MONTEL'S SECOND THEOREM

ANTHONY LEONG

Abstract. The development of mathematics is somewhat similar to a tree. From an idea, another idea may be formed, and these ideas tend to separate as if they were the branches of a tree. In this case, the paper will first prove a theorem about the Montel's Second Theorem or the fundamental normality test. However, the two main ideas of the modular forms and monodromy will be greatly expanded upon in the upcoming sections.

1. Montel's Second Theorem

Montel's Second Theorem first appeared when trying to prove the Riemann Mapping Theorem. There had not yet been a proof constructed at that time. This section will examine the proof of Montel's Second Theorem. Montel's Second Theorem was created by a mathematician by the name of Paul Montel who introduced the notion of normality. This section will show the steps to proving Montel's Second Theorem.

Definition 1.1. A family of functions is normal if every sequence of functions in the family contains a subsequence of functions that converges uniformly on all compact subsets of Ω.

Theorem 1.2 (Montel's Theorem)**.** *If $\mathcal{F}$ is a family of analytic functions defined on an open set $\Omega \subset \mathbb{C}$, uniformly bounded on every compact set of Ω, then $\mathcal{F}$ is a normal family.*

Theorem 1.3 (Montel's Second theorem (Also known as the fundamental normality test))**.** *. Let $\Omega \subset \mathbb{C}$ be an open subset, and let $\mathcal{F}$ be a family of holomorphic functions in Ω whose range omits two values a and b. Then $\mathcal{F}$ is a normal family.*

Before we begin the proof, there are some reductions we have to make. We may assume that the two omitted values are 0 and 1 considering the fraction $\frac{f(z)-a}{b-a}$. If the two omitted values of f were not already 0 and 1, a linear map in the form of the fraction above may transport where the omitted values occur. We may assume that the domain Ω is the unit disk since normality is a local property. This means that the only significant points are in a nearby neighborhood, so the domain may be restricted to a smaller set. Finally, it suffices to prove normality of the smaller family $\mathcal{F}_1$ such that $\mathcal{F}_1 = \{f \in \mathcal{F} : |f(0)| \leq 1\}$ or if $f \notin \mathcal{F}_1$, then $\frac{1}{f} \in \mathcal{F}_1$ also due to the locality of normality.

Definition 1.4. The modular function λ maps the open upper half plane onto $\mathbb{C} \setminus \{0, 1\}$. Another property is that λ is invariant under the action of the congruence subgroup of the modular group, meaning $\lambda(\frac{az+b}{cz+d}) = \lambda(z)$ when a, b, c, and d are integers such that $ad-bc = 1$ and $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \pmod 2$.

As a reminder, the original definition of the modular function $\lambda(z)$ is the fraction $\frac{e_3-e_2}{e_1-e2}$ where $e_1 = \wp(\frac{\omega_1}{2}), e_2 = \wp(\frac{\omega_2}{2}), and e_3 = \wp(\frac{\omega_1+\omega_2}{2})$.

We are examining the lattice $\Lambda = \{m\omega_1 + n\omega_2 : m, n \in \mathbb{Z}$ for an elliptical function. Now, we may prove why the modular function is invariant under the congruent subgroup of the modular group.

For some lattice with the periods of ω_1 and ω_2, we might apply the modular group matrix, yielding a new lattice where the new periods are Ω_1 and Ω_2.

$$\begin{pmatrix} \Omega_1 \\ \Omega_2 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} = \begin{pmatrix} a\omega_1 + b\omega_2 \\ c\omega_1 + d\omega_2 \end{pmatrix} \text{ where } a, b, c, d \equiv 1, 0, 0, 1 (mod 2).$$

$\Omega_1 \equiv a\omega_1 + b\omega_2 (mod \Lambda)$
$\Omega_2 \equiv c\omega_1 + d\omega_2 (mod \Lambda)$.
Additionally,
$\frac{\Omega_1}{2} \equiv \frac{\omega_1}{2} (mod \Lambda)$
$\frac{\Omega_2}{2} \equiv \frac{\omega_2}{2} (mod \Lambda)$.
This means that the modular transformation does not change the lattice but only moves around the specific lattice points. Since e_1, e_2, e_3 are not changed, the modular λ function is not changed by the matrix.

Definition 1.5. (The Monodromy Theorem). If f is analytic in a disk contained in a simply connected domain D, and f can be analytically continued along every polygonal arc in D, then f can be analytically continued to a single-valued analytic function on all of D.

Definition 1.6. Alternate Monodromy Theorem. Let two continuous paths $\gamma(s)$,$0 \geq s \geq 1$ and $\delta(s)$,$0 \geq s \geq 1$ be given, which have the same end points $\gamma(0) = \delta(0)$, $\gamma(1) = \delta(1)$ and which are homotopic. This means is a map from the curve γ and δ given by a continuous function F(s,t), $0 \geq s, t \geq 1$ such that $F(s, 0) = \gamma(s)$, $F(s, 1) = \delta(s)$. Then analytic continuation along γ yields the same result as analytic continuation along δ.

This means that if we have a function f which admits analytic continuation over the whole domain, and the domain is simply connected, then f extends to a single valued function.

Lemma 1.7. *If f maps the unit disc into $\mathbb{C}\setminus\{0,1\}$, then $\lambda^{-1}f$ has a local branch (the issue of the possibility of two values has been eliminated) defined near f(0), and this function element admits unrestricted analytic continuation in the disc. This means that the function is still well-defined.*

Proof. By the monodromy theorem, there exist a function $\hat{f}$ from the unit disc into the upper half-plane such that $\lambda\hat{f} = f$. Remember that the modular function, λ can map the upper half plane to $\mathbb{C} \setminus \{0, 1\}$. The function of $\hat{f}$ can be analytically continued all across the disc which allows $\hat{f}$ to be defined at all points in the disc and not just a proper, open subset. ■

Proof. In the first part of the proof, we will examine the first subsequence. Suppose $\{f_n : n \in \mathbb{N}\}$ is a sequence of functions in the family $\mathcal{F}_1$. We need to prove a normally convergent subsequence.

Since the numbers $f_n(0)$ lie in a bounded set, there is a subsequence $\{n(k) : k \in \mathbb{N}\}$ such that the numbers $f_{n(k)}(0)$ converge to some complex number L.

Suppose first that $L \neq 0 or 1$. (These cases will be examined later.) We fix a branch of λ^{-1} in a neighborhood of L and use it to define the function $\hat{f_{n(k)}}$ consistently.

Next, we will look at the second subsequence. Since each $\hat{f_{n(k)}}$ has a range contained in the upper half-plane, the sequence $\{\hat{f_{n(k)}} : k \in \mathbb{N}\}$ is a normal family. Let $\{n(k(j)) : j \in \mathbb{N}\}$ be a subsequence such that the functions $\hat{f_{n(k(j))}}$ converge normally to a limit function g.

The range of the limit function g is certainly contained in the closed upper half plane. Since $g(0) = \lambda^{-1}(L)$, the open mapping principle implies that the range of g is contained in the open upper half-plane.

Consequently, λg is defined, and $f_{n(k(j))} = \lambda \hat{f_{n(k(j))}}$. $\lambda \hat{f_{n(k(j))}} \longrightarrow \lambda g$ when $j \longrightarrow \infty$. (This is it for the main case.)

Here is the proof for the edge cases when the limit is either 0 or 1. Suppose that $\lim_{k\longrightarrow\infty} f_{n(k)}(0) \longrightarrow 1$. Let h_k be a holomorphic square-root of the non-vanishing function $f_{n(k)}$ with the branch chosen such that $\lim_{k\longrightarrow\infty} h_k(0) = -1$. Clearly the range of each function h_k omits the values 0 and 1.

The preceding analysis applies to the sequence $\{h_k : k \in \mathbb{N}\}$ and shows that there is a normally convergent subsequence $\{h_{k(j)} : j \in \mathbb{N}\}$. Squaring shows that the sequence

$\{f_{n(k(j))} : j \in \mathbb{N}\}$ converges.

Finally, suppose $\lim_{k \longrightarrow \infty} f_{n(k)}(0) \longrightarrow 0$. The preceding case applies to the functions $1 - f_{n(k)}$.

We are now done. ■

The proof of Montel's Second Theorem has many similarities with the proof of Picard's Little Theorem. The monodromy theorem was able to take the place of the theory of covering spaces in Picard's Little Theorem. In fact, there is actually a proof of Picard's Little Theorem where the monodromy theorem is used. To outline that proof, we first define a function to be the fraction $\frac{f(z)-a}{b-a}$. Next, we take similar steps to define a composition with this fraction to be analytic. Finally, we will use the monodromy theorem to extend this function to be entire and then prove that it is bounded and hence constant using Liouville's Theorem.

To expand on the idea of analytical continuation, this idea is currently thought of through the lenses of sheaves. To give some history, the theory of sheaves were created after the theory of complex analysis. Leray invented this concept during World War II to analyze the topological obstructions to determine global solvability after local solvability is proven. A global analytic function can be constructed when the sheaf of germs are path-connected and maximal. We will not explore this topic further in the research paper. However, it can be a possible topic for further research later.

2. More on Modular Forms

We have only tapped the surface of the properties and applications of modular forms. We only examined the invariance of the SL subgroup on the modular function and how the modular form maps $\mathbb{H}$ to $\mathbb{C} \setminus \{0, 1\}$. There are many more important properties. In this section, we will define the congruence subgroup and the properties which are important to it. We will then discuss its connections to differential forms and how the orders of zeros and poles are related to each other.

Definition 2.1. For some $N \geq 0$, let us define

$$\Gamma(N) = \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} : a \equiv 1, b \equiv 0, c \equiv 0, d \equiv 1 (modN)\}$$

as the principal congruence subgroup of level N. Earlier in this paper, we examined the properties of this group of level 2 for the modular λ function.

One important property of this subgroup is that is it meromorphic as the cusps. A known property of modular functions is that $f(z) = f(z+1)$. We may rewrite any function which satisfies this property in the form $f(z) = f^*(q)$ where $q = e^{2\pi iz}$. Notice that as z goes around the upper half plane, q(z) ranges over a disc with the 0 removed. The term meromorphic at the cusp means that $f^*(q)$ is meromorphic on the full plane and can be rewritten as $f(z) = \sum_{n \geq -N_0} a_n q^n$.

Definition 2.2. A modular function f for Γ is a function on $\mathbb{H}$ satisfying these following conditions:

- f(z) is invariant under Γ, i.e., $f(\gamma z) = f(z)$ for all $\gamma \in \Gamma$;
- f(z) is meromorphic in $\mathbb{H}$;
- f(z) is meromorphic at the cusps.

Let the last condition be clarified. The cusp might be the point at infinity $i\infty$. A subgroup of $\Gamma(1)$ may be generated by $\begin{pmatrix} 1 & h \\ 0 & 1 \end{pmatrix}$ where $h \in \mathbb{N}$. Therefore, $f(z+h) = f(z)$ since f(z) is invariant under $\begin{pmatrix} 1 & h \\ 0 & 1 \end{pmatrix}$. f(z) may be expressed as a function $f^*(q)$ for a $q = e^{\frac{2\pi iz}{h}}$ on a punctured disc where $0 < |q| < \epsilon$. For f to be meromorphic at $i\infty$, f^* must be meromorphic at $q = 0$.

If the cusp is not at the point at infinity, it can be mapped there by a subgroup of $\Gamma(1)$. f is invariant at the subgroups, so the point of infinity will still be in the domain of f.

There is an analogous characteristic of holomorphism for modular function where the holomorphism at a neighborhood is guaranteed by the boundedness. f(z) is meromorphic at $i\infty$ if and only if for some $A > 0$, $e^{Aiz}f(z)$ is bounded as $z \longrightarrow i\infty$.

Definition 2.3. A modular form for Γ of weight 2k is a function on $\mathbb{H}$ such that:

- $f(\gamma z) = (cz+d)^{2k}f(z)$, for all $z \in \mathbb{H}$;
- f(z) is holomorphic in $\mathbb{H}$;
- f(z) is holomorphic at the cusps of Γ.

Next, let us examine the structure of the group $SL_2(\mathbb{C})$. The group $SL_2(\mathbb{C})$ acts on $\mathbb{C}^2$, and hence on the set $\mathbb{P}^1(\mathbb{C})$ of lines through the origin in $\mathbb{C}^2$. When we examine the slope of the line, $\mathbb{P}^1(\mathbb{C})$ becomes identified with $\mathbb{C} \cup \{\infty\}$.

Definition 2.4. $\begin{pmatrix} a & b \\ c & d \end{pmatrix}(z) = \frac{az+b}{cz+d}$, $\begin{pmatrix} a & b \\ c & d \end{pmatrix}(\infty) = \frac{a}{c}$.

These mappings are defined as linear fractional transformations of $\mathbb{P}^1(\mathbb{C}) = \mathbb{C} \cup \{\infty\}$

Some basic observations of this mapping is that it maps circles and lines into circles and lines. The identity is any matrix in the form $\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$.

Now, we will view the modular forms as k-form differentials. The reason we are doing this is because differential forms are a motivation for the idea of modular form outside of elliptical functions.

Definition 2.5. A differential form on an open subset U of $\mathbb{C}$ is an expression of the form f(z)dz where f is a meromorphic function on U. With any meromorphic function f(z) on U, we associate the differential form $df = \frac{df}{dz}dz$.

Example. Let $w : U \longrightarrow V$ be a mapping from U to another open subset V of $\mathbb{C}$. For any $z \in U$ and $z_0 \in V$ we can write it $z_0 = w(z)$. Let $\omega = f(z_0)dz_0$ be a differential form on V. Then $w^*(\omega)$ is the differential form $f(w(z))\frac{dw(z)}{dz}dz$ on U.

Let us examine how to construct a differential form on a Riemann surface.

Example. Let X be a Riemann surface, and let (U_i, z_i) be a coordinate covering of X. To give a differential form on X is to give differential forms $\omega_i = f(z_i)dz_i$ on $z_i(U_i)$ for each i that agree on overlaps in the following sense: let $z_i = w_{ij}(z_j)$, so that w_{ij} is the conformal mapping $z_i \circ z_j^{-1} : z_j(U_i \cap U_j) \longrightarrow z_i(U_i \cap U_j)$; then $w_{ij}^*(\omega_i) = \omega_j$. Overall, $f_j(z_j)dz_j = f_i(w_{ij}(z_j)){w_{ij}}'(z_j)dz_j$.

Proposition 2.6. *A meromorphic function f on a compact Riemann surface has the same number of poles as it has zeros (counting multiplicities). Additionally, let ω be a differential form on a compact Riemann surface; then the sum of the residues of ω at its poles is zero.*

Proof. In proving part b, we know that for some closed path C not passing through any poles, $\int_C f dz = 2\pi i \sum_{poles}(f;p)$ where p represents the poles. We may rewrite f dz as ω as a differential form. $\int_C \omega = 2\pi i \sum_{poles}(\omega;p)$. If we fix a finite covering (U_i, z_i) of the surface and fix a triangulation, then the integrals on the paths would cancel out since there cannot be any poles due to how the surface is bounded. For the first part, we can just define the form to be $\omega = \frac{df}{f}$. ∎

Definition 2.7. On a compact Riemann surface, X, the group of divisors Div(X) on X is the additive abelian group generated by the points on X; thus an element of Div(X) is a finite sum $\sum n_i P_i, n_i \in \mathbb{Z}$.

Definition 2.8. Let f be a nonzero meromorphic function on X. For any point $P \in X$, let $ord_P(f) = m, -m$, or 0 according as f has a zero of order m at P, a pole of order m at P, or neither a pole nor a zero at P.

Definition 2.9. $div(f) = \sum ord_p(f)P$.

We may attach a divisor to a differential form ω. let $P \in X$, and let (U_i, z_i) be a coordinate neighbourhood containing P. The differential form ω is described by a differential $f_i dz_i$ on U_i, and we set $ord_p(\omega) = ord_p(f_i)$. Then $ord_p(\omega)$ is independent of the choice of the coordinate neighbourhood U_i.

Now, let us consider what happens to a differential $\omega = f(z)dz$ on $\mathbb{H}$ for a meromorphic function f(z) under the action of Γ. Is it still ω invariant?
Let $\gamma = \frac{az+b}{cz+d}$. Then,

$$\gamma\omega = f(\gamma z)d(\tfrac{az+b}{cz+d}) = f(\gamma z)\tfrac{a(cz+d)-c(az+b)}{(cz+d)^2}dz = f(\gamma z)(cz+d)^{-2}dz.$$

Remember, $ad - cb = 1$.

In conclusion, ω is invariant if and only if f(z) is a meromorphic differential form of weight 2.

Remark 2.10. There is a notion of a k-fold differential form on a Riemann surface. Locally, it can be written $\omega = f(z)(dz)^k$, and if $w = w(z)$, then

$$w\omega = f(w(z))(dw(z))^k = f(w(z))(w\prime(z))^k(dz)^k.$$

Modular forms of weight 2k correspond to Γ-invariant k-fold differential forms on $\mathbb{H}^*$, and hence to meromorphic k-fold differential forms on $\Gamma \setminus \mathbb{H}^*$. $\mathbb{H}^*$ is $\mathbb{H} \cup \mathbb{P}^1(\mathbb{Q})$ where $\mathbb{P}^1(\mathbb{Q})$ is the set of cusps for $\Gamma(1)$.

A zero or pole of order m for $\omega = f(z)(dz)^k$ at $z = 0$ is dependent on whether f(z) has a zero or pole of order m at $z = 0$.

As the last analysis of the k-fold differentials, we will be identifying a couple of interesting properties about the orders at points that relate the modular forms and k-fold differential forms.

Theorem 2.11. *Let f be a (meromorphic) modular form of weight 2k, and let ω be the corresponding k-fold differential form on $\Gamma \setminus \mathbb{H}^*$. Let $Q \in \mathbb{H}^*$ map to $P \in \Gamma \setminus \mathbb{H}^*$.*

- *If Q is an elliptic point with multiplicity e, then $ord_Q(f) = eord_P(\omega) + k(e-1)$.*
- *If Q is a cusp, then $ord_Q(f) = ord_P(\omega) + k$.*
- *For the remaining points, $ord_Q(f) = ord_P(\omega)$.*

Proof. A point $z \in \mathbb{H}$ is called an elliptic point if it is the fixed point of an elliptic element γ of Γ where $\gamma \in SL_2(\mathbb{R})$. To prove the first case, let w be a map between the unit discs D such that $w : z \longrightarrow z^e$. Suppose that P and Q are both zero. Let f be a function on D and $f^* = f \circ w$. If f has a zero f order m, then f^* has a zero of order em. Thus, $ord_Q(f^*) = eord_P(f)$. Now, let us consider a k-fold differential form ω on D. Then $\omega = f(z)(dz)^k$ and $\omega^* = f(z^e)(dz^e)^k = f(z^e)(ez^{e-1}dz)^k = e^k f(z^e)z^{k(e-1)}(dz)^k$ for $\omega^* = w^*(\omega)$. f is isomorphic to ω^* as a result of the invariance under the action of γ.

For the second case, consider the map $q : H \longrightarrow (punctureddisk)$, $q(z) = e^{\frac{2\pi iz}{h}}$, and let $\omega^* = g(q)(dq)^k$ be a k-fold differential form on the punctured disk. Then $dq = (\frac{2\pi i}{h})qdz$, and so the inverse image of ω^* on $\mathbb{H}$ is

$$\omega = g(q(z))q(z)^k(dz)^k,$$

and so ω^* corresponds to the modular form $f(z) = g(q(z))q(z)^k$. Thus $f^*(q) = g(q)qk$. The zeros already at g(q) for f will have the zeros at q added to it for the k-fold differential form. And we are done!

The last case is trivial due to the isomorphism in the map between a modular form of weight 2k and corresponding differential form. ■

We have now finished analyzing the properties of generalized modular forms for the scope of this paper. There are other topics which may be researched further within modular forms. For example, more can be said about the impact of including the concept of the genus which is the number of holes. This can help define the dimension of the space of modular forms. There is also a method to give a more direct count of the number of zeros and poles on the differential form. This would involve results from the Riemann-Roch theorem which describes the number of functions there are on a compact surface given a number of zeros and poles. This is beyond the scope of the paper. However, it may be a topic of further research.

3. Monodromy and Group Theory

In this section, we will talk about the monodromy group. This concept arises from the idea of how multiple-valued functions can be rewritten as single-valued functions on Riemann surfaces. As a result of how the functions are defined on the specific surface, there may be a permutation group involved. This permutation group will be known as the permutation monodromy group and can be used to prove the Abel-Ruffini Theorem about the insolvability of a polynomial of degree 5 or greater.

Definition 3.1. So, what is a monodromy? A monodromy is the idea of how an object reacts when it circles a singularity.

For example, when a parametrized curve circles 0 in the square root function, something interesting happens. There may two values which are positive and negative.

Definition 3.2. Another way of thinking of monodromies is in terms of the function. A function is monodromic if it admits single values. A function may also be thought of polydromic if it admits multiple values. Although this is not that important for this section, it should be noted that many textbooks view monodromies this way.

The sets of monodromies may form a group which will be explored in this section.

Definition 3.3. Just a reminder, a group is a set which is closed under the binary operation. The properties of associativity, existence of an inverse, and existence of an identity are satisfied.

Definition 3.4. A symmetric group is a group under the operation of permutations (switching the position of elements). Let us define a bijective function $f[n] \longrightarrow [n]$ such that $[n] = \{1, 2, \ldots n\}$.

Before looking further into groups of monodromies, we must revisit the Riemann Surfaces.

First, we will define algebraic functions since the inspiration comes from functions with multiple values.

Definition 3.5. An algebraic function is defined as

$$f : \mathbb{C}^n \longrightarrow \mathbb{C}$$
$$(a_0, \ldots a_n) \longrightarrow \{z : p(a_0, \ldots a_n, z) = a_0 z^n + a_1 z^{n-1} + \ldots + a_n = 0\}.$$

Notice that this algebraic function extracts the roots from a polynomial when given the coefficient. So how is this relevant to our analysis of monodromies? Let us examine the algebraic function for the polynomial $p(z) = z^2 - a$. The functions that helps find the z when $p(z) = 0$ are $f(a) = \sqrt{a}$ or $f(a) = -\sqrt{a}$.

For a visual representation on a complex plane. Let us define a loop $a(t)$ such that $t \in [0, 1]$. Let the starting point be $a_0 \in \mathbb{C} \setminus 0$. When we apply our functions, we find that $\sqrt{a(1)} = -\sqrt{a(0)}$ and $-\sqrt{a(1)} = \sqrt{a(0)}$.

We may now create our Riemann Surfaces. For the two branches of f(a), we may define two parametrized curves as $f_1(a) = \sqrt{r}e^{\frac{i\theta}{2}}$ and $f_2(a) = \sqrt{r}e^{\frac{i\theta}{2}+\pi}$. Now, we may view f(a) to be defined two different sheets $\mathbb{C} \setminus 0$ and not just the complex plane. We know that the sheets must be abstractly glued together where the branch cut occurs. But how do we find branch points?

The square root function has a branch point due to its multiple values; however, all radical functions are bound to have singularities. Singularities differ from branch points because they may also include non-uniqueness points.

Example. The surface for $\sqrt{a^2}$. This function has two single valued functions a and $-a$. Both functions actually return the same value when passing through 0. When going around 0, the argument of a varies by 2π. Therefore, the argument of a^2 varies by 4π. However, the argument of $\sqrt{a^2}$ varies by 2π again. The two sheets would only connect at the one point 0, making it unnecessary to have two different sheets.

Is there a way to more easily tell whether a singularity is a branch point or a non-uniqueness point? There is!

Definition 3.6. For a curve given by $a(t) : [0, 1] \longrightarrow \mathbb{C} \setminus \{0\}$, define $\theta(t)$ to be the argument of the curve at a time t. This function will be continuous since the loop is continuous with respect to the variable t. The variation of the argument is the angle that the loop travels with respect to the singularity or more formally, $\theta(1) - \theta(0)$.

Definition 3.7. The winding number is the variation divided by 2π. The winding number counts how many times the loop travels around the singularity.

Example. Let us examine the winding number of the square root function. Recall that our parametric curves were defined as $f_1(a) = \sqrt{r}e^{\frac{i\theta}{2}}$ and $f_2(a) = \sqrt{r}e^{\frac{i\theta}{2}+\pi}$. The argument of f(a) is half the argument of a(t). The variation of the function of the curve is also only half of the variation of the curve. So, the image must have an twice the variation of the curve to be considered closed; therefore, the winding number of the image of a curve has to be even.

If the variance of the loop is not a multiple of 2π after wrapping and the images at the starting and ending points are not equal, then there is a branch point. For example, this happens with the square root function since the variance may just be π. This will cause the image to be additive inverses of each other around 0 which we already know is a branch point.

Definition 3.8. Let f(a) be a function associated with the polynomial p(a,z).
The Riemann surface is defined as M of f(a) along the map $\phi : \{(a, z)|p(a, z) = 0\} \longrightarrow \{a\}$, as a covering space of $\mathbb{C} \setminus \{$singular points of f(a)$\}$.

Let use construct a couple of surfaces as examples.

Example. $f(a) = \sqrt[n]{(a - a_0)^{i_0} \dots (a - a_m)^{i_m}}$ where $n, m \in \mathbb{N}$ and $i_0, \dots i_m \in \mathbb{Z}$. Each a_j will be either branch points or non-uniqueness points if we separate the function onto n sheets. To distinguish between branch points and non-uniqueness points, we take a loop around the point in question. It will be a branch point if the difference of the function between the ending an starting point is not a multiple of 2π. Each function can be labeled as f_i where f_1 is chosen to be a function, and each $f_i = e^{\frac{2\pi(i-1)}{n}} f_1$.

Under the field operations like multiplication and addition, two functions with Riemann sheets can be composed. The branches of the composed function will just be the composition of the branches of the two functions for all the possible combinations. Let us define the functions as f and g. The branches of f are $\{f_1, f_2, \dots f_m\}$. The branches of g are $\{g_1, g_2, \dots g_n\}$. The branches of h if $h = f * g$ where $*$ is some field operation are $\{f_1 * g_1, \dots f_1 * g_n, \dots f_m * g_1, \dots f_m * g_n\}$.

Example. $g(a) = f(a)^n$. g(a) has the same amount of sheets as f(a) except that the branches are squared. The connection points of the sheets are the same.

Example. $g(a) = \sqrt[n]{f(a)}$ where we already know the Riemann surface of f(a). The single valued functions from f differ by a multiplication by a root of unity, so maximal amount of single-valued functions is n times the amount from f.

We may now associate the different sheets to the permutation group. What this means is that the different sheets informally switch places with each other on the function. The permutations form a group based on the logic for why a symmetric group is a group.

Here is a way to visualize the permutations mathematically. Take a point a_0 which is not a singularity (a branch or a non-uniqueness point). Next, define a continuous loop L given by a(t). The starting point is given by $a(0) = a_0$. The function $f(a(0)) = f_i$ can move across the loop L to the value $f_j = f(a(1))$. The different values of f_j can be determined by the different values of f_i. Therefore, there is a permutation between the functions.

Definition 3.9. We may define two groups as the permutations of the values of $f(a_0)$ and the permutations of $f(a_1)$. These are both considered the permutation monodromy group.

Despite the abstractness within this structure, there are some applications. We can use the ideas to solve cubic and quartic equations and prove the a polynomial that is a quintic or of higher degree is unsolvable. We will only prove that there exists a monodromy group for cubics and that this is actually impossible for quintics or greater.

Let us set up a cubic equation. $p(z) = z^3 + \alpha z^2 + \beta z + a$ where f(a) is the algebraic function which is associated with it. The branch points are where the double roots occur. This is because f(a) admits less than 3 values.

So, how do we figure out where the double roots are? We can take advantage of the derivative. We know that the geometric representation of a cubic equation has at most two points where the derivative is zero. We can only intersect the y-axis twice since one root is single and the other is double. When taking the derivative, we must have two distinct roots; therefore, $\alpha^2 - 3\beta$ cannot be 0.

f(a) will admit two values at $a = a_1 = -((z_1)^3 + \alpha(z_1)^2 + \beta(z_1))$ and at $a = a_2 = -((z_2)^3 + \alpha(z_2)^2 + \beta(z_2))$ for the roots z_1 and z_2.

Now, we must prove that we can construct a Riemann surface since this will determine whether the monodromy group exists for the cubic. We will construct the Riemann surface with the branch points a_1 and a_2 where a function f(a) is continuous (being a function implies that it is single-valued).

Let us define a curve that does not go through a_1 and a_2. Since f is continuous, and the curve is connected, the image of the curve under f will also be connected. Therefore, f will be a continuous function on this Riemann surface. There are 3 roots with two possible transpositions (permutations which switch the places of the branch points since each point connects two Riemann surfaces). The permutation monodromy group will be congruent to S_3.

Although we will not prove it in this paper, the permutation monodromy group for quartic polynomials will be congruent to S_4.

To determine whether the quintic or a polynomial of a degree greater than the quintic is solvable, we will have to prove two lemma.

Definition 3.10. A group is abelian if it satisfies the commutative property.

Definition 3.11. A group is solvable if it can be constructed from existing abelian groups.

Lemma 3.12. *Let f(a),g(a) be two algebraic functions with abelian permutation monodromy groups F and G. Then $h(a) = f(a) * g(a)$, where the operation $*$ is a field operation, has an abelian permutation monodromy group.*

Proof. To prove this lemma, observe that the monodromy group H_1 of the Riemann surface of h(a) before merging equal sheets, is isomorphic to a subgroup of $F \times G$. This is because each surface of h corresponds with a combination of the surfaces in f and g as shown in a previous example. Next, observe that there exists a surjective homomorphism from H_1 to H_2 where H_2 is the monodromy group associated to the Riemann surface with merged sheets. Hence because Fas well as H_1 and H_2.

■

The permutation monodromy group of an algebraic function of the form $h(a) = \sqrt[n]{(aa_0)^{i_0} \dots (aa_m)^{i_m}}$ is always abelian. To each branch point, there corresponds a permutation of all the sheets of the form

$$h_1 \longrightarrow h_2 \longrightarrow \dots \longrightarrow h_n \longrightarrow h_1$$

This means that all the branch points are associated to the same permutation and the monodromy group is cyclic.

In conclusion, a non-abelian permutation monodromy group implies the algebraic function involves nesting of roots. S_3 and S_4 are actually non-abelian.

Lemma 3.13. *If a multivalued function h(x) is representable by radicals, its permutation monodromy group is solvable.*

Proof. We will want to prove that given f(a) with solvable permutation monodromy groups, the monodromy group of $\sqrt[n]{f(a)}$ is also solvable.

Let F the monodromy group of f(a) and H be the monodromy group of $\sqrt[n]{f(a)}$. For every sheet of the Riemann surface for f(a), there are n-sheets in the Riemann surface for $\sqrt[n]{f(a)}$. If we go around a branchpoint of $\sqrt[n]{f(a)}$, then the n sheets are moved to another set of the n sets of sheets. As a result, the sets of sheets are preserved under the permutations of H. We may define a permutation group of the packs as Γ of H. Γ is a surjective homomorphism and its image is an isomorphism with F. The kernel (the elements which map to the identity) of Γ is the permutations in H which transform each set of n sheets back onto itself. As a result, if two elements are within the kernel, they are commutative. The kernel is abelian.

We know that the quotient group of the H over the kernel of Γ is isomorphic to F. The kernel is abelian, and F is solvable; therefore, H is solvable. ■

Now, we may move onto the Abel-Ruffini Theorem which is where we will prove that a polynomial with a degree of 5 or greater is unsolvable.

Theorem 3.14. *Given $n \geq 5$, the general algebraic equation of degree n*

$$a_0 z^n + a_1 z^{n-1} + \ldots + a_n = 0$$

is not solvable by radicals.

Proof. We only need to prove that a polynomial of degree 5 is unsolvable. We know that the monodromy group will be S_5 by using the previous methods for the quartic and cubic polynomials. This cannot be represented by radicals due to the previous lemma. ■

Now, we have seen an application of the monodromy group. In this case, it is a substitute for the Galois groups in proving the unsolvability of the quintic. In fact, there is actually a theorem which proves that the monodromy group and Galois group are identical.

There are a couple notable expansions for the concept of the monodromy group which has not been talked about in this section. One example is the braid group. The way braid groups would be viewed in this case is as particles dancing through time. Another notable property which has not been mentioned in this paper is the topological construction of the monodromy group through the actions of a fundamental group and a covering map between topological spaces.

4. Conclusion

From a simple theorem, mathematics about different tools used in the proof were expanded upon. One main idea that is a great possibility for further research is the generalized idea of normality. Overall, the goal of this paper was to learn about the different possibilities and implications of Montel‘s Second Theorem and possibly open new doors for further exploration of the concepts discussed.

References

[1] Baran Can Oner, LATEX*Galois Theory, Monodromy Groups and Flexes of Plane Cubic Curves* July 2012, ULR: http://www.math.uni-bonn.de/people/huybrech/Oenerfinal.pdf Pages 5-7

[2] J. S. Milne, LATEX*Modular Functions and Modular Forms* 1990, ULR: https://www.jmilne.org/math/CourseNotes/MF110.pdf Pages 1-49

[3] Hannah Santa Cruz, LATEX*A Survey on the Monodromy Groups of Algebraic Functions* 2016, ULR: http://math.uchicago.edu/ may/REU2016/REUPapers/SantaCruz.pdf Pages 1-15

[4] Fabrizio Catanese, LATEX*Monodromy and Normal Forms* July 2015, Bulletin (New Series) of the American Mathematical Society, ULR: https://arxiv.org/pdf/1507.00711.pdf Pages 5-10

[5] Harold P. Boas, LATEX*Montel's theorem* May 2004, ULR: https://www.math.tamu.edu/ boas/courses/618-2004a/may04.pdf Pages 1-5

[6] Daniel Milanovich, LATEX*Modular Functions and Picard's Little Theorem* ULR: https://sites.math.washington.edu/ morrow/$336_19/papers19/Daniel.pdf Pages1-10$