Elementary Aspects of the Fundamental Theorem of Algebra

Timothy W. Jones

October 31, 2019

Abstract

We motivate and give a proof of the fundamental theorem of algebra using high school algebra.

Introduction

How close can a typical high school student come to understanding the fundamental theorem of algebra? Currently some of the ingredients for a good understanding are present after a typical algebra 1, algebra 2, and pre-calc (or trigonometry) sequence, but the dots aren't connected. Thus students are familiar with quadratics and cubics and general polynomials, as well as Euler's and DeMoivre's formula and theorem; they are also told the fundamental theorem of algebra [1]; but, in no course are they encouraged to explore why, relative to their knowledge of algebra, the theorem might be true.

The standard proofs of the FTA use complex analysis. One proof uses Liouville's theorem and another Rouche's [3]; each of these theorems requires complex differentiation and integration and, hence, are not within the realm of high school algebra. But there is a proof given in Rudin [2] that doesn't reference these results. We make his proof simple enough that a good high school student could understand and appreciate it. There is no new proof here, exactly, except for one step (more of a gloss) that motivates the whole proof.

We will first show how high school topics can be re-glossed to make the theorem plausible. Then we will generalize and synthesize the plausible cases and find the pattern of interest. A proof of the FTA will be given.

There will be wholes in the argument. In the last part of the paper we will discuss these.

The FTA is plausible

How can you come to believe that all polynomials will have roots in the complex field? Well some examples where that is true will help. Consider

$$f(z) = z^n + a = 0. \tag{1}$$

(1) is a polynomial and we know its roots. They are

$$z = \sqrt[n]{-a} = \sqrt[n]{|-a|}(\cos\frac{\theta + 2\pi k}{n} + \sin\frac{\theta + 2\pi k}{n}),$$

where $0 \leq k \leq n-1$ and $-a = \cos\theta + i\sin\theta$. So at least one type of polynomial of degree arbitrary n can be solved with complex numbers. This result is just a re-gloss of DeMoivre's theorem: every number has nth roots.

If this is true, if we transform (1) shouldn't it remain true. Consider

$$f(z-b) = (z-b)^3 + a = 0. \tag{2}$$

Here we get a polynomial with all its terms via the binomial theorem. To wit

$$f(z) = (z-b)^3 + a = z^3 - 3bz^2 + 3b^2z - b^3 + a = 0.$$

But this can be solved with some algebra.

$$(z-b)^3 = -a \text{ implies } z - b = \sqrt[3]{-a} \text{ implies } z = \sqrt[3]{-a} + b.$$

Finding the key pattern

The quadratic case shows the essence of the problem of solving general polynomials. The graphs of x^2+1 indicates that there are no real zeros. The constant term has moved x^2 up by one and we know by the end behavior of polynomials, its parabola shape, no real value will move x^2+1 back to the x-axis for a root. But we also know x^2+1 can be solved with $\pm i$, complex numbers. This pattern is the same with the examples of the previous section: (1) and (2). Complex numbers enable a unsolvable polynomial in the real numbers to drop below their constants and reach the origin of the complex plane.

We can use this to construct a proof of the FTA.

Proof

After two easy lemmas, we give a proof of the FTA.

Lemma 1. *For every real non-zero number a there exists a real θ such that $ae^{i\theta} = -|a|$. This is also true for complex a.*

Proof. If $a < 0$ then let $\theta = 0$ and $ae^{0i} = -|a|$. If $a > 0$ then let $\theta = \pi$ and $ae^{\pi i} = -a = -|a|$.

For complex a, we just note $ae^{i\theta} = -|a|$ implies

$$e^{i\theta} = \frac{-|a|}{a}.$$

Not that it is necessary to note, we note $|-|a|/a| = 1$, so r in the polar coordinate representation of this number is 1. □

Lemma 2. *$|p(z)|$ can be made less than the absolute value of its constant.*

Proof. Cubic case: Let $p(z) = a_0 + a_1 z + a_2 z^2 + a_3 z^3$. We always will have a non-zero constant and here we assume a_1 is the first non-zero coefficient. It could be a_2 or a_3. The argument won't change. Then

$$p(re^{i\theta}) = a_0 + a_1 re^{i\theta} + a_2 r^2 e^{2i\theta} + a_3 r^3 e^{3i\theta}.$$

Using Lemma 1, there exists θ such that $a_1 e^{i\theta} = -|a_1|$. So now we have

$$|p(re^{i\theta})| = |a_0 - |a_1|r + a_2 r^2 e^{2i\theta} + a_3 r^3 e^{3i\theta}|$$

and taking the absolute value of the first constant term and the terms after $-|a_1|r$ increases the value of the right hand side. So

$$|p(re^{i\theta})| \leq |a_0| - |a_1|r + |a_2|r^2 + |a_3|r^3. \tag{3}$$

We've used $r > 0$ and $|e^{ik\theta}| = 1$. Rearranging (3),

$$|p(re^{i\theta})| \leq |a_0| - r\{|a_1| - |a_2|r^1 - |a_3|r^2\}.$$

Now for small enough r the value in the braces is positive, so the right hand side drops below $|a_0|$, as needed. □

Theorem 1. *If $p(z)$ is a polynomial it has complex root.*

Proof. Suppose $p(z)$ is a polynomial with no complex root. Then $|p(z)|$ has a non-zero minimum at a value, call it z_0. Consider the polynomial

$$Q(z) = \left|\frac{p(z + z_0)}{p(z_0)}\right|.$$

Then $Q(0) = 1$ and for all z, $Q(z) \geq 1$. But this says that the absolute value of a polynomial never is below the absolute value of its constant. This contradicts Lemma 2. □

Wholes in argument

The proof is simple, but we are assuming that the absolute value of a polynomial has a minimum. This in turn is dependent on polynomials being continuous functions on some closed set. This is easily understood for $p(x)$, p a polynomial, and x real. The closed intervals are of the form $[a, b]$, by inspection a pencil never is lifted off the page (the function is continuous), so there is a strong intuition that this result is true. But when one moves to complex arguments, what is the equivalent of a closed interval and can we assume $p(z)$ and $|p(z)|$ are continuous? We can't visually see maximums and minimums with $p(z)$ as we can in the real case $p(x)$.

Rudin fills in the gaps [2]. Here are quick statements of the theorems necessary. A theorem says that if $f(z)$ is continuous then $|f(z)|$ and $1/|f(z)|$ are too, provided in the latter case $f(z) \neq 0$ – which is good with our assumption that $p(z)$ has no roots. A bounded continuous real function on a closed interval reaches its minimum; i.e. there is a value in the domain that gives the minimum for the function. A function $1/|f(z)|$ is bounded below by 0, so there is a value which gives its minimum. The first statement in the proof are thus made good, but isn't all plausible enough to be stated to a student early on?

Conclusion

Should curious students wait for a course in real and complex analysis before getting a pretty good feel for why the FTA is true?

References

[1] R. Blitzer, *Algebra and Trigonometry*, 4th ed., Upper Saddle, NJ, 2010.

[2] W. Rudin, *Principles of Mathematical Analysis*, Wiley, New York, 1976.

[3] M. R. Spiegel, *Complex Variables: With an Introduction to Conformal Mapping and Its Applications*, McGraw Hill, New York, 1964.