Introduction

In 1878 a pair of brothers, the soon-to-become-infamous thieves Ahmed and Mohammed Abd er-Rassul, stumbled upon the ancient Egyptian burial site in the Valley of Kings, at Deir el-Bahri. They quickly had a thriving business going selling stolen relics, one of which was a mathematical papyrus; one of the brothers sold it to the Russian Egyptologist V. S. Golenishchev in 1893, who in turn gave it to the Museum of Fine Arts in Moscow in 1912.[1] There it remained, a mystery until its complete translation in 1930, at which time the scholarly world learned just how mathematically advanced the ancient Egyptians had been.

In particular, the fourteenth problem of the Moscow Mathematical Papyrus (MMP), as it is now called, is a specific numerical example of how to calculate the volume V of a truncated square pyramid, the so-called *frustum* of a pyramid. This example strongly suggests that the ancient Egyptians knew the formula

$$V = \frac{1}{3}h(a^2 + ab + b^2),$$

where a and b are the edge lengths of the bottom and top squares, respectively, and h is the height. One historian of science has called this knowledge “breath-taking” and “the masterpiece of Egyptian geometry.”[2] The derivation of this formula is a routine exercise for anyone who has had freshman calculus, but it is much less obvious how the Egyptians could have discovered it without a knowledge of integral calculus.[3]

While correct, this result does have one very slight stylistic flaw. The values of a and b are what a modern engineer or physicist would call an “observable,” i.e., they are lengths that can be directly determined simply by laying a tape measure out along the bottom and top edges of the frustum. The value of h, however, is not directly measurable, or certainly it isn’t for a solid pyramid. It can be calculated for any given pyramid, of course, using a knowledge of geometry and trigonometry, but how much more direct it would be to express the volume of the frustum in terms not of h, but of c, the *slant* edge length. That length *is* directly measurable. This was finally done but, as far as is known today, not until the first century A.D. by the great mathematician-engineer Heron of Alexandria, who is usually called a Greek but may have actually been an Egyptian. It is, in fact, an elementary problem in geometry to show that

$$h=\sqrt{c^2-2\left(\frac{a-b}{2}\right)^2}.$$

Now, let's skip ahead in time to 1897, to a talk given that year at a meeting of the American Association for the Advancement of Science by Wooster Woodruff Beman, a professor of mathematics at the University of Michigan, and a well-known historian of the subject. I quote from that address:

> We find the square root of a negative quantity appearing for the first time in the *Stereometria* of Heron of Alexandria . . . After having given a correct formula for the determination of the volume of a frustum of a pyramid with square base and applied it successfully to the case where the side of the lower base is 10, of the upper 2, and the edge 9, the author endeavors to solve the problem where the side of the lower base is 28, of the upper 4, and the edge 15. Instead of the square root of 81 − 144 required by the formula, he takes the square root of 144 − 81 . . . , i.e., he replaces $\sqrt{-1}$ by 1, and fails to observe that the problem as stated is impossible. Whether this mistake was due to Heron or to the ignorance of some copyist cannot be determined.[4]

That is, using $a = 28$, $b = 4$, and $c = 15$ in his formula for h, Heron wrote:

$$h=\sqrt{(15)^2-2\left(\frac{28-4}{2}\right)^2}=\sqrt{225-2(12)^2}=\sqrt{225-144-144}=\sqrt{81-144}.$$

The next, magnificent step would of course have been to write $h = \sqrt{-63}$, but the *Stereometria* records it as $h = \sqrt{63}$, and so Heron missed being the earliest known scholar to have derived the square root of a negative number in a mathematical analysis of a physical problem. If Heron really did fudge his arithmetic then he paid dearly for it in lost fame. It would be a thousand years more before a mathematician would even bother to take notice of such a thing—and then simply to dismiss it as obvious nonsense—and yet five hundred years more before the square root of a negative number would be taken seriously (but still be considered a mystery).

While Heron almost surely had to be aware of the appearance of the square root of a negative number in the frustum problem, his fellow Alexandrian two centuries later, Diophantus, seems to have completely missed a similar event when he chanced upon it. Diophantus is honored today as having played the same role in algebra that Euclid did in geometry. Euclid gave us his *Elements*, and Diophantus presented posterity with the *Arithmetica*. In both of these cases, the information contained was almost certainly the results of many previous, anonymous mathematicians whose identities are now lost forever to history. It was Euclid and Diophantus, however, who collected and organized this mathematical heritage in coherent form in their great works.

In my opinion, Euclid did the better job because *Elements* is a logical *theory* of plane geometry. *Arithmetica*, or at least the several chapters or books that have survived of the original thirteen, is, on the other hand, a collection of specific numerical solutions to certain problems, with no generalized, theoretical development of methods. Each problem in *Arithmetica* is unique unto itself, much like those on the Moscow Mathematical Papyrus. But this is not to say that the solutions given are not ingenious, and in many cases even diabolically clever. *Arithmetica* is still an excellent hunting ground for a modern teacher of high school algebra looking for problems to challenge, even stump, the brightest of students.[5]

In book 6, for example, we find the following problem (number 22): Given a right triangle with area 7 and perimeter 12, find its sides. Here's how Diophantus derived the quadratic equation $172x = 336x^2 + 24$ from the statement of the problem. With the sides of the right triangle denoted by P_1 and P_2, the problem presented by Diophantus is equivalent to solving the simultaneous equations

$$P_1P_2 = 14,$$
$$P_1 + P_2 + \sqrt{P_1^2 + P_2^2} = 12.$$

These can be solved by routine, if somewhat lengthy, algebraic manipulation, but Diophantus' clever idea was to immediately reduce the number of variables from two to one by writing

$$P_1 = \frac{1}{x} \text{ and } P_2 = 14\text{x}.$$

Then the first equation reduces to the identity $14 = 14$, and the second to

$$\frac{1}{x} + 14x + \sqrt{\frac{1}{x^2} + 196x^2} = 12,$$

which is easily put into the form given above,

$$172x = 336x^2 + 24.$$

It is a useful exercise to directly solve the original P_1, P_2 equations, and then to show that the results are consistent with Diophantus' solution.

Diophantus wrote the equation the way he did because it displays all the coefficients as positive numbers, i.e., the ancients rejected negative numbers as being without meaning because they could see no way physically to interpret a number that is "less than nothing." Indeed, elsewhere in *Arithmetica* (problem 2 in book 5) he wrote, of the equation $4x + 20 = 4$, that it was "absurd" because it would lead to the "impossible" solution $x = -4$. In accordance with this position, Diophantus used only the positive root when solving a quadratic. As late as the sixteenth century we find mathematicians referring to the negative roots of an equation as *fictitious* or *absurd* or *false*.

And so, of course, the square root of a negative number would have simply been beyond the pale. It is the French mathematician René Descartes, writing fourteen centuries later in his 1637 *La Geometrie,* whose work I will discuss in some detail in chapter 2, to whom we owe the term *imaginary* for such numbers. Before Descartes' introduction of this term, the square roots of negative numbers were called *sophisticated* or *subtle*. It is just such a thing, in fact, that Diophantus' quadratic equation for the triangle problem results in, i.e., the quadratic formula quickly gives the solutions

$$x = \frac{43 \pm \sqrt{-167}}{168}.$$

But this is not what Diophantus wrote. What he wrote was simply that the quadratic equation was not possible. By that he meant the equation has no rational solution because "half the coefficient of x multiplied into itself, minus the product of the coefficient of x^2 and the units" must make a square for a rational solution to exist, while

$$\left(\frac{172}{2}\right)^2 - (336)(24) = -668$$

certainly is not a square. As for the square root of this negative number, Diophantus had nothing at all to say.

Six hundred years later (circa 850 A.D.) the Hindu mathematician Mahaviracarya wrote on this issue, but then only to declare what Heron and Diophantus had practiced so long before: "The square of a positive as well as of a negative (quantity) is positive; and the square roots of those (square quantities) are positive and negative in order. As in the nature of things a negative (quantity) is not a square (quantity), *it has therefore no square root* [my emphasis]."[6] More centuries would pass before opinion would change.

At the beginning of George Gamow's beautiful little book of popularized science, *One Two Three . . . Infinity,* there's the following limerick to give the reader a flavor both of what is coming next, and of the author's playful sense of humor:

There was a young fellow from Trinity
Who took $\sqrt{\infty}$.
But the number of digits
Gave him the fidgets;
He dropped Math and took up Divinity.

This book is not about the truly monumental task of taking the square root of infinity, but rather about another task that a great many very clever mathematicians of the past (certainly including Heron and Diophantus) thought an even more absurd one—that of figuring out the meaning of the square root of minus one.

CHAPTER ONE

The Puzzles of Imaginary Numbers

1.1 The Cubic Equation

At the end of his 1494 book *Summa de Arithmetica, Geometria, Proportioni et Proportionalita,* summarizing all the knowledge of that time on arithmetic, algebra (including quadratic equations), and trigonometry, the Franciscan friar Luca Pacioli (circa 1445–1514) made a bold assertion. He declared that the solution of the cubic equation is "as impossible at the present state of science as the quadrature of the circle." The latter problem had been around in mathematics ever since the time of the Greek mathematician Hippocrates, circa 440 B.C. The quadrature of a circle, the construction by straightedge and compass alone of the square equal in area to the circle, had proven to be difficult, and when Pacioli wrote the quadrature problem was still unsolved. He clearly meant only to use it as a measure of the difficulty of solving the cubic, but actually the quadrature problem is a measure of the *greatest* difficulty, since it was shown in 1882 to be impossible.

Pacioli was wrong in his assertion, however, because within the next ten years the University of Bologna mathematician Scipione del Ferro (1465–1526) did, in fact, discover how to solve the so-called *depressed cubic*, a special case of the general cubic in which the second-degree term is missing. Because his solution to the depressed cubic is central to the first progress made toward understanding the square root of minus one, it is worth some effort in understanding just what del Ferro did.

The general cubic contains all the powers of the unknown, i.e.,

$$x^3 + a_1x^2 + a_2x + a_3 = 0$$

where we can take the coefficient of the third-degree term to be unity without loss of any generality. If that coefficient is not one, then we just divide through the equation by the coefficient, which we can always do unless it is zero—but then the equation isn't really a cubic.

The cubic solved by del Ferro, on the other hand, has the general form of

$$x^3 + px = q,$$

where p and q are non-negative. Just like Diophantus, sixteenth-century mathematicians, del Ferro included, avoided the appearance of negative coefficients in their equations.[1] Solving this equation may seem to fall somewhat short of solving the general cubic, but with the discovery of one last ingenious trick del Ferro's solution is general. What del Ferro somehow stumbled upon is that solutions to the depressed cubic can be written as the sum of two terms, i.e., we can express the unknown x as $x = u + v$. Substituting this into the depressed cubic, expanding, and collecting terms, results in

$$u^3 + v^3 + (3uv + p)(u + v) = q.$$

This single, rather complicated-looking equation, can be rewritten as two individually less complicated statements:

$$3uv + p = 0$$

which then says

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

How did del Ferro know to do this? The Polish-American mathematician Mark Kac (1914–84) answered this question with his famous distinction between the ordinary genius and the magician genius: "An ordinary genius is a fellow that you and I would be just as good as, if we were only many times better. There is no mystery as to how his mind works. Once we understand what he has done, we feel certain that we, too, could have done it. It is different with the magicians . . . the working of their minds is for all intents and purposes incomprehensible. Even after we understand what they have done, the process by which they have done it is completely dark." Del Ferro's idea was of the magician class.

Solving the first equation for v in terms of p and u, and substituting into the second equation, we obtain

$$u^6 - qu^3 - \frac{p^3}{27} = 0.$$

At first glance this sixth-degree equation may look like a huge step backward, but in fact it isn't. The equation is, indeed, of the sixth degree, but it is also quadratic in u^3. So, using the solution formula for quadratics, well-known since Babylonian times, we have

$$u^3 = \frac{q}{2} \pm \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}.$$

or, using just the positive root,[2]

$$u = \sqrt[3]{\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}.$$

Now, since $v^3 = q - u^3$, then

$$v = \sqrt[3]{\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}.$$

Thus, a solution to the depressed cubic $x^3 + px = q$ is the fearsome-looking expression

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

Alternatively, since $\sqrt[3]{-1} = -1$, then the second term in this expression can have a -1 factor taken through the outer radical to give the equivalent

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

You can find both forms stated in different books discussing cubics, but there is no reason to prefer one over the other.

Since p and q were taken by del Ferro to be positive, it is immediately obvious that these two (equivalent) expressions for x will always give a real result. In fact, although there are three solutions or *roots* to any cubic (see appendix A), it is not hard to show that there is always exactly one real, positive root and therefore two complex roots to del Ferro's cubic (see box 1.1).

Now, before continuing with the cubic let me say just a bit about the nature of complex numbers. A *complex* number is neither purely real nor purely imaginary, but rather is a composite of the two. That is, if a and b are both purely real, then $a + b\sqrt{-1}$ is complex. The form used by mathematicians and nearly everybody else is $a + ib$ (the great eighteenth-century Swiss mathematician Leonhard Euler, about whom much more is said in chapter 6, introduced the i symbol for $\sqrt{-1}$ in 1777). This is written as $a + jb$ by electrical engineers. The reason electrical engineers generally opt for j is that $\sqrt{-1}$ often occurs in their problems when electric currents are involved, and the letter i is traditionally reserved for that quantity. Contrary to popular myth, however, I can assure you that most electrical engineers are *not* confused when they see an equation involving complex numbers written with $i = \sqrt{-1}$ rather than with j. With that said, however, let me admit that in chapter 5 I, BOX 1.1

The One Real, Positive Solution to del Ferro's Cubic Equation

To see that there is exactly one real, positive root to the depressed cubic $x^3 + px = q$, where p and q are both non-negative, consider the function

$$f(x) = x^3 + px - q.$$

Del Ferro's problem is that of solving for the roots of $f(x) = 0$. Now, if you calculate the derivative of $f(x)$ [denoted by $f'(x)$] and recall that the derivative is the slope of the curve $f(x)$, then you will get

$$f'(x) = 3x^2 + p,$$

which is always non-negative because x^2 is never negative, and we are assuming p is non-negative. That is, $f(x)$ always has non-negative slope, and so never decreases as x increases. Since $f(0) = -q$, which is never positive (since we are assuming that q is non-negative), then a plot of $f(x)$ must look like figure 1.1. From the figure it is clear that the curve crosses the x-axis only once, thus locating the real root, and that the crossing is such that the root is never negative (it is zero only if $q = 0$).

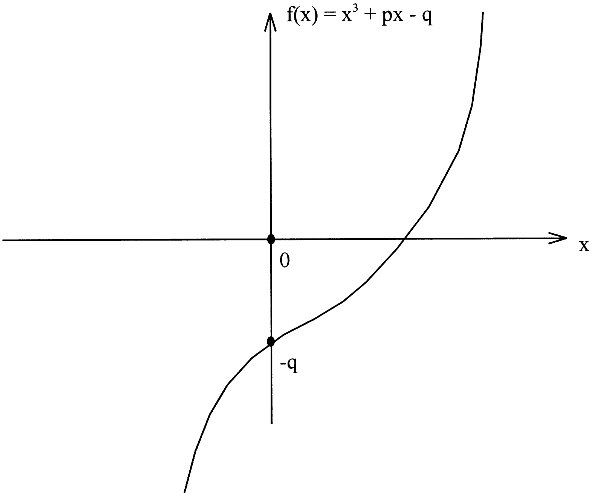

Figure 1.1. Plot of $f(x) = x^3 + px - q$, p and $q \geq 0$.

too, will use j rather than i for $\sqrt{-1}$ when I show you a nice little electrical puzzle from the nineteenth century.

Complex numbers obey many of the obvious rules, e.g., $(a + ib)(c + id) = ac + iad + ibc + i^2bd = ac - bd + i(ad + bc)$. But you do have to be careful. For example, if a and b can both only be positive, then $\sqrt{ab} = \sqrt{a}\sqrt{b}$. But if we allow negative numbers, too, this rule fails, e.g., $\sqrt{(-4)(-9)} = \sqrt{36} = 6 \neq \sqrt{-4}\sqrt{-9} = (2i)(3i) = 6i^2 = -6$. Euler was confused on this very point in his 1770 *Algebra*.

One final, very important comment on the reals versus the complex. Complex numbers fail to have the ordering property of the reals. *Ordering* means that we can write statements like $x > 0$ or $x < 0$. Indeed, if x and y are both real, and if $x > 0$ and $y > 0$, then their product $xy > 0$. If we try to impose this behavior on complex numbers, however, then we get into trouble. An easy way to see this is by a counterexample. That is, let us suppose we *can* order the complex numbers. Then, in particular, it must be true that either $i > 0$ or $i < 0$. Suppose $i > 0$. Then, $-1 = i \cdot i > 0$, which is clearly false. So we must suppose $i < 0$, which when we multiply through by -1 (which flips the sense of the inequality) says $-i > 0$. Then, $-1 = (-i) \cdot (-i) > 0$, just as before, and it is still clearly false. The conclusion is that the original assumption of ordering leads us into contradiction, and so that assumption must be false. Now back to cubics.

Once we have the real root to del Ferro's cubic, then finding the two complex roots is not difficult. Suppose we denote the real root given by del Ferro's equation by r_1. Then we can factor the cubic as

$$(x - r_1)(x - r_2)(x - r_3) = 0 = (x - r_1)[x^2 - x(r_2 + r_3) + r_2r_3].$$

To find the two additional roots, r_2 and r_3, we can then apply the quadratic formula to

$$x^2 - x(r_2 + r_3) + r_2r_3 = 0.$$

For example, consider the case of $x^3 + 6x = 20$, where we have $p = 6$ and $q = 20$. Substituting these values into the second version of del Ferro's formula gives

$$x = \sqrt[3]{10 + \sqrt{108}} - \sqrt[3]{-10 + \sqrt{108}}.$$

Now, if you look at the original cubic long enough, perhaps you'll have the lucky thought that $x = 2$ works $(8 + 12 = 20)$. So could that complicated-looking thing with all the radical signs that I just wrote *actually be* 2? Well, yes, it is. Run it through a hand calculator and you will see that

$$x = \sqrt[3]{20.392305} - \sqrt[3]{0.392305} = 2.7320508 - 0.7320508 = 2.$$

Then, to find the other two roots to $f(x) = 0 = x^3 + 6x - 20$, we use the fact that one factor of $f(x)$ is $(x - 2)$ to find, with some long division, that

$$(x - 2)(x^2 + 2x + 10) = x^3 + 6x - 20.$$

Applying the quadratic formula to the quadratic factor quickly gives the two complex roots (solutions to the original cubic) of

$$r_2 = -1 + 3\sqrt{-1}$$

and

$$r_3 = -1 - 3\sqrt{-1}.$$

1.2 Negative Attitudes about Negative Numbers

But this is all getting ahead of the story. Del Ferro and his fellow mathematicians did not, in fact, do any of the above sort of factoring to get the complex roots—the finding of a single, real, positive number for the solution of a cubic was all they were after. And, as long as mathematicians concerned themselves with del Ferro's original depressed cubic, then a single, real, positive root is all there is, and all was well. But what of such a cubic as $x^3 - 6x = 20$, where now we have $p = -6 < 0$? Del Ferro would never have written such a cubic, of course, with its negative coefficient, but rather would have written $x^3 = 6x + 20$ and would have considered this an entirely new problem. That is, he would have started over from the beginning to solve

$$x^3 = px + q$$

with, again, both p and q non-negative. This is totally unnecessary, however, as at no place in the solution to $x^3 + px = q$ did he ever actually use the non-negativity of p and q. That is, such assumptions have no importance, and were explicitly made simply because of an unwarranted aversion by early mathematicians to negative numbers.

This suspicion of negative numbers seems so odd to scientists and engineers today, however, simply because they are used to them and have forgotten the turmoil they went through in their grade-school years. In fact, intelligent, nontechnical adults continue to experience this turmoil, as illustrated in the following wonderful couplet, often attributed to the poet W. H. Auden:

Minus times minus is plus.
The reason for this we need not discuss.

The great English mathematician John Wallis (1616–1703), for example, whom you will meet in more detail later in the next chapter as the individual who made the first rational attempt to attach physical significance to $\sqrt{-1}$, also made some incredible assertions concerning negative numbers. In his 1665 book *Arithmetica Infinitorum*, an influential book read with great interest by the young Isaac Newton, Wallis made the following argument. Since $a \div 0$, with $a > 0$, is positive infinity, and since $a \div b$, with $b < 0$, is a negative number, then this negative number must be *greater* than positive infinity because the denominator in the second case is less than the denominator in the first case (i.e., $b < 0$). This left Wallis with the astounding conclusion that a negative number is simultaneously both less than zero and greater than positive infinity, and so who can blame him for being wary of negative numbers? And, of course, he was not alone. Indeed, the great Euler himself thought Auden's concern sufficiently meritorious that he included a somewhat dubious "explanation" for why "minus times minus is plus" in his famous textbook *Algebra* (1770).

We are bolder today. Now we simply say, okay, p is negative (so what?) and plug right into the original del Ferro formula. That is, replacing the negative p with $-p$ (where now p itself is non-negative) we have

$$x = \sqrt[3]{\frac{q}{2} + \sqrt{\frac{q^2}{4} - \frac{p^3}{27}}} - \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} - \frac{p^3}{27}}}$$

as the solution to $x^3 = px + q$, with p and q both non-negative. In particular, the formula tells us that the solution to $x^3 = 6x + 20$ is

$$x = \sqrt[3]{10 + \sqrt{92}} - \sqrt[3]{-10 + \sqrt{92}} = 3.4377073$$

which is indeed a solution to the cubic, as can be easily verified with a handheld calculator.

1.3 A Rash Challenge

The story of the cubic now takes a tortured, twisted path. As was the tradition in those days, del Ferro kept his solution secret. He did this because, unlike today's academic mathematicians who make their living publishing their results to earn first appointment to a junior professorship and later promotion and tenure, del Ferro and his colleagues were more like independently employed businessmen. They earned their livelihoods by challenging each other to public contests of problem solving, and the winner took all—prize money, maybe, certainly "glory," and with luck the support of an admiring and rich patron. One's chances of winning such contests were obviously enhanced by knowing how to solve problems that others could not, so secrecy was the style of the day.

In fact, del Ferro almost took the secret of how to solve depressed cubics to the grave, telling at most only a small number of close friends. As he lay dying he told one more, his student Antonio Maria Fior. While Fior was not a particularly good mathematician, such knowledge was a formidable weapon and so, in 1535, he challenged a far better known and infinitely more able mathematician, Niccolo Fontana (1500–57). Fontana had come to Fior's attention because Fontana had recently announced that he could solve cubics of the general form $x^3 + px^2 = q$. Fior thought Fontana was bluffing, that he actually had no such solution, and so Fior saw him as the perfect victim, ripe for the plucking of a public contest.

Fontana, who is better known today as simply Tartaglia ("the stammerer," because of a speech impediment caused by a terrible sword wound to the jaw that he received from an invading French soldier when he was twelve), suspected Fior had received the secret of the depressed cubic from del Ferro. Fearing it would be such cubics he would be challenged with, and not knowing how to solve them, Tartaglia threw himself with a tremendous effort into solving the depressed cubic; just before the contest day he succeeded in rediscovering del Ferro's solution of $x^3 + px = q$ for himself. This is an interesting example of how, once a problem is *known* to have a solution, others quickly find it, too—a phenomenon related, I think, to sports records, e.g., within months of Roger Bannister breaking the four-minute mile it seemed as though every good runner in the world started doing it. In any case, Tartaglia's discovery, combined with his ability to really solve $x^3 + px^2 = q$ (he had *not* been bluffing), allowed him to utterly defeat Fior. Each proposed thirty problems for the other, and while Fior could solve none of Tartaglia's, Tartaglia solved all of Fior's.

1.4 The Secret Spreads

All of this is pretty bizarre, but the story gets even better. Like del Ferro, Tartaglia kept his newly won knowledge to himself, both for the reasons I mentioned before and because Tartaglia planned to publish the solutions to both types of cubics himself, in a book he thought he might one day write (but never did). When the news of his rout of Fior spread, however, it quickly reached the ear of Girolamo Cardano (1501–76), otherwise known simply as Cardan. Unlike Fior, Cardan was an outstanding intellect who, among his many talents, was an extremely good mathematician.[3] Cardan's intellectual curiosity was fired by the knowledge that Tartaglia knew the secret to the depressed cubic, and he begged Tartaglia to reveal it. After initially refusing, Tartaglia eventually yielded and told Cardan the rule, but not the derivation, for calculating solutions—and even then only after extracting a vow of secrecy.

Cardan was not a saint, but he also was not a scoundrel. He almost certainly had every intention of honoring his oath of silence, but then he began to hear that Tartaglia was not the first to solve the depressed cubic. And once he had actually seen the surviving papers of del Ferro, Cardan no longer felt bound to keep his silence. Cardan rediscovered Tartaglia's solution for himself and then published it in his book *Ars Magna* (*The Great Art*—of algebra, as opposed to the lesser art of arithmetic) in 1545. In this book he gave Tartaglia *and* del Ferro specific credit, but still Tartaglia felt wronged and he launched a blizzard of claims charging Cardan with plagiarism and worse.[4] This part of the story I will not pursue here, as it has nothing to do with $\sqrt{-1}$, except to say that Tartaglia's fear of lost fame in fact came to pass. Even though he and del Ferro indeed had priority as the true, independent discoverers of the solution to the depressed cubic, ever since *Ars Magna*, it has been known as the "Cardan formula."

Cardan was not an intellectual thief (plagiarists don't give attributions), and in fact he showed how to extend the solution of the depressed cubic to *all* cubics. This was a major achievement in itself, and it is all Cardan's. The idea is as inspired as was del Ferro's original breakthrough. Cardan started with the general cubic

$$x^3 + a_1x^2 + a_2x + a_3 = 0$$

and then changed variable to $x = y - \frac{1}{3}a_1$. Substituting this back into the general cubic, expanding, and collecting terms, he obtained

$$y^3 + \left(a_2 - \frac{1}{3}a_1^2\right)y = -\frac{2}{27}a_1^3 + \frac{1}{3}a_2a_1 - a_3.$$

That is, he obtained the depressed cubic $y^3 + py = q$ with

$$p = a_2 - \frac{1}{3}a_1^2,$$

$$q = -\frac{2}{27}a_1^3 + \frac{1}{3}a_1a_2 - a_3.$$

The depressed cubic so obtained can now be solved with the Cardan formula. For example, if you start with $x^3 - 15x^2 + 81x - 175 = 0$ and then make Cardan's change of variable $x = y + 5$, you will get

$$p = 81 - \frac{1}{3}(15)^2 = 6,$$

$$q = -\frac{2}{27}(-15)^3 + \frac{1}{3}(81)(-15) - (-175) = 20,$$

and so $y^3 + 6y = 20$. I solved this equation earlier in this chapter, getting $y = 2$. Thus, $x = 7$ is the solution to the above cubic, as a hand calculation will quickly confirm.

So it looks as if the cubic equation problem has finally been put to rest, and all is well. Not so, however, and Cardan knew it. Recall the solution to $x^3 = px + q$,

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

There is a dragon lurking in *this* version of the Cardan formula! If $q^2/4 - p^3/27 < 0$ then the formula involves the square root of a negative number, and the great puzzle was not the imaginary number itself, but something quite different. The fact that Cardan had no fear of imaginaries themselves is quite clear from the famous problem he gives in *Ars Magna,* that of dividing ten into two parts whose product is forty. He calls this problem "manifestly impossible" because it leads immediately to the quadratic equation $x^2 - 10x + 40 = 0$, where x and $10 - x$ are the two parts, an equation with the complex roots—which Cardan called *sophistic* because he could see no physical meaning to them—of $5 + \sqrt{-15}$ and $5 - \sqrt{-15}$. Their sum is obviously ten because the imaginary parts cancel, but what of their product? Cardan boldly wrote "nevertheless we will operate" and formally calculated

$$(5 + \sqrt{-15})(5 - \sqrt{-15}) = (5)(5) - (5)(\sqrt{-15}) + (5)(\sqrt{-15}) - (\sqrt{-15})(\sqrt{-15}) = 25 + 15 = 40.$$

As Cardan said of this calculation, "Putting aside the mental tortures involved" in doing this, i.e., in manipulating $\sqrt{-15}$ just like any other number, everything works out. Still, while not afraid of such numbers, it is clear from his next words that he viewed them with more than a little suspicion: "So progresses arithmetic subtlety the end of which, as is said, is as refined as it is useless." But what *really* perplexed Cardan was the case of such square roots of negative numbers occurring in the Cardan formula for cubic equations that clearly had only real solutions.

1.5 How Complex Numbers Can Represent Real Solutions

To see what I mean by this, consider the problem treated by Cardan's follower, the Italian engineer-architect Rafael Bombelli (1526–72). Bombelli's fame among his contemporaries was as a practical man who knew how to drain swampy marshes, but today his fame is as an expert in algebra who explained what is really going in Cardan's formula. In his *Algebra* of 1572, Bombelli presents the cubic $x^3 = 15x + 4$ and, with perhaps just a little pondering, you can see that $x = 4$ is a solution. Then, using long division/factoring, you can easily show that the two other solutions are $x = -2 \pm \sqrt{3}$. That is, *all three* solutions are real. But look at what the Cardan formula gives, with $p = 15$ and $q = 4$. Since $q^2/4 = 4$ and $p^3/27 = 125$, then

$$x = \sqrt[3]{2+\sqrt{-121}} - \sqrt[3]{-2+\sqrt{-121}} = \sqrt[3]{2+\sqrt{-121}} + \sqrt[3]{2-\sqrt{-121}}.$$

Cardan's formula gives a solution that is the sum of the cube roots of two complex conjugates (if this word is strange to you, then you should read appendix A), and you might think that if anything isn't real it will be something as "complex" as that, right? Wrong. Cardan did not realize this; with obvious frustration he called the cubics in which such a strange result occurred "irreducible" and pursued the matter no more. It is instructive, before going further, to see why he used the term "irreducible."

Cardan was completely mystified by how to actually calculate the cube root of a complex number. To see the circular loop in algebra that caused his confusion, consider Bombelli's cubic. Let us suppose that, whatever the cube root in the solution given by the Cardan formula is, we can at least write it most generally as a complex number. For example, let us write

$$\sqrt[3]{2+\sqrt{-121}} = u + \sqrt{-v}.$$

We wish to find both u and v (where $v > 0$). Cubing both sides gives

$$2 + \sqrt{-121} = u^3 + 3u^2\sqrt{-v} - 3uv - v\sqrt{-v}.$$

Equating the real and imaginary parts on both sides, we then get

$$u^3 - 3uv = 2,$$
$$3u^2\sqrt{-v} - v\sqrt{-v} = \sqrt{-121}.$$

Squaring both of these equations gives us another pair:

$$u^6 - 6u^4v + 9u^2v^2 = 4,$$
$$-9u^4v + 6u^2v^2 - v^3 = -121.$$

and subtracting the second equation from the first equation results in

$$u^6 + 3u^4v + 3u^2v^2 + v^3 = 125.$$

Both sides of this are perfect cubes, i.e., taking cube roots gives $u^2 + v = 5$, or $v = 5 - u^2$. Substituting this back into the $u^3 - 3uv = 2$ equation above results in $4u^3 = 15u + 2$, another cubic equation in a single variable. And, in fact, dividing through by 4 to put it in the form $u^3 = pu + q$, we have $p = \frac{15}{4}$ and $q = \frac{1}{2}$ and so, using the formula at the end of section 1.2,

$$\frac{q^2}{4} - \frac{p^3}{27} = \frac{1}{16} - \frac{3{,}375}{(27)(64)},$$

which is clearly negative.

That is, $4u^3 = 15u + 2$ is an irreducible cubic and will, when "solved" by the Cardan formula, result in having to calculate the cube roots of complex numbers. So we are right back where we started, faced with the problem of how to calculate such a thing. The problem seems to be stuck in a loop. No wonder Cardan called this situation "irreducible." Later, in chapter 3, you will see how mathematicians eventually discovered how to calculate *any* root of a complex number.

It was Bombelli's great insight to see that the weird expression that Cardan's formula gives for x is real, but expressed in a very unfamiliar manner (see box 1.2 for what is going on geometrically in irreducible cubics). This insight did not come easily. As Bombelli wrote in his *Algebra*, "It was a wild thought in the judgement of many; and I too for a long time was of the same opinion. The whole matter seemed to rest on sophistry rather than on truth. Yet I sought so long, until I actually proved this to be the case." Here's how he did it, beginning with the observation that if the Cardan formula solution is actually real then it must be that $\sqrt[3]{2+\sqrt{-121}}$ and $\sqrt[3]{2-\sqrt{-121}}$ are complex conjugates,[5] i.e., if a and b are some yet to be determined real numbers, where

$$\sqrt[3]{2+\sqrt{-121}} = a + b\sqrt{-1},$$

$$\sqrt[3]{2-\sqrt{-121}} = a - b\sqrt{-1},$$

then we have $x = 2a$, which is indeed real. The first of these two statements says that

$$2 + \sqrt{-121} = (a + b\sqrt{-1})^3.$$

From the identity $(m + n)^3 = m^3 + n^3 + 3mn(m + n)$, with $m = a$ and $n = b\sqrt{-1}$, we get

BOX 1.2

The Irreducible Case Means There Are Three Real Roots

To study the nature of the roots to $x^3 = px + q$, where p and q are both non-negative, consider the function

$$f(x) = x^3 - px - q.$$

Calculating $f'(x) = 3x^2 - p$, we see that the plot of $f(x)$ will have tangents with zero slope at $x = \pm\sqrt{p/3}$, i.e., the local extrema of the depressed cubic that can lead to the irreducible case are symmetrically located about the vertical axis. The values of $f(x)$ at these two local extrema are, if we denote them by M_1 and M_2,

$$M_1 = \frac{p}{3}\sqrt{\frac{p}{3}} - p\sqrt{\frac{p}{3}} - q = -\frac{2}{3}p\sqrt{\frac{p}{3}} - q, \text{ at } x = +\sqrt{\frac{p}{3}},$$

$$M_2 = -\frac{p}{3}\sqrt{\frac{p}{3}} + p\sqrt{\frac{p}{3}} - q = \frac{2}{3}p\sqrt{\frac{p}{3}} - q, \text{ at } x = -\sqrt{\frac{p}{3}}.$$

Notice that the local minima $M_1 < 0$, always (as p and q are both non-negative), while the local maximum M_2 can be of either sign, depending on the values of p and q. Now, if we are to have three real roots, then $f(x)$ must cross the x-axis three times and this will happen only if $M_2 > 0$, as shown in figure 1.2. That is, the condition for all real roots is $\frac{2}{3}p\sqrt{p/3} - q > 0$, or $\frac{4}{27}p^3 > q^2$, or, at last, $q^2/4 - p^3/27 < 0$. But this is precisely the condition in the Cardan formula that leads to imaginary numbers in the solution. That is, the occurrence of the irreducible case is always associated with three real roots to the cubic $f(x) = 0$. As the figure also makes clear, these three roots are such that two are negative and one is positive. See if you can also show that the sum of the three roots must be zero.*

* This is a particular example of the following more general statement. Suppose we write the nth-degree polynomial equation $x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_1x + a_0 = 0$ in factored form. That is, if we denote the n roots of the equation by $r_1, r_2, \ldots, r_n$, then we can write $(x - r_1)(x - r_2)\cdots(x - r_n) = 0$. By successively multiplying the factors together, starting at the left, you can easily show that the coefficient of the x^{n-1} term is the negative of the sum of the roots, i.e., that $a_{n-1} = -(r_1 + r_2 + \cdots + r_n)$. In the case of the depressed cubic, with no x^2 term, we have $a_2 = 0$ by definition, i.e., the sum of the roots of any depressed cubic equation is zero.

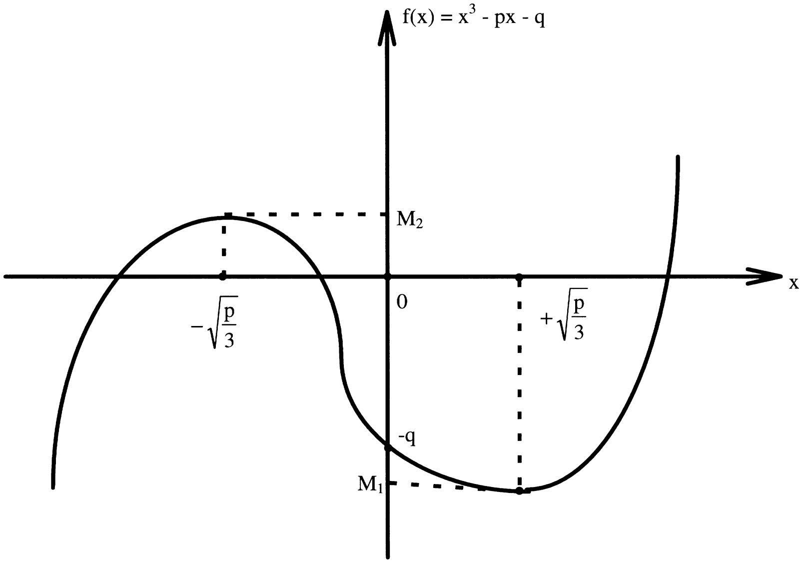

Figure 1.2. Plot of $f(x) = x^3 - px - q$, p and $q \geq 0$.

$$\begin{aligned}(a + b\sqrt{-1})^3 &= a^3 - b^3\sqrt{-1} + 3ab\sqrt{-1}\,(a + b\sqrt{-1})\\ &= a^3 - b^3\sqrt{-1} + 3a^2b\sqrt{-1} - 3ab^2\\ &= a(a^2 - 3b^2) + b(3a^2 - b^2)\sqrt{-1}.\end{aligned}$$

If this complex expression is to equal the complex number $2 + \sqrt{-121}$, then the real and imaginary parts must be separately equal, and so we arrive at the following pair of conditions:

$$a(a^2 - 3b^2) = 2,$$
$$b(3a^2 - b^2) = 11.$$

If we assume a and b are both integers (there is no a priori justification for this, but we are always free to try something and see where it goes), then perhaps you will notice that $a = 2$ and $b = 1$ work in both conditions. There are also ways to work this conclusion out more formally. For example, notice that 2 and 11 are prime, ask yourself what are the integer factors of any prime, and notice that if a and b are integers then so are $a^2 - 3b^2$ and $3a^2 - b^2$. For our purposes here, however, it is sufficient to see that

$$\sqrt[3]{2 + \sqrt{-121}} = 2 + \sqrt{-1},$$

$$\sqrt[3]{2 - \sqrt{-121}} = 2 - \sqrt{-1},$$

statements that are easily verified by cubing both sides. With these results Bombelli thus showed that the mysterious Cardan solution *is* $x = 4$, and this *is* correct. As shown in box 1.2, for the irreducible case with all three roots real, there is just one positive root; that is, the root given by the Cardan formula (see if you can prove this—read the last half of appendix A if you need help).

1.6 Calculating the Real Roots without Imaginaries

Still, while the Cardan formula works in all cases, including the irreducible case, you might be wondering why there is not a formula that directly produces a real answer for the positive real root in the irreducible case. And, in fact, there is. Discovered by the great French mathematician Francoise Viète[6] (1540–1603), it gives all the roots of the irreducible cubic in terms of the cosine and arccosine (or inverse cosine) trigonometric functions. This discovery is all the more remarkable when one considers that Viète was not a professional mathematician, but rather was a lawyer in service to the state, under the kings Henri III and Henri IV. He did his mathematics when he could steal time away from his "more important" duties, such as decoding intercepted, encrypted letters written by the Spanish court during France's war with Spain. While clever, Viète's solution (published posthumously in 1615) seems not to be very well known, and so here is what he did.

Viète started his analysis with the cubic equation $x^3 = px + q$, with p and q written as $p = 3a^2$ and $q = a^2b$. That is, he started with the cubic

$$x^3 = 3a^2x + a^2b, \text{ with } a = \sqrt{\frac{p}{3}} \text{ and } b = \frac{3q}{p}.$$

Then, he used the trigonometric identity

$$\cos^3(\theta) = \frac{3}{4}\cos(\theta) + \frac{1}{4}\cos(3\theta).$$

If you don't recall this identity, just accept it for now—I will derive it for you in chapter 3 using complex numbers. Viète's next step was to suppose that one can always find a θ such that $x = 2a\cos(\theta)$. I'll now show you that this supposition is in fact true by actually calculating the required value of θ. From the supposition we have $\cos(\theta) = x/2a$, and if this is substituted into the above trigonometric identity then you can quickly show that $x^3 = 3a^2x + 2a^3\cos(3\theta)$. But this is just the cubic we are trying to solve if we write $2a^3\cos(3\theta) = a^2b$. That is,

$$\theta = \frac{1}{3}\cos^{-1}\left(\frac{b}{2a}\right).$$

Inserting this result for θ into $x = 2a\cos(\theta)$ immediately gives us the solution

$$x = 2a\cos\left\{\frac{1}{3}\cos^{-1}\left(\frac{b}{2a}\right)\right\}$$

or, in terms of p and q,

$$x = 2\sqrt{\frac{p}{3}}\cos\left\{\frac{1}{3}\cos^{-1}\left(\frac{3\sqrt{3}q}{2p\sqrt{p}}\right)\right\}.$$

For this x to be real, the argument of the $\cos^{-1}$ must be no greater than one, i.e., $3\sqrt{3q} \leq 2p^{3/2}$. (Later in this book, in chapter 6, I will discuss what happens when the magnitude of the argument in the inverse cosine function *is* greater than one.) But this condition is easily shown to be equivalent to $q^2/4 - p^3/27 \leq 0$, which is precisely the condition that defines the irreducible case. Notice that imaginary quantities do not appear in Viète's formula, unlike the Cardan formula.

Does Viète's formula work? As a test, recall Bombelli's cubic $x^3 = 15x + 4$, with $p = 15$ and $q = 4$. Viète's formula gives

$$x = 2\sqrt{5}\cos\left\{\frac{1}{3}\cos^{-1}\left(\frac{12\sqrt{3}}{30\sqrt{15}}\right)\right\}.$$

This rather fearsome-looking expression is easily run through a hand calculator to give $x = 4$, which is correct. This root is found by taking $\cos^{-1}(12\sqrt{3}/30\sqrt{15}) = 79.695°$. But a quick sketch of the cosine function will show that the angles 280.305° and 439.695° are equally valid. Evaluating x for these two angles will give the other real roots -0.268 and -3.732, i.e., $-2 \pm \sqrt{3}$. Viète himself, however, paid no attention to negative roots. And for another quick check, consider the special case when $q = 0$. Then, $x^3 - px = 0$ which by inspection has the three real roots $x = 0$, $x = \pm\sqrt{p}$. That is, $x = \sqrt{p}$ is the one positive root. Viète's formula gives, for $q = 0$,

$$x = 2\sqrt{\frac{p}{3}}\cos\left\{\frac{1}{3}\cos^{-1}(0)\right\} = 2\sqrt{\frac{p}{3}}\cos(30°)$$

since $\cos^{-1}(0) = 90°$. But $(2/\sqrt{3})\cos(30°) = 1$ and so Viète's formula does give $x = \sqrt{p}$. And since $\cos^{-1}(0) = 270°$ (and 450°), too, you can easily verify that the formula gives the $x = 0$ and $x = -\sqrt{p}$ roots, as well. Technically, this is not an irreducible cubic, but Viète's formula still works. Notice that the roots in these two specific cases satisfy the last statement made in box 1.2.

Viète knew very well the level at which his analytical skills operated. As he himself wrote of his mathematics, it was "not alchemist's gold, soon to go up in smoke, but the true metal, dug out from the mines where dragons are standing watch." Viète was not a man with any false modesty. If his solution had been found a century earlier, would Cardan have worried much over the imaginaries that appeared in his formula? Would Bombelli have been motivated to find the "realness" of the complex expressions that appear in the formal solution to the irreducible cubic? It is interesting to speculate about how the history of mathematics might have been different if some genius had beaten Viète to his discovery. But there was no such genius, and it was Bombelli's glory to unlock the final secret of the cubic.

Bombelli's insight into the nature of the Cardan formula in the irreducible case broke the mental logjam concerning $\sqrt{-1}$. With his work, it became clear that manipulating $\sqrt{-1}$ using the ordinary rules of arithmetic leads to perfectly correct results. Much of the mystery, the near-mystical aura, of $\sqrt{-1}$ was cleared away with Bombelli's analyses. There did remain one last intellectual hurdle to leap, however, that of determining the *physical* meaning of $\sqrt{-1}$ (and that will be the topic of the next two chapters), but Bombelli's work had unlocked what had seemed to be an unpassable barrier.

1.7 A Curious Rediscovery

There is one last curious episode concerning the Cardan formula that I want to tell you about. About one hundred years after Bombelli explained how the Cardan formula works in all cases, including the irreducible case where all roots are real, the young Gottfried Leibniz (1646–1716) somehow became convinced the issue was still open. This is all the more remarkable because Leibniz is known to have studied Bombelli's *Algebra,* and yet he thought there was still something left to add to the Cardan formula. Leibniz was a genius, but this occurred at about age twenty-five, when, as one historian put it, "Leibniz had but little of any competence in what was then modern mathematics. Such firsthand knowledge as he had was mostly Greek."[7]

Leibniz had, at that time, just met the great Dutch physicist and mathematician Christian Huygens (1629–95), with whom he began a lifelong correspondence. In a letter written sometime between 1673 and 1675 to Huygens,[8] he began to rehash what Bombelli had done so long before. In this letter he communicated his famous (if anticlimactic) result

$$\sqrt{1+\sqrt{-3}}+\sqrt{1-\sqrt{-3}}=\sqrt{6},$$

of which Leibniz later declared, "I do not remember to have noted a more singular and paradoxical fact in all analysis; for I think I am the first one to have reduced irrational roots, imaginary in form, to real values. . . ." Of course, it was Bombelli who was the first, by a century.

When the imaginary number $\sqrt{-1}$ is first introduced to high school students it is common to read something like the following (which, actually, I've taken from a college level textbook[9]): "The real equation $x^2 + 1 = 0$ led to the invention of i (and also $-i$) in the first place. That was declared to be the solution and the case was closed." Well, of course, this is simple to read and easy to remember but, as you now know, it is also not true. When the early mathematicians ran into $x^2 + 1 = 0$ and other such quadratics they simply shut their eyes and called them "impossible." They certainly did not invent a solution for them. The breakthrough for $\sqrt{-1}$ came not from quadratic equations, but rather from cubics which clearly had real solutions but for which the Cardan formula produced formal answers with imaginary components. The basis for the breakthrough was in a clearer-than-before understanding of the idea of the conjugate of a complex number. Before continuing with Leibniz, then, let me show you a pretty use of the complex conjugate.

Consider the following statement, easily shown to be correct with a little arithmetic on the back of an envelope:

$$(2^2 + 3^2)(4^2 + 5^2) = 533 = 7^2 + 22^2 = 23^2 + 2^2.$$

And this one, which is only just a bit more trouble to verify:

$$(17^2 + 19^2)(13^2 + 15^2) = 256{,}100 = 64^2 + 502^2 = 8^2 + 506^2.$$

What is going on here?

These are two examples of a general theorem that says the product of two sums of two squares of integers is always expressible, in two different ways, as the sum of two squares of integers. That is, given integers a, b, c, and d, we can always find two pairs of positive integers u and v such that

$$(a^2 + b^2)(c^2 + d^2) = u^2 + v^2.$$

Therefore, says this theorem, it must be true that there are two integer solutions to, for example,

$$(89^2 + 101^2)(111^2 + 133^2) = 543{,}841{,}220 = u^2 + v^2.$$

Can you see what u and v are? Probably not. With complex numbers, and the concept of the complex conjugate, however, it is easy to analyze this problem. Here is how to do it.

Factoring the above general statement of the theorem to be proved, we have

$$[(a + ib)(a - ib)][(c + id)(c - id)] = [(a + ib)(c + id)][(a - ib)(c - id)].$$

Since the quantities in the right-hand brackets are conjugates, we can write the right-hand side as $(u + iv)(u - iv)$. That is,

$$u + iv = (a + ib)(c + id) = (ac - bd) + i(bc + ad)$$

and so

$$u = |ac - bd| \text{ and } v = bc + ad.$$

But this isn't the only possible solution. We can also write the factored expression as

$$[(a + ib)(c - id)][(a - ib)(c + id)] = [u + iv][u - iv]$$

and so a second solution is

$$u + iv = (a + ib)(c - id) = (ac + bd) + i(bc - ad)$$

or,

$$u = ac + bd \text{ and } v = |bc - ad|.$$

These results prove the theorem by actually constructing formulas for u and v, and in particular they tell us that

$$(89^2 + 101^2)(111^2 + 133^2) = 3{,}554^2 + 23{,}048^2 = 626^2 + 23{,}312^2.$$

This problem is quite old (it was known to Diophantus), and a discussion of it, one *not* using complex numbers, can be found in the 1225 book *Liber quadratorum* (*The Book of Squares*[10]) by the medieval Italian mathematician Leonardo Pisano (circa 1170–1250), i.e., Leonardo of Pisa, a town best known today for its famous leaning tower. Leibniz, no doubt, would have found the concept of the complex conjugate to be just what was needed to explain his "paradoxical fact."

As Leibniz expressed his confusion, "I did not understand how . . . a quantity could be real, when imaginary or impossible numbers were used to express it." He found this so astonishing that after his death, among some unpublished papers, several such expressions were found, as if he had calculated them endlessly. For example, solving the cubics $x^3 - 13x - 12 = 0$ and $x^3 - 48x - 72 = 0$, respectively, led him to the additional discoveries that

$$\sqrt[3]{6+\sqrt{-\frac{1225}{27}}}+\sqrt[3]{6-\sqrt{-\frac{1225}{27}}}=4$$

and

$$\sqrt[3]{-36+\sqrt{-2800}}+\sqrt[3]{-36-\sqrt{-2800}}=6.$$

The realness of the literally complex expressions on the left would, today, be considered by a good high school algebra student to be trivially obvious. Such has been the progress in mathematics in understanding $\sqrt{-1}$. Indeed, using the conjugate concept, we know today that a plot of *any* function $f(x)$ contains, in its very geometry, *all* the roots to the equation $f(x) = 0$, real *and* complex. Let me conclude this chapter by showing you how this is so, in particular, for quadratics and for cubics.

1.8 How to Find Complex Roots with a Ruler

When an nth-degree polynomial $y = f(x)$ with real coefficients is plotted, the geometrical interpretation is that the plot will cross the x-axis once for each real root of the equation $f(x) = 0$. Crossing the x-axis is, in fact, where the zero on the right-hand side comes from. If there are fewer than n crossings, say $m < n$, then the interpretation is that there are m real roots, given by the crossings, and $n - m$ complex roots. The value of $n - m$ is an even number since, as shown in appendix A, complex roots always appear as conjugate *pairs*. This is not to say, however, that there is no physical signature in the plot for the complex roots. The signature for the real roots, the x-axis crossings, is simple and direct, but if you are willing to do just a bit more work you can read off the complex roots, too.

First, consider the quadratic $f(x) = ax^2 + bx + c = 0$. The two roots to this equation are either both real or a complex conjugate pair, depending on the algebraic sign of the quantity $b^2 - 4ac$. If this is non-negative, then the roots are real and there are either two x-axis crossings or a touching of the axis (if $b^2 - 4ac = 0$, giving a double root). If $b^2 - 4ac$ is negative then the roots are complex and there are no x-axis crossings, which is the case shown in figure 1.3. Let us assume that this is the case, and that the roots are $p \pm iq$. Then, writing $f(x)$ in factored form,

$$f(x) = a(x - p - iq)(x - p + iq) = a[(x - p)^2 + q^2],$$

it is clear that $f(x) \geq aq^2$ if $a > 0$, and $f(x) \leq aq^2$ if $a < 0$. That is, $f(x)$ takes on its minimum value at $x = p$ if $a > 0$ (as shown in figure 1.3), or its

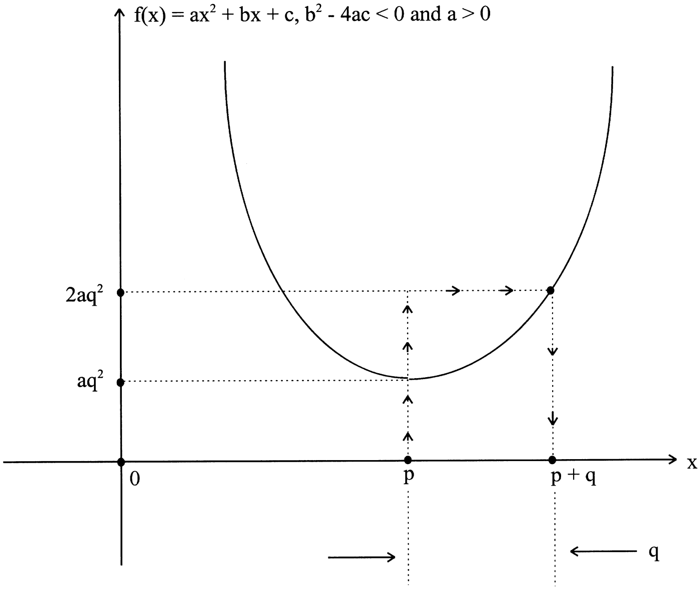

Figure 1.3. A quadratic equation with no real roots.

maximum value at $x = p$ if $a < 0$. We can, therefore, measure p from the plot of $f(x)$ as the x-coordinate of the local extremum.

Next, to measure the value of q from the plot, first measure the y coordinate of the minimum (I'm assuming $a > 0$, but the $a < 0$ case is a trivial variation), i.e., measure aq^2. Then, at $x = p$, first move upward $2aq^2$, then over to the right until you intersect the plot. The x-value of this intersection point (call it $\hat{x}$), when plugged into the quadratic equation, gives

$$f(\hat{x}) = 2aq^2 = a[(\hat{x} - p)^2 + q^2] = a(\hat{x} - p)^2 + aq^2$$

or,

$$aq^2 = a(\hat{x} - p)^2 \text{ or } q = \hat{x} - p.$$

Thus, q can be directly measured off the plot of $f(x)$, as shown in figure 1.3.

Concentrating next on cubics, observe first that there will be either (a) three real roots or (b) one real root and two complex conjugate roots. Be sure you are clear in your mind why all three roots cannot be complex, and why there cannot be two real roots and one complex root. If you're not clear on this, see appendix A. Case (b) is the one of interest for us. Call the real root $x = k$, and the conjugate pair of roots $x = p \pm iq$. Then, we can write $f(x)$ in factored form as

$$y = f(x) = (x - k)(x - p + iq)(x - p - iq)$$

or, expanding and collecting terms, as

$$f(x) = (x - k)(x^2 - 2xp + p^2 + q^2).$$

The plot of a cubic with a single real root, which means one x-axis crossing, will have the general appearance of figure 1.4. Construct the triangle AMT, where A is the intersection point of $y = f(x)$ with the x-axis, T is the point of tangency to $y = f(x)$ of a straight line passing through A, and M is the foot of the perpendicular through T perpendicular to the x-axis. Of course, the real root is $k = OA$.

Now, consider the straight line $y = \lambda(x - k)$, which clearly passes through A as $y = 0$ when $x = k$. Imagine that the slope of this line, λ, is adjusted until it just touches $y = f(x)$, i.e., until it is tangent to $y = f(x)$. This then gives us T, and since the x-value at T is common to both $y = f(x)$ and $y = \lambda(x - k)$, calling this x-value $\hat{x}$ says that

$$\lambda(\hat{x} - k) = (\hat{x} - k)(\hat{x}^2 - 2p\hat{x} + p^2 + q^2).$$

Since $\hat{x} - k \neq 0$, we can divide through both sides of this equation to obtain a quadratic in $\hat{x}$,

$$\lambda = \hat{x}^2 - 2p\hat{x} + p^2 + q^2.$$

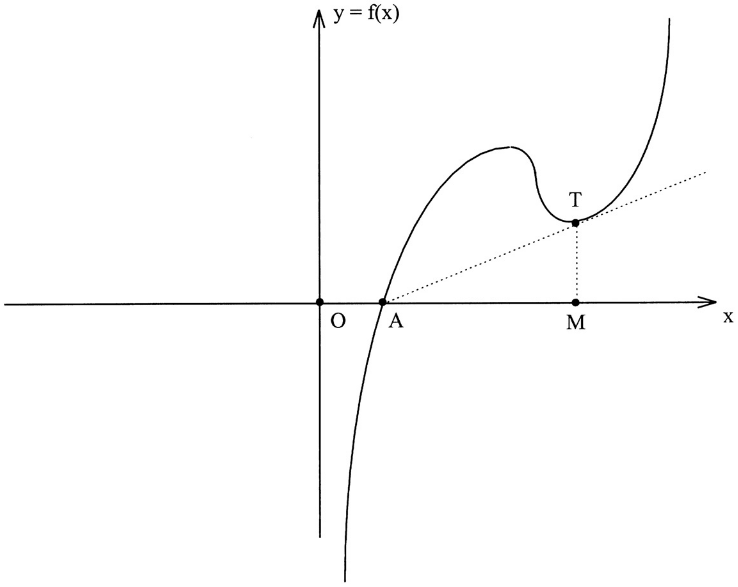

Figure 1.4. A cubic equation with one real root.

In fact, since T is a point of tangency, there must be exactly one value of $\hat{x}$. That is,

$$\hat{x}^2 - 2p\hat{x} + p^2 + q^2 - \lambda = 0$$

must have two equal, or double, roots. Now, in general,

$$\hat{x} = \frac{2p \pm \sqrt{4p^2 - 4(p^2 + q^2 - \lambda)}}{2},$$

and to have double roots the radical must be zero. That is,

$$4p^2 - 4(p^2 + q^2 - \lambda) = 0$$

or, $\lambda = q^2$. That is, the tangent line AT has slope $q^2 = TM/AM$. The value of $\hat{x}$ is, from the general expression for $\hat{x}$, then just $\hat{x} = p = OM$.

So, to find all the roots of the cubic, you need only plot $y = f(x)$ and then:

1. Read off the real root by measuring OA $(= k)$.
2. Place a straightedge at A as a pivot point and swing the edge slowly until it just touches the plotted function (thus "locating" T).
3. Measure TM and AM, and then calculate

$$q = \sqrt{\frac{TM}{AM}}.$$

4. Measure OM to give p.
5. The two imaginary roots are $p + iq$ and $p - iq$.