INFORMATION TO USERS

This reproduction was made from a copy of a manuscript sent to us for publication and microfilming. While the most advanced technology has been used to photograph and reproduce this manuscript, the quality of the reproduction is heavily dependent upon the quality of the material submitted. Pages in any manuscript may have indistinct print. In all cases the best available copy has been filmed.

The following explanation of techniques is provided to help clarify notations which may appear on this reproduction.

1. Manuscripts may not always be complete. When it is not possible to obtain missing pages, a note appears to indicate this.

2. When copyrighted materials are removed from the manuscript, a note appears to indicate this.

3. Oversize materials (maps, drawings, and charts) are photographed by sectioning the original, beginning at the upper left hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is also filmed as one exposure and is available, for an additional charge, as a standard 35mm slide or in black and white paper format.*

4. Most photographs reproduce acceptably on positive microfilm or microfiche but lack clarity on xerographic copies made from the microfilm. For an additional charge, all photographs are available in black and white standard 35mm slide format.*

***For more information about black and white slides or enlarged paper reproductions, please contact the Dissertations Customer Services Department.**

Kim, Myong-Hi

COMPUTATION COMPLEXITY OF THE EULER ALGORITHMS FOR THE ROOTS OF COMPLEX POLYNOMIALS

City University of New York

PH.D. 1986

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark __✓__.

1. Glossy photographs or pages _____
2. Colored illustrations, paper or print _____
3. Photographs with dark background _____
4. Illustrations are poor copy _____
5. Pages with black marks, not original copy _____
6. Print shows through as there is text on both sides of page _____
7. Indistinct, broken or small print on several pages ✓
8. Print exceeds margin requirements _____
9. Tightly bound copy with print lost in spine _____
10. Computer printout pages with indistinct print _____
11. Page(s) _____ lacking when material received, and not available from school or author.
12. Page(s) _____ seem to be missing in numbering only as text follows.
13. Two pages numbered _____. Text follows.
14. Curling and wrinkled pages _____
15. Dissertation contains pages with print at a slant, filmed as received ✓
16. Other _____

_____

_____

University
Microfilms
International

COMPUTATION COMPLEXITY OF THE EULER ALGORITHMS FOR THE ROOTS OF COMPLEX POLYNOMIALS

by

MYONG-HI KIM

A disseration submitted to the Graduate Faculty in Mathematics in Partial fullfilment of the Requirements for the degree of Doctor of Philosophy, The City University of New York.

1986

This manuscript has been read and accepted for the Graduate Faculty in Mathematics in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

11/6/85
Date

Michael Shub
Chair of Examining Committee

11/1/85
Date

E. Falcon Oq
Executive Officer

Professor Linda Keen (L)

Professor Alphonse Vasquez (GC)

Professor Michael Anshel (C)

Professor Michael Shub (Q)
Supervisory Committee

The City University of New York

Abstract

COMPUTATIONAL COMPLEXITY OF THE EULER TYPE ALGORITHMS FOR THE ROOTS OF COMPLEX POLYNOMIALS

by

Myong-Hi Kim

Adviser:Professor Mike Shub

In this thesis we show that the Euler type algorithms find

i) a complex number z such that $|f(z)|<\xi$, for any $\xi > 0$, approximately,

with $6672(d(d+\log|\log\xi|)M(26(d+|\log\xi|)$ binary bit operations,

ii) an approximate zero for a polynomial of degree d without multiple roots with average bit complexity approximately,

$6672(d\ M(26d))$ binary bit operations,

where M(n) denotes the bit complexity for the multiplication of two n-digit numbers.

To my Parents, with love

ACKNOWLEDGEMENTS

I thank my advisor , Professor Mike Shub ; without him my dream of being a Doctor in Mathematics would not have come true.

I am very grateful to Professors L.Keen , A. Vasquez and M. Anshell B. Randol for their interest and efforts in carefully reading my thesis.

I also thank Professors H. Wozniaski and S. Smale for their interest in my work.

I am very thankful to Professors M.Kuga , L.Goldberg and L.Blum for their encuragement and help through difficult times.

I give my special thank to Dennis Sullivan for providing me a good environment to work and for his friendship.

I give my love to my friends in New York Les.,Kathy, Misha, Raquel, Terry, and my brother Seung-Sik, for their being with me.

Finally, for Michael , for his support and love.

TABLE OF CONTENTS

Chapter 0. Introduction

1. Problems

In this thesis, we study the computational complexity of Shub & Smale's algorithmic approach to the Fundamental Theorem of Algebra (FTA). For a polynomial f of degree d , the FTA guarantees the existence of a complex solution of the equation $f(z) = 0$. However, most existence proofs in mathematics are not constructive in the sense that they do not lead to algorithms that can be used to find zeros.

In numerical analysis, one is interested in constructing algorithms, in particular efficient algorithms, to locate approximate solutions of complex polynomials. The total cost, or running time on a computer is called the computational complexity of the algorithm.(Borodin & Munro, Blum&Shub)

One approach to the analysis of the computational complexity of algebraic problems presupposes a model of computation with arithmetic exact and of unit cost,for example the real number model. But actual machines use a finite discrete process of computing. In other words, the machine can represent only a finite set of numbers and the arithmetic operations defined for the machine differs from the usual arithmetics of mathematics.For example, machine multiplication takes two n-digit numbers (n is called the machine precision) as input and produces an n-digit number output. The error between machine arithmetics and mathematical arithmetic is called the round-off error.

In the actual machine, all the arithmetic operations in algorithms are replaced by machine arithmetic. As the algorithm proceeds, round-off errors are accumulated and may be harmful.Neverthless the study of tolerable round off errors and achievable accuracy is often neglected in the algorithmic approach to a numerical problem and the theory of its cost. This problem was raised with respect to the FTA in Smale[1], Shub & Smale [I] [II]; They proved the FTA by an algorithmic approach based upon a modification of Newton's method (called Euler's method) and estimated the cost of their algorithm on a real number model machine.

More precisely, the main concerns of this thesis are;

Problem 1.(Smale 1981) Analyze the Euler and related algorithms for the FTA with respect to round-off errors, i.e. in the machine.

Problem 2.(Shub&Smale[2]) Let $P_d(1) = \{ \sum_{i=0}^{d} a_i z^i \; ; \; a_d=1 \; , |a_i| \leq 1 \}$. Find the average cost over $P_d(1)$ of an Euler iterational scheme to locate an approximate zero of f in the machine, where average is is taken over the Lebesque measure on $P_d(1) \subset C^d$.

Roughly speaking, z is called an approximate zero of f if $|f(z_n)| < 1/2^{2^n} |f(z)|$, where $z_n = N^n(z)$ and N is Newton's iteration. This notion is due to Smale.

The cost depends on the input size (eg, the number of decimal digits) and the machine precision as well as the machine model. Meanwhile the machine precision and the input size should be determined by the sensitivity of the algorithm with respect to input perturbation and round-off errors, for the success of the algorithm.

2.Results.

We solve problem 1 and use the result to answer problem 2.

By relative error we mean the following ; Suppose a is represented by some number $\tilde{a}$. Then the relative error of $\tilde{a}$ is $\frac{|\tilde{a} - a|}{|a|}$

Then for problem 1, we have

Theorem 2.19. There is an algorithm based on the Euler iteration which locates ζ such that $|f(\zeta)|<\varepsilon$, for any $\varepsilon > 0$, provided $f(z)$ and $f'(z)$ are computed with a relative error less than 1/4000.

In particular, each coefficient a_i of f is to be represented by at least $(11+ d\log^{+}|z| + \log d - \log \mathrm{Min}(|f'|,|f|)$) exact digits.

Our main theorem concerns the algorithm FAST-ROOT.(See Chapter 3)

Let M(n) denote the number of bit operations to multiply two n-bit numbers. See Chapter 4.

Main Theorem. The algorithm FAST-ROOT locates an approximate zero of of any f with no multiple root with average cost over $P_d(1)$ less than

i) $C_1(d^2\, M(d))$ bit operations,

ii) $C_0 d$ Euler iterations , where $C_0 \approx 846$, $C_1 \approx 6672$.

Remark. For the special case of the real number model, $O(d^2)$ arithmetic operations may be obtained by taking $M(n) = 1$.

The algorithm referred to in the above theorem is FAST-ROOT which we constructed in this thesis ; it is Newton's method applied to a translation of a given polynomial f, and related to the algorithm developed in SS[II],and also studied in Smale [II]. Smale also proved linear dependency of the number of iterations in Smale[II].

We note that FAST-ROOT refines the estimates of SS[II]; this is due to the subroutine Approximate-Test,an application of the following theorem.

Theorem A.2. Let f(z) be a polynomial of degree d. Then z is an approximate zero of f if

$$\left| \frac{f(z)}{f'(z)} \right| \left| \frac{f^{(j)}(z)}{j!f'(z)} \right|^{1/j-1} < \frac{1}{42} \quad \text{,for all } j = 2,..,d.$$

Conversations with Dennis Sullivan were useful in formulating this result. Smale has recently extended it to Banach Spaces with a better constant.

CHAPTER 1 . Preliminaries.

Let $f : C \longrightarrow C$ be a complex polynomial of degree d. Topologically, f is a branched covering of degree d with branch points at the critical points $\{\theta_i\}$ of f,ie those points $f'(\theta_i) = 0$. In particular, f does not admit a global inverse if $d > 1$. However, if $z_0 \in C$ is not a critical point of f , i.e. $f'(z_0) \neq 0$, then there is a localy well defined analytic inverse $f_{z_0}^{-1}$ of f such that $f_{z_0}^{-1}(f(z_0)) = z_0$.

Let D_{f,z_0} be the maximal open disk about $f(z_0)$ with radius R_{f,z_0} where $f_{z_0}^{-1}$ is well defined. Thus R_{f,z_0} is the radius of convergence of the power series expansion for $f_{z_0}^{-1}$ about $f(z_0)$.

It is well known that ;(cf,Smale 1)

Fact 1. $R_{f,z_0} = |f(z_0) - f(\theta_i)|$ for some critical point θ_i of f.

Hence $R_{f,z_0} \geq \min_{f'(\theta)=0} |f(z_0) - f(\theta)|$.

In the situation where $0 \in D_{f,z_0}$, we can compute the location of a root by evaluating the power series expansion of $f_{z_0}^{-1}$. However, this will not occur in general. The idea of Shub and Smale is to continue to invert f by a power series along the ray $(0, f(z_0)]$ (that is by analytic continuation) until one obtains a z such that $0 \in D_{f,z}$, or come as close to a root as desired for the case of multiple roots. (We note that there is always such a z_0 since there are at most (d-1) critical points of f).

More precisely, let $w_1 = (1-h)f(z_0)$ for $h < \dfrac{R_{f,z_0}}{|f(z_0)|}$. Note that $w_1 \in D_{f,z}$.

and we obtain $z_1 = f_{z_0}^{-1}(w_1) = f_{z_0}^{-1}((1-h)f(z_0))$ such that $1-h = \dfrac{f(z_1)}{f(z_0)}$.

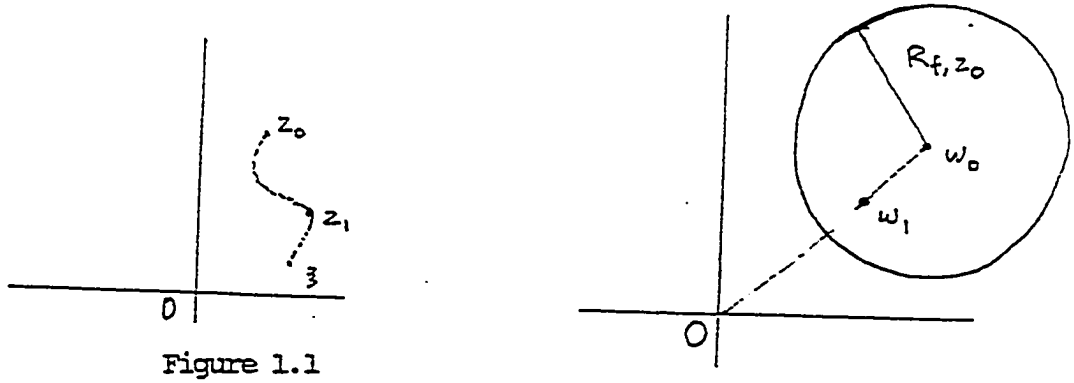


Figure 1.1

We continue this process (z_1, w_1) replacing the pair (z_0, w_0). To understand this process in a more consistent way, we consider a wedge shaped domain. Let W_{f,z_0} be the maximal open wedge domain centered at $f(z_0)$ with anlge $A_{f,z_0} \leq \pi/2$, where $f_{z_0}^{-1}$ is analytic.

$W_{f,z} \cong W_{f,z,A}$, where $A = \sup\{\alpha \leq \pi/2 : f_z^{-1} \text{ is analytic on } W_{f,z,\alpha}\}$. $\cong$

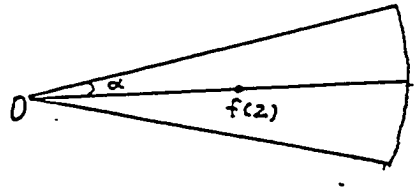


figure 1.2 wedge $W_{f,z,\alpha} = \{w : |w| < 2|f(z)|, |\arg w/f(z)| < \alpha\}$

Let $H_{f,z} = \dfrac{R_{f,z}}{|f(z)|}$.

Lemma 1.1 $H_{f,z} > \sin A_{f,z}$, for any $z \in f_{z_0}^{-1}[(0,f(z_0)]$.

Proof. Since f_z^{-1} is analytic on $W_{f,z}$, from Trigonometry,

$$R_{f,z} \geq |f(z)| \sin A_{f,z} \geq |f(z)| \sin A_{f,z_0}.$$

Let $h < \sin A_{f,z_0}$. Define inductively, $z_{n+1} = f_{z_n}^{-1}((1-h)f(z_n))$.

Corollary 1.2. $(1-h)f(z_n) \in D_{f,z_n}$.

Proof. This is immediate since the wedge contains the circle of radius $\sin A_{f,z_n} |f(z_n)|$ centered at $f(z_n)$ and $\sin A_{f,z_n} \geq \sin A_{f,z_0}$

Theorem 1.3 For any $\varepsilon > 0$, let $N < \dfrac{\log[|f(z_0)|/\varepsilon]}{h}$, then

$$|f(z_N)| < \varepsilon .$$

Proof. Since $|f(z_{n+1})| = (1-h)|f(z_n)|$ we have $|f(z_n)| = (1-h)^n f(z_0)|$.

Since $(1-h)^{1/h} \uparrow e^{-1}$, as $h \downarrow 0$, we have

$$(1-h)^N < (1-h)^{1/h(\log|f(z_0)|+|\log \varepsilon|)} < e^{-\log|f(z_0)+\log\varepsilon} = \frac{\varepsilon}{|f(z_0)|} .$$ QED.

Clearly, the step size N depends on the initial point z_0: in particular, on its angle size A_{f,z_0}, and the size of $|f(z_0)|$.

Thus it is reasonable to analyze the average cost , in terms of probability of getting an initial point of a certain property,eg. a point with a good angle size $A_{f,z}$.(cf,SSII, Section 1).

Shub and Smale introduced the following ideas from probability theory.

Let $P_d(1) = \{ \sum_{i=0}^{d} a_i z^i : a_d = 1, |a_i| \leq 1 \}$.

Let S_3^1 be the circle of radius of 3 and endow it with uniform probability measure. Let Ω_3 be the countable product of S_3^1 with itself and impose the product measure on it. For a fixed polynomial f, define $m_f((z_i)_1^{\infty}) = m$ if m is the smallest number such that $A_{f,z_m} > \pi/12$.

Theorem 1.4 (Shub&Smale) The following is true for all f $P_d(1)$.

i) Probabilistic estimates : $\int_{\Omega_3} m_f \leq 6$.

In other words, on the average, the sixth try of the points on S_3^1 will give $A_{f,z} > \pi/12$, where $S_3^1 = \{ z : |z| = 3 \}$.

ii) Deterministic estimates. At least 2d out of 24d points spread evenly around S_3^1 have $A_{f,z} > \pi/12$.

We now introduce an implementing the discussion above and estimate its cost.

Hereafter, we will call z an initial point good if $A_{f,z} > \pi/12$.

Theorem.1.5 Given > 0, the following algorithm IDEAL produces $z_N \in C$ such that $|f(z_N)| < \epsilon$ with the average cost over Ω_3,

less than $24((d+1)\log 3 + |\log \epsilon|)$

Proof.We note that we can use h = 1/4 since $\sin \pi/12 > 1/4$.

Hence if the chosen point z_0 has angle $> \pi/12$ then by Theorem 1.3, we find such z_N with $N = 4((d+1)\log 3 + |\log \delta|)$ and $|f(z_N)| < \epsilon$, since $|f(z_0)| < 3^{d+1}$.

Since this process terminates at sith times on the average from theorem 1.4,

the proof is complete. QED.

IDEAL

1 CHOOSE $|z_0| = 3$. Set $N = 4(d + |\log \xi|)$.

2 DO n=1 to N

$$z_{n+1} = f_{z_n}^{-1} \ (3/4f(z_n))$$

3 If $|f(z_N)| > \xi$, THEN GO TO 1.

4 TERMINATE WITH AN OUTPUT z_N.

In practice, IDEAL has the following difficulties;

Q 1. f_z^{-1} is given as an infinite power series and this is computationally infeasible. One must approximate f_z^{-1}. How does one approximate f_z^{-1} ?

Q 2. What about the round-off error ?

Q 3. What should be the criteria that z is a root ?

Q 1. is answered in SS I and II. They define a k-th order Euler's iterational scheme and estimate the resulting cost.

Definition.1.1 A k-th order Euler's iteration is the map parametrized by a complex number h and an integer k, $E_{k,h}$: C ----> C

where $E_{k,h}(z) = T_k f_z^{-1}((1-h)f(z))$, where $T_k \sum_{i=0}^{k} a_i z^i = \sum_{i=0}^{k} a_i z^i$.

Theorem (Shub&Smale) There is an algorithm based on the Euler's iteration which finds a z such that $|f(z)| < \varepsilon$ with the average cost over Ω_3,

$$O\,(\,d + |\log \varepsilon|\,)$$ k-th order Euler's iteration,

where $k = \text{Max}\,(\log d,\ \log|\log \varepsilon|\,)$.

A more detailed discussion and some improved results will be discussed in chapter 2.

Q 2. will be discussed in chapter 3. In fact we show that Newton's method applied to a translation of a given polynomial is cheap and stable.

Now, we consider Q 3.

Definition 1.3 (Shub& Smale) z is called an approximate zero for E_k

if $\dfrac{|f(z_n)|}{|f(z_0)|} < b^{(k+1)^n}$, where $z_n = E^n_{k,1}(z)$, $b < 1$.

Let $\rho_{f,z}$ = radius of convergence of f_z^{-1} at the origin.

We note that $\rho_f = \min_z \rho_{f,z} \leq \rho_{f,z}$. Also we note that

$$0 \in D_{f,z} \text{ if } |f(z)| < 1/2\,\rho_{f,z}.$$

In fact one can show that

Lemma Suppose $|f(z)| < \frac{1}{7}\rho_{f,z}$ Then z is an approximate zero of f for all k.

Proof See chapter 2 and Proposition 3.3.

Using the following probability estimates on $P_d(1)$, Shub & Smale obtained an average cost estimate of an algorithm based on the Euler's iteration to find an approximate zero of f.

We impose a normalized Lebesgue measure on $P_d(1) \subset C^d$.

Lemma (Shub & Smale)

i) $\int_{P_d(1)} |\log \rho_f| < 1/2 \log d + 1$

ii) Let $\varepsilon_f = \frac{Df}{(2d)^{4d}}$, where Df is the resultant of f and f'. (cf. Lang).

Then $\varepsilon_f < \rho_f$

and $\int_{P_d(1)} |\log \varepsilon_f| = O(d \log d)$

Proof. See SSI chapter 3.

Their main result is then :

Theorem There is an algorithm based on the Euler iteration to find an approximate zero of f with average cost less than

i) $O(d \log d)$ k-th order Euler's iterations, where

$k = \text{Max}(\log d, \log|\log \varepsilon|)$.

ii) $O((d \log d)^2)$ arithmetic operations.

Proof) See SSI chapter 3.

Chapter 2. Euler Iteration and the Distortion Lemma.

Section 1.

The main purpose of this section is to prove the three Theorems below. We remark that the basic result of theorem 2.A is done in SS1, but here we improve the result and present a different proof.

$$\text{Let } B_k(r) = (k+1)\frac{(1+r)^3 r^k}{(1-r)^5}.$$

Let r_k and $\tilde{r}_k$ be the smallest solutions respectively to

$$B_k(r) \text{ Max } (1, r/1-r) = 1$$

and $$B_k(r) \text{ Max } (1, r/1-r) = 1/2.$$

Note that $r_k \uparrow \tilde{r}_k$ |1 as $k \uparrow \infty$. A table of values of r_k and $\tilde{r}_k$ are presented below.

Recall $H_{f,z} = \frac{R_{f,z}}{|f(z)|}$. See Chapter 1.

Theorem 2.A. For any $r < r_k$ and $|h| = r H_{f,z}$, let $z' = E_{k,h}(z)$. Then we have

$$\frac{f(z')}{f(z)} = 1 - h(1 - \epsilon), \text{ where } |\epsilon| < B_k(r).$$

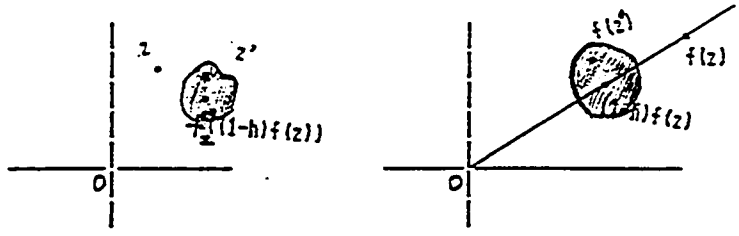


Figure 2.1

k	1	2	5	10	20	40	100	200	1000
r_k	.148	.225	.366	.495	.624	.737	.852	.908	.973
$\tilde{r}_k$	.106	.188	.338	.474	.601	.720	.836	.898	.969

Table of r_k and $\tilde{r}_k$

Remark. We remark that r_k is better than γ_k in SSI , where $\gamma_k \uparrow .178$

Theorem 2.B. For any $r < \tilde{r}_k$ and $|h| = r H_{f,z}$, let $z' = E_{k,h}(z)$.

Suppose $|\tilde{z} - z'| < |F(z)h| \frac{(k+1)r^k}{(1-r)^2}$, $F(z) = -f(z)/f'(z)$.

Then we have

$$\frac{\tilde{f}(z)}{f(z)} = 1 - h(1 - \epsilon), \text{ where } |\epsilon| < 2B_k(r).$$

More generally, we have ;

Let $z' = f_z^{-1}((1-h)f(z))$ for $|h| = rH_{f,z}$, $r < 1$.

Theorem 2.1. Suppose $|\tilde{z} - z'| < |F(z)|M$ for $M < H_{f,z} \frac{(k+1)r^k}{(1+r)^3} \text{Min}[r, 1-r]$.

Then $\tilde{z} = f^{-1}((1-\tilde{h})f(z))$, for some $\tilde{h}$ such that $|\tilde{h} - h| < M \frac{(1-r)^3}{(1+r)^3}$ |||

Now we begin the details.

We recall the definition of $E_{k,h}(z)$ from chapter 1.

Definition. $E_{k,h}(z) = T_k f_z^{-1}[(1-h)f(z)]$.

Here T_k is a k-th order truncation of a power series.

I.e. $T_k \sum_{i=0}^{\infty} a_i z^i = \sum_{i=0}^{k} a_i z^i$.

Following Shub and Smale, it is advantageous to introduce an auxillary function for a given polynomial f of degree d and a given point z. We define, $\sigma(w) = w + \sigma_2 w^2 + \ldots + \sigma_d w^d$.

where $\sigma_i = \left(\frac{-f(z)}{f'(z)}\right)^{i-1} \frac{f^{(i)}(z)}{i!f'(z)}$

The following facts are proven in SS1, .

Fact 2 . i) Let $R_{f,z}$ be the radius of convergence of f_z^{-1} at f(z), then σ^{-1} has a radius of convergence $H_{f,z} = \frac{R_{f,z}}{|f(z)|}$ about the origin.

ii) σ^{-1} is univalent. I.e. σ^{-1} is one to one and analytic on $D_{H_{f,z}}(0)$.

iii) $\sigma^{-1}(0) = 0$, $\sigma^{-1\prime}(0) = 1$.

iv) $f_z^{-1}((1-h)f(z)) = z + F(z)\sigma^{-1}(h)$, where $F(z) = -f(z)/f'(z)$.

and $T_k f_z^{-1}((1-h)f(z)) = z + F(z)T_k\sigma^{-1}(h)$.

Now we have a following definition equivalent to Definition 1.1 .

Definition 2.1 (Shub&Smale).

$$E_{k,h}(z) = z + F(z)\, T_k\sigma^{-1}(h).$$

Then Theorem 2.1 can be reduced as follows.

Let $z' = E_{k,h}(z) = T_k f_z^{-1}((1-h)f(z))$.

We will show that $z' = f_z^{-1}(w)$, for some $w = (1-h')f(z)$

such that $|h' - h| < hB_k(r)$.

By the Fact 2 iv) it is reduced to find h' such that $T_k\bar{\sigma}^{-1}(h) = \bar{\sigma}^{-1}(h')$.

Recall that $\bar{\sigma}^{-1}$ is univalent and $\bar{\sigma}^{-1}(0) = 0$, $\bar{\sigma}^{-1\prime}(0) = 1$.

The main tools used in this chapter are the Bieberbach conjecture and the Koebe distortion theorem for a schlicht function. f is called schlicht if it is one to one analytic on $D_1(0)$ and $f(0)=0$, $f'(0)=1$.

Theorem 2.3. (Biberbach Theorem) Let $g(z) = z + g_2z^2 + g_3z^3 + \ldots$ be schlicht. Then $|g_k| \leq k$.

Proof. See Acta Mathematica 1985.

The following Lemma is the Corollary to Theorem 3.

Lemma 2.4. (Shub&Smale) Let g be schlicht. Then

$$|g(h) - T_k g(h)| \leq \frac{(k+1)r^{k+1}}{(1-r)^2}, \quad |h| = r<1.$$

Proof. From Theorem 1.3, we have

$$|g(h) - T_k g(h)| < \sum_{i=k+1} |ir^i| \leq r\left(\frac{r^{k+1}}{1-r}\right)' < \frac{(k+1)r^{k+1}}{(1-r)^2}. \quad \text{Q.E.D.}$$

Theorem 2.5. (Distortion Theorem) Let g be schlicht. Then for $|h|=r$,

$$\frac{r}{(1+r)^2} \le |g(h)| \le \frac{r}{(1-r)^2}$$

$$\frac{1-r}{(1+r)^3} \le |g'(h)| \le \frac{1+r}{(1-r)^3}$$

Proof. See Hille[5].

This gives the following corollary.

Lemma 2.6. Let g be univalent on $D_R(z)$. Then for $|h| = r < R$

$$\frac{|g'(z)h|}{(1+a)^2} \le |\, g(z+h) - g(z)\,| \le \frac{|g'(z)h|}{(1-a)^2}$$

where $a = r/R$.

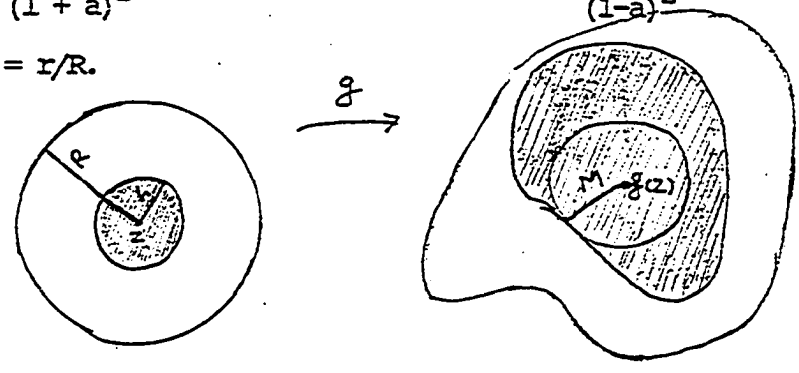


Figure 2.2 ; $M = |g'(z)h|/(1+a)^2$

Proof. Let $f(h) = \frac{1}{g'(z)} \frac{1}{R} [\, g(z+Rh) - g(z)\,]$

Then it is easy to check f is schlicht.

Hence by theorem 2.4, for any $|h|<R$, $a = |h/R|$, we have

$$\frac{|g'(z)h|}{(1+a)^2} \le |\, f(h/R)\,| \le \frac{|g'(z)h|}{(1-a)^2}.$$

Since $g(z + h) - g(z) = Rg'(z)f(h/R)$, we have

$$\frac{|g'(z)R|a}{(1 + a)^2} \leqq | g(z+h) - g(z) | \leqq \frac{|g'(z)R|a}{(1-a)^2}$$

Since $|h|=Ra$ we have the claim. Q.E.D.

Proposition 2.7. Let g be schlicht. Then for a given point $z \in D_1(0)$, and $r \leqq 1 - |z|$, we have $g(D_r(z)) \supset D_M(g(z))$,

where $M = \frac{r}{(1-|z|+r)^2} \frac{(1-|z|)^3}{(1+|z|)^3}$.

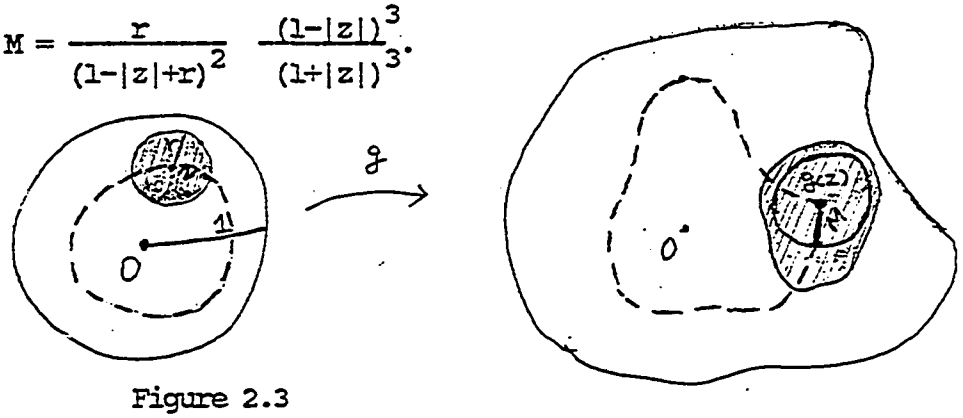


Figure 2.3

Proof. By Lemma 2.6 we have $g(D_r(z)) \supset D_M(g(z))$, where

$M = \frac{|g'(z)|r}{(1 + \frac{r}{1-|z|})^2}$. Since $|g'(z)| \geqq \frac{(1-|z|)}{(1+|z|)^3}$ by Theorem 2.5 we have

$$M \geqq \frac{(1-|z|)}{(1+|z|)^3} \frac{r}{(1 + \frac{r}{1-|z|})^2}$$

$$= \frac{(1-|z|)^3}{(1+|z|)^3} \frac{r}{(1-|z|-r)^2}$$ Q.E.D.

From the Proposition 2.7, we obtain the Quarter Theorem at an arbitrary point.

Corollary 2.8. Suppose $M \leq \frac{1}{4} \frac{(1-|z|)^2}{(1+|z|)^3}$

Then $D_M(g(z)) \subset g(D_{1-|z|}(z))$.

Proof. Apply Proposition 2.7 with $r = 1-|z|$. Q.E.D.

In the work to follow, we are particularly interested in the case of $r < |z|$.

Corollary 2.9. Supopose that $r \leq \text{Min}(|z|, 1-|z|)$. Then

$$g(D_r(z)) \supset D_M(g(z)), \text{ where } M = r \frac{(1-|z|)^3}{(1+|z|)^3}$$

Proof. Immediate from Proposition 2.7 since

$$\frac{(1-|z|)^3}{(1+|z|)^3} \frac{r}{(1-|z|+r)^2} > r \frac{(1-|z|)^3}{(1+|z|)^3} \qquad \text{Q.E.D.}$$

Corollary 2.10. Let g be univalent on $D_1(0)$ and $g(0) = 0$, $g'(0)=1$. Then

for $r < |z|$, $g(D_r(0)) \supset D_M(g(z))$, where $M = r \frac{(1-a)^3}{(1+a)^3}$, $a = |z|/H$.

Proof. $f(z) = 1/H\, g(Hz)$ is schlicht. Apply Corollary 2.9 to f.

The Corollary 2.10 is stated as the following in terms of σ^{-1}.

Recall $H_{f,z}$ denote the radius of convergence of σ^{-1} at 0.

For any $\epsilon < H_{f,z}\text{Min}(r,1-r)$, we have

$$\sigma^{-1}(D_\epsilon(h)) \supset D_M(\sigma^{-1}(h)) \text{ , where } M = \epsilon \frac{(1-r)^3}{(1+r)^3} \text{ , } r = |z|/H_{f,z}$$

In other words, if let $M < \frac{(1-r)^3}{(1+r)^3}$ Min $(r,1-r)$, then

if $|w - \bar{\sigma}^{-1}(h)| < M$ then $w = \bar{\sigma}^{-1}(h')$ for $|h'-h| < \epsilon = M \frac{(1+r)^3}{(1-r)^3}$

Proof of Theorem 2.1. From the Fact 2, we have for any $|h| < H_{f,z}$,

$$f_z^{-1}((1-h)f(z)) = z + F(z)\,\bar{\sigma}^{-1}(h).$$

Fix h and let $z' = f^{-1}((1-h)f(z)) = z + F(z)\bar{\sigma}^{-1}(h)$.

By Remark 2.11 we have

if $\tilde{z} = z + F(z)(\bar{\sigma}^{-1}(h) + \delta)$ for some $|\delta| < M$, then

$\tilde{z} = z + F(z)\,\bar{\sigma}^{-1}(\tilde{h}) = f_z^{-1}((1-\tilde{h})f(z))$ for some $\tilde{h}$ such that

$$|\tilde{h} - h| < M \frac{(1+r)^3}{(1-r)} \quad . \qquad \text{Q.E.D.}$$

Proof of Theorem 2.A and 2.B.

Since $\frac{1}{H_{f,z}} \bar{\sigma}^{-1}(H_{f,z}h)$ is schlicht, from Lemma 2.4 one can easily show

that $|\bar{\sigma}^{-1}(h) - T_k \bar{\sigma}^{-1}(h)| \le H_{f,z}(k+1) \frac{r^{k+1}}{(1-r)^2} = h(k+1) \frac{r^k}{(1-r)^2} \equiv M.$

Since $B_k(r) = \frac{(k+1)r^k}{(1+r)^5}(1+r)^3 \le \text{Min}[1,1-r/r]$

and $M = hB_k(r) \frac{(1+r)^3}{(1-r)^3}$,

we note $M \le h \frac{(1+r)^3}{(1-r)^3}$ Min$[1,1-r/r]$, for $r < r_k$.

Now by Theorem 2.1. we have that

$$| \tilde{h} - h | < M \frac{(1+r)^3}{(1-r)^3} = h B_k(r).$$

Similarly, one may prove Theorem 2.B. Q.E.D.

Section 2.

Let $B_k(r) = (k+1)\dfrac{(1+r)^3 r^k}{(1-r)^5}$ as before.

Let $\mu_1 = 0.0525$, $\tilde{\mu}_1 = 0.0333$ $\mu_k, \tilde{\mu}_k$, $k>1$ be the solutions to

$$B_k\left(\frac{k+2}{k} r\right) = \frac{1-r}{k+2}, \qquad (1)$$

$$B_k\left(\frac{k+2}{k} r\right) = \frac{1-r}{2(k+2)} . \qquad (2)$$

Recall that $A_{f,z} \leq \pi / 2$ denotes the angle of the largest wedge centered at $f(z)$ where f_z^{-1} is analytic.

Theorem 2.12 For a polynomial f and a complex number z_0, suppose that $A = A_{f,z_0} > 0$. Then for $h = r \sin A$, $r < \mu_k$, set $w_n = (1-h)^n f(z_0)$, $n=0,1,\ldots$

Let $z_{n+1} = T_k f_{z_n}^{-1}(w_{n+1})$. Then we have

i) $\dfrac{f(z_n)}{w_n} = 1 + h\delta_n$, $|\delta_n| < \dfrac{1}{k+1}$

ii) For any $\varepsilon > 0$, let $N = \dfrac{1}{h}(\log|f(z_0)| + |\log\varepsilon|)$.

Then $|f(z_n)| < \varepsilon$ for all $n > N$.

Proof will be given later.

Theorem 2.13 For a polynomial f and a complex number z_0, suppose that $A = A_{f,z_0} > 0$. Then for $h = r \sin A$ and $r < \tilde{\mu}_k$.

Suppose $\tilde{z}_0 = f_{z_0}^{-1}((1-\delta_0)f(z_0))$ for $|\delta_0| < \frac{h}{k+1}$. Let $w_n = (1-h)^n f(\tilde{z}_0)$, $n = 0,,,.$

and $\tilde{z}_{n+1}$ satisfy $|\tilde{z}_{n+1} - z_{n+1}| < |F(\tilde{z}_n)| \, h\frac{(k+1)r^k}{(1-r)^2}$,

where $z_{n+1} = T_k f_{\tilde{z}_n}^{-1}(w_n)$, $F = f(z)/f'(z)$. Then we have

i) $\frac{f(\tilde{z}_n)}{w_n} = 1 + h\delta_n$, $|\delta_n| < \frac{1}{k+1}$

ii) For any $\varepsilon > 0$, let $N = 1/h \, (\log|f(z_0)| + |\log\varepsilon|)$.

Then $|f(\tilde{z}_n)| < \varepsilon$ for all $n > N$.

Proof will be given later.

k	1	2	5	10	20	40	80	100
μ_k	.054	.110	.242	.370	.521	.666	.779	.809
$\tilde{\mu}_k$	.357	.089	.221	.360	.512	.658	.774	.805

Table of μ_k and $\tilde{\mu}_k$

The folowing Lemma is useful to have.

Lemma 2.14 i) Let f be a polynomial and let $g(z) = f(z) - c$ for some complex number c. Let R_{f,z_0}, R_{g,z_0} denote the radius of convergence of

$f_{z_0}^{-1}$ and $g_{z_0}^{-1}$ at $f(z_0)$ and $g(z_0)$ respectively. Then

i) $R_{f,z_0} = R_{g,z_0}$

ii) Further if $| w - f(z_0) | < R_{f,z_0}$ then $f^{-1}_{z_0}(w) = g^{-1}_{z_0}(w-c)$, and

$$T_k f^{-1}_{z_0}(w) = T_k g^{-1}_{z_0}(w-c)$$

Proof. This is immediate by noting that

if $z = f^{-1}_{z_0}(w)$ then $z = g^{-1}_{z_0}(w-c)$. Q.E.D.

Remark 2.15. Let $g_n(z) = f(z) - w_n$, where w_n as in Theorem 2.12 and 2.13.

Then $$T_k f^{-1}_{z_n}(w_{n+1}) = T_k g^{-1}_{n+1} \quad (0)$$

and hence $$E_{k,h_n,f}(z_n) = E_{k,1,g_{n+1}}(z_n)$$

Note that computation of $E_{k,1}$ is simpler than $E_{k,h}$. We will use $E_{k,1,g_n}$ when evaulating z_n. For example, when using k=1, we have

$$z_{n+1} = z_n - \frac{f(z_n) - w_{n+1}}{f'(z_n)} .$$

Proof of Theorem 2.12 and Theorem 2.13.

We will prove the theorems by an induction on n.

The case n=0 is trivial because $\frac{f(z_0)}{w_0} = 1$.

We recall that $z_{n+1} = T_k f^{-1}_{z_n}(w_{n+1}) = E_{k,h_n,f}(z_n)$

where $h_n = \frac{f(z_n) - w_{n+1}}{f(z_n)}$. (3)

Now suppose that $\frac{f(z_n)}{w_n} = 1 + h\delta_n, \quad |\delta_n| < \frac{1}{k+1}$.

Then from the trigonometry and by the triangle inequality,

$$(4) \quad |f(z_n) - w_{n+1}| \leq |f(z_n) - w_n| + |w_n - w_{n+1}|$$

$$\leq |w_n|h + |w_n| \frac{h}{k+1} \quad \text{by an induction hypothesis.}$$

$$\leq \frac{k+2}{k+1} |w_n| h = \frac{k+2}{k+1} r |w_n| \sin A.$$

$$(5) \quad R_{f,z_n} \geq |w_n| \sin A - |f(z_n) - w_n|$$

$$\geq |w_n| \sin A - |w_n| \frac{h}{k+1}$$

$$\geq |w_n| \sin A \left(1 - \frac{r}{k+1}\right), \text{ since } h = r \sin A.$$

Hence $r_n = \left| \frac{h_n}{H_{f,z_n}} \right| = \frac{|f(z_n) - w_{n+1}|}{R_{f,z_n}} < \frac{\frac{k+2}{k+1} r}{1 - \frac{r}{k+1}}$

$$= \frac{(k+2)r}{k+1-r} < \frac{(k+2)r}{k} \qquad (6)$$

Note that $r_n < r_k$, since $\frac{k+2}{k} r < r_k$ by equation (1).

Now by Theorem 2.A, we have

$$\frac{f(z_{n+1})}{f(z_n)} = 1 - h_n (1+\delta), \quad |\delta| < B(r_n), \qquad (7)$$

where $z_{n+1} = E_{k,h_n,f}(z_n)$, $h_n = \frac{f(z_n) - w_{n+1}}{f'(z_n)}$.

Hence $f(z_{n+1}) = f(z_n)\ (1-h_n) + f(z_n)\ h_n \delta$

$= w_{n+1} + f(z_n) h_n \delta$

$= w_{n+1} + (f(z_n) - w_{n+1})\delta$ by (3).

Thus $\dfrac{f(z_{n+1})}{w_{n+1}} = 1 + \dfrac{f(z_n) - w_{n+1}}{w_{n+1}} \delta$, $|\delta| < B_k(r_n)$.

Now $\left| \dfrac{f(z_n) - w_{n+1}}{w_{n+1}} \delta \right| < \left| \dfrac{k+2}{k+1} \dfrac{w_n h}{w_{n+1}} \delta \right|$ by (4)

$= \left| \dfrac{k+2}{k+1} h \dfrac{\delta}{1-h} \right| < h \dfrac{k+2}{k+1} \left| \dfrac{B_k(r_n)}{1-h} \right|$ by (7)

$< h \dfrac{k+2}{k+1} \dfrac{B_k(r_n)}{1-r}$

$< h \dfrac{k+2}{k+1} \dfrac{B_k(\frac{k+2}{k} r)}{1 - r}$ by (6)

$< h \dfrac{1}{k+1}$ by (1) a definiton of k.

Thus we established $\dfrac{f(z_{n+1})}{w_{n+1}} = 1 + h\delta_{n+1}$, $|\delta_{n+1}| < \dfrac{1}{k+1}$.

This proves Theorem i) of Theorem 2.12.

One proves i) of Theorem 2.13 in a similar way.

For ii) note that $(1 - h)^{1/h} \uparrow e^{-1}$ as $h \downarrow 0$. Thus for any $\varepsilon > 0$,

$$(1 - h)^{1/h \log[|f(z_0)|/\varepsilon]} < e^{-\log|f(z_0)| + \log\varepsilon} = \frac{\varepsilon}{|f(z_0)|}$$

Let $N = \frac{1}{h}(\log|f(z_0)|/\varepsilon)$. Then for all $n > N$,

$$|w_n| = (1-h)^n |f(z_0)| < (1-h)^N |f(z_0)| = \frac{\varepsilon}{|f(z_0)|} |f(z_0)| = \varepsilon$$

In other words for all $n > N$, we have $|f(z_n)| < \varepsilon$.

Similarly one proves ii) of Theorem 2.13.

We remark that for the case of $k = 1$, we use $\frac{(k+2)r}{k+1-r}$ rather than $\frac{k+2}{k}r$

in (3) since the loss is significant. i.e. $\mu_1 = .0525$, $\tilde{\mu}_1 = .034$ are obtained from

$$\frac{B_1(\frac{3r}{2-r})}{1-r} = \frac{1}{3}, \quad \frac{B_1(\frac{3r}{2-r})}{1-r} = \frac{1}{6}, \text{ respectively.}$$

In this thesis, frequently we are work on the assumption that $A_{f,z_0} > \pi/12$. Let us call z_0 is a good initial point if $A_{f,z_0} > \pi/12$. Note that

$\sin \pi/12 \sim .258 > 1/4$.

The following are the special case of Theorem 2.12 and Theorem 2.13, when $|z_0| = 3$ is a good initial point.

Corollary 2.16. Suppose z_0, $|z_0| = 3$ is a good initial point. Let $f \in P_d(1)$. Let $h = \frac{\mu_k}{4}$. Define z_n as in Theorem 2.12. Then for any $1 > \varepsilon > 0$, $|f(z_n)| < \varepsilon$, for all $n > N$, where $N = \frac{1}{h}((d+1)\log 3 + |\log\varepsilon|)$.

Corollary 2.17. Suppose z_0 is a good initial point. Let $h = \frac{\tilde{M}_k}{4}$.

Define $\tilde{z}_n$ as in Theorem 2.13. Then for any $1 > \xi > 0$,

$$|f(\tilde{z}_n)| < \xi \text{ , for all } n > \frac{1}{h}\left((d+1)\log 3 + |\log \xi|\right).$$

Proof. Just note that $|f(z_0)| \leq |z_0|^d + \ldots + |a_i z^i| + \ldots + |a_0|$

$$\leq \sum |z_0|^i \text{ since } |a_i| \leq 1.$$

$$\leq 1/2\ 3^{d+1} \quad \text{on } |z| = 3$$

Hence $\log|f(z_0)| \leq (d+1)\log 3$.

Thus it is immediate from Theorem 2.12. Q.E.D.

Remark 2.18. Special case of Corollary 2.17 with k=1.

Since $1/32 < \tilde{M}_k \approx 0.033$, we use $h = \frac{1}{4}\ \frac{1}{32} = \frac{1}{128}$.

Suppose $\tilde{z}_0 = f_{z_0}^{-1}((1-\delta)f(z_0))$, for $|\delta| < \frac{h}{2}$.

Let $z_{n+1} = \tilde{z}_n - \frac{f(\tilde{z}_n) - w_{n+1}}{f'(\tilde{z}_n)}$, where $w_{n+1} = \frac{127}{128}^{n+1} f(\tilde{z}_0)$,

and $|\tilde{z}_n - z_n| < \frac{|f(\tilde{z}_n)|}{|f'(z)|}\ \frac{1}{128}\ \frac{2\ 1/32}{(1-1/32)^2} = \frac{1}{1922}\ \frac{|f(\tilde{z}_n)|}{|f'(z_n)|}$,

Then for any $1 > \xi > 0$, $|f(\tilde{z}_n)| < \xi$, for all $n > N$,

where $N = 128\ ((d+1)\log 3 + |\log \xi|)$.

We recall the algorithm IDEAL in Chapter 1. The following algorithm which is based on Theorem 2.12 generalizes IDEAL.

ROOT (k,ε)

0. Set $h = \mu_k / 4$.

1 Choose $|z_0| = 3$. Set $w_0 = f(z_0)$.

2. Do n = 1 to $N = 1/h[(d+1)\log 3 + |\log \varepsilon|]$

$$w_{n+1} = (1-h)w_n \ , \quad z_{n+1} = E_{k,1,f-w_{n+1}}(z_n)$$

3. If $|f(z_N)| 7\varepsilon$ then GO TO 1.

4. If $f(z_N) \lesssim \varepsilon$ then Terminate with z_N.

We note that ROOT (∞,ε) is IDEAL.

One can design a similar algorithm based on Theorem 2.14. In particular, when using k=1 , we have the following result.

Theorem 2.19 There is an algorithm based on Euler algorithm which find such that $|f(\xi)| < \varepsilon$,for any $\varepsilon > 0$, provided for all the iterates z , f(z) and f'(z) are computed with a relative error less than 1/4000.

Proof. Immediate from Remark 2.18. cf proof of Theorem 4.3 in page 43.

Chapter 3. FAST-ROOT

The goal of this chapter is to construct an efficient algorithm and analyze its complexity. This algorithm is closely related to ROOT in Chapter 2 and called FAST-ROOT. Its advantage is that it proceeds fast once an iterate becomes an approximate zero.

We state the main complexity results.

We impose a normalized product measure on $\Omega_3 x P_d(1) \subset \Omega_3 x\, C^d$

Theorem 3.1 FAST-ROOT finds an approximate zero for any $f \in P_d(1)$ without multiple roots with average cost over $\Omega_3 \times P_d(1)$ less than

i) $C_1 d$ Euler Iterations,

ii) $C_2 d^2$ Arithmetic operations.

Theorem 3.2 FAST-ROOT finds z such that $|f(z)| < \zeta$ for any $f \in P_d(1)$ with the average cost over $\Omega_3 x\, P_d(1)$ less than

i) $C_1 (d + \log |\log \epsilon|)$ Euler iterations,

ii) $C_2(d^2 + d \log|\log\epsilon|)$ arithmetic operations.

Here $C_1 \approx 528$, $C_2 \approx 1056$

FAST-ROOT mainly has the three parts.

1. Choose a starting point.
2. Evaluate an Euler Iteration , $E_{k,h}$
3. Test Approximate zero.

Once the algorithm reaches an approximate zero , we will use h = 1. i.e. $z_{n+1} = E_{k,1}(z_n)$. Hence it is important to determine if a given iterate is an approximate zero. We refer the reader to the Appendix 1 that the test below , TEST-APPROXIMATE, suffices to determine an approximate zero.

TEST- APPROXIMATE (f,z)

$$\text{Set } T_j = \frac{f^{(j)}(z)}{j!\, f'(z)} \left| \frac{f(z)}{f'(z)} \right|^{j-1}$$

$$\text{Is } |T_j| < \left(\frac{r_k}{6}\right)^{j-1} \quad \text{for } j = 2, \ldots, d \ ?$$

Here r_k is as in Chapter 2, and $(r_k/6) \uparrow 1/6$ as $k \uparrow \infty$.

Proposition 3.3 . If TEST-APPROXIMATE (f,z) = YES then z is an approximate zero of f.

Proof. If TEST-APPROXIMATE = YES then $|T_j| < \left(\frac{r_k}{6}\right)^{j-1}$, for all $j= 2,..,d$.

Hence by Corollary A.2 in Appendix 1, we have

$$H_{f,z} = \frac{R_{f,z}}{|f(z)|} > \frac{6}{r_k} \frac{1}{3-2\sqrt{2}} > \frac{1}{0.972 r_k}$$

Hence we can apply Theorem 2.A to z with $h=1$.

Inductively define

$$z_{n+1} = E_{k,1}(z_n) \text{ . Then } \left|\frac{f(z_{n+1})}{f(z_n)}\right| < B_k\left(\frac{1}{H_{f,z_n}}\right) \approx \left(\frac{1}{H_{f,z_n}}\right)^k$$

We note that $H_{f,z_{n+1}} = \frac{R_{f,z_{n+1}}}{|f(z_{n+1})|} \approx \frac{R_{f,z_n}}{|f(z_n)|} \frac{|f(z_n)|}{|f(z_{n+1})|}$

Hence $H_{f,z_{n+1}} \approx H_{f,z_n}^{k+1}$, and we have

$$\left|\frac{f(z_{n+1})}{f(z_n)}\right| < b^{(k+1)^n}, \quad b < 0.972 . \qquad \text{Q.E.D.}$$

The following Corollary is immediate from Proposition 3.3.

Corollary 3.4. Suppose z is an approximate zero of f. Then for $\varepsilon > 0$,

$$|f(z_n)| < \varepsilon, \text{ for all } n > N ,$$

where $z_n = E_{k,1,f}^n(z)$ and $N = \log \log[|f(z)|/\varepsilon]$.

Now we describe the algorithms.

Approximate -Root.

0 Set $h = \mu_k/4$.

1 Choose $|z_0| = 3$. Set $w_0 = f(z_0)$

2 WHILE ($|f(z_n)| > 1$) DO

REDUCE (h).

If $|f(z_{n+1})| > |w_n|$ GO TO 1.

3 WHILE (TEST-APPROXIMATE = NO) DO

REDUCE (h)

IF $|f(z_{n+1})| > |w_n|$ GO TO 1

4 SUCCESS: Ho ho ho!"Found an approximare zero" !!!

5 REDUCE(h) : Set $w_{n+1}=(1-h)w_n$, $z_{n+1} = E_{k,1,f-w_{n+1}}(z_n)$.

6 TEST-APPROXIMATE : Set $T_j = \frac{f^{(j)}}{j!f'} \left|\frac{f(z)}{f'(z)}\right|^{j-1}$

IS $|T_j| < \left(\frac{r_k}{6}\right)^{j-1}$ for all j =2,..d ?

Remark. In step 2, the control is over to step 1 if $|f(z_{n+1})| \geq |w_n|$. Because this indicates the chosen initial point is a bad initial point.

FAST-ROOT (ε).

0 Set $h = M_k/4$.

1 Choose $|z_0| = 3$. Set $w_0 = f(z_0)$

2 WHILE ($|f(z_n)| > 1$) DO

REDUCE (h).

If $|f(z_{n+1}) > |w_n|$ GO TO 1.

3 WHILE (TEST-APPROXIMATE = NO) DO

REDUCE (h)

IF $|f(z_{n+1})| > |w_n|$ GO TO 1

IF $|f(z_{n+1})| < \varepsilon$ GO TO SUCCESS

4 Set h = 1.

5 WHILE ($|f(z_n)| > \varepsilon$) DO

REDUCE (h).

6 SUCCESS: Ho ho ho! Found $|f(z)| < \varepsilon$.

7 REDUCE(h) : Set $w_{n+1} = (1-h)w_n$, $z_{n+1} = E_{k,1,f-w_{n+1}}(z_n)$.

8 TEST-APPROXIMATE : Set $T_j = \dfrac{f^{(j)}(z)}{j!f'(z)} \left|\dfrac{f(z)}{f'(z)}\right|^{j-1}$

IS $|T_j| < \dfrac{r_k}{6}^{j-1}$ for all j =2,..d ?

Section of Proofs.

Let us take a look of the computational aspects of the evaluation of $E_{k,h}$; For a given polynomial f and a point z, $E_{k,h}$ is described as

$$E_{k,h} : C \times P_d(1) \longrightarrow C^{k+1} \longrightarrow C$$

$$(z, f) \longrightarrow (f(z), f'(z), \ldots, f^{(k)}(z)) \longrightarrow E_{k,h,f}(z)$$

There are two ingredients in this evaluations :

i) Computation of the first k Taylor coefficients of f at z.

ii) Evaluation of the inverse map f_z^{-1} at f(z).

Algorithms computing the inverse of the power series are called reversion algorithms. The best known algorithm is that of Brent & Kung. Their result is;

Lemma 3.5. (Brent& Kung)

The evaluation of an nth order reversion at a given point costs O(nlogn) arithmatic operations.

Proof. See [3].

Now we discuss computing the Taylor coefficints. There are many algorithms computing Taylor coefficients of a polynomial of degree d. Among the known algorithms, the Shaw-Traub algorithm has the least number of multiplication and division for the first m derivatives of the polynomial for m large.(For m=0, m=1, Horner's method and Munro's method are known to be better.)

Lemma 3.6. (Shaw Traub)

Let $f(z) = \sum_{i=0}^{d} a_i z^i$. Then the algorithm below produces

$$T_d^j = \frac{f^j(z)}{j!} z^j, \quad j = 0, 1, \ldots k, \text{ and uses } 2d \text{ multiplications}$$

and kd additions.

Algorithm. Set

$$T_i^{-1} = a_{d-i-1} z^{d-i-1}, \quad i = 0, 1, \ldots d-1$$

$$T_j^j = a_d z^d, \quad j = 0, 1, \ldots k.$$

$$T_i^j = T_{i-1}^{j-1} + T_{i-1}^j \qquad \begin{matrix} j=0,\ldots k \\ i=j+1,\ldots d. \end{matrix}$$

Proof. See [6].

First we will establish the following estimate on ROOT (ϵ).

Let $f \in P_d(1)$.

Theorem 3.7 On the average over the choices of $|z_0| = 3$, with respect to normalized Lebesgue measure on $|z_0| = 3$, ROOT (k, ϵ) terminates with less than

i) $\frac{24}{\mu_k}$ ($d\log 3 + |\log \epsilon|$) evaluations of E_k.

ii) $\frac{48}{\mu_k} d\,(d\log 3 + |\log \varepsilon| + O(k\log k))$ multiplications

$\frac{48}{\mu_k} kd\,(d\log 3 + |\log\varepsilon| + O(k\log k))$ additions,

where μ_k is as in chapter 2. ($\mu_1 \approx 0.05$ and $\mu_k \uparrow 1$ as $k \uparrow \infty$)

Proof. If z_0 is a good initial point, it takes $N = \frac{4}{\mu_k}((d+1)\log 3 + |\log\varepsilon|)$ evaluations of $E_{k,h}$ to obtain z such that $|f(z)| < \varepsilon$, by Corollary 2.16. By Theorem 1.3 , on the average one out of six choices of $|z|=3$ is a good initial point, ROOT finds z such that $|f(z)| < \varepsilon$, with an average number of iterations less than $6\frac{4}{\mu_k}((d+1)\log 3 + |\log\varepsilon|)$

$$\approx \frac{24}{\mu_k}(d\log 3 + |\log\varepsilon|)$$

Thus we have i).

For ii) note that evaluation of the first k Taylor coefficients of f costs less than 2d multiplication and dK additions, and that evaluation of reversion uses O(klogk) arithmetic operations (see Brent-Kung).QED .

We emphasize that these estimates depend only on the degree of f.

Remark. For $k=1$, ROOT(ε) terminates with less than

i) $480\,(d\log 3 + |\log \varepsilon|)$ Newton iterations.

ii) $960\,d\,(d\log 3 + |\log \varepsilon|)$ additions and multiplications.

Proof. Estimate Theorem 3.7 with $\mu_1 \leq 0.05$. Q.E.D.

Proof of theorem 3.1) First we note that the result in Theorem 3.7 is independent of f. By Proposition 3.3 and Corollary A2 in Appendix 2, Approximate-Root terminates when $\rho_{f,z}/42 < |f(z)| < \rho_{f,z}/7$.

Hence by Theorem 3.5 the number of iterations used to terminate with an approximate zero is $\frac{4}{\mu_k} \log|42\, f(z_0)/\rho_f| = \frac{4}{\mu_k}(d\log 3 + 6 + |\log \varepsilon|)$.

Now by the estimates in Theorem 1.7, we have

$$(*) \int_{\substack{P_d(1) \\ \rho_f < 1}} |\log \rho_f| < 1/2 \log d + 1.$$

Hence we have i.

Let us estimates the operation numbers used for TEST-APPROXIMATE. Note that the number of TEST-APPROXIMATE called is $|\log \rho_f/42|$, recalling that one starts the TEST from the point $|f(z)| < 1$.

Since TEST- APPROXIMATE uses 2d multiplications once the Taylor coefficients are obtained. Hence the average number of multiplications used is less than

$$\frac{48}{\mu_k} d(d\log 3 + |\log \rho_f/42| + \log \rho_f/42)$$

$$< \frac{48}{0.0525} d\,(d\log 3 + |\log \rho_f/42) < 960\, d\,(d\log 3 + |\log \varepsilon|).$$

Thus on the average, the dominant term is O(d) which gives i) and the constant are approximately, $C_1 < .528$, $C_2 < 1.056$.

Q.E.D.

Proof of Theorem 3.2

By Corollary 3.4, the number of Euler's iterations to reach z such that $|f(z)| < \varepsilon$ after the algorithm reaches an approximate zero is

$$< \log | \log \rho_f / \varepsilon | = \log[\log \rho_f - \log \varepsilon] \quad \text{Q.E.D.}$$

Together with Theorem 3.1 we are done. Q.E.D.

Chapter 4. Bit Complexity of PRECISION-ROOT

In chapter 3, we studied the algorithm ROOT ;which presuppose the use of a "real number" model machine. As we discussed in chapter 0, there is no guarantee that ROOT (FAST-ROOT) will succeed on a finite machine model ; (We recall Q1 and Q2 in chapter 0.)

In this chapter, we design an algorithm (called PRECISION-ROOT) which effectively models ROOT (FAST-ROOT) on a finite machine model and analyze its complexity.

Precision -ROOT is designed with variable precision. In other words, each stage of an algorithm uses a different precision depending on the sensitivity of error at that stage. (See the flow diagram on page 49). Those who prefer a fixed precision, may take the maximum precision of all the intermediate precisions. The final result obtained here is for a fixed precision and hence we expect the complexity for a variable precision to be better.

The complexity result will depend on the machine model as well as the precision used.

Let M(n) denote the number of bit operations to multiply two n-bit numbers.

Lemma 1. $M(n) = \frac{3}{2}n^2$ with hand calculation.

Lemma 2. $M(n) = O(n \log n)$ on a Turing machine when the Fast Fourier Transform is used.

Proof. See Knuth[6].

Lemma 3. $M(n) = O(n)$ on a pointer type machine . .

Proof. See[?].

The main results of this chapter are the following.

Theorem 4.1 There is an algorithm which produces z such that $|f(z)| < \epsilon$ with the average number of bit operations over $\Omega_3 \times P_d(1)$ less than

$$C_3[d^2 + d\log|\log\epsilon|]\, M[26(d+|\log\epsilon|)]. \qquad :::$$

Theorem 4.2 There is an algorithm which produces an approximate zero with an average number of bit operations less than

$$C_3[d^2 M(26d)]. \qquad :::$$

,where $C_3 \approx 6672$

Here,the average is taken over $\Omega_3 \times P_d(1) \cong \Omega_3 \times C^d$ with a normalized Lebgesgue measure, and the algorithm we refer to is Precision-ROOT.

Analysis shows that taking k>1 in the algorithm E_k improves the bit complexity at each stage at most by a factor of 30,provided when $|f(z)|<|f'(z)|$ where $z = z_n = E_k(z_{n-1})$ and worsens the complexity if $|f(z)|>|f'(z)|$.(See section 2). In this chapter we will emphasize the case of k=1, and obtain an upper bound of the bit complexity of Euler type algorithms.

We refer the reader to read appendix 2 for the arithmetic details , especially for the complex floating arithmetic and its error.

Section 1.

The main goal of this section is to prove Theorem 4.1 and Theorem 4.2.

Recall that z_0 is a good initial point if $A_{f,z_0} \geqq \pi/12$ and that on the average over $|z|=3$, one out of six choices on $|z|=3$ is a good initial point. See Theorem 1.3.

Also we recall the following from Remark 2.18

Suppose z_0 is a good initial point.

Suppose (1) $|\tilde{z}_{n+1} - z_{n+1}| < \frac{1}{1922} \frac{|f(\tilde{z}_n)|}{|f'(\tilde{z}_n)|}$, where $z_{n+1} = \tilde{z}_n - \frac{f(\tilde{z}_n) - w_{n+1}}{f'(\tilde{z}_n)}$

$$w_{n+1} = \frac{127}{128}^{n+1} f(\tilde{z}_0) \text{ , and } \tilde{z}_0 = f_{z_0}^{-1}((1-\delta_0)f(z_0)) \text{ , } |\delta_0| < \frac{1}{256}$$

Then for any $\varepsilon > 0$, $|f(z_N)| < \varepsilon$ for $N = 128[d\log 3 - \log\varepsilon]$ (2)

Hence the main task is to show that such $\tilde{z}_n$ may be obtained as the computed value of z_n on a finite type machine, eg. a machine using floating point arithmetic. The crucial estimate is obtained in Corollary 4.7, whose proof will be given later.

Corollary 4.7. Let $s_\varepsilon = 17 + 26(d + |\log\varepsilon|)$.

If z_0 is a good initial point, then the s_ε binary floating point arithmetic computation of z_n satisfies (1).

Now we begin the proof of Theorem 4.1.

Proof of Theorem 4.1.

Suppose z_0 is a good initial point. By Corollary 4.7 and Remark 2.18, we obtain z_N such that $|f(z_N)| < \epsilon$, where

(3) $N = 128\,[\,d\log 3 - \log\epsilon\,] < 141\,(d + |\log\epsilon|)$,

(4) Each arithmetic is performed with less than $17 + 26(d+|\log\epsilon|)$ binary binary digits.

I.e. One multiplication costs less than $M(17+26(d+|\log\epsilon|))$ binary bit operations.

(5) Each computaion of z_n costs $2d+2$ complex mutiplications, since computation of $f(z)$ and $f'(z)$ costs $2d$ complex multiplications. This is equivalent to $8d+8$ real multiplications.

Hence the total number of bit operations used for this process is

$< (8d+8)(141(\,d + |\log\epsilon|)\, M(26(d+|\log\epsilon|))$

$\sim 1112\, d\,(d + |\log\epsilon|)\, M(26(d+|\log\epsilon|)$ binary bit operations. (6 *)

If z_0 is not a good initial point it may not produce such z_N. Then we we will try with another initial point. Since it takes only six choices on the average, the average estimate over $S_3^1 = \{|z|=3\}$ is six times (6 *). I.e. For any f in $P_d(1)$, the average cost to find z such that $|f(z)| < \epsilon$, is

$$6672\, d(d + |\log\epsilon|)M(26(d+|\log\epsilon|) \text{ bit operations,}$$

where the average refers to product measure over Ω_3.

Note that this estimate is independent of f.

Now, recall that Approximate-Root termiates when

(7) $\quad 1/42\rho_{f,z} < |f(z)| < 1/7\rho_{f,z}$. (See Proof of Theorem 3.1).

Also from the proof of Theorem 3.1 we recall that the cost used for Test-Approximate is less than $2d|\log \rho_f/42|$ complex multiplications, which is equvalen to $8d|\log \rho_f/42|$ real multiplications.

Hence the total avrage cost over $P_d(1)$ is obtained as

$$\int_{\substack{P_d(1) \\ \rho_f<1}} 6672(d+2|\log \rho_f/42|)M(26(d+|\log \rho_f/42|)$$

$$\approx 6672\, d\, M(26d)\ ,$$

since $\int_{\substack{P_d(1) \\ \rho_f<1}} |\log \rho_f| < \log d + 1$.(See Theorem 1.3)

The proof of Theorem 4.2 is immediate from Theorem 4.1 and Theorem 3.2.

Q.E.D.

The following sequence of results 4.3 through 4.6 will be used to prove Corollary 4.7. We note that Theorem 4.3 and Lemma 4.6 are of interest in their right.

Theorem 4.3. The computed value z_{n+1} of z_{n+1} with s_n binary floating point arithmetic satisfies the condition (*), provided

$$s_n = 3/2[11 + \log d + d\log^+|z_n| - \log \mathrm{Min}[|f(z_n)|, |f'(z_n)|)]$$

where $\log^+ a = \mathrm{Max}\ (0, \log a)$.

Proof From the following lemma one can easily show that

$$\tilde{f}(z) = f(z)(1+\eta) \ , \ |\eta| < 1/4000,$$

$$\tilde{f}'(z) = f'(z)(1+\eta') \ , \ |\eta'| < 1/4000,.$$

where $\tilde{f}(z)$ and $\tilde{f}'(z)$ denote the computed value in s floating point arithmetic. By the property of floating point arithmetic we have

$$\tilde{F}(z) = F(z) \ (1+\delta), \quad |\delta| < 1/1922,$$

where $F(z)= f(z)/f'(z)$, and $\tilde{F}(z)$ is the computed value of $F(z)$.

Q.E.D.

An error analysis of computing the derivatives of real polynomials using Shaw-Traub has been carried out. If complex arithmetic is performed with a double accumulater as in Appendix 2, one obtains the same error estimate .

Lemma 4.4. (Wozniakowski)

Let $\tilde{T}_j$ be the computed value of T_j in fl t computation where $T_j = \frac{f^{(j)}(z)}{j!}$.

i) $\tilde{T}_j = \sum_{k=j}^{d} \binom{k}{j} a_k (1-\eta_{j,k}) z^k, \quad |\eta_{j,k}| < 2k2^{-t}.$

ii) If $f \in P_d(1)$, then

$$|\tilde{T}_j z^j - T_j z^j| < 2^{-t} d^{j+1} \max(|z|^j, |z|^d).$$

Proof For i) see W[Q]. Now by i) , we have

$$|\tilde{T}_j z^j - T_j z^j| < \sum_{k=j}^{d} \binom{k}{j} |\eta_{j,k}||a_k z^k|$$

$$< d^{j+1} 2^{-t} \text{ Max } (|z|^j, |z|^d), \text{ since } |a_i| \leq 1. \text{ Q.E.D.}$$

From Theorem 4.3 we note that the bit complexity depends crucially on the size of $|f|$ and $|f'|$. We need the followin Lemma to estimate $|f'|$.

Let σ be the induced polynomial from f at z , as in Chapter 2.

let $z' = f_z^{-1}[(1-h)f(z)]$, for $h = a\, H_{f,z}$, for $a<1$.

Lemma 5.i) $$\frac{f'(z)}{f'(z')} = \sigma^{-1\prime}(h) .$$

and ii) $$\frac{(1-a)}{(1+a)^3} \leq \left|\frac{f'(z)}{f'(z')}\right| \leq \frac{1+a}{(1-a)^3}$$

Proof. Recall that for $h < H_{f,z}$,

$$z' = f_z^{-1}((1-h)f(z)) = z - \frac{f(z)}{f'(z)}\sigma^{-1}(h) \qquad (2^*)$$

By differentiating (2 *) with respect to h, we have

$$-f(z)\left(f_z^{-1}\right)'((1-h)f(z)) = \frac{-f(z)}{f'(z)}\sigma^{-1\prime}(h)$$

Since $\left(f_z^{-1}\right)'((1-h)f(z)) = \left(f_z^{-1}\right)'(f(z')) = \frac{1}{f'(z')}$,

we have i).

ii) is immediate from the distortion theorem. Q.E.D.

Recall that $A_{f,z}$ denote the angle of the largest wedge where f_z^{-1} can be analitically continued.

Lemma 4.6 Suppose $\left|\frac{f(z')}{f(z)}\right| = L$ and $z' = f_z^{-1}(w)$, for $w \in (o, f(z)]$.

Then we have $e^{-4\log L/\sin A} < \left| \frac{f'(z')}{f'(z)} \right| < e^{4\log L/\sin A}$

Proof. Define $z_{n+1} = f_z^{-1}((1-h)f(z_n))$, $h=a\sin A$ and $a<1$.

By applying Lemma 4.4 recusively, one obtains

$$\frac{|f'(z_n)|}{|f'(z_0)|} > \left[\frac{1-a}{(1+a)^3}\right]^n \quad \text{, where } h = a \sin A,$$

Since $z' = z_N$ with $N = \log L/h$, one has

$$\frac{|f'(z')|}{|f'(z')|} > \left[\frac{1-a}{(1+a)^3}\right]^{\log L/h},$$

This is true for all $0<a<1$, $h=a\sin A$, and hence

$$\frac{|f'(z')|}{|f'(z')|} \geq \lim_{a\to 0} \left[\frac{(1-a)}{(1+a)^3}\right]^{(\log L)\,/\,a\sin A}$$

$$= e^{-4(\log L)/\sin A}$$

Similarly, for the the upper bound. Q.E.D.

Corollary 4.7 . Suppose z_0 is a good initial point.

i) For all n such that $|f(z_n)| < \tilde{\varepsilon}$ we have

$$\log|f'(z_n)| > \log d + d\log 3 - 16\,(d\log 3 - \log\varepsilon), \text{ and}$$

$$s_n < 3/2[11 + 24\,(d+|\log\varepsilon|)] < 17 + 26(d+|\log\varepsilon|), \text{ for all such n.}$$

ii) $|f(z_{n+88})| < |f(z_n)|$ for any n ,and for $1 \leq m < 88$

$s_{n+m} < s_n + 24$.

Proof First note that $|\log |f'(z_0)| > \log d + (d-1)\log 3 - 1$,

$$\log|f(z_0)| < (d+1)\log 3 ,$$

because for a polynomial $f(z) = z^d + ..+ a_i z^i + .. + a_0$, with $|a_i| \leq 1$,

$$|f'(z)| > \frac{d}{2} 3^{d-1} \text{ and } |f(z_0)| > \frac{1}{2}3^d \text{ on } |z| = 3 .$$

Also note that $\sin A_{f,z_n} > 1/4$, for all n ,

since $\sin \arg \dfrac{f(z_n)}{w_n} < \dfrac{1}{128}\dfrac{1}{3} < 0.02$, and $\sin A_{f,z_0} > 0.258.$

Hence by Lemma 4.6 with $\sin A = 1/4$ we have

$$\log|f'(z_n)| > |\log f'(z_0)| - 16\,[\log|f(z_0)| - \log|f(z_n)|] .$$

$$> \log d + d\log 3 - 16\,(d\log 3 - \log \varepsilon),$$

for all z such that $|f(z_n)| > \varepsilon$.

Thus we obtain the estimate s_n in Theorem 4.3 as a binary digit,

$$s_n < 3/2[11 + \log d + d\log 3 - \text{Min}[\log|f'(z_n)|, |f(z_n)|]]$$

$$< 17 + 26\,[d + |\log \varepsilon|], \text{ for all such n.}$$

For ii), the first statement is immediate from Theorem 2.13, since we use

$h = 1/128$ and $\left(\dfrac{127}{128}\right)^{88} < \dfrac{1}{2}$. The second statement is from Lemma 4.6

with $L = 1/2$ and from Theorem 4.3.

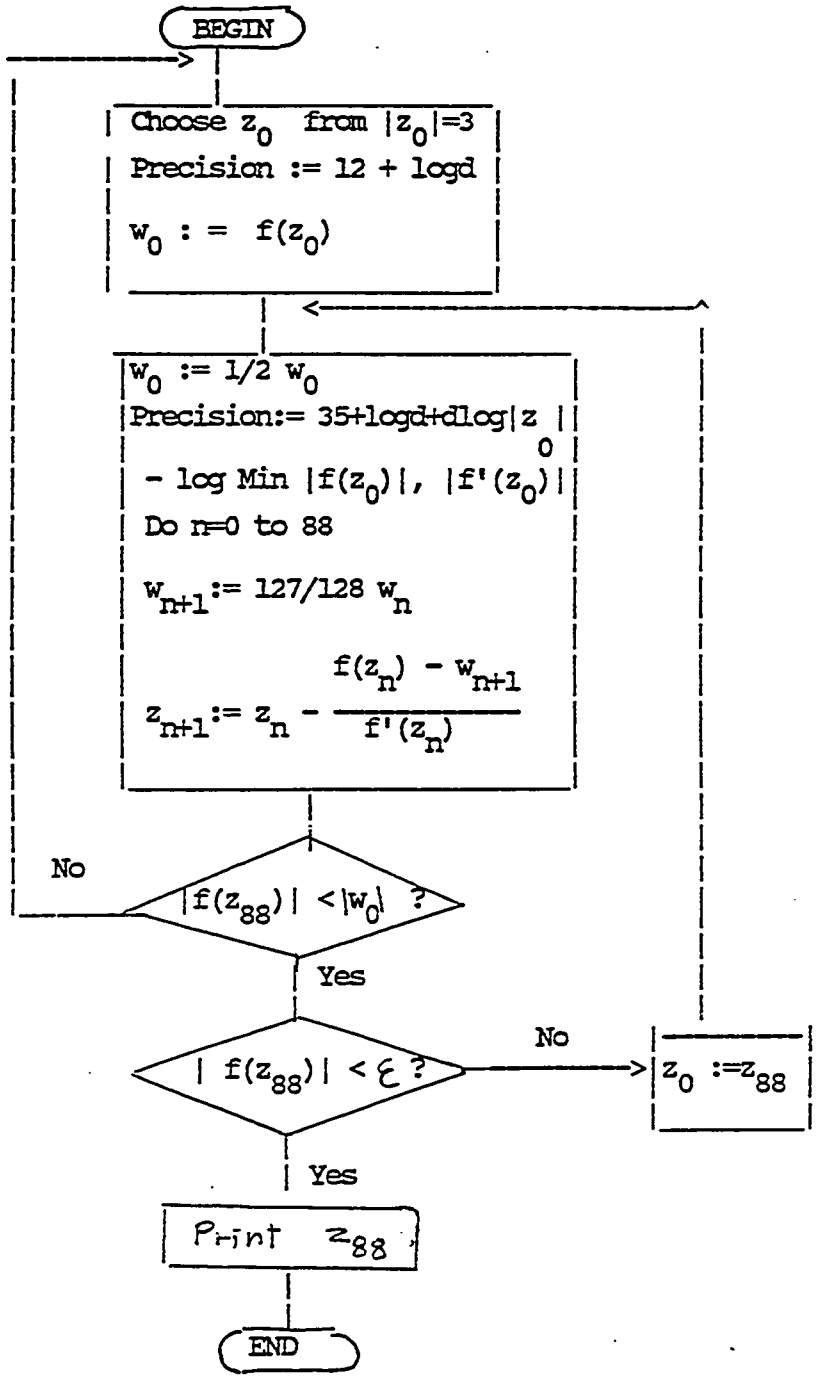
Remark. We design PRECISION-ROOT to use the computational precision as follows. For the flow-Diagram see page 49.

1. Initial Precision $s_0 = 12 + \log d$.

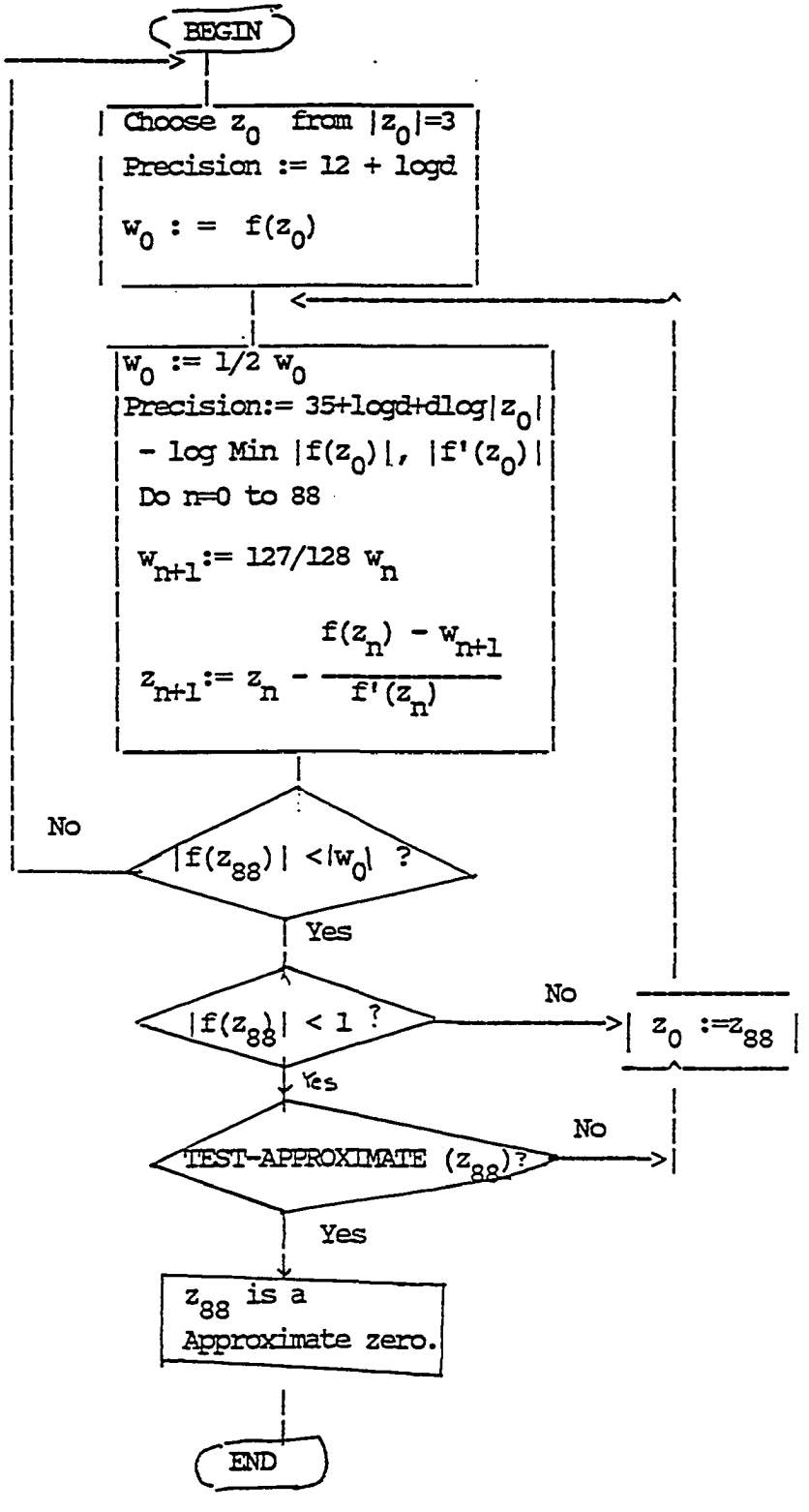
2. Update the precision every 88 steps.

   Set $s = 24 + 11 + \log d + d\log^{+}|z| - \mathrm{Min}[\log|f(z)|, \log|f'(z)|]$.

Precision-Root



Precision-Approximate -Root



section 2

In this section we will generalize the analysis of section 1 to the case of $k > 1$.

The main result in this section is the following.

Impose a normalized product measure on Ω_3 and $\Omega_3 \times P_d(1) \subset S_3^1 \times C^d$.

Theorem 4.8 i) Fast-Root finds z such that $|f(z)| < \varepsilon$, for any $f \in P_d(1)$, for any with the average cost over Ω_3 less than

$$O(d^2 + d|\log|\log\varepsilon|| + k^3)(M(26k(d + |\log\varepsilon|)$$ bit operations.

ii) Fast-Root finds an approximate zero for any $f \in P_d(1)$ without multiple roots with an average cost over $\Omega_3 \times P_d(1)$ less than

$$O(d^2 M(kd) + dk^3 M(kd))$$ bit operations.

Proof will be given later.

Remark. Recall that $ROOT_k$ takes $\frac{4}{\tilde{\mu}_k}(\log|f(z_o)|/\varepsilon)$ iterations to obtain $|f(z)| < \varepsilon$, where $\tilde{\mu}_1 = .0333$, and $\tilde{\mu}_k \uparrow 1$ as $k \uparrow \infty$. See Theorem 3.7. We compare the above result with that of theorem 4.1 and 1.2 and conclude that the advantage of taking a higher k can be only by a factor of 30, while the precision required is increased by a factor of k.

As in Section 1 our aim is to obtain $\tilde{z}_n$ in Theorem 2.13 as the computed value on a machine using floating point arithmetic. Assuming that we have a good initial point, we let $h=r/4$, $r<\tilde{\mu}_k$, and compute $E_{k,h}(z)$ so that the computed value $\tilde{E}_{k,h}(z)$ satisfies

$$(1\ *)\quad |\tilde{E}_{k,h}(z) - E_{k,h}(z)| < |F(z)| \frac{h(k+1)r^k}{(1-r)^2}. \quad (1*)$$

The following Theorem estimates the sufficient precision to obtain such $\tilde{E}_{k,h}(z)$.

Theorem 4.9 Computation of $E_{k,h}$ with t floating point arithmetic satisfies (1 *), provided

$$t = (k+1)[3 + \log d - \log\tilde{\mu}_k + \log^+ |f(z)|/|f'(z)|]$$
$$+d\log^+ |z| - \text{Min}\,[\log|f(z)|, \log|f'(z)|].$$

Proof.

We recall that $E_{k,h}(z) = z - F(z)\, T_k \sigma^{-1}(z)$.

When computing $E_{k,h}$ we use the following sequence of algorithms;

$$(z,f) \xrightarrow[ST]{} (f_j = f^j/j!) \longrightarrow (\sigma_j) \xrightarrow[REV]{} E_{k,h}$$

where $\sigma_j = \left(\frac{-f(z)}{f'(z)}\right)^{j-1} \frac{f^{(j)}(z)}{j!\, f'(z)}$, and ST abbreviates the Shaw-Traub and REV1 is described in appendix 3.

Let $\tilde{f}$, $\tilde{f}'$, $\tilde{F}$ be the computed value of $f(z)$, $f'(z)$ and $F(z)=\frac{-f(z)}{f'(z)}$ respectively, and $(\tilde{\sigma}_i)$, $\tilde{\tilde{REV}}(\sigma)$ be the computed values of $\sigma=(\sigma_i)$ and $REV(\sigma)$.

Then $\tilde{E}_{k,h}(z) = z + \tilde{F}(z)\tilde{\tilde{REV}}(\tilde{\sigma})$, and hence

$$\begin{aligned} |\tilde{E}_{k,h}(z) - E_{k,h}(z)| &= |FREV(\sigma) - \tilde{\tilde{FREV}}(\tilde{\sigma})| \\ &< |F||REV(\sigma) - REV(\tilde{\sigma})| \\ &+ |F||REV(\tilde{\sigma}) - \tilde{\tilde{REV}}(\tilde{\sigma})| \\ &+ |F - \tilde{F}||\tilde{\tilde{REV}}(\tilde{\sigma})|. \end{aligned}$$

Hence in order to satisfy (1 *) it is sufficient to have

$$(2\ *)\quad |\tilde{\tilde{REV}}(\tilde{\sigma}) - REV(\tilde{\sigma})| < 1/2\ h\Delta,$$

$$(3\ *)\quad |REV(\tilde{\sigma}) - REV(\sigma)| < \tfrac{1}{4} h\Delta,$$

and $$(4\ *)\quad \left|\frac{\tilde{F} - F}{F}\right| |\tilde{\tilde{REV}}(\sigma)| < \tfrac{1}{4} h\Delta,$$

where $\Delta = \frac{(k+1)r^k}{(1-r)^2}$.

To obtain (3 *) , it is enough to have

$$(5\ *)\quad \frac{|\tilde{\sigma}_j - \sigma_j|}{(3a)^{j-1}} < \delta\ ,$$

where $\delta = \frac{1}{2} \frac{\Delta}{\sum_{n=1}^{K} (6ah)^{n-1}}$, $a = \text{Max}_i\, |\sigma_i|^{1/i-1}$.

This is because by Theorem A in Appendix 3, (5 *) implies that

$$|REV(\tilde{\sigma}) - REV(\sigma)| < \delta \sum_{n=1}^{k} h^n (6ah)^{n-1} = \tfrac{1}{4} h\Delta$$

Hence (3*) is reduced to obtain $\tilde{\sigma}_j$ satisfying (5 *).

We need the following Lemma 4.10 for (5*).

Let $a = \max |\sigma_i|^{1/i-1}$

Lemma 4.10 Let $\tilde{f}_j = f_j + \delta_j$ where $|\delta_j| < |\delta f_1 (a/F)^{j-1}|$,j=2,...,d.

and $|\delta_0| < \delta |f|$.

Then $$\frac{|\tilde{\sigma}_j - \sigma_j|}{(3a)^{j-1}} < \delta .$$

Proof. By the Mean Value Theorem (See Appedix 3)

$$|\tilde{\sigma}_j - \sigma_j| < \left|\delta_j \frac{\sigma_j}{f_j}\right| + \left|(j-1)\delta_0 \frac{\sigma_j}{f}\right| + \left|j\delta_1 \frac{\sigma_j}{f_1}\right|$$

Note that $\frac{\sigma_j}{f_j} = \frac{1}{f_1} F^{j-1}$, since $\sigma_j = F^{j-1} \frac{f_j}{f_1}$.

By the hypothesis of δ_j we have

$$|\delta_j| \frac{|\sigma_j|}{|f_j|} < \delta \left| f_1 \frac{a}{F}^{j-1} \frac{1}{f_1} F^{j-1} \right| = \delta a^{j-1},$$

$$|\delta_0| \left|\frac{\sigma_j}{f_0}\right| < \delta \frac{|f||\sigma_j|}{|f|} < \delta a^{j-1},$$

and $$|\delta_1| \left|\frac{\sigma_j}{f_1}\right| < \delta \frac{|f||\sigma_j|}{|f_1|} < \delta a^{j-1}.$$

Hence $$\frac{|\tilde{\sigma}_j - \sigma_j|}{(3a)^{j-1}} < \frac{\delta (1+j-1+j)}{(3a)^{j-1}} < \delta .$$ Q.E.D. LEMMA.

Now by Lemma 4.4 and by Lemma 4.10 above, for (5*) it is sufficient to have

$$2^{-t} < \frac{1}{d^2} \frac{1}{\mathrm{Max}(1,z^d)} \min_{2\le j\le k} [|f|,|f'|,(a/dF)^{j-1}].$$

$$= \frac{1}{4} \frac{\Delta}{\sum_{n=1}^{k} (6ah)^{n-1}} \frac{1}{d^2} \frac{1}{\mathrm{Max}(1,z^d)} \min_{2\le j\le k} [|f|,|f'|,(a/dF)^{j-1}].$$

Since $\frac{1}{6 H_{f,z}} < a < \frac{4}{H_{f,z}}$ we have $(6ah) < \frac{24h}{H_{f,z}} < 24r$,

and $\Delta < (k+1)r^k < \frac{(k+1)r^k}{(1-r)^2}$, it is enough to have t such that

$$2^{-t} < \frac{1}{4} \frac{(k+1)r^k}{k\,(24r)^{k-1}} \frac{1}{d^2} \frac{1}{\mathrm{Max}(1,z^d)} \min_{2\le j\le k} [|f|,|f'|,(a/dF)^{j-1}].$$

Hence for the condition (3 *) it is enough to take t,

$$t > (k-1)\log 24 + (K+1)\log d + d\log^{+}|z| - k\log_k$$

$$k \log (|f(z)|/|f'(z)|) - \mathrm{Min}[\log|f|,\log|f'|]$$

For the condition (2 *) it is enough to have (See Theorem C in Appendix 3),

$$t > \log + 2.5k > k\log\tilde{\mu}_k + \log k+1 +2.3k, \text{ since } \log 3/2ah < 0.$$

For the condition (2*), we note that the hypothesis of Lemma 4.10 implies

that $\dfrac{|\tilde{f} - f|}{|f|} < \delta$ and $\dfrac{|\tilde{f}' - f'|}{|f'|} < \delta$, and hence (5*) implies

that $\dfrac{|\tilde{F} - F|}{|F|} < 2\delta$.

Recalling that δ^{-1} is univalent and $\delta^{-1}(0)=0$ and $\delta^{-1\prime}(0) =1$,

by Lemma 2.4 we have $|T_k\delta^{-1}(h)| < 3/2\ h$. Togethe with (3*), we have

that $|\widetilde{REV}(\tilde{\delta})| < 2h$. Thus (2*) follows from (5*).

Hence it is sufficient to have

$$t = (k+1)[3 + \log d - \log\mu_k + \log^{+}|f(z)|/|f'(z)|]$$
$$+d\log|z| - \text{Min}\,[\log|f(z)|, \log|f'(z)|].$$ Q.E.D.

Proof of Theorem 4.8 Note that the estimate in Theorem 4.9 gives

$s_n < O(k(d+|\log\xi|)) \approx M(26k(d + |\log\xi|))$ binary digit,

for all the iterates such that $|f(z_n)| < \xi$. (cf. the proof of Theorem 4.1 and 4.3). Together with Theorem 3.1 we complete the proof. Q.E.D.

Section 3.

In this section , we analyze the bit complexity of N-E_ϵ (N-E) algorithm which are studied in SSII. These algorithms have the advantage of being iterative ,but employ increasing k.

Algorithm N - E_ϵ

1. Set k = Max [logd, log|logϵ|] ,h = 1/500, N= 600 (d +|logϵ|)
2. CHOOSE $|z_0|$=3.
3. REDUCE : DO n=1 TO N

$$z_{n+1} = E_{k,h}(z_n).$$

4. If $|f(z)| < \epsilon$ THEN TERMINATE.

   ELSE GO TO 2.

END.

Algorithm N - E

0. $\epsilon = 1$
1. $\epsilon = \epsilon/2$
2. CHOOSE $|z_0|$=3.
3. Set k = Max [logd, log|logϵ|] ,h = 1/500, N= 600 (d +|logϵ|)
4. REDUCE : DO n=1 TO N

$$z_{n+1} = E_{k,h}(z_n).$$

5. If TEST-APPROXIMATE = YES THEN TERMINATE.

   ELSE GO TO 1.

END.

Impose a normalized Lebesgue measure on $\Omega_3 \times P_d(1)$.

THEOREM 10. i) N-E uses $O(d^2+d|\log\xi|+k^3)\, M(32(\log d+\log|\log\xi|)(d+|\log\xi|)$ bit operations to locate an ξ value of f.

ii) N-E finds an approximate zero of any f without multiple roots and with an average cost over $\Omega_3 \times P_d(1)$ of

$$O\,[d^2+d\log d)M(32d\log d)]$$ bit operations,.

where the constant are 24000.

The proofs are slight modifications of theorem 4.8 and Theorem 4.9, and we will be brief.

Proof. We note that the only situation here different from FAST-ROOT $r=\mu_k$ is replaced by a 1/125, and A_{f,z_n} assumes $>\pi/24$ upon all iterates if an initail point z_0 has $A_{f,z_0} > \pi/12$. And the estimate of $|f'|$ can be at worst $32\,(d\log 3+|\log\xi|)$. (See Lemma 4.6) . Thus this can be dealt with as a special case of $k = \text{Max}\,[\log d, \log|\log\xi|]$ of Theorem 4.8 and Theorem 4.9. Now as in Theorem 4.8 it is sufficient to compute $z_n = E_{k,h}(z_{n-1})$ with a computational precision

$$s_n = (k+1)(8+\log d+\log^+|f|/|f'|) +d\log^+|z| - \text{Min}[\log|f|,\log|f'|].$$

Now the result is straight corollary of Theorem 9. Q.E.D.

APPENDIX 1. DOMAIN OF INJECTIVITY AND ITS APPLICATION TO AN ALGORITHM.

SECTION 1. Domain of Injectivity.

Theorem A1. Let $f(z) = z + a_2z^2 + \dots$ be a power series and g be the compostional inverse of f . Then g is well defined and one to one on $D_R(0)$, where $\frac{1}{6a} < R < \frac{4}{a}$ and $a = \sup_i |a_i|^{1/i-1}$

Proof. Suppose on $|z| = r$ that

i) $|f(z)-z| < r$,

ii) $|f(z)| > R$.

Then g is well defined on $D_R(0)$, because i) and ii) implies that the winding number of the curve $f(|z| = r)$ with respect to each point at $|z| < R$ is one. Now

$$|f(z)| = |z|\,|1 + a_2z + a_3z^2 + \dots|$$

$$> r\,|1 - ((ar)^2 + (ar)^3 + \dots)|$$

$$> r\left(1 - \frac{ar}{1-ar}\right) \text{ on } |z| = r.$$

But $r\left(1 - \frac{ar}{1-ar}\right)$ achieves the maximum $\frac{3 - 2\sqrt{2}}{a}$ when $r = \frac{2-\sqrt{2}}{2a}$

Also $|f(z) - z| = |a_2z + a_3z^3 + \dots|$

$$= |z|\,|a_2z + a_3z^3 + \dots|$$

$$\leq r\,\frac{ar}{1-ar} < r \text{ on } |z| = r\,\frac{2-\sqrt{2}}{2a}$$

Hence g is well defined on $D_R(0)$, where $R = \frac{3 - 2\sqrt{2}}{a}$ $\frac{1}{5.83a} > \frac{1}{6a}$.

For the upper bound see Smale [1]. Q.D.E.

The following is a simple corollary.

Corollary A2. Let f(z) be a polynomial of degree d and z be a complex number. Let f_z^{-1} be the inverse of f such that $f_z^{-1}f(z)=z$. Then f_z^{-1} considered as a power series at f(z) has a radius of convergence

$$R_{f,z} > \frac{|f'(z)|}{6a}, \text{ where } a = \max|T_i|^{1/i-1}$$

and $T_j = \dfrac{f^{(j)}(z)}{f'(z)j!}$

Proof. Let $\sigma(w) = w + \sigma_2 w^2 + \ldots + \sigma_d w^d$.

Then it is known that $|f(z)|R_{\sigma,0} = R_{f,z}$ (See Chapter 2 Fact 2.)

But $R_{\sigma,0} > 1/6a$ by the previous Theorem. Q.E.D.

Section 2. Application to an algorithm.

Applying the estimates in Section 1, we design an algorithm. It is purely iterative and it always converge to a root or a critical point of f when applied to a polynomial f.

ALGORITHM.

$$\text{Define } I(z) = z - h_z \frac{f(z)}{f'(z)},$$

$$\text{where } h_z = \text{Min}\left[\, 1, \frac{1}{42}\left|\frac{f'(z)}{f(z)}\right|, \left|\frac{f'(z)\, j!}{f^{(j)}(z)}\right|^{1/j-1}\right].$$

Theorem 3. $\{I(z)\}$ is a purely iterational scheme and always converges to a zero or a critical point of f, when applied to a polynomial f.

Proof. First note that the fixed points of are exactly zeros and critical points . Also note that h_z satisfies

$$\frac{h_z}{H_{f,z}} < \frac{1}{7} < a_1 \qquad \text{by Theorem A1,}$$

where $H_{f,z} = R_{f,z} / |f(z)|$ and $R_{f,z}$ is the radius of convergence of f_z^{-1} at $f(z)$ as before. Hence by Theorem A in Chapter 2 we have that $|f(z_n)|$ is strictly decreasing, where $z_n = I^n(z)$ for any z with $f'(z)=0$. Thus if it does not converge to a zero then it converges to a critical point of f where the step size h_z tends to a zero.

A note on complex arithmetic and its error.

We refer the reader to W[ ], Knuth[ ] for the details and concepts of of floating point arithmetic. In this Appendix we find that floating point floating point arithmetic with double accumulator (usually denoted by fl_2) is highly satisfactory. Roughly speaking fl_2

performs +,- as a double precision number.

performs x,/ as a single precision number.

stores numbers as a single precision number.

The following lemmas are the standard results and can be found in W[6].

Lemma 1 For real numbers x and y

$$fl(x\ y) = x\ y(1+\varepsilon),\quad |\varepsilon| \leq 2^{-t}$$

where = +,-,x,- and t is the machine precision and division by zero is not permissible. :::

Lemma 2 i) $fl_2(ab+cd) = (ab+cd)(1+\varepsilon),\quad |\varepsilon| \leq 2^{-t}$.

ii) Suppose $3/2\ n\ 2^{-2t} < 0.1$. Then

$$| fl_2(\sum_{i=1}^{n} a_i b_i) - \sum_{i=1}^{n} a_i b_i | \leq | \varepsilon \sum_{i=1}^{n} a_i b_i | + \sum_{i=1}^{n} |a_i b_i \varepsilon_i|,$$

where $|\varepsilon| < 2^{-t}$, and $|\varepsilon_i| < 3/2\ 1.06\ (n+2-i)2^{-2t}$ :::

We consider a complex number as a pair of real numbers and complex arithmetic is carried out in the usual way, i.e.

Let $X = (x_1,x_2)$, $Y = (y_1,y_2)$ be complex numbers.

i) $X + Y = (x_1 + y_1,\ x_2 + y_2)$,carries two real additions.

ii) $XY = (x_1y_1 - x_2y_2,\ x_1y_2 + x_2y_1)$, carries 4 real multiplications and two real additions.

iii) $X/Y = X\tilde{Y} / (Y \tilde{Y})$, carries 6 real multiplications and one real division and two real additions. :::

We define $fl(X) = (flx_1, flx_2)$, for $X = (x_1, x_2)$,

and $fl(Z) = fl(X\ Y) = (flz_1, flz_2)$, for $Z = X \circ Y = (z_1, z_2)$,

where $\circ = +, -, \times, /$. :::

Then we have the following result.

Theorem 1 1) $fl(X \pm Y) = (X \pm Y)(1+\varepsilon)$

where ε a complex number, $|\varepsilon| \leq 2^{-t}$

2) $fl(XY) = XY(1 + \varepsilon)$,

where ε a complex number, $|\varepsilon| \leq (5/2)2^{-t}$

3) $fl(X/Y) = X/Y(1 + \varepsilon)$, $\quad |\varepsilon| \leq 6\ 2^{-t}$

(Proof) Let $X = (x_1, x_2)$ and $Y = (y_1, y_2)$.

1) Addition and Subtraction ; Let $Z = X + Y = (z_1, z_2)$

$$fl(X + Y) = (fl(x_1 + y_1), fl(x_2 + y_2))$$

$$= ((x_1 + y_1)(1 + \varepsilon_1), (x_2 + y_2)(1 + \varepsilon_2))$$

$$= (x_1 + y_1, x_2 + y_2) + (\varepsilon_1(x_1 + y_1), \varepsilon_2(x_2 + y_2))$$

$$= (z_1, z_2) + (\varepsilon_1 z_1, \varepsilon_2 z_2),$$

where $|(\varepsilon_1 z_1, \varepsilon_2 z_2)| < \text{Max}(|\varepsilon_1|, |\varepsilon_2|)\,|(z_1, z_2)| < 2\,|Z|$.

Similarly for the subtraction.

2) Muliplication

$$Z = (z_1, z_2) = XY = (x_1y_1 - x_2y_2 \ , \ x_1y_2 + x_2y_1)$$

$$fl(z_1) = fl(x_1y_2 - x_2y_1)$$

$$= (x_1y_1(1+\varepsilon_1) - x_2y_2(1+\varepsilon_2))(1+\varepsilon_3) \ , \ |\varepsilon_i| < 2^{-t}$$

$$= (x_1y_1 - x_2y_2)(1+\varepsilon) + (x_1y_1\varepsilon_1 - x_2y_2\varepsilon_2)(1+\varepsilon)$$

$$= z_1(1+\varepsilon) + (x_1y_1\varepsilon_1 - x_2y_2\varepsilon_2)(1+\varepsilon)$$

Similarly,

$$fl(z_2) = fl(x_1y_2 + x_2y_1)$$

$$= z_2(1+\eta) + (x_1y_2\eta_1 + x_2y_1\eta_2)(1+\eta).$$

Thus we have

$$fl(Z) = Z + [(\varepsilon z_1, \eta z_2) + ((x_1y_1\varepsilon_1 - x_2y_2\varepsilon_2)(1+\varepsilon), \ (x_1y_2\eta_1 + x_2y_1\eta_2)(1+\eta))]$$

We note that $|x_1y_1 + x_2y_2| \leq |XY|$,

because $|x_1y_1 + x_2y_2| = |Re(XY)| \leq |X\bar{Y}| = |XY|$

and $|x_1y_1 - x_2y_2| = |ReXY| \leq |XY|$.

Hence $|\varepsilon_1 x_1y_1 - \varepsilon_2 x_2y_2| \leq Max\,[|\varepsilon_1|, |\varepsilon_2|]|Z| \leq 2^{-t}|Z|$

Similarly, $|\eta_1 x_1y_2 + \eta_2 x_2y_1| \leq Max\,(|\eta_1|, |\eta_2|)\,|Z| \leq 2^{-t}|Z|$.

Hence $|((1+\varepsilon)(x_1y_1\varepsilon_1 - x_2y_2\varepsilon_2), (1+\eta)(x_1y_2\eta_1 + x_2y_1\eta_2)|$

$$< |(1+\varepsilon, 1+\eta)|2^{-t}|Z| \leq 3/2\ 2^{-t}|Z|. \qquad (1)$$

Also note that $|(z_1, z_2)| \leq Max[|\varepsilon|, |\eta|]|Z| \leq 2^{-t}|Z|$ (2)

Thus $fl(Z) = Z + \Delta$, where $|\Delta| < 5/2\ 2^{-t}|Z|$.

3) Division.

$X/Y = (X\tilde{Y})/Y\tilde{Y} = X\tilde{Y}/(y_1^2 + y_2^2)$.

Since $fl(y_1^2 + y_2^2) = (y_1^2(1+\varepsilon_1) + y_2^2(1+\varepsilon_2))(1+\varepsilon)$

$= (y_1^2 + y_2^2)(1+\varepsilon) + (\varepsilon_1 y_1^2 + \varepsilon_2 y_2^2)(1+\varepsilon)$,

where $|\varepsilon| < 2^{-t}$, $|\varepsilon_i| < 2^{-t}$,

we have $fl(y_1^2 + y_2^2) = (y_1^2 + y_2^2)(1+\varepsilon)$, $|\varepsilon| < 2 \cdot 1.06 \cdot 2^{-t}$

Also $fl(X\tilde{Y}) = X\tilde{Y}(1+\varepsilon)$, $|\varepsilon| < 5/2\ 2^{-t}$ from 2).

Hence we have

$$fl(Z) = fl\left[\frac{fl(XY)}{fl(y^2+y^2)}\right] = \frac{fl(XY)}{fl(y^2+y^2)}(1+\gamma),$$

$= Z(1+\eta)$, $|\eta| < 6\ 2^{-t}$ Q.E.D.

Theorem 2. i) $fl_2(X \circ Y) = (X \circ Y)(1+\varepsilon)$, $|\varepsilon| < 2^{-t}$

where $\circ = +, -, \times$

ii) $fl_2(X/Y) = X/Y\ (1+\varepsilon)$, $|\varepsilon| < 3\ 2^{-t}$

Proof.i) Addition and Subtraction are trivial.

Multiplication .

By Lemma 2 we have

$fl_2(XY) = ((x_1y_1 - x_2y_2)(1+\varepsilon), (x_2y_1 + x_1y_2)(1+\gamma))$

$= XY(1+\delta)$, $|\delta| < \text{Max}[|\varepsilon|, |\gamma|] < 2^{-t}$

ii) Division

Since $fl_2(X/Y) = fl_2\dfrac{fl_2(XY)}{fl_2(\tilde{Y}\tilde{Y})} = \dfrac{fl_2(XY)}{fl_2(\tilde{Y}\tilde{Y})}(1+\varepsilon)$, $|\varepsilon| < 2^{-t}$.

Now $fl_2(XY) = XY(1+\xi_1)$, $|\xi_1| < 2^{-t}$, by i)

and $fl_2(Y\bar{Y}) = (y_1^2 + y_2^2)(1+\xi_2)$, $|\xi_2| < \bar{2}^t$ by Lemma 2.

Hence we have

$$fl_2(X/Y) = X/Y\ (1+\mu), |\mu| < 3\cdot 2^{-t}.$$ Q.E.D.

Appendix 3. The reversion Algorithm and its Error Analysis

The functional inverse of a formal power series is called a 'Reversion Series'. The problem is to solve the equation

(1) $z = t + v_2 t^2 + v_3 t^3 + \ldots$ for t,

and obtain the coefficients of the power series

(2) $t = z + w_2 z^2 + w_3 z^3 + \ldots$

The main purpose of this chapter is to study the error behavior of a reversion series. More precisely, we study

I) The behavior of the reversion series (2) under a perturbation of the coefficients v_i of the power series (1).

II) The round-off error behavior of a reversion algorithm.

We note that I) does not depend on a specific choice of algorithm, while II) depends on the specific algorithm to compute a reversion series. We will analyze the algorithm REV which we construct in the next section. We will see that REV is stable in the sense that the round -off error effect is comparable with the input error effect.

Section 1. Formulation and Algorithm

We recall Lagrange's Inversion Formular;

Lemma 1. Let $z = V(z) = t + v_2 t^2 + \ldots$ and

$t = W(z) = z + w_2 z^2 + \ldots$ be a reversion of V.

I.e. $W(V(t)) = t$ and $V(W(z)) = z$.

Then $w_n = \frac{1}{n} u_{n-1}$, where

$$u_0 + u_1 t + \ldots + u_{n-1} t^{n-1} + \ldots = \left[\frac{v(t)}{t}\right]^{-n}$$

Proof See [13] !!

We need the following lemma.

Let $P = z + z^2 + z^3 + \ldots$ and let $P^k = \sum_{n=k} P_n^k z^n$

Lemma 2. $P_n^k = \binom{n-1}{k-1}$ for all $n \geq k$.

Proof. The proof will be proceed by an induction on k.

k=1; $P_n^1 = 1 = \binom{n-1}{0}$.

In general, $P^{k+1} = P^k P$.

$$= \left(\sum_{n=k} P_n^k z^n\right)(z + z^2 + \ldots)$$

A3-3

$$P_n^{k+1} = \sum_{j=k}^{n-1} P_j^k$$

$$= \binom{n-2}{k-1} + \ldots + \binom{k-1}{k-1} \quad \text{by an induction step.}$$

$$= \binom{n-1}{k} \quad \text{by Vandermonde's Formular.} \quad \text{Q.E.D.}$$

Let $Q(t) = -(v_2 t + v_3 t^2 + \ldots) = \frac{t - V(t)}{t}$ and Q_n^j denote the t^n coefficient of $Q(t)^j$.

Proposition 3. $W_n = \frac{1}{n} \sum_{j=1}^{n-1} \binom{n-1+j}{n-1} Q_{n-1}^j$

Proof. By Lemma 1 we have

$$u_{n-1} = t^{n-1} \text{ coefficient of } \left(\frac{1}{1-Q}\right)^n$$

$$= t^{n-1} \text{ coefficients of } (1 + Q + Q^2 + \ldots)^n$$

$$= t^{n-1} \text{ coefficient of } \sum_{j=1}^{n-1} \binom{n-1+j}{n-1} Q^j \quad \text{by Lemma 2,.}$$

$$\sum_{j=1}^{n-1} \binom{n-1+j}{n-1} Q_{n-1}^j,$$

where $w_n = \frac{1}{n} u_{n-1}$. Q.E.D.

Based on Proposition 3 we construct the following algorithm called REV.

REV

INPUT $v_2,\ldots,v_K$

OUTPUT $w_2,\ldots,w_K$

BEGIN ; DO STEP 1 TO STEP K-1

STEP 1 SET $Q = (Q_1, Q_2, \ldots, Q_{K-1}) = (-v_2, \ldots, -v_K)$

SET $S_1^1 = Q_1^1 = Q$

OUTPUT $W_2 = S_1^1$

STEP j $n=j,\ldots,K-1$

$$Q_n^j = \sum_{i=1}^{n-j+1} Q_{n-i}^{j-1} Q_i$$

$$S_n^j = S_n^{j-1} + \frac{1}{n+1} \binom{n+j}{n} Q_n^j$$

OUTPUT $W_{j+1} = S_j^j$

END

Remark. In REV, $\{Q_n^j ; n = j,\ldots,k-1\} = T_{k-1}(Q)^j$

Theorem 4. i) REV is an reversion algorithm.

ii) Let $U(z)$ be a reversion series of $\sum_{i=1}^{\infty} v_i h^{i-1} z^i$. Then for any k

$$h^k \sum_{n=1}^{\infty} U_n^k = \sum_{n=1}^{\infty} W_n h^n,$$ where $W(z)$ is the reversion series of $\sum_{i=1}^{\infty} v_i z^{i-1}$.

Proof i) is a simple consequence of Propostion 3.

ii) is immediate because $1/hW(hz)$ is the reversion series of $1/hV(hz)$.

Q.E.D.

Remark 5 By Proposition 3,

$$W_{n+1}^{j} = G^j(Q)_n = \frac{1}{n+1} \sum_{m=j}^{n-1} \sum_{i=1}^{n-m} \binom{n+j+i}{n} Q_{n-m}^{i} Q_m^{j}$$

$$= \frac{1}{n+1} \sum_{m=j}^{n-1} \sum_{i=1}^{n-m} \binom{n+j+i}{n} Q_n^{i+j}$$

We call the map G^j a j-th tail of REV, and we will define them precisely in the next section. :::

Section 2. Error Analysis and Mean Value Theorem

Consider the numerical problem of calculating $y=g(x)$, for the given input $x = (x_1,\ldots,x_m)$, and the function $g:D_i \longrightarrow C^n$ and $D_i \subset C^m$.

Definition 1. An algorithm for the given numerical problem $y=g(x)$ is a sequnce of maps $\left(g^i : D_i \longrightarrow D_{i+1},\ D_i \subset C^{n_i}\quad i=0,\ \ldots,r,\right)$ such that $g = g^r \circ g^{r-1} \circ \ldots \circ g^1 \circ g^0$ :::

In other words, an algorithm is a factorization of g into a sequece of maps whose composite is g.

Suppose $G = (g^r, \ldots, g^0)$ is the algorithm to compute $y=g(x)$. In practice, the actual input by which x is computed is $\tilde{G} = (\tilde{g}^r,\ldots,\tilde{g}^0)$, where $\tilde{g}^i$ is a perturbations of g^i. Hence the computed output is dependent on the error behavior of the chosen algorithm.

It is essential for us to develop a method of estimating the error of algorithms. We will need the following notions.

Definition 2. For an algorithm $G = (g^r,\ldots,g^0)$, $G^i = g^r \circ \ldots \circ g^i$ is called an i-th tail of the algorithm G.

Definition 3. An algorithm H is called an ϵ-approximation of G at x if $|H(x) - G(x)| \leq \epsilon$.

For a given input x, consider the algorithm $G = (g^r, \ldots, g^0)$ where $g = g^r \ldots g^0$. Assume that the g^i are C^1 maps and D_i are convex domains in C^{n_i}. Suppose that $\tilde{g}^i$ are ϵ^i-approximations of g^i and that $G = (g^r, \ldots, g^0)$. Set $g^0(x) = x^1$, $\tilde{g}^0(x) = y^1$.

Inductively, set $x^{i+1} = g^i(y^i)$, and

$$y^{i+1} = \tilde{g}^i(y^i) = \tilde{g}^i \circ \tilde{g}^{i-1} \ldots \circ \tilde{g}^0(x)$$

$$\alpha^i = y^i - x^i \quad \text{with} \quad \|\alpha^i\| < \epsilon^{i-1}.$$

Further we assume that $y^i \in D_i$.

The following diagram illustrates the algorithms G and $\tilde{G}$.

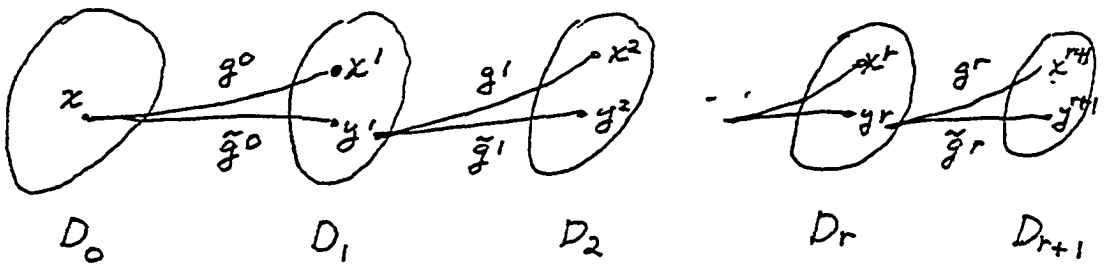


Figure 1.

Then we have the following proposition.

Proposition 5. $\|\tilde{G}(x) - G(x)\| \leq \sum_{i=1}^{r} \sup_t \|DG_t^i\| \|\alpha^i\|$

$$\leq \sum_{i=1}^{r} \sup_t \|DG_t^i\| \epsilon^{i-1},$$

ii) $\| \tilde{G}_n(x) - G_n(x) \| \leq \sum_{i=1}^{r} \sup_t \| (DG_n^i)_t \| \varepsilon^{i-1}$,

where G_n and $\tilde{G}_n$ denote the component maps of G and $\tilde{G}$,

and t runs over the line segment joining x^i and y^i .

We will use the following standard Mean Value Theorem to prove Proposition 6.

Mean Value Theorem Let $F: D_1 \longrightarrow D_2$ be a C^1 map between convex domains $D_1 \subset C^p$ and $D_2 \subset C^q$. Then

$$\| F(p_1) - F(p_2) \| \leq \sup_t \| DF_t \| \, \| p_1 - p_2 \|,$$

where t runs over the line segment joining p_1 to p_2.

Proof of Proposition 6:

We have $g(x) = G^1(x^1) = G^1(y^1 - \alpha^1)$,

so that $\| G^1(y^1) - g(x) \| < \sup_t \| DG^1 \| \| \alpha^1 \|$

$$< \sup_t \| DG^1_t \| \varepsilon^0 ,$$

where t runs over the line segment between x^1 and y^1.

Inductively, $G^i(x^i) = G^i(y^i - \alpha^i)$. Hence we have

$$\|G^i(y^i) - G^i(x^i)\| \leq \sup_t \|DG^i_t\| \varepsilon^{i-1},$$

where t runs over the line segment between x^i and y^i.

Since $G^i(y^i) = G^{i+1}(x^{i+1})$, by the trangle inequality, we have

$$\begin{aligned}\|\tilde{G}(x) - G(x)\| &= \|y^{r+1} - g(x)\| \\ &= \|\sum_{i=1}^{r} G^{i+1}(x^{i+1}) - G^i(x^i)\| \\ &= \sum_{i=1}^{r} \|G^i(y^i) - G^i(x^i)\| \\ &\leq \sum_{i=1}^{r} \sup_t \|DG^i\| \|\alpha^i\| \\ &\leq \sum_{i=1}^{r} \sup_t \|DG^i\| \varepsilon^{i-1}.\end{aligned}$$

ii) is immeadiate from 1). Q.E.D.

Section 3 Error Analysis of Reversion.

We need the following lemmas

Lemma 7 i) $$\sum_{k=1}^{n-1} \binom{n+k-1}{n-1}\binom{n-2}{k-1} < \frac{2}{3}^{n-1}\cdot 6 .$$

ii) $$\sum_{k=1}^{n-m-1} \binom{n+k}{n}\binom{n-m-1}{k-1} < 2^{n}\,3^{n-1-m} .$$

iii) $$\sum_{k=1}^{m} \binom{n+k+j}{n}\binom{m-1}{k-1} < 2^{n+j}\,3^{m-1} .$$

Proof. First note that

$$\frac{d^{n-1}}{dx^{n-1}} x^{n}(1+x)^{n-2} = \frac{d^{n-1}}{dx}\sum_{k=0}^{n-2}\binom{n-1}{k}x^{k+n}$$

$$= (n-1)!\sum_{k=1}^{n-2}\binom{n+k}{n-1}\binom{n-2}{k}x^{k+1}$$

$$\sum_{k=0}^{n-2}\binom{n+k}{n-1}\binom{n-2}{k} = \frac{1}{(n-1)!}\frac{d^{n-1}}{dx^{n-1}}\, x^{n}(1+x)^{n-2}\Big|_{x=1}$$

$$= \frac{1}{2\pi i}\oint \frac{x^{n}(1+x)^{n-2}}{(x-1)^{n}}$$

$$\leq \frac{1}{2\pi}\, 2\pi \operatorname*{Max}_{|x-1|=1}\left|\frac{x^{n}(1+x)^{n-2}}{(x-1)^{n}}\right|$$

$$\leq 2^n 3^{n-2} = \frac{2}{3} 6^{n-1}$$

Remark 6 can be replaced by $3 + 2\sqrt{2} = 5.83 < 6$

ii) $$\sum_{k=1}^{n-m-1} \binom{n+k}{n} \binom{n-m-1}{k-1} = \frac{d^n}{dx} x^n (1+x)^{n-m-1} \Big|_{x=1}$$

$$= \frac{1}{2\pi i} \oint \frac{x^n (1+x)^{n-m-1}}{(x-1)^{n+1}}$$

$$\leq 2^n 3^{n-m-1}$$ Q.E.D.

ii) $$\sum_{k=1}^{m} \binom{n+k+j}{n} \binom{m-1}{k-1} = \frac{d^n}{dx} x^{n+j} (1+x)^{m-1} \Big|_{x=1}$$

$$= \frac{1}{2\pi i} \oint \frac{x^{n+j} (1+x)^{m-1}}{(x-1)^{n+1}}$$

$$\leq 2^{n+j} 3^{m-1}$$ Q.E.D.

Lemma 8. $$\frac{\partial Q_n^k}{\partial v_j} = k\, Q_{n-j}^{k-1}$$

Proof Let $Q^k = (v_2 z + \ldots + v_K z^{K-1})^k$

$$= \sum_{n=k} Q_n^k \, z^n$$

Then we have $Q^k_{n-1} = \sum_{\vec{a} \in A_{n,k}} \frac{k!}{a_2! a_3! \ldots a_n!} \prod v_i^{a_i}$ by the multinomial formula,

where $\vec{a} = (a_2, \ldots, a_n) \in A_{n,k}$ iff

$$\begin{cases} a_2 + \ldots + a_n = k \\ a_2 + \ldots + (n-1)a_n = n-1 . \end{cases}$$

But $\vec{a} = (a_2, \ldots, a_n) \in A_{n,k}$ with $a_j \neq 0$ iff

$$\begin{cases} a_2 + \ldots + (a_j - 1) + \ldots + a_n = k - 1 \\ a_2 + \ldots + (j-1)(a_j - 1) + \ldots + (n-1)a_n = n-j \end{cases}$$

That is iff

$$(a_2, \ldots, a_j - 1, \ldots, a_n) \in A_{n-j+1,k-1} .$$

Hence $$\frac{\partial Q^k_{n-1}}{\partial v_j} = \sum_{\substack{\vec{a} \, A_{n,k} \\ a_j \neq 0}} \frac{k!}{a_2! a_3! \ldots a_n!} \prod v_i^{a_i} \frac{a_j}{v_j}$$

$$= \sum_{\substack{\vec{a} \in A_{n,k} \\ a_j \neq 0}} \frac{(k-1)! \, k}{a_2! \ldots (a_j - 1)! \ldots a_n!} \prod \frac{v_i^{a_i}}{v_j}$$

$$= k \sum_{\vec{a} \in A_{n-j+1,k-1}} \frac{(k-1)!}{a_2!\, a_3! \cdots a_n!} \prod v_i^{a_i}$$

$$= k\, Q_{n-j}^{k-1} \qquad \text{Q.E.D.}$$

Now we are ready to give an error estimate of reversions. The input error and the computational error effect will be treated separately in the following two Theorems.

Theorem A. (INPUT ERROR effect of REV)

Let $V(z) = z + v_2 z^2 + \ldots$ be an almost unit formal power series and

$\tilde{V}(z) = z + \tilde{v}_2 z^2 + \ldots +$, where $\tilde{v}_i = v_i + \Delta_i$.

Let $W(z) = z + w_2 z^2 + \ldots$ be the reversion of $V(z)$ and

let $\tilde{W}(z) = z + \tilde{w}_2 z^2 + \ldots$ be the reversion of $\tilde{V}(z)$. Then we have the following estimates;

$$| \tilde{W}_n - W_n | < \sum_{i=2}^{n} \frac{2}{6} (3a)^{n-i} |\Delta_i|$$

where $a = \max_{i=2}^{n} | |v_i| + |\Delta_i| |^{1/i-1}$.

Proof. From the mean value Theorem, we have

$$| \tilde{W}_n - W_n | \leq \int_0^1 | (DW_n)_t (\tilde{V} - V) |\, dt$$

$$\leq \sum_{i=2}^{n} \sup_{t} \left| \frac{\partial W_n}{\partial V_i} \right|_t |\Delta_i|$$

Using Lemma 8, we have

$$\frac{W_n}{V_i} = \frac{1}{n} \sum_{k=1}^{n-1} \binom{n-1+k}{k} Q_{n-1}^{k}$$

$$= \frac{1}{n} \sum_{k=2}^{n-i} \binom{n-1+k}{n-1} k\, Q_{n-i}^{k-1}$$

$$= \sum_{k=2}^{n-i} \binom{n-1+k}{n-1} Q_{n-i}^{k-1}$$

$$= \sum_{k=1}^{n-i-1} \binom{n-1+k}{n-1} Q_{n-i}^{k-1}$$

Since $|Q_{n-i}^{k}| < P_{n-i}^{k} a^{n-i} = \binom{n-i-1}{k-1} a^{n-i}$, from Lemma 7,

we have $\left| \frac{\partial W_n}{\partial V_i} \right| = \sum_{k=2}^{n-i-1} \binom{n+k}{n} \binom{n-i-1}{k-1} a^{n-i}$

$$< \frac{1}{6} 2^n 3^{n-i} a^{n-i} = \frac{1}{6} 2^n (3a)^{n-i} \quad \text{by Lemma 7.}$$

Hence we have the desired result. Q.E.D.

Corollary. If $|\Delta_i| < \delta (3a)^{i-1}$ then

$$|\tilde{W}_n - W_n| \leq n\, 2^n (3a)^n = \frac{n}{6} (6a)^{n-1} \delta.$$

Proof. Immediate. Q.E.D.

The following Theorem shows that REV is stable in the senge that the round off error is compatible with the input error on REV.

Let g^i be the map performed in step i in REV and let a be as in Theorem A.

Theorem B. (Round-off error effect on REV)

Let $V(z)=z + v_2z^2+ \ldots$ be an almost unit formal power series and

Suppose $\widetilde{\overline{REV}} = (\tilde{g}^{K-1},\ldots,\tilde{g}^1)$, where $\tilde{g}^i$ are ε-approximation of g^i,

Let $\tilde{W} = \{\tilde{W}_2,\tilde{W}_3,\ldots,\tilde{W}_K\} = \widetilde{\overline{REV}}_k(V)$. Then we have

$$| \tilde{W}_n - W_n| < \text{Max} [(6a), 4^{n-1} a].$$

Proof .We recall from Remark 5, that

$$G_n^j (Q^j) = \frac{1}{n+1} \sum_{m=j}^{n-1} \sum_{i=1}^{n-m} \binom{n+i+j}{n} Q_{n-m}^i Q_m^j$$

We note that G^j are linear in Q^j, and hence $DG^j = G^j$.

$$| G_n^j | = \frac{1}{n+1} \sum_{m=j}^{n-1} \sum_{i=1}^{n-m} \binom{n+i+j}{n} | Q_{n-m}^i |$$

$$\leq \frac{1}{n+1} \sum_{m=j}^{n-1} \sum_{i=1}^{n-m} \binom{n+i+j}{n} \binom{n-m-1}{i-1} a^{n-m}, \text{ by Lemma 2,}$$

$$\leq \frac{1}{n+1} \sum_{m=1}^{n-j} \sum_{i=1}^{m} \binom{n+i+j}{n} \binom{m-1}{i-1} a^m,$$

by a change of variable,

$$\leq \frac{1}{n+1} \sum_{m=1}^{n-j} 2^{n+j} 3^{m-1} a^m$$

by Lemma 7.

By Proposition 6,

$$|\tilde{W}_{n+1} - W_{n+1}| < \varepsilon \sum_{j=1}^{n} |G_n^j|$$

$$\leq \frac{1}{3}\frac{\varepsilon}{n+1} \sum_{j=1}^{n} \sum_{m=1}^{n-j} 2^{n+j} (3a)^m$$

$$= \frac{1}{3}\frac{\varepsilon}{n+1} 2^n \sum_{j=1}^{n} \sum_{m=1}^{n-j} 2^j (3a)^m$$

$$< \frac{1}{3}\frac{\varepsilon}{n+1} 2^n \sum_{m=1}^{n} (3a)^m \sum_{j=1}^{n-m} 2^j$$

$$= \frac{1}{3}\frac{\varepsilon}{n+1} 2^n \sum_{m=1}^{n} (3a)^m 2^{n-m+1}$$

$$= 4^n \frac{\varepsilon}{n+1} \sum_{m=1}^{n} (3a/2)^m$$

$$\leq \begin{cases} 2/3\ 4\ (6a)^n & \text{if } 3a/2 > 1 \\ 4^n a & \text{if } 3a/2 \leq 1 \end{cases}$$ Q.E.D.

Recall Theorem 4 that $\sum_{n=1} W_n h^n = h \sum_{n=1} U_n$,

where $\{U_n\}$ is the reversion series of $\{v_i h^{i-1}\}$

Theorem C. Let $\{\tilde{U}_n\} = \text{REV}\ (v_i h^{i-1})$ using t-fl$_2$ computation. Then

$$|h \sum_{n=1}^{K} \tilde{U}_n - \sum_{n=1}^{K} W_n h^n| < \rho 2^{-s}, \text{ provided that } t > s + 2.5(K-1) + 2(K-1)\log^+ 3/2 \ldots$$

Proof. Consider the computaion of X^j in REV ;

$$X_n^j = \sum_{i=1}^{n-j+1} X_{n-i}^{j-1} Q_i .$$

By Lemma 2 in Appendix 2, we have

$$| \tilde{X}_n^j - X_n^j | < (1+\varepsilon) \sum | X_{n-i}^{j-1} Q_i \xi_i | + X_n^j \varepsilon,$$

where $|\varepsilon| < 2^{-t}$, and $|\xi_i| < 3/2\ (n+2-i) 2^{-2t}$.

Since $|X_n^j| < \binom{n}{j} a^n$, $\sum | X_{n-i}^{j-1} | |Q_i| < \binom{n}{j} a^n$, we have

$$| \tilde{X}_n^j - X_n^j | \le \binom{n}{j} (ah)^n 2^{-t} (1 + 2 \cdot 2^{-2t})$$

$$< \binom{n}{j} (ah)^n 2^{-t} , \quad n=j,\ldots,K-1$$

$$\le 2^{-t} 4^{K-1} \mathrm{Max}\,[1, (ah)^{K-1}], \text{ for all } n,j.$$

Now by Theorem B,

$$| h\tilde{\Sigma U}_n - h\Sigma U_n | < \varepsilon h\, 4^{K-1} \mathrm{Max}\,[\, 3/2\ ah,\ (3/2\ ah)^{K-1}]$$

Hence it is sufficient to have

$$2^{-t} \mathrm{Max}[1, (ah)^{K-1}] \cdot \mathrm{Max}[\, 4^{K-1},\ (6ah)^{K-1}] < 2^{-s}.$$

Hence it is sufficient to have

$$2^{-t} 4^{K-1} \mathrm{Max}[1, (ah)^{K-1}] \cdot 4^{K-1} \mathrm{Max}\,[4^{K-1},\ (6ah)^{K-1}] < 2^{-s}.$$

Or sufficient to have

$$t > s + 2(K-1)\log 4 + (K-1)\log^+ ah + (K-1)\log^+ 3/2ah.$$

Or sufficient to have

$$t > s + 2.5(K-1) + 2(K-1)\log^+ 3/2\ ah + \log h. \qquad \text{Q.E.D.}$$

References

1) L. Blum, M. Shub ," Evaluating Rational Functions:Infinite Precision Is Finite and Tractible". Preprint.

2) A. Borodin, I. Munro, The Computaitonal Complexity of Algebraic and Algebraic and Numeric Problems. American Elsevier, New York , 1975

3) Brent , Kung,"Algorithms for Composition and Reversion of Power Series", in Analytic Computational Complexity, 1975

4) P. Henrichi,Applied and Computational Complex Analysis, Wiley ,1977

5)E.Hille, Analytic Function Theory, Vol I,II,Ginn and Company, Boston,1962.

6)D.Knuth,The Art of Computer Programming,Vol II,Addison Wiley, 1962.

7)A. Schonhage,"The Fundamental Theorem of Algebra and Complexity " Preprint.

8)M. Shub ,S.Smale, I,"Computational Complexity : On the Geometry of Polynomials and a Theory of Cost, Part I, Annales Scientifique de l'Ecole Normale Superieure, 1985
Part II,SIAM journal on Computing, to apperar.

9)S.Smale, "The Fundamental Theorm of Algebra and Complexity Theory",Bull.AMS,4,pp 1-36,1981.

10)S.Smale 2,"On the Efficiency of Algorithms of Analysis", Preprint.

11)J.Wilkinson, Rounding Error Analysis in Algebraic Process, Prentice Hall, Englewood Cliff, 1963.

12)H.Wozniakowski, "Rounding Error Analysis for a Polynomial and some of its Derivative", SIAM J.Num. Anal. 11 pp 780-787, 1974.

13)Lagrange,"Memorire Acad. Royale des Sciences et Belle-Lettres de Berlin ,24,1768.