# III. Concerning Infinitesimal Calculus Proper

Of course I shall assume that you all know how to differentiate and integrate, and that you have frequently used both processes. We shall be concerned here solely with more general questions, such as the logical and psychological foundations, teaching, and the like.

## 1. General Considerations in Infinitesimal Calculus

I should like to make a *general preliminary remark concerning the range of mathematics*. You can hear often from non mathematicians, especially from philosophers, that *mathematics consists exclusively in drawing conclusions from clearly stated premises;* and that, in this process, it makes no difference what these premises signify, whether they are true or false, provided only that they do not contradict one another. But the researcher who has done productive mathematical work will talk quite differently. In fact those persons are thinking only of the crystallized form into which finished mathematical theories are finally cast. The *researcher* himself, however, in mathematics, as in every other science, does not work in this rigorous deductive fashion. On the contrary, *he makes essential use of his phantasy and proceeds inductively, aided by heuristic expedients*. One can give numerous examples of mathematicians who have discovered theorems of the greatest importance, which they were unable to prove. Should one, then, refuse to recognise this as a great accomplishment and, in deference to the above definition, insist that this is not mathematics, and that only the successors who supply polished proofs are doing real mathematics? After all, it is an arbitrary thing how the word is to be used, but no judgment of value can deny that the *inductive work of the person who first announces the theorem is at least as valuable as the deductive work of the one who first proves it*. For both are equally necessary, and the discovery is the presupposition of the later conclusion. **[224]**



F. Klein, *Elementary Mathematics from a Higher Standpoint*,
DOI 10.1007/978-3-662-49442-4_11

## *Emergence of the Infinitesimal Calculus by the Specificity of Our Sense Intuition*

It is precisely in the discovery and in the development of the infinitesimal calculus that this inductive process, built up without compelling logical steps, played such a great role; and the *most effective heuristic aid was very often sense intuition*. And I mean here the *immediate* sense intuition, with all its inexactness, for which a curve is a stroke of definite width, *not* the *abstract* intuition, which postulates a completed passage to the limit, yielding a one-dimensional line. I should like to corroborate this statement by outlining to you how the ideas of the infinitesimal calculus were developed historically.

If we take up first the *concept of an integral*, we notice that it begins historically with the *problem of measuring areas and volumes (quadrature and cubature)*. The abstract logical definition determines the integral $\int_a^b f(x)dx$, i.e., the area bounded by the curve $y = f(x)$, the $x$-axis, and the ordinates $x = a, x = b$, as the *limit of the sum of narrow rectangles inscribed in this area* when their number increases and their width decreases without bound. Sense intuition, however, makes it natural to define this area, not as this exact limit, but simply as the *sum of a large number of quite narrow rectangles*. In fact, the necessary inexactness of the drawing would
**[225]** inevitably set bounds to the further narrowing of the rectangles (see Fig. 94).

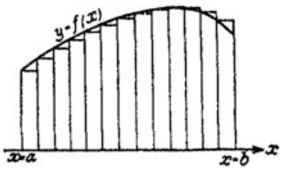


**Figure 94**

This naïve method characterizes, in fact, the thinking of the greatest researchers in the early period of infinitesimal calculus. Let me mention, first of all, *Kepler* who in his *Nova stereometria doliorum vinariorum*[133] was concerned with the volumes of bodies. His chief interest here was in the measuring of casks, and in determining their most suitable shape. He took precisely the naïve standpoint indicated above.

He thought of the volume of the barrel, as of every other body (see Fig. 95), as made up of *numerous thin leaves* suitably ranged in layers, and considered it as the sum of the volumes of these leaves, each of which was a cylinder. In a similar way he calculated the *simple geometric bodies*, e. g., the *sphere*. He thought of this as made up of a great many *small pyramids* with common vertex at the centre (see Fig. 96). Then its volume, according to the well-known formula for the pyramid, would be $r/3$ times the sum of the bases of all the small pyramids. By writing for

[133] Linz on the Danube, 1615. German in Ostwalds Klassikern, No. 165. Leipzig, 1908.

the sum of these little facets simply the surface of the sphere, or $4\pi r^2$, he obtained $4\pi r^3/3$, the correct formula for the volume.

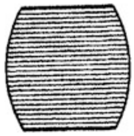

**Figure 95**

**Figure 96**

Moreover, Kepler emphasizes explicitly the practical heuristic value of such considerations, and refers, so far as rigorous mathematical proofs are concerned, to the so-called *method of exhaustion*. This method, which had been used by Archimedes, determines, for example, the area of the circle by following carefully the approximations to the area by means of inscribed and circumscribed polygons with an increasing number of sides. The essential difference between it and the modern method lies in the fact that it tacitly assumes, as self-evident, the existence of a number which measures the area of the circle, whereas the modern infinitesimal calculus declines to accept this intuitive evidence, but has recourse to the abstract concept of limit and defines this number as the limit of the numbers that measure the areas of the inscribed polygons. Granted, however, the existence of this num- **[226]**
ber, the method of exhaustion is an exact process for approximating to areas by means of the known areas of rectilinear figures, one which satisfies rigorous modern demands. The method is, however, very tedious in many cases, and ill suited to the *discovery* of areas and volumes. One of Archimedes writings[134], discovered by Johan Ludvig Heiberg in 1906, shows, in fact, that he did not use the method of exhaustion at all in his investigations. After he had first obtained his results by some other method, he developed the proof by exhaustion in order to meet the demands of that time as to rigour. For the discovery of his theorems he used a method which included considerations of the centre of gravity and the law of the lever, and also of intuition, such as, for example, that triangles and parabolic segments consist of series of parallel chords, or that cylinders, spheres, and cones are made up of series of parallel circular discs.

[134] Already referred to on p. [80].

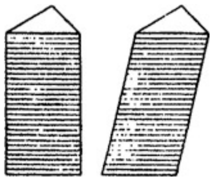

**Figure 97**

Returning now to the seventeenth century, we find considerations analogous to those of Kepler in the book of the Jesuit Bonaventura Cavalieri: *Geometria indivisibilibus continuorum nova quadam ratione promota*[135] where he sets up the principle called today by his name: *Two bodies have equal volumes if plane sections equidistant from their bases have equal areas.* This principle of Cavalieri is, as you know, much used in our schools. It is believed there that integral calculus can be avoided in this way, whereas this principle belongs, in fact, entirely to the calculus. Its establishment by Cavalieri amounts precisely to this, that he thinks of both solids as built up of layers of thin leaves which, according to the hypothesis, are congruent in pairs, i.e., one of the bodies could be transformed into the other by translating its individual leaves (see Fig. 97); but this could not alter the volume, since this consists of the same summands before and after the translation.

**Figure 98**

Naïve sense intuition leads in the same way to the *derivative of a function*, i. e., to the tangent to the curve. In this case, one replaces (and this is the way it was actually done) the curve by a polygonal line (see Fig. 98), which has on the curve a sufficient number of points, as vertices, taken close together. From the nature of our
**[227]** sense intuition we can hardly distinguish the curve from this aggregate of points and still less from the polygonal line. The *tangent* is now defined outright as the *line joining two successive points*, that is, as the prolongation of one of the sides of the polygon. From the abstract logical standpoint, this line remains only a *secant*, no matter how close together the points are taken; and the *tangent is only the limiting position approached by the secant when the distance between the points approaches zero*. Again, from this naïve standpoint, the *circle of curvature* is thought of as *the circle, which passes through three successive polygon vertices*, whereas the exact procedure defines the circle of curvature as the *limiting position of this circle when the three points approach each other*.

[135] Bononiae, 1635. First edition, 1653.

The *force of conviction* inherent in such naïve guiding reflections is, of course, different for different individuals. Some – and I include myself here – find them very satisfying. Others, again, who are gifted only on the purely logical side, find them thoroughly meaningless and are unable to see how anyone can consider them as a basis for mathematical thought. Yet considerations of this sort have often formed the beginnings of new and fruitful approaches.

Moreover, these naïve methods always rise to unconscious importance whenever in *mathematical physic, mechanics, or differential geometry* a preliminary theorem is to be set up. You all know that they are very serviceable then. To be sure, the pure mathematician is not sparing of his scorn on these occasions. When I was a student it was said that the differential, for a physicist, was a piece of brass which he treated as he did the rest of his apparatus.

In this connection, I should like to commend the *Leibniz notation*, the leading one today, because it combines with a *suitable suggestion of naïve intuition*, a certain reference to the abstract *limit process*, which is implicit in the concept. Thus, the Leibniz symbol $dy/dx$, for the derivative, reminds one, first that it comes from a *quotient;* but the $d$, as opposed to the $\Delta$ which is the usual symbol for finite difference, indicates that something new has been added, namely, the *passage to the limit.* In the same way, the integral symbol $\int y dx$ suggests the origin of the integral from a *sum* of small quantities. However, one does not use the usual sign $\sum$ for a sum, but rather a conventionalized $S^*$, which indicates here that something new has entered the process of summation. **[228]**

## *The Logical Foundation of Differential and Integral Calculus by Means of the Limit Concept (Newton and His Successors up to Cauchy)*

We shall now discuss with some detail the *logical foundation of differential and integral calculus*, and begin this by considering it in its *historical development*.

1. The *principal idea*, as the subject is taught, in general, in higher education (I need only briefly to refresh your memory here) is that infinitesimal calculus is *only an application of the general notion of limit. The derivative is defined as the limit of the quotient of corresponding finite increments of variable and function*

$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}, \quad \Delta x \neq 0$$

provided that this limit exists; and *not at all as a quotient* in which $dy$ and $dx$ have an independent meaning. In the same way, the *integral is defined as the limit of a sum*:

$$\int_a^b y\, dx = \lim_{\Delta x_i \to 0} \sum_{(i)} y_i \cdot \Delta x_i$$

* It is remarkable that many are unaware that $\int$ has this meaning.

where the $\Delta x_i$ are finite parts of the interval $a \leqq x \leqq b$, the $y_i$ corresponding arbitrary values of the function in that interval, and all the $\Delta x_i$ are to converge toward zero; *but* $y\,dx$ *does not have any actual significance as, say, a summand of a sum*. These designations are retained for the reasons of expediency which we mentioned above.

2. The conception as we have thus characterized it is set forth in precise form already by Newton himself. I refer you to a place in his principal work, the *Philosophiae Naturalis Principia Mathematica*[136] of 1687: "Ultimae rationes illae, quibuscum quantitates evanescunt, revera non sunt rationes quantitatum ultimarum, sed limites, ad quos quantitatum sine limite descrescentium rationes semper appropinquant, et quos propius assequi possunt, quam pro data quavis differentia, nunquam vero transgredi neque prius attingere quam quantitates diminuuntur in infinitum." Moreover, Newton avoids the infinitesimal calculus, as such, in the discussions in this work, although he certainly had used it in deriving his results. For, the fundamental work in which he developed his method of infinitesimal calculus was written in 1671, although it did not appear until 1736. It bears the title *Methodus Fluxionum et Serierum Infinitarum*[137].

In this, Newton develops the new calculus in *numerous examples*, without going into fundamental explanations. He makes connection here with a *phenomenon of daily life*, which suggests a passage to a limit. If one considers, namely, a motion
**[229]** $x = f(t)$ on the $x$-axis in the time $t$, then everyone has a notion as to what is meant by the *velocity* of this motion. If we analyse this motion it turns out that we mean the limiting value of the difference quotient $\Delta x/\Delta t$. *Newton made this velocity of* $x$ *with respect to the time the basis of his developments. He called it the "fluxion" of* $x$ *and wrote it* $\dot{x}$. He considered all the variables $x$, $y$ as dependent on this fundamental variable $t$, the time. Accordingly the derivative $dy/dx$ appears as the quotient of two fluxions $\dot{y}/\dot{x}$ which we now should write more fully $(dy/dt : dx/dt)$.

3. These ideas of Newton were accepted and developed by a long series of *mathematicians of the eighteenth century*, who built up the infinitesimal calculus, with more or less precision, upon the notion of limit. I shall select only a few names: Colin Maclaurin, in his *Treatise of Fluxions*[138], which as a textbook certainly had a wide influence; then Jean le Rond d'Alembert, in the great French *Encyclopédie Méthodique;* and finally Abraham Gotthelf Kästner[139], in Göttingen, in his lecture courses and books. Euler belongs primarily in this group although, with him, other tendencies also came to the front.

4. It was necessary to fill out an essential gap in all these developments, before one could speak of a *consistent system of infinitesimal calculus*. To be sure, the derivative was defined as a limit, but there was lacking a method for estimating, from it, the *increment of the function in a finite interval*. This was supplied

[136] New edition by W. Thomson and H. Blackburn, Glasgow, 1871, p. 38.
[137] J. Newtoni, *Opuscula Mathematica, philosophica, et philologica.* vol. I, p. 29. Lausanne, 1744.
[138] Edinburgh, 1742.
[139] Abraham G. Kästner, A.G., *Anfangsgründe der Analysis des Unendlichen,* Göttingen, 1760.

by the *mean value theorem;* and it was Cauchy's great service to have recognized its fundamental importance and to have made it the *starting point accordingly of differential calculus*. And it is not saying too much if, because of this, we adjudge Cauchy as the *founder of exact infinitesimal calculus* in the modern sense. The fundamental work in this connection, based on his Paris lecture courses, is his *Résumé des Leçons sur le Calcul Infinitésimal*[140], together with its second edition, of which only the first part, *Leçons sur le Calcul Différentiel*[141], was published.

The *mean value theorem*, as you know, may be stated as follows. *If a continuous function* $f(x)$ *possesses a derivative* $f'(x)$ *everywhere in a given interval, then there must be a point* $x + \vartheta h$ *between* $x$ *and* $x + h$ *such that*

$$f(x+h) = f(x) + h \cdot f'(x + \vartheta h), (0 < \vartheta < 1).$$

Note here the appearance of that $\vartheta$, peculiar to the mean value theorems, and which **[230]** to beginners often seems so strange at first. Geometrically, the theorem is fairly intuitive. It says, merely, *that between the points* $x$ *and* $x + h$ *on the curve there is a point* $x + \vartheta h$ *on the curve at which the tangent is parallel to the secant joining the points* $x$ *and* $x + h$ (see Fig. 99).

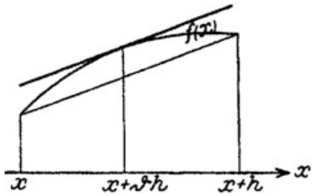


**Figure 99**

5. How can one give an *exact arithmetic proof of the mean value theorem*, without appealing to geometric intuition? Such a proof could only mean, of course, throwing the theorem back upon arithmetic definitions of variable, function, continuity etc., which would have to be set up in advance in abstract, precise form. For this reason such a rigorous proof had to wait for *Weierstraß and his followers*, to whom, also, we owe the spread of the modern arithmetic concept of the number continuum. I shall try to give you the characteristic points of the argument.

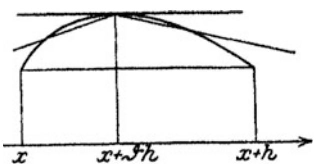


**Figure 100**

[140] Paris, 1823. *Œuvres complètes*, 2nd series, vol. 4. Paris, 1899.
[141] Paris, 1829. *Œuvres complètes*, 2nd series, vol. 4, Paris, 1899.

In the first place, it is easy to make this theorem depend on the case where the secant is horizontal, i.e. $f(x) = f(x+h)$ (see Fig. 100). One must then prove the existence of a place where the tangent is *horizontal*. To do this we can use the *theorem of Weierstraß that every function, which is continuous throughout a closed interval, actually reaches a maximum, and also a minimum value, at least once in that interval*. Because of our assumption, one of these extreme values of our function must lie *within* the interval $(x, x+h)$, provided we exclude the trivial case in which $f(x)$ is a constant. Let us suppose that there is a maximum (the case of a minimum is treated in the same way) and that it occurs at the place $x + \vartheta h$. It follows that $f(x)$ cannot have larger values, either to the right or to the left, i.e., the difference quotient to the right is negative, or zero, and to the left, positive or zero. Since the derivative exists, by hypothesis, at every point in the interval, its value at $x + \vartheta h$ can be looked upon as the limit of values, which are either not positive or not negative, according as one thinks of it as limit of a progressive or a regressive quotient of differences. Therefore it must have the value *zero*, the tangent at $x = \vartheta h$ is horizontal, and the theorem is proved.

**[231]** The scientific mathematics of today is built upon the sequence of developments, which we have been outlining. But an *essentially different conception of infinitesimal calculus* has been running parallel with this through the centuries.

## *Construction of the Infinitesimal Calculus Based on "Differentials" (Leibniz and His Followers)*

1. This conception harks back to *old metaphysical speculations concerning the structure of the continuum* according to which this is made up of ultimate indivisible "infinitely small" parts. There were already, in *ancient times*, suggestions of these indivisibles and they were widely cultivated by the *scholastics* and still further by the *Jesuit philosophers*. As a characteristic example I recall the title of *Cavalieri*'s book, mentioned on p. 226 *Geometria Indivisibilibus Continuorum Promota*, which indicates its true nature. As a matter of fact, he considers mathematical approximation in a secondary way only. He actually considers space as consisting of ultimate indivisible parts, the "indivisibilia". In this connection it would be interesting and important to know the various analyses to which the notion of the continuum has been subjected in the course of centuries (and milleniums).

2. *Leibniz*, who shares with Newton the distinction of having invented the infinitesimal calculus, also made use of such ideas. The primary thing for him was not the *derivative thought of as a limit*. The *differential dx of the variable x had for him actual existence as an ultimate indivisible part of the axis of abscissas*, as a quantity smaller than any finite quantity and still not zero (*"actually" infinitely small quantity*). In the same way, the *differentials of higher order* $d^2x, d^3x, \ldots$ are defined as infinitely small quantities of second, third, ... order, each of which is "infinitely small in comparison with the preceding". Thus one had a series of

systems of qualitatively different systems of magnitudes. According to the *theory of indivisibles*, the area bounded by the curve $y = y(x)$ and the axis of abscissas is the direct sum of all the individual ordinates. It is because of this view that Leibniz, in his first manuscript on integral calculus (1675), writes $\int y$ and not $\int y dx$.

This conception, however, is by no means the only one practiced by Leibniz. Sometimes he uses the notion of *mathematical approximation*, where, for example, the *differential dx is a finite segment but so small that, for that interval, the curve is not appreciably different from the tangent*. The above metaphysical speculations are surely only idealizations of this simple psychological fact here implied.

But there is a third direction for the mathematical ideas of Leibniz, one that is especially characteristic of him. It is his formal conception. I have frequently reminded you that we can look upon *Leibniz* as the *founder of formal mathematics*. His thought here is as follows. It makes no difference what meaning we attach to **[232]** the differentials, or whether we attach any meaning whatever to them. If we define appropriate rules of operation for them, and if we employ these rules properly, it is certain that something reasonable and correct will result. Leibniz refers repeatedly to the analogy with complex numbers, concerning which he had corresponding notions. As to these *rules of operation for differentials* he was concerned chiefly with the formula

$$f(x + dx) - f(x) = f'(x) \cdot dx.$$

The mean value theorem shows that this is correct only if one writes $f'(x + \vartheta \cdot dx)$ instead of $f'(x)$; but the error which one commits by writing $f'(x)$ outright is *infinitely small of higher* (*second*) *order, and such quantities should be neglected* (this is the most important formal rule) *in operations with differentials*.

The most important *publications* of Leibniz are contained in that famous first scientific journal, the *Acta Eruditorum*[142]; in the years 1684, 1685, and 1712. In the first volume, you find, under the title *Nova methodus pro maximis et minimis* (p. 467 et seq.), the very first publication concerning differential calculus. In this Leibniz merely develops the rules for differentiation. The later works give also *expositions of principles*, where preference is given to the *formal standpoint*. In this connection, the short article of the year 1712[143], one of the last years of his life, was especially characteristic. In this he speaks outright of theorems and definitions which are only "*toleranter vera*" or – in French – "passables": "Rigorem quidem non sustinent, habent tamen usum magnum in calculando et ad artem inveniendi universalesque conceptus valent." He has reference here to complex numbers as well as to the infinite. If we speak, perhaps, of the infinitely small, then "commoditati expressionis seu breviloquio mentalis inservimus, sed non nisi toleranter vera loquimur, quae explicatione rigidantur."

---

[142] Translated, in part, in *Ostwalds Klassiker* No. 162. Edited by Gerhard Kowalewski, Leipzig, 1908. Also in Leibniz, Mathematische Schriften. Edited by Carl Immanuel Gerhardt, from 1849 on.

[143] Observatio ...; et de vero sensu Methodi infinitesimalis, p. 167–169.

3. From Leibniz as centre the new calculus spread rapidly over the continent and we find each of his three points of view represented. I must mention here the *first textbook of differential calculus* that ever appeared, the *Analyse des Infiniment Petits pour l'Intelligence des Courbes*[144] by Marquis de 1'Hospital, a pupil of Johann Bernoulli, who for his part, had absorbed the new ideas from Leibniz with surprising speed and had himself published the first textbook on the
**[233]** integral calculus[145]. Both books represent the conception of approximation mathematics. For example, a curve is thought of as a polygon with short sides, a tangent as the prolongation of one of these sides. In Germany, the differential calculus according to Leibniz was spread widely by *Christian Wolff*, of Halle, who published the contents of his lecture courses in *Elementa matheseos universae*[146]. He introduces the differentials of Leibniz immediately, at the beginning of the differential calculus, although he emphasizes expressly that they have no *real equivalent of any kind.* And, indeed, as an aid to our intuition he develops his views concerning the infinitely small in a manner which savours thoroughly of mathematics of approximation. Thus he says, by way of example, that for purposes of practical measurement, the height of a mountain is not noticeably changed by adding or removing a particle of dust.

4. You will also frequently find the *metaphysical view*, which ascribes an actual existence to the differentials. It has always had supporters, especially on the *philosophical side*, but also among *mathematical physicists*. One of the most prominent here is Simeon-Denis Poisson, who, in the preface to his celebrated *Traite de Mécanique*[147], expressed himself in a very crass manner to the effect that the infinitely small magnitudes are not merely an aid in investigation but that they have a thoroughly real existence.

5. Due probably to the philosophic tradition, this conception went over into textbook literature and plays a marked rôle there even today. As an example, I mention the textbook by Heinrich Lübsen *Einleitung in die Infinitesimalrechnung*[148], which appeared first in 1855 and which had for a long time an extraordinary influence among a large part of the public. Everyone, in my day, certainly had Lübsen's book in his hand, either when he was a pupil, or later, and many received from it the first stimulation to further mathematical study. Lübsen defined the derivative first by means of the limit notion; but along side of this he placed (since the second edition) what he considered to be the *true infinitesimal calculus* – a *mystical scheme of operating with infinitely small quantities*. These chapters are marked with an asterisk to indicate that they bring nothing new in the way of result. There, the differentials are introduced as ultimate parts which arise, for example, by continued

[144] Paris, 1696; second edition, 1715.

[145] [Translated in Ostwalds Klassiker No. 194. Edited by Gerhard Kowalewski. Johann Bernoulli's Differentialrechnung was discovered and discussed a short time ago by Paul Schafheitlein. Verhandlungen der Naturforscher-Gesellschaft in Basel, vol. 32 (1921).]

[146] Appeared first in 1710. – Editio nova Hallae, Magdeburgiae, 1742, p. 545.

[147] Part I, second edition, p. 14. Paris, 1833.

[148] Eighth edition, Leipzig, 1899.

halving of a finite quantity an infinite, non assignable number of times; and each of these parts "although different from absolute zero is nevertheless not assignable, but an infinitesimal magnitude, a breath, an instant". And then follows an English quotation: "An infinitesimal is the ghost of a departed quantity" (p. 59, 60)[149]. Then in another place (p. 76): "The infinitesimal method is, as you see, very subtle, but correct. If this is not manifest from what has preceded, together with what follows, it is the fault only of inadequate exposition." It is certainly very interesting to read these passages. **[234]**

As companion piece to this I put before you the sixth edition of the widely used *Lehrbuch der Experimentalphysik* by Adolf Wüllner[150]. The first volume contains a brief preliminary exposition of infinitesimal calculus for the benefit of those students of natural science or medicine who have not acquired, at the Gymnasium, that knowledge of calculus, which is indispensable for physics. Wüllner begins (p. 31) with the explanation of the meaning of the infinitely small quantity $dx$, later follows with the explanation for the second differential $d^2x$, which, of course, is more difficult. I urge you to read this introduction with the eye of the mathematician and to reflect upon the absurdity of suppressing infinitesimal calculus in the schools because it is too difficult, and then of expecting a student in his first semester to gain an understanding of it from this ten page presentation, which is not only far from satisfying, but very hard to read!

The reason why such reflections could so long hold their place abreast of the mathematically rigorous method of limits, must be sought probably in the widely spread need of achieving a deeper feeling, beyond the abstract logical formulation of the method of limits, *of the intrinsic nature of continuous magnitudes*, and of forming more definite representations of them than were supplied by emphasis solely upon the psychological moment which determined the concept of limit. There is one formulation, which is characteristic, which is due, I believe, to the philosopher *Georg W. F. Hegel*, and which formerly was frequently used in textbooks and lectures. It declares that *the function $y = f(x)$ represents the being, the derivative $dy/dx$, however, the becoming of things*. There is assuredly something impressive in this, but one must recognize clearly that such words do not promote further mathematical development because this must be based upon precise concepts.

[149] Berkeley's original is (at the end of his section XXXV): "They are neither finite Quantities nor Quantities infinitely small, nor yet nothing. May we not call them the Ghosts of departed Quantities?"

[150] Leipzig, 1907.

## *The Actual Infinitely Small Quantities in the Axiomatics of Geometry*

In the most recent mathematics, "actually" infinitely small quantities have come to the front again, but in entirely different context, namely *in the geometric investiga-*
**[235]** *tions of Giuseppe Veronese* and also in *Hilbert's Grundlagen der Geometrie*[151]. The guiding thought of these investigations can be stated briefly as follows: A geometry is considered in which by indicating $x = a$ ($a$ an ordinary real number) not only *one* point on the $x$-axis is determined, but infinitely many points, whose abscissas differ by finite multiples of infinitely small quantities of different orders $\eta, \zeta, \ldots$ A point is thus determined only when one assigns

$$x = a + b\eta + c\zeta + \cdots,$$

where $a, b, c, \ldots$ are ordinary real numbers, and the $\eta, \zeta, \ldots$ actually infinitely small quantities of decreasing orders. Hilbert uses this guiding idea by subjecting these new quantities $\eta$, $\zeta$, ... to such axiomatic assumptions as will make it evident that one can operate with them consistently. To this end it is of chief importance to determine appropriately the relation as to size between $x$ and a second quantity $x_1 = a_1 + b_1\eta + c_1\zeta + \cdots$. The first assumption is that $x >$ or $< x_1$ if $a >$ or $< a_1$; but if $a = a_1$, the determination as to size rests with the second coefficient, so *that* $x \gtreqless x_1$ *according as* $b \gtreqless b_1$; *and if, in addition,* $b = b_1$, *the decision lies with the c, etc.* These assumptions will be clearer to you if you refrain from attempting to associate with the letters any sort of concrete representation.

Now it turns out that, after imposing upon these new quantities these rules, together with certain others, it is possible to operate with them as with finite numbers. One essential theorem, however, which holds in the system of ordinary real numbers, now loses its validity, namely the theorem: *Given two positive numbers e, a, it is always possible to find a finite integer n such that* $n \cdot e > a$, *no matter how small e is nor how large a may be.* In fact, it follows immediately from the above definition that an *arbitrary finite multiple* $n \cdot \eta$ *of* $\eta$ *is smaller than any positive finite number* $a$, and it is precisely this property that characterizes the $\eta$ as an infinitely small quantity. In the same way $n \cdot \zeta < \eta$, that is, $\zeta$ is an infinitely small quantity of higher order than $\eta$.

This number system is called *non-Archimedean*. The above theorem concerning finite numbers is called, namely, the *axiom of Archimedes*, because he emphasised it as an unprovable assumption, or as a fundamental one which did not need proof, in connection with the numbers which he used. The denial of this axiom *characterises the possibility of actually infinitely small quantities*. The name *Archimedean axiom*, however, like most personal designations, is historically inexact. Euclid gave prominence to this axiom more than half a century before Archimedes; and it
**[236]** is said not to have been invented by Euclid, either, but, like so many of his theorems, to have been taken over from Eudoxus of Knidos. The study of non-Archimedean

[151] Fifth edition, Leipzig, 1922.

quantities[152], which have been used especially as coordinates in setting up a non-Archimedean geometry, aims at deeper knowledge of the nature of continuity and belongs to the large group of investigations concerning the logical dependence of different axioms of ordinary geometry and arithmetic. For this purpose, the method is always to set up artificial number systems for which only a part of the axioms hold, and to infer the logical independence of the remaining axioms from these.

The question naturally arises *whether, starting from such number systems, it would be possible to modify the traditional foundations of infinitesimal calculus, so as to include actually infinitely small quantities in a way that would satisfy modern demands as to rigour;* in other words, to construct a non-Archimedean analysis. The first and chief problem of this analysis would be to prove the mean value theorem

$$f(x+h) - f(x) = h \cdot f'(x + \vartheta h)$$

from the assumed axioms. I will not say that progress in this direction is impossible, but it is true that none of the many researchers who have busied themselves with actually infinitely small quantities have achieved anything positive.

I remark for your orientation that, since Cauchy's time, the words *infinitely small* are used in modern textbooks in another sense. One never says, namely, that a quantity *is* infinitely small, but rather that it *becomes* infinitely small; which is only a convenient expression implying that the quantity decreases without bound toward zero.

We must bear in mind the *reaction*, which was evoked by the use of infinitely small quantities in infinitesimal calculus. People soon sensed the mystical, the unproven, in these ideas, and there arose often a prejudice, as though the differential calculus were a *particular philosophical system* which could not be proved, which could only be believed or, to put it bluntly, a *fraud.* One of the keenest critics, in this sense, was the philosopher Bishop *George Berkeley*, who in the little book *The Analyst*[153] assailed in an amusing manner the lack of clearness which prevailed in the mathematics of his time. Claiming the privilege of exercising the same freedom in criticizing the principles and methods of mathematics, "which the mathematicians employed with respect to the mysteries of religion", he launched a violent attack **[237]** upon all the methods of the new analysis, the calculus with fluxions as well as the operation with differentials. He came to the conclusion that the entire structure of analysis was obscure and thoroughly unintelligible.

Similar views have often maintained themselves even up to the present time, especially on the philosophical side. This is due, perhaps, to the fact that acquaintance here is confined to the operation with differentials; the rigorous method of limits, a rather recent development, has not been comprehended. As an example, let me quote from Johann Julius Baumann's *Raum, Zeit und Mathematik*[154] which

[152] [The so-called horn-shaped angles, known already to Euclid, are examples of non-Archimedean quantities. Compare also the excursus, in the second volume of this work, after the critique of Euclid's *Elements.*]
[153] London, 1734.
[154] Vol. 2, p. 55, Berlin, 1869.

appeared in the sixties: "Thus we discard the logical and metaphysical justification, which Leibniz gave to calculus, but we decline to touch this calculus itself. We look upon it as an ingenious invention, which has turned out well in practice; as an art rather than a science. It cannot be constructed logically. It does not follow from the elements of ordinary mathematics..."

## *The Reaction: the Derivative Calculus of Lagrange*

This reaction against differentials accounts also for the attempt by Lagrange, already mentioned, but appearing now in a new light, in his *Théorie des Fonctions Analytiques*, published in 1797, to eliminate from the theory not only infinitely small quantities, but also every passage to the limit. He confined himself, namely, to those functions, which are defined by power series

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots ,$$

and he defines formally the "derived function $f'(x)$" (he avoids characteristically the expression derivative and the sign $dy/dx$) by means of a new power series

$$f'(x) = a_1 + 2a_2 x + 3a_3 x^2 + \cdots .$$

Consequently he talks of *derivative calculus* instead of *differential calculus.*

This presentation, of course, could not be permanently satisfactory. In the first place, the concept of function used here is, as we have shown, much too limited. More than that, however, such thoroughly formal definitions make a deeper comprehension of the nature of the differential coefficient impossible, and take no account of what we called the *psychological moment* – they leave entirely unexplained just why one should be interested in a series obtained in such a peculiar way. Finally, one can get along without giving any thought to a limit process *only by disregarding*
**[238]** *entirely the convergence of these series and the question within what limits of error they can be replaced by finite partial sums.* As soon as one begins a consideration of these problems, which is essential, of course, for any actual use of the series, it is necessary to have recourse precisely to that concept of limit, the avoidance of which was the purpose of inventing the system.

It would be fitting, perhaps, to say a few words about the differences of opinion concerning the foundations of calculus, as these come up, even today, beyond the narrow circle of professional mathematicians. I believe that we can often find here the preliminary conditions for an agreement, in considerations very similar to those, which we set forth respecting the foundations of arithmetic (p. [15] et seq.). In every discipline of mathematics one must separate sharply the question as to the inner logical consistency of its structure from that as to the justification for applying its axiomatically and (so to speak) arbitrarily formulated concepts and theorems to objects of our external or internal perception. Georg Cantor[155] makes the distinction, with reference to integer numbers, between *immanent reality*, which belongs

[155] *Mathematische Annalen*, vol. 21 (1883), p. 562.

to them by virtue of their logical definability, and *transient reality*, which they possess by virtue of their applicability to concrete things. In the case of infinitesimal calculus, the first problem is completely solved by means of those theories, which the science of mathematics has developed in logically complete manner (through the use of the limit concept). The second question belongs entirely to the theory of knowledge, and the mathematician contributes only to its precise formulation by separating the first part and solving it. No pure mathematical work can, from its very nature, supply any immediate contribution to the solution of the epistemological part. (See the analogous remarks on arithmetic, p. [15] et seq.) All disputes concerning the foundations of infinitesimal calculus labour under the disadvantage that these two entirely different phases of the problem have not been sharply enough separated. In fact, the first, the purely mathematical part, is established here precisely as in all other branches of mathematics, and the difficulties lie in the second, the philosophical part. The value of investigations which press forward in this second direction take on especial importance in view of these considerations; but it becomes imperative to make them depend upon exact knowledge of the results of the purely mathematical work upon the first problem.

I am concluding with this excursus of our short historical sketch of the development of infinitesimal calculus. In it I was obliged of course to confine myself to an **[239]** emphasis of the most important guiding ideas. It should be extended, naturally, by a thorough-going study of the entire literature of that period. You will find many interesting references in the lecture given by Max Simon at the Frankfurt meeting of the natural scientists of 1896: *Zur Geschichte und Philosophie der Differentialrechnung*.

## *Form and Importance of the Infinitesimal Calculus in the Present State of Teaching*

If we now examine, finally, the attitude towards infinitesimal calculus in school teaching, we shall see that the entire course of its historical development is mirrored there to a certain extent. In earlier times, when infinitesimal calculus was taught in the schools, there existed *by no means a clear notion of its exact scientific structure as based on the method of limits*. At least this was manifest in the textbooks, and it was doubtless the same in teaching itself. This method cropped up in a vague way at most, whereas operations with infinitely small quantities and sometimes also derivative calculus, in the sense of Lagrange, came to the front. Such instruction, of course, lacked not only rigour but intelligibility as well, and it is easy to see why a marked aversion arose to the treatment of infinitesimal calculus at all in the schools. This culminated in the seventies and eighties in an official order forbidding this instruction even in the "realist" school types.

To be sure this did not entirely prevent (as I indicated earlier) the use of the *limit method* in the schools, where it was necessary – one merely avoided that name, or one even thought sometimes that something else was being taught. I shall mention here only *three examples*, which most of you will recall from your school days.

a) The well-known *calculation of the perimeter and the area of the circle* by an approximation, which uses the inscribed and circumscribed regular polygons is obviously nothing but an *integration.* It was employed, even in ancient times, and was used particularly by *Archimedes;* in fact, it is owing to this classical heritage that is has been retained in the schools.

b) Instruction in physics, and particularly in mechanics, necessarily involves the notions of *velocity* and *acceleration*, and their use in various deductions, including the *laws of falling bodies.* But the derivation of these laws is essentially identical with the *integration of the differential equation* $z'' = g$ by means of the function $z = \frac{1}{2}gt^2 + at + b$, where $a, b$ are constants of integration. The schools *must* solve this problem, under pressure of the demands of physics, and the means, which they employ are more or less exact methods of integration, of course disguised.

**[240]** c) In many North German schools the *theory of maxima and minima* was taught according to a method which bore the name of *Karl Heinrich Schellbach*, the prominent mathematical pedagogue of whom you all must have heard. According to this method one puts

$$\lim_{x \to x_1} \left( \frac{f(x) - f(x_1)}{x - x_1} \right) = 0$$

in order to obtain the extremes of the function $y = f(x)$. But that is precisely the method of differential calculus, only that the word “Differentialquotient” is not used. It is certain that Schellbach used the above expression only because differential calculus was prohibited in the schools and he nevertheless did not want to miss these important notions. His disciples, however, took it over unchanged, called it by his name, and so it came about that methods, which Fermat, Leibniz, and Newton had possessed were put before the pupils under the name of Schellbach!

Let me now indicate, finally, the attitude toward these things of our *reform tendency*, which is now gaining ground more and more in Germany, as well as elsewhere, especially in France, and which we hope will dominate the mathematical teaching of the next decades. *We desire that the concepts which are expressed by the symbols* $y = f(x), dy/dx, \int y dx$ *be made familiar to pupils, under these designations; not, indeed, as a new abstract discipline, but as an organic part of the total teaching; and that one advance slowly, beginning with the simplest examples.* Thus one might begin, with pupils of the age of fourteen and fifteen (*Obertertia* and *Untersekunda*), by treating extensively the functions $y = ax + b$ ($a, b$ definite numbers) and $y = x^2$, drawing them on millimetre paper, and letting the concepts *slope* and *area* arise slowly by these means. But one should hold to concrete examples. During the last three years (*Oberstufe* of the Gymnasia) this knowledge could be systematised, the result being that the pupils would come into *complete possession of the elements of infinitesimal calculus*. It is essential here to make it clear to the pupil that he is dealing, not with something mystical, but with simple things that anyone can understand.

The irrefutable necessity of such reforms lies in the fact that they are concerned with those mathematical notions, which govern completely the applications of mathematics today in every possible field, and without which all studies in higher education, even the simplest studies in experimental physics, are suspended in mid air. I can be content with these few hints, chiefly because this subject is fully discussed in Klein-Schimmack (referred to on p. [3]).

In order to supplement these general considerations with something, which again is concrete I shall now discuss in some detail an especially important subject in infinitesimal calculus. **[241]**

## 2. Taylor's Theorem

I shall proceed here in a manner analogous to the plan I followed with trigonometric series. I shall depart, namely, from the usual treatment in the textbooks by bringing to the foreground the *finite series*, so important in practice, and by aiding the *intuitive grasp of the situation by means of graphs*. In this way it will all seem elementary and easily comprehensible.

### *The First Parabolas of Osculation*

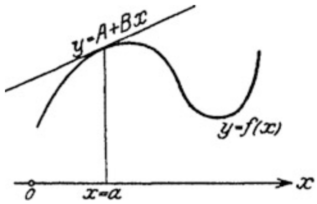


**Figure 101**

We begin with the question whether we can make a suitable approximation to the shape of an arbitrary curve $y = f(x)$, for a short part of it, by means of curves of the simplest kind. The most obvious thing is to replace the curve in the neighbourhood of a point $x = a$ by its rectilinear tangent

$$y = A + Bx,$$

just as in physics and in other applications, we often discard the higher powers of the independent variable in a series expansion (see Fig. 101). In a similar manner we can obtain better *approximations* by making use of *parabolas of second, third, ... order*

$$y = A + Bx + Cx^2, \quad y = A + Bx + Cx^2 + Dx^3, \ldots$$

or, in analytic terms, by using *polynomials of higher degree*. Polynomials are especially suitable because they are so easy to calculate. We shall give all these curves a special position, so *that at the point $x = a$ they snuggle as close as possible to the curve, i.e., so that they shall be parabolas of osculation*. Thus the quadratic parabola will coincide with $y = f(x)$ not only in its ordinate but also in its first and second derivatives (i.e., it will "osculate"). A simple calculation shows that the analytic expression for the parabola having osculation of order $n$ will be

$$y = f(a) + \frac{f'(a)}{1}(x-a) + \frac{f''(a)}{1\cdot 2}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{1\cdot 2\cdot \ldots \cdot n}(x-a)^n\,,$$
$$(n = 1, 2, 3, \ldots)$$

and these are precisely the *first* $n+1$ *terms of Taylor's series.*

The investigation as to *whether and how far these polynomials represent usable*
[242] *curves of approximation* will be started by a *somewhat experimental method*, such as we used in the case (p. [209]) of the trigonometric series. I shall show you a few drawings of the first osculating parabolas of simple curves, which were made[156] by Schimmack. The first are the four following functions, all having a singularity at $x = -1$, drawn with their parabolas of osculation at $x = 0$ (see Figs. 102, 103, 104, 105).

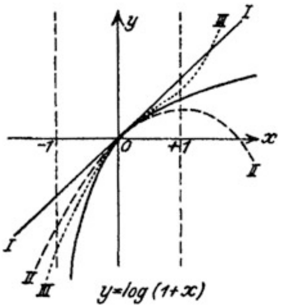


**Figure 102**

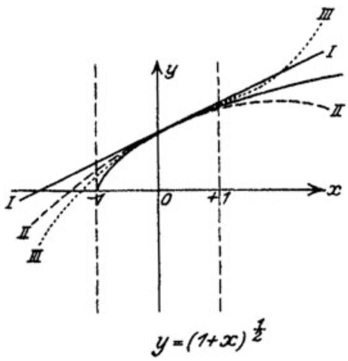


**Figure 103**

[156] Four of these drawings accompanied Schimmack's report on the Göttingen vacation course, Easter, 1908: *Über die Gestaltung des mathematischen Unterrichts im Sinne der neueren Reformideen,* Zeitschrift für den mathematischen und naturwissenschaftlichen Unterricht, vol. 39 (1908), p. 513; also separate reprints. Leipzig, 1908.

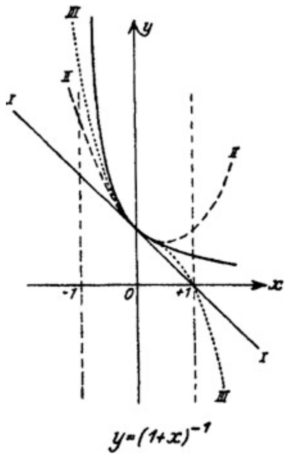


**Figure 104**

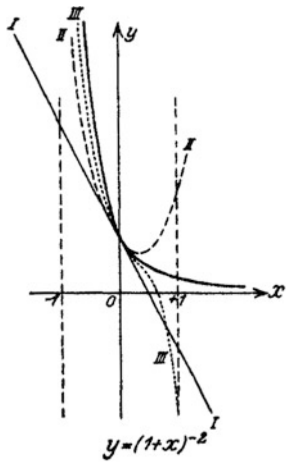


**Figure 105**

**[243]**

1. $$\log(1+x) \approx x - \frac{x^2}{2} + \frac{x^3}{3} - + \cdots,$$

2. $$(1+x)^{\frac{1}{2}} \approx 1 + \frac{x}{2} - \frac{x^2}{8} + \frac{x^3}{16} - + \cdots,$$

3. $$(1+x)^{-1} \approx 1 - x + x^2 - x^3 + - \cdots,$$

4. $$(1+x)^{-2} \approx 1 - 2x + 3x^2 - 4x^3 + - \cdots.$$

In the interval $(-1, +1)$ the osculating parabolas approach the original curve more and more as the order increases; but to the right of $x = +1$ they deviate from it increasingly, now above, now below, in a striking way.

*At the singular point* $x = -1$, in cases 1, 3, 4, where the original function becomes infinite, the ordinates of the successive parabolas assume *always greater values*. In case 2, where the branch of the original curve which appears in the drawing, ends in $x = -1$ at a vertical tangent, all the parabolas extend beyond this point but approach the original curve more and more at $x = -1$, by becoming ever steeper. At the *point* $x = +1$, symmetrical to $x = -1$, the parabolas in the first two cases approach the original curve more and more closely. In case 3, their ordinates are alternately equal to 1 and 0, while that of the original curve has the value $\frac{1}{2}$. In case 4, finally, the ordinates increase indefinitely with the order, and alternate in sign.

We shall examine, now, sketches of the osculating parabolas of two integer transcendental functions (see Fig. 106, 107)

5. $$e^x \approx 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots,$$

6. $$\sin x \approx x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots.$$

[244] You notice that as their order increases, the parabolas give usable approximations to the original curve for a greater and greater interval. It is especially striking in the case of $\sin x$ how the parabolas make the effort to share more and more oscillations with the sine curve.

I call your attention to the fact that the drawing of such curves in simple cases is perhaps a suitable topic even for the schools. After we have thus assembled our experimental material we must consider it mathematically.

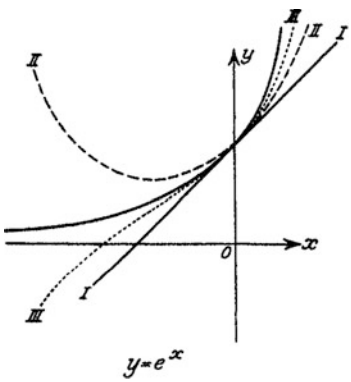


**Figure 106**

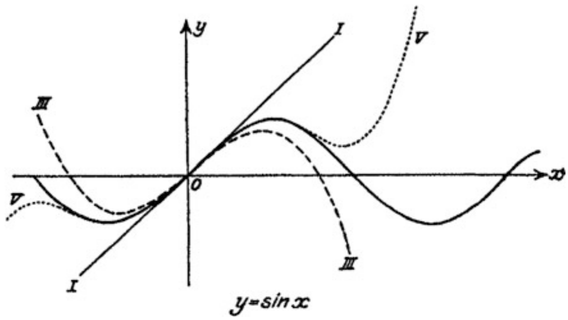


**Figure 107**

## *Increasing the Order: Questions of Convergence*

After having thus collected material from experience we have now to approach the *mathematical investigation*. The first question here is the extremely important one in practice as to the *exactness with which the n-th parabola of osculation represents the original curve*. This implies an estimate of the remainder for the values of the ordinate, and is connected naturally with the passage of $n$ to infinity. *Can the curve be represented exactly, at least for a part of its course, by an infinite power series?*

It will be sufficient to state the commonest of the theorems concerning the remainder:

$$R_n(x) = f(x) - \left\{ f(a) + \frac{x-a}{1!} f'(a) + \cdots + \frac{(x-a)^{n-1}}{(n-1)!} f^{(n-1)}(a) \right\}.$$

The proof of the theorem is given in all the books and I shall revert to it later, anyway, from a more general standpoint. The theorem is: *There is a value* $\xi$ *between* $a$ *and* $x$ *such that* $R_n$ *can be represented in the form*

$$R_n(x) = \frac{(x-a)^n}{n!} f^{(n)}(\xi), \quad (a < \xi < x).$$

The question as to the justification of the transition to an infinite series is now reduced to that as to whether this $R_n(x)$ has the limit 0 or not when $n$ becomes infinite.

Returning to our examples, it appears, as you can verify by reading in any textbook, that in cases 5 and 6 the infinite series converges for all values of $x$. In cases 1 to 4, it turns out that the series *converges*, *between* $-1$ *and* $+1$, to the original function, but that it diverges *outside* this interval. For $x = -1$ we have, in case 2, convergence to the function value; in cases $1, 3, 4$, the limiting value of the series as well as that of the function is infinite, so that one could speak of convergence here also, but it is not customary to use this word with a series that has a definitely infinite limit. For $x = +1$, finally, we have convergence in the first two examples, divergence in the last two. All this is in fullest agreement with our figures.

We may now raise the question, as we did with the trigonometric series, as to the *limit values toward which the approximating parabolas converge, thought of as*
**[245]** *complete curves*. They cannot, of course, break off suddenly at $x = \pm 1$. For the case of $\log(1 + x)$ I have sketched for you the limit curve (Fig. 108). The even and odd parabolas have different limiting positions, (indicated in the figure by dashes and dots) which consist of the logarithm curve between $-1$ and $+1$ together with the lower and upper portions, respectively, of the vertical line $x = +1$. The other three cases are similar.

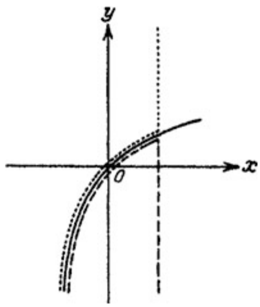


**Figure 108**

The theoretical consideration of Taylor's series cannot be made complete without going over to the complex variable. It is only then that one can understand the sudden ceasing of the power series to converge at points where the function is entirely regular. To be sure, one might be satisfied, in the case of our examples, by saying that the series cannot converge any farther to the right than to the left, and that the convergence must cease at the left because of the singularity at $x = -1$. But such reasoning would no more fit a case like the following. The Taylor's series expansion for the branch of $\tan^{-1} x$, which is regular for all real $x$

$$\tan^{-1} x \approx x - \frac{x^3}{3} + \frac{x^5}{5} - + \cdots$$

converges only in the interval $(-1, +1)$, and the parabolas of osculation converge alternately to two different limiting positions (see Fig. 109). The first consists, in the figure, of the long dotted parts of the vertical lines $x = +1$, $x = -1$ together with the portion of the inverse tangent[157] curve lying between these verticals. The

[157] [Translator's note: Inverse tangent also called arc tangent. See Klein's explanation in the part on trigonometric functions.]

second limiting position is obtained from the first by taking the short dotted parts of the vertical lines instead of the long dotted parts. The convergence is toward the first of these limit curves when we take an odd number of terms in the series, toward the second when we take an even number. In the figure, the long dotted curve represents $y = x - x^3/3 + x^5/5$, the short dotted curve is $y = x - x^3/3$. The sudden cessation **[246]**
of convergence at the thoroughly regular points $x = \pm 1$ is incomprehensible if we limit ourselves to real values of $x$ regarding the behaviour of the function. The explanation is to be found in the important theorem on the *circle of convergence*, the most beautiful of Cauchy's function-theoretic achievements, which can be stated as follows. *If one marks on the complex x-plane all the singular points of the analytic function $f(x)$, when $f(x)$ is single-valued, and on the Riemann surface belonging to $f(x)$ when $f(x)$ is many-valued, then the Taylor's series corresponding to a regular point $x = a$ converges inside the largest circle, which one can place on the respective sheet of the Riemann surface in such a manner about a that no singular point lies in its interior* (i.e., so that at least one singular point lies on its periphery). *The series converges for no point outside this circle* (see Fig. 110).

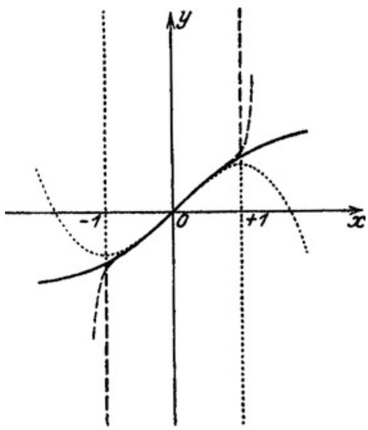


**Figure 109**

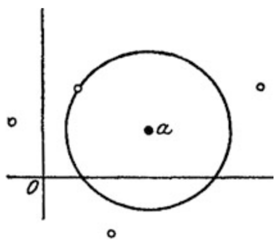


**Figure 110**

Now our example $\tan^{-1} x$ has, as you know, singularities at $x = \pm i$, and the circle of convergence of the expansion in powers of $x$ is consequently the unit circle about $x = 0$. The convergence must cease therefore at $x = \pm 1$, since the real axis leaves the circle of convergence at these points (see Fig. 111).

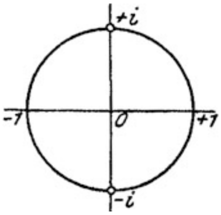


**Figure 111**

Finally, as to the convergence of the series *on the unit circle itself*, I shall give you the reference, which came up when we were talking about the *connection between power series and trigonometric series*. The convergence depends upon whether or not the real and the imaginary part of the function, in view of the singularities that must necessarily exist on the circle of convergence, can be expanded there into a convergent trigonometric series or not.

## *Generalising Taylor's Theorem to a Theorem of Finite Differences*

I should like now to enliven the discussion of Taylor's theorem by showing its relations to *the problems of interpolation and of finite differences. There, also, we are concerned with the approximation to a given curve by means of a parabola*. But instead of trying to make the parabola snuggle as closely as possible at *one* point, we require it to *cut the given curve in a number of pre-assigned points;* and the question is, again, as to how far this "interpolation parabola" gives a reasonable approximation. In the simplest case, this amounts to replacing the curve by a
**[247]** *secant* instead of the tangent (see Fig. 112). Similarly one passes a quadratic parabola through three points of the given curve, then a cubic parabola through four points, and so on.

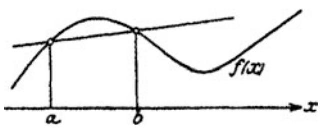


**Figure 112**

This is a natural way of approaching interpolation, one that is very often employed, e.g., in the use of logarithmic tables. There we assume that the logarithmic curve runs rectilinearly between two given tabular values and we interpolate "linearly" in the well known way, which is facilitated by the "difference tables". If this approximation is not close enough, we apply quadratic interpolation.

*From this broad statement of the general problem, we get a determination of the osculating parabolas in Taylor's theorem as a special case*, that is, when we simply allow the intersections with the interpolation parabolas to coincide in one point. To be sure, the replacing of the curve by these osculating parabolas is not properly expressed by the word "interpolation", except that one includes "extrapolation" in the problem of interpolation. For example, the curve is compared not only with the part of the secant lying *between* its points of intersection, but also with the part *beyond*. For the entire process the comprehensive word *approximation* seems more suitable.

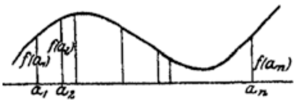


**Figure 113**

I shall now give the *most important formulas of interpolation*. Let us first determine the parabolas of order $n-1$ which cut the given function in the points $x = a_1, a_2, \ldots, a_n$, that is, whose ordinates in these points are $f(a_1), f(a_2), \ldots, f(a_n)$, (see Fig. 113). This problem, as you know, is solved by Lagrange's interpolation formula

$$(1)\qquad \begin{cases} y = \dfrac{(x-a_2)(x-a_3)\cdots(x-a_n)}{(a_1-a_2)(a_1-a_3)\cdots(a_1-a_n)} \cdot f(a_1) \\[2ex] \quad + \dfrac{(x-a_1)(x-a_3)\cdots(x-a_n)}{(a_2-a_1)(a_2-a_3)\cdots(a_2-a_n)} \cdot f(a_2) \\[2ex] \quad + \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{cases}$$

It contains $n$ terms with the factors $f(a_1), f(a_2), \ldots, f(a_n)$. The numerators lack in succession the factors $(x-a_1), (x-a_2), \ldots, (x-a_n)$. It is easy to verify the correctness of the formula. For, each summand of $y$, and hence $y$ itself, is a polynomial in $x$ of degree $n-1$. If we put $x = a_1$ all the fractions vanish except the first, which reduces to 1, so that we get $y = f(a_1)$. Similarly we get $y = f(a_2)$ for $x = a_2$, etc.

From this formula it is easy to derive, by specialization, one that is often called *Newton's formula*. This has to do with the case where the abscissas $a_1, \ldots, a_n$ are **[248]** equidistant (see Fig. 114). As the notation of the calculus of finite differences is advantageous here we shall first introduce it.

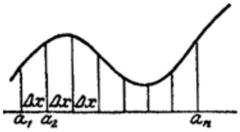


**Figure 114**

Let $\Delta x$ be any increment of $x$ and let $\Delta f(x)$ be the corresponding increment of $f(x)$ so that

$$f(x + \Delta x) = f(x) + \Delta f(x).$$

Now $\Delta f(x)$ is also a function of $x$ which, if we change $x$ by $\Delta x$, will have a definite difference called the second difference, $\Delta^2 f(x)$, so that

$$\Delta f(x + \Delta x) = \Delta f(x) + \Delta^2 f(x).$$

In the same way we have

$$\Delta^2 f(x + \Delta x) = \Delta^2 f(x) + \Delta^3 f(x), \quad \text{etc.}$$

This notation is precisely analogous to that of differential calculus, except that one is concerned here with finite quantities and there is no passing to the limit.

From the above definitions of differences there follows at once for the values of $f$ at the successive equidistant places

$$(2) \quad \left\{ \begin{aligned} f(x + \Delta x) &= f(x) + \Delta f(x), \\ f(x + 2\Delta x) &= f(x + \Delta x) + \Delta f(x + \Delta x) \\ &= f(x) + 2\Delta f(x) + \Delta^2 f(x), \\ f(x + 3\Delta x) &= f(x + 2\Delta x) + \Delta f(x + 2\Delta x) \\ &= f(x) + 3\Delta f(x) + 3\Delta^2 f(x) + \Delta^3 f(x), \\ f(x + 4\Delta x) &= f(x) + 4\Delta f(x) + 6\Delta^2 f(x) + 4\Delta^3 f(x) + \Delta^4 f(x) \end{aligned} \right. .$$

This table could be continued, the values at equidistant points being expressed by means of successive differences taken at the initial point $x$ and involving the *binomial coefficients* as factors.

*Newton's formula for the interpolation parabola of order* $(n - 1)$ *belonging to the n equidistant points of the x-axis*,

$$a_1 = a, \quad a_2 = a + \Delta x, \quad \ldots, \quad a_n = a + (n-1)\Delta x,$$

that is, which has at these points the same ordinates as $f(x)$, will be

$$(3) \quad \left\{ \begin{aligned} y = f(a) &+ \frac{(x-a)}{1!}\frac{\Delta f(a)}{\Delta x} + \frac{(x-a)(x-a-\Delta x)}{2!}\frac{\Delta^2 f(a)}{(\Delta x)^2} + \cdots \\ &+ \frac{(x-a)(x-a-\Delta x)\cdots(x-a-(n-2)\Delta x)}{(n-1)!}\frac{\Delta^{n-1} f(a)}{(\Delta x)^{n-1}} \end{aligned} \right. .$$

This is, in fact, a polynomial in $x$ of order $n - 1$. For $x = a$ it reduces to $f(a)$; for $x = a + \Delta x$ all the terms, except the first two, become zero and there remains
[249] $y = f(a) + \Delta f(a)$, which by (2) is equal to $f(a + \Delta x)$; and so on. Thus the table (2) yields a polynomial, which assumes the correct values at all the $n$ points.

## *Cauchy's Estimate of the Error*

If we wish to use this interpolation formula to real advantage, however, we must know something as to the correctness with which it represents $f(x)$, that is, we must be able to *estimate the remainder. Cauchy* gave[158] the formula for this in 1840, and I should like to derive it. I shall start from the *more general Lagrange formula*. Let $x$ be any value between the values $a_1, a_2, \ldots, a_n$, or beyond them (interpolation or extrapolation). We denote by $P(x)$ the ordinate of the interpolation parabola given by the formula and by $R(x)$ the remainder

$$f(x) = P(x) + R(x). \tag{4}$$

According to the definition of $P(x)$ the remainder $R$ vanishes for $x = a_1, a_2, \ldots, a_n$ and we therefore set

$$R(x) = \frac{(x-a_1)\,(x-a_2)\cdots(x-a_n)}{n!}\psi(x).$$

It is convenient to take out the factor $n!$ Then it turns out, in complete analogy with the remainder term of Taylor's series, *that* $\psi(x)$ *is equal to the n-th derivative of* $f(x)$ *taken for a value* $x = \xi$ *lying between the* $n-1$ *points* $a_1, a_2, \ldots, a_n, x$. This assertion that the deviation of $f(x)$ from the polynomial of order $n-1$ depends upon the entire course of the function $f^{(n)}(x)$ seems entirely plausible, if we reflect that $f(x)$ is equal to that polynomial when $f^{(n)}(x)$ vanishes.

As to the *proof of the remainder formula*, we derive it by the following device. Let us set up, as a function of a new variable $z$, the expression

$$F(z) = f(z) - P(z) - \frac{(z-a_1)\,(z-a_2)\cdots(z-a_n)}{n!}\psi(x),$$

where the variable $x$ remains as a parameter in $\psi(x)$. Now $F(a_1) = F(a_2) = \cdots = F(a_n) = 0$, since $P(a_1) = f(a_1), P(a_2) = f(a_2), \ldots, P(a_n) = f(a_n)$ by definition. Furthermore $F(x) = 0$ because the last summand goes over into $R(x)$, for $z = x$, so that the right side vanishes by (4). We know, therefore, $n+1$ zeros $z = a_1, a_2, \ldots, a_n, x$, of $F(z)$. Now apply the *extended mean value theorem*, which one gets by repeated application of the ordinary theorem (p. [230]), namely: *If a continuous function, together with its first n derivatives, vanishes at* $n+1$ *points, then the n-th derivative vanishes at one point, at least, which lies in the interval containing all the zeros.* Hence if $f(z)$, and therefore also $F(z)$, has $n$ continuous **[250]** derivatives, there must be a value $\xi$ between the extremes of the values $a_1, a_2, \ldots, a_n, x$ for which

$$F^{(n)}(\xi) = 0.$$

But we have

$$F^{(n)}(z) = f^{(n)}(z) - \psi(x),$$

[158] *Comptes Rendus*, vol. 11, pp. 775–789. – *Œuvres*, 1st series, vol. 5, pp. 409 to 424, Paris, 1885.

since the polynomial $P(z)$ of degree $n-1$ has 0 for its $n$-th derivative and since only $z^n \cdot \psi(x)/n!$, the highest term of the last summand, has an $n$-th derivative which does not vanish. Therefore we have, finally

$$F^{(n)}(\xi) = f^{(n)}(\xi) - \psi(x) = 0, \quad \text{or} \quad \psi(x) = f^{(n)}(\xi),$$

which we wished to prove.

I shall write down *Newton's interpolation formula with its remainder term*

$$(5)\quad \left\{ \begin{aligned} f(x) = f(a) &+ \frac{x-a}{1!}\frac{\Delta f(a)}{\Delta x} + \frac{(x-a)(x-a-\Delta x)}{2!}\frac{\Delta^2 f(a)}{(\Delta x)^2} + \cdots \\ &+ \frac{(x-a)\cdots[x-a-(n-2)\Delta x]}{(n-1)!}\frac{\Delta^{(n-1)} f(a)}{(\Delta x)^{n-1}} \\ &+ \frac{(x-a)\cdots[x-a-(n-1)\Delta x]}{n!} f^{(n)}(\xi) \end{aligned} \right. ,$$

*where $\xi$ is a mean value in the interval containing the $n+1$ points $a, a+\Delta x$, $a+2\Delta x, \ldots, a+(n-1)\Delta x, x$.* The formula (5) is, in fact, indispensable in the applications. I have already alluded to *linear interpolation when logarithmic tables are used*. If $f(x) = \log x$ and $n = 2$, we find, from (5)

$$\log x = \log a + \frac{x-a}{1!}\frac{\Delta \log a}{\Delta x} - \frac{(x-a)(x-a-\Delta x)}{2!}\frac{M}{\xi^2}.$$

Since $d^2 \log x/dx^2 = -M/x^2$ where $M$ is the modulus of the logarithmic system. Hence we have an expression for the error, which we commit when we interpolate linearly between the tabular logarithms for $a$ and $a+\Delta x$. This error has different signs according as $x$ lies between $a$ and $a+\Delta x$ or outside this interval. Everyone who has to do with logarithmic tables should really know this formula.

I shall not devote any more attention to applications, but shall now *draw your attention to the marked analogy between the interpolation formula of Newton and the formula of Taylor*. There is a substantial reason for this analogy. *It is easy to give an exact deduction of Taylor's theorem from the Newtonian formula*, corresponding to the passage to the limit from interpolation parabolas to osculating parabolas.
**[251]** Thus, if we keep $x$, $a$, and $n$ fixed and let $\Delta x$ converge to zero, then, since $f(x)$ has $n$ derivatives, the $n-1$ difference quotients in (5) go over into the derivatives

$$\lim_{\Delta x \to 0} \frac{\Delta f(a)}{\Delta x} = f'(a), \quad \lim_{\Delta x \to 0} \frac{\Delta^2 f(a)}{\Delta x^2} = f''(a), \ldots$$

In the last term of (5), the value of $\xi$ can change with decreasing $\Delta x$. Since all the other terms on the right have definite limits, however, and the left side has the fixed value $f(x)$ during the entire limit process, it follows that the values of $f^{(n)}(\xi)$ must converge to a definite value, too, and that this value, furthermore, must, because of the continuity of $f^{(n)}$, be a value of this function for some point between $a$ and $x$. If we denote this again by $\xi$ we have

$$f(a) = f(a) + \frac{x-a}{1!} f'(a) + \cdots + \frac{(x-a)^{n-1}}{(n-1)!} f^{(n-1)}(a) + \frac{(x-a)^n}{n!} f^{(n)}(\xi)\,,$$
$$(a < \xi < x)$$

Thus we have *obtained a complete proof of Taylor's theorem with the remainder term and at the same time have given it an ordered place in the theory of interpolation*.

It seems to me that this proof of Taylor's theorem, which brings it into wider relation with very simple questions and which provides such a smooth passage to the limit, is the *very best possible one*. But not all the mathematicians to whom these things are familiar (it is remarkable that they are unknown to many, including perhaps even some authors of textbooks) do think so. They are accustomed to confront a passage to a limit with a very grave face and would therefore prefer a direct proof of Taylor's theorem to one linking it with the calculus of finite differences.

## *Historical Remarks About Taylor and Maclaurin*

I must emphasize however that the historical source for the discovery of Taylor's theorem is actually the calculus of finite differences. I have already mentioned that Brook Taylor first published it in his *Methodus incrementorum*[159]. He first deduces there Newton's formula, evidently without the remainder, of course, and then lets pass in it simultaneously $\Delta x$ to 0 and $n$ to $\infty$. He thus gets correctly from the first terms of Newton's formula the first terms of his new series:

$$f(a) = f(a) + \frac{x-a}{1!}\frac{df(a)}{da} + \frac{(x-a)^2}{2!}\frac{d^2 f(a)}{da^2} + \cdots$$

The continuation of this series, according to the same law, seems to him self-evident, and he gives no thought either to a remainder term or to convergence. We have here, in fact, a *passage to the limit of unexampled audacity.* The first terms, in which $x-a-\Delta x, x-a-2\Delta x, \ldots$ appear, offer no difficulty, because these **[252]**
finite multiples of $\Delta x$ approach zero with $\Delta x$; but with increasing $n$ there appear terms in ever increasing number, presenting more factors $x-a-k\Delta x$ with larger and larger $k$, and one is not justified in treating these forthwith in the same way and in assuming that they go over into a convergent series.

Taylor really operates here with infinitely small quantities (differentials) in the same unquestioning way as the Leibnizians. It is interesting to reflect that although,

[159] Londini, 1715, p. 21–23.

as a young man of twenty-nine, he was under the eye of Newton, he departed from the latter's method of limits. Yet, he succeeded thus in achieving this discovery.

You will find an excellent critical presentation of the entire development of Taylor's theorem in Alfred Pringsheim's memoir: *Zur Geschichte des Taylorschen Lehrsatzes*[160]. Furthermore, I should like to speak here about the customary distinction between Taylor's series and that of Maclaurin. As is well known, many textbooks make a point of putting $a = 0$ and of calling the obvious special case of Taylor's series which thus arises:

$$f(x) = f(0) + \frac{x}{1!} f'(0) + \frac{x^2}{2!} f''(0) + \cdots$$

the series of Maclaurin; and many persons may think that this distinction is important. Anybody who understands the situation, however, sees that it is comparatively unimportant mathematically. But it is not so well known that, considered historically, it is pure nonsense. For Taylor had undoubted priority with his general theorem, deduced in the way indicated above. More than this, he emphasises at a later place in his book (p. 27) the special form of the series for $a = 0$ and remarks that it could be derived directly by the method which is called today that of undetermined coefficients. Furthermore, Maclaurin took over[161] this deduction in 1742 in his *Treatise of Fluxions* (which we mentioned on p. [229]) where he quoted Taylor expressly and made no claim whatever of offering anything new. But the quotation seems to have been disregarded and the author of the book seems to have been looked upon as the discoverer of the theorem. Errors of this sort are common. It was only later that people went back to Taylor and named the general theorem, at least, after him. It is difficult, if not impossible, to overcome such deep-rooted
**[253]** absurdities. At best, one can only spread the truth in the small circle of those who have historical interests.

I shall now supplement our discussion of infinitesimal calculus with some remarks of a general nature.

## 3. Historical and Pedagogical Considerations

### *Remarks About Textbooks for the Infinitesimal Calculus*

I should like to mention, first of all, that the bond which Taylor established between difference calculus and differential calculus held for a long time. These two branches always went hand in hand, still in the analytic developments of Euler, and the formulas of differential calculus appeared as limiting cases of elementary relations that occur in the difference calculus. This natural connection was first

160 Bibliotheca Mathematica, 3rd series, vol. I (1900), p. 433–479.
161 Edinburgh, 1742, vol.11, p. 610.

broken by the often mentioned formal definitions of Lagrange's derivative calculus. I should like to show you a compilation from the end of the eighteenth century which, closely following Lagrange, brings together all the facts then known about infinitesimal calculus, namely the *Traité du Calcul Différentiel et du Calcul Intégral* of Sylvestre-François Lacroix[162]. As a characteristic sample from this work, consider the definition of the derivative (vol. I, p. 145): A function $f(x)$ is defined by means of a power series. By using the binomial theorem (and rearranging the terms) one has

$$f(x+h) = f(x) + hf'(x) + \frac{1}{2}h^2 f''(x) + \cdots .$$

Lacroix now denotes the term of this series, which is linear in $h$ by $d\,f(x)$, and, writing $dx$ for $h$ itself, he has for the derivative, which he calls *differential coefficient*

$$\frac{df(x)}{dx} = f'(x).$$

Thus this formula is deduced in a manner thoroughly superficial even if unassailable. Within the range of these thoughts, Lacroix could no longer, of course, use the calculus of differences as a starting point. However, since this branch seemed to him too important in practice to be omitted, he adopted the expedient of developing it independently, which he did very thoroughly in a third volume, but without any connecting bridge between it and differential calculus.

This "large Lacroix" is historically significant as the *proper source of the many textbooks of infinitesimal calculus which appeared in the nineteenth century*. In the **[254]** first rank of these I should mention his own textbook, the "small Lacroix"[163].

Since the twenties of the last century the textbooks have been strongly influenced also by the method of limits which Cauchy raised to such an honourable place. Here we should first think of the many French textbooks, most of which, as *Cours d'Analyse de l'Ecole Polytechnique*, were prepared expressly for higher education. Directly or indirectly, German textbooks also have depended on them, with the single exception, perhaps, of the one by Oscar Schlömilch. From the long list of books, I shall single out only Joseph Alfred Serret's *Cours de Calcul Différentiel et Intégral*, which appeared first in 1869 in Paris. It was translated into German in 1884 by Axel Harnack and has been since then one of our most widely used textbooks. Due to a succession of various revisers, it suffered of some incoherent parts. The editions,[164] which have appeared since 1906, however, have been subjected to a thoroughgoing revision by Georg Scheffers of Charlottenburg, the result being an again homogeneous work. I am glad to mention also an entirely new

[162] Three volumes, Paris, 1797 – 1800, with many later editions.

[163] *Traitié Elémentaire du Calcul Différentiel et Intégral.*, Paris, 1802.

[164] Since 1906: Joseph Alfred Serret, & Georg Scheffers, *Lehrbuch der Differential- und Integralrechnung,* vol. I, sixth edition. Leipzig 1915; vol. II, 6 – 7 edition; vol. III, fifth edition, 1914.

French book, the *Cours d′ Analyse Mathématique* by Édouard Goursat[165] in three volumes, which is fuller in many ways than Serret and contains, in particular, a long series of entirely modern developments. Furthermore it is a very readable book.

In all these recent books, the derivative and the integral are based entirely upon the *concept of limit. There is never any question as to difference calculus or interpolation.* One sees the things in a clearer light, perhaps, in this way, but, on the other hand, the field of view is considerably narrowed – as it is when we use a microscope. Difference calculus is now left entirely to the practical calculators, who are obliged to use it, especially the astronomers; and the mathematician hears
**[255]** nothing of it. We may hope that the future will bring a change[166] here.

## *Characterising Our Proper Presentation*

As a conclusion of my discussion of infinitesimal calculus I should like to bring up again for emphasis *four points, in which my exposition differs especially from the customary presentation in the textbooks*:

1. *Visualisation of abstract considerations by means of figures* (curves of approximation, in the case of Fourier's and Taylor's series).

2. *Emphasis upon its relation to neighbouring fields*, such as calculus of differences and of interpolation, and finally to philosophical investigations.

3. *Emphasis upon historical growth.*

4. *Exhibition of samples of popular literature to mark the deviation of the thus induced view points in the public at large from those of the professional mathematician.*

It seems to me extremely important that precisely the prospective teacher should take account of all of these. As soon as you begin teaching you will be confronted with the popular views. If you lack orientation, if you are not well informed con-

[165] Paris 1902 – 1907, vol. I, third edition. 1917; vol. II, third edition. 1918; vol. III, second edition. 1915. (Translated into English: vol. I by Earle Raymond Hedrick, 1904, Ginn and Co.; vol. II by Earle Raymond Hedrick and Otto Dunkel, 1916, Ginn and Co.). [Of the most recent German calculus textbooks should be mentioned: I) "Vorlesungen über Differential- und Integralrechnung" by *Richard Courant*, in two volumes (second edition 1930/31). 2) "Einführung in die mathematische Behandlung naturwissenschaftlicher Fragen" by *Alwin Watther* (1928). Both textbooks are following Felix Klein's pedagogical conceptions]

[166] [In order to make a beginning here, Klein had then induced Friesendorff and Prüm to translate Markoff's *Differenzenrechnung* into German (Leipzig, 1896). There is a series of articles in the Enzyklopädie. A work on *Differenzenrechnung* by Niels Erik Nörlund has just appeared (Berlin, Julius Springer, 1924) which exhibits the subject in new light. Alwin Walther, who had cooperated in Nörlund's textbook, gave a lecture, at the summer vacation course on mathematics and physics in Göttingen in 1926 on issues of the calculus of finite differences, which are important for teaching the calculus. Unfortunately, this lecture is not yet published. A second lecture by *Walther* at the same course, dealt with „Begriff und Anwendungen des Differentials". It was published as *Beiheft* 14 (Berlin 1929, B. G. Teubner) of the *Zeitschrift für den mathematischen und naturwissenschaftlichen Unterricht*. The explanations given in this lecture complement in a precious manner what is given here on the pp. [223] to [255].]

cerning the intuitive elements of mathematics as well as the vital relations with neighbouring fields, if, above all, you do not know the historical development, your footing will be very insecure. You will then either withdraw to the ground of the most modern pure mathematics, and fail to be understood in the school, or you will succumb to the assault, give up what you learned in higher education, and even in your teaching allow yourself to be buried in the traditional routine. The discontinuity between school and university, of which I have often spoken, is greatest precisely in the field of infinitesimal calculus. I hope that my words may contribute to its removal and that they may provide you with useful armour in your teaching.

This brings me to the end of the conventional analysis. By way of supplement I shall discuss a few *theories of modern mathematics* to which I have referred occasionally and with which I think the teacher should have some acquaintance. **[256]**