A study of the Fundamental Theorem of Algebra

and its parallels to the integers.

Justin Boyer

A thesis presented for the degree of
Masters of Mathematics

Department of Mathematics
University of Utah
United States
June 02, 2017

Introduction

In mathematics, symmetry is often analyzed, utilized, or identified. The early part of this paper sets out to help the reader identify the mathematical analogy between the system of integers and the system of polynomials with real coefficients. We will then see parallels between prime numbers and irreducible polynomials, 'clock arithmetic' built from integers using prime numbers and 'clock arithmetic' built from polynomials using irreducible polynomials. Having explored this analogy in some detail, we will turn our attention to the Fundamental Theorem of Algebra, which states that any polynomial with complex coefficients has at least one complex root. Our main goal, the one that will occupy the remainder of this paper, is to use the understanding of the analogy explored in the earlier sections to give a complete and accessible exposition of an algebraic proof of the Fundamental Theorem of Algebra. This paper seeks to expose at an undergraduate college level one of the beautiful mathematical analogies and to show how it is used to establish one of the most important results in algebra.

Comparing the integers and the polynomials with coefficients in the rational, real or complex number fields

Suppose we want to solve a polynomial equation

$$\frac{a_d}{b_d}x^d + \frac{a_{d-1}}{b_{d-1}}x^{d-1} + \ldots + \frac{a_0}{b_0} = 0$$

with rational numbers $\frac{a_i}{b_i}$ (where a_i, b_i are integers) as coefficients. We could turn it into a polynomial equation with integer coefficients just by multiplying both sides of the above equation by a common multiple of $\{b_d, b_{d-1}, \ldots, b_0\}$. So every root of a polynomial with rational coefficients is a root of a polynomial with integer coefficients. A deeper fact, called Gauss's Lemma, says that any factorization of a polynomial

$$a_d x^d + a_{d-1}x^{d-1} + \ldots + a_0$$

with integer coefficients a_i that cannot be factored into two polynomial factors with integer coefficients cannot be factored into two polynomial factors with rational coefficients. These two facts led mathematicians to study systems of polynomials with coefficients in a number system in which you could add, subtract, multiply *and* divide. They called such a system a *field*. Besides the field of rational numbers,

denoted $\mathbb{Q}$, other examples of fields are the field $\mathbb{R}$ of real numbers and the field $\mathbb{C}$ of complex numbers. They also noticed that the system $\mathbb{F}[x]$ of polynomials with coefficients in a field $\mathbb{F}$ behaves a lot like the most familiar number system, the integers. Both have unique factorization into prime factors, greatest common divisors, least common multiples, etc.

Parallels

There are many parallels between the integers and the polynomials with integer coefficients. For example, we can add using very similar algorithms:

$$\begin{array}{cccc} & 1 & 2 & 3 \\ + & & 4 & 5 \\ \hline & 1 & 6 & 8 \end{array}$$

Likewise if we add the polynomials $(x^2 + 2x + 3) + (4x + 5)$

$$\begin{array}{cccc} & x^2+ & 2x+ & 3 \\ + & & 4x+ & 5 \\ \hline & x^2+ & 6x+ & 8 \end{array}$$

Similarly with the vertical multiplication algorithm allowing one modification we do not carry over the tens place when we multiply the polynomials:

$$\begin{array}{cccc} & 1 & 2 & 3 \\ \times & & 4 & 5 \\ \hline & 6 & 1 & 5 \\ 4 & 9 & 2 & 0 \\ \hline 5 & 5 & 3 & 5 \end{array}$$

Likewise if we multiply the polynomials $(x^2 + 2x + 3) \times (4x + 5)$

$$\begin{array}{cccc} & x^2+ & 2x+ & 3 \\ \times & & 4x+ & 5 \\ \hline & 5x^2+ & 10x+ & 15 \\ 4x^3+ & 8x^2+ & 12x & \\ \hline 4x^3+ & 13x^2+ & 22x+ & 15 \end{array}$$

Notice if we substitute in 10 for x we will arrive at the integer solution. Lastly the typical division algorithm:

$$
\begin{array}{r}
112 \\
11\overline{)\,1234} \\
\underline{1100} \\
134 \\
\underline{110} \\
24 \\
\underline{22} \\
2
\end{array}
\qquad\qquad
\begin{array}{r}
x^2 + x + 2 \\
x+1\overline{)\;x^3 + 2x^2 + 3x + 4} \\
\underline{-x^3 - x^2} \\
x^2 + 3x \\
\underline{-x^2 - x} \\
2x + 4 \\
\underline{-2x - 2} \\
2
\end{array}
$$

Again the two algorithms are very similar. These algorithms are not the only that have these similarities, for example multiplying using the box method also maintains these parallels. The reason for these similarities reduces to the fact that the variables in the polynomial keep track of the place values in the corresponding integer. Given this we can create a mapping and show that addition, multiplication and in some sense division are preserved.

The Mapping Φ_b

What we want to do is create a function (also called a mapping) that takes a polynomial with integer coefficients to the (corresponding) number system formed by its coefficients. For example we could ask that the mapping substitute an integer b for every "x", thereby taking the given polynomial with integer coefficients to an integer. As an example of this, if $b = 10$, the mapping would send $2x^2 + 3x + 4 \underset{\Phi_{10}}{\longrightarrow} 234$. It is common to use the letter $\mathbb{Z}$ to denote the integer number system and $\mathbb{Z}[x]$ to denote the ring of polynomials with integer coefficients.

Example We would then denote the function (mapping) described above as

$$\begin{array}{c} \Phi_{10} : \mathbb{Z}[x] \to \mathbb{Z} \\ f(x) \mapsto f(10). \end{array}$$

In other words, if $f(x) = 5x^4 + 4x^3 + 2x + 1$,

$$\Phi_{10}(f(x)) = f(10) = 5(10)^4 + 4(10)^3 + 2(10) + 1 = 5421.$$

More generally

$$\Phi_b(f(x)) = f(b)$$

Properties of Φ_b

The following proofs use two polynomials f_1, f_2, with coefficients in any of the number systems $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$.

- Φ_b preserves addition

Proof:

$$\begin{aligned}\Phi_b(f_1(x) + f_2(x)) &= f_1(b) + f_2(b) \\ &= \Phi_b(f_1(x)) + \Phi_b(f_2(x))\end{aligned}$$

- Φ_b preserves multiplication

Proof:

$$\begin{aligned}\Phi_b(f_1(x)f_2(x)) &= f_1(b)f_2(b) \\ &= \Phi_b(f_1(x))\Phi_b(f_2(x))\end{aligned}$$

- A polynomial maps to zero if and only if the polynomial is divisible by $x - b$

Proof:

Given $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$, and a number r such that

$$f(r) = 0.$$

By long division of polynomials

$$a_n x^n + \cdots + a_0 = (x - r)(c_{n-1}x^{n-1} + \cdots c_0) + \text{constant}$$

Substituting $x = r$ into the both sides of the equation above,

$$0 = 0 + \text{constant}.$$

So by necessity constant $= 0$, therefore

$$f(x) = (x - r)(c_{n-1}x^{n-1} + \cdots c_0).$$

A rough comparison of prime factorization

This section will compare prime factorization in the ring of integers and the ring of polynomials with coefficients in a field. Prime factorization is decomposing something into its constituent primes. Among the integers we have the *the fundamental theorem of arithmetic*, which says that every positive integer has a unique prime factorization. In math speak this states that any positive integer $k \geq 2$, k may be rewritten as

$$k = p_1^{l_1} p_2^{l_2} \cdots p_n^{l_n}$$

where the p_i's are the n distinct prime factors, each of order l_i.

Do the polynomials with coefficients in a field $\mathbb{F}$ have a similar analogue? In this set-up, polynomials of degree zero, that is, polynomials with only constant terms a_0, play the role of ± 1 in the integers, so that prime factors are always polynomials of degree $d \geq 1$. In grade school, polynomials are often factored by breaking up the polynomial into several terms, each of which can't be broken up further. Similarly these factored terms may be multiplied together to retrieve the original polynomial.

For example in $\mathbb{Q}[x]$

$$x^2 - 1 = (x+1)(x-1)$$

but

$$x^2 - 2$$

is "prime" in $\mathbb{Q}[x]$ since $\sqrt{2}$ is not in the number system $\mathbb{Q}$. Since $x^2 - 2$ is a polynomial not an integer instead of prime it is more common to say irreducible. In $\mathbb{R}[x]$, $x^2 - 2$ is reducible since $\sqrt{2}$ is an element of the reals, $\mathbb{R}$, i.e.,

$$x^2 - 2 = \left(x + \sqrt{2}\right)\left(x - \sqrt{2}\right).$$

Similarly the prime factorization of $f(x) = x^2 + 1$ in $\mathbb{R}[x]$ is $x^2 + 1$. Since the square of any real number is greater or equal to zero thus,

$$x^2 + 1$$

is irreducible in $\mathbb{R}[x]$. However in the system of polynomials $\mathbb{C}[x]$, $x^2 - 1$ factors as

$$x^2 + 1 = (x+i) \cdot (x-i).$$

In general, mathematicians decided that a polynomial $f(x)$ in $\mathbb{F}[x]$ is called irreducible if the polynomial cannot be written as a product

$$f(x) = g(x) \cdot h(x)$$

where $g(x)$ and $h(x)$ are polynomials in the same system $\mathbb{F}[x]$ and the degrees of $g(x)$ and $h(x)$ are both less than the degree of $f(x)$. In other words, $f(x)$ is irreducible if division by any polynomial $g(x)$ in the same system $\mathbb{F}[x]$ always leaves a remainder. So the only polynomials that divide an irreducible polynomial are the irreducible polynomial itself and the constant polynomials a_0. Remind you of anything?

Exercise: If you don't know already, try to figure out how to do prime factorization in any system $\mathbb{F}[x]$ of polynomials where $\mathbb{F}$ is one of our fields.

The 'miracle' is that every polynomial $f(x)$ in $\mathbb{R}[x]$ can be factored into irreducible factors of degrees 1 and 2. But then each irreducible factor

$$ax^2 + bx + c$$

of $f(x)$ with a, b and c real can be considered as a polynomial in $\mathbb{C}[x]$. (Remember that real numbers are also complex numbers, it's just that their imaginary part is zero). It is then possible to factor $ax^2 + bx + c$ in $\mathbb{C}[x]$ by the quadratic formula,

$$a\left(x - \frac{-b + \sqrt{b^2 - 4ac}}{2a}\right)\left(x - \frac{-b - \sqrt{b^2 - 4ac}}{2a}\right).$$

In fact, any polynomial $f(x)$ in $\mathbb{R}[x]$ be factored completely in $\mathbb{C}[x]$ as

$$f(x) = a_d \cdot (x - r_1)^{l_1}(x - r_2)^{l_2} \cdots (x - r_i)^{l_i},$$

the same is true for any $f(x)$ in $\mathbb{C}[x]$. We know this fact as the *Fundamental Theorem of Algebra*. The Fundamental Theorem of Algebra states that the only irreducible polynomials in $\mathbb{C}[x]$ are the polynomials of degree one! So, roughly speaking, every d-th degree polynomial in $\mathbb{C}[x]$ has d complex roots. These facts considered together with long division of polynomials enable use to rewrite the polynomial

$$f(x) = a_d x^d + a_{d-1}x^{d-1} + \ldots + a_0$$

as a product

$$f(x) = a_d(x - s_1) \cdot \cdots \cdot (x - s_d).$$

I.e., if the r_i's are the roots of $f(x)$ and the l_i are the multiplicities of the roots, then each root r_i occurs in the list

$$s_1, s_2, \ldots, s_d$$

exactly l_i times.

Exercise: For some small values of d, multiply out the right-hand-side of the equality

$$a_d x^d + a_{d-1} x^{d-1} + \ldots + a_0 = a_d (x - s_1) \cdot \cdots \cdot (x - s_d)$$

so that you can give a formula for each quantity $\frac{a_i}{a_d}$ as a function of the quantities $s_1, s_2, \ldots, s_d$. Can you guess the general formula for all values of d? Those formulas are called the elementary symmetric functions of $s_1, s_2, \ldots, s_d$. They will figure in an important way later on in this story.

Toward a proof of the Fundamental Theorem of Algebra

The Fundamental Thereon of Algebra is critical to much in algebra, and the proof is often waved away as being beyond the scope of college mathematics courses. In order to motivate the proof we start with an example that highlights necessary components of the proof, while giving some concreteness.

How many roots does $x^3 - r_1 x^2 + r_2 x - r_3$ have?

Consider the following polynomial

$$p(x) = x^3 - r_1 x^2 + r_2 x - r_3$$

how many roots does it have? Let's suppose that r_1, r_2, r_3 are real numbers and prove that $p(x)$ has three roots in the complex field. We will take as a given that we know that, as x goes to $+\infty$, $p(x)$ becomes positive and, as x goes to $-\infty$, $p(x)$ becomes negative. A slightly more complicated fact is the fact that $p(x)$ is a continuous function of x, which is often informally explained as the fact that you can draw the graph of

$$y = p(x)$$

without lifting your pencil from the page. So as a consequence of the fact that $p(x)$ is continuous, your pencil cannot go from negative y to positive y without crossing a place where $y = p(x) = 0$. That is, there is a real number x where $p(x) = 0$. So

$$x^3 - r_1 x^2 + r_2 x - r_3$$

has a real root s_1. As we have shown earlier, this means that

$$x^3 - r_1 x^2 + r_2 x - r_3 = (x - a_1)\left(x^2 + bx + c\right).$$

But now, again as had already been shown, this means that $x^2 + bx + c$ can be factored into linear factors in $\mathbb{C}[x]$ using the quadratic formula.

Suppose now that we don't know that the coefficients of $x^3 - r_1x^2 + r_2x - r_3$ are real but just lie in some field $\mathbb{F}$, so the operations on the coefficients of addition, subtraction, multiplication, and division hold. And suppose we know that its roots lie in that same field. Let's call these roots a_1, a_2, a_3.

Typically one would rewrite

$$x^3 - r_1x^2 + r_2x - r_3$$

as

$$(x - a_1)(x - a_2)(x - a_3).$$

Why can this polynomial be written as a product of linear factors? Recall the previous section on the relationship between prime numbers and irreducible factors of polynomials. It turns out that it is always possible to factor a polynomial in $\mathbb{F}[x]$ if $\mathbb{F}$ is a large enough field.

Field Extension

The idea of a field extension boils down to increasing the set of numbers one is able to use, while maintaining the properties of addition, subtraction and multiplication and division. For example the Pythagoreans worked in the field of rational numbers, therefore when they came across $\sqrt{2}$ they did not believe it could be a solution. We can now extend our field to all real values and now $\sqrt{2}$ is a perfectly fine solution.

5-hour Clock Arithmetic

The basics of field extension can be most easily understood by seeing first how the integers can be turned into a field through the world of clock (modular) arithmetic. Consider an algebraic structure that consists of five elements, $\{0, 1, 2, 3, 4\}$. We will call this a 5-clock arithmetic, because its arithmetic is just like what we use for the hour hand on the clock, except that there are only five hours in our 'day,' that is, the set of hours has only five elements. It is possible to add in this structure, for example, $1+1 = 2$, $1+2 = 3$, and $0+n = n$ thus 0 is the additive identity. But what about $4 + 1 =$? On the 5-hour clock we are forced to make $4 + 1 = 0$ since the hour after four o'clock is the place where the hour-hand starts over. In other words in this field the number 5 is mapped to the number 0, this implies the number 6 would be mapped to 1, and $7 \to 2$, $8 \to 3$, $9 \to 4, \ldots$. Using these mappings and induction

we can add any values in this structure and still have values in this structure. For example addition in 5-clock arithmetic looks like:

$$2 + 4 = 1$$

$$3 + 4 = 2$$

$$1 + 2 + 3 + 4 = 0$$

$$3 + 3 + 3 + 3 + 4 = 1$$

so on and so forth. The last equation in the series highlights repeated addition, we can use this model of multiplication (repeated addition) to understand how multiplication functions in 5-clock arithmetic.

$$\begin{aligned} 3 + 3 + 3 + 3 + 4 &= 1 \\ 4 \cdot 3 + 4 &= 1 \\ 4 \cdot (3 + 1) &= 1 \text{ using the distributive property} \\ 4 \cdot 4 &= 1 \end{aligned}$$

Thus the multiplicative inverse of 4 is 4. Further if $4 \cdot 4 = 1$, then multiplying both sides by 4, we get

$$\begin{aligned} \underbrace{4 \cdot 4}_{=1} \cdot 4 &= 4 \\ 1 \cdot 4 &= 4 \end{aligned}$$

There is a multiplicative identity in this structure, it is the number 1.

Let's take a closer look at multiplication of twos.

$$2 \cdot 0 = 0$$

$$2 \cdot 1 = 2$$

$$2 \cdot 2 = 4$$

$$2 \cdot 3 = 1$$

$$2 \cdot 4 = 3$$

We have a multiplicative inverse for 2 it is 3. I.e., $2 \cdot 3 = 1$, therefore $2^{-1} = 3$. Likewise there is a multiplicative inverse for $1, 3, 4$ namely $1^{-1} = 1$, $3^{-1} = 2$, $4^{-1} = 4$. We will use the multiplicative inverse to understand division in this system.

What does $2 \div 4 =?$ in 5-clock arithmetic? Rephrasing as a multiplication problem, $2 =? \cdot 4$. We know 4 has a multiplicative inverse namely itself, so we can multiply the previous equation by 4,

$$2 \cdot 4 =? \cdot 4 \cdot 4$$

$2 \cdot 4 = 3$ and $4 \cdot 4 = 1$ in this clock 5 arithmetic, therefore $3 =?$. So

$$2 \div 4 = 3$$

We can divide, because we have a multiplicative inverse.

A more efficient method for computing division exists. In fact it was the Egyptians who were one of the first peoples to record this more efficient method of division. Similar to the above process they framed it as a multiplication problem with the multiplier (or multiplicand)[1] missing. But then they used a multiplication table to deduce the solution.

From here it will help if we have a times-table to reference:

$\times$	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

What the Egyptians did was re-frame the division problem as multiplication, then read off the answer. For example, $3 \div 2 =?$ re-framed as multiplication becomes $3 =? \cdot 2$. Now we look at the times table and see that 4 times 2 gives me 3, therefore $? = 4$ and $3 \div 2 = 4$. Remember division is multiplication inverted. In other words, each division problem is really just a multiplication problem in reverse. Next we will show you a structure that does not have a multiplicative inverse.

4-hour clock arithmetic

Consider a structure whose set of numbers consists of four elements $\{0, 1, 2, 3\}$, we will call this a 4-clock arithmetic. Similar to the 5-clock structure it is possible to add these numbers, for example, $1+1 = 2$, $1+2 = 3$, $0+1 = 1$ and likewise $1+3 = 0$. In

[1]Exercise: Why does it not matter if it is the multiplier or multiplicand that is missing?

this structure we map the number 4 to 0, $5 \to 1, 6 \to 2, \ldots$etc. We can also consider multiplication take for example:

$$2 + 2 + 2 + 3 = 1$$

Again we can utilize repeated addition to understand multiplication.

$$\begin{aligned} 2 + 2 + 2 + 3 &= 1 \\ 2 \cdot 3 + 3 &= 1 \\ (2 + 1) \cdot 3 &= 1 \\ 3 \cdot 3 &= 1 \end{aligned}$$

This system has a multiplicative identity. Does it have a multiplicative inverse, can we divide? From here it will be helpful to reference a times table in this 4-clock structure:

$\times$	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

Suppose we wanted to find $3 \div 2 =?$. Previously we used the multiplicative inverse to find the solution to a division problem, so we'll try that same strategy again. we'll re-frame the problem as a multiplicative one $3 =? \cdot 2$, now we want to multiply both sides by the multiplicative inverse of 2, however we are not able to, because there is no number, which we can multiply 2 by and get 1 in this 4-clock structure. There does not exist a solution in the $4-$clock arithmetic to $3 \div 2$, for more on this see the appendix on Euclidean division.

The reason that we get a field modulo 5 but we don't get a field modulo 4 is that $4 = 2 \cdot 2$ whereas 5 is a prime number, that is, it is not the product of two integers smaller than itself. Thus in division there always exists a unique remainder. The important thing to take away from this is that clock or modulo arithmetic gives us a number system in which we can add, subtract, multiply and divide if and only if the modulus or number of hours on the clock is a prime (also called irreducible) number.

Applying the idea of field extension to polynomials

Suppose that we are in the real numbers with any polynomial of the form

$$x^3 + x^2 + x + 1$$

and we want to factor this polynomial. It looks like $x = -1$ might be a root. So we use long division to find

$$x^3 + x^2 + x + 1 = (x + 1)(x^2 + 1)$$

The polynomial $x^2 + 1$ cannot be factored in $\mathbb{R}[x]$, just like the number 5 cannot be factored in the number system $\mathbb{Z}$. Just as we found that setting the prime number 5 equal to zero led to the fact that every other number in the system has a multiplicative inverse and so the 5-hour clock system is a field, we can apply the same logic to polynomials in $\mathbb{R}[x]$, i.e., we can set

$$x^2 + 1 = 0$$

which then implies that we must set all polynomial multiples of $x^2 + 1$ equal to zero and see what happens. Well, if $g(x) \in \mathbb{R}[x]$ is any polynomial with real coefficients, then we can do long division with remainder

$$g(x) \div \left(x^2 + 1\right) = ?$$

to get

$$g(x) = h(x) \cdot \left(x^2 + 1\right) + r(x)$$

where $r(x)$ is a polynomial of degree less than two. This equation says that in our system, $x^2 + 1 = 0$ so:

$$g(x) = r(x).$$

Every element in this number system can be represented by a polynomial of degree less than two, just like every integer can be represented in 5-clock arithmetic by either 0, 1, 2, 3, or 4.

If $r(x) = 0$ then $g(x)$ is a multiple of $x^2 + 1$ and so is also zero. If $g(x)$ is not zero in this system, either $r(x)$ is a polynomial of degree one or a non-zero constant. If $r(x)$ is a non-zero constant polynomial $a_0 \in \mathbb{R}$, then it has a multiplicative inverse in this system, namely $a_0^{-1} \in \mathbb{R}$. If $r(x)$ is a polynomial of degree one, then we can do long division with remainder therefore there exists a $k(x)$ and a b_0 such that

$$\left(x^2 + 1\right) = k(x) \cdot r(x) + b_0$$

where b_0 is a constant polynomial. Notice that $b_0 \neq 0$ since, if it were zero, $x^2 + 1$ would could factor in $\mathbb{R}[x]$. So in this system

$$0 = k(x) \cdot r(x) + b_0$$

but we already know that $g(x) = r(x)$ so

$$0 = k(x) \cdot g(x) + b_0.$$

So dividing both sides by $-b_0$ and adding one we get

$$1 = \underbrace{\left(-b_0^{-1} \cdot k(x)\right)}_{g(x)^{-1}} \cdot g(x)$$

that is $\left(-b_0^{-1} \cdot k(x)\right)$ is the multiplicative inverse of $g(x)$. The notation for this system is

$$\mathbb{R}[x]/(x^2+1).$$

In short we "extend the field" by setting the un-factorable term to 0, from that it follows that the polynomial

$$x \in \mathbb{F} = \mathbb{R}[x]/(x^2+1)$$

is a root of the polynomial

$$y^2 + 1 \in \mathbb{F}[y]$$

since, substitution $x \in \mathbb{F}$ for y, we get $x^2 + 1 \in \mathbb{F}$ and in $\mathbb{F}$, $x^2 + 1 = 0$. This would be a field extension of $\mathbb{R}$, namely $\mathbb{F}$ is a field and $\mathbb{F}$ contains the field $\mathbb{R}$.[2]

For example multiplication in this field extension, $\mathbb{R}[x]/(x^2+1)$ is

$$\begin{aligned}(a+bx)(c+dx) &= (a+bx)c + (a+bx)dx \\ &= ac + bcx + adx + bdx^2 \\ &= (ac - bd + bd\left(x^2+1\right)) + (ad+bc)x \\ &= (ac-bd) + (ad+bc)x\end{aligned}$$

Usually we replace the letter x by the letter i and call this system the complex numbers. So we have obtained the complex numbers as a field extension of the real number system. We are doing this because it will turn out that we can split any polynomial up in this way (i.e. find the splitting field).

Splitting field of $x^3 - r_1x^2 + r_2x - r_3$

Suppose that $x^3 - r_1x^2 + r_2x - r_3$ is irreducible in some system of polynomials $\mathbb{F}[x]$ with coefficients in some field $\mathbb{F}$ about which we know nothing. ($\mathbb{F}$ can't be the field

[2]Effectively we are creating a new number, x, such that x is the solution to $x^2 + 1 = 0$.

of real numbers because we have seen above that any polynomial of degree three in $\mathbb{R}[x]$ has at least one real root and so can be factored in $\mathbb{R}[x]$.)

What we want to do is construct a clock arithmetic where

$$x^3 - r_1x^2 + r_2x - r_3 \equiv 0,$$

We will call this new field $\mathbb{G}$ and write $\mathbb{G}[y]$ for the system of polynomials with coefficients in $\mathbb{G}$. As we saw above, the polynomial $f(y) = y^3 - r_1y^2 + r_2y - r_3$ has the root $x \in \mathbb{G}$ since $f(x) = x^3 - r_1x^2 + r_2x - r_3 = 0$ therefore x is root of $y^3 - r_1y^2 + r_2y - r_3$.[3] For the sake of convenience with notation, we set $x = s_1$. We can then reduce the degree of the initial polynomial by long division[4] so that $y^3 - r_1y^2 + r_2y - r_3$ becomes

$$(y - s_1)(y^2 + by + c)$$

where s_1, b and c are some values in $\mathbb{G}[x]$. We complete the square on the quadratic term to find the other remaining two roots.

$$\begin{aligned}
y^2 + by + c &= 0 \\
\left(y + \frac{b}{2}\right)^2 - \frac{b^2}{4} + c &= 0 \\
\left(y + \frac{b}{2}\right)^2 &= \frac{b^2}{4} - c \\
\left(y + \frac{b}{2}\right)^2 &= \frac{b^2 - 4c}{4} \\
\left|y + \frac{b}{2}\right| &= \sqrt{\frac{b^2 - 4c}{4}} \\
y + \frac{b}{2} &= \pm\sqrt{\frac{b^2 - 4c}{4}} \\
y + \frac{b}{2} &= \pm\frac{\sqrt{b^2 - 4c}}{2} \\
y &= \frac{-b}{2} \pm \frac{\sqrt{b^2 - 4c}}{2} \\
y &= \frac{-b \pm \sqrt{b^2 - 4c}}{2}
\end{aligned}$$

[3]Since x is not divisible by $y^3 - r_1y^2 + r_2y - r_3$, it cannot be a multiple of it.

[4]Long division of polynomials is possible provided the coefficients are in a field. See the appendix on Euclidean division.

If we can solve the equation

$$y^2 - \left(b^2 - 4c\right) = 0$$

for some $y = s_0 \in \mathbb{G}$ then for simplification we write

$$s_0 = \sqrt{b^2 - 4c}.$$

Then $y^2 + by + c$ can be factored in $\mathbb{G}[y]$ and we are done. Just let $s_2 = \frac{-b+s_0}{2}$ and $s_3 = \frac{-b-s_0}{2}$, therefore the splitting field in $\mathbb{G}[y]$ is

$$y^3 - r_1 y^2 + r_2 y - r_3 = (y - s_1)(y - s_2)(y - s_3).$$

If $y^2 + by + c$ is not reducible, i.e., if we cannot solve $y^2 - (b^2 - 4c)$ in $\mathbb{G}[y]$, then we need to expand the field again. This time we set

$$y^2 + by + c \equiv 0$$

and call this new field

$$\mathbb{J} = \mathbb{G}[y] / \left(y^2 + by + c\right).$$

Then just like before $y \in \mathbb{G}$ is a root of the polynomial

$$z^2 + bz + c \in \mathbb{J}[z].$$

Again for convenience set $y = s_2$. Thus after long division of $(z^2 + bz + c) \div (z - s_2)$, we will have a first degree polynomial, so we will not need to repeat the procedure again.

To summarize: We split

$$z^3 - r_1 z^2 + r_2 z - r_3$$

into

$$(z - s_1)(z - s_2)(z - s_3)$$

and s_1, s_2, s_3 are elements of this field $\mathbb{J}$. This new field $\mathbb{J}$ may look very different than my original field, $\mathbb{F}$. We may have had to make two field extensions

$$\mathbb{F} \subseteq \mathbb{G} \subseteq \mathbb{J}$$

in order to be able to factor $x^3 - r_1 x^2 + r_2 x - r_3$ completely.

Symmetric Polynomials

A symmetric polynomial of several variables is a polynomial whose variables can be interchanged in any way without affecting the polynomial. For example,

$$x + y$$

is symmetric, if we switch x with y the expression $y + x$ is equivalent. Likewise

$$x + y + z$$

is also symmetric, we may replace x with z and z with x (or any other combination of variables) and maintain equivalency. However

$$x^2 + y$$

is not symmetric alternating the variables results in

$$y^2 + x \not\equiv x^2 + y$$

The left hand side is not equivalent to the right hand side.

The polynomial $(z-s_1)(z-s_2)(z-s_3)$ is symmetric in the three variables s_1, s_2, s_3. E.g.,

We know that in $\mathbb{J}[z]$ all the usual properties of addition and multiplication hold (except multiplicative inverse). Therefore there is nothing holding me back from distributing and multiplying:

$$(z - s_1)(z - s_2)(z - s_3) = z^3 - (s_1 + s_2 + s_3)z^2 + (s_1s_2 + s_1s_3 + s_2s_3)z - s_1s_2s_3$$

Inspecting the right hand side of the previous equation notice how s_1 and s_2 could interchange positions and maintain an equivalent expression, likewise with s_3. So each of the coefficients on the right-hand side

$$\begin{aligned} e_1(s_1, s_2, s_3) &= s_1 + s_2 + s_3 \\ e_2(s_1, s_2, s_3) &= s_1s_2 + s_1s_3 + s_2s_3 \\ e_3(s_1, s_2, s_3) &= s_1s_2s_3 \end{aligned}$$

is a polynomial whose value is unchanged if the three variables s_1, s_2, s_3 are permuted in any way. These polynomials

$$p(s_1, s_2, s_3)$$

are symmetric.

The three polynomials

$$\begin{aligned} e_1 &= s_1 + s_2 + s_3 \\ e_2 &= s_1 s_2 + s_1 s_3 + s_2 s_3 \\ e_3 &= s_1 s_2 s_3 \end{aligned}$$

are called elementary symmetric polynomials in the three variables s_1, s_2, s_3 . In the same way, the n coefficients of the polynomial

$$(x-s_1)(x-s_2)\cdots(x-s_n) = x^n - e_1(s_1,\ldots,s_n)x^2 + \cdots \pm e_{n-1}(s_1,\ldots,s_n)x \mp e_n(s_1,\ldots,s_n)$$

are called the elementary symmetric polynomials in the n variables $s_1, \ldots, s_n$.

Viéte proved that every symmetric polynomial

$$p\left(s_1, \ldots, s_n\right)$$

can be written as a polynomial

$$q\left(e_1, \ldots, e_n\right)$$

whose "variables' are elementary symmetric polynomials in the variables $s_1, \ldots, s_n$. Viéte discovered that the necessary formulas to prove this only utilize the basic properties of addition, subtraction, multiplication, and division, so that they apply for coefficients in any of the fields we might be interested in. Said otherwise, if the coefficients of $p\left(s_1, \ldots, s_n\right)$ lie in some number system $\mathbb{S}$, then the coefficients in the polynomial $q\left(e_1, \ldots, e_n\right)$ lie in the same number system $\mathbb{S}$.

Example: Using Viéte's formulas to rewrite

$$s_1^2 + s_2^2 + s_3^2$$

as the previously defined elementary symmetric polynomials. First notice how $s_1^2 + s_2^2 + s_3^2$ is symmetric in s_1, s_2, s_3. We can interchange any of the variables and we will have an equivalent expression. Next we will need the sum of squares:

$$e_1^2 = s_1^2 + s_2^2 + s_3^2 + 2s_1 s_2 + 2s_1 s_3 + 2s_2 s_3$$

Now we need to remove the cross terms, e_2 should work nicely for that

$$e_1^2 - 2e_2 = s_1^2 + s_2^2 + s_3^2 + 2s_1 s_2 + 2s_1 s_3 + 2s_2 s_3 - 2(s_1 s_2 + s_1 s_3 + s_2 s_3)$$

after zeroing out terms we arrive at

$$e_1^2 - 2e_2 = s_1^2 + s_2^2 + s_3^2.$$

Returning to the polynomial in question, $z^3 - r_1 z^2 + r_2 z - r_3$, the elementary symmetric polynomials e_1, e_2, e_3 are the same as the coefficients, r_1, r_2, r_3 respectively. For the sake of argument let us suppose the coefficients r_i are real.[5] This means we can relate the roots in our strange field $\mathbb{J}$ with the coefficients of the original polynomial that lie in $\mathbb{R}$. Since the r_1, r_2, r_3 lie in $\mathbb{R}$, i.e. the e_1, e_2, e_3 lie in $\mathbb{R}$. But we still can't make the leap to s_1, s_2, s_3 lying in $\mathbb{R}$. For example suppose

$$\begin{aligned} s_1 &= 1 - i \\ s_2 &= 1 + i \\ s_3 &= 1 \end{aligned}$$

Then the elementary symmetric polynomials would become

$$\begin{aligned} e_1 &= 3 \\ e_2 &= 4 \\ e_3 &= 2 \end{aligned}$$

The elementary symmetric polynomials are real, however only one of the three roots is real valued.

In our proof of the Fundamental Theorem of Algebra, we will need to use induction on polynomial equations whose coefficients are symmetric polynomials. For example, if we start with

$$z^3 - r_1 z^2 + r_2 z - r_3 = (z - s_1)(z - s_2)(z - s_3)$$

with the left-hand side in $\mathbb{R}[z]$ and the right-hand side in $\mathbb{J}[z]$, we will want to form the polynomial

$$G_t(z) = (z - s_1 - s_2 - t s_1 s_2)(z - s_1 - s_3 - t s_1 s_3)(z - s_2 - s_3 - t s_2 s_3).$$

Where t is an external variable that only takes real values. When we expand $G_t(z)$ we will get a polynomial in the variable z whose coefficients are polynomials in s_1, s_2, s_3

[5] As shown this would then imply $\mathbb{J} = \mathbb{G} = \mathbb{C}$

and those coefficient polynomials in the s_1, s_2, s_3 are symmetric polynomials in the s_1, s_2, s_3 and their coefficients are in the number system

$$\mathbb{S} = \mathbb{Z}[s].$$

Let's check. Here is what we get:

$$\begin{aligned}
G_t(z) = z^3 & \\
& -z^2(2s_1 + 2s_2 + 2s_3 + s_1s_2t + s_1s_3t + s_2s_3t) \\
& \quad +z(s_1s_2s_3^2t^2 + s_1s_2^2s_3t^2 + s_1^2s_2s_3t^2 + s_1s_2^2t + s_1s_3^2t + s_2s_3^2t + s_1^2s_2t \\
& \qquad +s_1^2s_3t + s_2^2s_3t + 6s_1s_2s_3t + s_1^2 + s_2^2 + s_3^2 + 3s_1s_2 + 3s_1s_3 + 3s_2s_3) \\
& \quad -s_1^2s_2^2s_3^2t^3 - 2s_1s_2^2s_3^2t^2 - 2s_1^2s_2s_3^2t^2 - 2s_1^2s_2^2s_3t^2 - s_1^2s_2^2t - s_1^2s_3^2t - s_2^2s_3^2t \\
& \qquad -3s_1s_2s_3^2t - 3s_1s_2^2s_3t - 3s_1^2s_2s_3t - s_1s_2^2 - s_1s_3^2 - s_2s_3^2 - s_1^2s_2 - s_1^2s_3 \\
& \qquad -s_2^2s_3 - 2s_1s_2s_3
\end{aligned}$$

Notice how we could switch all the s_1 with s_2 (likewise with s_1 and s_3 or s_2 and s_3) and we would end up with an equivalent equation. Thus the polynomial viewed as a function of s_1, s_2, s_3 is symmetric and $G_t(z)$ is symmetric in the coefficients s_1, s_2, s_3. Since $G_t(z)$ is symmetric then we can use Viéte's formulas and rewrite $G_t(z)$ in terms of its elementary symmetric polynomials.[6] Recall that the first three elementary symmetric polynomials are

$$\begin{aligned}
e_1 &= s_1 + s_2 + s_3 \\
e_2 &= s_1s_2 + s_1s_3 + s_2s_3 \\
e_3 &= s_1s_2s_3
\end{aligned}$$

Using the elementary symmetric polynomials we can rewrite

$$\begin{aligned}
G_t(z) = z^3 & \\
& -z^2(2e_1 + e_2t) \\
& \quad +z(e_1e_3t^2 + e_1e_2t + e_1^2 - 2e_2 + 3e_2) \\
& \qquad -e_3^2t^3 - 2e_2e_3t^2 - (e_2^2 - 2e_1e_3)t - 3e_1e_3t - e_1e_2 + e_3
\end{aligned}$$

[6] This is known as the Fundamental Theorem of Symmetric Polynomials, the proof of which is a bit beyond this paper. Please see the excellent paper by Ben Blum-Smith and Samuel Coskey, "The Fundamental Theorem on Symmetric Polynomials: History's First Whiff of Galois Theory": `https://arxiv.org/abs/1301.7116` for more exposition.

The coefficients of the elementary symmetric polynomials are real valued. Furthermore the elementary symmetric polynomials are real valued. Since $G_t(z)$ is a polynomial in the variables z and t with coefficients that are symmetric in the variables s_1, s_2, s_3, Viéte's theorem tells us that $G_t(z)$ may be rewritten as a polynomial in the variables z and t with coefficients in the number system $S = \mathbb{Z}[e_1, e_2, e_3]$. Now recall that the e_1, e_2, e_3 are just coefficients of the original polynomial,

$$x^3 - r_1x^2 + r_2x - r_3$$

i.e., $e_1 = r_1, e_2 = r_2, e_3 = r_3$. Since e_1, e_2, e_3 are real numbers, that means that $S = \mathbb{Z}[e_1, e_2, e_3]$ is a subset of $\mathbb{R}$, i.e., $G_t(z)$ is a polynomial in the variables z and t with real valued coefficients!

Complex number review

Recall that to take the conjugate of a complex number you just switch the sign of the imaginary part (e.g. if $z = 1 + 2i$ then the conjugate of z is $\bar{z} = 1 - 2i$). So a complex number $a + bi$ is in fact a real number if and only if $b = 0$, that is, if and only if

$$a + bi = \overline{a + bi}.$$

Also, in the complex numbers, the sum of conjugates equals the conjugate of the sum and the product of conjugates equals the conjugate of the product.

Something neat also happens when you multiply a complex number with its conjugate. Let's let $z = a + bi$ be any complex number, then the conjugate is $\bar{z} = a - bi$ and if we multiply them together we get

$$z\bar{z} = a^2 + abi - abi - (bi)^2$$

which becomes

$$z\bar{z} = a^2 + b^2$$

which is a real number, i.e., the imaginary part is 0 (or there is no imaginary part). Furthermore it is a positive real number.

The Fundamental Theorem of Algebra

So now we are ready to begin our proof of the Fundamental Theorem of Algebra, namely the theorem that says that any polynomial of degree $d > 0$ with complex coefficients has at least one complex root. This proof originally appeared in Samuel (2013), pg. 45.

Reduction to polynomials with real coefficients

We start with any polynomial of degree d with coefficients which are complex numbers. We can always divide through by the coefficient of x^d to reduce our polynomial to one of the form

$$p(x) = x^d + a_{d-1}x^{d-1} + \ldots + a_1 x + a_0 \in \mathbb{C}[x].$$

We need to show that there is a complex number z_1 that is a root of this polynomial. We start by forming the polynomial

$$P(x) = \left(x^d + a_{d-1}x^{d-1} + \ldots + a_1 x + a_0\right)\left(x^d + \overline{a_{d-1}}x^{d-1} + \ldots + \overline{a_1}x + \overline{a_0}\right).$$

Multiplying out the right-hand side we get

$$\begin{gathered} P(x) = x^{2d} + \left(a_{d-1} + \overline{a_{d-1}}\right)x^{2d-1} + \\ \ldots \\ + \left(a_1\overline{a_0} + a_0\overline{a_1}\right) + a_0\overline{a_0}. \end{gathered}$$

Now each coefficient in $P(x)$ has the property that its conjugate is itself–that just follows from the fact that the conjugate of a sum is the sum of the conjugates and the conjugate of a product is the product of the conjugates. So all the coefficients in the polynomial $P(x)$ are real! Also, it suffices to find a complex root z_1 of $P(x)$ since such a z_1 is either a root of

$$x^d + a_{d-1}x^{d-1} + \ldots + a_1 x + a_0$$

or it is a root of

$$x^d + \overline{a_{d-1}}x^{d-1} + \ldots + \overline{a_1}x + \overline{a_0}.$$

But in this second case

$$\begin{gathered} z_1^d + \overline{a_{d-1}}z_1^{d-1} + \ldots + \overline{a_1}z_1 + \overline{a_0} = 0 \\ \overline{z_1^d + \overline{a_{d-1}}z_1^{d-1} + \ldots + \overline{a_1}z_1 + \overline{a_0}} = \overline{0} = 0 \\ \overline{z_1}^d + a_{d-1}\overline{z_1}^{d-1} + \ldots + a_1\overline{z_1} + a_0 = 0 \end{gathered}$$

so we can put $z_0 = \overline{z_1}$. So it suffices to show that every polynomial with real coefficients has a complex root.

Roots of polynomials with real coefficients

Now suppose that we have any polynomial

$$x^d + r_{d-1}x^{d-1} + \ldots + r_1 x + r_0$$

with real coefficients. We can always factor

$$d = 2^k \cdot m$$

with m odd. We will show that for any value k there is at least one complex root of this polynomial. We will do this by inducting on k.

Base Case

For the base case $k = 0$, the polynomial has odd degree, and, as we have explored above, the end behavior of odd polynomials with positive leading coefficient is $x \to -\infty$ the value of the polynomial goes to negative infinity and as $x \to \infty$ the value of the polynomial goes to positive infinity. Recall that a polynomial is continuous, there are no jumps, therefore the polynomial must cross the $x-$axis at some place in between $(-\infty, \infty)$ therefore there is at least one real root. (We give a more rigorous proof in the Appendix.) This real root of course counts as our complex root since every real number is a complex number. This argument takes care of any polynomials with odd degree.

Induction Step

Now for the induction step, where we will use the induction hypothesis that every polynomial with real coefficients and degree $d = 2^{k-1}m'$ (where m' is odd) has at least one complex root.

As we have seen above, there is *some* field $\mathbb{F} \supseteq \mathbb{C} \supseteq \mathbb{R}$ so that we can factor

$$x^d + r_{d-1}x^{d-1} + \ldots + r_1 x + r_0 = (x - x_1) \cdot \ldots \cdot (x - x_d)$$

with all the $x_i \in \mathbb{F}$.

Here comes the ingenious step! Let s be an arbitrary real number and let $y_{s,i,j} = x_i + x_j + s x_i x_j$, where $1 \leq i < j \leq d$. Now define:

$$G_s(x) = (x - y_{s,1,2})(x - y_{s,1,3}) \cdots (x - y_{s,1,d})(x - y_{s,2,3}) \cdots (x - y_{s,2,d}) \cdots (x - y_{s,d-1,d})$$

This may be succinctly written as:

$$G_s(x) = \Pi_{1 \leq i < j \leq d}(x - y_{s,i,j})$$

$G_s(x)$ is a polynomial in the variables x and s whose coefficients are in the number system $\mathbb{Z}[x_1, \ldots, x_d]$. But each of those coefficients is a symmetric polynomial in $x_1, \ldots, x_d$ since it doesn't change under any permutation of the x_i. So, as in the previous example, by Viéte's theorem each coefficient in $G_s(x)$ is a polynomial in the variables x and s with coefficients in the elementary symmetric functions $e_1, \ldots, e_d$ of $x_1, \ldots, x_d$. Again as in the previous example, $e_1, \ldots, e_d$ are real numbers! So $G_s(x)$ is a polynomial in the variables x and s with coefficients that are real numbers!

But what is the degree of $G_s(x)$ in the variable x? We get exactly one term for each way of choosing two distinct numbers out of the set $\{1, \ldots, d\}$. You may recognize this as the 'choose number' $\binom{d}{2}$.

Recall that $\binom{d}{2} = \frac{d(d-1)}{2}$ (see appendix for derivation), then by substitution:

$$\frac{d \cdot (d-1)}{2} = \frac{2^k \cdot m \cdot \left(2^k \cdot m - 1\right)}{2} = 2^{k-1} \cdot m \cdot \left(2^k \cdot m - 1\right).$$

But m is odd and, since $k > 0$, $\left(2^k \cdot m - 1\right)$ is also odd and so $m' = m \cdot \left(2^k \cdot m - 1\right)$ is also odd. So, by the induction hypothesis, for any fixed real number s, the polynomial $G_s(x) \in \mathbb{R}[x]$ has a complex root! So, for each real number s, at least one of the $y_{s,i,j}$ must be a complex number!

We do not know which

$$x_i + x_j + s x_i x_j$$

is complex for a given s but there are infinitely many real numbers s and only finitely many pairs ij so there must be some ij such that

$$x_i + x_j + s x_i x_j$$

is a complex number, z_s, for an infinite number of real numbers s. Pick two of those, say s' and s'' . So we have a system of two linear equations

$$\begin{aligned} \left(x_i + x_j\right) + s' x_i x_j &= z' \in \mathbb{C} \\ \left(x_i + x_j\right) + s'' x_i x_j &= z'' \in \mathbb{C} \end{aligned}$$

in two unknowns $b = \left(x_i + x_j\right)$ and $c = x_i x_j$. Solving the system of two linear equations in two unknowns, we conclude that since s', s'', z', and z'' all lie in the complex number system, so do the unknowns $b = \left(x_i + x_j\right)$ and $c = x_i x_j$. Finally consider the quadratic equation

$$x^2 - (x_i + x_j)x + x_i x_j = x^2 - b \cdot x + c = 0$$

with complex coefficients. Applying the quadratic formula, the solutions are

$$x = \frac{b \pm \sqrt{b^2 - 4c}}{2}.$$

On the other hand

$$\begin{aligned} x^2 - (x_i + x_j)x + x_i x_j &= (x - x_i)\left(x - x_j\right) \\ &= \left(x - \frac{b \pm \sqrt{b^2 - 4c}}{2}\right)\left(x - \frac{b \mp \sqrt{b^2 - 4c}}{2}\right) \end{aligned}$$

So

$$\begin{aligned} x_i &= \tfrac{b \pm \sqrt{b^2 - 4c}}{2} \\ x_j &= \tfrac{b \mp \sqrt{b^2 - 4c}}{2}. \end{aligned}$$

We know b and c are complex so to show that x_j and/or x_i is complex we only need to show that the square root of a complex number $z = b^2 - 4c$ is again a complex number.

Proof the square root of a complex number is complex

Write z in polar coordinates as $p = r \cdot (\cos(\theta) + i \sin(\theta))$. We seek a complex number w such that $w^2 = p$. Using the sum of angles formula from trigonometry, we compute

$$\begin{gathered} \left(r^{1/2} \cdot (\cos(\theta/2) + i \sin(\theta/2))\right)^2 = \\ \left(r^{1/2}\right)^2 \cdot ((\cos(\theta/2) + i \sin(\theta/2)))^2 = \\ r \cdot \left(\left(\cos^2\left(\theta/2\right) - \sin^2\left(\theta/2\right)\right) + 2i \sin(\theta/2) \cdot \cos(\theta/2)\right) \\ r \cdot (\cos(\theta) + i \sin(\theta)) = p. \end{gathered}$$

So the complex number

$$w = r^{1/2} \cdot (\cos(\theta/2) + i \sin(\theta/2))$$

is the square root of the complex number p. Likewise both x_i and x_j are complex.

So we have finished the proof of the Fundamental Theorem of Algebra. Namely we have shown that every polynomial with coefficients which are complex numbers

$$p\left(x\right) = x^d + a_{d-1}x^{d-1} + \ldots + a_1 x + a_0 \in \mathbb{C}\left[x\right]$$

has at least one root that is a complex number. So we are almost done with the story. But we can say a bit more.

Every single variable polynomial of degree d in $\mathbb{C}[x]$ has exactly d roots Let's prove this as a corollary of the Fundamental Theorem of Algebra. Set:

$$f_0(x) = a_d x^d + a_{d-1} x^{d-1} + \cdots + a_1 x + a_0$$

where the a_i are complex. We know by the Fundamental Theorem of Algebra that $f_0(x)$ has at least one complex root, let's call it r_1. But since $\mathbb{C}$ is a field, we can divide polynomials in $\mathbb{C}[x]$ by the linear polynomial $x - r_1$ so that we can write

$$f_0(x) = (x - r_1) \cdot \underbrace{f_1(x)}_{\text{quotient}} + \underbrace{b_0}_{\text{remainder}} .$$

Substituting we have

$$0 = f_0(r_1) = (r_1 - r_1) f_1(r_1) + b_0 = b_0.$$

So in fact

$$f_0(x) = (x - r_1) \cdot f_1(x)$$

with $f_1(x) \in \mathbb{C}[x]$.

The degree of $f_0(x)$ was d, therefore the degree of $f_1(x)$ must be $d - 1$, since

$$\underbrace{f_0(x)}_{\text{degree } d} = \underbrace{(x - r_1)}_{\text{degree } 1} f_1(x)$$

and exponents add.

Now we apply the Fundamental Theorem of Algebra to $f_1(x)$ this gives us a root we will call r_2. We divide $f_1(x)$ by $(x - r_2)$ to get $f_2(x)$. Again the remainder must be zero and the degree of $f_2(x)$ will be $d - 2$. Now we have found two roots and we've reduced the polynomial by two degrees. Therefore if we repeat this process d times we will have exactly d roots and we will have reduced the polynomial to degree $d - d = 0$, for which no roots will exist. Thus we have found exactly d roots for a polynomial of degree d. In fact in $\mathbb{C}[x]$ we have the complete factorization

$$p(x) = x^d + a_{d-1} x^{d-1} + \ldots + a_1 x + a_0 = (x - r_1) \cdot \cdots \cdot (x - r_d).$$

Another way to say this is that the only irreducible polynomials in $\mathbb{C}[x]$ are linear polynomials $ax + b$.

Conclusion

Analogies often play critical roles in mathematics. This paper explored the analogies between polynomials and integers. We reviewed prime factorization and the fundamental theorem of arithmetic. This led us to find an analogue of the fundamental theorem of arithmetic for polynomials, which turned out to be the key to to establish the fundamental theorem of algebra!

Appendices

Division

Division in the language of abstract algebra is often defined as the solution, s to the equation $a \cdot s = t$.[7] This is similar to how the first peoples framed division. When we use the word division we often think of "chunking" values or items, yet considering division in this way is only true for integers. When we compute division we are calculating the quotient and the remainder. The reason we know we are able to do this for any integer is attributable to Euclidean division.

Euclidean Division

Euclidean division for integers

Euclidean division is a theorem that states:

Given two positive integers $a \geq b$, there is a unique positive integer q and non-negative integer r such that $r < b$ and $a = q \cdot b + r$

Proof of Euclidean division for integers

We will induct on n, such that $a \leq n$

Base case $n = 1$**:** $1 = 1 \cdot 1 + 0$

[7] Technically this would be the definition for left division, right division could be defined as the solution, k to the equation $k \cdot a = t$.

Induction step: Assume there is a unique positive integer q and non-negative integer r such that $r < b$ and $a = q \cdot b + r$ for $a = n - 1$. This implies $n - 1 = q \cdot b + r$, which may be rewritten as $n = q \cdot b + (r + 1)$.

The Euclidean division theorem tells us that for any integer division problem, not only does a quotient and remainder exist but there is a unique quotient and remainder. This is a very important fact, one that is not true for the $4-$clock number system where we saw a specific instance where a solution did not exist. This theorem is often taken for granted because we have had so much experience with division of integers, but the contrast with $4-$clock arithmetic highlights its importance.

Euclidean division of polynomials

Euclidean division has also been generalized for division of polynomials. It is typically stated as:

Given two polynomials $a(x)$ and $b(x)$ such that $deg(a(x)) \geq deg(b(x))$, there are unique polynomials $q(x)$ and $r(x)$ such that $deg(r(x)) < deg(b(x))$ and $a(x) = q(x) \cdot b(x) + r(x)$

Proof of Euclidean division of polynomials

We will induct on n such that $deg(a(x)) \leq n$.

Base case $n = 1$**:** Then $deg(a(x)) = 1$, this implies $b(x) = b_0 =$ constant. Further, $a_1x + a_0 = b_0q(x) + r(x)$ implies $q(x) = a_1b_0^{-1}x + a_0b_0^{-1}$ and $r(x) = 0$. Then $deg(r(x)) = 0 < deg(b(x)) = 1$.

Induction step: For $a'(x)$ with degree $n - 1$ and $b(x)$ with $deg(b(x)) \leq n - 1$, assume there are unique polynomials $q'(x)$ and $r'(x)$ such that $deg(r'(x)) < deg(b(x))$ and $a'(x) = q'(x) \cdot b(x) + r'(x)$.

We start by subtracting $\frac{a_n}{b_m} \cdot x^{n-m} \cdot b(x)$ from $a(x)$, where $deg(a(x)) = n$ and

$deg(b(x)) = m \leq n-1$.

$$\begin{aligned} a(x) - \frac{a_n}{b_m} \cdot x^{n-m} \cdot b(x) &= a_n x^n + \cdots + a_0 - \frac{a_n}{b_m} \cdot x^{n-m} \cdot b(x) \\ &= a_n x^n + \cdots + a_0 - \frac{a_n}{b_m} x^{n-m}(b_m x^m + b_{m-1} x^{m-1} \cdots + b_0) \\ &= a_n x^n - a_n x^n + (a_{n-1} - \frac{a_n}{b_m} b_{m-1}) x^{n-1} + \cdots \\ &\quad + (a_{n-m} - \frac{a_n}{b_m} b_0) x^{n-m} + a_{n-m-1} x^{n-m-1} + \cdots + a_0 \\ &= (a_{n-1} - \frac{a_n}{b_m} b_{m-1}) x^{n-1} + \cdots + a_0 \end{aligned}$$

The right hand side has degree $n-1$ therefore by the induction hypothesis there exists some $q'(x), r'(x)$ such that

$$\begin{aligned} a(x) - \frac{a_n}{b_m} \cdot x^{n-m} \cdot b(x) &= q'(x) \cdot b(x) + r'(x) \\ \Rightarrow a(x) &= (q'(x) + \frac{a_n}{b_m} \cdot x^{n-m}) \cdot b(x) + r'(x) \end{aligned}$$

Thus $q(x) = (q'(x) + \frac{a_n}{b_m} \cdot x^{n-m})$ and $r(x) = r(x)'$.

Note that Euclidean division theorem does not state how to perform the division. That is the topic of division algorithms.

Division Algorithm

Many division algorithms exist, the one with which most individuals are familiar with is long division, which is commonly confused with the concept of division itself. Other algorithms for division exist. One of them is remarkably simple and often overlooked. For example, if a model for multiplication is repeated addition then a model for division could be repeated subtraction.

Example: Divide 21 by 4 using repeated subtraction. Subtract 4 from 21, $21-4 = 17$, repeat with the new result $17-4=13$, repeat $13-4=9$, $9-4=5$, $5-4=1$, stop because $4 > 1$. So 21 divided by 4 has a remainder 1, and the quotient is 5, because we subtracted 4 five times.

There are many other algorithms for computing the quotient and remainder in a division problem. All of which rely on Euclidean division.

Odd degree real valued polynomials with real coefficients have a real root

In the base case of the proof of the fundamental theorem of algebra it was argued that odd degree polynomials with real coefficients have a real root. A rigorous proof is provided here. First we offer some preliminaries: we define what it means to be continuous as well as the property that a non-empty set with an upper bound must have a least upper bound is given. From here we then prove that polynomials are continuous. Then by invoking the previously stated property we can show that the least upper bound is the root of the polynomial.

Preliminaries

Definition of a real valued function which is continuous at a point

A real valued function, f is continuous at some real number c if the limiting value of $f(x)$ as x approaches c is equal to $f(c)$ (Lang, 2013), i.e.,

$$\lim_{x \to c} f(x) = f(c).$$

Bounding values

An upper bound of a set is an element that is greater than or equal to every element of the set. For example in the set $\{0, 1, 2, 3, 4\}$ an upper bound is 4, but 5 is also an upper bound. The set of upper bounds (in the integers) for this example is the set, $U = \{4, 5, 6, \ldots\}$. This is where the least upper bound comes in. In order to differentiate between all the possible upper bounds; the least upper bound is defined to be the smallest of all the upper bounds. If we consider the previous example then the smallest value of U is 4, thus the least upper bound of the set $\{0, 1, 2, 3, 4\}$ is 4.

Least upper bound property

The least upper bound property is a statement that if certain sets have an upper bound then they must have a least upper bound. In terms of real values the least upper bound property is often stated as:

Any non-empty set of real numbers that has an upper bound must have a least upper bound in the real numbers.

Proof that odd degree real valued polynomials with real coefficients have a real root

Polynomials are continuous functions

Here we prove that polynomials are continuous. We will need the sum and product rule for limits:

Let f, g be real valued $\lim_{x\to c} f(x) = l$ and $\lim_{x\to c} g(x) = k$. Then: $\lim_{x\to c} (f(x) + g(x)) = l + k$.

Let f, g be real valued $\lim_{x\to c} f(x) = l$ and $\lim_{x\to c} g(x) = k$. Then: $\lim_{x\to c} (f(x)g(x)) = lk$.

We will also need that the identity function $f(x) = x$ is continuous and that any constant function $f(x) = c$.

But any polynomial is just made by taking sums and products many times, starting with the identity function and constant functions, so we are done.

Asymptotic behavior of an odd degree polynomial

Let $P(x) = x^d + r_{d-1}x^{d-1} + \cdots + r_1 x + r_0$, where d is odd. We want to show that $P(x)$ has a real root.

$$x^d + r_{d-1}x^{d-1} + \cdots + r_1 x + r_0.$$

Recall: Since d is odd,

$$\lim_{x\to-\infty} x^d \to -\infty$$

and furthermore

$$\lim_{x\to-\infty} 1 + \frac{r_{d-1}}{x^{d-1}} + \cdots + \frac{r_1}{x} + \frac{r_0}{x^d} = 1 + 0 + \ldots + 0 + 0 = 1.$$

Thus:

$$\begin{aligned}
\lim_{x\to-\infty} P(x) &= \lim_{x\to-\infty} x^d + r_{d-1}x^{d-1} + \cdots + r_1 x + r_0 \\
&= \lim_{x\to-\infty} x^d \left(1 + \tfrac{r_{d-1}}{x^{d-1}} + \cdots + \tfrac{r_1}{x} + \tfrac{r_0}{x^d}\right) \\
&= \lim_{x\to-\infty} x^d \cdot \lim_{x\to-\infty} \left(1 + \tfrac{r_{d-1}}{x^{d-1}} + \cdots + \tfrac{r_1}{x} + \tfrac{r_0}{x^d}\right) \\
&= 1 \cdot (-\infty) = -\infty.
\end{aligned}$$

The same argument shows that $\lim_{x\to\infty} P(x) = \infty$.

Having shown that $\lim_{x\to-\infty} P(x) = -\infty$ and that $\lim_{x\to\infty} P(x) = \infty$, we consider the set B of all x values, where $P(x) < 0$, i.e., $B = \{x \in \mathbb{R} | P(x) < 0\}$. We know that B is not the empty set, because $\lim_{x\to-\infty} P(x) = -\infty$. Furthermore we know B has an upper bound, because $\lim_{x\to\infty} P(x) = \infty$. Since B has an upper bound, it must therefore have a least upper bound, call this value b.

Putting it all together

Considering again $B = \{x \in \mathbb{R} | P(x) < 0\}$, for each counting number i, there is a number $b_i \in B$ such that

$$b_i > b - \frac{1}{i}$$

since otherwise $b - \frac{1}{i}$ would be an upper bound for B contradicting the fact that b is the least upper bound. But this means that, for each i,

$$b - \frac{1}{i} < b_i \leq b$$

$$b = \lim_{i \to \infty} \left(b - \frac{1}{i} \right) \leq \lim_{i \to \infty} b_i \leq \lim_{i \to \infty} b = b.$$

So

$$\lim_{i \to \infty} b_i = b$$

and so, since $P(x)$ is continuous,

$$\lim_{i \to \infty} P(b_i) = P(b).$$

Furthermore, since $P(b_i) < 0$ for all i,

$$\lim_{i \to \infty} P(b_i) \leq 0.$$

Putting these two last facts together, we conclude that

$$P(b) \leq 0.$$

On the other hand for each counting number i, there is a number c_i such that $b < c_i < b + \frac{1}{i}$. So

$$\lim_{i \to \infty} c_i = b$$

by the same argument we used above. And so, since $P(x)$ is continuous,

$$\lim_{i \to \infty} P(c_i) = P(b).$$

But $b < c_i$ means that $c_i \notin B$ and so $P(c_i) \geq 0$. But this last inequality implies that

$$\lim_{i \to \infty} P(c_i) \geq 0.$$

So

$$P(b) \geq 0.$$

But the only number that is both less than or equal to zero and greater than or equal to zero is 0. So

$$P(b) = 0$$

and b is the root of $P(x)$ that we were looking for.

Degree of $G_s(x)$

In the proof of the fundamental theorem of algebra we showed that the degree of $G_s(x)$ is $2^{k-1}m'$ using a combinatorial argument, provided here is an alternative way to derive it.

There are $d-1$ terms that have $i = 1$, then there are $d-2$ terms that have $i = 2$, because $i < j$. So there are $d-3$ terms that have $i = 3$ and so on.

$$G_s(x) = \underbrace{(x - y_{s,1,2})(x - y_{s,1,3}) \cdots (x - y_{s,1,d})}_{d-1} \underbrace{(x - y_{s,2,3}) \cdots (x - y_{s,2,d})}_{d-2} \cdots \underbrace{(x - y_{s,d-1,d})}_{1}$$

So the degree of $G_s(x)$ is

$$(d-1) + (d-2) + \ldots + 2 + 1.$$

But doubling the series gives:

$$\begin{gathered}(d-1) + (d-2) + \ldots + 2 + 1 + \\ 1 + 2 + \ldots (d-2) + (d-1) \\ = d \cdot (d-1)\end{gathered}$$

since we doubled the series we must divide by 2, thus there are:

$$\binom{d}{2} = \frac{d(d-1)}{2}$$

terms in $G_s(x)$.

The Mapping Φ

Let $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$, where $a_i \in \mathbb{Z}$, for $i = 0, 1, 2, \ldots, n$.

Define:

$$\Phi_b(f(x)) = \int f(x)\delta(b-x)dx = f(b)$$

where δ is the dirac delta, and b is the base of the integer representation we are mapping into.

So if we were mapping into the base-10 representation of the integers, denoted $\mathbb{Z}_{10}$, we would have:

$$\Phi_{10}(f(x)) = \int f(x)\delta(10-x)dx = f(10)$$

It should be stated that Φ maps the polynomials with integer coefficients to the integers.

Φ preserves addition and multiplication

Given two polynomials with integer coefficients f_1, f_2.

Preserves addition

$$\begin{aligned}
\Phi_b(f_1(x) + f_2(x)) &= \int (f_1(x) + f_2(x))\delta(b-x)dx \\
&= \int f_1(x)\delta(b-x)dx + \int f_2(x)\delta(b-x)dx \\
&= f_1(b) + f_2(b) \\
&= \Phi_b(f_1(x)) + \Phi_b(f_2(x))
\end{aligned}$$

Preserves multiplication

$$\begin{aligned}
\Phi_b(f_1(x)f_2(x)) &= \int f_1(x)f_2(x)\delta(b-x)dx \\
&= f_1(b)f_2(b) - \int f_1'(x)g_2(x)dx \text{ By integrating by parts} \\
&= f_1(b)f_2(b) \\
&= \Phi_b(f_1(x))\Phi_b(f_2(x))
\end{aligned}$$

Where $\int f_1'(x)g_2(x)dx = 0$ because $g_2(b) = f_2(b)$[8] when $x = b$, but $g_2(x)$ is zero everywhere else, so the integral has measure zero.

References

Lang, S. (2013). *Undergraduate analysis*. Springer Science & Business Media.

Samuel, P. (2013). *Algebraic theory of numbers: Translated from the French by Allan J. Silberger*. Courier Corporation.

[8] $f_2(b) \in \mathbb{Z}$