows that $f_*\mu$ is also absolutely continuous with respect to $\mathcal{L}$ and has a Radon-Nikodym derivative equal to

$$(19.9) \qquad \mathcal{R}(\rho)(w) = \sum_{z \in f^{-1}(w)} \rho(z) J_f(z)^{-1}$$

for $w \in \widehat{\mathbb{C}} \setminus \mathrm{post}(f)$. Note that if $w \in \widehat{\mathbb{C}} \setminus \mathrm{post}(f)$ and $z \in f^{-1}(w)$, then $z \in \widehat{\mathbb{C}} \setminus \mathrm{crit}(f)$. Hence $J_f = (f^\#)^2$ is a non-zero continuous function near such a point z. In (19.8) we saw that with the given notation we have the local representation

$$(19.10) \qquad \mathcal{R}(\rho)(w) = \sum_{j=1}^{d} \rho(g_j(w)) \cdot J_f(g_j(w))^{-1}.$$

It follows that $\mathcal{R}(\rho)(w)$ is a non-negative continuous function on $\widehat{\mathbb{C}} \setminus \mathrm{post}(f)$. In particular, $f_*\mu = \mu$ if and only if $\mathcal{R}(\rho) = \rho$ on $\widehat{\mathbb{C}} \setminus \mathrm{post}(f)$.

The idea for the construction of our desired measure λ is to turn this consideration around. In the following, we use the notation

$$\widehat{\mathbb{C}}_P = \widehat{\mathbb{C}} \setminus \mathrm{post}(f)$$

for the Riemann sphere "punctured" at the points in $\mathrm{post}(f)$, and denote by $C(\widehat{\mathbb{C}}_P)$ the space of continuous functions on $\widehat{\mathbb{C}}_P$. We introduce an operator, the *Ruelle* or *transfer operator* $\mathcal{R}$ that maps each function

$\rho \in C(\widehat{\mathbb{C}}_P)$ to the function $\mathcal{R}(\rho)$ as defined in (19.9). Finding our measure λ then amounts to finding a suitable fixed point of this operator. This is a more tractable problem.

We will only use some very basic properties of the Ruelle operator summarized in the next lemma. For a thorough treatment in a more general setting see for example [**PU10**, Chapter 5].

LEMMA 19.8. *The Ruelle operator has the following properties.*

(i) *If $\rho \in C(\widehat{\mathbb{C}}_P)$, then $\mathcal{R}(\rho) \in C(\widehat{\mathbb{C}}_P)$. If, in addition, $\rho > 0$ on $\widehat{\mathbb{C}}_P$, then $\mathcal{R}(\rho) > 0$ on $\widehat{\mathbb{C}}_P$.*

(ii) *If $\rho \in C(\widehat{\mathbb{C}}_P) \cap L^1(\widehat{\mathbb{C}})$, then $\mathcal{R}(\rho) \in C(\widehat{\mathbb{C}}_P) \cap L^1(\widehat{\mathbb{C}})$ and*

$$\int_{\widehat{\mathbb{C}}} \rho \, d\mathcal{L} = \int_{\widehat{\mathbb{C}}} \mathcal{R}(\rho) \, d\mathcal{L}.$$

(iii) *The operator $\mathcal{R} \colon C(\widehat{\mathbb{C}}_P) \to C(\widehat{\mathbb{C}}_P)$ is linear, and continuous in the following sense: if ρ and ρ_n for $n \in \mathbb{N}$ are functions in $C(\widehat{\mathbb{C}}_P)$ such that $\rho_n \to \rho$ locally uniformly on $\widehat{\mathbb{C}}_P$, then $\mathcal{R}(\rho_n) \to \mathcal{R}(\rho)$ locally uniformly on $\widehat{\mathbb{C}}_P$.*

(iv) *If μ is a Borel measure on $\widehat{\mathbb{C}}$ of the form $d\mu = \rho \, d\mathcal{L}$ with $\rho \in C(\widehat{\mathbb{C}}_P) \cap L^1(\widehat{\mathbb{C}})$, then $d(f_*\mu) = \mathcal{R}(\rho) \, d\mathcal{L}$. In particular, μ is f-invariant if and only if $\mathcal{R}(\rho) = \rho$.*

Here $L^1(\widehat{\mathbb{C}})$ denotes the space of functions that are almost everywhere defined on $\widehat{\mathbb{C}}$ and are integrable with respect to $\mathcal{L}$. Note that for questions of integrability it is irrelevant whether a function is only defined on $\widehat{\mathbb{C}}_P$ or the whole Riemann sphere.

PROOF. (i) If $\rho \in C(\widehat{\mathbb{C}}_P)$, then $\mathcal{R}(\rho) \in C(\widehat{\mathbb{C}}_P)$ as follows from the local representation of the function $\mathcal{R}(\rho)$ in (19.10). Note that here it is important that if $w \in \widehat{\mathbb{C}}_P$, then near each point $z \in f^{-1}(w)$, the function $J_f = (f^\#)^2$ does not vanish and is continuous. If ρ is positive, then it is clear that $\mathcal{R}(\rho)$ is also positive.

(ii)+(iv) Let $\rho \in C(\widehat{\mathbb{C}}_P) \cap L^1(\widehat{\mathbb{C}})$. If we consider the signed measure μ on $\widehat{\mathbb{C}}$ with $d\mu = \rho \, d\mathcal{L}$, then the same computation as in (19.8) shows that the measure $f_*\mu$ is absolutely continuous with respect to $\mathcal{L}$ and $d(f_*\mu) = \mathcal{R}(\rho) \, d\mathcal{L}$. In particular,

$$\int_{\widehat{\mathbb{C}}} \mathcal{R}(\rho) \, d\mathcal{L} = (f_*\mu)(\widehat{\mathbb{C}}) = \mu(f^{-1}(\widehat{\mathbb{C}})) = \mu(\widehat{\mathbb{C}}) = \int_{\widehat{\mathbb{C}}} \rho \, d\mathcal{L}.$$

Thus $\mathcal{R}(\rho) \in C(\widehat{\mathbb{C}}_P) \cap L^1(\widehat{\mathbb{C}})$. Statements (ii) and (iv) follow immediately from these considerations.

(iii) By what we have seen in (i), we can consider the Ruelle operator as a map $\mathcal{R}\colon C(\widehat{\mathbb{C}}_P) \to C(\widehat{\mathbb{C}}_P)$. It is clear that $\mathcal{R}$ is a linear map. If we equip $C(\widehat{\mathbb{C}}_P)$ with the topology of locally uniform convergence, then the continuity of $\mathcal{R}$ immediately follows from the local representation formula (19.10). Note that the inverse branches g_j of f map a compact neighborhood $K \subset \widehat{\mathbb{C}}_P$ of the point $p \in \widehat{\mathbb{C}}_P$ near which these branches were defined again to compact subsets of $\widehat{\mathbb{C}}_P$. □

Our goal now is to construct a suitable fixed point of $\mathcal{R}$. This is based on an iteration and averaging procedure. In order to be able to pass to limits, we want to apply the Arzelà-Ascoli theorem for a certain subfamily of $C(\widehat{\mathbb{C}}_P)$. For this we will introduce a control function M that will allow us to establish the desired equicontinuity and uniform boundedness properties.

For the definition of M we consider the functions $\rho_n = \mathcal{R}^n(1)$ for $n \in \mathbb{N}_0$. Here 1 represents the function with constant value 1 on $\widehat{\mathbb{C}}_P$, and $\mathcal{R}^n$ is the n-iterate of $\mathcal{R}$ considered as an operator on $C(\widehat{\mathbb{C}}_P)$. We use the convention that $\mathcal{R}^0$ is the identity map on $C(\widehat{\mathbb{C}}_P)$ and so $\rho_0 \equiv 1$.

The definition of the Ruelle operator and the chain rule for Jacobians imply that

$$\rho_n(z) = \sum_{z' \in f^{-n}(z)} J_{f^n}(z')^{-1}$$

for $n \in \mathbb{N}_0$ and $z \in \widehat{\mathbb{C}}_P$. It follows from Lemma 19.8 (i) and (ii) that for $n \in \mathbb{N}_0$ the function ρ_n is a positive continuous function on $\widehat{\mathbb{C}}_P$ that is normalized such that

$$(19.11) \qquad \int_{\widehat{\mathbb{C}}} \rho_n \, d\mathcal{L} = \int_{\widehat{\mathbb{C}}} \mathcal{R}^n(1) \, d\mathcal{L} = \int_{\widehat{\mathbb{C}}} 1 \, d\mathcal{L} = 1.$$

For $z, w \in \widehat{\mathbb{C}}_P$ we now define

$$(19.12) \qquad M(z, w) = \sup_{n \in \mathbb{N}_0} \frac{\rho_n(z)}{\rho_n(w)}.$$

A priori it is not clear that M is everywhere finite, but we will verify this momentarily. It immediately follows from the definition of M that

$$(19.13) \qquad \rho_n(z) \leq M(z, w)\rho_n(w)$$

for all $n \in \mathbb{N}_0$ and $z, w \in \widehat{\mathbb{C}}_P$. This combined with the properties of M as formulated in the next lemma will give us the desired control for the behavior of the functions ρ_n (see the proof of Lemma 19.10).

LEMMA 19.9. *Let M be defined as in (19.12). Then the following statements are true:*

(i) *The function M is continuous on $\widehat{\mathbb{C}}_P \times \widehat{\mathbb{C}}_P$ and takes values in $[1, \infty)$.*

(ii) *For each $w_0 \in \widehat{\mathbb{C}}_P$ the map $z \mapsto M(z, w_0)$ is integrable (with respect to $\mathcal{L}$).*

PROOF. (i) It is clear that $M(z, w) \geq \rho_0(z)/\rho_0(w) = 1$ and so $M(z, w) \in [1, \infty]$ for $z, w \in \widehat{\mathbb{C}}_P$. We also have $M(z, z) = 1$ for $z \in \widehat{\mathbb{C}}_P$.

Let $w \in \widehat{\mathbb{C}}_P$ be arbitrary. We first show that $M(z, w) \to 1$ as $z \to w$. To see this, we fix $\delta > 0$ such that $B' := B(w, \delta) \subset \widehat{\mathbb{C}}_P$. Then for each $n \in \mathbb{N}_0$ the map f^n is a covering map over B'. In particular, f^n has d^n inverse branches $g_1, \ldots, g_{d^n}$, where $d = \deg(f)$. Each map g_j is a conformal map of B' onto its image. As in the proof of Theorem 19.1, one sees that the diameters of these images tend to 0 as $n \to \infty$. So with at most finitely exceptions these inverse branches of iterates f^n, $n \in \mathbb{N}_0$, send B' to a set contained in a hemisphere. So we can apply Koebe's distortion theorem (see (A.8) in Theorem A.1 and the discussion after the proof of Theorem A.1) and conclude that for $z \in B := B(w, \frac{1}{2}\delta)$, we have

$$(19.14) \qquad \frac{J_{g_j}(z)}{J_{g_j}(w)} = \frac{J_{f^n}(g_j(z))^{-1}}{J_{f^n}(g_j(w))^{-1}} \asymp 1,$$

where $C(\asymp)$ is independent of n and the choice of the inverse branch g_j. In addition, $C(\asymp) \to 1$ as $z \to w$. Based on (19.7), it follows that

$$\rho_n(z) = \sum_{z' \in f^{-n}(z)} J_{f^n}(z')^{-1} = \sum_{j=1}^{d^n} J_{f^n}(g_j(z))^{-1} = \sum_{j=1}^{d^n} J_{g_j}(z)$$

$$\leq C(z) \sum_{j=1}^{d^n} J_{g_j}(w) = C(z)\rho_n(w),$$

where $C(z) \to 1$ as $z \to w$. In particular,

$$1 \leq M(z, w) \leq C(z)$$

and so $M(z, w) \to 1$ as $z \to w$. A similar estimate based on (19.14) also shows that $M(w, z) \to 1$ as $z \to w$.

This shows that for each point $w \in \widehat{\mathbb{C}}_P$ there exists a neighborhood B of w such that $M(z, w)$ is finite for $z \in B$. From this and the (obvious) inequality

$$(19.15) \qquad M(u, w) \leq M(u, v)M(v, w)$$

for $u, v, w \in \widehat{\mathbb{C}}_P$ in combination with a chaining argument, the finiteness of $M(z, w)$ follows for all $z, w \in \widehat{\mathbb{C}}_P$.

Inequality (19.15) implies that

$$\frac{1}{M(z_0, z)M(w, w_0)} \leq \frac{M(z, w)}{M(z_0, w_0)} \leq M(z, z_0)M(w_0, w)$$

for $z, z_0, w, w_0 \in \widehat{\mathbb{C}}_P$. By what we have seen, the first and the third expression in this inequality approach 1 as $z \to z_0$ and $w \to w_0$. The continuity of M easily follows.

(ii) By the first part of the proof we know that for fixed $w_0 \in \widehat{\mathbb{C}}_P$ the function $z \to M(z, w_0)$ is continuous on $\widehat{\mathbb{C}}_P$. In order to show its integrability, we need an integrable upper bound for this function near each of the singularities in $\mathrm{post}(f)$. The following claim provides such a bound.

Claim. If $w_0 \in \widehat{\mathbb{C}}_P$ and $p \in \mathrm{post}(f)$, then there exists $\alpha \in (0, 2)$ such that

$$M(z, w_0) \lesssim |z - p|^{-\alpha} \text{ for } z \text{ near } p.$$

Once we know that the claim is true, the integrability of $z \to M(z, w_0)$ follows, because for $\alpha < 2$ the function $z \mapsto |z - p|^{-\alpha}$ is locally integrable near p.

One can say more about the exponent α here. Indeed, for $p \in \widehat{\mathbb{C}}$ we define

(19.16) $\beta_f(p) = \max\{\deg(f^n, q) : n \in \mathbb{N} \text{ and } f^n(q) = p\}.$

Since f has no periodic critical points, the local degrees $\deg(f^n, q)$ are uniformly bounded by Lemma 18.6; in particular, there exists $N \in \mathbb{N}$ such that $\beta_f(p) \leq N$ for all $p \in \widehat{\mathbb{C}}$. Moreover, it is clear that $\beta_f(p) = 1$ for $p \in \widehat{\mathbb{C}}_P = \widehat{\mathbb{C}} \setminus \mathrm{post}(f)$ and $\beta_f(p) > 1$ for $p \in \mathrm{post}(f)$. If $p \in \mathrm{post}(f)$, then, as we will see, the above claim is true with the exponent $\alpha = 2 - 2/\beta_f(p)$.

To prove the claim, we fix $w_0 \in \widehat{\mathbb{C}}_P$ and $p \in \mathrm{post}(f)$. We pick a Jordan curve $\mathcal{C} \subset \widehat{\mathbb{C}}$ with $\mathrm{post}(f) \subset \mathcal{C}$, and consider cells for $(f, \mathcal{C})$.

Let $U = W^0(p)$ be the 0-flower of p. This is a simply connected region whose complement contains more than two points (this complement contains the 0-vertices distinct from p, of which there are $\# \mathrm{post}(f) - 1 \geq 2$). So there exists a conformal map $\psi \colon \mathbb{D} \to U$ such that $\psi(0) = p$. Let $n \in \mathbb{N}_0$ and consider a component V of the preimage $f^{-n}(U)$. By Lemma 5.29 (ii) we know that V is an n-flower and so there exists an n-vertex q such that $V_q := W^n(q) = V$. Then necessarily $f^n(q) = p$, and we have

$$\bigcup_{q \in f^{-n}(p)} V_q = f^{-n}(U).$$

For such a component $V = V_q$ there also exists a conformal map $\varphi\colon \mathbb{D} \to V_q$ with $\varphi(0) = q$. Since V_q is a component of $f^{-n}(U)$, the map $f^n|V_q\colon V_q \to U$ is proper (see Lemma A.8 (ii)). It follows that the map $h := \psi^{-1} \circ (f^n|V_q) \circ \varphi\colon \mathbb{D} \to \mathbb{D}$ is also proper, and hence a finite Blaschke product (see [**Bu79**, Exercise 6.12]). Now $(f^n|V_q)^{-1}(p) = \{q\}$ which implies that $h^{-1}(0) = \{0\}$. It follows that $h(z) = cz^k$ with $c \in \mathbb{C}$, $|c| = 1$, and $k \in \mathbb{N}$. By replacing the original map φ by a precomposition with a rotation if necessary, we may assume that $c = 1$. We then obtain the following commutative diagram

$$(19.17) \qquad\qquad \begin{array}{ccc} V_q & \xrightarrow{\ f^n\ } & U \\ {\scriptstyle\varphi}\big\uparrow & & \big\uparrow{\scriptstyle\psi} \\ \mathbb{D} & \xrightarrow{\ h(u)=u^k\ } & \mathbb{D}. \end{array}$$

Here we have the uniform bound $k \le \beta_f(p) \le N$, where $\beta_f(p) < \infty$ is defined as in (19.16).

In the commutative diagram (19.17) the map ψ is fixed for the given point p under consideration. On the other hand, the map φ (and also k) depend on n and the chosen preimage $q \in f^{-n}(p)$. In the ensuing argument it will be crucial that we obtain uniform estimates for all such maps φ (as well as for the points under consideration).

We now fix a compact subset K of U, say $K = \psi(\overline{B}_{\mathbb{C}}(0, 1/2)) \subset U$, and use $w_1 = \psi(1/2) \in K$ as a base point in K. Note that K is a compact neighborhood of p. We want to estimate $M(z, w_1)$ for $z \in K \setminus \{p\}$. For this we take preimages of $z \in K$ under maps as in the diagram (19.17) and compare the values of Jacobians at such preimages with corresponding values at preimages of the basepoint w_1. Here it will be important that the preimage of K under a map $f^n \circ \varphi = \psi \circ h$ as in (19.17) lies in the fixed compact subset

$$A := \overline{B}_{\mathbb{C}}(0, 2^{-1/N}) \supset \overline{B}_{\mathbb{C}}(0, 2^{-1/k}) = h^{-1}(\overline{B}_{\mathbb{C}}(0, 1/2)) = h^{-1}(\psi^{-1}(K))$$

of $\mathbb{D}$.

So let $z \in K$ be arbitrary and suppose $w', z' \in V_q$ are points such that $f^n(z') = z$ and $f^n(w') = w_1$. Let $u' = \varphi^{-1}(z')$ and $v' = \varphi^{-1}(w')$. Finally, define $u = h(u') = (u')^k = \psi^{-1}(z)$ and $v = h(v') = (v')^k = \psi^{-1}(w_1)$. Then u' and v' lie in the fixed compact subset A of $\mathbb{D}$. The situation is represented by the following commutative diagram:

$$\begin{array}{ccc} z', w' \in V_q & \xrightarrow{\ f^n\ } & z, w_1 \in K \subset U \\ {\scriptstyle\varphi}\big\uparrow & & \big\uparrow{\scriptstyle\psi} \\ u', v' \in A \subset \mathbb{D} & \xrightarrow{\ h\ } & u, v \in \overline{B}_{\mathbb{C}}(0, 1/2) \subset \mathbb{D}. \end{array}$$

It follows from Koebe's distortion theorem (as formulated in (A.8) of Theorem A.1) that

$$(19.18) \qquad\qquad J_\varphi(u') \asymp J_\varphi(v')$$

with $C(\asymp)$ independent of u', v', and φ. Note that the flower $V_q = \varphi(\mathbb{D})$ is not necessarily contained in a hemisphere and so the assumption in Theorem A.1 may not always be true; it is true with possibly finitely many exceptions $n \in \mathbb{N}_0$ and $q \in f^{-n}(p)$. Then (19.18) still holds with a uniform constant if we adjust the constant to account for the finitely many exceptional cases (see the discussion after the proof of Theorem A.1).

Since $u, v \in \psi^{-1}(K) = \overline{B}_\mathbb{C}(0, 1/2)$ we also have

$$J_\psi(u) \asymp J_\psi(v)$$

with $C(\asymp)$ independent of u and v. In addition,

$$|z - p| \asymp |u|$$

with $C(\asymp)$ independent of z. Finally, note that

$$J_h(u') \asymp |u'|^{2k-2}$$

and

$$J_h(v') \asymp 1$$

with $C(\asymp)$ independent of the choices, because k is uniformly bounded by $\beta_f(p)$.

Putting this all together, we arrive at

$$\frac{J_{f^n}(w')}{J_{f^n}(z')} = \frac{J_\psi(v) J_h(v') J_\varphi(v')^{-1}}{J_\psi(u) J_h(u') J_\varphi(u')^{-1}}$$

$$\asymp \frac{J_h(v')}{J_h(u')} \asymp |u'|^{-2k+2} = |u|^{-2+2/k}$$

$$\le |u|^{-2+2/\beta_f(p)} \asymp |z - p|^{-2+2/\beta_f(p)}.$$

It follows that

$$\rho_n(z) = \sum_{z' \in f^{-n}(z)} J_{f^n}(z')^{-1} = \sum_{q \in f^{-n}(p)} \sum_{z' \in V_q \cap f^{-n}(z)} J_{f^n}(z')^{-1}$$

$$\lesssim |z - p|^{-2+2/\beta_f(p)} \sum_{q \in f^{-n}(p)} \sum_{w' \in V_q \cap f^{-n}(w_1)} J_{f^n}(w')^{-1}$$

$$= |z - p|^{-\alpha} \rho_n(w_1),$$

where $\alpha = 2 - 2/\beta_f(p) \in (0, 2)$. Here we used the previous estimate and also the fact that w_1 and z have the same number of preimages

in each flower V_q, namely $k = \deg(f^n, q)$ preimages. Since the implicit constants here are independent of $z \in K$ and $n \in \mathbb{N}_0$, we conclude that

$$M(z, w_1) = \sup_{n \in \mathbb{N}_0} \frac{\rho_n(z)}{\rho_n(w_1)} \lesssim |z - p|^{-\alpha}$$

for z near p. Hence

$$M(z, w_0) \le M(z, w_1) M(w_1, w_0) \lesssim M(z, w_1) \lesssim |z - p|^{-\alpha}$$

for z near p. The claim and statement (ii) follow. $\square$

LEMMA 19.10 (Fixed point of $\mathcal{R}$). *There exists a positive function* $\rho \in C(\widehat{\mathbb{C}}_P)$ *with* $\int_{\widehat{\mathbb{C}}} \rho \, d\mathcal{L} = 1$ *such that* $\mathcal{R}(\rho) = \rho$.

A fixed point ρ of $\mathcal{R}$ with these properties is actually unique. We will not show this directly, but it will immediately follow from the uniqueness of the measure λ_f in Theorem 19.2.

PROOF. We define

$$\widetilde{\rho}_n = \frac{1}{n} \sum_{i=0}^{n-1} \rho_i = \frac{1}{n} \sum_{i=0}^{n-1} \mathcal{R}^i(1)$$

for $n \in \mathbb{N}$. Then each $\widetilde{\rho}_n$ is a positive function in $C(\widehat{\mathbb{C}}_P)$.

Claim 1. The functions $\widetilde{\rho}_n$ are locally uniformly bounded on $\widehat{\mathbb{C}}_P$ for $n \in \mathbb{N}$.

To prove this claim, it suffices to show that the functions $\rho_n = \mathcal{R}^n(1)$, $n \in \mathbb{N}_0$, are locally uniformly bounded on $\widehat{\mathbb{C}}_P$. To see this, pick a point $w_0 \in \widehat{\mathbb{C}}_P$ and a small $\delta > 0$ such that the disk $B := \overline{B}(w_0, \delta)$ is contained in $\widehat{\mathbb{C}}_P$. We will first produce an upper bound for $\rho_n(w_0)$.

Let M be the function defined in (19.12). Lemma 19.9 (i) implies that the function $z \mapsto M(w_0, z)$ is uniformly bounded from above for $z \in B$, say by $C_0 \ge 1$. Then for each $n \in \mathbb{N}_0$ we have

$$\rho_n(w_0) \le \frac{1}{\mathcal{L}(B)} \int_B M(w_0, z) \rho_n(z) \, d\mathcal{L}(z)$$

$$\le \frac{C_0}{\mathcal{L}(B)} \int_{\widehat{\mathbb{C}}} \rho_n \, d\mathcal{L} = \frac{C_0}{\mathcal{L}(B)} =: C_1.$$

Here we used (19.13) and the normalization (19.11). Hence

$$\rho_n(z) \le M(z, w_0) \rho_n(w_0) \le C_1 M(z, w_0)$$

for $n \in \mathbb{N}_0$ and $z \in \widehat{\mathbb{C}}_P$.

Since M is bounded on compact subsets of $\widehat{\mathbb{C}}_P \times \widehat{\mathbb{C}}_P$ as follows from Lemma 19.9 (i), this last inequality implies that the functions ρ_n, $n \in \mathbb{N}_0$, are locally uniformly bounded on $\widehat{\mathbb{C}}_P$. Claim 1 follows.

Claim 2. At each point $z_0 \in \widehat{\mathbb{C}}_P$ the functions $\widetilde{\rho}_n$, $n \in \mathbb{N}_0$, are equicontinuous.

First note that (19.13) implies that

(19.19) $$\widetilde{\rho}_n(z) \leq M(z, w)\widetilde{\rho}_n(w)$$

for all $n \in \mathbb{N}$ and $z, w \in \widehat{\mathbb{C}}_P$.

Let $z_0 \in \widehat{\mathbb{C}}_P$ be arbitrary. Then it follows from (19.19) that for $n \in \mathbb{N}$ and $z \in \widehat{\mathbb{C}}_P$ we have

$$\left| \frac{\widetilde{\rho}_n(z)}{\widetilde{\rho}_n(z_0)} - 1 \right| \leq \max\{M(z, z_0) - 1, |M(z_0, z)^{-1} - 1|\}.$$

By Claim 1 we know that there exists a constant $C > 0$ such that $\widetilde{\rho}_n(z_0) \leq C$ for $n \in \mathbb{N}$. It follows that for each $z \in \widehat{\mathbb{C}}_P$ and $n \in \mathbb{N}$ we have

$$|\widetilde{\rho}_n(z) - \widetilde{\rho}_n(z_0)| = |\widetilde{\rho}_n(z_0)| \cdot \left| \frac{\widetilde{\rho}_n(z)}{\widetilde{\rho}_n(z_0)} - 1 \right|$$
$$\leq C \max\{M(z, z_0) - 1, |M(z_0, z)^{-1} - 1|\}.$$

Since $M(z, z_0), M(z_0, z) \to M(z_0, z_0) = 1$ as $z \to z_0$, Claim 2 follows.

By what we have seen, the functions $\widetilde{\rho}_n$, $n \in \mathbb{N}$, form a locally uniformly bounded family of continuous functions on $\widehat{\mathbb{C}}_P$ that is equicontinuous at each point in $\widehat{\mathbb{C}}_P$. So the Arzelà-Ascoli theorem implies that there exists a subsequence $\{\widetilde{\rho}_{n_k}\}$ that converges locally uniformly on $\widehat{\mathbb{C}}_P$ to a non-negative function $\rho \in C(\widehat{\mathbb{C}}_P)$.

Claim 3. $\mathcal{R}(\rho) = \rho$.

To see this, let $z \in \widehat{\mathbb{C}}_P$. As we have shown in the proof of Claim 1, the functions ρ_n are uniformly bounded at z, say by the constant $C > 0$. So for $n \in \mathbb{N}$ we have

$$|\mathcal{R}(\widetilde{\rho}_n)(z) - \widetilde{\rho}_n(z)| = \frac{1}{n}|\mathcal{R}^n(1)(z) - \mathcal{R}^0(1)(z)|$$
$$= \frac{1}{n}|\rho_n(z) - \rho_0(z)| \leq \frac{2C}{n} \to 0 \text{ as } n \to \infty.$$

Hence the continuity of $\mathcal{R}$ (see Lemma 19.8 (iii)) implies that

$$\mathcal{R}(\rho)(z) = \lim_{k \to \infty} \mathcal{R}(\widetilde{\rho}_{n_k})(z) = \lim_{k \to \infty} \widetilde{\rho}_{n_k}(z) = \rho(z).$$

Claim 3 follows.

The proof will be complete if we establish the last claim.

Claim 4. The function ρ is positive and satisfies $\int_{\widehat{\mathbb{C}}} \rho \, d\mathcal{L} = 1$.

By the normalization (19.11) and the definition of $\widetilde{\rho}_n$ we have

$$\int_{\widehat{\mathbb{C}}} \widetilde{\rho}_n \, d\mathcal{L} = 1$$

for $n \in \mathbb{N}$. Pick a point $w_0 \in \widehat{\mathbb{C}}_P$. Then by Claim 1 there exists $C > 0$ such that $\widetilde{\rho}_n(w_0) \leq C$ for $n \in \mathbb{N}$, and so by (19.19) we have

$$0 \leq \widetilde{\rho}_n(z) \leq CM(z, w_0)$$

for $n \in \mathbb{N}$ and $z \in \widehat{\mathbb{C}}_P$. Hence by Lemma 19.9 (ii) the functions $\widetilde{\rho}_n$, $n \in \mathbb{N}$, are majorized by an integrable function and so by Lebesgue's dominated convergence theorem we conclude that

$$\int_{\widehat{\mathbb{C}}} \rho \, d\mathcal{L} = \lim_{k \to \infty} \int_{\widehat{\mathbb{C}}} \widetilde{\rho}_{n_k} \, d\mathcal{L} = 1.$$

We know that ρ is non-negative on $\widehat{\mathbb{C}}_P$. To see that ρ is actually positive, we argue by contradiction and assume that there exists a point $w \in \widehat{\mathbb{C}}_P$ such that $\rho(w) = 0$. In inequality (19.19) we can pass to sublimits and conclude that then

$$0 \leq \rho(z) \leq M(z, w)\rho(w) = 0$$

for all $z \in \widehat{\mathbb{C}}_P$, and so $\rho \equiv 0$. This is impossible, since $\int_{\widehat{\mathbb{C}}} \rho \, d\mathcal{L} = 1$. $\square$

We are now ready to prove the existence and uniqueness of an f-invariant probability measure that is absolutely continuous with respect to $\mathcal{L}$.

PROOF OF THEOREM 19.2. Let ρ be a fixed point of the Ruelle operator as provided by Lemma 19.10 and let $\lambda_f = \lambda$ be the measure with $d\lambda = \rho \, d\mathcal{L}$. Then λ is a probability measure that is absolutely continuous with respect to $\mathcal{L}$. Actually, since ρ is a positive continuous function on $\widehat{\mathbb{C}}$ outside the finite set $\mathrm{post}(f)$, the measures λ and $\mathcal{L}$ are mutually absolutely continuous. By Lemma 19.8 (iv) the measure λ is f-invariant. The existence of a measure with the desired properties follows.

By Corollary 19.7 each f-invariant measure that is absolutely continuous with respect to $\mathcal{L}_{\widehat{\mathbb{C}}}$ is ergodic for f. If μ is another f-invariant probability measure that is absolutely continuous with respect to $\mathcal{L}$, then it is also absolutely continuous with respect to λ. On the other hand, both measures are ergodic and so necessarily $\mu = \lambda$. The uniqueness of λ follows. $\square$

We now take a closer look at the unique measure λ_f in the case where our rational expanding Thurston map has a parabolic orbifold

$\mathcal{O}_f = (\widehat{\mathbb{C}}, \alpha_f)$. Since f has no periodic critical points, this is the case precisely when f is a Lattès map (see Theorem 3.1 (i)).

By Theorem 3.1 (ii) we know that there exists a map $A\colon \mathbb{C} \to \mathbb{C}$ of the form $A(u) = \alpha u + \beta$ with $\alpha, \beta \in \mathbb{C}$, $\alpha \neq 0$, such that $f \circ \Theta = \Theta \circ A$, where $\Theta\colon \mathbb{C} \to \widehat{\mathbb{C}}$ is the universal orbifold covering map of $\mathcal{O}_f$. Moreover, $d := \deg(f) = |\alpha|^2$ (see Lemma 3.16).

We first want to construct a measure Ω on $\widehat{\mathbb{C}}$ so that f has constant Jacobian with respect to Ω. Since the lift A of f under Θ has constant Jacobian $J_{A,\mathcal{L}_{\mathbb{C}}} = |\alpha|^2 = d$ with respect to Lebesgue measure $\mathcal{L}_{\mathbb{C}}$ on $\mathbb{C}$, we want to (locally) push-forward $\mathcal{L}_{\mathbb{C}}$ by Θ to define Ω. The measure obtained in this way is actually the canonical orbifold measure $\Omega = \Omega_f$ of $\mathcal{O}_f$

We will quickly review the definition of Ω, but refer to Section A.10 for the details. Let J_Θ be the Jacobian of Θ with $\mathcal{L}_{\mathbb{C}}$ being the underlying measure on the source $\mathbb{C}$ and $\mathcal{L}_{\widehat{\mathbb{C}}}$ being the measure on the target space $\widehat{\mathbb{C}}$ of Θ. Then

$$J_\Theta(u) = \frac{|\Theta'(u)|^2}{\pi(1 + |\Theta(u)|^2)^2}$$

for $u \in \mathbb{C}$. Let $\kappa\colon \widehat{\mathbb{C}} \to (0, \infty)$ be the function defined as

$$\kappa(w) = J_\Theta(u)^{-1}$$

for $w \in \widehat{\mathbb{C}}_P$, where $u \in \Theta^{-1}(w)$. One can show that this function is well-defined, positive and continuous on $\widehat{\mathbb{C}}_P$, and integrable with respect to $\mathcal{L}_{\widehat{\mathbb{C}}}$. The map Θ is only unique up to a precomposition by a conformal automorphism of $\mathbb{C}$. Choosing this automorphism appropriately, we may assume that $\int_{\widehat{\mathbb{C}}} \kappa \, d\mathcal{L}_{\widehat{\mathbb{C}}} = 1$. Now let Ω be the unique measure on $\widehat{\mathbb{C}}$ with Radon-Nikodym derivative κ with respect to $\mathcal{L}_{\widehat{\mathbb{C}}}$, meaning that $d\Omega = \kappa \, d\mathcal{L}_{\widehat{\mathbb{C}}}$. This measure is normalized so that $\Omega(\widehat{\mathbb{C}}) = 1$.

By using the identity $f \circ \Theta = \Theta \circ A$ and the chain rule for Jacobians (where $\mathcal{L}_{\widehat{\mathbb{C}}}$ is the measure on $\widehat{\mathbb{C}}$ and $\mathcal{L}_{\mathbb{C}}$ the measure on $\mathbb{C}$) we see that

$$(19.20) \qquad J_f(\Theta(u)) \cdot J_\Theta(u) = |\alpha|^2 \cdot J_\Theta(A(u)) = d \cdot J_\Theta(A(u))$$

for $u \in \mathbb{C}$. Here $J_f = (f^\sharp)^2$ (see (19.3)).

Now suppose $w \in \widehat{\mathbb{C}}_P$ and $z \in f^{-1}(w)$. If we pick a point $u \in \mathbb{C}$ such that $\Theta(u) = z$, then for $v := A(u)$ we have

$$\Theta(v) = (\Theta \circ A)(u) = (f \circ \Theta)(u) = f(z) = w.$$

It follows that

$$\kappa(z) = J_\Theta(u)^{-1} \text{ and } \kappa(w) = J_\Theta(v)^{-1},$$

and so by taking reciprocals in (19.20) we obtain

$$(19.21) \qquad \kappa(z) \cdot J_f(z)^{-1} = \frac{1}{d}\kappa(w)$$

whenever $w \in \widehat{\mathbb{C}}_P$ and $z \in f^{-1}(w)$.

From this and (19.2) we conclude that if we change the underlying measure on $\widehat{\mathbb{C}}$ from $\mathcal{L}_{\widehat{\mathbb{C}}}$ to $\Omega = \Omega_f$, then the Jacobian of f is given by

$$(19.22) \qquad J_{f,\Omega}(z) = \frac{\kappa(f(z))}{\kappa(z)} J_f(z) = d = \deg(f)$$

for $z \in \widehat{\mathbb{C}} \setminus f^{-1}(\text{post}(f))$. So $J_{f,\Omega}(z) = d$ for Ω-almost every $z \in \widehat{\mathbb{C}}$ and $J_{f,\Omega}$ is indeed constant.

Since each point $w \in \widehat{\mathbb{C}}_P$ has precisely d preimages under f, it also follows from (19.21) that

$$\mathcal{R}(\kappa)(w) = \sum_{z \in f^{-1}(w)} \kappa(z)J_f(z)^{-1} = \kappa(w).$$

This shows that κ is a fixed point of the Ruelle operator with properties as in Lemma 19.10. Since $\kappa = d\Omega/d\mathcal{L}$ it follows from the uniqueness part of Theorem 19.2 that $\Omega = \lambda_f$ whenever f is a Lattès map.

The proof of Theorem 19.3 is now easy.

PROOF OF THEOREM 19.3. Let $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a Lattès map. As we have seen in the previous discussion, then $\lambda_f = \Omega_f$, where λ_f is the f-invariant measure provided by Theorem 19.2 and $\Omega = \Omega_f$ is the normalized canonical orbifold measure of $\mathcal{O}_f$.

By Rokhlin's formula (19.5) and (19.22) the measure-theoretic entropy h_Ω of Ω is given by

$$h_\Omega = \int_{\widehat{\mathbb{C}}} \log(J_{f,\Omega})\, d\Omega = \log(\deg(f)) = h_{top}(f).$$

So Ω is a measure of maximal entropy. Since the measure of maximal entropy ν_f is uniquely determined for an expanding Thurston map (see Theorem 17.1), it follows that $\nu_f = \Omega_f = \lambda_f$ as desired. $\qquad\square$

19.4. Lattès maps, the measure of maximal entropy, and Lebesgue measure

In this section we prove Theorem 19.4, the special case of Zdunik's theorem. Let $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a rational expanding Thurston map and μ be an f-invariant probability measure on $\widehat{\mathbb{C}}$. We know from Chapter 17 that the (measure-theoretic) entropy h_μ of μ satisfies $h_\mu \leq \log(\deg(f))$ (see Corollary 17.2 and (17.7)). Moreover, equality is attained precisely

for $\mu = \nu_f$, the measure of maximal entropy of f as constructed in Proposition 17.11.

We first establish a characterization of Lattès maps that uses the measure provided by Theorem 19.2.

THEOREM 19.11. *Let $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a rational expanding Thurston map and $\lambda = \lambda_f$ be the unique f-invariant probability measure that is absolutely continuous with respect to $\mathcal{L}_{\widehat{\mathbb{C}}}$. Then for the entropy h_λ we have*

$$h_\lambda = \log(\deg(f)) \text{ if and only if } f \text{ is a Lattès map.}$$

In other words, λ_f is equal to the measure of maximal entropy ν_f if and only if f is a Lattès map.

We require the following lemma.

LEMMA 19.12. *Let $f\colon S^2 \to S^2$ be a Thurston map. Then f has a parabolic orbifold if and only if*

$$(19.23) \qquad\qquad \deg(f^n, q) = \deg(f^n, q'),$$

whenever $p \in \operatorname{post}(f)$, $n \in \mathbb{N}$, and $q, q' \in f^{-n}(p) \setminus \operatorname{post}(f)$.

PROOF. If f has a parabolic orbifold, then it follows from condition (iii) in Proposition 2.13 by induction that

$$\alpha_f(q) \deg(f^n, q) = \alpha_f(p)$$

whenever $p \in S^2$, $n \in \mathbb{N}$, and $q \in f^{-n}(p)$. If in addition $q \notin \operatorname{post}(f)$, then $\alpha_f(q) = 1$ and so

$$\deg(f^n, q) = \alpha_f(p).$$

The relation (19.23) follows.

To show the other implication, assume that (19.23) holds. In order to show that f has parabolic orbifold, we want to verify condition (iii) in Proposition 2.13. We first establish several claims.

Claim 1. Let $p \in S^2$. Then we have $\deg(f^n, q) = \deg(f^m, q')$, whenever $n, m \in \mathbb{N}$, $q \in f^{-n}(p) \setminus \operatorname{post}(f)$, and $q' \in f^{-m}(p) \setminus \operatorname{post}(f)$.

Indeed, suppose q and q' are as in this statement for a given point $p \in S^2$. If $p \in S^2 \setminus \operatorname{post}(f)$, then $\deg(f^n, q) = 1 = \deg(f^m, q')$ and the claim follows. So we may assume $p \in \operatorname{post}(f)$.

We then choose $k, l \in \mathbb{N}$ such that $n + k = m + l$, and pick points $u \in f^{-k}(q)$ and $u' \in f^{-l}(q')$. Then $\deg(f^k, u) = 1$ (otherwise $q \in \operatorname{post}(f)$), and so

$$\deg(f^{n+k}, u) = \deg(f^n, q) \deg(f^k, u) = \deg(f^n, q).$$

Similarly, $\deg(f^{m+l}, u') = \deg(f^m, q')$.

Now $p \in \text{post}(f)$ and $u, u' \in f^{-(n+k)}(p) \backslash \text{post}(f)$. So (19.23) implies that

$$\deg(f^n, q) = \deg(f^{n+k}, u) = \deg(f^{m+l}, u') = \deg(f^m, q').$$

Claim 1 follows.

Claim 2. If $p \in S^2$ and $\bigcup_{n \in \mathbb{N}} f^{-n}(p) \subset \text{post}(f)$, then p is contained in a critical cycle of f.

If $p \in S^2$ is as in this statement, then each preimage of p under f is in $\text{post}(f)$. Since $\text{post}(f)$ is f-invariant, it follows that $p \in \text{post}(f)$. Hence there exists $q \in \text{crit}(f)$ and $n \in \mathbb{N}$ such that $f^n(q) = p$. To prove the claim, it suffices to show that q is periodic, because then p is contained in the critical cycle generated by q.

To see that q is periodic, let $q_1 := q$ and inductively choose points $q_k \in S^2$ such that $f(q_{k+1}) = q_k$ for $k \in \mathbb{N}$. Each point q_k is a preimage of p under some iterate of f. Hence $q_k \in \text{post}(f)$. Since $\text{post}(f)$ is a finite set, not all the points q_k, $k \in \mathbb{N}$, can be distinct. So there exist $k, l \in \mathbb{N}$ such that $q_{k+l} = q_k = f^l(q_{k+l})$. We see that q_{k+l} is a periodic point of f. This implies that $q = q_1 = f^{k+l-1}(q_{k+l})$ is also a periodic point of f and Claim 2 follows.

Claim 3. The ramification function of f satisfies

(19.24) $$\alpha_f(p) = \deg(f^n, q)$$

whenever $p \in S^2$, $n \in \mathbb{N}$, and $q \in f^{-n}(p) \setminus \text{post}(f)$.

To see this, let $p \in S^2$ be arbitrary. We may assume that the set $\bigcup_{l \in \mathbb{N}} f^{-l}(p)$ is not contained in $\text{post}(f)$, because otherwise there is nothing to prove.

We know that $\alpha_f(p)$ is the least common multiple of all numbers $\deg(f^k, u)$, where $k \in \mathbb{N}$ and $u \in f^{-k}(p)$ (see Definition 2.7). Moreover, by Claim 1 the right hand side in (19.24) is independent of the choices of n and q. So in order to prove (19.24), it suffices to show that if $k \in \mathbb{N}$ and $u \in f^{-k}(p)$ are arbitrary, then there exists $n \in \mathbb{N}$ and $q \in f^{-n}(p) \setminus \text{post}(f)$ such that $\deg(f^n, q)$ is a multiple of $\deg(f^k, u)$.

If $\bigcup_{l \in \mathbb{N}} f^{-l}(u) \subset \text{post}(f)$, then by Claim 2 the point u belongs to a critical cycle of f. This cycle then contains also p. Therefore, p is a preimage of u under some iterate of f. This in turn gives

$$\bigcup_{l \in \mathbb{N}} f^{-l}(p) \subset \bigcup_{l \in \mathbb{N}} f^{-l}(u) \subset \text{post}(f),$$

which is a contradiction to our assumption on p.

So $\bigcup_{l \in \mathbb{N}} f^{-l}(u)$ is not contained in $\text{post}(f)$. Then we can find $l \in \mathbb{N}$ and a point $q \in f^{-l}(u)$ that is not a postcritical point of f. Setting

$n = k + l$, we have $q \in f^{-n}(p) \setminus \mathrm{post}(f)$. Moreover, $\deg(f^n, q) = \deg(f^k, u) \deg(f^l, q)$ is a multiple of $\deg(f^k, u)$. Claim 3 follows.

After these preparations we will now show that condition (iii) in Proposition 2.13 is true. So let $p, q \in \widehat{\mathbb{C}}$ with $f(q) = p$ be arbitrary. If $\bigcup_{n \in \mathbb{N}} f^{-n}(q) \subset \mathrm{post}(f)$, then by Claim 2 the point q, and hence also $p = f(q)$, belongs to a critical cycle of f. Then $\alpha_f(p) = \infty = \alpha_f(q) = \alpha_f(q) \cdot \deg(f, q)$ by Proposition 2.9 (ii).

If $\bigcup_{n \in \mathbb{N}} f^{-n}(q)$ is not contained in $\mathrm{post}(f)$, then we can find $n \in \mathbb{N}$ and a point $u \in f^{-n}(q)$ that is not a postcritical point. Then $f^n(u) = q$ and $f^{n+1}(u) = p$. Thus $\alpha_f(q) = \deg(f^n, u)$ and $\alpha_f(p) = \deg(f^{n+1}, u)$ by Claim 3. Therefore,

$$\alpha_f(p) = \deg(f^{n+1}, u) = \deg(f^n, u) \deg(f, q) = \alpha_f(q) \deg(f, q).$$

We see that condition (iii) in Proposition 2.13 is always satisfied. This shows that f has a parabolic orbifold. $\qquad\square$

PROOF OF THEOREM 19.11. Let $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a rational expanding Thurston map, $\lambda = \lambda_f$ be the unique measure on $\widehat{\mathbb{C}}$ given by Theorem 19.2, and h_λ be the entropy of λ.

By Rokhlin's formula (19.5) we have

$$h_\lambda = \int_{\widehat{\mathbb{C}}} \log(J_{f,\lambda}) \, d\lambda.$$

Combined with (19.4) and Jensen's inequality this gives

$$h_\lambda \leq \log\left(\int_{\widehat{\mathbb{C}}} J_{f,\lambda} \, d\lambda \right) = \log(\deg(f)),$$

where equality is achieved if and only if $J_\lambda \equiv \deg(f)$ λ-almost everywhere and hence $\mathcal{L}$-almost everywhere on $\widehat{\mathbb{C}}$.

If we assume that f is a Lattès map, then by Theorem 19.3 and by (19.22) we have $J_{f,\lambda} \equiv \deg(f)$ λ-almost everywhere. Thus $h_\lambda = \log(\deg(f))$ by Rokhlin's formula (19.5).

Conversely, assume that $h_\lambda = \log(\deg(f))$ and so $J_{f,\lambda} \equiv h := \log(\deg(f))$ λ-almost everywhere on $\widehat{\mathbb{C}}$. Let $\rho = d\lambda/d\mathcal{L}$ be the Radon-Nikodym derivative of λ with respect to Lebesgue measure $\mathcal{L}$. We know by Theorem 19.2 that this is a positive continuous function on $\widehat{\mathbb{C}} \setminus \mathrm{post}(f)$. Since the Jacobian is given by (see (19.2))

$$J_{f,\lambda}(z) = f^\sharp(z)^2 \frac{\rho(f(z))}{\rho(z)} \equiv h,$$

we conclude that

$$\rho(f(z)) = h \frac{\rho(z)}{f^\sharp(z)^2}$$

for $z \in \widehat{\mathbb{C}} \setminus f^{-1}(\text{post}(f))$.

If we iterate this relation and use the chain rule for the spherical derivative, we arrive at

$$(19.25) \qquad \rho(f^n(z)) = h^n \frac{\rho(z)}{(f^n)^\sharp(z)^2}$$

for $z \in \widehat{\mathbb{C}} \setminus f^{-n}(\text{post}(f))$.

We will use this relation to derive the asymptotic behavior of ρ near a point $p \in \text{post}(f)$ in order to verify the condition in Lemma 19.12. For this let $n \in \mathbb{N}$, and consider an arbitrary point $q \in f^{-n}(p) \setminus \text{post}(f)$.

Let $k = \deg(f^n, q)$. If $z \in \widehat{\mathbb{C}}$ is a point near q, then $w = f^n(z)$ is a point near p. By considering local power series expansions of f^n in holomorphic coordinates, we see that

$$|w - p| \asymp |z - q|^k$$

and

$$(f^n)^\sharp(z) \asymp |z - q|^{k-1} \asymp |w - p|^{1-1/k},$$

where the constants $C(\asymp)$ are independent of z near q (recall that we use "Polish notation" $|u - v|$ for the chordal distance between points $u, v \in \widehat{\mathbb{C}}$).

Now we know from the second part of Theorem 19.2 that $\rho(z) \asymp 1$ for z near q, because $q \notin \text{post}(f)$. So from the relation (19.25) we conclude that

$$\rho(w) \asymp \frac{\rho(z)}{(f^n)^\sharp(z)^2} \asymp |w - p|^{-2+2/k}$$

for all z near q, and hence for all w near p.

If q' is another point with $q' \in f^{-n}(p) \setminus \text{post}(f)$ and $k' := \deg(f^n, q')$, then the same argument shows that

$$\rho(w) \asymp |w - p|^{-2+2/k'}$$

for all w near p. These two estimates can only be valid if $k = k'$.

We conclude that $\deg(f^n, q) = \deg(f^n, q')$ whenever $p \in \text{post}(f)$, $n \in \mathbb{N}$, and $q, q' \in f^{-n}(p) \setminus \text{post}(f)$. So Lemma 19.12 implies that f has a parabolic orbifold. Since f is a rational expanding Thurston map, it has no periodic critical points (see Proposition 2.3). Hence f is a Lattès map by Theorem 3.1 (i). $\qquad\square$

Before we proceed to the proof of Theorem 19.4, we record a statement about the asymptotics of the Radon-Nikodym derivative $\rho = d\lambda_f/d\mathcal{L}_{\widehat{\mathbb{C}}}$ near its singularities that easily follows from our previous

considerations. For the formulation we use the function $\beta_f\colon \widehat{\mathbb{C}} \to \mathbb{N}$ given by

$$\beta_f(p) = \max\{\deg(f^n, q) : n \in \mathbb{N} \text{ and } f^n(q) = p\}$$

for $p \in \widehat{\mathbb{C}}$ (see (19.16)).

PROPOSITION 19.13. *Let* $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ *be a rational expanding Thurston map and* $\rho = d\lambda_f/d\mathcal{L}_{\widehat{\mathbb{C}}}$ *be Radon-Nikodym derivative of the measure given by Theorem 19.2. Then*

(19.26) $$\rho(w) \asymp |w - p|^{-2+2/\beta_f(p)}$$

for w *near* $p \in \mathrm{post}(f)$.

As we pointed out in the proof of Lemma 19.9 (ii), for a rational expanding Thurston map f the function β_f is bounded on $\widehat{\mathbb{C}}$ and we have $\beta_f(p) = 1$ for $p \in \widehat{\mathbb{C}} \setminus \mathrm{post}(f)$ and $\beta_f(p) \geq 2$ for $p \in \mathrm{post}(f)$. Accordingly, the asymptotics (19.26) also makes sense near points in $\widehat{\mathbb{C}} \setminus \mathrm{post}(f)$, where it should be interpreted as $\rho(w) \asymp 1$ for w near $p \in \widehat{\mathbb{C}} \setminus \mathrm{post}(f)$. This is in accordance with the fact (see Theorem 19.2) that ρ is a positive continuous function on $\widehat{\mathbb{C}} \setminus \mathrm{post}(f)$.

PROOF. Since ρ is a fixed point of the Ruelle operator (see the proof of Theorem 19.2), we have

$$\rho(w) = \sum_{z \in f^{-n}(w)} \rho(z) J_{f^n}(z)^{-1}$$

for $w \in \widehat{\mathbb{C}} \setminus \mathrm{post}(f)$.

Fix $p \in \mathrm{post}(f)$. We can find $n \in \mathbb{N}$ and $q \in f^{-n}(p)$ such that $\deg(f^n, q) = \beta_f(p)$. Then clearly $q \notin \mathrm{post}(f)$, and so $\rho(z) \asymp 1$ for z near q.

A point w near p has at least one preimage z near q under f^n. As in the previous proof one sees that

$$J_{f^n}(z) = [(f^n)^\sharp(z)]^2 \asymp |z - q|^{2(\beta_f(p)-1)} \asymp |w - p|^{2-2/\beta_f(p)}.$$

Thus we obtain the lower bound

$$\rho(w) \geq \rho(z) J_{f^n}(z)^{-1} \gtrsim |w - p|^{-2+2/\beta_f(p)}.$$

For an inequality in the other direction, we note that the proof of Lemma 19.9 (ii) showed that

$$M(w, w_0) \lesssim |w - p|^{-2+2/\beta_f(p)}$$

for w near p, where M is the function defined in (19.12) and $w_0 \in \widehat{\mathbb{C}} \setminus \operatorname{post}(f)$ is a base point. Hence

$$\rho(w) \lesssim |w - p|^{-2 + 2/\beta_f(p)}$$

as follows from the proof of Lemma 19.10 (we have to pass to the sublimit ρ in (19.19)). The claim follows. $\square$

Theorem 19.4 is an easy consequence of Theorem 19.11.

PROOF OF THEOREM 19.4. Let f be a rational expanding Thurston map, ν_f be its measure of maximal entropy, and λ_f be the unique f-invariant probability measure that is absolutely continuous with respect to Lebesgue measure.

If f is a Lattès map, then $\nu_f = \lambda_f$ by Theorem 19.3 and so ν_f is absolutely continuous with respect to $\mathcal{L}$.

If f is not a Lattès map, then $\nu_f \neq \lambda_f$ by Theorem 19.11. Since both ν_f and λ_f are ergodic f-invariant probability measures on $\widehat{\mathbb{C}}$, they must be mutually singular in this case. Since λ_f and $\mathcal{L}$ lie in the same measure class, it follows that ν_f and $\mathcal{L}$ are mutually singular. In particular, ν_f is not absolutely continuous with respect to $\mathcal{L}$. $\square$

Before we turn to the proof of Theorem 18.1 (iii), we require an elementary lemma.

LEMMA 19.14. *Let* $\varphi \colon X \to X'$ *be a snowflake homeomorphism between metric spaces* (X, d) *and* (X', d') *such that*

$$d'(\varphi(x), \varphi(y)) \asymp d(x, y)^\beta$$

for all $x, y \in X$*, where* $\beta > 0$ *and* $C(\asymp)$ *are constants independent of x and y.*

Then (X, d) *is Ahlfors Q-regular for $Q > 0$ if and only if* (X', d') *is Ahlfors Q'-regular, where* $Q' = Q/\beta$*. Moreover, for the corresponding Hausdorff measures we have*

$$\mathcal{H}_{d'}^{Q'}(\varphi(M)) \asymp \mathcal{H}_d^Q(M)$$

for each Borel set $M \subset X$ with $C(\asymp)$ independent of M.

PROOF. Let $Q > 0$ be arbitrary and $Q' = Q/\beta$. Then a straightforward covering argument shows that

$$\mathcal{H}_{d'}^{Q'}(\varphi(M)) \asymp \mathcal{H}_d^Q(M)$$

for each Borel set $M \subset X$.

Suppose that (X', d') is Ahlfors Q'-regular, and let B be a closed ball in X of radius $R \leq \operatorname{diam}_d(X)$. Then there exist closed balls B' and B'' in X' with $B' \subset \varphi(B) \subset B''$ whose radii are comparable to

$R' := R^\beta \lesssim \operatorname{diam}_{d'}(X')$. Then the Ahlfors Q'-regularity of (X', d') implies that

$$\mathcal{H}_{d'}^{Q'}(\varphi(B)) \asymp (R')^{Q'} \asymp R^Q,$$

and so

$$\mathcal{H}_d^Q(B) \asymp \mathcal{H}_{d'}^{Q'}(\varphi(B)) \asymp R^Q.$$

This shows that (X, d) is Ahlfors Q-regular.

The other implication is obtained by reversing the roles of X and X'. $\qquad\square$

PROOF OF THEOREM 18.1 (iii). Let ϱ be a visual metric for the expanding Thurston map $f\colon S^2 \to S^2$.

Assume first that f is topologically conjugate to a Lattès map. Since such a topological conjugacy is in fact a snowflake homeomorphism with respect to visual metrics (see Proposition 8.8), we can assume that $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is a Lattès map. Let σ be the chordal metric on $\widehat{\mathbb{C}}$, and ω be the canonical orbifold metric for f (see Section A.10 for the definition of ω).

We know from Proposition 8.5 that ω is a visual metric for f Since two visual metrics are snowflake equivalent according to Proposition 8.3 (iv), we are further reduced to the case $\varrho = \omega$. Now for a Lattès map the spaces $(\widehat{\mathbb{C}}, \omega)$ and $(\widehat{\mathbb{C}}, \sigma)$ are bi-Lipschitz equivalent, because the orbifold of f has no punctures (see Lemma A.32 (i)). It follows that $(S^2, \varrho) = (\widehat{\mathbb{C}}, \omega)$ is snowflake equivalent to $(\widehat{\mathbb{C}}, \sigma)$ as desired.

To prove the other implication, assume that (S^2, ϱ) is snowflake equivalent to $(\widehat{\mathbb{C}}, \sigma)$. In particular, these spaces are quasisymmetrically equivalent. Thus, by Theorem 18.1 (ii), the map f is topologically conjugate to a rational map. So as before, we can assume that $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is in fact a rational map (which is postcritically-finite and has no periodic critical points). In order to prove that f is a Lattès map, we will verify the condition in Theorem 19.4 and show that the measure of maximal entropy ν_f is absolutely continuous with respect to Lebesgue measure $\mathcal{L}$ on $\widehat{\mathbb{C}}$.

According to our hypotheses there exists a snowflake homeomorphism $\varphi\colon (\widehat{\mathbb{C}}, \varrho) \to (\widehat{\mathbb{C}}, \sigma)$. Then

$$\sigma(\varphi(x), \varphi(y)) \asymp \varrho(x, y)^\beta,$$

for all $x, y \in \widehat{\mathbb{C}}$, where $\beta > 0$ and $C(\asymp)$ are constants independent of x and y.

Let $\mathcal{H}^2_\sigma$ denote 2-dimensional Hausdorff measure on $\widehat{\mathbb{C}}$ with respect to the chordal metric σ, and $\mathcal{H}^{2\beta}_\varrho$ denote (2β)-dimensional Hausdorff measure on $\widehat{\mathbb{C}}$ with respect to the metric ϱ.

Since $(\widehat{\mathbb{C}}, \sigma, \mathcal{H}^2_\sigma)$ is Ahlfors 2-regular, $(\widehat{\mathbb{C}}, \varrho, \mathcal{H}^{2\beta}_\varrho)$ is Ahlfors (2β)-regular by Lemma 19.14. Moreover,

$$\mathcal{H}^2_\sigma(\varphi(M)) \asymp \mathcal{H}^{2\beta}_\varrho(M)$$

for each Borel set $M \subset \widehat{\mathbb{C}}$. Since ϱ is a visual metric and $\mathcal{H}^{2\beta}_\varrho$ is Ahlfors regular, the measures ν_f and $\mathcal{H}^{2\beta}_\varrho$ are comparable (see the discussion after Proposition 18.2).

Now let $M \subset \widehat{\mathbb{C}}$ be an arbitrary Borel set with $\mathcal{L}(M) = 0$, and define $N = \varphi(M)$. By Lemma 18.10 the map $\mathrm{id}_{\widehat{\mathbb{C}}} \colon (\widehat{\mathbb{C}}, \sigma) \to (\widehat{\mathbb{C}}, \varrho)$ is a quasisymmetry. This implies that the composition

$$(\widehat{\mathbb{C}}, \sigma) \xrightarrow{\mathrm{id}} (\widehat{\mathbb{C}}, \varrho) \xrightarrow{\varphi} (\widehat{\mathbb{C}}, \sigma),$$

i.e., the map $\varphi \colon (\widehat{\mathbb{C}}, \sigma) \to (\widehat{\mathbb{C}}, \sigma)$, is also a quasisymmetry. Since quasisymmetries on $\widehat{\mathbb{C}}$ preserves sets of (Lebesgue-) measure zero (see [**AIM09**, Theorem 3.4.1 and Theorem 3.1.2]), we have $\mathcal{L}(N) = 0$. It follows that

$$\nu_f(M) \asymp \mathcal{H}^{2\beta}_\varrho(M) \asymp \mathcal{H}^2_\sigma(\varphi(M)) \asymp \mathcal{L}(\varphi(M)) = \mathcal{L}(N) = 0.$$

So ν_f is indeed absolutely continuous with respect to Lebesgue measure, and f is a Lattès map by Theorem 19.4. $\qquad\square$

CHAPTER 20

A combinatorial characterization of Lattès maps

In this chapter we characterize Lattès maps in terms of their combinatorial expansion behavior. This is based on results by Qian Yin (see[**Yi15a**]).

Let $f\colon S^2 \to S^2$ be an expanding Thurston map, and $\mathcal{C} \subset S^2$ be a Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}$. We defined the quantity $D_n(f,\mathcal{C})$ for $n \in \mathbb{N}_0$ as the minimal cardinality of a set of n-tiles for $(f,\mathcal{C})$ whose union is connected and joins opposite sides of $\mathcal{C}$ (see (5.14) and Definition 5.32). The combinatorial expansion factor $\Lambda_0(f)$ of f as introduced in Chapter 16 is related to the growth rate of $D_n(f,\mathcal{C})$, and given by

$$\Lambda_0(f) = \lim_{n\to\infty} D_n(f,\mathcal{C})^{1/n}.$$

We have already seen that $\Lambda_0(f) \le \deg(f)^{1/2}$ in the case that f has no periodic critical points; see the discussion after Proposition 18.2. This implies that $D_n(f,\mathcal{C})$ cannot grow much faster than $\deg(f)^{n/2}$ as $n \to \infty$. As the following statement shows, up to a multiplicative constant this is actually a precise upper bound for the growth rate of $D_n(f,\mathcal{C})$.

PROPOSITION 20.1. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map and $\mathcal{C} \subset S^2$ be a Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}$. Then there exists a constant $c > 0$ such that*

$$(20.1) \qquad\qquad D_n(f,\mathcal{C}) \le c\deg(f)^{n/2}$$

for all $n \in \mathbb{N}$. Moreover, we have $\Lambda_0(f) \le \deg(f)^{1/2}$.

This will be proved in Section 20.3. In view of Proposition 20.1 one can ask whether there are maps for which $D_n(f,\mathcal{C})$ actually grows as fast as $\deg(f)^{n/2}$ as $n \to \infty$. It turns out that Lattès maps are essentially characterized by this property. This is the main result of this chapter.

THEOREM 20.2. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map. Then f is topologically conjugate to a Lattès map if and only if the following conditions are true:*

    (i) *f has no periodic critical points.*

(ii) *There exists $c > 0$, and a Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}$ such that*

(20.2) $$D_n(f, \mathcal{C}) \geq c \deg(f)^{n/2}$$

for all $n \in \mathbb{N}_0$.

This theorem is due to Qian Yin (see [**Yi15a**]). The proof will occupy most of this chapter. We will follow Yin's approach with some modifications and simplifications.

If f is a rational map, then we get a slightly stronger statement.

COROLLARY 20.3. *Let $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a rational expanding Thurston map. Then f is a Lattès map if and only if condition* (ii) *in Theorem 20.2 is satisfied.*

We do not know whether a similar improvement of Theorem 20.2 is possible for arbitrary (not necessarily rational) expanding Thurston maps.

If (20.2) is true for *some* Jordan curve $\mathcal{C}$, then the same condition holds for *each* Jordan curve $\mathcal{C}' \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}'$ (in general with a different constant $c' > 0$ depending on $\mathcal{C}'$). This follows from the fact that $D_n(f, \mathcal{C}) \asymp D_n(f, \mathcal{C}')$ which was shown in Lemma 16.5 (see (16.3)).

By Proposition 20.1 the inequality $D_n(f, \mathcal{C}) \lesssim \deg(f)^{n/2}$ is always true, so (20.2) says that $D_n(f, \mathcal{C}) \asymp \deg(f)^{n/2}$ as $n \to \infty$. This asymptotic behavior of $D_n(f, \mathcal{C})$ implies that $\Lambda_0(f) = \deg(f)^{1/2}$. This equality is slightly weaker than the requirement (20.2). It would be very interesting to characterize the expanding Thurston maps whose combinatorial expansion factor is maximal in this sense. Besides for expanding Thurston that are topologically conjugate to Lattès maps, it is also satisfied for certain Lattès-type maps (similar to Example 16.8 where the associated linear map on $\mathbb{R}^2$ has a shear component).

20.1. Visual metrics, 2-regularity, and Lattès maps

Here we begin the proof of Theorem 20.2. Some parts are somewhat technical. We delegate them to the next two sections, and finish the proof here assuming these technical results.

The main implication (i.e., the sufficiency) in Theorem 20.2 will be a consequence of the following statement.

THEOREM 20.4. *Let $f \colon S^2 \to S^2$ be an expanding Thurston map. Then f is topologically conjugate to a Lattès map if and only if there is a visual metric ϱ for f such that (S^2, ϱ) is Ahlfors 2-regular.*

PROOF. First suppose that f is topologically conjugate to a Lattès map. To see the existence of a visual metric with the desired property, we can in fact assume that $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is a Lattès map. Then the canonical orbifold metric $\varrho = \omega$ for f is a visual metric with expansion factor $\Lambda = \deg(f)^{1/2}$ (see Proposition 8.14). In addition, f has no periodic critical points (see Theorem 3.1 (i)). By Proposition 18.2 the space $(\widehat{\mathbb{C}}, \varrho)$ is Ahlfors 2-regular.

Conversely, assume that $f\colon S^2 \to S^2$ is an expanding Thurston map and ϱ is a visual metric such that (S^2, ϱ) is Ahlfors 2-regular. Here the underlying measure is 2-dimensional Hausdorff measure $\mathcal{H}^2_\varrho$. Since every Ahlfors regular space is doubling, it follows from Theorem 18.1 (i) that f has no periodic critical points.

The metric space (S^2, ϱ) is linearly locally connected by Proposition 18.5 (iii). Hence Theorem 4.2 applies and (S^2, ϱ) is quasisymmetrically equivalent to the Riemann sphere $\widehat{\mathbb{C}}$ equipped with the chordal metric σ. This in turn implies by Theorem 18.1 (ii) that f is topologically conjugate to a rational map. Since our assumptions are preserved under such a conjugacy, we are reduced to the case where $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is a *rational* expanding Thurston map.

Recall from Lemma 18.10 that the identity map $\mathrm{id}_{\widehat{\mathbb{C}}}\colon (\widehat{\mathbb{C}}, \sigma) \to (\widehat{\mathbb{C}}, \varrho)$ is a quasisymmetry. Hence by Theorem 4.3 the Hausdorff 2-measure on $\widehat{\mathbb{C}}$ with respect to the chordal metric σ (i.e., the standard Lebesgue measure on $\widehat{\mathbb{C}}$) and the Hausdorff 2-measure $\mathcal{H}^2_\varrho$ on $\widehat{\mathbb{C}}$ with respect to the visual metric ϱ are absolutely continuous with respect to each other. On the other hand, by Proposition 18.2 the measure of maximal entropy ν_f of f and $\mathcal{H}^2_\varrho$ are comparable. Thus, ν_f is absolutely continuous with respect to Lebesgue measure on $\widehat{\mathbb{C}}$. Zdunik's theorem (i.e., Theorem 19.4) now implies that f is a Lattès map. $\qquad\square$

The previous theorem and its proof combined with Proposition 18.2 also give the following statement.

COROLLARY 20.5. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map. Then f is topologically conjugate to a Lattès map if and only if*

(i) *f has no periodic critical points, and*

(ii) *there is a visual metric ϱ for f with expansion factor $\Lambda = \deg(f)^{1/2}$.*

Thus in order to prove that conditions (i) and (ii) in Theorem 20.2 imply that f is topologically conjugate to a Lattès map, it is enough to construct a visual metric ϱ with expansion factor $\Lambda = \deg(f)^{1/2}$.

We will construct such a metric from combinatorial data. To this end, we fix a Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) \subset \mathcal{C}$ and consider tiles for $(f, \mathcal{C})$. For $x, y \in S^2$ we then define

(20.3) $N_n(x, y) :=$

$$\min \{\text{length}(P) : P \text{ is a chain of } n\text{-tiles joining } x \text{ and } y\}.$$

See Definition 5.19 for our terminology. The following lemma will now provide the main step in the proof of Theorem 20.2.

LEMMA 20.6. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map that has an invariant Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) \subset \mathcal{C}$. Suppose that $\Lambda := \deg(f)^{1/2} > 2$ and that condition (20.2) in Theorem 20.2 is satisfied for $\mathcal{C}$. Then*

$$N_n(x, y) \asymp \Lambda^{n-m(x,y)}$$

for all $x, y \in S^2$ and $n \in \mathbb{N}$ with $n \geq m(x, y) + 1$. Here the constant $C(\asymp)$ is independent of x, y, and n.

We use the notation $m(x, y) = m_{f,\mathcal{C}}(x, y)$, where $m_{f,\mathcal{C}}$ is as in Definition 8.1. Note that $m(x, y) = \infty$ if $x = y$. In this case, the statement is vacuous as there is no integer with $n \geq m(x, y) = \infty$.

The proof of this lemma will occupy the bulk of the next two sections. The assumptions that $\mathcal{C}$ is invariant and that $\deg(f)^{1/2} > 2$ are not essential and the lemma is still true without these hypotheses, but they will simplify the proof.

The previous lemma gives us the following consequence.

LEMMA 20.7. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map that satisfies condition (20.2) in Theorem 20.2. Then there is a visual metric ϱ for f with expansion factor $\Lambda = \deg(f)^{1/2}$.*

PROOF. Let $f\colon S^2 \to S^2$ be an expanding Thurston map that satisfies (20.2) for some Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) \subset \mathcal{C}$.

We assume first that $\mathcal{C}$ is f-invariant and $\Lambda := \deg(f)^{1/2} > 2$. Then for $x, y \in S^2$ we define

(20.4) $\varrho(x, y) = \limsup_{n \to \infty} \Lambda^{-n} N_n(x, y).$

Note that $N_n(x, x) = 1$ for $x \in S^2$ and $n \in \mathbb{N}$, and so $\varrho(x, x) = 0$. This together with Lemma 20.6 shows that

$$\varrho(x, y) \asymp \Lambda^{-m(x,y)},$$

where $m = m_{f,\mathcal{C}}$ and $C(\asymp)$ is independent of x and y. Note that in particular $\varrho(x, y) \in [0, \infty)$ for $x, y \in S^2$.

So ϱ will be a visual metric for f with expansion factor $\Lambda = \deg(f)^{1/2}$ if we can show that ϱ is indeed a metric. For $x, y \in S^2$ we obviously have $\varrho(x, y) = \varrho(y, x)$ and $\varrho(x, y) = 0$ if and only if $x = y$. The triangle inequality for ϱ follows immediately from the inequality

$$N_n(x, z) \leq N_n(x, y) + N_n(y, z)$$

valid for all $n \in \mathbb{N}_0$ and $x, y, z \in S^2$. Thus ϱ is a visual metric with expansion factor $\Lambda = \deg(f)^{1/2}$.

We now consider the general case without assuming that $\mathcal{C}$ is f-invariant and $\deg(f)^{1/2} > 2$. By Theorem 15.1 we can pick an iterate $F = f^k$ of f that has an F-invariant Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) = \mathrm{post}(F) \subset \mathcal{C}$. By picking k large enough we may also assume that $\deg(F)^{1/2} > 2$.

Recall that by (16.3) in Lemma 16.5 condition (20.2) is essentially independent of the chosen Jordan curve. So we assume that f satisfies this condition for the F-invariant curve $\mathcal{C}$. Since the n-tiles for $(F, \mathcal{C})$ are precisely the (nk)-tiles for $(f, \mathcal{C})$ (see Proposition 5.17 (vii)), it follows that

$$D_n(F, \mathcal{C}) = D_{nk}(f, \mathcal{C}) \gtrsim \deg(f)^{nk/2} = \deg(F)^{n/2}$$

for all $n \in \mathbb{N}$. So by the first part of the proof, there exists a visual metric ϱ for F with expansion factor $\Lambda_F = \deg(F)^{1/2}$. Proposition 8.3 (v) implies that ϱ is a visual metric for f with expansion factor $\Lambda = \Lambda_F^{1/k} = \deg(F)^{1/(2k)} = \deg(f)^{1/2}$. $\qquad\square$

Assuming that Lemma 20.6 is valid, we can now finish the proof of Theorem 20.2.

PROOF OF THEOREM 20.2. The sufficiency part of Theorem 20.2 is now easy. Indeed, let $f\colon S^2 \to S^2$ be an expanding Thurston map without periodic critical points and $\mathcal{C} \subset \widehat{\mathbb{C}}$ be a Jordan curve with $\mathrm{post}(f) \subset \mathcal{C}$ such that (20.2) is satisfied. By Lemma 20.7 there is a visual metric ϱ for f with expansion factor $\Lambda = \deg(f)^{1/2}$. Corollary 20.5 shows that f is topologically conjugate to a Lattès map as desired.

To prove the reverse implication, let $f\colon S^2 \to S^2$ be topologically conjugate to a Lattès map. We want to show that f satisfies the conditions (i) and (ii) in Theorem 20.2. Since these conditions are invariant under topological conjugacy (in a suitable sense), we may assume that f is a Lattès map on the Riemann sphere $\widehat{\mathbb{C}}$. Then f has no periodic critical points (see Theorem 3.1 (i)), and so (i) is true.

We also know that f has a parabolic orbifold. Let ω be the canonical orbifold metric of f on $\widehat{\mathbb{C}}$ (see Section A.10). Then by Proposition 8.14 this metric is a visual metric for f with expansion factor $\Lambda = \deg(f)^{1/2}$.

Now we pick a Jordan curve $\mathcal{C} \subset \widehat{\mathbb{C}}$ with $\mathrm{post}(f) \subset \mathcal{C}$ and consider tiles for $(f, \mathcal{C})$. Fix $n \in \mathbb{N}$. Then, according to the definition of $D_n(f, \mathcal{C})$, we can find a connected union K of n-tiles that joins opposite sides of $\mathcal{C}$ and consists precisely of $D_n(f, \mathcal{C})$ tiles. Then $\mathrm{diam}_\omega(K) \geq \delta_0$, where $\delta_0 > 0$ is the constant in (5.13) for the underlying base metric ω on $\widehat{\mathbb{C}}$. Among the n-tiles that form K we can find a chain $X_1, \ldots, X_N$ of distinct n-tiles that joins two points $x, y \in K$ with $\omega(x, y) = \mathrm{diam}_\omega(K)$. Then $N \leq D_n(f, \mathcal{C})$, and

$$\sum_{i=1}^{N} \mathrm{diam}_\omega(X_i) \geq \omega(x, y) \geq \delta_0.$$

Since ω is a visual metric for f with expansion factor $\Lambda = \deg(f)^{1/2}$, by Proposition 8.4 (ii) we have $\mathrm{diam}_\omega(X_i) \leq C\Lambda^{-n}$ with a constant $C > 0$ independent of n. Hence $CN \deg(f)^{-n/2} \geq \delta_0$, and so

$$D_n(f, \mathcal{C}) \geq N \gtrsim \deg(f)^{n/2},$$

where $C(\gtrsim)$ is independent of n. This shows that condition (ii) in Theorem 20.2 is also true. $\qquad\square$

20.2. Separating sets with tiles

In the previous section we have seen that in order to prove Theorem 20.2 we have to establish Lemma 20.6. This means that for given $x, y \in S^2$ we have to estimate $N_n(x, y)$, the minimal number of n-tiles in a chain joining x and y, in terms of $m_{f,\mathcal{C}}(x, y)$ (see (20.3)). Here we are of course assuming that (20.2) holds. The lower bound easily follows from (20.2) together with Lemma 5.36.

LEMMA 20.8. *Let $f \colon S^2 \to S^2$ be an expanding Thurston map that has an invariant Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}$. Suppose that* (20.2) *holds and define $\Lambda := \deg(f)^{1/2}$. Then*

$$N_n(x, y) \gtrsim \Lambda^{n - m(x,y)},$$

for all $x, y \in S^2$ and $n \in \mathbb{N}$ with $n \geq m(x, y) + 1$. Here the constant $C(\gtrsim)$ is independent of x, y, and n.

Here again $m(x, y) = m_{f,\mathcal{C}}(x, y)$ and it is understood that we use n-tiles for $(f, \mathcal{C})$ in the definition $N_n(x, y)$.

PROOF. Let $x, y \in S^2$ be two distinct points and $m := m(x, y)$. We pick $(m + 1)$-tiles X^{m+1} and Y^{m+1} (for $(f, \mathcal{C})$) with $x \in X^{m+1}$ and $y \in Y^{m+1}$. Then $X^{m+1} \cap Y^{m+1} = \emptyset$ by definition of $m(x, y)$.

If $n \geq m + 1$ and P is an arbitrary chain of n-tiles joining x and y, then P also joins X^{m+1} and Y^{m+1}. Since these are two disjoint $(m+1)$-cells, it follows from Lemma 5.36 that

$$\text{length}(P) \geq D_{n-m-1}(f, \mathcal{C}).$$

So by using assumption (20.2), we obtain

$$N_n(x, y) \geq D_{n-m-1}(f, \mathcal{C}) \gtrsim \Lambda^{n-m-1} \asymp \Lambda^{n-m}. \qquad \square$$

We have to prove an inequality in the other direction and show that two points $x, y \in S^2$ can be joined by a rather short chain of n-tiles. For this we use a duality argument that will give the existence of short chains provided certain separating sets of tiles do not have too small cardinality. The key ingredient of this duality argument is a well-known graph-theoretical statement, namely Menger's theorem. Before we formulate this result, we first record some definitions.

We consider a finite graph G. Here we take the combinatorial point of view; so G is just a pair (V, E), where V is a finite set called the *set of vertices* of G and $E \subset V \times V$ is a subset of the set of pairs in V, called the *set of edges* of G. We assume that E is symmetric and disjoint from the diagonal $\{(x, x) : x \in V\}$. For $x, y \in V$ we write $x \sim y$ if $(x, y) \in E$. In this case, we say that x and y are *joined by an edge* (in G). Since E is symmetric, we consider edges as non-oriented.

A *path* in G is a finite sequence $x_1, \dots, x_n$ of vertices in G so that successive vertices are distinct and joined by an edge. If $A, B \subset V$, then a path in G *joins* A and B, if the first vertex of the path lies in A, and its last vertex lies in B. A path joining A and B in G is called an *A-B-path*. A set $A \subset V$ is *connected* if for all $a, a' \in A$ there exists a path in A joining a and a'. A set $K \subset V$ *separates* A and B if every A-B-path contains an element in K. Given these definitions, we have the following statement.

THEOREM 20.9 (Menger's theorem). *Let G be a finite graph with vertex set V, and $A, B \subset V$. Then the minimal cardinality of a set separating A and B in G is equal to the maximal number of pairwise disjoint A-B-paths in G.*

For the proof and more background see [**Di10**, Section 3.3]. This theorem can be seen as a special case of the max-flow min-cut theorem (see [**Di10**, Section 6.2]).

An important consequence for us will be that if M is a lower bound for the cardinality of a set separating A and B, then there exists an A-B-path containing at most $\#V/M$ vertices.

The graphs G that we will consider in our context will have sets of tiles as vertex sets. Thus sets of tiles may be viewed in two ways, as sets of vertices in G, or as closed Jordan regions on the underlying 2-sphere. In this situation we want to use the topology of the sphere to obtain information on separating sets in the graph. For this we will invoke some well-known topological facts related to Janiszewski's lemma (see Lemma A.3). Actually, we will require a more refined version stated in Lemma A.4. Since its proof is somewhat long and technical and would distract from our present considerations, we included it in the appendix (see Section A.3).

Let us return to expanding Thurston maps. In the following, we fix such a map $f\colon S^2 \to S^2$ and a Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}$. All cells considered below are defined for $(f, \mathcal{C})$.

Given $k \in \mathbb{N}_0$, we consider a simple e-chain of k-tiles. Recall from Definition 5.20 that this is a sequence $Z_1, \ldots, Z_N$ of distinct k-tiles, where $N \in \mathbb{N}$, and we suppose that there exist k-edges $E_1, \ldots, E_{N-1}$ with $E_i \subset Z_i \cap Z_{i+1}$ for $i = 1, \ldots, N-1$. We now consider the "interior" of this e-chain, i.e., the set

$$(20.5) \qquad \Omega := \bigcup_{i=1}^{N} \mathrm{int}(Z_i) \cup \bigcup_{i=1}^{N-1} \mathrm{int}(E_i).$$

Note that in general, Ω is not the interior of the underlying set $\bigcup_{i=1}^{N} Z_i$; see Figure 20.1 for an example.

LEMMA 20.10. *The set Ω defined in* (20.5) *is a simply connected region in S^2.*

PROOF. For each $i = 1, \ldots, N-1$ the set

$$U_i := \mathrm{int}(Z_i) \cup \mathrm{int}(E_i) \cup \mathrm{int}(Z_{i+1})$$

is an open region (see Lemma 5.9 (iv)). This implies that Ω is open and connected, and hence a region.

The additional statement that Ω is simply connected follows from the fact that every loop γ in Ω can be homotoped to a constant loop inside Ω. To see this, define

$$\Omega_l := \bigcup_{i=1}^{l} \mathrm{int}(Z_i) \cup \bigcup_{i=1}^{l} \mathrm{int}(E_i) \subset \Omega$$

for $l = 1, \ldots, N$, where we set $E_N = \emptyset$. For each $i = 1, \ldots, N-1$ there exists a deformation retraction of the set

$$\mathrm{int}(E_i) \cup \mathrm{int}(Z_{i+1}) \cup \mathrm{int}(E_{i+1})$$

onto $\operatorname{int}(E_i)$ which implies that there exists a deformation retraction of Ω_{i+1} onto Ω_i. Here it is important that

$$(\operatorname{int}(Z_{i+1}) \cup \operatorname{int}(E_{i+1})) \cap \Omega_i = \emptyset$$

which follows from the facts that the k-tiles $Z_1, \ldots, Z_N$ are all distinct and that the set $\operatorname{int}(E_i)$ only meets the k-tiles Z_i and Z_{i+1} for $i = 1, \ldots, N-1$ (this follows from Lemma 5.9 (iv)).

By using these deformation retractions successively, every closed loop in $\Omega = \Omega_N$ can be homotoped inside Ω to a loop in $\Omega_1 \subset \Omega$. The set Ω_1 is the union of an open Jordan region with an open arc on its boundary if $N > 1$ and an open Jordan region if $N = 1$. Hence Ω_1 is contractible, and the simple connectivity of Ω follows. $\square$

We now assume in addition that the Jordan curve $\mathcal{C}$ is f-invariant. As before, $Z_1, \ldots, Z_N$ denotes an e-chain of k-tiles.

Fix $n \in \mathbb{N}_0$ with $n \geq k$. Since $\mathcal{C}$ is f-invariant, the k-cells are subdivided into n-cells. We form a graph G whose vertex set V consists of all n-tiles contained in any of the k-tiles Z_i. Then V is the set of all n-tiles X with $\operatorname{int}(X) \subset \Omega$. In G we join two distinct vertices in V given by n-tiles X and Y by an edge if there exists an n-edge $e \subset X \cap Y$ with $\operatorname{int}(e) \subset \Omega$.

Note that if $X, Y \in V$, $X \neq Y$, and X and Y share an n-edge e, then there are two possibilities. Namely, X and Y may lie in the same k-tile Z_i. Then necessarily $\operatorname{int}(e) \subset \operatorname{int}(Z_i) \subset \Omega$ and so the vertices X and Y are joined by an edge in G. If X and Y lie in different k-tiles Z_i, then the only situation where X and Y are joined by an edge in G is when X and Y lie in consecutive k-tiles of the e-chain, say $X \subset Z_i$ and $Y \subset Z_{i+1}$, and $\operatorname{int}(e) \subset \operatorname{int}(E_i) \subset \Omega$. So we only join "across" the k-edges $E_1, \ldots, E_{N-1}$, but not across any other k-edge contained in $Z_1 \cup \cdots \cup Z_N$.

Given a set of n-tiles M we use the notation $|M|$ for the underlying point set, i.e.,

$$|M| := \bigcup_{X \in M} X.$$

The next lemma relates connectedness and separation properties of sets of n-tiles considered as sets of vertices in G with the corresponding properties of the underlying sets.

LEMMA 20.11. *With the given assumptions, the following statements are true:*

(i) *Let γ be a path in Ω and $M \subset V$ be the set of all n-tiles X with $\operatorname{int}(X) \subset \Omega$ and $X \cap \gamma \neq \emptyset$. Then M is connected in G, and $M' = |M| \cap \Omega$ is a connected subset of Ω.*

(ii) *Let γ be a path in the boundary ∂Z_i of one of the k-tiles Z_i, and $M \subset V$ be the set of all n-tiles X with $X \subset Z_i$ and $X \cap \gamma \neq \emptyset$. Then M is connected in G, and $M' = |M| \cap \Omega$ is a connected subset of Ω.*

(iii) *If $A, B, K \subset V$, then K separates A and B in G if and only if $K' = |K| \cap \Omega$ separates $A' = |A| \cap \Omega$ and $B' = |B| \cap \Omega$ in Ω.*

Here we say that the set K' *separates* A' and B' in Ω if every path γ in Ω joining A' and B' meets K' (see Section A.3 for a more detailed discussion).

PROOF. Paths in G precisely correspond to e-chains consisting of n-tiles $X_1, \ldots, X_l \in V$ for which there exist n-edges $e_1, \ldots, e_{l-1}$ with $e_i \subset X_i \cap X_{i+1}$ and $\mathrm{int}(e_i) \subset \Omega$ for $i = 1, \ldots, l-1$, where $l \in \mathbb{N}$. For the rest of the proof we call such an e-chain of n-tiles *admissible*. The admissible e-chain $X_1, \ldots, X_l$ *connects* two points $p, q \in \Omega$ if $p, q \in \bigcup_{i=1}^{l} X_i$.

(i) The argument for this is along the lines of the proof of Lemma 11.16 with small modifications. We first analyze connectivity properties of admissible e-chains near points in Ω.

For fixed $p \in \Omega$, we define

$$\Omega^n(p) = \bigcup \{\mathrm{int}\,(c) : c \text{ is an } n\text{-cell with } p \in c\}.$$

Recall that by definition $\mathrm{int}\,(c) = c$ if c is a 0-dimensional n-cell (i.e., c is a singleton set consisting of an n-vertex). We know that each point in S^2 is contained in the interior of a unique n-cell (see Lemma 5.2). Thus, there are three types of sets $\Omega^n(p)$ depending on whether p is contained in the interior of an n-tile X_p, the interior of an n-edge e_p, or is an n-vertex.

In the first case, $\Omega^n(p) = \mathrm{int}(X_p)$. In the second case, $\Omega^n(p)$ is the union of $\mathrm{int}(e_p)$ with the interiors of the two n-tiles that contain e_p (see Lemma 5.9 (iv)). In the third case, $\Omega^n(p)$ is the n-flower of p (see Definition 5.27 and Lemma 5.28 (i)). In particular, $\Omega^n(p)$ is always an open set.

Claim 1. $\Omega^n(p) \subset \Omega$.

Recall from Lemma 5.7 that the interior of each n-cell is contained in the interior of a unique k-cell. Since Ω is a union of interiors of k-cells, we have $\mathrm{int}(c) \subset \Omega$ for an n-cell c if and only if $\mathrm{int}(c)$ contains a point in Ω. Now every n-cell c with $p \in c$ has points in its interior arbitrary close to p. Since $p \in \Omega$ and Ω is open, we conclude that $\mathrm{int}(c) \subset \Omega$ for such a cell c. Claim 1 follows.

The same argument also shows that $\text{int}(X) \subset \Omega$ for every n-tile X with $p \in X$; in this case, X represents a vertex in the graph G.

Claim 2. If $x, y \in \Omega^n(p)$, then there exists an admissible e-chain connecting x and y.

Indeed, since $\Omega^n(p) \subset \Omega$, the tile X_p forms such an admissible e-chain if $\Omega^n(p)$ is of the first type. For the second type, the two n-tiles containing e_p form such an e-chain. Finally, for the third type the n-tiles $X_1, \ldots, X_d$ containing p labeled cyclically around p form such an admissible e-chain (see Lemma 5.9 (v)). Claim 2 follows.

Now let $M \subset V$ be the set associated to a path γ as in the statement. The proof that M is connected in G amounts to showing that every two n-tiles in M can be joined by an e-chain of n-tiles that is admissible. For this in turn it is enough to assume that $\gamma\colon [a, b] \to \Omega$ is defined on a compact interval $[a, b] \subset \mathbb{R}$ and show that if $X, Y \in V$, and $\gamma(a) \in X$, $\gamma(b) \in Y$, then there is an admissible e-chain of n-tiles joining X and Y.

Based on Claim 2 it actually suffices to show that $\gamma(a)$ and $\gamma(b)$ can be connected by an admissible e-chain. Now let $I \subset [a, b]$ be the set of all numbers $s \in [a, b]$ such that there exists an admissible e-chain connecting $\gamma(a)$ and $\gamma(s)$. Clearly $a \in I$, and so I is non-empty. As in the proof of Lemma 11.16, one shows that I is closed. On the other hand, I is open in $[a, b]$ as follows from Claim 2 together with the fact that $\Omega^n(p)$ for each $p \in \Omega$ is open. So $I = [a, b]$, and it follows that M is connected in G.

The connectivity of the set $M' = |M| \cap \Omega$ can easily be derived from this. Indeed, to connect two points $p, q \in M'$, one chooses n-tiles $X, Y \in M$ with $p \in X$ and $q \in Y$. By what we have just seen, one can find an admissible e-chain consisting of tiles $X_1 = X, \ldots, X_l = Y$ in M. There are n-edges $e_1, \ldots, e_{l-1}$ with $e_i \subset X_i \cap X_{i+1}$ and $\text{int}(e_i) \subset \Omega$ for $i = 1, \ldots, l - 1$. To find a path γ that stays in $|M| \cap \Omega$ and joins p and q, we travel from p to a point in $\text{int}(e_1)$ along an arc whose interior stays in $\text{int}(X_1)$. We cross over to the interior of the next tile X_2 and travel along an arc whose interior stays in $\text{int}(X_2)$ to a point in $\text{int}(e_2)$, etc., until we finally join a suitable point in $\text{int}(e_{l-1})$ to q by an arc whose interior stays in $\text{int}(X_l)$. The concatenation of these arcs gives a path in $M' = |M| \cap \Omega$ joining p and q.

(ii) Let γ be a curve in the boundary of one of the n-tiles Z_i. Similar to the previous argument, for $p \in \partial Z_i$ we define

$$\Omega^n(p) = \bigcup \{\text{int}(c) : c \text{ is an } n\text{-cell with } p \in c \text{ and } c \subset Z_i\}.$$

If p is contained in the interior of an n-edge $e_p \subset \partial Z_i$, then $\Omega^n(p)$ is the union of $\text{int}(e_p)$ with the interior of the unique n-tile that contains e_p and is contained in Z_i. If p is an n-vertex, there are exactly two n-edges belonging to the cycle of p and contained in ∂Z_i. This implies that the Jordan curve ∂Z_i splits the n-tiles and the n-edges in the cycle of p into two families, namely those with interior contained in Z_i and those with interior disjoint from Z_i. Moreover, the interiors of the n-tiles and n-edges in the first family are all contained in $\Omega^n(p)$.

In any case, $\Omega^n(p)$ is again a relatively open neighborhood of p in Z_i and any two points in $\Omega^n(p) \cap \Omega$ can be connected by an admissible e-chain. The argument is now identical to the one in case (i).

(iii) "$\Leftarrow$" With the given setup suppose that $K' = |K| \cap \Omega$ separates $A' = |A| \cap \Omega$ and $B' = |B| \cap \Omega$ in Ω. A path in G joining A and B corresponds to an admissible e-chain P consisting of n-tiles $X_1, \ldots, X_l \in V$ with $X_1 \in A$ and $X_l \in B$ for which there exist n-edges $e_1, \ldots, e_{l-1}$ with $e_i \subset X_i \cap X_{i+1}$ and $\text{int}(e_i) \subset \Omega$ for $i = 1, \ldots, l-1$. We have to show that $P \cap K \neq \emptyset$.

Similar to the last part of the proof of (i), one can find a path γ that lies in the set

$$\bigcup_{i=1}^{l} \text{int}(X_i) \cup \bigcup_{i=1}^{l-1} \text{int}(e_i) \subset \Omega,$$

and has one endpoint in $\text{int}(X_1)$ and one in $\text{int}(X_N)$. Then γ does not meet any tile in V that does not belong to P. The path γ joins A' and B' in Ω and hence meets $K' \subset |K|$ by our assumptions. By choice of γ we then must have $P \cap K \neq \emptyset$ as desired.

(iii) "$\Rightarrow$" Suppose K separates A and B in G, and suppose γ is a path in Ω joining $A' = |A| \cap \Omega$ and $B' = |B| \cap \Omega$. Let M be the set of all n-tiles that meet γ. Since $\gamma \subset \Omega$ we have $M \subset V$. By (i) the set M is connected in G, and so there exists a path P in G that starts in A and ends in B and consists of n-tiles in $M \subset V$. By our assumptions we have $P \cap K \neq \emptyset$. In particular, M and K have a tile in common which implies that $\gamma \cap K' \neq \emptyset$, where $K' = |K| \cap \Omega$. Hence K' separates A' and B' in Ω. $\qquad \square$

The following lemma provides the main estimate of this section. Recall that we defined $D_n(f, \mathcal{C})$ as the minimal number of n-tiles that form a connected set joining opposite sides of $\mathcal{C}$ (see (5.14)). We will use this quantity to estimate the minimal number of n-tiles that are required to separate certain sets.

The setting is still the same as in the previous lemma. So we are given an e-chain $Z_1, \ldots, Z_N$ of k-tiles and Ω is as in (20.5). The k-tiles
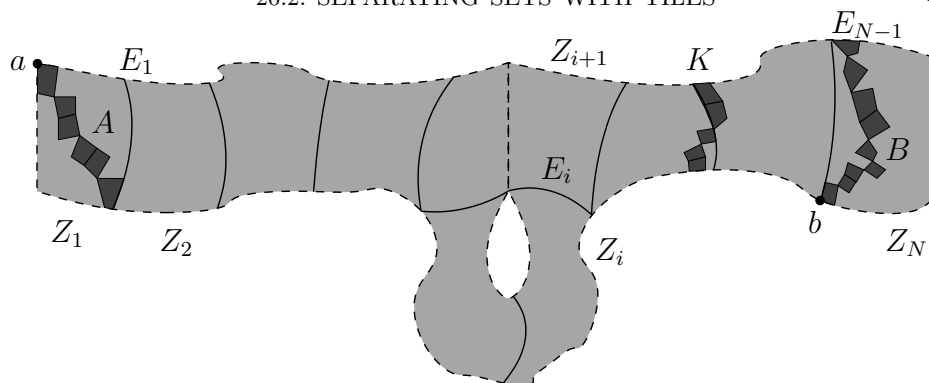
FIGURE 20.1. Lemma 20.12

Z_i are subdivided into n-tiles, and G is the graph defined above with its vertex set V given by the set of all n-tiles contained in one of the tiles Z_i.

LEMMA 20.12. *Let $A, B \subset V$, and suppose that each of the sets $|A|$ and $|B|$ contains at least two distinct k-vertices and that $A' = |A| \cap \Omega$ and $B' = |B| \cap \Omega$ are connected. Let $K \subset V$ be a set that separates A and B in G. Then*

$$(20.6) \qquad\qquad \#K \geq D_{n-k}(f, \mathcal{C}).$$

PROOF. Let $K \subset V$ be a set of minimal cardinality that separates sets A and B in V satisfying the given hypotheses. The existence of such a set K follows from the fact that the whole vertex set V separates A and B according to our definition. The setup is illustrated in Figure 20.1. Note that the dashed edges are not part of the region Ω. The main idea of the proof is now to show that $|K|$ is a connected set not contained in a k-flower. This will lead to the desired bound.

By Lemma 20.11 (iii) the set $K' = |K| \cap \Omega$ separates $A' = |A| \cap \Omega$ and $B' = |B| \cap \Omega$ in Ω. Since Ω is a simply connected region by Lemma 20.10, we can invoke Lemma A.4 to find a component of K' that separates A' and B' in Ω. Since the interiors of tiles in V are connected subsets of Ω, this component of K' is of the form $L' = |L| \cap \Omega$ with $L \subset K$. Using Lemma 20.11 (iii) again, we see that the set L separates A and B in G. Hence $L = K$ by the minimality of K. This shows that $K' = L' = |K| \cap \Omega$ is connected, and so the same is true for the set $|K| \subset \overline{K'}$.

We will now show that $|K|$ is not contained in a k-flower. We argue by contradiction, and assume that $|K| \subset W^k(p)$ for some k-vertex p. Each of the sets $|A|$ and $|B|$ contains at least two distinct k-vertices. So we can pick k-vertices $a \in |A|$ and $b \in |B|$ with $a, b \neq p$. There

are corresponding n-tiles $X \in A$ and $Y \in B$ with $a \in X$ and $b \in Y$. Since $X, Y \in V$, each of these n-tiles must be contained in one of the k-tiles Z_i. To keep the notation in the following argument simple, let us assume that $X \subset Z_1$ and $Y \subset Z_N$ (the general case requires inessential modifications). Then $a \in \partial Z_1$ and $b \in \partial Z_N$.

By the same argument as in the proof of Lemma 11.14, we can now find a path α in the (topological) graph $G' = \partial Z_1 \cup \cdots \cup \partial Z_N$ that joins a and b and avoids p. Namely, we can join a to one of the endpoints of the k-edge $E_1 \subset Z_1 \cap Z_2$ by a (possibly degenerate) path $\alpha_1 \subset \partial Z_1$ consisting of k-edges that avoid p. Then we join the endpoint of α_1 in E_1 to one of the endpoints of E_2 by a path $\alpha_2 \subset \partial Z_2$ of k-edges that avoids p, etc. In this way, we obtain paths $\alpha_i \subset \partial Z_i$ for $i = 1, \ldots, N$ whose concatenation gives a path α of k-edges in G' that joins a and b and does not contain p. Then α consists of k-cells that do not contain p, and so $\alpha \cap W^k(p) = \emptyset$. It follows that the compact sets α and $|K| \subset W^k(p)$ are disjoint, and so these sets have positive distance with respect to some base metric on S^2.

By a small modification we can push α inside Ω to obtain a path β in Ω that is disjoint from $|K|$ and joins A' and B'. To do this, we slightly move the initial point a of α_1 into $\text{int}(X) \subset \text{int}(Z_1) \subset \Omega$ and the other endpoint of α_1 to $\text{int}(E_1) \subset \Omega$. We can join these new endpoints by a path β_1 that follows the original path α_1 closely and stays inside $\text{int}(Z_1) \cup \text{int}(E_1) \subset \Omega$. We slightly move the last point of α_2 into $\text{int}(E_2)$, and join the endpoint of β_1 to this point by a path β_2 that follows α_2 closely and stays inside $\text{int}(E_1) \cup \text{int}(Z_2) \cup \text{int}(E_2) \subset \Omega$. Continuing in this way, we get a collection of paths $\beta_1, \ldots, \beta_N$ whose concatenation is a path β in Ω that joins A' and B' and stays so close to the original path α that $\beta \cap |K| = \emptyset$.

This contradicts the fact that $K' \subset |K|$ separates A' and B' in Ω. So $|K|$ is not contained in a k-flower.

Inequality (20.6) now easily follows. Indeed, $M = f^k(|K|)$ is a connected union of $(n - k)$-tiles. This set joins opposite sides of $\mathcal{C}$. For otherwise, M is contained in a 0-flower (Lemma 5.33) which in turn implies by Lemma 5.29 (iii) that the connected set $|K| \subset f^{-k}(M)$ is contained in a k-flower; but we have just seen that this is not the case. Since M joins opposite sides of $\mathcal{C}$, the number of $(n - k)$-tiles in this set must be at least $D_{n-k}(f, \mathcal{C})$. Since the set K contains at least as many n-tiles as M contains $(n - k)$-tiles, inequality (20.6) follows. $\square$

20.3. Short e-chains

After the preparation in the previous section, we can now prove Proposition 20.1 and complete the proof of Lemma 20.6 (and hence of Theorem 20.2). In this section we will denote the length of an e-chain or a tile chain P by $\#P$ instead of $\text{length}(P)$.

PROOF OF PROPOSITION 20.1. We first assume that $\mathcal{C}$ is f-invariant.

Now we apply Lemma 20.12 in the following setting. Let $k = 0$ and our e-chain of 0-tiles consist of $Z_1 = X_{\text{w}}^0$ and $Z_2 = X_{\text{b}}^0$, i.e., the two 0-tiles attached along some 0-edge E_1 (which is necessarily on the boundary of both Z_1 and Z_2). Then Ω is equal to S^2 with the union of the 0-edges distinct from E_1 removed. We can pick 0-edges $E_0 \subset Z_1$ and $E_2 \subset Z_2$, so that E_0, E_1, E_2 are distinct. Moreover, if $\# \text{post}(f) \geq 4$, we may assume that $E_0 \cap E_2 = \emptyset$.

Let $n \in \mathbb{N}_0$ and let G be the graph as defined before Lemma 20.11. In the present situation, the vertex set V actually is equal to the set of all n-tiles. So $\#V = 2 \deg(f)^n$. Let $A \subset V$ be the set of all n-tiles that are contained in Z_1 and meet E_0, and $B \subset V$ be the set of all n-tiles that are contained in Z_2 and meet E_2. Both $|A|$ and $|B|$ contain two distinct 0-vertices, namely the endpoints of E_0 and E_2, respectively. Moreover, $A' = |A| \cap \Omega$ and $B' = |B| \cap \Omega$ are connected as follows from Lemma 20.11 (ii) applied to the path γ given by parametrizations of the arcs E_0 and E_2, respectively.

If $K \subset V$ separates A and B in G, then $\#K \geq D_n(f, \mathcal{C})$ by Lemma 20.12. So by Theorem 20.9 (i.e., Menger's theorem) there are at least $D_n(f, \mathcal{C})$ pairwise disjoint A-B-paths in G. Then one of them must have cardinality $\leq \#V/D_n(f, \mathcal{C})$. This path gives an e-chain P consisting of n-tiles $X_1, \ldots, X_M$ with $X_1 \in A$, $X_M \in B$, and

$$\#P = M \leq \#V/D_n(f, \mathcal{C}) = 2 \deg(f)^n/D_n(f, \mathcal{C}).$$

Then X_1 meets E_0 and X_M meets E_2. If $\# \text{post}(f) \geq 4$, then E_0 and E_2 are disjoint 0-edges which implies that $|P|$ joins opposite sides of $\mathcal{C}$. Hence $M \geq D_n(f, \mathcal{C})$, and so

$$D_n(f, \mathcal{C}) \leq M \leq 2 \deg(f)^n/D_n(f, \mathcal{C}),$$

which gives

(20.7) $$D_n(f, \mathcal{C}) \leq \sqrt{2} \deg(f)^{n/2}.$$

This is an upper bound for $D_n(f, \mathcal{C})$ as desired.

If $\# \operatorname{post}(f) = 3$, the argument is along similar lines, but slightly more subtle. In this case, E_0, E_1, E_2 are the three 0-edges. The underlying set $|P|$ of our e-chain P meets E_0 and E_2. We want to show that it also meets E_1. To see this, we choose a path γ that starts in an interior point of X_1, ends in an interior point X_M, and stays inside $|P| \cap \Omega$. This is possible, since P forms an e-chain where successive tiles have a common n-edge e whose interior is contained in Ω.

Now E_1 splits our simply connected region Ω into the complementary components $\operatorname{int}(Z_1)$ and $\operatorname{int}(Z_2)$. Since γ stays in Ω, starts in $\operatorname{int}(Z_1)$, and ends in $\operatorname{int}(Z_2)$, it must meet E_1. Hence $|P| \supset \gamma$ also meets E_1. So $|P|$ meets all three 0-edges, which means that this set joins opposite sides of $\mathcal{C}$ (see Definition 5.32). Again we have $M \geq D_n(f, \mathcal{C})$ and derive (20.7). This completes the proof of the statement when $\mathcal{C}$ is invariant.

We now consider the general case when $\mathcal{C}$ is not necessarily f-invariant. Then we can find an iterate $F = f^N$ of f that has an F-invariant Jordan curve $\mathcal{C}' \subset S^2$ with $\operatorname{post}(F) = \operatorname{post}(f) \subset \mathcal{C}'$ (see Theorem 15.1). Then by the first part of the proof

$$D_k(F, \mathcal{C}') \lesssim \deg(F)^{k/2} = \deg(f)^{kN/2}.$$

Since the k-tiles for $(F, \mathcal{C}')$ are the (kN)-tiles for $(f, \mathcal{C}')$ (see Proposition 5.17 (vii)), this means

$$D_{kN}(f, \mathcal{C}') \lesssim \deg(f)^{kN/2}.$$

If $n \in \mathbb{N}_0$ is arbitrary, we can write it as $n = kN + l$, where $k \in \mathbb{N}_0$ and $l \in \{0, \dots, N-1\}$. Now we know by (16.2) in Lemma 16.5 that

$$D_{n+1}(f, \mathcal{C}') \lesssim D_n(f, \mathcal{C}').$$

Applying this inequality at most $(N-1)$-times, we are led to

$$\begin{aligned} D_n(f, \mathcal{C}') &= D_{kN+l}(f, \mathcal{C}') \lesssim D_{kN}(f, \mathcal{C}') \\ &\lesssim \deg(f)^{kN/2} \leq \deg(f)^{n/2}. \end{aligned}$$

Moreover, by (16.3) in Lemma 16.5 we know that

$$D_n(f, \mathcal{C}) \asymp D_n(f, \mathcal{C}').$$

So

$$D_n(f, \mathcal{C}) \asymp D_n(f, \mathcal{C}') \lesssim \deg(f)^{n/2}.$$

Since in all these inequalities the implicit multiplicative constants are independent of n (or of k as in some of the previous inequalities), the statement follows. $\qquad \square$

In the following two lemmas we make the assumption that $f\colon S^2 \to S^2$ is an expanding Thurston map with an f-invariant Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}$ that satisfies condition (20.2) in Theorem 20.2. We allow the possibility that f has periodic critical points. Tiles in these statements will be for $(f, \mathcal{C})$.

LEMMA 20.13. *Let $n, k \in \mathbb{N}$, $n \geq k$, and suppose that X^k and Y^k are two k-tiles that are both contained in a $(k-1)$-tile U^{k-1}.*

Then there exists an e-chain P consisting of n-tiles that starts in an n-tile contained in X^k, ends in an n-tile contained in Y^k, and satisfies

$$(20.8) \qquad\qquad \#P \leq c' \deg(f)^{(n-k)/2},$$

where $c' > 0$ only depends on f and $\mathcal{C}$.

The lemma essentially says that under our assumptions, k-tiles with a common parent can be joined by an e-chain of n-tiles with controlled length.

PROOF. Since $f^{k-1}|U^{k-1}$ is a homeomorphism that maps k-tiles to 1-tiles, the number of k-tiles contained in U^{k-1} is bounded by $N_0 := 2 \deg(f)$, the number of 1-tiles.

We can pick a path $\gamma \subset \mathrm{int}(U^{k-1})$ that joins X^k and Y^k. Then γ only meets k-tiles contained in U^{k-1}. By Lemma 20.11 (i) this implies that there exists a simple e-chain consisting of k-tiles

$$Z_1 = X^k, Z_2, \dots, Z_N = Y^k$$

with $N \leq N_0$.

Let $n \geq k$ be arbitrary. Then we are again in the situation of Lemma 20.11. Let G be the graph as defined before this lemma for $Z_1, \dots, Z_N$. Its vertex set V is given by the set of all n-tiles contained in any of the tiles Z_i. Let $A \subset V$ and $B \subset V$ consist of all n-tiles contained in $Z_1 = X^k$ and $Z_N = Y^k$, respectively. Then $|A| = X^k$ and $|B| = Y^k$ contain the k-vertices on the boundary of X^k and Y^k, respectively, and hence both contain at least two distinct k-vertices. Moreover, if we define Ω as in (20.5), and $A' = |A| \cap \Omega$ and $B' = |B| \cap \Omega$, then $\mathrm{int}(X^k) \subset A' \subset X^k$ and $\mathrm{int}(Y^k) \subset B' \subset Y^k$. This implies that A' and B' are connected.

So Lemma 20.12 applies and shows in combination with (Menger's) Theorem 20.9 that the number of disjoint A-B-paths in G is bounded below by $D_{n-k}(f, \mathcal{C})$. Let P be such a path of minimal length. Note that P forms an e-chain of n-tiles whose first n-tile is contained in X^k and its last in Y^k. The number $\#V$ of vertices in G is equal to the number of n-tiles contained in any of the k-tiles Z_i, and hence bounded
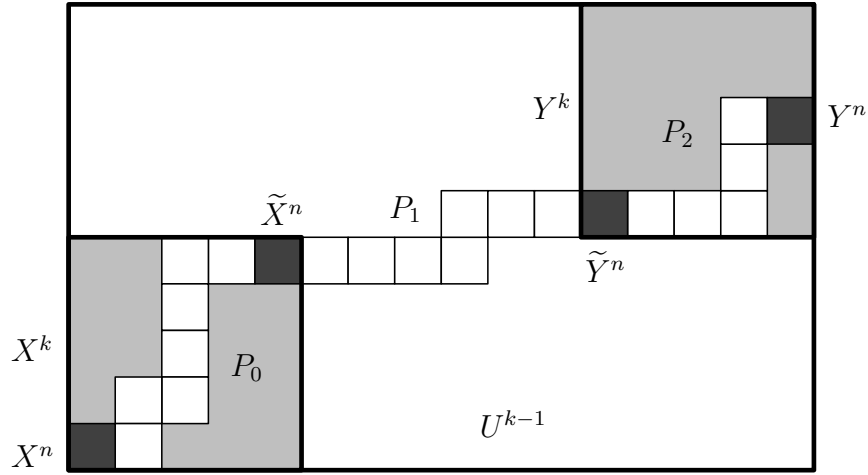
FIGURE 20.2. Connecting tiles by short e-chains.

by $2N_0 \deg(f)^{n-k}$. It follows that

$$\#P \cdot D_{n-k}(f,\mathcal{C}) \leq \#V \leq 2N_0 \deg(f)^{n-k}.$$

Using the lower bound (20.2) we conclude

$$\#P \leq c' \deg(f)^{(n-k)/2},$$

where $c' > 0$ only depends on f and $\mathcal{C}$. $\qquad\square$

LEMMA 20.14. *Let $n, k \in \mathbb{N}_0$, $n \geq k$, X^n and Y^n be two n-tiles that are both contained in a k-tile U^k. Then there exists an e-chain P of n-tiles that starts in X^n, ends in Y^n, and satisfies*

$$(20.9) \qquad\qquad \#P \leq c' \sum_{i=0}^{n-k} 2^{n-k-i} \deg(f)^{i/2},$$

where $c' \geq 1$ is a constant only depending on f and $\mathcal{C}$.

As we will see in the proof, we can take the same constant c' in (20.9) as in (20.8) if $c' \geq 1$ as we may assume.

PROOF. We prove this for fixed $n \in \mathbb{N}$ (and arbitrary tiles X^n and Y^n) by downward induction on $k = n, n-1, \ldots, 0$.

The statement is true for $k = n$. Indeed, in this case $X^n = Y^n = U^k$, and so the single tile $X^n = Y^n$ forms a suitable chain P. We have $\#P \leq 1$ which gives a bound as in (20.9) for $n = k$ if $c' \geq 1$.

Now we assume that the statement is true for some number $0 < k \leq n$. We need to show that it is also true for $k - 1$. To see this, suppose that we have n-tiles X^n and Y^n, and a $(k-1)$-tile U^{k-1} with

$X^n, Y^n \subset U^{k-1}$. Then there exist unique k-tiles X^k and Y^k with $X^n \subset X^k \subset U^{k-1}$ and $Y^n \subset Y^k \subset U^{k-1}$. See Figure 20.2 for illustration.

By Lemma 20.13 there exists an e-chain P_1 of n-tiles with

$$\#P_1 \le c' \deg(f)^{(n-k)/2}$$

that starts in an n-tile $\widetilde{X}^n \subset X^k$ and ends in an n-tile $\widetilde{Y}^n \subset Y^k$. We can now apply the induction hypothesis to the n-tiles $X^n, \widetilde{X}^n \subset X^k$ to find an e-chain P_0 of n-tiles that starts in X^n, ends in $\widetilde{X}^n$, and satisfies

$$\#P_0 \le c' \sum_{i=0}^{n-k} 2^{n-k-i} \deg(f)^{i/2}.$$

Similarly, we can find an e-chain P_2 of n-tiles that starts in $\widetilde{Y}^n$, ends in Y^n, and satisfies

$$\#P_2 \le c' \sum_{i=0}^{n-k} 2^{n-k-i} \deg(f)^{i/2}.$$

Concatenating P_0, P_1, P_2, leads to an e-chain P of n-tiles that starts in X^n, ends in Y^n, and satisfies

$$\begin{aligned}
\#P &\le \#P_0 + \#P_1 + \#P_2 \\
&\le 2c' \sum_{i=0}^{n-k} 2^{n-k-i} \deg(f)^{i/2} + c' \deg(f)^{(n-k)/2} \\
&\le c' \sum_{i=0}^{n-k+1} 2^{n-k+1-i} \deg(f)^{i/2}.
\end{aligned}$$

The statement follows. $\qquad\square$

If we assume $\deg(f)^{1/2} > 2$, the estimate (20.9) can be simplified and gives an upper bound as needed for the proof of Lemma 20.6. This is the reason why this assumption was made in the lemma.

COROLLARY 20.15. *Suppose in Lemma 20.14 we make the additional assumption that* $\deg(f)^{1/2} > 2$. *Then the e-chain P in (20.9) satisfies*

$$\#P \le C \deg(f)^{(n-k)/2},$$

where C is a constant depending only on $\mathcal{C}$ and f.

PROOF. If $\deg(f)^{1/2} > 2$, then the right hand side in (20.9) is a geometric sum with terms that increase with i; so up to a constant this sum is dominated by its last term. $\qquad\square$

We are now ready to prove Lemma 20.6.

PROOF OF LEMMA 20.6. Let $f\colon S^2 \to S^2$ be an expanding Thurston map with $\deg(f)^{1/2} > 2$ that has an invariant Jordan curve $\mathcal{C} \subset S^2$ with $\operatorname{post}(f) \subset \mathcal{C}$ and satisfies (20.2). Set $\Lambda := \deg(f)^{1/2}$.

Let $x, y \in S^2$ be distinct. In Lemma 20.8 we already saw that $N_n(x,y) \gtrsim \Lambda^{n-m(x,y)}$ for $n \geq m(x,y) + 1$. This is the desired lower bound.

To obtain the upper bound, set $k := m(x,y) \in \mathbb{N}_0$. Then by definition of $k = m(x,y)$ there exist k-tiles X^k and Y^k with $x \in X^k$, $y \in Y^k$, and $X^k \cap Y^k \neq \emptyset$. For each $n \geq k$ we can pick n-tiles $X^n, \widetilde{X}^n, \widetilde{Y}^n, Y^n$ with $X^n, \widetilde{X}^n \subset X^k$, and $Y^n, \widetilde{Y}^n \subset Y^k$, as well as

$$x \in X^n, y \in Y^n, \text{ and } \widetilde{X}^n \cap \widetilde{Y}^n \neq \emptyset.$$

Then by Corollary 20.15 there exists an e-chain P_1 of n-tiles with $\#P_1 \lesssim \Lambda^{n-k}$ whose first tile is X^n and whose last tile is $\widetilde{X}^n$. Similarly, there exists an e-chain P_2 of n-tiles with $\#P_2 \lesssim \Lambda^{n-k}$ whose first tile is $\widetilde{Y}^n$ and whose last tile is Y^n. Since $\widetilde{X}^n \cap \widetilde{Y}^n \neq \emptyset$, the union $P = P_1 \cup P_2$ is a chain of n-tiles (not necessarily an e-chain) with $\#P \lesssim \Lambda^{n-k}$ whose first tile is X^n and whose last tile is Y^n.

This implies that

$$N_n(x,y) \leq \#P \lesssim \Lambda^{n-k} = \Lambda^{n-m(x,y)}$$

for $n \geq k$. This is the required upper bound, finishing the proof. $\square$

By establishing Lemma 20.6 we have also completed the proof of Theorem 20.2 (see the end of Section 20.1).

CHAPTER 21

Outlook and open problems

In this final chapter we give an outlook on further results that are related to the major themes in this book. We do not try to be exhaustive, but rather intend the discussion as a first entry point into various other investigations. We combine this with a presentation of some open problems that will hopefully stimulate future research.

Markov partitions for Thurston maps. The basis of our combinatorial approach to the study of an expanding Thurston map f is to consider the cell decompositions induced by a Jordan curve containing the postcritical points of f as in Chapter 5. In the case when this Jordan curve is f-invariant we obtain a Markov partition and can describe our map in a combinatorial fashion by a two-tile subdivision rule (see Section 12.2). Recall that Theorem 15.1 ensures the existence of such an f^n-invariant Jordan curve for each sufficiently large $n \in \mathbb{N}$.

We know (see Example 15.11) that an expanding Thurston map f itself does not necessarily have an f-invariant Jordan curve and that in general one has to pass to a suitable iterate in order to obtain such a curve. This naturally leads to the question whether one can bound the order of this iterate in terms of some natural invariants of the map.

PROBLEM 1. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map. Is there a number $N_0 \in \mathbb{N}$, depending on some natural data such as $\deg(f)$, $\#\operatorname{post}(f)$, and $\Lambda_0(f)$ such that for all $n \geq N_0$ there exists an f^n-invariant Jordan curve $\mathcal{C} \subset S^2$ with $\operatorname{post}(f) \subset \mathcal{C}$?*

Recall that $\Lambda_0(f)$ denotes the combinatorial expansion factor of f (see Chapter 16).

If $f\colon S^2 \to S^2$ is an expanding Thurston map and $\mathcal{C} \subset S^2$ an f^n-invariant Jordan curve with $\operatorname{post}(f) \subset \mathcal{C}$, then the set

$$\mathcal{C}_n := \mathcal{C} \cup f^{-1}(\mathcal{C}) \cup \cdots \cup f^{-n+1}(\mathcal{C})$$

is easily seen to be f-invariant. This gives some type of Markov partition for f, but it will not be cellular as defined in Section 5.1. In particular, $S^2 \setminus \mathcal{C}_n$ may have infinitely many components and in general we have very little control over the geometric shapes of the "tiles" in this partition.

PROBLEM 2. *Does every expanding Thurston map $f: S^2 \to S^2$ have a (finite) cellular Markov partition?*

We conjecture that the answer should be affirmative. To prove this statement, one essentially has to construct a finite connected graph $G \subset S^2$ with $\mathrm{post}(f) \subset G$ that is f-invariant. A different way to phrase the problem is to ask whether f (and not some iterate f^n) can be described by a (suitably defined) *k-tile subdivision rule*. Here k would be the number of components of $S^2 \setminus G$.

An interesting special case is when the invariant graph is a tree T and so $S^2 \setminus T$ is connected (then we have a 1-tile subdivision rule).

Cannon-Floyd-Parry [**CFP10**, Theorem 3.1] showed that such an invariant tree exists for each sufficiently high iterate of a Lattès map with signature $(2, 2, 2, 2)$. They also observed [**CFP10**, Section 4] that it is indeed necessary to take an iterate here, because they found examples of such Lattès maps f for which no f-invariant tree T with $\mathrm{post}(f) \subset T$ exists.

Farrell and Jones [**FJ79**] constructed finite cellular Markov partitions in a more general context, but they also had to pass to sufficiently high iterates of the maps considered to guarantee existence of the Markov partition (their definition of this concept differs from ours; for yet another definition of a Markov partition, see [**PU10**, Definition 4.5.1]).

Recall from Chapter 9 that each expanding Thurston map f can be described as a factor of the left shift on the space of finite words with $d = \deg(f)$ elements. This can be used to obtain a Markov partition for f, but again we will have very little geometric control for the geometry of corresponding "tiles".

Markov partitions for Lattès maps and how they behave under certain perturbations were discussed in [**Be94**]. Related results can also be found in [**Re15**].

Another way to find Markov partitions is via the iterated monodromy group (the concept of a *limit space* is relevant here; see [**Ne05**, Chapter 3]) or by using invariant Peano curves (see the discussion below). These methods give "tiles" with a complicated geometric structure.

In this work we mostly considered Thurston maps that are expanding. One may ask whether combinatorial descriptions as for these maps exist for more general types of maps.

PROBLEM 3. *Let $f: S^2 \to S^2$ be a Thurston map (not necessarily expanding). Is there a Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}$ that is invariant for some iterate f^n? Are there other natural decompositions*

of the sphere S^2 that are invariant (in a suitable sense) under f or some iterate f^n?

Related to this is a variant of Theorem 15.1 proved in [**G–Z**]. Namely, let $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a rational Thurston map whose Julia set is a Sierpiński carpet. Then for each sufficiently large $n \in \mathbb{N}$ there is a Jordan curve $\mathcal{C} \subset \widehat{\mathbb{C}}$ with $\mathrm{post}(f) \subset \mathcal{C}$ that is invariant for f^n.

When f is a polynomial, natural decompositions of $\widehat{\mathbb{C}}$ can be obtained from *external rays* (see [**DH84**]) and the related *Yoccoz puzzles* (see [**Hu93**] and [**Mi00**]). A lack of similar combinatorial methods for general rational maps is one of the main reasons why their study is harder than the study of polynomials.

Another question is whether and how our results extend to maps that are not necessarily postcritically-finite. The theory of *coarse expanding dynamical systems* developed in [**HP09**] should be relevant here.

PROBLEM 4. *Let $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a rational map (not necessarily post-critically-finite) whose Julia set is the whole Riemann sphere $\widehat{\mathbb{C}}$. Does there exist a natural combinatorial description of f or some iterate f^n?*

The rational maps $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ of a given degree $d \geq 2$ form a complex manifold $\mathcal{R}_d$ of dimension $2d + 1$. Rees showed [**Re86**] that the set of points in $\mathcal{R}_d$ where the corresponding rational map f has a Julia set equal to the whole sphere has positive measure with respect to the natural measure class on $\mathcal{R}_d$. Such points in $\mathcal{R}_d$ and the corresponding maps can be obtained by a slight perturbation of certain expanding Thurston maps. It would be interesting to find combinatorial descriptions of expanding Thurston maps that change under such perturbations in a controlled manner.

Instead of asking whether good combinatorial models exist for (expanding) Thurston maps, one may ask for good analytical models. As we have seen, the dynamics of an expanding Thurston map f generates a class of visual metrics, and so a fractal geometry on the underlying 2-sphere. This does not rule out that the map f can actually be described by a smooth model. Li showed [**Li16b**] that no expanding Thurston map with periodic critical points is conjugate to a smooth map, but the following problem is still open.

PROBLEM 5. *Is every expanding Thurston map with no periodic critical points topologically conjugate to a smooth expanding Thurston map on $\widehat{\mathbb{C}}$?*

This question was raised by K. Pilgrim. We expect this to be true for at least every sufficiently high iterate of the given map.

If we assume that a Thurston map is given by a two-tile subdivision rule or in some other combinatorial way, then one wants to know which information about the map can be extracted from this combinatorial description. This general question is a major theme in the study of the dynamics of polynomials pioneered by Douady and Hubbard (see [**DH84**]).

We know that every two-tile subdivision rule is realized by a Thurston map that is unique up to Thurston equivalence. In contrast, a Thurston map may be realized by combinatorially different two-tile subdivision rules.

PROBLEM 6. *Suppose two Thurston maps f and g realize different two-tile subdivision rules. How can one decide from combinatorial data whether the maps are Thurston equivalent?*

Of course, there are several simple necessary conditions, such as $\deg(f) = \deg(g)$ and $\#\operatorname{post}(f) = \#\operatorname{post}(g)$, whose validity can easily be verified from the subdivision rules. In addition, the maps must have the same ramification portrait (see Section 2.2). A related question, namely when polynomials with the same critical portrait are Thurston equivalent, was answered in [**BN06**]. The major tool used there was the iterated monodromy group as discussed below. Closely related is the biset associated with a Thurston map, which may also be used to study Thurston equivalence (see in particular [**BD**]). The questions of whether two Thurston maps are equivalent or whether a Thurston map is equivalent to a rational map are decidable (see [**BBY12**]).

PROBLEM 7. *Let f be an expanding Thurston map that realizes a two-tile subdivision rule. Is there an effective way to compute the combinatorial expansion factor $\Lambda_0(f)$ from the combinatorial description?*

Since $\Lambda_0(f)$ is defined as a limit, a priori one cannot expect to find $\Lambda_0(f)$ by a finite procedure. However, if f is a Lattès or Lattès-type map, then $\Lambda_0(f)$ is the smallest absolute value of the two eigenvalues of the matrix describing the underlying torus endomorphism [**Yi15a**]. In general, one may speculate that if f realizes a two-tile subdivision rule with underlying cell decompositions $\mathcal{D}$ and $\mathcal{D}'$, then $\Lambda_0(f)$ is related to the spectral radius of a matrix that is obtained from the incidence relations of the cells in $\mathcal{D}'$ and their images under f in $\mathcal{D}$.

Conformal dimension of the visual sphere. Recall from Theorem 18.1 (ii) that if an expanding Thurston map f is topologically

conjugate to a rational map, then its visual sphere is a quasisphere. In other words, there is a quasisymmetry $(S^2, \varrho) \to (\widehat{\mathbb{C}}, \sigma)$, where ϱ is a visual metric for f and σ is the chordal metric on $\widehat{\mathbb{C}}$. In particular, (S^2, ϱ) is then quasisymmetrically equivalent to an Ahlfors 2-regular space, namely the Riemann sphere $(\widehat{\mathbb{C}}, \sigma)$

A closely related question is how much the Hausdorff dimension of a metric space (X, d) can be lowered by a quasisymmetric map. This can be measured by the infimum of all Hausdorff dimensions of metric spaces (X', d') that are quasisymmetrically equivalent to (X, d). Actually, often a more relevant quantity is the *(Ahlfors regular) conformal dimension* of (X, d), where one takes the corresponding infimum only over Ahlfors regular metric spaces (X', d') (see [**MT10**]). For metric spheres (S^2, d) that are not quasispheres the conformal dimension measures in a sense by how much (S^2, d) fails to be a quasisphere.

If $f : S^2 \to S^2$ is an expanding Thurston map without periodic critical points, then S^2 equipped with a visual metric ϱ is Ahlfors regular (Proposition 18.2). Since all visual metrics for f are quasisymmetrically equivalent, the conformal dimension of the visual sphere (S^2, ϱ) only depends on f and not on the choice of the visual metric ϱ. This is an important numerical invariant of the fractal geometry of (S^2, ϱ).

PROBLEM 8. *Is it possible to determine the (Ahlfors regular) conformal dimension of the visual sphere of an expanding Thurston map in terms of its dynamical data?*

Bonk-Geyer-Pilgrim [**Bo06**] stated a conjecture expressing this conformal dimension in terms of eigenvalues of certain matrices related to the dynamics of the map. One of the inequalities relating the quantities in this conjecture was established by Haïssinsky-Pilgrim [**HP08**].

In general, the infimum defining the conformal dimension of a metric space may not be attained as a minimum. For metric spheres that arise as boundaries at infinity of Gromov hyperbolic groups the following related result was proved in [**BK05**, Theorem 1.1].

THEOREM 21.1. *Let G be a Gromov hyperbolic group whose boundary at infinity $\partial_\infty G$ is homeomorphic to a 2-sphere. If the conformal dimension Q of $\partial_\infty G$ (equipped with a visual metric) is attained as a minimum, then $Q = 2$ and $\partial_\infty G$ is a quasisphere.*

As we discussed in Section 4.3, then there exists an action of G on hyperbolic 3-space $\mathbb{H}^3$ that is geometric, i.e., isometric, properly discontinuous, and cocompact.

A corresponding result for expanding Thurston maps was established in [**HP14**].

THEOREM 21.2. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map without periodic critical points and suppose the conformal dimension Q of its visual sphere is attained as a minimum. Then either $Q = 2$ and f is topologically conjugate to a rational map, or f is a Lattès-type map with signature $(2,2,2,2)$ and the map L_A in (3.4) has two distinct real eigenvalues > 1 (in which case $Q > 2$).*

Haïssinsky and Pilgrim actually proved their result in greater generality for *topologically coarse expanding conformal maps*. These maps cannot have periodic critical points, but are not necessarily postcritically-finite.

Equivalence to a rational map. As we already discussed in the introduction, Theorem 18.1 (ii) gives a criterion for an expanding Thurston map to be topologically conjugate to a rational map quite different from Thurston's theorem (Theorem 2.17). The latter theorem is proved by methods fundamentally different from our combinatorial approach: one considers a suitable Teichmüller space $\mathcal{T}$ and studies a map $f^*\colon \mathcal{T} \to \mathcal{T}$ (the *Thurston pull-back map*) induced by the given Thurston map f (with hyperbolic orbifold). Then f is equivalent to a rational map if and only if the induced map f^* has a fixed point in $\mathcal{T}$ [**DH93**]. It is very desirable to reconcile these points of view.

PROBLEM 9. *Is it possible to give a proof of Thurston's theorem that does not use Teichmüller theory and is based on a combinatorial description of the given map?*

As we know, such combinatorial descriptions are given, for example, by two-tile subdivision rules.

PROBLEM 10. *Suppose an expanding Thurston map f realizes a two-tile subdivision rule. Is it possible to decide from the subdivision rule whether f is Thurston equivalent to a rational map?*

Of course, one should interpret this as asking for a criterion that is easier to check in practice than the one provided by Thurston's theorem.

An additional motivation for considering these problems is that by Theorem 18.1 (ii) they are closely related to the question of characterizing quasispheres. So their solution may give new ideas that could also be used in the group setting and possibly be applied for progress on Cannon's conjecture (see Section 4.3).

Thurston's theorem (Theorem 2.17) is most useful as a negative criterion, allowing one to decide when a Thurston map is not equivalent to a rational map by finding a Thurston obstruction. For an

obstructed Thurston map, i.e., a Thurston map that is not equivalent to a rational map, in general there may be several different Thurston obstructions. Pilgrim showed that in this case one can single out a "canonical" Thurston obstruction [**Pi01**].

For a Thurston map for which each cycle in the postcritical set contains a critical point, a sufficient criterion for the map to be equivalent to a rational map was established by Dylan Thurston in joint work with Kahn and Pilgrim (see [**Th16**] and [**KPT**]).

Special classes of maps. Since Lattès maps form the best understood subclass of Thurston maps, it is natural to investigate maps that are closely related to Lattès maps. One such class, the nearly Euclidean Thurston maps, was introduced and studied in [**C–P12**]. By definition a Thurston map $f\colon S^2 \to S^2$ is called *nearly Euclidean* if it has exactly four postcritical points and every critical point has local degree equal to 2. These maps may have a hyperbolic orbifold, or periodic critical points.

One can construct such maps as follows. Let $g\colon S^2 \to S^2$ be a Lattès-type map with signature $(2, 2, 2, 2)$ and $h\colon S^2 \to S^2$ be an orientation-preserving homeomorphism with $h(\operatorname{post}(f)) \subset f^{-1}(\operatorname{post}(f))$. Then $f = h \circ g$ is a nearly Euclidean Thurston map.

In general, it is difficult to use Thurston's theorem and check whether a given Thurston map (with hyperbolic orbifold) is equivalent to a rational map, because the map may have infinitely many invariant multicurves each of which could be a Thurston obstruction. So potentially one has to verify infinitely many conditions. For nearly Euclidean Thurston maps, however, it is possible to give an explicit algorithm that decides whether the map is equivalent to a rational map.

A *Belyǐ map* is a holomorphic map $F\colon S \to \widehat{\mathbb{C}}$, defined on a compact Riemann surface S, that is ramified over three points. This means that F has exactly three critical values which are usually taken to be $\{0, 1, \infty\}$. In this case, the set $G := F^{-1}([0, 1]) \subset S$ is a topological graph with vertex set $F^{-1}(\{0, 1\})$. Its edges are the closures of the components of $F^{-1}((0, 1))$. The graph G is bipartite if one distinguishes the vertices in $F^{-1}(0)$ and $F^{-1}(1)$. They are often marked by black and white dots, respectively. The resulting diagram is called the *dessin d'enfant* of F (introduced by Grothendieck in [**Gro97**]). It determines F up to pre- and postcomposition by conformal maps. In particular, it defines S up to conformal equivalence. Belyǐ's theorem says that each non-singular algebraic curve defined over the field $\overline{\mathbb{Q}}$ of algebraic numbers can be represented by a Belyǐ map. The *absolute Galois group* $\operatorname{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ (i.e., the group of field automorphisms of $\overline{\mathbb{Q}}$ that fix $\mathbb{Q}$) acts

on these algebraic curves, and so on the set of dessins d'enfants. See [**LZ04**] for an introduction to this subject.

Let $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a Belyĭ map defined on $\widehat{\mathbb{C}}$ whose set of critical values is given by $\{0, 1, \infty\}$. In general, f is not a dynamical object, because the iterates of f will not be Belyĭ maps in general. This is true if $\{0, 1, \infty\} \subset f^{-1}(\{0, 1, \infty\})$. Note that this can always be achieved by precomposing f by a suitable Möbius transformation. In this case, f is a rational Thurston map with $\mathrm{post}(f) = \{0, 1, \infty\}$. The representation of such a map f by its dessin d'enfant is closely related to our description of f by cell decompositions. In fact, if we choose $\mathcal{C} = \widehat{\mathbb{R}}$ then the 1-skeleton $f^{-1}(\mathcal{C}) = f^{-1}(\widehat{\mathbb{R}}) \supseteq f^{-1}([0, 1])$ of $\mathcal{D}^1(f, \mathcal{C})$ contains the dessin d'enfant of f.

The relation between Hubbard trees and dessins d'enfants for polynomials P (with $\mathrm{post}(P) = \{0, 1, \infty\}$ equal to the set of critical values of P) was investigated in [**Pi00**].

Invariant Peano curves and mating of polynomials. If we denote by $\mathbb{S}^1 = \partial \mathbb{D}$ the unit circle in $\mathbb{C}$ and by S^2 a 2-sphere (as before), then a *Peano curve* in S^2 is a continuous and surjective map $\gamma\colon \mathbb{S}^1 \to S^2$. The following result was established in [**Me13**].

THEOREM 21.3. *Let* $f\colon S^2 \to S^2$ *be an expanding Thurston map. Then for each sufficiently high iterate* $F = f^n$ *there is a Peano curve* $\gamma\colon \mathbb{S}^1 \to S^2$ *such that* $F(\gamma(z)) = \gamma(z^d)$ *for all* $z \in \mathbb{S}^1$, *where* $d = \deg(F)$.

This means that the following diagram commutes:

$$
\begin{array}{ccc}
\mathbb{S}^1 & \xrightarrow{\ z \mapsto z^d\ } & \mathbb{S}^1 \\
{\scriptstyle \gamma}\big\downarrow & & \big\downarrow{\scriptstyle \gamma} \\
S^2 & \xrightarrow[\ F\ ]{} & S^2.
\end{array}
$$

A Peano curve γ as in this diagram is called *F-invariant*.

On a more intuitive level, Theorem 21.3 can be phrased as follows: if we wrap $\mathbb{S}^1$ around itself d times, then we obtain the map F through the parametrization of S^2 by the Peano curve γ.

One can actually say more about the Peano curve γ in Theorem 21.3; namely, if we identify S^2 and $\widehat{\mathbb{C}}$ so that $\mathbb{S}^1 \subset \widehat{\mathbb{C}} \cong S^2$, then there exists a *pseudo-isotopy* $H\colon \mathbb{S}^1 \times [0, 1] \to S^2$ (i.e., a homotopy such that H_t is an embedding for $t \in [0, 1)$) with $H_0 = \mathrm{id}_{\mathbb{S}^1}$ and $H_1 = \gamma$.

The construction of the invariant Peano curve γ in Theorem 21.3 is very similar to the iterative construction of invariant Jordan curves in Section 15.2.

According to "Sullivan's dictionary" there is a close correspondence between the dynamics of rational maps and of Kleinian groups [**Su85**]. In [**CT07**] Cannon-Thurston constructed Peano curves related to the fundamental group of a hyperbolic 3-manifold M^3 that fibers over the circle. Theorem 21.3 may be viewed as the corresponding result in the case of rational maps. This provides another entry in Sullivan's dictionary.

There is a converse to Theorem 21.3 (see [**Me13**]); namely, if for a Thurston map $f\colon S^2 \to S^2$ there exists an iterate $F = f^n$ that has an F-invariant Peano curve, then f is expanding.

In the early 1980s Douady and Hubbard observed that there are rational maps with Julia sets that "contain" the Julia sets of some polynomials. This motivated them to introduce the notion of a *mating of polynomials*. This operation combines two polynomials geometrically, often giving a rational map. In fact, Thurston's characterization of rational maps (Theorem 2.17) was in part motivated by the question when a map arising as a mating "is" a rational map. The notion of Thurston equivalence also appears naturally in this context.

There are many different variants of matings. Here we will only define the one most relevant for us. An introduction to matings can be found in [**Mi04**] and an overview of the different constructions in [**MP12**].

Let $P(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_0$ be a monic polynomial with complex coefficients. Then the *filled Julia set* $\mathcal{K} = \mathcal{K}_P$ of P is the set of all points $z \in \mathbb{C}$ with bounded orbit $\{P^n(z)\}_{n\in\mathbb{N}}$. We assume that $\mathcal{K}$ is connected and locally connected. Then there is a conformal map $\phi\colon \widehat{\mathbb{C}} \setminus \overline{\mathbb{D}} \to \widehat{\mathbb{C}} \setminus \mathcal{K}$ that satisfies $\phi(z^n) = P(\phi(z))$ for all $z \in \widehat{\mathbb{C}} \setminus \overline{\mathbb{D}}$ (this is Böttcher's theorem; see [**Mi06a**, Section 9] or [**CG93**, Section II.4]). By Carathéodory's theorem (see, for example, [**Mi06a**, Theorem 17.14]) the map ϕ extends to the unit circle $\mathbb{S}^1 = \partial\mathbb{D}$. We call the restriction $\sigma\colon \mathbb{S}^1 \to \mathcal{K}$ of this extension to the unit circle the *Carathéodory loop*. Then $\sigma(\mathbb{S}^1) = \partial\mathcal{K} = \mathcal{J}$ is the Julia set of P.

Since σ is the extension of ϕ, it follows that $\sigma(z^d) = f(\sigma(z))$ for all $z \in \mathbb{S}^1$, i.e., the following diagram commutes

(21.1)
$$
\begin{array}{ccc}
\mathbb{S}^1 & \xrightarrow{\ z\mapsto z^d\ } & \mathbb{S}^1 \\
{\scriptstyle\sigma}\big\downarrow & & \big\downarrow{\scriptstyle\sigma} \\
\mathcal{J} & \xrightarrow{\ f\ } & \mathcal{J}.
\end{array}
$$

In general, the map σ is not injective and so we only obtain a *semi-conjugacy* here. The existence of this semi-conjugacy is one of the main

reasons why the dynamics of polynomials is much better understood than the dynamics of arbitrary rational maps. In particular, it can be used to describe the dynamics of a polynomial on its Julia set in combinatorial terms.

A *(topological) mating* is now defined as follows. Let $P_{\mathtt{w}}$ and $P_{\mathtt{b}}$ be two monic polynomials of the same degree $d \geq 2$ (our use of the indices $\mathtt{w}$ and $\mathtt{b}$ is motivated by the close connection to the coloring of tiles as discussed in Section 5.3). We assume that their filled Julia sets $\mathcal{K}_{\mathtt{w}}$ and $\mathcal{K}_{\mathtt{b}}$ are connected and locally connected (equivalently, one can impose these conditions on the Julia sets of the polynomials). Let $\sigma_{\mathtt{w}} \colon \mathbb{S}^1 \to \mathcal{K}_{\mathtt{w}}$ and $\sigma_{\mathtt{b}} \colon \mathbb{S}^1 \to \mathcal{K}_{\mathtt{b}}$ be the corresponding Carathéodory loops. We now consider the (topological) disjoint union $\mathcal{K}_{\mathtt{w}} \sqcup \mathcal{K}_{\mathtt{b}}$. Then a map $P_{\mathtt{w}} \sqcup P_{\mathtt{b}}$ is naturally defined on this set by letting it act on $\mathcal{K}_{\mathtt{w}}$ as $P_{\mathtt{w}}$ and on $\mathcal{K}_{\mathtt{b}}$ as $P_{\mathtt{b}}$. Then $P_{\mathtt{w}} \sqcup P_{\mathtt{b}}$ is clearly a continuous map on $\mathcal{K}_{\mathtt{w}} \sqcup \mathcal{K}_{\mathtt{b}}$.

Let $\sim$ be the equivalence relation on $\mathcal{K}_{\mathtt{w}} \sqcup \mathcal{K}_{\mathtt{b}}$ generated by (i.e., the smallest equivalence relation satisfying) the relation

$$\text{(21.2)} \qquad\qquad\qquad \sigma_{\mathtt{w}}(z) \sim \sigma_{\mathtt{b}}(\bar{z})$$

for $z \in \mathbb{S}^1 = \partial \mathbb{D}$. Then the *mating* of $\mathcal{K}_{\mathtt{w}}$ and $\mathcal{K}_{\mathtt{b}}$ is defined as $\mathcal{K}_{\mathtt{w}} \perp\!\!\!\perp \mathcal{K}_{\mathtt{b}} := \mathcal{K}_{\mathtt{w}} \sqcup \mathcal{K}_{\mathtt{b}}/\sim$. Moreover, based on Lemma A.20 it follows from (21.1) that the map $P_{\mathtt{w}} \sqcup P_{\mathtt{b}} \colon \mathcal{K}_{\mathtt{w}} \sqcup \mathcal{K}_{\mathtt{b}} \to \mathcal{K}_{\mathtt{w}} \sqcup \mathcal{K}_{\mathtt{b}}$ descends to the quotient by $\sim$, i.e., to a map

$$P_{\mathtt{w}} \perp\!\!\!\perp P_{\mathtt{b}} \colon \mathcal{K}_{\mathtt{w}} \perp\!\!\!\perp \mathcal{K}_{\mathtt{b}} \to \mathcal{K}_{\mathtt{w}} \perp\!\!\!\perp \mathcal{K}_{\mathtt{b}}.$$

This map is called the *(topological) mating* of $P_{\mathtt{w}}$ and $P_{\mathtt{b}}$.

The space $\mathcal{K}_{\mathtt{w}} \perp\!\!\!\perp \mathcal{K}_{\mathtt{b}}$ may not be a "nice" topological space; in fact, it may not even be Hausdorff. Surprisingly often though, the mating results in a map $P_{\mathtt{w}} \perp\!\!\!\perp P_{\mathtt{b}}$ that is topologically conjugate to a rational map. This is particularly striking in cases when $\mathcal{K}_{\mathtt{w}}$ and $\mathcal{K}_{\mathtt{b}}$ are *dendrites* and have no interior points.

The situation is best understood for quadratic polynomials. The following statement is currently the best result on the existence of matings.

THEOREM 21.4. *Let $P_{\mathtt{w}}(z) = z^2 + c_{\mathtt{w}}$ and $P_{\mathtt{b}}(z) = z^2 + c_{\mathtt{b}}$ be postcritically-finite quadratic polynomials such that $c_{\mathtt{w}}$ and $c_{\mathtt{b}}$ are not contained in conjugate limbs of the Mandelbrot set. Then the mating $P_{\mathtt{w}} \perp\!\!\!\perp P_{\mathtt{b}}$ is topologically conjugate to a (postcritically-finite) rational map.*

For the terminology and the proofs see [**Ta92**], [**Re92**], and [**Sh00**]. Instead of asking when two polynomials can be mated, one can also investigate when a rational map f arises as a mating of two polynomials $P_{\mathtt{w}}$ and $P_{\mathtt{b}}$. If this is the case, we say that f *unmates* into $P_{\mathtt{w}}$ and $P_{\mathtt{b}}$.

The following result about unmatings was established in [**Me09b**]. An overview of the construction, as well as several examples, can be found in [**Me14**].

THEOREM 21.5. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map without periodic critical points. Then every sufficiently high iterate $F = f^n$ is topologically conjugate to the mating of two monic polynomials $P_{\mathtt{w}}$ and $P_{\mathtt{b}}$.*

Here the polynomials $P_{\mathtt{w}}$ and $P_{\mathtt{b}}$ are postcritically-finite and have the same degree as F. Their Julia sets are dendrites.

Theorem 21.3 and Theorem 21.5 are closely related. It is not hard to see that Theorem 21.5 implies Theorem 21.3, but for the proof one actually first establishes the latter theorem and then derives the former as a consequence.

We also remark that there is a version of Theorem 21.5 for expanding Thurston maps that do have periodic critical points.

In [**Me14**] a sufficient criterion was given when an expanding Thurston map f unmates into two polynomials and an algorithm was provided to determine the polynomials. However, the criterion was not necessary.

PROBLEM 11. *Is it possible to give a necessary and sufficient condition when an expanding Thurston map unmates into two polynomials? Can the polynomials be determined by an algorithm in this case?*

The iterated monodromy group. An important question is how to decide when two given Thurston maps are equivalent. A prominent example where this is relevant and hard to decide is in the following situation. There are exactly three distinct (up to conjugacy by a Möbius transformation) quadratic polynomials whose critical point is periodic with period 3. These maps are known as the "rabbit", the "anti-rabbit", and the "airplane". If we postcompose such a map with a Dehn twist around two of its postcritical points, then this results in a Thurston map that is equivalent to one of these three maps. Deciding whether the map is equivalent to the rabbit, the anti-rabbit, or the airplane is known as the *twisted rabbit problem.* It was solved by Bartholdi and Nekrashevych [**BN06**] by using the concept of the *iterated monodromy group.* This is an important group associated with every Thurston map. It was first considered by Kameyama [**Ka03**] and systematically studied by Nekrashevych [**Ne05**] in a more general setting.

To define this group, we consider a Thurston map $f\colon S^2 \to S^2$ and a point $p \in S^2 \setminus \mathrm{post}(f)$. For each $n \in \mathbb{N}_0$ let $V_n := f^{-n}(p)$ be the

preimage set of p under f^n. Then the *preimage tree* of p with respect to f is the graph $T = (V, E)$, whose set of vertices is the disjoint union $V = \bigsqcup_{n \in \mathbb{N}_0} V_n$. Moreover, if $x \in V_n \subset V$ with $n \geq 1$, then we connect the vertices x and $f(x) \in V_{n-1}$ by an edge and all edges arise in this way. It is clear that the graph T is indeed a tree.

Now consider a loop $\gamma \subset S^2 \setminus \mathrm{post}(f)$ starting (and ending) at p. Then γ represents an element $g = [\gamma]$ in the fundamental group $G = \pi_1(S^2 \setminus \mathrm{post}(f), p)$ of $S^2 \setminus \mathrm{post}(f)$. If $x \in f^{-n}(p) = V_n \subset V$, then γ can be lifted under f^n to a path γ_x starting at x. The endpoint y of γ_x will also belong to $V_n \subset V$. By the homotopy lifting theorem ([**Ha02**, Proposition 1.30]) y depends only on the homotopy class $g = [\gamma]$ of γ. If we set $g(x) = y$, then one can show that g induces an automorphism $\varphi(g)$ of the tree T. This defines an action of G on T, and, if we denote by $\mathrm{Aut}(T)$ the automorphism group of T, we obtain a group homomorphism $\varphi \colon G \to \mathrm{Aut}(T)$. The *iterated monodromy group* $\mathrm{img}(f)$ of f is defined as the quotient of G that acts effectively on T, or more precisely,

$$\mathrm{img}(f) := G/\ker(\varphi) \cong \varphi(G),$$

where $\ker(\varphi)$ denotes the kernel of φ.

The iterated monodromy group is invariant under Thurston equivalence in the sense that two equivalent Thurston maps have the same iterated monodromy group up to isomorphism. In fact, with some additional data, it is a complete invariant for Thurston equivalence (see [**Ne05**, Theorem 6.5.2]). The solution of the twisted rabbit problem by Bartholdi and Nekrashevych was based on this fact.

Iterated monodromy groups are *self-similar groups* (see [**Ne05**]). They can be quite complicated, even for very simple maps. In general, their algebraic properties (such as torsion or amenability) are poorly understood. Here we will only discuss one particularly interesting aspect of iterated monodromy groups in more detail, namely their growth behavior. We first have to recall some definitions.

Let S be a finite and symmetric set of generators for a finitely-generated group G. For $g \in G$ let $\ell(g) = \ell_{G,S}(g)$ be the minimal length of a word in the alphabet S that represents g. This is equal to the distance of g from the neutral element e of g in the Cayley graph $\mathcal{G}(G, S)$ (recall these concepts from Section 4.3).

Let $N = N_{G,S} \colon \mathbb{N}_0 \to \mathbb{N}$ be the *growth function* of G given by

$$N(n) = \#\{g \in G : \ell(g) \leq n\} \quad \text{for } n \in \mathbb{N}_0.$$

We say that G is of *polynomial growth* if $N(n)$ is bounded from above by a polynomial in n, and of *exponential growth* if $N(n)$ is bounded

from below by an exponential function of the form $C \exp(\alpha n)$ with $C, \alpha > 0$. If G is neither of polynomial nor of exponential growth, then we say that it is of *intermediate growth*. The growth behavior of N is independent of the choice of the generating set S, and can therefore be considered as a property of the group G.

For example, free groups and fundamental groups of closed hyperbolic manifolds are of exponential growth. A celebrated theorem due to Gromov says that a group is of polynomial growth if and only if it is virtually nilpotent (see [**Gr81**] and [**Kl10**]). This answered a question raised by Milnor in [**Mi68**]. In the same note, Milnor asked whether groups of intermediate growth actually exist. First examples of such groups were later found by Grigorchuk [**Gr84**].

It is quite striking that iterated monodromy groups of very simple rational maps can be groups of intermediate growth. For example, this is the case for $\text{img}(P)$ where $P(z) = z^2 + \boldsymbol{i}$ (see [**BP06**]). Note that for this polynomial the Julia set $\mathcal{J}$ is a dendrite and each postcritical point is a leaf of $\mathcal{J}$ (i.e., it does not separate $\mathcal{J}$). While intermediate growth of the iterated monodromy group has been shown for some other polynomials, at present no general condition is known that would implies this.

If a postcritically-finite quadratic polynomial P has two distinct Fatou components whose closures intersect in a single point, then it is not hard to show that its iterated monodromy group is of exponential growth. For example, this is true for $P(z) = z^2 - 1$.

Lattès maps and Lattès-type maps have iterated monodromy groups that are virtually isomorphic to $\mathbb{Z}^2$. Apart from some special cases such as the examples discussed, very little is known in general about the growth of iterated monodromy groups of postcritically-finite polynomials, and even less for Thurston maps that are non-polynomial (i.e., not Thurston polynomials; see Section 6.2). Some examples of non-polynomial Thurston maps that have iterated monodromy groups of exponential growth were found in [**HM**].

PROBLEM 12. *Are there non-polynomial Thurston maps with iterated monodromy groups of intermediate growth?*

Ergodic theory of expanding Thurston maps. The ergodic theory of expanding Thurston maps was developed further by Zhiqiang Li. In [**Li16a**] it was shown that the measure of maximal entropy of an expanding Thurston map can be obtained as a weak*-limit of point masses at periodic points, or at preimages of any point. Similar results for rational maps had been established before by Lyubich [**Ly83**].

Li also investigated *equilibrium states* for expanding Thurston maps [**Li15b**]. These are measures obtained from Hölder continuous functions, called *potentials*. For the precise definition let $f\colon S^2 \to S^2$ be an expanding Thurston map and $\phi\colon S^2 \to \mathbb{R}$ be a Hölder continuous function. We define the *topological pressure* of ϕ with respect to f as

$$(21.3) \qquad P(\phi, f) = \sup_{\mu} \left\{ h_{\mu}(f) + \int \phi \, d\mu \right\},$$

where the supremum is taken over all f-invariant Borel probability measures on S^2 (recall that $h_{\mu}(f)$ denotes the measure-theoretic entropy of f with respect to μ; see Section 17.1).

An f-invariant measure μ_{ϕ} for which the supremum in (21.3) is attained is called an *equilibrium state*. Li showed the existence and uniqueness of equilibrium states for any Hölder continuous potential ϕ. These measures can also be described as weak*-limits of suitably weighted point masses at periodic points or preimage points of a given point.

Bowen [**Bo72**] introduced the concept of an *h-expanding* map and Misiurewicz [**Mi76**] the weaker notion of an *asymptotically h-expanding* map. Roughly speaking, these notions mean that the map is expanding in a strong sense except on a set of topological entropy 0. We will not give the precise definitions here, because they are somewhat technical. Li showed [**Li15a**] that no expanding Thurston map is h-expanding and that an expanding Thurston map is asymptotically h-expanding if and only if it has no periodic critical points. A comprehensive account of Li's work on the ergodic theory of expanding Thurston maps can be found in [**Li16b**].

For a rational expanding Thurston map $R\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ there exists a unique R-invariant (Borel) measure λ_R that is absolutely continuous with respect to Lebesgue measure $\mathcal{L}$ on $\widehat{\mathbb{C}}$ (see Theorem 19.2 and Section 19.3). Suppose that an expanding Thurston map $f\colon S^2 \to S^2$ is topologically conjugate to a rational map $R\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$, i.e., there is a homeomorphism $h\colon S^2 \to \widehat{\mathbb{C}}$ such that $f = h^{-1} \circ R \circ h$. Then we can pull back the measure λ_R by h to obtain a measure λ_f on S^2. More explicitly, λ_f is defined by setting $\lambda_f(A) = \lambda_R(h(A))$ for each Borel set $A \subset S^2$. One can show that the measure λ_f only depends on f and not on the choice of the map h that conjugates f to a rational map (this easily follows from Theorem 2.19).

Often f is known to be topologically conjugate to a rational map, even though we do not have an explicit conjugating map h. A simple example for this situation is when the expanding Thurston map f has

precisely three postcritical points, but no periodic critical point (see Theorem 7.3 (ii)).

PROBLEM 13. *Assume an expanding Thurston map $f : S^2 \to S^2$ is topologically conjugate to a rational map. Is it possible to construct the measure λ_f intrinsically?*

In other words, we would like to obtain the measure λ_f without the use of the conjugating map h. If one can find a good characterization of the measure λ_f if it exists, it might be possible to decide whether f is topologically conjugate to a rational map by measure-theoretic methods. This is related to Thurston's theorem.

APPENDIX A

In this appendix we collect various facts whose discussion would have interrupted the main flow of our presentation. Among other things we discuss branched covering maps (Section A.6) and orbifolds (Sections A.9 and A.10) in quite some detail, because it is hard to find the statements relevant for us in the literature.

A.1. Conformal metrics

Here we summarize some standard metric space terminology and record facts related to conformal metrics.

Let (X, d) be a metric space. A *path* γ in X is a continuous map $\gamma \colon [a, b] \to X$ defined on some interval $I = [a, b] \subset \mathbb{R}$. Sometimes one considers also paths defined on half-open or open intervals $I \subset \mathbb{R}$. As is common, we also use the notation γ for the image set $\gamma(I) \subset X$ of a path.

A path $\gamma \colon [a, b] \to X$ *joins* two points $x, y \in X$ if $\gamma(a) = x$ and $\gamma(b) = y$. The *length* of γ is given as

$$\text{length}_d(\gamma) := \sup \sum_{k=1}^{n} d(\gamma(t_{k-1}), \gamma(t_k)) \in [0, \infty],$$

where the supremum is taken over all $n \in \mathbb{N}$ and all points $t_0 = a < t_1 < \cdots < t_n = b$. The path γ is called *rectifiable* (with respect to d) if $L := \text{length}_d(\gamma) < \infty$. In this case, we define $L(t) := \text{length}_d(\gamma|[a, t])$ for $t \in [a, b]$. Then there is a unique path $\widetilde{\gamma} \colon [0, L] \to X$, called the *arclength parametrization* of γ, such that $\widetilde{\gamma}(L(t)) = \gamma(t)$ for all $t \in [a, b]$.

If $\rho \colon X \to [0, \infty]$ is a Borel function, we define the path integral of ρ along the rectifiable path γ as

$$\int_{\gamma} \rho \, ds := \int_{0}^{L} \rho(\widetilde{\gamma}(s)) \, ds.$$

The metric d is called a *length metric* or *path metric* if

$$d(x, y) := \inf_{\gamma} \text{length}_d(\gamma)$$

for $x, y \in X$, where the infimum is taken over all paths γ in X joining x and y. The metric is a *geodesic metric* if this infimum is attained as a minimum. A path realizing this infimum is called a *geodesic segment* joining x and y.

Suppose Ω is a region in the complex plane $\mathbb{C}$ equipped with the Euclidean metric and $\rho\colon \Omega \to (0, \infty)$ is a positive and continuous function on Ω. Then we can define a metric on Ω by setting

$$(A.1) \qquad d(u, v) = \inf_{\gamma} \int_{\gamma} \rho(z)\, |dz|$$

for $u, v \in \Omega$, where the infimum is taken over all rectifiable paths γ in Ω joining u and v and $|dz|$ refers to integration with respect to Euclidean arclength. We say that d is the *conformal metric* on Ω with *length element* $ds = \rho(z)\, |dz|$ and call ρ the *conformal factor* of d.

The Riemann sphere $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ can be equipped with two natural metrics that are essentially equivalent. The *spherical metric* is a conformal metric on $\widehat{\mathbb{C}}$ given by the length element

$$(A.2) \qquad d\sigma = \frac{2|dz|}{1 + |z|^2}$$

(strictly speaking, this gives the restriction of the spherical metric to $\mathbb{C}$). This is actually a geodesic metric on $\widehat{\mathbb{C}}$.

One can identify $\widehat{\mathbb{C}}$ with the unit sphere in $\mathbb{R}^3$ via stereographic projection. The *chordal metric* σ on $\widehat{\mathbb{C}}$ is the metric that corresponds to the Euclidean metric in $\mathbb{R}^3$ under this identification. More explicitly,

$$(A.3) \qquad \sigma(z, w) = \frac{2|z - w|}{\sqrt{1 + |z|^2}\sqrt{1 + |w|^2}}$$

for $z, w \in \mathbb{C}$, and

$$\sigma(\infty, z) = \sigma(z, \infty) = \lim_{w \to \infty} \sigma(z, w) = \frac{2}{\sqrt{1 + |z|^2}}$$

for $z \in \mathbb{C}$.

The spherical and the chordal metrics on $\widehat{\mathbb{C}}$ are comparable up to a uniform factor that approaches 1 for small distances. Accordingly, the length of paths are the same for both metrics. Usually, we equip $\widehat{\mathbb{C}}$ with the chordal metric σ and consider the spherical metric as an "infinitesimal version" of σ.

The *spherical derivative* measures the expansion of a rational map $R\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ with respect to the chordal metric. It is given by

$$(A.4) \qquad R^{\sharp}(z) = \lim_{w \to z} \frac{\sigma(R(w), R(z))}{\sigma(w, z)} = \frac{(1 + |z|^2)|R'(z)|}{1 + |R(z)|^2}$$

for all $z \in \widehat{\mathbb{C}}$. If $z = \infty$ or $R(z) = \infty$, the last expression has to be understood as a suitable limit.

For a holomorphic map f we usually denote its derivative by f', but for a differentiable map that is not necessarily holomorphic we use different notation. If f is a map on $\widehat{\mathbb{C}}$ that is differentiable at a point $p \in \widehat{\mathbb{C}}$, we denote by $Df(p)$ its derivative, considered as a linear map between the tangent spaces at p and $f(p)$. We denote by $\|Df(p)\|_\sigma$ the norm of $D(p)$ with respect to the spherical metric; if $p, f(p) \in \mathbb{C}$ and $\|Df(p)\|$ denotes the Euclidean norm, then

$$(A.5) \qquad \|Df(p)\|_\sigma = \frac{(1 + |p|^2)\|Df(p)\|}{1 + |f(p)|^2}.$$

For a holomorphic map f this agrees with the spherical derivative of f at p.

Let $\mathbb{D} := \{z \in \mathbb{C} : |z| < 1\}$ be the unit disk in the complex plane $\mathbb{C}$. Then the *hyperbolic metric* d_0 on $\mathbb{D}$ is defined as the conformal metric with length element

$$(A.6) \qquad ds = \frac{2|dz|}{1 - |z|^2}.$$

The space $(\mathbb{D}, d_0)$ is geodesic and a model of the hyperbolic plane. The conformal automorphisms of $\mathbb{D}$ are precisely the orientation-preserving isometries of $(\mathbb{D}, d_0)$.

Similarly, if $\mathbb{H} = \{z \in \mathbb{C} : \operatorname{Im}(z) > 0\}$ is the upper half-plane in $\mathbb{C}$, then the hyperbolic metric on $\mathbb{H}$ is given by the length element

$$(A.7) \qquad ds = \frac{|dz|}{\operatorname{Im}(z)}.$$

If we equip $\mathbb{H}$ with this metric, then $\mathbb{H}$ and $(\mathbb{D}, d_0)$ are isometric spaces.

If $\Omega \subset \widehat{\mathbb{C}}$ is a region and ρ is a positive continuous function on Ω, then one can define a conformal metric with length element $ds = \rho \, d\sigma$ as in (A.1), where we use integration with respect to spherical arc-length $d\sigma$ as in (A.2) instead of $|dz|$. It is convenient (see Sections A.7 and A.10) to allow *singular conformal metrics* with a continuous conformal factor $\rho \colon \Omega \setminus P \to (0, \infty)$, where $P \subset \Omega$ is a discrete set in Ω (i.e., it has no limit points in Ω) such that for each $p \in P$ we have $\rho(z) \to 0$ or $\rho(z) \to \infty$ as $z \to p$. If in the latter case $\rho(z) \lesssim \sigma(z, p)^{-\alpha}$ for z near p with $\alpha < 1$, then the conformal metric with length element $ds = \rho \, d\sigma$ on $\Omega \setminus P$ extends to a length metric on Ω.

A.2. Koebe's distortion theorem

In this section we discuss some distortion estimates for conformal maps that can be derived from the classical Koebe distortion theorem. Since the conformal maps we are interested in are usually defined on subregions of the Riemann sphere, it is most natural to formulate the estimates in terms of spherical derivatives and chordal distances (see Section A.1). This mostly amounts to a straightforward translation of the corresponding classical distortion estimates for the Euclidean metric with one *caveat*: to get uniform control for the constants it is important to require that the image region of the conformal map is not too large in the Riemann sphere. We will impose the condition that the image region of the map is contained in a hemisphere of $\widehat{\mathbb{C}}$, i.e., a chordal disk of radius $\sqrt{2}$.

THEOREM A.1 (Spherical version of Koebe's distortion theorem). *Let $0 < r < R < \operatorname{diam}_\sigma(\widehat{\mathbb{C}}) = 2$, $z_0 \in \widehat{\mathbb{C}}$, $B := B_\sigma(z_0, R) \subset \widehat{\mathbb{C}}$, and $g\colon B \to g(B) \subset \widehat{\mathbb{C}}$ be a conformal map such that its image $g(B)$ is contained in a hemisphere of $\widehat{\mathbb{C}}$. Then for all $z, w \in B_\sigma(z_0, r)$ we have*

(A.8) $g^\sharp(z) \asymp g^\sharp(w)$ *and*

(A.9) $\sigma(g(z), g(w)) \asymp g^\sharp(z)\sigma(z, w).$

Here $C(\asymp) = C(r/R)$, and $C(r/R) \to 1$ as $r/R \to 0$. Moreover, there exist constants $c_1 = c_1(r/R) > 0$ and $c_2 = c_2(r/R) > 0$ such that for $w_0 := g(z_0)$ we have

(A.10) $c_1 B_\sigma(w_0, g^\sharp(z_0)r) \subset g(B_\sigma(z_0, r)) \subset c_2 B_\sigma(w_0, g^\sharp(z_0)r).$

Here we use the notation $\lambda B_\sigma(w_0, r) := B_\sigma(w_0, \lambda r)$ for $\lambda > 0$.

If we use the Euclidean metric and the usual derivative, then for conformal maps defined on Euclidean disks and with images in $\mathbb{C}$ the statements (A.8), (A.9), and (A.10) (with $c_1 = 1/4$) immediately follow from the classical Koebe distortion theorem (see [**Po92**, Theorem 1.3 and Corollary 1.4]).

By considering $g(z) = nz$ with $n \in \mathbb{N}$ on the unit disk $\mathbb{D}$ one can see that statement (A.8), for example, is not true with a constant independent of the map if one does not impose some restriction on its image.

PROOF. We will derive the spherical versions of the distortion statements from the Euclidean versions as follows.

By pre- and postcomposing the map g with auxiliary rotations of $\widehat{\mathbb{C}}$ (which can be realized by Möbius transformations), we may assume that $z_0 = 0$ and $g(B) \subset \mathbb{D}$. Let r' and R' be the Euclidean radii of

the disks $\widetilde{B} = B_\sigma(0, r)$ and $B = B_\sigma(0, R)$, respectively. Then it follows from (A.3) that

(A.11) $\qquad r' = r/\sqrt{4 - r^2}$ and $R' = R/\sqrt{4 - R^2}.$

Since $R \le 2$, and so $r \le 2r/R$, we obtain the estimate

(A.12) $\qquad r' \le \dfrac{r}{2\sqrt{1 - (r/R)^2}} \le \dfrac{1}{\sqrt{1 - (r/R)^2}}.$

Thus on $\widetilde{B}$ the chordal and the Euclidean metrics differ by a multiplicative constant only depending on r/R. On $g(B) \subset \mathbb{D}$ chordal and Euclidean metrics differ by a uniform multiplicative constant.

Note that (A.11) gives

(A.13) $\qquad\qquad\qquad\qquad r'/R' \le r/R.$

Thus the statements (A.8)–(A.10) follow from their Euclidean counterparts. Moreover, in (A.8) and (A.9) we indeed have $C(r/R) \to 1$ as $r/R \to 0$ which again easily follows from the Euclidean counterpart of this statement. $\qquad\qquad\square$

Of course, for a single conformal map g we always have an estimate as in (A.8) if we allow the constant also to depend on the map, because the spherical derivative of a conformal map is a positive continuous function. In applications we often consider families of maps where all maps have images contained in a hemisphere with possibly finitely many exceptions. Then one still obtains an estimate as in (A.8) with a uniform constant for the whole family if one uses the uniform constant in (A.8) and adjusts it so that the estimate remains valid also for the finitely many exceptional maps.

We also require a version of Koebe's distortion theorem for conformal maps on multiply connected subregions of $\widehat{\mathbb{C}}$. To get uniform distortion estimates, we again assume that the image of the conformal map is contained in a hemisphere.

LEMMA A.2. *Let $\Omega \subset \widehat{\mathbb{C}}$ be a region and $A, B \subset \Omega$ be continua. Then for all conformal maps $h\colon \Omega \to \Omega' := h(\Omega) \subset \widehat{\mathbb{C}}$ whose image Ω' is contained in a hemisphere of $\widehat{\mathbb{C}}$ we have*

(A.14) $\qquad\qquad \operatorname{diam}_\sigma(h(A)) \asymp h^\sharp(a)$ *and*

(A.15) $\qquad\qquad \operatorname{dist}_\sigma(h(A), \partial\Omega') \gtrsim h^\sharp(a)$ *for each $a \in A$,*

where $C(\asymp) = C(A, \Omega)$ and $C(\gtrsim) = C(A, \Omega)$. Moreover,

(A.16) $\qquad\qquad \operatorname{diam}_\sigma(h(A)) \asymp \operatorname{diam}_\sigma(h(B)),$

where $C(\asymp) = C(A, B, \Omega).$

Here a *continuum* $M \subset \widehat{\mathbb{C}}$ is a compact and connected set that contains more than one point.

PROOF. In the following, all metric notions refer to the chordal metric σ. If $D = B(z_0, r)$ is a disk, we use the notation $2D = B(z_0, 2r)$ for the disk with the same center and twice the radius. Note that $\widehat{\mathbb{C}} \setminus \Omega$ necessarily contains more than two points; so if $2D = B(z_0, 2r) \subset \Omega$, then $2r < 2 = \operatorname{diam}(\widehat{\mathbb{C}})$, and we can apply the distortion estimates of Theorem A.1 for the disks $\widetilde{B} = D$ and $B = 2D$.

A *Harnack chain* (in Ω) is a sequence $D_1, \ldots, D_n$ of disks with $2D_i \subset \Omega$ for $i = 1, \ldots, n$ and $D_i \cap D_{i+1} \neq \emptyset$ for $i = 1, \ldots, n-1$. We call n the length of the Harnack chain, and say that it joins two points $u, v \in \Omega$ if $u \in D_1$ and $v \in D_n$. Note that if u and v are two points in Ω that can be joined by a Harnack chain of length n, then repeated application of (A.8) leads to $h^{\sharp}(u) \asymp h^{\sharp}(v)$ with $C(\asymp) = C_0^n$, where $C_0 \geq 1$ is a universal constant.

Now any two points in a compact subset K of Ω can be joined by a Harnack chain whose length is uniformly bounded only depending on K and Ω. This implies that

(A.17) $$h^{\sharp}(u) \asymp h^{\sharp}(v) \quad \text{for all } u, v \in A,$$

where $C(\asymp) = C(A, \Omega)$ is independent of u, v, and h.

Let $a, u, v \in A$ be arbitrary. Then there exists a Harnack chain in Ω that joins u and v and has length uniformly bounded from above. Then (A.9), the triangle inequality, and (A.17) give

$$\sigma(h(u), h(v)) \lesssim h^{\sharp}(a)$$

with $C(\lesssim) = C(A, \Omega)$. Hence $\operatorname{diam}(h(A)) \lesssim h^{\sharp}(a)$ with an implicit multiplicative constant only depending on A and Ω. This gives one of the estimates in (A.14).

To show the other estimate in (A.14), we pick a disk D with $2D \subset \Omega$ that contains two distinct points $u_0, u_1 \in D \cap A$. Then by (A.9) and (A.17) we have

$$\operatorname{diam}(h(A)) \geq \sigma(h(u_0), h(u_1)) \gtrsim h^{\sharp}(u_0)\sigma(u_0, u_1) \asymp h^{\sharp}(u_0) \asymp h^{\sharp}(a)$$

for each $a \in A$ with implicit multiplicative constants only depending on A and Ω.

To prove (A.15), we note that by (A.10) and (A.14) we have

$$\operatorname{dist}(h(a), \partial\Omega') \gtrsim h^{\sharp}(a)\operatorname{dist}(a, \partial\Omega) \asymp h^{\sharp}(a)$$

for all $a \in A$, where $C(\asymp) = C(A, \Omega)$. Inequality (A.15) follows.

Finally, in order to establish (A.16) pick $a \in A$ and $b \in B$. By similar arguments as above one sees that $\operatorname{diam}(h(B)) \asymp |h^{\sharp}(b)|$ with

$C(\asymp) = C(B, \Omega)$, and that $|h^{\sharp}(a)| \asymp |h^{\sharp}(b)|$ with $C(\asymp) = C(A, B, \Omega)$. Hence

$$\operatorname{diam}(h(A)) \asymp |h^{\sharp}(a)| \asymp |h^{\sharp}(b)| \asymp \operatorname{diam}(h(B))$$

with $C(\asymp) = C(A, B, \Omega)$ as desired. $\qquad\square$

A.3. Janiszewski's lemma

In this section we discuss some topological facts related to separation of sets. We will establish Lemma A.4 that is required for the proof of Theorem 20.2.

In the following, S^2 will be a topological 2-sphere. If $\Omega \subset S^2$ is a region, and $A, B \subset \Omega$, then a set $K \subset \Omega$ *separates* the sets A and B in Ω if for every path γ in Ω joining A and B (i.e., γ has one endpoint in A and one in B), we have $\gamma \cap K \neq \emptyset$. Note that this is meaningful even if K is not disjoint from A or B. We omit the phrase "in Ω" if Ω is understood.

The set K *separates* $x \in \Omega$ and $y \in \Omega$ (or $x \in \Omega$ and a set $B \subset \Omega$) if K separates $A = \{x\}$ and $B = \{y\}$ (or $\{x\}$ and B) in Ω.

The following fact is well known.

LEMMA A.3 (Janiszewski's lemma). *Let* $K, L \subset \mathbb{R}^2$ *be two closed sets with* $K \cap L = \emptyset$. *If two points* $x, y \in \mathbb{R}^2$ *are separated by* $K \cup L$, *then they are separated by* K *or by* L.

A version of this can be found in [**Bi83**, Theorem III.4.A]; exactly the same statement is true if $\mathbb{R}^2$ is replaced by a 2-sphere S^2.

We need a slightly more sophisticated lemma in the same spirit.

LEMMA A.4. *Let* $\Omega \subset S^2$ *be a simply connected region,* $A, B \subset \Omega$ *connected sets, and* $K \subset \Omega$ *a set that is relatively closed in* Ω *and has finitely many connected components. If* K *separates* A *and* B *in* Ω, *then one of the components of* K *separates* A *and* B *in* Ω.

This is a rather straightforward consequence of Janiszewski's lemma if we make the additional assumption that the sets A and B do not meet K. This assumption is not convenient in our application of the lemma though (see the proof of Lemma 20.12). Possible crossings of the sets A or B with K complicate the situation and require a somewhat more involved argument.

PROOF. If $\Omega \neq S^2$, then Ω is homeomorphic to $\mathbb{R}^2$, and, as we have a purely topological statement, we may assume $\Omega = \mathbb{R}^2$. We will first present the proof in this case, and comment on the minor changes necessary for $\Omega = S^2$ after the argument.

We first establish the statement under an additional hypothesis.

Special Case: The set B is a singleton set, i.e., $B = \{b\}$, where $b \in \mathbb{R}^2$.

We run an induction on the number n of components of K. The induction beginning $n = 1$ is clear. For the induction step, we assume that the statement is true (for given A and B) for sets K with at most $n \in \mathbb{N}$ components. Now let K be a closed set in $\mathbb{R}^2$ with $n + 1$ components that separates A and B. Then we can decompose K as $K = C \cup K'$, where C is a component of K, and K' consists of the other n components of K. Since K is closed, the sets K' and C are also closed.

Let $S \subset A$ be the set of all points $a \in A$ that are separated from B by C. Similarly, let S' be the set of all points $a \in A$ that are separated from B by K'. Janiszewski's lemma now implies that $A = S \cup S'$. Here we are using the assumption that B is a singleton set.

If $S = A$, then C separates A and B, and we are done. Assuming from now on that $S \neq A$, we will show in the following that $S' = A$. Since K' has n components, our induction hypothesis then applies. This means there exists a component of K', and hence also a component of K, that separates A and B. This will finish the argument in this case.

If $S = \emptyset$, then $S' = A$, which is our desired statement. So we are reduced to the case where $S \neq \emptyset$ and $S \neq A$.

Claim 1. The sets S and S' are relatively closed in A.

Assume first that S is not relatively closed in A. Then there is a sequence $\{a_n\}$ of points in S with $a_n \to a$ as $n \to \infty$, where $a \in A \setminus S$. Then C does not separate a and B, and so there is a path γ in $\mathbb{R}^2$ joining a and B that does not meet C. In particular, $a \notin C$, and so there exists a small path-connected neighborhood of a in $\mathbb{R}^2$ disjoint from C; but then by traveling first from a_n to a in this neighborhood and then along γ, for large n we can join a_n and B by a path that avoids C, contradicting our assumption that $a_n \in S$. Thus S is relatively closed in A. The argument that S' is relatively closed in A is completely analogous. Claim 1 is proved.

Claim 2. $S \cap S' \cap C \neq \emptyset$.

Recall that $S \neq \emptyset$ and $S \neq A$. Since $S \subset A$ is relatively closed in the connected set A, the set S cannot be relatively open in A.

Hence there exists a point $a \in S$ and a sequence $\{a_n\}$ of points in $A \setminus S$ with $a_n \to a$ as $n \to \infty$. Thus the sequence $\{a_n\}$ is contained in S'. Since S' is relatively closed in A, it follows that $a \in S'$.

Moreover, $a \in C$; for otherwise, we can again find a small path-connected neighborhood of a in $\mathbb{R}^2$ disjoint from C. Then for large
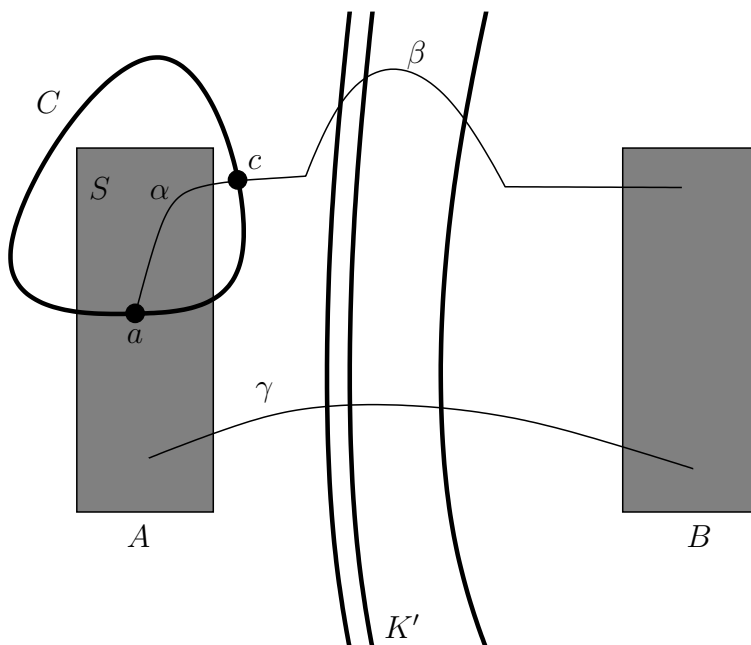
FIGURE A.1. Proof of Lemma A.4.

n, we could travel from a to a_n in this neighborhood, and then, since $a_n \in A \setminus S$, from a_n to B along a path disjoint from C. This contradicts the fact that $a \in S$. So $a \in S \cap S' \cap C$ and Claim 2 follows.

Claim 3. K' separates C and B.

The following argument is illustrated in Figure A.1. Assuming this claim is not true, we can find a path β that avoids K' and joins a point $c \in C$ to B. Now K' is closed and so $\mathbb{R}^2 \setminus K'$ is a union of open regions. Since $C \subset \mathbb{R}^2 \setminus K'$ is connected, the set C is contained in one of these regions. Since regions are path-connected, we can find a path α that joins $c \in C$ and a point $a \in S \cap S' \cap C$ (as provided by Claim 2), and avoids K'. Concatenating α and β, we obtain a path in $\mathbb{R}^2 \setminus K'$ joining a and B. This is impossible since $a \in S'$, meaning that K' separates a and B. This finishes the proof of Claim 3.

Claim 4. K' separates A and B.

Indeed, suppose γ is a path joining A and B. We claim that it meets K'. Since $K = K' \cup C$ separates A and B, it must meet K' or C. If it meets C, then it also meets K' as K' separates C and B by Claim 3. So γ meets K' in any case. Claim 4 follows.

We can now apply the induction hypotheses to K'. Since K' separates A and B, and has only n components, there exists a component of K', and hence also a component of K, that separates A and B.

General Case: $A, B \subset \mathbb{R}^2$ are arbitrary connected sets.

First note that the statement in the above special case (with the roles of A and B reversed) gives the following version of Janiszewski's lemma: Let A be a singleton set in $\mathbb{R}^2$, $B \subset \mathbb{R}^2$ be connected, and $K, L \subset \mathbb{R}^2$ be closed sets with finitely many components. If $K \cap L = \emptyset$ and $K \cup L$ separates A and B, then K or L separates A and B.

Indeed, by what we have seen, one of the components of $K \cup L$ separates A and B, which implies that K or L separates A and B.

The proof in the general case is now a repetition of the proof in the special case. The only difference is that we apply the above version of Janiszewski's lemma instead of the original version. We used it only once: to show that $A = S \cup S'$. In the general case, where $B \subset \mathbb{R}^2$ is connected, but not necessarily a singleton set, each point $a \in A$ is separated from B by $K = C \cup K'$. So by the modified version of Janiszewski's lemma a is separated from B by C or by K'. Thus $a \in S$ or $a \in S'$, and so again $A = S \cup S'$. The rest of the proof is concluded as above.

This completes the proof if Ω is homeomorphic to $\mathbb{R}^2$. The proof in the case $\Omega = S^2$ is the same. Here we apply the S^2-version of Janiszewski's lemma mentioned after the formulation of the $\mathbb{R}^2$-case.
□

A.4. Orientations on surfaces

Orientation is a subject that is easy to grasp on an intuitive level, but is notoriously difficult to discuss rigorously without some sophisticated mathematical concepts or facts. We first recall the fairly standard way of introducing orientation for surfaces by using homology groups (see [**Ha02**] for general background), and then discuss an alternative and very intuitive approach to orientation based on the concept of a flag.

Let M be a compact and connected n-dimensional topological manifold (without boundary). If the singular homology group $H_n(M)$ (with coefficients in $\mathbb{Z}$) is isomorphic to $\mathbb{Z}$, then we call M *orientable* (see [**Ha02**, Section 3.3] for a more detailed discussion). This is true, for example, if M is a 2-sphere or a 2-dimensional torus (the only cases we are interested in).

We say that M is *oriented* if one of the two generators of $H_n(M) \cong \mathbb{Z}$ has been chosen as the *fundamental class* $[M]$ of M. If $f \colon M \to$

N is a homeomorphism between compact and connected oriented n-dimensional topological manifolds M and N, then f induces an isomorphism $f_*\colon H_n(M) \to H_n(M)$, and so $f_*([M]) = [N]$ or $f_*([M]) = -[N]$. In the first case we say that f is *orientation-preserving*, and in the second that f is *orientation-reversing*.

In this framework we can also define the *(topological) degree* of a continuous map. Namely, if M and N are oriented n-dimensional topological manifolds with fundamental classes $[M]$ and $[N]$, respectively, and $f\colon M \to N$ is a continuous map, then its *degree* $\deg(f) \in \mathbb{Z}$ is the unique integer such that $f_*([M]) = \deg(f)[N]$, where $f_*\colon H_n(M) \to H_n(N)$ is the map between homology groups induced by f. Note that the sign of $\deg(f)$ depends on the orientations chosen on M and N. If $M = N$ and we choose the same orientation in source and target, then $\deg(f)$ is independent of this choice.

The degree is multiplicative in the following sense: if $f\colon M \to N$ and $g\colon N \to K$ are continuous maps between oriented n-manifolds, then

(A.18)
$$\deg(g \circ f) = \deg(g) \cdot \deg(f).$$

This immediately follows from the relation $(g \circ f)_* = g_* \circ f_*$ for the induced maps on homology (see [**Ha02**, p. 134]).

For open manifolds or manifolds with boundary one has to resort to suitable relative homology groups to give precise definitions for concepts related to orientation. This is somewhat technical and we will discuss this only in a simple relevant case to give the general idea.

Let M be a *surface*, i.e., a 2-dimensional topological manifold. We assume that M is compact, connected, and oriented. Then the orientation on M induces an orientation on every Jordan region $X \subset M$ which in turn induces an orientation on ∂X and on every arc $\alpha \subset \partial X$. These orientations are represented by generators in the homology groups $H_2(X, \partial X)$, $H_1(\partial X)$, and $H_1(\alpha, \partial\alpha)$, respectively.

To see how to get canonical generators in these groups from the fundamental class of M, first note that we have natural isomorphisms

$$H_2(M) \cong H_2(M, M \setminus \mathrm{int}(X)) \cong H_2(X, \partial X) \cong \mathbb{Z}$$

induced by the inclusion map and excision (see [**Ha02**, Section 2.1] for the relevant terminology and facts). Hence we get an induced orientation on X as represented by a generator of $H_2(X, \partial X)$ obtained as the image of $[M]$.

Similarly, we have natural isomorphisms $H_2(X, \partial X) \cong H_1(\partial X)$ (from the long exact sequence of relative homology) and $H_1(\partial X) \cong$

$H_1(\alpha, \partial\alpha)$. They give us canonical generators of the relevant homology groups once we have an orientation on M.

On a more intuitive level, an orientation of an arc is just a selection of one of the endpoints as the *initial point* and the other endpoint as the *terminal point*. This can easily be reconciled with the homological viewpoint if one uses the isomorphism $H_1(\alpha, \partial\alpha) \cong \widetilde{H}_0(\partial\alpha)$ for reduced homology.

Suppose the Jordan region $X \subset M$ in the oriented surface M is equipped with the induced orientation. If $\alpha \subset \partial X$ is an arc with a given orientation, then we say that X lies *to the left* or *to the right* of α depending on whether the orientation on α induced by the orientation of X agrees with the given orientation on α or not. Similarly, we say that with a given orientation of ∂X the Jordan region X lies to the left or right of ∂X.

An alternative way to introduce orientation is by using the notion of a flag. We will outline this only for surfaces. Let M be a connected (possibly open) surface. By definition a *(topological) flag* on M is a triple (c_0, c_1, c_2), where $c_i \subset M$ is an i-dimensional cell for $i = 0, 1, 2$ with $c_0 \subset \partial c_1$ and $c_1 \subset \partial c_2$. So a flag in M is a closed Jordan region c_2 with an arc c_1 contained in its boundary, where the point in c_0 is one of the endpoints of c_1. We orient the arc c_1 so that the point in c_0 is the initial point in c_1. If we already have an orientation on M, then the flag is called *positively-* or *negatively-oriented* (for the given orientation on M) depending on whether c_2 lies to the left or to the right of the oriented arc c_1.

One can turn this around to give an alternative definition of orientation. Namely, we call two flags (c_0, c_1, c_2) and (c_0', c_1', c_2') in M *equivalent* if there exists a homeomorphism $f\colon M \to M$ that is isotopic to id_M and satisfies $f(c_i) = c_i'$ for $i = 0, 1, 2$. On every connected surface M there are at most two equivalence classes of flags. This can easily be derived from the fact that if $X, Y \subset M$ are Jordan regions, then there exists a homeomorphism $f\colon M \to M$ that is isotopic to id_M and satisfies $f(X) = Y$. To get such a homeomorphism, one shrinks X and Y by isotopies on M into small neighborhoods of points $x \in \mathrm{int}(X)$ and $y \in \mathrm{int}(Y)$, and moves the neighborhood of x to the neighborhood of y by an isotopy. In the shrinking process it is important that for every Jordan region $Z \subset M$ there is a Jordan region $Z' \subset M$ such that $Z \subset \mathrm{int}(Z')$. This easily follows from the fact that the topological circle ∂Z is "tame" and so has a neighborhood that is homeomorphic to an annulus.

This outline of the argument also makes it obvious that the homeomorphism f on M that is isotopic to id_M and satisfies $f(X) = Y$ can be constructed so that it agrees with id_M outside a suitable compact subset of M.

We call M *orientable* if there exist precisely two such equivalences classes of flags in M. An *orientation* on M is a choice of one of the equivalence classes as a family of distinguished flags. We say that the flags in this class are *positively-oriented* and the flags in the other class are *negatively-oriented*.

Any positively-oriented flag determines the orientation uniquely. So on orientable surfaces such as the plane $\mathbb{C}$ or a 2-sphere we can think of an orientation just as a choice of some flag as positively-oriented.

The standard orientation on $\mathbb{C}$ or on $\widehat{\mathbb{C}}$ is the one for which the *standard flag* (c_0, c_1, c_2) is positively-oriented, where $c_0 = \{0\}$, $c_1 = [0, 1] \subset \mathbb{R}$, and

$$c_2 = \{z \in \mathbb{C} : 0 \le \mathrm{Re}(z) \le 1, \ 0 \le \mathrm{Im}(z) \le \mathrm{Re}(z)\}.$$

Let M be an oriented surface, and Ω be a region in M. Then Ω is orientable. Essentially, this follows from the fact that an isotopy on Ω between flags in Ω can be made to fix points outside a sufficiently large compact subset on Ω. This allows one to extend the isotopy to M.

We can represent the orientation on M by a flag in Ω. This flag represents a unique orientation on Ω, called the *induced orientation* on Ω.

An orientation on a not necessarily connected surface is a choice of an orientation on each of its connected components (if each of these components is orientable). If U is an arbitrary open subset of an oriented surface M, then each of the components of U is contained in a component of M. We equip each of these components of U with the induced orientation from the corresponding component of M. This defines the induced orientation on U.

If $f\colon M \to N$ is a homeomorphism between connected and oriented surfaces M and N, then either f maps all positively-oriented flags in M to positively-oriented flags in N, or all positively-oriented flags in M to negatively-oriented flags in N. We say that f is *orientation-preserving* in the first case, and *orientation-reversing* in the second.

A continuous map $f\colon M \to N$ is called a *local homeomorphism* if each point $p \in M$ has an open neighborhood $U \subset M$ such that $f|U\colon U \to V := f(U)$ is a homeomorphism of U onto V. It follows from the "invariance of domain" (see [**Ha02**, Theorem 2B.3, p. 172]) that then V is an open subset of N. By shrinking U if necessary, we can always assume here that U and V are topological disks. If, in

addition, these homeomorphisms $f|U$ preserve orientation, then we call f *orientation-preserving*. Roughly speaking, this means that f sends a small positively-oriented flag near a point in M to a positively-oriented flag in N.

LEMMA A.5. *Let M, N, K be oriented and connected surfaces, and $f\colon M \to K$, $g\colon N \to K$, and $h\colon M \to N$ be local homeomorphisms such that $f = g \circ h$. If two of the maps f, g, h are orientation-preserving, then the third map is orientation-preserving as well.*

PROOF. We have to consider three cases depending on which two of the maps f, g, h are orientation-preserving. We will only consider the case when f and h are orientation-preserving, and show that then g has the same property. The other two cases are very similar and we leave the details to the reader.

Since N is connected, g either preserves orientation near all points in N, or reverses it. In order to decide this, it suffices to consider g near one point $q \in N$ and verify that g preserves the orientation of one positively-oriented flag F contained in a small topological disk V with $q \in V$ such that the map $g|V$ is a homeomorphism onto its image.

We may assume that $q \in h(X)$. Then there exists $p \in M$ with $h(p) = q$, and we can find a topological disk $U \subset M$ with $p \in U$ such that $f|U$ and $h|U$ are orientation-preserving homeomorphisms onto their images. By replacing V with a smaller topological disk if necessary, we may assume that $h(U) = V$ and that there is a flag $F' \subset U$ with $h(F') = F$. Since $h|U$ is orientation-preserving and F is positively-oriented, the flag F' must also be positively-oriented. Since $f|U$ is orientation-preserving, the image flag $f(F') = g(h(F')) = g(F)$ of F' under f which agrees with the image of F under g is positively-oriented. Hence g is orientation-preserving. $\qquad\square$

A.5. Covering maps

Before we turn to *branched* covering maps, we remind the reader of some well-known facts about covering maps. They are true in great generality, but we restrict ourselves mostly to covering maps between surfaces (see [**Ha02**, Section 1.3] and [**Fo81**, Chapter 1] for a more detailed discussion). Here and also in the following section a *surface* is a connected and orientable 2-dimensional topological manifold. So in contrast to Section A.4 we use this term in a more restrictive sense. We assume that a surface is oriented by specifying an equivalence class of positively-oriented flags (as discussed in Section A.4).

Let X and Y be (oriented) surfaces, and $\pi\colon X \to Y$ be a continuous and surjective map. Then π is called a *covering map* if every point

$p \in Y$ has an open and connected neighborhood $V \subset Y$ such that $\pi^{-1}(V)$ can be written as a disjoint union

$$\pi^{-1}(V) = \bigcup_{i \in I} U_i$$

of open and connected sets $U_i \subset X$ such that $\pi|U_i$ is an orientation-preserving homeomorphism of U_i onto V for each $i \in I$. Here I is some index set. We say that a set V as in this definition is *evenly covered* by π. By possibly shrinking the set, one can always assume that V is a topological disk.

Usually, one does not insist on the maps $\pi|U_i$ being orientation-preserving; this additional requirement is motivated by our definition of a branched covering map in the next section: without it a covering map would not necessarily be a branched covering map.

If $\pi\colon X \to Y$ is a covering map and we want to emphasize Y, then we say that π is a *covering map over Y*. The covering map $\pi\colon X \to Y$ is called *finite* if π is *finite-to-one* in the sense that every point in $q \in Y$ has only finitely many preimages in X. In this case, the cardinality $\#\pi^{-1}(q)$ is constant and independent of $q \in Y$.

A covering map is a local homeomorphism, i.e., an open and continuous map such that for every point $x \in X$ there exists an open neighborhood U such that $\pi|U$ is a homeomorphism of U onto $\pi(U)$. Moreover, π is orientation-preserving in addition. Conversely, if X and Y are compact, then every orientation-preserving local homeomorphism $\pi\colon X \to Y$ is a covering map.

If $\pi\colon X \to Y$ is a covering map, then a homeomorphism $g\colon X \to X$ is called a *deck transformation* of π if $\pi = \pi \circ g$. These maps g form a group G called the *deck transformation group* of π. If X is simply connected, then G is isomorphic to the fundamental group of Y (see the discussion below and [**Fo81**, Theorem 5.6]).

Let $\pi\colon X \to Y$ be a covering map, Z a topological space, and $f\colon Z \to Y$ be a continuous map. A continuous map $g\colon Z \to X$ is called a *lift* of f (under π) if $\pi \circ g = f$. In this case, we have the commutative diagram:

(A.19)

$$\begin{array}{ccc} & & X \\ & \overset{g}{\nearrow} & \downarrow{\scriptstyle \pi} \\ Z & \overset{f}{\longrightarrow} & Y. \end{array}$$

The next lemma is a standard fact about existence and uniqueness of lifts (for the terminology and the proofs see [**Ha02**, Section 1.3,

Proposition 1.34, and Proposition 1.33]; see also [**Fo81**, Section 1.4 and Theorem 4.17]).

LEMMA A.6 (Existence and uniqueness of lifts). *Let X and Y be (oriented) surfaces, $\pi\colon X \to Y$ be a covering map, and Z be a path-connected and locally path-connected topological space.*

  (i) *Suppose $g_1, g_2\colon Z \to Y$ are two continuous maps such that $\pi \circ g_1 = \pi \circ g_2$. If there exists $z_0 \in Z$ with $g_1(z_0) = g_2(z_0)$, then $g_1 = g_2$.*

  (ii) *Suppose Z is simply connected, $f\colon Z \to Y$ is a continuous map, and $z_0 \in Z$ and $x_0 \in X$ are points such that $f(z_0) = \pi(x_0)$. Then there exists a continuous map $g\colon Z \to X$ such that $g(z_0) = x_0$ and $f = \pi \circ g$.*

In (i) the maps g_1 and g_2 are lifts of $f := \pi \circ g_1 = \pi \circ g_2$. So the statement says that lifts of maps are uniquely determined by the image of one point.

Statement (ii) guarantees the existence of a lift g of f with $g(z_0) = x_0$. By (i) this lift g of f satisfying $g(z_0) = x_0$ is unique.

A special and important case is if $Z = [0, 1]$, and $z_0 = 0$. Then f is a path in Y, and the statement says that we can lift it to a unique path g in X if we prescribe any point in the fiber $\pi^{-1}(f(0))$ as the initial point of the lift.

Let X be a path-connected and locally path-connected topological space, and $x_0 \in X$ be a basepoint in X. Then the *fundamental group* $\pi_1(X, x_0)$ of X with respect to x_0 consists of all homotopy classes of loops in X based (i.e, starting and ending) at x_0 (see [**Ha02**, Section 1.3] for precise definitions). Simple connectivity of X means that $\pi_1(X, x_0) = 0$, and so $\pi_1(X, x_0)$ is the trivial group only consisting of the unit element.

Suppose Y is another path-connected and locally path-connected topological space with basepoint y_0, and $f\colon X \to Y$ a continuous map that is basepoint-preserving in the sense that $f(x_0) = y_0$. If we assign to each class $[\gamma] \in \pi_1(X, x_0)$ represented by a loop γ in X based at x_0, the class $[f \circ \gamma]$ represented by the image loop $f \circ \gamma$, then we get a well-defined induced group homomorphism $f_*\colon \pi_1(X, x_0) \to \pi_1(Y, y_0)$.

Suppose X and Y are surfaces, $\pi\colon X \to Y$ is a covering map, and X is simply connected. Then $\pi\colon X \to Y$ is a *universal covering map*: if $f\colon Z \to Y$ is another covering map from a surface Z, $x_0 \in X$ and $z_0 \in Z$ with $\pi(x_0) = f(z_0)$, then there exists a covering map $g\colon X \to Z$ such that $\pi = f \circ g$ and $g(x_0) = z_0$. A universal covering map $\pi\colon X \to Y$ exists for each surface and is unique up to equivalence: if $\widetilde{\pi}\colon \widetilde{X} \to Y$

is another universal covering map, then there exists a homeomorphism $\varphi\colon X \to \widetilde{X}$ such that $\pi = \widetilde{\pi} \circ \varphi$ (see [**Fo81**, Section 1.5]).

A.6. Branched covering maps

In this section we discuss branched covering maps between surfaces. Since it is difficult to find references for this topic in the literature, our exposition is rather detailed and we provide proofs for the statements discussed. We will sometimes skip details if they are straightforward to fill in.

We will first define the relevant terminology; in particular, we will give a precise definition of a *branched covering map* and what it means for a neighborhood of a point to be *evenly covered* in this context. Useful criteria for verifying the relevant conditions are provided by Lemmas A.9 and A.10.

Lemma A.11 shows that a conformal structure can be pulled back by a branched covering map. This implies that one can often reduce to the holomorphic case if one studies such maps. In particular, every branched covering map on a 2-sphere can be represented by a rational map on the Riemann sphere up to suitable homeomorphic coordinate changes in source and target (see Corollary A.12).

The main difficulty in the proof of Lemma A.11 is the behavior of the given map near branch points; this is resolved by what can be viewed as a variant of Riemann's removability theorem.

Lemma A.15 is another useful criterion if one wants to check whether a map is a branched covering map. It says that if three continuous maps f, g, and h between compact surfaces satisfy $f = g \circ h$, and if two of the maps are branched covering maps, then the third one is a branched covering map as well. A similar statement is true about the holomorphicity of the maps f, g, and h.

In the last part of this section we consider existence and uniqueness statements for lifts under branched covering maps (see Lemma A.17 and Lemma A.18).

As in the previous section, we again make the standing assumption that each surface is connected and oriented. Let X and Y be compact surfaces, and $f\colon X \to Y$ be a continuous and surjective map. Recall from Section 2.1 that f is a *branched covering map* if for each point $p \in X$ there exists $d \in \mathbb{N}$, topological disks $U \subset X$ and $V \subset Y$ with $p \in U$, $q := f(p) \in V$, and orientation-preserving homeomorphisms $\varphi\colon U \to \mathbb{D}$ and $\psi\colon V \to \mathbb{D}$ with $\varphi(p) = 0$ and $\psi(q) = 0$ such that

$$(A.20) \qquad\qquad (\psi \circ f \circ \varphi^{-1})(z) = z^d$$

for all $z \in \mathbb{D}$.

So branched covering maps are modeled on non-constant holomorphic maps between compact Riemann surfaces. Every such map is a branched covering map.

For maps between surfaces that are not necessarily compact, one has to adjust the definition of a branched covering map. Recall that for (unbranched) covering maps we require that each point in the target has a neighborhood that is evenly covered. For branched covering maps we impose a similar condition. It is always true for compact surfaces (see the discussion after the proof of Lemma A.10). Accordingly, we make the following definition.

DEFINITION A.7 (Branched covering maps). Let X and Y be (connected and oriented) surfaces, and $f\colon X \to Y$ be a continuous map. Then f is a *branched covering map* if for each point $q \in Y$ there exists an open neighborhood $V \subset Y$ that is *evenly covered* by f in the following sense: for some index set $I \neq \emptyset$ we can write $f^{-1}(V)$ as a disjoint union

$$f^{-1}(V) = \bigcup_{i \in I} U_i$$

of open sets $U_i \subset X$ such that U_i contains precisely one point $p_i \in f^{-1}(q)$. Moreover, we require that for each $i \in I$ there exists $d_i \in \mathbb{N}$, and orientation-preserving homeomorphisms $\varphi_i\colon U_i \to \mathbb{D}$ and $\psi_i\colon V \to \mathbb{D}$ with $\varphi_i(p_i) = 0$ and $\psi_i(q) = 0$ such that

(A.21) $(\psi_i \circ f \circ \varphi_i^{-1})(z) = z^{d_i}$

for all $z \in \mathbb{D}$.

Note that the sets U_i and V are necessarily topological disks, and the sets U_i, $i \in I$, are the connected components of $f^{-1}(V)$. Moreover, (A.21) implies that $f(U_i) = V$ for each $i \in I$.

For given f the number d_i is uniquely determined by $p = p_i$ and called the *local degree* of f at p, denoted by $\deg_f(p)$ or $\deg(f, p)$. Our definition allows different local degrees at points in the same fiber $f^{-1}(q)$. Note that if q' is a point close to, but distinct from $q := f(p)$, then $\deg(f, p)$ is equal to the number of distinct preimages of q' under f close to p. In particular, near p the map f is d-to-1, where $d := \deg(f, p)$.

Every branched covering map $f\colon X \to Y$ is surjective, *open* (images of open sets are open), and *discrete* (the preimage set of every point is *discrete in X*, i.e., it has no limit points in X). Every covering map is also a branched covering map.

A *critical point* of a branched covering map $f\colon X \to Y$ is a point $p \in X$ with $\deg_f(p) \geq 2$. A *critical value* is a point $q \in Y$ such that the fiber $f^{-1}(q)$ contains a critical point of f. The set of critical points of f is discrete in X; indeed, if $p \in X$ is arbitrary, then there exists an open neighborhood U of p such that f is a local homeomorphism on $U \setminus \{p\}$. So the set of critical points of f cannot have a limit point in X. Similarly, the set of critical values of f is discrete in Y, because if $V \subset X$ is an evenly covered neighborhood of a point $q \in Y$, then q is the only possible critical value of f in V.

If $f\colon X \to Y$ is a branched covering map, then f is an orientation-preserving local homeomorphism near each point $p \in X$ that is not a critical point of f. Moreover, if $P \subset Y$ is the set of critical values of f, then $f\colon X \setminus f^{-1}(P) \to Y \setminus P$ is an (unbranched) covering map as defined in Section A.5.

A continuous map $f\colon X \to Y$ between surfaces is called *proper* if $f^{-1}(K)$ is compact for every compact set $K \subset Y$. We record the following useful fact.

LEMMA A.8. *Let $f\colon X \to Y$ be an open and continuous map between surfaces X and Y.*

(i) *If f is proper, then $f(X) = Y$.*

(ii) *Suppose $V \subset Y$ is a region and U a connected component of $f^{-1}(V)$. If f is proper or if $\overline{U}$ is compact, then $f|U\colon U \to V$ is a proper map, $f(U) = V$, and $f(\partial U) \subset \partial V$.*

In particular, a proper, open, and continuous map $f\colon X \to Y$ between surfaces is surjective.

PROOF. (i) It follows from our hypotheses that $f(X)$ is a nonempty open set and from our definition of a surface that Y is connected. So it suffices to show that $f(X)$ is closed. To see this, let $\{y_n\}$ be a sequence in $f(X)$ and suppose that $y_n \to y \in Y$ as $n \to \infty$. Then for each $n \in \mathbb{N}$ there exists $x_n \in X$ with $f(x_n) = y_n$.

Now the set $K := \{y\} \cup \{y_n : n \in \mathbb{N}\} \subset Y$ is compact. Since f is proper, the set $f^{-1}(K) \subset X$ is also compact. Since $\{x_n\}$ is a sequence in $f^{-1}(K)$, it has a convergent subsequence. By passing to a subsequence if necessary, we may assume that $\{x_n\}$ converges itself, say $x_n \to x \in X$ as $n \to \infty$. Then by continuity of f we have

$$y = \lim_{n\to\infty} y_n = \lim_{n\to\infty} f(x_n) = f(x).$$

Hence $y \in f(X)$ and so $f(X)$ is indeed closed.

(ii) Let $K \subset V$ be compact. To see that $f|U$ is proper, we have to show that $(f|U)^{-1}(K) = U \cap f^{-1}(K)$ is compact. To this end, let $\{x_n\}$

be an arbitrary sequence in $U \cap f^{-1}(K)$. If f is proper, then $f^{-1}(K)$ is compact, and so $\{x_n\}$ has a convergent subsequence. This is also true if $\overline{U}$ is compact. By passing to a subsequence, we may assume that $\{x_n\}$ itself converges, say $x_n \to x \in \overline{U}$. We have to show that actually $x \in U \cap f^{-1}(K)$.

By continuity of f the point $f(x)$ is the limit of the sequence $\{f(x_n)\}$ which lies in K. Hence $f(x) \in K$. So if $x \in U$, then $x \in U \cap f^{-1}(K)$ as desired.

The other alternative, $x \in \partial U$ is impossible. Indeed, since $f(x) \in K \subset V$, there exists a small connected neighborhood N of x with $f(N) \subset V$. Since $x \in \partial U$, the set N meets U and so $N \cup U$ is a connected subset of $f^{-1}(V)$. Since U is a connected component of $f^{-1}(V)$, this implies $N \subset U$; but then x would be an interior point of U and not a boundary point of U.

The set U is also a region which implies that $f|U \colon U \to V$ is an open and continuous map between the surfaces U and V. Since $f|U$ is also proper by what we have just seen, it follows from (i) that $f(U) = V$.

Finally, if $x \in \partial U$, then $f(x) \in \overline{V}$ by continuity of f. The argument above shows that $f(x) \in V$ is impossible, and so $f(x) \in \partial V$. Hence $f(\partial U) \subset \partial V$ as desired. $\qquad\square$

Suppose a topological disk $V \subset Y$ is evenly covered by a continuous map $f \colon X \to Y$ and U_i is a component of $f^{-1}(V)$ as in Definition A.7. Then the map $f|U_i \colon U_i \to V$ is proper, because up to homeomorphic changes in source and target the map is given by a power map $z \mapsto z^d$ which is a proper map on $\mathbb{D}$.

The following statement provides a convenient criterion that allows us to verify the conditions in Definition A.7.

LEMMA A.9. *Let U be a surface, V be a topological disk, $p \in U$, $q \in V$, and $f \colon U \to V$ be a proper and continuous map. Suppose that $f^{-1}(q) = \{p\}$ and that f is a local homeomorphism near each point in $U \setminus \{p\}$.*

Then U is also a topological disk, and for each homeomorphism $\psi \colon V \to \mathbb{D}$ with $\psi(q) = 0$, there exists a homeomorphism $\varphi \colon U \to \mathbb{D}$ with $\varphi(p) = 0$ such that

$$(A.22) \qquad\qquad (\psi \circ f \circ \varphi^{-1})(z) = z^d$$

for all $z \in \mathbb{D}$, where $d \in \mathbb{N}$.

If, in addition, ψ is orientation-preserving and f is orientation-preserving near each point in $U \setminus \{p\}$, then φ is orientation-preserving as well.

PROOF. Since f is proper, each point in V has finitely many preimages. Then our assumptions imply that the map f is a covering map from $U \setminus \{p\}$ onto $V \setminus \{q\}$. If ψ is a homeomorphism as in the statement, then $\psi \circ f$ is a covering map from $U \setminus \{p\}$ onto $\mathbb{D} \setminus \{0\}$.

Now it is a standard fact that up to equivalence each covering map onto $\mathbb{D} \setminus \{0\}$ with finite fibers is a power map $P_d(z) := z^d$ on $\mathbb{D} \setminus \{0\}$ for some $d \in \mathbb{N}$ (essentially, this is proved in [**Fo81**, Theorem 5.10]). Here this means that there exists a homeomorphism $\varphi \colon U \setminus \{0\} \to \mathbb{D} \setminus \{0\}$ such that $\psi \circ f = P_d \circ \varphi$ on $U \setminus \{p\}$. This equation implies that we get a homeomorphic extension $\varphi \colon U \to \mathbb{D}$ by setting $\varphi(p) = 0$. Hence U is a topological disk. The first part of the statement follows.

Suppose, in addition, that ψ and f are orientation-preserving. Since P_d is orientation-preserving on $\mathbb{D} \setminus \{0\}$, the relation $\psi \circ f = P_d \circ \varphi$ on $U \setminus \{p\}$ implies that φ must have this property as well (this easily follows from Lemma A.5). $\square$

LEMMA A.10. *Let X and Y be surfaces, and $f \colon X \to Y$ be an open and continuous map. Suppose $q \in Y$ and $V \subset Y$ is a topological disk that is an evenly covered neighborhood of q as in Definition A.7. If $\widetilde{V}$ is a topological disk with $q \in \widetilde{V} \subset V$, then $\widetilde{V}$ is also evenly covered by f.*

PROOF. Suppose we have a decomposition

$$f^{-1}(V) = \bigcup_{i \in I} U_i$$

into connected components as in Definition A.7. Let $\widetilde{U} \subset X$ be a connected component of $f^{-1}(\widetilde{V})$. Then there exists a unique $i \in I$ such that $\widetilde{U} \subset U_i$. We know that the map $f|U_i \colon U_i \to V$ is proper, open, and continuous. Moreover, $\widetilde{U}$ is a component of $(f|U_i)^{-1}(\widetilde{V}) = U_i \cap f^{-1}(\widetilde{V})$. So by Lemma A.8 (ii) the map $f|\widetilde{U} \colon \widetilde{U} \to \widetilde{V}$ is proper and we have $f(\widetilde{U}) = \widetilde{V}$. This, together with $\widetilde{U} \subset U_i$, implies that $(f|\widetilde{U})^{-1}(q) = \{p\}$, where p is the unique point in U_i with $f(p) = q$.

Finally, since V is evenly covered by f, the map f is an orientation-preserving homeomorphism near each point in $\widetilde{U} \setminus \{q\} \subset U_i \setminus \{q\}$. It now follows from Lemma A.9 that up to orientation-preserving homeomorphic changes in source and target, the map $f|\widetilde{U} \colon \widetilde{U} \to \widetilde{V}$ can be represented by a power map $z \mapsto z^d$, where $d \in \mathbb{N}$. Since this is true for each component $\widetilde{U}$ of $f^{-1}(\widetilde{V})$, the topological disk $\widetilde{V}$ is evenly covered by f. $\square$

The arguments in the previous lemma imply that the Definition A.7 of a branched covering map $f \colon X \to Y$ is equivalent to the definition

given in Section 2.1 in case the surfaces X and Y are compact. Indeed, if in this case $f\colon X \to Y$ is a branched covering map according to the definition given in Section 2.1, then each point $q \in Y$ has finitely many distinct preimages $p_1, \ldots, p_n \in X$ under f. For each point p_i there exist topological disks $U_i \subset X$ and $V_i \subset Y$ with $p_i \in U_i$ and $q \in V_i$ such that $f|U_i\colon U_i \to V_i$ can be represented by a power map $z \mapsto z^{d_i}$ with $d_i \in \mathbb{N}$ up to orientation-preserving homeomorphic changes in source and target. We can choose a topological disk $V \subset V_1 \cap \cdots \cap V_n$ with $q \in V$. Then it easily follows from the arguments in the proof of Lemma A.10 that V is a neighborhood of q that is evenly covered by f.

Questions about branched covering maps can often be reduced to the holomorphic case due to the following statement.

LEMMA A.11. *Let X and Y be surfaces, and $f\colon X \to Y$ be a branched covering map. Then for each conformal structure on Y there exists a conformal structure on X such that with these conformal structures on X and Y the map $f\colon X \to Y$ is holomorphic.*

For the proof we will use some standard facts and terminology from the theory of Riemann surfaces. We will follow [**Fo81**, Section 1.1]. A *conformal structure* on a surface is represented by a *complex atlas* of *holomorphically compatible (complex) charts*. Every (orientable) surface admits a complex atlas and hence a conformal structure. A surface equipped with such a conformal structure is called a *Riemann surface*.

PROOF. Let $\mathcal{A}$ be a complex atlas representing the given conformal structure on Y.

Let $E \subset X$ be the set of points where f is not a local homeomorphism. This set is discrete in X. Near each point $X \setminus E$ the map f is an orientation-preserving local homeomorphism. If $p \in X \setminus E$, then we obtain a chart defined near p by composing f restricted to a sufficiently small neighborhood of p with a chart in $\mathcal{A}$ defined near $f(p)$. These charts form an atlas $\mathcal{A}'$ on $X \setminus E$. The charts in $\mathcal{A}'$ are holomorphically compatible, because the charts in $\mathcal{A}$ are. The atlas $\mathcal{A}'$ defines a conformal structure on $X \setminus E$ such that the restriction $f|(X \setminus E)\colon X \setminus E \to Y$ is holomorphic.

It remains to find suitable charts defined near the points in E. Let $p \in E$ be arbitrary, and $q = f(p)$. Since f is a branched covering map, we can find small topological disks $U \subset X$ and $V \subset Y$ with $p \in U$, $U \cap E = \{p\}$, and $q \in V$ such that the map $f\colon U \setminus \{p\} \to V \setminus \{q\}$ is a covering map with finite fibers (up to orientation-preserving

homeomorphic changes in source and image it is represented by a map of the form $z \mapsto z^d$ on $\mathbb{D} \setminus \{0\}$ with $d \in \mathbb{N}$).

We may assume that V is so small that with the given conformal structure on Y the topological disk V is conformally equivalent to $\mathbb{D}$ and hence the punctured disk $V \setminus \{q\}$ is conformally equivalent to $\mathbb{D} \setminus \{0\}$. Since we only have a conformal structure defined on $X \setminus E \supset U \setminus \{p\}$, it is a priori not clear that $U \setminus \{p\}$ is also conformally equivalent to $\mathbb{D} \setminus \{0\}$; but $U \setminus \{p\}$ is a topological annulus and hence with the given conformal structure is conformally equivalent to a Euclidean annulus of the form

$$A = \{z \in \mathbb{C} : r < |z| < R\},$$

where $0 \le r < R \le \infty$. Then from the finite covering map $f\colon U \setminus \{p\} \to V \setminus \{q\}$ we obtain by a conformal change a finite holomorphic covering map $g\colon A \to \mathbb{D} \setminus \{0\}$ that satisfies $|g(z)| \to 0$ as $|z| \to r$ and $|g(z)| \to 1$ as $|z| \to R$. This is only possible if $r = 0$ and $0 < R < \infty$. This follows from the fact that the *conformal modulus* (see [**Mi06a**, Appendix B])

$$\mathrm{mod}(A) = \frac{1}{2\pi} \log(R/r) \in (0, \infty]$$

of the annulus A only changes by a finite multiplicative constant under a finite holomorphic covering map.

We may assume that $R = 1$. In particular, there exists a conformal map $\varphi\colon U \setminus \{p\} \to \mathbb{D} \setminus \{0\}$ such that $\varphi(u) \to 0$ as $u \to p$. If we extend this map by setting $\varphi(p) = 0$, then φ is an orientation-preserving homeomorphism of U onto $\mathbb{D}$. Moreover, this map gives a chart on X defined near p. It is holomorphically compatible with the charts in $\mathcal{A}'$, because the map $\varphi | U \setminus \{p\}$ is holomorphic.

If we add these charts φ defined near points $p \in E$, then we obtain an atlas $\mathcal{A}''$ of holomorphically compatible charts on X. This defines a conformal structure on X and it is clear that with the conformal structures represented by $\mathcal{A}''$ on X and $\mathcal{A}$ on Y, the map $f\colon X \to Y$ is holomorphic. $\qquad\square$

The previous statement can be applied to branched covering maps on a 2-sphere S^2 and gives the following result.

COROLLARY A.12. *Let* $f\colon S^2 \to S^2$ *be a branched covering map, and* $\psi\colon S^2 \to \widehat{\mathbb{C}}$ *be an orientation-preserving homeomorphism. Then there exists an orientation-preserving homeomorphism* $\varphi\colon S^2 \to \widehat{\mathbb{C}}$ *such that* $R := \psi \circ f \circ \varphi^{-1}$ *is a rational map on* $\widehat{\mathbb{C}}$.

In particular, up to homeomorphic changes in source and target, every branched covering map on a 2-sphere S^2 can be represented by a rational map on the Riemann sphere $\widehat{\mathbb{C}}$.

PROOF. The given homeomorphism $\psi\colon S^2 \to \widehat{\mathbb{C}}$ gives a natural conformal structure on S^2. It is represented by an atlas obtained by pulling back charts in an atlas representing the conformal structure on $\widehat{\mathbb{C}}$. Hence by Lemma A.11 there exists another conformal structure on S^2 such that the map $f\colon S^2 \to S^2$ is holomorphic with respect to these two conformal structures on source and target. By the uniformization theorem the sphere S^2 equipped with the source conformal structure is conformally equivalent to $\widehat{\mathbb{C}}$ by a conformal map $\varphi\colon S^2 \to \widehat{\mathbb{C}}$. In particular, $\varphi\colon S^2 \to \widehat{\mathbb{C}}$ is a homeomorphism and $R = \psi \circ f \circ \varphi^{-1}$ is a holomorphic map on $\widehat{\mathbb{C}}$. So R is a rational map. Based on Lemma A.5 the last relation also implies that φ is orientation-preserving, because the maps R, ψ, and f are. $\qquad\square$

The proof of Theorem A.11 also leads to a criterion when a map on S^2 is a branched covering map.

COROLLARY A.13. *Let $f\colon S^2 \to S^2$ be a continuous, open, and discrete map with degree $\deg(f) > 0$. Suppose that there exists a finite set $C \subset S^2$ such that f is a local homeomorphism near each point in $S^2 \setminus C$. Then f is a branched covering map.*

Recall that $\deg(f) \in \mathbb{Z}$ is the unique number such that $f_*([S^2]) = \deg(f)[S^2]$, where $[S^2] \in H_2(S^2)$ is the fundamental class of S^2 (see Section A.4).

A statement much stronger than Corollary A.13 is actually true.

THEOREM A.14. *Let $f\colon S^2 \to S^2$ be a continuous, open, and light map with degree $\deg(f) > 0$. Then f is a branched covering map.*

Here f is called *light* if $f^{-1}(p)$ is totally disconnected for each $p \in S^2$. Theorem A.14 follows from a deep and more general characterization theorem for continuous, open, and light maps between surfaces (see [**Wh42**, Theorem 5.1, p. 198]; we are grateful to P. Haïssinsky for pointing out this reference). Its basic ideas go back to Stoïlow (see [**St28**]).

In particular, for each continuous, open, and discrete map f on S^2 we obtain the existence of local representations as in (A.20), except that now the homeomorphisms ψ and φ are not necessarily orientation-preserving. If f has positive degree in addition, then f is actually is a branched covering map according to our definition (as follows from the argument below). For our purposes the weak form of Theorem A.14 as provided by Corollary A.13 will suffice (it is used in the proof of Theorem 13.2).

PROOF OF COROLLARY A.13. Let $p \in S^2$ be arbitrary and $q = f(p)$. The closed set $f^{-1}(q)$ is discrete in S^2 and so necessarily finite. This and Lemma 5.15 imply that if we take a small enough topological disk V with $q \in V$, then the unique component U of $f^{-1}(V)$ that contains p does not contain any other preimage of q and no point in $C \setminus \{p\}$.

By Lemma A.8 (ii) the map $f|U \colon U \to V$ is proper and it is clear by choice of U and V that the assumptions of Lemma A.9 are satisfied. So near p the map f has a representation as in (A.22) and the set where f is not a local homeomorphism is finite. These assumptions are enough to argue as in the proof of Lemma A.11 that with suitable (and possibly different) conformal structures in source and target the map $f \colon S^2 \to S^2$ is holomorphic. As in the proof of Lemma A.12, it follows that there are homeomorphisms ψ and φ from S^2 to $\widehat{\mathbb{C}}$ such that $\psi \circ f \circ \varphi^{-1}$ is a rational map R on $\widehat{\mathbb{C}}$.

We may assume that ψ is orientation-preserving, but without additional assumptions we cannot guarantee that φ has the same property. Now if, as in our hypotheses, we assume that $\deg(f) > 0$, then one can see that φ is orientation-preserving as follows. Let $[S^2]$ and $[\widehat{\mathbb{C}}]$ be the fundamental classes on S^2 and $\widehat{\mathbb{C}}$, respectively, and φ_*, ψ_*, and f_* the induced maps on homology of degree 2. Then R has positive degree $\deg(R) > 0$, because R is rational, and $\psi_*([S^2]) = [\widehat{\mathbb{C}}]$, because ψ is an orientation-preserving homeomorphism. Moreover, $\varphi_*([S^2]) = \pm[\widehat{\mathbb{C}}]$ depending on whether φ preserves or reverses orientation. Now

$$\deg(f)[\widehat{\mathbb{C}}] = \psi_*(f_*([S^2])) = R_*(\varphi_*([S^2])) = \pm \deg(R)[\widehat{\mathbb{C}}].$$

Since both $\deg(f)$ and $\deg(R)$ are positive, this implies that $\varphi_*([S^2]) = [\widehat{\mathbb{C}}]$, and so φ is indeed orientation-preserving.

It is now clear that $f = \psi^{-1} \circ R \circ \varphi$ is a branched covering map. $\square$

LEMMA A.15 (Compositions of branched covering maps). *Let X, Y, and Z be surfaces, and $f \colon X \to Z$, $g \colon Y \to Z$, and $h \colon X \to Y$ be continuous maps such that $f = g \circ h$.*

(i) *If g and h are branched covering maps, and Y and Z are compact, then f is also a branched covering map.*

(ii) *If f and g are branched covering maps, then h is a branched covering map. Similarly, if f and h are branched covering maps, then g is a branched covering map.*

If, in addition, X, Y, and Z are Riemann surfaces and two of the maps are holomorphic, then the third map is also holomorphic.

So in (ii) we can drop the assumption made in (i) that Y and Z are compact. On the other hand, in the last statement it is enough to assume that g and h are holomorphic and not necessarily branched covering maps in order to conclude that $f = g \circ h$ is holomorphic.

The surfaces and the maps in this statement are related as in the following commutative diagram:

$$X \xrightarrow{\ h\ } Y \xrightarrow{\ g\ } Z.$$

PROOF. (i) We assume that Y and Z are compact, and both g and h are branched covering maps.

This is the easiest case. First note that $g \colon Y \to Z$ is finite-to-one, because Y and Z are compact. If $z_0 \in Z$ is arbitrary, we can find a small topological disk $W \subset Z$ that for the map g is an evenly covered neighborhood of z_0 as in Definition A.7. Let $y_1, \ldots, y_n \in Y$ be the preimage points of z_0 under g, and $V_1, \ldots, V_n \subset Y$ be the components of $g^{-1}(W)$ with $y_i \in V_i$ for $i = 1, \ldots, n$. Each V_i is a topological disk. If we shrink W to z_0, then each disk V_i shrinks to y_i. So by Lemma A.10 we may assume that W is so small that for the map h each V_i is an evenly covered neighborhood of y_i. Then, based on Lemma A.9, one easily sees that W is an evenly covered neighborhood of z_0 for the map $f = g \circ h$. Hence f is a branched covering map.

If X, Y, and Z are Riemann surfaces, and g as well as h are holomorphic, then f is clearly holomorphic as well.

(iia) We assume that f and g are branched covering maps. In contrast to case (i) we do not assume that Y and Z are compact.

Let $y_0 \in Y$ be arbitrary. In order to show that h is a branched covering map, we have to find a neighborhood of y_0 that is evenly covered by h.

For this we set $z_0 = g(y_0)$. Since both f and g are branched covering maps, we can choose a small topological disk $W \subset Z$ with $z_0 \in W$ that is evenly covered by both f and g (this follows from Lemma A.10). Then we have decompositions into connected components of the form

$$g^{-1}(W) = \bigcup_{j \in J} V_j \subset Y$$

and

$$f^{-1}(W) = \bigcup_{i \in I} U_i \subset X$$

as in Definition A.7. Each set V_j contains precisely one point $p_j \in g^{-1}(z_0)$, and each U_i one point $q_i \in f^{-1}(z_0)$.

We denote the connected component of $g^{-1}(W)$ that contains $y_0 \in g^{-1}(z_0)$ by V; so $V = V_{j_0}$ and $y_0 = p_{j_0}$ for some $j_0 \in J$. We claim that V is either disjoint from the image $h(X)$ or is evenly covered by h.

To see this, note that for each $i \in I$, the set $h(U_i)$ is connected and

$$g(h(U_i)) = f(U_i) = W;$$

so $h(U_i) \subset g^{-1}(W)$. Hence $h(U_i)$ has to lie in one of the connected components V_j of $g^{-1}(W)$. Let I_0 be the possibly empty set of all $i \in I$ with $h(U_i) \subset V = V_{j_0}$. Then for each $i \in I \setminus I_0$ we have $h(U_i) \subset \bigcup_{j \in J \setminus \{j_0\}} V_j$, and so $h(U_i) \cap V = \emptyset$. This implies that

$$h^{-1}(V) = \bigcup_{i \in I_0} U_i.$$

So the sets U_i, $i \in I_0$, are the connected components of $h^{-1}(V)$. If $I_0 = \emptyset$, then $h^{-1}(V) = \emptyset$, and so V is disjoint from $h(X)$.

Now let us assume that $I_0 \neq \emptyset$, and fix $i \in I_0$. We will verify the conditions in Lemma A.9 for the map $h \colon U_i \to V$. First note that

$$h^{-1}(y_0) \cap U_i = f^{-1}(z_0) \cap U_i = \{q_i\}.$$

The map $h|U_i \colon U_i \to V$ is proper. Indeed, if $K \subset V$ is compact, then

$$(h|U_i)^{-1}(K) = h^{-1}(K) \cap U_i \subset f^{-1}(g(K)) \cap U_i = (f|U_i)^{-1}(g(K)).$$

Now $g(K) \subset W$ is compact and $f|U_i \colon U_i \to W$ is proper. Hence $(f|U_i)^{-1}(g(K)) \subset U_i$ is compact. Since $\widetilde{K} := (h|U_i)^{-1}(K)$ is relatively closed in U_i and a subset of the compact set $(f|U_i)^{-1}(g(K))$, the set $\widetilde{K}$ is also compact.

Finally, if $q \in U_i \setminus \{q_i\}$, then h is an orientation-preserving homeomorphism locally near q, because locally it is the composition of two such maps, namely, f near q followed by a local inverse of g near $p := h(q) \neq y_0$ (note that this local inverse of g maps $f(q) = g(p)$ back to $p = h(q)$). It now follows from Lemma A.9 that V is evenly covered by h.

We have seen that every point $y_0 \in Y$ has a neighborhood V that is evenly covered by h or is disjoint from $h(X)$. This implies that $h(X)$ is a non-empty open set with open complement. Since Y is connected, it follows that $h(X) = Y$. Therefore, actually every point in Y has a neighborhood that is evenly covered by h. Hence h is a branched covering map.

If, under the additional assumptions, g and f are holomorphic, then h is holomorphic as well. Indeed, if $x_0 \in X$ is not a critical point of f, then f is a local biholomorphism near x_0, and g must be a local biholomorphism near $y_0 = h(x_0)$. Then h is holomorphic near

x_0, because near this point h is the composition of f followed by a holomorphic inverse branch g^{-1} of g with $g^{-1}(f(x_0)) = y_0$.

If x_0 is a critical point of f, then there are no other critical points of f nearby, and so h is holomorphic in a punctured neighborhood of x_0. Since h is continuous at x_0, this is a removable singularity for h and so h is also holomorphic near x_0. It follows that h is holomorphic on X.

(iib) We assume that f and h are branched covering maps. Again we do not assume that Y and Z are compact.

To see that g is also a branched covering map, let $z_0 \in Z$ be arbitrary. Then we can find a neighborhood W of z_0 that is evenly covered by f. Then again we have a decomposition

$$f^{-1}(W) = \bigcup_{i \in I} U_i \subset X$$

into connected components as in Definition A.7. Moreover, there exists a unique point $q_i \in U_i \cap f^{-1}(z_0)$ for each $i \in I$.

Since h is surjective, we have

$$(A.23) \qquad g^{-1}(W) = h(h^{-1}(g^{-1}(W))) = h(f^{-1}(W)) = \bigcup_{i \in I} h(U_i).$$

Fix $i \in I$, and consider the connected component $V \subset Y$ of the open set $g^{-1}(W)$ that contains $h(q_i)$. Since $h(q_i) \in h(U_i) \subset g^{-1}(W)$, it follows that $h(U_i) \subset V$. Note that the map $h|U_i \colon U_i \to V$ is proper. Indeed, if $K \subset V$ is compact, then

$$(h|U_i)^{-1}(K) \subset (f|U_i)^{-1}(g(K)),$$

and, by a similar reasoning as in the proof of (iia), we see that the set $(h|U_i)^{-1}(K)$ is also compact. So we can apply Lemma A.8 (i) to the map $h|U_i \colon U_i \to V$ and it follows that $V = h(U_i)$. So for each $i \in I$ the set $h(U_i)$ is the connected component of $g^{-1}(W)$ that contains $h(q_i)$. Moreover, by (A.23) each connected component V of $g^{-1}(W)$ is such an image $h(U_i)$ for some $i \in I$.

We now claim that W is evenly covered by g. To see this, let V be a connected component of $g^{-1}(W)$. It suffices to verify the conditions of Lemma A.9 for the map $g|V \colon V \to W$. For this we pick U_i such that $h(U_i) = V$, and let $p_0 = h(q_i)$.

Then

$$g^{-1}(z_0) \cap V = h(f^{-1}(z_0) \cap U_i) = \{h(q_i)\} = \{p_0\}.$$

The map $g|V\colon V \to W$ is proper, because if $K \subset W$ is compact, then $f^{-1}(K) \cap U_i = (f|U_i)^{-1}(K)$ is compact, and so

$$(g|V)^{-1}(K) = g^{-1}(K) \cap V = h(f^{-1}(K) \cap U_i)$$

is compact as well.

Finally, if $p \in V \setminus \{p_0\}$, then g is an orientation-preserving homeomorphism near p; again this is true, because near p the map g is the composition of two such maps. To see this, we pick $q \in U_i$ with $h(q) = p$. Then $q \neq q_i$, and so f is an orientation-preserving homeomorphism near q. Since $f = g \circ h$, the map h is also locally injective near q, and hence an orientation-preserving homeomorphism near q, because h is a branched covering map. A local inverse h^{-1} of h defined near p with $h^{-1}(p) = q$ has the same property. Finally, $g = f \circ h^{-1}$ near p and we have a local representation of g as desired. It follows that g is indeed a branched covering map.

If, under the additional assumptions, f and h are holomorphic, then g is also holomorphic as can be verified by an argument very similar to the one in (iia): one first shows that g is holomorphic away from the critical values of h and then that the critical values of h are removable singularities for g. □

As the following lemma shows, the local degree behaves in the expected way under compositions of branched covering maps.

LEMMA A.16. *Let $f\colon X \to Y$ and $g\colon Y \to Z$ be branched covering maps, where X, Y, and Z are surfaces such that Y and Z are compact. Then*

$$(A.24) \qquad \deg(g \circ f, x) = \deg(g, f(x)) \cdot \deg(f, x)$$

for all $x \in X$.

PROOF. Let $x \in X$ be arbitrary. We know by Lemma A.15 that $g \circ f$ is a branched covering map, and so all terms in (A.24) are defined.

Let $y = f(x)$ and $z = g(y) = g(f(x))$. We consider a point $z' \neq z$ close to z. In order to determine $\deg(g \circ f, x)$, we have to count the number of preimages of z' under $g \circ f$ that are close to x.

Now if $z' \neq z$ is close to z, then it has $k := \deg(g, y)$ distinct preimages $y_1, \ldots, y_k \neq y$ under g near y. We may assume that z' is so close to z that the points $y_1, \ldots, y_k$ lie in a neighborhood of y that is evenly covered by f. Then each of the points y_i, $i = 1, \ldots, k$, has precisely $l := \deg(f, x)$ preimages under f close to x.

So z' has precisely $k \cdot l = \deg(g, f(x)) \cdot \deg(f, x)$ preimages under $g \circ f$ that are close to x and (A.24) follows. □

We next discuss existence and uniqueness for lifts under branched covering maps. The basic diagram is again (A.19), where π is now a given branched covering map. A map g as in (A.19), is called a *lift* of f (under π). We first record an important special case of this situation.

LEMMA A.17 (Lifting paths under branched covering maps). *Let X and Y be surfaces, $\pi\colon X \to Y$ be a branched covering map, $\gamma\colon [0,1] \to Y$ be a path in Y, and $x_0 \in \pi^{-1}(\gamma(0))$. Then there exists a path $\alpha\colon [0,1] \to X$ with $\alpha(0) = x_0$ and $\pi \circ \alpha = \gamma$.*

So every path can be lifted under a branched covering map. Moreover, any point in the fiber over the initial point of the original path can be prescribed as the initial point of the lift. In general, the lift is not uniquely determined, because it can make various "turns" as it runs through critical points of π. For a more general path-lifting statement see [**Fl50**, Theorem 2].

PROOF. We will only give an outline of the proof and leave some straightforward details to the reader.

If we break up γ into small subpaths, we can reduce to the situation where γ runs in an open subset of Y that is evenly covered by π. Using local coordinates, we can further make the assumption that X and Y are equal to the open unit disk $\mathbb{D}$ in $\mathbb{C}$, and $\pi(z) = z^n$ for $z \in X = \mathbb{D}$, where $n \in \mathbb{N}$.

Then the set $[0,1] \setminus \gamma^{-1}(0)$ can be written as a disjoint union of relatively open intervals $I_j \subset [0,1]$, $j \in J$, where J is a countable index set. Each path $\gamma|I_j$ runs in $\mathbb{D} \setminus \{0\}$ and hence has a (non-unique) lift α_j under the covering map $\pi\colon \mathbb{D} \setminus \{0\} \to \mathbb{D} \setminus \{0\}$. Moreover, if $0 \in I_j$ we can choose the lift α_j so that $\alpha_j(0) = x_0$ for a given point $x_0 \in \pi^{-1}(\gamma(0))$.

A continuous lift $\alpha\colon [0,1] \to \mathbb{D}$ of γ with $\alpha(0) = x_0$ is now defined by setting $\alpha(t) = \alpha_j(t)$ if $t \in [0,1]$ lies in one of the intervals I_j, and setting $\alpha(t) = 0$ if not. $\square$

The following statement is the analog of Lemma A.6 for branched covering maps.

LEMMA A.18 (Lifting branched covering maps). *Let X, Y, and Z be surfaces, and $\pi\colon X \to Y$ be a branched covering map.*

(i) *Suppose $g_1, g_2\colon Z \to X$ are two continuous and discrete maps such that $\pi \circ g_1 = \pi \circ g_2$. If there exists a point $z_0 \in Z$ such that $g_1(z_0) = g_2(z_0) =: x_0$, and $\pi(x_0) \in Y \setminus \pi(\mathrm{crit}(\pi))$, then $g_1 = g_2$.*

(ii) *Suppose Z is simply connected, $f\colon Z \to Y$ is a branched covering map, and $x_0 \in X$ and $z_0 \in Z$ are points such that $\pi(x_0) = f(z_0)$.*
   If for all $x \in X$ and $z \in Z$ with $\pi(x) = f(z)$ we have

$$\deg(\pi, x)|\deg(f, z),$$

then there exists a branched covering map $g\colon Z \to X$ such that $g(z_0) = x_0$ and $f = \pi \circ g$.

The situation is illustrated in the following commutative diagram:

$$
\begin{array}{ccc}
 & & X \\
 & \overset{g}{\nearrow} & \downarrow \pi \\
Z & \overset{f}{\longrightarrow} & Y.
\end{array}
$$

If π is a branched covering map as in the statement, then we call a fiber $\pi^{-1}(y_0)$, $y_0 \in Y$, *clean* if it does not contain critical points of π, or equivalently, if y_0 is not a critical value of π.

The maps g_1 and g_2 in (i) are lifts of $f := \pi \circ g_1 = \pi \circ g_2$ under π. Moreover, $y_0 = \pi(x_0) \in Y \setminus \pi(\mathrm{crit}(\pi))$ is not a critical value of π and so $\pi^{-1}(y_0) \cap \mathrm{crit}(\pi) = \emptyset$. This means that $x_0 \in \pi^{-1}(y_0)$ lies in a clean fiber of π. So (i) says that under our hypotheses lifts are uniquely determined by the image of a point that maps into a clean fiber of π.

PROOF. (i) The maps g_1 and g_2 are lifts of $f := \pi \circ g_1 = \pi \circ g_2$ under the branched covering map $\pi\colon X \to Y$. We are claiming that with the given normalization such a lift is unique.

To see this, let $P_Y := \pi(\mathrm{crit}(\pi))$, $P_X := \pi^{-1}(P_Y)$, $P_Z := f^{-1}(P_Y) = g_1^{-1}(P_X) = g_2^{-1}(P_X)$. Note that these are discrete sets in Y, X, and Z, respectively. Let $X' = X \setminus P_X$, $Y' = Y \setminus P_Y$, and $Z' = Z \setminus P_Z$. Then $x_0 \in X'$, $z_0 \in Z'$, $f(Z') \subset Y' = \pi(X')$, and $g_1(Z') = g_2(Z') \subset X'$. Moreover, if we restrict π to X', then we obtain a covering map in the usual sense. Now by Lemma A.6 (i) a lift of a continuous map on a connected surface under a covering map is uniquely determined by the image of one point. Since P_Z is a discrete set, Z' is a connected surface. So $g_1(z_0) = x_0 = g_2(z_0)$ implies that $g_1|Z' = g_2|Z'$. Moreover, Z' is dense in Z, and so it follows that $g_1 = g_2$ as desired.

(ii) This is an existence statement for lifts under branched covering maps. We will derive it from an analog of the classical monodromy theorem in complex analysis (see [**Fo81**, Section 1.7]).

For this we choose a complex structure on Y. Then by Lemma A.11 there exist complex structures on X and Z such that the maps π and f are holomorphic if we equip the surfaces with these complex structures.

A *germ* (of a lift of f) at a point $z \in Z$ is a holomorphic map $g_z \colon U_z \to X$ defined on an open and connected neighborhood $U_z \subset Z$ of z such that $f|U_z = \pi \circ g_z$. We consider two such germs at z as equivalent if they agree on a neighborhood of z. Actually, one should define a germ as an equivalence class of such local lifts at z, but for the sake of easier exposition it is convenient to ignore the distinction between a local lift and the equivalence class that it represents.

The value $g_z(z)$ does not determine g_z uniquely in general, because the fiber $\pi^{-1}(f(z))$ may not be clean; this is true for nearby points z' and so it follows from (i) that the value $g_z(z')$ for a point $z' \neq z$ sufficiently close to z determines g_z uniquely.

This fact allows one to define a notion of an *analytic continuation* of a germ g_z along a path $\alpha \colon [0,1] \to Z$ with $\alpha(0) = z$. Such a continuation is given by germs $g_{\alpha(t)}$ at $\alpha(t)$ for $t \in [0,1]$ such that $g_{\alpha(0)}$ is equivalent with the given germ g_z. Moreover, we require that these germs are compatible in the following sense: if $t_0 \in [0,1]$ is arbitrary, then each germ $g_{\alpha(t)}$ for $t \in [0,1]$ close enough to t_0 is equivalent to a germ obtained by restricting $g_{\alpha(t_0)}$ to a suitable neighborhood of $\alpha(t)$.

As in the classical monodromy theorem, one can show that if an analytic continuation of g_z along a path $\alpha \colon [0,1] \to X$ exists, then up to equivalence the germ $g_{\alpha(0)}$ uniquely determines $g_{\alpha(1)}$. In this case we say that the (equivalence class of the) germ $g_{\alpha(1)}$ is obtained by analytic continuation of $g_{\alpha(0)}$ along α.

Analytic continuation of a germ along homotopic paths with the same endpoints leads to equivalent germs. More precisely, let $z, z' \in Z$, $H \colon [0,1] \times [0,1] \to Z$ be continuous, and assume that $z = H(s,0)$ and $z' = H(s,1)$ for all $s \in [0,1]$. Suppose that we have an analytic continuation of a germ g_z at z along every path $t \mapsto \alpha^s(t) := H(s,t)$, $s \in [0,1]$. Let $g_{z'}^0$ and $g_{z'}^1$ be the two germs at z' obtained by analytic continuation of g_z along α^0 and α^1, respectively. Then $g_{z'}^0$ and $g_{z'}^1$ are equivalent.

Using these facts one can now give the proof of (ii) as follows. Let $P_l(z) := z^l$ for $z \in \mathbb{C}$ and $l \in \mathbb{N}$. Our hypotheses imply that $d := \deg(\pi, x_0)$ divides $k := \deg(f, z_0)$. Let $y_0 := \pi(x_0) = f(z_0)$ and $m := k/d \in \mathbb{N}$. If we choose suitable local conformal coordinates near x_0, y_0, and z_0, then we can represent the map π near x_0 and the map f near z_0 by the power maps P_d and P_k near 0, respectively. Now $k = dm$ and so $P_k = P_d \circ P_m$. Hence for the holomorphic map g_{z_0} represented in these local coordinates by P_m we have $g_{z_0}(z_0) = x_0$ and $\pi \circ g_{z_0} = f$ near x_0. In particular, g_{z_0} is a germ of a local lift of f near z_0.

Since $\deg(\pi, x) | \deg(f, z)$ whenever $x \in X$, $z \in Z$ and $\pi(x) = f(z)$, a similar argument shows that the analytic continuation of g_{z_0} along

any path $\alpha \colon [0,1] \to X$ with $\alpha(0) = z_0$ exists, because this condition ensures that we never encounter any singularities in the analytic continuation along α.

The fact that Z is simply connected and the homotopy invariance of analytic continuation as discussed guarantees that up to equivalence the initial choice of the germ g_{z_0} leads to a unique germ g_z at every point $z \in Z$. These germs are locally compatible. So if we define $g(z) = g_z(z)$ for $z \in Z$, then g is a holomorphic map from Z to X such that $g(z_0) = x_0$ and $f = \pi \circ g$. Since π and f are branched covering maps, Lemma A.15 (ii) implies that g is a branched covering map as well. $\qquad\square$

A.7. Quotients of spaces and group actions

In this section we discuss some general facts about quotient spaces, group actions and their associated quotient spaces, and how we can pass maps to quotients. All of this material is fairly standard.

Let $\sim$ be an equivalence relation on a set X. We denote by $X/\!\!\sim$ the *quotient space* consisting of all equivalence classes $[x] \coloneqq \{y \in X : x \sim y\}$, $x \in X$, and by $\pi \colon X \to X/\!\!\sim$ the natural projection that sends each point $x \in X$ to its equivalence class $[x]$. If X is a topological space, then we equip $X/\!\!\sim$ with the *quotient topology*. In this topology a set $U \subset X/\!\!\sim$ is open if and only if $\pi^{-1}(U) \subset X$ is open. Then π is a continuous map. Moreover, a map $f \colon X/\!\!\sim \to Z$ into another topological space Z is continuous if and only if $f \circ \pi \colon X \to Z$ is continuous. Actually, this functorial property characterizes the quotient topology on $X/\!\!\sim$.

Let $\Theta \colon X \to Y$ be a map between sets X and Y. The equivalence relation on X *induced* by Θ is defined by

$$x \sim x' \; :\Leftrightarrow \; \Theta(x) = \Theta(x')$$

for $x, x' \in X$. Clearly, this is indeed an equivalence relation on X. The following well-known fact allows one to identify the quotient $X/\!\!\sim$ with Y in the cases that are relevant for us.

LEMMA A.19. *Let $\Theta \colon X \to Y$ be a continuous and surjective map between topological spaces X and Y. Let $\sim$ be the equivalence relation on X induced by Θ and $\pi \colon X \to X/\!\!\sim$ be the quotient map. Assume that*

(i) *Θ is an open map, or*

(ii) *$X/\!\!\sim$ is compact and Y is Hausdorff.*

Then there exists a unique homeomorphism $\varphi \colon X/\!\!\sim \to Y$ such that $\Theta = \varphi \circ \pi$.

PROOF. We define $\varphi([x]) = \Theta(x)$ for $[x] \in X/\sim$. Since $\sim$ is induced by Θ, it is clear that $\varphi\colon X/\sim \to Y$ is a well-defined map satisfying $\Theta = \varphi \circ \pi$. This last identity determines φ uniquely. The map φ is injective, as follows from the fact that $\sim$ is induced by Θ. It is also surjective, because Θ is, and continuous by the functorial property of the quotient topology. So φ is a continuous bijection.

Let us first assume that Θ is an open map. To show that φ is a homeomorphism, it suffices to show that φ is an open map as well. To see this, let $U \subset X/\sim$ be an arbitrary open set. Then $V := \pi^{-1}(U) \subset X$ is also open. Since

$$\varphi(U) = \varphi(\pi(V)) = \Theta(V)$$

and Θ is an open map, it follows that $\varphi(U) \subset Y$ is open as desired.

Suppose as in (ii) that $X/\sim$ is compact and Y is Hausdorff. Let $A \subset X/\sim$ be an arbitrary closed, and hence compact, set. Then $\varphi(A) \subset Y$ is compact; so this set is closed since Y is Hausdorff. Thus φ is a closed bijection which implies that φ^{-1} is continuous. The first part of the proof implies that φ is a homeomorphism as desired. $\square$

We now study when maps descend to quotients. Let $\sim$ be an equivalence relation on a set X and $\pi\colon X \to X/\sim$ be the quotient map. If $f\colon X \to X$ is a map, then we say that f *descends* to the quotient $X/\sim$ if there exists a map $\widetilde{f}\colon X/\sim \to X/\sim$ such that $\widetilde{f} \circ \pi = \pi \circ f$. In this case, we have the following commutative diagram:

$$
\begin{array}{ccc}
X & \xrightarrow{\;f\;} & X \\
{\scriptstyle \pi}\downarrow & & \downarrow{\scriptstyle \pi} \\
X/\sim & \xrightarrow{\;\widetilde{f}\;} & X/\sim.
\end{array}
$$

Here necessarily $\widetilde{f}([x]) = [f(x)]$ for $x \in X$. This shows that if f descends to $X/\sim$, then $\widetilde{f}$ is uniquely determined.

It is easy to characterize when a map f descends. The relevant condition is that $\sim$ should be *invariant* under f (or *f-invariant*), which means the implication

$$x \sim y \Rightarrow f(x) \sim f(y)$$

is satisfied for all $x, y \in X$.

LEMMA A.20 (Quotients of maps). *Let $\sim$ be an equivalence relation on a set X with the quotient map $\pi\colon X \to X/\sim$. Suppose $f\colon X \to X$ is a map.*

(i) *Then f descends to a map $\widetilde{f}$ on $X/\!\sim$ if and only if $\sim$ is invariant under f.*

(ii) *Suppose X is a topological space, and f is a continuous map that descends to the map $\widetilde{f}\colon X/\!\sim \to X/\!\sim$. Then $\widetilde{f}$ is continuous.*

PROOF. (i) Suppose $\sim$ is invariant under f. Then $\widetilde{f}\colon X/\!\sim \to X/\!\sim$ given by $\widetilde{f}([x]) := [f(x)]$ for $[x] \in X/\!\sim$ is well-defined. Obviously, $\widetilde{f} \circ \pi = \pi \circ f$ which shows that f descends to $X/\!\sim$.

Conversely, suppose that f descends to the map $\widetilde{f}$ on $X/\!\sim$. Let $x, y \in X$ with $x \sim y$ be arbitrary. Then $[x] = [y]$, and so

$$[f(x)] = (\pi \circ f)(x) = (\widetilde{f} \circ \pi)(x) = \widetilde{f}([x]) = \widetilde{f}([y]) = [f(y)].$$

Hence $f(x) \sim f(y)$, and we see that $\sim$ is invariant under f.

(ii) Suppose the continuous map $f\colon X \to X$ descends to the map $\widetilde{f}\colon X/\!\sim \to X/\!\sim$. Then $\widetilde{f} \circ \pi = \pi \circ f$ is continuous. Hence $\widetilde{f}$ is continuous by the functorial property of the quotient topology.   □

We now consider equivalence relations that are induced by group actions. We first review some general terminology.

Let X be a topological space and G be a group of homeomorphisms acting on G. The equivalence relation $\sim$ on X *induced* by G is defined by

(A.25) $x \sim y \; :\Leftrightarrow \;$ there exists $g \in G$ such that $y = g(x)$

for $x, y \in X$. The equivalence class $[x]$ of a point $x \in X$ with respect to G is equal to its *G-orbit $Gx := \{g(x) : g \in G\}$* under G. We denote the quotient space by X/G, and equip it with the quotient topology. As before, we denote by π the quotient map $\pi\colon X \to X/G$ given by $\pi(x) = [x] = Gx$ for $x \in X$.

A *fundamental domain* for the action of G is a closed set $F \subset X$ such that every orbit of G has at least one point in F, and at most one point in the interior $\mathrm{int}(F)$. If F is a fundamental domain for G, then, at least on an intuitive level, one obtains the quotient space X/G from F by identifying the points on the boundary of F that lie in the same orbit.

The *stabilizer G_x* of a point $x \in X$ is the subgroup of G consisting of all elements $g \in G$ with $g(x) = x$. We call the action *cocompact* if there exists a compact set $K \subset X$ such that $X = \bigcup_{g \in G} g(K)$. So then the images of K under the elements in G cover X, and $X/G = \pi(K)$. In particular, the quotient space X/G is compact as it is the continuous image $\pi(K)$ of the compact set K.

Suppose in addition that X is a metric space and that X is *proper*, i.e., closed balls in X are compact. Then the action of G on X is called *properly discontinuous* if for each compact set $K \subset X$ the set $\{g \in G : g(K) \cap K \neq \emptyset\}$ is finite. We call the action of G on X *geometric* if it is cocompact and properly discontinuous, and each element of G acts as an isometry on X.

Suppose Y is another topological space and $\Theta \colon X \to Y$ is a continuous and surjective map. We say that Θ is *induced* by the action of a group G acting on X if $\Theta(x) = \Theta(y)$ for $x, y \in X$ if and only if there exists $g \in G$ such $y = g(x)$. In this case, the equivalence relation on X induced by Θ is the same as the equivalence relation induced by G.

The following statement immediately follows from Lemma A.19.

COROLLARY A.21. *Let X and Y be topological spaces, G be a group acting on X, $\pi \colon X \to X/G$ be the quotient map, and $\Theta \colon X \to Y$ be a continuous and surjective map induced by G. Assume*

(i) *Θ is an open map, or*

(ii) *G acts cocompactly on X and Y is Hausdorff.*

Then there exists a unique homeomorphism $\varphi \colon X/G \to Y$ such that $\Theta = \varphi \circ \pi$.

Note that in case (ii) the quotient space X/G is compact and so Lemma A.19 indeed applies.

Under the assumptions of Corollary A.21 one can identify Y with the quotient space X/G by the homeomorphism φ, and Θ with the quotient map $\pi \colon X \to X/G$.

Let us now consider when a map $f \colon X \to X$ descends to the quotient X/G. A relevant condition is that f is *G-equivariant*. This means that for each $g \in G$ there exists $h \in G$ such that

$$(A.26) \qquad\qquad\qquad f \circ g = h \circ f.$$

If $f \colon X \to X$ is a bijection, then obviously f is G-equivariant if and only if

$$(A.27) \qquad\qquad f \circ g \circ f^{-1} \in G \text{ for each } g \in G.$$

We now formulate a criterion in a special case relevant for us.

LEMMA A.22. *Let X and Y be surfaces, $f \colon X \to X$ be a homeomorphism, and $\Theta \colon X \to Y$ be a branched covering map induced by an action of a group G on X. Then there exists a continuous map $\widetilde{f} \colon Y \to Y$ such that $\Theta \circ f = \widetilde{f} \circ \Theta$ if and only if f is G-equivariant.*

Note that Θ is an open map, and so by Corollary A.21 we can identify Y with the quotient space X/G. Under this identification the lemma gives a condition for the homeomorphism f to descend to X/G.

As we will see in the proof, the "if" implication in this statement is valid in much greater generality.

PROOF. Suppose first that f is G-equivariant. Let $\sim$ be the equivalence relation on X induced by G. Then by Corollary A.21 and Lemma A.20 it suffices to show that $\sim$ is invariant under f. So let $x, y \in X$ and $x \sim y$. Then there exists $g \in G$ such that $y = g(x)$. Since f is G-equivariant, there exists $h \in G$ such that $f \circ g = h \circ f$. Hence

$$f(y) = f(g(x)) = h(f(x)) \sim f(x)$$

as desired.

For the proof of the other implication we assume that a continuous map $\widetilde{f} \colon Y \to Y$ with $\Theta \circ f = \widetilde{f} \circ \Theta$ exists. In order to show that f is G-equivariant, let $g \in G$ be arbitrary. We can pick a point $x_0 \in X$ so that the homeomorphism $f \circ g$ maps x_0 to a clean fiber of Θ, i.e., $(\Theta \circ f \circ g)(x_0)$ is not a critical value of Θ. Then

$$(\Theta \circ f \circ g)(x_0) = (\widetilde{f} \circ \Theta \circ g)(x_0)$$
$$= (\widetilde{f} \circ \Theta)(x_0) = (\Theta \circ f)(x_0).$$

Since Θ is induced by G, it follows that there exists $h \in G$ such that

$$(A.28) \qquad\qquad f(g(x_0)) = h(f(x_0)).$$

Now

$$\Theta \circ f \circ g = \widetilde{f} \circ \Theta \circ g = \widetilde{f} \circ \Theta = \Theta \circ f = \Theta \circ h \circ f,$$

and so by (A.28) the homeomorphisms $f \circ g$ and $h \circ f$ satisfy the conditions of Lemma A.18 (i) for the branched covering map Θ. Hence $f \circ g = h \circ f$. This shows that f is G-equivariant. $\square$

A.8. Tori and lattices

In this section we review some facts about and tori and lattices.

A *lattice* $\Gamma \subset \mathbb{R}^2$ is a non-trivial discrete subgroup of $\mathbb{R}^2$ (considered as a group with vector addition). Often it is more convenient to consider a lattice as a discrete (additive) subgroup of $\mathbb{C} \cong \mathbb{R}^2$. In the following, we will freely switch back and forth between these different viewpoints and real and complex notation.

The *rank* of a lattice is the dimension of the subspace of $\mathbb{R}^2$ (considered as a real vector space) spanned by the elements in Γ. The lattice can be a *rank-1 lattice*. Then, using complex notation, we can write

it in the form $\Gamma = \mathbb{Z}\omega$, where $\omega \in \mathbb{C} \setminus \{0\}$. The other case is that Γ is a *rank-2 lattice*. Then there exist generators $\omega_1, \omega_2 \in \mathbb{C} \setminus \{0\}$ with $\mathrm{Im}(\omega_2/\omega_1) > 0$ such that

$$\Gamma = \mathbb{Z}\omega_1 \oplus \mathbb{Z}\omega_2 := \{m\omega_1 + n\omega_2 : m, n \in \mathbb{Z}\}.$$

If $\Gamma \subset \mathbb{R}^2 \cong \mathbb{C}$ is a rank-2 lattice, then we have an equivalence relation $\sim$ on $\mathbb{R}^2$ given by

(A.29) $x \sim y :\Leftrightarrow x - y \in \Gamma$

for $x, y \in \mathbb{R}^2$. We denote the quotient space $\mathbb{R}^2/\!\sim$ by $\mathbb{R}^2/\Gamma$, and by $\pi\colon \mathbb{R}^2 \to \mathbb{R}^2/\Gamma$ the quotient map that sends a point $x \in \mathbb{R}^2$ to its equivalence class $[x]$. In complex notation, we use $\mathbb{C}/\Gamma$ to denote the quotient space; the quotient map is then a map $\pi\colon \mathbb{C} \to \mathbb{C}/\Gamma$.

For each $\gamma \in \Gamma$ we can define an associated translation

$$\tau_\gamma\colon \mathbb{R}^2 \to \mathbb{R}^2, \ u \in \mathbb{R}^2 \mapsto \tau_\gamma(u) := u + \gamma, \ \gamma \in \Gamma.$$

The translations τ_γ, $\gamma \in \Gamma$, form a group under composition that is isomorphic to Γ. The equivalence relation on $\mathbb{R}^2$ induced by the action of this group as in (A.25) is the same as in (A.29).

We equip $\mathbb{R}^2/\Gamma$ with the quotient topology. Then $T^2 = \mathbb{R}^2/\Gamma$ is a 2-dimensional torus, and $\pi\colon \mathbb{R}^2 \to T^2 = \mathbb{R}^2/\Gamma$ is a covering map; actually, since we insist on covering maps being orientation-preserving, we have to equip T^2 with a suitable orientation here. The lattice translations τ_γ, $\gamma \in \Gamma$, are deck transformations of the quotient map π and so $\pi = \pi \circ \tau_\gamma$ for $\gamma \in \Gamma$. We will see momentarily that actually every deck transformation of π has this form (Lemma A.23 (i)).

The conformal structure on $\mathbb{C}$ induces a unique conformal structure on $\mathbb{R}^2/\Gamma = \mathbb{C}/\Gamma$. It is represented by a complex atlas on $\mathbb{C}/\Gamma$ given by suitable local inverse branches of π. Then $\mathbb{T} = \mathbb{C}/\Gamma$ is a complex torus and $\pi\colon \mathbb{C} \to \mathbb{T}$ a holomorphic map. Such a complex torus will always be denoted by $\mathbb{T}$, whereas T^2 will denote a torus that is not equipped with a complex structure, i.e., a topological torus.

Every torus can be represented in the form $\mathbb{R}^2/\Gamma$ up to orientation-preserving homeomorphisms. Actually, we can choose $\Gamma = \mathbb{Z}^2$ if this is convenient. Up to conformal equivalence every complex torus has a representation of the form $\mathbb{C}/\Gamma$. Here two complex tori $\mathbb{T} = \mathbb{C}/\Gamma$ and $\mathbb{T}' = \mathbb{C}/\Gamma'$ obtained from rank-2 lattices $\Gamma, \Gamma' \subset \mathbb{C}$ are conformally equivalent if and only if there exists $\alpha \in \mathbb{C} \setminus \{0\}$ such that $\Gamma' = \alpha\Gamma$.

In the following lemma we collect various statements that are used in Chapter 3.

LEMMA A.23. *Let* $\Gamma \subset \mathbb{R}^2$ *be a rank-2 lattice,* $T^2 = \mathbb{R}^2/\Gamma$, *and* $\pi\colon \mathbb{R}^2 \to T^2 = \mathbb{R}^2/\Gamma$ *be the quotient map.*

(i) *For a continuous map $\varphi\colon \mathbb{R}^2 \to \mathbb{R}^2$ we have $\pi \circ \varphi = \pi$ if and only if there exists $\gamma \in \Gamma$ such that $\varphi = \tau_\gamma$.*

(ii) *If $\overline{A}\colon T^2 \to T^2$ is a torus endomorphism, then $\overline{A}$ can be lifted to a homeomorphism on $\mathbb{R}^2$, i.e., there exists a homeomorphism $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ such that $\overline{A} \circ \pi = \pi \circ A$. The homeomorphism A is orientation-preserving, and unique up to postcomposition by a translation τ_γ, $\gamma \in \Gamma$.*

(iii) *If $\overline{A}\colon T^2 \to T^2$ is a torus endomorphism, then there exists a unique $\mathbb{R}$-linear map $L\colon \mathbb{R}^2 \to \mathbb{R}^2$ with $L(\Gamma) \subset \Gamma$ such that for every lift A as in (ii) we have*

(A.30)
$$A \circ \tau_\gamma \circ A^{-1} = \tau_{L(\gamma)} = L \circ \tau_\gamma \circ L^{-1}$$

for all $\gamma \in \Gamma$.

(iv) *If $\overline{A}\colon T^2 \to T^2$ is a torus endomorphism and L the map as in (iii), then $\deg(\overline{A}) = \det(L)$.*

Statement (i) implies that the deck transformations of π are precisely the lattice translations τ_γ, $\gamma \in \Gamma$.

For (ii) recall that a *torus endomorphism* $\overline{A}\colon T^2 \to T^2$ is an orientation-preserving local homeomorphism on a torus (see the introduction of Chapter 3). From the statement we get a commutative diagram of the form:

(A.31)
$$\begin{array}{ccc} \mathbb{R}^2 & \xrightarrow{\;A\;} & \mathbb{R}^2 \\ {\scriptstyle \pi}\downarrow & & \downarrow{\scriptstyle \pi} \\ T^2 & \xrightarrow{\;\overline{A}\;} & T^2. \end{array}$$

The map L in (iii) can be viewed as the map induced by $\overline{A}$ on the fundamental group of $T^2 = \mathbb{R}^2/\Gamma$. Indeed, if $x_0 \in T^2$ and $y_0 := \overline{A}(x_0)$, then $\overline{A}$ induces a map $\overline{A}_*\colon \pi_1(T^2, x_0) \to \pi_1(T^2, y_0)$ (see the end of Section A.5). Moreover, we have a natural isomorphism $\pi_1(T^2, x_0) \cong \Gamma$. Namely, if $[\ell] \in \pi_1(T^2, x_0)$ is an element of the fundamental group represented by a loop ℓ based at x_0, then we can lift ℓ to a path $\tilde{\ell}$ on $\mathbb{R}^2$ under π. In general, $\tilde{\ell}$ is not a loop; if $u_0 \in \mathbb{R}^2$ is the initial point of $\tilde{\ell}$ and $v_0 \in \mathbb{R}^2$ the other endpoint of $\tilde{\ell}$, then $\gamma := v_0 - u_0 \in \Gamma$, and one can show that the map $[\ell] \mapsto \gamma$ gives a well-defined group isomorphism $\pi_1(T^2, x_0) \to \Gamma$. Similarly, we have a natural isomorphism $\pi_1(T^2, y_0) \to \Gamma$. Under these isomorphisms, the map $A_*\colon \pi_1(T^2, x_0) \to \pi_1(T^2, y_0)$ corresponds to the homomorphism $L\colon \Gamma \to \Gamma$.

To see this, note that in our setting $A \circ \tilde{\ell}$ is a lift of the image loop $\overline{A} \circ \ell$ representing $A_*([\ell])$. The path $A \circ \tilde{\ell}$ has the endpoints $A(u_0)$ and

$A(v_0)$. So under the isomorphism $\pi_1(T^2, x_0) \to \Gamma$ and $\pi_1(T^2, y_0) \to \Gamma$ the map $[\ell] \mapsto A_*([\ell]) = [\overline{A} \circ \ell]$ corresponds to the map

$$\gamma \in \Gamma \mapsto A(v_0) - A(u_0) = (A \circ \tau_\gamma)(u_0) - A(u_0)$$
$$= (\tau_{L(\gamma)} \circ A)(u_0) - A(u_0) = L(\gamma)$$

by (A.30).

PROOF OF LEMMA A.23. (i) Let $\varphi \colon \mathbb{R}^2 \to \mathbb{R}^2$ be a continuous map with $\pi \circ \varphi = \pi$. Then we can consider φ as a lift of $\pi \colon \mathbb{R}^2 \to T^2$ under the covering map $\pi \colon \mathbb{R}^2 \to T^2$, because we have the commutative diagram

$$\begin{array}{ccc} & & \mathbb{R}^2 \\ & {}^{\varphi}\nearrow & \downarrow{}^{\pi} \\ \mathbb{R}^2 & \xrightarrow{\;\pi\;} & T^2. \end{array}$$

Fix $u_0 \in \mathbb{R}^2$, and define $v_0 = \varphi(u_0)$. Then $\pi(v_0) = \pi(\varphi(u_0)) = \pi(u_0)$. This means that $u_0 \sim v_0$ are equivalent with respect to the equivalence relation $\sim$ induced by Γ. Thus, there exists $\gamma \in \Gamma$ such that $v_0 = u_0 + \gamma$. Consider the corresponding lattice translation τ_γ. Then $\pi \circ \tau_\gamma = \pi$, and so τ_γ is also a lift of π under the covering map π. Since $\tau_\gamma(u_0) = v_0 = \varphi(u_0)$, Lemma A.6 (i) implies that $\varphi = \tau_\gamma$.

(ii) We can lift the map $\overline{A} \circ \pi \colon \mathbb{R}^2 \to T^2$ under the covering map π to get a continuous map $A \colon \mathbb{R}^2 \to \mathbb{R}^2$ satisfying $\pi \circ A = \overline{A} \circ \pi$ (Lemma A.18 (ii)).

In order to find an inverse B of A we will reverse the roles of π and $\overline{A} \circ \pi$. First note that $\overline{A} \circ \pi \colon \mathbb{R}^2 \to T^2$ is also a covering map.

Fix $u_0 \in \mathbb{R}^2$, and define $v_0 := A(u_0) \in \mathbb{R}^2$. Then

$$\pi(v_0) = \pi(A(u_0)) = \overline{A}(\pi(u_0)).$$

So by lifting $\pi \colon \mathbb{R}^2 \to T^2$ under the covering map $\overline{A} \circ \pi$ we can find a continuous map $B \colon \mathbb{R}^2 \to \mathbb{R}^2$ with $B(v_0) = u_0$ and $\pi = \overline{A} \circ \pi \circ B$; so we obtain the commutative diagram

$$\begin{array}{ccc} u_0 \in \mathbb{R}^2 & \underset{\underset{B}{\longleftarrow}}{\overset{A}{\longrightarrow}} & v_0 \in \mathbb{R}^2 \\ {}_{\overline{A} \circ \pi}\searrow & & \swarrow{}_{\pi} \\ & T^2. & \end{array}$$

Then $(A \circ B)(v_0) = A(u_0) = v_0$ and

$$\pi \circ (A \circ B) = \overline{A} \circ \pi \circ B = \pi = \pi \circ \mathrm{id}_{\mathbb{R}^2}.$$

So we conclude $A \circ B = \mathrm{id}_{\mathbb{R}^2}$ by Lemma A.18 (i) applied to the covering map π. Similarly, $(B \circ A)(u_0) = B(v_0) = u_0$ and

$$\overline{A} \circ \pi \circ B \circ A = \pi \circ A = \overline{A} \circ \pi = \overline{A} \circ \pi \circ \mathrm{id}_{\mathbb{R}^2},$$

which implies $B \circ A = \mathrm{id}_{\mathbb{R}^2}$ by the same lemma applied to the covering map $\overline{A} \circ \pi$. It follows that A is a homeomorphism on $\mathbb{R}^2$ with inverse B.

The relation $\pi \circ A = \overline{A} \circ \pi$ shows that locally A can be written as a composition of $\overline{A} \circ \pi$ and a branch of π^{-1}. Since these maps are orientation-preserving, A has this property as well. These considerations show that $\overline{A}$ has a lift A as desired.

Suppose $A' \colon \mathbb{R}^2 \to \mathbb{R}^2$ is another continuous map with $\pi \circ A' = \overline{A} \circ \pi$. Then

$$\pi \circ (A' \circ A^{-1}) = \pi \circ A' \circ B = \overline{A} \circ \pi \circ B = \pi.$$

By (i) there exists $\tau \in \Gamma$ such that $A' \circ A^{-1} = \tau_\gamma$. Then $A' = \tau_\gamma \circ A$. This shows that A is unique up to postcomposition by a lattice translation. Note that clearly every map of the form $\tau_\gamma \circ A$, $\tau \in \Gamma$, is actually a lift of $\overline{A}$.

(iii) Let A be a homeomorphic lift of $\overline{A}$ as in (ii). If $\gamma \in \Gamma$ is arbitrary, then

$$\pi \circ A \circ \tau_\gamma \circ A^{-1} = \overline{A} \circ \pi \circ \tau_\gamma \circ A^{-1}$$
$$= \overline{A} \circ \pi \circ A^{-1} = \pi \circ A \circ A^{-1} = \pi.$$

So $A \circ \tau_\gamma \circ A^{-1}$ is a deck transformation of the covering map π. By (i) this implies that there exists a unique $\gamma' \in \Gamma$ such that

$$A \circ \tau_\gamma \circ A^{-1} = \tau_{\gamma'}.$$

Note that for $\gamma, \sigma \in \Gamma$ we have

$$A \circ \tau_{\gamma+\sigma} \circ A^{-1} = A \circ \tau_\gamma \circ \tau_\sigma \circ A^{-1} = \tau_{\gamma'} \circ A \circ \tau_\sigma \circ A^{-1}$$
$$= \tau_{\gamma'} \circ \tau_{\sigma'} = \tau_{\gamma'+\sigma'}.$$

Moreover, if

$$\tau_{\gamma'} = A \circ \tau_\gamma \circ A^{-1} = \tau_0 = \mathrm{id}_{\mathbb{R}^2},$$

then

$$\tau_\gamma = A^{-1} \circ \mathrm{id}_{\mathbb{R}^2} \circ A = \mathrm{id}_{\mathbb{R}^2},$$

and so $\gamma = 0$. This implies that the map $\gamma \mapsto \gamma'$ gives an injective group homomorphism $L \colon \Gamma \to \Gamma$ such that $A \circ \tau_\gamma \circ A^{-1} = \tau_{L(\gamma)}$ for all $\gamma \in \Gamma$. We can uniquely extend L to an invertible $\mathbb{R}$-linear map on $\mathbb{R}^2$, also denoted by L. Then $L(\Gamma) \subset \Gamma$ and the first equality in (A.30) is true by definition of L. The second equation in (A.30) is actually true for *every* invertible linear map L. Uniqueness of L is clear.

(iv) Let $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ be a lift of $\overline{A}$ as in (ii) and L be the linear map as in (iii). We define a homotopy $H\colon \mathbb{R}^2 \times [0,1] \to \mathbb{R}^2$ as

$$H(u,t) = (1-t)A(u) + tL(u), \quad u \in \mathbb{R}^2, \, t \in [0,1].$$

Let $H_t := H(\cdot, t)$ for $t \in [0,1]$. Then $H_0 = A$ and $H_1 = L$. For all $u \in \mathbb{R}^2$, $t \in [0,1]$, and $\gamma \in \Gamma$ we have

$$
\begin{aligned}
(H_t \circ \tau_\gamma)(u) &= (1-t)A(\tau_\gamma(u)) + tL(\tau_\gamma(u)) \\
&= (1-t)\tau_{L(\gamma)}(A(u)) + t\tau_{L(\gamma)}(L(u)) \\
&= (1-t)A(u) + tL(u) + L(\gamma) = (\tau_{L(\gamma)} \circ H_t)(u).
\end{aligned}
$$

Here we used (A.30). Hence $H_t \circ \tau_\gamma = \tau_{L(\gamma)} \circ H_t$ for all $t \in [0,1]$ and $\gamma \in \Gamma$. This implies that the equivalence relation $\sim$ induced by Γ is invariant under H_t, and so H_t descends to a continuous map $\overline{H}_t$ on $T^2 = \mathbb{R}^2/\Gamma$ (see Lemma A.20). Then $\pi \circ H_t = \overline{H}_t \circ \pi$ for all $t \in [0,1]$. In particular,

$$\overline{A} \circ \pi = \pi \circ A = \pi \circ H_0 = \overline{H}_0 \circ \pi,$$

which implies $\overline{H}_0 = \overline{A}$.

Define $\overline{H}\colon T^2 \times [0,1] \to T^2$ as $\overline{H}(x,t) = \overline{H}_t(x)$ for $(x,t) \in T^2 \times [0,1]$. Then $\overline{H}$ is continuous. Indeed, if $\{(x_n, t_n)\}$ is a sequence in $T^2 \times [0,1]$ with $(x_n, t_n) \to (x,t) \in T^2 \times [0,1]$ as $n \to \infty$, then by considering a local inverse of π near (x,t), we can find points $u \in \mathbb{R}^2$ and $u_n \in \mathbb{R}^2$ for $n \in \mathbb{N}$ with $\pi(u) = x$, $\pi(u_n) = x_n$ and $u_n \to u$ as $n \to \infty$. Then as $n \to \infty$,

$$
\begin{aligned}
\overline{H}(x_n, t_n) &= \overline{H}_{t_n}(\pi(u_n)) = \pi(H_{t_n}(u_n)) = \pi(H(u_n, t_n)) \to \pi(H(u,t)) \\
&= \pi(H_t(u)) = \overline{H}_t(\pi(u)) = H(x,t)
\end{aligned}
$$

by the continuity of H and π.

The map $\overline{H}$ is a homotopy on T^2 with the time-t maps $\overline{H}_t$. In particular, the maps $\overline{A} = \overline{H}_0$ and $\overline{L} := \overline{H}_1$ on T^2 are homotopic. Since degrees of maps are invariant under homotopies, it follows that $\deg(\overline{A}) = \deg(\overline{L})$. Note that $\overline{L} \circ \pi = \overline{H}_1 \circ \pi = \pi \circ H_1 = \pi \circ L$. So $\overline{L}$ is a map on T^2 induced by the linear map L.

It is a standard fact that then $\deg(\overline{L}) = \det(L)$. One can see this as follows by using some basic concepts from differential geometry. Namely, we can consider $\mathbb{T} = \mathbb{R}^2/\Gamma$ as a complex torus and hence as a smooth manifold. Then $\overline{L}$ is a smooth map on $\mathbb{T}$. There exists a unique 2-form α on $\mathbb{T}$ whose pull-back $\pi^*(\alpha)$ is equal to the standard volume form on $\mathbb{R}^2$ given by $\omega = dx \wedge dy$, where x and y are the standard coordinates on $\mathbb{R}^2$. If we pull ω back by the linear map L, we get $L^*(\omega) = \det(L)\omega$. This implies that $\overline{L}^*(\alpha) = \det(L)\alpha$. On the other

hand,

$$\int_{\mathbb{T}} \overline{L}^*(\alpha) = \deg(\overline{L}) \int_{\mathbb{T}} \alpha.$$

Hence

$$\det(L) \int_{\mathbb{T}} \alpha = \int_{\mathbb{T}} \overline{L}^*(\alpha) = \deg(\overline{L}) \int_{\mathbb{T}} \alpha.$$

Since $\int_{\mathbb{T}} \alpha \neq 0$ this implies that $\deg(\overline{L}) = \det(L)$. We conclude that

$$\deg(\overline{A}) = \deg(\overline{L}) = \det(L)$$

as desired. □

In Lemma A.23 we can consider a complex torus $\mathbb{T} = \mathbb{C}/\Gamma$. Then the quotient map $\pi \colon \mathbb{C} \to \mathbb{T} = \mathbb{C}/\Gamma$ is holomorphic. Suppose $\overline{A} \colon \mathbb{T} \to \mathbb{T}$ is a holomorphic torus endomorphism, and $A \colon \mathbb{C} \to \mathbb{C}$ is a homeomorphic lift of $\overline{A}$ as in Lemma A.23 (ii). Then $\pi \circ A = \overline{A} \circ \pi$. This shows that locally the map A is given as a composition of the holomorphic map $\overline{A} \circ \pi$ with a branch of π^{-1}, which is also holomorphic. We conclude that $A \colon \mathbb{C} \to \mathbb{C}$ is holomorphic itself (we can also apply Lemma A.15 here). Since this map is also a homeomorphism on $\mathbb{C}$ it must be of the form $A(z) = \alpha z + \beta$ for $z \in \mathbb{C}$, where $\alpha, \beta \in \mathbb{C}$, $\alpha \neq 0$. The associated linear map $L \colon \mathbb{C} \to \mathbb{C}$ as in Lemma A.23 (iii) is then given by $L(z) = \alpha z$ for $z \in \mathbb{C}$. By part (iv) of this lemma we have $\deg(\overline{A}) = \det(L) = |\alpha|^2$ (note that here we have to consider L as an $\mathbb{R}$-linear map).

A.9. Orbifolds and coverings

In this section we will discuss the relation of orbifolds and branched covering maps. Recall (see Section 2.5) that an *orbifold* is a pair $\mathcal{O} = (S, \alpha)$, where S is a surface and $\alpha \colon S \to \widehat{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$ a ramification function on S.

We will restrict ourselves to the case relevant for Thurston maps, where the underlying surface of the orbifold $\mathcal{O} = (S, \alpha)$ is an oriented topological 2-sphere $S = S^2$. Then $\operatorname{supp}(\alpha) = \{p \in S^2 : \alpha(p) \geq 2\}$ is a finite set. A point $p \in \operatorname{supp}(\alpha)$ is called a *cone point* if $2 \leq \alpha(p) < \infty$ and a *puncture* if $\alpha(p) = \infty$.

Often the underlying sphere S^2 of the orbifold will be the Riemann sphere $\widehat{\mathbb{C}}$. In this case, the orbifold $(\widehat{\mathbb{C}}, \alpha)$ has an underlying conformal structure, whereas in general we consider (S^2, α) as a purely topological object.

Recall (see Section 2.5) that the *Euler characteristic* of an orbifold $\mathcal{O} = (S^2, \alpha)$ is defined as

$$(A.32) \qquad \chi(\mathcal{O}) = 2 - \sum_{p \in S^2} \left(1 - \frac{1}{\alpha(p)} \right),$$

where we use the convention $1/\infty = 0$. The orbifold $\mathcal{O}$ is *parabolic* if $\chi(\mathcal{O}) = 0$ and *hyperbolic* if $\chi(\mathcal{O}) < 0$

One can give a geometric interpretation of $\chi(\mathcal{O})$ also as follows. Let $\mathcal{D}$ be a cell decomposition of S^2 (see Section 5.1) such that each point $p \in \operatorname{supp}(\alpha)$ is a vertex of $\mathcal{D}$. We count each vertex p in $\mathcal{D}$ with weight $1/\alpha(p)$, and define

$$(A.33) \qquad \#V = \sum \frac{1}{\alpha(p)},$$

where the sum is taken over all vertices p in $\mathcal{D}$. Then it easily follows from Euler's polyhedral formula that

$$\chi(\mathcal{O}) = \#F - \#E + \#V,$$

where $\#F$ is the number of faces (i.e., 2-cells) and $\#E$ is the number of edges (i.e., 1-cells) in $\mathcal{D}$. So with the slight modification (A.33) the Euler characteristic of an orbifold is given by the usual formula. Note that punctures do not contribute to $\#V$ in (A.33).

We will sometimes remove the punctures of the orbifold (S^2, α) from the 2-sphere. With the ramification function α understood, we set

$$(A.34) \qquad S_0^2 = S^2 \setminus \{p \in S^2 : \alpha(p) = \infty\}.$$

In particular, we use this notation if $S^2 = \widehat{\mathbb{C}}$ and so $\widehat{\mathbb{C}}_0$ denotes the Riemann sphere without the punctures of the given orbifold.

The following statement relates orbifolds and branched covering maps.

THEOREM A.24. *Let $\mathcal{O} = (S^2, \alpha)$ be an orbifold that is parabolic or hyperbolic. Then the following statements are true:*

(i) *There exist a simply connected surface X and a branched covering map $\Theta \colon X \to S_0^2$ such that*

$$\deg(\Theta, x) = \alpha(\Theta(x))$$

for each $x \in X$;

(ii) *if $S^2 = \widehat{\mathbb{C}}$, then we may in addition assume that $X = \mathbb{C}$ if $\mathcal{O}$ is parabolic and $X = \mathbb{D}$ if $\mathcal{O}$ is hyperbolic and that the map $\Theta \colon X \to \widehat{\mathbb{C}}_0$ is holomorphic.*

We will not provide a proof here. The statement follows from [**Be61**, Theorem 1]. See also [**La09**, 4.8.2], [**Th**, Proposition 13.2.4 and Theorem 13.3.6], and (for a much more general situation) [**BH99**].

Part (i) of the previous theorem implies that $\deg(\Theta, x) = \deg(\Theta, y)$ for all $x, y \in X$ with $\Theta(x) = \Theta(y)$, i.e., the local degree of Θ is constant in each fiber $\Theta^{-1}(p)$, $p \in S_0^2$. Note that Θ does not cover the punctures of $\mathcal{O}$.

Theorem A.24 (ii) explains why we call orbifolds parabolic or hyperbolic. Here the additional conformal structures are important, because $\mathbb{C}$ and $\mathbb{D}$ are topologically indistinguishable.

DEFINITION A.25. For a parabolic or hyperbolic orbifold $\mathcal{O} = (S^2, \alpha)$ the map $\Theta \colon X \to S_0^2$ from Theorem A.24 is called the *universal orbifold covering map* of $\mathcal{O}$.

We will momentarily see that up to equivalence the map Θ is uniquely determined. First we formulate the universal property of Θ (for a closely related statement see [**Th**, Proposition 13.2.4]).

THEOREM A.26. *Let $\mathcal{O} = (S^2, \alpha)$ be a parabolic or hyperbolic orbifold with universal orbifold covering map $\Theta \colon X \to S_0^2$, Z be a surface, and $f \colon Z \to S_0^2$ be a branched covering map such that*

$$\deg(f, z) | \alpha(f(z))$$

for each $z \in Z$.

Then for all points $x_0 \in X$ and $z_0 \in Z$ with $p_0 := \Theta(x_0) = f(z_0)$ there exists a branched covering map $\varphi \colon X \to Z$ such that $\varphi(x_0) = z_0$ and $\Theta = f \circ \varphi$. Moreover, if $\alpha(p_0) = 1$, then the map φ with these properties is unique.

So the following diagram commutes:

$$
\begin{array}{ccc}
 & X & \\
 \varphi \swarrow & & \downarrow \Theta \\
 Z & \xrightarrow{\ f\ } & S_0^2.
\end{array}
$$

PROOF. Note that if $x \in X$, $z \in Z$, and $p := \Theta(x) = f(z)$, then

$$\deg(f, z) | \alpha(p) = \deg(\Theta, x).$$

It follows that we may lift Θ by f to yield the desired map φ according to Lemma A.18 (ii) (note that f plays the role of π in this lemma).

If in addition $\alpha(p_0) = 1$, then $\deg(f, z) = 1$ for each point $z \in f^{-1}(p_0)$, and so the fiber $f^{-1}(p_0)$ is clean. Hence Lemma A.18 (i) implies that with the stated properties the map φ is uniquely determined. $\square$

COROLLARY A.27 (Uniqueness of the universal orbifold cover). *Let $\mathcal{O} = (S^2, \alpha)$ be an orbifold that is parabolic or hyperbolic and $\Theta \colon X \to S_0^2$ and $\widetilde{\Theta} \colon \widetilde{X} \to S_0^2$ be universal orbifold covering maps. Then for all points $x_0 \in X$ and $\tilde{x}_0 \in \widetilde{X}$ with $p_0 := \Theta(x_0) = \widetilde{\Theta}(\tilde{x}_0)$ there exists an orientation-preserving homeomorphism $A \colon X \to \widetilde{X}$ with $A(x_0) = \tilde{x}_0$ and $\Theta = \widetilde{\Theta} \circ A$. Moreover, if $\alpha(p_0) = 1$, then A is unique.*

So the universal orbifold covering map of $\mathcal{O}$ is unique up to precomposition by an orientation-preserving homeomorphism.

PROOF. We can apply Theorem A.26 to the universal covering map $\Theta \colon X \to S_0^2$ and the branched covering map $\widetilde{\Theta} \colon \widetilde{X} \to S_0^2$. This gives the existence of a branched covering map $A \colon X \to \widetilde{X}$ with $A(x_0) = \tilde{x}_0$ and $\Theta = \widetilde{\Theta} \circ A$.

To show that A is a homeomorphism, we have to construct an inverse for A. For this we want to reverse the roles of Θ and $\widetilde{\Theta}$, and apply the uniqueness statement in Theorem A.26. Here a complication is that the fibers of the maps above p_0 are not clean if $\alpha(p_0) \geq 2$. So we choose a point $p_1 \in S^2$ with $\alpha(p_1) = 1$ and a point $x_1 \in \Theta^{-1}(p_1)$ and set $\tilde{x}_1 = A(x_1)$. Then again by Theorem A.26 there exists a branched covering map $B \colon \widetilde{X} \to X$ with $B(\tilde{x}_1) = x_1$ and $\widetilde{\Theta} = \Theta \circ B$.

Then $\widetilde{\Theta} = \widetilde{\Theta} \circ A \circ B$ and $\Theta = \Theta \circ B \circ A$. Since $(A \circ B)(\tilde{x}_1) = \tilde{x}_1$, $(B \circ A)(x_1) = x_1$, and

$$\alpha(\Theta(x_1)) = \alpha(\widetilde{\Theta}(\tilde{x}_1)) = \alpha(p_1) = 1,$$

it follows from the uniqueness statement in Theorem A.26 that $A \circ B = \mathrm{id}_{\widetilde{X}}$ and $B \circ A = \mathrm{id}_X$. Therefore, A is a homeomorphism. As a branched covering map, A is necessarily orientation-preserving.

If $\alpha(p_0) = 1$, then the uniqueness of A again follows from Theorem A.26.   □

REMARK A.28. Theorem A.26 remains valid in the holomorphic setting with the obvious changes in the formulation. Namely, suppose that in addition to the hypotheses in the statement $S^2 = \widehat{\mathbb{C}}$, X and Z are Riemann surfaces, and Θ and f are holomorphic. Then the map φ with $\Theta = f \circ \varphi$ is also holomorphic. This follows from the last part of Lemma A.15.

Similarly, in the holomorphic setting in Corollary A.27 the map A will be holomorphic and hence a biholomorphism. In particular, if Θ is the universal orbifold covering map as in Theorem A.24 (ii), then Θ is unique up to precomposition by a suitable biholomorphism.

Let $\mathcal{O} = (S^2, \alpha)$ be a parabolic or hyperbolic orbifold and $\Theta \colon X \to S^2_0$ be the universal orbifold covering map. A homeomorphism $g \colon X \to X$ is called a *deck transformation* of Θ if $\Theta \circ g = \Theta$.

Each deck transformation is orientation-preserving. The deck transformations of Θ form a group under composition, denoted by $\pi_1(\mathcal{O})$ and called the *fundamental group* of the orbifold $\mathcal{O}$.

We collect properties of the group of deck transformations in the following statement.

PROPOSITION A.29 (Deck transformations of the universal orbifold cover). *Let $\mathcal{O} = (S^2, \alpha)$ be an orbifold that is parabolic or hyperbolic, and $\Theta \colon X \to \mathcal{O}$ be the universal orbifold covering map. Then for the group of deck transformations $G = \pi_1(\mathcal{O})$ the following statements are true:*

(i) *The map Θ is induced by G, meaning that for all $x, y \in X$ we have*

$\Theta(x) = \Theta(y)$ *if and only if there exists $g \in G$ with $y = g(x)$.*

(ii) *For all $x \in X$ the stabilizer $G_x = \{g \in G : g(x) = x\}$ is a finite cyclic group of order*

$$\#G_x = \deg(\Theta, x) = \alpha(\Theta(x)).$$

(iii) *G acts properly discontinuously on X.*

(iv) *If $\mathcal{O}$ has no punctures, then G acts cocompactly on X.*

Assume in addition that $S^2 = \widehat{\mathbb{C}}$, that $X = \mathbb{C}$ or $X = \mathbb{D}$ depending on whether $\mathcal{O}$ is parabolic or hyperbolic, and that $\Theta \colon X \to \widehat{\mathbb{C}}$ is holomorphic.

(v) *Each $g \in G$ acts as a biholomorphism on X. Moreover, if we equip $X = \mathbb{C}$ with the Euclidean metric and $X = \mathbb{D}$ with the hyperbolic metric, then each $g \in G$ acts on X as an isometry.*

Note in particular that by (i) the branched covering map Θ is *regular* in the sense that its group G of deck transformations acts transitively on the fibers of Θ.

PROOF. It is easiest to reduce to the holomorphic case. For this we choose a conformal structure on S^2, represented by an atlas of holomorphically compatible charts. With this conformal structure, S^2 is conformally equivalent to $\widehat{\mathbb{C}}$ and we can make the additional assumptions stated before (v) without loss of generality.

If $g \in G$, then g is an orientation-preserving homeomorphism on X. Since $\Theta = \Theta \circ g$, it follows from the last part of Lemma A.15 that g is holomorphic and hence a biholomorphism on $X = \mathbb{C}$ or $X = \mathbb{D}$.

(i) This immediately follows from the definition of a deck transformation and Corollary A.27.

In particular, the orbits of points under G are precisely the fibers of the map Θ. Since Θ is a branched covering map, these fibers, and hence the orbits of points under G, are discrete sets in X.

(v) We know that if $g \in G$, then g acts as a biholomorphism on X. So if $X = \mathbb{D}$, then g is a Möbius transformation with $g(\mathbb{D}) = \mathbb{D}$. Hence g preserves the hyperbolic metric d_0 on $\mathbb{D}$ and acts as an isometry on X.

If $X = \mathbb{C}$, then we can say that $g(z) = az + b$ for $z \in \mathbb{C}$, where $a, b \in \mathbb{C}$, $a \neq 0$. Here $|a| = 1$, and so g acts as an isometry on $\mathbb{C}$ equipped with the Euclidean metric; indeed, otherwise g or g^{-1} has an attracting fixed point and this would produce orbits under G that are not discrete in X. This contradicts what we have seen in the proof of (i).

(ii) We may assume that $x = 0$. Then an element $g \in G$ in the stabilizer G_x is necessarily a rotation of $X = \mathbb{C}$ or $X = \mathbb{D}$ around 0. Since G has discrete orbits, this implies that G_x is a finite cyclic group of rotations.

Two points u and v near x lie in the same orbit of G if and only if there exists $g \in G_x$ such that $v = g(u)$. Indeed, since orbits of G are discrete in X, an element $g \in G \setminus G_x$ moves x and hence also the nearby point u a definite distance away from x. Hence g cannot map u to the point v near x.

If $d = \deg(\Theta, x) = \alpha(\Theta(x))$, then Θ is d-to-1 near x; so for each point $x' \neq x$ near x there are precisely d points near x that are mapped to the same image $\Theta(x')$, or, equivalently, lie in the G-orbit of x'. By what we have seen, these d elements must be the points in the G_x-orbit of x'. Since $x' \neq x$, there are precisely $\#G_x$ such points, and we conclude $\#G_x = d$ as desired.

(iii) We argue by contradiction and assume that there exist a compact set $K \subset X$ and pairwise distinct elements $g_n \in G$ for $n \in \mathbb{N}$ with $g_n(K) \cap K \neq \emptyset$. Then we can find $x_n \in K$ such that $y_n = g_n(x_n) \in K$. By passing to subsequences if necessary, we may assume $x_n \to x$ and $y_n \to y$ as $n \to \infty$, where $x, y \in K$. Then $g_n(x) = y$ for all n large. Indeed, if $g_n(x) \neq y$, then by the discreteness of orbits g_n sends x and hence nearby points a definite distance away from y. This is impossible for large n, because g_n sends the point x_n near x to the point y_n near y.

By (ii) the stabilizer group G_x of x is finite. On the other hand, by what we have just seen, if n is large enough, G_x contains the infinitely many distinct elements $g_k^{-1} \circ g_n$, $k \geq n$. This is a contradiction.

(iv) If $\mathcal{O}$ has no punctures, then $S_0^2 = S^2 = \widehat{\mathbb{C}}$ and so $\Theta \colon X \to \widehat{\mathbb{C}}$ is a branched covering map with target $\widehat{\mathbb{C}}$. In particular, each point $p \in \widehat{\mathbb{C}}$ has an open neighborhood $V \subset \widehat{\mathbb{C}}$ that is evenly covered by Θ. By shrinking V if necessary (see Lemma A.10), we may assume that the connected components of $\Theta^{-1}(V)$ are compactly contained in X. In particular, we can find a set $U \subset X$ with $\Theta(U) = V$ such that $\overline{U}$ is compact. Finitely many of the sets V will cover $\widehat{\mathbb{C}}$. If we let K be the closure of the finitely many corresponding sets U, then $K \subset X$ is compact and $\Theta(K) = \widehat{\mathbb{C}}$. This implies that each fiber of Θ and hence also each orbit of G contains a point in K. So the sets $g(K)$, $g \in G$, cover X. The cocompactness of G follows. □

We now consider the orbifold $\mathcal{O}_f = (S^2, \alpha_f)$ of a Thurston map $f \colon S^2 \to S^2$. As the following statement shows, we can lift branches of f^{-1} to the universal orbifold cover of $\mathcal{O}_f$.

LEMMA A.30 (Existence of lifts to the universal orbifold cover). *Let $f \colon S^2 \to S^2$ be a Thurston map, $\mathcal{O}_f = (S^2, \alpha_f)$ be the associated orbifold, $\Theta \colon X \to S_0^2$ be the universal orbifold covering map, and $z_0, w_0 \in X$ be points with $(f \circ \Theta)(w_0) = \Theta(z_0)$.*
Then there exists a branched covering map $A \colon X \to X$ with $A(z_0) = w_0$ and
$$f \circ \Theta \circ A = \Theta.$$
If $z_0 \notin \operatorname{crit}(\Theta)$, then the map A is unique.
If we assume that $S^2 = \widehat{\mathbb{C}}$, X is a Riemann surface, and the maps f and Θ are holomorphic, then A is holomorphic as well.

So in this setting, we have the following commutative diagram:

(A.35)
$$
\begin{array}{ccc}
w_0 \in X & \xleftarrow{\ A\ } & z_0 \in X \\
{\scriptstyle \Theta}\downarrow & & \downarrow{\scriptstyle \Theta} \\
S^2 & \xrightarrow{\ f\ } & S^2.
\end{array}
$$

One should think of A as a lift of a suitable inverse branch of f^{-1} under the branched covering map Θ.

PROOF. We want to apply Theorem A.26 for the map $f \circ \Theta \colon X \to S^2$, which is a branched covering map by Lemma A.15 (i). To see that the hypotheses Theorem A.26 are satisfied, let $z, w \in X$ with $\Theta(z) = (f \circ \Theta)(w)$ be arbitrary. Define $p = \Theta(z)$ and $q = \Theta(w)$. Then

$p = f(q)$, and so $\deg(f,q)\alpha_f(q)$ divides $\alpha_f(p)$ (see Lemma 2.8 (i)). Since $\deg(\Theta,w) = \alpha_f(q)$ and $\deg(\Theta,z) = \alpha_f(p)$ by definition of the universal orbifold covering map Θ, this implies that

$$\deg(f,q)\alpha_f(q) = \deg(f,q)\deg(\Theta,w) = \deg(f \circ \Theta, w)$$

divides $\deg(\Theta,z) = \alpha_f(p)$. So by Theorem A.26 there exists a branched covering map $A\colon X \to X$ with $A(z_0) = w_0$ and $f \circ \Theta \circ A = \Theta$.

If $z_0 \notin \mathrm{crit}(\Theta)$, then $\deg(\Theta,z_0) = \alpha(\Theta(z_0)) = 1$. The uniqueness of A then follows from the uniqueness statement in Theorem A.26.

Finally, if f and Θ are holomorphic, then $f \circ \Theta$ is a holomorphic branched covering map. Since $(f \circ \Theta) \circ A = f$, the holomorphicity of A follows from Lemma A.15. $\square$

A.10. The canonical orbifold metric

In this section we discuss how to obtain an associated geometric structure from orbifold data given by a ramification function on a Riemann surface S. This geometric structure is represented by a metric on S, the *canonical orbifold metric*, and a measure, the *canonical orbifold measure*. We will restrict ourselves to the only case that is important for us, namely when S is the Riemann sphere $\widehat{\mathbb{C}}$. The underlying conformal structure on $\widehat{\mathbb{C}}$ is important here, because in this case the universal orbifold cover carries a natural metric and a measure that we will push-forward to the orbifold. We first discuss some related geometric facts that will be relevant for understanding the local geometry of an orbifold.

A *sector* is a set of the form

$$\Sigma = \{re^{i\theta} : 0 \le r < r_0,\, 0 \le \theta \le \theta_0\},$$

where $r_0 > 0$ and $\theta_0 \in (0, 2\pi]$. If we identify the points r and $re^{i\theta_0}$ on the boundary of Σ for $0 \le r < r_0$, we obtain a *cone* C of *cone angle* θ_0. The point 0 (or rather its image in the quotient C) is called the *cone point* or *conical singularity* of the cone. For the moment we only allow cones angles $0 < \theta_0 \le 2\pi$ (see below for the general case).

A cone C carries a natural length metric induced by the Euclidean metric on Σ. If $r_0 \le 1$, then $\Sigma \subset \mathbb{D}$ and one can equip Σ also with the hyperbolic metric on $\mathbb{D}$, and push it to a length metric on C. Depending on this choice of the metric, one calls C a *Euclidean* or *hyperbolic cone*. If we denote this metric on C by ω in both cases and by d_0 the Euclidean or hyperbolic metric on Σ, then the quotient map $(\Sigma, d_0) \to (C, \omega)$ is a *path isometry* in that it preserves lengths of paths. This property uniquely determines ω.

If we set $\alpha = 2\pi/\theta_0 \geq 1$, then the map $w \mapsto w^\alpha$ (with the branch chosen to be positive on the positive real axis) induces a well-defined homeomorphism of C onto the Euclidean disk $D = B_{\mathbb{C}}(0, r_0^\alpha)$. This defines a chart giving C a Riemann surface structure.

By using this chart one can push the metric on C to D. Then an easy computation shows that C is isometric to the disk D equipped with a (singular) conformal metric, where the length element on D is

$$(A.36) \qquad \frac{|dz|}{\alpha|z|^{1-1/\alpha}}$$

or

$$(A.37) \qquad \frac{2\,|dz|}{\alpha|z|^{1-1/\alpha}(1 - |z|^{2/\alpha})}$$

in the Euclidean or in the hyperbolic case, respectively. So the cone C can be viewed in two different ways: as a Riemann surface biholomorphic to a disk or as a smooth manifold equipped with a Riemannian metric as in (A.36) or (A.37) with a singularity at the cone point.

Note that the conformal metrics in (A.36) and (A.37) are defined for each parameter $\alpha > 0$. One can use this to define Euclidean or hyperbolic cones C for arbitrary cone angles $\theta_0 > 0$. Namely, such a cone C is given by a disk D as above equipped with the conformal metric (A.36) for Euclidean cones, and the conformal metric (A.37) for hyperbolic cones, where $\alpha = 2\pi/\theta_0$.

In many cases that are relevant for us, the cone angle has the form $\theta_0 = 2\pi/n$ for some $n \in \mathbb{N}$. Then one can obtain C also as a quotient under a group action. For this we consider the cyclic group G of order n generated by the rotation $z \in \mathbb{C} \mapsto e^{2\pi i/n}z$ and acting on $X = \{z \in \mathbb{C} : |z| < r_0\}$. Then we can identify C and X/G. Moreover, if we again denote by ω the metric on C and by d_0 the Euclidean metric or the hyperbolic metric on X (assuming $r_0 \leq 1$ in the latter case), then under the identification $C \cong X/G$ we have

$$\omega([z], [w]) = \inf\{d_0(u, v) : u \in Gz, v \in Gw\}$$

for $z, w \in X$, where we denote the image of $u \in X$ in $C \cong X/G$ by $[u]$.

Suppose S is a (connected and oriented) surface equipped with some path metric d. We call S *locally Euclidean* if each point $p \in S$ has a neighborhood U isometric to a Euclidean cone C so that p corresponds to the cone point of C under the isometry. Then the cone angle $\theta_0 > 0$ of C is uniquely determined by p. The points p with $\theta_0 \neq 2\pi$ form a closed and discrete subset of S and are called the *conical singularities* or *cone points* of the locally Euclidean surface.

A *(simple) Euclidean polygon* is a closed Jordan region X in $\mathbb{C}$ whose boundary consists of finitely many Euclidean line segments. These line segments are called the *edges* or *sides*, and their endpoints the *vertices* or *corners* of X.

A *(Euclidean) polyhedral surface* S is a surface obtained by gluing Euclidean polygons together along boundary edges by using isometries. More precisely, we require that S is a locally Euclidean surface that carries a cell decomposition $\mathcal{D}$ such that each tile X in $\mathcal{D}$ is isometric to a Euclidean polygon X' (with respect to the induced path metrics on X and X'). Note that each cone point of S is necessarily a vertex of $\mathcal{D}$.

A *pillow* P is a special case of a polyhedral surface. It is obtained by gluing two identical copies $X_{\mathtt{w}}$ and $X_{\mathtt{b}}$ of a Euclidean polygon X together so that corresponding points on $\partial X_{\mathtt{w}}$ and $\partial X_{\mathtt{b}}$ are identified. Then P is a topological 2-sphere. The Jordan curve $\partial X_{\mathtt{w}} = \partial X_{\mathtt{b}} \subset P$ is called the *equator* of the pillow.

Often it is useful to consider $X_{\mathtt{w}} \subset P$ as the top face of P colored "white" and $X_{\mathtt{b}} \subset P$ as the bottom face colored "black". The pillow P carries a natural cell decomposition with $X_{\mathtt{w}}$ and $X_{\mathtt{b}}$ as tiles, and edges and vertices on the equator $\partial X_{\mathtt{w}} = \partial X_{\mathtt{b}}$ of P that correspond to the edges and vertices of X.

The standard orientation on $\mathbb{C}$ induces a natural orientation on the polygon $X_{\mathtt{w}}$ under a fixed identification $X_{\mathtt{w}} \cong X$ (represented by a positively-oriented flag, for example; see Section A.4). We orient P so that the induced orientation agrees with this given orientation on $X_{\mathtt{w}} \subset P$. Moreover, we equip P with the unique path metric whose restriction agrees with the Euclidean path metrics on $X_{\mathtt{w}}$ and $X_{\mathtt{b}}$. Then P is a polyhedral surface. Each of its cone points correspond to a vertex of X.

It is a standard fact that every polyhedral surface S (such as a pillow) admits a natural Riemann surface structure (see, for example, [**Be84**, Section 3.3]). One can see this along the following lines. Each point $p \in S$ has a neighborhood U such that there exists an isometry $\varphi_p \colon U \to D$ with $\varphi_p(p) = 0$, where $D \subset \mathbb{C}$ is a Euclidean disk centered a 0 and equipped with a conformal metric as in (A.36). By postcomposing φ_p with complex conjugation if necessary, we may in addition assume that φ_p is orientation-preserving. One can then show that these charts φ_p, $p \in S$, are holomorphically compatible. So they form a complex atlas $\mathcal{A}$ on S. The analytic equivalence class of $\mathcal{A}$ is independent of the choices of the isometries φ_p. Hence S carries a natural Riemann surface structure.

If a polyhedral surface S is a topological 2-sphere, then by the uniformization theorem S (as a Riemann surface) is conformally equivalent to $\widehat{\mathbb{C}}$.

Now let $\mathcal{O} = (\widehat{\mathbb{C}}, \alpha)$ be an orbifold with the Riemann sphere as the underlying surface and a ramification function $\alpha \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{N}}$. We assume that $\mathcal{O}$ is parabolic or hyperbolic. Let $\Theta \colon X \to \widehat{\mathbb{C}}$ be the universal orbifold covering map (see Theorem A.24), where we assume that Θ is holomorphic with $X = \mathbb{C}$ in the parabolic and $X = \mathbb{D}$ in the hyperbolic case.

In the first case, $X = \mathbb{C}$ is equipped with the Euclidean metric, and in the second case $X = \mathbb{D}$ is equipped with the hyperbolic metric as in (A.6). To keep the notation simple, we denote these metrics both by d_0. We will remove the punctures of $\mathcal{O}$ from $\widehat{\mathbb{C}}$ and consider

$$\widehat{\mathbb{C}}_0 := \widehat{\mathbb{C}} \setminus \{p \in \widehat{\mathbb{C}} : \alpha(p) = \infty\}.$$

The universal orbifold covering map is then a holomorphic branched covering map $\Theta \colon X \to \widehat{\mathbb{C}}_0$. The associated group of deck transformations $G = \pi_1(\mathcal{O})$ acts properly discontinuously on X and each element $g \in G$ is an orientation-preserving isometry on X (see Proposition A.29).

The *canonical orbifold metric* ω on $\widehat{\mathbb{C}}_0$ is now defined as

$$(A.38) \qquad \omega(p, q) = \inf\{d_0(z, w) : z \in \Theta^{-1}(p), \, w \in \Theta^{-1}(q)\}$$

for $p, q \in \widehat{\mathbb{C}}_0$. Note that here the fibers of Θ are precisely the orbits of G. So these are discrete sets in X. Since G acts by isometries on (X, d_0) and transitively on $\Theta^{-1}(p)$, for each $z_0 \in \Theta^{-1}(p)$ there exists $w_0 \in \Theta^{-1}(q)$ such that $\omega(p, q) = d_0(z_0, w_0)$.

This implies that the infimum in (A.38) is attained as a minimum and that ω satisfies the triangle inequality. It easily follows that ω is actually a metric on $\widehat{\mathbb{C}}_0$.

The uniqueness of the universal orbifold covering map Θ (Corollary A.27 and Remark A.28) implies that the metric ω is uniquely determined by $\mathcal{O}$ if $\mathcal{O}$ is hyperbolic and unique up to a scaling factor if $\mathcal{O}$ is parabolic. Therefore, in the following we will refer to ω as the *canonical orbifold metric* of $\mathcal{O}$.

An equivalent way to view this metric is as follows. The quotient space X/G is homeomorphic to $\widehat{\mathbb{C}}_0 = \Theta(X)$ with a homeomorphism given by

$$\varphi \colon X/G \to \widehat{\mathbb{C}}_0, \quad \varphi \colon [z] = Gz \mapsto \Theta(z)$$

(see Corollary A.21). If we define

$$\widetilde{\omega}([z], [w]) = \inf\{d_0(g(z), h(w)) : g, h \in G\}$$

for $[z], [w] \in X/G$, then $\widetilde{\omega}$ is a metric on the quotient X/G and the map $\varphi \colon (X/G, \widetilde{\omega}) \to (\widehat{\mathbb{C}}_0, \omega)$ is an isometry.

Let $z \in X$ and $p = \Theta(z) \in \widehat{\mathbb{C}}_0$. We consider a small ball $B_z \subset X$ (with respect to d_0) centered at z. We know (see the proof of Proposition A.29 (ii)) that if $w \neq z$ is close to z, then the only points in the orbit Gw close to z are the points in $G_z w$, where $G_z \subset G$ is the stabilizer subgroup of z. So if B_z is small enough, then $B_z/G = B_z/G_z$.

Now G_z is a cyclic group of "rotations" fixing z. It follows that $B_z/G = B_z/G_z$ equipped with the metric $\widetilde{\omega}$ is a cone with cone angle $2\pi/\#G_z = 2\pi/\alpha(p)$. This cone is Euclidean or hyperbolic depending on whether the orbifold is parabolic or hyperbolic.

If $U_p = \varphi(B_z/G)$, then (U_p, ω) is a neighborhood of p that is isometric to such a cone. The points $p \in \widehat{\mathbb{C}}$ where $2 \leq \alpha(p) < \infty$ are precisely the cone points of the orbifold $\mathcal{O} = (\widehat{\mathbb{C}}, \alpha)$. Near all other points, $(\widehat{\mathbb{C}}_0, \omega)$ is locally isometric to the model space (X, d_0).

We have $\omega(\Theta(z), \Theta(w)) \leq d_0(z, w)$ for all $z, w \in X$. A stronger condition is true locally, namely Θ is a *local radial isometry* in the following sense: for each $z \in X$ there exists a neighborhood B_z of z in X such that

$$(A.39) \qquad\qquad \omega(\Theta(z), \Theta(w)) = d_0(z, w)$$

for all $w \in B_z$. This easily follows from the definition of ω and the fact that for $w \neq z$ near z the only points of $Gw = \Theta^{-1}(\Theta(w))$ near z are the points in $G_z w$ all of which have the same distance $d_0(z, w)$ to z.

The relation (A.39) together with a simple covering argument implies that the map Θ is a *path isometry*: if γ is a path in X, then

$$(A.40) \qquad\qquad \mathrm{length}_\omega(\Theta \circ \gamma) = \mathrm{length}_{d_0}(\gamma).$$

The metric space $(\widehat{\mathbb{C}}_0, \omega)$ is geodesic. Indeed, a geodesic segment joining two points $p, q \in \widehat{\mathbb{C}}_0$ can be obtained as follows. We can pick points $z \in \Theta^{-1}(p)$ and $w \in \Theta^{-1}(q)$ such that $\omega(\Theta(z), \Theta(w)) = d_0(z, w)$. Since (X, d_0) is geodesic, we can find a geodesic segment γ joining z and w. Since Θ is a path isometry, the path $\Theta \circ \gamma$ must then be a geodesic segment in $(\widehat{\mathbb{C}}_0, \omega)$ joining p and q.

In order to compare the canonical orbifold metric ω with the chordal metric σ, fix $p \in \widehat{\mathbb{C}}_0$ and $z \in \Theta^{-1}(p)$. Then $\deg(\Theta, z) = \alpha(p)$. This implies that if $w \in X$ is near z and $q = \Theta(w)$, then

$$\sigma(p, q) = \sigma(\Theta(z), \Theta(w)) \asymp d_0(z, w)^{\alpha(p)} = \omega(p, q)^{\alpha(p)}.$$

It follows that there is a neighborhood $U_p \subset \widehat{\mathbb{C}}_0$ of p such that

$$(A.41) \qquad \omega(p, q) \asymp \sigma(p, q)^{1/\alpha(p)},$$

for $q \in U_p$, where $C(\asymp) = C(p)$.

If $p \in \widehat{\mathbb{C}}$ is a puncture of $\mathcal{O}$ (i.e., if $\alpha(p) = \infty$), let $\gamma \colon [0, 1) \to \widehat{\mathbb{C}}_0$ be a path that ends at p in the sense that $\lim_{t \to 1} \gamma(t) = p$. If $\widetilde{\gamma}$ is a lift of γ under Θ (see Lemma A.17), then it must leave any compact subset of X. Thus $\widetilde{\gamma}$ has infinite d_0-length which by (A.40) (applied to $\widetilde{\gamma}$) implies that $\gamma = \Theta \circ \widetilde{\gamma}$ has infinite length with respect to ω. In other words, the punctures are infinitely far away from the points in $\widehat{\mathbb{C}}_0$ with respect to the metric ω.

If $\mathcal{O}$ has no punctures, then (A.41) and a covering argument implies that there exists a constant $C \geq 1$ such that

$$(A.42) \qquad \frac{1}{C} \sigma(p, q) \leq \omega(p, q) \leq C \sigma(p, q)^\epsilon$$

for $p, q \in \widehat{\mathbb{C}}$, where $\epsilon = \min\{1/\alpha(u) : u \in \widehat{\mathbb{C}}\}$. In this case, ω induces the standard topology on $\widehat{\mathbb{C}}$. The lower bound for ω in (A.42) is also true (on $\widehat{\mathbb{C}}_0$) if $\mathcal{O}$ has punctures (this can be shown by using the estimates from Lemma A.31 below). In particular, the map $\mathrm{id}_{\widehat{\mathbb{C}}_0} \colon (\widehat{\mathbb{C}}_0, \omega) \to (\widehat{\mathbb{C}}_0, \sigma)$ is always Lipschitz (but never bi-Lipschitz).

One can also describe the metric ω as a singular conformal metric with a conformal factor that is smooth everywhere except at the points in $\mathrm{supp}(\alpha)$. To see this, let $z \in X$. We define $\|\Theta'(z)\|$ as the norm of the derivative $\Theta'(z)$ with respect to the underlying metric d_0 on X and the chordal metric σ (or rather the spherical metric) on $\widehat{\mathbb{C}}$. More explicitly, if $X = \mathbb{C}$, then

$$\|\Theta'(z)\| = \frac{2|\Theta'(z)|}{1 + |\Theta(z)|^2},$$

and if $X = \mathbb{D}$, then

$$\|\Theta'(z)\| = \frac{(1 - |z|^2)|\Theta'(z)|}{1 + |\Theta(z)|^2}.$$

These expressions have to be understood as suitable limits if $\Theta(z) = \infty$. The function $z \mapsto \|\Theta'(z)\|$ is smooth and positive everywhere on $X \setminus \mathrm{crit}(\Theta)$.

Since $G = \pi_1(\mathcal{O})$ acts by isometries on X, we have

$$\|\Theta'(z)\| = \|\Theta'(g(z))\|$$

for $g \in G$ as follows from the chain rule. So if we set

(A.43) $$\rho(p) = \frac{1}{\|\Theta'(z)\|}$$

for $p \in \widehat{\mathbb{C}} \setminus \operatorname{supp}(\alpha)$ and $z \in \Theta^{-1}(p) = Gz$, then ρ is well-defined. Note that $\operatorname{supp}(\alpha)$ includes the punctures and the critical values of Θ. So on $\widehat{\mathbb{C}} \setminus \operatorname{supp}(\alpha)$ the function ρ is smooth and positive.

Now suppose β is a path in X with $\operatorname{length}_{d_0}(\beta) < \infty$ and define $\gamma = \Theta \circ \beta$. Then

$$\operatorname{length}_\sigma(\gamma) \lesssim \operatorname{length}_\omega(\gamma) = \operatorname{length}_{d_0}(\beta) < \infty.$$

If we denote by ds integration with respect to d_0-arclength for β and by $d\sigma$ integration with respect to σ-arclength for γ, then

$$\int_\gamma \rho \, d\sigma = \int_\beta (\rho \circ \Theta) \|\Theta'\| \, ds = \int_\beta ds$$
$$= \operatorname{length}_{d_0}(\beta) = \operatorname{length}_\omega(\gamma).$$

Since ω is a geodesic metric, this and a path lifting argument imply that for all $p, q \in \widehat{\mathbb{C}}_0$ we have

$$\omega(p, q) = \inf_\gamma \int_\gamma \rho \, d\sigma,$$

where the infimum is taken over all σ-rectifiable paths in $\widehat{\mathbb{C}}_0$ joining p and q. In other words, ω is the (singular) conformal metric on $\widehat{\mathbb{C}}$ with length element $\rho \, d\sigma$.

The local behavior of ρ near its singularities is described in the following statement.

LEMMA A.31. *Let ρ be the conformal density of the canonical orbifold metric ω of a parabolic or hyperbolic orbifold $\mathcal{O} = (\widehat{\mathbb{C}}, \alpha)$ as defined in (A.43). Let $p \in \widehat{\mathbb{C}}$. If $\alpha(p) < \infty$, then*

(A.44) $$\rho(q) \asymp \frac{1}{\sigma(p, q)^{1 - 1/\alpha(p)}}$$

for q near p.

If $\alpha(p) = \infty$, and $\mathcal{O}$ is parabolic, then

(A.45) $$\rho(q) \asymp \frac{1}{\sigma(p, q)},$$

and if $\mathcal{O}$ is hyperbolic, then

(A.46) $$\rho(q) \asymp \frac{1}{\sigma(p, q) \log(1/\sigma(p, q))},$$

for q near p.

In all these inequalities $C(\asymp) = C(p)$.

If $\alpha(p) = 1$, then (A.44) should be interpreted as $\rho(q) \asymp 1$ for q near p. This corresponds to the fact that ρ is smooth and positive on $\widehat{\mathbb{C}} \setminus \operatorname{supp}(\alpha)$.

PROOF. As before, we denote by $\Theta \colon X \to \widehat{\mathbb{C}}_0$ the holomorphic universal orbifold covering map of $\mathcal{O}$, where $X = \mathbb{C}$ if $\mathcal{O}$ is parabolic and $X = \mathbb{D}$ if $\mathcal{O}$ is hyperbolic.

Case 1: $\alpha(p) < \infty$. Then there exists $z \in X$ with $\Theta(z) = p$. If $w \neq z$ is near z and $q = \Theta(w)$, then we have

$$\sigma(p, q) \asymp d_0(z, w)^{\alpha(p)} \text{ and } \|\Theta'(w)\| \asymp d_0(z, w)^{\alpha(p)-1}.$$

Since $\rho(q) = 1/\|\Theta'(w)\|$, inequality (A.44) follows.

Case 2: $\alpha(p) = \infty$ and $\mathcal{O}$ is parabolic. This means that the signature of $\mathcal{O}$ is (∞, ∞) or $(2, 2, \infty)$ (see the list of parabolic orbifold signatures in Proposition 2.13 (ii)).

Without loss of generality we may assume that $p = 0$. Then near p chordal and Euclidean metrics are comparable, and so we will state our estimates in terms of the Euclidean metric.

If the signature is (∞, ∞), then we may assume that one of the corresponding punctures of $\widehat{\mathbb{C}}$ is at $p = 0$, the other at ∞, and $\Theta(z) = \exp(2\pi i z)$. Then Θ maps the upper half-plane $\mathbb{H}$ to the punctured neighborhood $U_p = \mathbb{D} \setminus \{0\}$ of $p = 0$. If $w \in \mathbb{H}$ and $q = \Theta(w)$, then

(A.47) $$|\Theta'(w)| \asymp |\exp(2\pi i w)| \asymp |\Theta(w)| = |q|.$$

Inequality (A.45) immediately follows.

If the signature of $\mathcal{O}$ is $(2, 2, \infty)$, then we may assume that the puncture is at $p = 0$, the two cone points of $\mathcal{O}$ are at -1 and 1, and that

$$\Theta(z) = 1/\cos(2\pi z) = \frac{2 \exp(2\pi i z)}{1 + \exp(4\pi i z)}.$$

Then Θ maps the half-plane $H = \{z \in \mathbb{C} : \operatorname{Im}(z) > C_0\}$ with $C_0 \gg 1$ to a small punctured neighborhood U_p of $p = 0$. For $w \in H$ and $q = \Theta(w)$ we again have inequalities as in (A.51) and (A.45) follows.

Case 3: $\alpha(p) = \infty$ and $\mathcal{O}$ is hyperbolic. As before we assume that $p = 0$ and work with the Euclidean rather than the chordal metric on $\mathbb{C} \subset \widehat{\mathbb{C}}$.

This case is much harder to analyse. As in Case 2 p has no preimage by Θ. The main difficulty however is that we do not have an explicit expression for Θ as in the parabolic case above. However, up to post-composition by a conformal map, Θ will be given as a specific function

Θ_0 (which will in fact be the exponential function). This will allow us to obtain our desired estimates.

Special Case: $\Theta_0 \colon \mathbb{H} \to \mathbb{D} \setminus \{0\}$, where $\Theta_0(z) = \exp(2\pi \boldsymbol{i} z)$.

Let $q_0 := \Theta_0(z)$ for $z = x + y\boldsymbol{i} \in \mathbb{H}$. Then

$$|\Theta_0'(z)| = 2\pi |q_0| \quad \text{and}$$

$$|q_0| = \exp(-2\pi y) \quad \text{or} \quad \operatorname{Im}(z) = \frac{1}{2\pi} \log(1/|q_0|).$$

If we equip $\mathbb{H}$ with the hyperbolic metric and $\mathbb{D} \setminus \{0\}$ with the Euclidean metric we obtain

$$\|\Theta_0'(z)\| = |\Theta_0'(z)| \operatorname{Im} z = |q_0| \log(1/|q_0|).$$

This means $1/\|\Theta_0'\|$ satisfies an estimate as in (A.46).

Note that Θ_0 is the (unbranched) universal covering map of $\mathbb{D} \setminus \{0\}$. The group of deck transformations G_0 of Θ_0 is given by all translations of the form $u \mapsto u + n$ for $n \in \mathbb{Z}$, meaning it is generated by $z \mapsto z + 1$. The space $\mathbb{H}/G_0$ carries naturally a conformal structure. Then $\mathbb{H}/G_0$ is biholomorphic to $\mathbb{D} \setminus \{0\}$. The biholomorphism $\psi_0 \colon \mathbb{H}/G_0 \to \mathbb{D} \setminus \{0\}$ is given by

$$(A.48) \qquad\qquad \psi_0 \colon [u] \in \mathbb{H}/G_0 \mapsto \Theta_0(u) \in \mathbb{D} \setminus \{0\},$$

(see Lemma A.19 and Corollary A.21).

Assume now that that $f \colon \mathbb{D} \to U$ is a holomorphic map with $f(0) = p$. We assume for simplicity that U is contained in a compact set in $\mathbb{C}$, so that the Euclidean metric is comparable to the chordal metric with some absolute mutiplicative constants. We may write $f(w) = p + a_n w^n + \dots$ where $a_n \in \mathbb{C} \setminus \{0\}$ and $n \in \mathbb{N}$ for w close to 0. Let $q_0 \in \mathbb{D} \setminus \{0\}$ and $f(q_0) = q$. Then

$$|q_0|^n \asymp |p - q| \quad \text{and} \quad |f'(q_0)| \asymp |q_0|^{n-1}.$$

Model Case: Let us consider the map $\Theta_1 \colon \mathbb{H} \to U$ that is given by $\Theta_1 := f \circ \Theta_0$. Here f is as above.

Let $z \in \mathbb{H}$ and $\Theta_0(z) = q_0 \in \mathbb{D} \setminus \{0\}$. Then

$$|\Theta_1'(z)| = |f'(q_0)||\Theta_0'(z)| \asymp |q_0|^{n-1}|q_0| \asymp |p - q|.$$

Furthermore, there is a constant $C > 0$ with

$$\frac{1}{n} \log|p - q| - C \leq \log|q_0| \leq \frac{1}{n} \log|p - q| + C.$$

It follows that

$$\log(1/|p - q|) \asymp \log(1/|q_0|) \asymp \operatorname{Im}(z)$$

for q close to p. Thus $\|\Theta_1'(z)\| = |\Theta_1'| \operatorname{Im}(z) \asymp |p - q| \log(1/|p - q|)$, meaning that $1/\|\Theta_1'\|$ satisfies the desired estimate (A.46).

After considering these model cases, let us now turn the attention back to the case we want to study. Let $p = 0$ be our point with $\alpha(p) = \infty$. Let $U = B_{\mathbb{C}}(0, \delta)$ and $U_p = U \setminus \{0\}$ be a small Euclidean disk, respective punctured disk, around 0. We may assume that $\delta > 0$ is so small that U_p does not contain any point in $\operatorname{supp}(\alpha)$. Let $f \subset \mathbb{D} \to U$ be given by $f(z) = \delta z$. The map $\Theta_1 \colon \mathbb{H} \to U_p$ is defined as above by $\Theta_1 = f \circ \Theta_0$. Note that Θ_1 is the (unbranched) universal covering map of U_p. Its group of deck transformations is again G_0.

Consider now $\Theta \colon X \to \widehat{\mathbb{C}}_0$, our universal orbifold covering map. Fix a component $V \subset X$ of $\Theta^{-1}(U_p)$. Recall that $U_p \cap \operatorname{supp}(\alpha) = \emptyset$. Therefore, $\Theta|V$ is also an (unbranched) covering map over U_p. This implies that there exists a holomorphic map $\varphi \colon \mathbb{H} \to V$ such that

$$(A.49) \qquad\qquad \Theta_1 = \Theta \circ \varphi$$

on $\mathbb{H}$ (see Section A.5).

Claim 1. The map $\varphi \colon \mathbb{H} \to V$ is bijective, thus a biholomorphism.

As a covering map, φ is surjective. Assume that φ is not injective. Let $u, v \in \mathbb{H}$ be distinct points. Then $\varphi(u) = \varphi(v)$ if and only if there is a deck transformation g of φ such that $v = g(u)$. In this case is follows from (A.49) that $\Theta_1(u) = \Theta_1(v)$, meaning that $v = u + n$ for a $n \in \mathbb{Z} \setminus \{0\}$. We may assume that $n \geq 1$ is the smallest such number. Let G_φ be the cyclic group generated by g. It follows that for any two points $u', v' \in \mathbb{H}$ we have $\varphi(u') = \varphi(v')$ if and only if they are in the same G_φ-orbit. This means that G_φ is induced by φ. Thus $\mathbb{H}/G_\varphi$ is conformally equivalent to $\varphi(\mathbb{H}) = V$.

On the other hand G_φ is induced by the map $z \mapsto \Theta_0(z/n)$. Thus $\mathbb{H}/G_\varphi$, hence V, is conformally equivalent to $\Theta_0(\mathbb{H}) = \mathbb{D} \setminus \{0\} \subset \mathbb{C}$. This means that V has a boundary point p_0 that is isolated, therefore $p_0 \in X$. Consider the sequence $\{k\boldsymbol{i}\}_{k \in \mathbb{N}} \subset \mathbb{H}$. Note that $\Theta_1(k\boldsymbol{i}) \to 0$ and $\varphi(k\boldsymbol{i}) \to p_0$ as $k \to \infty$. It follows that

$$\Theta(p_0) = \lim_{k \to \infty} \Theta \circ \varphi(k\boldsymbol{i}) = \lim_{k \to \infty} \Theta_1(k\boldsymbol{i}) = 0 = p.$$

However, p is not in the image of Θ by definition. This contradiction shows that φ is in fact injective. Claim 1 is proved.

Let $u, v \in \mathbb{H}$ be given, and $z = \varphi(u), w = \varphi(v) \in V$. It follows from (A.49) and Claim 1 that $\Theta_1(u) = \Theta_1(v)$ if and only if $u - v \in \mathbb{Z}$ if and only if $\Theta(z) = \Theta(w)$. This in turn is true if and only if there is a deck transformation $g \in \pi_1(\mathcal{O})$ of Θ such that $g(z) = w$. Recall that $\pi_1(\mathcal{O})$ is the fundamental group of the orbifold, see Proposition A.29. Note

that any such g satisfies $g(V) = V$. Let H be the subgroup of $\pi_1(\mathcal{O})$ of all such g.

Fix a $u_0 \in \mathbb{H}$ and let $z_0 := \varphi(u_0), z_1 := \varphi(u_0 + 1) \in V \subset X$. Let $h \in \pi_1(\mathcal{O})$ be the (unique) deck transformation with $h(z_0) = z_1$. Then both $u \mapsto \varphi(u + 1)$ and $h \circ \varphi$ are lifts of Θ_1 under $\Theta|V$ that send the point u_0 to z_1. By the uniqueness statement for lifts (Lemma A.6 (i)) this implies

$$(\text{A.50}) \qquad\qquad \varphi(u + 1) = h(\varphi(u)) \text{ for } u \in \mathbb{H}.$$

Thus φ conjugates (the deck transformation of Θ_1) $z \mapsto z + 1$ to h. Clearly it follows that $\varphi(u + n) = h^n(\varphi(u))$ for all $n \in \mathbb{Z}$. Thus H is generated by h.

Let us assume that $X = \mathbb{H}$. Then the map h is a biholomorphism on $X = \mathbb{H}$ and hence a Möbius transformation. It can be an elliptic element of finite order, hyperbolic, or parabolic (see [**Be83**, Section 4.3] for this standard terminology).

Claim 2. h is parabolic.

If h is elliptic, then we may assume that $X = \mathbb{D}$ and $h(z) = e^{2\pi i/n}z$ with $n \in \mathbb{N}$. In this case $g_o^n = \text{id}_\mathbb{D}$, meaning that h is of finite order. It follows from (A.52) that $\varphi(u + n) = \varphi(u)$ for $u \in \mathbb{H}$ contradicting Claim 1. Thus h is not elliptic.

If h is hyperbolic, then, up to conformal equivalence we may assume that $X = \{z \in \mathbb{C} : 0 < \text{Im}(z) < M\}$ for some $M > 0$ and $h(z) = z + 1$ (note that $\tau(z) = \exp(\pi z/M)$ maps X to $\mathbb{H}$ and conjugates h to $z \mapsto e^{\pi/M}z$). Then X/h is conformally equivalent to the annulus $\Omega = \{e^{-2\pi M} < |z| < 1\}$. In fact, the map $X/H \to \Omega$ may explicitly be given as $[z] \mapsto \exp(2\pi i z)$.

If h is parabolic, then, up to conformal equivalence, we may assume that X is the upper half-plane $\mathbb{H}$ and that $h(z) = z + 1$. Then $H = G_0$ and X/H is conformally equivalent to the punctured unit disk $\Omega = \mathbb{D} \setminus \{0\}$ (see (A.48)).

We conclude that in the hyperbolic and parabolic case there is a biholomorphic map $\psi_1: X/H \to \Omega$. The bounded region $\Omega \subset \mathbb{C}$ has an isolated boundary point precisely in the parabolic case.

Consider first the map $\widetilde{\varphi}: \mathbb{H}/G_0 \to X/H$ defined by $\widetilde{\varphi}([u]) := [\varphi(u)] \in X/H$ for all $[u] \in \mathbb{H}/G_0$. This map is well defined by (A.52). Furthermore, φ is holomorphic, it follows that $\widetilde{\varphi}$ is holomorphic with respect to the conformal structures on $\mathbb{H}/G_0$ and X/H.

In a similar fashion, the map $\widetilde{\Theta}: X/H \to \widehat{\mathbb{C}}_0$ given by $\widetilde{\Theta}([z]) := \Theta(z)$ for $[z] \in X/H$ is well defined and holomorphic.

Let us use the biholomorphisms $\psi_1 \colon X/H \to \Omega$ and $\psi_0 \colon \mathbb{H}/G_0 \to \mathbb{D} \setminus \{0\}$ (see (A.48)) to change the coordinates of domain and range of $\widetilde{\varphi}$. More precisely, we consider the holomorphic map

$$F := \psi_1 \circ \widetilde{\varphi} \circ \psi_0^{-1} \colon \mathbb{D} \setminus \{0\} \to \Omega.$$

Since $\Omega \subset \mathbb{C}$ is bounded, it follows that 0 is a removable singularity of F. There are two possibilities, namely $F(0) \in \Omega$ or $F(0) \notin \Omega$. We will show that the first possibility does not occur. To this end assume that $F(0) \in \Omega$; let $p_0 := \psi_1^{-1}(F(0)) \in X/H$. Recall that $\psi_0([k\boldsymbol{i}]) \to 0$ as $k \to \infty$. Then it follows that $\widetilde{\varphi}([k\boldsymbol{i}]) \to p_0 \in X/G_0$. But

$$\widetilde{\Theta}(p_0) = \lim_{k \to \infty} \widetilde{\Theta}(\widetilde{\varphi}([k\boldsymbol{i}])) = \lim_{k \to \infty} \Theta \circ \varphi(k\boldsymbol{i}) = \lim_{k \to \infty} \Theta_1(k\boldsymbol{i}) = p \notin \widehat{\mathbb{C}}_0.$$

This is a contradiction, meaning that $F(0) \notin \Omega$ is desired.

Therefore $F(0) \in \partial\Omega$. By the open mapping theorem, $F(0)$ must be an isolated boundary point on $\partial\Omega$. There are no such points on $\partial\Omega$ in the hyperbolic case. This shows that h is indeed parabolic. This finishes the proof of Claim 2.

With the knowledge that h is parabolic, we switch to the following more convenient situation. Namely, we may assume that X is the upper half-plane $\mathbb{H}$ equipped with the hyperbolic metric (given by the length element (A.7)) and that $H = G_0$ is generated by $h(z) = z + 1$. This can always be achieved by precomposing the original map Θ with a suitable Möbius transformation. Note that in the proof of Claim 2 then $\psi_1 = \psi_0$.

The map Θ can be given as the composition

$$\underbrace{X \to X/H \xrightarrow{\psi_0} \mathbb{D} \setminus \{0\}}_{\Theta_0} \underbrace{\xrightarrow{\psi_0^{-1}} X/H \xrightarrow{\widetilde{\Theta}} \widehat{\mathbb{C}}_0}_{f} .$$

Here the composition of the first two maps is exactly the map Θ_0 of the special case. Let $f := \widetilde{\Theta} \circ \psi_0^{-1} \colon \mathbb{D} \setminus \{0\} \to \widehat{\mathbb{C}}_0$, which is holomorphic. Then $\Theta = f \circ \Theta_0$. As in the proof of Claim 2, the map f extends holomorphically to 0 and $f(0) = p$. This means that Θ is exactly of the form considered in the model case. We have already seen that this means that Θ satisfies the desired estimate. $\qquad\square$

**** Here is the old proof for refence/backup ****

PROOF. As before, we denote by $\Theta \colon X \to \widehat{\mathbb{C}}_0$ the holomorphic universal orbifold covering map of $\mathcal{O}$, where $X = \mathbb{C}$ if $\mathcal{O}$ is parabolic and $X = \mathbb{D}$ if $\mathcal{O}$ is hyperbolic.

Case 1: $\alpha(p) < \infty$. Then there exists $z \in X$ with $\Theta(z) = p$. If $w \neq z$ is near z and $q = \Theta(w)$, then we have

$$\sigma(p,q) \asymp d_0(z,w)^{\alpha(p)} \text{ and } \|\Theta'(w)\| \asymp d_0(z,w)^{\alpha(p)-1}.$$

Since $\rho(q) = 1/\|\Theta'(w)\|$, inequality (A.44) follows.

The asymptotics of ρ near a puncture $p \in \widehat{\mathbb{C}}$ is much harder to analyze, because p has no preimage in X. Without loss of generality we may assume that $p = 0$. Then near p chordal and Euclidean metrics are comparable, and so we will state our estimates in terms of the Euclidean metric.

Case 2: $\alpha(p) = \infty$ and $\mathcal{O}$ is parabolic. Then the signature of $\mathcal{O}$ is (∞, ∞) or $(2, 2, \infty)$ (see the list of parabolic orbifold signatures in Proposition 2.13 (ii)).

If the signature is (∞, ∞), then we may assume that one of the corresponding punctures of $\widehat{\mathbb{C}}$ is at $p = 0$, the other at ∞, and $\Theta(z) = \exp(2\pi i z)$. Then Θ maps the upper half-plane $\mathbb{H}$ to the punctured neighborhood $U_p = \mathbb{D} \setminus \{0\}$ of $p = 0$. If $w \in \mathbb{H}$ and $q = \Theta(w)$, then

$$(A.51) \qquad\qquad |\Theta'(w)| \asymp |\exp(2\pi i w)| \asymp |\Theta(w)| \asymp |q|.$$

Since $\rho(q) = 1/\|\Theta'(w)\| \asymp 1/|\Theta'(w)|$, inequality (A.45) immediately follows.

If the signature of $\mathcal{O}$ is $(2, 2, \infty)$, then we may assume that the puncture is at $p = 0$, the two cone points of $\mathcal{O}$ are at -1 and 1, and that

$$\Theta(z) = 1/\cos(2\pi z) = \frac{2\exp(2\pi i z)}{1 + \exp(4\pi i z)}.$$

Then Θ maps the half-plane $H = \{z \in \mathbb{C} : \text{Im}(z) > C_0\}$ with $C_0 \gg 1$ to a small punctured neighborhood U_p of $p = 0$. For $w \in H$ and $q = \Theta(w)$ we again have inequalities as in (A.51) and (A.45) follows.

Case 3: $\alpha(p) = \infty$ and $\mathcal{O}$ is hyperbolic. As before, we may assume that $p = 0$ and can use the Euclidean rather than the chordal metric near p.

Let $U_p = B_{\mathbb{C}}(0, \delta) \setminus \{0\}$ be a small punctured Euclidean disk around 0. We may assume that $\delta > 0$ is so small that U_p does not contain any point in $\text{supp}(\alpha)$. Let

$$H = \{z \in \mathbb{C} : \text{Im}(z) > C_0\}$$

with $C_0 = \frac{1}{2\pi} \log(1/\delta)$. Then the map

$$\Theta_0 \colon H \to U_p, \ u \in H \mapsto \Theta_0(u) := \exp(2\pi i u)$$

is the universal covering map of U_p. Since $U_p \cap \text{supp}(\alpha) = \emptyset$, the map Θ restricted to any component of $\Theta^{-1}(U_p)$ is also a covering map over

U_p. This implies that there exists a holomorphic map $\varphi\colon H \to X$ such that $\Theta_0 = \Theta \circ \varphi$ on H.

In order to establish the desired asymptotics of ρ near p, we need good estimates for $|\Theta'|$. We will obtain these by establishing distortion bounds for φ. Then the estimates for $|\Theta'|$ are reduced to those for $|\Theta_0'|$.

We will first show an equivariance property of φ. Fix $u_0 \in H$. Then $q_0 := \Theta_0(u_0 + 1) = \Theta_0(u_0) \in U_p$. So if we set $z_0 = \varphi(u_0) \in X$ and $z_1 = \varphi(u_0 + 1) \in X$, then $\Theta(z_0) = \Theta(z_1) = q_0$. Hence there exists $g_0 \in G = \pi_1(\mathcal{O})$ such that $g_0(z_0) = z_1$. Then both $u \mapsto \varphi(u + 1)$ and $g_0 \circ \varphi$ are lifts of Θ_0 under Θ that send the point u_0 to z_1. By the uniqueness statement for lifts (Lemma A.6 (i)) this implies

$$(A.52) \qquad\qquad \varphi(u + 1) = g_0(\varphi(u)) \text{ for } u \in H.$$

The map g_0 is a biholomorphism on $X = \mathbb{D}$ and hence a Möbius transformation. It can be an elliptic element of finite order (where we allow $g_0 = \mathrm{id}_{\mathbb{D}}$), hyperbolic, or parabolic (see [**Be83**, Section 4.3] for this standard terminology).

Claim. g_0 is parabolic.

To see this, we consider the cyclic subgroup G_0 of G generated by g_0 and the quotient X/G_0. For all three possible types of g_0, the quotient X/G_0 carries a natural Riemann surface structure and is conformally equivalent to a bounded region $\Omega \subset \mathbb{C}$.

Indeed, if g_0 is elliptic, then up to conformal equivalence we may assume that $g_0(z) = e^{2\pi i/n}z$ with $n \in \mathbb{N}$. So $X/G_0 = \mathbb{D}/G_0$ is a cone and conformally equivalent to $\Omega = \mathbb{D}$.

If g_0 is hyperbolic or parabolic, then, up to conformal equivalence, we may assume that X is the upper half-plane $\mathbb{H}$ and that $g_0(z) = \lambda z$ with $\lambda > 1$ in the hyperbolic and $g_0(z) = z + 1$ in the parabolic case.

So if g_0 is hyperbolic or parabolic, then $X/G_0 = \mathbb{H}/G_0$ is conformally equivalent to the annulus $\Omega = \{z \in \mathbb{C} : 1 < |z| < \exp(2\pi^2/\log \lambda)\}$ or the punctured unit disk $\Omega = \mathbb{D} \setminus \{0\}$, respectively. In all cases, an explicit biholomorphism $\psi\colon X/G_0 \to \Omega$ can easily be obtained from a holomorphic branched covering map $X \to \Omega$ induced by G_0 (see [**Ne53**, pp. 17–19] for a related argument).

We conclude that only in the parabolic case the region $\Omega \cong X/G_0$ has an isolated boundary point.

To derive the claim from this, we define a holomorphic map $f\colon U_p \to \Omega$ as follows. If $q \in U_p$ is arbitrary, we pick $u \in H$ with $\Theta_0(u) = q$ and set $f(q) = \psi([\varphi(u)]) \in \Omega$. Here $[z] \in X/G_0$ denotes the orbit of a point $z \in X$ under G_0. As follows from (A.52), the map f is well-defined and holomorphic. Since $\Theta = \Theta \circ g$ for all $g \in G_0$, we can also define a

unique holomorphic map $\widetilde{\Theta}\colon X/G_0 \to \widehat{\mathbb{C}}_0$ by setting $\widetilde{\Theta}([z]) = \Theta(z)$ for $z \in X$.

Since Ω is a bounded region, the map f has a removable singularity at $p = 0$ and hence a holomorphic extension to the disk $D_p = U_p \cup \{p\}$. Then $f(0) \in \overline{\Omega}$. Here actually $f(0) \in \partial\Omega$. To see this, we pick a sequence $\{u_n\}$ in H with $\operatorname{Im}(u_n) \to +\infty$. Then $\Theta_0(u_n) \to 0$. The sequence $\{[\varphi(u_n)]\}$ has no limit point in X/G_0. Otherwise, by passing to a subsequence if necessary, we may assume that $[\varphi(u_n)] \to [z_0]$, where $z_0 \in X$. Then

$$\Theta(z_0) = \widetilde{\Theta}([z_0]) = \lim_{n\to\infty} \widetilde{\Theta}([\varphi(u_n)])$$
$$= \lim_{n\to\infty} \Theta(\varphi(u_n)) = \lim_{n\to\infty} \Theta_0(u_n) = 0,$$

contradicting the fact that $p = 0$ is a puncture.

Since the sequence $\{[\varphi(u_n)]\}$ has no limit point in X/G_0 and ψ is a biholomorphism of X/G_0 onto Ω, we have $\psi([\varphi(u_n)]) \to \partial\Omega$ as $n \to \infty$. It follows that

$$f(0) = \lim_{n\to\infty} f(\Theta_0(u_n)) = \lim_{n\to\infty} \psi([\varphi(u_n)]) \in \partial\Omega.$$

Since $f(0) \in \partial\Omega$, the open mapping theorem implies that $f(0)$ must be an isolated point on $\partial\Omega$. In the elliptic and hyperbolic case, there are no such points on $\partial\Omega$. This shows that g_0 is indeed parabolic. We also see that for the holomorphic extension of f we have $f(0) = 0$. This finishes the proof of the claim.

With the knowledge that g_0 is parabolic, we switch to the more convenient situation discussed above. Namely, we may assume that X is the upper half-plane $\mathbb{H}$ equipped with the hyperbolic metric (given by the length element (A.7)) and that $g_0(z) = z+1$. This can always be achieved by precomposing the original map Θ with a suitable Möbius transformation.

Then in (A.52) we have

$$\varphi(u + 1) = \varphi(u) + 1$$

for $u \in H$. A biholomorphism $\psi\colon X/G_0 \to \Omega = \mathbb{D} \setminus \{0\}$ is given by $[z] \mapsto \psi([z]) := \exp(2\pi \boldsymbol{i} z)$. So it follows from the proof of the claim that there is a non-constant holomorphic map f on the disk $D_p = U_p \cup \{p\}$ with $f(0) = 0$ such that

(A.53) $f(\exp(2\pi \boldsymbol{i} u)) = f(\Theta_0(u)) = \psi([\varphi(u)]) = \exp(2\pi \boldsymbol{i} \varphi(u))$

for $u \in H$. Near 0 the function f has a Taylor expansion of the form $f(q) = aq^n + \dots$, where $n \in \mathbb{N}$ and $a \in \mathbb{C} \setminus \{0\}$. Hence if $q = \exp(2\pi \boldsymbol{i} u)$

is near 0, then

$$|q|^n \asymp |f(q)| = \exp(-2\pi \operatorname{Im}(\varphi(u)))$$

and so

(A.54) $$\operatorname{Im}(\varphi(u)) \asymp \log(1/|q|).$$

If we differentiate in (A.53) with respect to u, we also see that

$$|q|^n \asymp 2\pi |f'(q)| \cdot |q| = |\varphi'(u)| \cdot |f(q)| \asymp |\varphi'(u)| \cdot |q|^n,$$

and so

(A.55) $$|\varphi'(u)| \asymp 1.$$

The inequalities (A.54) and (A.55) are valid for $u \in H$ with $q = \exp(2\pi \boldsymbol{i} u)$ near 0. We now use them to estimate

$$\|\Theta'(w)\| = \frac{2\operatorname{Im}(w)|\Theta'(w)|}{1 + |\Theta(w)|^2} \asymp \operatorname{Im}(w)|\Theta'(w)|$$

for $w \in \mathbb{H}$ with $q = \Theta(w)$ close to $p = 0$. Indeed, if $w = \varphi(u)$, then

$$\Theta(w) = \Theta(\varphi(u)) = \Theta_0(u) = \exp(2\pi \boldsymbol{i} u) = q;$$

so by differentiating with respect to u and using (A.55) we conclude that

$$|\Theta'(w)| \asymp |\Theta'(\varphi(u))| \cdot |\varphi'(u)| = |\Theta_0'(u)| \asymp |q|.$$

Hence by (A.54),

$$\rho(q) = \frac{1}{\|\Theta'(w)\|} \asymp \frac{1}{\operatorname{Im}(w)|\Theta'(w)|} \asymp \frac{1}{|q|\operatorname{Im}(\varphi(u))} \asymp \frac{1}{|q|\log(1/|q|)}$$

for q near $p = 0$. Inequality (A.46) follows. $\qquad\square$

We know that in the absence of punctures the orbifold metric ω is related to the chordal metric σ by an inequality as in (A.42). The following statement further clarifies the relation between these metrics.

LEMMA A.32. *Let $\mathcal{O} = (\widehat{\mathbb{C}}, \alpha)$ be a parabolic or hyperbolic orbifold without punctures. Then*

   (i) *$(\widehat{\mathbb{C}}, \omega)$ and $(\widehat{\mathbb{C}}, \sigma)$ are bi-Lipschitz equivalent;*
   (ii) *$\operatorname{id}_{\widehat{\mathbb{C}}} \colon (\widehat{\mathbb{C}}, \omega) \to (\widehat{\mathbb{C}}, \sigma)$ is a quasisymmetry.*

It follows from the behavior of the conformal density ρ of ω near the cone points of $\mathcal{O}$ that the bi-Lipschitz equivalence in (i) cannot be given by the identity map.

If $\mathcal{O}$ has punctures, then $(\widehat{\mathbb{C}}_0, \sigma)$ and $(\widehat{\mathbb{C}}_0, \omega)$ cannot be quasisymmetrically equivalent, since the first metric space is bounded while the other one is not, and a quasisymmetry preserves boundedness of a space.

Proof. (i) We will obtain the desired bi-Lipschitz map $(\widehat{\mathbb{C}}, \omega) \to$ $(\widehat{\mathbb{C}}, \sigma)$ from a quasiconformal map on $\widehat{\mathbb{C}}$ that behaves like a suitable radial stretch near each point in $\operatorname{supp}(\alpha)$.

The *radial stretch* R_β for exponent $\beta > 0$ is the quasiconformal homeomorphism $R_\beta \colon \mathbb{C} \to \mathbb{C}$ defined as

$$R_\beta(re^{i\theta}) = r^\beta e^{i\theta}$$

for $r \geq 0$, $\theta \in [0, 2\pi]$. The map R_β is smooth on $\mathbb{C} \setminus \{0\}$ and we have

$$\|DR_\beta(z)\|_\sigma \asymp |z|^{\beta - 1}$$

for z near 0 (see (A.5) for the notation used here). Note that R_β is the identity on $\partial \mathbb{D}$. This allows us to "cut and paste" radial stretches together to find a homeomorphism $\varphi \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ such that in a small chordal disk U_p centered at a point $p \in \widehat{\mathbb{C}}$ with $\alpha(p) \geq 2$, the map $\varphi|U_p$ conjugates to the radial stretch R_β on $\mathbb{D}$ with $\beta = 1/\alpha(p)$ under a suitable Möbius transformation that sends U_p onto $\mathbb{D}$ and p to 0. Moreover, we require that outside these neighborhoods U_p the map φ is the identity. Then φ is quasiconformal away from the boundaries ∂U_p of the disks where a conjugate of a radial stretch is pasted together with the identity. Since the union of these boundaries ∂U_p is a set of finite Hausdorff 1-measure and such sets are removable for quasiconformal maps (see [**Vä71**, Section 35]), the homeomorphism φ is quasiconformal on $\widehat{\mathbb{C}}$. Moreover, by construction we have $\|D\varphi(q)\|_\sigma \asymp 1$ for almost every q in the complement of a neighborhood of $\operatorname{supp}(\alpha)$ and

$$\|D\varphi(q)\|_\sigma \asymp \sigma(p, q)^{1/\alpha(p) - 1}$$

for q near p with $\alpha(p) \geq 2$.

In order to establish that $\varphi \colon (\widehat{\mathbb{C}}, \omega) \to (\widehat{\mathbb{C}}, \sigma)$ is bi-Lipschitz, we will show that it is a map of *bounded length distortion*. This means that for each path γ in $\widehat{\mathbb{C}}$ we have $\operatorname{length}_\sigma(\varphi(\gamma)) \asymp \operatorname{length}_\omega(\gamma)$, where $C(\asymp)$ is independent of γ. Since ω is a length metric and σ is comparable to a length metric (namely the spherical metric) on $\widehat{\mathbb{C}}$, this will imply that φ is bi-Lipschitz as desired. Since the universal orbifold covering map $\Theta \colon (X, d_0) \to (\widehat{\mathbb{C}}, \omega)$ is a path isometry and every path γ in $\widehat{\mathbb{C}}$ has a lift under Θ (see Lemma A.17), it suffices to show that $\psi := \varphi \circ \Theta$ is of bounded length distortion.

Now $\psi = \varphi \circ \Theta$ is a quasiregular map. If $w \in X$ is near a point $z \in \Theta^{-1}(p)$ with $\alpha(p) \geq 2$ and $q := \Theta(w)$, then

$$\|\Theta'(w)\| \asymp d_0(z, w)^{\alpha(p) - 1} = \omega(p, q)^{\alpha(p) - 1} \asymp \sigma(p, q)^{1 - 1/\alpha(p)},$$

by (A.39) and (A.41). So if we denote by

$$\|D\psi(w)\| := \limsup_{w' \to w} \frac{\sigma(\psi(w'), \psi(w))}{d_0(w', w)}$$

the norm of the differential $D\psi(w)$ with respect to the metric d_0 on X and the chordal metric σ on $\widehat{\mathbb{C}}$, then

$$\|D\psi(w)\| = \|D\varphi(q)\|_\sigma \cdot \|\Theta'(w)\| \asymp 1.$$

This is also true for almost every point w in the complement of a neighborhood of $\Theta^{-1}(\text{supp}(\alpha))$. Hence $\|D\psi(w)\| \asymp 1$ for almost every $w \in X$. Since this is true locally near each point and $\|D\psi(w)\|$ is invariant under precomposition with elements of the deck transformation group of Θ, which acts cocompactly on X (see Proposition A.29 (iv)), we conclude that here $C(\asymp)$ is independent of w.

Since ψ is quasiregular with $\|D\psi\| \asymp 1$, this map is of bounded length distortion (see [**MV88**, Theorem 2.16]), and it follows that $\varphi \colon (\widehat{\mathbb{C}}, \omega) \to (\widehat{\mathbb{C}}, \sigma)$ is indeed a bi-Lipschitz map.

(ii) Since the homeomorphism φ on $\widehat{\mathbb{C}}$ constructed in (i) is quasiconformal, it is a quasisymmetry on $(\widehat{\mathbb{C}}, \sigma)$. Hence $\varphi^{-1} \colon (\widehat{\mathbb{C}}, \sigma) \to (\widehat{\mathbb{C}}, \sigma)$ is a quasisymmetry.

This implies that the map $\text{id}_{\widehat{\mathbb{C}}} \colon (\widehat{\mathbb{C}}, \omega) \to (\widehat{\mathbb{C}}, \sigma)$ is a quasisymmetry, because it is the composition of the bi-Lipschitz map $\varphi \colon (\widehat{\mathbb{C}}, \omega) \to (\widehat{\mathbb{C}}, \sigma)$ followed by the quasisymmetry $\varphi^{-1} \colon (\widehat{\mathbb{C}}, \sigma) \to (\widehat{\mathbb{C}}, \sigma)$. $\qquad\square$

Associated with our orbifold $\mathcal{O} = (\widehat{\mathbb{C}}, \alpha)$ is also a natural Borel measure Ω on $\widehat{\mathbb{C}}$, the *canonical orbifold measure*. To define it, let $\mathcal{L}_{\widehat{\mathbb{C}}}$ be Lebesgue measure (i.e., spherical measure) on $\widehat{\mathbb{C}}$. Here (in contrast to Chapter 19) we do not impose a normalization on $\mathcal{L}_{\widehat{\mathbb{C}}}$ and so $\mathcal{L}_{\widehat{\mathbb{C}}}(\widehat{\mathbb{C}}) = 4\pi$. As before, let ρ be the conformal factor of the orbifold metric ω defined in (A.43). Then for a Borel set $M \subset \widehat{\mathbb{C}}$ we set

(A.56) $$\Omega(M) = \int_M \rho^2 \, d\mathcal{L}_{\widehat{\mathbb{C}}}.$$

We know that ρ is a smooth positive function on $\widehat{\mathbb{C}} \setminus \text{supp}(\alpha)$ and so (A.56) defines a measure on $\widehat{\mathbb{C}}$. Obviously, the measures Ω and $\mathcal{L}_{\widehat{\mathbb{C}}}$ are mutually absolutely continuous. In particular, Ω has no atoms even if $\mathcal{O}$ has punctures. The measure Ω is the natural (conformal) measure induced by the canonical orbifold metric ω with length element $\rho \, d\sigma$.

In the hyperbolic case, Ω is independent of the choice of Θ which underlies the definition of ρ and is hence unique; in the parabolic case Ω is only unique up to a positive multiplicative constant. From the

asymptotics of the conformal factor given in Lemma A.31, it follows that in the hyperbolic case Ω is always a finite measure. In the parabolic case Ω is finite if and only if $\mathcal{O}$ has no punctures. We mention without proof that one can show that actually $\Omega(\widehat{\mathbb{C}}) = -2\pi\chi(\mathcal{O})$ if $\mathcal{O}$ is hyperbolic (this is essentially a special case of [**Be83**, Theorem 10.4.3]).

Let $\mathcal{L}_X$ be the natural measure on the orbifold cover X, namely the Euclidean area in the parabolic case when $X = \mathbb{C}$, and the hyperbolic area in the hyperbolic case when $X = \mathbb{D}$. Then one can consider Ω as the "local" push-forward of $\mathcal{L}_X$ by the universal orbifold covering map Θ. This is made precise in the following statement.

PROPOSITION A.33. *The canonical orbifold measure Ω for a parabolic or hyperbolic orbifold $\mathcal{O} = (\widehat{\mathbb{C}}, \alpha)$ is the unique Borel measure on $\widehat{\mathbb{C}}$ without atoms and with the following property: if $M \subset X$ is a Borel set such that Θ is injective on M, then*

$$(\text{A.57}) \qquad\qquad \mathcal{L}_X(M) = \Omega(\Theta(M)).$$

We will see in the proof that $\Theta(M)$ is also a Borel set.

This proposition can be reformulated in terms of Jacobians (see Section 19.1 for a general discussion of Jacobians): Ω is the unique measure on $\widehat{\mathbb{C}}$ that is absolutely continuous with respect to $\mathcal{L}_{\widehat{\mathbb{C}}}$ such that for the Jacobian $J_{\Theta,\mathcal{L}_X,\Omega}$ of Θ with respect to $\mathcal{L}_X$ and Ω we have $J_{\Theta,\mathcal{L}_X,\Omega} \equiv 1$ $\mathcal{L}_X$-almost everywhere on X.

PROOF. It is clear that Ω has no atoms, i.e., points $p \in \widehat{\mathbb{C}}$ with $\Omega(\{p\}) > 0$, because Ω is absolutely continuous with respect to $\mathcal{L}_{\widehat{\mathbb{C}}}$.

Let M be a Borel set as in the statement. The map Θ is a local biholomorphism on $X \setminus \text{crit}(\Theta)$. So for each point $z \in X \setminus \text{crit}(\Theta)$ there exists a small open ball B centered at z such that Θ maps B biholomorphically onto $\Theta(B)$. This implies that $M \subset X$ is a countable disjoint union of Borel sets each of which is contained in such a ball B and a countable set $C \subset \text{crit}(\Theta)$.

It follows that $\Theta(M)$ is a Borel set. The set $C \subset \text{crit}(\Theta)$ is irrelevant, because $\mathcal{L}_X(C) = 0$ and $\Omega(\Theta(C)) = 0$. So in order to prove (A.57), we may assume that M is contained in a ball B on which Θ is a biholomorphism. Note that by definition of the conformal factor ρ we have

$$\rho(\Theta(z)) = \frac{1}{\|\Theta'(z)\|}$$

for $z \in B$. So the transformation formula implies

$$\Omega(\Theta(M)) = \int_{\Theta(M)} \rho^2 \, d\mathcal{L}_{\widehat{\mathbb{C}}} = \int_M (\rho \circ \Theta)^2 \|\Theta'\|^2 \, d\mathcal{L}_X$$

$$= \int_M \frac{1}{\|\Theta'(z)\|^2} \|\Theta'(z)\|^2 \, d\mathcal{L}_X(z)$$

$$= \int_M d\mathcal{L}_X = \mathcal{L}_X(M),$$

and so (A.57) follows.

With the stated properties the measure Ω is unique. Indeed, (A.57) uniquely determines $\Omega(A)$ for each Borel set A contained in an evenly covered neighborhood of a point $p \in \widehat{\mathbb{C}} \setminus \operatorname{supp}(\alpha)$. If B is an arbitrary Borel set with $B \subset \widehat{\mathbb{C}} \setminus \operatorname{supp}(\alpha)$, then it can be represented as a countable disjoint union of such sets A and so $\Omega(B)$ is uniquely determined. Finally, since we have no atoms, $\Omega(B)$ is uniquely determined for all Borel sets $B \subset \widehat{\mathbb{C}}$. $\qquad \square$

Suppose $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is a rational Thurston map with ramification function α_f. We know (see Proposition 2.12) that the orbifold $\mathcal{O}_f = (\widehat{\mathbb{C}}, \alpha_f)$ is parabolic or hyperbolic. So by the previous discussion we have a canonical orbifold metric $\omega = \omega_f$ for $\mathcal{O}_f$ on $\widehat{\mathbb{C}}_0$, which we will call the *canonical orbifold metric* of f. It is uniquely determined in the hyperbolic case and unique up to a scaling factor in the parabolic case.

One of the most important properties of ω is that the map f is expanding with respect to this metric if f has no periodic critical points, or equivalently, if $\mathcal{O}_f$ has no punctures (see Proposition 2.9 (ii)).

PROPOSITION A.34. *Let $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a rational Thurston map without periodic critical points, and let ω be the canonical orbifold metric of f. Then there exists a constant $\rho > 1$ such that*

(A.58) $$\operatorname{length}_\omega(f \circ \gamma) \geq \rho \operatorname{length}_\omega(\gamma)$$

for all paths γ in $\widehat{\mathbb{C}}$.

This lemma is essentially well known; see [**Mi06a**, Theorem 19.6] or [**CG93**, V.4.3.1], for example. As we will see, if f has a hyperbolic orbifold, then the main idea for the proof is to lift inverse branches of f^{-1} under the universal orbifold covering map Θ to $\mathbb{D}$, and use the fact that holomorphic maps of $\mathbb{D}$ into itself are contracting with respect to the hyperbolic metric.

PROOF. Let $\alpha = \alpha_f$ be the ramification function of f. If $\mathcal{O}_f = (\widehat{\mathbb{C}}, \alpha)$ is parabolic, then f is a Lattès map (see Theorem 3.1) and the

statement follows from Proposition 8.14. So we may assume that $\mathcal{O}_f$ is hyperbolic. Let $\Theta\colon X \to \widehat{\mathbb{C}}$ be the holomorphic universal orbifold covering map defined on $X = \mathbb{D}$ and ω be the canonical orbifold metric of $\mathcal{O}_f$. It is defined on $\widehat{\mathbb{C}}$, because f has no periodic critical points and so $\mathcal{O}_f$ has no punctures.

We will show that there is a constant $\rho > 1$ such that

$$(A.59) \qquad \|f'(q)\|_\omega := \liminf_{q' \to q} \frac{\omega(f(q'), f(q))}{\omega(q', q)} \geq \rho$$

for all $q \in \widehat{\mathbb{C}}$. This expression is the norm of the derivative of f (viewed as a linear map between tangent spaces) with respect to the metric ω. In order to establish this inequality, it is enough to show that for each point $q_0 \in \widehat{\mathbb{C}}$ there exists an open neighborhood N of q_0 and a constant $\rho' > 1$ such that $\|f'(q)\|_\omega > \rho'$ for all $q \in N$. Then the estimate (A.59) follows by covering $\widehat{\mathbb{C}}$ with finitely many such sets N.

So let $q_0 \in \widehat{\mathbb{C}}$ be arbitrary, and set $p_0 := f(q_0)$. We can find points $z_0, w_0 \in \mathbb{D}$ with $\Theta(w_0) = q_0$ and $\Theta(z_0) = p_0$. Then by Lemma A.30 we can find a holomorphic map $A\colon \mathbb{D} \to \mathbb{D}$ such that $A(z_0) = w_0$ and

$$(A.60) \qquad f \circ \Theta \circ A = \Theta.$$

So in this setting, we have the following commutative diagram:

$$(A.61) \qquad \begin{array}{ccc} w_0 \in \mathbb{D} & \xleftarrow{\ \ A\ \ } & z_0 \in \mathbb{D} \\ {\scriptstyle\Theta}\downarrow & & \downarrow{\scriptstyle\Theta} \\ q_0 \in \widehat{\mathbb{C}} & \xrightarrow{\ \ f\ \ } & p_0 \in \widehat{\mathbb{C}}. \end{array}$$

One should think of A as a lift of a suitable inverse branch of f^{-1} under the branched covering map Θ.

By the Schwarz-Pick lemma the derivative of A with respect to the hyperbolic metric d_0 on $\mathbb{D}$ satisfies

$$\|A'(z)\| := \frac{(1 - |z|^2)|A'(z)|}{1 - |A(z)|^2} = \lim_{z' \to z} \frac{d_0(A(z'), A(z))}{d_0(z', z)} \leq 1$$

for all $z \in \mathbb{D}$. Moreover, here $\|A'(z)\| = 1$ for some point $z \in \mathbb{D}$ if and only if A is an automorphism of $\mathbb{D}$.

Let us show that in fact $\|A'(z)\| < 1$ for all $z \in \mathbb{D}$. If not, then A is an automorphism of $\mathbb{D}$. Let $u \in \widehat{\mathbb{C}}$ be arbitrary and $v = f(u)$. Pick $w \in \Theta^{-1}(u)$ and let $z = A^{-1}(w)$. Then by (A.60) we have

$$\Theta(z) = (f \circ \Theta \circ A)(z) = (f \circ \Theta)(w) = f(u) = v.$$

Since Θ is the universal orbifold covering map of $\mathcal{O}_f$, it follows that

$$\alpha(v) = \deg(\Theta, z) = \deg(f \circ \Theta \circ A, z)$$
$$= \deg(f, u) \deg(\Theta, w) \underbrace{\deg(A, z)}_{=1} = \deg(f, u)\alpha(u).$$

This implies that $\mathcal{O}_f$ is parabolic by Proposition 2.13, which is a contradiction.

So in particular, $\|A'(z_0)\| < 1$. Since the map $z \mapsto \|A'(z)\|$ is continuous, we can find a neighborhood U of z_0 and a constant $k < 1$ such that $\|A'(z)\| \leq k < 1$ for all $z \in U$. The set $N := \Theta(A(U))$ is a neighborhood of $\Theta(A(z_0)) = q_0$. If $q \in N$ is arbitrary, then we can pick a point $w \in U$ with $\Theta(A(w)) = q$. Moreover, if $\{q_n\}$ is any sequence contained in $N \setminus \{q\}$ with $q_n \to q$ as $n \to \infty$, then there exists a sequence $\{w_n\}$ in U with $\Theta(A(w_n)) = q_n$ for all $n \in \mathbb{N}$ and $w_n \to w$ as $n \to \infty$. Hence

$$\liminf_{n \to \infty} \frac{\omega(f(q_n), f(q))}{\omega(q_n, q)} = \liminf_{n \to \infty} \frac{\omega((f \circ \Theta \circ A)(w_n), (f \circ \Theta \circ A)(w))}{\omega(\Theta(A(w_n)), \Theta(A(w)))}$$
$$= \liminf_{n \to \infty} \frac{\omega(\Theta(w_n), \Theta(w))}{\omega(\Theta(A(w_n)), \Theta(A(w)))}$$
$$= \liminf_{n \to \infty} \frac{d_0(w_n, w)}{d_0(A(w_n), A(w))}$$
$$= \frac{1}{\|A'(w)\|} \geq 1/k > 1.$$

In the third equality we used the fact that the map Θ is a local radial isometry (see (A.39)). We conclude that $\|f'(q)\|_\omega \geq \rho' := 1/k$ for all q belonging to the neighborhood N of q_0. Since q_0 was arbitrary, inequality (A.59) follows.

Now let γ be a path in $\widehat{\mathbb{C}}$. Inequality (A.59) in combination with a covering argument implies that $\operatorname{length}_\omega(f \circ \gamma) \geq \rho \operatorname{length}_\omega(\gamma)$. So f expands the lengths of paths with respect to the metric ω by the factor $\rho > 1$. $\qquad\square$

Bibliography

[Ah73]   L.V. Ahlfors, *Conformal Invariants: Topics in Geometric Function Theory*, McGraw-Hill, New York, 1973.

[Ah79]   L.V Ahlfors, *Complex Analysis*, 3rd ed., McGraw-Hill, New York, 1979.

[Ar91]   M. Artin, *Algebra*, Prentice Hall, Englewood Cliffs, NJ, 1991.

[AH35]   P.S. Alexandroff and H. Hopf, *Topologie I*, Springer, Berlin, 1935.

[AIM09] K. Astala, T. Iwaniec, and G. Martin, *Elliptic Partial Differential Equations and Quasiconformal Mappings in the Plane*, Princeton Univ. Press, Princeton, NJ, 2009.

[BD]   L. Bartholdi and D. Dudko, *Algorithmic aspects of branched coverings*, preprint, 2015, `http://arxiv.org/abs/1512.05948`.

[BN06]   L. Bartholdi and V. Nekrashevych, *Thurston equivalence of topological polynomials*, Acta Math. 197 (2006), 1–51.

[Be83]   A.F. Beardon, *The Geometry of Discrete Groups*, Springer, New York, 1983.

[Be84]   A.K. Beardon, *A Primer on Riemann Surfaces*, London Math. Soc. Lecture Note Ser. 78, Cambridge Univ. Press, Cambridge, 1984.

[Be87]   M. Berger, *Geometry I*, Universitext, Springer, Berlin, 1987.

[Be94]   J. Bernard, *Dynamique des pertubations d'un example de Lattès*, Ph.D. thesis, Univ. Paris-Sud Orsay, 1994.

[Be61]   L. Bers, *Uniformization by Beltrami equations*, Comm. Pure Appl. Math. 14 (1961), 215–228.

[Bi83]   R.H. Bing, *The Geometric Topology of 3-Manifolds*, Amer. Math. Soc., Providence, RI, 1983.

[Bo06]   M. Bonk, *Quasiconformal geometry of fractals*, in: *Proc. Internat. Congr. Math. (Madrid 2006)*, Vol. II, Eur. Math. Soc., Zürich, 2006, pp. 1349–1373.

[Bo11]   M. Bonk, *Uniformization of Sierpiński carpets in the plane*, Invent. Math. 186 (2011), 559–665.

[BK02]   M. Bonk and B. Kleiner, *Quasisymmetric parametrizations of two-dimensional metric spheres*, Invent. Math. 150 (2002), 127–183.

[BK05]   M. Bonk and B. Kleiner, *Conformal dimension and Gromov hyperbolic groups with 2-sphere boundary*, Geom. Topol. 9 (2005), 219–246.

[BS00]   M. Bonk and O. Schramm, *Embeddings of Gromov hyperbolic spaces*, Geom. Funct. Anal. 10 (2000), 266–306.

[BBY12] S. Bonnot, M. Braverman, and M. Yampolsky, *Thurston equivalence to a rational map is decidable*, Mosc. Math. J. 12 (2012), 747–763, 884.

[BK11]   M. Bourdon and B. Kleiner, *Combinatorial modulus, the combinatorial Loewner property, and Coxeter groups*, Groups Geom. Dyn. 7 (2013), 39–107.

589

[BP03]   M. Bourdon and H. Pajot, *Cohomologie ℓ_p and espaces de Besov*, J. Reine
          Angew. Math. 558 (2003), 85–108.

[Bo72]   R. Bowen, *Entropy-expansive maps*, Trans. Amer. Math. Soc. 164 (1972)
          323–331.

[BS97]   P.L. Bowers and K. Stephenson, *A "regular" pentagonal tiling of the
          plane*, Conform. Geom. Dyn. 1 (1997), 58–86.

[B–P00]   E. Brezin, R. Byrne, J. Levy, K.M. Pilgrim, and K. Plummer, *A census
          of rational maps*, Conform. Geom. Dyn. 4 (2000), 35–74.

[BH99]   M.R. Bridson and A. Haefliger, *Metric Spaces of Non-Positive Curvature*,
          Springer, Berlin, 1999.

[Bu79]   R.B. Burckel, *An Introduction to Classical Complex Analysis*, Birkhäuser,
          Basel, Stuttgart, 1979.

[Bu81]   R.B. Burckel, *Inessential maps and classical Euclidean topology*, in:
          Jahrbuch Überblicke Mathematik 1981, Eds. S.D. Chatterji et al., Bib-
          liographisches Inst., Mannheim, 1981, pp. 119–137.

[BBI01]   D. Burago, Y. Burago, and S. Ivanov, *A Course in Metric Geometry*,
          Amer. Math. Soc., Providence, RI, 2001.

[Bu92]   P. Buser, *Geometry and Spectra of Compact Riemann Surfaces*,
          Birkhäuser, Boston, 1992.

[BS07]   S. Buyalo and V. Schroeder, *Elements of Asymptotic Geometry*, Europ.
          Math. Soc., Zürich, 2007.

[BP06]   K.-U. Bux and R. Pérez, *On the growth of iterated monodromy groups*, in:
          Topological and Asymptotic Aspects of Group Theory, Contemp. Math.,
          394, Amer. Math. Soc., Providence, RI, 2006, pp. 61–76.

[Ca78]   J.W. Cannon, *The recognition problem: what is a topological manifold?*,
          Bull. Amer. Math. Soc. 84 (1978), 832–866.

[Ca94]   J.W. Cannon, *The combinatorial Riemann mapping theorem*, Acta Math.
          173 (1994), 155–234.

[C–P03]   J.W. Cannon, W.J. Floyd, R. Kenyon, and W.R. Parry, *Constructing
          rational maps from subdivision rules*, Conform. Geom. Dyn. 7 (2003),
          76–102.

[CFP01]   J.W. Cannon, W.J. Floyd, and W.R. Parry, *Finite subdivision rules*, Con-
          form. Geom. Dyn. 5 (2001), 153–196.

[CFP06a] J.W. Cannon, W.J. Floyd, and W.R. Parry, *Expansion complexes for
          finite subdivision rules. I*, Conform. Geom. Dyn. 10 (2006), 63–99.

[CFP06b] J.W. Cannon, W.J. Floyd, and W.R. Parry, *Expansion complexes for
          finite subdivision rules. II*, Conform. Geom. Dyn. 10 (2006), 326–354.

[CFP07]   J.W. Cannon, W.J. Floyd, and W.R. Parry, *Constructing subdivision
          rules from rational maps*, Conform. Geom. Dyn. 11 (2007), 128–136.

[CFP10]   J.W. Cannon, W.J. Floyd, and W.R. Parry, *Lattès maps and finite sub-
          divisison rules*, Conform. Geom. Dyn. 14 (2010), 113–140.

[C–P12]   J.W. Cannon, W.J. Floyd, W.R. Parry and K.M. Pilgrim, *Nearly Eu-
          clidean Thurston maps*, Conform. Geom. Dyn. 16 (2012), 209–255.

[CT07]   J.W. Cannon and W.P. Thurston, *Group invariant Peano curves*, Geom.
          Topol. 11 (2007), 1315–1355.

[CG93]   L. Carleson and T.H. Gamelin, *Complex Dynamics*, Springer, New York,
          1993.

[Co92]   J.H. Conway, *The orbifold notation for surface groups*, in: *Groups, Combinatorics & Geometry (Durham, 1990)*, London Math. Soc. Lecture Note Ser. 165, Cambridge Univ. Press, Cambridge, 1992, pp. 438–447.

[CF67]   G.E. Cooke and R.L. Finney, *Homology of Cell Complexes*, Princeton Univ. Press, Princeton, NJ, 1967.

[Co93]   M. Coornaert, *Mesures de Patterson-Sullivan sur le bord d'un espace hyperbolique au sens de Gromov*, Pacific J. Math. 159 (1993), 241–270.

[Da86]   R.J. Daverman, *Decompositions of Manifolds*, Academic Press, Orlando, FL, 1986.

[Di10]   R. Diestel, *Graph Theory*, 4th ed., Springer, Heidelberg, 2010.

[DS97]   G. David and S. Semmes, *Fractured Fractals and Broken Dreams*, Oxford Lecture Series in Mathematics and its Applications 7, Clarendon Press, Oxford Univ. Press, New York, 1997.

[DH84]   A. Douady and J.H. Hubbard, *Étude dynamique des polynômes complexes. I & II*, Publ. Math. d'Orsay, Univ. de Paris-Sud, Dépt. de Math., Orsay, 1984–85.

[DH93]   A. Douady and J.H. Hubbard, *A proof of Thurston's topological characterization of rational functions*, Acta Math. 171 (1993), 263–297.

[FM11]   B. Farb and D. Margalit, *A Primer on the Mapping Class Group*, Princeton Univ. Press, Princeton, NJ, 2011.

[FJ79]   F.T. Farrell and L.E. Jones, *Markov cell structures for expanding maps in dimension* 2, Trans. Amer. Math. Soc. 255 (1979), 315–327.

[Fl50]   E.E. Floyd, *Some characterizations of interior maps*, Ann. of Math. (2) 51 (1950), 571–575.

[Fo81]   O. Forster, *Lectures on Riemann surfaces*, Springer, New York, 1981.

[G–Z]   Y. Gao, P. Haïssinsky, D. Meyer, and J. Zeng, *Invariant Jordan curves of Sierpiński carpet rational maps*, to appear in Ergodic Theory Dynam. Systems.

[Gr84]   R.I. Grigorchuk, *Degrees of growth of finitely generated groups and the theory of invariant means*, Izv. Akad. Nauk SSSR Ser. Mat. 48 (1984), 939–985.

[Gr81]   M. Gromov, *Groups of polynomial growth and expanding maps*, Inst. Hautes Études Sci. Publ. Math. No. 53 (1981), 53–73.

[Gr87]   M. Gromov, *Hyperbolic Groups*, in *Essays in Group Theory (S. Gersten, Ed.)*, MSRI Publ. 8, Springer, New York, 1987, pp. 75–265.

[GT87]   M. Gromov and W.P. Thurston, *Pinching constants for hyperbolic manifolds*, Invent. Math. 89 (1987), 1–12.

[Gro97]   A. Grothendieck, *Esquisse d'un programme*, in: *Geometric Galois actions, 1*, London Math. Soc. Lecture Note Ser. 242, Cambridge Univ. Press, Cambridge, 1997, pp. 5–48.

[GH90]   E. Ghys and P. de la Harpe, Eds., *Sur les groupes hyperboliques d'après Mikhael Gromov*, Progress in Math. 83, Birkhäuser, Boston, 1990.

[GPS90]   P. Grzegorczyk, F. Przytycki, W. Szlenk, *On iterations of Misiurewicz's rational maps on the Riemann sphere*, Ann. Inst. H. Poincaré Phys. Théor. 53 (1990), 431–444.

[Ha02]   A. Hatcher, *Algebraic Topology*, Cambridge Univ. Press, Cambridge, 2002.

[HP08]   P. Haïssinsky and K.M. Pilgrim, *Thurston obstructions and Ahlfors regular conformal dimension*, J. Math. Pures Appl. (9) 90 (2008), 229–241.

[HP09]   P. Haïssinsky and K.M. Pilgrim, *Coarse Expanding Conformal Dynamics*, Astérisque 325 (2009).

[HP12a] P. Haïssinsky and K.M. Pilgrim, *Quasisymmetrically inequivalent hyperbolic Julia sets*, Rev. Mat. Iberoam. 28 (2012), 1025–1034.

[HP12b] P. Haïssinsky and K.M. Pilgrim, *An algebraic characterization of expanding Thurston maps*, J. Mod. Dyn. 6 (2012), 451–476.

[HP14]   P. Haïssinsky and K.M. Pilgrim, *Minimal Ahlfors regular conformal dimension of coarse expanding conformal dynamics on the sphere*, Duke Math. J. 163 (2014), 2517–2559.

[He01]   J. Heinonen, *Lectures on Analysis on Metric Spaces*, Springer, New York, 2001.

[HH02]   D. Heicklen and C. Hoffman, *Rational maps are d-adic Bernoulli*, Ann. of Math. (2) 156 (2002), 103–114.

[HM]   M. Hlushchanka and D. Meyer, *Exponential growth of some iterated monodromy groups*, preprint.

[Hu93]   J.H. Hubbard, *Local connectivity of Julia sets and bifurcation loci: three theorems of J.-C. Yoccoz*, in: *Topological Methods in Modern Mathematics (Stony Brook, NY, 1991)*, Publish or Perish, Houston, TX, 1993, pp. 467–511.

[Hu06]   J.H. Hubbard, *Teichmüller Theory and Applications to Geometry, Topology, and Dynamics. Vol. 1. Teichmüller Theory*, Matrix Editions, Ithaca, NY, 2006.

[Hu16]   J.H. Hubbard, *Teichmüller Theory and Applications to Geometry, Topology, and Dynamics. Vol. 2. Surface Homeomorphisms and Rational Functions*, Matrix Editions, Ithaca, NY, 2016.

[IM01]   T. Iwaniec and G. Martin, *Geometric Function Theory and Non-Linear Analysis*, Clarendon Press, Oxford, 2001.

[Jo98]   M. Jonsson, *Some properties of 2-critically finite holomorphic maps of P^2*, Ergodic Theory Dynam. Systems 18 (1998), 171–187.

[KPT]   J. Kahn, K.M. Pilgrim, and D.P. Thurston, *Conformal surface embeddings and extremal length*, preprint, 2015, `http://arxiv.org/abs/1507.05294`.

[Ka03]   A. Kameyama, *On Julia sets of postcritically finite branched coverings. I. Coding of Julia sets*, J. Math. Soc. Japan 55 (2003), 439–454.

[KK00]   M. Kapovich and B. Kleiner, *Hyperbolic groups with low-dimensional boundary*, Ann. Sci. École Norm. Sup. 33 (2000), 647–669.

[KB02]   I. Kapovich and N. Benakli, *Boundaries of hyperbolic groups*, in: *Combinatorial and Geometric Group Theory (New York, 2000/Hoboken, NJ, 2001)*, Contemp. Math. 296, Amer. Math. Soc., Providence, RI, 2002, pp. 39–93.

[KH95]   A. Katok and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems*, Cambridge Univ. Press, Cambridge, 1995.

[Ki14]   J. Kigami, *Quasisymmetric modification of metrics on self-similar sets*, in: *Geometry and analysis of fractals*, Springer Proc. Math. Stat. 88, Springer, Heidelberg, 2014, pp. 253–282.

[Kl06] B. Kleiner, *The asymptotic geometry of negatively curved spaces: uniformization, geometrization and rigidity*, in: *Proc. Internat. Congr. Math. (Madrid 2006)*, Vol. II, Eur. Math. Soc., Zürich, 2006, pp. 743–768.

[Kl10] B. Kleiner, *A new proof of Gromov's theorem on groups of polynomial growth*, J. Amer. Math. Soc. 23 (2010), 815–829.

[La09] K. Lamotke, *Riemannsche Flächen*, 2nd ed., Springer, Berlin, 2009.

[La18] S. Lattès, *Sur l'itération des substitutions rationnelles et les fonctions de Poincaré*, C. R. Acad. Sci. Paris 166 (1918), 26–28.

[Li15a] Z. Li, *Weak expansion properties and large deviation principles for expanding Thurston maps*, Adv. Math. 285 (2015), 515–567.

[Li15b] Z. Li, *Equilibrium states for expanding Thurston maps*, Preprint, 2015.

[Li16a] Z. Li, *Periodic points and the measure of maximal entropy of an expanding Thurston map*, to appear in Trans. Amer. Math. Soc., 2016.

[Li16b] Z. Li, *Ergodic Theory of Expanding Thurston Maps*, to appear in the Atlantis Series in Dynamical Systems, Springer, 2016.

[LZ04] S. K. Lando and A. K. Zvonkin, *Graphs on Surfaces and their Applications*, Springer, Berlin, 2004.

[Ly83] M. Lyubich, *Entropy properties of rational endomorphisms of the Riemann sphere*, Ergodic Theory Dynam. Systems 3 (1983), 351–385.

[MT10] J.M. Mackay and J.T. Tyson, *Conformal Dimension. Theory and Application,* Univ. Lecture Ser. 54, Amer. Math. Soc., Providence, RI, 2010.

[MSS83] R. Mañé, P. Sad, and D. Sullivan, *On the dynamics of rational maps*, Ann. Sci. Ec. Norm. Sup. 16 (1983), 193–217.

[MV88] O. Martio and J. Väisälä, *Elliptic equations and maps of bounded length distortion*, Math. Ann. 282 (1988), 423–443.

[McM94a] C.T. McMullen, *Complex Dynamics and Renormalization*, Princeton Univ. Press, Princeton, NJ, 1994.

[McM94b] C.T. McMullen, *Frontiers in complex dynamics*, Bull. Amer. Math. Soc. 31 (1994), 155–172.

[Mer10] S. Merenkov, *A Sierpiński carpet with the co-Hopfian property*, Invent. Math. 180 (2010), 361–388.

[Me02] D. Meyer, *Quasisymmetric embedding of self similar surfaces and origami with rational maps*, Ann. Acad. Sci. Fenn. Math. 27 (2002), 461–484.

[Me09a] D. Meyer, *Dimension of elliptic harmonic measure of snowspheres*, Illinois J. Math. 53 (2009), 691–721.

[Me09b] D. Meyer, *Expanding Thurston maps as quotients*, preprint, 2009, `http://arxiv.org/abs/0910.2003`.

[Me10] D. Meyer, *Snowballs are quasiballs*, Trans. Amer. Math. Soc. 362 (2010), 1247–1300.

[MP12] D. Meyer and C. L. Petersen, *On the notions of mating*, Ann. Fac. Sci. Toulouse Math. 21 (2012), 839–876.

[Me13] D. Meyer, *Invariant Peano curves of expanding Thurston maps*, Acta Math. 210 (2013), 95–171.

[Me14] D. Meyer, *Unmating of rational maps, sufficient criteria and examples*, in: *Frontiers in Complex Dynamics*, Princeton Math. Ser. 51, Princeton Univ. Press, Princeton, NJ, 2014, pp. 197–234.

[Mi68] J. Milnor, *Problem 5603*, Amer. Math. Monthly 75 (1968), 685–686.

[MT88] J. Milnor and W. Thurston, *On iterated maps of the interval*, in: *Dynamical systems (College Park, MD, 1986–87)*, Lecture Notes in Math. 1342, Springer, Berlin, 1988, pp. 465–563.

[Mi00] J. Milnor, *Local connectivity of Julia sets: expository lectures*, in: *The Mandelbrot Set, Theme and Variations*, London Math. Soc. Lecture Note Ser. 274, Cambridge Univ. Press, Cambridge, 2000, pp. 67–116.

[Mi04] J. Milnor, *Pasting together Julia sets: a worked out example of mating*, Experiment. Math. 13 (2004), 55–92.

[Mi06a] J. Milnor, *Dynamics in One Complex Variable*, 3rd ed., Princeton Univ. Press, Princeton, NJ, 2006.

[Mi06b] J. Milnor, *On Lattès maps*, in: *Dynamics on the Riemann Sphere*, Eur. Math. Soc., Zürich, 2006, pp. 9–43.

[Mi76] M. Misiurewicz, *Diffeomorphism without any measure with maximal entropy*, Bull. Acad. Pol. Sci. 21 (1973) 903–910.

[Mo77] E.E. Moise, *Geometric Topology in Dimensions 2 and 3*, Springer, New York, 1977.

[Mo25] R.L. Moore, *Concerning upper semicontinuous collections of compacta*, Trans. Amer. Math. Soc. 27 (1925), 416–426.

[Mo78] G.D. Mostow, *Strong Rigidity of Locally Symmetric Spaces*, Annals of Mathematics Studies 78, Princeton Univ. Press, Princeton, NJ, 1973.

[Ne05] V. Nekrashevych, *Self-similar Groups*, Amer. Math. Soc., Providence, RI, 2005.

[Ne53] R. Nevanlinna, *Eindeutige analytische Funktionen*, Springer, Berlin, 1953.

[Pi00] K.M. Pilgrim, *Dessins d'enfant and Hubbard trees*, Ann. Scient. Éc. Norm. Sup. 33 (2000), 671–693.

[Pi01] K.M. Pilgrim, *Canonical Thurston obstruction*, Adv. Math. 158 (2001), 154–168.

[PT98] K.M. Pilgrim and Tan Lei, *Combining rational maps and controlling obstructions*, Ergodic Theory Dynam. Systems 18 (1998), 221–245.

[Po92] C. Pommerenke, *Boundary Behaviour of Conformal Maps*, Springer, Berlin, 1992.

[Pr85] F. Przytycki, *Hausdorff dimension of harmonic measure on the boundary of an attractive basin for a holomorphic map*, Invent. Math. 80 (1985), 161–179.

[PU10] F. Przytycki and M. Urbański, *Conformal Fractals: Ergodic Theory Methods*, London Math. Soc. Lecture Note Ser. 371, Cambridge Univ. Press, Cambridge, 2010.

[Re86] M. Rees, *Positive measure sets of ergodic rational maps*, Ann. Sci. École Norm. Sup. 19 (1986), 383–407.

[Re92] M. Rees, *A partial description of parameter space of rational maps of degree two. I*, Acta Math. 168 (1992), 11–87.

[Re15] M. Rees, *Persistent Markov partitions for rational maps*, preprint, 2015, http://arxiv.org/abs/1306.6166.

[Re98] R. Remmert, *Classical Topics in Complex Function Theory*, Graduate Texts in Mathematics 172, Springer, New York, 1998.

[Ri93] S. Rickman, *Quasiregular Mappings*, Springer, Berlin, 1993.

[Ri90] T.J. Rivlin, *Chebyshev Polynomials. From Approximation Theory to Algebra and Number Theory*, 2nd ed., Wiley, New York, 1990.

[Ru87]   W. Rudin, *Real and Complex Analysis*, 3rd ed., McGraw-Hill, New York, 1987.

[Se93]   S. Semmes, *Bi-Lipschitz mappings and strong A_∞ weights*, Ann. Acad. Sci. Fenn. Ser. A I Math. 18 (1993), 211–248.

[Se96a]   S. Semmes, *On the nonexistence of bi-Lipschitz parameterizations and geometric problems about A_∞-weights*, Rev. Mat. Iberoamericana 12 (1996), 337–410.

[Se96b]   S. Semmes, *Good metric spaces without good parameterizations*, Rev. Mat. Iberoamericana 12 (1996), 187–275.

[Se99]   S. Semmes, *Metric spaces and mappings seen at many scales*, in: M. Gromov, *Metric Structures for Riemannian and Non-Riemannian Spaces*, Birkhäuser, Boston, 1999, pp. 401–518.

[Sh00]   M. Shishikura, *On a theorem of M. Rees for matings of polynomials*, in: *The Mandelbrot set, theme and variations*, London Math. Soc. Lecture Note Ser. 274, Cambridge Univ. Press, Cambridge, 2000, pp. 289–305.

[Sh69]   M. Shub, *Endomorphisms of compact differentiable manifolds*, Amer. J Math. 91 (1969), 175–199.

[StS05]   E.M. Stein and R. Shakarchi, *Real Analysis*, Princeton Univ. Press, Princeton, NJ, 2005.

[St28]   S. Stoïlow, *Sur les transformations continues et la topologique des fonctions analytiques*, Ann. Sci. École Norm. Sup. (3) 45 (1928), 347–382.

[Su85]   D. Sullivan, *Quasiconformal homeomorphisms and dynamics. II. Structural stability implies hyperbolicity for Kleinian groups*, Acta Math. 155 (1985), 243–260.

[Ta92]   Tan Lei, *Mating of quadratic polynomials*, Ergodic Theory Dynam. Systems 12 (1992), 589–620.

[Th16]   D.P. Thurston, *From rubber bands to rational maps: a research report*, Research in the Mathematical Sciences 3 (2016), 1–49.

[Th]   W. Thurston, *The Geometry and Topology of Three-Manifolds*, lecture notes available at `http://library.msri.org/books/gt3m/`.

[Ti16]   G. Tiozzo, *Continuity of core entropy of quadratic polynomials*, Invent. Math. 203 (2016), 891–921.

[TV80]   P. Tukia and J. Väisälä, *Quasisymmetric embeddings of metric spaces*, Ann. Acad. Sci. Fenn. Ser. A I Math. 5 (1980), 97–114.

[Vä71]   J. Väisälä, *Lectures on n-Dimensional Quasiconformal Mappings*, Springer, Berlin-Heidelberg-New York, 1971.

[Vä87]   J. Väisälä, *Porous sets and quasisymmetric maps*, Trans. Amer. Math. Soc. 299 (1987), 525–533.

[Vä88]   J. Väisälä, *Quasisymmetric maps of products of curves into the plane*, Rev. Roumaine Math. Pures Appl. 33 (1988), 147–156.

[Wa82]   P. Walters, *An Introduction to Ergodic Theory*, Springer, New York-Berlin, 1982.

[We52]   H. Weyl, *Symmetry*, Princeton Univ. Press, Princeton, NJ, 1952.

[Wi08]   K. Wildrick, *Quasisymmetric parametrizations of two-dimensional metric planes*, Proc. Lond. Math. Soc. 97 (2008), 783–812.

[Wh42]   G.T. Whyburn, *Analytic Topology*, AMS Colloquium Publications 28, Providence, RI, 1942.

[Yi15a]   Qian Yin, *Lattès maps and combinatorial expansion*, Ergodic Theory Dynam. Systems 6 (2015), 1–36.

[Yi15b]   Qian Yin, *Thurston maps and asymptotic upper curvature*, Geom. Dedicata 176 (2015), 271–293.

[Zd90]   A. Zdunik, *Parabolic orbifolds and the dimension of the maximal measure for rational maps*, Invent. Math. 99 (1990), 627–649.

Index