C. F. GAUSS'S PROOFS OF THE FUNDAMENTAL THEOREM OF ALGEBRA

HAREL CAIN, I.D. 036135184

1. Historical and mathematical exposition

The German mathematician Carl Friedrich Gauss (1777–1855), commonly regarded as one of the greatest mathematicians of all time, and who was even called *princeps mathematicorum*, made many outstanding tributes to diverse branches of mathematics, physics and geodesics. Among his many important achievements, one that Gauss gave special attention to was the proof of what is, somewhat mistakenly, known as the Fundamental Theorem of Algebra, a basic result of analysis the statement of which is simple enough to be understood by any layman with knowledge of high school mathematics.

The **Fundamental Theorem of Algebra** (hereafter, **FTA**) states that

> Every polynomial $f(x) \in \mathbb{C}[x]$ with complex coefficients can be factored into linear factors over the complex numbers.

Gauss used an alternative formulation that avoided the notion of complex numbers, stating that

> Every polynomial $f(x) \in \mathbb{R}[x]$ with real coefficients can be factored into linear and quadratic factors.

This formulation is equivalent to the first one, because if $f \in \mathbb{C}[x]$, then $g := f\bar{f} \in \mathbb{R}[x]$, and a factorization of $g(x)$ would yield a factorization of $f(x)$ as well, because every root of $f(x)$ would also be a root of $g(x)$.

The first attempt to prove the **FTA** is usually credited to d'Alembert [2], but Carl Friedrich Gauss is often considered to have offered its first satisfactory proof. Gauss returned to the elusive proof of this theorem time and time again. During his lifetime, he offered not less than four different proofs to the theorem, covering a timespan of fifty years spanning his entire adult life. His first proof materialized in October 1797 and was published in his doctoral dissertation [4] in 1799, aged 22. It contains critique of previous attempts to prove the theorem by d'Alembert, Euler, Fontenex and others. Gauss viewed these attempts as unsatisfactory, because they presupposed that the roots of the polynomial can be obtained as complex numbers. As we shall see, however, his proof of 1799, using geometric notions, had its own missing parts left to be rigorously proved (which were only completely proved in 1920!), and Gauss was not satisfied with it either. 1816 saw the publication by Gauss [5] of two more proofs to the **FTA**. The first of them was technical and almost strictly algebraic; the second — an easier proof using analysis. In 1849, just a few years before his death, Gauss offered a fourth proof that bore similarity to his first.

This essay will try to critically outline the proofs of the first three proofs Gauss gave to the **FTA**, with most emphasis given to the first and third proofs. Some attention will also be given to observing the similarities between the Gaussian proof and the earlier attempt by d'Alembert using *d'Alembert's lemma* (the principle of maximality), normally taught, in more modern guise, to undergraduate mathematics students. Wherever possible, and based on the limited translations of the proofs from Latin into English, the essay will try to remain faithful to the mathematical notation used by Gauss.

2. The first proof: a sketch

Gauss started with a **real** polynomial

$$X = x^m + Ax^{m-1} + Bx^{m-2} + \cdots + Lx + M$$

treating x as an "unbestimmte Größe", an indeterminate. A real linear factor can be written as $x \pm r$, with $r \geq 0$. An irreducible quadratic factor over the reals can be written as $x^2 - 2xr\cos\phi + r^2$ again with $r > 0$, the roots of which are $r(\cos\phi \pm i\sin\phi)$.

By substituting $x = r(\cos\phi + i\sin\phi)$ into the polynomial and separating into real and imaginary parts one can form a pair of expressions

$$U = r^m \cos m\phi + Ar^{m-1}\cos(m-1)\phi + \cdots Lr\cos\phi + M$$

$$T = r^n \sin m\phi + Ar^{m-1}\sin(m-1)\phi + \cdots Lr\sin\phi$$

Gauss now notes that if (r, ϕ) satisfy $T = 0$ and $U = 0$ simultaneously, then the polynomial X would be divisible by $x \pm r$ or $x^2 - 2xr\cos\phi + r^2$. He proves this directly, noting that a similar proof with imaginaries was given by Euler [3].

Gauss regards $T = 0$ and $U = 0$ as algebraic curves of order m given in the polar coordinates r and ϕ, drawn on an orthogonal plane with coordinates $(r\cos\phi, r\sin\phi)$. To see that the curves are algebraic, one just has to use standard trigonometric formulae to express U and T as algebraic expressions in $r\cos\phi$ and $r\sin\phi$. To prove the **FTA**, Gauss wants to prove that there exists a point of intersection between the two curves $T = 0$ and $U = 0$. Using his earlier lemma, such an intersection would be a root of the polynomial X, because if $U(r, \phi) = Re(X(x)) = 0$ and $T(r, \phi) = Im(X(x)) = 0$ then $X(x) = 0$.

To look for intersections of the two curves, Gauss now studies their intersections within a circle of radius R, and proves

> For a sufficiently large radius R there are exactly $2m$ intersections of the circle with $T = 0$ and $2m$ intersections with $U = 0$, and every point of intersection of the second kind lies between two of the first kind.

This result is proved with complete rigour, but only after the underlying idea is presented intuitively: as R tends to infinity, the curves $T = 0$ and $U = 0$ approach the curves $Re(x^m) = 0$ and $Im(x^m) = 0$, which are straight lines through the origin. Moreover, the lines where $Re(x^m) = 0$ *alternate* with those where $Im(x^m) = 0$. Fearing that "the readers might be offended by infinitely large quantities", Gauss then gives a rigorous proof of this result, not using the concept of limit. In the next step, Gauss notes that "it can easily be seen" that that the $4m$ points of intersection of the circle vary very little when R is slightly changed. In modern terms, we would say they are continuous functions of R, but the concept of continuity wasn't fully

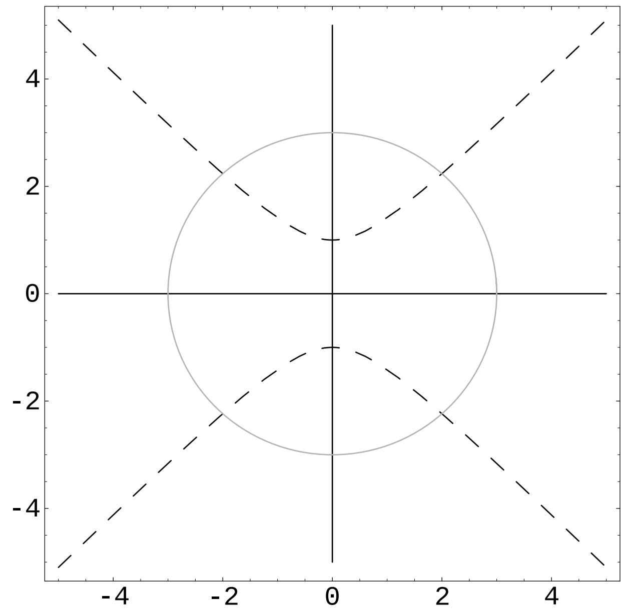

FIGURE 1. An example for $X = x^2 + 1$. The curve $T = 2r^2 \sin\phi\cos\phi = 0$ coincides with the axes, $U = 1 + r^2\cos\phi^2 - r^2\sin\phi^2 = 0$ is the dotted curve, and the $4m = 8$ intersections with the circle alternate as in the proof.

developed. Gauss now reaches the core of his proof: to show that the two curves intersect inside the circle. For this conclusion he gives an intuitive, geometrical proof: he numbers the intersections with the circle sequentially, starting with 0 for the negative x-axis of the plane (which is always part of the solution to $T = 0$) giving odd numbers to the intersections of $U = 0$ with the circle, and even numbers to the intersections of $T = 0$. He now claims that "if a branch of an algebraic curve enters a limited space, it necessarily has to leave it again". In a footnote he adds:

> "It seems to be well demonstrated that an algebraic curve neither ends abruptly (as it happens in the transcendental curve $y = 1/\log x$), nor lose itself after an infinite number of windings in a point (like a logarithmic spiral). As far as I know nobody has ever doubted this, but if anybody requires it, I take it on me to present, on another occasion, an indubitable proof." (quoted in [10])

If this remark is to be accepted, then every odd-numbered intersection point on the circle is connected with another odd-numbered point by a branch of the curve $U = 0$, and similarly for even-numbered points and branches of the curve $T = 0$. But then, however complicated these connections may be, one can show that a point of intersection exists, in the following way. Suppose no such intersection exists. The point 0 is connected with point $2m$ through the x-axis, which is always a branch of $T = 0$. Then 1 cannot be connected with any point on the other side of the axis, so it must be connected with an odd point n, $n < 2m$. In the same manner, 2 is connected with an even $n' < n$. Note that $n' - 2$ is even. Continuing this way, one ends with an intersection point h connected with $h+2$. But then the branch entering the circle at $h+1$ must intersect the branch connecting h and $h+2$, contrary to our hypothesis. Thus an intersection point for $Re(X) = T = 0$ and $Im(X) = U = 0$ exists, concluding the proof.

This outline of the proof shows that Gauss's first proof of the **FTA** is based on assumptions about the branches of algebraic curves, which might appear plausible to geometric intuition, but are left without any rigorous proof by Gauss. It took until 1920 for Alexander Ostrowski to show [7] that all assumptions made by Gauss can be fully justified.

2.1. **Comparing the proofs of Gauss and d'Alembert.** As mentioned above, it is d'Alembert who in 1746 first made an attempt [2] to prove the **FTA**. D'Alembert's proof shares some notions with Gauss's proof of 1799. The key to it is a proposition now known as as *d'Alembert's lemma*: if $p(z)$ is a polynomial function and $p(z_0) \neq 0$, then any neighborhood of z_0 contains a point z_1 such that $|p(z_1)| < |p(z_0)|$. To prove this lemma, d'Alembert used the method of fractional power series claimed by Newton back in 1671 [6], but made complete and rigorous only by Puiseux in 1850. Argand offered an elementary proof [1] of the lemma in 1806. It is clear, then, that in 1746 d'Alembert had a missing part in his proof, as did Gauss in his first proof.

However, if one is to accept *d'Alembert's lemma*, together with Weierstrass's extreme value theorem [11] (1874), then the **FTA** follows easily, in a manner resembling Gauss's first proof. Because for large $|z|$, $|p(z)|$ is dominated by the leading term in the polynomial, it increases outside a sufficiently large circle $|z| = R$. The objective is to prove that inside the circle, there is z_0 with $|p(z_0)| = 0$. But then by Weierstrass's extreme value theorem, $|p(z)|$ has a minimum M inside the circle. If the minimum is in the circle's interior, then d'Alembert's lemma leads to a contradiction. If it is on the boundary, then there is a point outside the circle with a smaller $|p(z)|$ value, also a contradiction. We can see that both proofs work by confining a circle inside of which they seek a point where $p(z)$ vanishes.

3. The second proof: a sketch

In 1816 Gauss published a second proof of the **FTA**. This proof was purely algebraic and very technical in nature. It presupposes only that every real equation of odd degree has a real root and that every quadratic equation with complex coefficients has two complex roots. This essay will give a very rough outline of the main steps of the proof.

Gauss starts with a *real* polynomial Y of degree m. Suppose for a moment that Y can be factored into linear factors

$$Y = (x-a)(x-b)(x-c)\dots$$

in some extension field. Every pair of roots, for example a and b, can be combined in a linear combination

$$(a+b)t - ab$$

in a new indeterminate t. Enumerating over all possible pairs, one can form $m' = \binom{m}{2}$ different linear combinations of this sort. One can construct an auxiliary equation of degree m', the roots of which would be the linear combinations $(a+b)t - ab$, with a and b ranging over all the roots of Y, and t specialized to leave different linear functions of this form different.

As soon as one root of the auxiliary equation is known, then $a+b$ and ab are known, and a and b, two roots of the original equation, can be computed as complex numbers by extracting a square root.

Now m can always be written as $m = 2^\mu k$ with k odd, so that $m' = \binom{m}{2} = 2^{\mu-1}k'$. Repeating the process of building an auxiliary equation, one can thus arrive at an equation of odd degree. The coefficients of that equation would be symmetric functions of $a, b, \ldots$ with real coefficients, so they would also be real numbers. Since the degree of this auxiliary real polynomial is odd, it has at least one real root. Climbing back through the series of auxiliary polynomials back to the original polynomial, one can compute at least one complex root of the original equation, as we wanted.

In this form, the proof works if the we know that the polynomial Y has m roots in some extension field of $\mathbb{R}$. Such an extension field can be formed by Kronecker's standard method of "symbolic adjunction", a basic technique of field extension. Gauss, however, doesn't take this road and constructs his auxiliary polynomials without assuming the existence of the roots.

4. The third proof: a sketch

The third proof by Gauss, published in 1816, is much simpler than the second. Starting with the same polynomial X, Gauss again substitutes $x = r(\cos\phi + i\sin\phi)$ and separates into real and imaginary parts, naming them t and u as opposed to U and T (in this order). He then introduces their derivatives with respect to ϕ

$$\begin{aligned} t' &= mr^m \cos m\phi + (m-1)Ar^{m-1}\cos(m-1)\phi + \cdots + Lr\cos\phi \\ u' &= mr^m \sin m\phi + (m-1)Ar^{m-1}\sin(m-1)\phi + \cdots + Lr\sin\phi \end{aligned}$$

Observing that the main term of $tt' + uu'$ is $mr^{2m}(\cos^2 m\phi + \sin^2 m\phi) = mr^{2m}$, Gauss concludes that $tt' + uu'$ is positive for a large enough value of r, which he calls R. He further introduces the second derivatives with respect to ϕ, called $-t''$ and u'':

$$\begin{aligned} t'' &= m^2 r^m \cos m\phi + \cdots + Lr\cos\phi \\ u'' &= m^2 r^m \sin m\phi + \cdots + Lr\sin\phi \end{aligned}$$

As in the first proof, the goal is to show there is a point in the $(r\cos\phi, r\sin\phi)$ plane where $t = 0$ and $u = 0$ meet, which would imply the existence of a complex root for the polynomial X. Suppose then that no point with $t = u = 0$ exists, so that $t^2 + u^2$ is always non-zero, and thus the function

$$y = \frac{(t^2+u^2)(tt''+uu'') + (tu'-ut')^2 - (tt'+uu')^2}{r(t^2+u^2)^2}$$

is finite everywhere. Next, Gauss considers the double integral

$$\Omega = \int_0^{360^\circ} \int_0^R y\,dr\,d\phi$$

The order of integration is immaterial. By differentiation, one observes that

$$\int y\,d\phi = \frac{tu' - ut'}{r(t^2+u^2)}$$

But the function on the right has the same value for $\phi = 0$ as for $\phi = 360^\circ$ (because it appears only in sines and cosines). So that integrating the indefinite integral from $\phi = 0$ to $\phi = 360°$ yields zero, and therefore also $\Omega = 0$. However, integrating first with respect to r, one obtains the indefinite integral

$$\int y dr = \frac{tt' + uu'}{t^2 + u^2}$$

Looking back at the definitions of t' and u', for $r = 0$ this expression is zero. But for $r = R$ it is positive as was argumented above. So that this indefinite integral, integrated from 0 to R, is positive, and therefore Ω is positive too, contrary to what we saw before. Therefore the hypothesis that t and u are never both zero leads to a contradiction, and the proof is completed.

4.1. **Looking behind the proof and into Gauss's mind.** Gauss's third proof is easily followed step by step. But how could it be obtained? How was the complicated function y constructed? One can make some reasonable guesses.

One can think of the polynomial X as defining a mapping $X(x) = X(r, \phi) = t+iu$ from the x plane to the X plane, where polar coordinates can also be introduced:

$$X = s(\cos\beta + i \sin\beta)$$

so that $\beta = \arctan \frac{u}{t}$. Differentiating β with respect to ϕ and to r one obtains

$$\frac{\partial\beta}{\partial\phi} = \frac{tt' + uu'}{t^2 + u^2} = V$$
$$\frac{\partial\beta}{\partial r} = \frac{tu' - ut'}{r(t^2 + u^2)} = U$$

If one differentiates again, one obtains

$$\frac{\partial U}{\partial r} = \frac{\partial V}{\partial \phi} = y$$

One might conclude, then that, Gauss arrived at the complicated function y by using $\frac{\partial U}{\partial r}$ and $\frac{\partial V}{\partial \phi}$, which he knew would be the same because U and V are the derivatives of the same function β with respect to ϕ and to r, respectively. The angle β is not uniquely defined, but only modulo 2π. It is likely that Gauss wanted to avoid the use of multi-value functions like β, and therefore chose to work only with its derivatives U, V and y in his proof, writing the integral $\int U d\phi = \int V dr$ as the double integral $\iint y dr d\phi$ and interchanging the order of integrations.

5. Concluding remarks: judging the proofs today

In retrospect, a complete and rigorous proof of the **FTA** eluded mathematicians for quite a long time. As we saw above, both d'Alembert's 1746 proof and Gauss's 1799 proof were based on unproved propositions which took a few more generations to settle. Both were premature, because they were based on geometric properties of the complex numbers and on the concept of continuity. That the complex numbers can be presented as a plane mysteriously escaped mathematicians until the end of the eighteenth century. Only with Argand in 1806 did this insight gain some ground and as a result d'Alembert's proof became more accepted. Gauss seems to have understood the geometric representation of complex numbers, but might have thought that his contemporaries were not ready for it.

The concept of continuity waited much longer. Bolzano proved the continuity of polynomials in 1817, and Weierstrass's theorems were properly established only in 1874 after the real numbers were defined in a suitable manner by using Dedekind cuts in the 1870's. Only then did d'Alembert's proof begin to be viewed in a more favorable manner, so that today it might be more highly regarded than Gauss's first proof, because the propositions it presupposed were more basic and easier to prove. Gauss's repeated and varied attempts to prove the **FTA**, however, still deserve both historical and mathematical attention.

References

[1] J. R. Argand. Essai sur une manière de représenter les quantités imaginaires dans les constructions géométriques. Paris, 1806.

[2] J. Le R. d'Alembert. Recherches sur le calcul integral. *Hist. Acad. Sci. Berlin*, 2:182–224, 1746.

[3] Leonhard Euler. *Introductio in analysin infinitorum*, volume I of *Opera Omnia*. 1748.

[4] Carl Friedrich Gauss. *Demonstratio nova theorematis omnem functionem algebraicam rationalem integram unius variabilis in factores reales primi vel secundi gradus resolvi posse*. PhD thesis, Universität Helmstedt, 1799. In Werke III, 1–30.

[5] Carl Friedrich Gauss. Demonstratio nova altera theorematis omnem functionem algebraicam rationalem integram unius variabilis in factores reales primi vel secundi gradus resolvi posse. *Comm. Recentiores (Gottingae)*, 3:107–142, 1816. In Werke III, 31–56.

[6] Isaac Newton. *De Methodis serierum et fluxionum*, volume 12 of *The Mathematical Papers of Isaac Newton*. Cambridge, 1967–1981.

[7] Alexander Ostrowski. Über den ersten und vierten Gauss'schen Beweis des Fundamentalsatzes der Algebra. In *Nachrichten der Gesellschaft der Wissenschaften Göttingen*, 1920.

[8] John Stillwell. *Mathematics and Its History*. Springer, New York, 1989.

[9] D. J. Struik. *A Source Book in Mathematics 1200–1800*. Harvard University Press, Cambridge, 1969.

[10] B. L. van der Waerden. *A History of Algebra*. Springer, Berlin, 1985.

[11] Karl Weierstreass. *Einleitung in die Theorie der analytischen Funktionen*, 1874. Notes by G. Hettner.