The Fundamental Theorem of Algebra

Unlike the *Fundamental Theorem of Arithmetic*, the history of the *Fundamental Theorem of Algebra* begins in medieval Europe. The Greeks and the Hindus had only vague ideas about polynomial equations, and no notion whatsoever of complex numbers. In fact, they tended to think of numbers as positive quantities: the idea of a negative number was foreign to them, and the idea of the square root of a negative number was literally unthinkable one or two thousand years ago.

In 1546 an Italian mathematician named Girolamo Cardano published *Artis Magnae Sive de Regulis Algebraicis Lib. Unus* containing, among other things, formulas for the solution of the general cubic and quartic equations. Commenting on one particular example he said "Dismissing mental tortures, and multiplying $5 + \sqrt{-15}$ by $5 - \sqrt{-15}$, we obtain $25 - (-15)$. Therefore the product is 40. ... and thus far does arithmetical subtlety go, of which this, the extreme, is, as I have said, so subtle that it is useless." So Cardano was willing to use imaginary numbers to solve an equation, even though they are "useless".

Cardano wasn't the only medieval mathematician to regard imaginary/complex numbers with suspicion. But as the attempts to solve algebraic equations of degree greater than four flourished, acceptance of the "impossible quantity" $\sqrt{-1}$ gained ground, becoming entirely respectable when Leonhard Euler in 1748 published *Introductio in Analysin Infinitorum*, in which he introduced the famous formula $e^{i\theta} = \cos\theta + i\sin\theta$. Euler probably suspected that the Fundamental Theorem is true, but never actually stated it. Carl Friedrich Gauss offered a proof in his doctoral dissertation (1799), but it was flawed. Jean Robert Argand gave a complete proof in 1806. We present Argand's proof here, since it is elementary and fairly easy to understand.

The Fundamental Theorem has been proved in dozens of different ways. Gauss himself offered four different proofs during his lifetime. Unfortunately, many of these proofs depend on results from an advanced branch of calculus known as *Functions of a Complex Variable*, and are beyond the grasp of most students.

Just for illustrative purposes, here is a simple proof of the Fundamental Theorem of Algebra that relies on Liouville's Theorem. I will not attempt to prove Liouville's Theorem; that proof depends on Cauchy's Integral Formula, which is an advanced result in complex analysis. We will define a few terms from complex analysis before we state Liouville's Theorem. A function $f(z)$ is said to be *holomorphic* at a point z_0 in the complex plane if the derivative $\frac{d}{dz}f(z_0)$ exists. A function that is holomorphic at every point $z \in \mathbb{C}$ is an *entire function*. A function (like $f(z) = 1/z$) that is differentiable almost everywhere is said to be a *meromorphic* function; the points at which a meromorphic function $f(z)$ is not differentiable are said to be the *poles* of $f(z)$. For example, $f(z) = 1/z$ has a pole at $z_0 = 0$; as $z \to 0$, $1/z$ approaches the point at infinity. Since $1/z$ has values that are very far apart within a vanishingly small radius around 0, the derivative of $f(z) = 1/z$ does not exist at the point $z = 0$.

Liouville's Theorem states that if an entire function $f(z)$ is bounded (i.e., $|f(z)| < M$ everywhere in $\mathbb{C}$), that function is a constant.

With Liouville's Theorem in hand we can prove the Fundamental Theorem of Algebra quite easily. Let $P(z)$ be a polynomial of degree ≥ 1, and suppose that $P(z)$ has no zeroes

in $\mathbb{C}$. Then $g(z) = 1/P(z)$ is an entire function. We will use Liouville's Theorem to arrive at a contradiction.

First, since the polynomial $P(z)$ is of some degree $n \geq 1$, for large values of z it is dominated by its leading term $a_n z^n$. Therefore we can find a radius R such that, if $|z| > R$, $|P(z)| > |\frac{a_n}{2}| * |z^n| > |\frac{a_n}{2}| * R^n$. Clearly, then, $|g(z)| = |1/P(z)| < |\frac{2}{a_n}| * 1/R^n$ when $|z| > R$. So $|g(z)|$ is bounded when $|z| > R$. But the set of z such that $|z| \leq R$ is closed, and therefore $|g(z)|$ is bounded there as well, by the extreme value theorem. Since $|g(z)|$ is bounded everywhere in $\mathbb{C}$, $g(z)$ is constant. But then $P(z) = 1/g(z)$ is also constant. But this contradicts the fact that a polynomial function of degree ≥ 1 assumes infinitely many values. Therefore $P(z) = 0$ for some complex number z_0, and $g(z) = 1/P(z)$ has a pole at z_0.

Argand's Proof of the Fundamental Theorem

In the discussion that follows, we will use *polar coordinates* to define certain quantities in the complex plane. Broadly speaking, there are two ways to specify a particular complex number $a + bi$. One way is to use *Cartesian coordinates*, as I have just done: a particular complex number z_0 is defined by an ordered pair (a, b) which is understood to be a point a units away from the origin in the x or real, direction, and b units away from the origin in the y, or imaginary, direction. The point z_0 can also be defined by an angle θ and a length r. If we draw an arrow from the origin to the point z_0, we define r to be the length of that arrow, and we define θ to be the angle between that arrow and the x-axis, going counter-clockwise. For convenience we restrict θ to a finite range, $0 \leq \theta < 2\pi$. We say the ordered pair (r, θ) are the polar coordinates for z_0, and we use Euler's formula to assert that $z_0 = re^{i\theta} = r(\cos\theta + i\sin\theta)$. Figure 1 below illustrates the two coordinate systems.

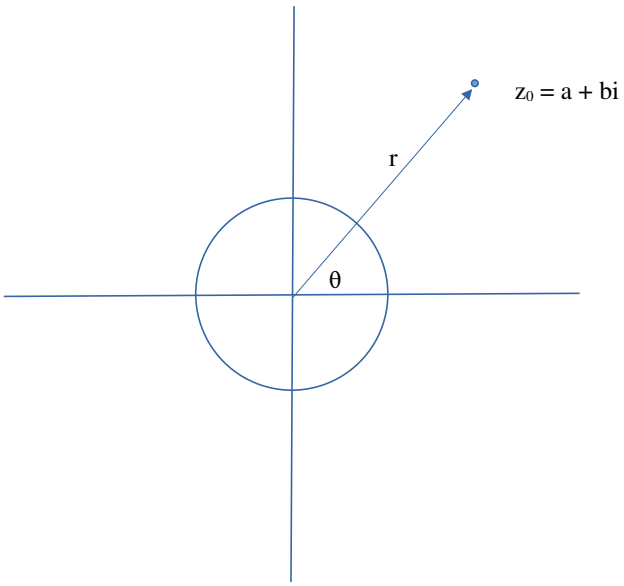

Figure 1: Two Coordinate Systems for $\mathbb{C}$

Before proceeding with Argand's proof we should first establish the Extreme Value Theorem for analytic functions of complex variables, on which Argand's proof depends. The Extreme Value Theorem for a real-valued function $f(x)$ states that if $f(x)$ is continuous on the closed interval $[a, b]$, it attains both a maximum value M and a minimum value m on $[a, b]$. This result follows directly from the hypotheses: f is continuous, so it must approach say, its maximum value very closely, and, since the domain of $f(x)$ is closed, it must actually reach that maximum value. For example, the function $f(x) = x + 1$ approaches the value 2 as $x \to 1$. On the half-open interval $[0, 1), f(x)$ does not actually attain the value 2; on the closed interval $[0, 1]$, it does.

The Extreme Value Theorem also holds for a continuous function with complex values, with a few modifications. Since there is no natural ordering of the complex numbers, we must consider the magnitude of the complex values, that is, if a complex-valued function $f(z)$ is continuous on a closed region Ω of the complex plane $\mathbb{C}$, $|f(z)|$ must attain both a maximum and a minimum value on Ω. As in the case of a function of a real variable, the proof follows immediately from the facts that (a) $f(z)$ is continuous and (b) Ω is closed.

The Fundamental Theorem of Algebra: Every polynomial $P(z)$ of degree ≥ 1 with complex coefficients has at least one root in $\mathbb{C}$. In other words, $\mathbb{C}$ is algebraically complete.

We may assume without loss of generality that $P(z)$ is monic; that is, $P(z) = z^n + a_{n-1}z^{n-1} + a_{n-2}z^{n-2} + \ldots + a_1 z + a_0$. Because $|P(z)|$ is dominated by the first term we can find a positive radius R such that on the boundary of $|z| \leq R, |P(z)| > 1$. Therefore, by the extreme value theorem, $|P(z)|$ must attain a minimum value on the closed disk $|z| \leq R$. Let us assume that $|P(z)|$ attains its minimum m at z_0, and that $|P(z_0)| = m > 0$. We will use the Binomial Theorem to show that there must then exist a point z_0' such that $|P(z_0')| < |P(z_0)|$; this contradiction establishes the theorem.

So let's apply the Binomial Theorem to expand $P(z_0 + h)$, where h is understood to be a small increment $\neq 0$ in z_0:

$$\begin{aligned}
P(z_0 + h) &= (z_0 + h)^n + a_{n-1}(z_0 + h)^{n-1} + \ldots + a_1(z_0 + h) + a_0 \\
&= z_0^n + \binom{n}{1} z_0^{n-1} h + \binom{n}{2} z_0^{n-2} h^2 + \ldots + h^n + \\
&\quad a_{n-1}\left(z_0^{n-1} + \binom{n-1}{1} z_0^{n-2} h + \binom{n-1}{2} z_0^{n-3} h^2 + \ldots + h^{n-1} \right) + \ldots + a_1 z_0 + a_1 h + a_0 \\
&= P(z_0) + \left(\binom{n}{1} z_0^{n-1} + a_{n-1}\binom{n-1}{1} z_0^{n-2} + \ldots + a_1 \right) h + \\
&\quad \left(\binom{n}{2} z_0^{n-2} + a_{n-1}\binom{n-1}{2} z_0^{n-3} + \ldots + a_2 \right) h^2 \\
&\quad + \ldots + h^n \\
&= P(z_0) + k_1 h + k_2 h^2 + k_3 h^3 + \ldots + h^n
\end{aligned}$$

where the k_i do not depend on h and remain constant so long as z_0 is constant. Now it may be that some of the k_i are equal to zero. But they can't all be zero, else we would have $P(z_0) = P(z_0 + h)$, violating the assumptions that $h \neq 0$ and that $P(z)$ is not a constant. Let r be the smallest value of i for which $k_i \neq 0$; let s be the next larger such value of i; and so forth. Then $P(z_0 + h) = P(z_0) + k_r h^r + k_s h^s + k_t h^t + \ldots + h^n$, where none of the coefficients involving h are equal to zero.

Now we will place some restrictions on h. Choose the direction of h^r so that h^r points from $p(z_0)$ toward the origin $z = 0$; that is, if $P(z_0) = me^{i\theta}$, then set $h = \sqrt[r]{\frac{m}{3k_r}}\, e^{-i\theta/r}$, so that $k_r h^r = \frac{1}{3} m e^{-i\theta}$ (the real value of the rth root is taken here). If necessary, reduce the magnitude of h so that $|h| < 1$. And then (if necessary) reduce the magnitude of h some more so that $|k_r h^r| > |k_s h^s + k_t h^t + \ldots + h^n|$; this is possible because $|h| < 1$, and $r < s < t < \ldots < n$. Define T by $T = k_s h^s + k_t h^t + \ldots + h^n$. With these choices made, $|P(z_0 + h)| < |P(z_0)|$, and the theorem is proved. See Figure 2, below.

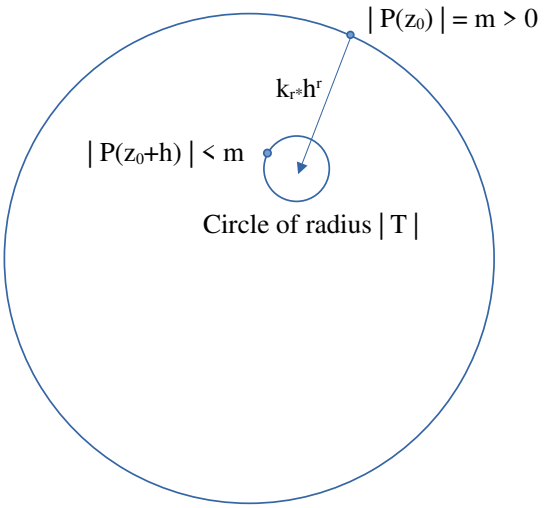

Figure 2: $P(z_0), k_r h^r$, and $P(z_0 + h)$

In Figure 2 the point $P(z_0)$ lies on a circle of radius m centered at the origin. The vector $k_r h^r$ points from $P(z_0)$ toward the origin. The value of $P(z_0 + h)$ must lie within a distance $|T|$ of the inner end of that vector, that is, it must lie somewhere on the perimeter of the small circle of radius $|T|$. Clearly $|P(z_0 + h)| < m$, and the theorem is proved.

Note that, crucially, this proof depends on the existence of a vector that points from $P(z_0)$ toward $z = 0$. When $P(z_0) = 0$, it is not possible to construct such a vector; all vectors of finite length originating at the origin point away from zero, and not towards it.