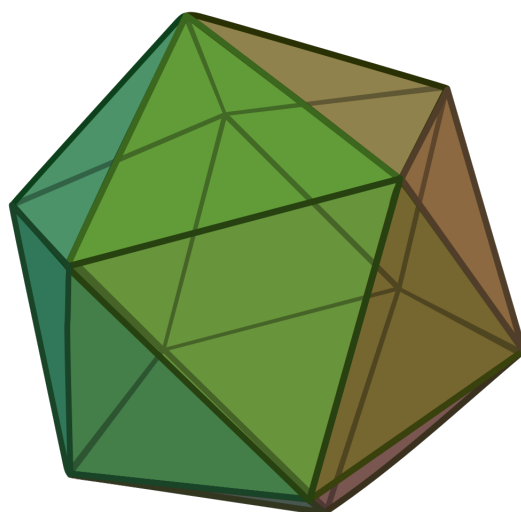


Lecture Notes for Math 3010

Symmetries and Symbols

Algebra from Ancient Times to Galois



Jason Underdown
April 14, 2014

Contents

Preface	iii
Chapter 1. Symmetries	1
1.1. Symmetries of the Equilateral Triangle	2
1.1.1. Properties of Symmetries	7
1.1.2. Associativity	7
Chapter 2. Groups	9
2.0.3. Cayley Tables	10
2.1. Groups and Algebra	12
2.2. The Symmetric Group: S_n	13
2.2.1. The Order of S_n	14
2.2.2. Composition of Permutations	14
2.2.3. Cycle Notation	14
2.2.4. Composing/Multiplying Cycles	17
2.2.5. Cycle Inverses	19
2.2.6. Equivalence of S_3 and D_6	19
2.3. Homomorphisms	21
2.3.1. Normal Subgroups	23
2.3.2. Isomorphisms	23
2.4. Equivalence Relations and Partitions	23
2.5. Cosets and Lagrange's Theorem	24
2.6. Simple Groups	25
Chapter 3. Ancient Algebra	31
3.1. Egypt	31
3.1.1. Multiplication by Doubling	32

3.1.2. Ahmes	32
3.2. Ancient Mesopotamian Mathematics	32
3.2.1. Systems of Equations	33
3.2.2. The Ancient Method of Completing the Square	35
3.2.3. The Modern Method of Completing the Square	36
3.3. Ancient Greece and Alexandria	38
3.3.1. Pythagoras	38
3.3.2. Euclid	39
3.3.3. Diophantus	43
Chapter 4. Medieval Algebra	45
4.1. Medieval Persia	45
4.1.1. Al-Khwārizmī	45
4.1.2. Omar Khayyam	46
4.2. Medieval Italy	48
4.2.1. Leonardo of Pisa [1170–1240]	48
Chapter 5. Renaissance Algebra	49
5.1. Italy	49
5.1.1. Luca Pacioli [1445–1517]	49
5.1.2. The Story of the Cubic	49
5.1.3. Scipione del Ferro [1456–1526]	51
5.1.4. Niccolo Fontana Tartaglia [1499–1557]	51
5.1.5. Girolamo Cardano [1501–1576]	51
5.2. Modern Derivation of Cardano’s Formula	52
5.3. Ferrari’s Solution to the Biquadratic	54
5.3.1. Lodovico Ferrari [1522–1565]	56
5.3.2. Rafael Bombelli [1526–1572]	56
5.4. Extending Cardano’s Formula with Complex Numbers	56
5.5. France	58
5.5.1. François Viète [1540–1603]	59
5.5.2. René Descartes [1596–1650]	60
Chapter 6. Symmetric Polynomials	61
6.1. Generators for S_n	62
6.2. Fundamental Theorem of Symmetric Polynomials	63
6.3. Generalizing the Solution Method	63
Bibliography	67
Index	68
Index	69

Preface

The study of polynomial equations dates back to ancient times. Every math student learns to factor polynomials as a tool for solving such equations. We also learn the quadratic formula, the Rational Roots Theorem, polynomial long division, and sometimes a few other algorithms. Armed with these tools, one is able to solve a good number of polynomial equations. However, the question naturally arises as to whether these tools, supplemented with the taking of roots ($\sqrt[n]{}$), suffice to solve all polynomial equations?

Before we proceed however, let's be careful about what we mean by, "solving" an equation. Numerical techniques for approximating the real roots of an equation have existed since ancient times, and the modern Newton-Raphson method of root finding is extremely fast and easy to implement in computer code. For nearly all applications, approximate solutions suffice. The question of "solving" or "solvability" boils down to whether or not we can express the roots or solutions of an equation via the ordinary algebraic operations of $+$, $-$, \times , \div and by taking roots, $\sqrt{}$, $\sqrt[3]{}$, $\sqrt[4]{}$, $\sqrt[5]{}$, \dots . For example, the solutions to the equation $x^2 = 2$ are $x = \pm\sqrt{2}$. Notice that these two roots are just convenient ways of writing irrational numbers, i.e. numbers with infinite decimal expansions. If you wish to place these numbers on the real number line you will still have to employ the square root algorithm to approximate the value of $\sqrt{2}$.

If numerical root finding techniques of solving equations suffice, then what is the point of asking whether the solutions of a general equation can be written via the four arithmetic operations and the taking of roots? From a purely practical perspective there is no reason. But mathematics is not just about practicality, curiosity is also an important driver of mathematical thought.

Curiosity mixed with a desire for prestige led Italian mathematicians to find formulas similar to the quadratic formula for solving cubic and quartic polynomials in the sixteenth century. Naturally, mathematicians believed that with more work similar formulas could be found for solving quintic equations and beyond. It wasn't until 1824 when a young Norwegian mathematician by the name of Niels Henrik Abel proved definitively that quintic polynomial equations do *not* have a general solution method.

At that time, some mathematicians were beginning to doubt whether a formula for quintic equations would ever be found, but most assumed it would eventually happen thus Abel's proof was quite a shock.

Interestingly, it was an even younger, French mathematician by the name of Evariste Galois who discovered around 1830 that there is no *general* formula for any polynomial equation of degree five or higher. More importantly, Galois discovered a criterion for deciding a polynomial equation's solvability.

To be clear, there are certainly polynomial equations of degree five and higher whose solutions can be written using just the normal algebraic operations of $+$, $-$, \times , \div and the taking of roots. For example $x^5 = 3$ has the single repeated root $\sqrt[5]{3}$. What Galois showed is that there is no formula or algorithm which will allow you to generate an algebraic expression for the roots of an arbitrary equation. In the degree two case, the quadratic formula tells us that the roots of the arbitrary equation $ax^2 + bx + c = 0$ are:

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

But when the degree of the equation is five or higher one cannot find a formula that works for every equation.

Galois effectively finished what was commonly called "equation theory", but the closing of this one door led to the opening of two new doors. Galois' criterion uses an algebraic object known as a *group*. His usage of groups led to what is now called "abstract algebra". Basically this is the study of groups and other related objects such as rings, fields, vector spaces and modules. The other door he opened was to finding applications of Galois Theory to questions other than the solvability of polynomial equations. It turns out there are several domains of math to which Galois Theory applies. Most notable is perhaps the study of differential equations started by Sophus Lie in the late nineteenth century and continuing to this day.

The purpose of this book is to trace the historical development of equation theory from ancient times up through Galois' amazing discovery. You might think of these notes as an historical study of Algebra. The topics are chronologically introduced with the exception of groups and fields. Since it is a bit much to digest Galois Theory without a thorough understanding of groups and fields, we begin with them in the hope that will afford

a pedagogically superior experience to the student. Also, groups are introduced from a geometric perspective (via Dihedral groups) rather than just the permutation perspective to reinforce their connection to symmetry. It seems a shame to not exploit the visual intuitiveness of how a group acts on a regular polygon.

It should be noted that most expositions of Galois Theory follow Artin's approach which relies heavily on Linear Algebra. This book instead teaches old fashioned Galois Theory without any need for the vector space notions which were not known at the time Galois wrote his memoir.

Symmetries

In this chapter we ask the question, “What structure is absolutely necessary to solve simple equations?” By simple, we mean an equation in one unknown and which uses only one operator, for example, $2 + x = 6$. This will lead us to the definition of a group. The group concept is one of the fundamental building blocks of algebra. The reason for examining this at the beginning of our story is to point out the amazing progress mankind has made along the path of abstraction. As Derbyshire states [3]:

The very first act of mathematical abstraction occurred several millennia ago when human beings discovered numbers, taking the imaginative leap from observed instances of (for example) “three-ness”—three fingers, three cows, three siblings, three stars—to *three*, a mental object that could be contemplated by itself, without reference to any particular instance of three-ness.

The second such act, the rise to a second level of abstraction, was the adoption, in the decades around 1600 CE, of literal symbolism—that is, the use of letter symbols to represent arbitrary or unknown numbers: *data* (things given) or *quaesita* (things sought). “Universal arithmetic,” Sir Isaac Newton called it. The long stumbling journey to this point had been motivated mainly by the desire to solve equations, to determine the *unknown quantity* in some mathematical situation.

Adopting literal symbolism was a major advance in the development of algebra. It fundamentally changed how math was done. Before the adoption of literal symbolism math problems were often posed geometrically and solved via geometric reasoning. With the advent of symbolism, we eventually learned a strict set of rules for how to manipulate these symbols, i.e.

how to solve for “ x ”. And geometric reasoning was mostly replaced by this succinct set of abstract rules. However, it should be noted that each abstract rule of symbol manipulation has a geometric analog.

This new abstract way of solving mathematically posable questions was a double edged sword. On one hand it opened mathematicians’ minds to the possibilities of further abstraction which led to entirely new techniques and mathematics. On the other hand, the economy of teaching only the abstract symbol manipulations to students mostly devoid of their geometric roots was just too tempting to math educators. While a small number of students take to the abstract “game” of algebra like a fish to water, most do not. In the opinion of the author it is pedagogical suicide to attempt to teach the abstract rules of algebra without their geometric context. However this seems to have become the norm in most math curricula in the United States.

Most people, including many math majors, are often unfamiliar with the next act or step in this progression of abstraction wherein the object of interest became the relationships between the symbols rather than some unknown quantity for which the symbol was merely a placeholder. As Derbyshire puts it:

During the 19th century though, these letter symbols began to detach themselves from the realm of numbers. Strange new mathematical objects were discovered: groups, matrices, manifolds, and many others. Mathematics began to soar up to new levels of abstraction. That process was a natural development of the use of literal symbolism, once that symbolism had been thoroughly internalized by everyone. It is therefore not unreasonable to regard it as a continuation of the history of algebra.

1.1. Symmetries of the Equilateral Triangle

If literal symbolism is the mind of algebra, then symmetry is its heart. Everyone is familiar with symmetry. The human body exhibits a bi-lateral symmetry. It is common in art, architecture and even nature, but it is not often associated with math by most people.

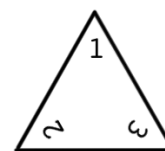
Here we wish to show that regular polygons, that is polygons with equal length sides and equal angles, give rise naturally to a certain algebraic structure called a group. First, let’s be more specific about what we mean by symmetry.

Definition 1.1. A *symmetry* is any transformation (function) that can be applied to a mathematical object which leaves the object in an equivalent state.

In this definition, we are using a relaxed meaning for the word “equivalent”. Here, equivalent does not mean “exactly the same”, but rather just *indistinguishable*.

For example we can rotate an equilateral triangle on an axis through its center by 120° or $2\pi/3$ radians and the triangle will look as though it was unchanged, or equivalent. In fact we can rotate an equilateral triangle by any integer multiple of $2\pi/3$ radians and it will look unchanged. Now obviously, a physical triangle made of paper or wood say, will have small distinguishing marks or defects that would allow us to determine whether or not it has been rotated, but we are only concerned with mathematical, equilateral triangles. That is, perfect or ideal triangles. These can only be imagined, but that is fine because we can still “manipulate” them in our minds.

Our first goal is to figure out how many symmetries the triangle possesses. To aid us, it is wise to temporarily add distinguishing marks to the triangle. Imagine an equilateral triangle with each vertex numbered from 1 to 3, both front and back so that the numbering on the front side corresponds to the numbering on the back side. Also imagine that this triangle is free to be lifted, flipped and manipulated in any way you wish. You might want to actually create a simple cardboard or paper triangle to help you visualize the various symmetry transformations. A symmetry transformation will be any series of rotations and flips that leave the triangle as it was before the transformation with the exception that the vertices may be in different positions, and thus the numbering may be different. So for example, any rotation that results in the triangle pointing downwards or in any direction other than up is disallowed.



Counting all of the various ways of rotating and flipping the triangle seems difficult. Perhaps the simplest way to determine the number of symmetries is to fix a starting state and then count the possible number of distinct end states. We can tell the various end states apart because the vertex labelling will be different. Let’s fix the start state of the triangle to be where the top vertex is labelled “1”, with “2” and “3” in the bottom left and bottom right vertices respectively. One way to count all of the symmetries is to count the number of positions to which a single vertex, say the “1” vertex, can be sent. Once you choose where the “1” vertex goes you have two choices for where the “2” and “3” vertices go. Since there are three places the “1” vertex can go, including its starting position, and since each other vertex has two possibilities, there are $3 \cdot 2 = 6$ possible permutations. These are pictured in figure 1.

If you carefully examine figure 1, you will notice that the first column corresponds to the three possible rotations about an axis that pierces the triangle in its center and comes out of the page. The first symmetry is

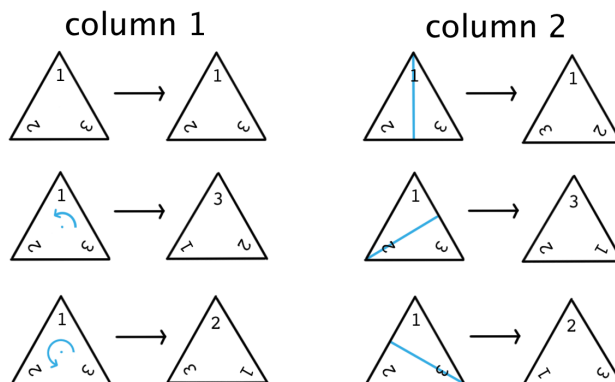


Figure 1. The six symmetries of an equilateral triangle.

a rotation by 0° about that axis or any integer multiple of 360° for that matter. The second symmetry in the first column is a counter-clockwise rotation by 120° or $2\pi/3$ radians, and finally the third one is a rotation by 240° or $2(2\pi/3)$ radians. Notice that $360^\circ = 3(2\pi/3)$ radians.

Each symmetry in the second column of figure 1 corresponds to flipping the triangle through an axis that starts at a vertex and intersects the opposite side midway. For example the first symmetry of column two corresponds to rotating the triangle 180° about a vertical axis which starts at the top vertex and cuts the bottom edge midway. Since we rotate by 180° this results in flipping the triangle over, but keeping one vertex in its original position. The first symmetry in the second column corresponds to preserving the original position of the “1” vertex, the second with preserving the “2” vertex and the third preserves the original position of the “3” vertex. Therefore there are three rotational symmetries and three flip symmetries for a total of six symmetries, which agrees with our previous count.

There are a few things to notice here. First, the six symmetries are functions which *act on* a set of labelled triangles. The phrase “act on” here simply means that each symmetry sends each triangle in the set of six possible triangles to some triangle in that same set. It is important to see that there is a distinction between the six possible end states of the triangle and the motions which achieve those end states. For example the second symmetry in the first column of figure 1 is rotation by 120° counter-clockwise around an axis through the center, but this symmetry can be applied to any triangle not just one in the designated start state. If we apply it to the triangle with “2” at the top and “1” and “3” in the bottom left and right vertices, then we end up with a triangle that has “3” at the top and “2” and “1” in the respective bottom vertices.

Second, we can combine or *compose* two symmetries to get a new symmetry. Since each symmetry is just a function which maps each triangle to some triangle in the set of possible triangles, these functions share the same domain and codomain, i.e. the set of six possible end states, and hence they can be composed. This is analagous to composing functions of a real variable. The difference is that now instead of having our domain and codomain equal to \mathbb{R} , the set of real numbers, it is a finite set consisting of six labelled triangles.

Third, each symmetry can be undone by some symmetry. For example, rotating by 120° is undone by rotating by 240° and vice versa. Interestingly, each one of the “flip” symmetries in the right hand colmn of figure 1 undoes itself. Also, the “do nothing” symmetry or rotation by 0° is undone by itself as well. Thus each symmetry has an *inverse symmetry* .

So far we have been discussing the symmetries of the equilateral triangle by referring to their physical movements, but this is awkward and tedious. We should name them. Naming something should never be taken lightly. Math and science are littered with several poor name choices which cause difficulty. One reasonable choice might be to use the letters R and F with subscripts. Subscripts on the letter R could denote the angle of rotation and subscripts on the letter F could denote the three different types of flips. Our set of symmetries would then be: $\{R_0, R_{120}, R_{240}, F_1, F_2, F_3\}$. These names are fairly easy to remember because the physical transformation which corresponds with each rotation is obvious. The subscripts on the flips also indicate which vertex is “preserved”, when applied to our “start state” triangle. Thus the motion corresponding with each symmetry can be easily remembered. This is good.

Using these new names we can write statements such as (check this):

$$(1.1) \quad (R_{120} \circ F_1)(\Delta) = R_{120}(F_1(\Delta)) = F_3(\Delta).$$

Here the Greek letter Δ , pronounced “delta”, is taking the place of our usual placeholder of x for unknowns. This is decent notation, but we can do better. Notice that all of the rotations of column one can be *generated* by R_{120} , specifically (check these):

$$\begin{aligned} R_0 &= R_{120} \circ R_{120} \circ R_{120}, \\ R_{120} &= R_{120} \\ R_{240} &= R_{120} \circ R_{120}. \end{aligned}$$

Also, the flips of column two can be *generated* by composing R_{120} and F_1 , (check these too):

$$\begin{aligned} F_1 &= F_1 \\ F_2 &= R_{120} \circ R_{120} \circ F_1, \\ F_3 &= R_{120} \circ F_1. \end{aligned}$$

Furthermore, since writing the “ \circ ” symbol for function composition is tedious let’s just drop it and instead use multiplicative style notation where juxtaposing two symbols implies function composition. For example, $R_{120} \circ R_{240} = R_{120}R_{240}$. This multiplicative style notation has a nice feature, it allows us to use exponents to represent functions or symmetries composed with themselves repeatedly. For example,

$$R_{120} \circ R_{120} \circ R_{120} = R_{120}R_{120}R_{120} = R_{120}^3.$$

Thus we see that it suffices to use these two symbols, R_{120} and F_1 to describe all of the symmetries. Or simpler yet, we could just use the lower case letters r and f where $r = R_{120}$ and $f = F_1$.

$$R_0 = R_{120} \circ R_{120} \circ R_{120} = rrr = r^3$$

$$R_{120} = r$$

$$R_{240} = R_{120} \circ R_{120} = rr = r^2$$

$$F_1 = f$$

$$F_2 = R_{120} \circ R_{120} \circ F_1 = rrf = r^2f$$

$$F_3 = R_{120} \circ F_1 = rf.$$

Above we are just being flexible with notation, but in order to make the analogy with multiplication complete it probably makes more sense to let $R_0 = 1$ instead of writing $R_0 = R_{120} \circ R_{120} \circ R_{120} = r \cdot r \cdot r = r^3$. This is because the number 1 is the *multiplicative identity*, which if you recall means that multiplying any number by 1 does not change that number. Similarly, composing R_0 with any other symmetry does not change that symmetry.

Figure 2 shows the six symmetries labelled with these new names.

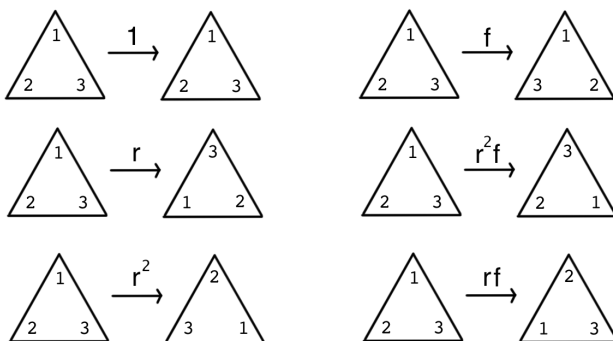


Figure 2. The six symmetries of an equilateral triangle.

Let’s recap what we have learned up to this point. We have found that that the set of symmetries of an equilateral triangle contains six elements,

$\{1, r, r^2, f, r^2f, rf\}$. Further we have seen that we are able to generate all six symmetries by composing together just a single counter-clockwise rotation of 120° with a single flip, in this case we used the flip that preserves the “1” vertex, F_1 . Could we have used F_2 or F_3 instead? For that matter could we generate all the symmetries with say two different flip symmetries, or perhaps two rotational symmetries?

Exercise 1.1. Determine which pairs of symmetries generate the entire set of symmetries, i.e. $\{R_0, R_{120}, R_{240}, F_1, F_2, F_3\}$. Show how to generate each element of the set using just elements in the pair. Can this set be generated by a single symmetry?

1.1.1. Properties of Symmetries. To summarize our findings so far, we have found the following:

- (1) **Action:** Symmetries are functions which “act” upon certain sets, by sending each element of the set to some other element in that set or possibly the same element.
- (2) **Closure:** When we compose two symmetries, we always end up with another symmetry from the set of all symmetries of an object.
- (3) **Identity Symmetry:** The identity symmetry, or “do nothing” symmetry is a member of the set of all possible symmetries of an object.
- (4) **Inverses:** Every symmetry has an opposite or inverse symmetry which undoes its action.
- (5) **Generators:** It might be possible to generate (via composition) the complete set of symmetries with only a proper subset of the whole set.

Not every set of symmetries acts on every set. For example the symmetries of the equilateral triangle do not act on squares and vice versa. That is to say that rotating a square by 120° will not leave it in an equivalent state, and neither will rotating an equilateral triangle by 90° leave it in an equivalent state.

We won’t discuss generators much further here, except to say that if you have studied Linear Algebra, then the notion of a *basis vector* is analogous to that of a *generator*.

1.1.2. Associativity. There is actually another subtle property that sets of symmetries possess which stems from the fact that they are functions. Function composition is *associative*. To illustrate, let f, g and h be three functions and suppose:

$$\begin{aligned} f &: A \rightarrow B, \\ g &: B \rightarrow C, \\ h &: C \rightarrow D. \end{aligned}$$

Since the domain of g matches the codomain of f we can compose them to get:

$$(g \circ f) : A \rightarrow C.$$

Similarly for g and h . But now we can create two new functions which both have domain A and codomain D , but are they equal? That is,

$$h \circ (g \circ f) \stackrel{?}{=} (h \circ g) \circ f.$$

We can look at the simple binary operation of subtraction to see that where you put parentheses can matter, to wit:

$$3 - (4 - 5) \neq (3 - 4) - 5.$$

Back to our specific example, it is actually very easy to prove that function composition is associative. We want to show that $h \circ (g \circ f) = (h \circ g) \circ f$. Recall that $(g \circ f)[x] = g[f[x]]$, where we use parentheses around functions and brackets around values to make this rewrite rule more clear. By applying this rewrite rule repeatedly, we see that,

$$\begin{aligned} (h \circ (g \circ f))[x] &= h[(g \circ f)[x]] \\ &= h[g[f[x]]] \\ &= (h \circ g)[f[x]] \\ &= ((h \circ g) \circ f)[x], \end{aligned}$$

which proves the claim.

Groups

We have discovered that the set of symmetries of an object must have certain features, namely:

- closure,
- associativity,
- an identity element,
- inverses.

The four properties above capture the essence of symmetry. Naturally anything that important deserves a name. Any set with a law of composition that satisfies these four properties is called a *group*. In short, you should think of a group as a “set with structure”. The structure derives from the law of composition, which in our only example, so far, is “o”, function composition.

Definition 2.1. A *group* is a set, G , with a binary law of composition, \cdot , often written (G, \cdot) , satisfying the following for all $a, b, c \in G$:

- (1) **closure:** $a \cdot b \in G$.
- (2) **associativity:** $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.
- (3) **identity element:** There exists an $e \in G$ such that $a \cdot e = a = e \cdot a$.
- (4) **inverses:** For every $a \in G$ there exists some $b \in G$ such that $a \cdot b = e$ and $b \cdot a = e$, we often denote b as $b = a^{-1}$.

Just from the definition of a group we can deduce that the identity element must be unique. To see this, suppose that G is a group and suppose that e and e' are two identity elements in G , then

$$e = e \cdot e' = e'.$$

The group of symmetries of the equilateral triangle has a special name, it is called the Dihedral group for the regular triangle, or (D_6, \circ) , or usually just D_6 , because the law of composition is well known for this group. In fact there are an infinite number of Dihedral groups, one for each regular polygon, and they are denoted D_{2n} where n is the number of sides of the regular polygon, thus the group of symmetries of the square is denoted D_8 , and the group of symmetries of the regular pentagon is D_{10} .

Notice that the group definition does not require the set to be finite. Our first group example, D_6 is a finite group, but there are many infinite groups. One important way of classifying a group is based upon the number of elements it contains.

Definition 2.2. The *order* of a group, G , is the number of elements it contains. The order of G is denoted by $|G|$.

The order of D_6 is 6, or $|D_6| = 6$. You are already familiar with some infinite groups, for example the real numbers, \mathbb{R} , form a group under multiplication. Well, this set is almost a group. The problem is that zero doesn't have a multiplicative inverse, but if we exclude zero: $\mathbb{R} - \{0\}$, then we have a true group. We have closure because whenever you multiply two real numbers you get a real number. Multiplication is associative. The multiplicative identity is just the number 1. Finally, every real number $x \in \mathbb{R} - \{0\}$ has inverse $1/x$. Thus $(\mathbb{R} - \{0\}, \times)$ is a group. Since this group comes up often, the shorthand name, \mathbb{R}^\times is usually preferred. The order of \mathbb{R}^\times is infinite and we write $|\mathbb{R}^\times| = \infty$.

If we change the operation from multiplication to addition, then we have a different group, $(\mathbb{R}, +)$, which again is usually shortened to just \mathbb{R}^+ . In this group, the identity is the number 0. Every number has an inverse, but in an additive group, i.e. a group where the operation is designated by "+", we often call inverses opposites. For example, we usually say -3 is the opposite of 3, rather than the inverse of 3.

2.0.3. Cayley Tables. It can be instructive to create a table which shows all possible ways of composing two symmetries. The first person to do so was the English mathematician Arthur Cayley, hence such tables are called Cayley tables. These are exactly analogous to multiplication tables.

To use a multiplication table to find the product, 9×12 , you find 9 in the leftmost vertical column and 12 in the top row. The answer of course lies at the intersection of the 9th row and 12th column (not counting the leftmost column and topmost row).

Similarly, we can create a table for composing symmetries. Table 1 is a Cayley table for the symmetries of the equilateral triangle.

Table 1. Cayley table for the symmetries of an equilateral triangle

\circ	1	r	r^2	f	r^2f	rf
1	1	r	r^2	f	r^2f	rf
r	r	r^2	1	rf	f	r^2f
r^2	r^2	1	r	r^2f	rf	f
f	f	r^2f	rf	1	r	r^2
r^2f	r^2f	rf	f	r^2	1	r
rf	rf	f	r^2f	r	r^2	1

There are some very interesting things to observe about this table. First notice that:

$$r \circ f = rf \neq r^2f = f \circ r.$$

In other words composition of symmetries in D_6 is non-commutative. Commutativity refers to order in space, i.e. position. Associativity refers to order in time, i.e. sequence. The law of composition for a group needs to be associative, but not necessarily commutative. If it is also commutative, then we have a special word to describe those groups.

Definition 2.3. An *abelian* group is a group where the law of composition is commutative. That is, if (G, \cdot) is abelian then for all $a, b \in G$, $a \cdot b = b \cdot a$. If the law of composition is non-commutative in a group, then we say the group is *non-abelian*.

It may seem like the mathematicians who came up with this term were just trying to be fancy, but that is not the case, we will soon learn that the term *commutative* is reserved for another group like object.

Another interesting thing to note about this table is that if you examine the upper left, 3×3 block of the table, the part with 1, r and r^2 , you can see that it forms a group all on its own. It is the group of rotations of the triangle. Naturally this subset is called a *subgroup*.

Definition 2.4. A *subgroup* is a subset of a group that is itself a group.

Notice that for a subset to be a subgroup it must contain the identity element. Subgroups play an important role in the study of polynomial equations. We will have much more to say about them later.

Exercise 1.2. Create a Cayley table for the symmetries of a square.

Hints: This group has eight symmetries. There are four rotational symmetries and four flip symmetries. Two of the flips are with respect to axes through diagonal vertices and two flips are with respect to axes through opposite midpoints i.e. one horizontal axis and one vertical axis.

Use similar notation as in the triangle case, let $r = R_{90}$ and let f be a flip through the vertical axis. In other words, your symmetries should be named only using the characters r and f with exponents only on r . Furthermore, any symmetry names with an f in them should have it at the end.

Cut out and label a paper square to help you.

2.1. Groups and Algebra

Groups are the fundamental building blocks of algebraic systems. You can think of a group as the smallest unit within which it is possible to solve symbolic equations. To illustrate, we will solve for x in the following equation:

$$rf \circ x = r^2f.$$

The rules of algebra and the Cayley table provide a systematic way of solving for the unknown, or isolating x . Each step will use one of the four group properties or a table lookup:

$$\begin{aligned} rf \circ x &= r^2f \\ (rf)^{-1} \circ (rf \circ x) &= (rf)^{-1} \circ r^2f && \text{inverses} \\ ((rf)^{-1} \circ rf) \circ x &= (rf)^{-1} \circ r^2f && \text{associativity} \\ 1 \circ x &= (rf)^{-1} \circ r^2f && \text{inverses} \\ x &= (rf)^{-1} \circ r^2f && \text{identity} \\ x &= rf \circ r^2f && \text{table lookup} \\ x &= r^2 && \text{table lookup} \end{aligned}$$

This is exactly analogous to how you might systematically solve a simple equation such as $3x = 6$.

$$\begin{aligned} 3x &= 6 \\ 3^{-1}(3x) &= 3^{-1}6 && \text{inverses} \\ (3^{-1}3)x &= 3^{-1}6 && \text{associativity} \\ 1x &= 3^{-1}6 && \text{inverses} \\ x &= 3^{-1}6 && \text{identity} \\ x &= 2 && \text{table lookup} \end{aligned}$$

If x were on the left instead of right such as: $x \circ rf = r^2f$, then we could begin to isolate x by multiplying both sides of the equation by $(rf)^{-1}$ but from the right hand side, i.e.

$$(x \circ rf) \circ (rf)^{-1} = r^2f \circ (rf)^{-1}.$$

Here our handy notation tells us that the answer is r without going through the algebraic steps.

2.2. The Symmetric Group: S_n

The next group we will examine is called the symmetric group on n objects or S_n for short. Actually this is an infinite family of groups, one group for each natural number $n \in \mathbb{N}$. If $n = 3$, then S_3 is the group of permutations of three objects.

Often we think of these groups as acting on the numbers $1 \dots n$, or even lists of numbers but there are many different objects this group can act on. One important class of objects are polynomials in more than one variable. These groups, especially S_5 are very important for understanding Galois' ideas.

Definition 2.5. A *permutation* in S_n is an invertible function where the domain and codomain are the same set, namely: $\{1, 2, 3, \dots, n\}$.

Recall that a function is invertible when it is *one-to-one* and *onto*. A good way to denote permutations is by a simple two row table. The following permutation is an element of the group S_3 :

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}.$$

The first row lists all domain values and the second row lists the image of each domain element directly below. Thus $1 \mapsto 3, 2 \mapsto 2$ and $3 \mapsto 1$. Each element in the codomain is the image of some element in the domain, thus this mapping is *onto*. Also, each element in the codomain is the image of a single element in the domain, thus this mapping is *one-to-one*. An example of a function that is not one-to-one, nor onto is the following:

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 3 & 2 \end{pmatrix}.$$

The above table *is* a function, because each element in the domain gets mapped to exactly one element in the codomain. However, it is neither one-to-one, nor onto. It is not one-to-one because two elements, 1 and 2 in the domain get mapped to the same element, namely 3. If you were to try to undo this mapping, you would face a dilemma on what to do with 3. Should the inverse function send 3 to 1 or 2? Furthermore, it is not onto, so we have no idea as to which element the inverse function should send the number 1. These ambiguities are what makes this function non-invertible, and thus it is not a permutation.

A simple criterion for deciding whether a given table is a permutation is to check whether each number that occurs in the top row occurs just once in the bottom row. This will ensure that the function is both one-to-one and onto.

In this new “table notation” for functions, the identity permutation of the group S_3 is denoted by:

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}.$$

It is the mapping which sends each element to itself.

2.2.1. The Order of S_n . Now that we have established workable notation for permutations, we should investigate exactly how many elements are in S_n . A single permutation makes a choice of where to send each and every element in the domain. We need to count how many choices are possible. Once you choose where the element “1” is sent to, you have $n - 1$ remaining choices for where to send $2, 3, \dots, n$. If we make a choice as to where “2” will be sent, then there will be $n - 2$ possible image values for $3, 4, \dots, n$. If we continue in this fashion we will eventually end up with

$$n(n - 1)(n - 2) \cdots 2 \cdot 1 = n!$$

possible choices, each of which corresponds to a unique permutation. Thus $|S_n| = n!$.

2.2.2. Composition of Permutations. It is fairly simple to compose permutations that are written in table notation. The key is to remember that the rightmost function is the first mapping to be applied, followed by the function directly left of it, and so on. Thus $f \circ g$, should be read: g followed by f . Get in the habit of reading it that way. For example,

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}.$$

Here is another example of the composition of two permutations from S_5 :

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 1 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 2 & 5 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 3 & 1 & 5 \end{pmatrix}.$$

Even with these small examples, we see that this notation can get unwieldy. If we were to compose permutations from say S_{11} the tables would be rather large and tedious to write. We need a better notation for permutations.

2.2.3. Cycle Notation. Table notation for permutations is good, but it requires a fair bit of writing. A more efficient notation is called cycle notation. The idea behind cycle notation is to write a permutation as an ordered list. Thus the table becomes a one row list:

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \mapsto (1 \mapsto 3 \quad 3 \mapsto 2 \quad 2 \mapsto 1) \mapsto (1 \ 3 \ 2).$$

We drop the arrows entirely in the last list, and remember that the last number in the list always cycles back to the first number in the list, hence the term cycle notation.

The first thing to notice about this notation is that it is *not* unique. There may be more than one way to write a cycle, based upon which number you choose to put first in the list. The following cycles all start with a different number but are equivalent.

$$(1\ 2\ 3) = (2\ 3\ 1) = (3\ 1\ 2)$$

They are equivalent because they all represent the mapping:

$$\begin{aligned} 1 &\mapsto 2 \\ 2 &\mapsto 3 \\ 3 &\mapsto 1. \end{aligned}$$

Another important observation to make is that disjoint cycles commute. Two cycles are *disjoint* if they share no numbers in common. For example,

$$(1\ 2)(3\ 4\ 5) = (3\ 4\ 5)(1\ 2) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 4 & 5 & 3 \end{pmatrix}.$$

Thus a permutation written in cycle notation is a product of disjoint cycles. If two or more cycles in a product of cycles are not disjoint, then the whole expression can be simplified via the cycle composition algorithm which is explained in the next section.

The fact that disjoint cycles commute further increases the number of equivalent ways in which a permutation can be written in cycle notation. For example, a permutation with three disjoint cycles such as:

$$(1\ 2)(3\ 4\ 5)(6\ 7),$$

has $3! = 6$ equivalent ways of ordering the three cycles, and there are two distinct ways of writing each 2-cycle and three distinct ways of writing the 3-cycle for a grand total of $3! \cdot 2 \cdot 2 \cdot 3 = 72$ distinct ways of writing this permutation in cycle notation! In practice this doesn't matter too much, but to make it easier for us to recognize equivalent permutations, we will declare that the standard way to write a cycle is to put the smallest number in that cycle first. Thus we will prefer $(1\ 2\ 3)$ to $(2\ 3\ 1)$ and $(3\ 1\ 2)$.

The above examples are fairly simple. They do not illustrate all the subtleties of converting a permutation into cycle notation. For example, often a permutation will need to be decomposed into more than one cycle like the following example shows.

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 2 & 1 & 5 & 4 \end{pmatrix} \longmapsto (1\ 3\ 2)(4\ 5)$$

We need to make the algorithm for writing a permutation in cycle notation explicit. In the following algorithm, when we refer to the domain list,

we simply mean the top row of the permutation when it is written in table notation.

Input: a permutation σ in table notation
Output: permutation σ in cycle notation

$D :=$ the domain list;
while D is not empty **do**
 $a :=$ the smallest number in the domain list;
 remove a from D ;
 $start := a$;
 print “ a ”;
 while $\sigma(a) \neq start$ **do**
 $a := \sigma(a)$;
 remove a from D ;
 print “ a ”;
 end
 print “)”;
end
remove any 1–cycles;

Algorithm 1: Cycle Generation Algorithm in Pseudocode

Let’s do a more complicated example. Convert the following permutation from table notation to cycle notation.

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ 3 & 4 & 5 & 7 & 9 & 6 & 8 & 2 & 1 & 12 & 11 & 10 \end{pmatrix}$$

This permutation becomes $(1\ 3\ 5\ 9)(2\ 4\ 7\ 8)(6)(10\ 12)(11)$, which after removing the unnecessary 1–cycles becomes:

$$(1\ 3\ 5\ 9)(2\ 4\ 7\ 8)(10\ 12).$$

Exercise 1.3. Convert the following permutation to cycle notation:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 4 & 7 & 6 & 3 & 8 & 1 & 2 & 5 \end{pmatrix}$$

Exercise 1.4. Convert the following permutation to cycle notation:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\ 9 & 10 & 11 & 7 & 1 & 8 & 2 & 4 & 6 & 5 & 3 \end{pmatrix}$$

Exercise 1.5. Convert the following permutation to cycle notation:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\ 11 & 9 & 5 & 3 & 8 & 2 & 1 & 4 & 6 & 7 & 10 \end{pmatrix}$$

Exercise 1.6. Convert the following permutation to cycle notation:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ 5 & 8 & 6 & 3 & 7 & 2 & 1 & 4 & 9 & 11 & 12 & 10 \end{pmatrix}$$

2.2.4. Composing/Multiplying Cycles. Composition of permutations written in cycle notation is straightforward. First, the composition symbol, \circ , is dropped in favor of multiplicative notation, thus $(1\ 2\ 3) \circ (1\ 2) = (1\ 2\ 3)(1\ 2)$, and we will often just refer to composition as multiplication.

The most important thing to remember when multiplying (composing) cycles is that they are functions and function application begins with the rightmost function and then proceeds leftward. In other words, to determine the composition of two permutations written in cycle notation, you start by writing “1”, and then determine the image of that input under the mapping of the first (rightmost) cycle. The output or image is then fed into the next cycle to the left and so on and so forth until there are no more cycles left. For example to see where “1” gets mapped under the composition $(1\ 2\ 3)(1\ 2)$, you start with the rightmost cycle which maps $1 \mapsto 2$, the next cycle then maps $2 \mapsto 3$, thus the composition in total sends $1 \mapsto 3$. Thus we have “(1 3)”. We repeat this procedure but this time we start with “3” and find that since “3” does not occur in the rightmost cycle that means it sends $3 \mapsto 3$. The next cycle to the left sends $3 \mapsto 1$ and that is the last cycle, so we write: “(1 3)”. We close the cycle to indicate that “3” cycles back to “1”. Next, since the next lowest number in the domain is “2”, we start a new cycle “(2)”, and repeat the whole process above. The rightmost cycle sends $2 \mapsto 1$, and the next cycle maps $1 \mapsto 2$, thus 2 gets mapped to itself and we write “(1 3)(2)”. Finally, we remove all 1-cycles (cycles of length one) from the product (composition). Thus,

$$(1\ 2\ 3)(1\ 2) = (1\ 3).$$


```

Input: a non-empty, finite sequence of cycles
Output: a finite sequence of disjoint cycles

// create the domain list
D := concatenate all cycles;
remove parentheses from D;
remove duplicates from D;

// prime the pump
a := the smallest number in D;
remove a from D;
start := a;
print "(a";
while D is not empty do
  while cycles remain do
    // cycles are chosen from right to left
     $\sigma$  := next cycle;
    a :=  $\sigma(a)$ ;
  end
  if a  $\neq$  start then
    remove a from D;
    print " a";
  else
    if D is not empty then
      a := the smallest number in D;
      remove a from D;
      start := a;
      print ") (a";
    end
  end
end
print " ";
remove any 1-cycles;

```

Algorithm 2: Cycle Multiplication Algorithm in Pseudocode

The cycle composition algorithm has similar aspects to the algorithm for writing a permutation in cycle notation. Notably, you start with the lowest number that occurs in all cycles, say a and begin by writing “(a”. Next, while there are cycles remaining you keep updating a according to the mapping defined in each cycle moving from right to left, until no cycles remaining. You then check to see if the resulting value matches the first number in the cycle. If it matches, you close the cycle by writing “)”. If it does not match you append a space and the current mapped value “ a”.

This loop continues until all values in the domain have been written exactly once. Finally, you remove one cycles.

Exercise 1.7. Compute $(1\ 2\ 3)(1\ 2)$.

Exercise 1.8. Compute $(1\ 2)(1\ 2\ 3)$.

Exercise 1.9. Compute $(1\ 2\ 3\ 5)(4\ 3\ 5)(3\ 4\ 5)$.

Exercise 1.10. Compute $(1\ 2\ 3)(2\ 3)(1\ 3\ 2)$.

Exercise 1.11. Compute $(2\ 5\ 3)(2\ 3)(2\ 3\ 5)$.

Exercise 1.12. Compute $(1\ 2)(1\ 2\ 5\ 3\ 4)(1\ 2)$.

Exercise 1.13. Compute $(2\ 3)(2\ 5\ 3\ 4)(2\ 3)$.

Exercise 1.14. Compute $(4\ 5)(3\ 6\ 5\ 4)(5\ 3\ 4\ 7)(3\ 6\ 7)$.

2.2.5. Cycle Inverses. Another useful feature of cycle notation is that it is very easy to find the inverse of any permutation written as a disjoint product of cycles. The inverse of a cycle is simply the cycle written in reverse order. For example, the inverse of $(1\ 3\ 2\ 4)$ is simply $(4\ 2\ 3\ 1)$ which if we follow our convention of writing cycles with the smallest number first would become $(1\ 4\ 2\ 3)$. You can check that indeed,

$$(1\ 3\ 2\ 4)(1\ 4\ 2\ 3) = (1)(2)(3)(4) = 1.$$

If a permutation consists of several disjoint cycles multiplied together, then the inverse will be the product of the inverses of each individual cycle. Thus

$$\begin{aligned} [(1\ 3\ 5)(2\ 4\ 6)(7\ 8)]^{-1} &= (1\ 3\ 5)^{-1}(2\ 4\ 6)^{-1}(7\ 8)^{-1} \\ &= (1\ 5\ 3)(2\ 6\ 4)(7\ 8), \end{aligned}$$

which you should verify for yourself.

2.2.6. Equivalence of S_3 and D_6 . Below is the Cayley table for S_3 with the permutations expressed in cycle notation, and the Cayley table for the symmetries of the equilateral triangle or D_6 . Compare them carefully.

If you look carefully at both tables, you will see that they have the exact same structure, just different names for each element or symmetry in the group. In other words we can create an invertible mapping, call it

Table 2. Cayley table for the symmetric group on three objects: S_3

\circ	1	(1 2 3)	(1 3 2)	(2 3)	(1 3)	(1 2)
1	1	(1 2 3)	(1 3 2)	(2 3)	(1 3)	(1 2)
(1 2 3)	(1 2 3)	(1 3 2)	1	(1 2)	(2 3)	(1 3)
(1 3 2)	(1 3 2)	1	(1 2 3)	(1 3)	(1 2)	(2 3)
(2 3)	(2 3)	(1 3)	(1 2)	1	(1 2 3)	(1 3 2)
(1 3)	(1 3)	(1 2)	(2 3)	(1 3 2)	1	(1 2 3)
(1 2)	(1 2)	(2 3)	(1 3)	(1 2 3)	(1 3 2)	1

\circ	1	r	r^2	f	r^2f	rf
1	1	r	r^2	f	r^2f	rf
r	r	r^2	1	rf	f	r^2f
r^2	r^2	1	r	r^2f	rf	f
f	f	r^2f	rf	1	r	r^2
r^2f	r^2f	rf	f	r^2	1	r
rf	rf	f	r^2f	r	r^2	1

Table 3. Cayley table for the symmetries of an equilateral triangle: D_6

$\varphi : S_3 \rightarrow D_6$ where,

$$\begin{aligned}
 1 &\mapsto 1, \\
 (1\ 2\ 3) &\mapsto r, \\
 (1\ 3\ 2) &\mapsto r^2, \\
 (2\ 3) &\mapsto f, \\
 (1\ 3) &\mapsto r^2f, \\
 (1\ 2) &\mapsto rf,
 \end{aligned}$$

and we see that these two groups are essentially the same group in different guises.

Another way to think of this is to suppose that I want to determine $(1\ 2\ 3) \circ (2\ 3)$. One way to do this is to rewrite it as $r \circ f$ which is of course just rf , then map backwards via the mapping above to get $(1\ 2)$. Thus

$$(1\ 2\ 3) \circ (2\ 3) = (1\ 2).$$

To better understand what is happening here, a diagram helps.

$$\begin{array}{ccc} (S_3, S_3) & \xrightarrow{\circ S_3} & S_3 \\ \downarrow \varphi \times \varphi & & \uparrow \varphi^{-1} \\ (D_6, D_6) & \xrightarrow{\circ D_6} & D_6 \end{array}$$

The symbol $\varphi \times \varphi$ is just a fancy way of saying that you have to apply φ to both of the elements from the pair (S_3, S_3) .

Equipped with this mapping we can answer any question regarding S_3 by considering the equivalent question in terms of elements of D_6 or vice versa. This idea of transforming a problem or question into an equivalent problem in a different realm, solving the problem there and then mapping the answer back to the original domain via an inverse mapping is the basis of much of mathematics. For example you may be familiar with the Laplace transform method of solving differential equations which uses this idea. Near the end of our investigations, we will see that this is exactly what Galois did. He found a way to associate a group with a polynomial equation such that the solvability of the equation was reflected in the structure of the group.

In order to study the structure of groups we need a better tool than the Cayley table. The problem with Cayley tables is twofold. First, they become unwieldy for large groups. For example, consider S_5 , $|S_5| = 5! = 120$ which would require a huge table too big to write on a single page! Second, groups often don't have a natural ordering, and thus the order in which you place the elements along the top row and left column of the Cayley table is rather arbitrary, but it drastically changes the appearance of the table. Thus using a Cayley table to determine when two groups are essentially equivalent is only feasible for groups of very small order. We need a better tool.

2.3. Homomorphisms

The primary tool for examining group structure is a special function or map called a *homomorphism*.

Definition 2.6. A *homomorphism* is a map from one group to another which preserves the group structure. That is, the map $\varphi : G \rightarrow G'$ is a homomorphism if for all $a, b \in G$,

$$\varphi(ab) = \varphi(a)\varphi(b).$$

There are two specific and important examples of what we mean by, a homomorphism “preserves the group structure”. First, a homomorphism, $\varphi : G \rightarrow G'$ must map the identity of G to the identity of G' . Let e be the identity of G and e' the identity of G' then by the definition of identity for

all $a \in G$, $ae = a = ea$. Thus,

$$\begin{aligned}\varphi(ae) &= \varphi(ea) \\ \varphi(a)\varphi(e) &= \varphi(e)\varphi(a).\end{aligned}$$

Notice that $\varphi(e)$ above exactly satisfies the definition of an identity, therefore $\varphi(e) = e'$, that is, it is the identity element of G' . Since we assumed nothing about φ except that it is a homomorphism we see that *every* homomorphism must preserve the identity element.

Second, homomorphisms preserve inverses. Let $a \in G$, then since G is a group a^{-1} is also in G . Since $aa^{-1} = e$, we see that:

$$\begin{aligned}\varphi(e) &= \varphi(aa^{-1}) \\ e' &= \varphi(a)\varphi(a^{-1}).\end{aligned}$$

But also,

$$\begin{aligned}\varphi(e) &= \varphi(a^{-1}a) \\ e' &= \varphi(a^{-1})\varphi(a).\end{aligned}$$

Therefore we see that $\varphi(a)$ and $\varphi(a^{-1})$ exactly satisfy the requirements to be inverses. This implies $\varphi(a^{-1}) = [\varphi(a)]^{-1}$.

Definition 2.7. Given a homomorphism $\varphi : G \rightarrow G'$, then the *kernel* of φ , denoted $\ker(\varphi)$ is the subset of all elements in the domain which get mapped to the identity in the codomain. In set notation,

$$\ker(\varphi) = \{a \in G \mid \varphi(a) = 1\}.$$

Proposition 2.1. *The kernel of a homomorphism is a subgroup of the domain of the homomorphism.*

Exercise 1.15. Prove that the kernel of a homomorphism is a subgroup of the domain of the homomorphism.

Hints: Since the domain is a group, and since any subgroup shares the same operation as its parent there is no need to show that the operation is associative. Let $\varphi : G \rightarrow G'$ be a homomorphism, then you must show three things:

- (1) **closure:** Pick two arbitrary elements say $a, b \in \ker(\varphi)$ and show that their product must necessarily also be an element of $\ker(\varphi)$, i.e. show $\varphi(ab) = 1$.
- (2) **identity:** Show that $1 \in \ker(\varphi)$.
- (3) **inverses:** Show that if $a \in \ker(\varphi)$, then $a^{-1} \in \ker(\varphi)$.

Exercise 1.16. Let G be a group, the map $f : G \rightarrow G$ given by $f : g \mapsto g^{-1}$ is not always a homomorphism. Why not? What property must G have to make it a homomorphism?

Hint: Consider the product of two elements in G say ab , then $f(ab) = [ab]^{-1}$, but you can actually figure out how to write $[ab]^{-1}$ in terms of a^{-1} and b^{-1} if you use the definition of inverses (found in the group definition).

Exercise 1.17. Let G be a group, prove that “the conjugation by g ” map $\varphi_g : G \rightarrow G$ given by $\varphi_g : a \mapsto gag^{-1}$ is a homomorphism.

Hint: You want to show that given $a, b \in G$, $\varphi_g(ab) = \varphi_g(a)\varphi_g(b)$. This can be done by inserting a special form of the identity element, namely, $g^{-1}g$, into the image of ab .

2.3.1. Normal Subgroups.

Definition 2.8. A subgroup N of G is a *normal subgroup* if for every $a \in N$ and every $g \in G$, the conjugate of a , $gag^{-1} \in N$.

Proposition 2.2. *The kernel of a homomorphism is a normal subgroup.*

Proof. Suppose $\varphi : G \rightarrow G'$ is a homomorphism. Assume $a \in \ker(\varphi)$ and let g be any element in G , then,

$$\begin{aligned} \varphi(gag^{-1}) &= \varphi(g)\varphi(a)\varphi(g^{-1}) \\ &= \varphi(g)1\varphi(g^{-1}) \\ &= \varphi(g)\varphi(g^{-1}) \\ &= \varphi(g)[\varphi(g)]^{-1} \\ &= 1. \end{aligned}$$

□

2.3.2. Isomorphisms.

2.4. Equivalence Relations and Partitions

Definition 2.9. A *partition* of a set S , is a collection of disjoint, non-empty subsets, E_1, E_2, \dots, E_n such that their union is S , i.e. $S = E_1 \cup E_2 \cup \dots \cup E_n$.

Definition 2.10. A *relation* between two sets S and T is a collection of ordered pairs, where the first element of each ordered pair is from S and the second is from T .

Notice that relations are similar to functions, but more general. That is every function is a relation but not every relation is a function. In order for a relation to be a function every element in the domain must be related to only one element in the range.

You can also have a relation from a single set S to itself.

Suppose $R = \{\{1, 1\}, \{2, 0\}, \{3, 1\}, \{4, 0\} \dots\}$ then we say that 2 is related to 0 and 3 is related to 1 or symbolically, $2 \sim 0$, and $3 \sim 1$.

Definition 2.11. An *equivalence relation* on a set S is a relation from S to itself that satisfies the following for all $a, b, c \in S$:

- (1) **reflexive:** $a \sim a$.
- (2) **symmetric:** If $a \sim b$, then $b \sim a$.
- (3) **transitive:** If $a \sim b$ and $b \sim c$, then $a \sim c$.

Proposition 2.3. A partition of a set E determines an equivalence relation on E and conversely (vice versa).

Proof. Suppose we have a partition of E , thus $E = E_1 \cup E_2 \cup \dots \cup E_n$. Pick any $a \in E_1$, we can define an equivalence relation on E by saying $a \sim b$ if $b \in E_1$. In other words, elements in the same subset are equivalent. (You should verify for yourself that this satisfies all the requirements of being an equivalence relation.)

Conversely, suppose we have an equivalence relation defined on E , we can create a partition of E by picking any element of E , say a and defining E_1 to be the set of all elements that are equivalent to a , that is:

$$E_1 = \{b \in S \mid b \sim a\}.$$

Continue this partitioning by picking another element, say $c \in S$ such that $c \notin E_1$ to create E_2 and so on and so forth. Since E is finite, this algorithm must terminate, and by design each set E_i will be nonempty and the union $E_1 \cup E_2 \cup \dots \cup E_n = E$. \square

2.5. Cosets and Lagrange's Theorem

Definition 2.12. Suppose H is a subgroup of G , and $a \in G$, then the subset

$$aH = \{ah \mid h \in H\},$$

is called a *left coset*.

Lemma 2.13. The left cosets of a subgroup, H of G , partition G .

Proof. We can define an equivalence relation on G as follows:

$$a \sim b \quad \text{if} \quad b = ah \text{ for some } h \in H.$$

We must verify that this is indeed an equivalence relation, but once we do, then by the previous proposition, this equivalence relation also defines a partition.

Reflexive: We must show $a \sim a$. Every subgroup must contain the identity, so $1 \in H$, therefore, $a \sim a$ because $a = a1$.

Symmetric: We must show that if $a \sim b$, then $b \sim a$. Suppose $a \sim b$, then by definition of the equivalence relation there is some $h \in H$ such that $b = ah$, then $a = bh^{-1}$ and h^{-1} is in H , so $b \sim a$.

Transitive: We must show that if $a \sim b$ and $b \sim c$, then $a \sim c$. Suppose $a \sim b$ and $b \sim c$ then $a = bh_1$ and $b = ch_2$. This implies $a = (ch_2)h_1 = c(h_2h_1)$, but this implies $c \sim a$, and by symmetry $a \sim c$. \square

Lemma 2.14. *All left cosets aH of a subgroup H have the same order.*

Proof. Define a map of sets

$$\begin{aligned} f : H &\rightarrow aH \\ f(h) &= ah. \end{aligned}$$

Notice that this map has an inverse, namely

$$\begin{aligned} f^{-1} : aH &\rightarrow H \\ f^{-1}(ah) &= a^{-1}(ah), \end{aligned}$$

but $a^{-1}(ah) = (a^{-1}a)h = h$, so $f^{-1}(ah) = h$. Therefore the sets H and aH have a bijection between them and therefore have an equal number of elements. Finally, the choice of a was arbitrary meaning that the same set of maps can be generated for any element of G . \square

Theorem 2.15 (Lagrange's Theorem). *Let H be a subgroup of a finite group G , then the order of H divides the order of G .*

Proof. Since the cosets of H all have the same order, and since they partition G we obtain the *counting formula*:

$$\begin{aligned} |G| &= |H| [G : H] \\ (\text{order of } G) &= (\text{order of } H)(\text{number of cosets}), \end{aligned}$$

which implies that the order of H divides the order of G . \square

Notice that Lagrange's theorem does not imply that if a number divides the order of a group that there necessarily must be a subgroup of that order. That notion is the converse of Lagrange's theorem which is not true.

2.6. Simple Groups

As was mentioned earlier, Galois' core idea was to translate the problem of solvability from the language of equations to the language of groups. What Galois noticed is that normal subgroups are special. They are special because whenever a group has a normal subgroup, we can decompose the parent group into a kind of product of smaller groups. This is exactly analogous to composite and prime numbers. A composite number can be decomposed into a product of primes: e.g. $57 = 3 \cdot 19$.

As we shall see later, if the Galois group associated with a particular polynomial equation can be decomposed in a particular way then the polynomial equation will be solvable. If the group can't be decomposed, then the equation will not be solvable in radicals.

Groups which are not decomposable, are akin to prime numbers and they form the building blocks of all finite groups. Since they are the simplest building blocks in the study of groups we call them *simple*.

Definition 2.16. A nontrivial group is *simple* if its only normal subgroups are the the trivial subgroup and itself.

Since we are going to associate subgroups of S_n with polynomials, we need to find the simple groups in S_n . We will show that when $n \geq 5$ a certain class of subgroups called *alternating groups* or A_n are simple. Before we can prove this, we need to build up a few ideas.

Every permutation in S_n can be thought of as a shuffle, just like when you shuffle a deck of cards. This is because if you think of each card as representing a number from 1 to 52, then if you start with an ordered deck, the shuffle will simply permute the cards. Furthermore, any shuffle can be effected simply by a series of two card swaps. That is to say, starting with an ordered deck of cards, one can permute them into any other order simply by repeatedly swapping just two cards in the deck. In other words, any permutation can be decomposed into a series of swaps or *transpositions*. For example, in cycle notation we can write a three cycle as the product of two non-disjoint cycles.

$$(1\ 2\ 3) = (1\ 3)(1\ 2).$$

Or even a 4-cycle can be decomposed this way, for example:

$$(1\ 2\ 3\ 4) = (1\ 4)(1\ 3)(1\ 2).$$

There is a pattern to this:

$$(1\ 2\ 3\ \dots\ n-1\ n) = (1\ n)(1\ n-1)\cdots(1\ 3)(1\ 2).$$

Thus we see that the set of all transpositions in S_n generate S_n .

However, these are not the only ways to write cycles as products of transpositions. There are many ways to write a particular permutation as a product of two cycles. For example:

$$(1\ 2\ 3) = (2\ 3)(1\ 3),$$

and even,

$$(1\ 2\ 3) = (1\ 2)(2\ 3)(2\ 1)(1\ 2).$$

The above examples show that a 3-cycle can be decomposed into a product of either two or four transpositions. By multiplying the above products by a special form of the identity permutation, e.g. $(1\ 2)(1\ 2)$, then we see that we could easily write any 3-cycle as a product of any *even* number of transpositions. The question naturally arises, “Is it possible to write a 3-cycle as a product of an *odd* number of transpositions?”. Notice that the 4-cycle above required three transpositions in its decomposition. Clearly we

could use the same “multiply by 1” trick to write it as a product of an odd number of transpositions.

every cycle might be fixed. By parity we mean whether there are an even or odd number of transpositions. This is indeed true, and we shall now prove it.

Recall how the symmetries of the triangle acted on the set of six possible configurations of a triangle. We will use the same idea but our group will be one of the symmetric groups from the family S_n and we will let these permutations act on special multivariable polynomials.

Suppose $f(x_1, x_2, x_3) = (x_1 - x_2)(x_1 - x_3)(x_2 - x_3)$. Then the group S_3 acts on f , by applying the permutation to the subscripts. Suppose $\sigma \in S_3$, then we will define:

$$\sigma \bullet f(x_1, x_2, x_3) = f(x_{\sigma(1)}, x_{\sigma(2)}, x_{\sigma(3)}).$$

An example should illustrate what we mean:

$$\begin{aligned} (1\ 2\ 3) \bullet f &= (1\ 2\ 3) \bullet (x_1 - x_2)(x_1 - x_3)(x_2 - x_3) \\ &= (x_2 - x_3)(x_2 - x_1)(x_3 - x_1) \\ &= (x_2 - x_3)(-1)(x_1 - x_2)(-1)(x_1 - x_3) \\ &= (-1)^2(x_1 - x_2)(x_1 - x_3)(x_2 - x_3) \\ &= f. \end{aligned}$$

Thus the cycle $(1\ 2\ 3)$ sends this particular polynomial back to itself instead of some other polynomial, but notice that the action of the permutation $(1\ 2)$ is different:

$$\begin{aligned} (1\ 2) \bullet f &= (1\ 2) \bullet (x_1 - x_2)(x_1 - x_3)(x_2 - x_3) \\ &= (x_2 - x_1)(x_2 - x_3)(x_1 - x_3) \\ &= (-1)(x_1 - x_2)(x_2 - x_3)(x_1 - x_3) \\ &= (-1)(x_1 - x_2)(x_1 - x_3)(x_2 - x_3) \\ &= -f. \end{aligned}$$

You can check that every permutation in S_3 amounts to one of two possible mappings: $f \mapsto f$ or $f \mapsto -f$. In fact if we define f as:

$$\begin{aligned} f &= (x_1 - x_2) \cdots (x_1 - x_n)(x_2 - x_3) \cdots (x_2 - x_n) \cdots (x_{n-1} - x_n) \\ f &= \prod_{i < j} (x_i - x_j) \end{aligned}$$

where $1 \leq i < n$ and $1 < j \leq n$, then we can prove the following lemma:

Lemma 2.17. *For every transposition $\tau \in S_n$, $\tau \bullet f = -f$.*

Proof. The proof is simply a counting argument. Assume $\tau = (i\ j)$ where $i < j$. We need to count the number of factors $(x_a - x_b)$ which get mapped to a factor $(x_c - x_d)$ where $c > d$, because these are exactly the factors which will produce a factor of (-1) when we rewrite the mapped version of f . There are exactly three types of factors which lead to a sign change:

- (1) $(x_a - x_j)$ when $i < a < j$, of which there are $j - i - 1$,
- (2) $(x_i - x_b)$ when $i < b < j$, of which there are $j - i - 1$,
- (3) the single term $(x_i - x_j)$.

The -1 at the end of the first two lines above is due to the fact that when $a = b$ we overcount by 1. Thus there $2(j - i - 1) + 1$ terms which lead to a sign change, but this is an odd number thus $\tau \bullet f = -f$. \square

The following calculation shows that the action defined above on our special polynomial f is associative. In other words, acting on f by a product of permutations is the same as first acting on f by the right permutation and then on the result of that by the left permutation. If $\sigma, \rho \in S_n$ then,

$$\begin{aligned} (\sigma\rho) \bullet f &= \prod_{i < j} (x_{(\sigma\rho)(i)} - x_{(\sigma\rho)(j)}) \\ &= \sigma \bullet \left(\prod_{i < j} (x_{\rho(i)} - x_{\rho(j)}) \right) \\ &= \sigma \bullet \left(\rho \bullet \left(\prod_{i < j} (x_i - x_j) \right) \right) \\ &= \sigma \bullet (\rho \bullet f) \end{aligned}$$

If σ is any permutation in S_n and $\sigma = \tau_1\tau_2 \cdots \tau_k$, where each τ_i is a transposition then:

$$\begin{aligned} \sigma \bullet f &= (\tau_1\tau_2 \cdots \tau_k) \bullet f \\ &= \tau_1 \bullet \tau_2 \bullet \cdots \bullet \tau_k \bullet f \\ &= \tau_1 \bullet (\tau_2 \bullet (\cdots \bullet (\tau_k \bullet f) \cdots)) \\ &= (-1)^k f \end{aligned}$$

Finally, since a cycle can be decomposed into a product of transpositions in multiple equivalent ways, we need to show that all such decompositions have the same parity. Suppose $\sigma \in S_n$ and τ_1, \dots, τ_j and μ_1, \dots, μ_k are all transpositions in S_n . Finally suppose,

$$\begin{aligned} \sigma &= \tau_1, \dots, \tau_j \\ \sigma &= \mu_1, \dots, \mu_k, \end{aligned}$$

with j not necessarily equal to k . Then it follows that

$$\sigma \bullet f = (\tau_1 \dots, \tau_j) \bullet f = (-1)^j f = (-1)^k f = (\mu_1 \dots, \mu_k) \bullet f = \sigma \bullet f$$

and thus j and k must have the same parity, that is they must both be odd or even.

Definition 2.18. A permutation will be called *even* if it can be decomposed into an even number of transpositions and *odd* otherwise.

Our work above proves the following theorem:

Theorem 2.19. *A permutation in S_n is either odd or even, but not both.*

Notice that the above definition means that 3-cycles, 5-cycles, 7-cycles, etc. are even while 2-cycles, 4-cycles, 6-cycles etc. are odd. This theorem allows us to make the following table.

◦	even	odd
even	even	odd
odd	odd	even

Table 4. Multiplication (composition) of even and odd permutations

Proposition 2.4. *The subset of S_n consisting of entirely even permutations is called A_n and is a subgroup of S_n .*

Proof. associativity: Recall that the law of composition in S_n is function composition which is associative, thus the law of composition for A_n is associative as well.

closure: If $\sigma, \rho \in A_n$, then the number of transpositions in both σ and ρ is even. Their product will have an even number of transpositions as well because one way to write the product of σ and ρ is to simply concatenate all of their transpositions. The sum of two even numbers is even, thus the product of σ and ρ can be written as a product of an even number of transpositions.

identity: The identity element of S_n is the empty cycle which has is composed of zero transpositions, and thus is in A_n .

inverses: Suppose $\sigma \in A_n$, then since σ^{-1} is equivalent to σ but with each cycle written in reverse order, we see that σ^{-1} will also be even and thus in A_n . \square

Proposition 2.5. *The group A_n is a normal subgroup of S_n .*

Proof. Let $\alpha \in A_n$. We must show that for every $\sigma \in S_n$, $\sigma\alpha\sigma^{-1} \in A_n$.

Notice that by definition α must consist of an even number of transpositions, but σ could have either an odd or even number of transpositions. However, as noted in the proof of proposition 2.4 both σ and σ^{-1} both have the same number of transpositions. Thus when we concatenate all the transpositions in the product $\sigma\alpha\sigma^{-1}$ we get either: odd + even + odd, or even + even + even. Either way the new permutation will have an even number of transpositions and hence will be even. \square

The above proof is rather straightforward, but we can prove that A_n is a normal subgroup of S_n via more elegant means if we recall proposition 2.2 which says that the kernel of a homomorphism is a normal subgroup of the domain. To use this theorem we will need to create an appropriate homomorphism $\varphi : S_n \rightarrow G$ such that $A_n = \ker(\varphi)$. Table 4, the even, odd multiplication table, holds the secret if we compare it to the multiplication table for the group of integers 1 and -1, under multiplication:

\times	1	-1
1	1	-1
-1	-1	1

Table 5. Cayley table for the group $\{\pm 1\}^\times$.

The homomorphism $\varphi : S_n \rightarrow \{\pm 1\}^\times$ which does the trick is given by

$$\varphi(\sigma) = \begin{cases} 1 & \text{if } \sigma \text{ is even,} \\ -1 & \text{if } \sigma \text{ is odd.} \end{cases}$$

By construction $A_n = \ker(\varphi)$. Also by construction this is a homomorphism, although that point may be a little harder to recognize. The key to seeing why it is a homomorphism boils down to the fact that we were able to partition the elements of S_n in a way such that the equivalence classes formed by the partition form a new group. In this case there were two equivalence classes, even and odd. They partition S_n because every permutation must be in one class or the other, but cannot be in both.

Ancient Algebra

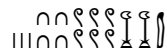
3.1. Egypt

Although the Egyptians used a base 10 number system, they did not employ a place value system like we do today. Instead they had special hieroglyphs to represent various values such as one, ten, one hundred, one thousand and so forth. Table 1 shows some of these hieroglyphs.

1	
10	∩
100	∞
1,000	⌋
10,000	⌋
100,000	⌋
1,000,000	⌋

Table 1. Values of hieroglyphic numerals

A number was represented by a jumble of these glyphs juxtaposed together. The usual practice was to put the largest numerals to the right with smaller numerals appended on going from right to left. For example, to represent the number 12,643, the Egyptians would write:



3.1.1. Multiplication by Doubling. The ancient Egyptians multiplied numbers by a method of repeated doubling. For example to perform the multiplication 13×12 , they would set up a table where they would repeatedly double the number on the right like so:

1'	12
2	24
4'	48
8'	96

Now the next number in the left hand column of the table would be 16 which is larger than 13, thus there is no need to continue the table. Notice that $1+4+8=13$, thus 13×12 is equivalent to summing the first, third and fourth entries in the right hand column of the table. That is,

$$13 \times 12 = 12 + 48 + 96 = 156.$$

Division was accomplished by changing the problem into the equivalent multiplicative statement which involved an unknown quantity which could then be deduced. The Egyptians also had a system for denoting fractional values and could multiply and divide these as well.

3.1.2. Ahmes.

- Lived during the Hyksos dynasty, about [1990-1780 BC].
- Earliest recorded name known to have some definite connection with mathematics.
- The name is found on the Rhind papyrus after A. Henry Rhind, a Scotsman who was vacationing in Egypt for his health—he had tuberculosis—in the winter of 1858. This document is now known as the Ahmes papyrus.

3.2. Ancient Mesopotamian Mathematics

Mesopotamia is a word of Greek origin which literally means “land between rivers”. It is an umbrella term for the area around the Tigris and Euphrates rivers in modern day Iraq. We have evidence that people have lived in this region since at least 4,000 BC. One of the most ancient civilizations of this region was that of the Sumerians. Later the Akkadians came to dominate but Sumerian language and culture coexisted with Akkadian language and culture. Sumerian was eventually only used by the scientific and religious scholars and during religious ceremonies.

Mesopotamians used a base 60 number system, but they did not use 60 separate symbols to represent the numbers 0-59. Instead they would use vertical marks to represent the unit and marks made at roughly 135° angle to represent 10. Thus the number 35 was written somewhat like:

\\ \\ |||||. Numbers were written according to a place value system similar to our decimal system but the columns now represented powers of 60: $\dots, 60^2, 60^1, 60^0, 60^{-1}, 60^{-2}, \dots$

The drawback of this system becomes evident when you try to multiply two numbers. A base 10 number system requires you to either memorize a 10×10 multiplication table or have one readily available. The Mesopotamian sexagesimal number system on the other hand requires a 60×60 multiplication table. Such a multiplication table has 3600 entries! As is evidenced by the high number of cuneiform multiplication table tablets that have been found, probably no one bothered to actually memorize the sexagesimal multiplication table.

Because it is cumbersome to read numbers written in cuneiform, we will adopt an equivalent but more convenient notation based upon our Arabic numerals 0–9. For example, we will write the base 10 number 65 as 1,5 where the comma separates the columns in our place value system. We can write larger numbers such as 212 like so: 3,32. A yet larger number such as 3726, will be written: 1,2,6, because

$$\begin{aligned} 3726 &= 3,600 + 120 + 6 \\ &= 1 \cdot 60^2 + 2 \cdot 60^1 + 6 \cdot 60^0 \\ &= 1,2,6 \end{aligned}$$

[Insert material about fractions here.]

3.2.1. Systems of Equations. Interestingly, very few cuneiform tablets have been found which have problems whose solution corresponds to quadratic equations in one variable. Instead most problems were posed such that solving them required you to solve a system of equations. For example, some problems correspond to the following system in the two unknowns x and y :

$$\begin{cases} x + y &= p \\ xy &= q. \end{cases}$$

This system actually corresponds to a quadratic equation in one variable. You can see this if you solve the second equation for y , yielding: $y = \frac{q}{x}$. Substituting this into the first equation yields:

$$\begin{aligned} x + \frac{q}{x} &= p \\ x^2 + q &= px. \end{aligned}$$

The key to solving this system is a special quadratic identity:

$$(3.1) \quad \boxed{\left(\frac{x+y}{2}\right)^2 - \left(\frac{x-y}{2}\right)^2 = xy.}$$

Notice that the identity has both the expressions $x+y$ and xy which occur in the original system. To solve the system, one makes the following two substitutions $x+y=p$ and $xy=q$ in the above identity giving:

$$(3.2) \quad \begin{aligned} \left(\frac{p}{2}\right)^2 - \left(\frac{x-y}{2}\right)^2 &= q \\ \left(\frac{x-y}{2}\right) &= \pm\sqrt{\left(\frac{p}{2}\right)^2 - q}. \end{aligned}$$

At this point your natural inclination might be to solve the last equation directly for x and y , but the ancients used two simple identities to finish the problem.

$$(3.3) \quad x = \left(\frac{x+y}{2}\right) + \left(\frac{x-y}{2}\right)$$

$$(3.4) \quad y = \left(\frac{x+y}{2}\right) - \left(\frac{x-y}{2}\right)$$

These new identities combined with equation (3.2) and the first equation of the original system give us the two solutions.

$$(3.5) \quad x = \left(\frac{p}{2}\right) + \sqrt{\left(\frac{p}{2}\right)^2 - q},$$

$$(3.6) \quad y = \left(\frac{p}{2}\right) - \sqrt{\left(\frac{p}{2}\right)^2 - q}.$$

With our modern symbols it is not hard to see that identity (3.1) is simply the product of equations (3.3) and (3.4), but the Mesopotamians did not use symbolic equations. Most likely they constructed the identity or something very similar via geometric means. Figure 1 provides the basis for one such construction. Although figure 1 does not appear on any ancient tablets, it does appear in Book II proposition 5 of Euclid's *Elements* which frequently drew from ancient Egyptian and Mesopotamian sources.

Lemma 3.1. *The area of rectangles $\mathcal{A} \cup \mathcal{B} \cup \mathcal{D}$ equals the area of the square formed by rectangles $\mathcal{B} \cup \mathcal{C} \cup \mathcal{D} \cup \mathcal{E}$.*

Proof. Rectangle \mathcal{A} has the same area as region $\mathcal{C} \cup \mathcal{E}$ because this region forms a rectangle with sides equal to that of rectangle \mathcal{A} \square

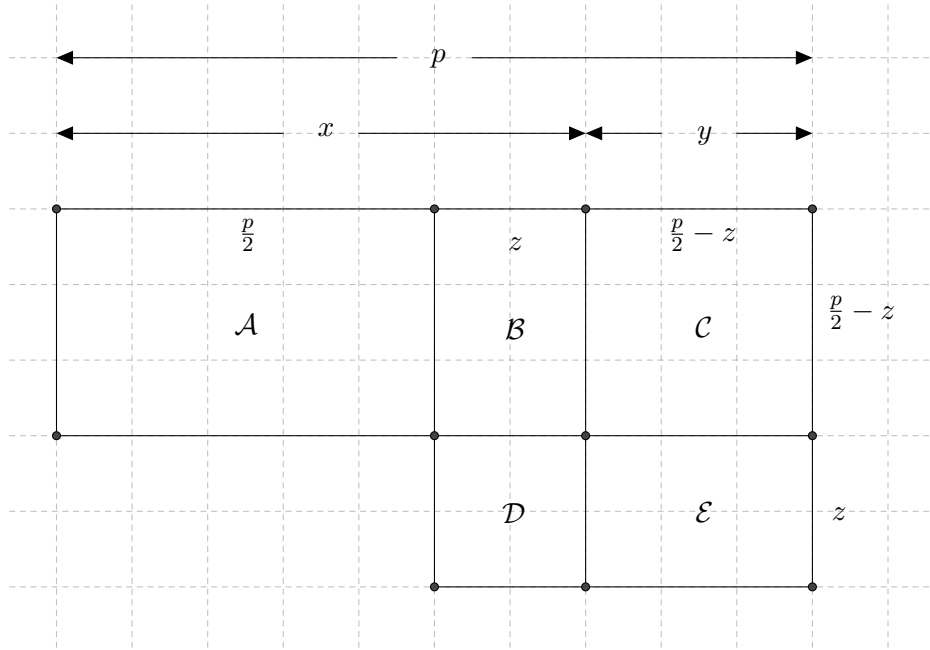


Figure 1. Geometric justification for identity (3.1).

Symbolically, the lemma can be stated as

$$q + z^2 = \left(\frac{p}{2}\right)^2,$$

because according to the original system of equations q is the area of rectangle $\mathcal{A} \cup \mathcal{B}$ because it has sides x and y , and z^2 is the area of rectangle \mathcal{D} . The proof could be expressed symbolically via

$$q + z^2 = xy + z^2 = \left(\frac{p}{2} + z\right) \left(\frac{p}{2} - z\right) + z^2 = \left(\frac{p}{2}\right)^2.$$

Finally, we see that lemma 3.1 along with figure 1 allow us to derive identity (3.1) like so:

$$\begin{aligned} \left(\frac{p}{2}\right)^2 - z^2 &= q \\ \left(\frac{p}{2}\right)^2 - \left(\frac{2z}{2}\right)^2 &= q \\ \left(\frac{x+y}{2}\right)^2 - \left(\frac{x-y}{2}\right)^2 &= xy \end{aligned}$$

3.2.2. The Ancient Method of Completing the Square. The ancients understood multiplication geometrically in terms of areas and volumes. That is, two numbers multiplied together, say ab , would be literally interpreted as the area of the rectangle formed by two segments of length a and b adjoined

at right angles. If the numbers were the same, such as in the case of x^2 , then the value would be interpreted as the area of a square. This extended up to three numbers to volumes of rectangular prisms and cubes. Thus it should come as no surprise that the best way to understand the ancient technique of completing the square is via a diagram, see figure 2.

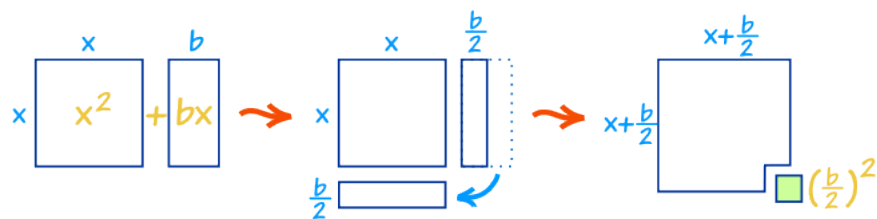


Figure 2. Geometrical interpretation of completing the square.

Before we use the method of completing the square to derive the quadratic formula, let's review how to use it to solve a quadratic equation.

A typical problem found on clay tablets is the following one: find the side of a square given that the area minus the side is 14,30. This corresponds to the modern day equation:

$$x^2 - x = 870,$$

because $14,30 = 870$. The Babylonian solution reads:

Take half of 1, which is 0;30, and multiply 0;30 by 0;30 which is 0;15. Add this to 14,30 to get 14,30;15. This is the square of 29;30. Now add 0;30 to 29;30. The result is 30, the side of the square.

3.2.3. The Modern Method of Completing the Square. Suppose we wish to solve the equation $x^2 + 5x + 7 = 0$. The first thing one always does is check to see if it factors, but since 7 is prime it only has factors of 1 and 7, and these can not be added or subtracted to yield the middle coefficient 5, thus this polynomial does not factor. Therefore we must either complete the square or use the quadratic formula to solve this system, which we will see are essentially the same.

Example 3.2.

$$\begin{aligned}
x^2 + 5x - 7 &= 0 \\
x^2 + 5x &= 7 \\
x^2 + 5x + \left(\frac{5}{2}\right)^2 &= 7 + \left(\frac{5}{2}\right)^2 \\
\left(x + \frac{5}{2}\right)^2 &= \frac{28}{4} + \frac{25}{4} \\
\left(x + \frac{5}{2}\right)^2 &= \frac{53}{4} \\
x + \frac{5}{2} &= \pm \sqrt{\frac{53}{4}} \\
x &= -\frac{5}{2} \pm \frac{\sqrt{53}}{2}
\end{aligned}$$

◇

Theorem 3.3 (The Quadratic Formula). *Given a quadratic equation of the form $ax^2 + bx + c = 0$, in the unknown x , where $a \neq 0$ (thus it actually is a quadratic equation), then there are two solutions for x given by:*

$$(3.7) \quad x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Proof.

$$\begin{aligned}
ax^2 + bx + c &= 0 \\
x^2 + \frac{b}{a}x &= -\frac{c}{a} \\
x^2 + \frac{b}{a}x + \left(\frac{b}{2a}\right)^2 &= -\frac{c}{a} + \left(\frac{b}{2a}\right)^2 \\
\left(x + \frac{b}{2a}\right)^2 &= \frac{b^2}{4a^2} - \frac{4ac}{4a^2} \\
x + \frac{b}{2a} &= \pm \sqrt{\frac{b^2 - 4ac}{4a^2}} \\
x + \frac{b}{2a} &= \pm \frac{\sqrt{b^2 - 4ac}}{2a} \\
x &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}
\end{aligned}$$

□

3.3. Ancient Greece and Alexandria

3.3.1. Pythagoras. It was one of the most surprising discoveries of the Pythagorean School of Greek mathematicians that there are irrational numbers. According to Courant and Robbins in “What is Mathematics”: This revelation was a scientific event of the highest importance. Quite possibly it marked the origin of what we consider the specifically Greek contribution to rigorous procedure in mathematics. Certainly it has profoundly affected mathematics and philosophy from the time of the Greeks to the present day.

Theorem 3.4. *The diagonal of a square whose sides are one unit long cannot be rational. That is $\sqrt{2}$ is irrational.*

Proof. (By Contradiction) Suppose $\sqrt{2}$ is rational. This means that we can write it as the ratio of two integers, p and q

$$(3.8) \quad \sqrt{2} = \frac{p}{q}$$

where p and q have no common factors.

Squaring both sides of (3.8) yields:

$$(3.9) \quad 2 = \frac{p^2}{q^2} \implies p^2 = 2q^2.$$

Therefore p^2 is even. This is only possible if p itself is even, because an odd times an odd is odd. Therefore p is even. But then p^2 is actually divisible by 4! Hence q^2 and thus q must be even. But this contradicts our initial assumption that p and q shared no common factors. Therefore our initial assumption that $\sqrt{2}$ is rational must be false. \square

Proposition 3.1. *The interior angles of a triangle sum to π radians.*

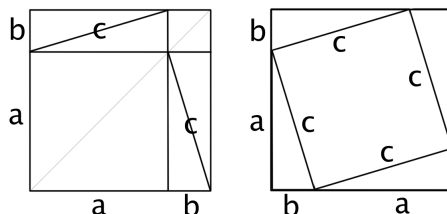
Theorem 3.5 (Pythagoras). *Given a right triangle with sides of length a and b and hypotenuse of length c , then:*

$$a^2 + b^2 = c^2.$$

Proof. The proof is by dissection. Take two equal squares, both with sides of length $a + b$. Dissect the first as shown on the left hand side of the figure. Each triangle in the right hand side square is equal to one of the triangles in the left hand square.

Since each triangle in the diagram is a right triangle, the two non-right angles in each one sum to $\pi/2$ radians. this forces each angle of the quadrangle with sides marked c on the right hand side to be $\pi/2$ radians or 90° . Thus the tilted quadrangle in the right hand figure is indeed a square with area c^2 , and not just a parallelogram.

Finally, by removing equal areas from both of the two equal squares, namely the four right triangles, we see that $a^2 + b^2 = c^2$.



□

3.3.2. Euclid. The *Elements* of Euclid is the most important mathematical text of Greek times and probably of all time. It has appeared in more editions than any work other than the *Bible*. It has been translated into countless languages and has been continuously in print in one country or another nearly since the beginning of printing. Yet to the modern reader the work is incredibly dull. There are no examples; there is no motivation; there are no witty remarks; there is no calculation. There are simply definitions, axioms, theorems and proofs. Nevertheless, the book has been intensively studied. Biographies of many famous mathematicians indicate that Euclid's work provided their initial introduction into mathematics. It provided them with a model of how "pure mathematics" should be written, with well-thought-out axioms, precise definitions, carefully stated theorems, and logically coherent proofs.

Besides his famous book, Euclid also has an important algorithm named after him—the Euclidean Algorithm—for finding the greatest common divisor or gcd of two whole numbers. Recall that the gcd of two whole numbers is simply the greatest (largest) whole number which divides both numbers. The gcd of 72 and 180 is 36 because $180 = 2^2 \cdot 3^2 \cdot 5$ and $72 = 2^3 \cdot 3^2$, and thus the greatest common divisor is $2^2 \cdot 3^2 = 36$.

The algorithm is simple, but often it is not taught because we can rely upon prime factorization as shown above. However, when the numbers are very, very large, for example a number with a hundred digits, then the prime factorization technique becomes slow because we don't have very fast algorithms for factorization. Euclid's algorithm on the other hand is very fast even for extremely large numbers. The algorithm is recursive (meaning that it is applied repeatedly to smaller and smaller inputs) and based upon division.

Example 3.6. Here are the steps required to compute $\text{gcd}(5463, 381)$:

$$\begin{array}{rcl}
5463 = 381 \cdot 14 + 129 & \gcd(5463, 381) = \gcd(381, 129) \\
381 = 129 \cdot 2 + 123 & \gcd(381, 129) = \gcd(129, 123) \\
129 = 123 \cdot 1 + 6 & \gcd(129, 123) = \gcd(123, 6) \\
123 = 6 \cdot 20 + 3 & \gcd(123, 6) = \gcd(6, 3) \\
6 = 3 \cdot 2 + 0 & \gcd(6, 3) = 3
\end{array}$$

Notice how the divisor becomes the dividend and the remainder becomes the divisor on each next line, where:

$$\text{dividend} = \text{divisor} \cdot \text{quotient} + \text{remainder}.$$

The quotient is discarded at each step, and the algorithm terminates when the remainder is 0. The gcd will equal the last nonzero remainder. \diamond

Proposition 3.2 (The Elements VII.2). *blah*

There is an equivalent and often more useful way of stating the Euclidean algorithm.

Corollary 3.1. *For every pair of whole numbers a and b , there exist two integers s, t (perhaps negative) such that:*

$$a \cdot s + b \cdot t = \gcd(a, b).$$

Proof. The proof is by induction on the number of steps in the Euclidean algorithm. Define $\text{Eulen}(a, b)$ to be the number of steps required to compute $\gcd(a, b)$, thus $\text{Eulen}(a, b)$ is a natural or counting number. Assuming $a > b$, we have to prove the basis step and the inductive step.

[$\text{Eulen}(a, b) = 1$]: If the algorithm terminates in one step, then $b \mid a$ (b divides a), and hence $a = bu$ where u is an integer. Hence,

$$a \cdot \underbrace{1}_s + b \cdot \underbrace{(1 - u)}_t = b = \gcd(a, b).$$

[$\text{Eulen}(a, b) = n$]: Apply the division algorithm to a and b to yield:

$$a = bq + r \quad q, r \in \mathbb{Z}.$$

$\text{Eulen}(b, r) = n - 1$, thus by the inductive hypothesis there exist $x, y \in \mathbb{Z}$ such that

$$bx + ry = \gcd(b, r) = \gcd(a, b).$$

But $r = a - bq$ thus $ry = ay - bqy$, hence

$$\begin{aligned}
bx + ay - bqy &= \gcd(a, b), \\
a \cdot \underbrace{y}_s + b \cdot \underbrace{(x - qy)}_t &= \gcd(a, b).
\end{aligned}$$

\square

If you analyze the proof carefully, you will see that it actually gives us a way to construct the linear combination as well. This is best illustrated by an example.

Example 3.7. How to compute s and t in the integral linear combination:

$$5463 \cdot s + 381 \cdot t = 3 = \gcd(5463, 381).$$

First, solve each equation from the Euclidean algorithm (except for the last one) for the remainder.

$$5463 = 381 \cdot 14 + 129 \quad \longrightarrow \quad 129 = 5463 - 381 \cdot 14 \quad (3.10a)$$

$$381 = 129 \cdot 2 + 123 \quad \longrightarrow \quad 123 = 381 - 129 \cdot 2 \quad (3.10b)$$

$$129 = 123 \cdot 1 + 6 \quad \longrightarrow \quad 6 = 129 - 123 \cdot 1 \quad (3.10c)$$

$$123 = 6 \cdot 20 + 3 \quad \longrightarrow \quad 3 = 123 - 6 \cdot 20 \quad (3.10d)$$

This last equation is our starting point for unrolling the recursive steps of the Euclidean algorithm. We substitute equation (3.10c) into equation (3.10d) to obtain the new equation:

$$3 = 123 - (129 - 123 \cdot 1) \cdot 20$$

$$(*) \quad 3 = 123 \cdot 21 - 129 \cdot 20$$

Next, substitute equation (3.10b) into equation (*):

$$3 = (381 - 129 \cdot 2) \cdot 21 - 129 \cdot 20$$

$$(**) \quad 3 = 381 \cdot 21 - 129 \cdot 62$$

Finally, substitute equation (3.10a) into equation (**):

$$3 = 381 \cdot 21 - (5463 - 381 \cdot 14) \cdot 62$$

$$(***) \quad 3 = 381 \cdot 889 + 5463 \cdot (-62)$$

Thus $s = -62$ and $t = 889$.

◇

Proposition 3.3 (The Elements VII.24). *If two numbers are relatively prime to any number, then their product is also relatively prime to the same.*

Proof. Suppose a is relatively prime to both b and c . Since a and b are relatively prime, there exist integers (perhaps negative) m and n such that $ma + nb = 1$. Similarly $ja + kc = 1$ for some j, k .

Multiplying these two equations together,

$$\begin{aligned} (ma + nb)(ja + kc) &= 1 \\ &= maja + makc + nbja + nbkc \\ &= (maj + mkc + nbj)a + (nk)bc \\ &= 1 \end{aligned}$$

so a and bc are relatively prime. □

Repeating the argument verifies that if a is relatively prime to b , then a is relatively prime to b^n for any positive integer n .

Proposition 3.4 (The Elements VII.30). *If two numbers, multiplied by one another make some number, and any prime number measures (divides) the product, then it also measures one of the original factors.*

Proof. Let a prime p divide the product ab . Assume $p \nmid a$. Then $\gcd(a, p) = 1$. By Corollary, $ax + py = 1$ for some x and y . Multiply by b : $abx + pby = b$. Now, $p \mid ab$ and $p \mid pby$. Hence, $p \mid b$. \square

With these tools in hand, we can now prove the Rational Root theorem.

Theorem 3.8 (Rational Root Theorem). *Let $P(x)$ be a polynomial with integer coefficients, say*

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

and suppose that $r = \frac{c}{d}$ is a rational root of P , that is $P(r) = 0$, expressed in lowest terms (so that c and d are relatively prime). Then c divides a_0 and d divides a_n .

Proof. Inserting the argument $x = \frac{c}{d}$ into the expression for $P(x)$ yields:

$$0 = a_n \frac{c^n}{d^n} + a_{n-1} \frac{c^{n-1}}{d^{n-1}} + \cdots + a_1 \frac{c}{d} + a_0$$

Multiplying through by d^n and isolating the first term yields:

$$-a_n c^n = a_{n-1} c^{n-1} d + \cdots + a_1 c d^{n-1} + a_0 d^n.$$

Since d is a factor of every term on the right hand side of this equation, d must divide $a_n c^n$. But c and d are relatively prime, so d and c^n are relatively prime, and it follows from proposition VII.30 that d divides a_n .

Isolating the last term instead of the first, we see that

$$a_n c^n + a_{n-1} c^{n-1} d + \cdots + a_1 c d^{n-1} = -a_0 d^n.$$

As before, since c is a factor of every term in the left side of this equation, c must divide $a_0 d^n$. Since c and d are relatively prime, c and d^n are relatively prime, and we conclude that c must divide a_0 . \square

Now consider the equation for the n th root of an integer t : $x^n - t = 0$. If $r = c/d$ is a rational n th root of t expressed in lowest terms, the Rational Root Theorem states that d divides 1, the coefficient of x^n . That is, that d must equal 1, and $r = c$ must be an integer, and t must be itself a perfect n th power.

3.3.3. Diophantus. Although we know little about Diophantus' life other than that he lived in Alexandria, we know that his book *Arithmetica* was known to Islamic scholars. *Arithmetica* like Euclid's *Elements* was divided into thirteen books. Only ten of these books have survived, six in Greek and four in Arabic.

Diophantus studied several higher order equations such as fourth order polynomial equations and even sixth order. Especially of note is that he is the first person to develop a symbolic notation for algebraic problems.

$$\begin{array}{c}
 K^{\gamma} \alpha \varsigma \gamma \Lambda \Delta^{\gamma} M^{\circ} \alpha \\
 x^3 \cdot 1 + x \cdot 3 - (x^2 \cdot 3 + 1) \\
 x^3 - 3x^2 + 3x - 1
 \end{array}$$

Medieval Algebra

4.1. Medieval Persia

4.1.1. Al-Khwārizmī. Muhammad ibn Mūsā al-Khwārizmī lived in the ninth century where present day Baghdad exists. He is known as the father of Algebra. He is famous for classifying the solvable (at that time) algebraic equations into six types.

(1)	$ax^2 = bx$	Squares are equal to roots
(2)	$ax^2 = c$	Squares are equal to a number
(3)	$bx = c$	Roots are equal to a number
(4)	$ax^2 + bx = c$	Squares and roots are equal to a number
(5)	$ax^2 = bx + c$	Roots and number are equal to squares
(6)	$ax^2 + c = bx$	Squares and a number are equal to roots

Table 1. Al-Khwārizmī's classification of equations

Although Al-Khwārizmī may have been familiar with Diophantus' symbolic way of writing a problem, he did not adopt it, but it is abundantly clear that the Greek way of doing mathematics had a strong influence upon him. Even though his book entitled *Al-kitāb al-muhtasar fī hisāb al-jabr wa-l-muqābala* (*The Condensed Book on the Calculation of al-Jabr and al-Muqābala*) was intended to primarily be a practical book of instruction he felt compelled to provide geometric proofs of many of his procedures. This book showed how to use *al-jabr* which can be translated as “restoring” and

in modern day terminology refers to the act of changing a negative quantity on one side of an equation to a positive quantity on the other side, to transform a problem into one of his six canonical forms. The word *al-muqābala* refers to eliminating a positive term by subtracting equal amounts from both sides of the equation. For example, the conversion of $5x+2=6$ to $5x=4$ is an example of *al-muqābala*. While the conversion of the equation $5x-2=4-2x$ to $3x-2=4$ is an example of *al-jabr*.

4.1.2. Omar Khayyam. Omar Khayyam lived in the 11th century and thus was not a contemporary of Al-Khwārizmī. He is known for finding a way to solve cubic equations geometrically. The basic idea of his method is to transform a single cubic equation into two conic section equations and then finding their point(s) of intersection.

The types of conic sections involved could be a circle, hyperbola or parabola depending on the form the equation takes. If we assume a monic equation (leading coefficient of one) and only allow all other coefficients to be positive or zero, then any cubic equation can be written in one of the fourteen forms given in table 2.

Equation	Solutions	Curves
(1) $x^3 = c$	$x_1 > 0; x_{2,3} \in \mathbb{C}$	P,P
(2) $x^3 + bx = c$	$x_1 > 0; x_{2,3} \in \mathbb{C}$	C,P
(3) $x^3 + c = bx$	$x_{1,2} > 0$ or $\in \mathbb{C}; x_3 < 0$	P,H
(4) $x^3 = bx + c$	$x_1 > 0; x_{2,3} < 0$ or $\in \mathbb{C}$	P,H
(5) $x^3 + ax^2 = c$	$x_1 > 0; x_{2,3} < 0$ or $\in \mathbb{C}$	P,H
(6) $x^3 + c = ax^2$	$x_{1,2} > 0$ or $\in \mathbb{C}; x_3 < 0$	P,H
(7) $x^3 = ax^2 + c$	$x_1 > 0; x_{2,3} \in \mathbb{C}$	P,H
(8) $x^3 + ax^2 + bx = c$	$x_1 > 0; x_{2,3} < 0$ or $\in \mathbb{C}$	C,H
(9) $x^3 + ax^2 + c = bx$	$x_{1,2} > 0$ or $\in \mathbb{C}; x_3 < 0$	H,H
(10) $x^3 + bx + c = ax^2$	$x_{1,2} > 0$ or $\in \mathbb{C}; x_3 < 0$	C,H
(11) $x^3 = ax^2 + bx + c$	$x_1 > 0; x_{2,3} < 0$ or $\in \mathbb{C}$	H,H
(12) $x^3 + ax^2 = bx + c$	$x_1 > 0; x_{2,3} < 0$ or $\in \mathbb{C}$	H,H
(13) $x^3 + bx = ax^2 + c$	$x_1 > 0; x_{2,3} > 0$ or $\in \mathbb{C}$	C,H
(14) $x^3 + c = ax^2 + bx$	$x_{1,2} > 0$ or $\in \mathbb{C}; x_3 < 0$	H,H

Table 2. Omar Khayyam's classification of cubic equations

Example 4.1. Let us show how to solve a type (2) equation: $x^3 + bx = c$, by intersecting a circle and parabola. First we let $p = \sqrt{b}$ and $q = \frac{c}{b}$. After making this substitution, equation (2) from the table becomes:

$$x^3 + p^2x = p^2q.$$

Next, we draw a semicircle of diameter q and a parabola of parameter p with vertex coinciding with the leftmost point of the semicircle as shown in figure 1. Recall that Cartesian coordinates had not been invented yet, so Khayyam would not have drawn axes. However, he did have tools with which he could accurately draw circles and all the other conic sections.

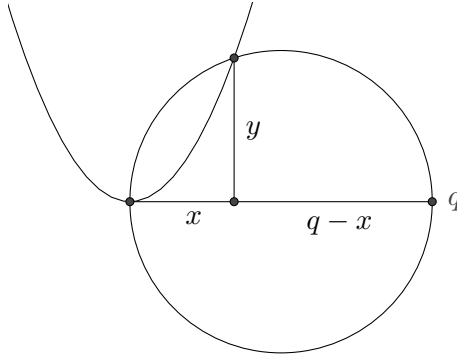


Figure 1. Omar Khayyam's geometric method for solving cubic equations.

The equation of the circle is $(x - \frac{q}{2})^2 + y^2 = (\frac{q}{2})^2$, which can be rewritten:

$$(4.1) \quad \frac{q-x}{y} = \frac{y}{x}.$$

It should be pointed out that Khayyam did not use the equation of a circle with center shifted to the right on the x axis. Most likely, he reasoned via Thale's theorem which states that any triangle inscribed in a circle with hypotenuse on a diameter is a right triangle. From this theorem one can deduce equation (4.1).

The equation of the parabola is $py = x^2 \implies y = \frac{1}{p}x^2$ or

$$(4.2) \quad \frac{y}{x} = \frac{x}{p}.$$

If we combine equation (4.1) with equation (4.2) we get:

$$(4.3) \quad \frac{q-x}{y} = \frac{x}{p}$$

Multiplying equation (4.3) by py we obtain:

$$(4.4) \quad pq - px = xy = \frac{x^3}{p}$$

This yields: $x^3 + p^2x = p^2q$, which is the original cubic equation we wished to solve. \diamond

4.2. Medieval Italy

4.2.1. Leonardo of Pisa [1170–1240].

- Leonardo's father was an official of the independent republic of Pisa, and was appointed to represent its merchants in the trading colony of Bugia on the North African coast in 1192.
- Leonardo accompanied his father and was exposed to the works of many Muslim scholars.
- Wrote: *Liber abacci*, "Book of Calculation" which introduced Arabic (actually Indian) numerals including zero to the West. The book borrowed many problems verbatim from Al-Khwārizmī, Abu Kamil and others.
- Later known as *Fibonacci*, literally son of Bonacci.

Renaissance Algebra

5.1. Italy

5.1.1. Luca Pacioli [1445–1517].

- Invented double-entry bookkeeping.
- Friend of Leonardo da Vinci
- Coined the term million
- Wrote: *Summa* in 1494. This book did not break any new ground, but standardized the current notation of the day. He used *co* for *cosa* (“thing”), *ce* for *census* (“property”), and *cu* for *cubus* (“cube”). Classified the following types of cubic equations as unsolvable:
 - (1) $n = ax + bx^3$
 - (2) $n = ax^2 + bx^3$
 - (3) $ax + n = bx^3$
- The third classification above was not actually listed in his book, but we’ll refer to this classification system below.

5.1.2. The Story of the Cubic. The following is verbatim from John Derbyshire’s book “Unknown Quantity: A Real and Imaginary History of Algebra”.

At some point in the early 16th century, a person named Scipione del Ferro found the general solution to the type 1 cubic. Del Ferro was professor of mathematics at the University of Bologna; his dates are ca. 1456–1526. We don’t know exactly when he got his solution or whether he also solved type 2. He never published his solution.

Before del Ferro died, he imparted the secret of his solution for “the cosa and the cube” to one of his students, a Venetian named Antonio Maria Fiore. This poor fellow has gone down in all the history books as a mediocre mathematician. I don’t doubt the judgment of the historians, but it seems a great misfortune for Fiore to have gotten mixed up—as a catalyst, so to speak—in such a great and algebraically critical affair, so that his mathematical mediocrity echoes down the ages like this. At any rate, having gotten the secret of the cosa and the cube, he decided to make some money out of it. This wasn’t hard to do in the buzzing intellectual vitality of northern Italy at the time. Patronage was hard to come by, university positions were not well paid, and there was no system of tenure. For a scholar to make any kind of living, he needed to publicize himself, for example, by engaging in public contests with other scholars. If some large cash prize was at stake in the contest, so much better the publicity.

One mathematician who had made a name for himself in this kind of contest was Nicolo Tartaglia, a teacher in Venice. Tartaglia came from Brescia, 100 miles west of Venice. When he was 13, a French army sacked Brescia and put the townsfolk to the sword. Nicolo survived but suffered a grievous saber wound on his jaw, which left him with a speech impediment: Tartaglia means “stutterer”—this was still the age when last names were being formed out of locatives, patronymics, and nicknames. Tartaglia was a mathematician of some scope, author of a book on the mathematics of artillery, and the first person to translate Euclid’s *Elements* into Italian.

In 1530, Tartaglia had exchanged some remarks about cubic equations with another native of Brescia, a person named Zuanne de Tonini da Coi, who taught mathematics in that town. In the course of those exchanges, Tartaglia claimed to have found a general rule for the solution of type 2 cubics, though he confessed he could not solve type 1.

Somehow Fiore, the mathematical mediocrity, heard of these exchanges and of Tartaglia’s claim. Either believing Tartaglia to be bluffing or confident that he was the only person who knew how to solve type 1 cubics (the secret he had gotten from del Ferro), Fiore challenged Tartaglia to a contest. Each was to present the other with 30 problems. Each was to deliver the 30 solutions to the other’s problems to a notary on February 22, 1535. The loser was to stand the winner 30 banquets.

Having no great regard for Fiore’s mathematical talents, Tartaglia at first did not bother to prepare for the contest. However,

someone passed on the rumor that Fiore, though no great mathematician himself, had learned the secret of solving “the cosa and the cube” from a master mathematician, since deceased. Now worried, Tartaglia bent his talents to finding a general solution of type 1 cubics. In the small hours of the morning of Saturday, February 13, he cracked it. As he had suspected, all of Fiore’s problems were type 1 cubics, the solution of which was Fiore’s sole claim to mathematical ability.

Tartaglia’s questions seem (we only have the first four) to have been a mix of types 2 and 3. It is plain that at this point Tartaglia had mastered all the cubics, of any type, having just one real solution—all the ones, that is, with a positive discriminant. Cubic equations with a negative discriminant (and therefore having three real solutions) can only be solved by manipulating complex numbers, which had not yet been discovered. At any rate, Tartaglia was able to solve all of Fiore’s problems, while Fiore could solve none of his. Tartaglia took the honor but waived the stake. Comments Cardano’s biographer: “The prospect of thirty banquets face to face with a sad loser may have been rather uninspiring to him.”

5.1.3. Scipione del Ferro [1456–1526].

- The first to solve the depressed cubic for the case where the discriminant is negative.
- A professor at the University of Bologna.
- Shared his secret solution of the cubic with his student Antonio Maria Fiore, who subsequently challenged Tartaglia.
- Never published his solution of the cubic.

5.1.4. Niccolo Fontana Tartaglia [1499–1557].

- Mathematician, engineer, surveyor and bookkeeper
- “Tartaglia” means stutterer in Italian.
- Translated Euclid’s *Elements* into Italian.
- Independently discovered how to solve cubic equations with negative discriminant.
- Confided his solution of the cubic to Cardano in the form of a poem, but later regretted it.

5.1.5. Girolamo Cardano [1501–1576].

- A physician by trade.
- Avid gambler and caster of horoscopes.

- Wrote at least 131 publications including an autobiography. Several of his books were bestsellers in Europe.
- Cajoled Tartaglia into divulging his secret solution to the cubic.
- Wrote *Ars Magna* [1545] in which he published a general method of solving cubic equations as well as quartic equations.

5.2. Modern Derivation of Cardano's Formula

We will derive the equation known as Cardano's formula. This formula was first discovered by Tartaglia on February 12, 1535. Tartaglia subsequently entrusted Cardano with his secret to solving the cubic after the latter promised to not publish it. Cardano broke this promise and published the formula in his book *Ars magna*. Consequently, it is usually known as Cardano's formula. This formula gives you one real root of any cubic equation with real coefficients such as the one given below.

$$(5.1) \quad Ay^3 + By^2 + Cy + D = 0$$

Before we do anything we notice that if this is truly a cubic equation then $A \neq 0$ and thus we can divide both sides of the equation by A to obtain a monic equation, that is a polynomial equation where the leading coefficient is 1.

$$(5.2) \quad y^3 + ay^2 + by + c = 0 \quad \text{where } a = \frac{B}{A}, b = \frac{C}{A}, d = \frac{D}{A}$$

And of course both of these equations share the same roots.

The derivation is done in two steps, first we *reduce* the cubic polynomial. This simply refers to a clever trick to rewrite the polynomial such that it has no quadratic term. We can accomplish this by letting $y = x + k$ in equation (5.2).

$$\begin{aligned} y^3 + ay^2 + by + c &= (x + k)^3 + a(x + k)^2 + b(x + k) + c \\ &= (x^3 + 3x^2k + 3xk^2 + k^3) + a(x^2 + 2xk + k^2) + b(x + k) + c \\ &= x^3 + (3k + a)x^2 + (3k^2 + 2ak + b)x + (k^3 + ak^2 + bk + c) \\ &= 0. \end{aligned}$$

We are free to choose whatever value we wish for k . If we choose $k = -\frac{a}{3}$, then the quadratic term will vanish leaving:

$$x^3 + \left(b - \frac{a^2}{3}\right)x + \left(2\left(\frac{a}{3}\right)^3 - \frac{ab}{3} + c\right) = 0.$$

Thus we have reduced the original equation to the somewhat simpler equation:

$$(5.3) \quad x^3 + px + q = 0.$$

Where,

$$p = b - \frac{a^2}{3}$$

$$q = 2 \left(\frac{a}{3}\right)^3 - \frac{ab}{3} + c$$

At this point everything we have done is reversible. If we can solve equation (5.3) for x , then we will have a solution for the original equation as well because $y = x + k$.

The next step of the derivation requires the counter-intuitive notion of rewriting the unknown x as the sum of two unknowns u and v . That is we will let $x = u + v$. At first, this seems ludicrous! How could changing a single variable problem into a problem with two unknowns possibly make our lives easier? Actually, this is an ancient trick. Omar Khayyam did it when he geometrically solved the cubic, and the Babylonians were certainly aware of it because they used it to solve quadratic equations, and Tartaglia was an excellent mathematician, so he was certainly aware of it.

The reason we choose to write $x = u + v$ is to exploit the identity:

$$(u + v)^3 = u^3 + 3u^2v + 3uv^2 + v^3,$$

which we will rewrite:

$$(u + v)^3 - 3uv(u + v) - (u^3 + v^3) = 0.$$

If $x = u + v$ is to be a solution, then by comparing this equation to equation (5.3), u and v must satisfy two new equations:

$$u^3 + v^3 = -q$$

$$uv = -\frac{p}{3}.$$

A second identity:

$$\left(\frac{u^3 - v^3}{2}\right)^2 = \left(\frac{u^3 + v^3}{2}\right)^2 - u^3v^3,$$

allows us to write:

$$(5.4) \quad \frac{u^3 + v^3}{2} = -\frac{q}{2}$$

$$(5.5) \quad \frac{u^3 - v^3}{2} = \pm \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}.$$

The sum and difference of equations (5.4) and (5.5) allows us to solve for u and v .

$$u = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}$$

$$v = \sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}$$

Cardano's formula is simply the sum of the two expressions for u and v :

$$(5.6) \quad x = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}$$

5.3. Ferrari's Solution to the Biquadratic

One might imagine that since quartic or biquadratic equations are more complicated than cubic equations, their solution would require a bold stroke of genius similar to Tartaglia's choice of letting the unknown x equal the sum of two new unknowns, $u + v$, but actually it only involves incorporating a single new variable.

We wish to solve the equation,

$$(5.7) \quad y^4 + ay^3 + by^2 + cy + d = 0.$$

Our first step, just like Tartaglia is to *reduce* the equation. Let $y = x + k$ in (5.7),

$$(x + k)^4 + a(x + k)^3 + b(x + k)^2 + c(x + k) + d = 0.$$

After expanding this expression and collecting like powers of x we get:

$$\begin{aligned} & x^4 + \\ & (a + 4k)x^3 + \\ & (b + 3ak + 6k^2)x^2 + \\ & (c + 2bk + 3ak^2 + 4k^3)x + \\ & (d + ck + bk^2 + ak^3 + k^4) = 0 \end{aligned}$$

Which tells us to choose $k = -\frac{a}{4}$, whereupon we get the reduced equation:

$$(5.8) \quad x^4 + px^2 + qx + r = 0.$$

The next step is the clever part. Add the expression $2zx^2 + z^2$ to both sides of the equation and rearrange. Since we are introducing a new variable z ,

into the equation it gives us a new degree of freedom which we will soon exploit.

$$\begin{aligned} x^4 + 2zx^2 + z^2 &= 2zx^2 + z^2 - px^2 - qx - r \\ (5.9) \quad (x^2 + z)^2 &= (2z - p)x^2 - qx + (z^2 - r). \end{aligned}$$

Adding this expression allows us to factor the left hand side of the equation into a perfect square, but it also allows us to factor the right hand side of the equation into a perfect square. To see this, consider the equation $ax^2 - qx + b = 0$. If $-q = 2\sqrt{a}\sqrt{b}$ then we can factor:

$$\begin{aligned} ax^2 + 2\sqrt{a}\sqrt{b}x + b &= 0 \\ (\sqrt{a}x + \sqrt{b})(\sqrt{a}x + \sqrt{b}) &= 0 \\ (\sqrt{a}x + \sqrt{b})^2 &= 0. \end{aligned}$$

We can apply this technique to our problem. If we choose z such that:

$$(5.10) \quad -q = 2\sqrt{2z - p}\sqrt{z^2 - r},$$

then the right hand side of equation (5.9) factors and we get,

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

Upon solving for x , we get:

$$(5.11) \quad x_{1,2} = \frac{1}{2}\sqrt{2z - p} \pm \sqrt{-\frac{1}{2}z - \frac{1}{4}p + \sqrt{z^2 - r}}$$

$$(5.12) \quad x_{3,4} = \frac{1}{2}\sqrt{2z - p} \pm \sqrt{-\frac{1}{2}z - \frac{1}{4}p - \sqrt{z^2 - r}}$$

You may wonder whether we can actually find a z such that equation (5.10) is satisfied. It turns out that equation (5.10) is actually a cubic equation in disguise.

$$\begin{aligned} 2\sqrt{2z - p}\sqrt{z^2 - r} &= -q \\ (2z - p)(z^2 - r) &= \frac{q^2}{4} \\ 2z^3 - pz^2 - 2rz + pr &= \frac{q^2}{4} \\ (5.13) \quad z^3 - \frac{p}{2}z^2 - rz + \left(\frac{pr}{2} - \frac{q^2}{8}\right) &= 0 \end{aligned}$$

Thus by Cardano's formula we are guaranteed to be able to find a z which satisfies equation (5.10).

Notice that Ferrari's solutions, equations (5.11) and (5.12) are expressed in terms of z which is obtained via Cardano's formula and may involve a complicated expression with square roots inside a cube root. Thus, although

Ferrari's solution does not appear too complicated, if you were to replace every occurrence of z in these formulas with Cardano's formula, the resulting formulas would fill an entire page! These are truly complicated formulas.

5.3.1. Lodovico Ferrari [1522–1565].

- Was a secretary to Cardano beginning at age 14.
- Discovered a way to solve the general quartic equation by transforming it to a cubic equation.
- Allowed Cardano to publish his solution to the quartic in *Ars Magna*.



5.3.2. Rafael Bombelli [1526–1572].

- A civil engineer who was responsible for draining marshland in central Italy.
- His friend Antonio Maria Pazzi introduced him to Diophantus' writings which were at the university in Rome.
- Wrote *l'Algebra* [1572] with the goal of making an easier to understand version of Cardano's *Ars Magna*.
 - (1) First clear usage of negative and complex numbers.
 - (2) Included 143 problems from Diophantus' writings.
 - (3) Was the first introduction to Diophantus' writings for most Europeans.
 - (4) Still lacked good symbolism.

5.4. Extending Cardano's Formula with Complex Numbers

Cardano's formula (5.6) gives us only one solution to a cubic equation, but of course by the Fundamental Theorem of Algebra, every cubic equation has three solutions (with perhaps repetition). In this section we will show how to find the other two solutions.

The key idea is to multiply the two cube roots in Cardano's formula by special numbers called the cube roots of unity. The cube roots of unity are exactly the complex valued solutions to the equation:

$$z^3 = 1,$$

and consequently yield 1 when cubed. Clearly this equation has solution $z_1 = 1$, thus to find the other two solutions we use polynomial long division to divide $z^3 - 1$ by $z - 1$ to obtain the following quadratic expression:

$$(5.14) \quad \frac{z^3 - 1}{z - 1} = z^2 + z + 1.$$

The quadratic equation, $z^2 + z + 1 = 0$ has two complex-conjugate solutions:

$$z_{1,2} = \frac{-1 \pm \sqrt{-3}}{2}.$$

We immediately rewrite these solutions in complex number notation: $a + bi$ where $a, b \in \mathbb{R}$ and $i = \sqrt{-1}$.

$$z_{1,2} = -\frac{1}{2} \pm \frac{\sqrt{3}}{2}i.$$

Recall that complex numbers are multiplied according to the following rule:

$$\begin{aligned} (a + bi)(c + di) &= ac + (bc + ad)i + bdi^2 \\ &= ac + (bc + ad)i - bd \\ &= (ac - bd) + (bc + ad)i. \end{aligned}$$

Let's check whether our solution is correct, that is whether $z_1^3 = z_2^3 = 1$?

$$\begin{aligned} z_1^3 &= \left(-\frac{1}{2} + \frac{\sqrt{3}}{2}i\right) \left(-\frac{1}{2} + \frac{\sqrt{3}}{2}i\right) \left(-\frac{1}{2} + \frac{\sqrt{3}}{2}i\right) \\ &= \left(-\frac{1}{2} - \frac{\sqrt{3}}{2}i\right) \left(-\frac{1}{2} + \frac{\sqrt{3}}{2}i\right) \\ &= 1 \end{aligned}$$

The middle equation shows that $z_1^2 = z_2$. How about z_2 ?

$$\begin{aligned} z_2^3 &= \left(-\frac{1}{2} - \frac{\sqrt{3}}{2}i\right) \left(-\frac{1}{2} - \frac{\sqrt{3}}{2}i\right) \left(-\frac{1}{2} - \frac{\sqrt{3}}{2}i\right) \\ &= \left(-\frac{1}{2} + \frac{\sqrt{3}}{2}i\right) \left(-\frac{1}{2} - \frac{\sqrt{3}}{2}i\right) \\ &= 1 \end{aligned}$$

The middle calculation here shows that $z_2^2 = z_1$! This is remarkably similar to the symmetries of the equilateral triangle. There we had three rotational symmetries, $\{1, R_{120}, R_{240}\}$ where the subscript stood for degrees of counter-clockwise rotation. Notice,

$$\begin{aligned} R_{120}^2 &= R_{120} \circ R_{120} = R_{240}, \text{ and} \\ R_{240}^2 &= R_{240} \circ R_{240} = R_{120}. \end{aligned}$$

[Insert derivation of Euler's Formula $e^{ix} = \cos x + i \sin x$ here.]

Now let's return to our original goal and modify Cardano's formula to obtain formula's for all three roots of the cubic equation and not just a single

real root. For the sake of short formulas, we will denote the two complex cube roots of unity by ω and ω^2 , where

$$\omega = \left(-\frac{1}{2} + \frac{\sqrt{3}}{2}i \right) \quad \omega^2 = \left(-\frac{1}{2} - \frac{\sqrt{3}}{2}i \right)$$

Nine Possibilities:

$$\begin{aligned} x_1 &= u + v \\ x_2 &= u + \omega v \\ x_3 &= u + \omega^2 v \\ x_4 &= \omega u + v \\ x_5 &= \omega u + \omega v \\ x_6 &= \omega u + \omega^2 v \\ x_7 &= \omega^2 u + v \\ x_8 &= \omega^2 u + \omega v \\ x_9 &= \omega^2 u + \omega^2 v \end{aligned}$$

This is kind of absurd, we have nine different possible roots, but a cubic equation only has three roots! It turns out that six of these solutions are not valid, because if you recall, when Tartaglia set $x = u + v$ he found that if the solutions were to have this form, then they must satisfy a system of two equations, namely:

$$\begin{cases} u^3 + v^3 &= -q \\ uv &= -\frac{p}{3}. \end{cases}$$

By design, all nine possible roots satisfy the first equation in this system, but only candidates x_1, x_6 and x_8 satisfy the second equation. After reindexing, these are the three solutions of the cubic.

$$(5.15) \quad x_1 = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}$$

$$(5.16) \quad x_2 = \omega \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} + \omega^2 \sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}$$

$$(5.17) \quad x_3 = \omega^2 \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} + \omega \sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}$$

5.5. France

5.5.1. François Viète [1540–1603].

- Born into a Huguenot family (French protestant).
- A lawyer by trade.
- Privy councillor to both Henry III and Henry IV.
- Wrote *In artem analyticem isagoge* or “Introduction to the Analytic Art”.
- This book had the first systematic use of letters to represent numbers. Unknowns were represented by uppercase vowels (A,E,I,O,U,Y) and data such as coefficients were represented by uppercase consonants.
- Discovered that polynomial coefficients can always be represented via symmetric polynomials in the roots. This led Lagrange and later Galois on to important discoveries.



The last bullet point above needs explanation. Let’s work backwards by supposing we know the roots of a polynomial. If these roots are labelled r_1 through r_4 , then we can write quadratic, cubic or quartic polynomials in these roots in the following ways:

$$\begin{aligned}
 (x - r_1)(x - r_2) &= x^2 \\
 &\quad - (r_1 + r_2)x \\
 &\quad + r_1r_2 \\
 (x - r_1)(x - r_2)(x - r_3) &= x^3 \\
 &\quad - (r_1 + r_2 + r_3)x^2 \\
 &\quad + (r_1r_2 + r_1r_3 + r_2r_3)x \\
 &\quad - r_1r_2r_3 \\
 (x - r_1)(x - r_2)(x - r_3)(x - r_4) &= x^4 \\
 &\quad - (r_1 + r_2 + r_3 + r_4)x^3 \\
 &\quad + (r_1r_2 + r_1r_3 + r_1r_4 + r_2r_3 + r_2r_4 + r_3r_4)x^2 \\
 &\quad - (r_1r_2r_3 + r_1r_2r_4 + r_1r_3r_4 + r_2r_3r_4)x \\
 &\quad + r_1r_2r_3r_4
 \end{aligned}$$

This pattern extends indefinitely and is known as Viète’s theorem. In what follows we will change notation and use z as the unknown or indeterminate in polynomial expressions and x_1, x_2, \dots, x_n will represent the n roots of an n th degree polynomial equation. The coefficient expressions above are

called *elementary symmetric polynomials in n variables*, and are denoted:

$$\begin{aligned}\sigma_1 &= x_1 + \cdots + x_n \\ \sigma_2 &= x_1x_2 + x_1x_3 + \cdots + x_{n-1}x_n = \sum_{i<j} x_ix_j \\ \sigma_3 &= \sum_{i<j<k} x_ix_jx_k \\ &\vdots \\ \sigma_n &= x_1x_2 \cdots x_n\end{aligned}$$

Theorem 5.1 (Viète's Theorem). *Let $p(z)$ be an n^{th} degree, monic polynomial with roots x_1, x_2, \dots, x_n . Let $\sigma_1, \sigma_2, \dots, \sigma_n$ be the n elementary symmetric polynomials in the x_i , then*

$$p(z) = z^n - \sigma_1z^{n-1} + \sigma_2z^{n-2} - \cdots + (-1)^n\sigma_n.$$



5.5.2. René Descartes [1596–1650].

- Famous philosopher, best known for saying, “Cogito ergo sum.” (I think therefore I am.).
- Wrote *La géométrie*.
 - (1) Introduced the xy coordinate system which is now named for him, i.e. Cartesian coordinates.
 - (2) Borrowed the $+$, $-$, $\sqrt{\quad}$ symbols from German mathematicians.
 - (3) Used superscripts for exponentiation.
 - (4) Used lowercase letters such as a, b, c, \dots to

represent data such as polynomial coefficients, and letters such as x, y, z to represent unknowns.

Descartes' introduction of Cartesian coordinates marked a subtle yet profound break with past thinking on geometry. Before Descartes mathematicians tended to connect unknowns with lengths, squares of unknowns with square areas, and cubes of unknowns as cubic volumes. And since there is no fourth spatial dimension, ancient and medieval mathematicians often viewed expressions such as x^4 and higher powers of an unknown as rather meaningless. In fact, this is why Cardano only devoted a single chapter in his great tome *Ars Magna*, he saw the solution of the quartic as a curious oddity.

Descartes' great idea was to envision multiplication as scaling.

Symmetric Polynomials

Definition 6.1. A function f of n variables is *symmetric* if for every permutation $\sigma \in S_n$, $\sigma \bullet f = f$. That is, if

$$\begin{aligned}\sigma \bullet f &= f(x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)}) \\ &= f(x_1, x_2, \dots, x_n) \\ &= f.\end{aligned}$$

In simple terms, a multivariable function is symmetric if you can swap any two variables in its definition and you get the exact same function. Recall that we can *generate* S_n via certain subsets of permutations. For example, S_n can be generated by its transpositions (2-cycles). This is exactly analogous to how any shuffle of a deck of cards can be accomplished by a series of swaps

It turns out that when determining whether a function on n variables, say f , is symmetric, it suffices to just check whether f is symmetric under the action of a set of generators for S_n . This is due to the way we defined how a group acts on a set, which in this case is permutations acting on multi-variable functions. In short it is due to the associativity of the action.

Suppose you have two permutations, say $\sigma, \tau \in S_n$, and f is any function (not necessarily symmetric) then

$$\begin{aligned}\sigma \bullet (\tau \bullet f) &= \sigma \bullet f(x_{\tau(1)}, x_{\tau(2)}, \dots, x_{\tau(n)}) \\ &= f(x_{\sigma(\tau(1))}, x_{\sigma(\tau(2))}, \dots, x_{\sigma(\tau(n))}) \\ &= f(x_{(\sigma \circ \tau)(1)}, x_{(\sigma \circ \tau)(2)}, \dots, x_{(\sigma \circ \tau)(n)}) \\ &= (\sigma \circ \tau) \bullet f.\end{aligned}$$

Thus acting on the function f by τ and then σ is equivalent to acting on f by their composition: $\sigma \circ \tau$. In general, acting on a function by several

permutations is the same as first composing all the permutations together and then applying the result to f .

The above fact allows us to realize that if a subset of S_n , say $\{\sigma_1, \dots, \sigma_k\}$ generates S_n , then when determining whether f is symmetric, it suffices to simply test the effect of each $\sigma_i, \dots, \sigma_k$ on f . This is a great time saver because recall that the size of S_n grows factorially, that is, $|S_n| = n!$.

Example 6.2.

- (1) $x_1^2 + x_2^2 - x_1x_2$ is a symmetric polynomial in two variables. To check this we need to check that all permutations in $S_2 = \{1, (1\ 2)\}$ leave it fixed. Clearly the identity permutation leaves it unchanged, so we only need to check $(1\ 2)$:

$$\begin{aligned} (1\ 2) \bullet (x_1^2 + x_2^2 - x_1x_2) &= x_2^2 + x_1^2 - x_2x_1 \\ &= x_1^2 + x_2^2 - x_1x_2 \end{aligned}$$

- (2) $x_1^2 + x_2^2 + x_3^2$ is symmetric because $S_3 = \langle (1\ 2), (1\ 2\ 3) \rangle$ and because:

$$\begin{aligned} (1\ 2) \bullet (x_1^2 + x_2^2 + x_3^2) &= x_2^2 + x_1^2 + x_3^2 \\ &= x_1^2 + x_2^2 + x_3^2, \text{ and} \\ (1\ 2\ 3) \bullet (x_1^2 + x_2^2 + x_3^2) &= x_2^2 + x_3^2 + x_1^2 \\ &= x_1^2 + x_2^2 + x_3^2. \end{aligned}$$

- (3) $5x_1x_2 + 5x_1x_3 + 5x_2x_3$ is symmetric, but if we change the coefficients to say: $5x_1x_2 + x_1x_3 + x_2x_3$, then it is no longer symmetric because:

$$\begin{aligned} (1\ 2\ 3) \bullet (5x_1x_2 + x_1x_3 + x_2x_3) &= 5x_2x_3 + x_2x_1 + x_3x_2 \\ &= 5x_2x_3 + x_1x_2 + x_2x_3 \\ &\neq 5x_1x_2 + x_1x_3 + x_2x_3 \end{aligned}$$

- (4) $x_1^2x_2 + x_2^2x_3 + x_3^2x_1$ is *not* symmetric because:

$$\begin{aligned} (1\ 2\ 3) \bullet (x_1^2x_2 + x_2^2x_3 + x_3^2x_1) &= x_2^2x_3 + x_3^2x_1 + x_1^2x_2 \\ &= x_1^2x_2 + x_2^2x_3 + x_3^2x_1, \text{ but} \\ (1\ 2) \bullet (x_1^2x_2 + x_2^2x_3 + x_3^2x_1) &= x_2^2x_1 + x_1^2x_3 + x_3^2x_2 \\ &\neq x_1^2x_2 + x_2^2x_3 + x_3^2x_1 \end{aligned}$$

◇

6.1. Generators for S_n

S_3 is generated by any 2-cycle and any 3-cycle. Also any two distinct 2-cycles will also generate S_3 .

Proposition 6.1. *The cycles $(1\ 2)$ and $(1\ 2 \cdots n)$ generate S_n .*

Proof. First we note that $(1\ 2\ \cdots\ n)^{n-1} = (1\ 2\ \cdots\ n)^{-1}$. This allows us to generate the following transpositions:

$$\begin{aligned}(2\ 3) &= (1\ 2\ \cdots\ n)(1\ 2)(1\ 2\ \cdots\ n)^{-1} \\(3\ 4) &= (1\ 2\ \cdots\ n)(2\ 3)(1\ 2\ \cdots\ n)^{-1} \\(4\ 5) &= (1\ 2\ \cdots\ n)(3\ 4)(1\ 2\ \cdots\ n)^{-1} \\&\vdots \\(n-1\ n) &= (1\ 2\ \cdots\ n)(n-2\ n-1)(1\ 2\ \cdots\ n)^{-1}\end{aligned}$$

We showed previously that these transpositions can be used to generate any cycle with up to n elements, thus they can be used to generate any permutation because the permutations in S_n consist of products of disjoint cycles of length less than or equal to n . \square

The situation with S_4 is slightly more complicated, for example,

$$\begin{aligned}S_4 &= \langle (1\ 2), (1\ 2\ 3\ 4) \rangle, \text{ but,} \\S_4 &\neq \langle (1\ 3), (1\ 2\ 3\ 4) \rangle.\end{aligned}$$

You must be careful because *not* any 2-cycle and 4-cycle will generate S_4 . In the case of S_4 it turns out that any 4-cycle paired with any 2-cycle that appears inside the 4-cycle will generate. So for example since 1 and 2 are adjacent in $(1\ 2\ 3\ 4)$ the first pair above generates. However, since 1 and 3 are not adjacent in $(1\ 2\ 3\ 4)$ the second pair above does not generate S_4 . Why this is the case is an interesting story, but beyond the scope of this book.

6.2. Fundamental Theorem of Symmetric Polynomials

6.3. Generalizing the Solution Method

If we let x_1 and x_2 represent the two roots of the monic quadratic equation:

$$\begin{aligned}x^2 + bx + c &= 0 \\x^2 - (x_1 + x_2)x + x_1x_2 &= 0\end{aligned}$$

then the solutions can be expressed very succinctly:

$$\begin{aligned}x_1 &= \frac{1}{2}[(x_1 + x_2) + (x_1 - x_2)] = \frac{1}{2} \left[(x_1 + x_2) + \sqrt{(x_1 - x_2)^2} \right] \\x_2 &= \frac{1}{2}[(x_1 + x_2) - (x_1 - x_2)] = \frac{1}{2} \left[(x_1 + x_2) - \sqrt{(x_1 - x_2)^2} \right].\end{aligned}$$

Notice that the two rightmost expressions correspond exactly to the familiar

$$x_1, x_2 = \frac{-b \pm \sqrt{b^2 - 4c}}{2}.$$

because $-b = x_1 + x_2$, and

$$\begin{aligned} b^2 - 4c &= [-(x_1 + x_2)]^2 - 4x_1x_2 \\ &= x_1^2 + 2x_1x_2 + x_2^2 - 4x_1x_2 \\ &= x_1^2 - 2x_1x_2 + x_2^2 \\ &= (x_1 - x_2)^2. \end{aligned}$$

Building on work by Euler, Alexandre–Théophile Vandermonde realized that the above technique of writing each solution in terms of all the solutions could be extended to higher degree equations if we take into account the various n^{th} roots of unity.

In the $n = 2$ case the two, second roots of unity, or solutions of $z^2 = 1$ are just ± 1 , hence the minus sign preceding the square root in the expression for x_2 above. If you recall from the section on extending Cardano's formula, there are three solutions to $z^3 = 1$ which we denoted: $1, \omega, \omega^2$ because they formed a group of three elements under multiplication, generated by

$$\omega = -\frac{1}{2} + \frac{\sqrt{3}}{2}i.$$

The first root of the cubic equation with roots x_1, x_2 , and x_3

$$\begin{aligned} x^3 + ax^2 + bx + c &= 0 \\ x^3 - (x_1 + x_2 + x_3)x^2 + (x_1x_2 + x_1x_3 + x_2x_3)x - x_1x_2x_3 &= 0, \end{aligned}$$

can be written:

$$\begin{aligned} x_1 &= \frac{1}{3} [(x_1 + x_2 + x_3) + (x_1 + \omega x_2 + \omega^2 x_3) + (x_1 + \omega^2 x_2 + \omega x_3)] \\ &= \frac{1}{3} \left[(x_1 + x_2 + x_3) + \sqrt[3]{(x_1 + \omega x_2 + \omega^2 x_3)^3} + \sqrt[3]{(x_1 + \omega^2 x_2 + \omega x_3)^3} \right]. \end{aligned}$$

Vandermonde's insight was that permuting the three roots only resulted in two unique expressions. Notice that the expression $x_1 + x_2 + x_3$ is symmetric and thus fixed by all six permutations in S_3 , but the expressions under the cube roots are not fixed by S_3 . Thus to understand what the six possible permutations of the roots do to this expression for x_1 we just need to understand what they do to u and v where:

$$\begin{aligned} u &= (x_1 + \omega x_2 + \omega^2 x_3)^3 \\ v &= (x_1 + \omega^2 x_2 + \omega x_3)^3. \end{aligned}$$

$$\begin{aligned}
(1\ 2) \bullet u &= (1\ 2) \bullet (x_1 + \omega x_2 + \omega^2 x_3)^3 \\
&= (x_2 + \omega x_1 + \omega^2 x_3)^3 \\
&= \omega^6 (x_2 + \omega x_1 + \omega^2 x_3)^3 \\
&= [\omega^2 (x_2 + \omega x_1 + \omega^2 x_3)]^3 \\
&= [\omega^2 x_2 + \omega^3 x_1 + \omega^4 x_3]^3 \\
&= [\omega^2 x_2 + x_1 + \omega x_3]^3 \\
&= [x_1 + \omega^2 x_2 + \omega x_3]^3 \\
&= v.
\end{aligned}$$

Similarly

$$\begin{aligned}
(1\ 2) \bullet v &= (1\ 2) \bullet (x_1 + \omega^2 x_2 + \omega x_3)^3 \\
&= (x_2 + \omega^2 x_1 + \omega x_3)^3 \\
&= \omega^3 (x_2 + \omega^2 x_1 + \omega x_3)^3 \\
&= [\omega (x_2 + \omega^2 x_1 + \omega x_3)]^3 \\
&= [\omega x_2 + \omega^3 x_1 + \omega^2 x_3]^3 \\
&= [\omega x_2 + x_1 + \omega^2 x_3]^3 \\
&= [x_1 + \omega x_2 + \omega^2 x_3]^3 \\
&= u.
\end{aligned}$$

Now we examine the action of the other permutation needed to generate S_3 : $(1\ 2\ 3)$.

$$\begin{aligned}
(1\ 2\ 3) \bullet u &= (1\ 2\ 3) \bullet (x_1 + \omega x_2 + \omega^2 x_3)^3 \\
&= (x_2 + \omega x_3 + \omega^2 x_1)^3 \\
&= \omega^3 (x_2 + \omega x_3 + \omega^2 x_1)^3 \\
&= [\omega (x_2 + \omega x_3 + \omega^2 x_1)]^3 \\
&= [\omega x_2 + \omega^2 x_3 + \omega^3 x_1]^3 \\
&= [x_1 + \omega x_2 + \omega^2 x_3]^3 \\
&= u.
\end{aligned}$$

Bibliography

1. Michael Artin, *Algebra*, Pearson Education Inc., 2011.
2. Jorg Bewersdorff, *Galois theory for beginners*, American Mathematical Society, 2006.
3. John Derbyshire, *Unknown quantity: A real and imaginary history of algebra*, National Academies Press, 2006.
4. Harold M Edwards, *Galois theory*, Springer-Verlag, New York, 1997.
5. Benjamin Fine, *The fundamental theorem of algebra*, Springer, 1997.
6. Peter E Hydon, *Symmetry methods for differential equations: a beginner's guide*, Cambridge University Press, 2000.
7. Victor J Katz, *A history of mathematics*, 3rd ed., Addison-Wesley, 2009.
8. Victor J Katz and Bill Barton, *Stages in the history of algebra with implications for teaching*, Educational Studies in Mathematics **66** (2007), no. 2, 185–201.
9. R Bruce King, *Beyond the quartic equation*, Springer, 2009.
10. Serge Lang, *Algebra*, vol. 211, Springer-Verlag, New York, 2002.
11. Joseph J Rotman, *Galois theory*, Springer-Verlag, New York, 1998.
12. Jacques Sesiano, *An introduction to the history of algebra: Solving equations from mesopotamian times to the renaissance*, AMS, Providence, 2009.
13. Jean Pierre Tignol, *Galois' theory of algebraic equations*, World Scientific Publishing Company Incorporated, 2001.

Index

- abelian, 10
- act on, 4
- action, 7

- Cayley table, 9
- closure, 7
- composition, 5
- coset
 - left, 23

- Dihedral group, 9
- disjoint, 14

- equivalence relation, 23

- generator, 5
- generators, 7
- group
 - definition, 8
 - dihedral, 9
 - order, 9

- identity, 7
- inverse, 5
- inverses, 7

- Lagrange's theorem, 24

- non-abelian, 10

- partition, 22
- permutation, 12, 14

- relation, 22
 - equivalence, 23

- S_5 , 12
- S_n , 12
- subgroup, 10
- symmetry, 2

- table notation, 12