New Proof of the Theorem That Every Algebraic Rational Integral Function In One Variable can be Resolved into Real Factors of the First or the Second Degree.

Which Carl Friedrich Gauss has presented,
In order to obtain the highest honors in philosophy,
To the famous faculty of philosophers
At the Julia Carolina Academy.

Helmstedt
At C.G. Fleckeisen's, 1799

Foreword

There is something wrong with the title "New Proof of the Theorem …", as Eric Temple Bell remarks in his paper "The Prince of Mathematics" (cf. in J. R. Newman's The World of Mathematics, Vol. I): It was not a new proof but the first proof of this theorem. All previous "proofs" were flawed and therefore not acceptable as proofs.

Gauss submitted this outstanding work to the University of Helmstedt, Germany, as his doctoral dissertation and was awarded the degree in 1799, at the age of 22. He published two further proofs, in Latin like this first proof, in 1816, and a fourth proof, in German, in 1850. The first proof ranks as one of the monumental achievements in the history of mathematics. English versions of the second and third proofs exist (e.g. D.E. Smith's A Sourcebook of Mathematics contains an English version of the essential part of the second proof). But an intensive and extended search of university libraries in the U.S., and the Library of Congress has failed to find an English translation of the first proof (though this proof is cited in countless works on algebra, analysis, and the history of mathematics). If an English version exists in this country, then it is certainly not readily available. To remedy this deficiency, this translation will be put on file and available through the Internet. It is obvious that the number of readers will always be very small. But that, after all, is the fate of many works found currently in our libraries, and not reason enough to do entirely without an English version of Gauss' celebrated work.

This translation has followed the rule: As precisely as possible, as freely as necessary[1]. It is consequently not in the style of 21st century English but rather like some books dating from the beginning of the 19th century. The reader who wants to familiarize himself with an important work of a former century will probably appreciate a translation that adheres to the style of the original. For that, Gauss' long Latin sentences have been rendered into long English sentences except where this was incompatible with the requirement of clarity. (A sentence that stretches over nine lines in print, as for instance in Article 9, can hardly be translated into intelligible English without splitting it into several sentences.) In the same vein, the punctuation of the original has been followed where this was possible without impairing clarity.[2] A clear representation of Gauss' ideas was, after all, the overriding goal. It must, however, be kept in mind that these ideas are not so simple that any translation – or the original for that matter – would make easy reading. Yet, though many modern proofs of the Fundamental Theorem of Algebra exist, it seems worthwhile to go back to that unique achievement of Gauss. It is hoped that this translation will facilitate such an endeavor.

Translated by:
Ernest Fandreyer, M.S., Ed.D.
Professor emeritus

Fitchburg State College
Department of Mathematics
Fitchburg, MA

---

1 Dr. W. Frahnert's German translation, of 1889, seems to follow the opposite rule: As freely as possible, as precisely as necessary.

2 Also, the abbreviations and the italics are those of the Latin original.

1.

Every determined algebraic equation can be reduced to the form $x^m + Ax^{m-1} + Bx^{m-2} + \dots M = 0$, such that m is a positive integer. When we denote the first part of this equation by X and suppose that the equation X = 0 is satisfied by several different values of x, say by setting $x = \alpha$, $x = \beta$, $x = \gamma$, etc., then the function X will be divisible by the product of the factors $x - \alpha$, $x - \beta$, $x - \gamma$, etc. Vice versa, when the product of several simple factors $x - \alpha$, $x - \beta$, $x - \gamma$, etc. is a divisor of the function X, then the equation X = 0 will be satisfied if this x is set equal to each of the quantities α, β, γ etc. Lastly, when X is equal to the product of m such simple factors (which may all be different, or some of which may be identical), then other simple factors besides these cannot divide the function X. For that reason, an equation of degree m cannot have more than m roots; at the same time it is indeed clear that an equation of m-th degree may have *fewer* roots, even if X may be resolved into m simple factors: namely if among these factors some are identical, then the number of different ones which fulfill the equation will necessarily be smaller than m. For the sake of agreement, the mathematicians prefer to say that also in this case the equation has m roots, only that some of these turn out to be equal; which to say, they may at any rate permit themselves.

2.

What has so far been set forth is sufficiently proved in algebraic books and does not in any way offend mathematical rigor. But analysts seem to have adopted far too quickly and without previous solid proof a theorem upon which almost all of the teaching of equations is built: *That any such function X can always be resolved into m simple factors* or, which agrees with that entirely, *that every equation of degree m has indeed m roots*. But already with equations of second degree such cases are quite often encountered which disagree with this theorem. In order to make these cases conform to it, the algebraists were forced to invent an imaginary quantity whose square is -1; and then they acknowledged, if quantities of the form $a + b\sqrt{-1}$ be admitted in the same manner as real quantities, that the theorem is true not only for equations of second degree but also for cubic and biquadratic equations. But it was by no means allowed to infer from this that by admitting quantities of the form $a + b\sqrt{-1}$ equations of fifth or higher degree can be satisfied or, as it is often expressed (though I commend a less slippery expression), that the roots of any equation can be reduced to the form $a + b\sqrt{-1}$. This theorem does not differ from the one stated in the title of this paper if one considers only the subject matter; it is the intent of this dissertation to give a new rigorous proof of this theorem.

Again, since the time when the analysts found out that there are infinitely many equations which had no roots at all unless quantities of the form $a + b\sqrt{-1}$ were admitted as a peculiar kind of quantities called *imaginaries* distinct from *reals*, these supposed quantities have been studied, and have been introduced in all of analysis: with what justification, I shall not discuss here. I shall free my proof from any help of imaginary quantities, although I might avail myself of that liberty of which all those dealing with analysis make use.

3.

Although the proofs for our theorem which are given in most of the elementary textbooks are so unreliable and so inconsistent with mathematical rigor that they are hardly worth mentioning, I shall nevertheless briefly touch upon them so as to leave nothing out.

"In order to demonstrate that any equation $x^m + Ax^{m-1} + Bx^{m-2}$ + etc. + M = 0, or X = 0, has indeed m roots, they undertake to prove that X can be resolved into m simple factors. To this end they assume m simple factors x - α, x - β, x - γ, etc., where α, β, γ etc. are as yet unknown, and set their product equal to the function X. Then they deduce m equations from the comparison of the coefficients and assert that from these the unknown quantities α, β, γ etc. can be determined, in as much as their number is also m. That is to say, m – 1 unknowns can be eliminated until an equation emerges which, as it seems proper, contains only one unknown."

I will not speak of what else might be objected to such an augmentation. I merely ask how can we be certain that this last equation has any root? Could it not be possible that in the whole realm of real and imaginary numbers there is not any quantity which satisfies either this last or the proposed equation? Moreover, experienced persons will easily perceive that this last equation must necessarily be *entirely identical* with the proposed one if only the calculations have been performed according to rule. In other words, after the unknown quantities β, γ etc. have been eliminated an equation $\alpha^m + A\alpha^{m-1} + B\alpha^{m-2}$ + etc. + M = 0 must appear. More to say about such reasoning is not necessary.

Certain authors who seem to have perceived the weakness of this method assume virtually as *an axiom* that an equation has indeed roots, if not possible ones, then impossible roots. What they want to be understood under possible and impossible quantities, does not seem to be set forth sufficiently clearly at all. If possible quantities are to denote the same as real quantities, impossible ones the same as imaginaries: then that axiom can on no account be admitted but needs a proof necessarily. Yet it does not appear that the expressions are to be accepted in this sense but that the meaning of the axiom seems rather to be this: "Although we are not yet certain that m real or imaginary quantities necessarily exist which satisfy any given equation of degree m, we will yet suppose this for the moment; for if by chance it should happen that that many real or imaginary quantities cannot be found, then certainly the escape remains open that we might say the others are impossible." If someone rather wants to use this phrase than simply to say the equation in this case does not have that many roots, then I have no objections: But if he then uses these impossible roots in such a manner as if they were something factual, and e.g. it is said that the sum of all roots of the equation $x^m + Ax^{m-1}$ + etc. = 0 is –A although among these there are impossible roots (which expression in fact signifies *although some are absent*), then I cannot at all approve of this. For the impossible roots, accepted in such a sense, are still roots, and then that axiom can in no way be admitted without proof; nor would it be inappropriate to ponder whether equations might not exist which do not even have impossible roots.

(The following long paragraph is a footnote in Gauss' dissertation.)

By imaginary quantity I always understand here a quantity comprised in the form $a+b\sqrt{-1}$, as long as b is not = 0. That expression has always been used in this sense by all the mathematicians of first rank; and I believe one should not listen to those who wanted to call the quantity $a+b\sqrt{-1}$ imaginary only in the case where a = 0, but impossible whenever a is not = 0; for such distinction is neither necessary nor of any use. --- If imaginary quantities are to be retained in analysis at all (which seems for several reasons more advisable than to abolish them, once they are established in a solid manner), then they must necessarily be considered equally possible as are real quantities; for which reason I would like to comprise the reals and the imaginaries under the common denomination of *possible quantities*: Against which I would call *impossible* a quantity that would have to fulfill conditions which could not even be fulfilled by allowing imaginaries. But that way the phrase would signify just the same as if it said that such a quantity does not exist in the whole realm of magnitudes. I cannot at all allow, however, from this to form a new species of quantities. If someone would say a rectilinear equilateral right triangle is impossible, there will be nobody to deny that. But if he intended to consider such an impossible triangle as a new species of triangles and to apply to it other qualities of triangles, would anyone refrain from laughing? That would be playing with words, or rather, misusing them.

Indeed even first-rate mathematicians have sometimes applied facts that presuppose the possibility of the quantities, which they consider, to such quantities also whose possibility was so far doubtful. Nor would I deny that licenses of this kind pertain mostly to the mere form of the computations, like a veil which the acumen of a true mathematician will soon be able to penetrate. Nevertheless, it seems more advisable and more worthy of the sublime science – which is rightly celebrated as a most perfect model of clarity and reliability – either to prohibit such liberties entirely or at least to use them more sparingly; and to use them only where even less experienced persons may be able to realize that the matter could have been treated without the help of such liberties, though perhaps less briefly but yet with equal rigor.

Furthermore, I do not at all deny that what I have said here against the misuse of the impossible quantities, may in a certain respect also be offered against the imaginary quantities. But I reserve the vindication of the latter, and in fact a fuller explanation of that whole matter, for another occasion. (End of footnote)

4.

Before I review the proofs of our theorem given by other mathematicians and lay open what, to my mind, must be criticized in each, I observe that it suffices to show only this: Every equation of whatever degree $x^m + Ax^{m-1} + Bx^{m-2} + etc. + M = 0$, or X = 0 (where the coefficients A, B, etc. are real numbers) will be satisfied at least once by a value of x of the form $a+b\sqrt{-1}$. For it is well known that then X is divisible by a real factor $x^2 + 2ax + a^2 + b^2$, if b is not = 0, and by a simple factor x – a if b = 0. In either case will the quotient be real, and of a lower degree than X. And since by the same reasoning this quotient must have a real factor of the first or second degree, it is clear that by continuing this process the function X will at last be resolved into simple or second-degree real factors; or into m simple factors if you prefer two simple imaginary factors to single second-degree ones.

5.

The first proof of the theorem we owe to the famous mathematician d'Alembert, *Recherches sur le calcul intégral, Histoire de l'Acad. de Berlin, Année 1746, p.182 ff.* The same proof is set forth in *Bougainville, Traité du calcul intégral, à Paris 1754, p. 47 ff.* The principal points of his method are these:

First he proves: Whatever function X of a variable x may be = 0 either for x = 0 or for x = ∞, and may obtain an infinitely small, positive, real value when x is assigned a real value, that function can also obtain an infinitely small negative value through a value of x which is either a real number or of the imaginary form $p+q\sqrt{-1}$. For let Ω designate an infinitely small value of X, and ω the corresponding value of x. Then he asserts that ω can be expressed by a rapidly converging series $a\Omega^{\alpha}+b\Omega^{\beta}+c\Omega^{\gamma}$ etc., where the exponents α,β,γ etc. are constantly increasing rational numbers that will be positive at least within a certain distance from the beginning of the series, and that will make the terms, in which they appear, infinitely small. If among all these exponents none occurs that is a fraction with even denominator, then all terms of the series will be real numbers for positive as well as for negative values of Ω. But if some fractions with even denominator are found among these exponents then the corresponding terms, for negative values of Ω, will be in the form $p+q\sqrt{-1}$. Because the series converges rapidly it suffices in the former case to consider only the first (i.e. the greatest) term; in the latter case one need not proceed beyond the term that first produces an imaginary part.

By similar reasoning it can be shown: If X can obtain an infinitely small negative real value for a real value of x, then the function can also obtain an infinitely small positive real value for a real value of x or for an imaginary x of the form $p+q\sqrt{-1}$.

From that he concludes, secondly, that also a finite real value of X exists in the former case negative, in the latter case positive, which can be produced by an imaginary value of x of the form $p+q\sqrt{-1}$.

Hence it follows, if X is such a function of x that it takes the real value V for the real value v of x and also obtains a real value greater or smaller by an infinitely small amount, for a real value of x, then this same function can also take a real value greater or smaller than V (resp.) by an infinitely small, and indeed even by a finite, quantity by assigning to x a value of the form $p+q\sqrt{-1}$. This can be deduced from the preceding without trouble if V + Y is substituted for X and v + y for x.

Finally d'Alembert asserts: If it is supposed that X can run through the whole interval between two real values R, S (i.e. be equal now to R, then to S and to all real values between) by always assigning to x values of the form $p+q\sqrt{-1}$, then the function X can moreover be increased or decreased (just as S > R or S < R) by any finite real quantity with x always having the form $p+q\sqrt{-1}$.

For if any real quantity U is given (with S supposed to fall between U and R) to which X could not be equal by such a value of x, then necessarily a maximum of X will exist (that is when S > R, or a minimum when S < R), say T, which it will acquire for a value $p+q\sqrt{-1}$ of x, so that no value of similar form could be assigned to x which would bring the function X closer than even the smallest deviation towards U. Now in an equation between X and x let the value $p+q\sqrt{-1}$ be substituted everywhere for x, and then set first the real part equal to zero, and next the part that involves the factor $\sqrt{-1}$ which is then omitted. From the two equations thus produced two others can be obtained by elimination, in one of which p, X, and constants appear again while the other, free of p, involves only q, X and constants. Thus when X runs through all the values from R to T for real values of p, q, then, by the preceding, X can approach the value of U ever more closely when such values $\alpha+\gamma\sqrt{-1}, \quad \beta+\partial\sqrt{-1}$ are assigned to p, q respectively. From that, however, x can be made $=\alpha-\partial+(\gamma+\beta)\sqrt{-1}$, i.e. it is yet of the form $p+q\sqrt{-1}$, contrary to the hypothesis.

Thus when it is supposed that X denotes a function of the form $x^m + Ax^{m-1} + Bx^{m-2}$ + etc. + M, it is perceived without difficulty that such real values can be assigned to x that X may run through the whole interval between two real values. Wherefore some value of the form $p+q\sqrt{-1}$ can be obtained for x so that X may be made = 0. QED*

6.

The objections to d'Alembert's proof can generally be rendered as these:

1. d'Alembert expresses no doubt about the existence of values of x to which given values of X correspond, but assumes their existence and investigates only the *form* of these values.

   Although this objection is in itself really quite grave, it pertains here nevertheless only to the form of the expression, which can easily be corrected in such a manner that the objection is completely demolished.

* It is proper to observe, that d'Alembert applied geometric considerations in the exposition of his proof and looked upon X as the abscissa , and x as the ordinate of a curve (according to the custom of all mathematicians of the first part of this century to whom the notion of functions was less familiar). But all his reasoning, if one considers only what is essential, rests not on geometric but on purely analytic principles, and an imaginary curve and imaginary ordinates are rather hard concepts and may offend a reader of our time. Therefore I have rather given here a purely analytic form of representation. This footnote I have added so that someone who compares d'Alembert's proof with this concise exposition may not mistrust that anything essential has been altered.

2. The assertion that ω can always be expressed by a series of the kind which he sets down, is certainly false if X may also designate any transcendental function (as d'Alembert admits in several places). This is manifest, for example, when X is set $= e^{\frac{1}{x}}$, or $x = \frac{1}{\ln X}$. If, however, we restrict the proof to the case where X is an algebraic function of x (which suffices for the present task) then the proposition is at any rate true. Beyond this, d'Alembert adduces nothing for the verification of his assumption. Bougainville stipulates X to be an algebraic function of x and recommends Newton's parallelogram for the development of the series.

3. He makes use of infinitely small quantities with greater freedom than is consistent with mathematical rigor or will, at any rate, be conceded by a scrupulous analyst in our days (when these quantities are deservedly regarded with disfavor). Nor has he explained sufficiently clearly the leap from the infinitely small value of Ω to the finite one. His proposition that Ω can also reach some finite value, he appears to conclude not so much from the possibility of Ω having an infinitely small value but rather from the fact that, for very small values of Ω and because of the strong convergence of the series, it approaches the true value of ω all the more, the more terms of the series are taken. Or, that the equation which exhibits the relation between ω and Ω or x and X will be made the more exact if the sum of more terms is taken for ω. Besides the fact that this whole argumentation appears too vague that one can draw a rigorous conclusion from it, I observe that anyhow there are series which, however small a value may be given to the quantity by whose powers they progress, nevertheless always diverge if only they are continued far enough, so that you may arrive at terms greater than any given quantity.* This happens when the coefficients of the series form a hypergeometric progression, therefore it must necessarily be demonstrated that in the present case such a hypergeometric series cannot appear.

For the rest it appears to me that d'Alembert did not suitably resort to infinite series and that these are not at all suitable to establish this fundamental theorem of the theory of equations.

---

* Here I note in passing that among these series there belong very many of those which at first glance seem to converge strongly, e.g. for the greatest part those which the illustrious Euler uses in the latter part of the Inst. Calc. Diff., chapter VI in order to approximate the sum of other series, p.441-474 (the remaining series, p.475-478, are indeed convergent); something that nobody has till now observed as far as I know. Therefore it is greatly to be desired that it will be clearly and rigorously revealed why these series which converge at first very quickly but then, gradually, more and more slowly and finally diverge more and more, may nevertheless supply a sum very near the correct one if not too many terms are taken, and in how far may such a sum be taken to be exact?

4. From the assumption that X can obtain a value S but not a value U, it does not follow that there falls necessarily a value T between S and U which X can attain but not exceed. There exists here also another case: namely, it could happen that there is a limit between S and U which X can approach as closely as you wish but which it yet never reaches. From the arguments supplied by d'Alembert it only follows that X can exceed any value which it will attain by a finite quantity. Namely, when X has become = S, it might still be increased by some finite quantity Ω; then a new increment Ω' may be added, then another addition Ω'' etc. So that, how many increments may have been added already, none need to be considered the last one, but always something new may be added. But although the *number* of possible increments may not be restricted by any limits, yet it may at any rate happen that the sum $S+\Omega+\Omega'+\Omega''$ etc. nonetheless never attains a limit even if the increments $\Omega, \Omega', \Omega''$ etc. continually decrease and no matter how many terms may be considered.

Although this case cannot occur when X signifies an integral algebraic function of x, yet without proof that this case cannot happen, the development must be considered incomplete. In fact, when X is a transcendental function or a rational function that case can take place, e.g. always when to some value of X corresponds an infinitely large value of x. Then d'Alembert's development cannot, as it seems, be brought down to indubitable principles without great digressions, and in some cases not at all.

For these reasons I am unable to consider d'Alembert's proof as satisfactory. Despite this it seems to me that the true substance of the proof under discussion is in no way impaired by all these objections. And I believe that on the same foundation (although in a far different way and at least with greater circumspection) a rigorous proof of our theorem may be constructed, and not only that, but everything may be obtained from it that can be asked of a theory of transcendental equations. I may treat of this very weighty matter at greater length on another occasion. Meanwhile cf. below art. 24.

7.

After d'Alembert, the illustrious Euler published his inquiries on this subject, *Recherches sur les racines imaginaires des équation, Hist. de l'Acad. de Berlin, A.1749, p 223 ff.* Euler delivered a two-fold development. A summary of the first one follows here.

First, Euler undertakes to prove: If m denotes any power of 2, then the function $x^{2m} + Bx^{2m-2} + Cx^{2m-3}$ + etc. + M = X (where the coefficient of the second term is = 0) can always be resolved into two real factors in which x rises up to m dimensions. To this purpose he assumes two factors, $x^m - ux^{m-1} + \alpha x^{m-2} + \beta x^{m-3}$ + etc. and $x^m + ux^{m-1} + \lambda x^{m-2} + \mu x^{m-3}$ etc., where the coefficients $u, \alpha, \beta,$ etc., $\lambda, \mu,$ etc. are as yet unknown, and sets their product equal to the function X. Then the comparison of the coefficients yields 2m – 1 equations, and manifestly it needs only to be proved that to the unknown coefficients u $, \alpha, \beta,$ etc., λ , μ , etc. (whose number is also 2m – 1) such real values can be given that may satisfy these equations. Now E. affirms: If at first u is considered as known so that the number of unknowns would be one less than the number of equations, then all the $\alpha, \beta,$ etc., $\lambda, \mu,$ etc. can be determined from these equations, combined according to the algebraic rules, by rational operations and without root extractions, through u and the coefficients B, C, etc.; and they will indeed become real values as soon as u is real. Furthermore all the $\alpha, \beta,$ etc., $\lambda, \mu,$ etc. can certainly be eliminated, so that an equation U = 0 will result where U will be an integral function of u and of known coefficients only. To solve this equation by the usual elimination method would be an immense task when the proposed equation X = 0 is of somewhat high degree; and for an indeterminate degree, it would be clearly impossible (as E. himself points out, p.239).

However here it suffices to know one property of this equation, namely that the last term in U (which does not contain u) is necessarily negative, whence it is well known to follow that the equation has at least one real root, or that u and consequently also $\alpha, \beta,$ etc., λ, μ , etc. can be determined as real numbers in at least one way. This property may indeed be proved by the following considerations: If it is supposed that $x^m - ux^{m-1} + \alpha x^{m-2}$ + etc. is a factor of the function X, then u will necessarily be the sum of the m roots of the equation X = 0, and it may have just as many values as m can be chosen in different ways from 2m roots or, according to the principles of the combinatorial calculus $\frac{2m \cdot (2m-1) \cdot 2(m-2) \cdot ..... \cdot (m+1)}{1 \cdot 2 \cdot 3 \cdot .... \cdot m}$ values. This number is always twice an odd number (I omit the not difficult proof); if we set it = 2k, then its half k will be odd. The equation U = 0 will indeed be of degree 2k. But because in the equation X = 0 there is no second term, the sum of all 2m roots will be 0. Therefore it is clear, if the sum of any m roots is +p, then the sum of the remaining roots must be –p. i.e. if +p is among the values of u, then so will be the value –p. Hence E. concludes that U is the product of k such double factors as $u \cdot u - p \cdot p$, $u \cdot u - q \cdot q,$ $u \cdot u - r \cdot r$ etc., where +p, -p, +q, -q etc. denote all the 2k roots of the equation U = 0. Because of the odd number of these factors, the last term in U will therefore be the square of the product $p \cdot q \cdot r$ etc., with negative sign. But this product $p \cdot q \cdot r$ etc. can always be computed by rational operations from the coefficients B, C, etc. and is therefore necessarily a real number. Consequently, its square preceded by a negative sign will certainly be a negative quantity. Q.E.D.

Because those two real factors of X are of degree m, and m is a power of 2, each of them can in the same manner be resolved again into two real factors of degree $\frac{1}{2} \cdot m$. Because by repeated halving of the number m one will at last necessarily arrive at a binary term, it is manifest that by continuing the operation the function X will eventually be resolved into real factors of the second degree. But if the function is proposed to be such that its second term is not missing, e.g. $x^{2m} + Ax^{2m-1} + Bx^{2m-2}$ + etc. + M, where 2m is still a binary power, then this function will by the substitution $x = y - \frac{A}{2m}$ be transformed into a similar function without second term. Hence it follows easily that such a function is also resolvable into real factors of the second degree.

Lastly, let us propose a function of degree n where the number n is not a binary power. Then we may set the nearest binary power greater than n, = 2m, and may multiply the proposed function by any 2m – n simple real factors. It follows without difficulty from the resolvability of the product into real factors of the second degree, that the proposed function can also be resolved into factors of the second or the first degree.

8.

Against this proof arise the following objections:

1. The rule from which E. concludes that 2m – 2 unknown quantities $\alpha, \beta,$ etc., $\lambda, \mu,$ etc. can be computed from 2m – 1 equations by rational operations, is by no means true in general but suffers quite often an exception. If someone considers e.g. in article 3, one of the unknown quantities as known and then tries to express the other quantities by that one and the given coefficients with rational operations, he would quickly discover that to be impossible. None of the unknown quantities can be determined otherwise than from an equation of degree m – 1. From what was said previously, this must necessarily follow, as can be seen here at once. Nevertheless it might very well be deliberated whether not also, in the present case, for some values of m the matter might be such that the unknowns $\alpha, \beta,$ etc., $\lambda, \mu,$ etc. cannot be determined from u, B, C, etc. other than from an equation of a degree perhaps greater than 2m. For the particular case where the equation X = 0 is of the fourth degree, E. has worked out the rational values of the coefficients from u and the given coefficients. Whether that can indeed be done in all higher equations, needed at least further analysis.

   Furthermore, it seems worthwhile to inquire more deeply and generally into those formulas which express $\alpha, \beta,$ etc. through u, B, C, etc. by rational operations. I intend to deal with this matter and with others that belong to the theory of elimination (this matter has by no means been exhausted) more thoroughly at another occasion.

2. Even if it had been proved that formulas could be found, for equations of whatever degree, with which those quantities $\alpha, \beta,$ etc., $\lambda, \mu,$ etc. might be computed from u, B, C, etc. by rational operations, yet truly these formulas may become *indeterminate* for certain well-defined values of the coefficients u, B, C, etc. This way it might not only be impossible to compute the unknown quantities from u, B, C, etc. by rational operations, but in certain cases there may be indeed no real values of $\alpha, \beta,$ etc., $\lambda, \mu,$ etc. corresponding to some real value of u. For the confirmation of this I refer the reader, for the sake of brevity, to the very dissertation of E., where on p.236 the equation of fourth degree is at length developed. Everyone will see immediately that the formulas for the coefficients $\alpha, \beta,$ etc. become indeterminate when C = 0 and the value of 0 is assumed for u. Then the values of those coefficients cannot be obtained without root extractions, and they are not even real numbers when the quantity BB – 4D is negative. To be sure, in this case u has still other real values to which real values of $\alpha, \beta,$ may respond, as is easily seen. Yet it is to be feared that for higher equations the solution of this difficulty (which E. does not touch upon at all) will cause much harder work. This matter certainly must not be passed over with silence in an exact proof.

3. E. tacitly assumes that the equation X = 0 has 2m roots and that their sum is = 0 because X has no second term. What I think of that license (which all authors use for this argument), I have made plain in art. 3 above. The assumption that the sum of all the roots of any equation is equal to the first coefficient, with opposite sign, does not seem applicable to other equations, but only to those that have roots. But as by this proof it must be shown that the equation X = 0 indeed has roots, it does not at all seem permissible to assume their existence. Without doubt, those who have not yet penetrated the fallaciousness of this paralogism will answer *that here it is not to be proved that the equation X = 0 can be satisfied* (for that is the meaning of the expression "it has roots") *but merely that the equation may be fulfilled by values of x of the form* $a+b\sqrt{-1}$. *The former statement they assume indeed like an axiom.* However, other forms for quantities beyond the real, and imaginary numbers $a+b\sqrt{-1}$ can not be conceived. And thus it does not appear clear in which way that which is to be proved may differ from that which is assumed like an axiom. If it were possible to devise yet other forms of quantities, say of the form F, F', F" etc., then we should still not be obliged without a proof to admit that any equation whatever would be satisfied by a value of x which is real or of the form $a+b\sqrt{-1}$ or of the form F, or F' etc. For which reason that axiom cannot have any other meaning than this: An equation may be satisfied by a real value of the unknown, or by an imaginary value of the form $a+b\sqrt{-1}$, of perhaps by a value of a hitherto unknown form, or by a value which is not contained in any form whatsoever. But it can certainly not be understood with that clarity which must always be insisted upon in mathematics how quantities of such a nature, of which you cannot have any idea, may be added or multiplied. They are merely a shadow of a shadow.*

* This whole matter will be amply explained in another inquiry which is already being printed, which deals with a very different but nevertheless analogous topic. There, I might have availed myself of a similar license with the same right certainly as has been done with equations by all mathematicians. The proofs of several statements could have been given in a few words with the help of such fictions. Without these they turned out to be quite difficult and required the most subtle art. But I have preferred to abstain from those fictions entirely, and I expect that this will give me more satisfaction than having followed the method of those mathematicians.

However, I do not want with these objections in any way to render as suspect the conclusions which E. has drawn from his assumption. Rather, I am certain that they can be made good in a way that is neither difficult nor very different from Euler's, so that not even the smallest scruple will remain. I merely reprehend the *form* which may indeed be very useful for the *discovery* of new theorems but seems hardly acceptable for *proofs* before the public.

4. E. brings nothing at all for the proof of the assertion that the product $p \cdot q \cdot r$ etc. can be determined from the coefficients in X *by rational operations.* All that he explains about this matter, for the fourth-degree equation, is this (where a, b, c, d are the roots of the proposed equation $x^4 + Bx^2 + Cx + D = 0$):

"Without doubt it will be objected that I have here assumed the quantity $p \cdot q \cdot r$ to be a real number and its square ppqqrr to be positive. But that may be doubtful seeing that the roots a, b, c might be imaginary numbers; it could then happen that the square of $p \cdot q \cdot r$, which is composed of them, becomes negative. To this I respond that this case can never happen. For if some of the roots a, b, c, d are imaginary numbers, we know nevertheless that we must have a + b + c + d = 0, ab + ac + ad + bc + bd + cd = B, abc + abd + acd + bcd = -C * , abcd = D, the quantities B, C, D being real numbers. But since p = a + b, q = a + c, r = a + d, their product pqr = (a + b)(a + c)(a + d) is determinable, *as we know*, from the quantities B, C, D and will consequently be a real number, which is in effect pqr = -C and ppqqrr = CC, just as we have seen. Just as easily it can be seen that in higher equations the same circumstances must take place and that one cannot make objections to me from that side."

The condition that the product pqr can be determined from B, C etc. *by rational oprations* Euler has not added anywhere but seems to have had in mind all along, for without it the proof could have no force. Now it is certainly true that in equations of the fourth degree one obtains aa(a + b + c + d) + abc + abd + acd + bcd = -C by developing the product (a + b)(a + c)(a + d). Yet it does not seem sufficiently clear how that product can be computed from the coefficients, by rational operations, in all equations of higher degree. The famous de Foncenex, who first noticed this *(Miscell. phil. Math. Soc. Taurin. T.I. p.117.)*, contends correctly that the procedure loses all force without a rigorous proof of this assumption. He admits that such a proof seems to him to be quite difficult and relates which way he has tried but in vain. **

* Euler erroneously has C, hence he later also states pqr = C, incorrectly.

** In that exposition an error appears to have crept in. On p.118, l.5, instead of the letter p (on choisissoit seulement celles où entroit p etc.) it is necessary to read <u>une même racine quelconque de l' équation proposée</u>, or something similar, because otherwise it makes no sense.

Yet this matter is settled without difficulty by the following method (of which I can here merely give a summary): Although in equations of the fourth degree it is not sufficiently clear that the product $(a + b)(a + c)(a + d)$ is determinable by the coefficients B, C, D, it is yet easily perceived that that product is also $= (b + a)(b + c)(b + d)$, and besides $= (c + a)(c + b)(c + d)$, and finally $= (d + a)(d + b)(d + c)$. Therefore the product pqr will be a quarter of the sum $(a + b)(a + c)(a + d) + (b + a)(b + c)(b + d) + (c + a)(c + b)(c + d) + (d + a)(d + b)(d + c)$, which, when it is worked out, will be a rational integral function of the roots a, b, c, d of the kind in which all terms enter in the same way as can be seen beforehand without trouble. Such functions can indeed always be expressed through the coefficients of the equation whose roots are a, b, c, d.

The same is also manifest when the product pqr is brought to this form: $\frac{1}{2}(a + b - c - d) \cdot \frac{1}{2}(a + c - b - d) \cdot \frac{1}{2}(a + d - b - c)$, and it is easy to see beforehand that this product when worked out will involve all a, b, c, d in the same way[1)]. At the same time, experienced mathematicians will infer from this how this method may be applied to higher-degree equations. I reserve for another occasion the complete exposition of this proof, which brevity does not permit to bring here, together with a fuller investigation of functions which involve several variables in the same manner.

Furthermore I observe that beyond these four objections some other points can still be criticized in Euler's proof over which I will pass in silence, however, so that I may not perchance be looked upon as too strict a critic. The proof in the very form in which Euler set it forth can in no way be considered complete.

After this proof Euler shows moreover another way how to reduce the theorem for equations, whose degree is not a binary power, to the solution of equations of the aforementioned sort: But as the latter method teaches nothing about equations whose degree is a binary power, it is not necessary here to explain it at length, all the more since it is equally liable to all the objections (other than the fourth) as the first, general proof.

9.

In the same treatise, Euler has striven to prove our theorem in yet another way (p. 263) whose substance is contained in this: For a given equation $x^n + Ax^{n-1} + Bx^{n-2}$ etc. $= 0$, it has so far indeed not been possible to find an analytic expression for its roots if $n > 4$. But still it appears certain (as Euler asserts) that such an expression can contain nothing other than arithmetic operations and root extractions, which will be the more complicated the greater n is.

[1)] i.e. symmetrically (translator's note)

If this is conceded, Euler shows very nicely that, however complicated the radicals may be among themselves, yet the value of the formula will always be represented in the form $M + N\sqrt{-1}$, where M, N are real quantities.

Against this reasoning one can object that after so much labor of such great mathematicians there is very little hope left ever to arrive at a general solution of algebraic equations. It seems more and more probable that such a solution is entirely impossible and contradictory. This must not at all be considered paradoxical, *as that which is commonly called the solution of an equation is indeed nothing other than its reduction to pure equations*. For the solution of pure equations is here not taught but presupposed; and if you express the roots of an equation $x^m = H$ by $\sqrt[m]{H}$, you have in no way solved it, and you have not done more than if you had devised some symbol to denote the root of an equation $x^h + Ax^{h-1} + \text{etc.} = 0$ and set the root equal to this.

It is true, pure equations stand out very much among all the others because of the ease of finding their roots by approximations and because of the beautiful relation which the roots together have among them. And it is therefore not to be censured that mathematicians have denoted their roots by a special symbol: But from the fact that this symbol is raised to the dignity, and comprised under the name, of *analytic expressions* just like the arithmetic symbols for addition, subtraction, multiplication, division, and powering, it does not at all follow that the root of any equation whatsoever can be expressed by these, unless it is tacitly presupposed, without sufficient reason, that the solution of any equation can be reduced to the solution of pure equations. It is perhaps not so difficult to demonstrate quite rigorously the impossibility already for the fifth degree; I shall report my investigations of this matter more fully in another place. Here it suffices that the general solution of equations, taken in that sense, is so far very doubtful. And a proof whose force depends entirely on this supposition has no weight.

10.

Later, the famous de Foncenex had noticed the defect in Euler's first proof (above, Article 8, objection 4) but had been unable to remove it. And so he tried yet another way and published it in his acclaimed treatise p.120*. It consists in the following.

Let an equation $Z = 0$ be given where Z designates a function of degree m in a variable z. If m is an odd number, then it is well known that this equation has a real root; if however m is even then Foncenex attempts to prove that the equation has at least one root of the form $p + q\sqrt{-1}$, in the following manner. Let $m = 2^n \cdot i$ where i designates an odd number, and let it be supposed that $zz + uz + M$ is a divisor of the function Z.

* The second volume of those Miscelleanorum, p. 337, contains explanations to this treatise. However, they do not deal with the present investigation but with logarithms of negative quantities about which that treatise discourses.

Then individual values of u will be the sums of roots of the equation $Z = 0$ taken in pairs (with interchanged signs). Therefore u will have $\frac{m\cdot(m-1)}{1\cdot 2} = m'$ values. And if u is supposed to be determined by an equation $U = 0$ (where U designates a function of u and of coefficients known from Z), then this equation will be of degree m'. It is quite easy to see that m' will be a number of the form $2^{n-1}\cdot i'$, where i' is an odd number. If m' is not yet odd, suppose $u\cdot u + u'\cdot u + M'$ in turn to be a divisor of U. Clearly by similar reasoning u' is determined by an equation $U' = 0$, where U' is a function of u' of degree $\frac{m'\cdot(m'-1)}{1\cdot 2}$. Setting $\frac{m'\cdot(m'-1)}{1\cdot 2} = m''$, then m'' will be a number of the form $2^{n-2}\cdot i''$, where i'' designates an odd number. If m'' is not yet odd, consider $u'\cdot u' + u''\cdot u' + M''$ to be a divisor of U'; then u'' will be determined by an equation $U'' = 0$. If this is taken to be of degree m''', then m''' will be a number of the form $2^{n-3}\cdot i'''$. It is manifest that in the sequence of equations $U = 0$, $U' = 0$, $U'' = 0$ etc., the n-th equation will be of odd degree and therefore has a real solution. Setting $n = 3$, for brevity, then the equation $U'' = 0$ has the real root u'', and one sees without trouble that the same reasoning holds for any other value of n. Consequently, says Foncenex, the coefficient M'' can be calculated from u'' and coefficients in U' (which will be integral functions of the coefficients in Z, as is easily seen), or from u'' and coefficients in Z, and is therefore real. From this it follows that the root of the equation $u'u' + u''u' + M'' = 0$ will be contained in the form $p + q\sqrt{-1}$; and these roots obviously fulfill the equation $U' = 0$. Therefore any value of u' will be in the form $p + q\sqrt{-1}$. Then the coefficient M' can be computed (in the same manner as before) from u' and coefficients in Z and will also be of the form $p + q\sqrt{-1}$. Therefore the roots of the equation $uu + u'u + M' = 0$ will also be of that form, and they will certainly satisfy the equation $U = 0$; i.e. this equation will have a root of the form $p + q\sqrt{-1}$. Finally, it follows from this similarly that also M and certainly a root of the equation $zz + uz + M = 0$ are of this form; also, this root will manifestly fulfill the given equation $Z = 0$. Wherefore every equation has at least one root of the form $p + q\sqrt{-1}$.

11.

The objections 1, 2, 3 which I raised against Euler's first proof (article 8.), have the same force against this method, but with this difference that the second objection, to which Euler's proof was liable only in certain special cases, strikes the present proof in all cases. For it can straight away be demonstrated, whenever a formula may be given which expresses the coefficient M by u' and coefficients in Z, that this formula will necessarily become indeterminate for several values of u'. Likewise a formula that expresses the coefficient M'' by u'' will become indeterminate for several values of u'' etc. This becomes very clear when we assume for example an equation of the fourth degree. Let us therefore set $m = 4$, and the roots of the equation $Z = 0$ may be $\alpha, \beta, \gamma, \delta$. Then it is clear that the equation $U = 0$ will be of sixth degree with roots $-(\alpha+\beta), -(\alpha+\gamma), -(\alpha+\delta), -(\beta+\gamma), -(\beta+\delta)$, and $-(\gamma+\delta)$. The equation $U' = 0$, then will be of the fifteenth degree and the values of u' these:

$$2\alpha+\beta+\gamma,\ 2\alpha+\beta+\delta,\ 2\alpha+\gamma+\delta,\ 2\beta+\alpha+\gamma,\ 2\beta+\alpha+\delta,\ 2\beta+\gamma+\delta,\ 2\gamma+\alpha+\beta,\ 2\gamma+\alpha+\delta,$$
$$2\gamma+\beta+\delta,\ 2\delta+\alpha+\beta,\ 2\delta+\alpha+\gamma,\ 2\delta+\beta+\gamma,\ \alpha+\beta+\gamma+\delta,\ \alpha+\beta+\gamma+\delta,\ \alpha+\beta+\gamma+\delta.$$

At this equation one must already stop because it is of odd degree, and it will indeed have the real root $\alpha+\beta+\gamma+\delta$ (which will equal the first coefficient of Z with opposite sign and will therefore be not merely real but rational if the coefficients of Z are rational). But one can see without difficulty: If a formula is given which expresses the value of M' through the corresponding value of u' by rational operations, then this formula will necessarily become indeterminate for $u' = \alpha+\beta+\gamma+\delta$. For this value will be three times the root of U' = 0 and to it correspond three values of M', namely $(\alpha+\beta)(\gamma+\delta)$, $(\alpha+\gamma)(\beta+\delta)$, and $(\alpha+\delta)(\beta+\gamma)$ which can all be irrational. But obviously in this case a rational formula can produce neither an irrational value of M' nor three different values. From this example it can be concluded sufficiently that Foncenex's method is not at all satisfactory; in order to deliver a method complete in every respect one must rather inquire much more deeply into the elimination theory.

12.

Finally, the illustrious LaGrange treats of our theorem in his work S*ur la forme des racines imaginaires des équation, Nouv. Mem. de l'Acad. de Berlin 1772, p.222 ff.* Here this great mathematician applies himself primarily to making good the defect in Euler's first proof, and indeed has examined particularly that which constitutes the third and fourth objection above (article 8) so profoundly that nothing more remains to be desired except that perhaps some doubts appear to be left in a former treatise on the elimination theory (on which the present investigation depends totally). However, he has not even touched the first objection; moreover the whole treatise is built upon the supposition that any equation of m-th degree has indeed m roots.

And so having thoroughly considered those proofs which have so far been published I hope that a new proof of this most important theorem, based on entirely different foundations, will not be unwelcome to the experts. This proof I shall at once undertake to explain.

13.

Lemma: Let m denote any positive whole number. Then the function $\sin\varphi\cdot x^m - \sin m\varphi\cdot r^{m-1}x + \sin(m-1)\varphi\cdot r^m$ is divisible by xx – 2 cos$\varphi\cdot$ rx + rr.

Proof: For m = 1, the function is = 0 and therefore divisible by any factor. For m = 2, the quotient is $\sin\varphi$, and for any greater value the quotient will be $\sin\varphi\cdot x^{m-2} + \sin 2\varphi\cdot rx^{m-3} + \sin 3\varphi\cdot rrx^{m-4} + etc. + \sin(m-1)\varphi\cdot r^{m-2}$. It is easily confirmed that the product of this function multiplied by xx – 2 cos$\varphi\cdot$ rx + rr is equal to the given function.

14.

Lemma: If the quantity r and the angle φ are determined in such a way that the equations hold

$r^m \cos m\varphi + Ar^{m-1}\cos(m-1)\varphi + Br^{m-2}\cos(m-2)\varphi + \ldots + Krr\cos 2\varphi + Lr\cos\varphi + M = 0$ (1)
$r^m \sin m\varphi + Ar^{m-1}\sin(m-1)\varphi + Br^{m-2}\sin(m-2)\varphi + \ldots + Krr\sin 2\varphi + Lr\sin\varphi = 0$ (2), then the function $x^m + Ax^{m-1} + Bx^{m-2} + \ldots Kxx + Lx + M = X$ will be divisible by the second-degree factor $xx - 2\cos\varphi \cdot rx + rr$, if only $r \cdot \sin\varphi$ is not = 0. But if $r \cdot \sin\varphi = 0$ then that function will be divisible by the simple factor $x - r\cos\varphi$.

Proof : I. As in the preceding article, the following quantities will all be divisible by $xx - 2\cos\varphi \cdot rx + rr$:

$$\sin\varphi \cdot rx^m \quad - \quad \sin m\varphi \cdot r^m x \quad + \quad \sin(m-1)\varphi \cdot r^{m+1}$$
$$A\sin\varphi \cdot rx^{m-1} - A\sin(m-1)\varphi \cdot r^{m-1}x + A\sin(m-2)\varphi \cdot r^m$$
$$B\sin\varphi \cdot rx^{m-2} - B\sin(m-2)\varphi \cdot r^{m-2}x + B\sin(m-3)\varphi \cdot r^{m-1}$$
$$\text{etc.} \qquad\qquad \text{etc.}$$
$$K\sin\varphi \cdot rxx - K\sin 2\varphi \cdot rrx + K\sin\varphi \cdot r^3$$
$$L\sin\varphi \cdot rx - L\sin\varphi \cdot rx$$
$$M\sin\varphi \cdot r \qquad + M\sin(-\varphi)r$$

Therefore the sum of these quantities will also be divisible by $xx - 2\cos\varphi \cdot rx + rr$. The first terms of these quantities yield the sum $\sin\varphi \cdot rX$; the second terms added together yield 0 because of (2); and indeed also the sum of the third terms vanishes, as is easily seen if one multiplies (1) by $\sin\varphi$, (2) by $\cos\varphi$ and subtracts the one product from the other. Thus it follows that the function $\sin\varphi \cdot rX$ is divisible by $xx - 2\cos\varphi \cdot rx + rr$ and likewise the function X, as long as $r\sin\varphi$ is not = 0. Q.E.P.

II. But if $r\sin\varphi = 0$, then either r = 0 or $\sin\varphi = 0$. In the first case M = 0 because of (1), and therefore X will be divisible by x or by $x - r\cos\varphi$. In the latter case $\cos\varphi = \underline{+}\,1$, $\cos 2\varphi = +1$, $\cos 3\varphi = \underline{+}\,1$ and generally $\cos n\varphi = \cos\varphi^n$. Therefore X will = 0 because of (1) when x is set = $r\cos\varphi$, and consequently the function X will be divisible by $x - \cos\varphi$. Q.E.S.

15.

The outstanding theorem is frequently proved with the help of imaginary numbers, cf. Euler *Introd. In Anal. Inf. T.I. p 110;* I consider it worth the trouble to show how it can easily be elicited without their help. It is quite manifest that for the proof of our theorem nothing more is required than to show: *When any function X of the form* $x^m + Ax^{m-1} + Bx^{m-2} + etc. + Lx + M$ *is given, then r and* φ *can be determined in such a way that the equations (1) and (2) hold.* For it follows thence that X has a real factor of the first or the second degree. And then division produces necessarily a real quotient of lower degree which for the same reason will have a factor of first or second degree, too. Continuing this operation, X will at last be resolved into simple or second-degree factors. It is therefore the main point of the following inquiries to prove that theorem.

16.

Consider an immovable infinite plane (the plane of the Table* Fig. 1) and in it a fixed straight line GC through the fixed point C. Choose some length as unit so that all straight lines can be expressed in numbers. At any point P of the plane whose distance from C is r and with angle GCP = φ erect the perpendicular with length equal to the expression $r^m \sin m\varphi + Ar^{m-1}\sin(m-1)\varphi + \text{etc.} + Lr\sin\varphi$ which for brevity I shall designate T from here on. The distance r I shall always take to be positive, and for points which fall on the other side of the axis the angle φ must be considered as increased by two right angles or as negative (which comes to the same thing). The end points of those perpendiculars will lie above the plane for positive values of T, below it for negative values, and on the plane itself if T vanishes; and they will be on a curved surface, continuous and in all directions infinite, which for brevity I shall call *the first surface* from here on.

* see last page

Again, in a similar manner a second surface may be referred to the same plane and center and axis whose altitude above any point of the plane shall be
$r^m \cos m\varphi + Ar^{m-1}\cos(m-1)\varphi$ + etc. + $Lr\cos\varphi + M$, which expression I shall always denote by U, for brevity. This surface, which will also be continuous and infinite in every direction, I shall distinguish from the former surface by the term *second surface.* Then it is manifest that the whole task is involved in showing that there exists at least one point which lies simultaneously in the plane, on the first surface and on the second surface.

17.

One can easily see that the first surface lies partly above and partly below the plane; for it is clear that the distance from the center can be taken so large that the remaining terms in T contribute nothing compared to the first $r^m \sin m\varphi$; but that one can be made positive as well as negative accordingly as the angle φ is determined. Therefore the fixed plane will necessarily be intersected by the first surface. This intersection of the plane with the first surface I shall call *the first curve*, which will therefore be determined by the equation T = 0. For the same reason, the plane will be intersected by the second surface. The intersection constitutes a curve determined by the equation U = 0, which I will call *the second curve*. Especially, either curve consists of several branches which generally can be separate, but each of which will be a continuous curve. In fact the first curve will always be such as is called complex, and the axis GC must be considered as being a part of that curve; for whatever value may be given to r, T will always = 0 whenever $\varphi = 0$ or $\varphi = 180^o$. But it is better to consider as one curve the composite of all the branches passing through any points where T = 0 (according to generally accepted use in higher mathematics). And to do likewise with all the branches passing through any points where U = 0. Obviously, the task has now been reduced to that of demonstrating that there exists in our plane at least one point where some branch of the first curve is intersected by a branch of the second curve. For this, it is imperative to survey the nature of these curves more closely.

18.

First of all, I observe that either curve is algebraic, and of order M if referred to orthogonal coordinates. If now the origin is assumed at C, the abscissa x taken in the direction of G, the corresponding y toward P, then $x = r\cos\varphi$, $y = r\sin\varphi$ and therefore generally for any n

$$r^n \sin n\varphi = nx^{n-1}y - \frac{n\cdot n-1\cdot n-2}{1\cdot 2\cdot 3}x^{n-3}y^3 + \frac{n\ldots.n-4}{1\cdot\ldots.\cdot.5}x^{n-5}y^5 - etc.,$$

$$r^n \cos n\varphi = x^n \frac{n\cdot n-1}{1\cdot 2}x^{n-2}yy + \frac{n\cdot n-1\cdot n-2\cdot n-3}{1\cdot 2\cdot 3\cdot 4}x^{n-4}y^4 - etc.$$

Wherefore T as well as U will consist of several terms of the form $ax^{\alpha}y^{\beta}$ where α, β denote positive integers whose sum at most = m. Moreover it is easily seen that all the terms of T contain a factor y; consequently the first curve is in the strict sense composed of a straight line (with equation y = 0) and a curve of order m – 1, though this distinction need not be regarded here.

Of greater import will be the investigation whether the first curve, and the second, have infinite legs, how many, and of which kind. At an infinite distance from the point C, the first curve, whose equation is sin m $\varphi + \frac{A}{r}\sin(m-1)\varphi + \frac{B}{rr}\sin(m-2)\varphi$ etc. = 0, will coincide with the curve whose equation is sin m φ = 0. Now this produces m straight lines intersecting each other in point C, the first of which is the axis GCG', the others being inclined towards the axis at angles

$\frac{1}{m}180, \frac{2}{m}180, \frac{3}{m}180$ etc. Wherefore the first curve has 2 m infinite branches which will divide the periphery of a circle with infinite radius into 2 m equal parts, so that the periphery will be cut by the first branch where the circle and the axis meet, by the second branch at the distance $\frac{1}{m}180^o$, by the third at the distance $\frac{2}{m}180^o$ etc. The same way, the second curve will have an asymptote of equation cos m φ = 0 at an infinite distance from the center. This second curve is the composite of m straight lines which intersect at point C at equal angles in such a way that the first line forms an angle of $\frac{1}{m}90^o$ with the axis, the second an angle of $\frac{3}{m}90^o$, the third an angle of $\frac{5}{m}90^o$ etc. Wherefore the second curve will also have 2 m infinite branches, each of which will occupy the space between the two adjacent branches of the first curve so that they intersect the periphery of a circle with infinite radius in points which differ from the axis by

$\frac{1}{m}90^o, \frac{3}{m}90^o, \frac{5}{m}90^o$ etc. Moreover it is clear that the axis itself always constitutes two infinite branches of the first curve, namely the first and the (m + 1)th. In a perceptually clear way, this position of the branches is shown in Figure 2, constructed for the case m = 4, where the branches of the second curve are dotted to distinguish them from the branches of the first curve; this is also to be noted in the fourth figure.*_____ Because these conclusions are indeed of the greatest importance, and because infinitely large quantities may offend some readers, I shall show how to develop these conclusions without the help of the infinite in the following articles.

---

* The fourth figure has been constructed assuming X = x^4 – 2xx + 3x + 10, by which readers less familiar with general and abstract investigations can then look at the position of both curves in a concrete case. The length of CG is taken as = 10.

19.

Theorem. *With everything the same as above, a circle can be described around C on whose periphery are 2 m points where T = 0 and as many where U = 0, and in such a way that the latter ones lie singly between pairs of the former points.*

If the sum of all the coefficients A, B, etc. K, L, M is taken as positive and = S, and also R taken $> S\sqrt{2}$ and > 1*, then I assert that what has been stated in the theorem will necessarily take place in a circle of radius R.

In other words; Let us for the sake of brevity designate by (1) that point of this circle's circumference which is $\frac{1}{m}$ 45 degrees away from the intersection of the circle and the left part of the axis, or for which $\varphi = \frac{1}{m}45^o$; similarly by (3) the point which is $\frac{3}{m}45^o$ away from that intersection, or for which $\varphi = \frac{3}{m}45^o$; next by (5) the point where $\varphi = \frac{5}{m}45^o$ etc. up to (8m – 1) which is $\frac{8m-1}{m}45^o$ away from that intersection if you proceed always in the direction of that part (or $\frac{1}{m}45^o$ towards the opposite part). Whereupon a total of 4 m points will be obtained on the periphery spaced at equal intervals. Then one point will fall between (8m – 1) and (1) for which T = 0, and certainly similar single points will be between (3) and (5), between (7) and (9), between (11) and (13) etc. Thus there are 2 m of them. In the same way, single points for which U = 0 will fall between (1) and (3), (5) and (7), (9) and (11), etc., wherefore their number is also 2 m. Finally, there will be no other points, on the whole periphery, beyond these 4 m points for which T or U = 0.

Proof: I. At point (1), mφ = 45^o and therefore

$$T = R^{m-1}(R\sqrt{\frac{1}{2}} + A\sin(m-1)\varphi + \frac{B}{R}\sin(m-2)\varphi + etc. + \frac{L}{R^{m-2}}\sin\varphi).$$

However, the sum A sin(m – 1) ϕ + $\frac{B}{R}$ sin(m – 2) φ etc. can certainly not be greater than S and is therefore necessarily less than $R\sqrt{\frac{1}{2}}$; whence it follows that the value of T at this point is certainly positive.

---

* When S > $\sqrt{\frac{1}{2}}$, the first condition includes the second; when S < $\sqrt{\frac{1}{2}}$, the second includes the first.

More importantly, T will consequently have a positive value as long as $m \cdot \varphi$ falls between 45^o and 135^o, i.e. from point (1) up to (3) the value of T will be positive. For the same reason will T have a positive value from point (9) and up to (11), and in general from any point (8k + 1) up to (8k + 3) where k denotes any integer. Similarly, T will have negative values everywhere between (5) and (7), between (13) and (15) etc. and generally between (8k + 5) and (8k + 7), and certainly will nowhere in all these intervals = 0. But because at (3) this value is positive, in (5) negative, it will necessarily be = 0 somewhere between (3) and (5); and certainly also between (7) and (9), between (11) and (13) etc. up to and including the interval between (8m – 1) and (1), so that altogether T = 0 at 2 m points. Q.E.P.

II. Beyond these 2 m points there are no others having this property, as can be seen thus: Because there are none between (1) and (3), (5) and (7) etc., such points could exist in no other way than that in some interval between (3) and (5), or between (7) and (9) etc. would lie at least two. Then T would necessarily be a maximum or a minimum in the same interval and therefore $\frac{dT}{d\varphi} = 0$. But $\frac{dT}{d\varphi} = mR^{m-2}(R\cos m\varphi + \frac{m-1}{m}A\cos(m-1)\varphi + \text{etc.})$, and $\cos m\varphi$ is always negative between (3) and (5) and $> \sqrt{\frac{1}{2}}$ *. Whence it is easily seen that in this whole interval $\frac{dT}{d\varphi}$ is a negative quantity; likewise it is positive everywhere between (7) and (9), negative between (11) and (13) etc., so that it can be 0 in none of these intervals, and therefore that supposition cannot be maintained. Therefore etc. Q.E.S.

III. Again, in a similar way it is proved that U has a negative value everywhere between (3) and (5), (11) and (13) etc. and generally between (8k + 3) and (8k + 5), but a positive value between (7) and (9), (15) and (17) etc. and generally between (8k + 7) and (8k + 9). It follows from this immediately that U must become = 0 somewhere between (1) and (3), between (5) and (7) etc., i.e. at 2 m points. Again, in none of these intervals can $\frac{dU}{d\varphi} = 0$ (which is easily proved in the same manner as above); wherefore more than those 2 m points where U = 0 will not be possible on the periphery of the circle. Q.E.T. and Q.

Moreover, that part of the theorem according to which more than 2 m points do not exist at which T = 0, nor more than 2m points at which U = 0, can also be proved from the fact that the equations T = 0, U = 0 describe curves of order m which by a circle as a curve of order 2 cannot be intersected in more than 2 m points as is established in higher mathematics.

---

* In Gauss' time, $\sqrt{\frac{1}{2}}$ was $\pm$ .7071…, and not just .7071 as it is generally defined today. (Translator's note)

20.

If another circle is described around the same center with a radius greater than R and is divided in the same manner, then on this circle one point will also fall between points (3) and (5) and likewise between (7) and (9) etc. where T = 0. And it is easily seen that the less the radius of this circle differs from radius R, the more closely together must these particular points between (3) and (5) be situated on the circumferences of these two circles. The same will also happen when a circle with a somewhat smaller radius than R but greater than $S\sqrt{2}$ and 1 is described. Whence it is perceived without difficulty that the circumference of the circle drawn with radius R is indeed *intersected* by some branch of the first curve in that point between (3) and (5) where T = 0; and the same holds true for the other points where T = 0. In the same way it is clear that the circumference of this circle will be intersected by some branch of the second curve in all those 2 m points where U = 0. These conclusions can also be expressed in the following manner: When a circle of the required magnitude is drawn around center C then 2 m branches of the first curve and as many of the second curve will enter it, and indeed in such a way that any two successive branches of the first curve will be separated by some one branch of the second curve alternately. See Figure 2 where the circle is, however, not of infinite but of finite magnitude. The numbers added to the individual branches are not to be confused with the numbers by which I have in brief form designated, in this and the preceding articles, the definite intersections on the periphery.

21.

Already from this relative position of the branches entering the circle it can be deduced in many ways that there must necessarily be an intersection of some branch of the first curve with a branch of the second curve within the circle. And I hardly know which method should preferably be chosen above the others. The most adroit one seems to be this: Let us designate by 0 (Figure 2) the point on the periphery of the circle where it is intersected by the left part of the axis (which itself is one of the 2 m branches of the first curve); by 1, the nearest point where a branch of the second curve enters; the point nearest to this one where the second branch of the first line enters, by 2, and so on up to $4m - 1$. In this way a branch of the first line enters the circle at every point marked by an even number, whereas a branch of the second line enters at all points represented by an odd number. But according to higher mathematics, any algebraic curve (or the individual parts of such an algebraic curve if it perhaps consists of several parts) either turns back into itself or extends to infinity.

Consequently, a branch of any algebraic curve which enters a limited space, must necessarily exit from this space somewhere. * From this it is easily concluded that any point marked by an even number (or in short, *any even point*) must be connected within the circle to another even point by a branch of the first curve, and likewise any point marked by an odd number to a similar point by a branch of the second curve. Now although this connection of two such point can be very varied according to the nature of the function X so that in general it cannot be clearly described, it is yet easy to prove *that an intersection of the first curve with the second always occurs whatever that connection may be eventually*.

---

* It seems to have been proved with sufficient certainty that an algebraic curve can neither be broken off suddenly anywhere (as happens e.g. with the transcendental curve whose equation is $y = \frac{1}{\ln x}$) nor lose itself, so to say, in some point after infinitely many coils (like the logarithmic spiral). As far as I know, nobody has raised any doubts about this. However, should someone demand it then I will undertake to give a proof that is not subject to any doubt, on some other occasion. In the present case this is really manifest: Suppose some branch, e.g. 2, does not exit the circle anywhere (Figure 3). Then you could enter the circle between 0 and 2, thereafter could go around this whole branch (which would have to lose itself within the space of the circle), and finally could exit again from the circle between 2 and 4 so that in the whole path you shall not have intersected the first curve. That is this is really absurd is clear from the fact that at the point where you enter the circle you had the first surface above you, at the exit below you. Therefore you must necessarily have somewhere come upon the first surface itself, that is to say: Upon a point of the first curve.________ Still, from this reasoning based on the principles of the geometry of position, which principles are no less true than those of the geometry of size, there follows only that much: If you enter the circle on any branch of the first curve, then you can in turn exit from the circle at some other place while remaining always on the first curve. It does not follow, however, that your whole path is a continuous curve in the sense which is understood in higher mathematics. But here it suffices that the path is a continuous curve in the usual sense, i.e. nowhere interrupted and everywhere hanging together.

22.

The proof that this happens necessarily seems to be most suitably given indirectly. That is to say, let us assume the junction of any two of the even points and any two of the odd points can be arranged in such a way that no intersection of a branch of the first curve with a branch of the second curve results. Because the axis is a part of the first curve the point 0 is obviously connected with the point 2 m. Point 1 can therefore not be connected with any point beyond the axis, i.e. with no point marked by a number greater than 2 m, for otherwise the connecting curve would necessarily intersect the axis. If it is therefore assumed that 1 is connected with point n, then n will be < 2 m. For a similar reason, if 2 is established to be connected to n', then n' will be < n, because otherwise the branch 2 . . . n' would necessarily intersect the branch 1 . . . n. For the same reason point 3 will be connected with one of the points falling between 4 and n'; and obviously when 3, 4, 5 etc. are taken as connected with n'', n''', n'''' etc., then n''' falls between 5 and n'', n'''' between 6 and n''' etc. Whence it is evident that at last there will be reached some point h connected with the point h + 2, and then the branch that enters the circle at point h + 1 will necessarily intersect the branch connecting points h and h + 2. But because one of these two branches belongs to the first curve, the other to the second curve, it is thus manifest that the assumption is contradicted and that indeed an intersection of the first curve with the second curve necessarily exists somewhere.

When this is combined with the preceding, it will be concluded from all the investigations set forth that the theorem has been proved with all rigor *that any algebraic rational whole function in one variable can be resolved into real factors of the first or the second degree.*

23.

Moreover, from those foundations it is not hard to deduce that there will be not just one but at least m intersections of the first line with the second, although it can also happen that the first line is intersected by several branches of the second line in the same point. In this case, the function X will have several equal factors. But as it may suffice here to have proved the necessity of one intersection, I shall for the sake of brevity not dwell upon this matter any further. For the same reason, I shall also not pursue other properties of these curves more fully here, e.g. an intersection is always made at right angles; or if several legs of either curve meet at the same point, then there will be as many legs of the first curve as of the second curve, and these will be placed alternately and will intersect one another at equal angles etc.

Finally I observe that it is not at all impossible to render the preceding proof, which I have here built on geometric principles, in purely analytic form. But that representation that I have set forth here, I believe to turn out less abstract; and I place here the very nerve of what was to be proved more clearly before the eyes than might be expected of an analytic proof.

To come to an end, I shall briefly indicate still another method for proving our theorem which at the first glance will appear to be quite different not merely from the preceding proof but from all the other proofs explained in detail above, and which nevertheless is in the strict sense the same as d'Alembert's if one looks at the essential. I commit it to the experts for whose sake it is added to compare this method with that one and to explore the parallelism between the two.

24.

Above the plane of Figure 4 relative to the axis CG and a fixed point I presuppose the first and the second surface in the same manner as before. Take any point on any branch of the first curve where T = 0, (e.g. any point M on the axis). If U is not also = 0 at this point progress from this point along the first curve toward that part in whose direction the absolute value of U decreases. If by chance at point M the absolute value of U decreases towards either part then it does not matter in which direction you may progress. I shall show what is to be done if U increases toward either part. It is manifest, when you always progress along the first curve, that you will necessarily arrive at a point eventually where U = 0 or at such a point where the value of U becomes a minimum, e.g. point N. In the first case we have found what we were looking for; in the latter case, however, one can prove that in this point several branches of the first line intersect (and indeed an even number of branches). Their halves are arranged in such a manner that the value of U continues to decrease into whichever half you may divert (whether this one or that one). (Of this theorem I must for brevity omit the proof which is more lengthy than difficult.) On this branch you can then again progress until U either becomes = 0 (as happens in Figure 4 at P) or again a minimum. When you divert in turn, you will at last necessarily arrive at a point where U = 0.

Against this proof a doubt can be cast. It may be possible, however far you may progress and although the value of U may always decrease, that this decrease may yet become continually slower, and that value may nowhere arrive at a limit just the same; which objection corresponds to the fourth one in article 6. But it is not difficult to assign some boundary so that as soon as you cross it the value of U must not only change ever faster, necessarily, but can also *decrease* no further, so that even before you arrive at that boundary the value of 0 must necessarily have occurred already. I reserve to myself pursing further, on another occasion, this matter and all that I have only been able to touch slightly in this proof.

Fig. 1.

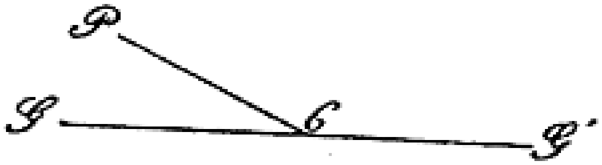

Fig. 2.

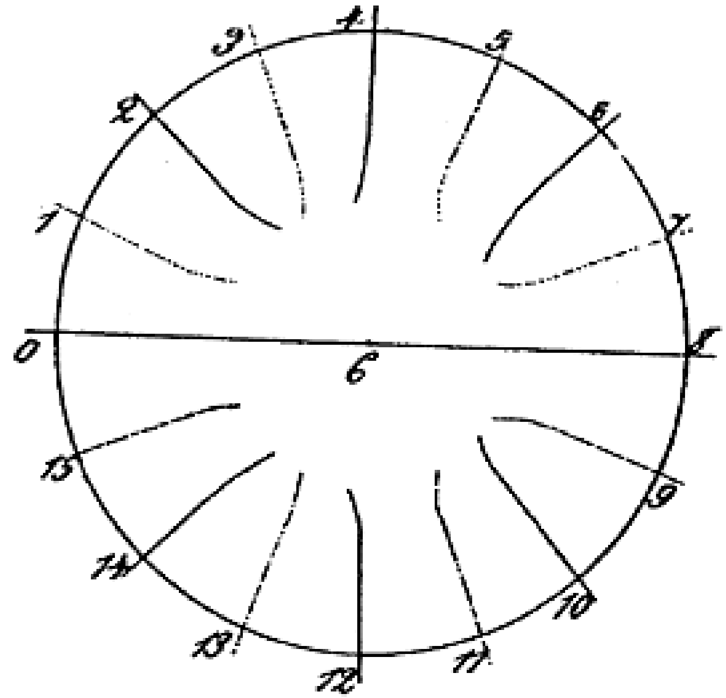

Fig. 4.

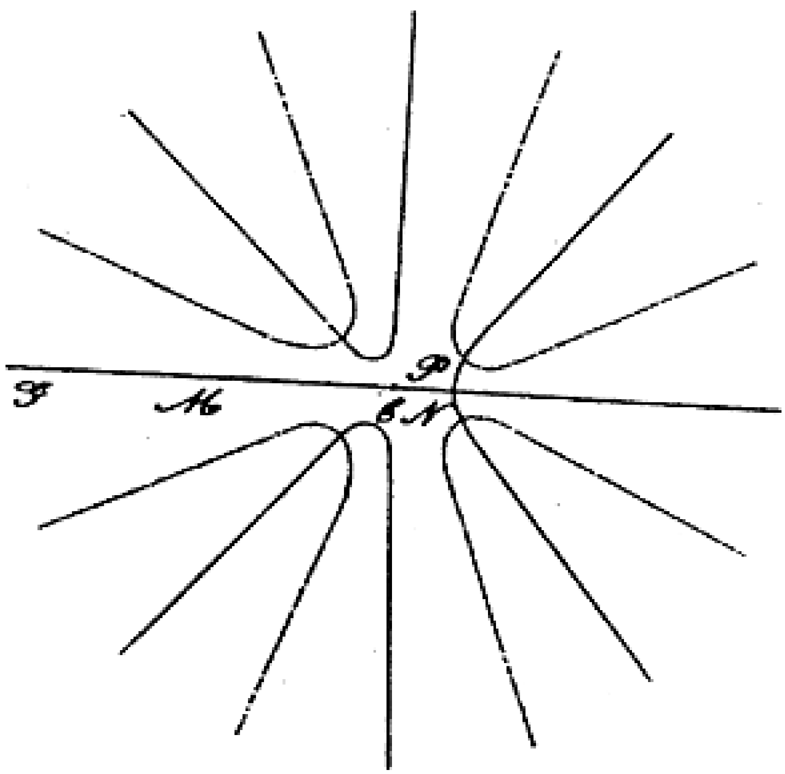

Fig. 3.