HOCHSCHULBÜCHER FÜR MATHEMATIK

HERAUSGEGEBEN VON
H. GRELL, K. MARUHN UND W. RINOW

BAND 9

ENZYKLOPÄDIE DER ELEMENTARMATHEMATIK

REDAKTION:

P. S. ALEXANDROFF
A. I. MARKUSCHEWITSCH
A. J. CHINTSCHIN

BAND III

ANALYSIS

VEB DEUTSCHER VERLAG DER WISSENSCHAFTEN
BERLIN 1958

Академия педагогических наук РСФСР

Энциклопедия элементарной математики

под редакцией

П. С. Александрова, А. И. Маркушевича и А. Я. Хинчина

книга третья ·
Функции и пределы (Основы анализа)

Государственное издательство
технико-теоретической литературы
Москва 1952 Ленинград

Übersetzung: Helmut Limberg, Karl-Heinz Rupp und Gerhard Tesch
Endgültige Abfassung des deutschen Textes und wissenschaftliche Redaktion:
Dr. Günter Asser

Herausgeber und Verlag danken Herrn Pr. Dr. I. P. Natanson
für die Durchsicht der Korrekturen seines Artikels

Verantwortlicher Verlagsredakteur: Ludwig Boll

Alle Rechte vorbehalten
Copyright 1958 by VEB Deutscher Verlag der Wissenschaften, Berlin
Printed in Germany
Lizenz-Nr. 206 · 435/14/57
Herstellung: (III/18/203) VEB Leipziger Druckhaus, Leipzig

VORWORT

Dieser dritte Band der „Enzyklopädie der Elementarmathematik" schließt den ersten großen Teilabschnitt dieses Werkes ab, der diejenigen Anfangsgründe der Mathematik systematisch behandelt, auf denen der Schulunterricht in Arithmetik, Algebra und Trigonometrie aufbaut. Während die ersten beiden Bände sich vorzugsweise auf Fragen der Arithmetik und der Algebra im engeren Sinne des Wortes, d. h. auf die Lehre von den Zahlen, ihren Verallgemeinerungen und den Rechenoperationen — genauer, den algebraischen Rechenoperationen Addition, Subtraktion, Multiplikation, Division — und auf die Theorie der algebraischen Gleichungen beschränkten, werden in diesem dritten Band Probleme der Analysis, nämlich Funktionen und Grenzwerte, dargelegt.

Neben der Lehre von den elementaren Funktionen und einer sehr ausführlichen Behandlung der Theorie der Grenzwerte wurden auch die elementarsten Tatsachen aus der Differential- und Integralrechnung, der Reihenlehre und eine Einführung in die komplexe Funktionentheorie aufgenommen.

Die Begriffe „Ableitung" und „Integral" verlangen nach Aufnahme in den Stoff der allgemeinbildenden Schule. Wie man auch zu ihrer faktischen Aufnahme in die Lehrpläne stehen mag, eine irgendwie befriedigende abgeschlossene Darlegung der elementaren Grundlagen der Mathematik ohne diese Grundbegriffe muß beim heutigen Stand der Wissenschaft als unmöglich angesehen werden.

Was die komplexe Funktionentheorie betrifft, so besteht weder die Möglichkeit noch die Notwendigkeit, in nächster Zukunft eine systematische Einführung in den Lehrplan aufzunehmen. Jedoch rechtfertigt die Tatsache, daß die elementaren Funktionen analytische Funktionen sind, die (mit eventueller Ausnahme einzelner Punkte) in der ganzen komplexen Ebene definiert sind, und daß daher ein volles Verständnis der Eigenschaften dieser Funktionen und der Zusammenhänge zwischen ihnen nur erreicht werden kann, wenn man sie als Funktionen einer komplexen Veränderlichen betrachtet, voll und ganz die Aufnahme eines kurzen Abrisses über analytische Funktionen einer komplexen Veränderlichen in diesen Band.

Die Redaktion

INHALTSVERZEICHNIS

Vorwort V

W. L. Gontscharow

Elementare Funktionen einer reellen Veränderlichen. Grenzwerte von Folgen und Funktionen. Der allgemeine Funktionsbegriff

Kapitel I. Allgemeine Betrachtungen über elementare Funktionen und die graphische Darstellung von Gleichungen

§ 1. Elementare Funktionen 1
§ 2. Graphische Darstellungen. Verfahren zur punktweisen Konstruktion . . 8
§ 3. Einfachste Transformationen von graphischen Darstellungen 17
§ 4. Funktion und Umkehrfunktion 24
§ 5. Elementare Diskussion von Funktionen (Fragestellung und einige allgemeine Verfahren) 27

Kapitel II. Überblick über die elementaren Funktionen und ihre graphischen Darstellungen

§ 6. Klassifikation der rationalen Funktionen 33
§ 7. Die Potenzfunktionen mit positivem ganzen Exponenten 34
§ 8. Polynome ersten Grades (lineare Funktionen) 37
§ 9. Polynome zweiten Grades 39
§ 10. Polynome dritten Grades 40
§ 11. Biquadratische Polynome 43
§ 12. Polynome höheren Grades 44
§ 13. Die Potenzfunktionen mit negativem ganzen Exponenten 46
§ 14. Gebrochene lineare Funktionen 49
§ 15. Gebrochene rationale Funktionen zweiten Grades 50
§ 16. Gebrochene rationale Funktionen (allgemeiner Fall) 55
§ 17. Irrationale algebraische Funktionen 58
§ 18. Einige Beispiele für die Diskussion von algebraischen Funktionen . . 59
§ 19. Elementare transzendente Funktionen 69
§ 20. Die Exponentialfunktionen 70
§ 21. Die hyperbolischen Funktionen 75
§ 22. Die logarithmischen Funktionen 79
§ 23. Die Umkehrfunktionen der hyperbolischen Funktionen 81
§ 24. Potenzen mit beliebigem reellen Exponenten 84
§ 25. Die trigonometrischen Grundfunktionen Sinus und Kosinus 86

§ 26. Einfache harmonische Schwingungen 93
§ 27. Trigonometrische Polynome 96
§ 28. Tschebyschewsche Polynome 98
§ 29. Der Tangens und andere gebrochene trigonometrische Funktionen . . 102
§ 30. Über die Darstellbarkeit von Funktionen, die rational von den trigonometrischen Funktionen abhängen, durch eine oder zwei trigonometrische Funktionen 104
§ 31. Beispiele für die Diskussion von rationalen Funktionen der trigonometrischen Funktionen. Trigonometrische Gleichungen 112
§ 32. Die Umkehrfunktionen der trigonometrischen Funktionen 119
§ 33. Diskussion der Tschebyschewschen Polynome. Ihre Minimaleigenschaft 125

Kapitel III. Grenzwerte von Zahlenfolgen und von Funktionen

§ 34. Endliche und unendliche Zahlenfolgen 131
§ 35. Die allgemeine Definition von unendlichen Zahlenfolgen 139
§ 36. Der Satz von Bolzano und Weierstrass über die Existenz eines Häufungspunktes 144
§ 37. Beispiele. Der Grenzwert als einziger Häufungspunkt 150
§ 38. Die klassische Definition des Grenzwerts einer Folge und seine wesentlichsten Eigenschaften 155
§ 39. Verallgemeinerung des Limesbegriffs („uneigentliche Grenzwerte") . . 162
§ 40. Der Grenzwert einer Funktion im Unendlichen 165
§ 41. Einseitiger Grenzwert einer Funktion in einem endlichen Punkte . . . 169
§ 42. Der zweiseitige Grenzwert und der Begriff der Stetigkeit 176
§ 43. Beispiele für stetige Funktionen 179
§ 44. Grenzwerte bei monotoner Änderung. Die Zahl e 184

Kapitel IV. Grenzwerte von Funktionenfolgen. Eigenschaften stetiger Funktionen

§ 45. Einfache Konvergenz von Funktionenfolgen 191
§ 46. Der allgemeine Begriff der Funktion einer reellen Veränderlichen . . 200
§ 47. Eigenschaften stetiger Funktionen 204
§ 48. Gleichmäßige Konvergenz von Folgen stetiger Funktionen 211
§ 49. Der Satz von Weierstrass-Bernstein über die Approximation stetiger Funktionen mit Hilfe von rationalen Polynomen 215
§ 50. Beweis des Satzes von Bernstein. 220
§ 51. Endgültige Definition der Exponentialfunktionen. Fortsetzung einer stetigen Funktion auf die Randpunkte einer überall dichten Menge . . 225
§ 52. Der Satz von Bolzano und das Problem der Existenz einer eindeutigen Umkehrfunktion 231
§ 53. Funktionalgleichungen und elementare Funktionen 234

Kapitel V. Der allgemeine Funktionsbegriff

§ 54. Zuordnungen zwischen Mengen 241
§ 55. Geometrische Gebilde in mehrdimensionalen Räumen 243
§ 56. Räumliche Abbildungen 246
§ 57. Metrische Räume 250
§ 58. Der Begriff des Grenzwertes in metrischen Räumen 255
§ 59. Topologische Räume 258

§ 60. Mengenalgebra. Die Ableitung einer Menge. Abgeschlossenheit und Zusammenhang . . . 260
§ 61. Stetige Abbildungen und ihre Eigenschaften . . . 264
§ 62. Homöomorphe Abbildungen . . . 268
§ 63. Obere und untere Grenze, Limes superior und Limes inferior von Zahlenmengen und Zahlenfolgen . . . 272

I. P. Natanson

Ableitungen, Integrale, Reihen

Einleitung . . . 283

Kapitel I. Ableitungen

§ 1. Ableitung und Differential . . . 287
§ 2. Wichtige Sätze über Ableitungen . . . 322
§ 3. Anwendung der Differentialrechnung bei der Untersuchung der Eigenschaften von Funktionen . . . 337

Kapitel II. Integrale

§ 4. Das unbestimmte Integral . . . 349
§ 5. Das bestimmte Integral . . . 360
§ 6. Anwendungen der Integralrechnung . . . 390

Kapitel III. Reihen

§ 7. Reihen mit konstanten Gliedern . . . 406
§ 8. Potenzreihen . . . 428

W. L. Gontscharow

Elementare Funktionen einer komplexen Veränderlichen

§ 1. Rationale Funktionen . . . 473
§ 2. Grenzwerte. Reihen . . . 476
§ 3. Die Exponentialfunktion, der Sinus und der Kosinus . . . 480
§ 4. Darstellung der trigonometrischen Funktionen mit Hilfe der Exponentialfunktion . . . 483
§ 5. Die hyperbolischen und die trigonometrischen Funktionen . . . 486
§ 6. Der Logarithmus . . . 488
§ 7. Potenzen mit beliebigem Exponenten . . . 489
§ 8. Die Umkehrfunktionen der trigonometrischen und der hyperbolischen Funktionen . . . 491
§ 9. Differentiation im Komplexen . . . 493
§ 10. Das Kurvenintegral . . . 497
§ 11. Die Approximation von Funktionen durch Polynome . . . 502
§ 12. Die Stammfunktion . . . 505
§ 13. Die Cauchysche Integralformel . . . 511
§ 14. Der Begriff der analytischen Funktion . . . 514
§ 15. Eigenschaften der analytischen Funktionen . . . 518
§ 16. Die geometrische Bedeutung der analytischen Funktionen . . . 523
§ 17. Beispiele für konforme Abbildungen . . . 526

Namenverzeichnis . . . 531
Sachverzeichnis . . . 532

W. L. GONTSCHAROW

ELEMENTARE FUNKTIONEN EINER REELLEN VERÄNDERLICHEN.

GRENZWERTE VON FOLGEN UND FUNKTIONEN.

DER ALLGEMEINE FUNKTIONSBEGRIFF

Kapitel I

ALLGEMEINE BETRACHTUNGEN ÜBER ELEMENTARE FUNKTIONEN UND DIE GRAPHISCHE DARSTELLUNG VON GLEICHUNGEN

§ 1. Elementare Funktionen

Die Zahlen und die über ihnen ausführbaren Rechenoperationen bilden den nächstliegenden Gegenstand mathematischer Untersuchungen. Sowohl der Begriff Zahl als auch der Begriff Operation lassen sich erheblich erweitern und verallgemeinern. Im vorliegenden Artikel werden, sofern keine besonderen Einschränkungen gemacht werden, unter Zahlen stets reelle Zahlen verstanden. An Operationen werden (jedenfalls in den ersten drei Kapiteln) in erster Linie diejenigen Operationen untersucht, die in der Elementarmathematik auftreten und daher üblicherweise *elementar* genannt werden. Hierher gehören vor allem die algebraischen Operationen: Addition, Subtraktion, Multiplikation, Division, ferner das Potenzieren, das Ausziehen von Wurzeln beliebigen Grades und das Logarithmieren in bezug auf eine beliebige positive Basis; schließlich betrachtet man in der Elementarmathematik noch die trigonometrischen Funktionen (Sinus, Cosinus, Tangens, Cotangens, Sekans und Cosekans) und die zugehörigen Umkehrfunktionen (Arcussinus, Arcuscosinus usw.).

Diese Operationen kann man nun, je nach den Umständen, in irgendeiner Reihenfolge auf vorgegebene Zahlen oder auf Veränderliche anwenden. Wir setzen in den Kapiteln I und II voraus, daß die Anzahl der auszuführenden Operationen stets *endlich* ist. Das Resultat einer Operation bezeichnet man mit einem neuen Buchstaben; z. B. schreibt man

$$v = u^{\operatorname{tg} u}. \tag{1}$$

Mitunter ist es zweckmäßig, diesem Buchstaben die in die Rechnung eingehenden Veränderlichen in einer bestimmten Reihenfolge und durch Kommata getrennt, in Klammern beizufügen, also z. B.

$$f(t) = \frac{2t}{t^2+1} \tag{2}$$

oder

$$F(x, y) = x^2 + y^2 - 25 \tag{3}$$

zu schreiben.

Durch die auf diese Weise erhaltenen mathematischen Ausdrücke (Formeln) wird jedem Zahlenwert der auftretenden (d. h. an der Rechnung beteiligten) Veränderlichen ein bestimmter „Funktionswert" zugeordnet; sie legen elementare Funktionen dieser Veränderlichen fest.

Die in den vorliegenden Formeln auftretenden „Veränderlichen" werden genauer *unabhängige Veränderliche* (Argumente), die Funktionen selbst *abhängige Veränderliche* genannt.

Die Funktionssymbole der Form $f(t)$ oder $F(x, y)$ sind in folgender Hinsicht besonders zweckmäßig: Ungeachtet dessen, was c, a, b bedeuten — Zahlen oder irgendwelche neuen „Buchstabenausdrücke" — bedeutet $f(c)$ bzw. $F(a, b)$ diejenige Zahl oder denjenigen Ausdruck, den man erhält, wenn man c an Stelle von t bzw. a an Stelle von x und gleichzeitig b an Stelle von y setzt. In diesem Sinne folgt aus (2) und (3)

$$f(3)=\frac{2\cdot 3}{3^2+1}=\frac{3}{5}, \qquad f(1)=\frac{2\cdot 1}{1^2+1}=1;$$

$$F(6,\,7)=6^2+7^2-25=60, \quad F(4,\,3)=4^2+3^2-25=0$$

und entsprechend

$$f\left(\frac{m}{n}\right)=\frac{2\cdot\frac{m}{n}}{\left(\frac{m}{n}\right)^2+1}=\frac{2mn}{m^2+n^2}, \quad f(-t)=\frac{2(-t)}{(-t)^2+1}=-\frac{2t}{t^2+1};$$

$$F(px,\,qy)=(px)^2+(qy)^2-25=p^2x^2+q^2y^2-25.$$

Werden die unabhängigen Veränderlichen nicht in Klammern beigefügt, so benötigt man zur Festlegung von bestimmten Werten für die unabhängigen Veränderlichen eine kompliziertere Schreibweise; z. B. kann man im Fall (1) $[v]_{u=\pi}=\pi^{\operatorname{tg}\pi}=1$ schreiben, was man etwa durch die folgende Redewendung wiedergeben kann: „Für $u=\pi$ nimmt die Funktion $v=u^{\operatorname{tg}u}$ den Wert 1 an."

Dem Leser ist sicherlich die Definition einer Funktion (einer Veränderlichen) als Zuordnung zwischen den Zahlenwerten der unabhängigen Veränderlichen und denen der abhängigen Veränderlichen bekannt.[1]) Nach dieser Definition ist *eine veränderliche Größe y eine Funktion der* (z. B. in einem gewissen Intervall *I*) *veränderlichen Größe x, wenn jedem Wert x aus I ein eindeutig bestimmter Wert von y entspricht.* In diesem allgemeinen Sinn findet sich der Funktionsbegriff bereits im Jahre 1834 bei N. I. Lobatschewski, der ihn mit folgenden Worten definierte: „Dieser allgemeine Begriff (einer Funktion — *die Red.*) verlangt, daß eine Zahl, die für jeden Wert von x gegeben ist und die sich zusammen mit x ändert, eine Funktion von x genannt wird. Der Wert der Funktion kann durch einen analytischen Ausdruck oder durch irgendeine andere Bedingung gegeben sein, nach der aus der Gesamtheit der Zahlen eine ausgewählt werden kann; schließlich kann sie in Form irgendeiner Abhängigkeit bestehen, die an sich unbekannt bleibt."[2])

Die Worte „es entspricht" (oder „es ist zugeordnet" — wie man auch häufig sagt) lassen die Frage, welcher Natur die Vorschrift ist, durch die die Zuordnung getroffen wird, vollkommen offen; wesentlich ist nur, daß es eine irgendwie geartete Vorschrift gibt. Zum Beispiel kann die Zuordnungsvorschrift rein „empirischen" Charakter tragen. Das ist etwa dann der Fall, wenn man die Temperatur als Funktion der Zeit betrachtet (hier besteht die Zuordnungsvorschrift darin, den in bestimmten Zeitabständen abgelesenen Thermometerstand festzulegen).

Bei den in der Elementarmathematik auftretenden Funktionen hat die Zuordnungsvorschrift fast ausschließlich „operativen" oder „analytischen" Charakter: Es wird

[1]) Vgl. I. W. Proskurjakow, Mengen, Gruppen, Ringe und Körper. Die theoretischen Grundlagen der Arithmetik. EdEM, Band 1, Seite 187 (deutsche Ausgabe) — *Anm. d. wissenschaftl. Red.*

[2]) N. I. Lobatschewski, Gesammelte Werke, Band V, Seite 43, Gostechisdat 1951.

angegeben, in welcher Reihenfolge man welche mathematischen Operationen auf die unabhängigen Veränderlichen anzuwenden hat, um den zugehörigen Funktionswert zu erhalten.

Es ist kein Grund dafür vorhanden, dem Begriff der durch einen eindeutigen analytischen Ausdruck gegebenen Funktion den Begriff der Funktion als Zuordnung gegenüberzustellen: Der erste Begriff ist ein Spezialfall des zweiten.[1])

Der Begriff der Funktion als analytischer Ausdruck wurde in der ersten Hälfte des 18. Jahrhunderts geprägt, und zwar von JOHANN BERNOULLI (1718) und L. EULER (1748). Bei EULER findet sich z. B. die folgende Definition: „Eine Funktion einer Veränderlichen ist ein analytischer Ausdruck, der irgendwie aus jener Veränderlichen und Konstanten zusammengesetzt ist."[2]) Es muß jedoch gesagt werden, daß bei EULER 1. die „zugelassenen" Operationen nicht klar genug abgegrenzt sind; 2. keine Ausdrücke ausgeschlossen sind, die unendlich viele Operationen enthalten.

Die genaue Definition der elementaren Funktion (im modernen Sinn) gründet sich auf den Begriff der Funktion als Zuordnung und kann folgendermaßen formuliert werden:

Eine Funktion heißt elementar, wenn man ihre Werte aus konstanten Zahlen und den jeweiligen Werten für die unabhängigen Veränderlichen durch endlich viele elementare Operationen erhalten kann.

Konkrete Beispiele für nichtelementare Funktionen werden in Kapitel IV gegeben. Zu diesem Zweck wird dort ein ganz natürliches Verfahren entwickelt (vgl. § 49).

Im vorliegenden Abschnitt werden wir mehrfach den Funktionsbegriff in voller Allgemeinheit (als Zuordnung) anzuwenden haben. Zunächst kann der Leser jedoch, sofern von Funktionen die Rede ist, darunter diejenigen elementaren Funktionen verstehen, die ihm von der Schule her bekannt sind.

Im folgenden (in den Kapiteln I bis IV) werden wir uns auf Funktionen in höchstens zwei Veränderlichen, die wir stets mit x und y bezeichnen wollen, beschränken.

Gegeben sei eine Gleichung der Form

$$F(x, y) = 0, \tag{4}$$

wobei $F(x, y)$ eine beliebige elementare Funktion in x und y ist.[3]) Es seien ferner x_0 und y_0 beliebige Zahlen. Sind die Zahlen x_0 und y_0 so gewählt, daß die Gleichung (4) erfüllt ist, wenn man x_0 für x und y_0 für y einsetzt[4]), also

$$F(x_0, y_0) = 0$$

---

[1]) Diese beiden Funktionsbegriffe (den engeren und den weiteren) kann man durch die folgenden beiden Verfahren einander annähern bzw. sogar gleichmachen: a) Man stellt fest, daß sehr umfangreiche Klassen von Zuordnungen eine analytische Darstellung besitzen (vgl. z. B. den Satz von WEIERSTRASS in § 49); b) man betrachtet als weitere mathematische Operation den Übergang von einem Zahlwert der unabhängigen Veränderlichen zu dem (auf Grund einer gegebenen Zuordnung) entsprechenden Funktionswert.

[2]) L. EULER „Introductio in Analysin infinitorum", Bd. I, Kap. I. — *Anm. d. wissenschaftl. Red.*

[3]) Sofern in einer Gleichung auf der rechten Seite noch nicht die Zahl Null steht, kann man den dort stehenden Ausdruck auf die andere Seite bringen.

[4]) Wir setzen hierbei stillschweigend voraus, daß diese Substitution einen Sinn hat, d. h., daß die mit $F(x, y)$ bezeichnete Zusammenfassung von Operationen für die Werte $x = x_0$ und $y = y_0$ ausgeführt werden kann.

gilt, so heißt das Paar (x_0, y_0) eine *Lösung* (eine der Lösungen) der gegebenen Gleichung. Die Funktion $F(x, y)$ kann nun so beschaffen sein, daß die Gleichung $F(x, y) = 0$ überhaupt keine Lösung besitzt (wie dies z. B. für $F(x, y) = \frac{1}{x+y}$ und $F(x, y) = 2^{x-y}$ der Fall ist). Ebenso kann der Fall eintreten, daß die Gleichung (4) nur eine oder im ganzen nur endlich viele Lösungen besitzt (z. B. besitzt die Gleichung $x^2 + y^2 = 0$ als einzige Lösung $x_0 = 0$, $y_0 = 0$). Schließlich ist auch das entgegengesetzte Extrem nicht ausgeschlossen, daß nämlich $F(x, y)$ identisch verschwindet, daß also jedes Wertepaar (x, y) die Gleichung (4) erfüllt.

Besonders wichtig ist der häufig anzutreffende Fall, daß die Gleichung $F(x, y) = 0$ unendlich viele Lösungen besitzt, und zwar derart, daß man zu jedem „willkürlich gewählten" Wert der einen Veränderlichen wenigstens einen Wert für die andere Veränderliche finden kann, so daß die Gleichung befriedigt wird. Man sagt dann, die Funktion $F(x, y)$ stelle eine *funktionale Abhängigkeit* zwischen den Veränderlichen x und y her.

Betrachten wir z. B. die Funktion

$$F(x, y) = 2x - 5y + 10. \tag{5}$$

Offenbar kann man zu jedem Wert für x genau einen Wert für y bestimmen, für den

$$2x - 5y + 10 = 0 \tag{6}$$

gilt, und zwar leistet

$$y = \frac{2x + 10}{5} \tag{7}$$

das Verlangte. Durch (6) wird also jedem Wert von x ein eindeutig bestimmter Wert von y zugeordnet, der sich auf Grund von (7) ergibt. Setzt man also

$$f(x) = \frac{2x + 10}{5},$$

so ist die Gleichung $F(x, y) = 0$ der Gleichung $y = f(x)$ gleichwertig.

Als zweites Beispiel betrachten wir die Funktion $F(x, y) = x^2 + y^2 - 1$. Die Gleichung

$$x^2 + y^2 - 1 = 0 \tag{8}$$

besitzt ganz andere Eigenschaften als die Gleichung (6). Hier kann man als Werte für x nur Zahlen aus dem Intervall $-1 \leqq x \leqq +1$ willkürlich wählen, erhält dann aber jeweils zwei verschiedene Werte für y, nämlich

$$y = \pm\sqrt{1 - x^2}.$$

Setzt man

$$f_1(x) = +\sqrt{1 - x^2}, \quad f_2(x) = -\sqrt{1 - x^2},$$

so sieht man leicht, daß jede der Gleichungen

$$y = f_1(x), \quad y = f_2(x) \tag{9}$$

die Gleichung (8) zur Folge hat; umgekehrt ist jede Lösung von (8) Lösung einer der Gleichungen (9).[1]) Im zweiten Beispiel können die Veränderlichen x und y ohne weiteres vertauscht werden.

Wir kommen nun von den speziellen Beispielen wieder zum allgemeinen Fall zurück. Wir haben gesehen, daß man unter Umständen eine funktionale Abhängigkeit zwischen x und y durch eine Gleichung ersetzen kann, welche x und y miteinander verknüpft, wobei es häufig noch gleichgültig ist, welche der Veränderlichen x, y man vorteilhafter als unabhängige bzw. abhängige Veränderliche nimmt.

Nehmen wir etwa x als unabhängige Veränderliche. Wenn die Gleichung $F(x, y) = 0$ für jeden Wert von x aus einem gewissen Intervall (z. B. dem Intervall $a < x < b$) stets nur eine Lösung y besitzt, die sich elementar durch x ausdrücken läßt, so ist y im gegebenen Intervall eine eindeutige Funktion von x:

$$y = f(x).$$

In diesem Fall ist (im betrachteten Intervall) die Gleichung $y = f(x)$ der gegebenen Gleichung $F(x, y) = 0$ gleichwertig. Entsprechen dagegen jedem Wert der Veränderlichen x aus einem gewissen Intervall mehrere (etwa k) Lösungen y der gegebenen Gleichung, von denen sich jede durch eine besondere Formel

$$y = f_1(x), \quad y = f_2(x), \ \ldots, \ y = f_k(x)$$

elementar ausdrücken läßt, so erhalten wir k verschiedene eindeutige Funktionen oder, wie man manchmal auch sagt, eine *mehrdeutige* (k-*deutige*) *Funktion*. In diesem Fall läßt sich die gegebene Gleichung $F(x, y) = 0$ in k Gleichungen „aufspalten".

Im oben angeführten Beispiel (6) ist y eine eindeutige Funktion von x für beliebige Werte von x ($-\infty < x < +\infty$); im Beispiel (8) ist y eine zweideutige Funktion von x im Intervall $-1 \leqq x \leqq +1$.

Zur Erläuterung der hierbei bestehenden Vielzahl von Möglichkeiten betrachten wir noch ein weiteres Beispiel[2]), und zwar

$$F(x, y) = x^4 + y^4 - x^2 - y^2. \tag{10}$$

Die Gleichung

$$x^4 + y^4 = x^2 + y^2 \tag{11}$$

(die der Gleichung $x^4 + y^4 - x^2 - y^2 = 0$ gleichwertig ist) ist eine biquadratische Gleichung in y. Ihre Lösungen sind

$$y = \pm \sqrt{\frac{1}{2} \pm \sqrt{\frac{1}{4} + x^2 - x^4}}.$$

[1]) Auf die Gleichung (8) kommen wir in § 46 ausführlicher zurück.
[2]) Vgl. Seite 68.

Man prüft leicht nach, daß für

$$1<|x|<\sqrt{\frac{1}{2}+\sqrt{\frac{1}{2}}},$$

d. h. in den Intervallen

$$-\sqrt{\frac{1}{2}+\sqrt{\frac{1}{2}}}<x<-1 \quad \text{und} \quad 1<x<\sqrt{\frac{1}{2}+\sqrt{\frac{1}{2}}},$$

der innere Radikand stets positiv und das zugehörige Radikal kleiner als $\frac{1}{2}$ ist. Daher ist in den betrachteten Intervallen unabhängig von der Wahl des Vorzeichens für das innere Radikal der äußere Radikand stets positiv. Die Gleichung (11) führt also in diesen Intervallen auf vier eindeutige elementare Funktionen

$$y=f_i(x) \qquad (i=1,2,3,4),$$

wobei

$$f_1(x)=-\sqrt{\frac{1}{2}+\sqrt{\frac{1}{4}+x^2-x^4}},$$

$$f_2(x)=-\sqrt{\frac{1}{2}-\sqrt{\frac{1}{4}+x^2-x^4}},$$

$$f_3(x)=+\sqrt{\frac{1}{2}-\sqrt{\frac{1}{4}+x^2-x^4}},$$

$$f_4(x)=+\sqrt{\frac{1}{2}+\sqrt{\frac{1}{4}+x^2-x^4}}.$$

Im Intervall $-1<x<+1$ ist der innere Radikand positiv und das innere Radikal größer als $\frac{1}{2}$. Daher erhalten wir in diesem Intervall nur die Funktionen

$$y=f_1(x) \quad \text{und} \quad y=f_4(x).$$

Für

$$|x|>\sqrt{\frac{1}{2}+\sqrt{\frac{1}{2}}},$$

d. h. in den Intervallen

$$-\infty<x<-\sqrt{\frac{1}{2}+\sqrt{\frac{1}{2}}} \quad \text{und} \quad \sqrt{\frac{1}{2}+\sqrt{\frac{1}{2}}}<x<\infty,$$

besitzt die Gleichung (11) überhaupt keine Lösung.

§ 2. Graphische Darstellungen. Verfahren zur punktweisen Konstruktion

Wir haben gesehen, daß eine Gleichung

$$F(x, y)=0 \tag{12}$$

unendlich viele Lösungen besitzen kann. Um einen anschaulichen Überblick über sämtliche Lösungen zu erhalten, bestimmt man in einer x, y-Ebene für

jede Lösung (x, y) den Punkt, der die Abszisse x und die Ordinate y besitzt. Die Menge der hierbei erhaltenen Punkte bildet eine graphische Darstellung der Lösungen der Gleichung. Genauer und einfacher: Die *graphische Darstellung einer Gleichung* ist die Gesamtheit (Menge) aller Punkte[1]) einer Ebene, deren Koordinaten der Gleichung genügen.

Das Aufsuchen aller Lösungen einer gegebenen Gleichung und die Konstruktion ihrer graphischen Darstellung sind im wesentlichen gleichwertige Aufgaben; sprechen wir jedoch statt von der Gesamtheit ihrer Lösungen von ihrer graphischen Darstellung, so verleihen wir nicht nur den uns interessierenden Fragen Anschaulichkeit, sondern vereinfachen auch erheblich die Ausdrucksweise.

Wenn eine vorgelegte Gleichung nach der abhängigen Veränderlichen y „aufgelöst" ist, also y in einem gewissen Intervall $a < x < b$ als eindeutige Funktion von x dargestellt ist,

$$y = f(x),$$

so sprechen wir, ohne einen Unterschied zu machen, von einer *graphischen Darstellung der Gleichung* $y = f(x)$ oder auch von einer *graphischen Darstellung (einem graphischen Bild) der Funktion* $f(x)$.

Die graphische Darstellung hat in diesem Fall die charakteristische Eigenschaft, daß (in den Grenzen des Intervalls) jede zur y-Achse parallele Gerade mit dem Bild der Funktion $f(x)$ genau einen Punkt gemeinsam hat.

Der zuletzt genannte Fall ist sowohl in praktischer als auch in theoretischer Hinsicht von besonderer Wichtigkeit; wir werden ihm daher im folgenden besondere Beachtung schenken.

Um für eine gegebene Gleichung eine graphische Darstellung anzufertigen, müßte man im allgemeinen sämtliche Punkte der graphischen Darstellung in einer gegebenen x, y-Ebene kennzeichnen. Wenn man von Spezialfällen absieht, so ist dies einfach deshalb unmöglich, weil diese Punkte eine unendliche Menge bilden. Daher beschränkt man sich gewöhnlich auf eine Näherungskonstruktion, die man folgendermaßen ausführt: Man markiert in einer Ebene eine genügende Anzahl von Punkten der graphischen Darstellung, wobei nach Möglichkeit die einzelnen Punkte einen hinreichend kleinen Abstand voneinander besitzen; sodann verbindet man aufeinanderfolgende Punkte durch „passende Kurvenstücke"; auf diese Weise ergibt sich im allgemeinen eine angenäherte graphische Darstellung der gegebenen Gleichung. Man nennt diese Methode das *Verfahren der punktweisen Konstruktion*. Je nach dem Charakter der angestellten Berechnungen und den verwendeten geometrischen Operationen sind die auf der Zeichnung markierten Punkte mehr oder weniger genau bestimmt. Wie genau jedoch die graphische Darstellung „zwischen" den markierten Punkten der Gleichung ist, hängt einerseits von den Eigenschaften der betrachteten Gleichung und zum andern von den Kenntnissen und Erfahrungen des Konstrukteurs ab.

[1]) Man sagt auch: „... der geometrische Ort der Punkte ...".

Ein allgemein bekanntes Verfahren, eine graphische Darstellung einer Funktion zu gewinnen, besteht darin, daß man für eine Folge von Werten für die unabhängige Veränderliche x die zugehörigen Funktionswerte $y = f(x)$ berechnet, die dabei gewonnenen Resultate übersichtlich in einer Tabelle zusammenstellt, danach (oder zugleich mit der Berechnung) die entsprechenden Punkte in einer Ebene markiert und schließlich die erhaltenen Punkte durch eine passende Kurve verbindet.[1])

In Abb. 1 und der zugehörigen Tabelle ist z. B. eine punktweise Konstruktion für die Funktion

$$y = \sqrt{x} \tag{13}$$

im Intervall $0{,}5 \leqq x \leqq 1{,}2$ dargestellt; die Werte von x sind dabei von Zehntel zu Zehntel genommen.

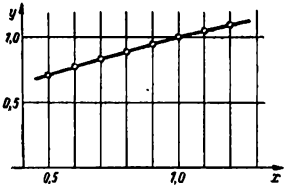


Abb. 1

x	y	x	y
0,5	0,71	0,9	0,95
0,6	0,77	1,0	1,00
0,7	0,84	1,1	1,05
0,8	0,89	1,2	1,10

Abb. 2 zeigt eine punktweise Konstruktion der Funktion

$$y = \frac{x(x+1)}{x+2} \tag{14}$$

im Intervall $0 \leqq x \leqq 1$; die Werte von x sind dabei von Fünftel zu Fünftel genommen.

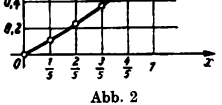


Abb. 2

x	y	x	y
0	0	$\frac{3}{5}$	$\frac{24}{65} \approx 0{,}37$
$\frac{1}{5}$	$\frac{6}{55} \approx 0{,}11$	$\frac{4}{5}$	$\frac{18}{35} \approx 0{,}51$
$\frac{2}{5}$	$\frac{7}{30} \approx 0{,}23$	1	$\frac{2}{3} \approx 0{,}67$

[1]) Häufig verwendet man dabei „äquidistante" Werte von x, d. h. Werte, die eine arithmetische Progression bilden.

Mitunter ist es zweckmäßig, in die Tabelle Spalten für Zwischenergebnisse einzuschieben, wie dies aus folgendem Beispiel ersichtlich ist:

$$y = \frac{1}{2}\sqrt{x^2+1}. \tag{15}$$

x	x^2	x^2+1	$\sqrt{x^2+1}$	$\frac{1}{2}\sqrt{x^2+1}$
0,0	0,00	1,00	1,00	0,50
0,5	0,25	1,25	1,12	0,56
1,0	1,00	2,00	1,41	0,70
1,5	2,25	3,25	1,80	0,90
2,0	4,00	5,00	2,24	1,12
2,5	6,25	7,25	2,69	1,34

Neben dem geschilderten „numerischen" Verfahren spielen auch geometrische Konstruktionen eine große Rolle. Wenn die rechte Seite der Gleichung

$$y = f(x)$$

keine anderen Operationen als die vier Spezies und das Ausziehen der Quadratwurzel enthält, so kann man, sofern man den Wert von x als Strecke vorgibt, den zugehörigen Wert von y als Strecke mit Zirkel und Lineal konstruieren. Sind für verschiedene Werte

$$x = x_1, \quad x = x_2, \quad \ldots, \quad x = x_n$$

die zugehörigen Funktionswerte „zu konstruieren", so empfiehlt es sich, systematisch zu arbeiten und die n Konstruktionen zusammen zu erledigen, indem man jeden Konstruktionsschritt für alle n Konstruktionen gleichzeitig vornimmt.

Betrachten wir z. B. die Funktion (15). Um für einen gegebenen Wert x den zugehörigen Funktionswert y zu konstruieren, nimmt man auf der horizontalen[1]) Achse den Punkt P mit der Abszisse x und auf der vertikalen Achse den Punkt Q mit der Ordinate 1 (Abb. 3); in P errichtet man die Senkrechte und trägt auf ihr von P aus die Strecke PQ ab; der Mittelpunkt M der dabei erhaltenen Strecke PR besitzt die Abszisse x und die Ordinate y. Sollen nun diejenigen Punkte der graphischen Darstellung von (15) konstruiert werden, welche zu den Abszissen-

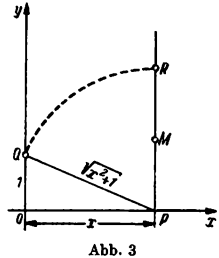


Abb. 3

[1]) Unter der horizontalen Achse verstehen wir im folgenden die x-Achse, unter der Vertikalachse die y-Achse; die Parallelen zu x-Achse werden kurz Horizontalen, die Parallelen zur y-Achse Vertikalen genannt.

werten $x_1, \ldots, x_n$ gehören, so geht man folgendermaßen vor: Man markiert zunächst auf der x-Achse die Punkte $P_1, \ldots, P_n$ mit den Abszissen $x_1, \ldots, x_n$; sodann errichtet man in jedem Punkt P_k die Senkrechte; als nächstes trägt man mit dem Zirkel auf der Senkrechten in P_1 von P_1 aus die Strecke P_1Q, auf der Senkrechten in P_2 von P_2 aus die Strecke P_2Q, usw., schließlich auf der Senkrechten in P_n von P_n aus die Strecke P_nQ ab; endlich ermittelt man den Mittelpunkt M_1 der Strecke P_1R_1, den Mittelpunkt M_2 der Strecke P_2R_2 usw. Die auf diese Weise erhaltenen Punkte $M_1, \ldots, M_n$ sind offenbar die gesuchten.

Bei der Funktion

$$y = \frac{x(x+1)}{x+2} \tag{16}$$

kann man zu gegebenem Wert von x den zugehörigen Wert von y etwa so konstruieren: Man markiert in der x, y-Ebene den Punkt S mit den Koordinaten $(-2, 0)$; ferner zeichnet man die Gerade $y = x + 1$, die durch die Punkte $(-1, 0)$ und $(0, 1)$ geht (Abb. 4). Auf der Horizontalachse wird im Punkt P mit der Abszisse x die Senkrechte errichtet und auf ihr der Schnittpunkt Q

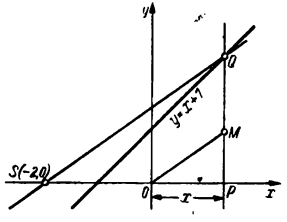


Abb. 4

mit der Geraden $y = x + 1$ markiert. Da (für $x > -1$) im Dreieck PQS die Kathete PQ gleich $x + 1$ und die Kathete PS gleich $x + 2$ ist, genügt es, durch den Punkt O eine Parallele zur Geraden SQ bis zum Schnittpunkt M mit der Geraden PQ zu zeichnen, um ein dem Dreieck PQS ähnliches Dreieck PMO zu erhalten, für das

$$\frac{PM}{OP} = \frac{PQ}{SP},$$

d. h.

$$PM = \frac{OP \cdot PQ}{SP} = \frac{x(x+1)}{x+2}$$

gilt, so daß der Punkt M der graphischen Darstellung von (16) angehört. Sind die zu den Abszissen $x_1, \ldots, x_n$ gehörenden Punkte $M_1, \ldots, M_n$ zu kon-

struieren, so bestimmt man zunächst auf der x-Achse die Punkte $P_1, \ldots, P_n$ mit den Abszissen $x_1, \ldots, x_n$, errichtet in diesen die Senkrechten und bestimmt deren Schnittpunkte $Q_1, \ldots, Q_n$ mit der Geraden $y = x + 1$, zeichnet die Geraden $SQ_1, \ldots, SQ_n$ und bestimmt schließlich die Schnittpunkte $M_1, \ldots, M_n$ der Parallelen zu $SQ_1, \ldots, SQ_n$ durch O mit $P_1Q_1, \ldots, P_nQ_n$. Die Punkte $M_1, \ldots, M_n$ sind die gesuchten.

Schließlich wollen wir uns noch mit der Funktion

$$y = \sqrt{x} \tag{17}$$

befassen. Offenbar kann der Funktionswert y als geometrisches Mittel von x und 1 aufgefaßt werden. Unter Berücksichtigung dessen markieren wir auf der x-Achse den Punkt S mit der Abszisse -1 und den Punkt P mit der Abszisse x, zeichnen den Halbkreis über dem Durchmesser SP und fällen vom Schnittpunkt N dieses Halbkreises mit der y-Achse das Lot auf die Vertikale durch P. Für den Fußpunkt M dieses Lotes gilt

$$PM = ON = \sqrt{x},$$

d. h., der Punkt M gehört zur graphischen Darstellung der Funktion (17) (Abb. 5).

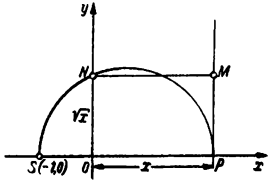


Abb. 5

Die punktweise Konstruktion einer graphischen Darstellung von (17) läuft also darauf hinaus, durch den Punkt S eine Reihe von Halbkreisen mit Mittelpunkten auf der x-Achse zu zeichnen und die Schnittpunkte dieser Halbkreise mit der y-Achse auf die Tangente des Kreises zu projizieren, die durch den S gegenüberliegenden Endpunkt des entsprechenden Durchmessers geht.

Weitere Beispiele von Funktionen der genannten Art, an denen man die Findigkeit in geometrischen Konstruktionen üben kann, lassen sich jederzeit angeben.

Wir wollen uns nun einer ganz natürlichen Verallgemeinerung der im vorangehenden geschilderten Konstruktionsverfahren zuwenden. Dazu nehmen wir an, es seien Funktionen

$$y = u(x), \quad y = v(x), \ldots$$

gegeben, deren graphische Darstellungen bereits vorliegen. Dann verfügt man auch ohne weiteres über eine graphische Darstellung jeder Funktion

$$y = f(x, u(x), v(x), \ldots), \tag{18}$$

wobei die rechte Seite aus den Argumenten $x, u, v, \ldots$ allein mittels der oben angegebenen fünf Operationen aufgebaut ist. In der Tat: Da graphische Darstellungen für die Funktionen $u(x), v(x), \ldots$ vorliegen, kann man für jeden

Wert von x stets Strecken der Länge $u(x), v(x), \ldots$ bestimmen, aus denen man dann mit Hilfe von Zirkel und Lineal die Strecke $f(x, u(x), v(x), \ldots)$ konstruieren kann.

Besonders interessieren uns die folgenden Spezialfälle:

1. Aus graphischen Darstellungen der Funktionen $y = u(x)$ und $y = v(x)$ kann man punktweise eine graphische Darstellung der Funktion

$$y = u(x) + v(x) \tag{19}$$

gewinnen (Abb. 6a).

2. Aus graphischen Darstellungen der Funktionen $y = u(x)$ und $y = v(x)$ kann man punktweise eine graphische Darstellung der Funktion

$$y = u(x) - v(x) \tag{20}$$

gewinnen.

3. Aus graphischen Darstellungen der Funktionen $y = u(x)$ und $y = v(x)$ kann man punktweise eine graphische Darstellung der Funktion

$$y = u(x)\, v(x) \tag{21}$$

gewinnen (Abb. 6b).

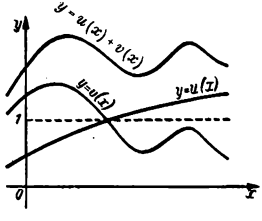


Abb. 6a

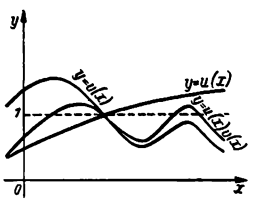


Abb. 6b

4. Aus graphischen Darstellungen der Funktionen $y = u(x)$ und $y = v(x)$ kann man punktweise eine graphische Darstellung der Funktion

$$y = \frac{u(x)}{v(x)} \tag{22}$$

gewinnen.

Man sieht unmittelbar, daß es hierbei auf die Addition, Subtraktion, Multiplikation und Division von Strecken ankommt. „Strecken" sind hierbei natürlich im algebraischen Sinn zu verstehen, denn $u(x)$ und $v(x)$ sind reelle Zahlen, die größer, kleiner oder gleich Null sein können. Die „Multiplikation" z. B. ist also folgendermaßen gemeint: Man multipliziert die Zahlen und nimmt

dann (als Ordinate des zu konstruierenden Punktes) diejenige Strecke, die der Länge und dem Vorzeichen nach mit dem erhaltenen Produkt übereinstimmt.[1])

Häufig ist es angebracht, die Subtraktion und die Division zu vermeiden, indem man die Subtraktion auf die Addition und die Division auf die Multiplikation zurückführt:

$$u - v = u + (-v),$$

$$\frac{u}{v} = u \cdot \frac{1}{v},$$

und sich hierbei auf folgenden Sachverhalt stützt:

5. Aus einer graphischen Darstellung der Funktion $y = v(x)$ kann man punktweise eine graphische Darstellung der Funktion $y = -v(x)$ gewinnen (Abb. 6c).

6. Aus einer graphischen Darstellung der Funktion $y = v(x)$ kann man punktweise eine graphische Darstellung der Funktion $y = \frac{1}{v(x)}$ gewinnen (Abb. 6d).

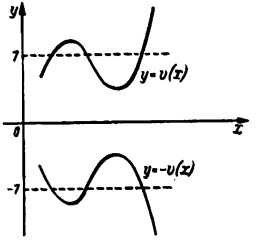


Abb. 6c

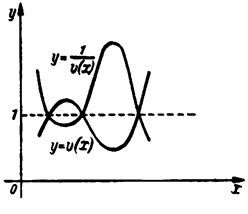


Abb. 6d

In der Tat: Aus einer graphischen Darstellung der Funktion $y = v(x)$ ergibt sich eine graphische Darstellung der Funktion $y = -v(x)$, wenn man jede Ordinate durch die Ordinate mit entgegengesetztem Vorzeichen ersetzt, d. h. die graphische Darstellung von $y = v(x)$ an der x-Achse spiegelt; was dagegen die graphische Darstellung von $y = \frac{1}{v(x)}$ betrifft, so ergibt sie sich aus einer graphischen Darstellung von $y = v(x)$, wenn man in dieser jede Ordinate durch ihren Kehrwert ersetzt.

[1]) Die geometrische Konstruktion in den Fällen 3. und 4. wird offenbar auf eine Bestimmung der sogenannten vierten Proportionalen zurückgeführt, wenn man der Gleichung $y = uv$ die Form $\frac{y}{u} = \frac{v}{1}$ und der Gleichung $y = \frac{u}{v}$ die Form $\frac{y}{1} = \frac{u}{v}$ gibt.

Die in 1. bis 4. erwähnten Operationen an graphischen Darstellungen kann man kurz „Addition", „Subtraktion", „Multiplikation" bzw. „Division" von graphischen Darstellungen nennen. Sie können entweder durch Rechnung und nachfolgende punktweise Konstruktion oder aber auch durch geometrische Konstruktionen allein vollzogen werden.

In diesem Zusammenhang sei auch noch auf folgendes hingewiesen, was nicht in dem oben genannten Prinzip (18) enthalten ist:

7. Aus graphischen Darstellungen der Funktionen $y = u(x)$ und $y = v(x)$ kann man punktweise eine graphische Darstellung der „zusammengesetzten" Funktion („Funktionsfunktion" oder „mittelbaren Funktion")

$$y = u(v(x)) \tag{23}$$

gewinnen.

Dieser Fall möge noch etwas genauer untersucht werden: Es sei P der Punkt der x-Achse, der die Abszisse x besitzt (Abb. 7). Der Schnittpunkt der Senkrechten in P mit dem graphischen Bild von $y = v(x)$ sei Q. Es sei

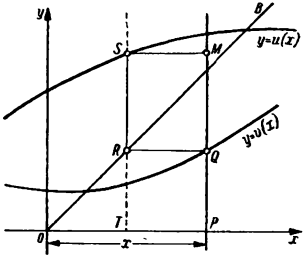


Abb. 7

ferner T der Punkt der x-Achse mit der Abszisse $v(x)$, so daß die Länge von OT gleich der Länge von PQ ist. Die Vertikale durch T schneidet das graphische Bild von $y = u(x)$ in einem Punkt S mit der Ordinate $u(v(x))$. Projizieren wir also den Punkt S auf die Vertikale durch P, so erhalten wir einen Punkt M, der die Abszisse x und die Ordinate $u(v(x))$ besitzt. Statt auf der x-Achse die Strecke OT abzutragen, kann man auch die Winkelhalbierende OB des Winkels zwischen den Koordinatenachsen benutzen (ihre Gleichung ist $x = y$). Man bestimmt den Schnittpunkt R der Horizontalen durch Q mit der Winkelhalbierenden OB und anschließend den Schnittpunkt S der Vertikalen durch R mit dem graphischen Bild von $y = u(x)$; dann verfährt man wie oben.

Hat man sich mit dieser Konstruktion einmal vertraut gemacht, so kann man sie auch sehr schnell für mehrere Punkte $P_1, \ldots, P_n$ durchführen: Man zeichnet zunächst die Strecken $P_\nu Q_\nu$, dann die Strecken $Q_\nu R_\nu$, daraufhin die Strecken $R_\nu S_\nu$ und schließlich die Strecken $S_\nu M_\nu$.

§ 3. Einfachste Transformationen von graphischen Darstellungen

Den Überlegungen dieses Paragraphen liegt folgende Vorstellung zugrunde: Auf irgendeine Weise sei bereits eine graphische Darstellung G der Gleichung

$$F(x, y) = 0 \tag{24}$$

gefunden; dann kann man in einer Reihe von Fällen, in denen sich eine Gleichung $F^*(x, y) = 0$ von der Gleichung $F(x, y) = 0$ nur durch ein bestimmtes, leicht feststellbares Merkmal unterscheidet, aus der graphischen Darstellung G der Gleichung $F(x, y) = 0$ durch eine leicht ausführbare geometrische Transformation eine graphische Darstellung G^* der Gleichung $F^*(x, y) = 0$ gewinnen.

Wir begnügen uns hier mit der Untersuchung der folgenden elementaren Transformationen:

I. a) Spiegelung an der x-Achse;
   b) Spiegelung an der y-Achse;
   c) Spiegelung am Ursprung O.

(Spiegelungen an einer beliebigen Geraden und an einem beliebigen Punkt werden nicht betrachtet.)

II. a) Verschiebung (Bewegung, Translation) um eine gegebene Strecke parallel zur x-Achse;
   b) Verschiebung um eine gegebene Strecke parallel zur y-Achse.

(Verschiebungen parallel zu einer beliebigen Geraden werden nicht betrachtet.)

III. a) Dehnung (oder Stauchung) mit gegebenem Dehnungsfaktor in Richtung der x-Achse;[1])
   b) Dehnung mit gegebenem Dehnungsfaktor in Richtung der y-Achse.

(Dehnungen in beliebiger Richtung werden nicht betrachtet.)

Andere als die hier genannten elementaren Transformationen sollen nicht untersucht werden; jedoch werden auch Transformationen betrachtet, die durch Hintereinanderausführung gewisser dieser elementaren Transformationen erhalten werden können. Auch die Drehungen sollen (bis auf einen unter IV zu behandelnden Spezialfall) außer acht bleiben.

Die genannten Transformationen besitzen die folgenden Eigenschaften (durch die man sie geradezu definieren kann):

I. a) Bei der Spiegelung an der x-Achse geht der Punkt mit den Koordinaten (x, y) in den Punkt mit den Koordinaten $(x, -y)$ über;
   b) bei der Spiegelung an der y-Achse geht der Punkt mit den Koordinaten (x, y) in den Punkt mit den Koordinaten $(-x, y)$ über;

[1]) Eine „Stauchung mit dem Stauchungsfaktor m" ist dasselbe wie eine Dehnung mit dem Dehnungsfaktor $\frac{1}{m}$.

c) bei der Spiegelung am Ursprung O geht der Punkt mit den Koordinaten (x, y) in den Punkt mit den Koordinaten $(-x, -y)$ über.

II. a) Bei einer Verschiebung parallel zur x-Achse um eine Strecke $a\,(a \gtrless 0)$ geht der Punkt mit den Koordinaten (x, y) in den Punkt mit den Koordinaten $(x + a, y)$ über;

b) bei einer Verschiebung parallel zur y-Achse um eine Strecke $b\,(b \gtrless 0)$ geht der Punkt mit den Koordinaten (x, y) in den Punkt mit den Koordinaten $(x, y + b)$ über.

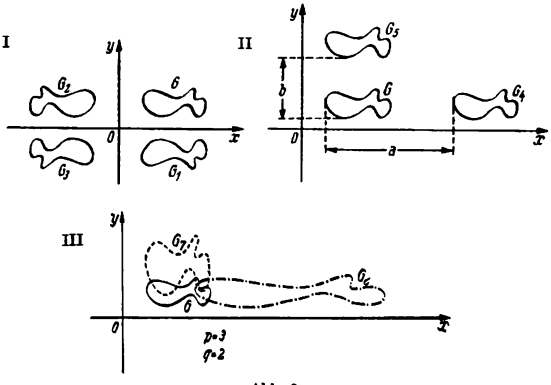


Abb. 8

III. a) Bei einer Dehnung in Richtung der x-Achse mit einem Dehnungsfaktor $p\,(p > 0,\ p \gtrless 1)$ geht der Punkt mit den Koordinaten (x, y) in den Punkt mit den Koordinaten (px, y) über;

b) bei einer Dehnung parallel zur y-Achse mit einem Dehnungsfaktor q $(q > 0, q \gtrless 1)$ geht der Punkt mit den Koordinaten (x, y) in den Punkt mit den Koordinaten (x, qy) über.

Man erkennt hieraus nun leicht, daß die folgenden Sätze gelten (Abb. 8):

I. a) Die graphische Darstellung G_1 der Gleichung

$$F(x, -y) = 0$$

liegt in bezug auf die x-Achse symmetrisch (spiegelbildlich) zur graphischen Darstellung der Gleichung $F(x, y) = 0$;

b) die graphische Darstellung G_2 der Gleichung

$$F(-x, y) = 0$$

liegt in bezug auf die y-Achse symmetrisch zur graphischen Darstellung G der Gleichung $F(x, y) = 0$;

c) die graphische Darstellung G_3 der Gleichung

$$F(-x, -y) = 0$$

liegt in bezug auf den Ursprung O symmetrisch zur graphischen Darstellung der Gleichung $F(x, y) = 0$.

II. a) Die graphische Darstellung G_4 der Gleichung

$$F(x - a, y) = 0$$

ergibt sich aus der graphischen Darstellung G von $F(x, y) = 0$, indem man diese parallel zur x-Achse um die Strecke a verschiebt;

b) die graphische Darstellung G_5 der Gleichung

$$F(x, y - b) = 0$$

ergibt sich aus der graphischen Darstellung G von $F(x, y) = 0$, indem man diese parallel zur y-Achse um die Strecke b verschiebt.

III. a) Die graphische Darstellung G_6 der Gleichung

$$F\left(\frac{x}{p}, y\right) = 0$$

ergibt sich aus der graphischen Darstellung G von $F(x, y) = 0$, indem man diese parallel zur x-Achse um den Faktor p dehnt;

b) die graphische Darstellung G_7 der Gleichung

$$F\left(x, \frac{y}{q}\right) = 0$$

ergibt sich aus der graphischen Darstellung G von $F(x, y) = 0$, indem man diese parallel zur y-Achse um den Faktor q dehnt.

Die graphischen Darstellungen G_6 und G_7 in Abb. 8 (III) entsprechen Dehnungsfaktoren $p = 3$ bzw. $q = 2$.

Beweis von I. a) Es sei $M(x_0, y_0)$ ein Punkt der graphischen Darstellung G, so daß

$$F(x_0, y_0) = 0.$$

Dann gehört der Punkt $M'(x_0, -y_0)$ zur graphischen Darstellung G_1, da seine Koordinaten der Gleichung $F(x, -y) = 0$ genügen. Der Punkt M' liegt jedoch in bezug auf die x-Achse symmetrisch zum Punkt M. Also enthält die graphische Darstellung G_1 sicher alle diejenigen Punkte, die in bezug auf die x-Achse symmetrisch zu irgendeinem Punkt M der graphischen Darstellung G liegen. Würde nun die graphische Darstellung G_1 darüber hinaus noch einen weiteren Punkt $N'(x_1, y_1)$ enthalten, so gälte

$$F(x_1, -y_1) = 0,$$

d. h., der Punkt $N(x_1, -y_1)$, der zu $N'(x_1, y_1)$ in bezug auf die x-Achse symmetrisch liegt, würde zu G gehören, was nicht der Fall sein sollte.

Beweis von II. a) Es sei $M(x_0, y_0)$ ein Punkt der graphischen Darstellung G, so daß

$$F(x_0, y_0) = 0.$$

Dann gehört der Punkt $M'(x_0 + a, y_0)$ der graphischen Darstellung G_a an, da seine Koordinaten der Gleichung $F(x - a, y) = 0$ genügen. Der Punkt M' wird jedoch aus dem Punkt M dadurch erhalten, daß man diesen parallel zur x-Achse um die Strecke a verschiebt; usw.

Beweis von III. a) Es sei $M(x_0, y_0)$ ein Punkt der graphischen Darstellung G, so daß

$$F(x_0, y_0) = 0.$$

Dann gehört der Punkt $M'(px_0, y_0)$ der graphischen Darstellung G_p an, da seine Koordinaten der Gleichung $F\left(\frac{x}{p}, y\right) = 0$ genügen. Der Punkt M' geht jedoch bei der Dehnung der x, y-Ebene parallel zur x-Achse mit dem Dehnungsfaktor p aus dem Punkt M hervor; usw.

Die anderen Sätze werden entsprechend bewiesen.

Aus diesen Sätzen erhält man mühelos entsprechende Sätze für das graphische Bild einer Funktion

$$y = f(x). \tag{25}$$

Dazu setzen wir

$$F(x, y) = y - f(x).$$

Dann ist z. B.

$$F(x, -y) = -y - f(x),$$

so daß die Gleichung $F(x, -y) = 0$ auch in der Form

$$y = -f(x)$$

geschrieben werden kann.

Hieraus folgt:

I'. a) Das graphische Bild der Funktion $-f(x)$ liegt in bezug auf die x-Achse symmetrisch zum graphischen Bild der Funktion $f(x)$ (vgl. § 2, 5).

Entsprechend ergeben sich die folgenden Sätze:

b) Das graphische Bild der Funktion $f(-x)$ liegt in bezug auf die y-Achse symmetrisch zum graphischen Bild der Funktion $f(x)$.

c) Das graphische Bild der Funktion $-f(-x)$ liegt in bezug auf den Ursprung O symmetrisch zum graphischen Bild der Funktion $f(x)$.

II'. a) Das graphische Bild der Funktion $f(x - a)$ erhält man aus dem graphischen Bild der Funktion $f(x)$, indem man dieses parallel zur x-Achse um die Strecke a verschiebt.

b) Das graphische Bild der Funktion $f(x) + b$ erhält man aus dem graphischen Bild der Funktion $f(x)$, indem man dieses parallel zur y-Achse um die Strecke b verschiebt.

III'. a) Das graphische Bild der Funktion $f\left(\frac{x}{p}\right)$ erhält man aus dem graphischen Bild der Funktion $f(x)$, indem man es parallel zur x-Achse mit dem Dehnungsfaktor p dehnt.

b) Das graphische Bild der Funktion $qf(x)$ erhält man aus dem graphischen Bild der Funktion $f(x)$, indem man es parallel zur y-Achse mit dem Dehnungsfaktor q dehnt.

Die bisher betrachteten Transformationen können nun, wie bereits erwähnt, miteinander kombiniert werden. So ergibt sich z. B. die Gleichung $F(x-a,\ y-b)=0$, wenn man in der Gleichung $F(x, y)=0$ zunächst x durch $x-a$ und danach y durch $y-b$ ersetzt (oder umgekehrt). Man erhält also eine graphische Darstellung der Gleichung

$$F(x-a,\ y-b)=0,$$

wenn man eine graphische Darstellung der Gleichung $F(x, y)=0$ parallel zur x-Achse um die Strecke a und parallel zur y-Achse um die Strecke b verschiebt (was natürlich zusammengenommen auf eine Parallelverschiebung längs einer geeignet gewählten Geraden hinausläuft).

Entsprechend erhält man aus einer graphischen Darstellung von $F(x, y)=0$ eine graphische Darstellung der Gleichung

$$F\left(\frac{x}{p},\ \frac{y}{q}\right)=0,$$

wenn man sie in Richtung der x-Achse mit dem Dehnungsfaktor p und in Richtung der y-Achse mit dem Dehnungsfaktor q dehnt.

Entsprechend erhält man aus einer graphischen Darstellung von $F(x, y)=0$ durch zwei hintereinander ausgeführte Spiegelungen eine graphische Darstellung der Gleichung

$$F(-x,\ -y)=0,$$

und zwar spiegelt man die graphische Darstellung von $F(x, y)=0$ zunächst an der x-Achse und danach die dabei erhaltene graphische Darstellung an der y-Achse (oder umgekehrt). Andererseits erhält man (auf Grund von I c)) eine graphische Darstellung von $F(-x, -y)=0$, wenn man eine gegebene graphische Darstellung von $F(x, y)=0$ am Ursprung spiegelt. Das ist nicht verwunderlich, denn man zeigt auch leicht direkt geometrisch, daß die Hintereinanderausführung von Spiegelungen an zueinander senkrechten Geraden jeden Punkt P in einen Punkt überführt, der zu P in bezug auf den Schnittpunkt der beiden Geraden symmetrisch liegt.

Besondere Beachtung verdient noch der Fall, daß sich $F(x, y)$ bei der Substitution von y durch $-y$ nicht ändert, d. h., daß

$$F(x, -y)\equiv F(x, y).$$

In diesem Fall stimmen die graphischen Darstellungen der Gleichungen

$$F(x, -y)=0 \quad \text{und} \quad F(x, y)=0$$

überein. Da andererseits (auf Grund von I a)) die Gleichung $F(x, -y) = 0$ eine graphische Darstellung besitzt, die zur graphischen Darstellung von $F(x, y) = 0$ in bezug auf die x-Achse symmetrisch liegt, bedeutet das Übereinstimmen der graphischen Darstellungen von $F(x, y) = 0$ und $F(x, -y) = 0$, daß die graphische Darstellung G der Gleichung $F(x, y) = 0$ die x-Achse als Symmetrieachse besitzt.

Entsprechend ergibt sich aus I b) bzw. I c):

Ist $F(-x, y) \equiv F(x, y)$, so besitzt die graphische Darstellung G von $F(x, y) = 0$ die y-Achse als Symmetrieachse; ist $F(-x, -y) \equiv F(x, y)$, so besitzt die graphische Darstellung G von $F(x, y) = 0$ den Ursprung O als Symmetriezentrum.

Wir wenden uns nun der Gleichung $y = f(x)$ zu. Ändert sich die Funktion $f(x)$ nicht, wenn man x durch $-x$ ersetzt, gilt also

$$f(-x) \equiv f(x),$$

so nennt man $f(x)$ eine *gerade Funktion*; ändert sich beim Ersetzen von x durch $-x$ nur das Vorzeichen der Funktion $f(x)$, gilt also

$$f(-x) \equiv -f(x),$$

so heißt $f(x)$ *ungerade*. Aus den Sätzen I' b) bzw. I' c) folgt unmittelbar:

Das graphische Bild einer geraden Funktion besitzt die y-Achse als Symmetrieachse, das graphische Bild einer ungeraden Funktion besitzt den Ursprung als Symmetriezentrum.

Ändert sich eine Funktion $F(x, y)$ nicht, wenn man gleichzeitig x durch $x - a$ und y durch $y - b$ ersetzt, gilt also

$$F(x - a, y - b) \equiv F(x, y),$$

so nennt man $F(x, y)$ *periodisch mit der Periode* (a, b); das gleiche sagt man in diesem Fall auch von der graphischen Darstellung der Gleichung $F(x, y) = 0$. Ändert sich eine Funktion $f(x)$ nicht, wenn man x durch $x - a$ ersetzt, so heißt $f(x)$ (und ebenso ihr graphisches Bild) *periodisch mit der Periode* a.

In Abb. 9a ist das graphische Bild einer Gleichung $F(x, y) = 0$ mit einer periodischen Funktion $F(x, y)$ der Periode $(1, 2)$ wiedergegeben, während in Abb. 9b eine periodische Funktion $y = f(x)$ mit der Periode 1 dargestellt ist.

Mit einer Zahl a sind auch die Zahlen $2a, 3a, \ldots, na, \ldots$ und die Zahlen $-a, -2a, \ldots$ Perioden einer periodischen Funktion $f(x)$. Üblicherweise versteht man jedoch unter der Periode einer Funktion die kleinste positive Zahl a, welche Periode von $f(x)$ ist. Entsprechendes gilt auch für Funktionen $F(x, y)$ und ihre graphische Darstellungen.

Man könnte sich auch speziell mit Funktionen $F(x, y)$ bzw. $f(x)$ befassen, die sich bei Substitution von x durch px und y durch qy bzw. x durch px nicht ändern, d. h., für die

$$F(px, qy) \equiv F(x, y), \quad f(px) \equiv f(x)$$

gilt. Die graphische Darstellung von $F(x, y) = 0$ bzw. $y = f(x)$ ist für derartige Funktionen gegenüber den entsprechenden Dehnungen „invariant". Diese Eigenschaft ist für uns hier jedoch ohne besondere Bedeutung.

Kann man von einer graphischen Darstellung sofort feststellen, daß sie symmetrisch bzw. periodisch sein muß, so bedeutet dies eine wesentliche Vereinfachung der eigentlichen Konstruktionsarbeit. Wenn man z. B. erkannt hat, daß die graphische Darstellung in bezug auf die y-Achse symmetrisch ist, so braucht man sie nur in einer Halbebene (z. B. der rechten Halbebene) genau zu konstruieren; liegt die graphische Darstellung in bezug auf beide Achsen symmetrisch, so kann man sich sogar auf einen Quadranten (etwa den ersten) beschränken; ist bekannt, daß die graphische Darstellung die Periode a besitzt, so braucht man nur ihren Verlauf in einem „Periodenstreifen" $\left(\text{z. B. im Intervall } 0 \leqq x < a \text{ oder im Intervall } -\frac{a}{2} < x \leqq \frac{a}{2}\right)$ zu betrachten, usw.

Zur Erläuterung wollen wir einige Beispiele betrachten.

Beispiel 1. Die graphische Darstellung der Funktion $y = 4x^2$ hat die y-Achse als Symmetrieachse. Da man der betrachteten Gleichung auch die Form $\frac{y}{4} = x^2$ geben kann, erhält man ihre graphische Darstellung aus einer graphischen Darstellung von $y = x^2$ durch Dehnung in Richtung der y-Achse mit dem Dehnungsfaktor 4. Andererseits kann man die betrachtete Gleichung aber auch in der Form $y = \left(\frac{x}{1/2}\right)^2$ schreiben, so daß man ihre graphische Darstellung auch aus der von $y = x^2$ durch Stauchung in Richtung der x-Achse mit dem Stauchungsfaktor 2 gewinnen kann.

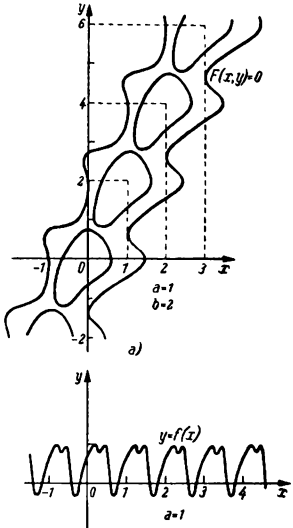


Abb. 9

Beispiel 2. Eine graphische Darstellung von $y = 2x - 1$ erhält man aus einer graphischen Darstellung von $y = 2x$ durch Parallelverschiebung längs der y-Achse um die Strecke -1. Bringt man die betrachtete Gleichung auf die Form $y = 2\left(x - \frac{1}{2}\right)$, so sieht man, daß man ihre graphische Darstellung aus der von $y = 2x$ auch durch Verschiebung längs der x-Achse um die Strecke $\frac{1}{2}$ gewinnen kann.

Beispiel 3. Da die Gleichung $x^2 + y^2 = 2(3x + 4y)$ auch in der Form

$$(x-3)^2 + (y-4)^2 = 25$$

geschrieben werden kann, läßt sich ihre graphische Darstellung aus der von $x^2 + y^2 = 25$ durch Verschiebung parallel zur x-Achse um die Strecke 3 und parallel zur y-Achse um die Strecke 4 gewinnen.

Beispiel 4. Eine graphische Darstellung von $y = 2^{x-3}$ erhält man aus der von $y = 2^x$ durch Parallelverschiebung längs der x-Achse um die Strecke 3. Da man die betrachtete Gleichung andererseits auch auf die Form $y = \frac{1}{8} \cdot 2^x$ bringen kann, läßt sich ihre graphische Darstellung auch aus der von $y = 2^x$ durch Stauchung in Richtung der y-Achse um den Faktor 8 gewinnen.

Beispiel 5. Die Funktionen $y = \cos x$, $y = \sin 2x$, $y = \cos x + \sin 2x$ haben die Periode 2π, π bzw. 2π. Die erste Funktion ist gerade, so daß ihre graphische Darstellung in bezug auf die y-Achse symmetrisch ist; die zweite Funktion ist ungerade, so daß ihre graphische Darstellung in bezug auf den Ursprung symmetrisch ist; die dritte Funktion ist weder gerade noch ungerade.

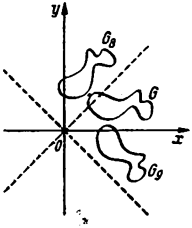


Abb. 10

Beispiel 6. Die Funktionen $y = \operatorname{tg} \pi x$ (ungerade) und $y = \operatorname{tg}^2 \pi x$ (gerade) besitzen beide die Periode 1.

Beispiel 7. Die graphische Darstellung der Funktion $\sin^2 x + \sin^2 y = \frac{1}{2}$ ist symmetrisch in bezug auf beide Achsen und daher auch in bezug auf den Ursprung. Sie hat die Perioden $(\pi, 0)$ und $(0, \pi)$ (und allgemein Perioden der Form $(m\pi, n\pi)$, wobei m und n beliebige ganze Zahlen sind).

Zum Abschluß erwähnen wir noch einen Satz, der sich auf Drehungen bezieht:

IV. Die graphischen Darstellungen G_s und G_g der Gleichungen

$$\text{a) } F\left(\frac{x+y}{\sqrt{2}}, \frac{-x+y}{\sqrt{2}}\right) = 0 \quad \text{und} \quad \text{b) } F\left(\frac{x-y}{\sqrt{2}}, \frac{x+y}{\sqrt{2}}\right) = 0$$

ergeben sich aus der graphischen Darstellung von $F(x, y) = 0$ durch Drehung um den Ursprung um 45° in positivem (d. h. von der x-Achse zur y-Achse gerichtetem) bzw. negativem (d. h. von der y-Achse zur x-Achse gerichtetem) Drehsinn (Abb. 10).

Das ergibt sich aus den Formeln

$$x = x' \cos\varphi - y' \sin\varphi,$$
$$y = x' \sin\varphi + y' \cos\varphi$$

für die Drehung des Koordinatensystems[1]) um einen Winkel φ, wenn man in ihnen $\varphi = \frac{\pi}{4}$ setzt.

§ 4. Funktion und Umkehrfunktion

Im Zusammenhang mit den Sätzen I bis IV des vorangehenden Paragraphen ergibt sich noch die Frage, was mit einer graphischen Darstellung geschieht, wenn man in der zugehörigen Gleichung $F(x, y) = 0$ die Veränderlichen x und y vertauscht.

[1]) Vgl. etwa O.-H. Keller, Analytische Geometrie und lineare Algebra, Berlin 1957.

Es gilt hier der folgende Satz:

V. Die graphische Darstellung G_{10} der Gleichung

$$F(y, x) = 0$$

liegt zur graphischen Darstellung G der Gleichung $F(x, y) = 0$ in bezug auf die Winkelhalbierende $x = y$ des ersten Quadranten symmetrisch (Abb. 11).

Gehört nämlich ein Punkt $M(x_0, y_0)$ zur graphischen Darstellung G, so gehört der entsprechende Punkt $M'(y_0, x_0)$ zur graphischen Darstellung G_{10}; der Punkt M' liegt aber offenbar zum Punkt M symmetrisch in bezug auf die Winkelhalbierende $x = y$.

Ist insbesondere die Funktion $F(x, y)$ „symmetrisch", d. h., ändert sie sich nicht bei Vertauschung von x und y,

$$F(y, x) \equiv F(x, y),$$

so besitzt die graphische Darstellung der Gleichung $F(x, y) = 0$ die Gerade $x = y$ als Symmetrieachse.

Beispiele für symmetrische Funktionen sind:

$$x + y = 1, \quad x^2 + y^2 = 25, \quad xy = 1,$$

$$x + y = xy, \quad \sqrt{x} + \sqrt{y} = 1$$

usw.

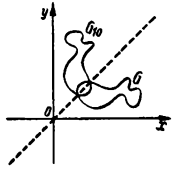


Abb. 11

Es sei $F(x, y) = 0$ eine Gleichung, durch die sowohl y als (eindeutige oder mehrdeutige, vgl. § 1) Funktion von x als auch x als eindeutige oder mehrdeutige Funktion von y festgelegt wird. Diese Funktionen, die im allgemeinen (vom Standpunkt der auszuführenden Operationen aus) verschieden sind, nennt man mitunter *Funktion* und *Umkehrfunktion*.

Wir wollen uns hier auf den einfachsten Fall beschränken und annehmen, daß in einem gewissen Teil der Ebene (oder evtl. der ganzen Ebene) die graphische Darstellung der Gleichung

$$F(x, y) = 0 \tag{26}$$

mit jeder Parallelen zur y-Achse und jeder Parallelen zur x-Achse nur einen Punkt gemeinsam hat. Dies hat zur Folge, daß man in dem betreffenden Teil der Ebene die Gleichung (26) sowohl nach y als auch nach x „auflösen" kann:

$$y = f(x), \tag{27}$$

$$x = g(y). \tag{28}$$

Wenn wir noch annehmen, daß man aus jedem in Frage kommenden Wert von x den zugehörigen Wert von y allein mit Hilfe von elementaren Operationen ausrechnen kann und entsprechendes auch für die Berechnung von x aus y gilt, so sind die durch f und g bezeichneten Funktionen elementare Funktionen. In diesem Fall kann man f als *Funktion* und g als *Umkehrfunktion* ansehen (oder umgekehrt). Da die Gleichungen (26), (27) und (28) auf Grund

unserer Voraussetzungen gleichwertig sind, besitzen sie alle dieselbe graphische Darstellung G. In Gleichung (28) spielt dabei jedoch, entgegen unseren generellen Verabredungen, die Variable y die Rolle der unabhängigen Veränderlichen, während x die abhängige Veränderliche bezeichnet. Vertauschen wir in (28) die Buchstaben x und y, so erhalten wir vom Standpunkt der auszuführenden Operationen aus dieselbe Gleichung

$$y = g(x), \tag{29}$$

die jedoch im allgemeinen eine ganz andere graphische Darstellung besitzt, nämlich die graphische Darstellung, die man erhält, wenn man G an der Geraden $x = y$ spiegelt.

Wir erhalten damit den folgenden wichtigen Satz: *Das graphische Bild der Umkehrfunktion $y = g(x)$ einer Funktion $y = f(x)$ liegt zum graphischen Bild von $y = f(x)$ in bezug auf die Gerade $x = y$ symmetrisch.*

Beispiel 1. Die Gleichung $2x - 5y + 10 = 0$ (vgl. § 1) ist jeder der Gleichungen

$$y = \frac{2x+10}{5} \quad \text{und} \quad x = \frac{5y-10}{2}$$

gleichwertig. Nimmt man

$$f(x) = \frac{2x+10}{5}$$

als Funktion, so wird

$$g(x) = \frac{5x-10}{2}$$

die zugehörige Umkehrfunktion.

Beispiel 2. Da man die Gleichung $x^2 - y = 0$ auch in der Form

$$y = x^2$$

schreiben kann, erhält man, wenn man sie nach x auflöst,

$$x = \begin{cases} +\sqrt{y} \text{ in der rechten Halbebene } (x > 0), \\ -\sqrt{y} \text{ in der linken Halbebene } (x < 0). \end{cases}$$

Nimmt man also $f(x) = x^2$ als Funktion, so erhält man $g(x) = \sqrt{x}$ als Umkehrfunktion (die Wurzel im algebraischen Sinn), oder genauer: in der oberen Halbebene ($y > 0$) ist $g(x) = +\sqrt{x}$ und in der unteren Halbebene ($y < 0$) ist $g(x) = -\sqrt{x}$ Umkehrfunktion (jetzt die Wurzeln im arithmetischen Sinne verstanden).[1])

Beispiel 3. Für die Funktion $f(x) = \frac{1}{100}(x^3 + 1)$ ist $g(x) = \sqrt[3]{100x - 1}$ die Umkehrfunktion.

Beispiel 4. Die Funktion

$$f(x) = \frac{x}{1+x}$$

---

[1]) In der deutschen Literatur bezeichnet man in der Regel die hier Wurzel im algebraischen Sinn genannte Wurzel mit unbestimmten Vorzeichen als Wurzel schlechthin, während man den hier Wurzel im arithmetischen Sinn genannten nicht negativen Wert der Wurzel als positive Wurzel oder Betrag der Wurzel bezeichnet. — *Anm. d. wissenschaftl. Red.*

besitzt die Funktion

$$g(x) = \frac{x}{1-x}$$

als Umkehrfunktion.

Beispiel 5. Nimmt man als Funktion

$$f(x) = \lg \cos x,$$

so erhält man als Umkehrfunktion

$$g(x) = \arccos 10^x.$$

§ 5. Elementare Diskussion von Funktionen (Fragestellung und einige allgemeine Verfahren)

Es sei eine Gleichung der Form

$$y = f(x) \tag{30}$$

gegeben. Unter „Diskussion" der Funktion $f(x)$ (oder ihrer graphischen Darstellung) kann man eine Erläuterung der verschiedensten Eigenschaften von $f(x)$ verstehen. Meistens sucht man jedoch, wenn man von „Diskussion" spricht, nur Antwort auf folgende Fragen zu erhalten:

A. Nimmt die Funktion $f(x)$ den Wert Null an und, wenn ja, an welchen Stellen? In welchen Intervallen ist die Funktion $f(x)$ positiv und in welchen negativ?

B. In welcher Weise „ändert" sich die Funktion $f(x)$? In welchen Intervallen wächst die Funktion $f(x)$ und in welchen fällt sie? Wo nimmt die Funktion $f(x)$ einen größten Wert (ein Maximum) und wo einen kleinsten Wert (ein Minimum) an?

Hierbei bedürfen zunächst die Bezeichnungen „wächst", „fällt", „Maximum" und „Minimum" einer näheren Erläuterung.

Man nennt eine Funktion $f(x)$ in einem Intervall *wachsend*, wenn für zwei beliebige Werte von x aus diesem Intervall, etwa für x' und x'', die Beziehung

$$x' < x'' \tag{31}$$

die Beziehung

$$f(x') < f(x'') \tag{32}$$

zur Folge hat. (Dann folgt natürlich aus der Ungleichung (32) auch die Ungleichung (31); denn wäre $x'' < x'$, so wäre $f(x'') < f(x')$, während die Gleichheit $x'' = x'$ die Gleichheit $f(x'') = f(x')$ zur Folge hätte.)

Anders ausgedrückt besagt dies: *Eine Funktion $f(x)$ ist wachsend, wenn mit den Argumentwerten* (den Werten für die unabhängige Veränderliche) *auch die Funktionswerte* (die Werte für die abhängige Veränderliche) *wachsen und umgekehrt; entsprechend werden mit den Argumentwerten auch die Funktionswerte kleiner und umgekehrt.* Bei einer wachsenden Funktion ändern sich also die Veränderlichen x und y „in gleichem Sinne".[1])

[1]) Beispiele für wachsende Funktionen sind: $f(x) = mx$ $(m > 0)$, $f(x) = x^3$ in jedem Intervall; $f(x) = x^2$, $f(x) = \sqrt{x}$ im Intervall $0 \leqq x < \infty$.

Man nennt eine Funktion $f(x)$ in einem Intervall *fallend*, wenn für zwei beliebige Zahlen x' und x'' aus diesem Intervall die Beziehung

$$x' < x'' \tag{33}$$

die Beziehung

$$f(x') > f(x'') \tag{34}$$

zur Folge hat (und dann natürlich auch das Umgekehrte gilt). In diesem Fall hat also ein Wachsen der Argumentwerte ein Fallen der Funktionswerte und ein Fallen der Argumentwerte ein Wachsen der Funktionswerte zur Folge; die Veränderlichen x und y ändern sich „in verschiedenem Sinne".[1])

Es sei $a < c < b$; eine Funktion, die im Intervall von a bis c wächst und im Intervall von c bis b fällt, besitzt an der Stelle c ein Maximum; eine Funktion, die im Intervall von a bis c fällt und im Intervall von c bis b wächst, besitzt an der Stelle c ein Minimum. Das sind jedoch nicht die üblichen Definitionen für ein Maximum bzw. Minimum. Diese lauten folgendermaßen: Eine Funktion $f(x)$ besitzt an der Stelle $x = c$ ein *Maximum*, wenn es Zahlen a und b mit $a < c < b$ gibt, so daß für jedes x mit $a < x < b$ und $x \neq c$ die Ungleichung

$$f(c) > f(x) \tag{35}$$

erfüllt ist; entsprechend für ein *Minimum*. Später (§ 46, Beispiel 4) wird an einem Beispiel gezeigt werden, daß die zweite Formulierung eine gewisse Verallgemeinerung der ersten ist.

Die geometrische Bedeutung der Begriffe „wachsend", „fallend", „Maximum" und „Minimum" ist so klar, daß sich eine weitere Erläuterung erübrigt.

Bemerkung 1. Die Begriffe „wachsende Funktion" und „fallende Funktion" können noch etwas erweitert werden, und zwar dadurch, daß man an die Stelle der Beziehung (32) als Folge der Ungleichung $x' < x''$ die weniger einschränkende Beziehung

$$f(x') \leqq f(x'') \tag{32'}$$

setzt. Hierdurch wird zum Ausdruck gebracht, daß beim Anwachsen der Argumentwerte die Funktionswerte ebenfalls größer werden oder aber sich nicht ändern (also jedenfalls nicht kleiner werden). Eine Funktion, die diese Eigenschaft besitzt, heißt *nicht-fallend*.

Jede wachsende Funktion ist nicht-fallend; die Umkehrung hiervon gilt nicht.

Hier ein Beispiel für eine nicht-fallende Funktion „aus dem Alltag": Es sei x die Zeit, $f(x)$ der Stand des Elektrizitätszählers. Bei Entnahme von elektrischer Energie steigt mit der Zeit auch der Zählerstand; wird jedoch im Laufe eines Zeitraumes keine Energie entnommen, so bleibt der Zählerstand unverändert.

Man kann zeigen, daß jede elementare nicht-fallende Funktion (die nicht in eine Konstante entartet) beständig wächst.

---

[1]) Beispiele für fallende Funktionen sind: $f(x) = mx$ $(m < 0)$ in jedem Intervall; $f(x) = \frac{1}{x}$, $\frac{1}{x^s}$ $(s > 0)$ für $0 < x < \infty$.

Entsprechend wird eine Funktion $f(x)$, die an Stelle der Beziehung (34) der Beziehung

$$f(x') \geqq f(x'') \tag{34'}$$

genügt, *nicht-wachsend* genannt.

Bemerkung 2. In ähnlicher Weise lassen sich der Begriff des Maximums und des Minimums erweitern: Eine Funktion $f(x)$ besitzt im Punkt $x = c$ ein *Maximum im weiteren Sinn* (schwaches Maximum), wenn an Stelle der Ungleichung (35) in einer gewissen „Umgebung" von c die Beziehung

$$f(c) \geqq f(x) \tag{35'}$$

gilt. Entsprechend für ein Minimum.

Bemerkung 3. Die nicht-fallenden und die nicht-wachsenden Funktionen werden zusammenfassend *monotone* Funktionen genannt.

Bemerkung 4. Die Begriffe „Maximum" und „Minimum" werden zum Begriff „Extremum" zusammengefaßt.

Bezüglich der Fragen vom Typ A (s. S. 27) beschränken wir uns hier auf eine allgemeine Bemerkung: Um Aufschluß über das Vorzeichen der Werte einer Funktion $f(x)$ zu erhalten, versucht man gewöhnlich, die Formel, die die Funktion $f(x)$ beschreibt, (durch sogenannte identische Umformungen) in ein Produkt von Faktoren umzuwandeln, deren Vorzeichenverhalten sich leicht übersehen läßt; alles weitere ergibt sich dann auf Grund der elementaren Eigenschaften eines Produktes:

1. Ein Produkt ist dann und nur dann gleich Null, wenn wenigstens einer der Faktoren verschwindet.

2. Sind alle Faktoren von Null verschieden, so ist das Produkt positiv bzw. negativ, je nachdem, ob die Anzahl der negativen Faktoren gerade oder ungerade ist.

In den Fällen, in denen eine solche Faktorzerlegung nicht gelingt, versucht man die Fragen vom Typ A auf Fragen vom Typ B zurückzuführen. Hat man z. B. festgestellt, daß eine Funktion in einem Intervall wächst und daß sie im Anfangspunkt des Intervalles einen positiven Wert annimmt, so kann man sofort behaupten, daß die Funktion im ganzen Intervall positiv ist. Zur selben Schlußfolgerung gelangt man, wenn man nur weiß, daß der kleinste Wert der Funktion in betrachteten Intervall positiv ist.

Die Stellen, an denen eine Funktion $f(x)$ den Wert Null annimmt, nennt man üblicherweise *Nullstellen* der Funktion $f(x)$.

Bezüglich der Fragen vom Typ B läßt sich allgemein sagen, daß ihre Beantwortung prinzipiell schwieriger ist; Methoden hierfür werden in der Differentialrechnung bereitgestellt. Insbesondere gehört die Frage nach dem Wachsen oder Fallen einer Funktion zu den Grundaufgaben der Differentialrechnung; sie wird dort dadurch gelöst, daß man das „Wachstumsverhalten" einer Funktion $f(x)$ auf die Untersuchung des Vorzeichens einer anderen Funktion, der sogenannten „Ableitung" von $f(x)$ zurückführt (vgl. S. 291).[1])

Es gibt jedoch auch verschiedene, wenn auch keineswegs erschöpfende, elementare Methoden zur Untersuchung der Art der Änderung einer Funktion;

[1]) Das heißt, eine Frage vom Typ B wird auf eine Frage von Typ A zurückgeführt.

ihnen kommt im Hinblick auf den Schulunterricht besondere Bedeutung zu, so daß es angebracht erscheint, auf einige von ihnen hier näher einzugehen. Darüber hinaus muß festgestellt werden, daß es in einigen Fällen überhaupt überflüssig ist, auf Hilfsmittel der Differentialrechnung zurückzugreifen, da sich diese Aufgabe auf einfachstem und natürlichstem Wege durch elementare Methoden lösen läßt; in anderen Fällen dagegen ist eine elementare Lösung, wenn sie auch möglich ist, viel komplizierter als eine Lösung mit Hilfsmitteln der Differentialrechnung und häufig sehr gekünstelt (vgl. den Schluß von § 10).

Die elementaren Methoden gründen sich darauf, daß man für einfachste Funktionen die „Wachstumsverhältnisse" direkt durch Rückgang auf die Definition des Wachsens und Fallens ermitteln kann. Von diesen kann man durch Anwendung einer Reihe von fast trivialen Sätzen auf die Art der Änderung komplizierterer Funktionen schließen.

Zur Abkürzung wollen wir sagen, daß sich Funktionen $u(x)$ und $v(x)$ in einem gewissen Intervall *im gleichen Sinn ändern*, wenn sie beide wachsen oder beide fallen; dagegen werden wir sagen, die Funktionen $u(x)$ und $v(x)$ *ändern sich im entgegengesetzten Sinn*, wenn die eine wächst und die andere fällt.

Es gelten dann die folgenden Sätze:

1. Die Funktionen $u(x)$ und $u(x) + C$, wobei C eine beliebige Konstante ist, ändern sich im gleichen Sinn.

2. und 2′. Die Funktionen $u(x)$ und $Cu(x)$ ändern sich im gleichen oder im entgegengesetzten Sinn, je nachdem, ob die Konstante C positiv oder negativ ist.

3. Ändern sich die Funktionen $u(x)$ und $v(x)$ im gleichen Sinn, so ändert sich auch die Funktion $u(x) + v(x)$ in diesem Sinn.

4. Ändern sich positive Funktionen $u(x)$ und $v(x)$ im gleichen Sinn, so ändert sich auch das Produkt $u(x)\,v(x)$ in eben diesem Sinn.

4′. Ändern sich negative Funktionen $u(x)$ und $v(x)$ im gleichen Sinn, so ändert sich das Produkt $u(x)\,v(x)$ in dem zu $u(x)$ und $v(x)$ entgegengesetzten Sinn.

5. und 5′. Ist $u(x)$ eine positive Funktion, so ändern sich $u(x)$ und $u^2(x)$ im gleichen Sinn; ist $u(x)$ eine negative Funktion, so ändern sich $u(x)$ und $u^2(x)$ im entgegengesetzten Sinn.

6. Die Funktionen $u(x)$ und $-u(x)$ ändern sich im entgegengesetzten Sinn.

7. Ist die Funktion $u(x)$ in einem gewissen Intervall positiv, so ändern sich in diesem Intervall die Funktionen $u(x)$ und $\frac{1}{u(x)}$ im entgegengesetzten Sinn.

7′. Ist die Funktion $u(x)$ in einem gewissen Intervall negativ, so ändern sich in diesem Intervall die Funktionen $u(x)$ und $\frac{1}{u(x)}$ ebenfalls im entgegengesetzten Sinn.

Man kann also sagen, daß in einem Intervall, in dem die Funktion $u(x)$ überall von Null verschieden ist und in dem sie ihr Vorzeichen nicht ändert, sich die Funktionen $u(x)$ und $\frac{1}{u(x)}$ im entgegengesetzten Sinn ändern.

8. Ist die Funktion $u(x)$ in einem gewissen Intervall positiv, so ändern sich in diesem Intervall die Funktionen $u(x)$ und $\sqrt{u(x)}$ im gleichen Sinn (unter $\sqrt{u(x)}$ die Wurzel im arithmetischen Sinn, d. h. als „positive Wurzel" verstanden).

9. Eine „zusammengesetzte Funktion" $u(v(x))$ wächst, wenn sich $u(x)$ und $v(x)$ im gleichen Sinn ändern; sie fällt, wenn sich $u(x)$ und $v(x)$ im entgegengesetzten Sinn ändern.[1])

Die Beweise dieser Sätze sind alle von derselben Struktur und äußerst elementar. Man hat dabei die Definition für das Wachsen bzw. Fallen einer Funktion und einfachste Sätze über das Rechnen mit Ungleichungen zu verwenden. Als Beispiel beweisen wir Satz 3 für den Fall wachsender Funktionen, d. h.:

Eine Summe von wachsenden Funktionen ist eine wachsende Funktion.

Beweis. Es sei $f(x) = u(x) + v(x)$, wobei $u(x)$ und $v(x)$ in einem gewissen Intervall wachsende Funktionen sind. Es seien ferner x' und x'' beliebige Zahlen aus diesem Intervall, und zwar sei $x' < x''$. Dann ist nach Voraussetzung

$$u(x') < u(x''),$$
$$v(x') < v(x'').$$

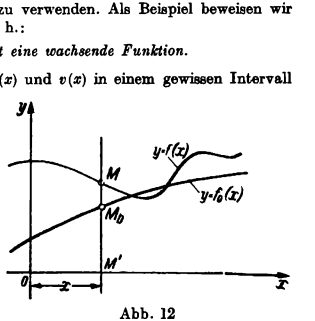


Abb. 12

Hieraus ergibt sich unmittelbar

$$u(x') + v(x') < u(x'') + v(x''),$$

d. h.

$$f(x') < f(x'').$$

Es sei an dieser Stelle noch erwähnt, daß man sich bei der „Diskussion" einer Funktion natürlich nicht streng nur an die unter A und B aufgezählten Fragen hält; die Antwort auf weitere interessierende Fragen kann aber häufig auf die Beantwortung von Fragen vom Typ A oder B zurückgeführt werden. Ein Beispiel möge dies erläutern:

Neben der graphischen Darstellung G einer zu diskutierenden Funktion $y = f(x)$ liege bereits die gut bekannte graphische Darstellung G_0 einer Funktion $y = f_0(x)$ vor (Abb. 12). Es kann dann unter Umständen die Beantwortung folgender Fragen wesentlich sein:

A'. Schneidet die graphische Darstellung G die graphische Darstellung G_0 und wenn ja, in welchen Punkten? In welchen Intervallen liegt G oberhalb G_0 und in welchen darunter?

[1]) Die Änderung der Funktion $u(x)$ bezieht sich hierbei auf das Intervall, dem bei veränderlichem x die Werte der Funktion $v(x)$ angehören.

B'. Wie ändert sich (mit wachsendem x) die Strecke M_0M zwischen den Schnittpunkten der vertikalen Geraden mit G_0 bzw. G? In welchen Intervallen wächst bzw. fällt sie? In welchen Punkten nimmt sie einen maximalen bzw. minimalen Wert an?

Man sieht unmittelbar, daß die Beantwortung der Fragen A' und B' auf die Beantwortung der Fragen A und B für die Funktion $f(x) - f_0(x)$ hinausläuft.

Bei der Diskussion der verschiedenen Eigenschaften einer Funktion $f(x)$ hat man häufig die Formel, die die Funktion $f(x)$ beschreibt, identischen Umformungen zu unterwerfen. Welche Umformungen im konkreten Fall vorzunehmen sind, hängt davon ab, welche Eigenschaft der Funktion man betrachtet.

Kapitel II

ÜBERBLICK ÜBER DIE ELEMENTAREN FUNKTIONEN UND IHRE GRAPHISCHEN DARSTELLUNGEN

§ 6. Klassifikation der rationalen Funktionen

Eine Funktion $f(x)$ heißt *rational*, wenn man jeden ihrer Werte $y = f(x)$ aus dem Argumentwert x und konstanten Zahlen durch alleinige Anwendung der vier Spezies (Addition, Subtraktion, Multiplikation und Division) berechnen kann;[1]) sie heißt eine *ganze rationale* (oder *ganzrationale*) *Funktion*[2]), wenn man allein mit Hilfe der drei Rechenoperationen Addition, Subtraktion und Multiplikation auskommt. Die Subtraktion braucht hierbei nicht unbedingt erwähnt zu werden, denn die Subtraktion einer Zahl ist ja nichts anderes als eine Addition nach vorheriger Multiplikation mit -1. Die rationalen Funktionen, die nicht ganzrational sind, heißen *gebrochene rationale Funktionen*. Beispiele für ganze rationale Funktionen sind:

$$2x-3, \quad (x-1)(x+4), \quad x^3, \quad \frac{x}{5}\left(=x\cdot\frac{1}{5}\right),$$

$$\sqrt{x^4+4x^2+4}\ (=x^2+2),^3) \quad \frac{x^2-1}{x-1}\ (=x+1),$$

$$\cos(2\arccos x)\ (=2x^2-1).^4)$$

Beispiele für gebrochene rationale Funktionen sind:

$$\frac{x+1}{x^2-2x+5}, \quad x+\frac{1}{x}, \quad x^{-3}\left(=\frac{1}{x^3}\right), \quad \operatorname{tg} 3\operatorname{arctg} x\left(=\frac{3x-x^3}{1-3x^2}\right).$$

Man beweist nun leicht durch vollständige Induktion:

1. *Jede ganzrationale Funktion $P(x)$ läßt sich als Polynom in x der Form*

$$P(x) = ax^n + bx^{n-1} + \cdots + kx + l$$

darstellen, wobei $a \neq 0$ und n eine nicht negative ganze Zahl ist. Die hierbei auftretende Zahl n heißt der *Grad* von $P(x)$.

---

[1]) Gemeint ist natürlich die Berechenbarkeit nach einer Formel, in die nur Zeichen für diese Operationen eingehen, was natürlich nicht bedeuten soll, daß die Funktion von vornherein durch eine solche Formel gegeben sein muß (vgl. die folgenden Beispiele). Zur notwendigen Präzisierung des Begriffs der rationalen Funktion vgl. L. J. Okunjew, Der Ring der Polynome und der Körper der rationalen Funktionen, EdEM Bd. II, insbes. Seite 153 und Seite 234. — *Anm. d. wissenschaftl. Red.*

[2]) Üblicherweise nennt man die ganzrationalen Funktionen auch *Polynome*.

[3]) Gemeint ist hier der nicht negative Wert der Wurzel.

[4]) Von der hier zunächst notwendigen Voraussetzung $|x| \leqq 1$ werden wir uns im Artikel „Elementare Funktionen einer komplexen Veränderlichen" befreien.

2. *Jede gebrochene rationale Funktion $R(x)$ läßt sich als Quotient zweier Polynome $P(x)$ und $Q(x)$ ohne gemeinsamen Faktor darstellen:*

$$R(x) = \frac{P(x)}{Q(x)}.$$

Der *Grad* einer gebrochenen rationalen Funktion ist das Maximum von Grad des Zählers $P(x)$ und Grad des Nenners $Q(x)$.[1])

Die vollständige Induktion beim Beweis von 1. wird über die Anzahl N der bei der Berechnung der Werte von $P(x)$ verwendeten Operationsschritte geführt. Dazu bemerken wir zunächst, daß a) die Summe, b) die Differenz und c) das Produkt von Polynomen der oben angegebenen Form als Polynome dieser Form dargestellt werden können. Wir nehmen nun an, daß die Behauptung 1. für jede ganzrationale Funktion gilt, die aus der Variablen x und Konstanten mit Hilfe von N Operationen „gebildet" ist, und zeigen, daß sie dann auch für jede ganzrationale Funktion richtig ist, die auf die angegebene Weise mittels $N+1$ Operationen erhalten wird: Die letzte, $(N+1)$-te Operation, die in der Formel auftritt, mit deren Hilfe die betrachtete Funktion berechnet wird, muß entweder eine Addition, Subtraktion oder Multiplikation sein. Da aber jede der beiden ganzrationalen Funktionen, an denen diese Operation auszuführen ist, mit nicht mehr als N Operationen „gebildet" ist, kann jede von ihnen als Polynom in x dargestellt werden. Folglich kann auf Grund der vorangehenden Bemerkung auch die gegebene Funktion als Polynom in x dargestellt werden.

Es bleibt zu erwähnen, daß die Behauptung für den Fall $N = 0$ trivialerweise richtig ist.

Die Behauptung 2. beweist man ebenfalls durch vollständige Induktion, diesmal über die Anzahl der auszuführenden Operationen einschließlich Division. Hier ist zuvor zu bemerken, daß a) die Summe, b) die Differenz, c) das Produkt und d) der Quotient von Brüchen, in denen Zähler und Nenner Polynome sind, sich ebenfalls als Brüche dieser Art darstellen lassen.

§ 7. Die Potenzfunktionen mit positivem ganzem Exponenten

Wir betrachten die graphischen Darstellungen der Funktionen

$$x,\ x^2,\ x^3,\ \ldots,\ x^n,\ \ldots. \tag{1}$$

Das Bild G_1 der Funktion

$$y = x \tag{2}$$

[1]) Man beachte den Unterschied zwischen dem Grad einer rationalen Funktion und der „Ordnung" (vgl. § 17) der durch sie dargestellten algebraischen Kurve (ihrer graphischen Darstellung). Vom geometrischen Standpunkt ist die Ordnung einer Kurve die Maximalzahl ihrer Schnittpunkte mit einer beliebigen Geraden, während der Grad einer Funktion die Maximalzahl der Schnittpunkte ihrer graphischen Darstellung mit einer beliebigen horizontalen Geraden ist (Schnittpunkte jeweils unter Einschluß der „komplexen" Schnittpunkte). Für ganze rationale Funktionen stimmen Grad und Ordnung überein, für gebrochene rationale Funktionen können sie verschieden sein. So ist z. B. die Funktion $y = \frac{1}{x}$ vom Grad Eins, während ihre graphische Darstellung, eine Hyperbel, eine Kurve zweiter Ordnung ist.

besteht aus allen den Punkten, deren Abszisse gleich ihrer Ordinate ist, d. h. aus den Punkten der Winkelhalbierenden des ersten und dritten Quadranten; es ist also die *Gerade*, die durch den Ursprung O geht und mit der positiven Richtung der x-Achse einen Winkel von 45° einschließt (Abb. 13a).

Das graphische Bild G_2 der Funktion

$$y = x^2 \tag{3}$$

(Abb. 13b) ist eine sogenannte *Parabel* (im weiteren Sinne nennt man zuweilen auch die graphischen Bilder beliebiger Polynome Parabeln). Es liegt symmetrisch zur y-Achse, denn die Funktion x^2 ist gerade. Mit Ausnahme eines Punktes, des Ursprungs O, des sogenannten *Scheitels* der Parabel, in dem

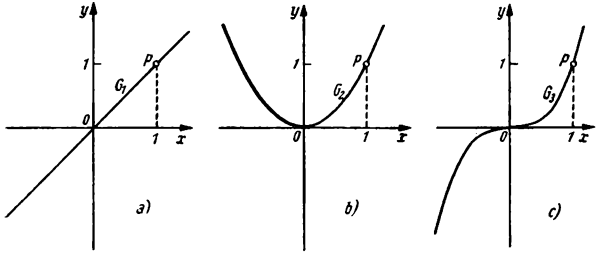


Abb. 13

die graphische Darstellung G_3 ihre Symmetrieachse schneidet, liegen sämtliche Punkte von G_3 in der oberen Halbebene; denn es ist stets $x^2 \geqq 0$, und es gilt $x^2 = 0$ nur für $x = 0$. Da die Funktion $y = x$ wachsend ist, ist für $x > 0$ die Funktion x^2 ebenfalls wachsend, während sie wegen ihrer Symmetrie zur y-Achse (oder auf Grund von Satz 4′ von Seite 30) für $x < 0$ fallend ist.

Um die gegenseitige Lage der graphischen Darstellungen von (2) und (3) zu ermitteln, betrachten wir die Differenz $x^2 - x$. Der Faktorzerlegung

$$x^2 - x = x(x-1)$$

entnimmt man, daß G_1 und G_2 genau zwei Punkte gemeinsam haben, den Ursprung O mit den Koordinaten (0, 0) und den Punkt P mit den Koordinaten (1,1). Für $x > 1$ ist $x^2 - x > 0$, so daß dort G_2 oberhalb G_1 liegt; für $0 < x < 1$ ist $x^2 - x < 0$, d. h., G_2 liegt dort unterhalb G_1; für $x < 0$ liegt schließlich G_2 wieder oberhalb G_1.

Für sehr kleine Werte von $|x|$ ist der Quotient

$$\frac{y}{x} = x$$

seinem Betrage nach sehr klein, d. h., die graphische Darstellung von G_2 schmiegt sich in der Umgebung des Ursprunges O (ihres Scheitels) eng an die x-Achse an (sie „berührt" sie im Punkte O, vgl. S. 293). Für große Werte von $|x|$ ist dagegen dieser Quotient sehr groß, was von einem schnellen Wachsen der Funktion x^2 zeugt.

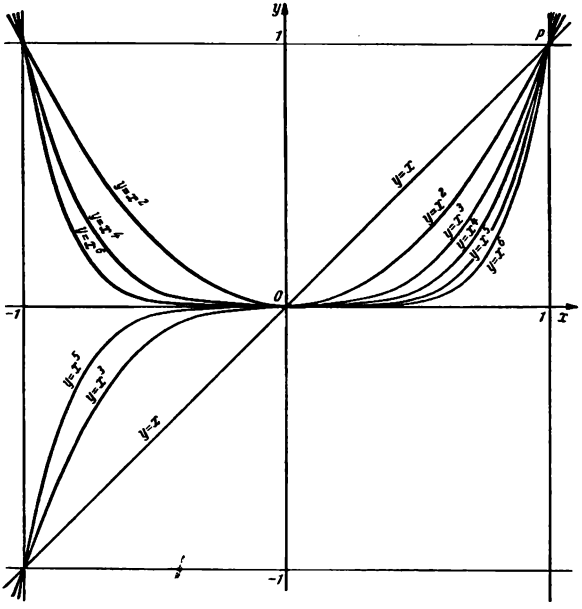


Abb. 14

Das graphische Bild G_3 der Funktion

$$y = x^3 \tag{4}$$

(die „kubische" Parabel) hat den Ursprung O als Symmetriezentrum (Abb. 13c); denn die Funktion x^3 ist ungerade. Ähnlich wie G_1 liegt es im ersten und dritten Quadranten. Mit G_2 hat G_3 die Punkte O und $P(1, 1)$ gemeinsam. Im Intervall $0 < x < 1$ liegt G_3 unterhalb, im Intervall $1 < x < \infty$ oberhalb G_2. In der Nähe des Ursprungs schmiegt sich G_3 noch enger an die x-Achse an als

G_2 (im Punkt O berührt G_3 die x-Achse „von zweiter Ordnung", wobei die Tangente die Kurve schneidet); bei unbeschränktem Wachsen von x wächst x^3 noch schneller als x^2.

Es sei allgemein (für beliebiges positives ganzes n) G_n das graphische Bild der Funktion

$$y = x^n.$$

Man stellt leicht fest, daß für jedes n das Bild G_n durch die Punkte O und $P(1, 1)$ geht. Für gerades n hat G_n die y-Achse als Symmetrieachse und liegt im ersten und zweiten Quadranten, während für ungerades n der Ursprung O Symmetriezentrum ist und G_n im ersten und dritten Quadranten liegt. Im Intervall $0 < x < 1$ bilden für jedes x die entsprechenden Werte der Funktionen (1) eine fallende geometrische Folge, d. h., in der Folge

$$G_1,\ G_2,\ G_3,\ \ldots,\ G_n,\ \ldots \tag{5}$$

der graphischen Bilder liegt im Intervall $0 < x < 1$ jedes vorangehende oberhalb aller folgenden, und die Kurven schmiegen sich immer enger an die x-Achse an. Im Intervall $1 < x < \infty$ bilden die betreffenden Werte eine wachsende geometrische Folge, und bei den graphischen Bildern liegt jedes vorangehende unterhalb der folgenden.

In Abb. 14 sind die graphischen Bilder der Funktionen x, x^2, x^3, x^4, x^5, x^6 im Intervall $-1 \leqq x \leqq 1$ dargestellt.

§ 8. Polynome ersten Grades (lineare Funktionen)

Unter einer *linearen Funktion* versteht man allgemein eine Funktion der Form

$$y = ax + b \qquad (a \neq 0). \tag{6}$$

Der Gleichung kann man entnehmen, daß man ihre graphische Darstellung aus der graphischen Darstellung der Gleichung $y = x$ durch Dehnung in Richtung der y-Achse mit dem Dehnungsfaktor $|a|$ (im Fall $a < 0$ mit nachfolgender Spiegelung an der x-Achse) und anschließender Verschiebung um die Strecke b parallel zur y-Achse erhalten kann. In der Tat: Wir betrachten nacheinander die graphischen Darstellungen der Gleichungen

1. $y = x$,
2. $y = |a|\,x$,
3. $y = ax$,
4. $y = ax + b$.

Die graphische Darstellung der Gleichung 1. ist uns bereits bekannt; es ist die Winkelhalbierende des ersten und dritten Quadranten. Der Gleichung 2. kann man auch die Form $\frac{y}{|a|} = x$ geben, woraus zu ersehen ist (vgl. § 3, III b), daß

ihre graphische Darstellung aus der graphischen Darstellung von 1. durch Dehnung mit dem Dehnungsfaktor $|a|$ erhalten werden kann. Die graphische Darstellung von 3. unterscheidet sich im Fall $a > 0$ nicht von der graphischen Darstellung von 2.; im Fall $a < 0$ kann man der Gleichung 3. die Form $y = -|a|\,x$ oder $(-y) = |a|\,x$ geben, woraus zu ersehen ist, daß in diesem Fall die graphische Darstellung von 3. aus der graphischen Darstellung von 2. durch Spiegelung an der x-Achse erhalten werden kann. Schließlich sieht man leicht ein (§ 3, II b), daß man die graphische Darstellung von 4. aus der von 3. durch Parallelverschiebung in Richtung der y-Achse um die Strecke b erhält.[1])

Entsprechend entnimmt man der Darstellung

$$y = a\left(x + \frac{b}{a}\right)$$

für (6), daß man die graphische Darstellung von (6) auch durch Verschiebung der graphischen Darstellung von $y = x$ parallel zur x-Achse um die Strecke $\left(-\frac{b}{a}\right)$ und nachfolgende Dehnung in Richtung der y-Achse mit dem Dehnungsfaktor $|a|$ (und im Fall $a < 0$ mit nachfolgender Spiegelung an der x-Achse) erhalten kann.

So, aber auch auf andere Weise, kann man sich klar machen, daß das graphische Bild jeder linearen Funktion eine geneigte (d. h. zu keiner Koordinatenachse parallele) Gerade ist. Der Koeffizient a in der Gleichung (6) ist ein Maß für die *Steigung* (oder den *Anstieg*) der durch (6) dargestellten Geraden; er wird deshalb auch *Steigungskoeffizient* genannt. Der Steigungskoeffizient ist der Tangens des Winkels, den die Gerade mit der positiven Richtung der x-Achse einschließt. Im Fall $a = 1$ ist dies klar, da in diesem Fall der Winkel 45° beträgt. Im Fall $a \neq 1$ ergibt sich diese Behauptung daraus, daß bei Dehnung in Richtung der y-Achse mit dem Dehnungskoeffizienten $|a|$ sich der Tangens im gleichen Maße ändert, während bei einer evtl. Spiegelung an der x-Achse der Winkel wie auch der Tangens ihr Vorzeichen ändern.

Das absolute Glied b ist der Wert der Funktion (6) für $x = 0$, d. h. die Strecke, welche die Gerade auf der y-Achse „abschneidet" oder genauer: die Ordinate des Schnittpunktes der Geraden mit der y-Achse.

Was den Schnittpunkt der Geraden mit der x-Achse betrifft, so ist seine Abszisse gleich $-\frac{b}{a}$.

Die Funktion (6) ist, wie man leicht sieht, im Fall $a > 0$ wachsend und im Fall $a < 0$ fallend.

Im Fall $a = 0$ entartet die Funktion (6) in eine Konstante (ein Polynom vom Grad Null); ihr graphisches Bild ist eine zur x-Achse parallele Gerade.

---

[1]) Im folgenden werden Überlegungen dieser Art nicht mehr im einzelnen ausgeführt, sondern mit Hinweisen auf die entsprechenden Sätze in § 3 abgetan.

§ 9. Polynome zweiten Grades

Ein *Polynom zweiten Grades* ist ein Ausdruck der Form

$$y = ax^2 + bx + c \qquad (a \neq 0). \tag{7}$$

Wir nehmen auf der rechten Seite die folgenden Umformungen vor: Zunächst ziehen wir a „vor die Klammer" und spalten dann ein vollständiges Quadrat ab:

$$y = a\left[\left(x + \frac{b}{2a}\right)^2 + \frac{4ac - b^2}{4a^2}\right]. \tag{8}$$

Die Gleichung (8) zeigt, daß man die graphische Darstellung der Gleichung (7) aus der Parabel $y = x^2$ durch Hintereinanderausführung folgender Transformationen erhält:

1. Eine Verschiebung parallel zur x-Achse um die Strecke $\left(-\frac{b}{2a}\right)$.

2. Eine Verschiebung parallel zur y-Achse um die Strecke $\frac{4ac - b^2}{4a^2}$.

3. Eine Dehnung in Richtung der y-Achse mit dem Dehnungsfaktor $|a|$ (und im Fall $a < 0$ anschließend eine Spiegelung an der x-Achse).

Die Gleichung (7) kann man aber auch in der Form

$$y = a\left(x + \frac{b}{2a}\right)^2 + \frac{4ac - b^2}{4a} \tag{9}$$

schreiben, so daß auch eine andere Folge von Transformationen möglich ist:

1. Eine Verschiebung parallel zur x-Achse um die Strecke $\left(-\frac{b}{2a}\right)$.

2. Eine Dehnung in Richtung der y-Achse (mit evtl. nachfolgender Spiegelung an der x-Achse) mit dem Dehnungsfaktor $|a|$.

3. Eine Verschiebung parallel zur y-Achse um die Strecke $\frac{4ac - b^2}{4a}$.

Die graphische Darstellung eines Polynoms zweiten Grades ist also eine Parabel, die kongruent ist zur Parabel, die man aus der Normalparabel $y = x^2$ durch Dehnung in Richtung der y-Achse mit dem Dehnungsfaktor $|a|$ erhält.

Unabhängig von den beschriebenen Transformationen ist aus den Gleichungen (8) und (9) ersichtlich (vgl. § 5), daß im Fall $a > 0$ die Funktion $ax^2 + bx + c$ für $x < -\frac{b}{2a}$ fällt und für $x > -\frac{b}{2a}$ wächst und daß sie im Punkt $x = -\frac{b}{2a}$ ein Minimum mit dem Wert $\frac{4ac - b^2}{4a}$ annimmt; im Fall $a < 0$ wächst sie dagegen für $x < -\frac{b}{2a}$ und fällt für $x > -\frac{b}{2a}$, während sie im Punkt $x = -\frac{b}{2a}$ ein Maximum mit dem Wert $\frac{4ac - b^2}{4a}$ annimmt.

Ist $b \neq 0$, so wird (bei der Transformation 1.) die Parabel nach links verschoben, wenn a und b gleiches Vorzeichen besitzen, und nach rechts, wenn die Vorzeichen verschieden sind. Ist $b = 0$, so fällt die Horizontalverschiebung fort.

Die Vertikalverschiebung fällt weg, wenn die *Diskriminante*

$$D = 4ac - b^2$$

verschwindet. Die Gleichungen (8) und (9) zeigen, daß dann die Funktion $ax^2 + bx + c$ für alle Werte von x mit Ausnahme von $x = -\frac{b}{2a}$ dasselbe Vorzeichen besitzt, und zwar gerade das Vorzeichen von a, während sie für $x = -\frac{b}{2a}$ der Wert Null annimmt, d. h., die Parabel berührt in diesem Fall die x-Achse.

Ist D positiv, so ist in (8) die Summe in den eckigen Klammern für alle Werte von x positiv, d. h., die Funktion $ax^2 + bx + c$ besitzt für alle Werte von x dasselbe Vorzeichen wie a, d. h., die Parabel hat mit der x-Achse keinen Punkt gemeinsam.

Ist D negativ, so kann man den Ausdruck in den eckigen Klammern in (8) als eine Differenz von Quadraten ansehen und, wenn man

$$x_1 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}, \quad x_2 = \frac{-b + \sqrt{b^2 - 4ac}}{2a}$$

setzt, das Polynom als Produkt

$$ax^2 + bx + c = a(x - x_1)(x - x_2) \qquad (x_1 < x_2)$$

darstellen. Dieser Gleichung kann man entnehmen, daß die Funktion $ax^2 + bx + c$ für $x = x_1$ und $x = x_2$ verschwindet, d. h., daß in diesen beiden Punkten die Parabel die x-Achse schneidet. Zum Vorzeichen der Funktionswerte ist zu sagen, daß es nicht überall dasselbe ist; im Fall $a > 0$ sind die Funktionswerte positiv für $x < x_1$ und $x > x_2$ und negativ für $x_1 < x < x_2$, während im Fall $a < 0$ die Verhältnisse genau umgekehrt liegen. Auch dies läßt sich natürlich leicht geometrisch deuten.

§ 10. Polynome dritten Grades

Ein *Polynom dritten Grades* ist ein Ausdruck der Form

$$y = ax^3 + bx^2 + cx + d \qquad (a \neq 0). \tag{10}$$

Ziehen wir hier a vor die Klammer und spalten wir einen vollständigen Kubus ab, so erhalten wir

$$\begin{aligned} y &= a\left[\left(x + \frac{b}{3a}\right)^3 + \left(\frac{c}{a} - \frac{1}{3}\,\frac{b^2}{a^2}\right)x + \left(\frac{d}{a} - \frac{1}{27}\,\frac{b^3}{a^3}\right)\right] \\ &= a\left[\left(x + \frac{b}{3a}\right)^3 + \left(\frac{c}{a} - \frac{1}{3}\,\frac{b^2}{a^2}\right)\left(x + \frac{b}{3a}\right) + \left(\frac{d}{a} - \frac{1}{3}\,\frac{bc}{a^2} + \frac{2}{27}\,\frac{b^3}{a^3}\right)\right] \qquad (11) \\ &= a\left[\left(x + \frac{b}{3a}\right)^3 + D\left(x + \frac{b}{3a}\right) + D'\right], \end{aligned}$$

wobei zur Abkürzung

$$D = \frac{c}{a} - \frac{1}{3}\,\frac{b^2}{a^2},$$

$$D' = \frac{d}{a} - \frac{1}{3}\,\frac{bc}{a^2} + \frac{2}{27}\,\frac{b^3}{a^3}$$

gesetzt ist. Ist $D \neq 0$, so kann man (11) noch folgendermaßen umformen:

$$y = a\,|D|^{3/2}\left\{\left(\frac{x+\frac{b}{3a}}{\sqrt{|D|}}\right)^3 \pm \left(\frac{x+\frac{b}{3a}}{\sqrt{|D|}}\right) + \frac{D'}{|D|^{3/2}}\right\}, \tag{12}$$

wobei vor dem zweiten Glied das Plus- bzw. Minuszeichen steht, je nachdem, ob $D > 0$ oder $D < 0$. Ist dagegen $D = 0$, so nimmt (11) die Form

$$y = a\left\{\left(x+\frac{b}{3a}\right)^3 + D'\right\} \tag{13}$$

an.

Den Gleichungen (12) bzw. (13) kann man entnehmen, daß das graphische Bild jeder Funktion (10) aus dem graphischen Bild eines der folgenden einfacheren Polynome durch elementare Transformationen (vgl. § 3, I bis III) gewonnen werden kann:

$$(\alpha)\quad y = x^3 + x,$$

$$(\beta)\quad y = x^3 - x,$$

$$(\gamma)\quad y = x^3.$$

Ist nämlich $D \neq 0$, so erhält man das graphische Bild von (10) aus dem graphischen Bild von (α) bzw. (β) (je nachdem, ob $D > 0$ oder $D < 0$) durch folgende Transformationen:

1. Eine Dehnung in Richtung der x-Achse mit dem Dehnungsfaktor $\sqrt{|D|}$.
2. Eine Parallelverschiebung längs der x-Achse um die Strecke $\left(-\frac{b}{3a}\right)$.
3. Eine Parallelverschiebung längs der y-Achse um die Strecke $\frac{D'}{|D|^{3/2}}$.
4. Eine Dehnung in Richtung der y-Achse mit dem Dehnungsfaktor $|a| \cdot |D|^{3/2}$ (und eine Spiegelung an der x-Achse im Fall $a < 0$).

Im Fall $D = 0$ erhält man das graphische Bild von (10) aus dem von (γ) durch

1. Eine Parallelverschiebung längs der x-Achse um die Strecke $\left(-\frac{b}{3a}\right)$.
2. Eine Parallelverschiebung längs der y-Achse um die Strecke D'.
3. Eine Dehnung in Richtung der y-Achse mit dem Dehnungsfaktor $|a|$ (und eine Spiegelung an der x-Achse im Fall $a < 0$).

Um also Klarheit über den Verlauf und die Lage des graphischen Bildes einer beliebigen Funktion der Form (10) zu erlangen und diese zu diskutieren, braucht man nur die graphischen Bilder der Funktionen (α), (β) und (γ) zu untersuchen.

Was die Untersuchung der Funktion (γ) betrifft, so können wir auf § 7 verweisen.

Die Funktionen (α) und (β) sind beide ungerade, d. h., ihre graphischen Bilder (Abb. 15) liegen symmetrisch in bezug auf den Ursprung. Die Funktion (α) ist für $x > 0$ wachsend (vgl. § 5); ihre graphische Darstellung liegt für $x > 0$ offenbar oberhalb der Geraden $x = y$ und schmiegt sich in der Umgebung von O eng an diese an (sie berührt die Gerade $x = y$ im Punkt O).

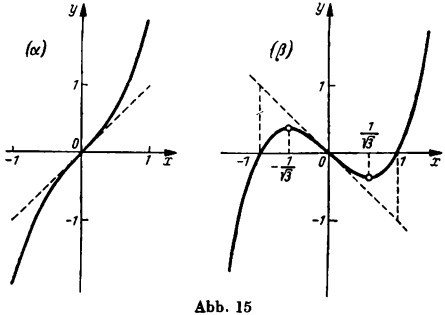


Abb. 15

Der Verlauf der graphischen Darstellung von (β) ist komplizierter. Die Faktorzerlegung

$$x^3 - x = x(x-1)(x+1)$$

zeigt, daß das Bild von (β) die x-Achse in den Punkten $x = 0$, $x = 1$ und $x = -1$ schneidet. Für das weitere beschränken wir uns auf positive Werte von x. Ersichtlich ist $x^3 - x > 0$ für $x > 1$ und $x^3 - x < 0$ für $0 < x < 1$. Ferner ist unmittelbar klar, daß (für $x > 0$) das graphische Bild von (β) oberhalb der Geraden $y = -x$ liegt und diese im Punkt O berührt. Der Identität

$$x^3 - x = -\frac{2}{3\sqrt{3}} + \left(x - \frac{1}{\sqrt{3}}\right)^2 \left(x + \frac{2}{\sqrt{3}}\right) \tag{14}$$

kann man entnehmen, daß die Funktion (β) für $x = \frac{1}{\sqrt{3}}$ ein Minimum mit dem Wert $\left(-\frac{2}{3\sqrt{3}}\right)$ annimmt. Aus (14) ergibt sich ferner, daß die Funktion (β) für $x > \frac{1}{\sqrt{3}}$ wächst (s. § 5).

Ebenfalls kann man, wenn man direkt auf die Definition zurückgeht, leicht elementar zeigen, daß die Funktion $f(x) = x^3 - x$ für $0 < x < \frac{1}{\sqrt{3}}$ fällt. Sind nämlich

$$x' = x$$

und

$$x'' = x + h$$

Zahlen aus diesem Intervall, die der Bedingung

$$0 \leqq x' < x'' \leqq \frac{1}{\sqrt{3}}$$

genügen, so ist

$$0 < h \leqq \frac{1}{\sqrt{3}},$$

$$0 \leqq x \leqq \frac{1}{\sqrt{3}} - h$$

und

$$3x^2 + 3hx + h^2 \leqq 3\left(\frac{1}{\sqrt{3}} - h\right)^2 + 3h\left(\frac{1}{\sqrt{3}} - h\right) + h^2$$

$$= 1 - h\sqrt{3} + h^2 = 1 - h(\sqrt{3} - h) < 1,$$

so daß in

$$f(x'') - f(x') = f(x+h) - f(x)$$

$$= [(x+h)^3 - (x+h)] - (x^3 - x) = -h\{1 - (3x^2 + 3hx + h^2)\}$$

der Ausdruck in den geschweiften Klammern positiv und daher die Differenz $f(x'') - f(x')$ negativ ist.[1])

§ 11. Biquadratische Polynome

Von den Polynomen vierten Grades wollen wir hier nur die sogenannten *biquadratischen* Polynome, d. h. die Polynome der Form

$$y = ax^4 + bx^2 + c \qquad (a \neq 0) \tag{15}$$

näher untersuchen.

Alle diese Polynome sind gerade Funktionen, d. h., ihre graphischen Bilder liegen symmetrisch in bezug auf die y-Achse. Die Umformung von (15) zu

$$y = \frac{b^2}{a}\left\{\left(\frac{x}{\sqrt{\left|\frac{b}{a}\right|}}\right)^4 \pm \left(\frac{x}{\sqrt{\left|\frac{b}{a}\right|}}\right)^2\right\} + c \tag{16}$$

läßt erkennen, daß man im Fall $b \neq 0$ die graphische Darstellung des Polynoms (15) aus der von

$$(\alpha) \quad y = x^4 + x^2$$

bzw.

$$(\beta) \quad y = x^4 - x^2$$

[1]) Es ist zunächst nicht ersichtlich, wie man auf die Identität (14) kommt. Das läßt die von uns verwendete elementare Methode sehr künstlich erscheinen. In diesem Zusammenhang ist es nützlich, sich zu vergegenwärtigen, daß die Differentialrechnung eine natürliche und einheitliche Methode zur Bestimmung von Maxima und Minima liefert und auch das Studium der „Wachstumsverhältnisse" einer Funktion wesentlich erleichtert.

durch passende Dehnungen in Richtung beider Achsen (mit evtl. Spiegelung an der x-Achse) und anschließende Verschiebung parallel zur y-Achse erhalten kann, während man im Fall $b = 0$ von einer graphischen Darstellung von

$$(\gamma) \quad y = x^4$$

ausgehen muß.

Da die Funktionen x^2 und x^4 positiv und für $x > 0$ wachsend sind und ihre graphischen Bilder im Punkt O die x-Achse berühren, besitzt auch ihre Summe, die Funktion (α), diese Eigenschaften (Abb. 16).

Was die Funktion (β) betrifft, so kann man zunächst der Gleichung

$$x^4 - x^2 = x^2(x-1)(x+1)$$

unmittelbar entnehmen, daß sie für $x = 0$ und $x = \pm 1$ verschwindet (im Punkt O berührt die graphische Darstellung von (β) die x-Achse; vgl. Abb. 16).

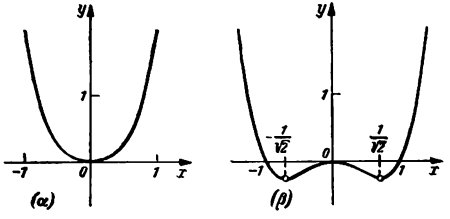


Abb. 16

Nun kann man die Funktion (β) aber auch in der Form $y = \varphi(x^2)$ darstellen, wobei $\varphi(u) = u^2 - u$ ist. Wächst hier u von 0 bis $\frac{1}{2}$, so fällt die Funktion $\varphi(u)$ von 0 bis $-\frac{1}{4}$, während für $\frac{1}{2} < u < \infty$ die Funktion $\varphi(u)$ von $-\frac{1}{4}$ bis ∞ wächst;[1]) wächst also $x\left(=\sqrt{u}\right)$ von 0 bis $\frac{1}{\sqrt{2}}$, so fällt die Funktion $y = \varphi(x^2)$ von 0 bis $-\frac{1}{4}$, während sie für $\frac{1}{\sqrt{2}} < x < \infty$ von $-\frac{1}{4}$ bis ∞ wächst (vgl. § 5).

§ 12. Polynome höheren Grades

Die Hauptschwierigkeit bei der Untersuchung, wie sich die Vorzeichen der Werte eines Polynoms

$$y = ax^n + bx^{n-1} + \cdots + kx + l \qquad (a \neq 0;\ n \geqq 4) \tag{17}$$

[1]) Es ist nämlich (vgl. § 9)

$$u^2 - u = \left(u - \frac{1}{2}\right)^2 - \frac{1}{4}.$$

verhalten, besteht darin, es in Faktoren zu zerlegen, was seinerseits darauf hinausläuft, alle (zumindest aber die reellen) Wurzeln der Gleichung

$$ax^n + bx^{n-1} + \cdots + kx + l = 0 \tag{18}$$

zu bestimmen. Nehmen wir an (und damit schränken wir die Allgemeinheit der Überlegungen nur wenig ein), der höchste Koeffizient a sei gleich 1. Aus der Algebra[1]) ist bekannt, daß das betrachtete Polynom nach Zusammenfassung der paarweise konjugiert komplexen Faktoren als Produkt der Form

$$y = (x-\alpha)^\lambda (x-\beta)^\mu (x-\gamma)^\nu \cdots (x^2+px+q)^\varrho \cdots \tag{19}$$

dargestellt werden kann, wobei die Zahlen $\alpha, \beta, \ldots$ die reellen (paarweise verschiedenen) Nullstellen des Polynoms und $\lambda, \mu, \ldots$ ihre Vielfachheiten sind, während die quadratischen Polynome $x^2+px+q, \ldots$ nur konjugiert komplexe Nullstellen besitzen. Da die Polynome $x^2+px+q, \ldots$ stets positiv sind, hängt das Vorzeichen der Funktionswerte y nur von den Vorzeichen der Faktoren $(x-\alpha)^\lambda, \ldots$ ab. Um etwas bestimmtes vor Augen zu haben, wollen wir annehmen, daß die Nullstellen $\alpha, \beta, \ldots$ der Größe nach angeordnet sind, etwa $\alpha > \beta > \ldots$. Wir untersuchen nun, was geschieht, wenn bei kleiner werdenden Argumentwerten x durch die Werte $\alpha, \beta, \ldots$ hindurchgeht. Solange x größer als α ist, sind alle Faktoren des Produktes und damit auch das Polynom selbst positiv. Hat x den Wert α durchlaufen, so ist die Differenz $x-\alpha$ negativ; ob der Faktor $(x-\alpha)^\lambda$ positiv oder negativ ist, hängt davon ab, ob der Exponent λ (d. h. die Vielfachheit von α) eine gerade oder ungerade Zahl ist. Entsprechendes gilt, wenn x den Wert β durchläuft, usw.

Aus diesen Überlegungen ergibt sich das folgende Resultat für die graphische Darstellung des Polynoms (17): Das graphische Bild des Polynoms (17) liegt für $x > \alpha$ oberhalb der x-Achse; für $\beta < x < \alpha$ liegt es oberhalb bzw. unterhalb der x-Achse, je nachdem, ob die Vielfachheit λ von α gerade oder ungerade ist; für $\gamma < x < \beta$ liegt es oberhalb bzw. unterhalb der x-Achse, je nachdem, ob $\lambda + \mu$ gerade oder ungerade ist, usw. Die Summe der Grade aller Faktoren in (19), d. h. $\lambda + \mu + \nu + \cdots + 2\varrho + \cdots$, ist gleich dem Grad n des gegebenen Polynoms. Daraus folgt, daß für gerades n das graphische Bild von (17) links von der kleinsten Wurzel (wie auch rechts von der größten Wurzel) oberhalb der x-Achse liegt, während es für ungerades n links von der kleinsten Wurzel unterhalb der x-Achse liegt. Alles dies gilt auch dann, wenn der höchste Koeffizient a eine beliebige positive Zahl ist; ist dagegen a negativ, so kehren sich alle Vorzeichen um.

Wir bemerken noch, daß in den Punkten, welche einfachen Nullstellen (d. h. Nullstellen der Vielfachheit Eins) des gegebenen Polynoms entsprechen, das graphische Bild die x-Achse schneidet, ohne sie zu berühren; in Punkten, welche Nullstellen einer geraden Vielfachheit entsprechen, liegt eine Berührung ohne Schnitt vor; eine Berührung mit Schnitt (wie z. B. bei der Funktion $y = x^3$) erfolgt in den Punkten, die Nullstellen ungerader Viel-

[1]) Vgl. EdEM, Bd. II, L. J. Okunjew, Der Ring der Polynome und der Körper der rationalen Funktionen, § 6. — *Anm. d. wiss. Red.*

fachheit $\lambda \geqq 3$ entsprechen. Alles dies läßt sich streng mit Methoden der Differentialrechnung beweisen.

Das Wachstumsverhalten eines Polynoms höheren Grades läßt sich mit elementaren Methoden nur in Spezialfällen ermitteln. Die Darstellung

$$y = a x^n \left(1 + \frac{b}{x} + \cdots + \frac{k}{x^{n-1}} + \frac{l}{x^n}\right) \tag{20}$$

zeigt lediglich, daß bei unbeschränktem Wachsen von $|x|$ auch der absolute Betrag der Funktionswerte y unbegrenzt wächst, und zwar um so schneller, je größer der Grad n ist.

Bei der Untersuchung von Polynomen sind gewisse Vereinfachungen dann möglich, wenn das Polynom eine gerade oder ungerade Funktion ist. Offenbar ist ein Polynom dann (und nur dann) eine gerade Funktion der Veränderlichen x, wenn es nur gerade Potenzen von x enthält, und eine ungerade Funktion, wenn es nur ungerade Potenzen von x enthält.

§ 13. Die Potenzfunktionen mit negativem ganzen Exponenten

Wir wollen jetzt die graphischen Bilder G_1, G_2, G_3, ..., G_n, ... der einfachsten gebrochenen rationalen Funktionen

$$\frac{1}{x}, \frac{1}{x^2}, \frac{1}{x^3}, \ldots, \frac{1}{x^n}, \ldots$$

untersuchen. Wir beginnen mit der Funktion

$$y = \frac{1}{x}, \tag{21}$$

die als Lösung der Gleichung

$$xy = 1 \tag{22}$$

zweiter Ordnung auftritt. Ihr graphisches Bild (Abb. 17a) kann man entsprechend § 2,6 punktweise aus der Geraden $x = y$ gewinnen. Es liegt symmetrisch in bezug auf den Ursprung O und in bezug auf die Gerade $x = y$ (vgl. § 4); der Punkt $x = 0$ ist dadurch ausgezeichnet, daß ihm kein y-Wert entspricht, d. h., auf der y-Achse liegt kein Punkt des graphischen Bildes von (21). Dieser Punkt ist eine „Unstetigkeitsstelle" der Funktion (vgl. Kap. III, § 42), in ihm „verliert die Funktion $y = \frac{1}{x}$ ihren Sinn". Es sei im folgenden $x > 0$. Da die Funktion $y = x$ wachsend ist, ist die Funktion $y = \frac{1}{x}$ als zu ihr reziproke Funktion fallend. Ihr graphisches Bild G_1 geht durch den Punkt mit den Koordinaten (1,1). Wächst x unbeschränkt, so wird der reziproke Wert $\frac{1}{x}$ beliebig klein, die graphische Darstellung nähert sich (wie man sagt, „asymptotisch") der x-Achse. Nähert sich x dem Wert 0, so wird der reziproke Wert $\frac{1}{x}$ beliebig groß, das graphische Bild geht unter („asymptotischer") Annäherung an die y-Achse „ins Unendliche". Wegen der Symmetrie zum

Ursprung besteht die Kurve aus zwei „Ästen" (oder „Zweigen"); der eine liegt im ersten, der andere im dritten Quadranten. Man nennt die Kurve eine *Hyperbel*. Die x- und die y-Achse sind ihre *Asymptoten*.[1])

Wir betrachten als nächstes die Funktion

$$y = \frac{1}{x^2}. \tag{23}$$

Ihre Funktionswerte sind reziprok zu den entsprechenden Funktionswerten der Funktion $y = x^2$. Eine Symmetrie zur Winkelhalbierenden $x = y$ liegt

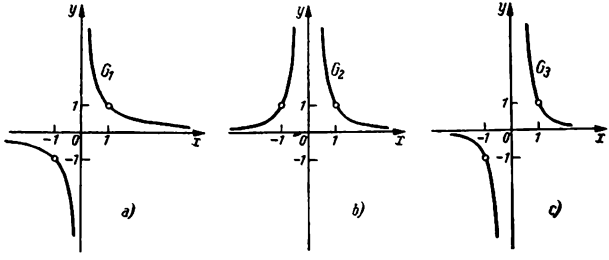


Abb. 17

beim graphischen Bild G_2 der Funktion (23) nicht vor. Für $x > 1$ ist $x^2 > x$ und daher $\frac{1}{x^2} < \frac{1}{x}$, so daß für $x > 1$ das graphische Bild G_2 unterhalb von G_1 liegt (sich schneller der x-Achse anschmiegt). Für $0 < x < 1$ ist dagegen $x^2 < x$ und daher $\frac{1}{x^2} > \frac{1}{x}$, so daß für $0 < x < 1$ das Bild von (23) oberhalb G_1 liegt (schneller anwächst). An die Stelle der Symmetrie von G_1 zum Ursprung tritt bei G_2 eine Symmetrie zur y-Achse; auch das graphische Bild G_2 besteht aus zwei „Ästen", von denen einer im ersten und der andere im zweiten Quadranten liegt (Abb. 17b).

Eine Untersuchung der weiteren Funktionen $y = \frac{1}{x^n}$ und ihrer graphischen Bilder G_n ($n = 3, 4, \ldots$; G_3 ist in Abb. 17c dargestellt) zeigt, daß alle diese Funktionen, wie auch die beiden bereits betrachteten, für $x = 0$ „ihren Sinn verlieren", ferner gerade oder ungerade sind, je nachdem, ob n gerade oder ungerade ist, und für $x > 0$ bei unbeschränktem Wachsen von x beliebig klein und bei Annäherung von x gegen Null beliebig groß werden. Wegen

$$x < x^2 < x^3 < \cdots \text{ für } x > 1$$

und

$$x > x^2 > x^3 > \cdots \text{ für } 0 < x < 1$$

[1]) Vgl. etwa O.-H. Keller, Analytische Geometrie und lineare Algebra, Berlin 1957

gilt

$$\frac{1}{x} > \frac{1}{x^2} > \frac{1}{x^3} > \cdots \text{ für } x > 1$$

und

$$\frac{1}{x} < \frac{1}{x^2} < \frac{1}{x^3} < \cdots \text{ für } 0 < x < 1.$$

Geometrisch bedeutet dies, daß für jedes n im ersten Quadranten für $x > 1$ die Kurve G_{n+1} unter der Kurve G_n und für $0 < x < 1$ die Kurve G_{n+1} über der

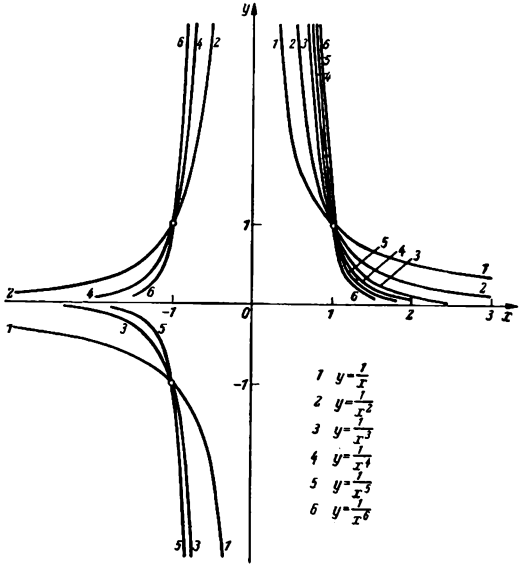


Abb. 18

Kurve G_n liegt. Alle Kurven nähern sich asymptotisch der x- und der y-Achse (Abb. 18). Jede der Kurven G_n besitzt noch einen zweiten Ast, der, je nachdem, ob n ungerade oder gerade ist, zum ersten in bezug auf den Ursprung O oder die y-Achse symmetrisch ist und im dritten oder zweiten Quadranten liegt. Die Kurven G_n ($n = 2, 3, 4, \ldots$) nennt man mitunter auch *Hyperbeln höherer Ordnung*.

§ 14. Gebrochene lineare Funktionen

Im vorliegenden Paragraphen betrachten wir Funktionen

$$y = \frac{ax+b}{cx+d}, \tag{24}$$

wobei wir annehmen können, daß der Bruch auf der rechten Seite nicht gekürzt ($ad - bc \neq 0$) und die Funktion nicht auf eine ganze lineare Funktion zurückgeführt werden kann ($c \neq 0$). Da man (24) offenbar auch in der Form

$$y = \frac{a}{c} - \frac{ad-bc}{c^2\left(x+\frac{d}{c}\right)} \tag{25}$$

schreiben kann, läßt sich das graphische Bild der Funktion (24) aus dem von $\frac{1}{x}$ durch folgende elementare Transformationen erhalten:

1. Eine Dehnung in Richtung der y-Achse mit dem Dehnungsfaktor $\left|\frac{ad-bc}{c^2}\right|$ (und eine Spiegelung an der x-Achse im Fall $ad - bc > 0$).

2. Eine Parallelverschiebung in Richtung der x-Achse um die Strecke $\left(-\frac{d}{c}\right)$.

3. Eine Parallelverschiebung in Richtung der y-Achse um die Strecke $\frac{a}{c}$.

Hieraus folgt, daß die Gleichung (24) ebenfalls eine (gewöhnliche) Hyperbel darstellt. Im Punkt $x = -\frac{d}{c}$ verliert die Funktion (24) ihren Sinn (sie hat dort keinen Wert, sie „existiert dort nicht"). In den Intervallen $\left(-\infty, -\frac{d}{c}\right)$ und $\left(-\frac{d}{c}, +\infty\right)$ wächst (im Fall $ad - bc > 0$) bzw. fällt (im Fall $ad - bc < 0$) sie ständig.[1]) Bei unbegrenztem Wachsen von $|x|$ kommen die Funktionswerte dem Wert $\frac{a}{c}$ beliebig nahe, wie man unmittelbar aus (25) oder noch einfacher aus

$$\frac{ax+b}{cx+d} = \frac{a+\frac{b}{x}}{c+\frac{d}{x}}$$

ersieht. Der Punkt mit dem Koordinaten $\left(-\frac{d}{c}, \frac{a}{c}\right)$ (der nicht auf der Kurve liegt) ist der *Mittelpunkt* (das Symmetriezentrum) der Hyperbel; die Geraden $x = -\frac{d}{c}$ und $y = \frac{a}{c}$ sind ihre Asymptoten.

---

[1]) Wir sagen hierfür auch kurz: Die Funktion wächst (oder fällt) überall mit Ausnahme der „Unstetigkeitsstelle".

§ 15. Gebrochene rationale Funktionen zweiten Grades

Auch das Wachstumsverhalten von Funktionen

$$y = \frac{ax^2 + bx + c}{dx^2 + ex + f} \qquad (a^2 + d^2 \neq 0)^{1)} \tag{26}$$

mit teilerfremdem Zähler und Nenner läßt sich mit elementaren Methoden beschreiben, obgleich die dabei verwendeten „identischen Umformungen" teilweise mehr oder weniger künstlich sind. Ist $d \neq 0$, so kann man die rechte Seite von (26) in der Form

$$\frac{ax^2 + bx + c}{dx^2 + ex + f} = \frac{a}{d} + \frac{1}{\dfrac{d(dx^2 + ex + f)}{(bd - ae)x + (cd - af)}} \tag{27}$$

darstellen (die Ausdrücke $bd - ae$ und $cd - af$ können nicht beide gleich Null sein, da sonst in (26) nicht Zähler und Nenner teilerfremd wären). Damit ist die Untersuchung der Funktion (26) auf die Untersuchung einer gebrochenen rationalen Funktion zurückgeführt, deren Zähler ein Polynom zweiten Grades und deren Nenner ein Polynom ersten Grades ist. Im Fall $d = 0$ ist (26) bereits von vornherein eine rationale Funktion dieser spezielleren Art.

Wir betrachten also eine Funktion

$$y = \frac{mx^2 + nx + p}{rx + s}, \tag{28}$$

wobei wir annehmen, daß $m \neq 0$ und $r \neq 0$. Durch diese Nebenbedingung werden lediglich bereits früher betrachtete Fälle ausgeschlossen (§ 14 und § 9). Setzen wir

$$\alpha = -\frac{s}{r},$$

so erhalten wir, wenn wir den Zähler von (28) nach Potenzen von $x - \alpha$ ordnen,

$$y = \frac{m(x-\alpha)^2 + (2m\alpha + n)(x-\alpha) + (m\alpha^2 + n\alpha + p)}{r(x-\alpha)},$$

d. h.

$$y = \frac{m}{r}\left\{(x-\alpha) + \left(2\alpha + \frac{n}{m}\right) + \frac{\alpha^2 + \dfrac{n}{m}\alpha + \dfrac{p}{m}}{x-\alpha}\right\}. \tag{29}$$

Die Zahl $A = \alpha^2 + \frac{n}{m}\alpha + \frac{p}{m}$ ist von Null verschieden[2)], so daß die Fälle

$$\text{I.} \quad A < 0,$$

$$\text{II.} \quad A > 0$$

alle Möglichkeiten erschöpfen.

[1)] Die Beziehung $a^2 + d^2 \neq 0$ bringt offenbar zum Ausdruck, daß mindestens eine der Zahlen a oder b von Null verschieden ist.

[2)] Da α Nullstelle des Nenners von (28) ist, hätte das Verschwinden von A zur Folge, daß α auch Nullstelle des Zählers ist, d. h., daß Zähler und Nenner in (28) nicht teilerfremd sind.

Im ersten Fall können wir (29) auf die Form

$$y=\frac{2m\alpha+n}{r}+\frac{m\sqrt{|A|}}{r}\left\{\frac{x-\alpha}{\sqrt{|A|}}-\frac{\sqrt{|A|}}{x-\alpha}\right\}$$

bringen. Hieraus ist zu ersehen, daß die graphische Darstellung der Funktion (28) in diesem Fall aus der graphischen Darstellung der Funktion

$$y=x-\frac{1}{x} \tag{I}$$

durch Anwendung folgender Transformationen erhalten werden kann:

1. Dehnung in Richtung der x-Achse mit dem Dehnungsfaktor $\sqrt{|A|}$;
2. Dehnung in Richtung der y-Achse mit dem Dehnungsfaktor $\frac{m\sqrt{|A|}}{r}$;
3. Verschiebung parallel zur x-Achse um die Strecke $\alpha=-\frac{s}{r}$;
4. Verschiebung parallel zur y-Achse um die Strecke $\frac{2m\alpha+n}{r}$.

Im zweiten Fall können wir, wenn wir beachten, daß dann $|A|=A$ ist, der Gleichung (29) die Form

$$y=\frac{2m\alpha+n}{r}+\frac{m\sqrt{A}}{r}\left\{\frac{x-\alpha}{\sqrt{A}}+\frac{\sqrt{A}}{x-\alpha}\right\}$$

geben. Hieraus ist zu ersehen, daß im Fall II die graphische Darstellung von (28) durch dieselbe Transformation wie im zuvor behandelten Fall I aus der graphischen Darstellung von

$$y=x+\frac{1}{x} \tag{II}$$

gewonnen werden kann.

Es genügt also, die graphischen Bilder der Funktionen (I) und (II) näher zu untersuchen.

Der Gleichung (I) kann man offenbar auch die Form

$$y=\frac{x^2-1}{x}$$

oder

$$y=\frac{(x-1)(x+1)}{x} \tag{30}$$

geben. Aus (30) ist zu ersehen, daß die Funktion (I) genau an den Stellen $x=1$ und $x=-1$ verschwindet und an der Stelle $x=0$ ihren Sinn verliert (eine „Unstetigkeitsstelle" hat) (Abb. 19a). Ihr Vorzeichenverhalten läßt sich unmittelbar übersehen. Weil die Funktion x überall und die Funktion $-\frac{1}{x}$ überall mit Ausnahme der Unstetigkeitsstelle $x=0$ wächst, ist (vgl. § 5, 3) auch die Funktion (I) mit Ausnahme der Unstetigkeitsstelle überall wachsend.

Auch die Funktion (II) hat eine Unstetigkeitsstelle im Punkt $x=0$; im Gegensatz zur Funktion (I) nimmt sie aber nirgends den Wert Null an (Abb. 19b). Um ihr Wachstumsverhalten für $x>0$ zu ermitteln, geben wir (II) die Form

$$y=2+\left(\sqrt{x}-\frac{1}{\sqrt{x}}\right)^2. \tag{31}$$

Diese Schreibweise läßt erkennen, daß die Funktion (II) im Punkt $x=1$ ein Minimum mit dem Wert $y=2$ annimmt. Da die Funktionen $\sqrt{x}$ und $-\frac{1}{\sqrt{x}}$ im Intervall $1<x<\infty$ wachsend sind (§ 5; 3, 6, 7, 8), wächst dort auch

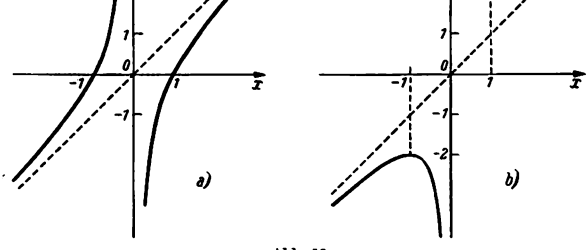


Abb. 19

ihre Summe $\sqrt{x}-\frac{1}{\sqrt{x}}$, und wegen $\sqrt{x}-\frac{1}{\sqrt{x}}>0$ sind dort auch die Funktion $\left(\sqrt{x}-\frac{1}{\sqrt{x}}\right)^2$ und schließlich die Funktion (31), d. h. die Funktion $x+\frac{1}{x}$ wachsend. Entsprechend sieht man, daß die Funktion (II) im Intervall $0<x<1$ fällt; hierzu braucht man nur der Gleichung (II) die Form

$$y=2+\left(\frac{1}{\sqrt{x}}-\sqrt{x}\right)^2 \tag{32}$$

zu geben.

Eine eingehende Untersuchung der Funktionen (I) und (II) für negative Werte von x erübrigt sich, weil beide Funktionen ungerade sind, d. h. ihre graphischen Bilder zum Ursprung O symmetrisch liegen.

Zu bemerken ist noch, daß die beiden graphischen Darstellungen die Gerade $x=y$ als Asymptote besitzen (vgl. Abb. 19a und b). Dies ergibt sich daraus, daß die Differenz $\left(x\pm\frac{1}{x}\right)-x$ beliebig klein wird, wenn $|x|$ unbeschränkt wächst.

Wir haben bisher das Verhalten von Funktionen der Form (28), d. h. von Funktionen (26) mit $d = 0$ untersucht. Zur Klärung des allgemeinen Falls genügt es unter Hinweis auf (27) zu bemerken, daß man das graphische Bild von (26) mit $d \neq 0$ aus dem graphischen Bild einer geeigneten Funktion (28) (im Fall $bd - ae \neq 0$) bzw. einer Funktion

$$y = ax^2 + bx + c \qquad (a \neq 0)$$

(im Fall $bd - ae = 0$, $cd - af \neq 0$) durch folgende Transformationen erhält:

1. Übergang von der graphischen Darstellung einer Funktion zur graphischen Darstellung der reziproken Funktion (vgl. § 2);

2. Parallelverschiebung längs der y-Achse um die Strecke $\frac{a}{d}$.

Damit ist auf Grund von § 5, 1 bis 9, auch das Wachstumsverhalten der betrachteten Funktion leicht zu ermitteln. Ebenso bereitet die Bestimmung des Vorzeichenverhaltens und die Ermittlung der Punkte, in denen die Funktion verschwindet und in denen sie Unstetigkeitsstellen besitzt, keine Schwierigkeiten.

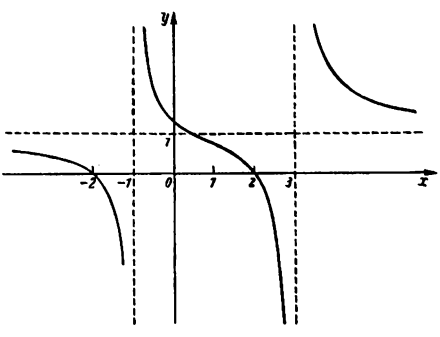


Abb. 20

Beispiel 1.

$$y = \frac{x^2 - 4}{x^2 - 2x - 3}. \quad (33)$$

Die Schreibweise

$$y = \frac{(x+2)(x-2)}{(x-3)(x+1)}$$

läßt erkennen, daß die graphische Darstellung der Funktion (33) die x-Achse in den Punkten $x = -2$ und $x = +2$ schneidet und in den Punkten $x = -1$ und $x = 3$ Unstetigkeiten besitzt; sie ermöglicht es auch, das Vorzeichenverhalten der Funktion zu ermitteln. Der Schreibweise

$$y = \frac{1 - \frac{4}{x^2}}{1 - \frac{2}{x} - \frac{3}{x^2}}$$

sieht man an, daß sich y dem Wert 1 nähert, wenn $|x|$ unbeschränkt wächst. Die Gerade $y = 1$ ist also Asymptote. Mit Hilfe von zwei bis drei Punktsubstitutionen kann man ein graphisches Bild entwerfen, wie es in Abb. 20 dargestellt ist.

Um sich jedoch davon zu überzeugen, daß die Funktion (33) mit Ausnahme der Unstetigkeitsstellen wirklich überall fällt, muß man auf die oben entwickelte Theorie zurückgreifen. Sie führt auf die folgende Darstellung der betrachteten Funktion:

$$y = 1 + \cfrac{2}{-1 + \left(x - \frac{1}{2}\right) + \cfrac{\left(-\frac{15}{4}\right)}{x - \frac{1}{2}}},$$

und man kann sich, um die erforderlichen Schlüsse zu ziehen, unmittelbar auf die entsprechenden Sätze aus § 5 berufen.

Beispiel 2.

$$y = \frac{x^2 + x + 1}{x^2 - x + 1}.$$

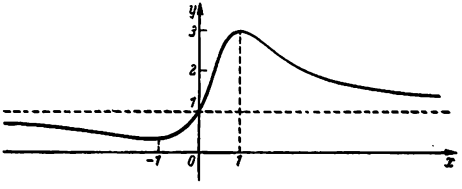


Abb. 21

Da der Zähler und der Nenner stets positiv sind, ist auch die Funktion überall positiv. Wie in Beispiel 1 ist auch hier die Gerade $y = 1$ Asymptote. Die dargelegte Theorie führt auf die Darstellung

$$y = 1 + \cfrac{2}{-1 + \left(x + \frac{1}{x}\right)}.$$

Da die Funktion $x + \frac{1}{x}$ für $|x| < 1$ (mit Ausnahme der Unstetigkeitsstelle) fällt und für $|x| > 1$ wächst, ist die betrachtete Funktion für $|x| < 1$ wachsend und für $|x| > 1$ fallend; Unstetigkeitsstellen treten nicht auf (Abb. 21).

Beispiel 3.

$$y = \frac{1}{x^2 + 1}.$$

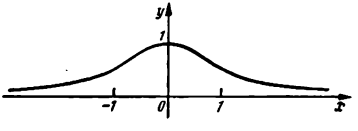


Abb. 22

Auch diese Funktion ist überall positiv. Ihr Wachstumsverhalten folgt unmittelbar aus den Sätzen des § 5 (Abb. 22).

Beispiel 4.

$$y = \frac{x^2 - 1}{x^2 + 1}.$$

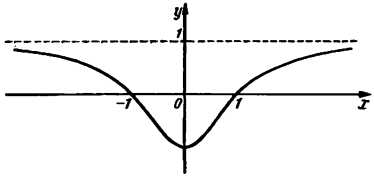


Abb. 23

Diese Funktion ändert ihr Vorzeichen in den Punkten $x = \pm 1$ (Abb. 23); zur Untersuchung ihres Wachstumsverhaltens genügt es, sie nach den Anweisungen der allgemeinen Theorie in der Form

$$y = 1 - \frac{2}{x^2 + 1}$$

darzustellen.

§ 16. Gebrochene rationale Funktionen (allgemeiner Fall)

Jede gebrochene rationale Funktion $y = R(x)$ kann dargestellt werden als Quotient zweier Polynome ohne gemeinsame Nullstelle:

$$R(x) = \frac{P(x)}{Q(x)},$$

wobei

$$\begin{aligned} P(x) &= ax^n + bx^{n-1} + \cdots + kx + l \\ &= a(x-\alpha)^\lambda (x-\beta)^\mu \cdots (x^2 + px + q)^\varrho \cdots, \\ Q(x) &= a' x^{n'} + b' x^{n'-1} + \cdots + k' x + l' \\ &= a' (x-\alpha')^{\lambda'} (x-\beta')^{\mu'} \cdots (x^2 + p' x + q')^{\varrho'} \cdots \end{aligned}$$

ist und die bei der Aufspaltung in Faktoren auftretenden quadratischen Polynome nur komplexe Nullstellen besitzen, also stets positiv sind. Die Nullstellen des Zählers werden, wie bereits gesagt, als verschieden von den Nullstellen des Nenners vorausgesetzt. Von den höchsten Koeffizienten a und a' können wir (und dadurch wird die Allgemeinheit der folgenden Überlegungen nur unwesentlich eingeschränkt) annehmen, daß sie gleich Eins sind. Also ist

$$y = R(x) = \frac{(x-\alpha)^\lambda (x-\beta)^\mu \cdots (x^2 + px + q)^\varrho \cdots}{(x-\alpha')^{\lambda'} (x-\beta')^{\mu'} \cdots (x^2 + p' x + q')^{\varrho'} \cdots}. \tag{34}$$

Wir stellen uns nun vor, daß die Zahlen $\alpha, \beta, \ldots, \alpha', \beta', \ldots$, also die reellen Nullstellen des Zählers und des Nenners, der Größe nach angeordnet sind. Durch sie wird dann die x-Achse in eine bestimmte Anzahl von Intervallen zerlegt. Ist nun x ein Wert aus irgendeinem dieser Intervalle, so kann man leicht die Anzahl derjenigen Faktoren des Zählers und des Nenners von (34) ermitteln, welche für diesen Wert x negativ sind; das Vorzeichen des zugehörigen Funktionswertes hängt dann davon ab, ob diese Anzahl gerade oder ungerade ist. Da die besagte Anzahl von Faktoren gleich der Anzahl derjenigen reellen Nullstellen des Zählers und des Nenners ist, deren Vielfachheit ungerade ist und die größer als der betrachtete Wert x sind, hängt sie nicht davon ab, wie man den Wert x in einem festen Intervall zwischen zwei aufeinanderfolgenden Nullstellen von Zähler oder Nenner wählt, sondern nur von dem Intervall selbst. Ausgehend von dieser Überlegung kann man nun aber auch sofort das Vorzeichen ermitteln, das die Funktion in einem gegebenen Intervall zwischen zwei aufeinanderfolgenden Nullstellen annimmt; dazu braucht man sich nur vorzustellen, daß x von $+\infty$ bis $-\infty$ fällt, und zu berücksichtigen, daß bei jedem Durchgang durch eine reelle Nullstelle ungerader Vielfachheit des Zählers oder des Nenners ein Vorzeichenwechsel auftritt.

Sind a und a' voneinander verschieden (also insbesondere nicht beide gleich Eins), so tritt vor dem Bruch auf der linken Seite von (34) noch ein Zahlkoeffizient auf; ist dieser positiv, so beeinflußt er das Vorzeichen der Funktion nicht (er bewirkt lediglich eine Dehnung des graphischen Bildes von (34) in Richtung der y-Achse); ist der Koeffizient dagegen negativ, so gelangt man zu einer entgegengesetzten Vorzeichenverteilung (zur Dehnung kommt noch eine Spiegelung an der x-Achse).

Wir wollen nun den Verlauf des graphischen Bildes der Funktion in der Umgebung der Punkte, die Nullstellen des Zählers oder des Nenners entsprechen, noch etwas eingehender studieren. Dazu sei ξ eine beliebige Nullstelle des Zählers oder des Nenners von (34). Wir können dann (34) auch in der Form

$$y = R(x) = (x - \xi)^{\sigma} R_1(x)$$

schreiben, wobei ξ die betrachtete Nullstelle und $|\sigma|$ ihre Vielfachheit ist; $R_1(x)$ ist das Produkt aller übrigen Faktoren.

Ist ξ eine Nullstelle des Zählers, also $\sigma > 0$, so nimmt die Funktion (34) an der Stelle ξ den Wert Null an, d. h., der zugehörige Punkt der graphischen Darstellung liegt auf der x-Achse. Die graphische Darstellung berührt oder schneidet[1]) in diesem Punkt die x-Achse, je nachdem, ob σ gerade oder ungerade ist. Wenn man noch bedenkt, daß $R_1(\xi)$ positiv oder negativ sein kann, so sieht man, daß eine der vier in Abb. 24a schematisch dargestellten Anordnungen vorliegen muß.

---

[1]) Im Fall $\sigma = 1$ ohne Berührung, in den Fällen $\sigma = 3, 5, \ldots$ mit Berührung.

Im Fall $\sigma < 0$, wo also ξ eine Nullstelle des Nenners (und daher eine Unstetigkeitsstelle der Funktion $R(x)$) ist, gelangt man zu vier entsprechenden Anordnungen, die in Abb. 24b dargestellt sind.

Um die Frage nach „dem Verhalten der Kurve im Unendlichen" zu klären, braucht man nur den Zähler und den Nenner der vorgegebenen Funktion durch x^N zu dividieren, wobei N das Maximum der Zahlen n und n' (des Grades

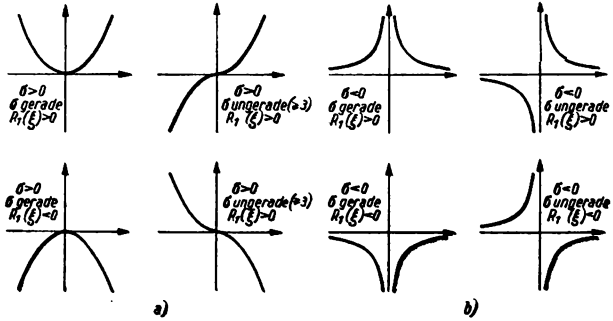


Abb. 24

von Zähler bzw. Nenner) ist. Man findet dann als Verhalten von $R(x)$ bei unbeschränkt wachsendem $|x|$:

1. Ist $n < n'$, so nähert sich die Kurve asymptotisch der x-Achse;

2. ist $n = n'$, so besitzt sie die Gerade

$$y = \frac{a}{a'}$$

als horizontale Asymptote;

3. ist $n > n'$, so wächst $|y|$ unbeschränkt; dabei gibt es für $n \geqq n' + 2$ keine geradlinige Asymptote, während im Fall $n = n' + 1$ eine „geneigte" Asymptote vorliegt, deren Gleichung man dadurch erhält, daß man den „ganzrationalen Anteil" (der im betrachteten Fall ein Polynom ersten Grades ist) abspaltet. Zum Beispiel besitzt die durch

$$y = \frac{2x^3 + 3x^2}{x^2 + x + 1}$$

beschriebene Kurve, wie man aus der Darstellung

$$y = 2x + 1 - \frac{3x + 1}{x^2 + x + 1}$$

ersieht, die Gerade $y = 2x + 1$ als Asymptote.

§ 17. Irrationale algebraische Funktionen

Unter einer ganzrationalen Funktion (einem Polynom) von zwei Veränderlichen x und y versteht man eine Funktion in den Veränderlichen x und y, deren Werte man aus den Werten dieser Veränderlichen und aus konstanten Zahlen mit Hilfe einer Formel berechnen kann, in die nur die Operationen Addition, Subtraktion und Multiplikation eingehen.

Ist $P(x, y)$ eine solche Funktion, so definiert die (sogenannte *algebraische*) Gleichung

$$P(x, y) = 0 \tag{35}$$

eine algebraische funktionale Abhängigkeit zwischen x und y.

Ist nun

$$y = f(x) \tag{36}$$

eine (eindeutige oder mehrdeutige) Funktion, die in einem gewissen Intervall einer algebraischen Gleichung (35) identisch in x genügt, d. h.

$$P(x, f(x)) \equiv 0,$$

so nennt man $f(x)$ eine *algebraische Funktion*. Zum Beispiel ist die Funktion $y = \sqrt{1 - x^2}$ eine algebraische Funktion, denn sie genügt im Intervall $-1 \leqq x \leqq 1$ der Gleichung $x^2 + y^2 = 1$.[1])

Jede rationale Funktion (speziell also jede ganzrationale Funktion) *ist algebraisch.*

In der Tat: Aus der Beziehung

$$y = \frac{P(x)}{Q(x)},$$

wobei $P(x)$ und $Q(x)$ Polynome (ganzrationale Funktionen) in x sind, folgt die Beziehung

$$Q(x)\, y - P(x) = 0,$$

deren linke Seite eine ganzrationale Funktion in den Veränderlichen x und y ist.

Ebenso wie die Polynome einer Veränderlichen werden auch die Polynome in zwei Veränderlichen x und y nach ihrem *Grad* klassifiziert. Man muß hierbei jedoch hinsichtlich des Grades in bezug auf x, des Grades in bezug auf y und des Grades in bezug auf beide Veränderlichen unterscheiden; unter letzterem wird folgendes verstanden: Zunächst versteht man unter dem Grad eines einzelnen Gliedes in bezug auf x und y die Summe der Exponenten von x und y, die in diesem Glied auftreten; der Grad des gesamten Polynoms in bezug auf x und y ist dann das Maximum der Grade seiner einzelnen Glieder.

Zum Beispiel ist das Polynom

$$P(x, y) = x^2 + 3x^5y^3 - y^4$$

---

[1]) Siehe auch Seite 203.

ein Polynom in x vom Grad 5, ein Polynom in y vom Grad 4 und ein Polynom in x und y vom Grad 8.

Hat die linke Seite einer algebraischen Gleichung (35) den Grad n in bezug auf x und y, so sagt man, die algebraische Gleichung sowie ihre graphische Darstellung seien von der *Ordnung* n.

Aus dem Vorangehenden folgt, daß jede rationale Funktion $y=\frac{P(x)}{Q(x)}$ einer algebraischen Gleichung genügt, die in bezug auf y den Grad Eins besitzt. Dies bedeutet jedoch nicht, daß auch umgekehrt jede algebraische Gleichung, die eine rationale Funktion festlegt, bezüglich y den Grad Eins besitzen muß. Zum Beispiel genügt die rationale Funktion $y=x$ der Gleichung

$$y^2=x^2, \tag{37}$$

die in bezug auf y den Grad 2 besitzt, und der Gleichung

$$y^3=x^3, \tag{38}$$

die in bezug auf y den Grad 3 hat; der Gleichung (37) genügt neben der rationalen Funktion $y=x$ auch noch die rationale Funktion $y=-x$, während der Gleichung (38) außer der Funktion $y=x$ keine andere Funktion genügt. Die algebraischen Funktionen, die nicht rational sind, heißen *irrational*.

Einfache Beispiele für irrationale algebraische Funktionen sind

$$y=\sqrt{x}$$

und

$$y=\sqrt{x^2+1}.^{1})$$

Um die Irrationalität einer gegebenen algebraischen Funktion zu beweisen, muß man zeigen, daß es unmöglich ist, sie als Quotienten zweier Polynome darzustellen. Wie jeder Unmöglichkeitsbeweis ist auch ein solcher Beweis mit gewissen Schwierigkeiten verbunden. Dem Leser sei empfohlen, die als Beispiel angeführten Funktionen nach dem Muster des Beweises für die Irrationalität von $\sqrt{2}$ als irrational nachzuweisen; man muß sich dabei auf den folgenden Hilfssatz stützen: *Ist das Quadrat eines Polynoms durch* x (*bzw.* x^2+1) *teilbar, so ist auch das Polynom selbst durch* x (*bzw.* x^2+1) *teilbar.*

§ 18. Einige Beispiele für die Diskussion von algebraischen Funktionen

Wir wollen hier nur einige charakteristische Beispiele für die Diskussion von algebraischen Funktionen betrachten.

Beispiel 1. Die Funktion

$$y=\sqrt{x} \qquad (x\geqq 0) \tag{39}$$

[1]) In diesem und dem folgenden Paragraphen werden, wenn nichts anderes gesagt wird, die Wurzeln stets im algebraischen Sinn verstanden, d. h., im Fall eines geraden Exponenten sind bei der Wurzel beide Vorzeichen zugelassen.

ist die zur Funktion $y = x^2$ gehörende Umkehrfunktion, denn sie genügt der Gleichung

$$y^2 - x = 0. \tag{40}$$

Vertauscht man in (40) die Veränderlichen x und y, so erhält man die Gleichung

$$y - x^2 = 0, \tag{41}$$

die wir bereits früher untersucht haben (§ 7). Die graphische Darstellung von (41) ist eine Parabel, die die y-Achse als Symmetrieachse besitzt und deren Scheitel im Ursprung O liegt; sie ist ganz in der oberen Halbebene enthalten (Abb. 13b). Die graphische Darstellung der Gleichung (40) ergibt sich aus der

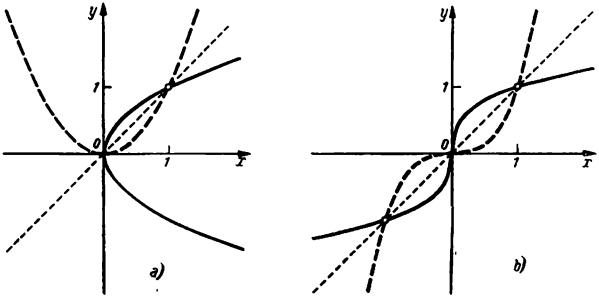


Abb. 25

von (41) durch Spiegelung an der Geraden $y = x$; sie hat die x-Achse als Symmetrieachse und liegt ganz in der rechten Halbebene (Abb. 25a).

Die graphische Darstellung von (41) hat in ihrem Scheitel O eine horizontale Tangente (die x-Achse), so daß die graphische Darstellung von (40) im Punkt O eine vertikale Tangente besitzt (die y-Achse).

Durch die Gleichung (40) wird eine doppeldeutige Funktion in y festgelegt; daher ist in (39) die Wurzel im algebraischen Sinn (mit beliebigem Vorzeichen) aufzufassen. Nimmt man nur das positive Vorzeichen (die Wurzel im arithmetischen Sinn), so erhält man als graphisches Bild nur eine „Hälfte" der Parabel, und zwar die in der oberen Halbebene liegende; entsprechendes gilt, wenn man nur das negative Vorzeichen zuläßt.

Beispiel 2. Ähnlich liegen die Dinge bei der Funktion

$$y = \sqrt[3]{x}, \tag{42}$$

der Umkehrfunktion zu

$$y = x^3. \tag{43}$$

Das graphische Bild von (43) ist eine kubische Parabel, deren Scheitel im Ursprung O liegt, die im ersten und dritten Quadranten verläuft und die x-Achse im Punkt O berührt (vgl. § 7; Abb. 13c). Daher ist das graphische Bild von (42) ebenfalls eine kubische Parabel, und zwar liegt sie zu der durch (43) beschriebenen kubischen Parabel spiegelbildlich in bezug auf die Gerade $y = x$, besitzt also ebenfalls den Punkt O als Scheitel, liegt ebenfalls im ersten und dritten Quadranten, hat aber im Punkt O die y-Achse als vertikale Tangente (Abb. 25b). Die Funktionen (42) und (43) sind beide eindeutig und für alle Werte von x erklärt.

Beispiel 3. Allgemein ist für jedes positive n die Funktion

$$y = \sqrt[n]{x} \tag{44}$$

Umkehrfunktion zu

$$y = x^n. \tag{45}$$

Das graphische Bild jeder Funktion (44) liegt zum graphischen Bild der entsprechenden Funktion (45) spiegelbildlich in bezug auf die Gerade $y = x$; alle Kurven haben also im Punkt O die y-Achse als vertikale Tangente, der sie sich um so enger anschmiegen, je größer n ist. Der Verlauf der Kurve ist für gerades n ähnlich dem der Kurve für $n = 2$ und für ungerades n ähnlich dem der Kurve für $n = 3$.

Beispiel 4. Wir wollen als nächstes die Potenzfunktion mit beliebigem positivem rationalem Exponenten, d. h. die Funktion

$$y = x^{\frac{p}{q}} \qquad (p \neq q) \tag{46}$$

oder

$$y = \sqrt[q]{x^p} \tag{47}$$

betrachten, wobei p und q positive ganze Zahlen ohne gemeinsamen Faktor sind. Die Funktion (46) genügt der algebraischen Gleichung

$$x^p - y^q = 0. \tag{48}$$

Beschränken wir uns zunächst auf die Betrachtung positiver Werte für x und y, so erkennen wir, daß der Quotient

$$\frac{y}{x} = x^{\frac{p}{q}-1},$$

wenn x gegen Null geht, entweder gleichfalls beliebig klein wird oder aber unbegrenzt wächst, je nachdem nämlich, ob $\frac{p}{q} > 1$ oder $\frac{p}{q} < 1$ ist; geometrisch bedeutet dies, daß sich das graphische Bild von (46) bei Annäherung an den Ursprung aus dem ersten Quadranten her entweder an die x-Achse (Abb. 26a) oder an die y-Achse (Abb. 26b) anschmiegt (die betreffende Achse „berührt"), je nachdem, ob der Exponent $\frac{p}{q}$ größer oder kleiner als Eins ist.

Über den Verlauf des graphischen Bildes in den anderen Quadranten erhält man leicht Aufschluß aus Gleichung (48): Sind p und q beide ungerade, so liegt das graphische Bild symmetrisch in bezug auf den Ursprung, es liegt also im ersten und dritten Quadranten; ist p gerade und q ungerade, so ist die graphische Darstellung symmetrisch zur y-Achse, sie liegt also im ersten und zweiten Quadranten; ist p ungerade und q gerade, so liegt Symmetrie in bezug auf die x-Achse vor, das graphische Bild verläuft im ersten und vierten Quadranten; der Fall, daß p und q beide gerade sind, kann nicht eintreten, da p und q als teilerfremd vorausgesetzt sind.

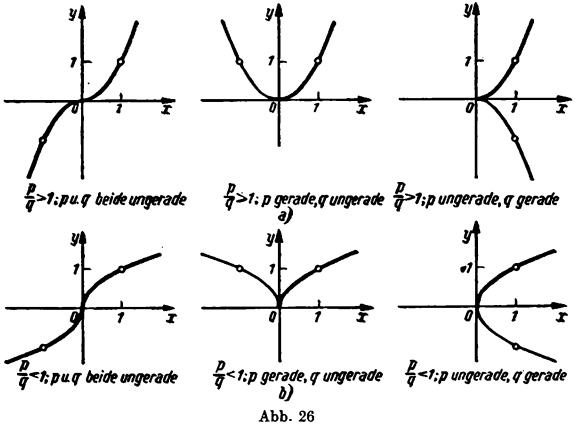


Abb. 26

Sind die Zahlen p und q nicht beide ungerade und ist die ungerade Zahl größer als die gerade, so tritt im Koordinatenursprung O eine „Spitze" auf. In Abb. 26 sind die möglichen Fälle schematisch dargestellt.

Beispiel 5. Die Funktionen

$$y = x^{-\frac{p}{q}}$$

oder

$$y = \frac{1}{\sqrt[q]{x^p}}$$

(für positives ganzes p und q) sind die reziproken Funktionen der im vorangehenden Beispiel behandelten Funktionen. Wir beschränken uns darauf (vgl. § 2, 6, und § 5, 7), den Verlauf ihrer graphischen Darstellungen in der Umgebung des Ursprungs in Abb. 27 schematisch darzustellen.

Beispiel 6. Das graphische Bild der Funktion

$$y = \sqrt{-x} \qquad (x \leqq 0) \tag{49}$$

liegt zu dem der Funktion (39) symmetrisch in bezug auf die y-Achse (vgl. Abb. 25a).

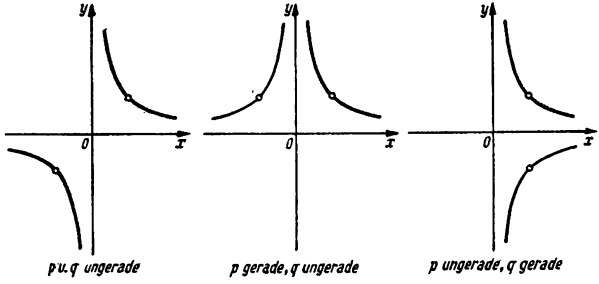


Abb. 27

Beispiel 7. Die Funktion

$$y = \sqrt{1 - x^2} \qquad (|x| \leqq 1) \tag{50}$$

genügt der algebraischen Gleichung

$$x^2 + y^2 = 1. \tag{51}$$

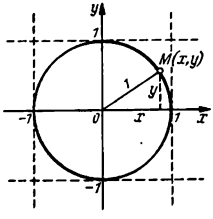


Abb. 28

Ihr graphisches Bild ist der Kreis um den Ursprung mit dem Radius 1 (Abb. 28); denn man ersieht aus Gleichung (51), daß die Koordinaten eines Punktes $M(x, y)$ dann und nur dann dieser Gleichung genügen, wenn sein Abstand vom Koordinatenursprung gleich Eins ist.

Bei einer Dehnung der Ebene in Richtung der x- und der y-Achse mit dem Dehnungsfaktor r erhalten wir die Gleichung

$$x^2 + y^2 = r^2, \tag{52}$$

deren graphisches Bild der Kreis um den Ursprung mit dem Radius r ist. Wenn wir nun noch eine Verschiebung parallel zur x-Achse um die Strecke a und eine Verschiebung parallel zur y-Achse um die Strecke b vornehmen, so gelangen wir zu der Gleichung

$$(x - a)^2 + (y - b)^2 = r^2, \tag{53}$$

die den Kreis vom Radius r um den Punkt mit den Koordinaten (a, b) beschreibt.

Beispiel 8. Bei Dehnung in Richtung der x-Achse mit dem Dehnungsfaktor p und Dehnung in Richtung der y-Achse um den Dehnungsfaktor q nimmt Gleichung (51) die Form

$$\frac{x^2}{p^2}+\frac{y^2}{q^2}=1 \tag{54}$$

oder

$$y=\frac{q}{p}\sqrt{p^2-x^2} \qquad (|x|\leqq p)$$

an. Diese Gleichungen beschreiben die *Ellipse* mit den Halbachsen p und q (Abb. 29). Die Gleichung (54) wird die *kanonische Gleichung* der Ellipse (auch *Hauptachsengleichung*) genannt. Für $p=q=r$ erhält man die Gleichung (52) für den Kreis um O mit dem Radius r.

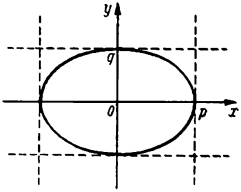


Abb. 29

Beispiel 9. Wendet man auf die Gleichungen

$$\text{a)}\quad x^2-y^2=1, \qquad y=\sqrt{x^2-1} \qquad (|x|\geqq 1), \tag{55}$$

bzw.

$$\text{b)}\quad x^2-y^2=-1, \qquad y=\sqrt{x^2+1} \tag{56}$$

die in § 3 betrachtete Transformation IVa bzw. IVb an, so gelangt man in beiden Fällen zur Gleichung

$$2xy=1.$$

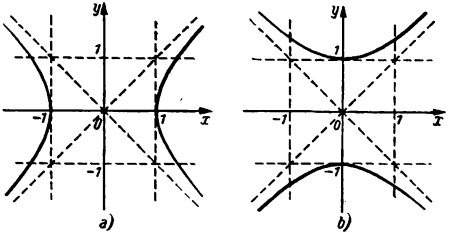


Abb. 30

Es genügt, eine Dehnung in Richtung beider Koordinatenachsen mit dem Dehnungsfaktor $\sqrt{2}$ auszuführen, um zu erkennen, daß die graphischen Bilder der genannten irrationalen algebraischen Funktionen Hyperbeln sind (vgl. § 13, Abb. 17). Sie sind in Abb. 30 dargestellt.

Beispiel 10. Wendet man auf die Gleichungen a) und b) des Beispiels 9 eine Dehnung in Richtung der x-Achse mit dem Dehnungsfaktor p und eine

Dehnung in Richtung der y-Achse mit dem Dehnungsfaktor q an, so gelangt man zu den Gleichungen

$$\text{a')}\quad \frac{x^2}{p^2}-\frac{y^2}{q^2}=1, \qquad y=\frac{q}{p}\sqrt{x^2-p^2} \qquad (|x|\geqq p), \tag{57}$$

$$\text{b')}\quad \frac{x^2}{p^2}-\frac{y^2}{q^2}=-1, \qquad y=\frac{q}{p}\sqrt{x^2+p^2} \tag{58}$$

(deren graphische Bilder zueinander konjugierte Hyperbeln mit den Halbachsen p und q sind; Abb. 31a und 31b).

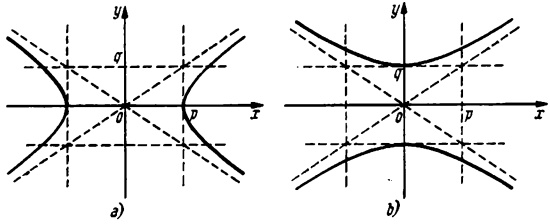


Abb. 31

Beispiel 11. Die Funktionen

$$\left.\begin{aligned} &\text{a)}\quad y=\frac{1}{\sqrt{1-x^2}} \qquad (|x|<1),\\ &\text{b)}\quad y=\frac{1}{\sqrt{x^2-1}} \qquad (|x|>1),\\ &\text{c)}\quad y=\frac{1}{\sqrt{x^2+1}} \end{aligned}\right\} \tag{59}$$

genügen den Gleichungen

$$x^2y^2+1=y^2, \quad x^2y^2=y^2+1 \quad \text{bzw.} \quad x^2y^2+y^2=1.$$

Sie sind die zu (50), (55) bzw. (56) reziproken Funktionen.

Durch die genannte Tatsache ist bereits ihr Wachstumsverhalten festgelegt. In Abb. 32 sind ihre graphischen Bilder dargestellt (ausgezogene Linien); zum Vergleich ist jeweils das graphische Bild der entsprechenden Funktion (50), (55) bzw. (56) eingezeichnet (gestrichelte Linien).

Beispiel 12.

$$\sqrt{x}+\sqrt{y}=1. \tag{60}$$

Beseitigt man in (60) in bekannter Weise die Wurzeln, so erhält man die Gleichung

$$(x-y)^2+1=2(x+y). \tag{60'}$$

Sie definiert die doppeldeutige algebraische Funktion

$$y = (1 - \sqrt{x})^2 \qquad (x \geqq 0). \tag{61}$$

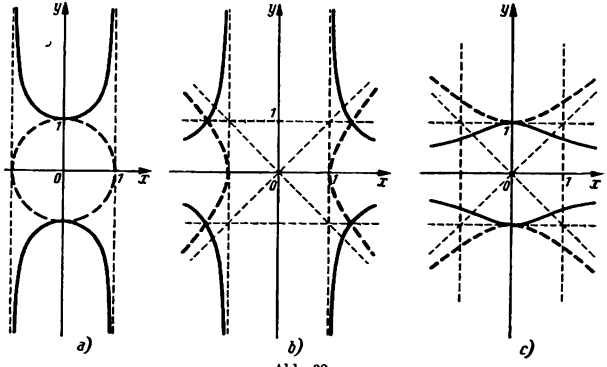


Abb. 32

Ihr graphisches Bild hat die Gerade $y = x$ als Symmetrieachse; denn die Gleichungen (60) und (60′) ändern sich nicht, wenn man dort x und y vertauscht. Es liegt vollständig im ersten Quadranten und hat mit den Achsen nur die Punkte mit den Koordinaten (0, 1) und (1, 0) gemeinsam. Eine Drehung der Ebene um $-45°$ (vgl. § 3) führt die Gleichung (60′) in die Gleichung

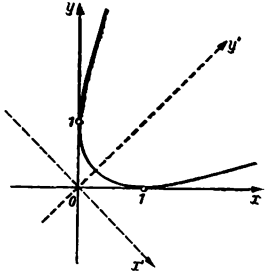


Abb. 33

$$y' - \frac{1}{2\sqrt{2}} = \frac{x'^2}{\sqrt{2}} \tag{62}$$

über. Der Gleichung (62) entnimmt man, daß das graphische Bild eine Parabel ist, die in den Punkten (0, 1) und (1, 0) die Koordinatenachsen berührt (Abb. 33).[1])

Beispiel 13. Die vierdeutige Funktion

$$y = \sqrt{x+1} + \sqrt{x} \qquad (x \geqq 0) \tag{63}$$

[1]) Die graphische Darstellung von (62) ergibt sich nämlich aus der Parabel $y = x^2$ durch Dehnung in Richtung der x-Achse mit dem Dehnungsfaktor $\sqrt{2}$ und Verschiebung parallel zur y-Achse um die Strecke $\frac{1}{2\sqrt{2}}$.

genügt der Gleichung

$$(y^2-1)^2=4xy^2, \tag{64}$$

zu der man gelangt, wenn man zunächst

$$u=\sqrt{x+1}, \quad v=\sqrt{x}, \quad y=u+v$$

setzt und dann u und v wieder eliminiert. Über den Verlauf des graphischen Bildes erlangt man Klarheit, wenn man y als unabhängige Veränderliche nimmt,

$$x=\frac{(y^2-1)^2}{4y^2},$$

oder indem man die Parabeln $y=\sqrt{x}$ und $y=\sqrt{x+1}$ „addiert" (vgl. § 2).

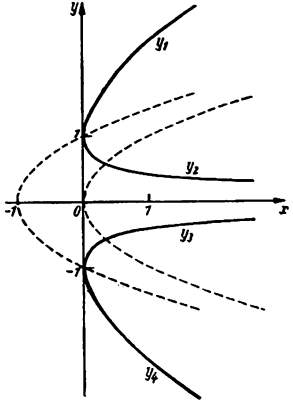


Abb. 34

Nimmt man die Wurzeln in (63) im arithmetischen Sinn, so spaltet sich die vierdeutige Funktion in vier eindeutige Funktionen auf, nämlich in

$$y_1=\sqrt{x+1}+\sqrt{x},$$

$$y_2=\sqrt{x+1}-\sqrt{x}=\frac{1}{\sqrt{x+1}+\sqrt{x}}=\frac{1}{y_1},$$

$$y_3=-y_2,$$

$$y_4=-y_1.$$

Die graphischen Darstellungen von y_1 und y_2 liegen beide im ersten Quadranten; y_1 ist überall wachsend, y_2 überall fallend; die graphischen Darstellungen von y_1 und y_2 gehen im Punkt (0, 1) ineinander über, wo sie eine vertikale Tangente besitzen (Abb. 34). Mit x wächst y_1 unbeschränkt, wenn auch ziemlich langsam[1]), während y_2 gegen Null strebt.[2]) Das graphische Bild von y_3 bzw. y_4 liegt zu dem von y_2 bzw. y_1 symmetrisch in bezug auf die x-Achse.

---

[1]) Genaueres sagt darüber die Beziehung

$$y_1-2\sqrt{x}=\frac{1}{\sqrt{x+1}+\sqrt{x}}<\frac{1}{2\sqrt{x}}$$

aus.

[2]) Es ist nämlich

$$y_2<\frac{1}{2\sqrt{x}}.$$

Beispiel 14. Aus der Gleichung

$$x^4+y^4=x^2+y^2 \tag{65}$$

(vgl. § 1, Seite 7) kann man bereits eine Reihe von Symmetrieeigenschaften des zugehörigen graphischen Bildes ablesen (§ 3): Die graphische Darstellung von (65) liegt symmetrisch in bezug auf die x-Achse, symmetrisch in bezug auf die y-Achse (und folglich symmetrisch zum Ursprung), aber auch symmetrisch in bezug auf die Winkelhalbierende $y=x$. Daher kann man sich bei ihrer Diskussion auf das durch die Ungleichungen

$$0 \leqq x \leqq y$$

beschriebene „Achtel" der Ebene beschränken. Man braucht also in der nach y aufgelösten Gleichung

$$y=\sqrt{\frac{1}{2}+\sqrt{\frac{1}{4}+x^2-x^4}} \tag{66}$$

(beide Wurzeln im arithmetischen Sinn genommen, denn das negative Vorzeichen vor der inneren Wurzel ergibt für $0<x\leqq 1$ keinen Sinn) nur solche Werte für x zu betrachten, die nicht größer als 1 sind (für $x>1$ wird nämlich $\sqrt{\frac{1}{4}+x^2-x^4}<\frac{1}{2}$ und damit $y<1$, was der Bedingung $x\leqq y$ widerspräche). Wenn nun x von 0 bis 1 wächst, so hängt die Änderung der zugehörigen Funktionswerte y nur von der Änderung von

$$z=\frac{1}{4}+x^2-x^4=\frac{1}{2}-\left(x^2-\frac{1}{2}\right)^2$$

ab (vgl. § 9). Offenbar ist z niemals größer als $\frac{1}{2}$ und nimmt den Wert $\frac{1}{2}$ nur für $x=\frac{1}{\sqrt{2}}$ an. Infolgedessen kann die Funktion (66) höchstens den Wert $\sqrt{\frac{1}{2}+\sqrt{\frac{1}{2}}}$ annehmen, und dieses Maximum wird im betrachteten Intervall $0\leqq x\leqq 1$ genau an der Stelle $x=\frac{1}{\sqrt{2}}$ angenommen. Wächst x von 0 bis $\frac{1}{\sqrt{2}}$, so fällt $\left(x^2-\frac{1}{2}\right)^2$ von $\frac{1}{4}$ bis 0, so daß z von $\frac{1}{4}$ bis $\frac{1}{2}$ und y von 1 bis $\sqrt{\frac{1}{2}+\sqrt{\frac{1}{2}}}$ wächst; wächst x weiter von $\frac{1}{\sqrt{2}}$ bis 1, so wächst $\left(x^2-\frac{1}{2}\right)^2$ von 0 bis $\frac{1}{4}$, so daß z von $\frac{1}{2}$ bis $\frac{1}{4}$ und y von $\sqrt{\frac{1}{2}+\sqrt{\frac{1}{2}}}$ bis 1 fällt.

Man beachte, daß man für $x=0$ und $x=1$ vor die innere Wurzel das Minuszeichen setzen kann. Dies ergibt noch einen „isolierten Punkt" im Ursprung O. In Abb. 35 ist der Gesamtverlauf des graphischen Bildes von (66) dargestellt.

Bemerkung. Ist die rechte Seite einer Gleichung $y = f(x)$ aus x und konstanten Zahlen allein mit Hilfe der rationalen Operationen und von Wurzeln mit ganzen Exponenten (in endlicher Anzahl) „aufgebaut", so ist y eine elementare und obendrein algebraische Funktion der Veränderlichen x. Durch „Beseitigen" der Wurzeln kann man die Gleichung $y = f(x)$ nämlich stets auf die Form (35) bringen.[1])

Gehen wir umgekehrt von einer algebraischen Gleichung der Form (35) aus, so ist, wenn es überhaupt eine Funktion $y = f(x)$ gibt, die dieser Gleichung genügt[2]), diese Funktion auch algebraisch; sie braucht jedoch keineswegs elementar zu sein. Zum Beispiel ist es nicht möglich, die Gleichung

$$y^5 + y = x$$

durch Wurzeln (über den Parameter x) nach y aufzulösen; trotzdem kann man leicht zeigen, daß in diesem Fall y stets eindeutig durch x festgelegt, also y eine eindeutige Funktion von x ist: Wächst nämlich x von $-\infty$ bis $+\infty$, so wächst auch die linke Seite der Gleichung von $-\infty$ bis $+\infty$ (Abb. 36) und nimmt daher für einen und nur einen Wert von y einen vorgegebenen Wert x an (vgl. § 52, Satz von Bolzano). Selbstverständlich wird hierbei der Begriff Funktion im Sinne von „eindeutiger Zuordnung" verwendet.

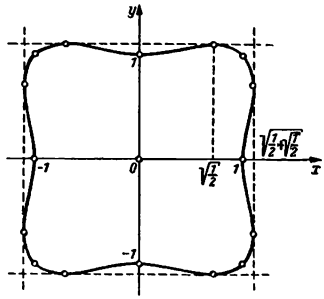


Abb. 35

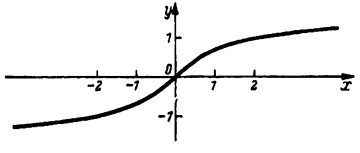


Abb. 36

§ 19. Elementare transzendente Funktionen

Der Begriff der transzendenten Funktion hat eine ganz präzise Bedeutung: Eine Funktion $y = f(x)$ heißt *transzendent*, wenn sie keiner algebraischen Gleichung $P(x, y) = 0$ genügt, wobei $P(x, y)$ ein Polynom (eine ganzrationale Funktion) in den Veränderlichen x und y ist. Im Mathematikunterricht der Oberschule werden an transzendenten Funktionen die Exponentialfunktionen, die trigonometrischen Funktionen und deren Umkehrfunktionen systematisch behandelt; daneben treten noch mannigfache Kombinationen dieser Funktio-

[1]) Hierzu führt man die Wurzeln als neue Veränderliche ein, die man anschließend wieder eliminiert (vgl. EdEM, Band II, A. I. Uskow, Vektorräume und lineare Transformationen, § 17).

[2]) Das ist nicht stets der Fall! So gibt es keine reellwertige Funktion einer reellen Veränderlichen, die der Gleichung $x^2 + y^2 + 1 = 0$ genügt.

nen auf. In der höheren Mathematik werden transzendente Funktionen der verschiedensten Typen betrachtet und studiert.

Das Wort „transzendent" bedeutet wörtlich: „übersteigend", und zwar versteht man nach EULER darunter: den Anwendungsbereich algebraischer Methoden übersteigend, was genau der oben angeführten Definition entspricht. Transzendenzbeweise für Funktionen sind „Unmöglichkeitsbeweise"; sie werden indirekt geführt und erfordern tieferliegende Verfahren verschiedener Art.

§ 20. Die Exponentialfunktionen

Jede Funktion der Form

$$y = a^x \qquad (a > 0) \tag{67}$$

heißt eine *Exponentialfunktion*. Die Bedeutung des Symbols a^x wird dabei in mehreren Schritten erklärt:

1. Bezeichnet x eine positive ganze Zahl n, so versteht man unter a^n das Resultat der n-maligen Multiplikation der durch a bezeichneten Zahl mit sich selbst:

$$a^n = \underbrace{a\,a\cdots a}_{n\text{ mal}}. \tag{68}$$

2. Ist x eine positive gebrochene rationale Zahl $\frac{p}{q}$, so ist definitionsgemäß

$$a^{\frac{p}{q}} = \sqrt[q]{a^p}, \tag{69}$$

wobei die Wurzel im arithmetischen Sinn gemeint ist.

3. Ist x eine negative rationale Zahl $-\frac{p}{q}$, so ist

$$a^{-\frac{p}{q}} = \frac{1}{a^{\frac{p}{q}}}. \tag{70}$$

4. Weiter setzt man

$$a^0 = 1. \tag{71}$$

5. Schließlich wird a^x von den rationalen Zahlen „stetig" auf die irrationalen Zahlen fortgesetzt (vgl. § 51).

Bemerkung. Im vorliegenden Paragraphen werden zwar die wichtigsten Eigenschaften der Exponentialfunktionen bereits formuliert, jedoch die Beweise nur für rationale Exponenten erbracht. Für irrationale Exponenten vgl. § 51.

Aus der Definition folgt, daß jede Exponentialfunktion für alle Werte der unabhängigen Veränderlichen x positive Werte annimmt. Insbesondere ist sie für jeden Wert x von Null verschieden, d. h., ihr graphisches Bild liegt ganz oberhalb der x-Achse.

Die Exponentialfunktionen auch für negative Basen a zu betrachten, ist aus folgendem Grunde nicht angebracht: Wenn a negativ ist, so darf die

unabhängige Veränderliche x sicher die Werte $\frac{2m+1}{2n}$, wobei m und n ganze Zahlen sind, nicht annehmen, von irrationalen Werten gar nicht zu sprechen (s. S. 489 ff.).

Die Fälle $a=0$ und $a=1$ sind ohne besonderes Interesse.

Ist die Basis a größer als Eins, so ist die zugehörige Exponentialfunktion wachsend, ist die Basis a kleiner als Eins, so ist sie fallend.

Es seien x' und x'' rationale Zahlen mit $x'<x''$. Ohne Beschränkung der Allgemeinheit können wir annehmen, daß sie auf denselben Nenner $q>0$ gebracht sind:

$$x'=\frac{p'}{q}, \quad x''=\frac{p''}{q}, \quad p'<p''.$$

Ist nun $a>1$, so ist $a^{p'}<a^{p''}$ (und zwar unabhängig vom Vorzeichen der Zahlen p' und p'' — *Anm. d. wiss. Red.*). Dann ist aber, da die Funktion $x^{\frac{1}{q}}$ für $x>0$ wachsend ist (vgl. § 18, Beispiel 3), auch

$$a^{\frac{p'}{q}}<a^{\frac{p''}{q}}, \quad \text{d. h.} \quad a^{x'}<a^{x''}.$$

Entsprechend ergeben sich im Fall $0<a<1$ die Ungleichungen $a^{p'}>a^{p''}$, $a^{\frac{p'}{q}}>a^{\frac{p''}{q}}$, d. h. $a^{x'}>a^{x''}$.

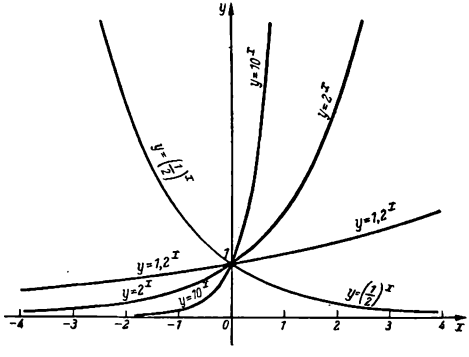


Abb. 37

Wir wollen nun die gegenseitige Lage der graphischen Bilder zweier Exponentialfunktionen a^x und b^x $(0<a<b)$ beschreiben. Dazu beachten wir zunächst, daß sie sich im Punkt (0, 1) schneiden; denn $a^0=b^0=1$. Ferner

stellen wir folgendes fest: Ist $0 < a < b$, so gilt für jedes $x > 0$ die Ungleichung $a^x < b^x$ und für jedes $x < 0$ die Ungleichung $a^x > b^x$. [1])

Schließlich können wir feststellen, daß die graphische Darstellung der Funktion $\left(\frac{1}{a}\right)^x$ zu der der Funktion a^x symmetrisch in bezug auf die y-Achse liegt; definitionsgemäß ist nämlich

$$\left(\frac{1}{a}\right)^x = a^{-x},$$

und unsere Behauptung ist eine unmittelbare Folge aus § 3, I b.

In Abb. 37 sind die graphischen Bilder einiger Exponentialfunktionen zu verschiedenen Basen dargestellt.

Ist $a > 1$, so wächst mit x auch die Exponentialfunktion a^x unbegrenzt, und zwar sehr schnell, schneller als jede beliebige Potenz von x.

Da die Funktion a^x wachsend ist, genügt es, die Funktion a^x für ganzzahlige Werte x zu betrachten; nimmt nämlich die Funktion für einen ganzzahligen Wert $x = n$ einen bestimmten Wert N an, so nimmt sie für jeden größeren Wert x (auch jeden Wert, der nicht ganz ist) einen Wert an, der größer als N ist.

Setzen wir nun

$$a - 1 = \delta, \quad a = 1 + \delta \qquad (\delta > 0),$$

so erhalten wir nach der binomischen Formel:

$$a^n = (1 + \delta)^n = 1 + n\delta + \cdots + \delta^n > n\delta$$

(denn eine Summe von positiven Zahlen ist größer als jeder Summand). Weil nun mit n auch $n\delta$ beliebig groß wird, gilt dies erst recht für a^n.

Es sei nun p eine beliebige positive ganze Zahl. Um zu zeigen, daß für alle genügend großen Werte n die Ungleichung

$$a^n > n^p$$

gilt, bedienen wir uns nochmals der vorangehenden Überlegungen, benutzen jetzt aber zur Abschätzung nicht das zweite, sondern das $(p+2)$-te Glied der binomischen Entwicklung:

$$a^n = (1 + \delta)^n = 1 + n\delta + \cdots + \frac{n(n-1)\cdots(n-p)}{1 \cdot 2 \cdots (p+1)} \delta^{p+1} + \cdots + \delta^n$$

$$> \frac{n(n-1)\cdots(n-p)}{1 \cdot 2 \cdots (p+1)} \delta^{p+1}. \tag{72}$$

Für hinreichend großes n ist nun aber

$$\frac{n(n-1)\cdots(n-p)}{1 \cdot 2 \cdots (p+1)} \delta^{p+1} > n^p; \tag{73}$$

[1]) Man braucht hierzu nur c für x zu setzen und zu beachten, daß die Funktion x^c für $c > 0$ wachsend und für $c < 0$ fallend ist (für rationales c folgt dies aus § 18, Beispiel 4, für irrationales c vgl. § 51).

denn diese Ungleichung ist gleichwertig der Ungleichung

$$n > \frac{1 \cdot 2 \cdots (p+1)}{\left(1-\frac{1}{n}\right) \cdots \left(1-\frac{p}{n}\right) \delta^{p+1}},$$

und diese ist tatsächlich für hinreichend großes n erfüllt, da die linke Seite der Ungleichung mit n über alle Grenzen wächst, während sich die rechte Seite dem festen Wert $\frac{1 \cdot 2 \cdots (p+1)}{\delta^{p+1}}$ nähert. Zur Vollendung des Beweises genügt es, die Ungleichungen (72) und (73) zu kombinieren.

Setzt man z. B. $\delta = 0{,}01$ und $p = 100$, so kann man behaupten, daß für hinreichend großes x

$$(1{,}01)^x > x^{100}$$

ist.

Aus der Identität

$$a^{-x} = \left(\frac{1}{a}\right)^x$$

folgt: *Ist $a < 1$ und wächst x, indem es stets negativ bleibt, dem Betrage nach über alle Grenzen, so werden die zugehörigen Werte der Exponentialfunktion $y = a^x$ beliebig klein, und zwar schneller als jede beliebige negative Potenz von x.*

Die Eigenschaft der Exponentialfunktion, schnell zu wachsen oder zu fallen, wenn die unabhängige Veränderliche wächst, kann zum Nachweis ihrer Transzendenz verwendet werden.

Beweis. Um etwas Bestimmtes vor Augen zu haben, setzen wir $a > 1$ voraus. Wir nehmen nun an, die Funktion $y = a^x$ wäre algebraisch. Dann gäbe es eine algebraische Gleichung der Form $P(x, y) = 0$, der die Funktion $y = a^x$ identisch genügt. Wir könnten dann das Polynom $P(x, y)$ nach fallenden Potenzen von y entwickeln und erhielten

$$P(x, y) \equiv P_0(x)\, y^n + P_1(x)\, y^{n-1} + \cdots + P_{n-1}(x)\, y + P_n(x),$$

wobei n eine positive ganze Zahl ist und $P_0(x)$, $P_1(x)$, ..., $P_{n-1}(x)$, $P_n(x)$ Polynome in x sind. Insbesondere sei

$$P_0(x) = A x^m + \cdots,$$

wobei m eine positive ganze Zahl oder gleich Null, A jedoch von Null verschieden ist. Ohne Beschränkung der Allgemeinheit könnten wir annehmen, daß $A > 0$ ist. Wir würden dann die Identität

$$P(x, a^x) \equiv 0$$

auf die Form

$$P_0(x)\, a^{nx} + P_1(x)\, a^{(n-1)x} + \cdots + P_{n-1}(x)\, a^x + P_n(x) \equiv 0$$

oder

$$P_0(x) \equiv -P_1(x)\, a^{-x} - \cdots - P_{n-1}(x)\, a^{-(n-1)x} - P_n(x)\, a^{-nx}$$

bringen. Sodann könnten wir x über alle Grenzen wachsen lassen. Dabei würde das Polynom auf der linken Seite der Identität ebenfalls beliebig groß (im Fall $m > 0$) werden, oder es wäre durchweg gleich der positiven Konstanten A (im Fall $m = 0$). Was nun die rechte Seite betrifft, so würde sie in eine Summe aus endlich vielen Gliedern zerfallen, von denen jedes die Gestalt $Cx^k a^{-kx}$ ($k > 0$) hätte. Infolgedessen strebte sie wegen der bewiesenen Eigenschaft der Exponentialfunktion gegen Null (man beachte, daß $a^{-k} = \frac{1}{a^k} < 1$ ist). Damit ergibt sich ein Widerspruch, und die Transzendenz der Funktion a^x ist bewiesen.

Unter den Eigenschaften der Exponentialfunktionen ist in erster Linie eine Funktionaleigenschaft zu nennen, die in der für beliebige Zahlen x' und x'' geltenden Beziehung

$$a^{x'+x''} = a^{x'} a^{x''} \tag{74}$$

zum Ausdruck kommt („Additionstheorem").

Diese Gleichung ist sicher dann richtig, wenn x' und x'' ganze Zahlen sind. Wir wollen nun zeigen, daß sie auch für beliebige rationale Zahlen x' und x'' gilt. Dazu denken wir uns x' und x'' als Brüche mit gemeinsamen Nenner $q > 0$ dargestellt:

$$x' = \frac{p'}{q}, \quad x'' = \frac{p''}{q}.$$

Dann ist zunächst offenbar

$$a^{p'+p''} = a^{p'} a^{p''}.$$

Hieraus ergibt sich auf Grund bekannter Eigenschaften der Wurzel:

$$a^{\frac{p'}{q}+\frac{p''}{q}} = \sqrt[q]{a^{p'+p''}} = \sqrt[q]{a^{p'} a^{p''}} = \sqrt[q]{a^{p'}} \sqrt[q]{a^{p''}} = a^{\frac{p'}{q}} a^{\frac{p''}{q}}.$$

Durch vollständige Induktion kann man (74) leicht auf den Fall verallgemeinern, daß auf der linken Seite im Exponenten eine Summe aus n Zahlen auftritt. Im Spezialfall, daß alle Summanden dieser Summe gleich (etwa gleich x) sind, ergibt dies die Identität

$$a^{nx} = (a^x)^n. \tag{75}$$

Diese Beziehung läßt sich nun leicht auf den Fall verallgemeinern, daß an Stelle von n eine beliebige (reelle) Zahl λ steht:

$$a^{\lambda x} = (a^x)^\lambda. \tag{76}$$

Ist λ eine ganze Zahl, so ergibt sich dies leicht aus (75) und evtl. (70) bzw. aus (71); ist dagegen λ eine gebrochene rationale Zahl $\frac{p}{q}$ mit $q > 0$, so ergibt sich (76) aus dem bereits behandelten Spezialfall $\lambda = p$, indem man auf beiden Seiten die q-te Wurzel zieht.

Aus (76) ergibt sich insbesondere folgendes: *Durchläuft bei einer Exponentialfunktion a^x die unabhängige Veränderliche x eine arithmetische Folge*

$$g,\ g+h,\ g+2h,\ \ldots,\ g+nh,\ \ldots,$$

so durchlaufen die zugehörigen Funktionswerte eine geometrische Folge

$$G,\ Gq,\ Gq^2,\ \ldots,\ Gq^n,\ \ldots.^{1})$$

Dabei ist

$$G = a^g, \quad q = a^h.$$

Für das graphische Bild der Funktion a^x bedeutet dies, daß zu einer Folge von äquidistanten Punkten der x-Achse eine Folge von Punkten des graphischen Bildes gehört, deren Ordinaten eine geometrische Folge bilden.

---

[1]) Es ist zu bemerken, daß neben den Exponentialfunktionen a^x auch die Funktionen der etwas umfassenderen Klasse ca^x (wobei c eine beliebige Konstante ist) diese Eigenschaft besitzen.

Setzt man in (74) $x' = x$, $x'' = -c$, so erhält man

$$a^{x-c} = Ca^x,$$

wobei $C = a^{-c}$. Dies bedeutet (vgl. § 3): *Eine Verschiebung des graphischen Bildes der Funktion a^x parallel zur x-Achse um die Strecke c ist gleichbedeutend seiner Dehnung in Richtung der y-Achse mit dem Dehnungsfaktor a^{-c}* (Stauchung mit dem Stauchungsfaktor a^c).

In (76) sind x und λ beliebige reelle Zahlen, so daß man in (76) ohne weiteres λ und x vertauschen kann:

$$a^{\lambda x} = (a^\lambda)^x.$$

Setzen wir hier $\lambda = \frac{1}{p}$ $(p > 0)$, so erhalten wir folgende Eigenschaft des graphischen Bildes der Funktion a^x: *Eine Dehnung in Richtung der x-Achse mit dem Dehnungsfaktor p ist gleichbedeutend dem Übergang zum graphischen Bild der Exponentialfunktion zur Basis $a^{\frac{1}{p}}$.*

Auch der Relation (74) wohnt eine ganz bestimmte geometrische Bedeutung inne. Setzen wir nämlich in (74) $x' = \lambda x$, $x'' = \mu x$, so erhalten wir

$$a^{\lambda x + \mu x} = a^{\lambda x} a^{\mu x},$$

woraus sich auf Grund von (76)

$$(a^\lambda)^x (a^\mu)^x = (a^{\lambda+\mu})^x \quad \text{oder} \quad a'^x a''^x \triangleq (a' a'')^x$$

ergibt, wobei

$$a' = a^\lambda,$$

$$a'' = a^\mu.$$

Somit gilt: *Die „Multiplikation"* (vgl. § 3) *der graphischen Bilder der Exponentialfunktionen zur Basis a' bzw. a'' liefert das graphische Bild einer Exponentialfunktion, und zwar der Exponentialfunktion zur Basis a'a''.*

§ 21. Die hyperbolischen Funktionen

Wir betrachten die Funktionen

$$f(x) = \frac{a^x + a^{-x}}{2}, \quad g(x) = \frac{a^x - a^{-x}}{2} \qquad (a > 1).$$

Das graphische Bild der Funktion $f(x)$ erhält man, wenn man die bezüglich der y-Achse zueinander symmetrischen Bilder der Exponentialfunktionen zur Basis a und $\frac{1}{a}$ „addiert" (vgl. § 2, 1) und anschließend eine Stauchung mit dem Stauchungsfaktor 2 vornimmt. Die auf diese Weise gewonnene graphische Darstellung könnte man die „Halbsumme" der gegebenen graphischen Bilder nennen. Die punktweise Konstruktion der graphischen Darstellung läuft natürlich darauf hinaus, daß man auf jeder vertikalen Geraden die durch ihre Schnittpunkte mit den graphischen Bildern von a^x und a^{-x} bestimmte Strecke halbiert.

In entsprechender Weise zeichnet man das Bild der Funktion $g(x)$, nur muß man hier von dem graphischen Bild von a^{-x} zunächst zum graphischen Bild von $(-a^{-x})$ übergehen, d. h., das graphische Bild von a^{-x} an der x-Achse spiegeln.

Man prüft leicht nach, daß *die Funktion $f(x)$ gerade und die Funktion $g(x)$ ungerade ist*; für die zugehörigen graphischen Bilder haben wir daher die bekannten Symmetrieeigenschaften (vgl. § 3).

Offenbar gilt: *Es ist* $f(0)=1$ *und* $g(0)=0$.

Es ist $f(x)>0$ *für alle* x *und* $g(x)>0$ *für alle* $x>0$, $g(x)<0$ *für alle* $x<0$. In der Tat ist für $a>1$ offenbar $a>a^{-1}$ und daher $a^x>a^{-x}$ (vgl. § 20) für $x>0$ usw.

Für alle Werte x gilt

$$[f(x)]^2-[g(x)]^2=1. \tag{77}$$

Wir wollen nun zeigen, daß *die Funktion $f(x)$ für $x>0$ wachsend und daher* (da sie eine gerade Funktion ist) *für $x<0$ fallend ist* und daß *die Funktion $g(x)$ überall wächst.* Zum Beweis genügt es zu bemerken, daß

$$f(x)=\frac{1}{2}\left(u+\frac{1}{u}\right), \quad g(x)=\frac{1}{2}\left(u-\frac{1}{u}\right),$$

wobei

$$u=a^x.$$

Ist nun $x>0$, so ist $u>1$ (da nach Voraussetzung $a>1$). Für $u>1$ sind aber (vgl. § 15, (I) und (II)) die Funktionen $u+\frac{1}{u}$ und $u-\frac{1}{u}$ wachsend. Da nun für $a>1$ die Exponentialfunktion $u=a^x$ ebenfalls wachsend ist, sind unsere Behauptungen evident (vgl. § 5, 9).

Aus der Funktionaleigenschaft (74) der Exponentialfunktionen kann man eine entsprechende Funktionaleigenschaft der Funktionen $f(x)$ und $g(x)$ ableiten.

Einerseits ist nämlich

$$f(x'+x'')=\frac{1}{2}\left(a^{x'+x''}+a^{-(x'+x'')}\right)=\frac{1}{2}\left(a^{x'}a^{x''}+a^{-x'}a^{-x''}\right),$$

andererseits ist

$$f(x')f(x'')+g(x')g(x'')$$
$$=\frac{a^{x'}+a^{-x'}}{2}\cdot\frac{a^{x''}+a^{-x''}}{2}+\frac{a^{x'}-a^{-x'}}{2}\cdot\frac{a^{x''}-a^{-x''}}{2}=\frac{1}{2}\left(a^{x'}a^{x''}+a^{-x'}a^{-x''}\right),$$

so daß

$$f(x'+x'')=f(x')f(x'')+g(x')g(x''). \tag{78}$$

Analog beweist man

$$g(x'+x'')=f(x')g(x'')+g(x')f(x''). \tag{79}$$

Als drittes untersuchen wir die Funktion

$$h(x)=\frac{g(x)}{f(x)}=\frac{a^x-a^{-x}}{a^x+a^{-x}} \qquad (a>1). \tag{80}$$

Da die Funktion $f(x)$ gerade und die Funktion $g(x)$ ungerade ist, ist die Funktion $h(x)$ ebenfalls ungerade, so daß $h(0) = 0$. Offenbar ist $h(x) > 0$ für $x > 0$ und $h(x) < 0$ für $x < 0$.

Für $x > 0$ ist die Funktion $h(x)$ wachsend. Es ist nämlich

$$h(x) = \frac{1-u}{1+u}$$

mit $u = a^{-2x}$; wächst nun x von Null bis Unendlich, so fällt u von 1 gegen Null; daher (vgl. § 5; 6, 1, 7, 4) wächst die Funktion $h(x)$ und nähert sich bei unbeschränktem Wachsen von x dem Wert 1. Aus der Tatsache, daß $h(x)$ ungerade ist, folgt sodann, daß $h(x)$ sogar überall wächst und sich dem Wert -1 nähert, wenn x gegen $-\infty$ strebt.

Für die Funktionen $f(x)$, $g(x)$ und $h(x)$ sind (üblicherweise allerdings nur für $a = \mathrm{e} = 2{,}718\ldots$, vgl. § 44) folgende Bezeichnungen und Namen gebräuchlich:

$$f(x) = \mathfrak{Cof}\,x \text{ („Kosinus hyperbolicus“)}$$

$$g(x) = \mathfrak{Sin}\,x \text{ („Sinus hyperbolicus“)}$$

$$h(x) = \mathfrak{Tg}\,x \text{ („Tangens hyperbolicus“).}^{1})$$

In diesen Bezeichnungen nehmen die Beziehungen (77) bis (80) folgende Form an:

$$\mathfrak{Cof}^2 x - \mathfrak{Sin}^2 x = 1, \tag{81}$$

$$\mathfrak{Cof}(x' + x'') = \mathfrak{Cof}\,x'\,\mathfrak{Cof}\,x'' + \mathfrak{Sin}\,x'\,\mathfrak{Sin}\,x'', \tag{82}$$

$$\mathfrak{Sin}(x' + x'') = \mathfrak{Cof}\,x'\,\mathfrak{Sin}\,x'' + \mathfrak{Sin}\,x'\,\mathfrak{Cof}\,x'',^{2}) \tag{83}$$

$$\mathfrak{Tg}\,x = \frac{\mathfrak{Sin}\,x}{\mathfrak{Cof}\,x}. \tag{84}$$

Hieraus ist die Analogie zu den üblichen („Kreis-“) Funktionen $\cos x$, $\sin x$, $\mathrm{tg}\,x$ klar zu erkennen; daher rühren auch die Namen „Kosinus“, „Sinus“ und „Tangens“.[3])

Was den Terminus „hyperbolische“ Funktionen betrifft, so erklärt er sich auf folgende Weise: Wie aus der Trigonometrie bekannt ist (vgl. auch § 25), genügen die Funktionen

$$X = \cos x, \quad Y = \sin x$$

der Gleichung

$$X^2 + Y^2 = 1.$$

Diese Gleichung stellt (in der X, Y-Ebene) einen Kreis dar, und deshalb heißen die Funktionen $\cos x$ und $\sin x$ „Kreisfunktionen“. Entsprechend genügen die Funktionen

$$X = \mathfrak{Cof}\,x \quad \text{und} \quad Y = \mathfrak{Sin}\,x$$

---

[1]) Man schreibt auch: $\cosh x$, $\sinh x$, $\tanh x$. — *Anm. d. wissenschaftl. Red.*

[2]) Die Beziehungen (82) und (83) nennt man die „Additionstheoreme“ (für die hyperbolischen Funktionen $\mathfrak{Cof}\,x$ und $\mathfrak{Sin}\,x$).

[3]) Der eigentliche Grund für diese formale Analogie zwischen Kreis- und Hyperbelfunktionen wird auf Seite 487 erläutert.

(vgl. Formel (81)) der Gleichung

$$X^2 - Y^2 = 1;$$

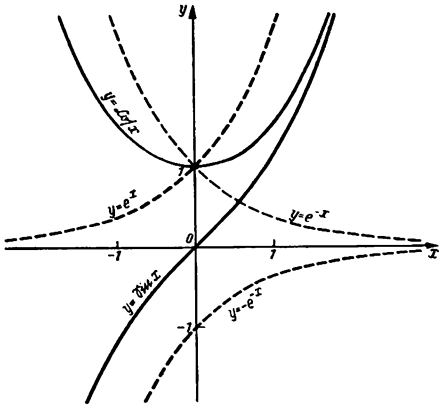


Abb. 38

weil diese Gleichung in der X, Y-Ebene eine Hyperbel beschreibt (§ 18, Beispiel 9), nennt man die Funktionen $\mathfrak{Cof}\, x$ und $\mathfrak{Sin}\, x$ „hyperbolische" Funktionen oder *Hyperbelfunktionen*.

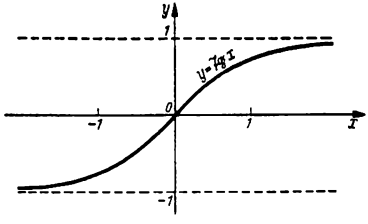


Abb. 39

Die graphischen Bilder der Funktionen $\mathfrak{Cof}\, x$ und $\mathfrak{Sin}\, x$ sind in Abb. 38, das graphische Bild der Funktion $\mathfrak{Tg}\, x$ ist in Abb. 39 dargestellt.

§ 22. Die logarithmischen Funktionen

Die *logarithmische Funktion* (kürzer: der *Logarithmus*) *zur Basis* a $(a > 0)$ ist definitionsgemäß die Umkehrfunktion der Exponentialfunktion zur Basis a.

Wir wollen annehmen, daß $a > 1$ ist. Dann erhält man aus der Exponentialfunktion

$$y = a^x \qquad (a > 1) \tag{85}$$

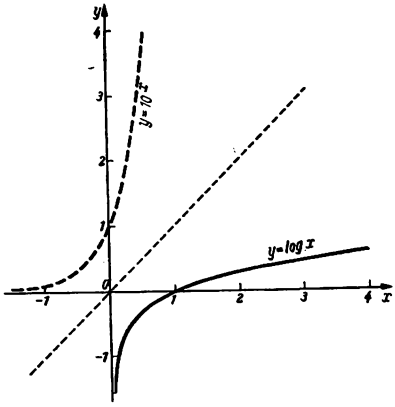


Abb. 40

den Logarithmus zur Basis a, indem man die Rolle der Veränderlichen x und y vertauscht (vgl. § 4); die Logarithmusfunktion zur Basis a

$$y = \log_a x \tag{86}$$

wird also durch die Gleichung $x = a^y$ definiert.

Das graphische Bild des Logarithmus liegt zu dem der entsprechenden Exponentialfunktion (85) in bezug auf die Winkelhalbierende $y = x$ symmetrisch (vgl. § 4). In Abb. 40 sind die Exponentialfunktion zur Basis 10 (gestrichelte Linie) und die zugehörige Logarithmusfunktion $y = \lg x$ („dekadischer Logarithmus") dargestellt.

Die Frage, zu welchen Zahlen x es eine Zahl y gibt, so daß $a^y = x$ ist, scheint eine präzise Antwort zuzulassen, und zwar sieht es so aus, als ob es genau für die positiven Zahlen x eine solche Zahl y gibt, während für negative Zahlen und für $x = 0$ eine solche Zahl y nicht existiert. Das ist auch tatsächlich der Fall,

jedoch kann eine exakte Begründung für die Existenz des Logarithmus zu jeder positiven (reellen) Zahl erst später gegeben werden (vgl. § 52). Sie dürfte jedoch auf Grund der geometrischen Vorstellungen als plausibel erscheinen.

Die bekannten Eigenschaften des Logarithmus kann man unmittelbar den in § 20 hergeleiteten Eigenschaften der Exponentialfunktionen entnehmen.

Sicher ist, daß *die Logarithmen nur für positive Zahlen existieren* (denn die zugehörige Exponentialfunktion nimmt nur positive Werte an).

Ferner kann man zeigen, daß *die Funktion* $\log_a x$ *für* $x > 0$ *wachsend ist* (vgl. § 52). *Sie wächst mit* x *über alle Grenzen, jedoch langsamer als jede positive Potenz von* x. In der Tat: Es sei δ eine beliebige positiv Zahl. Setzen wir

$$\log_a x = y, \quad x = a^y,$$

so ist

$$\frac{\log_a x}{x^\delta} = \frac{y}{a^{\delta y}} = \left(\frac{y^{\frac{1}{\delta}}}{a^y}\right)^\delta.$$

Nun wird die Exponentialfunktion a^y für hinreichend große Werte y größer als jede Potenz von y, also z. B. größer als $y^{\frac{1}{\delta}}$. Daraus folgt, daß der Ausdruck in der Klammer kleiner als Eins wird, so daß für hinreichend großes x

$$\log_a x < x^\delta$$

gilt.

Für $x < 1$ *ist* $\log_a x$ *negativ und strebt, wenn* x *gegen Null strebt, gegen* $-\infty$, *jedoch langsamer als jede negative Potenz von* x.

Aus der Funktionaleigenschaft (74) der Exponentialfunktion a^x ergibt sich unmittelbar die bekannte Funktionaleigenschaft

$$\log_a(x'x'') = \log_a x' + \log_a x'' \tag{87}$$

der Logarithmen, die dem logarithmischen Rechnen zugrunde liegt. Es erübrigt sich, auf die bekannten Folgerungen aus (87) einzugehen.

Wir wollen hier lediglich noch einige Eigenschaften der graphischen Bilder von logarithmischen Funktionen erwähnen.

Aus der Formel

$$\log_a \frac{x}{p} = \log_a x - \log_a p \tag{88}$$

folgt: *Die Dehnung des graphischen Bildes der Funktion* $\log_a x$ *in Richtung der* x*-Achse mit dem Dehnungsfaktor* p *ist gleichwertig seiner Verschiebung parallel zur* y*-Achse um die Strecke* $(-\log_a p)$.

Die Formel

$$q \log_a x = \log_{a^{1/q}} x \tag{89}$$

besagt, daß *die Dehnung der graphischen Darstellung des Logarithmus zur Basis* a *in Richtung der* y*-Achse mit dem Dehnungsfaktor* q *gleichbedeutend dem Übergang zur graphischen Darstellung des Logarithmus zur Basis* $a^{\frac{1}{q}}$ *ist.*

Logarithmieren wir die Identitäten

$$a^{\log_a x} = x, \quad b^{\log_b x} = x$$

nach den Basen b bzw. a, so erhalten wir

$$\log_a x \cdot \log_b a = \log_b x, \quad \log_b x \cdot \log_a b = \log_a x$$

oder

$$\log_a x = \frac{\log_b x}{\log_b a}, \quad \log_b x = \frac{\log_a x}{\log_a b}. \tag{90}$$

Diese Formeln geben an, wie sich der Logarithmus einer Zahl in bezug auf eine gegebene Basis durch den Logarithmus dieser Zahl in bezug auf eine andere Basis ausdrückt. Aus diesen Formeln ergibt sich insbesondere (wenn wir in der ersten $x = b$ und in der zweiten $x = a$ setzen), daß $\log_a b$ *und* $\log_b a$ *zueinander reziprok sind*:

$$\log_a b \cdot \log_b a = 1.$$

Es gilt also: *Um von der Logarithmusfunktion in bezug auf eine Basis zur Logarithmusfunktion in bezug auf eine andere Basis überzugehen, braucht man nur jene mit einem passenden konstanten Faktor* (dem „Übergangsmodul") *zu multiplizieren.*

Geometrisch bedeutet dies gerade eine Dehnung des graphischen Bildes in Richtung der y-Achse (man gelangt übrigens zum gleichen Ergebnis, wenn man in (89) $a^{\frac{1}{q}} = b$ setzt).

Bemerkung. Bisher haben wir vorausgesetzt, daß $a > 1$ ist. Der Fall $0 < a < 1$ erledigt sich durch Hinweis auf die Identität

$$\log_a x = -\log_{\frac{1}{a}} x$$

$\left(\text{ist nämlich } x = a^y, \text{ so ist } x = \left(\frac{1}{a}\right)^{-y}\right)$. Der Fall $a = 1$ ist ohne besonderes Interesse.

§ 23. Die Umkehrfunktionen der hyperbolischen Funktionen

1. Da die Funktionen $f(x)$, $g(x)$ und $h(x)$ aus § 21 sehr einfache Kombinationen von Exponentialfunktionen sind, dürfte es nicht verwunderlich erscheinen, daß ihre Umkehrfunktionen in bestimmter Weise mit dem Logarithmus zusammenhängen. Unser Ziel soll es sein, einen expliziten Ausdruck für die Umkehrfunktionen der Hyperbelfunktionen anzugeben.

Dazu müssen wir (vgl. § 4) in den Gleichungen

$$y = \frac{a^x + a^{-x}}{2}, \quad y = \frac{a^x - a^{-x}}{2}, \tag{91}$$

$$y = \frac{a^x - a^{-x}}{a^x + a^{-x}} \tag{92}$$

die Veränderlichen x und y vertauschen und die dabei entstehenden Gleichungen

$$x = \frac{a^y + a^{-y}}{2}, \quad x = \frac{a^y - a^{-y}}{2}, \tag{93}$$

$$x = \frac{a^y - a^{-y}}{a^y + a^{-y}} \tag{94}$$

nach y auflösen.

Multiplizieren wir die Gleichungen (93) mit $2a^y$, so erhalten wir nach einfachen Umformungen

$$a^{2y} - 2xa^y + 1 = 0, \quad a^{2y} - 2xa^y - 1 = 0,$$

d. h. quadratische Gleichungen für a^y, so daß

$$a^y = x \pm \sqrt{x^2 - 1}, \quad a^y = x + \sqrt{x^2 + 1}.$$

Im zweiten Fall ist das Minuszeichen vor der Wurzel überflüssig, da wir bei der Auflösung der quadratischen Gleichungen nach a^y natürlich nur ihre positiven Lösungen suchen und $x - \sqrt{x^2 + 1}$ offenbar stets negativ ist (von $x - \sqrt{x^2 - 1}$ kann man Entsprechendes nicht behaupten). Logarithmieren wir die zuletzt erhaltenen Gleichungen, so erhalten wir

$$y = \log_a(x \pm \sqrt{x^2 - 1}), \quad y = \log_a(x + \sqrt{x^2 + 1}).$$

Wegen

$$x - \sqrt{x^2 - 1} = \frac{1}{x + \sqrt{x^2 - 1}}$$

können wir schließlich die Umkehrfunktionen der Funktionen (91) in der Form

$$y = \pm \log_a(x + \sqrt{x^2 - 1}), \quad y = \log_a(x + \sqrt{x^2 + 1}) \tag{95}$$

darstellen. Die erste Funktion unterliegt der Einschränkung $x \geqq 1$ und ist doppeldeutig (wobei sich die beiden Werte nur durch ihr Vorzeichen unterscheiden), die zweite unterliegt keinen Einschränkungen und ist eindeutig.

Lösen wir schließlich die Gleichung (94) nach a^y auf, so erhalten wir

$$a^y = \sqrt{\frac{1 + x}{1 - x}}$$

(die Wurzel im arithmetischen Sinn), so daß die zu (92) gehörende Umkehrfunktion durch

$$y = \frac{1}{2} \log_a \frac{1 + x}{1 - x} \tag{96}$$

gegeben wird. Sie unterliegt der Einschränkung $-1 < x < 1$ und ist eindeutig.

Für $a = e$ ergeben die Formeln (95) und (96) explizite Ausdrücke für die Umkehrfunktionen des hyperbolischen Kosinus, des hyperbolischen Sinus und des hyperbolischen Tangens, nämlich

$$\mathfrak{Ar}\,\mathfrak{Cos}\,x = \pm \ln\left(x + \sqrt{x^2 - 1}\right) \text{ („Area Kosinus“)}$$

$$\mathfrak{Ar}\,\mathfrak{Sin}\,x = \ln\left(x + \sqrt{x^2 + 1}\right) \text{ („Area Sinus“)}$$

$$\mathfrak{Ar}\,\mathfrak{Tg}\,x = \frac{1}{2}\ln\frac{1+x}{1-x} \text{ („Area Tangens“).}^{1})$$

Die graphischen Bilder der Umkehrfunktionen der Hyperbelfunktionen (für $a = e$) ergeben sich aus denen der hyperbolischen Funktionen (vgl. Abb. 38 und 39) durch Spiegelung an der Winkelhalbierenden $y = x$.

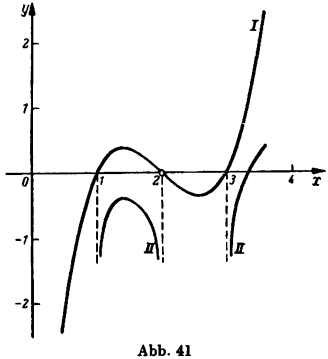


Abb. 41

2. Sehr häufig muß man sich über das Verhalten des Logarithmus einer Funktion Klarheit verschaffen, unter der Voraussetzung, daß man das Verhalten der Funktion selbst kennt. Mit anderen Worten: Aus dem graphischen Bild einer Funktion ist das graphische Bild des Logarithmus der Funktion zu konstruieren.

Das ist nicht schwer, wenn man beachtet, daß folgendes gilt:[2])

a) *Die Funktion* $\log_a f(x)$ *ist genau für die Werte x erklärt, für die die Funktion* $f(x)$ *positiv ist.*

b) *Die Funktion* $\log_a f(x)$ *ist gleich Null, positiv oder negativ, je nachdem, ob* $f(x)$ *gleich, größer oder kleiner als Eins ist; für* $a = 10$ (aber z. B. auch für $a = e$) *gilt*

$$\log_a f(x) < f(x).$$

c) *Die Funktion* $\log_a f(x)$ *wächst und fällt in jeweils denselben Intervallen wie die Funktion* $f(x)$ (vgl. § 5, 9); *sie hat ein Maximum bzw. Minimum an denselben Stellen wie* $f(x)$.

Die zuletzt erwähnte Eigenschaft ist insofern noch von besonderer Bedeutung, als es mitunter viel einfacher ist, ein Maximum oder Minimum des Logarithmus einer Funktion zu bestimmen, als für die Funktion selbst.

---

[1]) Mit ln bezeichnet man üblicherweise den Logarithmus zur Basis e („natürlicher Logarithmus“); vor allem in der höheren Mathematik, besonders in der Funktionentheorie, wird daneben die Bezeichnung log (ohne Angabe einer Basis!) verwendet. — *Anm. d. wissenschaftl. Red.*

[2]) Nach wie vor wird $a > 1$ vorausgesetzt.

In Abb. 41 sind die graphischen Bilder der Funktionen

$$f(x) = (x-1)\,(x-2)\,(x-3) \quad \text{(Kurve I)}$$

und

$$\log_{10} f(x) = \log_{10}[(x-1)\,(x-2)\,(x-3)] \quad \text{(Kurve II)}$$

dargestellt.

§ 24. Potenzen mit beliebigem reellen Exponenten

Bisher haben wir die Potenzfunktion

$$y = x^{\alpha} \tag{97}$$

für den Fall untersucht, daß der Exponent α eine positive oder negative ganze Zahl (vgl. § 7 und § 13) oder eine beliebige rationale Zahl $\frac{p}{q}$ (§ 18) ist. Im Fall eines irrationalen α versteht man unter (97) die durch „stetigen" Grenzübergang im Exponenten gewonnene Funktion. Vorausschickend bemerken wir, daß die Funktion x^{α} schon im Fall eines rationalen Exponenten (mit geradem Nenner) ihren Sinn für negative Werte von x verliert; ist der Exponent α irrational, so kann man dem Ausdruck x^{α} für negatives x weder direkt noch als Resultat eines Grenzüberganges einen vernünftigen Sinn beilegen. Aus diesem Grunde läßt man, wenn man eine Potenzfunktion x^{α} mit irrationalem Exponenten α betrachtet, die unabhängige Veränderliche x grundsätzlich nur nicht negative Werte durchlaufen, betrachtet also diese Funktion nur für $0 \leqq x < \infty$. Die Funktion (97) ist, wie man zeigen kann, für irrationales α eine transzendente Funktion, d. h., *eine Potenzfunktion x^{α} ist dann und nur dann algebraisch, wenn der Exponent α rational ist.*

Zwischen der Potenzfunktion x^{α} auf der einen Seite und der Exponential- und der Logarithmusfunktion zu einer beliebigen positiven Basis a auf der anderen Seite besteht folgender Zusammenhang:

$$x^{\alpha} = (a^{\log_a x})^{\alpha} = a^{\alpha \log_a x}. \tag{98}$$

Dieser Zusammenhang ist sowohl theoretisch als auch praktisch von großer Bedeutung.

Einerseits stellt nämlich ein stetiger Grenzübergang eine recht komplizierte Bildung (Operation) dar, die man, wenn man die Exponentialfunktionen und die Potenzfunktionen unabhängig voneinander definiert, sowohl bei den Exponentialfunktionen als auch bei den Potenzfunktionen durchzuführen hat. Es empfiehlt sich daher, den Grenzübergang nur einmal, und zwar etwa bei der Exponentialfunktion (vgl. § 52)[1]) vorzunehmen; er überträgt sich dann

[1]) Häufig wird auch ein etwas anderer Weg eingeschlagen: Man definiert zunächst für alle $x > 1$ den („natürlichen", d. h., zur Basis e gehörenden; vgl. Kap. III, § 44) Logarithmus gemäß

$$\ln x = \int_1^x \frac{d\xi}{\xi},$$

automatisch auf deren Umkehrfunktion, den Logarithmus, und dann auf Grund von (98) auch auf die Potenzfunktionen.[1])

Andererseits gibt Formel (98) eine einfache Möglichkeit, zu vorgegebenen Werten x die zugehörigen Werte x^α mit Hilfe einer Logarithmentafel zu berechnen.

Schließlich kann man aus (98) ein Verfahren für die punktweise Konstruktion des graphischen Bildes der Funktion x^α mit Hilfe graphischer Bilder der

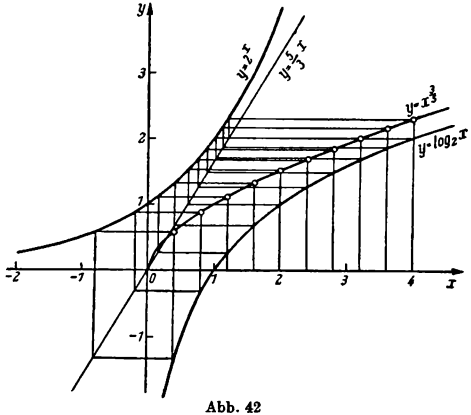


Abb. 42

Funktionen a^x und $\log_a x$ gewinnen. Der Verlauf der Konstruktion ist in Abb. 42 für den Fall $a = 2$, $\alpha = \frac{3}{5}$ dargestellt.

Bezüglich der graphischen Bilder der Potenzfunktionen zu verschiedenem Exponenten α sei noch folgendes bemerkt (vgl. Abb. 43):

1. Sie gehen alle durch den Punkt A mit den Koordinaten (1, 1).

2. Für $\alpha > 0$ ist die Funktion x^α wachsend; ihr graphisches Bild kommt aus dem Ursprung O und geht ins Unendliche; es liegt ganz in den Quadranten

d. h. als Inhalt der durch die Hyperbel $y = \frac{1}{x}$, die x-Achse und die Vertikalen durch die Punkte (1, 0) und (x, 0) begrenzten Fläche; sodann erklärt man die Exponentialfunktion zur Basis e als Umkehrfunktion des Logarithmus und die Exponentialfunktion zur Basis a mit Hilfe von Formel (90) (dort b = e gesetzt); schließlich definiert man die Potenzfunktion x^α mittels Formel (98).

[1]) Nimmt man (98) als Definition für x^α, so ergibt sich die Formel für das Logarithmieren einer Potenz als Folgerung.

OMAN und *SAT*. Ist $\alpha < 0$, so ist die Funktion x^α fallend; ihre graphische Darstellung liegt ganz in den Teilbereichen $xMAS$ und $yNAT$; sie nähert sich asymptotisch der x- und der y-Achse.

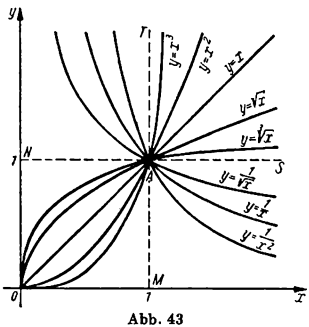


Abb. 43

3. Ist $\alpha' < \alpha''$, so liegt von den graphischen Darstellungen der Funktionen $x^{\alpha'}$ und $x^{\alpha''}$ für $0 < x < 1$ die erste oberhalb der zweiten und für $x > 1$ die zweite oberhalb der ersten.

4. Die graphischen Bilder der Funktionen $y = x^\alpha$ und $y = x^{\frac{1}{\alpha}}$ liegen zueinander symmetrisch in bezug auf die Winkelhalbierende $y = x$ (etwa x^3 und $\sqrt[3]{x}$; $\frac{1}{x^2}$ und $\frac{1}{\sqrt{x}}$).

§ 25. Die trigonometrischen Grundfunktionen Sinus und Kosinus

Wir betrachten in der X, Y-Ebene den „Einheitskreis"

$$X^2 + Y^2 = 1 \tag{99}$$

um den Ursprung O mit dem Radius 1.

Den Schnittpunkt A des Einheitskreises mit der positiven Halbgeraden der X-Achse wollen wir als „Anfangspunkt" ansehen. Wir stellen uns nun vor, daß sich ein Punkt M auf der Peripherie des Kreises bewegt (auf ihr um den Ursprung läuft). Wir sagen, *der Punkt M durchlaufe die Kreisperipherie in positivem Sinn*, wenn er sie in der Richtung durchläuft, die von der positiven Richtung der X-Achse zur positiven Richtung der Y-Achse führt, also entgegengesetzt dem Uhrzeigersinn verläuft (Abb. 44).

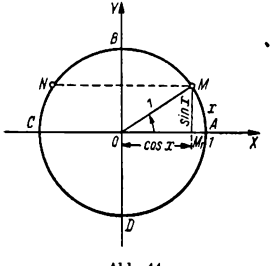


Abb. 44

Die Lage des Punktes M ist nun jeweils eindeutig durch die Länge x des Bogens bestimmt, den der Punkt M vom Punkte A aus im positiven Durchlaufungssinn der Peripherie des Einheitskreises durchlaufen mußte, um an die gegebene Stelle zu gelangen.

Wenn x von 0 bis 2π wächst, so führt der Punkt M einen vollen Umlauf auf dem Einheitskreis aus, gelangt also (vgl. Abb. 44) über B, C, D nach A

zurück. Durchläuft x die Intervalle $0 < x < \frac{\pi}{2}$, $\frac{\pi}{2} < x < \pi$, $\pi < x < \frac{3}{2}\pi$ bzw. $\frac{3}{2}\pi < x < 2\pi$, so durchläuft M die Kreisbögen $\overset{\frown}{AB}$, $\overset{\frown}{BC}$, $\overset{\frown}{CD}$ bzw. $\overset{\frown}{DA}$, d. h. die Teile der Kreisperipherie, die im ersten, zweiten, dritten bzw. vierten Quadranten der X, Y-Ebene liegen. Man nennt diese Kreisbögen (und im übertragenen Sinne auch die zugehörigen Intervalle) die „Quadranten" des Einheitskreises.[1])

Die Abszisse X bzw. die Ordinate Y des Punktes M, der so auf der Kreisperipherie liegt, daß die Länge des Bogens $\overset{\frown}{AM}$ gleich x ist, sind Funktionen von x, die man *Kosinus* bzw. *Sinus* nennt:

$$X = \cos x = OM_1, \quad Y = \sin x = M_1M. \tag{100}$$

Aus (99) ergibt sich unmittelbar die Identität

$$\cos^2 x + \sin^2 x = 1. \tag{101}$$

Diese Definition der trigonometrischen Funktionen, die im Prinzip mit der in der Schule gegebenen Definition übereinstimmt, trägt rein geometrischen Charakter. Weil sie mit der Betrachtung der Peripherie eines Kreises zusammenhängt, heißen die trigonometrischen Funktionen auch *Kreisfunktionen*. Daneben gibt es noch verschiedene analytische (in Formeln gefaßte) Definitionen des Sinus und Kosinus. Die einfachsten derartigen Definitionen sind die durch Potenzreihen (vgl. Seite 452 und Seite 480).

Eine bemerkenswerte Eigenschaft des Sinus und Kosinus, welche sie von allen Funktionen, die wir bisher untersucht haben, unterscheidet, ist ihre Periodizität. Geometrisch ist zunächst unmittelbar klar, daß der Punkt M nach einem vollen Umlauf des Einheitskreises wieder auf seinem früheren Platz erscheint; daraus ergeben sich die Identitäten

$$\sin(x + 2\pi) = \sin x, \quad \cos(x + 2\pi) = \cos x, \tag{102}$$

welche besagen, daß 2π *eine Periode der Sinus- und der Kosinusfunktion ist* (vgl. § 3).[2])

Aus der Periodizität der trigonometrischen Funktionen kann man leicht auf ihre Transzendenz schließen (vgl. § 19). In der Tat: Ist eine Funktion periodisch, so nimmt sie jeden Wert c, den sie überhaupt annehmen kann, an unendlich vielen Stellen an:

---

[1]) Verfolgt man dies weiter, so erkennt man, daß dem nächsten Intervall $2\pi < x < \frac{5}{2}\pi$ wiederum der Bogen $\overset{\frown}{AB}$ entspricht, es ist dem ersten „homolog", genauso ist das sechste Intervall $\frac{5}{2}\pi < x < 3\pi$ dem zweiten homolog usw.; das Intervall $-\frac{\pi}{2} < x < 0$ ist dem vierten Intervall homolog, das Intervall $-\pi < x < -\frac{\pi}{2}$ ist dem dritten homolog usw.

[2]) Die Funktionen $\sin x$ und $\cos x$ nehmen also an „homologen" Stellen jeweils denselben Wert an. Aus diesem Grund kann man sich in vielen Fällen auf die Betrachtung der Funktionen im Intervall $0 \leqq x < 2\pi$ (oder in einem anderen Intervall der Länge 2π) beschränken.

Zum Beispiel nimmt die Funktion $\cos x$ den Wert 1 an allen Stellen $2k\pi$ an.[1]) Eine Funktion, die diese Eigenschaft besitzt und die nicht in eine Konstante entartet, kann aber niemals algebraisch sein. Eine algebraische Funktion $y = f(x)$ genügt nämlich identisch einer Gleichung $P(x, y) = 0$, wobei P ein Polynom in x und y ist. Würde nun $f(x)$ den Wert c unendlich oft, aber nicht überall annehmen, so besäße die Gleichung

$$P(x, c) = 0,$$

die dann keine Identität in x ist, unendlich viele Wurzeln. Das ist aber unmöglich, denn sie ist eine algebraische Gleichung in x, und eine solche besitzt bekanntlich nur endlich viele Wurzeln. *Also sind die Funktionen Sinus und Kosinus transzendente Funktionen.*

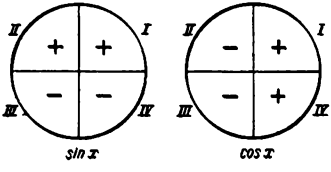
Um Klarheit über das Vorzeichen- und Wachstumsverhalten der trigonometrischen Grundfunktionen zu erhalten, muß man auf ihre Definition zurückgehen, d. h. sich auf die geometrische Anschauung stützen. Die Funktion $\sin x$ verschwindet dann und nur dann, wenn die Ordinate des Punktes M gleich Null ist, d. h., wenn M auf der X-Achse liegt. Also ist

$$\sin k\pi = 0. \tag{103}$$

Die Funktion $\cos x$ verschwindet dann und nur dann, wenn die Abszisse des Punktes M gleich Null ist, d. h., wenn M auf der Y-Achse liegt. Also ist

$$\cos\left(\frac{\pi}{2} + k\pi\right) = 0. \tag{104}$$

Das Vorzeichen der Funktionen $\sin x$ und $\cos x$ (in den verschiedenen Quadranten) ist durch das Vorzeichen der Ordinate bzw. der Abszisse des Punktes M bestimmt und aus folgendem Schema zu entnehmen:



Die Werte der Funktionen $\sin x$ und $\cos x$ variieren zwischen -1 und $+1$ einschließlich. Die Funktion $\sin x$ nimmt den maximalen Wert $+1$ an, wenn der Punkt M mit dem Punkt B zusammenfällt, so daß

$$\sin\left(\frac{\pi}{2} + 2k\pi\right) = 1, \tag{105}$$

und den minimalen Wert -1, wenn $M = D$, so daß

$$\sin\left(\frac{3}{2}\pi + 2k\pi\right) = -1. \tag{106}$$

[1]) Hier und fernerhin bedeutet k eine beliebige ganze Zahl.

Die Funktion $\cos x$ nimmt den maximalen Wert $+1$ an, wenn $M = A$, so daß

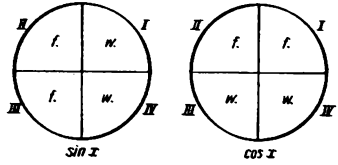
$$\cos 2k\pi = 1, \tag{107}$$

und den minimalen Wert -1, wenn $M = C$, so daß

$$\cos(\pi + 2k\pi) = -1 \tag{108}$$

(Abb. 44).

Über das Wachstumsverhalten der Funktionen $\sin x$ und $\cos x$ gibt das folgende Schema Aufschluß:



Am besten kann man sich die genannten Eigenschaften jederzeit dadurch in die Erinnerung zurückrufen, daß man direkt auf die Darstellung der trigonometrischen Funktionen am Einheitskreis zurückgreift (Abb. 44). Diese einfache Zeichnung läßt sich immer leicht auf einem Stück Papier oder in Gedanken rekonstruieren.

Stellt man die Schemata für das Vorzeichenverhalten den Schemata für das Wachstumsverhalten gegenüber, so sieht man, daß die Änderung des Sinus in „direktem Verhältnis" zum Vorzeichen des Kosinus steht (d. h., der Sinus wächst dort, wo der Kosinus positiv ist, und fällt dort, wo der Kosinus negativ ist); dagegen steht die Änderung des Kosinus in „umgekehrtem Verhältnis" zum Vorzeichen des Sinus.[1])

Wir tragen nun vom Punkt A aus einerseits den Bogen x und andererseits den Bogen $-x$ ab (Abb. 45). Die dabei erhaltenen Punkte seien M und M_1. Offenbar besitzen M und M_1 dieselbe Abszisse, während ihre Ordinaten zwar dem Betrage nach, aber nicht im Vorzeichen übereinstimmen. Hieraus folgt (vgl. § 3), daß *der Kosinus eine gerade und der Sinus eine ungerade Funktion ist*, d. h.

$$\cos(-x) = \cos x, \quad \sin(-x) = -\sin x. \tag{109}$$

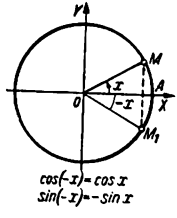


Abb. 45

[1]) Das ist nichts anderes als ein Teil des Inhalts der „Differentiationsregeln" $(\sin x)' = \cos x$, $(\cos x)' = -\sin x$ (vgl. S. 296 Formeln 5. und 6.).

Als nächstes betrachten wir die Punkte M und M_1, die sich als Endpunkte der Bögen x und $x+\pi$ ergeben (Abb. 46). Die Punkte M und M_1 liegen sich im Einheitskreis diametral gegenüber, also zum Ursprung O spiegelbildlich. Daher unterscheiden sich sowohl die Abszissen als auch die Ordinaten von M und M_1 nur durch ihr Vorzeichen. Folglich ist

$$\cos(x+\pi) = -\cos x, \quad \sin(x+\pi) = -\sin x. \tag{110}$$

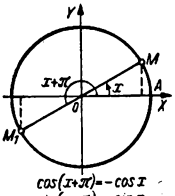


Abb. 46

Schließlich betrachten wir die Punkte M und M_1, die zu den Bögen x und $\frac{\pi}{2}-x$ gehören (Abb. 47). Man sieht leicht ein, daß die Punkte M und M_1 zueinander symmetrisch in bezug auf die Winkelhalbierende $y=x$ liegen. Also ist (§ 4) die Abszisse von M_1 gleich der Ordinate von M und die Ordinate von M_1 gleich der Abszisse von M. Somit gilt

$$\cos\left(\frac{\pi}{2}-x\right) = \sin x, \quad \sin\left(\frac{\pi}{2}-x\right) = \cos x. \tag{111}$$

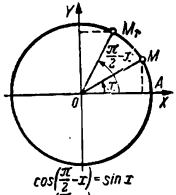


Abb. 47

Aus (111), (109) und (110) gewinnt man leicht

$$\cos x = \sin\left(x+\frac{\pi}{2}\right). \tag{112}$$

Bevor wir zur Beschreibung der graphischen Bilder der Funktionen $\sin x$ und $\cos x$ übergehen, bemerken wir, daß man nur die graphische Darstellung des Sinus im Intervall $0 \leqq x \leqq \frac{\pi}{2}$ benötigt, um sie unter Verwendung der Formeln (109) und (110) mittels elementarer Transformationen auf die ganze x-Achse fortzusetzen; (112) liefert dann die Darstellung des Kosinus.

In der Tat: Auf Grund von (109) liegt das graphische Bild symmetrisch zum Ursprung O, so daß man es durch Spiegelung am Ursprung unmittelbar vom Intervall $0 \leqq x \leqq \frac{\pi}{2}$ auf das Intervall $-\frac{\pi}{2} \leqq x \leqq 0$ fortsetzen kann. Ferner folgt aus (110), daß die graphische Darstellung der Funktion $\sin x$ symmetrisch zur Geraden $x=\frac{\pi}{2}$ liegt; ersetzt man nämlich in (110) x durch $x-\frac{\pi}{2}$ und beachtet man, daß der Sinus eine ungerade Funktion ist [vgl. (109)], so erhält man unmittelbar

$$\sin\left(\frac{\pi}{2}+x\right) = \sin\left(\frac{\pi}{2}-x\right).$$

Dies gestattet es, das graphische Bild von $\sin x$ auf das Intervall $-\frac{\pi}{2} \leqq x \leqq \frac{3}{2}\pi$ fortzusetzen. Die Fortsetzung auf die ganze x-Achse gelingt schließlich dadurch, daß man die Periodizität der Funktion $\sin x$ ausnutzt [Formel (102)]. Was

die graphische Darstellung der Funktion $\cos x$ anbetrifft, so gewinnt man sie, wie aus (112) ersichtlich ist, aus dem graphischen Bild des Sinus durch eine Parallelverschiebung längs der x-Achse um die Strecke $\left(-\frac{\pi}{2}\right)$.

Zur punktweisen Konstruktion des graphischen Bildes der Funktion $\sin x$ über dem Intervall $0 \leqq x \leqq \frac{\pi}{2}$ teilt man einerseits (Abb. 48a) den ersten Quadranten des Einheitskreises in der X, Y-Ebene und zum anderen (Abb. 48b)

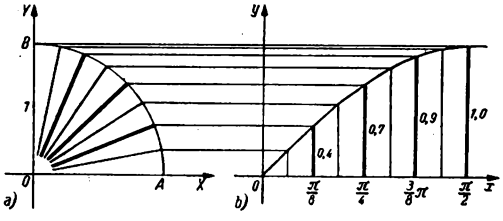


Abb. 48

das Intervall $0 \leqq x \leqq \frac{\pi}{2}$ in der x, y-Ebene in jeweils m gleiche Teile; sodann zieht man durch die Teilpunkte die vertikalen Geraden und trägt auf den Vertikalen in der x, y-Ebene die Ordinaten der entsprechenden Teilpunkte des ersten Quadranten des Einheitskreises ab. Die erhaltenen Punkte gehören dem graphischen Bild der Funktion $\sin x$ an.

Wenn es darum geht, schnell eine (nicht allzu genaue) graphische Darstellung des Sinus zu entwerfen, so ist die Konstruktion „aus vier Punkten", die dem Fall $m = 4$ entspricht, recht vorteilhaft (vgl. die fett gedruckten Ordinaten in Abb. 48b). Die Längen der Ordinaten (die zu den Werten von $\sin x$ für $x = 0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{3}{8}\pi, \frac{\pi}{2}$ gehören) sind annähernd gleich 0; 0,4; 0,7; 0,9; 1.[1])

Hierbei sei noch auf folgendes hingewiesen: Sind M, M', M'' drei Punkte des ersten Quadranten des Einheitskreises in der X, Y-Ebene, die in der Nähe des Punktes A liegen, also zu kleinen Werten der unabhängigen Veränderlichen x gehören, so sind zwar die Ordinaten dieser Punkte, wie man aus Abb. 49a erkennt, etwas kleiner als die Längen der zugehörigen Bögen, jedoch strebt das Verhältnis von Ordinate zu Bogenlänge, wenn der Punkt M gegen den Punkt A strebt, gegen Eins. Dem entspricht bei der graphischen Darstellung der Funktion $\sin x$ in der x, y-Ebene (Abb. 49b), daß zwar die Ordinate PM

[1]) Genauer: 0; 0,383; 0,707; 0,924; 1.

eines Punktes M der Sinuskurve ständig etwas kleiner als seine Abszisse OP (oder was dasselbe ist, als die Ordinate PQ des Punktes Q auf der Winkelhalbierenden $y = x$) ist, daß jedoch das Verhältnis $\frac{PM}{OP}$ $\left(\text{oder } \frac{PM}{PQ}\right)$ gegen Eins strebt, wenn der Punkt M gegen den Punkt O rückt.

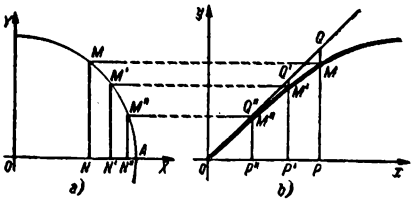


Abb. 49

Es gilt also: *Das graphische Bild der Funktion* $\sin x$ *liegt* (*für* $x > 0$) *unterhalb der Winkelhalbierenden* $y = x$ (d. h. $\sin x < x$), *schmiegt sich aber in der Nähe des Punktes O immer besser an diese an* („berührt sie im Punkte O").

In Abb. 50 sind die graphischen Bilder der Funktionen

$$y = \sin x \quad \text{und} \quad y = \cos x \tag{113}$$

dargestellt. Man nennt diese Kurven die „Sinuskurve" und die „Kosinuskurve" (auch „Sinuslinie" bzw. „Kosinuslinie").

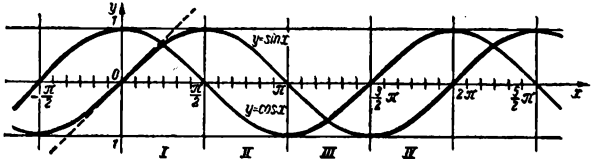


Abb. 50

Die „Additionstheoreme" für die Funktionen $\sin x$ und $\cos x$,

$$\sin(x' + x'') = \sin x' \cos x'' + \cos x' \sin x'',$$

$$\cos(x' + x'') = \cos x' \cos x'' - \sin x' \sin x'',$$

dürften allgemein bekannt sein, so daß wir uns hier nicht näher mit ihnen zu beschäftigen brauchen (vgl. Seite 484).

§ 26. Einfache harmonische Schwingungen

Die graphische Darstellung der Funktion

$$y = C\sin\lambda(x-c) \qquad (C>0,\quad \lambda>0) \tag{114}$$

erhält man aus der Sinuskurve mit Hilfe folgender Transformationen:

1. Stauchung in Richtung der x-Achse mit dem Stauchungsfaktor λ $\left(\text{Dehnung mit dem Dehnungsfaktor } \frac{1}{\lambda}\right)$;

2. Dehnung in Richtung der y-Achse mit dem Dehnungsfaktor C;

3. Verschiebung parallel zur x-Achse um die Strecke c.

Jede Funktion der Form (114) heißt ebenso wie ihr graphisches Bild eine *einfache harmonische Schwingung* (kurz: *harmonische Schwingung*). Die zugehörige Kurve nennt man häufig auch *sinusförmig*.

Für die Parameter C, λ, c in (114) sind die folgenden Bezeichnungen üblich: C heißt die *Amplitude*, λ die *Frequenz* und c die *Phase*.

Infolge der Stauchung mit dem Stauchungsfaktor λ hat die Funktion (114) nicht mehr die Periode 2π, sondern die Periode $\frac{2\pi}{\lambda}$. Setzt man $\omega = \frac{2\pi}{\lambda}$, so wird $\lambda = \frac{2\pi}{\omega}$, und Gleichung (114) bekommt die Form

$$y = C\sin\frac{2\pi}{\omega}(x-c) \qquad (C>0,\quad \omega>0). \tag{115}$$

Die Frequenz und die Periode einer harmonischen Schwingung sind umgekehrt proportional; ihr Produkt ist gleich 2π:

$$\lambda\omega = 2\pi. \tag{116}$$

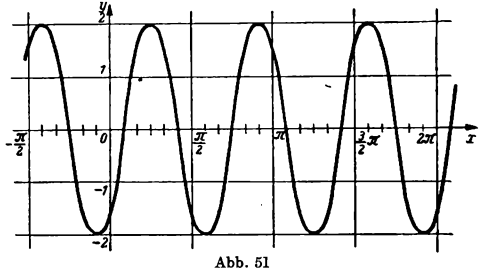


Abb. 51

In Abb. 51 ist das graphische Bild der harmonischen Schwingung

$$y = 2\sin\left(3x - \frac{\pi}{4}\right)$$

mit der Amplitude $C=2$, der Frequenz $\lambda=3$ $\left(\text{der Periode } \omega = \frac{2\pi}{3}\right)$ und der Phase $c = \frac{\pi}{12}$ dargestellt. Die sinusförmige Kurve ist dabei aus vier Punkten konstruiert.

Da wir uns im weiteren zunächst nur mit harmonischen Schwingungen gleicher Periode beschäftigen werden, nehmen wir der Einfachheit halber an, daß $\omega = 2\pi$, d. h. $\lambda = 1$ ist.

Beispiele für einfache harmonische Schwingungen mit der Periode 2π sind

$$u = A\cos x \quad \text{und} \quad v = B\sin x. \tag{117}$$

Man sieht leicht, daß sich die Phasen der harmonischen Schwingungen (117) (welches Vorzeichen auch A und B besitzen mögen) um eine Viertelperiode, also um $\frac{\pi}{2}$ unterscheiden.

Jede einfache harmonische Schwingung

$$y = C\sin(x - c)$$

mit der Periode 2π kann als Summe aus zwei harmonischen Schwingungen (117) *dargestellt werden*:

$$C\sin(x - c) = A\cos x + B\sin x. \tag{118}$$

Dazu genügt es zu beachten, daß

$$C\sin(x - c) = C(\sin x\cos c - \cos x\sin c);$$

setzt man

$$A = -C\sin c, \quad B = C\cos c, \tag{119}$$

so ergibt sich (118).

Umgekehrt ist aber auch jede Summe aus zwei einfachen harmonischen Schwingungen gleicher Periode eine einfache harmonische Schwingung mit eben dieser Periode.

Beweis. Zunächst nehmen wir an, daß die gegebenen harmonischen Schwingungen von der Form

$$u = A\cos x, \quad v = B\sin x$$

sind. Es ist zu zeigen, daß es eine Konstante $C > 0$ und eine Konstante c gibt, so daß (118) gilt. Hierfür haben wir die Gleichungen (119) bei gegebenem A und B nach C und c aufzulösen. Zu diesem Zweck quadrieren wir die Gleichungen (119) und addieren die erhaltenen Resultate. Es ergibt sich

$$C = \sqrt{A^2 + B^2}. \tag{120}$$

Dann kann man einen eindeutig bestimmten Wert c im Intervall $0 \leqq c < 2\pi$ so finden, daß

$$\sin c = -\frac{A}{\sqrt{A^2 + B^2}},$$

$$\cos c = \frac{B}{\sqrt{A^2 + B^2}}.$$

Die erhaltenen Zahlen C und c leisten das Verlangte.

Im allgemeinen Fall stellen wir die gegebenen Funktionen

$$y_1 = C_1\sin(x - c_1), \quad y_2 = C_2\sin(x - c_2)$$

zunächst in der Form (118) dar:

$$y_1 = (-C_1 \sin c_1) \cos x + (C_1 \cos c_1) \sin x,$$
$$y_2 = (-C_2 \sin c_2) \cos x + (C_2 \cos c_2) \sin x.$$

Die Summe der Funktionen ist dann

$$y = y_1 + y_2 = -(C_1 \sin c_1 + C_2 \sin c_2) \cos x + (C_1 \cos c_1 + C_2 \cos c_2) \sin x,$$

und diese ist von der Form (118), wenn man

$$A = -(C_1 \sin c_1 + C_2 \sin c_2), \quad B = C_1 \cos c_1 + C_2 \cos c_2$$

setzt.

Die Summe einfacher harmonischer Schwingungen mit verschiedenen Perioden ist bereits keine einfache harmonische Schwingung mehr. Sind nämlich die Frequenzen der zu addierenden Schwingungen kommensurabel, d. h. ist

$$\lambda_1 = m_1 \lambda, \quad \lambda_2 = m_2 \lambda,$$

wobei m_1 und m_2 teilerfremde positive ganze Zahlen sind, so ist die Summe der Schwingungen

$$y_1 = C_1 \sin \lambda_1 (x - c_1) \quad \text{und} \quad y_2 = C_2 \sin \lambda_2 (x - c_2)$$

gleich

$$y = C_1 \sin \lambda_1 (x - c_1) + C_2 \sin \lambda_2 (x - c_2),$$

und dies stellt eine sogenannte *zusammengesetzte harmonische Schwingung*[1]) mit der Periode $\omega = \frac{2\pi}{\lambda}$ dar. (Sind die Phasen inkommensurabel, so ist die Summe keine periodische Funktion mehr.)

In Abb. 52 ist die (punktweise konstruierte) Summe der Schwingungen

$$y_1 = \sin 2x \quad \text{und} \quad y_2 = \sin 3x$$

mit den Perioden 2 und 3 dargestellt (die Schwingungen y_1 und y_2 gestrichelt, die Summe durchgehend).

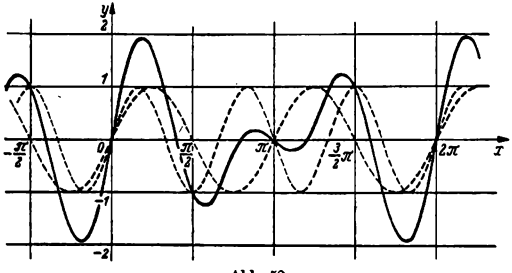


Abb. 52

[1]) Zur Definition der zusammengesetzten harmonischen Schwingungen vgl. § 27.

§ 27. Trigonometrische Polynome

Entsprechend wie jede Funktion

$$f(x) = c_0 + c_1 x + c_2 x^2 + \cdots + c_n x^n$$

(wobei n eine positive ganze Zahl und $c_0, \ldots, c_n$ konstante Koeffizienten sind) ein (rationales) Polynom in der Veränderlichen x genannt wird, hat sich für die Funktionen, die durch eine Gleichung der Form

$$f(x) = a_0 + (a_1 \cos x + b_1 \sin x) + (a_2 \cos 2x + b_2 \sin 2x) + \cdots + (a_n \cos nx + b_n \sin nx) \tag{121}$$

gegeben werden, wobei n eine positive ganze Zahl ist und $a_0, \ldots, a_n, b_1, \ldots, b_n$ konstante Koeffizienten sind, ein besonderer Name eingebürgert; man nennt sie *trigonometrische Polynome* (in der Veränderlichen x). Ist $a_n^2 + b_n^2 \neq 0$, so heißt n die *Ordnung* des trigonometrischen Polynoms (121). Der Koeffizient a_0 wird das *absolute Glied*, der Summand $a_1 \cos x + b_1 \sin x$ das *erste Glied*, der Summand $a_2 \cos 2x + b_2 \sin 2x$ das *zweite Glied* usw. genannt. Die graphische Darstellung des m-ten Gliedes eines trigonometrischen Polynoms ist (für $m \geqq 1$) eine einfache harmonische Schwingung der Frequenz m $\left(\text{also mit der Periode } \frac{2\pi}{m}\right)$. Die graphische Darstellung jedes trigonometrischen Polynoms eines Grades $n \geqq 2$ wird eine *zusammengesetzte harmonische Schwingung* genannt. Ein Beispiel für eine zusammengesetzte harmonische Schwingung findet sich im vorangehenden Paragraphen (Abb. 52).

Da mit einer Zahl ω auch alle Vielfachen von ω Perioden einer gegebenen Funktion sind, besitzt jedes Glied von (121) die Periode 2π. Weil andererseits jede Summe von Funktionen mit der Periode ω eine Funktion mit der Periode ω ist, hat das Polynom (121) selbst die Periode 2π.[1])

Man sieht leicht, daß folgendes gilt:

1. Die Summe zweier (oder mehrerer) trigonometrischer Polynome, deren Ordnung höchstens n ist, ist ein trigonometrisches Polynom, dessen Ordnung ebenfalls höchstens n ist.

---

[1]) Allgemeiner versteht man unter einem trigonometrischen Polynom bzw. einer zusammengesetzten harmonischen Schwingung auch Funktionen der Form

$$a_0 + \left(a_1 \cos \frac{2\pi x}{\omega} + b_1 \sin \frac{2\pi x}{\omega}\right) + \cdots + \left(a_n \cos \frac{2\pi n x}{\omega} + b_n \sin \frac{2\pi n x}{\omega}\right),$$

die aus den oben betrachteten durch Dehnung in Richtung der x-Achse mit dem Dehnungsfaktor $\frac{\omega}{2\pi}$ entstehen. Diese Funktionen haben natürlich die Periode ω. Wir wollen uns hier jedoch der Einfachheit halber auf den Fall $\omega = 2\pi$ beschränken.

2. Jedes Vielfache eines trigonometrischen Polynoms von der Ordnung n ist ein trigonometrisches Polynom von der Ordnung n.

Wir wollen nun zeigen, daß folgendes gilt: *Ist $P(u, v)$ ein rationales Polynom in den Veränderlichen u und v, so ist $P(\cos x, \sin x)$ ein trigonometrisches Polynom in x, wobei die Ordnung dieses Polynoms nicht größer als der Grad von $P(u, v)$ in bezug auf u und v ist.*

Der Beweis hierfür erfolgt in mehreren Schritten:

1. Die Funktionen

$$\cos px \cos qx, \quad \sin px \cos qx, \quad \cos px \sin qx, \quad \sin px \sin qx \tag{122}$$

(wobei p und q beliebige positive ganze Zahlen sind) sind trigonometrische Polynome der Ordnung $p + q$. Dies ergibt sich unmittelbar aus den elementaren Identitäten

$$\cos px \cos qx = \frac{1}{2}[\cos(p+q)x + \cos(p-q)x],$$

$$\sin px \cos qx = \frac{1}{2}[\sin(p+q)x + \sin(p-q)x],$$

$$\cos px \sin qx = \frac{1}{2}[\sin(p+q)x - \sin(p-q)x],$$

$$\sin px \sin qx = \frac{1}{2}[-\cos(p+q)x + \cos(p-q)x].$$

2. Das Produkt zweier trigonometrischer Polynome $f(x)$ und $g(x)$ der Ordnung r bzw. s ist ein trigonometrisches Polynom der Ordnung $r + s$. Zum Beweis sei

$$f(x) = a_0 + (a_1 \cos x + b_1 \sin x) + \cdots + (a_r \cos rx + b_r \sin rx)$$

und

$$g(x) = a_0' + (a_1' \cos x + b_1' \sin x) + \cdots + (a_s' \cos sx + b_s' \sin sx).$$

Das Produkt der trigonometrischen Polynome $f(x)$ und $g(x)$ ist eine Linearkombination aus endlich vielen Gliedern der Form (122) mit konstanten Koeffizienten und daher wegen der vorangehenden Bemerkungen ebenfalls ein trigonometrisches Polynom. Seine Ordnung ist offenbar nicht größer als $r + s$. Sie kann aber, wenn $f(x)$ die Ordnung r und $g(x)$ die Ordnung s besitzt, auch nicht kleiner sein, denn Glieder der Ordnung $r + s$ treten nur bei der Multiplikation der Glieder $a_r \cos rx + b_r \sin rx$ und $a_s' \cos sx + b_s' \sin sx$ auf; und zwar sind es genau die folgenden:

$$\frac{1}{2}\{(a_r a_s' - b_r b_s')\cos(r+s)x + (a_r b_s' + b_r a_s')\sin(r+s)x\};$$

die hier bei $\cos(r+s)x$ und $\sin(r+s)x$ auftretenden Koeffizienten können nicht beide gleich Null sein, denn wegen $a_r^2 + b_r^2 > 0$ würde aus

$$a_r a_s' - b_r b_s' = 0,$$

$$b_r a_s' + a_r b_s' = 0$$

sofort

$$a_s' = b_s' = 0$$

folgen, was nicht der Fall ist.

3. Jede Funktion

$$\cos^h x \sin^k x \tag{123}$$

(wobei h und k beliebige nicht negative ganze Zahlen sind) ist ein trigonometrisches Polynom der Ordnung $h + k$.

In der Tat: Jeder der $h + k$ Faktoren in (123) ist $\cos x$ oder $\sin x$, also ein trigonometrisches Polynom der Ordnung 1.

4. Da $P(\cos x, \sin x)$ für jedes rationale Polynom $P(u, v)$ eine Summe aus endlich vielen Gliedern der Form (123) mit konstanten Koeffizienten ist, ist auf Grund der vorangehenden Bemerkungen $P(\cos x, \sin x)$ ein trigonometrisches Polynom, dessen Ordnung nicht größer sein kann als das Maximum der Zahlen $h + k$ für die Glieder (123) aus $P(\cos x, \sin x)$, also nicht größer als der Grad von $P(u, v)$ in bezug auf u und v. Die Ordnung von $P(\cos x, \sin x)$ kann aber durchaus kleiner als dieser Grad sein (wie dies das Beispiel $P(u, v) = u^2 + v^2$ zeigt).

§ 28. Tschebyschewsche Polynome

Wir wollen nun zeigen, daß auch die Umkehrung des im vorangehenden Paragraphen bewiesenen Satzes gilt:
Jedes trigonometrische Polynom

$$f(x) = a_0 + (a_1 \cos x + b_1 \sin x) + (a_2 \cos 2x + b_2 \sin 2x) + \cdots + (a_n \cos nx + b_n \sin nx) \quad (a_n^2 + b_n^2 \neq 0) \tag{124}$$

der Ordnung n läßt sich in der Form $f(x) = P(\cos x, \sin x)$ darstellen, wobei $P(u, v)$ ein rationales Polynom in u und v genau vom Grad n ist.

Da eine Summe von Polynomen und das Produkt eines Polynoms mit einer konstanten Zahl wieder Polynome sind, genügt es, die Behauptung über die Darstellbarkeit für die trigonometrischen Polynome der Form $\cos nx$ und $\sin nx$ (wobei n eine beliebige positive ganze Zahl ist) zu beweisen und zu zeigen, daß allgemein auch die Behauptung über den Grad von $P(u, v)$ erfüllt werden kann.

Wir zeigen zunächst, daß es zu jeder positiven ganzen Zahl n ein Polynom $T_n(u)$ vom Grad n und ein Polynom $U_n(u)$ vom Grad $n - 1$ gibt, so daß identisch in x

$$\cos nx = T_n(\cos x), \tag{125}$$

$$\sin nx = U_n(\cos x) \sin x \tag{126}$$

ist.

Um die Art der hierbei auftretenden Polynome $T_n(u)$ und $U_n(u)$ genau zu erkennen, wollen wir zunächst die Spezialfälle $n = 1, 2, 3$ behandeln.

Im Fall $n = 1$ leisten offenbar die Polynome

$$T_1(u) = u, \quad U_1(u) = 1 \tag{127}$$

das Gewünschte.

Im Fall $n = 2$ ist bekanntlich

$$\cos 2x = \cos^2 x - \sin^2 x = 2\cos^2 x - 1,$$
$$\sin 2x = 2\sin x \cos x = 2\cos x \sin x,$$

so daß

$$T_2(u) = 2u^2 - 1, \quad U_2(u) = 2u \tag{128}$$

gesetzt werden kann.

Im Fall $n = 3$ gilt schließlich

$$\begin{aligned}\cos 3x &= \cos(2x + x) = \cos 2x \cos x - \sin 2x \sin x \\ &= (2\cos^2 x - 1)\cos x - 2\cos x \sin^2 x \\ &= (2\cos^2 x - 1)\cos x - 2\cos x(1 - \cos^2 x) = 4\cos^3 x - 3\cos x,\end{aligned}$$
$$\begin{aligned}\sin 3x &= \sin(2x + x) = \sin 2x \cos x + \cos 2x \sin x \\ &= 2\cos^2 x \sin x + (2\cos^2 x - 1)\sin x = (4\cos^2 x - 1)\sin x,\end{aligned}$$

so daß

$$T_3(u) = 4u^3 - 3u, \quad U_3(u) = 4u^2 - 1. \tag{129}$$

In dieser Weise kann man nun offenbar fortfahren.

Den strengen Nachweis führen wir durch vollständige Induktion. Wir nehmen also an, daß die Existenz der Polynome $T_n(u)$ und $U_n(u)$ bereits gesichert ist und daß sie bereits ausgerechnet vorliegen; wir zeigen, daß dann auch die Polynome $T_{n+1}(u)$ und $U_{n+1}(u)$ existieren und geben an, wie sie berechnet werden können.

Nach Induktionsvoraussetzung ist einerseits

$$\begin{aligned}\cos(n+1)x &= \cos(nx + x) = \cos nx \cos x - \sin nx \sin x \\ &= T_n(\cos x)\cos x - U_n(\cos x)\sin^2 x \\ &= T_n(\cos x)\cos x - U_n(\cos x)(1 - \cos^2 x)\end{aligned}$$

und andererseits

$$\begin{aligned}\sin(n+1)x &= \sin(nx + x) = \sin nx \cos x + \cos nx \sin x \\ &= U_n(\cos x)\sin x \cos x + T_n(\cos x)\sin x \\ &= \{U_n(\cos x)\cos x + T_n(\cos x)\}\sin x.\end{aligned}$$

Da die Polynome $T_n(u)$ und $U_n(u)$ den Grad n bzw. $n-1$ besitzen, haben die Polynome

$$T_n(u)\,u - U_n(u)(1 - u^2) \quad \text{und} \quad U_n(u)\,u + T_n(u)$$

den Grad $n+1$ bzw. n. Bezeichnen wir diese Polynome mit T_{n+1} bzw. U_{n+1},

$$\left.\begin{aligned}T_{n+1}(u) &= uT_n(u) + (u^2 - 1)\,U_n(u) \\ U_{n+1}(u) &= T_n(u) + u\,U_n(u),\end{aligned}\right\} \tag{130}$$

so gilt

$$\cos(n+1)x = T_{n+1}(\cos x),$$
$$\sin(n+1)x = U_{n+1}(\cos x)\sin x,$$

d. h., die so erhaltenen Polynome $T_{n+1}(u)$ und $U_{n+1}(u)$ leisten das Verlangte.

Die Polynome $T_1(u)$ und $U_1(u)$ sind direkt durch (127) gegeben. Die Polynome $T_n(u)$, $U_n(u)$ für $n \geqq 2$ können mittels der Rekursionsformeln (130) berechnet werden. Für $n = 2, 3, 4, 5, 6, \ldots$ erhält man aus (130) sofort (128), (129) und weiter

$$T_2(u) = 8u^4 - 8u^2 + 1, \qquad U_2(u) = 8u^3 - 4u,$$
$$T_5(u) = 16u^5 - 20u^3 + 5u, \qquad U_5(u) = 16u^4 - 12u^2 + 1,$$
$$T_6(u) = 32u^6 - 48u^4 + 18u^2 - 1, \qquad U_6(u) = 32u^5 - 32u^3 + 6u$$

usw.

Aus den Formeln (125) und (126) folgt unmittelbar, daß alle Polynome $T_{2n}(u)$ und $U_{2n+1}(u)$ gerade und alle Polynome $T_{2n+1}(u)$ und $U_{2n}(u)$ ungerade sind.

Ferner beweist man leicht durch vollständige Induktion, daß der höchste Koeffizient der Polynome $T_n(u)$ und $U_n(u)$ gleich 2^{n-1} ist, so daß

$$\begin{aligned} T_n(u) &= 2^{n-1}u^n + \cdots, \\ U_n(u) &= 2^{n-1}u^{n-1} + \cdots. \end{aligned} \tag{131}$$

Die Polynome $T_n(u)$ heißen Tschebyschewsche *Polynome erster Art*, die Polynome $U_n(u)$ Tschebyschewsche *Polynome zweiter Art*.

Unter Verwendung der Tschebyschewschen Polynome können wir nun dem Polynom (124) die Form

$$f(x) = \{a_0 + a_1 T_1(\cos x) + a_2 T_2(\cos x) + \cdots + a_n T_n(\cos x)\}$$
$$+ \sin x\{b_1 U_1(\cos x) + b_2 U_2(\cos x) + \cdots + b_n U_n(\cos x)\}$$

geben. Die Ausdrücke in den geschweiften Klammern auf der rechten Seite sind Polynome in $u = \cos x$:

$$\left.\begin{aligned} L(u) &= a_0 + a_1 T_1(u) + a_2 T_2(u) + \cdots + a_n T_n(u), \\ M(u) &= \quad b_1 U_1(u) + b_2 U_2(u) + \cdots + b_n U_n(u), \end{aligned}\right\} \tag{132}$$

so daß

$$f(x) = L(\cos x) + \sin x\, M(\cos x). \tag{133}$$

Ist nun $a_n \neq 0$, so ist $L(u)$ vom Grad n; ist $b_n \neq 0$, so ist $M(u)$ vom Grad $n-1$. Da nun nach Voraussetzung wenigstens einer der Koeffizienten a_n oder b_n von Null verschieden ist, ist ganz gewiß $L(u)$ vom Grad n oder $M(u)$ vom Grad $n-1$, so daß $L(u) + vM(u)$ in bezug auf u und v vom Grad n ist.

Das Polynom $L(u) + vM(u)$ aus (133) genügt also allen Anforderungen, die durch den behaupteten Satz an $P(u, v)$ gestellt werden; selbstverständlich ist es nicht das einzige Polynom, welches diesen Forderungen genügt.[1])

Wir wollen noch eine andere mögliche Form für das Polynom $P(u, v)$ angeben.

---

[1]) Dies folgt z. B. daraus, daß man an die Stelle von u^2 nach Belieben $1 - v^2$ setzen kann.

Ersetzen wir in (125) und (126) die Veränderliche x durch $\frac{\pi}{2} - x$, so erhalten wir die neuen Gleichungen

$$\left.\begin{aligned} \cos n\left(\frac{\pi}{2} - x\right) &= T_n(\sin x), \\ \sin n\left(\frac{\pi}{2} - x\right) &= U_n(\sin x)\cos x, \end{aligned}\right\} \tag{134}$$

welchen man durch Aufspaltung in die Fälle eines geraden und ungeraden n auch die folgende Form geben kann:

$$\cos 2nx = (-1)^n T_{2n}(\sin x), \quad \cos(2n+1)x = (-1)^n U_{2n+1}(\sin x)\cos x,$$
$$\sin 2nx = (-1)^{n+1} U_{2n}(\sin x)\cos x, \quad \sin(2n+1)x = (-1)^n T_{2n+1}(\sin x).$$

Hieraus ergibt sich

$$\begin{aligned} f(x) = \{a_0 &+ b_1 T_1(\sin x) - a_2 T_2(\sin x) - b_3 T_3(\sin x) + a_4 T_4(\sin x) + \cdots\} \\ &+ \cos x\{a_1 U_1(\sin x) + b_2 U_2(\sin x) - a_3 U_3(\sin x) - b_4 U_4(\sin x) \\ &\qquad + a_5 U_5(\sin x) + \cdots\} \end{aligned}$$

oder, wenn wir

$$\left.\begin{aligned} L_1(v) &= a_0 + b_1 T_1(v) - a_2 T_2(v) - b_3 T_3(v) + a_4 T_4(v) + \cdots, \\ M_1(v) &= \quad a_1 U_1(v) + b_2 U_2(v) - a_3 U_3(v) - b_4 U_4(v) + \cdots \end{aligned}\right\} \tag{135}$$

setzen,

$$f(x) = L_1(\sin x) + \cos x\, M_1(\sin x). \tag{136}$$

Abschließend wollen wir noch folgende Spezialfälle behandeln:

1. Ist $b_1 = b_2 = \cdots = b_n = 0$, so ist $f(x)$ ein trigonometrisches Polynom der Form

$$f(x) = a_0 + a_1\cos x + a_2\cos 2x + \cdots + a_n\cos nx;$$

in diesem Fall kann nach (133) $f(x)$ als Polynom in $\cos x$ dargestellt werden.

2. Ist $a_0 = a_1 = \cdots = a_n = 0$, so ist $f(x)$ ein trigonometrisches Polynom der Form

$$f(x) = b_1\sin x + b_2\sin 2x + \cdots + b_n\sin nx;$$

in diesem Fall kann nach (133) $f(x)$ in der Form

$$f(x) = \sin x \cdot P(\cos x)$$

dargestellt werden, wobei $P(\cos x)$ ein Polynom in $\cos x$ ist.

3. Ist $a_1 = a_3 = a_5 = \cdots = 0$ und $b_2 = b_4 = b_6 = \cdots = 0$, so ist $f(x)$ ein trigonometrisches Polynom der Form

$$f(x) = a_0 + b_1\sin x + a_2\cos 2x + b_3\sin 3x + \cdots;$$

in diesem Fall ist nach (136) $f(x)$ ein Polynom in $\sin x$.

4. Ist $a_0 = a_2 = a_4 = \cdots = 0$ und $b_1 = b_3 = b_5 = \cdots = 0$, so ist $f(x)$ ein trigonometrisches Polynom der Form

$$f(x) = a_1 \cos x + b_2 \sin 2x + a_3 \cos 3x + b_4 \sin 4x + \cdots;$$

in diesem Fall kann nach (136) $f(x)$ in der Form

$$f(x) = \cos x \cdot P(\sin x)$$

dargestellt werden, wobei $P(\sin x)$ ein Polynom in $\sin x$ ist.

Man prüft leicht nach, daß alle diese Behauptungen auch umkehrbar sind.

§ 29. Der Tangens und andere gebrochene trigonometrische Funktionen

Wir wollen nun zur Untersuchung *gebrochener trigonometrischer Funktionen* übergehen. Hierunter versteht man rationale Funktionen der trigonometrischen Grundfunktionen $\sin x$ und $\cos x$, bei denen mindestens einmal die Division auftritt. Die einfachsten (elementaren) gebrochenen trigonometrischen Funktionen sind die, welche nur eine Grundrechenart (die Division) enthalten und bei denen die Division auch nur einmal angewendet wird. Es sind dies die Funktionen Tangens, Kotangens, Sekans und Kosekans, die folgendermaßen definiert sind:

$$\operatorname{tg} x = \frac{\sin x}{\cos x}, \quad \operatorname{ctg} x = \frac{\cos x}{\sin x}, \quad \sec x = \frac{1}{\cos x} \quad \text{und} \quad \operatorname{cosec} x = \frac{1}{\sin x}. \tag{137}$$

Besonderes Interesse verdient der Tangens.

Aus den Formeln (110) folgt

$$\operatorname{tg}(x+\pi) = \frac{\sin(x+\pi)}{\cos(x+\pi)} = \frac{-\sin x}{-\cos x} = \frac{\sin x}{\cos x} = \operatorname{tg} x, \tag{138}$$

d. h., *der Tangens besitzt die Periode* π, also eine Periode, die halb so groß wie die des Sinus und des Kosinus ist.

Ferner ist klar, daß *der Tangens eine ungerade Funktion ist*:

$$\operatorname{tg}(-x) = \frac{\sin(-x)}{\cos(-x)} = \frac{-\sin x}{\cos x} = -\operatorname{tg} x. \tag{139}$$

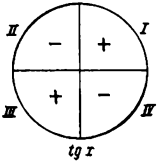
Es genügt also, das Verhalten des Tangens im Intervall $0 < x < \frac{\pi}{2}$ zu studieren. Verfügt man über das graphische Bild des Tangens in diesem Intervall, so kann man es zunächst wegen seiner Symmetrie in bezug auf den Ursprung O auf das Intervall $-\frac{\pi}{2} < x < 0$ und von dort aus wegen der Periodizität des Tangens auf die ganze x-Achse fortsetzen.

Für $x = 0$ erhalten wir

$$\operatorname{tg} 0 = \frac{\sin 0}{\cos 0} = 0.$$

Da der Zähler $\sin x$ im Intervall $0 \leqq x \leqq \frac{\pi}{2}$ von 0 bis 1 wächst und der Nenner $\cos x$ von 1 bis 0 fällt, wächst der Tangens im Intervall $0 \leqq x < \frac{\pi}{2}$, und zwar von Null über alle Grenzen. Für $x = \frac{\pi}{2}$ ist der Tangens nicht erklärt (verliert er seinen Sinn, besitzt er eine Unstetigkeit), weil nämlich der Nenner $\cos x$ dort verschwindet, während der Zähler $\sin x$ den Wert Eins annimmt. Somit hat das graphische Bild des Tangens mit der Geraden $x = \frac{\pi}{2}$ keinen Punkt gemein.

Der Tangens ist positiv im Intervall $0 < x < \frac{\pi}{2}$ und negativ (weil er ungerade ist) im Intervall $-\frac{\pi}{2} < x < 0$. Wegen der Periodizität ist er wiederum positiv im Intervall $\pi < x < \frac{3\pi}{2}$ und negativ für $\frac{\pi}{2} < x < \pi$. Schematisch kann man also das Vorzeichenverhalten des Tangens folgendermaßen beschreiben:



Zur geometrischen Deutung des Tangens kehren wir zum Einheitskreis zurück (Abb. 53): Ist M der Endpunkt des Bogens $\overset{\frown}{AM}$ der Länge x, so hat der Schnittpunkt P des Durchmessers OM mit der Tangente an den Kreis im Punkte A die Ordinate $\operatorname{tg} x$; denn aus der Ähnlichkeit der Dreiecke OAP und OM_1M folgt

$$\frac{AP}{OA} = \frac{M_1M}{OM_1},$$

d. h.

$$\frac{AP}{1} = \frac{\sin x}{\cos x}, \quad AP = \operatorname{tg} x.$$

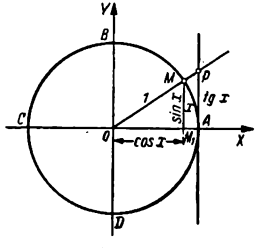


Abb. 53

In Abb. 54 ist dargestellt, wie man das graphische Bild des Tangens im Intervall $0 < x < \frac{\pi}{2}$ punktweise gewinnen kann. Das Intervall ist zu diesem Zweck in acht gleichlange Teilintervalle zerlegt.

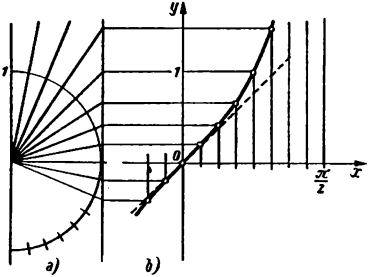


Abb. 54

Unter Beachtung der geometrischen Deutung des Tangens hat man bei der punktweisen Konstruktion des graphischen Bildes der Funktion $\operatorname{tg} x$ in der x, y-Ebene (Abb. 54) lediglich die Schnittpunkte der Durchmesser des Einheitskreises mit der vertikalen Tangente im Anfangspunkt A des Einheitskreises zu bestimmen und die Ordinaten dieser Punkte auf den entsprechenden vertikalen Geraden in der x, y-Ebene abzutragen.

Bemerkenswert ist noch folgende Eigenschaft des graphischen Bildes der Funktion $\operatorname{tg} x$. Wie der Abb. 53 zu entnehmen ist, ist die Länge des Bogens $\overset{\frown}{AM}$ stets kleiner als die Strecke AP,[1]) so daß der Quotient $\frac{AP}{\overset{\frown}{AM}}$ größer als Eins ist; wenn sich der Punkt M dem Punkt A nähert, so strebt dieser Quotient gegen Eins. In Abb. 54 entspricht dem folgendes: Die Abszisse jedes Punktes der graphischen Darstellung ist kleiner als seine Ordinate, und daher das Verhältnis der Ordinate zur Abszisse größer als Eins; strebt jedoch der Punkt der graphischen Darstellung gegen den Ursprung O, so nähert sich dieses Verhältnis dem Wert Eins. Wir können also feststellen, daß die Tangenskurve im Intervall $0 < x < \frac{\pi}{2}$ oberhalb der Winkelhalbierenden $y = x$ liegt,

$$x < \operatorname{tg} x \qquad \left(0 < x < \frac{\pi}{2}\right),$$

daß sie sich aber in unmittelbarer Nähe des Ursprungs eng an diese Gerade anschmiegt (sie im Koordinatenursprung berührt).

Ohne uns mit einer eingehenden Diskussion der Funktionen $\operatorname{ctg} x$, $\sec x$ und $\operatorname{cosec} x$ abzugeben, bemerken wir lediglich, daß sich die graphischen Bilder dieser Funktionen dadurch ergeben, daß man aus den graphischen Bildern der Funktionen $\operatorname{tg} x$, $\cos x$, $\sin x$ die graphischen Bilder der zugehörigen reziproken Funktionen konstruiert (vgl. § 4).

[1]) Bezeichnet man nämlich mit A' den Punkt des Einheitskreises, der in bezug auf die Gerade OM symmetrisch zum Punkt A liegt, so ist der Bogen $\overset{\frown}{AMA'}$ kleiner als der Streckenzug APA' (denn von zwei konvexen Kurven ist die „innere" kürzer als die „äußere"), so daß auch der Bogen $\overset{\frown}{AM}$ kleiner als die Strecke AP ist.

Die graphische Darstellung der Funktion $\operatorname{ctg} x$ kann allerdings aus der des Tangens noch durch ein anderes, einfacheres Verfahren gewonnen werden: Da nämlich

$$\operatorname{tg}\left(\frac{\pi}{2}-x\right)=\frac{\sin\left(\frac{\pi}{2}-x\right)}{\cos\left(\frac{\pi}{2}-x\right)}=\frac{\cos x}{\sin x}=\operatorname{ctg} x$$

oder

$$\operatorname{ctg} x=-\operatorname{tg}\left(x-\frac{\pi}{2}\right)$$

ist, braucht man am graphischen Bild des Tangens nur die folgenden Transformationen auszuführen:

1. eine Spiegelung an der x-Achse;

2. eine Parallelverschiebung längs der x-Achse um die Strecke $\frac{\pi}{2}$.

Man sieht leicht ein, daß diese Transformationen zusammen einer Spiegelung an der Gerade $x=\frac{\pi}{4}$ gleichwertig sind.[1])

In ähnlicher Weise, wie man die graphische Darstellung des Kosinus aus der des Sinus erhält, kann man auch die graphische Darstellung des Kosekans aus der des Sekans erhalten.[2])

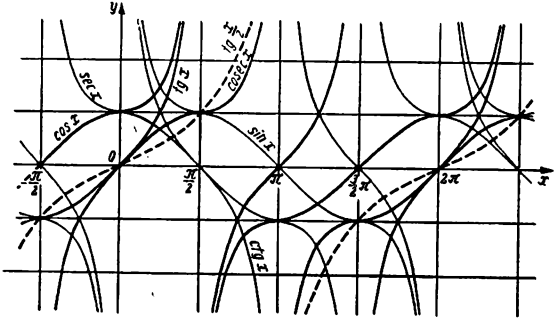


Abb. 55

In Abb. 55 sind die graphischen Bilder der sechs trigonometrischen Funktionen

$$\sin x,\quad \cos x,\quad \operatorname{tg} x,\quad \operatorname{cosec} x,\quad \sec x,\quad \operatorname{ctg} x$$

dargestellt.

[1]) Hierbei ist allerdings noch zu berücksichtigen, daß die graphische Darstellung des Tangens symmetrisch in bezug auf den Ursprung O liegt.

[2]) Daher stammt die Bezeichnung „Kofunktion".

Um über die Beziehungen zwischen den sechs trigonometrischen Funktionen Klarheit zu erlangen, setzen wir zur Abkürzung

$$\left.\begin{aligned} u &= \cos x, & u_1 &= \sec x, \\ v &= \sin x, & v_1 &= \operatorname{cosec} x, \\ w &= \operatorname{tg} x, & w_1 &= \operatorname{ctg} x. \end{aligned}\right\} \tag{140}$$

Die grundlegenden Beziehungen (101) und (137) kann man dann in der Form

$$\left.\begin{aligned} &1.\ u^2+v^2=1, && 3.\ uu_1=1, \\ &2.\ w=\frac{u}{x}, && 4.\ vv_1=1, \\ & && 5.\ ww_1=1 \end{aligned}\right\} \tag{141}$$

schreiben.

Das System dieser fünf Gleichungen besitzt die folgende Eigenschaft: Ist ein Wert für eine der sechs Veränderlichen vorgegeben, so kann man, ohne auf andere Rechenarten als auf die vier Grundrechenarten und das Ausziehen der Quadratwurzel zurückgreifen zu müssen, die Werte der übrigen fünf Veränderlichen bestimmen (das Gleichungssystem (141) nach diesen fünf Veränderlichen auflösen). Man muß jedoch, entsprechend der Problemstellung, die Werte der Quadratwurzeln im algebraischen Sinn nehmen, d. h., in den Fällen, in denen sie auftritt, verbleibt eine Unbestimmtheit im Vorzeichen, also eine Doppeldeutigkeit.

Wenn es also das System (141) auch gestattet, jede der sechs Größen u, v, w, u_1, v_1, w_1 als Funktion jeder anderen anzusehen, so sind, wie eine ausführliche Betrachtung zeigt, von den $5 \cdot 6 = 30$ hierbei auftretenden Funktionen doch nur 6 eindeutig (rational), während die übrigen eine Quadratwurzel enthalten und daher doppeldeutig sind. Wenn beispielsweise für den Sinus der Wert $v=\frac{1}{2}$ vorgegeben ist, so ergeben sich für den Kosinus die beiden Werte $u=\pm\frac{\sqrt{3}}{2}$. Das Vorzeichen kann nur dadurch bestimmt werden, daß man von irgendwelchen zusätzlichen Angaben ausgeht, die eine Entscheidung darüber gestatten, in welchem der vier Teilintervalle des Intervalls $0 \leqq x < 2\pi$ der Wert der unabhängigen Veränderlichen x liegt. Unter den sechs trigonometrischen Funktionen gibt es keine, durch deren Werte die Werte der übrigen Funktionen eindeutig bestimmt sind.

Lediglich der Tangens verdient in dieser Hinsicht ein besonderes Interesse. Ist nämlich $\operatorname{tg} x = w$, so ergeben sich die entsprechenden Werte des Sinus und des Kosinus gemäß

$$\left.\begin{aligned} \sin x &= \frac{w}{\pm\sqrt{1+w^2}}, \\ \cos x &= \frac{1}{\pm\sqrt{1+w^2}}. \end{aligned}\right\} \tag{142}$$

Auch hier ist zwar das Vorzeichen vor den Wurzeln noch willkürlich, jedoch muß in beiden Fällen dasselbe Vorzeichen gewählt werden. Daraus folgt,

daß nicht nur die Werte $\sin^2 x$ und $\cos^2 x$, sondern z. B. auch das Produkt $\sin x \cos x$ und die Werte $\sin 2x$ und $\cos 2x$ eindeutig (rational) durch w ausdrückbar sind:

$$\left.\begin{aligned} \sin 2x &= \frac{2w}{1+w^2}, \\ \cos 2x &= \frac{1-w^2}{1+w^2}. \end{aligned}\right\} \tag{143}$$

Hieraus ergibt sich nun, daß sich mittels einer neuen Funktion der Periode 2π, nämlich des Tangens des halben Winkels,

$$t = \operatorname{tg} \frac{x}{2},$$

der Sinus und der Kosinus und damit auch die übrigen vier trigonometrischen Funktionen rational ausdrücken lassen. Ersetzt man nämlich in (143) x durch $\frac{x}{2}$, so erhält man

$$\left.\begin{aligned} \sin x &= \frac{2t}{1+t^2}, \\ \cos x &= \frac{1-t^2}{1+t^2} \end{aligned}\right\} \tag{144}$$

(was man natürlich auch direkt ausrechnen kann).

Das graphische Bild der Funktion $\operatorname{tg} \frac{x}{2}$ (in Abb. 55 gestrichelt) ergibt sich aus der graphischen Darstellung der Funktion $\operatorname{tg} x$ durch Dehnung in Richtung der x-Achse mit dem Dehnungsfaktor 2. Der Abb. 55 ist zu entnehmen, daß jeder Wert t einem und nur einem Wert x (aus einem Intervall der Länge 2π) entspricht. Folglich gehört zu dem Wert t auch nur ein Wert jeder der Funktionen u, v, w, u_1, v_1, w_1.

Natürlich leistet die Funktion $t_1 = \frac{1}{t} = \operatorname{ctg} \frac{x}{2}$ genau dasselbe.

§ 30. Über die Darstellbarkeit von Funktionen, die rational von den trigonometrischen Funktionen abhängen, durch eine oder zwei trigonometrische Funktionen

Jede rationale Funktion der elementaren trigonometrischen Funktionen in ein und derselben unabhängigen Veränderlichen x kann als rationale Funktion in den beiden trigonometrischen Grundfunktionen

$$u = \cos x \quad \text{und} \quad v = \sin x$$

dargestellt werden.

Dies folgt daraus, daß sich auf Grund von (141) die Funktion $w = \operatorname{tg} x$, aber auch die Funktionen $u_1 = \sec x$, $v_1 = \operatorname{cosec} x$ und $w_1 = \operatorname{ctg} x$ rational durch die Funktionen $\cos x$ und $\sin x$ ausdrücken lassen. Hierbei hat man nur

zu beachten, daß eine rationale Funktion in einer oder mehreren Veränderlichen, deren jede rational von einer oder mehreren Veränderlichen abhängt, offenbar selbst eine rationale Funktion von eben dieser oder diesen letztgenannten Veränderlichen ist.

Beispiel.

$$\frac{1}{2}\left(\frac{\sec x}{1+\operatorname{tg} x}-\frac{\operatorname{cosec} x}{1-\operatorname{ctg} x}\right)=\frac{v}{u^2-v^2}=\frac{\cos x}{\cos^2 x-\sin^2 x}.$$

I. *Jede rationale Funktion der elementaren trigonometrischen Funktionen* (in ein und derselben Veränderlichen x) *kann als rationale Funktion von nur einer Funktion, dem Tangens des halben Winkels,*

$$t=\operatorname{tg}\frac{x}{2},$$

dargestellt werden. Dies ergibt sich unmittelbar aus den Formeln (144).

Beispiel.

$$\frac{\cos x}{\cos^2 x-\sin^2 x}=\frac{1-t^4}{1-6t^2+t^4}=\frac{1-\operatorname{tg}^4\frac{x}{2}}{1-6\operatorname{tg}^2\frac{x}{2}+\operatorname{tg}^4\frac{x}{2}}.$$

Während Satz I ganz allgemein gilt, enthalten die folgenden Sätze II, III und IV zusätzliche Voraussetzungen und sondern somit wichtige Spezialfälle aus.

II. *Ist $f(x)$ eine rationale Funktion der elementaren trigonometrischen Funktionen, die als Funktion von x gerade bzw. ungerade ist, so kann man $f(x)$ als rationale Funktion allein in* cos x *bzw. als Produkt aus* sin x *und einer rationalen Funktion von* cos x *darstellen.*

a) Es sei $f(x)$ gerade. Dann ist einerseits (nach Satz I)

$$f(x)=R_1\left(\operatorname{tg}\frac{x}{2}\right)^{1}) \tag{145}$$

und andererseits

$$f(-x)=f(x). \tag{146}$$

Daraus folgt

$$R_1\left(\operatorname{tg}\frac{(-x)}{2}\right)=R_1\left(\operatorname{tg}\frac{x}{2}\right),$$

d. h.

$$R_1\left(-\operatorname{tg}\frac{x}{2}\right)=R_1\left(\operatorname{tg}\frac{x}{2}\right) \tag{147}$$

oder

$$R_1(-t)=R_1(t),^{2}) \tag{148}$$

---

[1]) Mit R (evtl. mit Indizes) bezeichnen wir im folgenden rationale Funktionen.

[2]) Die Gleichung (148) folgt aus Gleichung (147), da man zu jedem t ein x so finden kann, daß $t=\operatorname{tg}\frac{x}{2}$.

so daß die rationale Funktion R_1 gerade ist. Demzufolge kann sie als rationale Funktion in t^2 dargestellt werden[1]):

$$R_1(t) = R_2(t^2).$$

Nun ist aber

$$t^2 = \operatorname{tg}^2 \frac{x}{2} = \frac{1-\cos x}{1+\cos x} = \frac{1-u}{1+u}$$ [2])

und daher

$$f(x) = R_1(t) = R_2(t^2) = R_2\left(\frac{1-u}{1+u}\right) = R_3(u),$$

was zu beweisen war.

b) Ist dagegen die Funktion $f(x)$ ungerade, so gilt an Stelle von (146)

$$f(-x) = -f(x)$$

und an Stelle von (148)

$$R_1(-t) = -R_1(t),$$

so daß die Funktion $R_1(t)$ ungerade ist. Dann ist aber die Funktion $\frac{R_1(t)}{t}$ gerade, und es gilt wie oben

$$\frac{R_1(t)}{t} = R_2(t^2) = R_3(u).$$

---

[1]) Ist $R(x)$ eine gerade rationale Funktion in x, so ist R eine rationale Funktion in x^2. Der Beweis dieses Satzes liegt auf der Hand, soll jedoch der Vollständigkeit halber hier angedeutet werden:

Es sei $R(x) = \frac{P(x)}{Q(x)}$, wobei $P(x)$ und $Q(x)$ Polynome in x sind. Trennen wir in $P(x)$ und $Q(x)$ die Glieder mit geraden und ungeraden Potenzen von x, so erhalten wir

$$P(x) = P_1(x^2) + xP_2(x^2), \quad Q(x) = Q_1(x^2) + xQ_2(x^2),$$

$$R(x) = \frac{P_1(x^2) + xP_2(x^2)}{Q_1(x^2) + xQ_2(x^2)},$$

wobei P_1, P_2, Q_1 und Q_2 neue Polynome sind. Aus der Beziehung $R(-x) = R(x)$ ergibt sich nun

$$\frac{P_1(x^2) - xP_2(x^2)}{Q_1(x^2) - xQ_2(x^2)} = \frac{P_1(x^2) + xP_2(x^2)}{Q_1(x^2) + xQ_2(x^2)},$$

d. h.

$$\frac{P_1(x^2)}{Q_1(x^2)} = \frac{P_2(x^2)}{Q_2(x^2)}.$$

Hier steht auf der linken Seite dieselbe rationale Funktion wie auf der rechten Seite. Bezeichnen wir sie mit $S(x^2)$, so erhalten wir

$$P_1(x^2) = Q_1(x^2)\,S(x^2), \quad P_2(x^2) = Q_2(x^2)\,S(x^2)$$

und damit

$$R(x) = S(x^2).$$

[2]) Mit u, v, t werden wie bisher die Funktionen $\cos x$, $\sin x$ bzw. $\operatorname{tg} \frac{x}{2}$ bezeichnet.

Im betrachteten Fall ist also

$$f(x) = R_1(t) = t\,R_2(u) = \operatorname{tg}\frac{x}{2}\cdot R_2(u).$$

Nun gilt aber

$$\operatorname{tg}\frac{x}{2} = \frac{\sin x}{1+\cos x}, \tag{149}$$

so daß

$$f(x) = \frac{\sin x}{1+\cos x}\,R_2(u) = \frac{v}{1+u}\,R_2(u) = v\,R_4(u),$$

was zu beweisen war.

III. *Ist $f(x)$ eine rationale Funktion der elementaren trigonometrischen Funktionen, die bei Substitution von x durch $\pi - x$ sich überhaupt nicht ändert bzw. die bei dieser Substitution nur ihr Vorzeichen ändert, so ist $f(x)$ eine rationale Funktion allein von $\sin x$ bzw. Produkt von $\cos x$ mit einer rationalen Funktion von $\sin x$.*

a) Ist für jedes x

$$f(x) = f(\pi - x), \tag{150}$$

so ist, wie man erkennt, wenn man x durch $x + \frac{\pi}{2}$ ersetzt, auch

$$f\left(\frac{\pi}{2}+x\right) = f\left(\frac{\pi}{2}-x\right).$$

Setzen wir nun

$$f\left(\frac{\pi}{2}-x\right) = f_1(-x),$$

so ist also

$$f_1(x) = f\left(x+\frac{\pi}{2}\right). \tag{151}$$

Die Funktion $f_1(x)$ ist also gerade und kann daher nach Satz II als rationale Funktion in $u = \cos x$ dargestellt werden,

$$f_1(x) = R_1(u),$$

d. h.

$$f\left(x+\frac{\pi}{2}\right) = R_1(\cos x).$$

Ersetzen wir nun hier x durch $x - \frac{\pi}{2}$, so erhalten wir

$$f(x) = R_1\left(\cos\left(x-\frac{\pi}{2}\right)\right) = R_1(\sin x) = R_1(v).$$

b) Gilt an Stelle von (150) die Beziehung

$$f(\pi - x) = -f(x),$$

so ist die in (151) definierte Funktion $f_1(x)$ ungerade, und es ergibt sich auf Grund von Satz II

$$f_1(x) = v\,R_1(u) = \sin x\cdot R_1(\cos x),$$

d. h.

$$f\left(x+\frac{\pi}{2}\right)=\sin x\cdot R_1(\cos x).$$

Ersetzt man hier x durch $x-\frac{\pi}{2}$, so erhält man

$$f(x)=-\cos x\cdot R_1(\sin x)=\cos x\cdot R_2(\sin x)=uR_2(v).$$

IV. *Ist $f(x)$ eine rationale Funktion der elementaren trigonometrischen Funktionen, die bei Substitution von x durch $x+\pi$ sich überhaupt nicht ändert* (also die Periode π hat) *bzw. die bei dieser Substitution nur ihr Vorzeichen ändert, so ist $f(x)$ eine rationale Funktion von $\operatorname{tg} x$ bzw. Produkt von $\cos x$ oder $\sin x$ mit einer rationalen Funktion von $\operatorname{tg} x$.*

a) Wir betrachten zunächst den Fall, daß $f(x)$ die Periode π besitzt, d. h., daß für jedes x

$$f(x+\pi)=f(x)$$

gilt. Offenbar können wir zunächst $f(x)$ auf die Form

$$f(x)=\frac{P(\cos x,\sin x)}{Q(\cos x,\sin x)}$$

bringen, wobei Zähler und Nenner Polynome in $\cos x$ und $\sin x$ sind. Auf Grund von

$$\cos^2 x=\frac{1}{1+\operatorname{tg}^2 x},\quad \sin^2 x=\frac{\operatorname{tg}^2 x}{1+\operatorname{tg}^2 x},\quad \sin x\cos x=\frac{\operatorname{tg} x}{1+\operatorname{tg}^2 x},\quad \sin x=\operatorname{tg} x\cos x$$

können wir dann $f(x)$ folgendermaßen schreiben:

$$f(x)=\frac{R_1(\operatorname{tg} x)+R_2(\operatorname{tg} x)\cos x}{R_3(\operatorname{tg} x)+R_4(\operatorname{tg} x)\cos x},$$

wobei R_1, R_2, R_3, R_4 rationale Funktionen sind. Wegen $\cos(x+\pi)=-\cos x$ gilt nun

$$f(x+\pi)=\frac{R_1(\operatorname{tg} x)-R_2(\operatorname{tg} x)\cos x}{R_3(\operatorname{tg} x)-R_4(\operatorname{tg} x)\cos x},$$

d. h., da $f(x+\pi)=f(x)$ ist,

$$\frac{R_1(\operatorname{tg} x)+R_2(\operatorname{tg} x)\cos x}{R_3(\operatorname{tg} x)+R_4(\operatorname{tg} x)\cos x}=\frac{R_1(\operatorname{tg} x)-R_2(\operatorname{tg} x)\cos x}{R_3(\operatorname{tg} x)-R_4(\operatorname{tg} x)\cos x}$$

und daher

$$\frac{R_1(\operatorname{tg} x)}{R_3(\operatorname{tg} x)}=\frac{R_2(\operatorname{tg} x)}{R_4(\operatorname{tg} x)}.$$

Setzen wir nun $\frac{R_1(w)}{R_3(w)}=R(w)$, so ist R eine rationale Funktion, und es gilt

$$R_1(\operatorname{tg} x)=R(\operatorname{tg} x)\,R_3(\operatorname{tg} x),\quad R_2(\operatorname{tg} x)=R(\operatorname{tg} x)\,R_4(\operatorname{tg} x)$$

und daher

$$f(x)=\frac{R(\operatorname{tg} x)\,(R_3(\operatorname{tg} x)+R_4(\operatorname{tg} x)\cos x)}{R_3(\operatorname{tg} x)+R_4(\operatorname{tg} x)\cos x}=R(\operatorname{tg} x),$$

was zu beweisen war.

b) Genügt dagegen die Funktion $f(x)$ der Bedingung

$$f(x+\pi) = -f(x),$$

so betrachten wir die Funktionen

$$f_1(x) = \frac{f(x)}{\cos x} \quad \text{und} \quad f_2(x) = \frac{f(x)}{\sin x}.$$

Man sieht leicht, daß $f_1(x)$ und $f_2(x)$ die Periode π besitzen, so daß nach dem oben Bewiesenen

$$f_1(x) = R_1(\operatorname{tg} x), \quad f_2(x) = R_2(\operatorname{tg} x)$$

ist. Dann ist aber

$$f(x) = f_1(x)\cos x = \cos x R_1(\operatorname{tg} x),$$

$$f(x) = f_2(x)\sin x = \sin x R_2(\operatorname{tg} x).$$

Bemerkung. Man prüft leicht nach, daß auch die Umkehrungen der Sätze I bis IV gelten. Für die Sätze II bis IV stellt man dies direkt fest; was Satz I betrifft, so genügt ein Hinweis auf die Beziehung (149). Die Sätze I bis IV liefern also notwendige und hinreichende Bedingungen dafür, daß eine Funktion in der jeweils angegebenen Form darstellbar ist.

§ 31. Beispiele für die Diskussion von rationalen Funktionen der trigonometrischen Funktionen. Trigonometrische Gleichungen

Die im vorangehenden Paragraphen bewiesenen Sätze I bis IV und die mit ihnen verwandten Sätze aus § 28 eröffnen die Möglichkeit für eine Diskussion von Funktionen, die rational von den trigonometrischen Grundfunktionen abhängen. Dies gelingt dadurch, daß man die gegebene Funktion $f(x)$ in der Form $f(x) = R(g(x))$ darstellt, wobei R eine rationale Funktion und $g(x)$ eine der Funktionen $\cos x$, $\sin x$, $\operatorname{tg} x$ oder, im allgemeinen Fall, die Funktion $\operatorname{tg} \frac{x}{2}$ ist. Da man das Verhalten der rationalen Funktion und das Verhalten der Funktion $g(x)$ im Prinzip als bekannt ansehen kann, ist damit auch das Verhalten der zusammengesetzten Funktion $f(x)$ festgelegt (vgl. § 5, 9).

Die genannten Sätze sind aber darüber hinaus auch nützlich bei der Lösung von trigonometrischen (goniometrischen) Gleichungen. Das Hauptproblem bei der Lösung solcher Gleichungen besteht darin, sie durch Einführung einer passenden neuen Veränderlichen zu „algebraisieren". Die genannten Sätze gestatten es nun, die Wahl der trigonometrischen Funktion, die als neue Veränderliche genommen wird, nach einfachen formalen Kriterien zu treffen, wodurch natürlich die Art der weiteren identischen Umformungen wesentlich beeinflußt wird.

Zu bemerken ist noch folgendes: Eine „Rationalisierung" mit Hilfe der Funktion $t = \operatorname{tg}\frac{x}{2}$ gelingt immer (§ 30, Satz I), die Wahl von $\operatorname{tg}\frac{x}{2}$ als neue Veränderliche ist sozusagen die „Universalsubstitution". Allerdings ist in den Fällen, in denen man mit den Funktionen $u = \cos x$, $v = \sin x$ oder $w = \operatorname{tg} x$ auskommt, die zu verwendende rationale Funktion im allgemeinen einfacher, so daß die „Universalsubstitution" nur „im äußersten Fall" zu empfehlen ist.

Beispiel 1.

$$y = f(x) = \frac{1}{3 + 2\cos x}.$$

Dies ist eine rationale Funktion in $u = \cos x$. Durchläuft x alle Zahlen, so variiert u im Intervall von -1 bis $+1$ (und zwar periodisch mit der Periode 2π). Die Funktion $\frac{1}{3+2u}$ ist in diesem Intervall $\left(\text{ja sogar überall mit Ausnahme der Unstetigkeitsstelle } u = -\frac{3}{2}\right)$ fallend, so daß (vgl. § 5,9) die Funktion $y = f(x)$ überall dort wächst, wo $\cos x$ fällt, und überall dort fällt, wo $\cos x$ wächst. Da die Funktion $f(x)$ die Periode 2π besitzt

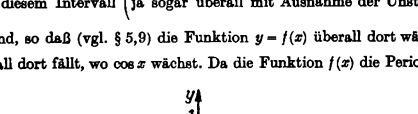


Abb. 56

und gerade ist, genügt es, ihr Verhalten im Intervall $0 \leqq x \leqq \pi$ zu untersuchen. Wächst x von 0 bis π, so fällt $\cos x$ von 1 bis -1, so daß die Funktion $f(x)$ im Intervall $0 \leqq x \leqq \pi$ wächst $\left(\text{und zwar von } \frac{1}{5} \text{ bis } 1\right)$. Für $x = \frac{\pi}{2}$ nimmt $f(x)$ den Wert $\frac{1}{3}$ an (Abb. 56).

Beispiel 2.

$$y = f(x) = \frac{\sin^2 x - \sin 2x}{\cos^2 x}.$$

Da die Werte der Funktion sich nicht ändern, wenn man x durch $x + \pi$ ersetzt ($f(x)$ also die Periode π hat), wird man natürlich versuchen, $w = \operatorname{tg} x$ als neue Veränderliche einzuführen. Einfache Umformungen ergeben

$$y = \operatorname{tg} x(\operatorname{tg} x - 2).$$

Die Funktion

$$w(w-2) = (w-1)^2 - 1$$

fällt für $-\infty < w < 1$ und wächst für $1 < w < \infty$. Die Funktion $w = \operatorname{tg} x$ ist durchweg wachsend; den Wert $w = 1$ nimmt sie (im Intervall $0 \leqq x \leqq \pi$) für $x = \frac{\pi}{4}$ an. An der Stelle $x = \frac{\pi}{2}$ besitzt sie eine Unstetigkeitsstelle. Wenn also x von 0 bis $\frac{\pi}{4}$ wächst, so wächst

w von 0 bis 1, so daß y von 0 bis -1 fällt; wächst x von $\frac{\pi}{4}$ bis $\frac{\pi}{2}$, so wächst w von 1 über alle Grenzen und y von -1 über alle Grenzen (der Wert $y = 0$ ergibt sich für $w = 2$, d. h. für $x \approx 63{,}5^0$); wächst x weiter von $\frac{\pi}{2}$ bis π, so wächst (hinter der Unstetigkeitsstelle) w von $-\infty$ bis 0, so daß y von $+\infty$ bis 0 fällt. Es sei noch erwähnt, daß sich für $x = \frac{3}{4}\pi$ die Werte $w = -1$, $y = 3$ ergeben (Abb. 57).

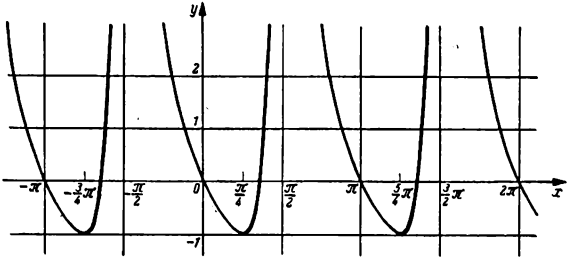


Abb. 57

Beispiel 3.

$$y = f(x) = \frac{1}{2}(3\cos x - \cos 3x).$$

Da die Funktion $f(x)$ die Periode 2π besitzt und gerade ist, kann man sich bei der Diskussion auf das Intervall $0 \leqq x \leqq \pi$ beschränken. Als Hilfsfunktion verwenden wir $u = \cos x$. Wegen

$$\cos 3x = 4\cos^3 x - 3\cos x = 4u^3 - 3u$$

(das Tschebyschewsche Polynom $T_4(u)$) ist

$$y = \frac{1}{2}[3u - (4u^3 - 3u)] = 3u - 2u^3.$$

Dieses Polynom dritten Grades bringen wir durch die Substitution

$$u = z\sqrt{\frac{3}{2}}, \quad z = u\sqrt{\frac{2}{3}}$$

(vgl. § 10, (β)) auf die „Normalform"

$$y = -3\sqrt{\frac{3}{2}}(z^3 - z).$$

Das Wachstumsverhalten der Funktion $z^3 - z$ ändert sich in den Punkten $z = \pm\sqrt{\frac{1}{3}}$ (vgl. § 10), denen die Punkte $u = \pm\sqrt{\frac{1}{2}}$ und damit $x = \frac{\pi}{4}$, $x = \frac{3}{4}\pi$ entsprechen. Bei Betrachtung der durch diese Punkte gebildeten Intervalle überzeugt man sich nun leicht

auf Grund allgemeiner Sätze über das Wachstumsverhalten zusammengesetzter Funktionen (vgl. § 5), daß folgendes gilt: Wächst x von 0 bis $\frac{\pi}{4}$, so fällt u von 1 bis $\sqrt{\frac{1}{2}}$, z von $\sqrt{\frac{2}{3}}$ bis $\sqrt{\frac{1}{3}}$, z^3-z von $-\frac{1}{3}\sqrt{\frac{2}{3}}$ bis $-\frac{2}{3}\sqrt{\frac{1}{3}}$ und y wächst von 1 bis $\sqrt{2}$; wächst x von $\frac{\pi}{4}$ bis $\frac{3}{4}\pi$, so fällt u von $\sqrt{\frac{1}{2}}$ bis $-\sqrt{\frac{1}{2}}$, z von $\sqrt{\frac{1}{3}}$ bis $-\sqrt{\frac{1}{3}}$, z^3-z wächst von $-\frac{2}{3}\sqrt{\frac{1}{3}}$ bis $\frac{2}{3}\sqrt{\frac{1}{3}}$ und y fällt von $\sqrt{2}$ bis $-\sqrt{2}$; wächst x von $\frac{3}{4}\pi$ bis π, so

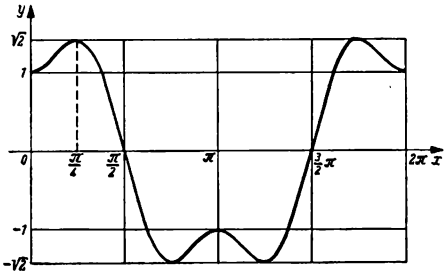


Abb. 58

fällt u von $-\sqrt{\frac{1}{2}}$ bis -1, z von $-\sqrt{\frac{1}{3}}$ bis $-\sqrt{\frac{2}{3}}$, z^3-z von $\frac{2}{3}\sqrt{\frac{1}{3}}$ bis $\frac{1}{3}\sqrt{\frac{2}{3}}$ und y wächst von $-\sqrt{2}$ bis -1 (Abb. 58). Für $x=\frac{\pi}{2}$ nimmt y den Wert 0 an. Man kann an diesem Beispiel die Ergebnisse der Diskussion mit einem punktweise konstruierten graphischen Bild der Funktion $f(x)$ vergleichen; die punktweise Konstruktion besteht in der Bildung der „Halbsumme" der graphischen Bilder von

$$y_1=\cos 3x \quad \text{und} \quad y_2=-3\cos x.$$

Beispiel 4.

$$y=f(x)=\frac{\cos 2x}{\sin x}.$$

Diese Funktion besitzt ebenfalls die Periode 2π und ist ungerade, so daß man auch hier die Funktion nur im Intervall $0\leqq x\leqq\pi$ zu untersuchen braucht. Die Funktion $f(x)$ ändert sich nicht, wenn man x durch $\pi-x$ ersetzt; man ersieht daraus, daß man vorteilhaft die Funktion $v=\sin x$ als Hilfsfunktion einführt. Ferner entnimmt man der genannten Tatsache, daß das graphische Bild der Funktion $f(x)$ symmetrisch in bezug auf die Gerade $x=\frac{\pi}{2}$ ist, so daß man sich bei der Diskussion sogar auf das Intervall $0\leqq x\leqq\frac{\pi}{2}$ beschränken kann.

Offenbar ist

$$y=\frac{\cos^2 x-\sin^2 x}{\sin x}=\frac{1-2\sin^2 x}{\sin x}=\frac{1}{\sin x}-2\sin x=\frac{1}{v}-2v \tag{152}$$

oder auch

$$y=2\frac{\left(\sqrt{\frac{1}{2}}+v\right)\left(\sqrt{\frac{1}{2}}-v\right)}{v}. \tag{153}$$

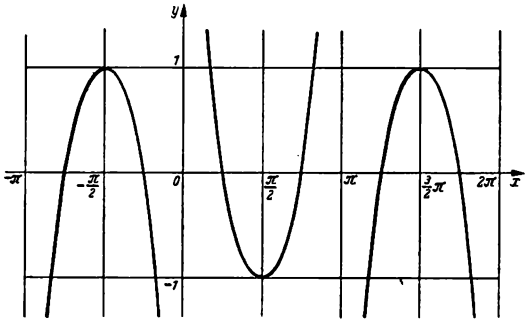


Abb. 59

Die Funktion $\frac{1}{v}-2v$ ist mit Ausnahme ihrer Unstetigkeitsstelle $v=0$ überall fallend, denn die Funktion $\frac{1}{v}$ ist fallend und die Funktion $2v$ ist wachsend. Wenn x von 0 bis $\frac{\pi}{2}$ wächst, so wächst $v=\sin x$ von 0 bis 1, und $y=\frac{1}{v}-2v$ fällt von ∞ bis -1. Aus (153) ersieht man, daß y $\left(\text{im Intervall } 0\leqq x\leqq\frac{\pi}{2}\right)$ den Wert 0 für $v=\sqrt{\frac{1}{2}}$, d. h. für $x=\frac{\pi}{4}$ annimmt (Abb. 59).

Beispiel 5.

$$y=f(x)=\frac{1+\cos x}{3-\sin x}.$$

Hier gelingt es nicht, Symmetrie- oder Periodizitätseigenschaften (außer der trivialen Periode 2π) zu erkennen. Wir werden daher die Substitution

$$t=\operatorname{tg}\frac{x}{2}$$

vornehmen. Drücken wir $\cos x$ und $\sin x$ gemäß (144) durch t aus, so erhalten wir

$$y = \frac{1 + \frac{1-t^2}{1+t^2}}{3 - \frac{2t}{1+t^2}} = \frac{2}{3 - 2t + 3t^2} \tag{154}$$

und weiter (vgl. § 9)

$$y = \frac{2}{3} \cdot \frac{1}{\left(t - \frac{1}{3}\right)^2 + \frac{8}{9}}. \tag{155}$$

Die Funktion $\left(t - \frac{1}{3}\right)^2 + \frac{8}{9}$ ändert ihr Wachstumsverhalten bei $t = \frac{1}{3}$. Diesem Wert t entspricht derjenige Wert x, der sich als Lösung ξ der Gleichung

$$\operatorname{tg} \frac{x}{2} = \frac{1}{3}$$

ergibt, das ist der Wert

$$x = \xi \approx 37°.$$

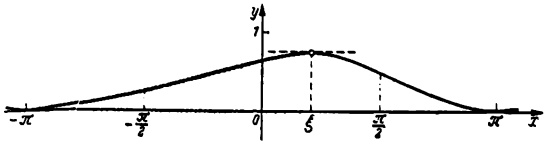


Abb. 60

Wächst x von $-\pi$ bis ξ, so wächst t von $-\infty$ bis $\frac{1}{3}$, und $\left(t - \frac{1}{3}\right)^2 + \frac{8}{9}$ fällt von ∞ bis $\frac{8}{9}$, so daß y von 0 bis $\frac{3}{4}$ wächst; wächst x weiter von ξ bis π, so wächst t von $\frac{1}{3}$ bis über alle Grenzen und $\left(t - \frac{1}{3}\right)^2 + \frac{8}{9}$ von $\frac{8}{9}$ über alle Grenzen, so daß y von $\frac{3}{4}$ bis 0 fällt. Wir bemerken, daß sich für $x = 0$, $x = \frac{\pi}{2}$, $x = -\frac{\pi}{2}$ die Werte $y = \frac{2}{3}$, $y = \frac{1}{2}$ bzw. $y = \frac{1}{4}$ ergeben (Abb. 60).

Bevor wir zu den Beispielen für die Lösung von trigonometrischen (goniometrischen) Gleichungen übergehen, weisen wir darauf hin (vgl. § 5), daß uns hierbei in erster Linie die Frage interessiert, für welche Werte der Veränderlichen x eine gegebene Funktion den Wert Null annimmt. Da es bei der Beantwortung dieser Frage hauptsächlich darauf ankommt, die gegebene Funktion in ein Produkt von hinreichend einfachen Faktoren zu zerlegen, können bei der Lösung dieser Aufgabe die Sätze II bis IV aus § 30 (oder die entsprechenden Sätze aus § 28) gute Dienste leisten. Nach Anwendung dieser Sätze, die angeben, welche trigonometrische Funktion als Hilfsfunktion am besten geeignet ist, ist es mitunter nicht schwer, die Art der vorzunehmenden identischen Umformungen zu erkennen.

Beispiel 6.

$$9 - 11\cos x + 13\cos 2x - 3\cos 3x = 0.$$

Da die linke Seite der Gleichung eine gerade Funktion in x ist, kann man (nach Satz II aus § 30) die linke Seite als Funktion in $u = \cos x$ darstellen. Wegen

$$\cos 2x = 2u^2 - 1, \quad \cos 3x = 4u^3 - 3u$$

(TSCHEBYSCHEWsche Polynome erster Art, Seite 100) ist

$$9 - 11\cos x + 13\cos 2x - 3\cos 3x = -2(6u^3 - 13u^2 + u + 2).$$

Man sieht leicht, daß man den Faktor $u - 2$ abspalten kann. Zerlegt man das verbleibende Polynom zweiten Grades noch in Linearfaktoren, so erhält man

$$6u^3 - 13u^2 + u + 2 = (u - 2)(6u^2 - u - 1) = (u - 2)(2u - 1)(3u + 1).$$

Also kann man der zu lösenden Gleichung die Form

$$(\cos x - 2)(2\cos x - 1)(3\cos x + 1) = 0$$

geben. Der Faktor $\cos x - 2$ kann nirgends den Wert Null annehmen, so daß die gegebene Gleichung nur für

$$\cos x = \frac{1}{2} \quad \text{und} \quad \cos x = -\frac{1}{3}$$

erfüllt ist. Dies führt (im Intervall $-\pi \leqq x \leqq \pi$) auf die Lösungen

$$x_1 = \frac{\pi}{3}, \quad x_2 = -\frac{\pi}{3}, \quad x_3 \approx 71°, \quad x_4 \approx -71°.$$

Beispiel 7.

$$3\sin 2x + 4\cos 3x - 3\sin 4x = 0.$$

Da die linke Seite der Gleichung lediglich das Vorzeichen ändert, wenn man x durch $\pi - x$ ersetzt, empfiehlt es sich (nach Satz III aus § 30), $v = \sin x$ als neue Veränderliche einzuführen. Mittels der TSCHEBYSCHEWschen Polynome ergibt sich

$$\sin 2x = \cos x \cdot 2v,$$

$$\cos 3x = 4\cos^3 x - 3\cos x = \cos x(4\cos^2 x - 3) = \cos x \cdot (1 - 4v^2),$$

$$\sin 4x = (8\cos^3 x - 4\cos x)\sin x = \cos x(8\cos^2 x - 4)\sin x = (4v - 8v^3)\cos x.$$

Wir können also der gegebenen Gleichung auch die Form

$$(12v^3 - 8v^2 - 3v + 2)\cos x = 0$$

geben. Zerlegt man das Polynom auf der linken Seite in Faktoren, so erhält man

$$(4v^2 - 1)(3v - 2)\cos x = 0$$

oder

$$(4\sin^2 x - 1)(3\sin x - 2)\cos x = 0.$$

Dies ergibt (für das Intervall $0 \leqq x \leqq 2\pi$) insgesamt acht Lösungen:

$$x_1 = \frac{\pi}{6}, \quad x_2 = \frac{5\pi}{6}, \quad x_3 = \frac{7\pi}{6}, \quad x_4 = \frac{11\pi}{6}, \quad x_7 = \frac{\pi}{2}, \quad x_8 = \frac{3\pi}{2},$$

x_5 und x_6 sind die beiden Lösungen der Gleichung

$$\sin x = \frac{2}{3}.$$

Beispiel 8. Gesucht sind die Lösungen der Gleichung

$$1 + \sin x \cos x = 35 \cos^4 x$$

im Intervall $0 \leqq x \leqq \frac{\pi}{2}$. Da auf beiden Seiten nur Glieder geraden Grades in bezug auf $\cos x$ und $\sin x$ auftreten, sich also die rechte und die linke Seite bei Substitution von x durch $x + \pi$ nicht ändert, empfiehlt es sich, $w = \operatorname{tg} x$ als neue Veränderliche einzuführen. Unter Verwendung von (142) kann man die gegebene Gleichung auf die Form

$$(1 + w^2)(1 + w + w^2) = 35 \tag{156}$$

bringen. Da $w = \operatorname{tg} x$ im Intervall $0 \leqq x \leqq \frac{\pi}{2}$ stets positiv ist, benötigen wir nur die positiven Lösungen der Gleichung (156). Da die linke Seite dieser Gleichung eine wachsende Funktion in w ist, kann es höchstens eine positive Lösung geben, und man prüft leicht nach, daß $w = 2$ auch wirklich eine positive Lösung ist. Es bleibt also nur noch der Wert x im Intervall $0 < x < \frac{\pi}{2}$ zu bestimmen, für den

$$\operatorname{tg} x = 2$$

gilt.

§ 32. Die Umkehrfunktionen der trigonometrischen Funktionen

Vertauschen wir in

$$y = \sin x \tag{157}$$

die Rolle der Veränderlichen x und y, so erhalten wir die Beziehung

$$x = \sin y, \tag{158}$$

die man auch in der Form

$$y = \operatorname{Arcsin} x \tag{159}$$

schreibt; die Gleichungen (159) und (158) besagen genau dasselbe, sind also als gleichwertig anzusehen.

Die *zyklometrische* Funktion $\operatorname{Arcsin} x$ („Arkussinus" x bedeutet wörtlich: der Bogen, dessen Sinus gleich x ist) ist nicht eindeutig. Um dies einzusehen, vergegenwärtigen wir uns die Verhältnisse am Einheitskreis, mit dessen Hilfe der Sinus definiert wurde (Abb. 44).

Jedem gegebenen Bogen $\overset{\frown}{AM}$ entspricht ein eindeutig bestimmter Wert des Sinus, und zwar ist dies die Ordinate des Punktes M. Ist nun umgekehrt ein gewisser Ordinatenwert x vorgegeben (der positiv oder negativ sein kann, aber dem Betrage nach nicht größer als Eins sein darf), so läßt sich offenbar die Frage, welcher Bogen einen Endpunkt besitzt, dessen Ordinate gleich x ist, nicht eindeutig beantworten. Mit einem Bogen $\overset{\frown}{AM}$ besitzen auch der Bogen $\overset{\frown}{ABN}$, der Bogen $\overset{\frown}{ABCDAM}$, der Bogen $\overset{\frown}{ABCDABN}$ usw., aber auch die „negativen" Bögen $\overset{\frown}{ADCN}$, $\overset{\frown}{ADCBM}$ usw. diese Eigenschaft (vgl. Abb. 44). Wenn also für einen gegebenen Wert x die Funktion Arcsin den Wert y besitzt, so besitzt sie damit auch im Punkt x stets den Wert $\pi - y$ und (allgemein) die Werte $y + 2k\pi$ und $(\pi - y) + 2k\pi$.

Die Mehrdeutigkeit der Funktion Arcsin x kann man übrigens auch unmittelbar aus ihrer graphischen Darstellung ablesen. Wie wir wissen, geht nämlich die graphische Darstellung von (158) (und damit (159)) aus der von (157) durch Spiegelung an der Winkelhalbierenden $y = x$ hervor (Abb. 61). Dem graphischen Bild von Arcsin x sieht man unmittelbar an, daß es für jedes x mit $|x| \leq 1$ unendlich viele Punkte enthält, die die Abszisse x besitzen; ist y die Ordinate eines dieser Punkte (ganz gleich welches), so sind

$$y + 2k\pi, \quad (\pi - y) + 2k\pi$$

die Ordinaten der anderen Punkte. Genauso, wie man bei der Quadratwurzel aus einer gegebenen positiven Zahl (der Umkehrfunktion der Funktion x^2) zwischen der doppeldeutigen „algebraischen" Wurzel und der eindeutigen „arithmetischen" Wurzel unterscheiden muß, muß man beim Arkussinus einer gegebenen Zahl x mit $|x| \leq 1$ angeben, ob ein beliebiger oder aber ein ganz bestimmter Wert (und zwar welcher) gemeint ist. Der einzige Unterschied in den genannten Beispielen besteht darin, daß die Quadratwurzel $\sqrt{x}$ eine doppeldeutige, die Funktion Arcsin x dagegen eine unendlich vieldeutige Funktion ist.

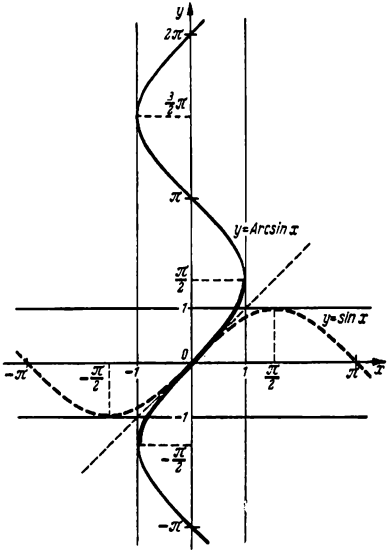


Abb. 61

Den eindeutig bestimmten Wert Arcsin x, der der Bedingung $-\frac{\pi}{2} \leqq y \leqq \frac{\pi}{2}$ genügt, nennt man den *Hauptwert*; man bezeichnet ihn mit arcsin x. Dann ist also

$$\operatorname{Arcsin} x = \begin{cases} \arcsin x + 2k\pi, \\ (\pi - \arcsin x) + 2k\pi. \end{cases} \tag{160}$$

In Abb. 61 ist das graphische Bild der Funktion arcsin x durch fetteren Druck hervorgehoben. Die Funktion arcsin x ist definiert im Intervall $-1 \leqq x \leqq 1$ und ist dort eindeutig, ungerade und überall wachsend (vgl. § 52).

Nachdem die Verhältnisse am Beispiel des Arkussinus genauer dargelegt sind, ist es nicht notwendig, auf die Umkehrfunktionen der anderen elementaren trigonometrischen Funktionen hier ebenso ausführlich einzugehen; wir beschränken uns daher auf kurze Angaben und graphische Illustrationen.

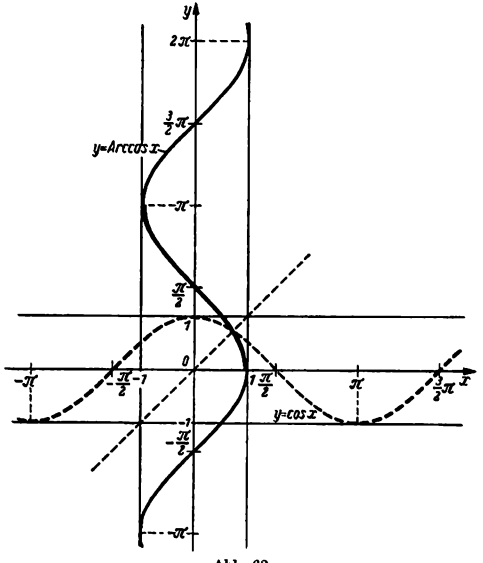


Abb. 62

Die Funktion

$$y = \operatorname{Arccos} x \tag{161}$$

(„Arkuskosinus") ist die Umkehrfunktion der Funktion

$$y = \cos x; \tag{162}$$

sie ist (ebenso wie die Funktion Arcsin x) definiert im Intervall $-1 \leqq x \leqq 1$ und dort unendlich vieldeutig. Ihr *Hauptwert* $y = \arccos x$ ist festgelegt durch die Bedingung $0 \leqq y \leqq \pi$; die anderen Werte der Funktion Arccos x ergeben sich aus dem Hauptwert durch

$$\operatorname{Arccos} x = \pm \arccos x + 2k\pi. \tag{163}$$

Die Funktion $\arccos x$ ist ebenfalls im Intervall $-1 \leqq x \leqq 1$ erklärt und dort eindeutig und überall fallend (Abb. 62).

Die Funktion

$$y = \operatorname{Arctg} x \tag{164}$$

(„Arkustangens") ist die Umkehrfunktion der Funktion

$$y = \operatorname{tg} x. \tag{165}$$

Sie ist definiert für alle Werte der unabhängigen Veränderlichen x (für $-\infty < x < \infty$) und (ebenso wie die Funktion Arcsin x) unendlich vieldeutig.

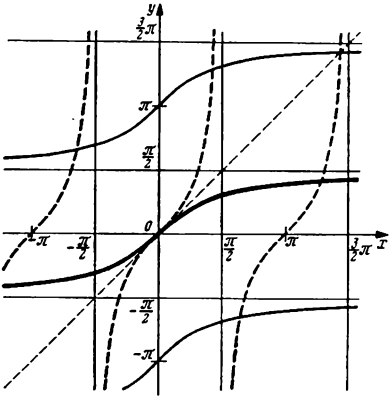


Abb. 63

Der *Hauptwert* arctg x ist festgelegt durch die Bedingung $-\frac{\pi}{2} \leqq y \leqq \frac{\pi}{2}$. Die anderen Werte des Arctg x ergeben sich aus dem Hauptwert durch

$$\operatorname{Arctg} x = \operatorname{arctg} x + k\pi.$$

Die Funktion arctg x ist eindeutig, ungerade und überall wachsend. Die Funktion Arctg x ist aller Werte mit Ausnahme der Werte $\frac{\pi}{2} + k\pi$ fähig. Hieraus folgt, daß ihr graphisches Bild aus unendlich vielen „Zweigen" besteht, die durch Verschiebung des graphischen Bildes des Hauptwertes arctg x in Richtung der y-Achse um Strecken der Form $k\pi$ aus diesem hervorgehen (Abb. 63).

Entsprechend werden die Funktionen Arkuskosekans, Arcussekans und Arkuskotangens definiert:

$$\left.\begin{aligned} y &= \operatorname{Arccosec} x, \\ y &= \operatorname{Arcsec} x, \\ y &= \operatorname{Arcctg} x. \end{aligned}\right\} \tag{166}$$

Sie sind bzw. die Umkehrfunktionen von

$$\left.\begin{aligned} y &= \operatorname{cosec} x, \\ y &= \sec x, \\ y &= \operatorname{ctg} x. \end{aligned}\right\} \tag{167}$$

Ihre Verwendung kann stets dadurch vermieden werden, daß man beachtet, daß[1])

$$\left.\begin{aligned} \operatorname{Arccosec} x = \operatorname{Arcsin}\frac{1}{x}, \quad \operatorname{Arcsec} x = \operatorname{Arccos}\frac{1}{x}, \\ \operatorname{Arctg} x = \operatorname{Arctg}\frac{1}{x} \end{aligned}\right\} \tag{168}$$

Bei Aufgaben der Elementarmathematik ist es nur ganz selten notwendig, die zyklometrischen Funktionen in expliziter Form zu verwenden. Zum Beispiel ist die Aufgabe, den Wert von $\operatorname{Arcsin} x$ für $x = \frac{1}{3}$ zu berechnen, gleichwertig der Aufgabe, alle Bögen (oder Winkel) zu finden, deren Sinus gleich $\frac{1}{3}$ ist. Bei den meisten Aufgaben geometrischen Inhalts gestattet in der Regel eine geeignete Wahl der Veränderlichen, die Umkehrfunktionen der Kreisfunktionen zu umgehen. Dagegen treten in der höheren Mathematik (in der Integralrechnung) die zyklometrischen Funktionen auf ganz natürlichem, direktem Wege auf (vgl. Seite 350), und es wäre dort außerordentlich unzweckmäßig, sie umgehen zu wollen.

Die trigonometrischen Funktionen stehen durch eine ganze Reihe von Relationen miteinander in Beziehung. Einen bedeutenden Teil von ihnen findet man in den üblichen Lehrbüchern der Trigonometrie; einige dieser Formeln sind so wichtig, daß man sie stets und ständig im Gedächtnis haben muß.[2]) Mit den Umkehrfunktionen der trigonometrischen Funktionen taucht eine große Anzahl weiterer Formeln auf. Die mathematische Praxis zeigt jedoch, daß Beziehungen, die zyklometrische Funktionen enthalten, nur gelegentlich auftreten und daß nur einige dieser Formeln besondere Beachtung verdienen. Es interessieren nur Formeln einiger spezieller Typen:

I. *Jede trigonometrische Funktion einer zyklometrischen Funktion* (die durchaus nicht von derselben Art zu sein braucht) *ist eine algebraische Funktion, und zwar läßt sie sich allein mit Hilfe der arithmetischen Operationen und von Quadratwurzeln ausdrücken.*

Handelt es sich bei den Funktionen um Funktionen „gleicher Art", so „heben sie sich auf", wie unmittelbar aus der Definition hervorgeht:[3])

$$\sin(\operatorname{Arcsin} x) = x, \quad \cos(\operatorname{Arccos} x) = x, \quad \operatorname{tg}(\operatorname{Arctg} x) = x.$$

Wir betrachten nun die Fälle, in denen Funktionen verschiedener Art vorliegen. Zum Beispiel ist

$$\sin \operatorname{Arccos} x = \sqrt{1 - x^2}, \tag{169}$$

---

[1]) Die Bögen, deren Kosekans gleich x ist, sind gleich den Bögen, deren Sinus gleich $\frac{1}{x}$ ist; usw.

[2]) Hierzu gehören z. B. die „Additionstheoreme" (für den Sinus und den Kosinus).

[3]) Wir fragen uns: „Wer ist Vater des Sohnes, dessen Vater Johann heißt?" Hier gibt es keinen Zweifel: Der Vater ist Johann.

denn der Sinus des Bogens, dessen Kosinus gleich x ist, ist gleich $\sqrt{1-x^2}$, da $\sin^2 x + \cos^2 x = 1$. Oder genauer: Ist y irgendein Bogen, dessen Kosinus gleich x ist,

$$\operatorname{Arccos} x = y,$$

so ist

$$x = \cos y$$

und daher[1])

$$\sin \operatorname{Arccos} x = \sin y = \sqrt{1-\cos^2 y} = \sqrt{1-x^2}.$$

Entsprechend gelangt man zu den folgenden Beziehungen:[2])

$$\sin \operatorname{Arctg} x = \frac{x}{\sqrt{1+x^2}}, \quad \cos \operatorname{Arcsin} x = \sqrt{1-x^2},$$

$$\cos \operatorname{Arctg} x = \frac{1}{\sqrt{1+x^2}}, \quad \operatorname{tg} \operatorname{Arcsin} x = \frac{x}{\sqrt{1-x^2}},$$

$$\operatorname{tg} \operatorname{Arccos} x = \frac{\sqrt{1-x^2}}{x}.$$

II. *Jede zyklometrische Funktion einer trigonometrischen Funktion* (gleicher oder verschiedener Art) *ist eine unendlich vieldeutige Funktion, die nur im Fall entsprechender Funktionen und in den Fällen, in denen die zyklometrische Funktion zur „Kofunktion" der trigonometrischen Funktion gehört, vereinfacht werden kann.*

Es gilt nämlich:[3])

$$\operatorname{Arcsin}(\sin x) = \begin{cases} x + 2k\pi, \\ (\pi - x) + 2k\pi, \end{cases}$$

$$\operatorname{Arccos}(\cos x) = \pm x + 2k\pi,$$

$$\operatorname{Arctg}(\operatorname{tg} x) = x + k\pi.$$

Ferner ist

$$\operatorname{Arcsin}(\cos x) = \left(\frac{\pi}{2} \mp x\right) + 2k\pi, \quad \operatorname{Arccos}(\sin x) = \pm\left(\frac{\pi}{2} - x\right) + 2k\pi$$

usw. In der Tat: Zum Beispiel bezeichnet $\operatorname{Arcsin}(\cos x)$ jeden Bogen, dessen Sinus gleich $\cos x$ ist. Wegen

$$\cos x = \sin\left(\frac{\pi}{2} - x\right)$$

---

[1]) Natürlich sind hierbei die Wurzeln im algebraischen Sinn zu nehmen; das Vorzeichen kann nur dann genau präzisiert werden, wenn bekannt ist, in welchem Quadranten sich der Endpunkt des Bogens befindet, der durch die zyklometrische Funktion gegeben wird.

[2]) Ist der Bogen, der durch die zyklometrischen Funktionen auf der linken Seite gegeben wird, jeweils derselbe, so besitzen die Wurzeln auf der rechten Seite jeweils dasselbe Vorzeichen.

[3]) Wir fragen uns: „Wer ist Sohn des Vaters, dessen Sohn Johann heißt?" Die Antwort hierauf ist nicht eindeutig: Es kann Johann, es kann aber auch einer seiner Brüder sein.

ist $\frac{\pi}{2} - x$ einer dieser Bögen. Ein anderer Bogen, der diese Eigenschaft hat, ist der Bogen $\frac{\pi}{2} + x$. Alle weiteren Bögen unterscheiden sich von einem der beiden genannten um ein Vielfaches von 2π.

III. Mitunter hat man schließlich *eine zyklometrische Funktion durch passende Änderung des Arguments in eine andere überzuführen.* Beispielsweise soll einer der Werte von $\operatorname{Arcsin} x$ als Arkuskosinus einer Größe y dargestellt werden,

$$\operatorname{Arcsin} x = \operatorname{Arccos} y.$$

In diesem Fall muß

$$y = \cos(\operatorname{Arcsin} x)$$

sein, so daß auf Grund von I

$$y = \sqrt{1 - x^2}$$

sein muß, d. h.

$$\operatorname{Arcsin} x = \operatorname{Arccos} \sqrt{1 - x^2}. \tag{170}$$

Die Formel (170) besagt, daß jeder Wert von $\operatorname{Arcsin} x$ (für $|x| \leqq 1$) gleich einem gewissen Wert von $\operatorname{Arccos} \sqrt{1 - x^2}$ ist; welcher Wert dies ist, bedarf einer genaueren Präzisierung, die von den gegebenen Umständen abhängt. Ähnlich ergibt sich

$$\operatorname{Arcsin} x = \operatorname{Arctg} \frac{x}{\sqrt{1 - x^2}},$$

$$\operatorname{Arctg} x = \operatorname{Arcsin} \frac{x}{\sqrt{1 + x^2}}.$$

§ 33. Diskussion der Tschebyschewschen Polynome. Ihre Minimaleigenschaft

Das Grundprinzip bei der elementaren Diskussion von trigonometrischen Polynomen (oder allgemeiner von Funktionen, die rational von den trigonometrischen Grundfunktionen abhängen) bestand (vgl. § 30) darin, diese Funktionen mit Hilfe einer trigonometrischen Substitution aus einer rationalen Funktion zu gewinnen.

In einigen Fällen läßt sich nun auch der umgekehrte Weg beschreiten, nämlich die Untersuchung eines rationalen Polynoms (oder einer rationalen Funktion) auf die Untersuchung eines trigonometrischen Polynoms (oder einer rationalen Funktion der trigonometrischen Funktionen) zurückführen. Es ist nämlich möglich, die elementaren Eigenschaften gewisser einfacher trigonometrischer Polynome direkt zu ermitteln. Hierzu gehören in erster Linie die trigonometrischen Polynome $\cos nx$ und $\sin nx$.

Als Beispiel betrachten wir das Tschebyschewsche Polynom $T_n(x)$ erster Art im Intervall $-1 \leqq x \leqq +1$. Ersetzt man in

$$\cos nx = T_n(\cos x)$$

$\cos x$ durch u, so ergibt sich

$$T_n(u) = \cos(n \operatorname{Arccos} u). \tag{171}$$

Hier ist es gleichgültig, welcher Wert des Arkuskosinus genommen wird, denn die Funktion $\cos nx$ ändert sich nicht, wenn man x durch $-x$ oder durch $x + 2k\pi$ ersetzt. Wir können also annehmen, daß es sich um den Hauptwert handelt.

Zunächst betrachten wir den Fall, daß n eine gerade Zahl ist: $n = 2m$.

Wächst u von -1 bis $+1$, so fällt $x = \arccos u$ von π bis Null und somit $nx = n \arccos u$ von $n\pi$ bis Null; wenn aber das Argument des Kosinus von $n\pi = 2m\pi$ bis Null fällt, so führt (wie man leicht einsieht, wenn man sich die graphische Darstellung des Kosinus vergegenwärtigt) der Kosinus selbst, d. h. das Polynom $T_n(u)$, insgesamt $\frac{n}{2} = m$ volle Schwingungen zwischen $+1$ und -1 aus und zurück. Es gilt also folgendes:

1. Den Wert $+1$ nimmt $T_n(u)$ überall dort an, wo $n \arccos u = 2k\pi$, d. h. für $u = \cos\frac{2k\pi}{n}$, also

$$T_n\left(\cos\frac{2k\pi}{n}\right) = 1 \qquad \left(0 \leqq k \leqq \frac{n}{2}\right). \tag{172}$$

2. Den Wert -1 nimmt $T_n(u)$ überall dort an, wo $n \arccos u = \pi + 2k\pi$, d. h. für $u = \cos\frac{(2k+1)\pi}{n}$, also

$$T_n\left(\cos\frac{(2k+1)\pi}{n}\right) = -1 \qquad \left(0 \leqq k < \frac{n}{2}\right). \tag{173}$$

3. Den Wert 0 nimmt $T_n(u)$ überall dort an, wo $n \arccos u = \frac{\pi}{2} + k\pi$, d. h. für $u = \cos\frac{\left(k+\frac{1}{2}\right)\pi}{n}$, also

$$T_n\left(\cos\frac{\left(k+\frac{1}{2}\right)\pi}{n}\right) = 0 \qquad (0 \leqq k \leqq n-1). \tag{174}$$

Ist n ungerade, $n = 2m + 1$, so erhalten wir ein ganz ähnliches Resultat. Der einzige Unterschied besteht darin, daß $n \arccos u$ von $n\pi = (2m+1)\pi$ bis Null fällt, wenn u von -1 bis $+1$ wächst, so daß $T_n(u)$ zunächst von -1 bis $+1$ wächst und dann noch m volle Schwingungen von $+1$ bis -1 und zurück ausführt. Die Formeln (172) bis (174) gelten auch hier.

In Abb. 64 ist das Tschebyschewsche Polynom

$$y = T_{12}(x)^{1})$$

im Intervall $-1 \leqq x \leqq 1$ dargestellt.

---

[1]) Hier und im folgenden bezeichnen wir die unabhängige Veränderliche wie üblich wieder mit x.

Eine bemerkenswerte Eigenschaft der Tschebyschewschen Polynome

$$y = T_n(x)$$

ist folgende: Durchläuft x das Intervall $-1 \leqq x \leqq 1$ von 1 bis -1, so nimmt $T_n(x)$, angefangen vom rechten Endpunkt ($x = 1$), abwechselnd das Maximum $+1$ und das Minimum -1 an, die dem Betrage nach übereinstimmen, und läuft dabei n-mal von $+1$ bis -1 oder umgekehrt von -1 bis $+1$; $T_n(x)$ nimmt also am linken Endpunkt des Intervall ($x = -1$) den Wert $+1$ oder -1 an, je nachdem, ob n gerade oder ungerade ist; außerhalb des Intervalls $-1 \leqq x \leqq +1$ wächst $|T_n(x)|$ schnell an.

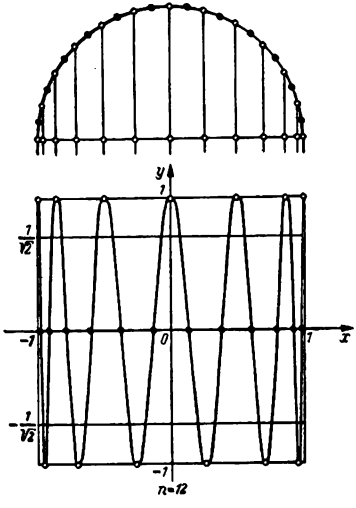


Abb. 64

Wir nehmen nun eine Stauchung in Richtung der y-Achse mit dem Stauchungsfaktor 2^{n-1} vor, gehen also zu dem Polynom

$$y = \frac{1}{2^{n-1}} T_n(x) \qquad (175)$$

über. Die graphische Darstellung dieses Polynoms besitzt entsprechende Eigenschaften wie die des Polynoms $T_n(x)$, nur sind hier die Beträge der Maxima und Minima gleich $\frac{1}{2^{n-1}}$. Gleichzeitig ist der höchste Koeffizient des neuen Polynoms gleich 1, denn der höchste Koeffizient von $T_n(x)$ ist gleich 2^{n-1} (vgl. § 28, (131)). Damit ist gezeigt, daß es unter den Polynomen n-ten Grades eines gibt,

$$P_n(x) = x^n + \cdots, \qquad (176)$$

das im Intervall $-1 \leqq x \leqq 1$ überall die Ungleichung

$$|P_n(x)| \leqq \frac{1}{2^{n-1}}$$

erfüllt $\left(\text{nämlich das Polynom } \frac{1}{2^{n-1}} T_n(x)\right)$.

Wir wollen sofort zeigen, daß man die damit erhaltene Grenze nicht noch weiter herabdrücken kann: Ein Polynom n-ten Grades mit höchstem Koeffi-

zienten 1 kann nicht für alle x mit $|x| \leqq 1$ Werte annehmen, die der Ungleichung

$$|P_n(x)| < \frac{1}{2^{n-1}} \tag{177}$$

genügen.

In Abb. 65 ist der Teil des graphischen Bildes der Funktion (175) dargestellt, der im Rechteck mit den Eckpunkten

$$A\left(+1, \frac{1}{2^{n-1}}\right), \quad B\left(-1, \frac{1}{2^{n-1}}\right), \quad C\left(-1, -\frac{1}{2^{n-1}}\right), \quad D\left(1, -\frac{1}{2^{n-1}}\right)$$

liegt (ausgezogene Linie). Durch die Vertikalen, die durch die Maxima und Minima gehen, wird dieses Rechteck in n Teile zerlegt, und zwar in n Rechtecke, die alle dieselbe Höhe wie das ursprüngliche Rechteck haben. Gäbe es

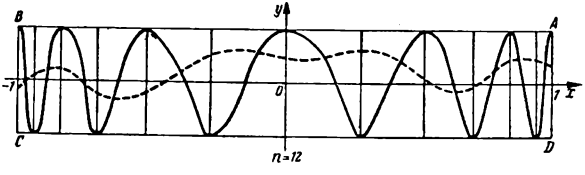


Abb. 65

nun ein Polynom (176), das im Intervall $-1 \leqq x \leqq 1$ überall der Ungleichung (177) genügt, so läge der Teil der graphischen Darstellung von

$$y = P_n(x), \tag{178}$$

der durch die Grenzen $-1 \leqq x \leqq 1$ bestimmt ist, ganz innerhalb des Streifens zwischen den Geraden AB und DC (gestrichelte Kurve in Abb. 65). Da er die Strecken BC und AD verbinden würde, hätte er mit dem graphischen Bild von (175) wenigstens n Schnittpunkte gemeinsam. In der Tat: Verfolgt man den Verlauf des graphischen Bildes von (178) von BC nach AB, so sieht man, daß es alle Vertikalen schneiden muß, die die Teilrechtecke voneinander trennen, und innerhalb jedes Teilrechteckes hätte die graphische Darstellung von (178) einen Punkt mit der von (175) gemeinsam; denn offenbar *ist es unmöglich, daß sich zwei innerhalb eines Rechtecks liegende Kurven nicht schneiden, wenn die eine von ihnen gegenüberliegende Eckpunkte und die andere innere Punkte gegenüberliegender Seiten verbindet.* Die Abszissen der n Schnittpunkte der graphischen Bilder von (178) und (175) (sie sind ohne Zweifel paarweise verschieden) wären nun Wurzeln der Gleichung

$$\frac{1}{2^{n-1}} T_n(x) - P_n(x) = 0. \tag{179}$$

Das ist aber unmöglich, denn die höchsten Koeffizienten der betrachteten Polynome sind gleich, so daß der Grad des Polynoms auf der linken Seite

von (179) kleiner als n ist. Wir sind damit zu einem Widerspruch gelangt, so daß unsere Annahme, es gäbe ein Polynom $P_n(x)$, für das überall im Intervall $-1 \leqq x \leqq 1$ die Ungleichung (177) gilt, falsch ist.

Die bewiesene Eigenschaft der TSCHEBYSCHEWschen Polynome formuliert man häufig folgendermaßen: *Von allen Polynomen $P_n(x)$ der Form* (176) *weicht im Intervall* $-1 \leqq x \leqq 1$ *das* TSCHEBYSCHEWsche *Polynom* $\frac{1}{2^{n-1}} T_n(x)$ *am wenigsten von Null ab.* Hierbei hat man unter der „Abweichung" eines Polynoms $P(x)$ von Null im Intervall $a \leqq x \leqq b$ den größten Wert von $|P(x)|$ in diesem Intervall zu verstehen oder, anders ausgedrückt, die (dem Betrage nach) größte Ordinate der entsprechenden graphischen Darstellung. Diese Eigenschaft ist der Ausgangspunkt für die von P. L. TSCHEBYSCHEW (in der Mitte des vorigen Jahrhunderts) begründete Theorie der besten Approximation von Funktionen gewesen.

Entsprechend könnte man die TSCHEBYSCHEWschen Polynome zweiter Art (vgl. § 28, (131))

$$U_n(x) = \frac{\sin(n \arccos x)}{\sin \arccos x} = \frac{\sin(n \arccos x)}{\sqrt{1-x^2}},$$

untersuchen.[1]) Wir wollen hier nicht näher darauf eingehen, sondern als zweites Beispiel eine gebrochene rationale Funktion behandeln, deren Untersuchung auf die Untersuchung einer gebrochenen trigonometrischen Funktion hinausläuft, und zwar wollen wir die Funktion

$$y = R_n(x) = \operatorname{tg}(n \operatorname{arctg} x) \tag{180}$$

betrachten[2]). Man sieht leicht, daß diese Funktion rational ist. Zunächst ist nämlich $R_1(x) = \operatorname{tg}(\operatorname{arctg} x) = x$ eine rationale Funktion. Ferner ist mit $R_n(x)$ wegen

$$R_{n+1}(x) = \operatorname{tg}((n+1) \operatorname{arctg} x) = \operatorname{tg}(n \operatorname{arctg} x + \operatorname{arctg} x)$$
$$= \frac{\operatorname{tg}(n \operatorname{arctg} x) + \operatorname{tg}(\operatorname{arctg} x)}{1 - \operatorname{tg}(n \operatorname{arctg} x) \cdot \operatorname{tg}(\operatorname{arctg} x)} = \frac{R_n(x) + x}{1 - R_n(x) \cdot x}$$

auch $R_{n+1}(x)$ rational, so daß für jedes n die Funktion $R_n(x)$ rational ist. Beispielsweise ist

$$R_2(x) = \frac{2x}{1-x^2}, \quad R_3(x) = \frac{3x - x^3}{1 - 3x^2}, \quad R_4(x) = \frac{4(x - x^3)}{1 - 6x^2 + x^4}.$$

Da $\operatorname{arctg} x$ und auch $\operatorname{tg} x$ mit Ausnahme ihrer Unstetigkeitsstellen überall wachsend sind, ist auch die Funktion $R_n(x)$ mit Ausnahme ihrer Unstetigkeitsstellen überall wachsend. Wächst nun x von $-\infty$ bis ∞, so wächst die Funktion $\operatorname{arctg} x$ von $-\frac{\pi}{2}$ bis $\frac{\pi}{2}$, also $n \operatorname{arctg} x$

---

[1]) Die Maxima und Minima von $U_n(x)$ liegen allerdings nicht auf horizontalen Geraden, sondern auf den Kurven

$$y = \pm \frac{1}{\sqrt{1-x^2}}.$$

[2]) Es ist gleichgültig, welchen Wert des Arkustangens man nimmt, da sich die Funktion $\operatorname{tg} x$ nicht ändert, wenn x durch $x + \pi$ ersetzt wird. Man kann also etwa den Hauptwert wählen.

von $-\frac{n\pi}{2}$ bis $\frac{n\pi}{2}$. Unstetigkeitsstellen treten für $n \operatorname{arctg} x = \frac{\pi}{2} + k\pi$, d. h. für $x = \operatorname{tg} \frac{\frac{\pi}{2} + k\pi}{n}$ auf, wobei k den Ungleichungen

$$-\frac{n\pi}{2} \leqq \frac{\pi}{2} + k\pi \leqq +\frac{n\pi}{2},$$

d. h.

$$-\frac{n+1}{2} \leqq k \leqq \frac{n-1}{2} \tag{181}$$

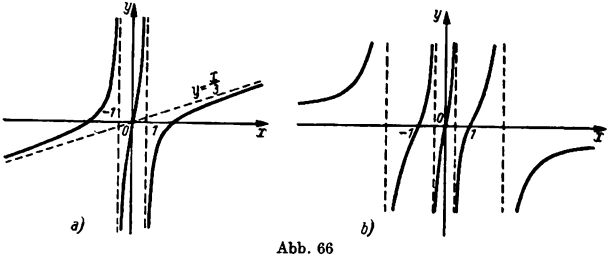


Abb. 66

genügt. Die Funktion $R_n(x)$ verschwindet überall dort, wo $n \operatorname{arctg} x = k\pi$, d. h. für $x = \operatorname{tg} \frac{k\pi}{n}$, wobei k den Ungleichungen

$$-\frac{n\pi}{2} \leqq k\pi \leqq +\frac{n\pi}{2},$$

d. h.

$$-\frac{n}{2} \leqq k \leqq \frac{n}{2} \tag{182}$$

genügt. Die Werte von k, die genau gleich den in (181) bzw. (182) angegebenen Grenzen sind, geben „Unstetigkeitsstellen" bzw. „Nullstellen" im Unendlichen. Geometrisch bedeutet eine „Unstetigkeit" für $x = \pm\infty$ eine geneigte Asymptote, während ein „Verschwinden" für $x = \pm\infty$ dem Fall entspricht, daß die x-Achse Asymptote ist. In Abb. 66 sind die graphischen Bilder von $R_3(x)$ und $R_4(x)$ dargestellt.

Kapitel III

GRENZWERTE VON ZAHLENFOLGEN UND VON FUNKTIONEN

§ 34. Endliche und unendliche Zahlenfolgen

Aus den Elementen einer gegebenen Menge (Gesamtheit) E von beliebigen Objekten kann man *Folgen* bilden. Dies geschieht auf folgende Weise: Man greift aus der Menge E ein beliebiges Element A_1 heraus und nimmt es als *erstes Glied* der Folge; sodann gibt man ein zweites Element A_2 an, das man in der Folge unmittelbar auf A_1 folgen läßt und welches das *zweite Glied* der Folge genannt wird; weiter gibt man ein Element A_3 an, das unmittelbar auf A_2 folgt und als *drittes Glied* der Folge bezeichnet wird, usw. Die Elemente A_1, A_2, A_3, ... brauchen nicht unbedingt paarweise verschieden zu sein, d. h., es kann mehrfach dasselbe Element genommen werden. Der Bildungsprozeß einer Folge besteht darin, daß zu einem bereits vorliegenden n-ten Glied, das mit der laufenden Nummer (dem *Index*) n behaftet ist, das ihm unmittelbar folgende $(n+1)$-te Glied mit dem Index $n+1$ angegeben wird.

Dieser Prozeß kann nun bei einem Objekt A_N, das einen gewissen Index N besitzt, abbrechen. Dieser Fall tritt ein, wenn kein $(N+1)$-tes Element benannt wird, das unmittelbar auf A_N folgt; dann heißt A_N das *letzte Glied* der Folge und der Index N gibt in diesem Fall die *Anzahl der Glieder* der Folge an. Die Folge selbst heißt im genannten Fall *endlich*.

Schreibt man die Glieder in ihrer Reihenfolge auf, so setzt man zwischen zwei aufeinanderfolgende Glieder üblicherweise ein Komma:

$$A_1,\ A_2,\ A_3,\ \ldots,\ A_N. \tag{1}$$

Die Punkte deuten im betrachteten Fall die nicht explizit aufgeführten Glieder der Folge, deren Index größer als 3 und kleiner als N ist, an.

Man kann sich aber auch den Bildungsprozeß einer Folge als einen *unbegrenzt fortgeführten Prozeß*, der also kein Ende hat, vorstellen. In diesem Fall existiert in der Folge kein letztes Glied; wie auch immer eine natürliche Zahl n gewählt wird, stets folgt auf das Folgeglied A_n mit dem Index n noch ein weiteres Glied A_{n+1}, dessen Index um Eins größer ist. Wir sagen dann, daß sich im Endergebnis des Bildungsprozesses eine *unendliche Folge* ergibt; das besagt nur, daß man, welche natürliche Zahl n man auch nimmt, immer ein Glied der

Folge finden kann, das den Index n besitzt, und daß umgekehrt auch jedes Glied der Folge eine gewisse natürliche Zahl als Index besitzt.[1])

Die Glieder einer Zahlenfolge werde im folgenden durch kleine Buchstaben bezeichnet. Das einfachste Verfahren, eine endliche Zahlenfolge anzugeben, besteht darin, daß man alle ihre Glieder, eines nach dem anderen, so wie sie in der Folge aufeinanderfolgen, hinschreibt:

$$a_1, a_2, a_3, \ldots, a_N. \tag{2}$$

Hier einige Beispiele für endliche Zahlenfolgen:

a) 1, 2, 3, 4, 5 $\quad (N=5)$,

b) 7, 2, 10 $\quad (N=3)$,

c) 1, 2, 3, 1, 2, 3, 1, 2 $\quad (N=8)$.

Für den Fall, daß die Anzahl N der Glieder sehr groß ist, wird natürlich dieses Hinschreiben aller Glieder eine recht mühsame Angelegenheit. Daher benutzt man zur Beschreibung auch von endlichen Folgen zuweilen eine analytische Methode. Hierbei verwendet man eine (elementare) Funktion $f(x)$, die wenigstens für die positiven ganzen Zahlen von 1 bis N erklärt ist und die so beschaffen ist, daß sie für $x=1$ als Wert gerade das erste Folgeglied a_1, für $x=2$ als Wert das Folgeglied a_2, usw., schließlich für $x=N$ als Wert das Folgeglied a_N annimmt. Um die Folge anzugeben, genügt es nun, die Funktion $f(x)$ und die Gliederanzahl N anzugeben; die Berechnung der Folgeglieder $f(1), f(2), f(3), \ldots, f(N)$ bereitet dann keine Schwierigkeiten.

---

[1]) Letzteres besagt folgendes: Wir betrachten z. B. die Gesamtheit der untereinander verschiedenen Zahlen der Form $1-\frac{1}{n}$, $1+\frac{1}{n}$, $2-\frac{1}{n}$, wobei für n jede natürliche Zahl außer 1 bzw. 2 gesetzt werden soll. Ordnen wir diese Zahlen nach ihrer Größe, so erhalten wir folgendes „geordnete" Zahlensystem:

$$\frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{5}{6}, \ldots, \frac{7}{6}, \frac{6}{5}, \frac{5}{4}, \frac{4}{3}, \frac{5}{3}, \frac{7}{4}, \frac{9}{5}, \frac{11}{6}, \frac{13}{7}, \ldots.$$

In diesem System gibt es ein erstes Glied, nämlich $\frac{2}{3}$, dem kein anderes Glied vorangeht; abgesehen hiervon geht jeder Zahl des Systems eine und nur eine Zahl unmittelbar voran, und jedem Glied folgt ein und nur ein Glied unmittelbar nach, so daß es u. a. im System kein letztes Glied gibt. Das betrachtete System bildet jedoch keine unendliche Folge. Ordnen wir nämlich dem Glied $\frac{2}{3}$ den Index 1, dem Glied $\frac{3}{4}$ den Index 2, allgemein dem Glied $1-\frac{1}{n}$ den Index $n-2$ zu, so würden wir alle natürlichen Zahlen als Indizes für die Glieder der Form $1-\frac{1}{n}$ verbrauchen, und für die Glieder der Form $1+\frac{1}{n}$ und $2-\frac{1}{n}$ würden die natürlichen Zahlen zur Indizierung nicht ausreichen. Geht man zu einer anderen Reihenfolge über, so kann man dieses System natürlich in mannigfacher Weise als unendliche Folge darstellen, z. B.:

$$\frac{2}{3}, \frac{4}{3}, \frac{5}{3}, \frac{3}{4}, \frac{5}{4}, \frac{7}{4}, \frac{4}{5}, \frac{6}{5}, \frac{9}{5}, \frac{5}{6}, \frac{7}{6}, \frac{11}{6}, \ldots.$$

Beispiele.

1. Für $f(x)=x^2$ und $N=5$ erhält man die Folge

$$1,\ 4,\ 9,\ 16,\ 25.$$

2. Für $f(x)=x^3-6x^2+11x-6$ und $N=4$ erhält man die Folge

$$0,\ 0,\ 0,\ 6.$$

3. Für $f(x)=(-1)^{x+1}$ und $N=8$ erhält man die Folge

$$1,\ -1,\ 1,\ -1,\ 1,\ -1,\ 1,\ -1.$$

Für eine endliche Folge mit N Gliedern kann man stets eine elementare Funktion angeben, die die gewünschte Bedingung erfüllt.[1])

Unendliche Folgen von beliebigen Objekten deutet man durch

$$A_1,\ A_2,\ A_3,\ \ldots \tag{3}$$

oder

$$A_1,\ A_2,\ A_3,\ \ldots,\ A_n,\ \ldots \tag{4}$$

an. Die ersten Punkte in der zweiten Schreibweise bezeichnen — wie früher — die nicht explizit hingeschriebenen endlich vielen Glieder zwischen A_3 und A_n, die Punkte am Ende der Zeile bedeuten nach einer allgemeinen Verabredung immer, daß unendlich viele Glieder weggelassen worden sind, oder besser gesagt, die Möglichkeit einer unbegrenzten Fortsetzung.

Häufig verwendet man auch die noch kürzere Schreibweise

$$\{A_n\}. \tag{5}$$

Das Folgeglied A_n, dessen Index durch einen neuen Buchstaben (im vorliegenden Fall n) angedeutet wird, heißt das *allgemeine Glied* der Folge.

Im vorliegenden Kapitel werden nur Zahlenfolgen, d. h. Folgen von (ausschließlich reellen) Zahlen, im folgenden Kapitel Folgen von Funktionen (Funktionenfolgen) behandelt. Später werden wir uns auch mit Folgen zu beschäftigen haben, die aus anderen Objekten gebildet sind.

Wie soll man nun eine unendliche Zahlenfolge angeben (definieren)? Am natürlichsten ist hier die analytische Methode, die in expliziter Form angibt, welche Operationen am Index n vorzunehmen sind, um das allgemeine Glied

$$a_n=f(n) \qquad (n=1,\ 2,\ 3,\ \ldots)$$

der Folge zu erhalten.

In jedem konkreten Fall kann es mehr oder weniger Mühe kosten, eine Funktion $f(x)$ zu finden, die das Verlangte leistet. Es ist nicht die Aufgabe dieses Artikels, diese Frage ausführlicher zu behandeln; ihr wird hier kein Platz eingeräumt. Wir bemerken lediglich

[1]) Eine solche ist z. B. für die Folge (2) die Funktion

$$f(x)=\sum_{m=1}^{N}(-1)^{N-m}\frac{(x-1)(x-2)\cdots(x-m+1)(x-m-1)\cdots(x-N)}{(m-1)!\,(N-m)!}\,a_m,$$

die ein rationales Polynom vom Grad $N-1$ ist (LAGRANGEsches Interpolationspolynom).

folgendes: In jedem Fall kann man die Funktion $f(x)$ noch auf verschiedene Weise abändern. So kann man im Beispiel 3 an die Stelle der Funktion $f(x) = (-1)^{x+1}$ etwa die Funktion $f(x) = \cos \pi x$ setzen; allgemein kann man an Stelle einer Funktion $f(x)$ stets auch die Funktion $f(x) + \sin \pi x$ nehmen usw.

Eine andere Methode für die Definition von Zahlenfolgen besteht darin, daß man sie durch eine *Rekursionsformel* beschreibt. Eine solche Rekursionsformel gibt an, welche Rechenoperationen an den bereits berechneten Folgegliedern

$$a_1, a_2, \ldots, a_n$$

auszuführen sind, um das folgende Glied a_{n+1} zu erhalten. Hierbei müssen (je nach dem Charakter der Rekursionsformel) noch ein oder mehrere Glieder der Folge als *Anfangsglieder* (Anfangsbedingung der Rekursion) vorgegeben werden.

Wir wollen die vorangehenden Ausführungen an einer Reihe von Beispielen erläutern, wobei wir der Einfachheit halber die entstehenden Folgen immer als unendliche Folgen ansehen wollen.

1. Eine *arithmetische Folge* (arithmetische Progression) wird durch die Rekursionsformel (Differenzengleichung)

$$a_{n+1} - a_n = d \qquad (n = 1, 2, 3, \ldots) \tag{6}$$

mit einer Anfangsbedingung der Form

$$a_1 = a$$

definiert. In der Schreibweise (4) stellt sich die Folge als

$$a, a + d, a + 2d, \ldots, a + (n-1)d, \ldots$$

dar. Ihr allgemeines Glied wird durch

$$a_n = a + (n-1)d \qquad (n = 1, 2, 3, \ldots)$$

beschrieben, und zwar gibt diese Gleichung sofort auch eine analytische Definition für die betrachtete Folge. Man sieht, daß die arithmetischen Folgen gerade diejenigen Folgen sind, die mit Hilfe einer linearen Funktion gewonnen werden können.

2. Wird das allgemeine Glied der Folge mittels einer Formel

$$a_n = P(n)$$

gegeben, wobei $P(x)$ ein Polynom vom Grad m in x ist, so nennt man die Folge $\{a_n\}$ eine *arithmetische Folge der Ordnung m*. Die gewöhnlichen arithmetischen Folgen sind somit die arithmetischen Folgen der Ordnung 1.

Beispiele für arithmetische Folgen höherer Ordnung sind

$$\{n^2\} \equiv 1, 4, 9, 26, 25, \ldots \qquad \text{Ordnung 2,}$$

$$\left\{\frac{n(n-1)}{2}\right\} \equiv 0, 1, 3, 6, 10, \ldots \qquad \text{Ordnung 2,}$$

$$\left\{\frac{n(n^2-1)}{6}\right\} \equiv 0, 1, 4, 10, 20, \ldots \qquad \text{Ordnung 3.}$$

Ist nun $P(x)$ ein Polynom vom Grad m, so ist

$$P(x+1)-P(x)$$

ein Polynom vom Grad $m-1$; denn ist

$$P(x)=ax^m+bx^{m-1}+\cdots+l \qquad (a\neq 0),$$

so ist

$$P(x+1)-P(x)=ma\,x^{m-1}+\left[\frac{m(m-1)}{2}a+(m-1)\,b\right]x^{m-2}+\cdots,$$

und mit a ist auch ma von Null verschieden. Hieraus folgt, daß *die „Differenzenfolge"*

$$\{a_{n+1}-a_n\}$$

einer arithmetischen Folge der Ordnung m eine arithmetische Folge der Ordnung $m-1$ *ist.* Man kann aber auch umgekehrt zeigen (worauf wir hier verzichten wollen), daß *die aus einer arithmetischen Folge der Ordnung m gebildete „Summenfolge"*

$$\{a_1+a_2+\cdots+a_n\}$$

eine arithmetische Folge der Ordnung $m+1$ *ist.* Man kann dies leicht an den oben genannten Beispielen nachprüfen. Diese Eigenschaft der arithmetischen Folgen kann man auch zur fortlaufenden Berechnung der Glieder einer arithmetischen Folge höherer Ordnung benutzen. Schreibt man z. B unter die oben aufgeschriebenen ersten fünf Glieder der Folge $\left\{\frac{n(n-1)}{2}\right\}$ die entsprechenden Differenzen, so erhält man den Anfang einer arithmetischen Folge erster Ordnung; setzt man diese fort und summiert dann, so erhält man die Fortsetzung der gegebenen Folge:

$$0,\ 1,\ 3,\ 6,\ 10,\ 15,\ 21,\ 28,\ \ldots$$
$$1,\ 2,\ 3,\ 4,\ \ 5,\ \ 6,\ \ 7,\ \ldots$$

Hieraus ersieht man, daß *sich die Glieder einer arithmetischen Folge beliebiger Ordnung allein mit Hilfe der Addition berechnen lassen.*

3. Eine *geometrische Folge* wird durch eine Rekursionsformel

$$\frac{a_{n+1}}{a_n}=q \qquad (n=1,\ 2,\ 3,\ \ldots) \tag{7}$$

mit einer Anfangsbedingung der Form

$$a_1=a$$

definiert. In der Schreibweise (4) stellt sich die Folge in der Form

$$a,\ aq,\ aq^2,\ \ldots,\ aq^{n-1},\ \ldots$$

dar. Ihr allgemeines Glied

$$a_n=aq^{n-1} \qquad (n=1,\ 2,\ 3,\ \ldots)$$

gibt zugleich eine analytische Definition für die geometrische Folge.

4. Die *Folge der Fakultäten* wird durch die Rekursionsformel

$$a_{n+1}=(n+1)\,a_n \qquad (n=1,\ 2,\ 3,\ \ldots)$$

mit der Anfangsbedingung

$$a_1=1$$

definiert. Sie lautet ausführlich geschrieben:

$$1, 2, 6, 24, 120, \ldots, n!, \ldots;$$

ihr allgemeines Glied wird durch

$$a_n = n! = 1 \cdot 2 \cdot 3 \cdots n$$

gegeben.

5. Eine *Fibonaccische Zahlenfolge*[1]) wird durch die Rekursionsformel

$$a_{n+2} = a_n + a_{n+1} \qquad (n = 1, 2, \ldots) \tag{8}$$

beschrieben, nach der jedes Glied (angefangen vom dritten) gleich der Summe der vorangehenden beiden Glieder ist. Als Anfangsbedingung muß man die ersten beiden Glieder vorgeben, z. B.

$$a_1 = 0, \quad a_2 = 1. \tag{9}$$

Dann erhält man auf Grund der Rekursionsformel (8) mühelos

$$a_3 = 0 + 1 = 1, \quad a_4 = 1 + 1 = 2, \quad a_5 = 1 + 2 = 3,$$

$$a_6 = 2 + 3 = 5, \quad a_7 = 3 + 5 = 8, \quad a_8 = 5 + 8 = 13,$$

$$a_9 = 8 + 13 = 21, \quad a_{10} = 13 + 21 = 34, \ldots .$$

Nachdem man auf diese Weise eine Anzahl von Fibonaccischen Zahlen erhalten hat,

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, \ldots, \tag{10}$$

ergibt sich natürlich die Frage, ob man diese Folge auch analytisch beschreiben kann, d. h., ob es möglich ist, das allgemeine Glied a_n explizit als Funktion des Index n (in „geschlossener Form"[2]) darzustellen. Das geht nun tatsächlich. Zuvor sei aber bemerkt, daß es nur eine Folge gibt, die den Forderungen (8) und (9) genügt; denn durch (9) sind die ersten beiden Glieder vorgegeben, daß dritte Glied ist durch die ersten beiden eindeutig bestimmt, das vierte Glied ergibt sich aus dem zweiten und dritten, usw.

Man prüft leicht nach, daß *das allgemeine Glied der Folge* (10) *der Fibonaccischen Zahlen durch*

$$a_n = \frac{1}{\sqrt{5}} \left\{ \left(\frac{1+\sqrt{5}}{2} \right)^{n-1} - \left(\frac{1-\sqrt{5}}{2} \right)^{n-1} \right\} \qquad (n = 1, 2, \ldots)$$

gegeben wird.

---

[1]) Benannt nach Fibonacci (Leonardo von Pisa) — 13. Jh. (Interessante Einzelheiten über die Fibonaccischen Zahlen finden sich in: N. N. Worobjow, Die Fibonaccischen Zahlen (Übers. aus dem Russ.), Deutscher Verlag der Wissenschaften, Berlin 1954).

[2]) Das bedeutet folgendes: In der Formel, die a_n durch n ausdrückt und die nur mit Hilfe der elementaren Operationen aufgebaut ist, darf die Anzahl der auszuführenden Operationen nicht von n abhängen. In diesem Sinn kann man von der Folge der Fakultäten nicht behaupten, ihr allgemeines Glied $a_n = n!$ sei „in geschlossener Form" geschrieben (obgleich die Schreibweise äußerst kurz ist).

Dem Leser, den diese Formel überrascht, geben wir gerne einen Wink, wie er sie gewinnen kann. Wir versuchen zunächst, ohne die Anfangsbedingungen (9) zu beachten, eine geometrische Folge zu finden, welche die Rekursionsgleichung (8) erfüllt. Setzen wir in (8) $a_n = aq^{n-1}$, so erhalten wir

$$aq^{n+1} = aq^{n-1} + aq^n,$$

was uns, wenn wir annehmen, daß $a \neq 0$ und $q \neq 0$, auf die quadratische Gleichung

$$q^2 = 1 + q \tag{11}$$

führt. Dies zeigt, daß der „Quotient" der Folge (wenn es überhaupt eine Folge der verlangten Art gibt) gleich einer der Zahlen

$$q = \frac{1+\sqrt{5}}{2} \quad \text{oder} \quad q' = \frac{1-\sqrt{5}}{2}$$

sein muß. Umgekehrt sieht man aber leicht, daß für beliebiges a und a' die Folgen

$$\left.\begin{array}{l} a,\ aq,\ aq^2,\ aq^3,\ \ldots,\ aq^{N-1},\ \ldots \\ a',\ a'q',\ a'q'^2,\ a'q'^3,\ \ldots,\ a'q'^{N-1},\ \ldots \end{array}\right\} \tag{12}$$

der Rekursionsgleichung (8) genügen. Mit den Folgen (12) erfüllt aber auch ihre „Summe"

$$a+a',\quad aq+a'q',\quad aq^2+a'q'^2,\quad aq^3+a'q'^3,\ \ldots,\ aq^{N-1}+a'q'^{N-1}$$

diese Rekursionsgleichung.[1]) Es genügt also, die Konstanten a und a' so zu wählen, daß auch die Anfangsbedingungen (9) erfüllt sind, d. h.

$$a + a' = 0,$$
$$aq + a'q' = 1$$

gilt. Dies erreicht man, indem man $a = \frac{1}{\sqrt{5}}$, $a' = -\frac{1}{\sqrt{5}}$ setzt.

6. Wir wollen nun die Folge betrachten, die sich dadurch von der Folge der Fibonaccischen Zahlen unterscheidet, daß jedes Glied nicht gleich der Summe, sondern gleich der „Halbsumme" der vorangehenden beiden Glieder ist,

$$a_{n+2} = \frac{a_n + a_{n+1}}{2} \qquad (n = 1,\ 2,\ 3,\ \ldots), \tag{13}$$

und die ebenfalls die Anfangsglieder

$$a_1 = 0, \quad a_2 = 1$$

besitzt. Wir erhalten

$$a_3 = \frac{0+1}{2} = \frac{1}{2}, \qquad a_4 = \frac{1+\frac{1}{2}}{2} = \frac{3}{4}, \qquad a_5 = \frac{\frac{1}{2}+\frac{3}{4}}{2} = \frac{5}{8},$$

$$a_6 = \frac{\frac{3}{4}+\frac{5}{8}}{2} = \frac{11}{16}, \quad a_7 = \frac{\frac{5}{8}+\frac{11}{16}}{2} = \frac{21}{32}, \quad a_8 = \frac{\frac{11}{16}+\frac{21}{32}}{2} = \frac{43}{64}, \ \ldots.$$

[1]) Addiert man nämlich die Gleichungen

$$aq^{n+1} = aq^{n-1} + aq^n \quad \text{und} \quad a'q'^{n+1} = a'q'^{n-1} + a'q'^n,$$

so erhält man

$$aq^{n+1} + a'q'^{n+1} = (aq^{n-1} + a'q'^{n-1}) + (aq^n + a'q'^n).$$

Das allgemeine Glied der Folge

$$0,\ 1,\ \frac{1}{2},\ \frac{3}{4},\ \frac{5}{8},\ \frac{11}{16},\ \frac{21}{32},\ \frac{43}{64},\ \ldots \tag{14}$$

kann man durch dasselbe Verfahren, wie es bei der Folge der FIBONACCIschen Zahlen verwendet wurde, gewinnen. An Stelle der quadratischen Gleichung (11) erhält man hier die quadratische Gleichung

$$q^2 = \frac{1+q}{2},$$

welche die Lösungen 1 und $-\frac{1}{2}$ besitzt. Wählt man die Konstanten in $a + (-1)^{n-1}\frac{a'}{2^{n-1}}$ dann wieder so, daß die Anfangsbedingungen erfüllt sind, so erhält man als allgemeines Glied der Folge (14)

$$a_n = \frac{2}{3}\left\{1 + \frac{(-1)^n}{2^{n-1}}\right\} \qquad (n = 1,\ 2,\ 3,\ \ldots).$$

7. Wird statt (13) verlangt, daß jedes Glied gleich dem arithmetischen Mittel der vorangehenden drei Glieder ist,

$$a_{n+3} = \frac{a_n + a_{n+1} + a_{n+2}}{3} \qquad (n = 1,\ 2,\ \ldots),$$

so muß man als Anfangsbedingung die ersten drei Glieder vorgeben. Die rekursive Berechnung der einzelnen Glieder bereitet keine Mühe; allerdings stößt man bei der Aufstellung der Formel für das allgemeine Glied a_n auf gewisse Schwierigkeiten.[1])

Eine endliche Zahlenfolge

$$a_1,\ a_2,\ a_3,\ \ldots,\ a_n,\ \ldots,\ a_N \tag{15}$$

kann man sich auf zweierlei Art geometrisch veranschaulichen.

Bei der zuerst zu schildernden geometrischen Darstellung benutzt man die x, y-Ebene. Man markiert dort die Punkte mit den Koordinaten

$$(n,\ a_n) \qquad (n = 1,\ 2,\ \ldots,\ N)$$

und zeichnet die vertikalen Strecken, die jeweils den Punkt (n, a_n) mit dem Punkt $(n, 0)$ auf der x-Achse verbinden. Die Glieder der Folge werden durch diese Strecken („Säulen") dargestellt. Ist ein Glied positiv, so ist die entsprechende Strecke nach oben gerichtet, ist es negativ, so ist sie nach unten gerichtet. Ist von zwei positiven Gliedern das eine größer als das andere, so ist die zum ersten gehörende Strecke länger als die zum zweiten gehörende, usw. Ist die Folge in der Form

$$a_n = f(n)$$

gegeben, wobei $f(x)$ eine elementare Funktion ist, so genügt es, das graphische Bild von $y = f(x)$ zu zeichnen und die vertikalen Strecken, die den ganzzahligen Abszissenwerten entsprechen, einzutragen.

---

[1]) Eine ähnliche Überlegung wie in den vorangehenden Beispielen führt auf eine Gleichung dritten Grades mit komplexen Wurzeln, die sich allerdings in der endgültigen Formel wieder ausschalten lassen.

Bei der zweiten Darstellungsweise für Zahlenfolgen kommt man mit einer Zahlengeraden aus. Auf der Zahlengeraden werden nacheinander die Punkte, die $a_1, a_2, \ldots$ entsprechen, markiert; damit man die Punkte nicht verwechselt, muß man jeden Punkt mit einem Index versehen. Ein Glied einer Folge ist dann und nur dann positiv bzw. negativ, wenn der zugeordnete Punkt auf dem positiven bzw. negativen Halbstrahl der Zahlengeraden liegt. Ist von zwei Gliedern das eine größer als das andere, so liegt der dem ersteren entsprechende Punkt rechts von dem zum zweiten gehörenden.

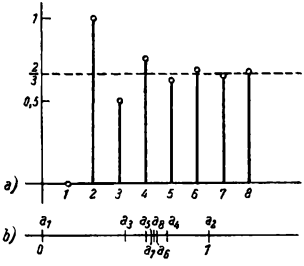


Abb. 67

Um von der zuerst genannten Darstellungsweise zur zweiten zu gelangen, braucht man nur die vertikalen Strecken auf die vertikale Achse zu projizieren und (nachdem man sie, wenn man will, um 90° gedreht hat) die Indizes an den Projektionen anzubringen.

In Abb. 67 sind die ersten acht Glieder der durch (13) definierten Zahlenfolge graphisch dargestellt. Ebenso kann man sich unendliche Zahlenfolgen durch entsprechende graphische Darstellungen veranschaulichen. Ein wesentlicher Unterschied besteht allerdings darin, daß es natürlich unmöglich ist, die angegebenen Konstruktionen jeweils für die gesamte unendliche Zahlenfolge auszuführen, sondern daß man stets bei einem gewissen Glied a_n abbrechen muß, von wo ab man alles weitere dem geometrischen Vorstellungsvermögen überläßt.

§ 35. Die allgemeine Definition von unendlichen Zahlenfolgen

Wenn wir im weiteren von Folgen sprechen, so meinen wir damit — dem allgemeinen mathematischen Sprachgebrauch folgend — stets unendliche Folgen.

Eine unendliche Folge unterscheidet sich von einer endlichen Folge grundsätzlich dadurch, daß eine endliche Folge (wie groß die Anzahl ihrer Glieder auch sei) durch direktes Aufzählen ihrer Glieder angegeben werden kann, während eine derartige Aufzählung bei unendlichen Zahlenfolgen prinzipiell unmöglich ist.

Dem genannten Sachverhalt wird ein unerfahrener Leser ohne weiteres zustimmen, und er wird auch nicht zögern, die Frage: „Wodurch kann man im Fall einer unendlichen Folge das direkte Aufzählen ersetzen?" dahingehend zu beantworten, daß für eine unendliche Folge eine „Gesetzmäßigkeit" charakteristisch ist, d. h., daß ein bestimmtes „Gesetz" angegeben werden muß, nach dem die Glieder der Folge, eines nach dem anderen, gebildet werden können. So findet man es auch vielfach in der Literatur.

Es kann nicht bestritten werden, daß diese Überlegungen im wesentlichen richtig sind; jedoch muß der Inhalt des Begriffs „Gesetz" bzw. „Gesetzmäßigkeit" einer ge-

naueren Analyse unterzogen werden, da Begriffsbildungen dieser Art (die zwar außerordentlich suggestiv sind) im mathematischen Sprachgebrauch keine ganz einheitliche Bedeutung haben.

Wenn man von „Gesetz" spricht, meint man meistens in erster Linie eine mathematische Formel, durch die (im betrachteten Fall der Folgen) das allgemeine Glied als Funktion des Index n gegeben wird. Wenn man den Begriff „Formel" oder „elementare Funktion" (so wie es in § 1 geschehen ist) näher präzisiert, so bedeutet die Angabe eines „Gesetzes" genau dasselbe wie die „analytische Definition einer Folge" in „geschlossener Form", etwa wie in § 34. Diese Methode, eine Folge zu definieren, ist sowohl im Fall der endlichen als auch im Fall der unendlichen Folgen recht brauchbar. Jedoch ist diese Erklärung für den Begriff „Gesetzmäßigkeit" ziemlich dürftig, werden doch durch sie augenscheinlich viele einfache „rekursive Folgen" nicht ohne weiteres erfaßt; denn bei einer rekursiven Definition bleibt doch stets die Frage offen, ob eine analytische Darstellung der Folge „in geschlossener Form" existiert oder nicht.

In vielen Fällen ist es nicht schwer, eine Funktion $f(x)$ zu finden, die eine irgendwie gegebene Folge analytisch beschreibt. So leistet z. B. für die Folge

$$(-1),\quad 1,\quad (-1),\quad 1,\quad (-1),\quad 1,\ \ldots \tag{16}$$

die Funktion $f(x) = \cos \pi x$ das Gewünschte. Schwieriger mag es schon für die Folge

$$1,\ 0,\ 0,\ 1,\ 0,\ 0,\ 1,\ 0,\ 0,\ \ldots \tag{17}$$

erscheinen, jedoch läßt sich auch hier eine Funktion $f(x)$ angeben, z. B. kann man

$$f(x) = \frac{1}{3}\left\{1 + 2\cos\left[\frac{2}{3}\pi(x-1)\right]\right\}$$

setzen. Ähnlich gelangt man auch bei jeder anderen „periodischen Folge"

$$a_1,\ a_2,\ \ldots,\ a_p,\quad a_1,\ a_2,\ \ldots,\ a_p,\quad a_1,\ a_2,\ \ldots,\ a_p,\ \ldots \tag{18}$$

mit einer beliebigen „Periode" p zum Ziel.

Jede derartige Folge wird andererseits durch die Rekursionsgleichung

$$a_{n+p} = a_n \qquad (n = 1,\ 2,\ 3,\ \ldots)$$

beschrieben, wobei die Glieder $a_1, \ldots, a_p$ noch als Anfangsbedingung beliebig vorgegeben werden können.

Ebenso definiert aber z. B. auch die Rekursionsgleichung

$$a_{n+1} - a_n = \frac{1}{n+1}$$

mit der Anfangsbedingung $a_1 = 1$ eine eindeutig bestimmte Folge, nämlich die Folge

$$1,\ 1+\frac{1}{2},\quad 1+\frac{1}{2}+\frac{1}{3},\ \ldots,\quad 1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n},\ \ldots, \tag{19}$$

und es liegt wohl kein Grund dafür vor, diese Folge nicht zu den „gesetzmäßigen" Folgen zu rechnen, obwohl keineswegs völlig klar ist, ob es eine elementare Funktion $f(x)$ gibt, die den unendlich vielen Bedingungen

$$f(n) = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \qquad (n = 1,\ 2,\ 3,\ \ldots)$$

genügt.

Zugleich muß man aber folgendes bedenken: Wenn man sich bei der Definition von Folgen nur auf die „analytische" Methode stützt, so macht man den Begriff „Folge"

von der zugelassenen Klasse von Funktionen $f(x)$ abhängig, denn es liegt natürlich kein Grund vor, sich nur auf die Klasse der „elementaren" Funktionen zu beschränken. Bei seiner weiteren Verallgemeinerung gewinnt der Funktionsbegriff jedoch so an Elastizität, daß man mit seiner Hilfe jede beliebige empirische Gesetzmäßigkeit widerspiegeln kann.[1]) Im Zusammenhang damit wird, wenn man sich so ausdrücken darf, die Vorstellung von „Gesetzmäßigkeit" auf eine höhere Stufe gehoben.

Nach den vorangehenden Überlegungen und den angeführten Beispielen wird es nicht mehr verwunderlich erscheinen, daß wir für den Begriff der Folge die nachstehende allgemeine Definition geben, die gegenwärtig in der Mathematik üblich ist:

Eine Zahlenfolge $\{a_n\}$ soll als gegeben angesehen werden, wenn eine Vorschrift angegeben ist, welche es gestattet, zu jedem vorgelegten Index, d. h. zu jeder beliebigen positiven ganzen Zahl n das Glied a_n, das an n-ter Stelle steht, eindeutig zu bestimmen. Mit anderen Worten: *Eine Folge $\{a_n\}$ ist gegeben, wenn jeder positiven ganzen Zahl n eine eindeutig bestimmte Zahl a_n zugeordnet ist.*

Über die Art der Vorschrift, welche die Zuordnung zwischen n und a_n herstellt, werden keine einschränkenden Voraussetzungen gemacht. Im einfachsten Fall, dem der „analytischen" Beschreibung $a_n = f(n)$, gibt die Funktion $f(x)$ an, welche elementaren mathematischen Operationen man auf die Zahl n anzuwenden hat, um die Zahl a_n zu erhalten. Die Zuordnungsvorschrift kann aber auch ganz anderer Natur sein.

Hier einige Beispiele, bei denen man die Zuordnungsvorschrift ohne Mühe erraten kann:

1. 0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, ...;
2. 1, 1, 2, 1, 2, 3, 1, 2, 3, 4, 1, 2, 3, 4, 5, 1, 2, 3, 4, 5, 6, ...;
3. 3, 1, 4, 1, 5, 9, 2, 6, 5, 3, 5, ...;
4. $\frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{3}{4}, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, \frac{1}{6}, \frac{5}{6}, \ldots$.

In jedem dieser Beispiele liegt eine „Gesetzmäßigkeit" (im weiteren Sinne) vor; die Folgen sind gegeben, wenn es auch keineswegs klar ist, ob eine elementare Funktion $f(x)$ existiert, für die $f(n) = a_n$ gilt, und es ebenfalls unmöglich ist, das Bestehen einfacher Rekursionsgleichungen zu erkennen.

In Beispiel 3 ist die Zuordnungsvorschrift folgende: Das Glied a_n ist die n-te Ziffer in der Dezimalbruchentwicklung der Zahl π, die das Verhältnis des Umfangs eines Kreises zu einem Durchmesser angibt.

In Beispiel 4 sind alle reduzierten echten Brüche in Form einer Folge aufgeschrieben, wobei die Reihenfolge durch folgende Bedingung festgelegt ist: Von zwei Brüchen $\frac{p}{q}$ und $\frac{p'}{q'}$ folgt der Bruch $\frac{p'}{q'}$ hinter $\frac{p}{q}$, wenn $q' > q$ oder wenn $q' = q$ und $p' > p$ ist. Um zu erfahren, welcher Bruch an der n-ten Stelle steht, muß man die Folge bis zu dieser Stelle fortsetzen und nachsehen, welcher Bruch dort auftritt.

Wir wollen nun einige wichtige Eigenschaften der unendlichen Zahlenfolgen behandeln.

---

[1]) Ein überzeugendes Beispiel hierfür wird in § 49 angegeben.

Aus endlich vielen vorgegebenen Zahlen, die paarweise voneinander verschieden sind, kann man stets die größte und die kleinste Zahl auswählen. Der Begriff der größten und der kleinsten Zahl erfordert eine geringfügige Präzisierung für den Fall, daß die vorgelegten Zahlen nicht notwendig paarweise voneinander verschieden sind. Zunächst betrachten wir eine endliche Zahlenfolge

$$a_1,\ a_2,\ a_3,\ \ldots,\ a_N. \tag{20}$$

Eine Zahl der Folge (20), die von keiner Zahl dieser Folge übertroffen wird, heißt *eine größte Zahl* (*ein Maximum*) der Folge und wird mit $\max(a_1, \ldots, a_N)$ bezeichnet. Offenbar ist

$$\max(a_1, a_2, \ldots, a_N) \geqq a_n \qquad (n = 1, 2, \ldots, N), \tag{21}$$

wobei aber für wenigstens einen Index n Gleichheit vorliegt.

Entsprechend heißt eine Zahl der Folge (20), die von keiner Zahl der Folge unterboten wird, *eine kleinste Zahl* (*ein Minimum*) der Folge und wird mit $\min(a_1, \ldots, a_N)$ bezeichnet. Für das Minimum gilt

$$\min(a_1, a_2, \ldots, a_N) \leqq a_n \qquad (n = 1, 2, \ldots, N), \tag{22}$$

wobei auch hier für wenigstens ein n das Gleichheitszeichen auftritt.

Die Existenz einer größten und einer kleinsten Zahl bei jeder endlichen Zahlenfolge unterliegt keinem Zweifel.[1])

Ganz anders verhalten sich die Dinge bei einer unendlichen Zahlenfolge: *In einer unendlichen Zahlenfolge braucht es weder ein größtes noch ein kleinstes Glied zu geben.* Zum Beispiel gibt es in der Folge $\{n\}$ der natürlichen Zahlen kein größtes Glied; in der Folge $\{-n\}$ der negativen ganzen Zahlen gibt es kein kleinstes Glied; in der Folge $\{(-1)^{n+1} n\}$ gibt es schließlich weder ein größtes noch ein kleinstes Glied.

Ein weiterer wesentlicher Unterschied zwischen endlichen und unendlichen Zahlenfolgen hängt mit dem Begriff der Beschränktheit zusammen.

Eine Folge (gleichgültig, ob endlich oder unendlich) heißt

a) *nach oben* (oder — wie man auch sagt — *nach rechts*) *beschränkt*, wenn alle ihre Glieder kleiner als eine gegebene Zahl M sind:

$$a_n < M \qquad (n = 1, 2, 3, \ldots); \tag{23}$$

b) *nach unten* (*nach links*) *beschränkt*, wenn alle ihre Glieder größer als eine gegebene Zahl m sind:

$$a_n > m \qquad (n = 1, 2, 3, \ldots); \tag{24}$$

c) (schlechthin) *beschränkt*, wenn sie sowohl nach oben als auch nach unten beschränkt ist, d. h., wenn es Zahlen M und m gibt, so daß

$$m < a_n < M \qquad (n = 1, 2, 3, \ldots). \tag{25}$$

Jede endliche Zahlenfolge ist offenbar beschränkt. Als obere Schranke M kann man jede Zahl M nehmen, die größer ist als das Maximum der Glieder,

[1]) Die größte und die kleinste Zahl einer Folge können natürlich durchaus übereinstimmen. Dieser Fall tritt dann und nur dann ein, wenn alle Glieder der Folge gleich sind.

als untere Schranke jede Zahl m, die kleiner ist als das Minimum der Glieder. Die entsprechende Aussage für unendliche Folgen ist wiederum falsch. Zum Beispiel ist die Folge $\{n\}$ der natürlichen Zahlen zwar nach unten, aber nicht nach oben beschränkt; die Folge $\{-n\}$ der negativen ganzen Zahlen ist nach oben, aber nicht nach unten beschränkt; die Folge $\{(-1)^{n+1}n\}$ schließlich ist weder nach oben noch nach unten beschränkt.

Jede unendliche arithmetische Folge erster Ordnung ist nach unten und nicht nach oben bzw. nach oben und nicht nach unten beschränkt, und zwar je nachdem, ob die „Differenz" der Folge positiv oder negativ ist. Jede geometrische Folge, deren „Quotient" größer als Eins ist, ist nach unten, aber nicht nach oben beschränkt; jede geometrische Folge, deren Quotient zwischen -1 und $+1$ liegt, ist sowohl nach oben als auch nach unten beschränkt; jede geometrische Folge, deren Quotient kleiner als -1 ist, ist weder nach oben noch nach unten beschränkt. Die unendliche Folge der Fibonaccischen Zahlen (10) ist nach unten, aber nicht nach oben beschränkt. Die Folge (14) ist sowohl nach oben als auch nach unten beschränkt. Beschränkt sind ferner alle periodischen Folgen, desgleichen auch die Folgen 1., 3. und 4. von Seite 141; die Folge 2. ist dagegen zwar nach unten, nicht aber nach oben beschränkt.

Die Folge

$$\left\{1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}\right\} \tag{19}$$

ist nach oben nicht beschränkt. Um dies zu beweisen, müssen wir zeigen, daß jede noch so große Zahl M von einem gewissen Glied a_n der Folge (19) übertroffen wird, d. h., man muß zu gegebenem M ein n so finden, daß

$$1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}\geqq M.$$

Dazu beachten wir, daß

$$1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{2^p}$$

$$=\left\{\begin{array}{l} 1+ \\ +\frac{1}{2}+ \\ +\frac{1}{3}+\frac{1}{4}+ \\ +\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}+ \\ +\ \cdots\cdots\cdots\cdots+ \\ +\frac{1}{2^{p-1}+1}+\cdots+\frac{1}{2^p} \end{array}\right\} > \left\{\begin{array}{l} 1+ \\ +\frac{1}{2}+ \\ +\frac{1}{4}+\frac{1}{4}+ \\ +\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+ \\ +\ \cdots\cdots\cdots\cdots+ \\ +\frac{1}{2^p}+\cdots+\frac{1}{2^p} \end{array}\right\} > \frac{p}{2}.^{1})$$

[1]) Die Summe in jeder Zeile ist nämlich $\geqq \frac{1}{2}$, und im ganzen sind es $p+1$ Zeilen.

Wählen wir also $p > 2M$, so ist

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{2^p} > M.$$

Eine unendliche Folge kann nach oben beschränkt sein, ohne daß sie ein Maximum besitzt. So ist z. B. die Zahl Eins eine obere Schranke für die Glieder der Folge

$$\left\{\frac{n}{n+1}\right\} \equiv \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \ldots,$$

welche jedoch kein Maximum besitzt.

Die Umkehrung hiervon ist selbstverständlich richtig: *Wenn die Folge* $\{a_n\}$ *ein maximales Glied besitzt, so ist sie nach oben beschränkt*; denn für jede Zahl M, die größer ist als das Maximum der Folge, gilt

$$a_n < M$$

für alle n.

§ 36. Der Satz von Bolzano und Weierstrass über die Existenz eines Häufungspunktes

Wir wollen zunächst einige der Begriffsbildungen, die bereits in früheren Artikeln und auch im vorliegenden Artikel vielfach benutzt wurden, genauer behandeln, um zu präzisieren, in welchem Sinn sie im weiteren verwendet werden.

Jede *reelle Zahl* x kann man sich in Form eines *unendlichen Dezimalbruchs*

$$x = x_0, x_1 x_2 x_3 \ldots$$

geschrieben denken. Dabei bedeutet x_0 die Anzahl der *Einer*, x_1 die Anzahl der *Zehntel*, x_2 die Anzahl der *Hundertstel*, usw., allgemein x_n die Anzahl der *Einheiten von der Ordnung* 10^{-n}, die „in der betrachteten Zahl enthalten sind". Auch die Zahlen, die sich in Form eines endlichen Dezimalbruchs darstellen lassen, kann man natürlich als unendliche Dezimalbrüche schreiben, und zwar als Dezimalbruch mit Periode Null, z. B.

$$4{,}83 = 4{,}83000 \ldots .$$

Daneben lassen diese Zahlen noch eine zweite Darstellungsweise zu, nämlich die als Dezimalbruch mit Periode Neun, z. B.

$$4{,}83 = 4{,}82999 \ldots .$$

Jede reelle Zahl kann also auf genau eine Art oder genau zwei Arten als unendlicher Dezimalbruch geschrieben werden. Umgekehrt stellt jeder unendliche Dezimalbruch genau eine reelle Zahl dar.

Sind a, b zwei reelle Zahlen mit $a < b$, so verstehen wir unter dem (abgeschlossenen) Intervall $[a, b]$ die Gesamtheit aller reellen Zahlen x, die

zwischen a und b liegen, einschließlich dieser beiden Zahlen (des *Anfangs-* und des *Endpunktes* des Intervalls), für die also

$$a \leqq x \leqq b$$

gilt. Die vom Anfangs- und Endpunkt des Intervalls verschiedenen Punkte aus dem Intervall heißen *innere Punkte*; Anfangs- und Endpunkt werden auch unter dem Begriff *Randpunkt* zusammengefaßt. Die Differenz $b - a$ wird die *Länge* des Intervalls $[a, b]$ genannt. Die Länge eines Intervalls ist stets positiv. Als *Mittelpunkt* des Intervalls $[a, b]$ bezeichnet man die Zahl (den Punkt) $\frac{a+b}{2}$.

Je zwei voneinander verschiedene reelle Zahlen a und b kann man als Endpunkte eines eindeutig bestimmten Intervalls ansehen, und zwar des Intervalls $[a, b]$ (falls $a < b$) oder des Intervalls $[b, a]$ (falls $b < a$). Die Länge dieses Intervalls ist gleich $|a - b|$. Die Zahl $|a - b|$ bezeichnet man auch als *Abstand* der Punkte a und b (Entfernung des Punktes a vom Punkt b oder des Punktes b vom Punkt a). Offenbar ist für beliebige reelle Zahlen a und b stets

$$|a - b| \geqq 0. \tag{26}$$

Nach Definition des absoluten Betrages bedeutet die Ungleichung

$$|x - c| < m \qquad (m > 0) \tag{27}$$

die gleichzeitige Gültigkeit der beiden Ungleichungen

$$x - c < m,$$
$$-(x - c) < m,$$

und dies ist gleichwertig dem Bestehen der Ungleichungskette

$$c - m < x < c + m. \tag{28}$$

Die Ungleichung (27) hat also stets die Ungleichungen (28) zur Folge und umgekehrt. Da c der Mittelpunkt des Intervalls $[c - m, c + m]$ der Länge $2m$ ist, kann man den genannten Sachverhalt auch folgendermaßen deuten: Die Entfernung eines Punktes x vom Punkt c ist dann und nur dann kleiner als m, wenn der Punkt x im Innern des Intervalls der Länge $2m$ um c als Mittelpunkt liegt.

Unter einer *Umgebung* des Punktes (der Zahl) c versteht man die Gesamtheit der inneren Punkte eines Intervalls, welches c im Innern enthält. Meistens hat man es mit Umgebungen zu tun, für die der gegebene Punkt sogar Mittelpunkt ist. Solche Umgebungen nennt man *m-Umgebungen*, und zwar versteht man unter der m-Umgebung ($m > 0$) des Punktes c genauer die Menge aller derjenigen Punkte, deren Abstand von c kleiner als m ist, d. h., für die (27) bzw. (28) gilt. Ist $0 < m_1 < m_2$, so gehört jeder Punkt der m_1-Umgebung von c auch der m_2-Umgebung von c an; die Umkehrung hiervon ist selbstverständlich nicht richtig.

Dem Leser sei empfohlen, sich den geometrischen Inhalt der vorangehenden Überlegungen auf der Zahlengeraden zu veranschaulichen. Wir erwähnen schließlich als Hilfssatz noch den folgenden Sachverhalt:

Vorgelegt sei eine „zweiseitige"[1] arithmetische Folge mit der Differenz $d > 0$. Dann enthält für jedes $m > d$ die m-Umgebung eines beliebigen Punktes c, der nicht der Folge angehört, beide Punkte der Folge, zwischen denen c liegt, und damit auch das ganze von ihnen gebildete Intervall. Ist dagegen der Punkt c selbst ein Punkt der Folge, so enthält seine m-Umgebung beide Nachbarpunkte von c in der Folge und auch die beiden zugehörigen Intervalle der Länge d.

Geometrisch ist dies vollkommen klar. Aber auch der analytische Beweis bereitet keinerlei Schwierigkeiten. Haben nämlich die Glieder der Folge allgemein die Form

$$a + (p-1)\,d \qquad (-\infty < p < +\infty),$$

so folgt für $m > d$ aus

$$a + (p-1)\,d \leqq c \leqq a + pd$$

unmittelbar

$$c - m \leqq a + pd - m < a + pd - d = a + (p-1)\,d,$$

$$c + m \geqq a + (p-1)\,d + m > a + (p-1)\,d + d = a + pd,$$

so daß

$$c - m < a + (p-1)\,d < a + pd < c + m$$

gilt.

Wir kommen nun zum Fundamentalsatz dieses Paragraphen:

Zu jeder beschränkten unendlichen Folge

$$\{a_n\} \equiv a_1,\ a_2,\ a_3,\ \ldots,\ a_n,\ \ldots \tag{29}$$

gibt es stets mindestens einen Punkt c, so daß in jeder (der Länge nach) noch so kleinen Umgebung von c unendlich viele Glieder der Folge (29) liegen.

Der Beweis dieses Satzes basiert auf einem wesentlichen arithmetischen Prinzip, das wir der größeren Anschaulichkeit wegen folgendermaßen formulieren:

Wenn in endlich vielen Schubfächern unendlich viele Dinge liegen, so liegen bereits in wenigstens einem dieser Schubfächer unendlich viele dieser Dinge.

Bevor wir zum Beweis des Fundamentalsatzes übergehen, wollen wir folgendes verabreden: Wenn unter den Gliedern a_n der Folge negative Zahlen auftreten sollten, so wollen wir sie uns als Dezimalbrüche mit negativer „Kennziffer", aber positiver „Mantisse" geschrieben denken, z. B.

$$-4{,}52038\ldots = \bar{5}{,}47961\ldots.$$

Wir denken uns nun die Folgeglieder a_n auf endlich viele Klassen („Schubfächer") verteilt, und zwar sollen Glieder a_n und a_m dann und nur dann derselben Klasse angehören, wenn sie in ihrer Dezimalbruchentwicklung dieselbe Kennziffer, d. h. denselben ganzzahligen Anteil besitzen (dieselbe ganze Zahl

[1] Das ist eine „Folge", die sich nicht nur nach rechts, sondern auch nach links unbeschränkt fortsetzt.

links vom Komma steht). Die Anzahl der hierbei auftretenden Klassen ist in der Tat endlich; denn da die Folge $\{a_n\}$ beschränkt ist, gibt es Zahlen m und M, so daß

$$m < a_n < M \qquad (n = 1, 2, 3, \ldots);$$

nehmen wir nun ohne Beschränkung der Allgemeinheit an, daß die Zahlen m und M ganz sind, so können wir behaupten, daß es höchstens so viele Klassen gibt wie Kennziffern

$$m,\ m+1,\ m+2,\ \ldots,\ M-1, \tag{30}$$

also $N = M - m$. Weil nun die Folgeglieder a_n (mehrfach auftretende Glieder auch mehrfach gezählt) eine unendliche Menge bilden, liegen in wenigstens einer der Klassen unendlich viele Glieder der Folge. Das heißt mit anderen Worten: Unter den Zahlen (30) gibt es wenigstens eine, wir bezeichnen sie mit c_0, so daß unendlich viele Glieder der Folge $\{a_n\}$ die Kennziffer c_0 besitzen, also für unendlich viele Glieder der Folge die Ungleichungen

$$c_0 \leqq x < c_0 + 1$$

erfüllt sind.

Wir zerlegen nun die Klasse, die zur Kennziffer c_0 gehört, ihrerseits in 10 Klassen, und zwar sollen zwei Folgeglieder aus der Klasse zu c_0 dann und nur dann derselben Klasse der neuen Klasseneinteilung angehören, wenn in ihren Dezimalbruchdarstellungen die erste Ziffer hinter dem Komma übereinstimmt. Da in der Klasse zu c_0 unendlich viele Glieder der Folge liegen, müssen auch in wenigstens einer Klasse der neuen Klasseneinteilung unendlich viele Glieder der Folge enthalten sein. Nehmen wir an, daß zu dieser Klasse die Ziffer c_1 als erste Ziffer hinter dem Komma gehört, so können wir behaupten, daß für unendlich viele Glieder der Folge die Ungleichungen

$$c_0, c_1 \leqq x < c_0, c_1 + \frac{1}{10}$$

erfüllt sind.

Im nächsten Schritt denken wir uns die Klasse, die alle die Folgeglieder enthält, deren Dezimalbruchentwicklung mit c_0, c_1 beginnt, ihrerseits in 10 Klassen zerlegt, und zwar so, daß Folgeglieder aus der betrachteten Klasse dann und nur dann derselben Klasse der neuen Klasseneinteilung angehören, wenn ihre Dezimalbruchentwicklungen in der zweiten Stelle hinter dem Komma übereinstimmen. Wieder enthält eine der Klassen, es sei die Klasse, die der Ziffer c_2 an der zweiten Stelle hinter dem Komma entspricht, unendlich viele Glieder der Folge, so daß also für unendlich viele Glieder der Folge die Ungleichungen

$$c_0, c_1 c_2 \leqq x < c_0, c_1 c_2 + \frac{1}{100}$$

erfüllt sind.

Denken wir uns diesen Prozeß unbeschränkt fortgesetzt, so erhalten wir eine unendliche Folge von Ziffern[1])

$$c_0,\ c_1,\ c_2,\ \ldots,\ c_n,\ \ldots,$$

[1]) Bis auf die „Ziffer" c_0, die eine beliebige ganze Zahl bezeichnet, sind alles Ziffern im üblichen Sinne, d. h. Zahlzeichen für die Zahlen von 0 bis 9.

die so beschaffen sind, daß in jedem der Intervalle

$$\left.\begin{array}{c} [c_0;\ \ c_0+1], \\ [c_0,c_1;\ \ c_0,c_1+10^{-1}], \\ [c_0,c_1c_2;\ \ c_0,c_1c_2+10^{-2}], \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ [c_0,c_1c_2\ldots c_n;\ \ c_0,c_1c_2\ldots c_n+10^{-n}] \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \end{array}\right\} \tag{31}$$

unendlich viele Glieder der Folge $\{a_n\}$ liegen.

Wir betrachten nun die Zahl c mit der Dezimalbruchdarstellung

$$c = c_0,c_1c_2\ldots c_n\ldots . \tag{32}$$

Wir behaupten, daß diese Zahl die verlangte Eigenschaft besitzt, d. h., daß in jeder Umgebung von c unendlich viele Glieder der Folge $\{a_n\}$ liegen. Zunächst ist klar, daß die Zahl c jedem der Intervalle (31) angehört. Wir betrachten nun irgendeine Umgebung des Punktes c, d. h. ein beliebig kleines Intervall, welches den Punkt c im Innern enthält. Ohne Beschränkung der Allgemeinheit können wir annehmen, daß c der Mittelpunkt des betrachteten Intervalls ist.[1]) Bezeichnen wir die Länge dieses Intervalls mit 2ε, so haben wir es mit einer ε-Umgebung des Punktes c zu tun, d. h., die betrachtete Umgebung besteht aus allen den Zahlen x, für die

$$|x-c| < \varepsilon$$

oder

$$c-\varepsilon < x < c+\varepsilon$$

gilt.

Wir wählen nun die Zahl n so groß, daß

$$\varepsilon > 10^{-n}, \tag{33}$$

und betrachten die „zweiseitige" arithmetische Folge

$$\ldots, -3\cdot 10^{-n},\ -2\cdot 10^{-n},\ -1\cdot 10^{-n},\ 0,\ 1\cdot 10^{-n},\ 2\cdot 10^{-n},\ 3\cdot 10^{-n}, \ldots .$$

Nach dem oben bewiesenen Hilfssatz enthält wegen (33) die ε-Umgebung von c die beiden c benachbarten Glieder der betrachteten arithmetischen Folge, also die Zahlen

$$c_0,c_1c_2\ldots c_n \quad \text{und} \quad c_0,c_1c_2\ldots c_n+10^{-n}$$

und damit auch das von ihnen gebildete Intervall

$$[c_0,c_1c_2\ldots c_n;\ \ c_0,c_1c_2\ldots c_n+10^{-n}] \tag{34}$$

---

[1]) Wenn bewiesen ist, daß jedes Intervall mit Mittelpunkt c unendlich viele Punkte der Folge $\{a_n\}$ enthält, so gilt diese Behauptung auch für jedes andere Intervall, welches c im Innern enthält, denn in jedes Intervall, welches c im Innern enthält, kann man stets ein Intervall „einbetten", das c als Mittelpunkt hat.

(dies ist auch dann der Fall, wenn c mit einer der Zahlen zusammenfällt). Da aber im Intervall (34) unendlich viele Glieder der Folge $\{a_n\}$ liegen, enthält dann auch die betrachtete ε-Umgebung von c unendlich viele Glieder der Folge, was zu beweisen war.

Es bleibt noch zu bemerken, daß die Ungleichung (33) mit der Ungleichung

$$n > \lg \frac{1}{\varepsilon} \tag{35}$$

gleichwertig ist.

Einen Punkt mit der Eigenschaft, daß in jeder seiner Umgebungen unendlich viele Glieder einer gegebenen Folge liegen, nennt man einen *Häufungspunkt* der betrachteten Folge. Unter Verwendung dieser Begriffsbildung können wir den zuletzt bewiesenen Satz auch folgendermaßen formulieren:

Satz von Bolzano und Weierstrass. *Jede beschränkte unendliche Folge besitzt wenigstens einen Häufungspunkt.*

Bemerkung 1. Beim obigen Beweis des Satzes von Bolzano und Weierstrass wurde nirgends davon Gebrauch gemacht, daß die Menge der betrachteten Zahlen a_n ($n = 1, 2, 3, \ldots$) als Folge angeordnet, d. h., daß sie abzählbar (vgl. EdEM, Band I, S. 76) ist. Daher gilt der Satz nicht nur für Zahlenfolgen, sondern für beliebige unendliche beschränkte Mengen von Zahlen.

Bemerkung 2. Bei Mengen[1]) kann man die Definition des Häufungspunktes noch dahingehend abschwächen, daß man nur verlangt, daß *in jeder noch so kleinen Umgebung des betrachteten Punktes wenigstens ein von ihm verschiedener Punkt der gegebenen Menge liegt.*

Hierzu beweisen wir: *Wenn in jeder Umgebung eines Punktes c wenigstens ein von c verschiedener Punkt einer gegebenen Menge E liegt, so liegen in jeder Umgebung von c sogar unendlich viele Punkte der Menge E.*

In der Tat: Wir betrachten für beliebiges $\varepsilon > 0$ die ε-Umgebung von c. Nach Voraussetzung gibt es in dieser Umgebung wenigstens einen von c verschiedenen Punkt c_1 der Menge E. Wir betrachten dann eine ε_1-Umgebung von c, wobei ε_1 der Bedingung $\varepsilon_1 < |c_1 - c|$ genüge. Auch in dieser Umgebung gibt es nach Voraussetzung einen von c verschiedenen Punkt c_2 der Menge E, und dieser ist wegen $\varepsilon_1 < |c_1 - c|$ offenbar von c_1 verschieden. Weiterhin betrachten wir dann eine ε_2-Umgebung von c mit $\varepsilon_2 < |c_2 - c|$. Auch in dieser Umgebung gibt es einen von c verschiedenen Punkt c_3 der Menge E, der offenbar von c_1 und c_2 verschieden ist, usw. Indem wir in der angegebenen Weise fortfahren, erhalten wir unendlich viele (verschiedene) Punkte der Menge E, die sämtlich in der ε-Umgebung von c liegen, womit unsere Behauptung bewiesen ist.

---

[1]) Für Folgen ist die entsprechende Abschwächung im allgemeinen nicht zulässig. Zum Beispiel ist die Zahl 1 Häufungspunkt der Zahlenfolge 1, 1, 1, 1, ... (denn in jeder Umgebung von 1 liegen unendlich viele — sogar alle — Glieder der Folge), ohne daß auch nur in einer Umgebung von 1 wenigstens ein von 1 verschiedenes Glied der Folge liegt (denn ein solches existiert ja gar nicht). Die Bedingung „ein von ihm verschiedener Punkt" ist selbstverständlich in der abgeschwächten Fassung nicht entbehrlich, da sonst jeder Punkt der Menge Häufungspunkt wäre. — *Anm. d. wissenschaftl. Red.*

§ 37. Beispiele. Der Grenzwert als einziger Häufungspunkt

Jede unendliche Folge besitzt eine gewisse Anzahl von Häufungspunkten. Diese Anzahl kann gleich Null sein; es gibt aber auch Folgen, die unendlich viele Häufungspunkte besitzen. Dabei kann sowohl der Fall eintreten, daß ein Häufungspunkt der Folge „angehört", d. h. gleich einem ihrer Glieder ist (vielleicht sogar mehrfach), als auch, daß er ihr nicht angehört. Von der Vielfalt der hier bestehenden Möglichkeiten sollen die folgenden Beispiele einen Eindruck vermitteln. Dem Leser sei empfohlen, sich die jeweiligen Verhältnisse dadurch zu veranschaulichen, daß er sich die Folgeglieder als Punkte auf der Zahlengeraden darstellt.

1. Die Folge

$$\left\{\frac{1}{n}\right\} \equiv 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$$

hat als einzigen Häufungspunkt die Zahl Null (vgl. Abb. 68a).

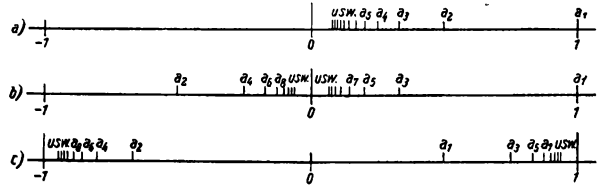


Abb. 68

2. Dasselbe gilt von der Folge

$$0, 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots,$$

die sich von der vorangehenden Folge nur dadurch unterscheidet, daß die Zahl Null als erstes Folgeglied hinzugekommen ist, während alle anderen Glieder um eine Stelle nach rechts gerückt sind. Ebenso besitzt aber auch die Folge

$$0, 1, 0, \frac{1}{2}, 0, \frac{1}{3}, 0, \frac{1}{4}, \ldots$$

die Zahl Null als einzigen Häufungspunkt. Sie unterscheidet sich von der Folge $\left\{\frac{1}{n}\right\}$ dadurch, daß vor das erste Glied und zwischen je zwei aufeinanderfolgende Glieder die Zahl Null getreten ist.

3. Auch alle Folgen

$$\left\{\frac{1}{n^\alpha}\right\} \equiv 1, \frac{1}{2^\alpha}, \frac{1}{3^\alpha}, \frac{1}{4^\alpha}, \ldots$$

mit $\alpha > 0$ besitzen die Zahl Null als einzigen Häufungspunkt.

4. Ebenso besitzt die „alternierende" Folge

$$\left\{\frac{(-1)^{n+1}}{n}\right\} \equiv 1, -\frac{1}{2}, \frac{1}{3}, -\frac{1}{4}, \dots$$

die Zahl Null als einzigen Häufungspunkt (Abb. 68b), wie allgemeiner jede Folge, die man aus der Folge $\left\{\frac{1}{n}\right\}$ dadurch erhält, daß man jedem Glied dieser Folge noch ein beliebiges Vorzeichen gibt.

5. Jede geometrische Folge, deren Quotient q der Bedingung $|q| < 1$ genügt, hat die Zahl Null als einzigen Häufungspunkt.

6. Die Folgen

$$\left\{a+\frac{1}{n}\right\}, \quad \left\{a+\frac{(-1)^n}{n}\right\}, \quad \left\{a+\frac{1}{n^\alpha}\right\} \text{ mit } \alpha>0, \quad \{a+q^n\} \text{ mit } |q|<1$$

besitzen sämtlich die Zahl a als einzigen Häufungspunkt.

7. Die Folge

$$\left\{(-1)^{n+1}\frac{n}{n+1}\right\} \equiv \frac{1}{2}, -\frac{2}{3}, \frac{3}{4}, -\frac{4}{5}, \dots$$

besitzt genau zwei Häufungspunkte, nämlich 1 und -1 (Abb. 68c).

8. Dasselbe gilt von der Folge

$$\{\cos n\pi\} \equiv \{(-1)^n\} \equiv -1, 1, -1, 1, \dots .$$

9. Besitzt die Folge $\{a_n\}$ die Zahl a und die Folge $\{b_n\}$ die Zahl b jeweils als einzigen Häufungspunkt, so besitzt, falls $a \neq b$ ist, die Folge

$$a_1, b_1, a_2, b_2, a_3, b_3, \dots \tag{36}$$

genau die beiden Häufungspunkte a und b, während im Fall $a = b$ die genannte Folge genau einen Häufungspunkt besitzt, nämlich $a = b$. Zum Beispiel besitzt also die Folge

$$\frac{1}{1}, \frac{2}{1}, \frac{1}{2}, \frac{2}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{2}{4}, \frac{1}{5}, \frac{2}{5}, \dots$$

die Zahl Null als einzigen Häufungspunkt.

Hieran ändert sich nichts, wenn man die Glieder der Folgen $\{a_n\}$ und $\{b_n\}$ nicht wie in (36) „abwechselnd" nimmt, sondern die neue Folge in beliebiger anderer Weise aus den Folgen $\{a_n\}$ und $\{b_n\}$ „zusammensetzt" (natürlich so, daß dabei alle Glieder jeder der beiden Folgen verbraucht werden).

10. Die Folge

$$\left\{\sin n\frac{\pi}{2}\right\} \equiv 1, 0, -1, 0, 1, 0, -1, 0, \dots$$

besitzt drei Häufungspunkte, nämlich 1, 0, -1. Allgemein hat jede periodische Zahlenfolge genau diejenigen Zahlen als Häufungspunkte, die als Glieder innerhalb der Periode auftreten.

11. Besitzt die Folge $\{a_n\}$ die Zahl a, die Folge $\{b_n\}$ die Zahl b, usw., die Folge $\{l_n\}$ die Zahl l jeweils als einzigen Häufungspunkt, so besitzt die Folge

$$a_1, b_1, \ldots, l_1, \quad a_2, b_2, \ldots, l_2, \quad a_3, b_3, \ldots, l_3, \ldots$$

genau die Zahlen $a, b, \ldots, l$ als Häufungspunkte.

12. Die Folgen in den Zeilen des Schemas

$$\begin{array}{llll}
\frac{1}{1}+\frac{1}{1^2}, & \frac{1}{1}+\frac{1}{2^2}, & \frac{1}{1}+\frac{1}{3^2}, & \frac{1}{1}+\frac{1}{4^2}, \ldots \\
\frac{1}{2}+\frac{1}{1^2}, \nearrow & \frac{1}{2}+\frac{1}{2^2}, \nearrow & \frac{1}{2}+\frac{1}{3^2}, \nearrow & \frac{1}{2}+\frac{1}{4^2}, \ldots \\
\frac{1}{3}+\frac{1}{1^2}, \nearrow & \frac{1}{3}+\frac{1}{2^2}, \nearrow & \frac{1}{3}+\frac{1}{3^2}, \nearrow & \frac{1}{3}+\frac{1}{4^2}, \ldots \\
\frac{1}{4}+\frac{1}{1^2}, \nearrow & \frac{1}{4}+\frac{1}{2^2}, \nearrow & \frac{1}{4}+\frac{1}{3^2}, \nearrow & \frac{1}{4}+\frac{1}{4^2}, \ldots \\
\ldots & \ldots & \ldots & \ldots
\end{array}$$

besitzen der Reihe nach $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$ jeweils als einzigen Häufungspunkt. Bilden wir aus den Gliedern dieser unendlich vielen unendlichen Folgen eine neue Folge, indem wir die Glieder entsprechend der eingezeichneten Pfeile aufeinander folgen lassen, so erhalten wir die Diagonalfolge

$$2, \frac{3}{2}, \frac{5}{4}, \frac{4}{3}, \frac{3}{4}, \frac{10}{9}, \frac{5}{4}, \frac{7}{12}, \frac{11}{18}, \frac{17}{16}, \ldots,$$

die neben den Zahlen $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$ noch die Zahl Null als Häufungspunkt besitzt (vgl. dazu den Satz von Seite 153).

13. Die Folge 1. von Seite 141 besitzt die beiden Häufungspunkte 0 und 1.

14. Die Folge 3. von Seite 141 kann höchstens die zehn Zahlen

$$0, 1, 2, 3, 4, 5, 6, 7, 8, 9$$

als Häufungspunkte besitzen. Ob diese Zahlen alle Häufungspunkte sind, können wir hier nicht entscheiden.

15. Die Folge 4. von Seite 141 hat alle Zahlen des Intervalls $[0, 1]$ als Häufungspunkte, denn jedes Intervall enthält unendlich viele rationale Zahlen (die rationalen Zahlen liegen überall dicht).[1])

17. Ist θ eine beliebige irrationale Zahl, so besitzt die Folge $\{\sin n\pi\theta\}$ alle Zahlen des Intervalls $[-1, 1]$ als Häufungspunkte.

Den Beweis des zuletzt genannten Satzes führen wir indirekt. Wir stellen uns dazu vor, daß auf der Peripherie des Einheitskreises diejenigen Punkte P_n markiert sind, deren

[1]) Vgl. EdEM, Band 1, A. J. Chintschin, Elemente der Zahlentheorie, Kap. IV und V.

zugehöriger Zentriwinkel die Form $n\pi\theta$ besitzt. Ferner seien die Projektionen P_n' der Punkte P_n auf den vertikalen Durchmesser eingezeichnet.

Gäbe es nun auf diesem Durchmesser ein Intervall, in dem kein Punkt P_n' liegt, so gäbe es auch auf der Peripherie einen Kreisbogen, der keinen Punkt P_n enthält. Es wäre zunächst denkbar, daß es mehrere solcher Kreisbögen gibt; dann gibt es aber, weil die Summe ihrer „Längen" den Umfang 2π des Einheitskreises nicht übertreffen kann, einen Bogen L maximaler Länge λ (wobei $0 < \lambda < 2\pi$). Drehen wir nun den Bogen L um den Winkel $2n\pi$ um den Mittelpunkt, so erhalten wir (auf der Peripherie des Kreises) denselben Bogen. Geht umgekehrt bei der Drehung um einen Winkel φ der Bogen L in sich über, so muß der Winkel φ ein Vielfaches von 2π sein.

Wir wollen nun Bögen *homolog* nennen, wenn sie durch Drehung um einen Winkel, der ein Vielfaches von $\pi\theta$ ist, ineinander übergeführt werden können. Offenbar enthält jeder zu L homologe Bogen ebenfalls keinen der Punkte P_n.

Wir betrachten nun die Menge der Bögen $L_n (n = 1, 2, 3, \ldots)$, die homolog zum Bogen $L_0 = L$ sind. Da die Länge jedes dieser Bögen gleich $\lambda > 0$ ist und es ihrer unendlich viele gibt, haben wenigstens zwei von ihnen, etwa L_p und L_q $(p < q)$ einen Bogen M einer Länge $\mu > 0$ gemeinsam. Es sind nun zwei Fälle möglich: (1) $\mu = \lambda$; (2) $\mu < \lambda$. Der erste Fall kann aber offenbar nicht eintreten. Ist nämlich $\mu = \lambda$, so stimmen bei den Bögen L_p und L_q sowohl die Anfangs- als auch die Endpunkte überein, so daß der Drehwinkel, der L_p in L_q überführt, ein Vielfaches von 2π sein muß. Dieser Winkel ist aber gleich $(q-p)\pi\theta$. Im Fall (1) müßte also

$$(q-p)\pi\theta = 2r\pi \qquad (r \text{ ganz}),$$

d. h.

$$\theta = \frac{2r}{q-p}$$

sein, im Widerspruch zur Irrationalität von θ.

Aber auch der zweite Fall kann nicht eintreten. Da nämlich weder L_p noch L_q einen der Punkte P_n enthält, enthält auch die Vereinigung dieser Bögen (d. h. der Bogen, der aus den Punkten der Kreisperipherie besteht, die auf wenigstens einem der Bögen L_p und L_q liegen) keinen der Punkte P_n. Die Länge dieses Bogens ist aber gleich

$$2\lambda - \mu > \lambda,$$

was der Wahl des Bogens L widerspricht (L sollte unter den Bögen, die keinen der Punkte P_n enthalten, maximale Länge besitzen).

Satz. *Jeder Häufungspunkt der Menge der Häufungspunkte einer Punktmenge E ist zugleich Häufungspunkt von E.*

Beweis. Es sei E' die Menge der Häufungspunkte von E und c'' ein Häufungspunkt dieser Menge. Nach Definition des Häufungspunktes gibt es dann in jeder Umgebung von c'' unendlich viele Punkte der Menge E', während es in jeder Umgebung eines Punktes c' von E' unendlich viele Punkte der Menge E gibt. Es ist zu zeigen, daß es in jeder Umgebung von c'' wenigstens einen von c'' verschiedenen Punkt der Menge E gibt. Dazu betrachten wir die ε-Umgebung von c''. In der $\frac{\varepsilon}{2}$-Umgebung von c'' finden wir zunächst einen von c'' verschiedenen Punkt c' der Menge E', für den also

$$|c' - c''| < \frac{\varepsilon}{2}$$

gilt. Ferner gibt es in der $\frac{\varepsilon}{2}$-Umgebung von c' unendlich viele Punkte der Menge E, also sicher einen Punkt c der Menge E, der von c'' verschieden ist. Für den Punkt c gilt dann

$$|c - c'| < \frac{\varepsilon}{2}.$$

Auf Grund einer bekannten Eigenschaft des absoluten Betrages[1]) ergibt sich sodann

$$|c - c''| \leqq |c - c'| + |c' - c''| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

womit der behauptete Satz bereits bewiesen ist.

Bemerkung. Der vorangehende Satz und sein Beweis bleiben richtig, wenn man dort die Menge E durch eine Folge $\{a_n\}$ ersetzt.

Besitzt eine beschränkte Folge nur einen Häufungspunkt, so nennt man diesen den *Limes* (oder *Grenzwert*) der Folge.

Von einer unbeschränkten Folge und einer beschränkten Folge mit mehreren Häufungspunkten sagt man, sie besitze keinen Limes.

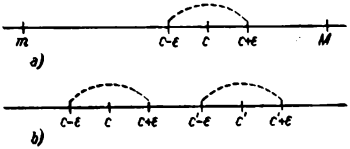


Abb. 69

In diesem Sinn besitzen die in den Beispielen 1 bis 6 behandelten Folgen einen Limes, und zwar ist es in den Beispielen 1 bis 5 die Zahl Null und in Beispiel 6 die Zahl a, während die in den Beispielen 7 bis 16 behandelten Folgen keinen Grenzwert besitzen.[2])

Wenn ein Punkt c nicht nur Häufungspunkt, sondern sogar Limes (d. h. einziger Häufungspunkt) einer gegebenen beschränkten Folge $\{a_n\}$ ist, so liegen nicht nur in jeder Umgebung von c unendlich viele, sondern sogar *fast alle* Glieder der Folge. „Fast alle" bedeutet hierbei — nach einem allgemeinen mathematischen Sprachgebrauch — alle Glieder mit Ausnahme von höchstens endlich vielen. Wie früher, kann man sich auch hier auf die Betrachtung von Intervallen beschränken, die c als Mittelpunkt besitzen, also auf ε-Umgebungen des Punktes c (vgl. Abb. 69a), wobei wir, wenn alle Glieder der Folge zwischen m und M liegen, noch annehmen können, daß $\varepsilon < M - c$ und $\varepsilon < c - m$, so daß das Intervall $[c - \varepsilon, c + \varepsilon]$ ganz innerhalb des Intervalls $[m, M]$ liegt. Dies sieht man leicht „indirekt" ein. Nach Voraussetzung liegen alle Glieder der Folge im Intervall $[m, M]$. Würden nun nicht fast alle Glieder der Folge bereits im Innern des Intervalls $[c - \varepsilon,\ c + \varepsilon]$ liegen, so würden unendlich viele Glieder der Folge in den Intervallen $[m, c - \varepsilon]$

[1]) Vgl. S. 160.

[2]) Dies trifft auch auf die in Beispiel 14 behandelte Folge zu. Man sieht nämlich leicht, daß die in diesem Beispiel angegebene Folge wenigstens zwei Häufungspunkte besitzen muß. Anderenfalls würde nämlich die Zahl π durch einen periodischen Dezimalbruch dargestellt und wäre demnach rational (vgl. EdEM, Band 1, S. 277, Satz 4), was aber nachweisbar nicht der Fall ist.

und $[c+\varepsilon, M]$ zu finden sein. Dann müßte aber bereits wenigstens eines dieser Intervalle undenlich viele Glieder der Folge enthalten, und dann müßte nach dem Satz von BOLZANO und WEIERSTRASS auch wenigstens ein Häufungspunkt der Folge in dem entsprechenden Intervall zu finden sein. Das ist aber unmöglich, wenn c der einzige Häufungspunkt der Folge ist. Wir haben damit gezeigt, daß fast alle Glieder der Folge bereits im Intervall $[c-\varepsilon, c+\varepsilon]$ liegen.

Liegen umgekehrt in jeder Umgebung eines Punktes c fast alle Glieder einer Folge $\{a_n\}$, so ist c der einzige Häufungspunkt dieser Folge. Auch dies beweisen wir „indirekt". Wir nehmen also an, daß es außer c noch einen weiteren Häufungspunkt c' $(\neq c)$ gibt. Ohne Beschränkung können wir annehmen, daß $c'>c$ ist (vgl. Abb. 69b). Da c' Häufungspunkt sein soll, liegen in jeder ε-Umgebung von c' unendlich viele Glieder der Folge $\{a_n\}$. Andererseits liegen jeweils in der entsprechenden ε-Umgebung von c fast alle Glieder der Folge (also alle, bis auf höchstens endlich viele Ausnahmen). Wählen wir nun ε so, daß $\varepsilon<\frac{c'-c}{2}$, so wird $c+\varepsilon<c'-\varepsilon$, d. h., die entsprechenden ε-Umgebungen von c bzw. c' haben keinen Punkt gemeinsam; dies ergibt offenbar einen Widerspruch.

Wir wollen schließlich den Begriff „fast alle" noch etwas variieren. Wenn fast alle Glieder einer Folge in einer gewissen ε-Umgebung eines Punktes c liegen, so bedeutet dies, daß nur endlich viele Glieder der Folge außerhalb dieser Umgebung liegen. Jedes dieser Folgeglieder hat nun einen ganz bestimmten Index, und da die Anzahl dieser Folgeglieder endlich ist, gibt es unter diesen Indizes einen größten N. Dieser Index N hängt natürlich von der gewählten Umgebung, d. h. von ε ab. Um dies hervorzuheben, hängt man manchmal an N das ε als Index an, schreibt also statt N ausführlicher N_ε. Von den Gliedern der Folge $\{a_n\}$, deren Index nicht größer als N ist, liegen eventuell einige in der ε-Umgebung von c, andere außerhalb dieser Umgebung; von den Gliedern der Folge, deren Index größer als N ist, kann man jedoch mit Sicherheit sagen, daß sie alle in der ε-Umgebung von c zu finden sind.

Auf Grund dieser Überlegungen kann man den Begriff des Limes einer Folge anders fassen. Dies führt zu einer zweiten Definition des Grenzwertes, die der oben angeführten Definition logisch gleichwertig ist und mit der wir uns sogleich genauer befassen werden (sie ist historisch gesehen die ältere).[1])

§ 38. Die klassische Definition des Grenzwertes einer Folge und seine wesentlichsten Eigenschaften

Eine Zahl a heißt *Grenzwert* oder *Limes* („im eigentlichen Sinn") einer Zahlenfolge

$$\{a_n\} \equiv a_1,\ a_2,\ a_3,\ \ldots,\ a_n,\ \ldots,$$

[1]) Diese Definition geht auf A. CAUCHY zurück. Sie ist heute allgemein verbreitet und wird in der mathematischen Praxis laufend verwendet.

wenn man zu jeder noch so kleinen positiven Zahl ε eine Zahl $N = N_\varepsilon$ so finden kann, daß für alle Folgeglieder a_n mit

$$n > N \tag{37}$$

die Ungleichung

$$|a_n - a| < \varepsilon \tag{38}$$

gilt.[1])

Häufig drückt man dies auch folgendermaßen aus (was aber viel ungenauer ist): Die Zahl a ist Grenzwert der Folge $\{a_n\}$, wenn für hinreichend großen Index n der Abstand des Folgegliedes a_n von a beliebig klein wird.

Zur Bezeichnung dafür, daß a Limes der Folge $\{a_n\}$ ist, sind die folgenden beiden Schreibweisen gebräuchlich:

1. $\lim a_n = a$ (ältere Schreibweise),
2. $a_n \to a$ (neuere Schreibweise).

Beide Schreibweisen sind gegenwärtig gleichermaßen üblich.

Statt „a ist Grenzwert der Folge $\{a_n\}$", sagt man häufig auch, „a_n strebt gegen a". Will man nur zum Ausdruck bringen, daß eine Folge einen Limes besitzt, ohne anzugeben, welche Zahl dies ist, so sagt man, die Folge sei *konvergent*. Manchmal sagt man bei Angabe des Limes auch, die Folge *konvergiere gegen a* oder, noch ausführlicher, sie *konvergiere gegen den Grenzwert a*.

Besitzt eine Folge keinen Grenzwert, so nennt man sie *divergent*.

Wir wollen zunächst die Beispiele 1 bis 5 des vorangehenden Paragraphen vom Standpunkt der hier angegebenen Definition des Grenzwertes noch einmal betrachten. In allen diesen Beispielen ist der Grenzwert a gleich Null. Daher nimmt in diesen Fällen die Ungleichung (38) die einfachere Form

$$|a_n| < \varepsilon \tag{39}$$

an.

In Beispiel 1 war $a_n = \frac{1}{n}$, so daß die Ungleichung (39) für $n > \frac{1}{\varepsilon}$ erfüllt ist, d. h., man muß als N eine beliebige Zahl nehmen, die nicht kleiner als $\frac{1}{\varepsilon}$ ist.

Bei der ersten Folge aus Beispiel 2 ist (für $n \geqq 2$) $a_n = \frac{1}{n-1}$, und die Ungleichung (39) ist erfüllt, wenn $n > \frac{1}{\varepsilon} + 1$, d. h., man kann als N jede Zahl nehmen, die wenigstens um Eins größer als $\frac{1}{\varepsilon}$ ist.

Bei der zweiten Folge aus Beispiel 2 ist $a_{2m} = \frac{1}{m}$, $a_{2m-1} = 0$. Die Ungleichung (39) ist für alle die Folgeglieder, deren Index ungerade ist, von selbst erfüllt; für gerades n ist sie erfüllt, wenn $\frac{n}{2} = m > \frac{1}{\varepsilon}$, d. h. für $n > \frac{2}{\varepsilon}$. Man kann also als N jede Zahl nehmen, die nicht kleiner als $\frac{2}{\varepsilon}$ ist.

---

[1]) Zur übersichtlicheren Darstellung des Folgenden ist der Buchstabe c des vorangehenden Paragraphen hier durch den Buchstaben a ersetzt.

In Beispiel 3 ist $a_n = \frac{1}{n^\alpha}$, und als N kann jede Zahl genommen werden, die nicht kleiner als $\varepsilon^{-\frac{1}{\alpha}}$ ist.

In Beispiel 4 liegen die Dinge genauso wie in Beispiel 1, denn es ist $\left|\frac{(-1)^{n+1}}{n}\right| = \frac{1}{n}$.

In Beispiel 5 ist schließlich $a_n = aq^{n-1}$, so daß $|a| \cdot |q|^{n-1} < \varepsilon$, sobald

$$n > \frac{\lg|a| + \lg\frac{1}{\varepsilon}}{\lg\frac{1}{|q|}} + 1.$$

In diesem Fall muß also N so gewählt werden, daß es nicht kleiner ist als die auf der rechten Seite dieser Ungleichung stehende Zahl.

Die angegebenen Beispiele zeigen, in welchem Sinne die bei der Definition des Limes einer Folge erwähnte Abhängigkeit der Zahl N von der Zahl ε zu verstehen ist. Es muß allerdings gesagt werden, daß man in der Praxis nur selten auf die „Epsilontik" zurückgreifen muß, wenn es gilt, den Grenzwert einer vorgelegten Folge zu bestimmen. Häufig gelingt es nämlich mittels der unten genannten Sätze, komplizierte Fälle auf einfachere zurückzuführen. Nur selten erweist sich eine Rückführung dieser Art als unmöglich, so daß man auf die „Epsilontik" als Beweismittel zurückgreifen muß.

Als erstes beweisen wir einen Satz, den wir im folgenden häufiger werden anzuwenden haben:

Aus jeder beschränkten Folge kann man eine konvergente Teilfolge „aussondern".

Dies bedeutet genauer folgendes: Ist die Folge $\{a_n\}$ beschränkt, so kann man eine wachsende Folge $\{p_n\}$ von positiven ganzen Zahlen so finden, daß die Folge $\{a_{p_n}\}$ einen Grenzwert besitzt.

Beweis. Nach dem Satz von Bolzano und Weierstrass (§ 36) besitzt die Folge $\{a_n\}$ wenigstens einen Häufungspunkt a. Aus der Definition des Häufungspunktes ergibt sich, daß in jeder ε-Umgebung von a unendlich viele Glieder der Folge $\{a_n\}$ liegen. Es sei nun $\{\varepsilon_n\}$ eine fallende Folge von positiven Zahlen, die gegen Null konvergiert. Zu jedem Glied ε_n $(n = 1, 2, \ldots)$ der Folge können wir dann ein Glied a_{p_n} der Folge $\{a_n\}$ so wählen, daß a_{p_n} in der ε_n-Umgebung von a liegt, daß also

$$|a_{p_n} - a| < \varepsilon_n \qquad (n = 1, 2, 3, \ldots) \tag{40}$$

gilt. Da es in der ε_n-Umgebung von a sogar unendlich viele Glieder der Folge gibt, kann man noch dafür sorgen, daß jeder folgende Index p_{n+1} größer als der vorangehende p_n ist. Aus den Relationen (40) ergibt sich sofort

$$a_{p_n} \to a.$$

Bemerkung. Wir werden uns im vorliegenden Paragraphen nur mit beschränkten Folgen befassen, weisen aber bereits hier darauf hin, daß eine unbeschränkte Folge keinen (endlichen) Grenzwert besitzen kann, d. h., daß *jede konvergente Folge* $\{a_n\}$ *beschränkt ist.*

In der Tat: Konvergiert die Folge $\{a_n\}$ gegen a, so gibt es zu jedem $\varepsilon > 0$ ein N, so daß vom Glied mit dem Index $N+1$ an alle Glieder der Folge $\{a_n\}$ der Ungleichung

$$a_n < a + \varepsilon$$

genügen. Außerdem kann man aber unter den verbleibenden endlich vielen Gliedern $a_1, a_2, \ldots, a_N$ ein größtes auswählen. Ist dann M eine Zahl, die sowohl größer als dieses maximale Glied als auch größer als $a + \varepsilon$ ist, so gilt für jedes n die Ungleichung

$$a_n < M,$$

d. h., die Folge $\{a_n\}$ ist nach oben beschränkt. Ähnlich sieht man, daß die Folge $\{a_n\}$ auch nach unten beschränkt ist.

Den Übergang von einer Folge zu ihrem Grenzwert (den sogenannten Grenzübergang) kann man als eine gewisse *Operation* ansehen. Sie hat manches mit den arithmetischen Operationen gemeinsam, unterscheidet sich aber auch in vieler Hinsicht merklich von diesen. Wie bei den arithmetischen Operationen wird hier zu gegebenen Zahlen

$$a_1, a_2, \ldots, a_n, \ldots$$

eine neue Zahl $a = \lim a_n$ gebildet; wie bei den arithmetischen Operationen ist a durch die gegebenen Zahlen eindeutig bestimmt. Jedoch ist der Übergang zum Limes bei weitem nicht immer „möglich" (eine gegebene Folge von Zahlen braucht keinem Limes zu besitzen). Ein weiterer wesentlicher Unterschied zu den arithmetischen Operationen besteht darin, daß man hier nicht von zwei (oder endlich vielen) „gegebenen" Zahlen ausgeht, sondern von unendlich vielen. Dabei ist es überaus bedeutsam, daß *sich das Konvergenzverhalten einer Folge und ihr Grenzwert* (wenn vorhanden) *nicht ändern, wenn man endlich viele Glieder der Folge fortläßt, hinzufügt oder durch beliebige andere Zahlen ersetzt.*[1]) Ist nämlich a der einzige Häufungspunkt einer Folge $\{a_n\}$, so ist offenbar a auch der einzige Häufungspunkt jeder Folge, die man auf eine der genannten Arten aus $\{a_n\}$ erhält.

Sehr wichtig ist auch, daß man bei einer konvergenten Folge, ohne den Grenzwert zu ändern, sogar unendlich viele Glieder fortlassen kann, wenn nur unendlich viele übrigbleiben. Es gilt nämlich der folgende Satz:

Es sei

$$\{p_n\} \equiv p_1, p_2, \ldots, p_n, \ldots$$

eine beliebige wachsende Folge von positiven ganzen Zahlen. Besitzt eine gegebene Folge $\{a_n\}$ *den Grenzwert* a,

$$a_n \to a,$$

so hat auch die „Teilfolge" $\{a_{p_n}\}$ *einen Limes, und dieser ist ebenfalls gleich* a*:*

$$a_{p_n} \to a.$$

Beweis. Wenn die Folge $\{a_n\}$ den Limes a besitzt (und daher beschränkt ist), so ist a der einzige Häufungspunkt der Folge $\{a_n\}$. Weil nun die Glieder der Folge $\{a_{p_n}\}$ unter denen der Folge $\{a_n\}$ vorkommen, ist die Folge $\{a_{p_n}\}$ ebenfalls beschränkt und besitzt

[1]) D. h., eine konvergente Folge bleibt konvergent, eine divergente Folge bleibt divergent.

daher wenigstens einen Häufungspunkt (vgl. § 36). Jeder Häufungspunkt der Folge $\{a_{p_n}\}$ ist aber offenbar erst recht Häufungspunkt der Folge $\{a_n\}$. Daher kann die Folge $\{a_{p_n}\}$ höchstens die Zahl a als Häufungspunkt besitzen. Also ist a der einzige Häufungspunkt der Folge $\{a_{p_n}\}$, d. h. ihr Grenzwert.

Die folgenden Sätze stellen einen Zusammenhang her zwischen der Operation des Grenzüberganges und den arithmetischen Operationen. Sie spielen darüber hinaus eine wesentliche Rolle bei der Berechnung der Grenzwerte von gegebenen Zahlenfolgen.[1])

Satz I. *Wenn die Folgen* $\{a_n\}$ *und* $\{b_n\}$ *einen Grenzwert besitzen, also etwa*

$$a_n \to a, \quad b_n \to b,$$

so besitzt auch die „Summenfolge" $\{a_n + b_n\}$ *einen Grenzwert, und zwar gilt:*

$$a_n + b_n \to a + b.$$

Satz II. *Unter den gleichen Voraussetzungen besitzt auch die „Differenzenfolge"* $\{a_n - b_n\}$ *einen Grenzwert, und zwar gilt:*

$$a_n - b_n \to a - b.$$

Satz III. *Unter eben diesen Voraussetzungen besitzt auch die „Produktfolge"* $\{a_n b_n\}$ *einen Grenzwert, und zwar gilt:*

$$a_n b_n \to ab.$$

Satz IV. *Unter den genannten Voraussetzungen und den zusätzlichen Voraussetzungen* $b_n \neq 0$ *für* $n = 1, 2, 3, \ldots$ *und* $b \neq 0$ *besitzt auch die „Quotientenfolge"* $\left\{\frac{a_n}{b_n}\right\}$ *einen Grenzwert, und zwar gilt:*

$$\frac{a_n}{b_n} \to \frac{a}{b}.$$

Häufig spricht man den in Satz I wiedergegebenen Sachverhalt kurz folgendermaßen aus: „Der Grenzwert einer Summe ist gleich der Summe der Grenzwerte", und schreibt dafür

$$\lim (a_n + b_n) = \lim a_n + \lim b_n,$$

wobei man sich die Voraussetzungen des Satzes hinzudenkt. Wenn die Folgen $\{a_n\}$ und $\{b_n\}$ keinen Grenzwert besitzen, so kann man im allgemeinen nichts über den Grenzwert der Folge $\{a_n + b_n\}$ aussagen. Zu bemerken ist, daß die Folge $\{a_n + b_n\}$ durchaus einen Grenzwert besitzen kann, ohne daß dies für $\{a_n\}$ und für $\{b_n\}$ der Fall zu sein braucht.[2])

Entsprechendes gilt für die Sätze II bis IV.

---

[1]) Beweise für diese Sätze sind bereits in EdEM, Band 1, S. 168/69 angegeben. Wir beweisen sie im folgenden nochmals mit Hilfe eines anderen Verfahrens.

[2]) Beispielsweise sind die Folgen

$$1, -1, 1, -1, 1, -1, \ldots$$

und

$$-1, 1, -1, 1, -1, 1, \ldots$$

beide divergent, während ihre „Summenfolge"

$$0, 0, 0, 0, 0, 0, \ldots$$

konvergent ist.

Wir beweisen zunächst Satz I. Vorausgesetzt ist dort, daß der Abstand $|a_n - a|$ für hinreichend großes n beliebig klein wird und daß dasselbe für $|b_n - b|$ gilt; zu zeigen ist, daß auch der Abstand $|(a_n + b_n) - (a + b)|$ für hinreichend großes n beliebig klein wird.

Auf Grund einer bekannten Eigenschaft des absoluten Betrages[1]) ist

$$|(a_n + b_n) - (a + b)| = |(a_n - a) + (b_n - b)| \leqq |a_n - a| + |b_n - b|.$$

Wenn uns nun eine beliebige Zahl $\varepsilon > 0$ vorgegeben ist, so können wir einerseits eine Zahl N'_ε so finden, daß für $n > N'_\varepsilon$

$$|a_n - a| < \frac{\varepsilon}{2},$$

und ebenso können wir eine Zahl N''_ε so finden, daß für $n > N''_\varepsilon$

$$|b_n - b| < \frac{\varepsilon}{2}.$$

Bezeichnen wir mit N_ε das Maximum der Zahlen N'_ε und N''_ε, so gilt für alle $n > N_\varepsilon$

$$|(a_n + b_n) - (a + b)| \leqq |a_n - a| + |b_n - b| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

womit Satz I bewiesen ist.

Als nächstes beweisen wir Satz III. Hier ist unter den Voraussetzungen von Satz I zu zeigen, daß der Abstand $|a_n b_n - ab|$ für hinreichend großes n kleiner als ein beliebig vorgegebenes $\varepsilon > 0$ wird. Offenbar ist

$$a_n b_n - ab = (a_n - a)\,b + a(b_n - b) + (a_n - a)(b_n - b),$$

so daß

$$|a_n b_n - ab| \leqq |a_n - a| \cdot |b| + |a| \cdot |b_n - b| + |a_n - a| \cdot |b_n - b|. \tag{41}$$

Es sei nun K das Maximum der Zahlen $|a|$, $|b|$ und 1 (letzteres ist für den Fall $a = b = 0$ erforderlich); ferner wollen wir für die vorgegebene Zahl ε annehmen, daß

$$\varepsilon < 3K^2$$

gilt. Wir denken uns nun die Zahlen N'_ε und N''_ε so gewählt, daß

$$|a_n - a| < \frac{\varepsilon}{3K} \quad \text{für} \quad n > N'_\varepsilon$$

und

$$|b_n - b| < \frac{\varepsilon}{3K} \quad \text{für} \quad n > N''_\varepsilon.$$

---

[1]) Gemeint ist die sogenannte Dreiecksungleichung

$$|x + y| \leqq |x| + |y|.$$

Ferner benutzen wir im folgenden die Beziehung

$$|xy| = |x| \cdot |y|.$$

Siehe EdEM, Band 1, Seite 108.

Dann ist für $n > N_\varepsilon$, wobei $N_\varepsilon = \max(N_\varepsilon', N_\varepsilon'')$,

$$|a_n - a| \cdot |b| + |a| \cdot |b_n - b| + |a_n - a| \cdot |b_n - b|$$
$$\leqq \frac{\varepsilon}{3K} \cdot K + K \cdot \frac{\varepsilon}{3K} + \frac{\varepsilon}{3K} \cdot \frac{\varepsilon}{3K} = \varepsilon \left(\frac{2}{3} + \frac{\varepsilon}{9K^2} \right) < \varepsilon \left(\frac{2}{3} + \frac{3K^2}{9K^2} \right) = \varepsilon. \quad (42)$$

Aus den Ungleichungen (41) und (42) ergibt sich die Behauptung von Satz III.

Als Spezialfall von Satz III (für den Fall, daß alle Glieder der Folge $\{a_n\}$ gleich a sind) ergibt sich: *Wenn die Folge $\{b_n\}$ gegen b konvergiert, so konvergiert die Folge $\{ab_n\}$ gegen ab* („ein konstanter Faktor kann vor das Limeszeichen gezogen werden"). Insbesondere ergibt sich für $a = -1$: *Konvergiert die Folge $\{b_n\}$ gegen b, so konvergiert die Folge $\{-b_n\}$ gegen $-b$.*

Hieraus folgt mit Hilfe von Satz I sofort Satz II:

$$a_n - b_n = a_n + (-b_n) \to a + (-b) = a - b.$$

Wir wenden uns schließlich dem Beweis von Satz IV zu, und zwar zunächst für den Spezialfall, daß alle Glieder der Folge $\{a_n\}$ gleich 1 sind, d. h., *konvergiert die Folge $\{b_n\}$ gegen die Zahl $b \neq 0$, so konvergiert die Folge $\left\{\frac{1}{b_n}\right\}$ gegen $\frac{1}{b}$.*

Hierzu haben wir den Abstand des Punktes $\frac{1}{b_n}$ vom Punkt $\frac{1}{b}$ abzuschätzen:

$$\left| \frac{1}{b_n} - \frac{1}{b} \right| = \frac{|b_n - b|}{|b_n| \cdot |b|}. \quad (43)$$

Der Abstand $|b_n - b|$ wird nach Voraussetzung für hinreichend großes n kleiner als jede positive Zahl, insbesondere also kleiner als $\frac{|b|}{2}$:

$$|b_n - b| < \frac{|b|}{2} \qquad (\text{für} \quad n > N').$$

Es ist jedoch

$$|b| = |(b - b_n) + b_n| \leqq |b - b_n| + |b_n|,$$

so daß

$$|b - b_n| \geqq |b| - |b_n|.$$

Daher ist

$$|b| - |b_n| < \frac{|b|}{2},$$

d. h.

$$|b_n| > \frac{1}{2} |b|. \quad (44)$$

Wählen wir nun die Zahl N_ε'' so groß, daß für $n > N_\varepsilon''$ die Ungleichung

$$|b_n - b| < \frac{1}{2} |b|^2 \varepsilon \quad (45)$$

erfüllt ist, so ist wegen (43) und (44) für $n > N_\varepsilon$, wobei $N_\varepsilon = \max(N', N_\varepsilon'')$,

$$\left| \frac{1}{b_n} - \frac{1}{b} \right| < \frac{\frac{1}{2} |b|^2 \cdot \varepsilon}{|b| \cdot \frac{1}{2} |b|} = \varepsilon,$$

womit der betrachtete Spezialfall von Satz IV bewiesen ist.

Aus diesem bewiesenen Spezialfall von Satz IV und aus Satz III ergibt sich unmittelbar Satz IV selbst. Wenn nämlich $a_n \to a$, $b_n \to b$ gilt, so gilt auch

$$\frac{a_n}{b_n} = a_n \cdot \frac{1}{b_n} \to a \cdot \frac{1}{b} = \frac{a}{b}.$$

Schließlich beweisen wir noch den folgenden Satz über den Grenzübergang in einer Ungleichung:

Wenn die Folgen $\{a_n\}$ *und* $\{b_n\}$ *einen Grenzwert besitzen, und zwar*

$$a_n \to a, \; b_n \to b,$$

und wenn außerdem für alle n *die Ungleichung*

$$a_n \leqq b_n \tag{46}$$

erfüllt ist, so gilt auch die Ungleichung

$$a \leqq b. \tag{47}$$

Wir betrachten zunächst den Spezialfall, daß alle Glieder der Folge $\{a_n\}$ und damit auch a gleich 0 sind. Hier ist zu zeigen, daß *aus* $b_n \geqq 0$ $(n = 1, 2, 3, \ldots)$ *und* $b_n \to b$ *die Ungleichung* $b \geqq 0$ *folgt*. Dies beweisen wir indirekt, d. h., wir nehmen an, es wäre $b < 0$. Setzen wir $\varepsilon = |b|$, so finden wir, daß für hinreichend großes n die Ungleichungen

$$-|b| < |b_n - b| < |b|, \tag{48}$$

d. h.

$$b - |b| < b_n < b + |b| \tag{49}$$

erfüllt sein müssen. Wäre $b < 0$, so wäre $b + |b| = 0$, und (49) nähme die Form $b_n < 0$ an, im Widerspruch zur Voraussetzung $b_n \geqq 0$ $(n = 1, 2, 3, \ldots)$.

Wir wenden uns nun dem allgemeinen Fall zu. Nach Voraussetzung ist $b_n \geqq a_n$ $(n = 1, 2, 3, \ldots)$, also $b_n - a_n \geqq 0$ $(n = 1, 2, 3, \ldots)$. Ferner gilt auf Grund von Satz II die Beziehung $b_n - a_n \to b - a$. Hieraus ergibt sich durch Anwendung des bereits bewiesenen Spezialfalles: $b - a \geqq 0$, also $b \geqq a$.

Ersetzt man die Voraussetzung (46) durch die stärkere Voraussetzung

$$a_n < b_n,$$

so kann man ebenfalls nur auf $a \leqq b$ (und nicht etwa auf $a < b$) schließen.

§ 39. Verallgemeinerung des Limesbegriffs („uneigentliche Grenzwerte")

Wir wollen uns nun mit einigen Redeweisen vertraut machen, die in der mathematischen Praxis weitverbreitet sind.

Unter einer „Umgebung des Punktes $+\infty$", wollen wir jedes Intervall verstehen, welches nach oben nicht beschränkt ist (also keinen Endpunkt besitzt). Insbesondere nennen wir die Menge aller Zahlen

$$x > M \quad (\text{oder } M < x < +\infty) \tag{50}$$

die „M-Umgebung des Punktes $+\infty$".

Man sagt, eine Folge besitze den „Häufungspunkt $+\infty$", wenn in jeder „Umgebung des Punktes $+\infty$" unendlich viele Glieder der Folge enthalten sind. Dies bedeutet offenbar gerade, daß die Folge nach oben unbeschränkt ist.

Entsprechend wird jedes Intervall, welches nach unten nicht beschränkt ist (keinen Anfangspunkt besitzt), eine „Umgebung des Punktes $-\infty$" genannt. Insbesondere bezeichnen wir das Intervall

$$x < -M \quad (\text{oder } -\infty < x < -M) \tag{51}$$

als die „$(-M)$-Umgebung des Punktes $-\infty$". Die Aussage, eine Folge besitzt „den Häufungspunkt $-\infty$", ist gleichwertig der Feststellung, daß die betrachtete Folge nach unten unbeschränkt ist.

Den Satz von Bolzano und Weierstrass können wir dann kurz folgendermaßen aussprechen:

Jede unendliche Folge $\{a_n\}$ besitzt wenigstens einen Häufungspunkt (im „eigentlichen" oder im „uneigentlichen" Sinne).

In der Tat: Ist die Folge nach oben unbeschränkt, so hat sie den Häufungspunkt $+\infty$; ist sie nach unten beschränkt, so hat sie den Häufungspunkt $-\infty$. Ist sie sowohl nach oben als auch nach unten beschränkt, so hat sie nach dem Hauptsatz aus § 36 wenigstens einen „eigentlichen" („endlichen") Häufungspunkt.

Da wir nach wie vor einen Punkt, der einziger Häufungspunkt einer Folge ist, als ihren Limes bezeichnen, hat mit der Erweiterung des Begriffs „Häufungspunkt" auch der Begriff „Limes" eine Verallgemeinerung erfahren, und zwar dadurch, daß jetzt auch die Punkte $+\infty$ und $-\infty$ als („uneigentliche") Grenzwerte von Folgen zugelassen werden. Wir wollen zunächst die Tatsache, daß $+\infty$ einziger Häufungspunkt einer gegebenen Folge ist, mit anderen Worten ausdrücken. Da in diesem Fall $-\infty$ kein Häufungspunkt ist, muß die Folge nach unten beschränkt sein, so daß in einem bestimmten Intervall $(-\infty, m)$ kein Punkt der Folge liegt; weil kein „endlicher" Punkt Häufungspunkt der Folge ist, sind in jedem endlichen Intervall $[m, M]$ mit $M > m$ nur endlich viele Glieder der Folge enthalten. Es sei nun $N = N_M$ der größte Index von diesen Folgegliedern. Dann liegen offenbar alle Folgeglieder a_n, deren Index größer als N ist, in der M-Umgebung des Punktes $+\infty$, d. h., sie genügen der Ungleichung $a_n > M$. Selbstverständlich gilt hiervon auch die Umkehrung: Wenn für jedes noch so große M alle Glieder $\{a_n\}$ mit hinreichend großem n in der M-Umgebung des Punktes $+\infty$ liegen, so besitzt die Folge $\{a_n\}$ den Punkt $+\infty$ als einzigen Häufungspunkt, d. h. als Grenzwert. Es gilt also: *Eine Folge $\{a_n\}$ besitzt dann und nur dann den Grenzwert $+\infty$, wenn man zu jeder noch so großen Zahl M eine Zahl $N = N_M$ finden kann derart, daß für alle $n > N$ die Ungleichung $a_n > M$ gilt.* Man schreibt dafür

1. $\lim a_n = \infty$,
2. $a_n \to \infty$.

Statt „$\{a_n\}$ hat den Grenzwert $+\infty$" sagt man auch „$\{a_n\}$ strebt gegen $+\infty$" oder auch „$\{a_n\}$ divergiert gegen $+\infty$" oder auch „$\{a_n\}$ divergiert bestimmt gegen $+\infty$" („bestimmte" Divergenz im Gegensatz zu „unbestimmter" Divergenz, bei der mehrere Häufungspunkte vorhanden sind).

Entsprechend ergibt sich, daß *eine Folge* $\{a_n\}$ *dann und nur dann den Grenzwert* $-\infty$ *besitzt, wenn man zu jeder noch so großen Zahl* M *eine Zahl* $N = N_M$ *so finden kann, daß für alle* $n > N$ *die Ungleichung* $a_n < -M$ *besteht.* Man schreibt dafür: $\lim a_n = -\infty$ oder $a_n \to -\infty$.

Einige Beispiele mögen das Vorangehende erläutern:

1. Die Folge der natürlichen Zahlen $\{n\}$ besitzt $+\infty$ als einzigen Häufungspunkt und damit als Grenzwert. Dasselbe gilt für die Folgen $\{n^\alpha\}$ mit $\alpha > 0$.

2. Die Folge

$$\{(-1)^{n+1} n\} \equiv 1, -2, 3, -4, 5, -6, \ldots$$

besitzt die beiden Häufungspunkte $+\infty$ und $-\infty$ und infolgedessen nicht einmal einen Grenzwert im „uneigentlichen" Sinne.

3. Besitzt $\{a_n\}$ keinen endlichen Häufungspunkt, so ist $\lim |a_n| = +\infty$.

4. Die arithmetische Folge $\{a + (n-1)d\}$ besitzt den Grenzwert $+\infty$ oder $-\infty$, je nachdem, ob $d > 0$ oder $d < 0$.

5. Die geometrische Folge $\{q^n\}$ besitzt für $q > 1$ den Grenzwert $+\infty$ (siehe Seite 72 ff.), während sie für $q < -1$ keinen Grenzwert besitzt.

6. $\lim\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right) = \infty$ (siehe § 35).

7. Die Folge 2. von Seite 141 hat alle natürlichen Zahlen und $+\infty$ als Häufungspunkte. Einen Grenzwert besitzt sie nicht.

8. Die Folge $\{\operatorname{tg} n\pi\theta\}$ hat für irrationales θ alle reellen Zahlen, $+\infty$ und $-\infty$ als Häufungspunkte. Gäbe es nämlich in einem Intervall (α, β) keinen Punkt der Folge $\{\operatorname{tg} n\pi\theta\}$, so läge im Intervall $\left(\frac{\alpha}{\sqrt{1+\alpha^2}}, \frac{\beta}{\sqrt{1+\beta^2}}\right)$ kein Glied der Folge $\{\sin n\pi\theta\}$ im Widerspruch zu Beispiel 16 aus § 37.

Die Sätze I bis IV aus § 38 sind bei unendlichen Limites nur mit Vorsicht anzuwenden, da sie hier nur teilweise richtig sind. Es gelten die folgenden Sätze (der Beweis bleibe dem Leser überlassen):

I'. *Strebt die Folge* $\{a_n\}$ *gegen* $+\infty$ *und ist die Folge* $\{b_n\}$ *nach unten beschränkt, so strebt die Folge* $\{a_n + b_n\}$ *gegen* $+\infty$.

I''. *Strebt die Folge* $\{a_n\}$ *gegen* $-\infty$ *und ist die Folge* $\{b_n\}$ *nach oben beschränkt, so strebt die Folge* $\{a_n + b_n\}$ *gegen* $-\infty$.

Die Sätze vom Typ II kann man durch entsprechende Wahl des Vorzeichens aus den Sätzen vom Typ I erhalten.

Dagegen ist es unmöglich, z. B. auf folgende Fragen eine allgemeingültige Antwort zu geben:

1. Wie verhält sich die Folge $\{a_n - b_n\}$ für $a_n \to \infty$, $b_n \to \infty$?
2. Wie verhält sich die Folge $\{a_n b_n\}$ für $a_n \to \infty$, $b_n \to 0$?
3. Wie verhält sich die Folge $\left\{\frac{a_n}{b_n}\right\}$ für $a_n \to \infty$, $b_n \to \infty$?
4. Wie verhält sich die Folge $\left\{\frac{a_n}{b_n}\right\}$ für $a_n \to 0$, $b_n \to 0$?

Diese und ähnliche Fragen können nur dann beantwortet werden, wenn die Folgen $\{a_n\}$ und $\{b_n\}$ konkret vorgegeben sind. Sehr häufig schafft dann eine einfache identische Umformung eine Situation, in der bereits Sätze vom Typ I bis IV anwendbar werden.[1]) Mitunter muß man jedoch auch besondere Untersuchungen anstellen (Beispiele siehe unten, § 43 bis 45).

§ 40. Der Grenzwert einer Funktion im Unendlichen

Der Begriff des Grenzwertes einer Funktion wird auf den Begriff des Grenzwertes einer Folge zurückgeführt. Als erstes behandeln wir den Begriff des Grenzwertes einer Funktion „im Unendlichen" $(+\infty)$.

Vorgegeben sei eine Funktion $f(x)$ der reellen Veränderlichen x, die in einer gewissen Umgebung des Punktes $+\infty$, etwa im Intervall (a, ∞), d. h. für alle x mit $a < x < \infty$ erklärt ist (dort nirgends ihren Sinn verliert). Im vorangehenden hatten wir es bereits mit Folgen zu tun, deren allgemeines Glied a_n die Form $f(x_n)$ mit $x_n = n$ hatte:

$$\{f(n)\} \equiv f(1),\ f(2),\ \ldots,\ f(n),\ \ldots\ .$$

Im Gegensatz dazu sei jetzt $\{x_n\}$ eine beliebige Folge von Zahlen aus dem Intervall (a, ∞), die den Grenzwert $+\infty$ besitzt:

$$x_n \to +\infty. \tag{52}$$

Es kann nun der Fall eintreten, daß auch die Folge der zugehörigen Funktionswerte

$$\{f(x_n)\} \equiv f(x_1),\ f(x_2),\ \ldots,\ f(x_n),\ \ldots$$

gegen einen gewissen endlichen oder unendlichen[2]) Grenzwert L strebt,

$$f(x_n) \to L, \tag{53}$$

und zwar unabhängig davon, wie die Folge $\{x_n\}$ mit der Eigenschaft (52) ausgewählt ist. In diesem Fall sagt man, *die Funktion $f(x)$ strebe für unbeschränkt wachsendes x gegen den Grenzwert (Limes) L.* Man bringt dies durch folgende Schreibweisen zum Ausdruck:

$$\left.\begin{array}{ll} 1. & \lim\limits_{x\to\infty} f(x) = L, \\ 2. & f(x) \to L \quad \text{für} \quad x \to \infty. \end{array}\right\} \tag{54}$$

Wenn dagegen zwei verschiedene Folgen der Form $\{f(x_n)\}$ verschiedene Grenzwerte besitzen oder wenn für wenigstens eine Folge $\{x_n\}$ die zugehörige

---

[1]) Die verbreitete Ausdrucksweise „Auswertung unbestimmter Ausdrücke" (die Fälle: $\infty - \infty$, $\infty \cdot 0$, $\frac{\infty}{\infty}$, $\frac{0}{0}$) ist irreführend, da die „Unbestimmtheit" doch lediglich darin besteht, daß eine unmittelbare Anwendung z. B. der Sätze I–IV nicht möglich ist.

[2]) In beiden Fällen wollen wir den Grenzwert mit L bezeichnen.

Folge $\{f(x_n)\}$ keinen Grenzwert besitzt[1]), so sagt man, *die Funktion $f(x)$ strebe für unbeschränkt wachsendes x nicht gegen einen Grenzwert* bzw. *der Grenzwert* $\lim\limits_{x\to\infty} f(x)$ *existiere nicht.*

Ähnlich definiert man den Begriff des Grenzwertes von $f(x)$ für $x \to -\infty$.

Beispiele.

1. Die Funktion $f(x) = \frac{1}{x}$ besitzt für $x \to \infty$ den Grenzwert 0:

$$\lim_{x\to\infty} \frac{1}{x} = 0.$$

Wenn nämlich $x_n \to \infty$, so $\frac{1}{x_n} \to 0$ (IV‴). Ebenso ergibt sich

$$\lim_{x\to-\infty} \frac{1}{x} = 0.$$

2. Für die Funktion $f(x) = x^n$ (n natürliche Zahl) gilt

$$\lim_{x\to\infty} x^n = +\infty, \quad \lim_{x\to-\infty} x^n = \begin{cases} +\infty, & \text{falls } n \text{ gerade ist,} \\ -\infty, & \text{falls } n \text{ ungerade ist.} \end{cases}$$

3. Für die Funktion $f(x) = a^x$ ($a > 1$) gilt

$$\lim_{x\to\infty} a^x = \infty, \qquad \lim_{x\to-\infty} a^x = 0.$$

Für ganzzahlige Werte von x ist dies auf Seite 73 gezeigt; für alle anderen Werte folgt es daraus, daß die Funktion a^x wachsend ist.

4. $\lim\limits_{x\to\infty} \log_a x = \infty$ (für $a > 1$).

5. Die Funktion $f(x) = \sin x$ (Abb. 70a) besitzt für unbeschränkt wachsendes x keinen Grenzwert. Setzen wir nämlich $x_n = \pi n$, so erhalten wir $f(\pi n) = \sin \pi n \to 0$; setzen wir dagegen $x_n = \frac{\pi}{2} + 2\pi n$, so erhalten wir $f\left(\frac{\pi}{2} + 2\pi n\right) = \sin\left(\frac{\pi}{2} + 2\pi n\right) \to 1$; setzen wir $x_n = \frac{3}{2}\pi + 2\pi n$, so erhalten wir $f\left(\frac{3}{2}\pi + 2\pi n\right) = \sin\left(\frac{3}{2}\pi + 2\pi n\right) \to -1$, usw.

Ebenso ergibt sich, daß $\lim\limits_{x\to-\infty} \sin x$ nicht existiert.

6. Die Funktion $f(x) = x \sin x$ (Abb. 70b) besitzt ebenfalls weder für $x \to \infty$ noch für $x \to -\infty$ einen Grenzwert. Auch dieses Mal gilt $f(\pi n) \to 0$, während $f\left(\frac{\pi}{2} + 2\pi n\right) \to +\infty$, $f\left(\frac{3}{2}\pi + 2\pi n\right) \to -\infty$.

---

[1]) Der Fall, daß alle möglichen Folgen der Form $\{f(x_n)\}$ einen Grenzwert besitzen, dabei aber nicht immer denselben, kann nicht eintreten. Besäße nämlich die Folge $\{f(x'_n)\}$ den Grenzwert L' und die Folge $\{f(x''_n)\}$ den davon verschiedenen Grenzwert L'', so würde die Folge

$$f(x'_1), f(x''_2), f(x'_3), f(x''_4), \ldots, f(x'_{2n-1}), f(x''_{2n}), \ldots$$

auf Grund von § 37, Beispiel 9, keinen Grenzwert besitzen.

7. Die Funktion $f(x)=\frac{\sin x}{x}$ (Abb. 70c) besitzt für $x\to\infty$ und für $x\to-\infty$ den Grenzwert 0:

$$\lim_{x\to\pm\infty}\frac{\sin x}{x}=0.$$

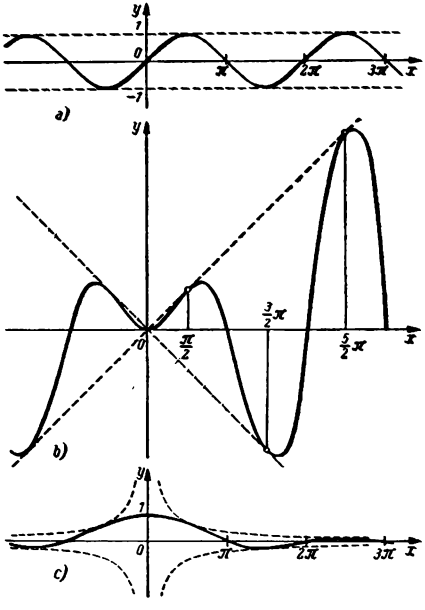


Abb. 70

Dies ergibt sich auf Grund der Ungleichungen

$$\left|\frac{\sin x}{x}\right|=\frac{|\sin x|}{|x|}\leq\frac{1}{|x|}$$

aus Beispiel 1.

8. Die Funktion $f(x)=2^x\sin 2\pi x$ besitzt für $x\to\infty$ keinen Grenzwert, während sie für $x\to-\infty$ gegen 0 strebt:

$$\lim_{x\to-\infty}2^x\sin 2\pi x=0.$$

Entsprechend besitzt die Funktion $f(x) = 2^{-x} \sin 2\pi x$ (Abb. 71) für $x \to \infty$ den Grenzwert 0:

$$\lim_{x \to +\infty} 2^{-x} \sin 2\pi x = 0,$$

während sie für $x \to -\infty$ keinen Grenzwert besitzt.

Der oben angegebenen Definition des Grenzwertes einer Funktion im Unendlichen sind für den Fall des Punktes $+\infty$ die folgenden Bedingungen gleichwertig.

1. Für endliches L: *Zu jedem noch so kleinen* $\varepsilon > 0$ *gibt es eine Zahl* $X \equiv X_\varepsilon$, *so daß für jedes* $x > X$ *die Ungleichung*

$$|f(x) - L| < \varepsilon \qquad (55)$$

erfüllt ist.

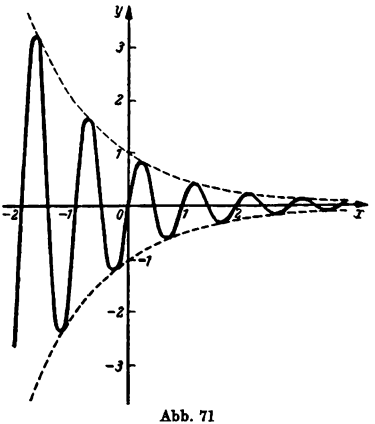


Abb. 71

2. Für $L = +\infty$: *Zu jeder noch so großen Zahl* M *kann man eine Zahl* $X \equiv X_M$ *so finden, daß für alle* $x > X$ *die Ungleichung*

$$f(x) > M \qquad (56)$$

erfüllt ist.

3. Für $L = -\infty$: *Zu jeder noch so großen Zahl* M *kann man eine Zahl* $X \equiv X_M$ *so finden, daß für alle* $x > X$ *die Ungleichung*

$$f(x) < -M \qquad (57)$$

erfüllt ist.

Wir beschränken uns hier darauf, die Gleichwertigkeit für den Fall eines endlichen Grenzwertes L zu beweisen.

Zunächst setzen wir voraus, daß aus $x_n \to \infty$ stets $f(x_n) \to L$ folgt, und zeigen, daß man dann zu jedem $\varepsilon > 0$ eine Zahl $X \equiv X_\varepsilon$ mit der unter 1. angegebenen Eigenschaft finden kann. Angenommen, dies wäre nicht der Fall, man könnte also nicht zu jedem ε eine solche Zahl X finden. Dann existiert eine gewisse Zahl $\varepsilon^* > 0$, zu der es keine Zahl X mit der geforderten Eigenschaft gibt, d. h., wie groß auch X sei, stets kann man einen Wert $x > X$ so finden, daß $|f(x) - L| \geqq \varepsilon^*$ ist. Es sei nun

$$\{X_n\} \equiv X_1, X_2, \ldots, X_n, \ldots$$

eine Folge von Zahlen mit $X_n \to \infty$. Zu jeder Zahl X_n können wir dann eine Zahl x_n so finden, daß

$$x_n > X_n \qquad (58)$$

und gleichzeitig

$$|f(x_n) - L| \geqq \varepsilon^* \tag{59}$$

gilt. Wir betrachten nun die Folge

$$\{x_n\} \equiv x_1, x_2, \ldots, x_n, \ldots .$$

Aus (58) folgt, daß $x_n \to \infty$. Andererseits folgt aus (59), daß $f(x_n) \to L$ nicht gelten kann, da ja wegen (59) der Abstand $|f(x_n) - L|$ größer als die betrachtete positive Zahl ε^* bleibt. Damit sind wir zu einem Widerspruch gelangt, so daß also unsere Annahme falsch war.

Umgekehrt setzen wir nun voraus, daß man zu jedem $\varepsilon > 0$ eine Zahl $X \equiv X_\varepsilon$ so finden kann, daß aus der Ungleichung $x > X$ stets die Ungleichung (55) folgt, und zeigen, daß aus $x_n \to \infty$ stets $f(x_n) \to L$ folgt. Wenn nun $x_n \to \infty$ gilt, so gibt es zu jedem M ein $N \equiv N_M$, so daß für alle $n > N$ die Ungleichung $x_n > M$ erfüllt ist; setzen wir $M = X$, so ist für alle $n > N$ dann $x_n > X$, woraus sich auf Grund unserer Voraussetzung $|f(x_n) - L| < \varepsilon$ ergibt.

§ 41. Einseitiger Grenzwert einer Funktion in einem endlichen Punkte

Wir wollen nun den Begriff des Grenzwertes einer Funktion $f(x)$ an einer endlichen Stelle $x = c$ behandeln, d. h., wir fragen nach dem Grenzwert, gegen den möglicherweise der Wert $f(x)$ strebt, wenn x dem Wert c beliebig nahe kommt. Wir beginnen mit einer Reihe von Beispielen, welche die Vielzahl der hierbei auftretenden Möglichkeiten zeigen sollen.

Beispiel 1. Wir wollen das Verhalten der Funktion

$$f(x) = 2^{\frac{1}{x}} \tag{60}$$

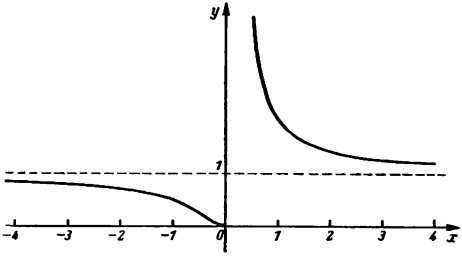


Abb. 72

untersuchen und dabei unsere Aufmerksamkeit besonders auf die Umgebung des Punktes $x = 0$ richten. Die Funktion $f(x)$ ist für alle Werte von x mit Ausnahme von $x = 0$ definiert.

Sie ist offenbar für alle $x \neq 0$ positiv. Ferner sieht man leicht, daß

$$\lim_{x \to \infty} f(x) = \lim_{x \to -\infty} f(x) = 1.$$

Fällt nun x von $+\infty$ gegen Null, so wächst $\frac{1}{x}$ von 0 über alle Grenzen und $2^{\frac{1}{x}}$ von 1 über alle Grenzen. Wächst andererseits x von $-\infty$ gegen Null, so fällt $\frac{1}{x}$ von 0 unter alle Grenzen und $2^{\frac{1}{x}}$ von 1 gegen 0 (wobei der Wert 0 für keinen Wert x angenommen wird). Das graphische Bild der Funktion (60) ist in Abb. 72 dargestellt.

Eine für uns hier bemerkenswerte Eigenschaft der Funktion (60) besteht darin, daß ihr graphisches Bild im Punkte $x = 0$ eine Unstetigkeit („Sprungstelle") besitzt; wenig rechts vom Punkt $x = 0$ sind die Funktionswerte überaus groß, während sie ein wenig links davon nahe bei Null liegen; im Punkt $x = 0$ selbst ist die Funktion nicht erklärt (d. h., auf der y-Achse liegt kein Punkt ihres graphischen Bildes).

Beispiel 2.

$$f(x) = \frac{1}{1 + 2^{\frac{1}{x}}}. \tag{61}$$

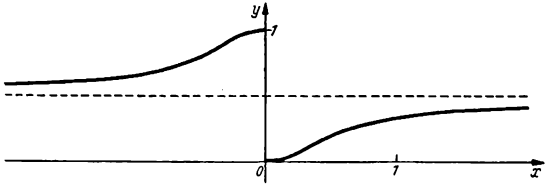


Abb. 73

Das Wachstumsverhalten dieser Funktion ermittelt man ähnlich wie das der Funktion (60). Man gelangt dabei zu einem graphischen Bild, wie es in Abb. 73 dargestellt ist. Auch hier liegt eine Unstetigkeit im Punkt $x = 0$ vor; während etwas links vom Punkte $x = 0$ die Funktionswerte $f(x)$ nahe bei 1 liegen, sind sie ein wenig rechts vom Punkte $x = 0$ nahe bei 0 gelegen („endlicher Sprung"); auf der y-Achse liegt kein Punkt des graphischen Bildes der Funktion (61), da die Funktion für $x = 0$ nicht definiert ist. Man sieht unmittelbar, daß

$$f(x) + f(-x) = 1,$$

so daß der Punkt mit den Koordinaten $\left(0, \frac{1}{2}\right)$ Symmetriezentrum des graphischen Bildes der Funktion (61) ist.

Beispiel 3. Die graphischen Bilder der Funktionen

$$\text{a) } f(x) = \frac{1}{x} \quad \text{und} \quad \text{b) } f(x) = \frac{1}{x^2}$$

besitzen gleichfalls Sprungstellen im Punkte $x = 0$. In beiden Fällen sind die Funktionswerte beiderseits der Stelle $x = 0$ dem Betrage nach beliebig groß, im ersten Fall mit verschiedenen Vorzeichen, im zweiten Fall mit gleichem Vorzeichen.

Beispiel 4.

$$f(x) = \sin \frac{1}{x}. \tag{62}$$

Die Funktionswerte schwanken bei dieser Funktion offenbar immer zwischen $+1$ und -1. Es ist klar, daß $\lim\limits_{x \to \infty} f(x) = 0$. Wenn nun x von $+\infty$ bis $\frac{2}{\pi}$ fällt, so wächst $\frac{1}{x}$ von 0 bis $\frac{\pi}{2}$ und $f(x)$ von 0 bis 1; fällt x von $\frac{2}{\pi}$ bis $\frac{2}{3\pi}$, so wächst $\frac{1}{x}$ von $\frac{\pi}{2}$ bis $\frac{3}{2}\pi$ und $f(x)$ fällt von 1 bis -1, wobei $f(x)$ für $x = \frac{1}{\pi}$ den Wert Null annimmt usw. Weil die Funktion $f(x)$ ungerade ist, hat ihr in Abb. 74 dargestelltes graphisches Bild den Ursprung O

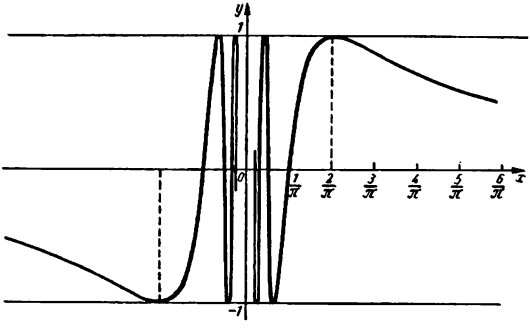


Abb. 74

als Symmetriezentrum. Je näher x an Null kommt, um so mehr „häufen" sich die „Schwankungen" der Funktionswerte. Im Punkte $x = 0$ ist die Funktion (62) nicht definiert.

Beispiel 5.

$$f(x) = \frac{1}{x} \sin \frac{1}{x}. \tag{63}$$

Diese Funktion charakterisiert eine „Schwingung", bei der sich bei Annäherung von x an den Punkt 0 die Schwankungen „häufen", wobei die Amplitude immer größer wird. Da für $x > 0$

$$|f(x)| \leqq \frac{1}{x},$$

d. h.

$$-\frac{1}{x} \leqq f(x) \leqq +\frac{1}{x}$$

gilt, wird das graphische Bild der Funktion (63) von den Zweigen der Hyperbeln $y = \frac{1}{x}$ und $y = -\frac{1}{x}$ „eingeschlossen" (Abb. 75). Die Berührungspunkte der graphischen

Darstellung der Funktion (63) mit der oberen Hyperbel ergeben sich aus der Gleichung $\sin\frac{1}{x}=1$, d. h., sie liegen bei $x=\frac{2}{(2k+1)\pi}$; entsprechend erhält man als Berührungspunkte der graphischen Darstellung von (63) mit der unteren Hyperbel $x=\frac{2}{(2k+3)\pi}$ $\left(\text{das sind die Lösungen der Gleichung } \sin\frac{1}{x}=-1\right)$. Da die Funktion (63) gerade ist, liegt ihr graphisches Bild symmetrisch in bezug auf die y-Achse. Offenbar schneidet es die y-Achse nicht; denn die Funktion (63) ist für $x=0$ nicht erklärt.

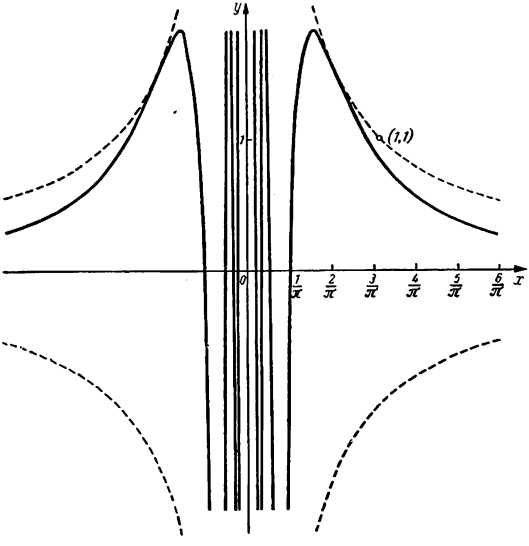


Abb. 75

Beispiel 6.

$$f(x)=x\sin\frac{1}{x}. \tag{64}$$

Auch diese Funktion charakterisiert eine Schwingung, die dadurch gekennzeichnet ist, daß ihre Amplituden bei Annäherung an den Ursprung O beliebig klein werden. Für $x>0$ ist offenbar

$$|f(x)|\leqq x,$$

so daß

$$-x\leqq f(x)\leqq +x. \tag{65}$$

Gleichheit tritt rechts für $x = \frac{2}{(2k+1)\pi}$ und links für $x = \frac{2}{(2k+3)\pi}$ ein. Die Maxima und Minima der Funktion liegen etwas rechts von diesen Punkten.[1]) Da die Funktion (64) gerade ist, liegt ihre graphische Darstellung symmetrisch in bezug auf die y-Achse (Abb. 76). Trotz Bestehens der Ungleichungen (65) ist die Funktion (64) im Punkte $x = 0$ unstetig; denn im Punkte $x = 0$ ist die Funktion (64) nicht erklärt, so daß auf der y-Achse kein Punkt ihrer graphischen Darstellung liegt.[2])

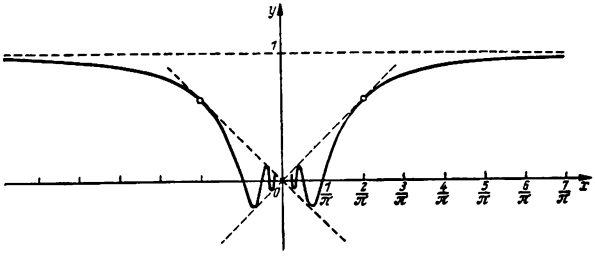


Abb. 76

Die graphischen Bilder der in Beispiel 5 und Beispiel 6 betrachteten Funktionen zeichnet man am einfachsten nach den in § 2 angegebenen systematischen Verfahren.

---

[1]) Man erhält dies am einfachsten mit Hilfe der Differentialrechnung. Offenbar ist

$$f'(x) = \sin\frac{1}{x} - \frac{1}{x}\cos\frac{1}{x}.$$

Daraus folgt

$$f'\left(\frac{2}{(2k+1)\pi}\right) = (-1)^k + 0.$$

Die Gleichung $f'(x) = 0$ löst man am einfachsten, indem man sie auf die Form $\operatorname{tg}\frac{1}{x} = \frac{1}{x}$ oder $\operatorname{tg}\xi = \xi$ mit $\xi = \frac{1}{x}$ bringt und diese Gleichung graphisch löst.

[2]) Wir setzen $\frac{1}{x} = u$. Da $\sin u < u$ für $u > 0$, ist $x \sin\frac{1}{x} < 1$ für $x > 0$. Ferner folgt aus

$$\lim_{u \to 0} \frac{\sin u}{u} = 1$$

(vgl. Formel (89) auf Seite 181)

$$\lim_{x \to \infty} x \sin\frac{1}{x} = 1.$$

Daher ist die Gerade $y = 1$ eine Asymptote des graphischen Bildes der Funktion (64); das graphische Bild selbst nähert sich ihr von unten her.

Beispiel 7.

$$f(x) = 2^{-\frac{1}{x^2}}. \tag{66}$$

Diese Funktion ist gleichfalls für $x = 0$ nicht erklärt. Sie nähert sich jedoch, wenn man von rechts oder von links her gegen $x = 0$ geht, ohne „Schwingungen" sehr schnell dem Wert Null (Abb. 77).

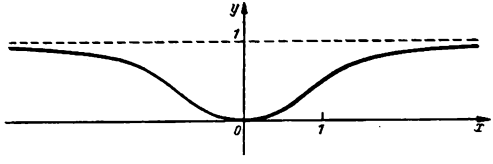


Abb. 77

Die Beispiele der genannten Art geben zu folgenden präzisierenden Definitionen Anlaß: Die Funktion $f(x)$ sei definiert in einem gewissen Intervall mit Ausnahme evtl. des Endpunktes c, wo sie entweder definiert oder auch nicht definiert ist. Wenn dann für jede Zahlenfolge $\{x_n\}$ mit

$$\left.\begin{array}{l} 1.\ x_n < c \\ 2.\ x_n \to c \end{array}\right\} \quad (n = 1,\ 2,\ \ldots) \tag{67}$$

aus diesem Intervall die Folge $\{f(x_n)\}$ der zugehörigen Funktionswerte gegen ein und denselben endlichen oder unendlichen Grenzwert L' strebt,

$$f(x_n) \to L', \tag{68}$$

so sagt man, *die Funktion $f(x)$ strebe bei linksseitiger Annäherung von x an den Punkt c gegen den Grenzwert L'.* Man schreibt hierfür

$$\lim_{\substack{x \to c \\ x < c}} f(x) = L',$$

$$f(x) \to L' \text{ für } x \to c,\ x < c,$$

$$\lim_{x \to c-} f(x) = L',$$

$$\lim_{x \to c-0} f(x) = L'$$

und gelegentlich

$$f(c-0) = L'.$$

Ist die Funktion $f(x)$ in einem gewissen Intervall mit Ausnahme evtl. des Anfangspunktes c definiert und strebt für jede Zahlenfolge $\{x_n\}$ mit

$$\left.\begin{array}{l} 1.\ x_n > c \\ 2.\ x_n \to c \end{array}\right\} \quad (n = 1,\ 2,\ \ldots) \tag{69}$$

aus diesem Intervall die Folge der zugehörigen Funktionswerte $f(x_n)$ gegen ein- und denselben Grenzwert L'',

$$f(x_n) \to L'', \tag{70}$$

so sagt man, *die Funktion $f(x)$ strebe bei rechtsseitiger Annäherung von x an c gegen den Grenzwert L''*. Man schreibt dafür:

$$\lim_{\substack{x \to c \\ x > c}} f(x) = L'',$$

$$f(x) \to L'' \text{ für } x \to c,\ x > c,$$

$$\lim_{x \to c+} f(x) = L'',$$

$$\lim_{x \to c+0} f(x) = L''$$

und gelegentlich

$$f(c+0) = L''.$$

In den oben betrachteten Beispielen war stets $c = 0$. Wir erhalten dann

in Beispiel 1: $L' = 0$, $L'' = +\infty$;
in Beispiel 2: $L' = 1$, $L'' = 0$;
in Beispiel 3: a) $L' = -\infty$, $L'' = +\infty$; b) $L' = L'' = +\infty$;
in Beispiel 4 und Beispiel 5 existieren weder L' noch L'';
in Beispiel 6 und Beispiel 7: $L' = L'' = 0$.

Es sei ausdrücklich darauf hingewiesen, daß die Zahl $f(c)$, der Funktionswert der Funktion $f(x)$ für $x = c$, für die Grenzwerte L' und L'' keinerlei Rolle spielt; die Funktion $f(x)$ braucht unter Umständen im Punkte $x = c$ „keinen Sinn zu haben".

Gleichwertig der oben angegebenen Definition des Grenzwertes L' sind die folgenden Bedingungen:

1. Für endliches L': *Zu jedem noch so kleinen $\varepsilon > 0$ gibt es eine Zahl $\delta = \delta_\varepsilon > 0$, so daß für alle x mit*

$$x < c, \quad x > c - \delta \tag{71}$$

die folgende Ungleichung erfüllt ist:

$$|f(x) - L'| < \varepsilon. \tag{72}$$

2. Für $L' = +\infty$: *Zu jeder noch so großen Zahl M gibt es eine Zahl $\delta = \delta_M > 0$, so daß für alle x, für die* (71) *gilt, die folgende Ungleichung erfüllt ist:*

$$f(x) > M. \tag{73}$$

3. Für $L' = -\infty$: *Zu jeder noch so großen Zahl M gibt es eine Zahl $\delta = \delta_M > 0$, so daß für jedes x, für das* (71) *gilt, die folgende Ungleichung erfüllt ist:*

$$f(x) < -M. \tag{74}$$

Durch analoge Eigenschaften kann auch der Grenzwert L'' charakterisiert werden.

Wir wollen uns hier nicht mit dem Gleichwertigkeitsbeweis aufhalten. Er wird ganz entsprechend wie der Gleichwertigkeitsbeweis im vorangehenden Paragraphen geführt.

§ 42. Der zweiseitige Grenzwert und der Begriff der Stetigkeit

Die Größen L' und L'' des vorangehenden Paragraphen werden auch *linksseitiger* bzw. *rechtsseitiger Grenzwert* der Funktion $f(x)$ im Punkte $x = c$ genannt. Man bezeichnet sie zusammenfassend auch als *einseitige Grenzwerte*.

Wenn in einem gewissen Punkte $x = c$ der linksseitige und der rechtsseitige Grenzwert einer Funktion $f(x)$ übereinstimmen,

$$L' = L'' = L,$$ [1])

so nennt man ihren gemeinsamen Wert L den *zweiseitigen Grenzwert* oder kurz den *Grenzwert* der Funktion $f(x)$ im Punkte $x = c$. Um anzudeuten, daß L der Grenzwert der Funktion $f(x)$ im Punkte $x = c$ ist, schreibt man

$$\lim_{x \to c} f(x) = L \tag{75}$$

oder

$$f(x) \to L \quad \text{für} \quad x \to c. \tag{76}$$

Unabhängig von den einseitigen Grenzwerten kann man den zweiseitigen Grenzwert auch folgendermaßen charakterisieren: Die Funktion $f(x)$ sei in einer gewissen Umgebung, des Punktes $x = c$, mit Ausnahme evtl. von c selbst, definiert. Konvergiert für jede Zahlenfolge $\{x_n\}$ aus dieser Umgebung, welche die Eigenschaften

$$\begin{aligned} x_n &\neq c, \\ x_n &\to c \end{aligned} \tag{77}$$

besitzt, die Folge $\{f(x_n)\}$ der zugehörigen Funktionswerte stets gegen ein- und denselben endlichen oder unendlichen Grenzwert L,

$$f(x_n) \to L, \tag{78}$$

so sagt man, *die Funktion $f(x)$ strebe bei beliebiger Annäherung von x an den Punkt c gegen den Grenzwert L.*

Dies ist nun offenbar notwendig und hinreichend dafür, daß L der Grenzwert der Funktion $f(x)$ im Punkte $x = c$ im Sinne der oben angegebenen Definition ist. In der Tat: Wenn die Beziehung (78) für jede Folge $\{x_n\}$ gilt, die der Forderung (77) genügt, so ist sie insbesondere für jede Folge erfüllt, die der Forderung (67) oder der Forderung (69) genügt, so daß aus der Existenz des Grenzwertes L speziell die des linksseitigen Grenzwertes L' und die des rechtsseitigen Grenzwertes L'' folgt, wobei $L' = L'' = L$. Existieren umgekehrt der linksseitige Grenzwert L' und der rechtsseitige Grenzwert L'' und sind diese Grenzwerte gleich, $L' = L'' = L$, so kann man jede Folge $\{x_n\}$, die der Forderung (77) genügt, als eine Folge auffassen, die aus zwei Folgen „zusammengesetzt" ist, von denen die eine die Forderungen (67) und die andere die Forderungen (69) erfüllt; entsprechend ist dann die zugehörige Folge $\{f(x_n)\}$ aus zwei Folgen „zusammengesetzt", die beide den Grenzwert L besitzen, so daß auch $f(x_n) \to L$.

---

[1]) Hierbei seien auch der Fall $L = +\infty$ und der Fall $L = -\infty$ zugelassen.

Schließlich läßt sich der Grenzwert L auch noch durch folgende Bedingungen charakterisieren.

1. Im Fall eines endlichen L: *Zu jedem noch so kleinen* $\varepsilon > 0$ *gibt es eine Zahl* $\delta = \delta_\varepsilon > 0$, *so daß für alle* x *mit*

$$x \neq c, \quad |x - c| < \delta \tag{79}$$

die Ungleichung

$$|f(x) - L| < \varepsilon \tag{80}$$

erfüllt ist.

2. Für $L = +\infty$: *Zu jeder noch so großen Zahl* M *kann man eine Zahl* $\delta = \delta_M > 0$ *so finden, daß für alle* x, *die der Bedingung* (79) *genügen, die Ungleichung*

$$f(x) > M \tag{81}$$

erfüllt ist.

3. Für $L = -\infty$: *Zu jeder noch so großen Zahl* M *kann man eine Zahl* $\delta = \delta_M > 0$ *so finden, daß für alle* x, *die der Bedingung* (79) *genügen, die Ungleichung*

$$f(x) < -M \tag{82}$$

erfüllt ist.

Die in den Beispielen 1, 2, 3a, 4 und 5 des vorangehenden Paragraphen betrachteten Funktionen besitzen im Punkte $x = 0$ keinen Grenzwert; für die in den Beispielen 6 und 7 betrachteten Funktionen ist $L = \lim_{x \to 0} f(x) = 0$, während für die in Beispiel 3b betrachtete Funktion $L = +\infty$ ist.

Es sei darauf hingewiesen, daß es sowohl für den numerischen Wert des „zweiseitigen" Grenzwertes

$$L = \lim_{x \to c} f(x)$$

als auch für seine Existenz ohne Belang ist, welchen Wert $f(c)$ die Funktion $f(x)$ im Punkte $x = c$ annimmt, ja, ob sie dort überhaupt definiert ist.

Mit dem Begriff des Grenzwertes einer Funktion in einem Punkte hängt eng der Begriff der Stetigkeit der Funktion in diesem Punkte zusammen. Dazu nehmen wir an, daß die Funktion $f(x)$ in allen Punkten einer gewissen Umgebung des Punktes $x = c$ definiert ist, daß die Funktion im Punkte $x = c$ einen linksseitigen Grenzwert L' und einen rechtsseitigen Grenzwert L'' besitzt und daß die Größen L', L'' und $f(c)$ übereinstimmen:

$$L' = L'' = f(c). \tag{83}$$

Ist dies der Fall, so nennt man die Funktion $f(x)$ im Punkte $x = c$ *stetig*.

Auf Grund des oben Bewiesenen *ist die Funktion* $f(x)$ *dann und nur dann im Punkte* $x = c$ *stetig, wenn sie im Punkte* $x = c$ *einen endlichen Grenzwert* L *besitzt und wenn dieser Grenzwert* L *mit dem Funktionswert* $f(c)$ *übereinstimmt*:

$$L = f(c). \tag{84}$$

Keine der in den Beispielen 1—7 des vorangehenden Paragraphen behandelten Funktionen ist im Punkte $x = 0$ stetig, und zwar aus folgenden Gründen: In den Beispielen 1,

2, 3a ist $L' \neq L''$, so daß diese Funktionen im Punkte $x = 0$ keinen Grenzwert besitzen; in Beispiel 3b existiert zwar der Grenzwert im Punkte $x = 0$, ist jedoch nicht endlich; in den Beispielen 4 und 5 existieren weder L' noch L''; in den Beispielen 6 und 7 existiert zwar der Grenzwert L, ist auch endlich, jedoch sind die dort angegebenen Funktionen im Punkte $x = 0$ nicht definiert.

Übrigens trifft der zuletzt genannte Sachverhalt auf alle anderen Funktionen auch zu, so daß schon aus diesem Grunde keine der in den genannten Beispielen behandelten Funktionen im Punkte $x = 0$ stetig ist.

Die angegebene Definition der Stetigkeit einer Funktion $f(x)$ in einem Punkte c macht vom Begriff des Grenzwertes Gebrauch. Das kann man dadurch vermeiden, daß man den Inhalt dieses Begriffes in die Formulierung der Stetigkeitsdefinition einbezieht. Auf diese Weise gelangt man zu den folgenden Definitionen, die der oben angegebenen Definition äquivalent sind.

1. *Eine Funktion $f(x)$, die in allen Punkten einer gewissen Umgebung des Punktes $x = c$ definiert ist, ist im Punkte $x = c$ stetig, wenn für jede Folge $\{x_n\}$ aus dieser Umgebung, die gegen c konvergiert,*

$$x_n \to c, \text{ [1]} \tag{85}$$

die Folge $\{f(x_n)\}$ der zugehörigen Funktionswerte gegen $f(c)$ konvergiert,

$$f(x_n) \to f(c) \tag{86}$$

(Definition nach Heine).

2. *Eine Funktion, die in einer gewissen Umgebung des Punktes $x = c$ definiert ist, ist im Punkte $x = c$ stetig, wenn man zu jeder noch so kleinen Zahl $\varepsilon > 0$ eine Zahl $\delta = \delta_\varepsilon$ so finden kann, daß für alle x, die der Bedingung*

$$|x - c| < \delta \text{ [2]} \tag{87}$$

genügen, die Ungleichung

$$|f(x) - f(c)| < \varepsilon \tag{88}$$

erfüllt ist (Definition nach Cauchy).

Mit dem Begriff der Stetigkeit einer Funktion in einem Punkte ist eng der Begriff der Stetigkeit einer Funktion in einem Intervall verknüpft, und zwar nennt man eine Funktion *stetig in einem gewissen Intervall*[3], wenn sie in jedem Punkte dieses Intervalls stetig ist.

Zum Beispiel ist die Funktion $f(x) = \frac{1}{x}$ im Intervall $0 < x \leqq 1$ (den Anfangspunkt $x = 0$ also nicht mitgerechnet) stetig, während sie im Intervall $0 \leqq x \leqq 1$ nicht stetig ist, da sie für $x = 0$ nicht erklärt ist.

Eine Funktion $f(x)$, die in einer gewissen Umgebung des Punktes $x = c$ definiert ist, heißt *unstetig im Punkte $x = c$*, wenn sie dort nicht definiert oder nicht stetig ist.

---

[1]) Eine Beschränkung auf $x_n \neq c$ ist hier nicht mehr notwendig.

[2]) Die Einschränkung $x \neq c$ ist hier überflüssig.

[3]) Anfangs- und Endpunkt nach Belieben hinzugenommen.

§ 43. Beispiele für stetige Funktionen

Um nachzuweisen, daß eine gegebene Funktion $f(x)$ in einem Punkte c stetig ist, braucht man häufig weder alle Folgen $\{f(x_n)\}$ mit $x_n \to c$ zu untersuchen (was im allgemeinen praktisch gar nicht möglich ist) noch eine „Epsilon-Konstruktion" durchzuführen.

Zunächst bemerken wir, daß die Funktion

$$f(x) \equiv C,$$

wobei C irgendeine Konstante ist, und auch die Funktion

$$f(x) \equiv x$$

in jedem Punkte stetig sind. Im ersten Fall ist nämlich $|f(x) - f(c)|$ für jedes x gleich Null, während im zweiten Fall für alle x offenbar $|f(x) - f(c)| = |x - c|$ ist. Daher braucht man in (87) nur $\delta = \varepsilon$ zu setzen, damit (88) erfüllt ist.

Mit Hilfe der in § 38 bewiesenen Sätze über den Grenzwert von Zahlenfolgen kann man aus der Stetigkeit einiger Funktionen sofort auf die Stetigkeit einer ganzen Reihe von weiteren Funktionen schließen. Insbesondere gelten die folgenden Sätze:

Sind die Funktionen $u(x)$ und $v(x)$ beide im Punkte c stetig, so sind es dort auch a) *die Summe $u(x) + v(x)$*, b) *die Differenz $u(x) - v(x)$*, c) *das Produkt $u(x)\,v(x)$ und* d), *falls $v(c) \neq 0$, der Quotient $\frac{u(x)}{v(x)}$.*

Als Beispiel beweisen wir die Behauptung a). Sind $u(x)$ und $v(x)$ im Punkte c stetig, so gilt für jede Folge $\{x_n\}$: Wenn $x_n \to c$, so auch $u(x_n) \to u(c)$ und $v(x_n) \to v(c)$. Dann gilt aber nach Satz I aus § 38

$$u(x_n) + v(x_n) \to u(c) + v(c),$$

womit a) bewiesen ist.

Ferner gilt der folgende Satz über die Stetigkeit einer „zusammengesetzten" Funktion:

e) *Ist die Funktion $u(x)$ stetig im Punkte $x = c$ und die Funktion $f(u)$ stetig im Punkte $u = u(c)$, so ist die Funktion $F(x) = f(u(x))$ stetig im Punkte $x = c$.*

In der Tat: Wenn $x_n \to c$, so gilt wegen der Stetigkeit der Funktion $u(x)$

$$u(x_n) \to u(c),$$

und hieraus folgt wegen der Stetigkeit der Funktion $f(u)$

$$f(u(x_n)) \to f(u(c)),$$

d. h. $F(x_n) \to F(c)$.

Aus e) folgt sofort:

Ist die Funktion $v(x)$ in einem ganzen Intervall $a < x < b$ stetig, wobei die Funktionswerte von $v(x)$ alle in einem Intervall $c < x < d$ liegen, in dem eine andere Funktion $u(x)$ (überall) stetig ist, so ist die zusammengesetzte Funktion $u(v(x))$ im ganzen Intervall $a < x < b$ stetig. Wenn also insbesondere die Funktionen $u(x)$ und $v(x)$ für alle reellen Werte x definiert und stetig sind, so gilt dasselbe auch für die Funktion $u(v(x))$.

Aus a) bis c) folgt sofort, daß *jede ganzrationale Funktion* (*jedes Polynom*) $P(x)$ *in allen Punkten* (d. h. im Intervall $-\infty < x < +\infty$) *stetig ist.*

Hieraus und aus d) ergibt sich sodann:

Jede gebrochene rationale Funktion

$$R(x) = \frac{P(x)}{Q(x)}$$

(wobei $P(x)$ und $Q(x)$ Polynome ohne gemeinsamen Teiler sind) *ist überall stetig, mit Ausnahme der Punkte, in denen der Nenner* $Q(x)$ *verschwindet.* In diesen Ausnahmepunkten sind die Limites L' und L'' (vgl. den vorangehenden Paragraphen) unendlich, und zwar mit demselben oder mit verschiedenen Vorzeichen, je nach der Vielfachheit der Nullstelle von $Q(x)$ (vgl. § 16). Eine rationale Funktion besitzt also nur endlich viele Unstetigkeiten, und alle sind von einem der genannten Typen.

Die Wurzelfunktion

$$f(x) = \sqrt[n]{x},$$

wobei n eine beliebige natürliche Zahl ist, ist überall dort stetig, wo sie definiert ist.

Als Beispiel betrachten wir die positive Quadratwurzel ($n = 2$). Für jedes $c > 0$ ist

$$\left|\sqrt{x_n} - \sqrt{c}\right| = \left|\frac{x_n - c}{\sqrt{x_n} + \sqrt{c}}\right| = \frac{|x_n - c|}{\sqrt{x_n} + \sqrt{c}} \leqq \frac{|x_n - c|}{\sqrt{c}}.$$

Wenn nun $x_n \to c$, so strebt $\frac{|x_n - c|}{\sqrt{c}}$ und damit $\left|\sqrt{x_n} - \sqrt{c}\right|$ gegen Null. Im Punkte $c = 0$ kann natürlich nur von einer rechtsseitigen Stetigkeit der Funktion $f(x) = \sqrt{x}$ die Rede sein. Hierzu ist zu zeigen, daß $\sqrt{x_n} \to 0$, falls $x_n \to 0$ mit $x_n > 0$. Um aus

$$|x_n| < \delta$$

auf

$$\sqrt{x_n} < \varepsilon$$

schließen zu können, braucht man aber nur $\delta < \delta_\varepsilon = \varepsilon^2$ zu wählen.

Bezüglich des allgemeinen Falles ($n \geqq 3$) vgl. S. 233, Beispiel 2.

Die trigonometrischen Funktionen $\sin x$ *und* $\cos x$ *sind überall stetig.*

In der Tat: Wir betrachten auf dem Einheitskreis die Bögen $\widehat{AC} = c$ und $\widehat{AM} = x$ $(x \neq c)$. Unter C_1 bzw. M_1 verstehen wir die Projektion des Punktes C bzw. M auf die x-Achse und unter P die Projektion des Punktes C auf die Gerade M_1M. Dann ist offenbar

$$\sin x - \sin c = M_1M - C_1C = PM,$$
$$\cos x - \cos c = OM_1 - OC_1 = -PC;$$

hieraus folgt, wie man leicht sieht,

$$\left.\begin{aligned} |\sin x - \sin c| &= |PM| \\ |\cos x - \cos c| &= |PC| \end{aligned}\right\} < CM < \widehat{CM} = |x - c|,$$

so daß also, wenn $x_n \to c$, auch $\sin x_n \to \sin c$ und $\cos x_n \to \cos c$.

Aus der Stetigkeit der Polynome einerseits und der der trigonometrischen Funktionen andererseits folgt auf Grund von e) sofort, daß *alle trigonometrischen Polynome* (ganzrationalen Funktionen in $\sin x$ und $\cos x$, vgl. § 27) *überall stetig sind.*

Für die gebrochenen trigonometrischen Funktionen gilt ähnliches wie für die gebrochenen rationalen Funktionen, jedoch haben jene im Unterschied zu diesen, wenn sie überhaupt eine Unstetigkeitsstelle (Nullstelle des Nenners) besitzen, stets unendlich viele Unstetigkeiten, wobei allerdings in jeder Periode nur endlich viele Unstetigkeitsstellen liegen. So besitzen z. B. die trigonometrischen Funktionen $\operatorname{tg} x$, $\sec x$ und $\operatorname{cosec} x$ in einer Periode jeweils zwei Unstetigkeitsstellen.

Als Produkt überall stetiger Funktionen sind die Funktion $x \sin x$ und allgemein alle Funktionen der Form $P(x) \sin x$ und $P(x) \cos x$, wobei $P(x)$ eine ganzrationale Funktion ist, überall stetig.

Die Funktion $f(x) = \dfrac{\sin x}{x}$ ist als Quotient stetiger Funktion überall mit Ausnahme des Punktes $x = 0$ stetig. Für $x = 0$ ist die Funktion $f(x)$ nicht definiert, so daß dort auch nicht von der Stetigkeit der Funktion $f(x)$ die Rede sein kann.[1]) Trotzdem besitzt die Funktion $f(x)$ im Punkte $x = 0$ einen beiderseitigen Grenzwert, und zwar ist

$$\lim_{x \to 0} \frac{\sin x}{x} = 1. \tag{89}$$

In der Tat: Es sei zunächst $x > 0$. Dann ist für alle $x < \frac{\pi}{2}$ (vgl. § 25 und § 29)

$$\sin x < x < \operatorname{tg} x. \tag{90}$$

Dividieren wir die Ungleichung (90) gliedweise durch $\sin x$ und gehen wir anschließend zum reziproken Wert über, so erhalten wir

$$\cos x < \frac{\sin x}{x} < 1. \tag{91}$$

Wegen der Stetigkeit der Funktion $\cos x$ kann man nun zu jedem $\varepsilon > 0$ ein δ so finden, daß für alle $|x| < \delta$ die Ungleichung

$$|\cos x - 1| < \varepsilon$$

erfüllt ist, also

$$\cos x > 1 - \varepsilon$$

gilt. Dann ergibt sich jedoch aus (91)

$$1 - \varepsilon < \frac{\sin x}{x} < 1,$$

also

$$\left| \frac{\sin x}{x} - 1 \right| < \varepsilon.$$

[1]) Vgl. übrigens S. 483, Aufgabe 5.

Damit ist gezeigt, daß die betrachtete Funktion $f(x)$ im Punkte $x=0$ den rechtsseitigen Grenzwert $L''=1$ besitzt. Die Existenz des linksseitigen Grenzwerts L' und die Gleichheit $L'=L''=1$ ergibt sich daraus, daß die Funktion $\frac{\sin x}{x}$ gerade ist.

Die Exponentialfunktion $f(x)=a^x$ ist überall stetig.[1])

Ohne Beschränkung der Allgemeinheit können wir annehmen, daß $a>1$; im Fall $a<1$ braucht man nur zu beachten, daß $a^x=\frac{1}{(a^{-1})^x}$, wobei $a^{-1}>1$.

Als erstes beweisen wir, daß die Funktion a^x in jedem Punkt c einen rechtsseitigen Grenzwert L'' besitzt und daß

$$L''=a^c. \tag{92}$$

Dazu sei $\{x_n\}$ eine beliebige Zahlenfolge mit $x_n>c$ für $n=1, 2, \ldots$ und $x_n\to c$. Dann gilt

$$a^{x_n}=a^c\cdot a^{x_n-c}. \tag{93}$$

Wir setzen nun $x_n=c+h_n$. Offenbar ist $h_n>0$ für $n=1, 2, \ldots$, und es gilt

$$h_n\to 0. \tag{94}$$

Zu zeigen ist

$$a^{h_n}\to 1;$$

denn dann strebt die rechte und damit auch die linke Seite von (93) gegen a^c.

Der Beweis erfolgt in mehreren Schritten:

1. Es sei $h_n=\frac{1}{n}$ $(n=1, 2, \ldots)$. In diesem Fall ist zu zeigen, daß

$$\sqrt[n]{a}\to 1. \tag{95}$$

Auf S. 72 wurde gezeigt, daß für beliebiges $b>1$ die Ungleichung

$$b^n>1+(b-1)n,$$

d. h.

$$b-1<\frac{b^n-1}{n}$$

gilt. Insbesondere gilt dies, da wegen $a>1$ auch $\sqrt[n]{a}>1$ ist, für $b=\sqrt[n]{a}$:

$$\sqrt[n]{a}-1<\frac{a-1}{n}.$$

Da nun die linke Seite dieser Ungleichung stets positiv ist, während die rechte gegen Null strebt, geht auch die linke Seite gegen Null. Hieraus folgt (95).

[1]) Der nachfolgende Beweis ist sowohl richtig, wenn nur rationale Werte für x betrachtet werden, als auch dann, wenn beliebige reelle Werte zugelassen sind. Bezüglich der Definition der Exponentialfunktion a^x für irrationale Argumentwerte und der Übertragung ihrer wesentlichen Eigenschaften auf den Fall irrationaler Argumente vgl. § 51.

2. Es sei $h_n = \frac{1}{p_n}$, wobei $\{p_n\}$ eine beliebige Folge von positiven ganzen Zahlen mit $p_n \to \infty$ ist. In diesem Fall ergibt sich aus (95) unmittelbar

$$\sqrt[p_n]{a} \to 1$$

(vgl. S. 158).

3. Es sei schließlich $\{h_n\}$ eine beliebige Folge von positiven Zahlen mit $h_n \to 0$. Wir bestimmen dann zu jeder Zahl h_n diejenige ganze Zahl p_n, für die

$$p_n \leqq \frac{1}{h_n} < p_n + 1 .$$

Offenbar gilt $p_n \to \infty$. Da die Funktion a^x wachsend ist, können wir aus

$$h_n \leqq \frac{1}{p_n}$$

unmittelbar auf

$$a^{h_n} \leqq a^{\frac{1}{p_n}}$$

schließen. Da die linke Seite dieser letzten Ungleichung stets größer als 1 ist, während die rechte Seite gegen 1 strebt, gilt auch $a^{h_n} \to 1$.

Die Existenz des linksseitigen Grenzwerts L' der Funktion a^x im Punkte c und die Beziehung

$$L' = a^c \tag{96}$$

erhält man entsprechend, wenn man von der Beziehung

$$a^{x_n} = \frac{a^c}{a^{c-x_n}} \qquad (x_n < c)$$

Gebrauch macht und

$$c - x_n = h_n > 0$$

setzt.

Da a^c der Wert der Funktion a^x im Punkte $x = c$ ist, ergibt sich aus (92) und (96) sofort die Stetigkeit der Funktion a^x im Punkte c.

Die Funktion $\log_a x$ $(a > 1)$ *ist überall dort stetig, wo sie definiert ist, d. h. für* $x > 0$.

Beim Beweis machen wir wesentlich davon Gebrauch, daß die Funktion a^x wachsend ist. Wir betrachten die Funktion $\log_a x$ im Punkte $x = c > 0$ und zeigen, daß sie dort einen rechtsseitigen Grenzwert L'' besitzt, wobei $L'' = \log_a c$. Es sei $\{x_n\}$ eine beliebige Zahlenfolge mit $x_n > c$ für $n = 1, 2, \ldots$ und $x_n \to c$. Aus $x_n > c$ folgt, daß $\log_a x_n > \log_a c$. Zu zeigen ist, daß $\log_a x_n \to \log_a c$. Wir nehmen an, dies wäre nicht der Fall. Dann gäbe es eine Zahl $\varepsilon^* > 0$, so daß für beliebig große n die Ungleichung

$$\log_a x_n > \log_a c + \varepsilon^*$$

erfüllt werden kann. Für eben diese Werte n gilt dann aber (da die Funktion a^x wachsend ist)

$$a^{\log_a x_n} > a^{\log_a c + \varepsilon^*},$$

d. h.

$$x_n > c a^{\varepsilon^*} > c$$

im Widerspruch zu $x_n \to c$.

Damit ist die Existenz des rechtsseitigen Grenzwerts im Punkte c und die Beziehung $L''=\log_a c$ bewiesen. Entsprechend beweist man die Existenz des linksseitigen Grenzwerts L' und die Beziehung $L'=\log_a c$. Daraus ergibt sich dann $L'=L''=\log_a c$ und damit die Stetigkeit der Funktion $\log_a x$ im Punkte c.[1])

Die Stetigkeit der allgemeinen Potenzfunktion x^α für beliebiges α in jedem Punkte $x=c>0$ ergibt sich dann als unmittelbare Folgerung mit Hilfe des Satzes über die Stetigkeit einer zusammengesetzten Funktion. Es ist nämlich (vgl. § 24)

$$x^\alpha = a^{\alpha \log_a x}.$$

Setzen wir $\alpha \log_a x = v(x)$ und $a^x = u(x)$, so ist $x^\alpha = u(v(x))$, wobei die Funktion $u(x)$ überall und die Funktion $v(x)$ für $x>0$ stetig ist.

Entsprechend ergibt sich, daß die Funktion x^x für $x>0$ stetig ist ($x^x = a^{x \log_a x}$).

Aus dem Gesagten ergibt sich, daß bei den elementaren Funktionen die Stetigkeitsstellen sozusagen die Regel sind, während die Unstetigkeitsstellen Ausnahmen bilden. Viele elementare Funktionen sind überall dort stetig, wo sie definiert sind, andere sind bis auf „einzelne" Punkte überall stetig (in früheren Zeiten hat man sich häufig so ausgedrückt, ohne daß man näher präzisierte, was unter diesen „einzelnen Punkten" zu verstehen sei).

Als letztes Beispiel betrachten wir die Funktion

$$f(x)=\frac{1}{\sin\dfrac{1}{x}}. \tag{97}$$

Sie besitzt Unstetigkeiten in allen Punkten $x=\dfrac{1}{k\pi}$, wobei k eine beliebige von Null verschiedene ganze Zahl ist, und außerdem noch eine Unstetigkeit im Punkte $x=0$.

§ 44. Grenzwerte bei monotoner Änderung. Die Zahl e

Satz von Weierstrass. *Jede wachsende oder zumindest nicht-fallende nach oben beschränkte Folge $\{a_n\}$ besitzt einen endlichen Grenzwert.*

Hierbei hat man unter einer *wachsenden Folge* natürlich eine solche zu verstehen, bei der jedes Folgeglied a_{n+1} größer ist als das vorangehende Folgeglied a_n; ist jedes Folgeglied a_{n+1} nicht kleiner als das vorangehende Glied a_n, d. h. gilt

$$a_1 \leqq a_2 \leqq \cdots \leqq a_n \leqq \cdots, \tag{98}$$

so heißt die Folge $\{a_n\}$ *nicht-fallend.*

[1]) Ein allgemeiner Satz über die Stetigkeit von Umkehrfunktionen wird in § 52 bewiesen.

Ist nun die Folge $\{a_n\}$ wachsend oder nicht-fallend, so ist sie sicher nach unten beschränkt, da alle ihre Glieder größer als jede Zahl m sind, die kleiner als a_1 ist. Weil sie andererseits nach oben beschränkt sein sollte, ist sie also beschränkt. Dann besitzt sie aber nach dem Satz von BOLZANO und WEIERSTRASS (§ 36) wenigstens einen Häufungspunkt. Es ist zu zeigen, daß sie auch nur höchstens einen Häufungspunkt besitzen kann. Dazu nehmen wir an, sie besäße zwei Häufungspunkte a und a'. Ohne Beschränkung der Allgemeinheit können wir annehmen, daß $a < a'$. Wir betrachten dann eine beliebige Zahl ε, welche der Bedingung

$$0 < \varepsilon < \frac{a' - a}{2}$$

genügt. Offenbar ist dann

$$a + \varepsilon < a' - \varepsilon. \tag{99}$$

Da a' Häufungspunkt der Folge $\{a_n\}$ ist, liegt in der ε-Umgebung von a' wenigstens ein Glied a_N der Folge $\{a_n\}$:

$$a' - \varepsilon < a_N < a' + \varepsilon.$$

Da die Folge $\{a_n\}$ wachsend oder wenigstens nicht-fallend ist, sind auch alle auf a_N folgenden Glieder a_n größer als $a' - \varepsilon$:

$$a_n \geqq a_N > a' - \varepsilon \qquad (n = N+1,\ N+2, \ldots).$$

Es können also nur solche Glieder der Folge kleiner als $a' - \varepsilon$ sein, deren Index kleiner als N ist, und das sind endlich viele. Daher können wegen (99) in der ε-Umgebung von a nur endlich viele Glieder der Folge $\{a_n\}$ liegen, im Widerspruch dazu, daß a Häufungspunkt sein sollte. Damit ist gezeigt, daß unter den angegebenen Voraussetzungen die Folge $\{a_n\}$ genau einen Häufungspunkt besitzt, der dann definitionsgemäß ihr Grenzwert ist; die Folge $\{a_n\}$ ist konvergent.

Wir merken an, daß beim Satz von WEIERSTRASS die Folge $\{a_n\}$ natürlich nicht durchweg nicht-fallend zu sein braucht; es genügt vorauszusetzen, daß sie von einem bestimmten Glied an nicht-fallend (bzw. sogar wachsend) ist.

Was im Satz von WEIERSTRASS über nicht-fallende nach oben beschränkte Folgen ausgesagt wurde, gilt natürlich genauso für nicht-wachsende nach unten beschränkte Folgen.

Indem wir, wie es üblich ist, den Begriff der nicht-fallenden Folge und den der nicht-wachsenden Folge zum Begriff der *monotonen Folge*[1]) zusammenfassen, können wir beide Fälle in einer gemeinsamen Formulierung wiedergeben:

Jede monotone beschränkte Folge besitzt einen endlichen Grenzwert.

Oder allgemeiner: *Jede monotone Folge besitzt einen Grenzwert*, und zwar ist dieser Grenzwert endlich, wenn die Folge beschränkt ist, und unendlich, wenn sie unbeschränkt ist.

[1]) Vgl. § 5, S. 29.

Beispiel 1. Die Folge

$$\left\{1+\frac{1}{2^2}+\frac{1}{3^2}+\cdots+\frac{1}{n^2}\right\} \equiv 1,\quad 1+\frac{1}{2^2},\quad 1+\frac{1}{2^2}+\frac{1}{3^2},\ldots \tag{100}$$

ist monoton wachsend (trivial) und nach oben beschränkt. Vom Glied a_{2^p-1} kann man nämlich feststellen, daß

$$a_{2^p-1}=1+\frac{1}{2^2}+\frac{1}{3^2}+\cdots+\frac{1}{(2^p-1)^2}=$$

$$=\left\{\begin{array}{l}1+\\ +\frac{1}{2^2}+\frac{1}{3^2}+\\ +\frac{1}{4^2}+\frac{1}{5^2}+\frac{1}{6^2}+\frac{1}{7^2}+\\ +\cdots\cdots\cdots\cdots+\\ +\left(\frac{1}{2^{p-1}}\right)^2+\cdots+\left(\frac{1}{2^p-1}\right)^2\end{array}\right\}<\left\{\begin{array}{l}1+\\ +\frac{1}{2^2}+\frac{1}{2^2}+\\ +\frac{1}{4^2}+\frac{1}{4^2}+\frac{1}{4^2}+\frac{1}{4^2}+\\ +\cdots\cdots\cdots\cdots+\\ +\left(\frac{1}{2^{p-1}}\right)^2+\cdots+\left(\frac{1}{2^{p-1}}\right)^2\end{array}\right\}$$

$$<1+\frac{1}{2}+\frac{1}{4}+\cdots+\frac{1}{2^{p-1}}=2-\frac{1}{2^{p-1}}<2.$$

Wenn aber für jedes p

$$a_{2^p-1}<2 \tag{101}$$

gilt, so gilt auch für jedes n

$$a_n<2. \tag{102}$$

In der Tat: Wir wählen zunächst eine Zahl p so, daß

$$2^p-1\geqq n;$$

dann ist

$$a_n\leqq a_{2^p-1},$$

da die Folge $\{a_n\}$ wächst; hieraus und aus (101) ergibt sich jedoch sofort (102).

Damit ist klar, daß die Folge (100) einen endlichen Grenzwert besitzt, wobei

$$a=\lim a_n=\lim\left(1+\frac{1}{2^2}+\frac{1}{3^2}+\cdots+\frac{1}{n^2}\right)\leqq 2.$$

Es bereitet keine wesentlichen Schwierigkeiten, diesen Limes jeweils auf eine gegebene Anzahl von Dezimalstellen zu berechnen:

$$a=1{,}644\ldots.$$

Jedoch gelingt es erst mit Hilfsmitteln der Differential- bzw. Integralrechnung, den folgenden merkwürdigen Zusammenhang zwischen der Zahl a und der Zahl π festzustellen:

$$a=\frac{\pi^2}{6}$$

(dieses Ergebnis stammt von Euler).

Beispiel 2. Besonders bedeutungsvoll ist in der Analysis die Folge $\{a_n\}$ mit

$$a_n = \left(1+\frac{1}{n}\right)^n. \tag{103}$$

Wir empfehlen dem Leser, die Folgeglieder a_n für $n = 1, 2, 3, 4, 5$ mit einer Genauigkeit von 0,001 zu berechnen; dann berechne man a_n mit Hilfe der Logarithmentafel für $n = 10, 50, 100$ und 1000.

Die Folge (103) ist wachsend. Entwickelt man nämlich a_n nach dem binomischen Satz, so erhält man

$$\begin{aligned} a_n = 1 + 1 + \frac{1}{2!}\left(1-\frac{1}{n}\right) + \frac{1}{3!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right) \\ + \cdots + \frac{1}{n!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{n-1}{n}\right) \end{aligned} \tag{104}$$

und entsprechend

$$\begin{aligned} a_{n+1} = 1 + 1 + \frac{1}{2!}\left(1-\frac{1}{n+1}\right) + \frac{1}{3!}\left(1-\frac{1}{n+1}\right)\left(1-\frac{2}{n+1}\right) \\ + \cdots + \frac{1}{n!}\left(1-\frac{1}{n+1}\right)\left(1-\frac{2}{n+1}\right)\cdots\left(1-\frac{n-1}{n+1}\right) + \frac{1}{(n+1)^{n+1}}. \end{aligned}$$

Vom dritten Summanden an ist jeder Summand der zweiten Summe größer als der entsprechende Summand der ersten Summe, wobei in der zweiten Summe zusätzlich noch ein weiterer Summand vorhanden ist. Also gilt

$$a_n < a_{n+1}.$$

Weiter erkennt man sofort, daß die Folge $\{a_n\}$ nach oben beschränkt ist; denn aus (104) ergibt sich, daß für jedes n

$$\begin{aligned} a_n &< 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!} \\ &< 1 + \left(1 + \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^{n-1}}\right) < 1 + 2 = 3 \end{aligned}$$

ist.

Daher besitzt die Folge (103) einen endlichen Grenzwert, welcher höchstens gleich 3 ist und den man mit e bezeichnet:

$$\mathrm{e} = \lim \left(1+\frac{1}{n}\right)^n. \tag{105}$$

Der numerische Wert von e wird durch

$$\mathrm{e} = 2{,}718\,28\ldots$$

gegeben.

Die Zahl e spielt in der höheren Mathematik eine fundamentale Rolle. Sie bildet u. a. auch die Basis für die sogenannten „natürlichen" Logarithmen[1]) (vgl. S. 301).

[1]) Die Logarithmen zur Basis e deutet man häufig durch $\ln x$ oder $\log x$ an.

Die vorangehenden Ausführungen über monotone Folgen lassen sich sinngemäß auch auf monotone Funktionen übertragen:

Wenn die Funktion $f(x)$ in einer gewissen Umgebung des Punktes $+\infty$ definiert und dort wachsend (oder zumindest nicht-fallend) *und nach oben beschränkt ist, so besitzt sie einen endlichen Grenzwert*

$$L = \lim_{x \to \infty} f(x).$$

Wir haben zu zeigen, daß für jede Folge $\{x_n\}$ aus der betreffenden Umgebung von $+\infty$, für welche $x_n \to \infty$ gilt, die Folge $\{f(x_n)\}$ gegen jeweils denselben endlichen Grenzwert L konvergiert.

Zunächst ist klar, daß die Folge $\{f(n)\}$ wachsend und beschränkt ist (jedenfalls von einem gewissen Glied an) und somit einen Grenzwert besitzt. Wir setzen dann

$$L = \lim f(n). \tag{106}$$

Es sei nun $\{x_n\}$ eine Folge der genannten Art, wobei wir annehmen können, daß alle Folgeglieder größer als Eins sind. Wir bestimmen dann für jedes x_n diejenige positive ganze Zahl p_n, für welche

$$p_n \leqq x_n < p_n + 1 \qquad (n = 1, 2, \ldots) \tag{107}$$

gilt. Aus (107) folgt sofort

$$f(p_n) \leqq f(x_n) < f(p_n + 1). \tag{108}$$

Weil nun offenbar $p_n \to \infty$ und $p_n + 1 \to \infty$, erhalten wir aus (106) unmittelbar $f(p_n) \to L$ und $f(p_n + 1) \to L$, so daß wegen (108) auch

$$f(x_n) \to L.$$

Dasselbe gilt natürlich, falls die Funktion $f(x)$ in einer gewissen Umgebung von $+\infty$ fallend (nicht-wachsend) und nach unten beschränkt ist. Entsprechendes gilt ferner für den Punkt $-\infty$. Schließlich gelten entsprechende Sätze auch für die einseitige Annäherung an einen endlichen Punkt:

Wenn die Funktion $f(x)$ in einem gewissen Intervall $c < x < c + \delta$ bzw. $c - \delta < x < c$ mit $\delta > 0$ definiert ist und sich dort monoton verhält (nicht-fallend oder nicht-wachsend ist), *so besitzt die Funktion $f(x)$ im Punkte c einen rechtsseitigen bzw. linksseitigen Grenzwert.* Dieser Grenzwert ist genau dann endlich, wenn die Funktion im betreffenden Intervall beschränkt ist.

Es dürfte sich erübrigen, den Beweis für diese Behauptung hier vorzuführen.

Beispiel 3. Wenn man mit elementaren Hilfsmitteln leicht beweisen könnte, daß die Funktion

$$f(x) = \left(1 + \frac{1}{x}\right)^x$$

wachsend ist, so könnte man mit Hilfe des vorangehenden Satzes sofort feststellen, daß

$$\lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x = e. \tag{109}$$

Da wir jedoch über diese Eigenschaft der Funktion $f(x)$ nicht verfügen, müssen wir zum Beweis von (109) etwas anders vorgehen. Zunächst wissen wir, daß $f(n)\to e$. Wir wählen nun entsprechend wie oben

$$p_n \leqq x_n < p_n + 1.$$

Dann ist einerseits

$$\left(1+\frac{1}{x_n}\right)^{x_n} < \left(1+\frac{1}{p_n}\right)^{p_n+1} = \left(1+\frac{1}{p_n}\right)^{p_n}\cdot\left(1+\frac{1}{p_n}\right) \tag{110}$$

und andererseits

$$\left(1+\frac{1}{x_n}\right)^{x_n} > \left(1+\frac{1}{p_n+1}\right)^{p_n} = \frac{\left(1+\frac{1}{p_n+1}\right)^{p_n+1}}{1+\frac{1}{p_n+1}}. \tag{111}$$

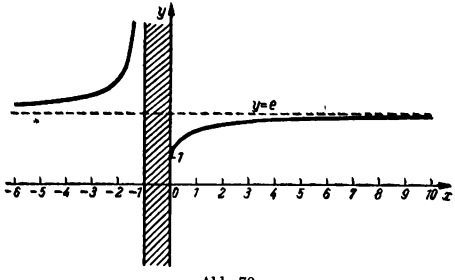


Abb. 78

Man sieht nun leicht, daß die rechte Seite sowohl von (110) als auch von (111) gegen e konvergiert; dann gilt aber auch

$$\lim\left(1+\frac{1}{x_n}\right)^{x_n} = e$$

und damit (109).

Als nächstes wollen wir $\lim\limits_{x\to-\infty} f(x)$ berechnen. Dazu sei $\{x_n\}$ eine beliebige Folge mit $x_n\to-\infty$. Wir setzen $|x_n| = -x_n = \xi_n$. Dann ist

$$f(x_n) = \left(1+\frac{1}{x_n}\right)^{x_n} = \left(1-\frac{1}{\xi_n}\right)^{-\xi_n} = \left(1+\frac{1}{\xi_n-1}\right)\cdot\left(1+\frac{1}{\xi_n-1}\right)^{\xi_n-1}.$$

Weil nun $\xi_n \to \infty$ und ebenso $\xi_n - 1 \to \infty$, konvergiert der erste Faktor auf der rechten Seite gegen 1, während der zweite, wie bereits bewiesen wurde, gegen e strebt. Damit erhalten wir neben (109) sofort auch

$$\lim_{x \to -\infty} \left(1 + \frac{1}{x}\right)^x = e. \tag{112}$$

In Abb. 78 ist die graphische Darstellung der Funktion $y = \left(1 + \frac{1}{x}\right)^x$ wiedergegeben.[1])

[1]) Im schraffierten Streifen liegen noch „einzelne" Punkte der graphischen Darstellung. So erhält man für $x = -\frac{1}{2}$ keinen (reellen) Wert für $f(x)$, während sich für $x = -\frac{1}{3}$ ein negativer und für $x = -\frac{2}{3}$ ein positiver Wert ergibt usw. Dieser zunächst etwas sonderbare Sachverhalt findet seine Erklärung in der Theorie der Funktionen einer komplexen Veränderlichen.

Kapitel IV

GRENZWERTE VON FUNKTIONENFOLGEN. EIGENSCHAFTEN STETIGER FUNKTIONEN

§ 45. Einfache Konvergenz von Funktionenfolgen

Bei der Behandlung der Zahlenfolgen haben wir erkannt (§ 38), daß man den Grenzübergang $a_n \to a$ als eine Operation auffassen kann, durch welche unter Umständen gegebenen Zahlen a_n $(n = 1, 2, \ldots)$ eine neue Zahl a, der Grenzwert der Folge $\{a_n\}$, zugeordnet wird:

$$a = \lim a_n .$$

Ganz entsprechend betrachten wir jetzt Funktionenfolgen

$$\{f_n(x)\} \equiv f_1(x),\ f_2(x), \ldots,\ f_n(x), \ldots, \tag{1}$$

wobei wir voraussetzen, daß alle Funktionen $f_n(x)$ in einem gegebenen (endlichen oder unendlichen) Intervall erklärt sind, und definieren auch hier einen Grenzübergang

$$f_n(x) \to f(x),$$

durch den unter Umständen den Funktionen $f_n(x)$ $(n = 1, 2, \ldots)$ eine neue Funktion $f(x)$, der Grenzwert der Funktionenfolge $\{f_n(x)\}$, zugeordnet wird:

$$f(x) = \lim_{n \to \infty} f_n(x). \tag{2}$$

Die Frage nach der Konvergenz einer Funktionenfolge kann man von verschiedenen Gesichtspunkten aus behandeln. Wir beschäftigen uns zunächst mit dem allgemeineren Begriff der *einfachen Konvergenz* und wenden uns dann (§ 48) dem spezielleren (und praktisch wichtigeren) Begriff der *gleichmäßigen Konvergenz* zu.

Vorgegeben sei also eine Funktionenfolge (1), wobei wir annehmen wollen, daß alle Funktionen $f_n(x)$ $(n = 1, 2, \ldots)$ im Intervall I definiert sind. Dann können wir für jede Zahl x_0 aus I die Zahlenfolge

$$\{f_n(x_0)\} \equiv f_1(x_0),\ f_2(x_0), \ldots,\ f_n(x_0), \ldots \tag{3}$$

bilden. Es kann nun der Fall eintreten, daß einige oder sogar alle diese Zahlenfolgen $\{f_n(x_0)\}$ einen endlichen Grenzwert besitzen, der natürlich im allgemeinen von der Wahl des Punktes x_0 abhängen wird und den wir daher mit $f(x_0)$ bezeichnen wollen.

Wenn nun für jedes x_0 aus I die Folge $\{f_n(x_0)\}$ im eigentlichen Sinne (vgl. § 38) gegen einen Grenzwert $f(x_0)$ konvergiert,

$$f_n(x_0) \to f(x_0),$$

so sagt man, *die Funktionenfolge $\{f_n(x)\}$ konvergiere in I gegen die Grenzfunktion (den Grenzwert) $f(x)$*; man schreibt dafür

$$f_n(x) \to f(x). \tag{4}$$

Es muß sofort bemerkt werden, daß hierbei $f(x)$ durchaus nicht immer eine elementare Funktion (§ 1) sein wird (auch dann nicht, wenn alle Funktionen $f_n(x)$ elementar sind). Wir können nur feststellen, daß jedem Wert x_0 aus I im betrachteten Fall eine eindeutige Zahl $f(x_0)$ zugeordnet ist. Zur Bestimmung von $f(x_0)$ wird dabei neben den elementaren Operationen die Operation des Grenzüberganges (für Zahlenfolgen) verwendet, und es ist nicht gesagt, daß man dasselbe Resultat auch ohne Verwendung dieser Operation erhalten kann. Wenn wir also den Grenzwert der Folge (1) als eine Funktion der Veränderlichen x ansehen, so verlassen wir zwangsläufig den Bereich der elementaren Funktionen. Alle Überlegungen sind natürlich vollständig unproblematisch, wenn man von vornherein den allgemeinen Funktionsbegriff verwendet. Die folgenden Beispiele werden zeigen, von welcher Natur die Funktionen sind, die man aus den elementaren Funktionen auf die angegebene Weise erhält.

Allgemein sind für einen Zahlenwert $x = x_0$ aus I die folgenden Fälle möglich:

1. Die Folge (3) besitzt einen endlichen Grenzwert.

2. Die Folge (3) besitzt einen uneigentlichen (unendlichen) Grenzwert ($+\infty$ oder $-\infty$).

3. Die Folge (3) besitzt keinen (weder einen endlichen noch einen unendlichen) Grenzwert.

Bezeichnen wir die entsprechenden Mengen von Punkten $x = x_0$ mit E_1, E_2, E_3, so können wir feststellen, daß jeder Punkt des Intervalls I in genau einer der Mengen E_1, E_2 oder E_3 liegt. Im ersten Fall können wir entsprechend wie oben den Grenzwert mit $f(x_0)$ bezeichnen; liegt x_0 in E_3, so hat $f(x_0)$ keinen Sinn; im zweiten Fall könnte man

$$f(x_0) = +\infty \quad \text{oder} \quad f(x_0) = -\infty$$

schreiben, was jedoch nicht üblich ist.[1]) Die Menge E_1, in der im allgemeinen Fall die Funktion $f(x)$ erklärt ist, nennt man den *Konvergenzbereich* der Folge (1) (im Intervall I).

---

[1]) Geometrisch besteht jedoch zwischen den Fällen x_0 aus E_2 und x_0 aus E_3 ein wesentlicher Unterschied.

Im folgenden wird die Schreibweise

$$f_n(x) \to f(x)$$

oder

$$\lim_{n\to\infty} f_n(x) = f(x)$$

nur dann verwendet, wenn E_1 mit dem Intervall I übereinstimmt, d. h., wenn für jedes x_0 aus I die zugehörige Folge (3) gegen einen endlichen Grenzwert konvergiert.

Wir betrachten nun einige Beispiele, wobei I stets die Menge aller reellen Zahlen ist.

Beispiel 1. Wir betrachten als erstes die Folge $\{f_n(x)\}$, wobei

$$f_n(x) = \left(1 + \frac{x}{n}\right)^n \qquad (n = 1, 2, 3, \ldots).$$

Es sind drei Fälle zu unterscheiden:

1. $x = 0$. In diesem Fall gilt $f_n(x) \to 1$.

2. $x > 0$. Hier setzen wir bei festgehaltenem x

$$\frac{n}{x} = \omega_n.$$

Da mit $n \to \infty$ auch $\omega_n \to \infty$, gilt nach § 44, Beispiel 3, auf Grund der Stetigkeit der Potenzfunktion

$$f_n(x) = \left\{\left(1 + \frac{x}{n}\right)^{\frac{n}{x}}\right\}^x = \left\{\left(1 + \frac{1}{\omega_n}\right)^{\omega_n}\right\}^x \to e^x.$$

3. $x < 0$. Setzen wir auch hier

$$\frac{n}{x} = \omega_n,$$

so gilt $\omega_n \to -\infty$ für $n \to \infty$, und wir erhalten (vgl. den Schluß von § 44)

$$f_n(x) \to e^x.$$

Es gilt also in allen drei Fällen

$$f(x) = \lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n = e^x. \tag{5}$$

Im betrachteten Beispiel konvergiert also die Funktionenfolge für alle x, und die Grenzfunktion ist eine elementare Funktion.

Beispiel 2. Es sei nun $f_n(x) = x^n$ $(n = 1, 2, 3, \ldots)$. Die Folge

$$\{x^n\} \equiv x,\ x^2,\ x^3,\ \ldots,\ x^n,\ \ldots$$

(vgl. § 37, Beispiel 5, und § 39, Beispiel 5) konvergiert für $-1 < x < 1$ gegen 0, für $x = 1$ gegen 1, divergiert für $x > 1$ gegen $+\infty$ und besitzt für $x \leqq -1$ keinen Grenzwert. Im vorliegenden Fall besteht also die Menge E_1 aus allen

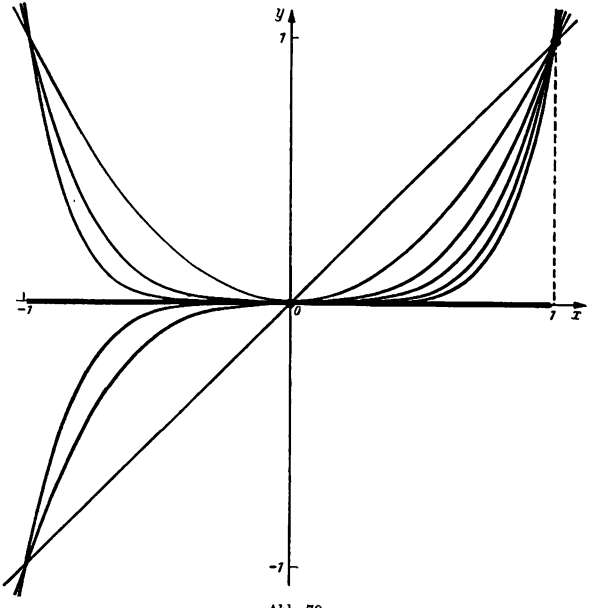


Abb. 79

Zahlen x mit $-1 < x \leqq 1$, die Menge E_2 aus allen Zahlen x mit $x > 1$ und die Menge E_3 aus allen Zahlen x mit $x \leqq -1$. Daher ist die Grenzfunktion $f(x) = \lim\limits_{n\to\infty} x^n$ für genau die x mit $-1 < x \leqq 1$ definiert, und zwar ist

$$f(x) = \begin{cases} 0 & \text{für } -1 < x < +1, \\ 1 & \text{für } x = 1. \end{cases} \tag{6}$$

Ihr graphisches Bild ist in Abb. 79 als fettgedruckte Linie mit einer Unstetigkeit (einem Sprung) für $x = 1$ dargestellt; zu beachten ist, daß die Punkte $(1, 0)$ und $(-1, 0)$ nicht dem graphischen Bild angehören.

Beispiel 3. Als nächstes betrachten wir die Funktionenfolge $f_n(x) = \frac{1}{1+n^2x^2}$ $(n = 1, 2, 3, \ldots)$. Die Folge

$$\left\{\frac{1}{1+n^2x^2}\right\} \equiv \frac{1}{1+x^2}, \quad \frac{1}{1+4x^2}, \quad \frac{1}{1+9x^2}, \ldots$$

konvergiert für jedes x. Wie man leicht sieht, wird die Grenzfunktion $f(x)$ durch

$$f(x) = \begin{cases} 1 & \text{für } x = 0, \\ 0 & \text{für } x \neq 0 \end{cases} \tag{7}$$

gegeben. Sie besitzt im Punkte $x = 0$ (vgl. Abb. 80) eine Unstetigkeitsstelle,

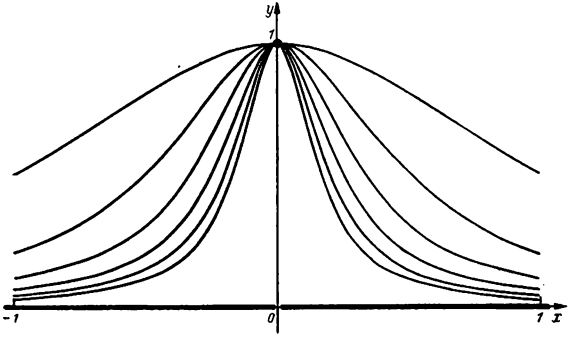


Abb. 80

es gilt nämlich (vgl. § 41)

$$L' = \lim_{\substack{x \to 0 \\ x < 0}} f(x) = 0, \quad L'' = \lim_{\substack{x \to 0 \\ x > 0}} f(x) = 0,$$

während

$$f(0) = 1.$$

Im betrachteten Beispiel erhält man die graphische Darstellung von $f_n(x)$ aus der von $f_1(x)$ durch Stauchung in Richtung der x-Achse mit dem Stauchungsfaktor n.

Zu einem ähnlichen Resultat gelangt man bei der Funktionenfolge

$$f_n(x) = \left(\frac{1}{1+x^2}\right)^n.$$

Beispiel 4. Es sei $f_n(x) = \cos^{2n} \pi x$ $(n = 1, 2, 3, \ldots)$.

Auch hier liegt Konvergenz auf der ganzen x-Achse vor. Die Grenzfunktion $f(x)$ wird gegeben durch

$$f(x) = \begin{cases} 1, & \text{falls } x \text{ eine ganze Zahl ist,} \\ 0 & \text{sonst.} \end{cases} \tag{8}$$

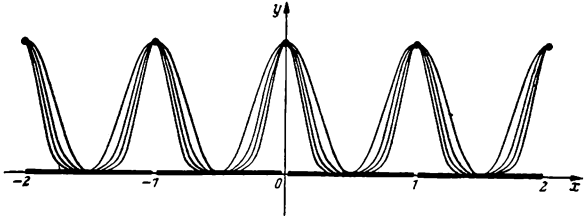


Abb. 81

Sie besitzt unendlich viele Sprungstellen, und zwar genau für die ganzzahligen Argumentwerte x (Abb. 81).

Beispiel 5. Bei der Funktionenfolge $f_n(x) = \cos^n \pi x$ $(n = 1, 2, 3, \ldots)$ gehören genau die Punkte $x = k$, wobei k eine ungerade Zahl ist, zu E_3,

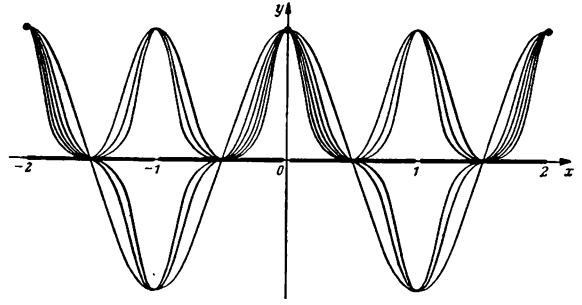


Abb. 82

während alle anderen Punkte der Menge E_1 angehören. Die Grenzfunktion $f(x)$ (vgl. Abb. 82) wird gegeben durch

$$f(x) = \begin{cases} 1, & \text{falls } x \text{ eine gerade ganze Zahl ist,} \\ 0, & \text{falls } x \text{ keine ganze Zahl ist.} \end{cases} \tag{9}$$

Beispiel 6. Wir betrachten die Funktionenfolge

$$f_n(x) = \frac{2}{\pi} \operatorname{arctg} nx \qquad (n = 1, 2, 3, \ldots).$$

Abb. 83 zeigt die graphischen Bilder der Funktionen $f_n(x) = \frac{2}{\pi} \operatorname{arctg} nx$. Man erhält die graphische Darstellung der Funktion $f_n(x)$ aus der von $f_1(x)$ durch Stauchung in Richtung der x-Achse mit dem Stauchungsfaktor n. Die Grenzfunktion $f(x)$ ist für alle Werte von x definiert und wird durch

$$f(x) = \begin{cases} -1 & \text{für } x < 0, \\ 0 & \text{für } x = 0, \\ +1 & \text{für } x > 0 \end{cases} \tag{10}$$

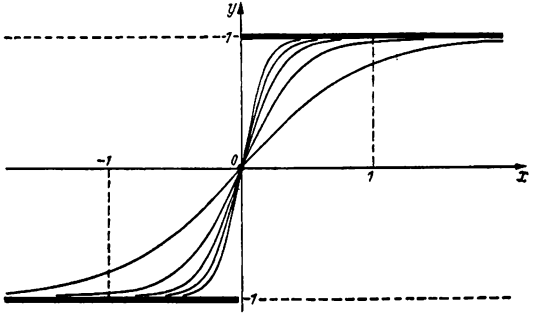


Abb. 83

gegeben. Die Funktion $f(x)$ ist im Punkte $x = 0$ unstetig, da im Punkte 0 der linksseitige Grenzwert gleich -1 und der rechtsseitige Grenzwert gleich $+1$ ist, während $f(0) = 0$. Die durch (10) definierte Funktion $f(x)$ wird häufig die *Signumfunktion* (Vorzeichenfunktion) genannt und mit $\operatorname{sgn} x$ bezeichnet.

Beispiel 7. Es sei

$$f_n(x) = \frac{2}{\pi} x \operatorname{arctg} nx \qquad (n = 1, 2, 3, \ldots). \tag{11}$$

Durch Grenzübergang findet man

$$\begin{aligned} f(x) &= \lim_{n \to \infty} f_n(x) \\ &= \lim_{n \to \infty} \frac{2}{\pi} x \operatorname{arctg} nx \\ &= x \cdot \lim_{n \to \infty} \frac{2}{\pi} \operatorname{arctg} nx \\ &= x \cdot \operatorname{sgn} x = |x|. \end{aligned}$$

Das graphische Bild der Funktion $f_n(x)$ erhält man aus dem von $f_1(x) = \frac{2}{\pi} x \operatorname{arctg} x$ (vgl. Abb. 84) durch gleichzeitige Stauchung in Richtung der x- und der y-Achse mit dem Stauchungsfaktor n, d. h. durch n-fache Verkleinerung. In der Grenze ergibt sich dabei eine aus zwei Halbstrahlen, den Winkelhalbierenden des ersten und zweiten Quadranten, bestehende Figur, und das ist ja die graphische Darstellung der Funktion $|x|$.

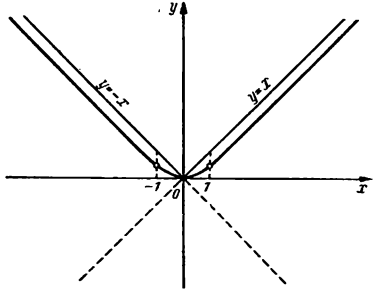


Abb. 84

Beispiel 8. Die Funktionenfolge $\{f_n(x)\}$ mit

$$f_n(x) = \frac{1}{1+2^{nx}}$$

besitzt als Grenzfunktion die Funktion

$$f(x) = \begin{cases} 1 & \text{für} \quad x < 0, \\ \frac{1}{2} & \text{für} \quad x = 0, \\ 0 & \text{für} \quad x > 0. \end{cases} \tag{12}$$

Beispiel 9. Es seien $u(x)$ und $v(x)$ zwei beliebige, für alle x erklärte Funktionen. Dann besitzt die Funktionenfolge $\{f_n(x)\}$ mit

$$f_n(x) = u(x) \cdot \frac{1}{1+2^{nx}} + v(x) \cdot \frac{1}{1+2^{-nx}}$$

die Grenzfunktion

$$f(x) = \begin{cases} u(x) & \text{für} \quad x < 0, \\ \frac{u(x)+v(x)}{2} & \text{für} \quad x = 0, \\ v(x) & \text{für} \quad x > 0. \end{cases}$$

Dieselbe Funktion erhält man, wenn man

$$f_n(x) = \frac{u(x)+v(x)}{2} - \frac{u(x)-v(x)}{2} \cdot \frac{2}{\pi} \operatorname{arctg} n x$$

setzt.

Beispiel 10. Ist $F(x)$ eine Funktion, welche für alle x definiert ist, so erhält man als Grenzfunktion der Funktionenfolge $\{f_n(x)\}$ mit

$$f_n(x) = \left[\frac{1}{1+F^2(x)}\right]^n$$

diejenige Funktion $f(x)$, die überall dort gleich 1 ist, wo $F(x) = 0$, und die überall dort gleich 0 ist, wo $F(x) \neq 0$.

Ist

$$f_n(x) = \frac{2}{\pi} \operatorname{arctg} nF(x),$$

so erhält man als Grenzfunktion die Funktion $f(x) = \operatorname{sgn} F(x)$, die den Wert -1, 0 oder $+1$ annimmt, je nachdem, ob $F(x)$ negativ, gleich Null oder positiv ist.

Beispiel 11. Die Funktion

$$f_p(x) = \lim_{q \to \infty} \cos^{2q}(10^p \pi x)$$

unterscheidet sich von der Grenzfunktion aus Beispiel 4 nur dadurch, daß ihr graphisches Bild in Richtung der x-Achse mit dem Faktor 10^p gestaucht ist. Sie nimmt also genau dann den Wert 1 an, wenn x sich als Dezimalbruch der Form $a_0, a_1 \ldots a_p$ darstellen läßt, und ist überall sonst gleich Null.

Wir betrachten nun die Funktion

$$f(x) = \lim_{p \to \infty} f_p(x),$$

d. h. den Grenzwert der Funktionenfolge

$$f_1(x), \quad f_2(x), \quad f_3(x), \ldots .$$

Ist x_0 eine Zahl, die sich nicht als endlicher Dezimalbruch darstellen läßt, so enthält die Folge $f_1(x_0), f_2(x_0), \ldots$ nur die Zahl Null, so daß auch ihr Grenzwert gleich Null ist; läßt sich dagegen die Zahl x_0 als endlicher Dezimalbruch darstellen, so enthält die Folge $\{f_n(x_0)\}$ endlich viele Glieder, die gleich Null sind, während alle übrigen Glieder gleich 1 sind, so daß auch der Grenzwert der Folge gleich 1 ist. Die Funktion $f(x)$ ist also überall definiert; sie nimmt für ein gegebenes x dann und nur dann den Wert 1 bzw. 0 an, wenn sich x durch einen endlichen Dezimalbruch darstellen läßt bzw. wenn dies nicht der Fall ist.

Das graphische Bild der Funktion $y = f(x)$ läßt sich natürlich nicht mehr zeichnen, jedoch kann man es sich vorstellen als eine Punktmenge, die sowohl auf der Geraden $y = 0$ als auch auf der Geraden $y = 1$ „überall dicht" liegt.

Man beachte, daß bei der Definition der Funktion $f(x)$ im zuletzt betrachteten Beispiel die Operation des Grenzübergangs zweimal ausgeführt wurde:

$$f(x) = \lim_{p \to \infty} \left[\lim_{q \to \infty} \cos^{2q}(10^p \pi x)\right]. \tag{13}$$

§ 46. Der allgemeine Begriff der Funktion einer reellen Veränderlichen

Die vorangehenden Beispiele zeigen, daß man, wenn man die Operation des Grenzübergangs gleichberechtigt zu den elementaren Operationen hinzunimmt, unbegrenzte Möglichkeiten für die Bildung neuer Funktionen erhält. So gewinnt man Funktionen, die Unstetigkeitsstellen der verschiedensten Typen besitzen[1]) (vgl. Beispiel 2 bis 6 in § 45), Funktionen, die aus mehreren gegebenen Funktionen „zusammengesetzt" sind (vgl. Beispiel 9), Funktionen, deren Wert davon abhängt, welcher Menge der zugehörige Argumentwert angehört (Beispiel 10 und 11) usw. Aus den angegebenen Gründen erscheint es künstlich und nur schwer zu rechtfertigen, wenn man sich auf die alleinige Betrachtung von „elementaren" Funktionen beschränkt.

Hinzu kommt noch ein weiterer entscheidender Gesichtspunkt: Beim Funktionsbegriff kommt es in erster Linie darauf an, daß jedem Argumentwert x ein ganz bestimmter Funktionswert $f(x)$ entspricht, und nicht darauf, mit Hilfe welcher Operationen man zufällig den Funktionswert berechnen kann.

Die angeführten Überlegungen dienen nicht nur zur Erklärung, sondern auch zur eigentlichen Rechtfertigung der folgenden Definition des allgemeinen Funktionsbegriffs, die heute in der Wissenschaft einzig üblich ist:

Man sagt, auf einer Menge E von Zahlen sei eine Funktion

$$y = f(x)$$

gegeben, wenn jeder Zahl x aus der Menge E auf irgendeine Weise ein und nur ein Wert $f(x)$ zugeordnet ist.

Es ist mit anderen Worten eine beliebige Vorschrift angegeben, vermöge der jedem Argumentwert x (aus E) ein eindeutig bestimmter Funktionswert $f(x)$ entspricht.

Durch diese Definition wird der Begriff der Funktion gegenüber dem Begriff der elementaren Funktion in ungewöhnlichem Maße erweitert, weil der Charakter der genannten „Vorschrift" durch nichts eingeschränkt wird. Natürlich kann diese „Vorschrift" in einer Formel bestehen, wie dies bei den elementaren Funktionen der Fall ist; dies braucht aber keineswegs der Fall zu sein.

Zugleich wird aber der Begriff der Funktion durch diese Definition gegenüber der früheren Definition der elementaren Funktion eingeengt, und zwar dadurch, daß jedem Wert x nur ein Wert $f(x)$ entsprechen soll, also „mehrdeutige Funktionen" ausgeschlossen werden.

Die Beschränkung auf *eindeutige* Funktionen hat für die Theorie der Funktionen einer reellen Veränderlichen eine grundlegende methodische Bedeutung: Wenn man sich auf die Betrachtung von eindeutigen Funktionen beschränkt (also evtl. auftretende mehrdeutige Funktionen — soweit dies möglich ist — in eine Reihe von eindeutigen Funktionen „aufspaltet"), vereinfachen sich die Formulierungen vieler Aussagen ganz erheblich.

Den hier eingeführten allgemeinen Funktionsbegriff haben wir bereits in § 1 erwähnt, wobei wir uns jedoch dort auf den Fall beschränkt haben, daß die Funktion in einem gewissen Intervall, nicht aber in einer beliebigen „Punktmenge" definiert ist.

---

[1]) Insbesondere erhält man also Funktionen, bei denen rechtsseitiger Grenzwert, linksseitiger Grenzwert und Funktionswert voneinander verschieden sind.

Wir wollen hier noch einige Beispiele für Funktionen geben, die nicht durch eine „geschlossene Formel" gegeben sind.

Beispiel 1. Unter

$$f(x) = [x] \tag{14}$$

versteht man üblicherweise diejenige Funktion, welche einer gegebenen Zahl x diejenige ganze Zahl g zuordnet, für die gilt: $g \leqq x < g+1$, d. h. die größte ganze Zahl, welche nicht größer als x ist. Beispielsweise ist also

$$\left[7\frac{1}{3}\right] = 7, \quad [3] = 3, \quad [-2{,}47] = -3.$$

Das graphische Bild dieser Funktion $f(x)$ ist in Abb. 85 dargestellt. Für jede ganze Zahl g ist im Punkte g der linksseitige Grenzwert von $f(x)$ gleich $g-1$ und der rechtsseitige Grenzwert gleich g.

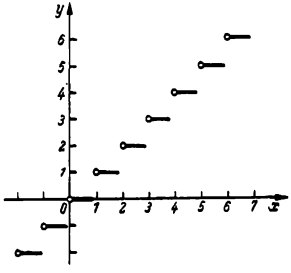


Abb. 85

Beispiel 2. Als nächstes erklären wir eine Funktion $f(x)$ im Intervall $0 \leqq x \leqq 1$ dadurch, daß wir sie zunächst auf der Menge E aller derjenigen Brüche definieren, deren Nenner die Form 2^n hat, und zwar sei

$$f(0) = 0, \quad f(1) = 1, \quad f\left(\frac{1}{2}\right) = \frac{f(0) + 3f(1)}{4},$$

sodann

$$f\left(\frac{1}{4}\right) = \frac{f(0) + 3f\left(\frac{1}{2}\right)}{4},$$

$$f\left(\frac{3}{4}\right) = \frac{f\left(\frac{1}{2}\right) + 3f(1)}{4}$$

usw.; allgemein sei für jedes m

$$f\left(\frac{2m+1}{2^{n+1}}\right) = \frac{f\left(\frac{m}{2^n}\right) + 3f\left(\frac{m+1}{2^n}\right)}{4}.$$

(Die Definition von $f(x)$ in den Punkten des Intervalls $0 \leqq x \leqq 1$, die nicht zu E gehören, erfolgt dann weiter „durch stetige Fortsetzung"; vgl. § 51.)

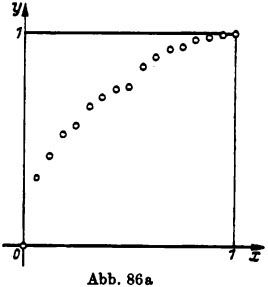


Abb. 86a

In Abb. 86a sind alle die Punkte des graphischen Bildes von $f(x)$ eingezeichnet, die den Werten x aus E mit $n \leqq 4$ entsprechen.

Beispiel 3. Ähnlich wie im vorangehenden Beispiel sei zunächst $f(0) = 0$, $f(1) = 1$. Ist dann $f\left(\frac{m}{2^n}\right) = \frac{p}{q}$, $f\left(\frac{m+1}{2^n}\right) = \frac{p'}{q'}$, wobei $\frac{p}{q}$ und $\frac{p'}{q'}$ reduzierte Brüche sind $\left(0 = \frac{0}{1}\right.$ gesetzt$\left.\right)$, so sei

$$f\left(\frac{2m+1}{2^{n+1}}\right) = \frac{p+p'}{q+q'}.$$

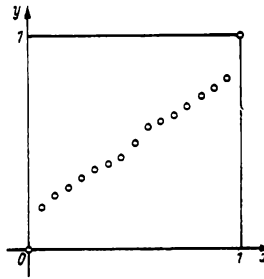


Abb. 86b

Man sieht leicht, daß

$$f\left(\frac{1}{2}\right) = \frac{0+1}{1+1} = \frac{1}{2},$$

$$f\left(\frac{1}{4}\right) = \frac{0+1}{1+2} = \frac{1}{3}, \quad f\left(\frac{3}{4}\right) = \frac{1+1}{2+1} = \frac{2}{3},$$

$$f\left(\frac{1}{8}\right) = \frac{0+1}{1+3} = \frac{1}{4}, \quad f\left(\frac{3}{8}\right) = \frac{1+1}{3+2} = \frac{2}{5},$$

$$f\left(\frac{5}{8}\right) = \frac{1+2}{2+3} = \frac{3}{5}, \quad f\left(\frac{7}{8}\right) = \frac{2+1}{3+1} = \frac{3}{4}$$

usw. (vgl. Abb. 86b). (Auch hier kann $f(x)$ von E stetig auf das Intervall $0 \leqq x \leqq 1$ fortgesetzt werden.)

Beispiel 4. Die Funktion $f(x)$ werde im Intervall $-1 \leqq x \leqq 1$ durch folgende Bedingungen definiert:

1. $f(0) = 0$;
2. $f(x) = x$, falls x die Form $\frac{1}{2g}$ (g ganz) hat, und $f(x) = 2x$, falls x die Form $\frac{1}{2g+1}$ hat;
3. $f(x)$ ist in jedem der Intervalle $\frac{1}{n+1} \leqq x \leqq \frac{1}{n}$ $(n = 1, 2, 3, \ldots)$ linear;
4. $f(x)$ ist gerade.

Der Leser fertige eine graphische Darstellung der Funktion $f(x)$ an und überzeuge sich davon, daß diese Funktion im Punkte $x = 0$ ein Minimum besitzt, ohne daß es möglich ist, eine Zahl $\varepsilon > 0$ so anzugeben, daß im Intervall $-\varepsilon < x < 0$ die Funktion $f(x)$ fällt und im Intervall $0 < x < \varepsilon$ wächst.[1])

Beispiel 5. Die Funktionen $f_1(x)$ und $f_2(x)$ seien für $x \neq 0$ durch

$$f_1(x) = 2^{-\frac{1}{x^2}}, \quad f_2(x) = \frac{\sin x}{x}$$

und durch

$$f_1(0) = 0, \quad f_2(0) = 1$$

definiert. Man sieht sofort, daß $f_1(x)$ und $f_2(x)$ überall (d. h. für $-\infty < x < \infty$) stetig sind (vgl. § 41, Beispiel 7, und § 43).

---

[1]) Dieselbe Eigenschaft besitzen auch die Funktionen

$$f(x) = 2|x| + x \sin\frac{1}{x}, \quad f(x) = x^{\frac{2}{3}} + x \sin\frac{1}{x} \qquad (x \neq 0).$$

Beispiel 6. Auf Seite 6 haben wir die Gleichung

$$x^2 + y^2 = 1$$

betrachtet und festgestellt, daß es im Intervall $-1 \leqq x \leqq 1$ zwei Funktionen gibt, die diese Gleichung identisch erfüllen, nämlich

$$f_1(x) = \sqrt{1-x^2} \quad \text{und} \quad f_2(x) = -\sqrt{1-x^2}.$$

Wir wollen hier nochmals auf diese Frage zurückkommen und untersuchen, ob dies die einzigen Funktionen sind oder ob es noch weitere Funktionen gibt, die der angegebenen Gleichung genügen. Von dem in diesem Paragraphen eingenommenen Standpunkt aus müssen wir feststellen, daß $f_1(x)$ und $f_2(x)$ keineswegs die einzigen Funktionen sind, die das Verlangte leisten, sondern daß es daneben noch unendlich viele weitere derartige Funktionen gibt. Zum Beispiel leistet auch die Funktion

$$f_3(x) = \begin{cases} \sqrt{1-x^2} & \text{für } -1 \leqq x \leqq 0, \\ -\sqrt{1-x^2} & \text{für } 0 < x \leqq +1 \end{cases}$$

das Verlangte. Diese Funktion ist nicht elementar; sie besitzt im Punkte $x = 0$ eine Sprungstelle.

Wir nehmen nun an, das Intervall $-1 \leqq x \leqq 1$ sei auf irgendeine Weise in zwei Mengen A und B eingeteilt, die keinen Punkt gemeinsam haben. Dann genügt auch die Funktion

$$f(x) = \begin{cases} \sqrt{1-x^2}, & \text{falls } x \text{ in } A \text{ liegt}, \\ -\sqrt{1-x^2}, & \text{falls } x \text{ in } B \text{ liegt}, \end{cases}$$

der betrachteten Gleichung.

Unter den genannten Funktionen sind allerdings nur die Funktionen $f_1(x)$ und $f_2(x)$ stetig.

Der zu Beginn dieses Paragraphen skizzierte Gedankengang spiegelt auch ungefähr den geschichtlichen Entwicklungsprozeß wider. Bei der geschichtlichen Entwicklung des Funktionsbegriffs trat allerdings der Grenzübergang nicht sofort in der allgemeinen (augenscheinlichen) Form, sondern in Form der Summation von Funktionsreihen auf. Durch die Arbeiten von FOURIER (im ersten Viertel des vorigen Jahrhunderts) wurden die Vorbedingungen für eine Erweiterung des Funktionsbegriffs geschaffen, indem nämlich in ihnen Methoden (in Gestalt von Formeln mit Grenzübergängen) für die analytische Darstellung auch solcher Funktionen gegeben wurden, die man bis dahin nicht als analytisch darstellbar ansah und die man deshalb auch nicht näher untersucht hatte. Danach wurde eine neue, im wesentlichen die heute allgemein übliche Definition des Funktionsbegriffs unumgänglich; sie wurde zuerst von LOBATSCHEWSKI (vgl. S. 4) und LEJEUNE DIRICHLET formuliert und von RIEMANN verwendet.

Die in der oben angegebenen Definition auftretende Menge E, also die Gesamtheit derjenigen Zahlen x, für welche die Funktion $f(x)$ erklärt ist, wird der *Definitionsbereich* der Funktion genannt. Sehr häufig ist der Definitionsbereich einer Funktion ein Intervall. Dabei hat man zwischen *abgeschlossenen Intervallen* der Gestalt $a \leqq x \leqq b$ und *offenen Intervallen* der Gestalt $a < x < b$ zu unterscheiden. Im abgeschlossenen Intervall $a \leqq x \leqq b$ ist die Zahl a kleinste und die Zahl b größte Zahl, während es in einem offenen Intervall weder eine größte noch eine kleinste Zahl gibt. Daneben hat man es mitunter

auch mit Intervallen zu tun, die an einem Ende offen und am anderen Ende abgeschlossen sind (z. B. dem Intervall $-1 < x \leqq 1$ usw.). Abgeschlossene Intervalle werden auch *Segmente* genannt.

Unendliche Intervalle werden an dem Ende, das sich ins Unendliche erstreckt, als offen angesehen.

Heute sind die folgenden Bezeichnungen allgemein üblich:

$[a, b]$ für das abgeschlossene Intervall $a \leqq x \leqq b$;

(a, b) für das offene Intervall $a < x < b$;

$[a, b)$ für das halboffene Intervall $a \leqq x < b$ usw.

Bemerkung. Der Leser wird wahrscheinlich eine Analogie zwischen der Definition der Zahlenfolge in § 35 und der Definition der Funktion im vorliegenden Paragraphen bemerkt haben. Hier liegt jedoch nicht nur eine Analogie vor, vielmehr ist der allgemeine Folgenbegriff ein Spezialfall des allgemeinen Funktionsbegriffs. Vom Standpunkt des vorliegenden Paragraphen aus ist eine Folge nichts anderes als eine auf der Menge aller positiven ganzen Zahlen erklärte Funktion.

Wie aus der angegebenen Definition ersichtlich ist, kann der Definitionsbereich einer Funktion eine ganz beliebige Menge von Zahlen sein; wir werden es allerdings im folgenden hauptsächlich mit Funktionen zu tun haben, die in einem Intervall einer der genannten Arten erklärt sind.

§ 47. Eigenschaften stetiger Funktionen

Um zu einer richtigen Vorstellung des allgemeinen Funktionsbegriffs zu gelangen, müssen wir noch kurz etwas über die graphische Darstellung einer Funktion sagen. Definitionsgemäß wird durch eine Funktion $f(x)$ jedem Wert x (eventuell aus einem gewissen Intervall, allgemein aus dem Definitionsbereich der Funktion) ein eindeutig bestimmter Wert $y = f(x)$ zugeordnet. Geometrisch bedeutet dies, daß auf jeder Vertikalen durch die gegebene „Vorschrift" ein bestimmter Punkt ausgezeichnet wird (praktisch ist es natürlich im allgemeinen nicht möglich, alle diese Punkte zu markieren; man kann sich jedoch „vorstellen", daß dies geschehen ist). Die Menge der auf diese Weise entstehenden Punkte braucht selbstverständlich keine „zusammenhängende Kurve" zu bilden (vgl. etwa Beispiel 11 aus § 45); das wird genau dann der Fall sein, wenn die Funktion $f(x)$ (im betrachteten Intervall) stetig ist.[1]) Obwohl in der Mathematik das Studium von Funktionen mit „Unstetigkeitsstellen" nicht nur in theoretischer Hinsicht, sondern auch für viele Anwendungen von grundlegender Bedeutung ist, bilden die stetigen Funktionen (und ihre graphischen Bilder, die stetigen Kurven) eine besonders wichtige Klasse von Funktionen, die bemerkenswerte Eigenschaften besitzen. Mit einigen dieser Eigenschaften wollen wir uns im vorliegenden Paragraphen beschäftigen. Auf weitere wesentliche Eigenschaften kommen wir in den §§ 49 und 50 zu sprechen.

Satz I. *Nimmt eine in einem Intervall stetige Funktion in zwei Punkten dieses Intervalls Werte mit verschiedenen Vorzeichen an, so verschwindet sie in (wenigstens) einem Punkte zwischen diesen Punkten.*

Dieser Satz liegt der unter dem Namen Tabellenverfahren[2]) bekannten Näherungsmethode zum Auflösen von Gleichungen zugrunde und wird durch ein ähnliches Verfahren bewiesen.

[1]) Vgl. § 42. Der Begriff der Stetigkeit ist offenbar unabhängig davon, in welcher Form die Funktion gegeben ist.

[2]) Vgl. EdEM, Band II, S. 229.

Beweis. Die Funktion $f(x)$ sei im Intervall I stetig, und es seien a und b $(a < b)$ Punkte dieses Intervalls, in denen die Funktionen $f(x)$ Werte mit verschiedenen Vorzeichen besitzt. Um etwas Bestimmtes vor Augen zu haben, wollen wir annehmen, daß $f(a) < 0$ und $f(b) > 0$. Wir betrachten nun den Punkt $\frac{a+b}{2}$, d. h. den Mittelpunkt des Intervalls $a \leqq x \leqq b$. Es sind dann folgende Fälle möglich: $f\left(\frac{a+b}{2}\right) = 0$ oder $f\left(\frac{a+b}{2}\right) > 0$ oder $f\left(\frac{a+b}{2}\right) < 0$. Im ersten Fall sind wir fertig. Im zweiten Fall setzen wir $a_1 = a$ und $b_1 = \frac{a+b}{2}$, während wir im dritten Fall $a_1 = \frac{a+b}{2}$ und $b_1 = b$ setzen. Dann ist offenbar $f(a_1) < 0$, $f(b_1) > 0$ mit $a \leqq a_1 < b_1 \leqq b$ und $b_1 - a_1 = \frac{b-a}{2}$. In entsprechender Weise konstruieren wir (ausgehend von a_1, b_1) Punkte a_2, b_2 mit $f(a_2) < 0$ $f(b_2) > 0$, $a \leqq a_1 \leqq a_2 < b_2 \leqq b_1 \leqq b$ und $b_2 - a_2 = \frac{b_1 - a_1}{2} = \frac{b-a}{4}$, usw. Die auf diese Weise entstehenden Folgen $\{a_n\}$ und $\{b_n\}$ sind beide monoton und beschränkt, so daß beide (vgl. § 44) einen Grenzwert besitzen. Wegen $b_n - a_n = \frac{b-a}{2^n} \to 0$ besitzen aber beide Folgen sogar denselben Grenzwert, den wir mit ξ bezeichnen wollen. Wegen der Stetigkeit der Funktion $f(x)$ folgt aus

$$a_n \to \xi \quad \text{und} \quad b_n \to \xi,$$

daß

$$f(a_n) \to f(\xi) \quad \text{und} \quad f(b_n) \to f(\xi).$$

Aus $f(a_n) < 0$ $(n = 1, 2, \ldots)$ folgt $f(\xi) \leqq 0$ (vgl. § 38, Seite 162), während sich aus $f(b_n) > 0$ entsprechend $f(\xi) \geqq 0$ ergibt. Beides zusammen liefert

$$f(\xi) = 0.$$

Folgerung. *Eine in einem Intervall stetige Funktion $f(x)$ nimmt dort mit zwei Werten A und B auch alle zwischen A und B gelegenen Werte an*; genauer: *Ist $f(a) = A$ und $f(b) = B$, wobei $a < b$ und $A \neq B$, so gibt es für jede zwischen A und B gelegene Zahl X wenigstens ein ξ derart, daß $a < \xi < b$ und $f(\xi) = X$.*

Zum Beweis dieser Behauptung genügt es, den bereits bewiesenen Satz auf die Funktion

$$f_1(x) \equiv f(x) - X$$

anzuwenden, die in demselben Intervall stetig ist wie $f(x)$.

Diese Eigenschaft der stetigen Funktionen charakterisiert noch deutlicher als die eigentliche Stetigkeitsdefinition den „Zusammenhang" des Kurvenverlaufs. Sie ist jedoch nicht der üblichen Stetigkeitsdefinition gleichwertig. Zum Beispiel erfüllt auch die Funktion

$$f(x) = \begin{cases} \sin\frac{1}{x} & \text{für} \quad x \neq 0, \\ 0 & \text{für} \quad x = 0, \end{cases}$$

die doch eine Unstetigkeitsstelle für $x = 0$ aufweist (vgl. § 41, Beispiel 4), die angegebene Bedingung.

Das soeben bewiesene Resultat geht bereits auf Bolzano (1817) zurück.

Mit Hilfe dieses Satzes von BOLZANO läßt sich leicht zeigen, daß es genau eine positive reelle Zahl gibt, deren Quadrat gleich 2 ist. Die Funktion $f(x) = x^2$ ist nämlich stetig (§ 43) und wächst im Intervall $0 \leqq x \leqq 2$ monoton von 0 bis 4. Sie muß daher in einen bestimmten Punkte ξ den Wert 2 annehmen, d. h., es gibt wenigstens ein ξ, für das

$$f(\xi) = \xi^2 = 2.$$

Es kann aber auch nur eine solche Zahl ξ geben. Würde es nämlich zwei derartige Zahlen ξ' und ξ'' ($\xi' \neq \xi''$) geben, so müßte eine von ihnen kleiner als die andere sein, etwa $\xi' < \xi''$. Hieraus ergäbe sich jedoch sofort: $\xi'^2 < \xi''^2$, im Widerspruch zu $\xi'^2 = \xi''^2 = 2$. Es gibt also eine und nur eine positive Zahl ξ mit $\xi^2 = 2$, die man üblicherweise mit $\sqrt{2}$ bezeichnet.

Entsprechend beweist man die Existenz aller anderen (irrationalen) Wurzeln wie auch der Nullstellen von Polynomen. Genauso kann man dann (ausgehend von der Stetigkeit der Exponentialfunktionen) die Existenz der Logarithmen beweisen. Allgemein kann man den Satz von BOLZANO dazu verwenden, die Existenz der Umkehrfunktion einer beliebigen stetigen und monotonen Funktion zu beweisen (vgl. § 52) usw.

Satz II. *Eine in einem abgeschlossenen Intervall stetige Funktion ist dort auch beschränkt.*

Beweis. Die Funktion $f(x)$ sei stetig im Intervall $a \leqq x \leqq b$. Zu zeigen ist unter anderem, daß es eine Zahl M gibt, so daß

$$f(x) \leqq M$$

für alle x mit $a \leqq x \leqq b$. Wir nehmen an, es gäbe keine derartige Zahl M. Dann gäbe es zu jedem noch so großen M ein $x = x_M$ aus dem betrachteten Intervall, für welches

$$f(x) > M$$

gilt. Es sei nun $\{M_n\}$ eine Folge von Zahlen mit $M_n \to \infty$. Zu jedem M_n wählen wir eine Zahl x_n aus dem betrachteten Intervall mit

$$f(x_n) > M_n.$$

Dann gilt

$$f(x_n) \to \infty. \tag{15}$$

Aus der beschränkten Folge $\{x_n\}$ können wir nun (vgl. § 38, Seite 157) eine konvergente Teilfolge $\{x_{p_n}\}$ auswählen:

$$x_{p_n} \to \xi.$$

Dann ist wegen $a \leqq x_{p_n} \leqq b$ (vgl. § 38) auch $a \leqq \xi \leqq b$. Da die Funktion $f(x)$ stetig ist, strebt die Folge $\{f(x_{p_n})\}$ gegen den endlichen Grenzwert $f(\xi)$:

$$f(x_{p_n}) \to f(\xi).$$

Andererseits folgt aus (15)

$$f(x_{p_n}) \to \infty.$$

Damit haben wir einen Widerspruch erhalten. Es gibt also eine Zahl M, welche die gewünschte Eigenschaft besitzt, die Funktion $f(x)$ ist im Intervall $a \leqq x \leqq b$ nach oben beschränkt. Entsprechend zeigt man, daß sie auch nach unten beschränkt ist.

Wir weisen darauf hin, daß der Satz nicht mehr gilt, wenn man die Voraussetzung über die Abgeschlossenheit des betrachteten Intervalls fallen läßt. Ein Gegenbeispiel liefert bereits die Funktion

$$f(x) = \frac{1}{x};$$

sie ist im Intervall $0 < x \leqq 1$ stetig, aber dort — wie man leicht sieht — nicht nach oben beschränkt.

Satz III (Weierstrass). *Eine in einem abgeschlossenen Intervall stetige Funktion nimmt dort sowohl einen größten als auch einen kleinsten Wert an.*[1])

Mit anderen Worten: Ist die Funktion $f(x)$ im Intervall $a \leqq x \leqq b$ stetig, so gibt es Zahlen ξ_1 und ξ_2 mit $a \leqq \xi_1 \leqq b$ und $a \leqq \xi_2 \leqq b$ derart, daß $f(\xi_2) \leqq f(x) \leqq f(\xi_1)$ für alle x mit $a \leqq x \leqq b$ (wobei natürlich im allgemeinen weder ξ_1 noch ξ_2 eindeutig bestimmt ist).

Beweis. Wir beweisen als Beispiel die Existenz von ξ_1; die Existenz von ξ_2 läßt sich dann entsprechend zeigen.

Nach Satz II ist die Funktion $f(x)$ im Intervall $a \leqq x \leqq b$ beschränkt, d. h., es gibt Zahlen m und M, so daß für alle x aus dem betrachteten Intervall gilt:

$$m < f(x) < M.$$

Offenbar können wir nun unter den rationalen Zahlen, die eine Dezimalbruchdarstellung der Form g, a_1 besitzen, zwei aufeinanderfolgende y_1' und y_1'' $(y_1' < y_1'')$ so finden, daß für alle x aus dem betrachteten Intervall

$$f(x) < y_1'',$$

daß aber nicht für alle derartigen x

$$f(x) < y_1'$$

gilt.

Insbesondere gibt es dann im Intervall $a \leqq x \leqq b$ eine Zahl $x = x_1$, so daß

$$y_1' \leqq f(x_1) < y_1''.$$

Entsprechend findet man unter den rationalen Zahlen, die eine Dezimalbruchdarstellung der Form $g, a_1 a_2$ besitzen, zwei aufeinanderfolgende y_2' und y_2'' $(y_2' < y_2'')$ derart, daß

$$f(x) < y_2''$$

für alle, aber

$$f(x) < y_2'$$

nicht für alle x aus dem betrachteten Intervall gilt. Insbesondere gibt es eine Zahl $x = x_2$, so daß

$$y_2' \leqq f(x_2) < y_2'',$$

usw.

---

[1]) Ein größter Wert einer Funktion $f(x)$ (in einem Intervall) ist ein Funktionswert, der von keinem anderen Funktionswert übertroffen wird; ein kleinster Funktionswert ist ein Funktionswert, welcher von keinem anderen Funktionswert unterboten wird.

Die auf diese Weise gewonnenen Folgen $\{y_n'\}$ und $\{y_n''\}$ sind beide monoton und beschränkt, so daß jede von ihnen einen Grenzwert besitzt. Aus

$$y_n'' - y_n' = 10^{-n} \to 0 \tag{16}$$

folgt, daß diese Grenzwerte übereinstimmen. Bezeichnen wir den gemeinsamen Grenzwert der Folgen $\{y_n'\}$ und $\{y_n''\}$ mit η, so gilt also

$$y_n' \to \eta \tag{17}$$

und

$$y_n'' \to \eta. \tag{18}$$

Nach Konstruktion ist nun für jede natürliche Zahl n und für jedes x aus dem Intervall $[a, b]$

$$f(x) < y_n''.$$

Hieraus ergibt sich für $n \to \infty$

$$f(x) \leqq \eta.$$

Wir werden nun zeigen, daß es eine Zahl $\xi_1 \equiv \xi$ aus dem betrachteten Intervall gibt, für die

$$f(\xi) = \eta.$$

Nach Konstruktion gibt es zu jeder natürlichen Zahl n eine Zahl x_n aus dem Intervall $[a, b]$, so daß

$$y_n' \leqq f(x_n) < y_n''$$

oder, wenn man (16) beachtet,

$$f(x_n) < y_n'' \leqq f(x_n) + 10^{-n}.$$

Aus der beschränkten Folge $\{x_n\}$ können wir nun eine konvergente Teilfolge $\{x_{p_n}\}$ auswählen, und zwar gelte

$$x_{p_n} \to \xi \qquad (a \leqq \xi \leqq b).$$

Offenbar ist dann auch

$$f(x_{p_n}) < y_{p_n}'' \leqq f(x_{p_n}) + 10^{-p_n}.$$

In dieser Ungleichung vollziehen wir nun den Grenzübergang (für $n \to \infty$). Wegen der Stetigkeit der Funktion $f(x)$ im Punkte $x = \xi$ gilt einerseits

$$f(x_{p_n}) \to f(\xi).$$

Andererseits gilt wegen (18)

$$y_{p_n}'' \to \eta.$$

Damit erhalten wir

$$f(\xi) \leqq \eta \leqq f(\xi),$$

also

$$f(\xi) = \eta,$$

was zu beweisen war.

Der Satz von Weierstrass ist nicht mehr allgemein richtig, wenn man auf die Voraussetzung der Abgeschlossenheit des betrachteten Intervalls verzichtet. Als Gegenbeispiel kann wieder die Funktion $f(x) = \frac{1}{x}$ im Intervall $0 < x \leqq 1$ dienen. Ein noch einfacheres Gegenbeispiel liefert die Funktion $f(x) = x$ im Intervall $0 \leqq x < 1$.

Der Satz ist auch nicht mehr allgemein richtig, wenn man die Voraussetzung der Stetigkeit von $f(x)$ fallen läßt. Ein Gegenbeispiel liefert die Funktion

$$f(x)=\begin{cases} x & \text{für} \quad 0 \leqq x<1, \\ 0 & \text{für} \quad x=1. \end{cases}$$

Diese Funktion besitzt eine Unstetigkeit im Punkte $x=1$. Im abgeschlossenen Intervall $0 \leqq x \leqq 1$ nimmt sie keinen größten Wert an.

Satz IV (Satz von der „gleichmäßigen" Stetigkeit). *Ist die Funktion $f(x)$ stetig im abgeschlossenen Intervall I, so gibt es zu jedem $\varepsilon>0$ eine Zahl $\delta=\delta_\varepsilon>0$ derart, daß für alle x' und x'' aus dem Intervall I die Ungleichung*

$$|x'-x''|<\delta \tag{19}$$

die Ungleichung

$$|f(x')-f(x'')|<\varepsilon \tag{20}$$

zur Folge hat.

Wir überlassen es dem Leser, sich den geometrischen Inhalt dieses Satzes zu veranschaulichen.

Beweis. Wir nehmen an, der genannte Satz wäre nicht richtig, d. h., man könnte nicht zu jeder Zahl ε eine Zahl δ mit der erwähnten Eigenschaft finden. Dann gäbe es eine gewisse positive Zahl ε^*, so daß es zu jeder noch so kleinen Zahl $\delta>0$ zwei Zahlen x' und x'' aus I gibt, derart daß zwar die Ungleichung

$$|x'-x''|<\delta,$$

aber nicht die Ungleichung

$$|f(x')-f(x'')|<\varepsilon^*$$

erfüllt ist, daß also die Ungleichung

$$|f(x')-f(x'')| \geqq \varepsilon^*$$

besteht. Wir betrachten dann eine Folge $\{\delta_n\}$ von positiven Zahlen mit

$$\delta_n \to 0$$

und bestimmen zu jedem δ_n ein Zahlenpaar x_n', x_n'', für welches

$$|x_n'-x_n''|<\delta_n \tag{21}$$

und

$$|f(x_n')-f(x_n'')| \geqq \varepsilon^* \tag{22}$$

gilt. Aus der beschränkten Folge $\{x_n'\}$ wählen wir sodann eine konvergente Teilfolge $\{x_{p_n}'\}$ aus, und zwar gelte

$$x_{p_n}' \to \xi. \tag{23}$$

Die Zahl ξ gehört dabei, wie man sofort sieht, zu I. Aus (21) und (23) folgt, daß wegen $\delta_n \to 0$ auch

$$x_{p_n}'' \to \xi. \tag{24}$$

Wegen der Stetigkeit der Funktion $f(x)$ im Intervall I folgt aus (23) und (24)

$$f(x'_{p_n}) \to f(\xi),$$

$$f(x''_{p_n}) \to f(\xi).$$

Aus (22) folgt andererseits, daß für jedes n

$$\left|f(x_{p_n}) - f(x''_{p_n})\right| \geqq \varepsilon^*$$

gilt. Vollziehen wir in dieser Ungleichung den Grenzübergang $n \to \infty$, so erhalten wir einen Widerspruch, denn die linke Seite strebt gegen Null, während die rechte Seite stets gleich der positiven Zahl ε^* ist.

Bemerkung 1. Der wesentliche Unterschied zwischen der „gleichmäßigen" und der „gewöhnlichen" Stetigkeit besteht darin, daß im Fall der „gewöhnlichen" Stetigkeit in einem Punkte $x = x_0$ die Zahl δ in der Ungleichung $|x - x_0| < \delta$ nicht nur von der Zahl ε in der Ungleichung $|f(x) - f(x_0)| < \varepsilon$, sondern auch von der Zahl x_0 abhängt:

$$\delta = \delta(\varepsilon, x_0),$$

während bei der „gleichmäßigen" Stetigkeit in einem gegebenen Intervall die Zahl δ in der Ungleichung (19) nur von der Zahl ε aus (20) und von dem Intervall selbst abhängt (und nicht von Punkt zu Punkt des Intervalls variiert).

Bemerkung 2. Aus der gleichmäßigen Stetigkeit einer Funktion in einen gegebenen (abgeschlossenen oder nicht abgeschlossenen) Intervall I folgt offenbar sofort ihre Stetigkeit (im gewöhnlichen Sinne) in jedem Punkte x_0 des Intervalls I. Gilt nämlich die Ungleichung (20) für irgendwelche x' und x'', welche der Beziehung (19) genügen, so kann man insbesondere $x'' = x_0$ setzen und erhält, daß für alle $x = x'$, die der Ungleichung

$$|x - x_0| < \delta$$

genügen, die Ungleichung

$$|f(x) - f(x_0)| < \varepsilon$$

erfüllt ist, wie dies gerade bei der Definition der „gewöhnlichen" Stetigkeit gefordert wird.

Bemerkung 3. Satz IV ist nicht mehr allgemein richtig, wenn man die Voraussetzung der Abgeschlossenheit des Intervalls I fallen läßt. Man sieht dies wieder z. B. an der Funktion $f(x) = \dfrac{1}{x}$ im Intervall $0 < x \leqq 1$. Setzt man hier $\varepsilon = 1$, $x' = \delta$, $x'' = \dfrac{\delta}{2}$, so erhält man für $\delta < 1$

$$|x' - x''| = \frac{\delta}{2} < \delta,$$

während

$$|f(x') - f(x'')| = \left|\frac{1}{\delta} - \frac{2}{\delta}\right| = \frac{1}{\delta} > 1.$$

Damit haben wir erhalten: *Eine Funktion $f(x)$ ist in einem abgeschlossenen Intervall I dann und nur dann gleichmäßig stetig, wenn sie dort stetig ist.* Im Falle eines nicht abgeschlossenen Intervalls ist dies dagegen nicht richtig.

§ 48. Gleichmäßige Konvergenz von Folgen stetiger Funktionen

In § 45 haben wir den Begriff des Grenzwertes einer Funktionenfolge mit Hilfe des Begriffs des Grenzwertes von Zahlenfolgen definiert. Dabei haben wir verlangt, daß für jedes x aus dem betrachteten Intervall die Folge $\{f_n(x)\}$ konvergiert; der Grenzwert der jeweiligen Zahlfolge wurde sodann als der entsprechende Funktionswert der Grenzfunktion angesehen.

Man kann an dieselbe Frage aber auch von einem anderen Standpunkt aus herangehen, indem man nämlich die Funktionen der gegebenen Funktionenfolge im betrachteten Intervall „als Ganzes" ansieht. Im folgenden wird eine der hierbei möglichen Betrachtungsweisen genauer diskutiert. Der Einfachheit halber setzen wir voraus, daß wir es zunächst nur mit Funktionen zu tun haben, die im betrachteten Intervall stetig sind.[1])

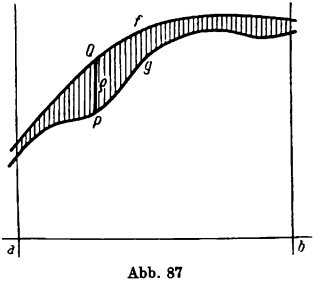


Abb. 87

Es seien $f(x)$ und $g(x)$ zwei in einem gegebenen abgeschlossenen (und damit endlichen) Intervall I $(a \leqq x \leqq b)$ stetige Funktionen. Unsere erste Aufgabe soll es sein, den „Abstand" der einen Funktion von der anderen (oder des graphischen Bildes der einen vom graphischen Bild der anderen) durch eine Zahl zu charakterisieren. Bekanntlich versteht man unter dem „Abstand" einer Zahl a von einer Zahl b (vgl. § 36) den absoluten Betrag ihrer Differenz, d. h. die Zahl $|a-b|$. Demgegenüber nehmen wir als Maß für den „Abstand" zweier Funktionen (oder kurz als „Abstand" der Funktionen) das Maximum (den größten Wert) des absoluten Betrages der Differenz der Funktionswerte im betrachteten Intervall:

$$\varrho(f, g) \equiv \varrho_I(f, g) = \max_I |f(x) - g(x)|. \tag{25}$$

Geometrisch ist der Abstand $\varrho(f, g)$, wie man leicht einsieht, die maximale Länge der durch die Punkte des graphischen Bildes von f bzw. g mit gleichen Abszissen gelegten Strecken (Abb. 87). Da die Funktionen $f(x)$ und $g(x)$ stetig sind, ist es auch die Differenz $f(x) - g(x)$ und damit auch der absolute Betrag davon, d. h. die Funktion $|f(x) - g(x)|$;[2]) daher gibt es (nach Satz III des vorangehenden Paragraphen) unter allen in Betracht kommenden Strecken wirklich eine größte.

---

[1]) Diese Einschränkung ist nur teilweise sachlich begründet; sie wird jedoch dadurch gerechtfertigt, daß es gemäß dem Vorhaben dieses Artikels lediglich darauf ankommt, jeweils die einfachsten und wichtigsten Fälle zu behandeln.

[2]) Wir überlassen es dem Leser, allgemein den folgenden Satz zu beweisen: *Ist die Funktion $F(x)$ im Punkte $x = c$ stetig, so ist es dort auch die Funktion $|F(x)|$.*

Der auf diese Weise definierte Abstand $\varrho(f, g)$ besitzt die folgende Eigenschaft:

Für jede Zahl x aus dem Intervall I ist

$$|f(x)-g(x)| \leqq \varrho(f, g), \tag{26}$$

wobei für wenigstens einen Wert x aus I Gleichheit eintritt.

Der Abstand $\varrho(f, g)$ ist nicht negativ,

$$\varrho(f, g) \geqq 0; \tag{27}$$

er ist dann und nur dann gleich Null, wenn im betrachteten Intervall die Funktionen $f(x)$ und $g(x)$ übereinstimmen (ihre graphischen Bilder zusammenfallen).

Man nennt nun eine Folge

$$\{f_n(x)\} \equiv f_1(x),\ f_2(x),\ \ldots,\ f_n(x),\ \ldots$$

von Funktionen, die alle in einem gegebenen abgeschlossenen Intervall I stetig sind, in diesem Intervall *gleichmäßig konvergent* gegen die stetige Funktion $f(x)$, wenn die Zahlenfolge

$$\{\varrho(f_n, f)\} \equiv \varrho(f_1, f),\ \varrho(f_2, f),\ \ldots,\ \varrho(f_n, f),\ \ldots$$

gegen Null konvergiert:

$$\varrho(f_n, f) \to 0. \tag{28}$$

Die gleichmäßige Konvergenz einer Funktionenfolge deutet man üblicherweise durch

$$f_n \Rightarrow f \tag{29}$$

an.

Satz. *Aus der gleichmäßigen Konvergenz* $(f_n \Rightarrow f)$ *folgt die Konvergenz im gewöhnlichen Sinne* $(f_n \to f)$.

In der Tat: Die mit (29) gleichwertige Beziehung (28) bedeutet doch, daß bei noch so kleinem $\varepsilon > 0$ für alle hinreichend großen n die Ungleichung

$$\varrho(f_n, f) < \varepsilon \tag{30}$$

erfüllt ist. Es sei nun x eine beliebige Zahl aus dem Intervall I. Wegen (26) gilt dann

$$|f_n(x)-f(x)| \leqq \varrho(f_n, f) \tag{31}$$

und damit

$$|f_n(x)-f(x)| < \varepsilon.$$

Es gilt also für jedes x aus I

$$f_n(x) \to f(x).$$

(Der Beweis läßt sich folgendermaßen kurz zusammenfassen: Wegen $f_n \Rightarrow f$ strebt für $n \to \infty$ die größte der „Strecken" $|f_n(x)-f(x)|$ und damit auch jede einzelne „Strecke" gegen Null.)

Die folgenden Beispiele zeigen, daß man umgekehrt aus der gewöhnlichen Konvergenz im allgemeinen nicht auf gleichmäßige Konvergenz schließen kann.

Beispiel 1.

$$f_n(x) = \frac{1}{1 + n^2\left(x - \frac{1}{\sqrt{n}}\right)^2} \qquad (n = 1, 2, 3, \ldots), \tag{32}$$

wobei wir uns auf das Intervall

$$0 \leqq x \leqq 1$$

beschränken wollen. Wir erhalten:

1. Für $x = 0$

$$f_n(0) = \frac{1}{1+n} \to 0;$$

2. für $x \neq 0$

$$f_n(x) = \frac{1}{n^2 x^2} \cdot \frac{1}{\left(1 - \frac{1}{x\sqrt{n}}\right)^2 + \frac{1}{n^2 x^2}} \to 0.$$

Daher gilt für jedes x

$$f_n(x) \to 0,$$

so daß die Grenzfunktion $f(x)$ identisch verschwindet. Jedoch liegt keine gleichmäßige Konvergenz vor, da

$$\varrho(f_n, f) = \max_x |f_n(x) - f(x)| = \max_x f_n(x) = 1.^{1})$$

Es ist nützlich, sich die Verhältnisse geometrisch zu veranschaulichen; dabei beachte man, daß das graphische Bild der Funktion $f_n(x)$ das gleiche ist, wie das der in § 45, Beispiel 3, betrachteten Funktion, nur ist es um $\frac{1}{\sqrt{n}}$ nach rechts verschoben.

Im folgenden Beispiel sind die Funktionen der Folge zwar keine elementaren Funktionen, jedoch lassen sie sich geometrisch einfacher übersehen als die im vorangehenden Beispiel betrachteten Funktionen.

Beispiel 2. Im Intervall $0 \leqq x \leqq 1$ seien die Funktionen $f_n(x)$ gegeben durch

$$f_n(x) = \begin{cases} nx & \text{für } 0 \leqq x \leqq \frac{1}{n}, \\ 2 - nx & \text{für } \frac{1}{n} \leqq x \leqq \frac{2}{n}, \\ 0 & \text{für } \frac{2}{n} \leqq x \leqq 1 \end{cases} \tag{33}$$

$(n = 1, 2, 3, \ldots)$

(in Abb. 88 ist die Funktion f_{10} dargestellt). Dann gilt:

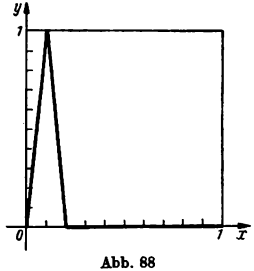


Abb. 88

1. Für $x = 0$

$$f_n(0) = 0 \to 0;$$

[1]) Aus (32) ersieht man sofort, daß die Funktion $f_n(x)$ für $x_n = \frac{1}{\sqrt{n}}$ ihren größten Wert $f_n(x_n) = 1$ annimmt.

2. für $0 < x \leqq 1$

$$f_n(x) = 0, \quad \text{falls} \quad n \geqq \frac{2}{x},$$

so daß auch in diesem Fall

$$f_n(x) \to 0.$$

Somit gilt also im gesamten Intervall $0 \leqq x \leqq 1$

$$f_n(x) \to f(x) \equiv 0.$$

Andererseits ist

$$\varrho(f_n, f) = \max f_n(x) = f_n\left(\frac{1}{n}\right) = 1.$$

Daher ist die Konvergenz $f_n \to f$ keine gleichmäßige Konvergenz.

Der geometrische Inhalt der gleichmäßigen Konvergenz ist leicht zu erkennen: Die Beziehung (29) bedeutet doch, daß bei noch so kleinem $\varepsilon > 0$ für alle hinreichend großen natürlichen Zahlen n die Ungleichung

$$\varrho(f_n, f) < \varepsilon$$

erfüllt ist. Diese Ungleichung ist jedoch auf Grund der oben angegebenen Eigenschaft des Abstandes von Funktionen gleichwertig dem gleichzeitigen Bestehen der unendlich vielen Ungleichungen

$$|f_n(x) - f(x)| < \varepsilon \qquad (34)$$

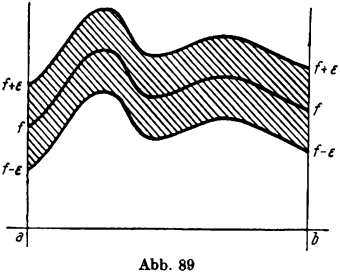


Abb. 89

(wobei x eine beliebige Zahl aus dem Intervall I ist).

Geben wir der Ungleichung (34) die Form

$$f(x) - \varepsilon < f_n(x) < f(x) + \varepsilon,$$

so können wir die gleichmäßige Konvergenz

$$f_n \Rightarrow f$$

auch folgendermaßen beschreiben: Wir betrachten zunächst das graphische Bild der Funktion $f(x) - \varepsilon$ und das der Funktion $f(x) + \varepsilon$, welches man aus dem von $f(x)$ durch Parallelverschiebung um die Strecke ε nach unten bzw. nach oben erhält (Abb. 89). Die gleichmäßige Konvergenz $f_n \Rightarrow f$ besagt dann, daß für hinreichend großes n ($n > n_0$) das graphische Bild der Funktion $f_n(x)$ in dem zwischen diesen beiden graphischen Darstellungen gelegenen Streifen enthalten ist. (Bezeichnen wir diesen Streifen als „ε-Umgebung" des graphischen Bildes der Funktion $f(x)$, so können wir auch sagen, daß das graphische Bild „fast jeder" Funktion $f_n(x)$ ganz in dieser ε-Umgebung gelegen ist, d. h. nur endlich viele von ihnen sich nicht ganz darin befinden; vgl. hierzu die Definition des Grenzwertes einer Zahlenfolge auf Seite 154.)

Oben wurde die gleichmäßige Konvergenz $f_n \Rightarrow f$ nur für den Fall definiert, daß sowohl alle Funktionen der Folge $\{f_n\}$ als auch die Grenzfunktion $f(x)$ in dem betrachteten Intervall stetig sind. Zu dieser Einschränkung liegt jedoch keine logische Notwendigkeit vor. Aus der Definition der gleichmäßigen Konvergenz ergibt sich nämlich unmittelbar, daß *genau dann* $f_n \Rightarrow f$, *wenn es zu jedem* $\varepsilon > 0$ *eine natürliche Zahl* n *gibt* (welche von ε, aber nicht von x abhängt) *derart, daß für jedes* x *aus dem Intervall* I *die Ungleichung* (34) *erfüllt ist.* Diese Aussage nimmt man gewöhnlich als Definition der gleichmäßigen Konvergenz, wobei man dann weder die Stetigkeit der Funktionen $f_n(x)$ noch die der Funktion $f(x)$ voraussetzt.

Für die so verstandene gleichmäßige Konvergenz gilt der

Satz. *Wenn eine Folge* $\{f_n\}$ *von in einem gegebenen Intervall* I *stetigen Funktionen dort gleichmäßig gegen eine Funktion* $f(x)$ *konvergiert, so ist diese Funktion* $f(x)$ *im Intervall* I *ebenfalls stetig.*

Beweis. Es sei c eine beliebige Zahl aus dem Intervall I und ε eine beliebige positive Zahl. Aus der Identität

$$f(x)-f(c)=[f(x)-f_n(x)]+[f_n(x)-f_n(c)]+[f_n(c)-f(c)]$$

ergibt sich sofort die Ungleichung

$$|f(x)-f(c)| \leqq |f(x)-f_n(x)|+|f_n(x)-f_n(c)|+|f_n(c)-f(c)|. \tag{35}$$

Hier wählen wir nun a) n so groß, daß für alle x aus I (insbesondere also für $x=c$)

$$|f(x)-f_n(x)|<\frac{\varepsilon}{3}$$

gilt, b) $\delta>0$ so klein, daß für eines der in Frage kommenden n für $|x-c|<\delta$ die Ungleichung

$$|f_n(x)-f_n(c)|<\frac{\varepsilon}{3}$$

erfüllt ist. Dadurch wird erreicht, daß jeder Summand auf der rechten Seite von (35) kleiner als $\frac{\varepsilon}{3}$ ist, so daß für alle x mit $|x-c|<\delta$ die Ungleichung

$$|f(x)-f(c)|<\varepsilon$$

erfüllt ist. Das besagt aber gerade, daß die Funktion $f(x)$ im Punkte c stetig ist. Dies gilt für jedes c aus I, so daß die Funktion $f(x)$ im ganzen Intervall I stetig ist, was zu beweisen war.

§ 49. Der Satz von Weierstrass-Bernstein über die Approximation stetiger Funktionen mit Hilfe von Polynomen

Eine fundamentale Aufgabe, welcher die Kapitel I und II des vorliegenden Artikels gewidmet waren und deren Lösung gleichermaßen Schüler, Studenten und Lehrerschaft interessiert, besteht in folgendem: Gegeben ist eine Gleichung, durch die eine veränderliche Größe y durch eine andere veränderliche Größe x

ausgedrückt wird; es soll die zugehörige Kurve dargestellt werden. Die Lösung dieser Aufgabe bereitet keine prinzipiellen Schwierigkeiten. Jedoch ist es wegen der Vielzahl der möglichen Fälle notwendig, technisch brauchbare Konstruktionsverfahren auszuarbeiten, und hierin besteht das eigentliche Problem.

Daneben interessiert aber sowohl den Praktiker als auch den Wissenschaftler häufig die umgekehrte Aufgabe, bei der Schwierigkeiten prinzipieller Art auftreten. Sie läßt sich kurz folgendermaßen formulieren: Gegeben ist eine graphische Darstellung; gesucht ist die „zugehörige" Gleichung.

Ein Beobachter z. B., welcher den Verlauf eines bestimmten Vorganges registriert, bei dem jeder „Zustand" durch die Werte zweier funktional abhängiger Größen (Veränderlicher oder Parameter) charakterisiert wird, erhält in der Koordinatenebene eine graphische Darstellung des beobachteten Vorganges in Form einer empirischen oder (wie man zu Zeiten EULERS sagte) „willkürlichen" Kurve; wenn man nun den beobachteten Vorgang mathematisch erfassen will, so muß man eine (genaue oder in einem gewissen Sinne näherungsweise) Gleichung für die erhaltene Kurve aufstellen.

Wie gelangt man nun zu einer solchen Gleichung?

Zunächst muß die Fragestellung genauer präzisiert werden, was auf mehrere Weisen möglich ist.

Wir wollen hier annehmen, daß die gegebene Kurve in den Grenzen des betrachteten Intervalls $I\,(a \leqq x \leqq b)$ mit jeder vertikalen Geraden nur einen Punkt gemeinsam hat, so daß man die Ordinate y jedes Kurvenpunktes als (eindeutige) Funktion seiner Abszisse x ansehen kann,

$$y = f(x)$$

(Funktion hierbei als „Zuordnung" aufgefaßt). Weiter setzen wir voraus, daß die Funktion $f(x)$ im genauen mathematischen Sinne stetig ist (vgl. § 42).[1])

Mitunter geht man nun so vor, daß man auf der vorgelegten Kurve endlich viele Punkte

$$M_i(x_i,\ y_i) \qquad (i = 1,\ 2,\ \ldots,\ m)$$

markiert und eine elementare stetige Funktion (meistens ein Polynom $P(x)$) bestimmt, deren graphisches Bild genau durch die markierten Punkte hindurchgeht (wobei man auf ein genaues Übereinstimmen der Kurven in den Zwischenpunkten verzichtet). Das Aufsuchen einer derartigen Funktion bezeichnet man als *Interpolation*.

Eine andere Aufgabenstellung, und zwar die, welche uns hier interessieren soll, besteht darin, eine elementare Funktion (ein Polynom) $P(x)$ so zu be-

[1]) In dieser letzten Forderung ist ein idealisierendes Element enthalten, da sie bei einer empirisch gegebenen Kurve nur bedingt erfüllt ist. Streng erfüllt ist sie aber z. B. bei jeder stetigen Kurve, die aus stetigen Kurvenbögen „zusammengesetzt" ist (vgl. § 45, Beispiel 7 und 9, unter der Voraussetzung, daß $u(x)$ und $v(x)$ stetige Funktionen sind und dabei $u(0) = v(0) = 0$ ist).

stimmen, daß der Abstand der Funktion $f(x)$ von $P(x)$ in allen Punkten des betreffenden Intervalls kleiner als eine gegebene Zahl ε ist, d. h., daß

$$|P(x)-f(x)|<\varepsilon$$

für alle x aus I gilt, also (mit den Bezeichnungen aus § 48)

$$\varrho(P, f)<\varepsilon$$

ist. Dies bezeichnet man als *Approximation* von $f(x)$ durch $P(x)$.

In erster Linie interessiert hierbei natürlich, ob man für jedes beliebig kleine $\varepsilon>0$ zu der gegebenen Funktion $f(x)$ ein Polynom $P(x)$ finden kann, welches das Verlangte leistet.

Eine positive Antwort auf diese Frage gibt ein Satz von WEIERSTRASS (1885), welchen man in gewissem Sinne als Umkehrung des am Ende des vorangehenden Paragraphen bewiesenen Satzes ansehen kann. Hier ist seine genaue Formulierung:

Zu jeder Funktion $f(x)$, die in einem abgeschlossenen[1]) *Intervall I ($a \leqq x \leqq b$) stetig ist, läßt sich eine Folge $\{P_n(x)\}$ von Polynomen angeben, welche in I gleichmäßig gegen $f(x)$ konvergiert.*

Die Aufgabe, eine gegebene Kurve durch eine „Gleichung" zu beschreiben, besitzt also die genaue Lösung

$$f(x)=\lim_{n\to\infty} P_n(x)$$

und die Näherungslösung

$$f(x)\approx P_n(x),$$

wobei die Abweichung $|P_n(x)-f(x)|$ die Zahl $\varrho(P_n, f)=\max|P_n(x)-f(x)|$, welche von x unabhängig ist, nicht übertrifft und wobei $\varrho(P_n, f)$ für $n\to\infty$ gegen Null strebt.

Da die Spitze jedes Schreibgerätes (Bleistift, Feder, Kreide) einen gewissen Durchmesser besitzt, ist jedes „reale" graphische Bild (in Form der Spur, welche auf dem Papier bei Bewegung des Schreibgerätes markiert wird) keine „ideale mathematische Kurve", sondern ein „Streifen", welcher eine gewisse „Breite" besitzt. Aus diesem Grunde kann man den Satz von WEIERSTRASS auch in der folgenden, zwar recht ungenauen, dafür aber sehr anschaulichen Form aussprechen: Zu jedem noch so „gewundenen" vorgegebenen „realen" Kurvenverlauf, der durch „Bewegung" eines Schreibgerätes erzeugt wird (vgl. Abb. 89), läßt sich ein rationales Polynom finden, dessen graphisches Bild ganz im gezeichneten „Streifen" verläuft.

Zunächst bemerken wir, daß die Folge $\{P_n(x)\}$ nicht eindeutig bestimmt ist. Verschiedene Autoren haben Beweise für diesen Satz angegeben, bei denen verschiedene Verfahren zur Konstruktion der approximierenden Polynome verwendet wurden.

Die ersten Beweise für den Satz von WEIERSTRASS waren so kompliziert, daß sie nicht allgemein verstanden wurden. Wir führen im folgenden einen

[1]) Diese Voraussetzung ist wesentlich.

Beweis durch, der im wesentlichen mit dem von S. N. BERNSTEIN im Jahre 1912 angegebenen Beweis übereinstimmt. Obgleich dieser Beweis vollkommen elementar ist, empfiehlt es sich, zunächst einige Vorbereitungen zu treffen.

Der Einfachheit halber nehmen wir an, daß die Approximation der Funktion $f(x)$ im Intervall

$$0 \leqq x \leqq 1$$

erfolgen soll.

Die von BERNSTEIN verwendeten Polynome bezeichnet man heute üblicherweise mit $B_n(x)$. Sie haben die Form

$$B_n(x) = \sum_{m=0}^{n} f\left(\frac{m}{n}\right) \binom{n}{m} x^m (1-x)^{n-m}, \tag{36}$$

wobei unten $\binom{n}{m}$ der Binomialkoeffizient

$$\binom{n}{m} = \frac{n!}{m!\,(n-m)!}$$

verstanden wird (hierbei ist $0! = 1$ zu setzen).

Satz von BERNSTEIN. *Es sei $f(x)$ eine im abgeschlossenen Intervall $0 \leqq x \leqq 1$ stetige Funktion. Dann konvergiert die Folge $\{B_n(x)\}$ der Polynome (36) in diesem Intervall gleichmäßig gegen $f(x)$:*

$$B_n(x) \Rightarrow f(x). \tag{37}$$

Wir beweisen zunächst das folgende

Lemma. *Ist $p+q=1$, so ist*

$$S^{(n)} = \sum_{m=0}^{n} (m-np)^2 \binom{n}{m} p^m q^{n-m} = npq \qquad (n=0,\ 1,\ 2,\ \ldots).$$

Beweis. Zunächst berechnen wir die Summen

$$S_0^{(n)} = \sum_{m=0}^{n} \binom{n}{m} p^m q^{n-m},$$

$$S_1^{(n)} = \sum_{m=0}^{n} m \binom{n}{m} p^m q^{n-m},$$

$$S_2^{(n)} = \sum_{m=0}^{n} m(m-1) \binom{n}{m} p^m q^{n-m}.$$

Was die erste Summe betrifft, so ist auf Grund des binomischen Satzes

$$S_0^{(n)} = (p+q)^n,$$

also nach Voraussetzung

$$S_0^{(n)} = 1 \qquad (n=0,\ 1,\ 2,\ \ldots).$$

Zur Berechnung von $S_1^{(n)}$ beachten wir zunächst, daß

$$m\binom{n}{m} = n\binom{n-1}{m-1} \qquad (m \geqq 1, \quad n \geqq 1).$$

Hieraus folgt für $n \geqq 1$

$$\begin{aligned} S_1^{(n)} &= \sum_{m=0}^{n} m\binom{n}{m} p^m q^{n-m} = \sum_{m=1}^{n} m\binom{n}{m} p^m q^{n-m} \\ &= n \sum_{m=1}^{n} \binom{n-1}{m-1} p^m q^{n-m} \\ &= np \sum_{m=1}^{n} \binom{n-1}{m-1} p^{m-1} q^{(n-1)-(m-1)} \\ &= np \sum_{m=0}^{n-1} \binom{n-1}{m} p^m q^{(n-1)-m} = np \cdot S_0^{(n-1)} = np. \end{aligned}$$

Zur Berechnung von $S_2^{(n)}$ beachten wir, daß

$$m(m-1)\binom{n}{m} = n(n-1)\binom{n-2}{m-2} \qquad (m \geqq 2, \quad n \geqq 2).$$

Daher ist für $n \geqq 2$

$$\begin{aligned} S_2^{(n)} &= \sum_{m=0}^{n} m(m-1)\binom{n}{m} p^m q^{n-m} = \sum_{m=2}^{n} m(m-1)\binom{n}{m} p^m q^{n-m} \\ &= n(n-1) \sum_{m=2}^{n} \binom{n-2}{m-2} p^m q^{n-m} \\ &= n(n-1)p^2 \sum_{m=2}^{n} \binom{n-2}{m-2} p^{m-2} q^{(n-2)-(m-2)} \\ &= n(n-1)p^2 \sum_{m=0}^{n-2} \binom{n-2}{m} p^m q^{(n-2)-m} \\ &= n(n-1)p^2 \cdot S_0^{(n-2)} = n(n-1)p^2. \end{aligned}$$

Damit erhalten wir

$$S_0^{(n)} = 1, \quad S_1^{(n)} = np, \quad S_2^{(n)} = n(n-1)p^2 \qquad (n = 0, 1, 2, \ldots).^{1})$$

Um nun $S^{(n)}$ zu berechnen, gehen wir von der Identität

$$(m - np)^2 = n^2p^2 - (2np - 1)m + m(m-1)$$

---

[1]) Daß diese Formeln auch für $S_1^{(0)}$, $S_2^{(0)}$ und $S_2^{(1)}$ richtig sind, bestätigt man leicht direkt.

aus. Mit ihrer Hilfe und wegen der bereits hergeleiteten Formeln erhalten wir (für $n \geqq 0$)

$$\begin{aligned} S^{(n)} &= \sum_{m=0}^{n} (m-np)^2 \binom{n}{m} p^m q^{n-m} \\ &= \sum_{m=0}^{n} \{n^2p^2 - (2np-1)\,m + m(m-1)\} \binom{n}{m} p^m q^{n-m} \\ &= n^2p^2 \sum_{m=0}^{n} \binom{n}{m} p^m q^{n-m} - (2np-1) \sum_{m=0}^{n} m \binom{n}{m} p^m q^{n-m} \\ &\qquad + \sum_{m=0}^{n} m(m-1) \binom{n}{m} p^m q^{n-m} \\ &= n^2p^2 \cdot S_0^{(n)} - (2np-1)\, S_1^{(n)} + S_2^{(n)} \\ &= n^2p^2 \cdot 1 - (2np-1)\, np + n(n-1)\, p^2 \\ &= np(1-p) = npq. \end{aligned}$$

§ 50. Beweis des Satzes von Bernstein

Wir setzen zunächst zur Abkürzung

$$u_m^{(n)}(x) = \binom{n}{m} x^m (1-x)^{n-m} \qquad (m = 0,\ 1,\ \ldots,\ n;\quad n = 0,\ 1,\ 2,\ \ldots),$$

so daß also

$$B_n(x) = \sum_{m=0}^{n} f\left(\frac{m}{n}\right) u_m^{(n)}(x). \tag{38}$$

Ersetzen wir in

$$S_0^{(n)} = \sum_{m=0}^{n} \binom{n}{m} p^m q^{n-m} = 1$$

p durch x und q durch $1-x$ (so daß also $p+q=1$), so erhalten wir die folgende Identität in x:

$$\sum_{m=0}^{n} u_m^{(n)}(x) = 1. \tag{39}$$

Hieraus folgt

$$f(x) = \sum_{m=0}^{n} f(x)\, u_m^{(n)}(x). \tag{40}$$

Subtrahieren wir nun (40) von (38), so erhalten wir

$$B_n(x) - f(x) = \sum_{m=0}^{n} \left[f\left(\frac{m}{n}\right) - f(x) \right] u_m^{(n)}(x).$$

Daher ist (auf Grund einer bekannten Eigenschaft des absoluten Betrages)

$$|B_n(x)-f(x)| \leqq \sum_{m=0}^{n} \left|f\left(\frac{m}{n}\right)-f(x)\right| u_m^{(n)}(x).^{1}) \tag{41}$$

Wir haben uns nun nur davon zu überzeugen, daß für hinreichend großes n die rechte Seite von (41) kleiner als jede beliebige Zahl $\varepsilon > 0$ wird. Hierbei werden wir maßgeblich die Stetigkeit der Funktion $f(x)$ auszunutzen haben.

Da die Funktion $f(x)$ im Intervall $0 \leqq x \leqq 1$ stetig ist, ist sie dort beschränkt (vgl. § 47, Satz II) und gleichmäßig stetig (§ 47, Satz IV). Das bedeutet:

1. Es gibt eine Zahl M, so daß

$$|f(x)| \leqq M \qquad (0 \leqq x \leqq 1). \tag{42}$$

2. Zu jeder vorgegebenen Zahl $\varepsilon > 0$ kann man eine Zahl $\delta > 0$ so finden, daß für jedes x' und x'' aus dem Intervall $0 \leqq x \leqq 1$ aus der Ungleichung

$$|x'-x''| < \delta$$

die Ungleichung

$$|f(x')-f(x'')| < \frac{\varepsilon}{2} \tag{43}$$

folgt.

Wir wenden uns nun der „Abschätzung" der rechten Seite von (41) zu. Hierfür zerlegen wir die dort stehende Summe in zwei Summanden, und zwar auf folgende Weise: Es sei x eine beliebige Zahl aus dem Intervall $0 \leqq x \leqq 1$. Bei der Summation in (41), bei der m von 0 bis n läuft, ist, wenn n genügend groß ist, für gewisse m die Ungleichung

$$\left|\frac{m}{n}-x\right| < \delta \tag{44}$$

erfüllt, während sie für die übrigen m nicht erfüllt ist. In die erste der beiden Summen, in die wir die rechte Seite von (41) zerlegen wollen, nehmen wir nun alle diejenigen Glieder auf, für welche die Ungleichung (44) erfüllt ist, während der zweite Summand aus allen den Gliedern besteht, für die dies nicht der Fall ist. Die auf diese Weise gewonnenen Teilsummen sollen durch den Index I bzw. II angedeutet werden[2]), so daß

$$\begin{aligned}\sum_{m=0}^{n} \left|f\left(\frac{m}{n}\right)-f(x)\right| u_m^{(n)}(x) \\ = \sum\nolimits_{\mathrm{I}} \left|f\left(\frac{m}{n}\right)-f(x)\right| u_m^{(n)}(x) + \sum\nolimits_{\mathrm{II}} \left|f\left(\frac{m}{n}\right)-f(x)\right| u_m^{(n)}(x).\end{aligned} \tag{45}$$

Die Teilsummen werden nun einzeln untersucht.

---

[1]) Offenbar ist $u_m^{(n)}(x) \geqq 0$ für $0 \leqq x \leqq 1$.

[2]) Diese Zerlegung hängt selbstverständlich noch von x ab, jedoch soll dies nicht explizit berücksichtigt werden.

Im ersten Summanden treten nur solche Zahlen m auf, für die die Ungleichung (44) erfüllt ist, so daß wegen 2. für alle diese m die Ungleichung

$$\left|f\left(\frac{m}{n}\right)-f(x)\right|<\frac{\varepsilon}{2}$$

erfüllt ist. Daher ist

$$\Sigma_{\mathrm{I}}\left|f\left(\frac{m}{n}\right)-f(x)\right|u_m^{(n)}(x)\leqq\Sigma_{\mathrm{I}}\frac{\varepsilon}{2}u_m^{(n)}(x)=\frac{\varepsilon}{2}\Sigma_{\mathrm{I}}u_m^{(n)}(x). \tag{46}$$

Wir betrachten nun die Summe $\sum_{m=0}^{n}u_m^{(n)}(x)$; diese kann man ebenso wie die vorher betrachtete Summe in zwei Teilsummen zerlegen:

$$\sum_{m=0}^{n}u_m^{(n)}(x)=\Sigma_{\mathrm{I}}u_m^{(n)}(x)+\Sigma_{\mathrm{II}}u_m^{(n)}(x).$$

Da nun die Glieder $u_m^{(n)}(x)$ alle nicht-negativ sind und die Gesamtsumme wegen (39) den Wert 1 hat, können wir feststellen, daß

$$\Sigma_{\mathrm{I}}u_m^{(n)}(x)=1-\Sigma_{\mathrm{II}}u_m^{(n)}(x)\leqq 1. \tag{47}$$

Damit erhalten wir aus (46)

$$\Sigma_{\mathrm{I}}\left|f\left(\frac{m}{n}\right)-f(x)\right|u_m^{(n)}(x)\leqq\frac{\varepsilon}{2}. \tag{48}$$

Wir betrachten nun die zweite Teilsumme

$$\Sigma_{\mathrm{II}}\left|f\left(\frac{m}{n}\right)-f(x)\right|u_m^{(n)}(x).$$

Wegen

$$\left|f\left(\frac{m}{n}\right)-f(x)\right|\leqq\left|f\left(\frac{m}{n}\right)\right|+|f(x)|\leqq M+M=2M$$

(vgl. (42)) ist

$$\Sigma_{\mathrm{II}}\left|f\left(\frac{m}{n}\right)-f(x)\right|u_m^{(n)}(x)\leqq\Sigma_{\mathrm{II}}2Mu_m^{(n)}(x)=2M\Sigma_{\mathrm{II}}u_m^{(n)}(x). \tag{49}$$

Nach Voraussetzung ist nun für die Glieder von Σ_{II} die Ungleichung (44) nicht erfüllt, d. h., es gilt die Ungleichung

$$\left|\frac{m}{n}-x\right|\geqq\delta,$$

der man auch die Form

$$|m-nx|\geqq\delta n$$

oder

$$(m-nx)^2\geqq\delta^2n^2$$

oder schließlich

$$1\leqq\left(\frac{m-nx}{\delta n}\right)^2$$

geben kann. Multiplizieren wir beide Seiten der zuletzt erhaltenen Ungleichung mit $u_m^{(n)}(x)$ und summieren wir anschließend im Sinne von $\sum_{\text{II}}$, so erhalten wir

$$\sum_{\text{II}} u_m^{(n)}(x) \leqq \sum_{\text{II}} \left(\frac{m-nx}{\delta n}\right)^2 u_m^{(n)}(x) = \frac{1}{\delta^2 n^2} \sum_{\text{II}} (m-nx)^2 u_m^{(n)}(x). \tag{50}$$

Die Summe $\sum_{m=0}^{n} (m-nx)^2 u_m^{(n)}(x)$ können wir nun ebenfalls in die Summen $\sum_{\text{I}}$ und $\sum_{\text{II}}$ zerlegen,

$$\sum_{m=0}^{n} (m-nx)^2 u_m^{(n)}(x) = \sum_{\text{I}} (m-nx)^2 u_m^{(n)}(x) + \sum_{\text{II}} (m-nx)^2 u_m^{(n)}(x),$$

was jetzt

$$\sum_{\text{II}} (m-nx)^2 u_m^{(n)}(x) \leqq \sum_{m=0}^{n} (m-nx)^2 u_m^{(n)}(x) \tag{51}$$

ergibt. Aus (50) und (51) folgt

$$\sum_{\text{II}} u_m^{(n)}(x) \leqq \frac{1}{\delta^2 n^2} \sum_{m=0}^{n} (m-nx)^2 u_m^{(n)}(x). \tag{52}$$

Die in (52) auf der rechten Seite auftretende Summe läßt sich nun aber sehr einfach berechnen. Hierzu brauchen wir nämlich nur in dem im vorangehenden Paragraphen bewiesenen Lemma $p = x$ und $q = 1 - x$ zu setzen; dies ergibt

$$\sum_{m=0}^{n} (m-nx)^2 u_m^{(n)}(x) = \sum_{m=0}^{n} (m-nx)^2 \binom{n}{m} x^m (1-x)^{n-m} = nx(1-x).$$

Damit nimmt die Ungleichung (52) die Form

$$\sum_{\text{II}} u_m^{(n)}(x) \leqq \frac{x(1-x)}{\delta^2 n}$$

an. Das hierbei auftretende Produkt $x(1-x)$ können wir nun sofort folgendermaßen abschätzen:

$$x(1-x) = \frac{1}{4} - \left(x - \frac{1}{2}\right)^2 \leqq \frac{1}{4};$$

hieraus folgt

$$\sum_{\text{II}} u_m^{(n)}(x) \leqq \frac{1}{4\delta^2 n}.$$

Aus dieser Ungleichung und der früher erhaltenen Ungleichung (49) ergibt sich

$$\sum_{\text{II}} \left| f\left(\frac{m}{n}\right) - f(x) \right| u_m^{(n)}(x) \leqq \frac{M}{2\delta^2 n}. \tag{53}$$

Fassen wir schließlich die Beziehungen (41), (45) und (48), (53) zusammen, so erhalten wir

$$|B_n(x) - f(x)| < \frac{\varepsilon}{2} + \frac{M}{2\delta^2 n}.$$

Die Zahl $\varepsilon > 0$ ist hierbei willkürlich vorgegeben, während die Zahl δ unter Berücksichtigung der gleichmäßigen Stetigkeit von $f(x)$ in Abhängigkeit von ε zu bestimmen ist. Bisher haben wir noch nichts über den Grad n des Polynoms $B_n(x)$ vorausgesetzt. Wir wollen nun annehmen, daß der Grad n so groß ist, daß die Ungleichung

$$\frac{M}{2\delta^2 n} < \frac{\varepsilon}{2}$$

erfüllt ist, d. h., es sei

$$n > \frac{M}{\delta^2 \varepsilon}. \tag{54}$$

Dann gilt

$$|B_n(x) - f(x)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Diese Ungleichung gilt für jedes x aus dem Intervall $0 \leqq x \leqq 1$. Nehmen wir für x speziell denjenigen Punkt aus dem Intervall $0 \leqq x \leqq 1$, in dem die stetige Funktion $|B_n(x) - f(x)|$ ihr Maximum im Intervall $0 \leqq x \leqq 1$ annimmt (vgl. § 47, Satz III), so erhalten wir

$$\varrho(B_n, f) = \max |B_n(x) - f(x)| < \varepsilon.$$

Damit haben wir gezeigt, daß es zu jedem $\varepsilon > 0$ eine Zahl n_ε gibt $\left(\text{z. B. } n_\varepsilon = \frac{M}{\delta^2 \varepsilon}\right)$, so daß für alle $n > n_\varepsilon$ die Ungleichung

$$\varrho(B_n, f) < \varepsilon$$

erfüllt ist. Das bedeutet aber gerade

$$\varrho(B_n, f) \to 0,$$

so daß im Intervall $0 \leqq x \leqq 1$

$$B_n(x) \rightrightarrows f(x)$$

gilt, womit der Satz von BERNSTEIN bewiesen ist.

Beispiel 1. Für $f(x) = a^x$ erhalten wir

$$B_n(x) = \sum_{m=0}^{n} a^{\frac{m}{n}} \binom{n}{m} x^m (1-x)^{n-m} = \left[(1-x) + a^{\frac{1}{n}} x\right]^n = \left[1 + \left(\sqrt[n]{a} - 1\right) x\right]^n.$$

Beispiel 2. Für $f(x) = \sin \pi x$ ist[1])

$$B_n(x) = \sum_{m=0}^{n} \sin \frac{m\pi}{n} \cdot \binom{n}{m} x^m (1-x)^{n-m}$$

$$= \frac{1}{2i}\left\{\left[\left(1 - 2x \sin^2 \frac{\pi}{2n}\right) + i\, x \sin \frac{\pi}{n}\right]^n - \left[\left(1 - 2x \sin^2 \frac{\pi}{2n}\right) - i\, x \sin \frac{\pi}{n}\right]^n\right\}.$$

[1]) Hierbei benutzen wir die Formel

$$\sin t = \frac{e^{it} - e^{-it}}{2i}$$

(vgl. S. 484). Die Koeffizienten von $B_n(x)$ sind natürlich reell.

Bemerkung. Aus (54) kann man entnehmen, welcher Grad n für B_n dafür ausreicht, daß $\varrho(B_n, f)$ kleiner als eine gegebene Zahl ε wird.

Im Falle der Funktion $f(x) = \sin \pi x$ kann man etwa $M = 1$ setzen; ferner genügt es wegen

$$|\sin \pi x' - \sin \pi x''| < \pi |x' - x''| \text{ [1]}$$

offenbar,

$$\delta = \frac{\varepsilon}{\pi}$$

zu setzen, damit für $|x' - x''| < \delta$ stets $|f(x') - f(x'')| < \varepsilon$ ist. Daher ist gemäß (54)

$$n > \frac{\pi^2}{\varepsilon^2}$$

zu wählen, damit $\varrho(B_n, f) < \varepsilon$ wird. Für $\varepsilon = 0{,}1$ ergibt dies, daß sicher dann $\varrho(B_n, f)$ kleiner als 0,1 ist, wenn der Grad von $B_n(x)$ größer als 3141 ist.

Dieses Resultat ist zunächst recht entmutigend, jedoch muß man berücksichtigen, daß ein so hoher Grad hinreichend dafür ist, daß $\varrho(B_n, f)$ kleiner als 0,1 wird; er ist aber keineswegs dafür notwendig, tatsächlich wird die gewünschte Approximation schon bei bedeutend kleineren Werten von n erreicht. Wir haben uns hier nur mit der prinzipiellen Seite der gestellten Aufgabe befaßt. Hätten wir es uns zur Aufgabe gemacht, den tatsächlichen Fehler der Approximation abzuschätzen, so wären dafür andere (und zwar wesentlich kompliziertere) Methoden notwendig gewesen.

§ 51. Endgültige Definition der Exponentialfunktionen. Fortsetzung einer stetigen Funktion auf die Randpunkte einer überall dichten Menge

Im vorliegenden Paragraphen wollen wir uns mit Fragen folgenden Typs beschäftigen: Ist es möglich, den Wert einer Funktion $y = f(x)$ in einem Punkte $x = c$ aus Werten der Funktion in gewissen von c verschiedenen Punkten zu bestimmen? Ist die Funktion $f(x)$ dabei keinen speziellen Bedingungen unterworfen, so ist dies offenbar nicht möglich, und zwar folgt das unmittelbar aus der Definition der Funktion als einer „willkürlichen Zuordnung" (vgl. § 46). Ganz anders liegen dagegen die Dinge, wenn man über die Funktion $f(x)$ passende Voraussetzungen macht, durch die aus der Klasse aller Funktionen eine engere Klasse ausgewählt wird.

Beschränkt man sich z. B. auf die Klasse der elementaren Funktionen (vgl. § 1), so kann man beweisen, daß man bei Kenntnis der Werte einer elementaren Funktion in allen Punkten eines Intervalls noch so kleiner Länge die Werte der Funktion in allen Punkten, in denen die Funktion erklärt ist, berechnen kann (daß alle Werte der Funktion durch ihre Werte in einem noch so kleinen Intervall eindeutig festgelegt sind). Hier noch ein anderes Beispiel: Jedem Leser ist die triviale Tatsache bekannt, daß zur vollständigen Definition einer linearen Funktion die Vorgabe ihrer Werte in nur zwei Punkten ausreicht; allgemein ist der Verlauf eines Polynoms vom Grade n eindeutig durch die Werte des Polynoms in $n + 1$ verschiedenen Punkten festgelegt.

Im folgenden wollen wir untersuchen, in welchem Maße die Werte einer stetigen Funktion in einer gewissen Punktmenge die Werte der Funktion in Punkten bestimmen, welche nicht dieser Menge angehören.

---

[1]) Vgl. S. 180.

Aus der Definition der Stetigkeit ergibt sich unmittelbar, daß für eine in einer gewissen Umgebung des Punktes $x = c$ definierte und in diesem Punkte stetige Funktion $f(x)$ der Wert $f(c)$ eindeutig durch die Werte der Funktion $f(x)$ in den Punkten irgendeiner gegen c konvergierenden Folge $\{x_n\}$ festgelegt ist. In diesem Fall ist nämlich

$$f(c) = \lim f(x_n).$$

Speziell ist also der Wert $f(c)$ eindeutig festgelegt durch die Werte der Funktion $f(x)$ auf jeder Menge E, aus der man eine gegen c konvergierende Folge $\{x_n\}$ auswählen kann.

Von diesem Gesichtspunkt aus verdienen die überall dichten Mengen ein besonderes Interesse. Man nennt eine Punktmenge *E überall dicht*, wenn in jedem (noch so kleinen) Intervall (α, β), $\alpha < \beta$, wenigstens ein Punkt der Menge E liegt. Allgemeiner nennt man eine Menge *E überall dicht in einem vorgegebenen Intervall I*, wenn die genannte Bedingung für jedes in I enthaltene Intervall (α, β) erfüllt ist.

Beispiele für überall dichte Mengen sind 1. die Menge aller rationalen Zahlen, 2. die Menge aller Zahlen, die eine endliche Dezimalbruchentwicklung besitzen, 3. die Menge aller Zahlen der Form $\frac{m}{2^n}$ (m und n ganz) usw.

Satz. *Ist eine Menge E überall dicht (in einem gegebenen Intervall I), so gibt es zu jedem c (aus I) eine Folge $\{x_n\}$ von Punkten aus E, die gegen c konvergiert:*

$$x_n \to c. \tag{55}$$

In der Tat: Es sei $\{\varepsilon_n\}$ eine monoton fallende Folge von positiven Zahlen, die gegen Null konvergiert. Wir wählen aus jedem Intervall $(c - \varepsilon_n,\ c + \varepsilon_n)$ $(n = 1, 2, \ldots)$ einen Punkt x_n aus E. Für die so erhaltene Folge $\{x_n\}$ gilt dann offenbar (55).

Nebenbei sei bemerkt, daß man sogar erreichen kann, daß die Folge $\{x_n\}$ monoton wächst oder monoton fällt. Um etwa eine monoton wachsende Folge zu erhalten, wählt man den Punkt x_1 aus dem Intervall $(c - \varepsilon_1, c)$ und anschließend jeden Punkt x_{n+1} innerhalb des Intervalls (x_n, c) $(n = 1, 2, \ldots)$.

Aus dem Vorangehenden ergibt sich:

Alle Werte einer in einem gegebenen Intervall I stetigen Funktion $f(x)$ sind eindeutig festgelegt durch die Werte der Funktion $f(x)$ in den Punkten einer beliebigen, in I überall dichten Menge E.

Dieses Ergebnis ermöglicht es uns, die bei der Definition der Exponentialfunktion $f(x) = a^x (a > 1)$ verbliebene Lücke zu schließen. Nach Formel (69) aus § 20 gilt

$$f\left(\frac{p}{q}\right) = a^{\frac{p}{q}} = \sqrt[q]{a^p}; \tag{56}$$

damit ist die Exponentialfunktion a^x in allen Punkten der überall dichten Menge der rationalen Zahlen definiert (gegeben). Unsere Aufgabe besteht nun darin, mit Hilfe der Operation des Grenzüberganges die Funktion a^x auch für alle irrationalen Argumentwerte zu erklären. Zu diesem Zweck sei c eine

beliebige irrationale Zahl. Die vorangehenden Ergebnisse legen es nahe, eine Folge $\{r_n\}$ von rationalen Zahlen mit $r_n \to c$ zu betrachten und den Funktionswert a^c durch die Beziehung

$$f(c) = a^c = \lim_{n \to \infty} a^{r_n} \tag{57}$$

zu definieren. Diese Überlegung bedarf aber noch einer wesentlichen Ergänzung, da zunächst keineswegs sicher ist, ob es eine stetige Funktion $f(x)$ gibt, welche in den rationalen Punkten $r = \frac{p}{q}$ die in (56) angegebenen Werte annimmt. Hierfür ist zu zeigen, daß 1. in (57) der Grenzwert auf der rechten Seite existiert und daß 2. dieser Grenzwert nicht von der Zahlenfolge $\{r_n\}$, sondern nur von deren Grenzwert c abhängt.

Daß die Funktion a^x im Bereich der rationalen Zahlen stetig ist, ist hierbei ohne Belang. Zum Beispiel sind für jedes irrationale c die für $x \neq c$ definierten Funktionen

$$f_1(x) = \begin{cases} 0, & \text{falls } x < c, \\ 1, & \text{falls } x > c, \end{cases}$$

und $f_2(x) = \sin \frac{1}{x - c}$ beide im Bereich der rationalen Zahlen stetig, ohne daß es möglich ist, ihnen im Punkte $x = c$ einen Wert so zuzuschreiben, daß sie auch im Punkte $x = c$ stetig sind.

Beim Nachweis von 1. und 2. empfiehlt es sich, die Monotonie der Exponentialfunktion im Bereich der rationalen Zahlen und ihr Additionstheorem zu verwenden (vgl. dazu § 20).

Wir betrachten zunächst eine wachsende Folge $\{r_n'\}$ und eine fallende Folge $\{r_n''\}$ jeweils von rationalen Zahlen mit $r_n' \to c$ und $r_n'' \to c$. Wegen

$$r_1' < r_2' < \cdots < r_n' < \cdots < c < \cdots < r_n'' < \cdots < r_2'' < r_1'' \tag{58}$$

gilt dann

$$f(r_1') < f(r_2') < \cdots < f(r_n') < \cdots < f(r_n'') < \cdots < f(r_2'') < f(r_1''). \tag{59}$$

Die Folge $\{f(r_n')\}$ ist also wachsend und nach oben beschränkt, die Folge $\{f(r_n'')\}$ fallend und nach unten beschränkt. Daraus folgt, daß die Folge $\{f(r_n')\}$ einen Grenzwert C' und die Folge $\{f(r_n'')\}$ einen Grenzwert C'' besitzt:

$$\left.\begin{aligned} f(r_n') &\to C', \\ f(r_n'') &\to C'', \end{aligned}\right\} \tag{60}$$

wobei wegen

$$f(r_n') < f(r_n'') \qquad (n = 1, 2, \ldots)$$

die Ungleichung

$$C' \leqq C''$$

besteht.

Um nun zu zeigen, daß $C' = C''$, beachten wir, daß

$$a^{r_n''} = a^{r_n'' - r_n'} \cdot a^{r_n'}. \tag{61}$$

Hierbei gilt offenbar $r_n'' - r_n' \to 0$, so daß $a^{r_n'' - r_n'} \to 1$.[1]) Damit ergibt sich aus (61) unmittelbar

$$C'' = C'. \tag{62}$$

Bezeichnen wir den gemeinsamen Wert von C' und C'' mit C, so gilt also $f(r_n') \to C$ und $f(r_n'') \to C$.

Wir wollen nun zeigen, daß für jede Folge $\{\varrho_n\}$ von rationalen Zahlen, die gegen c strebt, die Folge der zugehörigen Funktionswerte $f(\varrho_n)$ gegen C strebt. Hierfür ist zu beweisen, daß man zu jeder noch so kleinen Zahl $\varepsilon > 0$ eine Zahl n_0 so finden kann, daß für alle $n > n_0$ die Ungleichung

$$|f(\varrho_n) - C| < \varepsilon \tag{63}$$

erfüllt ist. Zu diesem Zweck wählen wir zunächst N so groß, daß

$$C - \varepsilon < f(r_N') \quad \text{und} \quad f(r_N'') < C + \varepsilon. \tag{64}$$

Da nun c im Innern des Intervalls (r_N', r_N'') liegt, können wir sodann eine Zahl n_0 so bestimmen, daß für alle $n > n_0$ die Ungleichung

$$r_N' < \varrho_n < r_N''$$

erfüllt ist. Dann gilt aber für alle $n > n_0$ auch

$$f(r_N') < f(\varrho_n) < f(r_N''). \tag{65}$$

Aus (64) und (65) folgt jedoch unmittelbar (63).

Damit hat für jeden irrationalen Punkt $x = c$ die Definition der Exponentialfunktion $f(x) = a^x$ durch Formel (57) ihre Rechtfertigung erfahren.

Man zeigt nun leicht, daß die Exponentialfunktion auch in allen irrationalen Punkten stetig ist. Dazu ist zu beweisen, daß für jede Folge $\{x_n\}$ mit $x_n \to c$ die Folge $\{f(x_n)\}$ der zugehörigen Funktionswerte gegen $C = f(c)$ konvergiert. Der Beweis hierfür verläuft genauso wie der oben angegebene Beweis für rationale Zahlenfolgen $\{\varrho_n\}$, wenn man berücksichtigt, daß die Hinzunahme irrationaler Punkte an der Monotonie der Exponentialfunktion nichts ändert.

Damit ist jetzt die Exponentialfunktion a^x für alle x definiert und gezeigt, daß sie überall stetig und monoton wachsend ist.

Es bleibt schließlich zu zeigen, daß auch das Additionstheorem

$$a^{x'+x''} = a^{x'} \cdot a^{x''} \tag{66}$$

für die Exponentialfunktion allgemein richtig ist, was wir früher nur für den Fall rationaler Zahlen x' und x'' begründet haben. Zum Beweis dafür betrachten wir zwei Folgen $\{x_n'\}$ und $\{x_n''\}$ von rationalen Zahlen mit $x_n' \to x'$ und $x_n'' \to x''$. Dann gilt

$$a^{x_n'+x_n''} = a^{x_n'} \cdot a^{x_n''}.$$

Gehen wir in dieser Gleichung zur Grenze $n \to \infty$ über, so erhalten wir (66).

Wir wollen hier noch ein anderes, allgemeineres Verfahren erwähnen, mit dessen Hilfe man eine Funktion $f(x)$, die zunächst nur für alle Punkte einer in einem Intervall I überall dichten Menge E erklärt ist, „stetig" auf das ganze Intervall I fortsetzen kann.

[1]) Vgl. S. 182.

Dieses Verfahren beruht auf dem folgenden Satz:

Es seien die Werte einer Funktion $f(x)$ in allen Punkten einer im Intervall I ($a \leqq x \leqq b$) überall dichten Menge E gegeben, wobei 1. *die Funktion $f(x)$ auf der Menge E stetig ist und monoton wächst und* 2. *die Werte der Funktion $f(x)$ für x aus E einem gewissen Intervall K ($A \leqq y \leqq B$) angehören, in dem sie überall dicht liegen. Dann kann die Funktion $f(x)$ so auf das gesamte Intervall I fortgesetzt werden, daß die entstehende Funktion auf ganz I stetig ist.*

Der Beweis für diesen Satz verläuft ähnlich wie der Beweis für die Fortsetzbarkeit der Exponentialfunktion von der Menge der rationalen Zahlen auf die Menge aller reellen Zahlen. An die Stelle der Menge der rationalen Zahlen tritt dabei die Menge E, während an die Stelle der irrationalen Zahl c eine Zahl tritt, die in I liegt und nicht zu E gehört. Es muß nur gezeigt werden, daß man die Gleichung (62) auch ohne Benutzung spezieller Eigenschaften der Exponentialfunktion (also ohne Verwendung des Additionstheorems) beweisen kann, wobei jetzt aber natürlich die Voraussetzung 2. verwendet werden darf. Dazu nehmen wir an, es wäre $C' < C''$. Dann gibt es wegen 2. eine Zahl $x = r$ aus E derart, daß

$$C' < f(r) < C''. \tag{67}$$

Da c nicht zu E gehört, ist $r \neq c$. Es sei etwa $r < c$; für hinreichend großes n ist dann $r < r_n' < c$ und daher $f(r) < f(r_n')$, woraus sich durch Grenzübergang $f(r) \leqq C'$ ergibt, was der Ungleichung (67) widerspricht. Entsprechend führt man den Fall $r > c$ zum Widerspruch. Also muß $C' = C''$ sein.

Beispiel 1. Um den bewiesenen Satz speziell auf die Funktion a^x ($a > 1$) anwenden zu können, müssen wir zeigen, daß die Menge der Werte, die die Funktion a^x für rationales x annimmt, überall dicht ist. Dazu sei (α, β) ein beliebiges Intervall mit $0 < \alpha < \beta$. Wegen

$$\lim_{n \to \infty} \left(\frac{\beta}{\alpha}\right)^n = \infty$$

können wir eine natürliche Zahl n so finden, daß

$$\left(\frac{\beta}{\alpha}\right)^n > a,$$

also

$$q \equiv a^{\frac{1}{n}} < \frac{\beta}{\alpha}.$$

Man überlegt sich nun leicht, daß wenigstens eine der Zahlen

$$\ldots, q^{-2}, q^{-1}, 1, q, q^2, \ldots, q^n, \ldots$$

im Intervall (α, β) liegt; ist etwa

$$\alpha < q^m < \beta,$$

so gilt

$$\alpha < a^{\frac{m}{n}} < \beta,$$

was zu beweisen war.

Beispiel 2. Bei der in § 46, Beispiel 2, betrachteten Funktion $f(x)$ liegt die Menge K der dort definierten Funktionswerte überall dicht im Intervall (0, 1). Bezeichnen wir nämlich mit K_n die Menge derjenigen Werte, welche die Funktion $f(x)$ in rationalen Punkten mit einem Nenner $\leqq 2^n$ annimmt, so sehen wir, daß das größte der in (0, 1) durch die Punkte von K_n gebildeten Intervalle die Länge $\left(\frac{3}{4}\right)^n$ hat. Ist also (α, β) ein beliebiges Intervall mit $0 < \alpha < \beta < 1$, so brauchen wir nur n so groß zu wählen, daß $\left(\frac{3}{4}\right)^n < \alpha - \beta$ ist, um zu erreichen, daß ein Punkt von K_n (und damit von K) in (α, β) liegt.

Beispiel 3. Auch für die in § 46, Beispiel 3, betrachtete Funktion $f(x)$ ist die Menge K der dort definierten Funktionswerte überall dicht im Intervall (0, 1). Den Beweis hierfür überlassen wir dem Leser als Übungsaufgabe.

Wir wollen schließlich noch eine dritte Konstruktionsmethode für die Exponentialfunktion angeben, die auf einer gleichmäßigen Approximation der Exponentialfunktion durch Polynome beruht. Dazu betrachten wir die Folge der in § 45 näher untersuchten Polynome

$$P_n(x) = \left(1 + \frac{x}{n}\right)^n \qquad (n = 1, 2, 3, \ldots).$$

Man sieht leicht, daß die Folge $\{P_n(x)\}$ in jedem Intervall $(0, R)$ mit $R > 0$ gleichmäßig gegen eine bestimmte Grenzfunktion konvergiert.

In der Tat: Entwickeln wir $P_n(x)$ und $P_{n+1}(x)$ nach Potenzen von x, so erhalten wir zunächst (ähnlich wie in § 44, Formel (104)), daß

$$0 < P_1(x) \leqq P_2(x) \leqq \cdots \leqq P_n(x) \leqq \cdots \tag{68}$$

und daß darüber hinaus für jedes n

$$\begin{aligned} P_n(x) &= 1 + \frac{n}{1}\left(\frac{x}{n}\right) + \frac{n(n-1)}{1 \cdot 2}\left(\frac{x}{n}\right)^2 + \cdots + \frac{n(n-1)\ldots 1}{1 \cdot 2 \ldots n}\left(\frac{x}{n}\right)^n \\ &< 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} \\ &\leqq 1 + \frac{R}{1!} + \frac{R^2}{2!} + \cdots + \frac{R^n}{n!}. \end{aligned}$$

Es sei nun N eine ganze Zahl mit $N > R$. Dann ist für jedes $n > N$

$$\begin{aligned} P_n(x) &\leqq 1 + R + \frac{R^2}{2!} + \cdots + \frac{R^N}{N!} + \frac{R^N}{N!}\left\{\left(\frac{R}{N}\right) + \left(\frac{R}{N}\right)^2 + \cdots + \left(\frac{R}{N}\right)^{n-N}\right\} \\ &\leqq 1 + R + \frac{R^2}{2!} + \cdots + \frac{R^N}{N!} + \frac{R^N}{N!}\,\frac{R}{N-R} \equiv M. \end{aligned} \tag{69}$$

Aus (68) und (69) folgt, daß für jedes x aus $[0, R]$ die Zahlenfolge $\{P_n(x)\}$ gegen einen gewissen endlichen Grenzwert $f(x)$ konvergiert:

$$P_n(x) \to f(x). \tag{70}$$

Wir wollen nun zeigen, daß es sich bei dieser Konvergenz sogar um eine gleichmäßige Konvergenz handelt. Dazu betrachten wir die Differenz $P_{m+1}(x) - P_m(x)$. Da in ihrer Entwicklung nach Potenzen von x nur positive Koeffizienten auftreten, können wir sie höchstens vergrößern, wenn wir in ihr x durch R ersetzen:

$$P_{m+1}(x) - P_m(x) \leqq P_{m+1}(R) - P_m(R).$$

Hieraus ergibt sich u. a.

$$\sum_{m=n}^{n+p-1} [P_{m+1}(x) - P_m(x)] \leqq \sum_{m=n}^{n+p-1} [P_{m+1}(R) - P_m(R)],$$

d. h.

$$P_{n+p}(x) - P_n(x) \leqq P_{n+p}(R) - P_n(R).$$

Vollziehen wir in dieser Ungleichung den Grenzübergang $p \to \infty$, so erhalten wir

$$f(x) - P_n(x) \leqq f(R) - P_n(R). \tag{71}$$

Es bleibt zu bemerken, daß bei vorgegebenem $\varepsilon > 0$ und beliebigem x aus dem Intervall $[0, R]$ die rechte (von x unabhängige) Seite der Ungleichung (71) und damit auch die linke Seite für hinreichend großes n kleiner als ε gemacht werden kann, so daß im Intervall $[0, R]$ die Folge $\{P_n(x)\}$ gleichmäßig gegen die Grenzfunktion $f(x)$ konvergiert:

$$P_n(x) \Rightarrow f(x).$$

Da nun die Funktionen $P_n(x)$ überall stetig sind, ist die Funktion $f(x)$ im Intervall $[0, R]$ stetig. Da hierbei die Zahl R beliebig groß gewählt werden kann, ist die Funktion $f(x)$ sogar im Intervall $[0, \infty)$ stetig.

Andererseits zeigen die in § 45, Beispiel 1, durchgeführten Überlegungen, daß für jedes rationale x (und nur für diese sind im gegenwärtigen Zusammenhang die dortigen Überlegungen stichhaltig) die Gleichung

$$f(x) = e^x$$

gilt. Daraus folgt, daß die überall stetige Funktion

$$f(x) = \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n \quad \text{für} \quad x \geqq 0,$$

$$f(x) = \frac{1}{f(-x)} \quad \text{für} \quad x < 0$$

für jedes rationale x den Wert e^x annimmt.

Damit haben wir zunächst die Exponentialfunktion e^x vollständig definiert. Man überlegt sich leicht, daß man die Exponentialfunktion a^x erhält, wenn man in der letzten Formel x durch $x \ln a$ ersetzt:

$$a^x = \lim_{n\to\infty} \left(1 + \frac{x \ln a}{n}\right)^n. \tag{72}$$

§ 52. Der Satz von Bolzano und das Problem der Existenz einer eindeutigen Umkehrfunktion

Vorgegeben sei eine im abgeschlossenen Intervall $I = [a, b]$ stetige Funktion $y = f(x)$. Nach dem Satz von Weierstrass (§ 47, Satz III) nimmt die Funktion $f(x)$ in bestimmten Punkten des Intervalls ihren kleinsten Wert A und ihren größten Wert B an. Der Fall $A = B$, in dem sich die Funktion auf eine Konstante reduziert, soll im folgenden ausgeschlossen werden. Auf Grund des Satzes von Bolzano (§ 47, Satz I) können wir dann behaupten, daß die Funktion $f(x)$ im Intervall I jeden zwischen A und B liegenden Wert μ (evtl. mehrmals) annimmt. Für jedes y mit $A \leqq y \leqq B$ besitzt also die Gleichung

$$f(x) = y \tag{73}$$

wenigstens eine Lösung x mit $a \leqq x \leqq b$. Ordnen wir nun jedem Wert y aus dem Intervall $[A, B]$ alle diejenigen x zu, welche Lösung der Gleichung (73) sind, so erhalten wir eine im allgemeinen mehrdeutige Funktion

$$x = g(y), \tag{74}$$

die man die *Umkehrfunktion* der Funktion $f(x)$ nennt und die im ganzen Intervall $[A, B]$ erklärt ist.

Besonderes Interesse verdient der Fall, daß die Umkehrfunktion $x = g(y)$ eindeutig ist. Eine hinreichende Bedingung hierfür liefert der folgende Satz:

Ist die Funktion $y = f(x)$ stetig und im Intervall $[a, b]$ wachsend (fallend), so ist die Umkehrfunktion $x = g(y)$ eindeutig, wachsend (fallend) und im Intervall $[A, B]$ stetig.[1])

Ist die Funktion $f(x)$ wachsend (fallend), so ist offenbar $A = f(a)$ und $B = f(b)$ $(A = f(b)$ und $B = f(a))$.

Wir beschränken uns auf den Beweis des Satzes für wachsende Funktionen.

Die Existenz einer Lösung der Gleichung (73) ist für $y = A$ und $y = B$ klar; für $A < y < B$ ergibt sie sich, wie bereits erwähnt wurde, aus dem Satz von Bolzano. Daß es auch nur eine Lösung gibt, erkennt man folgendermaßen: Angenommen, es gäbe für ein gegebenes y zwei Lösungen x' und x'' $(x' < x'')$; da die Funktion $f(x)$ als wachsend vorausgesetzt ist, folgt aus $x' < x''$ sofort

$$f(x') < f(x''),$$

im Widerspruch dazu, daß $f(x') = y$ und $f(x'') = y$, also $f(x') = f(x'')$ sein sollte.

Bezeichnen wir die für ein gegebenes y aus $[A, B]$ eindeutig bestimmte Lösung x der Gleichung (73) mit $g(y)$, so ist also

$$f(g(y)) = y \qquad (A \leqq y \leqq B).$$

Wir wollen nun zeigen, daß die Funktion $g(y)$ im Intervall $[A, B]$ monoton wächst. Dazu seien y' und y'' Zahlen aus dem Intervall $[A, B]$ mit

$$A \leqq y' < y'' \leqq B. \tag{75}$$

Setzen wir etwa $g(y') = x'$, $g(y'') = x''$, so ist

$$f(x') = y', \quad f(x'') = y''.$$

Zu zeigen ist, daß $x' < x''$. Dazu nehmen wir an, dies wäre nicht der Fall, d. h., es wäre $x' \geqq x''$. Da die Funktion $f(x)$ wächst, wäre dann $f(x') \geqq f(x'')$, d. h. $y' \geqq y''$, im Widerspruch zu (75).

Es bleibt also zu zeigen, daß die Funktion $g(y)$ im Intervall $[A, B]$ stetig ist. Dazu sei $y = d$ eine beliebige Zahl mit $A \leqq d \leqq B$. Da die Funktion $g(y)$, wie soeben bewiesen wurde, wächst, existieren (vgl. § 42 und § 44) der linksseitige und der rechtsseitige Grenzwert

$$L' = \lim_{\substack{y \to d \\ y < d}} g(y) \quad \text{und} \quad L'' = \lim_{\substack{y \to d \\ y > d}} g(y).$$[2])

---

[1]) Bezüglich der Umkehrung dieses Satzes vgl. den Schluß von § 62.

[2]) Ist d einer der Endpunkte A bzw. B, so haben im folgenden natürlich nur die in bezug auf L'' bzw. L' ausgesprochenen Behauptungen Gültigkeit.

Ist $y < d$, so ist $g(y) < g(d)$; durch Grenzübergang $y \to d$ ergibt sich hieraus $L' \leqq g(d)$. Entsprechend erhält man $g(d) \leqq L''$. Somit gilt

$$L' \leqq g(d) \leqq L''.$$

Zu zeigen ist, daß

$$L' = g(d) = L''.$$

Nehmen wir an, dies wäre nicht der Fall, es wäre also z. B. $L' < g(d)$. Dann betrachten wir eine beliebige Zahl ξ mit

$$L' < \xi < g(d). \tag{76}$$

Wegen $g(d) < g(B) = b$ und $L' > g(A) = a$ ist $a < \xi < b$. Daher entspricht dem Argumentwert $x = \xi$ ein bestimmter Funktionswert $f(\xi)$, wobei wegen (76)

$$f(L') < f(\xi) < f(g(d)) = d. \tag{77}$$

Andererseits folgt aus $\lim\limits_{\substack{y \to d \\ y < d}} g(y) = L'$ auf Grund der Stetigkeit der Funktion $f(x)$

$$\lim_{\substack{y \to d \\ y < d}} f(g(y)) = f(L').$$

Wegen

$$\lim_{\substack{y \to d \\ y < d}} f(g(y)) = \lim_{\substack{y \to d \\ y < d}} y = d$$

ist daher

$$f(L') = d.$$

Damit nimmt die Ungleichung (77) die Form

$$d < f(\xi) < d$$

an, und dies ist offenbar ein Widerspruch. Entsprechend führt man den Fall $g(d) < L''$ ad absurdum.

Ist das betrachtete Intervall I an einem seiner Enden oder an beiden Enden nicht abgeschlossen, z. B. I das Intervall $a \leqq x < b$ (wobei auch der Fall $b = \infty$ zugelassen ist), so kann man den zuvor bewiesenen Satz zunächst auf jedes Intervall $a \leqq x \leqq b_1$ mit $b_1 < b$ anwenden und anschließend zur Grenze $b_1 \to b$ übergehen. Dabei zeigt sich, daß der Satz auch in diesem (und den anderen Fällen) richtig ist, wobei im betrachteten Fall an die Stelle des abgeschlossenen Intervalls $[A, B]$ natürlich überall das halboffene Intervall $A \leqq y < B$ mit $B = \lim\limits_{x \to b} f(x)$ tritt.

Beispiel 1. Die Funktion $y = a^x (a > 1)$ ist stetig und wachsend (im eigentlichen Sinne) im Intervall $-\infty < x < +\infty$. Auf Grund des oben bewiesenen Satzes folgt hieraus, daß ihre Umkehrfunktion $x = \log_a y$ eindeutig, wachsend und stetig im Intervall $0 < y < \infty$ ist.

Beispiel 2. Die Funktion $y = x^n$ (n natürliche Zahl) ist stetig und wachsend für $0 \leqq x < \infty$. Also ist ihre Umkehrfunktion $x = \sqrt[n]{y}$ (die Wurzel im arithmetischen Sinne) stetig und monoton wachsend für $0 \leqq y < \infty$.

Beispiel 3. Die Funktion $y = \operatorname{tg} x$ ist stetig und wachsend für $-\frac{\pi}{2} < x < \frac{\pi}{2}$. Demzufolge ist die Funktion $x = \operatorname{arctg} y$ stetig und wachsend für $-\infty < y < \infty$.

Beispiel 4. Da die Funktion $y = x^5 + x$ im Intervall $0 \leqq x < \infty$ stetig und wachsend ist, ist ihre Umkehrfunktion $x = g(y)$ eine eindeutige, monoton wachsende und stetige Funktion im Intervall $0 \leqq y < \infty$. Diese Funktion $g(y)$ ist jedoch, wie hier nur bemerkt werden soll, keine elementare Funktion mehr im Sinne von § 1 (vgl. auch die Bemerkung in § 18).

§ 53. Funktionalgleichungen und elementare Funktionen

Es sei $f(t)$ eine beliebige in einem Intervall I erklärte Funktion. Es seien ferner x und y Zahlen aus I, für die auch $x + y$ in I enthalten ist, so daß wir die Funktionswerte

$$X = f(x), \quad Y = f(y), \quad Z = f(x + y) \tag{78}$$

bilden können. Wir setzen ferner voraus, daß sich aus den drei Gleichungen (78) die Veränderlichen x und y eliminieren lassen, d. h., daß zwischen X, Y, Z eine Beziehung etwa der Form

$$F(X, Y, Z) = 0 \tag{79}$$

besteht, in der x und y nicht explizit auftreten. Ersetzt man in (79) X, Y, Z durch die entsprechenden Werte (78), so erhält man eine Beziehung der Form

$$F(f(x), f(y), f(x + y)) = 0, \tag{80}$$

welche für alle x, y gilt, die der obigen Einschränkung genügen.

Die Beziehung (80), welche der Kategorie der sogenannten **Funktionalgleichungen** angehört, bezeichnet man als *Additionstheorem* der Funktion $f(t)$.[1])

Hätten wir an Stelle der Summe $x + y$ das Produkt xy gewählt, so wären wir in entsprechender Weise zu einer Funktionalgleichung

$$\Phi(f(x), f(y), f(xy)) = 0 \tag{81}$$

gelangt, die man *Multiplikationstheorem* der Funktion $f(t)$ nennt.

Analog würde man, wenn man an Stelle der Summe oder des Produkts eine andere Funktion $\omega(x, y)$ nimmt, eine Funktionalgleichung der Form

$$\Psi\big(f(x), f(y), f(\omega(x, y))\big) = 0 \tag{82}$$

erhalten. Derartige Verallgemeinerungen werden uns hier jedoch nicht interessieren; wir werden uns vielmehr auf die Untersuchung der Additions- und Multiplikationstheoreme einiger spezieller Funktionen beschränken.

[1]) Es wird verlangt, daß die Beziehung (80) eine **Identität** in bezug auf x und y ist. Daß man trotzdem von einer Funktional„gleichung" spricht, liegt daran, daß man in der Beziehung (80) die Funktion $f(x)$ als Unbekannte ansieht.

Wir betrachten einige Beispiele (das Intervall I ist dabei in allen Fällen der Definitionsbereich der betrachteten Funktion).

Beispiel 1. Es sei $f(t)$ die homogene lineare Funktion

$$f(t) = mt.$$

In diesem Fall haben die Gleichungen (78) die Form

$$X = mx, \quad Y = my, \quad Z = m(x+y),$$

und es gilt

$$Z = X + Y,$$

d. h.

$$f(x+y) = f(x) + f(y). \tag{I}$$

Das ist das Additionstheorem für die Funktion $f(t) = mt$. Entsprechend ergibt sich aus den Gleichungen

$$X = mx, \quad Y = my, \quad Z = mxy$$

das Multiplikationstheorem

$$mf(xy) = f(x)f(y). \tag{83}$$

Wir sehen also, daß jede der Funktionen mt ein eigenes Multiplikationstheorem besitzt, während für alle homogenen linearen Funktionen das Additionstheorem dasselbe ist.

Beispiel 2.

$$f(t) = a^t \qquad (a > 0).$$

Aus

$$X = a^x, \quad Y = a^y, \quad Z = a^{x+y}$$

folgt

$$Z = XY,$$

d. h.

$$f(x+y) = f(x)f(y). \tag{II}$$

Das ist das Additionstheorem für alle Exponentialfunktionen a^t. Das Multiplikationstheorem für die Funktion a^t besitzt entsprechend die Form

$$\log_a f(xy) = \log_a f(x) \log_a f(y), \tag{84}$$

hängt also von der Basis a ab.

Beispiel 3.

$$f(t) = \log_a t.$$

In diesem Fall hängt das Additionstheorem vom Parameter a ab:

$$a^{f(x+y)} = a^{f(x)} + a^{f(y)},$$

während das Multiplikationstheorem nicht von a abhängt:

$$f(xy) = f(x) + f(y). \tag{III}$$

Beispiel 4.

$$f(t) = t^{\alpha} \qquad (\alpha \neq 0).$$

Auch hier hängt das Additionstheorem vom Parameter α ab:

$$[f(x+y)]^{\frac{1}{\alpha}} = [f(x)]^{\frac{1}{\alpha}} + [f(y)]^{\frac{1}{\alpha}},$$

während das Multiplikationstheorem wieder von α unabhängig ist:

$$f(xy) = f(x)f(y). \tag{IV}$$

Mit den vier Funktionalgleichungen (I) bis (IV) wollen wir uns noch etwas eingehender beschäftigen. Wir wollen nämlich untersuchen, welche der uns bekannten Funktionen diese Funktionalgleichungen erfüllen. Dabei wird sich herausstellen, daß mit einigen wesentlichen Einschränkungen die in den Beispielen angegebenen Funktionen bereits alle Lösungen der entsprechenden Funktionalgleichung ausschöpfen. Es gelten nämlich die folgenden Sätze:

Satz 1. *Genügt eine für alle reellen Zahlen t $(-\infty < t < \infty)$ definierte und überall stetige Funktion $f(t)$ der Funktionalgleichung* (I), *so hat sie die Form $f(t) = mt$, wobei m eine Konstante ist.*

Satz 2. *Genügt eine für alle reellen Zahlen t definierte und überall stetige Funktion $f(t)$ der Funktionalgleichung* (II), *so hat sie die Form $f(t) = a^t$, wobei a eine nicht negative Konstante ist.*

Satz 3. *Ist $f(t)$ eine für alle positiven reellen Zahlen t definierte und stetige Funktion, die nirgends verschwindet und identisch in allen positiven x und y die Funktionalgleichung* (III) *erfüllt, so hat sie die Form $f(t) = \log_a t$, wobei a eine positive Konstante ist.*

Satz 4. *Ist $f(t)$ eine für alle positiven reellen Zahlen t definierte und stetige Funktion, die nirgends verschwindet und identisch in allen positiven x und y die Funktionalgleichung* (IV) *erfüllt, so hat sie die Form $f(t) = t^{\alpha}$, wobei α eine Konstante ist.*

Beweis von Satz 1. Durch vollständige Induktion über n beweist man zunächst aus (I), daß

$$f(x_1 + x_2 + \cdots + x_n) = f(x_1) + f(x_2) + \cdots + f(x_n). \tag{85}$$

Für

$$x_1 = x_2 = \cdots = x_n = x$$

ergibt sich hieraus

$$f(nx) = nf(x). \tag{86}$$

Setzen wir nun

$$m = f(1),$$

so erhalten wir aus (86) für $x = 1$, daß für alle natürlichen Zahlen n die Gleichung

$$f(n) = mn$$

gilt.

Ersetzen wir in (86) n durch q und x durch $\frac{p}{q}$ (p, q positive ganze Zahlen), so erhalten wir

$$f(p) = q f\left(\frac{p}{q}\right),$$

woraus sich wegen $f(p) = mp$ unmittelbar

$$f\left(\frac{p}{q}\right) = m\frac{p}{q} \tag{87}$$

ergibt.

Setzen wir in (I) $x = 1$ und $y = 0$, so erhalten wir

$$f(1) = f(1) + f(0)$$

und damit

$$f(0) = 0. \tag{88}$$

Ersetzen wir schließlich in (I) y durch $-x$, so finden wir

$$0 = f(x) + f(-x)$$

und damit

$$f(-x) = -f(x). \tag{89}$$

Aus (87), (88) und (89) folgt, daß auf der überall dichten Menge der rationalen Zahlen die Beziehung

$$f(x) = mx \tag{90}$$

gilt. Hieraus ergibt sich wegen der vorausgesetzten Stetigkeit der Funktion $f(t)$ und der Stetigkeit der Funktion mt auf Grund des in § 51 bewiesenen Satzes, daß die Beziehung (90) für alle reellen Zahlen x gilt.

Beweis von Satz 2. Ersetzen wir in (II) sowohl x als auch y durch $\frac{t}{2}$, so erhalten wir

$$f(t) = f^2\left(\frac{t}{2}\right) \geqq 0.$$

Wenn nun für irgendeinen Wert $t = x$ die Funktion $f(t)$ den Wert Null annimmt, so verschwindet sie wegen (II) auch für $t = x + y$, wobei y eine ganz beliebige reelle Zahl ist, und dies bedeutet, daß die Funktion $f(t)$ dann identisch verschwindet. In diesem Fall ist also $f(t) = 0^t$.

Wir können also annehmen, daß für alle t

$$f(t) > 0$$

gilt. Es sei dann $f(1) = a > 0$. Setzen wir

$$f(t) = a^{\varphi(t)}, \quad \varphi(t) = \log_a f(t),$$

so ergibt sich aus (II)

$$a^{\varphi(x+y)} = a^{\varphi(x)} a^{\varphi(y)},$$

d. h.

$$\varphi(x+y) = \varphi(x) + \varphi(y).$$

Das ist aber eine Funktionalgleichung vom Typ (I), und wir können dem bereits bewiesenen Satz 1 entnehmen, daß

$$\varphi(t) = mt$$

oder

$$f(t) = a^{mt}.$$

Setzen wir hier $t = 1$, so finden wir, daß $m = 1$, also $f(t) = a^t$ ist.

Beweis von Satz 4. Da hierbei x und y positive reelle Zahlen sein sollen, können wir

$$x = 10^{\xi}, \quad \xi = \lg x,$$
$$y = 10^{\eta}, \quad \eta = \lg y$$

setzen, wobei die Funktionalgleichung (IV) die Form

$$f(10^{\xi+\eta}) = f(10^{\xi}) f(10^{\eta}) \tag{IV'}$$

annimmt. Setzen wir

$$F(\tau) \equiv f(10^{\tau}), \tag{91}$$

so können wir (IV') auch in der Form

$$F(\xi + \eta) = F(\xi)\, F(\eta)$$

ausdrücken.

Hierbei sind nun ξ und η beliebige reelle Zahlen. Die Funktion $F(\tau)$ ist als Hintereinanderausführung der stetigen Funktionen $f(t)$ und 10^{τ} überall stetig. Daher können wir auf Grund von Satz 2 behaupten, daß

$$F(\tau) \equiv a^{\tau} \qquad (a \geqq 0).$$

Ersetzen wir in dieser Identität τ durch $\lg t$, so erhalten wir schließlich

$$f(t) \equiv a^{\lg t} \equiv \begin{cases} 0, \text{ falls } a = 0, \\ t^{\alpha} \text{ (wobei } \alpha = \lg a), \text{ falls } a > 0. \end{cases}$$

Beweis von Satz 3. Offenbar können wir die Funktionalgleichung (III) auch in der Form

$$10^{f(xy)} = 10^{f(x)}\, 10^{f(y)}$$

schreiben. Setzen wir

$$G(t) \equiv 10^{f(t)}, \tag{92}$$

so können wir ihr auch die Gestalt

$$G(xy) = G(x)\, G(y) \tag{III'}$$

geben. Auf Grund von Satz 4 können wir dann jedoch feststellen, daß entweder $G(t) \equiv 0$ ist oder aber $G(t) \equiv t^{\alpha}$ gilt. Da nun eine Exponentialfunktion nirgends den Wert Null annimmt, scheidet der erste Fall aus, d. h., es ist $G(t) \equiv t^{\alpha}$. Ist dabei $\alpha = 0$, so ist $G(t) \equiv 1$ und $f(t) \equiv 0$. Ist dagegen $\alpha \neq 0$, so ist

$$f(t) \equiv \alpha \lg t = \log_a t \qquad \left(\text{mit } a = 10^{\frac{1}{\alpha}}\right).$$

Die vorangehenden Untersuchungen zeigen, daß die Funktionalgleichungen (I) bis (IV) für die Funktionen mt, a^t, $\log_a t$ bzw. t^ε charakteristisch sind; d. h. mit anderen Worten, diese Gleichungen können unter zusätzlicher Voraussetzung der Stetigkeit der Funktionen zur Definition dieser Funktionen dienen.

Bemerkung 1. Fordert man an Stelle der Stetigkeit der Funktionen ihre Differenzierbarkeit (vgl. S. 293), so sind die Sätze 1 bis 4 erst recht richtig. Unter dieser Voraussetzung lassen sich jedoch die Beweise wesentlich vereinfachen. Zum Beispiel kann man dann Satz 1 folgendermaßen beweisen: Differenziert man die Gleichung (I) nach x, so erhält man

$$f'(x+y) = f'(x).$$

Da hierbei y ganz beliebig ist, muß $f'(x)$ eine Konstante sein:

$$f'(x) = m.$$

Durch Integration ergibt sich jedoch hieraus

$$f(x) = mx + C,$$

wobei C eine neue Konstante ist. Setzt man den so gefundenen Ausdruck für $f(x)$ in (I) ein, so erhält man, daß identisch in x und y die Gleichung

$$m(x+y) + C = (mx + C) + (my + C)$$

gelten muß, und daraus folgt, daß $C = 0$, also $f(x) = mx$ ist.

Unter den Funktionen, die durch die Funktionalgleichungen (I) bis (IV) erfaßt werden, befinden sich weder die trigonometrischen Funktionen noch deren Umkehrfunktionen.

Um z. B. das Additionstheorem für die Funktion $f(t) = \sin t$ aufzustellen, geht man von der aus der Trigonometrie bekannten Formel für den Sinus einer Summe aus und drückt den darin auftretenden Kosinus durch den Sinus aus; die dabei entstehende Formel hat die recht komplizierte Gestalt

$$f(x+y) = f(x) \cdot \sqrt{1 - f^2(y)} + f(y) \cdot \sqrt{1 - f^2(x)}.$$

An dieser Stelle ist die folgende Überlegung angebracht: In der Theorie der Funktionen einer komplexen Veränderlichen wird bewiesen, daß sich die trigonometrischen Funktionen im Komplexen durch die Exponentialfunktion ausdrücken lassen, und zwar auf folgende Weise:

$$\cos t \equiv \frac{e^{it} + e^{-it}}{2}, \quad \sin t \equiv \frac{e^{it} - e^{-it}}{2i}; \tag{93}$$

dadurch verlieren sie sozusagen das Recht auf selbständige Bedeutung; entsprechend kann man die zyklometrischen Funktionen durch Logarithmen und Wurzeln ausdrücken (vgl. S. 491). Hieraus folgt, daß man zur Konstruktion der elementaren Funktionen neben den vier Grundrechenarten (Addition, Subtraktion, Multiplikation und Division) lediglich noch die Exponentialfunktion, das Logarithmieren und das Potenzieren mit einem beliebigen Exponenten benötigt.[1])

[1]) Wir erinnern daran, daß sich auch das Potenzieren mit einem beliebigen Exponenten noch auf die übrigen Operationen zurückführen läßt (vgl. S. 84).

Damit ergibt sich die Richtigkeit folgender Behauptung:

Jede elementare Funktion kann man in endlich vielen Schritten aus der unabhängigen Veränderlichen x und konstanten Zahlen mit Hilfe der arithmetischen Grundrechenarten und derjenigen Operationen, die als (stetige) Lösungen der Funktionalgleichungen (I) *bis* (IV) *auftreten, aufbauen.*

Hieraus erkennt man die Bedeutung, die in theoretischer Hinsicht die Funktionalgleichungen (I) bis (IV) für die Elementar-Mathematik besitzen.

Bemerkung 2. Genau genommen folgt die soeben ausgesprochene Behauptung noch nicht aus den Sätzen 1 bis 4, weil bei Bildung der trigonometrischen Funktionen aus der Exponentialfunktion gemäß (93) noch die Operation $f(t) = it$ benötigt wird, während bisher in der Lösung $f(t) = mt$ der Gleichung (I) die Konstante m stets als reell vorausgesetzt wurde. Um unsere Behauptung endgültig zu beweisen, ist es noch nötig, die Sätze 1 bis 4 auch im Komplexen als richtig nachzuweisen. Hierbei ist natürlich die Voraussetzung der Stetigkeit durch die der Regularität der gesuchten Funktion $f(t)$ zu ersetzen. In dem zu Satz 1 analogen Satz im Komplexen ist also die Funktion f in der ganzen Ebene als regulär vorauszusetzen und zu verlangen, daß für alle komplexen Zahlen x, y die Beziehung (I) erfüllt ist, Da aus der Regularität speziell die Differenzierbarkeit folgt, kann man den weiteren Beweis so führen, wie dies in Bemerkung 1 angedeutet wurde, wobei dann aber m eine beliebige komplexe Zahl sein kann.

Wir empfehlen dem Leser, nach dem Studium des Artikel „Elementare Funktionen einer komplexen Veränderlichen" auch die Sätze 2 bis 4 in den Bereich der komplexen Zahlen zu übertragen und die zugehörigen Beweise im einzelnen durchzuführen.

Kapitel V

DER ALLGEMEINE FUNKTIONSBEGRIFF

§ 54. Zuordnungen zwischen Mengen

Die Definition einer Funktion als einer Zuordnung zwischen zwei Zahlenmengen auf Grund einer vollständig beliebigen Vorschrift war bereits in § 1 erwähnt und in § 46 näher betrachtet worden. So allgemein diese Definition auch ist, sie ist doch in einer Hinsicht recht beschränkt und läßt noch eine bemerkenswerte Erweiterung zu. Nach der bisherigen Definition ordnet eine Funktion

$$y = f(x)$$

jeder Zahl x (jedem Argumentwert aus einem gewissen Intervall) eine gewisse andere Zahl y (den zugehörigen Funktionswert) zu. An Stelle der Zahlen x und y kann man nun aber Objekte ganz beliebiger Art betrachten. Dies führt dann auf den folgenden ganz allgemeinen Funktionsbegriff, der in der modernen Mathematik eine überragende Bedeutung besitzt:

Gegeben sei eine Menge von beliebigen Objekten, die wir mit $\mathfrak{X}$ bezeichnen wollen und deren Elemente durch den Buchstaben X (evtl. mit Indizes) gekennzeichnet werden sollen. Weiter sei eine mit $\mathfrak{Y}$ bezeichnete Menge von Objekten vorgegeben, deren Elemente wir mit Y (evtl. mit Indizes) bezeichnen wollen. Die Menge $\mathfrak{Y}$ kann dabei vollkommen oder auch teilweise aus Elementen bestehen, die in $\mathfrak{X}$ vorkommen, sie kann aber auch durchweg aus Elementen bestehen, die nicht in $\mathfrak{X}$ liegen. Unter einer *Abbildung* der Menge $\mathfrak{X}$ in die Menge $\mathfrak{Y}$ versteht man dann eine Vorschrift, durch welche jedem Element X der Menge $\mathfrak{X}$ ein oder mehrere Elemente der Menge $\mathfrak{Y}$ zugeordnet werden.

Hierbei wird im allgemeinen nicht vorausgesetzt, daß jedes Element der Menge $\mathfrak{Y}$ auch wirklich einem Element der Menge $\mathfrak{X}$ entspricht, d. h., es kann vorkommen, daß in der Menge $\mathfrak{Y}$ „freie" Elemente vorkommen, die keinem Element aus $\mathfrak{X}$ zugeordnet sind.

Es wird selbstverständlich auch nicht verlangt, daß verschiedenen Elementen X' und X'' aus $\mathfrak{X}$ stets verschiedene Elemente aus $\mathfrak{Y}$ entsprechen, d. h., es ist durchaus möglich, daß verschiedenen Elementen X' und X'' dasselbe Element aus $\mathfrak{Y}$ zugeordnet ist.

Ist jedem Element der Menge $\mathfrak{X}$ nur ein Element der Menge $\mathfrak{Y}$ zugeordnet, so nennt man die Abbildung *eindeutig;* ist dies nicht der Fall, d. h., gibt es wenigstens ein Element aus $\mathfrak{X}$, dem mehrere Elemente zugeordnet sind, so heißt die Abbildung *mehrdeutig*. Die eindeutigen Abbildungen werden auch *Funktionen* genannt (mitunter bezeichnet man allerdings auch mehrdeutige Abbildungen als Funktionen). Bezeichnet f eine Funktion im hier angegebenen Sinne, so versteht man für jedes X aus $\mathfrak{X}$ unter $f(X)$ das dem Element X

durch die betrachtete Funktion (Abbildung) zugeordnete Element aus $\mathfrak{Y}$. Durch die Schreibweise

$$Y = f(X)^{1}) \tag{1}$$

wird angedeutet, daß das Element Y aus $\mathfrak{Y}$ dem Element X aus $\mathfrak{X}$ durch die Funktion f zugeordnet wird (daß Y das Bild von X bei der Abbildung f ist).

Im folgenden werden wir nur eindeutige Abbildungen betrachten.

Eine Abbildung f von einer Menge $\mathfrak{X}$ auf eine Menge $\mathfrak{Y}$ heißt *umkehrbar eindeutig* oder *eineindeutig*, wenn jedes Element aus $\mathfrak{Y}$ einem gewissen Element aus $\mathfrak{X}$ zugeordnet ist und dabei je zwei verschiedenen Elementen aus $\mathfrak{X}$ je zwei verschiedene Elemente aus $\mathfrak{Y}$ entsprechen, d. h., wenn aus der Beziehung

$$f(X') = f(X'')$$

die Beziehung

$$X' = X''$$

folgt.

In diesem Fall wird durch die Abbildung (1) nicht nur jedem Element X aus $\mathfrak{X}$ ein eindeutig bestimmtes Element Y aus $\mathfrak{Y}$ zugeordnet, sondern es entspricht bei dieser Abbildung auch umgekehrt jedes Element Y aus $\mathfrak{Y}$ einem eindeutig bestimmten Element X aus $\mathfrak{X}$. Wir können also in diesem Fall die *inverse Abbildung* (*Umkehrfunktion*)

$$X = g(Y) \tag{2}$$

bilden, und diese ist eindeutig.[2])

Sind die betrachteten Mengen $\mathfrak{X}$ und $\mathfrak{Y}$ speziell Mengen von reellen Zahlen, so gelangen wir zum Begriff der *reellen Funktion einer reellen Veränderlichen*, wie er in § 46 betrachtet wurde.

In den folgenden Paragraphen werden wir eine Reihe von Beispielen für Abbildungen kennenlernen, bei denen die Mengen $\mathfrak{X}$ und $\mathfrak{Y}$ nicht Mengen von Zahlen, sondern Mengen anderer Natur (z. B. Mengen von Zahlenpaaren, Mengen von Zahlentripeln usw., d. h. Zahlensysteme, aber auch Mengen von Funktionen einer bestimmten Klasse oder dgl.) sein werden.

Hierbei werden wir zu gewissen geometrischen oder mechanischen (kinematischen) Interpretationen gelangen, die (wie im einfachsten Fall bei der „graphischen Darstellung" einer Funktion) zur Veranschaulichung des allgemeinen Funktionsbegriff dienen können, andererseits aber zugleich konkrete Beispiele für den allgemeinen Funktionsbegriff darstellen. Allerdings wird dieser Überblick recht oberflächlich sein, nicht nur wegen des beschränkten Umfanges, den wir ihm im vorliegenden Artikel widmen können, sondern auch deshalb, weil nicht damit zu rechnen ist, daß die folgenden Darlegungen eingehend in der Schule behandelt werden können.

---

[1]) Das Gleichheitszeichen (=) bezeichnet hier die Identität der Objekte Y und $f(X)$ aus $\mathfrak{Y}$.

[2]) Läßt man auch mehrdeutige Abbildungen zu, so kann man natürlich die inverse Abbildung generell bilden; wesentlich ist hier, daß für umkehrbar eindeutige Abbildungen (und auch nur für diese) die inverse Abbildung eindeutig ist. — *Anm. d. wissenschaftl. Red.*

Abschließend wollen wir noch einige Beispiele für „funktionale Zuordnungen" oder Abbildungen erwähnen, bei denen die Mengen $\mathfrak{X}$ und $\mathfrak{Y}$ beide aus endlich vielen Elementen bestehen. In einem derartigen Fall ist die Anzahl der möglichen eindeutigen Abbildungen ebenfalls endlich, und zwar ist sie gleich n^m, falls $\mathfrak{X}$ aus m und $\mathfrak{Y}$ aus n Elementen besteht.[1])

1. Es sei $\mathfrak{X}$ die Menge der Finger den linken Hand, $\mathfrak{Y}$ die Menge der Finger der rechten Hand. Insgesamt sind $5^5 = 3125$ eindeutige Abbildungen von $\mathfrak{X}$ in $\mathfrak{Y}$ möglich, von denen aber nur $5! = 120$ umkehrbar eindeutig sind.

2. Es sei $\mathfrak{X}$ die Menge der Schüler in einer Klasse (Anzahl: 30), $\mathfrak{Y}$ die Menge der Zensuren (1—5). Bei jeder Zensierung (z. B. einer Klassenarbeit) wird jedem Schüler (Element aus $\mathfrak{X}$) eindeutig eine Zensur (Element aus $\mathfrak{Y}$) zugeordnet. Insgesamt sind 5^{30} derartige Zuordnungen möglich.

3. Jeder von drei Freunden wünscht bzw. wünscht nicht an einer Exkursion teilzunehmen. Es sei $\mathfrak{X}$ die Menge der Freunde, $\mathfrak{Y}$ die aus den beiden Zeichen + (Teilnahme) und − (keine Teilnahme) bestehende Menge. Insgesamt sind $2^3 = 8$ Zuordnungen möglich.

4. Es sind m verschiedene Gegenstände ($\mathfrak{X}$) auf n Kisten ($\mathfrak{Y}$) zu verteilen. Dies ist auf n^m Arten möglich.

§ 55. Geometrische Gebilde in mehrdimensionalen Räumen

Es sei $\mathfrak{X}$ die Menge aller reellen Zahlen z (evtl. aus einem gewissen Intervall). Es seien ferner

$$\left.\begin{aligned} x &= f(z), \\ y &= g(z) \end{aligned}\right\} \tag{3}$$

reelle Funktionen (im üblichen Sinne). Es sei schließlich $\mathfrak{Y}$ die Menge aller Zahlenpaare (x, y). Wir betrachten dann die Abbildung, die jeder Zahl z das entsprechende Zahlenpaar $(f(z), g(z))$ zuordnet. Offenbar erhalten wir auf diese Weise eine eindeutige Abbildung der Menge $\mathfrak{X}$ in die Menge $\mathfrak{Y}$. Veranschaulichen wir uns die Verhältnisse in einem räumlichen x, y, z-Koordinatensystem im dreidimensionalen Raum, so erkennen wir, daß durch die Beziehungen (3) in jeder zur x, y-Ebene parallelen Ebene ein eindeutig bestimmter Punkt ausgezeichnet wird. Sind die Funktionen $f(z)$ und $g(z)$ stetig, so bildet die Menge aller dieser Punkte eine gewisse *Raumkurve*.

Der Leser möge beispielweise die *Schraubenlinie* (Abb. 90)

$$\left.\begin{aligned} x &= \cos z, \\ y &= \sin z \end{aligned}\right\} \tag{4}$$

[1]) Jedem der Elemente aus $\mathfrak{X}$ kann ein beliebiges Element aus $\mathfrak{Y}$ zugeordnet werden. Ist $m = 1$, so gibt es offenbar n derartige Abbildungen. Ist $m = 2$, so erhält man n^2 Abbildungen, da man jede der n Abbildungen des ersten Elementes mit jeder der n Abbildungen des zweiten Elementes kombinieren kann. Allgemein beweist man den Satz durch vollständige Induktion über m.

eingehend betrachten und sich veranschaulichen, wie bei Änderung von z von 0 bis 2π ein vollständiger Umlauf der Kurve um die z-Achse zustande kommt.

Als nächstes sei $\mathfrak{X}$ die Menge aller Zahlenpaare (x, y) oder, geometrisch, die Menge aller Punkte der x, y-Koordinatenebene bzw. eine gewisse Teilmenge hiervon (z. B. die Menge aller Punkte, die im Innern einer geschlossenen Kurve in der x, y-Ebene liegen bzw. auf einer Seite einer aus dem Unendlichen kommenden und ins Unendliche gehenden Kurve oder dgl.). Die Menge $\mathfrak{Y}$ sei die Menge aller reellen Zahlen z. In diesem Fall ist eine eindeutige Abbildung von $\mathfrak{X}$ in $\mathfrak{Y}$ nichts anderes als eine reelle Funktion von zwei reellen Veränderlichen:

$$z = f(x, y). \tag{5}$$

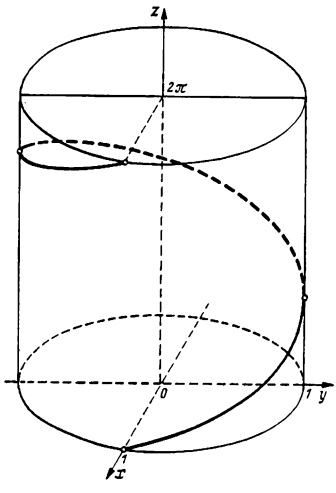


Abb. 90

Geometrisch bedeutet eine solche Funktion, daß im x, y, z-Raum auf jeder zur z-Achse parallelen Geraden, die durch einen Punkt der x, y-Ebene geht, der zu $\mathfrak{X}$ gehört, genau ein Punkt ausgezeichnet ist. Die Gesamtheit aller dieser Punkte bildet, falls die Funktion (5) eine gewisse Stetigkeitsbedingung erfüllt, eine *Fläche* im Raum.

Als Beispiel betrachten wir die Funktion

$$z = \sin \sqrt{x^2 + y^2}. \tag{6}$$

Berücksichtigt man, daß $\sqrt{x^2+y^2}$ der Abstand des Punktes (x, y) vom Koordinatenursprung O ist, so erkennt man, daß hier auf der durch den Punkt (x, y) gehenden Parallelen zur z-Achse derjenige Punkt ausgezeichnet ist, dessen Koordinate z gleich dem Sinus des genannten Abstandes ist. Hieraus ersieht man, daß die Menge der Punkte (x, y, z) des Raumes, welche die Gleichung (6) erfüllen, eine *Wellenfläche* darstellt (Abb. 91). Man kann diese Fläche offenbar dadurch erhalten, daß man die in der x, z-Ebene liegende Sinuskurve $z = \sin x$ $(0 \leqq x < \infty)$ um die z-Achse rotieren läßt.

Ein anderes Beispiel, dessen Untersuchung wir dem Leser empfehlen, ist die durch

$$z = \operatorname{Arctg} \frac{y}{x} \tag{7}$$

gegebene „Schraubenfläche" um die z-Achse.

In den bisherigen Beispielen war die Anzahl der auftretenden Veränderlichen gleich drei. Es sollen nun Zuordnungen betrachtet werden, bei denen mehr als drei Veränderliche vorkommen. Ist die Anzahl der Veränderlichen gleich vier, so hat man, wenn die Veränderlichen etwa mit x, y, z, u bezeichnet werden, drei wesentlich verschiedene Fälle zu unterscheiden:

1. Drei der Veränderlichen sind Funktionen der vierten, z. B.

$$\left.\begin{aligned} x &= f(u), \\ y &= g(u), \\ z &= h(u); \end{aligned}\right\} \tag{8}$$

in diesem Fall ist $\mathfrak{X}$ eine Menge von Zahlen und $\mathfrak{Y}$ ist die Menge aller[1]) Zahlentripel (x, y, z).

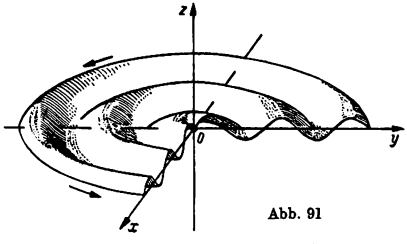


Abb. 91

2. Zwei der Veränderlichen sind Funktionen der anderen beiden, z. B.

$$\left.\begin{aligned} z &= f(x, y), \\ u &= f(x, y); \end{aligned}\right\} \tag{9}$$

in diesem Fall ist $\mathfrak{X}$ eine Menge von Zahlenpaaren (x, y) und $\mathfrak{Y}$ die Menge aller Zahlenpaare (z, u).

3. Eine der Veränderlichen ist Funktion der übrigen drei, z. B.

$$u = f(x, y, z); \tag{10}$$

in diesem Fall ist $\mathfrak{X}$ eine Menge von Zahlentripeln (x, y, z) und $\mathfrak{Y}$ die Menge aller reellen Zahlen.

Schließlich betrachten wir noch den in diesem Zusammenhang allgemeinsten Fall, wobei die Menge $\mathfrak{X}$ eine gewisse Menge von „p-Tupeln" $(x_1, \ldots, x_p)$ von Zahlen ist, die Menge $\mathfrak{Y}$ aus der Menge aller q-Tupel $(y_1, \ldots, y_q)$ von Zahlen besteht und die Zuordnung durch ein Gleichungssystem der Form

$$\left.\begin{aligned} y_1 &= f_1(x_1, x_2, \ldots, x_p), \\ y_2 &= f_2(x_1, x_2, \ldots, x_p), \\ &\cdots\cdots\cdots\cdots \\ y_q &= f_q(x_1, x_2, \ldots, x_p) \end{aligned}\right\} \tag{11}$$

beschrieben wird.

Um auch hier zu einer geometrischen Vorstellung zu gelangen, müßte man in den $(p+q)$-dimensionalen $(x_1, \ldots, x_p, y_1, \ldots, y_q)$-Raum gehen, in dem

[1]) Wir erinnern daran, daß nicht alle Elemente aus $\mathfrak{Y}$ als Bilder von Elementen aus $\mathfrak{X}$ aufzutreten brauchen. — *Anm. d. wissenschaftl. Red.*

durch das Gleichungssystem (11) eine gewisse p-dimensionale geometrische Figur bestimmt wird, die aus der Gesamtheit der „Punkte"

$$(x_1, x_2, \ldots, x_p, y_1, y_2, \ldots, y_q)$$

besteht, für welche die Gleichungen (11) simultan erfüllt sind.

Natürlich kann man auch geometrische Gebilde[1]) betrachten, die durch Beziehungen allgemeinerer Art beschrieben werden (z. B. „implizit" durch eine Gleichung der Form $F(x_1, \ldots, x_p, y_1, \ldots, y_q) = 0$). Hierbei muß man jedoch beachten, daß für gegebenes $(x_1, \ldots, x_p)$ keine „Lösung" $(y_1, \ldots, y_q)$ zu existieren braucht, während andererseits auch Mehrdeutigkeiten auftreten können. Als Beispiel erwähnen wir die Gleichung der „Kugel vom Radius R" im n-dimensionalen Raum:

$$x_1^2 + x_2^2 + \cdots + x_n^2 = R^2.$$

Diese Gleichung können wir zwar z. B. nach $x_n (= y_1)$ auflösen, wobei wir jedoch für $x_1^2 + x_2^2 + \cdots + x_{n-1}^2 \leqq R^2$ die doppeldeutige Funktion

$$y_1 = \pm\sqrt{R^2 - (x_1^2 + x_2^2 + \cdots + x_{n-1}^2)}$$

erhalten. Im betrachteten Fall ist $p = n - 1$ und $q = 1$.

Wir erwähnen zum Abschluß, daß ein und dasselbe Gebilde durch ganz verschiedene Gleichungssysteme beschrieben werden kann. So haben wir z. B. oben die Schraubenlinie betrachtet, die durch die Gleichungen (4) beschrieben wird. Geometrisch entsprechen die Gleichungen (4) der Gewinnung der Schraubenlinie durch Schnitt zweier sinusförmiger Zylinderflächen, deren Erzeugende parallel der y-Achse bzw. der x-Achse sind. Die Schraubenlinie läßt sich aber auch durch die Gleichungen

$$\left.\begin{aligned} x^2 + y^2 &= 1, \\ z &= \operatorname{Arctg}\frac{y}{x} \end{aligned}\right\} \tag{12}$$

erfassen, und dies entspricht geometrisch der Gewinnung der Schraubenlinie durch Schnitt eines Kreiszylinders um die z-Achse mit der oben betrachteten Schraubenfläche.

§ 56. Räumliche Abbildungen

Häufig ist es möglich, analytische Zusammenhänge geometrisch darzustellen, wobei man im allgemeinen noch eine gewisse Freiheit bei der Wahl der Interpretation hat: gegebene analytische Zusammenhänge können unter Umständen auf ganz verschiedene Weisen geometrisch dargestellt werden.

[1]) Der Begriff „Gebilde" dient hier als Zusammenfassung für alle möglichen Typen von geometrischen Figuren im n-dimensionalen Raum. Die „Dimension" eines solchen Gebildes ist dabei gleich der Anzahl der unabhängigen Veränderlichen. Unter „Kurven" versteht man gewöhnlich eindimensionale Gebilde, während man $(n-1)$-dimensionale Gebilde als Flächen oder Hyperflächen bezeichnet.

Das Verhalten einer Funktion

$$y = f(x) \tag{13}$$

kann man sich z. B., wie wir dies im vorangehenden bereits häufig getan haben, dadurch veranschaulichen, daß man in der x, y-Ebene alle diejenigen Punkte markiert, deren Abszisse x und deren Ordinate y in der Beziehung (13) stehen.

Neben dieser Darstellung im (x, y)-Koordinatensystem kann man jedoch auch eine Darstellung folgender Art betrachten: Vorgegeben sind zwei beliebige

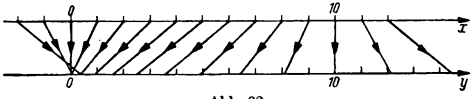


Abb. 92

Geraden als x- bzw. y-Achse; durch die Funktion (13) wird dann jedem Punkt der x-Achse ein eindeutig bestimmter Punkt der y-Achse zugeordnet. In Abb. 92 ist eine derartige Darstellung für die Funktion $y = \frac{1}{10} x^2$ skizziert. Einander entsprechende Punkte sind dabei durch Pfeile verbunden.

Genauso kann man sich jede Abbildung der Form (3) dadurch veranschaulichen, daß man jedem Punkt einer beliebig gelegenen z-Achse (oder eines gewissen Intervalls oder dgl. dieser Achse) den entsprechenden Punkt auf einer beliebig gelegenen x, y-Ebene zuordnet. In Abb. 93 ist auf diese Weise die Zuordnung (4) dargestellt, wobei in diesem Fall das Intervall $0 \leqq z < 2\pi$ und allgemein jedes Intervall $2k\pi \leqq z < 2(k+1)\pi$ auf den Kreis $x^2 + y^2 = 1$ abgebildet wird.

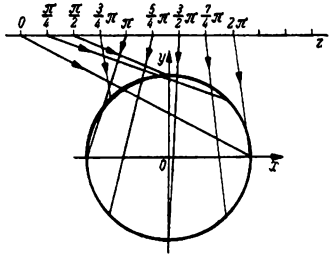


Abb. 93

Entsprechend kann man jede Funktion der Form (5) als eine Zuordnung zwischen den Punkten einer x, y-Ebene (oder eines Teils derselben) und denen einer z-Achse interpretieren. In Abb. 94 ist auf diese Weise die Funktion (6) dargestellt; hierbei gehen alle Punkte, die auf einem gegebenen Kreis um den Nullpunkt der x, y-Ebene liegen, in denselben Punkt der z-Achse über, und zwar wird z. B. allen Punkten des Kreises $x^2 + y^2 = \frac{\pi^2}{4}$ der Punkt $z = 1$ zugeordnet, während die Punkte, die im Innern dieses Kreises liegen, auf die Punkte des Intervalls $0 \leqq z < 1$ abgebildet werden.

Auf ähnliche Weise lassen sich im Fall von vier Veränderlichen die Abbildungen vom Typ (8), (9) und (10) der Reihe nach als Abbildungen einer

Geraden auf den Raum, einer Ebene auf eine Ebene und des Raumes auf eine Gerade interpretieren.

Allgemein kann man eine Abbildung vom Typ (11) als Abbildung des „p-dimensionalen $(x_1, \ldots, x_p)$-Raumes" auf den „q-dimensionalen $(y_1, \ldots, y_q)$-Raum" deuten.

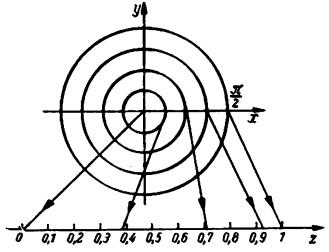


Abb. 94

Wir betrachten nun einige Beispiele für Abbildungen der genannten Art. Auf eine formelmäßige Beschreibung der Abbildungen wird dabei nicht unbedingt Wert gelegt, und zwar deshalb nicht, weil diese entweder sehr einfach ist oder aber eine genaue oder näherungsweise formelmäßige Wiedergabe mehr oder weniger prinzipielle oder praktische Schwierigkeiten bereitet.

1. Die Spiegelungen, Parallelverschiebungen, Drehungen, Dehnungen und Stauchungen (vgl. die Vielzahl von Beispielen in Kapitel I) sind spezielle umkehrbar eindeutige Abbildungen einer Ebene auf eine Ebene bzw. des Raumes auf den Raum.

2. Wir betrachten die folgende Abbildung einer Ebene auf eine Ebene (im vorliegenden Falle sogar der Ebene auf sich): Gegeben ist ein Kreis mit dem Mittelpunkt O und dem Radius R (Abb. 95); jedem Punkt $M (\neq O)$ wird derjenige Punkt M' der Halbgeraden OM zugeordnet, für den

$$OM \cdot OM' = R^2$$

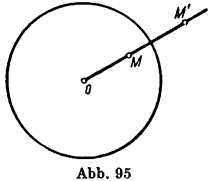


Abb. 95

gilt. Man sieht leicht, daß diese Abbildung umkehrbar eindeutig ist. Sie wird als *Inversion (am Kreis)* oder — im Fall $R = 1$ — als *Abbildung durch reziproke Radien* bezeichnet und besitzt eine Reihe von sehr bemerkenswerten Eigenschaften, auf die wir hier jedoch nicht eingehen können. In ähnlicher Weise kann man im Raum eine *Inversion an der Kugel* erklären.

Ist $R = 1$ und O der Ursprung des Koordinatensystems, so besitzt der betrachtete Kreis die Gleichung

$$x^2 + y^2 = 1 .$$

Bezeichnen wir mit x, y die Koordinaten des Punktes M und mit x', y' die des Punktes M', so gilt

$$\begin{cases} x' = \dfrac{x}{x^2 + y^2}, \\ y' = \dfrac{y}{x^2 + y^2}, \end{cases} \quad \text{oder} \quad \begin{cases} x = \dfrac{x'}{x'^2 + y'^2}, \\ y = \dfrac{y'}{x'^2 + y'^2}. \end{cases}$$

Dies ist eine analytische Beschreibung der Inversion.

3. Wir stellen uns ein Gefäß (V) vor, welches mit einer in Ruhe befindlichen (idealen) Flüssigkeit gefüllt sei; sodann denken wir uns die Flüssigkeit umgerührt und den Moment gekommen, in dem die Flüssigkeit wieder in Ruhe ist. Ordnen wir jedem Flüssigkeitsteilchen vor dem Umrühren dasselbe Teilchen nach dem Umrühren zu, so erhalten wir eine Abbildung des Teilraums (V) auf sich.

4. Bewegt sich ein Punkt im Laufe eines gewissen Zeitintervalls $t_1 \leqq t \leqq t_2$ a) auf der x-Achse, b) in der x, y-Ebene, c) im x, y, z-Raum, so kann man die Bewegung des Punktes durch eine Gleichung bzw. ein Gleichungssystem der Form

$$\text{a) } x = f(t), \quad \text{b) } \begin{cases} x = f(t), \\ y = g(t), \end{cases} \quad \text{c) } \begin{cases} x = f(t), \\ y = g(t), \\ z = h(t) \end{cases} \tag{14}$$

beschreiben. Dadurch wird das Zeitintervall $t_1 \leqq t \leqq t_2$ auf die Gerade, die Ebene bzw. den Raum abgebildet. Diese Abbildung ist natürlich im allgemeinen nicht umkehrbar eindeutig, da der Punkt zu verschiedenen Zeitpunkten dieselbe Lage besitzen kann.

Hier einige spezielle Beispiele dafür:

1. $\begin{cases} x = \cos t, \\ y = \sin t \end{cases}$ (gleichförmige Bewegung des Punktes auf dem Kreis $x^2 + y^2 = 1$);

2. $\begin{cases} x = \mathfrak{Cof}\, t, \\ y = \mathfrak{Sin}\, t \end{cases}$ (vgl. § 21, gleichförmige Bewegung des Punktes auf der Hyperbel $x^2 - y^2 = 1$);[1])

3. $\begin{cases} x = \cos t, \\ y = \sin t, \\ z = t \end{cases}$ (Bewegung des Punktes auf einer Schraubenlinie im Raum).

5. Es sei u irgendeine skalare physikalische Größe (Temperatur, atmosphärischer Druck usw.). Dann wird die Verteilung der Größe u a) auf der x-Achse, b) in der x, y-Ebene, c) im x, y, z-Raum (oder in irgendwelchen Teilen dieser Bereiche) durch eine Gleichung der Form

$$\text{a) } u = f(x), \quad \text{b) } u = f(x, y), \quad \text{c) } u = f(x, y, z)$$

beschrieben. Gleichzeitig wird dadurch eine Abbildung der x-Achse, der x, y-Ebene, des x, y, z-Raumes auf die u-Achse definiert. Die Gesamtheit der Punkte, in welchen u einen vorgegebenen Wert annimmt (d. h. die Gesamtheit der Punkte, denen ein gegebener Punkt der u-Achse entspricht) wird im Fall b) als *Niveaukurve* und im Fall c) als *Niveaufläche* bezeichnet; ist u die Temperatur, so sind dies die *Isothermen*, ist u der atmosphärische Druck, so sind dies die *Isobaren* usw.

Die folgenden Beispiele haben etwas allgemeineren Charakter, und zwar handelt es sich dabei um Abbildungen von Kurven auf Kurven bzw. Flächen auf Flächen.

---

[1]) Daher die Bezeichnung „Hyperbelfunktionen".

6. Es sei (c) ein Kreis, der im Innern eines anderen Kreises (C) liegt, und O ein Punkt im Innern von (c). Wir betrachten die Abbildung, welche jedem Punkt m des Kreises (c) denjenigen Punkt M des Kreises (C) zuordnet, der auf der Halbgeraden Om liegt. Man erhält auf diese Weise eine umkehrbar eindeutige Abbildung des Kreises (c) auf den Kreis (C).

Entsprechend kann man im Raum mit zwei Kugeln verfahren.

7. Wir umwickeln einen Zylinder mit einem Blatt Papier, das sich nicht überdeckt, und ordnen jedem Punkt der Mantelfläche des Zylinders denjenigen Punkt auf dem Blatt Papier zu, der direkt auf ihm liegt. Breiten wir das Blatt Papier in der Ebene aus, so erhalten wir damit eine umkehrbar eindeutige Abbildung der Mantelfläche des Zylinders auf die Ebene, bei der jede auf der Zylindermantelfläche gelegene Kurve in eine ebene Kurve gleicher Länge übergeht.

Entsprechend kann man mit einem Kegel verfahren.

Die genannte Abbildung bezeichnet man als *Abwicklung* des Zylinders bzw. des Kegels auf die Ebene.

8. Dasselbe kann man z. B. mit der Kugel nicht machen. In vielen Fällen liegt jedoch ein ausgesprochenes Bedürfnis vor, die Oberfläche einer Kugel (in der Kartographie z. B. der Erdkugel) oder einen Teil derselben auf die Ebene abzubilden. Das ist auf viele verschiedene Weisen möglich, wobei man der Abbildung noch verschiedene Bedingungen (z. B. Winkeltreue, Längentreue usw.) auferlegen kann.

Die wohl einfachste Methode besteht darin, daß man die geographische Länge u und die geographische Breite v der Punkte der Erdoberfläche als rechtwinklige Koordinaten, d. h. als Abszisse und Ordinate des Bildpunktes interpretiert. Wir empfehlen dem Leser, sich nach diesem Prinzip auf Millimeterpapier eine Karte von Europa anzufertigen.

§ 57. Metrische Räume

In § 38 wurde eine Definition des Grenzwertes einer Zahlfolge gegeben, bei der nur vom Begriff der Umgebung eines Punktes (auf der Zahlengeraden) Gebrauch gemacht wurde; unter der Umgebung eines Punktes war dabei die Gesamtheit aller der Punkte zu verstehen, deren Abstand (Entfernung) vom betrachteten Punkt kleiner als eine vorgegebene Zahl ist.

In ganz entsprechender Weise kann man nun den Begriff des Grenzwertes auch für Folgen von Elementen einer ganz beliebigen Menge erklären, sofern man vorher vereinbart, was unter dem „Abstand" zwischen je zwei Elementen dieser Menge verstanden werden soll. Mit anderen Worten: Es muß in der betrachteten Menge zunächst eine „Metrik" festgelegt werden.

Unter dem Abstand zweier Zahlen x' und x'' versteht man natürlich (vgl. § 36) den absoluten Betrag ihrer Differenz, also $|x' - x''|$, d. h. den Abstand im elementargeometrischen Sinne zwischen den entsprechenden Punkten M' und M'' auf der Zahlengeraden:

$$M'M'' = |x' - x''|. \tag{15}$$

Betrachtet man dagegen zwei Zahlenpaare (x', y') und (x'', y''), so kann man unter Berücksichtigung der üblichen Deutung solcher Zahlenpaare als Punkte einer (kartesischen) x, y-Koordinatenebene unter ihrem Abstand die Entfernung der entsprechenden Punkte M' und M'' verstehen, die bekanntlich durch

$$M'M'' = \sqrt{(x'-x'')^2+(y'-y'')^2} \tag{16}$$

gegeben wird.

Handelt es sich um Zahlentripel (x', y', z') und (x'', y'', z''), so erhält man bei Deutung als Punkte in einem räumlichen x, y, z-Koordinatensystem die Entfernung zwischen den entsprechenden Punkten M' und M'' gemäß

$$M'M'' = \sqrt{(x'-x'')^2+(y'-y'')^2+(z'-z'')^2}. \tag{17}$$

Allgemein kann man sich bei Betrachtung von n-Tupeln $(x_1', \ldots, x_n')$ und $(x_1'', \ldots, x_n'')$ auf den sogenannten n-dimensionalen Raum beziehen, in dem die Entfernung der zugehörigen Punkte M' und M'' durch

$$M'M'' = \sqrt{(x_1'-x_1'')^2+(x_2'-x_2'')^2+\cdots+(x_n'-x_n'')^2} \tag{18}$$

oder in kürzerer Schreibweise

$$M'M'' = \sqrt{\sum_{m=1}^{n}(x_m'-x_m'')^2} \tag{19}$$

definiert ist.[1])

In § 36 haben wir sodann die ε-Umgebung eines Punktes x_0 der x-Achse durch die Ungleichung

$$|x-x_0| < \varepsilon$$

definiert, was gerade auf das offene Intervall

$$x_0-\varepsilon < x < x_0+\varepsilon$$

führt. Entsprechend verstehen wir in der x, y-Ebene unter der ε-Umgebung des Punktes (x_0, y_0) die Menge aller Punkte (x, y), für die

$$\sqrt{(x-x_0)^2+(y-y_0)^2} < \varepsilon$$

gilt. Geometrisch bedeutet dies das Innere des Kreises um den Punkt (x_0, y_0) vom Radius ε:

$$(x-x_0)^2+(y-y_0)^2 < \varepsilon^2.$$

Entsprechend betrachtet man im dreidimensionalen x, y, z-Raum als ε-Umgebung eines Punktes (x_0, y_0, z_0) das Innere der Kugel um (x_0, y_0, z_0) vom Radius ε:

$$(x-x_0)^2+(y-y_0)^2+(z-z_0)^2 < \varepsilon^2,$$

während im n-dimensionalen $(x_1, \ldots, x_n)$-Raum unter der ε-Umgebung des Punktes $(x_1^0, \ldots, x_n^0)$ das Innere der „Hyperkugel" um $(x_1^0, \ldots, x_n^0)$ vom Radius ε verstanden wird:

$$(x_1-x_1^0)^2+(x_2-x_2^0)^2+\cdots+(x_n-x_n^0)^2 < \varepsilon^2.$$

[1]) Im Fall $n = 1$ nimmt (18) offenbar gerade die Form (15) an.

Es muß bemerkt werden, daß die durch die Formeln (15) bis (19) definierten Metriken nur durch geometrische Veranschaulichungen nahegelegt werden, daß sie aber in keiner Weise logisch zwingend sind. Man bezeichnet sie als *euklidische Metriken*. Für viele Untersuchungen erweist es sich als durchaus zweckmäßig, andere Metriken zu betrachten.

Wir wollen dies an einem Beispiel erläutern: Dazu betrachten wir zunächst den „Abstand" zweier Punkte einer Kugel, etwa der Erdkugel. Im Sinne der euklidischen Metrik wäre dies die Länge ihrer Verbindungsstrecke, d. h. der durch die Punkte bestimmten Sehne. Demgegenüber mißt man den Abstand zweier Punkte der Erdoberfläche auf der Oberfläche selbst, und zwar nimmt man als Abstand die Bogenlänge auf dem Großkreis, der durch die betrachteten Punkte hindurchgeht (dies ergibt den „kürzesten" Abstand auf der Kugel). Wir wollen nun annehmen, daß uns die beiden Punkte M' und M'' durch ihre „geographischen Koordinaten", die Länge u und die Breite v, gegeben sind. Weiter wollen wir voraussetzen, daß wir über eine Karte verfügen, auf der die Punkte (u, v) durch Punkte mit den kartesischen Koordinaten u (Abszisse) und v (Ordinate) dargestellt werden (vgl. Beispiel 8 des vorangehenden Paragraphen). In diesem Fall ergibt $\sqrt{(u'-u'')^2+(v'-v'')^2}$ den Abstand zwischen zwei Punkten auf der Karte, nicht jedoch (von trivialen Maßstabsumrechnungen abgesehen) auf der Erdoberfläche. Der Abstand der Punkte auf der Erdoberfläche ist vielmehr eine komplizierte Funktion der vier Koordinaten u', v', u'', v'', auf die wir hier nicht eingehen wollen.

An Stelle der Kugel können wir natürlich auch irgendeine andere gekrümmte Fläche im Raum nehmen, auf der in beliebiger Weise ein „krummliniges" Koordinatennetz aufgetragen ist. Als Abstand zweier Punkte M' und M'' auf einer derartigen Fläche nimmt man gewöhnlich die Länge der kürzesten Verbindung von M' und M'' auf der Fläche.[1]) Dieser Abstand ist eine mehr oder weniger komplizierte Funktion der Koordinaten von M' und M'' in bezug auf das gegebene krummlinige Koordinatensystem. Bilden wir die Fläche dadurch auf die Ebene ab, daß wir die krummlinigen Koordinaten eines Punktes der Fläche als kartesische Koordinaten in der Ebene deuten, so werden wir nur in ganz seltenen Spezialfällen erreichen, daß der Abstand auf der Fläche mit dem euklidischen Abstand in der Ebene übereinstimmt.

Wir wollen uns nun wieder der allgemeinen Theorie zuwenden. Man sagt, in einer Menge (einem „Raum") $\mathfrak{X}$ sei eine *Metrik* eingeführt, wenn je zwei Elementen X' und X'' aus $\mathfrak{X}$ eindeutig eine Zahl $\varrho(X', X'')$ als *Abstand* so zugeordnet ist, daß folgende Bedingungen erfüllt sind:

1. $$\varrho(X', X'') = \varrho(X'', X'). \tag{20}$$

Dies bedeutet, daß der von X' nach X'' gemessene Abstand gleich dem von X'' nach X' gemessenen Abstand ist, so daß man also vom Abstand zwischen (bzw. von) X' und X'' sprechen kann.

2. $$\left.\begin{aligned} &\varrho(X', X'') > 0, \text{ falls } X' \text{ und } X'' \text{ verschiedene Elemente aus } \mathfrak{X} \text{ sind.} \\ &\varrho(X', X'') = 0, \text{ falls } X' \text{ und } X'' \text{ übereinstimmen.} \end{aligned}\right\} \tag{21}$$

[1]) Da es auf der Fläche unendlich viele Verbindungen je zweier Punkte gibt, bedarf die Existenz einer kürzesten Verbindungslinie (einer sog. *geodätischen Linie*) eines Beweises.

Hieraus folgt, daß der Abstand zweier Elemente niemals negativ ist und daß er dann und nur dann verschwindet, wenn X' und X'' dasselbe Element aus $\mathfrak{X}$ bezeichnen.

3. Für beliebige Elemente X', X'' und X''' aus $\mathfrak{X}$ gilt

$$\varrho(X',\ X'') \leqq \varrho(X',\ X''') + \varrho(X''',\ X''). \tag{22}$$

Diese sogenannte *Dreiecksungleichung* besagt, daß der unmittelbar von X' nach X'' gemessene Abstand die Summe des von X' zu irgendeinem dritten Element X''' genommenen Abstandes und des von X''' bis X'' gemessenen Abstandes niemals übertrifft.

Beispiel. Ist $\mathfrak{X}$ der „Raum" aller Paare (x, y) von Zahlen (d. h. die Menge aller Punkte der x, y-Ebene), so haben wir bereits oben den Abstand der Punkte $M' = (x', y')$ und $M'' = (x'', y'')$ durch

$$\varrho(M',\ M'') = \sqrt{(x'-x'')^2 + (y'-y'')^2} \tag{23}$$

definiert. Der Leser prüft leicht nach, daß hierfür wirklich die angegebenen Bedingungen 1. bis 3. erfüllt sind; einige Schwierigkeiten bereitet nur der formale Beweis von 3., der aber ebenfalls völlig elementar geführt werden kann.[1])

Ebenso sind aber auch z. B. für

$$\varrho(M',\ M'') = |x'-x''| + |y'-y''| \tag{24}$$

und

$$\varrho(M',\ M'') = \max\{|x'-x''|,\ |y'-y''|\} \tag{25}$$

die angegebenen Bedingungen 1. bis 3. erfüllt. Der Leser mache sich klar, welche geometrische Bedeutung der durch (24) und (25) definierte „Abstand" besitzt.

Ist in einem Raum $\mathfrak{X}$ eine Metrik eingeführt, so ist damit auch der Begriff der *Umgebung* für jedes Element festgelegt: Unter der *ε-Umgebung* eines Elementes X_0 aus $\mathfrak{X}$ versteht man die Menge aller derjenigen Elemente X aus $\mathfrak{X}$, für welche die Ungleichung

$$\varrho(X,\ X_0) < \varepsilon$$

erfüllt ist.

Beispiel. Je nachdem, ob man in der x, y-Ebene die durch (23) oder (24) oder (25) definierte Metrik betrachtet, wird die ε-Umgebung eines Punktes (x_0, y_0) durch die Ungleichungen

$$(x-x_0)^2 + (y-y_0)^2 < \varepsilon^2$$

(Inneres des Kreises um (x_0, y_0) vom Radius ε),

$$\max\{|x-x_0|,\ |y-y_0|\} < \varepsilon$$

---

[1]) Dies führt auf die sogenannte MINKOWSKIsche Ungleichung

$$\sqrt{(x'-x'')^2 + (y'-y'')^2} \leqq \sqrt{(x'-x''')^2 + (y'-y''')^2} + \sqrt{(x'''-x'')^2 + (y'''-y'')^2}.$$

(Inneres des Quadrates mit dem Mittelpunkt (x_0, y_0) und der Seitenlänge 2ε, wobei die Seiten parallel den Koordinatenachsen sind) oder

$$|x - x_0| + |y - y_0| < \varepsilon$$

(Inneres des Quadrates mit dem Mittelpunkt (x_0, y_0) und den Eckpunkten $(x_0 + \varepsilon, y_0)$, $(x_0, y_0 + \varepsilon)$, $(x_0 - \varepsilon, y_0)$ und $(x_0, y_0 - \varepsilon)$) gegeben.

Im vorangehenden haben wir vorwiegend Räume betrachtet, deren Elemente Zahlen bzw. Zahlensysteme (Paare, Tripel usw.) waren. Man kann jedoch auch Räume betrachten, deren Elemente Geraden, Kreise, Ebenen, irgendwelche Kurven oder Flächen usw. sind. Auch in derartigen „Räumen" kann man Metriken einführen, die die angegebenen Bedingungen 1. bis 3. erfüllen. Wir beschränken uns hier auf das Studium gewisser spezieller Beispiele von „Funktionenräumen", d. h. von Räumen, deren Elemente Funktionen sind, wobei wir zeigen wollen, wie man auch in derartigen Räumen eine Metrik einführen kann.

Zunächst betrachten wir als Elemente des zu konstruierenden Funktionenraumes die Menge aller Funktionen $f(x)$ einer unabhängigen Veränderlichen x, die in einem gegebenen Intervall I ($a \leq x \leq b$) definiert sind, wobei wir uns der Einfachheit halber auf die in I stetigen Funktionen beschränken wollen. Diese Menge werde mit $\mathfrak{X}$ bezeichnet. Wie können wir nun in $\mathfrak{X}$ eine Metrik erklären? Wir haben bereits früher (vgl. § 48) den durch

$$\varrho(f, g) = \max_I \{|f(x) - g(x)|\} \tag{26}$$

definierten Abstand zweier Funktionen $f(x)$ und $g(x)$ kennengelernt. Geometrisch bedeutet dieser Abstand die maximale Länge der Strecken PQ (vgl. Abb. 87) für x aus I.

Es sei hier erwähnt, daß man als „Abstand" $\varrho(f, g)$ auch eines der Integrale

$$\int_a^b |f(x) - g(x)|\, dx \tag{27}$$

(Flächeninhalt des schraffierten „Streifens" in Abb. 87) oder

$$\int_a^b [f(x) - g(x)]^2\, dx \tag{28}$$

nehmen kann. (Der durch (28) gegebene „Abstand" ist besonders zweckmäßig, da hierbei die meisten Rechnungen sehr einfach werden.)

Der Nachweis der Bedingungen 1. bis 3. für den durch (26) gegebenen „Abstand", bereitet keine Schwierigkeiten, ebenso der Nachweis von 1. und 2. für den durch (27) und (28) gegebenen „Abstand", während zum Beweis von 3. für (27) und (28) die Kenntnis einiger tieferliegender Eigenschaften des bestimmten Integrals notwendig ist.[1])

[1]) Es besteht im Prinzip eine sehr wichtige formale Analogie zwischen den Metriken (26), (27) und (28) im Funktionenraum und den Metriken (25), (24) und (23) im (zweidimensionalen) Punktraum.

§ 58. Der Begriff des Grenzwertes in metrischen Räumen

Gegeben sei ein beliebiger metrischer Raum $\mathfrak{X}$. Wir betrachten eine Folge

$$\{A_n\} \equiv A_1,\ A_2,\ \ldots,\ A_n,\ \ldots \tag{29}$$

von Elementen aus $\mathfrak{X}$. Ein Element A aus $\mathfrak{X}$, für das es zu jeder noch so kleinen positiven Zahl ε unendlich viele Glieder der Folge $\{A_n\}$ gibt, welche die Ungleichung

$$\varrho(X,\ A) < \varepsilon \tag{30}$$

erfüllen, nennt man einen *Häufungspunkt* (oder *Häufungswert*) dieser Folge. Genügen für jedes $\varepsilon > 0$ sogar fast alle Glieder der Folge, d. h. alle von einem gewissen Glied an (alle mit Ausnahme von endlich vielen) der Ungleichung (30), so nennt man A einen *Grenzwert* der Folge $\{A_n\}$; man schreibt dafür:

$$A_n \to A \quad \text{oder} \quad \lim A_n = A.$$

Jede Folge von Elementen eines metrischen Raumes besitzt höchstens einen Grenzwert.

In der Tat: Wenn $A_n \to A$ und $A_n \to A'$, so gelten für jedes $\varepsilon > 0$ bei hinreichend großem n die Ungleichungen

$$\varrho(A_n,\ A) < \varepsilon \quad \text{und} \quad \varrho(A_n,\ A') < \varepsilon,$$

und daraus folgt (wegen der Bedingungen 3. und 1.)

$$\varrho(A,\ A') \leqq \varrho(A,\ A_n) + \varrho(A_n,\ A') = \varrho(A_n,\ A) + \varrho(A_n,\ A') < \varepsilon + \varepsilon = 2\varepsilon.$$

Da hierbei ε beliebig klein sein kann, gilt also

$$\varrho(A,\ A') = 0,$$

und daher stimmen (nach Bedingung 2.) A und A' überein.

Eine Folge $\{A_n\}$, die einen Grenzwert besitzt, heißt *konvergent* (gegen diesen Grenzwert).

Eine Folge $\{A_n\}$ heißt *beschränkt*, wenn für ein gewisses A' aus $\mathfrak{X}$ die Folge $\{\varrho(A_n, A')\}$ beschränkt ist. Man sieht unmittelbar, daß dies unabhängig von der speziellen Auswahl des Elementes A' ist, d. h., mit $\{\varrho(A_n, A')\}$ ist auch jede andere Folge $\{\varrho(A_n, A'')\}$ (A'' aus $\mathfrak{X}$) beschränkt.

In der Tat: Ist z. B.

$$\varrho(A_n,\ A') < M \qquad (n = 1,\ 2,\ 3, \ldots)$$

und

$$\varrho(A',\ A'') = m,$$

so ist (wegen 3.)

$$\varrho(A_n,\ A'') \leqq \varrho(A_n,\ A') + \varrho(A',\ A'') < M + m.$$

Es ergibt sich sofort die Frage, ob in einem beliebigen metrischen Raum $\mathfrak{X}$ jede beschränkte Folge $\{A_n\}$ unbedingt wenigstens einen Häufungspunkt besitzt, d. h., ob der Satz von Bolzano-Weierstrass (§ 36) in beliebigen metrischen Räumen gilt.

Diese Frage ist zu verneinen. Zum Beispiel gilt der Satz von BOLZANO-WEIERSTRASS nicht in dem mit der Metrik (26) versehenen Raum der im Intervall $I\,(a \leqq x \leqq b)$ stetigen Funktionen. Um dies einzusehen, betrachten wir die Folge

$$\{\sin nx\}. \tag{31}$$

Diese Folge ist beschränkt; denn wählen wir als „Vergleichselement" A' die Funktion $f(x) \equiv 0$, so erhalten wir

$$\varrho(\sin nx,\ 0) = \max|\sin nx| \leqq 1.$$

Dagegen besitzt, wie man leicht feststellt, die Folge (31) in keinem Intervall I auch nur einen Häufungspunkt.

Eine Menge $\mathfrak{A}$ von Elementen eines metrischen Raumes $\mathfrak{X}$ heißt *kompakt* (in $\mathfrak{X}$), wenn jede Folge von Elementen aus $\mathfrak{A}$ in $\mathfrak{X}$ einen Häufungspunkt besitzt.

Jede kompakte Menge $\mathfrak{A}$ ist beschränkt.

Wäre dies nämlich nicht der Fall, so könnte man aus $\mathfrak{A}$ eine Folge $\{A_n\}$ auswählen, für die

$$\varrho(A_n,\ A') \to \infty \tag{32}$$

gilt (wobei A' ein beliebiges Element aus $\mathfrak{A}$ ist). Andererseits besitzt aber die Folge $\{A_n\}$, weil $\mathfrak{A}$ kompakt ist, wenigstens einen Häufungspunkt A. Daher gilt bei beliebig vorgegebenem $\varepsilon > 0$ für unendlich viele Glieder der Folge $\{A_n\}$ die Ungleichung

$$\varrho(A_n,\ A) < \varepsilon.$$

Für alle diese Elemente gilt jedoch auf Grund der Dreiecksungleichung

$$\varrho(A_n,\ A') \leqq \varrho(A_n,\ A) + \varrho(A,\ A') < \varepsilon + \varrho(A,\ A') \equiv M,$$

im Widerspruch zu (32).

Das oben angegebene Beispiel zeigt, daß die Umkehrung dieses Satzes nicht allgemein richtig ist, d. h., daß es in gewissen metrischen Räumen beschränkte Mengen gibt, die nicht kompakt sind. Diese Umkehrung gilt jedoch, wie wir bereits früher gesehen haben (§ 36), im eindimensionalen „Punktraum". Sie gilt auch, und das bestätigt man leicht, in jedem anderen endlichdimensionalen „Punktraum".

Wir wollen die zuletzt ausgesprochene Behauptung für den Fall der Ebene beweisen; die Verallgemeinerung auf den n-dimensionalen Raum ist dann vollständig trivial.

Vorgegeben sei eine beschränkte Punktfolge

$$\{P_n(x_n,\ y_n)\}.$$

Nehmen wir als „Vergleichspunkt" A' den Koordinatenursprung O, so finden wir, daß es eine Zahl M gibt derart, daß für alle Punkte P_n die Ungleichung

$$\varrho(P_n,\ O) = \sqrt{x_n^2 + y_n^2} < M \qquad (n = 1,\ 2,\ 3, \ldots)$$

erfüllt ist, und hieraus folgt

$$|x_n| < M \qquad (n = 1,\ 2,\ 3, \ldots). \tag{33}$$

und

$$|y_n| < M \tag{34}$$

Insbesondere ist also wegen (33) die Folge $\{x_n\}$ beschränkt und besitzt daher einen Häufungspunkt ξ. Für jedes $\varepsilon > 0$ liegen also im Intervall

$$\xi - \frac{\varepsilon}{\sqrt{2}} < x < \xi + \frac{\varepsilon}{\sqrt{2}} \tag{35}$$

unendlich viele Glieder der Folge $\{x_n\}$, und zwar mögen dies die Glieder

$$x_{p_1}, x_{p_2}, \ldots, x_{p_n}, \ldots$$

sein. Wir betrachten nun die Zahlenfolge

$$y_{p_1}, y_{p_2}, \ldots, y_{p_n}, \ldots .$$

Diese ist wegen (34) beschränkt und besitzt daher einen Häufungspunkt η. Im Intervall

$$\eta - \frac{\varepsilon}{\sqrt{2}} < y < \eta + \frac{\varepsilon}{\sqrt{2}} \tag{36}$$

liegen dann unendlich viele Glieder der Zahlenfolge $\{y_{p_n}\}$. Hieraus folgt nun, daß in dem durch die Ungleichungen (35) und (36) festgelegten Bereich der x, y-Ebene unendlich viele Punkte der Folge $\{P_n\}$ liegen. Dieser Bereich ist aber, wie man unmittelbar sieht, ganz in der ε-Umgebung

$$(x-\xi)^2 + (y-\eta)^2 < \varepsilon^2$$

des Punktes (ξ, η) enthalten. Da hierbei die Zahl ε beliebig klein sein kann, ist (ξ, η) ein Häufungspunkt der Folge $\{P_n\}$, was zu beweisen war.

In einem kompakten metrischen Raum $\mathfrak{X}$ ist ein Element A dann und nur dann Grenzwert einer Folge $\{A_n\}$, wenn A einziger Häufungspunkt dieser Folge ist.

Beweis. Es sei A einziger Häufungspunkt der Folge $\{A_n\}$; in diesem Fall ist A auch Grenzwert dieser Folge, d. h., für jedes $\varepsilon > 0$ gibt es nur endlich viele Elemente A_n mit

$$\varrho(A_n, A) \geqq \varepsilon. \tag{37}$$

Wäre dies nämlich nicht der Fall, gäbe es also unendlich viele derartige Elemente, so müßte die aus ihnen gebildete Folge, weil $\mathfrak{X}$ kompakt ist, einen Häufungspunkt B besitzen, es würde also für unendlich viele Glieder der Folge $\{A_n\}$ sowohl die Ungleichung (37) als auch die Ungleichung

$$\varrho(A_n, B) < \frac{\varepsilon}{2}$$

erfüllt sein. Da nun auf Grund der Dreiecksungleichung

$$\varrho(A_n, A) \leqq \varrho(A, B) + \varrho(B, A_n)$$

ist, wäre dann

$$\varrho(A, B) \geqq \varrho(A_n, A) - \varrho(A_n, B) > \varepsilon - \frac{\varepsilon}{2} = \frac{\varepsilon}{2},$$

d. h., die Elemente A und B wären voneinander verschieden, die Folge $\{A_n\}$ besäße entgegen unserer Annahme zwei Häufungspunkte.

Es sei nun umgekehrt A Grenzwert der Folge $\{A_n\}$. Dann ist zunächst A auch Häufungspunkt dieser Folge. Besäße die Folge $\{A_n\}$ nun daneben noch einen zweiten Häufungspunkt B, so könnten wir eine Zahl ε so wählen, daß

$$0 < \varepsilon < \frac{1}{2}\varrho(A, B). \tag{38}$$

Für hinreichend großes n wäre dann sowohl $\varrho(A_n, A) < \varepsilon$ als auch $\varrho(A_n, B) < \varepsilon$, woraus

$$\varrho(A, B) \leqq \varrho(A, A_n) + \varrho(A_n, B) < 2\varepsilon$$

folgen würde, im Widerspruch zu (38).

Dagegen braucht, wenn $\mathfrak{X}$ nicht kompakt ist, der einzige Häufungspunkt einer Folge nicht auch zugleich ihr Grenzwert zu sein. Als Beispiel betrachten wir in dem mit der Metrik (26) versehenen Raum der in I stetigen Funktionen die Folge

$$\sin x,\ 0,\ \sin 2x,\ 0,\ \sin 3x,\ 0,\ \ldots,\ \sin nx,\ 0,\ldots\ . \tag{39}$$

Einziger Häufungspunkt dieser Folge ist die Funktion $f(x) \equiv 0$, der aber keineswegs Grenzwert der Folge ist.

Hinsichtlich der unbeschränkten Folgen sei noch bemerkt, daß sie einen Grenzwert im eigentlichen Sinne nicht besitzen können. Es sind jedoch in diesem Fall verschiedene Verallgemeinerungen des Grenzwertbegriffs möglich (uneigentliche Grenzwerte, vgl. § 39).

§ 59. Topologische Räume[1])

Ist in einem Raume eine Metrik eingeführt, so kann man, wie wir im vorangehenden Paragraphen gesehen haben, in diesem Raum auch den Begriff des Grenzwertes einer Folge von Elementen erklären. Der Begriff des Grenzwertes kann jedoch in einer Menge unter Umständen auch unabhängig von einer Metrik erklärt werden. So ist z. B. der Begriff der einfachen Konvergenz einer Folge $\{f_n(x)\}$ von in einem Intervall $I(a \leqq x \leqq b)$ definierten Funktionen (vgl. § 45) unabhängig von einer Metrik des Funktionenraumes.

Man nennt einen Raum $\mathfrak{X}$ einen *topologischen Raum*, wenn in ihm auf irgendeine Weise ein Grenzwertbegriff für Folgen von Elementen aus $\mathfrak{X}$ definiert ist, und zwar dadurch, daß jeder vorgegebenen Folge entweder gar kein Element oder genau ein Element A aus $\mathfrak{X}$ als „Grenzwert" zugeordnet ist, wobei nur verlangt wird, daß folgende Bedingung erfüllt ist: Wenn A Grenzwert der Folge $\{A_n\}$ ist, so ist A zugleich auch Grenzwert jeder Teilfolge $\{A_{p_n}\}$, d. h., wenn $A_n \to A$, so auch $A_{p_n} \to A$. Man sieht unmittelbar, daß die Menge aller in einem Intervall I definierten Funktionen (und auch jede Teilmenge hiervon) einen topologischen Raum bezüglich der einfachen Konvergenz von Funktionenfolgen bildet. Aber auch jeder metrische Raum

[1]) Im folgenden wird der Begriff des topologischen Raumes in einem etwas anderen Sinne verwendet, als dies in der modernen Topologie üblich ist.

ist (und zwar in bezug auf den aus der Metrik gewonnenen Grenzwertbegriff) ein topologischer Raum.

In einem topologischen Raum kann man in sehr einfacher Weise auch allgemeiner den Begriff des Häufungspunktes einer Folge erklären. Man nennt nämlich einen Punkt A genau dann *Häufungspunkt der Folge* $\{A_n\}$, wenn es eine Teilfolge $\{A_{p_n}\}$ von $\{A_n\}$ gibt, die A als Grenzwert besitzt. Es gilt dann: *Der Grenzwert einer Folge* $\{A_n\}$ *ist* (falls er überhaupt existiert) *auch Häufungspunkt der Folge, und zwar der einzige.*

Würde nämlich im betrachteten Fall noch ein anderer Häufungspunkt A' existieren, so müßte eine gewisse Teilfolge $\{A_{p_n}\}$ von $\{A_n\}$ den Grenzwert A' besitzen, zugleich aber nach der Definition des Grenzwertes auch den Grenzwert A, und das ist unmöglich.

Die hier angegebene Definition des Häufungspunktes einer Folge $\{A_n\}$ entspricht im Falle eines metrischen Raumes $\mathfrak{X}$ vollständig der früheren Definition des Häufungspunktes.

Ist nämlich A Häufungspunkt der Folge $\{A_n\}$ in einem metrischen Raum $\mathfrak{X}$, so liegen, wie klein auch ε gewählt wird, in jeder Umgebung $\varrho(X, A) < \varepsilon$ unendlich viele Elemente A_n. Wir betrachten dann eine Folge $\{\varepsilon_n\}$ von positiven Zahlen mit $\varepsilon_n \to 0$. Aus der Umgebung $\varrho(X, A) < \varepsilon_1$ wählen wir dann ein Element A_{p_1}, aus der Umgebung $\varrho(X, A) < \varepsilon_2$ ein Element A_{p_2} mit $p_2 > p_1$, aus der Umgebung $\varrho(X, A) < \varepsilon_3$ ein Element A_{p_3} mit $p_3 > p_2$ usw. Für die Elemente der Folge $\{A_{p_n}\}$ gilt

$$\varrho(A_{p_n}, A) < \varepsilon_n,$$

so daß

$$\varrho(A_{p_n}, A) \to 0,$$

d. h.

$$A_{p_n} \to A.$$

Besitzt umgekehrt die Folge $\{A_n\}$ eine Teilfolge $\{A_{p_n}\}$, die gegen A konvergiert, so ist natürlich A Häufungspunkt der Folge $\{A_n\}$; denn die Beziehung $A_{p_n} \to A$ bedeutet ja gerade, daß für hinreichend großes n die Ungleichung

$$\varrho(A_{p_n}, A) < \varepsilon$$

erfüllt ist, so daß jede ε-Umgebung von A unendlich viele Elemente der Folge $\{A_n\}$ enthält.

Der Begriff der *Kompaktheit* läßt sich unverändert in topologische Räume übernehmen.

Dabei ist jedoch zu beachten, daß in einem allgemeinen topologischen Raum der einzige Häufungspunkt einer Folge nicht notwendig ihr Grenzwert zu sein braucht, und zwar auch dann nicht, wenn alle Elemente der Folge einer gewissen kompakten Menge $\mathfrak{A}$ angehören.[1])

[1]) Der Begriff des Grenzwertes sei folgendermaßen definiert: Die Folge $\{A_n\}$ besitzt genau dann den Grenzwert A, wenn alle Elemente der Folge gleich A sind. Nun ist offenbar die aus zwei verschiedenen Elementen P, Q bestehende Menge kompakt. Wir betrachten dann die Folge, deren erstes Glied P und deren weitere Glieder gleich Q sind. Der einzige Häufungspunkt dieser Menge ist Q, sie besitzt aber keinen Grenzwert.

§ 60. Mengenalgebra. Die Ableitung einer Menge. Abgeschlossenheit und Zusammenhang

Hat man bei irgendwelchen Untersuchungen gleichzeitig verschiedene Mengen von Elementen eines gegebenes Raumes $\mathfrak{X}$ zu betrachten, so ist es häufig zweckmäßig, einige grundlegende Begriffsbildungen aus der allgemeinen Mengenlehre[1]) zu verwenden, die es gestatten, die Ausdrucksweise zu vereinfachen und die das Verständnis erleichtern. Wir wollen diese Begriffe hier in einer Tabelle zusammenstellen.

Sind $\mathfrak{A}$, $\mathfrak{B}$ usw. Mengen von Elementen aus $\mathfrak{X}$, so	bedeutet:
$\mathfrak{A} \subseteq \mathfrak{B}$ bzw. $\mathfrak{B} \supseteq \mathfrak{A}$ (gelesen: „$\mathfrak{A}$ ist enthalten in $\mathfrak{B}$" oder „$\mathfrak{A}$ ist Teilmenge von $\mathfrak{B}$")	Jedes Element der Menge $\mathfrak{A}$ ist auch Element der Menge $\mathfrak{B}$.
$\mathfrak{A} \subset \mathfrak{B}$ bzw. $\mathfrak{B} \supset \mathfrak{A}$ (gelesen: „$\mathfrak{A}$ ist echte Teilmenge von $\mathfrak{B}$")	Jedes Element der Menge $\mathfrak{A}$ ist auch Element der Menge $\mathfrak{B}$, und es gibt wenigstens ein Element in $\mathfrak{B}$, welches nicht zu $\mathfrak{A}$ gehört.
$\mathfrak{A} = \mathfrak{B}$	Die Mengen $\mathfrak{A}$ und $\mathfrak{B}$ enthalten dieselben Elemente, d. h., es gilt zugleich: $\mathfrak{A} \subseteq \mathfrak{B}$ und $\mathfrak{B} \subseteq \mathfrak{A}$.
$\mathfrak{A} + \mathfrak{B}$ („Summe" oder „Vereinigung" der Mengen $\mathfrak{A}$ und $\mathfrak{B}$)[2])	Die Menge aller derjenigen Elemente, die in wenigstens einer der Mengen $\mathfrak{A}$ oder $\mathfrak{B}$ liegen.
$\mathfrak{A}\mathfrak{B}$ („Produkt" oder „Durchschnitt" der Mengen $\mathfrak{A}$ und $\mathfrak{B}$)	Die Menge aller derjenigen Elemente, die sowohl in $\mathfrak{A}$ als auch in $\mathfrak{B}$ liegen.
$\mathfrak{O}$	Die „leere" Menge, welche kein Element enthält.

Bezüglich Klammersetzungen gelten dieselben Vereinbarungen wie in der Algebra.

[1]) Eine ausführlichere Darstellung der wesentlichsten Grundbegriffe der allgemeinen Mengenlehre ist in EdEM Bd. I, S. 65—82 gegeben. Wir verweisen den Leser ferner auf das ganz ausgezeichnete Lehrbuch von P. S. Alexandroff, Einführung in die Mengenlehre, Deutscher Verlag der Wissenschaften, Berlin 1956. — *Anm. d. wissenschaftl. Red.*

[2]) Statt $\mathfrak{A} + \mathfrak{B}$ und $\mathfrak{A}\mathfrak{B}$ sind in der modernen Literatur auch die Bezeichnungen $\mathfrak{A} \cup \mathfrak{B}$ bzw. $\mathfrak{A} \cap \mathfrak{B}$ üblich. — *Anm. d. wissenschaftl. Red.*

Man stellt unmittelbar fest, daß zwischen den hier eingeführten Begriffsbildungen folgende Beziehungen bestehen:

1. Wenn $\mathfrak{A} \subseteq \mathfrak{B}$ und $\mathfrak{B} \subseteq \mathfrak{C}$, so $\mathfrak{A} \subseteq \mathfrak{C}$;
2. wenn $\mathfrak{A} = \mathfrak{B}$ und $\mathfrak{B} = \mathfrak{C}$, so $\mathfrak{A} = \mathfrak{C}$;
3. $\mathfrak{A} + \mathfrak{B} = \mathfrak{B} + \mathfrak{A}$;
4. $\mathfrak{A}\mathfrak{B} = \mathfrak{B}\mathfrak{A}$;
5. $\mathfrak{A} + (\mathfrak{B} + \mathfrak{C}) = (\mathfrak{A} + \mathfrak{B}) + \mathfrak{C}$;
6. $\mathfrak{A}(\mathfrak{B}\mathfrak{C}) = (\mathfrak{A}\mathfrak{B})\mathfrak{C}$;
7. $\mathfrak{A}(\mathfrak{B} + \mathfrak{C}) = \mathfrak{A}\mathfrak{B} + \mathfrak{A}\mathfrak{C}$;
8. $\mathfrak{A} + \mathfrak{B}\mathfrak{C} = (\mathfrak{A} + \mathfrak{B})(\mathfrak{A} + \mathfrak{C})$;
9. $\mathfrak{A} \supseteq \mathfrak{A}\mathfrak{B}$;
10. $\mathfrak{A} \subseteq \mathfrak{A} + \mathfrak{B}$;
11. wenn $\mathfrak{A} \subseteq \mathfrak{B}$, so $\mathfrak{A}\mathfrak{B} = \mathfrak{A}$, und umgekehrt;
12. wenn $\mathfrak{A} \subseteq \mathfrak{B}$, so $\mathfrak{A} + \mathfrak{B} = \mathfrak{B}$, und umgekehrt;
13. $\mathfrak{A} + \mathfrak{A} = \mathfrak{A}\mathfrak{A} = \mathfrak{A}$;
14. wenn $\mathfrak{A} + \mathfrak{B} = \mathfrak{O}$, so $\mathfrak{A} = \mathfrak{B} = \mathfrak{O}$, und umgekehrt;
15. wenn $\mathfrak{A}\mathfrak{B} = \mathfrak{X}$, so $\mathfrak{A} = \mathfrak{B} = \mathfrak{X}$, und umgekehrt.

Es sei nun $\mathfrak{X}$ ein beliebiger topologischer Raum. Wir wollen als erstes erklären, was unter einem Häufungspunkt einer beliebigen Menge $\mathfrak{A}$ von Elementen aus $\mathfrak{X}$ verstanden werden soll: Ein Element A aus $\mathfrak{X}$ (welches der Menge $\mathfrak{A}$ angehören kann, aber nicht angehören muß) heißt ein *Häufungspunkt* (oder *Häufungswert*) der Menge $\mathfrak{A}$, wenn es eine Folge $\{A_n\}$ von paarweise verschiedenen Elementen aus $\mathfrak{A}$ gibt, die den Grenzwert A besitzt, für die also gilt:

$$A_n \to A.^{1})$$

Die Menge aller Häufungspunkte einer Menge $\mathfrak{A}$ nennt man die *Ableitung* der Menge $\mathfrak{A}$; sie wird üblicherweise mit $\mathfrak{A}'$ bezeichnet.

Hier einige Beispiele:

1. Ist $\mathfrak{A}$ das offene Intervall $a < x < b$, so ist in bezug auf die gewöhnliche Metrik $\mathfrak{A}'$ das abgeschlossene Intervall $a \leqq x \leqq b$.

2. Zu demselben Resultat gelangt man, wenn man als Menge $\mathfrak{A}$ die Menge aller rationalen Zahlen oder aller endlichen Dezimalbrüche im Intervall $a < x < b$ nimmt.

---

[1]) Man beachte, daß ein Unterschied besteht zwischen dem Begriff des Häufungspunktes einer Folge und dem Begriff des Häufungspunktes der Menge der Elemente einer Folge. Zum Beispiel besitzt die Folge 0, 1, 0, 1, 0, 1, ... die Häufungspunkte 0 und 1, während die aus ihnen gebildete Menge nur zwei Elemente enthält und daher (wie jede andere endliche Menge) keinen Häufungspunkt besitzt. Für Folgen ohne „mehrfache" Elemente stimmen natürlich beide Begriffe überein.

3. Ist $\mathfrak{A}$ die Menge aller Punkte (x, y), welche der Bedingung $\begin{cases} 0 < x < 1 \\ 0 < y < 1 \end{cases}$ genügen (Inneres des „Einheitsquadrates"), so ist $\mathfrak{A}'$ die Menge aller Punkte (x, y) mit $\begin{cases} 0 \leqq x \leqq 1 \\ 0 \leqq y \leqq 1 \end{cases}$ („Einheitsquadrat" unter Einschluß seiner Seiten).

4. Es sei $\mathfrak{A}$ die Menge aller Punkte der Ebene, die auf der Kurve $y = \sin\frac{1}{x}$ $(0 < x \leqq 1)$ liegen (vgl. § 41, Beispiel 4). Dann ist $\mathfrak{A}'$ die Vereinigungsmenge aus $\mathfrak{A}$ und der Menge aller Punkte (x, y) mit $\begin{cases} x = 0, \\ -1 \leqq y \leqq +1 \end{cases}$ (Intervall $-1 \leqq y \leqq 1$ auf der y-Achse).

5. Ist $\mathfrak{A}$ die Menge aller Zahlen $\frac{1}{n}$ $(n = 1, 2, 3, \ldots)$, so enthält $\mathfrak{A}'$ nur die Zahl 0.

6. Ist $\mathfrak{A}$ die Menge aller rationalen Zahlen $\frac{1}{p} + \frac{1}{q}$ (p, q positive ganze Zahlen), so ist $\mathfrak{A}'$ die Menge aus allen Zahlen $\frac{1}{p}$ und der Zahl 0.

7. Ist $\mathfrak{A}$ die Menge aller Punkte auf den Kreisen $x^2 + y^2 = \left(1 - \frac{1}{n}\right)^2$ $(n = 2, 3, 4, \ldots)$, so ist $\mathfrak{A}'$ die Vereinigungsmenge aus $\mathfrak{A}$ und der Menge aller Punkte, die auf dem Kreis $x^2 + y^2 = 1$ liegen.

8. Ist $\mathfrak{A}$ die Menge aller Polynome (betrachtet als Funktionen in einem gegebenen Intervall $a \leqq x \leqq b$), so ist in bezug auf die zur gleichmäßigen Konvergenz gehörenden Metrik $\mathfrak{A}'$ die Menge aller im Intervall $a \leqq x \leqq b$ stetigen Funktionen (vgl. den Satz von WEIERSTRASS in § 49).

9. Ist $\mathfrak{A}$ die Menge aller Funktionen $f_n(x) = \frac{1}{1 + n^2 x^2}$ $(n = 1, 2, 3, \ldots)$, so ist in bezug auf die Topologie der einfachen Konvergenz $\mathfrak{A}'$ die Menge, die nur aus der Funktion

$$f(x) = \begin{cases} 1 & \text{für} \quad x = 0, \\ 0 & \text{für} \quad x \neq 0 \end{cases}$$

besteht. In bezug auf die zur gleichmäßigen Konvergenz gehörende Metrik ist $\mathfrak{A}'$ die leere Menge.

Eine Menge $\mathfrak{A}$ von Punkten eines topologischen Raumes $\mathfrak{X}$ heißt *abgeschlossen*, wenn sie alle ihre Häufungspunkte enthält, d. h., wenn

$$\mathfrak{A}' \subseteq \mathfrak{A}.$$

Die Vereinigungsmenge von $\mathfrak{A}$ und der Ableitung $\mathfrak{A}'$ von $\mathfrak{A}$, also die Menge

$$\overline{\mathfrak{A}} = \mathfrak{A} + \mathfrak{A}',$$

bezeichnet man als *abgeschlossene Hülle* oder *Abschließung* von $\mathfrak{A}$. Offenbar stimmen für jede abgeschlossene Menge $\mathfrak{A}$ die Mengen $\mathfrak{A}$ und $\overline{\mathfrak{A}}$ überein. Gilt umgekehrt für eine Menge $\mathfrak{A}$ die Beziehung $\mathfrak{A} = \overline{\mathfrak{A}}$, so ist die Menge $\mathfrak{A}$ abgeschlossen (wegen 10., Seite 261). *Eine Menge $\mathfrak{A}$ ist also genau dann abgeschlossen, wenn sie mit ihrer abgeschlossenen Hülle übereinstimmt.*

Beispiele für abgeschlossene Mengen sind:

a) Die Menge aller Punkte des abgeschlossenen Intervalls $a \leqq x \leqq b$;

b) die Menge aller im Intervall $a \leqq x \leqq b$ stetigen Funktionen (in bezug auf die zur gleichmäßigen Konvergenz gehörende Metrik).

Ist $\mathfrak{A} \subseteq \mathfrak{B}$, so ist auch $\mathfrak{A}' \subseteq \mathfrak{B}'$.

Satz. *In einem metrischen Raum ist jede abgeleitete Menge* (jede Menge, die Ableitung einer anderen Menge ist) *abgeschlossen.*

In Gestalt einer Formel kann man diese Behauptung offenbar folgendermaßen ausdrücken:

$$\mathfrak{A}'' \subseteq \mathfrak{A}'.$$

Beweis. Es sei A'' irgendein Punkt der Menge $\mathfrak{A}''$. Zu zeigen ist, daß in jeder noch so kleinen Umgebung von A'' Punkte der Menge $\mathfrak{A}$ liegen, d. h., daß es zu jedem noch so kleinen $\varepsilon > 0$ ein Element A aus $\mathfrak{A}$ gibt, für welches

$$\varrho(A,\ A'') < \varepsilon$$

gilt.

Da A'' ein Häufungspunkt der Menge $\mathfrak{A}'$ ist, gibt es zunächst einen Punkt A' aus $\mathfrak{A}'$ mit

$$\varrho(A',\ A'') < \frac{\varepsilon}{2}.$$

Andererseits gibt es, da A' Häufungspunkt $\mathfrak{A}$ ist, ein Element A aus $\mathfrak{A}$ mit

$$\varrho(A,\ A') < \frac{\varepsilon}{2}.$$

Aus den angegebenen Ungleichungen ergibt sich jedoch auf Grund der Dreiecksungleichung sofort:

$$\varrho(A,\ A'') \leqq \varrho(A,\ A') + \varrho(A',\ A'') < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

was zu beweisen war.

Der angegebene Satz ist offenbar die Verallgemeinerung des auf Seite 153 formulierten und bewiesenen Satzes auf beliebige metrische Räume.

In beliebigen topologischen Räumen ist der Satz nicht richtig.

Eine Menge $\mathfrak{A}$ von Punkten eines topologischen Raumes $\mathfrak{X}$ heißt *zusammenhängend*, wenn es keine zwei Mengen $\mathfrak{M}$ und $\mathfrak{N}$ aus $\mathfrak{X}$ gibt, für die folgende Bedingungen erfüllt sind:

$$\left.\begin{array}{ll} 1. & \mathfrak{A} = \mathfrak{M} + \mathfrak{N}, \\ 2. & \mathfrak{M}\mathfrak{N} + \mathfrak{M}\mathfrak{N}' + \mathfrak{M}'\mathfrak{N} = \mathfrak{O}, \\ 3. & \mathfrak{M} \neq \mathfrak{O}, \quad \mathfrak{N} \neq \mathfrak{O}. \end{array}\right\} \tag{40}$$

Beispiele (zugrunde gelegt ist in allen Fällen die übliche euklidische Metrik).

10. Die aus zwei Zahlen a und b $(a \neq b)$ bestehende Menge ist nicht zusammenhängend (man braucht als Menge $\mathfrak{M}$ nur die aus dem einzigen Element a und als $\mathfrak{N}$ nur die aus dem einzigen Element b bestehende Menge zu nehmen).

11. Genauso verhält es sich bei der aus den Punkten der Kreisscheiben
$$(x+c)^2+y^2<R^2 \quad \text{und} \quad (x-c)^2+y^2<R^2$$
bestehenden Menge, falls $c>R>0$ (und sogar bei $c=R>0$).

12. Enthält eine Menge $\mathfrak{A}$ von Zahlen die Elemente a und c, aber nicht ein gegebenes Element b mit $a<b<c$, so ist die Menge $\mathfrak{A}$ nicht zusammenhängend. Als Menge $\mathfrak{M}$ kann man in diesem Fall die Menge aller Punkte x aus $\mathfrak{A}$ nehmen, für die $x<b$ gilt, und als Menge $\mathfrak{N}$ die Menge aller x aus $\mathfrak{A}$, für die $x>b$ gilt.

13. Das abgeschlossene Intervall $a \leqq x \leqq b$ ist zusammenhängend.

Diese Behauptung, die ganz trivial erscheint, ist rein formal gar nicht so einfach zu beweisen. Wir wollen hier den Verlauf des (indirekten) Beweises für diese Behauptung nur andeuten, die Durchführung im einzelnen jedoch dem Leser überlassen.

Man nimmt also an, die Menge $\mathfrak{A}$ aller Punkte des Intervalls $a \leqq x \leqq b$ wäre nicht zusammenhängend, es gäbe also Mengen $\mathfrak{M}$ und $\mathfrak{N}$, welche die Bedingungen (40) erfüllen. Aus 1. und 3. ergibt sich dann sofort, daß es in $\mathfrak{A}\mathfrak{M}$ einen Punkt m_1 und in $\mathfrak{A}\mathfrak{N}$ einen Punkt n_1 gibt. Da wegen 2. $\mathfrak{M}\mathfrak{N}=\mathfrak{O}$ ist, gilt $m_1 \neq n_1$. Es sei etwa $m_1<n_1$. Dann gehört sicher der Punkt $\frac{m_1+n_1}{2}$ zu $\mathfrak{A}$. Folglich liegt dieser Punkt in einer (und zwar wegen $\mathfrak{M}\mathfrak{N}=\mathfrak{O}$ sogar in genau einer) der Mengen $\mathfrak{M}$ oder $\mathfrak{N}$. Im ersten Fall setzen wir $m_2=\frac{m_1+n_1}{2}$, $n_2=n_1$, im zweiten $m_2=m_1$, $n_2=\frac{m_1+n_1}{2}$ und stellen mit m_2 und n_2 die gleichen Überlegungen wie vorher mit m_1 und n_1 an. Setzt man diesen Prozeß unbegrenzt fort, so erhält man zwei monotone Folgen von Zahlen, die beide gegen denselben Grenzwert ξ streben, der zu $\mathfrak{M}'\mathfrak{N}'$ gehört. Beachtet man noch, daß ξ entweder zu $\mathfrak{M}$ oder zu $\mathfrak{N}$ gehört, so erhält man einen Widerspruch zu $\mathfrak{M}\mathfrak{N}'=\mathfrak{M}'\mathfrak{N}=\mathfrak{O}$.

14. Was oben über abgeschlossene Intervalle gesagt wurde, läßt sich entsprechend auf beliebige andere Intervalle übertragen.

Umgekehrt ist jede zusammenhängende Menge auf der Zahlengeraden ein Intervall einer der möglichen Typen (endlich oder unendlich, offen oder abgeschlossen oder „halboffen"); vgl. Beispiel 12.

Eine Menge $\mathfrak{A}$ von Punkten eines Raumes $\mathfrak{X}$, welche in diesem Raume zusammenhängend und abgeschlossen ist, wird ein *Kontinuum* genannt.

Ist $\mathfrak{X}$ der „eindimensionale euklidische Raum", d. h. die Zahlengerade mit der gewöhnlichen Metrik, so sind die Kontinua genau die abgeschlossenen Intervalle und nur diese.

Im „zweidimensionalen euklidischen Raum", d. h. in der Ebene mit der gewöhnlichen Metrik, gibt es dagegen die verschiedensten Typen von Kontinua: Hierzu gehören z. B. alle „abgeschlossenen Quadrate" (d. h. Quadrate mit Berandung), die abgeschlossenen Kreisscheiben usw.

§ 61. Stetige Abbildungen und ihre Eigenschaften

Ein topologischer Raum $\mathfrak{X}$ möge durch eine Abbildung f auf einen topologischen Raum $\mathfrak{Y}$ abgebildet werden. Unter $Y=f(X)$ soll dasjenige Element aus $\mathfrak{Y}$ verstanden werden, das dem Element X aus $\mathfrak{X}$ bei der Abbildung f entspricht.

Ist $\mathfrak{A}$ irgendeine Menge von Elementen aus $\mathfrak{X}$, so verstehen wir unter $f(\mathfrak{A})$ die Gesamtheit der Elemente aus $\mathfrak{Y}$, die Elementen aus $\mathfrak{A}$ zugeordnet sind.

Eine Abbildung

$$Y = f(X)$$

heißt *stetig im Punkte* X, wenn aus der Beziehung

$$X_n \to X$$

stets die Beziehung

$$f(X_n) \to f(X)$$

folgt; die Abbildung heißt *stetig auf der Menge* $\mathfrak{A}$, wenn sie in allen Punkten von $\mathfrak{A}$ stetig ist.

Wir wollen hier einige Eigenschaften von stetigen Abbildungen beweisen:

Satz 1. *Ist die Menge* $\mathfrak{A}\,(\subseteq \mathfrak{X})$ *zusammenhängend und die Abbildung* $Y = f(X)$ *stetig auf der Menge* $\mathfrak{A}$, *so ist auch die Menge* $\mathfrak{B} = f(\mathfrak{A})$ $(\subseteq \mathfrak{Y})$ *zusammenhängend.*

Wir werden diesen Satz indirekt beweisen, d. h., wir nehmen an, die Menge $\mathfrak{B}$ wäre nicht zusammenhängend. Dann gibt es zwei Mengen $\mathfrak{M}$ und $\mathfrak{N}$ derart, daß

1. $\mathfrak{B} = \mathfrak{M} + \mathfrak{N}$,
2. $\mathfrak{M}\mathfrak{N} + \mathfrak{M}\mathfrak{N}' + \mathfrak{M}'\mathfrak{N} = \mathfrak{O}$,
3. $\mathfrak{M} \neq \mathfrak{O}, \quad \mathfrak{N} \neq \mathfrak{O}$.

Wegen $\mathfrak{M}\mathfrak{N} = \mathfrak{O}$ wird kein Element A aus $\mathfrak{A}$ gleichzeitig in ein Element aus $\mathfrak{M}$ und in ein Element aus $\mathfrak{N}$ abgebildet. Bezeichnen wir mit $\mathfrak{R}$ die Menge derjenigen Elemente aus $\mathfrak{A}$, die in $\mathfrak{M}$, und mit $\mathfrak{L}$ die Menge derjenigen Elemente aus $\mathfrak{A}$, die in $\mathfrak{N}$ abgebildet werden, so ist also $\mathfrak{R}\mathfrak{L} = \mathfrak{O}$. Dabei ist wegen $\mathfrak{M} \neq \mathfrak{O}$ und $\mathfrak{N} \neq \mathfrak{O}$ auch $\mathfrak{R} \neq \mathfrak{O}$ und $\mathfrak{L} \neq \mathfrak{O}$, und wegen $\mathfrak{B} = \mathfrak{M} + \mathfrak{N}$ auch $\mathfrak{A} = \mathfrak{R} + \mathfrak{L}$.

Wir beweisen als erstes, daß $\mathfrak{R}\mathfrak{L}' = \mathfrak{O}$. Angenommen, es wäre $\mathfrak{R}\mathfrak{L}' \neq \mathfrak{O}$. Dann gäbe es in der Menge $\mathfrak{R}$ ein Element A, welches Häufungspunkt der Menge $\mathfrak{L}$ ist. Man könnte also eine Folge $\{A_n\}$ von Elementen der Menge $\mathfrak{L}$ finden, die das Element A aus $\mathfrak{R}$ als Grenzwert besitzt:

$$A_n \to A.$$

Wegen der Stetigkeit der betrachteten Abbildung ergibt sich daraus

$$f(A_n) \to f(A).$$

Das bedeutet, daß man in $\mathfrak{N}$ eine Folge $\{f(A_n)\}$ von Elementen so finden kann, daß sie das Element $f(A)$ aus $\mathfrak{M}$ als Grenzwert besitzt. Dann wäre aber $\mathfrak{M}\mathfrak{N}' \neq \mathfrak{O}$, im Widerspruch zu 2. Somit ist also $\mathfrak{R}\mathfrak{L}' = \mathfrak{O}$. Entsprechend ergibt sich $\mathfrak{R}'\mathfrak{L} = \mathfrak{O}$. Damit erhalten wir

1. $\mathfrak{A} = \mathfrak{R} + \mathfrak{L}$,
2. $\mathfrak{R}\mathfrak{L} + \mathfrak{R}\mathfrak{L}' + \mathfrak{R}'\mathfrak{L} = \mathfrak{O}$,
3. $\mathfrak{R} \neq \mathfrak{O}, \quad \mathfrak{L} \neq \mathfrak{O}$,

d. h., die Menge $\mathfrak{A}$ wäre nicht zusammenhängend, entgegen unserer Voraussetzung. Damit ist Satz 1 bewiesen.

Satz 2. *Ist die Menge* $\mathfrak{A}(\subseteq \mathfrak{X})$ *kompakt und abgeschlossen, so ist auch, sofern die Abbildung* $Y = f(X)$ *auf* $\mathfrak{A}$ *stetig ist, die Menge* $\mathfrak{B} = f(\mathfrak{A})(\subseteq \mathfrak{Y})$ *kompakt.*

Satz 3. *Unter den gleichen Voraussetzungen ist die Menge* $\mathfrak{B} = f(\mathfrak{A})(\subseteq \mathfrak{Y})$ *abgeschlossen.*

Beweis. Es sei $\{B_n\}$ irgendeine Folge von Elementen aus $\mathfrak{B}$. Unter $\{A_n\}$ verstehen wir jeweils eines der Elemente aus $\mathfrak{A}$, welches auf B_n abgebildet wird, so daß also

$$B_n = f(A_n) \qquad (n = 1, 2, 3, \ldots).$$

Wir betrachten dann in $\mathfrak{A}$ die Folge $\{A_n\}$. Da die Menge $\mathfrak{A}$ kompakt ist, können wir aus der Folge $\{A_n\}$ eine Teilfolge $\{A_{p_n}\}$ auswählen, die in $\mathfrak{X}$ einen Grenzwert A besitzt. Wegen der Abgeschlossenheit von $\mathfrak{A}$ gehört A zu $\mathfrak{A}$. Da die betrachtete Abbildung stetig ist, folgt aus

$$A_{p_n} \to A,$$

daß

$$f(A_{p_n}) \to f(A). \tag{41}$$

Damit ist gezeigt, daß die Folge $\{f(A_{p_n})\}$, und somit auch $\{B_{p_n}\}$ konvergiert, d. h., jede Folge $\{B_n\}$ von Elementen aus $\mathfrak{B}$ besitzt einen Häufungspunkt. Folglich ist $\mathfrak{B}$ kompakt, womit Satz 2 bewiesen ist.

Zum Beweis von Satz 3 führen wir denselben Übergang aus, wobei wir jedoch zusätzlich voraussetzen, daß die Folge $\{B_n\}$ einen Grenzwert B besitzt:

$$B_n \to B, \tag{42}$$

wobei B ein gewisses Element aus $\mathfrak{Y}$ ist. Nun können wir der Beziehung (41) offenbar auch die Form

$$B_{p_n} \to f(A) \tag{43}$$

geben, und da nach Definition des Grenzwertes in einem topologischen Raum (vgl. S. 258) der Grenzwert einer Teilfolge nicht von dem der Gesamtfolge verschieden sein kann, erhalten wir aus (42) und (43)

$$B = f(A).$$

Nach Konstruktion ist aber A ein Element aus $\mathfrak{A}$, und da es in B abgebildet wird, gehört B folglich zur Menge $\mathfrak{B}$, was zu beweisen war.

Nehmen wir als topologische Räume $\mathfrak{X}$ und $\mathfrak{Y}$ speziell die Menge aller reellen Zahlen mit der gewöhnlichen Metrik und als Menge $\mathfrak{A}$ hinsichtlich Satz 1 ein beliebiges und hinsichtlich Satz 2 und 3 ein abgeschlossenes Intervall, so erhalten wir unmittelbar, daß die Menge der von einer stetigen Funktion $f(x)$ in diesem Intervall angenommenen Werte

1. zusammenhängend ist, d. h. ein gewisses Intervall bildet (vgl. § 60, Beispiel 14).

2. (falls $\mathfrak{A}$ ein abgeschlossenes Intervall ist) kompakt, d. h. beschränkt ist (vgl. § 36 und § 57).

3. (falls $\mathfrak{A}$ ein abgeschlossenes Intervall ist) abgeschlossen ist, so daß es unter den Werten einen größten und einen kleinsten gibt.

Das sind aber nicht anderes als die Sätze I bis III aus § 47.

Auch Satz IV aus § 47 über die gleichmäßige Stetigkeit läßt sich entsprechend verallgemeinern; da jedoch die in ihm gemachte Aussage im wesentlichen metrischer Natur ist, bringen wir hier nur seine Verallgemeinerung auf beliebige metrische und nicht auf beliebige topologische Räume.

Satz 4. *Es sei $\mathfrak{A}$ eine kompakte und abgeschlossene Menge eines metrischen Raumes $\mathfrak{X}$, die durch die stetige Abbildung $Y = f(X)$ auf eine gewisse Menge $\mathfrak{B} = f(\mathfrak{A})$ eines ebenfalls metrischen Raumes $\mathfrak{Y}$ abgebildet wird. Dann kann man zu jeder noch so kleinen Zahl $\varepsilon > 0$ eine Zahl δ so finden, daß für alle X' und X'' aus $\mathfrak{A}$, welche die Beziehung*

$$\varrho(X', X'') < \delta \tag{44}$$

erfüllen, die Ungleichung

$$\varrho(f(X'), f(X'')) < \varepsilon \tag{45}$$

gilt.

Der Beweis hierfür verläuft entsprechend dem Beweis von Satz IV in § 47. Wir nehmen also an, man könnte nicht zu jedem $\varepsilon > 0$ eine derartige Zahl δ finden. Dann gibt es ein gewisses $\varepsilon^* > 0$, so daß sich zu jeder Zahl δ_n aus einer Folge $\{\delta_n\}$ mit $\delta_n > 0$ und $\delta_n \to 0$ ein Paar X'_n, X''_n von Elementen aus $\mathfrak{A}$ so angeben läßt, daß die Ungleichungen

$$\varrho(X'_n, X''_n) < \delta_n$$

und

$$\varrho(f(X'_n), f(X''_n)) \geqq \varepsilon^* \tag{46}$$

gelten. Da nun $\mathfrak{A}$ kompakt ist, kann man aus der Folge $\{X'_n\}$ eine konvergente Teilfolge $\{X'_{p_n}\}$ auswählen, und zwar möge

$$X'_{p_n} \to \Xi \tag{47}$$

gelten. Man sieht unmittelbar, daß dann auch

$$X''_{p_n} \to \Xi; \tag{48}$$

denn es ist

$$\varrho(X''_{p_n}, \Xi) \leqq \varrho(X''_{p_n}, X'_{p_n}) + \varrho(X'_{p_n}, \Xi),$$

und da hier die rechte Seite gegen Null strebt, gilt dies auch für die linke Seite. Wegen der Abgeschlossenheit von $\mathfrak{A}$ gehört nun Ξ zu $\mathfrak{A}$. Daher folgt aus (47) und (48) auf Grund der Stetigkeit der Funktion $f(X)$ in $\mathfrak{A}$

$$f(X'_{p_n}) \to f(\Xi),\ f(X''_{p_n}) \to f(\Xi) \tag{49}$$

und damit

$$\varrho(f(X'_{p_n}), f(X''_{p_n})) \to 0.$$

Dies steht jedoch im Widerspruch zu der sich aus (46) ergebenden Beziehung

$$\varrho(f(X'_{p_n}), f(X''_{p_n})) \geqq \varepsilon^*.$$

§ 62. Homöomorphe Abbildungen

Im vorliegenden Paragraphen wollen wir den Fall betrachten, daß die Mengen $\mathfrak{X}$ und $\mathfrak{Y}$ durch eine Abbildung f umkehrbar eindeutig und in beiden Richtungen stetig aufeinander abgebildet werden, d. h. also:

1. Jedem Element X aus $\mathfrak{X}$ entspricht genau ein Element $Y = f(X)$ aus $\mathfrak{Y}$, wobei auch umgekehrt jedes Element Y aus $\mathfrak{Y}$ Bild eines und nur eines Elementes X aus $\mathfrak{X}$ ist (vgl. § 54).

2. Hängt eine Folge $\{X_n\}$ von Elementen aus $\mathfrak{X}$ mit einem ebenfalls zu $\mathfrak{X}$ gehörenden Element X durch die Beziehung

$$X_n \to X \tag{50}$$

zusammen, so besteht zwischen der Folge $\{Y_n\} \equiv \{f(X_n)\}$ und dem Element $Y = f(X)$, das zu $\mathfrak{Y}$ gehört, die Beziehung

$$Y_n \to Y, \tag{51}$$

d. h.

$$f(X_n) \to f(X),$$

und besteht umgekehrt zwischen einer Folge $\{Y_n\}$ von Elementen aus $\mathfrak{Y}$ und einem Element Y aus $\mathfrak{Y}$ die Beziehung (51), so besteht zwischen der entsprechenden Folge $\{X_n\}$ aus $\mathfrak{X}$ und dem entsprechenden Element X aus $\mathfrak{X}$ die Beziehung (50).

Eine Abbildung f der Menge $\mathfrak{X}$ auf die Menge $\mathfrak{Y}$, welche diese Eigenschaften besitzt, nennt man eine *homöomorphe* oder *topologische Abbildung* bzw. einen *Homöomorphismus*. Mengen $\mathfrak{X}$ und $\mathfrak{Y}$, die man durch eine homöomorphe Abbildung aufeinander abbilden kann, wollen wir kurz *homöomorph* nennen.

Einfache Beispiele für homöomorphe Mengen sind (ebene) Figuren F und F', die im elementargeometrischen Sinne ähnlich sind. Bekanntlich nennt man Figuren F und F' genau dann *ähnlich*, wenn man sie so aufeinander abbilden kann, daß für je zwei Punkte M und N von F und die entsprechenden Bildpunkte M' und N' von F'

$$\frac{M'N'}{MN} = \lambda$$

gilt, wobei $\lambda > 0$ ein konstanter *Ähnlichkeitskoeffizient* ist. Nehmen wir nämlich der Einfachheit halber an, daß die Figuren F und F' nicht nur untereinander ähnlich sind, sondern auch zueinander ähnlich in bezug auf einen gegebenen Koordinatenursprung liegen, so können wir die betrachtete Abbildung offenbar durch folgende Beziehungen zwischen den Koordinaten (x, y) eines Punktes M und den Koordinaten (x', y') seines Bildpunktes M' beschreiben:

$$\begin{cases} x' = \lambda x, \\ y' = \lambda x \end{cases} \quad \text{oder} \quad \begin{cases} x = \dfrac{x'}{\lambda}, \\ y = \dfrac{y'}{\lambda}. \end{cases} \tag{52}$$

Ist nun $\{(x_n, y_n)\}$ eine Folge von Punkten der Figur F, die den Punkt (x, y) der Figur F als Häufungspunkt besitzt, und bezeichnen wir mit (x'_n, y'_n) und (x', y') die entsprechenden Punkte auf F', so bedingen sich die Beziehungen

$$x_n \to x, \quad y_n \to y,$$

und

$$x'_n \to x', \quad y'_n \to y'$$

gegenseitig; dies bedeutet offenbar, daß die betrachtete Abbildung in beiden Richtungen stetig ist.

Ein anderes einfaches Beispiel für homöomorphe Mengen bilden ebene Figuren F und F', die zueinander spiegelbildlich in bezug auf den Einheitskreis liegen, d. h., die durch Inversion am Einheitskreis ineinander übergeführt werden können (vgl. § 56, S. 248). In diesem Fall besteht zwischen den Koordinaten (x, y) eines Punktes M auf F und den Koordinaten (x', y') seines Bildpunktes M' auf F' die Beziehung

$$\begin{cases} x' = \dfrac{x}{x^2 + y^2}, \\ y' = \dfrac{y}{x^2 + y^2} \end{cases} \quad \text{oder} \quad \begin{cases} x = \dfrac{x'}{x'^2 + y'^2}, \\ y = \dfrac{y'}{x'^2 + y'^2}. \end{cases} \tag{53}$$

Enthält die Figur (Punktmenge) F nicht den Koordinatenursprung O, so beschreiben, wie man sofort einsieht, die Formeln (53) eine umkehrbar eindeutige und in beiden Richtungen stetige Abbildung von F auf F'.

Bevor wir uns weiteren Beispielen von homöomorphen Abbildungen zuwenden (wobei wir uns wie in den vorangehenden Beispielen ausschließlich auf die Ebene mit der gewöhnlichen Metrik beschränken werden), wollen wir zunächst einen wichtigen Hilfssatz beweisen:

Wird durch eine Abbildung f eine Menge $\mathfrak{X}$ homöomorph auf die Menge $\mathfrak{Y}$ abgebildet, so wird durch f auch jede Teilmenge $\mathfrak{X}_1$ von $\mathfrak{X}$ homöomorph auf die $\mathfrak{X}_1$ durch f zugeordnete Teilmenge $\mathfrak{Y}_1 = f(\mathfrak{X}_1)$ von $\mathfrak{Y}$ abgebildet.

In der Tat: Ist $\{X_n\}$ eine Folge von Elementen aus $\mathfrak{X}_1$, die gegen das zu $\mathfrak{X}_1$ gehörende Element X konvergiert, so gilt zunächst in $\mathfrak{Y}$

$$f(X_n) \to f(X).$$

Da aber nach Konstruktion alle Elemente $Y_n = f(X_n)$ und auch das Element $Y = f(X)$ zu $\mathfrak{Y}_1$ gehören und alle diese Überlegungen umkehrbar sind, ist die durch f zwischen $\mathfrak{X}_1$ und $\mathfrak{Y}_1$ erklärte Zuordnung ein Homöomorphismus.

Ist insbesondere

$$\mathfrak{X} = \mathfrak{X}_1 + \mathfrak{X}_2 \quad \text{und} \quad \mathfrak{X}_1 \mathfrak{X}_2 = \mathfrak{O}$$

und f eine homöomorphe Abbildung von $\mathfrak{X}$ auf eine gewisse Menge $\mathfrak{Y}$, wobei $\mathfrak{X}_1$ auf $\mathfrak{Y}_1$ und $\mathfrak{X}_2$ auf $\mathfrak{Y}_2$ abgebildet werden möge, so ist f als Abbildung sowohl von $\mathfrak{X}_1$ auf $\mathfrak{Y}_1$ als auch von $\mathfrak{X}_2$ auf $\mathfrak{Y}_2$ ein Homöomorphismus. Nehmen wir speziell an, daß $\mathfrak{X}_1$ nur ein Element enthält, so enthält auch $\mathfrak{Y}_1$ nur ein Element, und wir erhalten:

Ist f eine homöomorphe Abbildung einer Menge $\mathfrak{X}$ auf eine Menge $\mathfrak{Y}$, so vermittelt f zugleich auch eine homöomorphe Abbildung zwischen den Mengen, die man erhält, wenn man aus der Menge $\mathfrak{X}$ einen Punkt und aus der Menge $\mathfrak{Y}$ den entsprechenden Bildpunkt herausnimmt.

Mit Hilfe dieses und der im vorangehenden Paragraphen bewiesenen Sätze werden wir in einer ganzen Reihe von Fällen zeigen können, daß es unmöglich ist, zwischen gewissen Mengen eine homöomorphe Abbildung herzustellen.

Zunächst ist klar, daß je zwei abgeschlossene Intervalle zueinander homöomorph sind (weil sie ähnlich sind). Genauso verhält es sich mit je zwei offenen Intervallen (Intervallen ohne Endpunkte). Dagegen sind ein abgeschlossenes und ein offenes (oder halboffenes) Intervall niemals homöomorph, da schon bei jeder stetigen Abbildung jede abgeschlossene und kompakte Menge in eine abgeschlossene Menge übergeht (vgl. § 61, Satz 3). Aber auch ein halbabgeschlossenes und ein offenes Intervall können niemals zueinander homöomorph sein. Wäre dies nämlich der Fall, so könnten wir von dem halbabgeschlossenen Intervall den einen Endpunkt und aus dem offenen Intervall dessen Bildpunkt fortnehmen und erhielten nach dem vorangehenden zwei Mengen, die ebenfalls homöomorph sind. Das ist aber unmöglich, da eine dieser Mengen (nämlich das von dem halboffenen Intervall verbleibende offene Intervall) zusammenhängend ist, während dies bei der anderen nicht der Fall ist (vgl. § 61, Satz 1).

Aus demselben Grunde kann auch ein Intervall niemals einer Vereinigungsmenge von zwei oder mehr Intervallen ohne gemeinsame Punkte homöomorph sein. Desgleichen ist eine Vereinigungsmenge von zwei Intervallen ohne gemeinsame Punkte niemals homöomorph einer Vereinigungsmenge von drei oder mehr Intervallen ohne gemeinsame Punkte usw. Dagegen ist aber die Vereinigungsmenge von zwei abgeschlossenen Intervallen, die einen Endpunkt gemeinsam haben (aber nicht auf einer Geraden zu liegen brauchen, also einen „Streckenzug" bilden können) jedem abgeschlossenen Intervall homöomorph, z. B. dem Intervall $[0, 1]$. Zur homöomorphen Abbildung des Streckenzuges ABC auf das Intervall $[0, 1]$ braucht man nämlich nur die Strecke (das Intervall) AB auf das Intervall $\left[0, \frac{1}{2}\right]$ und die Strecke BC auf das Intervall $\left[\frac{1}{2}, 1\right]$ abzubilden, was jeweils durch eine Ähnlichkeitstransformation erreicht werden kann. Allgemein ist jeder Streckenzug, der sich nicht selbst schneidet und kein geschlossenes Vieleck bildet, jedem abgeschlossenen Intervall homöomorph. So beschaffen sind z. B. die Buchstaben[1])

L, M, N, V, W, Z

Demgegenüber ist z. B. der Buchstabe T keinem Intervall homöomorph. Wäre dies nämlich der Fall, so erhielte man, wenn man in T den Mittelpunkt des horiozontalen „Balkens", der zugleich oberer Endpunkt des vertikalen Balkens ist, und in dem dazu homöomorphen Intervall dessen Bildpunkt fortließe, eine homöomorphe Abbildung einer Vereinigungsmenge von drei Intervallen ohne gemeinsame Punkte auf eine Vereinigungsmenge von zwei Intervallen ohne gemeinsame Punkte, was unmöglich ist.

---

[1]) Die „Buchstaben" sind hierbei natürlich in einem idealisierten Sinne aufzufassen, so nämlich, als wären sie mit einer „unendlich feinen" Feder geschrieben.

Die Buchstaben

E, F, Y

sind ebenfalls keinem Intervall homöomorph, wohl aber dem Buchstaben T und damit auch untereinander.

Der Buchstabe K ist dem Buchstaben X, doch weder einem Intervall noch einem der bisher genannten Buchstaben homöomorph.

Der Buchstabe O (Kreis) ist weder einem Intervall noch einem der bisher betrachteten Buchstaben homöomorph. Um dies einzusehen, nehmen wir an, dies wäre doch der Fall. Nehmen wir dann aus dem Intervall einen von den Endpunkten verschiedenen Punkt und aus O den ihm entsprechenden Punkt fort, so erhielten wir wiederum einerseits eine nicht zusammenhängende Menge und andererseits eine zusammenhängende Menge.

Dagegen sieht man unmittelbar, daß z. B. jede Ellipse jedem Kreis homöomorph ist.

Dem Leser dürfte es ein leichtes sein, weitere Beispiele zu finden. So könnte man, falls wieder „Buchstaben" in Betracht gezogen werden, das gesamte lateinische (oder russische oder griechische) Alphabet derart in eine Reihe von Buchstabenklassen zerlegen, daß die Buchstaben aus einer und derselben Klasse jeweils untereinander homöomorph sind, während dies für Buchstaben aus verschiedenen Klassen nicht der Fall ist.

Wir wollen uns hier zum Abschluß noch mit einer anderen Fragestellung befassen, und zwar wollen wir am einfachsten Beispiel homöomorpher Abbildungen, nämlich der Abbildung eines abgeschlossenen Intervalls auf ein abgeschlossenes Intervall, untersuchen, wie man derartige Abbildungen realisieren kann. Wir haben bereits früher (§ 52) gesehen, daß jede in einem Intervall $[a, b]$ monoton wachsende oder fallende stetige Funktion $y = f(x)$ eine homöomorphe Abbildung eines Intervalls $[a, b]$ der x-Achse auf ein Intervall $[A, B]$ der y-Achse vermittelt. Wir wollen nun zeigen, daß hiervon auch die Umkehrung gilt, d. h., daß *man jede homöomorphe Abbildung von einem Intervall $[a, b]$ der x-Achse auf ein Intervall $[A, B]$ der y-Achse durch eine im Intervall $[a, b]$ monotone und stetige Funktion $y = f(x)$ realisieren kann.*

Da jede homöomorphe Abbildung eine spezielle eindeutige Abbildung ist, ist zunächst klar, daß es eine Funktion $y = f(x)$ gibt, welche die vorgelegte Abbildung „realisiert". Ferner ist klar, daß diese Funktion auch stetig sein muß; denn jede homöomorphe Abbildung ist ja stetig. Es bleibt also zu zeigen, daß diese Funktion auch monoton ist.

Zunächst sieht man sofort ein, daß den Endpunkten des einen Intervalls die Endpunkte des anderen Intervalls zugeordnet werden müssen; würde nämlich z. B. dem Punkt a ein innerer Punkt des Intervalls $[A, B]$ zugeordnet, so würde man nach Fortlassen dieser beiden Punkte in dem einen Fall eine zusammenhängende und im anderen Fall eine nicht zusammenhängende Menge erhalten. Es sind also nur die folgenden Fälle möglich: 1. $f(a) = A$, $f(b) = B$, 2. $f(a) = B$, $f(b) = A$; wir beschränken uns hier auf den ersten Fall.

Da $a < b$ und $A < B$ ist, kann die Funktion nicht fallend sein; wir beweisen, daß sie monoton wächst.

Wäre dies nicht der Fall, so gäbe es Punkte x' und x'' mit

$$a \leqq x' < x'' \leqq b,$$

für die

$$f(x') > f(x'')$$

gilt.

Wir nehmen zunächst an, daß $a < x'$. Da die Abbildung umkehrbar eindeutig ist, kann dann der Fall $f(a) = f(x')$ nicht eintreten, es muß also $f(a) < f(x')$ oder $f(a) > f(x')$ sein. Im ersten Fall betrachten wir irgendeine Zahl C, die zwischen der größeren der beiden Zahlen $f(a)$ und $f(x'')$ und der Zahl $f(x')$ liegt. Wegen der Stetigkeit der Funktion $f(x)$ muß diese den Wert C wenigstens einmal im Intervall (a, x') und wenigstens einmal im Intervall (x', x'') annehmen, und dies ist unmöglich, da die betrachtete Abbildung umkehrbar eindeutig ist. Im zweiten Fall kann man dieselben Überlegungen bezüglich der Intervalle (x', x'') und (x'', b)[1]) anstellen, wobei als C eine zwischen $f(x'')$ und der kleineren der Zahlen $f(x')$ und $f(b)$ gelegene Zahl zu nehmen ist. Dieselbe Überlegung führt auch im Fall $a = x'$ zum Ziel. Entsprechend verläuft der Beweis, falls $f(a) = B$ und $f(b) = A$.

§ 63. Obere und untere Grenze, Limes superior und Limes inferior von Zahlenmengen und Zahlenfolgen

Die Zahlenmengen (d. h. Mengen von reellen Zahlen) zeichnen sich vor anderen Mengen dadurch aus, daß ihre Elemente in natürlicher Weise, nämlich ihrer Größe nach angeordnet sind: Zwischen je zwei Elementen x und y einer Menge von reellen Zahlen besteht stets genau eine der Beziehungen

$$x < y \quad \text{oder} \quad x > y.$$

Wir legen im folgenden die gewöhnliche Metrik auf der Zahlengeraden zugrunde, bei der

$$\varrho(x, y) = |x - y|$$

ist. In bezug auf diese Metrik ist eine Menge E von Zahlen genau dann nach oben beschränkt, wenn es eine Zahl M gibt, so daß für jede Zahl x aus E

$$x < M \tag{54}$$

gilt. Ein Element x_0 aus der Menge E heißt *größtes Element* von E (*Maximum* von E), wenn für alle Elemente x aus E die Beziehung $x \leqq x_0$ besteht (wenn alle von x_0 verschiedenen Elemente aus E kleiner als x_0 sind). Offenbar *ist jede Menge E, die ein größtes Element enthält, nach oben beschränkt.* Die Umkehrung hiervon ist keineswegs richtig. Dies zeigen die folgenden Mengen: Die nach rechts offenen Intervalle $(0, 1)$ und $[0, 1)$, die Menge der echten Brüche, die Menge aller Zahlen der Form $1 - \frac{1}{n}$, die Menge aller Zahlen der Form $1 - \frac{1}{m} - \frac{1}{n}$ (wobei m und n natürliche Zahlen sind) usw.

---

[1]) Der Fall $x'' = b$ ist nicht möglich, da nach Annahme $f(x'') < f(x') < f(a)$ und überdies $f(a) < f(b)$ gilt.

Eine Zahl G heißt *obere Grenze* oder *Supremum* einer gegebenen, nach oben beschränkten Menge E, wenn sie folgende Bedingungen erfüllt:

1. Für alle Zahlen x aus E gilt $x \leqq G$;

2. Zu jeder noch so kleinen Zahl $\varepsilon > 0$ gibt es ein Element x aus E mit $x > G - \varepsilon$.

Offenbar *kann eine gegebene Menge E keine zwei verschiedenen oberen Grenzen besitzen.* Wären nämlich G_1 und G_2 obere Grenzen von E und etwa $G_1 < G_2$, so würden bei beliebigem $\varepsilon > 0$ für ein gewisses Element aus E die Ungleichungen

$$x > G_2 - \varepsilon \quad \text{und} \quad x \leqq G_1$$

gelten, was für $\varepsilon < G_2 - G_1$ einen Widerspruch ergibt.

Besitzt eine Menge E ein größtes Element M_0, so ist dieses zugleich obere Grenze von E.

Zunächst erfüllt nämlich die Zahl M_0 sicher die Bedingung 1.; ferner ist aber für $x = M_0$ bei beliebigem $\varepsilon > 0$ stets die Ungleichung $x > M_0 - \varepsilon$ erfüllt, d. h., es gilt auch die Bedingung 2.

Satz. *Jede nicht leere und nach oben beschränkte Zahlenmenge E besitzt eine obere Grenze.*

Es sei M eine Zahl, für die bei beliebigem x aus E die Ungleichung $x < M$ erfüllt ist, und L eine Zahl, die kleiner ist als irgendein Element aus E. Dann besitzt das Intervall $[L, M]$ der Länge $\Delta = M - L$ folgende Eigenschaften:

(α) Es enthält wenigstens ein Element der Menge E;

(β) es gibt in E kein Element, welches größer als jede Zahl aus diesem Intervall ist (d. h., welche größer ist als dessen rechter Endpunkt).

Wir betrachten jetzt die Intervalle $\left[L, \frac{L+M}{2}\right]$ und $\left[\frac{L+M}{2}, M\right]$. Wenigstens eines der beiden neuen Intervalle, das wir mit $[L_1, M_1]$ bezeichnen wollen, besitzt ebenfalls die Eigenschaften (α) und (β).

Sodann betrachten wir die Intervalle $\left[L_1, \frac{L_1+M_1}{2}\right]$ und $\left[\frac{L_1+M_1}{2}, M_1\right]$ von denen wieder wenigstens eines, das wir mit $[L_2, M_2]$ bezeichnen wollen, die Eigenschaften (α) und (β) besitzt, usw.

Setzen wir dieses Verfahren unbegrenzt fort, so erhalten wir eine Folge $\{[L_n, M_n]\}$ von Intervallen, von denen jedes die Eigenschaften (α) und (β) besitzt und die so beschaffen sind, daß

1. jedes Intervall $[L_n, M_n]$ das nachfolgende Intervall $[L_{n+1}, M_{n+1}]$ enthält;

2. die Folge der Längen der Intervalle gegen Null strebt; denn es ist $M_{n+1} - L_{n+1} = \frac{1}{2}(M_n - L_n)$, also $M_n - L_n = \frac{\Delta}{2^n}$.

Hieraus ergibt sich, daß die Folgen $\{L_n\}$ und $\{M_n\}$ gegen denselben Grenzwert G konvergieren: $\lim L_n = \lim M_n = G$, wobei offenbar für jedes n die Ungleichung $L_n \leqq G \leqq M_n$ erfüllt ist.

Wir wollen zeigen, daß die Zahl G die oben angegebenen Bedingungen 1. und 2. erfüllt, also obere Grenze von E ist.

Dazu nehmen wir an, es gäbe ein Element x_0 aus E mit $x_0 > G$. Wählen wir dann n so groß, daß $\frac{\Delta}{2^n} < x_0 - G$, so finden wir $x_0 > G + \frac{\Delta}{2^n} \geqq L_n + \frac{\Delta}{2^n} = M_n$, was der Eigenschaft ($\beta$) des Intervalls $[L_n, M_n]$ widerspricht. Also erfüllt die Zahl G die Bedingung 1.

Zum Beweis von 2. sei ε eine beliebige positive Zahl. Wir nehmen an, daß für alle Elemente x aus E die Ungleichung $x \leqq G - \varepsilon$ erfüllt ist. Dann wählen wir n so groß, daß $\frac{\Delta}{2^n} < \varepsilon$; offenbar ist dann für alle x aus E

$$x < G - \frac{\Delta}{2^n} \leqq M_n - \frac{\Delta}{2^n} = L_n,$$

im Widerspruch zur Eigenschaft (α) des Intervalls $[L_n, M_n]$. Also erfüllt G auch die Bedingung 2.

Die obere Grenze G einer Menge E kann auch auf folgende Weise charakterisiert werden: *Enthält E ein größtes Element, so ist G dieses größte Element; enthält E kein größtes Element, so ist G die kleinste Zahl M, für die* (54) *für alle x aus E gilt.*

Im ersten Fall ist G offenbar ein Element der Menge E, und es gibt unter den der Ungleichung (54) genügenden Zahlen M keine kleinste.[1]) Im zweiten Fall gehört G nicht der Menge E an, da sonst die Ungleichung $x < M = G$ für $x = G$ erfüllt sein müßte.

Ferner kann man behaupten:

Die obere Grenze G einer nach oben beschränkten Menge E ist entweder das größte isolierte Element der Menge E oder Häufungspunkt (vgl. § 36) *von E.* (Unter einem *isolierten Element* einer Menge verstehen wir dabei ein Element der Menge, zu dem es eine gewisse Umgebung gibt, die keine anderen Elemente der Menge enthält.)

In der Tat: Es sei G nicht größtes isoliertes Element von E. Wir betrachten eine beliebige Folge $\{\varepsilon_n\}$ von positiven Zahlen mit $\varepsilon_n \to 0$, und wählen unter Berücksichtigung der Eigenschaft 2. der oberen Grenze zu jedem ε_n ein Element x_n der Menge E mit $x_n > G - \varepsilon_n$ und $x_n \neq G$.[2]) Da wegen 1. außerdem noch $x_n \leqq G$ ist, strebt, wie leicht einzusehen ist, die Folge $\{x_n\}$ gegen G. Nun enthält aber die Folge $\{x_n\}$ unendlich viele verschiedene Glieder; hat man nämlich x_{n_1}, so erhält man ein Glied x_{n_2}, welches von x_{n_1} verschieden ist, wenn man n_2 so groß wählt, daß $\varepsilon_{n_2} < G - x_{n_1}$, usw. Hieraus folgt jedoch unmittelbar, daß in jeder Umgebung des Punktes G unendlich viele verschiedene Punkte der Menge E liegen, d. h., G ist Häufungspunkt der Menge E.

Dabei ist natürlich nicht der Fall ausgeschlossen, daß G zugleich größtes (nicht isoliertes) Element und Häufungspunkt von E ist. Als Beispiele hierfür können dienen: Der Punkt 1 im Falle des Intervalls $[0, 1]$ und der gleiche

[1]) Man kann als M jede Zahl wählen, die größer als G ist; unter diesen Zahlen gibt es jedoch keine kleinste.

[2]) Das ist möglich, weil G kein isoliertes Element von E ist.

Punkt im Falle der Menge, die aus allen Punkten der Gestalt $1-\frac{1}{n}$ und dem Punkt 1 besteht.

Aus dem zuletzt bewiesenen Satz folgt sofort:

Jede beschränkte abgeschlossene Menge E enthält ein größtes Element, und zwar ist die obere Grenze G von E größtes Element von E.

Nach dem soeben bewiesenen Satz ist nämlich G entweder größte isolierte Zahl in E oder einer ihrer Häufungspunkte. Im ersten Fall sind wir dann fertig; im zweiten Fall gehört G zur Ableitung E', und da E abgeschlossen sein sollte (d. h. $E' \subseteq E$, vgl. § 60) auch zu E.

Wir führen als nächsten wesentlichen Begriff den des oberen Häufungswertes einer Menge E ein:

Eine Zahl L heißt *oberer Häufungswert* oder *Limes superior* einer nach oben beschränkten Menge E, wenn sie folgende Bedingungen erfüllt:

1. Für jedes noch so kleine $\varepsilon > 0$ genügen fast alle (d. h. alle bis auf endlich viele) Elemente x aus E der Ungleichung $x < L + \varepsilon$;

2. Für jedes noch so kleine $\varepsilon > 0$ genügen unendlich viele Elemente x aus E der Ungleichung $x > L - \varepsilon$.

Hieraus folgt unmittelbar, daß *keine endliche Menge einen Limes superior besitzt*. Daher ist die angegebene Definition des Limes superior der folgenden Definition gleichwertig:

Der Limes superior einer Menge E ist der größte Häufungswert dieser Menge (d. h. das größte Element der Ableitung E').

In der Tat: Erfüllt der Punkt L die oben angegebenen Bedingungen 1. und 2., so folgt sofort, daß in jeder Umgebung des Punktes L unendlich viele Elemente aus E liegen, d. h., L ist ein Häufungspunkt von E. Darüber hinaus ist aber L auch der größte Häufungspunkt, weil die Existenz eines größeren Häufungspunktes L' der Bedingung 1. widersprechen würde (wenn dort ε kleiner als $L' - L$ gewählt wird).

Ist umgekehrt L der größte Häufungspunkt von E, so kann die Bedingung $x \geqq L + \varepsilon$ nur von endlich vielen Elementen der Menge E erfüllt werden; denn anderenfalls würde nach dem Satz von Bolzano und Weierstrass (§ 36, Bemerkung 2) die Menge E einen Häufungspunkt L' besitzen, für den $L' \geqq L + \varepsilon$ gilt. Damit ist die Eigenschaft 1. bewiesen. Die Eigenschaft 2. ergibt sich sodann aus der Tatsache, daß L Häufungspunkt von E ist.

Satz. *Jede nach oben beschränkte unendliche Menge E besitzt einen Limes superior.*

In der Tat: Im betrachteten Fall ist die Ableitung E' von E nicht leer; ferner ist sie (vgl. § 37, S. 153) abgeschlossen; schließlich ist sie nach oben beschränkt. Daher enthält sie nach dem Vorangehenden ein größtes Element.

Trivial gilt: *Der Limes superior einer unendlichen Menge ist nicht größer als ihre obere Grenze*:

$$L \leqq G. \tag{55}$$

Man überlegt sich sofort, daß in (55) das Ungleichheits- oder Gleichheitszeichen steht, je nachdem, ob G isoliertes Element von E oder Häufungspunkt von E ist.

Vollkommen analog zu den oben angeführten Begriffen, Beziehungen und Sätzen sind folgende Begriffe, Beziehungen und Sätze: Eine Menge E von Zahlen ist genau dann **nach unten beschränkt**, wenn es eine Zahl m gibt, so daß für alle Zahlen x aus E die Beziehung

$$x > m \tag{56}$$

besteht.

Ein Element m_0 aus E, für welches für alle x aus E die Ungleichung $x \geqq m_0$ erfüllt ist, heißt ein *kleinstes Element* (*Minimum*) von E.

Eine Zahl g wird *untere Grenze* oder *Infimum* der Menge E genannt, wenn sie folgende Bedingungen erfüllt:

1. Für alle x aus E ist $x \geqq g$.

2. Zu jeder noch so kleinen Zahl $\varepsilon > 0$ gibt es wenigstens ein Element x_0 aus E mit $x_0 < g + \varepsilon$.

Besitzt eine Menge E ein kleinstes Element m_0, so ist dieses zugleich untere Grenze von E:

$$g = m_0. \tag{57}$$

Satz. *Jede nicht leere, nach unten beschränkte Menge E besitzt eine untere Grenze.*

Die untere Grenze einer Menge E ist entweder kleinstes Element von E oder die größte der Zahlen m, für welche die Ungleichung (56) *für alle x aus E erfüllt ist.*

Außerdem ist g entweder das kleinste isolierte Element der Menge E oder Häufungspunkt von E.

Daraus folgt: *Jede abgeschlossene Menge besitzt ein kleinstes Element.*

Eine Zahl l heißt *untere Häufungsgrenze* oder *Limes inferior* einer Menge E, wenn sie folgende Eigenschaften besitzt:

1. Für jedes noch so kleine $\varepsilon > 0$ genügen fast alle Elemente x aus E der Ungleichung $x > l - \varepsilon$.

2. Für jedes noch so kleine $\varepsilon > 0$ genügen unendlich viele Elemente x aus E der Ungleichung $x < l + \varepsilon$.

Ist E eine nach unten beschränkte unendliche Menge, so ist l genau der kleinste Häufungspunkt von E.

Daraus folgt der

Satz. *Jede nach unten beschränkte unendliche Menge E besitzt einen Limes inferior.*

Der Limes inferior einer Menge E ist (sofern er existiert) *nicht kleiner als die untere Grenze von E:*

$$l \geqq g. \tag{58}$$

Da der Limes inferior einer Menge E nicht größer sein kann als der Limes superior von E (vorausgesetzt, daß diese existieren), gelten insgesamt die Ungleichungen

$$g \leqq l \leqq L \leqq G. \tag{59}$$

Der Fall $l = L$ tritt dann und nur dann ein, wenn die Menge E nur einen Häufungspunkt besitzt. In diesem Fall bezeichnet man den gemeinsamen Wert von l und L mitunter als *Grenzwert* der Menge E.

Der Fall $g = G$ tritt dann und nur dann ein, wenn die Menge E nur aus einem einzigen Punkt besteht.

Wir haben bisher ausschließlich Mengen betrachtet, die sowohl nach oben als auch nach unten beschränkt sind. Ist eine Menge nicht nach oben beschränkt, so setzt man gelegentlich

$$L = G = +\infty;$$

ist sie nicht nach unten beschränkt, so setzt man entsprechend

$$l = g = -\infty.$$

Der Begriff der oberen und der unteren Grenze sowie die Begriffe des Limes superior und des Limes inferior lassen sich ähnlich auch für Folgen erklären. Dazu beachten wir, daß zu jeder Zahlenfolge $\{a_n\}$ die aus den Gliedern dieser Folge gebildete Menge E gehört.

Als *obere* bzw. *untere Grenze* von $\{a_n\}$ bezeichnet man dann gerade die obere bzw. die untere Grenze dieser Menge E, unabhängig davon, ob die Menge E endlich oder unendlich ist. Damit stimmt die Definition der oberen (bzw. unteren) Grenze einer Folge fast wörtlich mit der Definition der oberen (bzw. unteren) Grenze einer Menge überein; man erkennt nämlich sofort, daß die Zahl G dann und nur dann obere Grenze der Folge $\{a_n\}$ ist, wenn sie folgende Eigenschaften besitzt.

1. Für alle n $(n = 1, 2, 3, \ldots)$ gilt $a_n \leqq G$.

2. Zu jeder noch so kleinen Zahl $\varepsilon > 0$ gibt es ein a_n, welches der Ungleichung $a_n > G - \varepsilon$ genügt.

Bei der Definition des Limes superior und Limes inferior einer Folge $\{a_n\}$ müssen wir dagegen etwas vorsichtiger vorgehen. Eine Zahl L heißt *Limes superior* einer gegebenen nach oben beschränkten Folge $\{a_n\}$, wenn sie folgende Eigenschaften besitzt:

1. Für jede noch so kleine Zahl $\varepsilon > 0$ ist die Ungleichung $a_n < L + \varepsilon$ für fast alle Glieder a_n der Folge (d. h. alle Glieder der Folge von einem gewissen Glied a_N an) erfüllt;

2. Für jedes noch so kleine $\varepsilon > 0$ ist die Ungleichung $a_n > L - \varepsilon$ für unendlich viele Glieder a_n der Folge erfüllt.

Auf Grund dieser Definition ergibt sich sofort, daß *jede nach oben beschränkte Folge einen Limes superior besitzt*, auch dann, wenn die aus den Gliedern der Folge gebildete Menge E nur endlich viele Elemente enthält.

Entsprechendes gilt für den Limes inferior.

Zum Beispiel ist der Limes superior bzw. Limes inferior der Folge $\{(-1)^n\}$ gleich $+1$ bzw. -1, obwohl die Menge der Folgeglieder endlich ist, nämlich nur aus den beiden Zahlen $+1$ und -1 besteht.[1])

[1]) Bei dieser Gelegenheit sei erwähnt (was allerdings im hier betrachteten Zusammenhang ohne Bedeutung ist), daß man nicht jede Menge von Zahlen zu einer Folge anordnen kann; dies ist unmöglich, sobald die Menge überabzählbar viele Elemente enthält (vgl. EdEM Band I, Seite 76).

Für die obere bzw. untere Grenze einer Menge bzw. einer Folge sind die folgenden Bezeichnungen üblich:

$$G = \sup E, \quad g = \inf E$$

und

$$G = \sup\{a_n\}, \quad g = \inf\{a_n\}.$$

Für die obere bzw. die untere Häufungsgrenze verwendet man die Bezeichnungen

$$L = \overline{\lim} E, \quad l = \underline{\lim} E$$

und

$$L = \overline{\lim} a_n, \quad l = \underline{\lim} a_n.$$

Für die Menge oder Folge[1]) der Zahlen $\left\{(-1)^n \frac{n+1}{n}\right\}$ $(n = 1, 2, 3, \ldots)$ ist z. B.

$$G = \sup\left\{(-1)^n \frac{n+1}{n}\right\} = \frac{3}{2}, \qquad g = \inf\left\{(-1)^n \frac{n+1}{n}\right\} = -2,$$

$$L = \overline{\lim}\,(-1)^n \frac{n+1}{n} = 1, \qquad l = \underline{\lim}\,(-1)^n \frac{n+1}{n} = -1.$$

Die Schreibweise $\overline{\lim}\, a_n = L$ kann, wovon man sich leicht überzeugt, in dem Sinne ausgelegt werden, daß für eine gewisse Teilfolge $\{a_{n_p}\}$ der Folge $\{a_n\}$ die Beziehung $\lim a_{n_p} = L$ besteht, während es für kein $\varepsilon > 0$ eine Teilfolge $\{a_{n_p}\}$ gibt, für welche $\lim a_{n_p} = L + \varepsilon$ ist.

Entsprechend versteht man unter der Schreibweise

$$\overline{\lim_{x \to c}} f(x) = L, \tag{60}$$

daß es eine Folge $\{x_n\}$ mit $x_n \to c$ gibt, für welche $\lim f(x_n) = L$ gilt, während es für kein $\varepsilon > 0$ eine Folge $\{x_n\}$ mit $x_n \to c$ gibt, die die Bedingung $\lim f(x_n) = L + \varepsilon$ erfüllt.

Entsprechendes gilt für den Limes inferior.

Die Zahlen $\overline{\lim\limits_{x \to c}} f(x)$ und $\underline{\lim\limits_{x \to c}} f(x)$ bezeichnet man als *Limes superior* bzw. *Limes inferior* (oder auch als *oberen* bzw. *unteren Grenzwert*) der Funktion $f(x)$ im Punkte $x = c$. Offenbar ist stets

$$\underline{\lim_{x \to c}} f(x) \leqq \overline{\lim_{x \to c}} f(x),$$

wobei dann und nur dann das Gleichheitszeichen steht, wenn die Funktion $f(x)$ im Punkte $x = c$ stetig ist.

---

[1]) Hierbei ist es ohne Belang, ob man die Folge oder Menge betrachtet, da die Glieder der Folge paarweise verschieden sind.

Für die Funktion $f(x) = \sin\frac{1}{x}$ (vgl. § 41, S. 171) ist z. B.

$$\overline{\lim_{x\to 0}} \sin\frac{1}{x} = +1, \quad \underline{\lim_{x\to 0}} \sin\frac{1}{x} = -1.$$

Macht man in (60) die zusätzliche Einschränkung $x > c$ bzw. $x < c$, so gelangt man zum Begriff des *rechtsseitigen* bzw. *linksseitigen Limes superior* der Funktion $f(x)$ im Punkte $x = c$; entsprechend erklärt man den *rechtsseitigen* bzw. *linksseitigen Limes inferior*.

Es sei zum Abschluß darauf hingewiesen, daß sich bei Verwendung der im vorangehenden eingeführten Begriffe die Beweise vieler Sätze erheblich vereinfachen und vereinheitlichen lassen. Einige Beispiele mögen dies erläutern, wobei wir uns auf Sätze beschränken wollen, welche uns bereits bekannt sind. Wir geben hier nur kurze Beweisskizzen und überlassen die genaue Durchführung der Beweise dem Leser.

1. *Jede nach oben beschränkte unendliche Menge besitzt einen Limes superior* (vgl. S. 275).

Es sei E die gegebene Menge. Wir betrachten dann die Menge $\mathfrak{E}$ derjenigen Punkte x, die nur von endlich vielen Punkten aus E übertroffen werden. Die untere Grenze der Menge $\mathfrak{E}$ ist gerade Limes superior der Menge E.

2. *Jede in einem Intervall $[a, b]$ stetige Funktion $f(x)$, die in den Endpunkten a und b Werte mit verschiedenen Vorzeichen annimmt, verschwindet in wenigstens einem Punkte des Intervalls $[a, b]$* (Satz von Bolzano, § 47, S. 204).

Es sei etwa $f(a) < 0$ und $f(b) > 0$. Dann betrachten wir die Menge aller derjenigen Punkte x des Intervalls $[a, b]$, in denen $f(x) > 0$ ist. Die untere Grenze dieser Menge ist ein Punkt des Intervalls $[a, b]$, in welchem die Funktion $f(x)$ den Wert Null annimmt.

3. *Es sei $f(x)$ eine im Intervall $[a, b]$ stetige Funktion. Dann gibt es unter den Werten der Funktion $f(x)$ im Intervall einen größten* (Satz von Weierstrass, § 47, S. 207).

Es sei $\mathfrak{E}$ die Menge der Werte von $f(x)$ im Intervall $[a, b]$, G ihre obere Grenze. Wir nehmen an, es gäbe keinen Punkt $x = \xi$, in welchem $f(\xi) = G$ ist. Dann betrachten wir die Funktion

$$\varphi(x) = \frac{1}{G - f(x)}.$$

Sie ist ebenso wie die Funktion $f(x)$ stetig im Intervall $[a, b]$ und daher (Satz II auf S. 206) nach oben beschränkt:

$$\varphi(x) < M.$$

In diesem Fall ist für alle x aus dem Intervall $[a, b]$

$$f(x) < G - \frac{1}{M},$$

im Widerspruch dazu, daß G die obere Grenze der Werte von $f(x)$ im Intervall $[a, b]$ sein sollte.

4. *Jedes abgeschlossene Intervall $\mathfrak{A} = [a, b]$ ist zusammenhängend* (§ 60, S. 264).

Angenommen, die Menge $\mathfrak{A}$ wäre nicht zusammenhängend. Dann gäbe es Mengen $\mathfrak{M}$ und $\mathfrak{N}$ derart, daß

1. $\mathfrak{A} = \mathfrak{M} + \mathfrak{N}$,
2. $\mathfrak{M}\overline{\mathfrak{N}} + \overline{\mathfrak{M}}\mathfrak{N} + \overline{\mathfrak{M}}\overline{\mathfrak{N}} = \mathfrak{O}$,
3. $\mathfrak{M} \neq \mathfrak{O}$, $\mathfrak{N} \neq \mathfrak{O}$.

Es möge der Punkt a etwa der Menge $\mathfrak{M}$ angehören. In diesem Fall sei g die untere Grenze der Menge $\mathfrak{N}$. Der Punkt g gehört dann entweder zu $\mathfrak{N}$ oder zu $\mathfrak{M}$. Im ersten Fall liegen alle Punkte x, die der Ungleichung $a \leqq x < g$ genügen, in $\mathfrak{M}$, es wäre also g Häufungspunkt von $\mathfrak{M}$, im Widerspruch zu $\mathfrak{M}'\mathfrak{N} = \mathfrak{O}$. Im zweiten Fall gäbe es dagegen in $\mathfrak{N}$ Punkte, die beliebig nahe bei g liegen, so daß g Häufungspunkt von $\mathfrak{N}$ wäre, im Widerspruch zu $\mathfrak{M}\mathfrak{N}' = \mathfrak{O}$. Damit ist auch dieser Satz bewiesen.

Der Leser möge selbst entscheiden, in welchem Maße die Beweise der vorangehenden Sätze gegenüber den früheren Beweisen an Klarheit und Kürze dadurch gewonnen haben, daß systematisch der Begriff der oberen bzw. unteren Grenze verwendet wurde.

I. P. NATANSON

ABLEITUNGEN, INTEGRALE UND REIHEN

EINLEITUNG

Bei der Erforschung von Naturvorgängen und in der Technik hat man es fast immer mit veränderlichen Größen oder, wie man kurz sagt, mit Veränderlichen zu tun. So sind z. B. die Lufttemperatur, der Dampfdruck in einem Kessel, die Stromstärke in einer elektrischen Leitung, die Geschwindigkeit eines Flugzeuges usw. Größen, die sich mit der Zeit ändern und daher veränderliche Größen darstellen. Jedoch wurde erst im 16. und 17. Jahrhundert, und zwar unter dem Einfluß der sich stürmisch entwickelnden Naturwissenschaften und Technik, der allgemeine Begriff der veränderlichen Größe in die Mathematik aufgenommen. Seit dieser Zeit stellen veränderliche Größen ein grundlegendes Objekt mathematischer Forschung dar. Hierfür bilden die mathematischen Arbeiten von DESCARTES ein großartiges Beispiel.

F. ENGELS schreibt: „Der Wendepunkt in der Mathematik war DESCARTES' *variable Größe*. Damit die *Bewegung* und *damit die Dialektik* in der Mathematik, und *damit auch sofort mit Notwendigkeit die Differential- und Integralrechnung*, die auch sofort anfängt und durch NEWTON und LEIBNIZ im ganzen und großen vollendet, nicht erfunden."[1])

Man kann kurz sagen, daß die Analysis die Mathematik der veränderlichen Größen ist. Jedoch muß zu einer genaueren Charakterisierung der Analysis sofort gesagt werden, daß dort veränderliche Größen nicht isoliert, sondern in Abhängigkeit voneinander studiert werden. Der präzise mathematische Begriff, der den Gedanken einer Abhängigkeit veränderlicher Größen ausdrückt, ist der Funktionsbegriff. Dies ist der eigentlich fundamentale und zugleich wichtigste Begriff der Analysis.

Die grundlegenden Begriffsbildungen der Analysis, also der Begriff der veränderlichen Größe und der Funktionsbegriff, besitzen auch für die Elementarmathematik eine außerordentliche Bedeutung. So ist z. B. die gesamte Theorie der trigonometrischen Funktionen in der Hauptsache ein elementarer Abschnitt der Analysis. In der Schule werden bereits die rationalen, die einfachsten irrationalen algebraischen, aber auch schon gewisse nichtalgebraische (transzendente) Funktionen untersucht. Zu den letzteren gehören die Potenzfunktionen mit irrationalem Exponenten, die Exponentialfunktionen und die logarithmischen Funktionen, aber auch die trigonometrischen Funktionen. In der Schule werden ferner die Grundzüge der Theorie der Grenzwerte und der einfachsten Reihen vermittelt. In der Geometrie werden u. a. der Inhalt gewisser einfacher krummlinig begrenzter Figuren und von gekrümmten Flächen begrenzter Körper als

[1]) F. ENGELS, Dialektik der Natur, Berlin 1952, S. 275.

Grenzwert der Inhalte von anderen sie approximierenden Figuren und Körpern definiert.

Zu den wichtigsten Teilen der Analysis, welche vielfältige Anwendung in sämtlichen Gebieten der Mathematik, der Physik und der Technik finden, gehören die Differential- und Integralrechnung und die allgemeine Theorie der Reihen. Diese Teile sind heute (in der Sowjetunion — *Anm. d. wiss. Red.*) nicht in das Programm der allgemeinbildenden Schulen aufgenommenen. Jedoch sind der Begriff der Ableitung und der Integralbegriff schon seit langem grundlegende Begriffe der Mathematik, welche einen hervorragenden Bildungswert besitzen, so daß sie im Laufe der Zeit (wenn auch nur in sehr bescheidenem Maße) in den Mathematikunterricht der Mittelschulen aufgenommen werden sollten.

Der vorliegende Artikel enthält in systematischer Anordnung die grundlegenden Tatsachen über Ableitungen, Integrale und Reihen. Der gebotene Stoff übersteigt in vielen Fällen das, was man im Schulunterricht bringen kann. Leitende Gesichtspunkte bei der Stoffauswahl sind, wie auch in den anderen Artikeln der „Enzyklopädie der Elementarmathematik", die Erfordernisse beim Studium und der Untersuchung der elementaren Funktionen sowie bei der Lösung einer Reihe von geometrischen und physikalischen Aufgaben (wie z. B. der Konstruktion der Tangente an eine Kurve, der Bestimmung des Flächeninhalts ebener Figuren, der Länge von Kurven und der Oberfläche einfachster Körper, der Bestimmung der Geschwindigkeit bei gegebenem Wege und des Weges bei gegebener Geschwindigkeit).

Es seien hier einige Etappen in der Entwicklungsgeschichte der Analysis erwähnt, wobei nicht beabsichtigt ist, einen auch nur einigermaßen vollständigen Abriß ihrer Entwicklungsgeschichte zu geben.

Schon in der antiken Wissenschaft lassen sich einzelne wichtige Gedankengänge der Analysis beobachten. So stellt die klassische „Ausschöpfungsmethode" das Urbild einer Reihentheorie dar. Weiter findet sich in einigen Arbeiten von Archimedes in den ersten Anfängen ein Gedankengang, der für die spätere Integralrechnung charakteristisch ist.

Erst etwa in der Mitte des 17. Jahrhunderts wurden die Ergebnisse der altgriechischen Wissenschaft auf dem Gebiete der Mathematik wesentlich übertroffen, und bereits gegen Ende des 17. Jahrhunderts wurde der Aufbau der Differential- und Integralrechnung durch die Arbeiten von Newton (1642 bis 1727) und Leibniz (1646—1716) abgeschlossen.

Wenn wir hier von einem „Abschluß" des Entstehungsprozesses der Analysis sprechen, so meinen wir damit, daß nun die Leitprinzipien der Analysis aufgestellt und die Wechselbeziehungen zwischen den Problemen der Differentialrechnung und denen der Integralrechnung erkannt waren. Es wäre falsch, diesen Abschluß so zu verstehen, daß nach den Arbeiten von Newton und Leibniz die Entwicklung der Analysis beendet gewesen wäre. Genau das Gegenteil ist der Fall: Das wissenschaftliche Schaffen auf diesem Gebiet wurde im 18. und 19. Jahrhundert und bis auf den heutigen Tag intensiv und erfolgreich fortgesetzt. So werden gegenwärtig in den verschiedenen Teilgebieten der Analysis jährlich mehr als tausend Arbeiten veröffentlicht.

In Rußland ist der Beginn der Beschäftigung mit Problemen der Analysis untrennbar mit dem Namen des bekannten Mathematikers L. Euler (1707 bis 1783) verbunden. Euler, der von Geburt Schweizer war, verbrachte den größten Teil seines Lebens in Petersburg. Er war Mitglied der Petersburger Akademie der Wissenschaften. Die bedeutendsten Ergebnisse Eulers, der sich mit den verschiedensten Problemen der Mathematik und der Mechanik beschäftigt hat, liegen auf dem Gebiete der Differential- und Integralrechnung.

Eine Reihe von wichtigen Ergebnissen auf dem Gebiet der Integralrechnung (Integrationsverfahren, Theorie der mehrfachen Integrale, Differentialgleichungen, Variationsrechnung) wurden von M. W. Ostrogradski (1801—1861) gefunden. Seine bekannte Formel für die Umformung mehrfacher Integrale (es handelt sich dabei um den Satz, der in der deutschen Literatur als Gausssscher Integralsatz bezeichnet wird; — *Anm. d. wiss. Red.*) und ein von ihm geschaffenes Verfahren zur Integration rationaler Funktionen sind heute in allen Lehrbüchern der Differential- und Integralrechnung zu finden.

Außer Ostrogradski beschäftigte sich zu dieser Zeit noch V. J. Bunjakowski (1804—1889) mit Fragen der Analysis. Von ihm stammen eine ganze Reihe von Veröffentlichungen aus diesem Gebiet. Interessante Untersuchungen im Bereich der Analysis wurden auch von dem großen Geometer N. I. Lobatschewski (1792—1856) durchgeführt.

Begründer einer ganzen analytischen Schule war der geniale russische Mathematiker P. L. Tschebyschew (1821—1894). Er selbst ist Autor von Arbeiten zur Integralrechnung (und zwar sowohl zum Problem der Integrierbarkeit in elementaren Funktionen als auch hinsichtlich der näherungsweisen Berechnung von bestimmten Integralen). Sein Hauptverdienst liegt allerdings in seinen Untersuchungen zur Approximationstheorie. Die Weiterentwicklung dieser Untersuchungen führte zu einem neuen Zweig der Analysis, der konstruktiven Funktionentheorie, welche größte Bedeutung für die Anwendungen besitzt.

Die hervorragendsten Vertreter der Tschebyschewschen Schule waren A. A. Markoff (1856—1922) und A. M. Ljapunoff (1857—1918). Die Traditionen dieser Schule leben noch heute an der Leningrader Universität fort.

Im 20. Jahrhundert entwickelte sich unter N. N. Lusin (1883—1950) und D. F. Egorow (1869—1931) die Moskauer Schule. Sie kam in der sowjetischen Periode zur vollen Blüte und nimmt heute hinsichtlich der Vielzahl der behandelten Probleme und der Bedeutsamkeit der erzielten Resultate unbestreitbar den ersten Platz in der Welt ein. Ihre hervorragendsten Vertreter auf dem Gebiet der Analysis sind A. N. Kolmogoroff, I. G. Petrowski, I. I. Priwalow, D. E. Menschow und andere.

Einer der bedeutendsten lebenden Mathematiker ist S. N. Bernstein (geb. 1880). Er setzte die Untersuchungen Tschebyschews über die Theorie der Approximation von Funktionen (und über Wahrscheinlichkeitsrechnung) fort. Durch ihn wurden eine ganze Reihe von jungen sowjetischen Mathematikern herangebildet.

Der allgemeine Aufschwung der Kultur und der Wissenschaft machte sich auch in der Mathematik lebhaft bemerkbar. Oben wurde bereits die hervorragende Bedeutung der Moskauer mathematischen Schule erwähnt. Sehr bedeu-

tende und umfangreiche Untersuchungen werden aber auch in den anderen wissenschaftlichen Zentren der Sowjetunion geführt, so in Leningrad, Kiew, Charkow, Odessa, Tbilissi, Kasan, Erewan usw.

Eine ausführliche Darstellung der in der UdSSR erzielten Fortschritte in den mathematischen Wissenschaften (bis 1947) findet man in dem Gemeinschaftswerk „Математика в СССР за 30 лет" (30 Jahre Mathematik in der UdSSR), Gostechisdat 1948.

Kapitel I

ABLEITUNGEN

§ 1. Ableitung und Differential

1. Probleme, die auf den Begriff der Ableitung führen. Der wichtigste Begriff der Differentialrechnung ist der Begriff der Ableitung. Wir betrachten zunächst einige konkrete Aufgaben, die auf diesen Begriff führen.

A. Das Problem der Geschwindigkeit. Ein Punkt M bewege sich auf einer Geraden.[1]) Bekanntlich versteht man unter der *mittleren Geschwindigkeit* des Punktes in einem gegebenen Zeitintervall das Verhältnis des Weges, den der Punkt im Verlaufe des Zeitintervalls zurückgelegt hat, zu der dazu benötigten Zeit.

Man sieht unmittelbar ein, daß die mittlere Geschwindigkeit eines Punktes in einem Zeitintervall keinen Aufschluß über die Art der Bewegung in einem vorliegenden Moment zu geben vermag. Dadurch wird man in der Mechanik zur Betrachtung des Begriffs der *Geschwindigkeit eines Punktes in einem gegebenen Zeitmoment* geführt. Hierunter versteht man den Grenzwert der mittleren Geschwindigkeit des Punktes bei Übergang zu einem „unendlich kleinen"[2]) Zeitintervall, welches mit dem betrachteten Moment beginnt (oder mit diesem endet).

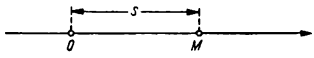


Abb. 1

Häufig spricht man, um den Unterschied dieses Geschwindigkeitsbegriffs vom Begriff der früher definierten mittleren Geschwindigkeit auszudrücken, von der „wahren" Geschwindigkeit des Punktes M im gegebenen Zeitpunkt.

Wir denken uns nun auf der Geraden, auf der sich der Punkt M bewegt, einen bestimmten Anfangspunkt O, eine bestimmte Längeneinheit und eine ausgezeichnete Richtung festgelegt. Dann ist die Lage des Punktes M in jedem Zeitpunkt eindeutig durch den Abstand OM des Punktes M vom Anfangspunkt O bestimmt (Abb. 1). Dieser Abstand ist gleich einer gewissen (positiven oder negativen oder verschwindenden) Zahl s, welche natürlich von dem betrachteten Zeitpunkt abhängt. Dieser Zeitpunkt kann dabei durch Angabe der Anzahl t der Zeiteinheiten beschrieben werden, welche ihn von

[1]) Im Fall einer krummlinigen Bewegung hat die Geschwindigkeit Vektorcharakter, so daß ihre Definition komplizierter wird.

[2]) Das heißt bei Übergang zu immer kleineren Intervallen, wobei die Länge der Intervalle (die Dauer der Zeitabschnitte) gegen Null strebt.

einem beliebig gewählten Anfangszeitpunkt trennen. Damit wird der Abstand $OM = s$ eine Funktion der Arguments t:

$$s = f(t), \tag{1}$$

wobei die Bewegung des Punktes vollständig durch diese Funktion beschrieben wird. Man nennt die Gleichung (1) die *Gleichung der Bahnkurve* für den betrachteten Punkt. Beispiele für derartige Gleichungen sind

$$s = t^2, \quad s = 2t^3 + 1, \quad s = \sin t, \quad s = \frac{gt^2}{2}.$$

Wir wollen nun untersuchen, wie man bei gegebener Bewegungsgleichung (1) die „wahre" Geschwindigkeit des Punktes M im Zeitpunkt t (d. h. in dem Zeitpunkt, in welchem seit dem Anfangszeitpunkt genau t Zeiteinheiten vergangen sind) ermitteln kann. Dazu betrachten wir neben dem Zeitpunkt t noch einen weiteren Zeitpunkt $t + \Delta t$. In diesem Zeitpunkt besitzt der Punkt M vom Punkt O den Abstand

$$s + \Delta s = f(t + \Delta t).$$

Daher wird der Weg Δs, den der Punkt M im Verlauf desjenigen Zeitintervalls durchlaufen hat, das zum Zeitpunkt t beginnt und zum Zeitpunkt $t + \Delta t$ endet, durch

$$\Delta s = f(t + \Delta t) - f(t)$$

gegeben. Da die Länge des betrachteten Zeitintervalls gleich Δt ist, erhalten wir als mittlere Geschwindigkeit in diesem Zeitintervall

$$\frac{\Delta s}{\Delta t} = \frac{f(t + \Delta t) - f(t)}{\Delta t}.$$

Die wahre Geschwindigkeit v des Punktes M im Zeitpunkt t ergibt sich dann, wenn wir den Grenzwert dieses Verhältnisses für $\Delta t \to 0$ berechnen:

$$v = \lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t} = \lim_{\Delta t \to 0} \frac{f(t + \Delta t) - f(t)}{\Delta t}. \tag{I}$$

Damit führt die physikalische Aufgabe, die wahre Geschwindigkeit eines Punktes M zu bestimmen, auf die rein analytische Aufgabe, den Grenzwert (I) zu berechnen.

Als nächstes betrachten wir eine Aufgabe vollkommen anderer Art, die uns jedoch auf die Bestimmung eines Grenzwertes ganz ähnlicher Natur führen wird.

B. Das Tangentenproblem. In den Anfängen der Elementargeometrie definiert man als Tangente an einen gegebenen Kreis eine Gerade, welche mit dem Kreis genau einen Punkt gemeinsam hat. Geht man zur Betrachtung allgemeinerer Kurven über, so reicht jedoch diese Definition nicht mehr aus. So wird man z. B. die y-Achse nicht als Tangente an die Parabel $y = x^2$ ansehen (Abb. 2), obwohl beide nur einen Punkt gemeinsam haben.

Aus diesem Grunde ist in der höheren Mathematik eine andere Definition für die Tangente üblich, und zwar versteht man unter der *Tangente* in einem Punkte M einer Kurve K diejenige Gerade MT (vgl. Abb. 3), welche als Grenzlage der Sekante MN auftritt, die durch M und einem anderen Punkt N der Kurve K führt, wenn sich N, ohne K zu verlassen, unbegrenzt dem Punkt M nähert.

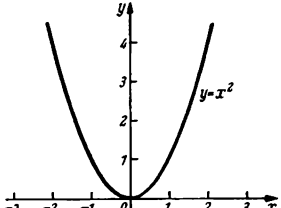


Abb. 2

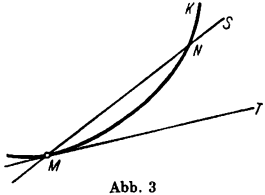


Abb. 3

Man sieht unmittelbar, daß im Fall eines Kreises diese Definition mit der oben angegebenen Definition gleichwertig ist.

Unsere Aufgabe soll es nun sein, die Gleichung für die Tangente in einem Punkte $M(x, y)$ einer Kurve

$$y = f(x) \tag{2}$$

zu bestimmen, wobei $f(x)$ eine beliebige stetige Funktion ist.

Da uns ein Punkt der Tangente, nämlich M, bereits bekannt ist, genügt es, den Steigungskoeffizienten k der Tangente zu bestimmen, d. h. den Tangens des Winkels α, unter dem die Tangente MT die x-Achse schneidet (Abb. 4).

Dazu wählen wir auf der Kurve einen zweiten Punkt $N(x + \Delta x, y + \Delta y)$ und legen durch M und N die Sekante MN. Wie aus der analytischen Geometrie bekannt ist, wird der Steigungskoeffizient der Sekante durch

$$k^* = \frac{\Delta y}{\Delta x}$$

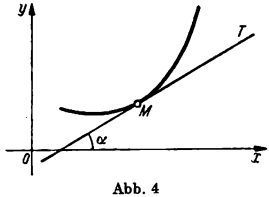


Abb. 4

gegeben. Da die Tangente die Grenzlage der Sekante ist, hat man daher

$$k = \lim_{N \to M} k^*. \text{ [1]} \tag{3}$$

Nun ist aber, da M und N beide auf der Kurve (2) liegen,

$$y = f(x), \quad y + \Delta y = f(x + \Delta x)$$

[1] Der strenge Beweis für (3) ergibt sich ohne Schwierigkeiten aus der Stetigkeit der Funktion tg α.

und daher

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

Wenn nun der Punkt N gegen den Punkt M strebt, so strebt Δx gegen Null.[1]) Also kann die Beziehung (3) auch in der Form

$$k = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \qquad \text{(II)}$$

geschrieben werden.

Damit ist das Tangentenproblem auf die Bestimmung des Grenzwertes (II) zurückgeführt, der sich nur in den Bezeichnungen vom Grenzwert (I) unterscheidet.

Wir wollen schließlich noch ein drittes Problem betrachten, welches ebenfalls auf einen Grenzwert dieses Typs führt.

C. Das Problem der Dichte. Unter der *mittleren Dichte* eines geradlinigen Stabes versteht man das Verhältnis seiner Masse zu seiner Länge (wir beschränken uns hier auf die Betrachtung der „linearen Dichte" eines Stabes; man wird hierzu geführt, wenn man die Dicke eines „realen" Stabes vernachlässigt). Um zu einer genauen Beschreibung der Massenverteilung längs des Stabes zu gelangen, führt man den Begriff der „wahren" Dichte des Stabes in seinen einzelnen Punkten ein. Hierunter versteht man den Grenzwert der mittleren Dichte bei Übergang zu einem auf diesen Punkt zusammenschrumpfenden „unendlich kleinen" Stabteil.

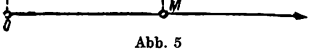


Abb. 5

Die Lage jedes Punktes M auf dem Stab kann durch den Abstand $OM = l$ des Punktes M von einem der Endpunkte O des Stabes charakterisiert werden (Abb. 5). Weiter bezeichne m die Masse des Stückes OM. Dabei ist offenbar m eine Funktion von l,

$$m = f(l), \qquad (4)$$

wobei die Verteilung der Masse auf dem Stab vollständig durch die Funktion (4) beschrieben wird.

Wir wollen nun untersuchen, wie bei gegebener Massenverteilung (4) die Stabdichte in einem gegebenen Punkte M des Stabes berechnet werden kann. Dazu betrachten wir neben dem Punkt M noch einen weiteren Punkt N des Stabes, und zwar sei $ON = l + \Delta l$, wobei $l = OM$ ist.[2]) Dann ist die Masse des Stückes ON gleich

$$m + \Delta m = f(l + \Delta l),$$

so daß die Masse des Stückes MN durch

$$\Delta m = f(l + \Delta l) - f(l)$$

[1]) Wegen der Stetigkeit der Funktion $f(x)$ gilt hiervon auch die Umkehrung, d. h., für $\Delta x \to 0$ strebt auch N gegen M.

[2]) Wir empfehlen dem Leser, sich eine Skizze anzufertigen.

gegeben wird. Daraus ergibt sich

$$\frac{\Delta m}{\Delta l} = \frac{f(l+\Delta l) - f(l)}{\Delta l}$$

als mittlere Dichte dieses Stückes; die wahre Dichte p des Stabes im Punkte M ist der Grenzwert dieses Verhältnisses, wenn N gegen M, d. h. wenn Δl gegen 0 strebt. Wir erhalten also

$$p = \lim_{\Delta l \to 0} \frac{\Delta m}{\Delta l} = \lim_{\Delta l \to 0} \frac{f(l+\Delta l) - f(l)}{\Delta l}, \tag{III}$$

wobei wiederum ein Grenzwert des oben angegebenen Typs zu berechnen ist.

2. Definition der Ableitung. Wir wollen jetzt allgemein Grenzwerte der Art betrachten, wie sie uns in den vorangehenden konkreten Aufgaben begegnet sind, und zwar wollen wir dabei von der konkreten Aufgabestellung abstrahieren und uns nur mit der zugehörigen mathematischen Problemstellung befassen.

Vorgegeben sei eine im offenen Intervall (a, b) definierte und dort stetige Funktion

$$y = f(x).$$

Wir führen dann folgende Operationen aus:

1. Wir betrachten einen beliebigen Punkt $x \in (a, b)$ und bestimmen den zugehörigen Funktionswert $y = f(x)$.

2. Sodann fügen wir zu dem Argumentwert x einen beliebigen von Null verschiedenen Zuwachs $\Delta x\ (\gtrless 0)$ hinzu, welcher jedoch nicht aus dem Intervall (a, b) hinausführen soll, und bestimmen den zu diesem neuen Argumentwert gehörenden Funktionswert $y + \Delta y = f(x + \Delta x)$.

3. Hieraus berechnen wir den dem Zuwachs Δx entsprechenden Zuwachs der Funktion:

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

4. Als nächstes bilden wir das Verhältnis

$$\frac{\Delta y}{\Delta x} = \frac{f(x+\Delta x) - f(x)}{\Delta x}.$$

5. Schließlich lassen wir den Zuwachs Δx gegen Null streben und berechnen den Grenzwert

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}. \tag{5}$$

Dieser Grenzwert wird, falls er existiert[1]), die *Ableitung der Funktion* $f(x)$ *im Punkte* x genannt.

[1]) Wir betrachten hier nur endliche Grenzwerte. Zuweilen führt man auch die „uneigentlichen Grenzwerte" $+\infty$ und $-\infty$ ein und kann dann auch von unendlichen Werten der Ableitung sprechen, was wir jedoch hier nicht tun wollen.

Die Ableitung ist also, kurz gesagt, der Grenzwert des Verhältnisses des Zuwachses der Funktion zum zugehörigen Zuwachs des Argumentes bei gegen Null strebendem Argumentzuwachs.

Die Ableitung der Funktion $y = f(x)$ im Punkte x wird mit $f'(x)$ oder y' bezeichnet. Die erste Bezeichnung ist insofern zweckmäßiger, als in ihr explizit der Punkt x auftritt, in welchem die Ableitung betrachtet wird.

Die Operation, welche die Bestimmung der Ableitung einer gegebenen Funktion zum Ziele hat, bezeichnet man als *Differentiation*. Der Punkt, in dem die Ableitung gebildet wird (d. h. der Wert x, der bei der ersten der oben genannten Operationen zur Auswahl kommt), nennt man den *Differentiationspunkt*.

Zur Illustration der gegebenen Definition mögen die folgenden Beispiele dienen:

Beispiel 1. Gesucht ist die Ableitung der Funktion $y = x^2$ im Punkte $x = 5$. Im betrachteten Fall ist $y = 25$, $y + \Delta y = (5 + \Delta x)^2 = 25 + 10\Delta x + (\Delta x)^2$, also

$$\Delta y = 10\Delta x + (\Delta x)^2 \quad \text{und} \quad \frac{\Delta y}{\Delta x} = 10 + \Delta x.$$

Lassen wir hier Δx gegen Null gehen, so erhalten wir

$$y' = 10.$$

Beispiel 2. Gesucht ist die Ableitung der Funktion $y = x^2$ in den Punkten $x = 4$, $x = 1$, $x = 0$, $x = 12$.

Um nicht für jeden einzelnen Punkt dieselben Überlegungen wiederholen zu müssen, bezeichnen wir den Differentiationspunkt allgemein mit x. Dann ist (die hierbei verwendeten Bezeichnungen verstehen sich von selbst)

$$y = x^2, \quad y + \Delta y = (x + \Delta x)^2 = x^2 + 2x\Delta x + (\Delta x)^2,$$

$$\Delta y = 2x\Delta x + (\Delta x)^2,$$

$$\frac{\Delta y}{\Delta x} = 2x + \Delta x,$$

$$y' = 2x. \tag{6}$$

Wir weisen darauf hin, daß der auf der rechten Seite in (6) auftretende Punkt x gerade der betrachtete Differentiationspunkt ist. Setzen wir in (6) der Reihe nach die speziellen Werte $x = 4$, $x = 1$, $x = 0$, $x = 12$ ein, so erhalten wir $y' = 8$, $y' = 2$, $y' = 0$ bzw. $y' = 24$.

Aus dem betrachteten Beispiel ist bereits ersichtlich, daß die Differentiation einer Funktion in allgemeiner Form, d. h. bei Bezeichnung des Differentiationspunktes durch einen Buchstaben, erhebliche Vorteile bietet.

Beispiel 3. Gesucht ist die Ableitung der Funktion $y = \sqrt{x}$ im Punkte $x\,(> 0)$. In diesem Falle ist

$$y = \sqrt{x}, \quad y + \Delta y = \sqrt{x + \Delta x}, \quad \Delta y = \sqrt{x + \Delta x} - \sqrt{x},$$

$$\frac{\Delta y}{\Delta x} = \frac{\sqrt{x + \Delta x} - \sqrt{x}}{\Delta x} = \frac{1}{\sqrt{x + \Delta x} + \sqrt{x}}, \quad y' = \frac{1}{2\sqrt{x}}.$$

Jede Funktion $y = f(x)$ kann nun in bekannter Weise graphisch dargestellt werden. Nachdem dies geleistet ist, kann man die Frage nach der Gleichung der Tangente in einem gegebenen Punkt der Kurve (d. h. der graphischen Darstellung) aufwerfen. Diese Frage haben wir bereits am Anfang dieses Paragraphen behandelt. Ein Vergleich des dabei erhaltenen Ausdrucks (II) für den Steigungskoeffizienten der Tangente mit der Definition der Ableitung führt auf folgende wichtige Aussage:

Satz. *Die Ableitung $y' = f'(x)$ bedeutet geometrisch den Steigungskoeffizienten der Tangente in demjenigen Punkte der graphischen Darstellung der Funktion, dessen Abszisse der Differentiationspunkt ist.*

Die große physikalische Bedeutung der Ableitung geht bereits deutlich aus dem oben betrachteten Problem A hervor. Dort wurde nämlich gezeigt, daß in dem Fall, in welchem die betrachtete Funktion den Weg beschreibt, den ein sich bewegender Punkt im Laufe der Zeit zurücklegt und wobei die Zeit die Rolle der unabhängigen Veränderlichen spielt, die Ableitung gerade die wahre Geschwindigkeit des Punktes in den einzelnen Zeitpunkten liefert. Ganz allgemein kann man aber auch bei jedem anderen physikalischen Prozeß, bei dem zwei voneinander abhängige veränderliche Größen x und y auftreten, von der Geschwindigkeit der Änderung der einen in bezug auf die andere sprechen. Man sieht leicht ein, daß dann die Ableitung y' gerade die Geschwindigkeit der Änderung von y bezüglich x charakterisiert. Es erübrigt sich, auf die Bedeutung hinzuweisen, die das Studium derartiger Änderungsgeschwindigkeiten in der Physik besitzt. Ist etwa x die Zeit und y die im Verlaufe der Zeit x durch den Querschnitt eines Drahtes fließende Elektrizitätsmenge, so ist y' nichts anderes als die Stromstärke. Es bereitet keine Schwierigkeiten, weitere Beispiele dieser Art zu finden.

3. Differenzierbarkeit und Stetigkeit. Einseitige Ableitungen. Die Ableitung einer Funktion in einem gegebenen Punkt wurde oben als Grenzwert einer gewissen veränderlichen Größe definiert. Nun strebt jedoch nicht jede veränderliche Größe stets gegen einen bestimmten Grenzwert. Aus diesem Grunde ist auch nicht zu erwarten, daß jede Funktion in jedem Punkte, in welchem sie definiert ist, eine Ableitung besitzt. Es ist vielmehr so, daß in jedem Punkt, in welchem die Ableitung vorhanden ist (der Grenzwert (5) also existiert), die Funktion notwendigerweise stetig ist, d. h., es gilt der

Satz. *Besitzt die Funktion $y = f(x)$ im Punkt x eine Ableitung,*[1]) *so ist die Funktion in diesem Punkt auch stetig.*

Beweis. Ist $\Delta x \neq 0$ ein gewisser Zuwachs des Argumentwertes x und Δy der entsprechende Zuwachs des Funktionswertes y, so ist

$$\Delta y = \frac{\Delta y}{\Delta x} \Delta x.$$

[1]) Man sagt hierfür auch, die Funktion $y = f(x)$ sei im Punkte x *differenzierbar*.

Für $\Delta x \to 0$ strebt nun das Verhältnis $\frac{\Delta y}{\Delta x}$ gegen den endlichen[1]) Grenzwert $y' = f'(x)$, so daß

$$\lim_{\Delta x \to 0} \Delta y = 0$$

ist; dies bedeutet offenbar gerade, daß die Funktion $f(x)$ im Punkte x (d. h. in dem Punkte, zu welchem der Zuwachs Δx hinzugefügt wurde) stetig ist.

Es ergibt sich nun naturgemäß die Frage, ob nicht vielleicht auch die Stetigkeit einer Funktion in einem Punkte für die Existenz der Ableitung in diesem Punkte (d. h. für die Differenzierbarkeit in diesem Punkte) hinreichend ist. Diese Frage muß nun durchaus verneint werden. Zum Beispiel besitzt die überall stetige Funktion $y = |x|$ im Punkte $x = 0$ keine Ableitung; denn es ist

$$\frac{\Delta y}{\Delta x} = \pm 1,$$

je nachdem, ob $\Delta x > 0$ oder $\Delta x < 0$ ist. Daher existiert im vorliegenden Fall der Grenzwert

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$$

nicht. Dies ist auch geometrisch vollkommen klar; denn der die Funktion $y = |x|$ darstellende Streckenzug (Abb. 6) besitzt im Koordinatenursprung keine Tangente.

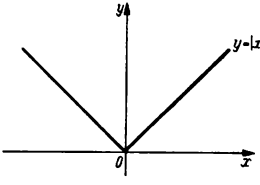


Abb. 6

Das betrachtete Beispiel ist insofern recht einfach, als bei der Funktion $y = |x|$ im Punkt $x = 0$ immer noch die sogenannte rechtsseitige und linksseitige Ableitung existieren. Dabei versteht man unter der *rechtsseitigen Ableitung* einer Funktion $f(x)$ im Punkte x den Grenzwert von

$$\frac{f(x + \Delta x) - f(x)}{\Delta x}$$

für $\Delta x \to 0$, wobei jedoch für Δx nur positive Werte zugelassen sind. Bezeichnen wir die rechtsseitige Ableitung im Punkte x mit $f'_+(x)$, so ist also

$$f'_+(x) = \lim_{\substack{\Delta x \to 0 \\ \Delta x > 0}} \frac{f(x + \Delta x) - f(x)}{\Delta x}.$$

Entsprechend wird die *linksseitige Ableitung* $f'_-(x)$ im Punkte x durch

$$f'_-(x) = \lim_{\substack{\Delta x \to 0 \\ \Delta x < 0}} \frac{f(x + \Delta x) - f(x)}{\Delta x}.$$

definiert.

[1]) Vgl. die Fußnote auf Seite 291.

Offenbar ist die Funktion[1] $f(x)$ *dann und nur dann im Punkte* x *differenzierbar, wenn sie im Punkte* x *sowohl eine rechtsseitige als auch eine linksseitige Ableitung besitzt und wenn diese beiden Ableitungen übereinstimmen.*

Für die Funktion $f(x) = |x|$ ist $f'_+(0) = +1$, $f'_-(0) = -1$, so daß sich auch hieraus das Fehlen der gewöhnlichen Ableitung $f'(0)$ erklärt.

Ein interessantes Beispiel für eine überall stetige Funktion, für welche weder $f'_+(0)$ noch $f'_-(0)$ existiert, ist die Funktion

$$f(x) = \begin{cases} x \sin \frac{1}{x} & (x \neq 0), \\ 0 & (x = 0). \end{cases}$$

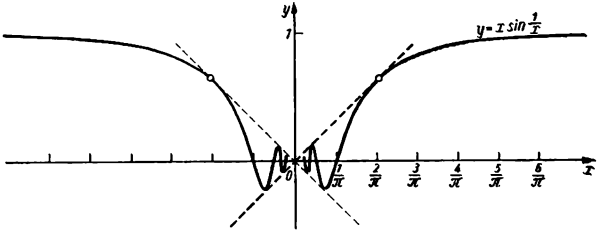


Abb. 7

Ist nämlich z. B. $\Delta x > 0$, so ist

$$\frac{f(0+\Delta x) - f(0)}{\Delta x} = \sin \frac{1}{\Delta x},$$

und dies strebt für $\Delta x \to 0$ gegen keinen bestimmten Grenzwert, sondern schwankt vielmehr unendlich oft zwischen $+1$ und -1, wobei diese Werte und jeweils auch alle dazwischenliegenden Werte angenommen werden. Man kann dies auch unmittelbar der Abb. 7 entnehmen, wo ein Teil des graphischen Bildes der Funktion $f(x)$ dargestellt ist; denn die durch den Koordinatenursprung und irgendeinen Punkt $N(x, y)$ des graphischen Bildes der Funktion $f(x)$ gehende Sekante führt bei Annäherung des Punktes N an den Koordinatenursprung unendlich viele Schwingungen zwischen den Geraden $y = +x$ und $y = -x$ aus, so daß sie gegen keine Grenzlage strebt.

In den betrachteten Beispielen fehlt die Ableitung nur in einem einzigen Punkt. Derartige Beispiele waren bereits im 18. Jahrhundert bekannt. Damals

[1] Hierbei setzen wir — wie oben — voraus, daß die Funktion $f(x)$ in einem den Punkt x enthaltenden offenen Intervall (a, b) erklärt ist. Für eine in einem abgeschlossenen Intervall $[a, b]$ erklärte Funktion hat man häufig auch die Ableitungen $f'_+(a)$ und $f'_-(b)$ zu betrachten. In einem derartigen Fall hebt man meistens die Einseitigkeit der Ableitungen nicht hervor, schreibt also einfach $f'(a)$ und $f'(b)$.

herrschte jedoch (wenn auch nicht klar ausgesprochen) die irrige Auffassung, daß jede stetige Funktion (auch dieser Begriff war, nebenbei gesagt, noch nicht genau präzisiert) überall mit Ausnahme von höchstens endlich vielen Punkten eine Ableitung besitze.

Erst im Jahre 1834 unterschied der große russische Mathematiker N. I. Lobatschewski deutlich zwischen der Stetigkeit und der Differenzierbarkeit einer Funktion. Noch in der Mitte des 19. Jahrhunderts versuchte man jedoch zu beweisen, daß eine stetige Funktion bis auf endlich viele Punkte überall eine Ableitung besitze. Diesen Versuchen bereitete im Jahre 1871 K. Weierstraß dadurch ein Ende, daß er eine Funktion konstruierte, welche überall stetig ist aber nirgends eine Ableitung besitzt. Das erste Beispiel für eine derartige Funktion wurde allerdings bereits wesentlich früher (nicht nach 1830) von dem bedeutenden tschechischen Mathematiker B. Bolzano angegeben. Dieses Beispiel blieb jedoch lange Zeit unbekannt und wurde erst in den zwanziger Jahren dieses Jahrhunderts veröffentlicht.

4. Die Ableitungen der einfachsten elementaren Funktionen. In diesem und den nächsten beiden Abschnitten wollen wir zu folgender Feststellung gelangen: *Jede elementare Funktion besitzt in allen Punkten ihres Definitionsbereiches, endlich viele Punkte evtl. ausgenommen, eine Ableitung, wobei die Ableitung ihrerseits wiederum eine elementare Funktion des Differentiationspunktes ist.*

Ziel des vorliegenden Abschnitts ist der Beweis folgender Formeln[1]):

1. $(C)' = 0$,
2. $(x)' = 1$,
3. $(x^n)' = n x^{n-1}$,
4. $(x^a)' = a x^{a-1}$,
5. $(\sin x)' = \cos x$,
6. $(\cos x)' = -\sin x$,
7. $(\operatorname{tg} x)' = \dfrac{1}{\cos^2 x}$,
8. $(\operatorname{ctg} x)' = -\dfrac{1}{\sin^2 x}$,
9. $(a^x)' = a^x \ln a$,
10. $(e^x)' = e^x$,
11. $(\ln x)' = \dfrac{1}{x}$,
12. $(\log_a x)' = \dfrac{1}{x \ln a}$.

Der Buchstabe „x" bezeichnet überall jeweils den Differentiationspunkt. In Formel 3. bedeutet n eine beliebige natürliche Zahl, während der in 4. auftretende Buchstabe „a" eine beliebige reelle Zahl bezeichnet. Dies hat zur Folge, daß die in 3. betrachtete Funktion x^n auf der ganzen x-Achse definiert ist, während die in 4. betrachtete Funktion x^a unter Umständen nicht überall erklärt ist. Ist nämlich a eine irrationale Zahl, so ist die Funktion nur für $x > 0$ definiert (falls $a > 0$ ist, auch noch für $x = 0$); ist dagegen a eine von Null verschiedene rationale Zahl, und zwar $a = \frac{p}{q}$, wobei p und q teilerfremd sind und

[1]) Auf die Funktionen sec x und cosec x soll hier nicht eingegangen werden, da sie im allgemeinen kaum gebraucht werden. Auf die zyklometrischen Funktionen kommen wir in Abschnitt 5 zu sprechen.

$q > 1$ ist, so ist bei ungeradem q die Funktion x^a sowohl für $x > 0$ als auch für $x < 0$ (und im Fall $p > 0$ auch noch für $x = 0$) erklärt, während sie bei geradem q nur für $x > 0$ (und im Fall $p > 0$ auch für $x = 0$) definiert ist. Ähnlich verlieren auch einige andere der oben aufgeführt Funktionen für gewisse Argumentwerte ihren Sinn, so etwa $\operatorname{tg} x$ für $x = \frac{\pi}{2} + k\pi (k = 0, \pm 1, \pm 2, \ldots)$, $\ln x$ für $x \leqq 0$ usw. Aus den nachfolgenden Beweisen ist jedoch ersichtlich, daß die Formeln 1. bis 12. für alle diejenigen Punkte x gelten, in denen sowohl die rechte als auch die linke Seite der entsprechenden Formel einen Sinn besitzt.

Für die Beweise benötigen wir die folgenden Formeln aus der Theorie der Grenzwerte:

$$\lim_{z \to 0} \frac{\ln(1+z)}{z} = 1, \tag{A}$$

$$\lim_{z \to 0} \frac{(1+z)^a - 1}{z} = a, \tag{B}$$

$$\lim_{z \to 0} \frac{a^z - 1}{z} = \ln a \quad (a > 0), \tag{C}$$

$$\lim_{z \to 0} \frac{\sin z}{z} = 1. \tag{D}$$

Die Formel (D) wurde bereits im vorangehenden Artikel (vgl. S. 181) angegeben. Wir brauchen also nur noch die Formeln (A) bis (C) zu beweisen.

Bekanntlich (vgl. S. 189) ist die Zahl e der Grenzwert von

$$\left(1 + \frac{1}{x}\right)^x$$

für $x \to +\infty$. Setzen wir hier

$$x = \frac{1}{z},$$

so ist also

$$\lim_{z \to 0} (1+z)^{\frac{1}{z}} = e.$$

Hieraus ergibt sich auf Grund der Stetigkeit des Logarithmus:

$$\lim_{z \to 0} \frac{\ln(1+z)}{z} = \lim_{z \to 0} \ln (1+z)^{\frac{1}{z}} = \ln e = 1,$$

womit Formel (A) bereits bewiesen ist.

Zum Beweis von (B) setzen wir

$$(1+z)^a - 1 = u \quad (a \neq 0).$$

Für $z \to 0$ gilt offenbar $u \to 0$. Ferner ist wegen

$$(1+z)^a = 1 + u$$

natürlich

$$a \ln (1+z) = \ln (1+u)$$

und daher

$$\frac{(1+z)^a-1}{z}=\frac{u}{z}=\frac{u}{\ln(1+u)}\cdot\frac{a\ln(1+z)}{z}.$$

Hieraus und aus (A) folgt jedoch sofort die Beziehung (B), die auch für $a=0$ richtig ist.

Setzen wir schließlich

$$a^z-1=u \quad (a>0),$$

so gilt für $z\to 0$ offenbar $u\to 0$. Hierbei ist aber

$$z=\frac{\ln(1+u)}{\ln a}$$

und daher

$$\frac{a^z-1}{z}=\frac{u}{\ln(1+u)}\ln a.$$

Daraus und aus (A) folgt unmittelbar (C).

Wir wenden uns nun dem Beweis der Formeln 1. bis 12. zu.

Die Formeln 1. und 2. sind offenbar trivial. Ist nämlich $y=C$, d. h. konstant, so ist der zu einem beliebigen Zuwachs Δx des Arguments gehörende Zuwachs Δy der Funktion gleich Null, so daß $(C)'=0$. Ist $y=x$, so ist stets $\Delta y=\Delta x$ und daher $(x)'=1$.

Nun zum Beweis von Formel 3. Für $y=x^n$ ist

$$y+\Delta y=(x+\Delta x)^n, \quad \Delta y=(x+\Delta x)^n-x^n$$

und daher

$$\frac{\Delta y}{\Delta x}=\frac{(x+\Delta x)^n-x^n}{(x+\Delta x)-x}=(x+\Delta x)^{n-1}+(x+\Delta x)^{n-2}x+\cdots+x^{n-1}.$$

Lassen wir hier Δx gegen Null gehen, so ergibt sich

$$\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=\underbrace{x^{n-1}+x^{n-1}+\cdots+x^{n-1}}_{n\,\text{mal}}=nx^{n-1},$$

womit Formel 3. bewiesen ist.

Zum Beweis von Formel 4. betrachten wir zunächst einen von Null verschiedenen Argumentwert x, für welchen die Funktion x^a erklärt ist. Dann ist sie auch für alle in hinreichender Nähe liegenden Argumentwerte $x+\Delta x$ erklärt, und es gilt

$$y+\Delta y=(x+\Delta x)^a=x^a\left(1+\frac{\Delta x}{x}\right)^a,$$

so daß also

$$y'=\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=\lim_{\Delta x\to 0}x^a\cdot\frac{\left(1+\frac{\Delta x}{x}\right)^a-1}{\Delta x}=x^{a-1}\lim_{\Delta x\to 0}\frac{\left(1+\frac{\Delta x}{x}\right)^a-1}{\frac{\Delta x}{x}}.$$

Setzen wir hier $\frac{\Delta x}{x}=z$, so ist 4. eine unmittelbare Folge aus (B).

Beim bisherigen Beweis war der Fall $x = 0$ ausdrücklich ausgeschlossen. Hierfür ist also eine besondere Überlegung notwendig. Gemäß der oben getroffenen Vereinbarung müssen wir voraussetzen, daß für $x = 0$ beide Seiten von 4. einen Sinn haben, daß also $a > 1$ ist. Dann ist aber für $x = 0$ auch $y = x^a = 0$ und $y + \Delta y = (\Delta x)^a$. Folglich ist also

$$\frac{\Delta y}{\Delta x} = (\Delta x)^{a-1}$$

und daher $y' = 0$, in Übereinstimmung mit 4. für $a > 1$ und $x = 0$ (wir weisen darauf hin, daß für $a = \frac{2m-1}{2n}$, wobei m und n natürliche Zahlen sind, die Funktion x^a nur für $x \geqq 0$ definiert ist, so daß uns 4. für solche a im Fall $x = 0$ lediglich die rechtsseitige Ableitung liefert).

Die Beweise für 5. und 6. laufen vollkommen analog, so daß wir uns hier auf einen Beweis für 5. beschränken.

Es sei also

$$y = \sin x, \quad y + \Delta y = \sin(x + \Delta x).$$

Dann ist

$$\Delta y = \sin(x + \Delta x) - \sin x = 2 \sin \frac{\Delta x}{2} \cos\left(x + \frac{\Delta x}{2}\right)$$

und

$$y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \left(\frac{\sin \frac{\Delta x}{2}}{\frac{\Delta x}{2}} \right) \cdot \lim_{\Delta x \to 0} \cos\left(x + \frac{\Delta x}{2}\right).$$

Der erste der Faktoren auf der linken Seite ist wegen (D) gleich 1, während der zweite (wegen der Stetigkeit der Funktion $\cos x$) gleich $\cos x$ ist. Also ist $y' = \cos x$, womit Formel 5. bewiesen ist.

Die Formeln 7. und 8. lassen sich ebenfalls analog beweisen, so daß wir uns auf einen Beweis für 7. beschränken. Es sei dazu $x \neq (2n-1)\frac{\pi}{2}$ und

$$y = \operatorname{tg} x, \quad y + \Delta y = \operatorname{tg}(x + \Delta x).$$

Dann ist[1])

$$\Delta y = \frac{\sin \Delta x}{\cos x \cdot \cos(x + \Delta x)},$$

woraus alles übrige trivial folgt.

Zum Beweis von 9. sei

$$y = a^x, \quad y + \Delta y = a^{x + \Delta x},$$

so daß

$$\Delta y = a^x (a^{\Delta x} - 1),$$

woraus auf Grund von (C) bereits die Behauptung folgt.

---

[1]) Bekanntlich ist

$$\operatorname{tg} A - \operatorname{tg} B = \frac{\sin(A - B)}{\cos A \cos B}.$$

Formel 10. ist sodann ein Spezialfall von 9.

Zum Beweis von 11. sei $x > 0$. Dann ist für

$$y = \ln x, \quad y + \Delta y = \ln(x + \Delta x)$$

offenbar

$$\Delta y = \ln(x + \Delta x) - \ln x = \ln\left(1 + \frac{\Delta x}{x}\right),$$

und wir erhalten auf Grund von (A)

$$y' = \lim_{\Delta x \to 0} \frac{\ln\left(1 + \frac{\Delta x}{x}\right)}{\Delta x} = \frac{1}{x} \lim_{\Delta x \to 0} \frac{\ln\left(1 + \frac{\Delta x}{x}\right)}{\frac{\Delta x}{x}} = \frac{1}{x}.$$

Ist schließlich $y = \log_a x$, so ist $a^y = x$, also $y = \frac{\ln x}{\ln a}$. Wiederholt man den soeben für $\ln x$ geführten Beweis für $\frac{\ln x}{\ln a}$, so erhält man genau Formel 12.

Zum Abschluß dieser Überlegungen wollen wir auf die Frage eingehen, warum in den Formeln der Analysis in überwiegendem Maße nicht der dekadische, sondern der natürliche Logarithmus auftritt. Auf den ersten Blick erscheint diese Tatsache verwunderlich; denn der dekadische Logarithmus besitzt gegenüber dem natürlichen Logarithmus unbestreitbare Vorteile, die ihn zu einem unentbehrlichen Hilfsmittel für das numerische Rechnen machen. Diese Vorteile bestehen in folgendem: Auf Grund der Definition des dekadischen Logarithmus ist

$$\lg 1 = 0, \quad \lg 10 = 1, \quad \lg 100 = 2, \quad \lg 1000 = 3, \ \ldots .$$

Soll nun der Logarithmus einer Zahl bestimmt werden, welche keine ganzzahlige Potenz von 10 ist, etwa von der Zahl 637, so verfährt man folgendermaßen: Die Zahl 637 liegt zwischen 100 und 1000. Also ist auf Grund der Monotonie des Logarithmus

$$\lg 100 < \lg 637 < \lg 1000$$

oder

$$2 < \lg 637 < 3,$$

d. h.

$$\lg 637 = 2, \ \ldots .$$

Dies zeigt, daß sich der ganzzahlige Anteil des Logarithmus (seine „Charakteristik" oder „Kennziffer") stets auf den ersten Blick erkennen läßt, so daß es nur noch notwendig ist, den gebrochenen Anteil (die „Mantisse") zu bestimmen.

Weiter hängen die Zahlen 637 und 63,7 durch die Beziehung

$$637 = 63{,}7 \cdot 10$$

zusammen, so daß

$$\lg 637 = \lg 63{,}7 + \lg 10 = \lg 63{,}7 + 1.$$

Ein weiterer Vorteil der dekadischen Logarithmen besteht also darin, daß die Mantisse gegenüber der Multiplikation der betrachteten Zahl mit Zehn oder einer Zehnerpotenz unabhängig ist.

Man sieht nun leicht ein, daß der Grund für diese rechnerischen Vorzüge des dekadischen Logarithmus in der Verwendung des Dezimalsystems zu suchen ist. Würde man an Stelle des Dezimalsystems z. B. ein Zwölfersystem verwenden, so würde nicht mehr der dekadische Logarithmus, sondern der Logarithmus zur Basis 12 die genannten Vorzüge besitzen. Die Verwendung der Zahl 10 als Basis unseres Zahlensystems erklärt sich aber nicht durch irgendwelche besonderen arithmetischen Eigenschaften dieser Zahl, sondern einfach durch die physiologische Tatsache, daß sich an den Händen des Menschen 10 Finger befinden. Beim Abzählen von Gegenständen durch Einbiegen der Finger war der Mensch einer frühen Kulturstufe gezwungen, wenn er bei 10 angekommen war, sich irgendeinen Vermerk zu machen, etwa eine Kerbe in einen Baumstamm oder dgl. Dieser Vermerk stellte dann die „Recheneinheit zweiter Ordnung" dar, und wir sehen, daß diese gerade gleich 10 ist.[1])

Die vorangehenden Ausführungen zeigen, daß die Verwendung der dekadischen Logarithmen durch keinerlei theoretische Erwägungen bedingt ist und daher auch keinen theoretischen Nutzen bringen kann. Das bedeutet natürlich in keiner Hinsicht eine Herabsetzung der rechnerischen Bequemlichkeiten, die die Verwendung der dekadischen Logarithmen mit sich bringt, so daß er bei allen numerischen Rechnungen benutzt wird. Nun spielen aber in der Mathematik nicht die numerischen, sondern die „buchstabenmäßigen" Rechnungen die beherrschende Rolle, und bei derartigen Rechnungen werden natürlich die Vorteile des dekadischen Logarithmus hinfällig. Gleichzeitig damit ergibt sich die Frage, bei welcher Basis der Logarithmen die allgemeinen Formeln möglichst einfach werden. Hierbei zeigt sich z. B., daß die natürlichen Logarithmen am geeignetsten sind, denn die Formel für die Ableitung des natürlichen Logarithmus,

$$(\ln x)' = \frac{1}{x},$$

ist nun einmal einfacher als die für den Logarithmus zu jeder anderen Basis,

$$(\log_a x)' = \frac{1}{x \ln a}.$$

Da insbesondere $\ln 10 = 2{,}30259\ldots$ ist, wird die Ableitung des dekadischen Logarithmus durch

$$(\lg x)' = \frac{1}{x \cdot 2{,}30259\ldots} = \frac{0{,}43429\ldots}{x}$$

gegeben.

5. Differentiation der Umkehrfunktionen. Es sei $y = f(x)$ eine im abgeschlossenen Intervall $[a, b]$ definierte, stetige und echt monotone Funktion.

[1]) Vgl. hierzu den Artikel von I. G. Baschmakowa und A. P. Juschkewitsch in EdEM Bd. I. — *Anm. d. wissenschaftl. Red.*

Dann existiert (vgl. S. 232) die zu $f(x)$ gehörende Umkehrfunktion $x = g(y)$, welche in dem abgeschlossenen Intervall zwischen $f(a)$ und $f(b)$ definiert ist und jedem y aus diesem Intervall den eindeutig bestimmten x-Wert aus dem Intervall $[a, b]$ zuordnet, für den $f(x) = y$ ist. Die Umkehrfunktion ist ebenfalls stetig und monoton.[1])

Unter diesen Voraussetzungen (und mit diesen Bezeichnungen) gilt:

Satz: *Existiert in einem Punkte x_0 die Ableitung $f'(x_0)$ und ist $f'(x_0) \neq 0$, so ist auch die Umkehrfunktion $g(y)$ im entsprechenden Punkte $y_0 = f(x_0)$ differenzierbar, und es ist*

$$g'(y_0) = \frac{1}{f'(x_0)}.$$

Beweis. Es sei $\Delta y \neq 0$, und es liege $y_0 + \Delta y$ zwischen $f(a)$ und $f(b)$. Dann ist, falls

$$g(y_0 + \Delta y) = x_0 + \Delta x$$

gesetzt wird, $\Delta x \neq 0$, $x_0 + \Delta x \in [a, b]$ und

$$f(x_0 + \Delta x) = y_0 + \Delta y.$$

Offenbar ist dann

$$\frac{g(y_0 + \Delta y) - g(y_0)}{\Delta y} = \frac{\Delta x}{f(x_0 + \Delta x) - f(x_0)}.$$

Weil nun die Funktion $x = g(y)$ stetig ist, strebt mit Δx auch Δy gegen Null. Daraus folgt, daß die linke Seite der zuletzt angegebenen Gleichung für $\Delta y \to 0$ gegen $\frac{1}{f'(x_0)}$ strebt, womit der angegebene Satz bewiesen ist.

Eine kurze aber prägnante Schreibweise für diesen Satz ist

$$y'_x = \frac{1}{x'_y}.$$

Als ein Beispiel für eine Anwendung dieses Satzes beweisen wir

$$(\arcsin x)' = \frac{1}{\sqrt{1 - x^2}} \qquad (-1 < x < 1).$$

Dazu betrachten wir die Funktion

$$y = \sin x,$$

wobei wir x nur in dem abgeschlossenen Intervall $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ variieren lassen. In diesem Intervall ist die Funktion $y = \sin x$ echt monoton wachsend, und

---

[1]) Ist die ursprüngliche Funktion $y = f(x)$ in einem offenen Intervall (a, b) erklärt, so ist auch die Umkehrfunktion in einem offenen, und zwar in dem offenen Intervall zwischen $f(a)$ und $f(b)$ definiert. Dabei können dann eines oder beide der erwähnten Intervalle unendlich sein $\left(\text{z. B. ist für } y = \operatorname{tg} x, -\frac{\pi}{2} < x < \frac{\pi}{2}, \text{ die Umkehrfunktion im unendlichen Intervall } (-\infty, +\infty) \text{ erklärt}\right)$.

zwar wächst sie dort von -1 bis $+1$. Außerdem ist die Funktion $y = \sin x$ im betrachteten Intervall stetig. Folglich besitzt sie dort eine Umkehrfunktion

$$x = \arcsin y,$$

welche im Intervall $[-1, +1]$ definiert ist und welche jedem $y \in [-1, +1]$ den eindeutig bestimmten Wert $x \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ zuordnet, für welchen $\sin x = y$ ist.

Beachten wir noch, daß die Funktion $y' = \cos x$ überall außer in den Punkten $x = \pm\frac{\pi}{2}$ von Null verschieden ist, so können wir schließen, daß die Umkehrfunktion $x = \arcsin y$ überall außer für $y = \pm 1$ eine Ableitung besitzt und daß

$$(\arcsin y)' = \frac{1}{\cos x},$$

wobei $x = \arcsin y$. Nun ist aber $\cos^2 x + \sin^2 x = 1$ und daher

$$\cos x = \pm\sqrt{1 - \sin^2 x} = \pm\sqrt{1 - y^2}.$$

Von den beiden Vorzeichen kommt im betrachteten Fall nur das Pluszeichen in Frage, da $-\frac{\pi}{2} < x < \frac{\pi}{2}$ und für diese Werte $\cos x > 0$ ist. Also ist

$$(\arcsin y)' = \frac{1}{\sqrt{1 - y^2}}.$$

Diese Formel unterscheidet sich jedoch von der zu beweisenden nur in der Bezeichnung des Differentiationspunktes.

Ganz entsprechend beweist man die Formeln

$$(\arccos x)' = \frac{-1}{\sqrt{1 - x^2}} \qquad (-1 < x < +1),$$

$$(\operatorname{arctg} x)' = \frac{1}{1 + x^2} \qquad (-\infty < x < +\infty),$$

$$(\operatorname{arcctg} x)' = \frac{-1}{1 + x^2} \qquad (-\infty < x < +\infty).$$

6. Formeln für die Differentiation von zusammengesetzten Funktionen. In den beiden vorangehenden Abschnitten haben wir gezeigt, daß die einfachsten elementaren Funktionen differenzierbar sind, wobei wir zugleich die Ausdrücke für die Ableitungen angegeben haben. In den Anwendungen hat man es jedoch in den meisten Fällen nicht nur mit den bisher betrachteten einfachsten Funktionen, sondern auch mit gewissen Kombinationen dieser Funktionen zu tun, insbesondere mit Funktionen, die man durch Anwendung arithmetischer Operationen aus diesen Funktionen erhält. Wir wollen uns daher jetzt der Frage nach der Differenzierbarkeit derartiger und ähnlicher Kombinationen der betrachteten Funktionen, insbesondere dem Problem der Differentiation von Summe, Differenz, Produkt und Quotient zuwenden.

Offenbar haben wir es hierbei mit einem Spezialfall der folgenden allgemeineren Aufgabe zu tun: Gegeben sind zwei (nicht unbedingt elementare) Funktionen $u = u(x)$ und $v = v(x)$, die beide in demselben Intervall (a, b) definiert und in einem gewissen Punkte x dieses Intervalls differenzierbar sind und dort die Ableitungen u' und v' besitzen. Gefragt ist dann nach der Differenzierbarkeit der Funktionen[1])

$$u+v, \quad u-v, \quad uv, \quad \frac{u}{v} \tag{7}$$

im Punkte x und wie sich gegebenenfalls die Ableitungen dieser Funktionen durch u' und v' ausdrücken lassen.

Diese Aufgabe läßt sich in allen vier Fällen durch fast dieselbe Überlegung lösen, so daß wir uns hier auf den Fall des Produktes uv beschränken wollen.

Es seien u bzw. v die Werte der Funktionen $u(x)$ bzw. $v(x)$ im betrachteten Differentiationspunkt x. Fügen wir zu x einen Zuwachs Δx ($\Delta x \neq 0$ und nicht aus (a, b) hinausführend) hinzu, so ändern wir damit die Werte von $u(x)$ und $v(x)$, und zwar seien $u + \Delta u$ und $v + \Delta v$ die neuen Werte. Der neue Wert des Produktes $y = u(x)v(x)$ wird dann

$$y + \Delta y = (u + \Delta u)(v + \Delta v).$$

Subtrahieren wir hiervon $y = uv$ und dividieren wir anschließend durch Δx, so erhalten wir

$$\frac{\Delta y}{\Delta x} = \frac{\Delta u}{\Delta x} v + u \frac{\Delta v}{\Delta x} + \frac{\Delta u}{\Delta x} \cdot \Delta v.$$

Lassen wir nun hier Δx gegen Null gehen, so gilt nach Definition der Ableitung:

$$\lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x} = u', \quad \lim_{\Delta x \to 0} \frac{\Delta v}{\Delta x} = v'.$$

Außerdem gilt aber auch, und das ist hier sehr wichtig,

$$\lim_{\Delta x \to 0} \Delta v = 0.$$

Da nämlich die Funktion $v(x)$ im Punkte x differenzierbar sein sollte, ist sie dort auch stetig; es strebt also mit dem Argumentzuwachs auch der entsprechende Funktionszuwachs gegen Null.

Fassen wir dies alles zusammen, so finden wir

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = u'v + uv'$$

oder, was dasselbe bedeutet,

$$(uv)' = u'v + uv'.$$

---

[1]) Im Fall des Quotienten $\frac{u}{v}$ muß natürlich vorausgesetzt werden, daß im betrachteten Differentiationspunkt x der Nenner $v(x)$ von Null verschieden ist.

Analog ergibt sich

$$(u+v)'=u'+v', \quad (u-v)'=u'-v', \quad \left(\frac{u}{v}\right)'=\frac{u'v-uv'}{v^2}.$$

Damit ist zugleich gezeigt, daß *mit den Funktionen $u(x)$ und $v(x)$ auch die Funktionen* (7) *im Punkte x differenzierbar sind.*

Ist beim Produkt speziell einer der Faktoren eine Konstante, so nimmt die obige Formel die Gestalt

$$(cu)'=cu'$$

an, d. h., eine Konstante kann beim Differenzieren „herausgezogen" werden.

Mit Hilfe der angegebenen Differentiationsregeln können wir jetzt bereits die verschiedensten Kombinationen der vorher betrachteten elementaren Funktionen differenzieren, z. B.[1])

$$y=\frac{\sin x}{\sqrt{x}}e^x,$$

$$y=(2x^3+5x^2-7x+2)\frac{\operatorname{arctg} x}{\ln x}$$

oder dergleichen.

Jedoch sind wir z. B. noch nicht in der Lage, die Ableitung der Funktion

$$y=e^{\sin x}$$

zu berechnen, da sie nicht das Resultat von arithmetischen Operationen über den einfachsten elementaren Funktionen ist. Hierfür benötigen wir noch eine weitere Differentiationsregel (die wohl im betrachteten Zusammenhang die wichtigste Regel ist), nämlich die sogenannte Kettenregel (Differentiationsregel für mittelbare Funktionen).

Ihr liegt folgende Fragestellung zugrunde: Vorgegeben sei eine im Intervall (A, B) definierte Funktion

$$y=f(z),$$

welche in einem gewissen Punkt z_0 dieses Intervalls differenzierbar ist und dort die Ableitung

$$y'_z=f'(z_0)$$

besitzt. Vorgegeben sei ferner eine zweite in einem Intervall (a, b) definierte Funktion

$$z=\varphi(x),$$

deren Funktionswerte der Ungleichung

$$A<\varphi(x)<B$$

genügen, so daß wir die sogenannte *mittelbare Funktion*

$$f(\varphi(x))$$

[1]) Vgl. S. 308.

bilden können, und zwar sei dem Punkt $x_0 \in (a, b)$ durch die Funktion $\varphi(x)$ gerade der oben betrachtete Punkt z_0 zugeordnet, d. h., es sei $\varphi(x_0) = z_0$. Wir wollen schließlich annehmen, daß die Funktion $\varphi(x)$ im Punkte x_0 eine Ableitung besitzt, und zwar sei $\varphi'(x_0) = z_x'$. Was kann man in diesem Fall über die Differenzierbarkeit der Funktion $f(\varphi(x))$ im Punkte x_0 aussagen, und wie läßt sich gegebenenfalls die Ableitung dieser Funktion im Punkte x_0 durch y_z' und z_x' ausdrücken?

Zur Beantwortung dieser Frage benötigen wir eine Hilfsformel, die übrigens auch an und für sich ein gewisses Interesse beanspruchen kann. Es sei dazu $f(x)$ eine Funktion, welche in einem gewissen Punkt x ihres Definitionsbereiches die Ableitung $y_x' = f'(x)$ besitzt. Es sei ferner Δx ein von Null verschiedener Argumentzuwachs, der nicht aus dem Definitionsbereich der Funktion herausführt, und Δy der entsprechende Zuwachs der Funktion. Da dann $\frac{\Delta y}{\Delta x}$ für $\Delta x \to 0$ gegen die Ableitung y_x' strebt, wird also die Differenz

$$\alpha = \frac{\Delta y}{\Delta x} - y_x' \tag{8}$$

zwischen $\frac{\Delta y}{\Delta x}$ und y_x' für $\Delta x \to 0$ beliebig klein. Wir schreiben nun (8) in der Form

$$\Delta y = y_x' \Delta x + \alpha \Delta x. \tag{9}$$

Diese neue Formel ist, wie auch (8), zunächst nur für $\Delta x \neq 0$ sinnvoll; denn die Größe α, die im wesentlichen eine Funktion von Δx ist, $\alpha = \alpha(\Delta x)$, ist für $\Delta x = 0$ überhaupt nicht erklärt. Da aber für $\Delta x = 0$ offenbar auch $\Delta y = f(x + \Delta x) - f(x) = 0$ ist, kann man für $\Delta x = 0$ die Funktion $\alpha(\Delta x)$ ganz beliebig festsetzen, wobei dann (9) auch noch für $\Delta x = 0$ richtig ist.

Wir setzen nun

$$\alpha(\Delta x) = \begin{cases} \dfrac{\Delta y}{\Delta x} - y_x' & \text{für} \quad \Delta x \neq 0, \\ 0 & \text{für} \quad \Delta x = 0. \end{cases} \tag{10}$$

Dann ist, wie bereits gesagt, die Formel (9) ausnahmslos gültig. Außerdem ist jetzt aber für jede gegen Null strebende Folge $\{h_n\}$

$$\lim \alpha(h_n) = 0,$$

und zwar unabhängig davon, ob alle Glieder der Folge $\{h_n\}$ von Null verschieden sind oder nicht.

Mit dieser Erkenntnis wenden wir uns jetzt wieder der Frage nach der Differenzierbarkeit der mittelbaren Funktion $y = f(\varphi(x))$ zu.

Unter Beibehaltung der oben benutzten Bezeichnungen geben wir dem Argument x zunächst den Wert x_0 und danach den Wert $x_0 + \Delta x$, wobei $\Delta x \neq 0$ und $x_0 + \Delta x \in (a, b)$. Dabei nimmt $z = \varphi(x)$ die Werte $z_0 = \varphi(x_0)$ und $z_0 + \Delta z = \varphi(x_0 + \Delta x)$ an, wobei allerdings nicht ausgeschlossen ist, daß $\Delta z = 0$ wird. Setzen wir nun $y_0 = f(z_0)$ und $y_0 + \Delta y = f(z_0 + \Delta z)$, so erhalten wir nach dem oben Bemerkten

$$\Delta y = y_z' \Delta z + \alpha \Delta z,$$

wobei α, wenn Δz eine gegen Null strebende Folge durchläuft, ebenfalls gegen Null strebt, und zwar unabhängig davon, ob von den Gliedern der betrachteten Folge gewisse verschwinden oder nicht.

Dividieren wir nun die zuletzt erhaltene Beziehung durch Δx und lassen wir anschließend Δx gegen Null gehen, so erhalten wir

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = y'_z \lim_{\Delta x \to 0} \frac{\Delta z}{\Delta x} + 0 \cdot \lim_{\Delta x \to 0} \frac{\Delta z}{\Delta x},$$

das heißt

$$y'_x = y'_z \cdot z'_x, \tag{11}$$

was wir auch in der Form

$$\{f(\varphi(x))\}' = f'(\varphi(x)) \cdot \varphi'(x)$$

schreiben können. In Worten besagt dies: *Die Ableitung einer mittelbaren Funktion ist gleich dem Produkt aus der Ableitung der Funktion nach der als Argument genommenen inneren Funktion und der Ableitung der inneren Funktion selbst.* In der nachfolgenden Tabelle sind nochmals alle von uns bewiesenen Differentiationsformeln und -regeln zusammengestellt:

1. $(C)' = 0$,
2. $(x)' = 1$,
3. $(x^n)' = n x^{n-1}$,
4. $(x^a)' = a x^{a-1}$,
5. $(\sin x)' = \cos x$,
6. $(\cos x)' = -\sin x$,
7. $(\operatorname{tg} x)' = \dfrac{1}{\cos^2 x}$,
8. $(\operatorname{ctg} x)' = -\dfrac{1}{\sin^2 x}$,
9. $(a^x)' = a^x \ln a$,
10. $(e^x)' = e^x$,
11. $(\ln x)' = \dfrac{1}{x}$,
12. $(\log_a x)' = \dfrac{1}{x \ln a}$,
13. $(\arcsin x)' = \dfrac{1}{\sqrt{1-x^2}}$,
14. $(\arccos x)' = -\dfrac{1}{\sqrt{1-x^2}}$,
15. $(\operatorname{arctg} x)' = \dfrac{1}{1+x^2}$,
16. $(\operatorname{arcctg} x)' = -\dfrac{1}{1+x^2}$,
17. $(u+v)' = u' + v'$,
18. $(u-v)' = u' - v'$,
19. $(uv)' = u'v + uv'$,
20. $(cu)' = cu'$,
21. $\left(\dfrac{u}{v}\right)' = \dfrac{u'v - uv'}{v^2}$,
22. $y'_x = \dfrac{1}{x'_y}$,
23. $y'_x = y'_z \cdot z'_x$.

(*)

Mit Hilfe dieser Formeln kann man die Ableitung jeder elementaren Funktion bestimmen, wobei die Ableitung selbst sich stets als elementare Funktion des Differentiationspunktes erweist.

Wir wollen die Anwendung dieser Formeln an einigen Beispielen illustrieren:

Beispiel 1. Gesucht ist die Ableitung y' der Funktion

$$y = 7x^3 + 8x^2 - 9x + 11.$$

Unter Verwendung der Regeln 17., 18. und 20. aus Tabelle (*) ergibt sich

$$\begin{aligned} y' &= (7x^3)' + (8x^2)' - (9x)' + (11)' \\ &= 7(x^3)' + 8(x^2)' - 9(x)' + (11)' = 21x^2 + 16x - 9. \end{aligned}$$

Analog beweist man ganz allgemein für ein beliebiges Polynom

$$y = a_0 x^n + a_1 x^{n-1} + a_2 x^{n-2} + \cdots + a_{n-1} x + a_n,$$

daß

$$y' = a_0 n x^{n-1} + a_1 (n-1) x^{n-2} + a_2 (n-2) x^{n-3} + \cdots + a_{n-2} 2x + a_{n-1}.$$

Beispiel 2. Gesucht ist die Ableitung y' der Funktion

$$y = \frac{\sin x}{\sqrt{x}} e^x.$$

Unter Verwendung der Regeln 19. und 21. sowie der Formeln 5., 4. und 10. aus Tabelle (*) findet man

$$y' = \frac{\cos x \cdot \sqrt{x} - \sin x \cdot \frac{1}{2} x^{-\frac{1}{2}}}{x} e^x + \frac{\sin x}{\sqrt{x}} e^x.$$

Beispiel 3. Gesucht ist die Ableitung y' der Funktion

$$y = (2x^3 + 5x^2 - 7x + 2) \frac{\operatorname{arctg} x}{\ln x}.$$

Man findet

$$y' = (6x^2 + 10x - 7) \frac{\operatorname{arctg} x}{\ln x} + (2x^3 + 5x^2 - 7x + 2) \frac{\frac{\ln x}{1+x^2} - \frac{\operatorname{arctg} x}{x}}{\ln^2 x}.$$

Beispiel 4. Gesucht ist die Ableitung der Funktion

$$y = e^{\sin x}.$$

Hier haben wir es im Gegensatz zu den vorangehenden Beispielen nicht mit dem Resultat von arithmetischen Operationen an einfachsten elementaren Funktionen, sondern mit einer mittelbaren Funktion zu tun.

Mit

$$y = e^z, \quad z = \sin x$$

finden wir

$$y'_x = e^z \cos x = e^{\sin x} \cos x.$$

Beispiel 5. Gesucht ist die Ableitung y' für

$$y = \sqrt{\operatorname{arctg} x}.$$

Setzen wir $y = \sqrt{z}$, $z = \operatorname{arctg} x$, so erhalten wir

$$y' = \frac{1}{2} (\operatorname{arctg} x)^{-\frac{1}{2}} \cdot \frac{1}{1+x^2}.$$

Beispiel 6. Gesucht ist die Ableitung y' für

$$y = \operatorname{tg}^3 \ln x.$$

Hier müssen wir die Kettenregel zweimal anwenden. Setzen wir nämlich

$$y = z^3, \quad z = \operatorname{tg} \ln x,$$

so erhalten wir

$$y' = 3z^2 \cdot z' = 3 \operatorname{tg}^2 \ln x \cdot z'.$$

Zur Berechnung von z' setzen wir sodann

$$z = \operatorname{tg} u, \quad u = \ln x$$

und erhalten

$$z' = \frac{1}{\cos^2 u} \cdot \frac{1}{x} = \frac{1}{x \cos^2 \ln x}.$$

Insgesamt ergibt sich also

$$y' = \frac{3 \operatorname{tg}^2 \ln x}{x \cos^2 \ln x}.$$

Das Auftreten von allzuvielen Hilfsbuchstaben bei mehrfacher Anwendung der Kettenregel kann man dadurch vermeiden, daß man das Argument der jeweils zu differenzierenden Funktion immer mit demselben Buchstaben, etwa z, bezeichnet. Ist z. B. die Funktion

$$y = \operatorname{tg}^5 \sqrt[4]{\ln \arcsin x}$$

zu differenzieren, so setzt man $y = z^5$ mit

$$z = \operatorname{tg} \sqrt[4]{\ln \arcsin x}.$$

Dann ist also

$$y' = 5 \operatorname{tg}^4 \sqrt[4]{\ln \arcsin x} \cdot \left(\operatorname{tg} \sqrt[4]{\ln \arcsin x}\right)'.$$

Sodann setzt man

$$z = \sqrt[4]{\ln \arcsin x}$$

und erhält

$$y' = 5 \operatorname{tg}^4 \sqrt[4]{\ln \arcsin x} \, \frac{1}{\cos^2 \sqrt[4]{\ln \arcsin x}} \left(\sqrt[4]{\ln \arcsin x}\right)'.$$

Anschließend versteht man unter z die Funktion

$$\ln \arcsin x,$$

was

$$y' = 5 \operatorname{tg}^4 \sqrt[4]{\ln \arcsin x} \cdot \frac{1}{\cos^2 \sqrt[4]{\ln \arcsin x}} \cdot \frac{1}{3} (\ln \arcsin x)^{-\frac{3}{4}} (\ln \arcsin x)'$$

ergibt. Schließlich setzt man

$$z = \arcsin x$$

und bekommt

$$y' = 5 \operatorname{tg}^4 \sqrt[4]{\ln \arcsin x} \cdot \frac{1}{\cos^2 \sqrt[4]{\ln \arcsin x}} \cdot \frac{1}{3} (\ln \arcsin x)^{-\frac{3}{4}} \frac{1}{\arcsin x} \cdot \frac{1}{\sqrt{1 - x^2}}.$$

Praktisch führt man natürlich alle diese Substitutionen nur im Gedächtnis aus und notiert sofort das Ergebnis der Differentiation.

Ist etwa $y = \sqrt{\operatorname{arctg} e^{\sin x}}$, so ist

$$y' = \frac{1}{2} (\operatorname{arctg} e^{\sin x})^{-\frac{1}{2}} \cdot \frac{1}{1 + e^{2 \sin x}} e^{\sin x} \cdot \cos x.$$

7. Differentiale. Mit dem Begriff der Ableitung hängt eng der Begriff des Differentials zusammen, von welchem die Bezeichnung „Differentialrechnung" herrührt.

Um zunächst das Wesen dieses Begriffs zu erläutern, betrachten wir eine im offenen Intervall (a, b) definierte und in einem gewissen Punkte x dieses Intervalls differenzierbare Funktion $y = f(x)$. Wir geben dann dem Argumentwert x einen von Null verschiedenen Zuwachs Δx, und zwar so, daß auch $x + \Delta x \in (a, b)$. Den entsprechenden Zuwachs der Funktion bezeichnen wir wie üblich mit Δy. Dann ist, wie wir im vorangehenden Abschnitt gesehen haben,

$$\Delta y = y' \Delta x + \alpha \Delta x,$$

wobei α mit Δx gegen Null strebt. Setzen wir nun

$$\alpha \Delta x = \varrho,$$

so wird mit Δx auch ϱ beliebig klein und strebt dabei schneller gegen Null als Δx, worunter genau zu verstehen ist, daß

$$\lim_{\Delta x \to 0} \frac{\varrho}{\Delta x} = 0$$

gilt. (Hat man allgemein zwei voneinander abhängige Größen λ und μ, wobei mit μ auch λ gegen Null strebt, so sagt man, λ strebe *schneller* oder *von höherer Ordnung* gegen Null als μ, wenn

$$\lim_{\mu \to 0} \frac{\lambda}{\mu} = 0$$

gilt.) Grob gesprochen bedeutet, daß ϱ von höherer Ordnung gegen Null strebt als Δx, daß bei kleinem Δx die Größe ϱ noch um Vieles kleiner ist als Δx. Daher kann in vielen Fällen die Größe $\varrho = \alpha \Delta x$ bei kleinem Δx vernachlässigt werden, was dann auf die Näherungsformel

$$\Delta y \approx y' \Delta x$$

führt. Diese Formel besagt, daß der Funktionszuwachs Δy bis auf einen kleinen Fehler höherer Ordnung (verglichen mit Δx) direkt proportional dem Argumentzuwachs Δx ist, wobei die Ableitung y' (im Punkte x) gerade die Rolle des Proportionalitätsfaktors spielt.

Das Produkt $y' \Delta x$, welches näherungsweise gleich dem Zuwachs Δy ist, wird nun das *dem Argumentzuwachs Δx entsprechende Differential der Funktion $y = f(x)$ im Punkte x* genannt. Das Differential von $y = f(x)$ wird folgendermaßen bezeichnet:

$$dy = y' \Delta x, \quad df(x) = f'(x) \Delta x. \tag{12}$$

Definitionsgemäß ist das Differential dy eine Funktion von zwei voneinander unabhängigen Veränderlichen, dem Differentiationspunkt x und dem Zuwachs Δx.[1]) In der Bezeichnung $df(x)$ tritt nur der Punkt x, in der Bezeichnung dy weder x noch Δx auf; daher sind beide Bezeichnungen unvollständig.

Das Differential spielt deshalb eine so große Rolle, weil es (bei festgehaltenem Differentiationspunkt) folgende bemerkenswerte Eigenschaften besitzt: 1. Es ist dem Argumentzuwachs direkt proportional. 2. Der Unterschied zwischen Differential und Funktionszuwachs (im betrachteten Differentiationspunkt) strebt von höherer Ordnung gegen Null als der Argumentzuwachs Δx.

Durch die erste Eigenschaft wird eine einfache Berechenbarkeit des Differentials gewährleistet, während die zweite es gestattet, bei hinreichend kleinem Zuwachs Δx den Zuwachs Δy angenähert durch dy zu ersetzen.

Zur Illustration des Gesagten betrachten wir folgendes einfache Beispiel: Es sei

$$y = 2x^3 + 3x^2 + 4x + 2.$$

Gesucht sind Δy und dy für $x = 2$, $\Delta x = 0{,}001$. Im betrachteten Fall ist $y = 38$, $y + \Delta y = 38{,}040015002$, also

$$\Delta y = 0{,}040015002.$$

Andererseits ist $y' = 6x^2 + 6x + 4$, für $x = 2$ ist also $y' = 40$ und daher

$$dy = 0{,}04.$$

In der früheren Bezeichnung ist also

$$\varrho = 0{,}000015002.$$

Der relative Fehler, den wir begehen, wenn wir Δy durch dy ersetzen, ist also

$$\delta = \frac{\varrho}{\Delta y} = \frac{0{,}000015002}{0{,}040015002} < \frac{0{,}000016}{0{,}04} = 0{,}0004 = 0{,}04\,\%.$$

Daraus folgt, daß hier die Approximation von Δy durch dy außerordentlich gut ist.

Wir wollen als nächstes Δy und dy für dieselbe Funktion

$$y = 2x^3 + 3x^2 + 4x + 2$$

in demselben Punkte $x = 2$, aber bezüglich des Zuwachses $\Delta x = 3$ berechnen. Hier ist $y = 38$, $y + \Delta y = 347$, also

$$\Delta y = 309;$$

andererseits ist $y' = 40$, also

$$dy = 120.$$

---

[1]) Am Anfang dieses Abschnitts haben wir Δx der Bedingung $x + \Delta x \in (a, b)$ unterworfen. Dadurch hängt der Variabilitätsbereich von Δx von der Wahl von x ab. Man kann sich jedoch von dieser Unbequemlichkeit sofort dadurch befreien, daß man dy durch $dy = y' \Delta x$ als für beliebiges Δx definiert ansieht.

Wir sehen also, daß jetzt Δy und dy durchaus nicht nahe beieinander liegen. Der recht bedeutende Unterschied zwischen den beiden Beispielen erklärt sich daraus, daß der im ersten Beispiel auftretende Zuwachs Δx recht klein war, während im zweiten Beispiel Δx zu groß gewählt wurde. Im allgemeinen ist also die Verwendung der oben angegebenen Näherungsformel ohne Abschätzung des begangenen Fehlers eine recht zweifelhafte Angelegenheit. Man kann z. B. sagen, daß 7 näherungsweise gleich 32 ist, und obwohl diese Aussage in praktischer Hinsicht völlig unsinnig ist, ist sie dennoch formal richtig. Wenn wir also im oben angegebenen Zusammenhang feststellten, daß für hinreichend kleines Δx die Größen Δy und dy annähernd gleich sind, so ist diese Aussage zumindest einer näheren Präzisierung bedürftig. Weiter unten werden wir zeigen, daß für eine recht umfangreiche Klasse von Funktionen der Fehler, der beim Ersetzen von Δy durch dy begangen wird, nicht größer ist als

$$M(\Delta x)^2,$$

wobei M die obere Grenze des absoluten Betrages der zweiten[1]) Ableitung $f''(x)$ im betrachteten Intervall ist. Bis dahin wollen wir uns mit der oben angegebenen, etwas verschwommenen Formulierung zufrieden geben.

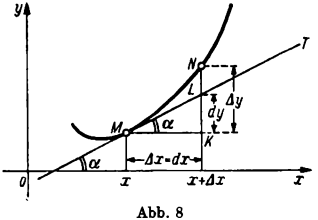


Abb. 8

Wir wollen noch kurz auf die geometrische Bedeutung des Differentials eingehen. Dazu betrachten wir das geometrische Bild der gegebenen Funktion $y = f(x)$ und denken uns an den dem Differentiationspunkt x entsprechenden Kurvenpunkt $M(x, y)$ die Tangente MT gelegt (Abb. 8). Daneben betrachten wir den Kurvenpunkt $N(x + \Delta x, y + \Delta y)$, wobei $y + \Delta y = f(x + \Delta x)$. Dann ist der Zuwachs Δy der Funktion gleich der Länge der Strecke KN in der Abbildung, während die Länge der Strecke KL durch

$$KL = MK \cdot \operatorname{tg}\alpha = y' \Delta x$$

gegeben wird, also gleich dem Differential dy ist, d. h., *das Differential bedeutet geometrisch den Zuwachs der Ordinate auf der Tangente an das geometrische Bild der Funktion im Punkte* $M(x, y)$. Da in unmittelbarer Nähe des Berührungspunktes die Tangente sich eng an die Kurve anschmiegt, wird auch hieraus klar, daß für kleines Δx das Ersetzen von Δy durch dy keinen bemerkenswerten Fehler verursacht.

Hier noch einige weitere Beispiele für die Bestimmung des Differentials:

1. Gesucht ist das Differential der Funktion $y = x^3$ für $x = 2$ und $\Delta x = 0{,}1$. Im betrachteten Beispiel ist allgemein $y' = 3x^2$, für $x = 2$ ist also $y' = 12$, so daß $dy = 1{,}2$.

[1]) Vgl. S. 316.

Definitionsgemäß hängt das Differential (außer von der betrachteten Funktion) vom Differentiationspunkt x und dem betrachteten Zuwachs Δx ab. Im folgenden wollen wir uns diese beiden Größen nicht zahlenmäßig vorgegeben denken, sondern allgemein dy formelmäßig durch x und Δx ausdrücken.

2. Gesucht ist dy für $y = e^x$. In diesem Fall ist $y' = e^x$ und daher $dy = e^x \Delta x$ oder kurz

$$de^x = e^x \Delta x.$$

Analog findet man

3. $d \operatorname{tg} x = \dfrac{\Delta x}{\cos^2 x}$,

4. $d \ln x = \dfrac{\Delta x}{x}$

und

5. $d \operatorname{tg}^6 \sqrt[5]{\ln \sin x} = 6 \operatorname{tg}^5 \sqrt[5]{\ln \sin x} \dfrac{1}{\cos^2 \sqrt[5]{\ln \sin x}} \cdot \dfrac{1}{3} (\ln \sin x)^{-\frac{1}{5}} \dfrac{1}{\sin x} \cos x \cdot \Delta x.$

Es ist klar, daß es bei der Berechnung der Differentiale im wesentlichen auf die Berechnung der zugehörigen Ableitung ankommt (und natürlich auch umgekehrt). Aus diesem Grunde wird das Bestimmen der Ableitung auch als „Differentiation" bezeichnet.

Ein sehr aufschlußreiches Beispiel für die Ermittlung des Differentials liefert die Funktion $y = x$. Hier ist nämlich $y' = 1$ und daher $dy = \Delta x$, d. h.

$$dx = \Delta x.$$

Bei der Funktion $y = x$ stimmt also das Differential mit dem Zuwachs überein.[1])

Hieraus folgt, daß die Formel (12), durch welche ja gerade das Differential definiert wurde, auch in der Form

$$dy = y' dx \tag{13}$$

geschrieben werden kann. Aus diesem Grunde kann man die Ableitung y' auch durch

$$\frac{dy}{dx}$$

bezeichnen, und diese Bezeichnungsweise ist tatsächlich sehr weit verbreitet.[2])

Bemerkenswert ist an Formel (13), daß sie auch dann noch richtig ist, wenn x nicht mehr die Rolle der unabhängigen Veränderlichen spielt, sondern seinerseits Funktion einer Veränderlichen t ist. In diesem Fall ist auch y eine (mittelbare) Funktion von t, und das Differential dy zum Zuwachs Δt wird durch

$$dy = y'_t \Delta t$$

[1]) Allgemein kommt diese Eigenschaft jeder linearen Funktion $y = ax + b$ zu.

[2]) Diese Tatsache erklärt, daß die Ableitung häufig auch als *Differentialquotient* bezeichnet wird. — *Anm. d. wissenschaftl. Red.*

gegeben. Auf Grund der Kettenregel ist nun aber

$$y_t' = y_x' x_t'$$

und daher

$$dy = y_x' x_t' \Delta t.$$

Berücksichtigen wir schließlich noch, daß $x_t' \Delta t = dx$ ist, so kommen wir unmittelbar zu Formel (13).

Um Irrtümer zu vermeiden, weisen wir darauf hin, daß eine entsprechende Eigenschaft der Formel (12) natürlich nicht zukommt. Ist nämlich x eine Funktion von t, so bezeichnen Δx und dx vollkommen verschiedene Begriffe, und (12) und (13) sind keineswegs mehr gleichwertig.

Der wesentliche Vorzug der Schreibweise (13) für das Differential dy besteht geradezu in der „Invarianz" gegenüber Ersetzung von x durch eine Funktion $x(t)$. So gilt z. B. für $y = x^3$ die Beziehung

$$dy = 3x^2 dx$$

unabhängig davon, ob x die Rolle der unabhängigen Veränderlichen spielt oder eine Funktion ist.

Wir kommen nun zu einigen Beispielen für die Anwendung des Differentials.

Beispiel 1. Man berechne $(1+h)^a$ für sehr kleines h. Zu diesem Zweck betrachten wir die Funktion $y = x^a$. Da der Funktionswert dieser Funktion für $x = 1$ bekannt (und zwar gleich 1) ist, genügt es, den Zuwachs Δy der Funktion für den Argumentzuwachs $\Delta x = h$ zu berechnen. Da h sehr klein sein sollte, können wir Δy durch dy ersetzen. Im betrachteten Fall ist nun allgemein $dy = a x^{a-1} dx = a x^{a-1} \Delta x$, was für $x = 1$, $\Delta x = h$ auf $dy = ah$ führt. Damit erhalten wir die Näherungsformel

$$(1+h)^a \approx 1 + ah.$$

Aus dieser Formel ergibt sich eine schnell zum Ziele führende Methode zur näherungsweisen Berechnung von Wurzeln. Ist nämlich n eine natürliche und B eine im Vergleich zu A^n kleine Zahl, so ist

$$\sqrt[n]{A^n + B} = A\left(1 + \frac{B}{A^n}\right)^{\frac{1}{n}} \approx A\left(1 + \frac{B}{nA^n}\right)$$

und daher

$$\sqrt[n]{A^n + B} \approx A + \frac{B}{nA^{n-1}}.$$

Zum Beispiel ist also

$$\sqrt[3]{131} = \sqrt[3]{125 + 6} \approx 5 + \frac{6}{3 \cdot 5^2} \approx 5{,}08$$

(tatsächlicher Wert: $\sqrt[3]{131} = 5{,}0788\ldots$).

Beispiel 2. Man berechne $\operatorname{tg} 31°$. Hierbei ist die Funktion $y = \operatorname{tg} x$ zu betrachten. Da der Wert der Funktion $\operatorname{tg} x$ für $x = 30° = \frac{\pi}{6}$ bekannt ist, genügt

es, den Zuwachs Δy der Funktion $y = \operatorname{tg} x$ zu berechnen, wenn man dem Argumentwert $x = 30°$ den Zuwachs $\Delta x = 1° = \frac{\pi}{180}$ erteilt. Statt des Zuwachses Δy berechnen wir einfach das Differential dy:

$$dy = \frac{\Delta x}{\cos^2 x} = \frac{1}{\cos^2 30°} \cdot \frac{\pi}{180} = \frac{\pi}{135} = 0{,}023\ldots.$$

Wegen

$$\operatorname{tg} 30° = \frac{\sqrt{3}}{3} = 0{,}577\ldots$$

ist dann

$$\operatorname{tg} 31° \approx 0{,}600\ldots$$

(tatsächlicher Wert: $\operatorname{tg} 31° = 0{,}60076\ldots$).

Beispiel 3. Es sei $y = uv$, wobei u und v Funktionen derselben Veränderlichen x sind. Wir wollen nun annehmen, daß x einen kleinen Zuwachs dx erfährt. Dann ist

$$dy = y'dx = (uv' + u'v)\,dx = u\,dv + v\,du$$

und daher

$$\frac{dy}{y} = \frac{du}{u} + \frac{dv}{x}.$$

Sind u und v zwei mit irgendeinem Meßgerät gemessene Größen und du und dv die entsprechenden Meßfehler, so bedeutet die erhaltene Beziehung, daß *der relative Fehler des Produkts gleich der Summe aus den relativen Fehlern der Faktoren ist.*[1])

Zum Abschluß sei erwähnt, daß das Differential dy einer Funktion $y = f(x)$ bereits durch die folgenden Eigenschaften vollständig definiert ist. 1. Es ist proportional zu Δx. 2. Der Quotient

$$\frac{\Delta y - dy}{\Delta x}$$

strebt mit Δx gegen Null. In der Tat: Es sei z eine beliebige Größe, welche diese beiden Eigenschaften besitzt, d. h., es sei $z = k\Delta x$, und es gelte

$$\lim_{\Delta x \to 0} \frac{\Delta y - z}{\Delta x} = 0. \tag{14}$$

Wegen

$$\frac{\Delta y - z}{\Delta x} = \frac{\Delta y}{\Delta x} - k$$

und (14) ist dann

$$k = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x},$$

also $k = y'$ und daher $z = dy$.

Aus diesem Grunde bedeutet jedes „Runden" von Δy, bei welchem Δy (bis auf eine von höherer Ordnung als Δx gegen Null strebende Größe) durch

[1]) Vgl. EdEM Band I, S. 351. — *Anm. d. wissenschaftl. Red.*

eine zu Δx proportionale Größe ersetzt wird, das Ersetzen von Δy durch das Differential dy.

Zum Beispiel wird der kubische Ausdehnungskoeffizient β eines Materials in der Physik folgendermaßen berechnet: Ein Würfel aus diesem Material mit der Kantenlänge 1 wird um 1° erhitzt. Dann vergrößert sich dabei die Länge jeder Kante auf $1+\alpha$, wobei α der lineare Ausdehnungskoeffizient des betr. Materials ist. Daher vergrößert sich bei der Erwärmung das Volumen auf

$$(1+\alpha)^3 = 1 + 3\alpha + 3\alpha^2 + \alpha^3,$$

so daß der Volumenzuwachs

$$3\alpha + 3\alpha^2 + \alpha^3$$

beträgt. Vernachlässigt man die Glieder $3\alpha^2$ und α^3 (dies ist möglich, weil α sehr klein ist), so kommt man zu der Formel

$$\beta = 3\alpha.$$

Mathematisch bedeutet dies gerade, daß der Volumenzuwachs durch das Differential ersetzt wird.

Dieses Verfahren, kleine Änderungen einer Funktion als proportional den entsprechenden Änderungen des Arguments anzusehen, ist dem Leser auch bereits aus der praktischen Arbeit mit der Logarithmentafel bekannt. Es werden also sogar schon auf der Schule, ohne daß es dort natürlich gesagt wird, Differentiale benutzt.

8. Ableitungen und Differentiale höherer Ordnung. Es sei $y = f(x)$ eine im Intervall $[a, b]$ erklärte Funktion, welche in jedem Punkt dieses Intervalls differenzierbar ist.[1]) Dann ist die Ableitung $y' = f'(x)$ ihrerseits eine Funktion von x (wobei x der jeweilige Differentiationspunkt ist). In diesem Fall kann man in den einzelnen Punkten des Intervalls $[a, b]$ nach der Existenz der Ableitung von $y' = f'(x)$ fragen. Man bezeichnet diese auch als *zweite Ableitung* (oder *Ableitung zweiter Ordnung*) der Funktion $y = f(x)$ und deutet sie (sofern sie existiert) durch y'' oder $f''(x)$ an.

Allgemein versteht man unter der $(n+1)$-*ten Ableitung* einer Funktion $y = f(x)$ die Ableitung ihrer n-ten Ableitung. Für die Ableitung dritter Ordnung sind die Bezeichnungen y''' und $f'''(x)$ üblich, während die weiteren Ableitungen mit $y^{(IV)}, y^{(V)}, \ldots, y^{(n)}, \ldots$ bzw. $f^{(4)}(x), f^{(5)}(x), \ldots, f^{(n)}(x), \ldots$ bezeichnet werden. Man sieht leicht ein, daß nicht jede Funktion Ableitungen beliebiger Ordnung besitzt; es kann auch der Fall eintreten, daß Ableitungen einer gewissen Ordnung in einer Reihe von Punkten existieren, während sie in anderen Punkten nicht vorhanden sind. Insbesondere muß man bedenken, daß die Existenz der Ableitung $f^{(n+1)}(x)$ in einem Punkte x_0 die Existenz der n-ten Ableitung nicht nur im Punkte x_0, sondern in einem ganzen, x_0 enthaltenden offenen[2]) Intervall voraussetzt.

---

[1]) In den Punkten a und b kann man natürlich sinngemäß nur von den einseitigen Ableitungen sprechen.

[2]) Dieses Intervall kann auch abgeschlossen sein und x_0 als Randpunkt besitzen. In diesem Fall ist $f^{(n+1)}(x_0)$ nur eine einseitige Ableitung.

Obwohl man definitionsgemäß zur Berechnung der $(n+1)$-ten Ableitung die n-te Ableitung selbst benötigt, kann man in einer Reihe von einfachen Fällen den Ausdruck für die Ableitungen höherer Ordnung sofort angeben, ohne erst die Ableitungen der vorangehenden Ordnungen zu notieren. So ist z. B. vollkommen klar, daß für $y = e^x$ für jedes n auch $y^{(n)} = e^x$ ist. Genauso einfach ergibt sich, daß für $y = \sin x$ wegen $y' = \cos x = \sin\left(x + \frac{\pi}{2}\right)$ für jedes n

$$y^{(n)} = \sin\left(x + n\frac{\pi}{2}\right)$$

ist. Eine weitere wichtige Tatsache ist, daß für jedes Polynom m-ten Grades

$$y = a_0 x^m + a_1 x^{m-1} + \cdots + a_{m-1} x + a_m$$

die m-te Ableitung $y^{(m)}$ konstant gleich $a_0 m!$ ist, während alle Ableitungen höherer als m-ter Ordnung identisch verschwinden.

Mit Hilfe der Ableitungen höherer Ordnung kann man in folgender Weise den binomischen Satz beweisen: Es sei n eine beliebige natürliche Zahl. Dann ist, wie man leicht einsieht, $(a+x)^n$ ein Polynom n-ten Grades in x:

$$(a+x)^n = A_0 + A_1 x + A_2 x^2 + \cdots + A_n x^n. \tag{15}$$

Unsere Aufgabe soll es nun sein, die Koeffizienten $A_0, A_1, \ldots, A_n$ zu bestimmen.

Indem wir zunächst $x = 0$ setzen, erhalten wir

$$a^n = A_0.$$

Differenzieren wir die Identität (15) nach x, so finden wir

$$n(a+x)^{n-1} = A_1 + 2A_2 x + \cdots + nA_n x^{n-1}, \tag{16}$$

woraus sich für $x = 0$

$$n a^{n-1} = A_1$$

ergibt. Differenzieren wir sodann (16), so erhalten wir

$$n(n-1)(a+x)^{n-2} = 2 \cdot 1 A_2 + 3 \cdot 2 A_3 x + \cdots + n(n-1) A_n x^{n-2} \tag{17}$$

und für $x = 0$

$$n(n-1)\, a^{n-2} = 2 \cdot 1 \cdot A_2.$$

Die Identität (16) ist das Resultat einmaliger, die Identität (17) das Resultat zweimaliger Differentiation von (15). Da bei jeder weiteren Differentiation der Exponent jedes Gliedes als Faktor vor dieses Glied tritt und sich der Exponent um Eins erniedrigt, erhalten wir nach m-maliger Differentiation von (15) $(m \leqq n)$

$$\underbrace{n(n-1)\cdots(n-m+1)}_{m \text{ Faktoren}}(a+x)^{n-m} = m(m-1)\cdots 1 \cdot A_m + (m+1)\, m \cdots 2 \cdot A_{m+1} x + \cdots + n(n-1)\cdots(n-m+1) A_n x^{n-m}.$$

Hierbei ist der erste Summand auf der rechten Seite übrigens gerade die m-te Ableitung von $A_m x^m$, der zweite Summand die m-te Ableitung von $A_{m+1} x^{m+1}$ usw. Setzen wir in der zuletzt erhaltenen Identität $x = 0$, so finden wir

$$n(n-1)\cdots(n-m+1)\,a^{n-m} = m(m-1)\cdots 1 \cdot A_m.$$

Die vorangehenden Überlegungen zeigen, daß man durch fortgesetztes Differenzieren und jeweiliges Einsetzen von $x = 0$ nach und nach alle Koeffizienten $A_0, A_1, \ldots, A_n$ bestimmen kann. Insgesamt ergibt sich dabei

$$A_0 = a^n, \quad A_1 = n a^{n-1}, \quad A_2 = \frac{n(n-1)}{2 \cdot 1}\, a^{n-2}, \quad \ldots,$$

$$A_m = \frac{n(n-1)\cdots(n-m+1)}{m(m-1)\cdots 1}\, a^{n-m}, \quad \ldots,$$

$$A_n = \frac{n(n-1)\cdots(n-n+1)}{n(n-1)\cdots 1}\, a^{n-n} = 1.$$

Setzen wir dies in (15) ein, so erhalten wir

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

Das ist aber gerade die Aussage des binomischen Satzes. Bemerkenswert ist, daß der vorangehende Beweis ganz ohne kombinatorische Hilfsmittel auskommt.

Ganz allgemein kann man durch Differentiation aus einer Identität stets eine neue Identität gewinnen. So ergibt sich z. B. aus der Identität

$$1 + x + x^2 + \cdots + x^n = \frac{1 - x^{n+1}}{1 - x}$$

durch Differentiation sofort die neue Identität

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

Mit dem Begriff der Ableitung höherer Ordnung hängt eng der Begriff des Differentials höherer Ordnung zusammen. Wir haben im vorangehenden Abschnitt gesehen, daß das Differential $dy = y'dx$ einer Funktion $y = f(x)$ von zwei Größen, nämlich dem Differentiationspunkt x und dem Zuwachs $\Delta x = dx$ abhängt (wir betrachten jetzt x als unabhängige Veränderliche). Wenn wir nun dx festhalten, so wird dy eine Funktion der einen Veränderlichen x. Das Differential dieser neuen Funktion nennt man dann das *zweite Differential* (oder *Differential zweiter Ordnung*) der Funktion $y = f(x)$. Es wird mit d^2y bzw. $d^2f(x)$ bezeichnet. Nach Definition muß man zur Ermittlung des zweiten Differentials die Ableitung der Funktion dy bestimmen und das erhaltene Resultat mit dx multiplizieren. Theoretisch könnte nun dieser zuletzt genannte Faktor dx durchaus von dem bei der Bestimmung von dy verwendeten Zuwachs dx, der bei der Differentiation festgehalten wird, verschieden sein. Die Definition des Differentials zweiter Ordnung ist jedoch

(und das ist genau genommen eine weitere Bedingung) so gemeint, daß in beiden Fällen derselbe Zuwachs zu nehmen ist. Dann ist

$$d^2y = d(dy) = d(y'dx) = (y'dx)'dx = y''(dx)^2.$$

Üblicherweise wird das Quadrat $(dx)^2$ kurz durch dx^2 bezeichnet (diese Bezeichnung kann allerdings leicht zu Irrtümern führen, weil dx^2 ebensogut das Differential der Funktion x^2 bezeichnen könnte). Daher ist schließlich

$$d^2y = y''dx^2.$$

Das Differential dieser Funktion (bei festgehaltenem dx und gleichem dx als Faktor) wird das *dritte Differential* oder *Differential dritter Ordnung* der Funktion $y = f(x)$ genannt und mit d^3y bzw. $d^3f(x)$ bezeichnet. Dann ist offenbar

$$d^3y = y'''dx^3.$$

Entsprechend bezeichnet man allgemein als *Differential* $(n+1)$*-ter Ordnung* der Funktion $y = f(x)$ das Differential des Differentials n-ter Ordnung, wobei wiederum dx festgehalten und dx für beide Differentiale in gleicher Weise gewählt wird. Man sieht unmittelbar, daß dann

$$d^ny = y^{(n)}dx^n \tag{18}$$

ist, so daß man die n-te Ableitung $y^{(n)} = f^{(n)}(x)$ einer Funktion $y = f(x)$ allgemein folgendermaßen schreiben kann:

$$\frac{d^ny}{dx^n} \quad \text{bzw.} \quad \frac{d^nf(x)}{dx^n}.$$

Im vorangehenden Abschnitt haben wir gesehen, daß die Formel

$$dy = y'dx$$

invariant ist gegenüber Ersetzung von x durch eine Funktion $x(t)$. Man sieht leicht ein, daß eine derartige Invarianz für die Differentiale höherer Ordnung nicht mehr vorhanden ist, d. h., daß die Formel (18) nicht richtig ist, wenn man dort die unabhängige Veränderliche x durch eine Funktion $x(t)$ ersetzt. In der Tat: Bei der Definition des Differentials d^2y wurde das Differential $dy = y'dx$ bei festgehaltenem dx als Funktion von x betrachtet. Ist nun x eine Funktion von t, so ist $dx = x'_t\,dt$ ebenfalls eine Funktion t, und das Festhalten von dx (bei veränderlichem t) ist unmöglich.

Dies läßt sich natürlich auch durch eine direkte Rechnung leicht bestätigen. Ist nämlich $y = f(x)$ und $x = \varphi(t)$, so ist y eine mittelbare Funktion von t, und es gilt

$$d^2y = y''_t\,dt^2.$$

Auf Grund der Differentiationsregel für mittelbare Funktionen ist nun

$$y''_t = (y'_x)'_t = (y'_x x'_t)'_t = (y'_x)'_t\,x'_t + y'_x x''_t.$$

Ferner ist $(y'_x)'_t = (y'_x)'_x \cdot x'_t = y''_x\,x'_t$. Daher gilt

$$d^2y = y''_x x'^2_t\,dt^2 + y'_x x''_t\,dt^2.$$

Berücksichtigen wir schließlich noch, daß

$$x_t' \, dt = dx, \quad x_t'' \, dt^2 = d^2x,$$

so gelangen wir zu der Formel

$$d^2y = y_x'' \, dx^2 + y_x' \, d^2x,$$

und diese ist von Formel (18) verschieden.

9. Partielle Ableitungen und totales Differential. Zwar ist der vorliegende Artikel im wesentlichen den Funktionen einer Veränderlichen gewidmet, jedoch soll hier kurz auch etwas über die Differentiation von Funktionen mehrerer Veränderlicher gesagt werden. Der Begriff der Ableitung für eine Funktion von mehreren Veränderlichen wird fast unmittelbar auf den Begriff der Ableitung einer Funktion einer Veränderlichen zurückgeführt, indem man nämlich bei einer Funktion von mehreren Veränderlichen alle Veränderlichen bis auf eine festhält und die Funktion der einen noch verbleibenden Veränderlichen betrachtet.

Sei z. B. $z = f(x, y)$ eine Funktion der beiden Veränderlichen x und y. Wir betrachten dann einen festen Punkt (x_0, y_0) und bestimmen den zugehörigen Funktionswert $f(x_0, y_0)$. Sodann geben wir bei festgehaltenem $y = y_0$ dem Argument x_0 einen Zuwachs $\Delta x \neq 0$. Der zugehörige Funktionswert ist dann $f(x_0 + \Delta x, y_0)$. Der Grenzwert

$$\lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x, \, y_0) - f(x_0, \, y_0)}{\Delta x} \tag{19}$$

wird die *partielle Ableitung der Funktion $f(x, y)$ bezüglich der Veränderlichen x im Punkte (x_0, y_0)* genannt. Offenbar ist dies nichts anderes als die gewöhnliche Ableitung der Funktion $f(x, y_0)$ einer Veränderlichen. Für den Grenzwert (19) sind folgende Bezeichnungen üblich:

$$f_x(x_0, y_0), \quad \frac{\partial f}{\partial x}, \quad z_x, \quad \frac{\partial z}{\partial x}.$$

Analog definiert man die partielle Ableitung $\dfrac{\partial z}{\partial y}$.

Wir weisen darauf hin, daß im Gegensatz zur Ableitung von einer Funktion einer Veränderlichen das Symbol $\dfrac{\partial z}{\partial x}$ nicht als Quotient von irgendwie definierten Größen ∂z und ∂x aufgefaßt werden kann. Diese Symbole sind für sich allein vollkommen sinnlos. Wir wollen an einem Beispiel zeigen, daß die Nichtbeachtung dieser Tatsache zu völlig unsinnigen Beziehungen führt.

Seien etwa die Veränderlichen x, y, z durch die Beziehung

$$xyz = 1 \tag{20}$$

miteinander verknüpft. Dann ist $z = \dfrac{1}{xy}$ und daher $\dfrac{\partial z}{\partial x} = -\dfrac{1}{x^2 y}$. Analog ergibt sich $\dfrac{\partial x}{\partial y} = -\dfrac{1}{y^2 z}$ und $\dfrac{\partial y}{\partial z} = -\dfrac{1}{z^2 x}$, so daß $\dfrac{\partial x}{\partial y} \cdot \dfrac{\partial y}{\partial z} \cdot \dfrac{\partial z}{\partial x} = \dfrac{-1}{x^3 y^3 z^3}$ und wegen (20)

$$\frac{\partial x}{\partial y} \cdot \frac{\partial y}{\partial z} \cdot \frac{\partial z}{\partial x} = -1.$$

Wären nun die angegebenen Ausdrücke Quotienten, so müßte natürlich ihr Produkt gleich $+1$ sein.

Der Leser überlegt sich leicht, daß *die partiellen Ableitungen geometrisch die Steigungskoeffizienten der Tangenten an die Schnittkurven der Fläche $z = f(x, y)$ mit den Ebenen $y = y_0$ und $x = x_0$ sind.*

Interessant ist, daß im Gegensatz zu den Verhältnissen bei den Funktionen einer Veränderlichen die Existenz beider partiellen Ableitungen $\frac{\partial z}{\partial x}$ und $\frac{\partial z}{\partial y}$ in einem Punkte (x_0, y_0) noch nicht die Stetigkeit der Funktion $f(x, y)$ im Punkte (x_0, y_0) gewährleistet. Betrachten wir z. B. die Funktion

$$f(x, y) = \begin{cases} 0 & \text{für } xy \neq 0, \\ 1 & \text{für } xy = 0. \end{cases}$$

Man sieht sofort, daß diese Funktion im Punkte (0,0) unstetig ist, obwohl beide partiellen Ableitungen in diesem Punkte vorhanden sind.

Wenn für eine Funktion $z = f(x, y)$ überall die partielle Ableitung $\frac{\partial z}{\partial x}$ existiert, so kann man die partielle Ableitung $\frac{\partial z}{\partial x}$ ihrerseits als Funktion von x und y ansehen und von ihr unter Umständen wiederum die partiellen Ableitungen bilden. Dies führt dann auf die *partiellen Ableitungen zweiter Ordnung*, die man mit

$$\frac{\partial^2 z}{\partial x^2}, \quad \frac{\partial^2 z}{\partial x \, \partial y}$$

bezeichnet.

Entsprechend kann man mit der Funktion $\frac{\partial z}{\partial y}$ verfahren und die Ableitungen $\frac{\partial^2 z}{\partial y \, \partial x}$ und $\frac{\partial^2 z}{\partial y^2}$ bilden. Dabei hat man zunächst zwischen den Ableitungen $\frac{\partial^2 z}{\partial x \, \partial y}$ und $\frac{\partial^2 z}{\partial y \, \partial x}$ zu unterscheiden. Es zeigt sich allerdings, daß *diese beiden partiellen Ableitungen, wenn sie existieren und stetig sind*[1]*, übereinstimmen*:

$$\frac{\partial^2 z}{\partial x \, \partial y} = \frac{\partial^2 z}{\partial y \, \partial x}.$$

Diesen kurzen Ausführungen sei noch folgendes hinzugefügt:

Wir haben oben gesehen, daß im Fall einer Funktion einer Veränderlichen der zu einem Zuwachs Δx des Arguments gehörende Zuwachs Δy einer differenzierbaren Funktion $y = f(x)$ in der Form

$$\Delta y = f'(x) \, \Delta x + \varrho \tag{21}$$

dargestellt werden kann, wobei ϱ von höherer Ordnung gegen Null strebt als Δx. Etwas ganz Ähnliches gilt nun auch für die Funktionen mehrerer Veränderlichen. Es sei z. B. $z = f(x, y)$ eine Funktion der beiden Veränderlichen x und y, für welche in allen Punkten des Rechtecks $a < x < b$, $c < y < d$ die partiellen Ableitungen $f_x(x, y)$ und $f_y(x, y)$ existieren, wobei diese partiellen Ableitungen selbst stetige Funktionen des Differentiationspunktes (x, y) seien. In diesem Fall kann der zu dem Zuwachs Δx und dem Zuwachs Δy gehörende Zuwachs Δz der Funktion $f(x, y)$ in der Form

$$\Delta z = f_x(x, y) \, \Delta x + f_y(x, y) \, \Delta y + \varrho \tag{22}$$

dargestellt werden, wobei x und y die betrachteten Ausgangswerte sind und ϱ von höherer Ordnung als $|\Delta x| + |\Delta y|$ gegen Null strebt. Den Beweis hierfür werden wir in Abschnitt 11 führen. Der Ausdruck

$$dz = f_x(x, y) \, \Delta x + f_y(x, y) \, \Delta y$$

[1]) Diese Voraussetzung ist im allgemeinen Fall nicht entbehrlich.

wird das *totale* oder *vollständige Differential* der Funktion $f(x, y)$ genannt. Das vollständige Differential einer Funktion von zwei Veränderlichen ist seinerseits eine Funktion der vier Veränderlichen x, y, Δx, Δy, die in den Veränderlichen Δx und Δy linear ist. Es gibt bis auf einen Fehler ϱ, der von höherer Ordnung als $|\Delta x| + |\Delta y|$ gegen Null geht, den Zuwachs Δz an, den die Funktion $z = f(x, y)$ erfährt, wenn man den Argumentwerten x, y bzw. den Zuwachs Δx, Δy erteilt. Damit haben wir eine fast vollständige Analogie zum Fall der Funktionen einer Veränderlichen. Der einzige Unterschied liegt darin, daß es für die Gültigkeit von Formel (21) genügt, daß die Ableitung $f'(x)$ für denjenigen Punkt existiert, zu welchem der Zuwachs Δx hinzugefügt wird, während im Fall einer Funktion von zwei Veränderlichen das Vorhandensein der Ableitungen $f_x(x, y)$ und $f_y(x, y)$ nur im Punkte (x, y) für die Gültigkeit von (22) nicht ausreicht.[1]) Vielmehr benötigt man ihre Existenz und Stetigkeit in einem ganzen Rechteck (oder Quadrat oder Kreis), welches diesen Punkt in seinem Innern enthält.

§ 2. Wichtige Sätze über Ableitungen

10. Die Sätze von Fermat und Rolle. In Abschnitt 2 haben wir bereits damit begonnen, die prinzipielle Bedeutung des Begriffs der Ableitung zu erläutern. Wir werden im folgenden sehen, daß mit Hilfe dieses Begriffs viele wichtige konkrete Aufgaben gelöst werden können. Dazu haben wir jedoch zunächst den Begriff der Ableitung genau zu studieren. Dieser Aufgabe ist der vorliegende Paragraph gewidmet.

Wir beginnen mit einem Satz, den man gewöhnlich als Satz von Fermat bezeichnet. Zwar hat Fermat noch nicht den Begriff der Ableitung verwendet, jedoch stellt dieser Satz eine Präzisierung von nicht ganz klar umrissenen Gedanken Fermats dar.

Satz von Fermat. *Es sei $f(x)$ eine in einem abgeschlossenen Intervall $[a, b]$ definierte Funktion. Nimmt diese Funktion in einem gewissen inneren*[2]) *Punkt x_0 dieses Intervalls ihren größten (kleinsten) Wert an und existiert in diesem Punkt die Ableitung $f'(x_0)$, so ist*

$$f'(x_0) = 0.$$

Beweis. Wir wollen annehmen, daß die Funktion $f(x)$ im Punkte x_0 ihren größten Wert annimmt. Da x_0 nach Voraussetzung ein innerer Punkt des Intervalls $[a, b]$ ist, muß die Ableitung $f'(x_0)$ zugleich rechtsseitige und linksseitige Ableitung sein. Aus der Tatsache, daß $f'(x_0)$ rechtsseitige Ableitung ist, folgt, daß $f'(x_0)$ gleich dem Grenzwert

$$\lim \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

für $\Delta x \to 0$ mit $\Delta x > 0$ ist. Da nun $f(x_0)$ größter Wert von $f(x)$ im Intervall $[a, b]$ sein sollte, ist

$$f(x_0 + \Delta x) - f(x_0) \leqq 0,$$

[1]) Wir haben oben gesehen, daß eine Funktion $f(x, y)$ in einem Punkte (x, y) beide partiellen Ableitungen besitzen und doch dort unstetig sein kann. Es ist klar, daß in einem solchen Fall die Formel (22) nicht gelten wird.

[2]) Das heißt, es sei $a < x_0 < b$.

so daß wegen $\Delta x > 0$ auch

$$\frac{f(x_0+\Delta x)-f(x_0)}{\Delta x} \leqq 0.$$

Aus diesem Grunde kann der Grenzwert hiervon für $\Delta x \to 0$ nicht positiv sein, d. h.

$$f'(x_0) \leqq 0.$$

Analog ergibt sich

$$f'(x_0) \geqq 0$$

aus der Tatsache, daß $f'(x_0)$ linksseitige Ableitung von $f(x)$ im Punkte x_0 ist. Damit ist der Satz von FERMAT bereits bewiesen.

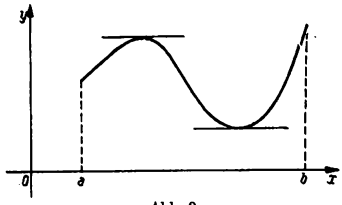


Abb. 9

Geometrisch besagt der Satz von FERMAT folgendes: *Wenn im höchsten bzw. im niedrigsten Punkt des graphischen Bildes der Funktion $f(x)$ die Tangente existiert, so ist sie, sofern die Abszisse dieses Punktes im Innern des Intervalls $[a, b]$ liegt, parallel zur x-Achse* (Abb. 9).

Es sei ausdrücklich darauf hingewiesen, daß auf die Voraussetzung $a < x_0 < b$ im allgemeinen nicht verzichtet werden kann. So wurde z. B. beim Beweis maßgeblich benutzt, daß $f'(x_0)$ sowohl rechtsseitige als auch linksseitige Ableitung ist. Man macht sich übrigens leicht klar, daß in dem Fall, daß $f(x)$ seinen größten bzw. kleinsten Wert im Punkte a oder b annimmt, die Existenz von $f'(x_0)$ (als einseitiger Grenzwert) nicht gewährleistet, daß $f'(x_0)$ gleich Null ist. Als Illustration hierfür kann Abb. 10 dienen, wo die Tangenten in den Punkten A und B keineswegs parallel zur x-Achse sind.

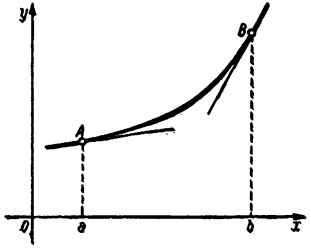


Abb. 10

Mit Hilfe des Satzes von FERMAT beweisen wir nun den folgenden Satz, der einer der fundamentalen Sätze der Analysis ist.

Satz von ROLLE. *Es sei $f(x)$ eine im abgeschlossenen Intervall $[a, b]$ definierte Funktion, welche in allen Punkten dieses Intervalls differenzierbar ist, für welche also in allen Punkten des Intervalls $[a, b]$ die Ableitung $f'(x)$ existiert. Stimmen dann die Werte der Funktion $f(x)$ in den Punkten a und b überein, d. h. ist*

$$f(a) = f(b),$$

so gibt es im Innern des Intervalls $[a, b]$ *wenigstens einen Punkt* c $(a < c < b)$, *in welchem die Ableitung verschwindet, d. h., in welchem*

$$f'(c) = 0$$

gilt.

Nehmen wir speziell an, daß unter den angegebenen Voraussetzungen $f(a) = f(b) = 0$ ist, so besagt der Satz von ROLLE, daß *zwischen je zwei Nullstellen der Funktion* $f(x)$ *stets wenigstens eine Nullstelle der Ableitung* $f'(x)$ *liegt.*

Bevor wir zum eigentlichen Beweis übergehen, bemerken wir, daß die Funktion $f(x)$ in allen Punkten des Intervalls $[a, b]$ stetig ist. Daher gibt es unter den Werten der Funktion $f(x)$ einen größten M und einen kleinsten m (vgl. S. 207). Ist $M = m$, so ist die Funktion $f(x)$ konstant, und es ist in allen Punkten des Intervalls $[a, b]$ dann $f'(x) = 0$, so daß jeder Punkt c des offenen Intervalls (a, b) die verlangte Eigenschaft besitzt. Ist dagegen $M > m$, so können wegen $f(a) = f(b)$ nicht beide Werte an den Enden des Intervalls $[a, b]$ angenommen werden, d. h., wenigstens einer dieser beiden Werte wird in einem Punkt c mit $a < c < b$ angenommen. Dann ist aber nach dem Satz von FERMAT $f'(c) = 0$, so daß dieser Punkt c die verlangte Eigenschaft besitzt.

Der Satz von ROLLE ist ein typischer „Existenzsatz". In ihm wird lediglich behauptet, daß es einen Punkt c mit den angegebenen Eigenschaften gibt, und es ist keine Rede davon, wie man diesen Punkt bestimmen kann.

11. Die Formeln von LAGRANGE und CAUCHY. Die L'HOSPITALsche Regel. Die folgenden beiden Sätze sind sukzessive Verallgemeinerungen des ROLLEschen Satzes.

Satz von LAGRANGE.[1]) *Es sei* $f(x)$ *eine im Intervall* $[a, b]$ *definierte Funktion, welche in allen Punkten dieses Intervalls differenzierbar ist. Dann gibt es im Innern des Intervalls* $[a, b]$ *wenigstens einen Punkt* c $(a < c < b)$, *so daß*

$$\frac{f(b) - f(a)}{b - a} = f'(c). \tag{1}$$

Die Beziehung (1) bezeichnet man auch als Formel von LAGRANGE.

Ist $f(a) = f(b)$, so geht (1) in $f'(c) = 0$ über, so daß der Satz von LAGRANGE eine echte Verallgemeinerung des Satzes von ROLLE ist. In entsprechender Weise ist der folgende Satz eine Verallgemeinerung des Satzes von LAGRANGE:

Satz von CAUCHY.[2]) *Es seien* $f(x)$ *und* $g(x)$ *zwei im Intervall* $[a, b]$ *definierte und dort überall differenzierbare Funktionen, wobei* $g'(x)$ *in keinem Punkte des offenen Intervalls* (a, b) *verschwinde. Dann gibt es im Innern des Intervalls* $[a, b]$ *wenigstens einen Punkt* c $(a < c < b)$, *so daß*

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(c)}{g'(c)}. \tag{2}$$

Die Beziehung (2) bezeichnet man auch als Formel von CAUCHY.

---

[1]) In der deutschen Literatur üblicherweise als Mittelwertsatz der Differentialrechnung bezeichnet. — *Anm. d. wissenschaftl. Red.*

[2]) In der deutschen Literatur üblicherweise als verallgemeinerter Mittelwertsatz der Differentialrechnung bezeichnet. — *Anm. d. wissenschaftl. Red.*

Für $g(x) = x$ geht die CAUCHYsche Formel in die LAGRANGEsche Formel über.

Offenbar genügt es, den Satz von CAUCHY zu beweisen. Dazu betrachten wir die Hilfsfunktion

$$\varphi(x) = [f(b) - f(a)][g(x) - g(a)] - [f(x) - f(a)][g(b) - g(a)].$$

Man sieht leicht ein, daß die Funktion $\varphi(x)$ in allen Punkten des Intervalls $[a, b]$ differenzierbar ist, und zwar ist

$$\varphi'(x) = [f(b) - f(a)]\, g'(x) - f'(x)\,[g(b) - g(a)].$$

Ferner ist, wie unmittelbar aus der Definition der Funktion $\varphi(x)$ hervorgeht, $\varphi(a) = \varphi(b) = 0$. Daher gibt es auf Grund des Satzes von ROLLE im Innern des Intervalls $[a, b]$ einen Punkt c mit $\varphi'(c) = 0$, d. h., in welchem

$$[f(b) - f(a)]\, g'(c) = f'(c)\,[g(b) - g(a)]. \tag{3}$$

Zum Beweis von (2) genügt es also, (3) durch $g'(c)\,[g(b) - g(a)]$ zu dividieren. Hierzu müssen wir uns nur überlegen, daß dieses Produkt von Null verschieden ist. Das ist aber tatsächlich der Fall; denn nach Voraussetzung ist im offenen Intervall zunächst überall $g'(x)$ von Null verschieden; hieraus folgt, daß auch der zweite Faktor $g(b) - g(a)$ von Null verschieden ist, da es anderenfalls auf Grund des Satzes von ROLLE zwischen a und b eine Nullstelle von $g'(x)$ geben müßte. Damit ist der Satz von CAUCHY bewiesen.

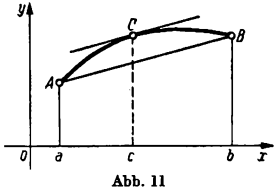


Abb. 11

Da offenbar

$$\frac{f(b) - f(a)}{b - a}$$

der Steigungskoeffizient der Sehne durch die Punkte $A\,(a, f(a))$, $B\,(b, f(b))$ des graphischen Bildes der Funktion $f(x)$ ist, können wir den Satz von LAGRANGE auch in der folgenden geometrischen Form aussprechen:

Besitzt das graphische Bild der Funktion $y = f(x)$ in allen Punkten eine Tangente, so gibt es auf jedem Bogen $\overset{\frown}{AB}$ der Kurve einen Punkt C, in welchem die Tangente parallel der Sehne AB ist (Abb. 11).[1])

Die Formel von LAGRANGE erweist sich als sehr brauchbar zum Auffinden von gewissen Ungleichungen und für angenäherte Berechnungen.

Zwei Beispiele mögen dies erläutern.

[1]) Darin besteht auch der geometrische Inhalt des CAUCHYschen Satzes. Um dies einzusehen, muß man sich die Kurve in Parameterdarstellung $x = f(t)$, $y = g(t)$ gegeben denken.

Beispiel 1. Wenden wir die Formel (1) auf die Funktion x^α an, so finden wir

$$\frac{b^\alpha - a^\alpha}{b-a} = \alpha c^{\alpha-1}, \qquad 0 < a < c < b.$$

Ist nun $\alpha > 1$, so ist $a^{\alpha-1} < c^{\alpha-1} < b^{\alpha-1}$, so daß

$$\alpha a^{\alpha-1} < \frac{b^\alpha - a^\alpha}{b-a} < \alpha b^{\alpha-1}$$

und daher

$$\alpha a^{\alpha-1}(b-a) < b^\alpha - a^\alpha < \alpha b^{\alpha-1}(b-a).$$

Ist dagegen $0 < \alpha < 1$, so ist $a^{\alpha-1} > c^{\alpha-1} > b^{\alpha-1}$, also

$$\alpha b^{\alpha-1} < \frac{b^\alpha - a^\alpha}{b-a} < \alpha a^{\alpha-1}$$

und daher

$$\alpha b^{\alpha-1}(b-a) < b^\alpha - a^\alpha < \alpha a^{\alpha-1}(b-a).$$

Beispiel 2. Mit $a = 1$, $b = 1 + x$ $(x > 0)$ ergibt sich für $f(x) = \ln x$ aus (1)

$$\frac{\ln(1+x) - \ln 1}{(1+x) - 1} = \frac{1}{c}, \quad \text{d. h.} \quad \ln(1+x) = \frac{x}{c},$$

wobei c zwischen 1 und $1 + x$ liegt. Daher ist

$$\frac{x}{1+x} < \ln(1+x) < x.$$

Beachten wir, daß $\lg(1+x) = 0{,}43429\ldots \cdot \ln(1+x)$ $(0{,}43429\ldots = \lg e)$, so erhalten wir aus den vorangehenden Ungleichungen sofort

$$\frac{0{,}43429\ldots \cdot x}{1+x} < \lg(1+x) < 0{,}43429\ldots \cdot x.$$

Wir finden damit also eine Abschätzung für den Wert von $\lg(1+x)$, wobei die Differenz zwischen der oberen und der unteren Schranke gleich $\frac{0{,}43429\ldots x^2}{1+x}$ also kleiner als $0{,}5\,x^2$ ist. Speziell ist für $0 < x < 0{,}1$ mit einem Fehler, der kleiner als 0,005 ist,

$$\lg(1+x) \approx 0{,}43x.$$

Ist $0 < x < 0{,}01$, so ist bis auf einen Fehler, welcher kleiner als 0,00005 ist,

$$\lg(1+x) \approx 0{,}434x.$$

Wir werden im folgenden eine Reihe von wichtigen theoretischen Anwendungen der Sätze von Lagrange und Cauchy kennenlernen. Im vorliegenden Abschnitt wollen wir lediglich noch drei Fragen spezielleren Charakters betrachten.

I. Es sei $y = f(x)$ eine im Intervall $[a, b]$ definierte Funktion, für welche in allen Punkten dieses Intervalls die zweite Ableitung $f''(x)$ existiert. Wir

betrachten dann einen Punkt $x \in [a, b]$ und einen davon verschiedenen Punkt $x + \Delta x$, der gleichfalls noch im Intervall $[a, b]$ liegt. Gefragt ist, welchen Fehler wir begehen, wenn wir den dem Zuwachs Δx entsprechenden Funktionszuwachs Δy durch das Differential $dy = y' \Delta x$ ersetzen.

Auf Grund des Satzes von LAGRANGE ist zunächst

$$\Delta y = f(x + \Delta x) - f(x) = f'(x_1)\, \Delta x,$$

wobei $x < x_1 < x + \Delta x$ (um etwas Bestimmtes vor Augen zu haben, nehmen wir an, daß $\Delta x > 0$). Daher ist wegen $dy = f'(x) \Delta x$

$$\Delta y - dy = [f'(x_1) - f'(x)]\, \Delta x.$$

Wenden wir nochmals den Satz von LAGRANGE an, so erhalten wir:

$$f'(x_1) - f'(x) = f''(x_2)\,(x_1 - x) \qquad (x < x_2 < x_1).$$

Nehmen wir nun an, daß die zweite Ableitung $f''(x)$ auf $[a, b]$ stets zwischen $-M$ und M liegt, und beachten wir, daß $|x_1 - x| < \Delta x$, so gelangen wir zu der Abschätzung

$$|\Delta y - dy| \leqq M (\Delta x)^2,$$

auf die schon in Abschnitt 7 hingewiesen wurde.

II. Es seien $f(x)$ und $g(x)$ zwei in einer gewissen Umgebung eines Punktes a stetige Funktionen, welche beide im Punkte a verschwinden, $f(a) = g(a) = 0$, wobei aber $g(x)$ in allen anderen Punkten dieser Umgebung von Null verschieden ist. Dann ist auch die Funktion

$$\frac{f(x)}{g(x)}$$

in allen Punkten dieser Umgebung mit Ausnahme von a stetig, während sie im Punkte a ihren Sinn verliert. Es ergibt sich dann die Frage, ob die Funktion $\frac{f(x)}{g(x)}$ im Punkte a einen gewissen Grenzwert besitzt, d. h., ob man „den unbestimmten Ausdruck $\frac{0}{0}$ auswerten kann".

Der Satz von CAUCHY gibt nun die Möglichkeit, diese Frage in gewissen Fällen mittels einer allgemeinen Regel zu beantworten, die man als L'HOSPITALsche Regel bezeichnet. Dieser Regel liegt der folgende Satz zugrunde:

Satz. *Erfüllen die Funktionen $f(x)$ und $g(x)$ die Voraussetzungen des Satzes von* CAUCHY *und ist dabei $f(a) = g(a) = 0$, so strebt, falls der Quotient $\frac{f'(x)}{g'(x)}$ für $x \to a$ gegen den endlichen oder unendlichen Grenzwert A konvergiert,*

$$\lim_{x \to a} \frac{f'(x)}{g'(x)} = A,$$

auch der Quotient $\frac{f(x)}{g(x)}$ gegen A,

$$\lim_{x \to a} \frac{f(x)}{g(x)} = A. \tag{4}$$

Beweis. Wenden wir bei festgehaltenem x mit $a < x \leqq b$ den Satz von CAUCHY auf das Intervall $[a, x]$ an, so erhalten wir

$$\frac{f(x)}{g(x)} = \frac{f(x)-f(a)}{g(x)-g(a)} = \frac{f'(x_1)}{g'(x_1)},$$

wobei x_1 zwischen a und x liegt. Hieraus ergibt sich, da mit $x \to a$ auch $x_1 \to a$ und daher

$$\frac{f'(x_1)}{g'(x_1)} \to A,$$

sofort die Beziehung (4).

Hier zwei Beispiele für die Anwendung der L'HOSPITALschen Regel:

1. $$\lim_{x\to 3} \frac{x^3-27}{x^2-4x+3} = \lim_{x\to 3} \frac{3x^2}{2x-4} = \frac{27}{2},$$

2. $$\lim_{x\to 0} \frac{x-\sin x}{x^3} = \lim_{x\to 0} \frac{1-\cos x}{3x^2}.$$

Der im zweiten Beispiel auftretende Ausdruck $\frac{1-\cos x}{3x^2}$ besitzt nun offenbar im Punkte $x = 0$ ebenfalls eine Unbestimmtheit der Form $\frac{0}{0}$. Um diese zu beseitigen, wenden wir nochmals die L'HOSPITALsche Regel an und finden

$$\lim_{x\to 0} \frac{x-\sin x}{x^3} = \lim_{x\to 0} \frac{\sin x}{6x} = \frac{1}{6}.$$

III. Wir wenden uns schließlich noch einmal dem vollständigen Differential zu, das bereits am Ende von Abschnitt 9 betrachtet wurde, und beweisen die dort ohne Beweis mitgeteilte Formel (22).

Es sei $z = f(x, y)$ eine in dem durch die Ungleichungen $a < x < b$, $c < y < d$ beschriebenen Rechteck Q definierte Funktion, für welche in allen Punkten dieses Rechtecks die partiellen Ableitungen $f_x(x, y)$ und $f_y(x, y)$ existieren, wobei diese partiellen Ableitungen selbst stetige Funktionen seien. Wir betrachten dann einen festen Punkt (x, y) aus diesem Rechteck und daneben einen zweiten Punkt $(x + \Delta x,\ y + \Delta y)$, der ebenfalls noch in Q liege. Zu untersuchen ist der Zuwachs

$$\Delta z = f(x + \Delta x,\ y + \Delta y) - f(x,\ y)$$

der Funktion $z = f(x, y)$ beim Übergang vom Punkte (x, y) zum Punkte $(x + \Delta x, y + \Delta y)$. Offenbar kann dieser Zuwachs auch in der Form

$$\Delta z = [f(x + \Delta x,\ y + \Delta y) - f(x,\ y + \Delta y)] + [f(x,\ y + \Delta y) - f(x,\ y)]$$

geschrieben werden. Dabei ist $f(x, y + \Delta y) - f(x, y)$ der Zuwachs von $f(x, y)$ als Funktion der einen Veränderlichen y bei festgehaltenem x. Daher können wir auf diese Differenz die Formel von LAGRANGE anwenden:

$$f(x,\ y + \Delta y) - f(x,\ y) = f_y(x,\ \eta)\, \Delta y.$$

Dabei ist η eine zwischen y und $y + \Delta y$ liegende Zahl.

Entsprechend ergibt sich

$$f(x + \Delta x,\ y + \Delta y) - f(x,\ y + \Delta y) = f_x(\xi,\ y + \Delta y)\, \Delta x,$$

wobei ξ zwischen x und $x+\Delta x$ liegt. Wegen der Stetigkeit der partiellen Ableitungen streben nun mit Δx und Δy auch

$$\alpha = f_x(\xi,\ y+\Delta y) - f_x(x,\ y), \quad \beta = f_y(x,\ \eta) - f_y(x,\ y)$$

gegen Null.[1]) Setzen wir

$$\alpha\Delta x + \beta\Delta y = \varrho.$$

so ist also wegen

$$\frac{|\varrho|}{|\Delta x| + |\Delta y|} \leqq |\alpha| + |\beta|$$

auch

$$\lim \frac{\varrho}{|\Delta x| + |\Delta y|} = 0.$$

Außerdem gilt, wie man sofort nachrechnet,

$$\Delta z = f_x(x,\ y)\,\Delta x + f_y(x,\ y)\,\Delta y + \varrho.$$

Das war aber gerade zu zeigen.[2])

12. Die Taylorsche Formel. Gemäß dem allgemeinen Funktionsbegriff ist eine Funktion eine Vorschrift, durch welche jedem Argumentwert x ein eindeutig bestimmter Funktionswert y zugeordnet wird. Dabei ist keine Rede davon, mit welchen Mitteln man etwa den einem Argumentwert x entsprechenden Funktionswert y bestimmen kann. Vielfach ist natürlich eine Funktion durch eine Formel gegeben, der man entnehmen kann, welche Rechenoperationen man anwenden muß, um aus einem gegebenen Wert x den zugehörigen Wert y zu erhalten (zum Beispiel $y = x^2$). In anderen Fällen ist dies dagegen keineswegs der Fall. Wird die betrachtete Funktion z. B. durch die Gleichung

$$y = \sin x$$

gegeben, so kann man zunächst den einem Wert x entsprechenden Funktionswert y nicht mit Hilfe einer Rechenmethode (oder, wie man häufiger sagt, mittels eines analytischen Verfahrens) bestimmen. Es entsteht damit die wichtige Aufgabe, analytische Hilfsmittel zu schaffen, mit deren Hilfe dies doch gelingt. In vielen Fällen führt hierbei eine bemerkenswerte Formel, die sogenannte Taylorsche Formel, zum Ziel.

Wir wollen diese Formel zunächst am Beispiel eines Polynoms

$$f(x) = c_0 + c_1 x + c_2 x^2 + \cdots + c_n x^n \tag{5}$$

erläutern.

Hierzu beachten wir zunächst, daß wir das Polynom (5) bei beliebig vorgegebener Zahl a in der Form

$$f(x) = A_0 + A_1(x-a) + A_2(x-a)^2 + \cdots + A_n(x-a)^n \tag{6}$$

---

[1]) Man beachte, daß für $\Delta x \to 0$, $\Delta y \to 0$ offenbar $\xi \to x$, $\eta \to y$.

[2]) Wie dem Beweis zu entnehmen ist, braucht man die Existenz der partiellen Ableitungen in allen Punkten von Q, während ihre Stetigkeit nur im Punkte (x, y) benötigt wird.

darstellen können, wobei die Koeffizienten $A_0, A_1, \ldots, A_n$ nicht von x abhängen, d. h. mit anderen Worten, daß wir $f(x)$ nach Potenzen von $(x-a)$ entwickeln können. Offenbar genügt es, dies für jede Funktion x^k $(k=1, 2, \ldots)$ zu beweisen, da es sich von diesen Funktionen unmittelbar auf beliebige Polynome überträgt. Für x^k ergibt sich jedoch unsere Behauptung sofort auf Grund der bekannten Formel

$$x^k=[a+(x-a)]^k=\sum_{i=0}^{k}\binom{k}{i}a^i(x-a)^{k-i}.$$

Unsere Aufgabe soll es nun sein, die Koeffizienten $A_0, A_1, \ldots, A_n$ effektiv zu berechnen. Den Koeffizienten A_0 erhält man sofort, wenn man in (6) $x=a$ setzt:

$$A_0=f(a).$$

Um A_1 zu bestimmen, bilden wir die Ableitung von (6):

$$f'(x)=A_1+2A_2(x-a)+3A_3(x-a)^2+\cdots+nA_n(x-a)^{n-1}. \tag{7}$$

Setzen wir hier $x=a$, so erhalten wir

$$A_1=f'(a).$$

Durch Differentiation von (7) ergibt sich

$$f''(x)=2A_2+3\cdot 2A_3(x-a)+\cdots+n(n-1)A_n(x-a)^{n-2}$$

und hieraus für $x=a$

$$A_2=\frac{f''(a)}{1\cdot 2}.$$

Entsprechend findet man allgemein

$$A_k=\frac{f^{(k)}(a)}{k!}.$$

Damit nimmt (6) die Form

$$f(x)=f(a)+\frac{f'(a)}{1!}(x-a)+\frac{f''(a)}{2!}(x-a)^2+\cdots+\frac{f^{(n)}(a)}{n!}(x-a)^n \tag{8}$$

an; *dies ist bereits die* Taylorsche *Formel für Polynome.*

Wir betrachten jetzt allgemein eine beliebige Funktion $f(x)$, von der wir jedoch voraussetzen, daß für sie in einem gegebenen Punkt a die Ableitungen $f'(a), f''(a), \ldots, f^{(n)}(a)$ existieren. Dann können wir das Polynom

$$T(x)=f(a)+\frac{f'(a)}{1!}(x-a)+\frac{f''(a)}{2!}(x-a)^2+\cdots+\frac{f^{(n)}(a)}{n!}(x-a)^n \tag{9}$$

bilden.

Natürlich stimmen im allgemeinen $f(x)$ und $T(x)$ nicht überein. Jedoch zeigt es sich, daß in vielen Fällen die Differenz zwischen $f(x)$ und $T(x)$ sehr klein wird, so daß man in guter Näherung $f(x)$ durch $T(x)$ ersetzen kann. Darüber hinaus zeigt sich, daß, falls die Funktion $f(x)$ im Punkte a Ableitungen beliebig hoher Ordnung besitzt und somit n beliebig groß gewählt werden kann, in vielen Fällen die Differenz $f(x)-T(x)$ für $n\to\infty$ gegen

Null strebt, so daß man dann $f(x)$ sogar mit beliebiger Genauigkeit berechnen kann.

Bevor wir uns einer genaueren Untersuchung dieser Fragen zuwenden, nehmen wir an, daß $f(x)$ in einem gewissen, den Punkt a enthaltenden Intervall $[A, B]$ definiert ist und dort überall Ableitungen bis zur $(n+1)$-ten Ordnung einschließlich besitzt.[1]) Betrachten wir dann das gemäß (9) definierte Polynom $T(x)$, so können wir feststellen, daß

$$T'(x) = f'(a) + \frac{f''(a)}{1!}(x-a) + \cdots + \frac{f^{(n)}(a)}{(n-1)!}(x-a)^{n-1},$$

$$T''(x) = f''(a) + \frac{f'''(a)}{1!}(x-a) + \cdots + \frac{f^{(n)}(a)}{(n-2)!}(x-a)^{n-2},$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$T^{(n)}(x) = f^{(n)}(a),$$

$$T^{(n+1)}(x) = 0.$$

Hieraus und aus (9) folgt

$$\left.\begin{aligned} &T(a) = f(a), \quad T'(a) = f'(a), \\ &T''(a) = f''(a), \ \ldots, \quad T^{(n)}(a) = f^{(n)}(a), \quad T^{(n+1)}(x) = 0. \end{aligned}\right\} \tag{10}$$

Wir betrachten nun die folgendermaßen definierte Hilfsfunktion $\varphi(x)$:

$$\varphi(x) = f(x) - T(x).$$

Aus (10) ergibt sich

$$\varphi(a) = \varphi'(a) = \varphi''(a) = \cdots = \varphi^{(n)}(a) = 0, \quad \varphi^{(n+1)}(x) = f^{(n+1)}(x).$$

Schließlich betrachten wir noch die folgende Hilfsfunktion $\psi(x)$:

$$\psi(x) = \frac{(x-a)^{n+1}}{(n+1)!}.$$

Offenbar ist

$$\psi(a) = \psi'(a) = \psi''(a) = \cdots = \psi^{(n)}(a) = 0, \quad \psi^{(n+1)}(x) = 1.$$

Wir weisen darauf hin, daß weder die Funktion $\psi(x)$ noch eine ihrer Ableitungen (bis zur $(n+1)$-ten einschließlich) in einem von a verschiedenen Punkt des Intervalls $[A, B]$ verschwindet.

Es sei nun x ein beliebiger, aber für das folgende fester Punkt aus dem Intervall $[A, B]$ mit $x \neq a$. Wegen $\varphi(a) = \psi(a) = 0$ ist

$$\frac{\varphi(x)}{\psi(x)} = \frac{\varphi(x) - \varphi(a)}{\psi(x) - \psi(a)}.$$

Wenden wir hierauf den CAUCHYschen Satz an, so erhalten wir

$$\frac{\varphi(x)}{\psi(x)} = \frac{\varphi'(x_1)}{\psi'(x_1)},$$

[1]) Ist a einer der Endpunkte A oder B, so sind als $f'(a)$, $f''(a)$, ... sinngemäß die einseitigen Ableitungen zu nehmen.

wobei x_1 zwischen a und x liegt. Nun ist aber auch $\varphi'(a) = \psi'(a) = 0$ und daher

$$\frac{\varphi'(x_1)}{\psi'(x_1)} = \frac{\varphi'(x_1) - \varphi'(a)}{\psi'(x_1) - \psi'(a)}.$$

Wenden wir hierauf die CAUCHYsche Formel an, so erhalten wir

$$\frac{\varphi(x)}{\psi(x)} = \frac{\varphi''(x_2)}{\psi''(x_2)},$$

wobei x_2 zwischen a und x_1 und daher erst recht zwischen a und x liegt. Indem wir dieses Verfahren hinreichend weit fortsetzen, erhalten wir

$$\frac{\varphi(x)}{\psi(x)} = \frac{\varphi^{(n+1)}(x_{n+1})}{\psi^{(n+1)}(x_{n+1})},$$

wobei x_{n+1} zwischen a und x liegt. Berücksichtigen wir nun, daß

$$\varphi^{(n+1)}(x) = f^{(n+1)}(x), \quad \psi^{(n+1)}(x) = 1,$$

so erhalten wir, wenn wir x_{n+1} kurz mit $\bar{x}$ bezeichnen,

$$\frac{\varphi(x)}{\psi(x)} = f^{(n+1)}(\bar{x})$$

oder, ausführlich,

$$f(x) - T(x) = \frac{f^{(n+1)}(\bar{x})}{(n+1)!}(x-a)^{n+1}.$$

Setzen wir schließlich noch den Ausdruck (9) für $T(x)$ ein, so gelangen wir zu der folgenden TAYLORschen Formel mit LAGRANGEschem Restglied:

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \frac{f^{(n+1)}(\bar{x})}{(n+1)!}(x-a)^{n+1}, \quad (11)$$

wobei $\bar{x}$ zwischen a und x liegt.

Ersetzt man also die Funktion $f(x)$ durch das Polynom $T(x)$, so begeht man den Fehler

$$R_n(x) = \frac{f^{(n+1)}(\bar{x})}{(n+1)!}(x-a)^{n+1}.$$

Es ergibt sich nun die Frage, wie sich dieser Fehler bei Änderung von n verhält.

Satz. *Besitzt die Funktion $f(x)$ in allen Punkten des Intervalls $[A, B]$ Ableitungen beliebig hoher Ordnung und gibt es eine Zahl K derart, daß für alle $x \in [A, B]$ und alle natürlichen Zahlen n die Ungleichung*

$$|f^{(n)}(x)| \leqq K$$

erfüllt ist, so gilt für alle Punkte $x \in [A, B]$

$$\lim_{n \to \infty} R_n(x) = 0.$$

Beweis. Unter den angegebenen Voraussetzungen ist offenbar

$$|R_n(x)| \leqq K \frac{M^{n+1}}{(n+1)!},$$

wobei zur Abkürzung $|x-a| = M$ gesetzt ist. Wir betrachten nun eine natürliche Zahl m mit

$$m+1 > 2M.$$

Dann ist für alle $n > m$

$$\frac{M^{n+1}}{(n+1)!} = \frac{M^m}{m!} \cdot \frac{M}{m+1} \cdot \frac{M}{m+2} \cdots \frac{M}{n+1}.$$

Dabei ist jeder der $n-m+1$ Brüche

$$\frac{M}{m+1}, \quad \frac{M}{m+2}, \quad \cdots, \quad \frac{M}{n+1}$$

kleiner als $\frac{1}{2}$, so daß

$$|R_n(x)| \leqq K \frac{M^m}{m!} \cdot \frac{1}{2^{n-m+1}},$$

das heißt

$$|R_n(x)| \leqq 2^{m-1} K \frac{M^m}{m!} \cdot \frac{1}{2^n}.$$

Bezeichnen wir nun die konstante Zahl (x sollte ja festgehalten werden!)

$$2^{m-1} K \frac{M^m}{m!}$$

mit C, so gilt also

$$|R_n(x)| \leqq \frac{C}{2^n},$$

woraus sich sofort die angegebene Behauptung ergibt.

Unter den Voraussetzungen des zuletzt angegebenen Satzes ist also

$$f(x) = \lim_{n \to \infty} \left[f(a) + \frac{f'(a)}{1!}(x-a) + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n \right],$$

wodurch die am Anfang des vorliegenden Abschnitts durchgeführten Überlegungen gerechtfertigt sind.

Wir betrachten im Anschluß an diese allgemeinen Ausführungen einige konkrete Beispiele.

Beispiel 1. Da die Funktion $f(x) = \sin x$ die Voraussetzungen des zuletzt bewiesenen Satzes erfüllt und für sie

$$f'(x) = \cos x, \quad f''(x) = -\sin x, \quad f'''(x) = -\cos x, \quad f^{(4)}(x) = \sin x, \quad f^{(5)}(x) = \cos x$$

und allgemein

$$f^{(n)}(x) = \sin\left(x + n\frac{\pi}{2}\right)$$

ist, erhalten wir, wenn wir $a = 0$ setzen, die Näherungsgleichung

$$\sin x \approx x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots + (-1)^p \frac{x^{2p+1}}{(2p+1)!},$$

deren Genauigkeit mit wachsendem p unbeschränkt verbessert werden kann. Der Fehler übersteigt dabei nicht die Zahl

$$\frac{x^{2p+2}}{(2p+2)!}.$$

Setzt man also z. B.

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!}, \tag{12}$$

so ist der Fehler nicht größer als

$$\frac{x^8}{8!} = \frac{x^8}{40320}.$$

Beschränkt man x auf das Intervall $0 < x \leqq \frac{\pi}{4}$, so ist bei Berechnung von $\sin x$ gemäß Formel (12) der Fehler durchweg kleiner als 0,000004.[1])

Beispiel 2. Entsprechend findet man

$$\cos x \approx 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots + (-1)^p \frac{x^{2p}}{(2p)!}.$$

Aus den angegebenen Beispielen ist zu ersehen, daß die Taylorsche Formel ein Mittel in die Hand gibt, die trigonometrischen Funktionen zu berechnen. Wir werden in dem Kapitel, welches der Theorie der Reihen gewidmet ist, näher darauf zurückkommen.

Beispiel 3. Wir betrachten bei gegebenem $c > 1$ die Funktion $f(x) = x^c$. Für $x > 1$, $a = 1$ und $n = 1$ ergibt Formel (11)

$$x^c = 1 + c(x-1) + \frac{c(c-1)}{1 \cdot 2} \bar{x}^{c-2}(x-1)^2, \quad 1 < \bar{x} < x$$

(es ist nämlich $(x^c)' = c x^{c-1}$, $(x^c)'' = c(c-1)x^{c-2}$). Wegen $\bar{x}^{c-2} = \frac{\bar{x}^c}{\bar{x}^2} > \frac{1}{x^2}$ ist also insbesondere

$$x^c > 1 + c(x-1) + \frac{c(c-1)}{2}\left(\frac{x-1}{x}\right)^2 \qquad (x > 1).$$

Hieraus erhält man speziell die etwas schwächere, aber einfachere Ungleichung

$$x^c - 1 > c(x-1) \qquad (x > 1).$$

Wir weisen darauf hin, daß dies alles für beliebiges (nicht notwendig ganzes) $c > 1$ gilt.

---

[1]) Dies ergibt sich folgendermaßen: Es ist

$$\lg \pi = 0{,}49715, \quad \lg 4 = 0{,}60206, \quad \lg \frac{\pi}{4} = \bar{1}{,}89509, \quad \lg \left(\frac{\pi}{4}\right)^8 = \bar{1}{,}16072,$$

also $\left(\frac{\pi}{4}\right)^8 < 0{,}15$ und

$$\frac{\left(\frac{\pi}{4}\right)^8}{40320} < \frac{1}{268800} < 0{,}000004.$$

Beispiel 4. Auf Seite 187 ist bereits streng bewiesen worden, daß $0 < e < 3$. Wir wollen jetzt zeigen, wie man die Konstante e genau berechnen kann.

Dazu wenden wir die TAYLORsche Formel auf die Funktion $f(x) = e^x$ an, wobei wir $a = 0$ setzen. Wegen $f^{(n)}(x) = e^x$ ist $f^{(n)}(0) = 1$ und daher

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + e^{\bar{x}} \frac{x^{n+1}}{(n+1)!},$$

wobei $\bar{x}$ ein gewisser Wert zwischen 0 und x ist. Insbesondere erhalten wir für $x = 1$

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} + \frac{e^{\bar{x}}}{(n+1)!} \qquad (0 < \bar{x} < 1). \tag{13}$$

Wählen wir hier $n = 8$, so finden wir wegen

$$\frac{e^{\bar{x}}}{9!} < \frac{e}{9!} < \frac{3}{9!} = \frac{3}{362880} < 0{,}00001,$$

daß mit einem Fehler, welcher kleiner als 0,00001 ist, die Zahl e durch

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \frac{1}{5!} + \frac{1}{6!} + \frac{1}{7!} + \frac{1}{8!} = 2{,}718275$$

gegeben wird, so daß

$$2{,}718275 < e < 2{,}718285.$$

Eine genauere Rechnung ergibt

$$e = 2{,}718281828459 \ldots .$$

Mit Hilfe von Formel (13) kann man leicht zeigen, daß e *irrational* ist.[1]) Angenommen nämlich, dies wäre nicht der Fall, d. h., es wäre $e = \frac{p}{q}$. Dann wählen wir in (13) die Zahl n größer als q und 2 und multiplizieren die entstehende Gleichung mit $n!$:

$$\frac{p\,n!}{q} = \left[1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!}\right] \cdot n! + \frac{e^{\bar{x}}}{n+1}.$$

Hiernach muß $\frac{e^{\bar{x}}}{n+1}$ eine ganze Zahl sein. Das ist aber nicht der Fall, denn wegen $0 < \bar{x} < 1$ ist $1 < e^{\bar{x}} < e < 3$, so daß $\frac{e^{\bar{x}}}{n+1}$ zwischen $\frac{1}{n+1}$ und $\frac{3}{n+1}$ liegt. Diese Brüche liegen aber (wegen $n > 2$) zwischen 0 und 1.

13. Die Untersuchungen von P. L. TSCHEBYSCHEW und S. N. BERNSTEIN. Bei der Behandlung der Frage nach der Güte der Approximation einer gegebenen Funktion $f(x)$ durch das TAYLORsche Polynom

$$T(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n$$

[1]) Mit Hilfe stärkerer Methoden kann man sogar beweisen, daß e transzendent ist, d. h. keiner algebraischen Gleichung mit ganzen Koeffizienten genügt.

haben wir im vorangehenden Abschnitt den Punkt x festgehalten und den Grad n laufend vergrößert. Es zeigte sich, daß man dabei in vielen Fällen zu einer beliebig genauen Approximation des Funktionswertes $f(x)$ gelangt. Hält man dagegen den Grad n fest, so liegt im allgemeinen keineswegs für alle x aus $[A, B]$ der Wert von $T(x)$ nahe bei dem entsprechenden Wert von $f(x)$. Nur für hinreichend dicht bei a gelegene Argumentwerte wird man ein gutes Übereinstimmen von $T(x)$ und $f(x)$ erwarten können, weil nämlich die stetigen Funktionen $T(x)$ und $f(x)$ im Punkte a denselben Wert annehmen. Das Polynom $T(x)$ wird also bei festem n nur eine lokale Approximation von $f(x)$ liefern. Eine Verbesserung der Güte der Approximation durch Vergrößern von n erweist sich dabei in praktischer Hinsicht nicht als sehr geeignet, da das Arbeiten mit Polynomen höheren Grades mit recht umfangreichen Rechnungen verbunden ist.

Diese Erwägungen führten den großen russischen Mathematiker P. L. Tschebyschew auf das Problem, Polynome zu konstruieren, deren Grad von vornherein beschränkt ist und die dennoch im Gegensatz zu den Taylorschen Polynomen im ganzen Intervall $[A, B]$ eine gute Approximation der Funktion $f(x)$ geben. Genauer formuliert, handelt es sich bei dem Problem von Tschebyschew um folgendes: Unter allen Polynomen, deren Grad höchstens n ist, ist dasjenige auszusuchen, welches von einer gegebenen stetigen Funktion $f(x)$ möglichst wenig abweicht, d. h., für welches

$$\max_{A \leqq x \leqq B} |f(x) - P(x)| \tag{14}$$

einen möglichst kleinen Wert annimmt. Es zeigt sich, daß *es stets ein und nur ein derartiges Polynom gibt*, welches man das *Polynom kleinster Abweichung* nennt. Die Größe (14) selbst wird die *minimale Abweichung der Polynome höchstens n-ten Grades von* $f(x)$ genannt und meistens mit $E_n(f)$ bezeichnet. Tschebyschew selbst untersuchte bereits eingehend die Eigenschaften der Polynome kleinster Abweichung und gab eine ganze Reihe von Anwendungen seiner Theorie, unter anderem in der Theorie der Maschinen und Mechanismen sowie in der Kartographie.

Die bemerkenswerten Untersuchungen Tschebyschews bildeten den Ausgangspunkt für viele bedeutende Arbeiten seiner Schüler (K. A. Posé, A. N. Korkin, E. I. Solotarjow, A. A. Markoff, W. A. Markow u. a.).

Wesentlich weitergeführt wurden die Gedanken Tschebyschews durch Untersuchungen von S. N. Bernstein. Dieser ging von der Tatsache aus, daß die minimale Abweichung $E_n(f)$ bei unbeschränktem Wachsen von n gegen Null strebt. Natürlich hängt die Art, in der $E_n(f)$ klein wird, von den Eigenschaften der Funktion $f(x)$ ab; je einfacher diese Funktion ist, um so genauer läßt sie sich durch ein Polynom ersetzen. Diese allgemeinen Überlegungen führten Bernstein zu einer konstruktiven Klassifikation der stetigen Funktionen. In neuerer Zeit sind die mit diesem Ideenkreis zusammenhängenden Fragen sehr eingehend bearbeitet worden. Sie bilden heute ein wichtiges Gebiet der Analysis, die sogenannte konstruktive Funktionentheorie, das für die Anwendungen von großer Bedeutung ist. Dabei spielen

die Untersuchungen sowjetischer Mathematiker eine erstrangige Rolle. Neben BERNSTEIN haben sich um dieses Gebiet A. N. KOLMOGOROFF, W. L. GONTSCHAROW, M. G. KREIN, N. I. ACHIESER, E. J. REMES, S. M. LOSINSKI, S. M. NIKOLSKI und andere große Verdienste erworben.

§ 3. Anwendung der Differentialrechnung auf das Studium der Eigenschaften von Funktionen

14. Kriterien für die Konstanz und Monotonie von Funktionen. Im vorliegenden Abschnitt wollen wir feststellen, welche Auswirkungen gewisse Eigenschaften der Ableitung einer Funktion auf die Funktion selbst haben.

Satz 1. *Eine in einem Intervall $[a, b]$ erklärte Funktion $f(x)$ ist dann und nur dann konstant, wenn in allen Punkten des Intervalls $[a, b]$ die Ableitung $f'(x)$ existiert und gleich Null ist.*

Wir haben bereits oben gesehen, daß diese Bedingung notwendig ist. Zum Beweis dafür, daß sie auch hinreichend ist, sei $f(x)$ eine beliebige Funktion, für welche in jedem Punkt x des Intervalls die Ableitung $f'(x)$ existiert und gleich Null ist. Wenden wir dann auf das Intervall $[a, x]$ den Satz von LAGRANGE an, so erhalten wir

$$f(x) - f(a) = f'(\bar{x})\,(x - a) \qquad (a < \bar{x} < x),$$

woraus sich wegen $f'(\bar{x}) = 0$ sofort $f(x) = f(a)$ ergibt, und das besagt gerade, daß die Funktion $f(x)$ im Intervall $[a, b]$ konstant ist.

Bevor wir zur Formulierung und zum Beweis der nächsten Sätze übergehen, erinnern wir daran, daß man eine Funktion $f(x)$ *nichtfallend* (*nichtwachsend*) nennt, wenn aus $x < y$ stets $f(x) \leqq f(y)$ $\left(f(x) \geqq f(y)\right)$ folgt; hat die Beziehung $x < y$ stets die Beziehung $f(x) < f(y)$ $\left(f(x) > f(y)\right)$ zur Folge, so nennt man $f(x)$ (*echt*) *wachsend* [(*echt*) *fallend*]. Die nichtfallenden und die nichtwachsenden Funktionen bezeichnet man auch zusammenfassend als *monotone Funktionen*. Da eine Funktion $f(x)$ genau dann wächst, wenn die Funktion $-f(x)$ fällt, können wir uns beim Beweis der folgenden Sätze stets auf wachsende Funktionen beschränken.

Satz 2. *Es sei $f(x)$ eine im Intervall $[a, b]$ erklärte Funktion, für welche in allen Punkten dieses Intervalls die Ableitung $f'(x)$ existiert. Die Funktion $f(x)$ ist dann und nur dann nichtfallend (nichtwachsend), wenn in allen Punkten des offenen Intervalls (a, b) die Ungleichung*

$$f'(x) \geqq 0 \qquad (f'(x) \leqq 0)$$

erfüllt ist.

Beweis. Wir nehmen zunächst an, daß die Funktion $f(x)$ nichtfallend ist. Es sei ferner x ein beliebiger Punkt aus dem Intervall (a, b). Wir geben dann dem Wert x einen Zuwachs $\Delta x > 0$, wobei noch $x + \Delta x \in [a, b]$. Wegen $f(x + \Delta x) \geqq f(x)$ ist

$$\frac{f(x + \Delta x) - f(x)}{\Delta x} \geqq 0.$$

Vollziehen wir in dieser Ungleichung den Grenzübergang $\Delta x \to 0$, so erhalten wir $f'(x) \geqq 0$, womit gezeigt ist, daß die angegebene Bedingung notwendig ist.

Es sei nun umgekehrt $f'(x) \geqq 0$ für alle Punkte $x \in (a, b)$. Es seien ferner x, y zwei Punkte aus dem Intervall $[a, b]$ mit $x < y$. Wenden wir den Satz von LAGRANGE auf das Intervall $[x, y]$ an, so erhalten wir

$$f(y) - f(x) = f'(z)\ (y - x).$$

Dabei ist $x < z < y$, also $z \in (a, b)$, so daß $f'(z) \geqq 0$ und daher

$$f(y) \geqq f(x),$$

d. h., die Funktion $f(x)$ ist nichtfallend.

Satz 3. *Es sei $f(x)$ eine im abgeschlossenen Intervall $[a, b]$ definierte Funktion, für welche in allen Punkten dieses Intervalls die Ableitung $f'(x)$ existiert. Die Funktion $f(x)$ ist dann und nur dann echt wachsend (echt fallend), wenn folgende Bedingungen erfüllt sind*: 1. *In allen Punkten des offenen Intervalls (a, b) ist*

$$f'(x) \geqq 0 \qquad (f'(x) \leqq 0).$$

2. *In keinem Teilintervall $[p, q]$ des Intervalls $[a, b]$ ist überall $f'(x) = 0$.*

Beweis. Ist die Funktion $f(x)$ echt wachsend, so ist sie erst recht nichtfallend, so daß auf Grund des vorangehenden Satzes sicher die Bedingung 1. erfüllt ist. Aber auch die Bedingung 2. ist erfüllt; denn gäbe es ein Teilintervall $[p, q]$, in welchem überall $f'(x) = 0$ ist, so wäre auf Grund von Satz 1 in diesem Intervall die Funktion $f(x)$ konstant, also $f(x)$ nicht echt wachsend, im Widerspruch zur Voraussetzung. Damit ist gezeigt, daß die angegebenen Bedingungen notwendig sind.

Wir nehmen nun umgekehrt an, daß 1. und 2. erfüllt sind. Dann ist zunächst auf Grund des vorangehenden Satzes die Funktion $f(x)$ nichtfallend. Wir betrachten nun zwei beliebige Punkte x, y aus dem Intervall $[a, b]$, wobei $x < y$ sei. Dann ist für alle Punkte z aus dem Intervall $[x, y]$ offenbar $f(x) \leqq f(z) \leqq f(y)$. Wäre nun $f(x) = f(y)$, so wäre also die Funktion $f(x)$ im Intervall $[x, y]$ konstant, und es würde in allen Punkten dieses Intervalls die Ableitung verschwinden, im Widerspruch zu 2. Es ist also $f(x) \neq f(y)$, und zwar $f(x) < f(y)$, womit Satz 3 vollständig bewiesen ist.

Die vorangehenden Ergebnisse werden vollständig plausibel, wenn wir uns vergegenwärtigen, daß der Wert der Ableitung nichts anderes ist als der Steigungskoeffizient der Tangente in demjenigen Punkt des graphischen Bildes der Funktion, welcher den Differentiationspunkt als Abszisse besitzt. Wenn nun die Funktion $f(x)$ wächst, so geht das graphische Bild der Funktion, falls man seinen Verlauf von links nach rechts verfolgt, nach oben (Abb. 12a), während es für eine fallende Funktion (Abb. 12b) nach unten geht. Daher ist im ersten Fall der Winkel zwischen der Tangente und der positiven Richtung der x-Achse spitz und im zweiten Fall stumpf. Dabei kann jedoch in einzelnen

Punkten die Tangente auch parallel zur x-Achse verlaufen (Abb. 12c). Analytisch bedeutet dies, daß auch bei einer wachsenden (fallenden) Funktion die Ableitung in einzelnen Punkten verschwinden kann.[1])

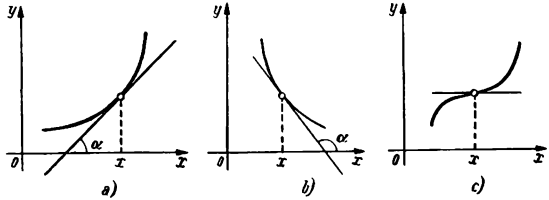


Abb. 12

Beispiel 1. Es sei

$$\varphi(x) = \frac{\sin x}{x}.$$

Da die Funktion $\varphi(x)$ im Punkte $x = 0$ ihren Sinn verliert, beschränken wir uns auf ein Intervall $\left[\delta, \frac{\pi}{2}\right]$ mit $0 < \delta < \frac{\pi}{2}$. In jedem Punkt eines solchen Intervalls existiert die Ableitung $\varphi'(x)$, und zwar ist

$$\varphi'(x) = \frac{x\cos x - \sin x}{x^2} = \frac{x - \operatorname{tg} x}{x^2}\cos x.$$

Nun ist aber für $0 < x < \frac{\pi}{2}$ bekanntlich $x < \operatorname{tg} x$, so daß im offenen Intervall $\left(\delta, \frac{\pi}{2}\right)$ überall $\varphi'(x) < 0$ ist und daher im Intervall $\left[\delta, \frac{\pi}{2}\right]$ die Funktion $\varphi(x)$ fällt. Es ist also $\varphi(x) \geqq \varphi\left(\frac{\pi}{2}\right) = \frac{2}{\pi}$ und daher

$$\sin x \geqq \frac{2}{\pi} x.$$

Diese Ungleichung gilt zunächst für alle x mit $\delta \leqq x \leqq \frac{\pi}{2}$, wobei δ eine beliebige positive reelle Zahl ist. Da hierbei δ beliebig klein gewählt werden kann, gilt sie also für $0 < x \leqq \frac{\pi}{2}$. Da sie schließlich für $x = 0$ trivial erfüllt ist, gilt sie für alle x aus dem Intervall $\left[0, \frac{\pi}{2}\right]$.

Beispiel 2. Wir wollen als nächstes zeigen, daß für $p > 1$ und jedes positive a und b die Ungleichung

$$(a+b)^p \leqq 2^{p-1}(a^p + b^p)$$

gilt. Ohne Beschränkung der Allgemeinheit können wir dabei annehmen, daß $a \leqq b$, so daß $0 < \frac{a}{b} \leqq 1$. Dividieren wir beide Seiten der vorgelegten

[1]) Zum Beispiel ist für die Funktion $f(x) = x^3$ offenbar $f'(0) = 0$, obwohl die Funktion $f(x)$ wächst.

Ungleichung durch b^p und ersetzen wir $\frac{a}{b}$ durch x, so nimmt die zu beweisende Ungleichung die Form

$$(1+x)^p \leqq 2^{p-1}(1+x^p)$$

oder

$$2^{p-1}(1+x^p)-(1+x)^p \geqq 0$$

an. Zum Beweis hierfür setzen wir

$$2^{p-1}(1+x^p)-(1+x)^p = f(x).$$

Dann ist offenbar

$$f'(x) = 2^{p-1}px^{p-1} - p(1+x)^{p-1} = p[(2x)^{p-1} - (1+x)^{p-1}],$$

so daß für $0 < x < 1$ und $p > 1$ die Ungleichung $f'(x) < 0$ erfüllt ist (denn für die angegebenen Werte ist $2x < 1 + x$ und daher $(2x)^{p-1} < (1+x)^{p-1}$). Daraus folgt, daß die Funktion $f(x)$ im Intervall $[0, 1]$ fällt, so daß

$$f(x) > f(1) \quad \text{für} \quad 0 < x < 1.$$

Nun ist aber $f(1) = 2^{p-1} \cdot 2 - 2^p = 0$ und daher

$$f(x) > 0 \quad \text{für} \quad 0 < x < 1;$$

ist $x = 1$, so geht diese Ungleichung in eine Gleichung über. Setzen wir für $f(x)$ die wirkliche Bedeutung ein und setzen wir anschließend $x = \frac{a}{b}$, so erhalten wir der Reihe nach

$$2^{p-1}(1+x^p) - (1+x)^p \geqq 0 \quad \text{und} \quad (a+b)^p \leqq 2^{p-1}(a^p + b^p).$$

Da die zuletzt erhaltene Ungleichung symmetrisch in bezug auf a und b ist, gelangen wir im Fall $b \leqq a$ zu demselben Resultat.

Wir überlassen es dem Leser, im Fall $p \leqq 1$ entsprechend die Ungleichung

$$(a+b)^p \leqq a^p + b^p$$

herzuleiten (aus der sich speziell $\sqrt[n]{a+b} \leqq \sqrt[n]{a} + \sqrt[n]{b}$ für beliebiges natürliches n und $a > 0$, $b > 0$ ergibt).

Beispiel 3. Wir beweisen folgenden Satz von HUYGENS: Es sei p_n der Umfang des dem Kreis mit dem Radius R einbeschriebenen und P_n der Umfang des dem Kreis mit dem Radius R umbeschriebenen regelmäßigen n-Ecks. Dann ist

$$\frac{2}{3}p_n + \frac{1}{3}P_n > 2\pi R.$$

Offenbar ist

$$p_n = 2Rn\sin\frac{\pi}{n} \quad \text{und} \quad P_n = 2Rn\operatorname{tg}\frac{\pi}{n},$$

so daß

$$\frac{2}{3}p_n + \frac{1}{3}P_n = \frac{2Rn}{3}\left(2\sin\frac{\pi}{n} + \operatorname{tg}\frac{\pi}{n}\right)$$

und daher die zu beweisende Ungleichung auch in der Form

$$2\sin\frac{\pi}{n}+\operatorname{tg}\frac{\pi}{n}>3\frac{\pi}{n}$$

geschrieben werden kann. Wir zeigen nun, daß sogar allgemein

$$2\sin x+\operatorname{tg}x>3x,\quad 0<x<\frac{\pi}{2}, \tag{1}$$

gilt, woraus dann speziell der Satz von HUYGENS folgt.

Dazu betrachten wir die Funktion

$$f(x)=2\sin x+\operatorname{tg}x-3x.$$

Offenbar ist

$$f'(x)=2\cos x+\frac{1}{\cos^2 x}-3=\frac{2\cos^3 x-3\cos^2 x+1}{\cos^2 x}$$

$$=\frac{(2\cos^3 x-2\cos^2 x)+(1-\cos^2 x)}{\cos^2 x}=\frac{(1-\cos x)(1+\cos x-2\cos^2 x)}{\cos^2 x}.$$

Da nun für $0<x<\frac{\pi}{2}$ die Ungleichungen $1-\cos x>0$, $\cos^2 x>0$ und $1+\cos x-2\cos^2 x=1-\cos^2 x+\cos x(1-\cos x)>0$ gelten, ist $f'(x)>0$ und daher die Funktion $f(x)$ echt wachsend, so daß

$$f(x)>0\quad\text{für}\quad 0<x<\frac{\pi}{2}.$$

Hieraus ergibt sich sofort die Behauptung. Aus (1) folgt ferner

$$\sin x+\operatorname{tg}x>3x-\sin x>3x-x=2x,$$

das heißt

$$\sin x+\operatorname{tg}x>2x,\quad 0<x<\frac{\pi}{2}.$$

Wir weisen darauf hin, daß man mit der in den behandelten Beispielen verwendeten Methode viele weitere Ungleichungen beweisen kann.

15. Extrema von Funktionen. Die Funktionen, mit denen man es in den Anwendungen zu tun hat, sind im allgemeinen nicht in ihrem ganzen Definitionsbereich monoton. Meistens zeigt ihr graphisches Bild einen Verlauf, der dem in Abb. 13 dargestellten ähnlich ist.

Die in Abb. 13 dargestellte Kurve weist nun in den Punkten B, C, D, E, F, G eine gewisse Besonderheit auf. Obwohl die zu diesen Punkten gehörenden Funktionswerte nicht größter bzw. kleinster Funktionswert auf dem Intervall $[a, b]$ sind, sind sie doch Extremwerte im

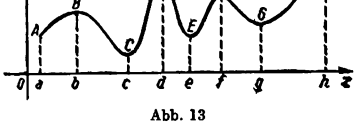


Abb. 13

Vergleich zu allen zu hinreichend benachbarten Argumentwerten gehörenden Funktionswerten. Zum Beispiel ist der Funktionswert für $x = b$ (welcher zum Punkt B gehört) der größte Wert der Funktion $f(x)$ im Intervall $[a, c]$ usw. Im Zusammenhang mit diesen Überlegungen geben wir die folgende Definition: Man sagt, die Funktion $f(x)$ besitze im Punkte $x = x_0$ ein *Maximum* bzw. *Minimum*, wenn es ein den Punkt x_0 im Innern enthaltendes Intervall $[p, q]$ gibt (d. h. $p < x_0 < q$), welches im Definitionsbereich der Funktion liegt und wobei $f(x) \leqq f(x_0)$ $(f(x) \geqq f(x_0))$ für alle $x \in [p, q]$. Die Maxima und Minima einer Funktion werden auch zusammenfassend deren *Extrema* oder *Extremwerte* genannt.

Wir weisen darauf hin, daß nach der angegebenen Definition die Punkte, in denen eine gegebene Funktion Extremwerte annimmt, grundsätzlich im Innern des Definitionsbereiches liegen. Man kann deshalb von der in Abb. 13 dargestellten Funktion nicht behaupten, daß sie im Punkte $x = h$ (in welchem sie ihren größten Wert annimmt) ein Maximum besitzt. Diese Einschränkung haben wir deshalb in die Definition aufgenommen, damit man auf die Punkte, in denen sich ein Extremum befindet, den Fermatschen Satz (Abschnitt 10) anwenden kann. Dieser gestattet, wie man leicht sieht, jetzt folgende Formulierung:

Satz von Fermat. *Nimmt die Funktion $f(x)$ im Punkte $x = x_0$ ein Extremum an und existiert in diesem Punkt die Ableitung $f'(x_0)$, so ist $f'(x_0) = 0$.*

Es versteht sich von selbst, daß die Umkehrung hiervon nicht richtig ist: Der Tatsache, daß in einem Punkt die Ableitung verschwindet, kann man keineswegs entnehmen, daß in diesem Punkt ein Extremum vorliegen muß. Zum Beispiel ist für die Funktion $f(x) = x^3$ offenbar $f'(0) = 0$, obwohl die Funktion $f(x)$ im Punkte $x = 0$ kein Extremum besitzt.

Verschwindet in einem Punkte x_0 die Ableitung $f'(x_0)$, so sagt man mitunter, die Funktion $f(x)$ verhalte sich im Punkte x_0 *stationär*. Die Punkte stationären Verhaltens sind gerade diejenigen Punkte, in welchen die Tangente an das geometrische Bild der Funktion parallel zur x-Achse ist.

Der Satz von Fermat besagt also, daß sich eine Funktion in jedem Punkte, in welchem sie einen Extremwert annimmt, stationär verhält. Jedoch hat man hierbei grundsätzlich die Existenz der Ableitung im betrachteten Punkt vorauszusetzen. Zum Beispiel besitzt die Funktion $y = |x|$ (vgl. Abb. 6) im Punkte $x = 0$ ein Minimum, ohne daß sie sich dort stationär verhält (denn in diesem Punkt existiert die Ableitung nicht).

Für die genauere Diskussion einer Funktion ist es nun wichtig, ihre Extremwerte zu kennen. Da sich allgemein hierüber nur sehr wenig sagen läßt, wollen wir uns im folgenden auf eine spezielle Klasse von Funktionen beschränken.

Wir sagen, die Funktion $f(x)$ gehöre der Klasse $K([a, b])$ an, wenn sie folgende Bedingungen erfüllt:

1. Die Funktion $f(x)$ ist im Intervall $[a, b]$ erklärt, und es existiert in allen Punkten dieses Intervalls die Ableitung $f'(x)$.[1])

[1]) Hieraus folgt bereits, daß $f(x)$ im Intervall $[a, b]$ stetig ist.

2. Die Ableitung $f'(x)$ ist im Intervall $[a, b]$ stetig.

3. Die Funktion $f(x)$ verhält sich nur in höchstens endlich vielen Punkten des Intervalls $[a, b]$ stationär.

Für die Funktionen dieser Klasse läßt sich die Frage nach den Extremwerten verhältnismäßig einfach beantworten.

Gibt es nämlich in (a, b) keinen Punkt stationären Verhaltens, so besitzt die Funktion im Intervall $[a, b]$ keinen Extremwert. In diesem Fall ist $f(x)$ monoton. In der Tat: Angenommen, es gäbe Punkte x' und x'' aus (a, b), in welchen die Ableitung $f'(x)$ verschiedene Vorzeichen besitzt; dann gäbe es wegen der Stetigkeit von $f'(x)$ zwischen x' und x'' einen Punkt, in welchem die Ableitung verschwindet, und das sollte ja nicht der Fall sein; folglich besitzt die Ableitung $f'(x)$ in (a, b) überall dasselbe Vorzeichen, und das bedeutet, daß die Funktion $f(x)$ monoton ist.

Gibt es dagegen im Intervall (a, b) Punkte $x_1 < x_2 < \cdots < x_m$, in welchen sich die Funktion $f(x)$ stationär verhält (und sind dies auch alle derartigen Punkte), so sorgt eine der vorangehenden entsprechende Überlegung, daß die Ableitung $f'(x)$ in den offenen Intervallen (a, x_1), (x_1, x_2), ..., (x_m, b) jeweils ihr Vorzeichen beibehält, so daß auf den zugehörigen abgeschlossenen Intervallen $[a, x_1]$, $[x_1, x_2]$, ..., $[x_m, b]$ die Funktion $f(x)$ monoton ist.

Auf diese Weise gelangen wir zu der folgenden

Regel. *Um für eine Funktion $f(x)$ der Klasse $K([a, b])$ die Extremwerte zu bestimmen, genügt es,*

1. *die Ableitung $f'(x)$ zu bestimmen;*

2. *die Lösungen der Gleichung $f'(x) = 0$ zu ermitteln;*

3. *für die im offenen Intervall (a, b) liegenden Lösungen $x_1, \ldots, x_m$ dieser Gleichung das Vorzeichen von $f'(x)$ in den Intervallen (a, x_1), (x_1, x_2), ..., (x_m, b) und damit das Verhalten von $f(x)$ in diesen Intervallen zu bestimmen.*

Offenbar sind dann nämlich für jedes i ($1 \leqq i \leqq m$) folgende Fälle möglich:[1])

	Vorzeichen von $f'(x)$ für $x_{i-1} < x < x_i$	Vorzeichen von $f'(x)$ für $x_i < x < x_{i+1}$	Verhalten der Funktion im Punkte x_i
1	+	+	kein Extremwert
2	+	−	Maximum
3	−	+	Minimum
4	−	−	kein Extremwert

In den Fällen 1 und 4 sagt man, die Funktion $f(x)$ besitze im Punkte x_i einen *Wendepunkt*.

In den Fällen, daß in den Punkten x_i, in welchen sich die Funktion $f(x)$ stationär verhält, auch die zweite Ableitung $f''(x_i)$ existiert, lassen sich die Untersuchungen häufig dadurch vereinfachen, daß man von folgendem Satz Gebrauch macht:

[1]) Es sei $a = x_0$, $b = x_{m+1}$ gesetzt.

Satz. *Es sei x_i ein Punkt des offenen Intervalls (a, b), in welchem sich die Funktion $f(x)$ stationär verhält. Ist dabei $f''(x_i) \neq 0$, so liegt im Punkte x_i mit Sicherheit ein Extremwert vor, und zwar ein Maximum, wenn $f''(x_i) < 0$ ist, und ein Minimum, wenn $f''(x_i) > 0$ ist.*

Beweis. Setzen wir $x_i + \Delta x = x$, so ist

$$f''(x_i) = \lim_{x \to x_i} \frac{f'(x) - f'(x_i)}{x - x_i},$$

also wegen $f'(x_i) = 0$

$$f''(x_i) = \lim_{x \to x_i} \frac{f'(x)}{x - x_i}.$$

Um etwas Bestimmtes vor Augen zu haben, wollen wir annehmen, daß $f''(x_i) > 0$. Dann ist für hinreichend kleines $|x - x_i|$

$$\frac{f'(x)}{x - x_i}$$

ebenfalls positiv, d. h., für ein hinreichend nahe bei x_i gelegenes x haben $f'(x)$ und $x - x_i$ dasselbe Vorzeichen, so daß $f'(x)$ im Intervall (x_{i-1}, x_i) negativ und im Intervall (x_i, x_{i+1}) positiv ist, und das bedeutet ja gerade, wie wir bereits wissen, daß im Punkte x_i ein Minimum vorliegt.

Beispiel 1. Die Funktion

$$y = \frac{x}{1 + x^2}$$

ist auf Extremwerte zu untersuchen.

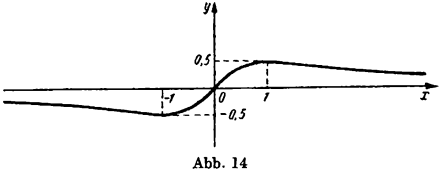


Abb. 14

Im betrachteten Beispiel ist

$$y' = \frac{1 - x^2}{(1 + x^2)^2}.$$

Man sieht sofort, daß y' genau in den Punkten $x_1 = -1$ und $x_2 = 1$ den Wert Null annimmt. Als Vorzeichen für y' in den Intervallen $(-\infty, -1)$, $(-1, 1)$, $(1, \infty)$ ergibt sich bzw. $-$, $+$, $-$. Daraus folgt, daß die betrachtete Funktion im Punkte $x = -1$ ein Minimum und im Punkte $x = 1$ ein Maximum besitzt. Beachten wir nun noch, daß $f(-1) = -\frac{1}{2}$, $f(+1) = +\frac{1}{2}$, daß das graphische Bild der Funktion $f(x)$ durch den Koordinatenursprung geht und daß $\lim f(x) = 0$ sowohl für $x \to +\infty$ als auch für $x \to -\infty$, so können wir uns bereits eine sehr gute Vorstellung vom Verlauf des graphischen Bildes der Funktion machen (Abb. 14).

Beispiel 2. Die Funktion

$$y = x^3(x-5)^2$$

ist auf Extremwerte zu untersuchen.

Im betrachteten Fall ist $y' = 5x^2(x-5)(x-3)$. Daraus ergeben sich als Punkte stationären Verhaltens: $x_1 = 0$, $x_2 = 3$, $x_3 = 5$. Das Vorzeichen der Ableitung y' in den Intervallen $(-\infty, 0)$, $(0,3)$, $(3,5)$, $(5,\infty)$ ist bzw. $+, +, -, +$. Daraus folgt, daß im Punkte $x = 0$ ein Wendepunkt, im Punkte $x = 3$ ein Maximum und im Punkte $x = 5$ ein Minimum vorliegt. Offenbar ist $f(0) = 0, f(3) = 108$, $f(5) = 0$. Daraus folgt, daß das graphische Bild der Funktion die x-Achse im Punkte $x = 0$ schneidet und im Punkte $x = 5$ berührt. Ferner ist klar, daß

$$\lim_{x \to -\infty} y = -\infty, \qquad \lim_{x \to +\infty} y = +\infty.$$

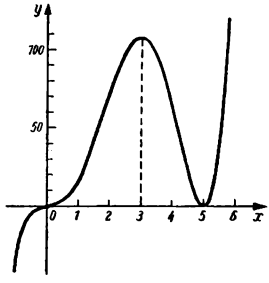


Abb. 15

In Abb. 15 ist der ungefähre Verlauf des graphischen Bildes der betrachteten Funktion dargestellt.[1])

Wir empfehlen dem Leser, die hier erarbeiteten Methoden auch auf die im vorangehenden Artikel (vgl. S. 41, 43 usw.) ohne Verwendung der Differentialrechnung diskutierten Funktionen (z. B. $y = x^3 - x$ usw.) anzuwenden.

16. Bestimmung des größten und des kleinsten Wertes einer Funktion auf einem abgeschlossenen Intervall. Es sei $f(x)$ eine Funktion der Klasse $K([a, b])$. Da sie stetig ist, besitzt sie (vgl. S. 207) auf $[a, b]$ einen größten Wert M und einen kleinsten Wert m. Unsere Aufgabe soll es nun sein, M und m zu bestimmen. Dabei beschränken wir uns der Einfachheit halber auf die Bestimmung von M; die Überlegungen für m verlaufen entsprechend.

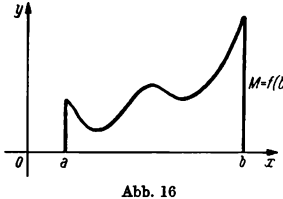


Abb. 16

Wir wollen annehmen, daß der Wert M im Punkte x_0 angenommen wird, d. h., es sei $f(x_0) = M$. Ist $a < x_0 < b$, so nimmt die Funktion $f(x)$ im Punkte x_0 offenbar ein Maximum an. Es kann aber natürlich auch der Fall eintreten, daß $x_0 = a$ oder $x_0 = b$ (Abb. 16); in diesem Fall braucht sich die Funktion $f(x)$ im Punkte x_0 nicht stationär zu verhalten. Aus dem Gesagten erhalten wir unmittelbar folgende

[1]) In Abb. 15 sind auf der x-Achse und auf der y-Achse verschiedene Maßstäbe gewählt, damit der Punkt mit der Ordinate 108 noch Platz auf der Abbildung findet.

Regel. *Um den größten Wert der Funktion $f(x)$ im Intervall $[a, b]$ zu finden, muß man alle Maxima der Funktion $f(x)$ in (a, b) bestimmen. Werden diese in den Punkten $x_1, x_2, \ldots, x_n$ angenommen, so ist der größte Wert der Funktion $f(x)$ im Intervall $[a, b]$ die größte der Zahlen*

$$f(x_1), f(x_2), \ldots, f(x_n), f(a), f(b).$$

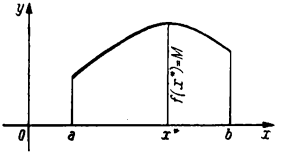


Abb. 17

Besonderes Interesse verdient der Fall, daß die Funktion $f(x)$ im offenen Intervall (a, b) nur einen Extremwert $f(x^*)$ annimmt und dieser Extremwert ein Maximum ist. In diesem Fall ist, wie man sofort sieht, *der zugehörige Funktionswert $f(x^*)$ der größte Wert der Funktion $f(x)$ im Intervall $[a, b]$*. Diese Verhältnisse sind in Abb. 17 angedeutet.

Wir betrachten nun zwei Aufgaben, die sich mit Hilfe der hier entwickelten Theorie lösen lassen.

Aufgabe 1. Vorgelegt ist (Abb. 18) ein rechteckiges Stück Blech im Format 8 dm × 5 dm. Aus den Ecken sollen Quadrate so ausgeschnitten werden, daß nach Hochklappen der überstehenden Teile ein oben offener Kasten möglichst großen Inhalts entsteht.

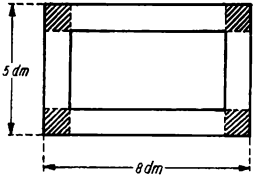


Abb. 18

Bezeichnen wir mit x die Seitenlänge des ausgeschnittenen Quadrats, so ist offenbar $0 < x < 2{,}5$. Das Volumen des entstehenden Kastens wird dann durch

$$V = x(8-2x)(5-2x) = 4x^3 - 26x^2 + 40x$$

gegeben, und unsere Aufgabe besteht darin, den größten Wert dieser Funktion im Intervall $\left[0, 2\frac{1}{2}\right]$ zu berechnen. Zu diesem Zweck differenzieren wir V zweimal:

$$V' = 12x^2 - 52x + 40, \qquad V'' = 24x - 52.$$

Als Lösungen von $V' = 0$ ergeben sich $x_1 = 1$ und $x_2 = 3\frac{1}{3}$, von denen nur $x_1 = 1$ im Intervall $\left(0, 2\frac{1}{2}\right)$ liegt. Wegen $V''(1) = -28 < 0$ liegt für $x = 1$ ein Maximum vor, und hier wird, wie oben bemerkt, gleichzeitig auch der größte Wert angenommen. Man erhält also den Kasten größten Inhalts, wenn man ein Quadrat der Seitenlänge 1 dm ausschneidet.

Aufgabe 2. Einem gegebenen Kegel soll ein Zylinder größten Inhalts einbeschrieben werden (Abb. 19).

Bezeichnen wir den Radius des gesuchten Zylinders mit r und seine Höhe mit h, so wird sein Volumen durch

$$V = \pi r^2 h$$

gegeben. Aus der Ähnlichkeit der Dreiecke ABC und MNC folgt nun

$$\frac{h}{H} = \frac{R-r}{R}, \tag{2}$$

wobei R der Radius und H die Höhe des gegebenen Kegels sind. Hieraus findet man

$$V = \pi \frac{H}{R}(Rr^2 - r^3).$$

Von dieser Funktion (von r) haben wir den größten Wert im Intervall $(0, R)$ zu bestimmen.

Differenzieren wir diese Funktion zweimal, so erhalten wir

$$V' = \pi \frac{H}{R}(2Rr - 3r^2),$$

$$V'' = \pi \frac{H}{R}(2R - 6r).$$

Die Gleichung $V' = 0$ besitzt nun offenbar genau die Lösungen $r = 0$ und $r = \frac{2}{3}R$. Von diesen beiden liegt nur die zweite im Intervall $(0, R)$, und da für sie $V''\left(\frac{2}{3}R\right) = -2\pi H < 0$ ist, liefert sie bereits den größten Wert von V. Also wird der Radius des gesuchten Zylinders durch $r = \frac{2}{3}R$ gegeben. Die Höhe h kann man sodann auf Grund von (2) berechnen.

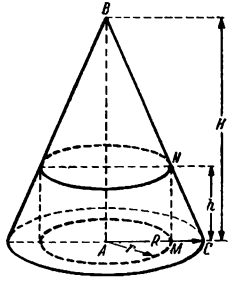


Abb. 19

Wir wollen hier auf die Angabe weiterer Beispiele verzichten, da bereits die betrachteten Beispiele die Wirksamkeit der Methoden der Differentialrechnung deutlich erkennen lassen.[1])

[1]) In vielen Fällen lassen sich unter Ausnutzung individueller Besonderheiten der gestellten Aufgabe der größte und der kleinste Wert einer Funktion auch mit Hilfe von elementaren Methoden bestimmen. Die Wirksamkeit der Methoden der Differentialrechnung besteht aber gerade darin, daß man die individuellen Besonderheiten der gestellten Aufgabe vollständig ignorieren kann; die Methoden der Differentialrechnung führen bei der Klasse $K[(a, b)]$ stets zum Ziel. Die Dinge liegen hier ganz entsprechend wie bei der Lösung arithmetischer Aufgaben durch Aufstellung von Gleichungen. Man kann auch dort mitunter ohne diese auskommen, jedoch geben die Gleichungen eine ganz allgemeine Methode zum Lösen arithmetischer Aufgaben.

Wir bemerken zum Abschluß, daß es eine ganze Reihe von speziellen Methoden gibt, mit deren Hilfe sich die Berechnung des größten und des kleinsten Wertes einer Funktion noch weiter vereinfachen läßt. Schließlich kann man eine ganz entsprechende Theorie auch für Funktionen von mehreren Veränderlichen aufbauen. Wir wollen jedoch auf diese Dinge hier nicht weiter eingehen. Man findet sie in allen Einzelheiten in sämtlichen Lehrbüchern der Differential- und Integralrechnung.

Kapitel II

INTEGRALE

§ 4. Das unbestimmte Integral

17. Grundbegriffe. Grundoperation der Differentialrechnung ist das Differenzieren, d. h. das Aufsuchen der Ableitung einer gegebenen Funktion. Dabei handelt es sich, wie wir bereits wissen, im wesentlichen darum, die „Geschwindigkeit" zu bestimmen, mit der sich die Funktionswerte im Vergleich zu den Argumentwerten ändern. Häufig hat man es nun in den Anwendungen auch mit der umgekehrten Aufgabe zu tun. Dies ist z. B. immer dann der Fall, wenn aus der gegebenen Ablaufgeschwindigkeit eines Prozesses der Prozeß selbst zu rekonstruieren ist. Mathematisch bedeutet diese Aufgabe das Bestimmen einer Funktion aus ihrer Ableitung. Mit diesem Problem beschäftigt sich im wesentlichen das zweite Hauptgebiet der Analysis, die sogenannte Integralrechnung.

Ist der Wert einer in einem gewissen Intervall[1]) $[a, b]$ definierten Funktion $f(x)$ in allen Punkten dieses Intervalls gleich der Ableitung der ebenfalls im Intervall $[a, b]$ definierten Funktion $F(x)$, so nennt man $F(x)$ eine *Stammfunktion* oder *primitive Funktion* zu $f(x)$ (im Intervall $[a, b]$).

Es gilt dann der bemerkenswerte

Satz 1. *Jede auf einem abgeschlossenen Intervall $[a, b]$ stetige Funktion besitzt dort eine Stammfunktion.*

Einen Beweis für diesen Satz werden wir später in Abschnitt 24 geben.

Man sieht sofort, daß *für eine beliebige Konstante C mit $F(x)$ auch $F(x) + C$ Stammfunktion zu $f(x)$ ist.* Umgekehrt *hat aber auch, wenn $F(x)$ Stammfunktion zu $f(x)$ ist, jede andere Stammfunktion die Form $F(x) + C$.* In der Tat: Ist $F_1(x)$ eine beliebige Stammfunktion zu $f(x)$, so ist die Ableitung von $F_1(x) - F(x)$ auf $[a, b]$ überall gleich Null, und daraus folgt (vgl. Abschnitt 14), daß diese Differenz konstant ist, d. h.

$$F_1(x) - F(x) = C \quad \text{und} \quad F_1(x) = F(x) + C.$$

Ist $F(x)$ eine Stammfunktion zu $f(x)$, so wird die Funktion $F(x) + C$ der beiden Veränderlichen x und C das *unbestimmte Integral der Funktion $f(x)$* genannt und mit

$$\int f(x)\, dx$$

bezeichnet.

---

[1]) Dieses Intervall kann offen, abgeschlossen oder halboffen sein. Um etwas Bestimmtes vor Augen zu haben, wollen wir im folgenden stets abgeschlossene Intervalle betrachten.

Das unbestimmte Integral einer Funktion ist also die allgemeine Form der Stammfunktion dieser Funktion. Die Größe C, die in die Definition des unbestimmten Integrals eingeht, bezeichnet man als „willkürliche Konstante" oder *Integrationskonstante*. Durch passende Spezialisierung von C kann man aus dem unbestimmten Integral jede Stammfunktion gewinnen.

Aus der Definition des unbestimmten Integrals erhält man sofort die folgende Aussage:

Satz 2. *Die Ableitung des unbestimmten Integrals einer Funktion $f(x)$ ist gleich der Funktion $f(x)$, d. h.*

$$\left(\int f(x)\,dx\right)' = f(x).$$

Zur erfolgreichen Anwendung der Integralrechnung beim Aufsuchen des unbestimmten Integrals elementarer Funktionen ist eine ausgefeilte Technik erforderlich. Der Leser, der diese erlangen möchte, muß auf die Spezialliteratur verwiesen werden, da hier nur ein allgemeiner Überblick über die dabei verwendeten Methoden gegeben werden kann.

Als Grundlage für die Integrationstechnik dienen eine Reihe von einfachen Formeln, von denen wir hier nur die folgenden nennen wollen:

1. $\int dx = x + C.$
2. $\int x^a\,dx = \frac{x^{a+1}}{a+1} + C \qquad (a \neq -1).$
3. $\int \frac{dx}{x} = \ln x + C.$
4. $\int a^x\,dx = \frac{a^x}{\ln a} + C.$
5. $\int e^x\,dx = e^x + C.$
6. $\int \cos x\,dx = \sin x + C.$
7. $\int \sin x\,dx = -\cos x + C.$
8. $\int \frac{dx}{\cos^2 x} = \operatorname{tg} x + C.$
9. $\int \frac{dx}{\sin^2 x} = -\operatorname{ctg} x + C.$
10. $\int \frac{dx}{\sqrt{a^2 - x^2}} = \arcsin \frac{x}{a} + C \qquad (a > 0).$
11. $\int \frac{dx}{a^2 + x^2} = \frac{1}{a} \operatorname{arctg} \frac{x}{a} + C \qquad (a \neq 0).$
12. $\int \frac{dx}{x^2 - a^2} = \frac{1}{2a} \ln \frac{x-a}{x+a} + C \qquad (a \neq 0).$

Die Gültigkeit dieser Formeln beweist man dadurch, daß man jeweils die rechte Seite differenziert; denn auf Grund der Definition des unbestimmten Integrals gilt die

Regel. *Um die Gültigkeit einer Gleichung*

$$\int f(x)\,dx = F(x) + C \qquad (*)$$

nachzuweisen, muß man ihre rechte Seite differenzieren. Ergibt sich dabei die auf der linken Seite auftretende Funktion $f(x)$, so ist die Gleichung (*) richtig.

Wir wollen hier darauf verzichten, die in der Tabelle angegebenen Gleichungen als richtig nachzuweisen. Es sollen lediglich einige Bemerkungen über den Gültigkeitsbereich dieser Formeln gemacht werden.

Die Formeln 1, 4, 5, 6, 7, 11 gelten auf der ganzen x-Achse (in Formel 4 muß natürlich vorausgesetzt werden, daß $a > 0$, $a \neq 1$; — *Anm. d. wissenschaftl. Red.*).

Formel 2 gilt in allen den Intervallen, in welchen beide Seiten einen Sinn haben; zum Beispiel gilt also die Formel

$$\int \frac{dx}{\sqrt{x}} = 2\sqrt{x} + C$$

in jedem Teilintervall des Intervalls $(0, \infty)$, die Formel

$$\int \sqrt[4]{x}\, dx = \frac{3}{4}\sqrt[4]{x^7} + C$$

in jedem Intervall usw.

Formel 3 gilt in jedem Teilintervall des Intervalls $(0, \infty)$. Ist dagegen $x \in (-\infty, 0)$ so tritt an die Stelle der Formel 3 die Formel

$$\int \frac{dx}{x} = \ln(-x) + C.$$

Formel 8 gilt in jedem Intervall, in welchem kein Punkt $(2n+1)\frac{\pi}{2}$ vorkommt, Formel 9 in jedem Intervall, welches keinen Punkt $n\pi$ enthält.

Formel 10 gilt im Intervall $(-a, +a)$ und Formel 12 schließlich in $(-\infty, -|a|)$ und $(|a|, \infty)$.

Die Integrale komplizierterer elementarer Funktionen kann man sodann unter Umständen durch passende Kombination der in der „Integraltabelle" angegebenen Formeln erhalten. Hierfür gibt es nun eine ganze Reihe von Methoden. Die einfachsten dieser Methoden bestehen in der Anwendung der folgenden beiden Sätze:

Satz 3. *Das Integral einer Summe oder Differenz mehrerer Funktionen ist gleich der Summe bzw. Differenz der Integrale der einzelnen Funktionen, d. h. zum Beispiel*

$$\int [f(x) + g(x) - h(x)]\, dx = \int f(x)\, dx + \int g(x)\, dx - \int h(x)\, dx. \qquad (*^*)$$

Satz 4. *Ein konstanter Faktor kann vor das Integralzeichen gezogen werden, d. h.*

$$\int c f(x)\, dx = c \int f(x)\, dx.$$

Wir wollen als Beispiel die Formel $(*^*)$ verifizieren. Zu diesem Zweck haben wir auf Grund der oben erwähnten Regel die rechte Seite zu differenzieren. Da die Ableitung einer Summe (Differenz) gleich der Summe (Differenz) der Ableitungen ist, erhalten wir die Ableitung der rechten Seite von $(*^*)$, wenn

wir die dort stehenden Integrale einzeln differenzieren. Auf Grund von Satz 2 ergibt dies aber gerade $f(x)+g(x)-h(x)$, d. h. die auf der linken Seite unter dem Integralzeichen stehende Funktion. Ähnlich beweist man Satz 4.

18. Integration durch Substitution. Ein sehr weittragendes Verfahren für die Reduktion komplizierterer Integrale auf einfache liefert die sogenannte Substitutionsregel. Sie kann in zwei verschiedenen Formen angewendet werden, welche aber beide auf dem folgenden Satz basieren:

Satz. *Es sei $F(z)$ im Intervall $[p, q]$ Stammfunktion zu $f(z)$. Dann ist für jede im Intervall $[a, b]$ definierte differenzierbare Funktion $\varphi(x)$, deren Werte die Ungleichung $p \leqq \varphi(x) \leqq q$ erfüllen, die mittelbare Funktion $F(\varphi(x))$ Stammfunktion zu $f(\varphi(x))\,\varphi'(x)$.*

Beweis. Zur Differentiation der mittelbaren Funktion $y = F(\varphi(x))$ führen wir das Hilfsargument $z = \varphi(x)$ ein. Dann ist $y = F(z)$, $z = \varphi(x)$ und daher $y'_x = y'_z z'_x = F'(z)\varphi'(x)$. Da nun nach Voraussetzung $F'(z) = f(z)$ ist, gilt $y'_x = f(z)\varphi'(x) = f(\varphi(x))\,\varphi'(x)$, womit der angegebene Satz bereits bewiesen ist.

Offenbar kann man den zuletzt genannten Satz auch kurz folgendermaßen formulieren:

Ist

$$\int f(z)\,dz = F(z) + C,$$

so ist

$$\int f(\varphi(x))\,\varphi'(x)\,dx = F(\varphi(x)) + C.$$

Hieraus folgt:

Erste Substitutionsregel. *Um ein Integral der Form*

$$I = \int f(\varphi(x))\,\varphi'(x)\,dx$$

zu berechnen, bringt man es auf die Form

$$I = \int f(\varphi(x))\,d\varphi(x),$$

ersetzt hier $\varphi(x)$ durch z, berechnet das entstehende Integral und setzt im Resultat für z wieder $\varphi(x)$ ein.

Hier einige Beispiele für die Anwendung der ersten Substitutionsregel:

1. $\displaystyle\int \frac{dx}{x(1+\ln^2 x)} = \int \frac{d\ln x}{1+\ln^2 x} = \int \frac{dz}{1+z^2} = \operatorname{arctg} z + C = \operatorname{arctg} \ln x + C;$

2. $\displaystyle\int \frac{e^x\,dx}{1+e^{2x}} = \int \frac{de^x}{1+e^{2x}} = \int \frac{dz}{1+z^2} = \operatorname{arctg} z + C = \operatorname{arctg} e^x + C;$

3. $\displaystyle\int \frac{\cos x\,dx}{1+\sin^2 x} = \int \frac{d\sin x}{1+\sin^2 x} = \int \frac{dz}{1+z^2} = \operatorname{arctg} z + C = \operatorname{arctg} \sin x + C.$

Die bisher angegebenen Beispiele zeigen, wie man aus der einen Formel

$$\int \frac{dz}{1+z^2} = \operatorname{arctg} z + C$$

beliebig viele weitere Integrationsformeln erhalten kann.

Entsprechend erhält man aus

$$\int \frac{dz}{z} = \ln z + C$$

die weiteren Integrationsformeln

$$\int \operatorname{tg} x\, dx = \int -\frac{d\cos x}{\cos x} = -\ln \cos x + C,$$

$$\int \frac{x^2\, dx}{x^3+2} = \frac{1}{3}\int \frac{d(x^3+2)}{x^3+2} = \frac{1}{3}\ln(x^3+2) + C,$$

$$\int \frac{dx}{(1+x^2)\operatorname{arctg} x} = \int \frac{d\operatorname{arctg} x}{\operatorname{arctg} x} = \ln \operatorname{arctg} x + C$$

usw.

Wir kommen nun zur zweiten Substitutionsregel. Dazu sei $f(x)$ eine im Intervall $[a, b]$ stetige Funktion. Es sei ferner $x = \varphi(t)$ eine auf einem Intervall $[\alpha, \beta]$ definierte differenzierbare Funktion, deren Werte die Ungleichung $a \leqq \varphi(t) \leqq b$ erfüllen und für die auf $[a, b]$ die Umkehrfunktion $t = \psi(x)$ existiert. Wir betrachten dann das Integral

$$I_1 = \int f\big(\varphi(t)\big)\, \varphi'(t)\, dt.$$

Auf Grund des oben Gesagten kann man dieses Integral dadurch berechnen, daß man es zunächst in der Form

$$\int f\big(\varphi(t)\big)\, d\varphi(t)$$

schreibt, hier $\varphi(t)$ durch x ersetzt und dann das Integral

$$I = \int f(x)\, dx$$

auswertet (dieses Integral existiert, weil die Funktion $f(x)$ nach Voraussetzung stetig ist). Ist etwa

$$\int f(x)\, dx = A(x) + C,$$

so ist nach der ersten Substitutionsregel

$$I_1 = A\big(\varphi(t)\big) + C.$$

Es sei nun $A\big(\varphi(t)\big) = F(t)$. Ersetzen wir hier t durch $\psi(x)$ und berücksichtigen wir, daß $\varphi\big(\psi(x)\big) = x$ ist, so erhalten wir

$$A(x) = F\big(\psi(x)\big).$$

Also ist

$$\int f(x)\, dx = F\big(\psi(x)\big) + C.$$

Beachten wir schließlich noch, daß $F(t)$ nichts anderes ist als die Stammfunktion zu $f\big(\varphi(t)\big)\varphi'(t)$, so finden wir:

Zweite Substitutionsregel. *Um das Integral*

$$I=\int f(x)\,dx$$

zu berechnen, kann man folgendermaßen vorgehen: Man setzt zunächst $x=\varphi(t)$, *wobei* $\varphi(t)$ *eine differenzierbare Funktion ist, die eine Umkehrfunktion* $\psi(x)$ *besitzt, berechnet das entstehende Integral und ersetzt im Resultat* t *durch* $\psi(x)$.

Wir betrachten zwei Beispiele für die Anwendung der zweiten Substitutionsregel:

Beispiel 1. Es sei $I=\int\sqrt{1-x^2}\,dx$. Setzen wir $x=\sin t$, so gelangen wir zu dem Integral

$$\int\cos^2 t\,dt=\int\frac{1+\cos 2t}{2}\,dt=\frac{t}{2}+\frac{1}{4}\int\cos 2t\,d(2t)=\frac{t}{2}+\frac{\sin 2t}{4}+C.$$

Also ist

$$I=\frac{\arcsin x}{2}+\frac{x\sqrt{1-x^2}}{2}+C.$$

Beispiel 2. Es sei $I=\int\frac{dx}{1+\sqrt{x}}$. Setzen wir $x=t^2$, so finden wir

$$I=\int\frac{2t\,dt}{1+t}=2\int\left(1-\frac{1}{1+t}\right)dt=2t-2\ln(1+t)+C$$
$$=2\sqrt{x}-2\ln(1+\sqrt{x})+C.$$

19. Partielle Integration. Ein weiteres wichtiges Verfahren zur Auswertung von Integralen ist die Methode der „partiellen Integration". Hierfür seien $u=u(x)$ und $v=v(x)$ zwei auf demselben Intervall $[a, b]$ definierte differenzierbare Funktionen. Dann ist bekanntlich in allen Punkten des Intervalls $[a, b]$

$$(uv)'=u'v+uv'.$$

Diese Beziehung können wir nun offenbar auch gleichwertig in der Form

$$\int(u'v+uv')\,dx=uv+C$$

schreiben. Beachten wir noch $u'dx=du$ und $v'dx=dv$, so ergibt sich

$$\int u\,dv=uv-\int v\,du. \tag{1}$$

Die Integrationskonstante C der vorangehenden Gleichung ist dabei offenbar in dem Integral $\int v\,du$ enthalten. Formel (1) bezeichnet man als Formel für die partielle Integration. Durch sie wird die Auswertung des Integrals $\int u\,dv$ auf die Auswertung des Integrals $\int v\,du$ zurückgeführt. Sie läßt sich also immer dann mit Erfolg anwenden, wenn das neu auftretende Integral einfacher als das Ausgangsintegral ist.

Dem Bau der Formel (1) läßt sich entnehmen, daß sie immer dann anwendbar ist, wenn die zu integrierende Funktion in Form eines Produktes $u\,dv$ einer Funktion u mit dem Differential dv einer anderen Funktion dargestellt ist. Nach Anwendung von (1) ist sodann ein Integral auszuwerten,

bei dem die zu integrierende Funktion gerade das Produkt aus v und dem Differential du ist. Der Übergang von dem einen zum anderen Integral besteht also mit anderen Worten darin, daß der Faktor dv integriert und der Faktor u differenziert wird. Die genannte Vereinfachung des Integrals kann nun sowohl bei der Integration als auch bei der Differentiation entstehen. In der Praxis ist es allerdings meistens so, daß die Vereinfachung durch die Differentiation des Faktors u zustande kommt. Ein gewisser Hinweis auf die Zweckmäßigkeit einer partiellen Integration kann daher das Vorhandensein eines sich bei Differentiation vereinfachenden Faktors in der zu integrierenden Funktion sein. Dieser Faktor wird dann als u genommen, während der Rest (einschließlich dx) als dv genommen wird.

Wir wenden uns nun wieder einigen Beispielen zu.

1. Es sei $I = \int x^3 \ln x \, dx$. Da sich $\ln x$ bei der Differentiation vereinfacht, setzen wir

$$\begin{array}{l|l} \ln x = u, & du = \dfrac{dx}{x}, \\ x^3 dx = dv, & v = \dfrac{x^4}{4}. \end{array}$$

Die Funktion $v = \frac{x^4}{4}$ wurde dabei durch Integration ihres Differentials[1]) $x^3 dx$ gefunden. Da wir in (1) v als eine bestimmte Stammfunktion ansehen können, fügen wir hier keine Integrationskonstante hinzu. Das ist auch im folgenden zu beachten. Also erhalten wir

$$I = \frac{x^4}{4} \ln x - \int \frac{x^4}{4} \cdot \frac{dx}{x} = \frac{x^4}{4} \ln x - \frac{x^4}{16} + C.$$

2. $\int \operatorname{arctg} x \, dx = x \operatorname{arctg} x - \int \frac{x \, dx}{1+x^2} = x \operatorname{arctg} x - \frac{1}{2} \ln (1 + x^2) + C.$

3. $\int x \cos x \, dx = \int x \, d \sin x = x \sin x - \int \sin x \, dx = x \sin x + \cos x + C.$

4. $\int x e^x \, dx = \int x \, d e^x = x e^x - \int e^x \, dx = x e^x - e^x + C.$

20. Allgemeine Bemerkungen zur Integration elementarer Funktionen. Wir haben bereits oben bemerkt, daß jede stetige Funktion $f(x)$ eine Stammfunktion besitzt. Dies ist insofern besonders erwähnenswert, weil nicht jede stetige Funktion eine Ableitung besitzt. Wenn wir also der Frage nach der Stammfunktion einer gegebenen stetigen Funktion einmal die Frage nach deren Ableitung gegenüberstellen, so müssen wir feststellen, daß die Antwort auf die zuerst genannte Frage stets positiv ausfällt, während dies bei der zweiten Frage nicht der Fall ist.

Beschränken wir uns nun aber auf die elementaren Funktionen und fragen wir, ob die Stammfunktion und die Ableitung einer elementaren Funktion stets wieder elementare Funktionen sind, so finden wir einen vollkommen anderen Sachverhalt vor. Wie wir bereits wissen, ist nicht nur jede elementare

[1]) Dieses Differential ist ein „Teil" des Integranden. Daher stammt auch die Bezeichnung „partielle Integration".

Funktion (d. h. jede Kombination der einfachsten elementaren Funktionen) differenzierbar, sondern in diesem Fall ist die Ableitung auch stets wieder eine elementare Funktion. Dagegen gibt es sehr einfache elementare Funktionen, deren Stammfunktion nicht mehr als Kombination der einfachsten elementaren Funktionen dargestellt werden kann. So kann man z. B. zeigen, daß keines der folgenden Integrale sich als endliche Kombination der einfachsten elementaren Funktionen darstellen läßt:

$$\left.\begin{gathered}\int e^{-x^2}\,dx, \quad \int x \operatorname{tg} x\,dx, \quad \int \sqrt{\sin x}\,dx, \quad \int e^{\operatorname{arctg} x}\,dx, \\ \int \frac{e^x}{x}\,dx, \quad \int \frac{dx}{\ln x}, \quad \int \frac{\sin x}{x}\,dx.\end{gathered}\right\} \tag{2}$$

Um zu einer Klärung dieser vielleicht auf den ersten Blick etwas seltsam anmutenden Situation zu gelangen, müssen wir uns zunächst vergegenwärtigen, daß die Abgrenzung des Begriffs der „elementaren" Funktion letzten Endes doch eine recht willkürliche Angelegenheit ist. Die Abgrenzung der Klasse der Funktionen, welche man heute üblicherweise als elementare Funktionen zu bezeichnen pflegt, ist in bedeutendem Maße durch die historische Entwicklung der Mathematik, und zwar nicht nur als Wissenschaft, sondern auch (und das sogar in hohem Grade) als Unterrichtsgegenstand, bedingt.

Wir wollen uns einmal vorzustellen versuchen, wie die Dinge lägen, wenn die historische Entwicklung etwas anders verlaufen wäre und man heute die Funktion $\ln x$ nicht zu den einfachsten elementaren Funktionen zählen würde. Es ist klar, daß man unabhängig davon die rationale Funktion $\frac{1}{x}$ als elementare Funktion ansehen müßte. Unter diesen Bedingungen wäre dann auch das Integral

$$\int \frac{dx}{x} \tag{3}$$

ein Integral einer elementaren Funktion, welches nicht durch die einfachsten elementaren Funktionen ausdrückbar ist.

Wir sehen also, daß die Antwort auf die Frage, welche Integrale von elementaren Funktionen sich durch elementare Funktionen ausdrücken lassen, ganz davon abhängt, in welcher Weise man den Begriff der elementaren Funktion abgrenzt. Erweitert man die Klasse der elementaren Funktionen dadurch, daß man noch weitere Funktionen als elementar ansieht, so wird man unter Umständen gewisse der Integrale von elementaren Funktionen (im alten Sinne), die zuvor nicht durch elementare Funktionen ausdrückbar waren, jetzt durch elementare Funktionen (im neuen Sinne) ausdrücken können. In dem zuvor genannten Beispiel braucht man nur die Funktion $\ln x$ als elementare Funktionen hinzuzunehmen, um das Integral (3) elementar „auswerten" zu können.

Jedoch wird man auch nach einer solchen Erweiterung wieder auf Integrale von elementaren Funktionen (im neuen Sinne) stoßen, deren Auswertung eine

erneute Erweiterung der Klasse der elementaren Funktionen erforderlich macht. Beachten wir z. B., daß das Integral

$$\int \frac{dx}{\ln x} \tag{4}$$

nicht elementar auswertbar ist, so können wir feststellen, daß im oben betrachteten Fall nach Hinzunahme der Funktion $\ln x$ zwar das Integral (3) auswertbar wird, daß aber die Auswertung des Integrals (4) eine erneute Erweiterung der Klasse der elementaren Funktionen erforderlich macht.

Angesichts der genannten Tatsachen ergibt sich nun die Frage, ob man nicht vielleicht durch eine Erweiterung der Klasse der elementaren Funktionen zu einer Funktionenklasse gelangen kann, für welche jedes Integral einer Funktion aus dieser Klasse sich durch Funktionen dieser Klasse auswerten läßt. Es zeigt sich, daß eine begrenzte Erweiterung der Klasse der elementaren Funktionen hier nicht zum Ziele führt. Dieses Problem läßt sich erst unter Verwendung von unendlichen Reihen lösen.

Die Abgrenzung der Klasse der elementaren Funktionen, wie sie sich im Laufe der historischen Entwicklung herausgebildet hat und wie sie heute üblicherweise verwendet wird, ist natürlich doch in einem gewissen Sinne gerechtfertigt; denn wenn auch diese Klasse noch ziemlich eng ist, so lassen sich doch bereits viele wichtige Probleme unter alleiniger Verwendung von elementaren Funktionen lösen. Hinzu kommt, daß man mit den elementaren Funktionen sehr bequem arbeiten kann, da für sie viele einfache Beziehungen gelten, wie z. B.

$$\sin^2 x + \cos^2 x = 1, \quad \sin(x+y) = \sin x \cos y + \sin y \cos x,$$

$$\log(xy) = \log x + \log y, \quad \frac{a^x}{a^y} = a^{x-y}$$

usw., die man leicht behalten kann und die zum großen Teil bereits dem Schüler bekannt sind. Es sei jedoch darauf hingewiesen, daß es viele Integrale gibt, die nicht elementar auswertbar sind und die doch zur Lösung der mannigfachsten Aufgaben notwendig sind. So spielt zum Beispiel das Integral $\int e^{-x^2} dx$ in der Wahrscheinlichkeitsrechnung eine bedeutende Rolle. In der Theorie der Lichtbeugung trifft man auf die Integrale

$$\int \sin x^2 \, dx, \qquad \int \cos x^2 \, dx,$$

die gleichfalls nicht elementar auswertbar sind. Bei den auf P. L. Tschebyschew zurückgehenden Untersuchungen über Primzahlverteilung spielt das Integral $\int \frac{dx}{\ln x}$ eine fundamentale Rolle. Wir werden später (in Abschnitt 27) auf die Frage nach der Auswertung von Integralen von einem anderen Gesichtspunkt aus zurückkommen.

Auf Grund der vorangehenden Überlegungen erlangt nun aber eine andere Frage ein besonderes Interesse, nämlich die Frage nach der Klasse derjenigen Funktionen, deren Integrale sich durch elementare Funktionen ausdrücken

lassen. In diesem Zusammenhang wollen wir vor allem den folgenden Satz erwähnen:

Satz. *Das Integral jeder rationalen Funktion*

$$\frac{a_0x^n+a_1x^{n-1}+\cdots+a_n}{b_0x^m+b_1x^{m-1}+\cdots+b_m} \tag{5}$$

mit reellen Koeffizienten a_i und b_k läßt sich durch rationale Funktionen, den Logarithmus und den Arkustangens ausdrücken.

In der Algebra zeigt man nämlich, daß sich jede Funktion der Form (5) (mit reellen Koeffizienten a_i und b_k) als Summe aus endlich vielen Summanden einer der folgenden fünf Typen darstellen läßt:

$$1.\ A_0x^r+A_1x^{r-1}+\cdots+A_r,\quad 2.\ \frac{A}{x-a},\quad 3.\ \frac{A}{(x-a)^r}\quad (r>1),$$

$$4.\ \frac{Ax+B}{x^2+px+q},\qquad 5.\ \frac{Ax+B}{(x^2+px+q)^r}\quad (r>1),$$

wobei die auftretenden Koeffizienten sämtlich reell sind und das in 4. und 5. im Nenner auftretende Polynom x^2+px+q keine reellen Nullstellen besitzt.

Was nun die Summanden der Form 1., 2. und 3. betrifft, so ergibt sich ihre Integration in elementaren Funktionen auf Grund der Formeln

$$\int\left(\sum_{k=0}^{r}A_kx^{r-k}\right)dx=\sum_{k=0}^{r}A_k\frac{x^{r-k+1}}{r-k+1}+C,$$

$$\int\frac{A}{x-a}\,dx=A\ln(x-a)+C,$$

$$\int\frac{A}{(x-a)^r}\,dx=\frac{A}{(1-r)(x-a)^{r-1}}+C.$$

Zur Integration der Summanden vom Typ 4. bringt man den Zähler $Ax+B$ auf die Form $\frac{A}{2}(2x+p)+\left(B-\frac{A}{2}p\right)$, so daß also

$$\int\frac{Ax+B}{x^2+px+q}\,dx=\frac{A}{2}\int\frac{2x+p}{x^2+px+q}\,dx+\left(B-\frac{A}{2}p\right)\int\frac{dx}{x^2+px+q}.$$

Das erste Integral auf der rechten Seite ist gleich $\ln(x^2+px+q)$. Zur Berechnung des zweiten Integrals führen wir die Substitution

$$\int\frac{dx}{x^2+px+q}=\int\frac{d\left(x+\frac{p}{2}\right)}{\left(x+\frac{p}{2}\right)^2+\left(q-\frac{p^2}{4}\right)}$$

durch. Beachten wir nun, daß $q-\frac{p^2}{4}>0$ (die Nullstellen von x^2+px+q sollten komplex sein), so erhalten wir als Wert für das zweite Integral (vgl. Formel 11 auf Seite 350)

$$\frac{1}{\sqrt{q-\frac{p^2}{4}}}\operatorname{arctg}\frac{x+\frac{p}{2}}{\sqrt{q-\frac{p^2}{4}}}+C.$$

Es bleibt also noch das Integral

$$I=\int\frac{Ax+B}{(x^2+px+q)^r}\,dx \qquad (r>1)$$

auszuwerten. Dazu gehen wir von der bereits erwähnten Identität

$$Ax+B=\frac{A}{2}(2x+p)+\left(B-\frac{A}{2}p\right)$$

aus, so daß also

$$I=\frac{A}{2}\int\frac{(2x+p)\,dx}{(x^2+px+q)^r}+\left(B-\frac{A}{2}p\right)\int\frac{dx}{(x^2+px+q)^r}.$$

Das erste Integral ist gleich

$$\frac{A}{2(1-r)}\cdot\frac{1}{(x^2+px+q)^{r-1}}+C,$$

während wir das zweite Integral auf die Form

$$\int\frac{d\left(x+\frac{p}{2}\right)}{\left[\left(x+\frac{p}{2}\right)^2+\left(q-\frac{p^2}{4}\right)\right]^r}$$

transformieren. Setzen wir hier $x+\frac{p}{2}=z$, $q-\frac{p^2}{4}=a^2$, so gelangen wir zu dem Integral

$$I_r=\int\frac{dz}{(z^2+a^2)^r}.$$

Offenbar ist nun

$$I_r=\frac{1}{a^2}\int\frac{z^2+a^2-z^2}{(z^2+a^2)^r}\,dz=\frac{1}{a^2}I_{r-1}-\frac{1}{a^2}\int\frac{z^2\,dz}{(z^2+a^2)^r}.$$

Zur Berechnung des Integrals

$$\int\frac{z^2\,dz}{(z^2+a^2)^r}$$

setzen wir $u=z$ und $dv=\frac{z\,dz}{(z^2+a^2)^r}$ (was auf $v=\frac{1}{2(1-r)}\cdot\frac{1}{(z^2+a^2)^{r-1}}$ führt) und integrieren partiell:

$$\int\frac{z^2\,dz}{(z^2+a^2)^r}=\frac{1}{2(1-r)}\cdot\frac{z}{(z^2+a^2)^{r-1}}-\frac{1}{2(1-r)}\int\frac{dz}{(z^2+a^2)^{r-1}}$$

$$=\frac{1}{2(1-r)}\cdot\frac{z}{(z^2+a^2)^{r-1}}-\frac{1}{2(1-r)}I_{r-1}.$$

Damit erhalten wir

$$I_r=\frac{1}{a^2}\left[1+\frac{1}{2(1-r)}\right]I_{r-1}-\frac{1}{2(1-r)a^2}\cdot\frac{z}{(z^2+a^2)^{r-1}}.$$

Wiederholen wir diese Reduktion hinreichend oft, so gelangen wir schließlich zu dem Integral

$$I_1=\int\frac{dz}{z^2+a^2}=\frac{1}{a}\operatorname{arctg}\frac{z}{a}+C.$$

In dem in Kleindruck gegebenen Beweis ist zugleich eine konstruktive Methode zur Berechnung des Integrals von Funktionen der Form (5) enthalten. In der Praxis verwendet man allerdings meistens kürzere Integrationsverfahren. Das wesentlichste derartige Integrationsverfahren stammt von M. W. Ostrogradski. Wir können hier jedoch nicht auf derartige spezielle Integrationsmethoden eingehen und verweisen dazu den Leser auf die entsprechende Fachliteratur.

Was die irrationalen Funktionen betrifft, so ist hier die Frage nach ihrer Integrierbarkeit in elementaren Funktionen wesentlich schwieriger zu beantworten. Wir wollen hier nur das folgende Resultat über die elementare Integrierbarkeit der sogenannten „binomischen Differentiale" mitteilen:

Satz von P. L. Tschebyschew. *Das Integral*

$$\int x^m (ax^n + b)^p \, dx$$

mit rationalen m, n und p ist genau dann elementar auswertbar, wenn entweder p oder $\frac{m+1}{n}$ oder $p + \frac{m+1}{n}$ ganz ist.

Auf den Beweis dieses Satzes wie auch auf weitere in diesem Zusammenhang auftretende Fragen, so die Frage nach der elementaren Integrierbarkeit transzendenter Funktionen, können wir hier nicht eingehen.

§ 5. Das bestimmte Integral

21. Probleme, die auf den Begriff des bestimmten Integrals führen. Der eigentlich grundlegende Begriff der Integralrechnung ist nicht der Begriff des unbestimmten, sondern der des bestimmten Integrals. Er ist wesentlich komplizierter als jener. Daher wollen wir zunächst eine Reihe von konkreten Problemen betrachten, die auf diesen Begriff führen.

I. Die Masse eines Stabes. In Abschnitt 1 haben wir bereits den Begriff der mittleren Dichte eines Stabes und den der wahren Dichte in einem Punkte betrachtet. Im Fall eines homogenen Stabes ist die wahre Dichte in jedem Punkte gleich der mittleren Dichte, während sich bei einem inhomogenen Stabe die Dichte p von Punkt zu Punkt ändert. Beschreibt man also die Lage eines Punktes M des Stabes durch seinen Abstand x von einem Ende des Stabes (Abb. 20), so kann man die Dichte p im Punkte x als eine Funktion $p = p(x)$ ansehen. Wir stellen uns jetzt die Aufgabe, die Masse m des Stabes bei gegebener Länge l und gegebener Funktion $p(x)$ zu bestimmen, wobei wir annehmen wollen, daß $p(x)$ eine stetige Funktion von x ist. Zu diesem Zweck denken wir uns den Stab durch

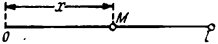


Abb. 20

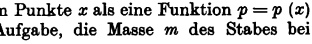


Abb. 21

Teilpunkte $x_1 < x_2 < \cdots < x_{n-1}$ $(0 < x_k < l)$ in n kleine Stücke unterteilt (Abb. 21). Der einheitlichen Bezeichnung wegen wollen wir sodann noch $x_0 = 0$ und $x_n = l$ setzen. Ferner sei λ der maximale Wert der Differenzen $x_{k+1} - x_k$. Ist λ hinreichend klein, so kann man die Teilstücke $[x_k, x_{k+1}]$ angenähert als homogen annehmen (da sich die stetige Funktion $p(x)$ auf diesen verhältnismäßig kleinen Teilstücken nur relativ wenig ändern kann). Dies bedeutet, daß auf jedem Teilstück die Funktion $p(x)$ als konstant angesetzt wird, etwa gleich $p(\xi_k)$, wobei ξ_k ein beliebiger Punkt aus dem Intervall $[x_k, x_{k+1}]$ ist. Die Masse des Teilstücks $[x_k, x_{k+1}]$ wird dann gleich $p(\xi_k)(x_{k+1} - x_k)$, und als Gesamtmasse des Stabes ergibt sich

$$m = \sum_{k=0}^{n-1} p(\xi_k)(x_{k+1} - x_k).$$

Der so erhaltene Wert für die Masse m stellt jedoch nur eine Näherung dar, denn in Wirklichkeit ist ja der Stab auch in den Teilstücken inhomogen. Je kürzer jedoch diese Teilstücke sind, d. h. je kleiner λ ist, um so genauer dürfte der erhaltene Wert den genauen Wert von m approximieren, so daß der genaue Wert von m durch

$$m = \lim_{\lambda \to 0} \sum_{k=0}^{n-1} p(\xi_k)(x_{k+1} - x_k) \tag{1}$$

gegeben wird.

II. Die Länge des bei einer Bewegung zurückgelegten Weges. Ein Punkt M bewege sich mit der Geschwindigkeit v auf einer gegebenen Geraden. Da sich im allgemeinen die Geschwindigkeit v im Laufe der Zeit ändern wird, ist v eine Funktion der Zeit t, also $v = v(t)$. Wir stellen uns nun die Aufgabe, die Länge s des Weges zu bestimmen, den der Punkt M im Verlaufe des Zeitintervalls $[a, b]$ zurücklegt, wobei wir wieder annehmen wollen, daß $v(t)$ eine stetige Funktion von t ist. Zu diesem Zweck teilen wir das Zeitintervall $[a, b]$ durch die Punkte $t_1 < t_2 < \cdots < t_{n-1}$ $(a < t_k < b)$ in n kurze Zeitintervalle. Der einheitlichen Bezeichnung wegen setzen wir dann $t_0 = a$ und $t_n = b$. Ferner sei $\lambda = \max(t_{k+1} - t_k)$. Da sich bei kleinem λ die Geschwindigkeit $v(t)$ (wegen der Stetigkeit der Funktion $v(t)$) in der kurzen Zeit von t_k bis t_{k+1} relativ wenig ändert, können wir sie für dieses Zeitintervall angenähert als konstant gleich $v(\tau_k)$ annehmen, wobei $\tau_k \in [t_k, t_{k+1}]$. Vom Standpunkt der Mechanik aus bedeutet dies, daß wir während der Zeit $[t_k, t_{k+1}]$ die Bewegung als gleichförmig annehmen. Der Weg, den der Punkt M während dieser Zeit zurücklegt, ist dann offenbar gleich $v(\tau_k)(t_{k+1} - t_k)$, so daß der Gesamtweg s durch

$$s = \sum_{k=0}^{n-1} v(\tau_k)(t_{k+1} - t_k)$$

gegeben wird. Dieser Näherungswert für s wird um so genauer sein, je kleiner λ ist. Daher wird der genaue Weg s durch

$$s = \lim_{\lambda \to 0} \sum_{k=0}^{n-1} v(\tau_k)(t_{k+1} - t_k) \tag{2}$$

gegeben.

III. Der Flächeninhalt eines krummlinigen Trapezes. Wir betrachten die ebene Figur, die durch die Kurven $y = 0$, $x = a$, $x = b$ und $y = f(x)$ begrenzt wird, wobei $f(x)$ eine im Intervall $a \leqq x \leqq b$ definierte stetige positive Funktion ist (Abb. 22). Eine solche Figur wird allgemein ein *krummliniges Trapez* genannt. Unsere Aufgabe soll es nun sein, den Flächeninhalt F eines derartigen krummlinigen Trapezes zu bestimmen.

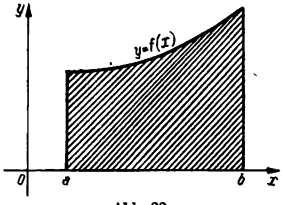


Abb. 22

Im Unterschied zu den vorangehenden Aufgaben müssen wir hier jedoch vor allem noch klären, was unter dem Flächeninhalt F verstanden werden soll. Wir werden in den folgenden Überlegungen hierauf besonders eingehen.

Zunächst denken wir uns das Intervall $[a, b]$ durch die Teilpunkte $a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b$ in n Teilintervalle zerlegt. Dabei sei $\lambda = \max(x_{k+1} - x_k)$. Durch die Geraden $x = x_k$ wird dann unser Trapez in n schmale Streifen zerlegt. Wegen der Stetigkeit der Funktion $f(x)$ ändern sich die Funktionswerte im Intervall $x_k \leqq x \leqq x_{k+1}$ relativ wenig, so daß man ohne einen allzugroßen Fehler $f(x)$ auf dem Intervall $[x_k, x_{k+1}]$ als konstant gleich $f(\xi_k)$ annehmen kann, wobei ξ_k ein beliebiger Punkt aus dem Intervall $[x_k, x_{k+1}]$ ist. Geometrisch bedeutet dies, daß man die Streifen des Trapezes als Rechtecke annimmt, also an Stelle des Trapezes die in Abb. 23 dargestellte treppenförmige Figur betrachtet. Der Flächeninhalt dieser treppenförmigen Figur wird nun offenbar durch

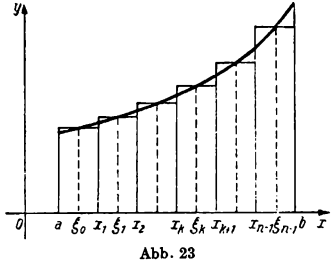


Abb. 23

$$\sum_{k=0}^{n-1} f(\xi_k)(x_{k+1} - x_k)$$

gegeben. Es dürfte nun als ganz natürlich erscheinen, daß man für kleines λ diesen Flächeninhalt als einen Näherungswert für den uns interessierenden Flächeninhalt F ansieht. Wir werden daher im Sinne einer Definition den Grenzwert

$$F = \lim_{\lambda \to 0} \sum_{k=0}^{n-1} f(\xi_k)(x_{k+1} - x_k) \tag{3}$$

als Flächeninhalt des betrachteten krummlinigen Trapezes bezeichnen, wobei natürlich gezeigt werden muß, daß dieser Grenzwert tatsächlich existiert (vgl. den folgenden Abschnitt).[1])

Ein Vergleich von (1), (2) und (3) zeigt, daß es sich in allen drei Fällen um die Lösung einer Aufgabe handelt, die vom mathematischen Standpunkt aus jeweils vom gleichen Typ ist. Wir werden uns daher im folgenden ganz allgemein mit dem Studium von Grenzprozessen dieses Typs, den sogenannten bestimmten Integralen, beschäftigen, ohne dabei auf ihre konkrete Bedeutung einzugehen.

22. Das bestimmte Integral. Wir kommen nun zur genauen Definition des Begriffs des bestimmten Integrals. Vorgegeben sei uns dazu eine beliebige, im abgeschlossenen Intervall $[a, b]$ definierte Funktion $f(x)$. Wir führen dann folgende Operationen aus:

1. Wir zerlegen das Intervall $[a, b]$ auf irgendeine Weise durch Einfügen von Teilpunkten

$$x_0 = a < x_1 < x_2 < \cdots < x_{n-1} < x_n = b$$

in endlich viele Teilintervalle $[x_k, x_{k+1}]$, wobei wir die maximale Länge dieser Teilintervalle, d. h. das Maximum der Differenzen $x_{k+1} - x_k$, mit λ bezeichnen.

2. Dann wählen wir aus jedem der Intervalle $[x_k, x_{k+1}]$ einen beliebigen Punkt ξ_k aus und bestimmen den Funktionswert $f(\xi_k)$.

3. Als nächstes multiplizieren wir den Wert $f(\xi_k)$ mit der Länge des entsprechenden Intervalls $[x_k, x_{k+1}]$.

4. Daraufhin wird die Summe aller dieser Produkte, d. h. die Summe

$$\sigma = \sum_{k=0}^{n-1} f(\xi_k)(x_{k+1} - x_k)$$

gebildet, die wir auch als eine „Integralsumme" der Funktion $f(x)$ bezeichnen wollen.

5. Schließlich ändern wir die willkürlich gewählte Zerlegung des Intervalls $[a, b]$ fortlaufend ab, und zwar so, daß dabei die Zahl λ gegen Null strebt (gehen also zu immer „feineren" Zerlegungen des Intervalls $[a, b]$ über).

Streben hierbei die zugehörigen Summen σ gegen einen gewissen endlichen Grenzwert I,

$$I = \lim_{\lambda \to 0} \sigma, \tag{4}$$

der von der Wahl der Zwischenpunkte ξ_k unabhängig ist, so nennt man diesen das *bestimmte Integral* der Funktion $f(x)$ im Intervall $[a, b]$ und bezeichnet ihn mit

$$\int_a^b f(x)\,dx.$$

[1]) In den vorangehenden Beispielen haben wir diesen Grenzwert als offensichtlich existent angenommen, weil die Masse m und die Länge s des zurückgelegten Weges objektiv existierende physikalische Größen sind.

Die Beziehung (4) soll also genau genommen folgendes besagen: Zu jedem $\varepsilon > 0$ gibt es ein $\delta > 0$, so daß bei jeder Zerlegung mit einer „Maximaldistanz" λ der Teilpunkte, welche kleiner als δ ist, für jede zugehörige Summe σ die Ungleichung

$$|\sigma - I| < \varepsilon$$

erfüllt ist, und zwar unabhängig davon, wie in der Summe σ die Zwischenpunkte $\xi_k \in [x_k, x_{k+1}]$ gewählt sind.

Wir sehen also, daß die Limesbeziehung (4) von recht eigentümlicher Natur ist. Es soll nun untersucht werden, für welche Funktionen dieser Begriff hinreichend natürlich ist.

Wenn für eine Funktion $f(x)$ dieser Grenzwert existiert (man sagt in diesem Fall, die Funktion $f(x)$ sei *integrierbar*), so bedeutet dies, daß für eine Zerlegung mit hinreichend kleinem λ jede Summe σ einer gewissen konstanten Zahl nahe kommt, die von der Auswahl der Zahlen ξ_k unabhängig ist, d. h., bei einer anderen Auswahl der Zwischenwerte ξ_k darf sich die Summe nicht wesentlich ändern. Das ist nun sicher dann der Fall, wenn eine Änderung der Punkte ξ_k keine bedeutende Änderung der Zahlen $f(\xi_k)$ bewirkt, wenn also die Funktion $f(x)$ stetig ist (man beachte, daß ξ_k jeweils nur in dem kleinen Intervall $[x_k, x_{k+1}]$ variieren kann). Wir werden also erwarten können, daß das bestimmte Integral für jede stetige Funktion existiert. Ist dagegen $f(x)$ unstetig, so wird man im allgemeinen nicht erwarten können, daß $f(x)$ integrierbar ist.

Die vorangehenden Überlegungen erhalten ihre Rechtfertigung durch den folgenden

Satz. *Für jede im Intervall $[a, b]$ stetige Funktion $f(x)$ existiert das bestimmte Integral*

$$\int_a^b f(x)\,dx.$$

Der strenge Beweis dieses Satzes ist recht verwickelt, so daß wir zunächst einige Hilfsbetrachtungen vorausschicken wollen.

Es sei $f(x)$ eine im Intervall $[a, b]$ stetige Funktion. Wir betrachten dann die folgende Zerlegung des Intervalls $[a, b]$

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b$$

und bezeichnen mit M_k bzw. m_k den größten bzw. kleinsten Wert der Funktion $f(x)$ im Intervall $[x_k, x_{k+1}]$. Die Summen

$$S = \sum_{k=0}^{n-1} M_k(x_{k+1} - x_k), \quad s = \sum_{k=0}^{n-1} m_k(x_{k+1} - x_k)$$

bezeichnet man als die zu der betrachteten Zerlegung gehörende *Ober-* bzw. *Untersumme* der Funktion $f(x)$.

Die Einführung dieser Summen wird dadurch nahegelegt, daß es sich bei ihnen um spezielle Summen σ handelt, die bereits eindeutig durch die betrach-

tete Zerlegung bestimmt sind, während im allgemeinen zur Festlegung einer Summe σ noch die Zwischenwerte ξ_k angegeben werden müssen. Dabei gilt für jede zur betrachteten Zerlegung gehörende Summe σ

$$s \leqq \sigma \leqq S. \tag{5}$$

Lemma. *Es sei $x_0 = a < x_1 < x_2 < \cdots < x_n = b$ eine beliebige Zerlegung des Intervalls $[a, b]$, und es seien S bzw. s die zu dieser Zerlegung gehörende Ober- bzw. Untersumme der Funktion $f(x)$. Sind dann S' und s' die Ober- bzw. Untersumme von $f(x)$ zu einer Zerlegung, die aus der gegebenen dadurch hervorgeht, daß man (unter Beibehaltung der alten Teilpunkte) neue Teilpunkte hinzufügt, so ist*

$$s \leqq s' \leqq S' \leqq S.$$

Mit anderen Worten: *Fügt man in eine Zerlegung weitere Teilpunkte ein, so werden die Untersummen höchstens größer und die Obersummen höchstens kleiner.*

Wir beweisen hier das angegebene Lemma nur für die Obersummen. Dabei können wir uns offenbar auf den Fall beschränken, daß ein einziger neuer Teilpunkt $\bar{x}$ hinzugenommen wird, da sich der allgemeine Fall als eine wiederholte Hinzunahme eines einzelnen Punktes auffassen läßt. Es sei dabei etwa $x_i < \bar{x} < x_{i+1}$. Dann entsteht die Obersumme S' dadurch aus der Obersumme S, daß man den in S auftretenden Summanden

$$M_i(x_{i+1} - x_i) \tag{6}$$

durch die Summe aus den beiden Summanden

$$M_i'(\bar{x} - x_i) + M_i''(x_{i+1} - \bar{x}) \tag{7}$$

ersetzt, wobei M_i' bzw. M_i'' der größte Wert von $f(x)$ im Intervall $[x_i, \bar{x}]$ bzw. $[\bar{x}, x_{i+1}]$ ist. Da nun $[x_i, \bar{x}]$ und $[\bar{x}, x_{i+1}]$ Teilintervalle von $[x_i, x_{i+1}]$ sind, ist $M_i' \leqq M_i$ und $M_i'' \leqq M_i$. Dann ist aber

$$M_i'(\bar{x} - x_i) + M_i''(x_{i+1} - \bar{x}) \leqq M_i[(\bar{x} - x_i) + (x_{i+1} - \bar{x})] = M_i(x_{i+1} - x_i),$$

d. h., (7) ist nicht größer als (6), so daß auch $S' \leqq S$.

Folgerung. *Keine Untersumme ist größer als irgendeine Obersumme.*

Wir betrachten dazu zwei beliebige Zerlegungen (I) und (II) des Intervalls $[a, b]$, und es sei s_1 die Untersumme zur Zerlegung (I) und S_2 die Obersumme zur Zerlegung (II). Wir betrachten dann die Zerlegung (III), deren Teilpunkte die Gesamtheit der Teilpunkte der Zerlegungen (I) und (II) ist. Die zu dieser Zerlegung gehörende Unter- und Obersumme sei s_3 bzw. S_3. Dann ist auf Grund des vorangehenden Lemma $s_1 \leqq s_3$ und $S_3 \leqq S_2$. Andererseits ist offenbar $s_3 \leqq S_3$. Hieraus folgt

$$s_1 \leqq S_2,$$

was zu beweisen war.

Wir betrachten nun eine feste Obersumme S_0. Dann gilt für jede Untersumme s:

$$s \leqq S_0.$$

Hieraus folgt, daß die Menge $\{s\}$ aller Untersummen, d. h. die Menge aller Zahlen, die als Untersumme bei einer beliebigen Zerlegung von $[a, b]$ auftreten, nach oben beschränkt ist, wobei S_0 eine obere Schranke für diese Menge ist. Daher besitzt diese Menge eine eindeutig bestimmte obere Grenze

$$I = \sup\{s\},$$

und es ist $I \leqq S_0$ für jede Obersumme S_0, d. h., es gilt

$$s \leqq I \leqq S, \tag{8}$$

wobei s und S beliebige Unter- und Obersummen von $f(x)$ sind.

Nach diesen Vorbereitungen können wir jetzt leicht zeigen, daß jede stetige Funktion integrierbar ist. Zu diesem Zweck seien s, S und σ zu einer beliebigen Zerlegung gehörende Summen (wobei die Summe σ zu einem beliebigen, aber im folgenden festen System von Zwischenwerten ξ_k gehört). Auf Grund von (5) und (8) ist dann

$$|\sigma - I| \leqq S - s. \tag{9}$$

Nun ist die nach Voraussetzung im Intervall $[a, b]$ stetige Funktion $f(x)$ sogar gleichmäßig stetig (vgl. S. 209). Daher gibt es zu jedem $\varepsilon > 0$ ein $\delta > 0$, so daß für beliebige Punkte x' und x'' aus $[a, b]$ gilt: Wenn $|x'' - x'| < \delta$, so

$$|f(x'') - f(x')| < \frac{\varepsilon}{b-a}.$$

Betrachten wir nun eine beliebige Zerlegung, für welche $\lambda < \delta$ ist, dann ist offenbar

$$M_k - m_k < \frac{\varepsilon}{b-a} \qquad (k = 0, 1, 2, \ldots, n-1)$$

und daher

$$S - s = \sum_{k=0}^{n-1} (M_k - m_k)(x_{k+1} - x_k) < \frac{\varepsilon}{b-a} \sum_{k=0}^{n-1} (x_{k+1} - x_k) = \varepsilon.$$

Hieraus und aus (9) folgt

$$|\sigma - I| < \varepsilon. \tag{10}$$

Damit ist gezeigt, daß es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt, so daß für jede Zerlegung, bei welcher $\lambda < \delta$ ist, die Ungleichung (10) gilt (und zwar unabhängig davon, wie in der Summe σ die Zwischenwerte ξ_k gewählt sind). Das bedeutet aber gerade, daß

$$I = \lim_{\lambda \to 0} \sigma,$$

wobei I die obere Grenze aller Untersummen ist.

Damit ist der behauptete Satz bewiesen.

Mit Hilfe dieses Satzes ergibt sich sofort, daß das im vorangehenden Abschnitt in Aufgabe III betrachtete krummlinige Trapez einen Flächeninhalt besitzt und daß dieser durch

$$F = \int_a^b f(x)\, dx$$

gegeben wird.

Lesen wir diese Formel von rechts nach links, so finden wir:

Geometrische Deutung des bestimmten Integrals. *Ist $f(x)$ eine im Intervall $[a, b]$ definierte stetige positive Funktion, so stellt*

$$\int_a^b f(x)\,dx$$

den Flächeninhalt des durch die Kurven $y = 0$, $x = a$, $x = b$ und $y = f(x)$ begrenzten krummlinigen Trapezes dar (Abb. 24).

Man darf nun aber nicht etwa glauben, daß die Voraussetzung der Stetigkeit der Funktion $f(x)$ für ihre Integrierbarkeit notwendig ist. Es gibt durchaus auch unstetige Funktionen, für welche das bestimmte Integral existiert.

Wir betrachten als Beispiel diejenige Funktion $f(x)$, welche in allen Punkten des Intervalls $[a, b]$ mit Ausnahme der Punkte $z_1, z_2, \ldots, z_N$ verschwindet, während sie in diesen Punkten beliebig vorgegebene Werte annimmt. Es sei dann σ eine zu einer beliebigen Zerlegung des Intervalls $[a, b]$ gehörende „Integralsumme", wobei etwa p der in die Summe eingehenden Punkte $\xi_0, \xi_1, \ldots, \xi_{n-1}$ mit gewissen der Punkte z_i zusammenfallen, während die übrigen von allen Punkten z_i verschieden sind. Dann sind in der Summe σ nur p Summanden von Null verschieden, und es ist daher, wenn K die größte der Zahlen $|f(z_i)|$ $(i = 1, 2, \ldots, N)$ bezeichnet,

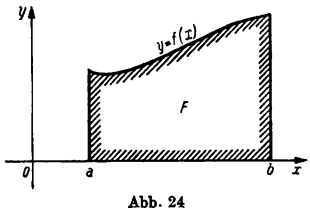


Abb. 24

$$|\sigma| \leqq Kp\lambda \leqq KN\lambda.$$

Hieraus folgt jedoch sofort, daß σ mit λ gegen 0 strebt. Also existiert das Integral

$$\int_a^b f(x)\,dx$$

und ist gleich Null.

Wir bringen schließlich noch ein Beispiel[1]) für eine nicht integrierbare Funktion.

Es sei auf dem Intervall $[0, 1]$

$$\varphi(x) = \begin{cases} 0, & \text{falls } x \text{ irrational ist,} \\ 1, & \text{falls } x \text{ rational ist.} \end{cases}$$

Verwenden wir dann bei der Bildung einer Summe σ nur irrationale Zwischenwerte ξ_k, so wird $\sigma = 0$; verwenden wir dagegen nur rationale Zwischenwerte,

[1]) Es stammt von dem deutschen Mathematiker G. P. Lejeune-Dirichlet (1805 — 1859).

so wird $\sigma = 1$. Aus diesem Grunde ist es unmöglich, daß σ für $\lambda \to 0$ gegen einen wohlbestimmten Grenzwert strebt, d. h., das Integral

$$\int_0^1 \varphi(x)\,dx$$

existiert nicht.

Man kennt heute notwendige und hinreichende Kriterien dafür, ob eine Funktion integrierbar ist oder nicht. Wir wollen uns jedoch hier auf den oben angegebenen Satz über die Integrierbarkeit der stetigen Funktionen beschränken.

23. Die wichtigsten Eigenschaften des Integrals. Im vorliegenden Abschnitt wollen wir die wichtigsten Eigenschaften des bestimmten Integrals zusammenstellen. Der größte Teil der folgenden Sätze gilt für beliebige integrierbare Funktionen; wir wollen uns jedoch hier stets nur auf stetige Funktionen beschränken.

Satz 1. *Es seien $f(x)$ und $g(x)$ zwei auf dem abgeschlossenen Intervall $[a, b]$ definierte und stetige Funktionen. Dann gilt*

$$\int_a^b [f(x) + g(x)]\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx,$$

d. h., das Integral einer Summe ist gleich der Summe der Integrale der Summanden.

Beweis. Offenbar gilt für jede zur Funktion $f(x) + g(x)$ gehörende Integralsumme

$$\sum_{k=0}^{n-1} [f(\xi_k) + g(\xi_k)]\,(x_{k+1} - x_k) = \sum_{k=0}^{n-1} f(\xi_k)\,(x_{k+1} - x_k) + \sum_{k=0}^{n-1} g(\xi_k)\,(x_{k+1} - x_k).$$

Gehen wir hier zur Grenze $\lambda \to 0$ über, so erhalten wir die Behauptung.

Analog beweist man:

Satz 2. *Für jede im Intervall $[a, b]$ stetige Funktion $f(x)$ und jede Konstante c gilt*

$$\int_a^b c f(x)\,dx = c \int_a^b f(x)\,dx,$$

d. h., ein konstanter Faktor kann vor das Integral gezogen werden.

Satz 3. *Es sei $f(x)$ eine im Intervall $[a, b]$ stetige Funktion. Zerlegt man das Intervall $[a, b]$ durch einen Punkt c in die beiden abgeschlossenen Teilintervalle $[a, c]$ und $[c, b]$, so ist das Integral über das ganze Intervall gleich der Summe der Integrale über die Teilintervalle, d. h.*

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx \qquad (a < c < b).$$

Beweis. Wir betrachten eine beliebige Zerlegung des Intervalls $[a, b]$, welche den Punkt c als Teilpunkt enthält, und zwar sei etwa $c = x_m$. Dann ist

$$\sum_{k=0}^{n-1} f(\xi_k)\,(x_{k+1} - x_k) = \sum_{k=0}^{m-1} f(\xi_k)\,(x_{k+1} - x_k) + \sum_{k=m}^{n-1} f(\xi_k)\,(x_{k+1} - x_k).$$

Jede der hier auftretenden Summen ist eine Integralsumme, und zwar bzw. für das Intervall $[a, b]$, $[a, c]$, $[c, b]$. Gehen wir in dieser Gleichung zur Grenze $\lambda \to 0$ über, so erhalten wir die Behauptung.[1])

Der zuletzt bewiesene Satz läßt sich in noch allgemeinerer Form aussprechen. Dazu empfiehlt es sich, den Sinn des Integralzeichens etwas zu erweitern. Bisher ist das Integral $\int_a^b f(x)\,dx$ nur für den Fall definiert, daß $a < b$ und die Funktion $f(x)$ im Intervall $[a, b]$ integrierbar ist.

Es sei nun $f(x)$ eine im Punkte a erklärte Funktion; dann definieren wir zusätzlich

$$\int_a^a f(x)\,dx = 0. \tag{11}$$

In Worten: *Ein Integral mit übereinstimmenden Integrationsgrenzen ist gleich Null.*

Ist die Funktion $f(x)$ im Intervall $[a, b]$ integrierbar, so definieren wir ferner

$$\int_b^a f(x)\,dx = -\int_a^b f(x)\,dx. \tag{12}$$

In Worten: *Vertauscht man bei einem bestimmten Integral die Integrationsgrenzen, so ändert das Integral sein Vorzeichen.*

Die erwähnte Verschärfung von Satz 3 kann dann folgendermaßen formuliert werden:

Satz 4. *Es sei $f(x)$ eine im Intervall $[A, B]$ stetige Funktion. Es seien ferner a, b, c irgendwelche Punkte aus diesem Intervall. Dann gilt*

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx. \tag{13}$$

Beweis. Fallen von den Punkten a, b, c zwei Punkte zusammen (oder sind sogar alle drei gleich), so gilt (13) trivial auf Grund von (11). Es seien also die Punkte a, b, c paarweise verschieden. Der Fall $a < c < b$ läuft offenbar gerade auf Satz 3 hinaus. In den anderen Fällen läßt sich dagegen die Be-

[1]) Die Beschränkung auf Zerlegungen, welche c als Teilpunkt enthalten, bedeutet dabei keine Einschränkung der Allgemeinheit; denn der Grenzwert der Integralsummen für $\lambda \to 0$ ändert sich offenbar nicht, wenn man sich in jede Zerlegung, welche c nicht als Teilpunkt enthält, c als neuen Teilpunkt eingeschaltet denkt. — *Anm. d. wissenschaftl. Red.*

hauptung leicht unter Verwendung von (12) auf Satz 3 zurückführen. In der Tat: Ist z. B. $c < b < a$, so ist

$$\int_c^a f(x)\,dx = \int_c^b f(x)\,dx + \int_b^a f(x)\,dx$$

und daher

$$-\int_b^a f(x)\,dx = -\int_c^a f(x)\,dx + \int_c^b f(x)\,dx,$$

woraus sich durch zweimalige Anwendung von (12) sofort (13) ergibt.

Die in den Sätzen 3 und 4 ausgesprochene Eigenschaft des bestimmten Integrals bezeichnet man auch als *Additivität des Integrals in bezug auf das Integrationsintervall.*

Satz 5. *Es sei $f(x)$ eine im Intervall $[a, b]$ definierte stetige Funktion. Dann gibt es stets einen Punkt $\xi \in [a, b]$, so daß*

$$\int_a^b f(x)\,dx = f(\xi)\,(b-a). \tag{14}$$

Beweis. Es bezeichne M bzw. m den größten bzw. kleinsten Wert, den die Funktion $f(x)$ im Intervall $[a, b]$ annimmt. Wir betrachten dann eine beliebige Integralsumme

$$\sigma = \sum_{k=0}^{n-1} f(\xi_k)\,(x_{k+1} - x_k).$$

Für jedes k ist hier $m \leqq f(\xi_k) \leqq M$, so daß wegen $x_{k+1} > x_k$ auch

$$m(x_{k+1} - x_k) \leqq f(\xi_k)(x_{k+1} - x_k) \leqq M(x_{k+1} - x_k).$$

Summieren wir alle diese Ungleichungen und beachten wir, daß

$$\sum_{k=0}^{n-1} (x_{k+1} - x_k) = b - a,$$

so erhalten wir

$$m(b-a) \leqq \sigma \leqq M(b-a).$$

Vollziehen wir hier den Grenzübergang $\lambda \to 0$ und dividieren wir anschließend durch $b - a$, so finden wir

$$m \leqq \frac{1}{b-a} \int_a^b f(x)\,dx \leqq M.$$

Das bedeutet also, daß die Zahl

$$h = \frac{1}{b-a} \int_a^b f(x)\,dx$$

zwischen dem kleinsten und dem größten Wert der als stetig vorausgesetzten Funktion $f(x)$ liegt. Dann gibt es aber bekanntlich (vgl. S. 205) zwischen den Punkten, in denen $f(x)$ den Wert m bzw. M annimmt, also erst recht im Intervall $[a, b]$ einen Punkt ξ, so daß $h = f(\xi)$. Das ist aber gerade die Behauptung von Satz 5.

Wir bemerken, daß (14) auch im Fall $a = b$ (in diesem Fall sind beide Seiten von (14) gleich Null) und im Fall $b > a$ (in diesem Fall führt eine Vorzeichenänderung auf beiden Seiten von (14) auf Satz 5 zurück) richtig ist. Im zuerst genannten Fall ist $\xi = a$, im zweiten ist $a \geqq \xi \geqq b$.

Satz 5 wird in der Literatur meistens als erster Mittelwertsatz der Integralrechnung bezeichnet. Aus ihm ergeben sich eine ganze Reihe von weiteren Eigenschaften des bestimmten Integrals.

Satz 6. *Es sei $f(x)$ eine nicht negative stetige Funktion. Dann ist im Fall $a \leqq b$*[1])

$$\int_a^b f(x)\,dx \geqq 0.$$

In der Tat: Unter den angegebenen Voraussetzungen sind in (14) beide Faktoren auf der rechten Seite nicht negativ.

Dieses zuletzt erhaltene Resultat kann man noch etwas verschärfen:

Satz 7. *Es sei $f(x)$ eine im Intervall $[a, b]$ stetige nicht negative Funktion, welche in wenigstens einem Punkte des Intervalls $[a, b]$ von Null verschieden ist. Dann ist im Fall $a < b$*

$$\int_a^b f(x)\,dx > 0.$$

Beweis. Es sei x_0 ein Punkt aus dem Intervall, in welchem $f(x_0) > 0$ ist. Dann wählen wir eine Zahl $\delta > 0$ so klein, daß für $|x - x_0| < \delta$ überall $f(x) > 0$ ist. Dies ist möglich, da nach Voraussetzung die Funktion $f(x)$ im Intervall $[a, b]$ stetig ist. Ohne Beschränkung der Allgemeinheit können wir annehmen, daß $a \leqq x_0 - \delta < x_0 + \delta \leqq b$. Dann ist

$$\int_a^b f(x)\,dx = \int_a^{x_0-\delta} f(x)\,dx + \int_{x_0-\delta}^{x_0+\delta} f(x)\,dx + \int_{x_0+\delta}^b f(x)\,dx.$$

Nach Satz 6 sind der erste und dritte Summand auf der rechten Seite nicht negativ, während der zweite Summand auf die Form

$$\int_{x_0-\delta}^{x_0+\delta} f(x)\,dx = f(\xi)\cdot 2\delta \qquad (x_0 - \delta \leqq \xi \leqq x_0 + \delta)$$

gebracht werden kann und daher echt positiv ist.

---

[1]) Erfüllen beim Integral $\int_a^b f(x)\,dx$ die Grenzen die Ungleichung $a \leqq b$, so wollen wir auch sagen, daß sich die Integrationsgrenzen *in normaler Anordnung* befinden.

Satz 7 kann man offenbar auch folgendermaßen formulieren:

Satz 8. *Es sei $f(x)$ eine im Intervall $[a, b]$ stetige nicht negative Funktion (wobei $a < b$). Ist dann*

$$\int_a^b f(x)\,dx = 0,$$

so ist $f(x)$ im Intervall $[a, b]$ identisch gleich Null.

In den Sätzen 7 und 8 kann man (im Unterschied zu Satz 6) nicht auf die Voraussetzung der Stetigkeit der Funktion $f(x)$ verzichten. Ist z. B. die Funktion $f(x)$ in endlich vielen Punkten des Intervalls $[a, b]$ gleich Eins und in den übrigen Punkten gleich Null, so ist sie offenbar nicht negativ und nicht identisch Null, während ihr Integral (wie bereits in Abschnitt 22 gezeigt wurde) verschwindet.

Satz 9. *Es seien $f(x)$ und $g(x)$ zwei im Intervall $[a, b]$ stetige Funktionen (wobei $a \leqq b$), für welche überall die Ungleichung $f(x) \leqq g(x)$ erfüllt ist. Dann gilt*

$$\int_a^b f(x)\,dx \leqq \int_a^b g(x)\,dx. \tag{15}$$

Das heißt mit anderen Worten: *Bei normaler Anordnung der Integrationsgrenzen kann man die Ungleichung $f(x) \leqq g(x)$ gliedweise integrieren.*

Beweis. Unter den angegebenen Voraussetzungen ist $g(x) - f(x)$ eine im Intervall $[a, b]$ stetige nicht negative Funktion, so daß auf Grund von Satz 6 gilt:

$$\int_a^b g(x)\,dx - \int_a^b f(x)\,dx = \int_a^b [g(x) - f(x)]\,dx \geqq 0.$$

Setzen wir voraus, daß in wenigstens einem Punkte des Intervalls $[a, b]$ sogar $f(x) < g(x)$ ist, so können wir im Fall $a < b$ in (15) auch noch das Gleichheitszeichen ausschließen.

Satz 10. *Es sei $f(x)$ eine im Intervall $[a, b]$ stetige Funktion (wobei $a \leqq b$). Dann gilt*

$$\left| \int_a^b f(x)\,dx \right| \leqq \int_a^b |f(x)|\,dx. \tag{16}$$

Das heißt mit anderen Worten: *Bei normaler Anordnung der Integrationsgrenzen ist der absolute Betrag eines Integrals nicht größer als das Integral über den absoluten Betrag des Integranden.*

Beweis. Wegen

$$-|f(x)| \leqq f(x) \leqq |f(x)|$$

ist auf Grund von Satz 9

$$-\int_a^b |f(x)|\,dx \leqq \int_a^b f(x)\,dx \leqq \int_a^b |f(x)|\,dx;$$

das ist aber gerade die Behauptung von Satz 10.

Aus dem ersten Mittelwertsatz erhalten wir schließlich noch unmittelbar die folgende wichtige Integralabschätzung:

Satz 11. *Wenn eine im Intervall $[a, b]$ stetige Funktion $f(x)$ in allen Punkten des Intervalls $[a, b]$ die Ungleichung $|f(x)| \leqq K$ erfüllt, so ist*

$$\left| \int_a^b f(x)\,dx \right| \leqq K\,|b-a|.$$

Wir weisen zum Abschluß noch einmal ausdrücklich darauf hin, daß das bestimmte Integral eine Zahl ist, welche durch die Integrationsgrenzen und den Integranden eindeutig bestimmt ist. Daraus folgt, daß beim bestimmten Integral die Bezeichnung der Integrationsveränderlichen ohne Belang ist, das heißt

$$\int_a^b f(x)\,dx, \quad \int_a^b f(z)\,dz, \quad \int_a^b f(t)\,dt, \quad \int_a^b f(u)\,du$$

bezeichnen alle dieselbe Zahl.

Es ist durchaus nicht überflüssig, dies hervorzuheben, denn beim unbestimmten Integral liegen die Dinge ganz anders. Zum Beispiel ist

$$\int x^2\,dx = \frac{x^3}{3} + C, \quad \int z^2\,dz = \frac{z^3}{3} + C.$$

Wenn wir also etwa das Integral

$$\int \sin^2 x \cos x\,dx$$

durch die Substitution $\sin x = z$ auf die Form

$$\int z^2\,dz$$

bringen, so erhalten wir als Resultat nicht etwa $\frac{x^3}{3} + C$, sondern $\frac{z^3}{3} + C$, wobei sodann z durch $\sin x$ zu ersetzen ist.

24. Das Integral als Funktion der oberen Grenze. Im vorangehenden haben wir die Eigenschaften des bestimmten Integrals bei konstanten Integrationsgrenzen untersucht. Im vorliegenden Abschnitt wollen wir sehen, wie sich das Integral bei veränderlichen Integrationsgrenzen verhält. Dazu sei $f(t)$ eine beliebige im Intervall $[a, b]$ definierte und dort stetige Funktion. Dann ist $f(t)$ insbesondere auch in jedem Teilintervall $[a, x]$ $(a \leqq x \leqq b)$ stetig, so daß wir für jedes x mit $a \leqq x \leqq b$ das Integral

$$\int_a^x f(t)\,dt$$

bilden können. Offenbar wird dadurch jeder Zahl x aus dem Intervall $[a, b]$ eine eindeutig bestimmte Zahl zugeordnet, d. h., wir können dieses Integral als Funktion des Arguments x ansehen (um Verwechslungen mit der Integrationsgrenze zu vermeiden, haben wir jetzt die Integrationsveränderliche

mit t bezeichnet; wie am Schluß des vorangehenden Abschnitts bemerkt wurde, ist die Art der Bezeichnung der Integrationsveränderlichen ohne Belang).

Es gilt nun der folgende grundlegende Satz[1]), den man häufig direkt als Hauptsatz der Differential- und Integralrechnung bezeichnet:

Satz. *Es sei $f(t)$ eine im Intervall $[a, b]$ stetige Funktion. Dann ist das bestimmte Integral*

$$\int_a^x f(t)\,dt$$

als Funktion der oberen (Integrations-) Grenze x in allen Punkten des Intervalls $[a, b]$ differenzierbar, und zwar ist die Ableitung im Punkte x gleich dem Wert der Funktion $f(t)$ an dieser Stelle, d. h. gleich $f(x)$.

In einer Formel ausgedrückt, besagt dies also:

$$\left(\int_a^x f(t)\,dt\right)' = f(x).$$

Setzen wir

$$\int_a^x f(t)\,dt = \Phi(x),$$

so können wir dafür auch kurz

$$\Phi'(x) = f(x). \tag{17}$$

schreiben.

Wir beginnen mit einer zwar nicht strengen, dafür aber sehr anschaulichen geometrischen Überlegung, die das Wesen dieses Satzes klarmachen soll.

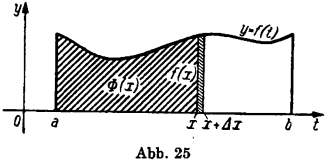


Abb. 25

Ist die als stetig vorausgesetzte Funktion $f(t)$ im Intervall $[a, b]$ überall positiv, so bedeutet $\Phi(x)$ geometrisch den Flächeninhalt des krummlinigen Trapezes, welches von den Kurven $y = 0$, $t = a$, $t = x$ und $y = f(t)$ begrenzt wird (Abb. 25). Erteilen wir nun dem Argument x einen gewissen Zuwachs Δx (der Einfachheit halber werde $\Delta x > 0$ angenommen), so erfährt dabei der Funktionswert $\Phi(x)$ einen entsprechenden Zuwachs $\Delta\Phi$, den wir uns geometrisch als Flächeninhalt eines schmalen Streifens veranschaulichen können (in der Zeichnung eng schraffiert). Ersetzen wir diesen Streifen angenähert durch das Rechteck mit der Grundlinie Δx und der Höhe $f(x)$ (wobei x der betrachtete Differentiationspunkt ist), so erhalten wir für $\Delta\Phi$ die Näherungsgleichung

$$\Delta\Phi \approx f(x)\,\Delta x,$$

[1]) Er wurde im Jahre 1669 von dem englischen Mathematiker I. Barrow (1630 – 1677) bewiesen.

die auch folgendermaßen geschrieben werden kann:

$$\frac{\Delta \Phi}{\Delta x} \approx f(x).$$

Die Genauigkeit dieser Gleichung wird nun um so größer sein, je kleiner Δx ist, d. h.

$$f(x) = \lim_{\Delta x \to 0} \frac{\Delta \Phi}{\Delta x},$$

und das ist gerade die Aussage von Formel (17).

Nach diesen Überlegungen wenden wir uns einem strengen Beweis für den genannten Satz zu. Dazu betrachten wir zwei beliebige Punkte x und $x + \Delta x$ aus dem Intervall $[a, b]$. Dann ist

$$\Phi(x) = \int_a^x f(t)\, dt,$$

$$\Phi(x + \Delta x) = \int_a^{x+\Delta x} f(t)\, dt = \int_a^x f(t)\, dt + \int_x^{x+\Delta x} f(t)\, dt.$$

Hieraus folgt auf Grund des ersten Mittelwertsatzes der Integralrechnung:

$$\Phi(x + \Delta x) - \Phi(x) = \int_x^{x+\Delta x} f(t)\, dt = f(\xi)\, \Delta x,$$

wobei ξ zwischen x und $x + \Delta x$ liegt. Also ist

$$\frac{\Phi(x + \Delta x) - \Phi(x)}{\Delta x} = f(\xi).$$

Strebt Δx gegen Null, so strebt offenbar ξ gegen x und daher, wegen der Stetigkeit der Funktion $f(t)$, auch $f(\xi)$ gegen $f(x)$. Also gilt:

$$\lim_{\Delta x \to 0} \frac{\Phi(x + \Delta x) - \Phi(x)}{\Delta x} = f(x).$$

Damit ist der genannte Satz bewiesen.

Aus dem Hauptsatz der Differential- und Integralrechnung folgt speziell, daß $\Phi(x)$ *stetig ist.*

Als Beispiele für den Hauptsatz nennen wir:

$$\left(\int_2^x e^{t^2} dt\right)' = e^{x^2}, \quad \left(\int_0^x \sqrt{\cos z}\, dz\right)' = \sqrt{\cos x}, \quad \left(\int_1^z \frac{dx}{x}\right)' = \frac{1}{z}.$$

In Abschnitt 17 haben wir ohne Beweis mitgeteilt, daß jede in einem Intervall stetige Funktion dort auch eine Stammfunktion besitzt. Das ist jetzt offensichtlich; denn ist die Funktion $f(x)$ im Intervall $[a, b]$ stetig, so ist

$$\Phi(x) = \int_a^x f(t)\, dt$$

eine Stammfunktion zu $f(x)$ im Intervall $[a, b]$. Ist die Funktion $f(x)$ positiv, so können wir aus einer gegebenen graphischen Darstellung für $y = f(t)$ leicht eine geometrische Darstellung für die Funktion $\Phi(x)$ erhalten (vgl. Abb. 26).

25. Berechnung des bestimmten mit Hilfe des unbestimmten Integrals. Es sei $f(x)$ eine im Intervall $[a, b]$ stetige Funktion. Wir wollen jetzt untersuchen, wie man das Integral

$$I = \int_a^b f(x)\,dx$$

berechnen kann.

Zu diesem Zweck führen wir zunächst eine neue Integrationsveränderliche ein, schreiben also das Integral I etwa in der Form

$$I = \int_a^b f(t)\,dt.$$

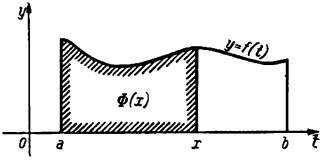


Abb. 26

Sodann betrachten wir allgemein die Funktion

$$\Phi(x) = \int_a^x f(t)\,dt.$$

Man sieht unmittelbar, daß $I = \Phi(b)$. Wenn wir also die Funktion $\Phi(x)$ kennen, so können wir mit ihrer Hilfe unmittelbar das Integral I berechnen. Auf den ersten Blick sieht es nun so aus, als ob das Problem, die Funktion $\Phi(x)$ zu bestimmen, wesentlich komplizierter ist als das Problem, lediglich die Zahl I zu berechnen. Das ist aber nicht unbedingt der Fall; denn wir werden sehen, daß wir zur Lösung des allgemeineren Problems Methoden zur Verfügung haben, die wir bei der ursprünglichen speziellen Fragestellung nicht verwenden können. Es sei in diesem Zusammenhang darauf hingewiesen, daß man in der Wissenschaft häufig vor die Situation gestellt wird, daß eine allgemeinere Fragestellung wesentliche neue Wege zur Lösung einer speziellen Fragestellung liefert.

In dem von uns betrachteten Fall können wir insbesondere die im vorangehenden Abschnitt bewiesene Beziehung

$$\Phi'(x) = f(x)$$

verwenden. Auf Grund dieser Beziehung ist also $\Phi(x)$ eine Stammfunktion des Integranden $f(x)$.

Wir wollen nun annehmen, daß wir bereits über irgendeine Stammfunktion $F(x)$ des Integranden $f(x)$ verfügen. Dann erhalten wir in

$$\int f(x)\,dx = F(x) + C \tag{18}$$

die Gesamtheit aller Stammfunktionen, so daß also

$$\Phi(x) = F(x) + C_0, \tag{19}$$

wobei C_0 eine passende Konstante ist.

Unter den angegebenen Bedingungen besteht also unsere einzige Aufgabe darin, diese Konstante C_0 zu bestimmen. Hierzu beachten wir, daß ein Integral mit übereinstimmenden Integrationsgrenzen gleich Null ist, so daß also

$$\Phi(a) = \int_a^a f(t)\,dt = 0.$$

Hieraus und aus (19) ergibt sich

$$C_0 = -F(a)$$

und damit

$$\Phi(x) = F(x) - F(a).$$

Setzen wir hier speziell $x = b$, so erhalten wir

$$\int_a^b f(x)\,dx = F(b) - F(a). \tag{20}$$

Durch diese Formel (die sogenannte Leibniz-Newtonsche *Formel*) wird die Berechnung des bestimmten Integrals einer stetigen Funktion $f(x)$ auf das Problem, eine Stammfunktion zu $f(x)$ zu bestimmen, zurückgeführt. Auf diese Weise wird ein wirklicher Zusammenhang zwischen den beiden Hauptgebieten der Analysis, der Differentialrechnung (zu welcher man ja genaugenommen den Begriff der Stammfunktion rechnen muß) und der Integralrechnung (in der im wesentlichen die Theorie des bestimmten Integrals behandelt wird), hergestellt. Bereits gegen Ende des 17. Jahrhunderts waren die wichtigsten Prinzipien sowohl der Differential- als auch der Integralrechnung ausgearbeitet, aber es war noch nicht klar, daß Differential- und Integralrechnung miteinander zusammenhängen. Es ist gerade das große Verdienst von Newton und Leibniz, daß sie diesen Zusammenhang erkannt haben. Es dürfte jetzt auch einleuchten, warum man den in Abschnitt 24 bewiesenen Satz, durch den ja gerade dieser Zusammenhang hergestellt wird, als Hauptsatz der Differential- und Integralrechnung bezeichnet.

Wegen der grundsätzlichen Bedeutung der vorangehenden Resultate wollen wir diese noch in Form einer Regel vermerken:

Regel. *Zur Berechnung des bestimmten Integrals einer stetigen Funktion genügt es, eine zu dieser Funktion gehörende Stammfunktion zu bestimmen und die Differenz der Werte dieser Stammfunktion an den Integrationsgrenzen zu bilden.*

Bei den vorangehenden Überlegungen haben wir stets stillschweigend vorausgesetzt, daß $a < b$ ist. Das ist aber nicht wesentlich. Im Fall $a = b$ ist nämlich die Formel (20) trivialerweise richtig, da hier beide Seiten gleich Null sind. Aber auch im Fall $a > b$ gilt Formel (20), und zwar kann man diesen Fall dadurch auf den Fall $a < b$ zurückführen, daß man beide Seiten von (20) mit (-1) multipliziert.

Bezeichnet man, wie es häufig üblich ist, die Differenz $F(b) - F(a)$ mit

$$[F(x)]_a^b,$$

so kann man (20) auch in der Form

$$\int_a^b f(x)\,dx = [F(x)]_a^b \tag{21}$$

schreiben. Da hierbei $F(x)$ eine beliebige Stammfunktion zu $f(x)$ sein darf, kann man hierfür schließlich auch

$$\int_a^b f(x)\,dx = \left[\int f(x)\,dx\right]_a^b$$

schreiben. In dieser Form kommt der Zusammenhang zwischen bestimmten und unbestimmten Integralen besonders deutlich zum Ausdruck.

Als Beispiele für das vorangehende nennen wir:

1. $\int_{\frac{\pi}{6}}^{\frac{\pi}{2}} \cos x\,dx = \left[\sin x\right]_{\frac{\pi}{6}}^{\frac{\pi}{2}} = \sin\frac{\pi}{2} - \sin\frac{\pi}{6} = 1 - \frac{1}{2} = \frac{1}{2}$,

2. $\int_1^4 x^2\,dx = \left[\frac{x^3}{3}\right]_1^4 = \frac{4^3}{3} - \frac{1^3}{3} = 21$,

3. $\int_0^1 \frac{dx}{1+x^2} = [\operatorname{arctg} x]_0^1 = \operatorname{arctg} 1 - \operatorname{arctg} 0 = \frac{\pi}{4}$.

Es versteht sich von selbst, daß man alle Berechnungsverfahren für unbestimmte Integrale auch als Berechnungsverfahren für bestimmte Integrale verwenden kann. Insbesondere wollen wir dies für die Substitution und die partielle Integration näher ausführen, da diese Verfahren bei Anwendung auf bestimmte Integrale noch etwas spezialisiert werden können.

Satz 1. *Es sei $f(z)$ eine im Intervall $[p, q]$ definierte und dort überall stetige Funktion. Es sei ferner $\varphi(x)$ eine im Intervall $[a, b]$ definierte und dort überall differenzierbare Funktion, deren sämtliche Werte die Ungleichung $p \leqq \varphi(x) \leqq q$ erfüllen. Dann gilt*

$$\int_a^b f(\varphi(x))\,\varphi'(x)\,dx = \int_{\varphi(a)}^{\varphi(b)} f(z)\,dz. \tag{22}$$

Formel (22) drückt die allgemeine Substitutionsregel für bestimmte Integrale aus. Sie erinnert in gewisser Weise an die Substitutionsregeln für unbestimmte Integrale, jedoch mit dem Unterschied, daß hier eine Rückkehr zur alten Veränderlichen unnötig ist, da Formel (22) eine Gleichung für Zahlen ist. Wir weisen darauf hin, daß diese eine Substitutionsregel für bestimmte Integrale beide Substitutionsregeln für unbestimmte Integrale ersetzt, da man bei praktischen Anwendungen die Formel (22) sowohl von links nach rechts als auch von rechts nach links lesen kann.

Wir kommen jetzt zum Beweis dieses Satzes. Dazu bezeichnen wir die linke und die rechte Seite von (22) mit I_1 bzw. I_2. Es sei dann $F(z)$ eine beliebige Stammfunktion zu $f(z)$. Dann ist

$$I_2 = F(\varphi(b)) - F(\varphi(a)). \tag{23}$$

Andererseits ist

$$I_1 = \left[\int f(\varphi(x))\,\varphi'(x)\,dx\right]_a^b.$$

Auf Grund des Satzes aus Abschnitt 18 ist nun aber

$$\int f(\varphi(x))\,\varphi'(x)\,dx = F(\varphi(x)) + C,$$

so daß

$$I_1 = F(\varphi(b)) - F(\varphi(a)).$$

Hieraus und aus (23) folgt $I_1 = I_2$, was zu beweisen war.

Beispiel 1.

$$\int_1^e \frac{dx}{x(1+\ln^2 x)} = \int_0^1 \frac{dz}{1+z^2} = \frac{\pi}{4}.$$

Hier haben wir die Substitution $\ln x = z$ angewendet (wobei Formel (22) von links nach rechts gelesen wurde).

Beispiel 2.

$$\int_0^R \sqrt{R^2 - x^2}\,dx = \int_0^{\frac{\pi}{2}} R^2 \cos^2 t\,dt.$$

Hier haben wir die Substitution $x = R\sin t$ angewendet (wobei Formel (22) von rechts nach links gelesen wurde). Die weitere Berechnung bereitet jetzt keinerlei Schwierigkeiten mehr. Es ist nämlich

$$\int_0^{\frac{\pi}{2}} \cos^2 t\,dt = \int_0^{\frac{\pi}{2}} \frac{1+\cos 2t}{2}\,dt = \left[\frac{t}{2} + \frac{\sin 2t}{4}\right]_0^{\frac{\pi}{2}} = \frac{\pi}{4}$$

und daher

$$\int_0^R \sqrt{R^2 - x^2}\,dx = \frac{1}{4}\pi R^2.$$

Satz 2. *Es seien $u(x)$ und $v(x)$ zwei im Intervall $[a, b]$ definierte Funktionen, welche dort stetige Ableitungen besitzen. Dann ist*

$$\int_a^b u\,dv = [uv]_a^b - \int_a^b v\,du. \tag{24}$$

Formel (24) ist die Formel für die partielle Integration bestimmter Integrale.

Der Beweis hierfür ist außerordentlich einfach. Zunächst ist

$$\int_a^b u\,dv = \left[\int u\,dv\right]_a^b.$$

Auf Grund der Formel für die partielle Integration eines unbestimmten Integrals ist nun

$$\int u\,dv = uv - \int v\,du,$$

so daß

$$\int_a^b u\,dv = \left[uv - \int v\,du\right]_a^b.$$

Hieraus folgt jedoch unmittelbar (24).

Beispiel 3. Zu berechnen ist

$$I = \int_0^\pi x\cos x\,dx.$$

Offenbar ist

$$I = \int_0^\pi x\,d\sin x = [x\sin x]_0^\pi - \int_0^\pi \sin x\,dx = [\cos x]_0^\pi = -2.$$

26. Die Wallissche Formel. Mit Hilfe der Ergebnisse des vorangehenden Abschnitts können wir einen recht interessanten Ausdruck für π herleiten. Dazu betrachten wir die Integrale

$$U_n = \int_0^{\frac{\pi}{2}} \cos^n x\,dx,$$

wobei n eine beliebige nicht negative ganze Zahl ist. Wir haben bereits gesehen, daß

$$U_0 = \frac{\pi}{2}, \quad U_1 = 1.$$

Es sei nun $n > 1$. Dann ist auf Grund von Satz 2 des vorangehenden Abschnitts

$$U_n = \int_0^{\frac{\pi}{2}} \cos^{n-1} x\,d\sin x = \left[\sin x\cos^{n-1} x\right]_0^{\frac{\pi}{2}} - \int_0^{\frac{\pi}{2}} \sin x\,d(\cos^{n-1} x).$$

Da das erste Glied der Differenz auf der rechten Seite gleich Null ist, gilt

$$U_n = (n-1)\int_0^{\frac{\pi}{2}} \cos^{n-2} x\sin^2 x\,dx = (n-1)\int_0^{\frac{\pi}{2}} \cos^{n-2} x(1-\cos^2 x)\,dx,$$

d. h.

$$U_n = (n-1)(U_{n-2} - U_n),$$

also

$$U_n = \frac{n-1}{n} U_{n-2}. \tag{25}$$

Wir nehmen nun an, daß n ungerade ist. Ist $n-2>1$, so können wir in (25) n durch $n-2$ ersetzen und erhalten

$$U_{n-2}=\frac{n-3}{n-2}U_{n-4}.$$

Diese Reduktion können wir so lange fortsetzen, bis wir zu

$$U_3=\frac{2}{3}U_1$$

gelangen. Drücken wir nun U_n durch U_{n-2}, sodann U_{n-2} durch U_{n-4} usw. aus, so finden wir

$$U_n=\frac{2\cdot 4\cdots(n-1)}{3\cdot 5\cdots n}U_1.$$

Beachten wir noch, daß $U_1=1$, und bezeichnen wir mit $m!!$ das Produkt aller natürlichen Zahlen, die nicht größer als m und ebenso wie m gerade bzw. ungerade sind, so können wir dafür auch

$$U_n=\frac{(n-1)!!}{n!!} \tag{26}$$

schreiben.

Entsprechend ergibt sich für gerades n

$$U_n=\frac{(n-1)!!}{n!!}\,\frac{\pi}{2}. \tag{27}$$

Nun ist für $0\leqq x\leqq\frac{\pi}{2}$ offenbar

$$\cos^{2n+2}x\leqq\cos^{2n+1}x\leqq\cos^{2n}x.$$

Daher gilt auf Grund von Satz 9 aus Abschnit 23

$$U_{2n+2}<U_{2n+1}<U_{2n}.$$

Diese Ungleichung können wir auf Grund von (26) und (27) auch als

$$\frac{(2n+1)!!}{(2n+2)!!}\,\frac{\pi}{2}<\frac{(2n)!!}{(2n+1)!!}<\frac{(2n-1)!!}{(2n)!!}\,\frac{\pi}{2}.$$

schreiben, so daß

$$\frac{2n+1}{2n+2}\cdot\frac{\pi}{2}<\left[\frac{(2n)!!}{(2n-1)!!}\right]^2\cdot\frac{1}{2n+1}<\frac{\pi}{2}. \tag{28}$$

Da nun

$$\lim_{n\to\infty}\frac{2n+1}{2n+2}=1,$$

erhalten wir aus (28)

$$\frac{\pi}{2}=\lim_{n\to\infty}\left[\frac{(2n)!!}{(2n-1)!!}\right]^2\cdot\frac{1}{2n+1},$$

und hieraus folgt

$$\pi=\lim_{n\to\infty}\frac{1}{n}\left[\frac{(2n)!!}{(2n-1)!!}\right]^2.$$

Diese Formel bezeichnet man üblicherweise als Wallissche *Formel*.[1]) Durch sie wird die Zahl π in sehr einfacher Weise durch natürliche Zahlen ausgedrückt.[2]) Diese Formel ist

[1]) J. Wallis (1616 — 1703), englischer Mathematiker.

[2]) Wir werden später noch andere Ausdrücke für π in natürlichen Zahlen angeben.

nun zwar vom theoretischen Standpunkt aus recht interessant, jedoch für eine praktische Berechnung von π nicht sehr geeignet. Um nämlich eine hinreichende Genauigkeit zu erreichen, muß man n ziemlich groß wählen; dann wird aber der Ausdruck $\frac{(2n)!!}{(2n-1)!!}$ sehr unhandlich.

27. Die näherungsweise Berechnung von bestimmten Integralen. Mit Hilfe der Formel von NEWTON und LEIBNIZ kann man ein bestimmtes Integral immer dann berechnen, wenn sich zu der betrachteten Funktion eine Stammfunktion angeben läßt, die durch endlich viele elementare Funktionen ausdrückbar ist. In den Fällen aber, in denen es keine Stammfunktion dieser Art gibt (oder ihr Aufsuchen mit außerordentlich umfangreichen Rechnungen verbunden ist), muß man andere Berechnungsverfahren für das bestimmte Integral verwenden.

In einigen Fällen gelingt es nun, den Wert eines bestimmten Integrals mit Hilfe eines Kunstgriffes zu bestimmen.

Wir betrachten z. B. das Integral[1])

$$I = \int_0^{\pi} \frac{x \sin x}{1 + \cos^2 x}\, dx.$$

Führen wir hier die Substitution $x = \pi - z$ aus, so finden wir

$$I = \int_0^{\pi} \frac{(\pi - z) \sin z}{1 + \cos^2 z}\, dz,$$

das heißt

$$I = \pi \int_0^{\pi} \frac{\sin z\, dz}{1 + \cos^2 z} - I.$$

Also ist

$$2I = \pi \int_0^{\pi} \frac{\sin z\, dz}{1 + \cos^2 z}.$$

Setzen wir hier $\cos z = u$, so erhalten wir

$$2I = \pi \int_{-1}^{1} \frac{du}{1 + u^2} = \pi \left[\operatorname{arctg} u\right]_{-1}^{+1} = \frac{\pi^2}{2}$$

und damit

$$\int_0^{\pi} \frac{x \sin x}{1 + \cos^2 x}\, dx = \frac{\pi^2}{4}.$$

Es ist klar, daß dieses Beispiel nicht typisch sein kann. Allgemein kann man jedoch stets das Integral einer Funktion, deren Stammfunktion nicht elementar ausdrückbar ist, mit Hilfe gewisser Näherungsformeln berechnen.

---

[1]) Das unbestimmte Integral $\int \frac{x \sin x}{1 + \cos^2 x}\, dx$ ist nicht elementar ausdrückbar.

Wir beschränken uns hier auf die Diskussion einer einzigen Näherungsformel, die als „Rechteckformel" bezeichnet wird. Alle anderen Näherungsformeln beruhen auf ganz ähnlichen Prinzipien, jedoch sind bei ihnen die Abschätzungen für die mit ihnen erreichbare Genauigkeit wesentlich komplizierter.

Wir haben bereits oben gesehen, daß für das bestimmte Integral einer stetigen Funktion allgemein die Beziehung

$$\int_a^b f(x)\,dx = f(\xi)\,(b-a)$$

gilt. Diese Formel stimmt genau, jedoch gibt sie uns kein Verfahren in die Hand, das bestimmte Integral zu berechnen, da sich allgemein über den Punkt ξ nur aussagen läßt, daß er zwischen a und b liegt. Ersetzen wir nun den Punkt ξ durch den Mittelpunkt des Intervalls $[a, b]$, so gelangen wir zu der folgenden, allerdings im allgemeinen nicht sehr genauen Näherungsformel

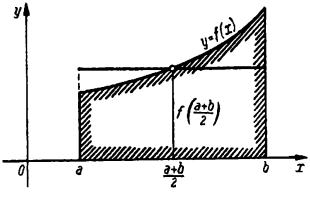


Abb. 27

$$\int_a^b f(x)\,dx \approx f\left(\frac{a+b}{2}\right)(b-a), \quad (29)$$

die man als „kleine Rechteckformel" bezeichnet. Man sieht leicht, daß im Fall einer positiven Funktion das Ersetzen des Integrals

$$\int_a^b f(x)\,dx$$

durch den Wert $f\left(\frac{a+b}{2}\right)(b-a)$ vom geometrischen Standpunkt aus darauf hinausläuft, daß man an Stelle des von den Kurven $y=0$, $y=f(x)$, $x=a$ und $x=b$ begrenzten krummlinigen Trapezes das Rechteck mit der Grundlinie $[a, b]$ und der Höhe $f\left(\frac{a+b}{2}\right)$ betrachtet (Abb. 27).

Wir wollen nun untersuchen, wie groß der maximale Fehler der Näherungsgleichung (29) ist. Dazu setzen wir voraus, daß für die Funktion $f(x)$ im Intervall $[a, b]$ überall sowohl die erste Ableitung $f'(x)$ als auch die zweite Ableitung $f''(x)$ existieren und stetig sind. Dann können wir $f(x)$ auf Grund der Taylorschen Formel folgendermaßen nach Potenzen von $x-\frac{a+b}{2}$ entwickeln:

$$f(x)=f\left(\frac{a+b}{2}\right)+f'\left(\frac{a+b}{2}\right)\left(x-\frac{a+b}{2}\right)+\frac{f''(\bar{x})}{2}\left(x-\frac{a+b}{2}\right)^2,$$

wobei $\bar{x}$ ein gewisser Punkt zwischen $\frac{a+b}{2}$ und x ist. Integrieren wir diese Gleichung, so erhalten wir wegen

$$\int_a^b \left(x-\frac{a+b}{2}\right) dx = 0,$$

daß

$$\int_a^b f(x)\,dx = f\left(\frac{a+b}{2}\right)(b-a) + \frac{1}{2}\int_a^b f''(\bar{x})\left(x-\frac{a+b}{2}\right)^2 dx. \tag{30}$$

Bezeichnen wir nun mit m bzw. M den kleinsten bzw. größten Wert der Ableitung $f''(x)$ im Intervall $[a, b]$, so ist

$$m \leqq f''(\bar{x}) \leqq M.$$

Multiplizieren wir diese Ungleichungskette gliedweise mit der (nicht negativen!) Zahl $\left(x-\frac{a+b}{2}\right)^2$ und integrieren wir die dabei entstehende Ungleichungskette gliedweise, so erhalten wir

$$m\int_a^b \left(x-\frac{a+b}{2}\right)^2 dx \leqq \int_a^b f''(\bar{x})\left(x-\frac{a+b}{2}\right)^2 dx \leqq M\int_a^b \left(x-\frac{a+b}{2}\right)^2 dx.$$

Das Integral

$$\int_a^b \left(x-\frac{a+b}{2}\right)^2 dx$$

läßt sich nun durch die Substitution $x = \frac{a+b}{2} + \frac{b-a}{2}t$ direkt auf das Integral

$$\frac{(b-a)^3}{8}\int_{-1}^{1} t^2\,dt = \frac{(b-a)^3}{12}$$

zurückzuführen, so daß

$$m\frac{(b-a)^3}{12} \leqq \int_a^b f''(\bar{x})\left(x-\frac{a+b}{2}\right)^2 dx \leqq M\frac{(b-a)^3}{12}.$$

Hieraus folgt, daß die Zahl

$$h = \frac{12}{(b-a)^3}\int_a^b f''(\bar{x})\left(x-\frac{a+b}{2}\right)^2 dx$$

zwischen m und M liegt. Daher nimmt die als stetig vorausgesetzte Funktion $f''(x)$ diesen Wert in einem gewissen Punkte $x = \eta$ zwischen a und b wirklich an, so daß

$$\int_a^b f''(\bar{x})\left(x-\frac{a+b}{2}\right)^2 dx = \frac{(b-a)^3}{12} f''(\eta).$$

Setzen wir dies in (30) ein, so gelangen wir zu der folgenden „kleinen Rechteckformel mit Restglied“:

$$\int_a^b f(x)\,dx = f\left(\frac{a+b}{2}\right)(b-a) + \frac{(b-a)^3}{24} f''(\eta). \tag{31}$$

Wir denken uns nun das Intervall $[a, b]$ durch die Teilpunkte

$$x_0 = a, \quad x_1 = a + \frac{b-a}{n}, \quad x_2 = a + 2\frac{b-a}{n}, \ \ldots, \quad x_n = b$$

in n Teilintervalle $[x_k, x_{k+1}]$ gleicher Länge zerlegt. Dann ist

$$\int_a^b f(x)\,dx = \sum_{k=0}^{n-1} \int_{x_k}^{x_{k+1}} f(x)\,dx. \tag{32}$$

Wir wenden nun auf die auf der rechten Seite in (32) auftretenden Integrale jeweils Formel (31) an:

$$\int_{x_k}^{x_{k+1}} f(x)\,dx = f\left(\frac{x_k + x_{k+1}}{2}\right)(x_{k+1} - x_k) + \frac{(x_{k+1} - x_k)^3}{24} f''(\eta_k), \tag{33}$$

wobei $x_k \leqq \eta_k \leqq x_{k+1}$. Beachten wir, daß $x_{k+1} - x_k = \frac{b-a}{n}$ und bezeichnen wir den Mittelpunkt $\frac{x_k + x_{k+1}}{2}$ des Intervalls $[x_k, x_{k+1}]$ jeweils mit $x_{k+1/2}$, so erhalten wir aus (32) und (33)

$$\int_a^b f(x)\,dx = \frac{b-a}{n} \sum_{k=0}^{n-1} f(x_{k+1/2}) + \frac{(b-a)^3}{24 n^3} \sum_{k=0}^{n-1} f''(\eta_k). \tag{34}$$

Da nun die Werte $f''(\eta_k)$ alle zwischen m und M liegen, ist offenbar

$$m \leqq \frac{1}{n} \sum_{k=0}^{n-1} f''(\eta_k) \leqq M.$$

Daher gibt es zwischen a und b einen Punkt ξ, so daß

$$f''(\xi) = \frac{1}{n} \sum_{k=0}^{n-1} f''(\eta_k).$$

Damit nimmt (34) die Form

$$\int_a^b f(x)\,dx = \frac{b-a}{n} \sum_{k=0}^{n-1} f(x_{k+1/2}) + \frac{(b-a)^3}{24 n^2} f''(\xi) \tag{35}$$

an. Das ist die sogenannte „große Rechteckformel mit Restglied“. Vernachlässigen wir hier den zweiten Summanden auf der rechten Seite, so erhalten wir die „angenäherte Rechteckformel“

$$\int_a^b f(x)\,dx \approx \frac{b-a}{n} \left[f(x_{1/2}) + f(x_{1/2}) + \cdots + f(x_{n-1/2})\right]. \tag{36}$$

Aus dem Vorangehenden folgt, daß der Fehler dieser Näherungsgleichung nicht größer ist als

$$K\frac{(b-a)^3}{24n^2}, \tag{37}$$

wobei $K=\max|f''(x)|$.

Da der Fehler (37) mit wachsendem n gegen Null strebt, ist

$$\int_a^b f(x)\,dx=\lim_{n\to\infty}\frac{b-a}{n}\left[f(x_{1/2})+f(x_{3/2})+\cdots+f(x_{n-1/2})\right].$$

Diese Formel ist übrigens für beliebige integrierbare Funktionen richtig[1]), während die Fehlerabschätzung (37) natürlich nur für Funktionen mit stetiger zweiter Ableitung möglich ist.

Wir betrachten nun einige Anwendungsbeispiele der vorangehenden Formeln.

Beispiel 1. Wir berechnen das Integral

$$\int_0^1 x^2\,dx$$

nach Formel (36) für $n=10$.

Hier ist $x_k=\frac{k}{10}$ und daher

$$x_{k+1/2}=\frac{x_k+x_{k+1}}{2}=\frac{2k+1}{20},$$

also

$$\left.\begin{aligned} &x_{1/2}=0{,}05,\quad x_{3/2}=0{,}15,\quad x_{5/2}=0{,}25,\quad x_{7/2}=0{,}35,\quad x_{9/2}=0{,}45,\\ &x_{11/2}=0{,}55,\quad x_{13/2}=0{,}65,\quad x_{15/2}=0{,}75,\quad x_{17/2}=0{,}85,\quad x_{19/2}=0{,}95. \end{aligned}\right\} \tag{38}$$

Für $f(x)=x^2$ ist dann

$$\begin{aligned} &f(x_{1/2})=0{,}0025,\quad f(x_{3/2})=0{,}0225,\quad f(x_{5/2})=0{,}0625,\\ &f(x_{7/2})=0{,}1225,\quad f(x_{9/2})=0{,}2025,\quad f(x_{11/2})=0{,}3025,\\ &f(x_{13/2})=0{,}4225,\quad f(x_{15/2})=0{,}5625,\quad f(x_{17/2})=0{,}7225,\\ &f(x_{19/2})=0{,}9025. \end{aligned}$$

---

[1]) In der Tat: Offenbar ist $\frac{b-a}{n}\left[f(x_{1/2})+\cdots+f(x_{n-1/2})\right]$ eine Integralsumme, die zu der Zerlegung

$$x_k=a+\frac{k}{n}(b-a)$$

gehört, und zwar gerade die Integralsumme mit den Zwischenwerten

$$\xi_k=x_{k+1/2}=\frac{x_k+x_{k+1}}{2}.$$

Addieren wir alle diese Werte, so erhalten wir

$$\sum_{k=0}^{9} f(x_{k+1/2}) = 3{,}325,$$

und es ist nach Formel (36)

$$\int_0^1 x^2\,dx \approx 0{,}3325. \tag{39}$$

Der genaue Wert des Integrals ist 0,33333. Daher ist der Fehler der Näherungsgleichung (39) kleiner als 0,001, während der relative Fehler kleiner als 0,003, d. h. kleiner als 0,3% ist.

Beispiel 2. Wir berechnen das Integral

$$\int_0^1 \frac{dx}{1+x^2}$$

nach Formel (36) für $n = 10$.

Da wir hier ebenso wie im vorangehenden Beispiel das Intervall $[0, 1]$ betrachten und ebenso wie dort $n = 10$ ist, können wir die Werte $x_{k+1/2}$ aus (38) entnehmen. Es sind nun die Werte der Funktion

$$f(x) = \frac{1}{1+x^2}$$

für diese Argumentwerte zu berechnen. Da wir zweckmäßigerweise mit Dezimalbrüchen rechnen werden, müssen wir uns zunächst überlegen, mit wieviel Dezimalstellen wir rechnen wollen. Dazu schätzen wir zunächst nach Formel (37) ab, welchen Fehler wir in unserem Fall bei $n = 10$ höchstens begehen. Offenbar ist

$$f''(x) = \frac{2(3x^2-1)}{(1+x^2)^3},$$

und dieser Quotient ist sicher nicht größer als 2.[1]) Daraus folgt, daß der Fehler bei Anwendung von (36) im betrachteten Fall kleiner als $\frac{1}{1200}$ ist. Berechnen wir nun die Funktionswerte $f(x_{k+1/2})$ gerundet auf vier Stellen hinter dem Komma, so begehen wir bei jedem Funktionswert einen Fehler, welcher kleiner als 0,00005 ist. Dann ist der Fehler der Summe der zehn Funktionswerte kleiner als 0,0005. Wenn wir sodann die Summe mit $\frac{b-a}{n} = 0{,}1$ multiplizieren, so geht auch der Fehler auf den zehnten Teil zurück. Daher ist der Gesamtfehler, der von der Ungenauigkeit der Formel und dem Runden herrührt, kleiner als 0,001.

[1]) Es ist nämlich $\left|\frac{3x^2-1}{(1+x^2)^3}\right| \leqq \frac{3x^2+1}{1+3x^2+3x^4+x^6} \leqq 1.$

Nach diesen Bemerkungen wollen wir zu den Berechnungen im einzelnen übergehen. Wegen

$$f(x_{1/2}) = 0{,}9975, \qquad f(x_{11/2}) = 0{,}7678,$$
$$f(x_{3/2}) = 0{,}9780, \qquad f(x_{13/2}) = 0{,}7030,$$
$$f(x_{5/2}) = 0{,}9412, \qquad f(x_{15/2}) = 0{,}6400,$$
$$f(x_{7/2}) = 0{,}8909, \qquad f(x_{17/2}) = 0{,}5806,$$
$$f(x_{9/2}) = 0{,}8316, \qquad f(x_{19/2}) = 0{,}5256$$

ist

$$\sum_{k=0}^{9} f(x_{k+1/2}) = 7{,}8562$$

und daher

$$\int_0^1 \frac{dx}{1+x^2} = 0{,}7856 (\pm 0{,}001).$$

Der genaue Wert des Integrals ist nun bekanntlich gleich $\frac{\pi}{4}$. Damit gelangen wir zu der Näherungsgleichung

$$\pi = 3{,}1424 (\pm 0{,}004),$$

das heißt

$$3{,}138 < \pi < 3{,}147$$

(genau ist $\pi = 3{,}14159 \ldots$).

Das zuletzt betrachtete Beispiel ist insofern von prinzipieller Bedeutung, weil hier ein praktisch anwendbares Verfahren zur beliebig genauen Berechnung von π (und π ist eine der wichtigsten Konstanten der Mathematik) gegeben wird (in Abschnitt 42 werden wir ein anderes Verfahren kennenlernen, welches noch schneller zum Ziele führt).

Beispiel 3. Wir berechnen das Integral

$$\int_0^1 \frac{dx}{1+x}$$

nach Formel (36) für $n = 10$. In diesem Fall sind die Werte der Funktion

$$f(x) = \frac{1}{1+x}$$

für die Argumentwerte (38) zu berechnen. Man erhält:

$$f(x_{1/2}) = 0{,}9524, \quad f(x_{11/2}) = 0{,}6452,$$
$$f(x_{3/2}) = 0{,}8696, \quad f(x_{13/2}) = 0{,}6061,$$
$$f(x_{5/2}) = 0{,}8000, \quad f(x_{15/2}) = 0{,}5714,$$
$$f(x_{7/2}) = 0{,}7407, \quad f(x_{17/2}) = 0{,}5405,$$
$$f(x_{9/2}) = 0{,}6897, \quad f(x_{19/2}) = 0{,}5128,$$

$$\sum_{k=0}^{9} f(x_{k+1/2}) = 6{,}9284,$$

$$\frac{b-a}{n} \sum_{k=0}^{9} f(x_{k+1/2}) = 0{,}69284.$$

Der Fehler bei den einzelnen Funktionswerten ist dabei jeweils kleiner als 0,00005, so daß der Fehler der Summe aller zehn Funktionswerte kleiner als 0,0005 und der von $\frac{b-a}{n}\sum f(x_{k+1/2})$ kleiner als 0,00005 ist. Ferner läßt sich der Fehler, der durch das Vernachlässigen des Restes in Formel (35) begangen wird, nach Formel (37) abschätzen. Wegen $f''(x)=\frac{2}{(1+x)^3}$ können wir hier $K=2$ setzen, so daß der Fehler der Näherungsgleichung (36) kleiner ist als

$$\frac{1}{1200}<0{,}000\,84.$$

Berücksichtigen wir, daß in unserem Beispiel $f''(x)$ stets positiv ist, so können wir feststellen, daß der Fehler R in (36) den Ungleichungen

$$0<R<0{,}000\,84$$

genügt, so daß der Gesamtfehler zwischen

$$-0{,}000\,05 \quad \text{und} \quad +0{,}000\,89$$

liegt. Damit finden wir für den genauen Wert des betrachteten Integrals, das ist $\ln 2$, die Abschätzung

$$0{,}692\,79<\ln 2<0{,}693\,73,$$

das heißt $\ln 2=0{,}693\,(\pm 0{,}001)$.

Entsprechend kann man wegen

$$\int_0^{N-1}\frac{dx}{1+x}=\ln N$$

mit Hilfe von Formel (36) mit hinreichend großem n den Logarithmus für jede positive reelle Zahl N mit beliebiger Genauigkeit berechnen (und den Fehler abschätzen!). Damit verfügen wir bereits über ein Verfahren, eine Logarithmentafel effektiv aufzustellen. Wir werden später (in Abschnitt 41) noch ein zweites derartiges Verfahren kennenlernen.

Wir wollen zum Abschluß dieser Überlegungen noch kurz darauf eingehen, wie die Werte einer Stammfunktion einer stetigen Funktion bestimmt werden können, wenn diese Stammfunktion nicht elementar ausdrückbar ist. Es ist leicht einzusehen, daß wir jetzt auch über eine Möglichkeit dafür verfügen. Es sei dazu $f(x)$ eine beliebige im Intervall $[a, b]$ stetige Funktion. Wir betrachten dann die durch

$$\Phi(x)=\int_a^x f(t)\,dt \tag{40}$$

gegebene Stammfunktion zu $f(x)$ (jede andere ergibt sich aus ihr durch Addition einer passenden Konstanten). Um einen bestimmten Wert $\Phi(x)$ dieser Funktion zu berechnen, braucht man nur das in (40) auf der rechten Seite stehende Integral auszurechnen. Eine Methode dafür haben wir aber gerade

kennengelernt. Damit man sich jedoch auf das erhaltene Resultat mit einiger Sicherheit verlassen kann, muß man den Fehler des berechneten Näherungswertes abschätzen. Ist $f(x)$ eine beliebige stetige Funktion, so haben wir hierzu kein allgemeines Verfahren. Besitzt jedoch die Funktion $f(x)$ eine stetige zweite Ableitung, so gelingt die Fehlerabschätzung unmittelbar mit Hilfe von Formel (37). Wir können also speziell die vorangehenden Überlegungen sicher dann mit befriedigendem Erfolg anwenden, wenn $f(x)$ eine elementare Funktion (mit stetiger zweiter Ableitung) ist.

§ 6. Anwendungen der Integralrechnung

28. Die Berechnung von Flächeninhalten. Wir haben bereits oben gesehen, daß für jede positive stetige Funktion $f(x)$ der Flächeninhalt des durch die Kurven $y = 0$, $x = a$, $x = b$ und $y = f(x)$ begrenzten krummlinigen Trapezes durch

$$F = \int_a^b f(x)\,dx$$

gegeben wird. Wir wollen hier zwei Anwendungsbeispiele dieser Formel geben.

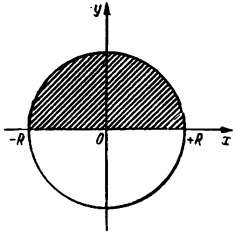


Abb. 28

Beispiel 1. Man berechne den Flächeninhalt des Kreises vom Radius R. Die Gleichung für den Kreis vom Radius R um den Koordinatenursprung (Abb. 28) lautet

$$x^2 + y^2 = R^2.$$

Daher wird der obere Halbkreis durch die Gleichung

$$y = +\sqrt{R^2 - x^2}$$

beschrieben, so daß der Flächeninhalt des in Abb. 28 schraffierten Halbkreises durch

$$\int_{-R}^{+R} \sqrt{R^2 - x^2}\,dx$$

gegeben wird. Durch die Substitution $x = R\sin t$ geht dieses Integral in das Integral $\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} R^2\cos^2 t\,dt$ über, und es ist

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} R^2\cos^2 t\,dt = R^2 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1+\cos 2t}{2}\,dt = \frac{R^2}{2}\left[t + \frac{1}{2}\sin 2t\right]_{-\frac{\pi}{2}}^{+\frac{\pi}{2}} = \frac{\pi R^2}{2}.$$

Daher ist der Flächeninhalt des ganzen Kreises gleich πR^2.

Beispiel 2. Man berechne den Flächeninhalt der Figur, die durch die x-Achse und eine Halbwelle der Sinuskurve $y = \sin x$ begrenzt wird (Abb. 29). Dieser Flächeninhalt wird durch

$$F = \int_0^{\pi} \sin x \, dx = \left[-\cos x\right]_0^{\pi} = 2$$

gegeben. Es ist interessant, daß dieser Flächeninhalt rational ist.

Wir wollen jetzt allgemeiner den Flächeninhalt der von den Kurven $x = a$, $x = b$, $y = f(x)$, $y = g(x)$ begrenzten Figur betrachten, wobei $f(x)$ und $g(x)$

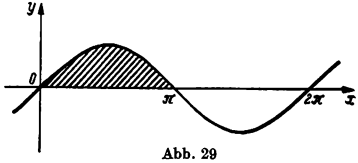


Abb. 29

zwei im Intervall $[a, b]$ stetige positive[1]) Funktionen sind, für welche überall $f(x) \leqq g(x)$ ist (Abb. 30).

Es liegt auf der Hand, daß der Flächeninhalt einer solchen Figur folgendermaßen definiert werden kann:

$$F = \int_a^b [g(x) - f(x)] \, dx .$$

Setzen wir $g(x) - f(x) = r(x)$, so können wir dafür auch kurz

$$F = \int_a^b r(x) \, dx \tag{1}$$

schreiben. Hieraus folgt, daß der Flächeninhalt von der speziellen Form der Figur weitgehend unabhängig ist, daß er nämlich nur von der jeweiligen Länge $r(x)$ des Ordinatenabschnitts zwischen den Kurven $y = f(x)$ und $x = g(x)$ abhängt. Sind also $f_1(x)$ und $g_1(x)$ irgendwelche andere Funktionen, für welche ebenfalls $g_1(x) - f_1(x) = r(x)$ gilt, so besitzt die von ihnen begrenzte Figur denselben Flächeninhalt wie die zuerst betrachtete. Dieses bereits im 17. Jahrhundert von dem Italiener B. Cavalieri (1591? — 1647), einem Vorläufer von Leibniz und Newton, aufgestellte Resultat gestattet die folgende rein geometrische Formulierung:

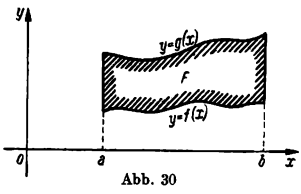


Abb. 30

[1]) Man sieht leicht ein, daß man auf diese Voraussetzung verzichten kann.

CAVALIERIsches Prinzip. *Es seien* (I) *und* (II) *zwei von den parallelen Geraden p und q eingeschlossene ebene Figuren* (Abb. 31), *für welche die Schnitte mit jeder zu p und q parallelen Geraden r jeweils gleiche Länge haben. Dann besitzen die Figuren* (I) *und* (II) *denselben Flächeninhalt.*

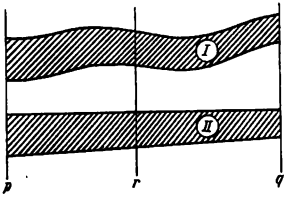


Abb. 31

Man zeigt leicht[1]), daß allgemeiner folgendes gilt: *Stehen die erwähnten Schnittlängen in einem gewissen konstanten Verhältnis, so stehen auch die Flächeninhalte von* (I) *und* (II) *in diesem Verhältnis.*

28. Die Berechnung von Volumina. Wir betrachten einen zwischen den parallelen Ebenen $x = a$ und $x = b$ liegenden Körper T, wobei wir annehmen wollen, daß sich als Schnitt von T mit jeder Ebene $x = x_0$ $(a \leq x_0 \leq b)$ eine Figur $T(x_0)$ ergibt, welche einen bestimmten Flächeninhalt $F(x_0)$ besitzt. Wir setzen ferner voraus, daß die dabei entstehende Funktion $F(x)$ eine stetige Funktion von x ist (Abb. 32). Gefragt ist nach dem Volumen des Körpers T. Neben der Frage nach der Berechnung dieses Volumens ist hier zunächst die Frage zu klären, was überhaupt unter dem Volumen des Körpers T verstanden werden soll.[2]) Zur Beantwortung dieser zweiten Frage denken wir uns das Intervall $[a, b]$ durch die Teilpunkte $x_0 = a < x_1 < x_2 < \cdots < x_n = b$ in n Teilintervalle zerlegt. Dann wird der Körper T durch die Ebenen $x = x_k$ in n dünne Schichten zerschnitten. In einfachen Fällen werden wir nun jede derartige Schicht angenähert als einen Zylinder mit dem Volumen

$$F(x_k)\,(x_{k+1} - x_k)$$

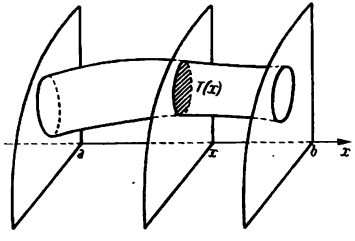


Abb. 32

[1]) Ist nämlich stets $r_1(x) = k r_2(x)$, so ist

$$F_1 = \int_a^b r_1(x)\,dx = \int_a^b k r_2(x)\,dx = k \int_a^b r_2(x)\,dx = k F_2.$$

[2]) Eine eingehende Analyse des Volumenbegriffs kann hier nicht gegeben werden. Die folgende Definition ist ausschließlich auf die Anwendung des (einfachen) Integrals auf Volumenberechnungen zugeschnitten.

ansehen können (wir verwenden hier ohne genauere Begründung, daß ein gerader Zylinder mit der Grundfläche F und der Höhe h das Volumen $V = Fh$ besitzt). In jedem solchen Fall werden wir

$$\sigma = \sum_{k=0}^{n-1} F(x_k)\,(x_{k+1} - x_k)$$

als angenäherten Wert des Volumens V des Körpers T ansehen können (wobei im Augenblick noch gar nicht genau feststeht, was eigentlich das Volumen eines Körpers ist). Dann liegt es aber nahe, den genauen Wert des Volumens V als Grenzwert von σ für $\lambda = \max(x_{k+1} - x_k) \to 0$ zu definieren. Diese Definition bedarf (wie auch jede andere Definition) selbstverständlich keines Beweises. Dagegen müssen wir uns davon überzeugen, ob dieser Grenzwert auch stets existiert. Das ist aber für die von uns betrachtete Klasse von Körpern der Fall, da es sich bei den Summen σ genau um die Integralsummen der als stetig vorausgesetzten Funktion $F(x)$ handelt. Auf Grund der vorangehenden Überlegungen gelangen wir ferner unmittelbar zu der Formel

$$V = \int_a^b F(x)\,dx, \tag{2}$$

mit deren Hilfe man das Volumen des Körpers T aus den Flächeninhalten seiner Schnitte $T(x)$ berechnen kann.[1])

Wir betrachten nun einige Beispiele für Anwendungen dieser Formel.

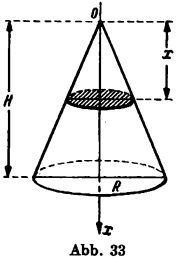


Abb. 33

Beispiel 1. Man berechne das Volumen eines geraden Kreiskegels vom Radius R und von der Höhe H.

Im vorliegenden Fall wählen wir als x-Achse die Symmetrieachse des Kegels und als Koordinatenursprung O dessen Spitze, wobei die positive Richtung der x-Achse von O aus in Richtung auf die Grundfläche des Kegels weise (Abb. 33). Die Schnitte $T(x)$ des Kegels sind dann Kreise, wobei der Radius des Kreises $T(x)$ offenbar gleich $\frac{R}{H}x$ ist. Hieraus folgt

$$V = \int_0^H \pi\left(\frac{R}{H}x\right)^2 dx = \frac{1}{3}\pi R^2 H.$$

[1]) Hieraus folgt, daß das Volumen V unabhängig von der speziellen Form des Schnittes $T(x)$ ist und nur vom jeweiligen Flächeninhalt $F(x)$ des Schnittes $T(x)$ abhängt. Damit erhalten wir:

Cavalierisches Prinzip für Volumina. *Es seien* (I) *und* (II) *zwei zwischen den parallelen Ebenen P und Q liegende Körper. Stimmt für jede zu P und Q parallele Ebene R der Flächeninhalt des Schnittes von* (I) *und R mit dem von* (II) *und R überein, so besitzen* (I) *und* (II) *dasselbe Volumen.*

Unterscheidet sich allgemeiner für jede derartige Ebene R der Flächeninhalt des Schnittes von (I) *und R von dem von* (II) *und R nur um einen von R unabhängigen Faktor k, so unterscheiden sich auch die Volumina von* (I) *und* (II) *nur um diesen Faktor.*

Beispiel 2. Man berechne das Volumen einer Kugel vom Radius R. Wir berechnen zunächst das Volumen einer Halbkugel. Mit den in Abb. 34 angedeuteten Bezeichnungen ist der Radius r des dort schraffiert gezeichneten

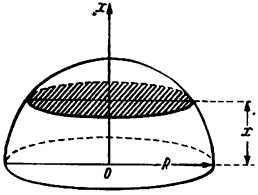


Abb. 34

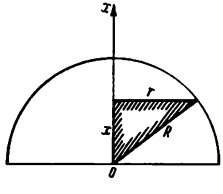


Abb. 35

Kreises gleich $\sqrt{R^2-x^2}$ (Abb. 35). Damit erhalten wir als Volumen der Halbkugel

$$\int_0^R \pi(R^2-x^2)\,dx = \frac{2}{3}\pi R^3,^{1})$$

so daß das Volumen der ganzen Kugel durch

$$V = \frac{4}{3}\pi R^3$$

gegeben wird.

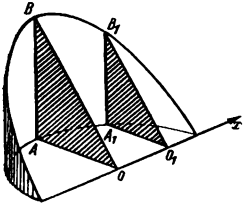


Abb. 36

Beispiel 3. Man berechne das Volumen eines Zylindersegments, das von einem geraden Kreiszylinder von einer durch einen Durchmesser der Grundfläche gehenden Ebene abgeschnitten wird.

Wir verwenden die Bezeichnungen der Abb. 36 und setzen $AB = H$ und $OA = R$; unsere Aufgabe soll es sein, das Volumen V des Zylindersegments durch H und R auszudrücken.

[1]) Diese Formel kann man übrigens auch mit Hilfe des CAVALLIERIschen Prinzips aus den Formeln für das Volumen eines Zylinders und eines Kegels herleiten. Zu diesem Zweck legen wir die Halbkugel vom Radius R zwischen die beiden Ebenen, die durch die obere und untere Grundfläche eines Zylinders vom Radius und von der Höhe R gebildet werden, aus dem ein Kegel gleichen Radius und gleicher Höhe herausgeschnitten ist (Abb. 37). Die in Abb. 37 eingezeichnete Ebene H schneidet dann aus der Kugel einen Kreis vom Radius $\sqrt{R^2-x^2}$, aus

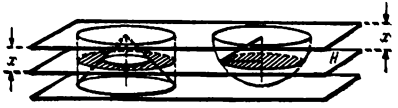


Abb. 37

Zu diesem Zweck betrachten wir das Dreieck $O_1A_1B_1$, welches aus dem Zylindersegment durch die Ebene ausgeschnitten wird, die zur Ebene OAB parallel ist und von dieser den Abstand $OO_1 = x$ besitzt. Der Flächeninhalt dieses Dreiecks wird durch $F(x) = \frac{1}{2} O_1A_1 \cdot A_1B_1$ gegeben. Dabei ist offenbar $O_1A_1 = \sqrt{R^2 - x^2}$; ferner besteht wegen der Ähnlichkeit der Dreiecke OAB und $O_1A_1B_1$ die Beziehung $A_1B_1 : AB = O_1A_1 : OA$, so daß

$$A_1B_1 = \frac{H}{R}\sqrt{R^2 - x^2}.$$

Also gilt

$$F(x) = \frac{H}{2R}(R^2 - x^2).$$

Hieraus folgt

$$V = \int_{-R}^{R} \frac{H}{2R}(R^2 - x^2)\,dx = \frac{H}{2R}\left[R^2 x - \frac{x^3}{3}\right]_{-R}^{+R},$$

d. h.

$$V = \frac{2}{3} R^2 H.$$

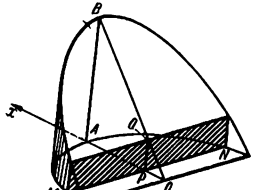


Abb. 38

Die zuletzt behandelte Aufgabe kann man aber auch noch auf andere Weise lösen. Dazu legen wir die x-Achse längs des Radius OA, der senkrecht auf dem Durchmesser steht, durch welchen die Ebene geht, die das betrachtete Zylindersegment abtrennt (Abb. 38). Dann schneidet die zu OA senkrechte Ebene durch den Punkt P, wobei $OP = x$, aus dem Zylindersegment ein Rechteck $T(x)$ aus, dessen Flächeninhalt durch

$$F(x) = MN \cdot PQ$$

gegeben wird. Dabei ist offenbar

$$MN = 2\sqrt{R^2 - x^2};$$

ferner ist wegen der Ähnlichkeit der Dreiecke OAB und OPQ

$$PQ : AB = OP : OA,$$

so daß

$$PQ = \frac{H}{R} x.$$

Daher ist

$$V = \int_0^R 2\frac{H}{R} x \sqrt{R^2 - x^2}\,dx = \frac{H}{R} \int_0^R \sqrt{R^2 - x^2}\,d(x^2).$$

---

dem Zylinder einen Kreis vom Radius R und aus dem Kegel einen Kreis vom Radius x aus. Daraus folgt, daß die Flächeninhalte der schraffierten Schnittflächen übereinstimmen, und zwar gleich $\pi(R^2 - x^2)$ sind. Daher ist das Volumen der Halbkugel gleich dem Volumen des Körpers, den man erhält, wenn man aus dem betrachteten Zylinder den betrachteten Kegel herausschneidet, d. h. gleich $\pi R^3 - \frac{1}{3}\pi R^3 = \frac{2}{3}\pi R^3$. Diese interessante Herleitung läßt sich übrigens auch gut in der Schule bringen.

Setzen wir hier $R^2 - x^2 = z$, so erhalten wir

$$V = -\frac{H}{R} \int_{R^2}^{0} \sqrt{z}\, dz = -\frac{H}{R} \left[\frac{2}{3} z^{\frac{3}{2}} \right]_{R^2}^{0}$$

und damit genau wie oben

$$V = \frac{2}{3} R^2 H.$$

Die Tatsache, daß wir im vorangehenden Beispiel auf zwei verschiedene Weisen zum selben numerischen Wert für das Volumen V gelangt sind, führt auf ein wesentliches neues Problem. Dazu beachten wir, daß bei der Berechnung des Volumens eines Körpers T nach Formel (2),

$$V = \int_a^b F(x)\, dx, \tag{2}$$

zunächst eine gewisse Gerade als x-Achse gewählt werden muß. Zeichnet man eine andere Gerade aus, so ändern sich dabei im allgemeinen die Schnitte $T(x)$ und damit auch ihr Flächeninhalt $F(x)$. Es erhebt sich nun naturgemäß die Frage, ob sich dabei nicht evtl. auch das Volumen ändert, das sich nach Formel (2) ergibt. Wäre dies der Fall, d. h., wäre das Volumen eines Körpers von einem subjektiven Element (wie z. B. der Auswahl der x-Achse) abhängig, so würde das den Wert unserer Volumendefinition stark beeinträchtigen. Daher müßten wir, genau genommen, unserer Definition noch einen Beweis für die Unabhängigkeit der Größe des Volumens von der Auswahl der x-Achse hinzufügen. Wir wollen jedoch darauf hier nicht eingehen, da der erwähnte Beweis recht kompliziert ist und eine bedeutende Einengung der Klasse der betrachteten Körper erfordern würde. Hinzu kommt, daß es sich bei dem genannten Problem weniger um ein Problem der Integralrechnung als um ein Problem der Geometrie handelt.

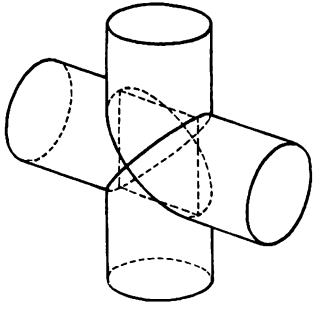


Abb. 39

Beispiel 4. Man berechne das Volumen V des Körpers T, der zwei geraden Kreiszylindern vom gleichen Radius R gemeinsam ist, deren Achsen sich unter einem rechten Winkel schneiden (Abb. 39).

Unabhängig von den Schwierigkeiten, mit denen eine genaue Beschreibung der Form dieses Körpers verbunden ist, läßt sich diese Aufgabe sehr einfach lösen. Dazu be-

zeichnen wir die Ebene, welche die Achsen beider Zylinder enthält, als „Achsenebene". Offenbar schneidet jede zur Achsenebene parallele Ebene aus jedem der beiden Zylinder ein Rechteck aus, dessen Breite, wie aus Abb. 40 zu entnehmen ist, durch

$$A_1B_1 = 2\sqrt{R^2 - x^2}$$

gegeben wird, wobei x den Abstand der betrachteten Ebene von der Achsenebene bezeichnet. Daher ist der Schnitt $T(x)$ dieser Ebene mit dem Körper T ein Quadrat mit der Seitenlänge A_1B_1, so daß das Volumen V durch

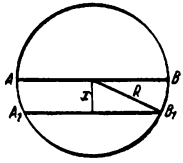


Abb. 40

$$V = \int_{-R}^{+R} 4(R^2 - x^2)\,dx = 4\left[R^2x - \frac{x^3}{3}\right]_{-R}^{+R},$$

d. h.

$$V = \frac{16}{3}R^3$$

gegeben wird.

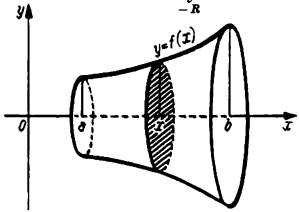


Abb. 41

Wir wollen zum Abschluß noch auf einen wichtigen Spezialfall von Formel (2) eingehen. Dazu betrachten wir eine im Intervall $a \leqq x \leqq b$ stetige Funktion $f(x)$. Wir stellen uns sodann vor, daß die Kurve $y = f(x)$ um die x-Achse rotiert, und betrachten den durch die Rotationsfläche und die Ebenen $x = a$ und $x = b$ begrenzten *Rotationskörper* T (Abb. 41). Im betrachteten Fall sind die Schnitte $T(x)$ Kreise, wobei der Kreis $T(x)$ jeweils den Radius $f(x)$ besitzt. Damit ergibt sich als Volumen des Rotationskörpers T:

$$V = \pi \int_a^b f^2(x)\,dx. \tag{3}$$

Betrachten wir zum Beispiel den durch eine Sinushalbwelle erzeugten Rotationskörper (Abb. 42), so erhalten wir nach Formel (3):

$$V = \pi \int_0^\pi \sin^2 x\,dx$$

$$= \pi \int_0^\pi \frac{1 - \cos 2x}{2}\,dx,$$

d. h.

$$V = \frac{\pi}{2}\left[x - \frac{1}{2}\sin 2x\right]_0^\pi = \frac{\pi^2}{2}.$$

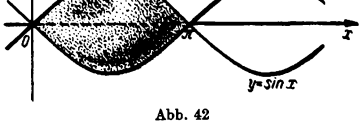


Abb. 42

30. Die Länge einer Kurve. Wir betrachten eine beliebige Kurve $y = f(x)$, wobei $f(x)$ eine im Intervall $[a, b]$ stetige Funktion ist. Wir zerlegen dann das Intervall $[a, b]$ durch die Teilpunkte $x_0 = a < x_1 < \cdots < x_n = b$ in Teilintervalle und bilden den Streckenzug, der die Punkte $M_k(x_k, f(x_k))$ $(k = 0, 1, \ldots, n)$ miteinander verbindet. Strebt nun bei gegen Null gehender Maximaldistanz aufeinanderfolgender Eckpunkte die Länge des Streckenzuges gegen einen endlichen Grenzwert s, so bezeichnen wir diesen als *Länge* der Kurve $y = f(x)$. Nicht jede stetige Kurve besitzt eine Länge (man kann zwar zeigen, daß dieser Grenzwert stets existiert, er kann jedoch unendlich sein). Diejenigen Kurven, die eine endliche Länge besitzen, werden auch *rektifizierbar* genannt.

Satz. *Besitzt die Funktion $f(x)$ im Intervall $[a, b]$ eine stetige Ableitung $f'(x)$, so ist die Kurve $y = f(x)$ rektifizierbar, und ihre Länge wird durch*

$$s = \int_a^b \sqrt{1 + f'^2(x)}\, dx \tag{4}$$

gegeben.

Beweis. Offenbar ist die Länge der Strecke $M_k M_{k+1}$ des oben betrachteten Streckenzuges gleich

$$\sqrt{(x_{k+1} - x_k)^2 + [f(x_{k+1}) - f(x_k)]^2}.$$

Nun ist aber auf Grund der Formel von LAGRANGE

$$f(x_{k+1}) - f(x_k) = f'(\xi_k)\,(x_{k+1} - x_k),$$

wobei $x_k \leqq \xi_k \leqq x_{k+1}$. Daher wird die gesamte Länge σ dieses Streckenzuges durch

$$\sigma = \sum_{k=0}^{n-1} \sqrt{1 + f'^2(\xi_k)}\,(x_{k+1} - x_k)$$

gegeben. Das ist offenbar gerade eine Integralsumme der stetigen Funktion

$$\sqrt{1 + f'^2(x)}.$$

Die Integralsummen streben aber, falls $\lambda = \max(x_{k+1} - x_k)$ gegen Null geht, gegen das Integral (4). Bezeichnen wir nun die Länge der größten Strecke $M_k M_{k+1}$ mit μ, so ist offenbar $\lambda \leqq \mu$. Daher strebt mit μ auch λ gegen Null. Damit ist der angegebene Satz bewiesen.

Beispiel. Wir betrachten einen beliebigen Kreis mit dem Radius R. In bezug auf ein Koordinatensystem, dessen Ursprung im Mittelpunkt des Kreises liegt, wird dieser Kreis durch die Gleichung

$$x^2 + y^2 = R^2$$

beschrieben. Die Gleichung für den oberen Halbkreis lautet dann

$$y = \sqrt{R^2 - x^2}.$$

In diesem Fall ist also

$$y' = \frac{-x}{\sqrt{R^2 - x^2}}.$$

Betrachten wir den ganzen oberen Halbkreis, d. h., lassen wir x auf dem Intervall $[-R, +R]$ variieren, so können wir den zuletzt bewiesenen Satz nicht anwenden, da die Ableitung y' für $x = \pm R$ nicht existiert. Wir betrachten daher zunächst nur denjenigen Bogen des oberen Halbkreises, der zwischen den Geraden $y = x$ und $y = -x$ liegt (d. h. den in Abb. 43 stark ausgezeichneten Bogen $\overset{\frown}{AB}$). Die Abszisse des Punktes A bzw. B ist nun gleich $-\frac{R\sqrt{2}}{2}$ bzw. $+\frac{R\sqrt{2}}{2}$, und da die Funktion $y = \sqrt{R^2 - x^2}$ im abgeschlossenen Intervall $\left[-\frac{R\sqrt{2}}{2}, +\frac{R\sqrt{2}}{2}\right]$ die Voraussetzungen des oben bewiesenen Satzes erfüllt, erhalten wir als Länge des Bogens $\overset{\frown}{AB}$

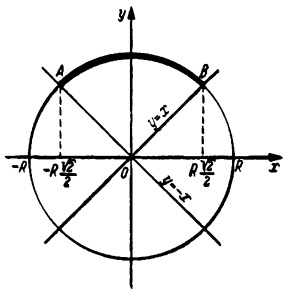


Abb. 43

$$s_{AB} = \int_{-\frac{R\sqrt{2}}{2}}^{+\frac{R\sqrt{2}}{2}} \sqrt{1 + y'^2}\, dx = \int_{-\frac{R\sqrt{2}}{2}}^{+\frac{R\sqrt{2}}{2}} \frac{R\, dx}{\sqrt{R^2 - x^2}},$$

d. h.

$$s_{AB} = R\left[\arcsin\frac{x}{R}\right]_{-\frac{R\sqrt{2}}{2}}^{+\frac{R\sqrt{2}}{2}} = \frac{\pi}{2} R.$$

Da aber der Bogen $\overset{\frown}{AB}$ genau ein Viertel der ganzen Kreisperipherie ist, ist auch diese rektifizierbar, und ihre Länge ist gleich $2\pi R$.

31. Der Flächeninhalt von Rotationsflächen. Wie oben, betrachten wir auch hier eine beliebige Kurve $y = f(x)$, wobei $f(x)$ eine im Intervall $[a, b]$ stetige Funktion ist. Der Einfachheit halber wollen wir jetzt jedoch annehmen, daß die betrachtete Funktion überall positiv ist. Sodann betrachten wir auch hier wieder einen Streckenzug mit den Ecken $M_k(x_k, f(x_k))$, wobei $x_0 = a < x_1 < x_2 < \cdots < x_n = b$. Wir lassen dann die Kurve und den Streckenzug um die x-Achse rotieren. Dabei beschreibt die Kurve eine gewisse Rotationsfläche (L). Der Streckenzug beschreibt ebenfalls eine Rotationsfläche (L'), und zwar ist diese aus n Kegelstümpfen (die unter Umständen in Zylinder entartet sein können) zusammengesetzt. Wir lassen nun die maximale Länge der Sehnen $M_k M_{k+1}$ gegen Null gehen. Strebt hierbei der Flächen-

inhalt[1]) L' der Fläche (L') gegen einen gewissen endlichen Grenzwert L, so bezeichnet man diesen als *Flächeninhalt* (Oberfläche) der Fläche (L).[2])

Satz. *Besitzt die Funktion $f(x)$ im Intervall $[a, b]$ eine stetige Ableitung $f'(x)$, so besitzt die Fläche (L) einen Flächeninhalt, und dieser wird durch*

$$L = 2\pi \int_a^b f(x)\sqrt{1+f'^2(x)}\,dx \tag{5}$$

gegeben.

Beweis. Offenbar ist der Flächeninhalt der Rotationsfläche, die von der Sehne M_kM_{k+1} erzeugt wird, gleich

$$\pi[f(x_k)+f(x_{k+1})]\sqrt{(x_{k+1}-x_k)^2+[f(x_{k+1})-f(x_k)]^2}.$$

Auf Grund der Formel von LAGRANGE ist nun

$$f(x_{k+1})-f(x_k)=f'(\xi_k)\,(x_{k+1}-x_k) \qquad (x_k<\xi_k<x_{k+1}).$$

Daher ist

$$L'=\pi\sum_{k=0}^{n-1}[f(x_k)+f(x_{k+1})]\sqrt{1+f'^2(\xi_k)}\,(x_{k+1}-x_k).$$

Wir betrachten daneben die Summe

$$\sigma=2\pi\sum_{k=0}^{n-1}f(\xi_k)\sqrt{1+f'^2(\xi_k)}\,(x_{k+1}-x_k),$$

die als Integralsumme der stetigen Funktion $2\pi f(x)\sqrt{1+f'^2(x)}$ für

$$\lambda=\max(x_{k+1}-x_k)\to 0$$

gegen das Integral (5) strebt. Da nun mit der Länge der größten Sehne M_kM_{k+1} auch λ gegen Null strebt, genügt es zu zeigen, daß

$$\lim_{\lambda\to 0}(L'-\sigma)=0. \tag{6}$$

Dazu betrachten wir eine beliebige Zahl $\varepsilon>0$. Zu dieser Zahl ε bestimmen wir eine Zahl $\delta>0$, so daß aus $|x'-x''|<\delta$ stets $|f(x')-f(x'')|<\varepsilon$ folgt. Es sei nun $\lambda<\delta$. Dann sind die Differenzen $|f(x_k)-f(\xi_k)|$ und $|f(x_{k+1})-f(\xi_k)|$ sämtlich kleiner als ε. Nun ist

$$L'-\sigma=\pi\sum_{k=0}^{n-1}\{[f(x_k)-f(\xi_k)]+[f(x_{k+1})-f(\xi_k)]\}\sqrt{1+f'^2(\xi_k)}\,(x_{k+1}-x_k).$$

---

[1]) Wir setzen hier als bekannt voraus, daß der Mantel eines Kegelstumpfes einen Flächeninhalt besitzt, der gleich dem Produkt aus der Länge einer Erzeugenden des Kegelstumpfes und dem Umfang des Schnittes des Kegelstumpfes mit der zur Grundfläche parallelen Ebene durch den Mittelpunkt der Höhe des Stumpfes ist.

[2]) Diese Definition ist nur für Rotationsflächen brauchbar. Im allgemeinen Fall ist die Definition des Flächeninhalts wesentlich komplizierter.

Bezeichnen wir also mit M das Maximum der stetigen Funktion $\sqrt{1+f'^2(x)}$, so ist für $\lambda < \delta$ offenbar

$$|L' - \sigma| \leqq 2M\pi\varepsilon \sum_{k=0}^{n-1} (x_{k+1} - x_k) = 2M\pi(b-a)\,\varepsilon.$$

Da nun die rechte Seite dieser Ungleichung mit ε beliebig klein gemacht werden kann, gilt (6), womit auch der behauptete Satz bewiesen ist.

Beispiel. Man berechne den Flächeninhalt einer Kugelzone.

Zu diesem Zweck lassen wir den Halbkreis

$$y = \sqrt{R^2 - x^2}$$

um die x-Achse rotieren. Wir berechnen den Flächeninhalt der Kugelzone, die von dem zwischen den Geraden $x = -h$ und $x = +h$ liegenden Bogen des Halbkreises erzeugt wird, wobei $0 < h < R$.[1]) Im betrachten Fall ist

$$y' = \frac{-x}{\sqrt{R^2 - x^2}}, \qquad \sqrt{1+y'^2} = \frac{R}{\sqrt{R^2 - x^2}},$$

so daß

$$L = 2\pi \int_{-h}^{+h} \sqrt{R^2 - x^2} \cdot \frac{R}{\sqrt{R^2 - x^2}}\, dx = 2\pi R \int_{-h}^{+h} dx = 4\pi R h.$$

Lassen wir hier h gegen R streben, so erhalten wir den Flächeninhalt der Kugeloberfläche, nämlich $L = 4\pi R^2$.[2])

32. Allgemeine Bemerkungen über Anwendungen der Integralrechnung und den Zusammenhang mit der Differentialrechnung. Ein großer Teil der Anwendungen der Theorie des bestimmten Integrals erfolgt nach demselben Schema. Wir wollen uns hier bemühen, die charakteristischen Züge dieses Schemas herauszuarbeiten.

Es möge jedem Teilintervall $[\alpha, \beta]$ eines gegebenen festen Intervalls $[a, b]$ ein bestimmter Wert einer physikalischen oder geometrischen Größe P entsprechen, den wir mit

$$P([\alpha, \beta])$$

bezeichnen wollen. In diesem Fall werden wir naturgemäß P eine Funktion des Intervalls $[\alpha, \beta]$, kurz eine *Intervallfunktion*, nennen.

---

[1]) Den vorangehenden Satz können wir nicht auf das ganze Intervall $[-R, +R]$ anwenden, weil die Funktion $y = \sqrt{R^2 - x^2}$ für $x = \pm R$ keine Ableitung besitzt.

[2]) Dieser letzte Schluß ist nicht vollkommen stichhaltig. Wir haben nämlich oben den Flächeninhalt L als Grenzwert der Flächeninhalte L' der Flächen (L') definiert, und es bedarf durchaus eines Beweises, daß der Grenzübergang von Kugelzonen zur ganzen Kugel zum selben Resultat führt. Man kann übrigens die Formel $L = 4\pi R^2$ auch dadurch beweisen, daß man den Flächeninhalt L' sofort für den Fall der ganzen Kugel studiert, was aber bedeutend schwieriger ist. Wir wollen hier jedoch auf diese Fragen nicht genauer eingehen.

Ist etwa $f(x)$ eine im Intervall $[a, b]$ definierte positive stetige Funktion, so können wir jedem Intervall $[\alpha, \beta]$ den Flächeninhalt $F([\alpha, \beta])$ des durch die Kurven $y = 0$, $y = f(x)$, $x = \alpha$, $x = \beta$ begrenzten krummlinigen Trapezes zuordnen. Ein anderes Beispiel für eine Intervallfunktion erhalten wir, wenn wir jedem Teilintervall $[\alpha, \beta]$ des Intervalls $[a, b]$ das Volumen $V([\alpha, \beta])$ des Rotationskörpers zuordnen, der durch Rotation des erwähnten krummlinigen Trapezes um die x-Achse entsteht. Ein drittes einfaches Beispiel ist schließlich folgendes: Das Intervall $[a, b]$ sei in stetiger Weise mit Masse belegt; dann können wir jedem Teilintervall $[\alpha, \beta]$ von $[a, b]$ die Menge $m([\alpha, \beta])$ der Masse zuordnen, welche sich auf diesem Intervall befindet.

Man nennt eine Intervallfunktion $P([\alpha, \beta])$ *additiv*, wenn für beliebiges γ mit $\alpha < \gamma < \beta$ gilt:

$$P([\alpha, \beta]) = P([\alpha, \gamma]) + P([\gamma, \beta]).$$

Man sieht unmittelbar, daß die im vorangehenden betrachteten Intervallfunktionen $F([\alpha, \beta])$, $V([\alpha, \beta])$ und $m([\alpha, \beta])$ in diesem Sinne additive Intervallfunktionen sind.

Es sei nun $P([\alpha, \beta])$ eine additive Intervallfunktion. Wir wollen annehmen, daß es eine im Grundintervall $[a, b]$ definierte und dort überall stetige Funktion $p(x)$ gibt, so daß

$$P([x, x + \Delta x]) = p(x)\,\Delta x + \varrho([x, x + \Delta x]), \tag{7}$$

wobei die Funktion $\varrho([x, x + \Delta x])$ folgende Eigenschaft besitze:

$$\lim_{\Delta x \to 0} \frac{\varrho([x, x + \Delta x])}{\Delta x} = 0.\ ^{1}) \tag{8}$$

Die Relation (7) besagt grob gesprochen, daß der Wert der Größe P in jedem hinreichend kleinen Intervall $[x, x + \Delta x]$ angenähert der Länge Δx proportional ist, wobei die Abweichung $\varrho([x, x + \Delta x])$ von höherer Ordnung gegen Null strebt als Δx. Wir wollen nun zeigen, daß unter dieser Voraussetzung *der Wert* $P([a, b])$ *der Größe* P *für das Gesamtintervall* $[a, b]$ *durch*

$$P([a, b]) = \int_a^b p(x)\,dx \tag{9}$$

gegeben wird.

Zu diesem Zweck betrachten wir eine beliebige Zerlegung

$$x_0 = a < x_1 < \cdots < x_n = b$$

des Intervalls $[a, b]$. Wegen der Additivität von P ist dann

$$P([a, b]) = \sum_{k=0}^{n-1} P([x_k, x_{k+1}]).$$

---

[1]) Das heißt also: Zu jedem $\varepsilon > 0$ gibt es ein $\delta > 0$ (welches nicht von x abhängt), so daß für $0 < \Delta x < \delta$ stets $\left|\frac{\varrho([x, x + \Delta x])}{\Delta x}\right| < \varepsilon$ ist.

Hieraus und aus (7) folgt

$$P([a,\ b]) = \sum_{k=0}^{n-1} p(x_k)\,(x_{k+1} - x_k) + \sum_{k=0}^{n-1} \varrho([x_k,\ x_{k+1}]). \tag{10}$$

Lassen wir nun in der Zerlegung die Maximaldistanz der Teilpunkte gegen Null gehen, so strebt die erste Summe gegen das Integral auf der rechten Seite in (9). Es genügt also zu zeigen, daß dabei die zweite Summe gegen Null geht. Zu diesem Zweck betrachten wir eine beliebige Zahl $\varepsilon > 0$. Zu ihr gibt es wegen (8) eine Zahl $\delta > 0$, so daß für $0 < \Delta x < \delta$ die Ungleichung

$$|\varrho([x,\ x + \Delta x])| < \varepsilon \Delta x$$

erfüllt ist. Wir betrachten nun zu dieser Zahl δ eine Zerlegung des Intervalls $[a,\ b]$ mit $\lambda = \max(x_{k+1} - x_k) < \delta$. Dann ist

$$\left|\sum_{k=0}^{n-1} \varrho([x_k,\ x_{k+1}])\right| < \varepsilon \sum_{k=0}^{n-1} (x_{k+1} - x_k) = \varepsilon(b - a).$$

Damit ist gezeigt, daß die zweite Summe auf der rechten Seite von (10) tatsächlich gegen Null strebt, was noch zu beweisen war.

Wenn es also gelingt, eine „angenäherte Proportionalität" zwischen der Größe $P([x,\ x + \Delta x])$ und der Länge Δx des Intervalls $[x,\ x + \Delta x]$ herzustellen, so ergibt sich damit die Möglichkeit, die Größe P nach Formel (9) zu berechnen. Hierin besteht aber gerade das oben genannte allgemeine Schema für die Anwendung der Integralrechnung. Das Problem liegt natürlich in jedem Fall darin, aus dem „elementaren" Summanden $P([x,\ x + \Delta x])$ angenähert (bis auf eine von höherer Ordnung gegen Null gehende Größe genau) eine Größe der Form $p(x)\Delta x$ abzuspalten.

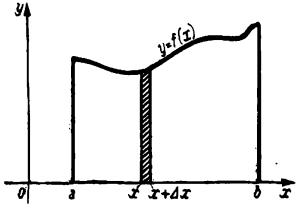


Abb. 44

So kann man z. B. den in Abb. 44 schraffierten Streifen angenähert als Rechteck mit der Grundlinie Δx und der Höhe $f(x)$ auffassen. Hieraus ergibt sich sofort

$$F([a,\ b]) = \int_a^b f(x)\,dx.$$

Wir wollen die Anwendung des allgemeinen Schemas noch an einem weiteren Beispiel erläutern: Wir stellen uns einen Kanal vor, der mit Wasser gefüllt und mit einem Schott verschlossen ist. Der Umriß des Schotts ist in Abb. 45 dargestellt. Die Wasseroberfläche befinde sich auf dem Niveau AB, die Tiefe des Kanals sei gleich h. Gesucht ist die Größe P des Druckes, der vom Wasser auf das Schott ausgeübt wird.

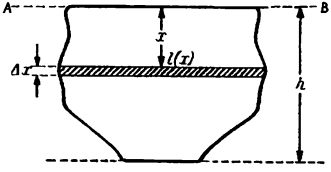


Abb. 45

Mit $l(x)$ bezeichnen wir die Breite des Schotts im Abstand x von AB. Der Druck $P([x,\ x+\Delta x])$, der auf den in Abb. 45 schraffiert gezeichneten Teil des Schotts ausgeübt wird, läßt sich nun leicht angenähert berechnen. Dazu ersetzen wir einerseits diesen Teil angenähert durch das Rechteck mit der Grundlinie $l(x)$ und der Höhe Δx und nehmen andererseits an, daß sich alle Punkte dieses Teils im Abstand x von der Wasseroberfläche AB befinden. Nach einem bekannten Gesetz der Hydrostatik, nach welchem der Druck des Wassers auf eine in ihm eingetauchte Fläche gleich dem Produkt aus dem Inhalt der Fläche und der Eintauchtiefe ist, erhalten wir, daß angenähert

$$P([x,\ x+\Delta x]) = x\,l(x)\,\Delta x$$

ist. Auf Grund von (9) ist dann aber

$$P = \int_0^h x\,l(x)\,dx.$$

Wir haben hier das gestellte Problem in ganz allgemeiner Form gelöst. Zur Berechnung des numerischen Wertes von P benötigt man natürlich noch die Funktion $l(x)$, die im einzelnen durch die Gestalt des Schotts gegeben wird.

In diesem Zusammenhang ist noch die folgende methodische Bemerkung von Bedeutung: Die vorangehenden Überlegungen geben nicht nur die Möglichkeit, eine Größe P zu berechnen, sondern können unter Umständen sogar zunächst den Weg zu ihrer formalen Definition weisen. Im Grunde haben wir diesen Weg bereits bei der Definition des Flächeninhalts in Abschnitt 28 und des Volumens in Abschnitt 29 beschritten. Damit gibt uns die Integralrechnung nicht nur die Möglichkeit, gewisse Größen effektiv zu berechnen, sondern wir sind gleichzeitig auch in der Lage, sie mit Hilfe einer allgemeinen Methode konstruktiv zu definieren.

Die Beziehung (7) gestattet übrigens auch noch eine andere Deutung, durch welche der Zusammenhang zwischen der Intervallfunktion $P([\alpha,\ \beta])$ und der Punktfunktion $p(x)$ von einer anderen Seite aus betrachtet wird. Aus (7) folgt nämlich unmittelbar, daß

$$p(x) = \lim_{\Delta x \to 0} \frac{P([x,\ x+\Delta x])}{\Delta x}. \tag{11}$$

Man sieht nun leicht ein, daß dieser Grenzübergang seinem Wesen nach eine Differentiation[1]) ist. Wenn also der Übergang von der Punktfunktion $p(x)$ zur Intervallfunktion $P([\alpha, \beta])$ mit Hilfe der Integration nach Formel (9) vollzogen werden kann, so läßt sich der umgekehrte Übergang mit Hilfe der Differentiation nach Formel (11) verwirklichen. Hier tritt nochmals der gegenseitig umkehrbare Zusammenhang zwischen Differentiation und Integration ganz anschaulich und dabei in sehr allgemeiner Form in Erscheinung.

Eine Charakterisierung einer Erscheinung durch eine Punktfunktion bedeutet nun stets eine lokale Charakterisierung dieser Erscheinung, d. h. eine Charakterisierung im Kleinen, während eine Charakterisierung durch eine Intervallfunktion eine globale Charakterisierung, eine Charakterisierung im Großen bedeutet. Daher kann man es auch als Aufgabe einer Integration ansehen, aus den lokalen Eigenschaften einer Erscheinung auf ihre globalen Eigenschaften zu schließen, während die Differentiation den Übergang von globalen zu lokalen Eigenschaften bedeutet. Diese wechselseitige Beziehung zwischen Integration und Differentiation bleibt auch selbst bei den verschiedensten Verallgemeinerungen erhalten.

---

[1]) Setzen wir nämlich $P([a, x]) = F(x)$, so ist, da wegen der Additivität der Funktion $P([\alpha, \beta])$

$$P([a, x + \Delta x]) = P([a, x]) + P([x, x + \Delta x])$$

ist, offenbar

$$P([x, x + \Delta x]) = F(x + \Delta x) - F(x),$$

so daß wegen (11) $p(x) = F'(x)$.

Kapitel III

REIHEN

§ 7. Reihen mit konstanten Gliedern

33. Grundbegriffe. Ein wesentliches Merkmal der Mathematik ist die enge Verbindung der abstrakt logischen und der numerischen Seite dieser Wissenschaft. Durch ihre gewaltigen numerischen Methoden ist die Mathematik ein unentbehrliches Hilfsmittel der Naturwissenschaften und der Technik geworden. Die Entwicklung dieser Methoden ist jedoch nur auf dem Boden einer sorgfältig ausgearbeiteten Theorie möglich. In der Theorie der Reihen tritt die erwähnte Verbindung besonders deutlich hervor. Diese Theorie ist nämlich einerseits außerordentlich gut durchgearbeitet und logisch aufgebaut, liefert aber andererseits auch die heute wesentlichsten numerischen Methoden. Es genügt, in diesem Zusammenhang darauf hinzuweisen, daß uns die Reihenlehre u. a. die Aufstellung von Logarithmentafeln und von Tafeln für die trigonometrischen Funktionen und die Berechnung von wichtigen Konstanten, wie z. B. der Zahl π usw., ermöglicht. Hierdurch wird klar, warum die Theorie der Reihen in der modernen Analysis eine so zentrale Stellung einnimmt, obwohl es sich doch bei den unendlichen Reihen, die den Gegenstand dieser Theorie bilden, um Ausdrücke sehr spezieller Gestalt handelt.

Definition. Unter einer *unendlichen Reihe* (oder kurz *Reihe*) versteht man einen Ausdruck der Form

$$a_1 + a_2 + a_3 + \cdots, \tag{1}$$

wobei $a_1, a_2, a_3, \ldots, a_n, \ldots$ (die *Glieder* der Reihe) irgendwelche festen Zahlen sind, deren Konstruktionsgesetz bekannt ist.

Häufig schreibt man auch den Ausdruck (1) kurz in der Form

$$\sum_{k=1}^{\infty} a_k. \tag{2}$$

Das „Wichtigste" bei der Bildung des Ausdrucks (1) sind die „Gedankenpunkte" am Ende des Ausdrucks. Sie sollen darauf hinweisen, daß die Anzahl der in die Definition der Reihe (1) eingehenden Glieder a_k unbegrenzt groß ist. Vom rein formalen Standpunkt aus sind also Reihen nichts anderes als Summen mit unendlich vielen Summanden.

Als erste Frage ergibt sich nun, ob man jedem derartigen Ausdruck einen numerischen Wert zuordnen kann. Es wird sich zeigen, daß es nicht immer gelingt, dies auf sinnvolle Weise zu tun. Das ist nur bei den sogenannten konvergenten Reihen[1]) stets möglich.

---

[1]) Beim weiteren Ausbau der Theorie zeigt es sich, daß man auch gewissen divergenten Reihen einen Zahlenwert zuordnen kann; wir können hier jedoch auf diese Dinge nicht eingehen.

Gegeben sei eine beliebige Reihe (1). Wir betrachten dann die Zahlenfolge $S_1, S_2, \ldots$, wobei

$$S_n = a_1 + a_2 + \cdots + a_n \qquad (n = 1, 2, 3, \ldots).$$

Die Zahlen dieser Folge nennt man die *Partialsummen* oder *Teilsummen* der gegebenen Reihe (1). Besitzt diese Folge einen endlichen Grenzwert

$$\lim_{n \to \infty} S_n, \tag{3}$$

so nennt man die Reihe (1) *konvergent*, wobei man den Grenzwert (3) als *Summe* der Reihe (1) bezeichnet. Die Summe einer Reihe ist also definitionsgemäß der (endliche) *Grenzwert der Folge ihrer Partialsummen*. Existiert dagegen der Grenzwert (3) nicht oder ist er unendlich, so heißt die Reihe (1) *divergent*. Besitzt die Reihe (1) eine Summe S, so schreibt man auch

$$S = a_1 + a_2 + a_3 + \cdots$$

oder

$$S = \sum_{k=1}^{\infty} a_k.$$

Divergiert dagegen eine Reihe, so wird ihr kein numerischer Wert zugeschrieben.

Wir wollen sogleich einige Beispiele betrachten, welche die bisher eingeführten Begriffsbildungen erläutern sollen.

Beispiel 1. Wir betrachten die Reihe

$$0 + 0 + 0 + 0 + \cdots. \tag{4}$$

Hier ist $S_n = 0$ und daher auch $\lim_{n \to \infty} S_n = 0$, d. h., die Reihe (4) konvergiert, und ihre Summe ist gleich Null.

Beispiel 2. Wir betrachten die Reihe

$$1 + 1 + 1 + 1 + \cdots. \tag{5}$$

Hier ist $S_n = n$ und daher $\lim_{n \to \infty} S_n = +\infty$, d. h., die Reihe (5) ist divergent.

Beispiel 3. Wir betrachten die Reihe

$$1 + (-1) + 1 + (-1) + \cdots. \tag{6}$$

Hier ist $S_n = 1$, falls n gerade ist, und $S_n = 0$, falls n ungerade ist. Hieraus folgt, daß die Reihe (6) divergiert.

Dieses letzte Beispiel ist insofern sehr lehrreich, als es zeigt, daß man dem formalen Apparat keine allzugroße Bedeutung zuschreiben darf. Man darf nämlich keineswegs glauben, daß jeder hinschreibbare analytische Ausdruck unbedingt einen Sinn haben muß und daß das Problem nur darin besteht, diesen Sinn zu finden. Es braucht nicht ausdrücklich betont zu werden, daß

diese Konzeption falsch ist[1]), und es gibt zahlreiche Beispiele, die das beweisen. Sehr gut eignet sich hierfür schon die Reihe (6). Ein Mathematiker des 18. Jahrhunderts „schloß" z. B. mit dieser Reihe folgendermaßen: Es sei

$$1+(-1)+1+(-1)+\cdots=S. \tag{7}$$

Fassen wir hier je zwei aufeinanderfolgende Glieder zusammen, so erhalten wir

$$S=[1+(-1)]+[1+(-1)]+[1+(-1)]+\cdots=0+0+0+\cdots=0.$$

Andererseits kann man aber S auch so schreiben:

$$S=1+[(-1)+1]+[(-1)+1]+\cdots=1+0+0\cdots=1.$$

Hieraus „folgt": $0=1$. Aus dieser „Entdeckung" folgerte er sodann, daß es nicht unmöglich ist, daß „Nichts" (d. h. 0) dasselbe ist wie „Etwas" (d. h. 1), und daß es daher angeblich auch nicht absurd ist, daß Gott die Welt aus dem Nichts schuf. Diese „Überlegungen" entbehren natürlich jeder wissenschaftlichen Bedeutung, charakterisieren aber sehr deutlich die Überschätzung des formalen Apparates, von der wir oben sprachen. Es ist klar, daß die Ursache für die unsinnige Entdeckung, daß $0=1$ sei, darin zu suchen ist, daß die Divergenz der betrachteten Reihe übersehen wurde. Dieses und ähnliche Mißverständnisse sind dadurch vollkommen gegenstandslos geworden, daß im ersten Viertel des vorigen Jahrhunderts die Theorie der Reihen durch die Einführung fester Begriffsbildungen auf eine sichere Grundlage gestellt wurde.

Beispiel 4. Wir betrachten die Reihe

$$\frac{1}{1\cdot 2}+\frac{1}{2\cdot 3}+\frac{1}{3\cdot 4}+\cdots \tag{8}$$

Wegen

$$\frac{1}{k(k+1)}=\frac{1}{k}-\frac{1}{k+1}$$

können wir die Partialsumme S_n offenbar folgendermaßen schreiben:

$$S_n=\left(1-\frac{1}{2}\right)+\left(\frac{1}{2}-\frac{1}{3}\right)+\left(\frac{1}{3}-\frac{1}{4}\right)+\cdots+\left(\frac{1}{n}-\frac{1}{n+1}\right).$$

Hieraus folgt

$$S_n=1-\frac{1}{n+1},$$

so daß die Folge der Partialsummen für $n\to\infty$ gegen 1 konvergiert. Daher konvergiert auch die Reihe (8), und ihre Summe ist gleich 1.

Beispiel 5. Von großer theoretischer Bedeutung ist die sogenannte *harmonische Reihe*

$$1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\cdots. \tag{9}$$

[1]) Mitunter gibt sie jedoch den Anstoß zu sehr wichtigen Entdeckungen (Beispiele sind die komplexen Zahlen und die Operatorenrechnung). Diese Konzeption, die also logisch nicht stichhaltig ist, kann daher unter Umständen doch ein wichtiges heuristisches Prinzip darstellen.

Wir wollen zeigen, daß sie *divergiert*. Dazu beachten wir, daß man ihre Partialsumme S_{2^m} folgendermaßen schreiben kann:

$$S_{2^m} = 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) + \cdots + \left(\frac{1}{2^{m-1}+1} + \frac{1}{2^{m-1}+2} + \cdots + \frac{1}{2^m}\right).$$

Hier ist wegen

$$\frac{1}{3} + \frac{1}{4} > \frac{1}{4} + \frac{1}{4} = \frac{1}{4} \cdot 2 = \frac{1}{2},$$

$$\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} > \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{1}{8} \cdot 4 = \frac{1}{2},$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\frac{1}{2^{m-1}+1} + \frac{1}{2^{m-1}+2} + \cdots + \frac{1}{2^m} > \frac{1}{2^m} \cdot 2^{m-1} = \frac{1}{2}$$

jeder der in Klammern zusammengefaßten Summanden größer als $\frac{1}{2}$, so daß

$$S_{2^m} > 1 + \frac{m}{2}.$$

Damit ist gezeigt, daß die Folge der Partialsummen S_n gegen keinen endlichen Grenzwert konvergieren kann, d. h., die Reihe (9) divergiert.

Beispiel 6. Als letztes Beispiel betrachten wir die *geometrische Reihe*

$$a + aq + aq^2 + aq^3 + \cdots \qquad (a \neq 0). \tag{10}$$

Im Fall $q = 1$ ist $S_n = an$, und die Reihe (10) divergiert. Ist dagegen $q \neq 1$, so ist nach einer bekannten Formel aus der Algebra

$$S_n = \frac{a - aq^n}{1 - q}.$$

Ist nun $|q| < 1$, so strebt bei unbeschränkt wachsendem n offenbar q^n gegen Null, so daß

$$S_n \to \frac{a}{1-q}.$$

Ist dagegen $|q| > 1$, so wächst q^n für $n \to \infty$ über alle Grenzen (bzw. divergiert), so daß auch die Folge der Partialsummen S_n gegen keinen endlichen Grenzwert konvergiert. Es bleibt also nur noch der Fall $q = -1$ zu betrachten. In diesem Fall nimmt die Reihe (10) die Form

$$a + (-a) + a + (-a) + \cdots$$

an, und diese Reihe ist ebenso wie die Reihe (6) divergent.

Somit gilt: *Die geometrische Reihe* (10) *konvergiert genau dann, wenn*

$$|q| < 1,$$

und zwar hat in diesem Fall ihre Summe S den Wert

$$S = \frac{a}{1-q}.$$

34. Die wesentlichsten Eigenschaften konvergenter Reihen.

Satz 1. *Werden die Glieder einer konvergenten Reihe bei unveränderter Reihenfolge in endlichen Gruppen zusammengefaßt und wird die Reihe aus den Summen dieser Gruppen gebildet, so ist diese Reihe ebenfalls konvergent, und ihre Summe stimmt mit der Summe der gegebenen Reihe überein, d. h.: Ist*

$$a_1 + a_2 + a_3 + \cdots = S, \tag{11}$$

so ist für $n_1 < n_2 < n_3 < \cdots$ *auch*

$$\begin{aligned}(a_1 + a_2 + \cdots + a_{n_1}) + (a_{n_1+1} + \cdots + a_{n_2}) \\ + (a_{n_2+1} + \cdots + a_{n_3}) + \cdots = S.\end{aligned} \tag{12}$$

Beweis. Wir bezeichnen mit S_n bzw. S_n^* die n-te Partialsumme der Reihe (11) bzw. (12). Dann ist, wie man leicht sieht,

$$S_k^* = S_{n_k}.$$

Hieraus folgt, daß die Folge $S_1^*, S_2^*, S_3^*, \ldots$ eine Teilfolge der Folge $S_1, S_2, S_3, \ldots$ ist. Da nun die zuletzt genannte Folge nach Voraussetzung gegen S konvergiert, gilt dies auch für ihre Teilfolge $\{S_k^*\}$.

Den soeben bewiesenen Satz kann man auch kurz folgendermaßen formulieren: *Die Glieder einer konvergenten Reihe können nach Belieben in Klammern zusammengefaßt werden.* Man sagt hierfür auch noch kürzer, daß *für konvergente Reihen das „assoziative Gesetz" gilt.*

Es sei jedoch ausdrücklich darauf hingewiesen, daß man aus der Konvergenz der Reihe (12) im allgemeinen nicht auf die Konvergenz der Reihe (11) schließen kann. Dies zeigt ganz deutlich folgendes Beispiel: Obwohl die Reihe

$$[1 + (-1)] + [1 + (-1)] + [1 + (-1)] + \cdots$$

konvergiert, ist die Reihe

$$1 + (-1) + 1 + (-1) + 1 + \cdots$$

divergent. Man darf also nicht immer die „Klammern fortlassen".

Satz 2. *Multipliziert man alle Glieder einer konvergenten Reihe mit einer gegebenen Zahl c, so erhält man eine Reihe, die ebenfalls konvergiert, und zwar ist die Summe der neuen Reihe gleich dem c-fachen der Summe der gegebenen Reihe. Es gilt also mit anderen Worten: Ist*

$$a_1 + a_2 + a_3 + \cdots = S, \tag{13}$$

so ist

$$ca_1 + ca_2 + ca_3 + \cdots = cS. \tag{14}$$

Beweis. Wir bezeichnen mit S_n bzw. S_n^* die n-te Partialsumme der Reihe (13) bzw. (14). Dann ist offenbar

$$S_n^* = cS_n,$$

so daß wegen $S_n \to S$ in der Tat $S_n^* \to cS$.

Satz 3. *Konvergente Reihen kann man gliedweise addieren, d. h., ist*

$$a_1+a_2+a_3+\cdots=A, \tag{15}$$

$$b_1+b_2+b_3+\cdots=B, \tag{16}$$

so ist

$$(a_1+b_1)+(a_2+b_2)+(a_3+b_3)+\cdots=A+B. \tag{17}$$

Beweis. Wir bezeichnen mit A_n, B_n bzw. S_n die n-te Partialsumme der Reihe (15), (16) bzw. (17). Wegen

$$S_n=A_n+B_n, \quad A_n\to A, \quad B_n\to B$$

gilt dann $S_n\to A+B$, was zu beweisen war.

Wir definieren als nächstes den wichtigen Begriff des Reihenrestes. Dazu betrachten wir eine beliebige Reihe

$$a_1+a_2+a_3+a_4+\cdots. \tag{18}$$

Unter dem *m-ten Rest* der Reihe (18) verstehen wir dann die Reihe

$$a_{m+1}+a_{m+2}+a_{m+3}+\cdots. \tag{19}$$

Satz 4. *Die Reihe* (18) *konvergiert dann und nur dann, wenn alle ihre Reste* (19) *konvergieren.*

Beweis. Wir bezeichnen mit S_n bzw. S_n^* die n-te Partialsumme der Reihe (18) bzw. (19). Offenbar ist dann für jedes $n>m$

$$S_n=S_m+S_{n-m}^*. \tag{20}$$

Wir nehmen nun an, daß die Reihe (19) konvergiert und die Summe R_m besitzt. Dann gilt

$$\lim_{n\to\infty} S_{n-m}=R_m,$$

so daß $S_n\to S_m+R_m$. Das bedeutet aber, daß auch die Reihe (18) konvergiert, wobei für ihre Summe S die Beziehung

$$S=S_m+R_m \tag{21}$$

besteht. Wir nehmen nun an, daß die Reihe (18) konvergiert und die Summe S besitzt. Dann ergibt sich ebenfalls aus (20), daß

$$\lim_{n\to\infty} S_{n-m}^*=S-S_m;$$

das bedeutet gerade, daß die Reihe (19) konvergiert.

Aus (21) folgt speziell, daß für eine konvergente Reihe die Summe R_m des m-ten Restes gleich der Differenz aus der Summe S der betrachteten Reihe und der m-ten Partialsumme S_m ist. Da nun bei unbeschränkt wachsendem m die Folge der Partialsummen gegen S strebt, erhalten wir hieraus sofort:

Satz 5. *Die Summen der m-ten Reste einer konvergenten Reihe streben bei unbeschränkt wachsendem m gegen Null*:

$$\lim_{m\to\infty}(a_{m+1}+a_{m+2}+a_{m+3}+\cdots)=0.$$

Schließlich wollen wir noch eine in den Anwendungen häufig brauchbare notwendige Bedingung für die Konvergenz einer Reihe beweisen:

Satz 6. *Die Glieder einer konvergenten Reihe streben mit wachsendem Index gegen Null, d. h., konvergiert die Reihe*

$$a_1+a_2+a_3+\cdots, \tag{22}$$

so gilt

$$\lim_{n\to\infty} a_n=0. \tag{23}$$

Beweis. Es sei S die Summe der Reihe (22). Dann streben bei unbeschränkt wachsendem n die Summen

$$S_n=a_1+a_2+\cdots+a_n, \quad S_{n-1}=a_1+a_2+\cdots+a_{n-1}$$

beide gegen S. Hieraus folgt

$$S_n-S_{n-1}\to S-S=0.$$

Offenbar ist aber $S_n-S_{n-1}=a_n$.

Wir weisen ausdrücklich darauf hin, daß diese Bedingung für die Konvergenz nur notwendig, im allgemeinen aber nicht hinreichend ist. So erfüllt zum Beispiel die harmonische Reihe (9) die Bedingung (23); sie ist aber, wie in Abschnitt 33 gezeigt wurde, divergent.

35. Reihen mit positiven Gliedern. Im vorliegenden Abschnitt wollen wir uns speziell mit Reihen beschäftigen, deren sämtliche Glieder positiv oder wenigstens nicht negativ sind. Erfüllen alle Glieder einer Reihe $a_1+a_2+a_3+\cdots$ die Ungleichung

$$a_n\geqq 0 \qquad (n=1,2,3,\ldots),$$

so wollen wir sie kurz eine *Reihe mit positiven Gliedern* nennen.[1]) Gilt für alle Glieder der Reihe sogar die Ungleichung $a_n>0$, so soll sie eine *Reihe mit echt positiven Gliedern* genannt werden.

Die Reihen mit positiven Gliedern haben viele Eigenschaften mit den endlichen Summen gemein.

Zunächst ist leicht einzusehen, daß *die Partialsummen*

$$S_n=a_1+a_2+\cdots+a_n$$

einer Reihe mit positiven Gliedern mit wachsendem n monoton wachsen (natürlich

[1]) Diese Bezeichnung ist nicht ganz konsequent (es müßte eigentlich „Reihe mit nicht negativen Gliedern" heißen). Jedoch hat sie sich in der deutschsprachigen Literatur weitgehend eingebürgert. Im russischen Original wird in diesem Zusammenhang von „positiven Reihen" gesprochen. — *Anm. d. wissenschaftl. Red.*

nicht notwendig echt monoton). Da nun jede monoton wachsende Zahlenfolge einen endlichen oder unendlichen Grenzwert besitzt (wobei die Glieder der Folge diesen Grenzwert nicht überschreiten), existiert für jede Reihe mit positiven Gliedern der Grenzwert

$$S = \lim_{n \to \infty} S_n.$$

Dieser Grenzwert ist endlich oder unendlich, je nachdem, ob die Folge der Partialsummen nach oben beschränkt ist oder nicht. Es gilt also:

Satz 1. *Eine Reihe mit positiven Gliedern ist dann und nur dann konvergent, wenn die Folge ihrer Partialsummen nach oben beschränkt ist.*

Es versteht sich von selbst, daß bei einer beliebigen Reihe die Beschränktheit der Folge der Partialsummen im allgemeinen nicht die Konvergenz der Reihe nach sich zieht. Dies zeigt bereits deutlich die Reihe $1+(-1)+1+(-1)+\cdots$.

Wir weisen nochmals ausdrücklich darauf hin, daß *bei einer konvergenten Reihe mit positiven Gliedern die Partialsummen stets höchstens so groß wie die Summe der Reihe sind.*

Durch den angegebenen Satz wird die Frage nach der Konvergenz einer Reihe mit positiven Gliedern auf die Frage nach der Beschränktheit der Folge ihrer Partialsummen zurückgeführt.

Als ein Beispiel für eine Anwendung von Satz 1 betrachten wir die Reihe

$$1+\frac{1}{2^\alpha}+\frac{1}{3^\alpha}+\frac{1}{4^\alpha}+\cdots, \tag{24}$$

wobei $\alpha > 1$ sei. Offenbar kann man die Partialsumme S_{2^m-1} dieser Reihe auch folgendermaßen schreiben:

$$S_{2^m-1} = 1+\left[\frac{1}{2^\alpha}+\frac{1}{3^\alpha}\right]+\left[\frac{1}{4^\alpha}+\frac{1}{5^\alpha}+\frac{1}{6^\alpha}+\frac{1}{7^\alpha}\right] + \cdots + \left[\frac{1}{(2^{m-1})^\alpha}+\cdots+\frac{1}{(2^m-1)^\alpha}\right].$$

Da die Summe

$$\frac{1}{(2^k)^\alpha}+\frac{1}{(2^k+1)^\alpha}+\cdots+\frac{1}{(2^{k+1}-1)^\alpha}$$

2^k Summanden enthält, von denen der erste der größte ist, ist diese Summe kleiner als

$$\frac{1}{(2^k)^\alpha}\cdot 2^k = \frac{1}{(2^{\alpha-1})^k}.$$

Daraus folgt, daß

$$S_{2^m-1} < 1+\frac{1}{2^{\alpha-1}}+\frac{1}{(2^{\alpha-1})^2}+\cdots+\frac{1}{(2^{\alpha-1})^{m-1}}.$$

Die hier auf der rechten Seite stehende Summe ist aber eine Partialsumme der geometrischen Reihe

$$1+\frac{1}{2^{\alpha-1}}+\frac{1}{(2^{\alpha-1})^2}+\frac{1}{(2^{\alpha-1})^3}+\cdots. \tag{25}$$

Auf Grund des in Abschnitt 33 Bewiesenen konvergiert diese Reihe (denn es ist nach Voraussetzung $\alpha > 1$), und ihre Summe ist gleich

$$\frac{2^{\alpha-1}}{2^{\alpha-1}-1}. \tag{26}$$

Da nun die Reihe (25) eine Reihe mit positiven Gliedern ist, ist keine ihrer Partialsummen größer als ihre Summe (26). Also gilt erst recht

$$S_{2^m-1} < \frac{2^{\alpha-1}}{2^{\alpha-1}-1}.$$

Diese Ungleichung gilt für beliebiges m. Nun kann man aber zu jeder Zahl n eine Zahl m finden, so daß

$$2^m - 1 > n.$$

Daher gilt auch für jedes n die Ungleichung

$$S_n < \frac{2^{\alpha-1}}{2^{\alpha-1}-1}.$$

Das besagt aber gerade, daß die Folge der Partialsummen der Reihe (24) nach oben beschränkt ist, so daß die Reihe (24) auf Grund von Satz 1 konvergiert.

Es muß in diesem Zusammenhang jedoch erwähnt werden, daß man einer unmittelbaren Anwendung von Satz 1 nur selten begegnet. Meistens verwendet man gewisse auf diesem Satz fußende, aber bequemer zu handhabende Konvergenzkriterien. Das einfachste von ihnen ist das sogenannte Vergleichskriterium.

Definition. Ist jedes Glied einer Reihe mit positiven Gliedern kleiner oder gleich dem entsprechenden Glied einer anderen Reihe mit positiven Gliedern, so bezeichnet man die zweite Reihe als eine *Majorante* der ersten und die erste Reihe als eine *Minorante* der zweiten Reihe, d. h., die Reihe

$$b_1+b_2+b_3+\cdots$$

ist z. B. eine Majorante der Reihe

$$a_1+a_2+a_3+\cdots,$$

wenn für jedes n die Ungleichung

$$a_n \leqq b_n$$

erfüllt ist.

Es ist ohne weiteres klar, daß jede Partialsumme einer gegebenen Reihe mit positiven Gliedern kleiner oder gleich der entsprechenden Partialsumme jeder ihrer Majoranten ist. Ist also die Folge der Partialsummen einer beliebigen Majorante der gegebenen Reihe nach oben beschränkt, so gilt dies erst recht für die Folge der Partialsummen der gegebenen Reihe. Wir erhalten damit:

Satz 2. Gibt es zu einer gegebenen Reihe mit positiven Gliedern eine konvergente Majorante, so ist auch die gegebene Reihe konvergent. Ist die gegebene Reihe divergent, so divergiert auch jede Majorante dieser Reihe.[1])

Wir betrachten als Beispiel die Reihe

$$1+\frac{1}{2^\alpha}+\frac{1}{3^\alpha}+\frac{1}{4^\alpha}+\cdots, \tag{27}$$

wobei $\alpha < 1$ sei. Es ist klar, daß diese Reihe eine Majorante der harmonischen Reihe ist, daß also die harmonische Reihe eine Minorante der Reihe (27) ist. Hieraus folgt auf Grund von Satz 2, daß die Reihe (27) divergiert.

Der Nachweis der Konvergenz oder Divergenz einer gegebenen Reihe mit Hilfe von Satz 2 führt immer auf das Problem, eine Reihe zu finden, deren Konvergenzverhalten bekannt ist und die sich im Sinne von Satz 2 mit der Ausgangsreihe vergleichen läßt. Das Auffinden einer solchen „Vergleichsreihe" hängt dabei in bedeutendem Maße von der Erfahrung und dem Scharfblick des Suchenden ab. Hierin liegt der Hauptmangel des Vergleichskriteriums. Daneben gibt es jedoch auch Konvergenzkriterien, die einen mehr algorithmischen Charakter besitzen. Wir beschränken uns hier auf den Beweis eines derartigen Kriteriums, und zwar des sogenannten D'ALEMBERTschen[2]) Quotientenkriteriums.

Lemma 1. *Ist*

$$a_1+a_2+a_3+\cdots \tag{28}$$

eine Reihe mit echt positiven Gliedern, für welche bei beliebigem n die Ungleichung

$$\frac{a_{n+1}}{a_n} > 1$$

erfüllt ist, so ist diese Reihe divergent.

In der Tat: Im betrachteten Fall ist offenbar stets $a_{n+1} > a_n > 0$, d. h., die Folge der Reihenglieder ist eine echt monoton wachsende Folge echt positiver Zahlen und strebt daher sicher nicht gegen Null; die in Abschnitt 34, Satz 6, bewiesene notwendige Bedingung für die Konvergenz der Reihe (28) ist also nicht erfüllt.

Lemma 2. *Ist für eine Reihe* (28) *mit echt positiven Gliedern bei beliebigem n die Ungleichung*

$$\frac{a_{n+1}}{a_n} \leqq q$$

erfüllt, wobei q eine feste Zahl kleiner als Eins ist,

$$0 < q < 1,$$

so ist die Reihe (28) *konvergent.*

---

[1]) Die zweite Behauptung dieses Satzes kann man offenbar auch folgendermaßen formulieren: *Gibt es zu einer gegebenen Reihe mit positiven Gliedern eine divergente Minorante, so ist auch die gegebene Reihe divergent. — Anm. d. wissenschaftl. Red.*

[2]) J. D'ALEMBERT (1717 — 1783), französischer Mathematiker.

In der Tat: Nach Voraussetzung ist

$$\frac{a_2}{a_1} \leqq q, \quad \frac{a_3}{a_2} \leqq q, \quad \ldots, \quad \frac{a_n}{a_{n-1}} \leqq q.$$

Multiplizieren wir alle diese Ungleichungen, so erhalten wir nach entsprechendem Kürzen

$$\frac{a_n}{a_1} \leqq q^{n-1}$$

oder

$$a_n \leqq a_1 q^{n-1}.$$

Aus dieser zuletzt erhaltenen Ungleichung folgt, daß die Glieder der Reihe (28) im betrachteten Fall nicht größer sind als die entsprechenden Glieder der Reihe

$$a_1 + a_1 q + a_1 q^2 + a_1 q^3 + \cdots. \tag{29}$$

Die Reihe (29) ist aber eine geometrische Reihe, deren Quotient kleiner als 1 ist, die also konvergiert. Hieraus folgt aber auf Grund des Vergleichskriteriums, daß die Reihe (28) erst recht konvergiert.

Wir weisen mit Nachdruck darauf hin, daß das Bestehen der Ungleichung

$$\frac{a_{n+1}}{a_n} < 1$$

im allgemeinen noch nicht die Konvergenz der Reihe (28) nach sich zieht. Man kann sich dies unmittelbar am Beispiel der harmonischen Reihe klarmachen.

Satz 3 (D'ALEMBERTsches Quotientenkriterium). *Es sei* (28) *eine Reihe mit echt positiven Gliedern, für welche der* (*endliche oder unendliche*) *Grenzwert*

$$l = \lim_{n \to \infty} \frac{a_{n+1}}{a_n} \tag{30}$$

existiert. Ist hierbei $l > 1$, *so divergiert die Reihe* (28), *ist* $l < 1$, *so konvergiert sie.*

Beweis. Es sei zunächst $l > 1$. In diesem Fall ist für hinreichend großes n

$$\frac{a_{n+1}}{a_n} > 1. \tag{31}$$

Es sei diese Ungleichung etwa für alle $n > m$ erfüllt. Dann gilt die Ungleichung (31) für je zwei aufeinanderfolgende Glieder der Reihe

$$a_{m+1} + a_{m+2} + a_{m+3} + \cdots, \tag{32}$$

so daß diese Reihe auf Grund von Lemma 1 divergiert. Da es sich bei der Reihe (32) um einen Rest der Reihe (28) handelt, divergiert dann auch die Reihe (28). Es sei nun $l < 1$. In diesem Fall betrachten wir eine beliebige Zahl q mit $l < q < 1$ $\left(\text{etwa } q = \frac{l+1}{2}\right)$. Offenbar gibt es dann eine Zahl m, so daß für alle $n > m$ die Ungleichung

$$\frac{a_{n+1}}{a_n} < q$$

gilt. Wir betrachten dann wiederum die Reihe (32) und können jetzt auf Grund von Lemma 2 behaupten, daß diese Reihe konvergiert. Dann konvergiert aber offenbar auch die Reihe (28).

Das D'ALEMBERTsche Kriterium hat offenbar vollständig algorithmischen Charakter. Um es anzuwenden, hat man nur die Quotienten $\frac{a_{n+1}}{a_n}$ zu bilden und ihr Verhalten bei unbeschränkt wachsendem n zu studieren. Irgendeine Hilfsreihe zum Vergleich mit der gegebenen Reihe ist nicht mehr erforderlich. Es kann natürlich der Fall eintreten, daß der Grenzwert (30) überhaupt nicht existiert oder daß er existiert und gleich Eins ist. In diesem Fall liefert Satz 3 keinen Aufschluß über das Konvergenzverhalten der Reihe (28).

Wir betrachten nun einige Beispiele.

Beispiel 1. Wir untersuchen die Reihe

$$\frac{1!}{1}+\frac{2!}{2^2}+\frac{3!}{3^3}+\frac{4!}{4^4}+\cdots.$$

Hier ist

$$a_n=\frac{n!}{n^n},\quad a_{n+1}=\frac{(n+1)!}{(n+1)^{n+1}},\quad \frac{a_{n+1}}{a_n}=\left(\frac{n}{n+1}\right)^n=\frac{1}{\left(1+\frac{1}{n}\right)^n}$$

und daher

$$\lim_{n\to\infty}\frac{a_{n+1}}{a_n}=\frac{1}{e}=\frac{1}{2{,}71828\ldots}<1.$$

Diese Reihe konvergiert also.

Beispiel 2. Für die Reihe

$$1+\frac{1}{2^\alpha}+\frac{1}{3^\alpha}+\frac{1}{4^\alpha}+\cdots \tag{33}$$

ist

$$a_n=\frac{1}{n^\alpha},\quad a_{n+1}=\frac{1}{(n+1)^\alpha},\quad \frac{a_{n+1}}{a_n}=\left(\frac{n}{n+1}\right)^\alpha\to 1.$$

Hier läßt sich also mit Hilfe des Quotientenkriteriums kein Aufschluß über das Konvergenzverhalten der Reihe gewinnen. Wir wissen jedoch bereits, daß diese Reihe für $\alpha>1$ konvergiert und für $\alpha\leqq 1$ divergiert. Es gibt also sowohl konvergente als auch divergente Reihen, bei denen $l=1$ ist.

Beispiel 3. Es ist zu untersuchen, für welche Werte $x\,(x>0)$ die Reihe

$$\frac{x}{1^2}+\frac{x^2}{2^2}+\frac{x^3}{3^2}+\frac{x^4}{4^2}+\cdots$$

konvergiert. Im betrachteten Fall ist

$$a_n=\frac{x^n}{n^2},\quad a_{n+1}=\frac{x^{n+1}}{(n+1)^2},\quad \frac{a_{n+1}}{a_n}=\left(\frac{n}{n+1}\right)^2 x\to x.$$

Daraus folgt, daß diese Reihe für $x<1$ konvergiert und für $x>1$ divergiert. Für $x=1$ gibt das D'ALEMBERTsche Kriterium keinen Aufschluß über das Konvergenzverhalten der Reihe. Man sieht jedoch leicht ein, daß in diesem Fall die Reihe gleichfalls konvergiert.

Neben dem Quotientenkriterium, welches am einfachsten ist, gibt es noch eine große Zahl weiterer Konvergenzkriterien für Reihen mit positiven Gliedern, z. B. das CAUCHYsche Kriterium, das RAABEsche Kriterium, das Integralkriterium u. a. Ein weiteres sehr allgemeines Kriterium stammt von dem bekannten russischen Mathematiker W. P. ERMAKOW (1845—1922).

Zum Abschluß dieser Überlegungen wollen wir noch eine einfache, aber sehr wesentliche Eigenschaft der Reihen mit positiven Gliedern beweisen, die wir weiter unten benötigen werden.

Satz 4. *Streicht man aus einer konvergenten Reihe mit positiven Gliedern beliebig viele Glieder und bildet man unter Erhaltung der Reihenfolge die Reihe aus den verbleibenden Gliedern, so ist diese ebenfalls konvergent.*

Beweis. Es sei

$$a_1 + a_2 + a_3 + a_4 + \cdots \tag{34}$$

eine beliebige konvergente Reihe mit positiven Gliedern, deren Summe gleich S sei. Es sei ferner

$$n_1 < n_2 < n_3 < \cdots$$

eine beliebige Folge von natürlichen Zahlen. Zu zeigen ist, daß dann auch die Reihe

$$a_{n_1} + a_{n_2} + a_{n_3} + \cdots \tag{35}$$

konvergiert. Zu diesem Zweck bezeichnen wir mit S_k bzw. S_k^* die k-te Partialsumme der Reihe (34) bzw. (35), d. h., es sei

$$S_k = a_1 + a_2 + \cdots + a_k, \quad S_k^* = a_{n_1} + a_{n_2} + \cdots + a_{n_k}.$$

Dann ist offenbar

$$S_k^* \leqq S_{n_k}.$$

Hieraus folgt, daß $S_k^* \leqq S$ für beliebiges k, d. h., die Folge der Partialsummen der Reihe (35) ist nach oben beschränkt. Das bedeutet aber gerade, daß die Reihe (35) konvergiert.

36. Alternierende Reihen. Wir wenden uns jetzt dem Studium von Reihen zu, bei denen nicht notwendig mehr alle Glieder nicht negativ sind. Dabei beschäftigen wir uns als erstes mit einem wichtigen Spezialfall derartiger Reihen, den sogenannten alternierenden Reihen, deren Theorie verhältnismäßig einfach ist.

Definition. Eine Reihe heißt *alternierend*, wenn je zwei aufeinanderfolgende Glieder der Reihe verschiedene Vorzeichen besitzen.

Für die folgenden Überlegungen empfiehlt es sich, die oben verwendete Bezeichnungsweise etwas abzuändern, nämlich unter a_n nicht mehr das n-te Glied der Reihe selbst, sondern dessen absoluten Betrag zu verstehen. Um etwas Bestimmtes vor Augen zu haben, wollen wir ferner annehmen, daß das erste Glied der betrachteten Reihe stets positiv ist (anderenfalls braucht man nur alle Glieder der Reihe mit -1 zu multiplizieren, was am Konvergenz-

verhalten der Reihe nichts ändert). Dann können wir die alternierende Reihe in der Form

$$a_1 - a_2 + a_3 - a_4 + a_5 - \cdots \text{ [1]} \tag{36}$$

schreiben.

Leibnizsche Regel. *Eine alternierende Reihe, bei welcher die Folge der absoluten Beträge der Reihenglieder monoton fällt und gegen Null strebt, ist konvergent.*

Beweis. Wir nehmen an, daß die Reihe (36) die genannten Voraussetzungen erfüllt:

$$a_1 > a_2 > a_3 > a_4 > \cdots, \tag{37}$$

$$\lim_{n\to\infty} a_n = 0. \tag{38}$$

Wir betrachten dann die Partialsummen S_{2n}:

$$S_2 = (a_1 - a_2),$$

$$S_4 = (a_1 - a_2) + (a_3 - a_4),$$

$$S_6 = (a_1 - a_2) + (a_3 - a_4) + (a_5 - a_6),$$

. . . . . .

Wegen (37) ist offenbar

$$S_2 < S_4 < S_6 < \cdots,$$

d. h., die Folge $\{S_{2n}\}$ ist monoton wachsend. Wegen

$$S_{2n} = a_1 - (a_2 - a_3) - (a_4 - a_5) - \cdots - (a_{2n-2} - a_{2n-1}) - a_{2n}$$

ist ferner

$$S_{2n} < a_1.$$

Hieraus folgt, daß die Folge $\{S_{2n}\}$ gegen einen endlichen Grenzwert S konvergiert:

$$S = \lim_{n\to\infty} S_{2n}.$$

Nun ist aber

$$S_{2n+1} = S_{2n} + a_{2n+1},$$

so daß wegen (38) auch die Folge $\{S_{2n+1}\}$ gegen S konvergiert. Damit ist gezeigt, daß die Partialsummen S_n, unabhängig davon, ob n gerade oder ungerade ist, bei hinreichend großem n der Zahl S beliebig nahe kommen, d. h., es gilt

$$\lim_{n\to\infty} S_n = S,$$

was zu beweisen war.

---

[1]) Strenggenommen stellt die Einführung dieser Schreibweise ein neues Übereinkommen dar, da wir nach Definition der Reihe statt (36) eigentlich

$$a_1 + (-a_2) + a_3 + (-a_4) + \cdots$$

schreiben müßten.

Wir weisen darauf hin, daß die Voraussetzung über das monotone Fallen der Folge der a_n nicht entbehrt werden kann. So divergiert z. B., wie man sofort sieht, die Reihe

$$1-\frac{1}{10}+\frac{1}{2}-\frac{1}{100}+\frac{1}{3}-\frac{1}{1000}+\cdots,$$

obwohl die Bedingung (38) erfüllt ist.[1])

Diesem Beweis ist übrigens sofort zu entnehmen, daß die Folge $\{S_{2n}\}$ monoton wächst und die Folge $\{S_{2n+1}\}$ wegen

$$\begin{aligned} S_1 &= a_1, \\ S_3 &= a_1-(a_2-a_3), \\ S_5 &= a_1-(a_2-a_3)-(a_4-a_5), \\ &\cdots\cdots\cdots\cdots\cdots \end{aligned}$$

monoton fällt, so daß S stets zwischen S_{2n} und S_{2n+1} liegt.

Diese zuletzt genannte Eigenschaft der alternierenden Reihen, welche die Voraussetzungen der Leibnizschen Regel erfüllen, kann man etwas anschaulicher auch folgendermaßen ausdrücken: Zunächst folgt aus $S_2 < S < S_1$, daß

$$0 < S < a_1,$$

d. h., die Summe der Reihe besitzt dasselbe Vorzeichen wie das erste Glied und ist dem Betrage nach kleiner als dieses Glied. Betrachten wir nun allgemein den n-ten Rest

$$(-1)^n[a_{n+1}-a_{n+2}+a_{n+3}-\cdots]$$

der Reihe (36), so handelt es sich bei diesem um eine Reihe vom gleichen Typ, und wir können auf Grund derselben Überlegung schließen, daß die Summe R_n dieses Restes das Vorzeichen $(-1)^n$ besitzt und $|R_n| < a_{n+1}$ ist. Beachten wir nun noch, daß

$$S = S_n + R_n,$$

so können wir folgendes feststellen: *Läßt man in einer alternierenden Reihe, welche die Voraussetzungen der* Leibnizschen *Regel erfüllt, die auf das n-te Glied folgenden Reihenglieder fort, betrachtet man also die Partialsumme*

$$S_n = a_1-a_2+a_3-\cdots\pm a_n$$

der Summe S, so begeht man einen Fehler[2]*), der dasselbe Vorzeichen wie das erste fortgelassene Glied $(-1)^n a_{n+1}$ besitzt und dem Betrage nach kleiner als a_{n+1} ist.*

---

[1]) Offenbar ist nämlich

$$S_{2n}=\left(1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}\right)-\left(\frac{1}{10}+\frac{1}{10^2}+\cdots+\frac{1}{10^n}\right).$$

Hierbei wächst der erste Anteil mit n über alle Grenzen, während der zweite Anteil stets kleiner als $\frac{1}{9}$ bleibt.

[2]) Als Fehler bezeichnen wir hier die Zahl, die man zu S_n hinzufügen muß, um S zu erhalten (d. h. R_n).

Diese Eigenschaft macht die Reihen, welche die Voraussetzungen der LEIBNIZschen Regel erfüllen, für Näherungsrechnungen besonders geeignet.

Beschränkt man sich z. B. bei der Berechnung von

$$S = 1 - \frac{1}{2^2} + \frac{1}{3^3} - \frac{1}{4^4} + \frac{1}{5^5} - \cdots$$

auf die hingeschriebenen Glieder, so begeht man einen Fehler mit negativem Vorzeichen, der dem Betrage nach kleiner als

$$\frac{1}{6^6} = \frac{1}{46656} < 0{,}000025$$

ist.

37. Absolut konvergente Reihen. Wir betrachten jetzt ganz beliebige Reihen, machen also über die Vorzeichen der Reihenglieder keinerlei einschränkende Voraussetzungen mehr. Hierbei bezeichnen wir mit $a_1, a_2, a_3, \ldots$ wieder die Reihenglieder selbst.

Satz 1. *Wir betrachten neben der Reihe*

$$a_1 + a_2 + a_3 + \cdots \tag{39}$$

die Reihe

$$|a_1| + |a_2| + |a_3| + \cdots \tag{40}$$

aus den absoluten Beträgen der Glieder der gegebenen Reihe. Konvergiert dann die Reihe (40), *so konvergiert auch die Ausgangsreihe* (39).

Beweis. Es sei

$$b_1 + b_2 + b_3 + \cdots \tag{41}$$

die Reihe aus den positiven (bzw. nicht negativen) Gliedern der Reihe (39) (in der Reihenfolge der Reihe (39)) und

$$c_1 + c_2 + c_3 + \cdots \tag{42}$$

die Reihe aus den Beträgen der negativen Glieder der Reihe (39), ebenfalls in der Reihenfolge der Reihe (39).[1]) Da sich jede der Reihen (41) und (42) aus der konvergenten Reihe (40), deren sämtliche Glieder nicht negativ sind, dadurch ergibt, daß man einen Teil ihrer Glieder fortläßt (um z. B. (41) aus (40) zu erhalten, muß man gerade die Glieder $c_1, c_2, c_3, \ldots$ fortlassen), sind diese Reihen auf Grund von Satz 4 aus Abschnitt 35 konvergent. Mit B bzw. C bezeichnen wir die Summe der Reihe (41) bzw. (42). Ferner wollen wir mit A_n, B_n bzw. C_n die n-te Partialsumme der Reihe (39), (41) bzw. (42) bezeichnen. Schließlich wollen wir annehmen, daß unter den Gliedern

$$a_1, a_2, a_3, \ldots, a_n$$

$m(n)$ nicht negativ und $p(n)$ negativ sind. Dann ist offenbar

$$A_n = B_{m(n)} - C_{p(n)}.$$

[1]) Ohne Beschränkung der Allgemeinheit können wir annehmen, daß es in der Reihe (39) unendlich viele positive und unendlich viele negative Glieder gibt, da andernfalls Satz 1 trivialerweise gilt.

Wächst nun n über alle Grenzen, so strebt die rechte Seite und damit auch die linke Seite dieser Beziehung gegen die Differenz $B - C$. Das bedeutet aber gerade, daß die Reihe (39) konvergiert, und zwar ist dabei, wie wir außerdem sehen, die Summe von (39) gleich $B - C$.

Wir weisen darauf hin, daß aus der Konvergenz der Reihe (39) im allgemeinen nicht die der Reihe (40) folgt. Zum Beispiel ist die Reihe

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$$

konvergent (wie man unmittelbar mit Hilfe der LEIBNIZschen Regel erkennt), während die Reihe der absoluten Beträge ihrer Glieder die divergente harmonische Reihe ist.

Die Forderung der Konvergenz der Reihe (40) ist also eine schärfere Forderung als die der Konvergenz der Reihe (39). Eine Reihe, die diese schärfere Forderung erfüllt, bei der also nicht nur die Reihe selbst, sondern sogar die Reihe der absoluten Beträge ihrer Glieder konvergiert, bezeichnet man als *absolut konvergent*. Konvergiert zwar die Reihe (39), aber nicht die Reihe (40), so nennt man die Reihe (39) *nicht-absolut konvergent*.

Aus dem Beweis von Satz 1 ergibt sich sofort:

Satz 2. *Ist die Reihe* (39) *absolut konvergent, so konvergieren auch die Reihe* (41) *ihrer positiven Glieder und die Reihe* (42) *der absoluten Beträge ihrer negativen Glieder. Die Summen A, B bzw. C der Reihen* (39), (41) *bzw.* (42) *sind dabei durch die Beziehung*

$$A = B - C \tag{43}$$

miteinander verknüpft.

Für Reihen mit beliebigen Gliedern läßt sich das D'ALEMBERTsche Konvergenzkriterium folgendermaßen formulieren:

Satz 3. *Es sei* (39) *eine beliebige Reihe, für welche der Grenzwert*

$$\lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right| = l$$

existiert. Ist $l < 1$, so konvergiert die Reihe (39), *ist $l > 1$, so divergiert sie.*

Beweis. Ist $l < 1$, so konvergiert auf Grund des D'ALEMBERTschen Kriteriums für Reihen mit positiven Gliedern die Reihe (40) und damit erst recht die Reihe (39). Ist dagegen $l > 1$, so gibt es eine natürliche Zahl m, so daß für $n \geqq m$

$$\left| \frac{a_{n+1}}{a_n} \right| > 1$$

ist. Dann ist aber

$$|a_m| < |a_{m+1}| < |a_{m+2}| < \cdots,$$

und die Folge der Glieder der Reihe (39) strebt nicht gegen Null. Hieraus folgt jedoch sofort, daß die Reihe (39) divergiert.

Auf Beispiele für diesen Satz kommen wir später zurück.

88. Die Umordnung und die Multiplikation von Reihen. Wir haben bereits oben bemerkt, daß die Reihen vom rein formalen Standpunkt aus nichts anderes als Summen mit unendlich vielen Summanden sind. Wir haben aber auch bereits darauf hingewiesen, daß man diesem formalen Standpunkt keine allzu große Bedeutung beimessen darf. Ein gutes Beispiel hierfür ist die Frage, ob man die Glieder einer Reihe nach Belieben umordnen darf. Ein wichtiges Gesetz der Arithmetik ist das kommutative Gesetz, nach welchem eine (endliche) Summe unabhängig von der Reihenfolge der Summanden ist. Es zeigt sich nun, daß dieses Gesetz für unendliche Reihen im allgemeinen nicht mehr gilt. Wir wollen dies zunächst an einem Beispiel erläutern.

Wir betrachten die Reihe

$$1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}-\frac{1}{6}+\cdots. \tag{44}$$

Nach der Leibnizschen Regel konvergiert diese, besitzt also eine gewisse Summe S. Dabei ist $S \neq 0$, da das Vorzeichen der Summe einer Reihe, welche die Voraussetzungen der Leibnizschen Regel erfüllt, gleich dem Vorzeichen ihres ersten Gliedes ist. Daneben betrachten wir die Reihe

$$1-\frac{1}{2}-\frac{1}{4}+\frac{1}{3}-\frac{1}{6}-\frac{1}{8}+\frac{1}{5}-\frac{1}{10}-\frac{1}{12}+\cdots, \tag{45}$$

die durch Umordnung der Glieder der Reihe (44) entsteht, wobei jetzt also auf jedes positive Glied zwei negative Glieder folgen. Mit S_n bzw. S_n^* bezeichnen wir die n-te Partialsumme der Reihe (44) bzw. (45). Indem wir die Summanden von S_{3n}^* zu je dreien zusammenfassen[1]), erhalten wir

$$S_{3n}^* = \left(1-\frac{1}{2}-\frac{1}{4}\right)+\left(\frac{1}{3}-\frac{1}{6}-\frac{1}{8}\right)+\cdots+\left(\frac{1}{2n-1}-\frac{1}{4n-2}-\frac{1}{4n}\right).$$

Offenbar ist nun

$$1-\frac{1}{2}=\frac{1}{2}, \quad \frac{1}{3}-\frac{1}{6}=\frac{1}{6}, \quad \ldots, \quad \frac{1}{2n-1}-\frac{1}{4n-2}=\frac{1}{4n-2}$$

und daher

$$S_{2n}^* = \left(\frac{1}{2}-\frac{1}{4}\right)+\left(\frac{1}{6}-\frac{1}{8}\right)+\cdots+\left(\frac{1}{4n-2}-\frac{1}{4n}\right),$$

so daß

$$S_{3n}^* = \frac{1}{2}\left[1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\cdots+\frac{1}{2n-1}-\frac{1}{2n}\right].$$

Es ist also

$$S_{3n}^* = \frac{1}{2}S_{2n}.$$

Daraus ergibt sich, daß die Folge $\{S_{3n}^*\}$ mit wachsendem n gegen $\frac{1}{2}S$ konvergiert. Aus

$$S_{3n+1}^* = S_{3n}^* + \frac{1}{2n+1}, \quad S_{3n+2}^* = S_{3n+1}^* - \frac{1}{4n+2}$$

folgt sodann, daß aber auch die Folgen $\{S_{3n+1}^*\}$ und $\{S_{3n+2}^*\}$ mit wachsendem n gegen $\frac{1}{2}S$ streben. Somit konvergiert allgemein die Folge $\{S_n^*\}$ mit wachsendem n gegen $\frac{1}{2}S$,

[1]) Weil $2k-1$ die k-te ungerade Zahl und $4k$ die $2k$-te gerade Zahl ist, steht in der k-ten Klammer der Ausdruck $\frac{1}{2k-1}-\frac{1}{4k-2}-\frac{1}{4k}$.

so daß die Reihe (45) die Summe $\frac{1}{2}S$ besitzt. Die Summe der Reihe (45) ist also (wegen $S \neq 0$) von der der Reihe (44) verschieden.

Wir sehen also, daß eine konvergente Reihe bei einer Umordnung ihrer Glieder ihren Wert ändern kann.

Da die hier angeschnittene Frage von großer methodischer Bedeutung ist, wollen wir sie noch etwas näher untersuchen. Wir wollen als erstes zeigen, daß bei Reihen mit positiven Gliedern ein derartiger „Unglücksfall" nicht eintreten kann.

Satz 1. *Bei einer Reihe mit positiven Gliedern wird durch eine Umordnung der Glieder weder die Konvergenz zerstört noch die Summe geändert.*

Beweis. Es sei

$$a_1 + a_2 + a_3 + \cdots \tag{46}$$

eine konvergente Reihe mit positiven Gliedern und

$$a_{n_1} + a_{n_2} + a_{n_3} + \cdots \tag{47}$$

eine „Umordnung" dieser Reihe, d. h. eine Reihe, die aus denselben Gliedern wie die Reihe (46) besteht, diese nur in irgendeiner anderen Reihenfolge genommen. Mit S bezeichnen wir die Summe der Reihe (46), während unter S_n bzw. S_n^* die n-te Partialsumme der Reihe (46) bzw. (47) verstanden werden soll. Es sei schließlich $m(k)$ die größte der Zahlen $n_1, n_2, \ldots, n_k$. Dann treten unter den Summanden der Summe

$$S_{m(k)} = a_1 + a_2 + \cdots + a_{m(k)}$$

alle Summanden der Summe

$$S_k^* = a_{n_1} + a_{n_2} + \cdots + a_{n_k}$$

auf, wobei allerdings die Summe $S_{m(k)}$ noch gewisse zusätzliche (nicht negative) Summanden enthalten kann. Daher ist

$$S_k^* \leqq S_{m(k)}$$

und damit erst recht

$$S_k^* \leqq S. \tag{48}$$

Also ist die Folge der Partialsumme der Reihe (47) nach oben beschränkt, so daß sie, da (47) eine Reihe mit positiven Gliedern ist, konvergiert. Damit ist gezeigt, daß bei einer Umordnung der Glieder die Konvergenz einer Reihe mit positiven Gliedern nicht zerstört wird.

Es bleibt zu zeigen, daß sich bei einer derartigen Umordnung auch die Reihensumme nicht ändert. Dazu bezeichnen wir die Summe der Reihe (47) mit S^*. Aus (48) folgt sofort

$$S^* \leqq S.$$

Also kann bei einer beliebigen Umordnung die Summe einer Reihe mit positiven Gliedern nicht vergrößert werden. Sie kann aber auch nicht verkleinert werden, da in diesem Fall der Übergang von der Reihe (47) zur Reihe (46) eine Umordnung wäre, bei der sich die Reihensumme vergrößern würde, was aber unmöglich ist. Damit ist Satz 1 bewiesen.

Dieser Satz läßt sich nun unmittelbar auf beliebige absolut konvergente Reihen übertragen.

Satz 2. *Bei einer absolut konvergenten Reihe wird durch eine Umordnung der Glieder weder die absolute Konvergenz zerstört noch die Summe geändert.*

Beweis. Es sei

$$a_1 + a_2 + a_3 + \cdots \tag{49}$$

eine beliebige absolut konvergente Reihe. Definitionsgemäß bedeutet dies, daß nicht nur die Reihe (49), sondern sogar die Reihe

$$|a_1| + |a_2| + |a_3| + \cdots \tag{50}$$

konvergiert. Es sei dann

$$a_{n_1} + a_{n_2} + a_{n_3} + \cdots \tag{51}$$

eine beliebige Umordnung der Reihe (49). Es ist klar, daß man die Reihe der absoluten Beträge der Glieder von (51) (in der Reihenfolge der Reihe (51)) durch eine Umordnung aus der Reihe (50) erhalten kann. Daher konvergiert (auf Grund von Satz 1) auch diese Reihe, und das bedeutet gerade, daß die Reihe (51) absolut konvergiert. Um zu zeigen, daß sich bei der betrachteten Umordnung die Summe nicht ändert, untersuchen wir die Reihe

$$b_1 + b_2 + b_3 + \cdots \tag{52}$$

der nicht negativen Glieder von (49) und die Reihe

$$c_1 + c_2 + c_3 + \cdots \tag{53}$$

der absoluten Beträge der negativen Glieder von (49) (beide Reihen in der Reihenfolge des Auftretens ihrer Glieder in (49) genommen). Auf Grund von Satz 2 aus Abschnitt 37 sind beide Reihen konvergent, wobei ihre Summen B und C mit der Summe A der Reihe (49) durch die Beziehung

$$A = B - C \tag{54}$$

verknüpft sind. Es ist nun klar, daß die entsprechenden Reihen zu (51) nichts anderes als gewisse Umordnungen der Reihe (52) bzw. (53) sind. Da nun (52) und (53) Reihen mit positiven Gliedern sind, ändern sich bei dieser Umordnung ihre Summen nicht, d. h., die den Reihen (52) bzw. (53) entsprechenden Reihen zu (51) besitzen ebenfalls die Summe B bzw. C. Beachten wir nun noch, daß sich die Summe von (51) gemäß (54) durch die Summen dieser Reihen ausdrücken läßt, so ist sofort ersichtlich, daß die Summe der Reihe (51) gleich der der Reihe (49) ist, was noch zu beweisen war.

Man kann sogar zeigen, daß die Voraussetzung der absoluten Konvergenz für die Möglichkeit einer beliebigen Umordnung nicht nur hinreichend, sondern auch notwendig ist.[1]) Es gilt nämlich:

Satz 3. *Ist die Reihe* (49) *nicht absolut konvergent, so kann man durch eine passende Umordnung ihrer Glieder erreichen, daß die entstehende Reihe eine beliebig vorgegebene Summe besitzt oder auch divergiert.*

Auf einen Beweis hierfür soll an dieser Stelle jedoch nicht eingegangen werden.

Wir wenden uns jetzt dem Problem der Multiplikation von Reihen zu.

Um zwei endliche Summen

$$A = a_1 + a_2 + \cdots + a_n, \quad B = b_1 + b_2 + \cdots + b_m$$

miteinander zu multiplizieren, muß man bekanntlich jeden Summanden der einen mit jedem Summanden der anderen Summe multiplizieren und sodann die Summe aller dabei auftretenden Produkte

$$a_i b_k \qquad (i = 1, 2, \ldots, n; \quad k = 1, 2, \ldots, m) \tag{55}$$

[1]) Üblicherweise bezeichnet man eine konvergente Reihe, deren Summe sich bei beliebiger Umordnung der Glieder nicht ändert, als *unbedingt konvergent*, während eine konvergente Reihe, die nicht unbedingt konvergent ist, *bedingt konvergent* genannt wird. Satz 3, den man auch als Satz von RIEMANN bezeichnet (B. RIEMANN (1826 — 1866), deutsch. Math.), lautet dann z. B.: *Jede nicht-absolut konvergente Reihe ist bedingt konvergent. — Anm. d. wissenschaftl. Red.*

bilden. Es ergibt sich nun naturgemäß die Frage, ob man diese Regel auch auf den Fall unendlicher Reihen übertragen kann. Die Hauptschwierigkeit hierbei ist, daß in diesem Fall unendlich viele Produkte auftreten und daher genauer gesagt werden muß, in welcher Weise diese Produkte „addiert" werden sollen. Das einfachste Verfahren für die Addition dieser Produkte liegt wohl der folgenden Art der Reihenmultiplikation zugrunde:

Definition. Unter dem *Produkt* der Reihen

$$a_1 + a_2 + a_3 + \cdots \tag{56}$$

und

$$b_1 + b_2 + b_3 + \cdots \tag{57}$$

verstehen wir die Reihe

$$c_1 + c_2 + c_3 + \cdots, \tag{58}$$

wobei

$$c_1 = a_1 b_1, \quad c_2 = a_1 b_2 + a_2 b_1, \quad c_3 = a_1 b_3 + a_2 b_2 + a_3 b_1$$

und allgemein

$$c_n = a_1 b_n + a_2 b_{n-1} + \cdots + a_n b_1. \tag{59}$$

Es zeigt sich nun, daß aus der Konvergenz der Reihen (56) und (57) im allgemeinen noch nicht die der Reihe (58) folgt. Multiplizieren wir z. B. die auf Grund der Leibnizschen Regel konvergente Reihe

$$1 - \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} - \frac{1}{\sqrt{4}} + \cdots$$

mit sich selbst, so erhalten wir als Produkt (58) eine Reihe, für deren allgemeines Glied (59) gilt:

$$|c_n| = 1 \cdot \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{n-1}} + \frac{1}{\sqrt{3}} \cdot \frac{1}{\sqrt{n-2}} + \cdots + \frac{1}{\sqrt{n}} \cdot 1.$$

Offenbar ist aber

$$|c_n| \geqq \frac{1}{\sqrt{n}} \cdot \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{n}} \cdot \frac{1}{\sqrt{n}} + \cdots + \frac{1}{\sqrt{n}} \cdot \frac{1}{\sqrt{n}} = 1,$$

so daß die Reihe (58) divergiert.

Es gilt aber:

Satz 4. *Sind die Reihen* (56) *und* (57) *absolut konvergent, so ist auch die Reihe* (58) *absolut konvergent. Ist dabei A die Summe von* (56), *B die von* (57) *und C die von* (58), *so ist*

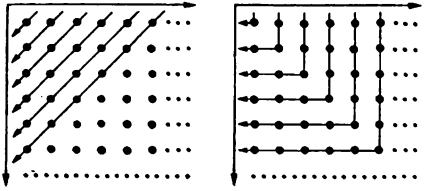
$$C = AB. \tag{60}$$

Mit anderen Worten besagt dies: *Absolut konvergente Reihen können wie endliche Summen miteinander multipliziert werden.*

Zum Beweis dieses Satzes ordnen wir alle Produkte (55) in folgender Weise zu einer unendlichen Matrix an:

$$\begin{array}{llllll} a_1 b_1 & a_1 b_2 & a_1 b_3 & \cdots & a_1 b_m & \cdots \\ a_2 b_1 & a_2 b_2 & a_2 b_3 & \cdots & a_2 b_m & \cdots \\ a_3 b_1 & a_3 b_2 & a_3 b_3 & \cdots & a_3 b_m & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_n b_1 & a_n b_2 & a_n b_3 & \cdots & a_n b_m & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{array} \tag{61}$$

Die Elemente dieser Matrix kann man nun auf die verschiedenste Art in einer Folge anordnen. Für uns sind besonders die in den nachfolgenden Schemata angedeuteten Verfahren wichtig:



Das erste Verfahren (Anordnung nach Schräglinien) führt auf die Reihe

$$a_1 b_1 + a_1 b_2 + a_2 b_1 + a_1 b_3 + a_2 b_2 + a_3 b_1 + a_1 b_4 + a_2 b_3 + \cdots \tag{62}$$

während das zweite Verfahren (Anordnung nach Quadraten) auf die Reihe

$$a_1 b_1 + a_1 b_2 + a_2 b_2 + a_2 b_1 + a_1 b_3 + a_2 b_3 + a_3 b_3 + a_3 b_2 + \cdots \tag{63}$$

führt.[1])

Es ist klar, daß jede Reihe, die in irgendeiner Reihenfolge aus den Produkten der Matrix (61) aufgebaut ist, aus jeder anderen derartigen Reihe durch Umordnung hervorgeht. Wir wollen nun zeigen, daß alle diese Reihen absolut konvergieren.

Dazu betrachten wir eine beliebige derartige Reihe, bilden die Reihe der absoluten Beträge ihrer Glieder und bezeichnen mit S_n^* die n-te Partialsumme dieser Reihe der absoluten Beträge. Bei gegebenem n können wir dann offenbar eine Zahl m so bestimmen, daß alle Summanden der Summe S_n^* unter den Zahlen des Quadrates

$$\boxed{\begin{array}{cccc} |a_1\, b_1| & |a_1\, b_2| & \ldots & |a_1\, b_m| \\ |a_2\, b_1| & |a_2\, b_2| & \ldots & |a_2\, b_m| \\ \cdot & \cdot & \cdot & \cdot \\ |a_m\, b_1| & |a_m\, b_2| & \ldots & |a_m\, b_m| \end{array}}$$

vorkommen. Hieraus folgt, daß die Summe S_n^* nicht größer sein kann als die Summe aller Zahlen dieses Quadrates, d. h. nicht größer als

$$(|a_1| + |a_2| + \cdots + |a_m|)\,(|b_1| + |b_2| + \cdots + |b_m|).$$

Bezeichnen wir nun mit A^* bzw. B^* die Summe der Reihe der absoluten Beträge von (56) bzw. (57), so ist also erst recht

$$S_n^* \leqq A^* B^*.$$

[1]) Man bezeichnet häufig (62) als CAUCHYsches und (63) als DIRICHLETsches Produkt. — *Anm. d. wissenschaftl. Red.*

Hieraus folgt aber unmittelbar, daß alle Reihen, die sich aus sämtlichen Produkten der Matrix (61) bilden lassen, absolut konvergieren. Insbesondere konvergiert also die Reihe (62) absolut. Hierdurch ist speziell die Konvergenz der Reihe

$$|a_1 b_1| + |a_1 b_2| + |a_2 b_1| + |a_1 b_3| + |a_2 b_2| + |a_3 b_1| + |a_1 b_4| + \cdots \tag{64}$$

gesichert. Vereinigen wir nun in (64) die Glieder zu Summen von ein, zwei, drei usw. Gliedern, so finden wir, daß auch die Reihe

$$d_1 + d_2 + d_3 + \cdots$$

mit

$$d_n = |a_1 b_n| + |a_2 b_{n-1}| + \cdots + |a_n b_1|$$

konvergiert. Nun ist aber offenbar $|c_n| \leqq d_n$. Daraus folgt, daß auch die Reihe (58) absolut konvergiert.

Es bleibt demnach zu zeigen, daß die Gleichung (60) besteht.

Da alle Reihen, die man aus allen Produkten der Matrix (61) bilden kann, absolut konvergieren und jede von ihnen durch Umordnung jeder beliebigen anderen erhalten werden kann, haben sie alle dieselbe Summe S. Insbesondere besitzen also die Reihen (62) und (63) die Summe S, und da (58) aus (62) durch Zusammenfassung der Glieder in passende Gruppen hervorgeht, besitzt auch (58) die Summe S, also $S = C$. Daher ist speziell C auch die Summe der Reihe (63). Diese letzte Summe kann man aber nun auch dadurch bestimmen, daß man die Glieder der Reihe (63) in Gruppen zu ein, drei, fünf usw. Gliedern zusammenfaßt, d. h. als Summe der Reihe

$$a_1 b_1 + (a_1 b_2 + a_2 b_2 + a_2 b_1) + (a_1 b_3 + a_2 b_3 + a_3 b_3 + a_3 b_2 + a_3 b_1) + \cdots.$$

Die n-te Partialsumme dieser Reihe ist nun offenbar nichts anderes als

$$\begin{aligned} &a_1 b_1 + a_1 b_2 + \cdots + a_1 b_n \\ &+ a_2 b_1 + a_2 b_2 + \cdots + a_2 b_n \\ &\cdots\cdots\cdots\cdots \\ &+ a_n b_1 + a_n b_2 + \cdots + a_n b_n, \end{aligned}$$

und das ist nichts anderes als das Produkt

$$(a_1 + a_2 + \cdots + a_n)(b_1 + b_2 + \cdots + b_n)$$

aus der n-ten Partialsumme von (56) und der n-ten Partialsumme von (57). Gehen wir hierbei zur Grenze $n \to \infty$ über, so erhalten wir aber offenbar gerade die Gleichung (60).

Wir führen zum Abschluß noch ohne Beweis die beiden folgenden Sätze an:

Satz 5. *Wenn die Reihen* (56) *und* (57) *konvergieren und wenigstens eine von ihnen absolut konvergent ist, so konvergiert auch (allerdings nicht notwendig absolut) die Reihe* (58), *und die Summen der Reihen* (56), (57) *und* (58) *hängen vermöge* (60) *miteinander zusammen.*

Satz 6. *Konvergieren alle drei Reihen* (56), (57) *und* (58), *so gilt* (60).

§ 8. Potenzreihen

39. Das Konvergenzintervall. Unter einer *Potenzreihe* versteht man eine Reihe der Form

$$c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \cdots, \tag{1}$$

wobei c_0, c_1, c_2, ... konstante Zahlen, die sogenannten *Koeffizienten* der Reihe sind. Mitunter bezeichnet man auch allgemeiner jede Reihe der Form

$$c_0 + c_1 (x - a) + c_2 (x - a)^2 + \cdots \tag{2}$$

als Potenzreihe. In diesem Fall nennt man eine Reihe der Form (1) ausführlicher eine *Potenzreihe in* x oder noch ausführlicher eine *Potenzreihe, die nach Potenzen von* x *geordnet* (bzw. *entwickelt*) *ist*. Entsprechend heißt dann eine Potenzreihe der Form (2) eine *Potenzreihe in* $(x-a)$ bzw. eine *Potenzreihe, die nach Potenzen von* $x-a$ *geordnet ist*. Da man offenbar jede Potenzreihe der Form (2) durch die Substitution $x-a=x'$ auf die Form (1) bringen kann, werden wir uns im folgenden vornehmlich mit Potenzreihen der Form (1) beschäftigen.

Während man die Reihen mit konstanten Gliedern rein formal als Summen mit unendlich vielen Summanden ansehen kann, stellen vom rein formalen Standpunkt aus die Potenzreihen „Polynome unendlichen Grades" dar.

Die Brauchbarkeit des Begriffs der Potenzreihe läßt sich bereits sehr schön an folgenden Beispielen erkennen: In Abschnitt 12 haben wir gesehen, daß für jede reelle Zahl x die Beziehung

$$\sin x = \lim_{p\to\infty}\left[x-\frac{x^3}{3!}+\frac{x^5}{5!}-\cdots+(-1)^p\frac{x^{2p+1}}{(2p+1)!}\right]$$

gilt. Unter Verwendung der Bezeichnungen des vorangehenden Paragraphen können wir hierfür kurz

$$\sin x = x-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\cdots \tag{3}$$

schreiben. Dies besagt aber gerade, daß sich die trigonometrische Funktion $\sin x$ als Potenzreihe der Form (1) darstellen läßt. Entsprechend erhält man auf Grund der Überlegungen aus Abschnitt 12:

$$\cos x = 1-\frac{x^2}{2!}+\frac{x^4}{4!}-\frac{x^6}{6!}+\cdots, \tag{4}$$

$$e^x = 1+\frac{x}{1!}+\frac{x^2}{2!}+\frac{x^3}{3!}+\frac{x^4}{4!}+\cdots. \tag{5}$$

Die Potenzreihen sind also analytische Ausdrücke, mit deren Hilfe sich viele wichtige Funktionen darstellen lassen.

Die Darstellung einer Funktion als Summe einer Potenzreihe bietet im allgemeinen erhebliche Vorteile. So kann man mit ihrer Hilfe die Werte der Funktion tatsächlich berechnen; denn die Partialsummen einer Potenzreihe sind gewöhnliche algebraische Polynome, die man ohne große Mühe berechnen kann, wobei diese Partialsummen bei hinreichend großer Gliederzahl die Funktion beliebig genau approximieren. Außerdem lassen sich aus der Potenzreihe häufig viele Eigenschaften der Funktion sehr einfach ablesen. So kann man z. B. aus den Darstellungen (3) und (4) unmittelbar entnehmen, daß

$$\sin(-x) = -\sin x, \quad \cos(-x) = \cos x,$$

d. h., daß $\sin x$ eine ungerade und $\cos x$ eine gerade Funktion ist.

Wir hoffen, daß die vorangehenden Ausführungen hinreichend das Interesse motivieren, das wir im vorliegenden Artikel der Theorie der Potenzreihen

entgegenbringen. Später wird die große Bedeutung dieser Theorie (nicht zuletzt für die Bedürfnisse der Elementarmathematik) noch deutlicher in Erscheinung treten.

Offenbar hat es bei einer Potenzreihe (wie auch bei jeder anderen Reihe, bei der die Glieder von einer Veränderlichen x abhängen) keinen Sinn, nach ihrer Konvergenz oder Divergenz zu fragen; denn es ist klar, daß eine gegebene Potenzreihe im allgemeinen für gewisse Werte x konvergieren und für andere divergieren wird. So ist z. B. die Reihe

$$1 + 2x + 3x^2 + 4x^3 + \cdots$$

offenbar für $x = 0$ konvergent und für $x = 1$ divergent.

Die vernünftige Fragestellung lautet hier selbstverständlich folgendermaßen: Für welche Werte x konvergiert die Reihe (1) und für welche divergiert sie? Es wird sich nun zeigen, daß die Menge derjenigen Werte x, für welche eine Potenzreihe konvergiert, stets eine sehr einfache Struktur besitzt. Es wird sich nämlich herausstellen, daß diese Menge stets ein zum Ursprung symmetrisches Intervall ist. Zum Beweis dieser Tatsache benötigen wir den folgenden Hilfssatz (der zuerst von dem norwegischen Mathematiker N. H. Abel, 1802—1829, bewiesen wurde):

Abelsches Lemma. *Konvergiert die Potenzreihe* (1) *für einen gewissen Wert* $x_0 \neq 0$, *so ist sie für jedes* x *mit*

$$|x| < |x_0| \tag{6}$$

absolut konvergent.

Beweis. Da die Glieder der nach Voraussetzung konvergenten Reihe

$$c_0 + c_1 x_0 + c_2 x_0^2 + c_3 x_0^3 + \cdots$$

mit wachsendem n gegen Null streben, ist

$$\lim_{n \to \infty} c_n x_0^n = 0.$$

Nun ist bekanntlich jede konvergente Folge beschränkt. Daher gibt es eine Zahl M, so daß

$$|c_n x_0^n| < M \qquad (n = 0, 1, 2, \ldots). \tag{7}$$

Wir betrachten nun die Reihe (1) für einen beliebigen Wert x, welcher die Ungleichung (6) erfüllt. Wegen $x_0 \neq 0$ können wir diese Reihe auch folgendermaßen schreiben:

$$c_0 + c_1 x_0 \frac{x}{x_0} + c_2 x_0^2 \left(\frac{x}{x_0}\right)^2 + c_3 x_0^3 \left(\frac{x}{x_0}\right)^3 + \cdots . \tag{8}$$

Auf Grund von (7) ist nun offenbar die Reihe

$$M + M\left|\frac{x}{x_0}\right| + M\left|\frac{x}{x_0}\right|^2 + M\left|\frac{x}{x_0}\right|^3 + \cdots$$

eine Majorante für die Reihe der absoluten Beträge der Glieder von (8). Da jene $\left(\text{als geometrische Reihe mit dem Quotienten } \left|\frac{x}{x_0}\right| < 1\right)$ konvergiert, ist auch die Reihe der absoluten Glieder von (8) konvergent, so daß also die Reihe (1) für den betrachteten Wert x absolut konvergiert, was zu beweisen war.

Wir betrachten nun eine beliebige Potenzreihe (1) und bilden die Menge A aller derjenigen Zahlen x, für welche die betrachtete Reihe konvergiert. Diese Menge ist nicht leer, da jede Potenzreihe für $x = 0$ konvergiert. Es sei ferner B die Menge aller Zahlen $|x|$ mit $x \in A$ und R die obere Grenze von B,[1])

$$R = \sup B.$$

Wir unterscheiden dann die folgenden drei Fälle:

$$1.\ R = 0, \quad 2.\ 0 < R < +\infty, \quad 3.\ R = +\infty.$$

Im ersten Fall konvergiert die Reihe (1) nur für $x = 0$. Würde sie nämlich für ein gewisses $x \neq 0$ konvergieren, so wäre $x \in A$ und daher $|x| \in B$, also $|x| \leqq R = 0$, im Widerspruch zu $x \neq 0$.

Wir untersuchen als nächstes den Fall 2. In diesem Fall konvergiert die Reihe (1) (und zwar absolut) für alle Werte x aus dem offenen Intervall $(-R, +R)$, während sie für alle x, welche außerhalb des abgeschlossenen Intervalls $[-R, +R]$ liegen, divergiert. In der Tat: Es sei zunächst $x \in (-R, +R)$. Dann ist $|x| < R$, so daß es in der Menge B eine Zahl gibt, welche größer als $|x|$ ist. Nach Definition der Menge B ist diese Zahl der absolute Betrag einer Zahl x_0 aus der Menge A, d. h., es ist $|x| < |x_0|$ mit $x_0 \in A$. Dann ist aber auf Grund des Abelschen Lemmas die Reihe (1) im Punkte x absolut konvergent, wie behauptet. Liegt dagegen x außerhalb des Intervalls $[-R, +R]$, so ist $|x| > R$. Daher liegt in diesem Fall $|x|$ nicht in B und daher x nicht in A, d. h., die Reihe (1) ist im Punkte x divergent.

Im Fall 3 ist schließlich die Menge B nach oben nicht beschränkt, d. h., es gibt für jede Zahl x ein Element aus B, so daß $|x|$ kleiner als dieses Element ist. Hieraus folgt, daß es zu jedem x ein $x_0 \in A$ gibt, so daß $|x| < |x_0|$. Auf Grund des Abelschen Lemmas ist daher im Fall 3 die Reihe (1) in jedem Punkte x absolut konvergent.

Damit erhalten wir:

Satz. *Zu jeder Reihe* (1) *gibt es eine nicht negative Zahl* R (den sogenannten *Konvergenzradius* der Reihe (1)), *so daß die Reihe* (1) *in allen Punkten des offenen Intervalls* $(-R, +R)$ *absolut konvergiert, während sie außerhalb des abgeschlossenen Intervalls* $[-R, +R]$ *überall divergiert.*

Bemerkungen.

1. Der angegebene Satz besagt, daß die Menge der Punkte x, in welchen die Potenzreihe (1) konvergiert, stets ein gewisses Intervall bildet, welches

[1]) Ist B nach oben nicht beschränkt, so setzen wir $R = +\infty$.

symmetrisch zum Punkte $x = 0$ liegt. Dieses Intervall bezeichnet man als *Konvergenzintervall* der Reihe (1).

2. In Spezialfällen kann das Konvergenzintervall auch in den einen Punkt $x = 0$ ausarten, was dem Fall $R = 0$ entspricht. Es gibt jedoch auch Reihen, bei denen das Konvergenzintervall die gesamte Zahlengerade ist (es ist dies der Fall $R = +\infty$).

3. Im angegebenen Satz ist nichts darüber ausgesagt, wie sich die Reihe in den Endpunkten des Konvergenzintervalls, d. h. in den Punkten $x = -R$ und $x = +R$ verhält. Wir werden im folgenden sehen, daß hier praktisch alle Möglichkeiten eintreten können, so daß man über das Konvergenzverhalten einer Potenzreihe in den Endpunkten des Konvergenzintervalls keine allgemeingültige Aussage machen kann.

4. Wir haben bereits oben bemerkt, daß man jede Potenzreihe der Form (2) durch die Substitution $x - a = x'$ auf die Form (1) bringen kann. Daher überträgt sich alles im vorangehenden über Reihen der Form (1) Gesagte sinngemäß auf Reihen der Form (2). Insbesondere bildet die Menge der Punkte, in welchen eine Reihe der Form (2) konvergiert, ebenfalls ein Intervall, welches jetzt jedoch symmetrisch zum Punkte $x = a$ liegt, so daß seine Endpunkte die Form $a - R$ bzw. $a + R$ haben.

Wir bringen jetzt einige Beispiele. An Hand dieser und ähnlicher Beispiele können wir uns davon überzeugen, daß alle Fälle, die auf Grund des angegebenen Satzes möglich sind, in der Praxis wirklich auftreten.

Beispiel 1. Die Reihe

$$\frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots$$

konvergiert für alle x (sie ist, wie man sagt, *beständig konvergent — Anm. d. wissenschaftl. Red.*). Dies ergibt sich unmittelbar mit Hilfe des D'ALEMBERTschen Quotientenkriteriums (Abschnitt 37, Satz 3); denn im vorliegenden Fall ist

$$a_n = \frac{x^n}{n!}, \quad a_{n+1} = \frac{x^{n+1}}{(n+1)!}, \quad \lim_{n\to\infty} \frac{a_{n+1}}{a_n} = \lim_{n\to\infty} \frac{x}{n+1} = 0.$$

Beispiel 2. Die Reihe

$$1!\,x + 2!\,x^2 + 3!\,x^3 + \cdots$$

konvergiert nur für $x = 0$; denn für $x \neq 0$ ist

$$a_n = n!\,x^n, \quad a_{n+1} = (n+1)!\,x^{n+1}, \quad \frac{a_{n+1}}{a_n} = (n+1)\,x \to \infty.$$

Beispiel 3. Die Reihe

$$1 + x + x^2 + x^3 + \cdots$$

Konvergiert bekanntlich für $|x| < 1$, während sie für $|x| \geqq 1$ divergiert. Im betrachteten Fall ist also Konvergenzintervall das offene Intervall $(-1, +1)$, in dessen Endpunkten die Reihe divergiert.

Beispiel 4. Für die Reihe

$$x+\frac{x^2}{2^2}+\frac{x^3}{3^2}+\frac{x^4}{4^2}+\cdots \tag{9}$$

ist

$$a_n=\frac{x^n}{n^2},\quad a_{n+1}=\frac{x^{n+1}}{(n+1)^2},\quad \frac{a_{n+1}}{a_n}=\left(\frac{n}{n+1}\right)^2 x\to x,$$

so daß auf Grund des D'ALEMBERTschen Kriteriums die Reihe (9) für $|x|<1$ konvergiert und für $|x|>1$ divergiert. Für $x=1$ erhalten wir die Reihe

$$1+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\cdots, \tag{10}$$

von der bereits in Abschnitt 35 gezeigt wurde, daß sie konvergiert. Für $x=-1$ erhalten wir schließlich eine Reihe, für welche die Reihe (10) die zugehörige Reihe der absoluten Glieder ist. Daher ist die Reihe (9) auch im Punkte $x=-1$ konvergent, und zwar sogar absolut konvergent. Konvergenzintervall der Reihe (9) ist also das abgeschlossene Intervall $[-1,+1]$. In den Endpunkten dieses Intervalls konvergiert die Reihe (9) absolut.

Beispiel 5. Eine entsprechende Untersuchung der Reihe

$$x^2-\frac{x^4}{2}+\frac{x^6}{3}-\frac{x^8}{4}+\cdots$$

zeigt, daß auch für sie das abgeschlossene Intervall $[-1,+1]$ Konvergenzintervall ist, wobei jetzt aber die Reihe in beiden Endpunkten nicht-absolut konvergiert.

Beispiel 6. Die Reihe

$$x+\frac{x^2}{2}+\frac{x^3}{3}+\frac{x^4}{4}+\cdots$$

konvergiert für $-1\leqq x<+1$, so daß in diesem Fall der linke Endpunkt des Konvergenzintervalls (mit nicht-absoluter Konvergenz) dem Konvergenzintervall angehört, während im rechten Endpunkt des Konvergenzintervalls die Reihe bereits nicht mehr konvergiert.

In sämtlichen Beispielen, bei denen $0<R<+\infty$ war, ergab sich als Konvergenzradius $R=+1$. Ersetzt man bei irgendeiner dieser Reihen x überall durch $\frac{x}{a}$ $(a>0)$, so erhält man eine Reihe, die in den Endpunkten ihres Konvergenzintervalls ein entsprechendes Verhalten wie die betrachtete Reihe zeigt, die jedoch den Konvergenzradius $R=a$ besitzt.

40. Eigenschaften der Summe einer Potenzreihe. Es sei

$$c_0+c_1x+c_2x^2+c_3x^3+\cdots \tag{11}$$

eine Potenzreihe mit einem von Null verschiedenen Konvergenzradius R (der Fall $R=+\infty$ soll dabei nicht ausgeschlossen sein). Die Summe dieser Reihe ist dann eine in allen Punkten des Konvergenzintervalls definierte Funktion von x, die wir mit $f(x)$ bezeichnen wollen. Unsere Aufgabe soll es hier sein, diese Funktion näher zu untersuchen.

Satz 1. *Die Funktion $f(x)$ ist in allen Punkten, die im Innern des Konvergenzintervalls liegen, stetig.*

Wir weisen darauf hin, daß hier nur von den Punkten die Rede ist, die im Innern des Konvergenzintervalls liegen. Die Endpunkte $x = +R$ schließen wir aus, obwohl die Reihe auch in diesen Punkten noch konvergieren kann.[1])

Beweis. Es sei x_0 ein beliebiger Punkt aus dem Innern des Konvergenzintervalls, d. h., es sei $-R < x_0 < +R$. Es ist zu zeigen, daß die Funktion $f(x)$ in diesem Punkte stetig ist. Zu diesem Zweck betrachten wir zunächst noch einen zweiten Punkt X mit

$$-X < x_0 < X, \quad 0 < X < R.$$

Dieser Punkt liegt ebenfalls im Innern des Konvergenzintervalls, so daß in ihm die Reihe (11) absolut konvergiert, d. h., die Reihe

$$|c_0| + |c_1| X + |c_2| X^2 + |c_3| X^3 + \cdots \tag{12}$$

ist konvergent. Wir betrachten nun einen beliebigen Punkt x aus dem Intervall $[-X, +X]$ (insbesondere kann dies auch der Punkt x_0 sein) und bilden die Reihe

$$c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \cdots . \tag{13}$$

Mit $\overline{R}_n$ bzw. $R_n(x)$ bezeichnen wir den n-ten Rest der Reihe (12) bzw. (13), d. h.

$$\overline{R}_n = |c_{n+1}| X^{n+1} + |c_{n+2}| X^{n+2} + \cdots ,$$

$$R_n(x) = c_{n+1} x^{n+1} + c_{n+2} x^{n+2} + \cdots .$$

Man sieht unmittelbar, daß

$$|R_n(x)| \leqq \overline{R}_n .^{2}) \tag{14}$$

Wir betrachten nun eine beliebige reelle Zahl $\varepsilon > 0$. Zu ihr bestimmen wir eine Zahl n so groß, daß

$$\overline{R}_n < \frac{\varepsilon}{3} .$$

Dies ist möglich, weil die Summen der Reste der konvergenten Reihe (12) mit wachsendem n gegen Null streben. Wegen (14) ist dann für dieses n und für alle x aus dem Intervall $[-X, +X]$

$$|R_n(x)| < \frac{\varepsilon}{3} .$$

---

[1]) Man kann zeigen, daß bei Konvergenz der Reihe in einem der Endpunkte die Funktion $f(x)$ auch noch in diesem Punkte (einseitig) stetig ist; wir wollen jedoch darauf hier nicht näher eingehen.

[2]) Man erhält dies, wenn man in der offenbar richtigen Ungleichung

$$\left| \sum_{k=n+1}^{n+m} c_k x^k \right| \leqq \sum_{k=n+1}^{n+m} |c_k| X^k$$

zur Grenze $m \to +\infty$ übergeht.

Insbesondere ist also auch $|R_n(x_0)| < \frac{\varepsilon}{3}$. Wir setzen ferner

$$S_n(x) = c_0 + c_1 x + \cdots + c_n x^n,$$

so daß

$$f(x) = S_n(x) + R_n(x), \quad f(x_0) = S_n(x_0) + R_n(x_0).$$

Offenbar ist dann

$$|f(x) - f(x_0)| \leqq |S_n(x) - S_n(x_0)| + |R_n(x)| + |R_n(x_0)|$$

und daher erst recht

$$|f(x) - f(x_0)| < |S_n(x) - S_n(x_0)| + \frac{2}{3}\varepsilon. \tag{15}$$

Diese Ungleichung gilt für alle x aus dem Intervall $[-X, +X]$. Nun ist aber $S_n(x)$ ein gewöhnliches algebraisches Polynom, also insbesondere stetig. Daher können wir zu der betrachteten Zahl $\varepsilon > 0$ eine Zahl $\delta > 0$ so finden, daß für $|x - x_0| < \delta$ stets

$$|S_n(x) - S_n(x_0)| < \frac{\varepsilon}{3} \tag{16}$$

wird. Ist also $x - x_0| < \delta$ und $x \in [-X, +X]$, so ist auf Grund von (15) und (16)

$$|f(x) - f(x_0)| < \varepsilon.$$

Das besagt aber gerade, daß die Funktion $f(x)$ im Punkte x_0 stetig ist.

Satz 2. *Es sei*

$$f(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \cdots. \tag{17}$$

Dann ist für jedes ganz im Innern des Konvergenzintervalls gelegene Intervall $[a, b]$

$$\int_a^b f(x)\,dx = c_0(b-a) + c_1 \frac{b^2 - a^2}{2} + c_2 \frac{b^3 - a^3}{3} + \cdots, \tag{18}$$

d. h., für jedes a und b mit $-R < a < b < R$ kann die Gleichung (17) *gliedweise integriert werden.*

Der Beweis hierfür verläuft ähnlich dem Beweis zu Satz 1. Wir beginnen wieder damit, daß wir zunächst eine Zahl X betrachten, für welche

$$-R < -X \leqq a < b \leqq X < R$$

gilt. Unter Verwendung der eben definierten Bezeichnungen gilt dann für jedes n und jedes $x \in [-X, +X]$ die Ungleichung

$$|R_n(x)| \leqq \overline{R}_n.$$

Daher ist für alle $x \in [-X, +X]$ und daher erst recht für alle $x \in [a, b]$

$$|f(x) - S_n(x)| \leqq \overline{R}_n.$$

Hieraus folgt (auf Grund von Satz 11 aus Abschnitt 23)

$$\left| \int_a^b \{f(x) - S_n(x)\}\, dx \right| \leqq \overline{R}_n (b-a),$$

d. h.

$$\left| \int_a^b f(x)\, dx - \int_a^b S_n(x)\, dx \right| \leqq \overline{R}_n (b-a).$$

Hier strebt nun die rechte Seite mit wachsendem n gegen Null. Daher ist

$$\int_a^b f(x)\, dx = \lim_{n \to \infty} \int_a^b S_n(x)\, dx.$$

Es bleibt jetzt lediglich noch zu bemerken, daß

$$\int_a^b S_n(x)\, dx = c_0 (b-a) + c_1 \frac{b^2 - a^2}{2} + \cdots + c_n \frac{b^{n+1} - a^{n+1}}{n+1},$$

daß also dieses Integral nichts anderes als die n-te Partialsumme der Reihe auf der rechten Seite von (18) ist.

Satz 3. *Die durch* (17) *definierte Funktion* $f(x)$ *ist in jedem Punkt aus dem Innern des Konvergenzintervalls der Reihe differenzierbar, wobei die Ableitung* $f'(x)$ *durch*

$$f'(x) = c_1 + 2c_2 x + 3c_3 x^2 + 4c_4 x^3 + \cdots$$

gegeben wird, d. h., die Gleichung (17) *darf man in jedem Punkt aus dem Innern des Konvergenzintervalls gliedweise differenzieren.*

Der Beweis dieses Satzes ist verhältnismäßig kompliziert, so daß es sich empfiehlt, zunächst zwei Hilfssätze vorauszuschicken.

Hilfssatz 1. *Die Reihe*

$$q + 2q^2 + 3q^3 + 4q^4 + \cdots \tag{19}$$

konvergiert für jedes q *mit* $0 < q < 1$.

Ein Beweis dieser Behauptung ergibt sich unmittelbar mit Hilfe des D'ALEMBERTschen Konvergenzkriteriums; denn hier ist

$$a_n = n q^n, \quad a_{n+1} = (n+1)\, q^{n+1},$$

$$\lim_{n \to \infty} \frac{a_{n+1}}{a_n} = \lim_{n \to \infty} \frac{n+1}{n}\, q = q < 1.$$

Hilfssatz 2. *Die Potenzreihen*

$$c_1 x + c_2 x^2 + c_3 x^3 + \cdots, \tag{20}$$

$$c_1 x + 2c_2 x^2 + 3c_3 x^3 + \cdots \tag{21}$$

besitzen beide denselben Konvergenzradius.

Beweis. Es sei R bzw. R' der Konvergenzradius der Reihe (20) bzw. (21). Wir nehmen zunächst an, daß $R' > 0$, und betrachten eine beliebige Zahl z mit $0 < z < R'$. Nach dem ABELschen Lemma konvergiert dann die Reihe (21) im Punkte z absolut, d. h., die Reihe

$$|c_1|\, z + 2\,|c_2|\, z^2 + 3\,|c_3|\, z^3 + \cdots$$

ist konvergent. Dann konvergiert aber erst recht die Reihe

$$|c_1|\, z + |c_2|\, z^2 + |c_3|\, z^3 + \cdots .$$

Damit ist gezeigt, daß die Reihe (20) im Punkte z (sogar absolut) konvergiert, so daß

$$z \leqq R. \tag{22}$$

Hierbei unterliegt z nur der Bedingung $0 < z < R'$. Lassen wir nun z gegen R' konvergieren, so erhalten wir aus (22) sofort

$$R' \leqq R. \tag{23}$$

Diese Ungleichung gilt aber nicht nur für $R' > 0$, sondern natürlich auch für $R' = 0$.

Wir zeigen nun, daß auch umgekehrt

$$R \leqq R'. \tag{24}$$

Hierbei können wir offenbar voraussetzen, daß $R > 0$, da im Fall $R = 0$ die Ungleichung (24) trivial erfüllt ist. Es sei dazu z eine beliebige Zahl mit $0 < z < R$. Wir zeigen, daß die Reihe (21) im Punkte z absolut konvergiert. Zu diesem Zweck betrachten wir eine beliebige Zahl y mit $z < y < R$. Dann konvergiert die Reihe (20) im Punkte y, so daß ihr allgemeines Glied $c_n x^n$ für $x = y$ mit wachsendem n gegen Null strebt. Daher gibt es eine Zahl M, so daß für alle n die Ungleichung

$$|c_n y^n| < M \tag{25}$$

erfüllt ist. Wir schreiben nun die Reihe

$$|c_1|\, z + 2\,|c_2|\, z^2 + 3\,|c_3|\, z^3 + \cdots \tag{26}$$

in der Form

$$|c_1 y|\,\frac{z}{y} + 2\,|c_2 y^2|\left(\frac{z}{y}\right)^2 + 3\,|c_3 y^3|\left(\frac{z}{y}\right)^3 + \cdots .$$

Wegen (25) sind die Glieder dieser Reihe kleiner als die entsprechenden Glieder der Reihe

$$M q + 2 M q^2 + 3 M q^3 + \cdots \qquad \left(q = \frac{z}{y}\right), \tag{27}$$

so daß also die Reihe (27) eine Majorante für die Reihe (26) ist. Nach Hilfssatz 1 ist aber die Reihe (27) konvergent, so daß auch die Reihe (26) konvergiert. Daraus folgt insbesondere, daß die Reihe (21) im Punkte z konvergiert. Also ist

$$z \leqq R',$$

woraus wir durch Grenzübergang $z \to R$ die Ungleichung (24) erhalten. Damit ist auch Hilfssatz 2 bewiesen.

Aus Hilfssatz 2 ergibt sich unmittelbar:

Folgerung. *Die Reihe* (11) *und die aus ihr durch gliedweise Differentiation entstehende Reihe*

$$c_1 + 2c_2 x + 3c_3 x^2 + \cdots \tag{28}$$

besitzen beide denselben Konvergenzradius.[1])

---

[1]) Man darf jedoch nicht glauben, daß die Konvergenzintervalle dieser Reihen übereinstimmen. Zum Beispiel besitzt die Reihe $\Sigma \frac{x^n}{n^2}$ das abgeschlossene Intervall $[-1, +1]$ als Konvergenzintervall, während die abgeleitete Reihe im Punkte $x = 1$ divergiert.

Beweis. Zunächst ist klar, daß die Reihe (11) denselben Konvergenzradius wie ihr Rest (20) besitzt. Nach Hilfssatz 2 stimmt aber der Konvergenzradius der Reihe (20) mit dem der Reihe (21) überein. Es genügt also zu zeigen, daß die Reihen (21) und (28) denselben Konvergenzradius besitzen. Es zeigt sich nun, daß diese Reihen sogar dasselbe Konvergenzintervall besitzen. In der Tat: Wenn die Reihe (28) für einen gegebenen Wert x konvergiert, so können wir die Konvergenz sicher nicht dadurch zerstören, daß wir alle Glieder von (28) mit x multiplizieren, d. h., es konvergiert auch die Reihe (21) für dieses x. Wenn dagegen die Reihe (28) für einen gegebenen Wert x divergiert, so ist $x \neq 0$ und wir können die Reihe (28) nicht dadurch in eine konvergente Reihe verwandeln, daß wir alle ihre Glieder mit dieser Zahl x multiplizieren.

Nach diesen Vorbereitungen kommen wir jetzt zum

Beweis von Satz 3. Zu diesem Zweck bezeichnen wir mit R den Konvergenzradius der Reihe auf der rechten Seite von (17), und zwar sei dabei $R > 0$. Wir differenzieren dann diese Reihe gliedweise, betrachten also die Reihe (28), die ebenfalls den Konvergenzradius R besitzt. Die Summe dieser Reihe werde mit $\varphi(x)$ bezeichnet. Nach Satz 1 ist dann die Funktion $\varphi(x)$ im Intervall $(-R, +R)$ stetig. Wir betrachten nun eine beliebige Zahl z mit $0 < z < R$. Offenbar ist dann das abgeschlossene Intervall $[0, z]$ ganz im Innern des Intervalls $(-R, +R)$ enthalten, so daß wir die Gleichung

$$\varphi(x) = c_1 + 2c_2 x + 3c_3 x^2 + \cdots$$

gliedweise integrieren können, d. h.

$$\int_0^z \varphi(x)\, dx = c_1 z + c_2 z^2 + c_3 z^3 + \cdots .$$

Dann ist aber auf Grund von (17)

$$f(z) = c_0 + \int_0^z \varphi(x)\, dx .$$

Hieraus folgt jedoch auf Grund des Satzes über die Differenzierbarkeit eines Integrals als Funktion der oberen Grenze (Abschnitt 24), daß die Funktion $f(x)$ im Punkte z differenzierbar ist, wobei $f'(z) = \varphi(z)$. Entsprechend schließt man im Fall $-R < z < 0$, wobei man an Stelle des Intervalls $[0, z]$ überall das Intervall $[z, 0]$ betrachtet.

Beachten wir nun, daß $\varphi(x)$ die Summe der Reihe (28) ist, so erhalten wir die Behauptung von Satz 3.

Aus Satz 3 ergibt sich unmittelbar

Satz 4. *Die durch die Potenzreihe* (11) *dargestellte Funktion* $f(x)$ *ist in jedem Punkt* x *aus dem Innern des Konvergenzintervalls beliebig oft differenzierbar. Dabei ist für jeden derartigen Punkt* x

$$\begin{aligned} f'(x) &= c_1 + 2c_2 x + 3c_3 x^2 + 4c_4 x^3 + \cdots , \\ f''(x) &= \quad 2c_2 + 3 \cdot 2 c_3 x + 4 \cdot 3 c_4 x^2 + \cdots , \\ f'''(x) &= \quad\quad 3 \cdot 2 \cdot 1 c_3 + 4 \cdot 3 \cdot 2 c_4 x + \cdots \end{aligned}$$

usw. Alle diese Reihen besitzen denselben Konvergenzradius R *wie die Ausgangsreihe* (11).

41. Entwicklung des Logarithmus und Aufstellung einer Logarithmentafel. Wir wollen jetzt die im vorangehenden erhaltenen Resultate auf eine

Frage anwenden, die von fundamentaler Bedeutung für die Elementarmathematik ist. Wir stellen uns nämlich die Aufgabe, eine Logarithmentafel zu schaffen.

Zu diesem Zweck betrachten wir zunächst die geometrische Reihe

$$1-x+x^2-x^3+\cdots .$$

Diese konvergiert genau dann, wenn ihr Quotient $q=-x$ dem Betrage nach kleiner als Eins ist, d. h. für $-1<x<1$, und besitzt dann die Summe $\frac{1}{1+x}$. Somit ist also für $-1<x<1$

$$\frac{1}{1+x}=1-x+x^2-x^3+\cdots .$$

Es sei nun z eine beliebige Zahl mit $-1<z<1$. Dann können wir auf Grund von Satz 2 aus Abschnitt 40 diese Gleichung über dem Intervall $[0, z]$ gliedweise integrieren und erhalten

$$\ln(1+z)=z-\frac{z^2}{2}+\frac{z^3}{3}-\frac{z^4}{4}+\cdots,$$

wofür wir auch

$$\ln(1+x)=x-\frac{x^2}{2}+\frac{x^3}{3}-\frac{x^4}{4}+\cdots \tag{29}$$

schreiben können. Diese Gleichung gilt zunächst für alle x mit $-1<x<1$. Für $x=-1$ verliert diese Gleichung ihren Sinn (denn links steht in diesem Fall $\log 0$, während rechts eine divergente Reihe auftritt). Dagegen ergibt sich auch im Fall $x=+1$ auf beiden Seiten von (29) ein ganz bestimmter numerischer Wert. Es ist jedoch keineswegs klar, ob die entstehende Gleichung tatsächlich richtig ist, d. h., ob die Beziehung

$$\ln 2=1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\cdots \tag{30}$$

besteht.

Wir wollen zeigen, daß dies wirklich der Fall ist. Zu diesem Zweck betrachten wir nicht sofort die unendliche Reihe

$$1-x+x^2-x^3+\cdots,$$

sondern zunächst die endliche Summe

$$1-x+x^2-x^3+x^4-x^5+\cdots+x^{2n-2}-x^{2n-1}=\frac{1-x^{2n}}{1+x} \qquad (x\neq -1).$$

Auf Grund dieser letzten Beziehung ist

$$\frac{1}{1+x}=(1-x+x^2-x^3+\cdots+x^{2n-2}-x^{2n-1})+\frac{x^{2n}}{1+x}$$

und daher

$$\ln 2=\int_0^1\frac{dx}{1+x}=1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\cdots+\frac{1}{2n-1}-\frac{1}{2n}+\int_0^1\frac{x^{2n}}{1+x}\,dx.$$

Nun ist für jedes $x > 0$ offenbar $\frac{x^{2n}}{1+x} < x^{2n}$, so daß

$$\int_0^1 \frac{x^{2n}}{1+x}\,dx < \int_0^1 x^{2n}\,dx = \frac{1}{2n+1}$$

und daher

$$\lim_{n\to\infty} \int_0^1 \frac{x^{2n}}{1+x}\,dx = 0, \quad \text{also} \quad \ln 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots .$$

Damit ist Gleichung (30) bewiesen.[1]) Also gilt die Beziehung (29) für alle x mit $-1 < x \leqq 1$.

Mit Hilfe von Formel (29) sind wir nun prinzipiell in der Lage, eine Logarithmentafel aufzustellen. Da jedoch Formel (29) nur für $-1 < x \leqq 1$ richtig ist, ist die unmittelbare Anwendung von Formel (29) zu umständlich.[2]) Wir werden daher mit ihrer Hilfe zunächst eine weitere Formel herleiten, die für unsere Zwecke besser geeignet ist.

---

[1]) Hier noch ein zweiter Beweis für Gleichung (30): Wir setzen $S_{2n} = 1 - \frac{1}{2} + \frac{1}{3} - \cdots - \frac{1}{2n}$. Indem wir hier auf der rechten Seite zunächst die Summe $1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$ addieren und dann sofort wieder subtrahieren (wobei wir beim Addieren die Summe vorteilhaft in der Form $2\left(\frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2n}\right)$ schreiben), erhalten wir

$$S_{2n} = \left(1 + \frac{1}{2} + \cdots + \frac{1}{2n}\right) - \left(1 + \frac{1}{2} + \cdots + \frac{1}{n}\right) = \frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n},$$

so daß also $S_{2n} = \sum_{k=1}^{n} \frac{1}{1+\frac{k}{n}} \cdot \frac{1}{n}$. Das ist aber offenbar nichts anderes als eine Integralsumme der Funktion $\frac{1}{1+x}$ im Intervall [0, 1], welche für $n \to \infty$ gegen $\int_0^1 \frac{dx}{1+x} = \ln 2$ konvergiert.

[2]) Diese Schwierigkeit läßt sich auch mit Hilfe einiger Kunstgriffe umgehen. Da z. B. die Reihe (30) sehr langsam konvergiert, ist sie zur Berechnung von $\ln 2$ nicht geeignet. Setzen wir aber in (29) $x = -\frac{1}{2}$, so erhalten wir

$$-\ln 2 = -\frac{1}{2} - \frac{1}{2 \cdot 2^2} - \frac{1}{3 \cdot 2^3} - \frac{3}{4 \cdot 2^4} - \cdots .$$

Diese Reihe konvergiert sehr gut, so daß wir bei gegebener Genauigkeit zur Berechnung von $\ln 2$ mit Hilfe dieser Formel wesentlich weniger Glieder berücksichtigen müssen als zur Berechnung von $\ln 2$ nach Formel (30).

Setzen wir in (29) $x = -\frac{1}{3}$, so erhalten wir $\ln \frac{2}{3}$, woraus sich dann (da $\ln 2$ bereits bekannt ist) $\ln 3$ ergibt usw.

Dazu ersetzen wir in (29) erst einmal x durch $-x$. Hierdurch erhalten wir, daß für $-1 \leqq x < 1$ die Beziehung

$$\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \cdots \tag{31}$$

besteht. Subtrahieren wir nun (31) von (29), so ergibt sich

$$\ln\frac{1+x}{1-x} = 2\left(x + \frac{x^3}{3} + \frac{x^5}{5} + \cdots\right).$$

Hier setzen wir nun

$$x = \frac{1}{2N+1},$$

wobei N eine beliebige natürliche Zahl ist. Dann wird $\frac{1+x}{1-x} = \frac{N+1}{N}$, und wir erhalten damit die „Rekursionsformel"

$$\ln(N+1) = \ln N + \frac{2}{2N+1} + \frac{2}{3(2N+1)^3} + \frac{2}{5(2N+1)^5} + \cdots. \tag{32}$$

Durch diese Formel wird die Berechnung von $\ln(N+1)$ auf die von $\ln N$ zurückgeführt. Da nun $\ln 1 = 0$ ist, können wir, indem wir in (32) nacheinander $N = 1$, $N = 2$, $N = 3$, ... setzen, der Reihe nach die Logarithmen aller natürlichen Zahlen bis zu einer uns gerade interessierenden Grenze berechnen.

Dazu ist es natürlich notwendig, daß man von Fall zu Fall den Fehler[1]) abschätzt, den man begeht, wenn man die in (32) auftretende Reihe durch eine ihrer Partialsummen ersetzt. Hierzu ist offenbar der jeweilige Reihenrest

$$\varrho_p(N) = \frac{2}{(2p+1)(2N+1)^{2p+1}} + \frac{2}{(2p+3)(2N+1)^{2p+3}} + \frac{2}{(2p+5)(2N+1)^{2p+5}} + \cdots$$

abzuschätzen. Es ist nun klar, daß

$$0 < \varrho_p(N) < \frac{2}{(2p+1)(2N+1)^{2p+1}} + \frac{2}{(2p+1)(2N+1)^{2p+3}} + \frac{2}{(2p+1)(2N+1)^{2p+5}} + \cdots.$$

Berechnen wir schließlich noch die Summe der hier auftretenden geometrischen Reihe, so finden wir

$$0 < \varrho_p(N) < \frac{1}{2N(N+1)} \cdot \frac{1}{2p+1} \cdot \frac{1}{(2N+1)^{2p-1}}. \tag{33}$$

---

[1]) Unter dem Fehler einer Näherungsgleichung verstehen wir hier ein für allemal diejenige Zahl, welche man zum Näherungswert hinzufügen muß, um den genauen Wert zu erhalten, so daß also der Fehler nichts anderes als die Korrektur des Näherungswertes ist. In diesem Sinne ist der Fehler der Näherungsgleichung $\frac{1}{3} = 0{,}333$ gleich $0{,}000\overline{3}$ und der der Näherungsgleichung $\frac{2}{3} = 0{,}667$ gleich $-0{,}000\overline{3}$.

Dabei ist also $\varrho_p(N)$ der Fehler der Näherungsgleichung

$$\ln(N+1) \approx \ln N + \frac{2}{2N+1} + \frac{2}{3(2N+1)^3} + \cdots + \frac{2}{(2p-1)(2N+1)^{2p-1}}. \quad (34)$$

Als Beispiel wollen wir $\ln 2$ nach Formel (34) berechnen. Hier ist $N = 1$, so daß die Abschätzung (33) die Form

$$0 < \varrho_p(1) < \frac{2}{4(2p+1)\,3^{2p-1}}$$

annimmt. Wählen wir hier $p = 6$, so finden wir, daß wegen

$$\frac{1}{52 \cdot 3^{11}} < 0{,}00000012$$

der Fehler $\varrho_6(1)$ der Näherungsgleichung

$$\ln 2 \approx \frac{2}{3} + \frac{2}{3 \cdot 3^3} + \frac{2}{5 \cdot 3^5} + \frac{2}{7 \cdot 3^7} + \frac{2}{9 \cdot 3^9} + \frac{2}{11 \cdot 3^{11}} \quad (35)$$

den Ungleichungen

$$0 < \varrho_6(1) < 0{,}00000012$$

genügt. Wir schreiben nun jedes Glied der rechten Seite von (35) als Dezimalbruch, und zwar jeweils gerundet auf acht Stellen hinter dem Komma, so daß der absolute Betrag der Korrektur, die man zu der Dezimalbruchentwicklung hinzufügen muß, um den genauen Wert zu erhalten, jeweils nicht größer als

$$0{,}000000005$$

ist. Dabei geben wir in jedem Fall das Vorzeichen der genannten Korrektur an.[1]) Wir finden:

$$\frac{2}{3} = 0{,}66666667\,(-),$$

$$\frac{2}{3 \cdot 3^3} = 0{,}02469136\,(-),$$

$$\frac{2}{5 \cdot 3^5} = 0{,}00164609\,(+),$$

$$\frac{2}{7 \cdot 3^7} = 0{,}00013064\,(+),$$

$$\frac{2}{9 \cdot 3^9} = 0{,}00001129\,(+),$$

$$\frac{2}{11 \cdot 3^{11}} = 0{,}00000103\,(-).$$

Die Summe aller dieser Dezimalzahlen ist gleich

$$0{,}69314708. \quad (36)$$

[1]) Das heißt: Ist die Dezimalzahl zu klein, so wird (+) hinzugefügt, ist sie zu groß, so wird (−) hinzugefügt.

Diese Zahl ist natürlich nicht die genaue Summe der links stehenden Zahlen. Subtrahieren wir jedoch von jeder Dezimalzahl auf der rechten Seite, welche (—) als Korrekturzeichen trägt, 0,000000005 und lassen wir die Dezimalbrüche, welche (+) als Korrekturzeichen tragen, ungeändert, so erhalten wir nach Addition dieser Zahlen eine Zahl, die mit Sicherheit kleiner als die Summe der Zahlen auf der linken Seite ist. Da in unserem Fall drei Dezimalbrüche das Korrekturzeichen (—) tragen, läuft dies gerade darauf hinaus, daß wir von (36) 0,000000015 subtrahieren. Die dabei entstehende Zahl 0,693147065 ist also kleiner als die Summe der Zahlen auf der rechten Seite von (35). Entsprechend können wir diese Summe nach oben abschätzen. Indem wir diese Abschätzungen mit der vorher durchgeführten Abschätzung für $\varrho_6(1)$ kombinieren, erhalten wir

$$0{,}693147065 < \ln 2 < 0{,}693147215, \tag{37}$$

so daß

$$\ln 2 = 0{,}693147\,(+), \tag{38}$$

wobei die hingeschriebenen sechs Ziffern hinter dem Komma richtig sind.

Damit dürften die Grundprinzipien, mit deren Hilfe man die Logarithmen berechnet, hinreichend klar herausgearbeitet sein. Wegen der Bedeutung dieser Frage für die Elementarmathematik wollen wir jedoch noch einige Beispiele behandeln, und zwar wollen wir uns die Aufgabe stellen, eine Tafel für die dekadischen Logarithmen der ganzen Zahlen von 1 bis 10 aufzustellen.

Zu diesem Zweck berechnen wir zunächst noch die natürlichen Logarithmen für die Zahlen 3, 4 und 5.

Setzen wir in (32) $N = 2$, so erhalten wir

$$\ln 3 = \ln 2 + \frac{2}{5} + \frac{2}{3 \cdot 5^3} + \frac{2}{5 \cdot 5^5} + \cdots, \tag{39}$$

wobei die Abschätzung (33) allgemein die Form

$$0 < \varrho_p(2) < \frac{1}{12} \cdot \frac{1}{(2p+1) \cdot 5^{2p-1}}$$

annimmt. Insbesondere ist also

$$0 < \varrho_4(2) < \frac{1}{108 \cdot 5^7} < 0{,}00000013. \tag{40}$$

Ferner ist

$$\frac{2}{5} = 0{,}40000000,$$

$$\frac{2}{3 \cdot 5^3} = 0{,}00533333\,(+),$$

$$\frac{2}{5 \cdot 5^5} = 0{,}00012800,$$

$$\frac{2}{7 \cdot 5^7} = 0{,}00000366\,(-),$$

so daß

$$\frac{2}{5} + \frac{2}{3 \cdot 5^3} + \frac{2}{5 \cdot 5^5} + \frac{2}{7 \cdot 5^7} = 0{,}40546499,$$

wobei der Fehler dieser Näherungsgleichung dem Betrage nach kleiner als 0,000000005 ist. Hieraus und aus (39), (40) und (37) folgt

$$1{,}098\,612\,050 < \ln 3 < 1{,}098\,612\,340, \tag{41}$$

so daß

$$\ln 3 = 1{,}098\,612\,(+), \tag{42}$$

wobei alle sechs Stellen hinter dem Komma genau sind.

Theoretisch könnte man nun $\ln 4$ nach demselben Verfahren bestimmen. Einfacher ist es jedoch, die Beziehung $\ln 4 = 2 \cdot \ln 2$ auszunutzen. Mit ihrer Hilfe ergibt sich aus (37),

$$1{,}386\,294\,130 < \ln 4 < 1{,}386\,294\,430, \tag{43}$$

so daß

$$\ln 4 = 1{,}386\,294\,(+),$$

wobei wiederum alle sechs Stellen hinter dem Komma genau sind.

Setzen wir schließlich in (32) $N = 4$, so finden wir

$$\ln 5 = \ln 4 + \frac{2}{9} + \frac{2}{3 \cdot 9^3} + \frac{2}{5 \cdot 9^5} + \cdots .$$

Dabei ist

$$0 < \varrho_3(4) < \frac{1}{280 \cdot 9^5} < 0{,}0000001. \tag{44}$$

Nun ist

$$\frac{2}{9} = 0{,}222\,222\,22\,(+),$$

$$\frac{2}{3 \cdot 9^3} = 0{,}000\,914\,49\,(+),$$

$$\frac{2}{5 \cdot 9^5} = 0{,}000\,006\,77\,(+),$$

so daß

$$\frac{2}{9} + \frac{2}{3 \cdot 9^3} + \frac{2}{5 \cdot 9^5} = 0{,}223\,143\,48,$$

wobei der Fehler dieser letzten Gleichung positiv und kleiner als 0,000000015 ist. Hieraus und aus (43) und (44) folgt:

$$1{,}609\,437\,610 < \ln 5 < 1{,}609\,438\,025. \tag{45}$$

Wir rechnen nun die natürlichen Logarithmen in dekadische um. Zu diesem Zweck sei zunächst N eine beliebige positive Zahl und $\lg N$ ihr dekadischer Logarithmus. Dann ist

$$N = 10^{\lg N}.$$

Logarithmieren wir diese Gleichung bezüglich der Basis e, so erhalten wir

$$\ln N = \lg N \cdot \ln 10,$$

das heißt

$$\lg N = \frac{\ln N}{\ln 10}.$$

Wir wollen nun als erstes $\ln 10$ berechnen. Dazu beachten wir, daß $\ln 10 = \ln 2 + \ln 5$. Daher ist auf Grund von (37) und (45)

$$2{,}302\,584\,675 < \ln 10 < 2{,}302\,585\,240$$

und daher erst recht

$$2{,}302\,584 < \ln 10 < 2{,}302\,586. \tag{46}$$

Jetzt bereitet es keine Schwierigkeiten, zunächst $\lg 2$ und $\lg 3$ zu berechnen. Um eine Gewähr dafür zu haben, daß die von uns gewonnenen Resultate auch absolut sicher sind, berechnen wir als erstes Grenzen für diese Logarithmen.[1])

Aus den Ungleichungen (37) und (46) folgt

$$\frac{0{,}693147}{2{,}302586} < \lg 2 < \frac{0{,}693148}{2{,}302584}.$$

Indem wir die äußeren Glieder dieser doppelten Ungleichung in Dezimalbrüche umwandeln (wobei wir die untere Schranke nach unten und die obere Schranke nach oben runden), erhalten wir

$$0{,}3010299 < \lg 2 < 0{,}3010305. \tag{47}$$

Multiplizieren wir diese doppelte Ungleichung mit 2 bzw. 3, so finden wir

$$0{,}6020598 < \lg 4 < 0{,}6020611, \tag{48}$$

$$0{,}9030897 < \lg 8 < 0{,}9030915. \tag{49}$$

Entsprechend erhalten wir aus (41) und (46)

$$\frac{1{,}098612}{2{,}302586} < \lg 3 < \frac{1{,}098613}{2{,}302584}$$

und damit

$$0{,}4771204 < \lg 3 < 0{,}4771224. \tag{50}$$

Multiplizieren wir diese doppelte Ungleichung mit 2, so finden wir

$$0{,}9542408 < \lg 9 < 0{,}9542447. \tag{51}$$

Ferner ergibt sich aus (47) und (50)

$$0{,}7781496 < \lg 6 < 0{,}7781525. \tag{52}$$

Indem wir schließlich die Ungleichungskette (47) gliedweise von 1 subtrahieren, finden wir

$$0{,}6989695 < \lg 5 < 0{,}6989710. \tag{53}$$

Damit haben wir bereits die dekadischen Logarithmen für alle ganzen Zahlen von 1 bis 10 mit Ausnahme von $\lg 7$ bestimmt.

Für $\lg 7$ müssen wir auf Formel (32) zurückgreifen. Setzen wir dort $N = 6$, so finden wir

$$\ln 7 = \ln 6 + \frac{2}{13} + \frac{2}{3 \cdot 13^3} + \frac{2}{5 \cdot 13^5} + \cdots.$$

Aus (37) und (41) folgt zunächst

$$1{,}7917591 < \ln 6 < 1{,}7917596.$$

Ferner ist auf Grund von (38)

$$0 < \varrho_2(6) < \frac{1}{84} \cdot \frac{1}{5 \cdot 13^3} < 0{,}0000011.$$

---

[1]) Unser Ziel ist es, die genannten Logarithmen mit einer völlig garantierten und nicht nur mit einer sehr großen Genauigkeit zu berechnen. Gewöhnlich führt man derartige Berechnungen mit anderen ökonomischeren Verfahren. Dabei wird aber häufig die letzte Stelle etwas zweifelhaft, und zwar auch dann, wenn man mit einer ganzen Reihe von überzähligen Dezimalstellen rechnet. Das soll natürlich nicht bedeuten, daß man die üblichen Rechenmethoden zugunsten der hier angewendeten Methoden aufgeben sollte.

Schließlich ist

$$\frac{2}{13} = 0{,}153\,8462\,(-),$$

$$\frac{2}{3 \cdot 13^3} = 0{,}000\,3034\,(+),$$

so daß mit einem Fehler, dessen absoluter Betrag nicht größer als 0,00000005 ist, die Gleichung

$$\frac{2}{13} + \frac{2}{3 \cdot 13^3} = 0{,}154\,1496$$

gilt. Daher ist

$$1{,}945\,908\,65 < \ln 7 < 1{,}945\,910\,35.$$

Hieraus und aus (46) folgt aber unmittelbar

$$\frac{1{,}945\,908}{2{,}302\,586} < \lg 7 < \frac{1{,}945\,911}{2{,}302\,584}$$

und damit

$$0{,}845\,096 < \lg 7 < 0{,}845\,099. \tag{54}$$

Aus (47) bis (57) erhalten wir schließlich sofort die folgenden Werte für die dekadischen Logarithmen der ganzen Zahlen von 1 bis 10:

N	$\lg N$	N	$\lg N$
1	0,000 00	6	0,778 15
2	0,301 03	7	0,845 10
3	0,477 12	8	0,903 09
4	0,602 06	9	0,954 24
5	0,698 97	10	1,000 00

Alle in dieser Tafel aufgeführten Ziffern sind dabei richtig.

Wir weisen nochmals darauf hin (vgl. die Fußnote auf Seite 445), daß man in der Praxis ganz andere Methoden zur Bestimmung der Genauigkeit verwendet, als wir im vorangehenden angewendet haben. Die Grundprinzipien der eigentlichen Berechnung stimmen dagegen mit den hier dargelegten vollkommen überein.

42. Entwicklung des Arkustangens und Berechnung der Zahl π. Als geometrische Reihe mit dem Quotienten $q = -x^2$ ist die Reihe

$$1 - x^2 + x^4 - x^6 + \cdots$$

für $|x| < 1$ konvergent, und zwar mit der Summe $\frac{1}{1+x^2}$. Hieraus folgt, daß für jedes z mit $0 \leqq z < 1$ die Beziehung

$$\int_0^z \frac{dx}{1+x^2} = \int_0^z dx - \int_0^z x^2\,dx + \int_0^z x^4\,dx - \cdots$$

besteht, d. h.

$$\operatorname{arctg} z = z - \frac{z^3}{3} + \frac{z^5}{5} - \frac{z^7}{7} + \cdots .$$

Ersetzen wir hier z durch x, so erhalten wir

$$\operatorname{arctg} x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots . \tag{55}$$

Diese Gleichung gilt zunächst für $0 \leqq x < 1$, dann aber, da die Funktion arctg x und die durch die auf der rechten Seite auftretende Reihe dargestellte Funktion ungerade Funktionen sind, auch für $-1 < x < 1$.

Für $x = 1$ nimmt (55) die Form

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \cdots \tag{56}$$

an, wobei zwar, wie man unmittelbar mit Hilfe der Leibnizschen Regel erkennt, die Reihe auf der rechten Seite konvergiert, jedoch zunächst keineswegs sicher ist, ob diese Gleichung tatsächlich richtig ist. Wir wollen nun zeigen, daß dies wirklich der Fall ist. Zu diesem Zweck betrachten wir an Stelle der unendlichen Reihe $1 - x^2 + x^4 - x^6 + \cdots$ die endliche Summe

$$1 - x^2 + x^4 - x^6 + \cdots + x^{2n-2} - x^{2n} = \frac{1 - x^{2n+2}}{1 + x^2} . \tag{57}$$

Auf Grund dieser letzten Gleichung ist

$$\frac{1}{1+x^2} = (1 - x^2 + x^4 - x^6 + \cdots + x^{2n-2} - x^{2n}) + \frac{x^{2n+2}}{1+x^2}$$

und daher

$$\operatorname{arctg} 1 = \int_0^1 \frac{dx}{1+x^2} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots + \frac{1}{2n-1} - \frac{1}{2n+1} + \int_0^1 \frac{x^{2n+2}\,dx}{1+x^2} . \tag{58}$$

Man sieht nun unmittelbar, daß

$$0 < \int_0^1 \frac{x^{2n+2}\,dx}{1+x^2} < \int_0^1 x^{2n+2}\,dx = \frac{1}{2n+3} ,$$

so daß

$$\lim_{n \to \infty} \int_0^1 \frac{x^{2n+2}\,dx}{1+x^2} = 0 \tag{59}$$

und daher

$$\operatorname{arctg} 1 = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots ,$$

was zu beweisen war.

Die theoretisch recht interessante Formel (56), durch die π in sehr einfacher Weise durch natürliche Zahlen ausgedrückt wird (auf einfachere Weise jeden-

falls als durch die Wallissche Formel), ist leider für die praktische Berechnung von π völlig unbrauchbar, da die Reihe viel zu langsam konvergiert.[1])

Eine besser geeignete Reihe erhält man, wenn man in (55) $x = \frac{\sqrt{3}}{3}$ setzt. Dies führt auf

$$\frac{\pi}{6} = \frac{\sqrt{3}}{3}\left(1 - \frac{1}{9} + \frac{1}{45} - \frac{1}{189} + \frac{1}{729} - \frac{1}{2673} + \cdots\right)$$

und damit auf

$$\pi = 2\sqrt{3}\left(1 - \frac{1}{9} + \frac{1}{45} - \frac{1}{189} + \frac{1}{729} - \frac{1}{2673} + \cdots\right). \tag{60}$$

Beschränken wir uns auf die explizit angegebenen Glieder, so begehen wir einen Fehler, der auf Grund der Bemerkung zur Leibnizschen Regel in Abschnitt 36 positiv und kleiner als das erste fortgelassene Glied ist, d. h. also kleiner als

$$\frac{2\sqrt{3}}{9477} < 0{,}0005.$$

Dies zeigt, daß die Reihe (60) für eine näherungsweise Berechnung von π schon recht gut geeignet ist. Um jedoch π mit einer sehr großen Genauigkeit zu berechnen, ist auch die Reihe (60) noch zu unhandlich. Eine bessere Konvergenz kann man, gleichfalls von Formel (55) ausgehend, durch gewisse Kunstgriffe erreichen. Wir wollen einen derartigen Kunstgriff hier vorführen.

Wir setzen

$$\alpha = \operatorname{arctg}\frac{1}{5}$$

und

$$\beta = 4\alpha - \frac{\pi}{4},$$

so daß

$$\pi = 16\alpha - 4\beta. \tag{61}$$

Wir wollen nun zeigen, daß man die Größen 16α und 4β mittels Formel (55) berechnen kann. In der Tat: Setzen wir in (55) $x = \frac{1}{5}$, so erhalten wir nach Multiplikation mit 16:

$$16\alpha = \frac{16}{5} - \frac{16}{3 \cdot 5^3} + \frac{16}{5 \cdot 5^5} - \frac{16}{7 \cdot 5^7} + \frac{16}{9 \cdot 5^9} - \frac{16}{11 \cdot 5^{11}} + \frac{16}{13 \cdot 5^{13}} - \cdots.$$

Beschränken wir uns auf die hingeschriebenen Glieder, so begehen wir einen Fehler Δ', der auf Grund der Bemerkung zur Leibnizschen Regel den Abschätzungen

$$-\frac{1}{2 \cdot 10^{10}} < -\frac{16}{15 \cdot 5^{15}} < \Delta' < 0$$

[1]) Um π auch nur mit zwei richtigen Stellen hinter dem Komma zu berechnen, müßte man fünfzig (!) Glieder der Reihe (56) berücksichtigen.

genügt. Die Addition der positiven Glieder ergibt:

$$\frac{16}{5} = 3{,}200000000000$$

$$\frac{16}{5 \cdot 5^5} = 0{,}001024000000$$

$$\frac{16}{9 \cdot 5^9} = 0{,}000000910220\ (+)$$

$$\frac{16}{13 \cdot 5^{13}} = 0{,}000000001010\ (-)$$

$$3{,}201024911230$$

wobei ein Fehler begangen wird, der zwischen $-\frac{1}{2 \cdot 10^{11}}$ und $\frac{1}{2 \cdot 10^{11}}$ liegt. Entsprechend wird die Summe der negativen Glieder durch

$$\frac{16}{3 \cdot 5^3} = 0{,}042666666670\ (-)$$

$$\frac{16}{7 \cdot 5^7} = 0{,}000029257140\ (+)$$

$$\frac{16}{11 \cdot 5^{11}} = 0{,}000000029790\ (-)$$

$$0{,}042695953600$$

gegeben, wobei hier der Fehler zwischen -10^{-11} und $\frac{1}{2}10^{-11}$ liegt. Damit erhalten wir

$$3{,}158328957570 < 16\alpha < 3{,}158328957650. \tag{62}$$

Bevor wir jetzt 4β berechnen, bemerken wir, daß β zwischen $-\frac{\pi}{2}$ und $+\frac{\pi}{2}$ liegt. In der Tat: Wegen $\alpha = \operatorname{arctg}\frac{1}{5}$ ist $0 < \alpha < \frac{1}{5}$ und daher $0 < 4\alpha < \frac{4}{5}$, so daß

$$-\frac{\pi}{4} < 4\alpha - \frac{\pi}{4} < \frac{4}{5} - \frac{\pi}{4}.$$

Daher ist

$$\beta = \operatorname{arctg}(\operatorname{tg}\beta).^{1})$$

Nun ist aber

$$\operatorname{tg}\beta = \operatorname{tg}\left(4\alpha - \frac{\pi}{4}\right) = \frac{\operatorname{tg}4\alpha - 1}{1 + \operatorname{tg}4\alpha}. \tag{63}$$

Beachten wir noch, daß $\operatorname{tg}\alpha = \frac{1}{5}$, so finden wir

$$\operatorname{tg}2\alpha = \frac{2\operatorname{tg}\alpha}{1 - \operatorname{tg}^2\alpha} = \frac{5}{12}, \quad \operatorname{tg}4\alpha = \frac{2\operatorname{tg}2\alpha}{1 - \operatorname{tg}^2 2\alpha} = \frac{120}{119}.$$

[1]) Wir weisen darauf hin, daß wir mit $\operatorname{arctg} x$ denjenigen Winkel bezeichnen, dessen Tangens gleich x ist und der zwischen $-\frac{\pi}{2}$ und $+\frac{\pi}{2}$ liegt.

Hieraus und aus (63) folgt

$$\operatorname{tg}\beta = \frac{1}{239}.$$

Daher ist auf Grund von Formel (55)

$$4\beta = \frac{4}{239} - \frac{4}{3\cdot 239^3} + \frac{4}{5\cdot 239^5} - \cdots.$$

Beschränken wir uns hier auf zwei (!) Glieder, so begehen wir einen Fehler Δ'', der den Ungleichungen

$$0 < \Delta'' < \frac{4}{5\cdot 239^5} < \frac{1}{2\cdot 10^{11}}$$

genügt. Wegen

$$\frac{4}{239} = 0{,}01673640167\ (+),$$

$$\frac{4}{3\cdot 239^3} = 0{,}00000009767\ (-)$$

ist die Differenz dieser beiden Zahlen gleich

$$0{,}01673630400,$$

wobei der Fehler dieser Differenz zwischen 0 und 10^{-11} liegt. Hieraus und aus den Abschätzungen für Δ'' folgt

$$0{,}01673630400 < 4\beta < 0{,}01673630402. \tag{64}$$

Aus (61), (62) und (64) erhalten wir aber sofort

$$3{,}14159265355 < \pi < 3{,}14159265365,$$

das heißt

$$\pi = 3{,}1415926536,$$

wobei alle hingeschriebenen Ziffern richtig sind.

43. Allgemeine Bemerkungen über die Entwicklung einer Funktion in eine Potenzreihe. Es sei $f(x)$ eine in einem gegebenen Intervall $[A, B]$ definierte Funktion. Es sei ferner a ein beliebiger Punkt aus diesem Intervall. Wir wollen untersuchen, unter welchen Voraussetzungen man die Funktion $f(x)$ als Potenzreihe in $x-a$ darstellen kann. Der Einfachheit halber wollen wir dabei annehmen, daß a ein Punkt aus dem Innern des Intervalls $[A, B]$ ist, d. h., es sei $A < a < B$.[1])

Wir wollen zunächst einmal annehmen, daß wir von der Funktion $f(x)$ bereits wissen, daß sie in einem gewissen Intervall $a-r < x < a+r\ (r>0)$ um a als Potenzreihe in $x-a$ darstellbar ist, und zwar sei für alle x aus diesem Intervall

$$f(x) = c_0 + c_1(x-a) + c_2(x-a)^2 + c_3(x-a)^3 + \cdots. \tag{65}$$

[1]) Im Fall $a = A$ (oder $a = B$) wären keine prinzipiellen Änderungen in den folgenden Ausführungen notwendig; es müßten lediglich einige Formulierungen etwas anders gefaßt werden.

In diesem Fall muß das Intervall $(a-r, a+r)$ einerseits im Intervall $[A, B]$ und andererseits im Konvergenzintervall der angegebenen Reihe enthalten sein (so daß also insbesondere der Konvergenzradius R dieser Reihe nicht kleiner als r sein kann). Setzen wir in (65) $x=a$, so finden wir

$$c_0=f(a). \tag{66}$$

Da eine Potenzreihe in allen Punkten aus dem Innern ihres Konvergenzintervalls differenzierbar ist, muß unter der obigen Annahme in allen Punkten x aus $(a-r, a+r)$ die Ableitung $f'(x)$ existieren, wobei

$$f'(x)=c_1+2c_2(x-a)+3c_3(x-a)^2+\cdots. \tag{67}$$

Setzen wir hier $x=a$, so finden wir

$$c_1=f'(a).$$

Mit Hilfe derselben Überlegungen erhalten wir aus (67), daß in allen Punkten des Intervalls $(a-r, a+r)$ auch die zweite Ableitung $f''(x)$ existiert und der Gleichung

$$f''(x)=2c_2+3\cdot 2c_3(x-a)+4\cdot 3\cdot c_4(x-a)^2+\cdots$$

genügt, so daß also

$$c_2=\frac{f''(a)}{2!}.$$

Entsprechendes gilt für die Ableitungen höherer Ordnung. Damit erhalten wir den

Satz 1. *Läßt sich die Funktion $f(x)$ in einem gewissen Intervall $(a-r, a+r)$ in eine Potenzreihe nach $x-a$ entwickeln, so ist sie im Intervall $(a-r, a+r)$ beliebig oft differenzierbar; die Entwicklung hat die Form*

$$f(x)=f(a)+\frac{f'(a)}{1!}(x-a)+\frac{f''(a)}{2!}(x-a)^2+\frac{f'''(a)}{3!}(x-a)^3+\cdots. \tag{68}$$

Man bezeichnet die Reihe (68) üblicherweise als TAYLORsche *Reihe* der Funktion $f(x)$. Unter Verwendung dieser Bezeichnung können wir den angegebenen Satz auch kurz folgendermaßen formulieren: *Ist eine Funktion überhaupt in eine Potenzreihe entwickelbar, so ist dies notwendigerweise ihre* TAYLORsche *Reihe.*

Hieraus folgt sofort, daß nicht jede Funktion in eine Potenzreihe entwickelbar ist, da man natürlich nicht für jede Funktion die TAYLORsche Reihe bilden kann. Zum Beispiel existiert für die Funktion

$$f(x)=\sqrt[5]{x}$$

für $a=0$ keine TAYLORsche Reihe, da die Ableitung

$$f'(x)=\frac{1}{3\sqrt[5]{x^2}}$$

für $x=0$ ihren Sinn verliert.

Man kann ferner zeigen, daß es Funktionen gibt, für die man zwar die TAYLORsche Reihe bilden kann, wobei aber der Konvergenzradius dieser Reihe gleich Null ist, so daß die Gleichung (68) nur für $x = a$ richtig ist. In diesem Fall ist die erhaltene Potenzreihe natürlich praktisch wertlos.

Schließlich kann aber auch noch der Fall eintreten, daß die gegebene Funktion Ableitungen beliebiger Ordnungen besitzt, daß die zugehörige TAYLORsche Reihe einen positiven Konvergenzradius besitzt und daß trotzdem die Gleichung (68) nur im Punkte $x = a$ gilt, daß also die Summe der TAYLORschen Reihe nur im Punkte $x = a$ mit $f(x)$ übereinstimmt.

Hieraus folgt, daß sogar die beliebig oft differenzierbaren Funktionen noch gewisse zusätzliche Bedingungen erfüllen müssen, damit sie in eine Potenzreihe entwickelbar sind. Wir beschränken uns hier auf die Angabe eines Satzes, den wir im Prinzip bereits in Abschnitt 12 bewiesen haben:

Satz 2. *Es sei $f(x)$ eine im Intervall $[A, B]$ definierte Funktion, welche dort Ableitungen beliebiger Ordnung besitzt. Gibt es dabei eine Zahl K, so daß für alle n und für alle $x \in [A, B]$ die Ungleichung*

$$|f^{(n)}(x)| \leqq K$$

gilt, so gilt (68) *für alle x und a aus dem Intervall $[A, B]$.*

Unter den angegebenen Voraussetzungen strebt nämlich, wie in Abschnitt 12 gezeigt wurde, das Restglied $R_n(x)$ der TAYLORschen Formel

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

mit wachsendem n gegen Null; weil aber hierbei

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

eine Partialsumme der TAYLORschen Reihe der Funktion $f(x)$ ist, ist dies vollkommen gleichwertig der Tatsache, daß die Beziehung (68) besteht.

Insbesondere folgt daraus[1]), daß die Funktionen $\sin x$, $\cos x$ bzw. e^x in folgender Weise als Potenzreihen darstellbar sind:

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots, \tag{69}$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots, \tag{70}$$

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots. \tag{71}$$

Die Formeln (69) und (70) kann man unmittelbar zur numerischen Berechnung der Werte der Funktionen $\sin x$ und $\cos x$ verwenden. Wir wollen hier jedoch darauf nicht näher eingehen, da die Grundprinzipien dabei dieselben wie bei der Berechnung der Logarithmen und des Arkustangens sind.

[1]) Wir haben hierauf bereits kurz in Abschnitt 39 hingewiesen.

Es sei jedoch ausdrücklich darauf hingewiesen, daß man zur Berechnung des Sinus (oder Kosinus) eines Winkels nach Formel (69) (bzw. (70)) für x die Größe des Winkels im Bogenmaß (und nicht im Gradmaß) nehmen muß.

Wir wollen zum Abschluß jedoch noch eine wichtige Anwendung von Formel (65) erwähnen. Gilt diese nämlich in einem Intervall $(a-r,\ a+r)$, so ist, wie wir in Abschnitt 40 gezeigt haben, für jedes in $(a-r,\ a+r)$ enthaltene abgeschlossene Intervall $[p,\ q]$

$$\int_p^q f(x)\,dx = c_0\int_p^q dx + c_1\int_p^q (x-a)\,dx + c_2\int_p^q (x-a)^2\,dx + \cdots. \tag{72}$$

Da sich hierbei alle Integrale auf der rechten Seite ohne Mühe berechnen lassen, erhalten wir damit eine Möglichkeit, das Integral

$$\int_p^q f(x)\,dx$$

näherungsweise zu berechnen. Bei dieser Methode ist es offenbar nicht notwendig, eine Stammfunktion zu $f(x)$ zu bestimmen, so daß diese Methode auch dann anwendbar ist, wenn sich die zugehörige Stammfunktion nicht elementar ausdrücken läßt.

Wir wollen als Beispiel das Integral

$$\int_0^1 \frac{\sin x}{x}\,dx$$

mit einer Genauigkeit von 0,0001 berechnen.[1])

Auf Grund von Formel (69) ist zunächst

$$\int_0^1 \frac{\sin x}{x}\,dx = \int_0^1 \left(1 - \frac{x^2}{3!} + \frac{x^4}{5!} - \frac{x^6}{7!} + \cdots\right) dx,$$

so daß

$$\int_0^1 \frac{\sin x}{x}\,dx = 1 - \frac{1}{3\cdot 3!} + \frac{1}{5\cdot 5!} - \cdots.$$

Beschränken wir uns hier auf die hingeschriebenen Glieder, so begehen wir auf Grund der Bemerkung zur Leibnizschen Regel einen Fehler Δ, der den Abschätzungen

$$0 > \Delta > -\frac{1}{7\cdot 7!} > -0{,}00003$$

[1]) Man kann zeigen, daß $\int \frac{\sin x}{x}\,dx$ nicht durch elementare Funktionen ausdrückbar ist.

genügt. Nun ist aber

$$1+\frac{1}{5\cdot 5!}=1{,}001\,67\,(-),$$

$$\frac{1}{3\cdot 3!}=0{,}055\,56\,(-).$$

Daher ist unter Berücksichtigung aller Fehler

$$\int_0^1 \frac{\sin x}{x}\,dx=0{,}9461,$$

wobei alle hingeschriebenen Ziffern richtig sind.

Dieses Verfahren für die angenäherte Berechnung eines bestimmten Integrals ist in den meisten Fällen wesentlich bequemer als das in Abschnitt 27 dargelegte.

44. Die Binomialreihe. Im vorliegenden Abschnitt wollen wir die sinngemäße Übertragung der bereits aus der Schule bekannten „binomischen Formel" auf den Fall beliebiger Exponenten behandeln, die bereits auf NEWTON zurückgeht.

Dazu stellen wir uns die Aufgabe, die Funktion

$$f(x)=(1+x)^{\mu}$$

nach Potenzen von x zu entwickeln. Hier ist

$$f'(x)=\mu(1+x)^{\mu-1},\quad f''(x)=\mu(\mu-1)(1+x)^{\mu-2},$$

$$f'''(x)=\mu(\mu-1)(\mu-2)(1+x)^{\mu-3}$$

und allgemein, wie man leicht durch vollständige Induktion zeigt,

$$f^{(n)}(x)=\mu(\mu-1)\cdots(\mu-n+1)(1+x)^{\mu-n},$$

so daß also

$$f(0)=1,\ f'(0)=\mu,\ f''(0)=\mu(\mu-1),\ \ldots,\ f^{(n)}(0)=\mu(\mu-1)\cdots(\mu-n+1),\ldots$$

Es ist daher

$$1+\mu x+\frac{\mu(\mu-1)}{2!}x^2+\frac{\mu(\mu-1)(\mu-2)}{3!}x^3+\cdots \tag{73}$$

die TAYLORsche Reihe der betrachteten Funktion. Man bezeichnet die Reihe (73) üblicherweise als *Binomialreihe* (oder *binomische Reihe*). Für $\mu=0, 1, 2$ nimmt (73) bzw. die Form

$$1,\quad 1+x,\quad 1+2x+x^2$$

an, d. h., in diesen Fällen wird aus der im allgemeinen unendlichen Reihe (73) eine endliche Summe. Man sieht nun sofort, daß dies überhaupt immer dann eintritt, wenn μ eine nicht negative ganze Zahl ist. In diesem Fall liefert die Reihe (73) (bei der also alle Glieder bis auf endlich viele gleich Null sind) als

Summe den Funktionswert $(1+x)^\mu$, d. h., es gilt für jedes reelle (und sogar komplexe) x die Gleichung

$$(1+x)^\mu = 1 + \mu x + \frac{\mu(\mu-1)}{2!}x^2 + \cdots + \frac{\mu(\mu-1)\cdots 1}{\mu!}x^\mu,$$

und das ist bekanntlich auch gerade die Aussage des üblichen „binomischen Satzes".

Ist dagegen μ keine nicht negative ganze Zahl, so erhalten wir in (73) mit Sicherheit eine unendliche Reihe, und es ergibt sich als erstes die Frage, für welche Werte x diese Reihe konvergiert.

Zur Beantwortung dieser Frage verwenden wir das D'ALEMBERTsche Quotientenkriterium. Im betrachteten Fall ist offenbar

$$a_n = \frac{\mu(\mu-1)\cdots(\mu-n+1)}{n!}x^n,$$

$$a_{n+1} = \frac{\mu(\mu-1)\cdots(\mu-n+1)(\mu-n)}{(n+1)!}x^{n+1}$$

und daher

$$\frac{a_{n+1}}{a_n} = \frac{\mu-n}{n+1}x,$$

so daß

$$\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = |x|.$$

Daraus folgt, daß die Reihe (73) für $|x|<1$ konvergiert und für $|x|>1$ divergiert, so daß der Konvergenzradius der Reihe (73) gleich Eins ist. Auf die Frage nach dem Verhalten der Reihe in den Endpunkten ± 1 des Konvergenzintervalls wollen wir hier nicht eingehen,[1]) sondern uns auf das Studium der Reihe im offenen Intervall $(-1, +1)$ beschränken.

Wir haben bereits oben darauf hingewiesen, daß es Fälle gibt, in denen die TAYLORsche Reihe zwar konvergiert, aber die Summe der Reihe nicht gleich dem entsprechenden Funktionswert ist. Daher ist es keineswegs sicher, ob die Reihe (73) für $-1<x<+1$ als Summe wirklich den Wert $(1+x)^\mu$ liefert. Dies kann auch nicht mit Hilfe von Satz 2 aus Abschnitt 43 erschlossen werden, weil die Voraussetzungen dieses Satzes im vorliegenden Fall nicht erfüllt sind. Man könnte nun die Darstellung von $(1+x)^\mu$ durch die binomische Reihe dadurch beweisen, daß man einen anderen Ausdruck für das Restglied der TAYLORschen Formel verwendet. Da wir jedoch über diese Form des Restgliedes hier nicht verfügen und der Beweis dafür wesentlich komplizierter ist als die Ableitung in Abschnitt 12, können wir auch diesen Weg nicht einschlagen. Nun ist aber die erwähnte Tatsache für das Folgende von großer Bedeutung. Daher müssen wir uns hier nach einem anderen Weg umsehen. Wir werden nachfolgend einen Beweis geben, der von einer Reihe von Kunstgriffen Gebrauch macht.

[1]) Die Konvergenz oder Divergenz der Reihe (73) in den Punkten $x = \pm 1$ hängt von der Wahl des Exponenten μ ab.

Mit $S(x)$ bezeichnen wir die Summe der Potenzreihe (73). Da man eine Potenzreihe gliedweise differenzieren kann, gilt für alle x aus $(-1, +1)$

$$S'(x) = \mu + \frac{\mu(\mu-1)}{1!}x + \frac{\mu(\mu-1)(\mu-2)}{2!}x^2 + \cdots$$

und damit

$$\frac{S'(x)}{\mu} = 1 + \frac{\mu-1}{1!}x + \frac{(\mu-1)(\mu-2)}{2!}x^2 + \cdots. \tag{74}$$

Wir multiplizieren nun die Gleichung (74) mit x und addieren sie gliedweise zu (74),

$$\frac{1+x}{\mu}S'(x) = \begin{cases} 1 + \dfrac{\mu-1}{1!}x + \dfrac{(\mu-1)(\mu-2)}{2!}x^2 + \dfrac{(\mu-1)(\mu-2)(\mu-3)}{3!}x^3 + \cdots \\ \quad + \qquad x + \qquad \dfrac{(\mu-1)}{1!}x^2 + \qquad \dfrac{(\mu-1)(\mu-2)}{2!}x^3 + \cdots, \end{cases}$$

das heißt

$$\frac{1+x}{\mu}S'(x) = 1 + \mu x + \frac{\mu(\mu-1)}{2!}x^2 + \frac{\mu(\mu-1)(\mu-2)}{3!}x^3 + \cdots.$$

Der Ausdruck auf der rechten Seite der zuletzt erhaltenen Gleichung ist nun aber gerade wieder $S(x)$, so daß

$$(1+x)\,S'(x) = \mu S(x). \tag{75}$$

Wir betrachten nun die Funktion

$$\psi(x) = \frac{S(x)}{(1+x)^\mu}.$$

Sie ist offenbar in allen Punkten des Intervalls $(-1, +1)$ differenzierbar, wobei

$$\psi'(x) = \frac{S'(x)(1+x)^\mu - S(x)\,\mu(1+x)^{\mu-1}}{(1+x)^{2\mu}} = \frac{(1+x)S'(x) - \mu S(x)}{(1+x)^{\mu+1}}.$$

Wegen (75) ist also

$$\psi'(x) = 0.$$

Daraus folgt, daß die Funktion $\psi(x)$ im Intervall $(-1, +1)$ konstant ist. Da nun $\psi(0) = 1$, ist im Intervall $(-1, +1)$ überall $\psi(x) = 1$, und dies bedeutet gerade, daß

$$S(x) = (1+x)^\mu.$$

Damit ist gezeigt:

Satz 1. *Für alle x mit $-1 < x < 1$ ist bei beliebigem μ*

$$(1+x)^\mu = 1 + \mu x + \frac{\mu(\mu-1)}{2!}x^2 + \frac{\mu(\mu-1)(\mu-2)}{3!}x^3 + \cdots. \tag{76}$$

Für $\mu = -1$ erhalten wir aus Satz 1 nochmals das bereits bekannte Ergebnis über die Summe der geometrischen Reihe:

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \cdots.$$

Wir wollen hier noch zwei weitere Spezialfälle von Formel (76) besonders erwähnen:

Für $\mu = -\frac{1}{2}$ wird

$$\mu(\mu-1)(\mu-2)\cdots(\mu-n+1) = (-1)^n \frac{1\cdot 3\cdot 5\cdots(2n-1)}{2^n}$$

$$= (-1)^n \frac{(2n-1)!!}{2^n},^{1})$$

so daß wegen

$$2^n \cdot n! = (2n)!!$$

Formel (76) hier folgendes ergibt:

$$\frac{1}{\sqrt{1+x}} = 1 - \frac{1!!}{2!!}x + \frac{3!!}{4!!}x^2 - \frac{5!!}{6!!}x^3 + \cdots. \tag{77}$$

Analog ergibt sich

$$\sqrt{1+x} = 1 + \frac{1}{2}x - \frac{1!!}{4!!}x^2 + \frac{3!!}{6!!}x^3 - \frac{5!!}{8!!}x^4 + \cdots.^{2}) \tag{78}$$

Wir bringen zum Abschluß noch einige praktische Anwendungen von Formel (76).

I. Die angenäherte Berechnung von Wurzeln. Wir stellen uns die Aufgabe, aus einer gegebenen natürlichen Zahl A die m-te Wurzel zu ziehen, wobei wir annehmen können, daß A keine m-te Potenz ist.

Dazu denken wir uns die Zahl A in der Form

$$A = a^m + b$$

dargestellt, wobei a eine natürliche und b eine ganze Zahl ist, und zwar so, daß $|b| < a^m$.[3]) Dann ist offenbar

$$\sqrt[m]{A} = a\sqrt[m]{1+\frac{b}{a^m}}.$$

Setzen wir hier $\frac{b}{a^m} = x$, so liegt x im Intervall $(-1, +1)$, und wir können daher die Funktion

$$\sqrt[m]{1+\frac{b}{a^m}} = (1+x)^{\frac{1}{m}}$$

---

[1]) Bezüglich der Definition von $k!!$ vgl. S. 381. — *Anm. d. wissenschaftl. Red.*

[2]) Der Koeffizient von x^n in (78) ist für $n \geqq 2$ gleich

$$(-1)^{n-1}\frac{(2n-3)!!}{(2n)!!}.$$

[3]) Eine derartige Darstellung ist auf unendlich viele Arten möglich. Es genügt zu diesem Zweck, eine natürliche Zahl a zu nehmen, für welche $a^m > A$ ist, und dann $b = A - a^m$ zu setzen.

als Potenzreihe in x darstellen. Diese Reihe wird um so besser konvergieren, je kleiner $|x|$ ist. Indem man von der erhaltenen Reihe hinreichend viele Glieder nimmt, kann man die gesuchte Wurzel mit beliebiger Genauigkeit berechnen.

Die Fehlerabschätzung wird stets dann besonders einfach, wenn $x > 0$ ist. In diesem Fall ist nämlich für $\mu = \frac{1}{m}$ mit $m > 1$ die Reihe

$$\mu x + \frac{\mu(\mu-1)}{2!}x^2 + \frac{\mu(\mu-1)(\mu-2)}{3!}x^3 + \cdots$$

offenbar alternierend, wobei das allgemeine Glied mit wachsendem n dem Betrage nach monoton fällt.[1]) Daher ist der Fehler dem Betrage nach kleiner als das erste vernachlässigte Glied, wobei das Vorzeichen des Fehlers gleich dem des ersten vernachlässigten Gliedes ist.

So ist z. B. der Fehler der Näherungsgleichung

$$\sqrt[5]{38} = 2\sqrt[5]{1+\frac{3}{16}} \approx 2\left(1+\frac{3}{80}-\frac{9}{3200}+\frac{81}{256000}\right)$$

negativ und dem Betrage nach kleiner als

$$\frac{1701}{20480000} < 0{,}0001\,.$$

Im Fall $x < 0$ ist die Fehlerabschätzung nicht ganz so einfach. Wir wollen uns hier auf eine Fehlerabschätzung der Näherungsgleichung

$$\sqrt{1+x} \approx 1 + \frac{x}{2}$$

beschränken, welche man aus (78) erhält, wenn man dort nur die ersten beiden Glieder berücksichtigt.[2])

Zu diesem Zweck[3]) betrachten wir die Funktion

$$\varphi(x) = \frac{1+\frac{x}{2}-\sqrt{1+x}}{x^2} \qquad (x \geqq -1,\quad x \neq 0).$$

Indem man hier Zähler und Nenner mit $1+\frac{x}{2}+\sqrt{1+x}$ multipliziert, erkennt man leicht, daß

$$\lim_{x\to 0}\varphi(x) = \frac{1}{8}\,.$$

---

[1]) Der Betrag des Quotienten zweier aufeinanderfolgender Glieder ist nämlich offenbar gleich $\left|\frac{\mu-n+1}{n+1}x\right|$, und diese Zahl ist $\left(\text{wegen } \mu = \frac{1}{m} \text{ und } |x|<1\right)$ kleiner als Eins.

[2]) Die folgende Abschätzung dieses Fehlers werden wir weiter unten noch benötigen.

[3]) Wir folgen hier einem brieflichen Vorschlag von Herrn Prof. Dr. D. K. Kazarinoff (Michigan, USA). In der russischen Ausgabe war durch umfangreichere Überlegungen eine schwächere Abschätzung hergeleitet worden. — *Anm. d. Autors.*

Setzen wir also noch $\varphi(0) = \frac{1}{8}$, so erhalten wir in $\varphi(x)$ eine Funktion, welche in Intervall $[-1, +\infty)$ stetig ist. Wir zeigen, daß $\varphi(x)$ in diesem Intervall monoton fällt. In der Tat: Zunächst ist für $x \neq 0$, $x > -1$ offenbar

$$\varphi'(x) = \frac{4x + 3x^2 - (x^2 + 4x)\sqrt{1+x}}{2x^4\sqrt{1+x}}.$$

Da der Nenner hier stets positiv ist, stimmt das Vorzeichen dieses Bruches mit dem seines Zählers überein. Nun ist aber

$$(x^2 + 4x)\sqrt{1+x} \geqq 4x + 3x^2;$$

ist nämlich $x > 0$, so ist diese Ungleichung gleichbedeutend mit der Ungleichung

$$(x+4)\sqrt{1+x} \geqq 4 + 3x,$$

deren Gültigkeit man sofort durch Quadrieren nachprüft; ist dagegen $x < 0$, so reduziert sich die zu beweisende Ungleichung auf die Ungleichung

$$(x+4)\sqrt{1+x} \leqq 4 + 3x,$$

deren Gültigkeit ebenfalls durch Quadrieren nachgeprüft werden kann (dies ist erlaubt, da für $x > -1$ beide Seiten positiv sind). Daher ist für $x > -1$, $x \neq 0$

$$\varphi'(x) \leqq 0,$$

so daß die Funktion $\varphi(x)$ in $[-1, +\infty)$ monoton fällt. Nun ist aber $\varphi(-1) = \frac{1}{2}$. Daher ist $1 + \frac{x}{2} - \sqrt{1+x} \leqq \frac{x^2}{2}$. Ferner ist für $x \geqq -1$ offenbar $\sqrt{1+x} \leqq 1 + \frac{x}{2}$. Also gilt:

Satz 2. *Für* $x \geqq -1$ *ist* $0 \leqq 1 + \frac{x}{2} - \sqrt{1+x} \leqq \frac{x^2}{2}$.

Wir wollen hier noch kurz ein interessantes Anwendungsbeispiel dieser Fehlerabschätzung bringen.

II. Die Tschebyschewsche Regel zur näherungsweisen Rektifikation eines Kreisbogens. Wir betrachten (Abb. 46) den Bogen $\overset{\frown}{ABC}$ eines Kreises vom Radius R, der von einem Zentriwinkel der Größe $2x$ erzeugt werden möge, wobei wir annehmen wollen, daß $2x$ kleiner als 180° ist. Bekanntlich ist dann die Länge s dieses Bogens gleich $2Rx$. Wir wollen nun eine von P. L. Tschebyschew stammende Näherungsmethode zur Konstruktion dieser Länge mit Zirkel und Lineal kennen lernen.

Wir fällen vom Mittelpunkt O des Kreises das Lot auf die Sehne AC und bezeichnen mit B den Schnittpunkt des Lotes mit dem Kreisbogen. Die Länge der Strecke BD bezeichnet man auch als Höhe des Kreisbogens. Es ist nun klar, daß die Summe $AB + BC$ der Längen der Schenkel des gleichschenkligen Dreiecks ABC (welches die Sehne AC als Grundlinie und die Strecke BD als Höhe besitzt) kleiner als s ist. Unsere Aufgabe soll es nun sein, auf der Geraden BD, ausgehend von den gegebenen Bestimmungsstücken, mit Hilfe von Zirkel und Lineal einen Punkt E zu konstruieren, so daß die Summe $AE + EC$ der Längen der Schenkel des gleichschenkligen Dreiecks AEC möglichst genau die Länge des Bogens $\overset{\frown}{ABC}$ wiedergibt (den Sinn dieser Forderung werden wir weiter unten noch genauer präzisieren).

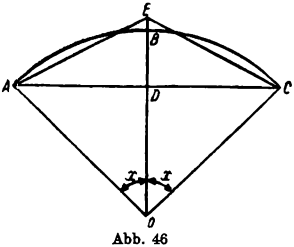


Abb. 46

Zu diesem Zweck verstehen wir unter h den Quotienten von ED und BD:

$$ED = h \cdot BD.$$

Wir setzen ferner $AE + EC = s^*$. Dann ist wegen

$$AD = R\sin x, \quad BD = R(1-\cos x)$$

offenbar

$$s^* = 2R\sqrt{\sin^2 x + h^2(1-\cos x)^2}.$$

Auf Grund von (69) und (70) ist nun

$$\sin x \approx x - \frac{x^3}{6}, \quad \cos x \approx 1 - \frac{x^2}{2},$$

wobei der Fehler dem Betrage nach kleiner als $\frac{x^5}{120}$ bzw. $\frac{x^4}{24}$ ist. Es ist also

$$\sin x = x - \frac{x^3}{6} + \vartheta_1 \frac{x^5}{120}, \quad \cos x = 1 - \frac{x^2}{2} + \vartheta_2 \frac{x^4}{24},$$

wobei die Zahlen ϑ_1 und ϑ_2 dem Betrage nach kleiner als Eins sind. Hieraus erhalten wir unmittelbar

$$\sin^2 x = x^2 - \frac{x^4}{3} + x^6\left(\frac{1}{36} + \frac{\vartheta_1}{60} + \frac{\vartheta_1^2 x^4}{14400} - \frac{\vartheta_1 x^2}{360}\right),$$

$$h^2(1-\cos x)^2 = h^2\frac{x^4}{4} + h^2 x^6\left(\frac{\left(\frac{\vartheta_2}{2}\right)^2 x^2}{576} - \frac{\vartheta_2}{24}\right).$$

Hierbei ist nun wegen $0 < x < \frac{\pi}{2} < 1{,}6$ sicher $x^2 < 3$, so daß

$$\left|\frac{1}{36} + \frac{\vartheta_1}{60} + \frac{\vartheta_1^2 x^4}{14400} - \frac{\vartheta_1 x^2}{360}\right| < \frac{1}{36} + \frac{1}{60} + \frac{1}{1600} + \frac{1}{120} < 0{,}055,$$

$$\left|\frac{\vartheta_2^2 x^2}{576} - \frac{\vartheta_2}{24}\right| < \frac{1}{192} + \frac{1}{24} < 0{,}048,$$

das heißt

$$s^* = 2R\sqrt{x^2 + \left(\frac{h^2}{4} - \frac{1}{3}\right)x^4 + \alpha x^6} \qquad (|\alpha| < 0{,}055 + 0{,}048\,h^2)$$

beziehungsweise

$$s^* = 2Rx\sqrt{1 + \left(\frac{h^2}{4} - \frac{1}{3}\right)x^2 + \alpha x^4}.$$

Da nun die tatsächliche Länge s des Bogens $\overset{\frown}{ABC}$ gleich $2Rx$ ist, nimmt Tschebyschew an, daß der bestmögliche Näherungswert derjenige ist, bei dem das im Radikanden auftretende Glied in x^2 verschwindet[1]), was auf

$$h = \frac{2}{\sqrt{3}}$$

führt.

---

[1]) Das ist eine spezielle Forderung, durch welche die oben angegebene verschwommene Forderung, daß die Summe $AE + EC$ die Länge des Bogens $\overset{\frown}{ABC}$ „möglichst genau" wiedergeben soll, einen vollkommen präzisen Sinn bekommt.

Man sieht sofort, daß der zu diesem Wert h gehörende Punkt E wirklich mit Zirkel und Lineal konstruierbar ist. Der zu diesem Wert h gehörende Näherungswert für s wird durch

$$s^* = 2Rx\sqrt{1+\alpha x^4}$$

gegeben, wobei

$$|\alpha| < 0{,}055 + 0{,}064 = 0{,}119 < 0{,}12.$$

Wegen

$$x^4 < (1{,}6)^4 < 7$$

ist nun $|\alpha x^4| < 1$. Daher können wir Satz 2 anwenden und erhalten

$$\sqrt{1+\alpha x^4} \approx 1 + \frac{\alpha x^4}{2},$$

wobei der Fehler dem Betrage nach kleiner als $\frac{\alpha^2 x^8}{2}$, d. h. sicher kleiner als $0{,}01\,x^8$ ist. Es ist also

$$\sqrt{1+\alpha x^4} = 1 + \frac{\alpha x^4}{2} + \beta x^8 \qquad (|\beta| < 0{,}01)$$

und daher

$$s^* = 2Rx + \alpha R x^5 + 2\beta R x^9.$$

Beachten wir schließlich noch, daß

$$|\alpha x^5 + 2\beta x^9| < x^5(0{,}12 + 0{,}02x^4) < 0{,}3x^5,$$

so finden wir, daß

$$s^* = s + \varrho,$$

wobei $|\varrho| < 0{,}3 \cdot Rx^5$.

Damit ist gezeigt:

Satz 3. *Die Länge eines Kreisbogens ist angenähert gleich der Summe der Längen der Schenkel des gleichschenkligen Dreiecks, welches die zum Bogen gehörende Sehne als Grundlinie hat und dessen Höhe gleich dem $\frac{2}{\sqrt{3}}$-fachen der Höhe des Kreisbogens ist. Der dabei begangene Fehler ist dem Betrag nach kleiner als*

$$0{,}3\,Rx^5,$$

wobei x der halbe Zentriwinkel des Bogens und R der Radius des betrachteten Kreises ist.

Es ist klar, daß die bei Anwendung der Tschebyschewschen Regel erzielte Genauigkeit um so besser ist, je kleiner x ist. Betrachten wir z. B. einen Zentriwinkel von 30°, so ist

$$x = \frac{\pi}{12}$$

und der bei Anwendung der Tschebyschewschen Regel begangene Fehler kleiner als $0{,}0006 \cdot R$.

Im betrachteten Fall wird bei Verwendung des in Abschnitt 42 erhaltenen Näherungswertes für π

$$s = 0{,}523\,599\,R,$$

während die Tschebyschewsche Regel auf

$$s \approx 2R\sqrt{\sin^2 15° + \frac{4}{3}(1-\cos 15°)^2},$$

also

$$s \approx 0{,}523\,585\,R$$

führt. Die erhaltenen Werte zeigen ein Übereinstimmen bis auf einen Fehler von $0{,}000014 \cdot R$, also einen Fehler, der unterhalb der angegebenen Fehlerschranke liegt.

III. Die Entwicklung des Arkussinus. Wir wollen schließlich noch die Funktion $\arcsin x$ als Potenzreihe in x darstellen. Die direkte Darstellung als Taylorsche Reihe ist hier zu schwierig, da die Ausdrücke für die höheren Ableitungen der Funktion $\arcsin x$ außerordentlich unhandlich sind. Diese Schwierigkeiten lassen sich jedoch durch Verwendung der Binomialreihe umgehen. Es ist nämlich, wie wir bereits gesehen haben,

$$\frac{1}{\sqrt{1+x}} = 1 + \sum_{n=1}^{\infty} (-1)^n \frac{(2n-1)!!}{(2n)!!} x^n.$$

Ersetzen wir hier x durch $-z$ und sodann z durch x^2, so finden wir

$$\frac{1}{\sqrt{1-x^2}} = 1 + \sum_{n=1}^{\infty} \frac{(2n-1)!!}{(2n)!!} x^{2n}.$$

Integrieren wir schließlich diese Gleichung über ein beliebiges Intervall $[0, x]$ mit $-1 < x < 1$, so erhalten wir

$$\arcsin x = x + \frac{1!!}{2!!}\frac{x^3}{3} + \frac{3!!}{4!!}\frac{x^5}{5} + \frac{5!!}{6!!}\frac{x^7}{7} + \cdots.$$

45. Ein kurzer Abriß der analytischen Theorie der trigonometrischen Funktionen. Für viele Untersuchungen der Differential- und Integralrechnung sind die trigonometrischen Funktionen von großer Bedeutung.[1]) Nun pflegt man in der Elementarmathematik die trigonometrischen Funktionen auf Grund von geometrischen Konstruktionen zu definieren. In diesem Zusammenhang ergibt sich die Frage, ob man nicht die Theorie dieser Funktionen auch ohne Rückgriff auf geometrische Überlegungen rein analytisch begründen kann. Dieses Problem ist noch insofern von grundsätzlicher Bedeutung, als neben der euklidischen Geometrie auch andere Geometrien existieren, so daß sich damit die Frage ergibt, ob nicht vielleicht gewisse Resultate der Analysis in irgendeiner Weise von der zugrundegelegten Geometrie abhängen oder, anders ausgedrückt, ob nicht bei Voraussetzung einer nichteuklidischen Geometrie eine Reihe von Sätzen geändert werden müßte.[2])

Es zeigt sich nun, daß das nicht der Fall ist, d. h., daß die Resultate der Analysis nicht von der Auswahl irgendeiner Geometrie abhängen. Wir werden nämlich zeigen, daß man die trigonometrischen Funktionen unabhängig von

---

[1]) Wir erinnern in diesem Zusammenhang nur an die trigonometrischen Substitutionen in der Integralrechnung (vgl. z. B. die Auswertung von $\int \sqrt{1-x^2}\,dx$ in Abschnitt 18).

[2]) In diesem Zusammenhang ist noch besonders zu erwähnen, daß es in der sogenannten Lobatschewskischen Geometrie keine ähnlichen Figuren gibt. Aber gerade auf den Ähnlichkeitssätzen beruhen viele Konstruktionen der Trigonometrie.

irgendwelchen geometrischen Überlegungen rein analytisch definieren und studieren kann, wobei wir hier allerdings nur auf die einfachsten Dinge eingehen wollen.

Wir betrachten zunächst die beiden Potenzreihen

$$1-\frac{x^2}{2!}+\frac{x^4}{4!}-\frac{x^6}{6!}+\frac{x^8}{8!}-\cdots, \tag{80}$$

$$\frac{x}{1!}-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\frac{x^9}{9!}-\cdots. \tag{81}$$

Mit Hilfe des D'ALEMBERTschen Quotientenkriteriums ergibt sich sofort, daß beide Reihen für beliebiges reelles x konvergieren. Die durch diese Reihen dargestellten Funktionen wollen wir den *Kosinus* bzw. *Sinus* nennen und mit $C(x)$ bzw. $S(x)$ bezeichnen[1]), so daß also

$$C(x)=1-\frac{x^2}{2!}+\frac{x^4}{4!}-\cdots, \quad S(x)=\frac{x}{1!}-\frac{x^3}{3!}+\frac{x^5}{5!}-\cdots.$$

Dieser Definition ist unmittelbar zu entnehmen, daß

$$C(-x)=C(x), \quad S(-x)=-S(x), \tag{82}$$

d. h., *der Kosinus ist eine gerade und der Sinus ist eine ungerade Funktion.* Ferner ergibt sich aus dem Satz über die gliedweise Differentiation einer Potenzreihe, daß

$$C'(x)=-S(x), \quad S'(x)=C(x) \tag{83}$$

und damit

$$\begin{aligned} C''(x)&=-C(x), & S''(x)&=-S(x),\\ C'''(x)&=S(x), & S'''(x)&=-C(x),\\ C^{(4)}(x)&=C(x), & S^{(4)}(x)&=S(x) \end{aligned}$$

und allgemein für beliebiges k

$$\left.\begin{aligned} C^{(4k)}(x)&=C(x), & S^{(4k)}(x)&=S(x),\\ C^{(4k+1)}(x)&=-S(x), & S^{(4k+1)}(x)&=C(x),\\ C^{(4k+2)}(x)&=-C(x), & S^{(4k+2)}(x)&=-S(x),\\ C^{(4k+3)}(x)&=S(x), & S^{(4k+3)}(x)&=-C(x). \end{aligned}\right\} \tag{84}$$

Als Grundlage für die weitere Theorie beweisen wir jetzt das Additionstheorem für den Kosinus.

Dazu bemerken wir zunächst, daß es zu jedem endlichen Intervall $[A, B]$

---

[1]) Wir wissen bereits, daß dies nichts anderes als die Funktionen $\cos x$ und $\sin x$ sind. Um jedoch zu verhindern, daß irgendwelche analytisch noch nicht begründete Eigenschaften dieser Funktionen rein gewohnheitsmäßig ungerechtfertigt benutzt werden, wollen wir hier $C(x)$ und $S(x)$ schreiben.

eine Konstante K gibt, so daß für alle n bei beliebigem $x \in [A, B]$ die Ungleichung

$$|C^{(n)}(x)| \leqq K$$

gilt.

In der Tat: Auf Grund von (84) ist für jedes n, je nachdem, ob n gerade oder ungerade ist, entweder $|C^{(n)}(x)| = |C(x)|$ oder $|C^{(n)}(x)| = |S(x)|$; jede der (nach Satz 1 aus Abschnitt 40) stetigen Funktionen $C(x)$ und $S(x)$ ist aber im Intervall $[A, B]$ beschränkt.

Daraus folgt auf Grund von Satz 2 aus Abschnitt 43, daß für beliebiges a und x aus irgendeinem Intervall $[A, B]$ die Beziehung

$$C(x) = C(a) + \frac{C'(a)}{1!}(x-a) + \frac{C''(a)}{2!}(x-a)^2 + \cdots \tag{85}$$

gilt. Da man aber zu beliebigen Punkten a und x stets ein endliches Intervall finden kann, welches beide Punkte enthält, gilt die Beziehung (85) überhaupt für jedes x und a.

Ersetzen wir nun in (85) x durch $x + a$, so erhalten wir

$$C(x+a) = C(a) + \frac{C'(a)}{1!}x + \frac{C''(a)}{2!}x^2 + \frac{C'''(a)}{3!}x^3 + \cdots.$$

Diese Beziehung können wir unter Verwendung von (84) offenbar auch folgendermaßen schreiben:

$$C(x+a) = C(a) - S(a)\frac{x}{1!} - C(a)\frac{x^2}{2!} + S(a)\frac{x^3}{3!} + C(a)\frac{x^4}{4!} - \cdots.$$

Hierbei treten als Koeffizienten von $\frac{x^n}{n!}$ abwechselnd $C(a)$ und $S(a)$ und zwar ihrerseits mit abwechselnden Vorzeichen auf. Fassen wir nun in der zuletzt erhaltenen Reihe je zwei aufeinanderfolgende Glieder zusammen, so erhalten wir

$$C(x+a) = \left[C(a) - S(a)\frac{x}{1!}\right] + \left[-C(a)\frac{x^2}{2!} + S(a)\frac{x^3}{3!}\right] + \left[C(a)\frac{x^4}{4!} - S(a)\frac{x^5}{5!}\right] + \cdots.$$

Da man jedoch konvergente Reihen gliedweise subtrahieren kann, erhält man dasselbe Resultat, wenn man von der Reihe

$$C(a) - C(a)\frac{x^2}{2!} + C(a)\frac{x^4}{4!} - \cdots$$

die Reihe

$$S(a)\frac{x}{1!} - S(a)\frac{x^3}{3!} + S(a)\frac{x^5}{5!} - \cdots$$

subtrahiert. Schließlich kann man aber in diesen Reihen noch $C(a)$ bzw. $S(a)$ herausziehen. Damit erhalten wir

$$C(x+a) = C(a)\left[1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots\right] - S(a)\left[\frac{x}{1!} - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots\right].$$

Die hierbei auftretenden Reihen sind aber gerade die Reihen (80) und (81). Also gilt

$$C(x+a) = C(a)\,C(x) - S(a)\,S(x). \tag{86}$$

Das ist aber bereits das Additionstheorem für den Kosinus.

Ersetzen wir in (86) jetzt a durch $-a$, so erhalten wir unter Verwendung von (82) das folgende „Subtraktionstheorem"

$$C(x-a) = C(a)\,C(x) + S(a)\,S(x). \tag{87}$$

Setzen wir in (87) speziell $a = x$ und beachten wir, daß auf Grund der Definition der Funktion $C(x)$

$$C(0) = 1 \tag{88}$$

ist, so erhalten wir die bekannte Formel

$$C^2(x) + S^2(x) = 1, \tag{89}$$

die man bei der geometrischen Begründung der Trigonometrie gewöhnlich mit Hilfe des Pythagoreischen Lehrsatzes beweist.

Aus (89) ergibt sich übrigens sofort, daß stets

$$|C(x)| \leqq 1, \quad |S(x)| \leqq 1.$$

Aus den Formeln (86) und (87) erhält man durch Differentiation sofort die folgenden Additionstheoreme für den Sinus:

$$S(x+a) = S(x)\,C(a) + S(a)\,C(x), \tag{90}$$

$$S(x-a) = S(x)\,C(a) - S(a)\,C(x). \tag{91}$$

Aus den Additionstheoremen ergeben sich dann unmittelbar die folgenden Formeln für den doppelten Winkel:

$$C(2x) = C^2(x) - S^2(x),$$

$$S(2x) = 2S(x)\,C(x).$$

In ähnlicher Weise lassen sich viele weitere Beziehungen beweisen, wie z. B. die bekannten Formeln für $C(3x)$ und $S(3x)$ usw.

Etwas schwieriger ist es schon, die Periodizität von $C(x)$ und $S(x)$ zu beweisen und die damit zusammenhängenden Reduktionsformeln aufzustellen.

Hierzu bemerken wir zunächst, daß für $0 < x \leqq 2$ stets $S(x) > 0$ ist. Es ist nämlich

$$S(x) = \frac{x}{1!}\left(1 - \frac{x^2}{2\cdot 3}\right) + \frac{x^5}{5!}\left(1 - \frac{x^2}{6\cdot 7}\right) + \frac{x^9}{9!}\left(1 - \frac{x^2}{10\cdot 11}\right) + \cdots,$$

wobei für $0 < x \leqq 2$ alle Glieder positiv sind.

Beachten wir nun, daß $C'(x) = -S(x)$, so können wir damit behaupten, daß die Funktion $C(x)$ im Intervall $[0, 2]$ echt monoton fallend ist. Schreiben wir nun speziell $C(2)$ in der Form

$$C(2) = 1 - \frac{2^2}{2!} + \frac{2^4}{4!} - \frac{2^6}{6!}\left(1 - \frac{2^2}{7\cdot 8}\right) - \frac{2^{10}}{10!}\left(1 - \frac{2^2}{11\cdot 12}\right) - \cdots,$$

so erkennen wir, daß $C(2) < 0$ ist (denn in der angegebenen Darstellung für $C(2)$ sind die Differenzen in den Klammern alle positiv, während

$$1 - \frac{2^2}{2!} + \frac{2^4}{4!} = -\frac{1}{3}$$

ist). Hieraus und aus (88) folgt, daß die Funktion $C(x)$ im Intervall $(0, 2)$ genau eine Nullstelle besitzt, die wir mit $\frac{\tau}{2}$ bezeichnen wollen[1]),

$$C\left(\frac{\tau}{2}\right) = 0 \qquad \left(0 < \frac{\tau}{2} < 2\right). \tag{92}$$

Aus (89) und (92) folgt unmittelbar

$$S\left(\frac{\tau}{2}\right) = +1 \tag{93}$$

(der Fall $S\left(\frac{\tau}{2}\right) = -1$ kann nicht eintreten, weil die Funktion $S(x)$ im Intervall $(0, 2)$ positiv ist). Damit erhalten wir aber aus (86) und (90) sofort

$$C\left(x + \frac{\tau}{2}\right) = -S(x), \quad S\left(x + \frac{\tau}{2}\right) = C(x). \tag{94}$$

Hieraus folgt dann aber

$$\begin{aligned} C(x + \tau) &= -S\left(x + \frac{\tau}{2}\right) = -C(x), \\ S(x + \tau) &= C\left(x + \frac{\tau}{2}\right) = -S(x) \end{aligned} \tag{95}$$

und schließlich

$$C(x + 2\tau) = C(x), \quad S(x + 2\tau) = S(x), \tag{96}$$

d. h., *die Funktionen $C(x)$ und $S(x)$ besitzen die Periode 2τ.*

Gleichzeitig haben wir damit auch bereits einige der bekannten Reduktionsformeln (z. B. (94) und (95)) bewiesen.

Ersetzen wir in (87) nun x durch $\frac{\tau}{2}$ und a durch x, so erhalten wir ferner die Reduktionsformel

$$C\left(\frac{\tau}{2} - x\right) = S(x), \tag{97}$$

aus der sich wegen (89) die Beziehung

$$C\left(\frac{\tau}{4}\right) = S\left(\frac{\tau}{4}\right) = \frac{\sqrt{2}}{2} \tag{98}$$

ergibt.

---

[1]) Die hier eingeführte Zahl τ stimmt natürlich mit der Zahl π überein. Wir wollen jedoch auch hier auf die gewohnte Bezeichnung verzichten, um die Unabhängigkeit unserer Überlegungen von allen geometrischen Vorstellungen mit aller Deutlichkeit hervortreten zu lassen.

Mit Hilfe der im vorangehenden bewiesenen Formeln zeigt man nun leicht, daß *genau die Zahlen der Form* $n\tau$ $(n = 0, \pm 1, \pm 2, \ldots)$ *Nullstellen der Funktion* $S(x)$ *und genau die Zahlen der Form* $n\tau + \frac{\tau}{2}$ $(n = 0, \pm 1, \pm 2, \ldots)$ *Nullstellen der Funktion* $C(x)$ *sind.*

Nachdem auf die angegebene Weise die Theorie der Funktionen Kosinus und Sinus aufgebaut ist, können wir auf folgende Weise die Funktion „*Tangens*" definieren:

$$T(x) = \frac{S(x)}{C(x)}.$$

Die Funktion $T(x)$ ist ungerade, sie ist im offenen Interval $\left(-\frac{\tau}{2}, +\frac{\tau}{2}\right)$ stetig und echt monoton wachsend, und ihre Werte streben gegen $-\infty$ bzw. $+\infty$, falls x in diesem Intervall gegen $-\frac{\tau}{2}$ bzw. $+\frac{\tau}{2}$ strebt. Ferner ist

$$T(x+\tau) = \frac{S(x+\tau)}{C(x+\tau)} = \frac{-S(x)}{-C(x)} = \frac{S(x)}{C(x)} = T(x),$$

d. h., *die Funktion* $T(x)$ *besitzt die Periode* τ.

Wir wollen hier auf die Theorie der Funktion $T(x)$ nicht näher eingehen. Es sei lediglich auf folgende Beziehungen hingewiesen:

$$T\left(\frac{\tau}{4}\right) = 1, \quad T'(x) = \frac{1}{C^2(x)}.$$

Während die erste dieser Formeln unmittelbar aus (84) folgt, ergibt sich die zweite mit Hilfe der bekannten Differentiationsregel für einen Quotienten aus (83) und (89).

Beschränken wir uns bei der Funktion $T(x)$ auf Argumentwerte aus dem Intervall $\left(-\frac{\tau}{2}, +\frac{\tau}{2}\right)$, in welchem die Funktion $T(x)$ stetig und echt monoton wachsend ist, so können wir die Umkehrfunktion

$$x = a(y)$$

bilden, die für $-\infty < y < +\infty$ erklärt ist und deren Werte im Intervall $\left(-\frac{\tau}{2}, +\frac{\tau}{2}\right)$ liegen. Diese Funktion ist ebenfalls stetig und echt monoton wachsend. Auf Grund des Satzes über die Differentiation einer Umkehrfunktion (vgl. Abschnitt 5) ist sie überall differenzierbar, wobei

$$a'(y) = \frac{1}{T'(x)}.$$

Hierbei ist $T(x) = y$, so daß nach der Formel für $T'(x)$

$$a'(y) = C^2(x)$$

ist. Daher gilt

$$a'(y) = \frac{C^2(x)}{C^2(x) + S^2(x)} = \frac{1}{1 + \dfrac{S^2(x)}{C^2(x)}} = \frac{1}{1 + T^2(x)} = \frac{1}{1 + y^2}.$$

Bezeichnen wir schließlich die unabhängige Veränderliche wie üblich mit x, so finden wir

$$a'(x) = \frac{1}{1 + x^2}.$$

Hieraus folgt

$$\int_0^1 \frac{dx}{1 + x^2} = \Big[a(x)\Big]_0^1 = a(1) - a(0) = \frac{\tau}{4}. \tag{99}$$

Dies zeigt, daß wir die Konstante τ mit beliebiger Genauigkeit mit Hilfe des Integrals

$$\int_0^1 \frac{dx}{1 + x^2}$$

berechnen können. Dieses Integral haben wir aber bereits in Abschnitt 27 genauer untersucht. Auf Grund der dort erhaltenen Resultate können wir behaupten, daß mit einem Fehler, der dem Betrage nach kleiner als 0,005 ist, die Näherungsgleichung

$$\tau \approx 3{,}14$$

gilt. Eine andere Methode zur Berechnung von τ besteht darin, daß wir die Umkehrfunktion von $S(x)$ betrachten, wobei wir uns auf Argumentwerte aus dem Intervall $\left[-\frac{\tau}{2}, +\frac{\tau}{2}\right]$ beschränken. Entsprechend wie (99) findet man dann

$$\int_{-\frac{\sqrt{2}}{2}}^{+\frac{\sqrt{2}}{2}} \frac{dx}{\sqrt{1 - x^2}} = \frac{\tau}{2}. \tag{100}$$

Zum Abschluß wollen wir uns noch kurz der Frage zuwenden, wie man den Zusammenhang zwischen den Funktionen $S(x)$, $C(x)$, $T(x)$, $a(x)$ mit den üblichen trigonometrischen Funktionen $\sin x$, $\cos x$, $\operatorname{tg} x$, $\operatorname{arctg} x$ und den Zusammenhang zwischen τ und π herstellen kann. Diese Frage kann man nun von zwei verschiedenen Standpunkten aus beantworten:

Man kann sich einmal auf den Standpunkt stellen, daß die ganze Theorie der trigonometrischen Funktionen auf geometrischer Grundlage bereits entwickelt ist. In diesem Fall kann man alle aus der üblichen Theorie bekannten Eigenschaften der trigonometrischen Funktionen verwenden. Insbesondere kann man also auch die Potenzreihendarstellung von $\cos x$ und $\sin x$ als bekannt voraussetzen. In diesem Fall ergibt sich der genannte Zusammen-

hang unmittelbar auf Grund der Tatsache, daß die Funktionen $C(x)$ und $S(x)$ durch dieselben Potenzreihen wie $\cos x$ und $\sin x$ definiert wurden, so daß also $C(x) = \cos x$ und $S(x) = \sin x$ ist. Die Gleichheit von τ und π ergibt sich in diesem Fall aus den Beziehungen

$$\frac{\tau}{4} = \int_0^1 \frac{dx}{1+x^2}, \qquad \frac{\pi}{4} = \int_0^1 \frac{dx}{1+x^2}.$$

Von größerer methodischer Bedeutung ist jedoch der andere Gesichtspunkt, bei dem wir uns die Theorie der Funktionen $C(x)$ und $S(x)$, so wie dies im vorliegenden Abschnitt angedeutet wurde, rein analytisch aufgebaut denken und die Frage nach dem Zusammenhang dieser Funktionen mit der Kreisgeometrie ohne Rückgriff auf Tatsachen aus der üblichen Trigonometrie beantworten.

Hierzu gehen wir aus von der Beziehung

$$C^2(t) + S^2(t) = 1,$$

nach der für beliebiges t der Punkt mit den Koordinaten $(C(t), S(t))$ auf dem Kreis $x^2 + y^2 = 1$ liegt. Hieraus folgt sofort, daß $C(t)$ und $S(t)$ die Längen der Strecken OC und CM in Abb. 47 sind. Das gibt uns aber noch lange nicht das Recht, diese Strecken mit jenen Strecken zu identifizieren, die man in der üblichen Theorie mit $\cos t$ und $\sin t$ bezeichnet, da wir noch nicht wissen, daß t die in Abb. 47 angegebene Bedeutung hat, wenn M der Punkt $(C(t), S(t))$ ist. Um diese Frage zu klären und gleichzeitig festzustellen, welche geometrische Bedeutung τ hat, betrachten wir den oberen Halbkreis des Einheitskreises, der durch die Gleichung

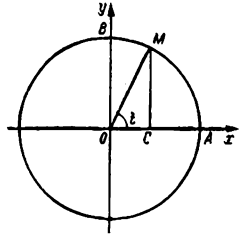


Abb. 47

$$y = \sqrt{1-x^2}$$

beschrieben wird. Nach Formel (4) auf Seite 398 ist die Länge des Kreisbogens über der Strecke $\left[-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right]$ der x-Achse gleich

$$\int_{-\frac{\sqrt{2}}{2}}^{+\frac{\sqrt{2}}{2}} \frac{dx}{\sqrt{1-x^2}},$$

d. h. (vgl. Formel (100)) gleich $\frac{\tau}{2}$. Daher ist aus Symmetriegründen der Umfang des gesamten Einheitskreises gleich 2τ, d. h., es ist $\tau = \pi$.

Ist nun M der Punkt mit den Koordinaten $C(t)$ und $S(t)$, so ist die Länge des Bogens $\overset{\frown}{BM}$ gleich

$$s_{\overset{\frown}{BM}} = \int_0^{C(t)} \frac{dx}{\sqrt{1-x^2}} = \int_0^{S\left(\frac{\pi}{2}-t\right)} \frac{dx}{\sqrt{1-x^2}} = \frac{\pi}{2} - t.$$ [1])

Daraus folgt, daß die Länge des Bogens $\overset{\frown}{AM}$ gleich $\frac{\pi}{2} - \left(\frac{\pi}{2} - t\right) = t$ ist. Also ist t die Größe des Winkels ACM.

Um also den Punkt $M = (C(t),\ S(t))$ zu konstruieren, müssen wir den Strahl ziehen, der im Punkte O die x-Achse unter dem Winkel t schneidet, und dessen Schnittpunkt mit der Kreisperipherie bestimmen. Daraus folgt, daß $C(t)$ und $S(t)$ die üblichen Funktionen $\cos t$ und $\sin t$ sind.

---

[1]) Man betrachtet hierzu für $-\frac{\pi}{2} \leqq x \leqq \frac{\pi}{2}$ die Umkehrfunktion $x = \alpha(y)$ von $y = S(x)$. Man zeigt, daß $\alpha'(x) = \frac{1}{\sqrt{1-x^2}}$, so daß

$$\int_0^{S\left(\frac{\pi}{2}-t\right)} \frac{dx}{\sqrt{1-x^2}} = \Big[\alpha(x)\Big]_0^{S\left(\frac{\pi}{2}-t\right)} = \frac{\pi}{2} - t.$$

W. L. GONTSCHAROW

ELEMENTARE FUNKTIONEN EINER KOMPLEXEN VERÄNDERLICHEN

§ 1. Rationale Funktionen

Im folgenden verwenden wir, wie dies weitgehend üblich ist, den Buchstaben z als Variable für komplexe Zahlen, wobei wir den Realteil von z mit x und den Imaginärteil von z mit y bezeichnen, so daß also

$$z = x + iy.$$

Sind in einem gewissen Zusammenhang mehrere komplexe Zahlen zu unterscheiden, so deuten wir dies durch Indizes an oder verwenden andere Buchstaben:

$$z_1 = x_1 + iy_1, \quad z_2 = x_2 + iy_2, \ \ldots, \quad \zeta = \xi + i\eta \text{ usw.}$$

Die Rechengesetze für die arithmetischen Grundoperationen an komplexen Zahlen setzen wir als bekannt voraus. Wir erwähnen hier nur, daß bezüglich der Addition und Multiplikation von komplexen Zahlen die üblichen formalen Rechengesetze der Algebra gelten, wobei man zudem i^2 jeweils nach Belieben durch -1 ersetzen darf. Zum Beispiel ist also

$$z_1 + z_2 = (x_1 + iy_1) + (x_2 + iy_2) = (x_1 + x_2) + i(y_1 + y_2),$$
$$z_1 z_2 = (x_1 + iy_1)(x_2 + iy_2) = (x_1 x_2 - y_1 y_2) + i(x_1 y_2 + x_2 y_1).$$

Ferner gelten die folgenden Rechengesetze:

$$z_1 + z_2 = z_2 + z_1, \quad z_1 + (z_2 + z_3) = (z_1 + z_2) + z_3,$$
$$z_1 z_2 = z_2 z_1, \qquad z_1(z_2 z_3) = (z_1 z_2) z_3,$$
$$z_1(z_2 + z_3) = z_1 z_2 + z_1 z_3.$$

Was die Umkehroperationen von Addition und Multiplikation, die Subtraktion und die Division betrifft, so sind diese, die Division durch Null selbstverständlich ausgenommen, stets eindeutig ausführbar. Bei der Division geht man dabei praktisch stets so vor, daß man Zähler (Dividend) und Nenner (Divisor) mit der zum Nenner konjugierten Zahl erweitert,

$$\frac{z_1}{z_2} = \frac{x_1 + iy_1}{x_2 + iy_2} = \frac{(x_1 + iy_1)(x_2 - iy_2)}{(x_2 + iy_2)(x_2 - iy_2)} = \frac{x_1 x_2 + y_1 y_2}{x_2^2 + y_2^2} + i\frac{-x_1 y_2 + x_2 y_1}{x_2^2 + y_2^2}.$$

Aus den genannten Tatsachen folgt, daß *alle Regeln für die rationalen algebraischen Operationen im Komplexen genauso gelten wie im Reellen.*

Wir können daher auch im Komplexen genau wie im Reellen den Begriff der *rationalen Funktion* erklären.

Eine gegebene Funktion $f(z)$ einer komplexen Veränderlichen z (die im allgemeinen auch komplexe Werte annimmt), d. h. die abhängige Veränderliche, wollen wir im folgenden mit w bezeichnen, wobei wir den Real- und den Imaginärteil von w mit u bzw. v andeuten wollen:

$$\left.\begin{aligned} w &= f(z), \\ z = x + iy, \quad w &= u + iv. \end{aligned}\right\} \tag{1}$$

Eine rationale Funktion $w = f(z)$ der komplexen Veränderlichen z ist nun offenbar eindeutig festgelegt durch gewisse rationale Operationen und die Reihenfolge, in der man diese Operationen anwenden muß, um aus einem gegebenen Argumentwert z und evtl. gegebenen konstanten (im allgemeinen komplexen) Zahlen den zugehörigen Funktionswert w zu berechnen. Daher kann man bei einer rationalen Funktion den zu einem Argumentwert z gehörenden Funktionswert w stets rein formal nach den Regeln der Algebra ausrechnen (wobei man natürlich nicht durch Null dividieren darf). Ist z. B.

$$w = f(z) = \frac{z}{z-1},$$

so erhält man für

$$z = 2 + 3i$$

den Funktionswert

$$w = \frac{2+3i}{(2+3i)-1} = \frac{2+3i}{1+3i} = \frac{(2+3i)(1-3i)}{(1+3i)(1-3i)} = \frac{11}{10} - \frac{3}{10}i.$$

Allgemein kann man bei einer rationalen Funktion in die Formel für $f(z)$ beliebig vorgegebene Argumentwerte einsetzen und auf diese Weise eine umfassende Wertetabelle aufstellen:

z	w
$2+3i$	$\frac{11-3i}{10}$
$1+i$	$1-i$
$3-2i$	$\frac{5+i}{4}$
i	$\frac{1-i}{2}$
$\frac{3-i}{2}$	$2+i$
.....	.....

Spalten wir bei einer gegebenen Funktion $w = f(z)$ sowohl z als auch w in Real- und Imaginärteil auf, so können wir, wenn wir, wie verabredet, $z = x + iy$ und $w = u + iv$ setzen, u und v als (reelle) Funktionen der reellen Veränderlichen x und y ansehen:

$$\left.\begin{aligned} u &= u(x, y), \\ v &= v(x, y). \end{aligned}\right\} \tag{2}$$

Es zeigt sich dabei, daß *im Fall einer rationalen Funktion* $w = f(z)$ *auch* u *und* v *rationale Funktionen der Veränderlichen* x *und* y *werden.* So ist z. B. für die bereits oben betrachtete Funktion $w = \frac{z}{z-1}$

$$u + iv = \frac{x+iy}{(x+iy)-1},$$

also

$$u + iv = \frac{(x^2+y^2-x)-iy}{(x-1)^2+y^2},$$

so daß hier

$$u = \frac{x^2+y^2-x}{(x-1)^2+y^2}, \quad v = -\frac{y}{(x-1)^2+y^2}.$$ [1])

Wir können daher anstelle einer $(z \to w)$-Tabelle auch eine $((x, y) \to (u, v))$-Tabelle aufstellen:

x	y	u	v
2	3	$\frac{11}{10}$	$-\frac{3}{10}$
1	1	1	-1
3	-2	$\frac{5}{4}$	$\frac{1}{4}$
0	1	$\frac{1}{2}$	$-\frac{1}{2}$
$\frac{3}{2}$	$-\frac{1}{2}$	2	1
...	.....	...	.....

Wir weisen darauf hin, daß im betrachteten Beispiel z nicht den Wert 1 annehmen darf, daß also nicht gleichzeitig $x = 1$ und $y = 0$ sein darf.

Wir empfehlen dem Leser, sich die Zerlegung einer Funktion in Real- und Imaginärteil noch an folgenden Beispielen klarzumachen:

1. $w = z^2$ $\quad (u = x^2 - y^2, \quad v = 2xy)$;
2. $w = z^3$ $\quad (u = x^3 - 3xy^2, \quad v = 3x^2y - y^3)$;
3. $w = \frac{1}{z}$ $\quad \left(u = \frac{x}{x^2+y^2}, \quad v = -\frac{y}{x^2+y^2}\right)$;
4. $w = \frac{z-i}{z+i}$ $\quad \left(u = \frac{x^2+y^2-1}{x^2+(y+1)^2}, \quad v = -\frac{2x}{x^2+(y+1)^2}\right)$.

Interpretieren wir die komplexen Zahlen in bekannter Weise als Punkte der „komplexen" z-Ebene bzw. der reellen x, y-Ebene, so erhalten wir durch

[1]) Komplexe Zahlen sind dann und nur dann gleich, wenn ihre Realteile und Imaginärteile untereinander gleich sind, d. h., *eine Gleichung im Komplexen ist stets zwei Gleichungen im Reellen gleichwertig.*

die Funktionalbeziehung (1) bzw. (2) eine geometrische „Abbildung" der z-Ebene in die w-Ebene bzw. der x, y-Ebene in die u, v-Ebene (vgl. S. 247).

Wir weisen darauf hin, daß wir bisher nur über die vier arithmetischen Grundoperationen verfügen, so daß wir an Funktionen im „operativen" Sinne im Augenblick nur rationale Funktionen betrachten können; Bildungen wie

$$2^i, \quad \cos i, \quad \lg(3i+2) \text{ usw.}$$

sind also zur Zeit vollkommen sinnlos, da wir bisher noch nicht definiert haben, was unter der Exponentialfunktion, den trigonometrischen Funktionen, den Logarithmen usw. im Komplexen zu verstehen ist.[1]) Es wird nun gerade eine Hauptaufgabe dieses Artikels sein, die Definitionen der elementaren Funktionen auf den Fall einer komplexen Veränderlichen auszudehnen. Dieser Aufgabe sind die nächsten Paragraphen gewidmet (§ 3 bis § 8). Erst im Anschluß daran wollen wir auf die Prinzipien, die dieser Ausdehnung zu Grunde liegen und sie logisch als empfehlenswert erscheinen lassen, eingehen (vgl. § 9 bis § 16).

Da sich die uns interessierenden Definitionen am einfachsten mittels Reihen geben lassen, wollen wir im nachfolgenden Paragraphen zunächst auf den Begriff des Grenzwertes und den der Reihe im Komplexen eingehen.

§ 2. Grenzwerte. Reihen

Die Definition des Grenzwertes einer Zahlenfolge[2])

$$\{z_n\} \equiv z_1, z_2, z_3, \ldots$$

ist im Komplexen wörtlich dieselbe wie im Reellen: Man sagt, *die Zahlenfolge* $\{z_n\}$ *strebe* bzw. *konvergiere gegen den Grenzwert* c,

$$\lim z_n = c, \quad z_n \to c,$$

wenn es zu jeder noch so kleinen reellen Zahl $\varepsilon > 0$ eine natürliche Zahl N_ε gibt, so daß für alle $n > N_\varepsilon$ die Ungleichung

$$|z_n - c| < \varepsilon \tag{3}$$

gilt. Der absolute Betrag auf der linken Seite dieser Ungleichung ist dabei die sinngemäße Verallgemeinerung des gewöhnlichen absoluten Betrages.[3])

---

[1]) Gelegentlich führt man in der elementaren Algebra neben den algebraischen Operationen auch andere Operationen, wie z. B. das Ausziehen der Wurzel n-ten Grades,

$$f(z) = \sqrt[n]{z},$$

ein.

[2]) Über den Begriff des Grenzwertes einer Funktion werden wir erst in § 9 sprechen.

[3]) Der absolute Betrag von z ist bekanntlich definiert als

$$|z| = |x + iy| = \sqrt{x^2 + y^2}.$$

Vgl. hierzu EdEM, Bd. 1, S. 209.

Man macht sich leicht die geometrische Bedeutung der Ungleichung (3) klar. Ist nämlich

$$z_n = x_n + iy_n, \quad c = a + ib,$$

so ist

$$|z_n - c| = \sqrt{(x_n - a)^2 + (y_n - b)^2}$$

nichts anderes als der Abstand des Punktes z_n vom Punkte c in der komplexen z-Ebene. Daher kann man die Konvergenz der Zahlenfolge $\{z_n\}$ gegen c geometrisch folgendermaßen interpretieren: In jedem Kreis um c mit einem noch so kleinen Radius ε (ε-Kreis, ε-Umgebung) liegen von einem gewissen Gliede ab alle Glieder der Folge $\{z_n\}$. Neben den endlichen Grenzwerten betrachtet man noch einen unendlichen Grenzwert ∞ (ohne Vorzeichen!), und zwar bedeutet

$$\lim z_n = \infty, \quad z_n \to \infty,$$

daß es zu jeder noch so großen Zahl M eine Zahl N_M gibt, so daß für alle $n > N_M$ die Ungleichung

$$|z_n| > M \tag{4}$$

gilt. Geometrisch besagt dies, daß bei noch so großem M von einem gewissen (von M abhängigen) Index ab alle Glieder der Folge $\{z_n\}$ außerhalb des Kreises um O vom Radius M liegen.

Auf Grund der formalen Analogie zwischen dem Begriff des Grenzwertes im Komplexen und dem des Grenzwertes im Reellen ergibt sich sofort, daß für jenen dieselben Rechengesetze wie für diesen gelten:

Wenn $z_n' \to c'$ *und* $z_n'' \to c''$, *so*

I. $z_n' + z_n'' \to c' + c''$,

II. $z_n' - z_n'' \to c' - c''$,

III. $z_n' z_n'' \to c' c''$,

IV. $\dfrac{z_n'}{z_n''} \to \dfrac{c'}{c''}$ (*sofern* $c'' \neq 0$).

Bei der Übertragung der Beweise hat man nur zu beachten, daß der absolute Betrag im Komplexen die folgenden aus dem Reellen bekannten Eigenschaften besitzt:

$$|z' + z''| \leqq |z'| + |z''|,$$
$$|z' z''| = |z'| \cdot |z''|.$$

Gelegentlich kann man auch die folgenden Beziehungen gut gebrauchen, deren geometrischer Inhalt hinreichend klar sein dürfte. Durch sie wird die Frage nach der Konvergenz einer komplexen Zahlenfolge auf die nach der Konvergenz von zwei reellen Zahlenfolgen zurückgeführt. Setzen wir nämlich

$$z_n = x_n + iy_n, \quad c = a + ib,$$

so gilt:

1. *Wenn* $x_n \to a$ *und* $y_n \to b$, *so* $z_n \to c$.
2. *Wenn* $z_n \to c$, *so* $x_n \to a$ *und* $y_n \to b$.

Die Beweise hierfür ergeben sich unmittelbar mit Hilfe der folgenden Ungleichungen:

$$\left.\begin{matrix}|x_n-a|\\|y_n-b|\end{matrix}\right\} \leqq |z_n-c| \leqq |x_n-a|+|y_n-b|$$

(jede Kathete eines rechtwinkligen Dreiecks ist kleiner als die Hypotenuse, und die Hypotenuse ist kleiner als die Summe der Katheten).

Man kann aber auch noch auf eine andere Weise die Frage nach der Konvergenz einer komplexen Zahlenfolge auf die nach der Konvergenz von zwei reellen Zahlenfolgen zurückführen. Setzen wir nämlich

$$z_n = r_n(\cos\vartheta_n + i\sin\vartheta_n), \quad c = \varrho(\cos\omega + i\sin\omega),$$

so gilt:

3. *Wenn* $\vartheta_n \to \omega$ *und* $r_n \to \varrho$ $(\varrho \neq 0)$,[1] *so* $z_n \to c$.

In der Tat gilt unter den angegebenen Voraussetzungen

$$x_n = r_n\cos\vartheta_n \to \varrho\cos\omega = a,$$
$$y_n = r_n\sin\vartheta_n \to \varrho\sin\omega = b,$$

und die Behauptung ist eine unmittelbare Folge aus Satz 1.

4. *Wenn*

$$z_n \to c \qquad (c \neq 0),$$

so ist $\varrho \neq 0$, *und es gilt*

$$r_n \to \varrho$$

und bei geeigneter Normierung des Arguments ϑ_n[2] *auch*

$$\vartheta_n \to \omega.$$

In der Tat: Auf Grund von

$$|r_n-\varrho| = \big||z_n|-|c|\big| \leqq |z_n-c|$$

ergibt sich zunächst unmittelbar, daß $r_n \to \varrho$. Ferner erhalten wir nach Satz 2, daß

$$r_n\cos\vartheta_n \to \varrho\cos\omega,$$
$$r_n\sin\vartheta_n \to \varrho\sin\omega,$$

so daß

$$\cos\vartheta_n \to \cos\omega,$$
$$\sin\vartheta_n \to \sin\omega.$$

Hieraus folgt jedoch bei passender Auswahl der Argumente ϑ_n sofort

$$\vartheta_n \to \omega.$$

---

[1]) Wenn $r_n \to 0$, so gilt (ohne jegliche Voraussetzungen über ϑ_n): $z_n \to 0$, und umgekehrt.

[2]) Bekanntlich ist das Argument einer von Null verschiedenen komplexen Zahl nur bis auf Vielfache von 2π eindeutig bestimmt. Wir können daher annehmen, daß die Argumente der Zahlen z_n und c etwa folgenden Bedingungen genügen: Für reelles $c > 0$ sei

$$\omega = 0, \quad -\pi < \vartheta_n \leqq +\pi,$$

während in allen anderen Fällen

$$0 < \omega < 2\pi, \quad 0 \leqq \vartheta_n < 2\pi$$

sei.

Es ist klar, daß man den Begriff der unendlichen Reihe rein formal auf den Fall übertragen kann, daß die Glieder der Reihe komplexe Zahlen sind. Dann sind wir aber jetzt auch in der Lage, den Begriff der Konvergenz von Reihen aus dem Reellen unmittelbar ins Komplexe zu übernehmen.

Wir werden also die Reihe

$$\sum_{n=1}^{\infty} z_n = z_1 + z_2 + \cdots + z_n + \cdots$$

genau dann *konvergent* nennen, wenn die Folge ihrer Partialsummen

$$s_n = z_1 + z_2 + \cdots + z_n \qquad (n = 1, 2, \ldots)$$

gegen einen gewissen Grenzwert s konvergiert,

$$\lim s_n = s. \tag{5}$$

Die (im allgemeinen komplexe) Zahl s werden wir dann sinngemäß als *Summe* der betrachteten Reihe ansehen.

Entsprechend wie im Reellen nennt man im Komplexen eine Reihe $\sum_{n=1}^{\infty} z_n$ *absolut konvergent*, wenn nicht nur sie selbst, sondern auch die aus den absoluten Beträgen ihrer Glieder gebildete Reihe $\sum_{n=1}^{\infty} |z_n|$ konvergiert.

Für $z_n = x_n + i y_n$ ist die Reihe $\sum_{n=1}^{\infty} z_n$ dann und nur dann absolut konvergent, wenn sowohl die Reihe $\sum_{n=1}^{\infty} x_n$ als auch die Reihe $\sum_{n=1}^{\infty} y_n$ absolut konvergieren.

Dieser Satz ergibt sich unmittelbar daraus, daß einerseits die Ungleichungen $|x_n| \leqq |z_n|$ und $|y_n| \leqq |z_n|$ und andererseits die Ungleichung $|z_n| \leqq |x_n| + |y_n|$ gelten.[1])

Aus diesem Satz folgt sofort, daß *aus der Konvergenz der Reihe $\sum_{n=1}^{\infty} |z_n|$ die der Reihe $\sum_{n=1}^{\infty} z_n$ folgt* (es genügt hierzu, auf Satz 1 von Seite 421 und auf Satz 1 von Seite 477 zu verweisen).

Die Frage nach der absoluten Konvergenz einer Reihe mit komplexen Gliedern ist damit auf die Frage nach der Konvergenz einer Reihe mit nicht negativen reellen Gliedern zurückgeführt.

Man zeigt entsprechend wie im Reellen, daß *die absolut konvergenten Reihen im großen und ganzen dieselben Eigenschaften wie endliche Summen besitzen.* So wird z. B. ihre Konvergenz nicht gestört und ihre Summe nicht geändert, wenn man ihre Glieder in irgendeiner Weise vertauscht oder in irgendeiner

---

[1]) Man beachte, daß $\sum_{n=1}^{\infty} |z_n|$ eine Reihe mit nicht negativen reellen Gliedern ist. — *Anm. d. wissenschaftl. Red.*

Weise Klammern einfügt oder fortläßt. Ferner kann man absolut konvergente Reihen in der bekannten Weise miteinander multiplizieren:

$$\sum_{n=1}^{\infty} z_n' \cdot \sum_{n=1}^{\infty} z_n'' = \sum_{p=1}^{\infty} \sum_{q=1}^{\infty} z_p' z_q''.^{1)} \tag{6}$$

Alles dies trifft für nicht-absolut konvergente Reihen im allgemeinen nicht zu. Bei den folgenden Darlegungen werden wir es jedoch meistens mit absolut konvergenten Reihen zu tun haben.

§ 3. Die Exponentialfunktionen, der Sinus und der Kosinus

In der Schulmathematik wird in der Regel keine Vereinbarung darüber getroffen, was unter der Exponentialfunktion mit einer positiven Basis bei komplexen Exponenten verstanden werden soll. Ebensowenig wird in der Schule auf die Frage eingegangen, was man unter den Werten der trigonometrischen Funktionen bei komplexem Argument zu verstehen hat.

Diese Fragen ergeben sich jedoch zwangsläufig, wenn man die Potenzreihen dieser Funktionen betrachtet:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots,$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots,$$

$$\sin x = \frac{x}{1!} - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots.$$

Diese Beziehungen gelten zunächst, wie wir bereits gesehen haben (vgl. S. 429), für beliebiges reelles x, sind also im Reellen Identitäten. Man sieht nun aber unmittelbar, daß die rechten Seiten dieser Beziehungen auch dann ihren Sinn nicht verlieren, wenn man für x eine beliebige komplexe Zahl einsetzt. In diesem Fall erhält man jeweils (und zwar bei jeder derartigen Substitution) eine absolut konvergente Reihe, die eine ganz bestimmte (komplexe) Zahl als Summe besitzt.[2)]

Auf Grund dieser Tatsache dürfte es nicht als verwunderlich erscheinen, daß in der höheren Mathematik folgende Definition üblich ist:

[1)] Alle diese Eigenschaften müssen natürlich bewiesen werden. Es ist keineswegs so, daß sich die angegebenen Gesetze von den endlichen Summen automatisch auf unendliche Reihen übertragen (vgl. dazu Seite 425ff).

[2)] Bezeichnen wir nämlich den Betrag der betrachteten komplexen Zahl mit r, so ist auf Grund des D'ALEMBERTschen Kriteriums (vgl. Seite 432) zunächst die Reihe

$$1 + \frac{r}{1!} + \frac{r^2}{2!} + \frac{r^3}{3!} + \cdots$$

konvergent. Dann konvergieren aber erst recht die Reihen

$$1 + \frac{r^2}{2!} + \frac{r^4}{4!} + \cdots \quad \text{und} \quad \frac{r}{1!} + \frac{r^3}{3!} + \frac{r^5}{5!} + \cdots,$$

die sich aus der ursprünglichen durch Streichen gewisser Glieder ergeben.

Unter e^z, $\cos z$, $\sin z$ versteht man bei beliebigem (reellen oder komplexen) z die Summe der absolut konvergenten Reihe:

$$\left.\begin{aligned} e^z &= 1+\frac{z}{1!}+\frac{z^2}{2!}+\frac{z^3}{3!}+\cdots,\\ \cos z &= 1-\frac{z^2}{2!}+\frac{z^4}{4!}-\frac{z^6}{6!}+\cdots,\\ \sin z &= \frac{z}{1!}-\frac{z^3}{3!}+\frac{z^5}{5!}-\frac{z^7}{7!}+\cdots, \end{aligned}\right\} \tag{7}$$

wofür man auch kürzer (und prägnanter)

$$e^z=\sum_{n=0}^{\infty}\frac{z^n}{n!},\quad \cos z=\sum_{n=0}^{\infty}(-1)^n\frac{z^{2n}}{(2n)!},\quad \sin z=\sum_{n=0}^{\infty}(-1)^n\frac{z^{2n+1}}{(2n+1)!} \tag{8}$$

schreiben kann. Nach dieser Definition werden also die angegebenen Gleichungen als Identitäten in der ganzen komplexen Ebene angesehen.

Die Zweckmäßigkeit dieser Definition (im Vergleich zu anderen äquivalenten Definitionen) wird später noch erläutert werden (vgl. § 15).

Auf Grund der angegebenen Formeln ist es möglich, den Wert jeder der drei Funktionen bei beliebig vorgegebenem Argumentwert z mit verlangter Genauigkeit näherungsweise zu berechnen. Will man etwa e^{3+2i} ermitteln, so hat man in der ersten Formel auf der rechten Seite für z den Wert $3+2i$ einzusetzen und dann eine hinreichend große Anzahl von Gliedern zu summieren. Die Potenzen von $3+2i$ können dabei nach dem binomischen Satz berechnet werden.

Man kann diese Berechnung dadurch wesentlich vereinfachen, daß man von dem sogenannten Additionstheorem der Funktion $f(z)=e^z$ Gebrauch macht. Dieses Additionstheorem besagt, daß für beliebiges z_1 und z_2 die Beziehung

$$f(z_1+z_2)=f(z_1)f(z_2) \tag{9}$$

gilt. In der ausführlicheren Form

$$e^{z_1+z_2}=e^{z_1}e^{z_2} \tag{10}$$

ist dieses Additionstheorem dem Leser sicher bereits für reelles z_1 und z_2 bekannt. Wir wollen nun zeigen, daß diese Beziehung für beliebiges komplexes z_1 und z_2 richtig ist. In der Tat ist

$$\begin{aligned} f(z_1+z_2)=e^{z_1+z_2}&=\sum_{n=0}^{\infty}\frac{(z_1+z_2)^n}{n!}\\ &=\sum_{n=0}^{\infty}\frac{1}{n!}\sum_{p+q=n}\frac{n!}{p!\,q!}z_1^p z_2^q=\sum_{p=0}^{\infty}\sum_{q=0}^{\infty}\frac{1}{p!\,q!}z_1^p z_2^q\\ &=\sum_{p=0}^{\infty}\frac{z_1^p}{p!}\cdot\sum_{q=0}^{\infty}\frac{z_2^q}{q!}=\sum_{n=0}^{\infty}\frac{z_1^n}{n!}\cdot\sum_{n=0}^{\infty}\frac{z_2^n}{n!}=e^{z_1}e^{z_2}=f(z_1)f(z_2).\ ^{1}) \end{aligned}$$

[1]) Wir verwenden hier die folgende Schreibweise für die binomische Formel:

$$(z_1+z_2)^n=\sum_{p+q=n}\frac{n!}{p!\,q!}z_1^p z_2^q$$

(p und q durchlaufen dabei alle nicht negativen ganzen Zahlen, die der Gleichung $p+q=n$ genügen).

Setzen wir in (10) speziell

$$z_1 = x, \quad z_2 = iy, \quad z_1 + z_2 = x + iy = z,$$ [1]

so erhalten wir

$$e^z = e^{x+iy} = e^x \cdot e^{iy}. \tag{11}$$

Dabei ist

$$\begin{aligned} e^{iy} &= 1 + \frac{iy}{1!} + \frac{(iy)^2}{2!} + \frac{(iy)^3}{3!} + \frac{(iy)^4}{4!} + \frac{(iy)^5}{5!} + \frac{(iy)^6}{6!} + \frac{(iy)^7}{7!} + \cdots \\ &= 1 + \frac{iy}{1!} - \frac{y^2}{2!} - \frac{iy^3}{3!} + \frac{y^4}{4!} + \frac{iy^5}{5!} - \frac{y^6}{6!} - \frac{iy^7}{7!} + \cdots \\ &= \left(1 - \frac{y^2}{2!} + \frac{y^4}{4!} - \frac{y^6}{6!} + \cdots\right) + i\left(\frac{y}{1!} - \frac{y^3}{3!} + \frac{y^5}{5!} - \frac{y^7}{7!} + \cdots\right). \end{aligned}$$

Die Ausdrücke in den Klammern sind dabei aber auf Grund von (7) nichts anderes als die Reihen für $\cos y$ bzw. $\sin y$, so daß also

$$e^{iy} = \cos y + i \sin y. \tag{12}$$

Das ist die sogenannte Eulersche Identität, *durch welche die Potenz* e^{iy} *mit rein imaginärem Exponenten durch Kosinus und Sinus mit reellem Argument ausgedrückt wird.*

Wir kommen im folgenden Paragraphen nochmals auf die Eulersche Identität zurück.

Setzen wir den Ausdruck (12) für e^{iy} in die Formel (11) für e^z ein, so erhalten wir

$$e^z = e^x(\cos y + i \sin y). \tag{13}$$

Nach dieser Formel läßt sich nun e^z für beliebiges komplexes z sehr einfach berechnen. Setzen wir in (13) etwa $x = 3$ und $y = 2$, so finden wir

$$\begin{aligned} e^{3+2i} &= e^3(\cos 2 + i \sin 2) \approx (2{,}71\ldots)^3(\cos 114° 35' + i \sin 114° 35') \\ &\approx -8{,}23 + 18{,}27\,i. \end{aligned}$$

Ebenso wie Formel (7) kann man auch Formel (13) zur Definition der Funktion $w = e^z$ für komplexes z verwenden, und zwar ist sie im gewissen Sinne noch bequemer als Formel (7). Diese Formel hat den Vorteil, daß man aus ihr ohne weiteres den Realteil u und den Imaginärteil v der Funktion w ablesen kann, und zwar ist

$$u = e^x \cos y, \quad v = e^x \sin y.$$

Außerdem kann man aber aus (13) sofort auch den Betrag ϱ und das Argument φ von w erkennen. Da nämlich $\cos y + i \sin y$ offenbar den Betrag Eins besitzt, ist

$$\varrho = |e^z| = e^x, \quad \varphi = \arg e^z = y. \tag{14}$$

Der Eulerschen Identität ist übrigens unmittelbar zu entnehmen, daß die Exponentialfunktion e^z noch folgende sehr bemerkenswerte Eigen-

[1] Wir wollen verabreden, daß hier und im folgenden die Buchstaben x und y stets reelle Zahlen bezeichnen sollen. Wir weisen jedoch darauf hin, daß man in den Formeln (11) und (12) x und y auch als komplexe Zahlen auffassen kann.

schaften besitzt (die den Leser sicher überraschen wird): *Die Exponentialfunktion e^z ist periodisch, und zwar besitzt sie die rein imaginäre Periode*

$$\omega = 2\pi i.$$

Setzen wir nämlich in (12) $y = 2\pi$, so erhalten wir

$$e^{2\pi i} = 1, \tag{15}$$

so daß auf Grund von (10) (für $z_1 = z$ und $z_2 = \omega = 2\pi i$)

$$e^{z+2\pi i} = e^z e^{2\pi i} = e^z.$$

Allgemein ist für beliebiges ganzes n

$$e^{z+2n\pi i} = e^z. \tag{16}$$

Bezüglich der Exponentialfunktion mit irgendeiner positiven (reellen) Zahl a als Basis bemerken wir hier nur, daß man entsprechend wie im Reellen auch im Komplexen folgende Definition trifft:

$$a^z = e^{z \ln a}. \tag{17}$$

Hieraus und aus (13) folgt sofort

$$a^z = a^x(\cos(y \ln a) + i \sin(y \ln a)). \tag{18}$$

Aufgaben.

1. Man zeige, daß $2^i \approx 0{,}77 + 0{,}64i$.
2. Man zeige, daß für die Funktion $w = e^{z^2}$ Real- und Imaginärteil durch

$$u = e^{x^2 - y^2} \cos 2xy, \quad v = e^{x^2 - y^2} \sin 2xy$$

gegeben werden.

3. Man zeige, daß für $w = ze^z$

$$u = e^x(x \cos y - y \sin y), \quad v = e^x(x \sin y + y \cos y)$$

ist.

4. Man bestimme Betrag und Argument von

$$e^{3+2i}, \quad 2^i, \quad 5^{2+3i}, \quad e^{\pi i}, \quad e^{\frac{\pi i}{2}}.$$

5. Es sei $f(z) = \dfrac{\sin z}{z}$. Man beweise, daß $f(0) = 1$.

§ 4. Darstellung der trigonometrischen Funktionen mit Hilfe der Exponentialfunktion

Es sei z eine beliebige komplexe Zahl. Dann ist auf Grund der Eulerschen Identität

$$e^{iz} = \cos z + i \sin z.^{1)} \tag{19}$$

Ersetzen wir hier z durch $-z$, so erhalten wir

$$e^{-iz} = \cos z - i \sin z.^{2)} \tag{20}$$

---

1) Vgl. Fußnote 1 auf Seite 482.

2) Die Beziehungen

$$\cos(-z) = \cos z, \quad \sin(-z) = -\sin z$$

ergeben sich unmittelbar daraus, daß in der Potenzreihenentwicklung des Kosinus nur gerade Potenzen und in der des Sinus nur ungerade Potenzen auftreten.

Lösen wir die Gleichungen (19) und (20) nach $\cos z$ und $\sin z$ auf, so finden wir

$$\cos z = \frac{e^{iz} + e^{-iz}}{2}, \quad \sin z = \frac{e^{iz} - e^{-iz}}{2i}. \tag{21}$$

Dies sind Identitäten in z, durch welche die trigonometrischen Funktionen $\cos z$ und $\sin z$ durch die Exponentialfunktion e^z ausgedrückt werden. Die Beziehungen (21) sind sowohl in theoretischer als auch in praktischer Hinsicht von großer Bedeutung.

Vom theoretischen Standpunkt aus ist es zunächst interessant, daß im Komplexen die trigonometrischen Funktionen ihre selbständige Bedeutung verlieren: Man kann immer „ohne sie auskommen", da man statt ihrer überall die entsprechenden Ausdrücke in der Exponentialfunktion verwenden kann. Allerdings ist dies in der Schule normalerweise nicht möglich, weil man sich dort mit der üblichen Definition der trigonometrischen Funktionen begnügen muß, die nur im Reellen sinnvoll ist, dafür aber den Vorteil größerer Anschaulichkeit und unmittelbarer Anwendbarkeit besitzt.

Was die praktische Bedeutung der Beziehungen (21) betrifft, so wollen wir hier nur bemerken, daß sich mit Hilfe dieser Beziehungen viele Formeln wesentlich kürzer schreiben lassen, wobei sich häufig auch die Beweise erheblich vereinfachen. Wir wollen dies an einigen Beispielen zeigen:

1. Wir beweisen als erstes das Additionstheorem für den Kosinus. Aus der Identität

$$\frac{e^{iz_1} + e^{-iz_1}}{2} \cdot \frac{e^{iz_2} + e^{-iz_2}}{2} - \frac{e^{iz_1} - e^{-iz_1}}{2i} \cdot \frac{e^{iz_2} - e^{-iz_2}}{2i} = \frac{e^{i(z_1+z_2)} + e^{-i(z_1+z_2)}}{2}$$

folgt sofort

$$\cos z_1 \cos z_2 - \sin z_1 \sin z_2 = \cos(z_1 + z_2). \tag{22}$$

(entsprechend für den Sinus).

2. Auf Grund des Additionstheorems für die Exponentialfunktion ist zunächst

$$e^{i(z_1+z_2+\cdots+z_n)} = e^{iz_1} e^{iz_2} \cdots e^{iz_n}.$$

Setzen wir nun hier alle Zahlen $z_k\,(k = 1, 2, \ldots, n)$ gleich einer gegebenen Zahl z, so finden wir

$$(e^{iz})^n = e^{inz}. \tag{23}$$

Das ist aber nichts anderes als eine kurze Schreibweise für die bekannte Moivresche Formel

$$(\cos z + i \sin z)^n = \cos nz + i \sin nz. \tag{24}$$

3. Gesucht ist eine Darstellung von $\cos^n z$ bzw. $\sin^n z$ als Linearkombination in $1, \cos z, \sin z, \cos 2z, \sin 2z, \ldots, \cos nz, \sin nz$.

Dazu schreiben wir z. B. im Fall des Kosinus $\cos^n z$ in der Form

$$\cos^n z = \left(\frac{e^{iz}+e^{-iz}}{2}\right)^n = \frac{1}{2^n}\left\{e^{inz}+\binom{n}{1}e^{i(n-2)z}+\binom{n}{2}e^{i(n-4)z}+\cdots+e^{-inz}\right\}$$

$$= \frac{1}{2^{n-1}}\left\{\frac{e^{inz}+e^{-inz}}{2}+\binom{n}{1}\frac{e^{i(n-2)z}+e^{-i(n-2)z}}{2}+\binom{n}{2}\frac{e^{i(n-4)z}+e^{-i(n-4)z}}{2}+\cdots\right\}$$

$$= \frac{1}{2^{n-1}}\left\{\cos nz+\binom{n}{1}\cos(n-2)z+\binom{n}{2}\cos(n-4)z+\cdots\right\}. \qquad (25)$$

Das letzte Glied in der Klammer ist dabei gleich

$$\frac{1}{2}\binom{n}{n/2}, \qquad \text{falls } n \text{ gerade ist.}$$

$$\binom{n}{(n-1)/2}\cos z, \qquad \text{falls } n \text{ ungerade ist.}$$

4. Als nächstes soll umgekehrt $\cos nz$ bzw. $\sin nz$ durch $\cos z$ und $\sin z$ ausgedrückt werden. Dies gelingt etwa im Fall des Kosinus unmittelbar auf Grund von

$$\cos nz = \frac{e^{inz}+e^{-inz}}{2} = \frac{1}{2}\{(e^{iz})^n+(e^{-iz})^n\}$$

$$= \frac{1}{2}\{(\cos z+i\sin z)^n+(\cos z-i\sin z)^n\},$$

wobei man evtl. die Klammern noch nach dem binomischen Satz auflösen kann.[1])

5. Man berechne die Summe

$$S_n = \frac{1}{2}+\cos z+\cos 2z+\cdots+\cos nz.$$

Durch Einführung der Exponentialfunktion wird diese Summe zu einer geometrischen Summe, die man nach der bekannten Formel summieren kann:

$$S_n = \frac{1}{2}+\frac{e^{iz}+e^{-iz}}{2}+\frac{e^{2iz}+e^{-2iz}}{2}+\cdots+\frac{e^{niz}+e^{-niz}}{2}$$

$$= \frac{1}{2}(e^{-inz}+\cdots+e^{-iz}+1+e^{iz}+\cdots+e^{inz})$$

$$= \frac{1}{2}\,\frac{e^{-inz}-e^{i(n+1)z}}{1-e^{iz}} = \frac{1}{2}\,\frac{\sin\left(n+\frac{1}{2}\right)z}{\sin\frac{z}{2}}.$$ [2])

[1]) Da $\cos nz$ der Realteil von e^{inz}, d. h. von $(e^{iz})^n$, also von $(\cos z+i\sin z)^n$ ist, kann man auch so vorgehen, daß man $(\cos z+i\sin z)^n$ auflöst und vom Ergebnis den Realteil nimmt. Dieser Weg stimmt im wesentlichen mit dem angegebenen überein.

[2]) Man kann diese Aufgabe auch dadurch lösen, daß man die Summe $\frac{1}{2}+e^{iz}+e^{2iz}+\ldots+e^{niz}$ berechnet und vom Ergebnis den Realteil bestimmt.

Die Funktion $\operatorname{tg} z$ definiert man wie im Reellen als Quotient aus $\sin z$ und $\cos z$, so daß also

$$\operatorname{tg} z = \frac{\sin z}{\cos z} = \frac{1}{i}\,\frac{e^{iz}-e^{-iz}}{e^{iz}+e^{-iz}}. \tag{26}$$

Bekanntlich kann man jede komplexe Zahl $z = x + iy\,(\neq 0)$ auch in der „trigonometrischen Form"

$$z = r(\cos\theta + i\sin\theta),$$

darstellen, wobei $r = |z|$, $\theta = \arg z$, so daß also die Zahlen $r\,(>0)$ und θ mit den Zahlen x und y durch die Beziehungen

$$r\cos\theta = x,$$
$$r\sin\theta = y$$

zusammenhängen. Die Zahl r ist dabei eindeutig bestimmt, während für θ unendlich viele Werte möglich sind, die sich jedoch paarweise nur um ein ganzzahliges Vielfaches von 2π unterscheiden.

Auf Grund der Eulerschen Identität ist klar, daß man die trigonometrische Darstellung einer komplexen Zahl auch kurz folgendermaßen in „exponentieller Form" schreiben kann:

$$z = re^{i\theta}. \tag{27}$$

Aufgaben.

1. Man schreibe

$$\cos i,\quad \cos(i\lg 2),\quad \sin\frac{1+i}{\sqrt{2}},\quad \operatorname{tg} i$$

in anderer Form.

2. Man zerlege die folgenden Funktionen in Real- und Imaginärteil:

$$\sin z,\quad \cos z,\quad \operatorname{tg} z.$$

3. Man kontrolliere die folgenden Gleichungen, durch welche die gegebenen komplexen Zahlen durch ihren Betrag und ihr Argument ausgedrückt werden:

$$i = e^{i\frac{\pi}{2}},\quad (-1) = e^{i\pi},\quad 1+i = \sqrt{2}\,e^{i\frac{\pi}{4}},$$

$$1+i\sqrt{3} = 2e^{i\frac{\pi}{3}},\quad 1-i\sqrt{3} = 2e^{-i\frac{\pi}{3}}.$$

§ 5. Die hyperbolischen und die trigonometrischen Funktionen

Bekanntlich (vgl. S. 78) bezeichnet man im Reellen die Funktionen

$$\left.\begin{aligned} \mathfrak{Cof}\, z &= \frac{e^{z}+e^{-z}}{2},\\ \mathfrak{Sin}\, z &= \frac{e^{z}-e^{-z}}{2} \end{aligned}\right\} \tag{28}$$

als *hyperbolischen Kosinus* bzw. *hyperbolischen Sinus*. Es ist bekannt, welche Analogien zwischen den Eigenschaften der trigonometrischen und der hyperbolischen Funktionen bestehen.

Indem wir nun die Definition (28) der hyperbolischen Funktionen in der angegebenen Form ins Komplexe übernehmen, können wir auch den tieferen Grund für diese formalen Analogien aufzeigen. Aus den Formeln (21) und (28) folgt nämlich sofort

$$\left.\begin{aligned} \cos z &= \mathfrak{Cof}(iz), \\ i \sin z &= \mathfrak{Sin}(iz) \end{aligned}\right\} \tag{29}$$

oder auch[1])

$$\left.\begin{aligned} \mathfrak{Cof}\, z &= \cos(iz), \\ i\, \mathfrak{Sin}\, z &= \sin(iz). \end{aligned}\right\} \tag{30}$$

Die hyperbolischen Funktionen lassen sich also sehr einfach durch die trigonometrischen Funktionen ausdrücken und umgekehrt, woraus sich sofort die genannten Analogien ergeben. Man kann nämlich aus jeder Identität in den trigonometrischen Funktionen eine entsprechende Identität in den hyperbolischen Funktionen herleiten, wenn man in den trigonometrischen Funktionen die Argumente mit dem Faktor i versieht.

So ergibt sich z. B. aus der Identität

$$\cos^2 z + \sin^2 z = 1,$$

wenn man dort z durch iz ersetzt, die Identität

$$\cos^2(iz) + \sin^2(iz) = 1,$$

d. h.

$$\mathfrak{Cof}^2 z - \mathfrak{Sin}^2 z = 1. \tag{31}$$

Entsprechend folgt aus der Identität

$$\cos(z_1 + z_2) = \cos z_1 \cos z_2 - \sin z_1 \sin z_2$$

die Identität

$$\cos(iz_1 + iz_2) = \cos(iz_1)\cos(iz_2) - \sin(iz_1)\sin(iz_2),$$

d. h.

$$\mathfrak{Cof}(z_1 + z_2) = \mathfrak{Cof}\, z_1\, \mathfrak{Cof}\, z_2 + \mathfrak{Sin}\, z_1\, \mathfrak{Sin}\, z_2. \tag{32}$$

Aus den Formeln (29) bzw. (30) folgt ferner, daß *die Funktion* $\mathfrak{Cof}\, z$ *für rein imaginäre* z*-Werte dieselben Werte wie die Funktion* $\cos z$ *für reelle* z*-Werte annimmt* (*und umgekehrt*), *während sich die Werte der Funktionen* $\mathfrak{Sin}\, z$ *und* $\sin z$ *dabei nur um den Faktor* i *unterscheiden*.

Die hyperbolischen Funktionen sind, ähnlich wie die trigonometrischen Funktionen, *periodisch*, jedoch ist ihre Periode nicht (wie bei den trigonometrischen Funktionen) reell, sondern rein imaginär (wie bei der Funktion e^z), und zwar ist sie gleich $2\pi i$.

[1]) Man kann die Formeln (30) auch unmittelbar dadurch aus den Formeln (29) erhalten, daß man dort z durch $-iz$ ersetzt.

Aufgaben.

1. Man zeige, daß für $z = x + iy$ die folgenden Formeln gelten:

$$\cos z = \cos x \operatorname{Cof} y - i \sin x \operatorname{Sin} y,$$
$$\sin z = \sin x \operatorname{Cof} y + i \cos x \operatorname{Sin} y$$

und

$$\operatorname{Cof} z = \operatorname{Cof} x \cos y + i \operatorname{Sin} x \sin y,$$
$$\operatorname{Sin} z = \operatorname{Sin} x \cos y + i \operatorname{Cof} x \sin y.$$

2. Man zeige, daß auch im Komplexen die Zahlen der Form $n\pi$ und nur diese Nullstellen der Funktion $\sin z$ sind.

3. Man bestimme sämtliche Lösungen der Gleichung

$$\operatorname{Cof} z = 0.$$

§ 6. Der Logarithmus

Die *Logarithmusfunktion*[1])

$$w = \operatorname{Ln} z$$

kann man im Komplexen, entsprechend wie im Reellen, als Umkehrfunktion der Funktion $w = e^z$ definieren. Um also festzustellen, welchen Wert die Funktion $w = \operatorname{Ln} z$ für einen gegebenen Argumentwert z annimmt, muß man die Gleichung

$$e^w = z \tag{33}$$

nach w auflösen. Setzen wir $z = re^{i\theta}$, $w = u + iv$, so können wir (33) auch in der Form

$$e^{u+iv} = re^{i\theta} \tag{34}$$

schreiben. Bekanntlich sind nun komplexe Zahlen dann und nur dann gleich, wenn sie denselben Betrag besitzen und ihre Argumente sich höchstens um ein ganzzahliges Vielfaches von 2π unterscheiden.

Die eine komplexe Gleichung (34) ist also den beiden reellen Gleichungen

$$\left.\begin{aligned} e^u &= r, \\ v &= \theta + 2k\pi \quad (k \text{ beliebige ganze Zahl}) \end{aligned}\right\} \tag{35}$$

gleichwertig. Hieraus folgt jedoch unmittelbar, daß (für $z \neq 0$) $u = \ln r$, d. h.

$$\operatorname{Ln} z = w = u + iv = \ln r + i(\theta + 2k\pi) \tag{36}$$

oder

$$\operatorname{Ln} z = \ln |z| + i(\arg z + 2k\pi). \tag{37}$$

[1]) Im Unterschied zum gewöhnlichen natürlichen Logarithmus bezeichnen wir den natürlichen Logarithmus im Komplexen (mit dem wir uns hier nur beschäftigen wollen) mit $\operatorname{Ln} z$.

Als Logarithmus einer beliebigen von Null verschiedenen komplexen Zahl ergibt sich also jede Zahl, deren Realteil gleich dem gewöhnlichen Logarithmus des Betrages dieser Zahl und deren Imaginärteil gleich einem ihrer Argumente ist. Hieraus folgt insbesondere, daß *jede von Null verschiedene komplexe Zahl unendlich viele Logarithmen besitzt, wobei sich je zwei verschiedene Werte um ein ganzzahliges Vielfaches von $2\pi i$ unterscheiden.* Bezeichnet man also einen beliebigen Wert des Logarithmus von z mit w_0, so erhält man in

$$\ldots, w_0 - 2n\pi i, \ldots, w_0 - 2\pi i, w_0, w_0 + 2\pi i, \ldots, w_0 + 2n\pi i, \ldots$$

sämtliche Werte von $\operatorname{Ln} z$.

Insbesondere gilt:

1. Ist z eine positive reelle Zahl, so kann man als w_0 den gewöhnlichen (reellen) Logarithmus nehmen, und alle übrigen Werte des Logarithmus sind komplex.

2. Ist z eine negative reelle Zahl oder eine echt komplexe Zahl, so treten unter ihren Argumenten keine ganzzahligen Vielfachen von 2π auf, so daß in diesem Fall sämtliche Werte des Logarithmus echt komplex sind.

Wir weisen schließlich noch darauf hin, daß $\operatorname{Ln} z$ für $z = 0$ nicht erklärt ist, da die Gleichung $e^w = r$ für $r = 0$ keine Lösung besitzt.

Wir sehen also, daß der Begriff des Logarithmus bei seiner Ausdehnung ins Komplexe eine wesentliche Änderung (Verallgemeinerung) erfährt. Man kann sagen, daß die Funktion $w = \operatorname{Ln} z$ in jedem Gebiet, welches nicht den Ursprung $z = 0$ enthält[1]), *unendlich vieldeutig* ist. Beschränkt man sich auf die Werte der Funktion $w = \operatorname{Ln} z$, die sich aus (37) für einen festen Wert k ergeben, so erhält man eine eindeutige Funktion, die man als einen *Zweig* des Logarithmus bezeichnet.

Aufgaben.

1. Man drücke $\operatorname{Ln} z$ durch Real- und Imaginärteil von z aus.
2. Man berechne sämtliche Werte des Logarithmus der Zahlen

$$1, \quad -1, \quad i, \quad -i, \quad 1+i, \quad 1+i\sqrt{3}, \quad 3+4i, \quad 5-6i.$$

§ 7. Potenzen mit beliebigem Exponenten

Die allgemeine *Potenzfunktion* $w = z^\alpha$, wobei α eine beliebige komplexe Zahl ist, definiert man im Komplexen entsprechend wie im Reellen durch

$$z^\alpha = (e^{\operatorname{Ln} z})^\alpha = e^{\alpha \operatorname{Ln} z} \qquad (z \neq 0). \tag{38}$$

Der wesentliche Unterschied gegenüber dem Reellen besteht hier jedoch darin, daß dabei der Logarithmus im verallgemeinerten Sinn, d. h. unendlich vieldeutig, zu verstehen ist. Daher wird also auch die Funktion $w = z^\alpha$ im

[1]) Für $z = 0$ ist die Funktion $w = \operatorname{Ln} z$ nicht erklärt.

allgemeinen unendlich vieldeutig sein. Bezeichnen wir mit $\ln z$ irgendeinen der Werte von $\operatorname{Ln} z$, so ist

$$\operatorname{Ln} z = \ln z + 2k\pi i$$

und daher nach Formel (38)

$$w = z^\alpha = e^{\alpha(\ln z + 2k\pi i)} = (e^{\ln z})^\alpha e^{2k\pi i\alpha} = w_0 e^{2k\pi i\alpha},$$

wobei w_0 einer der Werte von z^α ist. Daher erhalten wir in

$$\ldots,\ w_0 e^{-2n\pi i\alpha},\ \ldots,\ w_0 e^{-2\pi i\alpha},\ w_0,\ w_0 e^{2\pi i\alpha},\ \ldots,\ w_0 e^{2n\pi i\alpha},\ \ldots \tag{39}$$

sämtliche Werte von $w = z^\alpha$.

Hieraus folgt speziell:

1. Ist α eine ganze Zahl, so ist $e^{2\pi i\alpha} = e^{2\pi in} = 1$, d. h., sämtliche Zahlen (39) stimmen überein: *Die Potenzfunktionen mit ganzzahligem Exponenten sind* (wie zu erwarten war) *eindeutig.*

2. Ist α eine gebrochene rationale Zahl, $\alpha = \frac{p}{q}$ (reduzierter Bruch), so gibt es unter den Zahlen (39) genau q verschiedene. Man erhält sie z. B., wenn man in $e^{2k\pi i\frac{p}{q}}$ für k die Zahlen $0, 1, \ldots, q-1$ nimmt; für $k = q$ ergibt sich dann

$$e^{2k\pi i\frac{p}{q}} = e^{2\pi ip} = 1,$$

d. h., die Werte beginnen sich zu wiederholen. *Die Funktion* $w = z^{\frac{p}{q}} = \sqrt[q]{z^p}$ *ist also „q-wertig".* Die zu einem Wert z gehörenden q Werte von w sind, wie man sofort sieht, die Lösungen der Gleichung $w^q = z^p$.

3. Ist α eine irrationale reelle Zahl, so sind die Werte (39) paarweise verschieden. Sie haben jedoch alle denselben Betrag.

4. Ist α eine rein imaginäre Zahl, $\alpha = i\gamma$, wobei γ eine von Null verschiedene reelle Zahl ist, so ist

$$e^{2\pi i\alpha} = e^{-2\pi\gamma}$$

eine positive reelle Zahl. Daraus folgt, daß in diesem Fall die Werte (39) paarweise verschieden sind, und zwar genauer, daß sie bei paarweise verschiedenen Beträgen jeweils dasselbe Argument besitzen.

5. Ist schließlich $\alpha = \beta + i\gamma$ mit $\beta \neq 0$ und $\gamma \neq 0$, so sind die Werte (39) paarweise verschieden, und zwar sowohl dem Betrage als auch dem Argument nach.

Aufgaben.

1. Man zeige, daß $i^i = e^{-\frac{\pi}{2}} e^{2k\pi}$.[1])

2. Man zeige, daß $(-1)^{\sqrt{2}} = \cos(2k+1)\pi\sqrt{2} + i\sin(2k+1)\pi\sqrt{2}$.

3. Man zeige, daß $(-1)^i = e^{(2k+1)\pi}$.

[1]) Hier und im folgenden soll k stets eine beliebige ganze Zahl sein.

4. Man zeige, daß

$$(1+i)^{1+i} = \sqrt{2}\,e^{-\left(\frac{\pi}{4}+2k\pi\right)}\left\{\cos\left(\ln\sqrt{2}+\frac{\pi}{4}\right)+i\sin\left(\ln\sqrt{2}+\frac{\pi}{4}\right)\right\}.$$

5. Man zerlege für $z = x + iy = re^{i\theta}$ die Funktion $w = z^i$ in Real- und Imaginärteil (Resultat: $|w| = e^{-(\theta+2k\pi)}$, $\arg w = \ln r$).

6. Wieviel verschiedene Werte besitzt $(-1)^z$ für a) ganzes z, b) gebrochen rationales z, c) irrationales reelles z, d) rein imaginäres z? Man gebe Beispiele an.

§ 8. Die Umkehrfunktionen der trigonometrischen und der hyperbolischen Funktionen

Da sich die trigonometrischen und die hyperbolischen Funktionen in einfacher Weise mittels der Exponentialfunktion ausdrücken lassen, dürfte es jetzt nicht verwunderlich sein, wenn sich die Umkehrfunktionen der trigonometrischen und der hyperbolischen Funktionen leicht durch die Umkehrfunktion der Exponentialfunktion, also den Logarithmus ausdrücken lassen. Das ist nun tatsächlich der Fall.

Wir betrachten als erstes den *Arkussinus*.

Unter

$$w = \operatorname{Arcsin} z$$

wollen wir die Lösung (oder besser gesagt, die Gesamtheit der Lösungen) der Gleichung

$$\sin w = z$$

verstehen, die wir auch in der Form

$$\frac{e^{iw}-e^{-iw}}{2i} = z$$

oder

$$e^{2iw} - 2iz e^{iw} - 1 = 0$$

schreiben können. Die zuletzt erhaltene Gleichung können wir nun aber offenbar als eine quadratische Gleichung in e^{iw} ansehen, so daß also

$$e^{iw} = iz + \sqrt{1-z^2}.^{1})$$

Das besagt aber nichts anderes, als daß iw einer der Werte des Logarithmus von $iz + \sqrt{1-z^2}$ sein muß:

$$iw = \operatorname{Ln}\left(iz + \sqrt{1-z^2}\right).$$

---

[1]) Die quadratische Gleichung $z^2 + pz + q = 0$ kann man im Komplexen auf dieselbe Weise lösen wie im Reellen, und zwar z. B. durch „quadratische Ergänzung“: $z^2 + pz + \frac{p^2}{4} = \frac{p^2}{4} - q$, $\left(z+\frac{p}{2}\right)^2 = \frac{p^2}{4} - q$, $z + \frac{p}{2} = \sqrt{\frac{p^2}{4} - q}$ usw. Die Wurzel ist dabei doppeldeutig (die Operation des Radizierens haben wir bereits in § 7 eingeführt).

Damit erhalten wir

$$w \equiv \operatorname{Arcsin} z = -i \operatorname{Ln}(iz + \sqrt{1-z^2}). \tag{40}$$

Durch eine ähnliche Überlegung findet man

$$\operatorname{Arccos} z = -i \operatorname{Ln}(z + i\sqrt{1-z^2}). \tag{41}$$

Schließlich ergibt sich auf Grund von (26)

$$\operatorname{Arctg} z = \frac{1}{2i} \operatorname{Ln} \frac{1+iz}{1-iz}. \tag{42}$$

Die Formeln (40) bis (42) zeigen, daß *die Umkehrfunktionen der trigonometrischen Funktionen sämtlich unendlich vieldeutig sind.* Man darf dabei auch nicht übersehen, daß die Wurzeln in (40) und (41) doppeldeutig sind.

Um die Umkehrfunktion des hyperbolischen Kosinus zu erhalten, muß man die Gleichung

$$\mathfrak{Co}\, w = z$$

lösen, die man auf Grund von (30) auch in der Form

$$\cos(iw) = z$$

schreiben kann. Hieraus folgt

$$iw = \operatorname{Arccos} z,$$

so daß nach Formel (41)

$$w = -i \operatorname{Arccos} z = -\operatorname{Ln}(z + i\sqrt{1-z^2})$$

oder

$$\operatorname{Ar}\mathfrak{Co}\, z = \operatorname{Ln}(z + \sqrt{z^2-1}).^{1}) \tag{43}$$

Entsprechend findet man

$$\operatorname{Ar}\mathfrak{Sin}\, z = -i \operatorname{Arcsin}(iz) = \operatorname{Ln}(z + \sqrt{z^2+1}), \tag{44}$$

$$\operatorname{Ar}\mathfrak{Tg}\, z = -i \operatorname{Arctg}(iz) = \frac{1}{2} \operatorname{Ln} \frac{1+z}{1-z}. \tag{45}$$

Die Formeln (43) bis (45) kann man übrigens auch ohne Verwendung des Zusammenhanges zwischen den hyperbolischen und den trigonometrischen Funktionen direkt herleiten, indem man unmittelbar von den Darstellungen der hyperbolischen Funktionen mittels der Exponentialfunktion ausgeht.

Die hier angegebenen Formeln haben einen allgemeineren Inhalt als die auf Seite 83 hergeleiteten Formeln, und zwar deshalb, weil 1. die unabhängige Veränderliche z hier beliebige komplexe Werte annehmen kann,

---

[1]) In der Tat ist

$$-\operatorname{Ln}(z + i\sqrt{1-z^2}) = -\operatorname{Ln}(z + \sqrt{z^2-1}) = \operatorname{Ln} \frac{1}{z + \sqrt{z^2-1}} = \operatorname{Ln}(z - \sqrt{z^2-1}).$$

Welches Vorzeichen wir der Wurzel geben, ist gleichgültig, da diese sowieso doppeldeutig ist.

2. der Logarithmus im „komplexen" Sinn gemeint ist (vgl. § 6), 3. die Wurzeln in (43) und (44) doppeldeutig sind. Auf Grund dessen besitzt z. B. die Funktion $\mathfrak{Ar}\,\mathfrak{Sin}\,z$ für reelles $z\,(z=x)$ neben den Werten der Form $\ln\left(x+\sqrt{x^2+1}\right)+2k\pi i$ auch noch die Werte

$$(2k+1)\,\pi i-\ln\left(x+\sqrt{x^2+1}\right).$$

Aufgaben.

1. Man berechne, ausgehend von den Formeln (43) bis (45), die Werte

$$\operatorname{Arcsin}0,\quad \operatorname{Arccos}0,\quad \operatorname{Arctg}0,\quad \operatorname{Arcsin}1,\quad \operatorname{Arccos}1,\quad \lim_{z\to\infty}\operatorname{Arctg}z.$$

2. Setzt man z als reell voraus $(z=x)$, so nimmt die rechte Seite von Formel (42) die Gestalt

$$\frac{1}{2i}\operatorname{Ln}\frac{1+ix}{1-ix}=\frac{1}{2i}\left\{\ln\left|\frac{1+ix}{1-ix}\right|+i\arg\frac{1+ix}{1-ix}\right\}$$

an. Man schreibe in entsprechender Weise $\operatorname{Arcsin}x$ und $\operatorname{Arccos}x$ (unter der Voraussetzung $|x|<1$).

3. Man zeige durch Elimination der trigonometrischen und der zyklometrischen Funktion gemäß (21) bzw. (41), daß sich das TSCHEBYSCHEWsche Polynom

$$T_n(z)=\cos(n\operatorname{Arccos}z)$$

auch in der Form

$$T_n(z)=\frac{1}{2}\left[(z+\sqrt{z^2-1})^n+(z-\sqrt{z^2-1})^n\right]$$

schreiben läßt.

§ 9. Differentiation im Komplexen

Gegeben sei eine eindeutige Funktion $w=f(z)$ einer komplexen Veränderlichen, die in einer gewissen Umgebung des Punktes z erklärt sei. Man sagt, die Funktion $f(z)$ sei *im Punkte z differenzierbar* (bzw. *ableitbar*), wenn für den betrachteten Punkt z der Grenzwert

$$\lim_{\Delta z\to 0}\frac{f(z+\Delta z)-f(z)}{\Delta z} \tag{46}$$

existiert und endlich ist. Diesen Grenzwert nennt man dann die *Ableitung* der Funktion $f(z)$ im Punkte z und bezeichnet ihn mit $f'(z)$. Setzen wir

$$\Delta z=h,\quad \Delta w=\Delta f(z)=f(z+h)-f(z),$$

so können wir (46) auch in der Form

$$f'(z)=\lim_{h\to 0}\frac{f(z+h)-f(z)}{h}=\lim_{\Delta z\to 0}\frac{\Delta f(z)}{\Delta z}=\lim_{\Delta z\to 0}\frac{\Delta w}{\Delta z}=\frac{dw}{dz} \tag{46'}$$

schreiben. Der Quotient $\dfrac{f(z+h)-f(z)}{h}$ ist dabei eine Funktion von h, so daß es sich bei dem zu bildenden Grenzwert um den Grenzwert einer

Funktion handelt. Dieser Begriff wird dabei im Komplexen rein formal genauso wie im Reellen definiert. Die Schreibweise

$$\lim_{z \to z_0} F(z) = L$$

oder

$$F(z) \to L \quad \text{für} \quad z \to z_0$$

bedeutet also, daß es zu jeder noch so kleinen positiven reellen Zahl ε eine reelle Zahl $\delta > 0$ gibt, so daß die Ungleichung $|z - z_0| < \delta$ stets die Ungleichung $|F(z) - L| < \varepsilon$ zur Folge hat (vgl. S. 175), bzw. daß für jede Folge $\{z_n\}$, für welche $z_n \to z_0$ gilt, stets die Beziehung $F(z_n) \to L$ erfüllt ist.

In diesem Sinne ist auch Formel (46) zu verstehen.

Im Reellen bedeutet nun die Ungleichung $|h| < \delta$, daß der Punkt $x + h$ im Innern des Intervalls von $x - \delta$ bis $x + \delta$ liegt. Dagegen besagt dieselbe Ungleichung im Komplexen, daß der Punkt $z + h$ im Innern des Kreises um z vom Radius δ zu finden ist. Hieraus ist unmittelbar zu verstehen, daß die Forderung der Differenzierbarkeit im Komplexen eine wesentlich einschränkendere Forderung als im Reellen ist. Während nämlich bei der Bildung des Grenzwertes (46) im Reellen nur Annäherungen von $z + h$ an z auf der reellen Achse zu betrachten sind, müssen im Komplexen alle Annäherungen von $z + h$ an z in einem gewissen Kreis um z betrachtet werden. Ist also $f(z)$ eine Funktion, die in einer gewissen Umgebung des reellen Punktes $z = x$ definiert ist und für reelle Argumentwerte auch stets reelle Funktionswerte annimmt, so folgt aus der Differenzierbarkeit der Funktion $f(z)$ im Punkte x im Sinne der reellen Analysis im allgemeinen noch nicht die Differenzierbarkeit der Funktion $f(z)$ im Punkte x im Sinne der Theorie der Funktionen einer komplexen Veränderlichen.

Es zeigt sich jedoch, daß *die elementaren Funktionen*[1]) *einer komplexen Veränderlichen überall differenzierbar sind, einige* („*singuläre*") *Punkte evtl. ausgeschlossen.* Hierbei gelten im Komplexen überdies dieselben Differentiationsregeln wie im Reellen.

Letzteres erklärt sich daraus, daß der formale Apparat, der bei der Herleitung der Differentiationsregeln verwendet wird, beim Übergang ins Komplexe unverändert erhalten bleibt (wenn sich auch dabei der Inhalt der Formeln zum Teil ändert).

Wir wollen als ein einfaches Beispiel die Regel für die Differentiation der Funktion $f(z) = z^2$ herleiten. Dazu beachten wir, daß im Komplexen wie im Reellen für beliebiges von Null verschiedenes h die Formel

$$\frac{f(z+h) - f(z)}{h} = \frac{(z+h)^2 - z^2}{h} = 2z + h$$

gilt. Lassen wir hier nun die komplexe Veränderliche h auf irgendeine Weise gegen Null streben, so strebt der betrachtete Ausdruck stets gegen den Grenzwert $2z$. Daher gilt

$$(z^2)' = 2z.$$

[1]) Zur Definition der elementaren Funktion vgl. Seite 3 und 5.

Allgemein basieren die bekannten Differentiationsregeln auf den Sätzen über Grenzwerte, und diese beruhen ihrerseits auf den Eigenschaften des absoluten Betrages. Der absolute Betrag besitzt aber im Komplexen im wesentlichen dieselben Eigenschaften wie im Reellen.

Hieraus folgt, daß die bekannten Regeln für die Differentiation einer Summe, einer Differenz, eines Produktes und eines Quotienten mit ihren Folgerungen auch im Komplexen gelten.

Es gilt aber auch die Regel für die Differentiation einer zusammengesetzten (mittelbaren) Funktion,

$$[f(\varphi(z))]' = f'(\varphi(z))\,\varphi'(z),$$

und die bekannte Regel über die Differentiation der Umkehrfunktion (auf die Frage nach der Aussonderung der hierbei evtl. auftretenden „singulären" Punkte wollen wir hier nicht eingehen).

Wenn wir jetzt zur Behandlung der Differentiationsregeln für die Exponentialfunktion und die trigonometrischen Funktionen übergehen, müssen wir beachten, daß wir für diese Funktionen in § 3 eine ganz bestimmte Definition gegeben haben. Wir sind daher gezwungen, diese Differentiationsregeln im einzelnen zu beweisen.

Aus dem Additionstheorem für die Funktion $f(z) = e^z$ folgt zunächst

$$\frac{f(z+h)-f(z)}{h} = \frac{e^{z+h}-e^z}{h} = e^z \cdot \frac{e^h-1}{h}.$$

Es genügt daher zu zeigen, daß $\lim\limits_{h\to 0} \frac{e^h-1}{h} = 1$, d. h.,

$$\lim_{z\to 0} \frac{e^z-1}{z} = 1. \tag{47}$$

Dies ergibt sich jedoch sofort auf Grund der Ungleichung

$$\left|\frac{e^z-1}{z}-1\right| = \left|\frac{\left(1+\frac{z}{1!}+\frac{z^2}{2!}+\cdots\right)-1}{z}-1\right|$$

$$= \left|\frac{z}{2!}+\frac{z^2}{3!}+\frac{z^3}{4!}+\cdots\right| \leqq r\left(\frac{1}{2!}+\frac{r}{3!}+\frac{r^2}{4!}+\cdots\right),$$

wobei die Summe der Reihe auf der rechten Seite für $r<1$ kleiner als

$$\frac{1}{2!}+\frac{1}{3!}+\frac{1}{4!}+\cdots = e-2$$

ist, so daß die linke Seite dieser Ungleichung in der Tat für $z\to 0$ gegen Null strebt. Es ist also auch im Komplexen

$$(e^z)' = e^z. \tag{48}$$

Der bekannte Beweis für die Differentiationsregeln für den Sinus und den Kosinus im Reellen gründet sich auf die Beziehung

$$\lim_{z\to 0} \frac{\sin z}{z} = 1, \tag{49}$$

die wir jetzt ins Komplexe übertragen wollen. Dazu beachten wir, daß

$$\left|\frac{\sin z}{z}-1\right|=\left|\frac{1}{z}\left(\frac{z}{1!}-\frac{z^3}{3!}+\frac{z^5}{5!}-\cdots\right)-1\right|\leqq\frac{r^3}{3!}+\frac{r^4}{5!}+\cdots$$

$$=r^2\left(\frac{1}{3!}+\frac{r^2}{5!}+\cdots\right),$$

woraus alles Weitere zu ersehen ist.[1])

Man kann sich übrigens ganz allgemein davon überzeugen, daß eine absolut konvergente Reihe auch im Komplexen gliedweise differenziert werden kann (und dann die Ableitung ihrer Summe liefert).[2])

Unter Verwendung dieses Satzes ergibt sich ebenfalls unmittelbar:

$$(e^z)'=\left(1+\frac{z}{1!}+\frac{z^2}{2!}+\frac{z^3}{3!}+\cdots\right)'=\left(\frac{z}{1!}\right)'+\left(\frac{z^2}{2!}\right)'+\left(\frac{z^3}{3!}\right)'+\cdots$$

$$=1+\frac{z}{1!}+\frac{z^2}{2!}+\cdots=e^z \tag{50}$$

und analog

$$(\cos z)'=\left(1-\frac{z^2}{2!}+\frac{z^4}{4!}-\frac{z^6}{6!}+\cdots\right)'=-\frac{z}{1!}+\frac{z^3}{3!}-\frac{z^5}{5!}+\cdots=-\sin z, \tag{51}$$

$$(\sin z)'=\left(\frac{z}{1!}-\frac{z^3}{3!}+\frac{z^5}{5!}-\frac{z^7}{7!}+\cdots\right)'=1-\frac{z^2}{2!}+\frac{z^4}{4!}-\cdots=\cos z. \tag{52}$$

Zur Herleitung der Differentiationsregel für den Logarithmus[3]) $\operatorname{Ln} z$ empfiehlt es sich, den Logarithmus als Umkehrfunktion der Exponentialfunktion abzuleiten:

$$e^w=z,\quad e^w w'=1,\quad w'=e^{-w},\quad \text{d. h.}\quad (\operatorname{Ln} z)'=\frac{1}{z}. \tag{53}$$

Entsprechend kann man auch die Differentiationsregeln für die Umkehrfunktionen der trigonometrischen Funktionen erhalten.

Wir haben bisher noch nichts bezüglich des Begriffs der *Stetigkeit* bei Funktionen einer komplexen Veränderlichen gesagt. Hierzu genügt es, kurz zu bemerken, daß dieser Begriff im Komplexen genau wörtlich wie im Reellen definiert wird (und zwar zunächst als Stetigkeit in einem Punkte, sodann als Stetigkeit in einem Gebiet), wobei nur überall der absolute Betrag der entsprechenden komplexen Zahlen zu nehmen ist.

Man beweist dann genau wie im Reellen (vgl. S. 293), daß *jede* (in einem Punkte bzw. einem Gebiet) *differenzierbare Funktion* (in diesem Punkt bzw. in diesem Gebiet) *auch stetig ist.*

---

[1]) Man kann die Differentiationsregeln für den Sinus und den Kosinus auch unter Verwendung von (48) durch Differentiation der Identitäten (21) erhalten. — *Anm. d. wissenschaftl. Red.*

[2]) Die Sätze 2 und 3 von Seite 435 und 436 gelten auch im Komplexen, und zwar läßt sich im wesentlichen auch der dort angegebene Beweis übertragen.

[3]) Wir beschränken uns im folgenden auf einen bestimmten (im übrigen willkürlich gewählten) Zweig des Logarithmus.

§ 10. Das Kurvenintegral

Die sinngemäße Übertragung des bestimmten Integrals ins Komplexe führt auf den Begriff des sogenannten *Kurvenintegrals*, dem wir uns jetzt zuwenden wollen.

Es sei (L) eine beliebige Kurve[1]) in der komplexen z-Ebene, die von einem gewissen Anfangspunkt a zu einem gewissen Endpunkt b führen möge (Abb. 1), und $w = f(z)$ eine Funktion, welche in allen Punkten von (L) definiert ist. Es seien ferner z_k $(k = 1, 2, \ldots, n-1)$ irgendwelche Punkte dieser Kurve, die entsprechend der „Orientierung" der Kurve geordnet seien, wobei wir gleichzeitig noch $z_0 = a$ und $z_n = b$ setzen. Wir führen schließlich noch folgende Bezeichnung ein:

$$\Delta z_k = z_k - z_{k-1} \quad (k = 1, 2, \ldots, n).$$

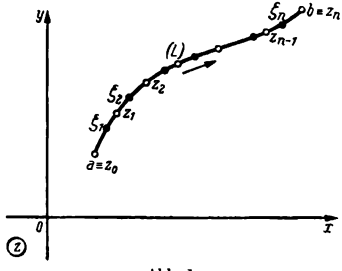


Abb. 1

Wir wählen nun auf jedem Bogen $z_k z_{k+1}$ einen gewissen Punkt ζ_k (der auch mit einem der Endpunkte z_k bzw. z_{k+1} zusammenfallen kann) und bilden die „Integralsumme"

$$S = \sum_{k=1}^{n} f(\zeta_k)\,(z_k - z_{k-1}) = \sum_{k=1}^{n} f(\zeta_k)\,\Delta z_k. \tag{54}$$

Mit δ bezeichnen wir das Maximum der Zahlen $|\Delta z_k|$ $(k = 1, 2, \ldots, n)$, also die Länge der größten Seite des Streckenzuges mit den Eckpunkten z_k.

Wir stellen uns vor, daß wir von der betrachteten Zerlegung der Kurve (L) durch die Punkte z_k $(k = 0, 1, 2, \ldots, n)$ durch Einschaltung weiterer Teilpunkte zu immer feineren Zerlegungen übergehen, wobei schließlich die Maximaldistanz δ der Teilpunkte gegen Null strebe. Es kann sich nun herausstellen, daß die Folge der zugehörigen Integralsummen dabei gegen einen endlichen Grenzwert I strebt, der von der Auswahl der Zerlegungsfolge und der Zwischenpunkte unabhängig ist. In diesem Fall nennt man den Grenzwert

$$I = \lim_{\delta \to 0} S$$

[1]) Unter einer „Kurve" verstehen wir hier und im folgenden eine Kurve, die in jedem Punkte eine Tangente besitzt bzw. aus endlich vielen Kurvenbögen zusammengesetzt ist, welche in jedem Punkte eine Tangente besitzen. Insbesondere kann also eine „Kurve" auch ein Streckenzug (Polygonzug) sein.

das *Kurvenintegral der Funktion $f(z)$ längs des Bogens (L)* und schreibt

$$I = \int_{(L)} f(z)\,dz, \tag{55}$$

so daß also definitionsgemäß

$$\int_{(L)} f(z)\,dz = \lim_{\delta \to 0} \sum_{k=1}^{n} f(\zeta_k)\,\Delta z_k. \tag{56}$$

Es zeigt sich nun, daß *der genannte Grenzwert sicher dann existiert, daß also die Funktion $f(z)$ längs (L) sicher integrierbar ist, wenn $f(z)$ auf (L) stetig ist.*

Wir wollen hier einen Beweis für diesen Satz angeben, bei dem das komplexe Kurvenintegral auf gewöhnliche reelle Integrale zurückgeführt wird. Dazu wollen wir uns der Einfachheit halber auf den Fall beschränken, daß jede Gerade, die parallel zur reellen oder imaginären Achse ist, die Kurve (L) in höchstens einem Punkte schneidet, so daß wir also die Kurve (L) sowohl durch eine Gleichung der Form $y = y(x)$ als auch durch eine Gleichung der Form $x = x(y)$ beschreiben können, wobei $y(x)$ und $x(y)$ stetig sind. Setzen wir

$$z_k = x_k + i y_k, \quad \Delta z_k = \Delta x_k + i\,\Delta y_k, \quad \zeta_k = \xi_k + i\,\eta_k,$$
$$f(z) \equiv u(x, y) + i\,v(x, y),$$

so können wir die Integralsumme S folgendermaßen schreiben:

$$\begin{aligned} S &= \sum_{k=1}^{n} \{u(\xi_k, \eta_k) + i\,v(\xi_k, \eta_k)\}\,(\Delta x_k + i\,\Delta y_k) \\ &= \sum_{k=1}^{n} \{u(\xi_k, \eta_k)\,\Delta x_k - v(\xi_k, \eta_k)\,\Delta y_k\} \\ &\quad + i \sum_{k=1}^{n} \{v(\xi_k, \eta_k)\,\Delta x_k + u(\xi_k, \eta_k)\,\Delta y_k\}. \end{aligned}$$

Berücksichtigen wir, daß $\eta_k = y(\xi_k)$, $\xi_k = x(\eta_k)$, so finden wir

$$\begin{aligned} S &= \left\{ \sum_{k=1}^{n} u(\xi_k, y(\xi_k))\,\Delta x_k - \sum_{k=1}^{n} v(x(\eta_k), \eta_k)\,\Delta y_k \right\} \\ &\quad + i \left\{ \sum_{k=1}^{n} v(\xi_k, y(\xi_k))\,\Delta x_k + \sum_{k=1}^{n} u(x(\eta_k), \eta_k)\,\Delta y_k \right\}. \end{aligned}$$

Hierbei ist jede der vier Summen eine Integralsumme einer stetigen[1]) Funktion, und zwar bzw. der Funktion

$$u(x, y(x)), \quad v(x(y), y), \quad v(x, y(x)), \quad u(x(y), y). \tag{57}$$

[1]) Es ist ohne weiteres klar, daß mit $f(z)$ auch die Funktionen $u(x, y)$ und $v(x, y)$ auf (L) stetig sind. Dann sind es aber auch (vgl. Seite 179) die Funktionen (57).

Daher strebt für $\delta \to 0$ jede dieser Summen gegen einen gewissen Grenzwert[1]), der gleich dem entsprechenden reellen Integral ist. Hieraus folgt jedoch unmittelbar, daß auch das (komplexe) Integral $I = \lim S$ existiert.

Man zeigt leicht, daß das Kurvenintegral die folgenden Eigenschaften besitzt:

I. $$\int_{(L)} (af(z) + bg(z))\,dz = a\int_{(L)} f(z)\,dz + b\int_{(L)} g(z)\,dz \tag{58}$$

(a und b sind dabei beliebige komplexe Konstanten).

II. *Stimmt der Anfangspunkt der Kurve* (L_2) *mit dem Endpunkt der Kurve* (L_1) *überein und ist* (L) *die Kurve, die aus den Kurven* (L_1) *und* (L_2) *„zusammengesetzt" ist, so gilt*

$$\int_{(L)} f(z)\,dz = \int_{(L_1)} f(z)\,dz + \int_{(L_2)} f(z)\,dz. \tag{59}$$

III. *Unterscheidet sich die Kurve* (L') *von der Kurve* (L) *nur durch ihre Orientierung* (d. h. handelt es sich um dieselbe Kurve, wobei nur b als Anfangs- und a als Endpunkt angesehen wird), *so ist*

$$\int_{(L')} f(z)\,dz = -\int_{(L)} f(z)\,dz. \tag{60}$$

IV. *Es ist stets*

$$\left|\int_{(L)} f(z)\,dz\right| \leqq LM, \tag{61}$$

wobei M *eine beliebige Zahl ist, die für kein* z *des Bogens* (L) *von* $|f(z)|$ *übertroffen wird, und* L *die Länge*[2]) *der Kurve* (L) *bezeichnet.*

Letzteres ergibt sich unmittelbar durch Grenzübergang aus der Ungleichung

$$|S| = \left|\sum_{k=1}^{n} f(\zeta_k)\,\Delta z_k\right| \leqq \sum_{k=1}^{n} |f(\zeta_k)| \cdot |\Delta z_k| \leqq M \sum_{k=1}^{n} |\Delta z_k|,$$

wenn man beachtet, daß die Summe auf der rechten Seite geometrisch die Länge des Streckenzuges mit den Eckpunkten $z_k\,(k = 0, 1, \ldots, n)$ bedeutet.

Ist die Kurve (L) ein Intervall der reellen Achse und nimmt $f(z)$ auf (L) nur reelle Werte an, so stimmt das Kurvenintegral mit dem üblichen (reellen) Integral überein. Hieraus erkennt man, daß die angegebenen Eigenschaften I bis IV sinngemäße Verallgemeinerungen von bekannten Eigenschaften des reellen Integrals sind.

---

[1]) Man beachte, daß mit δ auch das Maximum der $|\Delta x_k|$ und das der $|\Delta y_k|$ gegen Null strebt. — *Anm. d. wissenschaftl. Red.*

[2]) Man kann zeigen, daß unter den in Fußnote 1 auf S. 497 angegebenen Voraussetzungen über (L) diese Länge wirklich existiert (vgl. Seite 398). — *Anm. d. wissenschaftl. Red.*

Wir wollen die Funktion $f(z)$ *in einem gegebenen einfach zusammenhängenden*[1]) *Gebiet* (D) *integrierbar* nennen, wenn sie längs jeder ganz innerhalb (D) verlaufenden Kurve integrierbar ist, wobei für je zwei in (D) verlaufende Kurven (L_1) und (L_2), die beide denselben Anfangs- und Endpunkt besitzen, gilt:

$$\int_{(L_1)} f(z)\,dz = \int_{(L_2)} f(z)\,dz. \tag{62}$$

Diese letzte Forderung ist offenbar der Forderung äquivalent, daß *für jede ganz in* (D) *verlaufende geschlossene*[2]) *Kurve* (C)

$$\int_{(C)} f(z)\,dz = 0 \tag{63}$$

ist.

In der Tat: Ist das Integral längs jeder geschlossenen Kurve gleich Null, so ist

$$\int_{AnBmA} f(z)\,dz = 0,$$

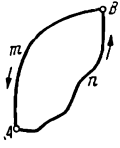


Abb. 2

wenn $A\,m\,B$ und $A\,n\,B$ irgend zwei die Punkte A und B verbindende, in (D) verlaufende Kurven sind (vgl. Abb. 2), d. h. (nach Eigenschaft II)

$$\int_{AnB} f(z)\,dz + \int_{BmA} f(z)\,dz = 0$$

oder (nach Eigenschaft III)

$$\int_{AnB} f(z)\,dz - \int_{AmB} f(z)\,dz = 0, \qquad \int_{AmB} f(z)\,dz = \int_{AnB} f(z)\,dz.$$

Ist umgekehrt der Wert des Kurvenintegrals jeweils nur vom Anfangspunkt A und Endpunkt B des Weges abhängig, über den integriert wird, nicht aber vom Verlauf des Weges selbst, so betrachten wir einen beliebigen geschlossenen Weg und auf ihm zwei Punkte A und B, wobei wir die dabei entstehenden Bögen mit $A\,m\,B$ und $A\,n\,B$ bezeichnen wollen. Dann ist

$$\int_{AnB} f(z)\,dz = \int_{AmB} f(z)\,dz$$

und daher (wenn wir das Integral auf der rechten Seite nach links bringen und beide Integrale zusammenfassen)

$$\int_{AnBmA} f(z)\,dz = 0,$$

was zu zeigen war.

Die Funktion $f(z) = z^p$ *ist für jedes positive ganze* p *in jedem Gebiet* (D) (d. h. in der ganzen Ebene) *integrierbar.*

Beweis. Bilden wir die Integralsumme von $f(z)$ bezüglich eines ge-

[1]) Wir nennen hier ein Gebiet einfach zusammenhängend, wenn es entweder aus sämtlichen Punkten besteht, die im Innern einer gewissen stetigen geschlossenen Kurve liegen, welche sich nicht selbst schneidet, oder wenn es aus sämtlichen Punkten besteht, die auf einer Seite einer „aus dem Unendlichen kommenden und ins Unendliche gehenden Kurve" (z. B. einer Parabel) liegen, oder wenn es aus allen Punkten der Ebene besteht. In allgemeinerer Form werden wir hier diesen Begriff nicht benötigen.

[2]) Das heißt, Anfangs- und Endpunkt der Kurve fallen zusammen.

schlossenen Weges (C), so fallen bei der Zerlegung Anfangs- und Endpunkt zusammen, d. h.

$$z_n = z_0 .$$

Daher ist

$$\sum_{k=1}^{n} z_k^{p+1} = \sum_{k=1}^{n} z_{k-1}^{p+1},$$

also

$$\sum_{k=1}^{n} \left(z_k^{p+1} - z_{k-1}^{p+1}\right) = 0$$

und damit

$$\sum_{k=1}^{n} \left(z_k^p + z_k^{p-1} z_{k-1} + \cdots + z_{k-1}^p\right) \Delta z_k = 0,$$

wofür man auch

$$\sum_{k=1}^{n} \left\{ \sum_{m=0}^{p} z_k^m z_{k-1}^{p-m} \right\} \Delta z_k = 0$$

oder

$$\sum_{m=0}^{p} \left\{ \sum_{k=1}^{n} z_k^m z_{k-1}^{p-m} \Delta z_k \right\} = 0$$

schreiben kann. Führt man nun den Integrationsprozeß aus, so strebt jede der Summen

$$S'_m \equiv \sum_{k=1}^{n} z_k^m z_{k-1}^{p-m} \Delta z_k \tag{64}$$

gegen das Integral $\int_{(C)} z^p dz$,[1] und da es insgesamt $p+1$ derartige Summanden sind, erhalten wir

$$(p+1) \int_{(C)} z^p dz = 0,$$

---

[1] Die Summen S'_m sind zwar selbst keine Integralsummen, lassen sich jedoch mit der gewöhnlichen Integralsumme

$$S_m = \sum_{k=1}^{n} z_{k-1}^p \Delta z_k \qquad (\zeta_k = z_{k-1})$$

sehr leicht vergleichen. Ist nämlich R der maximale Abstand der Punkte der Kurve (C) vom Ursprung, so ist

$$\begin{aligned} |S'_m - S_m| &= \left| \sum_{k=1}^{n} z_{k-1}^{p-m} (z_k^m - z_{k-1}^m) \Delta z_k \right| \\ &= \left| \sum_{k=1}^{n} z_{k-1}^{p-m} (z_k^{m-1} + z_k^{m-2} z_{k-1} + \cdots + z_{k-1}^{m-1}) \Delta z_k \right| \\ &\leqq \sum_{k=1}^{n} |z_{k-1}|^{p-m} (|z_k|^{m-1} + |z_k|^{m-2} |z_{k-1}| + \cdots + |z_{k-1}|^{m-1}) \cdot |\Delta z_k| \\ &\leqq m R^{p-1} \sum_{k=1}^{n} |\Delta z_k|. \end{aligned}$$

Nun ist aber $\sum_{k=1}^{n} |\Delta z_k| < \delta C$, wobei C die Länge der Kurve (C) ist. Hieraus folgt $|S'_m - S_m| < m R^{p-1} \delta C < p R^{p-1} \delta C$, so daß S'_m für $\delta \to 0$ denselben Grenzwert wie S_m besitzt.

so daß auch

$$\int_{(C)} z^p \, dz = 0. \tag{65}$$

Wenden wir auf das erhaltene Resultat Satz I an, so finden wir: *Jedes Polynom* (jede ganz-rationale Funktion)

$$f(z) = P(z) \quad (\equiv az^n + bz^{n-1} + \cdots + kz + l)$$

ist in jedem Gebiet (D) (d. h. in der ganzen Ebene) *integrierbar.*

Welche Funktionen noch und in welchen Gebieten integrierbar sind, werden wir im folgenden Paragraphen sehen.

§ 11. Die Approximation von Funktionen durch Polynome

Wir betrachten eine beliebige Folge $\{P_n(z)\}$ $(n = 1, 2, 3, \ldots)$ von Polynomen, die auf einer gewissen geschlossenen Kurve (C) gleichmäßig konverf gieren möge. Die Grenzfunktion $f(z)$ ist dann, wie man leicht zeigt, auder Kurve (C) stetig.[1])

Der Zusammenhang zwischen den Polynomen $P_n(z)$ und der Grenzfunktion $f(z)$ ist dabei folgender: Zu jedem noch so kleinen $\varepsilon > 0$ gibt es eine natürliche Zahl N_ε, so daß für alle $n > N_\varepsilon$ bei beliebigem z auf (C) die Ungleichung

$$|P_n(z) - f(z)| < \varepsilon \tag{66}$$

gilt. In diesem Fall gilt dann aber auch auf Grund von IV aus § 10 (man setze dort $M = \varepsilon$)

$$\left| \int_{(C)} \{P_n(z) - f(z)\} \, dz \right| < \varepsilon C,$$

wobei C die Länge von (C) ist. Nun ist aber $\int_{(C)} P_n(z) \, dz$ sicher gleich Null, so daß wir die vorangehende Gleichung auch folgendermaßen schreiben können:

$$\left| \int_{(C)} f(z) \, dz \right| < \varepsilon C.$$

Da nun aber die rechte Seite dieser Ungleichung beliebig klein gemacht werden kann, ist der Betrag des Integrals und damit auch das Integral selbst gleich Null.

---

[1]) Der Begriff der Funktion wird hier und im folgenden im Sinne von Zuordnung benutzt: Jedem Punkt z der Kurve (C) ist eine gewisse Zahl $f(z)$ zugeordnet, im vorliegenden Fall der Grenzwert der Zahlenfolge $\{P_n(z)\}$. Daß die Funktion $f(z)$ auf (C) stetig ist, wird genau wie im Reellen bewiesen. Man hat dabei nur zu beachten, daß die Polynome $P_n(z)$ in der ganzen Ebene stetig sind.

Wir finden also:

Gibt es eine Folge $\{P_n(z)\}$ von Polynomen, welche auf der geschlossenen Kurve (C) gleichmäßig gegen die Funktion $f(z)$ konvergiert, so ist

$$\int_{(C)} f(z)\,dz = 0. \tag{67}$$

Wir nehmen nun an, daß die Folge $\{P_n(z)\}$ sogar in einem ganzen einfach zusammenhängenden Gebiet (D) gleichmäßig gegen die Funktion $f(z)$ konvergiert. Dann gilt (67) für jede ganz in (D) liegende geschlossene Kurve (C), d. h., die Funktion $f(z)$ ist im Gebiet (D) integrierbar.

Wir wollen zunächst einige Beispiele betrachten:

Beispiel 1. Es sei

$$P_n(z) = 1 + \frac{z}{1!} + \frac{z^2}{2!} + \cdots + \frac{z^n}{n!} \qquad (n = 1, 2, \ldots).$$

Man erkennt leicht, daß die Folge $\{P_n(z)\}$ in jedem endlichen Gebiet (D) gleichmäßig gegen die Funktion $f(z) = e^z$ konvergiert. Liegen nämlich alle Punkte des Gebietes (D) innerhalb des Kreises vom Radius R um den Ursprung, so ist

$$|P_n(z) - e^z| = \left| \sum_{k=n+1}^{\infty} \frac{z^k}{k!} \right| \leqq \sum_{k=n+1}^{\infty} \frac{R^k}{k!},$$

wobei der Reihenrest auf der rechten Seite für hinreichend großes n kleiner als eine beliebig kleine positive Zahl ε wird.

Beispiel 2. Es sei

$$\Pi_n(z) = \left(1 + \frac{z}{n}\right)^n \qquad (n = 1, 2, \ldots).$$

Wir wollen die Polynome $\Pi_n(z)$ mit den entsprechenden Polynomen $P_n(z)$ vergleichen. Man erkennt ohne weiteres, daß das absolute und das lineare Glied in $\Pi_n(z)$ und $P_n(z)$ übereinstimmen, während (für $k \geqq 2$) der Koeffizient von z^k in der Differenz $P_n(z) - \Pi_n(z)$ durch

$$\frac{1}{k!} - \frac{\binom{n}{k}}{n^k} = \frac{1}{k!}\left[1 - \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)\right]$$

gegeben wird. Dieser Ausdruck ist jedoch kleiner als

$$\frac{1}{k!}\left[1 - \left(1 - \frac{k-1}{n}\right)^{k-1}\right] < \frac{(k-1)^2}{k!\,n} < \frac{1}{(k-2)!\,n},$$[1])

[1]) Man braucht nur in der für $0 < a < b$ geltenden Ungleichung

$$b^m - a^m < m\,b^{m-1}(b-a)$$

die Substitution

$$m = k-1, \quad a = 1 - \frac{k-1}{n}, \quad b = 1$$

auszuführen.

so daß für $|z| \leqq R$

$$|\Pi_n(z) - P_n(z)| < \frac{1}{n} \sum_{k=2}^{n} \frac{R^k}{(k-2)!} < \frac{R^2}{n} \sum_{n=2}^{\infty} \frac{R^{k-2}}{(k-2)!} = \frac{1}{n} \cdot R^2 e^R.$$

Hier strebt nun für $n \to \infty$ die rechte Seite gegen Null, so daß die Folge $\Pi_n(z)$ in jedem beschränkten Gebiet (D) gleichmäßig gegen denselben Grenzwert wie die Folge $P_n(z)$ konvergiert, d. h.,

$$\left(1 + \frac{z}{n}\right)^n \Rightarrow e^z \qquad (|z| \leqq R).\ ^1)$$

Beispiel 3. Entsprechend findet man, daß in jedem endlichen Gebiet (D) die Beziehungen

$$\sum_{k=0}^{n} (-1)^k \frac{z^{2k}}{(2k)!} \Rightarrow \cos z, \quad \frac{\left(1 + \frac{iz}{n}\right)^n + \left(1 - \frac{iz}{n}\right)^n}{2} \Rightarrow \cos z$$

gelten (analog für den Sinus).

Beispiel 4. Die Folge $\{z^n\}$ konvergiert für $|z| < 1$ gegen die Grenzfunktion $f(z) = 0$, und zwar konvergiert sie gleichmäßig in jedem Kreis $|z| \leqq \varrho$ mit $\varrho < 1$, denn

$$|z^n| \leqq \varrho^n \to 0.$$

Beispiel 5. Die Folge $\{1 + z + z^2 + \cdots + z^n\}$, d. h. die Reihe

$$1 + z + z^2 + \cdots + z^n + \cdots$$

konvergiert für $|z| \leqq \varrho$ mit $\varrho < 1$ gleichmäßig gegen die Grenzfunktion $f(z) = \frac{1}{1-z}$, da

$$\left|(1 + z + \cdots + z^n) - \frac{1}{1-z}\right| = \frac{|z|^{n+1}}{|1-z|} \leqq \frac{\varrho^{n+1}}{1-\varrho} \to 0.$$

Beispiel 6. Die Folge

$$\left\{-\frac{A}{a-c}\left[1 + \frac{z-c}{a-c} + \left(\frac{z-c}{a-c}\right)^2 + \cdots + \left(\frac{z-c}{a-c}\right)^n\right]\right\},$$

d. h. die Reihe

$$-\frac{A}{a-c}\left[1 + \frac{z-c}{a-c} + \left(\frac{z-c}{a-c}\right)^2 + \cdots + \left(\frac{z-c}{a-c}\right)^n + \cdots\right] \qquad (c \neq a)$$

konvergiert im Innern des Kreises mit dem Mittelpunkt c und dem Radius $|a-c|$ gegen die Funktion $f(z) = \frac{A}{z-a}$, und zwar gleichmäßig in jedem zum genannten Kreis konzentrischen Kreis mit kleinerem Radius. Dies ergibt sich aus dem im vorangehenden Beispiel erhaltenen Resultat, wenn man dort z durch $\frac{z-c}{a-c}$ ersetzt. Hierbei kann der Punkt $c \neq a$ beliebig gewählt werden.

---

[1]) Vgl. Seite 193.

Wir wollen die Funktion $f(z)$ *im Punkte* z_0 *durch Polynome approximierbar* nennen, wenn es eine Umgebung (D) des Punktes z_0 gibt (z. B. einen Kreis mit z_0 als Mittelpunkt), in welchem eine gewisse Folge von Polynomen gleichmäßig gegen die Funktion $f(z)$ konvergiert. Diese Eigenschaft einer Funktion ist von sehr großer Bedeutung in der Funktionentheorie. *Wenn zwei* (oder mehr) *Funktionen diese Eigenschaft besitzen, so besitzen auch ihre Summe und ihr Produkt diese Eigenschaft.* Dies ergibt sich unmittelbar aus den grundlegenden Sätzen über den Grenzwert einer Summe und eines Produktes (vgl. S. 159) unter zusätzlicher Beachtung der gleichmäßigen Konvergenz.

Aus Beispiel 6 folgt sofort, daß jede gebrochen rationale Funktion der Form $\frac{A}{z-a}$ in jedem Punkt außer im „Pol" a durch Polynome approximierbar ist. Daraus folgt jedoch weiter, daß *überhaupt jede rationale Funktion diese Eigenschaft besitzt*; denn man kann jede rationale Funktion als Summe von „Stammbrüchen" darstellen, die alle die angegebenen Eigenschaften besitzen.[1]) Man hat nur sämtliche Pole auszuschließen, d. h. die Punkte, in welchen der Nenner verschwindet.

Aus den angegebenen Beispielen folgt weiter, daß *auch die Funktionen* e^z, $\cos z$ *und* $\sin z$ *in jedem Punkte der Ebene die uns interessierende Eigenschaft besitzen.*

Eine große Bedeutung besitzt auch folgender Satz: *Sind die Funktion* $\varphi(z)$ *im Punkte* z_0 *und die Funktion* $f(w)$ *im Punkte* $w_0 = \varphi(z_0)$ *durch Polynome approximierbar, so ist die Funktion* $F(z) = f(\varphi(z))$ *im Punkte* z_0 *durch Polynome approximierbar.*

Beim Beweis hat man entscheidend davon Gebrauch zu machen, daß ein Polynom von Polynomen (als mittelbare Funktion) wiederum ein Polynom ist.

Aus dem zuletzt genannten Satz folgt, daß z. B. die Funktion $e^{\cos z}$ in allen Punkten und die Funktion $e^{\frac{1}{z}}$ in allen Punkten außer $z = 0$ durch Polynome approximierbar sind.

§ 12. Die Stammfunktion

Hauptsatz. *Ist eine Funktion* $f(z)$ *in jedem Punkt eines einfach zusammenhängenden Gebietes* (D) *durch Polynome approximierbar, so ist sie in diesem Gebiet integrierbar.*

Im vorangehenden Paragraphen haben wir bereits den Fall behandelt, daß die Approximation in allen Punkten des Gebietes durch dieselbe Folge von Polynomen geleistet wird (was im hier angegebenen Satz nicht mehr verlangt wird). Ohne daß wir auf den Beweis des Hauptsatzes in allen Einzel-

[1]) Falls der Nenner eine mehrfache Nullstelle besitzt, ist noch eine kleine Zusatzüberlegung notwendig; man muß nämlich zeigen, daß auch alle Funktionen der Form $\frac{A}{(z-a)^n}$ $(n \geqq 2)$ in jedem von a verschiedenen Punkt durch Polynome approximierbar sind.

heiten eingehen, erwähnen wir nur, daß er sich auf folgende Tatsache gründet: *Ist eine Funktion $f(z)$ in jedem der beiden sich schneidenden Kreise (Γ_1) und (Γ_2) integrierbar, so ist sie es auch im Gebiet $(D) = (\Gamma_1) + (\Gamma_2)$, das aus allen den Punkten besteht, welche wenigstens einem der beiden Kreise angehören* (Abb. 3).

In der Tat: Nach Voraussetzung ist das Integral $\int f(z)dz$ längs jeder geschlossenen Kurve gleich Null, welche ganz im Innern eines der beiden Kreise verläuft. Wir betrachten nun eine beliebige geschlossene Kurve (C), welche ganz im Innern von (D) verläuft, ohne daß sie ganz im Innern eines der beiden Kreise enthalten ist (Abb. 3). Dann ist

$$\int_{(C)} f(z)\,dz = \int_{(C_1)} f(z)\,dz + \int_{(C_2)} f(z)\,dz,$$

wobei (C_1) eine geschlossene Kurve ist, die aus einem Bogen $\overset{\frown}{QMP}$ der Kurve (C) und einem neu eingefügten (in der Abbildung gestrichelt gezeichneten) Bogen $\overset{\frown}{PQ}$ besteht, welcher ganz im „Durchschnitt" der Kreise (Γ_1) und (Γ_2) liegt; entsprechend ist (C_2) dann die Kurve, die aus dem Bogen $\overset{\frown}{PNQ}$ der Kurve (C) und dem mit entgegengesetzter Orientierung genommenen Bogen $\overset{\frown}{PQ}$ (also dem Bogen $\overset{\frown}{QP}$) besteht. Da nun jedes der Integrale auf der rechten Seite gleich Null ist, gilt dies auch für das Integral auf der linken Seite, was zu beweisen war.

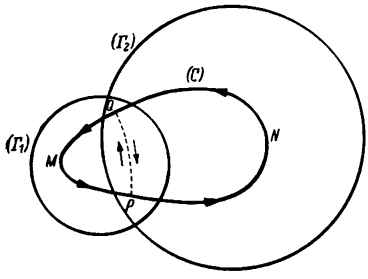


Abb. 3

Aus den weiteren Ausführungen wird sich ergeben, daß die Voraussetzung des einfachen Zusammenhanges sehr wesentlich ist.

Es sei nun $f(z)$ eine in einem gewissen einfach zusammenhängenden Gebiet (D) definierte Funktion, die in (D) stetig und integrierbar ist. Dann hängt das Integral $I = \int_{(L)} f(z)dz$, welches längs eines gewissen von einem Punkt a zu einem Punkt b führenden Weges (L) genommen wird, nur von den Punkten a und b, nicht aber von der speziellen Wahl des Integrationsweges ab (jedenfalls solange der Integrationsweg das Gebiet (D) nicht verläßt). Daher erübrigt sich im angegebenen Fall die Angabe des Integrationsweges, und man kann kurz

$$I = \int_a^b f(z)\,dz$$

schreiben.

Es zeigt sich nun, daß *das Integral*

$$\Phi(z)=\int_a^z f(\zeta)\,d\zeta \quad [1] \tag{68}$$

(wobei a ein beliebiger fester Punkt und z ein variabler Punkt aus (D) ist) *eine Stammfunktion zu* $f(z)$ *im Gebiet* (D) *bildet, d. h., es gilt*

$$\Phi'(z)=\frac{d}{dz}\int_a^z f(\zeta)\,d\zeta=f(z). \quad [2] \tag{69}$$

Beweis. Offenbar ist[3])

$$\begin{aligned}\frac{\Phi(z+h)-\Phi(z)}{h}&=\frac{1}{h}\left\{\int_a^{z+h} f(\zeta)\,d\zeta-\int_a^z f(\zeta)\,d\zeta\right\}=\frac{1}{h}\int_z^{z+h} f(\zeta)\,d\zeta\\&=\frac{1}{h}\int_z^{z+h}[f(z)+(f(\zeta)-f(z))]\,d\zeta\\&=\frac{1}{h}\left\{f(z)\int_z^{z+h} d\zeta+\int_z^{z+h}(f(\zeta)-f(z))\,d\zeta\right\}\\&=f(z)+\frac{1}{h}\int_z^{z+h}(f(\zeta)-f(z))\,d\zeta.\end{aligned}$$

Liegt nun der Punkt $z+h$ hinreichend nahe bei z (Abb. 4), so wird der zweite Summand im letzten Ausdruck dem Betrage nach kleiner als eine beliebig vorgegebene Zahl ε; wählen wir nämlich z. B. von z nach $z+h$ einen geradlinigen Integrationsweg (vorausgesetzt, daß dies in (D) möglich ist) und wenden wir auf das Integral

$$\int_z^{z+h}(f(\zeta)-f(z))\,d\zeta$$

Satz IV aus § 10 an, so erhalten wir

$$\left|\frac{1}{h}\int_z^{z+h}(f(\zeta)-f(z))\,d\zeta\right|\leqq M,$$

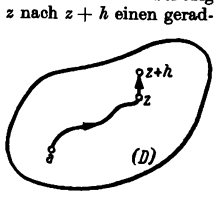


Abb. 4

[1]) Um Verwechslungen mit der Integrationsgrenze zu vermeiden, haben wir hier die Integrationsveränderliche mit ζ bezeichnet (bekanntlich hängt das Integral nicht von der Bezeichnung der Integrationsveränderlichen ab).

[2]) Dieser Satz ist die naturgemäße Verallgemeinerung des Satzes von Seite 374.

[3]) Wir setzen hierbei voraus, daß sich der Punkt $z+h$ im Gebiet (D) befindet und daß der Integrationsweg von a nach $z+h$ durch den Punkt z geht.

wobei M das Maximum der Funktion $|f(\zeta)-f(z)|$ auf dem Wege von z nach $z+h$ ist. Wegen der Stetigkeit von $f(z)$ kann aber die Differenz $f(\zeta)-f(z)$ dem Betrage nach kleiner als jedes positive ε gemacht werden, wenn man h hinreichend klein wählt. Also ist

$$\left|\frac{\Phi(z+h)-\Phi(z)}{h}-f(z)\right|<\varepsilon,$$

woraus unsere Behauptung unmittelbar zu entnehmen ist.

Als Beispiel betrachten wir das Integral

$$\Phi(z)=\int_1^z \frac{d\zeta}{\zeta}. \tag{70}$$

Die Funktion $f(z)=\frac{1}{z}$ ist in jedem einfach zusammenhängenden Gebiet, welches den Ursprung nicht enthält, integrierbar. Ohne daß wir zunächst ein derartiges Gebiet näher festlegen, betrachten wir folgenden Integrationsweg: Wir gehen vom Punkte 1 längs der reellen Achse zum Punkte r und von dort aus längs des Kreises $|z|=r$ zum Punkte z (Abb. 5). Dann ist also

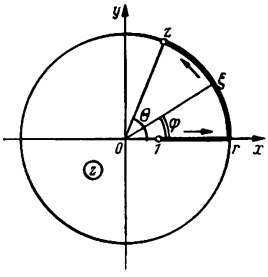


Abb. 5

$$\Phi(z)=\int_1^r \frac{d\zeta}{\zeta}+\int_r^z \frac{d\zeta}{\zeta}.$$

Das erste Integral ist ein gewöhnliches reelles Integral, und daher gilt

$$\int_1^r \frac{d\zeta}{\zeta}=\int_1^r \frac{dx}{x}=\ln r.$$

Zur Auswertung des zweiten Integrals setzen wir $z=re^{i\theta}$, $\zeta=re^{i\varphi}$ (Abb. 5) und wählen φ als neue Integrationsveränderliche. Dann ist

$$d\zeta=ire^{i\varphi}d\varphi, \quad \frac{d\zeta}{\zeta}=id\varphi,$$

und da sich φ in den Grenzen zwischen 0 und θ ändert, gilt also

$$\int_r^z \frac{d\zeta}{\zeta}=i\int_0^\theta d\varphi=i\theta.$$

Diese Substitution in einem komplexen Integral bedarf allerdings noch eines Beweises. Zu diesem Zweck setzen wir

$$e^z-1=z(1+\varepsilon). \tag{71}$$

Dann ist für $|z| \leqq 1$ offenbar $|\varepsilon| < 1$; es ist nämlich

$$\varepsilon = \frac{e^z - 1}{z} - 1 = \left(\frac{1}{2!} + \frac{z}{3!} + \frac{z^2}{4!} + \cdots\right) z$$

und daher

$$|\varepsilon| < \left|\frac{1}{2!} + \frac{z}{3!} + \frac{z^2}{4!} + \cdots\right| \cdot |z| < \left(\frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \cdots\right) \cdot |z| < (e-2) \cdot |z| < |z|.$$

Wir fügen nun auf dem Bogen zwischen r und z die Teilpunkte

$$z_k = r e^{i\varphi_k} \qquad (k = 0,\ 1,\ \ldots,\ n)$$

ein, wobei

$$0 < \varphi_1 < \cdots < \varphi_{n-1} < \varphi_n = \theta,$$

und setzen $z_0 = r$ und $z_n = z$. Dann ist auf Grund der Definition des Integrals ($\zeta_k = z_{k-1}$ gesetzt)

$$\int_r^z \frac{d\zeta}{\zeta} = \lim \Sigma \frac{\Delta z_k}{z_{k-1}}.$$

Mit (71) ist nun

$$\Delta z_k = z_k - z_{k-1} = r(e^{i\varphi_k} - e^{i\varphi_{k-1}}) = r e^{i\varphi_{k-1}}(e^{i\Delta\varphi_k} - 1) = i r e^{i\varphi_{k-1}} \Delta\varphi_k (1 + \varepsilon_k),$$

wobei

$$|\varepsilon_k| < |\Delta\varphi_k| \qquad (k = 1,\ \ldots,\ n).$$

Ferner ist

$$z_{k-1} = r e^{i\varphi_{k-1}}$$

und daher

$$\frac{\Delta z_k}{z_{k-1}} = i \Delta\varphi_k (1 + \varepsilon_k).$$

Also erhalten wir das uns interessierende Integral $\int_r^z \frac{d\zeta}{\zeta}$ auch als Grenzwert von

$$\sum_{k=1}^{n} i \Delta\varphi_k (1 + \varepsilon_k) = i \sum_{k=1}^{n} \Delta\varphi_k + i \sum_{k=1}^{n} \varepsilon_k \Delta\varphi_k.$$

Der erste Summand auf der rechten Seite ist gleich $i\theta$. Was den zweiten Summanden betrifft, so erhalten wir mit $|\Delta\varphi_k| < \delta$ wegen $|\varepsilon_k| < |\Delta\varphi_k| < \delta$ für seinen Betrag die Abschätzung

$$\left| i \sum_{k=1}^{n} \varepsilon_k \Delta\varphi_k \right| \leqq \sum_{k=1}^{n} |\varepsilon_k| \cdot |\Delta\varphi_k| < \delta \cdot \sum_{k=1}^{n} |\Delta\varphi_k| < \delta \cdot \theta,$$

d. h., er strebt beim Integrationsprozeß gegen Null. Daher ist

$$\int_r^z \frac{d\zeta}{\zeta} = i\theta.$$

Die im Haupttext angegebenen Umformungen sind nur als eine abgekürzte Darstellung dieses ganzen Prozesses aufzufassen.

Wir haben damit erhalten:

$$\Phi(z) = \ln r + i\theta.$$

Dieses Ergebnis entspricht genau der Definition der Funktion $\operatorname{Ln} z$ im Komplexen (§ 6). Im betrachteten Fall ist jedoch das Argument θ eindeutig bestimmt. Gehen wir, wie dies in Abbildung 5 angedeutet ist, von r aus auf dem Kreise entgegen dem Uhrzeigersinn direkt zum Punkte z, so liegt das Argument θ zwischen 0 und 2π. Führen wir dagegen zuvor einen vollen Umlauf auf dem Kreis (im positiven Sinne) aus, so erhalten wir den zusätzlichen Summanden

$$i\int_0^{2\pi} d\varphi = 2\pi i.$$

Allgemein tritt bei k vollen Umläufen ($k \gtrless 0$) der zusätzliche Summand $2k\pi i$ auf. Man kann also auf diese Weise sämtliche „Zweige" des komplexen Logarithmus erhalten. Die Mehrdeutigkeit des Logarithmus erklärt sich hier also aus dem Vorhandensein der „Singularität" bei $z = 0$, um welche herum die Integration ausgeführt wurde.

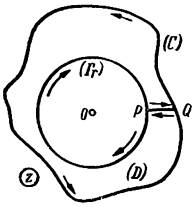


Abb. 6

Beschränken wir uns auf ein einfach zusammenhängendes Gebiet (D), welches den Punkt $z = 0$ nicht enthält, so fällt die Mehrdeutigkeit fort.

Man erkennt ohne Schwierigkeit, daß das Integral $\int \frac{d\zeta}{\zeta}$ nicht nur längs des Kreises (Γ_r) vom Radius r um den Ursprung den Wert $2\pi i$ besitzt, sondern daß dies, wie der Abb. 6 zu entnehmen ist, für jede geschlossene doppelpunktfreie Kurve (C) gilt, welche den Punkt $z = 0$ umschließt. Denken wir uns nämlich das in Abb. 6 mit (D) bezeichnete Gebiet längs der Strecke PQ „aufgeschnitten" und integrieren wir über den Rand dieses Gebietes[1]), so ist das Integral gleich Null, denn das „aufgeschnittene" Gebiet (D) ist einfach zusammenhängend und enthält den Ursprung $z = 0$ nicht; gleichzeitig heben sich die Integrale längs der Strecken PQ und QP gegenseitig auf, so daß

$$\int_{(C)} \frac{d\zeta}{\zeta} - \int_{(\Gamma_r)} \frac{d\zeta}{\zeta} = 0$$

und daher

$$\int_{(C)} \frac{d\zeta}{\zeta} = \int_{(\Gamma_r)} \frac{d\zeta}{\zeta} = 2\pi i. \tag{72}$$

Wir wollen schließlich noch das Integral $\int_{(C)} \frac{d\zeta}{\zeta - a}$ betrachten, wobei (C) eine beliebige geschlossene doppelpunktfreie Kurve ist, welche den Punkt a,

[1]) Genaugenommen haben wir hierbei einen Grenzübergang auszuführen (wir müssen nämlich die Strecke PQ zunächst wirklich „herausschneiden" und dann die Schnittränder gegen die Strecke PQ streben lassen — *Anm. d. wissenschaftl. Red.*).

den einzigen Pol der Funktion $f(z)=\frac{1}{z-a}$, umschließt. Mit Hilfe derselben Überlegungen wie im vorangehenden, wobei nur ein Kreis mit dem Mittelpunkt a zu nehmen und $\zeta-a=re^{i\varphi}$ zu setzen ist, findet man, daß dieses Integral ebenfalls den Wert $2\pi i$ besitzt.

Wir wollen noch einmal hervorheben, daß auf Grund des Hauptsatzes der Wert des Integrals $\int\limits_{(C)}\frac{d\zeta}{\zeta-a}$ gleich Null ist, wenn der Punkt a außerhalb der geschlossenen Kurve (C) liegt.[1])

Wir erhalten damit, daß *das Integral* $\int\limits_{(C)}\frac{d\zeta}{\zeta-a}$ *längs einer geschlossenen doppelpunktfreien Kurve* (C) *gleich Null oder gleich* $2\pi i$ *ist, je nachdem, ob die Kurve* (C) *den Punkt* a *nicht umschließt oder umschließt.*[2])

Aufgabe.

Man zeige (indem man die Überlegungen nach dem oben angegebenen Vorbild durchführt), daß das Integral $\int\limits_{(C)}\frac{d\zeta}{(\zeta-a)^m}$ für jede von 1 verschiedene ganze Zahl m gleich Null ist, wenn (C) irgendeine geschlossene Kurve ist, die nicht durch den Punkt a geht.

§ 13. Die Cauchysche Integralformel

Es ist bekannt, daß man bei Division eines beliebigen Polynoms $P(z)$ durch $z-a$ (a sei dabei eine beliebige komplexe Zahl) als Rest die Zahl $P(a)$ erhält.[3]) Es gilt also die Identität

$$P(z)\equiv(z-a)\,Q(z)+P(a), \tag{73}$$

wobei $Q(z)$ wiederum ein Polynom ist $\left(\text{der Quotient }\frac{P(z)-P(a)}{z-a}\right)$. Integrieren wir nun den Quotienten $\frac{P(a)}{z-a}$ längs einer beliebigen geschlossenen doppelpunktfreien Kurve (C), welche den Punkt a umschließt, so finden wir

$$\int\limits_{(C)}\frac{P(z)\,dz}{z-a}=\int\limits_{(C)}Q(z)\,dz+P(a)\int\limits_{(C)}\frac{dz}{z-a}.$$

Das erste Integral auf der rechten Seite ist gleich Null, während das zweite von Null verschieden ist und den Wert $2\pi i$ besitzt (§ 12). Lösen wir also die erhaltene Gleichung nach $P(a)$ auf, so finden wir

$$P(a)=\frac{1}{2\pi i}\int\limits_{(C)}\frac{P(z)\,dz}{z-a}. \tag{74}$$

---

[1]) Die Funktion $\frac{1}{z-a}$ läßt sich nämlich in jedem von a verschiedenen Punkt durch Polynome approximieren (vgl. Beispiel 6 in § 11).

[2]) Das Integral verliert seinen Sinn, wenn die Kurve (C) durch den Punkt a hindurchgeht, weil die Funktion $f(z)$ im Punkte a nicht erklärt ist.

[3]) Nach dem sogenannten Satz von Bézout.

Man bezeichnet diese Formel als CAUCHYsche *Integralformel* und das Integral auf der rechten Seite von (74) als CAUCHYsches *Integral*.

Die CAUCHYsche *Integralformel gilt nun aber keineswegs nur für Polynome, sondern auch allgemein für alle Funktionen $f(z)$, welche sich auf (C) und in dem von (C) umschlossenen einfach zusammenhängenden Gebiet (D) durch eine Folge von Polynomen approximieren lassen.* Ist nämlich $\{P_n(z)\}$ eine Folge von Polynomen, die auf (C) und in (D) gleichmäßig gegen die Funktion $f(z)$ konvergiert,

$$P_n(z) \Rightarrow f(z),$$

so können wir in

$$P_n(a) = \frac{1}{2\pi i} \int_{(C)} \frac{P_n(z)\,dz}{z-a}$$

zur Grenze übergehen[1]) und finden

$$f(a) = \frac{1}{2\pi i} \int_{(C)} \frac{f(z)\,dz}{z-a}. \tag{75}$$

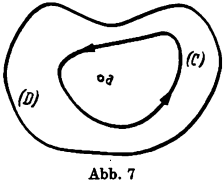


Abb. 7

Mit Hilfe des in § 12 beim Beweis des Hauptsatzes benutzten Verfahrens *kann man schließlich dieses Ergebnis auf alle Funktionen übertragen, welche in einem gewissen, die Kurve (C) ganz in ihrem Innern enthaltenden Gebiet (D) integrierbar sind* (Abb. 7).

Da nun a ein beliebiger Punkt innerhalb von (C) sein kann, können wir a als unabhängige Veränderliche z auffassen, so daß wir, wenn wir noch die Integrationsveränderliche mit ζ bezeichnen, (75) auch in der Form

$$f(z) = \frac{1}{2\pi i} \int_{(C)} \frac{f(\zeta)\,d\zeta}{\zeta - z} \tag{76}$$

schreiben können.

Durch die CAUCHYsche *Integralformel* (76) *werden die Werte der Funktion $f(z)$ im Innern von (C) durch die Werte $f(\zeta)$ auf der Kurve (C) ausgedrückt.*

Aus der CAUCHYschen Integralformel ergeben sich sofort einige wesentliche Folgerungen.

1. *Die Funktion $f(z)$ ist im Innern der Kurve (C) differenzierbar, und zwar erhält man die Ableitung $f'(z)$, wenn man den Integranden des* CAUCHYschen *Integrals nach z differenziert*:

$$f'(z) = \frac{1}{2\pi i} \int_{(C)} \frac{f(\zeta)\,d\zeta}{(\zeta - z)^2}. \tag{77}$$

---

[1]) Dies ergibt sich aus der Gültigkeit der Ungleichungen

$$\left| \int_{(C)} \frac{P_n(z) - f(z)}{z-a}\,dz \right| \leqq C \max_{(C)} \left| \frac{P_n(z) - f(z)}{z-a} \right| \leqq \frac{C}{\eta} \max_{(C)} |P_n(z) - f(z)|;$$

hierbei ist η eine beliebige Konstante, die kleiner als der Minimalabstand des Punktes a von der Kurve (C) ist.

Zum Beweis von (77) genügt es zu zeigen, daß der Quotient

$$\frac{f(z+h)-f(z)}{h}$$

für $h \to 0$ gegen das Integral auf der rechten Seite von (77) konvergiert. Dies ergibt sich jedoch unmittelbar aus

$$\frac{f(z+h)-f(z)}{h} = \frac{1}{2\pi i}\int_{(C)} \frac{f(\zeta)\,d\zeta}{(\zeta-z-h)(\zeta-z)}.$$ [1])

2. *Die Funktion $f(z)$ besitzt innerhalb von (C) Ableitungen beliebiger Ordnung, und zwar ergibt sich $f^{(n)}(z)$ durch n-malige Differentiation des Integranden von* (77) *nach* z:

$$f^{(n)}(z) = \frac{n!}{2\pi i}\int_{(C)} \frac{f(\zeta)\,d\zeta}{(\zeta-z)^{n+1}} \qquad (n=1,\ 2,\ \ldots). \tag{78}$$

Diese Behauptung läßt sich leicht durch vollständige Induktion beweisen. Man hat dazu nur zu zeigen, daß die Differentiation von (78) auf eine analoge Formel führt, bei der anstelle von n dann $n+1$ steht.[2])

3. *Die Funktion $f(z)$ läßt sich für jeden Punkt c aus dem Innern von (C) in eine Potenzreihe nach $z-c$ entwickeln, und zwar ist diese Reihe dann die entsprechende* TAYLORsche *Reihe. Die Konvergenz der Reihe gegen $f(z)$ ist dabei in jedem Kreis um c gesichert, dessen Radius kleiner als der Abstand des Punktes c vom nächsten Punkt der Kurve (C) ist.*

Die Gültigkeit dieses Satzes beruht auf der Tatsache, daß die Funktion $\frac{1}{\zeta-z}$, welche im Integranden von (76) auftritt, als Potenzreihe in $z-c$ dargestellt werden kann:

$$\frac{1}{\zeta-z} = \frac{1}{\zeta-c} + \frac{z-c}{(\zeta-c)^2} + \frac{(z-c)^2}{(\zeta-c)^3} + \cdots \tag{79}$$

(vgl. § 11, Beispiel 6). Dies ist gleichzeitig eine geometrische Reihe, und zwar mit dem Quotienten $\frac{z-c}{\zeta-c}$. Bezeichnen wir nun die Entfernung des Punktes c

[1]) Man braucht dann nämlich nur noch

$$\left|\frac{1}{2\pi i}\int_{(C)} \frac{f(\zeta)\,d\zeta}{(\zeta-z-h)(\zeta-z)} - \frac{1}{2\pi i}\int_{(C)} \frac{f(\zeta)\,d\zeta}{(\zeta-z)^2}\right| = \frac{|h|}{2\pi}\left|\int_{(C)} \frac{f(\zeta)\,d\zeta}{(\zeta-z)^2(\zeta-z-h)}\right|$$

abzuschätzen. Ist nun C die Länge der Kurve (C), M das Maximum von $|f(\zeta)|$ auf (C) und η der Minimalabstand des Punktes z von den Punkten der Kurve (C), so ist für $|h| < \delta$ der betrachtete Ausdruck kleiner als

$$\frac{\delta}{2\pi}\cdot\frac{CM}{\eta^2(\eta-\delta)},$$

woraus alles Weitere leicht zu ersehen ist.

[2]) Hierbei ist ein etwas komplizierteres Integral abzuschätzen, was wir jedoch dem Leser überlassen wollen.

vom nächsten Punkt der Kurve (C) mit η (Abb. 8), so konvergiert die Entwicklung (79), und zwar gleichmäßig für alle z mit $|z-c|<\theta\eta$, wobei θ eine beliebige Zahl zwischen Null und Eins ist. Wir multiplizieren nun die Gleichung (79) mit $f(\zeta)$ und integrieren längs der Kurve (C).[1]) Dividieren wir schließlich noch durch $2\pi i$, so finden wir

$$\frac{1}{2\pi i}\int_{(C)}\frac{f(\zeta)\,d\zeta}{\zeta-z}=\frac{1}{2\pi i}\int_{(C)}\frac{f(\zeta)\,d\zeta}{\zeta-c}+\frac{z-c}{2\pi i}\int_{(C)}\frac{f(\zeta)\,d\zeta}{(\zeta-c)^2}+\frac{(z-c)^2}{2\pi i}\int_{(C)}\frac{f(\zeta)\,d\zeta}{(\zeta-c)^3}+\cdots. \tag{80}$$

Nach Formel (76) ist der Ausdruck auf der linken Seite gleich $f(z)$, so daß

$$f(z)=a_0+a_1(z-c)+a_2(z-c)^2+\cdots\equiv\sum_{n=0}^{\infty}a_n(z-c)^n, \tag{81}$$

wobei

$$a_n=\frac{1}{2\pi i}\int_{(C)}\frac{f(\zeta)\,d\zeta}{(\zeta-c)^{n+1}}\qquad(n=0,\ 1,\ 2,\ \ldots). \tag{82}$$

Wir haben also damit die Funktion $f(z)$ bereits als Potenzreihe in $z-c$ dargestellt. Die gleichmäßige Konvergenz ist dabei für

$$|z-c|<\theta\eta$$

gesichert, wobei θ eine beliebige Zahl zwischen 0 und 1 ist, so daß also, da sich die Zahl θ beliebig wenig von 1 unterscheiden kann, die Darstellbarkeit von $f(z)$ gemäß (81) für alle z mit $|z-c|<\eta$ gesichert ist.

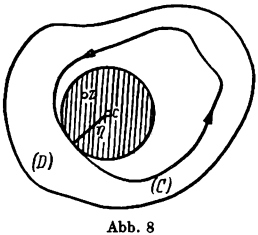


Abb. 8

Vergleichen wir schließlich die Formeln (82) mit den Formeln (76) und (78), so finden wir

$$a_n=\frac{f^{(n)}(c)}{n!}\qquad(n=0,\ 1,\ 2,\ \ldots). \tag{83}$$

Daher können wir (81) auch in der Form

$$f(z)=\sum_{n=0}^{\infty}\frac{f^{(n)}(c)}{n!}(z-c)^n \tag{84}$$

schreiben; das ist aber gerade die Taylorsche Reihe.

§ 14. Der Begriff der analytischen Funktion

Ausgehend von der Auffassung der Funktion als einer durch eine endliche Formel beschriebenen Gesetzmäßigkeit, sind wir zu der verhältnismäßig engen Klasse der elementaren Funktionen einer komplexen Veränderlichen gelangt. Demgegenüber führt uns die allgemeinere Auffassung der Funktion einer komplexen Veränderlichen mit Hilfe einer Zuordnungsvorschrift

$$w=f(z)\qquad(z=x+iy,\quad w=u+iv)$$

[1]) Vgl. Fußnote 2 auf Seite 496.

zwischen den Punkten eines gewissen Gebietes in der z-Ebene und Punkten der w-Ebene zu einer Zuordnung zwischen Paaren (x, y) von reellen Zahlen und Paaren (u, v) von reellen Zahlen, d. h. auf die Betrachtung von Systemen von zwei reellen Funktionen in zwei reellen Veränderlichen (im allgemeinsten Sinne):

$$u = u(x, y),$$
$$v = v(x, y).$$

Dieser allgemeinste Funktionsbegriff überschreitet jedoch bei weitem die Grenzen der Methoden der sogenannten Funktionentheorie.

Um nun zu einer gesetzmäßigen Erweiterung des Begriffs der (elementaren) Funktion im Komplexen zu gelangen, kann man folgendermaßen verfahren: Man fügt, ohne die operative Grundlage zu verlassen, zu den vier rationalen Operationen noch eine weitere Operation, den Grenzübergang bei gleichmäßiger Konvergenz, hinzu. Die auf diese Weise entstehende Funktionenklasse bezeichnet man als Klasse der analytischen Funktionen. Das Studium der Eigenschaften der Funktionen dieser Klasse ist der eigentliche Gegenstand der Funktionentheorie. Zwischen den verschiedenen Eigenschaften der analytischen Funktionen bestehen eine Reihe von sehr bemerkenswerten Beziehungen. Die Klasse der analytischen Funktionen ist in gewissem Sinne abgeschlossen, d. h., man gelangt durch Anwendung der genannten Operationen nicht aus der Klasse der analytischen Funktionen heraus.

Die vorangehenden Darlegungen erfordern jedoch noch einige Präzisierungen und Erläuterungen.

Dazu führen wir zunächst den Begriff der in einem gewissen Gebiet regulären Funktion ein. Was dabei unter einem Gebiet ungefähr zu verstehen ist, wurde oben bereits andeutungsweise gesagt;[1]) allgemein versteht man unter einem *Gebiet* eine zusammenhängende[2]) Punktmenge, welche nur aus inneren Punkten besteht (also „offen" ist).[3]) Um Mißverständnissen vorzubeugen, weisen wir darauf hin, daß ein Gebiet in der komplexen Ebene nachweisbar stets eine „zweidimensionale Ausdehnung" besitzt, daß also z. B. eine Strecke einer Geraden[4]) (oder einer Kurve) niemals ein Gebiet ist.[5]) Man sagt nun, eine in einem Gebiet (D) erklärte (eindeutige) stetige Funktion $f(z)$ sei in diesem Gebiet *regulär* (oder *holomorph* oder *analytisch* oder *regulär analytisch*), wenn sie eine der folgenden Eigenschaften besitzt:

A. *In jedem Punkt des Gebietes (D) ist die Funktion $f(z)$ durch Polynome approximierbar.*

---

[1]) Vgl. Fußnote 1 auf Seite 500.

[2]) Vgl. Seite 500.

[3]) Fügt man also zu der Menge der Punkte eines Gebietes (z. B dem Innern eines Kreises) noch einen Randpunkt (im Beispiel einen Punkt der Peripherie) hinzu, so ist die dabei entstehende Punktmenge kein Gebiet mehr.

[4]) Das heißt das, was wir in der Theorie der Funktionen einer reellen Veränderlichen als einen „Bereich" oder ein „Intervall" bezeichnet haben.

[5]) Das liegt daran, daß alle Punkte des Gebietes innere Punkte sein sollen. — *Anm. d. wissenschaftl. Red.*

B. *Die Funktion $f(z)$ ist in jedem im Gebiet (D) enthaltenen einfach zusammenhängenden Gebiet integrierbar.*

C. *Die Funktion $f(z)$ ist in jedem Punkte des Gebietes (D) differenzierbar.*

D. *Die Funktion $f(z)$ ist bezüglich jedes Punktes c aus (D) in eine Reihe nach Potenzen von $z-c$ entwickelbar, die in einem gewissen Kreis um c gleichmäßig konvergiert.*

Die Bedeutung dieser einzelnen Eigenschaften haben wir bereits früher kennengelernt: die von A in § 11, die von B in § 10, die von C in § 9 und die von D in § 13.

Es zeigt sich nun, daß alle vier Eigenschaften logisch gleichwertig sind, d. h., daß man aus jeder von ihnen die drei übrigen folgern kann.

Bei unseren bisherigen Ausführungen haben wir meistens A vorausgesetzt. Wir haben in § 12 gesehen, daß B aus A folgt. In § 13 haben wir ferner mit Hilfe der CAUCHYschen Integralformel gezeigt, daß C und D aus A folgen. Es ist schließlich unmittelbar einzusehen, daß A aus D folgt, denn die Partialsummen der Reihe liefern eine Folge von Polynomen, welche die Funktion $f(z)$ approximiert.

Auf einen Beweis dafür, daß B aus C folgt, können wir hier nicht eingehen.[1])

Unter Verwendung dieser Tatsache zeigt man jedoch leicht, daß auch D (und damit A) aus C folgt. Zu diesem Zweck genügt es, die CAUCHYsche Integralformel direkt (d. h. ohne Approximation von $f(z)$ durch Polynome) aus C zu beweisen. Wir nehmen also an, daß $f(z)$ im betrachteten Gebiet differenzierbar ist. Dann gilt dies, da $\frac{1}{z-a}$ im Gebiet überall außer im Punkte a differenzierbar ist, mit Ausnahme von a auch für das Produkt $\frac{f(z)}{z-a}$ (§ 9). Hieraus folgt jedoch auf Grund von B, daß man bei Integration dieser Funktion über eine beliebige, den Punkt a umschließende doppelpunktfreie Kurve (C) den Integrationsweg durch den Kreis (Γ_ϱ) um a vom Radius ϱ ersetzen kann, wobei ϱ nur so gewählt werden muß, daß der Kreis (Γ_ϱ) ganz im betrachteten Gebiet enthalten ist. Damit erhalten wir

$$\frac{1}{2\pi i}\int_{(C)}\frac{f(z)\,dz}{z-a}=\frac{1}{2\pi i}\int_{(\Gamma_\varrho)}\frac{f(z)\,dz}{z-a}=\frac{1}{2\pi i}\int_{(\Gamma_\varrho)}\frac{f(a)+[f(z)-f(a)]}{z-a}\,dz$$

$$=f(a)+\frac{1}{2\pi i}\int_{(\Gamma_\varrho)}\frac{f(z)-f(a)}{z-a}\,dz.$$

Beachten wir nun, daß die Funktion $f(z)$ stetig ist, so kann man das letzte Integral leicht abschätzen, und zwar kann man sich davon überzeugen, daß es bei hinreichend kleinem ϱ dem Betrage nach kleiner als jede noch so kleine

[1]) Vgl. etwa A. И. Маркушевич, Теория аналитических функций (A. I. MARKUSCHEWITSCH, Theorie der analytischen Funktionen), Gostechisdat 1950, Kap. III, § 2. (Der deutsche Leser sei z. B. verwiesen auf K. KNOPP, Funktionentheorie I, Sammlg. Göschen Bd. 908, Kap. 4. — *Anm. d. wissenschaftl. Red.*)

positive Zahl ε gemacht werden kann, so daß es für $\varrho \to 0$ gegen Null strebt. Damit ist die CAUCHYsche Integralformel bewiesen, aus der sich sofort D ergibt.

Die zuletzt erwähnte Tatsache, daß nämlich D aus C folgt, ist besonders erwähnenswert, zeigt sie doch, daß *im Komplexen eine Funktion, die in einem gewissen Gebiet einmal differenzierbar ist, automatisch beliebig oft differenzierbar ist*, d. h. Ableitungen $f^{(n)}(z)$ beliebiger Ordnung besitzt. Insbesondere kann man dann für diese Funktion die TAYLORsche Reihe

$$f(c) + \frac{f'(c)}{1!}(z-c) + \frac{f''(c)}{2!}(z-c)^2 + \cdots + \frac{f^{(n)}(c)}{n!}(z-c)^n + \cdots$$

bilden; diese konvergiert in einer gewissen Umgebung des Punktes c und besitzt dort die Summe $f(z)$.

Wir weisen darauf hin, daß im Reellen Entsprechendes keineswegs gilt.

Schließlich wollen wir noch zeigen, wie man aus B die Eigenschaft C erhalten kann: Ist in einem gewissen einfach zusammenhängenden Gebiet (D) das Kurvenintegral $\int\limits_{(C)} f(\zeta)\,d\zeta$ vom Wege unabhängig, so kann man (vgl. § 10) auf die Angabe des Weges verzichten und das Integral $\int\limits_{z_0}^{z} f(\zeta)\,d\zeta$ als Funktion der oberen Integrationsgrenze studieren:

$$\Phi(z) = \int_{z_0}^{z} f(\zeta)\,d\zeta.$$

In diesem Fall besitzt (vgl. § 12) die in ganz (D) erklärte Funktion $\Phi(z)$ überall eine Ableitung, und zwar ist

$$\Phi'(z) = f(z).$$

Dann existiert aber auch (s. oben) die Ableitung

$$\Phi''(z) = f'(z),$$

d. h., $f(z)$ ist differenzierbar, was zu beweisen war.[1])

Um also eine Funktion $f(z)$ als regulär nachzuweisen, genügt es zu zeigen, daß sie eine der Eigenschaften A bis D erfüllt; alle anderen Eigenschaften gelten dann ebenfalls. Am einfachsten ist es in der Regel, die Differenzierbarkeit zu beweisen. *So sind z. B. die elementaren Funktionen in allen den Punkten regulär*[2]), *in deren Umgebung sie eindeutig definiert und differenzierbar sind.*[3])

Beispiele.

1. Die Funktion $f(z) = \operatorname{tg} z$ ist in der ganzen Ebene, die Punkte $(2n+1)\frac{\pi}{2}$ ausgenommen, regulär.

---

[1]) Das ist der sogenannte Satz von MORERA.

[2]) Man sagt, eine Funktion sei in einem Punkte regulär, wenn sie in einer gewissen Umgebung dieses Punktes (z. B. in einem gewissen Kreis mit diesem Punkt als Mittelpunkt) regulär ist.

[3]) Ist die Funktion nicht eindeutig definiert, so kann man sie (wie z. B. den Logarithmus) in eindeutige „Zweige" zerlegen.

2. Die Funktion $f(z) = e^{\frac{1}{z}}$ ist in der ganzen Ebene mit Ausnahme des Punktes $z = 0$ regulär.

Wir wollen schließlich noch die bereits oben formulierte Abgeschlossenheit der Klasse der analytischen Funktionen beweisen.

Jede in einem Gebiet reguläre Funktion ist zunächst (jedenfalls in der Umgebung jedes Punktes des Regularitätsgebietes) als Limes einer gleichmäßig konvergenten Folge von Polynomen darstellbar (Eigenschaft A). Die regulären Funktionen sind also, wenn man sich so ausdrücken darf, „verallgemeinerte" Polynome. Es ergibt sich nun die Frage, ob nicht als Grenzfunktion einer in einem gewissen Gebiet gleichmäßig konvergenten Folge von regulären Funktionen eine Funktion auftreten kann, welche nicht mehr regulär ist (ob es also nicht vielleicht möglich ist, den Begriff der regulären Funktion in dieser Hinsicht noch zu verallgemeinern). Wir wollen zeigen, daß dies nicht der Fall ist, d. h.: *Wenn eine Folge* $\{f_n(z)\}$ *von in einem Gebiet* (D) *regulären Funktionen in diesem Gebiet gleichmäßig gegen die Grenzfunktion* $f(z)$ *konvergiert,*

$$f(z) = \lim_{n\to\infty} f_n(z),$$

so ist diese in (D) *ebenfalls regulär.*[1])

Ein Beweis dieses Satzes ergibt sich leicht mit Hilfe der CAUCHYschen Integralformel. Ist nämlich (C) eine beliebige ganz in (D) verlaufende doppelpunktfreie Kurve, so ist für jedes n

$$f_n(z) = \frac{1}{2\pi i}\int_{(C)} \frac{f_n(\zeta)\,d\zeta}{\zeta - z},$$

und der Grenzübergang liefert:

$$f(z) = \frac{1}{2\pi i}\int_{(C)} \frac{f(\zeta)\,d\zeta}{\zeta - z},$$

d. h., die Funktion $f(z)$ ist im Innern von (C) nach der CAUCHYschen Integralformel darstellbar. Hieraus folgt dann aber sofort, daß $f(z)$ innerhalb (D) überall differenzierbar ist usw.

§ 15. Eigenschaften der analytischen Funktionen

Einige besonders interessante Eigenschaften der analytischen Funktionen erhält man unmittelbar aus ihrer Darstellbarkeit als Potenzreihe (TAYLORsche Reihe)

$$f(z) = \sum_{n=0}^{\infty} a_n (z-c)^n, \quad a_n = \frac{f^{(n)}(c)}{n!}, \tag{85}$$

$$|z-c| < \varrho, \quad \varrho > 0 \qquad (n = 0,\ 1,\ 2,\ \ldots). \tag{86}$$

[1]) Satz von WEIERSTRASS.

1. *Wenn die im Punkte c reguläre Funktion* (85) *selbst und alle ihre Ableitungen im Punkte c verschwinden, so ist die Funktion $f(z)$ im Kreis* (86) *identisch gleich Null.* Ist nämlich

$$f^{(n)}(c) = 0 \qquad (n = 0,\ 1,\ 2,\ \ldots), \tag{87}$$

so ist in (85) $a_n = 0$ $(n = 0, 1, 2, \ldots)$ und daher

$$f(z) \equiv 0.$$

Wir weisen darauf hin, daß man im Reellen, ohne den Begriff der Regularität einzuführen, Funktionen konstruieren kann, welche die Bedingung (87) erfüllen, aber nicht identisch verschwinden.[1])

2. Der vorangehende Satz zeigt, daß *man die Nullstellen einer analytischen Funktion nach ihrer „Vielfachheit" klassifizieren kann.* Dabei versteht man allgemein unter einer *Nullstelle* der Funktion $f(z)$ einen Punkt c, für welchen $f(c) = 0$ gilt. Ist nun $f(c) = 0$, aber $f(z)$ nicht identisch gleich Null, so gibt es auf Grund des vorangehenden Satzes eine positive ganze Zahl p, so daß

$$f^{(n)}(c) = 0 \quad \text{für} \quad n = 0,\ 1,\ \ldots,\ p-1, \quad \text{aber} \quad f^{(p)}(c) \neq 0;$$

---

[1]) Wir setzen etwa $f(x) = 0$ für $x \leqq 0$ und $f(x) = e^{-\frac{1}{x^2}}$ für $x > 0$. Dann ist definitionsgemäß $f^{(n)}(0) = 0$ für $n = 0$. Wir nehmen nun an, daß $f^{(n)}(0) = 0$, und zeigen, daß dann auch $f^{(n+1)}(0) = 0$. Hierzu bemerken wir, daß

$$\frac{f^{(n)}(h) - f^{(n)}(0)}{h} = \frac{f^{(n)}(h)}{h}.$$

Offenbar ist zunächst für $h \leqq 0$ sicher $f^{(n)}(h) = 0$, so daß für diese h

$$\lim_{h \to 0} \frac{f^{(n)}(h)}{h} = 0.$$

Wir zeigen nun, daß dies auch für $h > 0$ gilt. Hierzu beachten wir, daß $f^{(n)}(x)$ für $x > 0$ die Form $e^{-\frac{1}{x^2}} P_n\left(\frac{1}{x}\right)$ besitzt, wobei P_n ein ganzrationales Polynom ist. Setzen wir nun $u P_n(u) = Q_n(u)$, $\frac{1}{h} = u$, so finden wir, daß für $h < 1$, d. h. für $u > 1$

$$\frac{f^{(n)}(h)}{h} = e^{-u^2} Q_n(u) < e^{-u} Q_n(u)$$

ist. Hier geht nun aber auf Grund einer bekannten Eigenschaft der Exponentialfunktion (vgl. S. 72) die rechte und damit auch die linke Seite mit h gegen Null. Also ist

$$\lim_{h \to 0} \frac{f^{(n)}(h) - f^{(n)}(0)}{h} = 0,$$

was für eine gegen Null strebende Folge aus von Null verschiedenen Zahlen auch h durchläuft, d. h., $f^{(n+1)}(0) = 0$.

anderenfalls wäre nämlich $f(z) \equiv 0$. Man bezeichnet die Zahl p als *Vielfachheit* der Nullstelle c von $f(z)$. Ist nun c eine Nullstelle von $f(z)$ der Vielfachheit p, so nimmt (85) die Form

$$f(z) = (z-c)^p \left\{ a_p + \sum_{n=p+1}^{\infty} a_n (z-c)^{n-p} \right\} \qquad \left(a_p = \frac{f^{(p)}(c)}{p!} \neq 0 \right)$$

an, d. h., es ist

$$f(z) \equiv (z-c)^p f_1(z), \tag{88}$$

wobei $f_1(z)$ eine im Punkte c reguläre Funktion ist[1]), die im Punkte c von Null verschieden ist. Man sieht sofort, daß hiervon auch die Umkehrung gilt: Kann man $f(z)$ in einer gewissen Umgebung des Punktes c in der Form (88) darstellen, so besitzt $f(z)$ im Punkte c eine Nullstelle der Vielfachheit p.

Der Begriff der Vielfachheit einer Nullstelle ist für den Fall eines Polynoms bereits aus der Schule bekannt; hier haben wir ihn sinngemäß auf beliebige analytische Funktionen übertragen.

3. *Eine nicht identisch verschwindende analytische Funktion besitzt nur isolierte Nullstellen, d. h., ist die Funktion $f(z)$ im Punkte c regulär und ist sie nicht identisch gleich Null, so kann man einen Kreis $|z-c| < \delta$ ($\delta > 0$) angeben, in welchem außer c keine Nullstelle enthalten ist.*

In der Tat: Es sei c eine p-fache Nullstelle der Funktion $f(z)$. Dann gilt (88). Die Funktion $f_1(z)$ ist dabei im Punkte c stetig (weil sie dort regulär ist) und von Null verschieden (weil $f_1(c) = a_p \neq 0$). Wir betrachten nun eine Zahl $\delta > 0$ derart, daß für $|z-c| < \delta$ die Ungleichung $|f_1(z) - f_1(c)| < \varepsilon \equiv |f_1(c)|$ gilt. Dann ist $f_1(z) \neq 0$ für alle z mit $|z-c| < \delta$,[2]) woraus alles Weitere zu ersehen ist.

4. *Nehmen die im Gebiet (D) regulären Funktionen $f(z)$ und $g(z)$ in allen Punkten eines gewissen Teilgebietes (D_1) von (D) oder in allen Punkten einer gewissen ganz in (D) liegenden Kurve (C) jeweils denselben Wert an, so stimmen ihre Werte in allen Punkten von (D) überein.*

Wäre dies nämlich nicht der Fall, so wäre

$$h(z) \equiv f(z) - g(z)$$

eine in (D) reguläre Funktion, die in allen Punkten von (D_1) bzw. in allen Punkten der Kurve (C) verschwindet, deren Nullstellen also nicht isoliert sind, was auf Grund des vorangehenden Satzes unmöglich ist.

Aus dem zuletzt genannten Satz folgt insbesondere, daß die oben vorgenommenen Erweiterungen der Funktionen e^x, $\cos z$ und $\sin z$ ins Komplexe die einzig vernünftigen sind. Zum Beispiel haben wir (§ 3) die Funktion e^z als Summe der Reihe

$$f(z) \equiv 1 + \frac{z}{1!} + \frac{z^2}{2!} + \cdots + \frac{z^n}{n!} + \cdots \tag{89}$$

---

[1]) Weil sie als Potenzreihe darstellbar ist.

[2]) Ist $|b-a| < |a|$ und $a \neq 0$, so ist $b \neq 0$.

definiert. Diese Summe ist nun 1. eine in der ganzen Ebene reguläre Funktion und stimmt 2. für reelle Werte der Veränderlichen $z = x + iy$ (d. h. für $y = 0$) mit der aus der Elementarmathematik geläufigen Funktion

$$f(x) \equiv e^x \qquad (-\infty < x < +\infty)$$

überein (vgl. S. 452).

Aus Eigenschaft 4 folgt nun ohne weiteres, daß die Funktion $f(z)$ die einzige Funktion ist, welche die angegebenen Bedingungen 1 und 2 erfüllt, und zwar sogar in der schwächeren Form, daß 1 nur in einem gewissen Gebiet (D) gilt, welches einen gewissen Abschnitt der reellen Achse in seinem Innern enthält, auf welchem die Funktion mit der Funktion e^x übereinstimmt (Abb. 9).

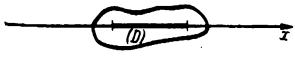


Abb. 9

Wir wollen hier noch einige weitere Anwendungsbeispiele für Eigenschaft 4 bringen.

In der elementaren Trigonometrie wird die Beziehung

$$\sin(a+b) = \sin a \cos b + \cos a \sin b \tag{90}$$

zunächst für den Fall $a > 0$, $b > 0$, $a + b < \frac{\pi}{2}$ bewiesen und dann durch eine Reihe von mehr oder weniger komplizierten Überlegungen auf alle möglichen reellen Werte von a und b ausgedehnt. Mit Hilfe von Eigenschaft 4 gelingt es uns jetzt auf einen Schlag, die Beziehung sofort auf alle komplexen Zahlen auszudehnen. Es sei nämlich zunächst b eine beliebige reelle Zahl mit $0 < b < \frac{\pi}{2}$. Wir betrachten dann die Funktion

$$f(z) \equiv \sin(z+b) - \{\sin z \cos b + \cos z \sin b\}. \tag{91}$$

Sie ist, wie man sofort sieht, in der ganzen Ebene regulär und auf Grund von (90) im Intervall $0 < x < \frac{\pi}{2} - b$ identisch gleich Null. Daraus folgt, daß die Beziehung (90) für $0 < b < \frac{\pi}{2}$ und alle komplexen Werte a richtig ist. Da nun die Formel für $f(z)$ in z und b symmetrisch ist, ist also $f(z)$ für beliebiges komplexes b und alle z, welche die Ungleichung $0 < z < \frac{\pi}{2}$ erfüllen, gleich Null. Das hat dann aber zur Folge, daß $f(z)$ für beliebiges z und b gleich Null ist, so daß die Beziehung (90) in der ganzen komplexen Ebene gilt.

Hier noch ein weiteres Beispiel:

Nehmen wir an, daß die Beziehung

$$a^{\frac{1}{p}} \cdot a^{\frac{1}{q}} = a^{\frac{1}{p}+\frac{1}{q}}$$

für alle positiven ganzzahligen p und q bewiesen ist, so können wir sie sofort auf den Fall beliebiger komplexer Exponenten z_1, z_2 ausdehnen, erhalten also die Identität

$$a^{z_1} a^{z_2} = a^{z_1+z_2}.$$

Hierzu genügt es, zu beachten, daß alle Zahlen der Form $\frac{1}{p}$ Nullstellen der in der ganzen Ebene regulären Funktion

$$f(z) \equiv a^z a^{\frac{1}{q}} - a^{z+\frac{1}{q}}$$

sind und daß die Menge aller dieser Zahlen die Zahl $z = 0$ als Häufungspunkt besitzt.

Die Eigenschaft 4 liegt auch der sogenannten analytischen Fortsetzung einer Funktion zugrunde.

Wir setzen dabei voraus, daß uns eine gewisse Funktion $f(z)$ vorgegeben ist, die in einem Gebiet (D) definiert und dort regulär ist. Wir setzen ferner voraus, daß (D) nicht die ganze Ebene ist und daß uns ein Gebiet (D_1) vorgegeben ist, welches das Gebiet (D) umfaßt (Abb. 10). Es ergibt sich dann das Problem, ob man eine Funktion $f_1(z)$ finden kann, welche in (D_1) regulär ist und welche in (D) mit $f(z)$ übereinstimmt. Es zeigt sich, daß es eine derartige Funktion nicht zu geben braucht. Wenn sie aber existiert, so ist sie auf Grund von Eigenschaft 4 eindeutig bestimmt. Das hat zur Folge, daß man dann nicht zwischen den Funktionen $f(z)$ und $f_1(z)$ zu unterscheiden braucht, d. h., man kann die Werte der Funktion $f_1(z)$ in den Punkten von (D_1), welche nicht zu (D) gehören, als Werte der Funktion $f(z)$ in diesen Punkten auffassen. Läßt man dies als „Prinzip der analytischen Fortsetzung" in der Funktionentheorie zu, so kann man unter Umständen auch von dem Wert einer Funktion in solchen Punkten sprechen, die außerhalb des ursprünglichen Definitionsbereichs der Funktion liegen. Andererseits ist es aber in der Funktionentheorie auch üblich, eine Funktion nur in den Punkten als definiert anzusehen, in welchen die Funktion regulär ist (vgl. § 14).

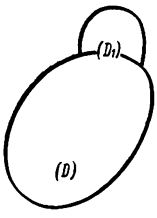


Abb. 10

Man nimmt also in der Funktionentheorie an, daß eine Funktion außer in dem Gebiet, in dem sie unmittelbar gegeben ist, auch überall dort definiert ist, wohin man sie analytisch fortsetzen kann. Die Gesamtheit aller dieser Punkte macht erst das *volle Existenzgebiet* der analytischen Funktion aus.

Dem Prinzip der analytischen Fortsetzung widerspricht es in keiner Weise, daß eine analytische Funktion „Verzweigungen" und damit Mehrdeutigkeiten besitzt. Es kann nämlich z. B. vorkommen, daß man eine Funktion aus dem doppelt schraffierten Gebiet $(D) = abcda$ (Abb. 11) in das senkrecht schraffierte Gebiet $(D_1) = aefghibcda$ und ebenso in das waagerecht schraffierte Gebiet $(D_2) = abcjhkfda$ analytisch fortsetzen kann, wobei die Funktionswerte in dem Gebiet $fghkf$ bei beiden Überführungen durchaus nicht übereinzustimmen brauchen.

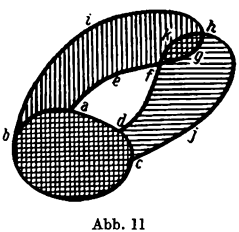


Abb. 11

Die folgenden Beispiele sollen noch deutlich den krassen Unterschied zwischen der Natur der Funktionen einmal vom Standpunkt der Theorie der Funktionen einer reellen Veränderlichen (TRF) und andererseits vom Standpunkt der Funktionentheorie (FT) aus hervorheben.

Beispiel 1. Betrachtet man die Funktion $y = x^2$ im Intervall $-1 \leqq x \leqq +1$ vom Standpunkt der TRF aus, so ist sie in diesem Intervall und sonst nirgends definiert.

Betrachtet man sie dagegen vom Standpunkt der FT aus, so kann man sie nicht nur auf die ganze reelle Achse, sondern sogar in die ganze komplexe Ebene hinein fortsetzen, und zwar ist diese Fortsetzung auf eine und nur eine Weise möglich.

Beispiel 2. Betrachtet man die Funktion $y = |x|$ im Intervall $-1 \leq x \leq +1$ vom Standpunkt der TRF aus, so gibt es hier nur eine Funktion, die im ganzen Intervall erklärt ist. Vom Standpunkt der FT aus liefert die analytische Fortsetzung von der rechten Halbgeraden in die ganze Ebene die Funktion $w = z$, während die Fortsetzung von der linken Halbgeraden aus die Funktion $w = -z$ ergibt. Hier kommt man also zu zwei verschiedenen analytischen Funktionen, d. h. zu keiner Fortsetzung der gegebenen Funktion.

Beispiel 3. Wir betrachten die Gleichung $y^2 = x$. In der TRF gibt es zwei verschiedene auf der Halbachse $0 \leq x < \infty$ definierte Funktionen, welche dieser Gleichung genügen, nämlich $y = \sqrt{x}$ und $y = -\sqrt{x}$ (die Wurzeln hier im arithmetischen Sinne verstanden). In der FT gibt es nur eine Funktion, welche dieser Gleichung genügt, die Funktion $w = \sqrt{z}$ (die Wurzel ist hier im algebraischen Sinne zu verstehen), die für jedes komplexe $z \neq 0$ zwei Werte besitzt. Die beiden Zweige gehen in diesem Fall bei analytischer Fortsetzung um die „Singularität" $z = 0$ ineinander über.

Beispiel 4. Entsprechendes gilt für die Funktion $y = \sqrt{1 - x^2}$, welche Lösung der Gleichung $x^2 + y^2 = 1$ ist (vgl. S. 7 und S. 203). Vom Standpunkt der FT aus erhalten wir auch hier nur die eine Funktion $w = \sqrt{1 - z^2}$ (die Wurzel ist im algebraischen Sinne zu verstehen), welche für jeden von $+1$ und -1 verschiedenen Argumentwert z zwei Werte besitzt. Hier geht ein Zweig in den anderen bei Umlauf um einen der Punkte $+1$ oder -1 über.

§ 16. Die geometrische Bedeutung der analytischen Funktionen

Es sei $w = f(z)$ eine in einem gewissen Gebiet (D) der z-Ebene definierte Funktion. Dann wird (vgl. § 1) vermöge der Abbildung $f(z)$ jedem Punkt z aus dem Gebiet (D) ein gewisser Punkt der w-Ebene zugeordnet. Man kann also kurz sagen, daß die Funktion $w = f(z)$ eine geometrische Abbildung des Gebietes (D) der z-Ebene in die w-Ebene erzeugt.

Uns interessiert nun zunächst die Frage, welche geometrische Eigenschaft der Regularität der Funktion $f(z)$ entspricht.

Die Regularität der Funktion $f(z)$ ist dafür charakteristisch, daß die Funktion $f(z)$ in allen Punkten des Gebietes (D) differenzierbar ist. Wir brauchen also nur zu untersuchen, was diese letzte Eigenschaft geometrisch bedeutet. Dazu wollen wir zunächst feststellen, welche geometrische Bedeutung a) der Betrag $|f'(z)|$ und b) das Argument $\arg f'(z)$ der Ableitung besitzen. Zu diesem Zweck setzen wir wie üblich

$$\Delta z = h, \quad \Delta w = \Delta f(z) = f(z + h) - f(z)$$

und außerdem

$$\Delta z = \varrho e^{i\varphi},$$

so daß also

$$|\Delta z| = \varrho, \quad \arg \Delta z = \varphi.$$

Gemäß der Definition der Ableitung existiert der Grenzwert

$$w' = \lim \frac{\Delta w}{\Delta z},$$

ist endlich und hängt nicht davon ab, auf welche Weise Δz gegen Null strebt. Dann existieren insbesondere die Grenzwerte

$$\text{a)}\ |w'| = \lim \frac{|\Delta w|}{|\Delta z|} \quad \text{und} \quad \text{b)}\ \arg w' = \lim \{\arg \Delta w - \arg \Delta z\},$$

sind endlich und vom Winkel φ unabhängig.[1])

Es bedeutet also keine Einschränkung der Allgemeinheit, wenn wir beim Differentiationsprozeß den Punkt $z + h$ geradlinig gegen den Punkt z streben

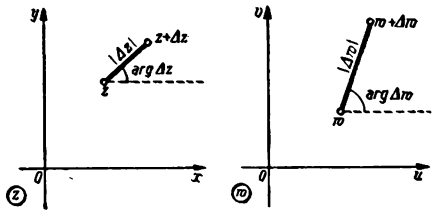


Abb. 12

lassen, d. h. also so, daß das Argument φ des Zuwachses Δz konstant bleibt, während sein Betrag ϱ gegen Null strebt.

Ist nun w' von Null verschieden, so ist (jedenfalls für hinreichend kleines Δz) auch Δw von Null verschieden, und der Quotient $\frac{|\Delta w|}{|\Delta z|}$ liefert das Verhältnis zwischen dem Abstand der Punkte $w + \Delta w$ und w einerseits und dem der Punkte $z + \Delta z$ und z andererseits (Abb. 12). Wir wollen daher den Grenzwert $|w'|$ als *Maßstab* der betrachteten Abbildung im Punkte z bezeichnen. Entsprechend zeigt die Differenz $\arg \Delta w - \arg \Delta z$ an, um welchen Winkel man den Vektor vom Punkte z zum Punkte $z + \Delta z$ drehen muß, um in die Richtung des Vektors vom Punkte w zum Punkte $w + \Delta w$ zu gelangen. Wir wollen daher den Grenzwert $\arg w'$ die *Drehung* der betrachteten Abbildung im Punkte z nennen.

Wir können damit feststellen, daß *bei einer im Gebiet* (D) *regulären Funktion* $w = f(z)$ *für die von ihr erzeugte Abbildung der Maßstab und die Drehung in keinem Punkte des Gebietes* (D) *davon abhängen, aus welcher Richtung sich der Punkt* $z + \Delta z$ *dem Punkt* z *nähert.*

---

[1]) Vgl. Seite 477 ff.

Wir wollen uns nun vorstellen, daß sich unabhängig voneinander zwei Punkte $z_1 = z + \Delta z_1$ und $z_2 = z + \Delta z_2$ aus verschiedenen Richtungen dem Punkte z nähern. Wir denken uns ferner in der w-Ebene die entsprechenden Punkte $w_1 = w + \Delta w_1$ und $w_2 = w + \Delta w_2$ markiert (Abb. 13). Die Punkte w_1 und w_2 nähern sich dann bei Annäherung von z_1 und z_2 gegen z beide dem Punkt w, allerdings im allgemeinen nicht geradlinig, sondern auf gewissen Kurven, welche im Punkte w gewisse Tangenten wT_1 und wT_2 besitzen. Wir betrachten nun die Dreiecke zz_1z_2 und ww_1w_2. Wegen der Unabhängigkeit des Maßstabes von der Richtung der Annäherung stimmen die Seitenverhält-

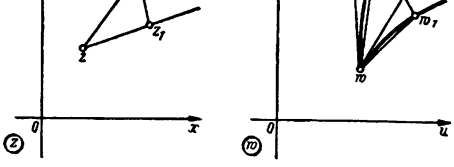


Abb. 13

nisse $ww_1 : zz_1$ und $ww_2 : zz_2$ im Limes überein, d. h., die Seitenlängen sind im Limes proportional. Dann sind aber, wie man sagen könnte, die Dreiecke zz_1z_2 und ww_1w_2 „im Limes ähnlich". Die von einer regulären Funktion $w = f(z)$ erzeugte Abbildung ist also „in der unmittelbaren Umgebung" jedes Punktes z aus (D) („im Infinitesimalen") eine Ähnlichkeitstransformation. Diese bemerkenswerte Eigenschaft der von regulären Funktionen erzeugten Abbildungen bezeichnet man als *Konformität*.

Man kann nun zeigen, daß *auch umgekehrt jede konforme Abbildung eines gewissen Gebietes (D) der z-Ebene von einer gewissen Funktion $w = f(z)$ erzeugt wird, die in diesem Gebiet regulär ist.*[1])

Folglich kann man die analytischen Funktionen auch charakterisieren als diejenigen Funktionen, deren zugehörige Abbildungen konform sind (vgl. § 14).

---

[1]) Unter der zusätzlichen Voraussetzung, daß die Abbildung die Orientierung erhält, d. h., daß der Umlaufsinn des Bildes einer geschlossenen Kurve stets gleich dem der gegebenen Kurve ist.

§ 17. Beispiele für konforme Abbildungen

Beispiel 1. $w = z + c$.

Diese Abbildung bedeutet geometrisch die Parallelverschiebung um den Vektor $\overrightarrow{Oc}$. Setzt man nämlich $c = a + ib$ und spaltet man die gegebene Gleichung in Real- und Imaginärteil auf, so findet man

$$\begin{cases} u = x + a, \\ v = y + b, \end{cases}$$

und das sind die aus der analytischen Geometrie bekannten Formeln für die Parallelverschiebung.

Beispiel 2. $w = mz$ mit $|m| = 1$, $m = e^{i\omega}$. Diese Abbildung bedeutet geometrisch die Drehung um O mit dem Winkel ω. Spaltet man nämlich die Gleichung

$$w = e^{i\omega} z$$

in Real- und Imaginärteil auf, so findet man

$$\begin{cases} u = x \cos\omega - y \sin\omega, \\ v = x \sin\omega + y \cos\omega, \end{cases}$$

und das sind gerade die aus der analytischen Geometrie bekannten Formeln für die Drehung um den Winkel ω.

Beispiel 3. $w = mz$, wobei m eine positive reelle Zahl ist.

Diese Abbildung ist wegen

$$\begin{cases} u = mx, \\ v = my \end{cases}$$

eine Dehnung in beiden Achsenrichtungen mit dem Dehnungsfaktor m (d. h. eine Streckung auf das m-fache, eine Ähnlichkeitstransformation ohne Drehung).

Beispiel 4. $w = mz$, wobei m eine beliebige von Null verschiedene komplexe Zahl ist.

Setzen wir

$$m = \mu e^{i\omega} \qquad (\mu > 0),$$

so können wir die zu untersuchende Abbildung als Hintereinanderausführung der beiden Abbildungen

$$w = \mu t,$$
$$t = e^{i\omega} z$$

ansehen. Hieraus folgt, daß diese Abbildung Hintereinanderausführung der Drehung um den Winkel ω und der Dehnung auf das μ-fache ist. Man bezeichnet eine solche Abbildung häufig als eine „loxodrome" Abbildung (Ähnlichkeitstransformation mit Drehung).

Beispiel 5. $w = mz + n$ mit $|m| = 1$, $m = e^{i\omega}$.

Diese Abbildung kann man in die beiden Abbildungen

$$w = t + n,$$
$$t = e^{i\omega} z$$

aufspalten, sie ist also Hintereinanderausführung der Drehung um den Winkel ω und der Parallelverschiebung um den Vektor $\overrightarrow{On}$.

Beispiel 6. $w = mz + n$ mit $m \neq 0$.

Mit $m = \mu e^{i\omega}$ können wir diese Abbildung in die drei Abbildungen

$$w = t_2 + n,$$
$$t_2 = \mu t_1,$$
$$t_1 = e^{i\omega} z$$

aufspalten, d. h., sie ist Hintereinanderausführung einer Drehung, einer Dehnung und einer Parallelverschiebung, also eine Ähnlichkeitstransformation (allgemeiner Fall).

Beispiel 7. $w = \frac{1}{z}$.

Eine Aufspaltung in Real- und Imaginärteil ergibt

$$\begin{cases} u = \dfrac{x}{x^2 + y^2}, \\ v = \dfrac{-y}{x^2 + y^2}. \end{cases}$$

Hieraus und aus den auf Seite 248 erhaltenen Formeln folgt, daß die betrachtete Abbildung Hintereinanderausführung der Spiegelung am Kreis $|z| = 1$ und der Spiegelung an der x-Achse ist.

Setzen wir $z = re^{i\vartheta}$, $w = Re^{i\varphi}$, so lauten die entsprechenden Formeln

$$\begin{cases} R = \dfrac{1}{r}, \\ \varphi = -\vartheta, \end{cases}$$

woraus die genannte geometrische Bedeutung noch deutlicher zu ersehen ist.

Beispiel 8. $w = z^\alpha$ $(\alpha > 0)$. Setzen wir hier $z = re^{i\vartheta}$, $w = Re^{i\varphi}$, so finden wir

$$\begin{cases} R = r^\alpha, \\ \varphi = \alpha\vartheta. \end{cases}$$

Diese Abbildung führt offenbar jeden Kreis um den Ursprung der z-Ebene in einen Kreis um den Ursprung der w-Ebene über, wobei die Änderung des Radius durch die erste Gleichung beschrieben wird. Die zweite Gleichung besagt entsprechend, daß jede Gerade durch den Ursprung der z-Ebene in eine Gerade durch den Ursprung der w-Ebene übergeführt wird, und zwar

vergrößert sich dabei der Winkel zwischen der Geraden und der positiven Richtung der x-Achse jeweils auf das α-fache. In Abb. 14 sind ein Teil des Polarnetzes in der z-Ebene und sein Bild in der w-Ebene dargestellt (für $\alpha = 2$).

Beispiel 9. $w = \operatorname{Ln} z$.

Setzen wir $z = re^{i\theta}$, $w = u + iv$ (d. h. führen wir in der z-Ebene Polarkoordinaten und in der w-Ebene kartesische Koordinaten ein), so erhalten wir (vgl. § 6)

$$\begin{cases} u = \ln r, \\ v = \theta + 2k\pi, \end{cases}$$

wobei k eine beliebige ganze Zahl ist. Das Auftreten der Konstanten k zeugt davon, daß die Abbildung der z-Ebene auf die w-Ebene unendlich vieldeutig

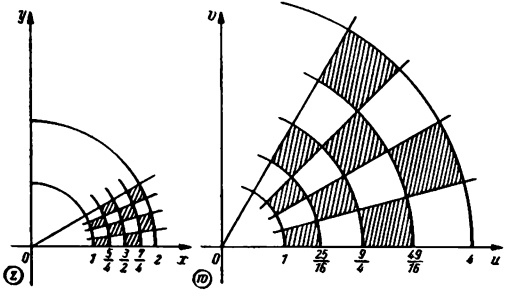


Abb. 14

ist. Setzen wir beispielsweise $k = 0$, so erhalten wir eine Abbildung der ganzen z-Ebene mit Ausnahme des Punktes $z = 0$, d. h. des Gebietes

$$0 < r < \infty, \quad 0 \leqq \theta < 2\pi,$$

auf den Streifen der w-Ebene, der durch die Ungleichungen

$$-\infty < u < +\infty, \quad 0 \leqq v < 2\pi$$

beschrieben wird. Dabei gehen offenbar die Kreise $r = r_0$ in die Geraden $u = \ln r_0$ und die Geraden $\theta = \theta_0$ in die Geraden $v = \theta_0$ über. Setzt man $k = 1$, so erhält man eine entsprechende Abbildung desselben Gebietes der z-Ebene auf den Streifen

$$-\infty < u < +\infty, \quad 2\pi \leqq v < 4\pi$$

der w-Ebene. (Analog für beliebiges k.)

Man sieht unmittelbar, daß insgesamt jeder Punkt $z = z_0 (+ 0)$ der z-Ebene in eine beiderseits unendliche „arithmetische Punktreihe" mit der „Differenz" $2\pi i$ abgebildet wird.

In Abb. 15 sind ein Teil des Polarnetzes der z-Ebene und sein Bild in der w-Ebene (in Form eines rechtwinkligen Netzes) dargestellt.

Beispiel 10. $w = \sin z$.

Hier empfiehlt es sich, in der z-Ebene kartesische Koordinaten und in der w-Ebene Polarkoordinaten einzuführen:

$$z = x + iy, \quad w = Re^{i\varphi},$$

so daß also

$$Re^{i\varphi} = \sin(x + iy),$$

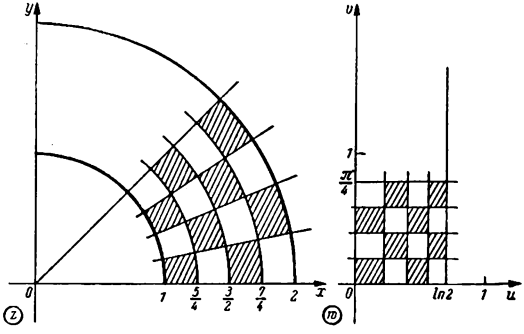


Abb. 15

d. h. (vgl. § 5)

$$Re^{i\varphi} = \sin x \operatorname{Cof} y + i \cos x \operatorname{Sin} y.$$

Hieraus folgt

$$\begin{cases} R = \sqrt{(\sin x \operatorname{Cof} y)^2 + (\cos x \operatorname{Sin} y)^2}, \\ \varphi = \operatorname{Arctg} \dfrac{\cos x \operatorname{Sin} y}{\sin x \operatorname{Cof} y}, \end{cases}$$

das heißt

$$\begin{cases} R = \sqrt{\operatorname{Cof}^2 y - \cos^2 x}, \\ \varphi = \operatorname{Arctg} \dfrac{\operatorname{Tg} y}{\operatorname{tg} x}. \end{cases}$$

Auf Grund dieser letzten Beziehung können wir nun den Betrag und das Argument von $w = \sin z$ unmittelbar aus Real- und Imaginärteil von z berechnen. Lösen wir diese Gleichungen nach y auf, so finden wir den geometrischen Ort

aller Punkte der z-Ebene, für welche a) $|\sin z|$ einen gegebenen Wert $R\,(\neq 0)$ oder b) $\arg\sin z$ einen gegebenen Wert $\varphi\,(\neq 0)$ besitzt:

$$\text{a)}\quad y = \pm \ln\left(\sqrt{R^2 + \cos^2 x} + \sqrt{R^2 - \sin^2 x}\right),$$

$$\text{b)}\quad y = \frac{1}{2} \ln \frac{1 + \operatorname{tg}\varphi \operatorname{tg} x}{1 - \operatorname{tg}\varphi \operatorname{tg} x}.$$

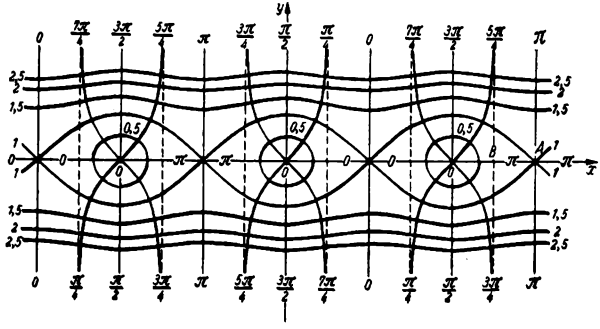


Abb. 16

Für $R = 1$ nimmt Gleichung a) beispielsweise die Form

$$y = \pm \ln\left(\cos x + \sqrt{1 + \cos^2 x}\right)$$

an; für $\varphi = \frac{\pi}{4}$ erhält man aus b)

$$y = \frac{1}{2} \ln \operatorname{ctg}\left(\frac{\pi}{4} - x\right)$$

usw.

In Abb. 16 ist ein krummliniges Netz in der z-Ebene dargestellt, welches die Kurven a) und b) enthält, die bei der betrachteten Abbildung in die w-Ebene auf die Kreise mit dem angegebenen Radius R um den Ursprung bzw. in die Geraden durch den Ursprung mit den angegebenen Winkeln φ gegen die positive reelle Achse übergeführt werden.

NAMENVERZEICHNIS

Abel, N. H. 430
Achieser, N. I. 337
d'Alembert, J. 415
Archimedes 284

Barrow, I. 374
Bernoulli, J. 5
Bernstein, S. N. 218, 285, 336
Bolzano, B. 205, 296
Bunjakowski, V. J. 285

Cauchy, A. 155, 178, 418
Cavalieri, B. 391

Dirichlet, P. Lejeune 203, 367

Egorow, D. F. 285
Ermakow, W. P. 418
Euler, L. 5, 70, 186, 285

Fermat, P. 322
Fibonacci 136
Fourier, J. B. J. 203

Gontscharow, W. L. 337

Heine, E. 178
Huygens, Ch. 340

Kazarinoff, D. K. 458
Kolmogoroff, A. N. 285, 337
Korkin, A.N. 336
Krein, M. G. 337

Leibniz, G. W. 284, 377
Leonardo von Pisa 136
Ljapunoff, A. M. 285
Lobatschewski, N. I. 4, 203, 285, 296
Losinski, S. M. 337
Lusin, N. N. 285

Markoff, A. A. 285, 336
Markow, W. A. 336
Menschow, D. E. 285

Newton, I. 284, 377, 454
Nikolski, S. M. 337

Ostrogradski, M. W. 285, 360

Petrowski, I. G. 285
Pose, K. A. 336
Priwalow, I. I. 285

Raabe, J. L. 418
Remes, E. J. 337
Riemann, B. 203, 425

Solotarjow, E. I. 336

Tschebyschew, P. L. 129, 285, 336, 357, 459

Wallis, J. 381
Weierstrass, K. 217, 296

SACHVERZEICHNIS

Abbildung 241
— durch reziproke Radien 248
—, eindeutige 241
—, eineindeutige 242
—, homöomorphe 268
—, inverse 242
—, konforme 525
—, loxodrome 526
—, mehrdeutige 241
—, stetige 265
—, topologische 268
—, umkehrbar eindeutige 242
abgeschlossene Hülle 262
— Menge 262
abgeschlossenes Intervall 144, 203, 264, 270
abhängige Veränderliche 3
ableitbare Funktion 493
Ableitung einer Funktion 291, 493
— einer Menge 261
— höherer Ordnung 316
—, linksseitige 294
—, partielle 320
—, rechtsseitige 294
Abschließung 262
absolut konvergente Reihe 421
Abstand 145, 211, 250, 252
Abwicklung 250
Additionstheorem 234, 463, 484
additive Intervallfunktion 402
Additivität eines Integrals in bezug auf das Integrationsintervall 370
ähnliche Figuren 268
Ähnlichkeitskoeffizient 268
Ähnlichkeitstransformation 526
algebraische Funktion 58
— Gleichung 58
allgemeines Glied einer Folge 133
alternierende Folge 151
— Reihe 418
Amplitude 93
analytische Fortsetzung 522
— Funktion 515
Anfangsglied(er) einer Folge 134
Anfangspunkt eines Intervalls 145
Anstieg 38
Approximation 217
Area-Funktionen 81
Areakosinus 83, 492
Areasinus, 83, 492
Areatangens 83, 492
arithmetische Folge 134, 143, 164
— — der Ordnung m 134
Arkuskosinus 121, 303, 492
Arkuskotangens 122, 303
Arkussinus 119, 303, 462, 491
Arkustangens 122, 303, 446, 492
Asymptoten 47

Bahnkurve 288
bedingt konvergente Reihe 425
beschränkte Folge 142, 255
— Funktion 206
beständig konvergente Reihe 432
bestimmtes Integral 363
Bild, graphisches 9
Binomialkoeffizient 218
Binomialreihe 454
binomische Reihe 454
binomischer Satz 317
biquadratisches Polynom 43

CAUCHYsche Integralformel 511
CAUCHYsches Integral 512
— Produkt 427
CAVALIERIsches Prinzip 392

Darstellung, graphische 9
Definitionsbereich 203
Dehnung 17
dekadischer Logarithmus 79, 300
Dezimalbruch, unendlicher 144
Dichte 290, 360
—, mittlere 290
Differential 310
—, totales 322, 328
—, vollständiges 322, 328
Differentiale höherer Ordnung 316
Differentialquotient 313
Differentiation 292
Differentiationspunkt 292
Differenzenfolge 135
differenzierbare Funktion 293, 436, 493
DIRICHLETsches Produkt 427
Diskriminante 40
Diskussion von Funktionen 27
divergente Folge 156
— Reihe 407

Drehung 24, 526
Dreiecksungleichung 253
Durchschnitt 620

e 187, 355
echte Teilmenge 260
eindeutige Abbildung 241
— Funktion 200
eineindeutige Abbildung 242
einfache harmonische Schwingung 93
einseitiger Grenzwert 176
Element, größtes 272
—, isoliertes 274
—, kleinstes 276
elementare Funktion 3, 225, 296
Ellipse 64
endliche Folge 131
— Zahlenfolge 132
Endpunkt eines Intervalls 145
euklidische Metrik 252
EULERsche Identität 482
Exponentialfunktion 70, 182, 299, 452, 480
Extremum 29, 342

fallende Funktion 28, 337
fast alle 154
FIBONACCIsche Zahlen 136
— Zahlenfolge 136
Figuren, ähnliche 268
Fläche 444
Flächeninhalt 362, 366, 390, 400
Folge 131
—, allgemeines Glied einer 133
—, alternierende 151
—, Anfangsglied(er) einer 134
—, arithmetische 134, 143, 164
—, —, der Ordnung m 134
—, beschränkte 142, 255
—, divergente 156
—, endliche 131
—, geometrische 135, 143, 164
Folge, gleichmäßig konvergente 212
—, Glied einer 131
—, konvergente 156, 255
—, Maximum einer 142
—, Minimum einer 142
—, monotone 185
—, nicht-fallende 184
—, unbeschränkte 163
—, unendliche 131
—, wachsende 184
Frequenz 93
Funktion 4, 200, 241
—, ableitbare 493
—, Ableitung einer 291, 493
—, algebraische 58
—, analytische 515
—, beschränkte 206
—, differenzierbare 293, 436, 493
—, durch Polynome approximierbare 505
—, eindeutige 200
—, elementare 3, 225, 296
—, fallende 28, 337
—, ganze rationale 33, 58, 180
—, gebrochene lineare 49
—, — rationale 33, 55, 180
—, — —, zweiten Grades 50
—, — trigonometrische 102
—, gerade 22
—, holomorphe 515
—, hyperbolische 75, 486
—, in einem Intervall stetige 178
—, in einem Punkt stetige 177
—, integrierbare 364, 435
—, irrationale algebraische 58
—, lineare 37, 134
—, logarithmische 79, 299
—, Maximum einer 28, 342
—, mehrdeutige 7, 200
—, Minimum einer 28, 342
—, mittelbare 305
—, monotone 29, 188, 337
—, nicht-fallende 28, 188, 337
—, nicht-wachsende 29, 337
Funktion, Nullstelle einer 29
—, periodische 22, 57
—, primitive 349
—, rationale 33, 358, 473
—, reelle 242
—, regulär analytische 515
—, reguläre 515
—, stationäre 342
—, stetige 204, 225, 293, 364, 434
—, transzendente 69
—, trigonometrische 86, 180, 452, 462, 480
—, ungerade 22
—, unstetige 178
—, wachsende 27, 188, 337
—, zyklometrische 119
Funktionalgleichung 234
Funktionen, Diskussion von 27
Funktionenfolge 191
Funktionenraum 254
Funktionswert 200

ganze rationale Funktion 33, 58, 180
gebrochene lineare Funktion 49
— rationale Funktion 33, 55, 180
— — — zweiten Grades 50
— trigonometrische Funktion 102
Gebiet 515
geodätische Linie 252
geometrische Folge 135, 143 164
— Reihe 409
Geschwindigkeit 287, 361
—, mittlere 287
Gerade 35
gerade Funktion 22
gleichmäßig konvergente Folge 212
gleichmäßige Stetigkeit 209
Gleichung, algebraische 58
—, trigonometrische (goniometrische) 113
Glied einer Folge 131
— einer Reihe 406
global 405

goniometrische Gleichung 113
Grad 33, 58
graphische Darstellung 9
graphisches Bild 9
Grenze, obere 273
—, untere 276
Grenzfunktion 192
Grenzwert einer Folge 154, 155, 163, 184, 255
— einer Funktion 165, 176, 493
— einer Menge 276
— einer Zahlenfolge 476
—, einseitiger 176
—, linksseitiger 176, 188
—, rechtsseitiger 176, 188
—, uneigentlicher 163
—, zweiseitiger 176
größtes Element einer Menge 272
große Rechteckformel 385

halboffenes Intervall 203
Häufungsgrenze, untere (obere) 270
Häufungspunkt 149, 163, 255, 259, 261
Häufungswert 255, 261
—, unterer (oberer) 275
harmonische Reihe 408
— Schwingung, einfache 93
— —, zusammengesetzte 95
Hauptachsengleichung 64
Hauptsatz der Differential- und Integralgleichung 374
Hauptwert 120
holomorphe Funktion 515
homöomorphe Abbildung 268
— Mengen 268
Homöomorphismus 268
L'Hospitalsche Regel 327
Hülle, abgeschlossene 262
Hyperbel 47, 64
— höherer Ordnung 48
—, Mittelpunkt einer 49
hyperbolische Funktion 75, 486
hyperbolischer Kosinus, 77, 487
— Sinus, 77, 487
— Tangens 77, 487

Imaginärteil 474
Index 131
Infimum 276
innerer Punkt eines Intervalls 145
Integral, bestimmtes 363
—, unbestimmtes 349
Integralsumme 363
Integration, partielle 354, 379
Integrationskonstante 350
integrierbare Funktion 364, 435
Interpolation 216
Intervall 203
—, abgeschlossenes 144, 203, 264, 270
—, halboffenes 203
—, nach oben nicht beschränktes 162
—, offenes 203, 270
Intervalls, Anfangspunkt eines 145
—, Endpunkt eines 145
—, innerer Punkt eines 145
—, Länge eines 145
—, Mittelpunkt eines 145
—, Randpunkt eines 145
Intervallfunktion 401
—, additive 402
inverse Abbildung 242
Inversion am Kreis 248, 269
— an der Kugel 248
irrationale algebraische Funktion 58
Isobare 249
isoliertes Element einer Menge 274
Isotherme 249

Kettenregel 305
kleine Rechteckformel 383
kleinstes Element 276
Koeffizienten einer Reihe 428
kompakte Menge 256
Kompaktheit 259
komplexe Zahl 473
konforme Abbildung 52
konjugiert komplexe Zahl 473
Kontinuum 264
konvergente Folge 156, 255, 476
— Reihe 407, 479
Konvergenz, einfache 191
Konvergenzbereich 193
Konvergenzintervall 432
Konvergenzradius 431
Kosekans 102
Kosinus 86, 299, 452, 463, 480
—, hyperbolischer 77, 487
Kosinuskurve 92
Kotangens 102, 299
Kreisfunktionen 87
krummliniges Trapez 362
kubische Parabel 61
Kurve 497
—, Länge einer 398
—, rektifizierbare 398
—, sinusförmige 93
—, stetige 204
Kurvenintegral 497

Länge eines Intervalls 451
— einer Kurve 398
Lagrangesches Interpolationspolynom 133
— Restglied 332
leere Menge 260
Leibnizsche Regel 419
Leibniz-Newtonsche Formel 377
Limes einer Folge 154, 155, 163
— inferior 276
— superior 275
lineare Funktion 37, 134
linksseitige Ableitung 294
linksseitiger Grenzwert 176, 188
logarithmische Funktion 79, 299
Logarithmus 79, 183, 300, 438, 488

Logarithmus, dekadischer 79, 300
—, natürlicher 83, 187, 300
lokal 405
loxodrome Abbildung 526

Majorante 414
Masse 360
Maximum einer Folge 142
— einer Funktion 28, 342
— einer Menge 272
mehrdeutige Abbildung 241
— Funktion 7, 200
Menge, abgeschlossene 262
—, Ableitung einer 261
—, kompakte 256
—, leere 260
—, Maximum einer 272
—, Minimum einer 276
—, nach oben beschränkte 272
—, nach unten beschränkte 276
—, zusammenhängende 263
Mengen, homöomorphe 268
Metrik 252
—, euklidische 252
Minimum einer Folge 142
— einer Funktion 28, 342
— einer Menge 276
Minkowskische Ungleichung 253
Minorante 414
mittelbare Funktion 305
Mittelpunkt einer Hyperbel 49
— eines Intervalls 145
Mittelwertsatz der Differentialrechnung 324
—, erster, der Integralrechnung 371
—, verallgemeinerter, der Differentialrechnung 324
mittlere Dichte 290
— Geschwindigkeit 287
Moivresche Formel 484
monotone Folge 185
— Funktion 29, 188, 337
Multiplikationstheorem 234

nach oben beschränkte Menge 272
nach unten beschränkte Menge 276
natürlicher Logarithmus 83, 187, 300
nicht-absolut konvergente Reihe 422
nicht-fallende Funktion 28, 184, 188, 337
nicht-wachsende Funktion 29, 337
Niveaufläche 249
Nullstelle 29, 519
—, Vielfachheit einer 520

obere Grenze 273
oberer Häufungswert 275
Oberfläche 400
Obersumme 364
offenes Intervall 203, 270
Ordnung 59
— eines trigonometrischen Polynoms 96

Parabel 35, 60
—, kubische 61
—, Scheitel einer 35
Parallelverschiebung 526
Partialsumme 407, 479
partielle Ableitung 320
— Integration 354, 379
periodische Funktion 22, 87
Phase 93
Polynom 33, 58, 180, 217, 308, 502
—, biquadratisches 43
— dritten Grades 40
—, trigonometrisches 96, 181
— zweiten Grades 39
Potenzfunktion 34, 46, 61, 84, 184, 298, 428, 489
primitive Funktion 349
Produkt von Reihen 426
punktweise Konstruktion 5
π 446

Quotientenkriterium 416, 422

Randpunkte eines Intervalls 145
Raum, topologischer 258
Raumkurve 243
rationale Funktion 33, 358, 473
Realteil 474
Rechteckformel, große 385
—, kleine 383
rechtsseitige Ableitung 294
rechtsseitiger Grenzwert 176, 188
reelle Funktion 242
regulär analytische Funktion 515
reguläre Funktion 515
Reihe 406
—, absolut konvergente 421
—, alternierende 418
—, bedingt konvergente 425
—, beständig konvergente 432
—, binomische 454
—, divergente 407
—, geometrische 409
—, Glieder einer 406
—, harmonische 408
—, Koeffizienten einer 428
—, konvergente 407, 479
— mit positiven Gliedern 412
—, nicht-absolut konvergente 422
—, Rest einer 411
—, Summe einer 407, 479
—, Taylorsche 451, 514
—, Umordnung einer 423
—, unbedingt konvergente 425
—, unendliche 406, 479
Reihen, Produkt von 426
rektifizierbare Kurve 398
Rekursionsformel 134
Rest einer Reihe 411
Restglied Lagrangesches 332
Rotationsfläche 399
Rotationskörper 397

Satz von Bernstein 218
— von Bolzano und Weierstrass 149, 163, 255

Satz von CAUCHY 324
— von FERMAT 322, 342
— von LAGRANGE 324
— von MORERA 517
— von RIEMANN 425
— von ROLLE 323
— von TSCHEBYSCHEW 360
— von WEIERSTRASS 184, 518
— von WEIERSTRASS-BERNSTEIN 215
Scheitel einer Parabel 35
Schraubenlinie 243
Schwingung, einfache harmonische 93
—, zusammengesetzte harmonische 95
Sekans 102
Signumfunktion 197
Sinus 86, 299, 452, 463, 480
—, hyperbolischer 77, 487
sinusförmige Kurve 93
Sinuskurve 92
Spiegelung 17
Stammfunktion 349, 505
stationäre Funktion 342
Stauchung 17
Steigung 38
Steigungskoeffizient 38
stetige Abbildung 265
— Funktion 204, 225, 293, 364, 434
— Kurve 204
Stetigkeit 496
—, gleichmäßige 209
Substitution 352, 378
Summe einer Reihe 407, 479
Supremum 273
Symmetrieachse 22
Symmetriezentrum 22

Tangens 102, 299, 467, 486
Tangente 288, 293
TAYLORsche Formel 329
— Reihe 451, 514
Teilfolge 157
Teilmenge 260
—, echte 260
Teilsumme 407
topologische Abbildung 268
topologischer Raum 258
totales Differential 320, 328
Translation 17
transzendente Funktion 69
Trapez, krummliniges 363
trigonometrische Funktion 86, 180, 452, 462, 480
— (goniometrische) Gleichung 113
trigonometrisches Polynom 96, 181
TSCHEBYSCHEWsche Polynome 98

überall dichte Punktmenge 226
Umgebung eines Punktes 145, 162, 251, 253, 477
umkehrbar eindeutige Abbildung 242
Umkehrfunktion 25, 231, 302
Umordnung einer Reihe 423
unabhängige Veränderliche 3
unbedingt konvergente Reihe 425
unbeschränkte Folge 163
unbestimmtes Integral 349
uneigentlicher Grenzwert 163
unendliche Folge 131
— Reihe 406, 479
unendlicher Dezimalbruch 144
ungerade Funktion 22
unstetige Funktion 178
untere Grenze 276
unterer Häufungswert 276
Untersumme 364

Veränderliche, abhängige 3
—, unabhängige 3
Vereinigung 260
Verfahren der punktweisen Konstruktion 5
Vergleichskriterium 414
Verschiebung 17
Vielfachheit einer Nullstelle 520
vollständiges Differential 322, 328

wachsende Folge 184
— Funktion 27, 188, 337
WALLISsche Formel 380
Wendepunkt 343

Zahlen, FIBONACCIsche 136
Zahlenfolge 132
—, endliche 132
—, FIBONACCIsche 136
—, konvergente 476
Zuordnung 4, 200, 241
zusammengesetzte harmonische Schwingung 95
zusammenhängende Menge 263
zweiseitiger Grenzwert 176
zyklometrische Funktion 119