EULER'S FORMULA AND THE FUNDAMENTAL THEOREM OF ALGEBRA

MATTHEW BOND

ABSTRACT. The Fundamental Theorem of Algebra states that each non-constant polynomial has at least one zero. It is often taught to young students with no hint whatsoever what its proof might be. To add insult to injury, they usually add a remark stating that Gauss proved it five different ways, leaving the student to meditate on how nice it would be to be smart like him.

Euler's Formula states the following: $e^{i\theta} = \cos(\theta) + i\sin(\theta)$. It is often given to students in a second Calculus course as a bit of trivia and a novel consequence of Taylor Series. In particular, it has a popular (but overrated) corollary: $e^{i\pi} + 1 = 0$. This was first popularized by textbooks, and then popularized by Justin Bieber.

The goal of this essay is to clearly detail a proof of the Fundamental Theorem of Algebra, showing how the bit of "trivia" called Euler's Formula is, in fact, one of the first and most fundamental facts about complex numbers. Without it, multiplication of complex numbers is difficult to visualize; with it, a proof of the FToA is within reach.

1. Some notations

The following notations are quite familiar to mathematicians, but are often not used in courses given before the undergraduate math major level.

:=

"[the thing on the left] is defined to be [the thing on the right]". Can also be read as part of a written sentence as "be defined to be", such as if I say, "Let $f(x) := x^2$"; that is, "Let $f(x)$ be defined to be x^2." It can be written backwards, such as $(x+1)^2 = x^2 + 2x + 1 =: g(x)$, or "$(x+1)^2 = x^2 + 2x + 1$, and let's call that expression $g(x)$."

$\mathbb{R}$:= the set of real numbers.

$\in$

This symbol means "is in", or more formally, "is an element of [the set named afterward]". $x \in \mathbb{R}$ means x belongs to the set of real numbers; that is, x is a real

number; $x \in [0,1]$ means $0 \leq x \leq 1$. It is also possible to have several objects separated by commas first, in which case all of them belong to the set that follows. That is, $a, b, c \in (0,1]$ means that a, b, and c are all real numbers which are both larger than 0 and at most 1.

$\{x : [\text{whatever}]\}$

"The set of x such that [whatever]". For example, $D = \{\text{weekdays} : \text{I don't work}\}$ is the set of weekdays such that I don't work; depending on my job, I could have $D = \{\text{Saturday, Sunday}\}$. One may also see $\{x \in A : [\text{whatever}]\}$, or "the set of x in A such that [whatever]". Here, $\in$ is read as "in" rather than "is in" or "is an element of", so that Thanksgiving $\in \{\text{days} \in \text{November} : \text{I don't work}\}$ says, "Thanksgiving belongs to the set of days in November on which I do not work."

$\subseteq$

"Is a subset of". $A \subseteq B$ means "A is a subset of B". It is possible that maybe $A = B$. Sometimes one writes $A \subset B$ to mean "A is a **proper** subset of B", meaning $A \subseteq B$ **and** we know that B has something that isn't in A. $\{1,2,3\} \subseteq \{1,2,3\}$ and $\{1,2\} \subseteq \{1,2,3\}$ are both correct statements, while $\{1,2,3\} \subset \{1,2,3\}$ is not. The extra line underneath, making $\subset$ into $\subseteq$, immitates visually the underline making $<$ into $\leq$, making it reasonable to read $\subseteq$ as "is a subset of or is equal to" or simply "subset or equal to".[1]

$i :=$ the imaginary number, $i = \sqrt{-1}$.

($-i$ is also **a** square root of -1, but as a matter of convention, one is free to refer to i as **the** square root of -1, for the sake of having a convention. Each is as good as the other.)

$\mathbb{C} :=$ the complex numbers $= \{a + bi : a, b \in \mathbb{R}\}$

Note that $\mathbb{R} \subset \mathbb{C}$, as the definition above allows one can have $b = 0$.

Remark: It is common for a student to assume that $\mathbb{C}$ **only** refers to those numbers that have a non-zero imaginary part, so that $1 + i$ is complex, while 3 is real, but not complex. That is, the student thinks of the complex numbers as the set $\mathbb{C}' := \{a + bi : a, b \in \mathbb{R} \text{ and } b \neq 0\}$. This is a matter of convention, but the student, in this case, is not using the generally accepted convention, which says that $1 + i \in \mathbb{C}$ and $3 \in \mathbb{C}$, so they are both complex numbers, while 3 is a real number, and $1 + i$ is not a real number. The more one works with complex

[1] Some authors prefer to have $A \subset B$ mean "A is a subset B", that is, they use it instead of $\subseteq$; in such cases, they may choose to use $A \subsetneq B$ to mean "A is a **proper** subset of B", which I denote by $\subset$.

numbers, the more it is convenient to adopt one of two stances: either (i) the real numbers are a special subset of the complex numbers, or (ii) the complex numbers do not exist at all. The second stance was popular before the mid-19th century, but has fallen out of fashion with working mathematicians since then. Students are justified in feeling strange about the complex numbers, but are likely to adopt the first stance if they study the subject for long enough.

$f : A \to B$

This means "f, a function from the set A to the set B". A is called the **domain** of f. The outputs of f do not have to take on every possible value of the set B; f is allowed to take on some, but not all, values of B. It is legitimate, for example, to write $f : \mathbb{R} \to \mathbb{R}$ with $f(x) := x^2$, even though $f(x)$ does not ever take on any of the negative values. It would be legitimate (and more informative) to speak of $f : \mathbb{R} \to [0, \infty)$. If we set the domain to be $\mathbb{C}$ instead of $\mathbb{R}$, we have a different function, in some formal sense, and we could write $g : \mathbb{C} \to \mathbb{C}$ as $g(z) = z^2$. It would be incorrect, then, to write that $g : \mathbb{C} \to \mathbb{R}$ or $g : \mathbb{C} \to [0, \infty)$, since for example, $g(1 + i) = 2i$.

Remark: For the most part, I will try to have the letters x, y be real numbers, and use z, w to denote complex numbers (which might or might not also be real). This is also common practice.

2. The Fundamental Theorem of Algebra - Statement of theorem and related results

Recall that a **polynomial** (with complex coefficients) p is a function $p : \mathbb{C} \to \mathbb{C}$ of the form

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n = \sum_{j=0}^{n} a_j z^j,$$

where $a_0, a_1, \cdots, a_n \in \mathbb{C}$. (If $a_0, a_1, \cdots, a_n \in \mathbb{R}$, then more specifically, p is a polynomial with real coefficients.) If at least one of the a_j are not zero, then we can choose to write p in a way so that $a_n \neq 0$, and then n is called the **degree** of p. That is, $3 - 4x + 5x^2 - 2x^3 + 0x^4 + 0x^5$ should be written as $3 - 4x + 5x^2 - 2x^3$, and the degree is 3 rather than 5.

If $a_j = 0$ for all j, then $p(z) = 0$, and p is called "the zero polynomial", and it has degree 0. $p(z) = a_0$ is a degree zero polynomial, called a "constant polynomial". If the degree of p is at least 1, then p is a non-constant polynomial.

z_0 is called a **zero** (also called a **root**) of a polynomial p if $p(z_0) = 0$. If $p(z) = z^2 + 3z + 2$, then the zeroes of p are -2 and -1.

Here is the theorem we eventually want to prove in this essay:

Theorem 1. The Fundamental Theorem of Algebra (FToA)

Each non-constant polynomial p has a complex zero.

Remark: Here we see the convenience of the standard convention defining $\mathbb{C}$: of course many polynomials only have real zeroes. But real numbers are also complex by our chosen definition, and nothing is wrong here.

Remark: Note that in the definition of a polynomial, $p : \mathbb{C} \to \mathbb{C}$. The notation used allows for the possibility that p does not attain every possible output of the space $\mathbb{C}$, and indeed constant polynomials attain only one output value. However, see the corollary below.

Let us also look at some consequences of the FToA (called **corollaries**):

Corollary 2. *Each non-constant polynomial p attains each complex number $w \in \mathbb{C}$ as an output.*

Proof. This is the "subtraction trick." We want a z_0 such that $p(z_0) = w$. This is the same as having $p(z_0) - w = 0$. Create a new function $q(z) := p(z) - w$. Then q is a non-constant polynomial. By the FToA, q has some zero z_0. Then $0 = q(z_0) = p(z_0) - w$, which is the same as $p(z_0) = w$. □

Corollary 3. *Each polynomial p of degree n factors uniquely as*

$$p(z) = a_n(z - z_1)^{m_1} \cdot (z - z_2)^{m_2} \cdots (z - z_k)^{m_k}, \text{ where } m_1 + \cdots + m_k = n.$$

(One simply writes $p(z) = a_0$ in the case $n = 0$, and the "empty product" to the right of a_n in the above expression is simply erased from the equation, or equivalently, regarded as the number 1.)

Proof. (Omitted for now; it is not the emphasis of this essay. Essentially, one repeatedly factors out $z - z_k$ for each zero z_k. Uniqueness is a somewhat more subtle point.) □

3. Rough outline of proof

So far, we don't know a lot. But let's sketch the argument out.

Let $A(z) := |p(z)|$. $A : \mathbb{C} \to \mathbb{R}$ must attain an absolute minimum somewhere in the plane for two reasons: First of all, A is continuous, and second, A goes to $+\infty$

as $|z|$ becomes large instead of, say, gradually approaching any sort sort of ceiling (a plane at a height M playing a two-dimensional role analogous to an asymptote when looking at functions from $\mathbb{R} \to \mathbb{R}$). The reasoning in this step is somewhat less delicate, using only the facts about A discussed in this paragraph. The part about continuity is carried out in Section 9. The part about growth rates is Section 6.

So after that, we know that A has a minimum. Surely $A \geq 0$ everywhere. We want to show that this minimum is zero. If $p(z_0) = 0$, then any change in $p(z)$ points away from the origin, making $A(z)$ increase away from zero. It turns out that A cannot have any other local minimum output. That is, it is impossible for a A to have a valley bottoming out at some positive value. Pick an initial z-value z_0. The reason A can't bottom out at z_0 what $A(z_0) > 0$ is this: for a polynomial, a small change in z can produce small changes in $p(z)$ that point in any complex direction. If $p(z_0) \neq 0$, then a small change in z in a carefully chosen direction produces a small change in $p(z)$ that brings $p(z)$ closer to the origin. That is, $A(z) < A(z_0)$.

How do we choose z that decreases $A(z)$ as described? First, we need Euler's Formula to understand how z^m behaves (Section 4). By a change of variable trick, we can think that $z_0 = 0$, at which point a small change in z is the same thing as a small value of z, which means we need to understand why the small powers of z matter the most when z is close to 0 (Section 6).

4. **Euler's formula and complex multiplications**

To understand polynomials in the complex plane, it is necessary to understand addition and multiplication of complex numbers.

There are two ways to represent complex numbers.

Rectanuglar representation of a complex number: $z = a + bi$

Let us start with what is usually taught to students first, the form $z = a + bi$. This is a form which is very good for intuitively understanding how addition works. If $w = c + di$, then $z + w = a + bi + c + di = (a + c) + (b + d)i$. The real and imaginary parts play out separately with no interaction. Positive and negative numbers cancel as usual, and it isn't too hard to build an intuition for complex addition and subtraction immediately from the definition.

Multiplication, however, is not so easy to visualize with this representation. $(1+i)\cdot(3-2i)$ looks like... well, what? A first time student only knows to multiply

this out and then plot the result, $5 + i$. There seems to be nothing anyone can do other than churn through the algebra; just about any numbers could have shown up in the answer, for all the student can tell, at this point.

To make matters worse, polynomials inevitably will ask the student to understand powers of z. Who really wants to compute the following by hand?

$s := (3 - 4i)^{18}$

But there is no hope of understanding complex polynomials without understanding how z^n depends on z. It simply won't do to conclude that s is whatever number results from a tedious computation. We need a new point of view.

Fortunately, we have another way of representing numbers in the complex plane.

First, let's just look at the size of a complex number. Plot the point $a + bi$ in the complex plane, and define its **modulus** (plural, **moduli**) as follows:

$|a + bi| := \sqrt{a^2 + b^2}$

This is just the distance from the origin of the complex plane, which generalizes the notion of absolute value from $\mathbb{R}$ to all of $\mathbb{C}$. (Because if $b = 0$, then $a + bi = a \in \mathbb{R} \subseteq \mathbb{C}$, and $|a + bi| = |a|$ interpreted as a modulus of a complex number is the same value as $|a|$, the absolute value of the real number a.)

Polar representation of a complex number: $z = r\cos(\theta) + ri\sin(\theta) = r[\cos(\theta) + i\sin(\theta)]$

The polar representation uses the usual trigonometric functions sine and cosine to express z in terms of the **modulus** $r = |z|$ and the angle θ. Specifically, θ is the counterclockwise angle from the positive real axis to the ray going from the origin to z. This angle is called an **argument**[2] of z. (In particular, $\theta + 2k\pi$ is another argument for z for any $k \in \mathbb{Z}$)

The following should not be so easily taken for granted. It is a first step toward understanding multiplication of complex numbers.

Exercise: Let $z = a + bi$, $w = c + di$. Then $|z \cdot w| = |z| \cdot |w|$. (Compute $|z|$, $|w|$, and $|z \cdot w|$ separately in terms of a, b, c, and d to see this. Simplify.)

This exercise tells us at least something about $s = (3 - 4i)^{18}$. Since $|3 - 4i| = 5$, one has that $|s| = |(3 - 4i)^{18}| = |3 - 4i|^{18} = 5^{18}$. So s lies on a circle centered at the origin having radius 5^{18}. We do not yet know where on this circle s lies, which now amounts to finding the argument s.

[2]Not **the** argument of z, as there are many.

We can write any non-zero complex number as $z = ru$, where $r = |z| \in (0, \infty)$ and $u = \frac{z}{|z|} \in \mathbb{C}$. u has the property $|u| = 1$. u lies on the unit circle, and so u can be written in the form

$$E(\theta) := \cos(\theta) + i \sin(\theta)$$

$E(\theta)$ traces out a lap around the unit circle as θ ranges from 0 to 2π. $r = |z|$

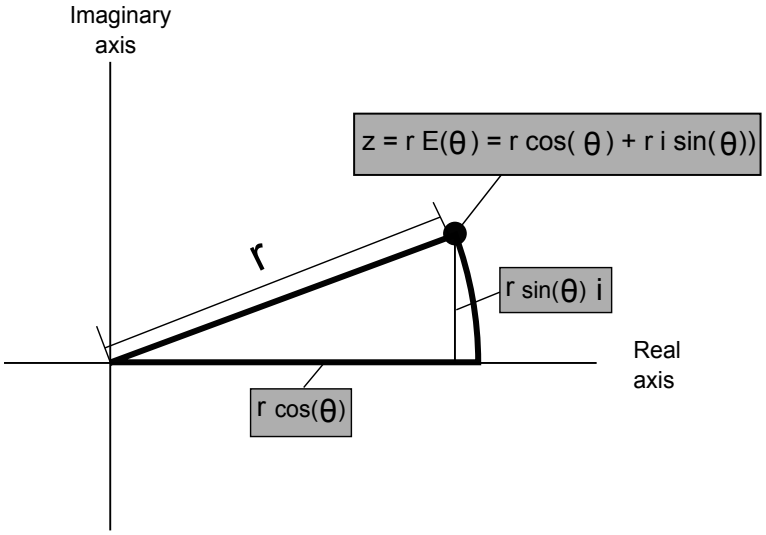

FIGURE 1. Good old trigonometry. The real and imaginary parts of a complex number are plotted as the two coordinates of a point in the plane, with the positive imaginary axis represented as the "up" direction, and the positive real axis represented as the "right" direction.

describes the size of z, and the argument θ describes the direction of z. So we have the representation $z = rE(\theta)$

Consider non-zero complex numbers, $z_1 = r_1 u_1$ and $z_2 = r_2 u_2$. Since $|z_1 z_2| = |r_1||r_2||u_1||u_2| = r_1 r_2$, we know the modulus of $z_1 z_2$ is $r_1 r_2$. We can multiply this as $z_1 z_2 = r_1 u_1 r_2 u_2 = (r_1 r_2)(u_1 u_2)$, which is now written in the form $rE(\theta_3)$, where $E(\theta_3) = u_1 u_2$. What we would like, then, is a nice formula for θ_3 in terms of θ_1 and θ_2. If the formula is simple, then polar form is good for understanding multiplication; if not, then it remains difficult to understand complex multiplication intuitively.

We are fortunate enough to have it work out nicely for us.

Definition:[3] For $z \in \mathbb{C}$,

$$e^z := \sum_{n=0}^{\infty} \frac{z^n}{n!}$$

$$\sin(z) := \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{(2n+1)!}$$

$$\cos(z) := \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!}$$

(If $z = x \in \mathbb{R}$, then these are the Taylor series expansions for e^x, $\sin(x)$, and $\cos(x)$.)

Theorem 4. Euler's Formula *For all $\theta \in \mathbb{R}$,*

$$e^{i\theta} = \cos(\theta) + i\sin(\theta) = E(\theta)$$

(This follows directly from the three definitions above.)

Corollary 5. Bieber's Theorem

$$e^{i\pi} + 1 = 0$$

Proof. Exercise. □

By itself, Euler's Formula is not helpful. We need to know that this strange new exponential function behaves, in important ways, like the real one.

Theorem 6. Exponential identity *For any $z, w \in \mathbb{C}$,*

$$e^{z+w} = e^z e^w.$$

(It follows that $(e^z)^n = e^{nz}$)

(Proof postponed.)

Corollary 7. Multiplication on the unit circle

$$E(\theta_1)E(\theta_2) = e^{i\theta_1}e^{i\theta_2} = e^{i\theta_1 + i\theta_2} = e^{i(\theta_1+\theta_1)} = E(\theta_1 + \theta_2)$$

In other words, to see geometrically what happens when one multiplies two unit complex numbers, simply add the angles together to get the new angle.

Let's put everything together.

[3]Sometimes this is stated as a theorem or identity rather than as the definition; if there are many formulas for the same thing, then the choice of which is the definition and which is an alternate expression is left as a matter of taste.

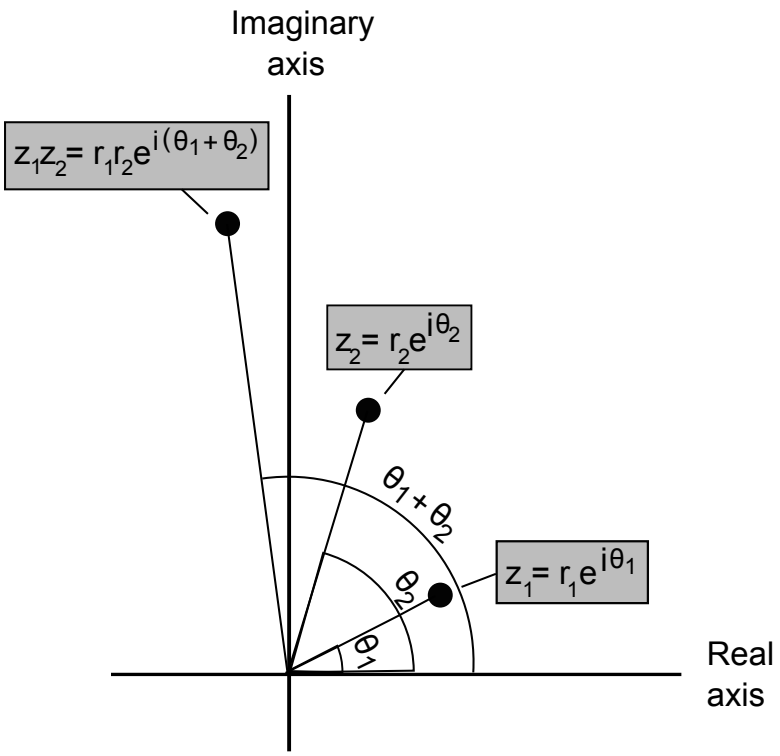

FIGURE 2. The geometric picture of how multiplication of complex numbers works.

Theorem 8. How multiplying complex numbers works *To multiply two complex numbers together,* ***multiply*** *their moduli and* ***add*** *their arguments.*

That is, $r_1e^{i\theta_1}r_2e^{i\theta_2} = (r_1r_2)e^{i(\theta_1+\theta_2)}$

In future sections, we will write $e^{i\theta}$, and never again use the notation $E(\theta)$.

4.1. **Optional - trigonometric identities playing the role of Euler's formula.** It wasn't strictly necessary to appeal to Euler's Formula. We really just wanted to know that

$$E(\theta_1) \cdot E(\theta_2) = E(\theta_1 + \theta_2)$$

Behold **the angle sum formulas:**

$$\cos(\theta_1 + \theta_2) = \cos(\theta_1)\cos(\theta_2) - \sin(\theta_1)\sin(\theta_2)$$

$$\sin(\theta_1 + \theta_2) = \sin(\theta_1)\cos(\theta_2) + \cos(\theta_1)\sin(\theta_2)$$

Thus, without knowing about Euler's formula, one could figure out the following:

$$E(\theta_1) \cdot E(\theta_2) = [\cos(\theta_1) + i\sin(\theta_1)][\cos(\theta_2) + i\sin(\theta_2)]$$

$$= \big[\cos(\theta_1)\cos(\theta_2) - \sin(\theta_1)\sin(\theta_2)\big] + i\big[\sin(\theta_1)\cos(\theta_2) + \cos(\theta_1)\sin(\theta_2)\big]$$
$$= \cos(\theta_1+\theta_2) + i\sin(\theta_1+\theta_2) = E(\theta_1+\theta_2)$$

Though in fact, it's a two-way street. If you do not know the angle sum formulas, then it is much easier to remember Euler's Formula and the exponential identity $e^{z+w} = e^z e^w$, which implies the angle sum formulas just by rearranging the above computation:

$$e^{i(\theta_1+\theta_2)} = \cos(\theta_1+\theta_2) + i\sin(\theta_1+\theta_2) \tag{4.1}$$

$$\begin{aligned} e^{i\theta_1}\cdot e^{i\theta_2} &= [\cos(\theta_1) + i\sin(\theta_1)][\cos(\theta_2) + i\sin(\theta_2)] \\ &= \big[\cos(\theta_1)\cos(\theta_2) - \sin(\theta_1)\sin(\theta_2)\big] + i\big[\sin(\theta_1)\cos(\theta_2) + \cos(\theta_1)\sin(\theta_2)\big] \end{aligned} \tag{4.2}$$

These are equal:

$$e^{i(\theta_1+\theta_2)} = e^{i\theta_1}\cdot e^{i\theta_2}$$

Therefore the real and imaginary parts on the right hand side of (4.1) and (4.2) are equal, giving the angle sum formulas.

A strong reason for using Euler's Formula here is for intuition. The following identity looks at first mysterious and unfamiliar, even if it's a convenient formula that is simple to apply:

$$E(\theta_1)\cdot E(\theta_2) = E(\theta_1+\theta_2)$$

By comparison, the following identity is just a matter of allowing yourself to plug imaginary numbers into a formula that you should already know:

$$e^{i\theta_1}\cdot e^{i\theta_2} = e^{i\theta_1+i\theta_2} = e^{i(\theta_1+\theta_2)}$$

It can also be argued that understanding Euler's Formula is a first step toward understanding the complex function e^z, and toward understanding complex analysis as a whole.

4.2. **Optional -** $\sin(z)$**,** $\cos(z)$**,** e^z**.** Other identities exist between the three functions. For example,

$$e^{iz} = \cos(z) + i\sin(z)$$

for any $z \in \mathbb{C}$; the case important to this essay is the case $z = \theta \in \mathbb{R}$.

$$e^{-iz} = \cos(z) - i\sin(z),$$

thus

$$\cos(z) = \frac{1}{2}\big[e^{iz} + e^{-iz}\big]$$

$$\sin(z) = \frac{-i}{2}\left[e^{iz} - e^{-iz}\right]$$

(As before, sometimes this is stated as a definition rather than an identity; if there are many formulas for the same thing, then which is the definition and which is an alternate expression is a matter of taste.)

5. n-th roots

Before proving the FToA, let's look at a special case, n-th roots. I claim that this is closer to the heart of the general case then one might expect.

Fix any complex number $w = c + di$, and any positive integer n. Does w have a an n-th root $z_0 = a + bi$? That is, is there a z_0 such that $z_0^n = w$? In other words, for $c, d \in \mathbb{R}$, are there $a, b \in \mathbb{R}$ such that $(a+bi)^n = c+di$? Note that the statement $z_0^n = w$ is the same as saying that z_0 is a root of the polynomial $p(z) = z^n - w$, so that we're investigaing whether the Fundamental Theorem of Algebra works in this one special case.

Hopefully the reader did not skip the previous section. The answer to n-th root question should be within reach if the reader uses the polar representation of both z and w. It's worth trying to do so before continuing on.

Exercise - Find out whether roots exist, what they are, and how many there are.

Solution: Let $\rho := |w|$ (ρ is the lower-case Greek letter sometimes written out as "rho", and pronounced as "row".) Let α (Greek "alpha") be the argument of $w = \rho e^{i\alpha}$. As usual, let $z_0 = re^{i\theta}$. We want to find all z such that $z_0^n = w$. That is, such that

$$(re^{i\theta})^n = \rho e^{i\alpha}$$

$$r^n e^{i\theta \cdot n} = \rho e^{i\alpha}$$

If $\rho = 0$, then the only solution is $z_0 = 0$. In all other cases, we arrive at a solution z_0 **if and only if** we have $r^n = \rho$ and $n\theta = \alpha + 2k\pi$ for some $k \in \mathbb{Z}$. In other words, $r = \rho^{\frac{1}{n}}$ and $\theta = \frac{\alpha}{n} + \frac{2k\pi}{n}$, for some $k \in \mathbb{Z}$.

The solutions, then, are $z_0 = \rho^{\frac{1}{n}} e^{i[\frac{\alpha}{n} + \frac{2k\pi}{n}]} = \rho^{\frac{1}{n}} e^{i\frac{\alpha}{n}} e^{i\frac{2k\pi}{n}}$, $k \in \mathbb{N}$. However, different k can give the same z_0 many times. Distinct z_0 are arrived at for $k = 0, ..., n-1$, and these n different z_0 are repeated periodically as k ranges over all other integers. Thus the solution set can be written

$$\{z_0 : z_0^n = w\} = \{\rho^{\frac{1}{n}} e^{i\frac{\alpha}{n} + i\frac{2k\pi}{n}} : k = 0, 1, ..., n-1\}.$$

(To elaborate slightly: note that $e^{2\pi it}$ traces a lap around the unit circle as t ranges from 0 to 1, and then retraces the same path when t ranges from m to $m+1$, for any $m \in \mathbb{Z}$; this is because of Euler's Formula and the periodicity of sin and cos. Since we used $t = \frac{k}{n}$ in the above paragraph, we visit the same n points over and over again.)

□

This result is rather nice. It says that we can be rather choosey about what we want z^n to do. It can have whatever size we would like, and it can point in whatever direction we would like. Both of these are rather important.

6. **Dominant terms**

Next, let us understand what a polynomial does when $|z|$ is large, and when $|z|$ is small. If you've ever sat down and really thought about real polynomials, then to a large extent, this is a complex repetition of a real-number phenomenon.

Claim: When $|z|$ is small, the smallest power of z is the most important. When $|z|$ is large, the largest power of z is the most important. In particular, $|z|^m$ is large when $|z|$ is large, and even larger when m is large. On the other hand, $|z|^m$ is small when $|z|$ is small, and even smaller when m is large.

Let us say something more concrete. First, consider the triangle inequalities for complex numbers:

$$|z_1 + z_2 + \cdots + z_m| \leq |z_1| + |z_2| + \cdots + |z_m| \text{ } \textit{(Trianle inequality)}$$

$$|z + w| \geq |z| - |w| \text{ } \textit{(Reverse triangle inequality)}$$

The reverse inequality follows from the first: $|z| = |(z + w) - w| \leq |z - w| + |-w| = |z - w| + |w|$, then subtract $|w|$ from both sides. Also, one may have $w = w_1 + \cdots + w_m$, and then it follows that $|z + w| \geq |z| - |w_1| - |w_2| - \cdots - |w_m|$ (take the negative of the triangle inequality; remember that the $\leq$ flips to become $\geq$).

We will use the expanded reverse triangle inequality to compare the dominant term of a polynomial to all other terms. That is, if z is a "dominant" term of a sum and the w_j are much smaller terms, we want to control how much effect the lesser terms can possibly have. The reverse triangle inequality essentially says that you can't do worse than the "worst case", where the w_j all point exactly opposite to z.

Let's write a polynomial p in a form where we skip over terms with coefficient 0. That is, $p(z) = \sum_{j=1}^{m} a_j z^{k_j}$, where $k_j \in \mathbb{N}$ form an increasing sequence, and $a_j \in \mathbb{C}$ are all non-zero.

The case where r is large: For a non-constant polynomial p, we want to show that z^{k_m} is the largest term by far when $|z| = r$ is large. In particular, for any large positive number N, we want to find some large value R so $|p(z)| > N$ whenever $r = |z| \geq R$. This should definitely be believable! The sizes of the different terms depend on $|z|$ exactly as they do in the real case, where the largest term is so large that the others all put together cannot cancel it when $|z|$ is too big. This is the relevant and familiar feature; the fact that possible cancellation may occur in the two-dimensional complex plane rather than on the one-dimensional real line does not change the important facts here.

(Note that the value R depends on N and on the particular polynomial. That is, all polynomials grow large eventually, but some may take longer than others.)
Computation: One has that

$$|P(z)| = |\sum_{j=1}^{m} a_j z^{k_j}| = |a_m z^{k_m} + \sum_{j=1}^{m-1} a_j z^{k_j}|$$

$$\geq |a_m| \cdot |z^{k_m}| - \sum_{j=1}^{m-1} |a_j| \cdot |z^{k_j}| = |a_m| r^{j_m} - \sum_{j=1}^{m-1} |a_j| r^{k_j}$$

Note that $r^{k_j} \leq r^{k_{j'}}$ for $j \leq j'$ (since r "is large"; in particular, $r \geq R \geq 1$). So $-|a_j| r^{k_j} \geq -|a_j| r^{j_{m-1}}$ for each $j = 1, 2, ..., m-1$. Let $M := \sum_{j=1}^{m-1} |a|_j$.

Continuing,

$$|P(z)| \geq |a_m| r^{j_m} - r^{j_{m-1}} \sum_{j=1}^{m-1} |a_j| = |a_m| r^{j_m} - r^{j_{m-1}} M = r^{j_{m-1}} \left[r^{j_m - j_{m-1}} |a_m| - M \right]$$

For example, one can have $R \geq N^{\frac{1}{j_{m-1}}}$ and $R \geq \left(\frac{M+1}{|a_m|}\right)^{\frac{1}{r^{j_m - j_{m-1}}}}$, so that

$$|P(z)| \geq r^{j_{m-1}} \left[r^{j_m - j_{m-1}} |a_m| - M \right] \geq R^{j_{m-1}} \left[R^{j_m - j_{m-1}} |a_m| - M \right] \geq N \cdot 1 = N$$

The case where r is small: We want to show that, for a small enough number r_0, we have that $P(z) \approx a_1 z^{k_1}$. In particular, fix a small number $\varepsilon > 0$. We will show that when $r < r_0$ is small enough, the higher power terms $j = 2, ..., m$ all added together are at most ε (Greek "epsilon") times as big as the $j = 1$ term alone.

Computation:

$$P(z) = a_1 z^{k_1} + \sum_{j=2}^{m} a_j z^{k_j}$$

This time, let $M := \sum_{j=2}^{m} |a_j|$. Note that $r \leq r_0 \leq 1$, so that $r^{k_j} \leq r^{k_2}$ for all $j = 2, ..., m$.

$$|\sum_{j=2}^{m} a_j z^{k_j}| \leq \sum_{j=2}^{m} |a_j| r^{k_j} \leq \sum_{j=2}^{m} |a_j| r^{k_2}$$

$$\leq M r^{k_2} = M r^{k_1} \cdot r^{k_2 - k_1} = |a_1| r^{k_1} \cdot \big[\frac{M}{|a_1|} r^{k_2 - k_1}\big]$$

The lowest-order term is $a_1 z^{k_1}$, which has modulus $|a_1| r^{k_1}$. So we need r small enough so that $\frac{M}{|a_1|} r^{k_2 - k_1} \leq \varepsilon$. That is, $r \leq = \big(\frac{|a_1|\varepsilon}{M}\big)^{\frac{1}{k_2 - k_1}} =: r_0$.

7. Continuous functions

Lemma 9. *Each polynomial p is continuous. Additionally, the function $A(z) := |p(z)|$ is a continuous funtion. (All of this remains true regardless of whether the domain is $\mathbb{C}$ or any subset of $\mathbb{C}$).*

The formal definition of "continuous" and the proof of this lemma are sometimes addressed in a first calculus course, though most often only for the domain $\mathbb{R}$, and the proof tends to be emphasized more heavily a few courses later, in a first real analysis course. Informal definitions popular in earlier courses are adequate for the purposes of thinking about this essay. We use the lemma above because we want to use the next lemma.

Lemma 10. ***Maxima and minima on a closed rectangle***

For $C > 0$, let $B := \{x + iy : x \in [-C, C], y \in [-C, C]\}$. If $f : R \to \mathbb{R}$ is a continuous function, then f attains an absolute maximum and an absolute minimum. (Sometimes these are instead called a global maximum and a global minimum)

Proof. (Postponed.) □

Remark: We will apply the above lemma in the case where $f(z) = A(z) = |p(z)|$. We regard B as the domain of A for the purposes of the above lemma, even though one can also have $A : \mathbb{C} \to \mathbb{R}$. One can easily have z_0 such that $A(z_0)$ minimizes $A(z)$ for all $z \in B$, but not for all $z \in \mathbb{C}$.

Remark: Veterans of a first real analysis course should recognize the above lemma as a consequence of two standard theorems: (i) The continuous image of a compact set is compact, and (ii) (The Heine-Borel theorem:) Subsets of $\mathbb{R}^d$ are compact **if and only if** they are closed and bounded.

Remark: The above lemma is the least constructive part of this proof of the FToA. Standard proofs of the lemma argue that an absolute minimum is attained **somewhere**, but do not illuminate a way to actually find it, in practice. Thus the FToA, as we prove it here, isn't particularly constructive, either.

Now we are ready to take a large step toward proving the FToA.

Corollary 11. *Let p be a non-constant polynomial. Let $A : \mathbb{C} \to \mathbb{C}$ be defined by $A(z) := |p(z)|$. Then A attains an absolute minimum.*

Proof. The main point is that $A : \mathbb{C} \to \mathbb{C}$ in this corollary, where the previous lemma can only be applied to A with a restricted domain, that is, $A : B \to B$.

If there is no constant term, then of course $A(0) = 0$, and this is the absolute minimum. Otherwise, write $p(z) = \sum_{j=1}^{m} a_j z^{k_j}$, where $k_1 = 0$ and all the a_j are non-zero. Recalling Section 6, we know that there is an R large enough so that whenever $|z| = r \geq R$, then $|p(z)| > |a_1| = |p(0)|$. We know that polynomials are continuous by Lemma 9, so apply Lemma 10, letting $C = R$, and letting the continuous function $A : B \to \mathbb{C}$. Thus $A : B \to \mathbb{C}$ has an absolute minimum at some $z_0 \in B$, and $A(z_0) \leq A(0) = |a_1|$. For all z outside of the box B, one has $|z| > R$, and it follows that $A(z) > |a_1|$. Thus $A(z_0) < A(z)$ for all $z \in \mathbb{C}$, not just for all $z \in B$. That is, z_0 also minimizes $A : \mathbb{C} \to \mathbb{C}$. □

We are not done because we do not yet know whether $A(z_0) = 0$ or if $A(z_0) > 0$. We want to prove that the former case is true.

We want to prove that if p is a non-constant polynomial and A has a local minimum at z_0, then $p(z_0) = 0$. The **contrapositive**[4] statement is as follows:

Theorem 12. Reduced Fundamental Theorem of Algebra *Let p be a non-constant polynomial. Then whenever $p(z_0) \neq 0$, A does **not** attain an absolute minimum at z_0.*

[4]Let A and B be mathematical statements which may be true or false under whatever circumstances. "A implies B" is the **contrapositive** of the statement "not-B implies not-A". These statements are totally identical. For example, "if it rains (A), then I will get wet (B)" becomes "if I am dry (not-B), then I was not rained on (not-A)."

So now we need to start at a point z_0, and show that we can find another z where A is even smaller.

Reduction step 1: Change of variables. We want to be able to assume that $z_0 = 0$, for convenience.

To do this, we consider a related polynomial. Suppose $A(z_0) = |p(z_0)| \neq 0$. Then let $\tilde{p}(z) := p(z + z_0)$. First of all, note that $\tilde{p}(z)$ is a polynomial; to see this, write out $p(z + z_0)$, expanding out each $(z + z_0)^{k_j}$, and collect like powers of z. Also, $\tilde{p}$ and p have the exact same set of outputs, and $|p|$ and $|\tilde{p}|$ have the exact same absolute minimum output. So to show that $|p|$ is not minimized at $z = z_0$, we need to show that $|\tilde{p}|$ is not minimized at $z = 0$.

Reduction step 2: Renormalization. We want to assume that $\tilde{p}(0) = 1$.

We have that $\tilde{p}(0) = a_1 \neq 0$. Then let $p^*(z) := \frac{1}{a_1}\tilde{p}(z)$. p^* defines another polynomial when you distribute the $\frac{1}{a_1}$, and $|p^*(z)| = \frac{1}{|a_1|}|\tilde{p}(z)|$. These two expressions are clearly minimized at exactly the same z values.

Combining these two reduction steps, we may consider the question about p^* at $z = 0$ instead of the question about p at $z = z_0$. Let us state the new theorem to prove.

Theorem 13. Re-reduced Fundamental Theorem of Algebra *Let p be a non-constant polynomial. If $p(0) = 1$, then A does **not** attain an absolute minimum at $z = 0$.*

Proof. To prove this, we will need to combine ideas: the idea that gave us n-th roots, and the fact that when $|z|$ is small, the larger powers of z do not matter much.

Write $p(z) = 1+a_1z^{k_1}+\sum_{j=2}^{m} a_jz^{k_j}$. Write $a_1 = \rho e^{i\alpha}$. We are looking for z, which we express in the form $z = re^{i\theta}$. Also define the remainder $R(z) := \sum_{j=2}^{m} a_jz^{k_j}$, and let $p_1(z) := 1 + a_1z^{k_1}$.

When $z = 0$, we start at $p(0) = 1$. From the section on n-th roots, we now know that $a_1z^{k_1}$ can point in any direction that we'd like. From Section 6 using $\varepsilon = \frac{1}{2}$, we also know that if r is small enough, then $|R(z)| \leq \frac{1}{2}|a_1|r^{k_1}$.

Perhaps now is a good time to take note of the figures.

First, let's arrange to have $|p_1(z)| < 1$. Write

$$p_1(z) = 1 + a_1z^{k_1} = 1 + \rho e^{i\alpha}r^{k_1}e^{ik_1\theta}$$

$$= 1 + \rho r^{k_1}e^{i(\alpha+k_1\theta)}$$

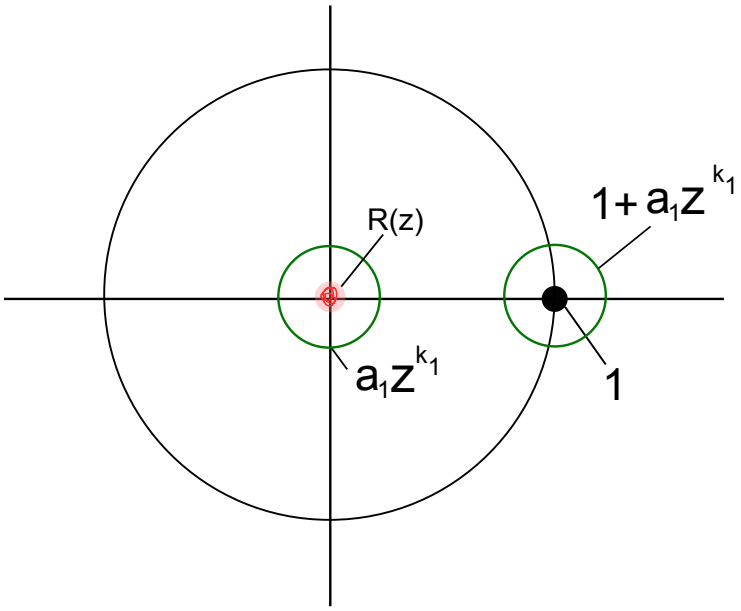

FIGURE 3. $z = re^{i\theta}$, r is small. Before we choose θ, the $p_1(z) = 1 + a_1 z^{k_1}$ could lie anywhere indicated.

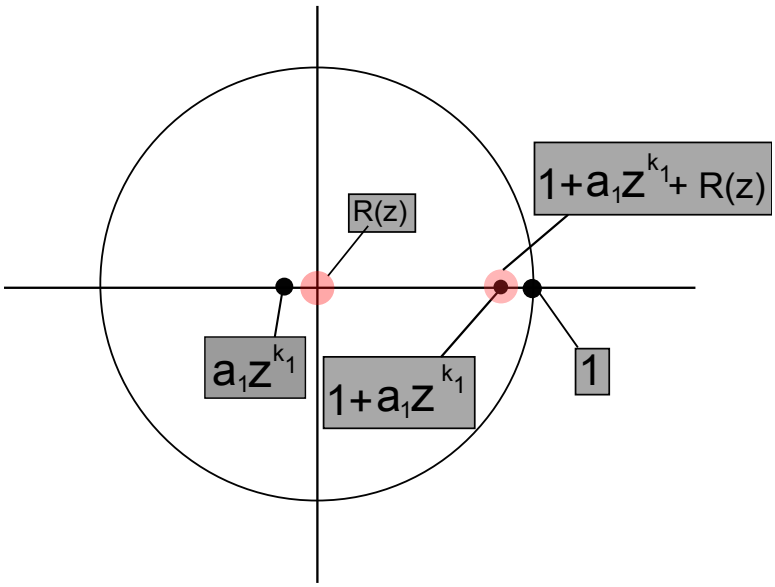

FIGURE 4. After we pick θ, we know what $p_1(z)$ is, and we have control over the size of $R(z)$; thus $|p(z)| < 1$, as desired.

For these two terms partially to cancel each other, we want to have $e^{i(\alpha+k_1\theta)} = -1$, which (by Bieber's Theorem) occurs if $\alpha + k_1\theta = \pi$. So let $\theta := \frac{\pi-\alpha}{k_1}$.

Then we get

$$p_1(z) = 1 + \rho r^{k_1} e^{i\pi}$$

By Bieber's Theorem, $e^{i\pi} = -1$, so:

$$p_1(z) = 1 - \rho r^{k_1}$$

$$|p_1(z)| = |1 - \rho r^{k_1}| = 1 - \rho r^{k_1}$$

(the last equality assumes that r is small enough. Let's assume r is small enough so that $\rho r^{k_1} < 1$.)

Recall that $|R(z)| \leq \frac{1}{2}\rho_1 r^{k_1}$. Then we can finish the FToA as follows:

$$|p(z)| = |p_1(z) + R(z)| \leq |p_1(z)| + |R(z)| \leq 1 - \rho r^{k_1} + \frac{1}{2}\rho r^{k_1} = 1 - \frac{1}{2}\rho r^{k_1} < 1$$

□

That's the Fundamental Theorem of Algebra for you.

Matthew Bond, Department of Mathematics, University of British Columbia bondmatt@math.ubc.ca