SIAM J. COMPUT.
Vol. 15, No. 1, February 1986

011

COMPUTATIONAL COMPLEXITY: ON THE GEOMETRY OF POLYNOMIALS AND A THEORY OF COST: II*

M. SHUB† AND S. SMALE‡

Abstract. This paper deals with traditional algorithms, Newton's method and a higher order generalization due to Euler. These iterations schemes and their modifications have had a great success in solving nonlinear systems of equations. We give some understanding of this phenomenon by giving estimates of efficiency. The problem we focus on is that of finding a zero of a complex polynomial.

Key words. Newton, Euler, approximate zero, polynomial, average

1. This paper deals with traditional algorithms, Newton's method and a higher order generalization due to Euler. These iteration schemes and their modifications have had a great success in solving nonlinear systems of equations. We give some understanding of this phenomenon by giving estimates of efficiency. The problem we focus on is that of finding a zero of a complex polynomial.

Following the work of Newton and Euler we define a rational map (or iteration scheme) $E:\mathbb{C}\to\mathbb{C}$ ($\mathbb{C}$ the complex numbers) which depends on three parameters:

(a) f, a polynomial, $f(z)=\sum_{i=0}^{d} a_i z^i$, $a_d\neq 0$. Often times we take f to be in the space $P_d(1)$ where

$$P_d(1)=\left\{f\ \middle|\ f(z)=\sum_{i=0}^{d} a_i z^i,\ a_d=1,\ |a_j|\leqq 1\right\};$$

(b) a positive integer k (which amounts to the number of derivatives used);
and

(c) A number h, $0<h\leqq 1$.

Then define $E=E_{k,h,f}$, $E:\mathbb{C}\to\mathbb{C}$ by

$$E(z)=T_k(f_z^{-1}((1-h)f(z))).$$

Here f_z^{-1} is the branch of the inverse of f which takes $f(z)$ into z, given as an analytic function in a neighborhood of $f(z)$ (provided $f'(z)\neq 0$).

T_k is the truncation of the power series expansion in h about $h=0$ at degree k. It is easy to check that $E_{1,1,f}$ is Newton's method. One can see a full discussion in Shub-Smale (1982) (hereafter referred to as [S-SI]).

Consider first the problem: Given (f,ε), $f\in P_d(1)$, $\varepsilon>0$, produce a $z\in\mathbb{C}$ with $|f(z)|<\varepsilon$. For this we particularize the Newton-Euler iteration scheme by choosing k and h to depend only on f and ε, in a certain way. Let

$$k=\lceil \max(\log|\log\varepsilon|, \log d)\rceil$$

where $\lceil x\rceil$ is the least integer greater than or equal to x. We will define in § 2, universal constants H and K, approximately $\frac{1}{512}$ and 512 respectively. Then we will take

$$h=H.$$

* Received by the editors November 17, 1983, and in revised form September 15, 1984. This work was supported in part by grants from the National Science Foundation.

† Mathematics Department, Queens College and the Graduate School, City University of New York, New York, New York 11367.

‡ Mathematics Department, University of California, Berkeley, California 94720.

Thus with these specializations the Newton–Euler iteration scheme $E:\mathbb{C}\to\mathbb{C}$ depends only on (ε, f) and we write $E_\varepsilon = E$. With $\varepsilon > 0$ define

ALGORITHM $(\text{N-E})_\varepsilon$. Let $f \in P_d(1)$ and $n = K(d+|\log \varepsilon|)$.

(1) Choose $z_0 \in \mathbb{C}$, $|z_0| = 3$ at random and set for $i = 1, 2, 3, \cdots$ (an *iteration*) $z_i = E_\varepsilon(z_{i-1})$ terminating if ever $|f(z_i)| < \varepsilon$

(2) If $i = n$, go to (1) (a *cycle*).

THEOREM A. *For each f, ε, $(\text{N-E})_\varepsilon$ terminates with probability one and produces a z satisfying $|f(z)| < \varepsilon$. The average number of cycles is less than or equal to 6. Hence the average number of iterations is less than $6K(d+|\log \varepsilon|)$.*

Here average and probability refer to the choice of the sequence of z_0 in (1) of $(\text{N-E})_\varepsilon$.

Remark. With certainty it only takes about twice as long. See § 2 for an elucidation of this remark. In practice one can obviously do better by trying and testing $h = 1, \frac{1}{2}, \cdots, H$. We have not analyzed this. Also see § 2 for the total number of arithmetic operations required.

Next consider sharpening the goal $|f(z)| < \varepsilon$. Machine or discrete processes will not generally succeed in finding exact zeros of polynomials. For our theory we use the notion of approximate zero z of a polynomial f, Smale (1981), [S-SI]. This complex number z is one close to an actual zero, where closeness is defined without any arbitrary choices. The justification of close is given in both theoretical and practical terms. More precisely define

$$\rho_f = \min_{\substack{\theta \\ f'(\theta)=0}} |f(\theta)|.$$

There is a universal constant c (about $\frac{1}{12}$) and z is an *approximate zero* of f if $|f(z)| < c\rho_f$. Then this proposition follows.

PROPOSITION. *Let $E = E_{k,f,1}$ be the Newton–Euler scheme with arbitrary k and $h = 1$, and E^l the composition $E \circ \cdots \circ E$ l times. Let z be an approximate zero of f, so $|f(z)| < bc\rho_f$ with $b < 1$. Then*

$$|f(E^l(z))| < b^{(k+1)^l} c\rho_f.$$

The extremely rapid convergence gives some good justification for "approximate zero".

In Algorithm $(\text{N-E})_\varepsilon$ there was a random element, the choice of z_0. Now probability enters into our analysis in a second way. We average over $f \in P_d(1)$ with respect to a uniform distribution; that is we normalize Lebesgue measure on $P_d(1) \subset \mathbb{C}^d = \mathbb{R}^{2d}$. We use these probabilities since speedy algorithms are not usually infallible.

Define for each $f \in P_d(1)$

$$\varepsilon_f = \frac{1}{(2d)^{4d}} |D_f|$$

where D_f is the discriminant of f (see Lang (1965)). With K as above let

$$n = \lceil K(d + |\log \varepsilon_f|) \rceil.$$

Let E be the Euler–Newton iteration scheme with $h = H$, and $k = \lceil \max(\log|\log \varepsilon_f|, \log d) \rceil$, so that E depends only on f.

ALGORITHM (N-E). Let $f \in P_d(1)$, satisfy $\varepsilon_f > 0$.

(1) Set $m = 1$;

($2m$) Choose $z_0 \in \mathbb{C}$, $|z_0| = 3$ at random and set $z_n = E^n(z_0)$. If $|f(z_n)| < \varepsilon_f$ terminate and print: "z_n is an approximate zero;"

(3) Otherwise let $m = m+1$ and go to (2_m).

THEOREM B. *Algorithm* (N-E) *terminates* (*and hence produces an approximate zero*) *with probability* 1 *and the average number of iterations is less than* $K_1 d \log d$ *where* K_1 *is a universal constant.*

We make the probability considerations a bit more precise.

Let S_R^1 be the circle in $\mathbb{C}$ defined by $|z| = R$ and endow it with the uniform probability measure (Lebesgue measure normalized to 1). Set $R = 3$ and denote by Ω the product of S_R^1 with itself a countable number of times. Thus a point z_0 of Ω is a sequence $\bar{z} = (\bar{z}_1, \bar{z}_2, \cdots)$ with $|\bar{z}_i| = 3$. Endow Ω with the product measure as well as $P_d(1) \times \Omega$. Let $T: P_d(1) \times \Omega \to Z^+$ be defined by: $T(f, \bar{z})$ is the first m such that $E^n(\bar{z}_m) < \varepsilon_f$.

Thus the total number of iterations of Algorithm (N-E) for a given f is of the form $S(f, \bar{z}) = nT(f, \bar{z})$, $n = K(d + |\log \varepsilon_f|)$. Theorem B asserts than when $\varepsilon_f > 0$, $S(f, \bar{z})$ is defined for almost all $\bar{z} \in \Omega$. Moreover $S(f) = \int_{\bar{z} \in \Omega} S(f, \bar{z})$ is defined and finite for almost all f and

$$\int_{f \in P_d(1)} S(f) \leqq K_1 d \log d.$$

By Fubini's theorem, we could equally well assert that

$$\int_{(f, \bar{z}) \in P_d(1) \times \Omega} S(f, \bar{z}) \leqq K_1 d \log d.$$

Remark 1. We are assuming exact arithmetic in the theory here. In general, because of the robust properties of Algorithms $(\text{N-E})_\varepsilon$ and (N-E), this is reasonable. However the calculation of ε_f in Algorithm (N-E) is not so robust. In that respect, Theorem A is more satisfying than Theorem B.

Remark 2. Our work emphasizes the theoretical side, and the understanding of classic algorithms, rather than the design of new practical algorithms. Yet the results do have some implications for the latter. For example they suggest calculating derivatives up to order $\lceil \log d \rceil$ and/or $\lceil \log |\log \varepsilon| \rceil$ could give speedier routines, especially for one complex polynomial. We have not tested our algorithms on the machine.

Remark 3. The number of arithmetic operations in contrast to the number of iterations is approximately quadratic in d. This is proved in § 2.

Remark 4. Questions of variance arising in these theorems can be handled. See § 2.

At this point we review some of the motivation from Smale (1981), Hirsch and Smale (1979) and [S-SI]. Let f be a polynomial, $f: \mathbb{C} \to \mathbb{C}$, let $z \in \mathbb{C}$ and $w = f(z)$. The ray from w to 0 is the segment in the target space from w to 0. Let R_w denote this ray and f_z the branch of the inverse of f taking $f(z)$ back to z. If f_z^{-1} is defined on all of R_w then $\zeta = f_z^{-1}(0)$ is one of the zeros of f (Fig. 1). Since a polynomial maps a

FIG. 1

neighborhood of infinity to a neighborhood of infinity and has only finitely many critical points it is a fairly simple calculus exercise to see that except for a finite number of rays (at most $d-1$) f_z^{-1} is defined on all of R_w. Thus we attempt to follow the curves $f_z^{-1}(R_w)$ from an initial starting point z to a zero ζ of t. One way to do this is to parameterize the R_w as $(1-h)f(z)$ for $0 \leqq h \leqq 1$. Then try to follow the ray by analytic continuation in h, $f_z^{-1}((1-h)f(z))$. Finally, truncate the power series at degree k in h to make the computation finite, $T_k(f_z^{-1}((1-h)f(z))$. This is $E_{k,h,f}(z)$. If $P(h)$ is positive for small positive h then $(1-P(h))$ is also further down the ray and we may try $T_k f_z^{-1}(1-P(h))f(z))$. The inverse images of the rays are also solution curves of the differential equations $\dot{z} = -f(z)/f'(z)$ and $\dot{z} = -\frac{1}{2}\operatorname{grad}|f(z)|^2$ (see Smale (1981)). Thus we may attempt to solve these equations numerically with step size h to attempt to follow the ray. These examples are given in greater detail in [S-SI].

We analyze a class of fast algorithms broader than the Euler iterations, but which still agree with the inverse of the ray to high order. First we extend the $E_{k,h,f}(z) = T_k(f_z^{-1}((1-h)f(z)))$ by replacing h on the right-hand side by $P(h) = \sum_{i=1}^k c_i h^i$ where c_i is real for all i and $c_1 > 0$. These generalized Euler iterations are

$$\mathrm{GE}_{p,k,h,f}(z) = T_k(f_z^{-1}((1-P(h))f(z)).$$

Thus $E_{k,h,f}$ is given by $P(h) = h$. The $\mathrm{GE}_{P,k,h,f}$ are polynomials in h of degree k. We allow modifications of these polynomial iterations by addition of a well bounded remainder term of order $k+1$ in h. We denote this largest class of iterations we consider by GEM_k (GEM = Generalized Euler with Modification). These iterations are fast.

An important ingredient in the analysis is the function $P_d \times \mathbb{C} \to R^+$. $(f, z) \to \Theta_{f,z}$ which was introduced in [S-SI]. We recall the definition of this function. Given f, z and $0 \leqq \alpha \leqq \pi/2$, let

$$W_{f,z,\alpha} = \left\{ w \in \mathbb{C} \,|\, 0 < |w| < 2|f(z)|, \left| \arg \frac{w}{f(z)} \right| < \alpha \right\}.$$

W is an open wedge of angle α centered at $f(z)$. Let $\Theta_{f,z}$ be the max $\alpha \leqq \pi/2$ on which f_z^{-1} is defined by analytic continuation (Fig. 2).

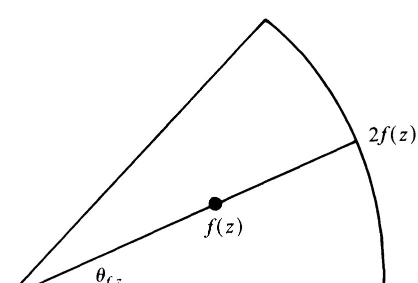

FIG. 2. *f_z^{-1} is defined on this wedge.*

In § 3 we prove Theorem C.

THEOREM C. *Suppose that $z' = I_{h,f}(z)$ is a GEM_k iteration. Then there is a constant k depending only on I such that: If $\Theta_{f,z_0} > 0$ and $|f(z_0)| > L > 0$ then there is an h given explicitly such that*

$$|f(z_n)| < L \quad \textit{for } n = k\left[\frac{\log |f(z_0)/L|}{\Theta_{f,z_0}}\right]^{(k+1)/k}$$

and $z_n = (I_{h,f})^n(z_0)$.

Theorem C can be used to show that any GEM_k iteration can be adapted to produce fast algorithms as in Theorems A and B of this paper. Finally, in § 3 we show that the GEM_k iterations are precisely the efficiency k incremental algorithms defined in [S-SI] which satisfy an additional "smallness" condition.

2. The main goal of this section is to prove Theorems A and B of § 1. We require some of the main results of [S-SI]. First Proposition 1 is a special case of [S-SI, Thm. 2] at least after a short translation of constants. The constants K, K' are universal, not very large and well estimated via [S-SI]. Let $k=1,2,\cdots$, and $\varepsilon>0$.

PROPOSITION 1. *There exist* K, $K'>0$ *so that if* $0<h\leqq K/(d+|\log \varepsilon|)^{1/k}$, $n=(K'/h)(d+|\log \varepsilon|)$, $f\in P_d(1)$ *and* $|z_0|=3$ *with* $\Theta_{f,z_0}\geqq \pi/12$ *then*

$$|f(E^i(z_0))|<\varepsilon \quad \textit{for some } 0\leqq i<n.$$

Here E^i *is* $E\circ\cdots\circ E$, *i times and* E *is the* Euler_k *iteration of* § 1.

Next by specializing k to $k=\max(\lceil\log d\rceil, \lceil\log|\log \varepsilon|\rceil)$ we obtain

COROLLARY. *There exist universal constants* H, K *so that for* $n=K(d+|\log \varepsilon|)$, $E=E_{k,H}$, $f\in P_d(1)$ *and* $|z_0|=3$ *with* $\Theta_{f,z_0}\geqq \pi/12$

$$|f(E^i(z_0))|<\varepsilon \quad \textit{for some } 0\leqq i<n.$$

Finally we require the following proposition which is obtained from [S-SI, Proposition 3, § 4]. For $f\in P_d(1)$, let $V_f=\{z\,\|z|=3 \text{ and } \Theta_{f,z}>\pi/12\}$. Then using the uniform probability measure on $S=\{z\,\|z|=3\}$ we have Proposition 2.

PROPOSITION 2. *The measure of* $V_f\geqq \frac{1}{6}$ *for any* f.

We recall a bit of probability theory. The set S has a probability measure. Impose the product measure on Ω the (ordered) countable product of S with itself. Suppose $V\subset S$ has measure v. For $\bar{z}\in\Omega$, $\bar{z}=(\bar{z}_1,\bar{z}_2,\cdots)$. Let $m(\bar{z})$ be the m such that $\bar{z}_i\notin V$ for $i<m$ but $\bar{z}_m\in V$.

PROPOSITION 3.

$$\int_{\bar{z}\in\Omega} m(\bar{z})=\frac{1}{v}.$$

Proof. Let $V_i=\{\bar{z}\,|\,m(\bar{z})=i\}$, $i=1,2,\cdots$. Let v_i be the measure of V_i. Then $v_i=v(1-v)^{i-1}$ and moreover

$$\int_{\bar{z}\in\Omega} m(\bar{z})=\sum_{i=1}^{\infty} iv_i=\sum_{i=1}^{\infty} iv(1-v)^{i-1}=\frac{1}{v}.$$

Now we can prove Theorem A of § 1. In fact it is an immediate consequence of the corollary of Proposition 1, and Propositions 2 and 3. Q.E.D.

We will need a few more facts to prove Theorem B. We begin with another elementary result of probability theory.

DEFINITION. Let (X,μ) be a probability space with no atoms. Let $S: X\to R_+$ be a real valued nonnegative measurable function and let $f:(0,1)\to R$ be decreasing and Riemann integrable. We say that $S(x)\leqq f(\mu)$ with probability $1-\mu$ if $\mu\{x\,|\,s(x)\leqq f(y)\}\geqq 1-y$ for all $0<y<1$.

PROPOSITION 4. *Suppose as above that* $S(x)<f(\mu)$ *with probability* $1-\mu$. *Then*

1) $E(S)=\displaystyle\int_x S(x)\mu(dx)\leqq \int_0^1 f(\mu)\,d\mu.$

2) $\mathrm{Var}(S)=\displaystyle\int_x (S(x)-E(S))^2\mu\,(dx)\leqq \int_0^1 f^2(\mu)\,d\mu-(E(S))^2.$

Proof. We only prove 1). Let $y_i \in (0,1)$ for $-\infty < i < \infty$ be a decreasing sequence with 0 and 1 as limit points. Let $\cdots \supset M_{y_i} \supset M_{y_{i-1}} \supset \cdots$ be constructed so that $\mu(M_{y_i}) = 1 - y_i$ and $S(x) \leqq f(y_i)$ for $x \in M_{y_i}$. Then $\int S(x)\mu(dx) \leqq \sum f(y_i)\mu(M_{y_i} - M_{y_{i-1}}) = \sum f(y_i)(y_{i-1} - y_i)$ which converges to the Riemann integral of f.

Recall that $\rho_f = \min_{\Theta, f'(\Theta)=0} |f(\Theta)|$. From Smale (1981) we have Proposition 5.

PROPOSITION 5.

$$\operatorname{Vol}\{f \in P_d(1) \mid \rho_f < \alpha\} < d\alpha^2.$$

Here Vol means normalized volume so that $\operatorname{Vol}(P_d(1)) = 1$ and Vol is a probability measure on $P_d(1)$.

PROPOSITION 6.

$$\int_{\substack{f \in P_d(1) \\ \rho_f < 1}} |\log \rho_f| \leqq \tfrac{1}{2} \log d + 1.$$

Proof. Let $\mu = d\alpha^2$ so from Proposition 5, $\operatorname{Vol}\{f \in P_d(1) \mid \rho_f < \sqrt{\mu/d}\} < \mu$ and

$$\operatorname{Vol}\{f \in P_d(1), \rho_f < 1 \mid |\log \rho_f| < \tfrac{1}{2} \log d + \tfrac{1}{2} |\log \mu|) > 1 - \mu.$$

Now apply Proposition 4 to finish the proof.

PROPOSITION 7. *There is a constant K_2 and for $f \in P_d(1)$,*

a) $d^d \rho_f^{d-1} \leqq D_f$,

b) $\varepsilon_f \leqq c\rho_f$,

c) $\int_{P_d(1)} |\log \varepsilon_f| \leqq K_2 d \log d$.

Proof. $D_f = d^d \prod_{\theta_i \text{s.t.}, f'(\theta_i)=0} f(\theta_i)$ (see Lang (1965)) so

$$|D_f| \geqq d^d (\min_{\substack{\theta_i \text{s.t.} \\ f'(\theta_i)=0}} |f(\theta_i)|)^{d-1} = d^d \rho_f^{d-1}.$$

which proves a). We will use a lemma to prove b). First recall that the discriminant is given as the determinant of a $(2d-1)$ by $(2d-1)$ matrix, see Lang (1965). $D_f = R(f, f') = d^d R(f, f'/d)$.

LEMMA 1. 1) *For a polynomial f of degree d with $|a_i| \leqq 1$ for $i \geqq 1$ then*

$$\left| \frac{\partial R(f, f'/d)}{\partial a_0} \right| \leqq (2d-1)! \sum_{j=1}^{d-1} j \binom{d-1}{j} |a_0|^{j-1}.$$

2) *If $f \in P_d(1)$ then $(2d-1)!(d-1)(\rho_f + 2)^{d-2} \rho_f \geqq R(f, f'/d)$.*

Proof. The resultant is a polynomial of degree $d-1$ in a_0. Crude estimates give that the modulus of the coefficient of a_0^j by expanding the determinant is $\leqq \binom{d-1}{j}(2d-1-j)!$ so

$$\left| \frac{\partial R(f, f'/d)}{\partial a_0} \right| \leqq \sum_{j=1}^{d-1} j \binom{d-1}{j} (2d-1-j)! |a_0|^{j-1}$$

$$\leqq (2d-1)! \sum_{j=1}^{d-1} j \binom{d-1}{j} |a_0|^{j-1}.$$

Let $f'(\theta) = 0$ and $|f(\theta)| = \rho_f$.

Let $f_0 = f - f(\theta)$. Then $a_0(f_0) = a_0 - f(\theta)$, $a_i(f_0) = a_i(f)$ for $i > 1$ and $R(f_0, f_0'/d) = 0$.

Applying the mean value theorem to the line segment between f_0 and f gives

$$\begin{aligned} |R(f,f'/d)| &= |R(f,f'/d)-R(f_0,f_0'/d)| \\ &\leqq (2d-1)!\sum_{j=1}^{d-1} j\binom{d-1}{j}(|a_0|+\rho_f)^{j-1}|f-f_0| \\ &\leqq (2d-1)!\sum_{j=1}^{d-1} j\binom{d-1}{j}(\rho_f+1)^{j-1}\rho_f \\ &= (2d-1)!\rho_f\left(\left(\sum_{j=0}^{d-1}\binom{d-1}{j}x^j\right)'\Big|_{\rho_f+1}\right) \\ &= (2d-1)!\rho_f((1+x)^{d-1})'|_{\rho_f+1} \\ &= (2d-1)!(d-1)(\rho_f+2)^{d-2}\rho_f. \end{aligned}$$

Now to prove b), note that

$$\rho_f < d+1 \quad \text{and} \quad d^d(2d-1)!(d-1)(d+3)^{d-2} < c(2d)^{4d}.$$

For the proof of c), note that $\varepsilon_f < 1$.

By definition

$$|\log \varepsilon_f| = \left|\log \frac{D_f}{(2d)^{4d}}\right|.$$

Therefore by using Proposition 7a,

$$|\log \varepsilon_f| \leqq K_3 d|\log d + |\log \rho_f|| \quad \text{and}$$

$$\int |\log \varepsilon_f| \leqq K_3 d\left(\log d + \int_{\rho_f<1} |\log \rho_f|\right) + K_3 d\left(\log d + \int_{\rho_f>1} |\log \rho_f|\right).$$

Now the first term on the left is estimated by Proposition 6. The second is estimated using the fact that $\rho_f < d+1$. This yields Proposition 7.

With these preparations we now prove Theorem B from Theorem A. Note first that by Proposition 7b, Algorithm (N-E) does terminate with an approximate zero if it terminates. Now apply Theorem A with Algorithm $(\text{N-E})_\varepsilon$ where ε is chosen to be ε_f, and then integrate over $P_d(1)$. This yields

$$\int_{P_d(1)\times\Omega} S(f,\bar{z}) \leqq 6Kd + 6K\int_f |\log \varepsilon_f|.$$

Now apply Proposition 7c. Q.E.D.

Remark 1. There are various ways to compute the Euler iterations

$$z' = E_{k,h}(z_0) = T_k(f_z^{-1}((1-h)f(z))).$$

a) The Taylor series expansions of f at z, f_z may be calculated by the algorithm of Shaw and Traub with $(2d-1)$ multiplications, $d-1$ divisions and $d^2+d/2$ additions, after writing $f(w)=(w-z_0)Q(w)+R$. See Knuth (1981, p. 470) or Borodin-Munro (1975, p. 33). Moreover f_z may be calculated in $O(d \log d)$ arithmetic operations and perhaps in $O(d)$ operations, see Borodin-Munro (1975, p. 106).

b) $T_k(f_z^{-1})$ may then be computed in $k^3/6$ multiplications by Lagrangian power series reversion (Knuth (1981, p. 508)). The algorithm of Brent-Kung (1976) (see also Knuth (1981, p. 510)) computes $T_k f_z^{-1}$ in $O(k \log k)^{3/2}$ operations and the value $T_k f_z^{-1}((1-h)f(z))$ in $O(k \log k)$ operations once the Taylor series of f at z is known.

c) Putting together a) and b) to get a good asymptotic estimate gives that $T_k f_z^{-1}((1-h)f(z))$ is computable in $O(d+k\log k)$ operations.

d) Thus in Theorem A the average of the total number of operations is

$$O((d+|\log\varepsilon|)(d+k\log k)),$$

where $k=\max\lceil\log d, \log|\log\varepsilon|\rceil$, or simply

$$O(d^2+d|\log\varepsilon|).$$

e) For Theorem B, the situation is a bit more subtle since k depends on f. The average total number of operations is

$$O\left(\int_{P_d(1)\times\Omega} S(f,z_0)(d+k\log k)\right)$$

where $k=\max\lceil\log d, \log|\log\varepsilon_f|\rceil$, which gives eventually

$$O(d^2(\log d)^2\log\log d).$$

f) D_f may be computed in $O(d(\log d)^2)$ arithmetic operations according to J. T. Schwartz (see Knuth (1981, p. 619)).

Remark 2. An estimate for the variance of the numer of iterates in Theorem B is simple to calculate by Proposition 4. The variance is $O(\int_f (d\log d+|\log\rho_f|)^2)$ which is $O(d^2(\log \mathrm{d})^2)$.

Remark 3. It is possible to devise deterministic algorithms in place of Algorithms (N-E)$_\varepsilon$ and (N-E). By [S-SI, Prop. 3, § 4] $\Theta_{f,z_0}\geqq\pi/12$ except for at most $2(d-1)$ arcs of S_3^1 of angle $10\pi/12d$ each. Thus if we place $24d$ points evenly around S_3^1 at least one-twelfth of them will have Θ bigger than $\pi/12$. Now for the sake of Theorem A or B we can pick from this finite set at random eventually exhausting the set and still not take more than 12 choices on the average. Thus for Theorems A and B a deterministic version is possible with a factor of 2 in the number of steps. Now the algorithms always terminate. It is not necessary to say with probability one (in the case of Theorem B as long as $\varepsilon_f>0$).

3. We begin this section with a description of the GEM (Generalized Euler with Modification) iterations that we consider. Throughout this section $P(h)=\sum_{i=1}^k c_i h^i$ is a polynomial with c_i real and $c_1>0$. Thus for small real h, $(1-P(h))f(z)$ is on the line segment between $f(z)$ and 0. The Generalized Euler iterations are $\mathrm{GE}_{p,k,h,f}(z)=T_k(f_z^{-1}((1-P(h))f(z)))$.

DEFINITION 1. $z'=I_{n,f}(z)$ is a GEM_k iteration iff there is a polynomial $P(h)$ and constants $c>0$, $\delta>0$ such that:

$$I_{h,f}(z)=\mathrm{GE}_{p,k,h,f}(z)+FR_{k+1}(h,f,z)$$

where

$$F=-\frac{f(z)}{f'(z)}$$

and $|R_{k+1}(h,f,z)|\leqq Ch^{k+1}\max(1, 1/h_1^k)$ for $0<h<\delta\min(1,h_1)$.

The number $h_1=h_1(f,z)$ is the radius of convergence of $f_z^{-1}((1-h)f(z))$ as a power series in h around zero.

Examples of GE iterations without modification are described in [S-SI] these include incremental Newton's method, kth order incremental Euler, and kth order Taylor's method for the solution of the differential equation $dz/dt=-f(z)/f'(z)$.

It is sometimes more convenient to express the GE iterations and the modifications in terms of the polynomials $\sigma_{f,z}$ introduced in [S-SI].

$$\sigma(w)=\sigma_{f,z}(w)=\sum \sigma_i w^i$$

where

$$\sigma_0=0, \quad \sigma_1=1, \quad \sigma_i=\frac{(-1)^{i+1}f^{(i)}(z)f^{i-1}(z)}{i!(f'(z))^i}.$$

We recall some basic facts about σ. Given the iteration $z'=I_{h,f}(z)$ write

$$I_{h,f}(z)=z+FR_{(h,f)}(z).$$

We frequently write $R(h,f,z)$ or just $R(h)$, $R(z)$ or R. By Taylor's formula (see [S-SI, § 1])

$$f(z')=f(z)(1-\sigma\circ R)$$

or

$$\frac{f(z')}{f(z)}=1-\sigma\circ R.$$

Thus $R_{(h,f)}(z)\in\sigma^{-1}(1-f(z')/f(z))$. σ is a polynomial of the same degree d as f, thus there are d points z' which give the same value for $f(z')/f(z)$. If $I_{h,f}(z)$ is continuous then at least for small values of h

$$I_{h,f}(z)=z+F\left(\sigma^{-1}\left(1-\frac{f(z')}{f(z)}\right)\right)$$

where σ^{-1} is the branch of the inverse of σ taking 0 to 0. The radius of convergence of σ^{-1} is h_1. Comparing coefficients of powers of h (see [S-SI]) shows that

$$\mathrm{GE}_{p,k,h,f}(z)=z+FT_k(\sigma^{-1}(P(h))).$$

Consequently we may restate Definition 1.

DEFINITION 1'. $z'=I_{h,f}(z)$ is a GEM_k iteration iff there is a polynomial $P(h)$ and constants $c>0$, $\delta>0$ such that

$$I_{h,f}(z)=z+F(T_k(\sigma^{-1}(P(h))+R_{k+1}(h,f,z))$$

where $|R_{k+1}(h,f,z)|\leqq ch^{k+1}\max(1, 1/h_1^k)$ for $0<h<\delta\min(1,h_1)$.

One may also express the Generalized Euler iterations in the source as follows:

PROPOSITION 1. [S-SI]. *There are universal polynomials* $P_j=P_j(\sigma_1,\cdots,\sigma_{j+1}, c_1,\cdots,c_{j+1})$ *such that given any Generalized Euler iteration (without modification)*:

$$\mathrm{GE}_{p,k,h,f}(z)=z+F\sum_{j=0}^{k-1}P_jh^{j+1},$$

$$P_0=c_1,$$

$$P_1=c_2-\sigma_1c_1^2,$$

$$P_2=c_3-2\sigma_2c_1c_2-(\sigma_3-2\sigma_2^2)c_1^3.$$

One can write down P_j *explicitly inductively, in terms of* P_i, $i<j$, σ_{j+1} *and* c_{j+1}.

Proof. The coefficients of σ^{-1} are computed from the σ_i and $P_j(c,\sigma)$ is defined by

$$\sum_{j=0}^{k-1}P_j(c,\sigma)h^{j+1}=T_k\left(\sum_{l=1}^{\infty}\frac{(\sigma^{-1})^{(l)}}{l!}P(h)^l\right).$$

The GEM_k iterations may be similarly written with the addition of a remainder term. Examples of Generalized Euler with Modification are the simple Runge-Kutta approximation to the solution of $dz/dt=-f(z)/f'(z)$ (see [S-SI]) and an incremental Laguerre method. For the latter do the following. Instead of letting $R=T_k\sigma^{-1}(h)$ as for Euler we let $(T_k\sigma\circ R)=h$ and solve for R. We can do this for $k=2$. That is we solve $\sigma_2R^2+R=h$ for R yielding $R=(-1+\sqrt{1+4\sigma_2h})/2\sigma_2$. For $h=1$ this is Laguerre's method of Henrici (1977, p. 53) with $\gamma=2$.

We now turn to an analysis of the GEM iterations with the goal of showing that GEM's are cheap, Theorem C.

Given $P(h)$ there is a $\delta>0$ such that P is injective on the disc of radius δ, $D(\delta)$. Let K be the Lipschitz constant of P on $D(\delta)$, $K=\sup_{z\in D(\delta)}|P'(z)|$. Let $\gamma_k(c_1)$ be the first positive root of

$$(1-\gamma)^2-4c_1\gamma(1+B(k+1)\gamma^k)$$

where $1\leqq B<1.07$ and B is a constant which makes the Bieberbach conjecture true (see [S-SI]). Finally, let

$$\gamma_k(P)=\min\left(\delta,\frac{1}{K},\gamma_k(c_1)\right).$$

LEMMA 1. *Suppose that* $0<h_*\leqq\min(1,h_1(f,z))$ *and that* $h=ah_*$ *for some complex number* a *with* $|a|=\gamma$ *and* $0<\gamma<\gamma_k(P)$. *Then*:

a) $|\sigma^{-1}P(h)|\leqq c_1\gamma h_*/(1-\gamma)^2$;

b) $|\sigma^{-1}P(h)-T_k\sigma^{-1}P(h)|\leqq c_1h_*B(k+1)\gamma^{k+1}/(1-\gamma)^2$;

c) $|T_k(\sigma^{-1}P(h))|<h_*/4$;

d) $T_k(\sigma^{-1}P(h)\in\sigma^{-1}(D(h_*))$.

Here $D(h)$ *is the disc of radius* h *around* 0.

Proof. Since $h_*\leqq(1/K)h_1$ $P(D(h_*))\subset D(h_1)$ and since $h_*<\delta h_1$ $\sigma^{-1}\circ P$ is defined and injective on $D(h_*)$. $(\sigma^{-1}\circ P)_0'=c_1$ so $(1/c_1)(\sigma^{-1}\circ P)$ is defined and injective on $D(h_*)$. It has derivative 1 at 0. Now [S-SI, Lemma 7, § 2] applied to $(1/c_1)(\sigma^{-1}P)$ proves a) and b). By the triangle inequality

$$|T_k(\sigma^{-1}(P(h))|\leqq\frac{c_1\gamma h_*}{(1-\gamma)^2}+\frac{c_1h_*B(k+1)\gamma^{k+1}}{(1-\gamma)^2}$$

and the right-hand side is $<h_*/4$ since $\gamma<\gamma_k(c_1)$. This proves c). Since $|T_k\sigma^{-1}(P(h)|<h_*/4$, [S-SI, Lemma 5, § 2] proves d).

LEMMA 2. *Let* $I_{h,f}(z)=z+F(z)R_{(h,f)}(z)$ *be a* GEM *iteration then there is a* δ *such that*

$$|R_{(h,f)}z)|<\frac{h_1(f,z)}{4}$$

for $0<h<\delta\min(1,h_1(f,z))$.

Proof. There is a $P(h)$ and R_{k+1} such that $R_{h,f}=T_k(\sigma^{-1}(P(h))+R_{k+1}$. Now use Lemma 1c with $h_*=\frac{1}{2}\min(1,h_1)$ and further choose $\delta<\gamma_k(P)$ small enough so that $|R_{k+1}|<h_1/8$ for $0<h<\delta\min(1,h_1)$.

LEMMA 3. *Let* $I_{h,f}(z)=z+F(z)R_{(h,f)}(z)$ *be a* GEM_k *iteration. Then there is a* $\delta>0$ *such that for* $0<h<\delta\min(1,h_1)$, $\sigma^{-1}(\sigma\circ R)=R$.

Proof. By the Koebe theorem

$$\sigma^{-1}(D(h_1))\supset D\left(\frac{h_1}{4}\right)$$

where $D(r)$ denotes the open disc of radius r around 0 in $\mathbb{C}$. By Lemma 2 there is a δ such that

$$|R_{(h,f)}(z)| < \frac{h_1(f,z)}{4}$$

for $0 < h < \delta \min(1, h_1(f,z))$. Thus $R_{(h,f)}(z) = \sigma^{-1}(h')$ for some $h' \in D(h_1)$ $\sigma \circ R_{(h,f)}(z) = h'$ and $\sigma^{-1}(\sigma \circ R_{(h,f)}(z)) = \sigma^{-1}(h') = R_{(h,f)}(z)$.

PROPOSITION 2. *If $z' = I_{h,f}(z) = z + F(T_K\sigma^{-1}(P(h)) + R_{k+1}(h))$ is a* GEM$_k$ *iteration, then there are constant $K > 0$, $\varepsilon > 0$ such that*

$$(*) \qquad \frac{f(z')}{f(z)} = 1 - P(h) + S_{k+1}(h)$$

where $|S_{k+1}(h)| < Kh^{k+1} \max(1, 1/h_1^k)$ for $0 < h < \varepsilon \min(1, h_1)$.

Proof.

$$\frac{f(z')}{f(z)} = 1 - \sigma \circ R = 1 - \sigma(T_k\sigma^{-1}P(h) + R_{k+1}(h)).$$

There are K, $\delta > 0$ such that for $0 \leqq h < \delta \min(1, h_1)$

$$S_{k+1}(h) = \sigma(\sigma^{-1}(P(h)) - \sigma(T_k\sigma^{-1}P(h) + R_{k+1}(h))$$

and

$$|\sigma^{-1}(P(h))| < 2c_1 h,$$

$$|\sigma^{-1}P(h) - (T_k\sigma^{-1}(P(h)) + R_{k+1}(h))| < Kh^{k+1} \max\left(1, \frac{1}{h_1^k}\right)$$

by Lemma 1, the definition of $R_{k+1}(h)$ and the triangle inequality. Now by the generalized Loewner theorem (see Smale (1981) say) $|\sigma_i|^{1/i-1} < 4/h_1$. Now apply [S-SI, Lemma 3, § 2] with $a = 4/h_1$, $b = 2c_1 h$, $c = (k/2c_1)h^{k+1} \max(1, 1/h_1^k)$. Make sure that δ is small enough to guarantee $(1+c)ab < 1$.

Iterations satisfying $(*)$ were called efficiency k in [S-SI]. One of the fundamental estimates on speed is proven for iterations of efficiency k [SS-I, Thm. 4]. Since a polynomial f of degree d is generally d to 1 a point z' is only determined by $f(z')$ up to this d to 1 ambiguity. Efficiency k iterations are determined by the ratios $f(z')/f(z)$ and thus are determined up to this same ambiguity. We impose a continuity condition on iterations.

DEFINITION. The iteration $I_{h,f}(z) = z + FR(h, f, z)$ is called small iff there is a $\delta > 0$ such that $\sigma^{-1}(\sigma(R)) = R$ for all $0 < h < \delta \min(1, h_1)$.

Here, as above, σ^{-1} takes 0 to 0 and is defined and convergent on the disc of radius $h_1 = h_1(x, z)$.

THEOREM D. *The small iterations of efficiency k are precisely the* GEM$_k$ *iterations.*

Proof. Proposition 2 proves that the GEM$_k$ are efficiency k and Lemma 3 that the GEM$_k$ are small. Now to prove the converse. If

$$z' = I_{h,f}(z) = z + FR$$

and

$$\frac{f(z')}{f(z)} = 1 - (P(h) + S_{k+1}(h))$$

where $|S_{k+1}(h)| < Kh^{k+1} \max(1, 1/h_1^k)$ for $0 < h < \varepsilon \min(1, 1/h_1^k)$ then

$$\sigma \circ R = P(h) + S_{k+1}(h);$$

and there is a $0 < \delta < \varepsilon$ such that

$$R = \sigma^{-1}(P(h) + S_{k+1}(h))$$

for $0 < h < \delta \min(1, h_1)$ since I is small. Thus it only remains to prove that for

$$R_{k+1} = \sigma^{-1}(P(h) + S_{k+1}(h)) - T_k \sigma^{-1} P(h),$$

there are $c > 0$, $\delta > 0$ s.t.

$$|R_{k+1}| < ch^{k+1} \max\left(1, \frac{1}{h_1^k}\right)$$

for $0 < h < \delta \min(1, h_1)$.

$$|R_{k+1}(h)| \leqq |\sigma^{-1}(P(h) + S_{k+1}(h)) - \sigma^{-1}(P(h))| + |\sigma^{-1}(P(h)) - T_k \sigma^{-1}(P(h))|.$$

We estimate the two terms on the right. For $h_* = \min(1, h_1)$ and $h = \gamma h_*$ with $\gamma < \gamma_k(P)$ Lemma 1b asserts that

$$|\sigma^{-1}(P(h)) - T_k \sigma^{-1}(P(h))| \leqq \frac{c_1 h_* B(k+1)\gamma^{k+1}}{(1-\gamma)^2}$$

so for $0 < h < \gamma_k(P)/2 \min(1, h_1)$

$$|\sigma^{-1}(P(h)) - T_k \sigma^{-1}(P(h))| \leqq 4c_1 \frac{h^{k+1}}{h_*^k} = 4c_1 h^{k+1} \max\left(1, \frac{1}{h_1^k}\right).$$

Now estimate

$$|\sigma^{-1}(P(h) + S_{k+1}(h)) - \sigma^{-1}(P(h))|$$

as in Proposition 2.

$$|\sigma_i^{-1}|^{1/i-1} < \frac{(Bi)^{1/i-1}}{h_1} < \frac{4}{h_1}.$$

For $0 < h < (\gamma_k(P)/2) \min(1, h_1) = (\gamma_k(P)/2) h_*$,

$$|P(h)| < \frac{1}{\gamma_k(P)} h \quad \text{and} \quad |\sigma^{-1} P(h))| < 4c_1 h$$

by Lemma 1a and there are δ_1, K such that

$$|S_{k+1}(h)| < Kh^{k+1} \max\left(1, \frac{1}{h_1^k}\right) = K \frac{h^{k+1}}{h_*^k}$$

for $0 < h < \delta_1 \min(1, h_1)$.

Thus by [S-SI, Lemma 3]

$$|\sigma^{-1}(P(h) + S_{k+1}(h)) - \sigma^{-1}(P(h))|$$

$$\leqq \frac{Kh^{k+1}/h_*^k}{(1 - (4/\gamma_k(P))h/h_1)(1 - (1 + K\gamma_k(P)(h/h_*)^k)((4/\gamma_k(P))h/h_1))}.$$

Now it is easy to produce a δ such that for $0 < h < \delta h_*$, K divided by the denominator is bounded.

LEMMA 4. *Suppose* $|f(z_1)|\leqq|f(z)|$ *and that* $z_1=f_z^{-1}((1-h)f(z))$ *for* $0<|h|<\sin\Theta_{f,z_0}$, *then*

$$\Theta_{f,z_1}\geqq\Theta_{f,z}-\left|\arg\frac{f(z_1)}{f(z)}\right|.$$

Proof. Since $|f(z_1)|<|f(z)|$ it suffices to see that $z_1\in f_z^{-1}(W_{f,z})$ see [SS-I, § 3] for then $f_{z_1}^{-1}$ is defined on a wedge of angle at least the angle of $W_{f,z}$ minus $|\arg(f(z_1)/f(z))|$.

But the open disc of radius $|f(z)|\sin\Theta_{f,z}$ centered at $f(z)$ is contained in $W_{f,z}$ as Fig. 3 shows. Thus $(1-h)f(z)$ is in this disc for $0\leqq|h|<\sin\theta_{f,z}$ and $z_1\in f_z^{-1}(W_{f,z})$.

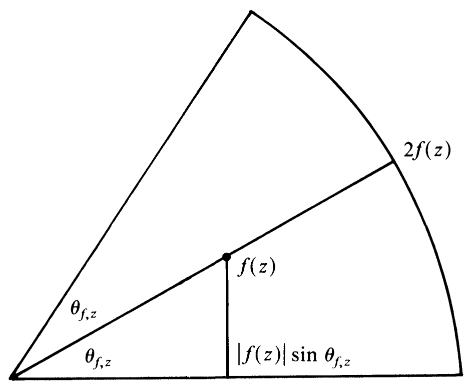

FIG. 3

LEMMA 5. *Let* $I_{h,f}(z)$ *be a* GEM_k *iteration. Then there is a constant* a, $1\geqq a>0$ *depending only on I such that: If* $0<h<a\sin\Theta_{f,z}$, *then*

$$\Theta_{f,z'}\geqq\Theta_{f,z}-\left|\arg\frac{f(z')}{f(z)}\right].$$

Proof. By Lemma 4 we need only show that there is an a such that if $0<h<a\sin\Theta_{f,z}$ then $z'=f_z^{-1}((1-h)f(z))$ for some h with $0<|h|<\sin\Theta_{f,z}$. Now $f_z^{-1}((1-h)f(z))=z+F\sigma^{-1}(h)$. Thus it suffices to show for $z'=I_{h,f}(z)=z+FR$ that $R(h,f,z)=\sigma^{-1}(h')$ for some h' with $0<|h'|<\sin\Theta_{f,z}$. Since a GEM_k iteration is small, there is a $\delta_1>0$ such that for $0<h<\delta_1\min(1,h_1)$, $\sigma^{-1}(\sigma R(h,f,z))=R$. We will let $h'=\sigma R(h,f,z)$. Since $\sin\Theta_{f,z}<\min(1,h_1)$ we know that $\delta_1\sin\Theta_{f,z}<\delta_1\min(1,h_1)$. Now we claim that there is an a, $0<a\leqq\delta_1$ such that if $0<h<a\sin\Theta_{f,z}$ then $|h'|=|\sigma R(h,f,z)|<\sin\Theta_{f,z}$. The last claim finishes the proof, for then $h'=\sigma R(h,f,z)$ satisfies $0<|h'|<\sin\Theta_{f,z}$ and $\sigma^{-1}(h')=R(h,f,z)$. To verify the claim we may assume $\delta_1\leqq 1$. Let L be the Lipschitz constant of $P(h)$ on $D(1)$ $\sigma R(h,f,z)=P(h)+S_{k+1}(h)$ where C, $\mu>0$ with the property that

$$|S_{k+1}(h)|<Ch^{k+1}\max\left(1,\frac{1}{h_1^k}\right)\quad\text{for } 0<h<\mu\min(1,h_1),$$

$$\max\left(1,\frac{1}{h_1^k}\right)=\frac{1}{\min(1,h_1)^k}<\frac{1}{\sin\Theta_{f,z}}.$$

Thus for $0<h<\mu\sin\theta_{1,z}$

$$|S_{k+1}(h)|<\mu^{k+1}\sin\theta_{f,z}.$$

We may assume $\mu < \min(1/2L, 1/2)$ for $0 < h < \mu \sin \Theta_{f,z} |P(h)| < \frac{1}{2} \sin \Theta_{f,z} |S_{k+1}(h)| < (\frac{1}{2})^{k+1} \sin \Theta_{f,z}$ and we are done. □

We now can prove that GEM's are cheap. This is the analogy of [S-SI, Theorem 4] but for GEM's Λ_{f,z_0} can be replaced by Θ_{f,z_0}.

THEOREM C. *Suppose that $z' = I_{h,f}(z)$ is a* GEM_k *iteration. Then there is a constant K depending only on I such that: If $\Theta_{f,z_0} > 0$ and $|f(z_0)| > L > 0$ then there is an h given explicitly such that $|f(z_n)| < L$ for*

$$n = \left[K \left(\frac{\log |f(z_0)/L|}{\theta_{f,z_0}} \right)^{(k+1)/k} \right]$$

and $z_n = (I_{h,f})^n(z_0)$.

Proof. The proof is the same as [S-SI, Thm. 4]. Take a to be the min of the a considered there and the a of Lemma 5 above, and replace Λ_{f,z_0} by Θ_{f,z_0}.

Remark. It is easy to adapt the algorithms of Theorem A and B to GEM_k iterations and prove analogous theorems. We state a particular theorem which is the analogue of the "Main Theorem" of [SS-I] and which has the same proof starting from Theorem C.

THEOREM E. *If $z' = I_{h,f}(z)$ is a* GEM_k *iteration, there are positive constants K_1, K_2 depending only on $P(h)$ with the following true: Given $d > 1$, $1 > \mu > 0$ there are R, h such that: If $(z_0, f) \in S_R^1 \times P_d(1)$ then $z_s = (I_{h,f})^s(z_0)$ will be an approximate zero of f with probability $1 - \mu$ for any $s \geqq K_1(d(|\log \mu|/\mu))^{(k+1)/k} + K_2$.*

4. This section consists of a series of problems and remarks.

(1) There is the general problem of comparing speeds of different algorithms for root finding of complex polynomials. See Dejon-Henrici (1969) and Henrici (1977) for a number of such algorithms. Comparison by experiment on machines can often be done readily. But the theoretical study of which algorithms are faster is another story. Such an analysis must deal seriously with round off errors and eventually consideration of the question: What are appropriate models of algorithms for this problem? Perhaps even the question must be confronted; what is the best model for the machine for this kind of study? Is a Turing machine always the right model? See the Traub-Wozniakowski (1982) critique of Khachiyan's work on linear programming for a good perspective on related questions.

For the comparison of algorithms lower bounds on speed are important. See Traub-Wozniakowski (1980).

Finally in this discussion, we note that we have just received the paper of Schönhage (1982). This takes a different approach to the study of speed related to the fundamental theorem of algebra. It seems to be quite an interesting work.

(2) We only have results for one variable. It is a wide open and central problem to extend the analysis to more than one variable, especially to polynomial maps from $\mathbb{C}^n$ to $\mathbb{C}^n$ (or $\mathbb{R}^n$ to $\mathbb{R}^n$). Since the theory of schlicht functions, used heavily here, is a one variable theory, finding the right estimates in several variables is a challenging problem. Relevant to this problem is that a mild algebraic condition on a polynomial map $f: \mathbb{C}^n \to \mathbb{C}^n$ guarantees that f is proper and hence has a zero, see Hirsch-Smale (1979).

(3) We have studied the problem of efficiency for finding one zero of a polynomial. What about the problem of finding all zeros of a polynomial f? The natural algorithms would follow paths in the space $I \times \mathbb{C}$, $I = [0, 1]$. Define $f_d(z) = \prod_{i=1}^d (z - \zeta_i)$ where $\zeta_1, \cdots, \zeta_d$ are the dth roots of unity. Let $F: I \times \mathbb{C} \to \mathbb{C}$ be defined by $F(t, z) = (1-t)f_d(z) + tf(z)$. Generally $F^{-1}(0) \subset I \times C$ consists of d curves leading from the

known ζ_i to zeros of f. One obtains good algorithms by following these curves simultaneously. We suspect that the speed in this case can be understood by the methods in these papers.

(4) We have used the space

$$P_d(1)=\{(a_0,\cdots,a_{d-1},a_d\}\,||a_i|\leqq 1,\ a_d=1\}$$

of coefficients of $f(z)=\sum_{i=0}^d a_iz^i$. While this parameterization has a simple immediacy, the projective space $\mathbb{C}^{d+1}/(\mathbb{C}-0)$ is more natural. Here $(a_0,\cdots,a_d)\in\mathbb{C}^{d+1}$ is equivalent to $(\lambda a_0,\cdots,\lambda a_d)$, each nonzero complex λ. It would seem reasonable for the results to go over. Also we have used one particular probability measure, the uniform one. It would be useful to make a generalization to a wide class of probability measures, say given axiomatically as in Smale (1982b). Finally the problem suggests itself to replace polynomials by other classes of comlex analytic functions. For example much of the analysis applies to rational functions which are also invertible up to the "first" critical value.

(5) We recall a problem, yet open, from Smale (1981) which has received some attention. For any polynomial f, $\deg f>1$, and complex number z, $f'(z)\neq 0$, it seems likely that there exists a critical point $\theta(f'(\theta)=0)$ such that

$$\left|\frac{f(z)-f(\theta)}{z-\theta}\right|\leqq|f'(z)|.$$

It is proved there that

$$\min_{\substack{\theta\\ f'(\theta)=0}}\left|\frac{f(z)-f(\theta)}{z-\theta}\right|\leqq K|f'(z)|\quad\text{with } K=4.$$

One can express the conjecture in a slightly sharper form by making K a function of d, $K=K_d=d-1/d$. This conjecture is the best possible as can be seen by choosing $f(z)=z^d-dz$, and $z=0$. The conjecture is false for entire functions such as $f(z)=e^z$.

Dick Palais first pointed out to Smale that the estimate $(K=1)$ was true for f with real zeros. Nan Boultbee confirmed Smale's early calculations for degree $f\leqq 4$. And recently David Tischler (1982) has proved the conjecture when one root of f is zero and the others have the same absolute value. Linda Keen and Tischler have produced some supportive numerical evidence, but as mentioned above, the general conjecture remains open.

(6) We remark that although detailed techniques are different, there are basic similarities between the main theorems here and in Smale (1982a). Each gives a good estimate for the average number of iterations of a well-known algorithm or variation thereof. Moreover the underlying geometry of the algorithm in each case is following the inverse image of a segment in the target space.

The work of Kuhn-Zeke-Senlin (1982) and Renegar (1982) on the speed of piecewise linear algorithms to find zeros of polynomials relates to both our paper here and Smale (1982a).

(7) The algorithms in [S-SI] and in this paper start with $z_0\in\mathbb{C}$ satisfying $|z_0|\gg 1$, e.g. $|z_0|=3$. This is necessary for our analysis since the rough behaviour of $f\in P_d(1)$ on points z_0 with $|z_0|$ large enough is independent of f. On the other hand the large starting value contributes eventually to the d in the estimates of the main theorems. This suggests that if one started with $|z_0|\leqq 1$, e.g. $z_0=0$ at least that factor of d would be eliminated, sharpening the theorem drastically.

The problem here is to obtain information on the behavior of $\Theta_{f,z}$ for $|z|=1$ or even $z=0$. Consider the integral Θ_d of $\Theta_{f,z}$ over the space $P_d(1)\times S_1^1$ with the usual uniform probability measure. Is there an $\varepsilon>0$ independent of d such that $\Theta_d>\varepsilon>0$? A related question is: Do there exists universal constants $\varepsilon_1, \varepsilon_2>0$ such that $\int_{f\in P_d(1)}$ (measure of $\{z\in S^1 | \Theta_{f,z}>\varepsilon_1\})^{-1}<\varepsilon_2^{-1}$. An affirmative answer would imply that the d in $d \log d$ of Theorems A and B could be eliminated.

In the case of Theorem A this is very direct; Theorem B actually requires a slightly different algorithm. The idea is to switch to $h=1$ at some point. We develop such an algorithm a bit.

Let $\tilde{\rho}_f=\min(\frac{1}{2}c, \frac{1}{2}c\rho_f^{3/2})$ and $j=j_k=\lceil \log_{k+1}(8d\log_2(d))\rceil$. Here c is about $\frac{1}{12}$ as in the definition of approximate zero. It is a simple computation to check from the proposition of the introduction that:

LEMMA. *Suppose* $f\in P_d(1)$ *and* $\varepsilon_f>0$. *If* $|f(z_0)|<\tilde{\rho}_f$ *then* $|f(E_k^j(z_0))|<\varepsilon_f$.

Proof. If $\rho_f\leqq 1$ then $|f(z_0)|<(\rho_f^{1/2}/2)c\rho_f$. Thus

$$|f(E_k^j(z_0))|<\left(\frac{\rho_f^{1/2}}{2}\right)^{8d\log_2 d} c\rho_f \leqq \frac{c\rho_f^d}{2^{8d\log_2 d}}$$

$$\leqq \frac{c\rho_f^d}{2^{4d}2^{4d\log_2 d}}=\frac{c\rho_f^d}{(2d)^{4d}}<\varepsilon f.$$

If $\rho_f\geqq 1$ then $|f(E_k^j(z_0)|<c/(2d)^{4d}<\varepsilon_f$ by a similar calculation.

We are now ready to describe the Algorithm (N$-$E)$'$:

0) $m=1$
1) $k=\max(\lceil \log d\rceil, m)$
   $j=\lceil \log_{k+1}(8d\log_2 d)\rceil$
   $n=K(d+m)$
2) Pick z_0 with $|z_0|=3$ at random.
3) $z=E^j_{k,l,f}(E^n_{k,H,f}z_0)$.

If $|f(z)|<\varepsilon_f$ terminate and print "z is an approximate zero of f".

4) If not $m=m+1$ and go to 1.

Once $2^{-m}<\tilde{\rho}_f$ all that is needed for this algorithm to produce an approximate zero is to pick a z_0 with $\Theta_{f,z_0}\geqq\pi/12$. So in the mean the algorithm goes through six additional cycles of the loop. Thus for fixed f with $\varepsilon_f>0$, Algorithm (N-E)$'$ terminates with probability one.

Let $m(f)=\lceil \log\tilde{\rho}_f\rceil$.

$$S(f)\leqq m(f)[K(d+m(f)+j_2]+\sum_{k=1}^{\infty}\frac{1}{6^k}(K(d+m(f)+k)+j_{m(f)+k})$$

$$\int_{f\in P_d(1)} m(f)<C_1\log d \quad \text{for } C_1 \text{ a constant.}$$

This shows that Theorem B applies to Algorithm (N-E)$'$. The extra computations involved in increasing k do not seriously effect the total number of arithmetic operations either.

REFERENCES

A. BORODIN AND I. MUNRO, (1975), *The Computational Complexity of Algebraic and Numerical Problems*, American Elsevier, New York.

R. BRENT AND KUNG (1976), *$O((n \log n)^{3/2})$ algorithms for composition and reversion of power series*, in Analytic Computational Complexity, J. F. Traub, ed., Academic Press, New York, pp. 217-225.

——— (1978), *Fast algorithms for manipulating formal series*, J. Assoc. Comput. Mach., 25, pp. 581-595.

B. DEJON AND P. HENRICI (1969), *Constructive Aspects of the Fundamental Theorem of Algebra*, John Wiley, New York.

P. HENRICI 1977, *Applied and Computational Complex Analysis*, John Wiley, New York.

M. HIRSCH AND S. SMALE (1979), *On algorithms for solving $f(x)=0$*, Comm. Pure Appl. Math., 32, pp. 281-312.

D. KNUTH (1981), *The Art of Computer Programming, Vol. 2, Seminumerical Algorithms*, 2nd ed., Addison-Wesley, Reading, MA.

H. KUHN, W. ZEKE AND X. SENLIN (1982), *On the cost of computing roots of polynomials*, preprint.

S. LANG (1965), *Algebra*, Addison-Wesley, Reading MA.

J. RENEGAR (1982), *On the complexity of a piecewise linear algorithm for approximating roots of complex polynomials.*

A. SCHÖNHAGE (1982), *The fundamental theorem of algebra in terms of computational complexity-preliminary report*, Math. Inst. der Univ. Tübingen.

M. SHUB AND S. SMALE (1982), *Computational complexity; on the geometry of polynomials and a theory of cost* (*Part* I), Ann. Scient. Ec. Norm. Sup. 4 serie, 18 (1985).

S. SMALE (1981), *The fundamental theorem of algebra and complexity theory*, Bull. Amer. Math. Soc., (New Series), 4, pp. 1-36.

——— (1982a), *On the average speed of the simplex method of linear programming*, preprint.

——— (1982b), *The problem of the average speed of the simplex method*, to appear in the Proceedings of the International Symposium of Mathematical Programming, Bonn, 1982, Springer, Heidelberg.

D. TISCHLER (1982), preprint.

J. TRAUB AND H. WOZNIAKOWSKI (1980), *A General Theory of Optimal Algorithms*, Academic Press, New York.

——— (1982), *Complexity of linear programming*, Oper. Res. Lett., 1, pp. 59-62.