PLANE TREES AND SHABAT POLYNOMIALS

DAOJI HUANG, LEO OLIVARES, BEN STRASSER, AND ADAM ZWEBER

ABSTRACT. In their expository paper *Plane Trees and Algebraic Numbers* [7], Shabat and Zvonkin discuss how each combinatorial bicolored plane tree can be realized as the preimage of a segment between two critical values of a complex polynomial. This 'Shabat polynomial' is unique up to automorphism of the complex plane. Shabat and Zvonkin illustrate in broad strokes many surprising and useful consequences of this correspondence; both in how the world of plane trees and influence can inform the world of polynomials, and vice-verse.

This paper provides a natural companion to Shabat and Zvonkin, by filling in many of the details omitted or glossed over in the original paper. Specifically, we develop the necessary theory to prove the above correspondence from basic undergraduate-level ideas and tools. Following this theme, we provide detailed exposition on the composition of Shabat polynomials and their corresponding trees, the Galois-theoretic properties of bicolored plane trees, and how trees determine solutions to the Pell equation for polynomials and continued fractions of polynomials. The theory in this paper is as self-contained and elementary as possible.

CONTENTS

1. INTRODUCTION

Our paper investigates a strong connection between the combinatorial and the algebraic. Specifically, we study combinatorial bicolored plane trees and Shabat polynomials. For the first step in our investigation, let us study the preimages of segments in $\mathbb{C}$ under polynomials. If n is the degree of a polynomial $f \in \mathbb{C}[z]$, then in general, each point in $\mathbb{C}$ will have n preimages. Therefore, the preimage of an arbitrary segment in the image of f will in general be n disconnected (not necessarily linear) curves, as illustrated in Figure 1.

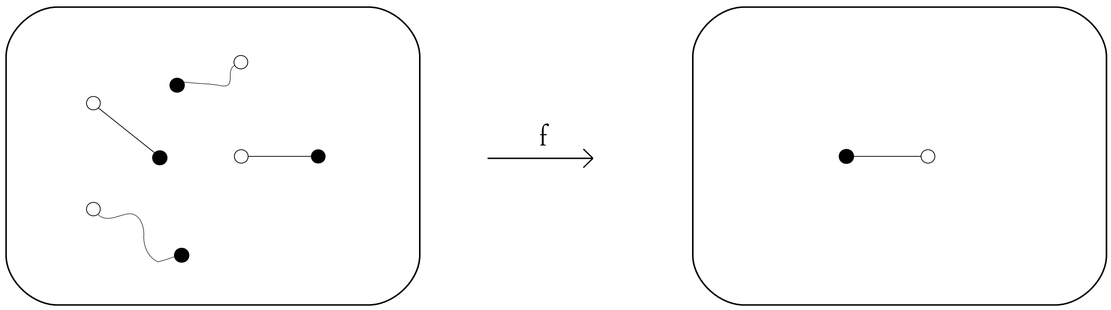


FIGURE 1. The Preimage of an Arbitrary Segment

We can make this situation more interesting if we examine the so called critical values of f.

Definition 1.1. *For a polynomial f, the* critical values *of f are the images of its critical points, that is, the set of points $f(c) \in \mathbb{C}$ such that $f'(c) = 0$.*

An important observation to make is that around the critical points of f, the map f is many-to-one, and thus not bijective. When a segment lies between the two critical values of f, some of its preimages will therefore 'link up', as in Figure 2.

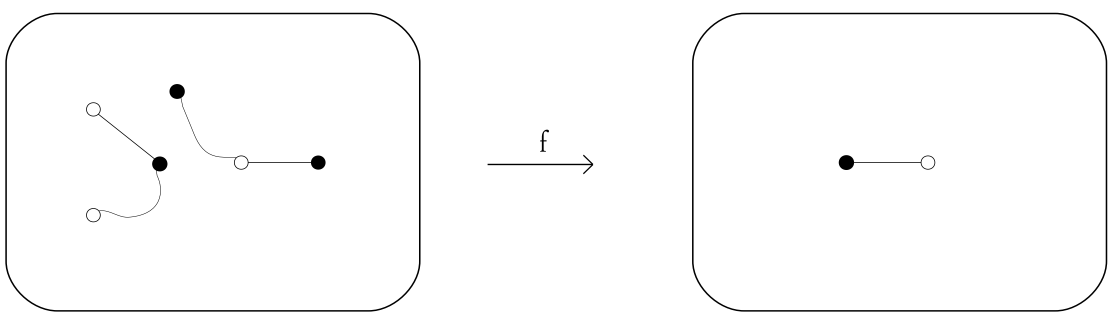


FIGURE 2. The Preimage of a Segment Between Two Critical Values

To get an even more interesting picture, we can restrict the critical values of f.

Definition 1.2. *A polynomial f over $\mathbb{C}$ is* Shabat *if it has at most two critical values.*

When f has exactly two critical values (i.e. f is Shabat), then the preimage of a segment between the two critical values of f is a bicolored plane tree, as seen in Figure 3.

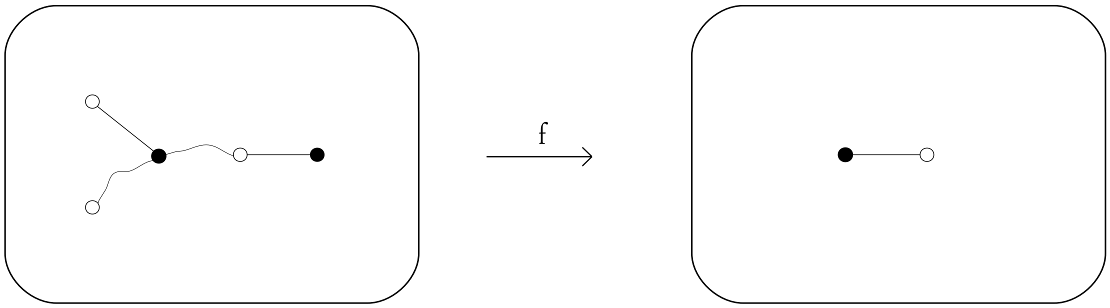


FIGURE 3. The Preimage of a Segment Between the Two Critical Values of a Shabat Polynomial

Definition 1.3. *A* bicolored plane tree *is a tree T embedded in a plane (with no edge crossings) such that each vertex of T has one of two colors, and each edge contains one vertex of each color.*

In our paper, we will study the connections between plane trees and Shabat polynomials, generally understood with natural notions of equivalence described in Section 2. In Section 2, we develop the basic tools of covering spaces and Riemann surfaces in order to prove Theorem 1.1 of [7], that is, up to equivalence, there is a bijection between bicolored plane trees and Shabat polynomials such that each tree is the preimage of a segment between the two critical values of the Shabat polynomial. In Section 3, we will compute numerous examples absent from Sections 2 and 5 of [7], illustrating an algorithmic approach one can take to calculate polynomials from their corresponding trees. Later, in Section 4, we provide a detailed exposition of the Galois-theoretic results described Section 4 of [7], showing how the algebraic properties of Shabat polynomials can affect their corresponding trees by defining a faithful action of the absolute Galois group $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$ on the set of bicolored plane trees. Finally, in Section 5, we further develop the links bewteen Abel's equation, polynomial continued fractions, and bicolored plane trees which are hinted at in Section 7 of [7].

We will begin by proving a proposition which we alluded to earlier.

Proposition 1.4. (Theorem 1.1 of [7]) *For each Shabat polynomial f, the inverse image of a segment between the two critical values of f is a bicolored plane tree, colored by the critical values which each vertex maps to.*

Proof. Let T be the preimage of a segment under f, and let c_1 and c_2 be the two critical value of f. If we understand the vertices of T to be the points which map to critical values of f and we color them based on which critical values they map to, T is naturally a bicolored plane tree. Our proof proceeds in two parts. First, we will show that T is acyclic, and second we will show that T has exactly one more vertex than edge. The second claim completes the proof since an acyclic graph with one more vertex than edge is necessarily a tree.

Claim 1: T is acyclic.

Assume by way of contradiction that there is some cycle in T, as illustrated in Figure 4.

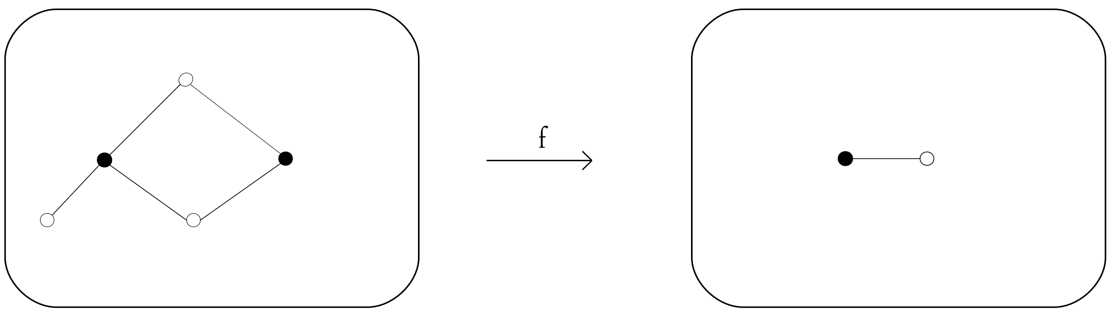


FIGURE 4. A Cycle as the Preimage of a Segment

Consider the region R bounded by a cycle in T (where R includes the boundary). Since f is continuous, each edge in the boundary of R must map to an edge in the boundary of $f(R)$, and $f(R)$ is compact. However, each edge understood with an orientation, either white to black or black to white. We will say a positive edge is one which, when viewed from inside R, has the white vertex to the left of the black vertex. In other words, when traversing the boundary of R in the positive direction, the black vertex is the first point in the edge which is hit. Since the boundary of R is a bicolored cycle, it has both positive and negative edges, meaning the image of R has both positive and negative edges. Since each edge in T maps to a single edge, that edge is necessarily contained in R, contradicting it being on the boundary. To see this, imagine a thin, open band in the subspace topology of R around two edges incident to the same vertex. This open band must contain the segment in $f(R)$, as in Figure 5.

Claim 2: If n is the degree of f, then T has n edges and $n+1$ vertices.

Since f has degree n and the segment between c_1 and c_2 has only noncritical values in it, the tree T associated with f must have n edges, since noncritical values of f have n preimages. As each vertex in T is the preimage of c_1 or c_2, the vertices of T are

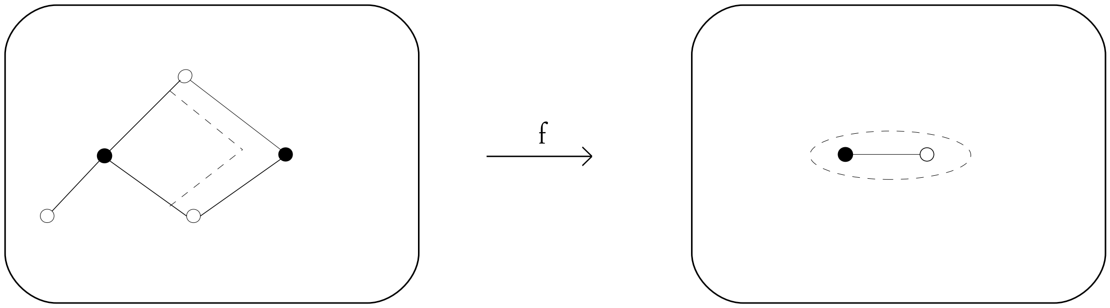


FIGURE 5. An Open Band in the Subspace Topology of R

exactly the roots of the equations

$$\begin{aligned} f(z) &= c_1, \qquad \text{and} \\ f(z) &= c_2. \end{aligned}$$

Counting multiplicity, there are a total of $2n$ solutions to the above equations. Around each solution, the map f is k to 1, for $k \in \mathbb{N}$. The multiplicity of each solution is k, which will be the degree of the corresponding vertex. Therefore, if V is the vertex set in T,

$$|V| = 2n - \sum_{v \in V} (\mathrm{Deg}(v) - 1)$$

where the final sum counts the repeated roots of the above equations. However, each vertex maps to a critical value, and vertices of degree greater than 1 are the critical points of f, since around these points f is many-to-one (and not a local homeomorphism). These points are the solutions to both $f(z) = c_i$ and its derivative, $f'(z) = 0$. If they have multiplicity k in $f(z) = c_i$, then as vertices they have degree k, and they necessarily have multiplicity $k-1$ in $f'(z) = 0$. Therefore, the sum in the above equation is actually the total number of roots to the equation $f'(z) = 0$, which is $n-1$. From here, we get

$$\begin{aligned} |V| &= 2n - (n-1) \\ &= n+1. \end{aligned}$$

Therefore, T is a connected graph with n edges and $n+1$ vertices, so T is a tree. As T is embedded in a plane and has a natural 2-coloring (with color classes being the points mapping to c_1 and those mapping to c_2), we have shown that T is a bicolored plane tree. $\square$

2. The Link Between Plane Trees and Shabat Polynomials

Before we begin, we must establish some notions of equivalence for both Shabat polynomials and bicolored plane trees.

Definition 2.1. *Two bicolored plane trees are* isomorphic *if there exists an orientation-preserving transformation of the plane which takes one tree to the other, respecting the edges, vertices, and coloring. An isomorphism class of bicolored plane trees is a* combinatorial bicolored plane tree.

Note that bicolored plane trees have a stronger sense of isomorphism than ordinary trees, as illustrated by Figures 6 and 7.

FIGURE 6. Two Nonequivalent Combinatorial Plane Trees

FIGURE 7. Two Nonequivalent Combinatorial Bicolored Plane Trees

Definition 2.2. *Two Shabat polynomials f and g are equivalent (written $f \sim g$) if there exist constants a, b, A, and B $(a, A \neq 0)$ such that*

$$Af(az + b) + B = g(z)$$

for all $z \in \mathbb{C}$.

While the above notions of equivalence may or may not seem natural, they are in part motivated by the following theorem, which we shall prove in this section.

Theorem 2.3. (Theorem 1.1 of [7]) *For every combinatorial bicolored plane tree there exists a unique Shabat polynomial (up to equivalence) which has that tree as the preimage of a segment between its critical values.*

By Proposition 1.4, we already know that every Shabat polynomial produces such a tree as the preimage of a segment between its critical values, so combining these two results we will be able to say

Theorem 2.4. *There is a bijection between combinatorial bicolored plane trees and equivalence classes of Shabat polynomials such that each tree can be realized as the preimage of a segment between the two critical values of the corresponding polynomial.*

This proof of Theorem 2.3 will follow in two main phases:

Phase 1: The Topological Map:
For any combinatorial bicolored plane tree T there exists a unique topological covering from a sphere X punctured at the vertices of T (and infinity) to a sphere Y punctured at three points such that the edges in T all map to a single segment in Y between two of these punctured points.

Phase 2: The Analytic Map:
For a specific topological map given above, there exists a unique way to assign coordinates such that the map is holomorphic on the punctured sphere. There is then a unique way to fill in the holes so that the map becomes meromorphic, and any meromorphic function on the Riemann sphere fixing infinity is a polynomial.

2.1. **The Topological Map.** We proceed in two parts. First, we show the existence of the topological map by explicit construction. Then we will show the uniqueness of that same map by developing the theory of covering spaces.

2.1.1. *Constructing a Covering from a Combinatorial Bicolored Plane Tree.* We begin by rigorously defining the objects we are working with.

Definition 2.5. *Let X and Y be two path connected topological spaces. An* (unramified) covering *of Y by X is a continuous function $f : X \to Y$ (denoted (X, f)) such that for all $y \in Y$ there exists an open $V_y \subseteq Y$ containing y such that $f^{-1}(V_y) \subset X$ is a disjoint union of open sets, each mapped homeomorphically onto V_y by f. Each connected component of $f^{-1}(V_y)$ is a* sheet *of the covering, and $|S|$ is the* degree *of the covering, which is the number of sheets. For any $y \in Y$, the set $f^{-1}(y)$ is called the* fiber *of y.*

Definition 2.6. *Two coverings of Y by X and X' are* isomorphic *if there exists a homeomorphism from X to X' respecting the covering maps $f : X \to Y$ and $f' : X' \to Y$.*

Let T be a combinatorial bicolored plane tree. By a stereographic projection, we can imagine T drawn on a sphere X. Puncture X at every vertex of T and infinity, and call the result X'. Let Y be a sphere, and puncture Y at three points (one of which we will call infinity), calling the result Y'. We will construct a continuous map (presented in [7]) $f : X \to Y$ such that every edge in T maps to a single segment between y_1 and y_2, the two non-infinite holes in Y, and where f is a covering of Y' when restricted to X'. For convenience, let the outer boundary of our figures be the point at infinity. As a running example, we will show the construction for the tree in Figure 8.

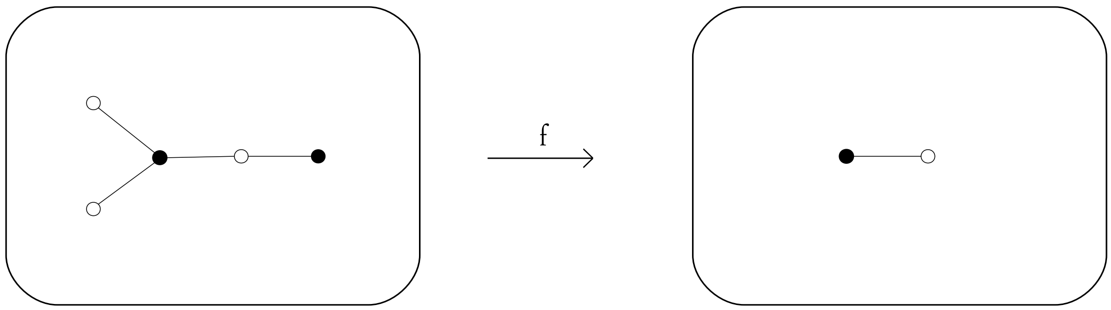


Figure 8. A Combinatorial Bicolored Plane Tree as the Preimage of a Segment

First, we will triangulate X by drawing lines from each vertex to infinity in such a way that each edge in T is incident to at least one triangle, and that no two of them are incident to the same triangle. In other words, each triangle has a black vertex, a white vertex, and a vertex at infinity. We can do the same with Y, by drawing a triangle between y_1, y_2, and ∞. This triangulation is illustrated in Figure 9

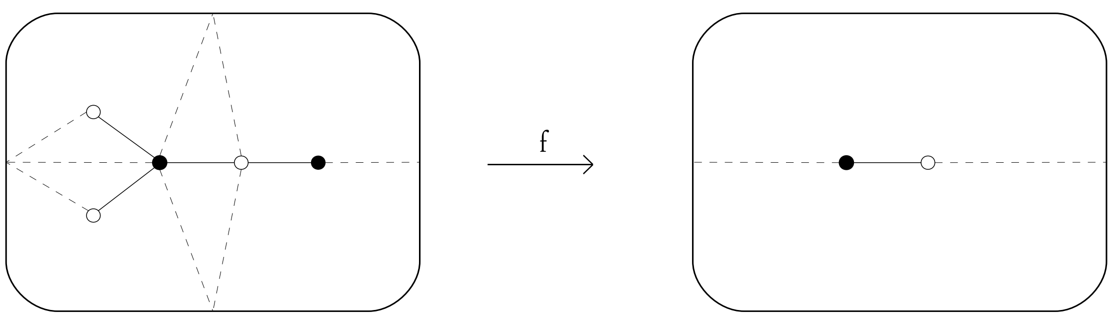


Figure 9. A Triangulation of the Sphere Respecting the Map

Remember, the boundary of each figure is a single point at infinity, so each triangle has a single edge from the tree and two new edges. Therefore, the image has been partitioned into exactly two triangles. We will homeomorphically map each triangle in the preimage to a triangle in the image in such a way that the edges are preserved (with orientation). We will call a 'positive' triangle any triangle which has its vertices in the counter-clockwise order of black, white, infinity. Similarly, 'negative' triangles will be those with the counter-clockwise order white, black, infinity. This designation is illustrated in Figure 10.

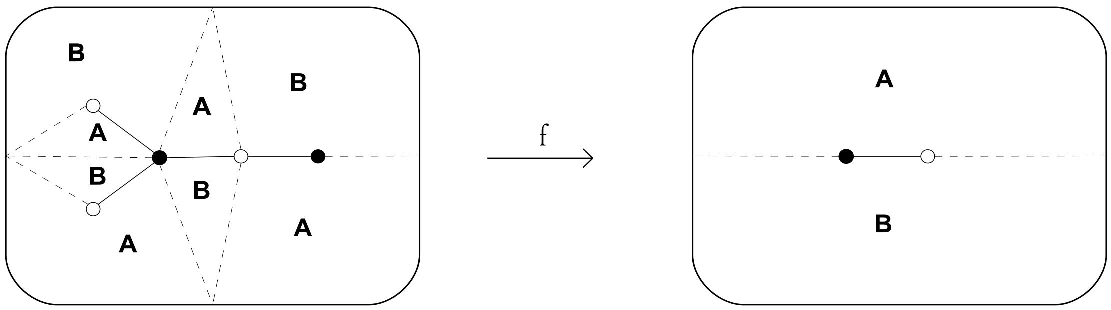


FIGURE 10. The Labeled Triangulation

The positive triangles in each image have been labeled **A**, and the negative have been labeled **B**. We construct our map by sending the **A** regions to the **A** region, and the **B** regions to the **B** region. By construction, each positive triangle will be adjacent to only negative triangles and vice-versa. Every positive triangle only shares edges with negative triangles and vice-versa because any triangle sharing an edge with a positive triangle will necessarily have the opposite orientation along that edge.

As each triangle is a simply connected compact set, we can certainly map each individual triangle in X homeomorphically its associated half plane in Y in such a way that preserves the edges. Furthermore, our mapping respects the orientation of the triangles, so we can insist that our mapping agrees on all of the edges. Therefore, our mapping is indeed a continuous function from X to Y. If we restrict our construction to X' and Y' (the two punctured spheres), then each point in Y' has half as many preimages in X' as there are triangles in our triangulation of X. Any point $y \in Y'$ contained in one of the triangles of the image will have a simply connected open neighborhood U entirely contained in that triangle, so the preimage of U will be a disjoint union of simply connected open sets, each of which is contained in a triangle in the preimage and homeomorphic to U by f, illustrated in Figure 11. Note that the preimage of a point in a positive triangle is a point in each positive triangle.

If $y \in Y'$ is on the edge of a triangle, then there is a single preimage of y on each preimage of the edge. Since X' is Hausdorff, we can give each point in $E = f^{-1}(y)$ a simply connected open neighborhood disjoint from the open neighborhood of each other point in E. The intersection of the images of these open neighborhoods is a simply connected open set U in Y' containing y. The preimage of the set U will be a disjoint union of open sets in X', each of which is homeomorphic to U by f. Therefore,

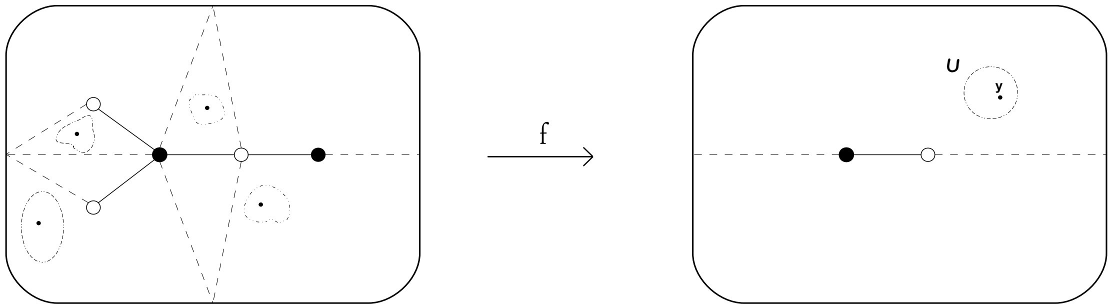


FIGURE 11. An Illustration of the Local Homeomorphism

in all cases we satisfy the covering property, so f is a covering of X' by Y' which maps every edge in the tree to a single edge in such a way that respects orientation.

Definition 2.7. *Given a tree T, the covering (X', f) of Y' described above is the covering* determined *by T.*

We will now show how to identify the covering constructed above with a unique subgroup $M \leq \pi_1(Y', y_0)$ where $M \cong \pi_1(X', x_0)$ and $x_0 \in f^{-1}(y_0)$. Later, when we begin our study of covering spaces, we will show that this uniquely determines our covering up to homeomorphism.

The following definitions and constructions are from [6].

Definition 2.8. *A sequence of permutations $[g_1, \ldots, g_k]$ on n letters is called a* constellation *(or* k-constellation*) if $G = \langle g_1, \ldots, g_k \rangle$ acts transitively on n letters and $g_1 g_2 \ldots g_k = e$. The group G is called the* cartographic group *of the constellation.*

Let $f : X' \to Y'$ be a covering of Y' by X'. Let $y_0 \in Y'$ be given, and consider $E = f^{-1}(y_0)$, the fiber of y_0. Note that if $\gamma \in \pi_1(Y', y_0)$, then $f^{-1}(\gamma)$ is a set of paths beginning and ending at points in E. As these paths are oriented (by the orientation of f), this set of paths induces a natural map $g : E \to E$ sending each path's starting point to its ending point. This induced map is necessarily invertible as f is locally bijective and γ is invertible. Hence, g is a permutation. The group

$$G = \{g_i \text{ corresponding to } f^{-1}(\gamma_i) : \gamma_i \in \pi_1(Y, y_0)\}$$

acting on the set E is called the *monodromy group* of the covering. In our situation, Y' and X' are punctured spheres (and therefore path connected), so different choices of $y_0 \in Y'$ result in isomorphic monodromy groups. Furthermore, the path-connectedness of X' implies G acts transitively on E since for any $x_1, x_2 \in E$ there is a directed path

γ from x_1 to x_2, and $f(\gamma) \in \pi_1(Y', y_0)$.

If our space Y' is homeomorphic to a sphere punctured at n points, then $\pi_1(Y', y_0) \cong \mathbb{F}_{n-1}$, the free group on $n-1$ generators. We can take our generators of $\pi_1(Y', y_0)$ to be loops γ_i based at y_0 going around hole i in a counter-clockwise direction. With this convention we always have

$$\gamma_1 \gamma_2 \dots \gamma_n = e,$$

meaning $[\gamma_1, \dots, \gamma_n]$ is a natural constellation (permuting the singleton set $\{y_0\}$). As $G = \langle f^{-1}(\gamma_i) \rangle$ by definition, this naturally extends to a constellation $[g_1, g_2, \dots, g_n]$ in the monodromy group.

Proposition 2.9. *If $f : X' \to Y'$ is a covering of Y' by X' and G is the monodromy group of the covering, then for some $y_0 \in Y'$ and $x \in E = f^{-1}(y_0)$,*

$$M_x = \{\gamma_i \in \pi_1(Y', y_0) : g_i(x) = f^{-1}(\gamma_i)(x) = x\} \cong \pi_1(X', x).$$

Furthermore, if $f_* : \pi_1(X', x) \to \pi_1(Y', y_0)$ denotes the map between fundamental groups induced by f, then the image of $\pi_1(X', x)$ under f_* is M_x. If a different point $x' \in E$ is chosen, the stabilizer of that point $M_{x'}$ is conjugate to M_x in $\pi_1(Y, y_0)$.

Proof. Consider some loop $\gamma_x \in \pi_1(X', x)$. As $x \in E$, $f(\gamma_x) \in M_x$. Similarly, if $\gamma \in M_x$, then $f^{-1}(\gamma)$ contains a loop based at x by definition, and thus $f^{-1}(\gamma)$ stabilizes x under the action of $\pi_1(Y, y_0)$. Therefore, there is a natural bijection between M_x and $\pi_1(X', x)$. Furthermore, since f_* is just the restriction of the action of $\pi_1(Y, y_0)$ to a subgroup, it is necessarily a group isomorphism. The observation that $f(\gamma_x) \in M_x$ implies that $f_*(\pi_1(X, x)) = \pi_1(Y, y_0)$. The final claim follows from the transitivity of the action of $\pi_1(Y, y_0)$. $\square$

Note that the cosets of the group $M_x \leq G$ are in bijection with E. This is because, if $x_0 \in E$ such that $g_1(x) = f^{-1}(\gamma_1)(x) = x_0$, and $\gamma_1 \gamma_2^{-1} \in M_x$, then $g_2 = \gamma_2$ must also send x to x_0. Furthermore, since the constellation determines the image of a homomorphism of the fundamental group of Y', it also determines (up to conjugacy) the stabilizer of a point, M_x. Therefore, a constellation will uniquely determine, up to conjugacy, a subgroup of $\pi_1(Y', y_0)$ which is isomorphic to $\pi_1(X', x_0)$.

We now show how to topologically encode our combinatorial bicolored plane trees as unique constellations. As constellations uniquely determine a subgroup of $\pi_1(Y', y_0)$ and by the above propositions such a subgroup uniquely determines the covering, this will show that each combinatorial bicolored plane tree uniquely determines a covering of the punctured sphere.

Definition 2.10. *Bicolored graphs drawn on a surface are called* hypermaps*.*

We encode hypermaps as *3-constellations*, that is, constellations containing exactly three permutations, which we will call σ, α and ϕ. Assume all edges are labeled, with the label on the right side of an edge when viewed from a black vertex. The label of an edge is considered incident to a face if it lies inside that face. Our 3-constellation will be a permutation on the labels of the edges. For each black vertex in the plane tree, we add a cycle of all edges incident to that vertex in counter-clockwise order to σ. We call α the analogous permutation of edges around white vertices in the positive direction. We then define $\phi = \alpha^{-1}\sigma^{-1}$. ϕ also has a natural geometric interpretation, as shown in the following proposition.

Proposition 2.11. *ϕ permutes the labels of edges incident to a face in the positive direction.*

Proof. By definition, $\phi = \alpha^{-1}\sigma^{-1}$. First consider the permutation σ^{-1} on a set of labels, illustrated in Figure 12:

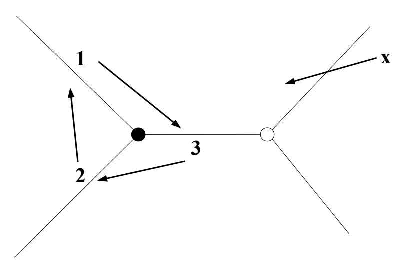


FIGURE 12. The Edge Labels Permuted by σ^{-1}

The labels **2**, **3** and **x** are for illustrative purposes only. Our primary goal is to show that the **1** labeled edge ends up permuted to the next edge on the upper face in the positive direction. This is clear after applying α^{-1} (illustrated in figure 13). Note that all labels return to the faces they were originally on, but at the next edge in the counterclockwise direction.

□

Proposition 2.12. *For a given hypermap H, the Euler characteristic χ of the surface it is drawn on is given by*

$$\chi(H) = c(\sigma) + c(\alpha) + c(\phi) - n$$

where $c(\pi)$ denotes the number of cycles in permutation π and n is the number of edges in H.

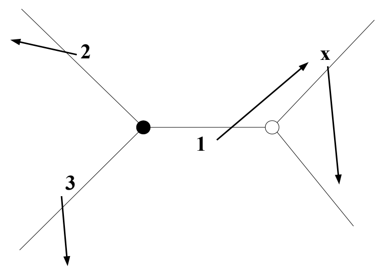


FIGURE 13. The Edge Labels Permuted by α^{-1}

Proof. The number of vertices in the hypermap is $c(\sigma)+c(\alpha)$, and by 2.11 the number of faces is $c(\phi)$. Therefore, the above formula is exactly $V - E + F = \chi(H)$, Euler's formula. □

We now know that a constellation of a monodromy permutation uniquely determines, up to isomorphism, a subgroup of the fundamental group of the target space which is isomorphic to the fundamental group of the covering space. We also know how to obtain a constellation from a combinatorial bicolored plane tree. The following propositions will complete the unique link (up to conjugacy) between a combinatorial bicolored plane tree and a subgroup $M_x \leq \pi_1(Y', y_0)$ which is isomorphic to $\pi_1(X', x_0)$ and which is mapped injectively onto M_x by f.

Proposition 2.13. *The punctures in the sphere X' are in bijection with the number of cycles in the constellation C associated to the tree (when they are written in disjoint cycle notation).*

Proof. First, we show that for each cycle in each $g_i \in C$, we have a distinct puncture in X. Let $g_i \in C$ and γ_i be the associated loop around the hole y_i in Y. Consider some element $x_0 \in E = f^{-1}(y)$. Let $x_1 = g(x_0)$, $x_2 = g(x_1)$, etc. As $|E|$ is finite, for some k, $x_0 = x_k$. In other words, $(x_1, \ldots, x_k)$ is a cycle in $\langle g_i \rangle$. The curve traced by the path from x_i to x_{i+1} by definition maps to the loop γ_i around x_i. For this reason, this is the same curve for each x_i. Therefore, the image (under f) of the loop formed by all such curves is a loop around exactly the hole in y_i. Therefore, this loop in X (corresponding to a cycle in g_i) contains exactly one hole. Since each g_i is associated with a different γ_i around a different puncture in Y, the cycles in distinct elements of C correspond to different punctures in X.

To see why different cycles in each g_i must correspond to different punctures in X, assume that two (disjoint) cycles, c_1 and c_2 in g_i map points $\{x_{1,1}, \ldots, x_{1,n}\}$ and $\{x_{2,1}, \ldots, x_{2,n}\}$ around the same puncture X, respectively. Fill in all other holes in

X, and all holes in Y except for the one associated with γ_i, the loop associated with g_i. Now X and Y are both homeomorphic to a plane. The inverse image of γ_i provides two natural closed loops connecting all of $\{x_{1,1}, \ldots, x_{1,n}\}$ with one loop and all of $\{x_{2,1}, \ldots, x_{2,n}\}$ with a different loop (both in X). Either some $x_{2,j}$ is in the interior of the $\{x_{1,1}, \ldots, x_{1,n}\}$ loop (the part which is now simply connected) or some $x_{1,k}$ is in the interior of the $\{x_{2,1}, \ldots, x_{2,n}\}$ loop. In either case, this point must map to the interior of the image of this loop, which is a contradiction since each $x_{i,j} \in E$, so they all map to the same point $y \in Y$.

Now we show that for each hole in X, there is at least one cycle associated with that hole. Fix a base-point $x_0 \in E$ and a loop in X around a given hole in X. This is a nontrivial element of $\gamma_X \in \pi_1(X, x_0)$ which must map to a nontrivial element $\gamma_Y \in \pi_1(Y, y_0)$. Hence, γ_X is homeomorphic to some loop in the inverse image of γ_Y which goes through some elements in E. The element g_Y in the monodromy group associated with γ_Y therefore has a cycle going around this particular puncture in X. □

Proposition 2.14. *Given a combinatorial bicolored plane tree T, the group $G = \langle \phi, \sigma, \alpha \rangle$ is equal to the monodromy group of the covering $f : X' \to Y'$ determined by T when the point y_0 is chosen to be on the edge in Y'.*

Proof. For the purposes of this proof, let $C = [g_1, g_2, g_3]$ be a constellation such that $G = \langle g_1, g_2, g_3 \rangle$ is the monodromy group, and where each g_i corresponds to a loop γ_i in $\pi_1(Y, y_0)$ around the i^{th} hole in Y'. We will also assume that the lines dividing regions **A** and **B**, which we constructed in our triangulation above, are drawn on both X' and Y'. The constellation we produced from T has the same number of cycles as there are holes in X' by Proposition 2.11. Further by the previous proposition, we know that each cycle in C corresponds to exactly one hole in X'.

Let $y_0 \in Y'$ be a point on the segment in Y'. $E = f^{-1}(y_0)$ contains exactly one point in each edge of T. First, let γ be a loop around either the black or the white hole of Y' (i.e. not around infinity) in the positive direction. Assume without loss of generality that γ is around the black hole in Y'. We can choose γ so that it intersects exactly one of the lines separating the upper and lower half of the plane (the dotted lines separating regions **A** and **B** in the earlier figures). The preimages of γ only intersect a single line separating preimages of **A** and **B** once, so each preimage is a curve between two edges incident to a single black vertex in the positive direction. Since the preimages of γ create loops around the black vertices, γ naturally extends into a permutation of the edges in T in a counterclockwise direction around the black vertices. The $g \in G$ associated with γ is the same as the permutation σ described above. An analogous argument will show that a loop around the white vertex is associated with α.

Since we know the loop around ∞ is the inverse of the products of the other two loops, for C to be a constellation the loop around infinity must be associated with the permutation $\phi = \alpha^{-1}\sigma^{-1}$. The claim follows. $\square$

Note that the assumption that y_0 lies on the edge in Y' is not actually very constraining, since for any $y \in Y'$, this edge is homeomorphic to a non-self intersecting path between the white and black holes of Y' going through y. Therefore, we can conclude that the constellation $[\phi, \sigma, \alpha]$ is the constellation given by the monodromy group of the covering determined by T. By Proposition 2.9, it follows that, up to conjugacy, T determines a subgroup M_x of $\pi_1(Y', y_0)$ which is the image of $\pi_1(X', x)$ under the induced map f_* between fundamental groups. We will now begin an in-depth investigation of covering spaces to show that the subgroup M_x determines a unique covering of Y. We will see that conjugate subgroups determine isomorphic coverings, and thus justifying why we can only know M_x up to conjugacy.

2.1.2. *The Uniqueness of the Covering.* We have shown how to explicitly construct a covering from punctured sphere X to punctured sphere Y so that a given bicolored plane tree is the preimage of a segment between two of the punctures in Y. Our theorem would seem to be proven, except for the single word *unique*. In order to establish the uniqueness of the covering constructed above, we need to further develop the theory of covering spaces, using arguments and propositions from [5].

For the following propositions, let Y be a path connected, locally path connected and semi-locally simply connected topological space, and let (X, p) be a covering space of Y. These assumptions are not needed for all the following propositions, but they are all necessary for Y to have a covering.

Definition 2.15. *A topological space Y is* semi-locally simply connected *if, for all $y \in Y$, there is an open set U containing y such that each loop in U is homotopic to a point in Y.*

Definition 2.16. *Let L be a topological space. A* lift *of a continuous map $f : L \to Y$ is a continuous map $\tilde{f} : L \to X$ such that $p\tilde{f} = f$, that is, a function $\tilde{f}$ such that the following diagram commutes:*

$$\begin{tikzcd} & X \arrow[d, "p"] \\ L \arrow[ur, dashed, "\tilde{f}"] \arrow[r, "f"] & Y \end{tikzcd}$$

The first order of business is to understand lifts in a general sense.

Proposition 2.17. The Homotopy Lifting Property (Proposition 1.30 of [5]) *Let L be a topological space and let $f : L \times [0,1] \to Y$ be a homotopy, and let $\tilde{f} : L \times \{0\} \to X$ be* *a continuous map lifting* $f|L \times \{0\}$*. There exists a unique homotopy* $\tilde{f} : L \times [0,1] \to X$ *such that* $p\tilde{f} = f$*.*

Proof. For the purposes of this proof, let U_α denote an open set in Y containing some point α such that $p^{-1}(U_\alpha)$ can be written as a disjoint union of open sets each homeomorphic to U_α. This is always possible as (X, p) is a covering of Y.

Since (X, p) is a covering space for Y, for all points $(l, t) \in L \times [0,1]$, there is an open neighborhood U_α of $f((l,t))$. By the continuity of f, $f^{-1}(U)$ is an open set in $L \times [0,1]$ containing (l, t). Therefore, there is an open neighborhood $N \times (a, b) \subseteq L \times [0,1]$ of (l, t) such that $f(N \times (a, b)) \subseteq U_\alpha$.

Fix some point $l_0 \in L$. As $\{l_0\} \times [0,1]$ is compact, we can cover $\{l_0\} \times [0,1]$ with finitely many sets of the form $N_t \times (a_t, b_t)$. Pick

$$0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = 1$$

so that for $N_0 = \cap_{t=0}^{n} N_t$, we have $N_0 \subseteq L$, and for some $\alpha \in Y$, $f(N_0 \times [t_i, t_{i+1}]) \subseteq U_{\alpha_i}$. From now on, let U_i denote this particular U_{α_i}. We will construct our lift $\tilde{f}$ in a small neighborhood of l_0 inductively, and we will consistently update N_i so that

$$N_0 \supseteq N_1 \supseteq \cdots \supseteq N_i \ni l_0$$

for all i.

Let $\tilde{f}_i$ denote $\tilde{f}$ restricted to set $N_i \times \{t_i\}$. Note that our base case is given by assumption, as $\tilde{f}_0$ is given. Assume we have lifted f to $\tilde{f}$ uniquely on $N_{i-1} \times [0, t_i]$. We have already shown that $f(N_i \times [t_i, t_{i+1}]) \subseteq U_i$, since $N_i \subseteq N_0$. There is a unique set $\tilde{U}_i \subseteq p^{-1}(U_i)$ which maps homeomorphically onto U_i by p and which contains $\tilde{f}(N_i \times [l_0, t_i])$. Setting $N_{i+1} = N_i \cap \tilde{f}_i^{-1}(\tilde{U}_i)$, we get that $\tilde{f}(N_{i+1} \times \{t_i\}) \subseteq \tilde{U}_i$. We may now define $\tilde{f} : N \times [t_i, t_{i+1}] \to X$ as $\tilde{f} = p^{-1} f$. This choice of definition of $\tilde{f}$ is unique on $\tilde{U}_{i+1}$ since p is a homeomorphism. As we perform only finitely many steps, the final neighborhood N_n of l_0 is open. Therefore, we can lift f uniquely in finitely many steps in an open neighborhood of any point in L. This naturally extends to a lifting of f on all of L. □

A very important application of Proposition 2.17 arises when one applies it to fundamental groups, as in Proposition 2.18.

Proposition 2.18. (Proposition 1.31 of [5]) *The map* $p_* : \pi_1(X, x_0) \to \pi_1(Y, y_0)$ *induced by* p *is injective. The image subgroup of* p_* *consists of homotopy classes of loops based at* y_0 *whose lifts to* X *are loops based at* x_0*.*

Proof. Consider some loop $\gamma \in \ker(p_*)$. By definition, $p(\gamma)$ is homotopic to the constant path at y_0, and by the homotopy lifting property, γ must be homotopic to the constant path at x_0 in X. Therefore, the map p_* is injective. For the second part, consider some loop γ' homotopic to a loop γ in the image of p_*. By the homotopy lifting property, there is a homotopy between the preimage of γ and the preimage of γ', so the map p_* respects homotopy classes. □

A zealous reader will note the similarity between Propositions 2.18 and 2.9. Proposition 2.18 is essentially the same as Proposition 2.9, when one has not defined the monodromy group of the covering. From this point on, for any topological spaces L and K and any continuous map $f : L \to K$, we will use the notation f_* to refer to the map from $\pi_1(L, l_0)$ to $\pi_1(K, k_0)$ induced by f. Now that we understand the power of lifts, a natural question becomes, when can a map $f : L \to Y$ be lifted to a map $\tilde{f} : L \to Y$? Towards that end, we proceed with Proposition 2.19.

Proposition 2.19. The Lifting Criterion (Proposition 1.33 of [5]) *Let L be a path connected and locally path connected topological space and let $f : L \to Y$ be a continuous map. Then a lift $\tilde{f} : L \to X$ exists iff $f_*(\pi_1(L, l_0)) \subseteq p_*(\pi_1(X, x_0))$.*

Proof. ($\Leftarrow$) If f lifts to a map $\tilde{f}$, then by definition $f_* = p_*\tilde{f}_*$. For an arbitrary loop γ in L, $\tilde{f}(\gamma) = \tilde{\gamma}$ is a loop in X, which is then mapped injectively into $\pi_1(Y, y_0)$ by p_*, so the claim follows.

($\Rightarrow$) Now assume $f_*(\pi_1(L, l_0)) \subseteq p_*(\pi_1(X, x_0))$. Let $l \in L$, and let γ be a path from l_0 to l in L. By the homotopy lifting property, the path $f(\gamma)$ lifts uniquely to a path $\tilde{f}(\gamma)$ in X starting at x_0. Define $\tilde{f}(l) = \tilde{f\gamma}(1)$. To verify that $\tilde{f}$ is well defined, consider any path γ' from l_0 to l. $f\gamma'(f\gamma)^{-1}$ is a loop h based at y_0, and by assumption $[h] \in p_*(\pi_1(X, x_0))$. Therefore, there is a loop in X whose image is homotopic to h. By Proposition 2.18, this loop is homotopic (in X), to a loop α with $p(\alpha) = h$. The first half of α is exactly $(\tilde{f\gamma'})^{-1}$, and the second half is $\tilde{f\gamma}$. The common intersection of these two paths means that $\tilde{f}(l)$ does not depend on the path γ to l, and so $\tilde{f}$ is well defined.

We show $\tilde{f}$ is continuous like so. Let $U \subseteq Y$ be an open neighborhood of $f(l)$ lifting to an open set $\tilde{U}$ (containing $\tilde{f}(l)$) such that $p : \tilde{U} \to U$ is a homeomorphism, and let $V \subseteq L$ be an open, path connected neighborhood of l with $f(V) \subseteq U$. For some $l' \in V$, define a path η from l to l' in V. The paths $(f\gamma)(f\eta)$ in Y lift uniquely to X, and since p is a homeomorphism between U and $\tilde{U}$, we get that $\tilde{f}(l') \in \tilde{U}$. As this is true for all $l' \in V$, it follows that $\tilde{f}(V) \subseteq \tilde{U}$, and on V $\tilde{f} = p^{-1}f$. Therefore, for all $l \in L$, f is continuous on an open neighborhood of l. As L is path connected (and therefore connected), we conclude that $\tilde{f}$ is continuous. □

As is a major theme in this section, we must now establish when and how a lift can be considered *unique*. The result of Proposition 2.20 is both surprisingly powerful and extremely versatile.

Proposition 2.20. The Unique Lifting Property (Proposition 1.34 of [5]) *Let L be a connected topological space and let $f : L \to Y$ be a continuous map. If two lifts $\tilde{f}_1$ and $\tilde{f}_2$ of f agree anywhere, they are equivalent.*

Proof. We proceed by showing $A = \{l \in L : \tilde{f}_1(l) = \tilde{f}_2(l)\}$ is a clopen set. The connectedness of L implies that any clopen set must be L or the empty set, so this is certainly sufficient.

Claim 1: $A \subseteq L$ is open.
Let $l \in L$ be given, and let U be an open neighborhood of $f(l)$ in Y so that $p^{-1}(U)$ is a disjoint union of sets mapped homeomorphically onto U by p. Let $\tilde{U}_1 \subseteq p^{-1}(U)$ be the neighborhood containing $\tilde{f}_1(l)$, and let $\tilde{U}_2 \subseteq p^{-1}(U)$ be the neighborhood containing $\tilde{f}_2(l)$. By the continuity of $\tilde{f}_1$ and $\tilde{f}_2$, there is an open set $N \subseteq L$ containing l such that $\tilde{f}_1(N) \subseteq \tilde{U}_1$ and $\tilde{f}_2(N) \subseteq \tilde{U}_2$. As $p : \tilde{U}_1 \to U$ and $p : \tilde{U}_2 \to U$ are homeomorphisms, if $\tilde{U}_1 = \tilde{U}_2$, then $p\tilde{f}_1 = p\tilde{f}_2$ implies $f_1(l) = f_2(l)$. Therefore, $\tilde{f}_1(l) = \tilde{f}_2(l)$ if and only if $\tilde{U}_1 = \tilde{U}_2$. Furthermore, $\tilde{f}_1(l) = \tilde{f}_2(l)$ implies $\tilde{f}_1(N) = \tilde{f}_2(N)$ as $\tilde{f}_1(N) \subseteq \tilde{U}_1$. Hence, $A = \{l \in L : \tilde{f}_1(l) = \tilde{f}_2(l)\}$ is open.

Claim 2: A is closed.
Consider some $l \in A^c = L \setminus A$. By the above arguments, $\tilde{f}_1(l) \neq \tilde{f}_2(l) \Rightarrow \tilde{U}_1 \neq \tilde{U}_2 \Rightarrow \tilde{f}_1(N) \cap \tilde{f}_2(N) = \emptyset$. Therefore, there is an open neighborhood N of l such that $\tilde{f}_1$ and $\tilde{f}_2$ disagree everywhere on N. By unioning over all such l, we conclude that $A^c = \{l \in L : \tilde{f}_1(l) \neq \tilde{f}_2(l)\}$ is open. Combining this with claim 1, we get that A is clopen, so $A = L$ or $A = \emptyset$. □

We now have enough machinery to construct a simply connected covering space of Y, which is often called the *universal cover* of Y. One should note that many authors (as in [4]) define the universal cover in terms of a universal property, and then show that it has the characteristics we describe. For simplicity sake, we will only show its existence and behavior in our specific context.

Theorem 2.21. The Universal Cover ([5]) *There exists a simply-connected, path connected topological space X and a map $p : X \to Y$, such that (X, p) is a covering of Y.*

Proof. We proceed by construction. For any path γ, we write $[\gamma]$ to denote the homotopy class of γ with the endpoints fixed. Fix a base-point $y_0 \in Y$, and define

$$X = \{[\gamma] : \gamma \text{ is a path in } Y \text{ satisfying } \gamma(0) = y_0\}.$$

Define the map $p : X \to Y$ so that

$$f([\gamma]) = \gamma(1).$$

The above map is well defined as each homotopy class of paths has a fixed endpoint. It is onto Y since Y is path connected. We will define a clever topology on X so that f is indeed a bona fide covering. Let

$$\mathcal{U} = \{U \subseteq Y : U \text{ is open, path connected and semi-locally simply connected}\}$$

As Y is locally path connected and semi-locally simply connected, $Y = \cup_{U \in \mathcal{U}} U$. Furthermore, if $U_1, U_2 \in \mathcal{U}$, then there is some path-connected $V \subseteq U_1 \cap U_2$. V is necessarily semi-locally simply connected since U_1 and U_2 are. For each $U \in \mathcal{U}$ and each path γ in Y satisfying $\gamma(1) \in U$, define

$$U_{[\gamma]} = \{\eta \circ \gamma : \eta \text{ is a path in } U \text{ and } \gamma(1) = \eta(0)\}.$$

We define our topology on X so that each $U_{[\gamma]}$ is open. Note that $f : U_{[\gamma]} \to U$ is necessarily a bijection between open sets.

Claim 1: if $[\alpha] \in U_{[\gamma]}$, then $U_{[\gamma]} = U_{[\alpha]}$.

Consider $[\alpha] \in U_{[\gamma]}$. By definition, there is a path η in U from $\gamma(1)$ to $\alpha(1)$. For any $[\gamma'] \in U_{[\gamma]}$, there is a path μ such that $[\gamma' = \mu\gamma]$, meaning $[\mu\gamma] = [\mu\overline{\eta}\alpha]$, so $U_{[\gamma]} \subseteq U_{[\alpha]}$. By a symmetric argument, we may show $U_{[\alpha]} \subseteq U_{[\gamma]}$. The claim follows.

Claim 2: $\{U_{[\gamma]} : \gamma \text{ is a path in } Y \text{ based at } y_0\}$ is a basis for a topology on X.

Let $U_{[\gamma]}$ and $V_{[\alpha]}$ be open sets in X and let $\gamma' \in U_{[\gamma]} \cap V_{[\alpha]}$. Under f, $U_{[\gamma]} \cap V_{[\alpha]}$ must map to an open set $U \cap V$ in Y. There is a semi-locally simply connected open set $W \subseteq U \cap Y$ containing $\alpha(1)$ since $\mathcal{U}$ is a basis for Y. $W_{[\gamma']} \subseteq X$ is a thus an open set in $U_{[\gamma]} \cap V_{[\alpha]}$.

Claim 3: (X, p) is a covering of Y.
$f : X \to Y$ is a local homeomorphism since f is bijective between each basis set U and each one of its preimages. To show f is a covering, let $y \in Y$ and let γ and γ' be paths from y_0 to y. Let U be a path connected and semi-locally simply connected open neighborhood of y. By claim 1, if $[\alpha] \in U_{[\gamma]} \cap U_{[\gamma']}$, then $U_{[\gamma]} = U_{[\alpha]} = U_{[\gamma']}$. Therefore, two preimage sets of U intersect if and only if they are exactly the same set, so we conclude that f is a covering.

Claim 4: X is path connected and simply connected.

For each path γ in Y, define

$$\gamma_t(a) = \begin{cases} a & \text{if} \quad a \leq t \\ t & \text{if} \quad a > t \end{cases}$$

The map $t \mapsto \gamma_t$ lifts the path γ to X, when we fix $x_0 \in f^{-1}(y_0)$. By the above lift, every element in X has a path to x_0, so X is path connected. Furthermore, by definition, any loop in X must therefore satisfy $[\gamma_1] = [\gamma_0] = x_0$, meaning it is homotopic to the constant path γ_0. Hence there are no nontrivial elements of $\pi_1(X, x_0)$. The final claim follows. ☐

Recall that in the start of this section, we explicitly constructed a covering of a punctured sphere and showed that this covering was fundamentally associated with a unique subgroup of $\pi_1(Y, y_0)$. Using the universal cover, we can show that we can construct a cover associated with any subgroup of $\pi_1(Y, y_0)$, and we can rigorously describe the properties it must have.

Proposition 2.22. (Proposition 1.36 of [5]) *For any $M \leq \pi_1(Y, y_0)$ there exists a path-connected covering $(X_M, \tilde{p})$ of Y such that $\tilde{p}_*(\pi_1(X_M, x_0)) = M$ for some $x_0 \in X_H$.*

Proof. By the construction in Theorem 2.21, we know there is a simply connected covering (X, p) of Y. Now let $M \leq \pi_1(Y, y_0)$ be given. We show how to introduce a quotient topology on X so that the fundamental group of the resulting space is isomorphic to M. For two paths γ and γ' based at y_0 (two points in the universal covering X), we say $[\gamma] \sim [\gamma]'$ if $\gamma(1) = \gamma'(1)$ and $\gamma\gamma'^{-1} \in M$. We call the induced map from $X/\sim$ to Y $\tilde{p}$. This is illustrated in the following commutative diagram.

$$\begin{tikzcd} & X \arrow[dl, dashed, "\sim"'] \arrow[dd, "p"] \\ X/\sim \arrow[dr, dashed, "\tilde{p}"] & \\ & Y \end{tikzcd}$$

Claim: $\sim$ is an equivalence relation on $\Gamma = \{[\gamma] : [0,1] \to Y | \gamma(0) = y_0, \gamma(1) = y_1\}$.

Let $[\gamma_1], [\gamma_2]$, and $[\gamma_3]$ be three paths in Γ. Since $\gamma_1\gamma_1^{-1}$ is homeomorphic to the constant path (the identity element of M), $[\gamma_1] \sim [\gamma_1]$. If $[\gamma_1] \sim [\gamma_2]$, then $\gamma_1\gamma_2^{-1} \in M$, so $(\gamma_1\gamma_2^{-1})^{-1} = \gamma_2\gamma_1^{-1} \in M$, meaning $[\gamma_2] \sim [\gamma_1]$. Finally, if $[\gamma_1] \sim [\gamma_2]$ and $[\gamma_2] \sim [\gamma_3]$, then $\gamma_1\gamma_2^{-1}$ and $\gamma_2\gamma_3^{-1}$ are in M. Their product $\gamma_1\gamma_2^{-1}\gamma_2\gamma_3^{-1} = \gamma_1\gamma_3^{-1} \in M$, meaning $[\gamma_1] \sim [\gamma_3]$.

The above claim shows that $\sim$ is an equivalence relation on X, since under $\sim$, paths can only be related if they share endpoints.

Define $X_M = X/\sim$. We will show $X/\sim$ is a covering of Y by showing that if two points $[\gamma]$ and $[\alpha]$ in X are glued together, then the associated sets $U_{[\gamma]}$ and $U_{[\alpha]}$ are also glued together. If $[\gamma] \sim [\gamma']$, then for any path η based at $\gamma(1)$, $\gamma\eta(\gamma'\eta)^{-1} = \gamma\gamma' \in M$. Let $\tilde{p} : X_M \to Y$ be defined as $\tilde{p}([\gamma]) = \gamma(1)$. If $p : X \to Y$ denotes the universal covering map, then we know for any open, path connected and semi-locally simply connected region U, $p^{-1}(U)$ is a disjoint union of open sets homeomorphic to U. By the previous observation, when we take the quotient any two of these open sets will be either mapped homeomorphically onto the same set or to two disjoint sets. Therefore, $\tilde{p}$ inherits the covering property from p, and so $(\tilde{p}, X_M)$ is a covering of Y.

Finally, we will show that $\tilde{p}_*(\pi_1(X_M, x_0)) = M$. Let x_0 be the equivalence class of the constant path in X_M (the image of the base-point of X under $\sim$). Consider some loop $\gamma \in \pi_1(X_M)$. Note that $\gamma(1) = \gamma(0) = x_0$ by definition. This means the path γ and the constant path were identified in our quotient map, so $\gamma \in M$. Therefore, $\pi_1(X_M, x_0) \leq M$. Any element of M will necessarily be identified with the constant path in the quotient map, and the map $\tilde{p} : X/\sim \to Y$ will map each of these loops to the corresponding element of M, so $\tilde{p}_* : \pi_1(X_M, x_0) \to \pi_1(Y, y_0)$ is a group isomorphism. □

Now we need only show the cover obtained above is unique, up to some understanding of equivalence. Recall that if (X_1, p_1) and (X_2, p_2) are two coverings of Y, then a homeomorphism $f : X_1 \to X_2$ is an isomorphism if $p_1 = p_2 f$.

Proposition 2.23. (Proposition 1.37 of [5]) *If Y is path connected and locally path connected then for any two connected covering spaces (X_1, p_1) and (X_2, p_2) of Y, $\pi_1(X_1, x_1) \cong \pi_1(X_2, x_2)$ if and only if there exists an isomorphism $f : X_1 \to X_2$ such that $f(x_1) = x_2$*

Proof. If there is an isomorphism $f : X_1 \to X_2$, then by definition $p_1 = p_2 f$ and $p_2 = p_1 f^{-1}$, so $p_{1*}(\pi_1(X_1, x_1)) = p_{2*}(\pi_1(X_2, x_2))$.

If $p_{1*}(\pi_1(X_1, x_1)) = p_{2*}(\pi_1(X_2, x_2))$, by the lifting criterion $p_1 : X_1 \to Y$ lifts uniquely to a $\tilde{p_1} : X_1 \to X_2$ when we choose $\tilde{p_1}(x_1) = x_2$, and similarly p_2 lifts uniquely to a $\tilde{p_2} : X_2 \to X_1$ with the choice that $\tilde{p_2}(x_2) = x_1$. By definition,

$$p_2 \tilde{p_1} = p_1 \quad \text{and} \quad p_1 \tilde{p_2} = p_2.$$

Combining these statements, we get that $\tilde{p_1}\tilde{p_2} : X_2 \to X_2$ is a map fixing x_2. Furthermore, this is a lift of $\tilde{p_1}$, so by the unique lifting property, it is the identity map on X_2. By symmetry, we see that $\tilde{p_2}\tilde{p_1}$ is the identity map on X_1. Therefore, $\tilde{p_1}$ and $\tilde{p_2}$ are inverse continuous bijections between X_1 and X_2, so $\tilde{p_1}$ is an isomorphism from X_1 to X_2. □

Recall that in the covering constructed at the beginning of this section, the subgroup of $\pi_1(Y, y_0)$ we obtained depended in part on how we chose our base-point in the covering space X. Specifically, different choices of base-points resulted in conjugate subgroups of $\pi_1(Y, y_0)$. Therefore, we cannot simply content ourselves with base-point preserving isomorphisms of covering spaces. Theorem 2.24 states a more general correspondence between subgroups of $\pi_1(Y, y_0)$ and covering spaces. Many have noted that theorem bears a remarkable similarity to the fundamental theorem of Galois theory, a fact which is not lost on [9].

Theorem 2.24. (Theorem 1.3.8 of [5]) *If Y is path connected, locally path connected and semi-locally simply connected then there is a bijection between the set of isomorphism classes of covering spaces (X, p) and conjugacy classes of subgroups of $\pi_1(Y, y_0)$.*

Proof. By Proposition 2.23, we need only show that conjugation of subgroups of $\pi_1(Y, y_0)$ results in isomorphic coverings of Y. Given a covering (X, p) of Y, we will show that changing the base-point inside of $E = f^{-1}(y_0)$ from x_0 to x_1 is equivalent to changing the covering (X, p) to a covering constructed with a conjugate subgroup of $\pi_1(Y, y_0)$. Let x_0 and x_1 be as described, and let γ be a path in X from x_0 to x_1, and let $M_0 = p_*(\pi_1(X, x_0))$ and $M_1 = p_*(\pi_1(X, x_1))$. The image g of γ under f is a loop in $\pi_1(Y, y_0)$. Conjugating M_0 by this loop gives an isomorphic subgroup of $\pi_1(Y, y_0)$. Each loop $g\alpha g^{-1} \in M$ lifts to a loop in X which follows a path from x_1 to x_0, does a loop based at x_0, and then follows the same path back to x_1. This loop is therefore a loop based at x_1, and so $gM_0g^{-1} \subseteq M_1$. By a symmetric argument, we can say $g^{-1}M_1g \subseteq M$, so $M_1 = g^{-1}M_0g$, meaning conjugating the subgroup corresponds exactly to changing the base-point in X. □

An immediate consequence of Theorem 2.24 and Proposition 2.23 is that a universal cover of a space (if it exists) is unique up to isomorphism. Therefore, one may talk about the universal cover of a space Y without ambiguity.

We now have a strong understanding of the correspondence between subgroups of $\pi_1(Y, y_0)$ and coverings of Y. However, one often wants to understand a covering (X, p) of Y in terms of the covering space X instead. This is especially useful when proving that our topological covering can indeed extend into a meromorphic mapping between Riemann surfaces. With this in mind, we will restate the above correspondence in terms of transformations on the covering space X.

Definition 2.25. *Let Y be a topological space and let (X, p) be a covering of Y. A map $f : X \to X$ is* fiber-preserving *if for all $x \in X$, $p(f(x)) = p(x)$.*

Definition 2.26. *Let Y be a topological space and let (X, p) be a covering of Y. A* deck transformation *of this covering is a fiber-preserving homeomorphism $f : X \to X$. The* *set of all such mappings, denoted* $\mathrm{Deck}(X/Y)$ *or* $\mathrm{Deck}(X \xrightarrow{p} Y)$*, is called the group of deck transformations for* Y*.*

Note that, by definition, all deck transformations are isomorphisms of covering spaces. We can illustrate the theory developed thus far with the following commutative diagram:

$$\begin{tikzcd}
X \arrow[rr, leftrightarrow, "\mathrm{Deck}(X/Y)"] \arrow[dr, "\sim"] \arrow[ddr, bend right, "p"'] & & X \arrow[dl, "\sim"'] \arrow[ddl, bend left, "p"] \\
& \tilde{X} \arrow[d, "\tilde{p}"] & \\
& Y &
\end{tikzcd}$$

Theorem 2.27. *Let* (X, p) *be the universal cover of* Y*. Let* $X/\!\sim\, = \tilde{X}$ *be any covering space of* Y*, where* $\sim$ *is the natural equivalence relation defined in Proposition 2.22 for some subgroup* M *of* $\pi_1(Y, y_0)$*. There is an isomorphism* $f : \mathrm{Deck}(X/Y) \to \pi_1(Y, y_0)$ *such that* $f^{-1}(M) = \mathrm{Deck}(X/\tilde{X})$*, where* f *is independent of the choice of* M*.*

Proof. First, we will explicitly construct an isomorphism $f : \mathrm{Deck}(X/Y) \to \pi_1(Y, y_0)$. By Theorem 2.24, we know the universal cover is unique. We can therefore assume the space X is as we constructed it in Theorem 2.21. Under this assumption, each element of X is a homotopy class of paths in Y based at y_0. Let $x_0 \in p^{-1}(y_0)$ be given, and let $\sigma \in \mathrm{Deck}(X/Y)$. Define

$$f(\sigma) = \sigma(x_0).$$

By the definition of X, $\sigma(x_0) = [\gamma]$ for some path in Y which starts at $p(x_0)$ and ends at $p(\sigma(x_0))$. Since σ is fiber-preserving, it follows that $\sigma(x_0) = [\gamma]$ is the homotopy class of a loop in Y based at y_0, which by definition is an element of $\pi_1(Y, y_0)$. To show f is one to one, consider some other element α of $\mathrm{Deck}(X/Y)$ for which $f(\alpha) = f(\sigma)$. σ and α both lift the map $p : X \to Y$ to X, so since X is connected and σ and α agree on x_0, by the unique lifting property, $\sigma = \alpha$ on all of X.

To show f is onto, consider some element $[\gamma]$ of $\pi_1(Y, y_0)$. We will show that there is an element $\sigma \in \mathrm{Deck}(X/Y)$ such that $\sigma([y_0]) = [\gamma]$. As Y is path connected, for each $y \in Y$, there is a path η from y_0 to y and loop α based at y such that $[\gamma] = [\eta^{-1}\alpha\eta]$. For each x in the fiber of y, let g_x be a function lifting α to X, where $\alpha(0) = x$. Define $\sigma(x) = g(\alpha(1))$. This is well-defined by the unique lifting property, and it is bijective since it has a well-defined inverse (which can be realized by lifting $\overline{\alpha}$). As every $x \in X$ is the preimage of some point in Y, we know σ is a fiber-preserving bijection on X.

We will now show σ constructed above is a homeomorphism, and thus that σ is indeed a deck transformation. Fix some y_1 in Y, and let U be a simply connected open neighborhood of y_1 such that $p^{-1}(U)$ can be expressed as a disjoint union of sets each homeomorphic to U by p. Let $V \subseteq X$ be one such set, and let $v \in V \cap p^{-1}(y_1)$. By definition of X, $v_1 = [\eta]$, where η is a path from y_0 to y_1. Let α be a loop based at y_1 such that $[\eta^{-1}\alpha\eta] = [\gamma]$, and let g be a lift of α to X such that $g(\alpha(0)) = v_1$. For any $v \in V$, $v = [\zeta\eta]$ for some path ζ in U. $\eta^{-1}(\zeta\alpha\zeta^{-1})\eta$ is a loop based at y_0 homotopic to γ, and $(p|_V)^{-1}(\zeta)$ is a lift of ζ starting at v_1. If $p'^{-1}(\zeta)$ is defined as the lift of ζ based at $\sigma(y_1)$, we get $\sigma(y) = p'^{-1}(\zeta)g(\alpha)(p|_V)^{-1}(\zeta)^{-1}(1) = p'^{-1}(\zeta)(g(\alpha)(1)) = p'^{-1}(\zeta)\sigma(y_1)$. Therefore, if $V' \subseteq p^{-1}(U)$ is the open set in X containing $\sigma(y_1)$ which is homeomorphic to U by p, then $\sigma(y) \in V'$. We conclude that σ is an open map. The analogous argument with $\overline{\gamma}$ shows that σ^{-1} is an open map, so we may conclude that σ is a homeomorphism and thus is a deck transformation. Therefore, f is onto and therefore an isomorphism between $\mathrm{Deck}(X/Y)$ and $\pi_1(Y, y_0)$.

To show the final part of the theorem, let $M \subseteq \pi_1(Y, y_0)$ and let $\tilde{X}$ the covering determined by M. By theorem 2.24, we may assume without loss of generality that $\tilde{X} = X/\sim$, where $\sim$ is the equivalence relation defined in the proof of Proposition 2.22. Observe that if $[\gamma] \in M$ and $\sigma = f^{-1}([\gamma])$, then $\sigma([y_0]) = [\gamma]$. As σ is a lift of the covering map p, by the unique lifting property there are no other deck transformations which send $[y_0]$ to $[\gamma]$. In particular, this means that a deck transformation $\sigma \in \mathrm{Deck}(X/Y)$ is in $\mathrm{Deck}(X/\tilde{X})$ if and only if it maps $x_0 = [y_0]$ to elements of $\pi_1(X, x_0)$. By the definition of $\sim$, these are precisely the deck transformations which map $[y_0]$ to loops in M, which is $f^{-1}(M)$, as desired. $\square$

We have shown that, given a combinatorial bicolored plane tree T, we can construct a covering from a sphere punctured at the vertices of T (and infinity) to a sphere punctured at three points, so that the edges in T all map to a single edge between two of the holes in the image. The tree T can be encoded as a constellation, which is equivalent to the constellation of the monodromy group of the covering. The stabilizer of a point in the monodromy group is a subgroup of the fundamental group of the target space which is uniquely determined by T. By Theorem 2.24, a subgroup M of the fundamental group determines a covering space $(\tilde{X}, \tilde{p})$, which is unique up conjugacy of the subgroup M. This subgroup M is also isomorphic to the group of deck transformations $\mathrm{Deck}(X/\tilde{X})$ for the target space Y. In conclusion, a combinatorial bicolored plane tree uniquely determines a covering of a punctured sphere by a punctured sphere.

2.2. **The Analytic Map.** To complete our result, We first establish the necessary analysis machinery, which mainly includes basics of Riemann surfaces and holomorphic mappings. We assume the reader is familiar with the definition of Riemann surface and holomorphic mapping, as in [4]. The results in this section come from sections 1.2, 1.4, and 1.5 of [4].

Theorem 2.28. *(Identity theorem) Let X, Y be Riemann surfaces, f_1, f_2 two holomorphic mappings from X to Y which coincide on a set $A \subseteq X$ having a limit point $a \in X$. Then f_1, f_2 are identically equal.*

Proof. Let $M = \{x \in X : x \text{ has an open neighborhood } W \text{ such that } f_1|W = f_2|W\}$. By definition M is open. We show that M is also closed. Since let b be a boundary point of M. Since f_1 and f_2 are continuous, we have $f_1(b) = f_2(b)$. Let $\phi : U \to V$ be a chart on X and $\psi : U' \to V'$ be chart on Y such that $f_i(U) \subseteq U'$ and $b \in U$. Assume also that U is connected. Then $g_i := \psi \circ f_i \circ \phi^{-1} : V \to V' \subseteq \mathbb{C}$ are holomorphic. Since b is a boundary point of M, we have $U \cap G \neq \emptyset$, and thus by the Identity Theorem for holomorphic functions on domains in $\mathbb{C}$ we have g_1 and g_2 are identically equal. Thus $f_1|U = f_2|U$. Hence $b \in M$ and thus M is closed. Since X is connected, $M = \emptyset$ or $M = X$. The first case is impossible since $a \in M$. So we must have $M = X$. □

Theorem 2.29. *Suppose X and Y are Riemann surfaces and $f : X \to Y$ is a non-constant holomorphic mapping. Suppose $a \in X$ and $b := f(a)$. Then there exists an integer $k \geq 1$ and charts $\phi : U \to V$ on X and $\psi : U' \to V'$ on Y with the following properties:*

(1) *$a \in U$, $\phi(a) = 0$; $b \in U'$, $\psi(b) = 0$.*
(2) *$f(U) \subseteq U'$.*
(3) *The map $F := \psi \circ f \circ \phi^{-1} : V \to V'$ is given by $F(z) = z^k$ for all $z \in V$.*

Proof. First note that there exist charts $\phi_1 : U_1 \to V_1$ and $\psi : U' \to V'$ such that (1) and (2) are satisfied. The function $f_1 := \psi \circ f \circ \phi_1^{-1}$ is non constant. Since if it is, by ϕ and ψ are homeomorphisms, we have f is constant on the U_1, and applying the identity theorem we get f is constant everywhere, contradicting the assumption.

Since $f_1(0) = 0$, there is a $k \geq 1$, such that $f_1(z) = z^k g(z)$, where g is holomorphic on V_1, and $g(0) \neq 0$. Therefore there exists a neighborhood V_2 of 0 and a holomorphic function h on this neighborhood such that $h^k = g$. Let $\alpha : V_2 \to V$ be a map onto an open neighborhood V such that $\alpha : z \mapsto zh(z)$. This is a biholomorphic map. Let $U := \phi_1^{-1}(V_2)$ and $\phi := \alpha \circ (\phi_1|_U)$. Replace (U_1, ϕ_1) by (U, V), and by construction $F = \psi \circ f \circ \phi^{-1}$ satisfies $F(z) = z^k$. □

From this theorem it follows directly

Corollary 2.30. *Suppose X, Y are Riemann surfaces and $f : X \to Y$ is a non-constant, holomorphic mapping. Then f is open.*

Theorem 2.31. *Suppose X and Y are Riemann surfaces, X is compact, and $f : X \to Y$ is a non-constant holomorphic mapping. Then Y is compact and f is surjective.*

Proof. $f(X)$ is open by 2.30. Since X is compact, $f(X)$ is compact, and thus closed. The only subsets which are both open and closed of a connected topological space are the empty set and the whole space, so $f(X) = Y$. □

Theorem 2.32. *Every holomorphic function on a compact Riemann surface is constant. (Recall that a holomorphic function on a Riemann surface X is a holomorphic mapping from X to $\mathbb{C}$.)*

Proof. $\mathbb{C}$ is not compact. □

We will need the following theorem to conclude the map determined by a bicolored plane tree is a polynomial.

Theorem 2.33. *Every meromorphic function f on $\mathbb{P}^1$ is rational.*

Proof. f can only have finitely many poles, since if it had infinitely many poles, they must have a limit point on $\mathbb{P}^1$. By the identity theorem f would equal ∞ everywhere.

We may assume that ∞ is not a pole, since if it is we could consider $1/f$ instead. Suppose $a_1, \cdots, a_n$ are poles of f, and let $h_v(z) = \sum_{j=-k_v}^{-1} c_{vj}(z-a_v)^j$ be the principal part of f at the pole a_v. Then $f - \sum_{j=1}^{n} h_v$ is holomorphic, and thus is a constant.

Notice that if f only has a single pole at ∞, $\frac{1}{f(z)} = \frac{c_n}{z^n} + \frac{c_{n-1}}{z^{n-1}} + \cdots + \frac{c_1}{z} + C$ for some constants $c_1, \cdots, c_n, C$. It follows that f is a polynomial.

□

The following theorem is needed for determining a complex structure of the covering space determined by a tree.

Theorem 2.34. *Let X be a Riemann surface, Y a Hausdorff topological space, $p : Y \to X$ a local homeomorphism. Then there is a unique complex structure on Y such that p is holomorphic.*

Proof. Suppose $\phi_1 : U_1 \to V \subset \mathbb{C}$ is a complex chart on X such that there exists an open set $U \subset Y$ with $p|U \to U_1$ a homeomorphism. Then $\phi = \phi_1 \circ p : U \to V$ is a complex chart on Y. Let $\mathcal{U}$ be the set of all complex charts obtained this way. Charts of $\mathcal{U}$ cover Y, since for any $y \in Y$, there is an open neighborhood V_y around y such that p maps V_y homeomorphically onto its image, which is an open neighborhood $V_{p(y)}$ of $p(y)$. We can choose V_y small enough so that $V_{p(y)}$ is contained in the domain of a chart. It is also easy to check that the charts are holomorphically compatible with each other. Let Y have the complex structure defined by $\mathcal{U}$. Then p is locally biholomorphic and hence holomorphic.

To show that $\mathcal{U}$ is unique, suppose there is another complex atlas on Y, $\mathcal{U}'$ such that $p : (Y,\mathcal{U}') \to X$ is holomorphic and thus locally biholomorphic. Then the identity mapping $(Y,\mathcal{U}) \to (Y,\mathcal{U})$ is locally biholomorphic. (For any point $y \in Y$, there is an open neighborhood V_y of y such that $\psi \circ p \circ \phi_1^{-1}$ from $\phi_1(V_y)$ onto its image is biholomorphic, where $\psi : U \to V$ is a chart on X containing $p(V_y)$ and ϕ_1 is a chart of $\mathcal{U}$. Similarly, there is a chart ϕ_2 of $\mathcal{U}'$ such that $\psi \circ p \circ \phi_2^{-1}$ on $\phi_2(V_y)$ is biholomorphic.

Thus $\phi_2 \circ p^{-1} \circ \psi^{-1} \circ \psi \circ p \circ \phi^{-1} = \phi_2 \circ \phi_1^{-1} = \phi_2 \circ \mathrm{id} \circ \phi_1^{-1}$ is biholomorphic.) It then follows that $\mathcal{U}$ and $\mathcal{U}'$ define the same complex structure on Y. □

Theorem 2.34 states that the covering determined by a combinatorial bicolored plane tree T as described in section 2.1 can be made holomorphic outside of the vertices of T and infinity. In order to extend this map, we must prove the following theorems.

Theorem 2.35. *Let X be a Riemann surface, $D^* = \{z \in \mathbb{C} : 0 < |z| < 1\}$ and let $f : X \to D^*$ be an unbranched holomorphic covering map. Then one of the following holds:*

(1) *If the covering has an infinite number of sheets, then there exists a biholomorphic mapping $\phi : X \to H$ of X onto the left half plane such that the following diagram commutes.*

$$\begin{tikzcd} X \arrow[rr, dashed, "\phi"] \arrow[dr, "f"'] & & H \arrow[dl, "exp"] \\ & D^* & \end{tikzcd}$$

(2) *If the covering is k-sheeted ($k < \infty$), then there exists a biholomorphic mapping $\phi : X \to D^*$ such that the following diagram commutes.*

$$\begin{tikzcd} X \arrow[rr, dashed, "\phi"] \arrow[dr, "f"'] & & D^* \arrow[dl, "p_k"] \\ & D^* & \end{tikzcd}$$

Proof. $\exp : H \to D^*$ is a universal covering, so there exists a holomorphic mapping $\psi : H \to X$ such that $\exp = f \circ \psi$. $\mathrm{Deck}(H/D^*) = \{\tau_n : n \in \mathbb{Z}\}$ where $\tau_n : H \to H$ denotes a translation $z \mapsto z + 2\pi i n$. Let $G \subset \mathrm{Deck}(H/D)$ be the subgroup such that $\psi(h) = \psi(h')$ if and only if there exists $\sigma \in G$ such that $\sigma(h) = h'$.

When $G = \{e\}$, $\psi(h) = \psi(h')$ if and only if $h = h'$. Therefore ψ is an injective holomorphic map, which means it is biholomorphic onto its image (cor 2.5 of [4]). The inverse map of ψ, $\phi : X \to H$, is what we are looking for in part (i).

For part (ii), we know that when $G \subseteq \mathrm{Deck}(H/D^*)$ is not the identity, there exists a natural number $k \geq 1$ so that $G = \{\tau_{nk} : n \in \mathbb{Z}\}$. Let $g : H \to D^*$ be the covering map defined by $g(z) = \exp(z/k)$. Then $g(z) = g(z')$ if and only if $z = z + 2mk\pi i$ for some $m \in \mathbb{Z}$. This is equivalent as saying there exists $\sigma \in G$ such that $\sigma(z) = z'$ (z and z' are equivalent modulo G). Since ψ also maps points that are equivalent modulo G to the same point in X, we can construct a bijective mapping $\phi : X \to D^*$, such that $\phi(x) = y$ if and only if $\psi^{-1}(x) = g^{-1}(y)$. Thus ϕ is a lift of g with respect to ψ. Since ψ

and g are locally biholomorphic, ϕ is locally biholomorphic. Now $\exp = f \circ \psi = p_k \circ g$ and the top diagram commutes, the bottom also does. □

Theorem 2.36. *Suppose X is a Riemann surface, $A \subset X$ is a closed discrete subset and let $X' = X \backslash A$. Suppose Y' is another Riemann surface and $\pi' : Y' \to X'$ is a finite-sheeted unbranched holomorphic covering. Then π' extends to a branched covering of X. In other words, there exists a Riemann surface Y, a holomorphic mapping $\pi : Y \to X$ and a fiber-preserving biholomorphic mapping $\pi : Y \setminus \pi^{-1}(A) \to Y'$.*

Proof. For every $a \in A$ we choose a coordinate neighborhood (U_a, z_a) on X such that $z_a(a) = 0$, $z_a(U_a)$ is the unit disk in $\mathbb{C}$ and $U_a \cap U_{a'}$ if $a \neq a'$. Let $U_a^* = U_a \setminus \{a\}$. Since $\pi' : Y' \to X'$ is finite sheeted, $\pi'^{-1}(U_a^*)$ consists of a finite number $n(a)$ of connected components V_{av}^*, $v = 1, \cdots, n(a)$. For every v the mapping $\pi'|V_{av}^* \to U_a^*$ is an unbranched covering. Let its covering number be k_{av}. By theorem [above] there exist biholomorphic mappings $\zeta_{av} : V_{av}^* \to D^*$ of V_{av}^* onto the punctured disk such that the diagram commutes, where $\pi_{av}(z) = z^{k_{av}}$.

We choose distinct points p_{av}, $a \in A$, $v = 1, \cdots, n(a)$ disjoint from Y', and let $Y := Y' \cup \{p_{av} : a \in A, v = 1 \cdots, n(a)\}$. We put a topology on Y as follows. For a basis set in Y' we make it a basis set of Y. If $W_i, i \in I$ is a neighborhood basis of a, then let $\{p_{av}\} \cup (\pi'^{-1}(W_i) \cap V_{av}^*)$ be a neighborhood basis of p_{av}. It is easy to check that this is indeed a topology on Y', and that it is Hausdorff. Define $\pi : Y \to X$ by $\pi(y) = \pi'(y)$ for $y \in Y'$ and $\pi(p_{av}) = a$.

To make Y into a Riemann surface, we add to the charts of complex structure of Y' the following charts. Let $V_{av} = V_{av}^* \cup \{p_{av}\}$ and let $\zeta_{av} : V_{av} \to D$ be the same as $\zeta_{av} : V_{av}^* \to D^*$ on V_{av}^* and $\zeta_{av}(p_{av}) = 0$. Since $\zeta_{av} : V_{av}^* \to D^*$ is biholomorphic with respect to complex structure on Y', the new charts $\zeta_{av} : V_{av} \to D$ is holomorphically compatible with the the charts of the complex structure of Y'. Then $\pi : Y \to X$ is holomorphic. We choose $\phi : Y \setminus \pi^{-1} \to Y'$ to be the identity mapping. This shows the existence of a continuation of the covering π'. □

We have now essentially established the theory we need to complete the proof of the bijection theorem (Theorem 2.4).

Proof of Theorem 2.3. By Propositions 2.9, 2.14, and Theorem 2.24, we know that a combinatorial bicolored plane tree T uniquely determines a covering map f (up to covering isomorphism) from the sphere punctured at the vertices of T and infinity (call in X') to the sphere punctured at three points (call it Y'). By Theorem 2.34, the complex structure of X' can be uniquely constructed (up to automorphism of the image sphere) by lifting the complex structure of the punctured Y', where we take the canonical coordinates on Y'. Then by Theorem 2.36, we can fill in holes of Y' and X' to uniquely extend the map to a meromorphic mapping. Since this is a map from a Riemann sphere to a Riemann sphere, it is a rational function by Theorem 2.33. By the construction of the map we see the function has a single pole at infinity, and therefore it has to be a polynomial.

The above argument glosses over the fact that we only know the covering of Y' up to homeomorphism. To see why this is sufficient, consider some covering $(\tilde{X}', \tilde{f})$ of Y' which is isomorphic (as a covering) to (X', f). If these are isomorphic covers, there is a homeomorphic map $g : \tilde{X}' \to X'$ such that $f \circ g = \tilde{f}$. Thus, $(\tilde{X}', g)$ is a covering of X', so we can uniquely lift the complex structure from X' to $\tilde{X}'$ by another application of Theorem 2.34. Thus, the covering of Y' is unique up by automorphisms of both the image sphere and the preimage sphere which fix the point at infinity. These are exactly the linear transformations $z \mapsto az+b$. Since Shabat polynomials are equivalent up to pre- and post- composition of linear maps, we conclude that each combinatorial bicolored plane tree T uniquely determines (up to equivalence) a Shabat polynomial which has that tree as the preimage of a segment between its critical values.

□

Proof of Theorem 2.4. By Proposition 1.4, each Shabat polynomial produces a bicolored plane tree as the preimage of a segment. By Theorem 2.3, each combinatorial bicolored plane tree is so produced by some Shabat polynomial, and further equivalent polynomials produce equivalent trees, and inequivalent polynomials produce inequivalent trees. The desired bijection follows. □

3. Constructing Shabat Polynomials from Trees

3.1. **Methods for Constructing Shabat Polynomials.** The correspondence between the equivalence classes of Shabat polynomials (polynomials that differ only by affine linear transformation of the variable and the entire polynomial itself) and equivalence classes of Bicolored Plane Trees allows us to exploit certain combinatorial and geometric facts about plane trees in order to construct actual Shabat polynomials. The general method for finding Shabat polynomials based on their trees goes as follows:

We begin by defining the type of the tree:

(1) We begin first by deciding the combinatorial nature of the plane tree which will correspond to our Shabat polynomial.

    Definition 3.1. *Given a bicolored plane tree T of N edges, we let $\langle B, W\rangle$ be a pair of partitions of N such that B corresponds to the list of valencies of the black vertices and define W similarly for the white vertices. We say that $\langle B, W\rangle$ is the type of T.*

    Choosing our type will allow us in a sense to narrow down which equivalence class the resulting Shabat polynomial will come from. Although different classes of bicolored plane trees may share the same type, in many cases we may narrow down the the corresponding equivalence class of Shabat polynomials to just one.

(2) We can normalize any Shabat polynomial to have the black vertices map to 0 and have the white vertices map to 1. Additionally we may take certain liberties in adjusting the geometric positioning of its corresponding plane tree in order to determine algebraic facts about our Shabat polynomial. So given a bicolored plane tree with black vertices a_1, a_2, ..., a_n and corresponding valencies α_1, α_2, ..., α_n, without loss of generality our Shabat polynomial will take on the form $P(z) = (z - a_1)^{\alpha_1}(z - a_2)^{\alpha_2}...(z - a_n)^{\alpha_n}$

(3) The derivative of our polynomial P will be of the form $P(z) = (z - a_1)^{\alpha_1 - 1}(z - a_2)^{\alpha_2 - 1}...(z - a_n)^{\alpha_n - 1}W(z)$, where $W(z)$ will have all the non-leaf white vertices as its only roots. Furthermore, the multiplicity of each root of $W(z)$ will be one less than its actual valency in T. Using certain algebraic facts about the polynomial W we may in some cases determine all of the roots of the original Shabat polynomial, or in other cases determines the roots up to equations describing them as the roots of some other polynomial.

3.2. **Canonical Geometric Form.** Any two BPT of the same equivalence class will be the same up to similarity. This fact follows from the equivalence relation on Shabat polynomials, and means that each equivalence class of BPT will have a canonical geometric Form. We begin by proving this basic theorem.

Theorem 3.2. *Let P and Q be equivalent Shabat polynomials, and let T_P and T_Q be their respective plane trees. Then T_P and T_Q are the same up to similarity (i.e. scaling and isometry).*

Proof. Let Q be equivalent to P, so for some $A, B, a, b \in \mathbb{C}$, $AP(az + b) + B = Q(z)$, and assume without loss of generality that P has critical points $0, 1$. We will first show that linear transformation of the polynomial from the outside will transform the segment $[0, 1]$ by similarity while fixing T_P. We will show this by dividing the proof into two cases that together will show that equivalent polynomials have the same image segment and their BPTs will be the same up to similarity.

First let $AP(z) + B = Q(z)$, and let r be a critical point of P, and hence a vertex of T_P. Therefore, $P'(r) = 0 \implies Q'(r) = AP'(r) = 0$, so r is a critical point of Q, and hence a vertex of T_Q. Now let E be an edge of T_P, and note that the corresponding critical value segment for $Q(z)$ will be $[0, A + B] = A[0, 1] + B$. Therefore $P(E) = [0, 1] \implies Q(E) = AP(E) + B = [0, A + B]$, so E is also an edge in T_Q. So we have that the linear transformation of the polynomial moves the critical value segment under linear transformation (i.e. similarity) and fixes T_P.

Next we will show that the linear transformation of the variable will fix the critical value segment while transforming the tree T_P up to similarity, so let $P(az + b) = Q(z)$.

The critical value segment will be the same for both P and Q. Let r be a critical point of P, then $(r-b)/a$ will be the corresponding critical point of Q. Furthermore, if E is an edge of T_P then $(E-b)/a$ will be an edge of Q since $Q((E-b)/a) = P(a((E-b)/a)+b) = P(E)$. Therefore $T_Q = (T_P - b)/a$, so the trees are equivalent up to similarity.

Since the linear transformation of the variable and polynomial are independent of each other, for any equivalent P and Q we can find a two suitable linear transformations to turn P into Q, the difference between the respective trees T_P and T_Q will be the same up to the similarity of the image segment and the critical value. □

As a result of this proof, when we attempt to find Shabat polynomials we may assume certain things about the position, size, and symmetry of the corresponding tree that will determine the Shabat polynomial. Also, this means that for every equivalence class of BPTs there is a canonical geometric form, that is each equvalence class of Shabat polynomials will correspond to a plane tree up to not only planar graph equivalence but also up to similarity.

3.3. **Using Geometry to Determine Shabat Polynomials.** For the rest of this paper we will assume that the critical values of any Shabat Polynomial, unless specified otherwise, will be $0, 1$. Notice that as a result of 3.2 we may assume that a Shabat polynomial has corresponding critical values $0, 1$, as we need only find a suitable affine linear transformation to make the critical value segment $[0, 1]$. Now we will give some useful definitions.

Definition 3.3. *Let N be the number of edges adjacent to a vertex v, then we say that N is the valency of v*

Definition 3.4. *Let v be a vertex of some bicolored plane tree T such that no other vertex of the same color has the same valency, then v is a bachelor of T.*

By theorem 3.2 we can narrow down our search for a corresponding Shabat polynomial by fixing any two bachelors. The reason for this fact is that for any equivalence class of Shabat polynomials there will be either one or finitely many equivalent polynomials with a corresponding tree T having those bachelor vertices in those places.

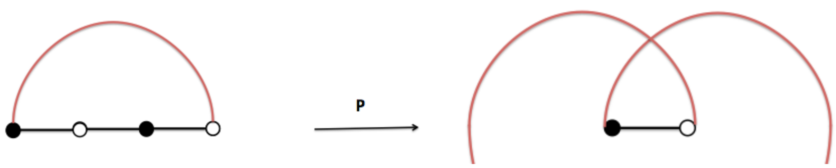
Additionally, we may even consider symmetries of Shabat Polynomials in a sense:

Definition 3.5. *A canonical geometric form is said to be* symmetric *if there exists a Shabat Polynomial P in its corresponding polynomial equivalence class such that P has only real coefficients.*

Theorem 3.6. *If P has a unique type and at least two vertices of the same color but different valency then $\overline{P}$ is symmetric.*

Proof. We begin this proof by first stating a Lemma about complex conjugation.

FIGURE 14. Example of path with self-intersecting image



Lemma 3.7. *If P is a Shabat polynomial, then its complex conjugate $\overline{P}$ will also be a Shabat polynomial.*

Proof. Let $P(z) = a_n z^n + ... + a_1 z + a_0$ be a Shabat Polynomial. $\overline{P}(z) = \overline{a_n} z^n + ... + \overline{a_1} z + \overline{a_0}$, so it follows that $\overline{P}'(z) = n\overline{a_n} z^{n-1} + ... + 2\overline{a_2} z + \overline{a_1}$. If $P'(z_0) = na_n z_0^{n-1} + ... + 2a_2 z + a_1 = 0$, then it follows that $\overline{P}'(\overline{z_0}) = n\overline{a_n}\overline{z_0}^{n-1} + ... + 2\overline{a_2}\overline{z} + \overline{a_1} = \overline{na_n z_0^{n-1} + ... + 2a_2 z + a_1} = \overline{0} = 0$, so $\overline{\text{CritP}(P)} = \text{CritP}(\overline{P})$, where $\text{CritP}(P)$ are the critical points of P. Furthermore, $\text{CritV}(\overline{P}) = \overline{P}(\text{CritP}(\overline{P})) = \overline{P}(\overline{\text{CritP}(P)}) = \overline{\text{CritV}(P)} = \overline{\{0,1\}} = \{0,1\}$, so $\overline{P}$ is also Shabat. □

Let P be a Shabat polynomial that has 2 same-colored vertices of different valency fixed at 0 and 1 with a unique type, so P and $\overline{P}$ will be equivalent, thus for some $A, B, a, b \in \mathbb{C}$, $AP(az+b) + B = \overline{P}(z)$. Since $CV(P) = CV(\overline{P})$, P and $\overline{P}$ will have the same critical points, and since the critical points of both P and $\overline{P}$ will map to zero, so $A = 1$ and $B = 0$. So $P(az+b) = \overline{P}$. If $\overline{P}'(z_0) = aP'(az_0 + b) = 0$, then $z_0 \in Crit(P(az+b)) = Crit(az+b) \cup (Crit(P) - b)/a = (Crit(P) - b)/a$, so $Crit(\overline{P}) = \overline{Crit(P)} = Crit(P) - b)/a$. $(z-b)/a$ is a linear transformation, but this linear transformation must fix 0 and 1 since they are real numbers (and thus self conjugates). Therefore, $(z-b)/a$ is trivial, so $b = 0$ and $a = 1$. Hence $P = \overline{P}$, so its coefficients must be real. By definition then P will be symmetric. □

Even if for an arbitrary BPT you find its Shabat polynomial, and the coordinates of both its black and white vertices, there is still a chance that you may not know whether two differently colored vertices are neighbors or not. One way to test this is to construct a path between two differently colored vertices, and look at the image of the path under the polynomial. It the image of the path intersects itself while the path itself is bijective (i.e. non-intersecting), then the two vertices cannot be adjacent.

Theorem 3.8. *Let v_0 and v_1 be differently colored vertices (black and white respectively) of the same tree T_P that are not adjacent. If $p(t)$ is a non-self-intersecting path connecting them (with no other vertices incident in it), then $P(p(t))$ is self-intersecting.*

Proof. Let $p : [0,1] \to \mathbb{C}$ be a continuous non-iself-ntersecting path from v_0 to v_1. v_0 and v_1 are non-adjacent but still part of the same BPT, so there is a unique graph path $v_0 x_0 ... x_n v_1$ (not to be confused with p) that connects them where x_0 will be white and

x_n will be black. Draw two edges, $[x_0, \infty]$ and $[x_n, \infty]$, such that both edges intersect with $p(t)$.

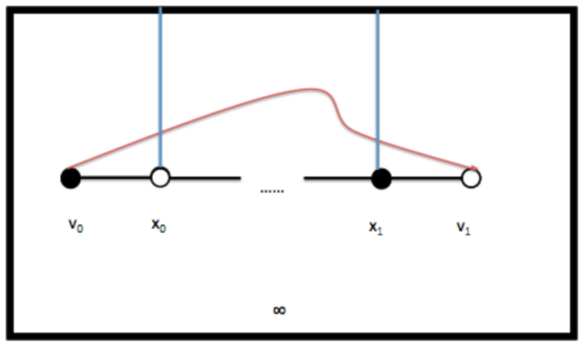


FIGURE 15. Preimage of path and edges

The image of $[x_0, \infty]$ and $[x_n, \infty]$ will connect with the segment $[0, 1]$ and will cut the w-plane into two halves in a way topologically equivalent to cutting the w-plane along the real line. Moving from v_0, p will intersect $[x_n, \infty]$, meaning that $P(p)$ will move from $P(v_0) = 0$ until it intersects a point on $[1, \infty]$, isolating 1 from $[x_0, \infty]$ on the first half-plane of the w-plane with $P(p(t))$ moving from 0 to $[x_n, \infty]$. When the path $P(p(t))$ returns to the first half-plane for the last time it will first intersect from $[x_0, \infty]$, but $[x_0, \infty]$ is isolated from $1 = P(v_1)$ on the first half-plane by the first segment of $P(p(t))$. Therefore $P(p(t))$ must be self-intersecting. □

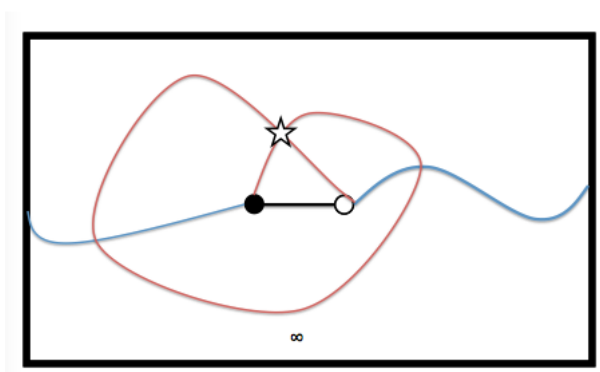


FIGURE 16. Image with path self-intersection

3.4. **Composition of Shabat Polynomials.** We know that the composition of two Shabat polynomials P, and Q with critical values $\{p_1, p_2\}$ and $\{q_1, q_2\}$ respectively will be Shabat if $P(q_1), P(q_2) \in \{p_1, p_2\}$

Theorem 3.9. *Let P and Q be Shabat Polynomials with critical values $\{p_1, p_2\}$ and $\{q_1, q_2\}$ respectively such that $P(q_1), P(q_2) \in \{p_1, p_2\}$, then $P \circ Q(z)$ is Shabat. Furthermore the tree corresponding to $P \circ Q$ is obtained by replacing the edges of the tree of Q by the tree of P in the following way which we outline in figures 17 and 18.*

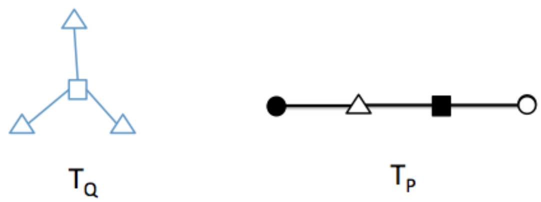


FIGURE 17. Trees of P and Q

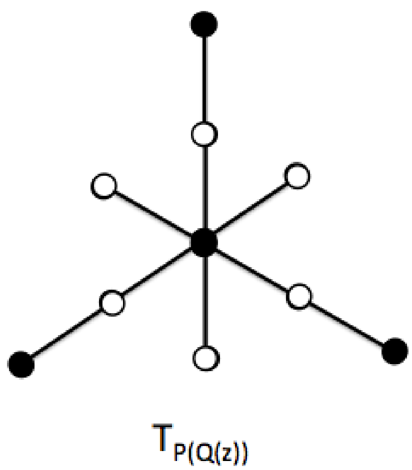


FIGURE 18. Resulting Tree of $P \circ Q$

Proof. Let CritV(P) be the critical values of P, and let CritP(P) be the critical points of P. Notice that the CritV$(P \circ Q) = P \circ Q(\text{CritP}(P \circ Q)) = P \circ Q(\text{CritP}(Q) \cup Q^{-1}(\text{CritP}(P))) = P(\text{CritP}(Q)) \cup P(\text{CritP}(P)) = P(\text{CritV}(Q)) \cup \text{CritV}(P) = P(\{q_1, q_2\}) \cup \{p_1, p_2\} \subseteq \{p_1, p_2\}$. Therefore $P \circ Q$ is Shabat. $\square$

The corresponding tree of $P \circ Q$ will be the tree corresponding to Q with the edges replaced by the tree corresponding to P, and more specifically each edge of T_Q will be replaced by the unique part between the vertices in T_P that are the image of all the white vertices and black vertices respectively of T_Q. In addition the parts of T_P which lie on either side of one of the image vertices of one color of vertices in T_Q (but not in between) will be pasted on in each instance of edge replacement by T_P in T_Q, outlined by example in 17 and 18. Notice that the number of edges of the composition tree will equal the number of edges of T_P times the number of edges of T_Q. Therefore the only trees that can be non-trivial compositions are those with a composite number of edges.

If you have a Shabat polynomial with only multiple roots, then you may take any power of it and still obtain a Shabat polynomial.

Theorem 3.10. *Let P be a Shabat polynomial with only multiple roots, then P^n will also be a Shabat Polynomial.*

Proof. Notice that $(P^n(z))' = nP'(z)(P^{n-1}(z))$, so $(P^n(z_0))' = 0$ if $P'(z_0) = 0$ or $P^{n-1}(z_0) = 0$. In the case that $P'(z_0) = 0$, then z_0 is a critical point of P, and thus $P(z_0) \in \{0, 1\}$. If $P^{n-1}(z_0) = 0$, then $P(z_0) = 0$, thus z_0 is a multiple root of P, which means that $P'(z_0) = 0$, so z_0 is a critical point of P, and thus $P(z_0) \in \{0, 1\}$. Therefore, P is a Shabat polynomial. $\square$

3.5. **Example of composition tree construction.** Let $P(z) = (z - i2\sqrt{2})(z-1)(z+1)$, and let $T_3(z) = 4z^3 - 3z$, that is, let T_3 be the third Chebyshev Polynomial. Notice that $P(-1), P(1) \in \{0, P(i\frac{\sqrt{2}}{2})\}$, thus $P \circ T_3$ must be a Shabat Polynomial. To find the corresponding tree of $P \circ T_3$, which we denote $T_{P \circ T_3}$, we go through the process outlined in the proof of Theorem 3.9.

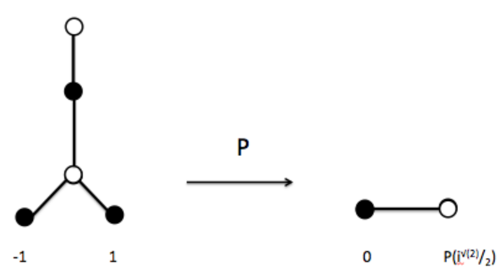


FIGURE 19. T_P as the preimage of $[0, P(i\frac{\sqrt{2}}{2})]$ under P

We see that two of the vertices of T_P happen to fall on the critical points of T_3, we then replace the edges of T_{T_3} with the component of T_P between 1 and -1, which happens to be all T_P.

When we replace the edges of T_{T_3} with copies of T_P, we do it in a manner that respects the orientation of the edges of T_{T_3}. The figure below shows the tree $T_{P \circ T_3}$. Notice that since $P(-1) = P(1) = 0$, the critical points of T_3, which map to 1 or -1, will map to 0. Therefore $\mathrm{CritP}(T_3) \subseteq \mathrm{CritV}(P \circ T_3)$, so the vertices of T_{T_3} will also be vertices of $T_{P \circ T_3}$

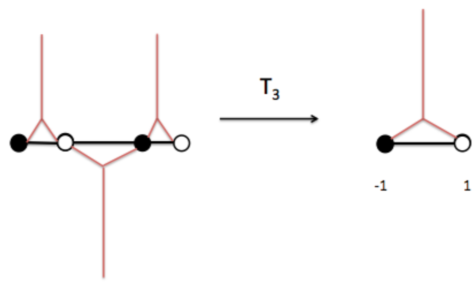


FIGURE 20. Tree for T_3 with T_P and its preimage under T_3 in red

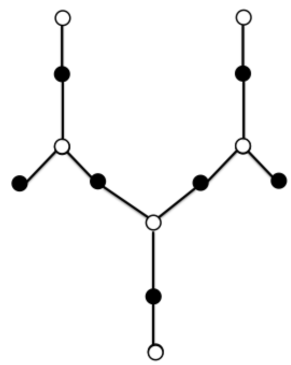


FIGURE 21. $T_{P \circ T_3}$

3.6. **Shabat polynomials on N distinct roots.** We assume from this point on that the given Shabat polynomial will correspond to the black vertices, unless stated otherwise. In this section we show computations of $W(z)$, and how we may exploit certain algebraic features in order to determine our Shabat polynomials

- 1 Distinct Root

Let $P(z) = (z - a)^{\alpha}$
$P'(z) = \alpha(z - a)^{\alpha-1}$

Any polynomial with only one distinct root will automatically be Shabat. The corresponding tree T_P will be a star, with a single black vertex surrounded by α white leaves. Its corresponding segment will be $[0, 1]$. Moreover we may fix $a = 0$, in order to get the unique up to equivalence Shabat polynomial family z^n.

- 2 Distinct Roots

Let $P(z) = (z - a)^{\alpha}(z - b)^{\beta}$.
$P'(z) = (z - a)^{\alpha-1}(z - b)^{\beta-1}(Az - B)$.
Where $A = \alpha + \beta$, and $B = \alpha b + \beta a$.

Any polynomial with two distinct roots will automatically be Shabat. The corresponding tree T_P will have two black vertices a and b with $\alpha - 1$ and $\beta - 1$ white leaves respectively. There will be one non-leaf white vertex B/A. Moreover we may fix $a = 0$, and $b = 1$, implying that $B = \alpha$, hence α and β determine the Shabat polynomial up to equivalence. The corresponding segment for P will be $[0, P(\frac{B}{A})]$

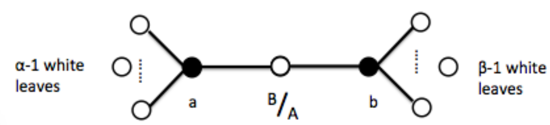


FIGURE 22. Tree of polynomial $P(z) = (z-a)^\alpha(z-b)^\beta$

- 3 Distinct Roots

Let $P(z) = (z-a)^\alpha(z-b)^\beta(z-c)^\gamma$.
$P'(z) = (z-a)^{\alpha-1}(z-b)^{\beta-1}(z-c)^{\gamma-1}(Az^2 - Bz + C)$.
Where $A = \alpha+\beta+\gamma$, $B = \alpha(b+c)+\beta(a+c)+\gamma(a+b)$, and $C = \alpha bc+\beta ac+\gamma ab$.
Let Δ be the discriminant of $Az^2 - Bz + C$, that is, let $\Delta = \sqrt{B^2 - 4AC}$. Notice that by the quadratic formula that $Az^2 - Bz + C = (z - \frac{B-\Delta}{2A})(z - \frac{B+\Delta}{2A})$, where the roots $\frac{B\pm\Delta}{2A}$ will correspond with the non-leaf white vertices of the tree T_P. In order for a polynomial of three distinct roots to be Shabat, it is necessary and sufficient that $P(\frac{B+\Delta}{2A}) = P(\frac{B-\Delta}{2A})$.

Case 1: We want $\Delta = 0$, then $P(\frac{B+\Delta}{2A}) = P(\frac{B-\Delta}{2A}) = P(\frac{B}{2A})$, which will make P Shabat. For P, fix $a = 0$ and $b = 1$, then $B^2 - 4AC = (\alpha(1+c)+\beta c+\gamma)^2 - 4(\alpha+\beta+\gamma)(\alpha bc) = (\alpha+\beta)^2c^2 + 2(\beta\gamma - \alpha(\alpha+\beta+\gamma))c + (\alpha+\gamma)^2$. In order for $\Delta = 0$, we will have to choose c to be root of $(\alpha+\beta)^2c^2 + 2(\beta\gamma - \alpha(\alpha+\beta+\gamma))c + (\alpha+\gamma)^2$, which can be done easily by application of the quadratic formula.

The corresponding tree T_P for this polynomial of zero discriminant will have black vertices 0, 1, and c with $\alpha - 1$, $\beta - 1$, and $\gamma - 1$ leaves respectively. The three black vertices will all be adjacent to the single non-leaf white vertex $\frac{B}{2A}$. The corresponding segment will be $[0, P(\frac{B}{2A})]$.

Case 2: Assume that $\Delta \neq 0$. In order for $P(\frac{B+\Delta}{2A}) = P(\frac{B-\Delta}{2A})$, it will require that $(B-2Aa\pm\Delta)^\alpha(B-2Ab\pm\Delta)^\beta(B-2Ac\pm\Delta)^\gamma$ be the same no matter which choice of $\pm$. Special attention must be paid to if (without loss of generality) α, β, and γ are all distinct; $\alpha = \beta$ and γ is distinct, or if $\alpha = \beta = \gamma$. In the case where you fix a and b, the polynomial given by solving $(B-2Aa+\Delta)^\alpha(B-2Ab+\Delta)^\beta(B-2Ac+\Delta)^\gamma = (B-2Aa-\Delta)^\alpha(B-2Ab-\Delta)^\beta(B-2Ac-\Delta)^\gamma$ may give different solutions for c but with equivalent polynomials.

The corresponding tree T_P will consist of a 4-path connecting the three black vertices a, b, and c together with the white vertices $\frac{B+\Delta}{2A}$ and $\frac{B-\Delta}{2A}$. The two non-leaf white vertices will have valency two, and without loss of generality the black vertices a, b,

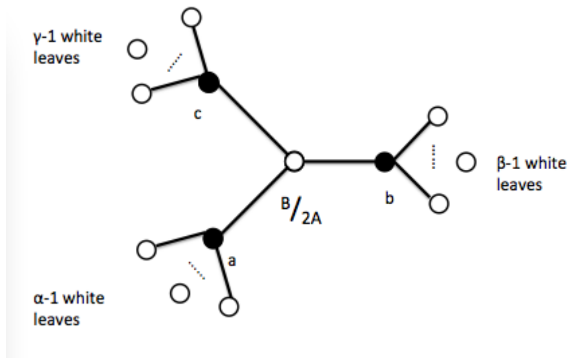
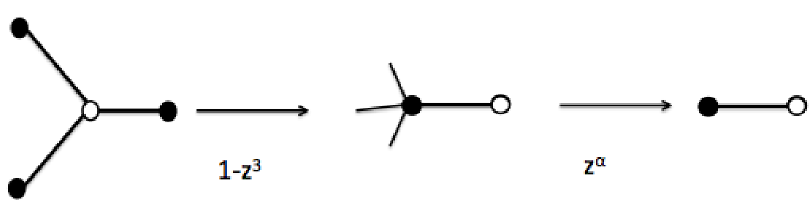


FIGURE 23. Tree of $P(z) = (z-a)^\alpha(z-b)^\beta(z-c)^\gamma$ when $\Delta = 0$

and c will have $\alpha - 1$, $\beta - 2$, and $\gamma - 1$ leaves respectively where b is the middle black vertex of the 4-path.

3.7. **Constructing Algebraic Families of Shabat Polynomials.** Using the machinery laid out in the previous section and using the fact that fixing any two bachelors will decide the Shabat polynomial up to the position of the second (non-black) vertex in the image segment. Therefore, we may give an explicit construction of families of polynomials based on the number of roots that they have. We will explicitly give a construction of two families of Shabat polynomials that correspond to a subset of planar trees having only three black vertices and a single non-leaf white vertex. These types of Shabat polynomials are described in the previous section as having 3 distinct roots, such that $P(z) = (z-a)^\alpha(z-b)^\beta(z-c)^\gamma$ where the discriminant of the polynomial of the roots unique to $P'(z)$ is zero. The construction is based on the degrees of the black vertices of T_P. In this construction we fix the non-leaf white vertex at 0 and one of the black vertices at 1

Example 3.11. *Let all the black vertices have the same degree α. In the first construction we give we use the composition of Shabat polynomials to form our P.*



Following from the construction of trees of composed Shabat polynomials, it is not to hard to see that $P(z) = (1 - z^3)^\alpha = (-1)^\alpha (z^3 - 1)^\alpha$ *will be Shabat and will correspond with our tree.*

Example 3.12. *Let two of the black vertices have degree* β*, but let the black vertex fixed at one have degree* $\alpha \neq \beta$*, hence* $P(z) = (z-1)^\alpha (z-b)^\beta (z-c)^\beta$*. First consider the complex conjugate* $\overline{P}(z) = (z-1)^\alpha (z-\overline{b})^\beta (z-\overline{c})^\beta$*, since* P *corresponds to a tree of a unique type,* P *and* $\overline{P}$ *must be equivalent, hence* $P = \overline{P}$ *and thus* $c = \overline{b}$*. Now, where* $A = \alpha + 2\beta$ *and* $B = \alpha(b+\overline{b}) + \beta(b+\overline{b}+2)$*,* $\frac{B}{2A} = 0 \implies 0 = B = \alpha(b+\overline{b}) + \beta(b+\overline{b}+2) \implies (\alpha+\beta)(b+\overline{b}) = -2\beta \implies b+\overline{b} = -\frac{2\beta}{\alpha+\beta}$*. Now in order for* T_P *to have the form we want, we need that* $0 = B^2 - 4AC$*. However,* $B = 0$ *and* $4A$ *is positive, so we only need that* $C = 0$ *where* $C = \alpha b\overline{b} + \beta(b+\overline{b}) = \alpha b\overline{b} - \beta(\frac{2\beta}{\alpha+\beta}) = 0$*. Now because of similarity we may assume that* $Re(b) = -1$*, hence* $\alpha b\overline{b} - \beta(\frac{2\beta}{\alpha+\beta}) = \alpha((Im(b))^2 + 1) - \beta(\frac{2\beta}{\alpha+\beta}) = 0$*, which implies that* $Im(b) = \sqrt{\frac{1}{\alpha}(\frac{2\beta^2}{\alpha+\beta} - \alpha)} = \sqrt{\frac{2\beta^2-\alpha\beta-\alpha^2}{\alpha(\alpha+\beta)}}$*. Hence* $b = 1 + i\sqrt{\frac{2\beta^2-\alpha\beta-\alpha^2}{\alpha(\alpha+\beta)}}$*, giving us the family of Shabat polynomials.*

Example 3.13. *We present in Figures 25 through 28 several pictures of combinatorial plane trees, along with representatives of the corresponding equivalence classes of Shabat polynomials. The polynomials were generated by taking the critical values to be zero (whose preimages are the black vertices) and one (whose preimages are the white vertices), choosing a black vertex to be zero and a white vertex to be one, and then solving the resulting systems of polynomial equations.*

FIGURE 24. $P(z) = z^2$ and $P(z) = \frac{1}{2}(z+1)(z-\frac{1}{2})^2$

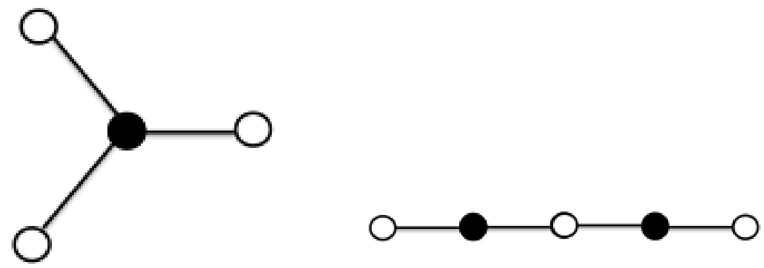


FIGURE 25. $P(z) = z^3$ and $P(z) = 4(z - \frac{\sqrt{2}}{2})^2(z + \frac{\sqrt{2}}{2})^2$

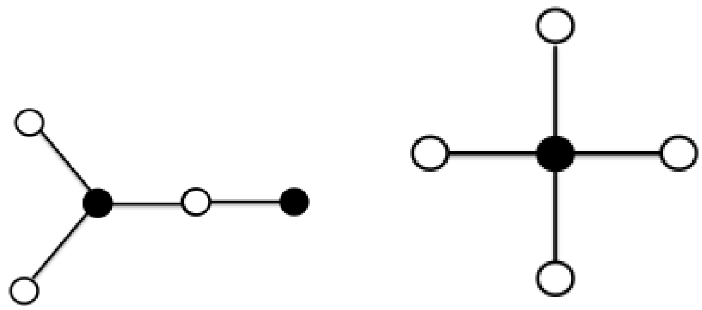


FIGURE 26. $\frac{-256}{27}z^3(z-1)$ and z^4

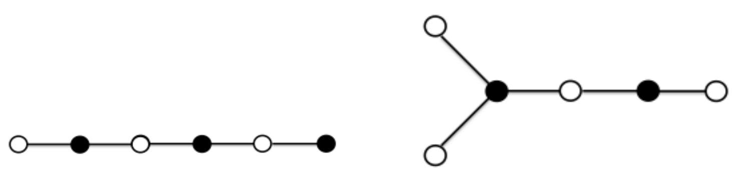


FIGURE 27. $(z-1)(z - \frac{1+\sqrt{5}}{4})^2(z - \frac{1-\sqrt{5}}{4})^2$ and $\frac{3125}{108}z^3(z-1)$

4. Algebraic Numbers and Combinatorial Bicolored Plane Trees

A surprising consequence of Theorem 2.4 is that our trees inherit certain algebraic properties from their corresponding polynomials. Since each tree is constructed by solving systems of algebraic equations, each tree can certainly be realized as the preimage of a segment between the critical values of a Shabat polynomial with algebraic coefficients. As the universal Galois group $\Gamma = \mathrm{Aut}(\overline{\mathbb{Q}}/\mathbb{Q})$ acts on the set of such

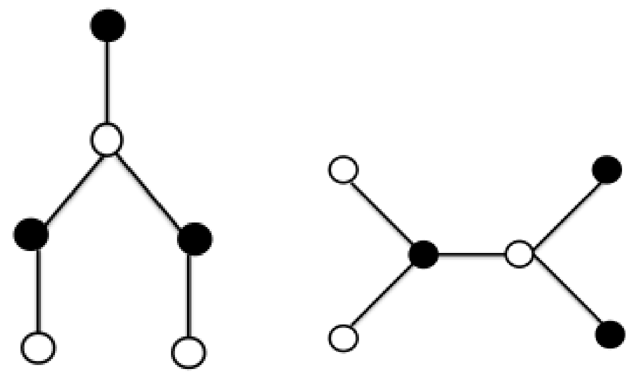


FIGURE 28. $\frac{-2000}{27\sqrt{5}}z^2(z-1)^2(z-\frac{2+\sqrt{5}}{4})$ and $\frac{135}{512}z^3(z-\frac{\sqrt{15}+i}{\sqrt{15}})(z-\frac{\sqrt{15}-i}{\sqrt{15}})$

Shabat polynomials by sending coefficients to conjugate coefficients, we can naturally define an action of Γ on the set of combinatorial bicolored plane trees. If T_1 and T_2 are two bicolored plane trees with corresponding polynomials f_1 and f_2 (over $\overline{\mathbb{Q}}$) such that the element g of Γ conjugates f_1 to f_2, then we say $g(T_1) = T_2$. We can further study the algebraic properties of trees by studying and understanding the field of definition of a tree.

Definition 4.1. *Let T be a bicolored plane tree. The* field of definition *of T is the Galois extension of the subfield K of $\overline{\mathbb{Q}}$ corresponding to the subgroup of Γ which fixes T.*

The above definition is partially motivated by the following theorem, which gives surprising algebraic information about bicolored plane trees.

Theorem 4.2. *For any bicolored plane tree T there exists a Shabat polynomial whose coefficients belong to the field of definition of T.*

Proof. Given a bicolored plane tree, let $p_k(z) = (z-b_{k1})\cdots(z-b_{kl})$ (resp. $q_k(z) = (z-a_{k1})\cdots(z-a_{km})$) denote the monic polynomial whose roots are black (resp $q_k(z)$) vertices of valency k. The coefficients of these polynomials are elementary symmetric functions of the coordinates of the vertices of the same color and valency. We call these symmetric functions vertex combinations. Note that a vertex combination includes only vertices of same color and valency.

As an example, for the given tree in figure 4, we have $p_3(z) = (z-b_{31})(z-b_{32})$, $q_1(z) = (z-a_{11})(z-a_{12})(z-a_{13})(z-a_{14})$, and $q_2(z) = (z-a_{21})$. The vertex combinations look like $b_{31}b_{32}$, $b_{31}+b_{32}$, $a_{11}a_{12}a_{13}+a_{11}a_{12}a_{14}+a_{11}a_{13}a_{14}+a_{12}a_{13}a_{14}$, etc.

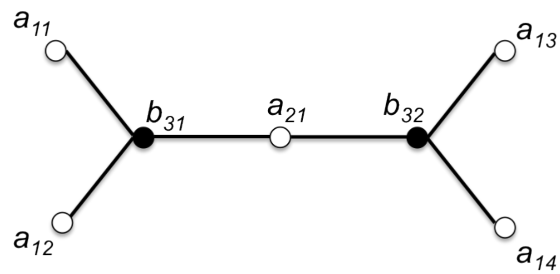


FIGURE 29. Vertex combination

We look for a polynomial of the form

$$P(z) = \lambda \prod p_k(z)^k$$

such that

$$P(z) - 1 = \lambda \prod q_k(z)^k.$$

The coefficients give us (#vertices−1) equations but we have (#vertices+1) unknowns. We will need to add two equations. Intuitively, this depends on where we put the tree on the complex plane. Let σ denote the sum of vertices of degree one with the same color (black or white), and set $\sigma = 0$. This means that the position of the tree is determined up to a transformation $z \mapsto Az$, for some $A \in \mathbb{C}$.

Now let m be the order of symmetry of the tree. This means m is the largest number such that $P(z) = P(\omega_m z)$, where ω_m is the mth root of unity. We show that m equals the smallest nonzero degree of all vertex combinations. Consider any $p_k(z) = (z - b_{k1}) \cdots (z - b_{kl})$. We have that $(z - b_{k1}) \cdots (z - b_{kl}) = (z - \omega_m b_{k1}) \cdots (z - \omega_m b_{kl})$. Since $\omega_m^n \neq 1$ for $n < m$, any vertex combination of degree less than m has to equal 0 for the equality to hold. Conversely, when m' is the smallest nonzero degree of all vertex combinations, we have $p_k(z) = p_k(\omega_{m'} z)$, so $m' \leq m$. The claim then follows. Now the position of the tree is determined up to $z \mapsto Az$, with $A^m = 1$.

Let $\overline{P}$ be another solution of the system of equations, and suppose $\overline{P}$ gives the same tree as P. This means $\overline{P} = P(Az) = P$. This shows the uniqueness of P.

□

Before we prove that the absolute Galois group acts faithfully on the set of combinatorial bicolored plane trees, we need to use the following lemmas.

Lemma 4.3. *Let f be a polynomial of degree n, and let $d|n$. Suppose there is a monic polynomial h of degree d such that $h(0) = 0$ and for some polynomial g, $f = g(h)$. Then h is unique.*

Proof. By definition,

$$f = a_m h^m + \cdots + a_1 h + a_0$$

where $m = n/d$. h is monic and has no constant term, so there are $d-1$ coefficients of h which are unknown. The highest d-degree terms in f (the terms of degree $n, \ldots n-d+1$) must equal the terms of corresponding degree in $a_m h^m$. This gives us d independent equations and d unknowns (the roots of h), so there is at most one solution. □

Lemma 4.4. *Let $g, h, \bar{g}, \bar{h}$ be polynomials such that $g(h) = \bar{g}(\bar{h})$ and deg $h =$ deg $\bar{h}$. Then for some constants c and d, $\bar{h} = ch + d$.*

Proof. For some constants c_1, d_1, c_2, d_2, the polynomials

$$h/c_1 - d_1 \quad \text{and} \quad \bar{h}/c_2 - d_2$$

are monic and have no constant term. By altering g and $\bar{g}$ accordingly into g_1 and $\bar{g}_2$, we get

$$g(h) = g_1(h/c_1 - d_1) = \bar{g}_2(\bar{h}/c_2 - d_2).$$

By Lemma 4.3,

$$h/c_1 - d_1 = \bar{h}/c_2 - d_2 \Rightarrow h = \frac{c_1}{c_2}\bar{h} + (d_1 - d_2)c_1.$$

□

This next lemma is a quite powerful result from the theory of Belyi functions, which are studied extensively in [6]. This allows us to easily produce Shabat polynomials which are acted upon by specific elements of the absolute Galois group, which is central to the proof that this group acts faithfully on the set of combinatorial bicolored plane trees.

Lemma 4.5. *For any polynomial P with algebraic coefficients there exists a polynomial f with rational coefficients such that $f \circ P$ is a Shabat polynomial.*

For any polynomial g, let $\mathrm{CritV}(g)$ be the set of critical values of g, and $\mathrm{CritP}(g)$ the set of critical points. First, observe that for any polynomials g and f,

$$(g \circ f)'(z) = f'(z)(g' \circ f)(z).$$

Hence,

$$\begin{aligned} \mathrm{CritP}(g \circ f) &= \mathrm{CritP}(f) \cup f^{-1}(\mathrm{CritP}(g)), \text{ and} \\ \mathrm{CritV}(g \circ f) &= g(\mathrm{CritV}(f)) \cup \mathrm{CritV}(g). \end{aligned}$$

From now on, call this observation $*$.

Let S_0 be the set of all irrational critical values of $P = P_0$, along with all of their algebraic conjugates. Let P_1 be the polynomial annihilating S_0. As S_0 contains the algebraic conjugates of each of its elements, P_1 has only rational coefficients. This is because the set S_0 by definition is invariant under any action by the Galois group Γ. Let $n = |S_0| = \deg(P_0)$, and let $S_1 = \mathrm{CritV}(P_1)$. As P_1 has only rational coefficients, S_1 must contain all of the algebraic conjugates of its elements. Furthermore, $\deg(P_1) = n$, so $|S_1| \leq n-1$. By $*$, $\mathrm{CritV}(P_1 \circ P_0) = \mathrm{CritV}(P_1) \cup \{0\}$ since P_1 annihilates S_0.

For $i < n$, let P_i be the polynomial annihilating S_{i-1}, and let $S_i = \mathrm{CritV}(P_i)$. By induction, $|S_i| \leq n - i$ and $\deg(P_i) \leq n - i + 1$. By $*$,

$$\mathrm{CritV}(P_i \circ P_{i-1} \circ \cdots \circ P_1 \circ P_0) = \mathrm{CritV}(P_i) \cup \mathrm{CritV}(P_{i-1} \circ \cdots \circ P_0).$$

As each P_i has rational coefficients and maps P_{i-1}'s irrational critical points to 0, a rational number. Therefore, the only irrational critical values resulting from the above composition can be those of P_i. Terminate this process at the first polynomial P_k which is linear. The resulting composition

$$g = P_k \circ P_{k-1} \circ \cdots \circ P_1 \circ P_0$$

will have all rational critical values.

With appropriate scaling by rationals, we can assume all critical values of g are contained in a unit length interval. With a translation by a rational, we may insist that the interval is $[0, 1]$. By abuse of notation, we will assume g only has critical values in $[0, 1]$.

While our composition still has a critical value which is not 0, write that critical value as $\frac{m}{m+n}$. Construct

$$P_{m,n}(z) = \frac{(m+n)^{m+n}}{m^m n^n} z^m (1-z)^n.$$

Assuming m and n are both greater than 1 (if they are not, simply multiply them both by a sifficiently large scalar), each $P_{m,n}$ will be Shabat. Furthermore, note that

$$\begin{aligned} P_{m,n}\left(\frac{m}{m+n}\right) &= \frac{(m+n)^{m+n}}{m^m n^n}\left(\frac{m}{m+n}\right)^m \left(1 - \left(\frac{m}{m+n}\right)\right)^n \\ &= \frac{(m+n)^m (m+n)^n}{m^m n^n}\left(\frac{m}{m+n}\right)^m \left(\frac{n}{m+n}\right)^n = 1. \end{aligned}$$

Therefore, $P_{m,n} \circ g$ will have one fewer rational critical value than g. If $\frac{m_1}{m_1+n_1}, \ldots, \frac{m_l}{m_l+n_l}$ are the rational critical values of $g, P_{m_1,n_1} \circ g, \ldots P_{m_{l-1},n_{l-1}} \circ \cdots \circ g$, then by $*$ the polynomial

$$f = P_{m_l,n_l} \circ \cdots \circ P_{m_1,n_1} \circ P_k \circ \cdots \circ P_1$$

is a polynomial with rational coefficients such that $f \circ P$ is Shabat.

We finally have the machinery needed to prove the much-anticipated theorem about the faithful action of $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$.

Theorem 4.6. *The action of $\Gamma = Gal(\overline{\mathbb{Q}}/\mathbb{Q})$, on the set of combinatorial bicolored plane trees is faithful.*

Proof. Let $K = \mathbb{Q}[\alpha]$ be a number field, and let $\beta \cong \alpha$ such that $\beta \neq \alpha$. Define

$$p_\alpha = \int z^3(z-1)^2(z-\alpha)dz$$
$$p_\beta = \int z^3(z-1)^2(z-\beta)dz$$

such that p_α and p_β have constant terms 0. We know there exists a polynomial f over the rationals such that $P_\alpha = f \circ p_\alpha$ is Shabat. In the proof of Lemma 4.5, the set S_0 contains all conjugates of α, including β, so the same polynomial f will satisfy $P_\beta = f \circ p_\beta$ is Shabat. Let T_α and T_β be the bicolored plane trees obtained from P_α and P_β, respectively.

Claim: T_α and T_β are distinct.

Suppose the contrary. By Theorem 2.3,

$$P_\alpha = AP_\beta(az+b) + B \Rightarrow f(p_\alpha(z)) = Af(p_\beta(az+b)) + B$$

Applying Lemma 4.4 to $g = f$, $h = p_\alpha$, $\overline{g} = Af + B$, and $\overline{h} = p_\beta(az+b)$, we know there exist constants c and d such that

$$p_\alpha(z) = cp_\beta(az+b) + d.$$

Furthermore,

$$\mathrm{CritP}(p_\beta(az+b)) = \left\{\frac{-b}{a}, \frac{1-b}{a}, \frac{\beta-b}{a}\right\}$$
$$\mathrm{CritP}(p_\alpha(z)) = \{0, 1, \alpha\}.$$

Since these polynomials are equivalent, $\mathrm{CritP}(p_\beta(az+b)) = \mathrm{CritP}(p_\alpha(z))$. Because each of these critical points has different multiplicity in the derivative, we may write

$$\frac{-b}{a} = 0, \quad \frac{1-b}{a} = 1, \quad \frac{\beta-b}{a} = \alpha,$$

implying $b = 0, a = 1$, and $\beta = \alpha$. We assumed $\beta \neq \alpha$, a contradiction. □

Theorem 4.6 illustrates how the algebraic richness of the world of polynomials can be translated into the world of combinatorial bicolored plane trees. The group $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$ is poorly understood, and mathematicians have struggled with the famous 'inverse Galois problem' for centuries. Results like Theorem 4.6 help to peel away the layers of abstraction associated with the group $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$, in the hopes of helping us better understand this group.

5. Continued Fractions and Plane Trees

5.1. **Introduction.** Continued fractions are best known as an interesting and often beautiful way of expressing real numbers. However, the concept of a continued fraction can be expanded to include analogous representations of functions. Whereas most of paper [7] deals with a bijection between plane trees and Shabat polynomials, the paper briefly explains a function from plane trees to a different subset of polynomials: those which display a certain symmetry in the continued fraction representation of their square roots. We will introduce continued fractions as a topic in number theory, generalize them to functions, and explain their relationship to plane trees.

5.2. **Continued fractions: the number case.**

5.2.1. *Finite continued fractions.*

Definition 5.1. (Definition 8.1 of [3]) *A finite continued fraction (of a real number) is an expression of the form*

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots + \cfrac{1}{a_n}}}}$$

where each of the a_i are real numbers and $a_1, \ldots, a_n$ are positive.

The a_is of a continued fraction are called *coefficients*. In the case in which each of the coefficients is an integer the continued fraction is called *simple*. Due to the cumbersome nature of such expressions we will often use the notation

$$[a_0, a_1, \ldots, a_n]$$

to represent the above continued fraction.

Example 5.2. *The number $\dfrac{2013}{4}$ can be written as a finite continued fraction in the following manner.*

$$2013 = 4 \times 503 + 1$$

So

$$\frac{2013}{4} = 503 + \frac{1}{4} = [503, 4]\,.$$

We will discuss generally how to compute such expressions below after we have seen why we may want to do such a thing.

5.2.2. *Infinite continued fractions.* Now in the finite case discussed above, each of the successive a_k's intuitively seems to be less and less relevant to the value of the expression. For instance, in example (5.2), 503 by itself is pretty close to $\frac{2013}{4}$. The below discussion seeks to capture and exploit this intuition.

Definition 5.3. (Definition 8.15 of [3]) *Let $a_0, a_1, a_2, \ldots$ be an infinite sequence of integers with each a_i positive for $i \geq 1$. Then*

$$\lim_{n\to\infty} [a_0, a_1, \ldots, a_n] = [a_0, a_1, a_2, \ldots]$$

is called an infinite simple continued fraction.

If the sequence $a_0, a_1, a_2, \ldots$ is eventually periodic, then the continued fraction is said to be periodic. If the sequence is purely periodic, then so is the continued fraction said to be. We have not answered whether or not this limit always exists. The affirmative is true, but to see this we will need to do some legwork first.

A useful concept to introduce at this point is that of a *convergent*. The n-th convergent, C_n, of the (finite or infinite) continued fraction $[a_0, a_1, a_2, \ldots]$ is simply the finite continued fraction $[a_0, a_1, \ldots, a_n] = C_n = \frac{x_n}{y_n}$ where x_n and y_n are not necessarily integers, though they are if the continued fraction is simple (Definition 8.5 of [3]). A key observation about these convergents is the following proposition.

Proposition 5.4. ([10]) *Let $[a_0, a_1, a_2, \ldots]$ be a (not necessarily simple) continued fraction whose n-th convergents are $C_n = \frac{x_n}{y_n}$. Then for all $n \geq 0$,*

$$\begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix}\cdots\begin{pmatrix} a_n & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} x_n & x_{n-1} \\ y_n & y_{n-1} \end{pmatrix}. \tag{5.1}$$

Proof. We will induct on n. First note that $\frac{x_0}{y_0} = a_0 = \frac{a_0}{1}$. We can define $x_{-1} = 1$ and $y_{-1} = 0$ so that

$$\begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} x_0 & x_{-1} \\ y_0 & y_{-1} \end{pmatrix}.$$

Now suppose for all continued fractions, equation (5.1) holds for all $k < n$. Then

$$\begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix}\cdots\begin{pmatrix} a_{n-1} & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} x_{n-1} & x_{n-2} \\ y_{n-1} & y_{n-2} \end{pmatrix}.$$

We'll let

$$\begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix}\cdots\begin{pmatrix} a_{n-2} & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} x_{n-2} & x_{n-3} \\ y_{n-2} & y_{n-3} \end{pmatrix} = M.$$

Then we can think of the n-th convergent as the $(n-1)$-st convergent of the continued fraction expansion that is identical to the expansion we are interested in, but with

$(n-1)$-st term equal to $a_{n-1} + \frac{1}{a_n}$, instead of a_{n-1}. By hypothesis

$$M \begin{pmatrix} a_{n-1} + \frac{1}{a_n} & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} x_n & x_{n-2} \\ y_n & y_{n-2} \end{pmatrix}.$$

So

$$\begin{aligned} M \begin{pmatrix} a_{n-1} + \frac{1}{a_n} \\ 1 \end{pmatrix} &= M \begin{pmatrix} a_{n-1} \\ 1 \end{pmatrix} + M \begin{pmatrix} \frac{1}{a_n} \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} x_{n-1} \\ y_{n-1} \end{pmatrix} + \frac{1}{a_n} \begin{pmatrix} x_{n-2} \\ y_{n-2} \end{pmatrix} \\ &= \begin{pmatrix} x_n \\ y_n \end{pmatrix}. \end{aligned}$$

Thus

$$\frac{x_n}{y_n} = \frac{x_{n-1} + \frac{x_{n-2}}{a_n}}{y_{n-1} + \frac{y_{n-2}}{a_n}} = \frac{a_n x_{n-1} + x_{n-2}}{a_n y_{n-1} + y_{n-2}}. \tag{5.2}$$

Therefore

$$\begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_n & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} x_{n-1} & x_{n-2} \\ y_{n-1} & y_{n-2} \end{pmatrix} \begin{pmatrix} a_n & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} x_n & x_{n-1} \\ y_n & y_{n-1} \end{pmatrix}.$$

□

There are a number of useful identies to be learned from this proof (The following corollaries and proposition can all be found in [3] Chp. 8). For instance, equation (5.2) implies a simple recurrence relation about the convergents.

Corollary 5.5. *Let $C_k = \frac{x_k}{y_k}$ denote the k-th convergent of the continued fraction $[a_0, a_1, a_2, \ldots]$. Then*

$$x_n = a_n x_{n-1} + x_{n-2}, \;\; x_0 = a_0, x_{-1} = 1 \tag{5.3}$$

$$y_n = a_n y_{n-1} + y_{n-2} \;\; y_0 = 1, y_{-1} = 0 \tag{5.4}$$

By taking the determinant of both sides of equation (5.1) we usefully obtain

Corollary 5.6. *Let x_k, y_k be as in Corollary 5.5. Then*

$$x_n y_{n-1} - y_n x_{n-1} = (-1)^{n-1}. \tag{5.5}$$

Dividing by $y_n y_{n-1}$, we get

Corollary 5.7. *Let x_k, y_k, C_k be as in Corollary 5.5. Then*

$$C_n - C_{n-1} = \frac{x_n}{y_n} - \frac{x_{n-1}}{y_{n-1}} = \frac{(-1)^{n-1}}{y_n y_{n-1}}. \tag{5.6}$$

We are now well-equipped to show that Definition 5.3 makes sense. That is, we can prove the following proposition

Proposition 5.8. *Let $a_0, a_1, a_2 \dots$ be a sequence of integers with each a_i positive for $i \geq 1$. The limit*

$$\lim_{n\to\infty} [a_0, a_1, \dots, a_n]$$

converges.

Proof. Let $\epsilon > 0$ be given. We will show that the sequence of convergents is Cauchy, that is, that there exists some N such that if $N \leq l < n$ then $|C_l - C_n| < \epsilon$. Notice that the triangle inequality and equation (5.6) give us

$$\begin{aligned} |C_l - C_n| &\leq \sum_{k=l+1}^{n} |C_k - C_{k-1}| \\ &= \sum_{k=l+1}^{n} \frac{1}{|y_k y_{k-1}|}. \end{aligned}$$

But because each a_i (except possibly a_0) is a positive integer, equation (5.4) implies that the y_i's form a strictly increasing sequence of integers with $y_0 = 1$. Therefore $y_k > k$, so

$$|C_l - C_n| \leq \sum_{k=l+1}^{n} \frac{1}{|y_k y_{k-1}|} < \sum_{k=l}^{n} \frac{1}{k^2}$$

The last sum converges so for some N, $\sum_{k=N}^{\infty} \frac{1}{k^2} < \epsilon$. Therefore, the sequence is Cauchy. □

This is great news for Definition 5.3, but as we will see, danger still looms large over the following attempt to compute one of these infinite continued fractions.

Example 5.9. *Consider solutions to the equation $x^2 - 2x - 1 = 0$, which has roots $x = 1 \pm \sqrt{2}$. Notice that $x(x-2) = 1$. So if $x \neq 0$ (which we know to be the case) then*

$$x = 2 + \frac{1}{x}.$$

Substituting this last expression into itself yields

$$x = 2 + \cfrac{1}{2 + \cfrac{1}{x}}.$$

And again

$$x = 2 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{x}}}.$$

And so on ad infinitum

$$x = 2 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{\ddots}}}.$$

But which x is it, $1+\sqrt{2}$ or $1-\sqrt{2}$? We may notice that the right hand side must be greater than 2 and conclude that we have a continued fraction expansion for $x = 1+\sqrt{2}$. But that fails to answer why our method somehow "preferred" this root over its conjugate. To gain some insight into these matters, first note that any purely periodic continued fraction can be written in the form

$$\alpha = \left[a_0, a_1, \ldots, a_{n-1}, \frac{1}{\alpha}\right],$$

whose n-th convergent is equal to α. Equation (5.2) then gives us

$$C_n = \alpha = \frac{\alpha x_{n-2} + x_{n-1}}{\alpha y_{n-2} + y_{n-1}},$$

which is equivalent to the quadratic

$$y_{n-2}\alpha^2 + (y_{n-1} - x_{n-2})\alpha - x_{n-1} = 0. \tag{5.7}$$

The following theorem about the roots of this quadratic tells us for which x we have found a continued fraction expansion in our example. Remember that throughout this discussion of periodicity we have said nothing about whether or not the a_i are integers, so our previous theorem about the convergence of infinite simple continued fractions is irrelevant.

Theorem 5.10. (Theorem 8.1 of [11]) *Let α_1, α_2 be the fixed points of the transformation*

$$f(\alpha) = \left[a_0, a_1, \ldots, a_{n-1}, \frac{1}{\alpha}\right] = \frac{\alpha x_{n-2} + x_{n-1}}{\alpha y_{n-2} + y_{n-1}},$$

i.e.$f(\alpha_1) = \alpha_1$ and $f(\alpha_2) = \alpha_2$. Then the continued fraction

$$[a_0, a_1, \ldots, a_{n-1}, a_0, a_1, \ldots, a_{n-1}, a_0, a_1 \ldots] \tag{5.8}$$

converges if and only if α_1, α_2 satisfy

$$\alpha_1 = \alpha_2,$$

or

$$|C_{n-1} - \alpha_2| > |C_{n-1} - \alpha_1|, \ C_p \neq \alpha_2$$
$$p = 0, 1, 2, \ldots, n-1.$$

If the continued fraction in (5.8) converges, it equals α_1.

Proof. Notice that
$$C_{nk+p} = f^k(C_p),\ p = 0, 1, 2, \ldots, n-1. \tag{5.9}$$
Thus C_p is a fixed point of f if and only if $C_p = C_{nk+p}$. We consider three cases.

- Case 1: $\alpha_1 = \alpha_2$. In this case the quadratic (5.7) can be solved for α_1. This gives us
$$\frac{1}{f(\alpha) - \alpha_1} = \frac{1}{\alpha - \alpha_1} + \frac{1}{C_{n-1} - \alpha_1},$$
which, when iterated k times gives
$$\frac{1}{f^k(\alpha) - \alpha_1} = \frac{1}{\alpha - \alpha_1} + \frac{k}{C_{n-1} - \alpha_1}.\ \text{(See [11])} \tag{5.10}$$
Substituting in C_p for α and making use of equation (5.9) we get
$$\frac{1}{f^k(C_p) - \alpha_1} = \frac{1}{C_{nk+p} - \alpha_1} = \frac{1}{C_p - \alpha_1} + \frac{k}{C_{n-1} - \alpha_1}. \tag{5.11}$$
If $C_p = \alpha_1$ we know that $C_{nk+p} = \alpha_1$, so the following limit will hold, regardless. We take the limit of both sides of equation (5.11):
$$\lim_{k\to\infty} \frac{1}{C_{nk+p} - \alpha_1} = \lim_{k\to\infty} \frac{1}{C_p - \alpha_1} + \frac{k}{C_{n-1} - \alpha_1}$$
We know that C_p, C_{n-1}, α_1 are fixed numbers so the right hand side goes to infinity. Therefore, the denominator of the left hand side must go to zero. Thus for all p,
$$\lim_{k\to\infty} C_{nk+p} = \alpha_1$$
so the convergents in fact converge to α_1, making this the value of the continued fraction.
- Case 2: $\alpha_1 \neq \alpha_2$ and $|C_{n-1} - \alpha_2| > |C_{n-1} - \alpha_1|$. Again using the quadratic formula and the fact that the roots of equation (5.7) are in this case distinct, we have the following (from [11]):
$$\frac{f(\alpha) - \alpha_1}{f(\alpha) - \alpha_2} = \frac{(C_n - \alpha_1)(\alpha - \alpha_1)}{(C_n - \alpha_2)(\alpha - \alpha_2)}.$$
This time the k-th iteration is given by
$$\frac{f^k(\alpha) - \alpha_1}{f^k(\alpha) - \alpha_2} = \frac{(C_{n-1} - \alpha_1)^k(\alpha - \alpha_1)}{(C_{n-1} - \alpha_2)^k(\alpha - \alpha_2)} = \sigma^k \frac{(\alpha - \alpha_1)}{(\alpha - \alpha_2)} \tag{5.12}$$
where $|\sigma| = |\frac{(C_{n-1} - \alpha_1)}{(C_{n-1} - \alpha_2)}| < 1$, by hypothesis. We let $\alpha = C_p$ and use equation (5.9) to obtain
$$\frac{C_{nk+p} - \alpha_1}{C_{nk+p} - \alpha_2} = \sigma^k \frac{(C_p - \alpha_1)}{(C_p - \alpha_2)}. \tag{5.13}$$

If $C_p \neq \alpha_2$, we now have

$$C_{nk+p} - \alpha_1 = \epsilon_k (C_{nk+p} - \alpha_2)$$

with $\epsilon_k = \sigma^k \frac{C_p - \alpha_1}{C_p - \alpha_2}$, $\lim_{k\to\infty} \epsilon_k = 0$. But now we have

$$(1 - \epsilon_k) C_{nk+p} = \alpha_1 - \epsilon_k \alpha_2$$

So

$$C_{nk+p} = \frac{\alpha_1 - \epsilon_k \alpha_2}{1 - \epsilon_k}.$$

Subtracting α_1 from both sides we find

$$C_{nk+p} - \alpha_1 = \frac{\epsilon_k (\alpha_1 - \alpha_2)}{1 - \epsilon_k}.$$

Upon taking the limit of both sides, we find that $\lim_{k\to\infty} C_{nk+p} - \alpha_1 = 0$ and thus the convergents converge to α_1. Now if we had for some p that $C_p = \alpha_2$, then we know $C_{nk+p} = \alpha_2$ for all k by equation (5.9). But for all other $m = 0, 1, 2, \ldots, n-1, m \neq p$ we have $\lim_{k\to\infty} C_{nk+p} = \alpha_1 \neq \alpha_2$, and thus the continued fraction no longer converges to one value, and thus diverges.

- Case 3: $\alpha_1 \neq \alpha_2$ and $|C_{k-1} - \alpha_1| = |C_{k-1} - \alpha_2|$. We have σ as in Case 2, but this time $|\sigma| = 1$. But because $\alpha_1 \neq \alpha_2$, we know that $\sigma \neq 1$, and thus the sequence of σ^k's must have at least two different limit points. We let $p = n-1$ in equation (5.13) and observe

$$\frac{C_{(k+1)n-1} - \alpha_1}{C_{(k+1)n-1} - \alpha_2} = \sigma^{k+1}$$

and thus the sequence $(\frac{C_{(k+1)n-1} - \alpha_1}{C_{(k+1)n-1} - \alpha_2})$ has at least two limit points which means that so must the sequence (C_{kn-1}) so the sequence (C_k) cannot converge.

□

Now this theorem shows us what happened in our example. We have

$$\alpha_1 = 1 + \sqrt{2}, \quad \alpha_2 = 1 - \sqrt{2}, \quad C_{n-1} = 2.$$

Which means we satisfy the second condition of the theorem because

$$|1 - \sqrt{2} - 2| > |1 + \sqrt{2} - 2|.$$

And thus

$$1 + \sqrt{2} = 2 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{\ddots}}}.$$

Alternatively, we could note that

$$1+\sqrt{2} = 2.414... = 2 + .414... = 2 + \frac{1}{r_1}$$

for some $r_1 > 1$. We seek the greatest integer less than or equal to r_1 so we can write the above expression in the form

$$1+\sqrt{2} = 2 + \cfrac{1}{a_1 + \cfrac{1}{r_2}}.$$

Solving $\frac{1}{r_1} = 1+\sqrt{2}-2$ reveals $r_1 = 1+\sqrt{2}$, which, to no one's surprise based off the above discussion, shows

$$1+\sqrt{2} = 2 + \cfrac{1}{2 + \cfrac{1}{r_2}}.$$

Repeating this process will give the desired formula:

$$1+\sqrt{2} = 2 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{\ddots}}}.$$

In general, to compute a simple continued fraction expansion of some real number r we use the following algorithm: Find the largest integer a_0 such that $a_0 \leq r$. Then $r = a_0 + \frac{1}{r_1}$ for some r_1. Next find the largest integer a_1 such that $a_1 \leq r_1$. Then $r = a_0 + \cfrac{1}{a_1 + \cfrac{1}{r_2}}$ for some r_2. Continue this process until your curiosity is satisfied.

5.2.3. *Continued fractions of functions.* This is all very wonderful for integers, but our primary concern will be representing continued fractions of functions. Specifically, we are seeking a representation for the square roots of polynomials. What exactly does this mean and why would we do it? The idea of representing a number as a continued fraction is to represent it as something resembling a rational number. Similarly the idea behind continued fractions of functions is to represent the function as something resembling a rational function. Specifically,

Definition 5.11. *A continued fraction of a function, f, is an expression of the form*

$$f(z) = a_0(z) + \cfrac{1}{a_1(z) + \cfrac{1}{a_2(z) + \cfrac{1}{\ddots}}} = [a_0, a_1, a_2, \ldots]$$

where each of the a_i are polynomials.

Now we wish to construct a continued fraction of a function, but it is unclear how to proceed. In the case of irrational numbers we started by taking the integer part of the number under consideration, but the analog to an integer part of a function is unclear. Van der Poorten and Tran describe the process as follows in [10]: Given a function, f, let $z = \frac{1}{x}$, such that $f(\frac{1}{x})$ has finitely many poles, and consider the Laurent Series of $f(\frac{1}{x})$, centered around 0. This gives us a series representation converging to $f(\frac{1}{x})$ for sufficiently small $\frac{1}{x}$, which is equivalent to a series representation for $f(z)$ for sufficiently large z. This series is

$$f(\frac{1}{x}) = \sum_{i=-m}^{\infty} b_i x^i$$

Substituting in $z = \frac{1}{x}$, we now take the polynomial part of $f(z)$, that is, $\sum_{i=0}^{m} b_i z^i$, and use it to start our continued fraction. We have

$$f(z) = \sum_{i=0}^{m} b_i z^i + \frac{1}{F_1(z)}$$

for some function $F_1(z)$. Let $a_0 = \sum_{i=0}^{m} b_i z^i$ so

$$f(x) = a_0 + \frac{1}{F_1(z)}.$$

We can rearrange this expression to give $F_1(z) = \frac{1}{f(z)-a_0}$ and then compute the polynomial part of $F_1(z)$ in the same manner we did for $f(z)$, and repeat. More explicitly we define $F_{i+1} = \frac{1}{F_i - a_i}$ and $F_0 = f$, where a_i is the polynomial part of F_i. We can then see that the continued fraction expansion of f is given by

$$f(z) = a_0(z) + \cfrac{1}{a_1(z) + \cfrac{1}{a_2(z) + \ddots}}$$

for z sufficiently large.

5.3. **The number case vs. the polynomial case.** One may notice that many of the results we proved in our discussion of continued fractions of real numbers actually generalize, with identical proofs, to the case in which each of the a_is is a polynomial (Proposition 5.4 and Corollaries 5.5, 5.6, and 5.7). However, there is at least one important fact about simple continued fractions that does not hold for continued fractions of functions. This is the well known result that for a non-square integer m the continued fraction expansion of $\sqrt{m}$ is always eventually periodic (Lemma 8.35 of [3]). The fact that this is not so if m is a polynomial raises the question of when exactly the continued fraction expansion of the square root of a polynomial is eventually periodic. To this end Abel proved in 1826 the following incredible theorem:

Theorem 5.12. ([1]) *The following are equivalent*

(a) *Given $D(z)$, there exist polynomials P, Q such that the Pell equation for polynomials, $P(z)^2 - D(z)Q(z)^2 = 1$, is satisfied.*
(b) *The square root $\sqrt{D(z)}$ may be represented as an eventually periodic continued fraction.*
(c) *An integral*

$$\int \frac{f(z)}{\sqrt{D(z)}} dz$$

for some $f(z)$ of deg $p \leq$ deg $D - 2$ can be computed in elementary functions.

We will show the path $b) \Rightarrow a) \Rightarrow c)$, the rest of the proof remaining unknown to us, but available to curious french-speakers in [1].

Remark 5.13. *In the literature it is generally assumed, though not proven that if*

$$\int \frac{f(z)}{\sqrt{D(z)}} dz$$

can be computed in elementary functions then it will take the form

$$\int \frac{f(z)}{\sqrt{D(z)}} dz = ln(p(z) + q(z)\sqrt{D(z)}).$$

From this we can see c) $\Rightarrow$ a) by computing

$$\begin{aligned} 0 &= \int \frac{f(z)}{\sqrt{D(z)}} dz - \int \frac{f(z)}{\sqrt{D(z)}} dz \\ &= \int \frac{f(z)}{\sqrt{D(z)}} dz + \int \frac{f(z)}{-\sqrt{D(z)}} dz \\ &= ln(p(z) + q(z)\sqrt{D(z)}) + ln(p(z) - q(z)\sqrt{D(z)}) \\ &= ln(p(z)^2 - D(z)q(z)^2). \end{aligned}$$

So $p(z)^2 - D(z)q(z)^2 = 1$.

Now for the proof:

Proof.

- $b) \Rightarrow a)$. Assuming $b)$, Proposition 3.3 of [10] gives us

$$\sqrt{D(z)} = [a_0, a_1, a_2, \ldots a_n, a_{n+1}]$$

Where

$$a_{n+1} = [2a_0, a_1, a_2, \ldots] = a_0 + \sqrt{D(z)}$$

So by equation (5.2), $\sqrt{D(z)} = \frac{a_{n+1}x_n + x_{n-1}}{a_{n+1}y_n + y_{n-1}}$ where $x_n, x_{n-1}, y_n, y_{n-1}$ represent the numerators and denominators respectively of the n-th and $n-1$-th convergents of $\sqrt{D(z)}$. Then substituting $a_0 + \sqrt{D(z)}$ for a_{n+1} yields

$$\sqrt{D(z)} = \frac{(a_0 + \sqrt{D(z)})x_n + x_{n-1}}{(a_0 + \sqrt{D(z)})y_n + y_{n-1}}$$

From which it follows that

$$\sqrt{D(z)}(a_0 + \sqrt{D(z)})y_n + \sqrt{D(z)}y_{n-1} = (a_0 + \sqrt{D(z)})x_n + x_{n-1}$$

and thus

$$D(z)y_n + (a_0 y_n + y_{n-1})\sqrt{D(z)} = (a_0 x_n + x_{n-1}) + x_n\sqrt{D(z)}.$$

But this means that

$$D(z)y_n = (a_0 x_n + x_{n-1})$$

and

$$a_0 y_n + y_{n-1} = x_n.$$

Solving for x_{n-1} and y_{n-1} gives

$$\begin{aligned} x_{n-1} &= D(z)y_n - a_0 x_n \\ y_{n-1} &= x_n - a_0 y_n \end{aligned}$$

We substitute these equations into equation (5.5) to get

$$x_n(x_n - a_0 y_n) - y_n(D(z)y_n - a_0 x_n) = (-1)^{n+1}$$

which we can rewrite as

$$x_n{}^2 - D(z)y_n{}^2 = (-1)^n.$$

If $n+1$ is even we are done. Otherwise, we use the second period in the expansion of $\sqrt{D(z)}$ instead of the first (so now when we repeat everything above we will get $x_{2n+1}{}^2 - D(z)y_{2n+1}{}^2 = (-1)^{2(n+1)}$.). This guarantees a solution to the Pell equation!

- $a) \Rightarrow c)$. Suppose $P(z)^2 - D(z)Q(z)^2 = 1$. We will show that

$$\int \frac{f(z)}{\sqrt{D(z)}}dz = \ln(P(z) + Q(z)\sqrt{D(z)})$$

where $f(z) = \frac{P'(z)}{Q(z)}$. First we must show that $f(z)$ is indeed a polynomial. Note that the Pell equation implies that P and Q are relatively prime. Further, taking derivatives of both sides of it yields

$$2P(z)P'(z) - 2D(z)Q(z)Q'(z) - D'(z)Q(z)^2 = 0.$$

Or

$$2P(z)P'(z) = Q(z)(2D(z)Q'(z) + D'(z)Q(z)).$$

So it must be the case that Q divides P' and that

$$f(z) = \frac{P'(z)}{Q(z)} = \frac{(2D(z)Q'(z) + D'(z)Q(z))}{2P(z)}.$$

Thus

$$\begin{aligned}
\frac{f(z)}{\sqrt{D(z)}} &= \frac{2D(z)Q'(z) + D'(z)Q(z)}{2P(z)\sqrt{D(z)}} \\
&= \frac{\sqrt{D(z)}Q'(z) + \frac{D'(z)Q(z)}{2\sqrt{D(z)}}}{P(z)} \\
&= \frac{\sqrt{D(z)}Q'(z) + \frac{D'(z)Q(z)}{2\sqrt{D(z)}}}{P(z)} \frac{(1 + \frac{Q\sqrt{D}}{P})}{(1 + \frac{Q\sqrt{D}}{P})} \\
&= \frac{\sqrt{D(z)}Q'(z) + \frac{D'(z)Q(z)}{2\sqrt{D(z)}} + \frac{2D(z)Q'(z)Q(z) + D'(z)Q(z)^2}{2P(z)}}{P(z) + Q(z)\sqrt{D(z)}} \\
&= \frac{\sqrt{D(z)}Q'(z) + \frac{D'(z)Q(z)}{2\sqrt{D(z)}} + P'(z)}{P(z) + Q(z)\sqrt{D(z)}} \\
&= \frac{(P(z) + Q(z)\sqrt{D(z)})'}{P(z) + Q(z)\sqrt{D(z)}}.
\end{aligned}$$

Lo and behold, from this it follows that

$$\int \frac{f(z)}{\sqrt{D(z)}}dz = \ln(P(z) + Q(z)\sqrt{D(z)}).$$

$\square$

Not only is this result surprising and interesting, but it helps if we find ourselves searching for things we can integrate or things with periodic continued fractions. We need only to find a polynomial, D, that satisfies the Pell Equation in order to find some weird integral that we can be sure has a solution in elementary functions, as was Abel's wont.

5.4. **Relationship to Plane Trees.** Combinatorial bicolored plane trees can help in an otherwise fruitless search for polynomials that satisfy one of the conditions of Theorem 5.12. Shabat and Zvonkin discuss how to do this in [7]. Suppose we are given a plane tree with corresponding Shabat polynomial, p. Consider the case where p has critical values -1 and 1. Say, without loss of generality, that p maps the white vertices of our tree to -1 and the black vertices to 1. Let the black vertices be given by $b_1, b_2, \ldots, b_n$ each with valency $\beta_1, \beta_2, \ldots, \beta_n$ and let the white vertices be given by $w_1, w_2, \ldots, w_m$ with corresponding valencies $\upsilon_1, \upsilon_2, \ldots, \upsilon_m$. Then

$$p(z) = \lambda \prod_i^n (z - b_i)^{\beta_i} - 1 = \lambda \prod_j^m (z - w_j)^{\upsilon_j} + 1.$$

So

$$p(z)^2 - 1 = (p(z) - 1)(p(z) + 1) = \lambda^2 \prod_i^n (z - b_i)^{\beta_i} \prod_j^m (z - w_j)^{\upsilon_j}. \tag{5.14}$$

We can think of this product though as the product of terms corresponding to vertices of odd valency, which all must be raised to odd powers, and terms corresponding to vertices of even valency, which all must be raised to even powers. Letting the product of terms corresponding to vertices of odd valency be equal to $D(z)$ and those of even valency be equal to $q(z)^2$, Equation (5.14) reads $p(z)^2 - 1 = D(z)q(z)^2$ so $p(z)^2 - D(z)q(z)^2 = 1$. So for a plane tree the "odd valency vertices" polynomial D satisfies the Pell equation and thus inherits all of the interesting properties implied by Theorem 5.12.

Remark 5.14. *If some of the odd valencies are greater than* 1*, we can get other solutions to the Pell equation by simply letting $D(z)$ be the square-free part of the right hand side of Equation (5.14).*

The natural thing to do now is to compute some continued fractions of these "$\sqrt{D}$"s. We start with the simplest tree, given by the pre-image of $[-1, 1]$ under $T_1(z) = z$.

The odd vertices are the end points -1 and 1 so

$$D(z) = (z + 1)(z - 1) = z^2 - 1.$$

Thus we'd like to compute the continued fraction of

$$\sqrt{D} = \sqrt{z^2 - 1}.$$

We'll do this in three ways.

5.4.1. *Method 1: Algebra and Divine Insight.* Knowing that the continued fraction expansion we seek is periodic, we can try to solve for it algebraically, in a way analogous to Example 5.9. In Example 5.9, we used a polynomial that had $1+\sqrt{2}$ as a zero to solve for the continued fraction expansion of this root. Following this lead and secretly already knowing the answer we seek, we will search for a quadratic with $z+\sqrt{z^2-1}$ and its algebraic conjugate as roots to compute the continued fraction expansion of $\sqrt{z^2-1}$. We have

$$(x-(z+\sqrt{z^2-1}))(x-(z-\sqrt{z^2-1})) = x^2-2xz+1.$$

Setting this equal to zero gives

$$-x(x-2z)=1$$

So

$$x = 2z+\frac{1}{-x}.$$

We can almost conclude that

$$z+\sqrt{z^2-1} = 2z+\cfrac{1}{-2z+\cfrac{1}{2z+\ddots}}$$

implying

$$\sqrt{z^2-1} = z+\cfrac{1}{-2z+\cfrac{1}{2z+\ddots}}.$$

But before we jump to such hasty conclusions, we may notice that plugging $z=\frac{1}{2}$ into this equation implies the astonishing equality

$$\frac{\sqrt{3}}{2}i = \frac{1}{2}+\cfrac{1}{-1+\cfrac{1}{1+\ddots}}.$$

This is concerning to say the least, as the left side is entirely imaginary, while the would-be convergents of the right side are all entirely real! It turns out that once again we need z sufficiently large for this continued fraction expansion to converge. Further, like in Example 5.9 we need to answer why we have obtained the continued fraction expansion of the root $z+\sqrt{z^2-1}$ and not its conjugate? The answer to both concerns, again like in Example 5.9, is provided by Theorem 5.10. If $|z|<1$, we have that $z+\sqrt{z^2-1}$ and its conjugate are complex conjugates and are thus equidistant from the $(n-1)$-st convergent, an entirely real number. If $|z|=1$, then we are in the case where $\alpha_1=\alpha_2=1$, and thus the first condition listed in Theorem 5.10 is satisfied. Finally, if $|z|>1$, then we have $\alpha_1 = z+\sqrt{z^2-1}, \alpha_2 = z-\sqrt{z^2-1}$ and

$C_{n-1} = C_1 = \frac{1-2z^2}{2z}$. We have then satisfied the second condition of the theorem, so we can conclude that the continued fraction converges to $z + \sqrt{z^2-1}$.

5.4.2. *Method 2: Laurent series.* We apply the algorithm for computing continued fractions of functions discussed above. Usually computations with Laurent series are messy, but in this case its not so bad, especially when we know what we're looking for. We first find that for $f(z) = \sqrt{z^2-1}$,

$$\begin{aligned} f(1/x) &= \sqrt{x^{-2}-1} \\ &= \sqrt{\frac{1-x^2}{x^2}} \\ &= \frac{1}{x}\sqrt{1-x^2} \\ &= \frac{1}{x}(1-\frac{x^2}{2}-\frac{x^4}{8}\ldots) \\ &= \frac{1}{x}-\frac{x}{2}-\frac{x^3}{8}+\ldots \end{aligned}$$

Changing the signs of the exponents, we find that the polynomial part we seek is $a_0 = z$. So $f(z) = z + \frac{1}{F_1(z)}$. We now want the polynomial part of $F_1(z) = \frac{1}{\sqrt{z^2-1}-z}$. Computing the Laurent series of $F_1(1/x)$ using a similar method as before we find that

$$F_1(1/x) = -\frac{2}{x} + \frac{x}{2} + \frac{x^3}{8} + \ldots.$$

So the polynomial part of $F_1(z)$ is $-2z$. Continuing, we have $f(z) = z + \frac{1}{-2z+\frac{1}{F_2(z)}}$. But here, when we solve $F_2(z) = \frac{1}{F_1(z)+2z}$ we find $F_2(z) = -F_1(z)$. We know that the polynomial part of $F_2(z)$ is therefore $2z$. We then observe that $F_3(z) = \frac{1}{F_1(z)+2z} = -F_2(z)$, and so on. This observation is equivalent to the algebra of method 1. Thus we have

$$\sqrt{z^2-1} = z + \cfrac{1}{-2z + \cfrac{1}{2z + \ddots}}.$$

5.4.3. *Method 3: Plane Trees and Luck.* Could we have done this without Laurent series or without knowing what we were looking for? By a fortunate accident in this case, yes. We'll need to use the following lemma from Van der Poorten.

Lemma 5.15. (Proposition 3.4 of [10]) *Suppose x, y, D are polynomials satisfying*

$$x(z)^2 - D(z)y(z)^2 = 1.$$

Then $\frac{x}{y}$ is a convergent of $\sqrt{D}$.

Now notice that in our case we have the simple plane tree given by the first Chebyshev polynomial, which has "odd vertex" polynomial D. But actually each of the plane trees corresponding to the Chebyshev polynomials has the same odd vertices (because the plane trees are "chains"), and thus each Chebyshev polynomial corresponds to the same D. But the reason this is relevant is because this means each Chebyshev polynomial, T_n and its corresponding "even vertex" polynomial, Q_n satisfy

$$T_n(z)^2 - D(z)Q_n(z)^2 = 1$$

So we are given an infinite number of convergents $\frac{T_n}{Q_n}$ for $\sqrt{D}$. Now if we are lucky enough to have obtained every convergent this way, then we can use equation (5.3) to compute the coefficients from the numerators of the convergents. Because the degrees of the Chebyshev polynomials are strictly increasing by 1 with every next polynomial, we may conclude that we do indeed have every convergent. Example 1.2 of [7] Chebyshev polynomials follow the relation

$$T_n = 2zT_{n-1} - T_{n-2},\ T_0 = 1, T_1 = z.$$

This observation in conjunction with Equation (5.3) provides yet another way to see that

$$\sqrt{z^2-1} = z + \cfrac{1}{-2z + \cfrac{1}{2z + \ddots}}.$$

5.4.4. *Generalization of Method 3.* Now generally, one may notice that Theorem 3.9 still holds if we impose the alternative stipulations that the critical values of P and Q are -1 and 1 (as opposed to 0 and 1), and that $P(-1), P(1) \in \{-1, 1\}$ (instead of $P(0), P(1) \in \{0, 1\}$). What's more, each of the Chebyshev polynomials, T_n, has the qualifications of such a P. This gives us a proposition.

Proposition 5.16. *Let $S(z)$ be a Shabat polynomial with $CritV(S) = \pm 1$ and "odd valency polynomial" $D(z)$. Then $\frac{X_n}{Y_n}$ is a convergent of the periodic continued fraction expansion of $\sqrt{D(z)}$ where X_n and Y_n are given by*

$$X_n(z) = T_n \circ S(z) = 2S(z)T_{n-1} \circ S(z) - T_{n-2} \circ S(z)$$

$$Y_n(z) = \sqrt{\frac{X_n^2 - 1}{D(z)}}.$$

Proof. We have seen the polynomial $D(z)$ must satisfy the Pell equation and therefore has a periodic continued fraction expansion by Theorem 5.12. Further, the polynomial $T_n \circ S$ is Shabat by Theorem 3.9 with the alternate conditions proposed above. The plane tree corresponding to $T_n \circ S$ amounts to substituting the tree corresponding to T_n in for each of the edges of S (See Sections 3.4 and 3.5 for an exposition of tree composition). But, as T_n is just a "chain tree" (i.e. one of the "chains" shown in figures 25 and 28, see Example 1.3 of [7]), this only adds verticies of even valency to the tree corresponding to S, not changing the odd valencies. The critical values of $T_n \circ S$ are still at ± 1 so by Lemma 5.15, $T_n \circ S(z)$ is another numerator of a convergent of $\sqrt{D(z)}$. For this convergent's denominator, the formula for Y_n can be obtained by solving the Pell equation. Further, the recursive formula for X_n is obtained by substituting $S(z)$ for z in the recursive formula for Chebyshev polynomials. $\square$

Unfortunately, these are not generally all of the convergents, and it is unclear to us at this time exactly which convergents can be obtained in this way. Thus, until a better understanding of these convergents is reached, Proposition 5.16 is of limited value.

References

[1] N.H. Abel. Sur l'integration de la formule differentielle $\frac{pdv}{\sqrt{R}}$, R et p etant des fonctions entieres. In *J. Reine Angew. Math.*, 1: 185-221,1826.

[2] James Ward Brown and Ruel V. Churchill. *Complex Variables and Applications, 8^{th} ed.* McGraw-Hill, New York. 2009.

[3] Martin Erickson and Anthony Vanazza. *Introduction to Number Theory.* Chapman and Hall, Boca Raton. 2007.

[4] Otto Forster. *Lectures on Riemann Surfaces.* Graduate Texts in Mathematics, New York. 1981.

[5] Allen Hatcher. *Algebraic Topology.* Cambridge University Press, 2002. Electronic Copy.

[6] S.K. Lando and A. K. Zvonkin. *Graphs on Surfaces and their Application.* Springer, Berlin Heidelberg. 2004.

[7] George Shabat and Alexander Zvonkin. Plane trees and algebraic numbers. In *Jerusalem combinatorics '93*, volume 178 of *Contemp. Math.*, pages 233–275. Amer. Math. Soc., Providence, RI, 1994.

[8] Ian Stewart and David Tall. *Complex Analysis: The Hitchhikers Guide to the Plane.* Cambridge University Press, New York. 1983.

[9] Tamás Szamuely. *Galois Groups and Fundamental Groups.* Cambridge University Press, New York. 2009.

[10] Alfred Van der Poorten and XC Tran. Quasi-Elliptic Integrals and Periodic Continued Fractions. In *Monatshefte fur Mathematik,* 131: 155-169, 2000.

[11] H.S. Wall *Analytic Theory of Continued Fractions, 8^{th} ed.* D. Van Nostrand, Princeton, NJ. 1948.