SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

Permutations of Roots of Complex Polynomials

av

Henrik Treadup

2010 - No 14

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET, 106 91 STOCKHOLM

Permutations of Roots of Complex Polynomials

Henrik Treadup

Självständigt arbete i matematik 30 högskolepoäng, grundnivå

Handledare: Torsten Ekedahl

2010

Abstract

A complex bivariate polynomial can be viewed as a continuous family of complex polynomials. If the parameter is moved along a continuous curve the roots of the generated polynomial will move along continuous curves. If the parameter is moved along a closed curve then each root will end up where it started except in the case when the curve goes around certain critical points. In this case the roots can swap places and the curve will generate a permutation of the roots.

The Predict Correct Algorithm can be used to numerically follow roots of the generated polynomial as the parameter is moved along a curve. A problem that can occur with the Predict Correct Algorithm is that the algorithm will jump and start following the wrong root. In this paper a modified version of the Predict Correct Algorithm is developed that guarantees that no root jumping occurs. The new algorithm is called the Predict Correct Verify Algorithm. An algorithm for calculating the critical points of a bivariate polynomial is presented.

An algorithm for automatically calculating all the permutations of the roots generated by a bivariate polynomial is developed. A program implementing the algorithm is written using the Scheme programming language.

Contents

1 Introduction

1.1 Purpose and Scope

A complex bivariate polynomial $f(z,w)$ can be written as

$$f(z,w) = \sum_{k=0}^{n} c_k(w) z^k, \tag{1.1}$$

where $c_1(w), c_2(w), \ldots, c_n(w)$ are complex polynomials in w. Hence a complex bivariate polynomial $f(z,w)$ can be viewed as a family of complex polynomials parameterized by w.

Each point w in the complex plane generates a polynomial. The generated polynomial will have roots. When the parameter w is moved in the complex plane the roots of the generated polynomial will also move. In fact a simple root will vary continuously with the parameter w.

If the parameter w is moved along a closed curve then the set of roots at the start of the path is equal to the set of roots at the end of the path. If while moving along a closed curve the parameter w goes around so called critical points then the roots can swap places. Otherwise each root will end up back where it started. It is possible to use the discriminant to calculate the critical points generated by a bivariate polynomial.

Given a bivariate polynomial $f(z,w)$ a curve can be thought of as a bijection between the roots at the start of the curve and the roots at the end of the curve generated by following the roots of $f(z,w)$ as the parameter w moves along the curve. The curve can be thought of as performing an action on the roots at the start of the curve. If the curve is closed then the generated bijection can be viewed as a permutation of the roots at the start of the curve.

Suppose that $c_1, c_2, \ldots, c_n$ are the critical points of $f(z,w)$. To determine the permutation generated by an arbitrary closed path starting at s it is enough to know the permutation generated by n paths starting at s where each path goes around a single distinct critical point of $f(z,w)$.

One approach for tracking the roots of a polynomial as the coefficients of the polynomial change continuously is to use Homotopy Continuation methods. Using the Predict and Correct algorithm it is possible to follow a root of $f(z,w)$ numerically as the parameter w is moved along a path. However there is no guarantee that the Predict and Correct Algorithm will produce the correct results. It is possible for the algorithm to jump and start following an incorrect root. For more information on Homotopy Continuation and the Predict and Correct algorithm see Sommese and Wampler [8] and Morgan [3].

The purpose of this paper is to create and implement a numerical algorithm for calculating the permutations of the roots generated by a bivariate polynomial $f(z,w)$ when the parameter w is moved around critical points of $f(z,w)$. This will involve developing an algorithm for following a root of $f(z,w)$ when the parameter w is moved along a closed path that produces a provably correct result. A restriction is that the path is made up of line segments. An algorithm for calculating the critical points of a bivariate polynomial will be presented. Some care is taken to make sure that the computed results are provably correct.

The implementation is written in the Scheme programming language. The program consists of the following steps.

1. Generate a random bivariate polynomial $f(z, w)$.

2. Calculate the critical points of the bivariate polynomial.

3. Generate linear approximations of circle paths paths that all have a common starting point and where each circle path goes around one and only one of the critical points of $f(z, w)$.

4. Calculate the initial roots of $f(z, w)$ at the common starting point of the paths.

5. For each path calculate the generated permutation of the roots.

1.2 Disposition

This paper is structured in the following manner. Section 2 introduces the reader to Complex Analysis. Several definitions and theorems are given leading up to Rouché's Theorem.

In Section 3 several algorithms for finding the roots of a polynomial numerically are given. An algorithm that can find all the roots of a polynomial is developed.

In Section 4 the concept of a family of polynomials generated by a bivariate polynomial is introduced. Several theorems regarding polynomials are stated and proved. The concept of a critical point of a bivariate polynomial is introduced and a numerical algorithm for calculating the critical points of a bivariate polynomial is given.

In Section 5 the concept of a path and a homotopy are introduced. Circle paths, N-gon paths and triangle paths are defined. The fact that an arbitrary path is homotopic to a path that is the composition of circle paths and inverse circle paths is stated and proved.

In Section 6 the Complex Implicit Function Theorem is introduced. The concept of analytic continuation is described. The Monodromy Theorem is stated and proved.

Section 7 is concerned with numerical algorithms for following the simple roots of a bivariate polynomial $f(z, w)$ when the parameter w is moved along a path. The Davidenko Differential Equation is described. The Predict Correct Algorithm is introduced. Euler Prediction and Newton Correction are described. The root jumping issue is described.

A new algorithm called the Predict Correct Verify Algorithm is developed. Euler Disc Prediction and Newton Rouché correction are described. A sufficient condition for no root jumping to occur is developed. A new algorithm called Rouché Verification is introduced.

In Section 8 a function that permutes the roots of a bivariate polynomial $f(z, w)$ when the parameter w is moved along a closed path is defined. Some basic properties about the permutation function are stated and proved. The fact that the permutations generated by two homotopic paths are equal to each other is proved.

An algorithm for automatically generating circle paths such that all circle paths have a common starting point and each circle path goes around one and only one critical point of a bivariate polynomial is described. An algorithm for calculating permutations of the roots generated by circle paths going around each of the critical points of a bivariate polynomial is described.

In Section 9 a program that implements some of the algorithms from the previous sections is discussed. The program consists of the following steps. The program generates a random bivariate polynomial $f(z, w)$. The critical points of the bivariate polynomial are calculated. The program calculates a list of paths that all share a common starting point and where each path goes around a single critical point. The program calculates how the roots at the start of the paths are permuted when the parameter w is moved along each of the paths.

2 Complex Analysis

The core concepts in real analysis are limits of real functions, continuity of real functions, the derivative of a real function and the integral of a real function. It is possible to define limits, continuity, derivatives and integrals of complex functions in such a manner that a lot of the theorems from real analysis are still valid for complex functions. In this section results from complex analysis that will be needed later on are presented. Most of the definitions and theorems in this section are based on definitions and theorems presented by Ablowitz and Fokas (see [1]) and Silverman (see [7]).

2.1 Limit, Continuity and Differentiability

The concept of a limit of a complex function is defined as follows.

Definition 1. *The* limit *of the complex function $f(z)$ at the point z_0 is equal to c if given ϵ there exists δ such that if $|z - z_0| < \delta$ then $|f(z) - c| < \epsilon$. The limit of $f(z)$ at the point z_0 is denoted by*

$$\lim_{z \to z_0} f(z).$$

The concept of continuity for a complex function is defined as follows.

Definition 2. *The complex function $f(z)$ is* continuous *at a point z_0 if*

$$\lim_{z \to z_0} f(z) = f(z_0).$$

A complex function $f(z)$ is continuous in a region *if it is continuous at every point in the region.*

The concept of differentiability of a complex function is defined as follows.

Definition 3. *The* derivative *of a complex function $f(z)$ at the point z_0 is defined as*

$$\lim_{h \to 0} \frac{f(z_0 + h) - f(z_0)}{h}.$$

If the limit does not exist the the derivative is undefined. The derivative of $f(z)$ at the point z_0 is denoted by $f'(z_0)$ or $\frac{df}{dz}(z_0)$. If the function $f(z)$ has a derivative at the point z_0 then $f(z)$ is said to be differentiable *at z_0.*

The reader should not be fooled by the fact that this definition looks very similar to the definition of the derivative of a real function. The existence of a complex derivative of a function is a much stronger statement than the existence of a real derivative.

2.2 Analytic Function

The next concept to be introduced is that of an analytic function. This is the core concept of complex analysis.

Definition 4. *A complex function $f(z)$ is said to be* analytic *at the point z_0 if $f(z)$ is differentiable in a neighborhood of z_0. A complex function is said to be* analytic in a region *if the derivative of the function exists at each point in the region.*

Just as for Real Analysis one can show that the sum, product and quotient of two analytic functions is analytic.

Lemma 1. *Let $f(z)$ and $g(z)$ be two complex functions that are differentiable at z_0. Then $f(z)+g(z)$ is differentiable at z_0 and the derivative is given by*

$$(f(z)+g(z))' = f'(z)+g'(z). \tag{2.1}$$

Proof.

$$\frac{d}{dz}(f(z)+g(z)) = \lim_{h\to 0}\frac{f(z+h)+g(z+h)-f(z)-g(z)}{h} \tag{2.2}$$

$$= \lim_{h\to 0}\frac{f(z+h)-f(z)}{h} + \lim_{h\to 0}\frac{g(z+h)-g(z)}{h} \tag{2.3}$$

$$= f'(z)+g'(z). \tag{2.4}$$

□

Lemma 2. *Let $f(z)$ and $g(z)$ be two complex functions that are differentiable at z_0. Then $f(z)g(z)$ is differentiable at z_0 and the derivative is given by*

$$(f(z)g(z))' = f'(z)g(z)+f(z)g'(z). \tag{2.5}$$

Proof.

$$\frac{d}{dz}(f(z)g(z)) = \lim_{h\to 0}\frac{f(z+h)g(z+h)-f(z)g(z)}{h} \tag{2.6}$$

$$= \lim_{h\to 0}\frac{f(z+h)g(z+h)-f(z)g(z+h)}{h} + \tag{2.7}$$

$$\lim_{h\to 0}\frac{f(z)g(z+h)-f(z)g(z)}{h} \tag{2.8}$$

$$= f'(z)g(z)+f(z)g'(z) \tag{2.9}$$

□

Lemma 3. *Let $f(z)$ and $g(z)$ be two complex functions that are differentiable at z_0 where $g(z)$ satisfies the condition that $g(z_0)\neq 0$. Then $f(z)/g(z)$ is differentiable at z_0 and the derivative is given by*

$$(f(z)/g(z))' = \frac{f'(z)g(z)-f(z)g'(z)}{g^2(z)}. \tag{2.10}$$

Proof.

$$\frac{d}{dz}(f(z)g(z)) = \lim_{h\to 0}\frac{f(z+h)/g(z+h)-f(z)/g(z)}{h} \tag{2.11}$$

$$= \lim_{h\to 0}\frac{f(z+h)/g(z+h)-f(z)/g(z+h)}{h} + \tag{2.12}$$

$$\lim_{h\to 0}\frac{f(z)/g(z+h)-f(z)/g(z)}{h} \tag{2.13}$$

$$= f'(z)/g(z)+f(z)\lim_{h\to 0}\frac{-1}{g(z+h)g(z)}\frac{g(z+h)-g(z)}{h} \tag{2.14}$$

$$= f'(z)/g(z)-\frac{f(z)g'(z)}{g^2(z)} \tag{2.15}$$

$$= \frac{f'(z)g(z)-f(z)g'(z)}{g^2(z)} \tag{2.16}$$

□

2.3 Cauchy-Riemann Equations

A complex function $f(z)$ can be written as

$$u(x, y) + iv(x, y),$$

where $z = x+iy$ and the functions $u(x, y)$ and $v(z, y)$ are the real and imaginary components of the complex function $f(z)$. If the function $f(z)$ is differentiable at a point then there is a relationship between the partial derivatives of $u(x, y)$ and $v(x, y)$ at that same point.

Cauchy-Riemann Equations. *If the complex function*

$$f(z) = u(x, y) + iv(x, y),$$

is differentiable at $z = x + iy$ *then*

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \tag{2.17}$$

$$\frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}. \tag{2.18}$$

Proof. Let $h = \triangle x$ where $\triangle x \in \mathbb{R}$. Then

$$\begin{aligned}
f'(z) &= \lim_{h\to 0} \frac{f(z+h) - f(z)}{h} && (2.19)\\
&= \lim_{\triangle x\to 0} \frac{f(z+\triangle x) - f(z)}{\triangle x} && (2.20)\\
&= \lim_{\triangle x\to 0} \frac{u(x+\triangle x, y) + iv(x+\triangle x, y) - u(x, y) - iv(x, y)}{\triangle x} && (2.21)\\
&= \lim_{\triangle x\to 0} \frac{u(x+\triangle x, y) - u(x, y)}{\triangle x} + \lim_{\triangle x\to 0} \frac{iv(x+\triangle x, y) - iv(x, y)}{\triangle x} && (2.22)\\
&= \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}. && (2.23)
\end{aligned}$$

Let $h = i\triangle y$ where $\triangle y \in \mathbb{R}$. Then

$$\begin{aligned}
f'(z) &= \lim_{h\to 0} \frac{f(z+h) - f(z)}{h} && (2.24)\\
&= \lim_{\triangle y\to 0} \frac{f(z+\triangle y) - f(z)}{i\triangle y} && (2.25)\\
&= \lim_{\triangle y\to 0} \frac{u(x, y+\triangle y) + iv(x, y+\triangle y) - u(x, y) - iv(x, y)}{i\triangle x} && (2.26)\\
&= \lim_{\triangle y\to 0} \frac{v(x, y+\triangle y) - v(x, y)}{\triangle y} - i \lim_{\triangle y\to 0} \frac{u(x, y+\triangle y) - u(x, y)}{\triangle y} && (2.27)\\
&= \frac{\partial v}{\partial y} - i\frac{\partial u}{\partial y}. && (2.28)
\end{aligned}$$

The theorem now follows from the fact that the real part of (2.23) is equal to the real part of (2.28) and the imaginary part of (2.23) is equal to the imaginary part of (2.28). □

2.4 Curves

An important concept in complex analysis is that of a curve in the complex plane.

Definition 5. *A* curve *is a complex valued function $z(t)$ that is defined on a real interval $[a, b]$. Let $x(t)$ and $y(t)$ be the real and imaginary components of $z(t)$. Then $z(t)$ can be written as*

$$z(t) = x(t) + iy(t).$$

The curve $z(t)$ is said to be continuous *if $x(t)$ and $y(t)$ are continuous functions. The curve $z(t)$ is said to be* piecewise continuous *if $x(t)$ and $y(t)$ are piecewise continuous functions. The curve $z(t)$ is said to be* differentiable *if $x(t)$ and $y(t)$ are differentiable functions. The* derivative *of the curve $z(t)$ is defined as*

$$z'(t) = x'(t) + iy'(t).$$

Next let us define some useful properties of curves.

Definition 6. *A curve $z(t) : [a, b] \to \mathbb{C}$ is* closed *if $z(a) = z(b)$.*

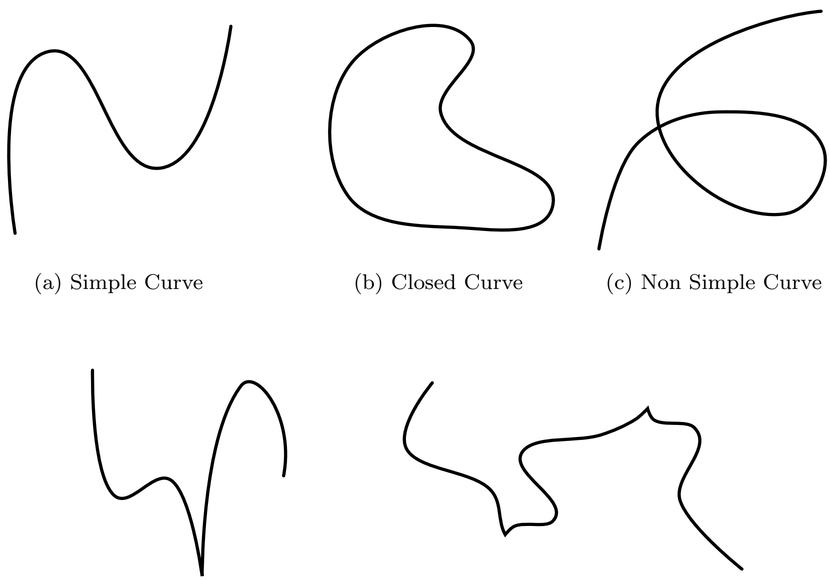
Definition 7. *A curve $z(t) : [a, b] \to \mathbb{C}$ is* simple *if for $t_0, t_2 \in [a, b]$*

$$z(t_0) = z(t_1) \implies t_0 = t_1,$$

with the exception that $z(a) = z(b)$ is allowed.

Note that with this definition a closed curve can be simple.

Definition 8. *A curve $z(t) : [a, b] \to \mathbb{C}$ is* smooth *if z(t) is continuous and $z'(t)$ is piecewise continuous. A curve is* piecewise smooth *if it can be split into a finite number of smooth pieces.*



(a) Simple Curve (b) Closed Curve (c) Non Simple Curve

(d) Non Smooth Curve (e) Piecewise Smooth Curve

Figure 1: Different kinds of curves.

2.5 Integrals

The next concept to be defined is that of an integral of a complex function.

Definition 9. *Let $f(t)$ be a complex valued function that is defined on the real interval $[a, b]$. Let $u(t)$ and $v(t)$ be the real and imaginary components of $f(t)$. Then $f(t)$ can be written as*

$$f(t) = u(t) + iv(t). \tag{2.29}$$

The function $f(t)$ is said to be integrable *if the following real integrals exist:*

$$\int_a^b u(t)\, dt \tag{2.30}$$

$$\int_a^b v(t)\, dt \tag{2.31}$$

If the function $f(t)$ is integrable then the integral *of the function $f(t)$ on the interval $[a, b]$ is defined as*

$$\int_a^b f(t)\, dt = \int_a^b u(t)\, dt + i \int_a^b v(t)\, du. \tag{2.32}$$

The above definition handles integrals along line segments on the real axis. The definition can be extended to handle integrals along smooth curves in the complex plane. To do this the concept of continuity of a function along a curve is needed.

Definition 10. *Let $u(z)$ and $v(z)$ be the real and imaginary components of $f(z)$. The function $f(z)$ is said to be* continuous on the curve *$z(t)$ if $u(z(t))$ and $v(z(t))$ are continuous. The function $f(z)$ is said to be* piecewise continuous on the curve *$z(t)$ if the curve can be split into a finite number of pieces such that $f(z)$ is continuous on each piece.*

Definition 11. *The* contour integral *of a piecewise continuous function along a smooth curve C is defined as*

$$\int_C f(z)\, dz = \int_a^b f(z(t))z'(t)\, dt.$$

The contour integral *of a piecewise continuous function along a piecewise smooth curve C is defined as*

$$\int_C f(z)\, dz = \sum_{k=1}^{m} \int_{C_k} f(z)\, dz,$$

where C is split at each point where it is not smooth, into the smooth curves $C_1, C_2, \ldots, C_m$. The integral of the function $f(z)$ along the closed curve C is denoted by

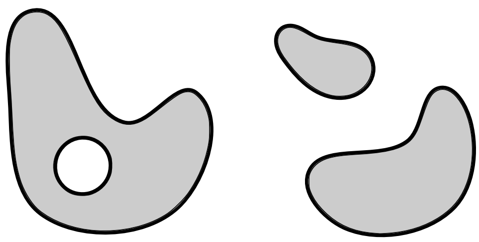
$$\oint_C f(z)\, dz.$$

By convention an integral along a simple closed curve C is taken in the direction such that the interior of C lies to the left of the curve.

2.6 Regions

The next concept to be defined is that of a region in the complex plane. First the concept of a connected set needs to be introduced.

Definition 12. *A set S is* connected *if given two arbitrary points a and b in S there exists a curve between a and b that lies in S. A connected set is called a* region.



(a) Connected Set (b) Non Connected Set

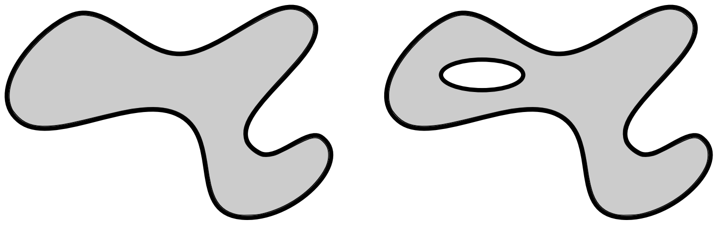
Figure 2: A connected and non connected set.

Another important concept is that of a simply connected region. Informally this can be thought of as a region that doesn't contain any holes.

Definition 13. *Let S be a set. Let $v(t) : [a, b] \to U$ and $u(t) : [a, b] \to U$ be two curves that lie in S. The curve $v(t)$ is* continuously deformable in S *into the curve $u(t)$ if there exists a continuous function $H(t, s)$ such that:*

1. *H(t,c) = v(t) and H(t,d) = u(t).*
2. *If $t_0 \in [a, b]$ and $s_0 \in [c, d]$ then $H(t_0, s_0) \in S$.*
3. *H(a,s) and H(b,s) are independent of s.*

Definition 14. *Let S be a set. Let $v(t)$ and $u(t)$ be two arbitrary curves in S that have a common starting point and a common ending point. Then the set S is* simply connected *if S is connected and the curve $v(t)$ can be continuously deformed in S into the curve $u(t)$.*

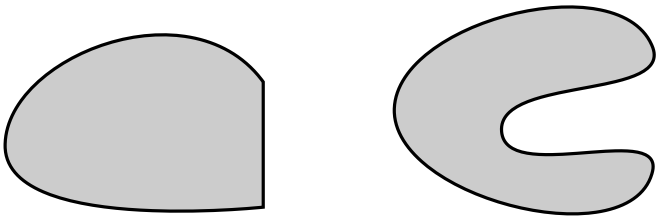


(a) Simply Connected Set (b) Non Simply Connected Set

Figure 3: A simply connected and non simply connected set.

The next concept that will be defined is that of a regular region. This requires that the concept of a standard region be defined first.

Definition 15. *An* interior point *of a region R is a point that lies in R but does not lie on the boundary of R. A region is a* standard region *if it is closed and bounded and all horizontal and vertical lines going through an interior point of the region R intersect the boundary of R at two points.*



(a) Standard Region (b) Non Standard region

Figure 4: A standard region and a non standard region.

Definition 16. *A region is a* regular region *if it is a standard region or it satisfies both of the following conditions:*

1. *The region can be split into a finite number of standard regions by splitting the region along a finite number of horizontal lines.*
2. *The region can be split into a finite number of standard regions by splitting the region along a finite number of vertical lines.*

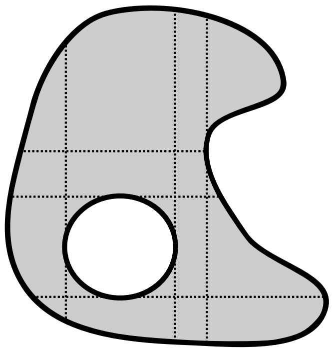


Figure 5: A regular region.

In the remainder of this section several theorems from Complex Analysis are presented. These theorems could perhaps be proved for more general sets than regular regions. There is however no need in this paper to prove these theorems for more general sets.

2.7 Cauchy's Theorem

An important theorem in Complex Analysis is Cauchy's Theorem. To prove Cauchy's Theorem it is necessary to first introduce Green's Theorem.

Green's Theorem. *Let C be a simple piecewise smooth closed curve that goes along the border of a simply connected regular region R in the counter clockwise direction. Let $u(x, y)$ and $v(x, y)$ be real functions that are continuous on R. Let the partial derivatives $\partial u/\partial x, \partial u/\partial y, \partial v/\partial x, \partial v/\partial y$ be continuous on R. Then*

$$\oint_C (u\, dx + v\, dy) = \iint_R (\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y})\, dx\, dy.$$

For a proof of Green's Theorem see [4]. The reason that Green's Theorem is relevant for Complex Analysis is due to the fact that the definition of a line integral in for a function $f(x, y) : \mathbb{R}^2 \to \mathbb{R}^2$ coincides with the definition of a contour integral of the same function when the x coordinate and y coordinate are interpreted as the real and imaginary components of a complex number.

Cauchy's Theorem. *Let C be a simple piecewise smooth closed curve that goes along the border of a simply connected regular region R. Let $f(z)$ be a complex function that is analytic on R. Then*

$$\oint_C f(z)\, dz = 0.$$

Proof. Assume that C goes around R in the counter clockwise direction. Then according to Green's Theorem

$$\oint_C (u\, dx + v\, dy) = \iint_R (\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y})\, dx\, dy.$$

According to the Cauchy-Riemann Equations the integrand in the double integral is 0 and therefore the value of the double integral is 0. A similar argument is used in the case where the curve C goes in the clockwise direction. □

2.8 Residue Theorem

The next concept to be defined is that of the residue of a complex function.

Definition 17. *Let the function $f(z)$ be analytic in a punctured disc D with center z_0. The* residue *of $f(z)$ at z_0 is defined as*

$$Res(f, z_0) = \frac{1}{2\pi i} \oint_C f(z)\, dz, \tag{2.33}$$

where C is a curve that lies in D that goes around a circle that is centered on z_0 in the counter clockwise direction.

The reader might be concerned that the residue is not well defined since the radius of the circle that the curve C goes around is not specified. The following lemma should alleviate that concern.

Lemma 4. *Let the function $f(z)$ be analytic in a punctured disc D with center z_0. Let C_1 be a simple piecewise smooth closed curve in D that goes around a circle centered on z_0 in the counter clockwise direction. Let C_2 be a simple piecewise smooth closed curve in D that goes around a circle centered on z_0 in the counter clockwise direction. Then*

$$\oint_{C_1} f(z)\,dz = \oint_{C_2} f(z)\,dz.$$

Proof. If the radius of the circles that the curves C_1 and C_2 go around are the same then the theorem is obviously true. Assume that the radius of the circle that the curve C_1 goes around is greater than the radius of the circle that the curve C_2 goes around. Pick a point a on C_1 and a point b on C_2 such that a and b lie on a ray from z_0. Let P be the path that is the composition of the following paths:

1. P_1: A curve starting at a and going around C_1 in the counter clockwise direction.
2. P_2: A path that goes along the line segment from a to b.
3. P_3: A path that starts at b and goes around C_2 in the clockwise direction.
4. P_4: A path that goes along the line segment from b to a.

It is obvious that P is a simple piecewise smooth closed path. Let R be the regular region enclosed by P. The function $f(z)$ is analytic on R. It is obvious that

$$\int_{P_2} f(z)\,dz + \int_{P_4} f(z)\,dz = 0. \tag{2.34}$$

According to Cauchy's Theorem

$$\int_{P_1} f(z)\,dz + \int_{P_2} f(z)\,dz + \int_{P_3} f(z)\,dz + \int_{P_4} f(z)\,dz = 0. \tag{2.35}$$

Substituting (2.34) into (2.35) yields

$$\int_{P_1} f(z)\,dz + \int_{P_3} f(z)\,dz = 0. \tag{2.36}$$

The Lemma now follows from the fact that

$$\int_{C_1} f(z)\,dz = \int_{P_1} f(z)\,dz,$$
$$\int_{C_2} f(z)\,dz = -\int_{P_3} f(z)\,dz.$$

□

An important theorem in Complex Analysis is the Residue Theorem. It can be viewed as an extension of Cauchy's Theorem that handles the case when the function that is being integrated is analytic everywhere inside and on the curve except at a finite number of points.

Residue Theorem. *Let C be a simple piecewise smooth closed curve that goes along the border of a simply connected regular region R in the counter clockwise direction. Let the function $f(z)$ be analytic in R except at the isolated interior points $z_1, z_2, \ldots, z_n$. Then*

$$\oint_C f(z)\,dz = 2\pi i \sum_{k=1}^{n} Res(f, z_k).$$

Proof. For each point z_k find a point u_k on the curve C such that none of the line segments intersect. For each of the points z_k create a curve C_k that goes along a circle centered on z_k in the counter clockwise direction and that does not intersect the curve C. Furthermore the radius of each circle should be sufficiently small to ensure that the curve C_k only intersects the line segment from z_k to u_k but not any of the other line segments.

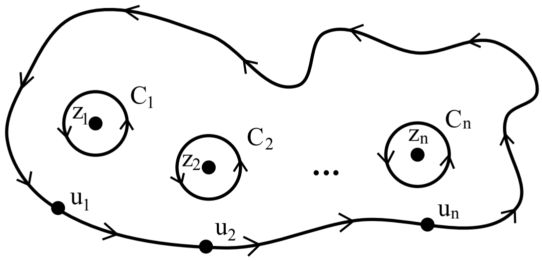


Figure 6: The curve C and the curves C_k.

Showing that the theorem is true is equivalent to showing that

$$\oint_C f(z)\,dz = \sum_{k=1}^{n} \oint_{C_k} f(z)\,dz.$$

Create a new curve T that goes along the curve C but that at each of the points u_k makes a detour along the line segment from u_k to z_k until it reaches the curve C_k. The curve T then goes along the curve C_k in the clockwise direction until it reaches the line segment again at which point the curve T goes back along the line segment to the point u_k. Let P_k be the part of the curve T that goes along the curve C_k.

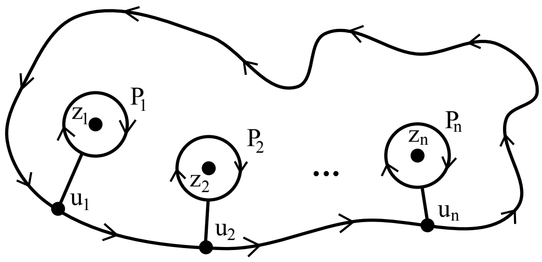


Figure 7: The curve T.

The curve C is a simple piecewise smooth closed curve and it is obvious that the same is true for the curve T. Furthermore the curve T lies on the border of a region R that is regular. It is obvious that the function $f(z)$ is analytic on R.

According to Cauchy's Theorem

$$\oint_T f(z)\,dz = 0. \tag{2.37}$$

It is obvious that for each line segment the integrals along the line segment cancel each other out and therefore the above equation can be transformed into

$$\oint_C f(z)\,dz + \sum_{k=1}^{n} \oint_{P_k} f(z)\,dz = 0. \tag{2.38}$$

The theorem now follows from the fact that

$$\oint_{C_k} f(z)\,dz = -\oint_{P_k} f(z)\,dz.$$

□

2.9 Neighborhood

The next concept to be defined is that of a neighborhood of a point.

Definition 18. *Let $N(c, r) = \{z \in \mathbb{C} | |z - c| < r\}$. The set $N(c, r)$ is known as a* neighborhood *of c.*

A neighborhood in $\mathbb{C}$ is an open disc.

2.10 Roots, Zeroes and Singularities

The next concept to be formalized is that of a root of a polynomial.

Definition 19. *Let $p(z)$ be a polynomial. The real or complex number c is said to be a* root *of $p(z)$ if*

$$p(c) = 0. \tag{2.39}$$

A root c of $p(z)$ is said to be a simple root *if*

$$p(z) = (z - c)q(z), \tag{2.40}$$

where $q(z)$ is a polynomial and $q(c) \neq 0$.

Let $n \in \mathbb{N}$ and $n \geq 1$. A root c of $p(z)$ is said to be a root of order n *if*

$$p(z) = (z - c)^n q(z), \tag{2.41}$$

where $q(z)$ is a polynomial and $q(c) \neq 0$. A root of order 2 or greater is called a multiple root.

The concept of a zero of an analytic function can be formalized in a similar manner to the concept of a root of a polynomial.

Definition 20. *Let $f(z)$ be an analytic function. The real or complex number c is said to be a* zero *of $f(z)$ if*

$$f(c) = 0. \tag{2.42}$$

A zero c of $f(z)$ is said to be a simple zero *if there is an analytic function $g(z)$ defined in a neighborhood of c such that*

$$f(z) = (z - c)g(z), \tag{2.43}$$

where $g(c) \neq 0$.

Let $n \in \mathbb{N}$ and $n \geq 1$. A zero c of $f(z)$ is said to be a zero of order n *if there is an analytic function $g(z)$ defined in a neighborhood of c such that*

$$f(z) = (z - c)^n g(z), \tag{2.44}$$

where $q(c) \neq 0$.

The next concept to be introduced is that of an isolated singular point of a complex function.

Definition 21. *Let f(z) be a function that is analytic in a neighborhood of z_0 except at the point z_0 itself. Then z_0 is called an* isolated singular point *of $f(z)$.*

A pole is an important type of isolated singular point.

Definition 22. *Let $f(z)$ be an analytic function. Let $n \in \mathbb{N}$ and $n \geq 1$. The complex number c is said to be a* pole of order n *if there is an analytic function $g(z)$ defined in a neighborhood of c such that*

$$f(z) = \frac{g(z)}{(z - c)^n}, \tag{2.45}$$

where $g(c) \neq 0$. If $f(z)$ has a pole of order n at c then $f(z)$ is said to have a pole *at c*

The complex number c is said to be a simple pole *if there is an analytic function $g(z)$ defined in a neighborhood of c such that*

$$f(z) = \frac{g(z)}{(z - c)},$$

where $g(x) \neq 0$.

A zero c of $f(z)$ is said to be a simple zero *if there is an analytic function $g(z)$ defined in a neighborhood of c such that*

$$f(z) = (z - c)g(z), \tag{2.46}$$

where $g(c) \neq 0$. A zero c of $f(z)$ is said to be a zero of order n *if there is an analytic function $g(z)$ defined in a neighborhood of c such that*

$$f(z) = (z - c)^n g(z), \tag{2.47}$$

where $q(c) \neq 0$.

The next concept to be defined is that of a meromorphic function.

Definition 23. *A complex function $f(z)$ is said to be* meromorphic *in a region R if it is analytic in R except at a finite number of isolated singular points where the function has poles.*

2.11 Cauchy's Integral Formula

An important theorem in Complex Analysis is Cauchy's Integral Formula. To prove Cauchy's Integral Formula it is necessary to first introduce some material concerning integrals. The proof of the following Lemma is due to Rudin (See [6]).

Lemma 5. *Let $f(x)$ be a piecewise continuous function that is defined on the interval $[a, b]$. Then*

$$\left| \int_a^b f(x)\,dx \right| \leq \int_a^b |f(x)|\,dx. \tag{2.48}$$

Proof. Let I be the magnitude of the integral of $f(x)$

$$I = \left| \int_a^b f(x)\,dx \right|. \tag{2.49}$$

Let c be defined as follows

$$c = \begin{cases} +1 & \text{if } I \geq 1. \\ -1 & \text{if } I < 1. \end{cases} \tag{2.50}$$

Then

$$|cf(x)| = |c||f(x)| = |f(x)| \tag{2.51}$$

and

$$I = \left| \int_a^b f(x)\,dx \right| = c\int_a^b f(x)\,dx = \int_a^b cf(x)\,dx \tag{2.52}$$

$$\leq \int_a^b |cf(x)|\,dx = \int_a^b |f(x)|\,dx. \tag{2.53}$$

□

The following theorem shows the magnitude of a curve integral of a piecewise continuous function along a curve is bounded by the length of the curve multiplied by an upper bound for the magnitude of the function along the curve.

Theorem 1. *Let $f(z)$ be a function that is continuous on a curve C. Let L be the arc length of the curve. Let M be an upper bound of $|f(z)|$ on C. Then*

$$\left| \int_C f(z)\,dz \right| \leq ML. \tag{2.54}$$

Proof. Let $z(t) : [a, b] \to \mathbb{C}$ be the curve C. Let I be the integral

$$I = \left| \int_C f(z)\,dz \right| = \left| \int_a^b f(z(t))z'(t)\,dt \right|.$$

Applying Lemma 5 to the above inequality results in

$$I \leq \int_a^b |f(z(t))z'(t)|\,dt = \int_a^b |f(z(t))||z'(t)|\,dt. \tag{2.55}$$

Since M is an upper bound for $|f(z)|$ the above inequality can be transformed into

$$I \leq \int_a^b M|z'(t)|\,dt = M\int_a^b |z'(t)|\,dt \tag{2.56}$$

All that remains now is to show that

$$\int_a^b |z'(t)|\,dt \leq L. \tag{2.57}$$

Let $x(t)$ and $y(t)$ be the real and imaginary components of $z(t)$. In other words

$$z(t) = x(t) + iy(t). \tag{2.58}$$

The magnitude of the derivative of $z(t)$ is equal to

$$|z'(t)| = |x'(t) + iy'(t)| = \sqrt{(x'(t))^2 + (y'(t))^2}. \tag{2.59}$$

Then

$$\int_a^b |z'(t)|\,dt = \int_a^b \sqrt{(x'(t))^2 + (y'(t))^2}\,dt \tag{2.60}$$

which is equal to the arc length of $z(t)$. □

Corollary 1. *Let $f(z)$ be a function that is piecewise continuous on a curve C. Let L be the arc length of the curve. Let M be an upper bound of $|f(z)|$ on C. Then*

$$\left|\int_C f(z)\,dz\right| \leq ML. \tag{2.61}$$

The proof of Cauchy's Integral Formula also uses the fact that an analytic function is continuous.

Lemma 6. *If the function $f(z)$ is analytic in a set S then $f(z)$ is continuous in S.*

Proof. Let z_0 be a point in S. The function $f(z)$ is continuous at z_0 if

$$\lim_{\zeta\to z_0} |f(\zeta) - f(z_0)| = 0. \tag{2.62}$$

The expression on the left hand side of the above equation can be rewritten as

$$\lim_{\zeta\to z_0} |f(\zeta) - f(z_0)| = \lim_{\zeta\to z_0}\left|\frac{f(\zeta) - f(z_0)}{\zeta - z_0}\right| \lim_{\zeta\to z_0} |\zeta - z_0| \tag{2.63}$$

The lemma follows from the following statements

1. The first limit on the left hand side of (2.63) is by definition equal to $f'(z_0)$. The limit exists since $f(z)$ is analytic.

2. The second limit on the left hand side of (2.63) is equal to 0.

□

Cauchy's Integral Formula relates the value of an analytic function at an interior point of a regular region to a curve integral going around the border of the region. The fact that the behavior of an analytic function at a point is determined by the behavior of the function along a curve enclosing the point is quite surprising.

Cauchy's Integral Formula. *Let C be a simple piecewise smooth closed curve that goes along the border of a regular region R in the counter clockwise direction. Then for any interior point z of R*

$$f(z) = \frac{1}{2\pi i} \oint_C \frac{f(\zeta)}{\zeta - z}\, d\zeta. \tag{2.64}$$

Proof. Let C_δ be a curve that goes in the counter clockwise direction along a small circle inscribed in the curve C centered on z with radius δ.

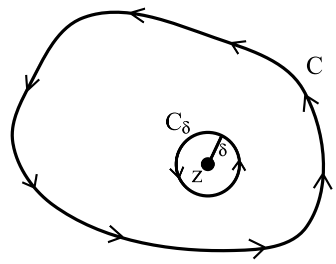


Figure 8: The curve C and the curve C_δ.

Then according to Cauchy's Theorem

$$\oint_C \frac{f(\zeta)}{\zeta - z}\, d\zeta = \oint_{C_\delta} \frac{f(\zeta)}{\zeta - z}\, d\zeta. \tag{2.65}$$

The right hand side of the above equation can be rewritten as

$$\oint_{C_\delta} \frac{f(\zeta)}{\zeta - z}\, d\zeta = f(z) \oint_{C_\delta} \frac{d\zeta}{\zeta - z} + \oint_{C_\delta} \frac{f(\zeta) - f(z)}{\zeta - z}\, d\zeta. \tag{2.66}$$

By performing the variable substitution $\zeta = z + \delta e^{i\theta}$ the first integral on the right hand side of (2.66) becomes

$$\oint_{C_\delta} \frac{d\zeta}{\zeta - z} = \int_0^{2\pi} \frac{i\delta e^{i\theta}}{\delta e^{i\theta}}\, d\theta = 2\pi i. \tag{2.67}$$

The function $f(z)$ is analytic and therefore according to Lemma 6 continuous. Given ϵ there exists r such that if $|\zeta - z| < r$ then $|f(\zeta) - f(z)| < \epsilon$.

According to Theorem 1 the second integral on the left hand side of (2.66) satisfies the following inequality

$$\left| \oint_{C_\delta} \frac{f(\zeta) - f(z)}{\zeta - z}\, d\zeta \right| \leq \oint_{C_\delta} \frac{|f(\zeta) - f(z)|}{|\zeta - z|}\, d\zeta. \tag{2.68}$$

Choose δ such that if $|\zeta - z| = \delta$ then $|f(\zeta) - f(z)| < \epsilon$. Then the left hand side of (2.68) satisfies the following inequality

$$\oint_{C_\delta} \frac{|f(\zeta) - f(z)|}{|\zeta - z|} \, d\zeta < \frac{\epsilon}{\delta} \oint_{C_\delta} d\zeta = 2\pi\epsilon. \tag{2.69}$$

The theorem now follow from the fact that the first integral of of the right hand side of (2.66) is equal to $2\pi i$ and as $\epsilon \to 0$ the second integral on the right hand side of equation (2.66) vanishes. □

Cauchy's Integral Formula can be used to show that the derivative of an analytic function is itself analytic.

Theorem 2. *Let C be a simple piecewise smooth closed curve that goes along the border of a regular region R in the counter clockwise direction. Let $f(z)$ be a function that is analytic on R. Let D be the set of interior points of R. Then $f'(z)$ is analytic on D. Furthermore*

$$f'(z) = \frac{1}{2\pi i} \oint_C \frac{f(\zeta)}{(\zeta - z)^2} \, d\zeta, \tag{2.70}$$

$$f''(z) = \frac{1}{2\pi i} \oint_C \frac{f(\zeta)}{(\zeta - z)^3} \, d\zeta. \tag{2.71}$$

Proof. Let L be the length of the curve C. If ζ lies on the curve C then let U be an upper bound of $|f(\zeta)|$

$$|f(\zeta)| < U. \tag{2.72}$$

If ζ lies on the curve C then let 2δ be a lower bound of $|\zeta - z|$

$$2\delta < |\zeta - z|, \tag{2.73}$$

where $\delta > 0$. If $|h| < \delta$ and ζ lies on the curve C then

$$|\zeta - (z + h)| \geq |\zeta - z| - |h| > 2\delta - \delta = \delta. \tag{2.74}$$

Let z be an arbitrary interior point of R. According to Cauchy's Integral Formula

$$\frac{f(z+h) - f(z)}{h} = \frac{1}{h}\frac{1}{2\pi i} \oint_C \frac{f(\zeta)}{\zeta - (z+h)} \, d\zeta - \frac{1}{h}\frac{1}{2\pi i} \oint_C \frac{f(\zeta)}{\zeta - z} \, d\zeta \tag{2.75}$$

$$= \frac{1}{h}\frac{1}{2\pi i} \oint_C f(\zeta) \left(\frac{1}{\zeta - (z+h)} - \frac{1}{\zeta - z} \right) d\zeta \tag{2.76}$$

$$= \frac{1}{2\pi i} \oint_C \frac{f(\zeta)}{(\zeta - (z+h))(\zeta - z)} \, d\zeta \tag{2.77}$$

$$= \frac{1}{2\pi i} \oint_C \frac{f(\zeta)(\zeta - z)}{(\zeta - (z+h))(\zeta - z)^2} \, d\zeta \tag{2.78}$$

$$= \frac{1}{2\pi i} \oint_C \frac{f(\zeta)(\zeta - (z+h) + h)}{(\zeta - (z+h))(\zeta - z)^2} \, d\zeta \tag{2.79}$$

$$= \frac{1}{2\pi i} \oint_C \frac{f(\zeta)}{(\zeta - z)^2} \, d\zeta + r(h), \tag{2.80}$$

where $r(h)$ is equal to

$$r(h) = \frac{h}{2\pi i} \oint_C \frac{f(\zeta)}{(\zeta - (z+h))(\zeta - z)^2} \, d\zeta. \tag{2.81}$$

To show that $f'(z)$ exists and is given by (y) it is sufficient to show that

$$\lim_{h \to 0} |r(h)| = 0. \tag{2.82}$$

If ζ lies on the curve C and $|h| < \delta$ then according to (2.72), (2.73) and (2.74) the following inequality holds

$$\left| \frac{f(\zeta)}{(\zeta - (z+h))(\zeta - z)^2} \right| \leq \frac{|f(\zeta)|}{|(\zeta - (z+h))||(\zeta - z)|^2} < \frac{U}{\delta (2\delta)^2}. \tag{2.83}$$

If $|h| < \delta$ then according to Theorem 1 and (2.83) the following inequality holds

$$|r(h)| = \frac{|h|}{2\pi} \left| \oint_C x \frac{f(\zeta)}{(\zeta - (z+h))(\zeta - z)^2} \, d\zeta \right| < \frac{|h|}{2\pi} \frac{U}{\delta (2\delta)^2} L. \tag{2.84}$$

Therefore according to the above inequality

$$\lim_{h \to 0} |r(h)| = 0, \tag{2.85}$$

and therefore (2.70) holds.

Let z be an arbitrary interior point of R. According to (2.70)

$$\frac{f'(z+h) - f'(z)}{h} = \frac{1}{h} \frac{1}{2\pi i} \oint_C \frac{f(\zeta)}{(\zeta - (z+h))^2} \, d\zeta - \frac{1}{h} \frac{1}{2\pi i} \oint_C \frac{f(\zeta)}{(\zeta - z)^2} \, d\zeta \tag{2.86}$$

$$= \frac{1}{h} \frac{1}{2\pi i} \oint_C f(\zeta) \left(\frac{1}{(\zeta - (z+h))^2} - \frac{1}{(\zeta - z)^2} \right) d\zeta \tag{2.87}$$

$$= \frac{1}{h} \frac{1}{2\pi i} \oint_C f(\zeta) \frac{(\zeta - z)^2 - ((\zeta - z) - h)^2}{(\zeta - (z+h))^2 (\zeta - z)^2} \, d\zeta \tag{2.88}$$

$$= \frac{1}{2\pi i} \oint_C f(\zeta) \frac{2(\zeta - z) - h}{(\zeta - (z+h))^2 (\zeta - z)^2} \, d\zeta \tag{2.89}$$

$$= \frac{1}{2\pi i} \oint_C f(\zeta) \frac{2(\zeta - (z+h)) + h}{(\zeta - (z+h))^2 (\zeta - z)^2} \, d\zeta \tag{2.90}$$

$$= q(h) + r(h), \tag{2.91}$$

$$\tag{2.92}$$

where $q(h)$ is equal to

$$q(h) = \frac{1}{2\pi i} \oint_C f(\zeta) \frac{2}{(\zeta - (z+h))(\zeta - z)^2} \, d\zeta, \tag{2.93}$$

$$\tag{2.94}$$

The function $q(h)$ can be split up as follows

$$q(h) = \frac{1}{2\pi i} \oint_C f(\zeta) \frac{2}{(\zeta - (z+h))(\zeta - z)^2} \, d\zeta, \tag{2.95}$$

$$= \frac{2}{2\pi i} \oint_C f(\zeta) \frac{(\zeta - z)}{(\zeta - (z+h))(\zeta - z)^3} \, d\zeta \tag{2.96}$$

$$= \frac{2}{2\pi i} \oint_C f(\zeta) \frac{(\zeta - (z+h)) + h}{(\zeta - (z+h))(\zeta - z)^3} \, d\zeta \tag{2.97}$$

$$= u(h) + v(h), \tag{2.98}$$

where $u(h)$ and $v(h)$ are equal to

$$u(h) = \frac{2}{2\pi i} \oint_C \frac{f(\zeta)}{(\zeta - z)^3} \, d\zeta, \tag{2.99}$$

$$v(h) = \frac{2h}{2\pi i} \oint_C \frac{f(\zeta)}{(\zeta - (z+h))(\zeta - z)^3} \, d\zeta. \tag{2.100}$$

To show that $f''(z)$ exists and is given by (y) it is sufficient to show that

$$\lim_{h \to 0} |r(h)| = 0, \tag{2.101}$$

$$\lim_{h \to 0} |v(h)| = 0. \tag{2.102}$$

However it has already been show that (2.101) holds. Therefore it is sufficient to show that (2.102) holds.

If ζ lies on the curve C and $|h| < \delta$ then according to (2.72), (2.73) and (2.74) the following inequality holds

$$\left| \frac{f(\zeta)}{(\zeta - (z+h))(\zeta - z)^3} \right| \leq \frac{|f(\zeta)|}{|(\zeta - (z+h))||(\zeta - z)|^3} < \frac{U}{\delta (2\delta)^3}. \tag{2.103}$$

If $|h| < \delta$ then according to Theorem 1 and (2.103) the following inequality holds

$$|v(h)| = \frac{|2h|}{2\pi} \left| \oint_C \frac{f(\zeta)}{(\zeta - (z+h))(\zeta - z)^3} \, d\zeta \right| < \frac{|h|}{\pi} \frac{U}{\delta (2\delta)^3} L. \tag{2.104}$$

Hence

$$\lim_{h \to 0} |v(h)| = 0. \tag{2.105}$$

□

Corollary 2. *Let C be a simple piecewise smooth closed curve that goes along the border of a regular region R in the counter clockwise direction. Let $f(z)$ be a function that is analytic on R. Let D be the set of interior points of R. Then all the derivatives $f^{(k)}(z)$ for $k = 1, 2, \ldots$ exist and are analytic on D.*

This is quite a surprising result and again shows that a complex function being analytic is much stronger than a real function being differentiable.

2.12 Winding Number

The next concept to be introduced is that of the winding number of a curve around a point. The material in this subsection is based on material presented in [13].

Definition 24. *The* winding number *of the curve C around the point z_0 is defined as*

$$W(C, z_0) = \frac{1}{2\pi i} \oint_C \frac{dz}{z - z_0}. \tag{2.106}$$

As the following theorem shows the winding number can be used to figure out how many times a curve winds or goes around a point.

Theorem 3. *Let C be a piecewise smooth closed curve that does not intersect z_0. Let $z(t) : [a, b]- > \mathbb{C}$ be a parametrization of the curve C defined as*

$$z(t) = z_0 + r(t)e^{i\theta(t)}, \tag{2.107}$$

where $r(t)$ and $\theta(t)$ are piecewise smooth functions and $r(t) > 0$. Then

$$W(C, z_0) = \frac{\theta(b) - \theta(a)}{2\pi}. \tag{2.108}$$

Proof. By definition

$$\oint_C \frac{dz}{z - z_0} = \int_a^b \frac{z'(t)}{z(t) - z_0}\, dt. \tag{2.109}$$

The derivative of $z(t)$ is equal to

$$z'(t) = r'(t)e^{i\theta(t)} + r(t)e^{i\theta(t)} i\theta'(t). \tag{2.110}$$

Substituting (2.107) and (2.110) into (2.109) results in

$$\oint_C \frac{dz}{z - z_0} = \int_a^b \frac{z'(t)}{z(t) - z_0}\, dt \tag{2.111}$$

$$= \int_a^b \frac{r'(t)e^{i\theta(t)} + r(t)e^{i\theta(t)} i\theta'(t)}{r(t)e^{i\theta(t)}}\, dt \tag{2.112}$$

$$= \int_a^b \frac{r'(t)}{r(t)}\, dt + i \int_a^b \theta'(t)\, dt \tag{2.113}$$

$$= [\ln(r(t))]_a^b + i[\theta(t)]_a^b \tag{2.114}$$

$$= \ln(r(b)) - \ln(r(a)) + i(\theta(b) - \theta(a)). \tag{2.115}$$

The above equation can be transformed into

$$\oint_C \frac{dz}{z - z_0} = i(\theta(b) - \theta(a)), \tag{2.116}$$

since $z(a) = z(b)$ and therefore

$$\ln(r(a)) = \ln(|z(a)|) = \ln(|z(b)|) = \ln(r(b)). \tag{2.117}$$

Dividing both sides of (2.116) by $2\pi i$ results in

$$\frac{1}{2\pi i} \oint_C \frac{dz}{z - z_0} = \frac{\theta(b) - \theta(a)}{2\pi}. \tag{2.118}$$

which by definition is equal to

$$W(C, z_0) = \frac{\theta(b) - \theta(a)}{2\pi}. \qquad (2.119)$$

□

As the following Lemma show a winding number will always be an integer.

Lemma 7. *Let C be a curve that does not intersect the point z_0. Then $W(C, z_0)$ is an integer.*

Proof. Let $z(t) : [a, b]- > \mathbb{C}$ be a parametrization of the curve C defined as

$$z(t) = z_0 + r(t)e^{i\theta(t)}, \qquad (2.120)$$

Since C is a closed curve

$$\theta(b) = \theta a + k2\pi, \qquad (2.121)$$

for some integer k. The above equation can be rewritten as

$$\theta(b) - \theta(a) = k2\pi, \qquad (2.122)$$

According to Theorem 3

$$W(C, z_0) = \frac{\theta(b) - \theta(a)}{2\pi}. \qquad (2.123)$$

Substituting (2.122) into (2.123) results in

$$W(C, z_0) = k. \qquad (2.124)$$

□

In other words the winding number $W(C, z_0)$ counts the number of times the curve C goes around the point z_0. There are three cases:

1. A positive non zero winding number means that the curve goes around the point z_0 in the counter clockwise direction more times than it goes around the point z_0 in the clockwise direction.

2. A negative winding number means that the curve goes around the point z_0 in the clockwise direction more times than it goes around the point z_0 in the counter clockwise direction.

3. A winding number of 0 means that the curve does not go around the point z_0 at all or that the curve goes around the point z_0 the in the counter clockwise direction the same number of times it goes around the point z_0 in the counter clockwise direction.

2.13 Argument Principle

An important theorem in Complex Analysis is the Argument Principle. Before the Argument Principle can be proved it is necessary to first introduce a couple of Lemmas related to the residues of analytic and meromorphic functions.

Lemma 8. *If $f(z)$ is analytic in a neighborhood N of z_0 then*

$$Res(f, z_0) = 0. \tag{2.125}$$

Proof. Let C be a curve that goes in the counter clockwise direction around a circle that lies in N and that has center z_0. Then by definition

$$Res(f, z_0) = \frac{1}{2\pi i} \oint_C f(z)\, dz. \tag{2.126}$$

According to Cauchy's Theorem

$$\oint_C f(z)\, dz = 0, \tag{2.127}$$

since $f(z)$ is analytic. □

Lemma 9. *Let $f(z)$ and $g(z)$ be two complex functions that are analytic in a neighborhood of c. Furthermore let $g(c) \neq 0$. Let $h(z)$ be the complex function*

$$h(z) = \frac{f(z)}{g(z)}. \tag{2.128}$$

Then $h(z)$ is analytic in a neighborhood of c.

Proof. The function $g(z)$ is analytic in a neighborhood N of c and therefore $f(z)$ is continuous in N. Since $g(z)$ is continuous in N and $g(c) \neq 0$ there is a neighborhood M of c such that $g(z) \neq 0$ in M. Hence there is a neighborhood of c where $g(z)$ is both analytic and non zero. □

Lemma 10. *Let $f(z)$ and $g(z)$ be two complex functions that are meromorphic on D. Then for any point $z_0 \in D$*

$$Res(af + bg, z_0) = aRes(f, z_0) + bRes(g, z_0). \tag{2.129}$$

Proof. Let C be a curve that goes in the counter clockwise direction around a small circle centered on z_0. Choose the radius r of the circle such that $f(z)$ is analytic in the punctured disc centered on z_0 with radius r.

$$\begin{aligned} Res(af + bg, z_0) &= \frac{1}{2\pi i} \oint_C af(z) + bg(z)\, dz && (2.130)\\ &= a\frac{1}{2\pi i} \oint_C f(z)\, dz + b\frac{1}{2\pi i} \oint_C g(z)\, dz && (2.131)\\ &= aRes(f, z_0) + bRes(g, z_0). && (2.132) \end{aligned}$$

□

Lemma 11. *Let $f(z)$ be the complex function*

$$f(z) = \frac{a}{z - z_0}. \tag{2.133}$$

Then the residue of $f(z)$ at z_0 is

$$Res(f, z_0) = a. \tag{2.134}$$

Proof. Let C be a curve that goes in the counter clockwise direction along a circle centered on z_0. Then

$$Res(f, z_0) = \frac{1}{2\pi i} \oint_C \frac{a}{z - z_0}\, dz = a \frac{1}{2\pi i} \oint_C \frac{1}{z - z_0}\, dz = aW(C, z_0). \tag{2.135}$$

The winding number $W(C, z_0) = 1$ since the curve C goes once around the point z_0 in the counter clockwise direction. Therefore

$$Res(f, z_0) = a. \tag{2.136}$$

□

The following Lemma gives an explicit formula for calculating the order of a root and the order of a pole.

Lemma 12. *If the meromorphic function $f(z)$ has a root of order k at c then the order of the root c can be calculated using*

$$Res\left(\frac{f'(z)}{f(z)}, c\right) = k. \tag{2.137}$$

If the meromorphic function $f(z)$ has a pole of order k at c then the order of the pole c can be calculated using

$$Res\left(\frac{f'(z)}{f(z)}, c\right) = -k. \tag{2.138}$$

Proof. Assume that the function $f(z)$ has a root of order k at c or a pole of order k at c. Then there exists a complex function $g(z)$ that is analytic in a neighborhood of c that satisfies $g(z) \neq 0$ such that the function $f(z)$ can then be written as

$$f(z) = (z - c)^{\pm k} g(z), \tag{2.139}$$

where the sign of $k \neq 0$ is positive if c is a root and negative if c is a pole. The derivative of $f(z)$ becomes

$$f'(z) = \pm k(z - c)^{\pm k - 1} g(z) + (z - c)^{\pm k} g'(z). \tag{2.140}$$

The function $f'(z)/f(z)$ can then be written as

$$\frac{f'(z)}{f(z)} = \frac{\pm k(z - c)^{\pm k - 1} g(z) + (z - c)^{\pm k} g'(z)}{(z - c)^{\pm k} g(z)} = \frac{\pm k}{z - c} + \frac{g'(z)}{g(z)}. \tag{2.141}$$

Taking the residue of both sides of the above equation and applying Lemma 10 results in

$$Res\left(\frac{f'(z)}{f(z)}, c\right) = Res\left(\frac{\pm k}{z - c}, c\right) + Res\left(\frac{g'(z)}{g(z)}, c\right). \tag{2.142}$$

According to Lemma 8 the above equation is equal to

$$Res\left(\frac{f'(z)}{f(z)}, c\right) = Res\left(\frac{\pm k}{z-c}, c\right), \tag{2.143}$$

since $g'(z)/g(z)$ is an analytic function. Applying Lemma 11 to the the above equation results in

$$Res\left(\frac{f'(z)}{f(z)}, c\right) = \pm k. \tag{2.144}$$

□

The next concept to be defined is that of the image of a function along a curve.

Definition 25. *Let C be a curve with the parametrization $z(t) : [a,b] \to \mathbb{C}$. Let $f(z)$ be a function that is continuous on C. Let $u(t)$ be a curve. Then $u(t)$ is the* image *of $f(z)$ on the curve C if*

$$u(t) = f(z(t)). \tag{2.145}$$

The argument principle relates the number of zeroes and poles of a function $f(z)$ in a region R bounded by a curve C to the winding number around the origin of the image of $f(z)$ on the curve C.

Argument Principle. *Let C be a simple piecewise smooth closed curve that goes in the counter clockwise direction along the border of a regular region R. Let $f(z)$ be a function that is meromorphic on R and that does not have any zeroes or poles on C.*

Let N be the number of zeroes of $f(z)$ in R and let P be the number of poles in of $f(z)$ in R where a multiple zero or pole is counted according to its order. Let C^* be the image of the function $f(z)$ on the curve C. Then

$$\frac{1}{2\pi i}\oint_C \frac{f'(z)}{f(z)}\,dz = N - P = W(C^*, 0). \tag{2.146}$$

Proof. If $f(z)$ has a zero or pole at c then there is a complex function $g(z)$ that is analytic in a neighborhood of c with $g(c) \neq 0$ such that $f(z)$ is equal to

$$f(z) = (z-c)^k g(z), \tag{2.147}$$

where $k \in \mathbb{N}$ and $k \neq 0$. The derivative of $f(z)$ is equal to

$$f'(z) = k(z-c)^{k-1}g(z) + (z-c)^k g'(z). \tag{2.148}$$

Let $h(z)$ be the complex function

$$h(z) = \frac{f'(z)}{f(z)}. \tag{2.149}$$

Substituting (2.147) and (2.148) into (2.149) results in

$$h(z) = \frac{k(z-c)^{k-1}g(z) + (z-c)^k g'(z)}{(z-c)^k g(z)} \tag{2.150}$$

$$= \frac{kg(z) + (z-c)g'(z)}{(z-c)g(z)} \tag{2.151}$$

$$= \frac{u(z)}{z-c}, \tag{2.152}$$

where $u(z)$ is equal to

$$u(z) = \frac{kg(z) + (z-c)g'(z)}{g(z)}. \tag{2.153}$$

According to Lemma 9 the complex function $h(z)$ will be analytic in a neighborhood of c since both the numerator and denominator of $u(z)$ are analytic in a neighborhood of c and the denominator of $u(z)$ is non zero at c.

Hence $h(z)$ is a meromorphic function that has a simple poles at the points where $f(z)$ has a zero or a pole.

Therefore according to the Residue Theorem

$$\frac{1}{2\pi i}\oint_C \frac{f'(z)}{f(iz)}\,dz = \sum_{k=1}^{a} Res(h, z_k) + \sum_{j=1}^{b} Res(h, p_j), \tag{2.154}$$

where $z_1, z_2, \ldots, z_a$ are the zeroes of $f(z)$ and $p_1, p_2, \ldots, p_b$ are the poles of $f(z)$. Let c_k be the order of the zero z_k. Let d_k be the order of the pole p_k.

According to Lemma 12

$$\sum_{k=1}^{a} Res(h, z_k) = \sum_{k=1}^{a} c_k = N. \tag{2.155}$$

According to Lemma 12

$$\sum_{j=1}^{b} Res(h, p_j) = \sum_{j=1}^{b} -d_j = -P. \tag{2.156}$$

Substituting (2.155) and (2.156) into (2.154) results in

$$\frac{1}{2\pi i}\oint_C \frac{f'(z)}{f(z)}\,dz = N - P. \tag{2.157}$$

Performing the variable substitution $w = f(z)$ on the integral in the above equation results in

$$N - P = \frac{1}{2\pi i}\oint_C \frac{f'(z)}{f(z)}\,dz = \frac{1}{2\pi i}\oint_{C^*} \frac{dw}{w} = W(C^*, 0). \tag{2.158}$$

□

2.14 Rouché's Theorem

An important theorem in Complex analysis is Rouché's Theorem.

Rouché's Theorem. *Let C be a simple piecewise smooth closed curve that goes in the counter clockwise direction along the border of a regular region R. Let $f(z)$ and $g(z)$ be analytic on R. If $|f(z)| > |g(z)|$ on C then $f(z)$ and $f(z)+g(z)$ will have the same number of zeroes in R where the zeroes are counted according to their order.*

Proof. Let the complex function $w(z)$ be

$$w(z) = \frac{f(z) + g(z)}{f(z)}. \tag{2.159}$$

If z lies on the curve C then $f(z) \neq 0$ since $|f(z)| > |g(z)| \geq 0$ on R. Therefore $w(z)$ is well defined on C and the curve integral

$$\frac{1}{2\pi i} \oint_C \frac{w'(z)}{w(z)}\, dz \tag{2.160}$$

is well defined. Let N be the number of zeroes of $w(z)$ in R where each zero is counted according to its order. Let P be the number of poles of $w(z)$ in R where each pole is counted according to its order. Let C^* be the image of $w(z)$ on the curve C. If z lies on the curve C then

$$|w(z) - 1| = \frac{|g(z)|}{|f(z)|} < 1. \tag{2.161}$$

Therefore C^* lies in the open disc centered on the point 1 with radius 1. The curve C^* can thus never wind around the origin and therefore $W(C^*, 0) = 0$. Then according to the Argument Principle $N = P$. The theorem now follows from the fact that the number of zeroes of $f(z) + g(z)$ in R is equal to the number of zeroes of $w(z)$ in R and the number of zeroes of $f(z)$ in R is equal to the number of poles of $w(z)$ in R. □

3 Numerical Root Finding

There are several well known methods for finding the roots of a real polynomial. Some of these methods can be used to find the roots of a complex polynomial. In this section a method for finding all the roots of a complex polynomial and all unique roots of a complex polynomial ignoring multiplicities are developed. This section is based on material presented by Press et al. (see [14]) and Ralston and Rabinowitz (see [5]).

3.1 Newton's Method

The most well known numerical method for finding the zeros of a function is Newton's method. Any method that can be used to find the zeros of a general function can obviously be used to find the roots of a polynomial. Newton's method is such a method that can be used to find the zeroes of a general function. The function needs to be differentiable near the zero and also have an invertible differential.

Newton's method works as follows. Given an initial guess z_0 of the value of the root create the number sequence $\{z_k\}$ for $k = 0, 1, 2...$ where z_k is defined by

$$z_{k+1} = z_k - \frac{f(z_k)}{f'(z_k)} \tag{3.1}$$

If the starting value z_k was chosen to be close enough to a root then the sequence $\{z_k\}$ will converge to the root.

A more formal argument for why Newton's method works follows. Begin by Taylor expanding the function $f(z)$ around the point z_k.

$$f(z) = \sum_{j=0}^{n} f^{(j)}(z_k)(z - z_k)^j \tag{3.2}$$

The best linear approximation of $f(z)$ at z_k is given by the first two terms of the Taylor series.

$$t(z) = f(z_k) + f'(z_k)(z - z_k) \tag{3.3}$$

Now make the assumption that is at the core of Newton's method, namely that the root of $t(z)$ will be a better approximation of the root than z_k. Let z_{k+1} be the root of $t(z)$. In other words z_{k+1} should satisfy the following equation.

$$t(z_{k+1}) = 0 \tag{3.4}$$

Expanding (4) using the definition of $t(z)$ we get the following equation.

$$f(z_k) + f'(z_k)(z_{k+1} - z_k) = 0 \tag{3.5}$$

Subtracting $f(z_k)$ from both sides of the equation, then dividing both sides of the equation with $f'(z_k)$ and finally adding z_k to both sides of the equation gives us

$$z_{k+1} = z_k - \frac{f(z_k)}{f'(z_k)} \tag{3.6}$$

To find a zero of the function $f(z)$ start with an initial guess z_0 and iterate until $f(z)/f'(z)$ becomes sufficiently small. It can be proved that for complex functions that are differentiable with $f(z_k) \neq 0$ near the root, z_k will converge to a zero if the initial guess z_0 is sufficiently close to a zero of the function. If z_0 is not close to a zero of $f(z)$ then Newton's method is not guaranteed to converge.

Newton's method is not suitable for use as a general purpose root finding algorithm since convergence is not guaranteed. For Newton's method to be guaranteed to converge converge some other method for obtaining information about the location of the root is needed. Once the root is located Newton's method can be used to refine the root to any desired accuracy.

3.2 Laguerre's Method

Laguerre's method is a commonly used numerical method for finding the roots of a polynomial. It can be applied to both real and complex polynomials and it will find both real and complex roots. The method works by constructing a number sequence $\{z_k\}$ that in practice always converges to the root.

The elements of the sequence are defined using the following recursive formula where z_0 is the initial guess at the value of the root.

$$z_{k+1} = z_k - \frac{n}{G(z_k) \pm \sqrt{(n-1)(nH(z_k) - G^2(z_k))}} \tag{3.7}$$

The sign is chosen to maximize the magnitude of the denominator. The helper functions $G(z)$ and $H(z)$ are defined as follows

$$G(z) = \frac{p'(z)}{p(z)} \tag{3.8}$$

$$H(z) = G^2(z) - \frac{p''(z)}{p(z)} \tag{3.9}$$

What follows is a motivation for why the method works that is based on material presented by Weisstein (see [10]). According to the fundamental theorem of algebra a polynomial of degree n can be written as

$$p(z) = (z - \alpha_1)(z - \alpha_2) \ldots (z - \alpha_n) \tag{3.10}$$

where α_1,α_2,...,α_n are the roots of the polynomial. Taking the derivative of both sides of (3.10) yields

$$p'(z) = (z - \alpha_2) \ldots (z - \alpha_n) + \cdots + (z - \alpha_1) \ldots (z - \alpha_n) + \ldots \tag{3.11}$$

$$= p(z) \left(\frac{1}{z - \alpha_1} + \frac{1}{z - \alpha_2} + \cdots + \frac{1}{z - \alpha_n} \right) \tag{3.12}$$

which after dividing both sides by $p(z)$ becomes

$$\frac{p'(z)}{p(z)} = \frac{1}{z - \alpha_1} + \frac{1}{z - \alpha_2} + ... + \frac{1}{z - \alpha_n} \tag{3.13}$$

and is equal to $G(z)$ by definition. Multiplying both sides of the above equation by -1 results in

$$\frac{1}{\alpha_1 - z} + \frac{1}{\alpha_2 - z} + ... + \frac{1}{\alpha_n - z} = -\frac{p'(z)}{p(z)}. \tag{3.14}$$

Taking the derivative of both sides of the above equation yields

$$\frac{1}{(z-\alpha_1)^2}+\cdots+\frac{1}{(z-\alpha_n)^2} = \frac{p'(z)p'(z)-p"(z)p(z)}{p(z)^2} \tag{3.15}$$

$$= \left(\frac{p'(z)}{p(z)}\right)^2 - \frac{p"(z)}{p(z)} \tag{3.16}$$

$$= G^2(z) - \frac{p"(z)}{p(z)} \tag{3.17}$$

which is equal to $H(z)$ by definition. Take note of the fact that the functions $G(z)$ and$H(z)$ can both be expressed in terms of differences between z and the roots $\alpha_1, \alpha_2, \ldots, \alpha_n$.

$$G(z) = \frac{1}{z-\alpha_1}+\frac{1}{z-\alpha_2}+...+\frac{1}{z-\alpha_n} \tag{3.18}$$

$$H(z) = \frac{1}{(z-\alpha_1)^2}+\frac{1}{(z-\alpha_2)^2}+\cdots+\frac{1}{(z-\alpha_n)^2} \tag{3.19}$$

Assume that the difference between z and α_1 is equal to a and that the difference between z and all other roots is b. Rewriting H(z) and G(z) using these assumptions yields the following.

$$G = \frac{1}{a}+\frac{n-1}{b} \tag{3.20}$$

$$H = \frac{1}{a^2}+\frac{n-1}{b^2} \tag{3.21}$$

Solving for a we get

$$a = \frac{n}{G(z_k) \pm \sqrt{(n-1)(nH(z_k)-G^2(z_k))}} \tag{3.22}$$

According to the assumption $a = z - \alpha_1$ which in turn means that $\alpha_1 = z - a$. Hoping that $z - a$ will be a better approximation to the root than z is we arrive at the following recursive formula for z_k

$$z_{k+1} = z_k - \frac{n}{G(z_k) \pm \sqrt{(n-1)(nH(z_k)-G^2(z_k))}} \tag{3.23}$$

The sign is chosen to minimize the denominator. This is to avoid loss of significance. To find a root start with an initial guess z_0 and iterate until a becomes sufficiently small.

3.3 Polynomial Deflation

If α is a root of the polynomial $p(z)$ then the polynomial $p(z)$ can be written as

$$p(z) = (z-\alpha)q(z) \tag{3.24}$$

where the degree of $q(z)$ will be 1 lower than the degree of $p(z)$. The process of calculating $q(z)$ is known as polynomial deflation. The division algorithm can be used to calculate

$$q(z) = \frac{p(z)}{(z-\alpha)} \tag{3.25}$$

but in this case since the divisor is a linear polynomial there is a better way to perform the above calculation: Horner's method can be used to deflate the polynomial using fewer operations than the division algorithm would require.

Horner's method is an algorithm for evaluating a polynomial. It can also be used to calculate the quotient and remainder of dividing a polynomial by a linear polynomial.

Horner's Method. *Assume that* $p(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0$. *Create a new set of numbers* b_k *defined by*

$$b_n = a_n \tag{3.26}$$
$$b_k = a_k + b_{k+1} z_0 \qquad \textit{for } k = n-1, n-2, \ldots, 1, 0 \tag{3.27}$$

Then $p(z_0) = b_0$ *and* $p(z) = (z - z_0)(b_n z^{n-1} + b_{n-1} z^{n-2} + \cdots + b_2 z + b_1) + b_0$.

Proof. Assume that

$$p(z) = a_n z^n + \sum_{k=0}^{n-1} a_k z^k \tag{3.28}$$

and

$$b_n = a_n \tag{3.29}$$
$$b_k = a_k + b_{k+1} z_0 \qquad \textit{for } k = n-1, n-2, \ldots, 1, 0 \tag{3.30}$$

Solving the above equations for a_n and a_k results in

$$a_n = b_n \tag{3.31}$$
$$a_k = b_k - b_{k+1} z_0 \qquad \textit{for } k = n-1, n-2, \ldots, 1, 0 \tag{3.32}$$

Substituting the above equations into (3.28) results in

$$p(z) = b_n z^n + \sum_{k=0}^{n-1} (b_k - b_{k+1} z_0) z^k \tag{3.33}$$
$$= b_n z^n + \sum_{k=1}^{n} (b_{k-1} - b_k z_0) z^{k+1} \tag{3.34}$$
$$= b_n z^n + \sum_{k=2}^{n} (b_{k-1} - b_k z_0) z^{k+1} + b_0 - b_1 z_0 \tag{3.35}$$

Splitting the summation into two parts yields

$$p(z) = b_n z^n + \sum_{k=2}^{n} b_{k-1} z^{k-1} + \sum_{k=2}^{n} -z_0 b_k z^{k-1} - z_0 b_1 + b_0 \tag{3.36}$$
$$= b_n z^n + \sum_{k=1}^{n-1} b_k z^k + \sum_{k=1}^{n} -z_0 b_k z^{k-1} + b_0 \tag{3.37}$$
$$= z \sum_{k=1}^{n} b_k z^{k-1} - z_0 \sum_{k=1}^{n} b_k z^{k-1} + b_0 \tag{3.38}$$
$$= (z - z_0) \sum_{k=1}^{n} b_k z^{k-1} + b_0 \tag{3.39}$$

From the above equation it is obvious that $p(z_0) = b_0$. □

When the above algorithm is used to calculate the quotient and remainder of dividing a polynomial by a linear polynomial then it is also known as synthetic division.

3.4 Finding All Roots

The methods discussed so far only find single roots of a polynomial. To find all the roots of the polynomial $p(z)$ the following approach can be used. Apply the following process recursively until all the roots have been found.

1. Use Laguerre's method to find a root α of the polynomial.
2. Factor out the root α using synthetic division.

The calculated roots are not exact and will contain errors. This in turn means that the coefficients of the deflated polynomials will also contain errors. The error will grow each time the polynomial is deflated. After all the roots have been found they are polished using Newton's method to remove these errors.

3.5 Multiple Roots

To find all the unique roots of a polynomial ignoring their multiplicities the following method can be used. Assume that $p(z)$ can be written as

$$p(z) = (z - \alpha_1)^{k_1}(z - \alpha_2)^{k_2} \dots (z - \alpha_n)^{k_n} \tag{3.40}$$

where $\alpha_1, \dots, \alpha_m$ are unique roots and $k_1, \dots, k_m$ their multiplicities. Finding the unique roots of the above equation is equivalent to finding all the roots of the following polynomial.

$$q(z) = (z - \alpha_1)(z - \alpha_2) \dots (z - \alpha_m) \tag{3.41}$$

The polynomial $q(z)$ is equal to $p(z)$ divided by the following polynomial

$$d(z) = (z - \alpha_1)^{k_1-1}(z - \alpha_2)^{k_2-1} \dots (z - \alpha_n)^{k_n-1} \tag{3.42}$$

Finding the unique roots of $p(z)$ is equivalent to finding all the roots of $p(z)$ divided by $d(z)$. Since a method for finding all the roots of a polynomial already exists all that remains is the calculation of $d(z)$.

Lemma 13.

$$gcd(p(z), p'(z)) = (z - \alpha_1)^{k_1-1}(z - \alpha_2)^{k_2-1} \dots (z - \alpha_m)^{k_m-1} \tag{3.43}$$

Proof. Assume that the polynomial p(z) can be written as

$$p(z) = \prod_{j=1}^{m} (z - \alpha_j)^{k_j} \tag{3.44}$$

Let $q(z)$ be defined as

$$q(z) = (z - \alpha_1)(z - \alpha_2) \dots (z - \alpha_m) \tag{3.45}$$

Let $d(z)$ be defined as

$$d(z) = (z - \alpha_1)^{k_1 - 1}(z - \alpha_2)^{k_2 - 1} \dots (z - \alpha_m)^{k_m - 1} \tag{3.46}$$

It is obvious that $p(z) = q(z)d(z)$. Taking the derivative of $p(z)$ results in

$$\begin{aligned} p'(z) &= k_1(z - \alpha_1)^{k_1 - 1}(z - \alpha_2)^{k_2} \dots (z - \alpha_m)^{k_m} + \\ &\quad (z - \alpha_1)^{k_1} k_2 (z - \alpha_2)^{k_2 - 1} \dots (z - \alpha_m)^{k_m} + \\ &\quad \dots \\ &\quad (z - \alpha_1)^{k_1}(z - \alpha_2)^{k_2} \dots k_m (z - \alpha)^{k_m - 1} \\ &= \sum_{i=1}^{m} \frac{k_i p(z)}{(z - \alpha_i)} \\ &= d(z) \sum_{i=1}^{m} \frac{k_i q(z)}{(z - \alpha_i)} \end{aligned}$$

The polynomial $d(z)$ is a divisor of both $p(z)$ and $p'(z)$. Let $r(z)$ be defined as

$$r(z) = \sum_{i=1}^{m} \frac{k_i q(z)}{(z - \alpha_i)} \tag{3.47}$$

Then $p'(z) = d(z)r(z)$. Showing that $d(z)$ is the greatest common divisor of $p(z)$ and $p'(z)$ is equivalent to showing that $q(z)$ and $r(z)$ do not have any common factors.

Assume that $(z - \alpha_j)$ divides $q(z)$. From the definition of $r(z)$ it is trivial to see that $(z - \alpha_j)$ will divide all but one of the terms of $r(z)$. The remaining non zero term is not divisible by $(z - \alpha_j)$. Therefore $(z - \alpha_j)$ will not divide $r(z)$. But this means that $q(z)$ and $r(z)$ do not have any common factors.

□

4 Families of Polynomials

First it is necessary to introduce the concept of a family of polynomials.

Definition 26. *A* family of polynomials *is a function $p(w) : \mathbb{C} \to \mathbb{C}[z]$. By convention a family of polynomials will be parameterized by the variable w and the generated polynomials will be polynomials in z. The domain of $p(w)$ is called the* parameter plane*.*

A polynomial in two variables with complex coefficients

$$f(z, w) = \sum c_{kj} w^k z^j \tag{4.1}$$

can be viewed as a polynomial in z

$$f(z, w) = \sum_{k=0}^{n} c_k(w) z^k, \tag{4.2}$$

where the coefficients $c_k(w)$ are complex polynomials in w. In other words a bivariate polynomial $f(z, w)$ can be viewed as a family of polynomials. In this section several theorems regarding complex univariate and bivariate polynomials are presented.

4.1 Multiple Roots

There is an interesting relationship between the derivatives of a polynomial and the roots of the same polynomial.

Lemma 14. *$p(z) = (z - \alpha)^m q(z) \implies p^{(j)}(z) = (z - \alpha)^{m-j} r(z)$ if $0 \leq j \leq m$.*

Proof. The proof is by induction on j. It is obvious that the lemma is true for $j = 0$. Assume that the lemma is true for $j = k \leq m$. Then for $k < m$

$$\begin{aligned}
p^{(k+1)}(z) &= \frac{d}{dz}\left(p^{(k)}(z)\right) \\
&= \frac{d}{dz}\left((z-\alpha)^{m-k} r(z)\right) \\
&= (m-k)(z-\alpha)^{m-k-1} r(z) + (z-\alpha)^{m-k} r'(z) \\
&= (z-\alpha)^{m-(k+1)}(m-k) r(z) + (z-\alpha)^{m-(k+1)}(z-\alpha) r'(z) \\
&= (z-\alpha)^{m-(k+1)} q(z)
\end{aligned}$$

which means that the lemma is true for $j = k + 1 \leq m$. □

Theorem 4. *Let $p(z)$ be a complex polynomial. Then $p(z) = (z - \alpha)^k q(z)$ if and only if $p(\alpha) = 0, p'(\alpha) = 0, \ldots, p^{(k-1)}(\alpha) = 0$.*

Proof. Assume that $p(z) = (z - \alpha)^k q(z)$. From Lemma 14 is can be seen that the first $k - 1$ derivatives of $p(z)$ will have a root at α.

Assume that $p(z) = 0, p'(z) = 0, \ldots, p^{(k-1)}(z) = 0$. Taylor expanding $p(z)$ around the point α results in

$$p(z) = \sum_{j=0}^{n} \frac{f^{(j)}(\alpha)}{j!}(z - \alpha)^j \tag{4.3}$$

Since the first $k-1$ derivatives of $p(z)$ are zero $p(z)$ can be rewritten as

$$\begin{aligned} p(z) &= \sum_{j=k}^{n} \frac{f^{(j)}(\alpha)}{j!}(z-\alpha)^j && (4.4) \\ &= (z-\alpha)^k \sum_{j=k}^{n} \frac{f^{(j)}(\alpha)}{j!}(z-\alpha)^{j-k} && (4.5) \\ &= (z-\alpha)^k q(z) && (4.6) \end{aligned}$$

□

Corollary 3. *A polynomial $p(z)$ has a multiple root at α if and only if $p(\alpha) = 0$ and $p'(\alpha) = 0$.*

4.2 Continuity

A simple root of a polynomial will vary continuously with the coefficients of the polynomial. In some sense the same is true for a root of order 2 or greater.

Theorem 5. *Let $c_0 + c_1 z + \ldots + c_{n-1}z^{n-1} + c_n z^n$ be a polynomial with the roots $\alpha_1, \alpha_2, \ldots, \alpha_n$. Given ε there exists δ such that if $|w_k| < \delta$ then $c_0 + \ldots + c_n z^n$ and $[c_0 + w_0] + \ldots + [c_n + w_n]z^n$ will have the same number of roots in $N(\alpha_k, \varepsilon)$.*

Proof. Assume that

$$p(z) = c_0 + c_1 z + \ldots + c_{n-1}z^{n-1} + c_n z^n, \tag{4.7}$$

has the roots $\alpha_1, \alpha_2, \ldots, \alpha_n$. Let

$$q(z) = [c_0 + w_0] + [c_1 + w_1]z + \ldots + [c_{n-1} + w_{n-1}]z^{n-1} + [c_n + w_n]z^n. \tag{4.8}$$

Let

$$h(z) = w_0 + w_1 z + \ldots + w_{n-1}z^{n-1} + w_n z^n. \tag{4.9}$$

Then $q(z) = p(z) + h(z)$. Pick r such that:

1. $r < \varepsilon$
2. The only distinct root of $p(z)$ in $N(\alpha_k, r)$ is α_k.
3. No other root α_j of $p(z)$ lies on the border of $N(\alpha_k, r)$.

According to Rouché's Theorem the polynomials $p(z)$ and $q(z)$ will have the same number of roots in $N(\alpha_k, r)$ if $|h(z)| < |p(z)|$ on the border of $N(\alpha_k, r)$.

Let L be a lower bound of $|p(z)|$ on the border of $N(\alpha_k, r)$. Then $L > 0$ since $p(z)$ is a continuous function that is non-zero on the border of $N(\alpha_k, r)$.

All that is left to do now is to show that there is a δ such that $|w_k| < \delta \implies |h(z)| < L$ when z lies on the border of $N(\alpha_k, r)$. Let $s = |\alpha_k| + r$. Then s is an upper bound for $|z|$ when z lies on the border of $N(\alpha, r)$. Let

$$U = \delta \frac{1 - s^{n+1}}{1 - s}.$$

U is an upper bound for $|h(z)|$ on the border of $N(\alpha_k, r)$ since

$$\delta\frac{1-s^{n+1}}{1-s} = \sum_{k=0}^{n} \delta s^k \geq \sum_{k=0}^{n} |w_k||z|^k \geq |\sum_{k=0}^{n} w_k z^k| = |h(z)|.$$

The upper bound U is less than L if

$$\delta < L\frac{1-s}{1-s^{n+1}}. \tag{4.10}$$

□

Corollary 4. *Let α be a simple root of the polynomial $c_0 + c_1 z + \ldots + c_n z^n$. Given ϵ there exists δ such that if $|w_k| < \delta$ then $[c_0 + w_0] + \ldots + [c_n + w_n]z^n$ will have a simple root in $N(\alpha, \epsilon)$.*

4.3 Critical Points

According to the fundamental theorem of algebra the polynomial

$$p(z) = c_0 + c_1 z + \ldots + c_n z^n. \tag{4.11}$$

will have n roots assuming that $c_n \neq 0$. The roots need not be distinct.

The bivariate polynomial $f(z, w)$ can be viewed as a family of polynomials parameterized by w

$$f(z, w) = \sum_{k=0}^{n} c_k z^k. \tag{4.12}$$

If $f(z, w)$ has degree n when viewed as a polynomial in z then the generated polynomials will have degree n except in the cases where the leading coefficient $c_n(w)$ of $f(z, w)$ is zero. Since $c_n(w)$ is a polynomial this will happen at a finite number of isolated points. The generated polynomial will therefore have n roots at points where $c_n(w) \neq 0$. The roots need not be distinct.

Definition 27. *A point w_0 is a* critical point *of $f(z, w)$ if the polynomial $f(z, w_0)$ has a root of order 2 or greater or if the leading coefficient of $f(z, w_0)$ vanishes. Let the critical points of $f(z, w)$ be denoted by $Critical(f)$.*

In other words given a bivariate polynomial $f(z, w)$ of degree n when viewed as a polynomial in z a point in the parameter plane is a critical point if the generated polynomial does not have n distinct roots.

Consider a bivariate polynomial $f(z, w)$. The behavior of the roots of $f(z, w)$ when the parameter $w$approaches a point where the leading coefficient of $f(z, w)$ is zero is quite interesting. To see what happens it is necessary to introduce the following Lemmas.

Lemma 15. *If $z_0 \neq 0$ is a root of $a_0 + a_1 z + \ldots + a_{n-1} z^{n-1} + a_n z^n$ then $1/z_0$ will be a root of $a_n + a_{n-1} z + \ldots + a_1 z^{n-1} + a_0 z^n$.*

Proof. Assume that $p(z) = a_0 + a_1 z + \ldots + a_{n-1} z^{n-1} + a_n z^n$. Let $q(y) = y^n p(1/y)$ be a function defined on $\mathbb{C} \setminus \{0\}$. Expanding $q(y)$ results in

$$q(y) = a_0 y^n + a_1 y^{n-1} + \ldots + a_{n-1} y + a_n i. \tag{4.13}$$

If $1/y_0 \neq 0$ is a root of $p(z)$ then y_0 is a root of $q(z)$ since $q(y_0) = y_0^n p(1/y_0) = y_0^n \cdot 0 = 0$. In other words after performing the variable substitution $z_0 = 1/y_0$, if $z_0 \neq 0$ is a root of $p(z)$ then $1/z_0$ is a root of $q(z)$. □

Lemma 16. *Given ϵ there exists δ such that if $|w_j| < \delta$ then*

$$w_0 + w_1 z + \ldots + w_{k-1} z^{k-1} + c_k z^k + c_{k+1} + \ldots + c_n z^n$$

will have at least k roots in $N(0, \epsilon)$.

Proof. Let $p(z)$ and $q(z)$ be the polynomials

$$p(z) = w_0 + w_1 z + \ldots + w_{k-1} z^{k-1} + c_k z^k + c_{k+1} + \ldots + c_n z^n, \quad (4.14)$$

$$q(z) = c_k z^k + c_{k+1} + \ldots + c_n z^n. \quad (4.15)$$

According to Theorem 5 given ϵ there exists δ such that $p(z)$ and $q(z)$ will have the same number of roots in $N(0, \epsilon)$. But $q(z)$ will have at least k roots in $N(0, \epsilon)$ since $q(z) = z^k(c_k + c_{k+1} z + \ldots + c_n z^{n-k})$. □

Definition 28. *Let $A(c, r) = \{z \in \mathbb{C} | |z - c| > r\}$.*

Lemma 17. *Given r there exists δ such that if $|w_j| < \delta$ and $w_0 \neq 0$ then*

$$c_n + \ldots + c_k z^{n-k} + w_{k-1} z^{n-k+1} + \ldots + w_0 z^n$$

will have at least k roots in $A(0, r)$.

Proof. Let the $p(z)$ and $q(z)$ be the polynomials

$$p(z) = c_n + \ldots + c_k z^{n-k} + w_{k-1} z^{n-k+1} + \ldots + w_0 z^n \quad (4.16)$$

$$q(z) = w_0 + \ldots + w_{k-1} z^{k-1} + c_k z^k + \ldots + c_n z^n. \quad (4.17)$$

According to Lemma 17 there exists δ such that if $|w_j| < \delta$ then $q(z)$ will have at least k roots in $N(0, 1/r)$. Furthermore since $w_0 \neq 0$ none of these roots will be 0. Let $\alpha_1, \alpha_2, \ldots, \alpha_k$ be k roots of $q(z)$ such that $0 < |\alpha_k| < 1/r$. This in turn means that $1/|a_k| > r$. According to Lemma 15 $\beta_k = 1/\alpha_k$ will be a root of $p(z)$ and $|\beta_k| = 1/|\alpha_k| \geq r$.

□

In other words if w approaches a point where the top k coefficients of $f(z, w)$ approach zero then k of the roots of $f(z, w)$ will approach infinity.

4.4 Discriminant

The next concept to be introduced is that of the discriminant of a polynomial. The discriminant can be used to see if a polynomial has a multiple root. The approach used here follows the one presented by Sommese and Wampler (see [8]).

Definition 29. *A polynomial $c_n z^n + c_{n-1} z^{n-1} + \ldots + c_1 z + c_0$ is* represented by *the row vector $[c_n, c_{n-1}, \ldots, c_0]$. The row vector $[c_n, c_{n-1}, \ldots, c_0]$* represents *the polynomial $c_n z^n + c_{n-1} z^{n-1} + \ldots + c_1 z + c_0$.*

Definition 30. *Let $p(z)$ be a polynomial of degree n*

$$p(z) = c_0 + c_1 z + \ldots + c_n z^n.$$

Then $S(p, k)$ is defined as the following $k \times (n + k)$ matrix

$$\begin{bmatrix} c_n & c_{n-1} & \ldots & c_0 & 0 & \ldots & 0 \\ 0 & c_n & c_{n-1} & \ldots & c_0 & 0 & \ldots \\ & & & \vdots & & & \\ 0 & \ldots & 0 & c_n & c_{n-1} & \ldots & c_0 \end{bmatrix}. \quad (4.18)$$

Lemma 18. *Multiplication of the polynomials* $a_n z^n + a_{n-1} z^{n-1} + \ldots + a_1 z + a_0$ *and* $b_m z^m + b_{m-1} z^{m-1} + \ldots + b_1 z + b_0$ *is equal to the polynomial represented by*

$$[a_n, a_{n-1}, \ldots, a_0] S(b_m z^m + b_{m-1} z^{m-1} + \ldots + b_1 z + b_0, n+1).$$

Proof. $(a_n z^n + a_{n-1} z^{n-1} + \ldots + a_0)(b_m z^m + b_{m-1} z^{m-1} + \ldots + b_0) =$

$$\begin{aligned} a_n b_m z^{n+m} &+ a_n b_{m-1} z^{n+m-1} + \ldots + a_n b_0 z^n \\ &+ a_{n-1} b_m z^{n+m-1} + a_{n-1} b_{m-1} z^{n+m-2} + \ldots + a_{n-1} b_0 z^{n-1} \\ &+ \ldots + \\ &+ a_0 b_m z^m + a_0 b_{m-1} z^{m-1} + \ldots + a_0 b_0, \end{aligned}$$

which can be represented in row vector form by

$$[1, 1, \ldots, 1] \begin{bmatrix} a_n b_m & a_n b_{m-1} & \ldots & a_n b_0 & 0 & \ldots & 0 \\ 0 & a_{n-1} b_m & a_{n-1} b_{m-1} & \ldots & a_{n-1} b_0 & 0 & \\ & & & \vdots & & & \\ 0 & \ldots & 0 & a_0 b_m & a_0 b_{m-1} & \ldots & a_0 b_0 \end{bmatrix},$$

which is equal to

$$[a_n, a_{n-1}, \ldots, a_0] \begin{bmatrix} b_m & b_{m-1} & \ldots & b_0 & 0 & \ldots & 0 \\ 0 & b_m & b_{m-1} & \ldots & b_0 & 0 & \ldots \\ & & & \vdots & & & \\ 0 & \ldots & 0 & b_m & b_{m-1} & \ldots & b_0 \end{bmatrix}. \tag{4.19}$$

But the above expression is by definition equal to

$$[a_n, a_{n-1}, \ldots, a_0] S(b_m z^m + b_{m-1} z^{m-1} + \ldots + b_1 z + b_0, n+1). \tag{4.20}$$

□

In other words vectors and matrices can be used to represent polynomials. Polynomial addition becomes vector addition. Polynomial multiplication becomes multiplication of a vector by a matrix.

The next concept to be introduced is that of the Sylvester Matrix.

Definition 31. *Assume that*

$$p(z) = a_0 + a_1 z + \ldots + a_n z^n \tag{4.21}$$

and

$$q(z) = b_0 + b_1 z + \ldots + b_m z^m. \tag{4.22}$$

Then $Syl(p, q)$ *is an* $(n+m) \times (n+m)$ *matrix that is defined as follows*

$$Syl(p, q) = \begin{bmatrix} a_n & a_{n-1} & \ldots & a_0 & 0 & \ldots & 0 \\ 0 & a_n & a_{n-1} & \ldots & a_0 & 0 & \ldots \\ & & & \vdots & & & \\ 0 & \ldots & 0 & a_n & a_{n-1} & \ldots & a_0 \\ b_m & b_{m-1} & \ldots & b_0 & 0 & \ldots & 0 \\ 0 & b_m & b_{m-1} & \ldots & b_0 & 0 & \ldots \\ & & & \vdots & & & \\ 0 & \ldots & 0 & b_m & b_{m-1} & \ldots & b_0 \end{bmatrix}.$$

The matrix $Syl(p, q)$ *is called a* Sylvester matrix.

The Sylvester Matrix can be written as a sum of the product of simpler matrices.

Definition 32. *Let $I_L(n,k)$ be the matrix that consists of the left k columns of an $n \times n$ identity matrix.*

Definition 33. *Let $I_R(n,k)$ be the matrix that consists of the right k columns of an $n \times n$ identity matrix.*

Lemma 19. *Let $p(z)$ be a polynomial of degree n. Let $q(z)$ be a polynomial of degree m. Then*

$$Syl(p,q) = I_L(n+m,m)S(p,m) + I_R(n+m,n)S(q,n).$$

Proof. The proof is by direct calculation. □

Lemma 20. *Let $v = [a_{m-1}, a_{m-2}, \dots, a_0, b_{n-1}, b_{n-2}, \dots, b_0]$. Then*

$$vI_L(n+m,m) = [a_{m-1}, a_{m-2}, \dots, a_0] \tag{4.23}$$

and

$$vI_R(n+m,n) = [b_{n-1}, b_{n-2}, \dots, b_0]. \tag{4.24}$$

Proof. The proof is by direct calculation. □

Corollary 5. *Let a and b be two vectors of length $n+m$ defined as follows*

$$a = [a_{m-1}, a_{m-2}, \dots, a_0, 0, \dots, 0],$$
$$b = [0, \dots, 0, b_{n-1}, b_{n-2}, \dots, b_0].$$

Then

$$aI_R(n+m,n) = 0, \tag{4.25}$$
$$bI_L(n+m,m) = 0. \tag{4.26}$$

If $p(z)$ and $q(z)$ are two polynomials and there exists polynomials $f(z)$ and $g(z)$ where the degrees of $f(z)$ and $g(z)$ depend on $p(z)$ and $g(z)$ such that

$$p(z)g(z) + q(z)f(z) = 0. \tag{4.27}$$

Lemma 21. *If the polynomials $p(z)$ and $q(z)$ have a common root then there exists non zero polynomials $f(z)$ and $g(z)$ such that*

$$q(z)f(z) = p(z)g(z),$$

where $deg(f) < deg(p)$ and $deg(g) < deg(q)$.

Proof. Assume that z_0 is a root of both $p(z)$ and $q(z)$. Then $(z - z_0)$ can be factored out of $p(z)$ and $q(z)$ yielding

$$p(z) = (z - z_0)f(z)$$

and

$$q(z) = (z - z_0)g(z).$$

But then

$$q(z)f(z) = (z - z_0)g(z)f(z) = (z - z_0)f(z)g(z) = p(z)g(z).$$

□

The converse is also true.

Lemma 22. *Let $p(z)$ and $q(z)$ be two non zero polynomials. If there exists polynomials $f(z)$ and $g(z)$ such that*

$$q(z)f(z) = p(z)g(z),$$

where $deg(f) < deg(p)$ and $deg(g) < deg(q)$ then the polynomials $p(z)$ and $q(z)$ have a common root.

Proof. Assume that

$$q(z)f(z) = p(z)g(z), \tag{4.28}$$

$deg(f) < deg(p)$ and $deg(g) < deg(q)$. Then there exists polynomials $u(z)$ and $v(z)$ such that

$$q(z)u(z) = p(z)v(z), \tag{4.29}$$

where $deg(u) < deg(p)$, $deg(v) < deg(q)$ and $u(z)$ and $v(z)$ are coprime.

Then $u|p$ since $u|pv$, u and v are relatively prime and $\mathbb{C}[z]$ is a unique factorization domain. The same argument can be used to show that $v|q$.

Since $u|p$ and $deg(u) < deg(p)$ there exists a polynomial $r(z)$ such that

$$p(z) = u(z)r(z), \tag{4.30}$$

and $deg(r) > 0$. The same argument can be used to show that there exists a polynomial $s(z)$ such that

$$q(z) = v(z)s(z), \tag{4.31}$$

and $deg(s) > 0$. Substituting (4.30) and (4.31) into (4.29) results in

$$v(z)s(z)u(z) = u(z)r(z)v(z), \tag{4.32}$$

which implies that

$$r(z) = s(z). \tag{4.33}$$

The polynomial $r(z)$ has at least one root α since $deg(r) > 0$. Furthermore α is a root of both $p(z)$ and $q(z)$ since $r(z)$ is a factor of both $p(z)$ and $q(z)$. □

Lemma 21 and Lemma 22 can be combined into a Theorem.

Theorem 6. *Let $p(z)$ and $q(z)$ be two non zero polynomials. The polynomials $p(z)$ and $q(z)$ will have a common root if and only if there exists polynomials $f(z)$ and $g(z)$ such that*

$$q(z)f(z) = p(z)g(z),$$

where $deg(f) < deg(p)$ and $deg(g) < deg(q)$.

Proof. This follows from Lemma 21 and Lemma 22. □

The linear combination $g(z)p(z)+f(z)q(z)$ can be represented using a vector constructed using the coefficients of the polynomials $f(z)$ and $g(z)$ and the the Sylvester Matrix of the polynomials $p(z)$ and $q(z)$.

Lemma 23. *Let $p(z)$ be a polynomial of degree n. Let $q(z)$ be a polynomial of degree m. Let $f(z)$ and $g(z)$ be polynomials defined as*

$$f(z) = f_0 + f_1 z + \ldots + f_{n-1} z^{n-1}, \tag{4.34}$$
$$g(z) = g_0 + g_1 z + \ldots + g_{m-1} z^{m-1}. \tag{4.35}$$

Then the polynomial represented by

$$\left[g_{m-1}, g_{m-2}, \ldots, g_0, f_{n-1}, f_{n-2}, \ldots, f_0\right] Syl(p,q),$$

is equal to $g(z)p(z) + f(z)q(z)$.

Proof. Let F and G be row vectors of length $n+m$ defined as follows

$$F = [0, \ldots, 0, f_{n-1}, f_{n-2}, \ldots, f_0], \tag{4.36}$$
$$G = [g_{m-1}, g_{m-2}, \ldots, g_0, 0, \ldots, 0]. \tag{4.37}$$

Then the expression

$$\left[g_{m-1}, g_{m-2}, \ldots, g_0, f_{n-1}, f_{n-2}, \ldots, f_0\right] Syl(p,q), \tag{4.38}$$

is equal to

$$(F+G)Syl(p,q). \tag{4.39}$$

According to Lemma 19 the above expression is equal to

$$(F+G)[I_L(n+m,m)S(p,m) + I_R(n+m,n)S(q,n)]. \tag{4.40}$$

According to Corollary 5 the above expression is equal to

$$GI_L(n+m,m)S(p,m) + FI_R(n+m,n)S(q,n). \tag{4.41}$$

According to Lemma 20 the above expression is equal to

$$[g_{m-1}, g_{m-2}, \ldots, g_0]S(p,m) + [f_{n-1}, f_{n-2}, \ldots, f_0]S(p,n). \tag{4.42}$$

The previous expression represents the polynomial

$$g(z)p(z) + f(z)q(z) \tag{4.43}$$

since according to Lemma 18 the polynomial $g(z)p(z)$ is represented by

$$[g_{m-1}, g_{m-2}, \ldots, g_0]S(p,m)$$

and the polynomial $f(z)q(z)$ is represented by

$$[f_{n-1}, f_{n-2}, \ldots, f_0]S(p,m)$$

□

The next concept to be introduced is that of the resultant.

Definition 34. *The* resultant *of the polynomials $p(z)$ and $q(z)$ is defined as*

$$Res(p,q) = det(Syl(p,q)).$$

The resultant of two polynomials is zero if and only if the polynomials have a common root.

Theorem 7. *The polynomials $p(z)$ and $q(z)$ will have a common root if and only if $Res(p,q) = 0$.*

Proof. According to Theorem 6 the polynomials $p(z)$ and $q(z)$ will have a common root if and only if there exists polynomials

$$f(z) = f_{n-1}z^{n-1} + f_{n-2}z^{n-2} + \ldots + f_0 \tag{4.44}$$

and

$$g(z) = g_{m-1}z^{m-1} + g_{m-2}z^{m-2} + \ldots + g_0, \tag{4.45}$$

such that

$$g(z)p(z) - f(z)q(z) = 0. \tag{4.46}$$

According to Theorem 23 the last equation can be written in matrix form as

$$[g_{m-1}, g_{m-2}, \ldots, g_0, -f_{n-1}, -f_{n-2}, \ldots, -f_0]Syl(p,q) = 0. \tag{4.47}$$

The above equation will have a non zero solution if and only if

$$det(Syl(p,q)) = 0. \tag{4.48}$$

□

The next concept to be introduced is that of the discriminant.

Definition 35. *The* discriminant *of the polynomial $p(z)$ is defined as*

$$Dis(p) = Res\left(p, \frac{d}{dz}p\right).$$

The discriminant of a polynomial is zero if and only if it has a multiple root.

Theorem 8. *The polynomial $p(z)$ has a multiple root if and only if $Dis(p) = 0$.*

Proof. According to Corollary 3 the polynomial $p(z)$ will have a multiple root at if and only if $p(z) = 0$ and $p'(z) = 0$. According to Theorem 7 $p(z)$ and $p'(z)$ will have a common root if and only if $Res(p,p') = 0$ which by definition will happen if and only if $Dis(p) = 0$. □

4.5 Problem Polynomials

In the later section of this paper we will be interested in the distinct roots of $f(z,w)$ when $f(z,w)$ is viewed as a polynomial in z.

If the polynomial $f(z,w)$ has a multiple factor $f(z,w) = g(z,w)^k h(z,w)$ then any root of $g(z,w)$ will be a multiple root of $f(z,w)$. This means that the polynomial $f(z,w)$ will never have distinct roots for any w. From this point forward no polynomial $f(z,w)$ will have a multiple factor.

If the leading coefficient of a polynomial goes to 0 then one of it's roots will go to infinity. This means that we will run into numerical issues at and near a point where the top coefficient is 0. From now on all polynomials $f(z,w)$ will have a constant leading coefficient.

4.6 Numerical Calculation of Critical Points

Consider the following bivariate polynomial

$$f(z, w) = \sum_{k=0}^{n} c_k(w) z^k, \tag{4.49}$$

where $c_1(w), c_2(w), \ldots, c_n(w)$ are polynomials in w.

By definition a point c is a critical point if the leading coefficient of the polynomial $f(z, c)$ vanishes or if the polynomial $f(z, c)$ has a root of order 2 or greater.

Finding points w where the equation

$$c_n(w) = 0 \tag{4.50}$$

is satisfied can be done using the algorithm described in Subsection 3.4.

The Discriminant can be used to find points where c where the polynomial $f(z, c)$ has a root of order 2 or greater. Let $p_c(z)$ be the polynomial $f(z, c)$. According to Theorem X the polynomial $p_c(z)$ will have a root of order 2 or greater if and only if $Dis(p_c) = 0$. By definition

$$Dis(p_c) = Res\left(p_c, \frac{d}{dz}p_c\right) = det\left(Syl\left(p_c, \frac{d}{dz}p_c\right)\right). \tag{4.51}$$

Hence the polynomial $p_c(z)$ will have a root of order 2 or greater if and only if

$$det\left(Syl\left(p_c, \frac{d}{dz}p_c\right)\right) = 0. \tag{4.52}$$

The above equation is a polynomial equation in w and can be solved numerically using the algorithm described in Subsection 3.4.

5 Homotopies

In this section the concept of a homotopy is introduced. A homotopy is closely related to the concept of a curve being continuous deformable into another curve. If a curve $z(t)$ is defined on an interval $[a, b]$ a variable substitution can be performed to define the curve on the interval $[0, 1]$ instead. The composition of two curves can also be scaled to be defined on the interval $[0, 1]$. The term path will be used interchangeably with the term curve with the exception that all paths are defined on the interval $[0, 1]$.

5.1 Definition and Basic Properties

First the concept of a homotopy is defined.

Definition 36. *A function $H(t, s) : [0, 1] \times [0, 1] \to \mathbb{C}$ is said to be a* homotopy *if it has the following properties:*

1. *$H(t, s)$ is continuous.*
2. *$H(0, s)$ is independent of s.*
3. *$H(1, s)$ is independent of s.*

A curve is homotopic to another curve in U if the first curve can be continuously deformed into the other curve without leaving U.

Definition 37. *Two paths $x(t)$ and $y(t)$ are said to be* homotopic in U *if there exists a homotopy $H(t, s)$ with the following properties:*

1. *$H(t, 0) = x(t)$.*
2. *$H(t, 1) = y(t)$.*
3. *If $(t_0, s_0) \in [0, 1] \times [0, 1]$ then $H(t_0, s_0) \in U$.*

$H(t, s)$ is said to be a homotopy in U between *the path $x(t)$ and the path $y(t)$.*

Homotopies can be composed.

Definition 38. *If two homotopies $H_1(t, s)$ and $H_2(t, s)$ satisfy the condition*

$$H_1(t, 1) = H_2(t, 0)$$

then the composition *$H(t, s)$ of $H_1(t, s)$ and $H_2(t, s)$ is defined as*

$$H(t, s) = \begin{cases} H_1(t, 2s), & \text{if } 0 \le s \le 0.5 \\ H_2(t, 2s - 1), & \text{if } 0.5 < s \le 1 \end{cases}$$

The composition of the homotopies $H_1(z, t)$ and $H_2(z, t)$ is denoted by $H_1 \circ H_2$.

Lemma 24. *$H_1 \circ H_2$ is a homotopy.*

Proof. Let $H = H_1 \circ H_2$. It is obvious that H is a continuous function. $H_1(0, s) = c$ since H_1 is a homotopy. $H_2(0, s) = d$ since H_2 is a homotopy. But then $c = d$ since $H_1(0, 1) = H_2(0, 0)$. It is now obvious that $H(0, s) = c$. The same type of argument can be used to show that $H(1, 0)$ is constant. But then $H(t, s)$ is a homotopy. □

Finally homotopies have inverses. If $v(t)$ is homotopic to $u(t)$ then $u(t)$ will be homotopic to $v(t)$.

Definition 39. *Let the* inverse *of the path $w(t)$ be the path $w(1-t)$. Denote the inverse of the path w by w^{-1}.*

Definition 40. *Let $H(t,s)$ be a homotopy between the paths $w(t)$ and $u(t)$. The* inverse *of the homotopy $H(t,s)$ is denoted by $H^{-1}(t,s)$ and is defined as*

$$H^{-1}(t,s) = H(t,1-s).$$

Lemma 25. *If $H(t,s)$ is a homotopy in U between $w(t)$ and $u(t)$ then $H^{-1}(t,s)$ is a homotopy in U between $u(t)$ and $w(t)$.*

Proof. $H^{-1}(t,s)$ is a continuous function since $H(t,s)$ is a continuous function. $H^{-1}(t,0) = H(t,1) = u(t)$. $H^{-1}(t,1) = H(t,0) = w(t)$. □

5.2 Special Paths

Definition 41. *Let $L(a,b)$ be the line segment from a to b.*

Definition 42. *Let $C(c,r)$ be the circle with center c and radius r.*

Definition 43. *Let $D(c,r)$ be the disc with center c and radius r.*

Definition 44. *Let $I(s,c,r)$ be the intersection point of C(c,r) and L(s,c).*

Definition 45. *Let $B(s,c,r)$ be the path composed of the following three paths.*

1. *The line segment from s to $I(s,c,r)$.*
2. *The path around the circle $C(c,r)$ starting at I(s,c,r) and going in the counter clockwise direction.*
3. *The line segment from $I(s,c,r)$ to s.*

This type of path is known as a circle path*.*

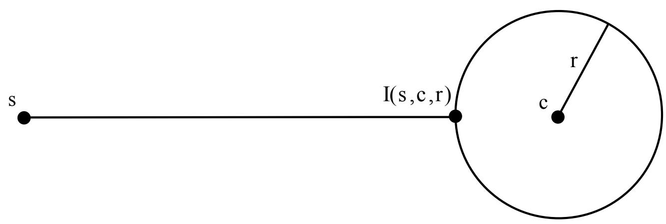


Figure 9: Circle path

Definition 46. *Let $A(s,c,r,n)$ be the path composed of the following three paths.*

1. *The line segment from s to $I(s,c,r)$.*
2. *The path around a regular n-gon inscribed inside the circle $C(c,r)$ starting at a vertex located at I(s,c,r) and going in the counter clockwise direction.*

3. *The line segment from $I(s,c,r)$ to s.*

This type of path is known as a n-gon path.

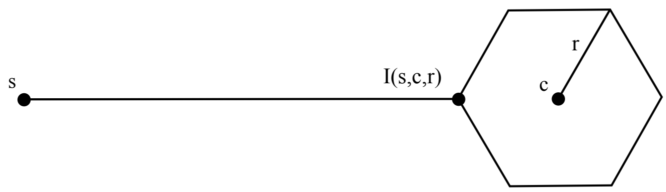


Figure 10: N-gon path

Definition 47. *Let $T(s,c,r)$ be the path $A(s,c,r,3)$. This type of path is known as a* triangle path.

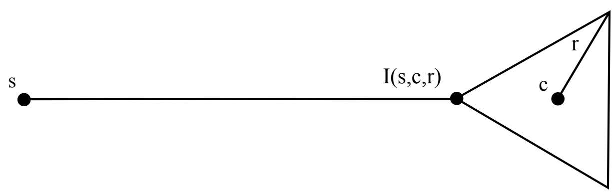


Figure 11: Triangle path

Definition 48. *Consider the triangle path $T(s,c,r)$. Number the vertexes of the triangle starting at $I(s,c,r)$ and going counter clockwise around the triangle. The names of the vertices of the triangle are defined as follows:*

1. *The first vertex is called the* inner vertex *of the triangle path.*

2. *The second vertex is called the* right vertex *of the triangle path.*

3. *The third vertex is called the* left vertex *of the triangle path.*

Lemma 26. *Let $U = \mathbb{C} \backslash \{c_1, c_2, \ldots, c_n\}$. Let c be one of the points $c_1, c_2, \ldots, c_n$. If c is the only one of the points $c_1, c_2, \ldots, c_n$ in $D(c,r)$ then the n-gon path A(s,c,r,n) is homotopic to the circle path B(s,c,r).*

Proof. Create a homotopy $H(s,t)$ defined as follows:

1. Each point on the n-gon is linearly interpolated radially from c until it hits the circle.

2. The points on the line segment part of the path do not move.

This is a homotopy in U since none of the points $c_1, c_2, \ldots, c_n$ lie between the n-gon and the circle. □

5.3 Homotopies of Closed Paths

It can be shown that an arbitrary closed path is homotopic in $\mathbb{C} \setminus \{c_1, c_2, \ldots, c_n\}$ to a path that is either a point or a composition of circle paths and inverse circle paths that all have a common starting point.

The approach that will be used is to construct several homotopies that when composed turns an arbitrary closed path into a path that is either a point or the composition of triangle paths and inverse triangle paths. Finally it is shown that a path that is the composition of triangle paths and inverse triangle paths is homotopic to a path that is the composition of circle paths and inverse circle paths.

5.3.1 Convex Hull Homotopy

The first homotopy to be introduced is the convex hull homotopy. Before the convex hull homotopy can be defined it is necessary to introduce the concept of the convex hull of a collection of triangle paths.

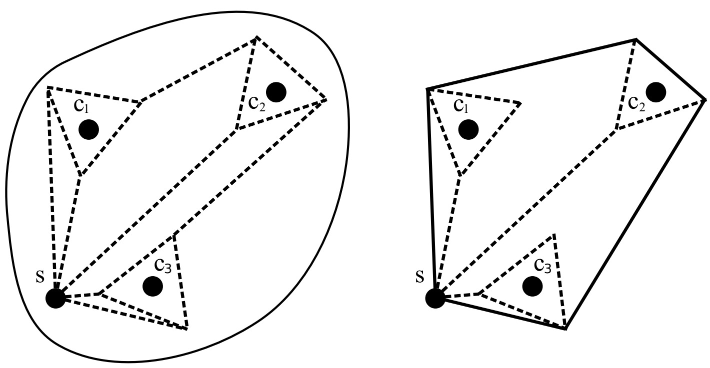
Definition 49. *Let $CH(s, r, c_1, c_2, \ldots, c_n)$ be the convex hull of s and the right and left vertices of the triangle paths $T(s, c_1, r), T(s, c_2, r), \ldots, T(s, c_n, r)$.*

Definition 50. *Let $Cent(P)$ be the centroid of the polygon P.*

Definition 51. *Let $H_{CH}[s, r, c_1, c_2, \ldots, c_n]$ be the homotopy defined as follows: Let $P = CH(s, r, c_1, c_2, \ldots, c_n)$.*

1. *All points that lie on P remain fixed.*
2. *All points outside P are interpolated radially towards $Cent(P)$ until they hit P.*

This type of homotopy is called a convex hull homotopy*.*



(a) $H_{CH}(t, 0)$ (b) $H_{CH}(t, 1)$

Figure 12: An arbitrary path is homotopic to a path that lies in the complex hull of the triangle paths.

5.3.2 Radial Triangle Homotopy

The next homotopy to be introduced is the radial triangle homotopy. It moves all paths that lie in a triangle radially outwards from the center of the triangle.

Definition 52. *Consider the triangle path $T(s, c, r)$. Let the homotopy $H_R[s, c, r]$ be defined as follows:*

1. *Points outside and on the triangle part of $T(s, c, r)$ remain fixed.*
2. *Points inside the triangle part of $T(s, c, r)$ are interpolated radially from c until they hit the triangle.*

This type of homotopy is called a radial triangle homotopy*.*

5.3.3 Triangle Homotopy

The next homotopy to be introduced is the triangle homotopy. It moves all paths that lie in a triangle to a polyline consisting of two of the edges of the triangle.

Definition 53. *Let T be the triangle with vertexes a, b, o. Let $L_{ab}(t)$ be the path*

$$L_{ab}(t) = (1 - t)a + tb.$$

Let the $u(t)$ and $v(t)$ be the following paths:

$$u(t) = L_{ab}(t), \tag{5.1}$$

$$v(t) = L_{ao}(t) \circ L_{ob}(t). \tag{5.2}$$

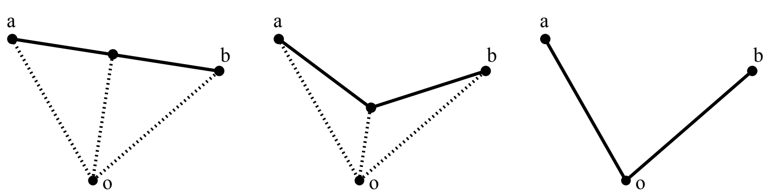
Let $H_T(t, s)$ be a homotopy defined as follows:

1. *Points outside the triangle T remain fixed.*
2. *For each $t \in [0, 1]$ the points on the line segment $L(u(t), v(t))$ are linearly interpolated along the line segment until they hit $v(t)$.*

This type of homotopy is known as a triangle homotopy*.*

Lemma 27. *Let $U = \mathbb{C}\ \{c_1, c_2, \ldots, c_n\}$. Let T be a triangle with vertices a, b, o that does not contain any of the points $c_1, c_2, \ldots, c_n$. A path in T is homotopic in U to a path that lies on the poly line a, o, b.*

Proof. Let $H_T(t, s)$ be the triangle triangle homotopy that moves the line segment $L(a, b)$ to the poly line a, o, b. A path $w(t)$ is homotopic in U to the poly line a, o, b since none of the points $c_1, c_2, \ldots, c_n$ lie in the triangle. □



(a) $H_T(t, 0)$ (b) $H_T(t, 0.5)$ (c) $H_T(t, 1)$

Figure 13: A path in the triangle with vertexes a, b, o is homotopic to a path that lies in the polyline a, o, b.

5.3.4 Convex Hull to C_s Homotopy

Before the next homotopy can be introduced it is necessary to define the set C_s.

Definition 54. *Let $w(t) = T(s, c, r)$. Let $T_s(s, c, r)$ be the set defined as follows*

$$T_s(s, c, r) = \{z \in \mathbb{C} | z \neq c \textit{ and } z \in L(w(t), c) \textit{ for } t \in [0, 1]\}.$$

Let $C_s(s, r, c_1, c_2, \ldots, c_n)$ be the set defined as follows

$$C_s(s, r, c_1, c_2, \ldots, c_n) = \bigcup_{k=1}^{n} T_s(s, c_k, r).$$

Informally the set $T_s(s, c, t)$ can be thought of as the image of a triangle path where the interior points of the triangle also belong to the set with the exception that the center point c of the triangle does not belong to the set. The set C_s is just the union of several T_s sets.

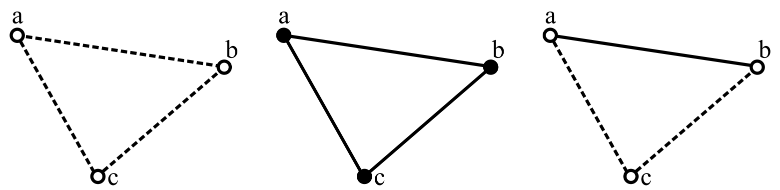
The next homotopy to be introduced is the convex hull to C_s homotopy. It takes a path that lies in the convex hull of a given set of triangle paths and moves it to a path that lies in C_s.

Definition 55. *Consider a triangle with vertexes a, b, c.*

1. *A set that contains all the points inside and on the triangle is called a* closed triangle*.*

2. *A set that contains all the points inside the triangle but none of the points on the edges of the triangle is called an* open triangle*.*

3. *A closed triangle where two of the edges have been removed is called a* semi-open triangle*.*



(a) Open Triangle (b) Closed Triangle (c) Semi-Open Triangle

Figure 14: Different types of triangles.

Lemma 28. *Let $U = \mathbb{C} \setminus \{c_1, c_2, \ldots, c_n\}$. Let $w(t)$ be a closed path that lies in $CH(s, r, c_1, c_2, \ldots, c_n)$ and that does not intersect any of the points $c_1, c_2, \ldots, c_n$. Then $w(t)$ is homotopic in U to a path that lies in $C_s(s, r, c_1, c_2, ..., c_n)$.*

Proof. By removing a finite number of open triangles the set $CH(s, r, c_1, \ldots, c_n)$ can be transformed into the set $C_s(s, r, c_1, c_2, \ldots, c_n)$ in such a manner that each intermediate set is closed.

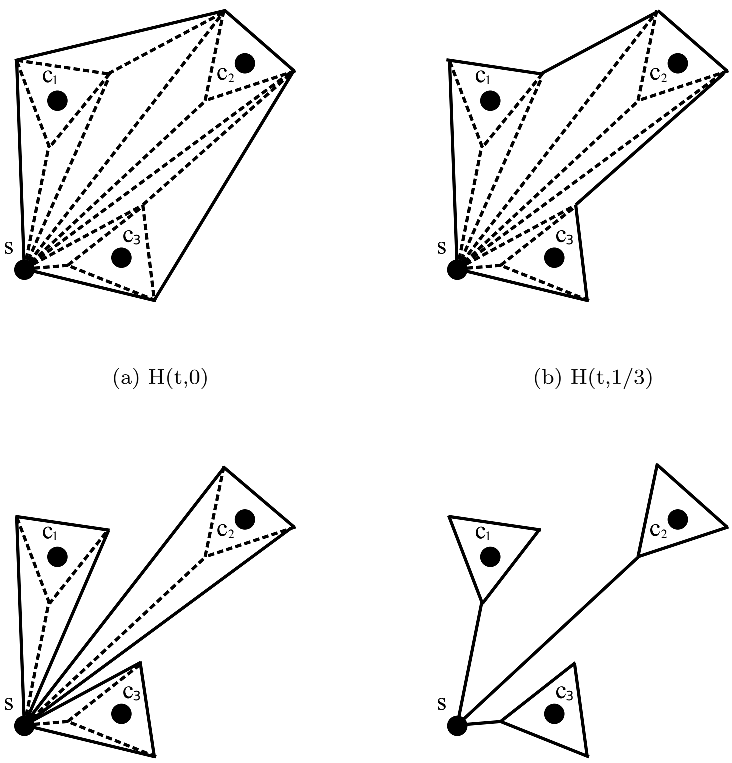
Let $T_1, T_2, \ldots, T_m$ be such a sequence of semi-open triangles. Let $S_0, S_1, \ldots, S_m$ be a sequence of $m+1$ sets defined by

$$S_0 = CH(s, r, c_1, c_2, \ldots, c_n), \tag{5.3}$$
$$S_k = S_k \setminus T_k. \tag{5.4}$$

Then $S_m = C_s(s, r, c_1, \ldots, c_n)$ since

$$S_m = S_0 \setminus \bigcup_{k=1}^{n} T_k \tag{5.5}$$
$$= CH(s, r, c_1, \ldots, c_n) \setminus \bigcup_{k=1}^{n} T_k \tag{5.6}$$
$$= C_s(s, r, c_1, \ldots, c_n). \tag{5.7}$$

According to Lemma 27 a path that lies in S_k will be homotopic in U to a path that lies in S_{k+1}. This in turn implies that a path that lies in S_0 will be homotopic in U to a path that lies in S_m. □



(a) H(t,0)

(b) H(t,1/3)

(c) H(t,2/3)

(d) H(t,1)

Figure 15: H(t,s_0) as s_0 moves from 0 to 1.

5.3.5 C_s to C_w Homotopy

Before the next homotopy can be introduced it is necessary to define the set C_w.

Definition 56. *Let $w(t) = T(s, c, r)$. Let $T_w(s, c, r)$ be the set defined as follows*

$$T_w(s, c, r) = \{z \in \mathbb{C} | z = w(t) \text{ for } t \in [0, 1]\}.$$

Let $C_w(s, r, c_1, c_2, \ldots, c_n)$ be the set defined as follows

$$C_w(s, r, c_1, c_2, \ldots, c_n) = \bigcup_{k=1}^{n} T_w(s, c_k, r).$$

In other words the set C_w is the composition of images of triangle paths. The next homotopy to be introduced is the C_s to C_w homotopy. It takes a path that lies in C_s and moves it to a path that lies in C_w.

Lemma 29. *Let $U = \mathbb{C} \setminus \{c_1, c_2, \ldots, c_n\}$. Let $w(t)$ be a closed path that lies in $C_s(s, r, c_1, c_2, \ldots, c_n)$ and that does not intersect any of the points $c_1, c_2, \ldots, c_n$. Then $w(t)$ is homotopic in U to a path that lies in $C_w(s, r, c_1, c_2, \ldots, c_n)$.*

Proof. Use a radial triangle homotopy on each of the triangle parts of C_s. □

5.3.6 Composite Homotopy

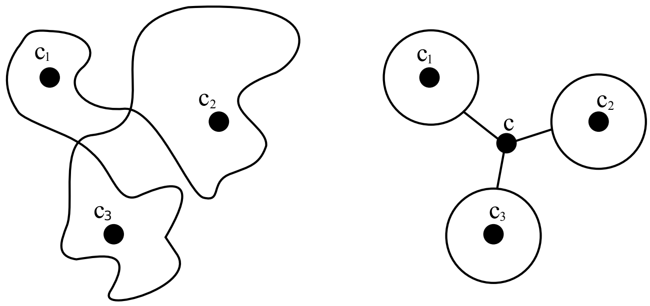
Finally all the pieces are put together to give the following Theorem.

Theorem 9. *A closed path $w(t)$ is homotopic in $\mathbb{C} \setminus \{c_1, c_2, \ldots, c_n\}$ to a path that is either a point or the composition of circle paths and inverse circle paths that all share a common starting point and where each circle path or inverse circle path encloses one and only one of the points $c_1, c_2, \ldots, c_n$.*

A formal proof of the above theorem will not be provided. Instead a sketch of the proof will be given. Using a convex hull homotopy it can be shown that an arbitrary path in $U = \mathbb{C} \setminus \{c_1, c_2, \ldots, c_n\}$ is homotopic to a path that lies in the convex hull of triangle paths starting at s and centered on the points c_k for $k = 1, 2, \ldots, n$. The radius of the circle that the triangles are inscribed in can be chosen such that each triangle path starting at s and centered on c_k is homotopic to a circle path starting at s and centered on c_k where the circles do not intersect.

Next it can be shown that a path that lies in the convex hull of the triangle paths is homotopic to a path that lies in C_s. Furthermore it can be shown that a path that lies in C_s that does not intersect any of the points c_k is homotopic to a path that lies in C_w.

Since C_w is just the composition of the images of triangle paths it is obvious that a path that lies in C_s is homotopic to a path that is either a point or a composition of triangle paths and inverse triangle paths. Finally each triangle path is homotopic to a circle path since the radius of the circle that the triangles are inscribed in were chosen such that the circles do not intersect.



(a) Arbitrary Closed Path (b) Composition of Circle Paths

Figure 16: An arbitrary path is homotopic to a composition of circle paths and inverse circle paths.

6 Following Roots

This section deals with how the roots of $f(z,w)$ move when the parameter w is moved in the parameter plane. A version of the Implicit Function Theorem that works for complex functions is introduced. Continuation of an analytic function along a path is defined. The Monodromy Theorem is stated and proved. Finally the behavior of the roots when the parameter w is moved along circle paths around critical points of $f(z,w)$ is discussed.

6.1 Implicit Function Theorem

An important theorem in Complex Analysis is the Implicit Function Theorem.

Implicit Function Theorem. *Let $f(z,w)$ be a function that is analytic in $D(z_0,r_1)\times D(w_0,r_2)$. If the function $f(z,w)$ satisfies the conditions*

$$f(z_0,w_0)=0 \text{ and } \frac{\partial f(z_0,w_0)}{\partial z}\neq 0, \tag{6.1}$$

then there exists a disc $D(w_0,r_3)$ and an analytic function $F(w)$ defined on $D(w_0,r_3)$ such that

$$f(F(w_0),w_0)=0. \tag{6.2}$$

For a proof of the above theorem see [2]. The following theorem shows that in the case that $f(z,w)$ is a bivariate polynomial a simple root of $f(z,w)$ will vary continuously with w and that a root of order 2 or greater will in some sense vary continuously with w as well.

Theorem 10. *Let $f(z,w)$ be a bivariate polynomial. Given ϵ there exists δ such that if $|w-w_0|<\delta$ then $f(z,w)$ and $f(z,w_0)$ will have the same number of roots in $N(\alpha_k,\epsilon)$ where α_k is a root of $f(z,w_0)$.*

Proof. According to Theorem 5 there exists δ_c such that if $|c_k(w)-c_k(w_0)|<\delta_c$ then $f(z,w)$ and $f(z,w_0)$ will have the same number of roots in $N(\alpha_k,\epsilon)$.

Each $c_k(w)$ is a continuous function and therefore there exists δ such that if $|w-w_0|<\delta$ then $|c_k(w)-c_k(w_0)|<\delta_c$. This in turn means that $f(z,w)$ and $f(z,w_0)$ will have the same number of roots in $N(\alpha_k,\epsilon)$ if $|w-w_0|<\delta$. □

Corollary 6. *A simple root of $f(z,w)$ will vary continuously with w.*

6.2 Analytic Continuation

The next concept to be defined is that of an analytic continuation of an analytic function. Before analytic continuation can be defined it is necessary to introduce the Identity Theorem.

Identity Theorem. *Let $f(z)$ and $g(z)$ be two analytic functions defined on a connected region R_1. Let R_2 be a subregion of R_1. If $f(z)=g(z)$ on R_2 then $f(z)=g(z)$ on R_1.*

For a proof of the identity theorem see [1].

The following definition of analytic continuation is from Wikipedia (see [11]).

Definition 57. *Let $f(z)$ an analytic function defined on an open subset U of the complex plane C. If V is a larger open subset of C, containing U, and $F(z)$ is an analytic function defined on V such that*

$$F(z) = f(z) \text{ for } z \in U, \tag{6.3}$$

then $F(z)$ is called an analytic continuation *of $f(z)$.*

The next concept to be introduced is that of an analytic continuation of a function along a curve. From now on curves will be scaled so that they are defined on the interval $[0, 1]$. Before analytic continuation can be defined it is first necessary to define the concept of an analytic function element.

Definition 58. *If $f(z)$ is an analytic function defined in a neighborhood D then the ordered pair (f, D) is called an* analytic function element*. If $w \in D$ then (f, D) is an analytic function element* at w*. Two analytic function elements (f_1, D_1) and (f_2, D_2) at w are said to be* equivalent at w *if $f_1 = f_2$ on $D_1 \cap D_2$. The remaining definitions and theorems in this section are based on theorems and definitions presented by Taylor (see [9]).*

Definition 59. *Let $w(t)$ be a curve. Let (f_0, D_0) be an analytic function element at $w(0)$. Suppose there exists a sequence of analytic function elements*

$$(f_1, D_1), (f_2, D_2), \ldots, (f_n, D_n), \tag{6.4}$$

and a partition of the interval [0,1]

$$0 = t_0 < t_1 < \ldots < t_{n+1} = 1, \tag{6.5}$$

that satisfy the following conditions:

1. *If $t \in [t_k, t_{k+1}]$ then $w(t) \in D_k$.*
2. *$D_k \cap D_{k+1} \neq \emptyset$.*
3. *$f_k = f_{k+1}$ on $D_k \cap D_{k+1}$.*

Then (f_n, D_n) is called an analytic continuation of (f_0, D_0) along $w(t)$.

The reader might be concerned that the above definition depends on the choice of partition and sequence of analytic function elements. The following Theorem should put those concerns to rest.

Theorem 11. *Let $w(t)$ be a curve. Let (f_0, D_0) be an analytic function element at $w(0)$. Then any two analytic continuations along w are equivalent as analytic function elements at $w(1)$.*

Proof. Let (f_n, D_n) be an analytic continuation of (f_0, D_0) along the curve $w(t)$ where $0 = t_0 < t_1 < \ldots < t_{n+1}$ is the associated partition of the interval $[0, 1]$ and $(f_1, D_1), (f_2, D_2), \ldots, (f_n, D_n)$ is the associated sequence of analytic function elements.

Given an arbitrary refinement of the partition of $[0, 1]$ a new sequence of analytic function elements can be constructed in such a manner that the new partition of $[0,1]$ and the new sequence of analytic function elements determines the same analytic continuation of (f_0, D_0) along the curve $w(t)$ as the original partition and sequence of analytic function elements does.

If the interval $[t_k, t_{k+1}]$ is refined into j_k subintervals then a new sequence of analytic function elements is created by adding the analytic function element (f_k, D_k) to the sequence j_k times for $k = 0, 1, \ldots, n$. It is obvious that the new partition of $[0,1]$ and the new sequence of analytic function elements determines the same analytic continuation along the curve $w(t)$ as the original partition and sequence of analytic function elements does.

It is now obvious that the new partition of $[0,1]$ and the new sequence of analytic function elements determines the same analytic continuation along $w(t)$ as the original partition of $[0,1]$ and the original sequence of analytic function elements.

Given two different analytic continuations of (f_0, D_0) along $w(t)$ it is possible to refine both of the associated partitions so that both analytic continuations use the same partition

$$0 = s_0 < s_1 < \ldots < s_{m+1} = 1. \tag{6.6}$$

Suppose that one of the analytic continuations is determined by the following sequence of analytic function elements

$$(f_1, D_1), (f_2, D_2), \ldots, (f_m, D_m). \tag{6.7}$$

Furthermore suppose that the other analytic continuation is determined by the following sequence of analytic function elements

$$(g_1, B_1), (g_2, B_2), \ldots, (g_m, B_m) \tag{6.8}$$

and let $(g_0, B_0) = (f_0, D_0)$. Let V_j be the set

$$V_j = B_j \cap D_j, \tag{6.9}$$

for $j = 0, 1, \ldots, m$. Then V_j is an open connected set containing $w(t_j)$ and $w(t_{j+1})$. To prove the theorem it is sufficient to show that $f_m = g_m$ on V_m.

By definition $f_0 = g_0$ on V_0. Suppose for $j < m$ that $f_j = g_j$ on V_j. Then there is a neighborhood $N_1 \subset V_j$ of $w(t_{j+1})$ where $f_j = g_j$.

The open sets D_j and D_{j+1} both contain $w(t_{j+1})$ and therefore so does $D_j \cap D_{j+1}$. Furthermore $f_j = f_{j+1}$ on $D_j \cap D_{j+1}$. Therefore there is a neighborhood $N_2 \subset D_j \cap D_{j+1}$ of $w(t_{j+1})$ where $f_j = f_{j+1}$.

The open sets B_j and B_{j+1} both contain $w(t_{j+1})$ and therefore so does $B_j \cap B_{j+1}$. Furthermore $g_j = g_{j+1}$ on $D_j \cap D_{j+1}$. Therefore there is a neighborhood $N_3 \subset B_j \cap B_{j+1}$ of $w(t_{j+1})$ where $g_j = g_{j+1}$.

Let $N = N_1 \cap N_2 \cap N_3$. The following equations hold on N

$$f_j = g_j, \tag{6.10}$$

$$f_j = f_{j+1}, \tag{6.11}$$

$$g_j = g_{j+1}. \tag{6.12}$$

Substituting (6.11) and (6.12) into (6.10) results in that the following equation holds on the set N

$$f_{j+1} = g_{j+1}. \tag{6.13}$$

According to the Identity Theorem this means that $f_{j+1} = g_{j+1}$ on V_j. □

6.3 Function Definition

Let $f(z,w)$ be a bivariate polynomial. Let α be a simple root of the polynomial $f(z,w_0)$. According to the Implicit Function Theorem there exists a function $F(w)$ defined in a neighborhood U of w_0 such that for $w \in U$

$$f(F(w),w) = 0. \tag{6.14}$$

Suppose that $w(t)$ is a path that starts at w_0 and avoids the critical points of $f(z,w)$. Then the analytic function element (F,U) can be analytically continued along the path $w(t)$.

Definition 60. *If $f(z,w)$ is a bivariate polynomial and $w(t)$ is a path that avoids the critical points of f then $R[f,w]$ is the bijection between the roots of $f(z,w(0))$ and $f(z,w(1))$ generated by performing an analytic continuation along the curve $w(t)$ of the function $F(w)$ that is implicitly defined by*

$$f(F(w),w) = 0. \tag{6.15}$$

Definition 61. *Let $Root(f,w,\alpha) := R[f,w](\alpha)$ be a complex valued function where:*

$f(z,w)$ is a bivariate polynomial.

$w(t)$ is a path that avoids the critical points of f.

α is a root of $f(z,w(0))$.

By convention the complex plane that the parameter w lies in will be called the *parameter plane.* The complex plane that the roots of $f(z,w)$ lie in will be called the *root plane.*

6.4 Monodromy Theorem

An important theorem in complex analysis is the Mondoromy Theorem. Before the Monodromy Theorem can be proved it is necessary to introduce a couple of Lemmas. The Lemmas and Theorems in this section are based on material presented by Taylor (see [9]).

Given ϵ for any path $H(t,s)$ in a homotopy there is a δ such that if r is within δ from s then the distance between the curves $H(t,s)$ and $H(t,r)$ will be less than ϵ.

Lemma 30. *If $H(t,s)$ is a homotopy then give ϵ there exists δ such that if $|s-s_0|<\delta$ then $|H(t,s)-H(t,s_0)|<\epsilon$.*

Proof. The homotopy $H(t,s)$ is a continuous function defined on $[0,1]\times[0,1]$. The set $[0,1]\times[0,1]$ is compact. The homotopy $H(t,s)$ is therefore uniformly continuous. Since $H(t,s)$ is uniformly continuous given ϵ there exists δ such that

$$|(t_x,s_x)-(t_y,s_y)|<\delta \implies |H(t_x,s_x)-H(t_y,s_y)|<\epsilon. \tag{6.16}$$

If t remains fixed given ϵ there exists δ such that

$$|s_x-s_y|<\delta \implies |H(t,s_x)-H(t,s_y)|<\epsilon. \tag{6.17}$$

□

Given an analytic function element (f_0, D_0) and ϵ for any path $H(t,s)$ in a homotopy there is a δ such that if r is within δ of s then the analytic continuation along $H(t,s)$ and $H(t,r)$ of (f_0, D_0) will be equivalent.

Lemma 31. *Let $H(t,s)$ be a homotopy. Let (f_0, D_0) be an analytic function element. Furthermore suppose that it is possible to perform an analytic continuation of (f_0, D_0) along all curves in $H(t,s)$. Then for every $r \in [0,1]$ there is a δ such that if $|s-r| < \delta$ then the analytic continuation of (f_0, D_0) along $H(t,s)$ is equivalent at $H(1,s) = H(1,r)$ to the analytic continuation of (f_0, D_0) along $H(t,r)$.*

Proof. Let (f_n, D_n) be the analytic continuation of (f_0, D_0) along H(t,r) that is determined by the partition $0 = t_0 < t_1 < \ldots < t_{n+1}$ and the sequence of analytic function elements $(f_1, D_1), (f_2, D_2), \ldots, (f_n, D_n)$.

If $t \in [t_j, t_{j+1}]$ then $H(t,r) \in D_j$ for $j = 0, 1, \ldots, n$. For $j = 0, 1, \ldots, n$ let ϵ_j be the minimum distance from the compact set $\{H(t,r) | t \in [t_j, t_{j+1}]\}$ to the boundary of the disc D_j.

For $t \in [t_j, t_{j+1}]$ if $|H(t,s) - H(t,r)| < \epsilon_j$ then $H(t,s) \in D_j$. Let $\epsilon = \min\{\epsilon_0, \epsilon_1, \ldots, \epsilon_n\}$. According to Lemma X there is a δ such that if $|s - r| < \delta$ then $|H(t,s) - H(t,r)| < \epsilon$. But if $|H(t,s) - H(t,r)| < \epsilon$ then $H(t,s) \in D_j$ for $t \in [t_j, t_{j+1}]$.

In other words given an analytic continuation of (f_0, D_0) along $H(t,r)$ defined by the partition $0 = t_0 < t_1 < \ldots < t_{n+1} = 1$ and the sequence $(f_1, D_1), (f_2, D_2), \ldots, (f_n, D_n)$ of analytic function elements there is a δ such that if $|s - r| < \delta$ then (f_n, D_n) is an analytic continuation of (f_0, D_0) along the curve $H(t,s)$. According to Theorem 11 any other analytic continuation of (f_0, D_0) along $H(t,s)$ will be equivalent to (f_n, D_n) at $H(1,s)$. □

The analytic continuation of an analytic function element along two paths that are homotopic will be equivalent if it is possible to perform an analytic continuation of the analytic function element for all paths in the homotopy.

Monodromy Theorem. *Let $H(t,s)$ be a homotopy. Suppose that it is possible to preform an analytic continuation of (f_0, D_0) along every path in $H(t,s)$. Then the analytic continuation of (f_0, D_0) along $H(t,0)$ is equivalent to the analytic continuation of (f_0, D_0) along $H(t,1)$.*

Proof. Let $h(s)$ be a function defined on the interval $[0,1]$ where $h(s) = 0$ if the analytic continuation of (f_0, D_0) along $H(t,s)$ is equivalent to the analytic continuation of (f_0, D_0) along $H(t,0)$, otherwise $h(s) = 1$. According to Lemma 31 the function $h(s)$ is continuous on $[0,1]$.

Hence $h(s)$ is equal to the constant function 1 or the constant function 0. But then $h(s) = 0$ since $h(0) = 0$. □

7 Numerical Root Following

In this section a numerical algorithm for following the roots of $f(z, w)$ as w is moved along a piecewise linear path is described. Euler Prediction is introduced. Newton Correction is described. A modified version of Newtons Method that has much stricter constraints on the error of the calculated root is developed. A modified algorithm that includes validation of the results is developed. Rouché Verification is described. The material in this section is based on material due to Sommese and Wampler (see [8]) and Morgan (see [3]).

7.1 Paths

To simplify the development of the algorithms the paths that will be considered are paths that are constructed using a finite number of line segments. It is enough to develop an algorithm for following the roots when the parameter w is moved along a line segment. The algorithm can then be applied to each line segment in the path with the initial roots being the roots that were calculated while traversing the previous line segment. Specifically this means that the algorithm will be able to handle N-gon paths since an N-gon path is the composition of a finite number of line segments.

7.2 Initial Roots

To be able to follow the roots as w moves along the path it is necessary to know the roots when w is at the start of the path. If s is the starting point of the path then the initial roots can be found by solving the following equation

$$f(z, s) = 0. \tag{7.1}$$

The above equation can be solved numerically using the algorithm described in the section on Numerical Root Finding.

7.3 Davidenko Differential Equation

According to the Implicit Function Theorem the function $f(z, w)$ will implicitly define a function $\alpha(w)$ in a neighborhood of w_0 assuming that

$$\frac{\partial f(z, w_0)}{\partial z} \neq 0. \tag{7.2}$$

Definition 62. *Let the partial derivatives of the function $f(z, w)$ be denoted as follows*

$$f_z(z, w) = \frac{\partial f(z, w)}{\partial z}, \tag{7.3}$$

$$f_w(z, w) = \frac{\partial f(z, w)}{\partial w}. \tag{7.4}$$

The implicit function $\alpha(w)$ will satisfy the equation

$$f(\alpha(w), w) = 0. \tag{7.5}$$

Taking the derivative of the above equation with respect to w results in

$$\frac{df(\alpha(w), w)}{dw} = 0. \tag{7.6}$$

Expanding the above equation results in

$$f_z(\alpha(w), w)\frac{d\alpha(w)}{dw} + f_w(\alpha(w), w) = 0. \tag{7.7}$$

Solving for the derivative of $\alpha(w)$ with respect to w results in

$$\frac{d\alpha(w)}{dw} = -\frac{f_w(\alpha(w), w)}{f_z(\alpha(w), w)}. \tag{7.8}$$

The above first order differential equation is known as the Davidenko differential equation.

7.4 Euler's Method

Consider the initial value problem

$$z' = f(z, w), \tag{7.9}$$
$$z_0 = z(w_0). \tag{7.10}$$

Using the above equations it is possible to calculate $z'(w_0)$.

$$z'(w_0) = f(z(w_0), w_0) = f(z_0, w_0). \tag{7.11}$$

The function $z(w)$ can be written as a Taylor series about the point w_0

$$z(w) = \sum_{k=0}^{\infty} \frac{z^{(k)}(w_0)}{k!}(w - w_0)^k \tag{7.12}$$
$$= z(w_0) + z'(w_0)(w - w_0) + \sum_{k=2}^{\infty} \frac{z^{(k)}(w_0)}{k!}(w - w_0)^k. \tag{7.13}$$

The function $z(x)$ can be approximated by the first two terms of the Taylor series

$$t(w) = z(w_0) + z'(w_0)(w - w_0). \tag{7.14}$$

Substituting (7.10) and (7.11) into (7.14) results in

$$t(w) = z_0 + f(z_0, w_0)(w - w_0). \tag{7.15}$$

Setting $z_1 = t(w_1)$ results in

$$z_1 = z_0 + f(z_0, w_0)(w_1 - w_0). \tag{7.16}$$

Suppose that $w(t)$ is a path that starts at w_0. Then it is possible to evaluate the solution of the initial value problem

$$z' = f(z, w), \tag{7.17}$$
$$z_0 = z(w_0), \tag{7.18}$$

along the path $w(t)$. Let $0 = t_0 < t_1 < \ldots < t_{n+1} = 1$ be a partition of the interval $[0, 1]$. Let w_k for $k = 0, 1, \ldots, n+1$ be defined as follows

$$w_k = w(t_k). \tag{7.19}$$

Note that the above definition does not conflict with the previous definition of w_0 since $w_0 = w(0)$. The variables in (7.16) can be relabeled resulting in

$$z_{k+1} = z_k + f(z_k, w_k)(w_{k+1} - w_k). \tag{7.20}$$

Equation (7.20) can be used to give an approximation of

$$z(w(t_k)) = z(w_k) \approx z_k, \tag{7.21}$$

for each t_k in the partition of $[0, 1]$. At each step of Euler's Method the error

$$E_{k+1} = \sum_{k=2}^{\infty} \frac{z^{(k)}}{k!}(w - w_0)^k, \tag{7.22}$$

is introduced. The errors are cumulative. Hence Euler's Method is numerically unstable.

7.5 Predict Correct Algorithm

Let $f(z, w)$ be a bivariate polynomial and L be a line segment from a to b. Let α be a simple root of $f(z, a)$. Let

$$w(t) = (1 - t)a + tb, \tag{7.23}$$

be a parametrization of the path going along the line segment from a to b. To see what happens to a simple root of $f(z, w(t))$ as t traverses the interval $[0, 1]$ the following algorithm can be used.

1. **Initialization.** Set $h = 1/10$, $\alpha_0 = \alpha$, $k = 0$, $s = 0$, and $t_0 = 0$.
2. **Loop.**
   (a) **Predict.** Set $t_{k+1} = \max(t_k + h, 1.0)$. Calculate a prediction α'_{k+1} of a root of $f(z, w(t_{k+1}))$ based on the fact that α_k is a root of $f(z, w(t_k))$.
   (b) **Correct.** Calculate a root α_{k+1} of $f(z, w(t_{k+1}))$ using α'_{k+1} as an initial guess.
   (c) **Adjust Step Length** If the correction step failed do the following:
      i. Set $h = h/2$.
      ii. Set $s = 0$.

      If the correction step succeeded do the following:
      i. If $t_{k+1} = 1.0$ then terminate the algorithm.
      ii. Increment s and k.
      iii. If $s \geq M$ set $h = 2h$ and $s = 0$.

According to Sommese and Wampler a value of M between 3 and 5 works well. The next two subsections will contain descriptions of a prediction step and a correction step.

7.6 Euler Predictor

A single step of Euler's Method can be used to calculate a prediction α'_{k+1} of a root of $f(z, w(t_{k+1}))$ based on the fact that α_k is a root of $f(z, w(t_k))$. According to the Davidenko differential equation

$$\frac{d\alpha(w)}{dw} = -\frac{f_w(\alpha(w), w)}{f_z(\alpha(w), w)}. \tag{7.24}$$

Using a single step of Euler's method to predict the value of α_{k+1} results in the following equation

$$\alpha_{k+1} = \alpha_k - \frac{f_w(\alpha_k, w(t_k))}{f_z(\alpha_k, w(t_k))}(w(t_{k+1}) - w(t_k)). \tag{7.25}$$

This type of predictor is known as an Euler Predictor.

7.7 Newton Corrector

Newton's Method can be used to calculate a root α_{k+1} of $f(z, w(t_{k+1}))$ given an initial guess α'_{k+1} of where the root is located. If α'_{k+1} is close enough to an actual root $\alpha_{k+1}$of $f(z, w(t_{k+1}))$ Newton's Method will converge to α_{k+1}. This type of corrector is known as a Newton Corrector.

Newton's Method can fail to converge. Hence the need for the part of the Predict and Correct Algorithm where the step length h is halved if the correct step fails. With a sufficiently small step size Euler's Method will generate a prediction α'_{k+1} for a root that is sufficiently close to the actual root α_{k+1} such that Newton's Method will converge to α_{k+1} using α'_{k+1} as an initial guess.

7.8 Root Jumping

There is one problem with using a Predict and Correct Algorithm that uses Euler Prediction and Newton Correction. The algorithm can start following the wrong root. A description of how this situation can occur will be described next.

Let $f(z, w)$ be a bivariate polynomial. Let $w(t)$ be a piecewise smooth path path that does not intersect any of the critical points of $f(z, w)$. Let α and β are two simple roots of $f(z, w(0))$. Then according to the Implicit Function Theorem

$$f(z, w) = 0, \tag{7.26}$$

will implicitly define two functions $\alpha(w)$ and $\beta(w)$ defined in a neighborhood of $w(0)$ such that $\alpha(w(0)) = \alpha$ and $\beta(w(0)) = \beta$. The functions $\alpha(w)$ and $\beta(w)$ can be analytically continued along the path $w(t)$. While using the Predict and Correct algorithm to follow the functions $\alpha(w)$ and $\beta(w)$ the following problem can occur.

Suppose that α_k is a root of $f(z, w(t_k)$. Then the Predict step can be used to find an approximate root α'_{k+1} of $f(z, w(t_{k+1}))$. The Correct step can be used to find a root α_{k+1} of $f(z, w(t_{k+1}))$ using α'_{k+1} as an initial guess. Usually $\alpha(w(t_{k+1})) = \alpha_{k+1}$. However it can happen that $\beta(w(t_{k+1})) = \alpha_k$. This situation occurs if the predicted root converges to a point on the curve $\beta(w(t))$ when the root is corrected. This situation is known as root jumping.

Usually this situation can be detected since the number of distinct roots of $f(z, w)$ at step $k + 1$ of the algorithm is one less than the number of distinct roots of $f(z, w)$ at step k of the algorithm. A situation can occur where two or more roots jump at the same time in such a manner that the number of distinct roots does not decrease. To avoid this situation an improved Predict and Correct Algorithm is needed.

7.9 Minimum Distance Between Roots

The following Theorem gives some insight into the minimum distance between roots of $f(z, w(t))$ for $t \in [0, 1]$.

Theorem 12. *Let $f(z, w)$ be a bivariate polynomial that has degree $n > 1$ when viewed as a polynomial in z. Let $w(t)$ be a path that avoids the critical points of $f(z, w)$. Let α_0 and β_0 be two roots of $f(z, w(0))$. Let $\alpha(t)$ be the implicit function defined by $\alpha(0) = \alpha_0$ and $f(\alpha(t), w(t)) = 0$. Let $\beta(t)$ be the implicit function defined by $\beta(0) = \beta_0$ and $f(\beta(t), w(t)) = 0$. Then there is an $\varepsilon > 0$ such that $|\beta(t) - \alpha(t)| > \varepsilon$ for all $t \in [0, 1]$.*

Proof. Let $d(t)$ be a real function defined as

$$d(t) = |\beta(t) - \alpha(t)|. \tag{7.27}$$

It is obvious that $d(t)$ is a continuous function and that $d(t) \geq 0$. Since $w(t)$ avoids the critical points of $f(z, w)$ there is no $t_0 \in [0, 1]$ such that $\alpha(t_0) = \beta(t_0)$. Therefore there is no $t_0 \in [0, 1]$ such that $d(t_0) = 0$. Hence

$$d(t) > 0. \tag{7.28}$$

According to the Extreme Value Theorem since $[0, 1]$ is closed and bounded there exists $c \in [0, 1]$ such that

$$d(t) \geq d(c). \tag{7.29}$$

Applying (7.28) to (7.29) results in

$$d(t) \geq d(c) > 0. \tag{7.30}$$

□

7.10 Newton Rouché's Method

A modified version of Newton's Method will now be introduced. For lack of a better name the new algorithm will be called Newton-Rouché's Method. Newton-Rouché's Method is used to find simple roots of a polynomial.

Given an initial guess z_0 at a root of the polynomial $f(z)$ Newton's Method works by creating a number sequence

$$z_{k+1} = z_k - \frac{f(z_k)}{f'(z_k)}, \tag{7.31}$$

that will converge to a root of $f(z)$ if z_0 is sufficiently close to a root of $f(z)$. Obviously only a finite number of elements in the sequence are calculated. It becomes necessary to have a criteria for when the algorithm should be halted. The algorithm is halted if the maximum number of iterations has been reached or the expression $|-f(z_k)/f'(z_k)|$ becomes smaller than a predetermined error ε. If the maximum number of iterations has been reached then the algorithm is considered to have failed to find a root of $f(z)$. If $|-f(z_k)/f'(z_k)| < \varepsilon$ then the algorithm is considered to have found a root of $f(z)$.

The term $-f(z_k)/f'(z_k)$ can be viewed as the error between z_k and the root of

$$t(z) = f(z_k) + f'(z_k)(z - z_k), \tag{7.32}$$

the best linear approximation of $f(z)$ at x_k. The fact that $|-f(z_k)/f'(z_k)| < \varepsilon$ does not guarantee that the disc $D(z_k, \varepsilon)$ contains a root of $f(z)$. To make such a guarantee it is necessary to consider the higher order terms of the Taylor expansion of $f(z)$ around the point z_k.

A fast way of calculating the Taylor expansion of a polynomial is given by the following Lemma.

Lemma 32. *The Taylor expansion of the polynomial $p(z)$ around the point z_0 is equal to $p(z + z_0)$.*

Proof. Taylor expanding the polynomial $p(z)$ around the point z_0 results in

$$q(z) = \sum_{k=0}^{n} \frac{p^{(k)}(z_0)}{k!}(z - z_0)^k, \tag{7.33}$$

where $p(z) = q(z - z_0)$. Substituting $x = z - z_0$ results in $p(x + z_0) = q(x)$. □

Newton-Rouché's Method works as follows. Given an initial guess z_0 at a root of $f(z)$ create the number sequence

$$z_{k+1} = z_k - \frac{f(z_k)}{f'(z_k)}. \tag{7.34}$$

The algorithm is halted when the number of iterations reaches a predetermined maximum or it is determined that the disc $D(z_k, \varepsilon)$ contains a simple root of $f(z)$.

The following method is used to determine if the disc $D(z_k, \varepsilon)$ contains a simple root of $f(z)$. The function $f(z)$ is Taylor expanded around the point z_k using the method in Lemma 32

$$f(z) = f(z_k) + f'(z_k)(z - z_k) + \sum_{j=2}^{n} \frac{f^{(j)}(z_k)}{j!}(z - z_k) \tag{7.35}$$

$$= t(z) + s(z), \tag{7.36}$$

where $t(z)$ and $s(z)$ are defined as

$$t(z) = f(z_k) + f'(z_k)(z - z_k), \tag{7.37}$$

$$s(z) = \sum_{j=2}^{n} \frac{f^{(j)}(z_k)}{j!}(z - z_k). \tag{7.38}$$

It is obvious that the linear polynomial $t(z)$ has a simple root α at

$$\alpha = z_k - \frac{f(z_k)}{f'(z_k)}, \tag{7.39}$$

and that $t(z)$ has no other roots. The distance d between α and the center of the disc $D(z_k, \varepsilon)$ is equal to

$$d = |\alpha - z_k| = \left| z_k - \frac{f(z_k)}{f'(z_k)} - z_k \right| = \left| \frac{f(z_k)}{f'(z_k)} \right| = \frac{|f(z_k)|}{|f'(z_k)|}. \tag{7.40}$$

Therefore if $|f(z_k)|/|f'(z_k)| < \varepsilon$ then the disc $D(z_k, \varepsilon)$ will contain the only root of $t(z)$.

The following Lemmas can be used to calculate a lower bound of $t(z)$ and an upper bound of $s(z)$.

Lemma 33. *Suppose that* $t(z) = a_0 + a_1(z - z_0)$. *Then* $||a_1|r - |a_0||$*is a lower bound of* $|t(z)|$ *on* $C(z_0, r)$.

Proof. Apply the reverse triangle inequality to $t(z)$. □

Lemma 34. *Suppose that* $s(z)$ *is the polynomial*

$$s(z) = \sum_{k=0}^{n} c_k (z - z_0)^k. \tag{7.41}$$

Then

$$\sum_{k=0}^{n} |c_k| r^k, \tag{7.42}$$

is an upper bound of $|s(z)|$ *on* $C(z_0, r)$.

Proof. Since $z \in C(z_0, r)$ it follows that $|z - z_0| = r$. But then

$$|s(z)| = |\sum_{k=0}^{n} c_k (z - z_0)^k| \leq \sum_{k=0}^{n} |c_k (z - z_0)^k| \tag{7.43}$$

$$= \sum_{k=0}^{n} |c_k||z - z_0|^k = \sum_{k=0}^{n} |c_k| r^k \tag{7.44}$$

□

To determine if the disc $D(z_k, \varepsilon)$ contains a simple root α of $f(z)$ do the following:

1. Verify that $D(z_k, \varepsilon)$ contains the root α of $t(z)$. If α is not in $D(z_k, \varepsilon)$ then it is not possible to determine if $D(z_k, \varepsilon)$ contains a simple root of $f(z)$. In this case the verification step is terminated.

2. Use Lemma 33 to calculate a lower bound L of $|t(z)|$ for $z \in C(z_k, \varepsilon)$. Use Lemma 34 to calculate an upper bound U of $|s(z)|$ for $z \in C(z_k, \varepsilon)$. If $U < L$ then according to Rouché's Theorem the disc $D(z_k, \varepsilon)$ contains a simple root of $f(z)$ since $D(z_k, \varepsilon)$ will contain one and only one root of $t(z)$. If $U \geq L$ then it can not be determined if the disc $D(z_k, \varepsilon)$ contains a root of $f(z)$.

7.11 Predict Correct Verify Algorithm

Let $f(z, w)$ be a bivariate polynomial that when viewed as a polynomial in z has degree n. Let L be a line segment from a to b that does not intersect any of the critical points of $f(z, w)$. Let $w(t)$ be a path that parameterizes the line segment L defined as follows

$$w(t) = (1 - t)a + tb. \tag{7.45}$$

For each $t \in [0, 1]$ the polynomial $f(z, w(t))$ will have n simple roots since there is no t_0 such that $w(t_0)$ is a critical point of $f(z, w)$. Let $z_1, z_2, \ldots, z_n$ be the simple roots of $f(z, a)$.

According to the Implicit Function Theorem there exists analytic functions $\alpha_1(w), \alpha_2(w), \ldots, \alpha_n(w)$ defined in a neighborhood of a such that

$$f(\alpha_k(a), a) = 0 \text{ for } k = 1, 2, \ldots, n. \tag{7.46}$$

The functions $\alpha_k(w)$ for $k = 1, 2, \ldots, n$ can be analytically continued along $w(t)$ since $w(t)$ avoids the critical points of $f(z, w)$.

The Predict and Correct Algorithm can be used to track the path of each of the functions $\alpha_k(w)$ for $k = 1, 2, \ldots, n$. The problem with this approach is that root jumping can occur. Instead a new algorithm that will be developed that avoids the problem of root jumping. For lack of a better name the new algorithm will be called the Predict, Correct and Verify Algorithm.

The Predict Correct and Verify algorithm tracks all the simple roots of the polynomial $f(z, w(t))$ as t traverses the interval $[0, 1]$. If there is a point t_0 such that any of the roots of $f(z, w(t_0))$ is not simple then the algorithm will fail. Therefore it is a requirement that the path $w(t)$ does not intersect any of the critical points of $f(z, w)$.

The following is a detailed description of the Predict, Correct and Verify Algorithm.

Input. *The following inputs are needed for the algorithm:*

1. *A bivariate polynomial $f(z, w)$ that when viewed as a polynomial in z has degree n.*

2. *A path $w(t) = (1 - t)a + tb$ that is a parametrization of the line segment from a to b.*

3. *A maximum error ε.*

4. *A minimum step size d.*

5. *A collection of non overlapping discs $D(z_0, \varepsilon), D(z_1, \varepsilon), \ldots, D(z_n, \varepsilon)$ where each disc contains a simple root of $f(z, a)$.*

6. *A maximum number of iterations N that should be preformed before Newton-Rouché's Method is considered to have failed.*

7. *The number of successful steps M that should be preformed before doubling the step size.*

Algorithm. *The Predict, Correct and Verify Algorithm consists of the following steps:*

1. ***Initialization.*** *For* $j = 1, 2, \ldots, n$ *set* $D_0^j = D(z_j, \varepsilon)$. *Set* $h = 1/10$, $k = 0$, $s = 0$, *and* $t_0 = 0$.

2. ***Loop.***

   (a) ***Predict.*** *Set* $t_{k+1} = \max(t_k + h, 1.0)$. *For* $j = 1, 2, \ldots, n$ *calculate a prediction* α_{k+1}^j *of a root of* $f(z, w(t_{k+1}))$ *based on the fact that* D_k^j *contains a root of* $f(z, w(t_k))$.

   (b) ***Correct.*** *For* $j = 1, 2, \ldots, n$ *calculate a disc* D_{k+1}^j *that contains a root of* $f(z, w(t_{k+1}))$ *using* α_{k+1}^j *as an initial guess. If the resulting discs overlap then the Correct step is considered to have failed.*

   (c) ***Verify.*** *If the correct step succeeded Verify that no root jumping has occurred between* D_k^j *and* D_{k+1}^j *when going along the path from* $w(t_k)$ *to* $w(t_{k+1})$.

   (d) ***Adjust Step Length*** *If the correction step failed do the following:*

      i. *Set* $h = h/2$.

      ii. *If* $h < d$ *then the algorithm has failed.*

      iii. *Set* $s = 0$.

   If the correction step succeeded do the following:

      i. *If* $t_{k+1} = 1.0$ *then terminate the algorithm.*

      ii. *Increment* s *and* k.

      iii. *If* $s \geq M$ *set* $h = 2h$ *and* $s = 0$.

Output. *The output of the algorithm is a collection of non overlapping discs* $D_k^1, D_k^2, \ldots, D_k^n$. *Each disc* D_k^j *contains the root of* $f(z, w(t))$ *that is the result of following the root of* $f(z, w(0))$ *that is located in* D_0^j *as* t *traverses the interval* $[0, 1]$.

Just as in the case for the Predict and Correct Algorithm a value of M between 3 and 5 works well.

7.12 Verified Initial Roots

Suppose that s is a non critical point of the bivariate polynomial $f(z, w)$. Furthermore suppose that the degree of the polynomial $f(z, w)$ when viewed as a polynomial in z has degree n. Then the polynomial $f(z, s)$ will have only simple roots.

The following algorithm can be used to calculate a collection of n discs with radius ε such that each disc contains a simple root of $f(z, s)$.

1. Use the following process to calculate approximations of all the simple roots of $f(z, s)$:

   (a) Use Laguerre's Method to calculate a root.

   (b) Use Horner's Method to synthetically divide out the root.

   (c) If n roots have been found then stop. Otherwise repeat the process.

2. Make sure that the error ε is smaller than one third of the minimum distance between the roots of $f(z, s)$. If this is not the case then the algorithm fails with an indication that the error ε is too large.

3. Use Newton-Rouché's Method to polish the approximate roots of $f(z, s)$. If Newton-Rouché's Method fails to converge then the algorithm fails with an indication of what has happened. If any of the discs overlap then the algorithm fails with an indication of what has happened.

The result of a successful run of the algorithm is a collection of n discs with radius ε that do not overlap. Each disc contains a simple root of $f(z, s)$. In the rare cases where the algorithm fails manual intervention is required to find non overlapping discs that each contain a simple root of $f(z, w)$.

7.13 Euler Disc Predictor

The roots of the polynomial $f(z, w(t_k))$ are located in the discs $D_k^1, D_k^2, \ldots, D_k^n$. Even though the exact location of the roots of $f(z, w(t_k))$ are not known a single step of Euler's Method can still be used to predict where the roots of $f(z, w(t_{k+1}))$ are located. This is done by making the assumption that the roots of $f(z, w(t_k))$ are located at the centers of the discs $D_k^1, D_k^2, \ldots, D_k^n$ and using an Euler predictor to predict where the roots of $f(z, w(t_{k+1}))$ are located.

7.14 Newton Rouché Correction

Let $f(z, w)$ be a bivariate polynomial. Let $w(t)$ be a path that avoids the critical points of $f(z, w)$. Suppose that $\alpha_1, \alpha_2, \ldots, \alpha_n$ are initial guesses for where the roots of $f(z, w(t_{k+1}))$ are located. Then Newton-Rouché's Method can be used to calculate discs with radius ε that contain the roots of $f(z, w(t_{k+1}))$. However Newton-Rouché's can fail to converge. In this case the step size is halved and the step is repeated.

7.15 Sufficient Condition for No Root Jumping

A sufficient condition for determining that no root jumping has occurred will now be developed. Before the sufficient condition can be stated and proved it is necessary to introduce a couple of Lemmas. The following Lemma gives a sufficient condition for a simple root of the bivariate polynomial $(f, w(t))$ to stay in a disc D as the parameter t traverses the interval [0,1].

Lemma 35. *Let $f(z, w)$ be a bivariate polynomial. Let $w(t)$ be a path that goes along a line segment from a to b*

$$w(t) = (1 - t)a + tb. \tag{7.47}$$

If $f(z, w)$ has one and only one root in $D(c, r)$ when $w \in D(a, |b - a|)$ then $f(z, w(t))$ has one and only one root in $D(c, r)$ for every $t \in [0, 1]$.

Proof. The Lemma follows from the fact that $w(t) \in D(a, |b - a|)$ for $t \in [0, 1]$. □

Next is a Theorem that gives a sufficient condition for no root jumping to occur.

Theorem 13. *Let $f(z, w)$ be a bivariate polynomial that has degree n when viewed as polynomial in z. Let $w(t)$ be a path that goes along a line segment from a to b and that avoids the critical points of $f(z, w)$.*

Let $\alpha_1, \alpha_2, \dots, \alpha_n$ be the roots of $f(z, a)$. Let $\beta_1, \beta_2, \dots, \beta_n$ be the roots of $f(z, b)$. Suppose that the discs $D(c_1, r_1), D(c_2, r_2), \dots, D(c_n, r_n)$ do not intersect and that for $k = 1, 2, \dots, n$ each disc $D(c_k, r_k)$ contains the roots α_k and β_k.

If for $k = 1, 2, \dots, n$ the disc $D(c_k, r_k)$ contains one and only one root of $f(z, w)$ when $w \in D(a, |b - a|)$ then β_k is the analytic continuation of α_k.

Proof. Suppose that β_k is the analytic continuation of α_k but that α_k and β_k do not lie in the same disc $D(c_k, r_k)$. Then according to the Implicit Function Theorem $f(z, w)$ implicitly defines a continuous function $\gamma(t)$ such that

$$f(\gamma(t), w(t)) = 0, \gamma(0) = \alpha_k, \gamma(1) = \beta_k, \tag{7.48}$$

where $k \neq j$.

The function $\gamma(t)$ has a starting point that lies in a disc $D(c_k, r_k)$ and an end point that lies in different disc $D(c_j, r_j)$. Since the discs $D(c_k, r_k)$ and $D(c_j, r_j)$ do not intersect and $\gamma(t)$ is a continuous function there has to be a t_0 such that $\gamma(t_0)$ does not lie in $D(c_k, r_k)$ or $D(c_j, r_j)$.

But this is a contradiction since for every $w \in D(a, |b - s|)$ each of the n discs $D(c_k, r_k)$ contains a simple root of $f(z, w)$. □

If D_1 and D_2 are two non overlapping discs then the following Lemma gives a way of constructing a larger disc D_3 that contains both of the smaller discs.

Lemma 36. *Let $D_1 = D(z_1, r)$ and $D_2 = D(z_2, r)$ be two discs that do not intersect. Let $d = |z_2 - z_1|$. Let $k = 3/2$. Then the disc $D_3 = D(z_1, kd)$ contains both D_1 and D_2.*

Proof. Since D_1 and D_2 do not intersect and the radius of both discs is r it is obvious that d satisfies the following inequality

$$d > 2r. \tag{7.49}$$

The above inequality can be transformed into the following inequality

$$\frac{3d}{2} > 3r. \tag{7.50}$$

From the above inequality it is obvious that

$$D_1 \subset D_3. \tag{7.51}$$

Dividing both sides of (7.50) by 3 results in

$$\frac{d}{2} > r. \tag{7.52}$$

Let s be the distance between z_2 and the border of D_3. It is obvious that

$$s = \frac{d}{2}. \tag{7.53}$$

Hence $s > r$ and therefore D_2 is contained in D_3. □

The following Theorem gives a sufficient condition for detecting that no root jumping has occurred during the Predict and Correct step.

Theorem 14. *Let $f(z,w)$ be a bivariate polynomial that has degree n when viewed as polynomial in z. Let $D(a_1,\varepsilon),\ldots,D(a_n,\varepsilon)$ be non intersecting discs that each contains a simple root of $f(z,a)$. Let $D(b_1,\varepsilon),\ldots,D(b_n,\varepsilon)$ be non intersecting discs that each contains a simple root of $f(z,b)$. Let $w(t)$ be a path that goes along a line segment from a to b and that avoids the critical points of $f(z,w)$.*

For $k=1,2,\ldots,n$ let D_k be the disc $D(a_k,3|b_k-a_k|/2)$. If the discs $D_1,D_2,\ldots,D_n$ do not intersect and each of the discs $D_1,D_2,\ldots,D_n$ contains one and only one root of $f(z,w)$ when $w\in D(a,|b-a|)$ then the root in $D(b_k,\varepsilon)$ is the analytic continuation along $w(t)$ of the root in $D(a_k,\varepsilon)$.

Proof. According to Lemma 36 for $k=1,2,\ldots,n$ the disc D_k contains the discs $D(a_k,\varepsilon)$ and $D(b_k,\varepsilon)$. This in turn means that for $k=1,2,\ldots,n$ the disc D_k contains α_k and β_k such that $f(\alpha_k,a)=0$ and $f(\beta_k,b)=0$. According to Theorem 13 β_k is the analytic continuation of α_k since the discs $D_1,D_2,\ldots,D_n$ do not intersect and each of the discs $D_1,D_2,\ldots,D_n$ contains one and only one root of $f(z,w)$ when $w\in D(a,|b-a|)$. □

7.16 Roots and Variable Substitution

Suppose that the bivariate polynomial $f(z,w)$ has one and only one root in the disc D when w lies in the disc S. Furthermore suppose that $g(z,w)$ is the following bivariate polynomial

$$g(z,w)=f(z+c,w+a). \tag{7.54}$$

Then there are discs D' and S' such that $g(z,w)$ has one and only one root in the disc D' when w lies in the disc S'.

Lemma 37. *Let $f(z,w)$ be a bivariate polynomial. Let g(z,w) be the bivariate polynomial*

$$g(z,w)=f(z,w+a). \tag{7.55}$$

The polynomial $f(z,w)$ has one and only one root in $D(c,r)$ when $w\in D(a,s)$ if and only if the polynomial $g(z,w)$ has one and only one root in D(c,r) when $w\in D(0,s)$.

Proof. The lemma follows from the variable substitution $w=u+a$. □

Lemma 38. *Let $f(z,w)$ be a bivariate polynomial. Let g(z,w) be the bivariate polynomial*

$$g(z,w)=f(z+c,w). \tag{7.56}$$

The polynomial $f(z,w)$ has one and only one root in $D(c,r)$ when $w\in D(a,s)$ if and only if the polynomial $g(z,w)$ has one and only one root in D(0,r) when $w\in D(a,s)$.

Proof. The proof of this Lemma is similar to that of Lemma 37. □

Theorem 15. *Let $f(z,w)$ be a bivariate polynomial. Let g(z,w) be the bivariate polynomial*

$$g(z,w) = f(z+c, w+a). \tag{7.57}$$

The polynomial $f(z,w)$ has one and only one root in $D(c,r)$ when $w \in D(a,s)$ if and only if the polynomial $g(z,w)$ has one and only one root in D(0,r) when $w \in D(0,s)$.

Proof. The Theorem follows from Lemma 38 and Lemma 37. □

7.17 Bounds of Polynomials

The following Lemma gives a lower bound of the magnitude of a linear polynomial on a circle centered on the origin.

Lemma 39. *Suppose that $h(z)$ is the polynomial*

$$h(z) = h_0 + h_1 z. \tag{7.58}$$

Then $\big||h_1|r - |h_0|\big|$ is a lower bound of $|h(z)|$ when z lies on the circle $C(0,r)$.

Proof. According to the reverse triangle inequality

$$|h(z)| = |h_1 z + h_0| \geq \big||h_1||z| - |h_0|\big| \tag{7.59}$$

The Lemma now follows from the fact that $|z| = r$ when $z \in C(0,r)$. □

The following Lemma gives an upper bound of the magnitude of a polynomial on a circle centered on the origin.

Lemma 40. *Suppose that $c(w)$ is the polynomial*

$$c(w) = \sum_{k=0}^{n} c_k w^k. \tag{7.60}$$

An upper bound of $|c(w)|$ when w lies on the circle $C(0,s)$ is given by

$$\sum_{k=0}^{n} |c_k| s^k. \tag{7.61}$$

Proof. According to the triangle inequality

$$|c(w)| = |\sum_{k=0}^{n} c_k w^k| \leq \sum_{k=0}^{n} |c_k||w|^k. \tag{7.62}$$

The Lemma now follows from the fact that $|w| = s$ when $w \in C(0,s)$. □

The following Lemma gives an upper bound of the magnitude of a bivariate polynomial $k(z,w)$ when (z,w) *in* $C(0,r) \times C(0,s)$.

Lemma 41. *Suppose that $k(z,w)$ is the bivariate polynomial*

$$k(z,w) = \sum_{k=0}^{n} c_k(w) z^k, \tag{7.63}$$

where $c_1(w), c_2(w), \ldots, c_n(w)$ are polynomials in w. Let M_k be an upper bound of $c_k(w)$ when w lies on the circle $C(0,s)$. An upper bound of $|k(z,w)|$ when $(z,w) \in C(0,r) \times C(0,s)$ is given by

$$\sum_{k=0}^{n} M_k r^k. \tag{7.64}$$

Proof. According to the triangle inequality

$$|k(z,w)| = |\sum_{k=0}^{n} c_k(w) z^k| \leq \sum_{k=0}^{n} |c_k(w)|\, |z|^k \tag{7.65}$$

Since M_k is an upper bound for $|c_k(w)|$ when $w \in C(0,s)$ the above inequality can be transformed into

$$|k(z,w)| \leq \sum_{k=0}^{n} M_k\, |z|^k \tag{7.66}$$

The Lemma now follows from the fact that $|z| = r$ when $z \in C(0,r)$. □

7.18 Sufficient Condition for Disc to Contain a Root

The following theorem gives a sufficient condition for the bivariate polynomial $g(z,w)$ to contains one and only one root in $D(0,r)$ when $w \in D(0,s)$.

Theorem 16. *Suppose that $g(z,w)$ is the bivariate polynomial*

$$g(z,w) = g_0(w) + g_1(w)z + g_2(w)z^2 + \ldots + g_n(w)z^n. \tag{7.67}$$

Let $h(z)$ be the polynomial

$$h(z) = g_0(0) + g_1(0)z. \tag{7.68}$$

Let $k(z,w)$ be the bivariate polynomial

$$k(z,w) = g(z,w) - h(z). \tag{7.69}$$

Suppose that $r > |g_0(0)/g_1(0)|$. If there exists a lower bound L of $h(z)$ for $z \in C(0,r)$ and an upper bound U of $|k(z,w)|$ for $(z,w) \in D(0,r) \times D(0,s)$ such that $U < L$ then $g(z,w)$ will have one and only one root in $D(0,r)$ when $w \in D(0,s)$.

Proof. It is obvious that $h(z)$ has a root at $z = -g_0(0)/g_1(0)$. Furthermore since $r > |g_0(0)/g_1(0)|$ the disc $D(0,r)$ contains one and only one root of $h(z)$. According to Rouché's Theorem the polynomials $g(z,w)$ and $h(z)$ will have the same number of roots in $D(0,r)$ when $w \in D(0,s)$ since $|k(z,w)| < |h(z)|$ when $z \in D(0,r)$ and $w \in D(0,s)$. □

7.19 Rouché Verification

Let $f(z, w)$ be a bivariate polynomial that has degree n when viewed as a polynomial in z. Let $w(t)$ be a path alone a line segment from a to b that avoids the critical points of $f(z, w)$. Let $D(a_1, \varepsilon), D(a_2, \varepsilon), \ldots, D(a_n, \varepsilon)$ be a collection of non intersecting discs where each disc contains a simple root of $f(z, a)$. Let $D(b_1, \varepsilon), D(b_2, \varepsilon), \ldots, D(b_n, \varepsilon)$ be a collection of non intersecting discs where each disc contains a simple root of $f(z, b)$.

The following method can be used to check if for $k = 1, 2, \ldots, n$ the root of $f(z, b)$ located in $D(b_k, \varepsilon)$ is the analytic continuation of the root of $f(z, a)$ located in $D(a_k, \varepsilon)$ when w traverses the line segment from a to b.

Create the discs $D_1 = D(a_1, 3|b_1 - a_1|/2), \ldots, D_n = D(a_n, 3|b_n - a_n|/2)$. If any of the discs overlap then the verification step is considered to have failed. Otherwise according to Theorem 14 if the discs $D_1, D_2, \ldots, D_n$ do not overlap and $f(z, w)$ has one and only one root in each of the discs $D_1, D_2, \ldots, D_n$ when $w \in D(a, |b - a|)$ then for $k = 1, 2, \ldots, n$ the root of $f(z, b)$ located in $D(b_k, \varepsilon)$ is the analytic continuation along $w(t)$ of the root of $f(z, a)$ located in $D(a_k, \varepsilon)$.

The following method can be used to check if for $k = 1, 2, \ldots, n$ the polynomial $f(z, b)$ has one and only one root in D_k when $w \in D(a, |b - a|)$.

Let $g(z, w)$ be the bivariate polynomial

$$g(z, w) = f(z + a_k, w + a). \tag{7.70}$$

Let r be the radius of the disc D_k. According to Theorem 15 the polynomial $f(z, w)$ will have one and only one root in $D(a_k, r)$ when $w \in D(a, |b - a|)$ if and only if the polynomial $g(z, w)$ has one and only one root in $D(0, r)$ when $w \in D(0, |b - a|)$.

The following method can be used to check if the polynomial $g(z, w)$ has one and only one root in $D(0, r)$ when $w \in D(0, |b - a|)$.

The bivariate polynomial $g(z, w)$ can be written as

$$g(z, w) = g_0(w) + g_1(w)z + g_2(w)z^2 + \ldots + g_n(w)z^n, \tag{7.71}$$

where $g_0(w), g_1(w), \ldots, g_n(w)$ are polynomials in w. Let $h(z)$ be the polynomial

$$h(z) = g_0(0) + g_1(0)z. \tag{7.72}$$

Let $k(z, w)$ be the bivariate polynomial

$$k(z, w) = g(z, w) - h(z). \tag{7.73}$$

Let L be the lower bound of $h(z)$ when $z \in C(0, r)$ calculated using Lemma 39. Let U be the upper bound of $k(z, w)$ when $z \in C(0, r)$ and $w \in D(0, |b - a|)$ calculated using Lemma 40 and Lemma 41. According to Theorem 16 if $U < L$ and $r \leq |g_0(0)/g_1(0)|$ then $g(z, w)$ has one and only one root in $D(0, r)$ when $w \in D(0, |b - a|)$. If $r \leq |g_0(0)/g_1(0)|$ or $U \geq L$ then the verification step is considered to have failed.

8 Permutations of Roots

In this section a function that takes a bivariate polynomial and a closed path and generates a permutation is defined. Several properties of the permutation function are stated and proved. The permutations generated by homotopic paths are discussed. It will be shown that the permutation generated by an arbitrary closed path is equal to the permutation generated by a path that is the composition of circle paths and inverse circle paths.

Given a bivariate polynomial a method is developed for creating a collection of circle paths that all have a common starting point and where each circle path goes around a critical point of the bivariate polynomial. A method for numerically calculating the permutation generated by a bivariate polynomial is described.

8.1 Permutation Function Definition

For an arbitrary closed path $w(t)$ and a bivariate polynomial $f(z,w)$ the function $R[f,w]$ will permute the roots of $f(z,w(0))$.

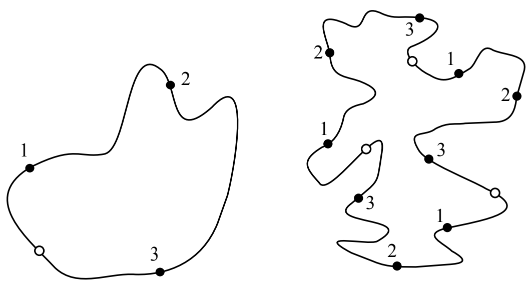
Theorem 17. *If $w(t)$ is a closed path that avoids the critical points of $f(z,w)$ then $R[f,w]$ will act as a permutation of the roots of $f(z,w(0))$.*

Proof. $R[f,w]$ will be a bijection between the roots of $f(z,w(0))$ and $f(z,w(1))$. Since $w(t)$ is a closed path we have that $w(0)=w(1)$. This in turn means that the roots of $f(z,w(0))$ are equal to the roots of $f(z,w(1))$. □

Corollary 7. *Let $f(z,w)$ be a bivariate polynomial and let $w(t)$ be a closed path. Furthermore let $\alpha_1,\alpha_2,\ldots,\alpha_n$ be the roots of $f(z,w(0))$. Then $Root(f,w,\alpha_k)=\alpha_j$ for some j.*

Given a bivariate polynomial and a closed path a new function can be defined that permutes the roots of the bivariate polynomial at the start of the path.

Definition 63. *Let $f(z,w)$ be a bivariate polynomial. Let $w(t)$ be a closed path that avoids the critical points of $f(z,w)$. Let $\alpha_1,\alpha_2,\ldots,\alpha_n$ be the simple roots of $f(z,w(0))$. Let $P(f,w)$ be the permutation of the roots of $f(z,w)$ generated by $Root(f,w,\alpha_k)$ for $k=1,2,\ldots,n$.*



(a) Parameter Plane (b) Root Plane

Figure 17: A closed loop in the parameter plane causes the roots to swap places.

8.2 Properties of the Permutation Function

The permutation generated by a bivariate polynomial along the composition of two paths is equal to the composition of the permutations generated by the individual paths.

Lemma 42. *If the end point of the closed path $u(t)$ coincides with the start point of the closed path $v(t)$ then $P(f, u \circ v) = P(f, u) \circ P(f, v)$.*

Proof. This is obvious if we consider what happens to the roots at the intersection of the closed paths u and v when moving along the closed path $u \circ v$. □

The permutations generated by a bivariate polynomial along the inverse of a path is equal to the inverse of the permutation generated by the path.

Lemma 43. *If w(t) is a closed path then $P(f, (w)^{-1}) = P(f, w)^{-1}$.*

Proof. Let I be the identity permutation. According to Lemma 8 $P(f, w \circ w^{-1}) = I$ and $P(f, w^{-1} \circ w) = I$ since the paths $w \circ w^{-1}$ and $w^{-1} \circ w$ are both homotopic to a point. According to Lemma 42 $P(f, w \circ w^{-1}) = P(f, w) \circ P(f, w^{-1})$ and $P(f, w^{-1} \circ w) = P(f, w^{-1}) \circ P(f, w)$. □

8.3 Permutations and Homotopic Paths

Consider a bivariate polynomial $f(z, w)$. Suppose that $w(t)$ and $u(t)$ are two paths that are homotopic in $\mathbb{C} \setminus Critical(f)$. Then $w(t)$ and $u(t)$ will generate the same permutation of the roots.

Theorem 18. *Let $f(z, w)$ be a bivariate polynomial. Let $w(t)$ and $u(t)$ be two closed paths that avoid the critical points of $f(z, w)$. If the paths $w(t)$ and $u(t)$ are homotopic in $\mathbb{C} \setminus Critical(f)$ then $P(f, w) = P(f, u)$.*

Proof. According to the Monodromy Theorem $R[f, w] = R[f, u]$ since the paths $w(t)$ and $u(t)$ are homotopic in $\mathbb{C} \setminus Critical(f)$. □

Corollary 8. *Let $f(z, w)$ be a bivariate polynomial. Let $w(t)$ be a closed path that avoids the critical points of $f(z, w)$. If $w(t)$ is homotopic to a point in $\mathbb{C} \setminus Critical(f)$ then $P(f, w)$ is equal to the identity permutation.*

If a path is the composition of circle paths and inverse circle paths then to determine the permutation generated by the path it is enough to know the permutation generated by the circle paths.

Lemma 44. *Let $w(t)$ be a composition of the paths $u_k(t)$ for $k = 1, \ldots, n$ where each u_k is a circle path or an inverse circle path path. Then to determine $P(f, w)$ it is enough to know $P(f, v_k)$ where $v_k(t) = u_k(t)$ if u_k is a circle path or $v_k(t) = u_k^{-1}(t)$ if u_k is an inverse circle path.*

Proof. This follows from Lemma 42 and Lemma 43. □

As the following Theorem shows the permutation generated by an arbitrary closed path is equal to the permutation generated by a closed path that is the composition of circle paths and inverse circle paths.

Theorem 19. *Let $f(z,w)$ be a bivariate polynomial. Let $w(t)$ be a path that avoids the critical points of $f(z,w)$. Let $c_1, c_2, \ldots, c_n$ be the critical points of $f(z,w)$. Let s be the starting point of the path $w(t)$. Suppose that there is no $k \neq j$ such that the points s, c_k, c_j are not colinear. Then there is a path $u(t)$ that is the composition of circle paths and inverse circle paths such that $P(f,w) = P(f,u)$.*

Proof. Let r be $1/4$ times the minimum distance between the critical points of $f(z,w)$. According to Theorem 18 the path $w(t)$ is homotopic in $\mathbb{C} \setminus Critical(f)$ to a path $u(t)$ that is a composition of the circle paths $B(s, c_1, r), \ldots, B(s, c_n, r)$ and their inverses. According to Theorem 18 $P(f,w) = P(f,u)$ since the paths $w(t)$ and $u(t)$ are homotopic in $\mathbb{C} \setminus Critical(f)$. □

8.4 Automatic Path Generation

Consider a bivariate polynomial $f(z,w)$ that has the critical points $c_1, c_2, \ldots, c_n$. The following method can be used to calculate a starting point s and a radius r such that the circle paths $B(s, c_1, r), B(s, c_2, r), \ldots, B(s, c_n, r)$ do not intersect. Furthermore the starting point of the paths will not lie in any of the discs $D(c_1, r), D(c_2, r), \ldots, D(c_n, r)$.

8.4.1 Calculating Circle Path Radius

The radius of the circle paths is calculated as follows. Let d be the minimum distance between the points $c_1, c_2, \ldots, c_n$. Let the radius $r = d/4$.

8.4.2 Valid and Nice Starting Points

A starting point has to satisfy certain constraints.

Definition 64. *A point s is a* valid starting point *for the points $c_1, c_2, \ldots, c_n$ and the radius r if it satisfies the following conditions:*

1. *The circle paths $B(s, c_1, r), \ldots, B(s, c_n, r)$ do not intersect except at s.*
2. *Each of the circle paths $B(s, c_1, r), \ldots, B(s, c_n, r)$ goes around one and only one of the points $c_1, c_2, \ldots, c_n$.*

The concept of a nice starting point is introduced to give a more visually appealing starting point.

Definition 65. *A point s is a* nice starting point *for the points $c_1, c_2, \ldots, c_n$ and the radius r if it satisfies the following conditions:*

1. *The point s is a valid starting point.*
2. *Let d be the minimum distance between the points $c_1, c_2, \ldots, c_n$. Let $r = d/4$. The point s does not lie in any of the discs $D(c_1, r), D(c_2, r), \ldots, D(c_n, r)$.*
3. *Let R_k be the ray starting at s and intersecting c_k. There is no R_k and R_j with $k \neq j$ such that the angle between R_k and R_j is less than $2°$.*

8.4.3 Candidate Starting Points

The next concept to be defined is that of a candidate starting point.

Definition 66. *Let C be the centroid of the points $c_1, c_2, \ldots, n$. Let M be the maximum distance between the point C and the points $c_1, c_2, \ldots, c_n$. A point s is a* candidate starting point *for the points $c_1, c_2, \ldots, c_n$ and the radius r if the following conditions are satisfied:*

1. *The point s lies in the disc $D(C, 5M)$.*
2. *The point s is a nice starting point.*

8.4.4 Calculating a Starting Point

The following method is used to calculate a starting point for the radius r and the points $c_1, c_2, \ldots, c_n$. A collection of candidate starting points s created and then the best candidate starting point is selected as the starting point.

To create a collection of N candidate starting points the following method is used. Let C be the centroid of the points $c_1, c_2, \ldots, c_n$. Let M be the maximum distance form C to the points $c_1, c_2, \ldots, c_n$. Let D be the disc $D(C, 5M)$. Select random points in D keeping the points that are nice starting points until N nice candidate starting points have been found.

Before the best candidate starting point can be selected it is necessary to introduce the concept of minimum angle between a point s and a collection of points $c_1, c_2, \ldots, c_n$.

Definition 67. *Let $c_1, c_2, \ldots, c_n$ be a sequence of distinct complex numbers. Let s be a complex number such that there is k such that $s = c_k$. For $k = 1, 2, \ldots, n$ let R_k be the ray starting at s and going through the point c_k. The* minimum angle *between the point s and the points $c_1, c_2, \ldots, c_n$ is equal to the minimum angle between the rays $R_1, R_2, \ldots, R_n$.*

Let $s_1, s_2, \ldots, s_N$ be the candidate starting points. For $k = 1, 2, \ldots, n$ let α_k be the minimum angle between the candidate starting point s_k and the points $c_1, c_2, \ldots, c_n$. .

Select the starting point c_k with the largest minimum angle α_k. If there are several candidate starting points that all have the largest minimum angle then pick one of these points at random.

8.5 Automatic Permutation Calculation

Consider a bivariate polynomial $f(z, w)$ that has degree n when viewed as a polynomial in z. A method for automatically calculating the permutations of the roots associated with each critical point of $f(z, w)$ will now be described.

1. Calculate the critical points $c_1, c_2, \ldots, c_m$ of $f(z, w)$ using the method described in Subsection 4.6.
2. Create a collection of N-gon paths $w_1(t), w_2(t), \ldots, w_m(t)$ that all share a common starting point s and where each N-gon path goes around one of the critical points $c_1, c_2, \ldots, c_m$ using the method described in Subsection 8.4.

3. Calculate discs with radius ε containing the initial roots $\alpha_1, \alpha_2, \ldots, \alpha_n$ of the polynomial $f(z, s)$ using the method described in Subsection 7.12.

4. Calculate the permutation of the roots $\alpha_1, \alpha_2, \ldots, \alpha_n$ that occurs when the parameter w is moved along each of the paths $w_1(t), w_2(t), \ldots, w_m(t)$ using the Predict, Correct and Verify algorithm.

9 Implementation

In this section an implementation of the Automatic Permutation Calculation Algorithm will be discussed. The implementation is written in the Scheme programming language. Output from the program will be shown. Several issues with the implementation will be discussed. The implementation differs in several places from the algorithms described in the earlier sections.

9.1 Rounding Errors

The results produced by the Predict Correct Verify Algorithm is guaranteed to be correct. However the current implementation of said algorithm uses floating point numbers. Because of this rounding errors are introduced each time an arithmetic operation is performed and therefore the results generated by the implementation are not guaranteed to be correct. If the implementation was modified to use rational arithmetic or interval arithmetic then the results would be provably correct.

A problem with rational arithmetic is that the size of the integers in the numerator and the denominator increase each time an arithmetic operation is performed. Performing arithmetic operations on arbitrarily large integers takes longer time than performing arithmetic operations on small integers. The size of the integers can be reduced by dividing out the greatest common divisor of the numerator and the denominator. Calculating the greatest common divisor of the numerator and denominator after each arithmetic operation becomes computationally expensive.

A problem with interval arithmetic is that the size of the intervals can grow so large that it is not possible to extract any useful information from a calculation. It becomes necessary to modify the algorithms to introduce a step where the size of the intervals are decreased. For more information about interval arithmetic see [12].

9.2 Scheme

The main selection criterion for the programming language used to implement the algorithms in this thesis was that it had to be licensed under a Free Software license. Being forced to stop working on the implementation due to licensing fees was not acceptable. Another selection criterion was that the language should support automatic memory management since this lowers the development time. Hence C and C++ were rejected.

An implementation using the Ruby programming language was started. However the performance was poor and the implementation was abandoned. The Python programming language was rejected due to Pythons similarity to Ruby.

The next programming language to be tried was Scheme. There are several Open Source implementations of Scheme. PLT Scheme was selected since it has a nice development environment and the performance of numerical code is adequate. Unfortunately the members of the Scheme community do not believe in code libraries. The implementation time was thus increased since all numerical code had to be developed from scratch. Choosing Common Lisp as the implementation language would have been a better choice.

9.3 The Program

Since the author could find no numerical libraries for Scheme all numerical code had to be written from scratch. The amount of time required to develop and test a numerical library should not be underestimated. The author estimates that about three quarters of the time spent developing the program was spent on writing and debugging generic numerical code. More effort should have been spent to find an environment with an existing numerical library suited for developing the algorithms in this paper.

A numerical library that is to be used in this problem space should satisfy the following requirements:

1. The library should have functions for working with Complex Polynomials and Complex Bivariate Polynomials.
2. The library should have functions for calculating the Discriminant of a Complex Polynomial or calculating the Determinant of a Complex Matrix.
3. The library should have an implementation of a Root Finding Algorithm for Complex Polynomials.

The implementation does not calculate the discriminant of a polynomial by using the method of taking the determinant of a Sylvester Matrix. This is due to the fact that no existing code was available to calculate the determinant of a complex matrix. Instead the implementation is hard coded to calculate the discriminant of a fourth degree polynomial which has coefficients that are themselves arbitrary complex polynomials.

At one point the program needed to be modified in some places to use discs in the complex plane instead of just complex numbers. This was due to the fact that the different components of the Predict Correct Verify Algorithm operate on discs in the complex plane and not complex numbers.

9.4 Program Disposition

The program consists of the following steps:

1. Generate a random bivariate polynomial where the coefficient polynomials are parameterized by w.
2. Calculate the critical points for the bivariate polynomial.
3. Calculate a set of n-gon paths that all have the same starting point and where each path goes around one and only one of the critical points.
4. Calculates the roots of the polynomial generated at the start of the path.
5. Calculates the permutation of the initial roots generated by the Predict Correct Validate Algorithm generated when the parameter w is moved along each of the generated paths.

9.5 Program Output

In this subsection the output of a single run of the program will be displayed. The following polynomial was generated by the program.

$$\begin{aligned}&(-411/10 - 11/2i) + (-287/5 + 51/10i)w+\\&(387/5 + 474/5i)z + (811/10 - 861/10i)wz+\\&(151/5 + 9/5i)z^2 + (107/2 - 184/5i)wz^2+\\&(171/2 + 179/5i)z^3 + (33 + 55/2i)wz^3 + z^4\end{aligned}$$

The following is the output of a single run of the program. The output has been modified to fit inside the page margins.

```
Generating Bivariate Polynomial:
([-411/10-11/2i] + [-287/5+51/10i]w) +
([387/5+474/5i] + [-811/10-861/10i]w)z +
([151/5+9/5i] + [107/2-184/5i]w)z^2 +
([171/2+179/5i] + [33+55/2i]w)z^3 + ([1])z^4

Number of critical points: 6

Paths starting point: -1.927626657061293-0.058552364424453884i
Radius of Paths: 0.04235898245399947
Initial Root Discs:
(-0.11638063013351323+0.18010875963876044i 1e-10)
(-0.8891436944496411+2.0271096296405227i 1e-10)
(3.944836091446111-2.2996872673207482i 1e-10)
(-26.437822105512765+19.234429973234008i 1e-10)

Path 1
Critical Point: 0.045060695757472444-0.21366997354760417i
Permutation: (3 2 1 4)

Path 2
Critical Point: -0.31863326476913445+2.0851015811800266i
Permutation: (2 1 3 4)

Path 3
Critical Point: -1.56626021995974+0.7854460438570761i
Permutation: (1 2 4 3)

Path 4
Critical Point: -2.4285205160327554+0.5928567378235748i
Permutation: (1 4 3 2)

Path 5
Critical Point: -2.5555546613898352+0.4807368902228575i
Permutation: (1 2 4 3)
```

```
Path 6
Critical Point: 6.814778920027891-2.3741733270473486i
Permutation: (2 1 3 4)
```

The program finds the following 6 critical points.

$$c_1 = 0.045060695757472444 - 0.21366997354760417i$$
$$c_2 = -0.31863326476913445 + 2.0851015811800266i$$
$$c_3 = -1.56626021995974 + 0.78544604385707661i$$
$$c_4 = -2.4285205160327554 + 0.5928567378235748i$$
$$c_5 = -2.5555546613898352 + 0.4807368902228575i$$
$$c_6 = 6.814778920027891 - 2.3741733270473486i$$

The calculated starting point for the paths is equal to

$$-1.927626657061293 - 0.058552364424453884i.$$

The program calculates that the initial roots of the polynomial at the start of the path are located in discs with radius 10^{-10} centered on the points

$$r_1 = -0.11638063013351323 + 0.18010875963876044i,$$
$$r_2 = -0.8891436944496411 + 2.0271096296405227i,$$
$$r_3 = 3.944836091446111 - 2.2996872673207482i,$$
$$r_4 = -26.437822105512765 + 19.234429973234008i.$$

Let $p(c_k)$ denote the permutation of the roots generated by the program when going around the critical point c_k. The following permutations are generated by the program

$$p(c_1) = (3\,2\,1\,4), \quad p(c_2) = (2\,1\,3\,4), \quad p(c_3) = (1\,2\,4\,3),$$
$$p(c_4) = (1\,4\,3\,2), \quad p(c_5) = (1\,2\,4\,3), \quad p(c_6) = (2\,1\,3\,4).$$

References

[1] Mark J. Ablowitz and Athanassios S. Fokas. *Complex Variables Introduction and Application.* Cambridge, New York, second edition, 2003.

[2] Stephen G. Kranz and Harold R. Parks. *The Implicit Function Theorem, History, Theory and Applications*, chapter 2, pages 27–28. Birkhäuser, Boston, 2002.

[3] Alexander Morgan. *Solving Polynomial Systems Using Continuation for Engineering and Scientific Problems.* SIAM, Philadelphia, 2009.

[4] Arne Persson and Lars-Christer Böiers. *Analys i Flera Variabler*, pages 335–337. Studentlitteratur AB, Lund, third edition, 2005.

[5] Anthony Ralston and Philip Rabinowitz. *A First Course in Numerical Analysis.* McGraw-Hill, second edition, 1978.

[6] Walter Rudin. *Principles of Mathematical Analysis.* McGraw-Hill, third edition, 1976.

[7] Richard A. Silverman. *Modern Calculus and Analytic Geometry.* Dover, 2002.

[8] Andrew J. Sommese and Charles W. Wampler, II. *The Numerical Solution of Systems of Polynomials Arising in Engineering and Science.* World Scientific Publishing Co., Singapore, 2005.

[9] Joseph L. Taylor. *Complex Variables.* `http://www.math.utah.edu/~taylor/complexanalysis`, 2.2 edition, 2009.

[10] E. W. Weisstein. Laguerre's method. MathWorld – A Wolfram Web Resource. `http://mathworld.wolfram.com/LaguerresMethod.html`, 2010.

[11] Wikipedia. Analytic continuation — wikipedia, the free encyclopedia, 2010. [Online; accessed 25-May-2010].

[12] Wikipedia. Interval arithmetic — wikipedia, the free encyclopedia, 2010. [Online; accessed 27-August-2010].

[13] Wikipedia. Winding number — wikipedia, the free encyclopedia, 2010. [Online; accessed 15-May-2010].

[14] William T. Vetterling William H. Press, Saul A. Teukolsky and Brian P. Flannery. *Numerical Recipies in C.* Cambridge University Press, second edition, 1992.