

14.6 The Fundamental Theorem of Algebra

The fundamental theorem of algebra is the statement that every polynomial equation $p(z) = 0$ has a solution in the complex numbers. As Descartes observed (Section 6.7), a solution $z = a$ implies that $p(z)$ has a factor $z - a$. The quotient $q(z) = p(z)/(z - a)$ is then a polynomial of lower degree; hence if every polynomial equation has a solution, we can also extract a factor from $q(z)$, and if $p(z)$ has degree n , we can go on to factorize $p(z)$ into n linear factors. The existence of such a factorization is of course another way to state the fundamental theorem.

Initially, interest was confined to polynomials $p(z)$ with real coefficients, and in this case d'Alembert (1746) observed that if $z = u + iv$ is a solution of $p(z) = 0$, then so is its conjugate $\bar{z} = u - iv$. Thus the imaginary linear factors of a real $p(z)$ can always be combined in pairs to form real quadratic factors:

$$(z - u - iv)(z - u + iv) = z^2 - 2uz + (u^2 + v^2).$$

This gave another equivalent of the fundamental theorem: each (real) polynomial $p(z)$ can be expressed as a product of real linear and quadratic factors. The theorem was usually stated in this way during the eighteenth century, when its main purpose was to make possible the integration of rational functions (see Section 14.5). This also avoided mention of $\sqrt{-1}$.

It has often been said that attempts to prove the fundamental theorem began with d'Alembert (1746), and that the first satisfactory proof was given by Gauss (1799). This opinion should not be accepted without question, as the source of it is Gauss himself. Gauss (1799) gave a critique of proofs from d'Alembert on, showing that they all had serious weaknesses, then offered a proof of his own. His intention was to convince readers that the new proof was the first valid one, even though it used one unproved assumption (which is discussed further in the next section). The opinion as to which of two incomplete proofs is more convincing can of course change with time, and I believe that Gauss (1799) might be judged differently today. We can now fill the gaps in d'Alembert (1746) by appeal to standard methods and theorems, whereas there is still no easy way to fill the gap in Gauss (1799).

Both proofs depend on the geometric properties of the complex numbers and the concept of continuity for their completion. The basic geometrical insight—that the complex number $x + iy$ can be identified with

the point (x, y) in the plane—mysteriously eluded all mathematicians until the end of the eighteenth century. This was one of the reasons that d'Alembert's proof was unclear, and the use of this insight by Argand (1806) was an important step in d'Alembert's reinstatement. Gauss seems to have had the same insight but concealed its role in his proof, perhaps believing that his contemporaries were not ready to view the complex numbers as a plane.

As for the concept of continuity, neither Gauss nor d'Alembert understood it very well. Gauss (1799) seriously understated the difficulties involved in the unproved step, claiming that “no one, to my knowledge, has ever doubted it. But if anybody desires it, then on another occasion I intend to give a demonstration which will leave no doubt” [translation from Struik (1969), p. 121]. Perhaps to preempt criticism, he gave a second proof [Gauss (1816)], in which the role of continuity was minimized. The second proof is purely algebraic except for the use of a special case of the intermediate value theorem. Gauss assumed that a polynomial function $p(x)$ of a real variable x takes all values between $p(a)$ and $p(b)$ as x runs from a to b . The first to appreciate the importance of continuity for the fundamental theorem of algebra was Bolzano (1817), who proved the continuity of polynomial functions and attempted a proof of the intermediate value theorem. The latter proof was unsatisfactory because Bolzano had no clear concept of real number on which to base it, but it did point in the right direction. When a definition of real numbers emerged in the 1870s (for example, with Dedekind cuts; Section 4.2), Weierstrass (1874) rigorously established the basic properties of continuous functions, such as the intermediate value theorem and extreme value theorem. This completed not only the second proof of Gauss but also the proof of d'Alembert, as we shall see in the next section.

EXERCISES

Complex roots of an equation with real coefficients occur in conjugate pairs because of the fundamental properties of conjugates.

14.6.1 Show directly from the definition $\overline{u + iv} = u - iv$ that

$$\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2} \quad \text{and} \quad \overline{z_1 \cdot z_2} = \overline{z_1} \cdot \overline{z_2}$$

for any complex numbers z_1, z_2 .

14.6.2 Deduce from Exercise 14.6.1 that $p(\bar{z}) = \overline{p(z)}$ for any polynomial $p(z)$ with real coefficients, and hence that the complex roots of $p(z) = 0$ occur in conjugate pairs.

14.7 The Proofs of d'Alembert and Gauss

The key to d'Alembert's proof is a proposition now known as *d'Alembert's lemma*: if $p(z)$ is a nonconstant polynomial function and $p(z_0) \neq 0$, then any neighborhood of z_0 contains a point z_1 such that $|p(z_1)| < |p(z_0)|$.

The proof of this lemma offered by d'Alembert depended on solving the equation $w = p(z)$ for z as a fractional power series in w . As mentioned in Section 9.5, such a solution was claimed by Newton (1671), but it was only made clear and rigorous by Puiseux (1850). Thus d'Alembert's argument did not stand on solid ground, and in any case it was unnecessarily complicated.

A simple elementary proof of d'Alembert's lemma was given by Argand (1806). Argand was one of the co-discoverers of the geometric representation of complex numbers [probably the first was Wessel (1797), but his work remained almost unknown for 100 years]), and he offered the following proof as an illustration of the effectiveness of the representation.

The value of $p(z_0) = x_0 + iy_0$ is interpreted as the point (x_0, y_0) in the plane, so that $|p(z_0)|$ is the distance of (x_0, y_0) from the origin. We wish to find a Δz such that $p(z_0 + \Delta z)$ is nearer to the origin than $p(z_0)$. If

$$p(z) = a_0 z^n + a_1 z^{n-1} + \cdots + a_n,$$

then

$$\begin{aligned} p(z_0 + \Delta z) &= a_0(z_0 + \Delta z)^n + a_1(z_0 + \Delta z)^{n-1} + \cdots + a_n \\ &= a_0 z_0^n + a_1 z_0^{n-1} + \cdots + a_n + A_1 \Delta z + A_2 (\Delta z)^2 + \cdots + A_n (\Delta z)^n \\ &\quad \text{for some constants } A_i \text{ depending on } z_0, \text{ not all zero,} \\ &\quad \text{because } p \text{ is not constant} \\ &= p(z_0) + A \Delta z + \varepsilon, \end{aligned}$$

where $A = A_i (\Delta z)^i$ contains the first nonzero A_i and $|\varepsilon|$ is small compared with $|A \Delta z|$ when $|\Delta z|$ is small (because ε contains higher powers of Δz). It is then clear (Figure 14.5) that by choosing the direction of Δz so that $A \Delta z$ is opposite in direction to $p(z_0)$, we get $|p(z_0 + \Delta z)| < |p(z_0)|$. This completes the proof of d'Alembert's lemma.

To complete the proof of the fundamental theorem of algebra, take an arbitrary polynomial p and consider the continuous function $|p(z)|$. Since $p(z) \simeq a_0 z^n$ for $|z|$ large, $|p(z)|$ increases with $|z|$ outside a sufficiently large circle $|z| = R$. We now get a z for which $|p(z)| = 0$ from the extreme value

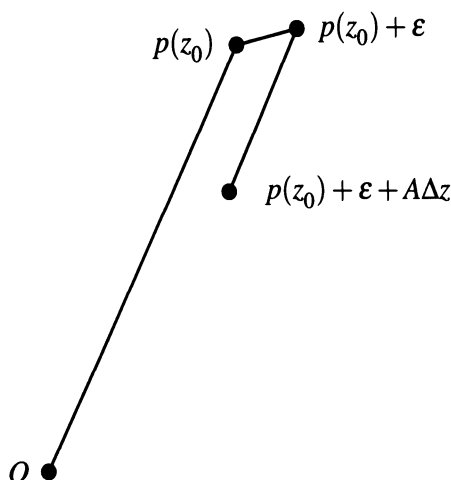


Figure 14.5: Construction for d'Alembert's lemma

theorem of Weierstrass (1874); a continuous function on a closed bounded set assumes maximum and minimum values. By this theorem, $|p(z)|$ takes a minimum value for $|z| \leq R$. The minimum is ≥ 0 by definition, and if it is > 0 we get a contradiction by d'Alembert's lemma: either a point z with $|z| \leq R$ where $|p(z)|$ takes a value less than its minimum or a point z with $|z| > R$ where $|p(z)|$ is less than its values on $|z| = R$. Thus there is a point z where $|p(z)| = 0$ and hence $p(z) = 0$.

The proof of Gauss also used the fact that $p(z)$ behaves like its highest-degree term $a_0 z^n$ for $|z|$ large and likewise relied on a continuity argument to show that $p(z) = 0$ inside some circle $|z| = R$. Gauss considered the real and imaginary parts of $p(z)$, $\text{Re}[p(z)]$, and $\text{Im}[p(z)]$ and investigated the curves

$$\text{Re}[p(z)] = 0 \quad \text{and} \quad \text{Im}[p(z)] = 0.$$

[These are easily seen to be algebraic curves $p_1(x, y) = 0$ and $p_2(x, y) = 0$ by expanding the powers $z^k = (x + iy)^k$ and collecting real and imaginary terms.] His aim was to find a point where these curves meet, because at such a point

$$0 = \text{Re}[p(z)] = \text{Im}[p(z)] = p(z).$$

For $|z|$ large, the curves are close to the curves $\text{Re}(a_0 z^n)$ and $\text{Im}(a_0 z^n) = 0$, which are families of straight lines through the origin. Moreover, the lines where $\text{Re}(a_0 z^n) = 0$ alternate with those where $\text{Im}(a_0 z^n) = 0$ as one makes

a circuit around the origin. For example, Figure 14.6 shows $\operatorname{Re}(z^2) = 0$ and $\operatorname{Im}(z^2) = 0$ as alternate solid and dashed lines. It follows that the curves $\operatorname{Re}[p(z)] = 0$ and $\operatorname{Im}[p(z)] = 0$ meet a sufficiently large circle $|z| = R$ *alternately*. Up to this point the argument is comparable to d'Alembert's lemma, and it can be made just as rigorous.

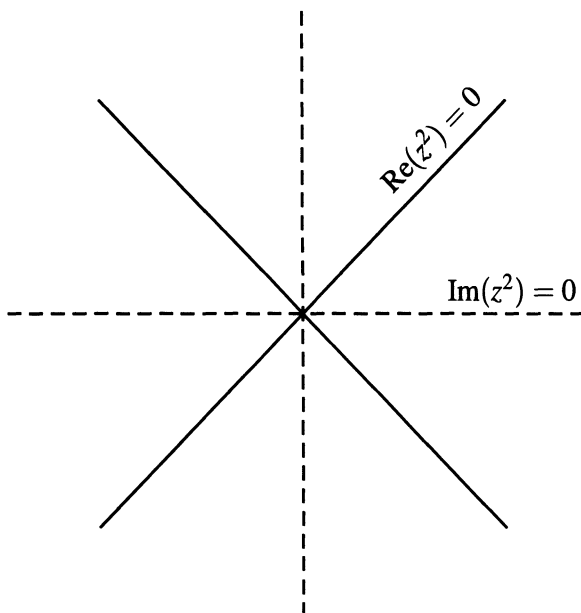


Figure 14.6: Lines for the Gauss proof

To complete this proof we only have to show that the curves meet inside the circle, and this is the step Gauss thought nobody could doubt. He assumed that the separate pieces of the algebraic curve $\operatorname{Re}[p(z)] = 0$ outside the circle $|z| = R$ would join inside the circle, as would the separate pieces of $\operatorname{Im}[p(z)] = 0$. Then, since the pieces of $\operatorname{Re}[p(z)] = 0$ alternate with those of $\operatorname{Im}[p(z)] = 0$ on $|z| = R$, it would be “patently absurd” for their connecting pieces inside the circle not to meet. One has only to visualize a situation like that seen in Figure 14.7 to feel sure that Gauss was right. However, the existence of the connecting pieces is extremely hard to prove (and proving that they meet is not trivial either, being at least as hard as the intermediate value theorem). The first proof was given by Ostrowski (1920).

From our present perspective, d'Alembert's route to the fundamental theorem of algebra seems basically easy because it proceeds through general properties of continuous functions. The route of Gauss, although ap-

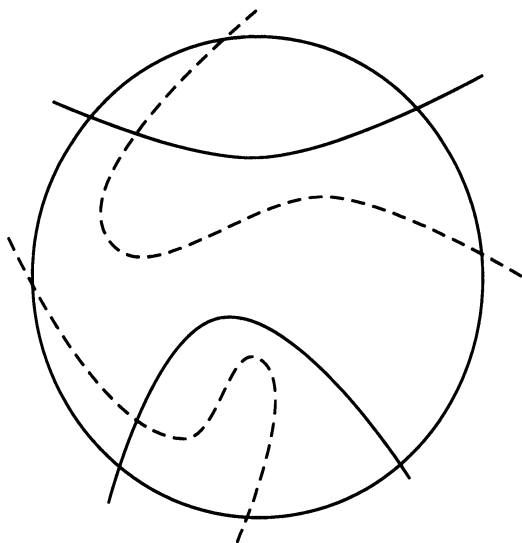


Figure 14.7: Curves for the Gauss proof

peering equally easy from a distance, goes through the still-unfamiliar territory of real algebraic curves. The intersections of real algebraic curves are harder to understand than the intersections of complex algebraic curves, and in retrospect they are harder to understand than the fundamental theorem of algebra. Indeed, as we shall see in the next chapter, the fundamental theorem gives us Bézout's theorem, which in turn settles the problem of counting the intersections of complex algebraic curves.

EXERCISES

The expression in d'Alembert's lemma for $p(z_0 + \Delta z)$ is an instance of *Taylor's series*, previously discussed in Section 10.2. When the function is a polynomial p , as here, its Taylor series is finite because p has only finitely many nonzero derivatives.

14.7.1 Show that $A_1 = na_0z_0^{n-1} + (n-1)a_1z_0^{n-2} + \cdots + a_{n-1}$ and that the latter expression is $p'(z_0)$.

14.7.2 Show that $A_2 = \frac{n(n-1)}{2}a_0z_0^{n-2} + \frac{(n-1)(n-2)}{2}a_1z_0^{n-3} + \cdots + a_{n-2}$ and that the latter expression is $p''(z_0)/2$.

14.7.3 Using the binomial theorem, show that $A_i = p^{(i)}(z_0)/i!$, and hence that

$$p(z_0 + \Delta z) = a_0z_0^n + a_1z_0^{n-1} + \cdots + a_n + A_1\Delta z + A_2(\Delta z)^2 + \cdots + A_n(\Delta z)^n$$

is an instance of the Taylor series formula.

14.8 Biographical Notes: d'Alembert

Jean le Rond d'Alembert (Figure 14.8) was born in Paris in 1717 and died there in 1783. He was the illegitimate son of the Chevalier Destouches-Canon, a cavalry officer, and salon hostess Madame de Tencin. His mother abandoned him at birth near the church of St. Jean-le-Rond in the cloisters of Notre Dame, and so he was christened Jean le Rond, following the custom for foundlings. He was subsequently located by his father, who found a home for him with a glazier named Rousseau and his wife. The name d'Alembert came later, for reasons that are unclear.

The Rousseaus must have been devoted foster parents, as d'Alembert lived with them until 1765. He received an annuity from his father, who also arranged for him to be educated at the Jansenist Collège de Quatre-Nations in Paris. There he received a good grounding in mathematics and developed a permanent distaste for theology. After brief studies in law and medicine he turned to mathematics in 1739.

In that year he began sending communications to the Académie des Sciences, and his ambition and talent rapidly carried him to fame. He became a member of the Académie in 1741 and published his best-known work, the *Traité de dynamique*, in 1743. Having struggled to the top from humble beginnings, d'Alembert did not want to lose his position. Once in the Académie, his struggle was to stay ahead of his rivals. Whether by accident or inborn competitiveness, d'Alembert always seemed to be working on the same problems as other top mathematicians—initially Clairaut, later Daniel Bernoulli and Euler. He was always fearful of losing priority and fell into a cycle of hasty publication followed by controversy over the meaning and significance of his work. Despite the fact that he was an excellent writer (elected to the Académie Française in 1754), his mathematics was almost always poorly presented. Many of his best ideas were not understood until Euler overhauled them and gave them masterly expositions. Since Euler often did this without giving credit, d'Alembert was understandably furious, but he squandered his energy in quarreling instead of giving his own work the exposition it deserved.

Another reason for d'Alembert's lack of attention to his mathematics was his involvement in the broader intellectual life of his time. When d'Alembert came on the scene in the 1740s, mathematics was enjoying great prestige in philosophical circles, largely because of Newton's success in explaining the motions of the planets. Mathematics was a model of



Figure 14.8: Jean Baptiste le Rond d'Alembert