On the Average Cost of Solving Polynomial Equations

by Mike Shub and Steve Smale*

Given a complex polynomial $f(z) = z^d + a_{d-1} z^{d-1} + \ldots + a_0$ and a complex number z_0 we can attempt to find a root of f by locally inverting f at $f(z_0)$, to $f_{z_0}^{-1}$ taking $f(z_0)$ to z_0, and analytically continuing $f_{z_0}^{-1}$ along the ray from $f(z_0)$ to 0. This process usually works (see Smale for a discussion of this and the history) and defines a curve leading from z_0 to a root ζ of f. This curve is given by

$$f_{z_0}^{-1}((1-h)f(z_0))$$

where $0 \leq h \leq 1$. We may think of h as a complex variable and then $f_{z_0}^{-1}((1-h)f(z_0))$ is an analytic function. Let $h_1 = h_1(f, z_0)$ be the radius of convergence of this function considered as a power series around h=0. If $h_1 > 1$ then substituting h=1 in the power series expansion of $f_{z_0}^{-1}((1-h)f(z_0))$ gives a power series expansion for ζ. But h_1 may not be bigger than one, and even if it is it is not usually computationally practical to evaluate an infinite series, so we truncate this series at some finite power of h. Let

$$E_k(z_0) = E_{k(h,f)}(z_0) = t_k f_{z_0}^{-1}((1-h)f(z_0))$$

where t_k stands for truncation of a power series at degree k

$(t_k(\sum_{i=0}^{\infty} a_i \cdot z^i) = \sum_{i=0}^{k} a_i \cdot z^i)$. Thus $z_0 \to E_{k(h,f)}(z_0)$ for $0 \leq h$ is a curve which has $k^{\underline{th}}$ order contact with $f_{z_0}^{-1}((1-h)f(z_0))$ at z_0. Our plan is to iterate E_k for small values of h obtaining $z_n = E_k(z_{n-1})$ which stay close to the curve $f_{z_0}^{-1}((1-h)f(z_0))$ and approach the root ζ. As in Smale we rely heavily on the theory of shlicht functions. A full detailed account of the work described here may be found in Shub-Smale which is in

The power series expansion of ζ was given by Euler and he iterated E_k with h=1 and $1 \le k \le 5$ to solve equations see Euler. Thus we call our iterative algorithms, $k^{\underline{th}}$ incremental Euler.

It is not difficult to compute that

$$E_{4(h,f)}(z) = z - \frac{f(z)}{f'(z)}\,(h - \sigma_2 h^2 + (2\sigma_2{}^2 - \sigma_3)h^3 - (5\sigma_2{}^3 - 5\sigma_2\sigma_3 + \sigma_4)h^4).$$

where $\sigma_i = (-1)^{i-1}\dfrac{f^{(i)}(z)\,f^{i-1}(z)}{i!(f'(z))^i}$. By keeping only the first k powers of **h**, k = 1,2,3 we obtain E_k for those values. In particular E_1 is just incremental Newton which was studied in Smale.

The first main result is the following:

Theorem 1 There is a universal constant B, $1 \le B \le 1.07$ such that for any polynomial f, complex number z with $f'(z) \neq 0$, $f(z) \neq 0$, $z' = E_{k(h,f)}(z)$

$$(*) \qquad \frac{f(z')}{f(z)} = 1 - h + Q(h,f,z)\,\frac{h^{k+1}}{h_1{}^k} \quad \text{and}$$

$|Q| \le \beta_k(\gamma)$ where

$$\beta_k(\gamma) = \frac{B(k+1)(1-\gamma)^2}{\left[(1-\gamma)^2 - 4\gamma\right]\left[(1-\gamma)^2 - 4\gamma(1+B(k+1)\gamma^k\right]}$$

Here $\gamma = \dfrac{h}{h_1}$ and is assumed to satisfy $0 < \gamma < \gamma_k$ where γ_k is the first positive number for which the denominator of $\beta_k(\gamma)$ vanishes.

Computation shows that γ_1 is approximately $\frac{1}{7}$ and that γ_k increases rapidly to $3 - \sqrt{8}$ which is approximately $\frac{1}{6}$. Thus the range of applicability of the estimate is around $\frac{h_1}{7}$.

We may rewrite (*) $\dfrac{f(z')}{f(z)} = 1 - h + T(h,f,z)h$

where $|T| \le \beta_k(\gamma)\gamma^k \equiv \alpha_k(\gamma)$.

Thus to make sure that $|f(z')| < |f(z)|$ we want $\alpha_k(\gamma) < 1$. Computation of $\alpha_k'(\gamma)$ shows that $\alpha_k(\gamma)$ is increasing. Thus there is a unique $\bar{\gamma}_k$, $0 < \bar{\gamma}_k < \gamma_k$ such that $\alpha_k(\bar{\gamma}_k) = 1$ and $\alpha_k(\gamma) < 1$ for $0 < \gamma < \bar{\gamma}_k$.

Theorem 1 admits the following Corollary. Let $\rho_f = \min\limits_{\substack{\theta \\ f'(\theta)=0}} |f(\theta)|$.

<u>Corollary</u>: Let $k > 0$. Suppose a polynomial f and a complex number z satisfy

$$|f(z)| = b\left(\frac{\bar{\gamma}_k}{1+\bar{\gamma}_k}\right)\rho_f \text{ for some } b < 1.$$

Then with $h = 1$, $(E_k)^{\ell}(z) = z_\ell$ is defined for all ℓ and $z_\ell \to z^*$ as $\ell \to \infty$ with $f(z^*) = 0$. Moreover, $|f(z_\ell)| \leq c|f(z_{\ell-1})|^{k+1}$ all $\ell > 0$ ("$(k+1)^{\underline{st}}$ order convergence") with $c = \left(\frac{b}{|f(z_0)|}\right)^k$.

Finally

$$|f(z_\ell)| \leq b^{((k+1)^\ell)} \frac{\bar{\gamma}_k}{1+\bar{\gamma}_k} \rho_f .$$

We will call z an <u>approximate zero</u> of f <u>relative</u> to <u>k</u> if $|f(z)| < \frac{\bar{\gamma}_k}{1+\bar{\gamma}_k}\rho_f$. An approximate zero of f relative to all $k > 0$ will simply be called an <u>approximate zero</u> of f

$$\frac{\bar{\gamma}_k}{1+\bar{\gamma}_k} \text{ increases with } k.$$

$$\frac{1}{12} > \frac{\bar{\gamma}_1}{1+\bar{\gamma}_1}, \quad \frac{\bar{\gamma}_k}{1+\bar{\gamma}_k} > \frac{1}{7} \text{ for } k \geq 5 \text{ and } \frac{\bar{\gamma}_k}{1+\bar{\gamma}_k} < \frac{1}{6} \text{ for all } k.$$

Thus if $|f(z)| < \frac{\rho_f}{12}$, z is an approximate zero of f.

<u>Our goal is to find approximate zeros of f!</u>

For $0 < \alpha \leq \frac{\pi}{2}$ let $w_{f,z,\alpha} = \{w \in C \,|\, |w| \leq 2|f(z)|,\ |\arg \frac{w}{f(z)}| < \alpha\}$

Define $w_{f,z}$ be the largest of the $w_{f,z,\alpha}$ to which f_z^{-1} may be analytically continued and $\theta_{f,z}$ to be the corresponding α. Since the $w_{f,z,\alpha}$ are simply connected, this definition makes sense.

We navigate in this wedge using Theorem 1 and its corollary.

$$\text{Let } K(k) = \frac{(k+1)^{\frac{k+1}{k}}}{k\,\bar{\gamma}_k\,(1-\bar{\gamma}_k)^{\frac{1}{k}}}$$

Theorem 2: Suppose given a polynomial f and a complex number z_0 such that $|f(z_0)| > \rho_f > L > 0$ and $\Theta_{f,z_0} > 0$. Let $C = \frac{1}{\Theta_{f,z_0}} \log \frac{|f(z_0)|}{L}$.

Then there is an $h_0 \geq \frac{\sin \Theta_{f,z_0}}{K(k)(c+1)^{\frac{1}{k}}}$ with this property. For each $0 < h \leq h_0$

and any $s \geq \frac{1}{h}\left(\log \frac{|f(z_0)|}{L} + \frac{\Theta_{f,z_0}}{k+1}\right)$

$$|f(z_n)| < L$$

K(k) decrease to $\frac{1}{3-\sqrt{8}}$ which is less than 6.

k =	1	2	3	4	5	6	7	8	9	10
K(k) <	47.1	21.1	14.7	12.1	10.7	9.9	9.3	8.9	8.6	8.4

Smale gives a better result for Newton's method than the K(1) achieved here.

Thus the problem of finding approximate zeros becomes the problem of controlling ρ_f and Θ_{f,z_0}.

This is done in the next two propositions

Let $P_d(1) = \{f \mid f(z) = z^d + a_{d-1} z^{d-1} + \ldots + a_0 \text{ where } |a_i| \leq 1,\ i=0,\ldots, d-1\}$

Proposition 1 (Smale)

$$\text{Vol } \{f \in p_d(1) \mid \rho_f < \alpha\} \leq (d-1)\alpha^2$$

where Vol means normalized Lebesgue measure.

Let S_r^1 be the circle of radius r around 0 in C.

Proposition 2 For $\alpha < \frac{\Pi}{}$ $f \in P_d(1)$

The last ingedient in the theory is a checking procedure. Let $R(f,\frac{f'}{d})$ be the resultant of f and $\frac{f'}{d}$ see Lang. $R(f,\frac{f'}{d})$ is computable from the coefficients of f alone.

Let $L(f) = \frac{1}{24} \min (1,\rho_f^{\frac{7}{6}})$

Lemma Let $f \in P_d(1)$, then

a) $$\rho_f \geq \frac{R(f,\frac{f'}{d})}{(d+3)^{2d-1}}$$

b) If $|f(z)| < L(f)$ then

$$|f(E^{\ell}_{k(1,f)}(z))| < \frac{\bar{\gamma}}{1+\bar{\gamma}} \frac{R(f,\frac{f'}{d})}{(d+3)^{2d-1}}$$

for $\ell = |\lceil \log_{k+1}(3d \log d)\rceil| \equiv \ell(k,d)$.

a) is a fairly naive estimate and b) follows from the Corollary.

The use of the lemma is the following.

If we suspect that

$|f(z)| < L(f)$ then

$$|f\ E^{\ell}_k(z)| < \frac{\frac{\bar{\gamma}}{1+\bar{\gamma}} R(f,\frac{f'}{d})}{(d+3)^{2d-1}}$$ which is checkable by direct calculation. If this last inequality holds then $E^{\ell}_r(z)$ is an approximate zero of f.

We use the results above to prove the following theorems. Let μ be normalized Lebesgue measure on $S^1_r \times P_d(1)$.

Theorem 3 Given $0 < \mu < 1$, $d > 1$ there is an iterative algorithm $E = E_{(k(\mu;d),h(\mu,d))}$ and an $r = r_{(\mu,d)}$ such that:

for (z_0,f) in $S^1_r \times P_d(1)$ and $s = L_1 d(\frac{|\log \mu|}{\mu})^{1+\frac{1}{\lceil \log d\rceil}} + L_2$;

$z = E^s(z_0)$ is an approximate zero for f with probability $1 - \mu$.

The exponents in Theorem 3 improve considerably on those of the Main Theorem of Smale. Yet it seems paradoxical that part of the improvement is achieved by making $r_{(\mu,d)}$ increase like $\frac{1}{\mu}$, which contributes the factor d in the theorem. It would seem more sensible to pick r small, close to 1. It is an excellent problem to carry out the analysis in this case; it becomes more difficult.

By attempting to find a "good" starting point z_0 for each given f we prove "average" theorems for iterative algorithms based on the $k^{\underline{th}}$ incremental Euler algorithms. By "average" we mean the integral with respect to normalized Lebesgue measure on $P_d(1)$.

Theorem 4: There are probabilistic and deterministic iterative algorithms for finding approximate zeros for f $P_d(1)$, with the average number of steps required $O(d)$ and $O(d^2 \log d)$ respectively.

The algorithms in Theorem 3 and 4 may be executed in $O(d \log d)$ arithmetic operations for each step in exact arithmetic with log and real $k^{\underline{th}}$ roots. They are robust.

References

L. Euler, Institutiones Calculi Differentialis, exp IX, Opera Omnia, série I, vol. X, pp. 422-455.

S. Lang, Albegra, Addison Wesley, Reading, Mass.

M. Shub and S. Smale, Computational Complexity: On the geometry of polynomials and a theory of cost: Part I and II, to appear.

S. Smale, 1981, The Fundamental Theorem of Algebra and Complexity Theory, Bull. Amer. Math. Soc.,Vol. 4, No. 1 pp. 36.