International Journal of Mathematical Analysis and
Optimization: Theory and Applications
Vol. 2020, No. 1, pp. 757 - 763

A Visualizable, Constructive Proof of the Fundamental Theorem of Algebra, and a Parallel Polynomial Root Estimation Algorithm

Christopher Thron [1*], Jordan Barry [2]

[1*], 2, Texas A& M University-Central Texas 79549 USA.
*Corresponding author: thron@tamuct.edu

Article Info

Received: 16 June 2020 Revised: 4 August 2020
Accepted: 24 August 2020 Available online: 27 August 2020

Abstract

This paper presents an alternative proof of the Fundamental Theorem of Algebra that has several distinct advantages. The proof is based on simple ideas involving continuity and differentiation. Visual software demonstrations can be used to convey the gist of the proof. A rigorous version of the proof can be developed using only single-variable calculus and basic properties of complex numbers, but the technical details are somewhat involved. In order to facilitate the reader's intuitive grasp of the proof, we first present the main points of the argument, which can be illustrated by computer experiments. Next we fill in some of the details, using single-variable calculus. Finally, we give a numerical procedure for finding all roots of an nth degree polynomial by solving $2n$ differential equations in parallel.

Keywords And Phrases: Fundamental Theorem of Algebra, calculus, chain rule, continuity.
MSC2010: 60E05; 30C15,65H04

1 Introduction

The fundamental theorem of algebra states that any polynomial function from the complex numbers to the complex numbers with complex coefficients has at least one root. There are several proofs of the fundamental theorem of algebra, which employ a number of different domains of mathematics, including complex analysis (Liouville's theorem, Cauchy's integral theorem or the mean value property) ([1][2], topology (Brouwer's fixed point theorem)[3], differential topology[4], calculus ([1],[5]), and "elementary" methods using meshes or lattices [6], [7]. For easily-accessible and readable web references that explain these proofs, see [8],[9],[10],[11]. Many of these proofs are beautiful and elegant. Most are not constructive and do not provide a practical method for finding roots ([6] and [7] are exceptions). The proof we present here is both constructive, and provides a practical method for finding all roots of any polynomial through the numerical solution of differential equations with different initial conditions. Furthermore, we have created a simple, intuitive visual display that demonstrates the construction of roots.

http://ijmao.unilag.edu.ng/article/view/274

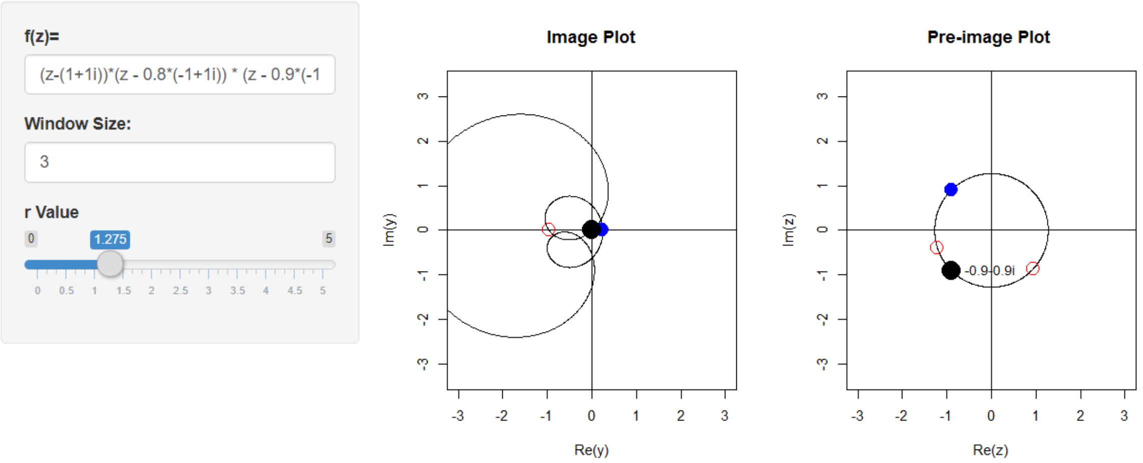
2 Empirical observations from computer experiments

Consider the polynomial $f(z) = \sum_{n=0}^{N} a_n z^n$, where N is a positive integer and a_n are complex. We want to show that $f(z) = 0$ has at least one complex solution. To approach this problem, we make some preliminary empirical observations on the behavior of polynomial functions.

First we may consider how $f(z)$ behaves for some specific values of z. When $z = 0$ we have $f(z) = a_0$, and when z has a very large magnitude then the terms $a_n z^n$ in $f(z)$ also have large magnitudes, especially the leading-order term $a_N z^N$. To understand the behavior of $f(z)$ between these two extremes, we isolate the behavior of $f(z)$ for different values of $|z|$, as described below.

A complex number can be written in polar form as $z = re^{i\theta}$, where $r > 0$ is the magnitude of z and $0 \leq \theta < 2\pi$. If we fix r and allow θ to vary, then the set of points $\{re^{i\theta}, 0 \leq \theta < 2\pi\}$ is a circle of radius r in the complex plane, which we denote as C_r. Since the function f is defined on all complex numbers, in particular it is defined on each circle C_r. The image of C_r under f is also a set (a curve, actually) in the complex plane, which we may denote as $f(C_r)$.

Using computer software, we may investigate the changes in the shape of $f(C_r)$ as r increases from 0, for different polynomials $f(z)$. For this purpose, an R Shiny code (listed in the Appendix) was developed that displays $f(C_r)$ and C_r as well as upcrossings and downcrossings, for any given value of r for an arbitrary polynomial with complex coefficients, as specified by the user. A screenshot of the interface is shown in Figure 1. A sequence of $f(C_r)$ plots for different values of r is shown in Figure 2.



f(z)=
(z-(1+1i))*(z - 0.8*(-1+1i)) * (z - 0.9*(-1

Window Size:
3

r Value
1.275

Figure 1: R Shiny interface for dynamic display of $f(C_r)$ and its preimage C_r, which shows upcrossings (solid blue dots) and downcrossings (hollow red dots). The function in this case is a cubic with roots at $1 + i$, $-0.8 + 0.8i$, and $-0.9 - 0.9i$. The large black dot in the preimage plot is the root $-0.9 - 0.9i$, which maps to 0 in the image plot. The R Shiny app used to create these plots is available online at https://github.com/jthomasbarry/complex_plot_r.

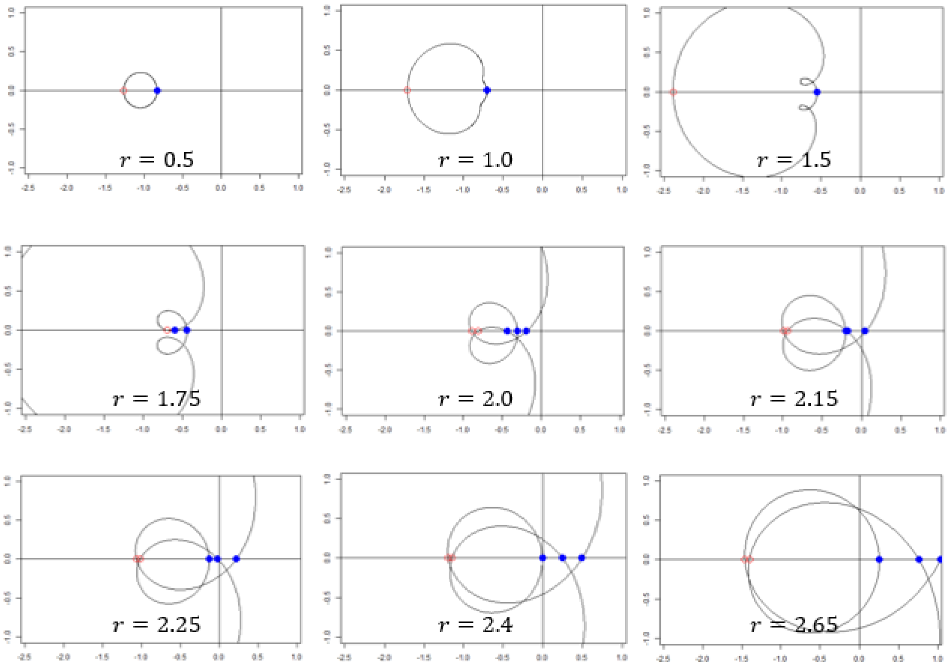


Figure 2: Curves $f(C_r)$ for different values of r, for the polynomial $f(z) = (a_1a_2a_3)^{-1}(z-a_1)(z-a_2)(z-a_3)$ where $a_1 = 1.6(1+i), a_2 = 1.7(-1+i), a_3 = 1.5(-1-i)$. The solid blue dots indicate upcrossings (which move to the right), while the hollow red dots indicate downcrossings (which move to the left). A new downcrossing-upcrossing pair is introduced when $r \approx 1.75$ (as a lobe in the upper half plane expands down across the real axs) and another pair is introduced when $1.75 < r < 2$ (when a lobe in the lower half plane expands up across the real axis). Roots are found at upcrossings for moduli $r \approx 2.15, 2.25, 2.4$ (compare $|a_3| = 2.12, |a_1| = 2.26, |a_2| = 2.40$).

Without loss of generality we may assume $a_0 = -1$: given any polynomial with $a_0 \neq 0$ we may obtain a polynomial with the same roots and having constant coefficient -1 by dividing by $-a_0$. If we look at several different polynomials f and see how the curve $f(C_r)$ evolves as r increases, we may make the following observations:

(i) When r is sufficiently small, then $f(C_r)$ has a nearly circular shape with center -1 and small radius. The curve $f(C_r)$ has multiple intersections with the real axis.

(ii) As r increases, these points of intersection betwee $f(C_r)$ and the real axis move continuously along the real axis (although sometimes they disappear: see point (v) below)

(iii) There are two types of intersections: some move consistently to the right as r increases, and others move consistently to the left. When r is small, the rightmost intersection is always rightward-moving.

(iv) New intersections with the x axis may appear as r increases. From a geometrical viewpoint, these new intersections occur when a lobe of $f(C_r)$ located in the upper (resp. lower) half-plane shfts downward (resp. upward) as r increases so that it intersects the axis. These new intersections always appear first as a single point that splits into a left-moving and right-moving intersection as r increases.

(v) A right-moving intersection continues to move to the right unless it runs into a left-moving intersection, in which case both intersections may disappear. From the two-dimensional viewpoint, this occurs when a lobe of $f(C_r)$ that intersects the real axis moves above or below the axis.

(vi) When r is very large, the shape of $f(C_r)$ approaches a large circle centered at the origin. In particular, the rightmost intersection between $f(C_r)$ and the real axis is large and positive.

(vii) The rightmost intersection when r is small is always continuously connected to the rightmost intersection when r is large by a series of right-moving or left-moving intersections. Since the origin is on the real axis between these two intersections, the origin must be either a right-moving or left-moving intersection for some value of r.

To understand the difference between right-moving and left-moving intersections, we may look more closely into the nature of the curve $f(C_r)$. As we mentioned above, the circle C_r is parametrized by the angle θ. As θ increases, the corresponding point on C_r (given by $re^{i\theta}$) moves counterclockwise around C_r, while the image of the point under the function f (given by $f(re^{i\theta}$ traces out the curve $f(C_r)$. As the tracing point crosses the real axis, we find there are two types of crossings: either upcrossings (from below to above), or downcrossings (from above to below). It may be observed experimentally (and we shall soon show mathematically) that the upcrossings correspond to the rightward-moving intersection points as noted above, and downcrossings correspond to leftward-moving intersection points.

We may summarize a systematic procedure for using the software to locate roots of f:

(I) Ensure that $a_0 = -1$ by dividing f by $-a_0$ (if $a_0 = 0$, then 0 is a root already);

(II) Start with a small value of r and locate an upcrossing point on C_r;

(III) Follow the upcrossing point as it moves rightward. Eventually it will either pass over the origin, or run into a leftward-moving downcrossing point and disappear.

(IV) If the latter holds, follow the leftward moving point backwards (i.e. decreasing r). Eventually, either it will pass over the origin, or it will merge with a rightward-moving point and disappear.

(V) Follow this rightward-moving point forward (increasing r) until it either passes through the origin or merges with a leftward-moving point.

(VI) Continue iterating Steps IV and V until the origin is reached.

3 Outline of a formal proof

This procedure is the basis for a formal proof of the theorem. Some of the technical detals are rather involved, but the guiding intuition is captured by the procedure described above.

The proof proceeds in several steps:

1. The roots of $f(z)$ are identical to the roots of $-f(z)/a_0$. So without loss of generality, we may assume that the constant coefficient a_0 is equal to -1.

2. Assume for the moment that $f'(z)$ has no zeros on the real axis between -1 and 0. (Later we will deal with the case where this is not true.)

3. For any value of $r > 0$, we define the curve $f_r(t) \equiv f(re^{it}), 0 \leq t \leq 2\pi$. Since $f_r(0) = f_r(2\pi)$, it follows that this is a closed (possibly self-intersecting) curve in the complex plane.

4. Denote by *upcrossing* (resp. *downcrossing*) a point where f_r crosses the real axis from below (resp. above). In other words, the real number x is an upcrossing for f_r if there exists t such that $f_r(t) = x$, and there exists $\delta > 0$ such that $\text{Im} f_r(s) \leq 0$ for $t - \delta \leq s \leq t$ and $\text{Im} f_r(s) \geq 0$ for $t \leq s \leq t + \delta$. By continuity, every root of $\text{Im} f_r$ is either an upcrossing, a downcrossing, or a point where the real line is tangent to f_r.

5. Suppose $x_r = f_r(t)$ is an upcrossing and ${f_r}'(t) \neq 0$, then the crossing point moves continuously to the right as a function of r. More precisely, there exist $\epsilon, \delta > 0$ and a continuous, real-valued function $g(s)$ that is strictly increasing on the interval $r - \epsilon < s < r + \epsilon$ such that $g(r) = x_r$ and $g(s) = f_s(t')$, for some t' in the interval $t - \delta < t' < t + \delta$. We call the function g an *upcrossing function.* A similar statement holds for the downcrossing case, except that g is decreasing: the function g in this case is called a *downcrossing function*).

6. The domain of any upcrossing function g may be extended to an open interval, such that either the range of g includes the origin, or the right endpoint b of the domain is such that $g(b)$ is a point of tangency of the curve f_b.

7. Every point of tangency that is the right endpoint of the domain of an upcrossing function is the left endpoint of the domain of a downcrossing function.

8. Every point in the interval $[-1, 0]$ is either an upcrossing, a downcrossing, or a point of tangency of f_r for some positive value of r. In particular, the origin is either an upcrossing, downcrossing, or point of tangency, and is thus equal to $f(re^{it})$ for some values of r and t.

Steps (1-8) handle the case where none of the roots of f' lie on the real segment $[-1, 0]$. If on the other hand f' does have a root on $[-1, 0]$, we may consider the ray $\theta = \pi + \nu$, for sufficiently small ν, which (by continuity) will have at least one upcrossing intersection with C_r if r, ν are sufficiently small. We denote this intersection as $\widetilde{z}$. Since the roots of f' are isolated, we can also choose the ν such that f' has no roots on the ray. We may then consider the function $\widetilde{f}(z) \equiv f(z)e^{-i\nu}$. Then $\widetilde{z}e^{-i\nu}$ is an upcrossing point for $\widetilde{f}$ of the negative real axis. Steps (1-8) above then goes through for $\widetilde{f}$: and the roots of $\widetilde{f}$ are identical with the roots of f.

4 Parallel numerical procedure for finding all roots of a polynomial

As above, we suppose $f(z) = \sum_{n=1}^{N} a_n z^n$ with $a_0 = -1$, where $z = re^{i\theta}$. We seek equations satisfied by upcrossing locations x as a function of r. Note that in order for $x = f(re^{i\theta})$ to be an upcrossing, the complex argument θ varies as r varies, so we must consider both θ and x as functions of r. To make this clear, we will use $\phi = \phi(r)$ to denote the complex argument, so that $x(r) = f(z(r))$ where $z(r) = re^{i\phi}$.

Using the chain rule, we have:

$$\frac{dx}{dr} = f'(z) \frac{d}{dr}(z) = f'(z)\, e^{i\phi} \left(1 + ir\frac{d\phi}{dr}\right) \tag{4.1}$$

Since x is real-valued, so $\frac{dx}{dr}$ is also a real function and $\frac{dx}{dr} = \left(\frac{dx}{dr}\right)^*$, where * denotes complex conjugate. This gives

$$f'(z)\, e^{i\phi} \left(1 + ir\frac{d\phi}{dr}\right) = f'(z)^*\, e^{-i\phi} \left(1 - ir\frac{d\phi}{dr}\right). \tag{4.2}$$

Solving for $r\frac{d\phi}{dr}$, we obtain

$$r\frac{d\phi}{dr} = \frac{\left(f'\left(z\right)^{*}e^{-i\phi} - f'\left(z\right)e^{i\phi}\right)}{i(f'\left(z\right)e^{i\phi} + f'\left(z\right)^{*}e^{-i\phi})} = \frac{-\operatorname{Im}\left(f'\left(re^{i\phi(r)}\right)e^{i\phi(r)}\right)}{\operatorname{Re}\left(f'\left(re^{i\phi(r)}\right)e^{i\phi(r)}\right)}, \tag{4.3}$$

where we have replaced z with $re^{i\phi(r)}$ to highlight the r dependence. From (4.1) and (4.3) we may calculate:

$$\frac{dx}{dr} = f'\left(z\right)e^{i\phi}\left(1 - i\frac{\operatorname{Im}(f'\left(z\right)e^{i\phi}}{\operatorname{Re}(f'\left(z\right)e^{i\phi}}\right) = \frac{|f'(re^{i\phi(r)})e^{i\phi(r)}|^2}{\operatorname{Re}\left(f'\left(re^{i\phi(r)}\right)e^{i\phi(r)}\right)} \tag{4.4}$$

Equations (4.3) and (4.4) express $\frac{d\phi}{dr}$ and $\frac{dx}{dr}$ respectively in terms of $f'\left(re^{i\phi(r)}\right)e^{i\phi(r)}$. Fortunately, this rather complicated expression turns out to have a relatively simple interpretation. The definition of f_r implies that $\frac{d}{d\theta}f_r(\theta) = f'(re^{i\theta})(ie^{i\theta})$, so if we define:

$$\alpha(r) \equiv \left.\frac{d}{d\theta}f_r(\theta)\right|_{\theta=\phi(r)} \tag{4.5}$$

then we may re-express the system (4.3)-(4.4) as:

$$\begin{aligned} \frac{d\phi}{dr} &= \frac{-\operatorname{Im}\left(-i\alpha(r)\right)}{r\operatorname{Re}\left(-i\alpha(r)\right)} = \frac{\operatorname{Re}\left(\alpha(r)\right)}{r\operatorname{Im}\left(\alpha(r)\right)}; \\ \frac{dx}{dr} &= \frac{|-i\alpha(r)|^2}{\operatorname{Re}\left(-i\alpha(r)\right)} = \frac{|\alpha(r)|^2}{\operatorname{Im}\left(\alpha(r)\right)}. \end{aligned} \tag{4.6}$$

For future reference, note that $\operatorname{Im}(\alpha(r))$ is positive or negative depending on whether $x(r)$ is an upcrossing or downcrossing.

Alternatively, we can pose the system such that θ is the independent variable, and r, x are the dependent variables. To clarify the dependence of r on θ, we use $\rho = \rho(\theta)$ here to represent the complex modulus as a function of θ, so that $f_\rho(\rho e^{i\theta})$ is an upcrossing point for the curve C_ρ. In analogy to (4.5), we define:

$$\beta(\theta) \equiv \left.\frac{d}{d\theta}f_r(\theta)\right|_{r=\rho(\theta)}, \tag{4.7}$$

and in analogy to (4.6) we obtain:

$$\begin{aligned} \frac{d\rho}{d\theta} &= \frac{-\operatorname{Re}\left(-i\beta(\theta)\right)}{\operatorname{Im}\left(-i\beta(\theta)\right)} = \frac{\operatorname{Im}\left(\beta(\theta)\right)}{\operatorname{Re}\left(\beta(\theta)\right)}; \\ \frac{dx}{d\theta} &= -\frac{|-i\beta(\theta)|^2}{\operatorname{Im}\left(-i\beta(\theta)\right)} = \frac{|\beta(\theta)|^2}{\operatorname{Re}\left(\beta(\theta)\right)}. \end{aligned} \tag{4.8}$$

It follows that given a crossing point $x = f_r(\theta)$, we can always make the crossing point 'move' continuously to the right by following this strategy:

1. If $\lim_{z\to re^{i\theta}} |\operatorname{Im} f'(re^{i\theta})/\operatorname{Re} f'(re^{i\theta})| > c$ (where $c < 1$ is a fixed positive parameter), then propagate x to the right using (4.6) with either increasing or decreasing r, depending on the sign of $\operatorname{Im} f'(re^{i\theta})$:

2. Otherwise, propagate x to the right using (4.8) with either increasing or decreasing θ, depending on the sign of $\operatorname{Im} f'(re^{i\theta})$.

Following this procedure will yield a monotonically increasing crossing point. If the initial crossing point is chosen such that it is chosen between -1 and 0 on the real axis, then eventually the crossing point will pass 0 and a root will be obtained.

The above procedure is only guaranteed to obtain a single root γ_1. Subsequent roots may be estimated by taking $f^{(1)}(z) \equiv f(z)(1 - z/\gamma_1)^{-1}$ and finding another root γ_2, then iterating the procedure with $f^{(j)}(z) \equiv f(z)(1 - z/\gamma_j)^{-1}, j = 1, 2, \ldots$ until all roots are found. However, it is possible there may be numerical stability problems, because due to numerical error $f^{(j)}(z)$ is no longer a polynomial for $j \geq 1$.

An alternative approach finds all roots in parallel as follows. If the polynomial has degree n, the z^n term dominates the behavior of $f(C_r)$ when r is large. It follows that for r sufficiently large, $f(C_r)$ must have at least n upcrossings on the positive real axis and at least n downcrossings of the negative real axis. All upcrossings may be followed leftwards using the reverse of the rightward-tracking procedure described above; and all downcrossings may be followed rightward by a similar procedure. Not all of these $2n$ tracks (which may be computed in parallel) will result in a root; however, it is guaranteed that all n roots will be obtained through the procedure.

References

[1] Schep, A.: A simple complex analysis and and advanced calculus proof of the fundamental theorem of algebra. American Mathematical Monthly 116(1), 67–68 (2009).

[2] Vyborny, R.: A simple proof of the fundamental theorem of algebra. Mathematica Bohemica 135(1), 57–61 (2010).

[3] Arnold, B.H: A topological proof of the fundamental theorem of algebra. The American Mathematical Monthly 56(7), 465–466 (1949).

[4] Guillemin, V., Pollack, A.: Differential Topology. 1st edn. Prentice-Hall (1947).

[5] Fefferman, C.: An easy proof of the fundamental theorem of algebra. The American Mathematical Monthly 74(7),854–855 (1967).

[6] Rosenbloom, P.C.: An elementary constructive proof of the fundamental theorem of algebra.The American Mathematical Monthly 52(10), 562–570 (1945).

[7] Brenner, J.L, Lyndon, R.C.: Proof of the fundamental theorem of algebra. American Mathematical Monthly 88(4), 253–256 (1981).

[8] File, D., Miller, S.: Fundamental theorem of algebra lecture notes from the reading classics (euler) working group autumn 2003, https://people.math.osu.edu/sinnott.1/ReadingClassics/FundThmAlg_DFile.pdf, last accessed 2020/2/1.

[9] Steed, M.: Proofs of the fundamental theorem of algebra, http://math.uchicago.edu/~may/REU2014/REUPapers/Steed.pdf, last accessed 2020/1/12.

[10] Linford, K.: An analysis of Charles Fefferman's proof of the fundamental theorem of algebra, http://commons.emich.edu/honors/504, last accessed 2020/2/1.

[11] Dunfield, N.: The fundamental theorem of algebra (class notes), https://faculty.math.illinois.edu/~nmd/classes/2015/418/notes/fund_thm_alg.pdf, last accessed 2020/1/15.