The Dolciani Mathematical Expositions

NUMBER THIRTY-FIVE

Uncommon Mathematical Excursions

Polynomia and Related Realms

Dan Kalman
American University

Published and Distributed by
The Mathematical Association of America

4

Solving Polynomial Equations

Finding the roots of polynomials is a problem with a long history in mathematics, and one that has led to a tremendous volume of mathematical knowledge. In elementary mathematics we first encounter quadratic polynomials, and learn to find roots using factoring or the quadratic formula. The outcome depends on what types of numbers you are willing to accept as answers. If you accept complex numbers, then the quadratic formula gives a complete solution. Every quadratic has two roots (sometimes equal), and complete factorization is always possible. In contrast, if you are only interested in real solutions, some quadratic equations are unsolvable. Here again the quadratic formula gives us a complete answer. It tells us whether roots exist, and finds them when they do.

It is natural to turn next to cubic equations. We can find solutions for quadratic equations, like $x^2 - 3x + 5 = 0$. The cubic equation $x^3 + 7x^2 - 2x + 3 = 0$ doesn't look much more complicated. Can we solve it? Is there some way to find the solutions as combinations of cube roots and square roots, perhaps, extending what works for quadratics? And beyond cubics, what about quartics, quintics, and higher degree polynomials?

The attempt to solve cubics dates back at least to the tenth century efforts of Arab mathematicians. The earliest work had a strong geometric component, both in how the equations were understood and in methods of solution. Omar Khayyam discovered how to find roots of cubics as intersections of parabolas [3]. An algebraic solution for one form of cubic was developed in the beginning of the sixteenth century by Cardano, among others. Complete solutions for general cubic and quartic equations followed soon after. The cubic and quartic solutions have a similar flavor. To solve a cubic, you must first solve a related quadratic, and the solutions of the cubic involve square and cube roots. Similarly, the solution of a quartic depends on solving a related cubic.

We will see these solutions in this chapter. By today's standards, the algebra they involve does not seem very forbidding. The solution to cubic equations can be followed by students at the level of precalculus. Of course, following the logic of an algebraic argument is far easier than inventing the argument in the first place. And the first discovered solutions to cubics and quartics involve clever algebraic tricks. Moreover, they came before the development of modern algebraic notation, which makes the solutions easier to follow. Still, it is tempting to wonder why the solutions took so long to be found, and to imagine that similarly simple methods for solving higher equations might yet await discovery.

Interestingly, the pattern of results for polynomials up to the quartic do not extend to higher degrees. More subtle patterns lurk behind the scenes that account for both the solutions of the low degree cases and the obstructions to finding similar solutions for higher degree. A key to understanding this deeper structure involves the interchangeability of the roots of a given polynomial. As we will see, this idea arises quite naturally at a very elementary level. We have already seen it emerge in the previous chapter's study of the relationships between roots and coefficients. The significance of the idea of interchangeability goes far beyond these simple beginnings, however. It ultimately unlocks the deep mysteries of polynomial equations in a mathematically breathtaking development. The end of the chapter will offer a glimpse of this subject.

The quest to solve polynomial equations has played a central role in the evolution of modern mathematics. Along the way, mathematicians have encountered important methodological and philosophical questions. What does it mean to *find* a root? If we can approximate a root to any specified accuracy, can we then claim to have found it? Can we even claim to know that the root exists? These questions go beyond solving equations to the very nature of numbers themselves, and historically led to ever richer number systems, including negatives, irrationals, and imaginary numbers.

In broad outlines, here is what we know today. Permitting complex numbers to be used, both for coefficients and roots, every polynomial can be expressed as a product of linear factors. Thus, for any polynomial of degree n, there exist n complex roots. When n is 4 or less, exact solutions can be found using known methods, which are analogous to the quadratic formula and can be applied to any cubic or any quartic. For polynomials of degree 5 or higher, it is known that the roots do not always exist in an exact form, or at least, not as an algebraic expression involving arithmetic operations and radicals (i.e., square roots, cube roots, fourth roots, and so on). This shows that a general method for solving quintic and higher degree equations cannot exist.

These conclusions are valid when working with the complex numbers. But polynomials often arise where coefficients are known to be real, or rational, or integers, or where we are interested only in real, rational, or integer roots. This leads to quite a few different possible problems and conclusions. For example, if we have integer coefficients and desire only rational roots, an exact solution method for arbitrary degree is known.

In this chapter we will look at a variety of ideas connected with solving polynomial equations, including:

- Existence questions for rational, real, and complex solutions,
- Algebraic solutions of cubic and quartic equations,
- Lagrange's analysis and permutations of roots,
- Insolubility of the general quintic.

4.1 Existence Questions

Before talking about how to solve a polynomial equation, we should clarify what it means for a solution to exist. If it is possible to display an integer or rational number that satisfies a given equation, then the question of existence of course evaporates. But what about equations with irrational or complex solutions?

Today, with the complex numbers a familiar and well understood part of mathematics, we have an advantage over mathematicians of earlier eras. Our current ideas about number systems, indeed about what numbers *are*, evolved slowly over centuries. At the time of Cardano, negative numbers were sometimes used, but were not universally accepted as legitimate numbers. Complex numbers were contemplated only by the most advanced thinkers, and even they did not know what to make of the idea. Little wonder if there was some ambiguity about what it means for an equation to have a solution.

To make this more concrete, let us consider a few examples. According to the modern viewpoint, the equation $x^2 = 1$ has two solutions. A mathematician who denies the existence of negative numbers would say it has a single solution. Cardano would have identified two solutions, calling one *false*, meaning not really a legitimate number. The equation $x^2 + 1 = 0$ would be considered to have no solutions whatever.

Similar issues arise with irrational numbers. As an example, consider the equation $3x^5 - 15x + 5 = 0$. It has no rational roots (which can be verified by methods to follow), but by trial and error (as well as more sophisticated methods), we can find approximate solutions. To illustrate one approach, let $p(x) = 3x^5 - 15x + 5$, so $p(0) = 5$ and $p(1) = -7$. This suggests that a root should occur somewhere between 0 and 1. We might guess that it falls at $x = .5$. This guess is incorrect, but with a true solution somewhere between 0 and 1, $x = .5$ differs from a correct answer by at most .5.

Next compute $p(.1)$, $p(.2)$, $p(.3)$, and so on, finding $p(.3) = .50729$ and $p(.4) = -.96928$. This tells us the root is between .3 and .4, and if we adopt .35 as an estimate, we know we will be off by at most .05. Continuing in this way, we can get as close to the root as we wish.

But does this establish the existence of an exact solution? If there is no specific number we can identify that is a solution, if we can only point to better and better approximations, how do we know that there *is* a solution?

The modern viewpoint rests on a sophisticated understanding of both the algebraic and geometric properties of our number systems. We have a concept of continuity, according to which the real numbers form a continuum with no holes or gaps. This idea can be formulated rigorously and, as long as you are willing to accept the assumptions inherent in the formulation, results such as the intermediate value theorem give precise conditions for the existence of solutions to equations.

Historically, solving polynomial equations meant finding exact algebraic expressions for the roots. In the next section the original methods for solving cubic and quartic equations in this sense will be presented. Generalizing them to higher degrees involves the concept of *solvability by radicals*. That means solutions that are expressible in terms of radicals (square roots, cube roots, etc.) and arithmetic operations. The methods for cubics and quartics show they are all solvable by radicals. For higher degree polynomials, some equations are solvable by radicals, and some are not.

Is a solution by radicals preferable to one by successive approximation? Aesthetically, an exact algebraic representation of the solution is highly appealing. But the more aesthetic approach is not always the more practical. Working on a computer or calculator with limited precision, successive approximation is typically more efficient and accurate than direct implementation of solutions by radicals, at least for degree 3 or 4. Even if there is an exact algebraic expression for a root, it most likely will involve radicals that are themselves only

approximately computable. And on the aesthetic side solution by radicals is not the only option, for we saw exact solutions to cubic equations elegantly expressed in terms of curly roots (page 26).

As these considerations show, solvability of polynomial equations is a many-faceted subject. We can approach it in a variety of ways, conceptually, procedurally, and philosophically. With that in mind, let us look at some important solvability results.

Rational roots. Suppose that a polynomial has coefficients that are rational numbers. What can we deduce about the roots? First, any such equation can be transformed into one with integer coefficients. Expressing each of the original coefficients as a fraction, we can find a common multiple A of all the denominators. Multiplying the entire equation by A then eliminates all the denominators, leaving a polynomial with integer coefficients.

Fair enough — let us restrict our attention to polynomials with integer coefficients. Solutions may be rational or irrational. But if they are rational, they have to take a special form, as specified in the following result.

The Rational Roots Theorem. *Let $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$ have integer coefficients, and for integers r and s let r/s be a rational root in lowest terms. Then r is a divisor of a_0 and s is a divisor of a_n.*

Proof. Since r/s is a root, $p(r/s) = 0$, hence $s^n p(r/s) = 0$. This gives us the equation

$$a_n r^n + a_{n-1}r^{n-1}s + a_{n-2}r^{n-2}s^2 + \cdots + a_1 r s^{n-1} + a_0 s^n = 0,$$

which can be written in the form

$$a_n r^n + a_{n-1}r^{n-1}s + a_{n-2}r^{n-2}s^2 + \cdots + a_1 r s^{n-1} = -a_0 s^n.$$

Because r is a divisor of the left side of this equation, it is a divisor of $a_0 s^n$. But r/s is in lowest terms means that r and s have no common divisors. Thus r is a divisor of a_0. A similar argument shows that s is a divisor of a_n. ■

Ideally, we would like our equations to have rational solutions. The preceding theorem provides a tool for investigating whether or not they do. Rational roots, if they exist, have to lie among the finitely many fractions whose numerators divide into a_0 and whose denominators divide into a_n. If we compute $p(x)$ for each of these fractions, we will find all the rational roots, or demonstrate that there are none.

As an example, consider $x^3 + 6x - 20 = 0$. The divisors of a_3 are just 1 and -1. The divisors of a_0 are $1, 2, 4, 5, 10, 20$, and their negatives. Therefore, if the equation has rational roots, they must lie among the numbers $\pm 1, \pm 2, \pm 4, \pm 5, \pm 10$, and ± 20. Direct calculation reveals that 2 is a rational root, and it is the only rational root.

This can be verified from another direction. Knowing that $x = 2$ is a root implies that $x - 2$ is a factor of the polynomial, and it is a short step to $x^3 + 6x - 20 = (x - 2)(x^2 + 2x + 10)$. The quadratic formula shows that the roots of the quadratic factor are $-1 \pm 3i$. This both verifies that 2 is the only rational root and illustrates the important idea of using a known root to reduce the degree of an equation.

The example also illustrates a corollary of the Rational Roots Theorem: if a *monic* polynomial has integer coefficients, any rational roots are actually integers. As the example shows, when the leading coefficient of a polynomial is 1, the denominator of any rational

root must be ± 1. In more advanced treatments of polynomials, roots of monic polynomials are referred to as *algebraic integers*, and play a special role in the development of the theory.

Real roots. As many polynomials with integer coefficients have irrational roots, it is natural next to extend our focus to real numbers. Whereas the rational case is characterized by discrete methods (integer factorization, enumerating a finite set of possible solutions), the real case has a decidedly continuous flavor. For example, using the continuity of the real line and of polynomial functions, we can invoke the intermediate value theorem to show that if $p(x)$ has opposite signs at a and at b then it has a root somewhere in between. This is an existential result. It does not tell us how to find a root, only that one must exist. But that is an important consideration in locating roots through successive approximation.

It is also customary to use methods from calculus in the real case. A function that is increasing in an interval, for example, can have at most one root in that interval. Here, an *increasing function* $f(x)$ obeys the maxim *the larger the* x, *the larger the* $f(x)$. If $f(x) = 0$ for a specific x, then it must be greater than 0 for all larger values of x, so there can be at most one root.

A familiar result from calculus tells us that a function is increasing if it has a positive derivative. Let us apply this idea to the earlier example, with $f(x) = x^3 + 6x - 20$. The derivative, $f'(x) = 3x^2 + 6$, is positive for all real x. Therefore, $f(x)$ is increasing over the entire real line, and can have at most one real root. This shows that $x = 2$ is not just the only rational root, but also the only *real* root.

Similar analyses lead to a number of conclusions for polynomials with real coefficients:

- An odd degree polynomial always has at least one real root.
- An even degree polynomial has an even number of real roots (where a double root is counted twice).
- For odd n, the equation $x^n - a_0 = 0$ always has exactly one real root.
- For even n, the equation $x^n - a_0 = 0$ has two real roots if a_0 is positive, the unique root $x = 0$ if $a_0 = 0$, and no real roots if a_0 is negative.

The last two of these give us familiar properties about nth roots.

Complex roots. The complex numbers are the ultimate setting for polynomial equations in a particular sense. Observe the progression of number systems. Simple (linear) equations with natural number coefficients can have solutions that are negative, or noninteger rationals. So we extend the number system to include negatives and fractions, arriving at the rational numbers. Polynomial equations with rational coefficients can have irrational solutions, so we extend the number system again, this time to the reals. But there are equations with real coefficients that do not have real solutions. This leads us to the complex numbers. Now the process stops, because every polynomial with complex coefficients has roots that are complex numbers. At least in terms of formulating and solving polynomial equations, the complex numbers form a closed system. Technically, the complex numbers are said to be *algebraically closed.*

This result is known as the *Fundamental Theorem of Algebra*. Here is a formal statement.

The Fundamental Theorem of Algebra. *A non-constant polynomial with complex coefficients has a root in the complex numbers.*

The existence of a root also implies a factorization. If a polynomial $p(x)$ has a complex root r_1, then we can factor it in the form $p(x) = (x - r_1)q(x)$. Since $q(x)$ will also have complex coefficients, it too must have a root (unless it is a constant), and it will have degree one less than p. By applying the Fundamental Theorem in a chain of factorizations, we will eventually reduce $p(x)$ to a factored form $A(x - r_1)(x - r_2)\cdots(x - r_n)$. Thus, the Fundamental Theorem shows that every polynomial equation with complex coefficients can be completely factored, and so has all of its roots among the complex numbers.

As in the real case, this is an existence result. And like the real case, it reflects not only the algebraic properties of the complex numbers, but also the important idea of continuity. This is a topological or analytic aspect of the complex numbers. Many proofs of the Fundamental Theorem are known, and they all involve analysis or topology in some way.

Earlier we described the complex numbers as the ultimate number system for polynomial equations. It really would be more accurate to say *an* ultimate system. In other contexts polynomial equations arise with coefficients that are outside the integer-rational-real-complex progression. For example, polynomial equations with matrix coefficients arise naturally in some areas of mathematics. Another example is provided by modular arithmetic. It makes perfect sense to discuss the solvability of polynomial equations whose coefficients are integers modulo five, and that leads to a completely different ultimate number system.

Even so, it remains a valid observation that in familiar number systems, namely those encompassing the integers, the complex number system is the one that has all the answers. It provides the proper perspective for understanding polynomials, even in cases that seem to involve only real numbers. For example, the algebraic solution of cubic equations requires complex numbers, even when all the coefficients and roots are real. Indeed, it has been argued that solving cubic equations provided a primary motivation for the historical development of complex numbers [99].

We close this section with an illustration of complex numbers intruding on a real issue. It is a theorem about real factorizations of real polynomials, but the proof uses complex numbers.

The Fundamental Theorem of Algebra, Real Case. *Every polynomial with real coefficients can be expressed as a product of linear and quadratic factors with real coefficients.*

This is a corollary of the Fundamental Theorem of Algebra, and has a simple proof using complex conjugation. (See Appendix A at the website for this book [87] for an explanation of conjugation and other aspects of complex numbers.) In particular, if $p(x)$ has real coefficients, then for any complex number z, $p(\overline{z}) = \overline{p(z)}$. This implies that r is a root of p if and only if $\overline{r}$ is also a root, because

$$p(\overline{r}) = \overline{p(r)} = \overline{0} = 0.$$

Over the complex numbers, a polynomial with real coefficients factors as $A(x - r_1) \cdot (x - r_2)\cdots(x - r_n)$ with roots r_j that are complex numbers. Some of these may be real. For those that are not, each factor $(x - r)$ combines with a corresponding factor $(x - \overline{r})$ to form a quadratic factor with real coefficients. This establishes the real case of the Fundamental Theorem of Algebra.

4.2 Historic Solution of Cubics and Quartics

In mathematics, the first solution of a problem is often far from the best solution. Later developments can lead to solutions that are shorter and more transparent than the one first discovered. However, this is not the case for solving cubic equations (in terms of radicals). The algebraic solution first published by Cardano in 1545 is as simple as any of the many variations and alternatives that have appeared since. An example will illustrate how this solution works. We will consider an example that Cardano used.

In $x^3 = 20 - 6x$, substitute $u + v$ for x to obtain

$$(u + v)^3 = 20 - 6(u + v). \tag{1}$$

The cube on the left is $u^3 + 3u^2v + 3uv^2 + v^3$. When we group the u^3 and v^3 terms together and factor $3uv$ from the remaining terms, we have

$$u^3 + v^3 + 3uv(u + v) = 20 - 6(u + v).$$

Now we separate this one equation into two:

$$\begin{aligned} u^3 + v^3 &= 20 \\ 3uv &= -6. \end{aligned}$$

Clearly, if we can find numbers u and v that satisfy these two equations, then $u + v$ will satisfy (1), and hence the original equation.

At first glance, this appears similar to our earlier attempts to solve a cubic using elementary symmetric functions (see page 50). But this time the system can be solved. In fact, a solution follows from our understanding of the elementary symmetric functions for a quadratic. To make this clearer, divide the second equation by 3, and then cube both sides. The system then takes the form

$$\begin{aligned} u^3 + v^3 &= 20 \\ u^3v^3 &= -8. \end{aligned}$$

Thinking of u^3 and v^3 as the unknowns, the system specifies their sum and product — exactly the requirements for solving a quadratic equation as discussed on page 50. So u^3 and v^3 must be the solutions of the quadratic equation

$$t^2 - 20t - 8 = 0.$$

Solving, we find that u^3 and v^3 must be $10 \pm 6\sqrt{3}$. Extracting cube roots gives us values for u and v, and leads to

$$x = \sqrt[3]{10 + 6\sqrt{3}} + \sqrt[3]{10 - 6\sqrt{3}} \tag{2}$$

as a solution to the original cubic.

In principle, once a root r is known, the factor $x - r$ can be divided out of the original cubic. The remaining roots can then be found using the quadratic formula. This is not a very practical way to proceed in general, and is quite unwieldy in the example at hand. An alternative is to recognize that extracting a cube root is itself the solving of a polynomial

Sidebar 4.1

The History of the Cubic

The solution of the cubic equation is accompanied by an intriguing human interest story, centered on a dispute between Cardano and Tartaglia. Scipione del Ferro made the first progress toward a general algebraic method for cubics sometime between 1500 and 1515. His method applied to equations of the form $x^3 + cx = d$, with both c and d assumed to be positive. Today we consider a single standard cubic $x^3 + cx + d = 0$, but in del Ferro's time negative numbers were not accepted and the terms of the equation had to be positioned so as to avoid negative coefficients. To del Ferro, $x^3 + 3x = 5$ and $x^3 + 5 = 3x$ were distinct problems, and called for distinct methods of solution.

Although del Ferro's discovery was a major breakthrough, he did not make it known within the mathematical community. Rather, he kept his method secret so that he would be able to solve equations that stumped mathematical rivals. But before his death in 1526 he passed his method on to two colleagues, Antonio Maria Fiore and Annibale della Nave. In 1535, Fiore challenged Tartaglia to a problem solving contest. Inspired by the problems in the contest, Tartaglia discovered his own method for solving cubic equations. Following del Ferro's example, he kept his method secret.

Now Cardano enters the picture. He was working on a book at the time, and wanted to include Tartaglia's method. Tartaglia eventually agreed to reveal his method to Cardano, but wishing to publish the method himself, Tartaglia made Cardano promise not to reveal it. This was in 1539. True to his word, Cardano left Tartaglia's method out of his book.

But over the next several years, he worked on the cubic himself, working out methods for all the various cases (what we would regard as possible combinations of signs of coefficients). He also learned of del Ferro's original method and obtained the

equation, and so leads potentially to three solutions. So once we know a value for u^3 we should be able to obtain three values for u, corresponding to three roots of the original cubic equation. We will follow up on this idea below.

Curiously, the solution (2) is a complicated way of writing $x = 2$, though there is no simple algebraic method to verify that fact. The simplest verification is to show directly that 2 is a root of the original equation, and, in fact, the only real root. Here we are observing the tip of an iceberg. There is much more to say about identities involving radical expressions and how they relate to roots of polynomials, but it would be too great a digression to explore this topic now. Suffice it to say that it is sometimes difficult to recognize when an algebraic expression involving radicals has a simpler equivalent form.

Does this single example convince you that Cardano's method will solve *any* cubic equation? One issue is the special form of the example, with no x^2 term. But we saw in Chapter 3 that any cubic can be brought to this reduced form by making a substitution for the variable (see page 51). Another objection could arise from the fact that u^3 and v^3 are supposed to be obtained as roots of a quadratic. What if the quadratic does not have any real roots? Then we can express u^3 and v^3 as complex numbers, but how do we find u and v?

The History of the Cubic (cont.)

Cardano

Tartaglia

details from della Nave. Since del Ferro's discovery preceded Tartaglia's, and since Cardano had extended the method to many new cases, Cardano considered himself no longer bound by his vow to keep Tartaglia's method secret. In 1545 he published the *Ars Magna* (or *The Great Art*), which included a comprehensive account of methods for all the possible cubics, as well as a method for the quartic discovered by his student, Lodovico Ferrari.

Tartaglia was outraged, and thought Cardano had betrayed his trust. To add insult to injury, in a later mathematical contest he fell to defeat at the hands of Ferrari, Cardano's protege. Intrigue, betrayal, triumph, defeat — did you ever imagine that the history of algebra could be as dramatic as a soap opera?

Sidebar 4.1 (cont.)

Algebra has developed far beyond what was known to Cardano and his contemporaries. Today we have better notation and symbolic methods, and can draw on the full power of the complex number system. And in large measure, these tools were developed for the investigation of polynomial equations.

Our modern vantage point augments the algebraic derivation of Cardano's solution with a complete understanding of the solution of the general cubic equation. We now know that a single valid choice for u and v can be used to completely factor the original cubic; that this will involve cube roots of complex numbers exactly when the cubic has only real roots; and how to construct those complex cube roots. These points will be considered in detail before we proceed to quartic equations.

As a first step of the analysis of the cubic, let us retrace Cardano's method for the general equation

$$x^3 = bx + c.$$

We introduce the substitution $x = u + v$ and derive the equations

$$\begin{aligned} u^3 + v^3 &= c \\ uv &= b/3. \end{aligned} \tag{3}$$

Cubing the second equation produces

$$\begin{aligned} u^3 + v^3 &= c \\ u^3 v^3 &= b^3/27, \end{aligned}$$

a system of equations in u^3 and v^3.

If we let $s = u^3$ and $t = v^3$, then the system becomes

$$\begin{aligned} s + t &= c \\ st &= b^3/27. \end{aligned}$$

Now it is apparent that s and t are roots of a quadratic. We extract cube roots of s and t to determine u and v, and hence a solution $u + v$ to the cubic. However, there are actually three complex cube roots of s. Let u_0 be a particular cube root of s. Then the other cube roots can be expressed in terms of u_0 and $\omega = (-1 + i\sqrt{3})/2$, a primitive cube root of unity. There are exactly three complex numbers with cube equal to s, namely $u_0, \omega u_0$, and $\omega^2 u_0$.

We do not have equal freedom to choose v among the cube roots of t. Referring to (3), once u has been selected, we must take $v = b/(3u)$, so the system of equations leads us to exactly three choices for the pair (u, v). That is, Cardano's method gives us three expressions for a root of the cubic equation: $u_0 + b/(3u_0)$, $\omega u_0 + b/(3\omega u_0)$, and $\omega^2 u_0 + b/(3\omega^2 u_0)$. If we let $u = u_0$ and $v = b/(3u_0)$, then since $\omega^3 = 1$, our three expressions for roots of the cubic become $u + v, \omega u + \omega^2 v$, and $\omega^2 u + \omega v$.

There is no reason to assume that the three expressions are distinct, and in fact they need not be. But we can show that they give a complete factorization for the original cubic. The elementary symmetric functions again are useful. We consider the product

$$q(x) = (x - u - v)(x - \omega u - \omega^2 v)(x - \omega^2 u - \omega v). \tag{4}$$

This can be put in descending form with coefficients given by the elementary symmetric functions of the three roots:

$$\begin{aligned} a_0 &= -(u + v)(\omega u + \omega^2 v)(\omega^2 u + \omega v) \\ a_1 &= (u + v)(\omega u + \omega^2 v) + (u + v)(\omega^2 u + \omega v) + (\omega u + \omega^2 v)(\omega^2 u + \omega v) \\ a_2 &= -(u + v) - (\omega u + \omega^2 v) - (\omega^2 u + \omega v). \end{aligned}$$

To simplify these algebraically, it helps to observe that $1 + \omega + \omega^2 = 0$, which follows from $0 = (\omega^3 - 1) = (\omega - 1)(\omega^2 + \omega + 1)$. The alternative form $\omega + \omega^2 = -1$ is also handy. Using these identities, the system simplifies to

$$\begin{aligned} a_0 &= -u^3 - v^3 \\ a_1 &= -3(uv) \\ a_2 &= 0. \end{aligned}$$

Since we know that $u^3 + v^3 = c$ and $3uv = b$, it follows that $a_0 = -c, a_1 = -b$, and the descending form of $q(x)$ is thus $x^3 - bx - c$. This demonstrates that the three roots of $q(x)$ are all the roots of the original equation $x^3 = bx + c$.

It is important to stress here the dependence of the factorization just derived on an arbitrary choice of three possible values of u. For each choice of u_0, we get three roots (and a factorization) of $x^3 - bx - c$. This might seem to suggest a total of nine possible roots, but we know that there must be exactly three. Therefore, each choice of u_0 leads to the same three roots, appearing in different orders. Earlier, we saw that the coefficients of a polynomial depend symmetrically on the roots, since permuting the roots has no effect on the coefficients. But now we see how the idea of permuting the roots arises naturally from an ambiguity in defining a complex cube root. This same phenomenon occurs any time a root of a polynomial is expressed in terms of radicals. Each radical can be interpreted as one of n different complex nth roots, any choice of which leads to a permutation of the roots produced by any other choice.

Expressing the three roots as in (4) is of interest for another reason. From it we can deduce that all three roots of the cubic are real precisely when s and t are complex (that is, are not real). Because s and t arise as roots of the quadratic equation

$$x^2 - cx + b^3/27 = 0,$$

the discriminant $D = c^2 - 4b^3/27$ tells us whether or not they are real.

If $D \leq 0$, then s and t are complex conjugates (and possibly real and equal). In this case, u and v are also complex conjugates. To see this, note that $v^3 = t = \bar{s} = \overline{u^3} = \overline{u}^3$. Thus, v and $\overline{u}$ are both complex cube roots of t, so are equal or differ by a factor of ω or ω^2. That is, $v = \overline{u}$ or $v = \omega\overline{u}$ or $v = \omega^2\overline{u}$. But we also know that uv is real. Thus, $v = \overline{u}$, showing that one of the roots, $u + v$, is real. Next, observe that ω and ω^2 are complex conjugates. Thus each of the remaining roots, $\omega u + \omega^2 v$ and $\omega^2 u + \omega v$, is again the sum of a complex number and its conjugate, and so real. Therefore, when $D \leq 0$ all three roots are real.

On the other hand, suppose $D > 0$. Then s and t are real and unequal. Choosing a real value for u, and thus ensuring that v is also real, we obtain one real root, $r_1 = u + v$. A second root, $r_2 = \omega^2 u + \omega v$, can be rewritten

$$\omega^2 u + \omega u - \omega u + \omega v = -u + \omega(v - u).$$

If this is real, then $v = u$ so $s = u^3 = v^3 = t$, contrary to assumption. Thus r_2 is not real, and by a similar argument, neither is the final root, $r_3 = \omega u + \omega^2 v$.

Putting all this together, we see that a cubic has three real solutions if and only if $D \leq 0$, and three distinct real solutions when $D < 0$. This latter case occurs precisely when s and t fail to be real, and then we can reach the real solutions only by operating with complex numbers.

It is natural to ask whether this is an artifact of Cardano's method, or an intrinsic property of cubic equations. Could there be an alternative method for solving cubics that avoids complex numbers, at least when the roots are all real? Motivated partly by this question, we will look at alternate solutions of the cubic in the next section.

Now let us turn to the quartic. As mentioned in Sidebar 4.1, Cardano's student Ferrari is credited with first discovering a method for solving quartic equations. In modern form, Ferrari's method proceeds as follows.

Begin with a reduced equation (no x^3 term) written in the form

$$x^4 = ax^2 + bx + c.$$

Now introduce a new variable y and add $2x^2y + y^2$ to both sides of the equation, producing

$$x^4 + 2x^2y + y^2 = ax^2 + bx + c + 2x^2y + y^2.$$

On the left we have a perfect square $(x^2 + y)^2$. Call the expression on the right $q(x)$, and arrange the terms in decreasing powers of x to get

$$q(x) = (a + 2y)x^2 + bx + (c + y^2).$$

Ferrari's key insight was that this might also be a perfect square. If so, with a perfect square on each side, our quartic equation will be solvable. Can we make this happen? That is, can we find a number y for which $q(x)$ is a perfect square? That will occur if the discriminant $b^2 - 4(a + 2y)(c + y^2)$ equals 0.

An example will clarify this. Let $a = 7$, $b = 10$, and $c = 16$. Then $q(x) = (7 + 2y)x^2 + 10x + (16 + y^2)$, which we want to be a perfect square. Try $y = 1$, so $q(x) = 9x^2 + 10x + 17$. Is that a perfect square? If so, then it has just one root. But the quadratic formula gives $(-10 \pm i\sqrt{512})/18$ so there are two roots. This shows that $q(x)$ is not a perfect square for $y = 1$. But it also tells us what must be done. We need to change y so that the quadratic formula involves $\sqrt{0}$ instead of $\sqrt{512}$.

For any choice of y, what appears inside the squareroot is the discriminant of q, given in our example by $10^2 - 4(7 + 2y)(16 + y^2)$. We want this to be zero. And it will be zero if $y = -3$. (This can be verified by substitution. Never mind, for now, how it was discovered.) And with $y = -3$, $q(x)$ becomes $x^2 + 10x + 25$, which is evidently a perfect square.

The example reveals the pattern for the general case: $q(x) = (a+2y)x^2 + bx + (c + y^2)$ will be a perfect square when its discriminant is zero. Thus, we want y to satisfy

$$b^2 - 4(a + 2y)(c + y^2) = 0,$$

or in descending form,

$$8y^3 + 4ay^2 + 8cy - b^2 + 4ac = 0.$$

This is a cubic equation in y, so Cardano's method will give us at least one real solution. Using it, the equation

$$x^4 + 2x^2y + y^2 = (a + 2y)x^2 + bx + (c + y^2)$$

has a perfect square on each side, and thus has the form $w^2 = z^2$. That means we can simplify to the equation $w = \pm z$. This gives us two equations, each of which is quadratic in x, and so leads us to four roots.

Let us apply this method to one of Cardano's examples: $x^4 = 12x - 3$, in which $a = 0$, $b = 12$, and $c = -3$. Introducing y, add $2x^2y + y^2$ to both sides of the original equation, producing

$$x^4 + 2x^2y + y^2 = 12x - 3 + 2x^2y + y^2.$$

Rearranging both sides then gives us

$$(x^2 + y)^2 = 2yx^2 + 12x + (y^2 - 3), \tag{5}$$

with a perfect square on the left. We make the right-hand side a perfect square as well by requiring that

$$144 - 8y(y^2 - 3) = 0.$$

In standard form, this becomes

$$y^3 - 3y - 18 = 0,$$

and $y = 3$ is a solution. (Here, we were fortunate to find a cubic with so simple a root, no doubt the result of careful planning by Cardano. However, even if that were not the case, at least one y would be produced by Cardano's method for cubic equations.)

Substituting in (5), we find

$$(x^2 + 3)^2 = 6x^2 + 12x + 6.$$

As anticipated the right-hand side is a perfect square, so the equation becomes

$$(x^2 + 3)^2 = 6(x + 1)^2,$$

and that leads to

$$x^2 + 3 = \pm(x + 1)\sqrt{6}.$$

Thus we arrive at two quadratic equations

$$x^2 + \sqrt{6}x + 3 + \sqrt{6} = 0$$
$$x^2 - \sqrt{6}x + 3 - \sqrt{6} = 0.$$

Solving them gives us the four roots of the original quartic, namely

$$\frac{\sqrt{6} \pm \sqrt{4\sqrt{6} - 6}}{2} \quad \text{and} \quad \frac{-\sqrt{6} \pm i\sqrt{4\sqrt{6} + 6}}{2}.$$

Ferrari's method leads to four solutions for any quartic. The cubic equation in y will always have a real solution, which leads directly to a pair of quadratic equations in x. Looked at another way, once we have a value of y, we get a factorization of the original quartic into two quadratic factors. In the example we can go from

$$(x^2 + 3)^2 = 6(x + 1)^2$$

to

$$(x^2 + 3)^2 - 6(x + 1)^2 = 0,$$

and hence to

$$[(x^2 + 3) + \sqrt{6}(x + 1)][(x^2 + 3) - \sqrt{6}(x + 1)] = 0.$$

By revealing all the roots in this way, Ferrari's solution for the quartic is more direct than Cardano's solution of the cubic. This runs contrary to the expectation that the algebra should get progressively more complicated as the degree increases. Of course, you have to solve a cubic as part of the solution of a quartic, so the latter is not really simpler than the former. But that point aside, the direct factorization of the quartic does seem simpler than the factorization based on complex cube roots for the cubic. Similarly, the analysis

of root permutations works out more simply for four roots than for three, confirming the impression that the quartic is somehow simpler than the cubic.

On first studying the methods of Cardano and Ferrari, one is struck by the lack of any unifying strategy. Both methods depend on clever algebraic tricks, but the tricks are completely unrelated. This suggests that other solutions might exist, based on different algebraic tricks, which has proven to be the case. Since the time of Cardano and Ferrari, many alternative approaches for solving cubic and quartic equations have been discovered. In the next section we will consider some, and observe that there does appear to be something inescapable in these solutions. In all of the cubic solutions, a common feature appears, and the same is true for the quartic. Why does this happen? Why do algebraic tricks that work so effectively for cubic and quartic equations fail for higher order equations? These are precisely the questions that inspired Lagrange to focus on symmetric functions of roots, and so to lay the foundations for key ideas in modern algebra, including group theory and Galois theory. At the end of this chapter, we will review some of Lagrange's ideas. For now, we proceed to alternate solutions of cubic and quartic equations.

4.3 Alternate Solutions for Cubics

For future reference, each solution in this section will be labeled with a descriptive phrase or with the name of its discoverer and the approximate date of discovery.

Viète, 1591. Begin with the equation $x^3 + ax + b = 0$. Introduce the substitution $x = y - a/(3y)$. After expanding and simplifying, the original equation becomes $y^6 + by^3 - a^3/27 = 0$. This is a quadratic in y^3, and so tells us that

$$y^3 = \frac{-b \pm \sqrt{b^2 + 4a^3/27}}{2}.$$

Given these values for y^3, extracting a complex cube root provides a possible value of y. For each choice of y, we obtain a root for the original cubic from $x = y - a/(3y)$.

Although the algebra follows a different path for this solution than for Cardano's, the equation for y is the same as the equation in Cardano's solution for u, and both approaches lead to equivalent equations for x.

Euler, 1770. Euler's solution is essentially identical to Cardano's except that he starts with the substitution $x = \sqrt[3]{s} + \sqrt[3]{t}$ rather than Cardano's $x = u + v$. The algebra is the same. However, Euler's notation is worth mentioning here because it so similar to what he used in the case of the quartic.

Cayley, 1877. Cayley proposed a modified version of Cardano's solution for the equation $x^3 = ax + b$, substituting $u^2v + uv^2$ for x, rather than $u + v$. Cayley's substitution leads to two equations in u and v, but this time only u^3 and v^3 appear. The system is

$$\begin{aligned} u^3 + v^3 &= 3b/a \\ u^3v^3 &= a/3. \end{aligned}$$

The right-hand side of the second equation is a bit simpler than in Cardano's approach, where in place of $a/3$ we would find $(a/3)^3$. There is another difference between the two

methods. In Cardano's approach, once we define u as one of the possible cube roots of u^3, we are left with a single choice of v. But in Cayley's solution, we can choose v as any cube root of v^3, independent of the choice of u. Nevertheless, Cayley's solution to $x^3 = ax + b$ is the same as making a change of variables $x = y\sqrt[3]{a/3}$ and then applying Cardano's method.

Equality of Two Cubes. The idea for this method is to recast the equation $x^3 = ax + b$ in the form $A(x + m)^3 = B(x + n)^3$. Then we can find a solution by taking a cube root of each side, leading to

$$x = \frac{B^{1/3}n - A^{1/3}m}{A^{1/3} - B^{1/3}}.$$

Thus, one root of the original cubic will be obtained as soon as we know A, B, m, and n.

To find them, express $A(x+m)^3 - B(x+n)^3$ in descending form and equate coefficients with $x^3 - ax - b$. That gives the system

$$\begin{aligned} A - B &= 1 \\ 3Am - 3Bn &= 0 \\ 3Am^2 - 3Bn^2 &= -a \\ Am^3 - Bn^3 &= -b. \end{aligned}$$

Using the first two equations, A and B can be found in terms of m and n. Substituting in the third and fourth equations leads eventually to

$$\begin{aligned} m + n &= \frac{3b}{a} \\ mn &= \frac{a}{3}. \end{aligned}$$

Once again we are led to equations for the sum and product of two unknowns, and hence to two solutions of a quadratic equation. This system is identical to the one that appears in Cayley's approach if we set $m = u^3$ and $n = v^3$. Further investigation proves that the algebraic expressions for the roots revealed by this approach are equivalent to those obtained in Cardano's approach.

An interesting story accompanies this solution of the cubic, as retold in the autobiography of mathematician Mark Kac. See Sidebar 4.2.

Two Cube Completions. Here is another approach using the idea of a perfect cube, due to Frink [55]. Rewrite the equation $x^3 + ax = b$ as

$$\frac{x^3}{8} + \frac{3x^3}{8} + \frac{ax}{2} = \frac{b}{2}.$$

Sidebar 4.2

Mark Kac and the Cubic

Mark Kac (1914–1984) was a leading mathematician who made pioneering contributions to the modern development of mathematical probability, and in particular its applications to statistical physics. His work in the latter is commemorated, in part, by the Feynman-Kac path integral, named after Kac and Richard Feynman. With Paul Erdős he introduced probabilistic methods in number theory. He was the author of several books, including popular and philosophical works in collaboration with figures such as Stanislaw Ulam and Gian-Carlo Rota. Kac won many awards, among them the Birkhoff prize (awarded jointly by the AMS and SIAM) and on two separate occasions the MAA's Chauvenet prize. One of the Chauvenet prizes was for the paper *Can One Hear the Shape of a Drum?* [73]. Among mathematicians, he may be best remembered in connection with that paper. Kac also was invited to deliver quite a few prestigious lectures, including SIAM's John von Neumann Lecture, the MAA's Hedrick Lectures, and the AMS's Gibbs Lecture.

Kac grew up in Poland. In his autobiography [74], he describes how an early fascination with cubic equations led him to become a mathematician. When he was 15, he recalls,

> ...I became obsessed with the problem of solving cubic equations. Now, I knew the answer, which Cardan had published in 1545, but what I could not find was a derivation that satisfied my need for understanding. When I announced that I was going to write my own derivation, my father offered me a reward of five Polish zlotys (a large sum and no doubt the measure of his scepticism). I spent the days, and some of the nights, of that summer feverishly filling reams of paper with formulas. Never have I worked harder. Well, one morning, there it was — Cardan's formula on the page. My father paid up without a word, and that fall

In preparation for completing the cube, set $3x^3/8 + ax/2$ equal to $3xy^2/2$, which means that $y^2 = x^2/4 + a/3$. Then we obtain

$$\frac{x^3}{8} + \frac{3xy^2}{2} = \frac{b}{2}. \tag{6}$$

We will complete the cube on the left side of this equation in two ways. Let $R = 3x^2y/4 + y^3$. Then adding R to both sides produces

$$\left(\frac{x}{2} + y\right)^3 = \frac{b}{2} + R,$$

while subtracting R from both sides gives

$$\left(\frac{x}{2} - y\right)^3 = \frac{b}{2} - R.$$

Mark Kac and the Cubic (cont.)

Kac

my mathematics teacher submitted the manuscript to "Mlody Matematyk" (The Young Mathematician).... When my gymnasium principal, Mr Rusiecki, heard that I was to study engineering, he said, "No, you must study mathematics; you have clearly a gift for it".

The solution Kac discovered was the one presented here under the heading, *Equality of Two cubes.*

As Kac continues his story, he followed the advice of his principal, and as a result escaped sure destruction in World War II. Had it not been for the opportunities he found for study abroad in mathematics, and in particular, the fortuitous timing of his travels, he would undoubtedly have perished alongside his parents and brother at the hands of the Nazis.

Sidebar 4.2 (cont.)

Now we take cube roots to obtain the equations

$$\frac{x}{2} + y = \sqrt[3]{\frac{b}{2} + R} \tag{7}$$

$$\frac{x}{2} - y = \sqrt[3]{\frac{b}{2} - R}, \tag{8}$$

whose sum gives x in terms of R:

$$x = \sqrt[3]{\frac{b}{2} + R} + \sqrt[3]{\frac{b}{2} - R}. \tag{9}$$

To complete the solution, we need to express R in terms of the coefficients a and b. So multiply (7) and (8) to obtain

$$\frac{x^2}{4} - y^2 = \sqrt[3]{\frac{b^2}{4} - R^2},$$

and observe from the definition of y that the left side is $-a/3$. Therefore

$$\frac{-a^3}{27} = \frac{b^2}{4} - R^2$$

and

$$R = \sqrt{\frac{b^2}{4} + \frac{a^3}{27}}.$$

Substitution in (9) thus gives the solution

$$x = \sqrt[3]{\frac{b}{2} + \sqrt{\frac{b^2}{4} + \frac{a^3}{27}}} + \sqrt[3]{\frac{b}{2} - \sqrt{\frac{b^2}{4} + \frac{a^3}{27}}}.$$

This derivation seems to avoid the quadratic equation that arose in every other solution. However, when we obtain the value of R by taking a square root, that is equivalent to solving a quadratic equation, though a simple one. Moreover, it is essentially what we would encounter if we solved the quadratics in the earlier approaches by completing the square. And the solutions found here are in exactly the same form as those obtained using Cardano's solution.

Factorization Identities. The quadratic formula can be understood as a consequence of the identity $r^2 - s^2 = (r - s)(r + s)$. Any monic quadratic can be put into the form $(x - h)^2 - s^2$ by completing the square, and then the identity provides a decomposition into linear factors. This idea can be extended to cubic equations.

On the right-hand side of the quadratic identity, we can think of the $+$ and $-$ as representing square roots of unity, that is, ± 1. For cubics, there is a similar identity that involves the three cube roots of unity, $1, \omega$, and ω^2, where $\omega = (-1 + i\sqrt{3})/2$. It can be expressed in various forms, including

$$r^3 + s^3 + t^3 - 3rst = (r + s + t)(r + \omega s + \omega^2 t)(r + \omega^2 s + \omega t)$$

and

$$\omega(r^3 + s^3 + t^3) - 3\omega^2 rst = (\omega r + s + t)(r + \omega s + t)(r + s + \omega t).$$

These are equivalent, and it is a matter of taste which form is more memorable, or more closely resembles the identity for the quadratic case. While either can be used to solve a cubic equation, we will look at the first version.

First replace r, s, and t with x, $-u$, and $-v$:

$$x^3 - u^3 - v^3 - 3xuv = (x - u - v)(x - \omega u - \omega^2 v)(x - \omega^2 u - \omega v).$$

This can be used to factor any cubic in the form $x^3 + ax + b$ by making $-(u^3 + v^3) = b$ and $-3uv = a$. Then the roots are $u + v$, $\omega u + \omega^2 v$, and $\omega^2 u + \omega v$. Though different in concept, this method is algebraically identical to Cardano's approach.

Change of Variables in Symmetric Equations. In Chapter 3, we considered the idea of solving a cubic by direct inversion of the equations expressing the coefficients in terms of

the roots. If the cubic is $x^3 + ax^2 + bx + c = 0$ and the roots are r, s, and t, the equations are

$$\begin{aligned} r + s + t &= -a \\ rs + rt + st &= b \\ rst &= -c, \end{aligned} \tag{10}$$

and all we have to do is find r, s, and t given a, b, and c.

This can be carried out if we make a change of variables using

$$\begin{aligned} r &= u + v + w \\ s &= u + \omega v + \omega^2 w \\ t &= u + \omega^2 v + \omega w, \end{aligned} \tag{11}$$

with ω a primitive cube root of unity as before. This will transform (10) into a system in u, v, and w. If we can determine values for u, v, and w, we will then be able to find r, s, and t.

So substitute the expressions on the right-hand side of (11) for r, s, and t in (10). The resulting system can be simplified, using the identities $\omega^3 = 1$, $1 + \omega + \omega^2 = 0$, and their variants, to obtain

$$\begin{aligned} 3u &= -a \\ 3u^2 - 3vw &= b \\ u^3 + v^3 + w^3 - 3uvw &= -c. \end{aligned} \tag{12}$$

Use the first equation to eliminate u from the other two. What remains are equations involving vw and $v^3 + w^3$, essentially the same as we have seen in many of the other approaches. They can be put into a form that specifies values for the sum and product of v^3 and w^3, thus giving v and w as cube roots of the solutions of a quadratic equation. With u, v, and w thus determined, the roots r, s, and t are given by system (11).

How does this method compare to Cardano's? In Cardano's approach, we must take the preliminary step of eliminating the quadratic term of the cubic. Imposing the same assumption here amounts to making $u = 0$. But then system (12) reduces to

$$\begin{aligned} vw &= -b/3 \\ v^3 + w^3 &= -c, \end{aligned}$$

the very equations that arise in Cardano's method.

Matrix Algebra. The preceding solution is intriguing. With a simple change of variables it permits direct inversion of the elementary symmetric functions. This approach has the advantage that it is conceptually transparent. Anyone might think of it. But the change of variables used seems to be unmotivated and mysterious. Even if you recognize the close links between the algebraic combinations we saw earlier and the definitions of u, v, and w, it is still not obvious that the change of variables will make system (10) solvable. It is legitimate to ask how anyone would find this change of variables.

One answer is provided by matrix algebra. The idea uses the fact that roots of polynomials can be realized as eigenvalues of matrices. Thus, a given matrix leads to both a set of roots (or eigenvalues) and the corresponding polynomial. This provides another alternate method for solving cubics and quartics. At its heart, this method corresponds to making a linear change of variables in the equations for the elementary symmetric functions. But in this setting the change of variables arises very naturally. These ideas are developed in detail in [92].

There are still other solutions for cubics. In Chapter 2 we considered an approach using not radicals, but the similarly defined curly root function. There are also geometric constructions (including the one mentioned earlier, attributed to Omar Khayyam), an approach using differential equations, and even a solution based on origami, or paper folding. But solutions that can be reduced to algebraic manipulation with radicals inevitably turn out to be equivalent to Cardano's method.

This is not just a trivial consequence of the fact that all of the methods have to produce the same roots, because those roots might conceivably appear in different forms. This is dramatically illustrated by Cardano's example $x^3 + 6x - 20 = 0$, where the real root 2 appears as $\sqrt[3]{10 + 6\sqrt{3}} + \sqrt[3]{10 - 6\sqrt{3}}$. Every other method we have seen for solving a cubic produces this root in exactly the same outlandish form. Apparently, there is only one way to skin a cubic, at least as far as algebraic manipulation is concerned. We turn next to solutions of the quartic, where we will observe a similar phenomenon.

4.4 Alternate Solutions for Quartics

Quartics can be solved by a variety of methods, several of which are analogs of cubic methods. As before, each method will be identified with a discoverer or a short descriptive phrase. Unless noted otherwise, we always assume that the equation to be solved is $p(x) = x^4 + ax^2 + bx + c = 0$.

Descartes, 1637. Factor the quartic into a product of two quadratics. The fact that $p(x)$ has no cubic term implies that the linear coefficients of the quadratic factors must be equal in magnitude and opposite in sign. Therefore, the factorization we seek has the form $(x^2 + ux + v)(x^2 - ux + w)$. Expand this into descending form and equate the coefficients with those of p. That produces the equations

$$\begin{aligned} v + w - u^2 &= a \\ wu - vu &= b \\ vw &= c. \end{aligned}$$

The first two equations can be rewritten as

$$\begin{aligned} v + w &= a + u^2 \\ v - w &= -b/u, \end{aligned}$$

leading to expressions for v and w in terms of u. When they are substituted into the remaining equation, we obtain the equation in u alone

$$u^6 + 2au^4 + (a^2 - 4c)u^2 - b^2 = 0. \tag{13}$$

This is a cubic equation in u^2, and so is solvable.

Although this looks different from Ferrari's solution of the quartic, the cubic equations that arise in each method are closely related. Substituting $2y - a$ for u^2 in (13) (and adjusting for the opposite signs of the coefficients in the two approaches) reproduces Ferrari's cubic equation in y exactly. Though Descartes' method uses a different approach, it leads to an equivalent cubic equation, and factors the quartic into the same product of quadratic factors.

Euler, 1770. Euler's solution to the quartic is an extension of his solution of the cubic. He begins by assuming $x = \sqrt{r} + \sqrt{s} + \sqrt{t}$. He squares this, simplifies, squares a second time, and eventually obtains

$$x^4 - 2(r + s + t)x^2 - 8\sqrt{rst}x + (r + s + t)^2 - 4(rs + rt + st) = 0.$$

Now his idea is to make this match the original quartic by choosing r, s, and t appropriately. That way, we will know that $x = \sqrt{r} + \sqrt{s} + \sqrt{t}$ is one root of the original quartic. Equate the coefficients of the two equations to derive

$$\begin{aligned} -2(r + s + t) &= a \\ -8\sqrt{rst} &= b \\ (r + s + t)^2 - 4(rs + rt + st) &= c. \end{aligned}$$

In these equations Euler recognized the elementary symmetric functions for three variables. With $r + s + t = \sigma_1$, $rs + rt + st = \sigma_2$, and $rst = \sigma_3$, the system can be rewritten

$$\begin{aligned} \sigma_1 &= -\frac{a}{2} \\ \sigma_3 &= \frac{b^2}{64} \\ \sigma_1^2 - 4\sigma_2 &= c. \end{aligned}$$

To simplify further, use the first equation to eliminate σ_1 from the last equation, yielding

$$\begin{aligned} \sigma_1 &= -\frac{a}{2} \\ \sigma_3 &= \frac{b^2}{64} \\ \sigma_2 &= \frac{a^2 - 4c}{16}. \end{aligned}$$

These equations specify the elementary symmetric functions of r, s, and t in terms of the known coefficients a, b, and c. This is exactly analogous to the situation where we specify the sum and product of two variables. And just as the two-variable case tells us that the unknown variables are roots of a particular quadratic, so we can conclude here that r, s, and t are the roots of the cubic polynomial having $-\sigma_1$, σ_2, and $-\sigma_3$ as coefficients. That is, they are the roots of

$$g(x) = x^3 + \left(\frac{a}{2}\right)x^2 + \left(\frac{a^2 - 4c}{16}\right)x - \frac{b^2}{64}.$$

Using known methods, we can solve this cubic to find r, s, and t, and thus derive a root of the quartic we began with.

Here we have a third algebraic approach, different in form from the other two. Once again, the analysis depends on solving a cubic. And once again, it is essentially the same cubic. In fact, with the substitution $x = u^2/4$ Euler's cubic becomes Descartes' cubic.

Change of Variables in Symmetric Equations. As for the cubic, a quartic equation can be solved by direct inversion of the symmetric functions, after making a suitable change of variables. If we label the roots of $x^4 + ax^3 + bx^2 + cx + d$ as q, r, s, and t, then the original system is

$$\begin{aligned} q + r + s + t &= -a \\ qr + qs + qt + rs + rt + st &= b \\ qrs + qrt + qst + rst &= -c \\ qrst &= d. \end{aligned}$$

Substitute $q = u+v+w+z, r = u+v-w-z, s = u-v+w-z$, and $t = u-v-w+z$. The first equation then shows that $u = -a/4$, and we can use that to eliminate u from the remaining equations. After simplification, they become

$$\begin{aligned} v^2 + w^2 + z^2 &= \frac{3a^2 - 8b}{16} \\ 8vwz + a(v^2 + w^2 + z^2) &= \frac{a^3 - 16c}{16} \\ (v^4+w^4+z^4) - 2(v^2w^2+v^2z^2+w^2z^2) - \frac{a^2}{8}(v^2+w^2+z^2) - 2a(vwz) &= d - \frac{a^4}{256}. \end{aligned}$$

This system can be simplified by using symmetric functions of v^2, w^2, and z^2: $\sigma_1 = v^2+w^2+z^2$, $\sigma_2 = v^2w^2+v^2z^2+w^2z^2$, and $\sigma_3 = (vwz)^2$. Also, $v^4+w^4+z^4 = \sigma_1^2-2\sigma_2$. Using these identities, the system becomes

$$\begin{aligned} \sigma_1 &= \frac{3a^2 - 8b}{16} \\ 8\sqrt{\sigma_3} + a\sigma_1 &= \frac{a^3 - 16c}{16} \\ \sigma_1^2 - 4\sigma_2 - \frac{a^2}{8}\sigma_1 - 2a\sqrt{\sigma_3} &= d - \frac{a^4}{256}. \end{aligned}$$

Although this appears to be complicated, it is a triangular system. The first equation tells us the value of σ_1. Substituting it into the second equation determines the value of σ_3. Then, substituting both σ_1 and σ_3 into the third equation establishes the value of σ_2. Thus, we get each σ in terms of a, b, c, and d. From this point the solution proceeds as in Euler's analysis. With known expressions for each σ_j, it follows that v^2, w^2, and z^2 are roots of a cubic. The solutions of that cubic lead to v, w, and z, and we already know that $u = -a/4$. At last, the values of u, v, w, and z lead to corresponding values for the roots, q, r, s, and t of the original quartic.

To relate this approach to the ones already considered, we impose the assumption $a = 0$. Then the system of equations is readily solved, revealing

$$\sigma_1 = \frac{-b}{2}$$
$$\sigma_2 = \frac{b^2 - 4d}{16}$$
$$\sigma_3 = \frac{c^2}{64}.$$

Thus, v^2, w^2, and z^2 are roots of

$$x^3 + \left(\frac{b}{2}\right)x^2 + \left(\frac{b^2 - 4d}{16}\right)x - \frac{c^2}{64}.$$

This is the same as the polynomial $g(x)$ that appeared in Euler's solution, except that here the coefficients b, c, and d, respectively, play the roles of the coefficients a, b, and c from the earlier analysis.

Matrix Algebra. The remarks following the change of variables solution for the cubic also apply to the quartic. It is intriguing that a simple change of variables permits the solution of a quartic by directly inverting the elementary symmetric functions. But the lack of a rationale for defining the new variables detracts from the appeal of this approach. As mentioned before, matrix algebra provides an alternative viewpoint for solving cubics, and leads in a natural way to an appropriate change of variables. This matrix algebra viewpoint works in the same way for the quartic as well.

Looking back over all of the solutions to both cubics and quartics, several patterns stand out. For all of the cubic solutions it is necessary to solve an auxiliary quadratic equation. Likewise, each solution of the quartic depends on finding the roots of an auxiliary cubic equation. Moreover, all of the cubic solutions depend on the roots of the same or closely related quadratics. Likewise, all of the quartic solutions involve variations of the same cubic. These observations suggest that beneath the surface appearance of fortuitous algebraic gimmicks, there is an underlying structure dictating the form of the solution. Such an idea occurred to Lagrange, who analyzed this structure. Lagrange wanted to understand the successes in solving cubics and quartics, as well as the failure of all efforts to solve quintics. Although he did not settle the question of the quintic, he laid the foundation that permitted others to do so. In the next section we will see how Lagrange's ideas of symmetry and root permutation can be used to solve quartic equations.

4.5 Solving Quartics with Symmetry

As detailed in Sidebar 4.3, Lagrange and Vandermonde analyzed the significance of symmetry in solutions of cubic and quartic equations, working independently and essentially simultaneously around 1770. Reportedly, although they started with the same approach, Lagrange went further and his work had the greater impact on later developments. The following discussion is based loosely on his approach.

We have seen many methods for solving cubic and quartic equations. They all have in common the solution of an extra equation. We set out to find the roots of one polynomial

$p(x)$ (let's call it the *original* polynomial), but along the way we discover we have to find the roots of an *auxiliary*[1] polynomial $A(x)$. When p is a cubic, A is a quadratic; when p is a quartic, A is a cubic. And in both cases, the coefficients of the auxiliary polynomial are polynomial combinations of the coefficients of the original polynomial.

Lagrange set out to understand how and why an auxiliary polynomial arises. He used ideas related to the elementary symmetric functions that we considered in Chapter 3. For each polynomial the coefficients are symmetric functions of its roots. In addition, the coefficients of A depend on the coefficients of p, while the roots of p depend on the roots of A.

To illuminate the power of Lagrange's ideas, we will depart from his path. Where he analyzed a known solution of the quartic, we will pretend to know no solution. Then, using Lagrange's insights, we will see how to manufacture an auxiliary polynomial, based not on an algebraic trick, but on the understanding of symmetry.

To begin, we introduce notation for the coefficients and the roots of the original and auxiliary polynomials. Let $p(x) = x^4 + ax^3 + bx^2 + cx + d$ with roots q, r, s, and t. We will assume that A is a cubic, writing $A(x) = x^3 + Ux^2 + Vx + W$, and denote the roots u, v, and w.

Now, what properties must an auxiliary polynomial have? First, it must be possible to obtain the roots of p from the roots of A. After all, that is the entire point of having an auxiliary polynomial. Second, its coefficients must be expressible as functions of the coefficients of p, and for simplicity, we will require that they be polynomial functions. At the same time, for both polynomials, the coefficients are known functions of the roots. All of these relationships are shown in Fig. 4.1.

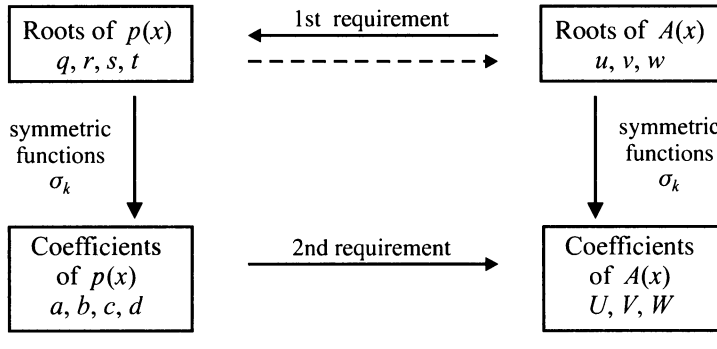


Figure 4.1. Related roots and coefficients for $p(x)$ and $A(x)$.

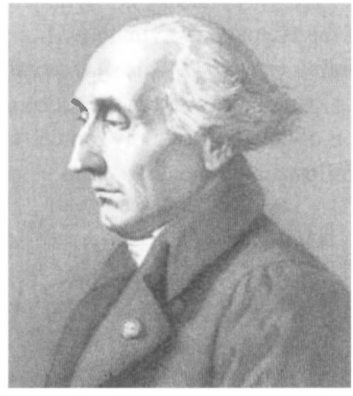
Inspired by these ideas, let us focus on how u, v, and w depend on q, r, s, and t, represented in the figure by the dashed arrow. Given equations

$$\begin{aligned} u &= f(q, r, s, t) \\ v &= g(q, r, s, t) \\ w &= h(q, r, s, t) \end{aligned} \tag{14}$$

we can satisfy the first requirement for an auxiliary polynomial by solving for q, r, s, and t in terms of u, v, and w. For the second requirement we make a key observation. The

[1] This is sometimes also referred to as a *resolvent* polynomial.

Lagrange and Vandermonde



Lagrange

Joseph-Louis Lagrange (1736–1813) and Alexandre-Théophile Vandermonde (1735–1796) made independent discoveries in the analysis of polynomial equations. A sketch of their contributions is given by Edwards [47, §15]. Edwards explains that although Vandermonde's early work in this area included promising insights, he (Vandermonde) went no further with it. The analyses of Vandermonde and Lagrange had much in common, but later developments primarily credited the contributions of Lagrange. Edwards' says

> Although Vandermonde had insightful ideas in other areas of mathematics as well, he does not seem to have followed these up either, and he is remembered today only because the name "Vandermonde determinant" was given to a determinant which, ironically, does not occur in his work at all.
>
> Unlike Vandermonde, who was French but did not have a French name, Lagrange had a French name but was not French. He was Italian (he was born with the name Lagrangia and his native city was Turin) and at the time of the publication of his [*Réflexions sur la Résolution Algébrique des Équations*, 1770–1771] he was a member of Frederick the Great's Academy in Berlin. When he left Berlin in 1787 he went to Paris, where he spent the rest of his life and where, of course, he was a leading member of the scientific community; this has tended to reinforce the impression that he was French. He was certainly the greatest mathematician of the generation between Euler and Gauss, and, indeed, has a secure place among the greatest mathematicians of all time.

Sidebar 4.3

coefficients of A will be expressible in terms of the coefficients of p if permutations of the roots of p leave the set $\{u, v, w\}$ unchanged.

To see why, note that (14) makes U, V, and W functions of q, r, s, and t, as suggested by the figure. Now consider what happens in (14) when we permute q, r, s, and t. We have not assumed that f, g, and h are symmetric functions, so the individual values of u, v, and w

may be altered. But this must amount to a permutation of u, v, and w if we assume the set $\{u, v, w\}$ remains unchanged. Then, U, V, and W will not change at all, since they depend symmetrically on u, v, and w. Thus U, V, and W are symmetric functions of q, r, s, and t, and so expressible in terms of the elementary symmetric functions σ_k by the fundamental theorem on symmetric polynomials (page 64). This shows that U, V, and W are expressible in terms of the coefficients of p, which differ from the σ_k by at most a change of sign.

To complete the construction of an auxiliary polynomial, we must specify the functions f, g, and h in (14). Here we follow Lagrange and define

$$\begin{aligned} u &= (q + r)(s + t) \\ v &= (q + s)(r + t) \\ w &= (q + t)(r + s). \end{aligned} \tag{15}$$

With these equations it is easy to check that every permution of q, r, s, and t leaves the set $\{u, v, w\}$ unchanged. Therefore, the coefficients of A will be expressible in terms of the the coefficients of p. And this is not just a theoretical result. Knowing neither the roots of p nor those of A, we can find the coefficients of A explicitly as functions of the coefficients of p. This echoes our earlier work with symmetric combinations of roots, for example the sum of the squares, which we expressed in terms of the coefficients. In essence, when we are dealing with explicit symmetric functions of the roots, the fundamental theorem on elementary symmetric polynomials allows us to invert the mapping from roots to coefficients of a polynomial. Thus, in the figure, we can add an arrow from the coefficients to the roots of p. This provides a path from the coefficients of p to the coefficients of A, and so satisfies the second requirement for an auxiliary polynomial.

However, before carrying out this step, we should verify that it will lead to a solution of the original quartic p. Since A is a cubic, once we know its coefficients we can find its roots, u, v, and w. But will we then be able to solve (15) for q, r, s, and t?

From u, v, w to q, r, s, t. Alone, the three equations of (15) are not enough to find the four roots of p. But we also know all of the coefficients of p. From our knowledge of the elementary symmetric functions,

$$q + r + s + t = -a. \tag{16}$$

With this additional equation, (15) can be solved for q, r, s, and t, as we shall now see.

Using (16), we eliminate t from (15) to obtain a solvable system of three equations in three unknowns. The algebra is simplified if the original quartic has no cubic term. That can always be arranged, we know, by making a change of variables. So without loss of generality, we assume that $a = 0$. Then $t = -q - r - s$, and 15 becomes

$$\begin{aligned} -(q + r)^2 &= u \\ -(q + s)^2 &= v \\ -(r + s)^2 &= w. \end{aligned}$$

This leads to

$$q + r = \alpha$$
$$q + s = \beta$$
$$r + s = \gamma$$

where $\alpha = \pm\sqrt{-u}$, $\beta = \pm\sqrt{-v}$, and $\gamma = \pm\sqrt{-w}$. Thus we obtain a linear system in q, r, and s, with solution

$$q = \frac{1}{2}(\alpha + \beta - \gamma)$$
$$r = \frac{1}{2}(\gamma - \beta + \alpha)$$
$$s = \frac{1}{2}(\gamma + \beta - \alpha).$$

And since $t = -(q + r + s)$, we also get

$$t = -(\alpha + \beta + \gamma)/2.$$

This verifies that we can find the roots of p once we know the roots of A.

Where does that leave us? We have seen that, with an understanding of symmetry, it is possible to formulate proposed roots of an auxiliary polynomial in terms of the unknown roots of a quartic. Without actually computing either the auxiliary polynomial or its roots, we have deduced that the auxiliary coefficients will be computable in terms of the known original coefficients, and that the roots of the original polynomial will be obtainable from the auxiliary roots. In this way, it is possible to engineer a method for solving quartics, without actually working out the steps of the method in detail.

To complete this analysis we should carry out the construction of $A(x)$. As usual, we may assume that the original quartic has no cubic term, so that $q + r + s + t = 0$. However, even with this simplifying assumption, the algebra required to find $A(x)$ remains forbidding. First, we know how the coefficients U, V, and W depend on the roots u, v, and w:

$$U = -(u + v + w)$$
$$V = uv + vw + wu$$
$$W = -uvw.$$

Next, using (15), we determine U, V, and W as functions of q, r, s, and t. They will be symmetric functions, and expressing them in terms of the elementary symmetric functions is the final step. That will give U, V, and W in terms of a, b, c, and d.

Although this program can be completed by hand, modern technology provides an easier alternative. We can use a computer algebra system, as discussed in Sidebar 3.3. This is practical even without the simplifying assumption $q + r + s + t = 0$. Using Maple and trial and error, it took me about half an hour to express U, V, and W in terms of the elementary symmetric functions in q, r, s, and t. Here are the results:

$$U = -2\sigma_2$$
$$V = \sigma_2^2 + \sigma_1\sigma_3 - 4\sigma_4$$
$$W = -\sigma_1\sigma_2\sigma_3 + \sigma_3^2 + \sigma_4\sigma_1^2.$$

How does this approach compare with those discussed in the preceding section? To compare the auxiliary polynomial derived here with the ones found earlier, we again impose the constraint $q+r+s+t = 0$, or equivalently, $\sigma_1 = 0$. Then the formulas for the coefficients U, V, and W simplify to

$$\begin{aligned} U &= -2\sigma_2 \\ V &= \sigma_2^2 - 4\sigma_4 \\ W &= \sigma_3^2. \end{aligned}$$

If we write the original polynomial as $x^4 + ax^2 + bx + c$, then $a = \sigma_2$, $b = -\sigma_3$, and $c = \sigma_4$. With those substitutions, the auxiliary polynomial becomes

$$A(x) = x^3 + Ux^2 + Vx + W = x^3 - 2ax^2 + (a^2 - 4c)x + b^2.$$

This is closely related to the auxiliary polynomials we saw in the previous section. For example, the auxiliary polynomial featured in Descartes' method is $-A(-u^2)$.

This shows that in all of the quartic methods we have considered, the auxiliary polynomial is essentially the same as the one found by Lagrange's analysis. Lagrange knew that the solvability of the cubic and quartic depended on the existence of auxiliary polynomials, which he showed could be discovered using the tools of symmetry. Attacking the quintic equation in the same way, Lagrange was led to an auxiliary polynomial of sixth degree, and he could find no way to reduce it. Ultimately he gave up the attempt to solve higher degree equations, concluding, in the words of Kline [101, p. 605] that *either the problem was beyond human capacities or the nature of the expressions for the roots must be different from all those thus far known.*

This conclusion was later proven correct. Our final topic for this chapter is the insolvability by radicals of equations of degree 5 and higher.

4.6 Quintic and Higher Degree Equations

It has already been mentioned that no general methods exist for solving quintics and higher degree equations in terms of radicals. The proof depends on an analysis of permutations of roots, and is fully developed in what today is called Galois Theory. It is beyond the scope of this book to give anything like a complete account of this topic. But having come so far in the discussion of polynomial equations, it would be a shame not to at least describe the main ideas in general terms.

We begin with the idea of a permutation group. As an instance, we might consider a set of four objects, say $\{q, r, s, t\}$. A permutation rearranges their order. Formally, a permutation is a function that maps the set of elements to itself, creating a one-to-one correspondence. One example of such a pairing is shown in Fig. 4.2.

Since a permutation is one-to-one, it is invertible. Reversing the directions of all the arrows defines a permutation, and applying the original and reverse permutations in order has the same effect as the identity function, $f(x) = x$. In general, when two permutations are composed, that is, applied one after the other, the result is again a permutation. If we denote the first permutation by α and the second by β, then for any element x of our set, the result of the combined operation will be $\beta(\alpha(x))$. It is customary to denote the combined operation $\beta\alpha$.

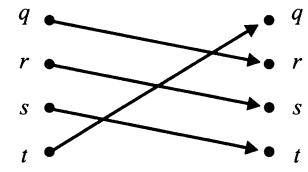


Figure 4.2. A sample permutation.

The set of permutations of any set forms an algebraic system called a *group*. Conceptually that means that the elements are invertible, and together they form a closed system. The same may be true of a subset of a group. In the set $\{q, r, s, t\}$ let α be the permutation that exchanges q and r, leaving s and t fixed. Then α is its own inverse, because applying it twice restores q and r to their original positions. The set consisting of α and the identity function thus constitute a closed system. It, too, is a permutation group, but not the group of all possible permutations of $\{q, r, s, t\}$, so it is called a *subgroup* of the full permutation group.

We have seen that the idea of permuting roots of a polynomial arises naturally when we express the roots in terms of radicals. In Cardano's solution of the cubic, there was an ambiguity in extracting a cube root of u. For complex u, there are three different cube roots, and switching from one to another permutes the roots of the original cubic. But we do not wish to deal only with polynomials that can be solved by radicals, so we need to understand how permutations might arise in other ways.

The modern viewpoint is to focus on number systems. We begin with the rational numbers, the system that contains the coefficients of our original polynomial. Next we extend the number system to include a root of our polynomial. If the root is, say, $3 + \sqrt{5}$, then we can form the set of numbers of the form $a + b\sqrt{5}$ with rational a and b. This formulation emphasizes $\sqrt{5}$ as a new element that must be incorporated into the number system. But we could equally well have incorporated the other square root of 5. In the end we arrive at the same number system. But the ambiguity in the choice of square roots of 5 has the effect of defining a transformation of the number system. The root $3+4\sqrt{5}$ defined as a result of one choice would instead have been $3 - 4\sqrt{5}$ had we made the other choice. This corresponds to the transformation $a+b\sqrt{5} \to a-b\sqrt{5}$, and is an analog of complex conjugation. There are two key properties of this transformation. First, it leaves all of the original elements of the number system unchanged. Second, it permutes the roots of our original polynomial. Thus, if $2 + 3\sqrt{5}$ is a root of a polynomial with rational coefficients, $2 - 3\sqrt{5}$ is also a root, and the transformation changes one into the other. More generally, permutations of roots arise any time we have a transformation of the extended number system that leaves elements of the original number system unchanged. The set of all transformations of this type gives rise to a group of permutations of the roots.

Galois theory is the study of number system transformations and the corresponding permutation groups. The root permutation group for a particular polynomial depends on the coefficients, and can either be the full permutation group or a subgroup. Distinguishing between these cases and understanding each group's structure are central aspects of the theory.

The idea of extending number systems also provides a way to consider roots that are expressed in terms of radicals. We begin again with the rational numbers, which we will

call system 1. We can extend this to a larger number system, system 2, by incorporating one new element, $\sqrt[m]{a}$, where a is some element of system 1. Next we extend the number system again by incorporating a second new element $\sqrt[m']{b}$, with m' possibly different from m, and with b some element of system 2. Continuing, we can build up a number system that incorporates any radical expression through a chain of extensions. At each stage of the chain, we incorporate one new mth root into the preceding number system.

What we add at each stage is a root to a very simple polynomial, one of the form $x^m - r$. In fact, all of these are roots of one master polynomial, the product of all the individual simple polynomials. Because the factors are very simple, Galois theory makes it possible to analyze the root permutation groups for the master polynomial, and the transformations of number systems at each stage of the process. Out of this comes a key result: If a fifth degree polynomial has roots expressible in terms of radicals, then the root permutation group for this polynomial cannot be the full group of permutations of the five roots (and similarly for polynomials of degree greater than 5).

At this point, we have enough of the background to consider an example: $p(x) = 3x^5 - 15x + 5$. Methods of calculus show that this polynomial has three real roots and two non-real complex roots. Using Galois theory, that is enough to imply that the root permutation group for this $p(x)$ is the full group of permutations of five roots. But it was stated earlier that this could not occur for a polynomial whose roots are expressible in terms of radicals. This shows that the roots of $3x^5 - 15x + 5$ cannot be expressible in terms of radicals.

What is the significance of this example? It shows that there cannot be a general method for solving quintics akin to the ones for cubics and quartics. If there were such a method, it would have to apply to $3x^5 - 15x + 5$, giving the roots in terms of radicals. But we know the roots have no such expression.

That does not mean that *no* quintics can be solved in terms of radicals. For example, we saw how to solve palindromic quintics with radicals in Chapter 2. But at least some quintics are not solvable using radicals, and that is enough to rule out the existence of a general method.

This completes our exploration of methods for solving polynomial equations. As mentioned early on, the idea of searching for roots in terms of radicals is somewhat arbitrary, and reflects the historical development of the subject. But this tradition is completely natural in the context of the elementary mathematics curriculum. As we have seen, cubics and quartics, like quadratics, have readily understood solutions using only elementary algebra. On the other hand, there are no such general solutions for equations of higher degree. Not surprisingly, it is easier to exhibit methods for solving cubics and quartics using radicals than it is to prove that corresponding methods cannot exist for higher degree equations. This is a central result of modern mathematics, and the ideas on which it rests remain a cornerstone for ongoing research in the field. Here, we have traced the beginnings of Galois theory, from properties of elementary symmetric functions and permutations of roots of polynomials arising out of the consideration of solutions for cubics and quartics .

Even without understanding all of the details of Galois theory, it is worthwhile to have some idea of what it involves. It is Galois theory, after all, that provides the complete picture of solvability in terms of radicals: there are known radical methods for degree four or less, and no such methods are possible for quintics or higher degree polynomials. That is certainly worth knowing.

4.7 History, References, and Additional Reading

For general historical reading on polynomial equations, both Katz [94] and Kline [101] are recommended. A more focussed treatment is provided by Edwards [47]. This work carefully develops the ideas of Lagrange, Galois, and the other key figures in the search for roots of polynomial equations and provides a rich source of historical information as well. Stewart makes the development of methods to solve polynomial equations the central theme of his account of mathematical symmetry [153].

Galois' life makes a dramatic tale, and some of the retellings apparently have emphasized drama over accuracy. For a fascinating account of the true story and some of the exaggerations, see Rothman [136]. Peterson [129] presents a lighter overview of the Galois story.

The history of polynomial equations is tightly connected to the history of algebraic methods and notation, which is nicely summarized by Gouvea and Berlinghoff [15]. In a related vein, Kleiner [99] argues that it was precisely the investigation of Cardano's method for cubics that forced the introduction of complex numbers.

In the discussion of Cardano's solution, we encountered a surprising arithmetic fact:

$$2 = \sqrt[3]{10 + 6\sqrt{3}} + \sqrt[3]{10 - 6\sqrt{3}}.$$

There is an article on identities of this sort by Osler [128].

References for the alternate solutions of quartics and cubics are: Viète's solution of the cubic, [101, p. 269]; Euler's solution of the cubic and quartic, [42]; Descartes' solution of the quartic and Cayley's solution of the cubic, [8, p. 20–22]; equality of two cubes, [49, 80, 156]; two cube completions, [55]; factorization identities, [91]; change of variables, [127, 160]; matrix algebra, [92].

Solutions to cubic and quartic equations can be constructed using paper folding techniques. An explanation of the solution of cubics can be found in [69, activity 6]. For more in depth discussions of cubics and quartics see [2, 46].

There are a great many other papers about the solution of cubics and quartics in expository mathematics journals, including [3, 65, 163, 168, 170].

The biographical information on Mark Kac in Sidebar 4.2 was taken from [126]. I learned about Kac's solution of the cubic in Roy's paper [137]. Although Kac describes solving the cubic in his autobiography, he does not say how he did it. Roy gives the details of the derivation that Kac published as a student.