STUDENT ARTICLE

5

Does Every Polynomial Root Have a Simple Approximation?

Bryan Gin–ge Chen '07†
Harvard University
Cambridge, MA 02138
bryangingechen@gmail.com

Abstract

The practice of neglecting small terms of an equation is analyzed in the case of polynomial root approximations. Our discussion centers on the following new result: The roots of a polynomial can be approximated self-consistently by roots of much simpler equations consisting of pairs of terms from the polynomial.

5.1 Introduction

The essence of mathematical modeling is to take a real–world question and translate it into a mathematical problem which can then be solved, yielding insight into the original question. In the course of such modeling, approximations must invariably be made.

It is not an exaggeration to say that all important equations take the schematic form

$$t_1(p_1, p_2, \dots) + t_2(p_1, p_2, \dots) + \cdots = 0,$$

where t_n are arbitrary terms and p_m are arbitrary parameters.

One commonly used approximation simplifies these equations by choosing some subset of terms deemed to be the most important and then neglecting all the others. If we choose only the two largest, we end up with

$$t_i(p_1, p_2, \dots) + t_j(p_1, p_2, \dots) \approx 0.$$

We will refer to such an approximation as a **dominant balance approximation**, since it seeks a solution which "balances" the two dominant terms against each other. The question we are interested in is: How often can solutions to an equation be approximated by the behavior of a few dominant terms?

We investigated this question for the case of polynomial equations of arbitrary order in a single variable.

†Bryan Gin-ge Chen, Harvard '07, is a physics concentrator in Adams House. He hails from Centerville, Ohio, where he attended Centerville High School. His mathematical interests extend freely to all that is unreasonably effective in the natural sciences, including calculus and linear algebra, scaling and similarity solutions, topology, and symmetry.

5.2 Example: $a_5x^5 + a_1x + a_0 = 0$

When are the roots of this quintic trinomial dominated by the behavior of just two of the terms? In other words, when are the roots of the equations

$$a_5x^5 + a_1x = 0$$
$$a_5x^5 + a_0 = 0$$
$$a_1x + a_0 = 0$$

close to the actual roots?

Note that if the roots of these equations are relatively close to the true roots, then we have indeed found simple solutions to the trinomial equation — these approximations depend on only two coefficients!

We say that a choice of terms yields a **self-consistent** root when those terms are larger (in absolute value) than any other terms when we set x equal to the root of the simplified equation. For instance, the roots of the equation for which we say that a_5x^5 and a_1x are dominant are $x = 0$ or $x = \left(-\frac{a_1}{a_5}\right)^{1/4}$ (all four of these 4th roots). We ignore the $x = 0$ root as this is obviously inconsistent with our assumption that a_5x^5 and a_1x are largest near the root. Therefore the self-consistency condition is

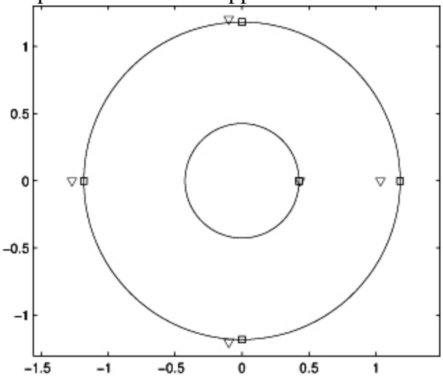
$$\left| a_5 \left(-\frac{a_1}{a_5}\right)^{5/4} \right| = \left| a_1 \left(-\frac{a_1}{a_5}\right)^{1/4} \right| > |a_0|,$$

or equivalently $\left|\frac{a_0^4 a_5}{a_1^5}\right| < 1$.

Working out the other cases yields the following result:

$$\left|\frac{a_0^4 a_5}{a_1^5}\right| < 1 \iff (a_1x, a_5x^5) \text{ and } (a_0, a_1x) \text{ yield self-consistent roots,}$$
$$\left|\frac{a_0^4 a_5}{a_1^5}\right| > 1 \iff (a_0, a_5x^5) \text{ yields self-consistent roots.}$$

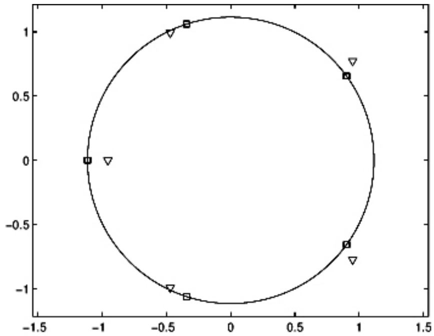
We plot true roots and approximate roots of one from each "class" of trinomial below:



$.832x^5 - 1.62x + .692 = 0,\ \left|\frac{a_0^4 a_5}{a_1^5}\right| = .0169$

true roots: .434, 1.04, -1.27, -.0993$\pm$1.20i;

approx. roots: .426, $\pm$1.18, $\pm$1.18i.



$.690x^5 + .669x + 1.19 = 0,\ \left|\frac{a_0^4 a_5}{a_1^5}\right| = 10.4$

true roots: -.956, -.471$\pm -$ .991i, .949$\pm$.773i;

approx. roots: -1.12, -.345$\pm$1.06i, .902$\pm$.656i.

These examples were generated using MATLAB with coefficients drawn from a normal distribution centered at zero with unit variance.

5.3 Generalization to Arbitrary Polynomials

It is possible to generalize the ideas in the previous column and prove the following result:

Theorem 1. *Given a nondegenerate*[1] *polynomial* $a_n x^n + a_{n_1} x^{n_1} + a_{n_2} x^{n_2} + \cdots + a_{n_p} x^{n_p} + a_0$ *where* $n = n_0 > n_1 > n_2 > \cdots > n_p > n_{p+1} = 0$*, all self-consistent approximations to the roots of this polynomial will come from pairs* $\{(a_n x^n,\ a_{n_{j_1}} x^{n_{j_1}}),\ (a_{n_{j_1}} x^{n_{j_1}},\ a_{n_{j_2}} x^{n_{j_2}}),\ \cdots,\ (a_{n_{j_s}} x^{n_{j_s}},\ a_0)\}$*, where* $\{n_{j_k}\}$ *is a subsequence of the* $\{n_l\}$ *above.*

The pairs of terms which give the self-consistent approximations can be bracketed together as below:

$$a_n x^n + a_{n_1} x^{n_1} + \ldots + a_{n_{j_1}} x^{n_{j_1}} + \ldots + a_{n_{j_s}} x^{n_{j_s}} + \ldots + a_{n_p} x^{n_p} + a_0$$

Theorem 1 essentially states that this series of brackets will not cross itself, and will reach from the $a_n x^n$ term to a_0. Note that since a pair $(a_j x^j,\ a_k x^k)$ yields $k - j$ different approximate roots (via the $\frac{1}{k-j}$th roots of $-\frac{a_k}{a_j}$), the total number of self-consistent approximate roots is guaranteed to be $(n_{j_s} - 0) + (n_{j_{s-1}} - n_{j_s}) + \cdots + (n_{j_2} - n_{j_1}) + (n - n_{j_1})$ which telescopes to n.

The nondegeneracy condition in Theorem 1 excludes polynomials which have pairs of terms which are not completely dominant at the approximate roots. In our quintic trinomial example, the degenerate polynomials would be those for which $\left|\frac{a_0^4 a_5}{a_1^5}\right| = 1$. This condition gives us the set of polynomials for which all three terms are equally large when evaluated at the approximate roots.

Note that Theorem 1 does not say anything about the accuracy of these self-consistent approximations — it merely states that they exist.

Sketch of proof. The key 'trick' is to transform the self-consistency inequalities by taking logarithms. In the quintic trinomial example, letting $A_j = \log |a_j|$, we have:

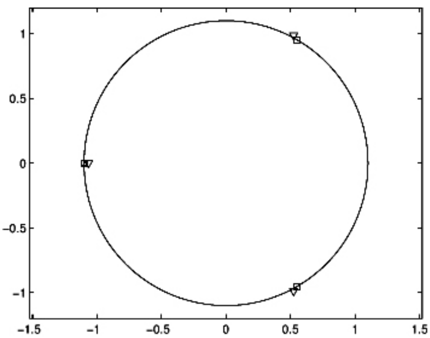
$$A_0 + A_2 - 2A_1 < 0 \Rightarrow (A_0, A_1, A_2) \cdot (1, -2, 1) < 0,$$
$$A_0 + A_2 - 2A_1 > 0 \Rightarrow (A_0, A_1, A_2) \cdot (1, -2, 1) > 0.$$

Here $\cdot$ is the ordinary dot product. The two classes have become half–spaces in $\mathbb{R}^3$.In the general case, the classes are cones in $\mathbb{R}^{j-2}$ (where j is the number of terms) defined by a set of dot–product inequalities. Using a 1958 result due to Samelson, Thrall and Wesler [STW], we can show that these cones partition all of $\mathbb{R}^{j-2}$.

5.4 Self-Consistent Approximation Picture Gallery

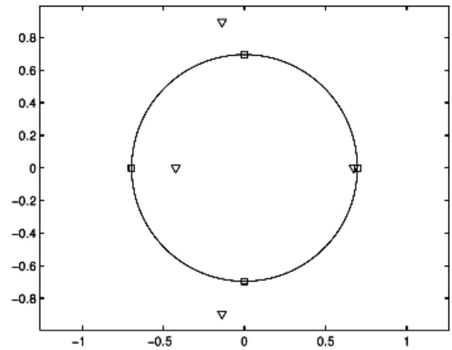
The following polynomials were generated in MATLAB by choosing coefficients from a normal distribution with unit variance. The approximate roots were found by an algorithm based on Theorem 1. The values ϵ are quantities analogous to $\left|\frac{a_0^4 a_5}{a_1^5}\right|$ in the example, but are now chosen so that a root is self-consistent if $\epsilon < 1$ for all ϵ.

[1] For the purposes of Theorem 1, a polynomial $a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0$ is **degenerate** if the vector $(\log |a_n|, \log |a_{n-1}|, \ldots, \log |a_0|)$ is in the linear space spanned by $(1, 1, \ldots, 1)$ and $(n, n-1, \ldots, 0)$.



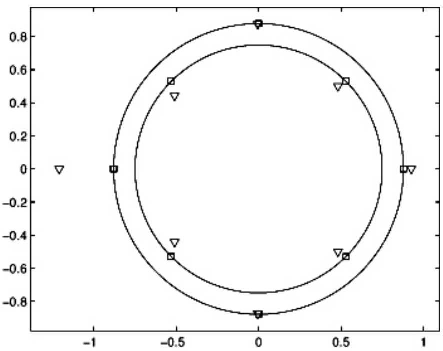
$1.20x^3 + .0198x^2 + .157x + 1.60 = 0$
$\epsilon_j = \{1.24 \times 10^{-3}, 3.34 \times 10^{-6}\}$
true roots: -1.07, .525$\pm .987i$;

approx. roots: -1.10, .550$\pm .953i$.



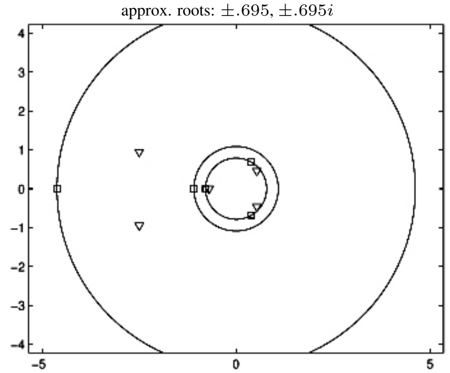
$-2.17x^4 + .0592x^3 - 1.01x^2 + .614x + .508 = 0$
$\epsilon_j = \{.501, .859, 2.36 \times 10^{-6}\}$
true roots: -.424, .673, -.138$\pm .895i$;

approx. roots: $\pm .695$, $\pm .695i$



8th order polynomial equation
$\max \epsilon_j = .0779, \min \epsilon_j = 1.07 \times 10^{-16}$

selected true roots: .927, -1.21, -.509$\pm .442i$; closest approx. roots: .879, -.879, -.530$\pm .530i$



$.262x^5 + 1.21x^4 + 1.32x^3 - .931x^2 - .0112x + .645 = 0$
$\epsilon_j = \{2.59 \times 10^{-6}, .719, .38, .235\}$

true roots: -.695, .534$\pm .459i$, -2.50$\pm .948i$; approx. roots: -1.09, .394$\pm .682i$, -4.62, -7.88.

The approximate roots we chose in this way are in general quite close to the actual roots. However, in the fourth plot there is a pair of complex roots that is approximated by a pair of real roots. It appears that partitioning using Theorem 1 does not place some polynomials correctly.

5.5 Self-Consistency Is Not Enough

Though the self-consistency condition gives us a simple criterion for choosing dominant terms, the choices do not always yield good approximations. The reason is that the self-consistency condition completely ignores the possibility of multiple roots.

If we apply Theorem 1 to the quadratic equation $a_2x^2 + a_1x + a_0 = 0$, we find:

$$\left|\frac{a_0a_2}{a_1^2}\right| < 1 \Longleftrightarrow (a_1x, a_2x^2) \text{ and } (a_0, a_1x) \text{ yield self-consistent roots,}$$
$$\left|\frac{a_0a_2}{a_1^2}\right| > 1 \Longleftrightarrow (a_0, a_2x^2) \text{ yields self-consistent roots.}$$

Thus the degenerate, borderline case is when $\left|\frac{a_0a_2}{a_1^2}\right| = 1$. However, by the quadratic formula

$x = \frac{-a_1 \pm \sqrt{a_1^2 - 4a_0a_2}}{2a_2}$, there is a multiple root when $\frac{a_0a_2}{a_1^2} = \frac{1}{4}$, so this should be the borderline case.

5.6 Series Expansions for Roots

Let us take a different approach now. Let our polynomial be $\sum_{j=0}^{n} a_j x^j = 0$. Choose two terms $a_k x^k$ and $a_j x^j$. We now rescale by $x \mapsto y \left(-\frac{a_j}{a_k}\right)^{\frac{1}{k-j}}$, which yields (after division by a simplifying factor):

$$\frac{a_n \left(-\frac{a_j}{a_k}\right)^{\frac{n}{k-j}}}{\left(-\frac{a_j^k}{a_k^j}\right)^{\frac{1}{k-j}}} y^n + \cdots + y^k + \cdots - y^j + \cdots + \frac{a_0}{\left(-\frac{a_j^k}{a_k^j}\right)^{\frac{1}{k-j}}} = 0.$$

We can rewrite this as

$$c_n y^n + c_{n-1} y^{n-1} + \cdots + y^k + \cdots - y^j + \cdots + c_0 = 0,$$

where for all $l \neq k, j$, $c_l = \left((-a_j)^{l-k} a_k^{j-l} a_l^{k-j}\right)^{\frac{1}{k-j}}$. It is clear that if the $n-1$ coefficients c_l are all sufficiently small, then $y \approx 1^{\frac{1}{k-j}}$ and $x \approx \left(-\frac{a_j}{a_k}\right)^{\frac{1}{k-j}}$. But what do we mean by sufficiently small? We can write a series solution for y by assuming an ansatz of the form

$$y_{(j,k)} = \sum_{s_0, s_2, \ldots, s_n = 0}^{\infty} A_{s_0, s_1, \ldots, s_n} c_n^{s_n} c_{n-1}^{s_{n-1}} \cdots c_0^{s_0},$$

where we have $n-1$ quantities c_l and s_l (no c_j, c_k or s_j, s_k). It makes sense to say that the terms are sufficiently small if this series converges. It is thus more natural to say that the pair $a_k x^k$, $a_j x^j$ is dominant at the root if the series for $y_{(j,k)}$ converges (rather than using the self-consistency conditions)!

The self-consistency conditions are equivalent to requiring that all $|c_l| < 1$; however, the domain of convergence of this series is in general a more complicated object.

Series solutions of polynomials can be written in terms of hypergeometric functions, but the domains of convergence are only known in some cases; see [St, PT].

5.7 What If We Iterate?

Suppose that we have some method of choosing pairs of terms which gives us approximate roots from dominant balances. Note that

$$f(z) = \frac{f^{(n)}(0)}{n!} z^n + \frac{f^{(n-1)}(0)}{(n-1)!} z^{n-1} + \cdots + f(0),$$

where $f^{(m)}$ is the m-th derivative of f.

The approximate root we get by assuming that the j-th and k-th terms are dominant is then

$$\left(-\frac{f^{(j)}(0)}{f^{(k)}(0)}(k-j)!\right)^{\frac{1}{k-j}}.$$

To improve on this root, instead of deriving the next term of a series (as in the previous section), consider $f(u - z_1)$ where $z_1 = \left(-\frac{f^{(j)}(0)}{f^{(k)}(0)}(k-j)!\right)^{\frac{1}{k-j}}$. Note that for any z_1, the roots of $f(z)$ are precisely $u + z_1$ for the roots u of the polynomial $f(u + z_1)$. Using Taylor's theorem, we have

$$f(u + z_1) = \frac{f^{(n)}(z_1)}{n!} u^n + \frac{f^{(n-1)}(z_1)}{(n-1)!} u^{n-1} + \cdots + f(z_1).$$

If the terms proportional to u^p and u^q (with $q > p$) are dominant terms of this polynomial, an approximate root of this polynomial will be $u_2 = \left(-\frac{f^{(p)}(z_1)}{f^{(q)}(z_1)}(q-p)!\right)^{\frac{1}{q-p}}$, which gives us

$$z_2 = u_2 + z_1 = z_1 + \left(-\frac{f^{(p)}(z_1)}{f^{(q)}(z_1)}(q-p)!\right)^{\frac{1}{q-p}}.$$

In general, there are multiple choices of pairs of terms which will give us dominant terms. Furthermore, each of these $(k-j)$th root expressions above will have $k-j$ different solutions — this shows that this process will **branch**. If we keep iterating, we have the recursive function:

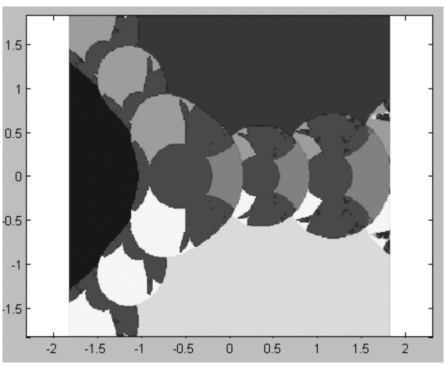
$$z_{m+1} = z_m + \left(-\frac{f^{(p_m)}(z_m)}{f^{(q_m)}(z_m)}(q_m - p_m)!\right)^{\frac{1}{q_m - p_m}}.$$

Note that if $q_m = 1$ and $p_m = 0$ for all m, this process no longer branches. In fact, we now have

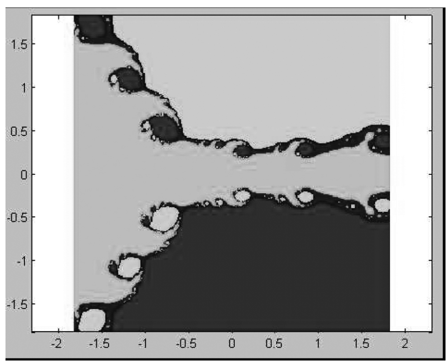
$$z_{m+1} = z_m - \frac{f(z_m)}{f'(z_m)},$$

which is precisely Newton's Method! Thus we may interpret Newton's method as an iterated dominant balance method which always assumes that the 1st order and 0th order terms dominate, or rather, we might interpret the iteration of a dominant balance method as a branching version of Newton's method.

If we color each point in $\mathbb{C}$ according to which point it converges to upon iterating Newton's method, we produce the Newton fractal. Similarly, with a branching algorithm, we can color the points of $\mathbb{C}$ according to which set of points we obtain. Below, we compare the "Newton fractals" of Newton's method, and iterating the self-consistent root algorithm based on Theorem 1.



This image shades each point in the plane according to which roots it goes to on iteration of the self-consistent roots method.



This image shades each point in the plane according to which roots it goes to after iterating Newton's method.

5.8 Conclusion

The results described in this paper give evidence in one case for a fact which equation–solvers have known intuitively for a long time: the solutions to equations are often determined in a large part by the behavior of the largest terms in the equation. Are there similar results for other equations — in particular, ordinary differential equations or partial differential equations?

Acknowledgments

I thank Professor Michael Brenner for introducing me to this problem and advising me throughout my work. Thanks to Takuya Kitagawa '08 for helpful discussions as well.

References

[PT] Mikael Passare and August Tsikh: Algebraic equations and Hypergeometric Series, *The Legacy of Niels Henrik Abel* (New York: Springer, 2004), 563–582.

[STW] Hans Samelson, R.M. Thrall and Oscar Wesler: A partition theorem for Euclidean n–space, *Proc. of the AMS* **9** (1958), 805–807.

[St] Bernd Sturmfels: Solving algebraic equations in terms of $\mathcal{A}$–hypergeometric series, *Discrete Math.* **210** (2000), 171–181.