Algebraic Systems of Equations and Computational Complexity Theory

Zeke Wang Senlin Xu and Tang'an Gao

SCIENCE PRESS
Beijing

Zeke Wang Senlin Xu and Tang'an Gao

Algebraic Systems of Equations and Computational Complexity Theory

SCIENCE PRESS
Beijing

Published by Science Press
16 Donghuangchenggen North Street
Beijing 100717, China
Second Printed in 2006

Printed in Beijing

ISBN 7-03-003944-0 (Science Press, Beijing)

Preface

In the past fifteen years, significant progress has been made in the solution of nonlinear systems, particularly in the areas of computing fixed points, solving nonlinear equation systems, and in applying these methods to equilibrium models. This progress has developed along two principal lines : simplicial and continuation methods. Simplicial methods stem from the pioneering work of Scarf on the approximation of fixed points. As shown by Kuhn, the essential idea is the use of a simplicial approximation of the map as is used in the proof of Brouwer's fixed point theorem by Sperner's lemma. Continuation methods originate in the the work of Kellogg, Li, and Yorke which converted the nonconstructive proof by Hirsch of Brouwer's theorem into a constructive algorithm.

The development of these two lines of approach has been parallel in many respects. For example, both have used the idea of a homotopy to make a transition from an easy problem to a difficult problem. On the other hand, the two points of view are contrasted in the behavior of numerical algorithms, where continuation methods normally work with probability near one (excluding "bad" cases) while simplicial methods frequently work without exception.

The book by Wang and Xu presents a self-contained exposition of recent work on simplicial and continuation methods applied to the solution of algebraic equations. For the case of the search for the roots of a single polynomial over the complex field, the simplicial algorithm studied is that proposed by Kuhn. Wang and Xu give a complete and self-contained exposition of this algorithm. This is followed by a discussion of error, cost, and efficiency, areas to which Wang and Xu have made original and interesting contributions. For the same problem approached by continuation methods, the starting point is Smale's recent study of a global Newton method.

Their exposition of Smale's work is notable for its clarity; in addition, they have corrected numerous errors in the only published version available and have filled several gaps in the arguments. This discussion is followed by original research that compares the cost estimates for Kuhn's algorithm and Smale's estimates for Newton's method.

The second half of the book deals with very recent research on systems of algebraic equations. It is notable for its exposition of the various tools from widely different mathematical subjects (such as algebraic, geometry and differential topology) that have been applied to this problem. As in the first part of the book, both continuation and simplicial methods are discussed. The final chapter contains contributions the authors have made to the design of a simplicial homotopy algorithm for the numerical solution of systems of nonlinear algebraic equations.

It is a pleasure to introduce this book to the reader and student. It is certain that, by their careful exposition of this active area of research, Wang and Xu will generate interest and make further progress on these problems possible.

H.W. Kuhn

Contents

Chapter 1

Kuhn's Algorithm for Algebraic Equations

This chapter is devoted to Kuhn's algorithm for algebraic equations and to the proof of its convergence.

The main references are [Kuhn, 1974; 1977].

Contrary to all traditional methods of iteration, Kuhn's algorithm is based on simplicial triangulation of the underlying space, an integer labelling, and a complementary pivoting procedure of computation. If its description is not so simple as, say, Newton method, after certain implementation, its use is, however, tremendously easy. To solve any algebraic equation by using Kuhn's algorithm, the only thing one should do is to input the complete set of its coefficients as well as the accuracy demand into the machine. Then the algorithm will find one by one all solutions with no more care needed. For Kuhn's algorithm, there are no difficult problems like the selection of initial values. It is a method with a very strong guarantee of global convergence.

Finally, for the purpose of only implementing the algorithm, it is enough to know the first two sections.

§1. Triangulation and Labelling

Let $f(x)$ be a monic polynomial of degree n in the complex variable z with complex numbers as its coefficients, that is, $f(z) = z^n + a_1 z^{n-1} + \cdots + a_n$, where n is a positive integer and $a_1, \cdots, a_n$ are complex constants. If a complex number ξ satisfies $f(\xi) = 0$, ξ is called a zero of the polynomial f or a solution of the algebraic equation $f(z) = 0$. Kuhn's algorithm is designed to find all zeros of f.

Denote by $\mathbf{C}$ the plane of complex $z = x + i\,y$ and by $\mathbf{C}'$ the plane of complex $w = u + i\,v$. Then $w = f(z)$ defines a polynomial mapping $f : \mathbf{C} \to \mathbf{C}'$.

To describe Kuhn's algorithm in the next section, we now introduce a triangulation of the half-space $\mathbf{C} \times [-1, +\infty)$ and a labelling rule for the vertices of the triangulation.

Let $\mathbf{C}_d$ denote the replica $\mathbf{C} \times \{d\}$ of the plane $\mathbf{C}$ for $d = -1, 0, 1, 2, \cdots$. Then $\mathbf{C}_d \subseteq \mathbf{C} \times [-1, +\infty)$. Given a center $\tilde{z}$ and a grid size h, we define the triangulation $\mathbf{T}$ of $\mathbf{C} \times [-1, \infty)$ as follows.

Triangulation $\mathbf{T}_d(\tilde{z}; h)$ or $\mathbf{T}_d$ of the plane $\mathbf{C}_d$.

The triangulation $\mathbf{T}_{-1}(\tilde{z}; h)$ of $\mathbf{C}_{-1}$ is illustrated in Fig. 1.1. A triangle in $\mathbf{T}_{-1}(\tilde{z}; h)$ is uniquely determined by a pair of integers (r, s) with $r + s$ even and $(a, b) \in \{(1,0), (0,1), (-1,0), (0,-1)\}$. The z-coordinates of its vertices are:

$$\tilde{z} + (r + i\,s)h;$$

$$\tilde{z} + ((r + a) + i\,(s + b))h;$$

$$\tilde{z} + ((r - b) + i\,(s + a))h.$$

The supremum of the diameters of its triangles is called the mesh of the triangulation $\mathbf{T}_{-1}(\tilde{z}; h)$. The mesh of the triangulation $\mathbf{T}_{-1}(\tilde{z}; h)$ is obviously $\sqrt{2}h$.

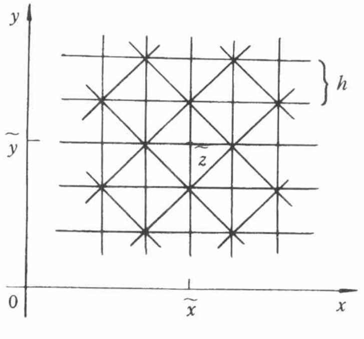

Figure 1.1

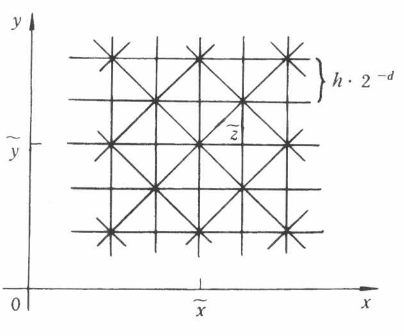

Figure 1.2

The triangulation $\mathbf{T}_d(\tilde{z}; h)$ of $\mathbf{C}_d$ is illustrated in Fig. 1.2, where $d = 0, 1, 2, \cdots$. A triangle in $\mathbf{T}_d(\tilde{z}; h)$, where $d \geq 0$, is uniquely specified by an $(a, b) \in \{(1,0), (0,1), (-1,0), (0,-1)\}$ and a pair of integers (r, s) with

$r + s$ odd. The z-coordinates of its vertices are

$$\tilde{z} + (r + i\,s)h2^{-d};$$
$$\tilde{z} + ((r + a) + i\,(s + b))h2^{-d};$$
$$\tilde{z} + ((r - b) + i\,(s + a))h2^{-d}.$$

With a similar definition, the mesh of $\mathbf{T}_d(\tilde{z}; d)$ is $\sqrt{2}h2^{-d}$.

Notice that every triangle in the triangulation $\mathbf{T}_d(\tilde{z}; h)$ is an isosceles right triangle with two right-angle sides paralleling to the x-axis and to the y-axis, respectively.

Remark 1.1 The triangulation $\mathbf{T}_d(\tilde{z}; h)$ can be easily portrayed by four families of parallel lines. But the advantage of the above description is that every vertex in the triangulation is specified by two pairs of integers r, s and a, b.

Triangulation $\mathbf{T}(\tilde{z}; h)$ or $\mathbf{T}$ of the half-plane $\mathbf{C} \times [-1, \infty)$.

By definition of the triangulation $\mathbf{T}_d(\tilde{z}; h)$, for every square in $\mathbf{C}_{-1}$ consisting of exactly two triangles in $\mathbf{T}_{-1}(\tilde{z}; h)$ with a common hypotenuse, there is a unique "above-opposite" square in $\mathbf{C}_0$ made of precisely two triangles in $\mathbf{T}_0(\tilde{z}; h)$ with a common hypotenuse, and the two hypotenuses are orthogonal each other. The cube between $\mathbf{C}_{-1}$ and $\mathbf{C}_0$ determined by two squares is subdivided into five tetrahedra with the manner shown in Fig. 1.3. All such cubes are treated in the same way.

Similarly, for $d \geq 0$, every square in $\mathbf{C}_d$ made of exactly two triangles in $\mathbf{T}_d$ and its four above-opposite squares in $\mathbf{C}_{d+1}$ determine an elementary guadrangular prism between $\mathbf{C}_d$ and $\mathbf{C}_{d+1}$, the prism is subdivided into fourteen tetrahedra as in Fig. 1.4

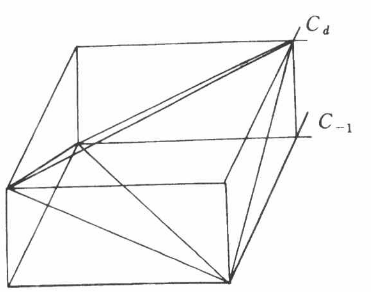

Figure 1.3

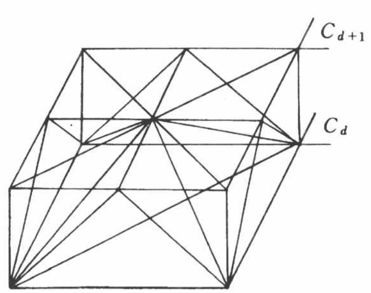

Figure 1.4

In this way we obtain the simplicial triangulation $\mathbf{T}(\tilde{z};h)$ of the half-space $\mathbf{C}\times[-1,\infty)$, shortly denoted by $\mathbf{T}$. Notice that no new vertices have been added to define the triangulation, that is, all vertices of the triangulation $\mathbf{T}(\tilde{z};h)$ are the vertices of the triangulation $\mathbf{T}_d(\tilde{z};h)$ for $d=-1,0,1,2,\cdots$. We denote the vertex set of $\mathbf{T}(\tilde{z};h)$ by $V(\mathbf{T}(\tilde{z};h))$ or $V(\mathbf{T})$.

In the algorithm to be presented in the next section, we shall be concerned very often with triples of points $\{(z_1,d_1),(z_2,d_2),(z_3,d_3)\}$ (or shortly $\{z_1,z_2,z_3\}$) which are the vertices of a triangular face of some tetrahedron in the triangulation. The name "triple" used in next several sections has always this meaning.

For the triangulation, the following lemma is obvious.

Lemma 1.2 *Suppose that $\{(z_1,d_1),(z_2,d_2),(z_3,d_3)\}$ is a triple in $\mathbf{T}$. Let $d=\min\{d_1,d_2,d_3\}$. Then $d\le d_k\le d+1$ for $k=1,2,3$.* ¶

In the case of Lemma 1.2, we say that the triple $\{z_1,z_2,z_3\}$ lies between levels $\mathbf{C}_d$ and $\mathbf{C}_{d+1}$. In particular, when $d_1=d_2=d_3$, we say that the triple $\{z_1,z_2,z_3\}$ lies in $\mathbf{C}_d$.

Let $\{(z_1,d_1),(z_2,d_2),(z_3,d_3)\}$ be a triple in $\mathbf{T}$. Define the diameter of the triple by

$$\mathrm{diam}\{(z_1,d_1),(z_2,d_2),(z_3,d_3)\}=\max\{|z_1-z_2|,|z_2-z_3|,|z_3-z_1|\}.$$

One may instead denote it shortly by $\mathrm{diam}\{z_1,z_2,z_3\}$. Notice that it is, in fact, a projective diameter.

Lemma 1.3 *Suppose that the triple $\{z_1,z_2,z_3\}$ lies between levels $\mathbf{C}_d$ and $\mathbf{C}_{d+1}$. Then*

$$\mathrm{diam}\{z_1,z_2,z_3\}\le\sqrt{2}h2^{-d}.$$

Proof. This inequality is obvious since the diameter of all possible triples lying between levels $\mathbf{C}_d$ and $\mathbf{C}_{d+1}$ is easily found from Figs. 1.3 and 1.4. ¶

This lemma shows the fact that the higher the level $\mathbf{C}_d$ is, the smaller is the diameter of the triple in $\mathbf{C}_d$ or above $\mathbf{C}_d$.

Now, we turn to present the labelling rule for the vertices of the triangulation $\mathbf{T}(\tilde{z};h)$.

Given any nonzero complex number $w=u+iv$, define $\arg w$ to be the unique real number α satisfying

$$-\pi<\alpha\le\pi,$$

$$\cos\alpha = \frac{u}{\sqrt{u^2+v^2}},$$

and

$$\sin\alpha = \frac{v}{\sqrt{u^2+v^2}}.$$

Definition 1.4 Define an assignment $l : \mathbf{C} \to \{1,2,3\}$ by

$$l(z) = \begin{cases} 1, & \text{if } -\pi/3 \leq \arg f(z) \leq \pi/3 \text{ or } f(z) = 0; \\ 2, & \text{if } \quad \pi/3 < \arg f(z) \leq \pi; \\ 3, & \text{if } \quad -\pi < \arg f(z) < -\pi/3. \end{cases}$$

l is called the labelling of the z-plane $\mathbf{C}$ induced by the polynomial f, and $l(z)$ is called the label of z. See Fig. 1.5 for an illustration.

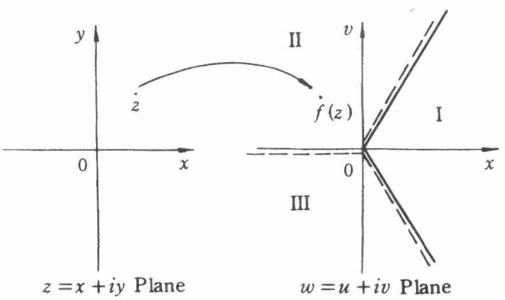

Figure 1.5

Definition 1.5 Let $f_{-1}(z) = (z-\tilde{z})^n$, and $f_d(z) = f(z)$ for $d = 0,1,2,\cdots$. Define $l : V(\mathbf{T}(\tilde{z};h)) \to \{1,2,3\}$ by

$$l(z,d) = \begin{cases} 1, & \text{if } -\pi/3 \leq \arg f_d(z) \leq \pi/3 \text{ or } f_d(z) = 0; \\ 2, & \text{if } \quad \pi/3 < \arg f_d(z) \leq \pi; \\ 3, & \text{if } \quad -\pi < \arg f_d(z) < -\pi/3. \end{cases}$$

l is called the labelling of $V(\mathbf{T}(\tilde{z};h))$ induced by the polynomial f, and $l(z,d)$ is called the label of the vertex (z,d).

Notice that we use the same notation l in Definition 1.4 and in Definition 1.5. No confusions could arise since, for every vertex $(z,d) \in V(\mathbf{T}(\tilde{z};h))$, it is clear that its label is either determined by $f(z)$ when $d \geq 0$ or determined by $(z-\tilde{z})^n$ when $d = -1$.

Based on the triangulation and the labelling, we can now introduce the concept of complete triples.

Definition 1.6 The triple $\{z_1, z_2, z_3\}$ is completely labelled by f if $\{l(z_1), l(z_2), l(z_3)\} = \{1, 2, 3\}$. In this case, we may also simply say that $\{z_1, z_2, z_3\}$ is a complete triple.

For simplicity of expressions and of specification, from now on we assume that $l(z_k) = k, k = 1, 2, 3$ for any given complete triple $\{z_1, z_2, z_3\}$.

The definition of complete triples does not point out whether the labels of its vertices are determined by $f(z)$ or by $(z - \tilde{z})^n$. In fact, for some triple $\{z_1, z_2, z_3\}$, it is possible that all of its vertices are labelled by $f(z)$, or all are labelled by $(z - \tilde{z})^n$, or partially labelled by $f(z)$ and partially labelled by $(z - \tilde{z})^n$.

The following result establishes certain connections between triples completely labelled by $f(z)$ and the zeros of $f(z)$.

Lemma 1.7 *Let $\{z_1, z_2, z_3\}$ be a complete triple whose labels are determined by $f(z)$, and $|f(z_k) - f(z_l)| \leq \eta$ for $k, l = 1, 2, 3$. Then for $k = 1, 2, 3$,*

$$|f(z_k)| \leq \frac{2\eta}{\sqrt{3}}.$$

Proof. The sectors 1, 2 and 3 of the w-plane illustrated in Fig. 1.6 are the three ranges whose preimages are respectively labelled 1, 2 and 3. The label of a point z is determined by the sector into which $f(z)$ falls. According to the convention $l(z_k) = k$ for $k = 1, 2, 3$, if $|f(z_k) - f(z_l)| \leq \eta$ for $k, l = 1, 2, 3$, then $f(z_1)$ must lie in Sector 1 and both the distances between $f(z_1)$ and Sectors 2 and 3 are smaller than η. Thus $f(z_1)$ must lie in the shadowed area as shown in Fig. 1.6. Hence, $|f(z_1)| \leq 2\eta/\sqrt{3}$.

Similarly, $|f(z_2)| \leq 2\eta/\sqrt{3}$ and $|f(z_3)| \leq 2\eta/\sqrt{3}$. ¶

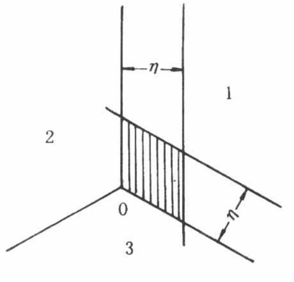

Figure 1.6

It is well-known that every polynomial function is uniformly continuous in any bounded domain. If we can find a triple completely labelled by f with small diameter, then the distances between the images of its vertices are also small and so are the distances between the images and the origin of the w-plane (by Lemma 1.7). If the distances are small enough, then every point of the triple can serve as a numerical zero of f with certain accuracy. Furthermore, we have learned that the higher the level $\mathbf{C}_d$ is, the smaller is the diameter of the triple in $\mathbf{C}_d$. These motivate to design an algorithm to locate complete triples whose projections always lie in a given bounded domain and the complete triples run through higher and higher levels. This will be done in the next section.

§2. Complementary Pivoting Algorithm

We first present several useful lemmas.

Let $Q_m(\tilde{z};h)$ denote the square in $\mathbf{C}$ with corners at $\tilde{z}+mh(\pm 1 \pm i)$, where m is a positive integer, that is, $Q_m(\tilde{z};h)$ is a square with center at $\tilde{z}$, the lengths of its four sides are all $2mh$ and each side parallels the x-axis or the y-axis (Fig. 1.7).

A side of some triangle in the triangulation $\mathbf{T}$ is called an edge of the triangulation. The boundary $\partial Q_m(\tilde{z};h)$ of $Q_m(\tilde{z};h)$ is oriented counterclockwise. With the notation $\{z', z''\}$ for an edge on $\partial Q_m(\tilde{z};h)$, we call $\{z', z''\}$ a positive edge if the direction of the edge from z' to z'' coincides with the direction of $\partial Q_m(\tilde{z};h)$, otherwise $\{z', z''\}$ is called a negative edge. The triangles of $\mathbf{T}_{-1}(\tilde{z};h)$ inside $Q_m(\tilde{z};h)$ are oriented in the customary counterclockwise cyclic order of their vertices. With the notation $\{z', z'', z'''\}$ for a triangle in $\mathbf{T}_{-1}(\tilde{z};h)$, we call $\{z', z'', z'''\}$ a positive triangle (triple) if the ordering of z', z'', z''' gives the positive direction of the triangle, otherwise it is a negative triangle. We may simply denote $Q_m(\tilde{z};h)$ and $\partial Q_m(\tilde{z};h)$ by Q_m and ∂Q_m respectively.

The angle spanned by z', z'' with respect to z^* is defined to be the unique angle no larger than π and spanned by two rays starting from z^* and passing through z' and z'' respectively. It is also called the angle spanned by the line segment $z'z''$ with respect to z^*.

By saying that an edge is labelled (k, l), we mean that its starting vertex is labelled k while the ending one is labelled l.

Lemma 2.1 *If* $m \geq \dfrac{3n}{2\pi}$, *then in the counterclockwise direction of*

∂Q_m, there are exactly n $(1,2)$-labelled edges on ∂Q_m and no $(2,1)$-labelled edges.

Proof. Let $\{z', z''\}$ be an edge on ∂Q_m. Denote by α the angle spanned by z' and z'' with respect to $\tilde{z}$. Refer to Fig. 1.7, it is evident that

$$0 < \alpha \leq \arctan\left(\frac{h}{mh}\right) < \frac{1}{m} \leq \frac{2\pi}{3n}.$$

Let β be the angle spanned by $w' = (z' - \tilde{z})^n$ and $w'' = (z'' - \tilde{z})^n$ with respect to the origin of the w-plane. Then

$$0 < \beta = n\alpha < \frac{n}{m} \leq \frac{2\pi}{3}.$$

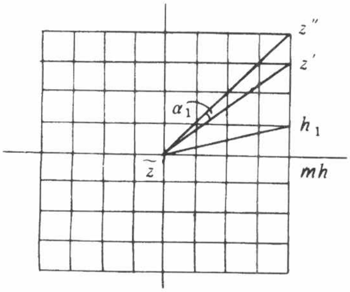

Figure 1.7

From the structure of Q_m and the properties of the function $w = (z - \tilde{z})^n$, the image of ∂Q_m runs around the origin of the w-plane n times. Since $0 < \beta < 2\pi/3$, the angle spanned by the images of some edge on ∂Q_m with respect to the origin is smaller than $2\pi/3$. Hence, starting from $w = (mh)^n$ in the w-plane, the image of ∂Q_m runs exactly n times from the sector 1 into the sector 2. Thus, in the z-plane, there are exactly n (1,2)-labelled edges on ∂Q_m.

Similarly, if $l(z') = 2$ then we have $l(z'') = 2$ or 3 and never $l(z'') = 1$ due to $0 < \beta < 2\pi/3$. Hence there is no (2,1)-labelled edges on ∂Q_m. ¶

Now, starting from $z = \tilde{z} + mh \in \partial Q_m$ and following ∂Q_m in its positive direction, we number the n (1,2)-labelled edges from 1 to n. This order of the (1,2)-labelled edges will be frequently used in the next sections.

Lemma 2.2 *If $m \geq \dfrac{3(1+\sqrt{2})n}{4\pi}$, then outside Q_m there is no complete triples labelled by $(z - \tilde{z})^n$.*

Proof. We first prove that if $z'z''$ is an edge on ∂Q_m or outside Q_m then the angle spanned by $z'z''$ with respect to $\tilde{z}$ is smaller than $\frac{2\pi}{3n}$. In fact, if $z'z''$ parallels the x-axis or the y-axis then the result follows directly from Lemma 2.1. Now, let $z'z''$ be the hypotenuse of a triangle in $\mathbf{T}_{-1}$ outside Q_m. Due to the structure of Q_m, the angle reaches its maximum only when the edge $z'z''$ intersecting with ∂Q_m. Without loss of generality, let k be a positive integer such that $z' = \tilde{z} + h(m + 1 + i\,k)$ and $z'' = \tilde{z} + h(m + i\,(k+1))$. Referring to Fig. 1.8, we have

$$\alpha = \arctan\frac{k+1}{m} - \arctan\frac{k}{m+1},$$

$$\tan\alpha = \frac{m+k+1}{m^2+m+k^2+k}.$$

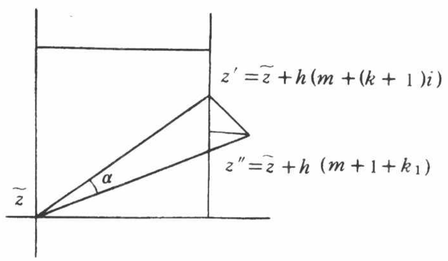

Figure 1.8

Hence, as a function of k, at $k = \sqrt{2m(m+1)} - m - 1$, $\tan\alpha$ takes its maximum

$$\begin{aligned}\frac{m+k+1}{m^2+m+k^2+k} &= \frac{1}{2\sqrt{2m(m+1)}-2m-1}\\ &< \frac{1}{2m(\sqrt{2}-1)}\\ &= \frac{1+\sqrt{2}}{2m}.\end{aligned}$$

Of course, the above inequality is true for any integer k. Notice also that $\alpha < \pi/2$, we obtain

$$\alpha < \tan\alpha < \frac{1+\sqrt{2}}{2m} \leq \frac{2\pi}{3n}.$$

Now let $\{z', z'', z'''\}$ be a triple in $\mathbf{T}_{-1}(\tilde{z}; h)$ outside Q_m. Then the angle spanned by each edge of the triple with respect to $\tilde{z}$ is less than $\frac{2\pi}{3n}$. Let $w' = (z' - \tilde{z})^n, w' = (z'' - \tilde{z})^n$ and $w''' = (z''' - \tilde{z})^n$. Then the angle spanned by a pair of w', w'', w''' with respect to the origin of the w-plane is less than $2\pi/3$. Combining with the following Lemma (2.3), $\{z', z'', z'''\}$ is not a triple completely labelled by $(z - \tilde{z})^n$. ¶

Lemma 2.3 *Let $\{z', z'', z'''\}$ be a triple in the z-plane, and let w', w'', w''' denote respectively the images of z', z'', z''' under mapping $w = (z-\tilde{z})^n$ or $w = f(z)$. If none of w', w'', w''' is zero and the angle spanned by any pair of w', w'', w''' with respect to the origin of the w-plane is less than $2\pi/3$, then the triple $\{z', z'', z'''\}$ is not complete.*

Proof. Suppose otherwise that $\{z', z'', z'''\}$ is complete and

$$l(z') = 1, \quad l(z'') = 2, \text{ and } l(z''') = 3.$$

In the w-plane, let α, β, γ denote respectively the less-than-2π angle from $0w'''$ to $0w'$, from $0w'$ to $0w''$, and from $0w''$ to $0w'''$.(cf. Fig. 1.9) Then $\alpha > 0, \beta > 0, \gamma > 0$ and $\alpha + \beta + \gamma = 2\pi$.

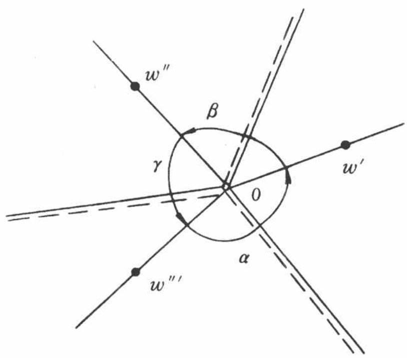

Figure 1.9

Now , if $\alpha > \pi$ then the angle spanned by $w'''w'$ with respect to the origin is $2\pi - \alpha$. By the assumption of this lemma, $2\pi - \alpha < 2\pi/3$, thus $\alpha > 4\pi/3$. But according to the labelling, $\alpha \leq 4\pi/3$. This is a contradiction. Similarly, $\beta > \pi$ or $\gamma > \pi$ will also lead to contradictions. Finally, if none of α, β, γ is larger than π, then α, β, γ are exactly the angles spanned respectively by the three pairs of w', w'' and w''' with respect to the origin, and thus are all smaller than $2\pi/3$. This contradicts