

Solving Polynomials: *The Roots of Modern Algebra*

Alexandre Bouchard-Côté (110228970)
McGill University
alexandre dot bouchard @ mail.mcgill.ca

November 23, 2004

1 Introduction

1.1 The Problem

A polynomial is a very unsophisticated mathematical object. It is simply a function which, given an input, transforms it using a finite sequence of scalings and additions¹. *Finding its roots* (or *solving it*) is a natural inversion problem: for any number, we want to find all the inputs (roots) that will produce this number.

The goal of this paper is to outline the history of this mathematical problem, and to show how it led mathematicians to study more and more abstract objects which made possible ultimately the emergence of abstract algebra.

For many cultures, $+$, $-$, $*$, the three basic tools required to construct polynomials, were available more than 3 millennia ago; thus it is easy to understand why the problem of finding roots of polynomials was also studied early in the history of many civilizations. It turns out, however, that solving this problem in its full generality is fairly intricate. It is a large enterprise that spans millennia of mathematical history and that progressed at an exponential pace and degree of abstraction. Historically, it is also the source of a very large tree of mathematics, which makes the study of the history of equation solving techniques very important.

The difficulty with this problem is twofold. First, it requires to understand and be able to manipulate much richer sets of numbers than the *natural numbers* $\mathbb{N} := \{1, 2, 3, \dots\}$. The first section will discuss how *formulae* involving only $+$, $*$, $-$, \div , $\sqrt[n]{}$ were discovered to solve polynomials of *degree* ≤ 4 (the definition of degree will be given soon). This kind of solution is called *solving in radicals*²

¹We are only interested in polynomial with one variable in this paper.

²Actually, *solving in radicals* refers to the situation in which every root of the polynomial can be obtained by a finite succession of algebraic operations ($+$, $-$, $*$, \div) and *root extractions* ($\sqrt[n]{}$), starting from the *base field* (a notion that will be defined later). Note however that \exists a formula in radical \Rightarrow can be solved in radicals (with the standard definition).

the polynomial. The focus will be made, however, on the second major difficulty, which is the shocking fact that *there are polynomials of degree ≥ 5 that cannot be solved in radicals* (the so-called insolubility of the quintic). The second part of this paper will discuss this in greater detail.

1.2 Some Terminology

For simplicity's sake, all the results will be described using modern terminology and notation. Let us define some basic concepts that will be very useful throughout the text. A *ring* is a generalization of the arithmetics with integers. It is a set (e.g., $\{\dots, -2, -1, 0, 1, 2, \dots\}$) equipped with two operations $(+, *)$ that satisfy some properties that enables us to do arithmetics in this set, parallel to the way arithmetics is done with the integers. These properties are just the axiomatization of the main admissible manipulations on integers one learnt at elementary school: associativity and existence of an identity for the $+, *$ operators ³, commutativity and invertibility of $+$ and distributivity of the two operators. A nontrivial example of a ring is the set of polynomials with usual addition and multiplication of polynomials. Indeed, if we redefine polynomials as the formal sums $a_n x^n + \dots + a_1 x + a_0$, where the *coefficients* a_n, \dots, a_0 belong to some commutative ring ⁴ R , one checks easily that if $p(x) := a_n x^n + \dots + a_1 x + a_0$, $q(x) := b_n x^n + \dots + b_1 x + b_0$, $c_n x^n + \dots + c_0 := p(x) + q(x)$, $d_n x^n + \dots + d_0 := p(x) * q(x)$, then setting ⁵:

$$c_i := a_i + b_i \tag{1}$$

$$d_i := \sum_{j,k:j+k=i} a_j * b_k \tag{2}$$

yields addition and multiplication operators with the required properties. A very important kind of rings is the *field*. We say that a ring is a field if it has the additional properties that it is commutative and that its $*$ operator is also invertible. Fields will play a central role in the proof of the insolubility of the quintic.

A *group* is a further generalization, in which we equip a set with only one operator, usually denoted $*$, that is required to be associative, invertible and to admit an identity. The $+$ operator of any ring, for instance, induces a group structure, and we can see then that we have much more control over the behavior of a ring compared to that of a group. One may ask how such a wild creature can help us in understanding polynomials over fields, which are very nicely behaved. The answer is a beautiful connection between the group and field theory called *Galois theory*.

³Some authors do not require the existence of an identity for the $*$ operation in the definition of a ring. We follow Serge Lang[2] and do include it.

⁴A commutative ring is a ring with the additional property that the operator $*$ also commutes.

⁵We can assume without loss of generality that the degrees, that is the index of the higher nonzero coefficient, match, otherwise put zeros for the a_i or b_i of the polynomial with lower degree until the degrees match.

2 First Part: Small Degrees

2.1 Linear and Quadratic Equations in Ancient Times

Solving linear equations is a fundamental requirement for all kinds of mathematics, as well as a practical problem that arise in many commercial, agricultural and geometrical situations. It is why all mathematical societies studied them very early in their history. For instance, the *Moscow Mathematical Papyrus* (copied by ancient Egyptians around 1650 BCE [1]), the *Jiuzhang Suanshu* (an ancient chinese text probably written around 1100 BCE [1]) and many Babylonian tablets (YBC 4652, VAT 8389, ... [1]) written about in the same time period describe examples of how these cultures solved linear equations. An important mathematical concept that was probably discovered while solving linear equations is that of fractional quantities. So excluding the notion of zero and negative numbers which were accepted much later, these ancient civilizations already started to realize the importance of the fields, although they were very far from being able to state it in such an abstract way.

Solving quadratic equations is a harder problem, but one that is still very geometric, so the study of this question also began early for many civilizations. All of the three civilizations cited above considered geometric versions of quadratic problems and solved them using geometric constructions [1]. In modern terminology, they considered only one positive solution of systems that admit at least one such solution. The greek mathematicians also used geometric solutions, but they realized the disturbing fact that their symbolism was not powerful enough to express exactly some of these solutions (the irrational solutions) [1]. This was a first step towards recognizing the importance of approximations in mathematics.

The first source of an almost modern solution of the quadratic comes from a text written by the hindu mathematician Brahmagupta (598-670) [1], [3]. In this text, the importance of negative numbers is recognized, as well as the convenient technique of using letters to represent unknown quantities. With this background, an equation using $+$, $-$, $*$, \div and a square root is given to find one solution of any quadratic equation allowing real solutions. Another hindu mathematician, Bhāskara (1114-1185), goes even farther and gives equations for the two solutions (again, when they are real) [3]. In other words, they solved in radicals the quadratic equation. The formulæ is of course:

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}, \quad (3)$$

where $x_{1,2}$ are the solutions of the polynomial $ax^2 + bx + c$ ⁶. It is not until the Renaissance, however, that the situations in which square roots of negative numbers were discussed.

⁶Of course, it was not expressed in modern notation, but rather in its textual equivalent.

2.2 The Development of the Cubic and Quartic during Italian Renaissance

The first extensive study of the solvability of the cubic known is due to Al-Khwarizmi (780-850). His solution was mostly geometric (using properties of conic sections)[5]. Rather, we will concentrate our attention on algebraic solutions that appeared during Renaissance in Italy.

In Italy, the first step towards a complete solution of the cubic was made by Scipione del Ferro (1465-1526), who solved equations of the form $x^3 + qx = q$ [5]. This is actually all that is required since given a cubic $x^3 - bx^2 + cy - d$, using the transformation $\phi(x) := x + b/3$, we get the above simplified form. Unfortunately, without the Hindu's knowledge of negative numbers, Ferro could not have been able to generalize this solution to all cubics[3]. Even more unfortunate is the fact that he kept his discovery secret (A common practice among the fifteenth century scholars: they would keep their discoveries secret in order to win public competitions). More than a decade later, Niccolò Fontana (1499-1557) (nicknamed "Tartaglia", the stutterer) rediscovered the method as well as two other methods needed to cover all the cases of cubics that must be considered when one does not know how to manipulate negative numbers.⁷ Tartaglia also kept this secret, and it is only many years later that another Italian mathematician, Gerolamo Cardano (1501-1576), finally published the method in his book *Ars Magna, sive de regulis algebraicis*. In modern notation, the three methods published in *Ars Magna* can be unified into the equation:

$$x_{1,2} = v - u - \frac{a}{3} \tag{4}$$

$$u := \sqrt[3]{\frac{q}{2} \pm \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} \tag{5}$$

$$v := \frac{p}{3u} \tag{6}$$

where $x_{1,2}$ is a subset of the roots of $x^3 + ax^2 + bx + c$.

An interesting observation about this method is that it sometimes requires to use *complex numbers* (that is, number of the form $a + b\sqrt{-1}$) in an intermediate step in order to get *real* roots. This was first noticed by Cardan but expressed in notation for the first time by Rafael Bombelli. Again, the development of a larger set of numbers was motivated by the problem of solving equations.

Shortly after the discovery of a solution for the cubic, Lodovic Ferrari (1522-1565), Cardano's servant, developed a method to reduce quartic equations into cubic, again using only field operations and root extractions. This method was also published in *Ars Magna*. This series of successes encouraged many mathematicians to improve the existing methods and develop new ones for equations of degree ≥ 5 . We will see what fate awaited these attempts in the second part of this paper.

⁷He made this discovery while doing a public competition against Fiore, Ferro's pupil, to whom Ferro told his secret before his death.

Noteworthy improvements and further contributions made to the methods described in this section include the work of Viète, Harriot, Tschirnhaus, Euler, Bézout, Descartes, Leibniz, Lagrange and Bring[3]. Descartes developed what is now called Descartes' Sign Rule, which gives upper bounds for the number of positive and negative real roots of a polynomial. Harriot made the important observation that if x_0, x_1, x_2 are the roots of a cubic, then the cubic is $(x - x_0)(x - x_1)(x - x_2)$. Using this, Leibniz reconstructed the cubic from its roots, producing the first purely algebraic proof of the cubic formula.

3 Second Part: Fifth Degree

3.1 Precursors

After the publication of *Ars Magna*, almost three hundred years passed before the first finished proof of the insolubility of the quintic appeared. Many great mathematicians (e.g., Euler or Bézout) attacked this problem inconclusively. Tschirnhaus claimed that he found a solution in radicals, but Leibniz later exhibited a flaw in his argumentation. During this period of time, the understanding of polynomials improved in several other ways though. For instance, Albert Girard conjectured in 1629 that a polynomial of degree n has n roots (this statement was formalized and proven by Gauss 150 years later), and Newton gave in 1669 the first iterative method for numerical approximation of roots. This method is called the Newton-Raphson method nowadays.

A crucial step was made by Lagrange (1736-1813), who recognized the importance of the *permutations*, the primitive ancestor of the modern notion of group[5]. Lagrange's work was built on the precursor work of Waring and Vandermonde. It is a student of Lagrange, Ruffini, who published a first tentative proof of the insolubility of the quintic. This proof, however, contained many gaps and attracted little attention.

Cauchy (1789-1857) defined the notion of "group of substitutions", and found many properties of this object. He was actually very close to the concept of group that we are familiar[5]. Helped by these new developments, the gifted young mathematician Niels Henrik Abel (1802-1829) succeeded in proving that a solution of the quintic by radicals is impossible (his proof is a particular case of Galois theory).

Two other important contributions to the question of solubility came from Carl Friedrich Gauss (1777-1855). First, he provided a complete solution in radicals of the *cyclotomic equations* $x^n - 1 = 0$. Second, he proved a version of what is now called the *Fundamental Theorem of Algebra*, which states that a polynomial of degree n with real coefficients can always be factored as a product of n linear terms $(x - x_i)$, where $x_i \in \mathbb{C}$.

Thus, at the untimely death of Abel (he was 27), it was known that some polynomials of degree 5 cannot be solved in radicals, but that some others (the cyclotomic equations, for instance) can. The next problem was to find a way of deciding whether a given polynomial can be solved in radicals or not.

3.2 The Work of Galois

The definitive answer to this question came from the notes of the french mathematician Évariste Galois (1811-1832), published posthumously thanks to Joseph Liouville, Galois' friend Auguste Chevalier, and Galois' brother[4]. It is quite remarkable that Galois managed to do such a breakthrough in his short and tumultuous existence (Galois also died ⁸ when he was very young: he was 20).

Let us now have a closer look to the role of these “group of substitutions” and to the theory of Galois[2], [4]. We begin with two definitions:

Definition 1. *Let F be a field. An automorphism of fields is a map $\varphi : F \rightarrow F$ ⁹ satisfying:*

$$\varphi(a + b) = \varphi(a) + \varphi(b) \quad \forall a, b \in F \quad (7)$$

$$\varphi(a * b) = \varphi(a) * \varphi(b) \quad \forall a, b \in F \quad (8)$$

In other words, the two above axioms require the map to preserve the structure induced by the two operations of the field. One easily checks that for any field, the set of all admissible automorphisms forms a group.

Definition 2. *Let K, L be fields and suppose that $L \supseteq K$. An automorphism ψ of L is a K -automorphism of L if*

$$\psi(k) = k \quad \forall k \in K \quad (9)$$

Note that $L \supseteq K$ is usually called a *field extension* ¹⁰. At the core of Galois theory is the observation that the set of K -automorphisms also forms a group. The theorem is actually quite easy to prove:

Theorem 1. *Let $L \supseteq K$ be a field extension. The set of K -automorphisms of L forms a group under composition ¹¹.*

Proof. Let ψ, φ be K -automorphisms of L . We have clearly that $\varphi \circ \psi$ is an automorphism. If $k \in K$, then $\varphi \circ \psi(k) = \varphi(k) = k$, so $\varphi \circ \psi$ is actually a K -automorphism. The identity map and φ^{-1} are also K -automorphisms ($k = \varphi^{-1}\varphi(k) = \varphi^{-1}(k)$). Composition of maps is always associative, so the theorem is proven. \square

This group is also called the *Galois group* ¹² of the extension. Galois investigated in more details the situation in which the field $K := \mathbb{Q}$ and L is taken

⁸He was killed in a duel, probably because of his relationship with Stéphanie-Felicie Poterin. As many other aspects of Galois' life, this duel was dramatized, and it has been theorized that it was actually an excuse used by political enemies. But this theory has no firm ground, and Galois himself wrote in his notes “I die the victim of an infamous coquette”[4].

⁹A function from F to F .

¹⁰The original work of Galois took a less general point of view, and it is the modern version of Galois theory that will be presented.

¹¹The composition of two functions is denoted $(g \circ f)$ and is defined to be the function such that $(g \circ f)(x) := g(f(x))$ for all x in the domain of f .

¹²Other authors use this terminology when an extra condition is satisfied: that the cardinality of the K -automorphism group be the same as the degree of the extension.

to be the smallest field containing K and the set of roots of some polynomial f with no multiple root and with coefficients in K . He discovered that in this case there are many correspondences between the structure of the Galois group and the fields:

Theorem 2 (Fundamental Theorem of Galois Theory). *Let $L \supseteq K$ be an extension as described above. Then there is a bijection between the fields contained in L (the subfields of L) and containing K and the groups contained in the Galois group of the extension (the subgroups of the Galois group).*

This theorem actually contains many other correspondences between the two structures, but the notions in group and field theory that are necessary in order to state and prove completely this theorem are outside the scope of this paper.

The last step was to develop a criterion *on the Galois group* that determines whether the polynomial f can be solved in radical. Groups that satisfy this criterion are called *solvable groups* for obvious reasons. Finally, using the tools and results on finite groups developed by Cauchy, Galois could finally characterize entirely whether a given polynomial can be solved in radicals or not.

4 Subsequent Developments

After the publication of Galois' paper, the modern definition of group and field gradually emerged and much effort was put for a better understanding of these abstract structures[5]. The approach of considering automorphisms of an object relative to a sub-object proved to be a fruitful one, and was extended in scope by Felix Klein's famous *Erlanger Program*[4]. Galois theory was extended and clarified by Camille Jordan and Ludwig Sylow[5] at the end of the nineteenth century and took the form it has now in most graduate algebra textbooks.

The investigations regarding polynomials solving did not stopped after the work of Galois though. Numerical and algebraic methods continue to be improved, and it is now very easy to solve polynomials of arbitrary degree with an arbitrary precision using freely available software packages. However, after the work of Galois, algebraists were no longer focusing exclusively on the problem of solving equations, and algebra evolved into the rich and abstract field we now know.

References

- [1] Victor Katz. *A History of Mathematics*.
- [2] Serge Lang. *Algebra*.
- [3] University of St Andrews Scotland. The mactutor history of mathematics archive. <http://www.gap-system.org/history/>.
- [4] Ian Stewart. *Galois Theory*.
- [5] B.L. Waerden. *A History of Algebra*.