786.

EQUATION.

[From the *Encyclopædia Britannica, Ninth Edition*, vol. VIII. (1878), pp. 497—509.]

THE present article includes Determinant and Theory of Equations; and it may be proper to explain the relation to each other of the two subjects. Theory of Equations is used in its ordinary conventional sense to denote the theory of a single equation of any order in one unknown quantity; that is, it does not include the theory of a system or systems of equations of any order between any number of unknown quantities. Such systems occur very frequently in analytical geometry and other parts of mathematics, but they are hardly as yet the subject-matter of a distinct theory; and even Elimination, the transition-process for passing from a system of any number of equations involving the same number of unknown quantities to a single equation in one unknown quantity, hardly belongs to the Theory of Equations in the above restricted sense. But there is one case of a system of equations which precedes the Theory of Equations, and indeed presents itself at the outset of algebra, that of a system of simple (or linear) equations. Such a system gives rise to the function called a Determinant, and it is by means of these functions that the solution of the equations is effected. We have thus the subject Determinant as nearly equivalent to (but somewhat more extensive than) that of a system of linear equations; and we have the other subject, Theory of Equations, used in the restricted sense above referred to, and as not including Elimination.

Determinant.

1. A sketch of the history of determinants is given under [the Article] Algebra; it thereby appears that the algebraical function called a determinant presents itself in the solution of a system of simple equations, and we have herein a natural source of the theory. Thus, considering the equations

$$a\,x+b\,y+c\,z=d\,,$$
$$a'\,x+b'\,y+c'\,z=d'\,,$$
$$a''x+b''y+c''z=d''\,,$$

and proceeding to solve them by the so-called method of cross multiplication, we multiply the equations by factors selected in such a manner that, upon adding the results, the whole coefficient of y becomes $=0$ and the whole coefficient of z becomes $=0$; the factors in question are $b'c''-b''c'$, $b''c-bc''$, $bc'-b'c$ (values which, as at once seen, have the desired property); we thus obtain an equation which contains on the left-hand side only a multiple of x, and on the right-hand side a constant term; the coefficient of x has the value

$$a(b'c''-b''c')+a'(b''c-bc'')+a''(bc'-b'c),$$

and this function, represented in the form

$$\begin{vmatrix} a, & b, & c \\ a', & b', & c' \\ a'', & b'', & c'' \end{vmatrix},$$

is said to be a determinant; or, the number of elements being 3^2, it is called a determinant of the third order. It is to be noticed that the resulting equation is

$$\begin{vmatrix} a, & b, & c \\ a', & b', & c' \\ a'', & b'', & c'' \end{vmatrix} x = \begin{vmatrix} d, & b, & c \\ d', & b', & c' \\ d'', & b'', & c'' \end{vmatrix},$$

where the expression on the right-hand side is the like function with d, d', d'' in place of a, a', a'' respectively, and is of course also a determinant. Moreover, the functions $b'c''-b''c'$, $b''c-bc''$, $bc'-b'c$ used in the process are themselves the determinants of the second order

$$\begin{vmatrix} b', & c' \\ b'', & c'' \end{vmatrix}, \quad \begin{vmatrix} b'', & c'' \\ b, & c \end{vmatrix}, \quad \begin{vmatrix} b, & c \\ b', & c' \end{vmatrix}.$$

We have herein the suggestion of the rule for the derivation of the determinants of the orders 1, 2, 3, 4, &c., each from the preceding one, viz. we have

$$|a| = a,$$

$$\begin{vmatrix} a, & b \\ a', & b' \end{vmatrix} = a|b'| - a'|b|,$$

$$\begin{vmatrix} a, b, c \\ a', b', c' \\ a'', b'', c'' \end{vmatrix} = a\begin{vmatrix} b', c' \\ b'', c'' \end{vmatrix} + a'\begin{vmatrix} b'', c'' \\ b, c \end{vmatrix} + a''\begin{vmatrix} b, c \\ b', c' \end{vmatrix},$$

$$\begin{vmatrix} a, b, c, d \\ a', b', c', d' \\ a'', b'', c'', d'' \\ a''', b''', c''', d''' \end{vmatrix} = a\begin{vmatrix} b', c', d' \\ b'', c'', d'' \\ b''', c''', d''' \end{vmatrix} - a'\begin{vmatrix} b'', c'', d'' \\ b''', c''', d''' \\ b, c, d \end{vmatrix} + a''\begin{vmatrix} b''', c''', d''' \\ b, c, d \\ b', c', d' \end{vmatrix} - a'''\begin{vmatrix} b, c, d \\ b', c', d' \\ b'', c'', d'' \end{vmatrix},$$

and so on, the terms being all + for a determinant of an odd order, but alternately + and − for a determinant of an even order.

2. It is easy, by induction, to arrive at the general results:—

A determinant of the order n is the sum of the $1.2.3 \ldots n$ products which can be formed with n elements out of n^2 elements arranged in the form of a square, no two of the n elements being in the same line or in the same column, and each such product having the coefficient $\pm$ unity.

The products in question may be obtained by permuting in every possible manner the columns (or the lines) of the determinant, and then taking for the factors the n elements in the dexter diagonal. And we thence derive the rule for the signs, viz. considering the primitive arrangement of the columns as positive, then an arrangement obtained therefrom by a single interchange (inversion, or derangement) of two columns is regarded as negative; and so in general an arrangement is positive or negative according as it is derived from the primitive arrangement by an even or an odd number of interchanges. This implies the theorem that a given arrangement can be derived from the primitive arrangement only by an odd number, or else only by an even number of interchanges,—a theorem the verification of which may be easily obtained from the theorem (in fact, a particular case of the general one), an arrangement can be derived from itself only by an even number of interchanges. And this being so, each product has the sign belonging to the corresponding arrangement of the columns; in particular, a determinant contains with the sign + the product of the elements in its dexter diagonal. It is to be observed that the rule gives as many positive as negative arrangements, the number of each being $= \frac{1}{2}.1.2 \ldots n$.

The rule of signs may be expressed in a different form. Giving to the columns in the primitive arrangement the numbers 1, 2, 3, ..., n, to obtain the sign belonging to any other arrangement we take, as often as a lower number succeeds a higher one, the sign $-$, and, compounding together all these minus signs, obtain the proper sign, + or $-$ as the case may be.

Thus, for three columns, it appears by either rule that 123, 231, 312 are positive; 132, 321, 213 are negative; and the developed expression of the foregoing determinant of the third order is

$$= ab'c'' - ab''c' + a'b''c - a'bc'' + a''bc' - a''b'c.$$

3. It further appears that a determinant is a linear function* of the elements of each column thereof, and also a linear function of the elements of each line thereof; moreover, that the determinant retains the same value, only its sign being altered, when any two columns are interchanged, or when any two lines are interchanged; more generally, when the columns are permuted in any manner, or when the lines are permuted in any manner, the determinant retains its original value, with the sign + or $-$ according as the new arrangement (considered as derived from the primitive arrangement) is positive or negative according to the foregoing rule of signs.

* The expression, a linear function, is here used in its narrowest sense, a linear function without constant term; what is meant is, that the determinant is in regard to the elements a, a', a'', ... of any column or line thereof, a function of the form $Aa + A'a' + A''a'' + \ldots$, without any term independent of a, a', a'', ...

It at once follows that, if two columns are identical, or if two lines are identical, the value of the determinant is $=0$. It may be added that, if the lines are converted into columns, and the columns into lines, in such a way as to leave the dexter diagonal unaltered, the value of the determinant is unaltered; the determinant is in this case said to be *transposed*.

4. By what precedes it appears that there exists a function of the n^2 elements, linear as regards the terms of each column (or say, for shortness, linear as to each column), and such that only the sign is altered when any two columns are interchanged; these properties completely determine the function, except as to a common factor which may multiply all the terms. If, to get rid of this arbitrary common factor, we assume that the product of the elements in the dexter diagonal has the coefficient $+1$, we have a complete definition of the determinant; and it is interesting to show how from these properties, assumed for the definition of the determinant, it at once appears that the determinant is a function serving for the solution of a system of linear equations. Observe that the properties show at once that if any column is $=0$ (that is, if the elements in the column are each $=0$), then the determinant is $=0$; and further that, if any two columns are identical, then the determinant is $=0$.

5. Reverting to the system of linear equations written down at the beginning of this article, consider the determinant

$$\begin{vmatrix} a\,x+b\,y+c\,z-d\,, & b\,, & c \\ a'x+b'y+c'z-d', & b', & c' \\ a''x+b''y+c''z-d'', & b'', & c'' \end{vmatrix};$$

it appears that this is

$$=x\begin{vmatrix} a\,, & b\,, & c \\ a', & b', & c' \\ a'', & b'', & c'' \end{vmatrix}+y\begin{vmatrix} b\,, & b\,, & c \\ b', & b', & c' \\ b'', & b'', & c'' \end{vmatrix}+z\begin{vmatrix} c\,, & b\,, & c \\ c', & b', & c' \\ c'', & b'', & c'' \end{vmatrix}-\begin{vmatrix} d\,, & b\,, & c \\ d', & b', & c' \\ d'', & b'', & c'' \end{vmatrix},$$

viz. the second and the third terms each vanishing, it is

$$=x\begin{vmatrix} a\,, & b\,, & c \\ a', & b', & c' \\ a'', & b'', & c'' \end{vmatrix}-\begin{vmatrix} d\,, & b\,, & c \\ d', & b', & c' \\ d'', & b'', & c'' \end{vmatrix}.$$

But if the linear equations hold good, then the first column of the original determinant is $=0$, and therefore the determinant itself is $=0$; that is, the linear equations give

$$x\begin{vmatrix} a\,, & b\,, & c \\ a', & b', & c' \\ a'', & b'', & c'' \end{vmatrix}-\begin{vmatrix} d\,, & b\,, & c \\ d', & b', & c' \\ d'', & b'', & c'' \end{vmatrix}=0;$$

which is the result obtained above.

We might in a similar way find the values of y and z, but there is a more symmetrical process. Join to the original equations the new equation

$$\alpha x + \beta y + \gamma z = \delta;$$

a like process shows that, the equations being satisfied, we have

$$\begin{vmatrix} \alpha, & \beta, & \gamma, & \delta \\ a, & b, & c, & d \\ a', & b', & c', & d' \\ a'', & b'', & c'', & d'' \end{vmatrix} = 0;$$

or, as this may be written,

$$\begin{vmatrix} \alpha, & \beta, & \gamma & \\ a, & b, & c, & d \\ a', & b', & c', & d' \\ a'', & b'', & c'', & d'' \end{vmatrix} - \delta \begin{vmatrix} a, & b, & c \\ a', & b', & c' \\ a'', & b'', & c'' \end{vmatrix} = 0;$$

which, considering δ as standing herein for its value $\alpha x + \beta y + \gamma z$, is a consequence of the original equations only. We have thus an expression for $\alpha x + \beta y + \gamma z$, an arbitrary linear function of the unknown quantities x, y, z; and by comparing the coefficients of α, β, γ on the two sides respectively, we have the values of x, y, z; in fact, these quantities, each multiplied by

$$\begin{vmatrix} a, & b, & c \\ a', & b', & c' \\ a'', & b'', & c'' \end{vmatrix},$$

are in the first instance obtained in the forms

$$\begin{vmatrix} 1 & & & \\ a, & b, & c, & d \\ a', & b', & c', & d' \\ a'', & b'', & c'', & d'' \end{vmatrix}, \quad \begin{vmatrix} & 1 & & \\ a, & b, & c, & d \\ a', & b', & c', & d' \\ a'', & b'', & c'', & d'' \end{vmatrix}, \quad \begin{vmatrix} & & 1 & \\ a, & b, & c, & d \\ a', & b', & c', & d' \\ a'', & b'', & c'', & d'' \end{vmatrix};$$

but these are

$$= \begin{vmatrix} b, & c, & d \\ b', & c', & d' \\ b'', & c'', & d'' \end{vmatrix}, \quad - \begin{vmatrix} c, & d, & a \\ c', & d', & a' \\ c'', & d'', & a'' \end{vmatrix}, \quad \begin{vmatrix} d, & a, & b \\ d', & a', & b' \\ d'', & a'', & b'' \end{vmatrix},$$

or, what is the same thing,

$$= \begin{vmatrix} b, & c, & d \\ b', & c', & d' \\ b'', & c'', & d'' \end{vmatrix}, \quad \begin{vmatrix} c, & a, & d \\ c', & a', & d' \\ c'', & a'', & d'' \end{vmatrix}, \quad \begin{vmatrix} a, & b, & d \\ a', & b', & d' \\ a'', & b'', & d'' \end{vmatrix}$$

respectively.

6. *Multiplication of two determinants of the same order.*—The theorem is obtained very easily from the last preceding definition of a determinant. It is most simply expressed thus—

$$\begin{array}{c|ccc|} & (\alpha,\ \alpha',\ \alpha''),\ (\beta,\ \beta',\ \beta''),\ (\gamma,\ \gamma',\ \gamma'') & \\ (a\ ,\ b\ ,\ c\ ) & \text{,,} & \text{,,} & \text{,,} \\ (a',\ b',\ c') & \text{,,} & \text{,,} & \text{,,} \\ (a'',\ b'',\ c'') & \text{,,} & \text{,,} & \text{,,} \end{array} = \begin{vmatrix} a\ , & b\ , & c \\ a', & b', & c' \\ a'', & b'', & c'' \end{vmatrix} \cdot \begin{vmatrix} \alpha\ , & \beta\ , & \gamma \\ \alpha', & \beta', & \gamma' \\ \alpha'', & \beta'', & \gamma'' \end{vmatrix},$$

where the expression on the left side stands for a determinant, the terms of the first line being $(a,\ b,\ c)(\alpha,\ \alpha',\ \alpha'')$, that is, $a\alpha + b\alpha' + c\alpha''$, $(a,\ b,\ c)(\beta,\ \beta',\ \beta'')$, that is, $a\beta + b\beta' + c\beta''$, $(a,\ b,\ c)(\gamma,\ \gamma',\ \gamma'')$, that is, $a\gamma + b\gamma' + c\gamma''$; and similarly the terms in the second and third lines are the like functions with $(a',\ b',\ c')$ and $(a'',\ b'',\ c'')$ respectively.

There is an apparently arbitrary transposition of lines and columns; the result would hold good if on the left-hand side we had written $(\alpha,\ \beta,\ \gamma)$, $(\alpha',\ \beta',\ \gamma')$, $(\alpha'',\ \beta'',\ \gamma'')$, or what is the same thing, if on the right-hand side we had transposed the second determinant; and either of these changes would, it might be thought, increase the elegance of the form, but, for a reason which need not be explained*, the form actually adopted is the preferable one.

To indicate the method of proof, observe that the determinant on the left-hand side, *qua* linear function of its columns, may be broken up into a sum of $(3^3 =)\ 27$ determinants, each of which is either of some such form as

$$\pm\alpha\beta\gamma' \begin{vmatrix} a\ , & a\ , & b \\ a', & a', & b' \\ a'', & a'', & b'' \end{vmatrix},$$

where the term $\alpha\beta\gamma'$ is not a term of the $\alpha\beta\gamma$-determinant, and its coefficient (as a determinant with two identical columns) vanishes; or else it is of a form such as

$$\pm\alpha\beta'\gamma'' \begin{vmatrix} a\ , & b\ , & c \\ a', & b', & c' \\ a'', & b'', & c'' \end{vmatrix},$$

that is, every term which does not vanish contains as a factor the *abc*-determinant last written down; the sum of all other factors $\pm\alpha\beta'\gamma''$ is the $\alpha\beta\gamma$-determinant of the formula; and the final result then is, that the determinant on the left-hand side is equal to the product on the right-hand side of the formula.

7. *Decomposition of a determinant into complementary determinants.*—Consider, for simplicity, a determinant of the fifth order, $5 = 2 + 3$, and let the top two lines be

$$\begin{array}{ccccc} a\ , & b\ , & c\ , & d\ , & e\ , \\ a'\ , & b'\ , & c'\ , & d'\ , & e'; \end{array}$$

* The reason is the connexion with the corresponding theorem for the multiplication of two matrices.

then, if we consider how these elements enter into the determinant, it is at once seen that they enter only through the determinants of the second order $\begin{vmatrix} a, & b \\ a', & b' \end{vmatrix}$, &c., which can be formed by selecting any two columns at pleasure. Moreover, representing the remaining three lines by

$$\begin{matrix} a'', & b'', & c'', & d'', & e'', \\ a''', & b''', & c''', & d''', & e''', \\ a'''', & b'''', & c'''', & d'''', & e'''', \end{matrix}$$

it is further seen that the factor which multiplies the determinant formed with any two columns of the first set is the determinant of the third order formed with the complementary three columns of the second set; and it thus appears that the determinant of the fifth order is a sum of all the products of the form

$$\pm \begin{vmatrix} a, & b \\ a', & b' \end{vmatrix} \cdot \begin{vmatrix} c'', & d'', & e'' \\ c''', & d''', & e''' \\ c'''', & d'''', & e'''' \end{vmatrix},$$

the sign $\pm$ being in each case such that the sign of the term $\pm ab' . c''d'''e''''$ obtained from the diagonal elements of the component determinants may be the actual sign of this term in the determinant of the fifth order; for the product written down the sign is obviously $+$.

Observe that for a determinant of the nth order, taking the decomposition to be $1+(n-1)$, we fall back upon the equations given at the commencement, in order to show the genesis of a determinant.

8. Any determinant $\begin{vmatrix} a, & b \\ a', & b' \end{vmatrix}$ formed out of the elements of the original determinant, by selecting the lines and columns at pleasure, is termed a *minor* of the original determinant; and when the number of lines and columns, or order of the determinant, is $n-1$, then such determinant is called a *first minor;* the number of the first minors is $=n^2$, the first minors, in fact, corresponding to the several elements of the determinant—that is, the coefficient therein of any term whatever is the corresponding first minor. The first minors, each divided by the determinant itself, form a system of elements *inverse* to the elements of the determinant.

A determinant is *symmetrical* when every two elements symmetrically situated in regard to the dexter diagonal are equal to each other; if they are equal and opposite (that is, if the sum of the two elements be $=0$), this relation not extending to the diagonal elements themselves, which remain arbitrary, then the determinant is *skew;* but if the relation does extend to the diagonal terms (that is, if these are each $=0$), then the determinant is *skew symmetrical;* thus the determinants

$$\begin{vmatrix} a, & h, & g \\ h, & b, & f \\ g, & f, & c \end{vmatrix}, \quad \begin{vmatrix} a, & \nu, & -\mu \\ -\nu, & b, & \lambda \\ \mu, & -\lambda, & c \end{vmatrix}, \quad \begin{vmatrix} 0, & \nu, & -\mu \\ -\nu, & 0, & \lambda \\ \mu, & -\lambda, & 0 \end{vmatrix},$$

are respectively symmetrical, skew, and skew symmetrical.

The theory admits of very extensive algebraic developments, and applications in algebraical geometry and other parts of mathematics; but the fundamental properties of the functions may fairly be considered as included in what precedes.

Theory of Equations.

9. In the subject "Theory of Equations," the term *equation* is used to denote an equation of the form $x^n - p_1 x^{n-1} + \ldots \pm p_n = 0$, where $p_1, p_2, \ldots, p_n$ are regarded as known, and x as a quantity to be determined; for shortness, the equation is written $f(x) = 0$.

The equation may be *numerical;* that is, the coefficients $p_1, p_2, \ldots, p_n$ are then numbers,—understanding by number a quantity of the form $\alpha + \beta i$ (α and β having any positive or negative real values whatever, or say each of these is regarded as susceptible of continuous variation from an indefinitely large negative to an indefinitely large positive value), and i denoting $\sqrt{-1}$.

Or the equation may be *algebraical;* that is, the coefficients are not then restricted to denote, or are not explicitly considered as denoting, numbers.

I. We consider first numerical equations. (Real theory, 10 to 14; Imaginary theory, 15 to 18.)

10. Postponing all consideration of imaginaries, we take in the first instance the coefficients to be real, and attend only to the real roots (if any); that is, $p_1, p_2, \ldots, p_n$ are real positive or negative quantities, and a root a, if it exists, is a positive or negative quantity such that $a^n - p_1 a^{n-1} + \ldots \pm p_n = 0$, or say, $f(a) = 0$. The fundamental theorems are given in the article Algebra, sections X., XIII., XIV.; but there are various points in the theory which require further development.

It is very useful to consider the curve $y = f(x)$,—or, what would come to the same, the curve $Ay = f(x)$,—but it is better to retain the first-mentioned form of equation, drawing, if need be, the ordinate y on a reduced scale. For instance, if the given equation be $x^3 - 6x^2 + 11x - 6{\cdot}06 = 0$,* then the curve $y = x^3 - 6x^2 + 11x - 6{\cdot}06$ is as shown in the figure at page 501, without any reduction of scale for the ordinate.

It is clear that, in general, y is a continuous one-valued function of x, finite for every finite value of x, but becoming infinite when x is infinite; i.e. assuming throughout that the coefficient of x^n is $+1$, then when $x = \infty$, $y = +\infty$; but when $x = -\infty$, then $y = +\infty$ or $-\infty$, according as n is even or odd; the curve cuts any line whatever, and in particular it cuts the axis of x, in at most n points; and the value of x, at any point of intersection with the axis, is a root of the equation $f(x) = 0$.

If β, α are any two values of x ($\alpha > \beta$, that is, α nearer $+\infty$), then if $f(\beta)$, $f(\alpha)$ have opposite signs, the curve cuts the axis an odd number of times, and therefore at least once, between the points $x = \beta$, $x = \alpha$; but if $f(\beta)$, $f(\alpha)$ have the same sign, then between these points the curve cuts the axis an even number of times, or it may be not at all. That is, $f(\beta)$, $f(\alpha)$ having opposite signs, there are between the limits β, α an odd number of real roots, and therefore at least one real

* The coefficients were selected so that the roots might be nearly 1, 2, 3.

root; but $f(\beta)$, $f(\alpha)$ having the same sign, there are between these limits an even number of real roots, or it may be there is no real root. In particular, by giving to β, α the values $-\infty$, $+\infty$ (or, what is the same thing, any two values sufficiently near to these values respectively) it appears that an equation of an odd order has always an odd number of real roots, and therefore at least one real root; but that an equation of an even order has an even number of real roots, or it may be no real root.

If α be such that for $x =$ or $> \alpha$ (that is, x nearer to $+\infty$) $f(x)$ is always $+$, and β be such that for $x =$ or $< \beta$ (that is, x nearer to $-\infty$) $f(x)$ is always $-$, then the real roots (if any) lie between these limits $x = \beta$, $x = \alpha$; and it is easy to find by trial such two limits including between them all the real roots (if any).

11. Suppose that the positive value δ is an inferior limit to the difference between two real roots of the equation; or rather (since the foregoing expression would imply the existence of real roots) suppose that there are not two real roots such that their difference taken positively is $=$ or $< \delta$; then, γ being any value whatever, there is clearly at most one real root between the limits γ and $\gamma + \delta$; and by what precedes there is such real root or there is not such real root, according as $f(\gamma)$, $f(\gamma + \delta)$ have opposite signs or have the same sign. And by dividing in this manner the interval β to α into intervals each of which is $=$ or $< \delta$, we should not only ascertain the number of the real roots (if any), but we should also *separate* the real roots, that is, find for each of them limits γ, $\gamma + \delta$ between which there lies this one, and only this one, real root.

In particular cases it is frequently possible to ascertain the number of the real roots, and to effect their separation by trial or otherwise, without much difficulty; but the foregoing was the general process as employed by Lagrange even in the second edition (1808) of the *Traité de la résolution des Équations Numériques**; the determination of the limit δ had to be effected by means of the "equation of differences" or equation of the order $\frac{1}{2}n(n-1)$, the roots of which are the squares of the differences of the roots of the given equation, and the process is a cumbrous and unsatisfactory one.

12. The great step was effected by Sturm's theorem (1835)—viz. here starting from the function $f(x)$, and its first derived function $f'(x)$, we have (by a process which is a slight modification of that for obtaining the greatest common measure of these two functions) to form a series of functions

$$f(x),\ f'(x),\ f_2(x), \ldots, f_n(x)$$

of the degrees n, $n-1$, $n-2, \ldots, 0$ respectively,—the last term $f_n(x)$ being thus an absolute constant. These lead to the immediate determination of the number of real roots (if any) between any two given limits β, α; viz. supposing $\alpha > \beta$ (that is, α nearer to $+\infty$), then substituting successively these two values in the series of functions, and attending only to the signs of the resulting values, the number of the changes of sign lost in passing from β to α is the required number of real roots

* The third edition (1826) is a reproduction of that of 1808; the first edition has the date 1798, but a large part of the contents is taken from memoirs of 1767—68 and 1770—71.

between the two limits. In particular, taking β, $\alpha = -\infty, +\infty$ respectively, the signs of the several functions depend merely on the signs of the terms which contain the highest powers of x, and are seen by inspection, and the theorem thus gives at once the whole number of real roots.

And although theoretically, in order to complete by a finite number of operations the separation of the real roots, we still need to know the value of the before-mentioned limit δ; yet in any given case the separation may be effected by a limited number of repetitions of the process. The practical difficulty is when two or more roots are very near to each other. Suppose, for instance, that the theorem shows that there are two roots between 0 and 10; by giving to x the values 1, 2, 3, ... successively, it might appear that the two roots were between 5 and 6; then again that they were between 5·3 and 5·4, then between 5·34 and 5·35, and so on until we arrive at a separation; say it appears that between 5·346 and 5·347 there is one root, and between 5·348 and 5·349 the other root. But in the case in question δ would have a very small value, such as ·002, and even supposing this value known, the direct application of the first-mentioned process would be still more laborious.

13. Supposing the separation once effected, the determination of the single real root which lies between the two given limits may be effected to any required degree of approximation either by the processes of Horner and Lagrange (which are in principle a carrying out of the method of Sturm's theorem), or by the process of Newton, as perfected by Fourier (which requires to be separately considered).

First as to Horner and Lagrange. We know that between the limits β, α there lies one, and only one, real root of the equation; $f(\beta)$ and $f(\alpha)$ have therefore opposite signs. Suppose any intermediate value is θ; in order to determine by Sturm's theorem whether the root lies between β, θ, or between θ, α, it would be quite unnecessary to calculate the signs of $f(\theta), f'(\theta), f_2(\theta), \ldots$; only the sign of $f(\theta)$ is required: for, if this has the same sign as $f(\beta)$, then the root is between β, θ; if the same sign as $f(\alpha)$, then the root is between θ, α. We want to make θ increase from the inferior limit β, at which $f(\theta)$ has the sign of $f(\beta)$, so long as $f(\theta)$ retains this sign, and then to a value for which it assumes the opposite sign; we have thus two nearer limits of the required root, and the process may be repeated indefinitely.

Horner's method (1819) gives the root as a decimal, figure by figure; thus, if the equation be known to have one real root between 0 and 10, it is in effect shown say that 5 is too small (that is, the root is between 5 and 6); next that 5·4 is too small (that is, the root is between 5·4 and 5·5); and so on to any number of decimals. Each figure is obtained, *not* by the successive trial of all the figures which precede it, but (as in the ordinary process of the extraction of a square root, which is in fact Horner's process applied to this particular case) it is given presumptively as the first figure of a quotient; such value may be too large, and then the next inferior integer must be tried instead of it, or it may require to be further diminished. And it is to be remarked that the process not only gives the approximate value α of the root, but (as in the extraction of a square root) it includes the calculation of the function $f(\alpha)$ which should be, and approximately is, $= 0$. The arrangement of the

calculations is very elegant, and forms an integral part of the actual method. It is to be observed that after a certain number of decimal places have been obtained, a good many more can be found by a mere division. It is in the progress tacitly assumed that the roots have been first separated.

Lagrange's method (1767) gives the root as a continued fraction $a+\frac{1}{b}+\frac{1}{c}+\dots$, where a is a positive or negative integer (which may be $=0$), but $b, c, \dots$ are positive integers. Suppose the roots have been separated; then (by trial if need be of consecutive integer values) the limits may be made to be consecutive integer numbers: say they are $a, a+1$; the value of x is therefore $=a+\frac{1}{y}$, where y is positive and greater than 1; from the given equation for x, writing therein $x=a+\frac{1}{y}$, we form an equation of the same order for y, and this equation will have one, and only one, positive root greater than 1; hence finding for it the limits $b, b+1$ (where b is $=$ or >1), we have $y=b+\frac{1}{z}$, where z is positive and greater than 1; and so on—that is, we thus obtain the successive denominators $b, c, d, \dots$ of the continued fraction. The method is theoretically very elegant, but the disadvantage is that it gives the result in the form of a continued fraction, which for the most part must ultimately be converted into a decimal. There is one advantage in the method, that a commensurable root (that is, a root equal to a rational fraction) is found accurately, since, when such root exists, the continued fraction terminates.

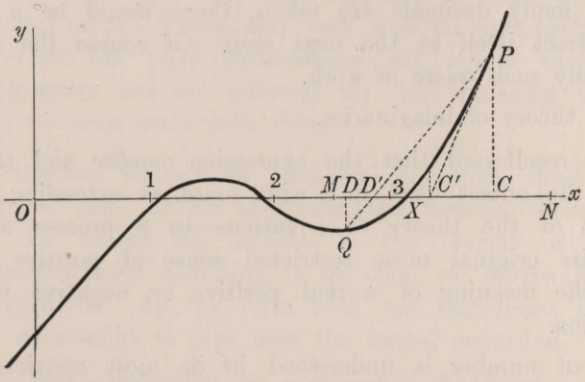
14. Newton's method (1711), as perfected by Fourier (1831), may be roughly stated as follows. If $x=\gamma$ be an approximate value of any root, and $\gamma+h$ the correct value, then $f(\gamma+h)=0$, that is,

$$f(\gamma)+\frac{h}{1}f'(\gamma)+\frac{h^2}{1.2}f''(\gamma)+\dots=0;$$

and then, if h be so small that the terms after the second may be neglected, $f(\gamma)+hf'(\gamma)=0$, that is, $h=-\frac{f(\gamma)}{f'(\gamma)}$, or the new approximate value is $x=\gamma-\frac{f(\gamma)}{f'(\gamma)}$; and so on, as often as we please. It will be observed that so far nothing has been assumed as to the separation of the roots, or even as to the existence of a real root; γ has been taken as the approximate value of a root, but no precise meaning has been attached to this expression. The question arises, what are the conditions to be satisfied by γ in order that the process may by successive repetitions actually lead to a certain real root of the equation; or say that, γ being an approximate value of a certain real root, the new value $\gamma-\frac{f(\gamma)}{f'(\gamma)}$ may be a more approximate value.

Referring to the figure, it is easy to see that, if OC represent the assumed value γ, then, drawing the ordinate CP to meet the curve in P, and the tangent PC' to meet the axis in C', we shall have OC' as the new approximate value of the root. But observe that there is here a real root OX, and that the curve beyond X

is convex to the axis; under these conditions the point C' is nearer to X than was C; and, starting with C' instead of C, and proceeding in like manner to draw a new ordinate and tangent, and so on as often as we please, we approximate continually, and that with great rapidity, to the true value OX. But if C had been taken on the other side of X, where the curve is concave to the axis, the new point C' might or might not be nearer to X than was the point C; and in this



case the method, if it succeeds at all, does so by accident only, i.e., it may happen that C' or some subsequent point comes to be a point C, such that OC is a *proper* approximate value of the root, and then the subsequent approximations proceed in the same manner as if this value had been assumed in the first instance, all the preceding work being wasted. It thus appears that for the proper application of the method we require *more* than the mere separation of the roots. In order to be able to approximate to a certain root a, $= OX$, we require to know that, between OX and some value ON, the curve is always convex to the axis: analytically, between the two values, $f(x)$ and $f''(x)$ must have always the same sign. When this is so, the point C may be taken anywhere on the proper side of X, and within the portion XN of the axis; and the process is then the one already explained. The approximation is in general a very rapid one. If we know for the required root OX the two limits OM, ON such that from M to X the curve is always *concave* to the axis, while from X to N it is always convex to the axis,—then, taking D anywhere in the portion MX and (as before) C in the portion XN, drawing the ordinates DQ, CP, and joining the points P, Q by a line which meets the axis in D', also constructing the point C' by means of the tangent at P as before, we have for the required root the new limits OD', OC'; and proceeding in like manner with the points D', C', and so on as often as we please, we obtain at each step two limits approximating more and more nearly to the required root OX. The process as to the point D', translated into analysis, is the ordinate process of interpolation. Suppose $OD = \beta$, $OC = \alpha$, we have approximately

$$f(\beta + h) = f(\beta) + \frac{h\{f(\alpha) - f(\beta)\}}{\alpha - \beta},$$

whence, if the root is $\beta + h$, then

$$h = -\frac{(\alpha - \beta) f(\beta)}{f(\alpha) - f(\beta)}.$$

Returning for a moment to Horner's method, it may be remarked that the correction h, to an approximate value α, is therein found as a quotient, the same or such as the quotient $f(\alpha) \div f'(\alpha)$ which presents itself in Newton's method. The difference is that with Horner the integer part of this quotient, is taken as the presumptive value of h, and the figure is verified at each step. With Newton the quotient itself, developed to the proper number of decimal places, is taken as the value of h; if too many decimals are taken, there would be a waste of work; but the error would correct itself at the next step. Of course the calculation should be conducted without any such waste of work.

Next as to the theory of imaginaries.

15. It will be recollected that the expression *number* and the correlative epithet *numerical* were at the outset used in a wide sense, as extending to imaginaries. This extension arises out of the theory of equations by a process analogous to that by which number, in its original most restricted sense of positive integer number, was extended to have the meaning of a real positive or negative magnitude susceptible of continuous variation.

If for a moment number is understood in its most restricted sense as meaning positive integer number, the solution of a simple equation leads to an extension; $ax - b = 0$, gives $x = \frac{b}{a}$, a positive fraction, and we can in this manner represent, not accurately, but as nearly as we please, any positive magnitude whatever; so an equation $ax + b = 0$ gives $x = -\frac{b}{a}$, which (approximately as before) represents any negative magnitude. We thus arrive at the extended signification of number as a continuously varying positive or negative magnitude. Such numbers may be added or subtracted, multiplied or divided one by another, and the result is always a number. Now from a quadric equation we derive, in like manner, the notion of a complex or imaginary number such as is spoken of above. The equation $x^2 + 1 = 0$ is not (in the foregoing sense, number = real number) satisfied by any numerical value whatever of x; but we assume that there is a number which we call i, satisfying the equation $i^2 + 1 = 0$; and then taking a and b any real numbers, we form an expression such as $a + bi$, and use the expression number in this extended sense: any two such numbers may be added or subtracted, multiplied or divided one by the other, and the result is always a number. And if we consider first a quadric equation $x^2 + px + q = 0$ where p and q are real numbers, and next the like equation, where p and q are any numbers whatever, it can be shown that there exists for x a numerical value which satisfies the equation; or, in other words, it can be shown that the equation has a numerical root. The like theorem, in fact, holds good for an equation of any order whatever. But suppose for a moment that this was not the case: say that there was a cubic equation $x^3 + px^2 + qx + r = 0$, with numerical coefficients, not satisfied by any numerical value of x, we should have to establish a new imaginary j satisfying some such equation, and should then have to consider numbers of the form $a + bj$, or perhaps $a + bj + cj^2$ (a, b, c numbers $\alpha + \beta i$ of the kind heretofore considered),—first we should be thrown back on the quadric equation $x^2 + px + q = 0$, p and q being now numbers

of the last-mentioned extended form—*non constat* that every such equation has a numerical root—and if not, we might be led to *other* imaginaries k, l, &c., and so on *ad infinitum* in inextricable confusion.

But in fact a numerical equation of any order whatever has always a numerical root, and thus numbers (in the foregoing sense, number = quantity of the form $\alpha + \beta i$) form (*what real numbers do not*) a universe complete in itself, such that starting in it we are never led out of it. There may very well be, and perhaps are, numbers in a more general sense of the term (quaternions are not a case in point, as the ordinary laws of combination are not adhered to): but in order to have to do with such numbers (if any), we must start with them.

16. The capital theorem as regards numerical equations thus is, every numerical equation has a numerical root; or for shortness (the meaning being as before), every equation has a root. Of course the theorem is the reverse of self-evident, and it requires proof; but provisionally assuming it as true, we derive from it the general theory of numerical equations. As the term root was introduced in the course of an explanation, it will be convenient to give here the formal definition.

A number a such that substituted for x it makes the function $x^n - p_1 x^{n-1} + \ldots \pm p_n$ to be $= 0$, or say such that it satisfies the equation $f(x) = 0$, is said to be a root of the equation; that is, a being a root, we have

$$a^n - p_1 a^{n-1} + \ldots \pm p_n = 0, \text{ or say } f(a) = 0;$$

and it is then easily shown that $x - a$ is a factor of the function $f(x)$, viz. that we have $f(x) = (x-a) f_1(x)$, where $f_1(x)$ is a function $x^{n-1} - q_1 x^{n-2} + \ldots \pm q_{n-1}$ of the order $n-1$, with numerical coefficients $q_1, q_2, \ldots, q_{n-1}$.

In general, a is not a root of the equation $f_1(x) = 0$, but it may be so—i.e., $f_1(x)$ may contain the factor $x - a$; when this is so, $f(x)$ will contain the factor $(x-a)^2$; writing then $f(x) = (x-a)^2 f_2(x)$, and assuming that a is not a root of the equation $f_2(x) = 0$, $x = a$ is then said to be a double root of the equation $f(x) = 0$; and similarly $f(x)$ may contain the factor $(x-a)^3$ and no higher power, and $x = a$ is then a triple root; and so on.

Supposing, in general, that $f(x) = (x-a)^\alpha F(x)$, α being a positive integer which may be $= 1$, $(x-a)^\alpha$ the highest power of $x - a$ which divides $f(x)$, and $F(x)$ being of course of the order $n - \alpha$, then the equation $F(x) = 0$ will have a root b which will be different from a; $x - b$ will be a factor, in general a simple one, but it may be a multiple one, of $F(x)$, and $f(x)$ will in this case be $= (x-a)^\alpha (x-b)^\beta \Phi(x)$, β a positive integer which may be $= 1$, $(x-b)^\beta$ the highest power of $x - b$ in $F(x)$ or $f(x)$, and $\Phi(x)$ being of course of the order $n - \alpha - \beta$. The original equation $f(x) = 0$ is in this case said to have α roots each $= a$, β roots each $= b$; and so on for any other factors $(x-c)^\gamma$, &c.

We have thus the *theorem*—A numerical equation of the order n has in every case n roots, viz. there exist n numbers $a, b, \ldots$, in general all distinct, but which may arrange themselves in any sets of equal values, such that $f(x) = (x-a)(x-b)(x-c)\ldots$ identically.

If the equation has equal roots, these can in general be determined: and the case is at any rate a special one which may be in the first instance excluded from consideration. It is therefore, in general, assumed that the equation $f(x) = 0$ has all its roots unequal.

If the coefficients $p_1, p_2, \ldots$ are all or any one or more of them imaginary, then the equation $f(x) = 0$, separating the real and imaginary parts thereof, may be written $F(x) + i\Phi(x) = 0$, where $F(x)$, $\Phi(x)$ are each of them a function with real coefficients; and it thus appears that the equation $f(x) = 0$, with imaginary coefficients, has not in general any real root; supposing it to have a real root a, this must be at once a root of each of the equations $F(x) = 0$ and $\Phi(x) = 0$.

But an equation with real coefficients may have as well imaginary as real roots, and we have further the theorem that for any such equation the imaginary roots enter in pairs, viz. $\alpha + \beta i$ being a root, then $\alpha - \beta i$ will be also a root. It follows that, if the order be odd, there is always an odd number of real roots, and therefore at least one real root.

17. In the case of an equation with real coefficients, the question of the existence of real roots, and of their separation, has been already considered. In the general case of an equation with imaginary (it may be real) coefficients, the like question arises as to the situation of the (real or imaginary) roots; thus if, for facility of conception, we regard the constituents α, β of a root $\alpha + \beta i$ as the coordinates of a point *in plano*, and accordingly represent the root by such point, then drawing in the plane any closed curve or "contour," the question is how many roots lie within such contour.

This is solved theoretically by means of a theorem of Cauchy's (1837), viz. writing in the original equation $x + iy$ in place of x, the function $f(x + iy)$ becomes $= P + iQ$, where P and Q are each of them a rational and integral function (with real coefficients) of (x, y). Imagining the point (x, y) to travel along the contour, and considering the number of changes of sign from $-$ to $+$ and from $+$ to $-$ of the fraction corresponding to passages of the fraction through zero, that is, to values for which P becomes $= 0$, disregarding those for which Q becomes $= 0$, the difference of these numbers gives the number of roots within the contour.

It is important to remark that the demonstration does not presuppose the existence of any root; the contour may be the infinity of the plane (such infinity regarded as a contour, or closed curve), and in this case it can be shown (and that very easily) that the difference of the numbers of changes of sign is $= n$; that is, there are within the infinite contour, or (what is the same thing) there are in all, n roots; thus Cauchy's theorem contains really the proof of the fundamental theorem that a numerical equation of the nth order (not only has a numerical root, but) has precisely n roots. It would appear that this proof of the fundamental theorem in its most complete form is in principle identical with Gauss's last proof (1849) of the theorem, in the form—A numerical equation of the nth order has always a root*.

* The earlier demonstrations by Euler, Lagrange, &c., relate to the case of a numerical equation with real coefficients; and they consist in showing that such equation has always a real quadratic divisor, furnishing two roots, which are either real or else conjugate imaginaries $\alpha + \beta i$: see Lagrange's *Équations Numériques*.

But in the case of a finite contour, the actual determination of the difference which gives the number of real roots can be effected only in the case of a rectangular contour, by applying to each of its sides separately a method such as that of Sturm's theorem; and thus the actual determination ultimately depends on a method such as that of Sturm's theorem.

Very little has been done in regard to the calculation of the imaginary roots of an equation by approximation; and the question is not here considered.

18. A class of numerical equations which needs to be considered is that of the binomial equations $x^n - a = 0$ ($a = \alpha + \beta i$, a complex number). The foregoing conclusions apply, viz. there are always n roots, which, it may be shown, are all unequal. And these can be found numerically by the extraction of the square root, and of an nth root, of *real* numbers, and by the aid of a table of natural sines and cosines*. For writing

$$\alpha + \beta i = \sqrt{\alpha^2 + \beta^2} \left\{ \frac{\alpha}{\sqrt{\alpha^2 + \beta^2}} + \frac{\beta}{\sqrt{\alpha^2 + \beta^2}} i \right\},$$

there is always a real angle λ (positive and less than 2π), such that its cosine and sine are $= \frac{\alpha}{\sqrt{\alpha^2 + \beta^2}}$ and $\frac{\beta}{\sqrt{\alpha^2 + \beta^2}}$ respectively; that is, writing for shortness $\sqrt{\alpha^2 + \beta^2} = \rho$, we have $\alpha + \beta i = \rho (\cos \lambda + i \sin \lambda)$, or the equation is $x^n = \rho (\cos \lambda + i \sin \lambda)$; hence observing that $\left(\cos \frac{\lambda}{n} + i \sin \frac{\lambda}{n}\right)^n = \cos \lambda + i \sin \lambda$, a value of x is $= \sqrt[n]{\rho} \left(\cos \frac{\lambda}{n} + i \sin \frac{\lambda}{n}\right)$. The formula really gives all the roots, for instead of λ we may write $\lambda + 2s\pi$, s a positive or negative integer, and then we have

$$x = \sqrt[n]{\rho} \left(\cos \frac{\lambda + 2s\pi}{n} + i \sin \frac{\lambda + 2s\pi}{n}\right),$$

which has the n values obtained by giving to s the values $0, 1, 2, \ldots, n-1$ in succession; the roots are, it is clear, represented by points lying at equal intervals on a circle. But it is more convenient to proceed somewhat differently; taking one of the roots to be θ, so that $\theta^n = a$, then assuming $x = \theta y$, the equation becomes $y^n - 1 = 0$, which equation, like the original equation, has precisely n roots (one of them being of course $= 1$). And the original equation $x^n - a = 0$ is thus reduced to the more simple equation $x^n - 1 = 0$; and although the theory of this equation is included in the preceding one, yet it is proper to state it separately.

The equation $x^n - 1 = 0$ has its several roots expressed in the form $1, \omega, \omega^2, \ldots, \omega^{n-1}$, where ω may be taken $= \cos \frac{2\pi}{n} + i \sin \frac{2\pi}{n}$; in fact, ω having this value, any integer power ω^k is $= \cos \frac{2\pi k}{n} + i \sin \frac{2\pi k}{n}$, and we thence have $(\omega^k)^n = \cos 2\pi k + i \sin 2\pi k$, $= 1$, that is, ω^k is a root of the equation. The theory will be resumed further on.

* The square root of $\alpha + \beta i$ can be determined by the extraction of square roots of positive real numbers, without the trigonometrical tables.

By what precedes, we are led to the notion (a numerical) of the radical $a^{\frac{1}{n}}$ regarded as an n-valued function; any one of these being denoted by $\sqrt[n]{a}$, then the series of values is $\sqrt[n]{a}$, $\omega\sqrt[n]{a}, \ldots, \omega^{n-1}\sqrt[n]{a}$; or we may, if we please, use $\sqrt[n]{a}$ instead of $a^{\frac{1}{n}}$ as a symbol to denote the n-valued function.

As the coefficients of an algebraical equation may be numerical, all which follows in regard to algebraical equations is (with, it may be, some few modifications) applicable to numerical equations; and hence, concluding for the present this subject, it will be convenient to pass on to algebraical equations.

II. We consider, secondly, algebraical equations (19 to 34).

19. The equation is

$$x^n - p_1 x^{n-1} + \ldots \pm p_n = 0,$$

and we here *assume* the existence of roots, viz. we assume that there are n quantities $a, b, c, \ldots$ (in general all of them different, but which in particular cases may become equal in sets in any manner), such that

$$x^n - p_1 x^{n-1} + \ldots \pm p_n = 0;$$

or looking at the question in a different point of view, and starting with the roots $a, b, c, \ldots$ as given, we express the product of the n factors $x-a, x-b, \ldots$ in the foregoing form, and thus arrive at an equation of the order n having the n roots $a, b, c, \ldots$. In either case we have

$$p_1 = \Sigma a, \quad p_2 = \Sigma ab, \ldots, \quad p_n = abc \ldots;$$

i.e., regarding the coefficients $p_1, p_2, \ldots, p_n$ as given, then we assume the existence of roots $a, b, c, \ldots$ such that $p_1 = \Sigma a$, &c.; or, regarding the roots as given, then we write p_1, p_2, &c., to denote the functions Σa, Σab, &c.

As already explained, the epithet algebraical is not used in opposition to numerical; an algebraical equation is merely an equation wherein the coefficients are not restricted to denote, or are not explicitly considered as denoting, numbers. That the abstraction is legitimate, appears by the simplest example; in saying that the equation $x^2 - px + q = 0$ has a root $x = \frac{1}{2}(p + \sqrt{p^2 - 4q})$, we mean that writing this value for x the equation becomes an identity, $\{\frac{1}{2}(p + \sqrt{p^2 - 4q})\}^2 - p\{\frac{1}{2}(p + \sqrt{p^2 - 4q})\} + q = 0$; and the verification of this identity in nowise depends upon p and q meaning numbers. But if it be asked what there is beyond numerical equations included in the term algebraical equation, or, again, what is the full extent of the meaning attributed to the term—the latter question at any rate it would be very difficult to answer; as to the former one, it may be said that the coefficients may, for instance, be symbols of operation. As regards such equations, there is certainly no proof that every equation has a root, or that an equation of the nth order has n roots; nor is it in any wise clear what the precise signification of the statement is. But it is found that the assumption of the existence of the n roots can be made without contradictory results; conclusions

derived from it, if they involve the roots, rest on the same ground as the original assumption; but the conclusion may be independent of the roots altogether, and in this case it is undoubtedly valid; the reasoning, although actually conducted by aid of the assumption (and, it may be, most easily and elegantly in this manner), is really independent of the assumption. In illustration, we observe that it is allowable to express a function of p and q as follows,—that is, by means of a rational symmetrical function of a and b; this can, as a fact, be expressed as a rational function of $a+b$ and ab; and if we prescribe that $a+b$ and ab shall then be changed into p and q respectively, we have the required function of p, q. That is, we have $F(\alpha, \beta)$ as a representation of $f(p, q)$, obtained as if we had $p=a+b$, $q=ab$, but without in any wise assuming the existence of the a, b of these equations.

20. Starting from the equation

$$x^n - p_1 x^{n-1} + \ldots = x - a \,.\, x - b \,.\, \&c.,$$

or the equivalent equations $p_1 = \Sigma a$, &c., we find

$$\begin{aligned} a^n - p_1 a^{n-1} + \ldots &= 0, \\ b^n - p_1 b^{n-1} + \ldots &= 0; \\ \vdots \qquad \vdots \qquad &\vdots \end{aligned}$$

(it is as satisfying these equations that a, b, ... are said to be the roots of $x^n - p_1 x^{n-1} + \ldots = 0$); and conversely from the last-mentioned equations, assuming that a, b, ... are all different, we deduce

$$p_1 = \Sigma a, \quad p_2 = \Sigma ab, \ \&c.,$$

and

$$x^n - p_1 x^{n-1} + \ldots = x - a \,.\, x - b \,.\, \&c.$$

Observe that if, for instance, $a=b$, then the equations $a^n - p_1 a^{n-1} + \ldots = 0$, $b^n - p_1 b^{n-1} + \ldots = 0$ would reduce themselves to a single relation, which would not of itself express that a was a double root,—that is, that $(x-a)^2$ was a factor of $x^n - p_1 x^{n-1} + \&c.$; but by considering b as the limit of $a+h$, h indefinitely small, we obtain a second equation

$$n a^{n-1} - (n-1) p_1 a^{n-2} + \ldots = 0,$$

which, with the first, expresses that a is a double root; and then the whole system of equations leads as before to the equations $p_1 = \Sigma a$, &c. But the existence of a double root implies a certain relation between the coefficients; the general case is when the roots are all unequal.

We have then the *theorem* that every rational symmetrical function of the roots is a rational function of the coefficients. This is an easy consequence from the less general theorem, every rational and integral symmetrical function of the roots is a rational and integral function of the coefficients.

In particular, the sums of the powers Σa^2, Σa^3, &c., are rational and integral functions of the coefficients.

The process originally employed for the expression of other functions $\Sigma a^\alpha b^\beta$, &c., in terms of the coefficients is to make them depend upon the sums of powers: for instance, $\Sigma a^\alpha b^\beta = \Sigma a^\alpha \Sigma a^\beta - \Sigma a^{\alpha+\beta}$; but this is very objectionable; the true theory consists in showing that we have systems of equations

$$\begin{aligned} p_1 &= \Sigma a, \\ \left\{\begin{aligned} p_2 &= \Sigma ab, \\ p_1^2 &= \Sigma a^2 + 2\Sigma ab, \end{aligned}\right. \\ \left\{\begin{aligned} p_3 &= \Sigma abc, \\ p_1 p_2 &= \Sigma a^2 b + 3\Sigma abc, \\ p_1^3 &= \Sigma a^3 + 3\Sigma a^2 b + 6\Sigma abc, \end{aligned}\right. \end{aligned}$$

where in each system there are precisely as many equations as there are root-functions on the right-hand side—e.g. 3 equations and 3 functions Σabc, $\Sigma a^2 b$, Σa^3. Hence in each system the root-functions can be determined linearly in terms of the powers and products of the coefficients:

$$\left\{\begin{aligned} \Sigma ab &= p_2, \\ \Sigma a^2 &= p_1^2 - 2p_2, \end{aligned}\right.$$

$$\left\{\begin{aligned} \Sigma abc &= p_3, \\ \Sigma a^2 b &= p_1 p_2 - 3p_3, \\ \Sigma a^3 &= p_1^3 - 3p_1 p_2 + 3p_3, \end{aligned}\right.$$

and so on. The older process, if applied consistently, would derive the originally assumed value Σab, $= p_2$, from the two equations $\Sigma a = p_1$, $\Sigma a^2 = p_1^2 - 2p_2$; i.e. we have $2\Sigma ab = \Sigma a \,.\, \Sigma a - \Sigma a^2$, $= p_1^2 - (p_1^2 - 2p_2)$, $= 2p_2$.

21. It is convenient to mention here the theorem that, x being determined as above by an equation of the order n, any rational and integral function whatever of x, or more generally any rational function which does not become infinite in virtue of the equation itself, can be expressed as a rational and integral function of x, of the order $n-1$, the coefficients being rational functions of the coefficients of the equation. Thus the equation gives x^n a function of the form in question; multiplying each side by x, and on the right-hand side writing for x^n its foregoing value, we have x^{n+1}, a function of the form in question; and the like for any higher power of x, and therefore also for any rational and integral function of x. The proof in the case of a rational non-integral function is somewhat more complicated. The final result is of the form $\frac{\phi(x)}{\psi(x)} = I(x)$, or say $\phi(x) - \psi(x) I(x) = 0$, where ϕ, ψ, I are rational and integral functions; in other words, this equation, being true if only $f(x) = 0$, can only be so by reason that the left-hand side contains $f(x)$ as a factor, or we must have identically $\phi(x) - \psi(x) I(x) = M(x) f(x)$. And it is, moreover, clear that the equation $\frac{\phi(x)}{\psi(x)} = I(x)$, being satisfied if only $f(x) = 0$, must be satisfied by each root of the equation.

From the theorem that a rational symmetrical function of the roots is expressible in terms of the coefficients, it at once follows that it is possible to determine an equation (of an assignable order) having for its roots the several values of any given (unsymmetrical) function of the roots of the given equation. For example, in the case of a quartic equation, having the roots (a, b, c, d), it is possible to find an equation having the roots ab, ac, ad, bc, bd, cd, being therefore a sextic equation: viz. in the product

$$(y-ab)(y-ac)(y-ad)(y-bc)(y-bd)(y-cd),$$

the coefficients of the several powers of y will be symmetrical functions of a, b, c, d and therefore rational and integral functions of the coefficients of the quartic equation; hence, supposing the product so expressed, and equating it to zero, we have the required sextic equation. In the same manner can be found the sextic equation having the roots $(a-b)^2$, $(a-c)^2$, $(a-d)^2$, $(b-c)^2$, $(b-d)^2$, $(c-d)^2$, which is the equation of differences previously referred to; and similarly we obtain the equation of differences for a given equation of any order. Again, the equation sought for may be that having for its n roots the given rational functions $\phi(a)$, $\phi(b)$, ... of the several roots of the given equation. Any such rational function can (as was shown) be expressed as a rational and integral function of the order $n-1$; and, retaining x in place of any one of the roots, the problem is to find y from the equations $x^n - p_1 x^{n-1} + \ldots = 0$, and $y = M_0 x^{n-1} + M_1 x^{n-2} + \ldots$, or, what is the same thing, from these two equations to eliminate x. This is, in fact, Tschirnhausen's transformation (1683).

22. In connexion with what precedes, the question arises as to the number of values (obtained by permutations of the roots) of given unsymmetrical functions of the roots, or say of a given set of letters: for instance, with roots or letters (a, b, c, d) as before, how many values are there of the function $ab+cd$, or better, how many functions are there of this form? The answer is 3, viz. $ab+cd$, $ac+bd$, $ad+bc$; or again we may ask whether, in the case of a given number of letters, there exist functions with a given number of values, 3-valued, 4-valued functions, &c.

It is at once seen that for any given number of letters there exist 2-valued functions; the product of the differences of the letters is such a function; however the letters are interchanged, it alters only its sign; or say the two values are $\Delta, -\Delta$. And if P, Q are symmetrical functions of the letters, then the general form of such a function is $P+Q\Delta$; this has only the two values $P+Q\Delta$, $P-Q\Delta$.

In the case of 4 letters there exist (as appears above) 3-valued functions: but in the case of 5 letters there does not exist any 3-valued or 4-valued function; and the only 5-valued functions are those which are symmetrical in regard to four of the letters, and can thus be expressed in terms of one letter and of symmetrical functions of all the letters. These last theorems present themselves in the demonstration of the non-existence of a solution of a quintic equation by radicals.

The theory is an extensive and important one, depending on the notions of *substitutions* and of *groups* *.

* A substitution is the operation by which we pass from the primitive arrangement of n letters to any other arrangement of the same letters: for instance, the substitution $\left(\begin{smallmatrix} bcda \\ abcd \end{smallmatrix}\right)$ means that a is to be changed

23. Returning to equations, we have the very important theorem that, given the value of any unsymmetrical function of the roots, e.g. in the case of a quartic equation, the function $ab+cd$, it is in general possible to determine rationally the value of any similar function, such as $(a+b)^3+(c+d)^3$.

The *a priori* ground of this theorem may be illustrated by means of a numerical equation. Suppose that the roots of a quartic equation are 1, 2, 3, 4, then if it is given that $ab+cd=14$, this in effect determines a, b to be 1, 2 and c, d to be 3, 4 (viz. $a=1$, $b=2$ or $a=2$, $b=1$, and $c=3$, $d=4$ or $c=4$, $d=3$) or else a, b to be 3, 4 and c, d to be 1, 2; and it therefore in effect determines $(a+b)^3+(c+d)^3$ to be $=370$, and not any other value; that is, $(a+b)^3+(c+d)^3$, as having a single value, must be determinable rationally. And we can in the same way account for cases of failure as regards particular equations; thus, the roots being 1, 2, 3, 4 as before, $a^2b=2$ determines a to be $=1$ and b to be $=2$; but if the roots had been 1, 2, 4, 16 then $a^2b=16$ does not uniquely determine a, b but only makes them to be 1, 16 or 2, 4 respectively.

As to the *a posteriori* proof, assume, for instance,

$$t_1=ab+cd, \quad y_1=(a+b)^3+(c+d)^3,$$
$$t_2=ac+bd, \quad y_2=(a+c)^3+(b+d)^3,$$
$$t_3=ad+bc, \quad y_3=(a+d)^3+(b+c)^3:$$

then

$$y_1+y_2+y_3, \quad t_1y_1+t_2y_2+t_3y_3, \quad t_1^2y_1+t_2^2y_2+t_3^2y_3,$$

will be respectively symmetrical functions of the roots of the quartic, and therefore rational and integral functions of the coefficients; that is, they will be known.

Suppose for a moment that t_1, t_2, t_3 are *all* known; then the equations being linear in y_1, y_2, y_3 these can be expressed rationally in terms of the coefficients and of t_1, t_2, t_3; that is, y_1, y_2, y_3 will be known. But observe further that y_1 is obtained as a function of t_1, t_2, t_3 symmetrical as regards t_2, t_3; it can therefore be expressed

into b, b into c, c into d, d into a. Substitutions may, of course, be represented by single letters α, β, ...; $\frac{abcd}{abcd}$, $=1$, is the substitution which leaves the letters unaltered. Two or more substitutions may be compounded together and give rise to a substitution; i.e., performing upon the primitive arrangement first the substitution β and then upon the result the substitution α, we have the substitution $\alpha\beta$. Substitutions are not commutative; thus, $\alpha\beta$ is not in general $=\beta\alpha$; but they are associative, $\alpha\beta.\gamma=\alpha.\beta\gamma$, so that $\alpha\beta\gamma$ has a determinate meaning. A substitution may be compounded any number of times with itself, and we thus have the powers α^2, α^3, ..., &c. Since the number of substitutions is limited, some power α^ν must be $=1$: or, as this may be expressed, every substitution is a root of unity. A group of substitutions is a set such that each two of them compounded together in either order gives a substitution belonging to the set; every group includes the substitution unity, so that we may in general speak of a group 1, α, β, ... (the number of terms is the order of the group). The whole system of the $1.2.3...n$ substitutions which can be performed upon the n letters is obviously a group: the order of every other group which can be formed out of these substitutions is a submultiple of this number; but it is not conversely true that a group exists the order of which is any given submultiple of this number. In the case of a determinant the substitutions which give rise to the positive terms form a group the order of which is $=\frac{1}{2}.1.2.3...n$. For any function of the n letters, the whole series of substitutions which leave the value of the functions unaltered form a group; and thence also the number of values of the function is $=1.2.3...n$ divided by the order of the group.

as a rational function of t_1 and of t_2+t_3, t_2t_3, and thence as a rational function of t_1 and of $t_1+t_2+t_3$, $t_1t_2+t_1t_3+t_2t_3$, $t_1t_2t_3$; but these last are symmetrical functions of the roots, and as such they are expressible rationally in terms of the coefficients; that is, y_1 will be expressed as a rational function of t_1 and of the coefficients; or t_1 (alone, not t_2 or t_3) being known, y_1 will be rationally determined.

24. We now consider the question of the algebraical solution of equations, or, more accurately, that of the *solution of equations by radicals.*

In the case of a quadric equation $x^2-px+q=0$, we can by the assistance of the sign $\sqrt{(\ )}$ or $(\ )^{\frac{1}{2}}$ find an expression for x as a two-valued function of the coefficients p, q such that, substituting this value in the equation, the equation is thereby identically satisfied; it has been found that this expression is

$$x=\tfrac{1}{2}\{p\pm\sqrt{p^2-4q}\},$$

and the equation is on this account said to be algebraically solvable, or more accurately solvable by radicals. Or we may by writing $x=-\frac{1}{2}p+z$, reduce the equation to $z^2=\frac{1}{4}(p^2-4q)$ viz. to an equation of the form $z^2=a$; and in virtue of its being thus reducible we say that the original equation is solvable by radicals. And the question for an equation of any higher order, say of the order n, is, can we by means of radicals, that is, by aid of the sign $\sqrt[m]{(\ )}$ or $(\ )^{\frac{1}{m}}$, using as many as we please of such signs and with any values of m, find an n-valued function (or any function) of the coefficients which substituted for x in the equation shall satisfy it identically.

It will be observed that the coefficients $p, q, \ldots$ are not explicitly considered as numbers, but even if they do denote numbers, the question whether a numerical equation admits of solution by radicals is wholly unconnected with the before-mentioned theorem of the existence of the n roots of such an equation. It does not even follow that in the case of a numerical equation solvable by radicals the algebraical solution gives the numerical solution, but this requires explanation. Consider first a numerical quadric equation with imaginary coefficients. In the formula $x=\frac{1}{2}(p\pm\sqrt{p^2-4q})$, substituting for p, q their given numerical values, we obtain for x an expression of the form $x=\alpha+\beta i\pm\sqrt{\gamma+\delta i}$, where $\alpha, \beta, \gamma, \delta$ are real numbers. This expression substituted for x in the quadric equation would satisfy it identically, and it is thus an algebraical solution; but there is no obvious *a priori* reason why $\sqrt{\gamma+\delta i}$ should have a value $=c+di$, where c and d are real numbers calculable by the extraction of a root or roots of real numbers; however the case is (what there was no *a priori* right to expect) that $\sqrt{\gamma+\delta i}$ has such a value calculable by means of the radical expressions $\sqrt{\{\sqrt{\gamma^2+\delta^2}\pm\gamma\}}$: and hence the algebraical solution of a numerical quadric equation does in every case give the numerical solution. The case of a numerical cubic equation will be considered presently.

25. A cubic equation can be solved by radicals. Taking for greater simplicity the cubic in the reduced form $x^3+qx-r=0$, and assuming $x=a+b$, this will be a solution if only $3ab=q$ and $a^3+b^3=r$, equations which give $(a^3-b^3)^2=r^2-\frac{4}{27}q^3$, a

quadric equation solvable by radicals, and giving $a^3 - b^3 = \sqrt{r^2 - \frac{4}{27}q^3}$, a two-valued function of the coefficients: combining this with $a^3 + b^3 = r$, we have $a^3 = \frac{1}{2}(r + \sqrt{r^2 - \frac{4}{27}q^3})$, a two-valued function: we then have a by means of a cube root, viz.

$$a = \sqrt[3]{\{\tfrac{1}{2}(r + \sqrt{r^2 - \tfrac{4}{27}q^3})\}},$$

a six-valued function of the coefficients; but then, writing $q = \dfrac{b}{3a}$, we have, as may be shown, $a + b$ a three-valued function of the coefficients; and $x = a + b$ is the required solution by radicals. It would have been wrong to complete the solution by writing

$$b = \sqrt[3]{\{\tfrac{1}{2}(r - \sqrt{r^2 - \tfrac{4}{27}q^3})\}},$$

for then $a + b$ would have been given as a 9-valued function having only 3 of its values roots, and the other 6 values being irrelevant. Observe that in this last process we make no use of the equation $3ab = q$, in its original form, but use only the derived equation $27a^3b^3 = q^3$, implied in, but not implying, the original form.

An interesting variation of the solution is to write $x = ab(a + b)$, giving $a^3b^3(a^3 + b^3) = r$ and $3a^3b^3 = q$, or say $a^3 + b^3 = \dfrac{3r}{q}$, $a^3b^3 = \frac{1}{3}q$; and consequently

$$a^3 = \frac{\frac{3}{2}}{q}(r + \sqrt{r^2 - \tfrac{4}{27}q^3}),\quad b^3 = \frac{\frac{3}{2}}{q}(r - \sqrt{r^2 - \tfrac{4}{27}q^3}),$$

i.e., here a^3, b^3 are each of them a two-valued function, but as the only effect of altering the sign of the quadric radical is to interchange a^3, b^3, they may be regarded as each of them one-valued; a and b are each of them 3-valued (for observe that here only a^3b^3, not ab, is given); and $ab(a + b)$ thus is in appearance a 9-valued function, but it can easily be shown that it is (as it ought to be) only 3-valued.

In the case of a numerical cubic, even when the coefficients are real, substituting their values in the expression

$$x = \sqrt[3]{\{\tfrac{1}{2}(r + \sqrt{r^2 - \tfrac{4}{27}q^3})\}} + \tfrac{1}{3}q \div \sqrt[3]{\{\tfrac{1}{2}(r + \sqrt{r^2 - \tfrac{4}{27}q^3})\}},$$

this may depend on an expression of the form $\sqrt[3]{\gamma + \delta i}$, where γ and δ are real numbers (it will do so if $r^2 - \frac{4}{27}q^3$ is a negative number), and then we *cannot* by the extraction of any root or roots of real positive numbers reduce $\sqrt[3]{\gamma + \delta i}$ to the form $c + di$, c and d real numbers; hence here the algebraical solution does *not* give the numerical solution, and we have here the so-called "irreducible case" of a cubic equation. By what precedes, there is nothing in this that might not have been expected; the algebraical solution makes the solution depend on the extraction of the cube root of a negative number, and there was no reason for expecting this to be a real number. It is well known that the case in question is that wherein the three roots of the numerical cubic equation are all real; if the roots are two imaginary, one real, then contrariwise the quantity under the cube root is real; and the algebraical solution gives the numerical one.

The irreducible case is solvable by a trigonometrical formula, but this is not a solution by radicals: it consists, in effect, in reducing the given numerical cubic (not to a cubic of the form $z^3 = a$, solvable by the extraction of a cube root, but) to a cubic of the form $4x^3 - 3x = a$, corresponding to the equation $4\cos^3\theta - 3\cos\theta = \cos 3\theta$ which serves to determine $\cos\theta$ when $\cos 3\theta$ is known. The theory is applicable to an algebraical cubic equation; say that such an equation, if it can be reduced to the form $4x^3 - 3x = a$, is solvable by "trisection"—then the general cubic equation is solvable by trisection.

26. A quartic equation is solvable by radicals: and it is to be remarked that the existence of such a solution depends on the existence of 3-valued functions such as $ab + cd$ of the four roots (a, b, c, d): by what precedes, $ab + cd$ is the root of a cubic equation, which equation is solvable by radicals: hence $ab + cd$ can be found by radicals; and since $abcd$ is a given function, ab and cd can then be found by radicals. But by what precedes, if ab be known then any similar function, say $a + b$, is obtainable rationally; and then from the values of $a + b$ and ab we may by radicals obtain the value of a or b, that is, an expression for the root of the given quartic equation: the expression ultimately obtained is 4-valued, corresponding to the different values of the several radicals which enter therein, and we have thus the expression by radicals of each of the four roots of the quartic equation. But when the quartic is numerical the same thing happens as in the cubic, and the algebraical solution does not in every case give the numerical one.

It will be understood, from the foregoing explanation as to the quartic, how in the next following case, that of the quintic, the question of the solvability by radicals depends on the existence or non-existence of k-valued functions of the five roots (a, b, c, d, e); the fundamental theorem is the one already stated, a rational function of five letters, if it has less than 5, cannot have more than 2 values, that is, there are no 3-valued or 4-valued functions of 5 letters: and by reasoning depending in part upon this theorem, Abel (1824) showed that a general quintic equation is not solvable by radicals; and *a fortiori* the general equation of any order higher than 5 is not solvable by radicals.

27. The general theory of the solvability of an equation by radicals depends fundamentally on Vandermonde's remark (1770) that, supposing an equation is solvable by radicals, and that we have therefore an algebraical expression of x in terms of the coefficients, then substituting for the coefficients their values in terms of the roots, the resulting expression must reduce itself to any one at pleasure of the roots $a, b, c, \ldots$; thus in the case of the quadric equation, in the expression $x = \frac{1}{2}(p + \sqrt{p^2 - 4q})$, substituting for p and q their values, and observing that $(a+b)^2 - 4ab = (a-b)^2$, this becomes $x = \frac{1}{2}\{a + b + \sqrt{(a-b)^2}\}$, the value being a or b according as the radical is taken to be $+(a-b)$ or $-(a-b)$.

So in the cubic equation $x^3 - px^2 + qx - r = 0$, if the roots are a, b, c, and if ω is used to denote an imaginary cube root of unity, $\omega^2 + \omega + 1 = 0$, then writing for shortness $p = a + b + c$, $L = a + \omega b + \omega^2 c$, $M = a + \omega^2 b + \omega c$, it is at once seen that LM,

$L^3 + M^3$, and therefore also $(L^3 - M^3)^2$ are symmetrical functions of the roots, and consequently rational functions of the coefficients: hence

$$\tfrac{1}{2}\{L^3 + M^3 + \sqrt{(L^3 - M^3)^2}\}$$

is a rational function of the coefficients, which when these are replaced by their values as functions of the roots becomes, according to the sign given to the quadric radical, $= L^3$ or M^3: taking it $= L^3$, the cube root of the expression has the three values $L, \omega L, \omega^2 L$; and LM divided by the same cube root has therefore the values $M, \omega^2 M, \omega M$; whence finally the expression

$$\tfrac{1}{3}[p \quad + \sqrt[3]{\{\tfrac{1}{2}(L^3 + M^3 + \sqrt{(L^3 - M^3)^2})\}}$$
$$+ LM \div \sqrt[3]{\{\tfrac{1}{2}(L^3 + M^3 + \sqrt{(L^3 - M^3)^2})\}}]$$

has the three values

$$\tfrac{1}{3}(p + L + M),\ \tfrac{1}{3}(p + \omega L + \omega^2 M),\ \tfrac{1}{3}(p + \omega^2 L + \omega M);$$

that is, these are $= a,\ b,\ c$ respectively. If the value M^3 had been taken instead of L^3, then the expression would have had the same three values $a,\ b,\ c$. Comparing the solution given for the cubic $x^3 + qx - r = 0$, it will readily be seen that the two solutions are identical, and that the function $r^2 - \frac{4}{27}q^3$ under the radical sign must (by aid of the relation $p = 0$ which subsists in this case) reduce itself to $(L^3 - M^3)^2$; it is only by each radical being equal to a rational function of the roots that the final expression *can* become equal to the roots $a,\ b,\ c$ respectively.

28. The formulæ for the cubic were obtained by Lagrange (1770—71) from a different point of view. Upon examining and comparing the principal known methods for the solution of algebraical equations, he found that they all ultimately depended upon finding a "resolvent" equation of which the root is $a + \omega b + \omega^2 c + \omega^3 d + \ldots$, ω being an imaginary root of unity, of the same order as the equation; e.g., for the cubic the root is $a + \omega b + \omega^2 c$, ω an imaginary cube root of unity. Evidently the method gives for L^3 a quadric equation, which is the "resolvent" equation in this particular case.

For a quartic the formulæ present themselves in a somewhat different form, by reason that 4 is not a prime number. Attempting to apply it to a quintic, we seek for the equation of which the root is $(a + \omega b + \omega^2 c + \omega^3 d + \omega^4 e)$, ω an imaginary fifth root of unity, or rather the fifth power thereof $(a + \omega b + \omega^2 c + \omega^3 d + \omega^4 e)^5$; this is a 24-valued function, but if we consider the four values corresponding to the roots of unity $\omega, \omega^2, \omega^3, \omega^4$, viz. the values

$$(a + \omega\, b + \omega^2 c + \omega^3 d + \omega^4 e)^5,$$
$$(a + \omega^2 b + \omega^4 c + \omega\, d + \omega^3 e)^5,$$
$$(a + \omega^3 b + \omega\, c + \omega^4 d + \omega^2 e)^5,$$
$$(a + \omega^4 b + \omega^3 c + \omega^2 d + \omega\, e)^5,$$

any symmetrical function of these, for instance their sum, is a six-valued function of the roots, and may therefore be determined by means of a sextic equation, the

coefficients whereof are rational functions of the coefficients of the original quintic equation; the conclusion being that the solution of an equation of the fifth order is made to depend upon that of an equation of the sixth order. This is, of course, useless for the solution of the quintic equation, which, as already mentioned, does not admit of solution by radicals; but the equation of the sixth order, Lagrange's resolvent sextic, is very important, and is intimately connected with all the later investigations in the theory.

29. It is to be remarked, in regard to the question of solvability by radicals, that not only the coefficients are taken to be arbitrary, but it is assumed that they are represented each by a single letter, or say rather that they are not so expressed in terms of other arbitrary quantities as to make a solution possible. If the coefficients are not all arbitrary, for instance, if some of them are zero, a sextic equation might be of the form $x^6 + bx^4 + cx^2 + d = 0$, and so be solvable as a cubic; or if the coefficients of the sextic are given functions of the six arbitrary quantities a, b, c, d, e, f, such that the sextic is really of the form

$$(x^2 + ax + b)(x^4 + cx^3 + dx^2 + ex + f) = 0,$$

then it breaks up into the equations $x^2 + ax + b = 0$, $x^4 + cx^3 + dx^2 + ex + f = 0$, and is consequently solvable by radicals; so also if the form is

$$(x - a)(x - b)(x - c)(x - d)(x - e)(x - f) = 0,$$

then the equation is solvable by radicals,—in this extreme case rationally. Such cases of solvability are self-evident; but they are enough to show that the general theorem of the non-solvability by radicals of an equation of the fifth or any higher order does not in any wise exclude for such orders the existence of particular equations solvable by radicals, and there are, in fact, extensive classes of equations which are thus solvable; the binomial equations $x^n - 1 = 0$ present an instance.

30. It has already been shown how the several roots of the equation $x^n - 1 = 0$ can be expressed in the form $\cos\frac{2s\pi}{n} + i\sin\frac{2s\pi}{n}$, but the question is now that of the algebraical solution (or solution by radicals) of this equation. There is always a root $= 1$; if ω be any other root, then obviously $\omega, \omega^2, \ldots, \omega^{n-1}$ are all of them roots; $x^n - 1$ contains the factor $x - 1$, and it thus appears that $\omega, \omega^2, \ldots, \omega^{n-1}$ are the $n - 1$ roots of the equation

$$x^{n-1} + x^{n-2} + \ldots + x + 1 = 0;$$

we have, of course,

$$\omega^{n-1} + \omega^{n-2} + \ldots + \omega + 1 = 0.$$

It is proper to distinguish the cases n prime and n composite; and in the latter case there is a distinction according as the prime factors of n are simple or multiple. By way of illustration, suppose successively $n = 15$ and $n = 9$; in the former case, if α be an imaginary root of $x^3 - 1 = 0$ (or root of $x^2 + x + 1 = 0$), and β an imaginary root of $x^5 - 1 = 0$ (or root of $x^4 + x^3 + x^2 + x + 1 = 0$), then ω may be taken $= \alpha\beta$; the successive powers thereof, $\alpha\beta$, $\alpha^2\beta^2$, β^3, $\alpha\beta^4$, α^2, β, $\alpha\beta^2$, $\alpha^2\beta^3$, β^4, α,

$\alpha^2\beta$, β^2, $\alpha\beta^3$, $\alpha^2\beta^4$, are the roots of $x^{14}+x^{13}+\ldots+x+1=0$; the solution thus depends on the solution of the equations $x^3-1=0$ and $x^5-1=0$. In the latter case, if α be an imaginary root of $x^3-1=0$ (or root of $x^2+x+1=0$), then the equation $x^9-1=0$ gives $x^3=1$, α, or α^2; $x^3=1$ gives $x=1$, α, or α^2; and the solution thus depends on the solution of the equations $x^3-1=0$, $x^3-\alpha=0$, $x^3-\alpha^2=0$. The first equation has the roots 1, α, α^2; if β be a root of either of the others, say if $\beta^3=\alpha$, then assuming $\omega=\beta$, the successive powers are β, β^2, α, $\alpha\beta$, $\alpha\beta^2$, α^2, $\alpha^2\beta$, $\alpha^2\beta^2$, which are the roots of the equation $x^8+x^7+\ldots+x+1=0$.

It thus appears that the only case which need be considered is that of n a prime number, and writing (as is more usual) r in place of ω, we have r, r^2, r^3, ..., r^{n-1} as the $(n-1)$ roots of the reduced equation

$$x^{n-1}+x^{n-2}+\ldots+x+1=0;$$

then not only $r^n-1=0$, but also $r^{n-1}+r^{n-2}+\ldots+r+1=0$.

31. The process of solution due to Gauss (1801) depends essentially on the arrangement of the roots in a certain order, viz. not as above, with the indices of r in arithmetical progression, but with their indices in geometrical progression; the prime number n has a certain number of prime roots g, which are such that g^{n-1} is the lowest power of g, which is $\equiv 1$ to the modulus n; or, what is the same thing, that the series of powers 1, g, g^2, ..., g^{n-2}, each divided by n, leave (in a different order) the remainders 1, 2, 3, ..., $n-1$; hence giving to r in succession the indices 1, g, g^2, ..., g^{n-2}, we have, in a different order, the whole series of roots r, r^2, r^3, ..., r^{n-1}.

In the most simple case, $n=5$, the equation to be solved is $x^4+x^3+x^2+x+1=0$; here 2 is a prime root of 5, and the order of the roots is r, r^2, r^4, r^3. The Gaussian process consists in forming an equation for determining the periods P_1, P_2, $=r+r^4$ and r^2+r^3 respectively,—these being such that the symmetrical functions P_1+P_2, P_1P_2 are rationally determinable: in fact,

$$P_1+P_2=-1, \quad P_1P_2=(r+r^4)(r^2+r^3), \ =r^3+r^4+r^6+r^7, \ =r^3+r^4+r+r^2, \ =-1.$$

P_1, P_2 are thus the roots of $u^2+u-1=0$; and taking them to be known, they are themselves broken up into subperiods, in the present case single terms, r and r^4 for P_1, r^2 and r^3 for P_2; the symmetrical functions of these are then rationally determined in terms of P_1 and P_2; thus $r+r^4=P_1$, $r.r^4=1$, or r, r^4 are the roots of $u^2-P_1u+1=0$. The mode of division is more clearly seen for a larger value of n; thus, for $n=7$ a prime root is $=3$, and the arrangement of the roots is r, r^3, r^2, r^6, r^4, r^5. We may form either 3 periods each of 2 terms,

$$P_1, P_2, P_3, \ =r+r^6, \ r^3+r^4, \ r^2+r^5,$$

respectively; or else 2 periods each of 3 terms, P_1, $P_2=r+r^2+r^4$, $r^3+r^6+r^5$ respectively; in each case the symmetrical functions of the periods are rationally determinable; thus in the case of the two periods $P_1+P_2=-1$, $P_1P_2=3+r+r^2+r^3+r^4+r^5+r^6$, $=2$;

and, the periods being known, the symmetrical functions of the several terms of each period are rationally determined in terms of the periods, thus

$$r + r^2 + r^4 = P_1, \quad r \cdot r^2 + r \cdot r^4 + r^2 \cdot r^4 = P_2, \quad r \cdot r^2 \cdot r^4 = 1.$$

The theory was further developed by Lagrange (1808), who, applying his general process to the equation in question, $x^{n-1} + x^{n-2} + \ldots + x + 1 = 0$, the roots a, b, c, ... being the several powers of r, the indices in geometrical progression as above, showed that the function $(a + \omega b + \omega^2 c + \ldots)^{n-1}$ was in this case a given function of ω with integer coefficients. Reverting to the before-mentioned particular equation $x^4 + x^3 + x^2 + x + 1 = 0$, it is very interesting to compare the process of solution with that for the solution of the general quartic the roots whereof are a, b, c, d.

Take ω, a root of the equation $\omega^4 - 1 = 0$ (whence ω is $= 1, -1, i$, or $-i$, at pleasure), and consider the expression

$$(a + \omega b + \omega^2 c + \omega^3 d)^4.$$

The developed value of this is

$$\begin{aligned} = \quad & a^4 + b^4 + c^4 + d^4 + 6(a^2c^2 + b^2d^2) + 12(a^2bd + b^2ca + c^2db + d^2ac) \\ + \omega \ & \{4(a^3b + b^3c + c^3d + d^3a) + 12(a^2cd + b^2da + c^2ab + d^2bc)\} \\ + \omega^2 \ & \{6(a^2b^2 + b^2c^2 + c^2d^2 + d^2a^2) + 4(a^3c + b^3d + c^3a + d^3b) + 24abcd\} \\ + \omega^3 \ & \{4(a^3d + b^3a + c^3b + d^3c) + 12(a^2bc + b^2cd + c^2da + d^2ab)\}; \end{aligned}$$

that is, this is a 6-valued function of a, b, c, d, the root of a sextic (which is, in fact, solvable by radicals; but this is not here material).

If, however, a, b, c, d denote the roots r, r^2, r^4, r^3 of the special equation, then the expression becomes

$$\begin{aligned} & r^4 + r^3 + r + r^2 + 6(1 + 1) + 12(r^2 + r^4 + r^3 + r) \\ + \omega \ & \{4(1 + 1 + 1 + 1) + 12(r^4 + r^3 + r + r^2)\} \\ + \omega^2 \ & \{6(r + r^2 + r^4 + r^3) + 4(r^2 + r^4 + r^3 + r)\} \\ + \omega^3 \ & \{4(r + r^2 + r^4 + r^3) + 12(r^3 + r + r^2 + r^4)\}; \end{aligned}$$

viz. this is

$$= -1 + 4\omega + 14\omega^2 - 16\omega^3,$$

a completely determined value. That is, we have

$$(r + \omega r^2 + \omega^2 r^4 + \omega^3 r^3)^4 = -1 + 4\omega + 14\omega^2 - 16\omega^3,$$

which result contains the solution of the equation. If $\omega = 1$, we have $(r + r^2 + r^4 + r^3)^4 = 1$, which is right; if $\omega = -1$, then $(r + r^4 - r^2 - r^3)^4 = 25$; if $\omega = i$, then we have $\{r - r^4 + i(r^2 - r^3)\}^4 = -15 + 20i$; and if $\omega = -i$, then $\{r - r^4 - i(r^2 - r^3)\}^4 = -15 - 20i$; the solution may be completed without difficulty.

The result is perfectly general, thus:—n being a prime number, r a root of the equation $x^{n-1} + x^{n-2} + \ldots + x + 1 = 0$, ω a root of $\omega^{n-1} - 1 = 0$, and g a prime root of $g^{n-1} \equiv 1 \pmod{n}$, then

$$(r + \omega r^g + \ldots + \omega^{n-2} r^{g^{n-2}})^{n-1}$$

is a given function $M_0+M_1\omega+\ldots+M_{n-2}\omega^{n-2}$ with integer coefficients, and by the extraction of $(n-1)$th roots of this and similar expressions we ultimately obtain r in terms of ω, which is taken to be known; the equation $x^n-1=0$, n a prime number, is thus solvable by radicals. In particular, if $n-1$ be a power of 2, the solution (by either process) requires the extraction of square roots only; and it was thus that Gauss discovered that it was possible to construct geometrically the regular polygons of 17 sides and 257 sides respectively. Some interesting developments in regard to the theory were obtained by Jacobi (1837); see the memoir "Ueber die Kreistheilung, u.s.w.," *Crelle*, t. XXX. (1846).

The equation $x^{n-1}+\ldots+x+1=0$ has been considered for its own sake, but it also serves as a specimen of a class of equations solvable by radicals, considered by Abel (1828), and since called Abelian equations, viz., for the Abelian equation of the order n, if x be any root, the roots are x, θx, $\theta^2 x, \ldots$, $\theta^{n-1}x$ (θx being a rational function of x, and $\theta^n x=x$); the theory is, in fact, very analogous to that of the above particular case. A more general theorem obtained by Abel is as follows:—If the roots of an equation of any order are connected together in such wise that *all* the roots can be expressed rationally in terms of any one of them, say x; if, moreover, θx, $\theta_1 x$ being any two of the roots, we have $\theta\theta_1 x=\theta_1\theta x$, the equation will be solvable algebraically. It is proper to refer also to Abel's definition of an *irreducible* equation:—an equation $\phi x=0$, the coefficients of which are rational functions of a certain number of known quantities $a, b, c, \ldots$, is called irreducible when it is impossible to express its roots by an equation of an inferior degree, the coefficients of which are also rational functions of $a, b, c, \ldots$ (or, what is the same thing, when ϕx does not break up into factors which are rational functions of $a, b, c, \ldots$). Abel applied his theory to the equations which present themselves in the division of the elliptic functions, but not to the modular equations.

32. But the theory of the algebraical solution of equations in its most complete form was established by Galois (born October 1811, killed in a duel May 1832; see his collected works, *Liouville*, t. XI., 1846). The definition of an irreducible equation resembles Abel's,—an equation is reducible when it admits of a rational divisor, irreducible in the contrary case; only the word *rational* is used in this extended sense that, in connexion with the coefficients of the given equation, or with the irrational quantities (if any) whereof these are composed, he considers any number of other irrational quantities called "adjoint radicals," and he terms rational any rational function of the coefficients (or the irrationals whereof they are composed) and of these adjoint radicals; the epithet irreducible is thus taken either absolutely or in a relative sense, according to the system of adjoint radicals which are taken into account. For instance, the equation $x^4+x^3+x^2+x+1=0$; the left-hand side has here no rational divisor, and the equation is irreducible; but this function is $=(x^2+\frac{1}{2}x+1)^2-\frac{5}{4}x^2$, and it has thus the irrational divisors $x^2+\frac{1}{2}(1+\sqrt{5})x+1$, $x^2+\frac{1}{2}(1-\sqrt{5})x+1$; and these, if we *adjoin* the radical $\sqrt{5}$, are rational, and the equation is no longer irreducible. In the case of a given equation, assumed to be irreducible, the problem to solve the equation is, in fact, that of finding radicals by the adjunction of which the equation

becomes reducible; for instance, the general quadric equation $x^2+px+q=0$ is irreducible, but it becomes reducible, breaking up into rational linear factors, when we adjoin the radical $\sqrt{\frac{1}{4}p^2-q}$.

The fundamental theorem is the Proposition I. of the "Mémoire sur les conditions de résolubilité des équations par radicaux"; viz. given an equation of which $a, b, c, \ldots$ are the m roots, there is always a group of permutations of the letters $a, b, c, \ldots$ possessed of the following properties:—

1. Every function of the roots invariable by the substitutions of the group is rationally known.

2. Reciprocally, every rationally determinable function of the roots is invariable by the substitutions of the group.

Here by an invariable function is meant not only a function of which the form is invariable by the substitutions of the group, but further, one of which the value is invariable by these substitutions: for instance, if the equation be $\phi x=0$, then ϕx is a function of the roots invariable by any substitution whatever. And in saying that a function is rationally known, it is meant that its value is expressible rationally in terms of the coefficients and of the adjoint quantities.

For instance, in the case of a general equation, the group is simply the system of the $1.2.3 \ldots n$ permutations of all the roots, since, in this case, the only rationally determinable functions are the symmetric functions of the roots.

In the case of the equation $x^{n-1}+\ldots+x+1=0$, n a prime number,

$$a, b, c, \ldots, k=r, r^g, r^{g^2}, \ldots, r^{g^{n-2}},$$

where g is a prime root of n, then the group is the cyclical group $abc \ldots k$, $bc \ldots ka, \ldots, kab \ldots j$, that is, in this particular case the number of the permutations of the group is equal to the order of the equation.

This notion of the group of the original equation, or of the group of the equation as varied by the adjunction of a series of radicals, seems to be the fundamental one in Galois's theory. But the problem of solution by radicals, instead of being the sole object of the theory, appears as the first link of a long chain of questions relating to the transformation and classification of irrationals.

Returning to the question of solution by radicals, it will be readily understood that by the adjunction of a radical the group may be diminished; for instance, in the case of the general cubic, where the group is that of the six permutations, by the adjunction of the square root which enters into the solution, the group is reduced to abc, bca, cab; that is, it becomes possible to express rationally, in terms of the coefficients and of the adjoint square root, any function such as $a^2b+b^2c+c^2a$ which is not altered by the cyclical substitution a into b, b into c, c into a. And hence, to determine whether an equation of a given form is solvable by radicals, the course of investigation is to inquire whether, by the successive adjunction of radicals, it is

possible to reduce the original group of the equation so as to make it ultimately consist of a single permutation.

The condition in order that an equation of a given prime order n may be solvable by radicals was in this way obtained—in the first instance in the form, scarcely intelligible without further explanation, that every function of the roots $x_1, x_2, \ldots, x_n$, invariable by the substitutions x_{ak+b} for x_k, must be rationally known; and then in the equivalent form that the resolvent equation of the order $1.2 \ldots \overline{n-2}$ must have a rational root. In particular, the condition in order that a quintic equation may be solvable is that Lagrange's resolvent of the order 6 may have a rational factor, a result obtained from a direct investigation in a valuable memoir by E. Luther, *Crelle*, t. XXXIV. (1847).

Among other results demonstrated or announced by Galois may be mentioned those relating to the modular equations in the theory of elliptic functions; for the transformations of the orders 5, 7, 11, the modular equations of the orders 6, 8, 12 are depressible to the orders 5, 7, 11 respectively; but for the transformation, n a prime number greater than 11, the depression is impossible.

The general theory of Galois in regard to the solution of equations was completed, and some of the demonstrations supplied, by Betti (1852). See also Serret's *Cours d'Algèbre supérieure*, 2nd ed. 1854; 4th ed. 1877—78.

33. Returning to quintic equations, Jerrard (1835) established the theorem that the general quintic equation is, by the extraction of only square and cubic roots, reducible to the form $x^5 + ax + b = 0$, or what is the same thing, to $x^5 + x + b = 0$. The actual reduction by means of Tschirnhausen's theorem was effected by Hermite in connexion with his elliptic-function solution of the quintic equation (1858) in a very elegant manner. It was shown by Cockle and Harley (1858—59) in connexion with the Jerrardian form, and by Cayley (1861), that Lagrange's resolvent equation of the sixth order can be replaced by a more simple sextic equation occupying a like place in the theory.

The theory of the modular equations, more particularly for the case $n = 5$, has been studied by Hermite, Kronecker, and Brioschi. In the case $n = 5$, the modular equation of the order 6 depends, as already mentioned, on an equation of the order 5; and conversely the general quintic equation may be made to depend upon this modular equation of the order 6; that is, assuming the solution of this modular equation, we can solve (not by radicals) the general quintic equation; this is Hermite's solution of the general quintic equation by elliptic functions (1858); it is analogous to the before-mentioned trigonometrical solution of the cubic equation. The theory is reproduced and developed in Brioschi's memoir, "Ueber die Auflösung der Gleichungen vom fünften Grade," *Math. Annalen*, t. XIII. (1877—78).

34. The great modern work, reproducing the theories of Galois, and exhibiting the theory of algebraic equations as a whole, is Jordan's *Traité des Substitutions et des Équations Algébriques*, Paris, 1870. The work is divided into four books—book I.,

preliminary, relating to the theory of congruences; book II. is in two chapters, the first relating to substitutions in general, the second to substitutions defined analytically, and chiefly to linear substitutions; book III. has four chapters, the first discussing the principles of the general theory, the other three containing applications to algebra, geometry, and the theory of transcendents; lastly, book IV., divided into seven chapters, contains a determination of the general types of equations solvable by radicals, and a complete system of classification of these types. A glance through the index will show the vast extent which the theory has assumed, and the form of general conclusions arrived at; thus, in book III., the algebraical applications comprise Abelian equations, equations of Galois; the geometrical ones comprise Hesse's equation, Clebsch's equations, lines on a quartic surface having a nodal line, singular points of Kummer's surface, lines on a cubic surface, problems of contact; the applications to the theory of transcendents comprise circular functions, elliptic functions (including division and the modular equation), hyperelliptic functions, solution of equations by transcendents. And on this last subject, solution of equations by transcendents, we may quote the result,—"the solution of the general equation of an order superior to five cannot be made to depend upon that of the equations for the division of the circular or elliptic functions"; and again (but with a reference to a possible case of exception), "the general equation cannot be solved by aid of the equations which give the division of the hyperelliptic functions into an odd number of parts."