Differential- und Integralrechnung

Von G. M. Fichtenholz

Mit 168 Abbildungen

Dreizehnte Auflage

VEB Deutscher Verlag der Wissenschaften
Berlin 1989

Г. М. Фихтенгольц
Курс дифференциального и интегрального исчисления
Том I
Физматгиз, Москва 1959

Die Übersetzung aus dem Russischen besorgte: Ludwig Boll, unter Mitarbeit von Renate Helle und Gisela Kürschner
Verfasser des Abschnitts Nr. 257: Dieter Ilse

ISBN 3-326-00398-6

ISSN 0073-2842

Verlagslektoren: Brigitte Mai, Erika Arndt
Umschlaggestaltung: Hartwig Hoeftmann

VEB Deutscher Verlag der Wissenschaften, DDR-1080 Berlin, Postfach 1216
Lizenz-Nr. 206 · 435/102/89
Printed in the German Democratic Republic
Gesamtherstellung: VEB Druckhaus „Maxim Gorki“, DDR-7400 Altenburg
LSV 1034
Bestellnummer: 5717048
02970

Vorwort

Es ist ein Wagnis — die Herausgeber sind sich dessen wohl bewußt —, neben so vielen ausgezeichneten Lehrbüchern der Analysis ein weiteres über diesen Gegenstand vorzulegen. Sie glauben jedoch aus guten Gründen, den „Fichtenholz“ dem Leser aus dem deutschen Sprachgebiet nicht vorenthalten zu dürfen.

Das Werk ist das Ergebnis der langjährigen erfolgreichen Tätigkeit eines Hochschullehrers, der neben seiner fruchtbaren wissenschaftlichen Arbeit einen großen Teil seiner Lebensaufgabe in der mathematischen Ausbildung des akademischen Nachwuchses sah. Es kam dem Autor darauf an, nicht nur die jungen Mathematiker, sondern auch die exakten Naturwissenschaftler, vor allem die Physiker und theoretisch arbeitenden Ingenieurwissenschaftler, in seine Wissenschaft einzuführen und dafür zu begeistern. Daß dies nicht einfach ist, wird jeder Dozent, der einmal Anfängervorlesungen gehalten hat, bestätigen. Daß es gelungen ist, beweisen die hohen Auflageziffern.

Durch behutsames Heran- und Hineinführen in die Analysis gelingt es, dem Leser die Angst vor für ihn neuen Abstraktionen und Begriffsbildungen zu nehmen und ihm beispielsweise spielend die gefürchtete „Epsilontik“ nach und nach in Fleisch und Blut übergehen zu lassen. Die zahlreichen sorgfältig ausgewählten Beispiele, die vielfach physikalischen Anwendungsgebieten entnommen sind, unterbrechen genau dort den Aufbau der Theorie, wo der Leser eine Ruhepause braucht.

Ein besonderer Vorzug des Werkes scheint den Herausgebern auch darin zu liegen, daß eine Reihe von Gegenständen, die in den Anfängervorlesungen und in den meisten Lehrbüchern keinen oder nur wenig Platz finden, behandelt werden, ohne daß der Leser dabei überfordert wird. Hierher gehören etwa weitgehende Kriterien für Grenzübergänge unter dem Integralzeichen (übrigens ein Forschungsgebiet des Autors), eine ausführliche Betrachtung der elliptischen Integrale, eine breitangelegte Behandlung mehrdimensionaler eigentlicher und uneigentlicher Integrale und vieles andere mehr. Auf diese Weise ist das Lehrbuch zugleich ein Nachschlagewerk von hohem Wert.

Das Werk ist nicht nur für den Studenten, der es bei der Überarbeitung seiner Vorlesungsnachschriften verwendet, von Nutzen, sondern es eignet sich überdies durchaus zum Selbststudium und könnte etwa auch als Grundlage eines Fernstudiums verwendet werden.

Den Leser, der die Materie schon kennt, mag es zunächst befremden, hie und da auf ungewöhnliche Begriffe zu stoßen, etwa die „unendlich kleinen“ oder „differentiellen“ Größen. Selbstverständlich werden solche Dinge völlig einwandfrei eingeführt, und man könnte sagen, daß gerade hierdurch bei dem fortgeschrittenen Leser das eigene Nachdenken angeregt wird und also der pädagogische Wert des Werkes einen besonderen Akzent erhält.

Wie in allen Werken dieser Reihe, soweit sie Übersetzungen sind, haben sich Übersetzer, Redakteure und Herausgeber vom Gesichtspunkt der präzisen mathematischen

Darstellung und auch der guten Lesbarkeit leiten lassen. Daß dabei hie und da die Übertragung nicht wörtlich bleiben konnte, dürfte das kleinere Übel sein.

Die Herausgeber danken den Übersetzern und Redakteuren für die großen Mühen bei der Herstellung des Manuskriptes und für die bereitwillige Erfüllung einer Reihe von Verbesserungsvorschlägen sowie dem VEB Deutscher Verlag der Wissenschaften für die Sorgfalt bei Herausgabe und Ausgestaltung des Werkes.

L. W. Kantorowitsch und I. P. Natanson meinen in ihrem Nachruf[1]) auf Fichtenholz, daß das Werk „in bezug auf Vollständigkeit, Klarheit und Straffheit des Aufbaus meisterhaft geschrieben ist und eines der besten Bücher über Analysis in der gesamten Weltliteratur darstellt". Die Herausgeber haben dem nichts hinzuzufügen und wünschen dem Buch einen großen Freundeskreis in aller Welt.

Die Herausgeber

[1]) Uspechi mat. nauk 14: 5 (89), 123–126 (1959).

Inhalt

Einführung. Die reellen Zahlen

§ 1. Der Bereich der rationalen Zahlen

1. Vorbemerkungen. Schon von der Schule her sind dem Leser die rationalen Zahlen und ihre Eigenschaften bekannt. Jedoch veranlassen uns bereits die Erfordernisse der Elementarmathematik, diesen Zahlenbereich zu erweitern. Denn nicht einmal die häufig vorkommenden Wurzeln aus ganzen positiven (natürlichen) Zahlen, z. B. $\sqrt{2}$, sind rationale Zahlen, d. h., *es existiert kein rationaler Bruch* $\frac{p}{q}$ (*wobei* p *und* q *natürliche Zahlen sind*), *dessen Quadrat gleich* 2 *ist.*

Wir beweisen dies indirekt, nehmen also an, es existiere ein solcher Bruch $\frac{p}{q}$, für den $\left(\frac{p}{q}\right)^2 = 2$ ist. Weiterhin sei dieser Bruch unkürzbar, d. h., p und q sollen keinen gemeinsamen Faktor haben. Also ist $p^2 = 2q^2$; somit ist p^2 und damit p eine gerade Zahl, $p = 2r$ (r ganz). Da p und q keinen gemeinsamen Teiler haben, ist q ungerade. Setzen wir jetzt diesen Ausdruck für p ein, so erhalten wir $q^2 = 2r^2$, und hieraus folgt, daß q eine gerade Zahl ist. Damit haben wir einen Widerspruch erhalten, und unsere Behauptung ist bewiesen.

Gleichzeitig ergibt sich folgendes: Würde man sich auf den Bereich der rationalen Zahlen beschränken, so könnte man in der Geometrie nicht mehr allen Strecken eine Länge zuschreiben.[1]) Wir betrachten z. B. das Quadrat über einer Seite der Länge 1. Die Länge seiner Diagonalen ist nicht durch eine rationale Zahl der Form $\frac{p}{q}$ auszudrücken; denn nach dem Satz von Pythagoras[2]) ist das Quadrat dieser Länge gleich 2. Wie wir gesehen haben, gibt es aber keine rationale Zahl, deren Quadrat gleich 2 ist.

In dieser Einführung stellen wir uns die Aufgabe, den Bereich der rationalen Zahlen zu erweitern, indem wir Zahlen anderer Art hinzufügen, nämlich die *irrationalen* Zahlen. Zugleich zeigen wir, daß in diesem erweiterten Bereich alle uns bekannten Eigenschaften der rationalen Zahlen, die sich auf die Grundrechenarten und den Vergleich von Zahlen beziehen, erhalten bleiben. Um diese Eigenschaften für den erweiterten Zahlenbereich nachzuweisen, werden wir ein minimales System grundlegender Eigenschaften auswählen, aus dem sich dann alle restlichen Eigenschaften als formal logische Folgerungen ergeben. Wir brauchen dann nur die Gültigkeit dieser grundlegenden Eigenschaften nachzuweisen. Hier gehen wir also nach der „axiomatischen Methode“ vor.

---

[1]) Es würden sogenannte *inkommensurable* Größen auftreten. Inkommensurabel nennt man zwei Strecken, deren Verhältnis nicht durch eine rationale Zahl ausdrückbar ist. — *Anm. d. Red.*

[2]) Pythagoras, 6. Jh. v. u. Z., griechischer Mathematiker und Philosoph. Der Satz wird zwar nach ihm benannt, ist aber viel älter.

In diesem Zusammenhang stellen wir nachstehend die grundlegenden Eigenschaften im Bereich der rationalen Zahlen zusammen. Gleichzeitig zeigen wir an einer Reihe von Beispielen, wie die anderen bekannten Eigenschaften aus diesen grundlegenden Eigenschaften rein formal hergeleitet werden können. Wenn wir im folgenden von „Zahlen“ sprechen, so verstehen wir darunter rationale Zahlen; wir bezeichnen sie mit $a, b, \ldots$

2. Die Ordnung des Bereichs der rationalen Zahlen. Wir vereinbaren zunächst folgendes: Wenn wir von *gleichen* Zahlen sprechen, so meinen wir ein und dieselbe Zahl, die in verschiedenen Formen dargestellt ist. Mit anderen Worten, für uns bedeutet der Begriff „gleich“ ($=$) soviel wie „identisch“. Daher führen wir die Eigenschaften gleicher Zahlen nicht besonders auf.

Der Bereich der rationalen Zahlen läßt sich mit Hilfe des Begriffs „größer“ ($>$) ordnen, und mit diesem Begriff hängt die erste Gruppe von Eigenschaften zusammen:

1a) *Für jedes Paar von Zahlen a und b gilt eine und nur eine der Beziehungen*

$$a = b, \quad a > b, \quad b > a.$$

1b) *Aus $a > b$ und $b > c$ folgt $a > c$* (Transitivität der Beziehung „größer“).

1c) *Wenn $a > b$ ist, so gibt es eine Zahl c derart, daß*

$$a > c \quad \textit{und} \quad c > b$$

ist.[1] („Die rationalen Zahlen liegen überall dicht.“)

Der Begriff „kleiner“ ($<$) wird entsprechend eingeführt. Man sagt, es sei $a < b$ dann und nur dann, wenn $b > a$ ist. Offenbar folgt aus $a < b$ und $b < c$, daß $a < c$ ist (Transitivität der Beziehung „kleiner“). In der Tat sind die Ungleichungen $a < b$ und $b < c$ nach Voraussetzung gleichbedeutend mit den Ungleichungen $b > a$ und $c > b$; aus diesen folgt aber nach 1b) die Beziehung $c > a$ oder, was dasselbe ist, $a < c$.

Weitere Eigenschaften des Größerbegriffs, die mit den Grundrechenarten an rationalen Zahlen zusammenhängen, werden später behandelt.

3. Addition und Subtraktion rationaler Zahlen. Die zweite Gruppe von Eigenschaften hängt mit der *Addition* zusammen, d. h. mit der Bestimmung der Summe zweier Zahlen. Zu jedem Paar von Zahlen a und b existiert eine (und nur eine) Zahl[2]), welche die *Summe* von a und b genannt und mit $a + b$ bezeichnet wird. Diese Summe hat folgende Eigenschaften:

2a) $a + b = b + a$ (Kommutativgesetz der Addition).

2b) $(a + b) + c = a + (b + c)$ (Assoziativgesetz der Addition).

Die besondere Rolle der Null wird durch die folgende Eigenschaft charakterisiert:

2c) $a + 0 = a$.

Außerdem gilt:

2d) *Zu jeder Zahl a existiert eine Zahl $-a$ (eine zu a entgegengesetzte Zahl) derart, daß $a + (-a) = 0$ ist.*

---

[1]) In diesem Fall sagt man auch, die Zahl c *liege zwischen* den Zahlen a und b. Offenbar gibt es unendlich viele solcher Zahlen c.

[2]) Wenn es eine und nur eine (d. h. genau eine) Zahl mit einer bestimmten Eigenschaft gibt, so sagt man, diese Zahl sei eindeutig bestimmt; statt „eindeutige Bestimmtheit“ gebraucht man auch die Bezeichnungen „Einzigkeit“ oder „Unität“ oder, etwas mißverständlich, „Eindeutigkeit“. — *Anm. d. Red.*

Auf Grund dieser Eigenschaften können wir zunächst die *Subtraktion* als Umkehrung der Addition behandeln. Wenn wir unter der *Differenz* der Zahlen a und b, wie üblich, eine Zahl c verstehen, für die $c + b = a$ ist,[1]) so entsteht die Frage, ob eine solche Zahl existiert und ob sie eindeutig bestimmt ist.

Wenn wir $c = a + (-b)$ setzen, so erhalten wir nach 2b), a), d), e):

$$\begin{aligned} c + b &= [a + (-b)] + b = a + [(-b) + b] \\ &= a + [b + (-b)] = a + 0 = a, \end{aligned}$$

so daß diese Zahl c wirklich der Definition der Differenz genügt.

Es sei jetzt umgekehrt c' Differenz der Zahlen a und b, also $c' + b = a$. Wenn wir auf beiden Seiten dieser Beziehung $-b$ hinzufügen und die linke Seite nach 2b, d), c) umformen, so erhalten wir

$$(c' + b) + (-b) = c' + [b + (-b)] = c' + 0 = c',$$

und wir können schließen, daß $c' = a + (-b) = c$ ist.

Somit haben wir Existenz und Eindeutigkeit der Differenz der Zahlen a und b bewiesen; man bezeichnet diese Differenz mit $a - b$.

Aus der Eindeutigkeit der Differenz ergibt sich eine Reihe von Folgerungen. Zunächst folgt aus 2c) also $0 = a - a$, und wir können schließen, daß außer der Zahl 0 keine Zahl existiert, die die Eigenschaft 2c) hat. Weiter folgt hieraus die Eindeutigkeit der zu einer Zahl entgegengesetzten Zahl: $-a = 0 - a$.

Da aus $a + (-a) = 0$ nach 2a) auch $(-a) + a = 0$ folgt, ergibt sich $a = -(-a)$, d. h., die Zahlen a und $-a$ sind zueinander entgegengesetzt. Wir geben noch eine weitere Eigenschaft entgegengesetzter Zahlen an:

$$-(a + b) = (-a) + (-b);$$

hierfür genügt es zu beweisen, daß $(a + b) + [(-a) + (-b)] = 0$ gilt, und das erhält man aus 2a), b), d), c).

Schließlich nennen wir noch eine Eigenschaft, die das Größerzeichen mit dem Pluszeichen verknüpft:

2e) *Aus $a > b$ folgt $a + c > b + c$.*

Hieraus folgt, daß man auf beiden Seiten einer Ungleichung dasselbe hinzuaddieren darf; auf diese Weise ergibt sich, daß die Ungleichungen

$$a > b \quad \text{und} \quad a - b > 0$$

gleichbedeutend sind.

Weiter folgt aus $a > b$ auch $-a < -b$. Aus $a > b$ folgt nämlich $a - b > 0$. Nun ist aber

$$\begin{aligned} a - b &= a + (-b) = (-b) + a \\ &= (-b) + [-(-a)] = (-b) - (-a), \end{aligned}$$

so daß die Ungleichung auch in der Form

$$(-b) - (-a) > 0$$

[1]) Nach 2a) kann man diese Beziehung, die die Differenz definiert, auch in der Form $b + c = a$ schreiben.

geschrieben werden kann; hieraus folgt aber

$$-b > -a \quad \text{oder} \quad -a < -b.$$

Insbesondere folgt aus $a > 0$ die Beziehung $-a < 0$, und aus $a < 0$ folgt $-a > 0$. Wenn $a \neq 0$ ist, so ist von den beiden entgegengesetzten Zahlen a und $-a$ eine und nur eine Zahl größer als 0; diese Zahl nennt man den *Absolutbetrag* (oder auch *Betrag*) der Zahl a bzw. der Zahl $-a$ und schreibt dafür

$$|a| \quad \text{bzw.} \quad |-a|.$$

Den Absolutbetrag der Zahl 0 setzt man gleich 0:

$$|0| = 0.$$

Nach der Eigenschaft 2e) kann man Ungleichungen gliedweise addieren: Aus $a > b$ und $c > d$ folgt $a + c > b + d$; denn aus $a > b$ folgt $a + c > b + c$, und aus $c > d$ folgt $c + b > d + b$, also nach 2a) $b + c > b + d$; damit erhalten wir nach 1b) schließlich $a + c > b + d$.

4. Multiplikation und Division rationaler Zahlen. Die dritte Gruppe von Eigenschaften hängt mit der *Multiplikation* zusammen, d. h. mit einer Operation, die sich mit dem Bestimmen des Produktes zweier Zahlen beschäftigt. Für jedes Paar von Zahlen a und b existiert genau eine Zahl, das *Produkt* von a und b (es wird mit $a \cdot b$ oder einfach mit ab bezeichnet). Das Produkt hat folgende Eigenschaften:

3a) $ab = ba$ (Kommutativgesetz der Multiplikation).

3b) $(ab)\,c = a(bc)$ (Assoziativgesetz der Multiplikation).

Die besondere Rolle der 1 bei der Multiplikation wird durch folgende Eigenschaft charakterisiert:

3c) $a \cdot 1 = a$.

Außerdem gilt:

3d) *Für jede von 0 verschiedene Zahl a existiert eine (zu ihr reziproke) Zahl $\frac{1}{a}$ derart, daß $a \cdot \frac{1}{a} = 1$ ist.*

Das Problem der *Division*, d. h. der zur Multiplikation entgegengesetzten Grundrechenart, kann man auf Grund der Eigenschaften der Multiplikation genauso lösen, wie wir das Problem der Subtraktion auf Grund der Eigenschaften der Addition gelöst haben. Die reziproke Zahl spielt hier dieselbe Rolle wie die entgegengesetzte Zahl bei der Addition.

Man nennt eine Zahl c, für welche

$$c \cdot b = a$$

ist, einen *Quotienten* der Zahlen a und b (wobei der Nenner b immer als von 0 verschieden vorausgesetzt wird.)[1])

Diese Definition ist erfüllt, wenn man $c = a \cdot \frac{1}{b}$ annimmt; denn nach 3b), a), d), c) ist

$$c \cdot b = \left(a \cdot \frac{1}{b}\right) \cdot b = a \cdot \left(\frac{1}{b} \cdot b\right) = a \cdot \left(b \cdot \frac{1}{b}\right) = a \cdot 1 = a.$$

[1]) Nach 3a) kann man diese Beziehung, die den Quotienten definiert, auch in der Form $b \cdot c = a$ schreiben.

Umgekehrt, wenn eine Zahl c' die Definition des Quotienten der Zahlen a und b erfüllt, wenn also $c' \cdot b = a$ ist, so erhalten wir, nachdem wir beide Seiten der Beziehung mit $\frac{1}{b}$ multipliziert und die linke Seite umgeformt haben [3b), d), c)]:

$$(c' \cdot b) \cdot \frac{1}{b} = c' \cdot \left(b \cdot \frac{1}{b}\right) = c' \cdot 1 = c',$$

also $c' = a \cdot \frac{1}{b} = c$.

Somit sind Existenz und Eindeutigkeit des Quotienten der Zahlen a und b (unter der Bedingung $b \neq 0$) bewiesen. Man bezeichnet diesen Quotienten mit $a:b$, $\frac{a}{b}$ oder a/b.

Aus der Eindeutigkeit des Quotienten folgt, daß es außer der Zahl 1 keine Zahl gibt, die einer zu 3c) analogen Eigenschaft genügt. Schließlich ergibt sich hieraus noch, daß die reziproke Zahl (als Quotient von 1 und a) eindeutig bestimmt ist; außerdem kann man sich leicht davon überzeugen, daß die Zahlen a und $\frac{1}{a}$ zueinander reziprok sind.

Die folgende Eigenschaft verknüpft die beiden grundlegenden arithmetischen Rechenoperationen, die Multiplikation und die Addition:

3e) $(a + b)\,c = ac + bc$ (Distributivgesetz der Multiplikation bezüglich der Addition).

Hieraus kann man leicht das Distributivgesetz der Multiplikation bezüglich der Subtraktion herleiten:

$$(a - b)\,c = ac - bc.$$

Nach der Definition der Differenz folgt das sofort aus

$$(a - b) \cdot c + b \cdot c = [(a - b) + b] \cdot c = a \cdot c.$$

Mit Hilfe der Eigenschaft 3e) beweisen wir noch, daß

$$b \cdot 0 = 0 \cdot b = 0$$

ist. Nach 2c) ist

$$a + 0 = a, \quad (a + 0) \cdot b = a \cdot b + 0 \cdot b = a \cdot b,$$

woraus $0 \cdot b = 0$ folgt; also ist nach 3a) auch $b \cdot 0 = 0$.

Umgekehrt, wenn $a \cdot b = 0$ und $b \neq 0$ ist, so muß $a = 0$ sein. In der Tat ist $a = \frac{0}{b}$, und gleichzeitig ist auch $0 = \frac{0}{b}$ (da $b \cdot 0 = 0$ ist); da der Quotient eindeutig bestimmt ist, ist also $a = 0$.

Schließlich geben wir noch eine Eigenschaft an, die das Größerzeichen mit dem Produktzeichen verknüpft:

3f) *Aus $a > b$ und $c > 0$ folgt $ac > bc$.*

Hierauf beruht die Möglichkeit, beide Seiten einer Ungleichung mit einer positiven Zahl zu multiplizieren; daraus folgt, daß für $a > 0$ und $b > 0$ auch $ab > 0$ ist.

Wir bemerken noch, daß $(-a) \cdot b = -(a \cdot b)$ ist; es ist nämlich

$$a \cdot b + (-a) \cdot b = [a + (-a)] \cdot b = 0 \cdot b = 0.$$

Jetzt sieht man leicht: Ist $a < 0$ und $b > 0$, also $a = -|a|$ und $b = |b|$, so ist

$$a \cdot b = (-|a|) \cdot |b| = -(|a| \cdot |b|) < 0;$$

dasselbe erhält man für $a > 0$ und $b < 0$. Wenn $a < 0$ und $b < 0$ ist, so gilt

$$\begin{aligned} a \cdot b &= (-|a|) \cdot (-|b|) = -[|a| \cdot (-|b|)] \\ &= -[-(|a| \cdot |b|)] = |a| \cdot |b| > 0. \end{aligned}$$

Somit haben wir die bekannten *Vorzeichenregeln für die Multiplikation* hergeleitet, und zwar als logische Folgerungen der angegebenen Eigenschaften der rationalen Zahlen. Mit anderen Worten, die Vorzeichenregeln erhalten wir zwangsläufig, wenn die erwähnten Eigenschaften erfüllt sind. Dasselbe kann man auch bezüglich der Regeln für die Multiplikation mit 0 sagen.

Nachdem wir die Eigenschaften der Addition und der Multiplikation angegeben haben, könnten wir jetzt die Eigenschaft beweisen, daß der Bereich der rationalen Zahlen überall dicht ist. Diese Eigenschaft formulierten wir oben als grundlegende Eigenschaft [1c]. In der Tat können wir z. B. zeigen, daß aus $a > b$ die Beziehung

$$a > \frac{a+b}{2} > b$$

folgt.

5. Das Archimedische[1]) Axiom. Wir beenden jetzt die Aufzählung der grundlegenden Eigenschaften der rationalen Zahlen mit der folgenden einfachen und wichtigen Aussage, die sich nicht aus den oben aufgeführten Eigenschaften ergibt.

4a) *Wie auch immer eine Zahl $c > 0$ gewählt wird, es existiert stets eine natürliche Zahl n, die größer als c ist* (Archimedisches Axiom).

Tatsächlich formulierte ARCHIMEDES diesen Sachverhalt als geometrischen Satz, der ebenfalls als „Archimedisches Axiom" bekannt ist:

Wenn auf einer Geraden zwei beliebige Strecken A und B gegeben sind, so können wir A so oft zu sich selbst addieren, daß die Summe größer als B wird:

$$\underbrace{A + A + \cdots + A}_{n\text{-mal}} = A \cdot n > B.$$

Wenn wir diese Aussage für positive Zahlen a und b formulieren, so liefert sie die um Existenz einer natürlichen Zahl n derart, daß

$$\underbrace{a + a + \cdots + a}_{n\text{-mal}} = a \cdot n > b$$

ist. Diese Ungleichung ist, wenn wir die Eigenschaften der rationalen Zahlen benutzen, gleichbedeutend mit $n > \frac{b}{a}$. Bezeichnen wir den Bruch $\frac{b}{a}$ mit c, so erhalten wir die obige Formulierung.

§ 2. Einführung der irrationalen Zahlen. Ordnung des Bereichs der reellen Zahlen

6. Definition der irrationalen Zahl. Die Gesamtheit der rationalen Zahlen mit allen ihren Eigenschaften, sie sie in § 1 aufgezählt sind, wird nun als gegeben vorausgesetzt.

[1]) ARCHIMEDES, um 250 v. u. Z., griechischer Mathematiker und Naturforscher.

Wir entwickeln die Theorie der irrationalen Zahlen nach der Methode von R. DEDEKIND[1]), die auf dem Begriff des Schnittes im Bereich der rationalen Zahlen beruht. Wir nennen eine Zerlegung der Menge aller rationalen Zahlen in zwei nichtleere (d. h. wenigstens eine Zahl enthaltende) Mengen A und A' einen *Schnitt*, wenn folgende Bedingungen erfüllt sind:

1. *Jede rationale Zahl liegt in einer und nur einer*[2]) *der Mengen A oder A'.*

2. *Jede Zahl a der Menge A ist kleiner als jede Zahl a' der Menge A'.*

Die Menge A nennen wir die *Unterklasse* des Schnittes, die Menge A' die *Oberklasse*. Den Schnitt bezeichnen wir mit $A \mid A'$.

Aus der Definition des Schnittes folgt, daß jede rationale Zahl, die kleiner als eine Zahl a der Unterklasse ist, ebenfalls in der Unterklasse liegt. Analog dazu liegt jede rationale Zahl, die größer als eine Zahl a' der Oberklasse ist, selbst in der Oberklasse.

Beispiel 1. Wir definieren A als Menge aller rationalen Zahlen a, die der Ungleichung $a < 1$ genügen, und zur Menge A' rechnen wir alle Zahlen a', für die $a' \geqq 1$ gilt.

Es läßt sich leicht verifizieren, daß wir auf diese Weise tatsächlich einen Schnitt erhalten. Die Zahl 1 gehört zur Klasse A' und ist offenbar die kleinste Zahl dieser Klasse. Dagegen gibt es keine größte Zahl in der Klasse A, so daß sich, wie immer auch die Zahl a aus A gewählt wird, stets eine rationale Zahl a_1 finden läßt, die zwischen ihr und 1 liegt, also größer als a ist und zur Klasse A gehört.

Beispiel 2. In die Unterklasse A nehmen wir alle rationalen Zahlen a, die kleiner als 1 oder gleich 1 sind, für die also $a \leqq 1$ gilt; in die Oberklasse A' nehmen wir alle rationalen Zahlen a', die größer als 1 sind, d. h. $a' > 1$.

Auch dies ist ein Schnitt, wobei hier in der Oberklasse keine kleinste, aber in der Unterklasse eine größte Zahl (nämlich 1) existiert.

Beispiel 3. In die Klasse A nehmen wir alle positiven rationalen Zahlen a, für die $a^2 < 2$ gilt, sowie die Zahl 0 und alle negativen rationalen Zahlen, in die Klasse A' dagegen alle positiven rationalen Zahlen a', für die $a'^2 > 2$ gilt.

Wie man sich leicht überzeugt, erhält man wieder einen Schnitt. Hier gibt es weder in der Klasse A eine größte Zahl noch in der Klasse A' eine kleinste. Wir beweisen z. B. die erste der beiden Behauptungen (der Beweis der zweiten folgt analog).

Es sei a eine beliebige positive Zahl der Klasse A, so daß $a^2 < 2$ ist. Wir zeigen, daß es möglich ist, eine ganze positive Zahl n zu finden, für die

$$\left(a + \frac{1}{n}\right)^2 < 2$$

gilt, so daß die Zahl $a + \frac{1}{n}$ zur Klasse A gehört.

Der eben genannten Ungleichung entsprechen die folgenden Relationen:

$$a^2 + \frac{2a}{n} + \frac{1}{n^2} < 2, \quad \frac{2a}{n} + \frac{1}{n^2} < 2 - a^2.$$

---

[1]) RICHARD DEDEKIND, 1831—1916, deutscher Mathematiker.

[2]) Daß jede rationale Zahl nur in eine dieser Klassen fällt, folgt übrigens aus der Forderung 2.

Die letzte Ungleichung ist sicher erfüllt, wenn n die Ungleichung

$$\frac{2a+1}{n} < 2 - a^2$$

befriedigt. Zu diesem Zweck genügt es,

$$n > \frac{2a+1}{2-a^2}$$

zu nehmen, was [nach dem Axiom von ARCHIMEDES; vgl. Nr. 5, Eigenschaft 4a)] immer möglich ist.

Zu jeder positiven Zahl a der Klasse A läßt sich also in der Klasse A eine größere Zahl finden; da für die Zahlen $a \leqq 0$ diese Behauptung trivial ist, existiert keine größte Zahl der Klasse A.

Es ist leicht einzusehen, daß *kein Schnitt existieren kann, für den gleichzeitig in der Unterklasse eine größte Zahl* a_0 *und in der Oberklasse eine kleinste Zahl* a_0' *gefunden werden kann.* Angenommen, es existierte tatsächlich ein solcher Schnitt. Dann läßt sich, da der Bereich der rationalen Zahlen dicht ist [Eigenschaft 1c)], eine rationale Zahl c angeben, die zwischen a_0 und a_0' liegt, für die also $a_0 < c < a_0'$ gilt. Die Zahl c kann nicht zur Klasse A gehören, denn sonst wäre a_0 nicht die größte Zahl in dieser Klasse. Analog kann c nicht zur Klasse A' gehören; das widerspricht aber der Eigenschaft 1 des Schnittes, die zur Definition dieses Begriffs benutzt wurde.

Somit kann man sagen, daß es nur drei Arten von Schnitten geben kann:

1. In der Unterklasse A gibt es keine größte Zahl, aber in der Oberklasse A' eine kleinste Zahl r;

2. in der Unterklasse A existiert eine größte Zahl r, aber in der Oberklasse A' keine kleinste;

3. in der Unterklasse existiert keine größte Zahl und in der Oberklasse keine kleinste.

In den ersten beiden Fällen sagen wir, *der Schnitt werde von einer rationalen Zahl r erzeugt* (die sich auf der Grenze zwischen den Klassen A und A' befindet) oder auch, der Schnitt *definiere* die rationale Zahl r. In den Beispielen 1 und 2 ist diese Zahl die Zahl 1. Im dritten Fall existiert eine solche „Randzahl" nicht, der Schnitt definiert keine rationale Zahl.

Wir führen jetzt neue Objekte ein — die *irrationalen Zahlen,* indem wir vereinbaren, jeder Schnitt der Form 3 definiere eine irrationale Zahl α. Diese Zahl α ersetzt die fehlende „Randzahl", die wir zwischen den Zahlen a der Klasse A und den Zahlen a' der Klasse A' eingefügt denken. Im Beispiel 3 ist diese eben neu eingeführte Zahl, wie man leicht errät, genau die Zahl $\sqrt{2}$.

Für die irrationalen Zahlen wird keine besondere, typische Bezeichnung[1]) eingeführt. Wir verknüpfen nach wie vor die irrationale Zahl α mit dem sie definierenden Schnitt $A \mid A'$ im Bereich der rationalen Zahlen.

Der Einheitlichkeit wegen werden wir oft genauso mit einer rationalen Zahl r verfahren. Jedoch existieren für jede rationale Zahl r zwei Schnitte, die sie definieren.

---

[1]) Hier ist von „endlichen" Bezeichnungen (d. h. Bezeichnungen in geschlossener Form) die Rede; eine Art „unendlicher" Bezeichnungen für irrationale Zahlen wird der Leser in Nr. 9 kennenlernen. Meistens werden irrationale Zahlen so bezeichnet, daß man erkennt, wie sie zustande kommen, etwa $\sqrt{2}$, log 5, sin 10°, usw.

In beiden Fällen gehören die Zahlen $a < r$ zur Unterklasse, die Zahlen $a' > r$ zur Oberklasse, jedoch kann die Zahl r nach Belieben entweder zur Unterklasse (dann ist r dort die größte Zahl) oder zur Oberklasse (dann ist r dort die kleinste Zahl) gerechnet werden.

Um etwas Bestimmtes vor Augen zu haben, setzen wir nun jedoch ein für allemal fest: Ist von einem Schnitt die Rede, der eine rationale Zahl r definiert, so soll diese Zahl r zur Oberklasse gehören.

Die rationalen und die irrationalen Zahlen werden unter der gemeinsamen Bezeichnung *reelle Zahlen* zusammengefaßt. Der Begriff der reellen Zahl ist einer der Grundbegriffe der Analysis.

7. Die Ordnung des Bereichs der reellen Zahlen. *Zwei irrationale Zahlen α und β, die durch die Schnitte $A \mid A'$ bzw. $B \mid B'$ definiert sind, heißen dann und nur dann gleich, wenn diese Schnitte identisch sind;* übrigens genügt es zu fordern, daß die Unterklassen A und B übereinstimmen, weil die Oberklassen A' und B' dann von selbst identisch sind.

Diese Definition behält auch für den Fall ihre Gültigkeit, daß die Zahlen α und β rational sind. Mit anderen Worten, *sind zwei rationale Zahlen α und β gleich, dann stimmen die sie definierenden Schnitte überein. Umgekehrt folgt aus der Identität der Schnitte die Gleichheit der Zahlen α und β.* Dabei versteht sich von selbst, daß man die Bedingung berücksichtigen muß, die wir hinsichtlich der rationalen Zahlen gemacht haben.[1])

Nun wollen wir den Begriff „größer" für reelle Zahlen definieren. Für die rationalen Zahlen wurde dieser Begriff bereits eingeführt.

Für eine rationale Zahl r und eine irrationale Zahl α wurde der Begriff „größer" eigentlich schon in Nr. 6 eingeführt. Wenn nämlich α durch den Schnitt $A \mid A'$ definiert wird, so sagen wir, α sei größer als alle rationalen Zahlen, die in der Klasse A liegen, und alle Zahlen der Klasse A' seien größer als α.

Wir betrachten jetzt zwei irrationale Zahlen α und β, wobei α durch den Schnitt $A \mid A'$ und β durch den Schnitt $B \mid B'$ definiert sei. Wir nennen jene Zahl größer, deren Unterklasse größer ist. Genauer, *wir schreiben $\alpha > \beta$ und nennen α größer als β, wenn die Klasse A die Klasse B umfaßt, ohne mit ihr identisch zu sein.* (Diese Bedingung ist offensichtlich damit äquivalent, daß die Klasse B' die Klasse A' umfaßt, ohne mit ihr identisch zu sein.) Es läßt sich leicht verifizieren, daß diese *Definition auch dann gültig bleibt, wenn eine der Zahlen α, β oder sogar beide rational sind.*

Wir zeigen, daß für die reellen Zahlen die folgenden Bedingungen 1.1 und 1.2 erfüllt sind.

1.1 *Für jedes Paar reeller Zahlen α und β gilt genau eine der Beziehungen*

$$\alpha = \beta, \quad \alpha > \beta, \quad \alpha < \beta.$$

Ist der Schnitt $A \mid A'$, der die Zahl α definiert, identisch mit dem Schnitt $B \mid B'$, der die Zahl β definiert, dann gilt $\alpha = \beta$. Sind diese Schnitte nicht identisch, dann enthält entweder A die Klasse B vollständig, und es gilt $\alpha > \beta$, oder das ist nicht der Fall, und es existiert ein Element b_0 der Klasse B, das in der Klasse A' enthalten ist. Dann gilt für jedes Element a der Klassse A die Beziehung $a < b_0$. Daher enthält die Klasse B die Klasse A, ohne mit ihr identisch zu sein, und es liegt der Fall $\beta > \alpha$ vor.

1.2 *Aus $\alpha > \beta$ und $\beta > \gamma$ folgt $\alpha > \gamma$.*

---

[1]) Ohne diese Bedingung wären z. B. die Schnitte, die wir in Nr. 6, Beispiel 1 und 2, betrachtet haben und die beide ein und dieselbe Zahl 1 definieren, nicht identisch.

Die Zahlen α, β und γ (unter denen sich auch rationale befinden können) seien durch die Schnitte $A|A'$, $B|B'$ bzw. $C|C'$ definiert. Ist $\alpha > \beta$, so enthält die Klasse A nach der Definition des Begriffs „größer" die Klasse B, ohne mit ihr identisch zu sein. Ihrerseits enthält, wenn $\beta > \gamma$ ist, die Klasse B die Klasse C, ohne mit ihr identisch zu sein. Folglich enthält die Klasse A die Klasse C vollständig, ohne mit ihr identisch zu sein, d. h., es gilt $\alpha > \gamma$.

Der Begriff „kleiner" läßt sich jetzt wie in Nr. 2 einführen. Wir sagen, es sei $\alpha < \beta$, wenn $\beta > \alpha$ ist. Ebenso wie die Beziehung „größer" ist auch die Beziehung „kleiner" transitiv.

8. **Hilfssätze.** Wir beweisen jetzt, daß der Bereich der reellen Zahlen [vgl. 1c)] dicht ist, d. h., wir zeigen:

Lemma 1. *Zu zwei beliebig gegebenen Zahlen α und β mit $\alpha > \beta$ läßt sich eine rationale Zahl r finden, die zwischen ihnen liegt, für die also $\alpha > r > \beta$ gilt* (folglich gibt es sogar eine unendliche Menge solcher rationaler Zahlen r).

Da $\alpha > \beta$ ist, umfaßt die Unterklasse A des Schnittes, der α definiert, die Unterklasse B der Zahl β, ohne mit B identisch zu sein. Daher läßt sich in der Klasse A eine rationale Zahl r angeben, die nicht in B enthalten ist, also zur Klasse B' gehört. Für diese Zahl gilt

$$\alpha > r \geqq \beta$$

(Gleichheit könnte bei rationalem β eintreten). Da jedoch in A keine größte Zahl existiert, kann man nötigenfalls, indem man r größer wählt, die Gleichheit ausschließen.

Bemerkung. Indem wir zeigten, daß zwischen den reellen Zahlen α und β (wenn $\alpha > \beta$) notwendigerweise eine rationale (und nicht nur eine reelle) Zahl liegt, haben wir faktisch eine stärkere Eigenschaft des Bereichs der reellen Zahlen bewiesen, als daß er überall dicht ist. Im folgenden werden wir diese stärkere Eigenschaft benötigen. Hieraus erhält man unmittelbar

Lemma 2. *Es seien zwei reelle Zahlen α und β gegeben. Wenn zu jeder Zahl $e > 0$ rationale Zahlen s und s' existieren, deren Differenz kleiner als e ist, für die also*

$$s' - s < e$$

ist und für die

$$s' \geqq \alpha \geqq s, \quad s' \geqq \beta \geqq s$$

gilt, dann sind α und β notwendigerweise identisch.

Den Beweis werden wir indirekt führen. Es sei etwa $\alpha > \beta$. Nach Lemma 1 lassen sich zwischen α und β zwei rationale Zahlen r und $r' > r$ einfügen, so daß $\alpha > r' > r > \beta$ ist. Daher gelten für zwei beliebige Zahlen s und s', zwischen welchen α und β liegen, offensichtlich die Ungleichungen

$$s' > r' > r > s,$$

woraus $s' - s > r' - r > 0$ folgt, so daß die Differenz $s' - s$ entgegen der Voraussetzung des Lemmas nicht kleiner gemacht werden kann als beispielsweise die Zahl $e = r' - r$. Dieser Widerspruch beweist die Richtigkeit des Lemmas.

9. Die Darstellung reeller Zahlen durch unendliche Dezimalbrüche. Wir wollen eine Darstellung betrachten, bei der der Bruchteil (die Mantisse) positiv ist, während der ganze Teil positiv oder negativ oder 0 sein kann.

Wir setzen zunächst voraus, die zu betrachtende reelle Zahl α sei weder eine ganze Zahl noch ein endlicher Dezimalbruch. Wir suchen eine angenäherte Darstellung in Form eines Dezimalbruchs. Ist α durch den Schnitt $A \mid A'$ definiert, dann läßt sich zunächst leicht verifizieren, daß sich in der Klasse A eine ganze Zahl M befindet und in der Klasse A' eine ganze Zahl $N > M$. Durch sukzessive Addition von Einsen zu M gelangt man notwendigerweise zu zwei aufeinanderfolgenden ganzen Zahlen C_0 und $C_0 + 1$ derart, daß

$$C_0 < \alpha < C_0 + 1$$

gilt. Hierbei kann C_0 positiv, negativ oder 0 sein.

Teilen wir nun das Intervall zwischen C_0 und $C_0 + 1$ mit Hilfe der Zahlen

$$C_0{,}1; \quad C_0{,}2; \quad \ldots; \quad C_0{,}9$$

in zehn gleiche Teile, dann fällt α in eines (und nur eines) dieser Teilintervalle, und wir haben zwei Zahlen gefunden, die sich um $\frac{1}{10}$ unterscheiden, nämlich $C_0{,}c_1$ und $C_0{,}c_1 + \frac{1}{10}$, und für welche $C_0{,}c_1 < \alpha < C_0{,}c_1 + \frac{1}{10}$ gilt. Setzen wir diesen Prozeß weiter fort, so bestimmen wir nach Festlegung der Ziffern $c_1, c_2, \ldots, c_{n-1}$ eine n-te Ziffer c_n durch die Ungleichungen

$$C_0{,}c_1c_2\ldots c_n < \alpha < C_0{,}c_1c_2\ldots c_n + \frac{1}{10^n}. \tag{1}$$

Auf diese Weise haben wir bei der Approximation der Zahl α durch Dezimalbrüche eine ganze Zahl C_0 und eine unendliche Folge $c_1, c_2, \ldots, c_n, \ldots$ von Ziffern gefunden.

Der aus ihnen gebildete unendliche Dezimalbruch, d. h. das Symbol

$$C_0{,}c_1c_2\ldots c_n\ldots, \tag{2}$$

kann als Darstellung der reellen Zahl α angesehen werden.

In dem bisher ausgeschlossenen Fall, daß α selbst eine ganze Zahl oder ein endlicher Dezimalbruch ist, lassen sich ebenfalls analog die Zahl C_0 und die Ziffern $c_1, c_2, \ldots, c_n, \ldots$ sukzessive bestimmen, wenn man von der Verallgemeinerung von (1), der Relation

$$C_0{,}c_1c_2\ldots c_n \leqq \alpha \leqq C_0{,}c_1c_2\ldots c_n + \frac{1}{10^n}, \tag{1a}$$

ausgeht. Dann stimmt irgendwann im Laufe des Prozesses die Zahl α mit einem der Endpunkte des Intervalls, in das wir sie eingeschlossen hatten, überein, und zwar je nach Belieben mit dem linken oder dem rechten; von da an wird bei Fortsetzung des Prozesses links bzw. rechts in der Relation (1a) bereits ständig Gleichheit herrschen.

Je nachdem, welche dieser Möglichkeiten realisiert wird, sind die folgenden Ziffern sämtlich 0 oder sämtlich 9. Somit gibt es für die Zahl α jetzt zwei Darstellungen, eine mit der Periode 0 und eine zweite mit der Periode 9, beispielsweise

$$3{,}826 = 3{,}826000\ldots = 3{,}825999\ldots,$$
$$-3{,}836 = \bar{4}{,}174000\ldots = \bar{4}{,}173999\ldots$$[1])

Es sei jetzt umgekehrt ein beliebiger unendlicher Dezimalbruch (2) vorgegeben. Wir zeigen, daß sich immer eine reelle Zahl α angeben läßt, die durch diesen Bruch dargestellt wird.

Zu diesem Zweck betrachten wir den Teil des Bruches (2),

$$C_n = C_0{,}c_1c_2\ldots c_n, \tag{3}$$

welcher beim Abrunden der gesuchten Zahl, sowie den Teil

$$C'_n = C_0{,}c_1c_2\ldots c_n + \frac{1}{10^n}, \tag{4}$$

welcher beim Aufrunden benutzt wird. Es ist leicht zu sehen, daß jedes C_n kleiner als jedes C'_m ist (nicht nur für $m = n$, sondern auch für $m \gtrless n$).

Wir wollen jetzt auf folgende Weise einen Schnitt im Bereich der rationalen Zahlen erzeugen. Zur Oberklasse A' rechnen wir die rationalen Zahlen a', die größer als alle C_n sind (z. B. alle Zahlen C'_n), und zur Unterklasse A alle übrigen (beispielsweise C_n selbst). Daß es sich hierbei um einen Schnitt handelt, ist leicht zu verifizieren; er definiert die reelle Zahl α, die zu bestimmen war. Da nämlich α die Randzahl zwischen den beiden Klassen ist, gilt insbesondere

$$C_n \leqq \alpha \leqq C'_n,$$

d. h., die Zahl α genügt allen Ungleichungen der Form (1a). Damit ist bewiesen, daß der beliebig angenommene Dezimalbruch (2) eine Darstellung der soeben bestimmten Zahl ist.

Die Differenz zwischen den beiden Näherungs-Dezimalbrüchen (4) und (3) beim Aufrunden bzw. Abrunden, die $\frac{1}{10^n}$ beträgt, kann mit wachsendem n kleiner als jede beliebige rationale Zahl $e > 0$ gemacht werden. Da es nämlich nur endlich viele natürliche Zahlen gibt, die die Zahl $\frac{1}{e}$ nicht übertreffen, gibt es auch nur endlich viele Zahlen n derart, daß die Ungleichung $10^n \leqq \frac{1}{e}$ oder die ihr äquivalente Ungleichung $\frac{1}{10^n} \geqq e$ erfüllt ist; für alle übrigen Werte von n gilt

$$\frac{1}{10^n} < e.$$

Diese Bemerkung gestattet unter Berücksichtigung von Lemma 2 die Schlußfolgerung, daß eine von α verschiedene Zahl β nicht denselben Ungleichungen (1) oder (1a) genügen kann wie α und somit eine von der Darstellung von α verschiedene Dezimalbruchdarstellung hat.

Hieraus folgt insbesondere, daß die Dezimalbruchentwicklung einer Zahl, die nicht nach endlich vielen Stellen abbricht, weder die Periode 0 noch die Periode 9 hat, da jeder Bruch mit der Periode 0 oder 9 offenbar ein endlicher Dezimalbruch ist.

Von nun an kann sich der Leser die reellen Zahlen als unendliche Dezimalbrüche vorstellen.

---

[1]) Der Querstrich über der 4 weist auf das Minuszeichen hin.

Aus der Schule ist bekannt, daß jeder periodische unendliche Dezimalbruch eine rationale Zahl darstellt und umgekehrt jede rationale Zahl in einen periodischen Dezimalbruch entwickelbar ist. *Somit können die nichtperiodischen unendlichen Dezimalbrüche zur Darstellung der neu eingeführten irrationalen Zahlen dienen.* Übrigens kann auch diese Darstellung zum Ausgangspunkt für den Aufbau der Theorie der irrationalen Zahlen genommen werden.

Bemerkung. Im folgenden werden wir die Approximation einer reellen Zahl α durch rationale Zahlen a, a',

$$a < \alpha < a',$$

deren Differenz beliebig klein ist, vielfach benutzen. Für rationales α ist die Existenz solcher Zahlen a und a' trivial, für irrationales α können für a und a' beispielsweise die Dezimalbrüche C_n und C_n' mit hinreichend großem n benutzt werden.

10. Die Stetigkeit des Bereichs der reellen Zahlen. Wir wenden uns nun der Untersuchung einer äußerst wichtigen Eigenschaft des Bereichs aller reellen Zahlen zu, durch die er sich grundlegend vom Bereich der rationalen Zahlen unterscheidet. Bei der Betrachtung der Schnitte im Bereich der rationalen Zahlen stellten wir fest, daß sich unter Umständen für Schnitte in diesem Bereich keine Randzahl angeben ließ, von der man sagen konnte, daß sie den Schnitt erzeuge. Gerade diese Unvollständigkeit des Bereichs der rationalen Zahlen — das Auftreten von Lücken — veranlaßte uns zur Einführung neuer Zahlen, eben der irrationalen Zahlen.

Wir wollen jetzt Schnitte im Bereich aller reellen Zahlen betrachten. Unter einem solchen Schnitt verstehen wir die Zerlegung dieses Bereichs in zwei nicht leere Mengen A, A', für welche folgendes gilt:

1. *Jede reelle Zahl liegt in einer und nur einer*[1]) *der Mengen A, A'.*
2. *Jede Zahl α der Menge A ist kleiner als jede Zahl α' der Menge A'.*

Es erhebt sich nun die Frage, ob sich für einen solchen Schnitt $A \mid A'$ (unter den reellen Zahlen) immer eine Randzahl angeben läßt, die diesen Schnitt erzeugt, oder ob Lücken in diesem Bereich vorhanden sind (die als Grundlage für die Einführung weiterer neuer Zahlen dienen könnten). Wie sich herausstellt, gibt es keine solchen Lücken.

Dedekindscher Hauptsatz. *Für jeden Schnitt $A \mid A'$ im Bereich der reellen Zahlen existiert eine reelle Zahl β, die diesen Schnitt erzeugt. Diese Zahl β ist entweder die größte in der Unterklasse A oder die kleinste in der Oberklasse A'.*

Man bezeichnet diese Eigenschaft des Bereichs der reeellen Zahlen als *Vollständigkeit* oder *Stetigkeit* (oder auch *Abgeschlossenheit*).

Beweis. Wir bezeichnen mit A die Menge aller rationalen Zahlen, die zu A gehören, und mit A' die Menge aller rationalen Zahlen, die zu A' gehören. Man überzeugt sich leicht davon, daß die Mengen A und A' einen Schnitt im Bereich der rationalen Zahlen erzeugen.

Dieser Schnitt $A \mid A'$ definiert eine reelle Zahl β. Diese muß in einer der Klassen A, A' liegen. Nehmen wir beispielsweise an, β liege in der Unterklasse A, so beweisen wir, daß dann der Fall 1 eintritt, nämlich β wirklich die größte Zahl in der Klasse A ist. In der Tat, wenn dies nicht so wäre, dann ließe sich eine andere Zahl α_0 dieser Klasse finden, die größer als β wäre. Nach Lemma 1 können wir eine rationale Zahl r

[1]) Vgl. die Fußnote 1 auf S. 21.

zwischen α_0 und β finden,

$$\alpha_0 > r > \beta.$$

Auch r gehört zur Klasse A, liegt also in A. Somit ergibt sich der Widerspruch, daß die rationale Zahl r, die in der Unterklasse des Schnittes liegt, der die Zahl β definiert, größer als diese Zahl ist, und damit ist unsere Behauptung bewiesen. Eine analoge Überlegung zeigt, daß der Fall 2 eintritt, wenn β in die Oberklasse A' fällt.

Bemerkung. Das gleichzeitige Auftreten einer größten Zahl in der Klasse A und einer kleinsten in der Klasse A' ist nicht möglich. Dies leitet man ebenso her wie für einen Schnitt in der Menge der rationalen Zahlen (mit Hilfe von Lemma 1).

11. Grenzen von Zahlenmengen. Wir benutzen den Dedekindschen Hauptsatz, um bestimmte Begriffe einzuführen, die in der modernen Analysis eine grundlegende Rolle spielen. (Sie werden auch bei der Betrachtung der Rechenoperationen mit reellen Zahlen notwendig sein.)

Wir stellen uns eine beliebige *unendliche Menge* reeller Zahlen vor, die auf beliebige Weise vorgegeben sein kann, etwa die Menge der natürlichen Zahlen, die Menge aller echten Brüche, die Menge aller reellen Zahlen zwischen 0 und 1, die Menge der Lösungen der Gleichung $\sin x = \frac{1}{2}$ usw. Eine beliebige Zahl der Menge bezeichnen wir mit x, so daß x die typische Bezeichnung für die Zahlen der Menge ist. Die Menge der Zahlen x selbst werden wir durch das Symbol $\mathcal{X} = \{x\}$ charakterisieren.

Existiert für eine Menge $\{x\}$ eine Zahl M mit der Eigenschaft $x \leqq M$ für alle x, so wollen wir sagen, die Menge sei (durch die Zahl M) *nach oben beschränkt*. Die Zahl M ist in diesem Fall eine *obere Schranke* der Menge $\{x\}$. Zum Beispiel ist die Menge der echten Brüche nach oben durch die Zahl 1 oder eine beliebige Zahl > 1 beschränkt, die Folge der natürlichen Zahlen ist nach oben nicht beschränkt.

Analog dazu ergibt sich folgendes: Wenn sich eine Zahl m finden läßt derart, daß $x \geqq m$ gilt für alle x, so sagt man, die Menge $\{x\}$ sei (durch die Zahl m) *nach unten beschränkt*, und die Zahl m selbst wird eine *untere Schranke* der Menge $\{x\}$ genannt.

Die natürlichen Zahlen sind beispielsweise nach unten durch die Zahl 1 oder eine beliebige Zahl < 1 beschränkt; die Menge der positiven echten Brüche ist nach unten durch die Zahl 0 oder eine beliebige Zahl < 0 beschränkt.

Eine nach oben (unten) beschränkte Menge kann dabei nach unten (oben) beschränkt oder auch nicht beschränkt sein. So ist die Menge der echten Brüche sowohl nach oben als auch nach unten beschränkt; dagegen ist die Folge der natürlichen Zahlen nach unten beschränkt, nach oben aber nicht beschränkt.

Ist eine Menge nach oben (unten) nicht beschränkt, dann versteht man unter ihrer oberen (unteren) Schranke die „uneigentliche Zahl" $+\infty$ $(-\infty)$. Bezüglich dieser uneigentlichen oder unendlichen Zahlen setzen wir folgendes fest:

$$-\infty < +\infty \quad \text{und} \quad -\infty < \alpha < +\infty$$

für jede reelle („endliche") Zahl α. Die Zeichen $+\infty$ und $-\infty$ liest man als „plus Unendlich" und „minus Unendlich".

Ist eine Menge nach oben beschränkt, d. h., hat sie eine endliche obere Schranke M, dann besitzt sie gleichzeitig eine unendliche Menge oberer Schranken (da ja jede Zahl $N > M$ offenbar ebenfalls eine obere Schranke ist). Von allen oberen Schranken interessiert besonders die kleinste, welche wir die *obere Grenze* nennen wollen. Analog dazu werden wir, wenn eine Menge nach unten beschränkt ist, die größte aller unteren Schranken die *untere Grenze* nennen. So sind z. B. für die Menge aller echten Brüche die Zahlen 0 und 1 die untere bzw. obere Grenze.

Es erhebt sich nun die Frage, ob für nach oben (unten) beschränkte Mengen immer eine obere (untere) Grenze existiert. Da es in diesem Fall eine unendliche Menge von oberen Schranken gibt und innerhalb einer unendlichen Menge von Zahlen keine größte oder kleinste Zahl zu existieren braucht[1]), muß tatsächlich die Existenz einer solchen kleinsten (größten) Zahl unter allen oberen (unteren) Schranken der zu betrachtenden Mengen bewiesen werden.

Satz. *Ist eine Menge $\mathcal{X} = \{x\}$ nach oben (unten) beschränkt, so hat sie auch eine obere (untere) Grenze.*

Beweis. Wir führen die Überlegung in bezug auf die obere Grenze durch und betrachten zwei Fälle:

1. Unter den Zahlen x der Menge $\mathcal{X}$ befindet sich eine größte $\bar{x}$. Dann erfüllen alle Zahlen der Menge die Ungleichung $x \leqq \bar{x}$, d. h., $\bar{x}$ ist eine obere Schranke für $\mathcal{X}$. Andererseits ist $\bar{x}$ in $\mathcal{X}$ enthalten. Folglich ist für jede obere Schranke M die Ungleichung $\bar{x} \leqq M$ erfüllt. Daher können wir schließen, daß $\bar{x}$ die obere Grenze der Menge $\mathcal{X}$ ist.

2. Unter den Zahlen x der Menge $\mathcal{X}$ gibt es keine größte. Wir führen dann auf folgende Weise einen Schnitt im Bereich aller reellen Zahlen. Zur Oberklasse A' nehmen wir alle oberen Schranken α' der Menge $\mathcal{X}$ und zur Unterklasse A alle übrigen reellen Zahlen α. Bei dieser Zerlegung fallen alle Zahlen x der Menge $\mathcal{X}$ in die Klasse A, da keine von ihnen (entsprechend der Voraussetzung) die größte ist. Somit sind beide Klassen A, A' nicht leer. Diese Zerlegung ist tatsächlich ein Schnitt, weil alle reellen Zahlen in Klassen verteilt sind und jede Zahl der Klasse A' größer ist als die Zahl der Klasse A. Nach dem Dedekindschen Hauptsatz (vgl. Nr. 10) muß eine reelle Zahl β existieren, die diesen Schnitt erzeugt. Alle Zahlen x überschreiten (als zur Klasse A gehörend) diese „Randzahl" β nicht, d. h., β ist obere Schranke für x, folglich liegt β selbst in der Klasse A' und ist dort die kleinste Zahl. Somit ist β als kleinste aller oberen Schranken gerade die gesuchte obere Grenze der Menge $\mathcal{X} = \{x\}$.

Auf die gleiche Weise wird auch die zweite Hälfte des Satzes (bezüglich der Existenz einer unteren Grenze) bewiesen.

Ist M^* die *obere Grenze* der Zahlenmenge $\mathcal{X} = \{x\}$, dann gilt *für alle* x

$$x \leqq M^*.$$

Wir wählen jetzt eine beliebige Zahl α, die kleiner ist als M^*. Da M^* die kleinste der oberen Schranken ist, kann α sicher keine obere Schranke der Menge $\mathcal{X}$ sein, d. h., es läßt sich eine Zahl x' aus $\mathcal{X}$ angeben derart, daß

$$x' > \alpha$$

ist. Durch diese beiden Ungleichungen ist die obere Grenze einer Menge $\mathcal{X}$ vollständig charakterisiert.

Analog dazu wird die *untere Grenze* m^* einer Menge $\mathcal{X}$ dadurch charakterisiert, daß *für alle* x aus $\mathcal{X}$

$$x \geqq m^*$$

gilt und sich zu jeder beliebig gewählten Zahl $\beta > m^*$ eine Zahl x'' aus $\mathcal{X}$ angeben

[1]) Wie beispielsweise in der Menge aller echten Brüche.

läßt derart, daß

$$x'' < \beta$$

gilt.

Zur Bezeichnung der oberen Grenze M^* bzw. der unteren Grenze m^* einer Zahlenmenge $\mathcal{X}$ benutzt man die Symbole

$$M^* = \sup \mathcal{X} = \sup \{x\}, \quad m^* = \inf \mathcal{X} = \inf \{x\}$$

(lateinisch: supremum = das Obere, infimum = das Untere).

Wir wollen noch eine einfache Schlußweise erwähnen, die wir im folgenden oft benutzen werden.

Genügen alle Zahlen x einer Menge der Ungleichung $x \leqq M$, so gilt auch $\sup \{x\} \leqq M$.

Die Zahl M ist nämlich eine der oberen Schranken der Menge; daher kann die kleinste aller oberen Schranken sie nicht übertreffen.

Analog *folgt aus der Ungleichung $x \geqq m$ auch* $\inf \{x\} \geqq m$.

Schließlich wollen wir noch folgendes vereinbaren. Besitzt die Menge $X = \{x\}$ keine obere Schranke, so sagen wir, ihre obere Grenze sei $+\infty$, also $\sup \{x\} = +\infty$. Wenn die Menge $X = \{x\}$ keine untere Schranke besitzt, so sagt man, ihre untere Grenze sei $-\infty$, also $\inf \{x\} = -\infty$.

§ 3. Die Rechenoperationen mit reellen Zahlen

12. Definition der Summe reeller Zahlen. Wir wollen jetzt die Rechenoperationen mit reellen Zahlen definieren. Griechische Buchstaben α, β, γ bezeichnen im folgenden reelle Zahlen, und zwar sowohl rationale als auch irrationale.

Es seien zwei reelle Zahlen α und β vorgegeben. Wir betrachten rationale Zahlen a, a' und b, b', die den folgenden Ungleichungen genügen:

$$a < \alpha < a' \quad \text{und} \quad b < \beta < b'. \tag{1}$$

Unter der *Summe* $\alpha + \beta$ der Zahlen α und β verstehen wir eine solche reelle Zahl γ, die zwischen allen Summen der Form $a + b$ einerseits und allen Summen der Form $a' + b'$ andererseits liegt:

$$a + b < \gamma < a' + b'. \tag{2}$$

Wir vergewissern uns zunächst, daß eine solche Zahl γ für jedes Paar reeller Zahlen α und β existiert, und betrachten dazu die Menge aller möglichen Summen $a + b$. Diese Menge ist nach oben beschränkt, beispielsweise durch eine beliebige Summe der Form $a' + b'$. Wir setzen nun (Nr. 11)

$$\gamma = \sup \{a + b\}.$$

Dann gilt $a + b \leqq \gamma$ und gleichzeitig $\gamma \leqq a' + b'$.

Da es bei jeder Wahl von rationalen Zahlen a, b, a', b', die der Beziehung (1) genügen, immer möglich ist, die Zahlen a und b zu vergrößern und die Zahlen a', b' zu verkleinern, ohne diese Bedingungen zu verletzen, kann in den eben erhaltenen Ungleichungen, in denen noch das Gleichheitszeichen vorkommt, in Wirklichkeit Gleichheit gar nicht eintreten. Also erfüllt γ die Summendefinition.

Es entsteht jedoch die Frage, ob die Summe $\gamma = \alpha + \beta$ eindeutig durch die Ungleichungen (2) definiert ist. Um uns von dieser Eindeutigkeit zu überzeugen, wählen wir entsprechend der Bemerkung in Nr. 9 die rationalen Zahlen a, a', b, b' so, daß

$$a' - a < e \quad \text{und} \quad b' - b < e$$

gilt, wobei e eine beliebig kleine rationale positive Zahl ist. Daraus folgt

$$(a' + b') - (a + b) = (a' - a) + (b' - b) < 2e,$$

d. h., diese Differenz kann beliebig klein gemacht werden.[1]) Dann existiert jedoch nach Lemma 2 nur eine einzige Zahl, die zwischen den Summen $a + b$ und $a' + b'$ liegt.

Schließlich wollen wir noch darauf hinweisen, daß die gewöhnliche Summe $\gamma = \alpha + \beta$ zweier rationaler Zahlen α und β offenbar der Ungleichung (2) genügt. Somit widerspricht die obige allgemeine Definition der Summe zweier reeller Zahlen nicht der alten Definition der Summe zweier rationaler Zahlen.

13. Eigenschaften der Addition. Man vergewissert sich leicht, daß auch für die reellen Zahlen folgende Eigenschaften erhalten bleiben:

2a) $\alpha + \beta = \beta + \alpha$;

2b) $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$;

2c) $\alpha + 0 = \alpha$.

Wir beweisen beispielsweise die letzte.

Sind die rationalen Zahlen a, a', b, b' so beschaffen, daß

$$a < \alpha < a', \quad b < 0 < b'$$

ist, so ist offenbar

$$a + b < a < \alpha < a' < a' + b'.$$

Somit ist α eine reelle Zahl, die zwischen Zahlen der Form $a + b$ und $a' + b'$ liegt, zwischen denen aber auch definitionsgemäß die Summe $\alpha + 0$ eingeschlossen ist. Es kann jedoch nur eine einzige solche Zahl geben, daher muß $\alpha + 0 = \alpha$ gelten, was zu beweisen war.

Wir wenden uns jetzt der Eigenschaft 2d) zu und beweisen, daß *zu jeder reellen Zahl α eine* (zu ihr entgegengesetzte) *Zahl $-\alpha$ existiert, welche die Relation $\alpha + (-\alpha) = 0$ erfüllt.* Hierbei genügt es, wenn wir uns auf den Fall beschränken, daß α eine irrationale Zahl ist. Unter der Voraussetzung, daß α durch den Schnitt $A \mid A'$ definiert wird, bestimmen wir die Zahl $-\alpha$ auf folgende Weise. Zur Unterklasse $\bar{A}$ der Zahl $-\alpha$ nehmen wir alle rationalen Zahlen $-a'$, wobei a' eine beliebige Zahl der Klasse A' ist, und zur Oberklasse $\bar{A}'$ dieser Zahl wählen wir die Zahlen $-a$, wobei a eine beliebige Zahl der Klasse A ist. Es ist leicht zu sehen, daß diese Zerlegung ein Schnitt ist und tatsächlich eine reelle (im gegebenen Fall eine irrationale) Zahl definiert. Diese Zahl bezeichnen wir mit $-\alpha$.

Wir zeigen jetzt, daß $-\alpha$ der Bedingung $\alpha + (-\alpha) = 0$ genügt. Unter Benutzung der Definition der Zahl $-\alpha$ selbst sehen wir, daß die Summe $\alpha + (-\alpha)$ diejenige eindeutig bestimmte reelle Zahl ist, die zwischen Zahlen der Form $a - a'$ und $a' - a$ eingeschlossen ist, wobei a und a' rational sind und $a < \alpha < a'$ gilt.

[1]) Die Zahl $2e$ wird kleiner als eine beliebige Zahl $e' > 0$, wenn man $e < \frac{e'}{2}$ wählt.

Nun ist offenbar

$$a - a' < 0 < a' - a,$$

so daß auch die Zahl 0 zwischen den eben erwähnten Zahlen eingeschlossen ist. Da aber eine solche Zahl eindeutig bestimmt ist, erhalten wir

$$\alpha + (-\alpha) = 0,$$

was zu beweisen war.

Schließlich leiten wir noch folgende Eigenschaft her:

2e) *Aus* $\alpha > \beta$ *folgt* $\alpha + \gamma > \beta + \gamma$.

Ist $\alpha > \beta$, so lassen sich zwischen diese beiden Zahlen zwei rationale Zahlen r_1 und r_2 einfügen: $\alpha > r_1 > r_2 > \beta$. Auf Grund der Bemerkung in Nr. 9 existieren auch zwei rationale Zahlen c und c' derart, daß

$$c < \gamma < c' \quad \text{und} \quad c' - c < r_1 - r_2$$

ist. Daraus folgt $r_1 + c > r_2 + c'$ und nach Definition der Summe

$$\alpha + \gamma > r_1 + c, \quad r_2 + c' > \beta + \gamma.$$

Diese Ungleichungen zusammen liefern die Behauptung.

Somit gelten im Bereich der reellen Zahlen bezüglich der Addition alle grundlegenden Eigenschaften 2a) bis 2e), welche in Nr. 3 ursprünglich für die rationalen Zahlen formuliert worden waren. Folglich können auch alle formal logischen Folgerungen aus diesen Eigenschaften automatisch auf die reellen Zahlen übertragen werden. Insbesondere kann für die reellen Zahlen alles wörtlich wiederholt werden, was in Nr. 3 unmittelbar nach der Darlegung der zweiten Gruppe von Eigenschaften gesagt wurde, d. h., man kann die Existenz und die Eindeutigkeit der Differenz $\alpha - \beta$ der Zahlen α und β beweisen, man kann den Begriff des Absolutbetrags einer Zahl α einführen (für den wir die Bezeichnung $|\alpha|$ beibehalten), usw.

14. Definition des Produktes reeller Zahlen. Wir gehen nun zur *Multiplikation* reeller Zahlen über, wobei wir uns zunächst auf *positive* Zahlen beschränken wollen. Es seien zwei solcher Zahlen α und β gegeben.

Wir werden auch hier alle möglichen rationalen Zahlen betrachten, die der Ungleichung (1) genügen, setzen sie aber jetzt sämtlich als positiv voraus.

Unter dem *Produkt* $\alpha\beta$ zweier positiver reeller Zahlen α und β verstehen wir eine reelle Zahl γ, welche zwischen allen Produkten der Form ab einerseits und allen Produkten der Form $a'b'$ andererseits liegt:

$$ab < \gamma < a'b'. \tag{3}$$

Zum Beweis der Existenz einer solchen Zahl γ betrachten wir die Menge aller möglichen Produkte der Form ab; sie ist nach oben durch jedes Produkt der Form $a'b'$ beschränkt. Wenn wir

$$\gamma = \sup \{ab\}$$

setzen, so ist natürlich $ab \leqq \gamma$, gleichzeitig auch $\gamma \leqq a'b'$.

Die Möglichkeit, die Zahlen a, b zu vergrößern und die Zahlen a', b' zu verkleinern (wie bei dem entsprechenden Beweis für die Summe), gestattet hier, das Gleichheitszeichen auszuschließen, so daß die Zahl γ tatsächlich die Bedingungen der Produktdefinition erfüllt.

Die Eindeutigkeit des Produktes ergibt sich aus folgender Überlegung. Wir wählen, entsprechend der Bemerkung in Nr. 9, die rationalen Zahlen a, a' und b, b' so, daß

$$a' - a < e \quad \text{und} \quad b' - b < e$$

gilt, wobei e eine beliebig kleine rationale positive Zahl ist. Dabei können wir annehmen, daß die Zahlen a und b positiv sind und die Zahlen a' und b' irgendwelche vorher fixierten Zahlen a_0' bzw. b_0' nicht übertreffen. Dann ergibt sich für die Differenz

$$a'b' - ab = a'(b' - b) + b(a' - a) < (a_0' + b_0')\, e,$$

d. h., sie kann beliebig klein gemacht werden[1]); dies genügt nach Lemma 2 zum Beweis der Behauptung, daß die Ungleichung (3) nur von einer einzigen Zahl γ erfüllt werden kann. Sind die positiven Zahlen α und β beide rational, so genügt ihr gewöhnliches Produkt $\gamma = \alpha\beta$ offenbar der Ungleichung (3), d. h., die allgemeine Definition des Produktes reeller Zahlen steht im Einklang mit der speziellen Definition des Produktes rationaler Zahlen.

Um schließlich das Produkt beliebiger Paare (nicht notwendig positiver) reeller Zahlen zu definieren, vereinbaren wir folgendes:

Zunächst gelte

$$\alpha \cdot 0 = 0 \cdot \alpha = 0$$

für jedes α. Wenn jedoch beide Faktoren von 0 verschieden sind, so legen wir die gewöhnliche „Vorzeichenregel" zugrunde: $\alpha\beta = |\alpha| \cdot |\beta|$, wenn α und β gleiche Vorzeichen haben $\alpha\beta = -(|\alpha| \cdot |\beta|)$, wenn α und β verschiedene Vorzeichen haben (was das Produkt der positiven Zahlen $|\alpha|$ und $|\beta|$ bedeutet, wissen wir bereits).

Diese Übereinkunft ist, wie wir in Nr. 4 gesehen haben, notwendig, wenn die Grundrechenarten mit reellen Zahlen alle grundlegenden Eigenschaften der Grundrechenarten mit rationalen Zahlen besitzen sollen.

15. Eigenschaften der Multiplikation. Wie auch im Fall der rationalen Zahlen bleiben für beliebige reelle Zahlen folgende Eigenschaften erhalten:

3a) $\alpha\beta = \beta\alpha$;

3b) $(\alpha\beta)\,\gamma = \alpha(\beta\gamma)$;

3c) $\alpha \cdot 1 = \alpha$.

Als Beispiel beweisen wir die zweite Eigenschaft, zunächst für den Fall, daß alle drei Zahlen α, β, γ positiv sind.

Es seien a, a', b, b', c, c' beliebige rationale Zahlen, die den Ungleichungen

$$0 < a < \alpha < a', \quad 0 < b < \beta < b', \quad 0 < c < \gamma < c'$$

genügen. Dann erhalten wir nach unserer Definition des Produktes zweier reeller Zahlen

$$ab < \alpha\beta < a'b' \quad \text{und} \quad bc < \beta\gamma < b'c'.$$

Ferner folgt

$$(ab)\,c < (\alpha\beta)\,\gamma < (a'b')\,c' \quad \text{und} \quad a(bc) < \alpha(\beta\gamma) < a'(b'c').$$

Da für rationale Zahlen die zu beweisende Eigenschaft bereits bekannt, also

$$(ab)\,c = a(bc) \quad \text{und} \quad (a'b')\,c' = a'(b'c')$$

---

[1]) Wir bemerken, daß $(a_0' + b_0')\,e$ kleiner als eine beliebige Zahl $e' > 0$ gemacht werden kann, wenn $e < \dfrac{e'}{a_0' + b_0'}$ gewählt wird.

ist, erweisen sich die reellen Zahlen $(\alpha\beta)\,\gamma$ und $\alpha(\beta\gamma)$ als zwischen ein und denselben Grenzen eingeschlossen.

Nun ist jedoch leicht zu zeigen, daß für Faktoren a und a', b und b', c und c', die sich wenig voneinander unterscheiden, auch die Differenz der Produkte $a'b'c' - abc$ beliebig klein gemacht werden kann (man schließt ähnlich wie in Nr. 14 beim Produkt zweier Faktoren). Hieraus ergibt sich nach Lemma 2, daß die Zahlen $(\alpha\beta)\,\gamma$ und $\alpha(\beta\gamma)$ einander gleich sind.

Der Fall, daß die Zahlen beliebiges Vorzeichen besitzen, läßt sich unter Berücksichtigung der gewöhnlichen Vorzeichenregel unmittelbar erledigen. Ist aber eine der Zahlen α, β, γ gleich 0, so sind auch beide Produkte 0.

Wir kommen jetzt zur Eigenschaft

3d) *Zu jeder von 0 verschiedenen reellen Zahl α existiert eine (zu ihr reziproke) Zahl $\frac{1}{\alpha}$, die der Bedingung*

$$\alpha \cdot \frac{1}{\alpha} = 1$$

genügt.

Wir können uns auf den Fall beschränken, daß α irrational ist. Es sei zunächst $\alpha > 0$. Wird α durch den Schnitt $A \mid A'$ definiert, so konstruieren wir auf folgende Weise einen Schnitt für die Zahl $\frac{1}{\alpha}$. In ihre Unterklasse $\tilde{A}$ nehmen wir alle negativen rationalen Zahlen und 0 sowie alle Zahlen der Form $\frac{1}{a'}$, wobei a' eine beliebige Zahl der Klasse A' ist; in die Oberklasse $\tilde{A}'$ geben wir alle Zahlen $\frac{1}{a}$, wobei a eine beliebige positive Zahl der Klasse A ist. Man überzeugt sich leicht davon, daß man auf diese Weise wirklich einen Schnitt erhält, der eine positive reelle (in diesem Fall irrationale) Zahl definiert. Diese Zahl bezeichnen wir mit $\frac{1}{\alpha}$. Wir zeigen nun, daß sie das Gewünschte leistet.

Berücksichtigt man die Art der Konstruktion der reziproken Zahl, so folgt (nach der Definition des Produktes), daß die Zahl $\alpha \cdot \frac{1}{\alpha}$ die eindeutig bestimmte reelle Zahl ist, die zwischen den Zahlen der Form $\frac{a}{a'}$ und $\frac{a'}{a}$ liegt, wobei a und a' positive rationale Zahlen sind, die der Ungleichung $a < \alpha < a'$ genügen. Nun liegt aber auch die Zahl 1 zwischen den erwähnten Zahlen:

$$\frac{a}{a'} < 1 < \frac{a'}{a};$$

folglich ist sie das gesuchte Produkt.

Ist $\alpha < 0$, so setzen wir

$$\frac{1}{\alpha} = -\frac{1}{|\alpha|};$$

dann ist nach der Vorzeichenregel

$$\alpha \cdot \frac{1}{\alpha} = |\alpha| \cdot \frac{1}{|\alpha|} = 1.$$

Nachdem wir uns davon überzeugt haben, daß in bezug auf die Multiplikation der Bereich der reellen Zahlen alle grundlegenden Eigenschaften 3a) bis 3d) besitzt, ist klar, daß für diesen Bereich alles in Nr. 4 über Existenz und Eindeutigkeit des Quotienten $\frac{\alpha}{\beta}$ der Zahlen α und β (unter der Voraussetzung $\beta \neq 0$) Gesagte gültig bleibt.

Das Distributivgesetz

3e) $(\alpha + \beta)\gamma = \alpha\gamma + \beta\gamma$

gilt ebenfalls für beliebige reelle Zahlen, was sich leicht für den Fall positiver Zahlen zeigen läßt (ebenso wie Eigenschaft 3b). Auf diesen Fall lassen sich alle übrigen zurückführen, entweder durch Änderung der Vorzeichen beider Seiten der Gleichung oder durch Herüber- bzw. Hinüberbringen der Glieder von einer Seite auf die andere. Eine Ausnahme bildet übrigens der Fall, daß eine der Zahlen $\alpha, \beta, \gamma, \alpha + \beta$ gleich 0 ist; jedoch ist dann die Gleichheit der beiden Seiten unmittelbar einzusehen.

Schließlich gilt noch:

3f) *Aus* $\alpha > \beta$ *und* $\gamma > 0$ *folgt* $\alpha\gamma > \beta\gamma$, wie sich leicht verifizieren läßt. Die Ungleichung $\alpha > \beta$ ist äquivalent mit $\alpha - \beta > 0$; dann gilt nach der Vorzeichenregel $(\alpha - \beta)\gamma > 0$. Da die Multiplikation auch bezüglich der Differenz distributiv ist, folgt

$$\alpha\gamma - \beta\gamma > 0, \quad \text{also} \quad \alpha\gamma > \beta\gamma.$$

16. Schlußbemerkungen. Es muß noch etwas zum Archimedischen Axiom gesagt werden.

4a) *Zu jeder reellen Zahl γ existiert eine natürliche Zahl n, die größer ist als γ.*

Dies läßt sich leicht nachweisen; denn in der Oberklasse des Schnittes $C \mid C'$, der die Zahl γ erzeugt, befindet sich eine rationale Zahl $c' > \gamma$, und für rationale Zahlen ist dieses Prinzip bereits als gültig angenommen.

Jetzt können wir endlich als erwiesen ansehen, daß *im Bereich der reellen Zahlen die* (zunächst für rationale Zahlen erarbeiteten) *Gesetze der elementaren Algebra, die sich auf die vier Grundrechenarten und auf die Verknüpfung von Gleichungen und Ungleichungen beziehen, sämtlich und in voller Allgemeinheit gültig bleiben.*

17. Absolute Beträge. Zum Verständnis des Folgenden wollen wir noch einige Bemerkungen über absolute Beträge machen. Zunächst stellen wir fest, daß die Ungleichung $|\alpha| < \beta$ (wobei natürlich $\beta > 0$ ist) mit der doppelten Ungleichung $-\beta < \alpha < \beta$ gleichwertig ist. Aus $|\alpha| < \beta$ folgt nämlich, daß gleichzeitig $\alpha < \beta$ und $-\alpha < \beta$, also $\alpha > -\beta$ ist. Ist umgekehrt bekannt, daß $\alpha < \beta$ und $\alpha > -\beta$ ist, so gilt gleichzeitig $\alpha < \beta$ und $-\alpha < \beta$; eine der Zahlen α, $-\alpha$ ist aber $|\alpha|$, so daß sicher $|\alpha| < \beta$ gilt. Analog ergibt sich, daß auch die Ungleichungen $|\alpha| \leqq \beta$ und $-\beta \leqq \alpha \leqq \beta$ äquivalent sind.

Ferner wollen wir die sogenannte *Dreiecksungleichung*

$$|\alpha + \beta| \leqq |\alpha| + |\beta|.$$

beweisen, die uns oft von Nutzen sein wird.

Addiert man nämlich die (trivialen) Ungleichungen $-|\alpha| \leqq \alpha \leqq |\alpha|$ und $-|\beta| \leqq \beta \leqq |\beta|$, so erhält man

$$-(|\alpha| + |\beta|) \leqq \alpha + \beta \leqq |\alpha| + |\beta|,$$

woraus sich auf Grund der obigen Bemerkung die gesuchte Ungleichung ergibt.

Durch Induktion läßt sie sich sukzessive auf den Fall beliebig vieler Summanden ausdehnen,

$$|\alpha + \beta + \cdots + \gamma| \leqq |\alpha| + |\beta| + \cdots + |\gamma|.$$

Ersetzt man in der bewiesenen Ungleichung β durch $-\beta$, so erhält man

$$|\alpha - \beta| \leqq |\alpha| + |\beta|.$$

Wegen $\alpha = (\alpha + \beta) - \beta$ ist $|\alpha| \leqq |\alpha + \beta| + |\beta|$ oder

$$|\alpha + \beta| \geqq |\alpha| - |\beta|.$$

Analog ist $|\alpha| - |\beta| \leqq |\alpha - \beta|$. Da gleichzeitig auch $|\beta| - |\alpha| \leqq |\alpha - \beta|$ gilt, ist offenbar

$$||\alpha| - |\beta|| \leqq |\alpha - \beta|.$$

Alle diese Ungleichungen werden uns in der Theorie der Grenzwerte von Nutzen sein

§ 4. Weitere Eigenschaften und Anwendungen der reellen Zahlen

18. Existenz der Wurzel. Potenz mit rationalem Exponenten. Die Definition der Multiplikation (und Division) reeller Zahlen führt unmittelbar, wie üblich, zur Definition der Potenz mit ganzen positiven (und negativen) Exponenten. Wir wollen nun zu Potenzen mit rationalen Exponenten übergehen und befassen uns zunächst mit dem Problem der Existenz der Wurzel.

Wie wir wissen, war gerade die Tatsache, daß im Bereich der rationalen Zahlen schon einfache (Quadrat-) Wurzeln nicht existieren, ein Anlaß zur Erweiterung dieses Bereichs. Wir wollen nun nachprüfen, in welchem Maße die durchgeführte Erweiterung die alten Lücken schließt (und keine neuen erzeugt).

Es sei α eine beliebige reelle Zahl, n eine natürliche Zahl. Bekanntlich versteht man unter der *n-ten Wurzel aus der Zahl α* eine reelle Zahl ξ, für die

$$\xi^n = \alpha$$

gilt. Wir beschränken uns auf den Fall, daß α positiv ist, und wollen ein ebenfalls positives ξ suchen, das dieser Beziehung genügt, d. h. einen sogenannten *arithmetischen Wurzelwert* bestimmen. Wir werden beweisen, daß eine solche Zahl ξ immer existiert und daß sie eindeutig bestimmt ist. Die Eindeutigkeit der Zahl ξ folgt übrigens sofort aus der Tatsache, daß für verschiedene positive Zahlen auch die Potenzen verschieden sind: Wenn $0 < \xi < \xi'$ gilt, so ist $\xi^n < \xi'^n$. Existiert eine rationale Zahl r, deren n-te Potenz gleich α ist, so ist sie genau die gesuchte Zahl ξ. Daher genügt es von nun an, sich auf die Annahme zu beschränken, daß es keine solche rationale Zahl gibt.

Wir konstruieren jetzt auf folgende Weise einen Schnitt $X \mid X'$ im Bereich aller rationalen Zahlen. Zur Klasse X nehmen wir alle negativen rationalen Zahlen und 0 sowie die positiven rationalen Zahlen x, für die $x^n < \alpha$ gilt. Zur Klasse X' nehmen wir alle positiven rationalen Zahlen x', für welche die Beziehung $x'^n > \alpha$ erfüllt ist.

Es ist leicht zu sehen, daß diese Klassen nicht leer sind und daß X auch positive Zahlen enthält: Wenn wir z. B. eine natürliche Zahl $m \geqq 2$ so wählen, daß $\frac{1}{m} < \alpha < m$ ist, so gilt um so mehr $\frac{1}{m^n} < \alpha < m^n$, so daß die Zahl $\frac{1}{m}$ in X liegt, während m zu X' gehört.

Daß auch die anderen Forderungen, die an einen Schnitt gestellt werden, erfüllt sind, läßt sich direkt verifizieren.

Es sei jetzt ξ die durch den Schnitt $X \mid X'$ definierte Zahl. Wir beweisen, daß $\xi^n = \alpha$, also $\xi = \sqrt[n]{\alpha}$ ist. Betrachtet man ξ^n als Produkt von n Faktoren ξ, so kann man auf Grund der Definition des Produktes positiver reeller Zahlen (vgl. Nr. 14) schließen, daß

$$x^n < \xi^n < x'^n$$

ist, wenn x und x' positive rationale Zahlen sind, für die

$$0 < x < \xi < x'$$

gilt. Da x offenbar zur Klasse X gehört und x' zur Klasse X', gilt gemäß der Definition dieser Klassen gleichzeitig auch

$$x^n < \alpha < x'^n.$$

Die Differenz $x' - x$ kann aber kleiner als eine beliebige Zahl $e > 0$ gemacht werden (vgl. die Bemerkung aus Nr. 9), wobei uns nichts daran hindert, x' kleiner anzunehmen als irgendeine vorher festgelegte Zahl x_0'. In diesem Fall gilt für die Differenz

$$x'^n - x^n = (x' - x)\,(x'^{n-1} + x \cdot x'^{n-2} + \cdots + x^{n-1}) < e \cdot n x_0'^{n-1}$$

d. h., sie kann ebenfalls beliebig klein gemacht werden.[1]) Hieraus folgt nach Lemma 2 die Gleichheit der Zahlen ξ^n und α.

Nachdem die Existenz der Wurzel bewiesen ist, kann man auf dem üblichen Wege zu dem Begriff der Potenz mit beliebigem rationalem Exponenten r übergehen und zeigen, daß für solche Potenzen die gewöhnlichen Regeln der elementaren Algebra gültig bleiben, nämlich

$$\alpha^r \cdot \alpha^{r'} = \alpha^{r+r'}, \quad \alpha^r : \alpha^{r'} = \alpha^{r-r'}, \quad (\alpha^r)^{r'} = \alpha^{rr'},$$

$$(\alpha\beta)^r = \alpha^r \beta^r, \quad \left(\frac{\alpha}{\beta}\right)^r = \frac{\alpha^r}{\beta^r} \quad \text{usw.}$$

Wir weisen noch darauf hin, daß die Potenz α^r für $\alpha > 1$ mit wachsendem rationalem Exponenten r wächst.

19. Die Potenz mit beliebigem reellem Exponenten. Wir wenden uns nun der Definition der Potenz einer beliebigen reellen (positiven) Zahl α mit beliebigem reellem Exponenten β zu.

Zu diesem Zweck betrachten wir die Potenzen

$$\alpha^b \quad \text{und} \quad \alpha^{b'}$$

mit rationalen Exponenten b und b', wobei diese der Ungleichung

$$b < \beta < b'$$

genügen.

---

[1]) Wir bemerken, daß die Zahl $e \cdot n x_0'^{n-1}$ kleiner als eine beliebige Zahl $e' > 0$ wird, wenn man $e < \dfrac{e'}{n x_0'^{n-1}}$ wählt.

Unter der *Potenz der Zahl* $\alpha > 1$[1]) *mit dem Exponenten* β versteht man eine reelle Zahl γ, die zwischen den Potenzen α^b und $\alpha^{b'}$ liegt (übrigens wird sich herausstellen, daß sie eindeutig bestimmt ist):

$$\alpha^b < \gamma < \alpha^{b'} \tag{1}$$

(man bezeichnet sie mit α^β).

Man überzeugt sich leicht davon, daß eine solche Zahl immer existiert: Die Menge der Potenzen $\{\alpha^b\}$ ist nach oben beschränkt, z. B. durch eine beliebige Potenz $\alpha^{b'}$. Wir setzen dann (vgl. Nr. 11)

$$\gamma = \sup_{b<\beta} \{\alpha^b\}.$$

Für diese Zahl gilt

$$\alpha^b \leqq \gamma \leqq \alpha^{b'}.$$

In Wirklichkeit ist das Gleichheitszeichen hier überflüssig, da man b vergrößern und b' verkleinern kann, so daß die konstruierte Zahl γ der Bedingung (1) genügt.

Nun kommen wir zum Beweis der Eindeutigkeit der so definierten Zahl γ. Zu diesem Zweck bemerken wir zunächst, daß Lemma 2 aus Nr. 8 auch dann gültig bleibt, wenn wir nicht mehr fordern, daß die Zahlen s, s' und e unbedingt rational sein müssen; der Beweisgedanke bleibt derselbe.

Alsdann wollen wir eine äußerst einfache, aber oft nützliche Ungleichung herleiten, die häufig als *Bernoullische Ungleichung*[2]) bezeichnet wird: *Ist n eine natürliche Zahl > 1 und ist $\gamma > 1$, so gilt*

$$\gamma^n > 1 + n(\gamma - 1). \tag{2}$$

Setzen wir nämlich $\gamma = 1 + \lambda$, wobei $\lambda > 0$ ist, so ist nach dem binomischen Satz

$$(1 + \lambda)^n = 1 + n\lambda + \cdots;$$

da die nicht ausgeschriebenen Glieder positiv sind, gilt

$$(1 + \lambda)^n > 1 + n\lambda,$$

was mit der Ungleichung (2) gleichwertig ist.

Setzen wir hier $\gamma = \alpha^{1/n}$ $(\alpha > 1)$, so erhalten wir die Ungleichung

$$\alpha^{1/n} - 1 < \frac{\alpha - 1}{n}, \tag{3}$$

die wir sofort ebenfalls benutzen werden.

Wir wissen, daß man die Zahlen b und b' so wählen kann, daß die Differenz $b' - b$ kleiner ist als $\frac{1}{n}$ bei beliebigem, vorher gewähltem natürlichem n, so daß nach Ungleichung (3)

$$\alpha^{b'} - \alpha^b = \alpha^b(\alpha^{b'-b} - 1) < \alpha^b(\alpha^{1/n} - 1) < \alpha^b \cdot \frac{\alpha - 1}{n}$$

---

[1]) Auf diesen Fall kann man sich beschränken. Im Fall $0 < \alpha < 1$ setzen wir beispielsweise $\alpha^\beta = \left(\frac{1}{\alpha}\right)^{-\beta}$.

[2]) Nach Jakob Bernoulli, 1654—1705, Schweizer Mathematiker.

gilt. Da b kleiner als ein beliebiges (aber festes) b_0' ist, genügt es,

$$n > \frac{\alpha^{b_0'}(\alpha - 1)}{\varepsilon}$$

zu wählen, wobei ε eine beliebig kleine positive Zahl ist, damit

$$\alpha^{b'} - \alpha^b < \varepsilon$$

gilt. In diesem Fall können nach dem verallgemeinerten Lemma 2 zwischen den Grenzen α^b und $\alpha^{b'}$ nicht zwei verschiedene Zahlen γ liegen.

Ist β rational, dann liefert die obige Definition der allgemeinen Potenz wieder die übliche Bedeutung des Symbols α^β.

Es ist leicht zu verifizieren, daß für Potenzen mit beliebigem reellem Exponenten alle für die gewöhnlichen Potenzen gültigen Gesetze erfüllt sind. Wir beweisen als Beispiel die Regel

$$\alpha^\beta \alpha^\gamma = \alpha^{\beta+\gamma}.$$

Es seien b, b', c, c' beliebige rationale Zahlen, für die

$$b < \beta < b', \quad c < \gamma < c'$$

gilt. Nach Definition der Summe (Nr. 12) ist

$$b + c < \beta + \gamma < b' + c'$$

und nach Definition der Potenz

$$\alpha^b < \alpha^\beta < \alpha^{b'}, \quad \alpha^c < \alpha^\gamma < \alpha^{c'} \quad \text{und} \quad \alpha^{b+c} < \alpha^{\beta+\gamma} < \alpha^{b'+c'}.$$

Multiplizieren wir die ersten beiden doppelten Ungleichungen gliedweise (unter Berücksichtigung der Tatsache, daß für rationale Exponenten die zu beweisende Regel bereits bekannt ist), so erhalten wir

$$\alpha^{b+c} < \alpha^\beta \alpha^\gamma < \alpha^{b'+c'}.$$

Somit liegen die beiden Zahlen $\alpha^{\beta+\gamma}$ und $\alpha^\beta\alpha^\gamma$ zwischen denselben Grenzen α^{b+c}, $\alpha^{b'+c'}$, welche, wie man leicht zeigt, beliebig einander angenähert werden können. Hieraus folgt (nach dem verallgemeinerten Lemma 2) die Gleichheit dieser Zahlen.

Wir weisen nun noch nach, daß für $\alpha > 1$ die Potenz α^β mit wachsendem reellem Exponenten β wächst.

Ist $\beta < \bar{\beta}$, so wähle man eine rationale Zahl r dazwischen: $\beta < r < \bar{\beta}$; nach Definition der Potenz mit reellem Exponenten erhält man $\alpha^\beta < \alpha^r$ und $\alpha^r < \alpha^{\bar{\beta}}$, also $\alpha^\beta < \alpha^{\bar{\beta}}$.

20. Der Logarithmus. Benutzt man die Definition der Potenz mit beliebigen reellen Exponenten, so läßt sich die *Existenz des Logarithmus* für jede reelle positive Zahl γ bei positiver, von 1 verschiedener Basis α leicht herleiten (wir nehmen beispielsweise $\alpha > 1$ an).

Existiert eine *rationale Zahl* r derart, daß

$$\alpha^r = \gamma$$

ist, dann ist sie der gesuchte Logarithmus.

Wir nehmen nun an, es gäbe keine solche rationale Zahl r. Dann läßt sich im Bereich aller rationalen Zahlen ein Schnitt $B \mid B'$ nach folgender Vorschrift durchführen. In die Klasse B nehmen wir die rationalen Zahlen b, für welche $\alpha^b < \gamma$ ist, und in die Klasse B' die rationalen Zahlen b', für welche $\alpha^{b'} > \gamma$ gilt.

Wir zeigen, daß die Klassen B und B' nicht leer sind. Nach Ungleichung (2) gilt

$$\alpha^n > 1 + n(\alpha - 1) > n(\alpha - 1),$$

und es genügt,

$$n > \frac{\gamma}{\alpha - 1}$$

zu wählen, damit $\alpha^n > \gamma$ ist; jede solche natürliche Zahl n gehört zur Klasse B'.

Andererseits ist

$$\alpha^{-n} = \frac{1}{\alpha^n} < \frac{1}{n(\alpha - 1)},$$

und es genügt,

$$n > \frac{1}{\gamma(\alpha - 1)}$$

zu wählen, damit $\alpha^{-n} < \gamma$ ist, und jede solche Zahl $-n$ liegt in der Klasse B. Die übrigen Forderungen, die an einen Schnitt gestellt werden, sind hier ebenfalls erfüllt.

Der Schnitt $B \mid B'$ definiert eine reelle Zahl β, welche die „Randzahl" zwischen den Zahlen beider Klassen bildet. Nach Definition der Potenz ist

$$\alpha^b < \alpha^\beta < \alpha^{b'} \quad (b < \beta < b'),$$

wobei α^β die eindeutig bestimmte Zahl ist, die alle solchen Ungleichungen erfüllt. Für die Zahl γ gilt (nach Konstruktion des Schnittes)

$$\alpha^b < \gamma < \alpha^{b'}.$$

Folglich ist

$$\alpha^\beta = \gamma \quad \text{und} \quad \beta = \log_\alpha \gamma \quad (\text{man schreibt auch } \beta = {}^\alpha\!\log \gamma).$$

Damit ist die Existenz des Logarithmus bewiesen.

21. Das Messen von Strecken. Die Tatsache, daß im Bereich der rationalen Zahlen nicht allen Strecken eine Länge zugeordnet werden konnte, war ebenfalls ein wichtiger Anlaß zur Einführung der irrationalen Zahlen. Wir zeigen jetzt, daß die durchgeführte Erweiterung des Bereichs der rationalen Zahlen (zum Bereich der reellen Zahlen) ausreicht, um das Problem der Streckenmessung zu lösen.

Zunächst formulieren wir die Aufgabe selbst.[1])

Es soll jeder geradlinigen Strecke A eine positive reelle Zahl $l(A)$ zugeordnet werden, welche man „Länge der Strecke A" nennt, und zwar so, daß

a) *irgendeine zuerst ausgewählte Strecke E (die Einheitsstrecke) die Länge 1 hat,* $l(E) = 1$;

[1]) Wir benutzen hier die in der Schulgeometrie übliche Darlegung und verzichten auf eine axiomatische Einführung.

b) *gleiche Strecken ein und dieselbe Länge haben;*

c) *beim Aneinanderfügen von Strecken die Länge der Summe stets gleich der Summe der Längen der addierten Strecken ist:*

$$l(A + B) = l(A) + l(B)$$

(*Additivität*).

Diese Bedingungen ermöglichen eine eindeutige Lösung der Aufgabe.

Aus b) und c) folgt, daß der q-te Teil der Einheitsstrecke die Länge $\frac{1}{q}$ haben muß; wird dieser Teil p-mal als Summand genommen, so erhalten wir nach c) eine Strecke, die die Länge $\frac{p}{q}$ haben muß. Wenn also die Strecken A und E ein gemeinsames Maß haben („kommensurabel" sind) und dieses q-mal in E und p-mal in A enthalten ist, so ist notwendigerweise

$$l(A) = \frac{p}{q}.$$

Man sieht leicht, daß diese Zahl nicht von dem gewählten gemeinsamen Maß abhängig ist und daß Strecken, die mit der Einheitsstrecke kommensurabel sind, nach dieser Vorschrift eine rationale Länge zuzuordnen ist; für solche Strecken ist also die Aufgabe des Messens vollständig gelöst.

Ist eine Strecke A größer als eine Strecke B, so daß also $A = B + C$ ist, wobei C ebenfalls eine bestimmte Strecke bedeutet, so muß auf Grund von c)

$$l(A) = l(B) + l(C)$$

sein, und wegen $l(C) > 0$ ist $l(A) > l(B)$. *Somit müssen ungleiche Strecken ungleiche Längen haben, und zwar die größere Strecke auch die größere Länge.*

Da jede positive rationale Zahl $\frac{p}{q}$ die Länge einer bestimmten Strecke ist, die mit der Einheitsstrecke E kommensurabel ist, ist nach unseren Ausführungen klar, daß eine Strecke, die nicht mit der Einheitsstrecke kommensurabel ist, keine rationale Länge haben kann.

Es sei nun Σ eine solche Strecke, die mit E inkommensurabel ist. Dann lassen sich beliebig viele Strecken S und S' finden, die mit E kommensurabel und kleiner bzw. größer als Σ sind.[1]) Wenn man ihre Längen mit s und s' bezeichnet, also $l(S) = s$, $l(S') = s'$ setzt, so muß die gesuchte Länge $l(\Sigma)$ der Ungleichung

$$s < l(\Sigma) < s'$$

genügen.[2]) Verteilt man alle rationalen Zahlen auf die beiden Klassen S^* und S'^*, indem man zur Unterklasse S^* alle Zahlen s (und außer ihnen alle negativen Zahlen und 0) nimmt und zur Oberklasse S'^* alle Zahlen s', so erhält man einen Schnitt im Bereich der rationalen Zahlen. Da in der Unterklasse offenbar keine größte Zahl und in der Oberklasse keine kleinste existiert, definiert dieser Schnitt eine irrationale Zahl σ, welche die einzige reelle Zahl ist, die der Ungleichung $s < \sigma < s'$ genügt. Diese Zahl müssen wir als die Länge $l(\Sigma)$ ansehen.

---

[1]) Dies läßt sich leicht beweisen, indem man von dem geometrischen Axiom des ARCHIMEDES ausgeht, von dem bereits in Nr. 5 die Rede war.

[2]) Es versteht sich, daß für die Länge einer mit E kommensurablen Strecke Σ diese Ungleichung ebenfalls erfüllt ist.

Wir nehmen nun an, wir hätten im Sinne unserer Vorschrift alle Strecken, sowohl den mit E kommensurablen als auch den mit E inkommensurablen, eine Länge zugeordnet.

Daß die Bedingungen a) und b) erfüllt sind, ist klar. Wir betrachten zwei Strecken P, Σ mit den Längen

$$\varrho = l(P), \quad \sigma = l(\Sigma)$$

und die Summe dieser Strecken $T = P + \Sigma$, deren Länge wir mit $\tau = l(T)$ bezeichnen.

Wir wählen beliebige rationale positive Zahlen r, r', s, s' so, daß

$$r < \varrho < r', \quad s < \sigma < s'$$

ist, und konstruieren Strecken R, R', S, S', welche die entsprechenden Längen r, r', s, s' haben.

Die Strecke $R + S$ (der Länge $r + s$) ist kleiner als T und die Strecke $R' + S'$ (der Länge $r' + s'$) größer als T. Daher gilt

$$r + s < \tau < r' + s'.$$

Nach Nr. 12 ist aber die einzige reelle Zahl, die zwischen den Zahlen der Form $r + s$ und $r' + s'$ liegt[1]), genau die Summe $\varrho + \sigma$. Folglich gilt $\tau = \varrho + \sigma$, was zu beweisen war.

Die „Additivität" läßt sich durch Induktion auf den Fall einer beliebigen endlichen Anzahl von Summanden ausdehnen.

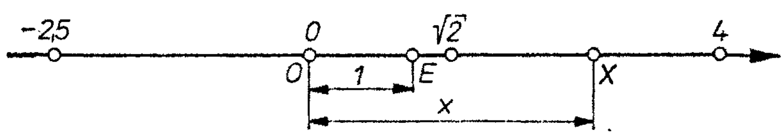

Abb. 1

Wenn man auf einer Achse (einer gerichteten Geraden; Abb. 1) einen Ursprung O und eine Einheitsstrecke der Länge OE abträgt, *so entspricht jedem Punkt X dieser Geraden eine bestimmte reelle Zahl*, seine *Abszisse* x, die gleich der Länge von OX ist, wenn X in positiver Richtung von O liegt, oder gleich der negativen Länge, wenn X in negativer Richtung von O liegt. Naturgemäß erhebt sich die Frage, ob auch das Umgekehrte gilt, d. h., *ob jeder reellen Zahl x hierbei ein bestimmter Punkt der Geraden entspricht*. Dieses Problem wird in der Geometrie in bejahendem Sinne gelöst, und zwar mit Hilfe des Stetigkeitsaxioms, das für die Gerade als Punktmenge eine Eigenschaft zum Ausdruck bringt, die der Stetigkeit des Bereiches der reellen Zahlen (Nr. 10) analog ist. Somit kann man zwischen allen reellen Zahlen und den Punkten einer gerichteten Geraden (einer Achse) eine umkehrbare eindeutige (d. h. eineindeutige) Zuordnung herstellen.

Die reellen Zahlen können auf die Punkte einer Achse abgebildet werden, welche in diesem Zusammenhang *Zahlengerade* genannt wird.

Diese Abbildung werden wir von nun an ständig benutzen.

---

[1]) Die Beschränkung auf *positive* Zahlen r und s ist natürlich nicht von Belang.

I. Theorie der Grenzwerte

§ 1. Folgen und ihre Grenzwerte

22. Veränderliche Größen, Folgen. In der Physik und in anderen Naturwissenschaften hat man es mit einer Vielfalt verschiedener Größen zu tun: Zeit, Länge, Volumen, Gewicht usw. Jede von ihnen nimmt den Umständen entsprechend verschiedene Werte an oder auch nur einen einzigen. Im ersten Fall haben wir es mit einer *veränderlichen* Größe zu tun, im zweiten mit einer *konstanten* Größe.

In der Mathematik sehen wir jedoch von der physikalischen Bedeutung einer solchen Größe ab, da hier nur die Zahl interessiert, durch welche sie ausgedrückt wird. Nur in den Anwendungen wird die physikalische Bedeutung einer Größe wieder wichtig. Demnach ist eine veränderliche Größe (oder kurz: *Veränderliche* oder *Variable*) für uns eine abstrakte oder eine rein numerische Veränderliche. Man bezeichnet sie mit irgendeinem Symbol (z. B. mit dem Buchstaben x), dem man Zahlenwerte zuschreibt.

Eine Veränderliche wird als gegeben angesehen, wenn die Menge $\mathcal{X} = \{x\}$ der Werte bekannt ist, welche sie annehmen kann. Eine konstante Größe (kurz: *Konstante*) läßt sich zweckmäßigerweise als Spezialfall einer Veränderlichen auffassen; und zwar entspricht dieser Fall der Annahme, daß die Menge $\mathcal{X} = \{x\}$ aus einem *einzigen* Element besteht.

Zur Definition des Begriffs *Grenzwert* einer Veränderlichen x reicht es nicht aus, daß man weiß, zu welcher Zahlenmenge $\mathcal{X}$ die von der Veränderlichen angenommenen Werte gehören; man muß außerdem wissen, welche Werte (unter denen auch gewisse Werte mehrfach auftreten können) sie tatsächlich annimmt und in welcher Reihenfolge diese Werte angenommen („durchlaufen") werden.

Wir verschieben die Behandlung des Problems der sogenannten gerichteten Veränderlichen und ihrer Grenzwerte in allgemeinster Darstellung an das Ende des dritten Bandes[1]) (bis der Leser genügend Erfahrung auf diesem Gebiet gesammelt hat) und widmen das vorliegende Kapitel dem Studium eines einzigen, und zwar des einfachsten und überdies wichtigsten Spezialfalls einer solchen veränderlichen Größe.

Wir beginnen mit der Festlegung des Begriffs der *Zahlenfolge*. Wir stellen uns die Folge

$$1, 2, 3, \ldots, n, \ldots, n', \ldots \tag{1}$$

der natürlichen Zahlen vor, in der die Zahlen wachsend angeordnet sind, so daß die größere Zahl n' auf die kleinere Zahl n folgt (oder die kleinere Zahl n der größeren n' vorangeht). Ersetzen wir jetzt in der Folge (1) nach irgendeiner Vorschrift jede natürliche Zahl n durch eine reelle Zahl x_n, so erhalten wir eine *Zahlenfolge*

$$x_1, x_2, x_3, \ldots, x_n, \ldots, x_{n'}, \ldots, \tag{2}$$

---

[1]) Vgl. dort S. 541 den Anhang: Der allgemeine Gesichtspunkt beim Limesbegriff.

deren Glieder oder Elemente x_n mit Hilfe aller natürlichen Zahlen numeriert und nach wachsenden Nummern (Indizes) angeordnet sind. Für $n' > n$ folgt das Glied $x_{n'}$ dem Glied x_n (x_n geht $x_{n'}$ voran), unabhängig davon, ob die Zahl $x_{n'}$ selbst größer, kleiner oder auch gleich der Zahl x_n ist.[1])

Eine Veränderliche x, die eine gewisse Folge (2) von Werten annimmt, nennen wir [nach Ch. Méray[2])] eine *diskrete Veränderliche* (oder *Variante*). Dies ist gerade der Typ von Veränderlichen, auf den wir unsere Betrachtung hier beschränken wollen.

Bereits in der Schulmathematik hat der Leser Veränderliche dieser Art, diskrete Veränderliche, kennengelernt. Er kennt z. B. Folgen (Progressionen) der Form

$$\begin{array}{cccccl} a, & a+d, & a+2d, & \ldots, & a+(n-1)\,d, \ldots & \text{(Glied der Folge)} \\ 1 & 2 & 3 & & n & \text{(Index)} \end{array}$$

(arithmetische Folgen) oder Folgen der Form

$$a, aq, aq^2, \ldots, aq^{n-1}, \ldots$$

(geometrische Folgen); das veränderliche Glied jeder dieser beiden Folgen ist eine diskrete Veränderliche.

Im Zusammenhang mit der Definition der Länge des Kreisumfangs betrachtet man im allgemeinen als Veränderliche den Umfang des dem Kreis einbeschriebenen regelmäßigen Vielecks, das man aus dem Sechseck durch fortlaufende Verdoppelung der Seitenzahl erhält; diese diskrete Veränderliche nimmt die Folge der Werte

$$\begin{array}{ccl} p_6 = 6R, & p_{12} = 12R\sqrt{2-\sqrt{3}}, & \text{(Glied der Folge)} \\ 1 & 2 & \text{(Index)} \end{array}$$

$$\begin{array}{ccl} p_{24} = 24R\sqrt{2+\sqrt{2+\sqrt{3}}}, & p_{48}, \ldots & \text{(Glied der Folge)} \\ 3 & 4 & \text{(Index)} \end{array}$$

an.

Wir erwähnen noch die Näherungsbrüche für $\sqrt{2}$, die diese irrationale Zahl (beispielsweise von unten) mit wachsender Genauigkeit approximieren. Sie bilden die Folge der Werte

$$\begin{array}{cccccl} 1{,}4; & 1{,}41; & 1{,}414; & 1{,}4142; & \ldots & \text{(Glied der Folge)} \\ 1 & 2 & 3 & 4 & & \text{(Index)} \end{array}$$

Wir haben also auch hier eine diskrete Veränderliche vor uns.

Eine Veränderliche x, welche eine Folge (2) durchläuft, wird oft mit x_n bezeichnet, indem man sie mit dem veränderlichen („allgemeinen") Glied dieser Folge identifiziert. Eine Folge, deren allgemeines Glied x_n ist, wird vielfach auch mit $\{x_n\}$ bezeichnet.

Manchmal wird eine diskrete Veränderliche dadurch vorgegeben, daß der Ausdruck für x_n direkt hingeschrieben wird; so haben wir im Fall der arithmetischen bzw. geometrischen Progression $x_n = a + (n-1)\,d$ bzw. $x_n = aq^{n-1}$. Hier läßt sich sofort jeder Wert der diskreten Veränderlichen durch Vorgabe des Index berechnen, ohne daß die vorhergehenden Werte bestimmt werden müssen.

---

[1]) Analog definiert man den Begriff einer *Folge* von Punkten auf einer Geraden oder von irgendwelchen Objekten anderer Art.

[2]) Hugues Charles Robert Méray, 1835—1911, französischer Mathematiker.

Für den Umfang des regelmäßigen einbeschriebenen Vielecks ist ein solcher allgemeiner Ausdruck nur möglich, wenn man die Zahl π (und trigonometrische Funktionen) einführt. Allgemein gilt für den Umfang p_m des regelmäßigen einbeschriebenen m-Ecks die Formel

$$p_m = 2mR \sin\frac{\pi}{m}.$$

In anderen Fällen kann es vorkommen, daß man keinen Ausdruck für das allgemeine Glied x_n einer Folge (2) kennt. *Nichtsdestoweniger sieht man eine Folge* (2) *und damit auch die ihr entsprechende diskrete Veränderliche als gegeben an, wenn man wenigstens die Vorschrift kennt, nach der jeder Wert der diskreten Veränderlichen berechnet werden kann, sobald sein Index bekannt ist.* Wenn wir also die Regel zur angenäherten Berechnung der Wurzel kennen, können wir die ganze Folge der Näherungsbrüche für $\sqrt{2}$ als gegeben ansehen, obgleich wir keinen Ausdruck für das allgemeine Glied kennen.

Ist eine diskrete Veränderliche in diesem Sinne gegeben, so ist damit nicht nur die Menge der Werte, die sie annimmt, im ganzen gegeben, sondern auch die Reihenfolge bestimmt, in der sie diese Werte durchläuft: Jedem Index entspricht ein Wert der diskreten Veränderlichen, und aus den beiden Werten läßt sich der folgende berechnen, dessen Index größer ist.

Wir betonen noch einmal, daß die Werte einer diskreten Veränderlichen nicht unbedingt voneinander verschieden sein müssen. Ist beispielsweise eine diskrete Veränderliche durch eine der Formeln

$$x_n = 1, \quad x_n = (-1)^{n+1}, \quad x_n = \frac{1+(-1)^n}{n}$$

gegeben, so lauten die entsprechenden Folgen

1,	1,	1,	1,	1,	1,	...;	(Werte, Glieder der Folge)
1	2	3	4	5	6		(Index)
1,	−1,	1,	−1,	1,	−1,	...;	(Werte, Glieder der Folge)
1	2	3	4	5	6		(Index)
0,	1,	0,	$\frac{1}{2}$,	0,	$\frac{1}{3}$,	....	(Werte, Glieder der Folge)
1	2	3	4	5	6		(Index)

Im ersten Fall haben wir einfach eine konstante Größe; die ganze „Menge" der Werte, die von der Veränderlichen angenommen werden, reduziert sich auf einen einzigen. Im zweiten Fall besteht die Menge aus zwei Werten, 1 und −1, die abwechselnd angenommen werden. Im dritten Fall schließlich ist die Menge der Werte der Veränderlichen unendlich; das hindert aber die Veränderliche nicht, ein ums andere Mal den Wert 0 anzunehmen, und wir stellen fest, daß der Wert 0 an der fünften Stelle nicht nur auf den Wert 1 an der zweiten Stelle, sondern auch auf den Wert 0 an der ersten Stelle folgt.

23. Der Grenzwert einer diskreten Veränderlichen. Diesen Begriff hat der Leser vermutlich ebenfalls bereits in der Schule kennengelernt. Hier soll er jedoch genau definiert werden.

Eine konstante Zahl a heißt *Grenzwert einer diskreten Veränderlichen* $x = x_n$, wenn zu jeder noch so kleinen positiven Zahl ε eine Zahl N existiert derart, daß alle Werte x_n, deren Index größer ist als N, der Ungleichung

$$|x_n - a| < \varepsilon \tag{3}$$

genügen.

Den Sachverhalt, daß a der Grenzwert einer diskreten Veränderlichen ist, drückt man folgendermaßen aus:

$$\lim_{n\to\infty} x_n = a \quad \text{bzw.} \quad \lim x_n = a \quad \text{bzw.} \quad \lim x = a$$

(lim ist die Abkürzung des lateinischen Wortes limes = Grenze).

Man sagt auch, *die Veränderliche x_n strebe gegen a*, und schreibt

$$x_n \to a \quad \text{bzw.} \quad x \to a.$$

Mitunter wird die Zahl a auch *Grenzwert der Folge* (2) statt Grenzwert der Veränderlichen genannt, und man sagt, diese Folge *konvergiere* (statt die Veränderliche *strebe*) gegen a.

Die gleiche Definition kann kurz folgendermaßen formuliert werden:

Die Zahl a ist *Grenzwert der diskreten Veränderlichen* $x = x_n$, wenn sich ihre Werte von einer bestimmten Stelle an (dafür sagt man auch „zuletzt" oder „schließlich") beliebig wenig von a unterscheiden.

Die Ungleichung (3), in der ε beliebig ist, ist gerade die genaue Umschreibung der Aussage, daß sich x_n von a „beliebig wenig unterscheidet", und die Zahl N gibt gerade die Stelle an, von der an dieser Sachverhalt eintritt.

Es ist wichtig, sich darüber klar zu werden, daß die Zahl N im allgemeinen nicht ein für allemal angegeben werden kann, sondern von der Wahl der Zahl ε abhängt. Um dies hervorzuheben, werden wir statt N mitunter N_ε schreiben. Eine Verkleinerung der Zahl ε hat im allgemeinen eine Vergrößerung von $N = N_\varepsilon$ zur Folge: Je näher die Werte der diskreten Veränderlichen x_n bei dem Wert a liegen sollen, desto größer werden die Indizes derjenigen Glieder in der Folge (2) sein, die dafür erst in Betracht kommen.

Eine Ausnahme bildet der Fall, daß alle Werte der diskreten Veränderlichen x_n gleich der konstanten Zahl a sind. Offenbar ist dann $a = \lim x_n$, und diesmal ist die Ungleichung (3) für beliebiges $\varepsilon > 0$ gleichzeitig für alle Werte x_n erfüllt.[1]) Die Ungleichung (3) ist, wie wir aus Nr. 17 wissen, gleichwertig mit den Ungleichungen $-\varepsilon < x_n - a < \varepsilon$ oder

$$a - \varepsilon < x_n < a + \varepsilon; \tag{4}$$

dies werden wir im folgenden oft benutzen.

Stellt man die Zahlen a, $a \pm \varepsilon$ und die Werte x_n unserer diskreten Veränderlichen durch Punkte auf der Zahlengeraden (Nr. 21) dar (Abb. 2), so erhält man eine anschauliche geometrische Deutung des Grenzwertes der Folge. Wie klein auch immer

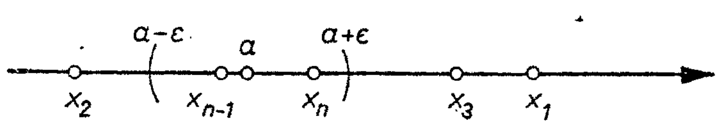

Abb. 2

[1]) Analog liegen die Verhältnisse bei einer diskreten Veränderlichen x_n, deren Werte von irgendeiner Stelle an gleich a sind.

eine Strecke (der Länge 2ε) mit dem Mittelpunkt im Punkt a gewählt sein möge, schließlich müssen alle Punkte x_n von einem bestimmten Punkt an in das Innere dieser Strecke fallen (so daß außerhalb dieser Strecke tatsächlich nur endlich viele dieser Punkte liegen).[1]) Der Punkt, der den Grenzwert a darstellt, ist sozusagen Sammelpunkt eines Haufens von Punkten, die die Werte der diskreten Veränderlichen darstellen.

24. Unendlich kleine Größen. Von besonderem Interesse ist der Fall, daß die Veränderliche gegen 0 strebt, $x_n \to 0$.

Eine diskrete Veränderliche x_n, die den Grenzwert 0 hat, nennt man eine *unendlich kleine Größe* oder auch kurz *unendlich klein*. Eine Folge mit dem Grenzwert 0 wird *Nullfolge* genannt.

Setzt man in der Definition des Grenzwertes einer diskreten Veränderlichen (Nr. 23) für a den Wert 0 ein, so erhält die Ungleichung (3) die Gestalt

$$|x_n - 0| = |x_n| < \varepsilon \quad (\text{für } n > N_\varepsilon).$$

Somit kann die obige Definition einer unendlich kleinen Größe auch formuliert werden, ohne daß der Terminus „Grenzwert" benutzt wird: Eine diskrete Veränderliche x_n wird *unendlich klein* genannt, wenn sie ihrem Absolutbetrag nach von einer bestimmten Stelle an kleiner als eine beliebig klein vorgegebene Zahl $\varepsilon > 0$ wird und bleibt.

Der nicht ganz glückliche (historisch entstandene) Terminus „unendlich kleine Größe" darf den Leser nicht zu einem Irrtum verleiten. Kein einzelner Wert dieser Größe kann — wenn er nicht gleich 0 ist — „klein" genannt werden. Der Kern der Sache besteht darin, daß es sich um eine *veränderliche* Größe handelt[2]), welche nur im Prozeß ihrer Änderung fähig ist, kleiner als jede beliebig gewählte positive Zahl ε zu werden. Statt *unendlich klein* sagt man vielfach auch *beliebig klein*.

Wir kehren zum allgemeinen Fall einer diskreten Veränderlichen x_n zurück, die den Grenzwert a besitzt. Die Differenz

$$\alpha_n = x_n - a$$

zwischen der Veränderlichen und ihrem Grenzwert ist offenbar unendlich klein; denn es gilt auf Grund von (3)

$$|\alpha_n| = |x_n - a| < \varepsilon \quad (\text{für } n > N_\varepsilon).$$

Ist umgekehrt jedoch α_n unendlich klein, so gilt $x_n \to a$. Dies führt uns auf folgende Aussage:

Eine diskrete Veränderliche x_n besitzt dann und nur dann (oder *genau dann*) *eine konstante Zahl a als Grenzwert, wenn die Differenz zwischen der Veränderlichen und ihrem Grenzwert, $\alpha_n = x_n - a$, unendlich klein ist* (*d. h. eine Nullfolge bildet*).

Man kann also für den Begriff „Grenzwert" noch eine andere (der ersten gleichwertige) Definition geben: Eine konstante Zahl a heißt *Grenzwert einer diskreten Veränderlichen x_n*, wenn die Differenz zwischen den Gliedern der Folge und der Zahl a eine unendlich kleine Größe ist.

---

[1]) Daher findet sich in der älteren Literatur auch folgende Definition des Grenzwertes einer Folge: a ist Grenzwert der Folge $\{x_n\}$, wenn für jedes $\varepsilon > 0$ fast alle x_n (das bedeutet in diesem Zusammenhang: alle, mit endlich vielen Ausnahmen) der Ungleichung $|x_n - a| < \varepsilon$ genügen. — *Anm. d. Red.*

[2]) Ausgenommen ist der uninteressante Fall, daß alle Glieder der Folge gleich 0 sind.

Gehen wir von dieser Definition des Grenzwertes aus, so versteht es sich von selbst, daß wir dann für den Begriff „unendlich klein“ die zweite der oben angegebenen Definitionen benutzen müssen (vgl. S. 47). Sonst würde man nämlich zu einem Zirkelschluß gelangen: Der Grenzwert wäre durch den Begriff „unendlich klein“ und der Begriff „unendlich klein“ durch den des „Grenzwertes“ definiert. Wenn also eine Veränderliche x_n gegen a strebt, dann kann sie auch in der Form

$$x_n = a + \alpha_n$$

dargestellt werden, wobei α_n unendlich klein ist. Läßt umgekehrt eine Veränderliche x_n eine solche Darstellung zu, so hat sie den Grenzwert a. Dies wird in der Praxis oft zur Bestimmung des Grenzwertes einer Veränderlichen benutzt.

25. Beispiele.

1. Wir betrachten die diskreten Veränderlichen

$$x_n = \frac{1}{n}, \quad x_n = -\frac{1}{n}, \quad x = \frac{(-1)^{n+1}}{n};$$

ihnen entsprechen die Wertefolgen

$$1, \quad \frac{1}{2}, \quad \frac{1}{3}, \quad \frac{1}{4}, \quad \ldots;$$

$$-1, \quad -\frac{1}{2}, \quad -\frac{1}{3}, \quad -\frac{1}{4}, \quad \ldots;$$

$$1, \quad -\frac{1}{2}, \quad \frac{1}{3}, \quad -\frac{1}{4},$$

Alle drei Veränderliche sind unendlich kleine Größen, d. h., sie haben den Grenzwert 0. Denn für sie gilt

$$|x_n| = \frac{1}{n} < \varepsilon,$$

sobald $n > \frac{1}{\varepsilon}$ ist. Somit kann man z. B. für N_ε die größte ganze Zahl wählen, die in $\frac{1}{\varepsilon}$ enthalten ist; wir bezeichnen sie mit $E\left(\frac{1}{\varepsilon}\right)$ (oder auch mit $\left[\frac{1}{\varepsilon}\right]$).[1])

Es sei noch erwähnt, daß die erste Veränderliche immer größer als ihr Grenzwert 0 ist, die zweite immer kleiner als dieser Grenzwert und die dritte abwechselnd größer und kleiner als dieser.

2. Setzt man

$$x_n = \frac{2 + (-1)^n}{n},$$

so durchläuft die Veränderliche die Wertefolge $1, \frac{3}{2}, \frac{1}{3}, \frac{3}{4}, \frac{1}{5}, \frac{3}{6}, \ldots$ Auch in diesem Fall gilt $x_n \to 0$, da $|x_n| \leqq \frac{3}{n} < \varepsilon$ für $n > \frac{3}{\varepsilon}$ ist, so daß man für N_ε den Wert $\left[\frac{3}{\varepsilon}\right]$ nehmen kann.

Wir haben es hier mit einer interessanten Besonderheit zu tun: Bei einem „Schritt“ nähert sich die Veränderliche ihrem Grenzwert 0, beim nächsten entfernt sie sich von ihm.

---

[1]) Allgemein bezeichnen wir mit $[p] = E(p)$ die größte ganze Zahl, die p nicht übertrifft, oder, mit anderen Worten, den *ganzen* Teil der Zahl p. (E ist der Anfangsbuchstabe des französischen Wortes entier = ganz.)

3. Es sei jetzt

$$x_n = \frac{1 + (-1)^n}{n}.$$

Mit dieser Folge haben wir uns schon am Schluß von Nr. 22 befaßt. Hier gilt ebenfalls $x_n \to 0$, weil

$$|x_n| \leqq \frac{2}{n} < \varepsilon$$

wird, sobald $n > N_\varepsilon = \left[\frac{2}{\varepsilon}\right]$ ist.

Wir weisen darauf hin, daß für alle ungeraden Werte von n die Veränderliche gleich ihrem Grenzwert ist.

Diese einfachen Beispiele sind deshalb interessant, weil sie die Vielfalt der Möglichkeiten charakterisieren, die durch unsere Definition des Grenzwertes einer Folge erfaßt wird. Es ist unwesentlich, ob sich die Werte der Veränderlichen nur von einer Seite her dem Grenzwert nähern oder nicht; ebenso ist es unwesentlich, ob sich die Veränderliche mit jedem Schritt ihrem Grenzwert nähert. Und schließlich ist es unwesentlich, ob die Veränderliche ihren Grenzwert erreicht, d. h., ob sie Werte annimmt, die gleich dem Grenzwert sind. Wesentlich ist nur das, was in der Definition zum Ausdruck kommt: Die Veränderliche darf sich zuletzt, d. h. für alle hinreichend großen Indizes, vom Grenzwert nur beliebig wenig unterscheiden.

4. Nehmen wir jetzt ein komplizierteres Beispiel für eine Folge:

$$x_n = \frac{n^2 - n + 2}{3n^2 + 2n - 4}.$$

Wir zeigen, daß ihr Grenzwert die Zahl $\frac{1}{3}$ ist. Zu diesem Zweck betrachten wir die Differenz

$$x_n - \frac{1}{3} = \frac{-5n + 10}{3(3n^2 + 2n - 4)}$$

und schätzen ihren absoluten Betrag ab. Für $n > 2$ gilt

$$\left|x_n - \frac{1}{3}\right| = \frac{5n - 10}{3(3n^2 + 2n - 4)} < \frac{5n}{3(3n^2 - 4)} < \frac{5n}{2 \cdot 2n^2} < \frac{1}{n},$$

so daß dieser Ausdruck kleiner als ε wird, wenn $n > N_\varepsilon = \left[\frac{1}{\varepsilon}\right]$ ist. Hiermit ist bewiesen, daß x_n gegen $\frac{1}{3}$ strebt.

5. Wir definieren eine Folge durch

$$x_n = a^{1/n} = \sqrt[n]{a} \qquad (a > 1)$$

und zeigen, daß $x_n \to 1$ gilt.

Benutzt man die Ungleichung (3) aus Nr. 19, so kann man schreiben:

$$|x_n - 1| = \sqrt[n]{a} - 1 < \frac{a-1}{n} < \varepsilon, \quad \text{sobald} \quad n > N_\varepsilon = \left[\frac{a-1}{\varepsilon}\right] \quad \text{ist.}$$

Es ist jedoch möglich, auch anders zu argumentieren. Die Ungleichung

$$|x_n - 1| = a^{1/n} - 1 < \varepsilon$$

ist gleichwertig mit

$$\frac{1}{n} < \log_\alpha (1 + \varepsilon) \quad \text{oder} \quad n > \frac{1}{\log_\alpha (1 + \varepsilon)},$$

so daß sie für $n > N_\varepsilon = \left[\frac{1}{\log_\alpha (1 + \varepsilon)}\right]$ erfüllt ist.

Je nach der Art der Überlegung kamen wir zu verschiedenen Ausdrücken für N_ε. Zum Beispiel erhalten wir für $a = 10$, $\varepsilon = 0{,}01$ nach der ersten Methode $N_{0,01} = \frac{9}{0{,}01} = 900$ und $N_{0,01} = \left[\frac{1}{0{,}00432\ldots}\right] = 231$ nach der zweiten. Mit Hilfe der zweiten Methode erhielten wir den kleinsten der möglichen Werte für $N_{0,01}$, denn schon $10^{1/231} = 1{,}010017\ldots$ unterscheidet sich von 1 um mehr als um $\varepsilon = 0{,}01$. Dasselbe tritt auch im allgemeinen Fall ein, da, wie leicht zu sehen ist, für $n \leqq \frac{1}{\log_\alpha (1 + \varepsilon)}$ notwendigerweise $a^{1/n} - 1 \geqq \varepsilon$ ist.

In diesem Zusammenhang bemerken wir, daß es gar nicht darauf ankommt, gerade den kleinstmöglichen Wert von N_ε zu kennen, wenn es sich nur um den Nachweis der Tatsache handelt, daß eine Folge einen Grenzwert hat. Es kommt nur darauf an, daß die Ungleichung (3) von einem gewissen Index an erfüllt ist, aber nicht darauf, ob dieser Index groß oder klein ist.

6. Ein wichtiges Beispiel für eine unendlich kleine Größe gibt die diskrete Veränderliche

$$\alpha_n = q^n \quad \text{für} \quad |q| < 1.$$

Um zu beweisen, daß α_n gegen 0 strebt, betrachten wir die Ungleichung

$$|\alpha_n| = |q|^n < \varepsilon;$$

sie ist gleichwertig mit $n \cdot \log |q| < \log \varepsilon$ oder $n > \frac{\log \varepsilon}{\log |q|}$.[1] Setzen wir also (für $\varepsilon < 1$, was wir ohne Beschränkung der Allgemeinheit annehmen können)

$$N_\varepsilon = \left[\frac{\log \varepsilon}{\log |q|}\right],$$

so ist für $n > N_\varepsilon$ die erwähnte Ungleichung sicher erfüllt.

Analog überzeugt man sich leicht davon, daß die diskrete Veränderliche $\beta_n = A \cdot q^n$, wobei wie eben $|q| < 1$ und A eine konstante Zahl ist, unendlich klein ist.

7. Wir betrachten ferner die unendliche fallende geometrische Folge

$$a, aq, aq^2, \ldots, aq^{n-1}, \ldots \qquad (|q| < 1)$$

und werfen das Problem auf, die Summe der entsprechenden unendlichen Reihe zu bilden (vgl. auch Beispiel 9).

Unter der *Summe* einer unendlichen Reihe versteht man bekanntlich[2] den Grenzwert, gegen den die Summe s_n der ersten n Glieder bei unbegrenzt wachsendem n strebt. Nun ist

$$s_n = \frac{a - aq^n}{1 - q} = \frac{a}{1 - q} - \frac{a}{1 - q} \cdot q^n,$$

so daß s_n von der festen Zahl $\frac{a}{1 - q}$ um die Größe $\alpha_n = -\frac{a}{1 - q} \cdot q^n$ abweicht, die, wie wir

[1]) Unter $\log x$ verstehen wir hier (und auch weiterhin) $\log_{10} x$. Es ist zu beachten, daß $|q| < 1$, also $\log |q| < 0$ ist; daher muß sich bei Division durch diese Zahl die Ungleichung umkehren.

[2]) In den Beispielen knüpfen wir vielfach an Schulkenntnisse an, obwohl der Schulstoff meist nicht wissenschaftlich einwandfrei dargeboten wird. Im Laufe des Studiums dieses Werkes wird der Leser die exakten Definitionen aller verwendeten Begriffe kennenlernen. — *Anm. d. Red.*

bereits sahen, unendlich klein ist. Folglich gilt nach der zweiten Definition des Grenzwertes für die gesuchte Summe der Reihe:

$$s = \lim s_n = \frac{a}{1-q}.$$

Demnach ist diese Zahl die Summe der unendlich vielen Glieder der Reihe[1]):

$$a + aq + aq^2 + \cdots + aq^{n-1} + \cdots = \frac{a}{1-q}.$$

8. Es seien zwei Zahlen a und b gegeben. Wir setzen $x_0 = a$ und $x_1 = b$ und definieren die folgenden Werte der Folge x_n durch

$$x_n = \frac{x_{n-2} + x_{n-1}}{2} \quad (n \geqq 2).$$

Dadurch ist die Folge der x_n tatsächlich gegeben, da sich, wenn wir $n = 2, 3, 4, \ldots$ annehmen, sukzessive alle ihre Werte bis zu jedem beliebigen einschließlich ermitteln lassen.

Subtrahieren wir auf beiden Seiten der vorstehenden Gleichung x_{n-1}, so erhalten wir

$$x_n - x_{n-1} = -\frac{1}{2} \cdot (x_{n-1} - x_{n-2}) \qquad (n = 2, 3, 4, \ldots).$$

Damit ergibt sich in der Folge der Differenzen

$$x_1 - x_0 = b - a, \quad x_2 - x_1, \quad \ldots, \quad x_{n-1} - x_{n-2}, \quad x_n - x_{n-1}, \ldots$$

jede (beginnend mit der zweiten) aus der vorhergehenden durch Multiplikation mit $-\frac{1}{2}$, d. h., es liegt eine geometrische Folge mit dem Quotienten $-\frac{1}{2}$ vor. Da die Summe von n ihrer Glieder gleich $x_n - a$ ist, erhalten wir unter Benutzung der aus Beispiel 7 bekannten Formel für die Summe einer Progression sofort die Beziehung

$$\lim (x_n - a) = \frac{b-a}{1 - \left(-\frac{1}{2}\right)} = \frac{2}{3} \cdot (b - a).$$

Daraus ergibt sich leicht

$$\lim x_n = a + \frac{2}{3} \cdot (b - a) = \frac{a + 2b}{3}.$$

9. Ähnlich wie die geometrische Folge können wir eine beliebige Zahlenfolge

$$a_1, a_2, a_3, \ldots, a_n, \ldots$$

betrachten und durch Addition (unter Beibehaltung der Reihenfolge) die „Teilsummen" bilden:

$$A_1 = a_1, \quad A_2 = a_1 + a_2, \quad A_3 = a_1 + a_2 + a_3, \ldots,$$
$$A_n = a_1 + a_2 + \cdots + a_n, \ldots$$

Strebt bei unbegrenzt wachsendem n der Ausdruck A_n gegen den (endlichen oder unendlichen) Grenzwert A, so nennt man diese Zahl A die *Summe* aller Zahlen a_n und schreibt

$$a_1 + a_2 + \cdots + a_n + \cdots = A.$$ [2])

Das Symbol auf der linken Seite der Gleichung nennt man eine *unendliche Reihe.*

---

[1]) Dafür schreibt man auch unter Verwendung des Summenzeichens $\sum\limits_{\nu=0}^{\infty} aq^\nu = \frac{a}{1-q}$.

[2]) oder $\sum\limits_{\nu=1}^{\infty} a_\nu = A$

Von Reihen, deren Summe *endlich* ist, dagt man, daß sie *konvergieren.*

Es sei z. B. die Reihe

$$\frac{1}{1\cdot 2}+\frac{1}{2\cdot 3}+\frac{1}{3\cdot 4}+\cdots+\frac{1}{n(n+1)}+\cdots$$

gegeben. Hier gilt[1])

$$a_1=\frac{1}{1\cdot 2}=1-\frac{1}{2},\quad a_2=\frac{1}{2\cdot 3}=\frac{1}{2}-\frac{1}{3},\ldots,$$

$$a_n=\frac{1}{n(n+1)}=\frac{1}{n}-\frac{1}{n+1},\ldots,$$

so daß

$$A_n=\left(1-\frac{1}{2}\right)+\left(\frac{1}{2}-\frac{1}{3}\right)+\cdots+\left(\frac{1}{n}-\frac{1}{n+1}\right)=1-\frac{1}{n+1}$$

ist. Offenbar strebt A_n gegen 1, so daß die vorgegebene Reihe konvergiert und die Summe 1 besitzt.

(Hat eine Reihe *keine endliche* Summe, so sagt man, sie *divergiere.* Dies ist z. B. bei der Reihe

$$a+1+\cdots+1+\cdots$$

der Fall.)

26. Einige Sätze über Folgen, die einen Grenzwert haben. Die diskrete Variable x_n habe den Grenzwert a. Für beliebiges $p < a$ (oder $q > a$) läßt sich leicht eine Zahl $\varepsilon > 0$ angeben derart, daß

$$a-\varepsilon > p \quad (\text{oder } a+\varepsilon < q)$$

ist; hierzu genügt es, ε kleiner als die Differenz $a - p$ (oder $q - a$) zu wählen. Nun läßt sich aber nach Definition des Grenzwertes (Nr. 23) eine Zahl N angeben derart, daß für $n > N$ die Ungleichung

$$x_n > a-\varepsilon \quad (x_n < a+\varepsilon)$$

erfüllt ist [vgl. (4)], also erst recht die Ungleichung

$$x_n > p \quad (\text{oder } x_n < q).$$

1. *Strebt eine diskrete Veränderliche x_n gegen den Grenzwert a und ist $a > p$ (bzw. $a < q$), so sind auch alle Werte der Veränderlichen von einer bestimmten Stelle an größer als p (bzw. kleiner als q).*

Aus diesem einfachen Satz läßt sich eine Reihe nützlicher Folgerungen ziehen.

2. *Strebt eine Folge $\{x_n\}$ gegen den Grenzwert $a > 0$ (bzw. < 0), so ist die Veränderliche x_n selbst von einer bestimmten Stelle an positiv (bzw. negativ).*

Zum Beweis braucht man im vorhergehenden Satz nur $p = 0$ (bzw. $q = 0$) zu setzen.

Damit läßt sich ein noch schärferes Ergebnis formulieren:

3. *Strebt eine Folge $\{x_n\}$ gegen einen Grenzwert $a \neq 0$, so sind wenigstens die Glieder mit hinreichend großen Indizes dem Absolutbetrag nach größer als eine positive Zahl r,*

$$|x_n| > r > 0 \quad (\text{für } n > N).$$

[1]) Dies ist eine sogenannte *Partialbruchzerlegung. — Anm. d. Red.*

In der Tat kann man bei $a > 0$ (bzw. < 0)

$$0 < p < a \quad (\text{bzw. } a < q < 0)$$

wählen und $r = p$ (bzw. $r = |q|$) setzen.

4. *Hat andererseits eine Folge $\{x_n\}$ den Grenzwert a, so ist sie beschränkt in folgendem Sinne: Alle ihre Glieder sind dem Absolutbetrag nach nicht größer als eine endliche Schranke:*

$$|x_n| \leqq M \quad (M = \text{const}; \; n = 1, 2, 3, \ldots).$$

Wir wählen eine Zahl $M' > |a|$, so daß also $-M' < a < M'$ gilt, und setzen $p = -M'$ und $q = M'$. Dann läßt sich eine Zahl N angeben derart, daß für $n > N$

$$-M' < x_n < M' \quad \text{oder} \quad |x_n| < M'$$

gilt. Diese Ungleichung ist sicher für $n = N + 1, N + 2, \ldots$ erfüllt, so daß ihr höchstens die ersten N Werte unserer Folge (oder irgendeiner von ihnen) nicht genügen.

Wählen wir also M gleich der größten der Zahlen

$$|x_1|, |x_2|, \ldots, |x_N|, M',$$

so gilt für alle Werte von x_n die Beziehung $|x_n| \leqq M$, was zu beweisen war.

Bemerkungen.

I. Man kann die Beschränktheit einer Veränderlichen x_n auch anders definieren; man fordert, daß die Ungleichungen

$$k \leqq x_n \leqq g \qquad (n = 1, 2, 3, \ldots)$$

erfüllt sind, wobei k und g zwei endliche Zahlen sind. Natürlich ist diese Definition der ersten gleichwertig. Aus diesen Ungleichungen folgt nämlich, wenn man M gleich der größeren der Zahlen $|k|$, $|g|$ setzt, $|x_n| \leqq M$. Gilt umgekehrt die Ungleichung $|x_n| \leqq M$ für alle n, so kann sie in der Form $-M \leqq x_n \leqq M$ gechrieben werden, so daß $-M$ die Rolle von k und M die Rolle von g spielt.

II. Die Aussage 4 ist nicht umkehrbar. Nicht jede beschränkte Folge hat einen Grenzwert. Nehmen wir als Beispiel $x_n = (-1)^{n+1}$, so ist diese Folge natürlich beschränkt: $|x_n| \leqq 1$; sie besitzt jedoch keinen Grenzwert. Die Glieder sind abwechselnd gleich $+1$ und -1.

Zum Schluß beweisen wir die Eindeutigkeit des Grenzwertes, indem wir uns auf Satz 1 stützen.

5. *Eine Folge $\{x_n\}$ kann nicht zugleich gegen zwei verschiedene Grenzwerte streben.*

Zum Beweis nehmen wir das Gegenteil an. Es gelte also gleichzeitig $x_n \to a$ und $x_n \to b$, wobei etwa $a < b$ sei. Wir wählen eine beliebige Zahl r zwischen a und b:

$$a < r < b.$$

Wegen $x_n \to a$ und $a < r$ gibt es eine Zahl N' derart, daß für alle $n > N'$ die Ungleichung $x_n < r$ erfüllt ist. Andererseits läßt sich wegen $x_n \to b$ und $b > r$ eine Zahl N'' angeben derart, daß für alle $n > N''$ die Beziehung $x_n > r$ gilt. Wählt man den Index n sowohl größer als N' als auch größer als N'', so müßte der entsprechende Wert der Veränderlichen x_n einerseits größer als r, andererseits kleiner als r sein, was unmöglich ist. Dieser Widerspruch beweist unsere Behauptung.

27. Unendlich große Größen. Den unendlich kleinen Größen lassen sich die unendlich großen Größen gegenüberstellen.

Eine diskrete Veränderliche x_n wird *unendlich groß* genannt, wenn ihr absoluter Betrag von einer bestimmten Stelle an größer als eine beliebig groß vorgegebene Zahl $E > 0$ ist und bleibt:

$$|x_n| > E \quad (\text{für } n > N_E).$$

Wie im Fall der unendlich kleinen Größen sei auch hier hervorgehoben, daß kein einzelner Wert einer unendlich großen Größe „groß" genannt werden kann. Wir haben es hier mit einer veränderlichen Größe zu tun, welche nur im Prozeß ihrer Änderung größer als jede beliebig gewählte Zahl wird.

Als Beispiele für unendlich große Größen können die Folgen

$$x_n = n, \qquad x_n = -n, \qquad x_n = (-1)^{n+1}\, n$$

dienen, welche die ganzen Zahlen durchlaufen: die erste die positiven, die zweite die negativen, während die dritte bei jedem Schritt das Vorzeichen wechselt.

Außerdem sei hier ein weiteres Beispiel einer unendlich großen Größe angegeben:

$$x_n = Q^n \quad \text{für} \quad |Q| > 1.$$

Wie auch immer die Zahl $E > 1$ gewählt sein mag, die Ungleichung

$$|x_n| = |Q|^n > E$$

ist erfüllt, sobald $n \cdot \log|Q| > \log E$ oder $n > \dfrac{\log E}{\log|Q|}$ ist[1]), so daß wir für N_E die Zahl $\left[\dfrac{\log E}{\log|Q|}\right]$ wählen können.

Ist eine diskrete Veränderliche x_n unendlich groß und behält sie (wenigstens für hinreichend großes n) ein bestimmtes Vorzeichen (+ oder −) bei, so sagt man, *die Folge* $\{x_n\}$ *besitze* (je nach dem Vorzeichen) *den Grenzwert* $+\infty$ *oder* $-\infty$, und man schreibt

$$\lim_{n\to\infty} x_n = +\infty \quad \text{oder} \quad \lim x_n = +\infty \quad \text{oder} \quad x_n \to +\infty$$

bzw.

$$\lim_{n\to\infty} x_n = -\infty \quad \text{oder} \quad \lim x_n = -\infty \quad \text{oder} \quad x_n \to -\infty.$$

Für diese Fälle lassen sich auch unabhängige Definitionen angeben, indem man die Ungleichung $|x_n| > E$ durch die Ungleichung

$$x_n > E \quad \text{bzw.} \quad x_n < -E$$

ersetzt, woraus bereits folgt, daß $x_n > 0$ bzw. $x_n < 0$ ist.

Offenbar läßt sich eine unendlich große Größe x_n im allgemeinen Fall durch die Relation $|x_n| \to +\infty$ charakterisieren.

Von den oben angeführten Beispielen unendlich großer Größen strebt offenbar die diskrete Veränderliche $x_n = n$ gegen $+\infty$ und $x_n = -n$ gegen $-\infty$. Dagegen kann man von $x_n = (-1)^{n-1}\, n$ weder sagen, daß sie gegen $+\infty$ noch daß sie gegen $-\infty$ strebt. Von $x_n = Q^n$ schließlich kann man sagen, daß sie für $Q > 1$ gegen $+\infty$ strebt und daß sie für $Q < -1$ keinen Grenzwert besitzt.

Mit den uneigentlichen Zahlen $\pm\infty$ haben wir uns bereits in Nr. 11 beschäftigt, man muß daran denken, daß ihre Verwendung auf reinen Vereinbarungen beruht,

---

[1]) Wegen $|Q| > 1$ ist $\log|Q| > 0$.

und sich hüten, mit diesen „Zahlen" irgendwelche arithmetischen Operationen auszuführen. Oft schreibt man statt $+\infty$ einfach ∞, und auch wir werden von jetzt an diese Schreibweise benutzen.

Die Einführung unendlicher Grenzwerte verstößt nicht gegen den Satz über die Eindeutigkeit des Grenzwertes, den wir in Nr. 26 bewiesen haben (vgl. Aussage 5). Wie dort gezeigt wurde (vgl. Aussage 4), ist eine Folge, die einen endlichen Grenzwert a besitzt, beschränkt und kann daher keinesfalls gleichzeitig gegen einen unendlichen Grenzwert streben.

Zum Schluß erwähnen wir noch einen einfachen Zusammenhang, der zwischen unendlich großen und unendlich kleinen Größen besteht: *Ist eine diskrete Veränderliche x_n unendlich groß, so ist ihr reziproker Wert $\alpha_n = \frac{1}{x_n}$ unendlich klein.*

Zum Beweis nehmen wir eine beliebige Zahl $\varepsilon > 0$. Wegen $|x_n| \to \infty$ läßt sich zu der Zahl $E = \frac{1}{\varepsilon}$ eine Zahl N derart angeben, daß

$$|x_n| > \frac{1}{\varepsilon}$$

gilt, sobald $n > N$ ist. Dann gilt für alle diese n offenbar

$$|\alpha_n| < \varepsilon,$$

womit unsere Behauptung bewiesen ist.

Analog beweist man die Umkehrung: *Ist eine nicht verschwindende diskrete Veränderliche α_n unendlich klein, so ist ihr reziproker Wert $x_n = \frac{1}{\alpha_n}$ unendlich groß.*

§ 2. Sätze über Grenzwerte, die ihre rechnerische Bestimmung erleichtern

28. Grenzübergänge bei Gleichungen und Ungleichungen. Sind zwei Folgen $\{x_n\}$ und $\{y_n\}$ durch ein Gleichheits- oder Ungleichheitszeichen miteinander verknüpft, so meinen wir stets, daß alle entsprechenden Glieder gleich bzw. ungleich sind, d. h., daß alle Glieder mit ein und demselben Index übereinstimmen bzw. daß für sie die entsprechende Ungleichheitsbeziehung gilt.

1. *Stimmen zwei Folgen $\{x_n\}$, $\{y_n\}$ in allen ihren Werten überein, d. h., gilt $x_n = y_n$ für $n = 1, 2, \ldots$, und besitzt jede der Folgen einen endlichen Grenzwert,*

$$\lim x_n = a, \quad \lim y_n = b,$$

so stimmen auch die Grenzwerte überein, $a = b$.

Dies folgt unmittelbar aus der Eindeutigkeit des Grenzwertes (Nr. 26, Aussage 5). Diesen Satz benutzt man gewöhnlich in der Form des Grenzübergangs in einer Gleichung: Aus $x_n = y_n$ folgert man $\lim x_n = \lim y_n$.

2. *Ist für zwei Folgen $\{x_n\}$, $\{y_n\}$ stets die Ungleichung $x_n \geqq y_n$ erfüllt und hat jede der Folgen einen endlichen Grenzwert,*

$$\lim x_n = a, \quad \lim y_n = b,$$

so gilt $a \geqq b$.

Wäre nämlich $a < b$, so könnten wir wie in Nr. 26, Aussage 5, schließen und eine Zahl r zwischen a und b wählen: $a < r < b$. Dann ließe sich einerseits eine Zahl N' angeben derart, daß für $n > N'$ die Beziehung $x_n < r$ gelten würde, andererseits eine solche Zahl N'', daß sich für $n > N''$ die Beziehung $y_n > r$ ergäbe. Ist N größer als die beiden Zahlen N', N'', so wären für $n > N$ gleichzeitig die beiden Ungleichungen

$$x_n < r, \quad y_n > r$$

erfüllt, woraus $x_n < y_n$ folgen würde, was im Widerspruch zur Annahme steht. Damit ist der Satz bewiesen.

Aus diesem Satz geht hervor, daß *man bei Ungleichungen (bei denen Gleichheit zugelassen ist) zur Grenze übergehen darf.* Aus $x_n \geqq y_n$ kann man schließen, daß $\lim x_n \geqq \lim y_n$ ist. Natürlich gilt die entsprechende Aussage, wenn man das Zeichen $\geqq$ durch das Zeichen $\leqq$ ersetzt.

Wir machen den Leser jedoch darauf aufmerksam, daß aus der echten Ungleichung $x_n > y_n$ im allgemeinen nicht auch die echte Ungleichung $\lim x_n > \lim y_n$ folgt, sondern nur $\lim x_n \geqq \lim y_n$. So ist z. B. $\frac{1}{n} > -\frac{1}{n}$ für alle n, aber

$$\lim \frac{1}{n} = \lim \left(-\frac{1}{n}\right) = 0.$$

Zum Nachweis der Existenz und zur Bestimmung des Grenzwertes einer Folge wird oft folgender Satz von Nutzen sein:

3. *Sind für die Folgen $\{x_n\}$, $\{y_n\}$, $\{z_n\}$ stets die Ungleichungen $x_n \leqq y_n \leqq z_n$ erfüllt und streben die Folgen $\{x_n\}$ und $\{z_n\}$ gegen den gemeinsamen Grenzwert a,*

$$\lim x_n = \lim z_n = a,$$

so besitzt auch die Folge $\{y_n\}$ den gleichen Grenzwert:

$$\lim y_n = a.$$

Wir geben ein beliebiges $\varepsilon > 0$ vor. Zu diesem ε können wir zunächst eine Zahl N' angeben derart, daß für $n > N'$

$$a - \varepsilon < x_n < a + \varepsilon$$

gilt. Damit läßt sich aber auch eine solche Zahl N'' angeben, daß für $n > N''$ die Beziehung

$$a - \varepsilon < z_n < a + \varepsilon$$

erfüllt ist. Nun sei N größer als die beiden Zahlen N' und N''. Dann sind für $n > N$ beide vorstehenden Ungleichungen erfüllt, und es gilt daher

$$a - \varepsilon < x_n \leqq y_n \leqq z_n < a + \varepsilon.$$

Also gilt insbesondere für $n > N$

$$a - \varepsilon < y_n < a + \varepsilon \quad \text{oder} \quad |y_n - a| < \varepsilon.$$

Also ist tatsächlich $\lim y_n = a$.

Aus diesem Satz folgt als Spezialfall: *Gilt $a \leqq y_n \leqq z_n$ für alle n und ist bekannt, daß $z_n \to a$, so gilt auch $y_n \to a$.* Übrigens ist dies sehr leicht auch direkt zu zeigen.

Die Sätze 1, 2 und 3 lassen sich leicht auf den Fall unendlicher Grenzwerte erweitern.

29. Hilfssätze über unendlich kleine Größen. In den folgenden Sätzen haben wir gleichzeitig zwei (oder mehrere) Folgen zu betrachten, die durch arithmetische Grundrechenarten miteinander verknüpft sind. Wie vorhin soll es sich dabei stets um Folgenglieder mit gleichen Indizes handeln.

Ist beispielsweise von der Summe zweier diskreter Veränderlicher x_n und y_n die Rede, die einzeln die Werte

$$x_1, x_2, x_3, \ldots, x_n, \ldots \quad \text{und} \quad y_1, y_2, y_3, \ldots, y_n, \ldots$$

durchlaufen, so haben wir unter $\{x_n + y_n\}$ die Folge der Werte

$$x_1 + y_1, x_2 + y_2, x_3 + y_3, \ldots, x_n + y_n, \ldots$$

zu verstehen.

Beim Beweis der Sätze, die sich auf die Resultate arithmetischer Rechenoperationen mit diskreten Veränderlichen beziehen, werden die beiden folgenden Hilfssätze über unendlich kleine Größen eine wichtige Rolle spielen.

Lemma 1. *Die Summe endlich vieler unendlich kleiner Größen ist eine unendlich kleine Größe.*

Wir führen den Beweis für den Fall zweier unendlich kleiner Größen α_n und β_n (der allgemeine Fall läßt sich analog behandeln oder durch Induktion erledigen). Es sei $\varepsilon > 0$ beliebig vorgegeben. Nach Definition der unendlich kleinen Größen läßt sich zur Zahl ε für unendlich kleines α_n eine Zahl N' angeben derart, daß für $n > N'$

$$|\alpha_n| < \frac{\varepsilon}{2}$$

gilt. Ebenso kann man für die unendlich kleine Größe β_n eine solche Zahl N'' finden, daß für $n > N''$

$$|\beta_n| < \frac{\varepsilon}{2}$$

gilt. Wählen wir die natürliche Zahl N größer als jede der beiden Zahlen N' und N'', so sind für $n > N$ diese beiden Ungleichungen gleichzeitig erfüllt, so daß

$$|\alpha_n + \beta_n| \leqq |\alpha_n| + |\beta_n| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

ist. Also ist $\alpha_n + \beta_n$ in der Tat eine unendlich kleine Größe.

Lemma 2. *Das Produkt einer beschränkten Veränderlichen x_n mit einer unendlich kleinen Größe α_n ist eine unendlich kleine Größe.*

Es sei für alle natürlichen Werte von n

$$|x_n| \leqq M \quad (M > 0).$$

Ist eine beliebige Zahl $\varepsilon > 0$ vorgegeben, so läßt sich zu der Zahl $\frac{\varepsilon}{M}$ für die unendlich kleine Größe α_n eine Zahl N angeben derart, daß für $n > N$

$$|\alpha_n| < \frac{\varepsilon}{M}$$

gilt. Dann ist offenbar auch für alle natürlichen $n > N$ die Beziehung

$$|x_n \cdot \alpha_n| = |x_n| \cdot |\alpha_n| < M \cdot \frac{\varepsilon}{M} = \varepsilon$$

erfüllt, und hieraus folgt, daß $x_n \alpha_n$ eine unendlich kleine Größe ist.

30. Arithmetische Operationen mit Veränderlichen. Die folgenden Sätze sind deshalb wichtig, weil sie es ermöglichen, daß man in vielen Fällen nicht immer wieder auf die Definition des Grenzwertes zurückzugreifen braucht, daß also nicht immer wieder zu gegebenem ε eine N angegeben werden muß, usw. Dadurch wird das Rechnen mit Grenzwerten bedeutend vereinfacht.

1. *Haben die beiden Folgen $\{x_n\}$ und $\{y_n\}$ endliche Grenzwerte, d. h., ist*

$$\lim x_n = a, \quad \lim y_n = b,$$

so besitzt auch ihre Summe (Differenz) einen endlichen Grenzwert, und es ist

$$\lim (x_n \pm y_n) = a \pm b.$$

Aus der Voraussetzung des Satzes folgt $x_n = a + a_n$ und $y_n = b + \beta_n$, wobei α_n und β_n unendlich kleine Größen sind. Dann ist

$$x_n \pm y_n = (a \pm b) + (\alpha_n \pm \beta_n).$$

Hier ist $\alpha_n \pm \beta_n$ nach Lemma 1 unendlich klein; daher kann man nach der zweiten Definition des Grenzwertes schließen, daß die Folge $\{x_n \pm y_n\}$ den Grenzwert $a \pm b$ besitzt, was zu beweisen war.

Dieser Satz und sein Beweis lassen sich auf den Fall endlich vieler Summanden übertragen.

2. *Haben die beiden Folgen $\{x_n\}$ und $\{y_n\}$ endliche Grenzwerte,*

$$\lim x_n = a, \quad \lim y_n = b,$$

so besitzt auch ihr Produkt einen endlichen Grenzwert, und es gilt

$$\lim x_n y_n = ab.$$

Wenn wir wieder von den Gleichungen (1) ausgehen, erhalten wir

$$x_n y_n = ab + (a\beta_n + b\alpha_n + \alpha_n \beta_n).$$

Der Ausdruck in der Klammer ist auf Grund von Lemma 1 und Lemma 2 eine unendlich kleine Größe. Daraus folgt aber, daß die Folge $\{x_n y_n\}$ tatsächlich den Grenzwert ab besitzt. Dieser Satz läßt sich auf den Fall einer beliebigen endlichen Anzahl von Faktoren ausdehnen (etwa durch Induktion).

3. *Haben die beiden Folgen $\{x_n\}$ und $\{y_n\}$ endliche Grenzwerte,*

$$\lim x_n = a, \quad \lim y_n = b,$$

und ist b ungleich 0, so besitzt auch ihr Quotient einen endlichen Grenzwert, und zwar ist

$$\lim \frac{x_n}{y_n} = \frac{a}{b}.$$

Wegen $b \neq 0$ ist nach Satz 3 aus Nr. 26 von einer bestimmten Stelle an also nicht nur $y_n \neq 0$, sondern sogar $|y_n| > r > 0$, wobei r eine Konstante ist.

Beschränken wir uns auf die Werte der Zahl n, für welche dies erfüllt ist, dann hat der Quotient $\frac{x_n}{y_n}$ offenbar einen Sinn.

Gehen wir, wie vorher, von den Gleichungen (1) aus, so erhalten wir

$$\frac{x_n}{y_n} - \frac{a}{b} = \frac{a + \alpha_n}{b + \beta_n} - \frac{a}{b} = \frac{1}{by_n}(b\alpha_n - a\beta_n).$$

Der Ausdruck in Klammern ist gemäß Lemma 1 und Lemma 2 unendlich klein.

Auf Grund des zu Anfang des Beweises Gesagten ist der Faktor vor der Klammer eine beschränkte Veränderliche,

$$\left|\frac{1}{by_n}\right| = \frac{1}{|b| \cdot |y_n|} < \frac{1}{|b|\, r}.$$

Folglich ist nach Lemma 2 das ganze Produkt auf der rechten Seite unendlich klein; es ist aber gerade die Differenz zwischen der diskreten Veränderlichen $\frac{x_n}{y_n}$ und der Zahl $\frac{a}{b}$. Also ist der Grenzwert der Folge $\left\{\frac{x_n}{y_n}\right\}$ die Zahl $\frac{a}{b}$, was zu zeigen war.

31. Unbestimmte Ausdrücke. In Nr. 30 betrachteten wir Ausdrücke der Gestalt

$$x_n \pm y_n, \quad x_n y_n, \quad \frac{x_n}{y_n}$$

unter der Voraussetzung, daß die Folgen $\{x_n\}$ und $\{y_n\}$ gegen einen endlichen Grenzwert streben (wobei im letzten Fall der Grenzwert von y_n nicht gleich 0 sein durfte), und berechneten den Grenzwert eines jeden dieser Ausdrücke.

Von der Betrachtung ausgeschlossen blieben die Fälle, daß die Grenzwerte einer der Veränderlichen x_n und y_n (oder beider) *unendlich* oder (im letzten Fall) der Grenzwert des Nenners 0 waren. Von diesen Fällen untersuchen wir hier nur vier, die eine wichtige und interessante Besonderheit aufweisen.

1. Wir betrachten zunächst den Bruch $\frac{x_n}{y_n}$ und nehmen an, daß beide Veränderliche x_n und y_n gleichzeitig gegen 0 streben.

Hier stoßen wir das erste Mal auf einen ganz besonderen Sachverhalt: Obgleich uns die Grenzwerte von x_n und y_n bekannt sind, können wir über den Grenzwert der Folge ihrer Quotienten (wenn wir diese Folge selbst nicht kennen) nichts aussagen. *Dieser Grenzwert kann je nach der Art und Weise, nach der sich beide Variablen ändern, ganz verschiedene Werte annehmen; er braucht nicht einmal zu existieren.* Die folgenden einfachen Beispiele sollen dies erläutern.

Es sei, sagen wir, $x_n = \frac{1}{n^2}$ und $y_n = \frac{1}{n}$. Beide Folgen streben gegen 0. Ihr Quotient $\frac{x_n}{y_n} = \frac{1}{n}$ strebt ebenfalls gegen 0.

Nehmen wir jedoch umgekehrt $x_n = \frac{1}{n}$ und $y_n = \frac{1}{n^2}$ an, so strebt, obgleich wie vorher beide Folgen den Grenzwert 0 haben, diesmal ihr Quotient $\frac{x_n}{y_n} = n$ gegen ∞.

Wenn wir eine beliebige von 0 verschiedene Zahl a nehmen und die beiden unendlich kleinen Größen $x_n = \frac{a}{n}$ und $y_n = \frac{1}{n}$ konstruieren, so sehen wir, daß ihr Quotient $\frac{x_n}{y_n}$ den Grenzwert a besitzt (da alle Glieder der Folge gleich a sind).

Ist schließlich $x_n = \frac{(-1)^{n+1}}{n}$, $y_n = \frac{1}{n}$ (beide haben den Grenzwert 0), so hat der Quotient $\frac{x_n}{y_n} = (-1)^{n+1}$ offenbar überhaupt keinen Grenzwert.

Somit erlaubt die bloße Kenntnis der Grenzwerte der Folgen $\{x_n\}$ und $\{y_n\}$ im gegebenen Fall noch kein Urteil über das Verhalten der Folge ihrer Quotienten. Man muß diese Folge selbst kennen, d. h. das Bildungsgesetz, und den Quotienten $\frac{x_n}{y_n}$ unmittelbar untersuchen.

Um diesen Sachverhalt zu charakterisieren, sagt man, im Fall $x_n \to 0$ und $y_n \to 0$ sei die diskrete Veränderliche $\frac{x_n}{y_n}$ ein *unbestimmter Ausdruck der Form* $\frac{0}{0}$.

2. In dem Fall, daß gleichzeitig $x_n \to \pm\infty$ und $y_n \to \pm\infty$ gilt, tritt ein ähnlicher Umstand ein. Kennt man nur die Folgen selbst, so ist es unmöglich, eine Aussage über das Verhalten der Folge der Quotienten zu machen. Diese Tatsache möge durch Beispiele veranschaulicht werden, die den in Fall 1 angegebenen völlig entsprechen:

$$x_n = n \to \infty, \qquad y_n = n^2 \to \infty, \qquad \frac{x_n}{y_n} = \frac{1}{n} \to 0;$$

$$x_n = n^2 \to \infty, \qquad y_n = n \to \infty, \qquad \frac{x_n}{y_n} = n \to \infty;$$

$$x_n = an \to \pm\infty \ (a \neq 0), \qquad y_n = n \to \infty, \qquad \frac{x_n}{y_n} = a \to a;$$

$$x_n = [2 + (-1)^{n+1}]\, n \to \infty, \quad y_n = n \to \infty, \qquad \frac{x_n}{y_n} = 2 + (-1)^{n+1},$$

und diese Folge besitzt überhaupt keinen Grenzwert. In diesem Fall sagt man, die diskrete Veränderliche $\frac{x_n}{y_n}$ sei ein *unbestimmter Ausdruck der Form* $\frac{\infty}{\infty}$.

Wenden wir uns jetzt der Untersuchung des Produkts $x_n y_n$ zu.

3. Strebt x_n gegen 0, während y_n gegen $\pm\infty$ strebt, so stoßen wir, wenn wir das Verhalten des Produkts $x_n y_n$ untersuchen, auf dieselbe Besonderheit wie in Fall 1 und 2. Hierüber geben die folgenden Beispiele Aufschluß:

$$x_n = \frac{1}{n^2} \to 0, \qquad y_n = n \to \infty, \qquad x_n y_n = \frac{1}{n} \to 0;$$

$$x_n = \frac{1}{n} \to 0, \qquad y_n = n^2 \to \infty, \qquad x_n y_n = n \to \infty;$$

$$x_n = \frac{a}{n} \to 0 \quad (a \neq 0), \quad y_n = n \to \infty, \quad x_n y_n = a \to a;$$

$$x_n = \frac{(-1)^{n+1}}{n} \to 0, \qquad y_n = n \to \infty, \quad x_n y_n = (-1)^{n+1},$$

und diese Folge besitzt überhaupt keinen Grenzwert.

In diesem Zusammenhang sagt man bei $x_n \to 0$ und $y_n \to \infty$, das Produkt $x_n y_n$ sei ein *unbestimmter Ausdruck der Form* $0 \cdot \infty$

Wir betrachten zum Schluß die Summe $x_n + y_n$.

4. Hier kann der besondere Fall eintreten, daß x_n und y_n gegen unendliche Werte mit verschiedenen Vorzeichen streben. Dann läßt sich im allgemeinen über die Summe $x_n + y_n$ nichts Bestimmtes sagen. Die verschiedenen Möglichkeiten, die hier auftreten können, werden durch die folgenden Beispiele veranschaulicht:

$$\begin{aligned}
&x_n = 2n \to \infty, && y_n = -n \to -\infty, && x_n + y_n = n \to \infty;\\
&x_n = n \to \infty, && y_n = -2n \to -\infty, && x_n + y_n = -n \to -\infty;\\
&x_n = n + a \to \infty, && y_n = -n \to -\infty, && x_n + y_n = a \to a;\\
&x_n = n + (-1)^{n+1} \to \infty, && y_n = -n \to -\infty, && x_n + y_n = (-1)^{n+1},
\end{aligned}$$

und diese Folge besitzt überhaupt keinen Grenzwert. In Anbetracht dessen sagt man bei $x_n \to \infty$ und $y_n \to -\infty$, die Veränderliche $x_n - y_n$ sei ein *unbestimmter Ausdruck der Form* $\infty - \infty$.

Wir hatten uns die Aufgabe gestellt, die Grenzwerte arithmetischer Ausdrücke (2) durch die Grenzwerte ihrer Komponenten (Summanden, Faktoren usw.) zu bestimmen; dabei haben wir vier Fälle gefunden, in denen dies nicht möglich war: Wir stießen auf unbestimmte Ausdrücke der Form[1])

$$\frac{0}{0}, \quad \frac{\infty}{\infty}, \quad 0 \cdot \infty, \quad \infty - \infty.$$

In diesen Fällen muß man unter Berücksichtigung der für die Folgen $\{x_n\}$ und $\{y_n\}$ geltenden Bildungsgesetze den interessierenden Ausdruck direkt untersuchen.

Dieses Verfahren wird *Auswertung unbestimmter Ausdrücke* genannt. Es ist bei weitem nicht immer so einfach wie in den oben angeführten schematischen Beispielen. Im folgenden erwähnen wir einige interessante Fälle dieser Art.

32. Beispiele für die Bestimmung von Grenzwerten.

1. Es sei $p(n)$ ein Polynom in n mit konstanten Koeffizienten,

$$p(n) = a_0 n^k + a_1 n^{k-1} + \cdots + a_{k-1} n + a_k \qquad (a_0 \neq 0).$$

Wir wollen den Grenzwert dieses Polynoms für $n \to \infty$ ermitteln. Sind alle Koeffizienten positiv (bzw. negativ), dann ist offenbar ∞ (bzw. $-\infty$) der Grenzwert. Ist ein Teil der Koeffizienten positiv, der andere negativ, so strebt ein Teil der Glieder gegen ∞, der andere gegen $-\infty$, und damit liegt ein unbestimmter Ausdruck der Form $\infty - \infty$ vor. Zu seiner Auswertung schreiben wir $p(n)$ in der Form

$$p(n) = n^k \left(a_0 + \frac{a_1}{n} + \cdots + \frac{a_{k-1}}{n^{k-1}} + \frac{a_k}{n^k}\right).$$

[1]) Natürlich bedeuten diese Symbole keine Zahlen. Jedes von ihnen ist nur eine Abkürzung für einen der vier Typen unbestimmter Ausdrücke.

Da alle Summanden in der Klammer vom zweiten an bei wachsendem n unendlich klein werden, hat der Ausdruck in der Klammer den Grenzwert a_0. Der erste Faktor strebt gegen ∞. In diesem Fall strebt also der ganze Ausdruck gegen ∞ oder $-\infty$, je nach dem Vorzeichen von a_0.

Dieses hier benutzte Verfahren, unbestimmte Ausdrücke auf dem Wege der Umformung auszuwerten, wird häufig angewandt.

2. Ist $q(n)$ ein ebensolches Polynom der Form

$$q(n) = b_0 n^l + b_1 n^{l-1} + \cdots + b_{l-1} n + b_l,$$

so liefert der Quotient $\frac{p(n)}{q(n)}$ für wachsendes n einen unbestimmten Ausdruck der Form $\frac{\infty}{\infty}$. Wenn wir hier jedes Polynom so umformen wie in Beispiel 1, so erhalten wir

$$\frac{p(n)}{q(n)} = n^{k-l} \cdot \frac{a_0 + \frac{a_1}{n} + \cdots + \frac{a_k}{n^k}}{b_0 + \frac{b_1}{n} + \cdots + \frac{b_l}{n^l}}.$$

Der zweite Faktor hat hier (im Fall $b_0 \neq 0$) den endlichen Grenzwert $\frac{a_0}{b_0}$. Haben die Polynome gleichen Grad, d. h., ist $k = l$, so ist der Grenzwert des Quotienten $\frac{p(n)}{q(n)}$ ebenfalls gleich $\frac{a_0}{b_0}$.[1]) Für $k > l$ strebt der erste Faktor gegen $+\infty$, so daß der betrachtete Quotient gegen $\pm\infty$ strebt (je nach dem Vorzeichen von $\frac{a_0}{b_0}$). Schließlich strebt für $l < l$ der erste Faktor und mit ihm der ganze Ausdruck gegen 0.

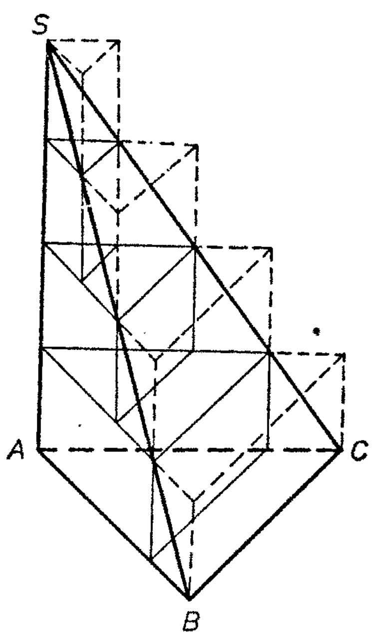

Abb. 3

3. Man bestimme das Volumen V der Dreieckspyramide $SABC$ (Abb. 3). Dabei wollen wir uns auf die Anschauung stützen. Wir teilen die Höhe H der Pyramide in n gleiche Teile und legen durch die Teilpunkte zur Grundfläche parallele Ebenen. Als Schnittflächen erhalten wir zur Grundfläche ähnliche Dreiecke. Auf ihnen errichten wir je ein System einbeschriebener und umbeschriebener Prismen. Erstere bilden einen Körper mit dem Volumen V_n, letztere einen Körper mit dem Volumen V_n', wobei offenbar

$$V_n < V < V_n'$$

[1]) Auf diese Weise hätte man in Nr. 25, Beispiel 4, sofort den Grenzwert $\frac{1}{3}$ erhalten.

gilt. Die Differenz $V'_n - V_n$ ist aber nichts anderes als das Volumen des unteren umbeschriebenen Prismas mit der Grundfläche $Q = \triangle ABC$ und der Höhe $\frac{H}{n}$. Also gilt für die Differenz:

$$V'_n - V_n = \frac{QH}{n} \to 0$$

bei unbegrenzt wachsendem n. Dann streben die Differenzen $V - V_n$ und $V'_n - V$ erst recht gegen 0, d. h., es gilt

$$V = \lim V_n = \lim V'_n.$$

Wir suchen jetzt einen Ausdruck für V'_n. Es handelt sich dabei um einen Körper, der aus einer Reihe umbeschriebener Prismen besteht. Ihre Grundflächen sind als Schnitte der Pyramide (Strahlensatz!), von oben nach unten aufgezählt, gleich

$$\frac{1}{n^2}Q, \quad \frac{2^2}{n^2}Q, \ldots, \quad \frac{i^2}{n^2}Q, \ldots, \quad \frac{n^2}{n^2}Q = Q,$$

während ihre Höhen sämtlich gleich $\frac{H}{n}$ sind. Daher ist

$$V'_n = \frac{Q}{n^2}(1^2 + 2^2 + \cdots + n^2)\frac{H}{n} = \frac{QH}{n^3}\frac{n(n+1)(2n+1)}{6}$$
$$= \frac{QH}{6}\frac{(n+1)(2n+1)}{n^2},[1]$$

also

$$V = \lim V'_n = \frac{QH}{3}.$$

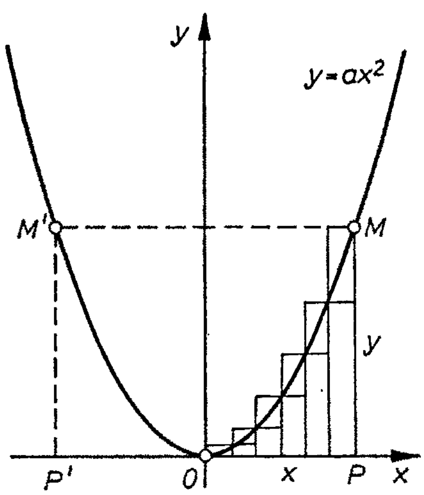

Abb. 4

4. Man bestimme den Flächeninhalt Q der Figur OPM, die von dem Stück OM der Parabel $y = ax^2$ $(a > 0)$, der Strecke OP der x-Achse und der Strecke PM begrenzt ist (Abb. 4).

Wir zerlegen die Strecke OP in n gleiche Teile und errichten über ihnen eine Reihe einbeschriebener und umbeschriebener Rechtecke. Die Flächeninhalte Q_n und Q'_n der aus ihnen bestehenden treppenförmigen Figuren unterscheiden sich um den Flächeninhalt $\frac{x}{n}y$ des größten Rechtecks. Daher strebt wie in Beispiel 3 die Differenz $Q'_n - Q_n$ gegen 0, und da

$$Q_n < Q < Q'_n$$

---

[1]) Hier benutzen wir die bekannte Formel für die Summe der Quadrate der ersten n natürlichen Zahlen.

ist, gilt offenbar

$$Q = \lim Q_n = \lim Q_n'.$$

Da die Höhen der einzelnen Rechtecke die Ordinaten der Parabelpunkte mit den Abszissen

$$\frac{1}{n}x,\ \frac{2}{n}x, \ldots, \frac{i}{n}x, \ldots, \frac{n}{n}x = x$$

sind, also gemäß der Gleichung der Parabel gleich

$$a \cdot \frac{1}{n^2}x^2,\ a \cdot \frac{2^2}{n^2}x^2, \ldots, a \cdot \frac{i^2}{n^2}x^2, \ldots, ax^2,$$

erhalten wir für Q_n' den Ausdruck

$$Q_n' = \frac{ax^2}{n^2}(1^2 + 2^2 + \cdots + n^2)\frac{x}{n} = \frac{ax^3}{6}\frac{(n+1)(2n+1)}{n^2}.$$

Hieraus folgt

$$Q = \lim Q_n' = \frac{ax^3}{3} = \frac{x \cdot ax^2}{3} = \frac{xy}{3}.$$

Somit ergibt sich leicht, daß der Flächeninhalt des Parabelsegments $M'OM$ gleich $\frac{4}{3}xy$ ist, d. h. $\frac{2}{3}$ des Flächeninhaltes des Rechtecks $M'P'PM$. (Dieses Resultat war schon ARCHIMEDES bekannt.)[1])

5. Man beweise, daß für $0 < k < 1$

$$\lim\left[(n+1)^k - n^k\right] = 0$$

ist. Hier liegt ein unbestimmter Ausdruck der Form $\infty - \infty$ vor. Wir formen ihn um, indem wir n^k ausklammern:

$$0 < (n+1)^k - n^k = n^k\left[\left(1 + \frac{1}{n}\right)^k - 1\right] < n^k\left[\left(1 + \frac{1}{n}\right) - 1\right] = \frac{1}{n^{1-k}}.$$

Wegen $\frac{1}{n^{1-k}} \to 0$ gilt erst recht $(n+1)^k - n^k \to 0$, was beweisen zu war.

6. Man bestimme den Grenzwert der diskreten Veränderlichen

$$x_n = \sqrt{n}\left(\sqrt{n+1} - \sqrt{n}\right),$$

einen unbestimmten Ausdruck der Form $\infty \cdot 0$ (nach Beispiel 5). Durch Erweitern mit $\sqrt{n+1} + \sqrt{n}$ transformieren wir den gegebenen Ausdruck auf die Form $\frac{\infty}{\infty}$:

$$x = \sqrt{n} \cdot \frac{\left(\sqrt{n+1} - \sqrt{n}\right)\left(\sqrt{n+1} + \sqrt{n}\right)}{\sqrt{n+1} + \sqrt{n}} = \frac{\sqrt{n}}{\sqrt{n+1} + \sqrt{n}};$$

wenn wir jetzt Zähler und Nenner durch $\sqrt{n}$ dividieren, erhalten wir

$$x_n = \frac{1}{\sqrt{1 + \frac{1}{n}} + 1};$$

---

[1]) Die allgemeine Definition des Flächeninhaltes krummlinig begrenzter Figuren wird erst in Band II, Kap. X, gegeben. Die hier verwendete Methode der Berechnung des Flächeninhaltes läßt sich auf andere krummlinige Figuren verallgemeinern.

offenbar gilt $1 < \sqrt{1 + \frac{1}{n}} < 1 + \frac{1}{n}$; da der Ausdruck auf der rechten Seite gegen 1 strebt, gilt dies auch für die Wurzel. Daher ist

$$\lim x_n = \frac{1}{2}.$$

7. Man bestimme die Grenzwerte der diskreten Veränderlichen

$$x_n = \frac{n}{\sqrt{n^2 + n}}, \quad y_n = \frac{n}{\sqrt{n^2 + 1}}$$

und schließlich den der diskreten Veränderlichen

$$z_n = \frac{1}{\sqrt{n^2 + 1}} + \frac{1}{\sqrt{n^2 + 2}} + \cdots + \frac{1}{\sqrt{n^2 + i}} + \cdots + \frac{1}{\sqrt{n^2 + n}}.$$

Die Veränderlichen x_n und y_n liefern unbestimmte Ausdrücke der Form $\frac{\infty}{\infty}$ (da beide Wurzeln größer als n sind, streben diese gegen ∞). Wir formen um, indem wir Zähler und Nenner durch n dividieren:

$$x_n = \frac{1}{\sqrt{1 + \frac{1}{n}}}, \quad y_n = \frac{1}{\sqrt{1 + \frac{1}{n^2}}}.$$

Da die im Nenner stehenden Wurzeln beide den Grenzwert 1 besitzen (vgl. Beispiel 6), gilt $x_n \to 1$ und $y_n \to 1$.

Der Ausdruck für z_n hat eine eigenartige Form: Jeder Summand dieser Summe hängt von n ab, und die Anzahl der Summanden wächst mit n.[1]) Da jeder Summand kleiner als der erste und größer als der letzte ist, gilt

$$\frac{n}{\sqrt{n^2 + n}} < z_n < \frac{n}{\sqrt{n^2 + 1}}, \quad \text{d. h.} \quad x_n < z_n < y_n.$$

Die Folgen $\{x_n\}$ und $\{y_n\}$ streben aber (wie bereits gezeigt) gegen den gemeinsamen Grenzwert 1. Folglich strebt nach Satz 3 aus Nr. 28 die Folge $\{z_n\}$ ebenfalls gegen diesen Grenzwert.

8. Es seien m positive Zahlen $a_1, a_2, \ldots, a_m$ gegeben. Es sei A die größte von ihnen; man zeige, daß

$$\lim \sqrt[n]{a_1^n + a_2^n + \cdots + a_m^n} = A$$

ist. Das ergibt sich aus den trivialen Ungleichungen

$$A \leqq \sqrt[n]{a_1^n + a_2^n + \cdots + a_m^n} \leqq A \cdot \sqrt[n]{m}$$

(vgl. Nr. 25, Beispiel 5).

9. Wir sahen in Nr. 27, daß für $a > 1$ die Potenz a^n (mit wachsendem n) gegen ∞ strebt. Nun wollen wir das Verhalten des Quotienten

$$\frac{a^n}{n^k}$$

(für $k > 0$) untersuchen, eines unbestimmten Ausdrucks der Form $\frac{\infty}{\infty}$. Dazu stellen wir eine Hilfsungleichung auf (vgl. die Bernoullische Ungleichung in Nr. 19). Wir setzen $a = 1 + \lambda$,

[1]) Dieselbe Besonderheit wiesen übrigens auch die Ausdrücke für V' und Q' in Beispiel 3 und 4 auf.

$\lambda > 0$; dann ist nach dem binomischen Satz

$$a^n = (1 + \lambda)^n = 1 + n\lambda + \frac{n(n-1)}{2}\lambda^2 + \cdots > \frac{n(n-1)}{2}\lambda^2.$$

Da für $n > 2$ offenbar $n - 1 > \frac{n}{2}$ ist, folgt

$$a^n > \frac{n^2}{4}\lambda^2 = \frac{(a-1)^2}{4} \cdot n^2. \tag{3}$$

Für $k = 1$ erhalten wir sofort

$$\frac{a^n}{n} > \frac{(a-1)^2}{4} n,$$

also

$$\lim \frac{a^n}{n} = \infty.$$

Da dieses Ergebnis für jedes $a > 1$ gilt, ergibt sich, da für $k > 1$ (wenigstens für hinreichend große n)

$$\frac{a^n}{n^k} = \left[\frac{(a^{1/k})^n}{n}\right]^k > \frac{(a^{1/k})^n}{n}$$

ist, die Beziehung

$$\lim \frac{a^n}{n^k} = \infty \qquad (a > 1).$$

Dieses somit für $k \geqq 1$ bewiesene Ergebnis gilt erst recht für $k < 1$.

10. Mit Hilfe der Ungleichung (3) kann man auch die Beziehung

$$\lim \sqrt[n]{n} = 1$$

herleiten. Setzen wir nämlich $a = \sqrt[n]{n}$, so erhalten wir

$$n > \frac{n^2}{4}\left(\sqrt[n]{n} - 1\right)^2;$$

hieraus folgt

$$0 < \sqrt[n]{n} - 1 < \frac{2}{\sqrt{n}},$$

was zu dem gesuchten Resultat führt.

11. Wir wollen nun einen anderen interessanten Grenzwert ermitteln, nämlich

$$\lim \frac{\log_a n}{n} \qquad (a > 1).$$

(Hier liegt wieder ein unbestimmter Ausdruck der Form $\frac{\infty}{\infty}$ vor, denn, wie man leicht zeigt, gilt $\log_a n \to \infty$.) Ist $\varepsilon > 0$ beliebig, so folgt wegen $a^\varepsilon > 1$ für hinreichend große n (Nr. 26, Satz 1) $\sqrt[n]{n} < a^\varepsilon$. Bilden wir den Logarithmus zur Basis a, so erhalten wir $\frac{\log_a n}{n} < \varepsilon$, und hieraus folgt

$$\lim \frac{\log_a n}{n} = 0.$$

33. Der Stolzsche Satz und seine Anwendung. Zur Bestimmung von Grenzwerten unbestimmter Ausdrücke des Typs $\frac{\infty}{\infty}$ ist der folgende Satz von O. STOLZ oft sehr dienlich.[1])

Für die Folge $\{y_n\}$ gelte $y_n \to \infty$, wobei wenigstens von einer Stelle an, mit wachsendem n auch y_n wächst, $y_{n+1} > y_n$. Dann ist

$$\lim \frac{x_n}{y_n} = \lim \frac{x_n - x_{n-1}}{y_n - y_{n-1}},$$

wenn die rechte Seite Sinn hat, d. h., wenn der Grenzwert rechts existiert (er kann endlich oder unendlich sein).

Nehmen wir zunächst an, dieser Grenzwert sei eine endliche Zahl l:

$$\lim \frac{x_n - x_{n-1}}{y_n - y_{n-1}} = l.$$

Dann gibt es zu jedem vorgegebenen $\varepsilon > 0$ eine Zahl N derart, daß für $n > N$

$$\left|\frac{x_n - x_{n-1}}{y_n - y_{n-1}} - l\right| < \frac{\varepsilon}{2} \quad \text{oder} \quad -\frac{\varepsilon}{2} < \frac{x_n - x_{n-1}}{y_n - y_{n-1}} < l + \frac{\varepsilon}{2}$$

ist. Das bedeutet, wie auch immer $n > N$ gewählt sein mag, daß alle Brüche

$$\frac{x_{N+1} - x_N}{y_{N+1} - y_N}, \frac{x_{N+2} - x_{N+1}}{y_{N+2} - y_{N+1}}, \ldots, \frac{x_{n-1} - x_{n-2}}{y_{n-1} - y_{n-2}}, \frac{x_n - x_{n-1}}{y_n - y_{n-1}}$$

zwischen diesen Grenzen liegen. Da ihre Nenner positiv sind (y_n sollte ja mit n wachsen), liegt auch der Bruch $\frac{x_n - x_N}{y_n - y_N}$, dessen Zähler die Summe aller Zähler der oben angegebenen Brüche und dessen Nenner die Summe aller ihrer Nenner ist, zwischen diesen Grenzen. Daher gilt für $n > N$

$$\left|\frac{x_n - x_N}{y_n - y_N} - l\right| < \frac{\varepsilon}{2}.$$

Wir schreiben jetzt die leicht zu beweisende Identität

$$\frac{x_n}{y_n} - l = \frac{x_N - l y_N}{y_n} + \left(1 - \frac{y_N}{y_n}\right)\left(\frac{x_n - x_N}{y_n - y_N} - l\right)$$

auf. Daraus folgt

$$\left|\frac{x_n}{y_n} - l\right| \leqq \left|\frac{x_N - l y_N}{y_n}\right| + \left|\frac{x_n - x_N}{y_n - y_N} - l\right|.$$

Der zweite Summand auf der rechten Seite wird für $n > N$, wir wie sahen, kleiner als $\frac{\varepsilon}{2}$; der erste Summand (dessen Zähler eine feste Zahl ist) wird wegen $y_n \to \infty$ ebenfalls kleiner als $\frac{\varepsilon}{2}$, etwa für $n > N'$. Wählen wir $N' > N$, so ist für $n > N'$ offenbar

$$\left|\frac{x_n}{y_n} - l\right| < \varepsilon,$$

womit unsere Behauptung bewiesen ist.

---

[1]) OTTO STOLZ, 1842—1905, österreichischer Mathematiker. Unter der speziellen Voraussetzung $y_n = n$ finden wir diesen Satz auch bei AUGUSTIN CAUCHY (1789—1857, französischer Mathematiker).

Der Fall eines unendlichen Grenzwertes läßt sich auf den des endlichen zurückführen. Es sei etwa

$$\lim \frac{x_n - x_{n-1}}{y_n - y_{n-1}} = \infty.$$

Hieraus ergibt sich zunächst (für hinreichend große n) $x_n - x_{n-1} > y_n - y_{n-1}$; folglich gilt mit $y_n \to \infty$ auch $x_n \to \infty$, wobei die Glieder der Folge $\{x_n\}$ mit zunehmendem n wachsen. In diesem Fall kann der schon bewiesene Teil des Satzes auf den reziproken Ausdruck $\frac{y_n}{x_n}$ angewandt werden:

$$\lim \frac{y_n}{x_n} = \lim \frac{y_n - y_{n-1}}{x_n - x_{n-1}} = 0$$

(weil hier der Grenzwert endlich ist), und hieraus folgt

$$\lim \frac{x_n}{y_n} = \infty,$$

was zu zeigen war.

Wir wenden uns nun weiteren Beispielen zu.

12. Wir sahen schon in Beispiel 9, daß für $a > 1$

$$\lim \frac{a^n}{n} = \infty$$

ist. Dieses Ergebnis erhält man sofort mit Hilfe des Stolzschen Satzes:

$$\lim \frac{a^n}{n} = \lim (a^n - a^{n-1}) = \lim a^n \left(1 - \frac{1}{a}\right) = +\infty.$$

Ebenso verhält es sich auch mit Beispiel 11.

13. Wir wenden den Stolzschen Satz zum Beweis des folgenden interessanten Satzes (der von Cauchy stammt) an:

Besitzt eine Folge $\{a_n\}$ einen (endlichen oder unendlichen) Grenzwert, so hat die Folge der

$$b_n = \frac{a_1 + a_2 + \cdots + a_n}{n}$$

(der arithmetischen Mittel der ersten n Glieder der Folge $\{a_n\}$) *denselben Grenzwert.*

In der Tat folgt, wenn wir im Stolzschen Satz

$$x_n = a_1 + a_2 + \cdots + a_n, \qquad y = n$$

setzen, die Beziehung

$$\lim b_n = \lim \frac{x_n}{y_n} = \lim \frac{x_n - x_{n-1}}{y_n - y_{n-1}} = \lim a_n.$$

Da wir schon wissen (vgl. Beispiel 10), daß $\sqrt[n]{n}$ gegen 1 strebt, gilt auch

$$\frac{1 + \sqrt{2} + \sqrt[3]{3} + \cdots + \sqrt[n]{n}}{n} \to 1.$$

14. Wir betrachten jetzt (k sei eine natürliche Zahl) die Folge der

$$z_n = \frac{1^k + 2^k + \cdots + n^k}{n^{k+1}},$$

die einen unbestimmten Ausdruck der Form $\frac{\infty}{\infty}$ liefert. Setzen wir im Stolzschen Satz

$$x_n = 1^k + 2^k + \cdots + n^k, \qquad y_n = n^{k+1},$$

so erhalten wir

$$\lim z_n = \lim \frac{n^k}{n^{k+1} - (n-1)^{k+1}}.$$

Nun ist aber $(n-1)^{k+1} = n^{k+1} - (k-1)\,n^k + \cdots$, also $n^{k+1} - (n-1)^{k+1} = (k+1)\,n^k + \cdots$ und (vgl. Beispiel 2, S. 62)

$$\lim z_n = \lim \frac{n^k}{(k+1)\,n^k + \cdots} = \frac{1}{k+1}.$$

15. Zum Schluß bestimmen wir den Grenzwert der diskreten Veränderlichen

$$u_n = n\left(z_n - \frac{1}{k+1}\right) = \frac{1^k + 2^k + \cdots + n^k}{n^k} - \frac{n}{k+1},$$

die in der ersten Gestalt ein unbestimmter Ausdruck der Form $\infty \cdot 0$, in der zweiten ein unbestimmter Ausdruck der Form $\infty - \infty$ ist.

Schreiben wir die Brüche auf einen Nenner, so erhalten wir einen unbestimmten Ausdruck der Form $\frac{\infty}{\infty}$:

$$u_n = \frac{(k+1)\,(1^k + 2^k + \cdots + n^k) - n^{k+1}}{(k+1)\,n^k}.$$

Setzen wir x_n gleich dem Zähler dieses Bruches und y_n gleich dem Nenner und wenden nochmals den gleichen Satz an, dann erhalten wir

$$\lim u_n = \lim \frac{(k+1)\,n^k - [n^{k+1} - (n-1)^{k+1}]}{(k+1)\,[n^k - (n-1)^k]}.$$

Nun ist aber

$$(k+1)\,n^k - [n^{k+1} - (n-1)^{k+1}] = \frac{(k+1)\,k}{2} \cdot n^{k-1} + \cdots$$

und

$$n^k - (n-1)^k = k n^{k-1} + \cdots,$$

so daß sich (vgl. Beispiel 2) schließlich

$$\lim u_n = \lim \frac{\frac{(k+1)\,k}{2}\, n^{k-1} + \cdots}{(k+1)\,k n^{k-1} + \cdots} = \frac{1}{2}$$

ergibt.

§ 3. Monotone Folgen

34. Grenzwert monotoner Folgen. Die Sätze über die Existenz von Grenzwerten diskreter Veränderlicher, die wir bisher behandelt haben, hatten folgenden Charakter: Unter der Voraussetzung, daß für einige der Folgen die Grenzwerte existieren, existieren auch die Grenzwerte für andere Folgen, die irgendwie mit den ersten zusammenhängen. Die Frage nach einem Kriterium für die Existenz eines endlichen Grenzwertes einer gegebenen Folge, unabhängig von anderen Folgen, wurde nicht gestellt.

Allgemein wird diese Frage in § 4 (Nr. 39 bis 42) beantwortet werden. Jetzt wollen wir einen einfachen und wichtigen Spezialfall von Veränderlichen betrachten, für den die Lösung leicht angegeben werden kann. Eine Folge $\{x_n\}$ heißt *wachsend*, wenn

$$x_1 < x_2 < \cdots < x_n < x_{n+1} < \cdots$$

ist, d. h., wenn aus $n' > n$ die Beziehung $x_{n'} > x_n$ folgt. Sie heißt *nicht fallend* (*nicht abnehmend*), wenn

$$x_1 \leqq x_2 \leqq \cdots \leqq x_n \leqq x_{n+1} \leqq \cdots$$

ist, d. h., wenn aus $n' > n$ nur $x_{n'} \geqq x_n$ folgt. Im letzten Fall kann man die Veränderliche ebenfalls (im weiteren Sinne) wachsend nennen; dann spricht man im ersten Fall von einer *im engeren Sinne* oder *echt wachsenden* oder *streng wachsenden Folge.*

Analog wird der Begriff einer (im weiteren Sinne) fallenden (abnehmenden) Folge eingeführt: Darunter versteht man eine Folge, für die

$$x_1 > x_2 > \cdots > x_n > x_{n+1} > \cdots$$

$$(\text{bzw.}\quad x_1 \geqq x_2 \geqq \cdots \geqq x_n \geqq x_{n+1} \geqq \cdots)$$

gilt, d. h., aus $n' > n$ die Beziehung $x_{n'} < x_n$ (bzw. nur $x_{n'} \leqq x$) folgt.

Folgen dieser Art, deren Glieder sich also nur in einer Richtung ändern, faßt man allgemein unter der Bezeichnung *monotone Folgen* zusammen. Meist sagt man, eine Folge sei *monton wachsend* bzw. *monoton fallend.*

Über monotone Folgen folgt folgender, sehr wichtiger

Satz. *Gegeben sei eine monoton wachsende Folge $\{x_n\}$. Ist sie nach oben beschränkt, d. h., ist*

$$x_n \leqq M \quad (M = \text{const}; \ n = 1, 2, 3, \ldots),$$

so existiert ein endlicher Grenzwert; anderenfalls strebt sie gegen ∞.

Ebenso hat eine monoton fallende Folge $\{x_n\}$ stets einen Grenzwert. Dieser Grenzwert ist endlich, wenn sie nach unten beschränkt ist, d. h., wenn

$$x_n \geqq m \quad (m = \text{const}; \ n = 1, 2, 3, \ldots)$$

ist; anderenfalls strebt sie gegen $-\infty$.

Beweis. Wir beschränken uns auf eine im weiteren Sinne wachsende Folge $\{x_n\}$; den Fall einer abnehmenden Folge kann man dann analog beweisen.

Wir nehmen an, daß diese Veränderliche nach oben beschränkt ist. Dann muß nach dem Satz aus Nr. 11 für die Menge $\{x_n\}$ ihrer Werte die (endliche) obere Grenze existieren,

$$a = \sup \{x_n\};$$

wie wir beweisen werden, *ist diese Zahl a auch der Grenzwert der Folge $\{x_n\}$.*

Wir erinnern an die charakteristischen Eigenschaften der oberen Grenze (Nr. 11). Erstens gilt für alle n

$$x_n \leqq a.$$

Zweitens existiert zu jedem $\varepsilon > 0$ eine Zahl N derart, daß

$$x_N > a - \varepsilon$$

ist.

Da die Folge monoton ist (das wird hier erstmalig benutzt), gilt für $n > N$ stets $x_n \geqq x_N$, d. h. erst recht $x_n > a - \varepsilon$, so daß für diese n die Ungleichung

$$0 \leqq a - x_n < \varepsilon \quad \text{oder} \quad |x_n - a| < \varepsilon$$

erfüllt ist; hieraus folgt $\lim x_n = a$.

Die Folge $\{x_n\}$ sei jetzt nicht nach oben beschränkt. Dann gibt es zu jeder noch so großen Zahl $E > 0$ immer noch einen Wert unserer Veränderlichen, der größer als E ist; dieser Wert sei x_N, also $x_N > E$. Da die Folge $\{x_n\}$ monoton ist, gilt für $n > N$ erst recht

$$x_n > E,$$

und das bedeutet $\lim x_n = \infty$.

Selbstverständlich gelten alle Schlußfolgerungen auch für Veränderliche, die erst von irgendeiner Stelle an monoton sind (weil man eine beliebige endliche Zahl von Anfangsgliedern einer Folge fortlassen kann, ohne daß dies den Grenzwert beeinflußt).

35. Beispiele.

1. Wir betrachten die diskrete Veränderliche $x_n = \frac{c^n}{n!}$; dabei sei $c > 0$ und wie üblich $n! = 1 \cdot 2 \cdot 3 \cdots n$. (Formal ist der Grenzwert für $c > 1$ ein unbestimmter Ausdruck der Form $\frac{\infty}{\infty}$).

Wegen

$$x_{n+1} = x_n \cdot \frac{c}{n+1}$$

ist die Veränderliche für $n > c - 1$ fallend; sie ist auch nach unten beschränkt, beispielsweise durch 0. Daher hat die Folge $\{x_n\}$ nach dem Satz aus Nr. 34 einen endlichen Grenzwert, den wir mit a bezeichnen.

Um ihn zu bestimmen, gehen wir in der obigen Gleichung zur Grenze über; da x_{n+1} dieselbe Wertefolge durchläuft wie x_n (das erste Glied ausgenommen), hat x_{n+1} auch denselben Grenzwert; wir erhalten auf diese Weise $a = a \cdot 0$, also $a = 0$ und schließlich, was für $0 < c \leqq 1$ trivial ist,

$$\lim \frac{c^n}{n!} = 0.$$

2. Wir setzen wieder $c > 0$ voraus und definieren jetzt die Folge $\{x_n\}$ folgendermaßen:

$$x_1 = \sqrt{c}, \quad x_2 = \sqrt{c + \sqrt{c}}, \quad x_3 = \sqrt{c + \sqrt{c + \sqrt{c}}}, \quad \ldots,$$

allgemein

$$x_n = \underbrace{\sqrt{c + \sqrt{c + \cdots + \sqrt{c}}}}_{n \text{ Wurzeln}}.$$

Somit erhalten wir x_{n+1} aus x_n durch

$$x_{n+1} = \sqrt{c + x_n}.$$

Offenbar wächst die Folge $\{x_n\}$ monoton. Sie ist, wie sogleich gezeigt wird, nach oben beschränkt, z. B. durch die Zahl $\sqrt{c} + 1$. In der Tat ist $x_1 = \sqrt{c}$ kleiner als diese Zahl. Nehmen wir jetzt an, es sei $x_n < \sqrt{c} + 1$, so erhalten wir auch für den folgenden Wert

$$x_{n+1} < \sqrt{c + \sqrt{c} + 1} < \sqrt{c + 2 \cdot \sqrt{c} + 1} = \sqrt{c} + 1.$$

Somit ist unsere Behauptung durch Induktion bewiesen.

Nach dem grundlegenden Satz aus Nr. 34 hat die Folge $\{x_n\}$ einen endlichen Grenzwert a. Um ihm zu bestimmen, gehen wir in der Beziehung

$$x_{n+1}^2 = c + x_n$$

zur Grenze über und finden, daß a der quadratischen Gleichung

$$a^2 = c + a, \quad \text{d. h.} \quad a^2 - a - c = 0$$

genügt. Die Lösungen dieser Gleichung enthalten Wurzeln verschiedenen Vorzeichens; da aber der uns interessierende Grenzwert a nicht negativ sein kann, ist er also gleich

$$\frac{\sqrt{4c+1}+1}{2}.$$

3. Wir nehmen einen beliebigen Wert x_0, $0 < x_0 < 1$, und definieren die Folge $\{x_n\}$ durch die Rekursionsformel

$$x_{n+1} = x_n(2 - x_n).$$

Wenn wir jetzt annehmen, daß $0 < x_n < 1$ ist (diese Bedingung war für $n = 0$ erfüllt), so erhalten wir

$$0 < x_n < x_{n+1} < 1.$$

Da nämlich $2 - x_n > 1$ ist, gilt $x_{n+1} > x_n$; nun ist aber $x_n(2 - x_n) = 1 - (1 - x_n)^2$, woraus $x_{n+1} < 1$ folgt. Somit ist durch Induktion bewiesen, daß die monoton wachsende Folge $\{x_n\}$ unterhalb der Zahl 1 bleibt. Daher hat sie einen endlichen Grenzwert $a \neq 0$. Gehen wir in der obigen Rekursionsformel zur Grenze über, so erhalten wir $a = 1$, also ist $\lim x_n = 1$.

Wir überlassen es dem Leser, selbst zu untersuchen, was geschieht, wenn x_0 außerhalb des Intervalls $(0, 1)$ liegt.

Bemerkung. Es sei c eine beliebige positive Zahl und $x_n = cy_n$. Dann hat die obige Rekursionsformel die Gestalt

$$y_{n+1} = y_n(2 - cy_n).$$

Wenn der Anfangswert y_0 der Bedingung $0 < y_0 < \frac{1}{c}$ genügt, so finden wir, daß die monoton wachsende Folge $\{y_n\}$ gegen $\frac{1}{c}$ strebt. Nach diesem Schema kann man auf Rechenmaschinen die zur Zahl c reziproke Zahl berechnen.

4. Es seien zwei positive Zahlen a und b gegeben $(a > b)$. Wir bilden ihr arithmetisches und ihr geometrisches Mittel:

$$a_1 = \frac{a+b}{2}, \quad b_1 = \sqrt{ab}.$$

Bekanntlich ist das arithmetische Mittel größer als das geometrische[1]); beide liegen zwischen den Ausgangswerten: $a > a_1 > b_1 > b$. Für die Zahlen a_1 und b_1 bilden wir wieder die beiden Mittel:

$$a_2 = \frac{a_1 + b_1}{2}, \quad b_2 = \sqrt{a_1 b_1},$$

wobei $a_1 > a_2 > b_2 > b_1$ gilt, usw. Sind die Zahlen a_n und b_n schon definiert, so ergeben sich

[1]) Dies folgt unmittelbar aus der Ungleichung

$$\frac{a+b}{2} - \sqrt{ab} = \frac{1}{2}\left(a - 2\sqrt{ab} + b\right) = \frac{\left(\sqrt{a} - \sqrt{b}\right)^2}{2} > 0 \quad (\text{für } a \neq b).$$

a_{n+1} und b_{n+1} durch

$$a_{n+1} = \frac{a_n + b_n}{2}, \qquad b_{n+1} = \sqrt{a_n b_n},$$

und es ist (wie oben)

$$a_n > a_{n+1} > b_{n+1} > b_n.$$

Somit erhalten wir zwei Folgen, $\{x_n\}$ und $\{b_n\}$, von denen die erste fällt und die zweite wächst (sie nähern sich einander). Nun ist

$$a > a_n > b_n > b,$$

so daß beide Folgen beschränkt sind; daher streben beide gegen einen endlichen Grenzwert:

$$\alpha = \lim a_n, \qquad \beta = \lim b_n.$$

Gehen wir in der Beziehung $a_{n+1} = \frac{a_n + b_n}{2}$ zur Grenze über, so erhalten wir $\alpha = \frac{\alpha + \beta}{2}$; hieraus folgt aber $\alpha = \beta$.

Somit streben beide Folgen, sowohl die der arithmetischen Mittel a_n als auch die der geometrischen Mittel b_n, gegen den gemeinsamen Grenzwert $\mu = \mu(a, b)$. Nach GAUSS[1]) nennt man ihn das *arithmetisch-geometrische Mittel* der Ausgangszahlen a und b. Der Ausdruck für die Zahl $\mu(a, b)$ ist jedoch auf Grund dieser Bemerkung vorläufig noch nicht zu bestimmen; das ist erst mit Hilfe eines sogenannten elliptischen Integrals möglich (vgl. Band II, Nr. 315).

5. Wir gehen wieder von zwei positiven Zahlen a und b ($a > b$) aus und bilden diesmal jeweils das arithmetische und das harmonische[2]) Mittel:

$$a_1 = \frac{a + b}{2}, \qquad b_1 = \frac{2ab}{a + b},$$

$$a_2 = \frac{a_1 + b_1}{2}, \qquad b_2 = \frac{2a_1 b_1}{a_1 + b_1},$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$a_{n+1} = \frac{a_n + b_n}{2}, \qquad b_{n+1} = \frac{2a_n b_n}{a_n + b_n},$$

Aus der uns schon bekannten Ungleichung $\frac{a + b}{2} > \sqrt{ab}$ (für $a \neq b$) erhalten wir

$$\left(\frac{a + b}{2}\right)^2 > ab$$

oder schließlich

$$\frac{a + b}{2} > \frac{2ab}{a + b},$$

so daß also das arithmetische Mittel größer als das harmonische Mittel ist. Die beiden Mittel liegen wieder zwischen den Ausgangszahlen. Wenn wir das auf a_n und b_n anwenden, finden wir $a_n > a_{n+1} > b_{n+1} > b_n$.

---

[1]) CARL FRIEDRICH GAUSS, 1777—1855, deutscher Mathematiker.

[2]) Eine Zahl c *heißt harmonisches Mittel* der beiden positiven Zahlen a und b, wenn die reziproke Zahl $\frac{1}{c}$ das arithmetische Mittel der reziproken Zahlen $\frac{1}{a}$ und $\frac{1}{b}$ ist:

$$\frac{1}{c} = \frac{1}{2}\left(\frac{1}{a} + \frac{1}{b}\right), \quad \text{also} \quad c = \frac{2ab}{a + b}.$$

Ganz analog wie im vorigen Beispiel überzeugen wir uns davon, daß die beiden Folgen $\{a_n\}$ und $\{b_n\}$ gegen einen gemeinsamen Grenzwert c streben, den man *arithmetisch-harmonisches Mittel* der Zahlen a und b nennen könnte.

Dieser Grenzwert c läßt sich jedoch in einfacher Weise durch a und b ausdrücken. Offenbar ist nämlich $a_1 b_1 = ab$; da analog auch $a_{n+1} b_{n+1} = a_n b_n$ ist, können wir schließen, daß für alle n

$$a_n b_n = ab$$

ist. Gehen wir hier zur Grenze über, so erhalten wir

$$c = \sqrt{ab},$$

d. h., das arithmetisch-harmonische Mittel zweier Zahlen ist gleich ihrem geometrischen Mittel.

6. Zum Schluß betrachten wir ein komplizierteres Beispiel. Wir gehen von einer beliebigen reellen Zahl c aus, setzen $x_1 = \frac{c}{2}$ und definieren die Werte x_n durch die Rekursionsformel

$$x_{n+1} = \frac{c}{2} + \frac{x_n^2}{2}. \tag{1}$$

Wir fragen jetzt nach dem Grenzwert dieser Folge unter zwei verschiedenen Voraussetzungen für c.

Wenn wir im voraus wüßten, daß ein endlicher Grenzwert

$$a = \lim x_n \tag{2}$$

existiert, so wäre er nicht schwer zu bestimmen. Man müßte nur in der Beziehung (1), die unsere Folge definiert, zur Grenze übergehen und erhielte

$$a = \frac{c}{2} + \frac{a^2}{2} \quad \text{oder} \quad a^2 - 2a + c = 0.$$

Aus dieser quadratischen Gleichung erhalten wir

$$a = 1 \mp \sqrt{1 - c}. \tag{3}$$

Hieraus ersieht man sofort, daß die Folge $\{x_n\}$ für $c > 1$ keinen endlichen Grenzwert haben kann.

a) Wir setzen zunächst $0 < c \leqq 1$ voraus. Dann ist $x_n > 0$. Wenn wir von (1) die analoge Beziehung $x_n = \frac{c}{2} + \frac{x_{n-1}^2}{2}$ subtrahieren, erhalten wir

$$x_{n+1} - x_n = \frac{x_n^2 - x_{n-1}^2}{2}. \tag{*}$$

Offenbar ist $x_2 > x_1 = \frac{c}{2}$, und aus der Gleichung (*) folgt für $x_n > x_{n-1}$ auch $x_{n+1} > x_n$. Somit haben wir durch Induktion bewiesen, daß die Folge $\{x_n\}$ monoton wachsend ist.

Analog beweisen wir, daß unsere Folge (nach oben) beschränkt ist:

$$x_n < 1.$$

Diese Ungleichung gilt offenbar für $n = 1$; wenn sie für irgendein n gilt, so muß sie nach (1) auch für $n + 1$ gelten. Somit existiert der Grenzwert (2) tatsächlich, und er kann durch (3) ausgedrückt werden, und zwar mit dem Minuszeichen vor der Wurzel, da er nicht größer als 1 sein kann.

b) Es sei jetzt $-3 \leqq c < 0$. Offenbar gilt für alle n

$$x_n \geqq \frac{c}{2}.$$

Wir zeigen jetzt, daß in diesem Fall $x_n < 0$ ist. Das ist für $n = 1$ richtig; nehmen wir an, diese

Behauptung sei für ein n erfüllt, so ist

$$|x_n| \leqq \frac{|c|}{2},$$

$$x_n^2 \leqq \frac{|c|^2}{4} < |c| \quad \left(\text{wegen } \frac{|c|}{4} < 1\right)$$

und x_{n+1} hat das Vorzeichen von $\frac{c}{2}$, d. h. wird negativ, was zu beweisen war.

In diesem Fall ist die Folge $\{x_n\}$ nicht monoton. Wenn wir jedoch für n in (1) $2k$ und $2k-2$ setzen und dann $2k+1$ und $2k-1$ und in beiden Fällen gliedweise subtrahieren, so erhalten wir

$$\left.\begin{aligned} x_{2k+1} - x_{2k-1} &= \frac{1}{2}(x_{2k}^2 - x_{2k-2}^2), \\ x_{2k+2} - x_{2k} &= \frac{1}{2}(x_{2k+1}^2 - x_{2k-1}^2). \end{aligned}\right\} \tag{4}$$

Hieraus kann man durch Induktion schließen, daß für alle k

$$x_{2k+1} > x_{2k-1} \quad \text{und} \quad x_{2k+2} < x_{2k}$$

ist. In der Tat ist $x_3 > x_1 = \frac{c}{2}$; dann ist aber $|x_3| < |x_1|$, $x_3^2 < x_1^2$, und nach der zweiten der Formeln (4) wird (für $k = 1$) $x_4 < x_2$. Folglich ist $|x_4| > |x_2|$, $x_4^2 > x_2^2$, und nach der ersten der Formeln (4) erhält man (für $k = 2$) $x_5 > x_3$, usw.

Somit sind in unserem Fall die Folgen $\{x_{2k-1}\}$ und $\{x_{2k}\}$ $(k = 1, 2, 3, \ldots)$ einzeln monoton; da sie innerhalb der endlichen Grenzen $\frac{c}{2}$ und 0 liegen, haben sie beide endliche Grenzwerte:

$$a' = \lim x_{2k-1}, \qquad a'' = \lim x_{2k}.$$

Wir müssen noch beweisen, daß $a' = a''$ ist. Um das zu zeigen, lassen wir in (1) den Index n gegen ∞ gehen, und zwar möge n zuerst gerade Werte, dann ungerade Werte annehmen. Wir erhalten in der Grenze die beiden Beziehungen

$$a' = \frac{c}{2} + \frac{a''^2}{2}, \quad a'' = \frac{c}{2} + \frac{a'^2}{2}. \tag{5}$$

Durch Subtraktion erhalten wir $a' - a'' = \frac{1}{2}(a''^2 - a'^2)$, also

$$(a' - a'')(a' + a'' + 2) = 0.$$

Wie wir sogleich feststellen werden, kann die zweite Klammer für $c > -3$ nicht verschwinden, so daß $a' = a''$ sein muß. Wenn wir nämlich $a'' = -a' - 2$ in die zweite Beziehung von (5) einsetzen, erhalten wir für a' die quadratische Gleichung

$$a'^2 + 2a' + (4 + c) = 0,$$

die für $c > -3$ keine reellen Wurzeln haben kann.

Schließlich verschwinden für $c = -3$ beide Klammern, weil dann sowohl $a' = -1$ als auch $a'' = -1$ ist.

Somit ist in allen Fällen $a' = a''$. Wenn man den gemeinsamen Wert dieser Grenzwerte mit a bezeichnet, so erhält man für a den Ausdruck (3), und zwar offenbar wieder mit dem Minuszeichen vor der Wurzel, weil der Grenzwert der negativen Folge $\{x_n\}$ nicht positiv sein kann.

Diese Beispiele geben zu folgender Bemerkung Anlaß. Der bewiesene Satz ist ein typischer „Existenzsatz“: Er liefert die Existenz des Grenzwertes, gibt aber kein Verfahren zu seiner Berechnung. Nichtsdestoweniger ist er sehr wichtig. Einerseits braucht man bei vielen theoretischen Fragen nur die Existenz des Grenzwertes; andererseits ist es in vielen Fällen wichtig, sich zunächst von der Existenz des Grenzwertes zu überzeugen, weil dadurch Wege zu seiner effektiven Berechnung erschlossen werden. So kann man in den Beispielen 1 bis 3, 5 und 6 durch Grenzübergang in gewissen Relationen den Grenzwert explizit bestimmen, wenn man weiß, daß er existiert.

In dieser Beziehung ist das Beispiel 6b) besonders lehrreich. Für $c < -3$ behält der Ausdruck (3) durchaus seinen Sinn; das bedeutet aber keineswegs, daß er auch weiterhin den Grenzwert der Folge $\{x_n\}$ liefert; dieser Grenzwert existiert hier im allgemeinen nicht. Beispielsweise durchläuft unsere Variable für $c = -4$ offenbar die Zahlenwerte

$$-2, 0, -2, 0, -2, 0, \ldots,$$

und diese Folge hat keinen Grenzwert.

Im Beispiel 4 haben wir für den Grenzwert keinen Ausdruck angegeben. Da wir aber wissen, daß er existiert, können wir ihn leicht mit beliebiger Genauigkeit berechnen; denn er liegt zwischen den Zahlen a_n und b_n, die von beiden Seiten gegen ihn streben.

In Nr. 36 werden wir ein weiteres wichtiges Beispiel für die Anwendung des Satzes über monotone Folgen behandeln.

36. Die Zahl e. Wir benutzen hier den Grenzübergang zur Definition einer uns bisher nicht bekannten Zahl.

Wir betrachten die diskrete Veränderliche

$$x_n = \left(1 + \frac{1}{n}\right)^n$$

und versuchen, auf sie den Satz aus Nr. 34 anzuwenden.

Da mit wachsendem Exponenten n die Basis der Potenz abnimmt, ist nicht unmittelbar erkennbar, daß die Folge monoton ist. Um uns davon zu überzeugen, entwickeln wir den Ausdruck für x_n nach der binomischen Formel:

$$\begin{aligned} x_n &= \left(1 + \frac{1}{n}\right)^n \\ &= 1 + n \cdot \frac{1}{n} + \frac{n(n-1)}{1 \cdot 2} \frac{1}{n^2} + \frac{n(n-1)(n-2)}{1 \cdot 2 \cdot 3} \frac{1}{n^3} + \cdots \\ &\quad + \frac{n(n-1)\cdots(n-k+1)}{1 \cdot 2 \cdots k} \frac{1}{n^k} + \cdots + \frac{n(n-1)\cdots(n-n+1)}{1 \cdot 2 \cdots n} \frac{1}{n^n} \\ &= 1 + 1 + \frac{1}{2!}\left(1 - \frac{1}{n}\right) + \frac{1}{3!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) + \cdots \\ &\quad + \frac{1}{k!}\left(1 - \frac{1}{n}\right)\cdots\left(1 - \frac{k-1}{n}\right) + \cdots + \frac{1}{n!}\left(1 - \frac{1}{n}\right)\cdots\left(1 - \frac{n-1}{n}\right). \end{aligned} \tag{6}$$

Wenn man jetzt von x_n zu x_{n+1} übergeht, also n um 1 vergrößert, so muß man zunächst ein neues $(n+2)$-tes (positives) Glied hinzufügen; jedes der angeschriebenen $n+1$ Glieder nimmt zu, weil jeder Faktor der Form $1 - \frac{s}{n}$ durch den größeren Faktor $1 - \frac{s}{n+1}$ ersetzt wird. Hieraus folgt

$$x_{n+1} > x_n,$$

d. h., die Folge $\{x_n\}$ ist wachsend.

Jetzt zeigen wir, daß sie nach oben beschränkt ist. Wenn wir in dem Ausdruck (6) alle Faktoren in Klammern, die ja kleiner als 1 sind, weglassen, so vergrößert sich der Ausdruck, und es ist

$$x_n < 2 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!} = y_n.$$

Wenn wir weiter jeden Faktor im Nenner der Brüche (von 3 an) durch die Zahl 2 ersetzen, so wird dieser Ausdruck nochmals vergrößert, so daß wir

$$y_n < 2 + \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^{n-1}}$$

erhalten. Die geometrische Reihe (mit dem Anfangsglied $\frac{1}{2}$) hat aber für jedes n eine Summe, die nicht größer als 1 ist; daher ist $y_n < 3$ und somit erst recht $x_n < 3$.

Hieraus folgt nach dem Satz auf S. 70 (Nr. 34), daß die Folge $\{x_n\}$ einen endlichen Grenzwert hat. Nach EULER[1]) wird er mit e bezeichnet. Diese Zahl

$$\mathrm{e} = \lim \left(1 + \frac{1}{n}\right)^n$$

ist sowohl für die Analysis selbst als auch für ihre Anwendungen von außerordentlicher Wichtigkeit. Wir notieren hier die ersten 15 Stellen ihrer Dezimalbruchentwicklung:

$$\mathrm{e} = 2{,}718281828459045\ldots$$

In Nr. 37 werden wir ein zweckmäßiges Verfahren zur angenäherten Berechnung der Zahl e angeben und gleichzeitig beweisen, daß e irrational ist.

Einige Eigenschaften der Zahl e, die wir später herleiten werden [Nr. 54, Formel (13)], lassen es als besonders vorteilhaft erscheinen, gerade diese Zahl als Basis eines Logarithmensystems zu wählen. Die Logarithmen zur Basis e werden *natürliche Logarithmen* genannt und mit ln (ohne weiteren Hinweis auf die Basis) bezeichnet. Bei theoretischen Untersuchungen werden ausschließlich natürliche Logarithmen benutzt.[2])

Wir erinnern daran, daß die gewöhnlichen dekadischen Logarithmen [Logarithmen zur Basis 10, auch *Briggssche Logarithmen* genannt[3])] mit den natürlichen Logarithmen durch die bekannte Formel

$$\log x = M \cdot \ln x$$

[1]) LEONHARD EULER, 1707—1783, Schweizer Mathematiker.

[2]) Diese Logarithmen werden manchmal auch *Napiersche* (oder *Nepersche*) *Logarithmen* genannt [nach dem schottischen Mathematiker JOHN NAPIER (NEPER), 1550—1617, einem der Entdecker der Logarithmen]. NAPIER selbst hatte den Begriff der Basis eines Logarithmensystems noch nicht (weil er sie nach einem anderen Prinzip konstruierte), aber seine Logarithmen beziehen sich auf eine Basis, die nahe bei $\frac{1}{\mathrm{e}}$ liegt. Die Logarithmen, mit denen sein Zeitgenosse, der Schweizer JOBST BÜRGI (1552—1632), rechnete, haben eine Basis, die sich nicht viel von e unterscheidet.

[3]) Nach HENRY BRIGGS, 1556—1630, englischer Mathematiker.

verknüpft sind, wo M der sogenannte *Modul*,

$$M = \log e = \frac{1}{\ln 10} = 0{,}434294\ldots,$$

ist. Das ergibt sich leicht, indem man die Identität

$$x = e^{\ln x}$$

zur Basis 10 logarithmiert.

37. Näherungsweise Berechnung der Zahl e. Wir wenden uns jetzt wieder der Beziehung (6) zu. Hält man k fest, so folgt für $n > k$, wenn man alle auf das $(k+1)$-te folgenden Glieder wegläßt, die Ungleichung

$$x_n > 2 + \frac{1}{2!}\left(1 - \frac{1}{n}\right) + \frac{1}{3!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) + \cdots + \frac{1}{k!}\left(1 - \frac{1}{n}\right)\cdots\left(1 - \frac{k-1}{n}\right).$$

Wir lassen hier n gegen ∞ streben und gehen zur Grenze über; da alle Klammern den Grenzwert 1 haben, ergibt sich

$$e \geqq 2 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{k!} = y_k.$$

Diese Ungleichung gilt für jedes natürliche k. Somit erhalten wir

$$x_n < y_n \leqq e;$$

hieraus folgt offenbar (nach Satz 3 aus Nr. 28) auch

$$\lim y_n = e.$$

Übrigens ist y_n die $(n+1)$*-te Partialsumme* der unendlichen Reihe (Nr. 25, Beispiel 9)

$$1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} + \cdots,$$

und die obige Limesrelation zeigt, daß e ihre Summe ist. Man sagt auch, daß e in diese Reihe *entwickelt* werden kann, und schreibt

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} + \cdots = \sum_{n=0}^{\infty} \frac{1}{n!}.$$

Die Folge $\{y_n\}$ ist zur angenäherten Berechnung der Zahl e geeigneter als $\{x_n\}$. Wir wollen einmal die Güte der Näherung von $\{y_n\}$ an e abschätzen. Zu diesem Zweck betrachten wir zuerst die Differenz zwischen einem beliebigen Wert y_{n+m} $(m = 1, 2, 3, \ldots)$, der auf y_n folgt, und y_n selbst,

$$\begin{aligned} y_{n+m} - y_n &= \frac{1}{(n+1)!} + \frac{1}{(n+2)!} + \cdots + \frac{1}{(n+m)!} \\ &= \frac{1}{(n+1)!}\left\{1 + \frac{1}{n+2} + \frac{1}{(n+2)(n+3)} + \cdots \right. \\ &\qquad \left. + \frac{1}{(n+2)(n+3)\cdots(n+m)}\right\}. \end{aligned}$$

Ersetzt man in der geschweiften Klammer alle Faktoren im Nenner der Brüche durch $n + 2$, so erhält man die Ungleichung

$$y_{n+m} - y_n < \frac{1}{(n+1)!}\left\{1 + \frac{1}{n+2} + \frac{1}{(n+2)!} + \cdots + \frac{1}{(n+2)^{m-1}}\right\};$$

sie bleibt sicher richtig, wenn man die Summe in der Klammer bis ins Unendliche erstreckt, und die Summe dieser unendlichen geometrischen Reihe läßt sich nach einer bekannten Formel ausdrücken; so erhalten wir

$$y_{n+m} - y_n < \frac{1}{(n+1)!} \cdot \frac{n+2}{n+1}.$$

Nun halten wir n fest und lassen m unbegrenzt wachsen; die diskrete Veränderliche y_{n+m} (nur der Index m „läuft") durchläuft die Folge der Werte

$$y_{n+1}, y_{n+2}, y_{n+3}, \ldots, y_{n+m}, \ldots,$$

die offenbar gegen e konvergiert. Daher erhalten wir in der Grenze

$$\mathrm{e} - y_n \leqq \frac{1}{(n+1)!} \cdot \frac{n+2}{n+1}$$

oder schließlich

$$0 < \mathrm{e} - y_n < \frac{1}{n!n}.^{1)}$$

Bezeichnet θ das Verhältnis der Differenz $\mathrm{e} - y_n$ zu der Zahl $\frac{1}{n!n}$ (es liegt offenbar zwischen 0 und 1), so können wir auch

$$\mathrm{e} - y_n = \frac{\theta}{n!n}$$

schreiben. Setzt man hier für y_n die obige Reihenentwicklung ein, so erhält man die wichtige Formel

$$\mathrm{e} = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!} + \frac{\theta}{n!n}, \tag{7}$$

die der Ausgangspunkt zur Berechnung von e ist. Lassen wir das letzte Glied, das sogenannte *Restglied*, fort und ersetzen jedes der übrigen Glieder durch einen Näherungs-Dezimalbruch, so erhalten wir gerade einen Näherungswert für e.

Wir stellen uns jetzt die Aufgabe, e mit Hilfe der Formel (7) zu berechnen, etwa auf $\frac{1}{10^7}$ genau. Zuerst müssen wir feststellen, wie groß die Zahl n zu wählen ist, damit diese Genauigkeit erreicht wird.

Wenn wir die reziproken Fakultäten berechnen (vgl. nachstehende Tabelle), so sehen wir, daß sich für $n = 10$ für das Restglied in Formel (7) schon

$$\frac{\theta}{n!n} = \frac{\theta}{10! \cdot 10} < 0{,}00000003$$

[1]) Es ist nämlich, wie man leicht sieht, $\frac{n+2}{(n+1)^2} < \frac{1}{n}$.

ergibt, so daß der Fehler, den wir machen, wenn wir es weglassen, bedeutend kleiner als die vorgeschriebene Genauigkeitsgrenze ist. Wir nehmen also $n = 10$. Jedes Glied verwandeln wir in einen Dezimalbruch, wobei wir die achte Stelle so runden, daß der absolute Betrag des Fehlers kleiner wird als die Hälfte der Einheit der achten Stelle, d. h. kleiner als $\frac{1}{2 \cdot 10^8}$. Wir geben die Resultate der Rechnung in einer Tabelle an:

$$\begin{array}{rl}
 & 2{,}00000000 \\
\frac{1}{2!} = & 0{,}50000000 \\
\frac{1}{3!} = & 0{,}16666667(-) \\
\frac{1}{4!} = & 0{,}04166667(-) \\
\frac{1}{5!} = & 0{,}00833333(+) \\
\frac{1}{6!} = & 0{,}00138889(-) \\
\frac{1}{7!} = & 0{,}00019841(+) \\
\frac{1}{8!} = & 0{,}00002480(+) \\
\frac{1}{9!} = & 0{,}00000276(-) \\
\frac{1}{10!} = & 0{,}00000028(-) \\
\hline
 & 2{,}71828181
\end{array}$$

Hinter die Näherungswerte wurden ein Plus- bzw. ein Minuszeichen in Klammern gesetzt, je nachdem, ob der Näherungswert kleiner oder größer ist als der genaue Wert. Wie wir sehen können, ist der Fehler, der durch Weglassen des Restgliedes entstanden ist, kleiner als $\frac{3}{10^8}$. Wenn wir jetzt noch die Rundungsfehler (mit Vorzeichen) berücksichtigen wollen, müssen wir beachten, daß wir fünfmal einen um weniger als $\frac{1}{10^8}$ zu großen Wert und dreimal einen um weniger als $\frac{1}{10^8}$ zu kleinen Wert angesetzt haben. Die Summe der Rundungsfehler liegt also zwischen

$$-\frac{3}{10^8} \quad \text{und} \quad +\frac{5}{10^8}.$$

Hieraus folgt, daß die Zahl e selbst zwischen den Zahlen

$$2{,}71828178 \quad \text{und} \quad 2{,}71828186$$

liegen muß, so daß sich

$$e = 2{,}7182818 \pm 0{,}0000001$$

ergibt.

Wir benutzen jetzt die Formel (7), um zu zeigen, daß *die Zahl* e *irrational ist.*

Wir beweisen diese Irrationalität indirekt, nehmen also an, es sei $e = \frac{m}{n}$; mit diesem n schreiben wir Formel (7) auf:

$$\frac{m}{n} = 1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} + \frac{\theta}{n!n} \quad (0 < \theta < 1).$$

Beide Seiten dieser Beziehung multiplizieren wir mit $n!$; dann können wir durch die Nenner aller Brüche (mit Ausnahme des letzten) kürzen. Links erhalten wir eine ganze Zahl, rechts aber die Summe einer ganzen Zahl und einer Zahl zwischen 0 und 1. Mit diesem Widerspruch ist die Irrationalität von e bewiesen.

38. Ein Lemma über Intervallschachtelungen. Zum Schluß dieses Paragraphen über monotone Folgen befassen wir uns mit zwei solcher Folgen, die „einander entgegenkommen".

Gegeben seien eine monoton wachsende Folge $\{x_n\}$ *und eine monoton fallende Folge* $\{y_n\}$, *und es sei stets*

$$x_n < y_n. \tag{8}$$

Wenn die Differenz $y_n - x_n$ *gegen* 0 *strebt, so haben beide Folgen einen gemeinsamen endlichen Grenzwert:*

$$c = \lim x_n = \lim y_n.$$

In der Tat, für alle n gilt $y_n \leqq y_1$; das bedeutet aber nach (8) auch $x_n < y_1$ ($n = 1, 2, 3, \ldots$). Die wachsende Folge $\{x_n\}$ ist nach oben beschränkt, hat also einen endlichen Grenzwert

$$c = \lim x_n.$$

Analog ist wegen

$$y_n > x_n \geqq x_1$$

die fallende Folge $\{y_n\}$ nach unten beschränkt, so daß auch sie gegen einen endlichen Grenzwert strebt:

$$c' = \lim y_n.$$

Nach Satz 1 aus Nr. 30 gilt für die Differenz beider Grenzwerte:

$$c' - c = \lim (y_n - x_n);$$

dieser Ausdruck ist aber nach Voraussetzung gleich 0, so daß $c' = c$ ist.

Diese Aussage läßt sich auch anders formulieren:

Wir nennen die Menge aller Zahlen (oder, wie man auch sagen kann, aller „Punkte") x, die der Ungleichung

$$a \leqq x \leqq b \quad (\text{für } a < b)$$

genügen, ein *abgeschlossenes Intervall* oder auch ein *Segment*; man schreibt dafür $[a, b]$. Gelegentlich findet sich in der Literatur statt der eckigen Klammer eine spitze, $\langle a, b\rangle$, oder auch eine stilisierte, $[a, b]$. Die Zahlen („Punkte") a und b heißen *linker* bzw. *rechter Endpunkt* des Intervalls, ihre Differenz $b - a$ heißt die *Länge* des Intervalls. Man sieht leicht, daß einem *Intervall* auf der Zahlengeraden eine *Strecke* (derselben Länge) entspricht.

Man sagt, das Intervall $[a', b']$ sei im Intervall $[a, b]$ *enthalten* oder *eingebettet* (*eingeschachtelt*), wenn alle Punkte des ersten Intervalls auch zum zweiten gehören oder, was dasselbe ist, wenn

$$a \leqq a' < b' \leqq b$$

ist. Die geometrische Bedeutung ist klar.

Es sei nun eine unendliche Folge ineinander enthaltener Intervalle (eine *Intervallschachtelung*)

$$[a_1, b_1], [a_2, b_2], \ldots, [a_n, b_n], \ldots$$

gegeben, so daß also jedes folgende im vorhergehenden enthalten ist, und die Längen dieser Intervalle mögen mit wachsendem n gegen 0 streben:

$$\lim (b_n - a_n) = 0.$$

Dann streben die Endpunkte a_n und b_n der Intervalle (von verschiedenen Seiten her) gegen den gemeinsamen Grenzwert

$$c = \lim a_n = \lim b_n,$$

und dies ist der einzige Punkt, der in allen Intervallen enthalten ist.

Diese Tatsache ist nur eine andere Formulierung des oben bewiesenen Satzes. Nach Voraussetzung ist

$$a_n \leqq a_{n+1} < b_{n+1} \leqq b_n,$$

so daß der linke Endpunkt a_n bzw. der rechte Endpunkt b_n des n-ten Intervalls hier die Rolle der monotonen diskreten Veränderlichen x_n bzw. y_n spielen.

Da a_n wachsend und b_n fallend gegen c strebt, ist

$$a_n \leqq c \leqq b_n \quad (n = 1, 2, 3, \ldots),$$

d. h., der Punkt c gehört tatsächlich zu allen Intervallen. Es kann kein von c verschiedener Punkt c' mit der gleichen Eigenschaft existieren, denn dann wäre

$$b_n - a_n \geqq |c' - c| > 0;$$

die Länge des n-ten Intervalls würde also nicht gegen 0 streben.

Später werden wir häufig auf diesen Intervallschachtelungssatz zurückgreifen.

§ 4. Das Konvergenzprinzip. Teilfolgen. Partielle Grenzwerte

39. Das Konvergenzprinzip. Es sei eine diskrete Veränderliche x_n gegeben, die die Folge der Werte

$$x_1, x_2, \ldots, x_n, \ldots, x_{n'}, \ldots \tag{1}$$

durchläuft. Wir untersuchen die Frage, ob es ein Kriterium für die Existenz eines endlichen Grenzwertes dieser Folge gibt. Die Definition des Grenzwertes selbst kann das nicht leisten, weil darin ja schon der Grenzwert vorkommt, um dessen Existenz es geht. Wir können in dem Kriterium nur das benutzen, was uns gegeben ist, nämlich die Wertefolge (1) der diskreten Veränderlichen.

Dieses Problem wird durch folgenden bemerkenswerten Satz gelöst, der von dem tschechischen Mathematiker B. BOLZANO (1781—1848) und dem französischen Mathematiker A. L. CAUCHY (1789—1857) stammt und der als allgemeines *Konvergenzprinzip* bezeichnet wird.

Satz. *Eine diskrete Veränderliche x_n hat genau dann einen endlichen Grenzwert, wenn zu jeder Zahl $\varepsilon > 0$ eine Zahl N existiert derart, daß die Ungleichung*

$$|x_n - x_{n'}| < \varepsilon \tag{2}$$

gilt, sobald $n > N$ und $n' > N$ ist.

Wie man sieht, besteht das Wesen der Sache darin, daß der Abstand zwischen den Werten der Veränderlichen mit wachsenden Indizes immer geringer und zuletzt beliebig klein wird. Solche Folgen werden auch *in sich konvergent* genannt, und das Konvergenzprinzip läßt sich dann folgendermaßen formulieren: *Eine in sich konvergente Folge hat einen endlichen Grenzwert* („konvergiert"), *und umgekehrt.*

Beweis. Zunächst zeigen wir, daß die Bedingung notwendig ist. Die Folge $\{x_n\}$ habe einen endlichen Grenzwert a. Nach *Definition* des Grenzwertes (Nr. 23) läßt sich dann zu jeder beliebig gewählten Zahl $\varepsilon > 0$ eine Zahl N finden derart, daß für alle $n > N$ die Ungleichung

$$|x_n - a| < \frac{\varepsilon}{2}$$

erfüllt ist.

Für zwei Zahlen $n > N$ und $n' > N$ gilt also gleichzeitig

$$|x_n - a| < \frac{\varepsilon}{2}$$

und

$$|a - x_{n'}| < \frac{\varepsilon}{2};$$

hieraus folgt

$$\begin{aligned}|x_n - x_{n'}| &= |(x_n - a) + (a - x_{n'})| \\ &\leqq |x_n - a| + |a - x_{n'}| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.\end{aligned}$$

Damit ist die Notwendigkeit der Bedingung bewiesen.

Der Beweis dafür, daß sie hinreicht, ist jedoch nicht ganz so einfach. Die Voraussetzung des Satzes sei also erfüllt; es muß nachgewiesen werden, daß dann für die Folge $\{x_n\}$ ein vollbestimmter endlicher Grenzwert existiert. Dazu konstruieren wir im Bereich aller reellen Zahlen einen Schnitt nach folgender Vorschrift. In der Unterklasse A nehmen wir jede reelle Zahl α, für die von einem gewissen Index an die Ungleichung

$$x_n > \alpha$$

gilt. In die Oberklasse A' nehmen wir dann alle übrigen (nicht in A liegenden) reellen Zahlen α'. Zunächst zeigen wir, daß diese Klassen nicht leer sind. Wenn eine beliebige Zahl $\varepsilon > 0$ vorgegeben ist, so können wir auf Grund der Voraussetzung eine ihr entsprechende Zahl N derart finden, daß für $n > N$ und $n' > N$ die Beziehung (2) erfüllt ist und somit

$$x_{n'} - \varepsilon < x_n < x_{n'} + \varepsilon \tag{3}$$

gilt. Jetzt sehen wir, daß für $n' > N$ insbesondere jede Zahl $x_{n'} - \varepsilon$ zur Klasse A gehört; denn sie wird für hinreichend großes n (nämlich für $n > N$) von x_n übertroffen. Da andererseits (für dieselben n) die Zahl x_n kleiner bleibt als jede Zahl der Form $x_{n'} + \varepsilon$ (für $n' > N$), kann offenbar keine solche Zahl zu A gehören; folglich liegt sie in der Klasse A'.

Die Vorschrift, nach der die Klassen A und A' definiert sind, ist so formuliert, daß folgendes unmittelbar klar ist: Jede reelle Zahl liegt in einer und nur in einer dieser Klassen. Außerdem ist offenbar jede Zahl α (aus A) kleiner als jede Zahl α' (aus A'); denn wäre $\alpha > \alpha'$, so würde die Veränderliche x_n von einem bestimmten n an auch die Zahl α' übertreffen, entgegen der Definition der Zahlen α'. Somit ist die oben definierte Einteilung des Bereichs der reellen Zahlen in die beiden Klassen tatsächlich ein Schnitt.

Nach dem Dedekindschen Hauptsatz (Nr. 10) existiert also eine reelle Zahl a, die zwischen den Zahlen der beiden Klassen liegt[1]):

$$\alpha \leqq a \leqq \alpha'.$$

Wie wir schon bemerkten, ist für jedes $n' > N$ die Zahl $x_{n'} - \varepsilon$ eine der Zahlen α und die Zahl $x_{n'} + \varepsilon$ eine der Zahlen α'. Daher ist insbesondere

$$x_{n'} - \varepsilon \leqq a \leqq x_{n'} + \varepsilon \quad \text{oder} \quad |a - x_{n'}| = |x_{n'} - a| \leqq \varepsilon$$

für jedes $n' > N$. Nach Definition des Grenzwertes (Nr. 23) bedeutet das aber

$$a = \lim x_n,$$

und damit ist der Satz bewiesen.

Dieses Kriterium werden wir im folgenden vielfach anwenden.

40. Teilfolgen und ihre Grenzwerte. Wir betrachten jetzt neben der Folge (1) irgendeine ihrer *Teilfolgen* (oder *Unterfolgen*):

$$x_{n_1}, x_{n_2}, x_{n_3}, \ldots, x_{n_k}, \ldots, \tag{4}$$

wobei $\{n_k\}$ eine Folge wachsender natürlicher Zahlen

$$n_1 < n_2 < n_3 < \cdots < n_k < n_{k+1} < \cdots \tag{5}$$

ist. Hier ist der Index, der alle natürlichen Werte annimmt, nicht n, sondern k; die n_k bilden selbst eine Folge, die aus natürlichen Zahlen besteht und offenbar mit wachsendem k gegen ∞ strebt.

Satz. *Wenn die Folge* (1) *einen bestimmten* (*endlichen oder unendlichen*) *Grenzwert* a *hat, so hat auch jede Teilfolge* (4) *denselben Grenzwert.*

Nehmen wir etwa beim Beweis den Fall eines endlichen a. Es möge sich also zu $\varepsilon > 0$ ein N finden lassen derart, daß für $n > N$ die Ungleichung

$$|x_n - a| < \varepsilon$$

[1]) Dort war sie mit β bezeichnet worden.

erfüllt ist. Wegen $n_k \to \infty$ existiert auch eine Zahl K derart, daß für alle $k > K$ die Beziehung $n_k > N$ gilt. Dann gilt für dieselben Werte von k die Ungleichung

$$|x_{n_k} - a| < \varepsilon,$$

und damit ist unsere Behauptung bewiesen.

Übrigens haben wir für diese Überlegung die Ungleichung (5) nicht gebraucht, d. h. die Monotonie der Folge $\{n_k\}$ nicht benutzt. Somit gilt unsere Behauptung auch dann, wenn die ganzzahlige Veränderliche n_k nach irgendeiner anderen Vorschrift (also nicht notwendig monoton) gegen ∞ strebt.

Wenn die Veränderliche x_n oder, was dasselbe ist, die Folge (1) keinen Grenzwert hat, so kann durchaus irgendeine Teilfolge (4), d. h. eine entsprechende Folge der $x'_k = x_{n_k}$, einen Grenzwert haben. Diesen Grenzwert nennen wir einen *partiellen Grenzwert* der Folge $\{x_n\}$ oder der Folge (1).

Es sei etwa $x_n = (-1)^{n+1}$; diese Folge hat keinen Grenzwert. Durchläuft jedoch n nur die ungeraden oder nur die geraden Zahlen, so haben die *Teilfolgen*

$$x_1 = 1,\ x_3 = 1, \ldots, x_{2k-1} = 1, \ldots$$

und

$$x_2 = -1,\ x_4 = -1, \ldots, x_{2k} = -1, \ldots$$

je einen Grenzwert, nämlich 1 bzw. -1. Diese Zahlen sind dann *partielle Grenzwerte* der Folge $\{x_n\}$. Analog folgt auch, daß die Folge der $x_n = (-1)^{n+1}\, n$ die partiellen Grenzwerte ∞ und $-\infty$ hat; die Folge der $x_n = n^{(-1)^{n+1}}$ hat die partiellen Grenzwerte ∞ und 0.

Man kann leicht Beispiele von Folgen konstruieren, die unendlich viele verschiedene partielle Grenzwerte besitzen. Wir definieren etwa die Veränderliche x_n durch folgende Vorschrift: Ist $\alpha\beta\ldots\nu$ (wo $\alpha, \beta, \ldots, \nu$ Ziffern sind) die Darstellung der Zahl n im Dezimalsystem, so sei

$$x_n = 0{,}\alpha\beta\ldots\nu,$$

also beispielsweise $x_{13} = 0{,}13$; $x_{4035} = 0{,}4035$ usw. Dabei kommt jeder endliche Dezimalbruch zwischen 0,1 und 1 in unsere Folge unendlich oft vor, beispielsweise 0,217 an der 217-ten, aber auch an der 2170-ten, der 21700-ten usw. Stelle. Mit anderen Worten, jeder *endliche* Dezimalbruch zwischen 0,1 und 1 ist ein partieller Grenzwert unserer Folge. Betrachtet man aber jetzt eine beliebige andere reelle Zahl α innerhalb dieser Grenzen, so braucht man sie nur als unendlichen Dezimalbruch darzustellen (Nr. 9),

$$\alpha = 0{,}c_1c_2\ldots c_k\ldots \qquad (c_1 \geqq 1),$$

damit klar wird, daß die Teilfolge

$$\begin{aligned} &x_{c_1} = 0{,}c_1,\\ &x_{c_1c_2} = 0{,}c_1c_2,\\ &\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\\ &x_{c_1c_2\cdots c_k} = 0{,}c_1c_2\ldots c_k\\ &\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \end{aligned}$$

diese Zahl α als Grenzwert hat. Somit füllen in unserem Fall die partiellen Grenzwerte das ganze Intervall $[0{,}1;\ 1]$ aus.

Existiert nun für eine Folge $\{x_n\}$ immer ein partieller Grenzwert? Diese Frage ist leicht zu bejahen, wenn die Menge $\{x_n\}$ unbeschränkt ist. Die Menge sei beispielsweise nach oben nicht beschränkt; dann existiert zu jedem natürlichen k in der Folge (1) ein Glied x_{n_k}, das größer als k ist:

$$x_{n_k} > k \quad (k = 1, 2, 3, \ldots)$$

(dabei kann man es leicht so einrichten, daß die Zahl n_k mit k wächst). Die Teilfolge

$$x_{n_1}, x_{n_2}, x_{n_3}, \ldots, x_{n_k}, \ldots$$

hat offenbar den Grenzwert ∞; er ist also ein partieller Grenzwert der Folge.

Die obige Frage ist aber auch im Fall beschränkter Folgen zu bejahen; dies erfordert jedoch weitere Überlegungen, die wir in Nr. 41 anstellen werden.

41. Satz von Bolzano[1]-Weierstraß[2]. *Aus jeder beschränkten Folge* (1) *kann man eine gegen einen endlichen Grenzwert konvergierende Teilfolge* (4) *auswählen.*

(Diese Formulierung schließt nicht aus, daß auch gleiche Zahlen in der gegebenen Folge vorkommen, was für die Anwendungen bequem ist.)

Beweis. Alle Zahlen x_n mögen zwischen den Grenzen a und b liegen. Wir halbieren dieses Intervall $[a, b]$; dann liegen mindestens in einer Hälfte unendlich viele Elemente der gegebenen Folge; denn sonst wären auch im ganzen Intervall $[a, b]$ nur endlich viele Elemente enthalten. Es sei jetzt $[a_1, b_1]$ diejenige Hälfte, die unendlich viele Zahlen x_n enthält (wenn beide Hälften diese Eigenschaft haben, so sei $[a_1, b_1]$ eine von ihnen).

Analog halbieren wir das Intervall $[a_1, b_1]$; es sei $[a_2, b_2]$ wieder eine Hälfte, die unendlich viele Zahlen x_n enthält, usw. Wir denken diesen Prozeß unbeschränkt fortgesetzt; nach k Schritten erhalten wir ein Intervall $[a_k, b_k]$, das unendlich viele Zahlen x_n enthält.

Jedes dieser so konstruierten Intervalle (vom zweiten an) ist im vorhergehenden enthalten, da es eine Hälfte davon ist. Außerdem ist die Länge des k-ten Intervalls gleich

$$b_k - a_k = \frac{b - a}{2^k},$$

strebt also mit wachsendem k gegen 0. Wenn wir jetzt den Satz aus Nr. 38 über Intervallschachtelungen anwenden, können wir folgern, daß $\{a_k\}$ und $\{b_k\}$ gegen einen gemeinsamen Grenzwert c streben.

Wir konstruieren jetzt folgendermaßen induktiv eine Teilfolge $\{x_{n_k}\}$. Als x_{n_1} nehmen wir ein beliebiges (z. B. das erste) der Elemente x_n unserer Folge, das in $[a_1, b_1]$ liegt. Als x_{n_2} nehmen wir ein beliebiges (z. B. das erste) auf x_{n_1} folgendes und in $[a_2, b_2]$ liegendes der Elemente x_n, usw. Allgemein nehmen wir als x_{n_k} ein beliebiges (z. B. das erste) der Elemente x_n, das auf alle vorhergehenden $x_{n_1}, x_{n_2}, \ldots, x_{n_{k-1}}$ folgt und in $[a_k, b_k]$ enthalten ist. Diese Wahl ist möglich, weil jedes der Intervalle $[a_k, b_k]$ unendlich viele Zahlen x_n enthält, d. h. Elemente x_n mit beliebig großem Index.

Da ferner

$$a_k \leqq x_{n_k} \leqq b_k$$

und

$$\lim a_k = \lim b_k = c$$

ist, gilt nach Satz 3 aus Nr. 28 auch

$$\lim x_{n_k} = c,$$

was zu beweisen war.

---

[1] Bernard Bolzano, 1781—1848, tschechischer Mathematiker und Philosoph.

[2] Karl Weierstrass, 1815—1897, deutscher Mathematiker.

Die beim Beweis dieses Satzes verwendete Methode, nach der die betrachteten Intervalle sukzessive halbiert werden, ist unter der Bezeichnung *Bolzanosche Methode*[1]) bekannt; sie wird uns in anderen Fällen auch noch oft begegnen.

Der Satz von BOLZANO-WEIERSTRASS erleichtert den Beweis vieler schwieriger Sätze, da mit seinem Beweis die Hauptschwierigkeit der entsprechenden Überlegungen sozusagen vorweggenommen ist. Als Beispiel beweisen wir mit seiner Hilfe nochmals das *Konvergenzprinzip,* und zwar die Tatsache, daß es hinreichend ist, was wir in Nr. 39 nur mit ziemlichem Aufwand nachweisen konnten.

Es sei also die Voraussetzung erfüllt; zu gegebenem $\varepsilon > 0$ existiere eine Zahl N derart, daß für $n > N$ und $n' > N$ die Ungleichungen (2) oder (3) gelten. Hält man dabei n' fest, so ist nach (3) klar, daß die Folge $\{x_n\}$ in jedem Fall beschränkt ist: Ihre Glieder liegen für $n > N$ zwischen den Zahlen $x_{n'} - \varepsilon$ und $x_{n'} + \varepsilon$, und es ist nicht schwer, diese Grenzen so auseinanderzuziehen, daß auch die ersten N Werte $x_1, x_2, \ldots, x_N$ dazwischenliegen. Nun können wir nach dem Satz von BOLZANO-WEIERSTRASS eine Teilfolge $\{x_k\}$ auswählen, die gegen den endlichen Grenzwert c konvergiert:

$$\lim x_{n_k} = c.$$

Wir zeigen, daß auch die Folge $\{x_n\}$ gegen diesen Grenzwert strebt. Wir können k so groß wählen, daß $|x_{n_k} - c| < \varepsilon$ und gleichzeitig $n_k > N$ ist. Folglich können wir in (2) $n' = n_k$ setzen,

$$|x_n - x_{n_k}| < \varepsilon,$$

und wenn wir beide Ungleichungen zusammenfassen, erhalten wir schließlich

$$|x_n - c| < 2\varepsilon \quad (\text{für } n > N);$$

damit ist unsere Behauptung bewiesen.[2])

42. Der größte und der kleinste partielle Grenzwert. Für jede Folge $\{x_n\}$, ob sie beschränkt oder nicht beschränkt ist, existieren Grenzwerte von Teilfolgen. Wir zeigen jetzt, daß unter diesen Grenzwerten von Teilfolgen notwendigerweise ein größter und ein kleinster existieren. Sie werden *oberer* bzw. *unterer Limes* der Folge $\{x_n\}$ genannt (auch limes superior bzw. limes inferior) und mit

$$\overline{\lim}\, x_n \text{ bzw. } \underline{\lim}\, x_n \quad \text{oder} \quad \limsup x_n \text{ bzw. } \liminf x_n$$

bezeichnet.

Satz. *Jede Folge $\{x_n\}$ besitzt einen größten und einen kleinsten partiellen Grenzwert. Notwendig und hinreichend für die Existenz des Grenzwertes einer Folge (im gewöhnlichen Sinne) ist die Übereinstimmung von limes superior und limes inferior.*[3])

Beweis. Wir untersuchen zuerst die Frage nach der Existenz des limes superior. Wir sahen schon in Nr. 40: Ist die Folge $\{x_n\}$ nicht nach oben beschränkt, so kann man aus (1) eine Teilfolge $\{x_{n_k}\}$ auswählen derart, daß

$$\lim x_{n_k} = \infty$$

[1]) Scherzhaft nennt man sie auch Löwenfangmethode. — *Anm. d. Red.*

[2]) Die Zahl 2ε ist von derselben Ordnung beliebig klein wie ε selbst. Wenn man will, kann man statt des ursprünglichen ε die Zahl $\frac{\varepsilon}{2}$ nehmen; dann würden wir hier ε erhalten haben. Von jetzt an werden wir auf solche Trivialitäten nicht mehr eingehen.

[3]) Dieser Satz, zu dessen Beweis der Satz von BOLZANO-WEIERSTRASS nicht benutzt wird, umfaßt diesen.

ist. Somit ist in diesem Fall ∞ ein partieller Grenzwert und offenbar auch der größtmögliche, so daß also $\overline{\lim}\, x_n = \infty$ ist. Wir setzen jetzt voraus, die Folge $\{x_n\}$ sei nach oben beschränkt: $x_n \leqq M$ $(n = 1, 2, 3, \ldots)$. Wir betrachten die obere Grenze der Werte x_n für $n > k$:

$$M_k = \sup_{n>k} \{x_n\} = \sup \{x_{k+1}, x_{k+2}, \ldots\} \leqq M.$$

Für wachsendes k kann der Wert M_k sicher nicht wachsen; folglich existiert nach dem Satz über monotone Folgen (Nr. 34) in jedem Fall der Grenzwert (für unbegrenzt wachsendes k)

$$\lim M_k,$$

der endlich oder gleich $-\infty$ sein kann.

Der Fall, daß er gleich $-\infty$ ist, ist einfach zu erledigen. Zu jedem $E > 0$ existiert ein Index $k = N$ derart, daß $M_N < -E$ ist, für $n > N$ ist aber offenbar $x_n \leqq M_N$, so daß für diesen Wert von n um so mehr

$$x_n < -E$$

ist. Das bedeutet aber, daß der Grenzwert (im gewöhnlichen Sinne)

$$\lim x_n = -\infty$$

existiert, der gleichzeitig größter und kleinster partieller Grenzwert ist.[1])

Wir betrachten jetzt den wichtigeren Fall, daß ein endlicher Grenzwert existiert,

$$\lim M_k = M^*,$$

und beweisen, daß diese Zahl M^* gerade gleich dem gesuchten limes superior von $\{x_n\}$ ist.

Zu diesem Zweck geben wir zwei charakteristische Eigenschaften der Zahl M^* an:

Ist die Zahl $\varepsilon > 0$ beliebig gewählt, so kann man eine Zahl $k = N'$ finden derart, daß $M_{N'} < M^* + \varepsilon$ ist; denn für $n > N'$ ist $x_n \leqq M_{N'}$, also erst recht $x_n < M^* + \varepsilon$. Somit gilt die erste Eigenschaft der Zahl M^*:

1. *Zu jedem $\varepsilon > 0$ existiert eine Zahl N' derart, daß für alle $n > N'$*

$$x_n < M^* + \varepsilon$$

gilt.

Andererseits gilt für jedes $\varepsilon > 0$ und jedes k die Beziehung $M_k \geqq M^* > M^* - \varepsilon$. Dann kann man aber nach der Eigenschaft der oberen Grenze (Nr. 11) unter den Zahlen x_n mit $n = k + 1, k + 2, k + 3, \ldots$ einen Wert $x_{n'}$ finden, für den auch $x_{n'} > M^* - \varepsilon$ ist.

Ersetzen wir die beliebig gewählte Zahl k durch N, so erhalten wir die zweite Eigenschaft.

2. *Zu jedem $\varepsilon > 0$ und jedem Index N kann man einen Wert $x_{n'}$ mit $n' > N$ finden derart, daß*

$$x_{n'} > M^* - \varepsilon$$

ist.

(Wir heben den Unterschied in den Formulierungen dieser beiden Eigenschaften hervor. Im ersten Fall ist die Ungleichung ausnahmslos für alle Werte x_n von einem bestimmten Wert des Index an erfüllt. Im zweiten Fall wird die Ungleichung nur von einzelnen Werten x_n erfüllt, unter denen es jedoch Werte mit beliebig großem Index gibt.)

Mit Hilfe dieser Eigenschaften beweisen wir zuerst, daß die Zahl M^* ein partieller Grenzwert für die Folge $\{x_n\}$ ist. Dafür muß eine Teilfolge $\{x_{n_i}\}$, die gegen M^* konvergiert, ausgewählt werden.

Wir wählen eine Folge positiver Zahlen ε_i, die gegen 0 strebt. Wir setzen $n_1 = 1$ und nehmen an, die Indizes

$$n_1 = 1 < n_2 < n_3 < \cdots < n_{i-1}$$

seien schon gewählt, und zeigen nun, wie n_i zu wählen ist. Nach der ersten Eigenschaft finden wir zu $\varepsilon = \varepsilon_i$ einen Index $N' = N_i$ derart, daß für alle $n > N_i$ die Beziehung $x_n < M^* + \varepsilon_i$

---

[1]) Falls der gewöhnliche Grenzwert einer Folge existiert, stimmen die Grenzwerte *aller* Teilfolgen mit diesem überein (Nr. 40).

gilt. Jetzt wenden wir uns der zweiten Eigenschaft zu, indem wir wie vorher $\varepsilon = \varepsilon_i$ annehmen und als N die größte der Zahlen n_{i-1} und N_i nehmen; dieser Wahl der Zahlen ε und N mögen die Zahl $n' = n_i$ entsprechen. Für sie gilt einerseits

$$x_{n_i} > M^* - \varepsilon_i,$$

andererseits, da $n_i > N_i$ ist, gleichzeitig auch

$$x_{n_i} < M^* + \varepsilon_i.$$

Außerdem ist $n_i > n_{i-1}$.

Für die Elemente x_{n_i}, die auf diese Weise induktiv konstruiert wurden, gilt

$$|x_{n_i} - M^*| < \varepsilon_i \qquad (i = 2, 3, 4, \ldots),$$

so daß also in der Tat x_{n_i} gegen M^* strebt.

Schließlich zeigen wir noch, daß *kein partieller Grenzwert größer als M^* sein kann.* Es gelte für eine gewisse Teilfolge $\{x_{n_i}\}$ die Beziehung $x_{n_i} \to a$, so daß a einer der partiellen Grenzwerte ist. Nach der ersten Eigenschaft von M^* wird für hinreichend große Indizes (größer als N')

$$x_{n_i} < M^* + \varepsilon.$$

Gehen wir hier zur Grenze über, so erhalten wir $a \leqq M^* + \varepsilon$ und, da ε beliebig war, schließlich

$$a \leqq M^*.$$

Somit ist M^* tatsächlich der größte partielle Grenzwert, d. h.

$$M^* = \overline{\lim}\, x_n.$$

Analog beweist man die Existenz des kleinsten partiellen Grenzwertes. Wir wiederholen nicht alle Überlegungen, sondern weisen nur auf folgendes hin.

Wenn dieser kleinste partielle Grenzwert gleich ∞ ist, so existiert der Grenzwert im gewöhnlichen Sinne,

$$\lim x_n = \infty.$$

Ist aber der kleinste partielle Grenzwert eine endliche Zahl M_*,

$$M_* = \underline{\lim}\, x_n,$$

so hat er Eigenschaften, die den oben für M^* angegebenen analog sind:

1. *Zu beliebigem $\varepsilon > 0$ existiert eine Zahl N'' derart, daß für alle $n > N''$*

$$x_n > M_* - \varepsilon$$

gilt.

2. *Zu beliebigem $\varepsilon > 0$ und jedem N kann man einen Wert $x_{n''}$ mit $n'' > N$ finden derart, daß*

$$x_{n''} < M_* + \varepsilon$$

ist.

Wir wenden uns jetzt dem Beweis der abschließenden Behauptung des Satzes zu. Wenn der (endliche oder unendliche) Grenzwert im gewöhnlichen Sinne, $\lim x_{n''}$, existiert, so stimmen alle partiellen Grenzwerte mit ihm überein (Nr. 40), so daß *für die Existenz eines Grenzwertes die Übereinstimmung von limes superior und limes inferior notwendig ist.*

Wir setzen jetzt voraus, es sei

$$\overline{\lim}\, x_n = \underline{\lim}\, x_n.$$

Ist ihr gemeinsamer Wert ∞ oder $-\infty$, so existiert, wie wir gesehen haben, der Grenzwert der Folge im gewöhnlichen Sinne und hat denselben Wert.

Schließlich seien beide Grenzwerte endlich:

$$M^* = M_* = a.$$

Wenn wir die ersten Eigenschaften der Zahlen M^* und M_* benutzen, dann finden wir zu vorgegebenem $\varepsilon > 0$ eine Zahl N derart, daß für alle $n > N$

$$a - \varepsilon < x_n < a + \varepsilon, \quad \text{d. h.} \quad |x_n - a| < \varepsilon$$

wird. Das bedeutet aber gerade, daß a der Grenzwert der Folge $\{x_n\}$ im gewöhnlichen Sinne ist. Der Satz ist damit bewiesen.

Wir bemerken noch, daß mit Hilfe dieses Satzes ganz einfach bewiesen werden kann, daß die Bolzano-Cauchysche Bedingung (Nr. 39) hinreichend ist. In der Tat entnehmen wir (in den bisherigen Bezeichnungen) der Ungleichung

$$x_{n'} - \varepsilon < x_n < x_{n'} + \varepsilon \quad (\text{für alle } n > N \text{ und alle } n' > N)$$

unmittelbar, daß der größte und der kleinste partielle Grenzwert der Folge $\{x_n\}$ endlich sind und sich nicht mehr als um 2ε unterscheiden; da ε beliebig war, stimmen sie also überein. Hieraus ergibt sich die Existenz eines endlichen Grenzwertes im gewöhnlichen Sinne.

II. Funktionen einer Veränderlichen

§ 1. Der Funktionsbegriff

43. Die Veränderliche und ihr Variationsbereich. In Nr. 22 haben wir den Begriff der Veränderlichen allgemein definiert. Eine *Veränderliche* x ist gegeben durch die Menge $\mathcal{X} = \{x\}$ derjenigen Werte, die sie (bei dem betrachteten Problem) annehmen kann. Diese Menge $\mathcal{X}$, in der jeder Wert x einmal vorkommt, heißt *Variationsbereich* der Veränderlichen x. Im allgemeinen kann jede Zahlenmenge als Variationsbereich einer Veränderlichen dienen.

Wir erwähnten schon, daß man Zahlen geometrisch als Punkte auf einer Zahlengeraden deuten kann. Der Variationsbereich $\mathcal{X}$ der Veränderlichen x ist dann eine bestimmte Punktmenge auf dieser Geraden. Im Zusammenhang damit nennt man die Zahlenwerte einer Veränderlichen vielfach auch einfach *Punkte*.

Oft hat man es mit einer Veränderlichen n zu tun, deren Variationsbereich die Menge N aller natürlichen Zahlen ist. Für die diskrete Veränderliche $x_n = \dfrac{1 + (-1)^n}{n}$ ist der Variationsbereich die Menge der Brüche $\dfrac{1}{2m}$ (mit $m = 1, 2, 3, \ldots$) und die Zahl 0; für eine konstante Größe ist der ganze Variationsbereich eine einzige Zahl.

In der Analysis betrachtet man jedoch gewöhnlich Veränderliche, die sich, wie man sagt, *stetig* oder *kontinuierlich* ändern: Ihr Urbild sind physikalische Größen, wie Zeit, Weg, sich bewegende Punkte usw. Der Variationsbereich solcher Veränderlicher ist ein Zahlenintervall. Meist ist es endlich, von zwei reellen Zahlen a und b $(a < b)$, seinen *Endpunkten*, begrenzt. Die Endpunkte selbst können dazugehören oder nicht. Je nachdem unterscheiden wir

das *abgeschlossene* Intervall	$[a, b]$, $a \leqq x \leqq b$ (beide Endpunkte gehören dazu);
das *halboffene* Intervall	$(a, b]$, $a < x \leqq b$ oder $[a, b)$, $a \leqq x < b$ (nur ein Endpunkt gehört dazu);
das *offene* Intervall	(a, b), $a < x < b$ (kein Endpunkt gehört dazu).

Unter der *Länge* des Intervalls versteht man in allen Fällen die Zahl $b - a$.

Das geometrische Analogon des Zahlenintervalls ist offenbar eine Strecke auf der Zahlengeraden, wobei man je nach dem Typ des Intervalls die Endpunkte hinzuzunehmen hat oder nicht.

Gelegentlich hat man auch *unendliche* Intervalle zu betrachten, bei denen einer der Endpunkte oder beide die „uneigentlichen Zahlen“ $-\infty$, ∞ sind. Man bezeichnet diese Intervalle analog wie oben. So ist $(-\infty, \infty)$ die Menge aller reellen Zahlen; (a, ∞) bezeichnet die Menge aller Zahlen x, die der Ungleichung $x > a$ genügen; das

Intervall $(-\infty, b]$ wird durch die Ungleichung $x \leqq b$ definiert. Geometrisch kann man die unendlichen Intervalle durch eine sich nach beiden Seiten ins Unendliche erstreckende Gerade bzw. durch einen Strahl darstellen.

44. Funktionale Abhängigkeit zwischen Veränderlichen. Beispiele. Hauptgegenstand der Analysis ist jedoch nicht die Untersuchung einer einzigen variierenden Veränderlichen, sondern die Untersuchung der *Abhängigkeit zwischen zwei oder mehreren Veränderlichen, die sich gleichzeitig ändern.* In diesem Kapitel beschränken wir uns auf den einfachsten Fall zweier Veränderlicher.

In verschiedenen Gebieten der Wissenschaft und der Praxis — in der Mathematik, in der Physik und in der Technik — hat man es vielfach mit solchen sich gleichzeitig ändernden Variablen zu tun. Sie können nicht unabhängig voneinander jedes Paar von Werten (aus ihren Variationsbereichen) annehmen: Wenn man einer von ihnen (der *unabhängigen Veränderlichen*) einen konkreten Wert zuschreibt, so ist damit auch der Wert der anderen (der *abhängigen Veränderlichen* oder der *Funktion*) bestimmt. Wir geben einige Beispiele an.

1. Der Flächeninhalt Q eines Kreises ist eine Funktion seines Radius R; sein Wert kann an Hand des Wertes des Radius mit Hilfe der bekannten Formel

$$Q = \pi R^2$$

berechnet werden.

2. Beim freien Fall eines schweren Massenpunktes besteht (bei Fehlen eines Widerstandes) zwischen der vom Anfang der Bewegung an gerechneten und in Sekunden gemessenen Zeit t und dem in dieser Zeit durchlaufenen Weg s (in Metern) die Beziehung

$$s = \frac{gt^2}{2},$$

wobei $g = 9{,}81\ \mathrm{m/s^2}$ die Erdbeschleunigung ist. Hieraus ist der Wert s zu bestimmen, der einem Zeitpunkt t entspricht: Der Weg s ist also eine Funktion der verflossenen Zeit t.

3. Wir betrachten die Menge eines (idealen) Gases, die sich unter dem Kolben eines Zylinders befindet. Unter der Annahme, daß die Temperatur konstant bleibt, genügen das Volumen V und der Druck p dieser Gasmenge dem Boyle-Mariotteschen Gesetz $pV = c =$ const. Wenn V beliebig verändert wird, so ist p als Funktion von V eindeutig durch die Formel

$$p = \frac{c}{V}$$

bestimmt.

4. Zum Schluß gehen wir noch auf die Abhängigkeit des Luftdrucks p von der Höhe h des Ortes über dem Meeresspiegel ein. In der Physik leitet man die barometrische Höhenformel

$$p = p_0 e^{-kh}$$

her, wobei p_0 der Druck auf dem Meeresspiegel und k eine Konstante ist. Nach dieser Formel ist der Wert von p eine Funktion von h und kann bestimmt werden, wenn der Wert von h gegeben ist.

Welche von zwei Veränderlichen als die unabhängige angesehen wird, ist oft gleichgültig oder durch Bequemlichkeitsgründe bestimmt. In den meisten Fällen richtet es sich nach dem Zweck der durchzuführenden Untersuchung.

Beispielsweise kann in dem letzten Beispiel der Zusammenhang zwischen dem Druck p und der Höhe h dazu benutzt werden, in einem Flugzeug aus dem beobachteten Druck die erreichte Höhe zu bestimmen, also die Rolle der Veränderlichen zu vertauschen und die barometrische Höhenformel in der Form

$$h = \frac{1}{k} \ln \frac{p_0}{p}$$

zu benutzen.

45. Die Definition des Funktionsbegriffs.[1]) Wir sehen jetzt, wie üblich, von der physikalischen Bedeutung der betrachteten Größen ab und geben eine abstrakte Definition des Begriffs der Funktion, eines Grundbegriffs der Analysis.

Es seien zwei Veränderliche x und y mit den Variationsbereichen $\mathcal{X}$ bzw. $\mathcal{Y}$ gegeben. Wir nehmen an, daß nach den Bedingungen des Problems der Veränderlichen x jeder Wert aus dem Bereich $\mathcal{X}$ zugeschrieben werden kann, ohne irgendwelche Einschränkung. Dann wird die Veränderliche y *Funktion der Veränderlichen* x im Variationsbereich $\mathcal{X}$ genannt, wenn nach einer Vorschrift oder einem Gesetz jedem Wert x aus $\mathcal{X}$ ein bestimmter Wert y (aus $\mathcal{Y}$) zugeordnet werden kann. Die unabhängige Veränderliche x nennt man auch *Argument* der Funktion.

In dieser Definition sind zwei Dinge wesentlich: erstens die Angabe des Gebietes $\mathcal{X}$, in dem das Argument x variiert (des sogenannten *Definitionsbereichs der Funktion*) und zweitens die Angabe der Vorschrift oder des Gesetzes, wodurch die Zuordnung zwischen den x- und den y-Werten beschrieben wird. (Der Wertebereich $\mathcal{Y}$ der Funktion y wird im allgemeinen nicht angegeben, weil die Zuordnungsvorschrift selbst schon die Menge der Funktionswerte, die angenommen werden können, festlegt.) Die Menge der tatsächlich angenommenen Funktionswerte wird *Wertevorrat* der Funktion genannt.

Man kann den Begriff der Funktion noch allgemeiner definieren, wenn man zuläßt, daß jedem Wert x aus $\mathcal{X}$ nicht nur ein Wert, sondern auch mehrere (sogar unendlich viele) Werte y entsprechen können. Dann nennt man die Funktion *mehrdeutig*, im Unterschied zu den oben definierten *eindeutigen* Funktionen. Übrigens werden in der reellen Analysis kaum mehrdeutige Funktionen auftreten, so daß wir von jetzt an unter Funktionen, sofern nicht ausdrücklich etwas anderes gesagt wird, stets eindeutige Funktionen verstehen.

Um anzudeuten, daß y eine Funktion von x ist, schreibt man:

$$y = f(x), \quad y = \varphi(x), \quad y = F(x) \quad \text{usw.}^{2)}$$

Die Buchstaben f, φ, F, ... charakterisieren gerade die Vorschrift, nach der man den Wert y erhält, wenn x gegeben ist. Wenn also gleichzeitig verschiedene Funktionen von ein und demselben Argument x betrachtet werden, die durch verschiedene Gesetze zusammenhängen, so müssen sie mit verschiedenen Buchstaben bezeichnet bzw. durch Indizes unterschieden werden.

Obwohl gerade der Buchstabe f (bzw. φ) an das Wort „Funktion" erinnern soll, kann man zur Bezeichnung eines funktionalen Zusammenhangs natürlich auch jeden

---

[1]) Vgl. dazu auch die Ausführungen in Nr. 257.

[2]) Dies spricht man folgendermaßen aus: „y ist gleich f von x" oder „y ist gleich φ von x" usw.

anderen Buchstaben benutzen; manchmal wiederholt man sogar den Buchstaben y, schreibt also $y = y(x)$.

Gelegentlich schreibt man das Argument als Index an das Funktionszeichen. Hierunter fällt auch die Bezeichnung einer diskreten Veränderlichen x_n, die (wie wir jetzt sagen können) eine Funktion der „unabhängigen Veränderlichen" n ist, die die Menge $\mathcal{N}$ der natürlichen Zahlen n durchläuft. Analog ist auch die Bezeichnung N_ε für den Index N (in der Definition des Grenzwertes für Folgen; vgl. Nr. 23) zu deuten, die auf die Abhängigkeit von ε weist, usw.

Wenn man eine Funktion $y = f(x)$ betrachtet und den *speziellen Wert* angeben will, den sie für $x = x_0$ annimmt, so benutzt man die Bezeichnung $f(x_0)$. Ist etwa

$$f(x) = \frac{1}{1+x^2}, \quad g(t) = \frac{10}{t}, \quad h(u) = \sqrt{1-u^2}, \ldots,$$

so bezeichnet $f(1)$ den Zahlenwert der Funktion $f(x)$ an der Stelle $x = 1$, d. h. die *Zahl* $\frac{1}{2}$; analog bezeichnet $g(5)$ die *Zahl* 2, $h\left(\frac{3}{5}\right)$ die *Zahl* $\frac{4}{5}$, usw.

Wir wenden uns jetzt der *Vorschrift* oder dem *Gesetz* selbst zu, das zwischen den Werten der Veränderlichen besteht und worin das Wesen der funktionalen Abhängigkeit zum Ausdruck kommt. Diese Vorschrift kann ganz verschiedenartiger Natur sein, da nichts Einschränkendes darüber gesagt ist.

Am einfachsten und natürlichsten kann diese Vorschrift in Form eines *analytischen Ausdrucks* oder in Gestalt von *Formeln* verwirklicht sein, die angeben, welche Rechenoperationen mit konstanten Zahlen und den x-Werten auszuführen sind, um die entsprechenden y-Werte zu erhalten. Diese analytische Form der Vorgabe einer Funktion ist für die Analysis am wichtigsten (wir werden in Nr. 46 darauf zurückkommen). Hiermit ist der Leser aus der Schulmathematik vertraut; übrigens benutzten wir die analytische Methode in den in Nr. 44 angegebenen Beispielen.

Es wäre aber falsch, würde man annehmen, diese Methode sei die einzige, Funktionen vorzugeben. In der Mathematik selbst kommt es vielfach vor, daß eine Funktion nicht durch eine Formel definiert wird. So ist es z. B. bei der Funktion $[x] = E(x)$, „dem ganzen Teil der Zahl x".[1]) Offenbar ist

$$E(1) = 1, \quad E(2{,}5) = 2, \quad E(\sqrt{13}) = 3, \quad E(-\pi) = -4 \quad \text{usw.},$$

obwohl es keine Formel für $E(x)$ gibt. Dasselbe gilt für zahlreiche „arithmetische Funktionen", d. h. Funktionen eines natürlichen Arguments, die nur natürliche Zahlen als Werte annehmen. Als Beispiel erwähnen wir die *Fakultät* der Zahl n,

$$n! = 1 \cdot 2 \cdot 3 \cdots n$$

(gesprochen: n Fakultät), ferner die Funktion $\tau(n)$, die die *Anzahl der Teiler von n* angibt, oder die *Eulersche Funktion* $\varphi(n)$, die *Anzahl der zu n teilerfremden Zahlen in der Folge* $1, 2, 3, \ldots, n$. (Diese sogenannten *zahlentheoretischen Funktionen* sind für ganzzahlige Werte des Arguments definiert.) Trotzdem kann man die Werte der Funktionen mit derselben Bestimmtheit berechnen, als ob sie durch Formeln gegeben seien. So ist z. B.

$$\tau(10) = 4, \quad \tau(12) = 6, \quad \tau(16) = 5, \quad \ldots,$$
$$\varphi(10) = 4, \quad \varphi(12) = 4, \quad \varphi(16) = 8,$$

[1]) Vgl. die Fußnote auf S. 48.

In den Naturwissenschaften und in der Technik ist eine funktionale Abhängigkeit zwischen Größen häufig experimentell, d. h. durch Beobachtungen gegeben. Unterwerfen wir beispielsweise Wasser einem beliebig gewählten Druck p, so kann man durch einen Versuch feststellen, bei welcher Temperatur θ das Wasser zum Sieden kommt: θ ist also eine Funktion von p. Diese funktionale Abhängigkeit ist aber nicht durch eine Formel, sondern nur durch eine Tabelle gegeben, in der einfach die aus dem Versuch erhaltenen Daten zusammengestellt sind. Beispiele für diese tabellarische Methode zur Vorgabe einer Funktion finden wir in jedem technischen Nachschlagewerk.

Schließlich sei noch erwähnt, daß in einigen Fällen die funktionale Abhängigkeit zwischen physikalischen Größen mit Hilfe eines selbstregistrierenden Gerätes unmittelbar durch eine Kurve beschrieben werden kann. So gibt das mit Hilfe des Indikators aufgenommene *Indikatordiagramm* die Abhängigkeit zwischen dem Volumen V und dem Druck p im Zylinder einer arbeitenden Dampfmaschine an; das mit Hilfe eines Barographen aufgenommene *Barogramm* gibt den vierundzwanzigstündigen Verlauf des atmosphärischen Druckes wieder, usw.

Wir gehen hier auf die tabellarischen und die graphischen Methoden zur Vorgabe einer funktionalen Abhängigkeit nicht näher ein, da wir sie in unserer Darstellung nicht benutzen werden.

46. Die analytische Methode zur Vorgabe von Funktionen. Wir geben nun einige Erläuterungen über die Vorgabe von Funktionen durch *analytische Ausdrücke* oder durch *Formeln*, die in der Analyse eine besonders wichtige Rolle spielen.

1. Zunächst fragen wir, welche *analytischen Operationen* oder *Rechenoperationen* dürfen in diesen Formeln vorkommen? An erster Stelle versteht man darunter alle in der elementaren Algebra und Trigonometrie auftretenden Operationen: die Grundrechenarten, das Potenzieren (und Radizieren), das Logarithmieren, der Übergang vom Winkel zu trigonometrischen Größen und umgekehrt (Näheres siehe Nr. 48 bis 51). Ferner gehören, und das sei besonders hervorgehoben, weitere Operationen dazu, die wir im Verlauf dieses Lehrgangs kennenlernen werden, in der Hauptsache der Grenzübergang, den wir schon in Kapitel I behandelt haben.

Somit wird der volle Inhalt der Termini „analytischer Ausdruck" bzw. „Formel" erst allmählich klarwerden.

2. Die zweite Bemerkung bezieht sich auf den *Definitionsbereich* von Funktionen, die durch analytische Ausdrücke oder Formeln gegeben sind.

Jeder analytische Ausdruck, der das Argument x enthält, hat sozusagen einen natürlichen Anwendungsbereich: die Menge aller derjenigen x-Werte, für die er sinnvoll bleibt, d. h. wohlbestimmte endliche reelle Werte hat. Wir erklären das an einfachsten Beispielen.

Für den Ausdruck $\frac{1}{1+x^2}$ ist dieser Bereich die Menge aller reellen Zahlen. Für den Ausdruck $\sqrt{1-x^2}$ reduziert sich dieser Bereich auf das abgeschlossene Intervall $[-1, 1]$, innerhalb dessen seine Werte reell sind. Dagegen hat der Ausdruck $\frac{1}{\sqrt{1-x^2}}$ als natürlichen Anwendungsbereich das offene Intervall $(-1, 1)$, weil an den Endpunkten sein Nenner 0 wird. Manchmal besteht der Wertebereich, für den ein Ausdruck sinnvoll ist, aus verschiedenen Intervallen: Für $\sqrt{x^2-1}$ sind das die Intervalle $(-\infty, -1]$ und $[1, \infty)$, für $\frac{1}{x^2-1}$ die Intervalle $(-\infty, -1)$, $(-1, 1)$ und $(1, \infty)$ usw.[1])

[1]) Natürlich sind Ausdrücke, die für keinen x-Wert einen Sinn haben, für uns uninteressant.

Als letztes Beispiel betrachten wir die Summe der unendlichen geometrischen Reihe

$$1 + x + x^2 + \cdots + x^{n-1} + \cdots = \lim\,(1 + x + x^2 + \cdots + x^{n-1}).$$

Für $|x| < 1$ existiert, wie wir wissen (vgl. Nr. 25, Beispiel 7), dieser Grenzwert und ist gleich $\frac{1}{1-x}$. Für $|x| > 1$ wird der Grenzwert entweder gleich ∞, oder er existiert nicht. Somit ist für diesen analytischen Ausdruck der natürliche Anwendungsbereich das offene Intervall $(-1, 1)$.

In den folgenden Darlegungen werden wir sowohl kompliziertere als auch allgemeinere analytische Ausdrücke zu betrachten haben, und wir werden uns öfter mit der Untersuchung der Eigenschaften von Funktionen beschäftigen, die durch einen Ausdruck in dem ganzen Bereich gegeben sind, in dem er sinnvoll bleibt, d. h. mit der Untersuchung des analytischen Ausdrucks an sich.

Jedoch ist auch eine andere Sachlage möglich, auf die wir von vornherein aufmerksam machen möchten. Wir stellen uns vor, irgendein konkretes Problem, in welchem die Veränderliche x vom Wesen der Sache her auf einen Variationsbereich $\mathscr{X}$ beschränkt ist, führe auf die Untersuchung einer Funktion $f(x)$, die durch einen analytischen Ausdruck angegeben werden kann. Obwohl es vorkommen kann, daß dieser Ausdruck auch außerhalb des Gebietes $\mathscr{X}$ sinnvoll ist, hat es dann keinen Zweck, ihn außerhalb dieser Grenzen zu untersuchen. Hier spielt der analytische Ausdruck eine untergeordnete Rolle, nämlich die eines Hilfsmittels.

Untersuchen wir z. B. den freien Fall eines schweren Massenpunktes von der Höhe h auf die Erdoberfläche, so kommen wir auf die Formel

$$s = \frac{gt^2}{2}$$

vgl. Nr. 44, Beispiel 2); es würde hier sinnlos sein, negative Werte von t zu betrachten oder solche Werte von t, der größer als $T = \sqrt{\frac{2h}{g}}$ sind; denn für $t = T$ ist, wie man leicht sieht, der Punkt schon auf der Erde angekommen. Abgesehen davon ist der Ausdruck $\frac{gt^2}{2}$ selbst natürlich für alle reellen t sinnvoll.

3. Es kann vorkommen, daß eine Funktion für alle Werte des Arguments nicht durch ein und dieselbe Formel definiert ist, sondern für gewisse Werte durch eine bestimmte Formel und für weitere Werte durch eine andere Formel. Als Beispiel einer solchen Funktion kann im Intervall $(-\infty, \infty)$ die Funktion dienen, die durch folgende drei Formeln definiert ist:

$$f(x) = 1 \quad \text{für} \quad |x| > 1 \quad (\text{d. h. für } x > 1 \text{ bzw. } x < -1),$$

$$f(x) = -1 \quad \text{für} \quad |x| < 1 \quad (\text{d. h. für } -1 < x < 1),$$

und schließlich

$$f(x) = 0 \quad \text{für} \quad x = \pm 1.$$

Wir erwähnen noch die *Dirichletsche Funktion*[1]), die folgendermaßen definiert ist:

$$\chi(x) = 1, \quad \text{wenn } x \text{ rational},$$

$$\chi(x) = 0, \quad \text{wenn } x \text{ irrational ist.}$$

[1]) Peter Gustav Lejeune Dirichlet, 1805—1859, deutscher Mathematiker.

Schließlich betrachten wir noch die *Kroneckersche Funktion*[1]), die KRONECKER „Signum x" nannte und mit $\operatorname{sgn} x$ bezeichnete[2]):

$$\operatorname{sgn} x = 1 \quad \text{für} \quad x > 0;$$
$$\operatorname{sgn} x = -1 \quad \text{für} \quad x < 0;$$
$$\operatorname{sgn} 0 = 0.$$

Übrigens darf man nicht glauben, daß ein prinzipieller Unterschied zwischen solchen Funktionen besteht, die durch eine einzige Formel für alle in Frage kommenden x-Werte gegeben sind, und solchen, die durch mehrere Formeln definiert werden. Vielfach läßt sich eine Funktion, die durch mehrere Formeln gegeben ist, auch durch eine einzige angeben. Allerdings muß man dabei in Kauf nehmen, daß der Ausdruck komplizierter wird.

Wenn wir z. B. den Grenzübergang zulassen, so kann die erste der oben angegebenen Funktionen $f(x)$ durch eine einzige Formel (für alle x) definiert werden:

$$f(x) = \lim \frac{x^{2n} - 1}{x^{2n} + 1}.$$

In der Tat strebt für $|x| > 1$ die Potenz x^{2n} gegen ∞, der reziproke Ausdruck also gegen 0 (Nr. 27); somit ist

$$\lim \frac{x^{2n} - 1}{x^{2n} + 1} = \lim \frac{1 - \dfrac{1}{x^{2n}}}{1 + \dfrac{1}{x^{2n}}} = 1.$$

Für $|x| < 1$ strebt die Potenz x^{2n} gegen 0 (vgl. Nr. 25, Beispiel 6); in diesem Fall ist also

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

Schließlich wird für $x = \pm 1$ offenbar $x^{2n} = 1$, und hieraus folgt

$$\frac{x^{2n} - 1}{x^{2n} + 1} = 0,$$

als Grenzwert ergibt sich somit 0. All dies steht mit der ursprünglichen Definition völlig in Einklang.

47. Graphische Darstellung von Funktionen. Obwohl man in der Analysis Funktionen nicht graphisch vorgibt, greift man zur Illustration gern zu ihrer graphischen Darstellung. Die Überschaubarkeit und Anschaulichkeit läßt die graphische Darstellung bei der Untersuchung der Eigenschaften von Funktionen zu einem unersetzlichen Hilfsmittel werden.

Es sei in einem Intervall $\mathcal{X}$ eine Funktion $y = f(x)$ gegeben. Wir stellen sie in der Ebene zweier (reeller) zueinander senkrechter Koordinatenachsen, der x- und der *y-Achse*, dar. Wir betrachten ein Paar entsprechender Werte x und y, wobei x dem Intervall $\mathcal{X}$ entnommen und $y = f(x)$ ist; das Bild dieses Paares in der Ebene ist der *Punkt* $M(x, y)$ mit der *Abszisse* x und der *Ordinate* y. Ändert sich die Variable x zwischen den Grenzen ihres Intervalls, so beschreibt dieser Punkt eine *Kurve* AB (Abb. 5), die das geometrische Bild unserer Funktion ist und ihre *graphische Darstellung* (ihre Kurve) genannt wird. Unter diesen Bedingungen heißt die Gleichung $y = f(x)$ die *Gleichung der Kurve* AB.

---

[1]) LEOPOLD KRONECKER, 1823—1891, deutscher Mathematiker.

[2]) von dem lateinischen Wort signum = Vorzeichen

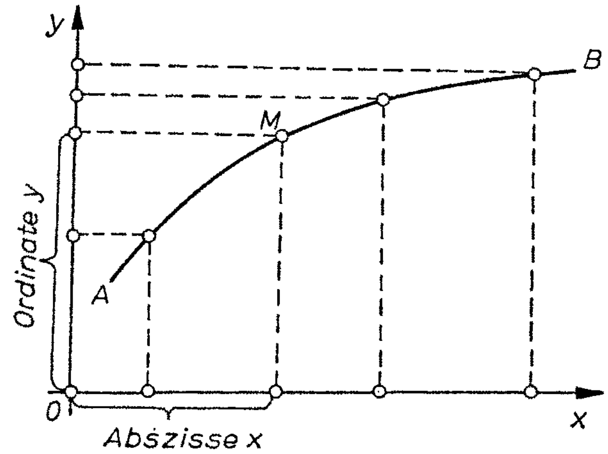

Abb. 5

Zum Beispiel sind in den Abb. 6 und 7 die Kurven der Funktionen

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

und

$$y = \pm\sqrt{x^2 - 1} \qquad (|x| \geqq 1)$$

angegeben, ein *Kreis* und eine *gleichseitige Hyperbel* (in jedem dieser Beispiele handelt es sich um zwei eindeutige Funktionen). Weitere Beispiele graphischer Darstellungen findet der Leser in den nächsten Nummern.

Man konstruiert eine Kurve gewöhnlich *punktweise*. Dazu nimmt man im Intervall $\mathscr{X}$ eine Reihe benachbarter x-Werte und berechnet nach der Formel $y = f(x)$ die entsprechenden y-Werte:

$$\begin{array}{c|ccccc} x & x_1 & x_2 & x_3 & \ldots & x_n \\ \hline y & y_1 & y_2 & y_3 & \ldots & y_n \end{array};$$

dann trägt man die Punkte

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_n, y_n)$$

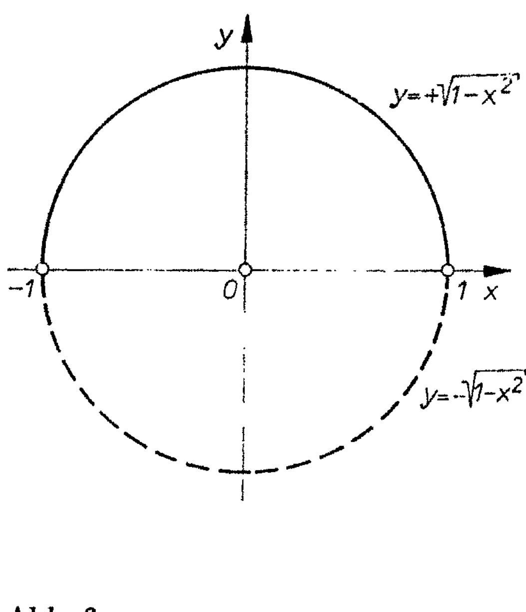

Abb. 6

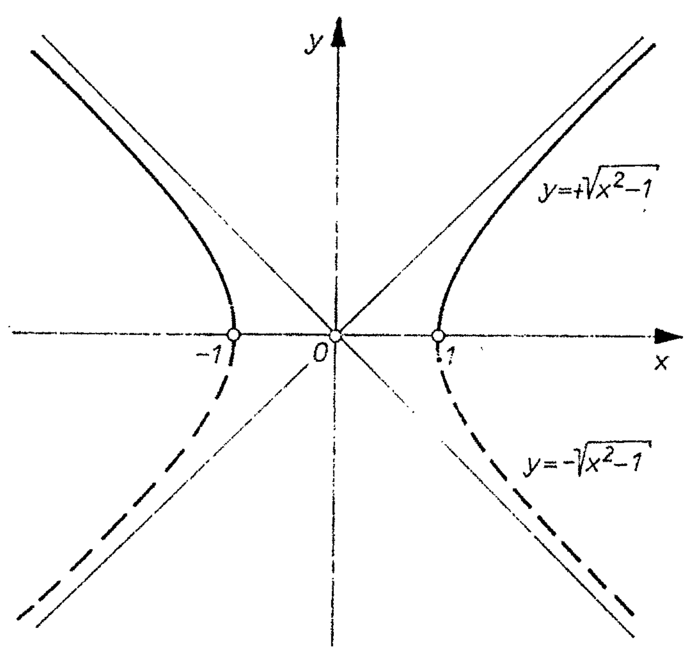

Abb. 7

in die Zeichnung ein. Durch diese Punkte zieht man mit der Hand oder mit Hilfe eines Kurvenlineals eine Kurve, die (natürlich nur in einer gewissen Näherung) die gesuchte graphische Darstellung ist. Je glatter der Verlauf der Kurve ist und je dichter beieinander die Punkte genommen wurden, desto genauer gibt die gezeichnete Kurve den Funktionsverlauf wieder.

Zwar kann man sich das geometrische Bild einer (nicht allzu „pathologischen“) Funktion meistens vorstellen, es wird aber nicht immer eine Kurve im gewöhnlichen anschaulichen Sinne sein.

Wir konstruieren beispielsweise die graphische Darstellung der Funktion $y = E(x) = [x]$. Da die Funktion in den Intervallen ..., $[-2, -1), [-1, 0), [0, 1), [1, 2), [2, 3), \ldots$ die konstanten Werte ..., $-2, -1, 0, 1, 2, \ldots$ annimmt, besteht die graphische Darstellung aus einer Reihe einzelner horizontaler Strecken, deren rechter Endpunkt fehlt (Abb. 8).[1])

Für die Dirichletsche Funktion $\chi(x)$ besteht die graphische Darstellung aus der Menge der Punkte mit *irrationaler* Abszisse auf der x-Achse und aus der Menge der Punkte mit *rationaler* Abszisse auf der Geraden $y = 1$; sie darzustellen, ist unmöglich.

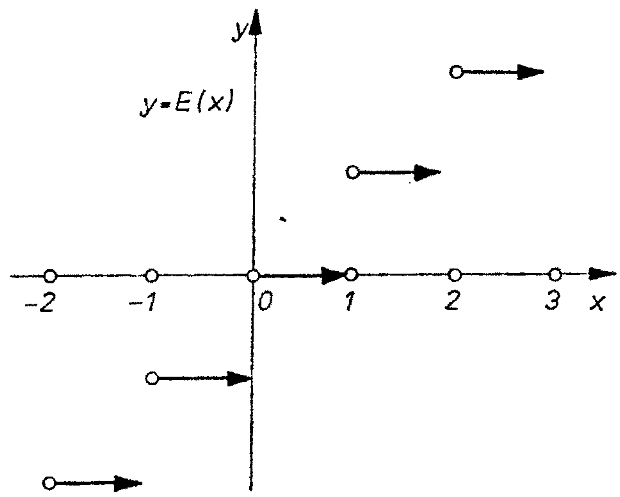

Abb. 8

48. Die wichtigsten Funktionenklassen. Wir zählen jetzt einige Klassen von Funktionen auf, die man *elementare* Funktionen nennt.

1. *Ganze und gebrochene rationale Funktionen.* Eine Funktion der Gestalt

$$y = a_0x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n$$

($a_0, a_1, a_2, \ldots$ Konstante) heißt *Polynom* oder *ganze rationale Funktion.*

Der Quotient zweier solcher Polynome,

$$y = \frac{a_0x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n}{b_0x^m + b_1x^{m-1} + \cdots + b_{m-1}x + b_m},$$

heißt *gebrochene rationale Funktion* oder einfach *rationale Funktion.* Er ist für alle Werte x definiert, für die der Nenner nicht 0 wird.

Als Beispiel ist in Abb. 9 die Funktion $y = ax^2$ (*Parabel*) für verschiedene Werte des Koeffizienten a graphisch dargestellt. In Abb. 10 ist die Funktion $y = \frac{a}{x}$ (*gleichseitige Hyperbel*) für verschiedene a-Werte graphisch dargestellt.

---

[1]) Das deuten wir durch Pfeile an, deren Spitzen auf die Punkte zeigen, die nicht mehr zur graphischen Darstellung dieser Funktion gehören.

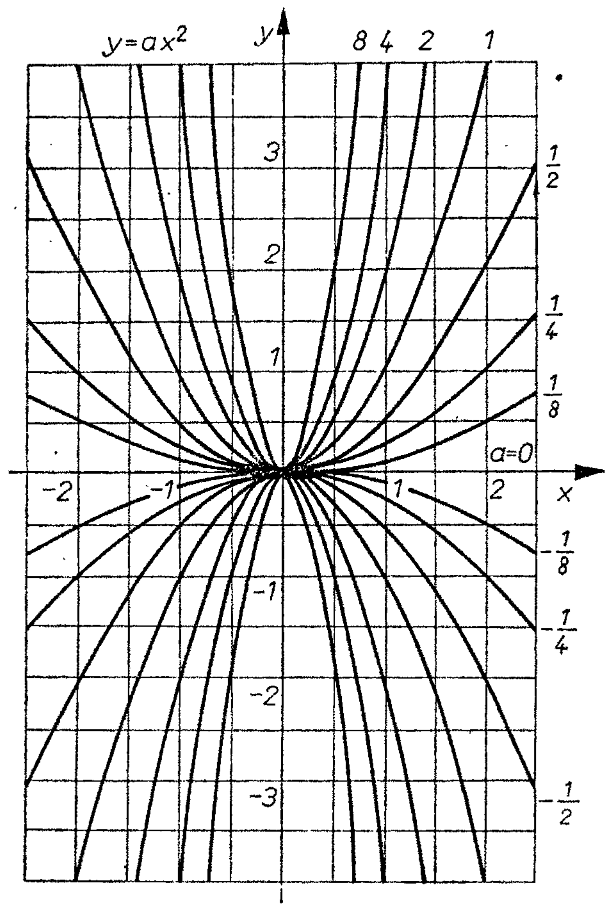

Abb. 9

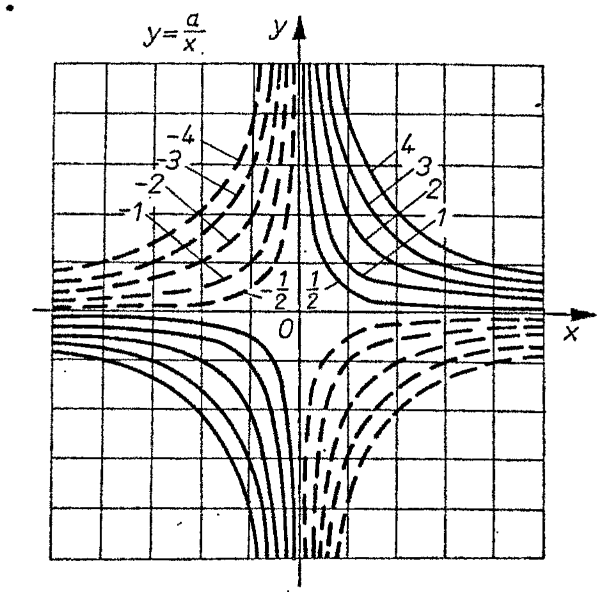

Abb. 10

2. *Die Potenzfunktion.* Die Funktion

$$y = x^\mu,$$

wobei μ eine beliebige konstante reelle Zahl ist, heißt *Potenzfunktion.* Für ganzes μ ist sie eine rationale Funktion. Ist μ ein Bruch, so erhält man eine *Wurzel.* Beispielsweise sei m eine natürliche Zahl und

$$y = x^{1/m} = \sqrt[m]{x};$$

diese Funktion ist für alle x definiert, wenn m ungerade ist, und nur für nichtnegative Werte, wenn m gerade ist (in diesem Fall meinen wir die *positive* Wurzel). Ist schließlich μ eine irrationale Zahl, so wollen wir $x > 0$ voraussetzen ($x = 0$ ist nur für $\mu > 0$ zulässig.)

Bemerkung. Oft wird für nicht ganzes μ der Einheitlichkeit halber x^μ nur für $x \geqq 0$ oder $x > 0$ definiert und zusätzlich für ganzes ungerades m

$$x^{1/m} = -|x|^{1/m} \qquad (x < 0)$$

gesetzt.

In Abb. 11 und 12 sind die Kurven der Potenzfunktion für verschiedene μ-Werte angegeben.

3. *Die Exponentialfunktion.* Hierunter versteht man die Funktion

$$y = a^x,$$

wobei a eine positive Zahl $\neq 1$ ist; x kann beliebige reelle Zahlenwerte annehmen.

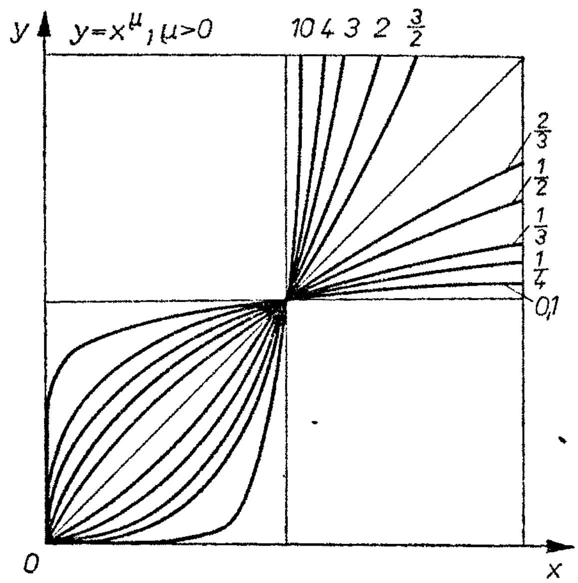

Abb. 11

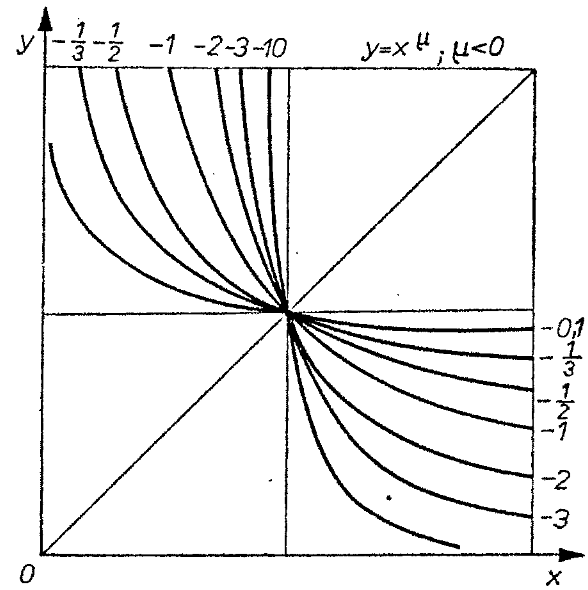

Abb. 12

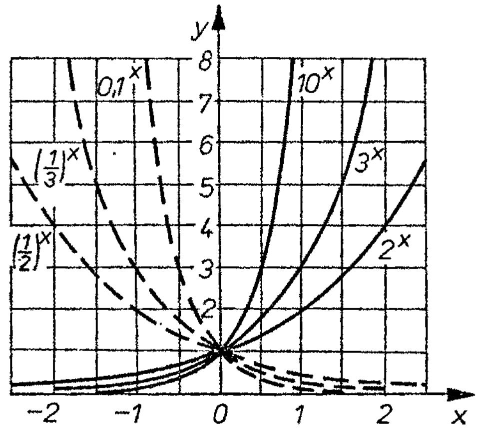

Abb. 13

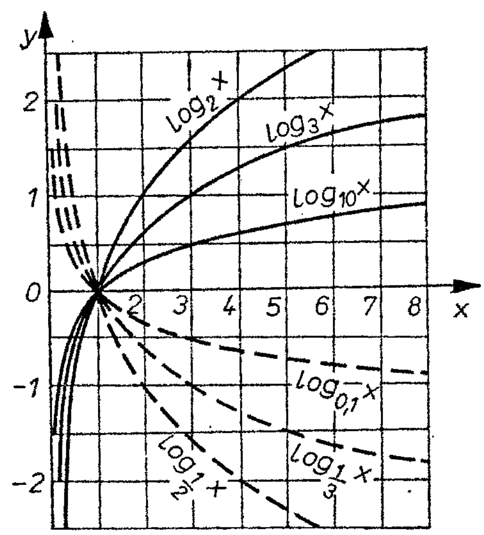

Abb. 14

Graphisch dargestellt ist diese Funktion für verschiedene a-Werte in Abb. 13.

4. *Die Logarithmusfunktion.* Die Funktion

$$y = \log_a x,$$

wobei a, wie schon in Nr. 20, eine positive Zahl $\neq 1$ ist und x nur positive Werte annimmt, heißt *Logarithmusfunktion.*

In Abb. 14 ist diese Funktion für verschiedene a-Werte graphisch dargestellt.

5. *Die trigonometrischen Funktionen* (*Kreisfunktionen*).

$$y = \sin x, \quad y = \cos x, \quad y = \tan x,$$

$$y = \cot x, \quad y = \sec x, \quad y = \operatorname{cosec} x.$$

Sehr wichtig ist es, ein für allemal darauf hinzuweisen, daß *die Argumente der trigonometrischen Funktionen, als Winkelmaße aufgefaßt, immer diese Winkel im Bogenmaß ausdrücken* (wenn nicht ausdrücklich etwas anderes gesagt ist). Für $\tan x$ und sec x schließen wir Werte der Form

$$x = (2k + 1)\frac{\pi}{2}$$

und für cot x und cosec x Werte der Form

$$x = k\pi \quad (k \text{ ganz})$$

aus.

Die Funktionen $y = \sin x$ ($y = \cos x$) und $y = \tan x$ ($y = \cot x$) sind in Abb. 15 und 16 graphisch dargestellt. Die graphische Darstellung des Sinus wird *Sinuskurve* genannt usw.

Manchmal, besonders bei technischen Problemen, sind von Interesse die

6. *Hyperbolischen Funktionen* (*Hyperbelfunktionen*). Die Funktionen

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

$$\tanh x = \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \qquad \coth x = \frac{\cosh x}{\sinh x} = \frac{e^x + e^{-x}}{e^x - e^{-x}},$$

heißen *hyperbolischer Sinus, Kosinus, Tangens, Kotangens,* ... (*sinus hyperbolicus* usw.); sie sind für alle x-Werte definiert, ausgenommen coth x, der für $x = 0$ sinnlos ist. *Diese Funktionen zeigen eine bemerkenswerte Analogie zu den trigonometrischen Funktionen.*

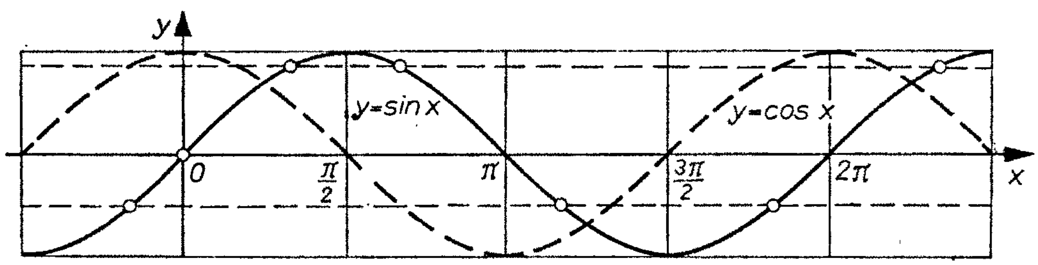

Abb. 15

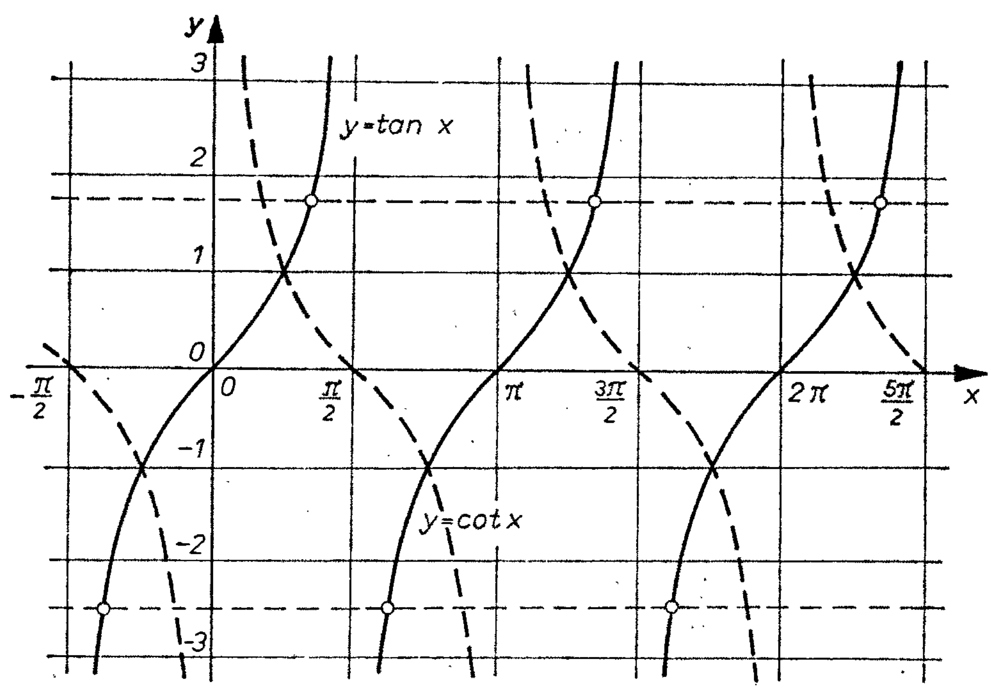

Abb. 16

So gelten die Formeln (man achte besonders auf das Vorzeichen!)

$$\begin{aligned}\cosh(x \pm y) &= \cosh x \cdot \cosh y \pm \sinh x \cdot \sinh y,\\ \sinh(x \pm y) &= \sinh x \cdot \cosh y \pm \cosh x \cdot \sinh y,\end{aligned} \tag{*}$$

aus denen für $y = x$ speziell

$$\cosh^2 x - \sinh^2 x = 1,$$

$$\cosh 2x = \cosh^2 x + \sinh^2 x,$$

$$\sinh 2x = 2 \sinh x \cdot \cosh x$$

folgt. Beispielsweise ergibt sich die erste der Formeln (*) aus der leicht nachzuprüfenden Identität

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

So lassen sich auch die restlichen Formeln verifizieren.

Die hyperbolischen Funktionen sind in Abb. 17 und 18 graphisch dargestellt.

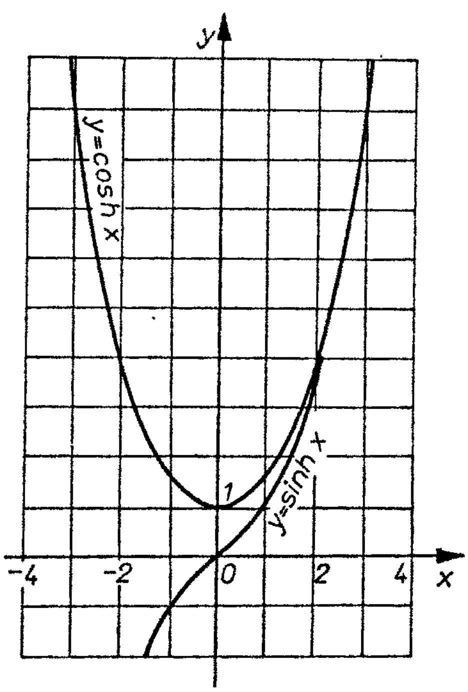

Abb. 17

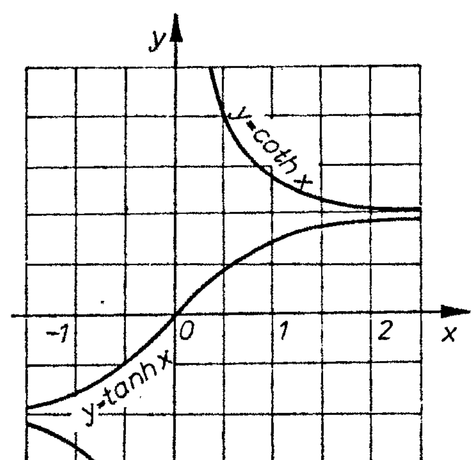

Abb. 18

49. Der Begriff der Umkehrfunktion. Ehe wir die Umkehrfunktionen der trigonometrischen Funktionen behandeln, erklären wir allgemein den Begriff der Umkehrfunktion (der inversen Funktion).

Wir nehmen an, die Funktion $y = f(x)$ sei in einem Bereich $\mathcal{X}$ gegeben, und $\mathcal{Y}$ sei die Menge aller Werte, die diese Funktion annimmt, wenn x innerhalb der Grenzen des Bereichs $\mathcal{X}$ variiert. (Wir können uns sowohl $\mathcal{X}$ als auch $\mathcal{Y}$ als Intervalle vorstellen.)

Wir wählen einen beliebigen Wert $y = y_0$ aus dem Bereich $\mathcal{Y}$; dann existiert notwendigerweise in dem Bereich $\mathcal{X}$ ein Wert $x = x_0$, für den unsere Funktion den Wert y_0 annimmt, so daß also

$$f(x_0) = y_0$$

ist; es kann auch mehrere solcher Werte x_0 geben. Somit ist jedem Wert y aus $\mathcal{Y}$ ein Wert oder mehrere Werte x zugeordnet; damit ist in dem Bereich $\mathcal{Y}$ eine eindeutige oder eine mehrdeutige Funktion $x = g(y)$ definiert; diese Funktion nennt man die *Umkehrfunktion* oder die *inverse Funktion* der Funktion $y = f(x)$.

Wir betrachten einige Beispiele:

1. Es sei $y = a^x$ $(a > 1)$, wobei x innerhalb des Intervalls $\mathcal{X} = (-\infty, \infty)$ variiert. Die Werte y füllen dann das Intervall $\mathcal{Y} = (0, \infty)$ aus, wobei jedem y aus diesem Intervall, wie wir wissen (Nr. 20), in $\mathcal{X}$ ein einziger Wert $x = \log_a y$ entspricht. In diesem Fall ist die inverse Funktion *eindeutig*.

2. Dagegen ist für die Funktion $y = x^2$, wenn x das Intervall $\mathcal{X} = (-\infty, \infty)$ durchläuft, die Umkehrfunktion *zweideutig*: Jedem Wert y aus $\mathcal{Y} = [0, \infty)$ entsprechen die beiden Werte $x = \pm\sqrt{y}$ aus dem Intervall $\mathcal{X}$. Statt dieser zweideutigen Funktion betrachtet man gewöhnlich die beiden eindeutigen Funktionen $x = +\sqrt{y}$ und $x = -\sqrt{y}$, die *Äste* (*Zweige*) der zweideutigen Funktion. Man kann sie auch einzeln als Umkehrfunktion von $y = x^2$ ansehen, wenn man annimmt, daß der Bereich, den x durchläuft, auf das Intervall $[0, \infty)$ bzw. $(-\infty, 0]$ beschränkt ist.

3. Betrachten wir jetzt $y = \cosh x$, wobei x wieder das Intervall $\mathcal{X} = (-\infty, \infty)$ durchläuft, so erhalten wir durch Auflösung der Gleichung

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

oder

$$e^{2x} - 2y \cdot e^x + 1 = 0$$

nach e^x für $y \geqq 1$ die beiden Werte

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

hieraus folgt

$$x = \ln\left(y \pm \sqrt{y^2 - 1}\right).$$

Das ist also wieder eine zweideutige Funktion, die in zwei eindeutige Äste zerfällt, die einzeln dem x-Intervall von 0 bis ∞ und dem x-Intervall von $-\infty$ bis 0 entsprechen. Man spricht hier von *Area-Funktionen*.

4. Für $y = \sinh x$ erhalten wir aus der Gleichung

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

oder

$$e^{2x} - 2y \cdot e^x - 1 = 0$$

für beliebiges y nur einen Wert für e^x, nämlich

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

Der zweite Wert mit dem Minuszeichen vor der Wurzel wäre nämlich negativ, was nicht möglich sein kann; er braucht also nicht berücksichtigt zu werden. Hieraus folgt

$$x = \ln\left(y + \sqrt{y^2 + 1}\right),$$

so daß hier die inverse Funktion eindeutig ist.

Wir bemerken noch, daß man an Hand der graphischen Darstellung einer Funktion $y = f(x)$ leicht feststellen kann, ob die Umkehrfunktion $x = g(y)$ eindeutig ist oder nicht. Der erste Fall tritt ein, wenn jede Parallele zur x-Achse die Kurve nur in einem einzigen Punkt schneidet. Umgekehrt, wenn einige dieser Geraden die Kurve in mehreren Punkten schneiden, so ist die inverse Funktion mehrdeutig. In diesem Fall kann man oft an Hand der Kurve das x-Intervall so in Teilintervalle zerlegen, daß jedem Teilintervall ein eindeutiger „Zweig" dieser inversen Funktion entspricht. Beispielsweise ist beim ersten Blick auf die Parabel in Abb. 4, der graphischen Darstellung von $y = x^2$, klar, daß die Umkehrfunktion zweideutig ist und daß man, um einen eindeutigen „Zweig" zu erhalten, nur den rechten und den linken Ast dieser Parabel einzeln zu betrachten braucht, d. h. positive und negative x-Werte einzeln.[1])

Ist $x = g(y)$ die Umkehrfunktion von $y = f(x)$, so stimmen offenbar die graphischen Darstellungen beider Funktionen überein. Man kann jedoch auch das Argument der inversen Funktion mit dem Buchstaben x bezeichnen, d. h. statt $x = g(y)$ die Funktion $y = g(x)$ betrachten. Dann muß man nur die horizontale Achse als y-Achse und die vertikale als x-Achse bezeichnen. Die graphische Darstellung bleibt die gleiche. Will man wie üblich die (neue) x-Achse horizontal und die (neue) y-Achse vertikal haben, so muß man diese Achsen vertauschen; das ändert auch die graphische Darstellung. Um das zu realisieren, klappen wir die x, y-Ebene in der Abbildung um 180° um die Winkelhalbierende des ersten Quadranten (Abb. 19).

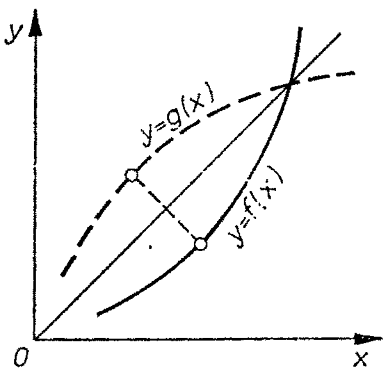

Abb. 19

Somit erhalten wir die graphische Darstellung von $y = g(x)$ als Spiegelbild der graphischen Darstellung von $y = f(x)$ bezüglich dieser Winkelhalbierenden. Aus Abb. 13 und 14 z. B. ist sofort zu sehen, daß die graphischen Darstellungen geradeso auseinander hervorgehen. Genauso kann man aus diesen Überlegungen leicht schließen, daß je zwei Kurven in Abb. 11 und 12 (bezüglich der Winkelhalbierenden) symmetrisch sind.

50. Die Umkehrfunktionen der trigonometrischen Funktionen. Als Ergänzung zur Klasse der elementaren Funktionen, die wir in Nr. 48 behandelt haben, betrachten wir jetzt

7. *Die Umkehrfunktionen der trigonometrischen Funktionen:*

$$y = \arcsin x, \quad y = \arccos x,$$
$$y = \arctan x, \quad y = \operatorname{arccot} x$$
$$(y = \operatorname{arcsec} x, \quad y = \operatorname{arccosec} x).$$

[1]) Später (Nr. 83) werden wir noch einmal auf das Problem der Existenz und der Eindeutigkeit der Umkehrfunktion zurückkommen.

Wir behandeln zunächst die erste von ihnen. Die Funktion $y = \sin x$ ist im Intervall $\mathcal{X} = (-\infty, \infty)$ definiert, die Funktionswerte füllen das Intervall $\mathcal{Y} = [-1, 1]$ aus. Jede Parallele zur x-Achse (im Abstand $\leqq 1$) schneidet die Sinuskurve, d. h. die graphische Darstellung der Funktion $y = \sin x$ (Abb. 15), in unendlich vielen Punkten; mit anderen Worten, jedem Wert y aus dem Intervall $[-1, 1]$ entsprechen unendlich viele x-Werte. Daher ist die inverse Funktion, die mit

$$x = \operatorname{Arcsin} y$$

bezeichnet wird[1]), (unendlich) vieldeutig.

Gewöhnlich betrachtet man nur den einen Zweig dieser Funktion, bei dem die Veränderliche x zwischen $-\frac{\pi}{2}$ und $\frac{\pi}{2}$ liegt; jedem y aus $[-1, 1]$ entspricht in diesen Grenzen ein einziger x-Wert; er wird mit

$$x = \arcsin y$$

bezeichnet und *Hauptwert des Arkussinus* genannt.

Wenn wir die Sinuskurve um die Winkelhalbierende des ersten Quadranten klappen (Abb. 20), so erhalten wir die graphische Darstellung der mehrdeutigen Funktion $y = \operatorname{Arcsin} x$; der ausgezogene Teil der Kurve gibt den Hauptzweig $y = \arcsin x$ an, der im x-Intervall $[-1, 1]$ eindeutig definiert ist; die y-Werte genügen dabei der Ungleichung

$$-\frac{\pi}{2} \leqq \arcsin x \leqq \frac{\pi}{2}.$$

die den Hauptzweig unter den anderen Zweigen charakterisiert.

Aus der elementaren Trigonometrie ist bekannt, wie sich alle Winkel, die denselben Sinus haben, durch einen dieser Werte des Hauptzweigs ausdrücken lassen; daher können wir die Formel, die alle Werte des Arkussinus liefert, gleich hinschreiben:

$$\operatorname{Arcsin} x = \arcsin x + 2k\pi \quad (k = 0, \pm 1, \ldots)$$

oder

$$\operatorname{Arcsin} x = (2k + 1)\pi - \arcsin x.$$

Aus dem Additionstheorem für den Sinus,

$$\sin(\alpha + \beta) = \sin\alpha \cos\beta + \cos\alpha \sin\beta,$$

kann man das Additionstheorem für den Arkussinus erhalten. Wir setzen hier $\alpha = \arcsin x$, $\beta = \arcsin y$ (wobei x und y zwischen -1 und 1 liegen) und finden

$$\sin\alpha = x, \qquad \sin\beta = y,$$

$$\cos\alpha = \sqrt{1 - x^2}, \qquad \cos\beta = \sqrt{1 - y^2},$$

wobei die Wurzeln mit dem Pluszeichen zu nehmen sind, weil die Winkel α und β nach der charakteristischen Eigenschaft des Hauptwertes des Arkussinus zwischen $-\frac{\pi}{2}$ und $\frac{\pi}{2}$ liegen, der Kosinus also positiv ist. Somit ist

$$\sin(\alpha + \beta) = x\sqrt{1 - y^2} + y\sqrt{1 - 2^2};$$

[1]) Wir haben schon früher (vgl. Nr. 48, Punkt 5) darauf hingewiesen, daß das Argument x der trigonometrischen Funktionen die Winkel im *Bogenmaß* ausdrückt; selbstverständlich werden auch die Werte der Umkehrfunktionen der trigonometrischen Funktionen, wenn man sie als Maß für den Winkel (oder des Bogens) betrachtet, sämtlich im Bogenmaß ausgedrückt.

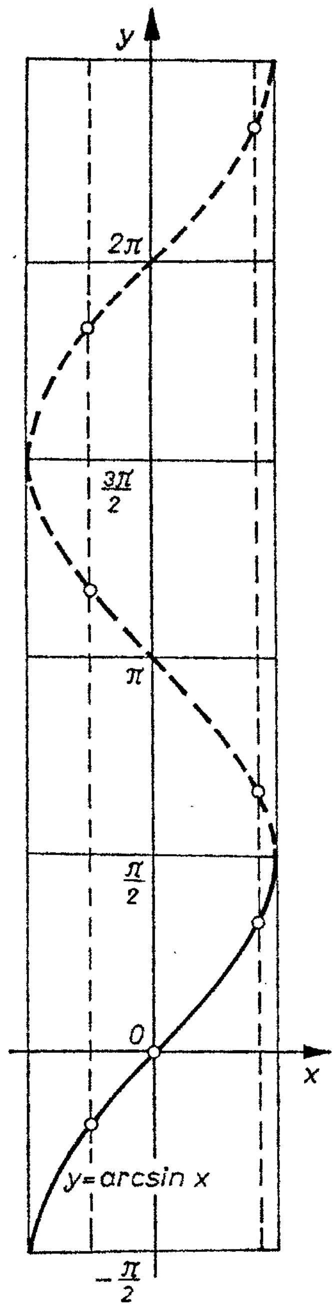

Abb. 20

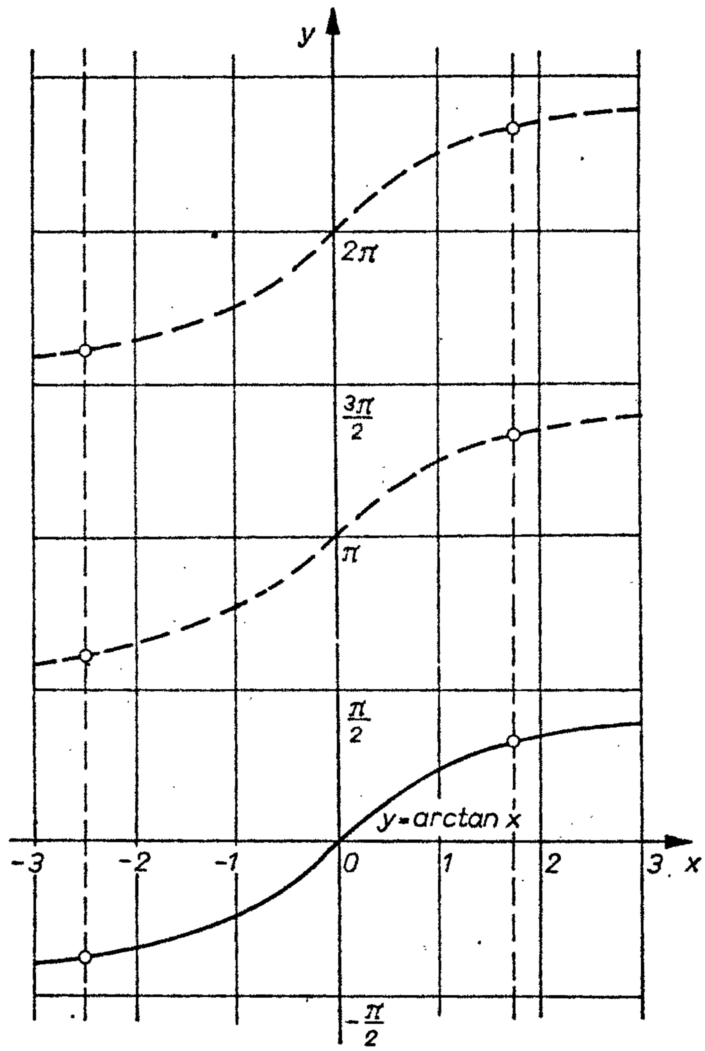

Abb. 21

hieraus folgt

$$\alpha + \beta = \arcsin x + \arcsin y = \operatorname{Arcsin}\left(x\sqrt{1-y^2} + y\sqrt{1-x^2}\right).$$

Die Formel kann man für den Fall, daß auch $\alpha + \beta$ innerhalb des Intervalls $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ liegt, einfacher so schreiben:

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

Diese Bedingung ist automatisch erfüllt, wenn die Argumente x und y (und mit ihnen α und β) verschiedene Vorzeichen haben. Für den Fall gleicher Vorzeichen ist, wie man leicht sieht, diese Bedingung mit $x^2 + y^2 \leqq 1$ äquivalent.

Ähnliche Überlegungen kann man auf die Funktion $y = \cos x$ $(-\infty < x < \infty)$ anwenden. Auch hier stellt sich heraus, daß die inverse Funktion

$$y = \operatorname{Arccos} x \quad (-1 \leqq x \leqq 1)$$

(unendlich) vieldeutig ist (vgl. Abb. 15). Um einen eindeutigen Zweig auszuwählen, unterwerfe man sie der Bedingung

$$0 \leqq \arccos x \leqq \pi;$$

das ist der Hauptzweig des Arkuskosinus.

Die Funktion $\arccos x$ hängt mit $\arcsin x$ offenbar durch die Beziehung

$$\arccos x = \frac{\pi}{2} - \arcsin x$$

zusammen; in der Tat ist nicht nur der Kosinus des Winkels $\frac{\pi}{2} - \arcsin x$ gleich $\sin(\arcsin x) = x$, sondern es liegt auch der Winkel selbst zwischen 0 und π. Die übrigen Werte von $\operatorname{Arccos} x$ lassen sich mit Hilfe seines Hauptwertes durch die Formel

$$\operatorname{Arccos} x = 2k\pi \pm \arccos x \quad (k = 0, \pm 1, \pm 2, \ldots)$$

ausdrücken.

Die Funktion $y = \tan x$ ist für alle x, mit Ausnahme der Werte $x = (2k+1)\frac{\pi}{2}$ $(k = 0, \pm 1, \pm 2, \ldots)$, definiert. Die y-Werte füllen das Intervall $(-\infty, \infty)$ aus, wobei jedem y wieder unendlich viele x-Werte entsprechen (vgl. Abb. 16). Daher ist die Funktion $x = \operatorname{Arctan} y$, die im Intervall $(-\infty, \infty)$ gegeben ist, (unendlich) vieldeutig. In Abb. 21 ist die Funktion $y = \operatorname{Arctan} x$ graphisch dargestellt, wie man sie aus der Kurve $y = \tan x$ durch Umklappen um 180° um die Winkelhalbierende des ersten Quadranten erhält. Als *Hauptwert* des Arkustangens, $\arctan x$, nimmt man den Zweig der Werte dieser mehrdeutigen Funktion, die der Ungleichung

$$-\frac{\pi}{2} < \arctan x < \frac{\pi}{2}$$

genügen.

Dadurch läßt sich eine eindeutige Funktion, der Hauptwert des Arkustangens, für alle x-Werte definieren. Die übrigen Werte des Arkustangens erhält man offenbar aus

$$\operatorname{Arctan} x = \arctan x + k\pi \quad (k = 0, \pm 1, \pm 2, \ldots).$$

Das Additionstheorem für den Tangens lautet:

$$\tan(\alpha + \beta) = \frac{\tan\alpha + \tan\beta}{1 - \tan\alpha \cdot \tan\beta}.$$

Wird $\alpha = \arctan x$, $\beta = \arctan y$ gesetzt, so ergibt sich (für $xy \neq 1$)

$$\tan(\alpha + \beta) = \frac{x + y}{1 - xy},$$

so daß

$$\alpha + \beta = \arctan x + \arctan y = \operatorname{Arctan} \frac{x + y}{1 - xy}$$

ist. Auch hier reduziert sich die Gleichung nur dann auf die einfache Form

$$\arctan x + \arctan y = \arctan \frac{x+y}{1-xy},$$

wenn $-\frac{\pi}{2} < \alpha + \beta < \frac{\pi}{2}$, d. h., wenn $xy < 1$ ist.

Man kann auch leicht den direkten Zusammenhang zwischen den Funktionen $\arctan x$ und $\arcsin x$ herleiten:

$$\arctan x = \arcsin \frac{x}{\sqrt{1+x^2}} \quad (-\infty < x < \infty)$$

oder

$$\arcsin x = \arctan \frac{x}{\sqrt{1-x^2}} \quad (-1 < x < 1).$$

Wenn wir z. B. $\alpha = \arctan x$ setzen, so daß $\tan \alpha = x$ ist, so wird

$$\sin \alpha = \frac{\tan \alpha}{\sqrt{1+\tan^2 \alpha}} = \frac{x}{\sqrt{1+x^2}},$$

wobei die Wurzel mit positivem Vorzeichen zu nehmen ist, weil $-\frac{\pi}{2} < \alpha < \frac{\pi}{2}$ ist; hieraus folgt gerade

$$\alpha = \arcsin \frac{x}{\sqrt{1+x^2}}.$$

Wir erwähnen noch die Funktion $\operatorname{Arccot} x$ $(-\infty < x < \infty)$; ihr Hauptwert ist durch die Ungleichung

$$0 < \operatorname{arccot} x < \pi$$

definiert und hängt mit $\arctan x$ durch die Beziehung

$$\operatorname{arccot} x = \frac{\pi}{2} - \arctan x$$

zusammen. Die übrigen Werte des Arkuskotangens haben die Form

$$\operatorname{Arccot} x = \operatorname{arccot} x + k\pi \quad (k = 0, \pm 1, \pm 2, \ldots).$$

Auf die Funktionen $\operatorname{arcsec} x$ $(-\infty < x \leqq -1$ und $1 \leqq x < \infty)$ und $\operatorname{arccosec} x$ (im gleichen Definitionsbereich) wollen wir hier nicht eingehen, sondern die Überlegungen dem Leser überlassen.

51. Verkettung von Funktionen. Schlußbemerkungen. Wir beschäftigen uns jetzt mit dem Begriff der Verkettung von Funktionen. Darunter versteht man folgendes: An Stelle des Arguments einer gegebenen Funktion wird eine Funktion eines anderen Arguments eingesetzt. Beispielsweise liefert die Verkettung der Funktionen $y = \sin x$ und $z = \log y$ die Funktion $z = \log \sin x$; analog erhält man die Funktion

$$\sqrt{1-x^2}, \quad \arctan \frac{1}{x} \quad \text{usw.}$$

Man spricht hier auch einfach von *zusammengesetzten* oder *mittelbaren Funktionen*.

Im allgemeinen Fall nehmen wir an, die Funktion $z = \varphi(y)$ sei in einem Gebiet $\mathcal{Y} = \{y\}$ definiert, die Funktion $y = f(x)$ in einem Gebiet $\mathcal{X} = \{x\}$, und ihre Werte seien sämtlich im Gebiet $\mathcal{Y}$ enthalten. Dann wird die Veränderliche z, wie man sagt, durch Vermittlung von y selbst eine Funktion von x:

$$z = \varphi\big(f(x)\big).$$

Zu gegebenem x aus $\mathcal{X}$ findet man zunächst den (nach der Vorschrift für f) entsprechenden Wert y aus $\mathcal{Y}$, und dann ermittelt man den diesem y-Wert (nach der Vorschrift für φ) entsprechenden Wert z; diesen nimmt man dann als den Wert, der durch die Verkettung der Funktionen φ und f dem Wert x zugeordnet ist. Diese so entstehende Funktion einer Funktion oder zusammengesetzte Funktion ist das Ergebnis der Verkettung der Funktionen $f(x)$ und $\varphi(x)$.

Die Voraussetzung, daß die Werte der Funktion $f(x)$ nicht außerhalb des Bereichs $\mathcal{Y}$ liegen, in dem die Funktion $\varphi(x)$ definiert ist, ist sehr wesentlich: Wenn sie weggelassen wird, so kann man auch sinnlose Ausdrücke erhalten. Wenn wir etwa $z = \log y$ und $y = \sin x$ annehmen, so können wir nur solche x-Werte betrachten, für die $\sin x > 0$ ist, weil der Ausdruck $\log \sin x$ sonst sinnlos werden würde.

Wir halten es für zweckmäßig, hier besonders darauf hinzuweisen, daß der Begriff der mittelbaren Funktion nichts mit der Art und Weise zu tun hat, in der z von x wirklich funktional abhängt, sondern nur mit der Art und Weise, in der diese Abhängigkeit analytisch ausgedrückt wird. Es sei etwa $z = \sqrt{1 - y^2}$ für y aus $[-1, 1]$ und $y = \sin x$ für x aus $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. Dann ist

$$z = \sqrt{1 - \sin^2 x} = \cos x.$$

Hier erscheint die Funktion $\cos x$ als mittelbare Funktion.

Nachdem wir jetzt auch den Begriff der Verkettung von Funktionen vollständig erklärt haben, können wir die einfachsten derjenigen Funktionenklassen charakterisieren, die in der Analysis untersucht werden: Das sind zunächst die oben angegebenen elementaren Funktionen 1 bis 7 und dann alle diejenigen, die man aus ihnen mit Hilfe der vier Grundrechenarten und der Verkettung erhält, wenn man diese Operationen endlich oft anwendet. Von ihnen sagt man, sie könnten durch elementare Funktionen *in geschlossener Form* ausgedrückt werden; manchmal nennt man auch die ganze Klasse der so erhaltenen Funktionen elementar.

Später, wenn wir kompliziertere analytische Hilfsmittel (unendliche Reihen, Integrale) zur Verfügung haben, werden wir auch andere Funktionen kennenlernen, die ebenfalls in der Analysis eine wichtige Rolle spielen, aber nicht mehr zur Klasse der elementaren Funktionen gehören.

§ 2. Grenzwert einer Funktion

52. Definition des Grenzwertes einer Funktion. Wir betrachten eine Zahlenmenge $\mathcal{X} = \{x\}$. Ein Punkt a heißt *Häufungspunkt* dieser Menge (man sagt auch „für diese Menge"), wenn in beliebiger Nähe des Punktes Werte x aus $\mathcal{X}$ liegen, die von a verschieden sind.

Um diese Definition genauer formulieren zu können, führen wir den Begriff der Umgebung eines Punktes a ein: Unter einer *δ-Umgebung* von a verstehen wir das offene Intervall $(a - \delta, a + \delta)$ mit dem Mittelpunkt a. Jetzt können wir sagen, ein Punkt a sei *Häufungspunkt* einer Menge $\mathfrak{X}$, wenn in jeder Umgebung des Punktes a von a verschiedene Punkte x aus $\mathfrak{X}$ liegen.

Dabei kann der Häufungspunkt selbst zu $\mathfrak{X}$ gehören oder auch nicht.

Es sei in einem Bereich $\mathfrak{X}$, für den a Häufungspunkt ist, eine Funktion $f(x)$ gegeben. Uns interessiert das Verhalten dieser Funktion, wenn sich x unbegrenzt dem Punkt a nähert. Man sagt, die Funktion $f(x)$ habe für gegen a strebendes x (oder im Punkt a) den Grenzwert A, wenn zu jedem $\varepsilon > 0$ ein $\delta > 0$ existiert derart, daß

$$|f(x) - A| < \varepsilon \quad \text{gilt, sobald} \quad |x - a| < \delta \quad \text{ist} \tag{1}$$

(dabei ist x aus $\mathfrak{X}$ und von a verschieden).[1]) Diesen Sachverhalt beschreibt man durch

$$\lim_{x \to a} f(x) = A. \tag{2}$$

Ist der Bereich $\mathfrak{X}$ so beschaffen, daß in beliebiger Nähe von a (aber nur rechts von a) von a verschiedene Werte x aus $\mathfrak{X}$ existieren (in diesem Fall heißt der Punkt a *rechtsseitiger Häufungspunkt* von $\mathfrak{X}$), so kann man die obige Definition des Grenzwertes einer Funktion spezialisieren, nämlich durch die Einschränkung $x > a$. In diesem Fall nennt man den Grenzwert der Funktion, wenn er existiert, *Grenzwert der Funktion $f(x)$ für von rechts gegen a strebendes x*, oder kürzer, *rechtsseitigen Grenzwert* (im Punkt a) und schreibt

$$\lim_{x \to a+0} f(x) \quad \text{oder} \quad f(a + 0).^{2)}$$

Analog kann man den Begriff des *linksseitigen Häufungspunktes* und des *Grenzwertes einer Funktion für von links gegen a strebendes x* bzw. des *linksseitigen Grenzwertes* (im Punkt a) einführen:

$$\lim_{x \to a-0} f(x) \quad \text{oder} \quad f(a - 0).^{2)}$$

Auch die Schreibweisen $\lim\limits_{x \downarrow a}$ bzw. $\lim\limits_{x \uparrow a}$ sind gebräuchlich.

Ist a gleichzeitig rechtsseitiger und linksseitiger Häufungspunkt von x, so ist, wie sich leicht zeigen läßt, für die Existenz des Grenzwertes (2) notwendig und hinreichend, daß die Grenzwerte von rechts und von links einzeln existieren und übereinstimmen:

$$\lim_{x \to a+0} f(x) = \lim_{x \to a-0} f(x) = A.$$

Wenn x gegen einen endlichen Grenzwert a strebt, so kann die Funktion auch einen unendlichen Grenzwert haben. Die Funktion $f(x)$ hat für gegen a strebendes x den Grenzwert ∞ (bzw. $-\infty$), wenn zu jeder Zahl $E > 0$ eine Zahl $\delta > 0$ existiert derart, daß

$$f(x) > E \quad (\text{bzw. } f(x) < -E) \quad \text{gilt, sobald} \quad |x - a| < \delta \tag{3}$$

ist (wie stets gehöre x zu $\mathfrak{X}$ und sei von a verschieden).

---

[1]) Da a Häufungspunkt für den Bereich $\mathfrak{X}$ ist, gibt es sicher solche Werte x in der Umgebung $(a - \delta, a + \delta)$ des Punktes a.

[2]) Ist $a = 0$, so schreibt man statt $0 + 0$ (bzw. $0 - 0$) einfach $+0$ (bzw. -0).

Analog (2) schreibt man

$$\lim_{x\to a} f(x) = \infty \quad (\text{bzw. } -\infty).$$

Auch in diesem Fall könnten wir die für rechtsseitige bzw. linksseitige (einseitige) Grenzwerte gemachten Bemerkungen wiederholen.

Enthält eine Menge $\mathfrak{X} = \{x\}$ (dem Absolutbetrag nach) beliebig große positive (negative) Werte x, so sagt man, ∞ (bzw. $-\infty$) sei *Häufungspunkt* für (von) $\mathfrak{X}$.

Unter dieser Annahme gilt: *Eine Funktion $f(x)$ hat für gegen ∞* (bzw. $-\infty$) *strebendes x den Grenzwert A, wenn zu jedem $\varepsilon > 0$ ein $\Delta > 0$ existiert derart, daß*

$$|f(x) - A| < \varepsilon \quad \textit{ist, sobald} \quad x > \Delta \quad (\textit{bzw. } x < -\Delta) \tag{4}$$

gilt (*x gehört zu $\mathfrak{X}$*). Hierfür schreibt man:

$$\lim_{x\to\infty} f(x) = A \quad \text{bzw.} \quad \lim_{x\to-\infty} f(x) = A.$$

Schließlich ist auch hier alles Gesagte leicht auf den Fall $A = \infty$ bzw. $A = -\infty$ zu übertragen.

Alle diese Definitionen haben eines gemeinsam: Die Funktion $f(x)$ muß ihrem Grenzwert A beliebig nahekommen, sobald die Veränderliche x ihrem Grenzwert a hinreichend nahekommt. Eine Veränderliche kommt aber einem endlichen Grenzwert beliebig nahe, wenn die Differenz zwischen ihr und diesem Grenzwert (absolut genommen) unendlich klein ist, und einem unendlichen Grenzwert, wenn sie selbst (absolut genommen) unendlich groß ist und außerdem dasselbe Vorzeichen wie der Grenzwert hat.

Die Zahl δ (bzw. Δ) hängt offenbar von ε (bzw. E) ab.

Schließlich sagt man, eine Funktion $f(x)$, die gegen 0 strebt, werde *unendlich klein*; sie wird *unendlich groß* genannt, wenn $|f(x)|$ gegen ∞ strebt. Gilt dies für $x \to a$, so sagt man, im Punkt a werde die Funktion ∞.

53. Zurückführung auf den Fall einer diskreten Veränderlichen. Betrachten wir eine diskrete Veränderliche als Funktion der unabhängigen Veränderlichen n, die die Folge der natürlichen Zahlen durchläuft, so stimmt der in Nr. 52 definierte Grenzwert dieser Funktion für $n \to \infty$ offenbar mit dem Grenzwert der diskreten Veränderlichen, wie er in Nr. 23 und Nr. 27 definiert wurde, überein (die Rolle von Δ spielt dort N). *Somit ist der Grenzwert einer diskreten Veränderlichen ein Spezialfall des Grenzwertes einer Funktion.*

Jedoch kann im gewissen Sinne auch umgekehrt der Grenzwert einer Funktion auf den Grenzwert einer diskreten Veränderlichen zurückgeführt werden.

Die Menge $\mathfrak{X} = \{x\}$ habe den Häufungspunkt a (hierbei kann a endlich oder auch ∞ oder $-\infty$ sein). Dann kann man auf $\mathfrak{X}$ (auf unendlich viele Arten) eine Folge

$$x_1, x_2, x_3, \ldots, x_n, \ldots \tag{6}$$

von x-Werten (die von a verschieden sind) auswählen, welche den Grenzwert a hat.

In der Tat, ist a endlich, so läßt sich, wenn eine gegen 0 strebende Folge positiver Zahlen δ_n vorgegeben ist, in jeder Umgebung $(a - \delta_n, a + \delta_n)$ $(n = 1, 2, 3, \ldots)$ des Punktes a ein von a verschiedener Punkt $x = x_n$ aus $\mathfrak{X}$ finden. Wegen $|x_n - a| < \delta_n$ gilt $x_n \to a$. Für $a = \infty$ (bzw. $-\infty$) gibt es eine unbegrenzt wachsende Folge positiver Zahlen $\Delta_n \to \infty$, und zu jedem Δ_n bestimmen wir einen Wert $x = x_n$ aus $\mathfrak{X}$, für den $x_n > \Delta_n$ (bzw. $x_n < -\Delta_n$) gilt; offenbar strebt x_n gegen ∞ (bzw. $-\infty$), usw.

Der Folge der Argumente (6) entspricht die Folge der Funktionswerte

$$f(x_1), f(x_2), f(x_3), \ldots, f(x_n), \ldots \tag{7}$$

Wir werden sogleich sehen, daß *diese Folge immer den Grenzwert A hat, wenn die Beziehung* (2) *vorliegt.* Als Beispiel behandeln wir den Fall, daß a und A endlich sind.

Ist eine beliebige Zahl $\varepsilon > 0$ gegeben, so nehmen wir zuerst ein solches $\delta > 0$, das nach Definition des Grenzwertes (2) der Zahl ε entspricht. Zur Zahl δ läßt sich auf Grund der Konvergenz der Folge (6) gegen a eine Zahl N finden (Nr. 23) derart, daß für $n > N$ die Ungleichung

$$|x_n - a| < \delta,$$

also [vgl. (1)] auch die Ungleichung

$$|f(x_n) - A| < \varepsilon$$

gilt. Damit ist die Konvergenz der Folge (7) gegen A bewiesen.

Nun wollen wir beweisen, daß auch die umgekehrte Aussage gilt:

Wir nehmen also an, für jede irgendwie gewählte Folge (6) von Werten (aus $\mathcal{X}$) der unabhängigen Veränderlichen x, die den Grenzwert a hat, strebe die entsprechende Folge (7) von Funktionswerten gegen A. *Dann ist diese Zahl A der Grenzwert der Funktion $f(x)$ im Sinne der Definition aus Nr. 52.*

Wir beschränken uns auch hier auf den Fall, daß a und A endlich sind. Den Beweis führen wir indirekt, nehmen also an, A sei nicht Grenzwert der Funktion in dem erwähnten Sinne. Dann gäbe es zu einem bestimmten $\varepsilon > 0$ kein entsprechendes δ, d. h., wie klein man auch δ wählen würde, stets könnte man einen von a verschiedenen Wert x' der Veränderlichen x finden, für den zwar

$$0 < |x' - a| < \delta, \quad \text{aber} \quad |f(x') - A| \geqq \varepsilon$$

wäre. Wir wählen nun eine Folge $\{\delta_n\}$ positiver Zahlen, die gegen 0 strebt. Auf Grund unserer Annahme könnte man zu jeder Zahl $\delta = \delta_n$ einen Wert $x' = x_n'$ finden derart, daß zwar

$$|x_n' - a| < \delta_n, \quad \text{aber} \quad |f(x_n') - A| \geqq \varepsilon$$

wäre. Aus diesen Werten könnte man somit eine Folge $x_1', x_2', x_3', \ldots, x_n', \ldots$ bilden, für die

$$|x_n' - a| < \delta_n \quad (n = 1, 2, 3, \ldots)$$

gilt; wegen $\delta_n \to 0$ würde also x_n' gegen a streben. Dann müßte die entsprechende Folge der Funktionswerte $f(x_1'), f(x_2'), f(x_3'), \ldots, f(x_n'), \ldots$ gegen A streben, da dies nach der Voraussetzung des Satzes für jede Folge von Funktionswerten der Fall ist, deren Argumente gegen a streben. Das widerspricht aber der aus unserer Annahme gefolgerten Tatsache, daß für alle $n = 1, 2, 3, \ldots$ die Beziehung $|f(x_n') - A| \geqq \varepsilon$ gilt. Durch diesen Widerspruch ist unsere Behauptung bewiesen.

Somit kommen wir im Grunde zu einer zweiten Definition des in Nr. 52 in der sogenannten „Epsilontik" („ε-δ-Sprache") eingeführten Begriffs Grenzwert einer Funktion. Wir können jetzt diesen Begriff in der „Sprache der Folgen" formulieren, indem wir nämlich die Formel (2) in dem Sinne verstehen, daß für jede Folge (6), die gegen a strebt, die entsprechende Folge (7) den Grenzwert A hat.

Zum Schluß sei noch bemerkt, daß man nur die Existenz eines Grenzwertes einer jeden Folge (7), die irgendeiner gegen a konvergierenden Folge (6) entspricht, vorauszusetzen braucht (und nicht auch die Übereinstimmung dieser Grenzwerte); daraus ergibt sich schon von selbst, daß alle diese Grenzwerte übereinstimmen. In der Tat, nehmen wir einmal an, für die beiden gegen a strebenden Folgen $x_1', x_2', \ldots, x_n', \ldots$ und $x_1'', x_2'', \ldots, x_n'', \ldots$ würde

$$f(x_n') \to A' \quad \text{und} \quad f(x_n'') \to A''$$

gelten, und es sei $A' \neq A''$. Durch Vereinigung der beiden Folgen läßt sich die neue Folge $x_1', x_1'', x_2', x_2'', \ldots, x_n', x_n'', \ldots$ konstruieren, die offenbar gegen a strebt, weil für hinreichend große n sich sowohl x_n' als auch x_n'' beliebig wenig von a unterscheiden. Die entsprechende Folge der Funktionswerte $f(x_1')$, $f(x_1'')$, $f(x_2')$, $f(x_2''), \ldots, f(x_n')$, $f(x_n''), \ldots$ könnte aber entgegen der Voraussetzung keinen Grenzwert besitzen, weil ihre Teilfolgen aus Gliedern mit geraden bzw. ungeraden Indizes gegen verschiedene Grenzwerte streben (Nr. 40). Dieser so erhaltene Widerspruch beweist gerade, daß die Folgen der Form (7) sämtlich gegen ein und denselben Grenzwert streben.

54. Beispiele.

1. Wir beweisen, daß für $a > 1$

$$\lim_{x\to\infty} a^x = \infty$$

gilt.

Für beliebiges $E > 0$ genügt es, $\Delta = \log_a E$ zu setzen, damit aus $x > \Delta$ die Beziehung $a^x > E$ folgt, womit unsere Behauptung bewiesen ist.[1])

Analog kann man

$$\lim_{x\to-\infty} a^x = 0 \qquad (\text{für } a > 1)$$

beweisen. In der Tat, wie auch immer $\varepsilon > 0$ $(\varepsilon < 1)$ gewählt sei, für $\Delta = \log_a \frac{1}{\varepsilon} = -\log_a \varepsilon$ folgt aus $x < -\Delta$ die Beziehung $a^x < \varepsilon$.

Ist aber $0 < a < 1$, so erhält man mit Hilfe der Transformation $a^x = \left(\frac{1}{a}\right)^{-x}$ leicht

$$\lim_{x\to\infty} a^x = 0, \qquad \lim_{x\to-0} a^x = \infty.$$

2. Für $a > 1$ gilt

$$\lim_{x\to\infty} \log_a x = \infty, \qquad \lim_{x\to+0} \log_a x = -\infty.$$

Für beliebig gegebenes $E > 0$ gilt bei $x > a^E$ die Beziehung $\log_a x > E$, und analog ist für $0 < x < a^{-E}$ die Ungleichung $\log_a x < -E$ erfüllt. Damit sind die beiden Beziehungen bewiesen.

3. Es gilt

$$\lim_{x\to\infty} \arctan x = \frac{\pi}{2}, \qquad \lim_{x\to-\infty} \arctan x = -\frac{\pi}{2}.$$

Wir beschränken uns auf den Beweis der ersten Grenzwertbeziehung. Für jedes $\varepsilon > 0$ genügt es, $x > \tan\left(\frac{\pi}{2} - \varepsilon\right)$ zu nehmen, damit $\arctan x > \frac{\pi}{2} - \varepsilon$ ist; dann ist aber

$$0 < \frac{\pi}{2} - \arctan x < \varepsilon.$$

---

[1]) Das speziellere Ergebnis (für natürliches n) $\lim a^n = \infty$ für $a > 1$ hatten wir schon in Nr. 27 erhalten.

4. Schärfer ist die Beziehung

$$\lim_{x \to \infty} \frac{a^x}{x} = \infty \quad (\text{für } a > 1).$$

Den Spezialfall (für natürliches n)

$$\lim_{n \to \infty} \frac{a^n}{n} = \infty$$

haben wir schon in Nr. 32, Beispiel 9, kennengelernt. Offenbar ist auch $\lim\limits_{n \to \infty} \frac{a^n}{n+1} = \infty$. Folglich kann man zu gegebenem $E > 0$ eine natürliche Zahl N finden derart, daß für $n > N$ die Ungleichung $\frac{a^n}{n+1} > E$ erfüllt ist.

Es sei jetzt $x > N + 1$; setzt man $n = [x]$, so ist $n > N$ und $n \leqq x < n + 1$, also

$$\frac{a^x}{x} > \frac{a^n}{n+1} > E,$$

und damit ist unsere Behauptung bewiesen.

Hieraus ergibt sich wie in Nr. 32, Beispiel 9, leicht

$$\lim_{x \to \infty} \frac{a^x}{x^k} = \infty \qquad (a > 1, k > 0).$$

5. Analog finden wir unter Benutzung der in Nr. 32, Beispiel 11, hergeleiteten Beziehung

$$\lim_{n \to \infty} \frac{\log_a n}{n} = 0 \qquad (a > 1)$$

allgemein

$$\lim_{n \to \infty} \frac{\log_a x}{x} = 0 \qquad (a > 1),$$

wenn x beliebige positive reelle Werte durchläuft. Ersetzt man hier x durch x^k $(k > 0)$, so kann man leicht zeigen, daß auch

$$\lim_{x \to \infty} \frac{\log_a x}{x^k} = 0 \qquad (a > 1, k > 0)$$

ist. In der Tat, ist $\varepsilon > 0$ beliebig vorgegeben, so wähle man Δ so, daß für $x > \Delta$ die Ungleichung $\frac{\log_a x}{x} < k\varepsilon$ erfüllt ist; für $x > \Delta_1 = \Delta^{1/k}$ ist dann $x^k > \Delta$ und $\frac{\log_a x}{x^k} < \varepsilon$. Ersetzt man hier x durch $\frac{1}{x}$, so kann man dieses Resultat in der Form

$$\lim_{x \to +0} x^k \log_a x = 0 \qquad (a > 1, k > 0)$$

schreiben.

6. Aus der in Nr. 25, Beispiel 5, bewiesenen Limesbeziehung $\lim\limits_{n \to \infty} a^{1/n} = 1$ kann man die allgemeinere Beziehung

$$\lim_{x \to 0} a^x = 1$$

erhalten.

Wir bemerken, daß offenbar auch

$$\lim_{n \to \infty} a^{-1/n} = \lim_{n \to \infty} \frac{1}{a^{1/n}} = 1$$

gilt. Daher läßt sich zu beliebigem $\varepsilon > 0$ eine natürliche Zahl n_0 finden derart, daß (etwa für $a > 1$)

$$1 - \varepsilon < a^{-1/n_0} < a^{1/n_0} < 1 + \varepsilon$$

gilt. Ist jetzt

$$|x| < \frac{1}{n_0} \quad \text{oder} \quad -\frac{1}{n_0} < x < \frac{1}{n_0},$$

so gilt

$$a^{-1/n_0} < a^x < a^{1/n_0};$$

hieraus folgt $1 - \varepsilon < a^x < 1 + \varepsilon$ oder $|a^x - 1| < \varepsilon$, was zu beweisen war.

7. Jetzt leiten wir folgende (auch für später wichtige) Beziehung her:

$$\lim_{x \to 0} \frac{\sin x}{x} = 1. \tag{8}$$

Vorher müssen wir jedoch die nützlichen Ungleichungen

$$\sin x < x < \tan x \quad \left(0 < x < \frac{\pi}{2}\right) \tag{9}$$

beweisen.

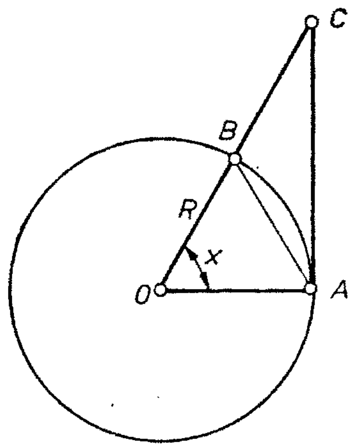

Abb. 22

Zu diesem Zweck betrachten wir in einem Kreis mit dem Radius R einen spitzen Winkel $\sphericalangle AOB$, die Sehne AB und die Tangente AC an den Kreis im Punkt A (Abb. 22). Dann gilt für den Flächeninhalt der Figuren: $\triangle AOB <$ Sektor AOB $< \triangle AOC$.[1])

Es sei x das Bogenmaß des Winkels AOB; dann ist die Länge des Bogens $\overset{\frown}{AB}$ gleich Rx. Daher kann man die Ungleichung für die Flächeninhalte in der Gestalt

$$\frac{1}{2} R^2 \cdot \sin x < \frac{1}{2} R^2 \cdot x < \frac{1}{2} R^2 \cdot \tan x$$

schreiben. Kürzt man durch $\frac{1}{2} R^2$, so erhält man hieraus die Ungleichung (9). Für $0 < x < \frac{\pi}{2}$ ist $\sin x$ positiv; daher können wir jedes der Glieder der Ungleichung (9) durch $\sin x$ dividieren. Wir erhalten $1 > \frac{\sin x}{x} > \cos x$ und hieraus

$$0 < 1 - \frac{\sin x}{x} < 1 - \cos x.$$

[1]) Dabei benutzen wir die aus der Schule bekannten Sätze über die Flächeninhalte elementarer Figuren.

Nun ist aber $1 - \cos x = 2 \sin^2 \frac{x}{2} < 2 \sin \frac{x}{2}$, und nach (9) gilt $2 \sin \frac{x}{2} < x$; folglich ist

$$0 < 1 - \frac{\sin x}{x} < x.$$

Hieraus ersibt sich die Ungleichung

$$\left|\frac{\sin x}{x} - 1\right| < |x|,$$

die offenbar richtig bleibt, wenn man x durch $-x$ ersetzt. Somit gilt

$$\left|\frac{\sin x}{x} - 1\right| < |x| \quad \text{für} \quad 0 < |x| < \frac{\pi}{2}.$$

Diese Ungleichung löst das Problem. In der Tat, ist $\varepsilon > 0$ beliebig gegeben, so wähle man für δ die kleinere der Zahlen ε und $\frac{\pi}{2}$. Für $|x| < \delta$ ist zunächst diese Ungleichung anwendbar $\left(\text{weil } \delta \leqq \frac{\pi}{2} \text{ ist}\right)$, und wegen $\delta \leqq \varepsilon$ ist

$$\left|\frac{\sin x}{x} - 1\right| < \varepsilon.$$

Nach Definition des Grenzwertes einer Funktion (Nr. 52) bedeutet das gerade, daß die Funktion $\frac{\sin x}{x}$ für $x \to 0$ den Grenzwert 1 hat, und damit ist die Beziehung (8) bewiesen.

7a) Die Limesbeziehung (8) kann nach Nr. 53 auch folgendermaßen verstanden werden: Falls x eine gegen 0 konvergierende Folge $\{x_n\}$ durchläuft, so strebt die Folge $\left\{\frac{\sin x_n}{x_n}\right\}$ gegen 1.

Wir benutzen diese Bemerkung, um den Grenzwert der diskreten Veränderlichen

$$\cos \frac{\varphi}{2} \cdot \cos \frac{\varphi}{2^2} \cdots \cos \frac{\varphi}{2^n}$$

für $n \to \infty$ zu bestimmen, wobei φ eine beliebige von 0 verschiedene Zahl ist. Offenbar ist

$$\sin \varphi = 2 \cos \frac{\varphi}{2} \cdot \sin \frac{\varphi}{2} = 2^2 \cos \frac{\varphi}{2} \cdot \cos \frac{\varphi}{2^2} \cdot \sin \frac{\varphi}{2^2}$$

$$= \cdots = 2^n \cos \frac{\varphi}{2} \cdot \cos \frac{\varphi}{2^2} \cdots \cos \frac{\varphi}{2^n} \cdot \sin \frac{\varphi}{2^n},$$

so daß der uns interessierende Ausdruck in der Form

$$\frac{\sin \varphi}{2^n \cdot \sin \frac{\varphi}{2^n}} = \frac{\sin \varphi}{\varphi} \cdot \frac{\frac{\varphi}{2^n}}{\sin \frac{\varphi}{2^n}}$$

geschrieben werden kann. Wegen $x_n = \frac{\varphi}{2^n} \to 0$ ist nach dem oben Bewiesenen

$$\lim \frac{\sin \frac{\varphi}{2^n}}{\frac{\varphi}{2^n}} = 1;$$

daher gilt

$$\lim_{n\to\infty} \cos\frac{\varphi}{2}\cdot\cos\frac{\varphi}{2^2}\cdots\cos\frac{\varphi}{2^n} = \frac{\sin\varphi}{\varphi}.$$ [1]

8. Wir untersuchen jetzt einen weiteren sehr wichtigen Grenzwert. In Nr. 36 haben wir die Zahl e als Grenzwert einer Folge definiert,

$$e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n. \tag{10}$$

Jetzt leiten wir die allgemeineren Resultate

$$\lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x = e \tag{11}$$

und

$$\lim_{x\to-\infty}\left(1+\frac{1}{x}\right)^x = e \tag{11a}$$

her. Diesmal benutzen wir die Definition des Grenzwertes in der „Sprache der Folgen" (Nr. 53).

Zunächst erinnern wir daran, daß neben (10) auch die Beziehung

$$\lim\left(1+\frac{1}{n_k}\right)^{n_k} = e \tag{12}$$

gilt, wenn $\{n_k\}$ eine beliebige Folge natürlicher Zahlen ist, die mit dem Index k unbegrenzt wachsen (Nr. 40). Jetzt möge x irgendeine Folge $\{x_k\}$ von Werten durchlaufen, die gegen ∞ streben; man kann ferner annehmen, für alle k sei $x_k > 1$. Wir setzen nun $n_k = [x_k]$, so daß

$$n_k \leqq x_k < n_k + 1 \quad \text{und} \quad n_k \to +\infty$$

gilt. Weil dabei $\dfrac{1}{n_k+1} < \dfrac{1}{x_k} \leqq \dfrac{1}{n_k}$ ist, gilt

$$\left(1+\frac{1}{n_k+1}\right)^{n_k} < \left(1+\frac{1}{x_k}\right)^{x_k} < \left(1+\frac{1}{n_k}\right)^{n_k+1}.$$

Die beiden äußeren Ausdrücke können wie folgt dargestellt werden:

$$\left(1+\frac{1}{n_k+1}\right)^{n_k} = \frac{\left(1+\dfrac{1}{n_k+1}\right)^{n_k+1}}{1+\dfrac{1}{n_k+1}},$$

$$\left(1+\frac{1}{n_k}\right)^{n_k+1} = \left(1+\frac{1}{n_k}\right)^{n_k}\cdot\left(1+\frac{1}{n_k}\right),$$

wobei nach (12)

$$\left(1+\frac{1}{n_k}\right)^{n_k} \to e \quad \text{und} \quad \left(1+\frac{1}{n_k+1}\right)^{n_k+1} \to e$$

[1] Damit haben wir den Wert eines „unendlichen Produkts" bestimmt.

gilt, während offenbar

$$1+\frac{1}{n_k}\to 1, \quad 1+\frac{1}{n_k+1}\to 1$$

ist; somit streben die beiden erwähnten Ausdrücke gegen den gemeinsamen Grenzwert e; dann strebt aber auch der dazwischenliegende Ausdruck gegen e (nach Satz 3 aus Nr. 28); also ist

$$\lim\left(1+\frac{1}{x_k}\right)^{x_k}=\mathrm{e}.$$

Damit ist die Beziehung (11) „in der Sprache der Folgen" vollständig bewiesen.

Um (11a) zu beweisen, setzen wir voraus, die Folge $\{x_k\}$ habe den Grenzwert $-\infty$ (wobei man $x_k<-1$ für alle k annehmen kann). Wenn wir $x_k=-y_k$ setzen, so strebt y_k gegen ∞ (und es ist $y_k>1$ für alle k). Offenbar ist

$$\left(1+\frac{1}{x_k}\right)^{x_k}=\left(1-\frac{1}{y_k}\right)^{-y_k}=\left(\frac{y_k}{y_k-1}\right)^{y_k}=\left(1+\frac{1}{y_k-1}\right)^{y_k-1}\cdot\left(1+\frac{1}{y_k-1}\right).$$

Nach dem oben Bewiesenen strebt der erste Faktor des letzten Ausdrucks gegen e, der zweite hat offenbar den Grenzwert 1, so daß auch der links stehende Ausdruck gegen e strebt. Damit ist (11a) bewiesen.

Wir ersetzen jetzt in dem Ausdruck $\left(1+\frac{1}{x}\right)^x$ die Veränderliche x durch $\frac{1}{\alpha}$. Durchläuft α eine Folge positiver oder negativer gegen 0 strebender Werte, so strebt $x=\frac{1}{\alpha}$ gegen $\pm\infty$. Daher kann man (11) und (11a) in der Gestalt

$$\mathrm{e}=\lim_{\alpha\to 0}(1+\alpha)^{1/\alpha} \tag{13}$$

schreiben.

Dieses bemerkenswerte Resultat liegt allen Awendungen der Zahl e *zugrunde.*

9. Interessant ist schließlich auch ein Beispiel, bei dem der Grenzwert einer Funktion nicht existiert: *Die Funktion* $\sin x$ *hat, wenn* x *gegen* ∞ (*bzw.* $-\infty$) *strebt, keinen Grenzwert.*

Davon überzeugt man sich am einfachsten, indem man Folgen untersucht. Es genügt zu bemerken, daß den beiden gegen ∞ strebenden Folgen

$$\left\{\left(2n-\frac{1}{2}\right)\pi\right\} \quad \text{und} \quad \left\{\left(2n+\frac{1}{2}\right)\pi\right\} \qquad (n=1,2,3,\ldots)$$

von x-Werten Folgen von Funktionswerten entsprechen, die gegen verschiedene Grenzwerte streben:

$$\sin\left(2n-\frac{1}{2}\right)\pi=-1\to -1, \qquad \sin\left(2n+\frac{1}{2}\right)\pi=1\to 1.$$

[Dasselbe kann man auch anders ausdrücken: Der gegen ∞ strebenden Folge

$$\left\{\left(n+\frac{1}{2}\right)\pi\right\} \qquad (n=1,2,3,\ldots)$$

von x-Werten entspricht die Folge der Funktionswerte

$$\sin\left(n+\frac{1}{2}\right)\pi=(-1)^n \qquad (n=1,2,3,\ldots),$$

die keinen Grenzwert hat.]

Wenn man sich vor Augen hält, daß die Sinuskurve „oszilliert", so ist auch anschaulich klar, daß kein Grenzwert existieren kann.

Analog hat auch die Funktion $\sin\frac{1}{\alpha}$ für (von rechts oder von links) gegen 0 strebendes α keinen Grenzwert. Das ist im Grunde nur eine andere Form des obigen Beispiels: Man braucht in der Funktion $\sin x$ nur den Wert x durch $\frac{1}{\alpha}$ zu ersetzen. Durchläuft α eine Folge von Werten, die von rechts (bzw. von links) gegen 0 streben, so strebt $x = \frac{1}{\alpha}$ gegen ∞ (bzw. gegen $-\infty$) und umgekehrt.

Wir ersetzen jetzt in dem Ausdruck $\sin\frac{1}{\alpha}$ wieder α durch x (um auf die ursprüngliche Bezeichnung der Abszisse zurückzukommen) und betrachten das lehrreiche Beispiel der graphischen Darstellung der Funktion

$$y = \sin\frac{1}{x} \qquad (x \neq 0),$$

wobei x die Werte zwischen 0 und $\frac{2}{\pi}$ (und zwischen $-\frac{2}{\pi}$ und 0) durchläuft.

Wir betrachten die gegen 0 abnehmenden x-Werte

$$\frac{2}{\pi}, \frac{1}{\pi}, \frac{2}{3\pi}, \frac{1}{2\pi}, \frac{2}{5\pi}, \frac{1}{3\pi}, \frac{2}{7\pi}, \ldots, \frac{2}{(2n-1)\pi}, \frac{1}{n\pi}, \frac{2}{(2n+1)\pi}, \ldots$$

und die entsprechenden Werte von $\frac{1}{x}$, die gegen $+\infty$ wachsen:

$$\frac{\pi}{2}, \pi, \frac{3\pi}{2}, 2\pi, \frac{5\pi}{2}, 3\pi, \frac{7\pi}{2}, \ldots, \frac{(2n-1)\pi}{2}, n\pi, \frac{(2n+1)\pi}{2}, \ldots$$

In den Intervallen zwischen diesen Werten nimmt (bei fallendem x) die Funktion von 1 bis 0 und von 0 bis -1 ab, danach wächst sie von -1 bis 0 und von 0 bis 1, usw.

Somit oszilliert die Funktion $\sin\frac{1}{x}$ unendlich oft, ähnlich der Funktion $\sin x$; während sich aber die Schwingungen von $\sin x$ über ein unendliches Intervall erstrecken, oszilliert $\sin\frac{1}{x}$ in einem endlichen Intervall, und die Schwingungen häufen sich um den Nullpunkt.

Ein Teil der Kurve ist in Abb. 23 dargestellt (natürlich ist es nicht möglich, unendlich viele Schwingungen abzubilden). Da $\sin\frac{1}{x}$ mit x das Vorzeichen wechselt, ist die linke Hälfte der Kurve zur rechten bezüglich des Ursprungs symmetrisch.

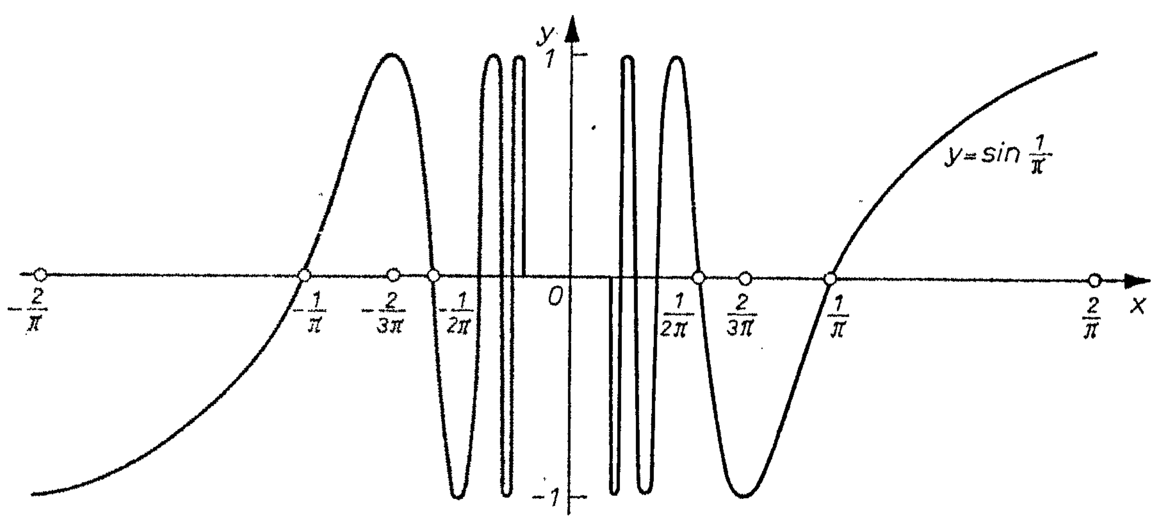

Abb. 23

10. Für die Funktion $x \cdot \sin \frac{1}{x}$ $(x \neq 0)$, die sich nur durch den Faktor x von $\sin \frac{1}{x}$ unterscheidet, existiert der Grenzwert für $x \to 0$:

$$\lim_{x \to 0} x \cdot \sin \frac{1}{x} = 0,$$

was sofort aus der Ungleichung $\left| x \cdot \sin \frac{1}{x} \right| \leq |x|$ folgt.

Für gegen 0 strebendes x oszilliert auch $x \cdot \sin \frac{1}{x}$ (ebenso wie $\sin \frac{1}{x}$) unendlich oft; auf Grund des Faktors x nimmt aber die Amplitude ab und strebt gegen 0. Dadurch ist die Existenz des Grenzwertes gewährleistet.

Die graphische Darstellung der Funktion

$$y = x \cdot \sin \frac{1}{x}$$

sehen wir in Abb. 24. Die Kurve liegt zwischen den beiden Winkelhalbierenden $y = x$ und $y = -x$.[1])

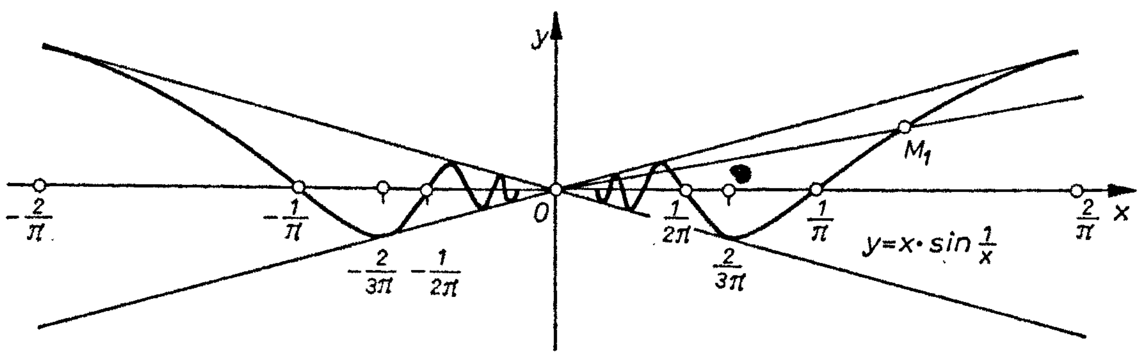

Abb. 24

Bemerkung. Wir hatten eine Reihe von Grenzwerten

$$\lim_{x \to 0} \frac{\sin x}{x} = 1, \quad \lim_{x \to 0} (1 + x)^{1/x} = e, \quad \lim_{x \to 0} x \cdot \sin \frac{1}{x} = 0,$$

die eine Besonderheit gemeinsam haben: Keine der hier betrachteten Funktionen ist für $x = 0$ definiert. Das hindert uns jedoch in keiner Weise, von ihren Grenzwerten für $x \to 0$ zu sprechen; denn nach der in Nr. 52 gegebenen Definition des Grenzwertes einer Funktion bleibt der Wert $x = 0$ außer Betracht.

Ähnlich ist es bei dem Problem, ob die Funktion $\sin \frac{1}{x}$ für $x \to 0$ einen Grenzwert hat. Die Frage ist sinnvoll, obwohl die Funktion für $x = 0$ nicht definiert ist. Allerdings existiert hier der Grenzwert nicht.

55. Erweiterung der Theorie der Grenzwerte. Natürlich erhebt sich die Frage, ob sich die im Kapitel I (§§ 1 und 2) für den Fall der Folgen (einer diskreten Veränderlichen) entwickelte Theorie der Grenzwerte auf den hier betrachteten allgemeinen Fall einer beliebigen Funktion erweitern läßt.

Das läßt sich auf zwei Wegen bewerkstelligen:

I. Man kann die dort angestellten Überlegungen in anderer Form wiederholen. Wir führen das als Beispiel für den Satz 1 aus Nr. 26 durch.

[1]) In den Abb. 23 und 24 haben wir der Übersichtlichkeit halber für die x-Achse einen größeren Maßstab gewählt, was eine Verzerrung ergibt.

Wir betrachten eine Funktion $f(x)$, die in einem Bereich $\mathcal{X}$ mit dem Häufungspunkt a gegeben ist.[1])

1. *Wenn für gegen a strebendes x die Funktion $f(x)$ den endlichen Grenzwert A hat und $A > p$ (bzw. $A < q$) ist, so genügt für hinreichend nahe bei a gelegene, von a verschiedene Werte von x die Funktion selbst der Ungleichung*

$$f(x) > p \quad (\textit{bzw. } f(x) < q). \tag{14}$$

Wir wählen eine positive Zahl $\varepsilon < A - p$ (bzw. $\varepsilon < q - A$); dann ist

$$A - \varepsilon > p \quad (\text{bzw. } A + \varepsilon < q).$$

Nach Definition des Grenzwertes gibt es aber zu diesem ε ein δ derart, daß für $|x - a| < \delta$ (wobei x aus $\mathcal{X}$ und von a verschieden ist) die Beziehung

$$A - \varepsilon < f(x) < A + \varepsilon$$

gilt. Für diese Werte von x ist aber (14) erst recht erfüllt.

Der Leser sieht, daß zum Beweis keine neuen Ideen herangezogen werden mußten.

Hieraus können auch unmittelbar die Behauptungen 2, 3 und 5 aus Nr. 26 bewiesen werden. Wenn wir in Satz 1 etwa $p = 0$ ($q = 0$) setzen, so erhalten wir:

2. *Wenn eine Funktion $f(x)$ für gegen a strebendes x einen endlichen positiven (negativen) Grenzwert hat, so ist auch die Funktion selbst wenigstens für diejenigen x-Werte, die in der Nähe von a liegen, aber von a verschieden sind, positiv (negativ).*

Es gilt auch eine Behauptung, die zu der Behauptung 4 aus Nr. 26 analog ist, allerdings in weniger allgemeiner Form:

4. *Wenn eine Funktion $f(x)$ für gegen a strebendes x den endlichen Grenzwert A hat, so ist die Funktion für hinreichend nahe bei a gelegene x-Werte beschränkt:*

$$|f(x)| \leqq M' \quad (M' = \text{const}; \; |x - a| < \delta).$$

Wir erinnern daran, daß sich ursprünglich auch für eine diskrete Veränderliche x_n, die einen endlichen Grenzwert hat, die Ungleichung $|x_n| \leqq M'$ nur für $n > N$ ergab; weil aber nur für endlich viele Werte die Veränderliche dieser Ungleichung nicht zu genügen brauchte, konnte man durch Vergrößerung von M' leicht erreichen, daß die Ungleichung für alle x_n erfüllt ist. Hier kann man das im allgemeinen nicht, weil die Menge der x, für die $|f(x)| > M'$ ist, unendlich sein kann. Beispielsweise strebt die Funktion $f(x) = \frac{1}{x}$ (für $x > 0$) für $x \to 1$ gegen 1; offenbar ist $f(x) < 2$ für $|x - 1| < \frac{1}{2}$; jedoch ist die Funktion $f(x)$ für alle in Frage kommenden x, d. h. für $x > 0$, keineswegs beschränkt.

II. Wenn wir jetzt zu den anderen Sätzen übergehen, in denen die Veränderlichen durch Gleichheitszeichen, Ungleichheitszeichen oder arithmetische Operationen verknüpft sind, so müssen wir zunächst verabreden, daß wir, wenn wir zwei oder mehrere Funktionen $f(x), g(x), \ldots$ (die in ein und demselben Gebiet $\mathcal{X}$ definiert sind) durch solche Zeichen verknüpfen, darunter stets verstehen, daß die Funktionswerte ein und demselben x-Wert entsprechen.

Alle diese Sätze könnte man auf analoge Weise beweisen, aber (und das ist wichtig zu betonen) es ist in Wirklichkeit gar nicht notwendig, es noch einmal zu tun. Wenn

---

[1]) Die Zahl a kann auch ∞ sein, wir wollen uns hier aber auf *endliches* a beschränken.

wir uns nämlich bei den Grenzwerten von Funktionen auf den „Folgen"-Standpunkt stellen, *so bleiben die Sätze für die Funktionen gültig, weil sie für Folgen bewiesen sind.*

Als Beispiel betrachten wir die Sätze 1, 2, 3 aus Nr. 30:

Es seien in einem Gebiet $\mathfrak{X}$ (mit dem Häufungspunkt a) zwei Funktionen $f(x)$ und $g(x)$ gegeben, die für gegen a strebendes x beide endliche Grenzwerte haben:

$$\lim f(x) = A, \qquad \lim g(x) = B.$$

Dann haben auch die Funktionen

$$f(x) \pm g(x), \quad f(x)\,g(x), \quad \frac{f(x)}{g(x)} \tag{15}$$

endliche Grenzwerte (im Fall des Quotienten nur unter der Voraussetzung $B \neq 0$), und zwar die Grenzwerte

$$A \pm B, \quad AB, \quad \frac{A}{B}.$$

In der „Sprache der Folgen" lassen sich die Voraussetzungen folgendermaßen formulieren: Ist $\{x_n\}$ eine beliebige Folge von Werten x aus $\mathfrak{X}$, die den Grenzwert a hat, so gilt

$$f(x_n) \to A, \quad g(x_n) \to B.$$

Wendet man auf diese beiden Folgen die schon bewiesenen Sätze an, so erhält man

$$\lim \bigl(f(x_n) \pm g(x_n)\bigr) = A \pm B, \quad \lim \bigl(f(x_n)\,g(x_n)\bigr) = AB,$$
$$\lim \frac{f(x_n)}{g(x_n)} = \frac{A}{B},$$

und das bringt in der Tat (in der „Sprache der Folgen") das zum Ausdruck, was bewiesen werden sollte.[1])

Auf diese Weise lassen sich auch alle Aussagen aus Nr. 31 über die „unbestimmten Ausdrücke"

$$\frac{0}{0}, \quad \frac{\infty}{\infty}, \quad 0 \cdot \infty, \quad \infty - \infty$$

auf den allgemeinen Fall übertragen. Wie auch im einfachsten Fall, wo wir es mit Funktionen eines natürlichen Arguments zu tun hatten, genügt es hier zur Auswertung nicht, nur die Grenzwerte der Funktionen $f(x)$ und $g(x)$ selbst zu kennen, sondern man muß auch hier wieder die durch die Verknüpfung entstehenden Funktionen untersuchen.

Der Leser prüft leicht nach, daß es sich in den Beispielen 4 und 5 aus Nr. 54 um unbestimmte Ausdrücke der Form $\frac{\infty}{\infty}$ und $0 \cdot \infty$ handelte und im Beispiel 7 um einen unbestimmten Ausdruck der Form $\frac{0}{0}$. In Nr. 56 führen wir einige weitere Beispiele an, wobei wir die einfachsten Sätze aus der Theorie der Grenzwerte schon verwenden.

---

[1]) Im Fall des Quotienten wäre noch zu bemerken, daß (analog wie bei Folgen) für hinreichend nahe bei a gelegene x der Nenner $g(x)$ von 0 verschieden ist, so daß der Bruch $\frac{f(x)}{g(x)}$ wenigstens für diese x-Werte sinnvoll ist.

Wir kommen auf diese Fragen in Kapitel IV, § 4, zurück, wo wir allgemeine Methoden zur Auswertung unbestimmter Ausdrücke mit Hilfe der Differentialrechnung bringen werden.

56. Beispiele.

1. Wir verallgemeinern die Beispiele 1 und 2 aus Nr. 32 und untersuchen das Verhalten des Polynoms

$$p(x) = a_0 x^k + a_1 x^{k-1} + \cdots + a_{k-1} x + a_k,$$

und dann das Verhalten des Quotienten zweier solcher Polynome,

$$\frac{p(x)}{q(x)} = \frac{a_0 x^k + a_1 x^{k-1} + \cdots + a_{k-1} x + a_k}{b_0 x^l + b_1 x^{l-1} + \cdots + b_{l-1} x + b_l},$$

für $x \to \pm\infty$.

Mittels der Transformation

$$p(x) = x^k \left(a_0 + \frac{a_1}{x} + \cdots + \frac{a_k}{x^k}\right)$$

findet man leicht

$$\lim_{x \to \pm\infty} p(x) = \pm\infty \quad (\text{unbestimmter Ausdruck vom Typ } \infty - \infty),$$

wobei das Vorzeichen des Grenzwertes bei geradem k nur vom Vorzeichen von a_0, bei ungeradem k jedoch auch von dem Vorzeichen von x abhängt.

2. Analog ergibt sich

$$\lim_{x \to \pm\infty} \frac{p(x)}{q(x)} = \pm\infty, \quad \frac{a_0}{b_0}, \quad 0 \quad \left(\text{unbestimmter Ausdruck vom Typ } \frac{\infty}{\infty}\right),$$

je nachdem, ob $k > l$, $k = l$ oder $k < l$ ist. Das Vorzeichen des Grenzwertes ergibt sich (für $x \to +\infty$) aus den Vorzeichen von a_0 und b_0, richtet sich aber auch (bei ungeradem $k - l$) nach dem Vorzeichen von x.

3. Wir beweisen für jeden positiven rationalen Exponenten r die Formel[1])

$$\lim_{x \to 0} \frac{(1+x)^r - 1}{x} = r \quad \left(\text{unbestimmter Ausdruck vom Typ } \frac{0}{0}\right).$$

Wir beginnen mit dem einfachsten Fall, daß der Exponent eine natürliche Zahl ist: $r = n$. Nach dem binomischen Satz ist

$$\frac{(1+x)^n - 1}{x} = \frac{nx + \frac{n(n-1)}{1 \cdot 2} x^2 + \cdots + x^n}{x} = n + \frac{n(n-1)}{1 \cdot 2} x + \cdots + x^{n-1};$$

da für $x \to 0$ alle Glieder rechts, außer dem ersten, gegen 0 streben, ergibt sich

$$\lim_{x \to 0} \frac{(1+x)^n - 1}{x} = n.$$

Es sei jetzt $r = \frac{1}{m}$ (m eine natürliche Zahl); wir betrachten den Ausdruck

$$\frac{\sqrt[m]{1+x} - 1}{x}.$$

[1]) Später (Nr. 77, Beispiel 5c) wird sie auf den Fall eines beliebigen reellen Exponenten verallgemeinert.

Wir setzen $\sqrt[m]{1+x} - 1 = y$; dann ist $x = (1+y)^m - 1$. Wegen

$$1 - |x| < \sqrt[m]{1+x} < 1 + |x| \quad (\text{für } |x| < 1)$$

ist

$$\lim_{x\to 0} \sqrt[m]{1+x} = 1,$$

so daß mit x auch y gegen 0 strebt. Dann ist aber nach dem schon Bewiesenen

$$\lim_{x\to 0} \frac{\sqrt[m]{1+x} - 1}{x} = \lim_{y\to 0} \frac{y}{(1+y)^m - 1} = \frac{1}{m}.$$

Schließlich läßt sich der allgemeine Fall $r = \frac{n}{m}$ durch Einführung derselben Veränderlichen y erledigen:

$$\frac{(1+x)^{n/m} - 1}{x} = \frac{(1+y)^n - 1}{(1+y)^m - 1} = \frac{(1+y)^n - 1}{y} \cdot \frac{y}{(1+y)^m - 1},$$

also

$$\lim_{x\to 0} \frac{(1+x)^{n/m} - 1}{x} = \frac{n}{m}.$$

4. Man bestimme den Grenzwert

$$\lim_{x\to 0} \frac{\sqrt[m]{1+x} - 1 - \frac{x}{m}}{x^2}.$$

Mit Hilfe derselben Substitution $\sqrt[m]{1+x} - 1 = y$ bringen wir diesen Ausdruck auf die Form

$$\frac{y - \frac{1}{m}\left[(1+y)^m - 1\right]}{\left[(1+y)^m - 1\right]^2} = \frac{-\frac{m-1}{2} \cdot y^2 + \cdots}{m^2 y^2 + \cdots} = \frac{-\frac{m-1}{2} + \cdots}{m^2 + \cdots},$$

und hieraus folgt unmittelbar, daß der gesuchte Grenzwert gleich $-\frac{m-1}{2m^2}$ ist.

5. Der Grenzwert

$$\lim_{x\to 0} \frac{\sin x}{x} = 1$$

(vgl. Nr. 54, Beispiel 7) läßt sich oft zur Berechnung anderer Grenzwerte benutzen.

$$\lim_{x\to 0} \frac{1 - \cos x}{x^2} = \frac{1}{2} \quad \left(\frac{0}{0}\right); \tag{a}$$

offenbar ist nämlich

$$\frac{1 - \cos x}{x^2} = \frac{2\sin^2 \frac{x}{2}}{x^2} = \frac{1}{2} \cdot \left(\frac{\sin \frac{x}{2}}{\frac{x}{2}}\right)^2,$$

und der Ausdruck in Klammern strebt gegen 1.

$$\lim_{x\to 0} \frac{\tan x - \sin x}{x^3} = \frac{1}{2} \quad \left(\frac{0}{0}\right). \tag{b}$$

Auch hier führt eine Umformung leicht auf schon betrachtete Grenzwerte:

$$\frac{\tan x - \sin x}{x^3} = \frac{1}{\cos x} \cdot \frac{1 - \cos x}{x^2} \cdot \frac{\sin x}{x}.$$

Für $x \to 0$ strebt $\cos x$ gegen 1; die Behauptung (b) ergibt sich z. B. aus (a).

$$\lim_{x\to\pi/2} (\sec x - \tan x) = 0 \qquad (\infty - \infty). \tag{c}$$

Hier ist es zweckmäßig, die Substitution $\alpha = \frac{\pi}{2} - x$ einzuführen; offenbar gilt $\alpha \to 0$ für $x \to \frac{\pi}{2}$. Wir erhalten

$$\sec x - \tan x = \operatorname{cosec} \alpha - \cot \alpha = \frac{1-\cos\alpha}{\sin\alpha} = \frac{1-\cos\alpha}{\alpha^2}\cdot\frac{\alpha}{\sin\alpha}\cdot\alpha \to 0\,.$$

57. Der Grenzwert einer monotonen Funktion. Das Problem der Existenz des Grenzwertes einer Funktion

$$\lim_{x\to a} f(x)$$

ist besonders einfach für Funktionen eines speziellen Typs zu lösen, der eine Verallgemeinerung des Begriffs der monotonen Folge ist (Nr. 34).

Die Funktion $f(x)$ sei in einem Bereich $\mathscr{X} = \{x\}$ definiert. Sie wird in diesem Bereich (*monoton*) *wachsend* (bzw. *fallend*) genannt, wenn für jedes Paar von Argumentwerten gilt:

Aus $x' > x$ folgt $f(x') > f(x)$ [bzw. $f(x') < f(x)$].

Wenn

aus $x' > x$ nur $f(x') \geqq f(x)$ [bzw. $f(x') \leqq f(x)$]

folgt, so heißt die Funktion *nicht fallend* (bzw. *nicht wachsend*). Manchmal spricht man auch in diesem Fall von einer (im weiteren Sinne) wachsenden (bzw. fallenden) Funktion.

Für monotone Funktionen gilt in voller Analogie zu dem Satz aus Nr. 34 der folgende

Satz. *Die Funktion $f(x)$ sei in einem Gebiet $\mathscr{X}$, das einen (endlichen oder unendlichen) Häufungspunkt a habe, der größer als alle x-Werte ist, monoton nicht fallend. Ist dann $f(x)$ nach oben beschränkt,*

$$f(x) \leqq M \quad (\textit{für alle } x \textit{ aus } \mathscr{X}),$$

so hat $f(x)$ für $x \to a$ einen endlichen Grenzwert; ist $f(x)$ nicht nach oben beschränkt, so strebt $f(x)$ gegen ∞.

Beweis. Zunächst nehmen wir an, $f(x)$ sei nach oben beschränkt, d. h., die Menge $\{f(x)\}$ der Funktionswerte, die allen x aus $\mathscr{X}$ entsprechen, sei nach oben beschränkt. Dann existiert nach Nr. 11 für diese Menge eine endliche obere Grenze A. Wir zeigen nun, daß diese Zahl A der gesuchte Grenzwert ist.

Es sei eine beliebige Zahl $\varepsilon > 0$ vorgegeben; dann gibt es nach einer Eigenschaft der oberen Grenze einen Wert $x' < a$ derart, daß $f(x') > A - \varepsilon$ ist. Da $f(x)$ monoton ist, gilt für $x > x'$ erst recht $f(x) > A - \varepsilon$. Weil andererseits stets $f(x) \leqq A < A + \varepsilon$ gilt, ist für die erwähnten x-Werte die Ungleichung

$$|f(x) - A| < \varepsilon$$

erfüllt.

Damit ist unsere Behauptung bewiesen; man braucht für endliches a nur $x' = a - \delta$ (d. h. $\delta = a - x'$) zu setzen und für $a = \infty$ nur $\Delta = x'$ zu nehmen.

Ist $f(x)$ nicht nach oben beschränkt, so läßt sich zu jeder Zahl E ein x' finden derart, daß $f(x') > E$ ist; dann ist für $x > x'$ erst recht $f(x) > E$, usw.

Wir überlassen es dem Leser, diesen Satz für den Fall eines Häufungspunktes a, der kleiner ist als alle x-Werte, zu beweisen; ebenso auch für den Fall einer monoton nichtwachsenden Funktion.

Man sieht leicht, daß der in Nr. 34 betrachtete Satz über monotone Folgen einfach ein Spezialfall dieses Satzes ist. Die unabhängige Veränderliche war dort mit n bezeichnet, ihr Wertebereich war die Folge der natürlichen Zahlen, $\mathscr{N} = \{n\}$, mit dem Häufungspunkt ∞.

Im folgenden werden wir meist als Bereich $\mathscr{X}$, in dem eine Funktion $f(x)$ untersucht wird, ein Intervall $[a', a)$, wobei $a' < a$ und a endlich oder ∞ ist, oder auch ein Intervall $(a, a']$ zu betrachten haben, wobei $a' > a$ und a endlich oder $-\infty$ ist.

58. Das allgemeine Kriterium von Bolzano-Cauchy. Jetzt betrachten wir den allgemeinen Fall einer Funktion $f(x)$, die in einem Bereich $\mathscr{X} = \{x\}$ mit einem Häufungspunkt a gegeben ist. Für die Existenz eines endlichen Grenzwertes dieser Funktion bei gegen a strebendem x läßt sich ein entsprechendes Kriterium angeben wie im Fall der Folgen (Nr. 39). Wir formulieren es gleichzeitig für endliches a und für $a = \infty$.

Satz. *Die Funktion $f(x)$ hat für gegen a strebendes x genau dann einen endlichen Grenzwert, wenn zu jedem $\varepsilon > 0$ ein $\delta > 0$ (bzw. $\Delta > 0$) existiert derart, daß die Ungleichung*

$$|f(x) - f(x')| < \varepsilon$$

erfüllt ist, sobald

$$|x - a| < \delta \quad \textit{und} \quad |x' - a| < \delta \quad (\textit{bzw.}\ x > \Delta \quad \textit{und} \quad x' > \Delta)$$

gilt.

Wir beweisen diesen Satz unter der Voraussetzung, daß a endlich ist, und zeigen zunächst, daß die Bedingung notwendig ist. Es existiere der endliche Grenzwert $\lim\limits_{x \to a} f(x) = A$. Dann läßt sich zu gegebenem $\varepsilon > 0$ ein $\delta > 0$ finden derart, daß

$$|f(x) - A| < \frac{\varepsilon}{2}$$

ist, sobald $|x - a| < \delta$ gilt. Es sei auch $|x' - a| < \delta$, so daß auch

$$|A - f(x')| < \frac{\varepsilon}{2}$$

gilt. Hieraus erhält man

$$\begin{aligned} |f(x) - f(x')| &= |(f(x) - A) + (A - f(x'))| \\ &\leqq |f(x) - A| + |A - f(x')| < \varepsilon \end{aligned}$$

unter der Voraussetzung, daß gleichzeitig $|x - a| < \delta$ und $|x' - a| < \delta$ gilt.

Daß die Bedingung hinreichend ist, kann man mit Hilfe einer Überlegung beweisen, die der für Folgen (Nr. 39) angewandten analog ist. Einfacher ist es jedoch, diese

Überlegungen nicht zu wiederholen, sondern das Problem auf den schon betrachteten Fall zurückzuführen. Dabei bedienen wir uns der zweiten Definition des Grenzwertes einer Funktion „in der Sprache der Folgen" (Nr. 53).

Es sei jetzt die Voraussetzung des Satzes erfüllt, zu beliebig gegebenem $\varepsilon > 0$ gebe es ein entsprechendes $\delta > 0$. Ist $\{x_n\}$ eine beliebige Folge von x-Werten aus $\mathcal{X}$, die gegen a konvergiert, so gibt es nach Definition des Grenzwertes einer Folge eine Zahl N derart, daß für $n > N$ die Beziehung $|x_n - a| < \delta$ gilt. Wir nehmen neben n noch eine andere Zahl $n' > N$ derart, daß gleichzeitig $|x_n - a| < \delta$ und $|x_{n'} - a| < \delta$ gilt. Dann ist auf Grund der Wahl von δ und der Voraussetzung des Satzes

$$|f(x_n) - f(x_{n'})| < \varepsilon.$$

Diese Ungleichung gilt somit unter der einzigen Voraussetzung, daß die beiden Zahlen n und n' größer als N sind. Das bedeutet, daß für die Folge $\{f(x_n)\}$ $(n = 1, 2, 3, \ldots)$ die Voraussetzung aus Nr. 39 erfüllt ist; somit hat die Folge $f(x_1), f(x_2), \ldots, f(x_n), \ldots$ einen endlichen Grenzwert.

Nun haben wir in Nr. 53 gesehen (vgl. dort die Bemerkung am Schluß), daß die Existenz eines Grenzwertes jeder solchen beliebigen Folge schon hinreichend dafür ist, daß alle diese Grenzwerte übereinstimmen, wie auch die gegen a strebende Folge $\{x_n\}$ gewählt sei; dieser Grenzwert ist gerade der Grenzwert der Funktion, dessen Existenz bewiesen werden sollte.

Übrigens ist die Tatsache, daß die Bedingung hinreicht, auch leicht aus dem Satz von Bolzano-Weierstrass zu beweisen, ähnlich wie es für Folgen am Schluß von Nr. 41 durchgeführt wurde.

59. Größter und kleinster Grenzwert einer Funktion. Auch wenn für gegen a strebendes x kein Grenzwert von $f(x)$ existiert, kann für einzelne Folgen von gegen a strebenden Werten x_n ein Grenzwert $\lim\limits_{n\to\infty} f(x_n)$ existieren; dann spricht man von einem *partiellen Grenzwert* der Funktion.

Beispielsweise füllen bei der Funktion $\sin x$ für $x \to \pm\infty$ (oder bei $\sin\frac{1}{x}$ für $x \to 0$) diese partiellen Grenzwerte das ganze Intervall von -1 bis 1 aus.

Unter den partiellen Grenzwerten einer Funktion gibt es immer einen größten und einen kleinsten; man bezeichnet sie mit

$$\overline{\lim_{x\to a}}\, f(x) \quad \text{bzw.} \quad \underline{\lim_{x\to a}}\, f(x).$$

Die Übereinstimmung von größtem und kleinstem Grenzwert ist notwendig und hinreichend dafür, daß der Grenzwert der Funktion im gewöhnlichen Sinne existiert.

Wir beschränken uns hier auf die Formulierung dieses Satzes, beweisen ihn aber nicht. Der Beweis kann nach dem Gedankengang des Beweises in Nr. 42 geführt werden.

§ 3. Klassifikation unendlich kleiner und unendlich großer Größen

60. Vergleich unendlich kleiner Größen. Wir nehmen an, bei einer Untersuchung seien gleichzeitig eine Reihe unendlich kleiner Größen $\alpha, \beta, \gamma, \ldots$ zu betrachten, die sämtlich Funktionen ein und derselben Veränderlichen, etwa von x, sind, die gegen einen endlichen oder unendlichen Grenzwert a strebt.

In vielen Fällen ist es von Interesse, zu vergleichen, in welcher Weise unendlich kleine Größen gegen 0 streben. Dem Vergleich zweier unendlich kleiner Größen α und

β legt man das Verhalten ihres Quotienten zugrunde.[1]) Zu diesem Zweck treffen wir folgende Vereinbarungen:

I. *Wenn der Quotient* $\frac{\beta}{\alpha}$ *(und mit ihm auch* $\frac{\alpha}{\beta}$*) einen endlichen und von 0 verschiedenen Grenzwert hat, so sagen wir, die unendlich kleinen Größen* α *und* β *haben dieselbe Ordnung (oder seien von derselben Ordnung).*

II. *Wenn der Quotient* $\frac{\beta}{\alpha}$ *unendlich klein ist (also* $\frac{\alpha}{\beta}$ *unendlich groß), so sagen wir,* β *sei eine unendlich kleine Größe von höherer Ordnung als* α (oder auch: β *werde von höherer Ordnung unendlich klein als* α*), und* α *sei eine unendlich kleine Größe niedrigerer (geringerer) Ordnung als* β.

Ist etwa $\alpha = x \to 0$, so haben folgende unendlich kleinen Größen dieselbe Ordnung:

$$\sin x, \quad \tan x, \quad \sqrt[m]{1+x} - 1;$$

denn, wie wir wissen (vgl. Nr. 54, Beispiel 7, und Nr. 56, Beispiel 3), ist

$$\lim_{x\to 0} \frac{\sin x}{x} = 1, \quad \lim_{x\to 0} \frac{\tan x}{x} = 1, \quad \lim_{x\to 0} \frac{\sqrt[m]{1+x} - 1}{x} = \frac{1}{m}.$$

Ist die unendlich kleine Größe α von derselben Ordnung wie die unendlich kleine Größe β, so schreibt man $\alpha = O(\beta)$.

Dagegen sind die unendlich kleinen Größen

$$\sqrt[m]{1+x} - 1 - \frac{x}{m}, \quad 1 - \cos x, \quad \tan x - \sin x \tag{1}$$

offenbar von höherer Ordnung als x [Nr. 56, Beispiele 4 und 5 (a), (b)].

Natürlich kann es auch vorkommen, daß der Quotient zweier unendlich kleiner Größen nicht gegen einen Grenzwert strebt; nehmen wir beispielsweise (vgl. Nr. 54, Beispiele 9 und 10)

$$\alpha = x$$

und

$$\beta = x \sin \frac{1}{x},$$

so hat ihr Quotient $\frac{\beta}{\alpha} = \sin \frac{1}{x}$ für $x \to 0$ keinen Grenzwert. In diesem Fall sagt man, die beiden unendlich kleinen Größen seien nicht vergleichbar.

Ist die unendlich kleine Größe β von höherer Ordnung als die unendlich kleine Größe α, so schreibt man

$$\beta = o(\alpha).$$

Beispielsweise kann man schreiben:

$$1 - \cos x = o(x), \quad \tan x - \sin x = o(x) \quad \text{usw.}$$

[1]) Wir wollen annehmen, daß die Veränderliche, durch die wir dividieren, wenigstens für alle hinreichend nahe bei a gelegenen x-Werte von 0 verschieden ist.

Somit ist das Symbol $o(\alpha)$ eine gemeinsame Bezeichnung für unendlich kleine Größen höherer Ordnung als α. Diese Bezeichnung werden wir von jetzt ab benutzen. Die Symbole o und O (das letzte für unendlich kleine Größen gleicher Ordnung) werden nach E. LANDAU[1]) *Landau-Symbole* genannt.

61. Die Skala der unendlich kleinen Größen. Manchmal ist es notwendig, das Verhalten unendlich kleiner Größen etwas genauer zu charakterisieren, und zwar ihre Ordnungen mit Hilfe von Zahlen zu beschreiben. In diesem Fall wählen wir zunächst als eine Art „Standard" eine der in der betreffenden Untersuchung auftretenden unendlich kleinen Größen (etwa α); wir nennen sie Grundgröße. Natürlich ist diese Wahl in gewissem Sinne willkürlich; gewöhnlich nimmt man die einfachste. Sind die betrachteten Größen, wie wir annehmen, Funktionen von x, die für gegen a strebendes x unendlich klein werden, so nimmt man zweckmäßigerweise als unendlich kleine Grundgröße

$$x, \quad x - a \quad \text{bzw.} \quad \frac{1}{x},$$

je nachdem, ob a gleich 0, gleich einer endlichen, von 0 verschiedenen Zahl bzw. gleich ∞ ist.

Ferner bilden von den Potenzen einer unendlich kleinen Grundgröße α (wir wollen $\alpha > 0$ annehmen) diejenigen mit verschiedenen positiven Exponenten k, α^k, sozusagen eine Skala zur Abschätzung unendlich kleiner Größen komplizierterer Natur.[2])

III. *Man nennt eine unendlich kleine Größe β unendlich klein von k-ter Ordnung (bezüglich einer unendlich kleinen Grundgröße α), wenn β und α^k $(k > 0)$ Größen derselben Ordnung sind, d. h., wenn der Quotient $\frac{\beta}{\alpha^k}$ einen endlichen, von 0 verschiedenen Grenzwert hat.*

Jetzt kann man beispielsweise, wenn man sich mit der Aussage, daß die unendlich kleinen Größen (1) (für $x \to 0$) Größen höherer Ordnung als $\alpha = x$ sind, nicht begnügt, genau sagen, daß die ersten beiden Größen von (1) unendlich kleine Größen zweiter Ordnung sind und die dritte Größe von (1) eine unendlich kleine Größe dritter Ordnung bezüglich $\alpha = x$ ist; denn es ist [vgl. Nr. 56, Beispiele 4 und 5 (a), (b)]

$$\lim_{x\to 0} \frac{\sqrt[m]{1+x} - 1 - \frac{1}{m}\cdot x}{x^2} = -\frac{m-1}{2m^2},$$

$$\lim_{x\to 0} \frac{1-\cos x}{x^2} = \frac{1}{2}, \qquad \lim_{x\to 0} \frac{\tan x - \sin x}{x^3} = \frac{1}{2}.$$

Als komplizierteres Beispiel betrachten wir den Ausdruck

$$\beta = \sqrt{x+1} + \sqrt{x-1} - 2\sqrt{x}.$$

Für $x \to \infty$ wird er unendlich klein; das ist unmittelbar klar, wenn wir ihn in der Form

$$\beta = (\sqrt{x+1} - \sqrt{x}) - (\sqrt{x} - \sqrt{x-1}) = \frac{1}{\sqrt{x+1}+\sqrt{x}} - \frac{1}{\sqrt{x}+\sqrt{x-1}}$$

[1]) EDMUND LANDAU, 1877–1938, deutscher Mathematiker.

[2]) Man sieht leicht, daß für $k > 0$ die Größe α^k gleichzeitig mit α unendlich klein wird.

darstellen. Wenn wir diese Brüche auf einen Nenner bringen und erweitern, so finden wir

$$\beta = \frac{\sqrt{x-1}-\sqrt{x+1}}{(\sqrt{x+1}+\sqrt{x})\,(\sqrt{x}+\sqrt{x-1})}$$

$$= -\frac{2}{(\sqrt{x+1}+\sqrt{x})\,(\sqrt{x}+\sqrt{x-1})\,(\sqrt{x-1}+\sqrt{x+1})}.$$

Setzen wir $\alpha = \frac{1}{x}$, so folgt leicht

$$\lim_{x\to\infty}\frac{\beta}{\alpha^{3/2}} = \lim_{x\to\infty}\frac{-2\,(\sqrt{x})^3}{(\sqrt{x+1}+\sqrt{x})\,(\sqrt{x}+\sqrt{x-1})\,(\sqrt{x-1}+\sqrt{x+1})}$$

$$= \lim_{x\to\infty}\frac{-2}{\left(\sqrt{1-\frac{1}{x}}+1\right)\cdot\left(1+\sqrt{1-\frac{1}{x}}\right)\cdot\left(\sqrt{1-\frac{1}{x}}+\sqrt{1+\frac{1}{x}}\right)}$$

$$= -\frac{1}{4}.\,^{1)}$$

Somit läßt sich hier die Ordnung durch die Zahl $\frac{3}{2}$ ausdrücken.

Man darf aber keineswegs denken, daß für jede unendlich kleine Größe β (selbst wenn sie mit allen Potenzen α^k vergleichbar ist) eine bestimmte Ordnung angegeben werden kann.

Interessante Beispiele, die sich hierauf beziehen, kann man aus den in Nr. 54, Beispiel 4 und 5 (für $a > 1$ und $k > 0$), aufgestellten Formeln erhalten:

$$\lim_{x\to\infty}\frac{a^x}{x^k} = \infty, \qquad \lim_{x\to\infty}\frac{\log_a x}{x^k} = 0. \tag{2}$$

Zunächst folgt hieraus

$$\lim_{x\to\infty}\frac{x^k}{a^x} = 0, \qquad \lim_{x\to\infty}\frac{x^k}{\log_a x} = \infty.$$

Wenn wir hierin x durch $\frac{1}{x}$ ersetzen und in der ersten dieser Beziehungen $a = \frac{1}{c}$, $0 < c < 1$, setzen, so erhalten wir

$$\lim_{x\to 0}\frac{c^{1/x}}{x^k} = 0, \qquad \lim_{x\to 0}\frac{\frac{1}{\log_a x}}{x^k} = \infty.$$

Somit wird die unendlich kleine Größe $c^{1/x}$ $(0 < c < 1)$ von höherer Ordnung klein als alle Potenzen x^k $(k > 0)$, während die unendlich kleine Größe $\frac{1}{\log_a x}$ $(a > 1)$ von niedrigerer Ordnung klein wird als alle diese Potenzen.

62. Äquivalente unendlich kleine Größen. Wir gehen jetzt auf einen besonders wichtigen Spezialfall unendlich kleiner Größen derselben Ordnung ein:

IV. *Wir nennen die unendlich kleinen Größen α und β äquivalent* (in Zeichen: $\alpha \sim \beta$), *wenn ihre Differenz $\gamma = \beta - \alpha$ von höherer Ordnung klein wird als jede der unendlich kleinen Größen α und β:*

$$\gamma = o(\alpha) \quad \textit{und} \quad \gamma = o(\beta).$$

---

[1]) Hier benutzen wir überall die Beziehung $\lim_{z\to 0}\sqrt{1+z} = 1$; sie wurde in Nr. 56, Beispiel 3 (sogar für m-te Wurzeln), bewiesen.

Übrigens genügt es zu fordern, daß γ von höherer Ordnung als eine dieser Größen klein wird; denn ist etwa γ von höherer Ordnung als α, so ist γ auch von höherer Ordnung als β. In der Tat folgt aus $\lim \frac{\gamma}{\alpha} = 0$, daß auch

$$\lim \frac{\gamma}{\beta} = \lim \frac{\gamma}{\alpha + \gamma} = \lim \frac{\frac{\gamma}{\alpha}}{1 + \frac{\gamma}{\alpha}} = 0$$

ist.

Wir betrachten zwei äquivalente unendlich kleine Größen α und β, so daß also mit $\beta = \alpha + \gamma$ die Beziehung $\gamma = o(\alpha)$ gilt. Wenn wir α durch β ersetzen, also $\beta \approx \alpha$ setzen[1]), so strebt, wenn beide Größen abnehmen, nicht nur der absolute Fehler bei dieser Substitution, also $|\gamma|$ gegen 0, sondern auch der relative Fehler, der gleich $\left|\frac{\gamma}{\alpha}\right|$ ist.

Mit anderen Worten, *für hinreichend kleine Werte von α und β kann man mit beliebig großer relativer Genauigkeit $\beta = \alpha$ setzen.* Das kann man benutzen, um bei Näherungsrechnungen komplizierte unendlich kleine Größen durch äquivalente einfache zu ersetzen.

Wir geben ein nützliches Kriterium für die Äquivalenz zweier unendlich kleiner Größen an, das eigentlich eine zweite, der früheren gleichwertige Definition der Äquivalenz ist:

Zwei unendlich kleine Größen α und β sind genau dann äquivalent, wenn

$$\lim \frac{\beta}{\alpha} = 1$$

ist.

Es sei zunächst diese Beziehung erfüllt, also $\delta = \frac{\beta}{\alpha} - 1 \to 0$. Dann ist $\gamma = \beta - \alpha = \delta \cdot \alpha$ eine Größe von höherer Ordnung als α; denn es ist

$$\lim \frac{\gamma}{\alpha} = \lim \delta = 0 .$$

Es seien jetzt umgekehrt α und β äquivalent, d. h., $\gamma = \beta - \alpha$ sei eine unendlich kleine Größe von höherer Ordnung als α. Infolgedessen gilt $\frac{\beta}{\alpha} - 1 = \frac{\gamma}{\alpha} \to 0$, also $\frac{\beta}{\alpha} \to 1$, was zu beweisen war.

Mit Hilfe dieses Kriteriums ergibt sich beispielsweise, daß für $x \to 0$ die unendlich kleinen Größen $\sin x$ und $\tan x$ der Größe x äquivalent sind und $\sqrt[m]{1 + x} - 1$ äquivalent $\frac{1}{m} x$ ist. Hieraus erhalten wir die Näherungsformeln

$$\sin x \approx x, \qquad \tan x \approx x,$$

$$\sqrt[m]{x + 1} - 1 \approx \frac{1}{m} x,$$

speziell

$$\sqrt{1 + x} - 1 \approx \frac{1}{2} x.$$

---

[1]) Das Zeichen $\approx$ bedeutet annähernd gleich.

Die eben bewiesene Eigenschaft der Äquivalenz unendlich kleiner Größen ermöglicht es, sie zur Auswertung unbestimmter Ausdrücke der Form $\frac{0}{0}$ zu benutzen, d. h. zur Bestimmung des Grenzwertes des Quotienten $\frac{\beta}{\alpha}$ zweier unendlich kleiner Größen. *Jede von ihnen kann dabei durch äquivalente unendlich kleine Größen ersetzt werden, ohne daß das die Existenz und den Wert des Grenzwertes beeinflußt.*

In der Tat, ist $\bar{\alpha} \sim \alpha$ und $\bar{\beta} \sim \beta$, d. h.

$$\lim \frac{\bar{\alpha}}{\alpha} = 1 \quad \text{und} \quad \lim \frac{\beta}{\bar{\beta}} = 1,$$

so hat der Quotient

$$\frac{\beta}{\alpha} = \frac{\beta}{\bar{\beta}} \cdot \frac{\bar{\beta}}{\bar{\alpha}} \cdot \frac{\bar{\alpha}}{\alpha},$$

der sich von dem Quotienten $\frac{\bar{\beta}}{\bar{\alpha}}$ nur durch Faktoren, die gegen 1 streben, unterscheidet, gleichzeitig mit diesem Quotienten ein und denselben Grenzwert.

Wenn es gelingt, $\bar{\alpha}$ und $\bar{\beta}$ hinreichend einfach zu wählen, so kann das die Lösung der Aufgabe bedeutend erleichtern, beispielsweise

$$\lim_{x\to 0} \frac{\sqrt{1 + x + x^2} - 1}{\sin 2x} = \lim_{x\to 0} \frac{\frac{1}{2}(x + x^2)}{2x} = \frac{1}{4}.$$

Aus dem Bewiesenen ergibt sich sofort, daß *zwei unendlich kleine Größen, die einer dritten äquivalent sind, einander äquivalent sind.*

63. Aussonderung des Hauptteils. Ist eine unendlich kleine Grundgröße α gewählt, so sind die einfachsten unendlich kleinen Größen Ausdrücke der Form $c\alpha^k$, wobei c ein konstanter Koeffizient und $k > 0$ ist. Die unendlich kleine Größe β sei bezüglich α von k-ter Ordnung, d. h.

$$\lim \frac{\beta}{\alpha^k} = c,$$

wobei c eine endliche und von 0 verschiedene Zahl ist. Dann ist

$$\lim \frac{\beta}{c\alpha^k} = 1,$$

und die unendlich kleinen Größen β und $c\alpha^k$ sind äquivalent:

$$\beta \sim c\alpha^k.$$

Diese einfachste, einer gegebenen unendlich kleinen Größe β äquivalente unendlich kleine Größe $c\alpha^k$ heißt ihr *Hauptteil* (oder *Hauptglied*).

Wenn wir die oben hergeleiteten Resultate benutzen, so können wir außer für die schon angegebenen einfachen Beispiele leicht die Hauptteile der Ausdrücke

$$1 - \cos x \sim \frac{1}{2} x^2, \qquad \tan x - \sin x \sim \frac{1}{2} x^3$$

bestimmen. Hier strebt x gegen 0, und $\alpha = x$ ist die unendlich kleine Grundgröße.

Strebt schließlich x gegen ∞ und nimmt man die unendlich kleine Größe $\alpha = \frac{1}{x}$ als Grundgröße, so folgt

$$\sqrt{x+1} + \sqrt{x-1} - 2\sqrt{x} \sim -\frac{1}{4}\left(\frac{1}{x}\right)^{3/2}.$$

Alle diese Resultate führen wieder zu Näherungsformeln.

Es sei $\beta \sim c\alpha^k$, d. h. $\beta = c\alpha^k + \gamma$, wobei $\gamma = o(\alpha^k)$ ist. Man kann jetzt von der unendlich kleinen Größe γ wieder das Hauptglied aussondern, $\gamma = c'\alpha^{k'} + \delta$, wobei $k' > k$ und $\delta = o(\alpha^{k'})$ ist, usw. Setzt man z. B. (für $x \to 0$) $\sqrt[m]{1+x} - 1 = \frac{1}{m}x + \gamma$, so ist, wie wir aus Nr. 56, Beispiel 4, wissen,

$$\lim_{x\to 0} \frac{\gamma}{x^2} = -\frac{m-1}{2m^2},$$

so daß der Hauptteil von γ gleich $-\frac{m-1}{2m^2}x^2$ ist. Hieraus folgt

$$\sqrt[m]{1+x} - 1 = \frac{1}{m}x - \frac{m-1}{2m^2}x^2 + o(x^2),$$

insbesondere

$$\sqrt{1+x} - 1 = \frac{1}{2}x - \frac{1}{8}x^2 + o(x^2).$$

Diesen Prozeß, aus einer unendlich kleinen Größe die einfachsten unendlich kleinen Größen wachsender Ordnung nacheinander auszusondern, kann man auch weiter fortführen.

Wir beschränken uns in diesem Paragraphen auf die Herleitung der allgemeinen Begriffe, die wir an einigen Beispielen erläutern. Im weiteren Verlauf werden wir ein systematisches Verfahren sowohl zur Bestimmung des Hauptteils einer gegebenen unendlich kleinen Größe angeben als auch zur weiteren Aussonderung der einfachsten unendlich kleinen Größen, wovon schon die Rede war (vgl. Nr. 104, 124).

Zum Schluß wollen wir noch auf die Frage eingehen, ob man dann, wenn von zwei unendlich kleinen Größen β und γ die Hauptglieder $c\alpha^k$ und $c'\alpha^{k'}$ bekannt sind, etwas über das Hauptglied ihrer Summe $\beta + \gamma$ aussagen kann.

Für $k \neq k'$ ist offenbar dasjenige der Glieder $c\alpha^k$ und $c'\alpha^{k'}$ das Hauptglied, in dem der Exponent kleiner ist. Es sei jetzt $k = k'$; dann ist das Hauptglied für $\beta + \gamma$ die Summe $(c + c')\,\alpha^k$, allerdings unter der Voraussetzung, daß $c + c' \neq 0$ ist. Wenn sich jedoch die beiden Hauptglieder gegenseitig aufheben, so ist die Summe $\beta + \gamma$ eine unendlich kleine Größe höherer Ordnung als jeder der einzelnen Summanden.

Das ist z. B. der Fall bei $x \to 0$ für die unendlich kleinen Größen

$$\beta = \sqrt{1+x} - 1 \sim \frac{1}{2}x \quad \text{und} \quad \gamma = \sqrt{1-x} - 1 \sim -\frac{1}{2}x.$$

Wenn wir hierin noch die Glieder

$$\beta = \frac{1}{2}x - \frac{1}{8}x^2 + o(x^2), \qquad \gamma = -\frac{1}{2}x - \frac{1}{8}x^2 + o(x^2)$$

aussondern, so ist offenbar

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

so daß $\beta + \gamma$ eine unendlich kleine Größe zweiter Ordnung mit dem Hauptglied $-\frac{1}{4}x^2$ ist.

64. Aufgaben. Zur Erläuterung dieser Überlegungen bringen wir einige Aufgaben.

1. Es sei ein geradliniger Abstand im Gelände mit Hilfe einer Meßlatte der Länge l zu vermessen. Da faktisch die Meßlatte nicht genau längs der zu vermessenden Geraden angelegt werden kann, wird das Resultat der Vermessung etwas größer sein als die wirkliche Länge. Wir machen die ungünstigste Annahme, nämlich die, daß die Meßlatte im Zickzack angelegt wird, so daß ihre Enden von der Geraden abwechselnd auf der einen und der anderen Seite den Abstand λ haben (Abb. 25). Wir wollen den Fehler abschätzen.

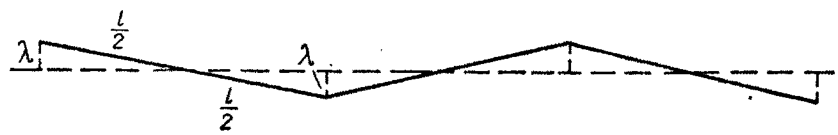

Abb. 25

Bei einem einmaligen Anlegen der Meßlatte ist der absolute Fehler gleich der Differenz der Länge l der Meßlatte und ihrer Projektion auf die zu vermessende Gerade. Die Länge der Projektion ist

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

Wenn wir die Näherungsformel $\sqrt{1+x} \approx 1 + \frac{1}{2}x$ für $x = -\frac{4\lambda^2}{l^2}$ benutzen (das ist erlaubt, da λ bezüglich l klein ist), erhalten wir für die Projektion den Ausdruck

$$l\left(1 - \frac{2\lambda^2}{l^2}\right) = l - \frac{2\lambda^2}{l}.$$

In diesem Fall ist der erwähnte Fehler gleich $\frac{2\lambda^2}{l}$, der relative Fehler offenbar $\frac{2\lambda^2}{l^2}$. Derselbe relative Fehler tritt auch bei mehrfachem Anlegen der Meßlatte auf.

Soll dieser relative Fehler kleiner als δ sein, d. h. $\frac{2\lambda^2}{l^2} < \delta$, so folgt $\lambda < l\sqrt{\frac{\delta}{2}}$.

Um also bei der Vermessung mit einer 2 m langen Meßlatte ($l = 2$) eine relative Genauigkeit von 0,001 zu erreichen, genügt es, die Abweichung λ nicht größer als $2\sqrt{0{,}0005} \approx 0{,}045$ m $= 4{,}5$ cm werden zu lassen.

2. Man gebe die Formel für die Länge l eines Treibriemens an, der um ein gegebenes Paar von Scheiben mit den Radien R und r läuft, deren Mittelpunkte den Abstand d haben (Abb. 26).

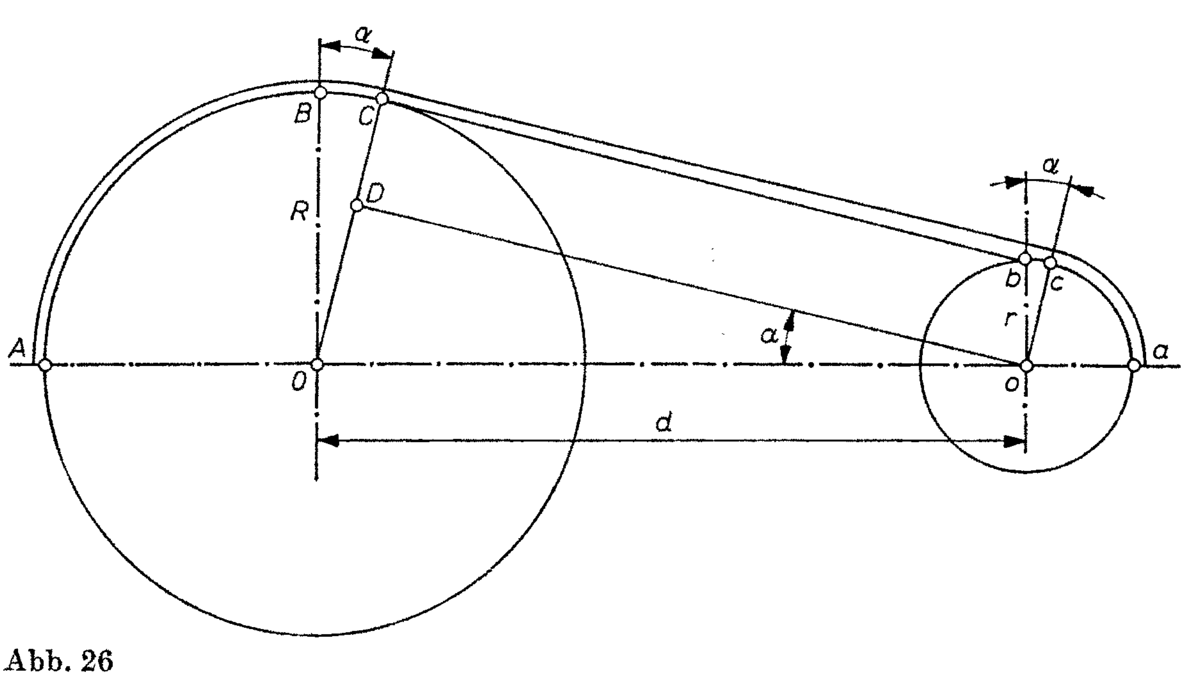

Abb. 26

Der Abb. 26 entnehmen wir $\frac{l}{2} = \widehat{AC} + \overline{Cc} + \widehat{ca}$. Nun ist aber $\widehat{AC} = R\left(\frac{\pi}{2} + \alpha\right)$, $\widehat{ca} = r\left(\frac{\pi}{2} - \alpha\right)$, wobei α die gleichen Winkel $\sphericalangle BOC$ und $\sphericalangle\, boc$ bezeichnet, und aus $\triangle\, ODo$ ergibt sich

$$\overline{Cc} = \overline{Do} = \sqrt{d^2 - (R - r)^2}.$$

Somit ist

$$l = \pi(R + r) + 2\alpha(R - r) + 2 \cdot \sqrt{d^2 - (R - r)^2}.$$

Zur Vereinfachung dieser Formel erinnern wir daran, daß

$$\alpha \approx \sin\alpha = \frac{\overline{OD}}{\overline{Oo}} = \frac{R - r}{d}$$

gilt, unter der Voraussetzung, daß $R - r$ bezüglich d klein ist. Unter derselben Voraussetzung ist

$$\sqrt{d^2 - (R - r)^2} = d\sqrt{1 - \left(\frac{R - r}{d}\right)^2} \approx d\left[1 - \frac{1}{2}\left(\frac{R - r}{d}\right)^2\right]$$

Setzen wir diese Werte ein und formen um, so erhalten wir schließlich

$$l \approx \pi(R + r) + 2d + \frac{(R - r)^2}{d}.$$

3. Beim Abstecken eines Kreisbogens im Gelände ist folgendes Problem von Bedeutung: Man bestimme das Verhalten der Pfeilhöhe $f = \overline{DB}$ des Kreisbogens ABC zu der Pfeilhöhe $f_1 = \overline{D_1B_1}$ des halben Bogens AB_1B (Abb. 27).

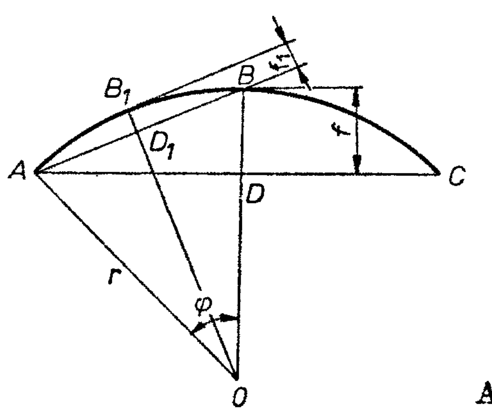

Abb. 27

Setzen wir den Radius des Kreisbogens gleich r und $\sphericalangle\, AOB = \varphi$, so ist $\sphericalangle\, AOB_1 = \frac{\varphi}{2}$ und

$$f = \overline{DB} = r(1 - \cos\varphi), \qquad f_1 = r\left(1 - \cos\frac{\varphi}{2}\right).$$

Somit ist das gesuchte Verhältnis gleich

$$\frac{f}{f_1} = \frac{1 - \cos\varphi}{1 - \cos\frac{\varphi}{2}}.$$

Dieser Ausdruck ist zu kompliziert, als daß er in der Praxis benutzt werden könnte. Wir bilden seinen Grenzwert für $\varphi \to 0$ (denn für hinreichend kleine φ kann dieser Ausdruck näherungsweise durch seinen Grenzwert ersetzt werden). Zu diesem Zweck ersetzen wir nun Zähler und

Nenner durch ihre Hauptteile und erhalten sofort

$$\lim \frac{f}{f_1} = \lim \frac{\frac{1}{2}\varphi^2}{\frac{1}{2}\left(\frac{1}{2}\varphi\right)^2} = 4.$$

Somit gilt für Kreisbögen über kleinen Zentriwinkeln, daß die Länge der Pfeilhöhe des Kreisbogens über den halben Winkel näherungsweise gleich dem vierten Teil der Länge der Pfeilhöhe des ursprünglichen Kreisbogens ist. In dieser Weise kann man nacheinander Punkte eines Kreisbogens konstruieren, dessen Endpunkte und dessen Mittelpunkt gegeben sind.

65. Klassifikation unendlich großer Größen. Natürlich können auch unendlich große Größen klassifiziert werden. Wie in Nr. 60 sehen wir die betrachteten unendlich großen Größen als Funktionen ein und derselben Veränderlichen x an, die für gegen a konvergierendes x gegen ∞ streben.

I. *Zwei unendlich große Größen y und z werden als Größen derselben Ordnung bezeichnet, wenn ihr Quotient $\frac{z}{y}$ (und damit auch $\frac{y}{z}$) einen endlichen und von 0 verschiedenen Grenzwert hat.*

II. *Wenn der Quotient $\frac{z}{y}$ unendlich groß (und der reziproke Quotient $\frac{y}{z}$ unendlich klein) wird, so heißt z eine unendlich große Größe von höherer Ordnung als y und y eine unendlich große Größe von niedrigerer Ordnung als z.*

Wenn der Quotient $\frac{z}{y}$ nicht gegen einen Grenzwert strebt, sind die unendlich großen Größen y und z nicht vergleichbar.

Betrachten wir eine Anzahl unendlich großer Größen gleichzeitig, so können wir unter diesen eine (sagen wir y) als Grundgröße wählen und die übrigen mit ihren Potenzen vergleichen. Sind beispielsweise (wie wir oben voraussetzten) alle diese Größen Funktionen von x und streben sie für $x \to a$ gegen ∞, so wählt man gewöhnlich $|x|$ als unendlich große Grundgröße, wenn $a = \pm\infty$ ist, und $\frac{1}{|x-a|}$ für endliches a.

III. *Eine unendlich große Größe z wird Größe k-ter Ordnung (bezüglich einer unendlich großen Grundgröße y) genannt, wenn z und y^k dieselbe Ordnung haben, d. h., wenn der Quotient $\frac{z}{y^k}$ einen endlichen und von 0 verschiedenen Grenzwert hat.*

Wir verzichten hier auf Beispiele, weil man leicht solche erhalten kann, wenn man die oben betrachteten unendlich kleinen Größen durch ihre reziproken Werte ersetzt. Wir erwähnen nur, daß für $x \to \infty$ die unendlich große Größe a^x $(a > 1)$ von höherer Ordnung und die unendlich große Größe $\log_a x$ $(a > 1)$ von niedrigerer Ordnung ist als jede Potenz x^k (mit positivem Exponenten k); dies folgt aus der Formel (2) aus Nr. 61.

§ 4. Stetigkeit (und Unstetigkeit) von Funktionen

66. Definition der Stetigkeit einer Funktion in einem Punkt. Mit dem Begriff des Grenzwertes hängt ein anderer wichtiger Begriff der Analysis eng zusammen, der Begriff der Stetigkeit einer Funktion.

Wir betrachten eine Funktion $f(x)$, die in einem Bereich $\mathscr{X} = \{x\}$ mit dem Häufungspunkt x_0 definiert ist; der Punkt x_0 selbst gehöre zum Definitionsbereich der Funktion, so daß diese in x_0 den wohlbestimmten Wert $f(x_0)$ hat.

Bei der Definition des Grenzwertes einer Funktion für gegen x_0 strebendes x (Nr. 52, 53),

$$\lim_{x \to x_0} f(x),$$

wurde mehrfach hervorgehoben, daß die Veränderliche x den Wert x_0 nicht annimmt; dieser Wert x_0 brauchte nicht einmal zum Definitionsbereich der Funktion zu gehören; wenn er aber dazu gehörte, wurde der Wert $f(x_0)$ bei der Bildung dieses Grenzwertes nicht berücksichtigt.

Ganz besonders wichtig ist nun der Fall, daß

$$\lim_{x \to x_0} f(x) = f(x_0) \tag{1}$$

gilt. Man sagt, die Funktion $f(x)$ sei in $x = x_0$ (oder im Punkt $x = x_0$) *stetig*, wenn die Beziehung (1) erfüllt ist; ist sie aber nicht erfüllt, so sagt man, die Funktion sei dort (in $x = x_0$) *unstetig.*[1])

Genau dann, wenn $f(x)$ in x_0 stetig ist, ist es bei der Berechnung des Grenzwertes von $f(x)$ für gegen x_0 strebendes x gleichgültig, ob dabei die Veränderliche den Wert x_0 annimmt oder nicht.

Man kann die Stetigkeit einer Funktion auch anders definieren. Den Übergang vom Wert x_0 zu einem anderen Wert x kann man sich auch so vorstellen, daß dem Wert x_0 der *Zuwachs* $\Delta x_0 = x - x_0$ erteilt wird.[2]) Der neue Wert der Funktion $y = f(x) = f(x_0 + \Delta x_0)$ unterscheidet sich von dem alten Wert $y_0 = f(x_0)$ um den Zuwachs

$$\Delta y_0 = f(x) - f(x_0) = f(x_0 + \Delta x_0) - f(x_0).$$

Die Funktion $f(x)$ ist nun im Punkt x_0 genau dann stetig, wenn ihr Zuwachs Δy_0 in diesem Punkt gemeinsam mit dem Zuwachs Δx_0 der unabhängigen Veränderlichen gegen 0 strebt. Mit anderen Worten: *Eine stetige Funktion ist dadurch charakterisiert, daß einem unendlich kleinen Zuwachs des Arguments ein unendlich kleiner Zuwachs der Funktion entspricht.*

Die grundlegende Definition (1) können wir auch in der „ε-δ-Sprache" ausdrücken (Nr. 52); und zwar läßt sich die Stetigkeit der Funktion $f(x)$ im Punkt x_0 folgendermaßen formulieren: Zu beliebig vorgegebenem $\varepsilon > 0$ läßt sich eine Zahl $\delta > 0$ finden derart, daß die Ungleichung

$$|x - x_0| < \delta$$

die Ungleichung

$$|f(x) - f(x_0)| < \varepsilon$$

---

[1]) Bekanntlich stellt man sich Funktionen *anschaulich* als Kurven vor. Eine Funktion ist dann, anschaulich gesprochen, stetig, wenn ihre Kurve stetig, d. h. ohne Unterbrechung, verläuft. Im Fall einer Unstetigkeit zeigt die Kurve eine Unterbrechung. In Wirklichkeit muß jedoch der Begriff einer stetigen Kurve *streng eingeführt* werden, und der einfachste Weg dahin führt über den Begriff der Stetigkeit einer Funktion.

[2]) In der Analysis bezeichnet man den Zuwachs einer Größe $x, y, t, \ldots$ mit $\Delta x, \Delta y, \Delta t, \ldots$ Diese Bezeichnung ist als ein einziges Symbol aufzufassen; Δ darf nicht von $x, y, t, \ldots$ getrennt werden.

zur Folge hat. Diese Ungleichung muß also in einer hinreichend kleinen Umgebung $(x_0 - \delta, x_0 + \delta)$ des Punktes x_0 erfüllt sein.

Schließlich kann man die Stetigkeit in der „Sprache der Folgen" auch in folgender Weise ausdrücken:

Für jede gegen x_0 konvergierende Folge von Werten x aus $\mathfrak{X}$, $x_1, x_2, \ldots, x_n, \ldots$, konvergiert die entsprechende Folge der Funktionswerte $f(x_1), f(x_2), \ldots, f(x_n), \ldots$ gegen $f(x_0)$.

Bemerkung. Der Punkt $x = x_0$, welcher Häufungspunkt des Definitionsbereichs $\mathfrak{X}$ der Funktion $f(x)$ ist, möge nun selbst nicht dem Bereich $\mathfrak{X}$ angehören, so daß also $f(x)$ in diesem Punkt nicht definiert ist. Wenn nun der endliche Grenzwert $\lim\limits_{x \to x_0} f(x)$ existiert, so braucht man die Definition der Funktion nur dadurch zu erweitern, daß man $f(x_0)$ gleich diesem Grenzwert setzt, um eine auch in dem Punkt $x = x_0$ stetige Funktion zu erhalten. Das wollen wir von jetzt ab in solchen Fällen so handhaben.

Wenn dagegen der erwähnte Grenzwert nicht existiert, so sagt man, obwohl die Funktion in x_0 nicht definiert ist, sie sei in diesem Punkt unstetig. Sie hat dort stets eine Unstetigkeit, welchen Wert man ihr für $x = x_0$ auch zusätzlich zuschreibt.

Gewöhnlich werden wir im folgenden Funktionen betrachten, die in einem offenen, halboffenen oder abgeschlossenen Intervall $\mathfrak{X}$ definiert sind; seine sämtlichen Punkte sind also Häufungspunkte, so daß man die Fragen nach der Stetigkeit für jeden seiner Punkte stellen kann. Zur Vereinfachung der Redeweise wollen wir vereinbaren, eine Funktion im Intervall $\mathfrak{X}$ stetig zu nennen, wenn sie in jedem einzelnen Punkt des Intervalls stetig ist. Für die Endpunkte des Intervalls vgl. Nr. 69.

67. Das Rechnen mit stetigen Funktionen. Bevor wir Beispiele stetiger Funktionen angeben, leiten wir folgenden einfachen Satz her, der es ermöglicht, die Anzahl der Beispiele zu vermehren.

Satz. *Sind die beiden Funktionen $f(x)$ und $g(x)$ in ein und demselben Intervall $\mathfrak{X}$ definiert und beide im Punkt x_0 stetig, so sind auch die Funktionen*

$$f(x) \pm g(x), \quad f(x)\,g(x), \quad \frac{f(x)}{g(x)}$$

in diesem Punkt stetig, die letzte natürlich unter der Bedingung $g(x_0) \neq 0$.

Dies folgt unmittelbar aus dem Satz über den Grenzwert der Summe, der Differenz, des Produktes und des Quotienten zweier Funktionen (vgl. Nr. 55).

Wir gehen hier nur auf den Quotienten zweier Funktionen näher ein. Die Stetigkeit der Funktionen $f(x)$ und $g(x)$ im Punkt x_0 ist gleichbedeutend mit der Gültigkeit der Beziehungen

$$\lim_{x \to x_0} f(x) = f(x_0) \quad \text{und} \quad \lim_{x \to x_0} g(x) = g(x_0).$$

Hieraus erhält man aber nach dem Satz über den Grenzwert des Quotienten (da der Grenzwert des Nenners von 0 verschieden ist)

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = \frac{f(x_0)}{g(x_0)};$$

das bedeutet aber gerade, daß $\dfrac{f(x)}{g(x)}$ im Punkt x_0 stetig ist.

68. Beispiele stetiger Funktionen.

1. *Ganze und gebrochene rationale Funktionen.* Die Funktion $f(x) = x$ ist offenbar im ganzen Intervall $(-\infty, \infty)$ stetig: Für $x_n \to x_0$ gilt die Beziehung $f(x_n) = x_n \to x_0 = f(x_0)$. Ebenso ergibt sich leicht, daß die Konstante eine stetige Funktion ist.

Hieraus folgt nach dem Satz aus Nr. 67 die Stetigkeit jedes Ausdrucks

$$ax^m = a \cdot \underbrace{x \cdot x \cdots x}_{m\text{-mal}}$$

als Produkt stetiger Funktionen und dann auch jedes Polynoms (jeder ganzen rationalen Funktion) $a_0x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n$ als Summe stetiger Funktionen. In allen diesen Fällen sind diese Funktionen im ganzen Intervall $(-\infty, \infty)$ stetig.

Offenbar ist schließlich auch der Quotient zweier Polynome (eine gebrochene rationale Funktion)

$$\frac{a_0x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n}{b_0x^m + b_1x^{m-1} + \cdots + b_{m-1}x + b_m}$$

für alle x-Werte, für die der Nenner ungleich 0 ist, stetig.

2. *Die Exponentialfunktion.* Wir beweisen jetzt die Stetigkeit der Exponentialfunktion a^x für jeden Wert $x = x_0$, mit anderen Worten, wir beweisen, daß

$$\lim_{x \to x_0} a^x = a^{x_0}$$

gilt. (Dabei genügt es, sich auf den Fall $a > 1$ zu beschränken.)

Wir haben in Nr. 54, Beispiel 6, gesehen, daß $\lim\limits_{x \to 0} a^x = 1$ ist. Da man per definitionem $a^0 = 1$ setzt, bringt diese Beziehung gerade die Stetigkeit der Exponentialfunktion im Punkt $x = 0$ zum Ausdruck. Wegen $a^x - a^{x_0} = a^{x_0}(a^{x-x_0} - 1)$ folgt hieraus, da für $x \to x_0$ offenbar $x - x_0 \to 0$ gilt, $a^{x-x_0} \to 1$, also in der Tat $a^x \to a^{x_0}$.

3. *Die Hyperbelfunktionen.* Ihre Stetigkeit folgt nach dem Satz aus Nr. 67 unmittelbar aus der soeben bewiesenen Stetigkeit der Exponentialfunktion, weil sie rational durch die Funktion e^x ausgedrückt werden können.

4. *Die trigonometrischen Funktionen.* Wir betrachten zunächst die Funktion $\sin x$. Sie ist für jedes $x = x_0$ stetig, d. h., es gilt

$$\lim_{x \to x_0} \sin x = \sin x_0.$$

Um das zu beweisen, bemerken wir, daß aus der in Nr. 54, Formel (9), für $0 < x < \frac{\pi}{2}$ bewiesenen Ungleichung $\sin x < x$ leicht folgt, daß die Ungleichung

$$|\sin x| \leqq |x|$$

für alle x-Werte gilt (für $|x| \geqq \frac{\pi}{2} > 1$ folgt sie aus $|\sin x| \leqq 1$). Nun ist aber

$$\sin x - \sin x_0 = 2 \sin \frac{x - x_0}{2} \cdot \cos \frac{x + x_0}{2},$$

also

$$|\sin x - \sin x_0| = 2 \cdot \left|\sin \frac{x - x_0}{2}\right| \cdot \left|\cos \frac{x + x_0}{2}\right|$$
$$\leqq 2 \cdot \left|\sin \frac{x - x_0}{2}\right| \leqq 2 \cdot \frac{|x - x_0|}{2}$$

und schließlich

$$|\sin x - \sin x_0| \leqq |x - x_0| \tag{2}$$

für alle Werte x und x_0.

Ist ein beliebiges $\varepsilon > 0$ vorgegeben, so setzen wir $\delta = \varepsilon$; für $|x - x_0| < \delta$ gilt

$$|\sin x - \sin x_0| < \varepsilon,$$

womit die Stetigkeit von $\sin x$ bewiesen ist. Analog beweist man auch die Stetigkeit der Funktion $\cos x$ für jedes x.

Hieraus ergibt sich nach dem Satz aus Nr. 67 sofort die Stetigkeit der Funktionen

$$\tan x = \frac{\sin x}{\cos x}, \quad \sec x = \frac{1}{\cos x}, \quad \cot x = \frac{\cos x}{\sin x}, \quad \operatorname{cosec} x = \frac{1}{\sin x}.$$

Eine Ausnahme bilden für die ersten beiden die Werte der Form $(2k + 1)\frac{\pi}{2}$, für die $\cos x$ gleich 0 wird, und für die letzten beiden die Werte der Form $k\pi$, für die $\sin x$ gleich 0 wird.

69. Einseitige Stetigkeit. Klassifikation der Unstetigkeitsstellen. Oben haben wir mit Hilfe der Beziehung (1) den Begriff der Stetigkeit einer Funktion $f(x)$ im Punkt x_0 definiert. Bei der Berechnung des Grenzwertes (1) konnte sich x sowohl von rechts als auch von links dem Punkt x_0 nähern. Wir definieren jetzt den Begriff der einseitigen Stetigkeit bzw. der einseitigen Unstetigkeit einer Funktion in einem gegebenen Punkt.

Man sagt, die Funktion $f(x)$ sei im Punkt x_0 *von rechts* (bzw. *von links*) *stetig*, wenn die Limesbeziehungen

$$\left.\begin{aligned} f(x_0 + 0) &= \lim_{x \to x_0 + 0} f(x) = f(x_0) \\ \text{bzw.} \quad f(x_0 - 0) &= \lim_{x \to x_0 - 0} f(x) = f(x_0) \end{aligned}\right\} \tag{3}$$

erfüllt sind. Ist jedoch eine dieser Beziehungen nicht erfüllt, so hat die Funktion $f(x)$ in dem Punkt x_0 von rechts (bzw. von links) eine Unstetigkeit.

Bezüglich des linken (bzw. rechten) als endlich angenommenen Endpunktes des abgeschlossenen Intervalls $\mathcal{X}$, in dem die Funktion definiert ist, kann man offenbar auch nur von der rechtsseitigen (bzw. linksseitigen) Stetigkeit oder Unstetigkeit sprechen. Ist aber x_0 ein innerer Punkt des Intervalls $\mathcal{X}$, d. h., stimmt x_0 mit keinem der Endpunkte überein, so ist die Gleichung (1), welche die Stetigkeit von $f(x)$ in x_0 im gewöhnlichen Sinne zum Ausdruck bringt, genau dann erfüllt, wenn gleichzeitig beide Gleichungen (3) gelten. Mit anderen Worten, die Stetigkeit einer Funktion im Punkt x_0 ist gleichbedeutend mit ihrer gleichzeitigen Stetigkeit von rechts und von links in diesem Punkt.

Wir wollen jetzt die Stetigkeit und die Unstetigkeit einer Funktion $f(x)$ im Punkt x_0 näher untersuchen. Dabei beschränken wir uns auf die rechtsseitige Stetigkeit bzw. Unstetigkeit. Wir setzen voraus, $f(x)$ sei in einem Intervall $[x_0, x_0 + h]$, $h > 0$, rechts von diesem Punkt definiert; alsdann sind folgende Bedingungen für die Stetigkeit notwendig und hinreichend: Erstens muß für von rechts gegen x_0 strebendes x der endliche Grenzwert $f(x_0 + 0)$ der Funktion $f(x)$ vorhanden sein, und zweitens muß dieser Grenzwert gleich dem Wert $f(x_0)$ der Funktion im Punkt x_0 sein.

Daher kann man sich leicht klar darüber werden, unter welchen Bedingungen die Funktion $f(x)$ im Punkt x_0 eine rechtsseitige Unstetigkeit hat. Es kann z. B. vorkommen, daß zwar ein endlicher Grenzwert $f(x_0 + 0)$ existiert, aber nicht gleich dem Wert $f(x_0)$ ist; eine solche Unstetigkeit heißt *gewöhnliche Unstetigkeit* oder *Unstetigkeit erster Art.*[1]) Es kann aber auch vorkommen, daß der Grenzwert $f(x_0 + 0)$ unendlich ist oder überhaupt nicht existiert; dann spricht man von einer *Unstetigkeit zweiter Art.*

In Nr. 70 geben wir einige Beispiele dieser Unstetigkeiten an.

Bemerkung. Wenn die Funktion $f(x)$ im Punkt $x = x_0$ nicht definiert ist (vgl. die Bemerkung in Nr. 66), so kann man sie nur dann zu einer in diesem Punkt stetigen Funktion machen, wenn die beiden Grenzwerte $f(x_0 + 0)$ und $f(x_0 - 0)$ existieren, endlich sind und übereinstimmen. Man spricht dann von einer *hebbaren Unstetigkeit.*

Ist einer dieser Grenzwerte unendlich oder nicht vorhanden, so liegt eine Unstetigkeit zweiter Art (von der entsprechenden Seite) vor.

70. Beispiele unstetiger Funktionen.

1. Wir betrachten die Funktion $y = E(x) = [x]$ (vgl. Abb. 8, S. 99). Ist x_0 keine ganze Zahl und $E(x_0) = m$, also $m < x_0 < m + 1$, so gilt auch für alle Werte von x in dem Intervall $(m, m + 1)$ die Beziehung $E(x) = m$, so daß die Stetigkeit der Funktion im Punkt x_0 unmittelbar klar ist.

Anders ist es, wenn x_0 gleich einer ganzen Zahl m ist. Von rechts ist $E(x)$ dort stetig, weil rechts von $x = m$, nämlich für Werte von x in $[m, m + 1)$, die Beziehung $E(x) = m$ gilt, so daß auch $E(m + 0) = m = E(m)$ ist. Dagegen ist links von $x = m$, d. h. für Werte von x in $[m - 1, m)$ offenbar $E(x) = m - 1$; hieraus folgt $E(m - 0) = m - 1$; also ist $E(m - 0) \neq E(m)$; daher hat $E(x)$ im Punkt $x = m$ eine linksseitige gewöhnliche Unstetigkeit, es liegt eine Sprungstelle vor.

2. Wir betrachten jetzt die schon in Nr. 46 untersuchte Funktion

$$y = f(x) = \lim_{n \to \infty} \frac{x^{2n} - 1}{x^{2n} + 1}$$

(ihre graphische Darstellung zeigt Abb. 28). Sie hat in den Punkten $x = \pm 1$ sowohl von rechts als auch von links eine gewöhnliche Sprungstelle; es ist nämlich

$$f(\pm 1) = 0, \quad f(-1 - 0) = f(1 + 0) = 1, \quad f(-1 + 0) = f(1 - 0) = -1.$$

3. Für die für $x \neq 0$ definierte Funktion

$$f(x) = \frac{1}{x^3}$$

ist der Punkt $x = 0$ eine Unstetigkeitsstelle zweiter Art, und zwar von beiden Seiten, da in diesem Punkt die Funktionen sowohl von rechts als auch von links ∞ wird:

$$f(+0) = \lim_{x \to +0} \frac{1}{x^3} = \infty, \qquad f(-0) = \lim_{x \to -0} \frac{1}{x^3} = -\infty.$$

---

[1]) In diesem Fall sagt man auch, die Funktion $f(x)$ habe im Punkt x_0 eine *Sprungstelle*; die Größe des Sprunges ist gleich $f(x_0 + 0) - f(x_0)$.

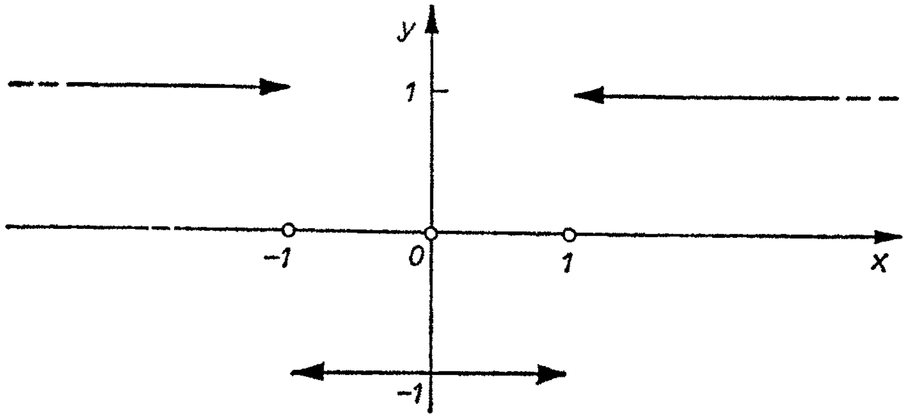

Abb. 28

4. Die in Nr. 54, Beispiel 9, für $x \neq 0$ betrachtete Funktion

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

hat im Punkt $x = 0$ eine Unstetigkeit zweiter Art, und zwar eine beiderseitige; ob x von rechts bzw. von links gegen 0 strebt, in keinem Fall existiert ein Grenzwert.

5. Betrachten wir dagegen für $x \neq 0$ die Funktion

$$f(x) = x \cdot \sin \frac{1}{x}$$

(vgl. Nr. 54, Beispiel 10), für die, wie wir wissen, der Grenzwert

$$\lim_{x \to 0} f(x) = 0$$

existiert, so sehen wir, daß wir sie durch die zusätzliche Definition $f(0) = 0$ zu einer auch in $x = 0$ stetigen Funktion machen können. Vgl. dazu die Bemerkung in Nr. 66.

6. Wir definieren durch

$$f_1(x)\, a^{1/x} \; (a > 1), \qquad f_2(x) = \arctan \frac{1}{x}$$

für $x \neq 0$ zwei Funktionen und setzen überdies $f_1(0) = f_2(0) = 0$.

Für die erste von ihnen gilt

$$f_1(+0) = \lim_{x \to +0} a^{1/x} = \lim_{z \to \infty} a^z = \infty, \qquad f_1(-0) = \lim_{x \to -0} a^{1/x} = \lim_{z \to -\infty} a^z = 0,$$

so daß im Punkt $x = 0$ von rechts eine Unstetigkeit zweiter Art und von links Stetigkeit vorliegt. Für die zweite ist

$$f_2(+0) = \lim_{x \to +0} \text{acrtan} \frac{1}{x} = \lim_{z \to \infty} \arctan z = \frac{\pi}{2}, \qquad f_2(-0) = -\frac{\pi}{2};$$

im Punkt $x = 0$ liegt eine beiderseitige Sprungstelle vor. Man vergleiche dazu Abb. 29 und 30.

7. Wir betrachten noch die Dirichletsche Funktion (Nr. 46):

$$\chi(x) = 1, \quad \text{wenn } x \text{ rational},$$

$$\chi(x) = 0, \quad \text{wenn } x \text{ irrational ist.}$$

Da sich in beliebiger Nähe eines rationalen Punktes irrationale Punkte befinden und umgekehrt, existiert, wie auch immer x_0 in dem Intervall $(-\infty, \infty)$ gewählt sein möge, der Grenzwert von $\chi(x)$ für $x \to x_0$ nicht; somit liegt in jedem Punkt eine (beiderseitige) Unstetigkeit zweiter Art vor.

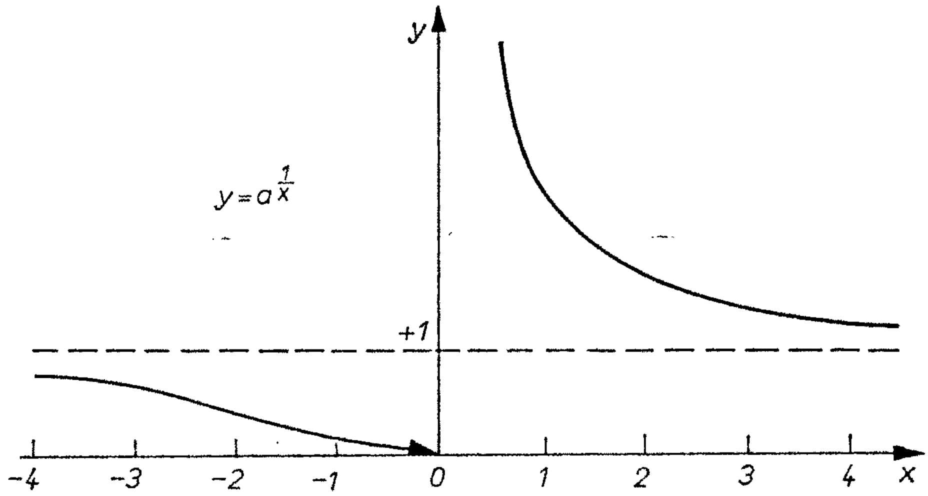

Abb. 29

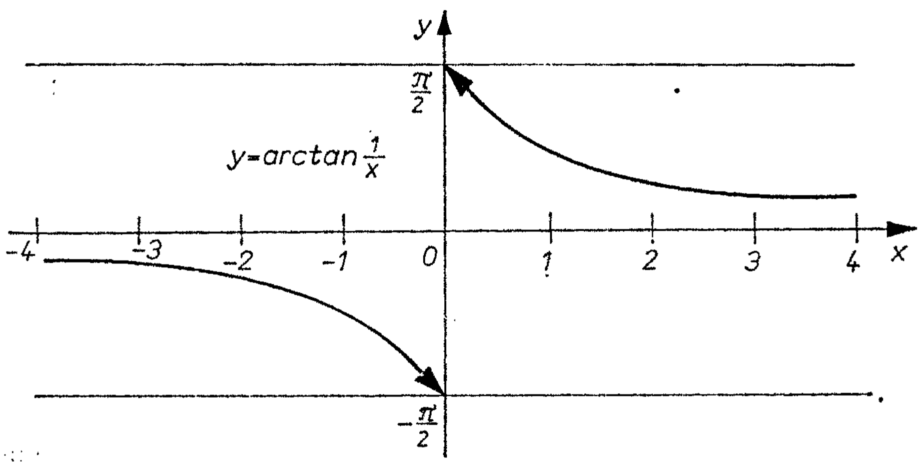

Abb. 30

8. Wir definieren schließlich die Funktion $f(x)$ in dem Intervall $[0, 1]$ folgendermaßen: Ist x rational und gleich dem unkürzbaren Bruch $\frac{p}{q}$, so sei $f(x) = \frac{1}{q}$; für irrationales x setzen wir $f(x) = 0$.[1]) Wir behaupten, daß diese Funktion in jedem rationalen Punkt gewöhnliche Unstetigkeiten hat, in jedem irrationalen Punkt jedoch stetig ist.

Es sei x_0 ein beliebiger Punkt in dem betrachteten Intervall. Wenn wir eine beliebige Zahl $\varepsilon > 0$ vorgeben, so existieren nur endlich viele natürliche Zahlen q, die nicht größer als $\frac{1}{\varepsilon}$ sind; das bedeutet, daß sich in dem Intervall nur endlich viele rationale Punkte $\frac{p}{q}$ befinden, für die $f\left(\frac{p}{q}\right) = \frac{1}{q} \geqq \varepsilon$ gilt. Daher kann man um den Punkt x_0 eine Umgebung $(x_0 - \delta, x_0 + \delta)$ derart beschreiben, daß darin keiner dieser Punkte liegt (mit eventueller Ausnahme des Punktes x_0 selbst). Ist dann $|x - x_0| < \delta$ $(x \neq x_0)$, so gilt dort sowohl für rationales als auch für irrationales x die Beziehung $|f(x)| < \varepsilon$. Somit existiert für jeden Punkt x_0

$$f(x_0 + 0) = f(x_0 - 0) = 0.$$

Ist x_0 irrational, so ist auch $f(x_0) = 0$, d. h., in diesem Punkt ist die Funktion stetig; ist aber x_0 rational, so ist $f(x_0) \neq 0$, und es liegt eine beiderseitige (gewöhnliche) Unstetigkeit vor. An diesem Beispiel erkennt man die vielfältigen, zum Teil völlig unanschaulichen Möglichkeiten, die in der abstrakten Definition der Stetigkeit stecken.

---

[1]) Diese Funktion wurde von Bernhard Riemann (1826—1866, deutscher Mathematiker) untersucht.

71. Stetigkeit und Unstetigkeit der monotonen Funktionen. Wir betrachten eine in einem Intervall $\mathcal{X}$ wenigstens im erweiterten Sinne (Nr. 57) monoton wachsende (bzw. fallende) Funktion $f(x)$. Das Intervall $\mathcal{X}$ kann dabei endlich oder auch unendlich, abgeschlossen oder auch halboffen bzw. offen sein. Über eine solche Funktion gilt folgender Satz:

1. *Eine monoton wachsende (bzw. abnehmende) Funktion $f(x)$ kann in $\mathcal{X}$ höchstens Unstetigkeiten erster Art, d. h. Sprünge, haben.*

Wir wählen einen beliebigen Punkt x aus $\mathcal{X}$, der aber nicht der linke Endpunkt sein soll. Auf den links von x liegenden Teil von $\mathcal{X}$ wenden wir den Satz aus Nr. 57 über den Grenzwert einer monotonen Funktion an. Da für $x < x_0$ nach Voraussetzung $f(x) \leqq f(x_0)$ gilt, existiert der endliche Grenzwert

$$f(x_0 - 0) = \lim_{x \to x_0 - 0} f(x) \leqq f(x_0).$$

Stimmt er mit dem Wert $f(x_0)$ überein, so ist $f(x)$ in x_0 von links stetig; anderenfalls liegt ein Sprung vor.

Analog überzeugen wir uns davon, daß in jedem Punkt x_0 aus $\mathcal{X}$, der nicht der rechte Endpunkt ist, $f(x)$ entweder von rechts stetig ist oder eine Sprungstelle hat.

Mit Hilfe des soeben bewiesenen Satzes läßt sich leicht ein Kriterium für die Stetigkeit einer monotonen Funktion herleiten, das für die Anwendungen nützlich ist:

2. *Sind die Werte einer in einem Intervall $\mathcal{X}$ monoton wachsenden (bzw. fallenden) Funktion $f(x)$ in einem Intervall $\mathcal{Y}$ enthalten und füllen sie es lückenlos aus, so daß also jeder Wert y aus $\mathcal{Y}$ von $f(x)$ mindestens einmal angenommen wird, so ist diese Funktion in $\mathcal{X}$ stetig.*[1])

Beweis. In irgendeinem Punkt x_0 aus $\mathcal{X}$ möge $f(x)$ eine Unstetigkeit besitzen, beispielsweise von links; wie wir gesehen haben, könnte es sich dabei nur um einen Sprung handeln. Es existiert also der Grenzwert $f(x_0 - 0)$, und dieser Wert ist kleiner als $f(x_0)$. Wegen $x < x_0$ ist $f(x) \leqq f(x_0 - 0)$, und für $x > x_0$ ist offenbar $f(x) \geqq f(x_0)$. Daher kann $f(x)$ keine Werte y annehmen, die zwischen den Zahlen $f(x_0 - 0)$ und $f(x_0)$ liegen und dem Intervall $\mathcal{Y}$ angehören. Das widerspricht aber der Voraussetzung des Satzes. Somit kann $f(x)$ keine Unstetigkeit besitzen.

In Nr. 72 geben wir eine Reihe von Beispielen für die Anwendung dieses äußerst nützlichen Satzes.

72. Die Stetigkeit der elementaren Funktionen. Für einige elementare Funktionen haben wir die Stetigkeit schon in Nr. 68 bewiesen. Unter Benutzung des Satzes 2 aus Nr. 71 läßt sich die Stetigkeit von a^x oder $\sin x$ leicht auf andere Art zeigen.

Die Funktion $y = a^x$ $(a > 1)$ wächst im Intervall $\mathcal{X} = (-\infty, \infty)$ monoton. Ihre Werte sind positiv und füllen das ganze Intervall $\mathcal{Y} = (0, \infty)$ aus, wie aus der Existenz der Funktion $x = \log_a y$ für jedes $y > 0$ folgt (vgl. Nr. 20). Daher ist die Exponentialfunktion für jedes x_0 stetig.

Analog folgt die Stetigkeit von $y = \sin x$ für jedes x, etwa aus $\mathcal{X} = \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, aus der Monotonie von $\sin x$ in diesem Intervall und der Tatsache, daß diese Funktion dabei jeden Wert zwischen -1 und $+1$ annimmt. (Das läßt sich geometrisch

[1]) Die hier formulierte Bedingung, daß die Funktionswerte das Intervall $\mathcal{Y}$ ausfüllen, ist für die Stetigkeit einer monotonen Funktion *hinreichend*; später (Nr. 82) werden wir uns davon überzeugen, daß sie auch *notwendig* ist.

leicht einsehen.) Das gilt auch für jedes Intervall der Form

$$\left[k\pi - \frac{\pi}{2}, k\pi + \frac{\pi}{2}\right] \quad (k = 0, \pm 1, \pm 2, \ldots).$$

Es sei darauf hingewiesen, daß wir uns im Grunde bei der Einführung der trigonometrischen Funktionen auf Schulkenntnisse stützen.

Interessanter sind für uns jedoch neue Resultate, die sich aus der Anwendung des erwähnten Satzes ebenso leicht ergeben. Wir setzen also die in Nr. 68 begonnene Aufzählung der elementaren Funktionen fort.

5. *Die Logarithmusfunktion* $y = \log_a x$ $(a > 0, a \neq 1)$. Für $a > 1$, worauf wir uns beschränken wollen, wächst sie für x aus $\mathcal{X} = (0, \infty)$. Dabei nimmt sie offenbar jeden Wert y aus $\mathcal{Y} = (-\infty, \infty)$ an, nämlich für $x = a^y$. Also ist sie stetig.

6. *Die Potenzfunktion* $y = x^\mu$ $(\mu \gtrless 0)$ wächst für wachsendes $0 < x < \infty$, falls $\mu > 0$ ist, und fällt für $\mu < 0$. Dabei nimmt sie jeden positiven Wert y an, nämlich für $x = y^{1/\mu}$. Also sie ist stetig.[1])

7. *Die Umkehrfunktionen der trigonometrischen Funktionen:*

$$y = \arcsin x, \quad y = \arccos x, \quad y = \arctan x, \quad y = \operatorname{arccot} x.$$

Die ersten beiden sind stetig in $[-1, 1]$, die letzten beiden in $(-\infty, \infty)$; den Beweis überlassen wir dem Leser.

Zusammenfassend können wir also sagen, daß *die elementaren Funktionen in allen Punkten ihres Definitionsbereichs* — in ihren natürlichen Definitionsbereichen — *stetig sind.*

73. Verkettung stetiger Funktionen. Umfassendere Klassen stetiger Funktionen können durch Verkettung von Funktionen (vgl. Nr. 51), deren Stetigkeit schon bekannt ist, erhalten werden.

Dabei ist folgender Satz grundlegend:

Satz. *Die Funktion $\varphi(y)$ sei im Intervall $\mathcal{Y}$ definiert, die Funktion $f(x)$ im Intervall $\mathcal{X}$; dabei sollen die Werte von $f(x)$ nicht außerhalb von $\mathcal{Y}$ liegen, wenn x das Intervall $\mathcal{X}$ durchläuft. Ist $f(x)$ im Punkt x_0 aus $\mathcal{X}$ und $\varphi(y)$ in dem entsprechenden Punkt $y_0 = f(x_0)$ aus $\mathcal{Y}$ stetig, so ist auch die zusammengesetzte Funktion $\varphi(f(x))$ in x_0 stetig.*

Beweis. Es sei $\varepsilon > 0$ vorgegeben. Da $\varphi(y)$ in $y = y_0$ stetig ist, gibt es zu diesem ε ein $\sigma > 0$ derart, daß aus $|y - y_0| < \sigma$ die Beziehung $|\varphi(y) - \varphi(y_0)| < \varepsilon$ folgt.

Da andererseits $f(x)$ in $x = x_0$ stetig ist, gibt es zu diesem σ ein $\delta > 0$ derart, daß aus $|x - x_0| < \delta$ die Beziehung $|f(x) - f(x_0)| = |f(x) - y_0| < \sigma$ folgt. Auf Grund der Wahl von σ folgt hieraus also für $|x - x_0| < \delta$

$$|\varphi(f(x)) - \varphi(y_0)| = |\varphi(f(x)) - \varphi(f(x_0))| < \varepsilon.$$

Damit ist in der ε-δ-Sprache die Stetigkeit von $\varphi(f(x))$ in x_0 bewiesen.

---

[1]) Im Fall $\mu > 0$ kann man den Wert 0 zum Variationsbereich von x hinzunehmen, ebenso dann zum Wertebereich $\mathcal{Y}$; für $\mu < 0$ muß man den Wert 0 ausschließen. Ist ferner μ eine ganze Zahl $\pm n$ oder ein Bruch $\pm \frac{p}{q}$ mit ungeradem Nenner, so kann man die Potenzfunktion auch für $x < 0$ betrachten. Die Stetigkeit für diese Werte läßt sich ganz analog beweisen. (Vgl. die Bemerkung auf S. 100).

Stellt man beispielsweise die Potenzfunktion x^μ $(x > 0)$ als zusammengesetzte Funktion in der Form

$$x^\mu = e^{\mu \ln x}$$

dar, also durch Verkettung der Logarithmus- und der Exponentialfunktion, so ergibt sich ihre Stetigkeit aus der Stetigkeit der letztgenannten Funktionen.

74. Lösung einer Funktionalgleichung. Zur Erleichterung der Darlegungen der nächsten Nummern befassen wir uns jetzt mit folgender Aufgabe, die auch für sich von Interesse ist[1]): *Man bestimme alle im Intervall* $(-\infty, \infty)$ *stetigen Funktionen* $f(x)$, *die für beliebige* x *und* y *der Bedingung*

$$f(x + y) = f(x) + f(y) \tag{A}$$

genügen.

Die Gleichung (A) ist das einfachste Beispiel einer sogenannten *Funktionalgleichung*, welche also für eine gesuchte Funktion eine gewisse Eigenschaft zum Ausdruck bringt, die diese Funktion zu bestimmen gestattet. Wir sollen also alle *stetigen Lösungen* von (A) ermitteln.

Man sieht leicht, daß die lineare Funktion

$$f(x) = cx \quad (c = \text{const}) \tag{a}$$

der Gleichung

$$c(x + y) = cx + cy$$

genügt. Die Frage besteht also weiter darin, ob dies (für beliebige c) die einzigen stetigen Lösungen von (A) sind.

Um zu zeigen, daß dies wirklich der Fall ist, nehmen wir an, wir hätten eine stetige Funktion $f(x)$ gefunden, welche die Gleichung (A) befriedigt, und zeigen, daß sie notwendigerweise die Gestalt (a) hat.

Zunächst läßt sich durch Induktion die Beziehung (A) leicht auf den Fall von n Summanden verallgemeinern:

$$f(\underbrace{x + y + \cdots + z}_{n}) = f(x) + f(y) + \cdots + f(z). \tag{4}$$

Nimmt man sie nämlich für irgendein $n \geqq 2$ als gültig an, so folgt für $n + 1$ Summanden

$$f(\underbrace{x + y + \cdots + z}_{n} + u) = f(\underbrace{x + y + \cdots + z}_{n}) + f(u)$$
$$= [f(x) + f(y) + \cdots + f(z)] + f(u).$$

Setzt man in (4) nun $x = y = \cdots = z$, so folgt

$$f(nx) = nf(x). \tag{5}$$

Ersetzt man x durch $\frac{1}{n}x$, so ergibt sich $f\left(\frac{1}{n}x\right) = \frac{1}{n}f(x)$, also, wenn $x = mz$ für natürliches m gesetzt wird,

$$f\left(\frac{m}{n}z\right) = \frac{m}{n}f(z). \tag{6}$$

Aus (A) erhalten wir ferner für $x = y = 0$

$$f(0) = 2f(0), \tag{7}$$

[1]) Über die Bedeutung von Funktionalgleichungen in der Mathematik und den Anwendungen vgl. der fortgeschrittene Leser etwa J. Aczél, Vorlesungen über Funktionalgleichungen und ihre Anwendungen, Berlin bzw. Basel/Stuttgart 1961. — *Anm. d. Red.*

also $f(0) = 0$. Setzt man jetzt $y = -x$ und berücksichtigt (7), so findet man $f(-x) = -f(x)$, so daß $f(x)$ mit x das Vorzeichen wechselt. Nun folgt aus (5)

$$f(-nx) = -f(nx) = -nf(x) \tag{8}$$

und allgemein aus (6)

$$f\left(-\frac{m}{n}x\right) = -\frac{m}{n}f(x). \tag{9}$$

Die Beziehungen (5) bis (9) lassen sich in der Form

$$f(rx) = rf(x)$$

zusammenfassen, die für jedes reelle x und jedes *rationale* r gilt. Setzt man hierin $x = 1$ und bezeichnet $f(1)$ mit c, so ergibt sich

$$f(r) = cr.$$

Damit haben wir die Gestalt der Funktion f wenigstens für rationale Argumente bestimmt. Dabei wurde lediglich die Bedingung (A), aber noch nicht die vorausgesetzte Stetigkeit von f benutzt.

Es sei nun ϱ ein *irrationaler* Argumentwert. Dann wählen wir eine gegen ϱ konvergierende Folge rationaler Zahlen

$$r_1, r_2, \ldots, r_n, \ldots,$$

beispielsweise die Folge der die Zahl ϱ approximierenden endlichen Dezimalbrüche. Nach dem Obigen ist

$$f(r_n) = cr_n \quad (n = 1, 2, 3, \ldots).$$

Gehen wir für $n \to \infty$ zur Grenze über, so erhalten wir rechts $c\varrho$, links, da f stetig sein sollte,

$$\lim f(r_n) = f(\varrho),$$

also in der Tat

$$f(\varrho) = c\varrho.$$

Somit hat die gesuchte Funktion für alle reellen Argumente die Gestalt (a). Diese Formel liefert also die allgemeinste Lösung der Gleichung (A) in der Klasse der stetigen Funktionen.

75. Charakterisierung der Exponential-, der Logarithmus- und der Potenzfunktion durch Funktionalgleichungen.

1. Ist

$$f(x) = a^x \quad (a > 0), \tag{b}$$

so gilt für alle reellen x und y die Beziehung

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

entsprechend der wohlbekannten Potenzregel

$$a^{x+y} = a^x a^y.$$

Es zeigt sich, daß in der Klasse der stetigen Funktionen die Bedingung (B) die Exponentialfunktion vollständig charakterisiert. Genauer:

Die einzige in $(-\infty, \infty)$ *definierte und dort stetige Funktion, die der Bedingung* (B) *genügt, ist* [abgesehen von der Funktion $f(x) \equiv 0$] *die Exponentialfunktion.*

Mit anderen Worten: Die Formel (b) liefert (mit der erwähnten Ausnahme) die allgemeinste Lösung der *Funktionalgleichung* (B) in der Klasse der stetigen Funktionen.

Zum Beweis betrachten wir eine von $f(x) \equiv 0$ verschiedene, überall definierte und stetige Funktion $f(x)$, für die (B) erfüllt ist.

Es gibt also ein x_0 mit $f(x_0) \neq 0$. Setzen wir in (B) für y den Wert $x_0 - x$, so folgt

$$f(x)\,f(x_0 - x) = f(x_0) \neq 0.$$

Also ist $f(x)$ für jedes x von 0 verschieden. Für $x = y = \frac{z}{2}$ folgt aus (B) die Beziehung

$$f(z) = \left[f\left(\frac{z}{2}\right)\right]^2,$$

und es zeigt sich, daß $f(x)$ überall positiv ist. Daher können wir (B) logarithmieren (etwa zur Basis e):

$$\ln f(x + y) = \ln f(x) + \ln f(y).$$

Setzt man

$$\varphi(x) = \ln f(x),$$

so ist φ als mittelbare Funktion stetig (Nr. 73) und genügt offenbar der Bedingung

$$\varphi(x + y) = \varphi(x) + \varphi(y),$$

deren allgemeine Lösung nach dem Vorhergehenden durch

$$\varphi(x) = \ln f(x) = cx \quad (c = \text{const})$$

geliefert wird. Hieraus folgt $f(x) = \mathrm{e}^{cx} = (\mathrm{e}^c)^x = a^x$, was zu beweisen war.

2. Ist

$$f(x) = \log_a x \quad (a > 0, a \neq 1), \tag{c}$$

so gilt für alle positiven x und y

$$f(xy) = f(x) + f(y). \tag{C}$$

Das entspricht der Regel für die Logarithmierung eines Produkts:

$$\log_a xy = \log_a x + \log_a y.$$

Diese Beziehung reicht — zusammen mit der Stetigkeit — zur Charakterisierung der Logarithmusfunktion aus:

Die einzige im Intervall $(0, \infty)$ *definierte und dort stetige Funktion, die der Bedingung* (C) *genügt, ist neben der Funktion* $f(x) \equiv 0$ *die Logarithmusfunktion, so daß die allgemeinste Lösung der Funktionalgleichung* (C) *die Funktion* (c) *ist.*

Zum Beweis betrachten wir eine für $x > 0$ stetige Funktion, die (C) genügt. Wir führen durch $x = \mathrm{e}^\xi$, also $\xi = \ln x$, eine im Intervall $(-\infty, \infty)$ variierende Variable ξ ein. Ferner setzen wir $\varphi(\xi) = f(\mathrm{e}^\xi)$, also $f(x) = \varphi(\ln x)$. Die nach Nr. 73 stetige Funktion $\varphi(\xi)$ genügt der Bedingung [vgl. (C)]

$$\varphi(\xi + \eta) = f(\mathrm{e}^{\xi+\eta}) = f(\mathrm{e}^\xi \mathrm{e}^\eta) = f(\mathrm{e}^\xi) + f(\mathrm{e}^\eta) = \varphi(\xi) + \varphi(\eta),$$

also einer Bedingung von Typ (A). Somit ist $\varphi(\xi) = c\xi$, also $f(x) = c \ln x$, Wegen $f(x) \not\equiv 0$ ist $c = 0$ ausgeschlossen; also kann man $a = \mathrm{e}^{1/c}$ setzen und

$$f(x) = \log_a x$$

schreiben. Damit ist alles bewiesen.

3. Für

$$f(x) = x^\mu \tag{d}$$

gilt offenbar die Funktionalgleichung

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

für alle positiven x und y, denn es ist $(xy)^\mu = x^\mu y^\mu$. Diese Beziehung charakterisiert (zusammen mit der Stetigkeit) die Potenzfunktion:

Die einzige von $f(x) \equiv 0$ *verschiedene, in* $(0, \infty)$ *definierte und dort stetige Funktion, die der Beziehung* (D) *genügt, ist die Potenzfunktion.*

Ist nämlich $f(x)$ eine für $x > 0$ stetige Funktion, die (D) genügt, so machen wir dieselbe Substitution $x = \mathrm{e}^\xi$ wie oben. Dann folgt für $\varphi(\xi)$

$$\varphi(\xi + \eta) = f(\mathrm{e}^{\xi+\eta}) = f(\mathrm{e}^\eta\, \mathrm{e}^\eta) = f(\mathrm{e}^\xi)\, f(\mathrm{e}^\eta) = \varphi(\xi)\, \varphi(\eta),$$

also eine Beziehung vom Typ (B). Wir wissen aber, daß, vom trivialen Fall abgesehen,

$$\varphi(\xi) = a^{\xi} \quad (a > 0)$$

ist. Daraus folgt

$$f(x) = a^{\ln x} = x^{\mu}$$

mit $\mu = \ln a$, was zu beweisen war.

76. Funktionalgleichungen der trigonometrischen Funktionen und der Hyperbelfunktionen.

4. Ist

$$f(x) = \cos ax \quad \text{bzw.} \quad f(x) = \cosh ax \quad (a \geqq 0), \tag{e}$$

so gilt für reelle beliebige x, y

$$f(y + x) + f(y - x) = 2f(x)\, f(y). \tag{E}$$

Das folgt sofort aus den Additionstheoremen

$$\cos(y \pm x) = \cos x \cos y \mp \sin x \sin y,$$

$$\cosh(y \pm x) = \cosh x \cosh y \pm \sinh x \sinh y$$

(vgl. Nr. 48, Funktionenklasse 6). Die Funktionalgleichung (E) genügt (zusammen mit der Stetigkeit) zur vollständigen Charakterisierung der beiden Kosinusfunktionen:

Die einzigen in $(-\infty, \infty)$ *definierten und dort stetigen Funktionen, die der Funktionalgleichung* (E) *genügen, sind der trigonometrische und der hyperbolische Kosinus* (e) [außer der Funktion $f(x) \equiv 0$].

Es sei also $f(x)$ stetig für alle x und genüge (E). Wir setzen $x = 0$ und nehmen für y einen Wert, für den $f(y) \neq 0$ ist; dann folgt

$$f(0) = 1. \tag{10}$$

Für $y = 0$ ergibt sich

$$f(-x) = f(x), \tag{11}$$

also ist $f(x)$ eine gerade Funktion.

Da die stetige Funktion $f(x)$ für $x = 0$ positiv ist, gibt es eine positive Zahl c derart, daß $f(x)$ im ganzen Intervall $[0, c]$ positiv ist. Die folgenden Überlegungen verlaufen auf zwei verschiedenen Wegen, je nachdem, ob (α) $f(c) \leqq 1$ oder (β) $f(c) > 1$ ist. Wir beginnen mit dem Fall (α).

Wegen $0 < f(c) \leqq 1$ gibt es ein θ mit $0 \leqq \theta < \dfrac{\pi}{2}$ derart, daß

$$f(c) = \cos \theta \tag{12}$$

ist. Nun schreiben wir (E) in der Form

$$f(y + x) = 2f(x)\, f(y) - f(y - x)$$

und setzen nacheinander

$$x = c, \quad y = c; \qquad x = c, \quad y = 2c; \qquad x = c, \quad y = 3c;$$

usw. Unter Berücksichtigung von (10) und (12) erhalten wir

$$f(2c) = 2\cos^2\theta - 1 = \cos 2\theta,$$

$$f(3c) = 2\cos\theta \cdot \cos 2\theta - \cos\theta = \cos 3\theta,$$

$$f(4c) = 2\cos\theta \cdot \cos 3\theta - \cos 2\theta = \cos 4\theta,$$

usw. Durch Induktion folgt für jedes natürliche m

$$f(mc) = \cos m\theta. \tag{13}$$

Setzt man in (E) nun $x = y = \frac{1}{2}c$, so erhält man, wieder unter Berücksichtigung von (10) und (12),

$$\left[f\left(\frac{1}{2}c\right)\right]^2 = \frac{f(0) + f(c)}{2} = \frac{1 + \cos\theta}{2} = \left[\cos\frac{1}{2}\theta\right]^2;$$

da $f(x)$ zwischen 0 und c und $\cos x$ zwischen 0 und θ positiv ist, kann man auf beiden Seiten die Wurzel ziehen und erhält

$$f\left(\frac{1}{2}c\right) = \cos\frac{1}{2}\theta.$$

Ebenso findet man aus (E) für $x = y = \frac{1}{2^2}c$

$$f\left(\frac{1}{2^2}c\right) = \cos\frac{1}{2^2}\theta$$

usw. Durch Induktion folgt

$$f\left(\frac{1}{2^n}c\right) = \cos\frac{1}{2^n}\theta \qquad (n = 1, 2, 3, \ldots). \tag{14}$$

Durch Wiederholung des Verfahrens, durch das wir von (12) zu (13) gelangten, kommen wir von (14) zu

$$f\left(\frac{m}{2^n}c\right) = \cos\frac{m}{2^n}\theta.$$

Für positive x-Werte der Form $\frac{m}{2^n}$ ist also

$$f(cx) = \cos\theta x. \tag{15}$$

Nun kann man aber jede positive Zahl als Limes von Zahlen dieser Form erhalten (als endlichen oder unendlichen „Dualbruch"), so daß man durch Grenzübergang [unter Berücksichtigung der Stetigkeit von $f(x)$ und $\cos x$] die Gültigkeit der Formel (15) für alle $x > 0$ nachgewiesen hat. Für $x < 0$ folgt sie aus (11) und für $x = 0$ aus (10). Ersetzt man in (15) x durch $\frac{x}{c}$ und setzt $\frac{\theta}{c} = a$, so ergibt sich schließlich

$$f(x) = \cos ax.$$

Im Fall (β) gilt $f(c) > 1$. Dann gibt es ein θ mit

$$f(c) = \cosh\theta.$$

Durch Wiederholung der Überlegungen erhält man unter Berücksichtigung der bekannten Formeln für den hyperbolischen Kosinus

$$f(x) = \cosh ax \qquad (a > 0).$$

Für $a = 0$ ergibt sich in beiden Fällen $f(x) \equiv 1$.

Die Funktionalgleichungen (A) bis (E) wurden zuerst von Cauchy untersucht, der auch die stetigen Lösungen fand.

77. Berechnung von Grenzwerten mit Hilfe der Stetigkeit von Funktionen. Die Stetigkeit von Funktionen wird vielfach zur Berechnung von Grenzwerten benutzt.[1])

Wir wollen nun einige Beispiele dieser Art behandeln.

---

[1]) Tatsächlich haben wir das ebenfalls schon getan. Beispielsweise haben wir in Nr. 56, Beispiel 3, die Stetigkeit von $\sqrt[m]{x}$ für $x = 1$ nachgewiesen und sie benutzt, und in Beispiel 5 (b) haben wir dasselbe für $\cos x$ in $x = 0$ durchgeführt.

1. Für jedes reelle x gilt

$$\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = e^x.$$

In der Tat kann man den betrachteten Ausdruck (wenn $x \neq 0$ vorausgesetzt wird) in der Form

$$\left[\left(1+\frac{x}{n}\right)^{n/x}\right]^x$$

darstellen. Da $\frac{x}{n}$ gegen 0 strebt, konvergiert die diskrete Veränderliche in der eckigen Klammer gegen e [vgl. Nr. 54, Formel (13)]; dann hat auf Grund der Stetigkeit der Potenzfunktion (hier ist x konstant) der ganze Ausdruck den Grenzwert e^x.

2. Man bestimme den Grenzwert

$$\lim_{x\to\infty}\left\{\sqrt[k]{(x+a_1)(x+a_2)\cdots(x+a_k)}-x\right\} \qquad (\infty-\infty),$$

wobei $a_1, a_2, \ldots, a_k$ gegebene konstante Zahlen sind.

Wir benutzen die Identität

$$y-z=\frac{y^k-z^k}{y^{k-1}+y^{k-2}z+\cdots+z^{k-1}}$$

und setzen hierin $y=\sqrt[k]{(x+a_1)\cdots(x+a_k)}$ und $z=x$. Dann läßt sich der betrachtete Ausdruck in der Gestalt

$$\frac{(x+a_1)\cdots(x+a_k)-x^k}{\left(\sqrt[k]{\cdots}\right)^{k-1}+x\left(\sqrt[k]{\cdots}\right)^{k-2}+\cdots+x^{k+1}}$$

$$=\frac{(a_1+\cdots+a_k)+\dfrac{a_1a_2+\cdots+a_{k-1}a_k}{x}+\cdots+\dfrac{a_1a_2\cdots a_k}{x^{k-1}}}{\left(\sqrt[k]{\left(1+\dfrac{a_1}{x}\right)\cdots\left(1+\dfrac{a_k}{x}\right)}\right)^{k-1}+\cdots+1}$$

darstellen. Für $x\to\infty$ strebt der unter der Wurzel stehende Ausdruck gegen 1; folglich hat die Wurzel selbst den Grenzwert $\sqrt[k]{1}=1$, da die Wurzel als Spezialfall der Potenzfunktion stetig ist. Da das im Nenner stehende Polynom $(k-1)$-ten Grades ebenfalls eine stetige Funktion ist, strebt der Nenner gegen k, und der Grenzwert des Bruches wird

$$\frac{a_1+a_2+\cdots+a_k}{k}.$$

3. Wir wenden uns wieder dem Satz aus Nr. 33, Beispiel 13, zu. Es sei $a_n>0$, und es gelte $a_n\to a$; wir beschränken uns auf $0<a<\infty$. Nun wenden wir den erwähnten Satz auf die Folge $\{\ln a_n\}$ an.

Wegen $\ln a_n\to\ln a$ (da die Logarithmusfunktion stetig ist) gilt

$$\lim\ln\sqrt[n]{a_1\cdots a_n}=\lim\frac{\ln a_1+\cdots+\ln a_n}{n}=\ln a.$$

Da die Exponentialfunktion stetig ist, ergibt sich

$$\sqrt[n]{a_1\ldots a_n}=e^{\ln\sqrt[n]{a_1\ldots a_n}}\to e^{\ln a}=a.$$

Mit Hilfe der Grenzwerte aus Nr. 54, Beispiel 1 und 2, läßt sich dieses Resultat auch auf die Fälle $a=0$ und $a=\infty$ erweitern.

Somit erhalten wir folgende Fassung des erwähnten Satzes:

Wenn eine positive diskrete Veränderliche a_n einen (endlichen oder unendlichen) Grenzwert hat, so hat die diskrete Veränderliche

$$b_n = \sqrt[n]{a_1 \cdot a_2 \cdots a_n}$$

denselben Grenzwert.

4. Wenn wir diesen Satz auf die Folge

$$a_1, \frac{a_2}{a_1}, \frac{a_3}{a_2}, \ldots, \frac{a_n}{a_{n-1}}, \frac{a_{n+1}}{a_n}, \ldots$$

anwenden, so erhalten wir die interessante Folgerung

$$\lim \sqrt[n]{a_n} = \lim \frac{a_{n+1}}{a_n},$$

unter der einzigen Voraussetzung, daß der zweite dieser Grenzwerte existiert.

Als Beispiel bestimmen wir den Grenzwert

$$\lim \frac{\sqrt[n]{n!}}{n}.$$

Wenn wir $a_n = \frac{n!}{n^n}$ setzen, erhalten wir

$$\frac{a_{n+1}}{a_n} = \frac{(n+1)!}{(n+1)^{n+1}} : \frac{n!}{n^n} = \frac{1}{\left(1 + \frac{1}{n}\right)^n} \to \frac{1}{\mathrm{e}}.$$

Somit ist der gesuchte Grenzwert gleich $\frac{1}{\mathrm{e}}$.

5. Wir bestimmen eine Reihe *wichtiger Grenzwerte*, die wir im folgenden Kapitel benutzen werden:

(a) $$\lim_{\alpha \to 0} \frac{\log_a (1 + \alpha)}{\alpha} = \log_a \mathrm{e} \qquad \left(\frac{0}{0}\right),$$

(b) $$\lim_{\alpha \to 0} \frac{a^\alpha - 1}{\alpha} = \ln a \qquad \left(\frac{0}{0}\right),$$

(c) $$\lim_{\alpha \to 0} \frac{(1 + \alpha)^\mu - 1}{\alpha} = \mu \qquad \left(\frac{0}{0}\right).$$

Es gilt

$$\frac{\log_a (1 + \alpha)}{\alpha} = \log_a (1 + \alpha)^{1/\alpha}.$$

Da der auf der rechten Seite unter dem Logarithmus stehende Ausdruck für $\alpha \to 0$ gegen e strebt [Nr. 54, Formel (13)], strebt (infolge der Stetigkeit der Logarithmusfunktion) ihr Logarithmus gegen $\log_a \mathrm{e}$, was zu beweisen war.

Als Spezialfall der soeben bewiesenen Formel vermerken wir noch den Fall, daß es sich um den *natürlichen* Logarithmus ($a = \mathrm{e}$) handelt:

$$\lim_{\alpha \to 0} \frac{\ln (1 + \alpha)}{\alpha} = 1.$$

In der Einfachheit dieser Beziehung liegt im Grunde der Vorteil, den das System der natürlichen Logarithmen bietet.

Wir wenden uns jetzt der Formel (b) zu und setzen $a^\alpha - 1 = \beta$; für $\alpha \to 0$ gilt dann auf Grund der Stetigkeit der Exponentialfunktion auch $\beta \to 0$.

Ferner gilt $\alpha = \log_a(1+\beta)$, so daß sich unter Benutzung der schon bewiesenen Ergebnisse die Beziehung

$$\lim_{\alpha\to 0} \frac{a^\alpha - 1}{\alpha} = \lim_{\beta\to 0} \frac{\beta}{\log_a(1+\beta)} = \frac{1}{\log_a \mathrm{e}} = \ln a$$

ergibt, die zu beweisen war.

Wenn wir insbesondere $\alpha = \frac{1}{n}$ $(n = 1, 2, 3, \ldots)$ setzen, so erhalten wir die interessante Formel

$$\lim_{n\to\infty} n\left(\sqrt[n]{a} - 1\right) = \ln a \quad (\infty \cdot 0).$$

Schließlich setzen wir zum Beweis der Formel (c)

$$(1+\alpha)^\mu - 1 = \beta.$$

Für $\alpha \to 0$ gilt (auf Grund der Stetigkeit der Potenzfunktion) auch $\beta \to 0$. Wenn wir die Beziehung $(1+\alpha)^\mu = 1 + \beta$ logarithmieren, erhalten wir

$$\mu \ln(1+\alpha) = \ln(1+\beta).$$

Mit Hilfe dieser Beziehung formen wir den Ausdruck folgendermaßen um:

$$\frac{(1+\alpha)^\mu - 1}{\alpha} = \frac{\beta}{\alpha} = \frac{\beta}{\ln(1+\beta)} \cdot \mu \cdot \frac{\ln(1+\alpha)}{\alpha}.$$

Nach dem Bewiesenen streben die beiden Quotienten

$$\frac{\beta}{\ln(1+\beta)} \quad \text{und} \quad \frac{\ln(1+\alpha)}{\alpha}$$

gegen 1, so daß das Produkt den Grenzwert μ hat, was zu beweisen war.

Der in Nr. 56, Beispiel 3, betrachtete Grenzwert ergibt sich hieraus als Spezialfall für $\mu = r$.

78. Potenz-Exponentialausdrücke. Wir betrachten jetzt den Potenz-Exponentialausdruck u^v, wobei $u > 0$ und v Funktionen von ein und derselben Veränderlichen x in dem Variationsbereich $\mathcal{X}$ sind, der den Häufungspunkt x_0 hat; insbesondere können das zwei Folgen $\{u_n\}$ und $\{v_n\}$ sein.

Es mögen die endlichen Grenzwerte

$$\lim_{x\to x_0} u = a \quad \text{und} \quad \lim_{x\to x_0} v = b$$

existieren, wobei $a > 0$ ist. Es ist der Grenzwert des Ausdrucks u^v zu bestimmen.

Wir stellen den Ausdruck in der Form

$$u^v = \mathrm{e}^{v \ln u}$$

dar. Die Funktionen v und $\ln u$ haben die Grenzwerte

$$\lim_{x\to x_0} v = b, \quad \lim_{x\to x_0} \ln u = \ln a$$

(hier wird die Stetigkeit der Logarithmusfunktion benutzt), so daß

$$\lim_{x \to x_0} v \ln u = b \ln a$$

ist. Hieraus folgt auf Grund der Stetigkeit der Exponentialfunktion schließlich

$$\lim_{x \to x_0} u^v = e^{b \ln a} = a^b.$$

Den Grenzwert des Ausdrucks u^v kann man immer in den Fällen bestimmen, in denen der (endliche oder unendliche) Grenzwert c des Produkts $v \ln u$ bekannt ist. Für endliches c ist der gesuchte Grenzwert offenbar e^c; ist jedoch $c = -\infty$ oder ∞, so wird dieser Grenzwert 0 bzw. ∞ (Nr. 54, Beispiel 1).

Die Bestimmung des Grenzwertes $c = \lim \{v \ln u\}$ ist bei gegebenen Grenzwerten a und b immer möglich, mit Ausnahme der Fälle, in denen dieses Produkt (für $x \to x_0$) ein unbestimmter Ausdruck der Form $\infty \cdot 0$ ist. Man kann sich leicht überlegen, daß die Ausnahmefälle folgenden Kombinationen der Werte a und b entsprechen:

$$a = 1, \quad b = \pm\infty,$$

$$a = 0, \quad b = 0,$$

$$a = \infty, \quad b = 0.$$

In diesen Fällen sagt man, der Ausdruck u^v sei ein *unbestimmter Ausdruck der Form* 1^∞, 0^0, ∞^0.[1]) Um hier die Frage nach dem Grenzwert des Ausdrucks u^v zu beantworten, genügt es nicht, die Grenzwerte der Funktionen u und v zu kennen, sondern man muß außerdem die Vorschrift berücksichtigen, nach der die Variablen selbst gegen ihre Grenzwerte streben.

Die diskrete Veränderliche $\left(1 + \frac{1}{n}\right)^n$ liefert für $n \to \infty$ [bzw. der allgemeinere Ausdruck $(1 + \alpha)^{1/\alpha}$ für $\alpha \to 0$] ein Beispiel für einen unbestimmten Ausdruck der Form 1^∞. Der Grenzwert ist in beiden Fällen e. In Nr. 77, Beispiel 4, haben wir die diskrete Veränderliche $\sqrt[n]{\frac{n!}{n^n}} = \left(\frac{n!}{n^n}\right)^{1/n}$ betrachtet, die einen unbestimmten Ausdruck der Form 0^0 liefert. Schließlich ergibt die in Nr. 32, Beispiel 10, untersuchte Variable $\sqrt[n]{n}$ einen unbestimmten Ausdruck der Form ∞^0.

Wir geben nun noch einige Beispiele für die Auswertung unbestimmter Ausdrücke anderer Form an.

79. Beispiele.

1. Man bestimme

$$\lim_{x \to \infty} (\ln x)^{1/x} \quad (\infty^0).$$

Wenn wir den gegebenen Ausdruck mit y bezeichnen, erhalten wir (vgl. Nr. 54, Beispiel 2 und 5)

$$\ln y = \frac{\ln(\ln x)}{x} = \frac{\ln(\ln x)}{\ln x} \cdot \frac{\ln x}{x} \to 0 \quad \left(\frac{\infty}{\infty}\right),$$

so daß $y \to e^0 = 1$ gilt.

---

[1]) Zu diesen Symbolen könnte man das in der Fußnote auf S. 61 Gesagte wiederholen.

2. Man bestimme

$$\lim_{x \to 0} x^{\sin x} \qquad (0^0).$$

Hier ist (vgl. Nr. 54, Beispiel 7 und 5) $\ln y = \sin x \cdot \ln x = \frac{\sin x}{x} \cdot x \ln x \to 0$, also wieder $y \to 1$.

3. Jetzt kann man das Beispiel 1 aus Nr. 76 folgendermaßen verallgemeinern: Strebt x_n gegen x (x endlich), so ist

$$\lim_{n \to \infty} \left(1 + \frac{x_n}{n}\right)^n = e^x \quad (1^\infty).$$

Zum Beweis genügt es, diesen Ausdruck in der Form

$$\left[\left(1 + \frac{x_n}{n}\right)^{n/x_n}\right]^{x_n}$$

darzustellen; die Basis strebt hier gegen e, der Exponent gegen x.

4. Hierauf kann man das folgende Beispiel zurückführen:

$$\lim_{n \to \infty} \left(\cos \frac{x}{n} + \lambda \sin \frac{x}{n}\right)^n = e^{\lambda x} \quad (1^\infty).$$

Setzen wir den Ausdruck in der Klammer gleich $1 + \frac{x_n}{n}$, so erhalten wir

$$x_n = n \cdot \left[\cos \frac{x}{n} - 1 + \lambda \sin \frac{x}{n}\right] = \lambda x \cdot \frac{\sin \frac{x}{n}}{\frac{x}{n}} - x \cdot \frac{1 - \cos \frac{x}{n}}{\frac{x}{n}} \to \lambda x$$

usw.

5. Analog läßt sich das Beispiel

$$\lim_{n \to \infty} \left(\frac{\sqrt[n]{a} + \sqrt[n]{b}}{2}\right)^n = \sqrt{ab} \quad (1^\infty)$$

für $a, b > 0$ erledigen. Hier setzen wir

$$x_n = n \cdot \left[\frac{\sqrt[n]{a} + \sqrt[n]{b}}{2} - 1\right] = \frac{1}{2}\left[n\left(\sqrt[n]{a} - 1\right) + n\left(\sqrt[n]{b} - 1\right)\right],$$

so daß sich auf Grund einer speziellen Folgerung aus Nr. 77, Beispiel 5 (b),

$$x_n \to \frac{1}{2}(\ln a + \ln b) = \ln \sqrt{ab}$$

ergibt; der gesuchte Grenzwert ist tatsächlich gleich $e^{\ln \sqrt{ab}} = \sqrt{ab}$.

6. Schließlich betrachten wir noch den Grenzwert

$$\lim_{x \to 0} (\cos x)^{\frac{1}{\sin^2 x}} = \lim_{x \to 0} \left[\left(1 - 2 \sin^2 \frac{x}{2}\right)^{-\frac{1}{2\sin^2 \frac{x}{2}}}\right]^{-\frac{1}{2\cos^2 \frac{x}{2}}} = e^{-\frac{1}{2}} = \frac{1}{\sqrt{e}} \quad (1^\infty).$$

Der Leser sieht, daß man im Fall eines unbestimmten Ausdrucks der Form 1^∞ zweckmäßigerweise immer auf die Exponentialform zurückgreift.

Wie schon gesagt, werden allemeine Methoden zur Auswertung unbestimmter Ausdrücke aller Formen in Kapitel IV (§ 4) angegeben.

§ 5. Eigenschaften der stetigen Funktionen

80. Der erste Zwischenwertsatz. Wir befassen uns jetzt mit der Untersuchung einiger grundlegender Eigenschaften von Funktionen, die in einem Intervall stetig sind. Diese Eigenschaften sind nicht nur an sich interessant, sondern liefern auch die Grundlage verschiedener Überlegungen, die wir später noch anzustellen haben.

Wir beginnen mit folgendem einfachen Satz, den man BOLZANO und CAUCHY verdankt.

Erster Satz von BOLZANO-CAUCHY. *Die Funktion $f(x)$ sei im abgeschlossenen Intervall $[a, b]$ definiert und stetig und habe in den Endpunkten Werte verschiedenen Vorzeichens. Dann gibt es einen Punkt c zwischen a und b, in dem die Funktion den Wert* 0 *annimmt:*

$$f(c) = 0 \quad (a < c < b).$$

Dieser Satz hat eine sehr einfache geometrische Bedeutung: Verläuft eine stetige Kurve auf verschiedenen Seiten der x-Achse, so muß sie diese *schneiden* (Abb. 31).

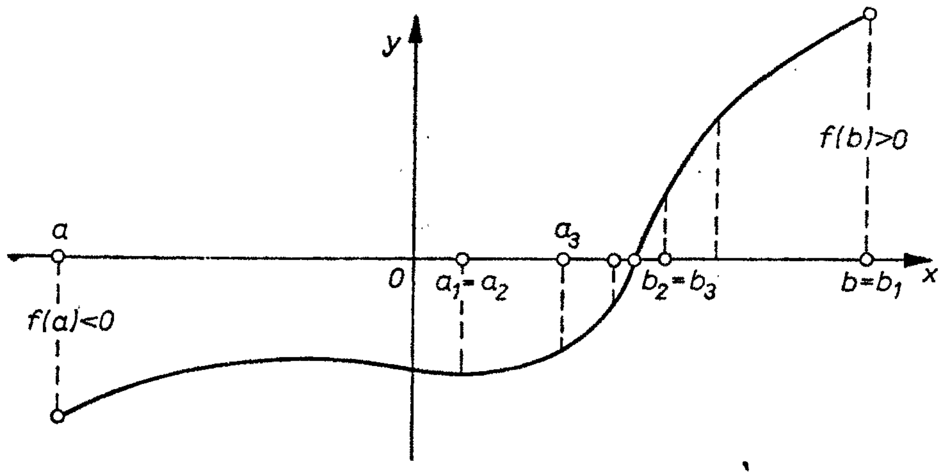

Abb. 31

Erster Beweis. Einen ersten Beweis führen wir nach der Bolzanoschen Methode (Nr. 41), durch Intervallschachtelung[1]). Um etwas Bestimmtes vor Augen zu haben, nehmen wir an, es sei $f(a) < 0$, $f(b) > 0$. Dann halbieren wir das Intervall; der Mittelpunkt ist offenbar $\frac{a+b}{2}$. Es kann sein, daß $f(x)$ dort den Wert 0 hat; dann ist $c = \frac{a+b}{2}$. Ist aber $f\left(\frac{a+b}{2}\right) \neq 0$, so muß $f(x)$ in den Endpunkten mindestens eines der Intervalle $\left[a, \frac{a+b}{2}\right]$, $\left[\frac{a+b}{2}, b\right]$ Werte verschiedenen Vorzeichens annehmen, und zwar den negativen am linken Ende, den positiven am rechten Ende. Bezeichnen wir dieses Intervall mit $[a_1, b_1]$, so ist also

$$f(a_1) < 0, \quad f(b_1) > 0.$$

---

[1]) In der Praxis ist dieses Verfahren zeitraubend und unzweckmäßig. In Kapitel IV, § 5, werden wir bessere Verfahren kennenlernen. — *Anm. d. Red.*

Dieses Intervall halbieren wir weiter. Hat $f(x)$ im Mittelpunkt den Wert 0, so sind wir fertig. Sonst bezeichnen wir mit $[a_2, b_2]$ das (bzw. eines der) Intervall(e), für welches

$$f(a_2) < 0, \quad f(b_2) > 0$$

ist. Diesen Prozeß setzen wir fort. Dann kommen wir entweder nach endlich vielen Schritten zu einem Mittelpunkt, in dem $f(x)$ den Wert 0 hat, und der Satz ist bewiesen, oder wir erhalten eine unendliche Intervallschachtelung. Nur auf diesen Fall müssen wir näher eingehen. Im n-ten Intervall $[a_n, b_n]$ $(n = 1, 2, \ldots)$ gilt

$$f(a_n) < 0, \quad f(b_n) > 0, \tag{1}$$

und offenbar ist

$$b_n - a_n = \frac{b-a}{2^n}. \tag{2}$$

Unsere Intervallschachtelung genügt aber dem Intervallschachtelungssatz aus Nr. 38; denn nach (2) ist $\lim (b_n - a_n) = 0$. Daher existiert ein Punkt c aus $[a, b]$, für den

$$\lim a_n = \lim b_n = c$$

ist. Nun zeigen wir, daß für dieses c die Funktion $f(x)$ den Wert 0 annimmt.

Gehen wir hierzu in den Ungleichungen (1) zur Grenze über und benutzen dabei die Stetigkeit von $f(x)$ im Punkt $x = c$, so erhalten wir

$$f(c) = \lim f(a_n) \leqq 0 \quad \text{und} \quad f(c) = \lim f(b_n) \geqq 0.$$

Also ist in der Tat $f(c) = 0$, und der Satz ist bewiesen.

Nachstehend geben wir für diesen Satz einen zweiten Beweis, dem ein anderer Gedanke zugrunde liegt. Zunächst erwähnen wir das folgende triviale Lemma.

Lemma. *Ist $f(x)$ in $x = x_0$* stetig *und ist $f(x_0) \neq 0$, so hat $f(x)$ in allen dem Punkt x_0 hinreichend benachbarten Punkten x dasselbe Vorzeichen wie in x_0.*

Das folgt aus der Aussage 2 aus Nr. 55, I, wobei der Wert $f(x_0)$ die Rolle des Grenzwertes A der Funktion übernimmt [das geht, da $f(x)$ stetig ist].

Zweiter Beweis. Wir betrachten alle Punkte $x = \bar{x}$ aus $[a, b]$, in denen $f(\bar{x}) < 0$ ist. Zu ihnen gehören beispielsweise der Punkt a und nach dem Lemma noch weitere, zu a benachbarte Punkte. Die Menge $\{\bar{x}\}$ ist durch die Zahl b nach oben beschränkt. Wir können also $c = \sup \{\bar{x}\}$ setzen (vgl. Nr. 11); wir behaupten, daß $f(c) = 0$ ist.

Andernfalls wäre entweder $f(c) < 0$ oder $f(c) > 0$. Im ersten Fall wäre sicher $c < b$; denn nach Voraussetzung ist $f(b) > 0$. Nach dem Lemma gäbe es dann rechts von c noch Werte $\bar{x}$, für die $f(\bar{x}) < 0$ wäre, und das widerspricht der Definition von c als einer oberen Schranke von $\{\bar{x}\}$. Wäre aber $f(c) > 0$, so wäre (wieder nach dem Lemma) auch für Punkte x links von c die Funktion positiv, etwa in $(c - \delta, c]$, und dann wäre wieder c nicht obere Grenze der Menge $\{\bar{x}\}$. Damit ist der Satz bewiesen.

Die Forderung, daß $f(x)$ im abgeschlossenen Intervall $[a, b]$ stetig ist, ist wesentlich: Eine Funktion, die in einem Punkt eine Sprungstelle hat, kann von negativen zu positiven Werten übergehen, ohne den Wert 0 anzunehmen. Das ist beispielsweise für die Funktion $f(x) = [x] - \frac{1}{2}$ der Fall, die den Wert 0 nicht annimmt, obwohl $f(0) = -\frac{1}{2}$, $f(1) = +\frac{1}{2}$ ist. Die Sprungstelle liegt in $x = 1$. Der Leser möge sich dieses Verhalten an einer Skizze veranschaulichen.

81. Anwendung auf die Lösung von Gleichungen. Zunächst kann man mit diesem Satz die Existenz einer Wurzel beweisen. Beispielsweise hat die Gleichung

$$2^x = 4x$$

offenbar die Lösung $x = 4$; aber schon der Nachweis, daß auch nur eine einzige weitere Lösung existiert, ist schwieriger. Die Funktion $f(x) = 2^x - 4x$ hat aber in $x = 0$ den Wert $1 > 0$ und in $x = \frac{1}{2}$ den Wert $\sqrt{2} - 2 < 0$; also nimmt sie als stetige Funktion zwischen 0 und $\frac{1}{2}$ einmal den Wert 0 an.

Ein anderes Beispiel: Wir betrachten eine algebraische Gleichung ungeraden Grades mit reellen Koeffizienten,

$$f(x) \equiv a_0 x^{2n+1} + a_1 x^{2n} + \cdots + a_{2n} x + a_{2n+1} = 0.$$

Für hinreichend große positive x hat das Polynom das Vorzeichen des Gliedes mit dem höchsten Exponenten, d. h. das Vorzeichen von a_0, für absolut hinreichend große, aber negative x das entgegengesetzte Vorzeichen. Da ein Polynom stetig ist, muß es dazwischen einmal den Wert 0 annehmen. *Also hat jede algebraische Gleichung ungeraden Grades mit reellen Koeffizienten mindestens eine reelle Lösung.*

Der Cauchysche Satz läßt sich aber nicht nur zum Nachweis der Existenz, sondern auch zur angenäherten Berechnung von Lösungen benutzen. Wir zeigen das an einem Beispiel. Es sei $f(x) = x^4 - x - 1$. Wegen $f(1) = -1$, $f(2) = 13$ hat das Polynom eine Nullstelle zwischen 1 und 2. Teilen wir das Intervall $[1, 2]$ in zehn gleiche Teile durch die Punkte 1,1; 1,2; 1,3; ... und rechnen die zugehörigen Werte aus, so finden wir

$$f(1{,}1) = -0{,}63\ldots; \quad f(1{,}2) = -0{,}12\ldots; \quad f(1{,}3) = +0{,}55,\ldots;$$

Also liegt eine Nullstelle zwischen 1,2 und 1,3. Durch Weiterteilung finden wir

$$f(1{,}21) = -0{,}06\ldots; \quad f(1{,}22) = -0{,}004\ldots; \quad f(1{,}23) = +0{,}058\ldots;$$

Also liegt eine Lösung zwischen 1,22 und 1,23; wir kennen sie also schon auf 0,01 genau.

Im Hinblick auf diese Bemerkungen ist es interessant, die beiden Beweise für den ersten Zwischenwertsatz zu vergleichen. Der zweite ist ein reiner Existenzbeweis für eine Lösung von $f(x) = 0$ und sagt nichts darüber aus, wie man sie finden kann. Der erste dagegen liefert ein Verfahren zur wirklichen Bestimmung einer Lösung, nämlich die Intervallschachtelung durch Halbierung (woraus wir uns der Einfachheit wegen beschränkten). Damit kann man die gesuchte Lösung der Gleichung in Intervalle beliebig kleiner Länge einschließen, d. h. sie mit beliebiger Genauigkeit berechnen.

82. Zweiter Zwischenwertsatz. Der in Nr. 80 bewiesene Satz läßt sich folgendermaßen verallgemeinern:

Zweiter Satz von Bolzano-Cauchy. *Die Funktion $f(x)$ sei in einem (offenen oder nicht offenen, endlichen oder unendlichen) Intervall definiert und stetig. Nimmt sie in zwei Werten $x = a$, $x = b$ $(a < b)$ die Werte*

$$f(a) = A, \quad f(b) = B \quad (A \neq B)$$

an, so gibt es zu jedem C zwischen A und B ein c zwischen a und b derart, daß

$$f(c) = C \quad (a < c < b)$$

ist.[1])

[1]) Offenbar ist der erste Zwischenwertsatz ein Spezialfall des zweiten: Haben A und B verschiedene Vorzeichen, so kann man $C = 0$ nehmen.

Beweis. Wir nehmen etwa $A < B$, also $A < C < B$ an. Im Intervall $[a, b]$ betrachten wir die Hilfsfunktion $\varphi(x) = f(x) - C$. Sie ist in $[a, b]$ stetig und hat in den Endpunkten Werte verschiedenen Vorzeichens:

$$\varphi(a) = f(a) - C = A - C < 0,$$

$$\varphi(b) = f(b) - C = B - C > 0.$$

Nach dem ersten Zwischenwertsatz gibt es ein c mit $a < c < b$ derart, daß $\varphi(c) = 0$, also

$$f(c) - C = 0 \quad \text{oder} \quad f(c) = C$$

ist, und das war zu beweisen.

Wir haben somit eine wichtige Eigenschaft der in einem Intervall stetigen Funktionen $f(x)$ bewiesen: *Beim Übergang von einem Wert zum anderen nimmt eine solche Funktion jeden dazwischenliegenden Wert mindestens einmal an.*

Diese Eigenschaft kann man auch folgendermaßen formulieren: *Die Werte, die von einer in einem Intervall stetigen Funktion $f(x)$ angenommen werden, wenn x ein Intervall $\mathcal{X}$ durchläuft, füllen ihrerseits ein Intervall $\mathcal{Y}$ lückenlos aus.*

Beweis. Es sei[1] $m = \inf\{f(x)\}$, $M = \sup\{f(x)\}$ und y_0 eine beliebige Zahl zwischen m und M, $m < y_0 < M$. Dann gibt es Funktionswerte $y_1 = f(x_1)$ und $y_2 = f(x_2)$, wobei x_1 und x_2 zu $\mathcal{X}$ gehören, derart, daß

$$m \leqq y_1 < y_0 < y_2 \leqq M$$

gilt; das folgt aus der Definition der oberen bzw. unteren Grenze einer Menge. Dann existiert aber nach unserem Satz eine zwischen x_1 und x_2 liegende Zahl x_0, die offenbar zu $\mathcal{X}$ gehört, derart, daß $f(x_0)$ genau gleich y_0 ist. Daher gehört y_0 zu $\mathcal{Y}$.

Somit ist $\mathcal{Y}$ ein Intervall mit den Endpunkten m und M, die selbst zu $\mathcal{Y}$ gehören können oder auch nicht; vgl. dazu Nr. 84.

In Nr. 71, Satz 2, haben wir gesehen, daß bei monotonen Funktionen aus dieser Eigenschaft die Stetigkeit folgt. Man darf jedoch nicht denken, daß das immer so ist; es lassen sich vielmehr leicht Funktionen konstruieren, die jeden Zwischenwert annehmen, aber nicht stetig sind. Beispielsweise füllen die Werte der Funktion (Nr. 70, Beispiel 4)

$$f(x) = \sin\frac{1}{x} \quad \text{für} \quad x \neq 0, \quad f(0) = 0,$$

wenn x in einem Intervall variiert, das die Unstetigkeitsstelle $x = 0$ enthält, das Intervall $[-1, 1]$ lückenlos aus.

83. Die Existenz der Umkehrfunktion. Wir benutzen nun die in Nr. 82 bewiesene Eigenschaft der stetigen Funktionen, um unter gewissen Voraussetzungen Existenz und Stetigkeit einer eindeutigen Umkehrfunktion nachzuweisen (vgl. Nr. 49).

Satz. *Die Funktion $y = f(x)$ sei in einem Intervall $\mathcal{X}$ definiert, monoton[2]) wachsend (bzw. fallend) und stetig. Dann existiert in dem entsprechenden Intervall $\mathcal{Y}$ der Werte dieser Funktion die eindeutige Umkehrfunktion $x = g(y)$ und ist ebenfalls monoton wachsend (bzw. fallend) und stetig.*

---

[1]) Wir erinnern daran, daß wir in Nr. 11 vereinbart hatten, für den Fall, daß die Menge $\{f(x)\}$ nicht nach oben (bzw. unten) beschränkt ist, $M = \infty$ (bzw. $m = -\infty$) zu setzen.

[2]) Monoton im *engeren* Sinne (das ist hier wesentlich!).

Beweis. Wir beschränken uns auf den Fall einer wachsenden Funktion. Oben haben wir gesehen, daß die Werte der stetigen Funktion $f(x)$ ein Intervall $\mathcal{Y}$ lückenlos ausfüllen, so daß zu jedem y_0 aus $\mathcal{Y}$ mindestens ein x_0 aus $\mathcal{X}$ existiert derart, daß

$$f(x_0) = y_0$$

ist. Da $f(x)$ monoton ist, kann es nur einen einzigen solchen Wert geben: Für $x_1 > x_0$ bzw. $x_1 < x_0$ ist $f(x_1) > f(x_0)$ bzw. $f(x_1) < f(x_0)$.

Wenn wir diesen Wert x_0 dem beliebig gewählten y_0 aus $\mathcal{Y}$ zuordnen, so erhalten wir die eindeutige Funktion

$$x = g(y),$$

die Umkehrfunktion von $y = f(x)$.

Man sieht leicht, daß diese Funktion $g(y)$ ebenso wie $f(x)$ monoton wächst. Es sei

$$y' < y'' \quad \text{und} \quad x' = g(y'), \quad x'' = g(y'').$$

Dann ist nach Definition von $g(y)$ gleichzeitig

$$y' = f(x') \quad \text{und} \quad y'' = f(x'').$$

Wäre $x' > x''$, so wäre, da $f(x)$ wächst, $y' > y''$, entgegen der Voraussetzung. Es kann auch nicht $x' = x''$ sein, denn dann wäre $y' = y''$, was ebenfalls der Voraussetzung widerspricht. Daher ist nur $x' < x''$ möglich, so daß tatsächlich $g(y)$ wächst.

Um schließlich noch die Stetigkeit von $x = g(y)$ zu beweisen, genügt es, den Satz 2 aus Nr. 71 heranzuziehen, dessen Voraussetzungen erfüllt sind. Die Funktion ist monoton, und ihre Werte füllen offenbar das Intervall $\mathcal{X}$ lückenlos aus.[1])

Alle Behauptungen des Satzes sind geometrisch einleuchtend und aus Abb. 32 sofort ablesbar.

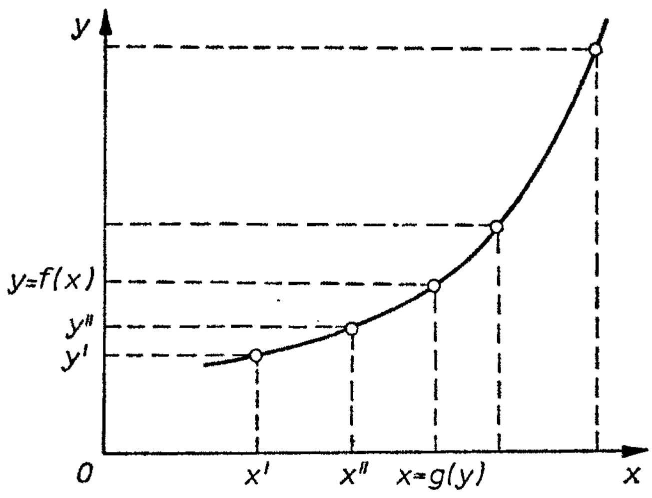

Abb. 32

Mit Hilfe dieses Satzes kann man eine Reihe schon bekannter Resultate von neuem herleiten. Wendet man ihn auf die Funktion x^n (n eine natürliche Zahl) in $\mathcal{X} = [0, \infty)$ an, so findet man Existenz und Stetigkeit der (arithmetischen) Wurzel $x = \sqrt[n]{y}$ für y aus $\mathcal{Y} = [0, \infty)$. Geht man von $y = a^x$ in $\mathcal{X} = (-\infty, \infty)$ aus, so ergibt sich Existenz und Stetigkeit von $x = \log_a y$ in $\mathcal{Y} = (0, \infty)$. Betrachtet man schließlich

[1]) Wie man auch x aus $\mathcal{X}$ wählen mag, man braucht nur $y = f(x)$ zu setzen, damit für dieses y die Funktion $g(y)$ diesen Wert x annimmt.

$y = \sin x$ in $\mathscr{X}_1 = \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ bzw. $y = \tan x$ in $\mathscr{X}_2 = \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, so überzeugt man sich von der Existenz und Stetigkeit von $x = \arcsin y$ in $\mathscr{Y}_1 = [-1, 1]$ bzw. $x = \arctan y$ in $\mathscr{Y}_2 = (-\infty, \infty)$.

Dabei wird vorausgesetzt, daß die Stetigkeit von x^n, a^x, $\sin x$, $\tan x$ schon bewiesen ist, und zwar ohne Berufung auf die Existenz der Umkehrfunktion, da man sonst einem Zirkelschluß verfiele. Solche Beweise gaben wir in Nr. 68; dagegen sind die Überlegungen aus Nr. 72 hier offenbar ungeeignet.

Wir wollen noch ein solches Beispiel betrachten. Für x aus $\mathscr{X} = (-\infty, \infty)$ sei

$$y = x - \varepsilon \sin x, \qquad 0 < \varepsilon < 1. \tag{3}$$

Man zeigt leicht, daß diese Funktion streng monoton wächst. Ist nämlich $x'' > x'$ und sind y', y'' die zugehörigen y-Werte, so ist

$$y'' - y' = (x'' - x') - \varepsilon(\sin x'' - \sin x').$$

Nun ist aber nach (2) aus Nr. 68

$$|\sin x'' - \sin x'| \leqq x'' - x',$$

und hieraus folgt

$$y'' - y' > 0, \quad \text{also} \quad y'' > y'.$$

Wendet man den Satz auf diesen Fall an, so überzeugt man sich davon, daß auch x eine eindeutige Funktion von y ist, usw.

Dieses Beispiel ist deshalb interessant, weil es mit einem Problem der theoretischen Astronomie zusammenhängt. Die Gleichung

$$E = M + \varepsilon \sin E \tag{3a}$$

ist die berühmte *Keplersche Gleichung*[1]), welche die mittlere Anomalie M eines Planeten mit seiner exzentrischen Anomalie E verknüpft (ε ist die Exzentrizität der Planetenbahn). Wir haben somit bewiesen, daß für jeden Wert der mittleren Anomalie die Keplersche Gleichung tatsächlich den Wert der exzentrischen Anomalie eindeutig bestimmt.

84. Der Satz über Beschränktheit einer Funktion. Ist $f(x)$ für alle x eines endlichen Intervalls definiert [nimmt $f(x)$ also endliche Werte an], so folgt daraus keineswegs die Beschränktheit der Funktion, d. h. die Beschränktheit der Menge $\{f(x)\}$ der Funktionswerte. Beispielsweise sei

$$f(x) = \frac{1}{x} \quad \text{für} \quad 0 < x \leqq 1,$$

$$f(0) = 0.$$

Diese Funktion nimmt nur endliche Werte an, ist aber nicht beschränkt; denn bei Annäherung von x an 0 werden die Funktionswerte beliebig groß. Übrigens ist $f(x)$ im halboffenen Intervall $(0, 1]$ stetig, aber in $x = 0$ unstetig.

Anders ist es bei Funktionen, die in einem abgeschlossenen Intervall stetig sind.

Erster Satz von WEIERSTRASS. *Ist $f(x)$ im abgeschlossenen Intervall $[a, b]$ definiert und stetig, so ist $f(x)$ beschränkt, d. h., es gibt konstante endliche Zahlen m und M derart, daß*

$$m \leqq f(x) \leqq M \quad \textit{für} \quad a \leqq x \leqq b$$

ist.

[1]) JOHANNES KEPLER, 1571—1630, deutscher Naturforscher und Astronom.

Den Beweis führen wir indirekt. Es sei also $f(x)$ in $[a, b]$ nicht beschränkt. Dann gäbe es zu jedem natürlichen n in $[a, b]$ ein $x = x_n$ derart, daß

$$|f(x_n)| \geqq n \tag{4}$$

wäre. Nach dem Satz von BOLZANO-WEIERSTRASS (Nr. 41) könnte man aus der Folge $\{x_n\}$ eine Teilfolge $\{x_{n_k}\}$ auswählen, die gegen einen endlichen Grenzwert konvergiert:

$$x_{n_k} \to x_0 \quad (\text{für } k \to \infty),$$

wobei offenbar $a \leqq x_0 \leqq b$ wäre. Da $f(x)$ in x_0 stetig ist, würde dann auch

$$f(x_{n_k}) \to f(x_0)$$

gelten. Das kann aber nicht sein; denn aus (4) folgt

$$|f(x_{n_k})| \to \infty.$$

Mit diesem Widerspruch ist der Satz bewiesen.

85. Größter und kleinster Wert einer Funktion. Wir wissen, daß eine unendliche Menge von Zahlen, selbst wenn sie beschränkt ist, kein größtes (bzw. kleinstes) Element zu enthalten braucht. Ist eine Funktion $f(x)$ in einem Intervall von x-Werten definiert, so braucht die Menge $\{f(x)\}$ ihrer Werte, selbst wenn sie beschränkt ist, kein größtes (bzw. kleinstes) Element zu enthalten. In diesem Fall wird die obere (bzw. untere) Grenze der Werte der Funktion $f(x)$ in diesem Intervall *nicht als Funktionswert angenommen.* Das gilt beispielsweise für

$$f(x) = x - [x]$$

(vgl. Abb. 33). Variiert x in einem beliebigen Intervall $[0, b]$, $b \geqq 1$, so ist die obere Grenze der Funktionswerte gleich 1, aber dieser Wert ist kein Funktionswert, so daß es keinen größten Funktionswert gibt.

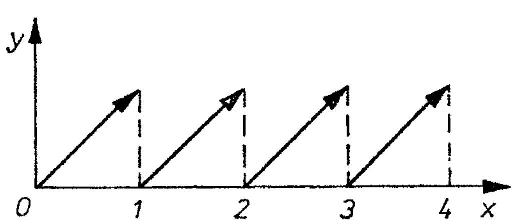

Abb. 33

Der Leser sieht natürlich den Zusammenhang zwischen dieser Tatsache und der Unstetigkeit von $f(x)$ für natürliche Werte von x. Dagegen gilt für stetige Funktionen in einem abgeschlossenen Intervall:

Zweiter Satz von WEIERSTRASS. *Ist eine Funktion $f(x)$ in einem abgeschlossenen Intervall $[a, b]$ definiert und dort stetig, so nimmt sie in diesem Intervall die obere und die untere Grenze ihrer Funktionswerte als Funktionswert an.*

Mit anderen Worten: *Im Intervall $[a, b]$ gibt es Punkte $x = x_0$ und $x = x_1$ derart, daß die Werte $f(x_0)$ bzw. $f(x_1)$ größter bzw. kleinster Wert von $f(x)$ sind.*

Erster Beweis. Wir setzen $M = \sup \{f(x)\}$; nach dem vorigen Satz ist M endlich. Wäre nun stets $f(x) < M$, würde also M nicht als Funktionswert angenommen, so könnte man die Hilfsfunktion

$$\varphi(x) = \frac{1}{M - f(x)}$$

betrachten, deren Nenner also nicht verschwinden würde; $\varphi(x)$ wäre stetig und nach dem vorhergehenden Satz beschränkt: $\varphi(x) \leqq \mu \, (\mu > 0)$. Hieraus ergäbe sich aber

$$f(x) \leqq M - \frac{1}{\mu},$$

so daß $M - \frac{1}{\mu} < M$ ebenfalls eine obere Schranke für alle Werte von $f(x)$ in $[a, b]$ wäre, entgegen der Definition von M als oberer Grenze. Dieser Widerspruch beweist den Satz: In $[a, b]$ gibt es ein x_0 derart, daß $f(x_0) = M$ der größte aller möglichen Werte von $f(x)$ in $[a, b]$ ist. Analog beweist man die Behauptung bezüglich des kleinsten Funktionswertes.

Zweiter Beweis. Man kann auch hier von dem Satz von Bolzano-Weierstrass (Nr. 41) ausgehen. Wir beschränken uns wieder auf den Beweis der Behauptung über den größten Wert. Ist wie eben $M = \sup \{(x)\}$, so läßt sich auf Grund der Definition der oberen Grenze (Nr. 11) zu jedem n ein x_n aus $[a, b]$ finden derart, daß

$$f(x_n) > M - \frac{1}{n} \tag{5}$$

ist. Dann läßt sich aus der Folge $\{x_n\}$ eine Teilfolge $\{x_{n_k}\}$ auswählen, die gegen einen Wert x_0 aus $[a, b]$ konvergiert und für die auf Grund der Stetigkeit von $f(x)$ auch

$$f(x_{n_k}) \to f(x_0)$$

gilt. Nach (5) ist

$$f(x_{n_k}) > M - \frac{1}{n_k},$$

also nach Grenzübergang $f(x_0) \geqq M$. Nun kann aber $f(x_0)$ nicht größer sein als die obere Grenze M der Menge der Funktionswerte. Also ist $f(x_0) = M$, was zu beweisen war.

Beide Beweise sind reine Existenzbeweise. Sie liefern kein Verfahren zur Berechnung etwa des Wertes $x = x_0$. Später, in Kapitel IV, § 1, werden wir, allerdings unter schärferen Voraussetzungen bezüglich der Funktion, Methoden zur Bestimmung der Werte der unabhängigen Veränderlichen kennenlernen, für welche die Funktion einen größten bzw. kleinsten Wert annimmt.

Variiert x in einem Intervall $\mathfrak{X}$ und ist die Funktion $f(x)$ dort definiert und beschränkt, so versteht man unter der *Schwankung* ω von $f(x)$ in diesem Intervall die Differenz

$$M - m.$$

Man kann die Schwankung ω auch als obere Grenze der Menge aller Differenzen $f(x'') - f(x')$ definieren, wenn x' und x'' unabhängig voneinander beliebige Werte aus $\mathfrak{X}$ annehmen:

$$\omega = \sup_{x', x'' \text{ aus } \mathfrak{X}} \{f(x'') - f(x')\}.$$

Handelt es sich um eine in einem abgeschlossenen endlichen Intervall $[a, b]$ stetige Funktion $f(x)$, so folgt aus dem bewiesenen Satz, daß *die Schwankung einfach die Differenz zwischen dem größten und dem kleinsten Funktionswert in diesem Intervall ist.*

In diesem Fall ist das Intervall $\mathfrak{Y}$ der Funktionswerte das abgeschlossene Intervall $[m, M]$ und die Schwankung seine Länge.

86. Die gleichmäßige Stetigkeit. Ist eine Funktion $f(x)$ in einem (abgeschlossenen, halboffenen oder offenen) Intervall $\mathfrak{X}$ definiert und in einem Punkt x_0 aus $\mathfrak{X}$ stetig, so ist

$$\lim_{x \to x_0} f(x) = f(x_0)$$

oder (in der ε-δ-Sprache; vgl. Nr. 66): *Zu jedem $\varepsilon > 0$ gibt es ein $\delta > 0$ derart, daß aus*

$$|x - x_0| < \delta$$

die Beziehung

$$|f(x) - f(x_0)| < \varepsilon$$

folgt.

Wir nehmen nun an, $f(x)$ sei im ganzen Intervall $\mathfrak{X}$, d. h. in jedem Punkt x_0 dieses Intervalls stetig. Dann gibt es für jeden einzelnen Punkt x_0 aus $\mathfrak{X}$ zu gegebenem ε ein δ, das das oben Verlangte leistet. Variiert x_0 in $\mathfrak{X}$, so wird sich die Zahl δ im allgemeinen selbst bei gleichbleibendem ε ändern. Ein einziger Blick auf Abb. 34 genügt, um sich davon zu überzeugen, daß ein δ, welches in einem Abschnitt ausreicht, in dem sich die Funktion langsam ändert (wo also die Kurve allmählich ansteigt), für einen anderen Abschnitt, in dem sich die Funktion schnell ändert (die Kurve steil ansteigt), viel zu groß ist. Mit anderen Worten, *δ hängt im allgemeinen nicht nur von ε, sondern auch von x_0 ab.*

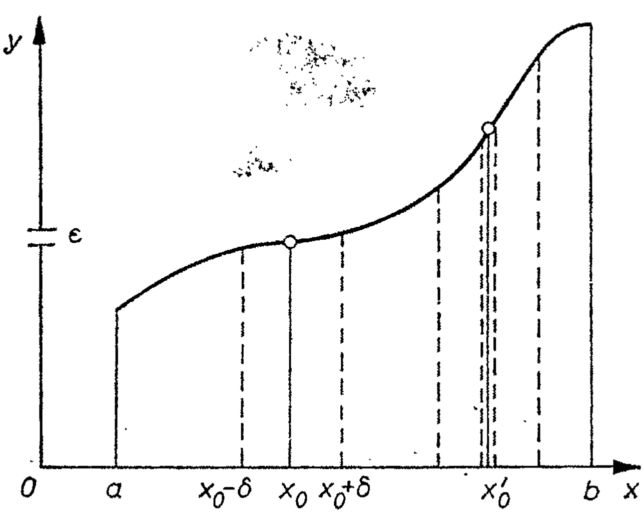

Abb. 34

Wenn es sich um *endlich* viele Werte x_0 (bei ein und demselben ε) handeln würde, so könnte man aus den endlich vielen Zahlen δ, die diesem ε für die einzelnen x_0 entsprechen, die kleinste auswählen, und diese wäre offenbar für alle betrachteten Punkte x_0 gleichzeitig geeignet.

So kann man aber in bezug auf die *unendliche* Menge der Werte x_0 im Intervall $\mathfrak{X}$ nicht schließen. Bei konstantem ε entspricht dieser Menge $\{x_0\}$ eine unendliche Menge von Zahlen δ, unter denen es auch beliebig kleine geben kann. Somit entsteht in bezug auf eine in einem Intervall $\mathfrak{X}$ stetige Funktion $f(x)$ die Frage, ob es zu gegebenem ε ein festes δ gibt, derart, daß für *alle* Punkte x_0 von $\mathfrak{X}$ aus $|x - x_0| < \delta$ die Beziehung $|f(x) - f(x_0)| < \varepsilon$ folgt.

Wenn zu jedem $\varepsilon > 0$ ein $\delta > 0$ gefunden werden kann derart, daß aus

$$|x - x_0| < \delta$$

die Beziehung

$$|f(x) - f(x_0)| < \varepsilon$$

folgt, wie auch immer x und x_0 in $\mathfrak{X}$ gewählt sein mögen, so nennt man $f(x)$ in $\mathfrak{X}$ *gleichmäßig stetig.*

In diesem Fall hängt also δ nur von ε ab und nicht vor der Wahl von x_0; diese Zahl δ ist für alle x_0 gleichzeitig geeignet.

Die gleichmäßige Stetigkeit bedeutet, daß in allen Teilen des Intervalls ein und derselbe Grad der Nachbarschaft zweier Argumentwerte ausreicht, um einen gegebenem Grad der Nachbarschaft der entsprechenden Funktionswerte zu erreichen.

Man kann an einem Beispiel zeigen, daß die Stetigkeit einer Funktion in allen Punkten eines Intervalls nicht mit Notwendigkeit ihre gleichmäßige Stetigkeit in diesem Intervall nach sich zieht. Es sei beispielsweise $f(x) = \sin \frac{1}{x}$ für $0 < x \leqq \frac{2}{\pi}$. Der Variationsbereich von x ist also das halboffene Intervall $\left(0, \frac{2}{\pi}\right]$, und in jedem seiner Punkte ist $f(x)$ stetig. Wir setzen nun $x_0 = \frac{2}{(2n+1)\pi}$, $x = \frac{1}{n\pi}$ (n eine natürliche Zahl). Dann ist

$$f(x_0) = \sin (2n+1)\frac{\pi}{2} = \pm 1, \quad f(x) = \sin n\pi = 0;$$

also $|f(x) - f(x_0)| = 1$, unabhängig davon, daß $|x - x_0| = \frac{1}{n(2n+1)\pi}$ mit wachsendem n beliebig klein gemacht werden kann. Hier gibt es zu $\varepsilon = 1$ kein δ, welches für alle x_0 in $\left(0, \frac{2}{\pi}\right]$ gleichzeitig geeignet wäre, obwohl auf Grund der Stetigkeit für jedes x_0 ein solches $\delta = \delta(\varepsilon, x_0)$ existiert.

Es ist nun äußerst bemerkenswert, daß in einem abgeschlossenen Intervall $[a, b]$ der eben geschilderte Sachverhalt nicht vorliegen kann, wie aus folgendem Satz von G. Cantor[1]) hervorgeht.

87. Der Satz von Cantor. *Ist $f(x)$ im abgeschlossenen Intervall $[a, b]$ definiert und stetig, so ist $f(x)$ dort gleichmäßig stetig.*

Den Beweis führen wir indirekt. Es gebe also zu einem bestimmten $\varepsilon > 0$ kein $\delta > 0$ mit den in der Definition der gleichmäßigen Stetigkeit verlangten Eigenschaften. Dann gäbe es bei jeder Wahl einer Zahl $\delta > 0$ in $[a, b]$ zwei Werte x_0' und x' derart, daß zwar

$$|x' - x_0'| < \delta,$$

aber

$$|f(x') - f(x_0')| \geqq \varepsilon$$

wäre.

Nun wählen wir eine gegen 0 strebende Folge $\{\delta_n\}$ positiver Zahlen δ_n. Dann gäbe es also zu jedem δ_n Werte $x_0^{(n)}$ und $x^{(n)}$ aus $[a, b]$ (die die Rolle von x_0' bzw. x' spielen) derart, daß für $n = 1, 2, 3, \ldots$ zwar

$$|x^{(n)} - x_0^{(n)}| < \delta_n,$$

aber

$$|f(x^{(n)}) - f(x_0^{(n)})| \geqq \varepsilon$$

---

[1]) Georg Cantor. 1845—1918. deutscher Mathematiker.

wäre. Nach dem Satz von BOLZANO-WEIERSTRASS (Nr. 41) könnte man aus der beschränkten Folge $\{x^{(n)}\}$ eine Teilfolge auswählen, die gegen einen bestimmten Punkt x_0 aus $[a, b]$ konvergiert. Um die Bezeichnungen nicht weiter zu komplizieren, wollen wir annehmen, schon die Folge $\{x^{(n)}\}$ konvergiere gegen x_0.

Aus $|x^{(n)} - x_0^{(n)}| < \delta_n$ und $\delta_n \to 0$ folgt $|x^{(n)} - x_0^{(n)}| \to 0$; daher würde auch die Folge $\{x_0^{(n)}\}$ gegen x_0 konvergieren. Aus der Stetigkeit von $f(x)$ in x_0 würde dann

$$f(x^{(n)}) \to f(x_0) \quad \text{und} \quad f(x_0^{(n)}) \to f(x_0),$$

also

$$f(x^{(n)}) - f(x_0^{(n)}) \to 0$$

folgen. Das widerspricht aber der Annahme, daß für alle n

$$|f(x^{(n)}) - f(x_0^{(n)})| \geqq \varepsilon$$

sein sollte. Damit ist der Satz bewiesen.

Daraus ergibt sich unmittelbar eine Folgerung, die uns später von Nutzen sein wird.

Folgerung. *Die Funktion $f(x)$ sei im abgeschlossenen Intervall $[a, b]$ definiert und stetig. Dann gibt es zu vorgegebenem $\varepsilon > 0$ ein $\delta > 0$ mit folgender Eigenschaft: Zerlegt man das Intervall in beliebiger Weise in Teilintervalle, deren Längen kleiner als δ sind, so ist in jedem dieser Teilintervalle die Schwankung von $f(x)$ kleiner als ε.*

Nimmt man nämlich bei gegebenem ε für δ die in der Definition der gleichmäßigen Stetigkeit vorkommende Zahl, so ist in jedem Teilintervall, dessen Länge kleiner als δ ist, die Differenz je zweier Funktionswerte dem absoluten Betrag nach kleiner als ε. Insbesondere gilt das für den größten und den kleinsten Funktionswert, deren Differenz nach der Definition in Nr. 85 gleich der Schwankung der Funktion in diesem Teilintervall ist.

88. Der Heine[1])-Borel-Lebesguesche Überdeckungssatz. Wir beweisen jetzt einen interessanten Hilfssatz, der ebenso wie der Satz von BOLZANO-WEIERSTRASS für viele scharfsinnige Überlegungen nützlich ist; er stammt von E. BOREL[2]).

Wir betrachten neben dem Intervall $[a, b]$ noch ein beliebiges System Σ offener Intervalle σ, das endlich viele oder auch unendlich viele solcher Intervalle σ enthalten kann. Wir sagen, das System Σ sei eine *Überdeckung des Intervalls* $[a, b]$ (oder dieses Intervall werde von Σ überdeckt), wenn zu jedem Punkt x aus $[a, b]$ ein Intervall σ aus Σ gefunden werden kann, das diesen Punkt x enthält. Diese Redeweise erleichtert sowohl Formulierung als auch Beweis unseres Satzes.

Überdeckungssatz. *Wird ein abgeschlossenes Intervall $[a, b]$ von einem unendlichen System $\Sigma = \{\sigma\}$ offener Intervalle überdeckt, so kann man aus Σ ein endliches Teilsystem*

$$\Sigma^* = \{\sigma_1, \sigma_2, \ldots, \sigma_n\}$$

aussondern, das zur Überdeckung von $[a, b]$ ausreicht.

---

[1]) EDUARD HEINE, 1821—1881, deutscher Mathematiker.

[2]) EMILE BOREL, 1871—1956, französischer Mathematiker.

Erster Beweis [indirekt, unter Benutzung der Bolzanoschen Methode (Nr. 41)]. Wir nehmen an, das Intervall $[a, b]$ könne nicht von endlich vielen σ aus Σ überdeckt werden. Dann halbieren wir $[a, b]$. Mindestens eine Hälfte kann dann nicht von endlich vielen σ überdeckt werden (denn würden $\sigma_1, \sigma_2, \ldots, \sigma_m$ aus Σ zur Überdeckung der einen Hälfte und $\sigma_{m+1}, \sigma_{m+2}, \ldots, \sigma_n$ aus Σ zur Überdeckung der anderen ausreichen, so könnte man aus ihnen das endliche System Σ^* bilden, welches ganz $[a, b]$ überdecken würde). Mit $[a_1, b_1]$ bezeichnen wir diese Hälfte (oder eine der beiden, wenn beide Hälften nicht von endlich vielen σ überdeckt werden können). Auf $[a_1, b_1]$ können wir dieselbe Überlegung anwenden und durch Halbierung zu einem Intervall $[a_2, b_2]$ gelangen, das ebenfalls nicht von einem endlichen System von Intervallen σ überdeckt wird, usw.

Führen wir diesen Prozeß unbeschränkt weiter fort, so erhalten wir eine unendliche Intervallschachtelung $\{[a_n, b_n]\}$, $n = 1, 2, 3, \ldots$; jedes dieser Intervalle ist eine Hälfte des vorhergehenden. Sie sind sämtlich so beschaffen, daß keines von ihnen von endlich vielen Intervallen σ aus Σ überdeckt wird. Nach dem Intervallschachtelungssatz der Nr. 38 existiert ein allen Intervallen $[a_n, b_n]$ gemeinsamer Punkt c, gegen den die Intervallendpunkte a_n und b_n streben.

Dieser Punkt c liegt, wie jeder Punkt aus $[a, b]$, in einem der Intervalle σ, sagen wir in $\sigma_0 = (\alpha, \beta)$ mit $\alpha < c < \beta$. Die diskreten Veränderlichen a_n und b_n, die gegen c streben, liegen aber von einem bestimmten n an selbst zwischen α und β (vgl. Nr. 26, Satz 1), so daß das durch sie definierte Intervall $[a_n, b_n]$ ganz von dem einzigen Intervall σ_0 überdeckt wird, entgegen der Definition der Intervalle $[a_n, b_n]$. Mit diesem Widerspruch ist der Satz bewiesen.

Nun führen wir einen zweiten Beweis an, der auf einer von H. Lebesgue[1]) stammenden Idee beruht.

Zweiter Beweis. Wir betrachten die Punkte x^* aus $[a, b]$ mit der Eigenschaft, daß das Intervall $[a, x^*]$ von endlich vielen Intervallen σ überdeckt werden kann. Solche Punkte x^* gibt es: Da beispielsweise a in einem der σ liegt, sind auch alle zu a hinreichend benachbarten Punkte in diesem σ enthalten, gehören also zu der Menge $\{x^*\}$.

Unsere Aufgabe besteht darin, zu beweisen, daß auch b zu der Menge der x^* gehört.

Da für alle x^*

$$x^* \leqq b$$

gilt, existiert nach Nr. 11 auch

$$\sup\{x^*\} = c \leqq b.$$

Wie jeder Punkt aus $[a, b]$ gehört c einem $\sigma_0 = (\alpha, \beta)$ an, $\alpha < c < \beta$. Nach einer Eigenschaft der oberen Grenze existiert ein x_0^* derart, daß $\alpha < x_0^* \leqq c$ ist. Das Intervall $[a, x_0^*]$ läßt sich von endlich vielen Intervallen σ überdecken (nach Definition der x^*). Fügt man das einzige Intervall σ_0 hinzu, so läßt sich auch das ganze Intervall $[a, c]$ mit endlich vielen Intervallen σ überdecken. Also ist c einer der Punkte x^*.

Nun ist aber klar, daß c nicht kleiner als b sein kann, denn sonst gäbe es zwischen c und β einen weiteren Punkt x^*, entgegen der Definition von c als oberer Grenze aller x^*. Also ist $b = c$; somit ist b einer der Punkte x^*, d. h., $[a, b]$ ist durch endlich viele σ überdeckbar, was zu beweisen war.

Bemerkung. Für die Gültigkeit des Satzes ist die Voraussetzung, daß das Grundintervall $[a, b]$ abgeschlossen ist, ebenso wichtig wie die Annahme, daß die über-

[1]) Henri Lebesgue, 1875—1941, französischer Mathematiker.

deckenden Intervalle σ des Systems Σ offen sind. Beispielsweise überdeckt das System der offenen Intervalle

$$\left(\frac{1}{2}, \frac{3}{2}\right), \left(\frac{1}{4}, \frac{3}{4}\right), \left(\frac{1}{8}, \frac{3}{8}\right), \ldots, \left(\frac{1}{2^n}, \frac{3}{2^n}\right), \ldots$$

das Intervall $(0, 1]$, aber man kann daraus kein endliches System aussondern, das diese Eigenschaft hätte. Analog überdeckt das System der abgeschlossenen Intervalle

$$\left[0, \frac{1}{2}\right], \left[\frac{1}{2}, \frac{3}{4}\right], \left[\frac{3}{4}, \frac{7}{8}\right], \ldots, \left[\frac{2^n - 1}{2^n}, \frac{2^{n+1} - 1}{2^{n+1}}\right], \ldots \quad \text{und } [1, 2]$$

das Intervall $[0, 2]$, aber auch hier gibt es kein endliches, dieses Intervall ebenfalls überdeckendes System.

89. Weitere Beweise der grundlegenden Sätze. Wir zeigen jetzt, wie der Überdeckungssatz zum Beweis der Sätze von Bolzano-Cauchy, Weierstrass und Cantor benutzt werden kann.

1. *Erster Zwischenwertsatz* (Nr. 80). Wir beweisen ihn diesmal indirekt. Alle Voraussetzungen des Satzes seien erfüllt, aber in keinem Punkt sei $f(x) = 0$. Dann könnte man nach dem Hilfssatz aus Nr. 80 jeden Punkt x' aus $[a, b]$ mit einer Umgebung $\sigma' = (x' - \delta', x' + \delta')$ versehen, innerhalb derer $f(x)$ das Vorzeichen beibehält.[1])

Das unendliche System $\Sigma = \{\sigma'\}$ dieser Umgebungen überdeckt somit das ganze vorgegebene Intervall $[a, b]$. Dann leistet dasselbe nach dem Überdeckungssatz schon ein endliches System dieser Umgebungen, das wir mit Σ^* bezeichnen.

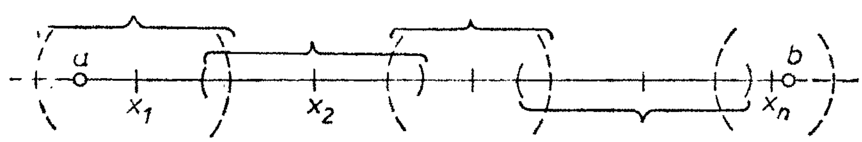

Abb. 35

Der linke Endpunkt a des Intervalls gehört einer der Umgebungen aus diesem System Σ^* an, sagen wir der Umgebung $\sigma_1 = (x_1 - \delta_1, x_1 + \delta_1)$. Ihr rechter Endpunkt $x_1 + \delta_1$ gehört einer Umgebung $\sigma_2 = (x_2 - \delta_2, x_2 + \delta_2)$ aus Σ^* an, $x_2 + \delta_2$ ist seinerseits in einer Umgebung $\sigma_3 = (x_3 - \delta_3, x_3 + \delta_3)$ aus Σ^* enthalten, usw. (vgl. Abb. 35). Nach endlich vielen Schritten nach rechts kommen wir zu einer Umgebung $\sigma_n = (x_n - \delta_n, x_n + \delta_n)$ aus Σ^*, die den rechten Endpunkt b des gegebenen Intervalls enthält. Enthält Σ^* noch weitere Intervalle außer

$$\sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_n, \tag{6}$$

so kann man sie offenbar weglassen.

In σ_1 behält $f(x)$ das Vorzeichen von $f(a)$ bei. Aber in σ_2 behält $f(x)$ ebenfalls das Vorzeichen bei, das natürlich mit dem von $f(a)$ übereinstimmen muß, da sich σ_1 und σ_2 überlappen. Das gilt natürlich für σ_3 und die folgenden Umgebungen ebenso, so daß es auch für σ_n gilt. Daher hätte $f(b)$ dasselbe Vorzeichen wie $f(a)$, und das widerspricht unserer Voraussetzung.

[1]) d. h. im Durchschnitt (gemeinsamen Teil) dieser Umgebung und des Intervalls $[a, b]$, in dem x variieren darf.

2. *Satz von der Beschränktheit einer stetigen Funktion* (Nr. 84). Da $f(x)$ in jedem Punkt x' von $[a, b]$ stetig ist, kann man zu vorgegebenem $\varepsilon > 0$ diesen Punkt in eine so kleine Umgebung $\sigma' = (x' - \delta', x' + \delta')$ einschließen, daß für Punkte x aus dieser Umgebung

$$|f(x) - f(x')| < \varepsilon \quad \text{oder} \quad f(x') - \varepsilon < f(x) < f(x') + \varepsilon$$

gilt. Somit ist $f(x)$ innerhalb dieser Umgebung beschränkt, nach unten durch $f(x') - \varepsilon$, nach oben durch $f(x') + \varepsilon$.

Auch hier kann man natürlich auf das unendliche System Σ der Umgebungen mit dieser Eigenschaft den Überdeckungssatz anwenden. Man kann also aus Σ endlich viele Umgebungen (6) auswählen, die ebenfalls ganz $[a, b]$ überdecken. Ist

$$\begin{aligned} m_1 &\leqq f(x) \leqq M_1 \quad \text{in} \quad \sigma_1, \\ m_2 &\leqq f(x) \leqq M_2 \quad \text{in} \quad \sigma_2, \\ &\ldots\ldots\ldots\ldots \\ m_n &\leqq f(x) \leqq M_n \quad \text{in} \quad \sigma_n, \end{aligned}$$

so nehme man für m das Minimum der Zahlen $m_1, m_2, \ldots, m_n$, für M das Maximum der Zahlen $M_1, M_2, \ldots, M_n$; dann ist offenbar $m \leqq f(x) \leqq M$ in ganz $[a, b]$, was zu beweisen war.

3. *Satz von der gleichmäßigen Stetigkeit* (Nr. 87). Es sei eine Zahl $\varepsilon > 0$ gegeben. Jeden Punkt x' aus $[a, b]$ schließen wir in eine Umgebung $\sigma' = (x' - \delta', x' + \delta')$ ein, innerhalb derer $|f(x) - f(x')| < \frac{\varepsilon}{2}$ gilt. Ist x_0 ebenfalls ein Punkt dieser Umgebung, so ist auch $|f(x') - f(x_0)| < \frac{\varepsilon}{2}$. Für je zwei Punkte x und x_0 aus σ' ist also

$$|f(x) - f(x_0)| < \varepsilon.$$

Nun ziehen wir jede Umgebung σ' auf die Hälfte zusammen, d. h., wir betrachten die Umgebung

$$\bar{\sigma}' = \left(x' - \frac{\delta'}{2}, x' + \frac{\delta'}{2}\right).$$

Aus diesen Umgebungen bilden wir ein System $\bar{\Sigma}$, das das Intervall $[a, b]$ überdeckt, und wenden darauf den Überdeckungssatz an; dann wird $[a, b]$ von endlich vielen Intervallen aus $\bar{\Sigma}$,

$$\bar{\sigma}_i = \left(x_i - \frac{\delta_i}{2}, x_i + \frac{\delta_i}{2}\right) \quad (i = 1, 2, \ldots, n)$$

überdeckt. Nun seien δ das Minimum der Zahlen $\frac{\delta_i}{2}$ und x_0, x zwei beliebige Punkte aus $[a, b]$, für die

$$|x - x_0| < \delta \tag{7}$$

gilt. Dann muß x_0 einer der Umgebungen $\bar{\sigma}_i$ angehören, etwa der Umgebung

$$\bar{\sigma}_{i_0} = \left(x_{i_0} - \frac{\delta_{i_0}}{2}, x_{i_0} + \frac{\delta_{i_0}}{2}\right),$$

so daß $|x_0 - x_{i_0}| < \frac{\delta_{i_0}}{2}$ ist.

Wegen $\delta \leqq \frac{\delta_{i_0}}{2}$ ist nach (7) auch $|x - x_0| < \frac{\delta_{i_0}}{2}$, also $|x - x_{i_0}| < \delta_{i_0}$, d. h., der Punkt x und erst recht der Punkt x_0 liegen in der ursprünglichen Umgebung $(x_{i_0} - \delta_{i_0}, x_{i_0} + \delta_{i_0})$, aus der wir $\bar{\sigma}_{i_0}$ durch Kontraktion erhalten haben. Dann ist aber

$$|f(x) - f(x_0)| < \varepsilon.$$

Da δ nicht von der Lage von x_0 abhing, ist die gleichmäßige Stetigkeit von $f(x)$ in $[a, b]$ bewiesen.

Wie aus diesen Überlegungen ersichtlich ist, läßt sich der Überdeckungssatz mit Erfolg dort anwenden, wo eine „lokale" Eigenschaft, die mit der Umgebung eines einzelnen Punktes verknüpft ist, für ein ganzes Intervall nachgewiesen werden soll.

III. Ableitungen und Differentiale

§ 1. Die Ableitung und ihre Berechnung

90. Berechnung der Geschwindigkeit eines sich bewegenden Punktes. Zuerst betrachten wir als spezielles Beispiel den freien Fall eines schweren Massenpunktes (der Einfachheit halber im Vakuum, um nicht den Luftwiderstand berücksichtigen zu müssen).

Ist seit Beginn des Falles eine Zeit t verstrichen, so ist nach dieser Zeit ein Weg s zurückgelegt, der sich durch die bekannte Formel

$$s = \frac{g}{2} t^2 \tag{1}$$

ausdrücken läßt, wobei g die Erdbeschleunigung ist. Hiervon ausgehend soll die Geschwindigkeit v des Punktes in einem gegebenen Zeitpunkt t bestimmt werden, zu dem sich der Punkt am Ort M befindet (Abb. 36).

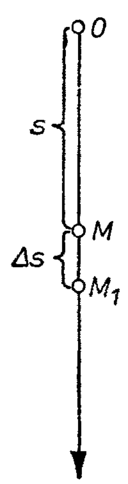

Abb. 36

Wir erteilen der Veränderlichen t den *Zuwachs* (das *Inkrement*) Δt und betrachten den Zeitpunkt $t + \Delta t$, zu dem sich der Punkt in M_1 befinde.[1]) Den Zuwachs MM_1 des Weges im Zeitintervall Δt bezeichnen wir mit Δs.

Wenn wir in (1) die Größe t durch $t + \Delta t$ ersetzen, so erhalten wir für den neuen Wert des Weges den Ausdruck

$$s + \Delta s = \frac{g}{2} \cdot (t + \Delta t)^2,$$

[1]) Wir weisen nochmals darauf hin, daß Δ kein Faktor ist, sondern das Zeichen für Differenz. — *Anm. d. Red.*

und hieraus folgt

$$\Delta s = \frac{g}{2} \cdot (2t \cdot \Delta t + \Delta t^2).$$

Dividieren wir Δs durch Δt, so erhalten wir die *mittlere Fallgeschwindigkeit* des Punktes auf der Strecke MM_1:

$$v_m = \frac{\Delta s}{\Delta t} = gt + \frac{g}{2} \cdot \Delta t.$$

Wie wir sehen, ändert sich diese Geschwindigkeit mit variierendem Δt, wobei der Zustand des fallenden Punktes im Zeitpunkt t um so besser charakterisiert wird, je kleiner das Intervall Δt ist, das nach diesem Zeitpunkt durchlaufen wird.

Unter der *Geschwindigkeit* v des Punktes im Zeitpunkt t versteht man den Grenzwert, gegen den die mittlere Geschwindigkeit v_m im Intervall Δt strebt, wenn Δt selbst gegen 0 strebt.

In unserem Fall ist offenbar

$$v = \lim_{\Delta t \to 0} \left(gt + \frac{g}{2} \cdot \Delta t\right) = gt.$$

Analog kann man die Geschwindigkeit v auch im allgemeinen Fall einer geradlinigen Bewegung eines Punktes berechnen. Die Lage des Punktes wird durch seinen Abstand s von einer Anfangslage O definiert; diesen Abstand nennt man auch den *zurückgelegten* (durchlaufenen) *Weg*. Die Zeit t wird von einem bestimmten Anfangszeitpunkt an gerechnet, wobei nicht unbedingt erforderlich ist, daß sich der Punkt zu diesem Zeitpunkt in O befindet. Die Bewegung gilt als gegeben, wenn die *Bewegungsgleichung* $s = f(t)$ bekannt ist, aus der die Lage des Punktes zu einem beliebigen Zeitpunkt bestimmt werden kann; in unserem Beispiel spielt die Gleichung (1) diese Rolle.

Zur Bestimmung der Geschwindigkeit v in einem gegebenen Zeitpunkt t hat man, wie oben, zu t den Zuwachs Δt hinzuzufügen; das entspricht einer Verlängerung des Weges s und Δs. Der Quotient

$$\frac{\Delta s}{\Delta t},$$

ein sogenannter *Differenzenquotient*, drückt die *mittlere Geschwindigkeit* v_m im Intervall Δt aus. Die wirkliche Geschwindigkeit v zum Zeitpunkt t erhält man hieraus durch Grenzübergang:

$$v = \lim_{\Delta t \to 0} v_m = \lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t},$$

sofern dieser Grenzwert existiert.

Wir betrachten jetzt eine weitere Aufgabe, die zu einer ähnlichen Grenzwertbildung führt.

91. Die Tangente an eine Kurve. Es sei eine Kurve (K) und auf ihr ein Punkt M gegeben (Abb. 37); wir definieren jetzt den Begriff einer Tangente an die Kurve im Punkt M.

Bereits in der Schule wurde die Tangente an einen Kreis definiert, und zwar als Gerade, die mit der Kurve nur einen einzigen Punkt gemeinsam hat. Aber diese

Definition hat ganz speziellen Charakter, sie trifft nicht das Wesentliche. Wollten wir sie z. B. auf die Parabel $y = ax^2$ (Abb. 38a) anwenden, so würden im Koordinatenursprung beide Koordinatenachsen unter diese Definition fallen; jedoch ist, wie vermutlich dem Leser unmittelbar klar ist, in Wirklichkeit nur die x-Achse im Punkt O Tangente an die Parabel.

Wir definieren nun die Tangente allgemein. Zu diesem Zweck wählen wir auf der Kurve (K) außer dem Punkt M noch den Punkt M_1 und ziehen die Sekante MM_1 (Abb. 37). Wenn sich M_1 längs der Kurve bewegt, dreht sich die Sekante um M.

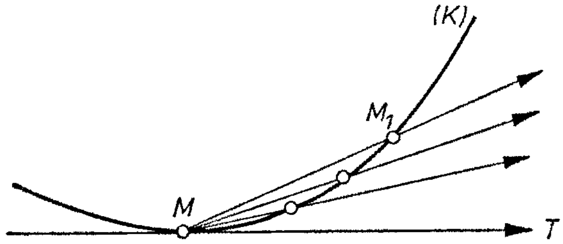

Abb. 37

Unter der *Tangente an die Kurve* (K) *im Punkt* M verstehen wir die Grenzlage MT (wenn eine solche existiert) der Sekante MM_1, wenn sich der Punkt M_1 längs der Kurve gegen M bewegt. (Der Sinn dieser Definition besteht darin, daß der Winkel $\sphericalangle M_1MT$ beliebig klein wird, sobald die Sehne $\overline{MM_1}$ hinreichend klein ist.)

Als Beispiel wenden wir diese Definition auf die Parabel $y = ax^2$ in einem beliebigen Parabelpunkt $M(x, y)$ an. Da die Tangente durch diesen Punkt geht, braucht man, um ihre Lage festlegen zu können, nur noch ihren *Richtungskoeffizienten* (*Steigungskoeffizienten*) zu kennen. Wir stellen uns also jetzt die Aufgabe, den Richtungskoeffizienten $\tan\alpha$ der Tangente im Punkt M zu finden.

Erteilen wir der Abszisse x den Zuwachs Δx, so kommen wir von dem Punkt M der Kurve zum Punkt M_1 mit der Abszisse $x + \Delta x$ und der Ordinate

$$y + \Delta y = a \cdot (x + \Delta x)^2$$

(Abb. 38a). Den Richtungskoeffizienten $\tan\varphi$ der Sekante $\overline{MM_1}$ erhält man aus dem rechtwinkligen Dreieck $\triangle MNM_1$. Darin ist die Kathete MN gleich dem Zuwachs Δx der Abszisse; die Kathete NM_1 ist offenbar gleich dem entsprechenden Zuwachs der Ordinate, $\Delta y = a(2x \cdot \Delta x + (\Delta x)^2)$, so daß

$$\tan\varphi = \frac{\Delta y}{\Delta x} = 2ax + a \cdot \Delta x$$

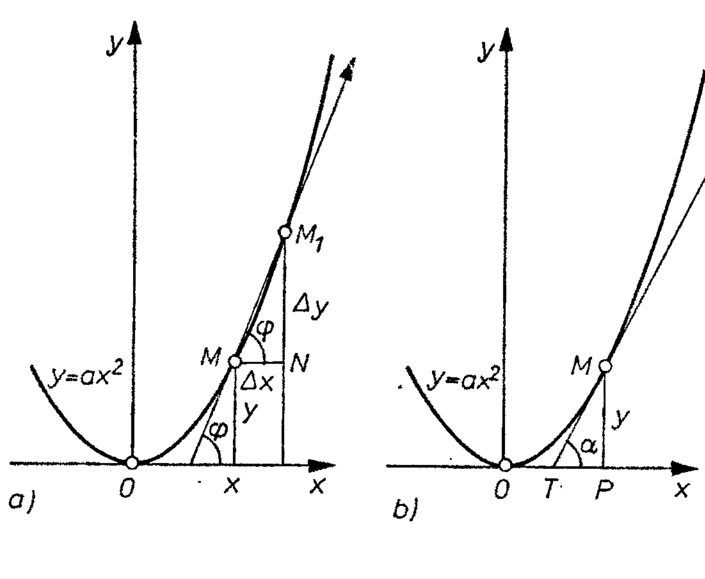

Abb. 38

ist. Um den Richtungskoeffizienten der Tangente zu erhalten, muß man, wie man leicht sieht, hier für $\Delta x \to 0$ zur Grenze übergehen. Somit kommt man zu dem Ergebnis

$$\tan \alpha = \lim_{\Delta x \to 0} (2ax + a \cdot \Delta x) = 2ax\,.$$

Übrigens folgt hieraus ein bequemes Verfahren zur Konstruktion der Tangente an eine Parabel. Aus dem $\triangle\, MPT$ (Abb. 38b) ergibt sich für die Strecke $\overline{TP}$, die sogenannte *Subtangente,*

$$\overline{TP} = \frac{y}{\tan \alpha} = \frac{ax^2}{2ax} = \frac{x}{2},$$

so daß T der Mittelpunkt der Strecke $\overline{OP}$ ist. Um also die Tangente an die Parabel im Punkt M zu erhalten, braucht man die Strecke $\overline{OP}$ nur zu halbieren und den Mittelpunkt mit dem Punkt M zu verbinden.

Im Fall einer beliebigen Kurve mit der Gleichung $y = f(x)$ erhält man den Richtungskoeffizienten der Tangente in ähnlicher Weise. Dem Zuwachs Δx der Abszisse entspricht der Zuwachs Δy der Ordinate, und der Quotient $\frac{\Delta y}{\Delta x}$ liefert den Richtungskoeffizienten der Sekante, $\tan \varphi$. Den Richtungskoeffizienten der Tangente erhält man hieraus durch Grenzübergang für $\Delta x \to 0$:

$$\tan \alpha = \lim_{\Delta x \to 0} \tan \varphi = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x},$$

falls dieser Grenzwert existiert.

92. Definition der Ableitung. Wenn wir die Operationen vergleichen, die wir bei der Lösung der oben betrachteten grundlegenden Aufgaben benutzt haben, so sehen wir leicht, daß in beiden Fällen — von der unterschiedlichen Deutung der Veränderlichen abgesehen — im Grunde dasselbe gemacht wurde: Der Zuwachs der Funktion wurde durch den Zuwachs der unabhängigen Veränderlichen dividiert; alsdann wurde der Grenzwert dieses Differenzenquotienten berechnet. Auf diese Weise kommen wir zu dem Grundbegriff der Differentialrechnung, dem Begriff der Ableitung.

Die Funktion $y = f(x)$ sei in dem Intervall $\mathcal{X}$ definiert. Ausgehend von einem Wert $x = x_0$ der unabhängigen Veränderlichen erteilen wir diesem den Zuwachs $\Delta x \gtrless 0$, der jedoch nicht aus dem Intervall $\mathcal{X}$ herausführen darf, so daß auch der neue Wert $x_0 + \Delta x$ in diesem Intervall liegt. Dann wird der Funktionswert $y = f(x_0)$ durch den neuen Wert $y + \Delta y = f(x_0 + \Delta x)$ ersetzt, d. h., man erhält den Zuwachs

$$\Delta y = \Delta f(x_0) = f(x_0 + \Delta x) - f(x_0)\,.$$

Den Grenzwert des Quotienten des Zuwachses Δy der Funktion y und des entsprechenden Zuwachses Δx der unabhängigen Veränderlichen x für gegen 0 strebendes x, also

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x},$$

nennt man die *Ableitung der Funktion $y = f(x)$ nach der unabhängigen Veränderlichen* x bei gegebenem Wert (oder in dem gegebenen Punkt) $x = x_0$.

Somit ist die Ableitung, wenn sie existiert, für einen gegebenen Wert $x = x_0$ eine bestimmte Zahl[1]); wenn die Ableitung im ganzen Intervall $\mathcal{X}$ existiert, d. h. für jeden Wert x des Intervalls, dann ist sie eine Funktion von x.

In Nr. 96 werden wir sehen, daß aus der Existenz der Ableitung die Stetigkeit der Funktion folgt.

Unter Benutzung des soeben eingeführten Begriffs kann man das in Nr. 90 über die Geschwindigkeit eines sich bewegenden Punktes Gesagte folgendermaßen formulieren: *Die Geschwindigkeit v ist die Ableitung des zurückgelegten Weges s nach der Zeit t.*

Faßt man das Wort „Geschwindigkeit" in allgemeinerem Sinne auf, so könnte man die Ableitung immer als eine Art „Geschwindigkeit" behandeln. Ist nämlich y eine Funktion der unabhängigen Veränderlichen x, so kann man die Frage stellen, wie schnell sich die Veränderliche y im Vergleich zur Veränderlichen x ändert (bei gegebenem Wert von x). Zieht der Zuwachs Δx, zu x addiert, den Zuwachs Δy von y nach sich, so erhält man in Analogie zu Nr. 90 für die mittlere Geschwindigkeit der Änderung von y bezüglich x, wenn sich x zum Δx ändert, die Beziehung

$$V_m = \frac{\Delta y}{\Delta x}.$$

Unter der Geschwindigkeit der Änderung von y für ein gegebenes x versteht man natürlich den Grenzwert dieses Quotienten für gegen 0 strebendes Δx:

$$V = \lim_{\Delta x \to 0} V_m = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x},$$

d. h. wieder die Ableitung von y nach x.

In Nr. 91 haben wir eine durch die Gleichung $y = f(x)$ gegebene Kurve betrachtet und das Problem gelöst, in einem gegebenen Punkt an die Kurve eine Tangente zu ziehen. Jetzt können wir das damalige Resultat folgendermaßen formulieren:

Der Richtungskoeffizient $\tan \alpha$ *der Tangente ist die Ableitung der Ordinate y nach der Abszisse x.*

Diese *geometrische Deutung* der Ableitung ist oft sehr nützlich.

Wir führen als Ergänzung noch einige weitere Beispiele an, die die Bedeutung des Begriffs der Ableitung zeigen.

Wenn die Geschwindigkeit v einer Bewegung nicht konstant ist und sich mit der Zeit t ändert, $v = f(t)$, so führt man die Beschleunigung als Geschwindigkeit der Geschwindigkeitsänderung ein. Entspricht nämlich dem Zuwachs Δt der Zeit t der Geschwindigkeitszuwachs Δv, so gibt der Quotient

$$a_m = \frac{\Delta v}{\Delta t}$$

die *mittlere Beschleunigung* im Zeitintervall Δt an, und der Grenzwert a, sofern er existiert, die *Beschleunigung* der Bewegung in einem gegebenen Zeitmoment:

$$a = \lim_{\Delta t \to 0} a_m = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t}.$$

Somit ist die Beschleunigung die Ableitung der Geschwindigkeit nach der Zeit.[2])

---

[1]) Wir beschränken uns zunächst auf den Fall, daß der oben erwähnte Grenzwert *endlich* ist (vgl. Nr. 101).

[2]) Der Buchstabe a ist die Abkürzung für das französische Wort accélération = Beschleunigung.

Wir wenden uns jetzt der Wärmelehre zu und führen mit Hilfe der Ableitung den Begriff der Wärmekapazität eines Körpers bei gegebener Temperatur ein.

Wir bezeichnen die hierbei auftretenden physikalischen Größen folgendermaßen: θ sei die Temperatur (in Grad Celsius), W die Wärmekapazität, die dem Körper zugeführt werden muß, um ihn von 0 °C auf θ °C zu erwärmen. Offenbar ist W eine Funktion von θ, also $W = f(\theta)$. Erteilen wir θ den Zuwachs $\Delta\theta$, so erhält W den Zuwachs ΔW. Die mittlere Wärmekapazität ist, wenn man von θ °C auf $(\theta + \Delta\theta)$ °C erwärmt,

$$c_m = \frac{\Delta W}{\Delta\theta}.$$

Da sich aber im allgemeinen bei einer Änderung von $\Delta\theta$ diese mittlere Wärmekapazität ändert, können wir sie nicht als Wärmekapazität bei gegebener Temperatur θ nehmen. Um diese zu erhalten, muß man zum Grenzwert übergehen:

$$c = \lim_{\Delta\theta\to 0} c_m = \lim_{\Delta\theta\to 0} \frac{\Delta W}{\Delta\theta}.$$

Somit kann man sagen, *die Wärmekapazität eines Körpers ist die Ableitung der Wärmemenge nach der Temperatur.*

Schließlich betrachten wir noch ein Beispiel aus der Elektrizitätslehre: Wir erarbeiten uns den Begriff der Stromstärke eines Wechselstroms in einem gegebenen Moment.

Wir bezeichnen mit t die Zeit, von einem Anfangsmoment an gerechnet, und mit Q die Elektrizitätsmenge, die in dieser Zeit durch den Querschnitt des Leiters fließt. Offenbar ist Q eine Funktion von t, also $Q = f(t)$. Wenn wir die vorigen Überlegungen wiederholen, finden wir: Die mittlere Stromstärke in dem Zeitintervall Δt ist

$$I_m = \frac{\Delta Q}{\Delta t},$$

die *Stromstärke* in einem gegebenen Moment ist dann der Grenzwert

$$I = \lim_{\Delta t\to 0} I_m = \lim_{\Delta t\to 0} \frac{\Delta Q}{\Delta t};$$

d. h., *die Stromstärke ist die Ableitung der strömenden (fließenden) Elektrizitätsmenge nach der Zeit.*

Alle diese Anwendungen des Begriffs der Ableitung (deren Anzahl sich leicht vergrößern ließe) lassen mit hinreichender Deutlichkeit erkennen, daß *dieser Begriff wesentlich mit den Grundbegriffen ganz verschiedener Wissensgebiete zusammenhängt.*

Das Berechnen von Ableitungen sowie die Untersuchung und Benutzung ihrer Eigenschaften bilden den Hauptgegenstand der *Differentialrechnung.*

Zur Bezeichnung der Ableitung gibt es verschiedene Symbole:

$\frac{\mathrm{d}y}{\mathrm{d}x}$ oder $\frac{\mathrm{d}f(x_0)}{\mathrm{d}x}$ (nach LEIBNIZ)[1]),

y' oder $f'(x_0)$ (nach LAGRANGE),

$\mathrm{D}y$ oder $\mathrm{D}f(x_0)$ (nach CAUCHY).

[1]) Wir betrachten hier die Leibnizschen Bezeichnungen als Ganzes, als einheitliche Symbole; später (Nr. 104) werden wir sehen, daß man sie auch als Brüche auffassen kann. (GOTTFRIED WILHELM LEIBNIZ, 1646—1716, deutscher Mathematiker.)

Wir werden vorzugsweise die einfachen Lagrangeschen Bezeichnungen benutzen. Verwendet man die funktionelle Bezeichnung (die zweite Spalte), so bedeutet der Buchstabe x_0 in den Klammern den Wert der unabhängigen Veränderlichen, für den die Ableitung gebildet wird.[1] Schließlich bemerken wir noch, daß in den Fällen, in denen nicht ohne weiteres klar ist, nach welcher Veränderlichen die Ableitung zu bilden ist (bezüglich welcher „die Geschwindigkeit der Änderung der Funktion" verglichen werden soll), diese Veränderliche als unterer Index angegeben wird. Dann läßt man den oberen Strich aber meistens fort:

$$y_x, f_x(x_0), \mathrm{D}_x y, \mathrm{D}_x f(x_0).$$

Dabei hat natürlich der Index x nichts zu tun mit dem speziellen Wert x_0 der unabhängigen Veränderlichen, für den die Ableitung zu berechnen ist.

In gewissem Sinn kann man sagen, daß die Symbole

$$\frac{\mathrm{d}f}{\mathrm{d}x}, \quad f' \text{ oder } f_x, \quad \mathrm{D}f \text{ oder } \mathrm{D}_x f$$

als Ganzes Bezeichnungen für die funktionale Abhängigkeit der Ableitung von der unabhängigen Veränderlichen sind, also gewissermaßen „Funktionszeichen", ähnlich wie etwa f oder φ die funktionale Abhängigkeit der abhängigen Variablen von der unabhängigen zum Ausdruck bringen.

Wir schreiben jetzt unter Benutzung der soeben für die Ableitungen eingeführten Symbole einige der oben erhaltenen Resultate noch einmal an. Für die Geschwindigkeit einer Bewegung hatten wir

$$v = \frac{\mathrm{d}s}{\mathrm{d}t} \quad \text{oder} \quad v = s_t,$$

für die Beschleunigung

$$a = \frac{\mathrm{d}v}{\mathrm{d}t} \quad \text{oder} \quad a = v_t$$

erhalten. Analog kann man für den Richtungskoeffizienten der Tangente an die Kurve $y = f(x)$

$$\tan\alpha = \frac{\mathrm{d}y}{\mathrm{d}x}$$

oder

$$\tan\alpha = y_x$$

schreiben, usw.

93. Beispiele für die Berechnung der Ableitung. Als Beispiele berechnen wir die Ableitung einiger elementarer Funktionen.

1. Zuerst geben wir folgendes offensichtliche Resultat an: Ist $y = c = \text{const}$, so ist $\Delta y = 0$, wie auch immer Δx beschaffen ist, so daß $y' = 0$ ist; ist $y = x$, so ist $\Delta y = \Delta x$, also $y' = 1$.

[1] Die exakte, jedoch etwas umständlichere Schreibweise für $\frac{\mathrm{d}f(x_0)}{\mathrm{d}x}$ ist $\left.\frac{\mathrm{d}f(x)}{\mathrm{d}x}\right|_{x=x_0}$ oder $\left.\frac{\mathrm{d}f(x)}{\mathrm{d}x}\right|_{x_0}$. — *Anm. d. Red.*

2. Es sei jetzt $y = x^n$ (n eine natürliche Zahl). Wir erteilen x den Zuwachs Δx;[1]) dann ergibt sich für den neuen Wert y:

$$\begin{aligned} y + \Delta y &= (x + \Delta x)^n \\ &= x^n + nx^{n-1} \cdot \Delta x + \frac{n(n-1)}{1 \cdot 2} x^{n-2} \cdot (\Delta x)^2 + \cdots + (\Delta x)^n, \end{aligned}$$

also

$$\Delta y = nx^{n-1} \cdot \Delta x + \frac{n(n-1)}{1 \cdot 2} x^{n-2} \cdot (\Delta x)^2 + \cdots$$

und

$$\frac{\Delta y}{\Delta x} = nx^{n-1} + \frac{n(n-1)}{1 \cdot 2} x^{n-2} \cdot \Delta x + \cdots + (\Delta x)^{n-1}.$$

Weil für $\Delta x \to 0$ alle Summanden, außer dem ersten, gegen 0 streben, ist

$$y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = nx^{n-1}.$$

3. Ist $y = \frac{1}{x}$, so ist $y + \Delta y = \frac{1}{x + \Delta x}$, also

$$\Delta y = \frac{1}{x + \Delta x} - \frac{1}{x} = \frac{-\Delta x}{x(x + \Delta x)} \quad \text{und} \quad \frac{\Delta y}{\Delta x} = -\frac{1}{x(x + \Delta x)}.$$

Hieraus folgt

$$y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} + -\frac{1}{x^2}.$$

Dabei muß natürlich $x \neq 0$ und $x + \Delta x \neq 0$ vorausgesetzt werden.

4. Wir betrachten die Funktion $y = \sqrt{x}$ (für $x > 0$). Es gilt mit $\Delta x > 0$

$$y + \Delta y = \sqrt{x + \Delta x},$$

$$\Delta y = \sqrt{x + \Delta x} - \sqrt{x} = \frac{\Delta x}{\sqrt{x + \Delta x} + \sqrt{x}},$$

$$\frac{\Delta y}{\Delta x} = \frac{1}{\sqrt{x + \Delta x} + \sqrt{x}};$$

da die Wurzel stetig ist, erhalten wir hieraus

$$y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \frac{1}{2\sqrt{x}}.$$

Alle diese Resultate sind Spezialfälle des folgenden Ergebnisses.

---

[1]) Wird die Ableitung für einen *beliebigen* Wert des Arguments berechnet, so bezeichnet man diesen gewöhnlich mit demselben Buchstaben wie das Argument (ohne irgendeinen Index).

5. *Die Potenzfunktion* $y = x^\mu$ (wobei μ eine beliebige reelle Zahl ist). Der Wertebereich von x hängt von μ ab; er wurde in Nr. 48, 2, angegeben. Es gilt (für $x > 0$ und $x + \Delta x > 0$)

$$\frac{\Delta y}{\Delta x} = \frac{(x + \Delta x)^\mu - x^\mu}{\Delta x} = x^{\mu-1} \cdot \frac{\left(1 + \frac{\Delta x}{x}\right)^\mu - 1}{\frac{\Delta x}{x}}.$$

Wenn wir den in Nr. 77, Beispiel 5(c), berechneten Grenzwert benutzen, so erhalten wir

$$y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \mu x^{\mu-1}.$$ [1]

Insbesondere gilt:

Ist $y = \frac{1}{x} = x^{-1}$, so ist $y' = (-1) \cdot x^{-2} = -\frac{1}{x^2}$,

ist $y = \sqrt{x} = x^{1/2}$, so ist $y' = \frac{1}{2} x^{-1/2} = \frac{1}{2\sqrt{x}}$.

6. *Die Exponentialfunktion* $y = a^x$ ($a > 0$; $-\infty < x < \infty$). Hier ist

$$\frac{\Delta y}{\Delta x} = \frac{a^{x+\Delta x} - a^x}{\Delta x} = a^x \cdot \frac{a^{\Delta x} - 1}{\Delta x}.$$

Wenn wir den in Nr. 77, Beispiel 5(b), berechneten Grenzwert benutzen, so finden wir

$$y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = a^x \cdot \ln a.$$

Insbesondere gilt:

Ist $y = e^x$, so ist auch $y' = e^x$.

Die Wachstumsgeschwindigkeit der Exponentialfunktion (für $a > 1$) ist also dem Wert der Funktion selbst proportional: Je größer der Funktionswert ist, desto schneller wächst die Funktion für das entsprechende Argument. Dadurch wird das Wachstum der Exponentialfunktion charakterisiert.

7. *Die Logarithmusfunktion* $y = \log_a x$ ($0 < a \neq 1$; $0 < x < \infty$). In diesem Fall ist

$$\frac{\Delta y}{\Delta x} = \frac{\log_a (x + \Delta x) - \log_a x}{\Delta x} = \frac{1}{x} \cdot \frac{\log_a \left(1 + \frac{\Delta x}{x}\right)}{\frac{\Delta x}{x}}.$$

Wenn wir den in Nr. 77, Beispiel 5(a), berechneten Grenzwert benutzen, so folgt

$$y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \frac{\log_a e}{x}.$$

[1] Ist $\mu > 1$, so erhält man für $x = 0$ leicht auf direktem Wege den Wert der Ableitung: $y' = 0$.

Insbesondere erhält man für den natürlichen Logarithmus das einfache Ergebnis:

Für $y = \ln x$ ist $y' = \frac{1}{x}$.

Das liefert ein weiteres (wenn auch im Grunde kein neues) Argument für die Bevorzugung des natürlichen Logarithmus bei theoretischen Untersuchungen.

Die Tatsache, daß die Geschwindigkeit des Anwachsens der Logarithmusfunktion (für $a > 1$) dem Wert des Arguments umgekehrt porportional ist und für unbeschränkt wachsendes Argument von positiven Werten her gegen 0 strebt, stimmt mit dem in Nr. 65 (S. 137) Gesagten überein.

8. *Die trigonometrischen Funktionen.* Es sei $y = \sin x$; dann ist

$$\frac{\Delta y}{\Delta x} = \frac{\sin(x + \Delta x) - \sin x}{\Delta x} = \frac{\sin\frac{\Delta x}{2}}{\frac{\Delta x}{2}} \cdot \cos\left(x + \frac{\Delta x}{2}\right).$$

Wenn wir die Stetigkeit der Funktion $\cos x$ und den in Nr. 54, Formel 8, hergeleiteten Grenzwert $\lim\limits_{\alpha \to 0} \frac{\sin \alpha}{\alpha} = 1$ benutzen, erhalten wir

$$y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \cos x.^{1)}$$

Analog finden wir:

Ist $y = \cos x$, so ist $y' = -\sin x$.

Im Fall $y = \tan x$ gilt

$$\frac{\Delta y}{\Delta x} = \frac{\tan(x + \Delta x) - \tan x}{\Delta x} = \frac{\frac{\sin(x + \Delta x)}{\cos(x + \Delta x)} - \frac{\sin x}{\cos x}}{\Delta x}$$

$$= \frac{\sin(x + \Delta x) \cdot \cos x - \cos(x + \Delta x) \cdot \sin x}{\Delta x \cdot \cos x \cdot \cos(x + \Delta x)}$$

$$= \frac{\sin \Delta x}{\Delta x} \cdot \frac{1}{\cos x \cdot \cos(x + \Delta x)}.$$

Hieraus folgt wie oben

$$y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \frac{1}{\cos^2 x} = \sec^2 x.$$

Analog gilt:

Ist $y = \cot x$, so ist $y' = -\frac{1}{\sin^2 x} = -\operatorname{cosec}^2 x$.

---

1) Diese Formel verdankt ihre Einfachheit der Tatsache, daß der Winkel im Bogenmaß gemessen wird. Würden wir x in Grad messen, so wäre der Grenzwert des Quotienten aus dem Sinus und dem Winkel nicht gleich 1, sondern, wie man leicht sieht, gleich $\frac{\pi}{180}$, und es wäre $(\sin x') = \frac{\pi}{180} \cos x$.

94. Die Ableitung der inversen Funktion. Bevor wir die Ableitungen der Umkehrfunktionen der trigonometrischen Funktionen berechnen, beweisen wir folgenden allgemeinen Satz.

Satz. *Die Funktion $f(x)$ genüge den Voraussetzungen des Satzes aus Nr. 83 über die Existenz der inversen Funktion; ferner habe $f(x)$ im Punkt x_0 eine endliche und von 0 verschiedene Ableitung $f'(x_0)$. Dann existiert auch für die inverse Funktion $g(y)$ in dem entsprechenden Punkt $y_0 = f(x_0)$ die Ableitung, und sie ist gleich* $\frac{1}{f'(x_0)}$.

Beweis. Wir geben dem Wert $y = y_0$ einen beliebigen Zuwachs Δy; dann erhält die Funktion $x = g(y)$ den entsprechenden Zuwachs Δx. Da die Funktion $f(x)$ eindeutig ist, folgt aus $\Delta y \neq 0$ auch $\Delta x \neq 0$. Dann ist

$$\frac{\Delta x}{\Delta y} = \frac{1}{\frac{\Delta y}{\Delta x}}.$$

Strebt jetzt Δy irgendwie gegen 0, so strebt infolge der Stetigkeit von $x = g(y)$ auch der Zuwachs Δx gegen 0. Dann strebt aber der Nenner der rechten Seite dieser Gleichung gegen den Grenzwert $f'(x_0) \neq 0$; daher existiert auch der Grenzwert der linken Seite, der gleich dem reziproken Wert $\frac{1}{f'(x_0)}$ ist. Das ist dann die Ableitung $g'(y_0)$.

Somit haben wir die einfache Formel

$$x_y = \frac{1}{y_x}.$$

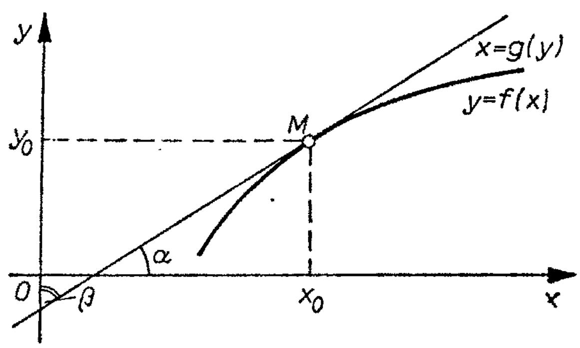

Abb. 39

Ihre geometrische Bedeutung läßt sich leicht erklären. Bekanntlich ist y_x der Tangens des Winkels α, den die Tangente an die Kurve der Funktion $y = f(x)$ mit der x-Achse bildet. Die inverse Funktion $x = g(y)$ wird aber durch dieselbe Kurve dargestellt, jedoch wird die unabhängige Veränderliche auf der y-Achse abgetragen. Daher ist die Ableitung x_y gleich dem Tangens des Winkels β, den dieselbe Tangente mit der y-Achse bildet (Abb. 39). Somit reduziert sich diese Formel auf die bekannte Beziehung

$$\tan \beta = \frac{1}{\tan \alpha},$$

welche die Tangenswerte zweier Winkel α und β miteinander verknüpft, deren Summe gleich $\frac{\pi}{2}$ ist.

Wir betrachten jetzt als Beispiel $y = a^x$. Die inverse Funktion hierfür ist $x = \log_a y$. Wegen $y_x = a^x \cdot \ln a$ (vgl. Nr. 93, 6) ist nach unserer Formel

$$x_y = \frac{1}{y_x} = \frac{1}{a^x \cdot \ln a} = \frac{\log_a \mathrm{e}}{y},$$

in Übereinstimmung mit Nr. 83, 7.

Wir berechnen nun die Ableitungen der Umkehrfunktionen der trigonometrischen Funktionen; der Bequemlichkeit halber vertauschen wir die Veränderlichen x und y und schreiben die oben bewiesene Formel in der Gestalt

$$y_x = \frac{1}{x_y}.$$

9. *Die Umkehrfunktionen der trigonometrischen Funktionen.* Wir betrachten die Funktion $y = \arcsin x$ $(-1 < x < 1)$, wobei $-\frac{\pi}{2} < y < \frac{\pi}{2}$ ist. Sie ist die inverse Funktion von $x = \sin y$, die für diese y-Werte die positive Ableitung $x_y = \cos y$ hat. In diesem Fall existiert auch die Ableitung y_x, und nach unserer Formel ist

$$y_x = \frac{1}{x_y} = \frac{1}{\cos y} = \frac{1}{\sqrt{1 - \sin^2 y}} = \frac{1}{\sqrt{1 - x^2}};$$

die Wurzel ist positiv zu nehmen, weil $\cos y > 0$ ist.

Wir haben die Werte $x = \pm 1$ ausgeschlossen, weil für die entsprechenden Werte $y = \pm\frac{\pi}{2}$ die Ableitung $x_y = \cos y$ gleich 0 ist.

Die Funktion $y = \arctan x$ $(-\infty < x < \infty)$ ist die inverse Funktion von $x = \tan y$. Nach unserer Formel ist

$$y_x = \frac{1}{x_y} = \frac{1}{\sec^2 y} = \frac{1}{1 + \tan^2 y} = \frac{1}{1 + x^2}.$$

Analog erhält man

für $y = \operatorname{arccos} x$ die Ableitung $y' = -\dfrac{1}{\sqrt{1 - x^2}}$ $\quad(-1 < x < 1)$,

für $y = \operatorname{arccot} x$ die Ableitung $y' = -\dfrac{1}{1 + x^2}$ $\quad(-\infty < x < \infty)$.

95. Formeln für Ableitungen. Wir stellen jetzt alle bewiesenen Formeln zusammen:

1.	$y = c$	$y' = 0$
2.	$y = x$	$y' = 1$
3.	$y = x^\mu$	$y' = \mu x^{\mu - 1}$
	$y = \dfrac{1}{x}$	$y' = -\dfrac{1}{x^2}$
	$y = \sqrt{x}$	$y' = \dfrac{1}{2\sqrt{x}}$

4. $y = a^x$ $\qquad y' = a^x \cdot \ln a$

$y = e^x$ $\qquad y' = e^x$

5. $y = \log_a x$ $\qquad y' = \dfrac{\log_a e}{x}$

$y = \ln x$ $\qquad y' = \dfrac{1}{x}$

6. $y = \sin x$ $\qquad y' = \cos x$

7. $y = \cos x$ $\qquad y' = -\sin x$

8. $y = \tan x$ $\qquad y' = \sec^2 x = \dfrac{1}{\cos^2 x}$

9. $y = \cot x$ $\qquad y' = -\operatorname{cosec}^2 x = -\dfrac{1}{\sin^2 x}$

10. $y = \arcsin x$ $\qquad y' = \dfrac{1}{\sqrt{1-x^2}}$

11. $y = \arccos x$ $\qquad y' = -\dfrac{1}{\sqrt{1-x^2}}$

12. $y = \arctan x$ $\qquad y' = \dfrac{1}{1+x^2}$

13. $y = \operatorname{arccot} x$ $\qquad y' = -\dfrac{1}{1+x^2}.$

96. Eine Formel für den Zuwachs der Funktion. Wir beweisen hier zwei einfache Aussagen, die wir im folgenden oft anwenden werden.

Die Funktion $y = f(x)$ sei im Intervall $\mathcal{X}$ definiert. Ausgehend von einem bestimmten Wert $x = x_0$ in diesem Intervall bezeichnen wir mit $\Delta x \gtrless 0$ einen beliebigen Zuwachs von x, für den nur die Einschränkung gemacht wird, daß der Punkt $x_0 + \Delta x$ ebenfalls im Intervall $\mathcal{X}$ liegt. Dann ist

$$\Delta y = \Delta f(x_0) = f(x_0 + \Delta x) - f(x_0)$$

der entsprechende Zuwachs der Funktion.

1. *Hat $y = f(x)$ im Punkt x_0 die (endliche) Ableitung $y_x = f'(x_0)$, so kann man den Zuwachs der Funktion in der Form*

$$\Delta f(x_0) = f'(x_0) \cdot \Delta x + \alpha \cdot \Delta x \tag{2}$$

oder kürzer,

$$\Delta y = y_x \cdot \Delta x + \alpha \cdot \Delta x \tag{2a}$$

darstellen, wobei die Größe α von Δx abhängt und mit Δx gegen 0 strebt.

Da nach Definition der Ableitung für $\Delta x \to 0$ die Beziehung

$$\frac{\Delta y}{\Delta x} \to y_x$$

gilt, folgt, daß

$$\alpha = \frac{\Delta y}{\Delta x} - y_x$$

gegen 0 geht. Wenn wir hieraus Δy bestimmen, so erhalten wir die Formel (2a).

Da die Größe $\alpha \cdot \Delta x$ (für $\Delta x \to 0$) von höherer Ordnung unendlich klein wird als Δx, kann man unter Benutzung der in Nr. 60 eingeführten Bezeichnungen diese Formeln auf die Gestalt

$$\Delta f(x_0) = f'(x_0) \cdot \Delta x + o(\Delta x) \tag{3}$$

bzw.

$$\Delta y = y_x \cdot \Delta x + o(\Delta x) \tag{3a}$$

bringen.

Bemerkung. Bisher haben wir $\Delta x \gtrless 0$ angenommen; die Größe α war für $\Delta x = 0$ nicht definiert. Als wir davon sprachen, daß α für $\Delta x \to 0$ gegen 0 strebt, setzten wir (wie üblich) voraus, daß Δx in beliebiger Weise gegen 0 strebt, *jedoch den Wert* 0 *nicht annimmt.* Wir setzen jetzt $\alpha = 0$ für $\Delta x = 0$; dann gilt natürlich Formel (2) auch für $\Delta x = 0$. Außerdem kann man die Beziehung $\alpha \to 0$ für $\Delta x \to 0$ auch in einem weiteren Sinne als früher auffassen, indem man zuläßt, daß Δx auch den Wert 0 annimmt, wenn Δx gegen 0 strebt.

Aus den bewiesenen Formeln folgt unmittelbar:

2. *Wenn eine Funktion $y = f(x)$ im Punkt x_0 eine (endliche) Ableitung hat, so ist sie in diesem Punkt notwendigerweise stetig.*

In der Tat wird aus (2a) klar, daß die Beziehung $\Delta x \to 0$ auch $\Delta y \to 0$ nach sich zieht.

97. Einfachste Regeln zur Berechnung der Ableitungen. In den vorhergehenden Nummern haben wir die Ableitungen für elementare Funktionen berechnet. Hier und in Nr. 98 geben wir eine Reihe von einfachen Regeln an, mit deren Hilfe man die Ableitungen beliebiger Funktionen berechnen kann, die man aus elementaren Funktionen mit Hilfe endlich vieler arithmetischer Operationen und durch Verkettung (Nr. 51) erhalten kann.

I. *Die Funktion $u = \varphi(x)$ habe (in einem bestimmten Punkt x) die Ableitung u'. Wir beweisen, daß auch die Funktion $y = cu$ ($c =$ const) dort eine Ableitung hat, und berechnen sie.*

Wenn die unabhängige Veränderliche x um Δx wächst, so wächst die Funktion u um Δu, d. h., sie geht aus dem Ausgangswert u in den Wert $u + \Delta u$ über. Der neue Funktionswert y wird $y + \Delta y = c(u + \Delta u)$.

Hieraus erhält man $\Delta y = c \cdot \Delta u$ und

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = c \cdot \lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x} = c \cdot u'.$$

Somit existiert die Ableitung, und es ist

$$y' = (c \cdot u)' = c \cdot u'.$$

Diese Formel bringt folgende Regel zum Ausdruck: *Ein konstanter Faktor kann vor das Differentiationszeichen gezogen werden.*

II. *Die Funktionen* $u = \varphi(x)$, $v = \varphi(x)$ *mögen (in einem bestimmten Punkt) die Ableitungen* u' *und* v' *haben. Wir beweisen, daß dann die Funktion* $y = u \pm v$ *(in demselben Punkt) ebenfalls eine Ableitung hat, und berechnen sie.*

Wir erteilen x den Zuwachs Δx; dann erhalten u, v und y die Zuwächse Δu, Δv bzw. Δy. Die neuen Werte $u + \Delta v$, $v + \Delta v$ und $y + \Delta y$ sind durch die Beziehung

$$y + \Delta y = (u - \Delta u) \pm (v + \Delta v)$$

miteinander verknüpft. Hieraus folgt

$$\Delta y = \Delta u \pm \Delta v, \quad \frac{\Delta y}{\Delta x} = \frac{\Delta u}{\Delta x} \pm \frac{\Delta v}{\Delta x}$$

und

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x} \pm \lim_{\Delta x \to 0} \frac{\Delta v}{\Delta x} = u' \pm v',$$

so daß die Ableitung y' existiert; es gilt

$$y' = (u \pm v)' = u' \pm v'.$$

Dieses Resultat kann man leicht (nach derselben Methode) auf beliebig viele Summanden verallgemeinern.

III. *Unter denselben Voraussetzungen bezüglich der Funktionen* u, v *beweisen wir, daß die Funktion* $y = u \cdot v$ *ebenfalls eine Ableitung hat, und berechnen sie.*

Dem Zuwachs Δx entsprechen, wie auch oben, die Zuwächse Δu, Δv und Δy; dabei ist $y + \Delta y = (u + \Delta u) \cdot (v + \Delta v)$, so daß

$$\Delta y = \Delta u \cdot v + u \cdot \Delta v + \Delta u \cdot \Delta v$$

und

$$\frac{\Delta y}{\Delta x} = \frac{\Delta u}{\Delta x} \cdot v + u \cdot \frac{\Delta v}{\Delta x} + \frac{\Delta u}{\Delta x} \cdot \Delta v$$

ist. Weil für $\Delta x \to 0$ nach Nr. 96, 2, auch $\Delta v \to 0$ gilt, ist

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x} \cdot v + u \cdot \lim_{\Delta x \to 0} \frac{\Delta v}{\Delta x} = u' \cdot v + u \cdot v',$$

d. h., es existiert die Ableitung y', und es ist

$$y' = (u \cdot v)' = u' \cdot v + u \cdot v'.$$

Ist $y = u \cdot v \cdot w$ und existieren u', v', w', so ist

$$y' = [(uv) \cdot w]' = (uv)' \cdot w + (uv) \cdot w' = u'vw + uv'w + uvw'.$$

Man kann sich leicht überlegen, daß für den Fall von n Faktoren die Regel

$$\underbrace{[uvw \cdots s]'}_{n} = \underbrace{u'vw \cdots s + uv'w \cdots s + uvw' \cdots s + \cdots + uvw \cdots s'}_{(n \text{ Summanden})} \tag{4}$$

gilt. Das beweisen wir durch Induktion. Wir nehmen an, die Formel (4) gelte für irgendein n, und beweisen, daß sie auch für $n + 1$ gilt:

$$\underbrace{[uvw \cdots st]'}_{n+1} = [\underbrace{(uvw \cdots s)}_{n} \cdot t]' = (uvw \cdots s)' \cdot t + (uvw \cdots s) \cdot t';$$

wenn die Ableitung $(uvw \cdots s)'$ nach (4) entwickelt wird, erhalten wir die Formel

$$[uvw \cdots st]' = u'vw \cdots st + uv'w \cdots st + \cdots + uvw \cdots s't + uvw \cdots st',$$

die zu (4) analog ist. Weil diese Formel (4) für $n = 2$ (und 3) richtig war, ist somit bewiesen, daß sie für beliebiges natürliches n gilt.

IV. *Wenn schließlich u, v denselben Bedingungen genügen und außerdem v von 0 verschieden ist, so können wir beweisen, daß die Funktion $y = \frac{u}{v}$ eine Ableitung hat, und diese werden wir berechnen.*

In denselben Bezeichnungen wie oben haben wir

$$y + \Delta y = \frac{u + \Delta u}{v + \Delta v},$$

also

$$\Delta y = \frac{\Delta u \cdot v - u \cdot \Delta v}{v(v + \Delta v)} \quad \text{und} \quad \frac{\Delta y}{\Delta x} = \frac{\frac{\Delta u}{\Delta x} \cdot v - u \frac{\Delta v}{\Delta x}}{v \cdot (v + \Delta v)}.$$

Wenn hier Δx gegen 0 strebt (wobei auch gleichzeitig $\Delta v \to 0$ gilt), können wir uns von der Existenz der Ableitung y' und der Gültigkeit der Formel

$$y' = \left(\frac{u}{v}\right)' = \frac{u' \cdot v - u \cdot v'}{v^2}$$

überzeugen.

98. Die Ableitung mittelbarer Funktionen (Kettenregel). Jetzt können wir eine für die Praxis äußerst wichtige Ableitungsregel herleiten, mit deren Hilfe man die Ableitung mittelbarer Funktionen berechnen kann, wenn die Ableitungen der einzelnen Funktionen bekannt sind.

V. *Die Funktion $u = \varphi(x)$ habe in einem Punkt x_0 die Ableitung $u_x = \varphi'(x_0)$; die Funktion $y = f(u)$ habe in dem entsprechenden Punkt $u_0 = \varphi(x_0)$ die Ableitung $y_u = f'(u)$. Dann hat die mittelbare Funktion $y = f(\varphi(x))$ in dem Punkt x_0 ebenfalls eine Ableitung, und sie ist gleich dem Produkt der Ableitungen der Funktionen $f(u)$ und $\varphi(x)$,*

$$[f(\varphi(x))]' = f_u(\varphi(x_0)) \cdot \varphi'(x_0),\text{[1]}$$

oder kürzer

$$y_x = y_u \cdot u_x.$$

Zum Beweis erteilen wir x_0 einen beliebigen Zuwachs Δx; es sei Δu der entsprechende Zuwachs der Funktion $u = \varphi(x)$ und schließlich Δy der Zuwachs der Funktion $y = f(u)$, der von dem Zuwachs Δu hervorgerufen wird. Wir benutzen die Beziehung (2a), die wir, wenn x durch u ersetzt wird, in der Form

$$\Delta y = y_u \cdot \Delta u + \alpha \cdot \Delta u$$

[1] Wir weisen darauf hin, daß das Symbol $f_u(\varphi(x_0))$ die Ableitung der Funktion $f(u)$ nach ihrem Argument u (und nicht nach x) für den Wert $u_0 = \varphi(x_0)$ dieses Arguments bedeutet.

schreiben können (α hängt von Δu ab und strebt mit Δu gegen 0). Wenn wir diese Beziehung jetzt gliedweise durch Δx dividieren, erhalten wir

$$\frac{\Delta y}{\Delta x} = y_u \cdot \frac{\Delta u}{\Delta x} + \alpha \cdot \frac{\Delta u}{\Delta x}.$$

Wenn Δx gegen 0 strebt, so wird auch Δu gegen 0 streben (Nr. 96, Satz 2); dann strebt aber auch, wie wir wissen, die von Δu abhängige Größe α gegen 0. Folglich existiert

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = y_u \cdot \lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x} = y_u \cdot u_x,$$

und das ist gerade die gesuchte Ableitung y_x.

Bemerkung. Hier zeigt sich, wie nützlich die Bemerkung in Nr. 96 bezüglich der Größe α für $\Delta x = 0$ ist: Solange Δx der Zuwachs der unabhängigen Veränderlichen war, konnten wir Δx als von 0 verschieden voraussetzen; wird aber Δx durch den Zuwachs der Funktion $u = \varphi(x)$ ersetzt, so dürfen wir sogar für $\Delta x \neq 0$ nicht mehr $\Delta u \neq 0$ annehmen.

99. Beispiele.[1]) Zunächst bringen wir einige Beispiele für die Anwendung der Regeln I bis IV.

1. Wir betrachten das Polynom

$$y = a_0 x^n + a_1 x^{n-1} + \cdots + a_{n-2} x^2 + a_{n-1} x + a_n.$$

Nach der Regel II und dann nach I erhalten wir

$$\begin{aligned} y' &= (a_0 x^n)' + (a_1 x^{n-1})' + \cdots + (a_{n-2} x^2)' + (a_{n-1} x)' + (a_n)' \\ &= a_0 (x^n)' + a_1 (x^{n-1})' + \cdots + a_{n-2} (x^2)' + a_{n-1} (x)' + (a_n)'. \end{aligned}$$

Wenn wir die Formeln 1, 2, 3 aus Nr. 95 benutzen, erhalten wir schließlich

$$y' = n a_0 x^{n-1} + (n-1)\, a_1 x^{n-2} + \cdots + 2 a_{n-2} x + a_{n-1}.$$

2. Nun sei $y = (2x^2 - 5x + 1) \cdot e^x$. Nach Regel III gilt

$$y' = (2x^2 - 5x + 1)' \cdot e^x + (2x^2 - 5x + 1) \cdot (e^x)'.$$

Analog zum vorigen Beispiel und mit Hilfe der Formel 4 aus Nr. 95 finden wir

$$y' = (4x - 5) \cdot e^x + (2x^2 - 5x + 1) \cdot e^x = (2x^2 - x - 4) \cdot e^x.$$

3. Jetzt betrachten wir $y = \frac{ax + b}{x^2 + 1}$. Nach Regel IV ist

$$\begin{aligned} y' &= \frac{(ax + b)'\,(x^2 + 1) - (ax + b)\,(x^2 + 1)'}{(x^2 + 1)^2} \\ &= \frac{a(x^2 + 1) - (ax + b) \cdot 2x}{(x^2 + 1)^2} = \frac{-ax^2 - 2bx + a}{(x^2 + 1)^2}. \end{aligned}$$

4. Wir berechnen noch einmal die Ableitung der Funktion $y = \tan x$, und zwar ausgehend von $\tan x = \frac{\sin x}{\cos x}$. Wenn wir die Regel IV (und die Formeln 6, 7 aus Nr. 95) benutzen, erhalten wir

$$y' = \frac{(\sin x)' \cdot \cos x - \sin x \cdot (\cos x)'}{\cos^2 x} = \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x}$$

(vgl. Nr. 95, Formel 8).

---

[1]) Mit x, y, u, v werden *Veränderliche* bezeichnet, mit anderen Buchstaben *konstante Größen*.

5. Zur Berechnung der Ableitung von $y = \frac{x \sin x + \cos x}{x \cos x - \sin x}$ benutzen wir zuerst die Regel IV und dann die Regeln II und III (sowie die Formeln 6, 7 aus Nr. 95):

$$y' = \frac{(x \sin x + \cos x)' \,(x \cos x - \sin x) - (x \sin x + \cos x)\,(x \cos x - \sin x)'}{(x \cos x - \sin x)^2}$$

$$= \frac{x \cos x \cdot (x \cos x - \sin x) - (x \sin x + \cos x) \cdot (-x \sin x)}{(x \cos x - \sin x)^2}$$

$$= \frac{x^2}{(x \cos x - \sin x)^2}.$$

Wir berechneten die Ableitungen des Zählers und des Nenners, ohne die einzelnen Schritte anzugeben. Durch Übung muß man dahin kommen, die Ableitungen gleich hinzuschreiben.

Wir geben jetzt einige Beispiele für die Berechnung der Ableitungen mittelbarer Funktionen an.

6. Es sei im Intervall $0 < x < \pi$ nun $y = \ln \sin x$, d. h. $y = \ln u$, wobei $u = \sin x$ positiv ist. Nach Regel V ist $y_x = y_u \cdot u_x$. In der Ableitung $y_u = (\ln u)_u = \frac{1}{u}$ (Formel 5) muß $\sin x$ für u genommen werden. Somit ist

$$y_x = \frac{1}{\sin x} \cdot (\sin x)' = \frac{\cos x}{\sin x} = \cot x \qquad \text{(Formel 6).}$$

7. $y = \sqrt{1 + x^2}$, d. h. $y = \sqrt{u}$ mit $u = 1 + x^2$; nach Regel V ist

$$y_x = \frac{1}{2\sqrt{1 + x^2}} \cdot (1 + x^2)' = \frac{x}{\sqrt{1 + x^2}} \qquad \text{(Formel 3; Beispiel 1).}$$

8. $y = e^{x^2}$, d. h. $y = e^u$ mit $u = x^2$;

$$y_x = e^{x^2} \cdot (x^2)' = 2x \cdot e^{x^2} \qquad \text{(V; 4 und 3).}$$

Natürlich braucht man die mittelbaren Funktionen im einzelnen nicht hinzuschreiben.

9. $y = \sin ax$;

$$y_x = \cos ax \cdot (ax)' = a \cdot \cos ax \qquad \text{(V; 7, 1, 2).}$$

10. $y = (x^2 + x + 1)^n$;

$$y_x = n(x^2 + x + 1)^{n-1} \cdot (x^2 + x + 1)'$$

$$= n(2x + 1)\,(x^2 + x + 1)^{n-1} \qquad \text{(V; 3, Beispiel 1).}$$

11. $y = 2^{\sin x}$;

$$y_x = 2^{\sin x} \cdot \ln 2 \cdot (\sin x)' = \ln 2 \cdot \cos x \cdot 2^{\sin x} \qquad \text{(V; 4, 6).}$$

12. $y = \arctan \frac{1}{x}$;

$$y_x = \frac{1}{1 + \left(\frac{1}{x}\right)^2} \cdot \left(\frac{1}{x}\right)' = \frac{x^2}{1 + x^2} \cdot \left(-\frac{1}{x^2}\right) = -\frac{1}{1 + x^2} \qquad \text{(V; 12, 3).}$$

Die Ableitungen von Funktionen, die durch mehrere Verkettungen entstanden sind, erhält man durch *sukzessive* Anwendung der Regel V:

13. $y = \sqrt{\tan \frac{1}{2} x}$; dann ist

$$y_x = \frac{1}{2\sqrt{\tan \frac{1}{2} x}} \cdot \left(\tan \frac{1}{2} x\right)_x \qquad \text{(V; 3)}$$

$$= \frac{1}{2\sqrt{\tan \frac{1}{2} x}} \cdot \sec^2 \frac{1}{2} x \cdot \left(\frac{1}{2} x\right)_x \qquad \text{(V; 8)}$$

$$= \frac{\sec^2 \frac{1}{2} x}{4\sqrt{\tan \frac{1}{2} x}}.$$

14. $y = e^{\sin^2 \frac{1}{x}}$; dann ist

$$y_x = e^{\sin^2 \frac{1}{x}} \cdot \left(\sin^2 \frac{1}{x}\right)_x \qquad \text{(V; 4)}$$

$$= e^{\sin^2 \frac{1}{x}} \cdot 2 \sin \frac{1}{x} \cdot \left(\sin \frac{1}{x}\right)_x \qquad \text{(V; 3)}$$

$$= e^{\sin^2 \frac{1}{x}} \cdot 2 \sin \frac{1}{x} \cdot \cos \frac{1}{x} \left(\frac{1}{x}\right)_x \qquad \text{(V; 6)}$$

$$= -\frac{1}{x^2} \cdot \sin \frac{2}{x} \cdot e^{\sin^2 \frac{1}{x}}. \qquad \text{(V; 3)}$$

Wir geben noch einige weitere Beispiele für die Anwendung aller Regeln an.

15. $y = \sinh x = \frac{e^x - e^{-x}}{2}$;

$$y' = \frac{1}{2}[(e^x)_x - (e^{-x})_x] = \frac{e^x + e^{-x}}{2} = \cosh x.$$

Ist umgekehrt $y = \cosh x$, so ist $y' = \sinh x$. Schließlich erhält man leicht, wie in Beispiel 4:

Ist $y = \tanh x = \frac{\sinh x}{\cosh x}$, so ist $y' = -\frac{1}{\cosh^2 x}$.

Ist $y = \coth x$, so ist $y' = -\frac{1}{\sinh^2 x}$.

16. $y = \ln\left(x + \sqrt{x^2 + 1}\right)$;

$$y_x = \frac{1}{x + \sqrt{x^2+1}} \cdot \left(x + \sqrt{x^2 + 1}\right)_x$$

$$= \frac{1}{x + \sqrt{x^2+1}} \cdot \left(1 + \frac{x}{\sqrt{x^2+1}}\right) = \frac{1}{\sqrt{x^2+1}}.$$

Dasselbe Resultat kann man auch durch andere Überlegungen erhalten. Wir sahen in Nr. 49, Beispiel 4, daß die Funktion $y = \ln\left(x + \sqrt{x^2+1}\right)$ die Umkehrfunktion von $x = \sinh y$ ist; daher ist (vgl. Nr. 94; das obige Beispiel 15; Nr. 48, 6)

$$y_x = \frac{1}{x_y} = \frac{1}{\cosh y} = \frac{1}{\sqrt{\sinh^2 y + 1}} = \frac{1}{\sqrt{x^2+1}}.$$

17. $y = \dfrac{x}{a^2\sqrt{x^2+a^2}}$;

$$y' = \frac{1}{a^2}\cdot\frac{1\cdot\sqrt{x^2+a^2} - x\cdot\dfrac{x}{\sqrt{x^2+a^2}}}{\left(\sqrt{x^2+a^2}\right)^2} = \frac{1}{(x^2+a^2)^{3/2}}.$$

18. $y = \dfrac{1}{2}\arctan\dfrac{2x}{1-x^2}\quad(-1 < x < 1)$;

$$y' = \frac{1}{2}\cdot\frac{1}{1+\left(\dfrac{2x}{1-x^2}\right)^2}\cdot 2\cdot\frac{1\cdot(1-x^2) - x\cdot(-2x)}{(1-x^2)^2} = \frac{1}{1+x^2}.$$

19. $y = \dfrac{1}{\sqrt{b-ac}}\ln\dfrac{\sqrt{ax+b}-\sqrt{b-ac}}{\sqrt{ax+b}+\sqrt{b-ac}}$ (wir setzen $b - ac > 0$ voraus);

$$y' = \frac{1}{\sqrt{b-ac}}\left[\frac{\dfrac{a}{2\sqrt{ax+b}}}{\sqrt{ax+b}-\sqrt{b-ac}} - \frac{\dfrac{a}{2\sqrt{ax+b}}}{\sqrt{ax+b}+\sqrt{b-ac}}\right] = \frac{1}{(x+c)\sqrt{ax+b}}.$$

20. $y = \dfrac{2}{\sqrt{ac-b}}\arctan\sqrt{\dfrac{ax+b}{ac-b}}$ (hier setzen wir $ac - b > 0$ voraus);

$$y' = \frac{2}{\sqrt{ac-b}}\cdot\frac{1}{1+\dfrac{ax+b}{ac-b}}\cdot\frac{1}{\sqrt{ac-b}}\cdot\frac{a}{2\sqrt{ax+b}} = \frac{1}{(x+c)\sqrt{ax+b}}.$$

21. $y = \dfrac{1}{\sqrt{a^2-b^2}}\arcsin\dfrac{a\sin x + b}{a + b\sin x}\quad\left(a > 0;\ |b| < a;\ -\dfrac{\pi}{2} < x < \dfrac{\pi}{2}\right)$;

$$y' = \frac{1}{\sqrt{a^2-b^2}}\cdot\frac{1}{\sqrt{1-\left(\dfrac{a\sin x+b}{a+b\sin x}\right)^2}}$$

$$\times\frac{a\cos x\cdot(a+b\sin x) - (a\sin x + b)\cdot b\cos x}{(a+b\sin x)^2} = \frac{1}{a+b\sin x}.$$

22. $y = \dfrac{1}{\sqrt{b^2-a^2}}\ln\dfrac{b + a\sin x - \sqrt{b^2-a^2}\cdot\cos x}{a+b\sin x}\quad\left(|a| < |b|;\ \sin x \neq -\dfrac{a}{b}\right)$;

$$y' = \frac{1}{\sqrt{b^2-a^2}}\left[\frac{a\cos x + \sqrt{b^2-a^2}\sin x}{b + a\sin x - \sqrt{b^2-a^2}\cos x} - \frac{b\cos x}{a+b\sin x}\right] = \frac{1}{a+b\sin x}.$$

23. Zur Übung untersuchen wir noch die Ableitung des Exponentialausdrucks $y = u^v$ $(u > 0)$, wobei u und v Funktionen von x sind, die in einem gegebenen Punkt die Ableitung u' bzw. v' haben.

Wir logarithmieren die Beziehung $y = u^v$ und erhalten

$$\ln y = v\cdot\ln u. \tag{5}$$

Somit kann man den Ausdruck für y in der Form

$$y = e^{v \ln u}$$

schreiben; hieraus ist schon klar, daß die Ableitung y' existiert. Sie ist am einfachsten zu berechnen, indem man beide Seiten von (5) nach x ableitet. Dabei benutzen wir die Regeln V und III (und beachten dabei, daß u, v und y Funktionen von x sind). Wir erhalten

$$\frac{1}{y} \cdot y' = v' \cdot \ln u + v \cdot \frac{1}{u} \cdot u';$$

hieraus folgt $y' = y\left(\frac{vu'}{u} + v' \ln u\right)$ oder, wenn wir den Ausdruck für y einsetzen,

$$y' = u^v\left(\frac{vu'}{u} + v' \ln u\right). \tag{6}$$

Diese Formel wurde zuerst von LEINBIZ und JOH. BERNOULLI[1]) hergeleitet.

Ist z. B. $y = x^{\sin x}$, so gilt

$$y_x = x^{\sin x}\left(\frac{\sin x}{x} + \cos x \cdot \ln x\right).$$

24. Unter der Voraussetzung, daß die Funktion $f(x)$ eine Ableitung $f'(x)$ hat, bestimme man die Ableitungen der Funktionen

(a) $\sin f(x)$, (b) $e^{f(x)}$, (c) $\ln f(x)$

nach x und der Funktionen

(d) $f(\sin t)$, (e) $f(e^t)$, (f) $f(\ln t)$

nach t.

Lösungen.

(a) $\cos f(x) \cdot f'(x)$, (b) $e^{f(x)} \cdot f'(x)$, (c) $\frac{f'(x)}{f(x)}$,

(d) $f'(\sin t) \cdot \cos t$, (e) $f'(e^t) \cdot e^t$, (f) $f'(\ln t) \cdot \frac{1}{t}$.

Hinsichtlich der letzten drei Beispiele (d), (e), (f) machen wir den Leser darauf aufmerksam, daß das Symbol $f'(\ldots)$ die Ableitung nach dem Argument x bezeichnet, von dem die Funktion $f(x)$ abhängt; dabei hängen natürlich die Werte dieses Arguments $x = \sin t$, e^t, $\ln t$ noch von t ab. Siehe Fußnote auf S. 187.

25. Eine Funktion $f(x)$, die in einem bezüglich des Nullpunktes symmetrischen Intervall definiert ist, heißt *gerade*, wenn $f(-x) = f(x)$ ist, und *ungerade*, wenn $f(-x) = -f(x)$ ist. Beispiele für gerade Funktionen sind die geraden Potenzen $x^2, x^4, \ldots$ oder die Funktionen $\cos x$, $\cosh x$, für ungerade Funktionen die ungeraden Potenzen $x, x^3, \ldots$ oder $\sin x$, $\sinh x$.

Man beweise, daß *die Ableitung einer geraden Funktion* (wenn sie existiert) *eine ungerade Funktion und die Ableitung einer ungeraden Funktion eine gerade Funktion ist.*

26. Man berechne die Ableitung der Funktion

$$y = \ln |x| \quad \text{für} \quad x \gtrless 0.$$

Für $x > 0$ ist offenbar $y' = \frac{1}{x}$. Diese Formel gilt auch für $x < 0$, denn berechnen wir die Ableitung der Funktion $y = \ln |x| = \ln(-x)$ $(x < 0)$ nach der Kettenregel, so erhalten wir auch in diesem Fall

$$y' = \frac{1}{-x} \cdot (-1) = \frac{1}{x}.$$

---

[1]) JOHANN BERNOULLI, 1667—1748, Schweizer Mathematiker.

27. Wir betrachten die Kurve

$$y = ax^m \qquad (m > 0, x > 0).$$

Der Richtungskoeffizient der Tangente in einem beliebigen Punkt (x, y) ist (Nr. 91, Nr. 92):

$$\tan \alpha = y' = max^{m-1}.$$

Aus Abb. 40 ist ersichtlich, daß für die Strecke $\overline{TP}$ (die sogenannte *Subtangente*) die Beziehung

$$\overline{TP} = \frac{y}{\tan \alpha} = \frac{ax^m}{max^{m-1}} = \frac{x}{m}$$

gilt. Hiernach kann man auch leicht die Tangente konstruieren (das ist eine Verallgemeinerung des Ergebnisses aus Nr. 91).

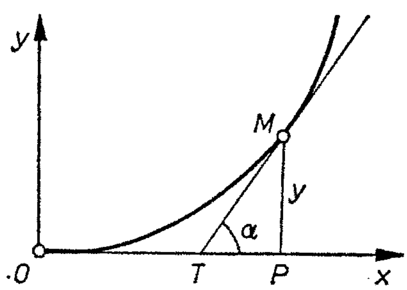

Abb. 40

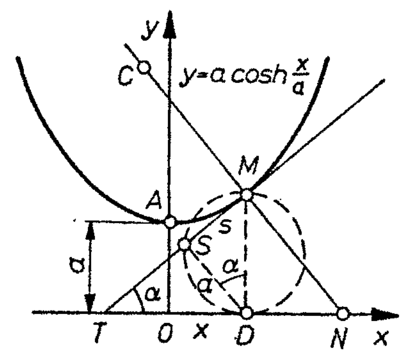

Abb. 41

28. Für die Kurve

$$y = a \cdot \cosh \frac{x}{a} \qquad (a > 0)$$

(*Kettenlinie*) erhalten wir

$$\tan \alpha = y' = \sinh \frac{x}{a}.$$

Wir bestimmen hierfür (für $x > 0$)

$$\cos \alpha = \frac{1}{\sqrt{1 + \tan^2 \alpha}} = \frac{1}{\sqrt{1 + \sinh^2 \frac{x}{a}}} = \frac{1}{\cosh \frac{x}{a}} = \frac{a}{y},$$

so daß $y \cdot \cos \alpha = a$ ist. Wenn wir vom Fußpunkt D der Ordinate $y = \overline{DM}$ (Abb. 41) das Lot $\overline{DS}$ auf die Tangente MT fällen, so ist die Strecke $\overline{DS}$ gleich a. Hieraus ergibt sich wieder eine einfache Methode zur Konstruktion der Tangente an die betrachtete Kurve: Um die Ordinate $\overline{DM}$ als Durchmesser schlagen wir einen Halbkreis, und von D aus schlagen wir einen Kreis mit dem Radius a; wir erhalten den Schnittpunkt S. Die Gerade MS ist dann die Tangente.

29. Ein Massenpunkt schwinge auf einer Achse um eine Mittellage nach dem Gesetz

$$s = A \sin (\omega t + \alpha) \qquad (A > 0, \omega > 0).$$

Man nennt dies eine *harmonische Schwingung*; A ist ihre *Amplitude*, ω die *Frequenz* und α die *Anfangsphase*.

Wenn wir die Ableitung des Weges s nach der Zeit t bilden, erhalten wir die Geschwindigkeit der Bewegung:

$$v = A\omega \cos (\omega t + \alpha).$$

Die größte Geschwindigkeit $\pm A\omega$ wird für $s = 0$ erreicht, d. h., wenn der Punkt durch die Mittellage geht. Hat dagegen der Punkt den größten Abstand von dieser Mittellage ($s = \pm A$), so ist $v = 0$.

Die Ableitung von v nach t,

$$a = -A\omega^2 \sin(\omega t + \alpha),$$

liefert die Beschleunigung des Punktes; offenbar ist

$$a = -\omega^2 s.$$

Ist m die Masse des sich bewegenden Punktes, so ergibt sich nach dem Newtonschen Gesetz für die Kraft F, unter deren Einwirkung die harmonische Schwingung ausgeführt wird,

$$F = -m\omega^2 s.$$

Wie man sieht, ist sie immer zur Mittellage hin gerichtet (weil sie ein zu s entgegengesetztes Vorzeichen hat) und proportional der Entfernung des Punktes von dieser Mittellage.

30. Eine Bewegung, die nach dem Gesetz

$$s = Ae^{-kt}\sin\omega t \qquad (A, k, \omega > 0)$$

verläuft, heißt *gedämpfte Schwingung*, weil das Vorhandensein des Faktors e^{-kt} den Punkt zwingt, im Verlauf der Schwingung um die Mittellage die Amplitude zu verkleinern, bis er dort ganz zur Ruhe kommt:

$$\lim_{t\to\infty} s = 0.$$

In diesem Fall ist

$$v = s_t = Ae^{-kt}(\omega\cos\omega t - k\sin\omega t)$$

und

$$a = v_t = -Ae^{-kt}(\omega^2\sin\omega t + 2\omega k\cos\omega t - k^2\sin\omega t).$$

Wenn wir in der Klammer das Glied $k^2\sin\omega t$ addieren und subtrahieren, erhalten wir nach trivialen Umformungen

$$a = -Ae^{-kt}[(\omega^2 + k^2)\sin\omega t + 2k(\omega\cos\omega t - k\sin\omega t)] = -(\omega^2 + k^2)\,s - 2kv.$$

Für die Kraft, unter deren Einwirkung die Bewegung vor sich geht, ergibt sich

$$F = -(\omega^2 + k^2)\,ms - 2kmv.$$

Wir sehen, daß sie sich aus zwei Kräften zusammensetzt: aus einer Kraft, die proportional dem Abstand des Punktes von der Mittellage ist und in Richtung dieser Mittellage wirkt (wie im Fall der harmonischen Schwingung), und aus einer die Bewegung hemmenden Kraft, die der Geschwindigkeit proportional und ihr engegengesetzt gerichtet ist.

100. Einseitige Ableitungen. Wir wollen uns zum Schluß eine Übersicht über eine Reihe von Sonderfällen verschaffen, die bei Ableitungen vorkommen können. Zunächst führen wir den Begriff der einseitigen Ableitung ein. Wenn der betrachtete x-Wert einer der Endpunkte des Intervalls $\mathcal{X}$ ist, in dem die Funktion $y = f(x)$ definiert ist, so muß man sich bei der Berechnung des Grenzwertes des Quotienten $\frac{\Delta y}{\Delta x}$ auf solche Δx beschränken, die (beim linken Endpunkt) von rechts her bzw. (beim rechten Endpunkt) von links her gegen 0 streben. In diesem Fall spricht man von der *einseitigen* (*rechtsseitigen* bzw. *linksseitigen*) *Ableitung*. In den entsprechenden Punkten hat die Kurve der Funktion nur eine einseitige Tangente (*Halbtangente*).

Es kann vorkommen, daß auch für innere Punkte x nur einseitige Grenzwerte des Quotienten $\frac{\Delta y}{\Delta x}$ (für $\Delta x \to +0$ oder $\Delta x \to -0$) existieren, die nicht übereinstimmen; auch diese nennt man einseitige Ableitungen. In den entsprechenden Punkten der Kurve der Funktion existieren dann nur einseitige Tangenten, die einen (nicht gestreckten) Winkel bilden; der Punkt ist eine *Ecke* (Abb. 42).

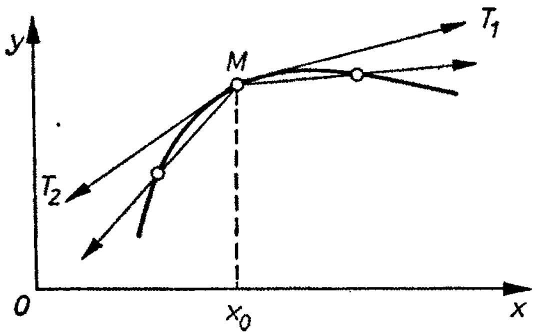

Abb. 42

Als Beispiel betrachten wir die Funktion $y = f(x) = |x|$. Ausgehend von dem Wert $x = 0$ erhalten wir

$$\Delta y = f(0 + \Delta x) - f(0) = f(\Delta x) = |\Delta x|.$$

Ist $\Delta x > 0$, so ist

$$\Delta y = \Delta x, \quad \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = 1.$$

Ist aber $\Delta x < 0$, so ist

$$\Delta y = -\Delta x, \quad \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = -1.$$

In der graphischen Darstellung ist der Koordinatenursprung ein Eckpunkt der Kurve, die aus den Winkelhalbierenden des ersten und zweiten Quadranten besteht.

101. Unendliche Ableitungen. Strebt der Quotient $\frac{\Delta y}{\Delta x}$ für $\Delta x \to 0$ gegen ∞ (bzw. $-\infty$), so wird diese uneigentliche Zahl ebenfalls Ableitung genannt (und wie üblich bezeichnet). Analog führt man den Begriff der einseitigen *unendlichen* Ableitung ein. Die geometrische Deutung der Ableitung als Richtungskoeffizient der Tangente gilt auch in diesem Fall, nur daß hier die Tangente parallel zur y-Achse verläuft (Abb. 43a, b, c, d).

In Abb. 43a und b ist diese Ableitung gleich ∞ bzw. $-\infty$ (beide einseitigen Ableitungen stimmen dem Vorzeichen nach überein), in Abb. 43c und d unterscheiden sich die einseitigen Ableitungen durch das Vorzeichen.

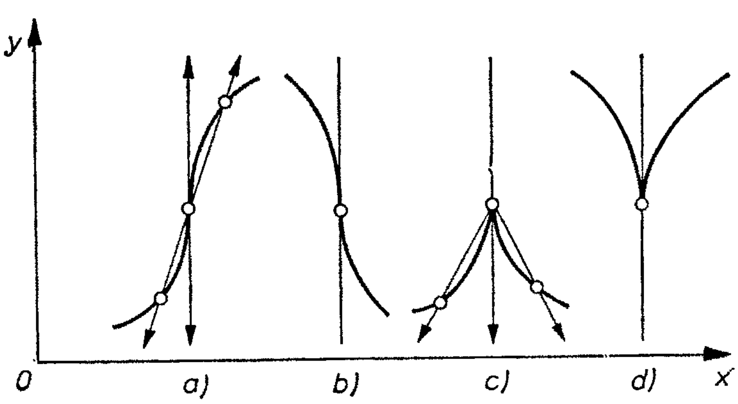

Abb. 43

Es sei beispielsweise

$$f_1(x) = x^{1/3};$$

für $x \neq 0$ ergibt die Formel (3) aus Nr. 95

$$f_1'(x) = \frac{1}{3}x^{-2/3} = \frac{1}{3x^{2/3}},$$

sie gilt jedoch nicht für $x = 0$. In diesem Punkt berechnen wir die Ableitung, indem wir unmittelbar von ihrer Definition ausgehen; wir bilden den Differenzenquotienten

$$\frac{f_1(0 + \Delta x) - f_1(0)}{\Delta x} = \frac{(\Delta x)^{1/3}}{\Delta x} = \frac{1}{(\Delta x)^{2/3}},$$

offenbar strebt er für $\Delta x \to 0$ gegen ∞. Analog überzeugen wir uns davon, daß für die Funktion $f_2(x) = x^{2/3}$ in $x = 0$ die linksseitige Ableitung gleich $-\infty$, die rechtsseitige gleich ∞ wird.

Wenn man den erweiterten Begriff der Ableitung benutzt, kann man den Satz aus Nr. 94 über die Ableitung der Umkehrfunktion durch die Aussage ergänzen, daß auch dann, wenn $f'(x_0)$ gleich 0 bzw. $\pm\infty$ ist, die Ableitung $g'(x_0)$ der inversen Funktion existiert und gleich $\pm\infty$ bzw. 0 ist; so hat z. B. die Funktion $\sin x$ für $x = \pm\frac{\pi}{2}$ die Ableitung $\cos\left(\pm\frac{\pi}{2}\right) = 0$, und für die inverse Funktion $\arcsin y$ existiert für $y = \pm 1$ eine unendliche Ableitung (nämlich ∞).

102. Weitere bemerkenswerte Beispiele.

1. *Beispiele dafür, daß die Ableitung nicht existiert.* Schon die Funktion $y = |x|$ hat im Punkt $x = 0$ (vgl. Nr. 100) keine gewöhnliche zweiseitige Ableitung. Interessanter ist jedoch das Beispiel der Funktion

$$f(0) = 0, \qquad f(x) = x \cdot \sin\frac{1}{x} \quad (\text{für } x \neq 0),$$

die in $x = 0$ stetig ist (Nr. 70, Beispiel 5), aber in diesem Punkt nicht einmal einseitige Ableitungen hat. In der Tat strebt der Quotient

$$\frac{f(0 + \Delta x) - f(0)}{\Delta x} = \frac{f(\Delta x)}{\Delta x} = \sin\frac{1}{\Delta x}$$

für $\Delta x \to \pm 0$ nicht gegen einen Grenzwert.

Aus der graphischen Darstellung dieser Funktion (Abb. 24, S. 121) kann man leicht sehen, daß die vom Ursprung O ausgehende Sekante OM_1 keine Grenzlage einnimmt, wenn M_1 gegen O strebt, so daß also im Ursprung keine Tangente (nicht einmal eine Halbtangente) an die Kurve existiert.

Später (in Band II) werden wir bemerkenswerte Beispiele von Funktionen kennenlernen, die für alle Argumentwerte stetig sind, aber für keinen dieser Werte eine Ableitung besitzen.

2. *Beispiele für Unstetigkeiten der Ableitung.* Wenn für eine gegebene Funktion $y = f(x)$ in jedem Punkt eines Intervalls $\mathscr{X}$ eine endliche Ableitung $y' = f'(x)$ existiert, so ist diese Ableitung ihrerseits in $\mathscr{X}$ eine Funktion von x. In den zahlreichen Beispielen, die wir bisher kennengelernt haben, war diese Funktion selbst stetig. Das braucht jedoch nicht so zu sein. Wir betrachten als Beispiel die Funktion

$$f(0) = 0, \qquad f(x) = x^2 \cdot \sin\frac{1}{x} \qquad (\text{für } x \neq 0).$$

Für $x \neq 0$ kann man die Ableitung in der üblichen Weise berechnen:

$$f'(x) = 2x \cdot \sin\frac{1}{x} - \cos\frac{1}{x};$$

dieses Resultat gilt aber nicht für $x = 0$. In diesem Fall gehen wir direkt auf die Definition der Ableitung zurück und erhalten

$$f'(0) = \lim_{\Delta x \to 0} \frac{f(0 + \Delta x) - f(0)}{\Delta x} = \lim_{\Delta x \to 0} \Delta x \cdot \sin \frac{1}{\Delta x} = 0.$$

Hieraus ersieht man, daß $f'(x)$ für $x \to 0$ nicht gegen einen Grenzwert strebt, so daß die Funktion $f'(x)$ in $x = 0$ unstetig ist.

Das gleiche gilt auch für jede Funktion

$$f(0) = 0, \qquad f(x) = x^{\alpha} \cdot \sin \frac{1}{x} \qquad (\text{für } x \neq 0),$$

sobald $2 > \alpha > 1$ ist.

In diesen Beispielen sind die Unstetigkeiten der Ableitungen von zweiter Art. Das ist kein Zufall: Später (Nr. 113) werden wir sehen, daß Ableitungen keine Unstetigkeiten erster Art, d. h. Sprünge, haben können. Hierzu steht nicht im Widerspruch, daß die Ableitung von $y = |x|$ für negative x gleich -1, für positive x dagegen gleich $+1$ ist (in $x = 0$ existiert y' ja gar nicht!).

§ 2. Das Differential

103. Definition des Differentials. Die Funktion $y = f(x)$ sei in einem Intervall $\mathcal{X}$ definiert und in dem Punkt x_0 stetig. Dann entspricht dem Zuwachs Δx des Arguments der Zuwachs

$$\Delta y = \Delta f(x_0) = f(x_0 + \Delta x) - f(x_0)$$

der Funktion, der mit Δx unendlich klein wird. Von großer Wichtigkeit ist die Frage, *ob zu Δy eine zu Δx proportionale unendlich kleine Größe $A \cdot \Delta x$* ($A =$ const) *existiert derart, daß die Differenz zwischen Δy und $A \cdot \Delta x$ von höherer Ordnung unendlich klein wird als Δx*, d. h. die Frage, ob Δy in der Gestalt

$$\Delta y = A \cdot \Delta x + o(\Delta x) \tag{1}$$

dargestellt werden kann.

Für $A \neq 0$ folgt aus dem Vorliegen von Gleichung (1), daß die unendlich kleine Größe $A \cdot \Delta x$ der unendlich kleinen Größe Δy äquivalent ist, d. h. der Hauptteil von Δy ist, wenn Δx als unendlich kleine Grundgröße gewählt wird (Nr. 62, 63). Ist Gleichung (1) erfüllt, so nennt man die Funktion $y = f(x)$ [für den gegebenen Wert (in dem Punkt) $x = x_0$] *differenzierbar* und den Ausdruck $A \cdot \Delta x$ selbst das *Differential* der Funktion; man bezeichnet das Differential mit $\mathrm{d}y$ oder mit $\mathrm{d}f(x_0)$.[1] (Dabei wird in Klammern auf den Ausgangswert x_0 hingewiesen.)

Wir wiederholen noch einmal, daß das Differential einer Funktion durch zwei Eigenschaften charakterisiert wird: a) Es ist eine lineare (homogene) Funktion des Zuwachses Δx des Arguments; b) es unterscheidet sich vom Zuwachs der Funktion um eine Größe, die für $\Delta x \to 0$ von höherer Ordnung als Δx unendlich klein wird.

Betrachten wir einige Beispiele.

1. Der Flächeninhalt Q eines Kreises mit dem Radius r wird durch die Formel $Q = \pi r^2$ gegeben. Vergrößert man den Radius r um Δr, so ist der entsprechende Zuwachs ΔQ der Größe Q der Flächeninhalt des Kreisringes, der zwischen den konzen-

[1]) Das Symbol $\mathrm{d}f$ als einheitliches Ganzes dient zur Bezeichnung einer funktionalen Abhängigkeit.

trischen Kreisen mit den Radien r und $r + \Delta r$ liegt. Der Relation

$$\Delta Q = \pi(r + \Delta r)^2 - \pi r^2 = 2\pi r \cdot \Delta r + \pi(\Delta r)^2$$

entnimmt man sofort, daß der Hauptteil von ΔQ bei $\Delta r \to 0$ die Größe $2\pi r \cdot \Delta r$ ist; das ist genau das Differential dQ. Geometrisch bedeutet es den Flächeninhalt des Rechtecks (sozusagen des „gestreckten" Ringes), dessen Grundlinie gleich der Länge des Kreisumfangs $2\pi r$ und dessen Höhe gleich Δr ist.

2. Analog erhält man für das Volumen $V = \frac{4}{3}\pi r^3$ einer Kugel mit dem Radius r bei einer Vergrößerung des Radius um Δr den Zuwachs

$$\Delta V = \frac{4}{3}\pi(r + \Delta r)^3 - \frac{4}{3}\pi r^3 = 4\pi r^2 \cdot \Delta r + 4\pi r \cdot (\Delta r)^2 + \frac{4}{3}\pi(\Delta r)^3,$$

dessen Hauptteil für $\Delta r \to 0$ offenbar $dV = 4\pi r^2 \cdot \Delta r$ ist. Dies ist das Volumen eines Zylinders, dessen Grundfläche gleich der Kugeloberfläche, also $4\pi r^2$, und dessen Höhe gleich Δr ist; es entspricht dem Volumen der „plattgewalzten" Schicht zwischen den beiden konzentrischen Kugelflächen mit den Radien r und $r + \Delta r$.

3. Zum Schluß betrachten wir den freien Fall eines Massepunktes gemäß dem Gesetz $s = \frac{gt^2}{2}$. Im Zeitintervall Δt von t bis $t + \Delta t$ durchläuft der Punkt den Weg

$$\Delta s = \frac{g(t + \Delta t)^2}{2} - \frac{gt^2}{2} = gt \cdot \Delta t + \frac{g}{2}(\Delta t)^2.$$

Für $\Delta t \to 0$ ist $ds = gt \cdot \Delta t$ sein Hauptteil. Berücksichtigen wir, daß die Geschwindigkeit im Zeitpunkt t gleich $v = gt$ ist (Nr. 90), so sehen wir, daß sich das Differential des Weges (das annähernd für den Zuwachs des Weges gesetzt werden kann) als der Weg eines Punktes ergibt, welcher sich während des ganzen Zeitintervalls Δt mit eben dieser Geschwindigkeit bewegt.

104. Zusammenhang zwischen der Existenz des Differentials und der Existenz der Ableitung. Nunmehr können wir folgende Aussage leicht beweisen:

Eine Funktion $y = f(x)$ hat genau dann im Punkt x_0 ein Differential, wenn sie in diesem Punkt eine endliche Ableitung $y' = f'(x_0)$ besitzt. Ist diese Bedingung erfüllt, so gilt Gleichung (1) *mit einer Konstanten A, die gleich dieser Ableitung ist:*

$$\Delta y = y_x \Delta x + o(\Delta x). \tag{1a}$$

Beweis. Die Bedingung ist notwendig. Ist nämlich (1) erfüllt, so folgt daraus

$$\frac{\Delta y}{\Delta x} = A + \frac{o(\Delta x)}{\Delta x},$$

so daß wir für $\Delta x \to 0$ tatsächlich

$$A = \lim \frac{\Delta y}{\Delta x} = y_x$$

erhalten.

Daß die Bedingung auch hinreichend ist, folgt sofort aus Nr. 96, Satz 1 [vgl. dort Formel (3a)].

Also gilt für das Differential einer Funktion $y = f(x)$ stets

$$dy = y_x \cdot \Delta x.\text{[1]} \tag{2}$$

Wir weisen hier nochmals darauf hin, daß wir unter Δx in diesem Ausdruck einen beliebigen Zuwachs der unabhängigen Veränderlichen verstehen, d. h. eine beliebige Zahl (die oft als von x unabhängig angesehen werden kann). Dabei braucht Δx keineswegs als unendlich klein angenommen zu werden. Gilt jedoch $\Delta x \to 0$, so wird auch das Differential dy unendlich klein und ist tatsächlich (für $y_x \neq 0$) der Hauptteil des unendlich kleinen Zuwachses Δy der Funktion. Das berechtigt zu der Näherungsgleichung

$$\Delta y \approx dy, \tag{3}$$

die um so genauer gilt, je kleiner Δx ist. Für lineare Funktionen gilt exakt $\Delta y = dy$. Wir werden auf die Näherungsgleichung (3) in Nr. 107 zurückkommen.

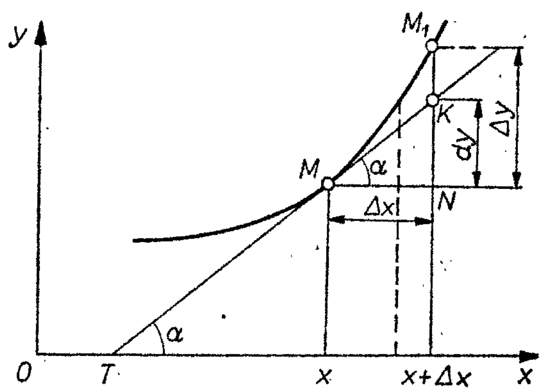

Abb. 44

Um das Differential dy und seinen Zusammenhang mit dem Zuwachs Δy der Funktion $y = f(x)$ geometrisch zu deuten, betrachten wir die graphische Darstellung dieser Funktion (Abb. 44). Durch den Wert x des Arguments und den Wert y der Funktion wird der Punkt M der Kurve bestimmt. Wir ziehen in diesem Kurvenpunkt die Tangente MT. Wie wir bereits in Nr. 92 gesehen haben, ist ihr Richtungskoeffizient $\tan\alpha$ gleich der Ableitung y_x. Erteilt man der Abszisse x den Zuwachs Δx, so erhält die Kurvenordinate y den Zuwachs $\Delta y = \overline{NM_1}$. Gleichzeitig erfährt die Ordinate der Tangente den Zuwachs $\overline{NK}$. Betrachten wir $\overline{NK}$ als Kathete des rechtwinkligen Dreiecks MNK, so finden wir

$$\overline{NK} = \overline{MN} \cdot \tan\alpha = y_x \cdot \Delta x = dy.$$

Während also Δy der Zuwachs der Ordinate der Kurve ist, stellt dy den entsprechenden Zuwachs der Ordinate der Tangente dar.

Zum Schluß wollen wir noch auf die unabhängige Veränderliche x selbst eingehen. Unter ihrem *Differential* versteht man den Zuwachs Δx, d. h., man setzt

$$dx = \Delta x. \tag{4}$$

Identifiziert man das Differential der unabhängigen Veränderlichen x mit dem Differential der Funktion $y = x$ (hierin liegt ebenfalls eine besondere Art der Überein-

[1]) Man prüft leicht nach, daß das Differential in den Beispielen von Nr. 103 tatsächlich so gebildet wurde. Es ist nämlich $Q = \pi r^2$, $Q = 2\pi r$, $dQ = 2\pi r \cdot \Delta r$, usw.

kunft), dann läßt sich Formel (4) sogar beweisen, indem man sich auf (2) stützt:

$$\mathrm{d}x = x_x \cdot \Delta x = 1 \cdot \Delta x = \Delta x.$$

Unter Berücksichtigung der Übereinkunft (4) können wir jetzt Formel (2) umschreiben und das Differential in der Form

$$\mathrm{d}y = y_x \cdot \mathrm{d}x \tag{5}$$

definieren, wie man es auch gewöhnlich tut. Hieraus erhält man

$$y_x = \frac{\mathrm{d}y}{\mathrm{d}x}, \tag{6}$$

so daß dieser Ausdruck, den wir früher als ein einziges Symbol angesehen haben, jetzt auch als Bruch behandelt werden kann.

Die Tatsache, daß hier auf der linken Seite eine wohldefinierte Zahl steht, rechts dagegen der Quotient zweier „unbestimmter" Zahlen $\mathrm{d}y$ und $\mathrm{d}x$ ($\mathrm{d}x = \Delta x$ ist nämlich willkürlich), darf uns nicht verwirren. Die Zahlen $\mathrm{d}x$ und $\mathrm{d}y$ ändern sich proportional, und die Ableitung y_x ist der Proportionalitätsfaktor.

Der Begriff des Differentials und auch die Bezeichnung „Differential"[1]) selbst stammen von Leibniz, der jedoch diesen Begriff nicht genau definiert hat. Neben den Differentialen betrachtete Leibniz auch *Differentialquotienten*, d. h. Quotienten zweier Differentiale, welche unserer Ableitung äquivalent sind. Jedoch war gerade das Differential bei Leibniz der ursprüngliche Begriff. Seit Cauchy, der mit seiner Theorie der Grenzwerte die gesamte Analysis begründete und als erster die Ableitung als Grenzwert definierte, begann man üblicherweise von der Ableitung auszugehen und den Begriff des Differentials auf der Basis der Ableitung aufzubauen.[2])

105. Grundlegende Formeln und Regeln für die Bildung von Differentialen. Da sich das Differential $\mathrm{d}y$ nur um den Faktor $\mathrm{d}x$ von der Ableitung y_x unterscheidet, läßt sich aus der Tabelle der Ableitungen für die elementaren Funktionen (Nr. 95) sofort eine Tabelle der Differentiale aufstellen:

1.	$y = c$	$\mathrm{d}y = 0$
2.	$y = x^\mu$	$\mathrm{d}y = \mu x^{\mu-1}\,\mathrm{d}x$
	$y = \dfrac{1}{x}$	$\mathrm{d}y = -\dfrac{\mathrm{d}x}{x^2}$
	$y = \sqrt{x}$	$\mathrm{d}y = \dfrac{\mathrm{d}x}{2\sqrt{x}}$
3.	$y = a^x$	$\mathrm{d}y = a^x \ln a\,\mathrm{d}x$
	$y = \mathrm{e}^x$	$\mathrm{d}y = e^x\,\mathrm{d}x$

[1]) Von dem lateinischen Wort differentia = Differenz.

[2]) Nachdem sich somit die Existenz der „Ableitung" und die Existenz des „Differentials" als praktisch äquivalent erwiesen haben, wird auch die Verwendung der Wörter „differenzieren" bzw. „Differentiation" sowohl für die Bildung der Ableitung als auch für die Bildung des Differentials verständlich. Im Deutschen (und im Russischen) wird kein Unterschied gemacht, während z. B. das Französische zwischen „dérivation" und „différentiation" unterscheidet.

4. $y = \log_a x \qquad dy = \dfrac{\log_a e\, dx}{x}$

$y = \ln x \qquad dy = \dfrac{dx}{x}$

5. $y = \sin x \qquad dy = \cos x\, dx$

6. $y = \cos x \qquad dy = -\sin x\, dx$

7. $y = \tan x \qquad dy = \sec^2 x\, dx = \dfrac{dx}{\cos^2 x}$

8. $y = \cot x \qquad dy = -\operatorname{cosec}^2 x\, dx = -\dfrac{dx}{\sin^2 x}$

9. $y = \arcsin x \qquad dy = \dfrac{dx}{\sqrt{1 - x^2}}$

10. $y = \arccos x \qquad dy = -\dfrac{dx}{\sqrt{1 - x^2}}$

11. $y = \arctan x \qquad dy = \dfrac{dx}{1 + x^2}$

12. $y = \operatorname{arccot} x \qquad dy = -\dfrac{dx}{1 + x^2}$

Die Regeln zur Bildung von Differentialen lauten:

I. $\quad d(cu) = c \cdot du,$

II. $\quad d(u \pm v) = du \pm dv,$

III. $\quad d(uv) = u \cdot dv + v \cdot du,$

IV. $\quad d\left(\dfrac{u}{v}\right) = \dfrac{v \cdot du - u \cdot dv}{v^2}.$

Alle diese Regeln ergeben sich direkt aus der entsprechenden Vorschrift für die Bildung der Ableitung. Wir beweisen als Beispiel die letzten beiden:

$$\begin{aligned} d(u \cdot v) &= (u \cdot v)'\, dx = (u' \cdot v + u \cdot v')\, dx \\ &= v \cdot (u' \cdot dx) + u \cdot (v' \cdot dx) = v \cdot du + u \cdot dv, \end{aligned}$$

$$\begin{aligned} d\left(\frac{u}{v}\right) &= \left(\frac{u}{v}\right)' \cdot dx = \frac{u'v - v'u}{v^2} \cdot dx \\ &= \frac{v(u' \cdot dx) - u(v' \cdot dx)}{v^2} = \frac{v \cdot du - u \cdot dv}{v^2}. \end{aligned}$$

106. Invarianz des Differentials. Die Regel für die Bildung des Differentials einer mittelbaren Funktion führt uns auf eine bemerkenswerte und wichtige Eigenschaft des Differentials.

Die Funktionen $y = f(x)$ und $x = \varphi(t)$ seien so beschaffen, daß aus ihnen die mittelbare Funktion $y = f(\varphi(t))$ gebildet werden kann. Existieren die Ableitungen

y_x und x_t, so ist nach Regel V aus Nr. 98 auch die Ableitung

$$y_t = y_x \cdot x_t \tag{7}$$

vorhanden.

Das Differential $\mathrm{d}y$ läßt sich, wenn x als unabhängige Variable aufgefaßt wird, durch Formel (5) ausdrücken. Wir gehen jetzt zur unabhängigen Veränderlichen t über. Unter dieser Annahme finden wir einen anderen Ausdruck für das Differential:

$$\mathrm{d}y = y_t \cdot \mathrm{d}t.$$

Ersetzen wir jedoch die Ableitung y_t durch den Ausdruck (7) und beachten, daß $x_t \cdot \mathrm{d}t$ das Differential von x als Funktion von t ist, so erhalten wir schließlich

$$\mathrm{d}y = y_x \cdot x_t \,\mathrm{d}t = y_x \cdot \mathrm{d}x,$$

d. h., wir kommen zur ursprünglichen Form des Differentials zurück.

Wir sehen also, daß *das Differential auch dann invariant bleibt, wenn die ursprüngliche unabhängige Veränderliche durch eine neue ersetzt wird.* Wir dürfen stets das Differential von y in der Form (5) schreiben, gleichgültig, ob x die unabhängige Veränderliche ist oder nicht. Der Unterschied besteht nur darin, daß $\mathrm{d}x$ nicht einen Zuwachs Δx bezeichnet, wenn t als unabhängige Veränderliche gewählt wird, sondern das Differential von x als Funktion von t. Diese Eigenschaft nennt man die *Invarianz des Differentials.*

Da sich aus (5) unmittelbar die Formel (6) ergibt, wenn man die Ableitung y_x durch die Differentiale $\mathrm{d}x$ und $\mathrm{d}y$ ausdrückt, bleibt auch die letzte Formel richtig, gleichgültig, in bezug auf welche unabhängige Veränderliche die betreffenden Differentiale gebildet wurden (natürlich in beiden Fällen in bezug auf dieselbe).

Es sei z. B.

$$y = \sqrt{1 - x^2} \quad (-1 < x < 1),$$

also $y_x = -\dfrac{x}{\sqrt{1 - x^2}}$. Setzen wir jetzt $x = \sin t$ $\left(-\dfrac{\pi}{2} < t < \dfrac{\pi}{2}\right)$, dann ist $y = \sqrt{1 - \sin^2 t} = \cos t$, und wir erhalten $\mathrm{d}x = \cos t \cdot \mathrm{d}t$, $\mathrm{d}y = -\sin t \cdot \mathrm{d}t$. Man überzeugt sich leicht, daß die Formel

$$y_x = \frac{-\sin t \cdot \mathrm{d}t}{\cos t \cdot \mathrm{d}t} = -\frac{\sin t}{\cos t}$$

nur ein anderer Ausdruck für die oben errechnete Ableitung ist. Man benutzt das zweckmäßigerweise insbesondere in solchen Fällen, bei denen die Abhängigkeit der Größen y und x nicht unmittelbar gegeben ist, sondern die beiden Veränderlichen x und y von irgendeiner dritten Hilfsveränderlichen (einem sogenannten *Parameter*) abhängen, z. B.

$$x = \varphi(t), \quad y = \psi(t). \tag{8}$$

Unter der Voraussetzung, daß jede dieser beiden Funktionen eine Ableitung besitzt und daß für die erste von ihnen die Umkehrfunktion $t = \theta(x)$ existiert, die ebenfalls eine Ableitung besitzt (Nr. 83, 94), ist leicht einzusehen, daß dann auch y eine Funktion von x ist, nämlich

$$y = \psi(\theta(x)) = f(x), \tag{9}$$

für welche gleichfalls eine Ableitung existiert. Diese Ableitung kann nach der oben angegebenen Regel berechnet werden,

$$y_x = \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{y_t \cdot \mathrm{d}t}{x_t \cdot \mathrm{d}t} = \frac{y_t}{x_t} = \frac{\psi'(t)}{\varphi'(t)}, \tag{10}$$

ohne daß die direkte Abhängigkeit der Größe y von x benutzt werden muß.

Ist etwa $x = \sin t$, $y = \cos t$ $\left(-\frac{\pi}{2} < t < \frac{\pi}{2}\right)$, so läßt sich die Ableitung y_x wie oben bestimmen, ohne daß die Beziehung $y = \sqrt{1 - x^2}$ herangezogen werden muß.

Sieht man x und y als rechtwinklige Koordinaten eines Punktes in der Ebene an, so ordnen die Gleichungen (8) jedem Wert des Parameters t des Definitionsbereichs von $\varphi(t)$ und $\psi(t)$ einen Punkt zu, welcher bei variierendem t eine Kurve in der Ebene beschreibt. Die Gleichungen (8) nennt man die *Parametergleichungen dieser Kurve* (*Parameterdarstellung der Funktion*).

Für den Fall, daß eine Kurve in Parameterdarstellung gegeben ist, ermöglicht Formel (10), den Richtungskoeffizienten der Tangente direkt aus den Gleichungen (8) aufzustellen, ohne daß man zu der Kurvengleichung (9) übergehen müßte, nämlich

$$\tan \alpha = \frac{y_t}{x_t}. \tag{11}$$

Bemerkung. Daß man die Ableitung durch Differentiale in bezug auf eine beliebige Veränderliche ausdrücken kann, führt insbesondere dazu, daß die Formeln

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{1}{\frac{\mathrm{d}x}{\mathrm{d}y}}, \quad \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}y}{\mathrm{d}u} \cdot \frac{\mathrm{d}u}{\mathrm{d}x},$$

die in der Leibnizschen Symbolik die Differentiationsregeln für die Umkehrfunktion und für eine mittelbare Funktion angeben, zu einfachen algebraischen Identitäten werden (da hier alle Differentiale in bezug auf ein und dieselbe Veränderliche gebildet werden können).

Man darf jedoch nicht glauben, daß damit ein neuer Beweis dieser Formeln gegeben worden sei. Erstens haben wir hier nicht die Existenz der Ableitung auf der linken Seite der Gleichungen bewiesen, und zweitens (und das ist die Hauptsache) haben wir die Invarianz des Differentials benutzt, welche ihrerseits eine Folgerung aus der Regel V ist.

107. Näherungsformeln mit Hilfe von Differentialen. Wir haben in Nr. 104 gesehen, daß für $\Delta x \to 0$ das Differential $\mathrm{d}y$ einer Funktion y (falls $y_x \neq 0$ ist) der Hauptteil des unendlich kleinen Zuwachses Δy der Funktion y ist. Somit ist

$$\Delta y \approx \mathrm{d}y \tag{3}$$

oder, ausführlicher,

$$\Delta f(x_0) = f(x_0 + \Delta x) - f(x_0) \approx f'(x_0) \cdot \Delta x, \tag{3a}$$

mit einem Fehler, der von höherer Ordnung als Δx klein wird. Das bedeutet (Nr. 62), daß der relative Fehler dieser Gleichung für hinreichend kleines Δx beliebig klein wird.

Betrachten wir ein einfaches Beispiel. Es sei $y = x^3$. Dann ist

$$\Delta y = (x_0 + \Delta x)^3 - x_0^3 = 3x_0^2 \cdot \Delta x + 3x_0 \cdot (\Delta x)^2 + (\Delta x)^3,$$

und der lineare Teil von Δy ist (wie wir oben allgemein zeigten) tatsächlich das Differential $\mathrm{d}y = 3x_0^2 \cdot \Delta x = y_x \cdot \Delta x$. Wir setzen nun beispielsweise $x_0 = 2{,}3$. Wählen wir $\Delta x = 0{,}1$, so erhalten wir

$$\Delta y = 2{,}4^3 - 2{,}3^3 = 1{,}657$$

und

$$\mathrm{d}y = 3 \cdot 2{,}3^2 \cdot 0{,}1 = 1{,}587,$$

so daß beim Ersetzen der ersten Zahl durch die zweite ein Fehler von 0,070 entsteht und der relative Fehler größer als 4% ist. Für $\Delta x = 0{,}01$ erhalten wir $\Delta y = 0{,}159391$ und $\mathrm{d}y = 0{,}1587$, was einen relativen Fehler ergibt, der bereits kleiner ist als 0,5%. Für $\Delta x = 0{,}001$ ist der relative Fehler kleiner als 0,05%, usw.

Das läßt sich schon unmittelbar aus Abb. 44 ablesen, die die geometrische Deutung des Differentials zeigt. Aus der graphischen Darstellung ist ersichtlich, daß man bei einer Verkleinerung von Δx tatsächlich mit stets größerer relativer Genauigkeit den Zuwachs der Ordinate der Kurve durch den Zuwachs der Ordinate der Tangente ersetzen kann.

Das Ersetzen des Zuwachses Δy der Funktion durch ihr Differential $\mathrm{d}y$ hat offenbar den Vorteil, daß $\mathrm{d}y$ von Δx linear abhängt, während Δy gewöhnlich eine kompliziertere Funktion von Δx ist.

Setzen wir $\Delta x = x - x_0$ oder $x_0 + \Delta x = x$, so erhält Gleichung (3a) die Form

$$f(x) - f(x_0) \approx f'(x_0) \cdot (x - x_0)$$

oder

$$f(x) \approx f(x_0) + f'(x_0)\,(x - x_0).$$

Durch diese Formel wird die Funktion $f(x)$ für solche Werte von x, die nahe bei x_0 liegen, näherungsweise durch eine lineare Funktion ersetzt. Geometrisch entspricht dies dem Ersetzen des dem Punkt $(x_0, f(x_0))$ benachbarten Teiles der Kurve $y = f(x)$ durch ein Stück der Tangente an die Kurve in diesem Punkt:

$$y = f(x_0) + f'(x_0) \cdot (x - x_0)$$ [1]

(vgl. Abb. 44). Nehmen wir der Einfachheit halber $x_0 = 0$ an und beschränken wir uns auf kleine Werte von x, so gelangen wir zu der Näherungsformel

$$f(x) \approx f(0) + f'(0) \cdot x.$$

Setzen wir hier für $f(x)$ verschiedene elementare Funktionen ein, so erhalten wir leicht die Näherungsformeln

$$(1 + x)^\mu \approx 1 + \mu x,$$

insbesondere

$$\sqrt{1 + x} \approx 1 + \frac{1}{2}x \qquad (1 + x > 0),$$

$$\mathrm{e}^x \approx 1 + x, \quad \ln(1 + x) \approx x \quad (1 + x > 0), \quad \sin x \approx x, \quad \tan x \approx x \quad \text{usw.},$$

von denen wir einige schon kennen.

[1] In der Tat lautet die Gleichung einer Geraden durch den Punkt (x_0, y_0) mit dem Richtungskoeffizienten k nämlich $y = y_0 + k(x - x_0)$; im Fall der Tagente hat man $y_0 = f(x_0)$, $k = f'(x_0)$ zu setzen.

Es seien hier noch einige weitere Beispiele von Näherungsformeln eines anderen Typs angeführt, die ebenfalls auf Gleichung (3) beruhen.

1. Bezeichnet man die Länge eines an beiden Enden aufgehängten schweren Fadens (Kabel, Seil, Riemen) mit $2s$, die Spannweite mit $2l$ und die Durchhangstrecke mit f (Abb. 45), so kann man zur Berechnung von s die (Näherungs-) Formel

$$s = l\left(1 + \frac{2}{3} \cdot \frac{f^2}{l^2}\right)$$

benutzen. Die Größe f wollen wir hier als unabhängige Veränderliche und s als Funktion von f ansehen.

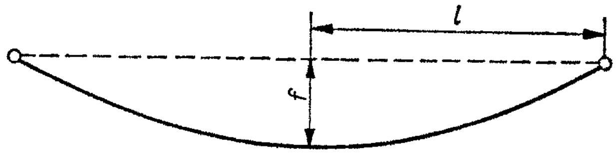

Abb. 45

Es soll der Zusammenhang zwischen der Änderung Δs der Länge s und der Änderung Δf der Durchhangstrecke f festgestellt werden. Wenn wir Δs durch ds ersetzen, so erhalten wir

$$\Delta s \approx \frac{4}{3} \cdot \frac{f}{l} \cdot \Delta f$$

und daraus

$$\Delta f \approx \frac{3}{4} \cdot \frac{l}{f} \cdot \Delta s.$$

Ist z. B. die Änderung der Länge eines Kabels bei einer Änderung der Temperatur oder der Belastung zu berücksichtigen, so läßt sich hieraus die Änderung der Durchhangstrecke angenähert vorherbestimmen.

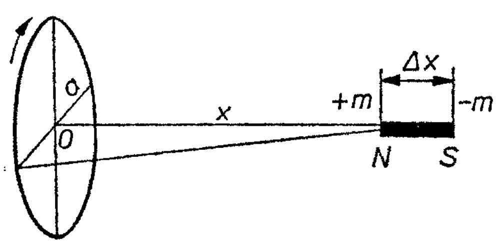

Abb. 46

2. Bekanntlich wirkt ein Kreisstrom (Abb. 46) auf die Einheit der sogenannten „magnetischen Ladung", die sich auf seiner Achse im Abstand x vom Mittelpunkt O befindet, mit der Kraft $\dfrac{k}{(a^2 + x^2)^{3/2}}$, wobei k eine Konstante und a der Radius ist. Gesucht wird ein Ausdruck für die Kraft, mit welcher der Kreisstrom auf einen Magneten NS der Länge Δx wirkt, der in der Achse des Stromes liegt. Hierbei nehmen wir an, daß im Pol N eine punktförmige positive „magnetische Ladung" m konzentriert ist und im Pol S eine gleich große negative „Ladung" $-m$. Die auf den Magneten wirkende Gesamtkraft F des Stromes läßt sich folgendermaßen ausdrücken:

$$F = \frac{km}{(a^2 + x^2)^{3/2}} - \frac{km}{[a^2 + (x + \Delta x)^2]^{3/2}} = -km \cdot \Delta\left[\frac{1}{(a^2 + x^2)^{3/2}}\right].$$

Ersetzen wir den Zuwachs der Funktion (unter der Voraussetzung, daß Δx klein ist) durch ihr Differential, so erhalten wir

$$F \approx -km \cdot \mathrm{d}\left[\frac{1}{(a^2 + x^2)^{3/2}}\right] = 3k \cdot m\Delta x \cdot \frac{x}{(a^2 + x^2)^{5/2}}.$$

108. Anwendung von Differentialen bei der Fehlerabschätzung. Besonders vorteilhaft ist der Begriff des Differentials bei Näherungsrechnungen zur Fehlerabschätzung anzuwenden. Beispielsweise werde die Größe x gemessen oder direkt berechnet und die von ihr abhängige Größe y nach einer Formel, etwa $y = f(x)$, bestimmt. Beim Messen von x ergibt sich gewöhnlich ein Fehler Δx, der einen Fehler Δy für y nach sich zieht. Da diese Fehler als klein angenommen werden, setzt man $\Delta y = y_x \cdot \Delta x$, d. h., man ersetzt den Zuwachs durch das Differential. Es sei δx der maximale absolute Fehler der Größe x:

$$|\Delta x| \leqq \delta x$$

(unter normalen Bedingungen ist diese Fehlergrenze bei einer Messung bekannt). Dann können wir offenbar als maximalen absoluten Fehler (als Fehlergrenze) für y

$$\delta y = |y_x| \cdot \delta x \tag{12}$$

ansetzen.

1. Es werde z. B. zur Bestimmung des Volumens einer Kugel ihr Durchmesser D (mit Hilfe eines Stangenzirkels, eines Dickenmessers, eines Mikrometers usw.) direkt gemessen und danach das Volumen V nach der Formel

$$V = \frac{\pi}{6} D^3$$

berechnet.

Wegen $V_D = \frac{\pi}{2} D^2$ haben wir in diesem Fall nach (12)

$$\delta V = \frac{\pi}{2} D^2 \cdot \delta D.$$

Wenn wir diese Gleichung durch die vorhergehende dividieren, so erhalten wir

$$\frac{\delta V}{V} = 3 \frac{\delta D}{D},$$

so daß der (maximale) relative Fehler des berechneten Volumenwertes dreimal so groß ist wie der (maximale) relative Fehler des für den Durchmesser gemessenen Wertes.

2. Wird eine Zahl x, deren dekadischer Logarithmus $y = \log x$ berechnet werden soll, mit irgendeinem Fehler behaftet vorgegeben, so wirkt sich dies auch auf den Logarithmus aus.

Hier gilt $y_x = \frac{M}{x}$ ($M \approx 0{,}4343$), so daß nach (12)

$$\delta y = 0{,}4343 \cdot \frac{\delta x}{x}$$

gilt. Somit ergibt sich der (maximale) absolute Fehler des Logarithmus einfach aus dem (maximalen) relativen Fehler der Zahl selbst, und umgekehrt.

Dieses Ergebnis wird vielfach angewandt. Zum Beispiel kann man damit eine Vorstellung von der Genauigkeit eines gewöhnlichen Rechenstabes der Länge 25 cm = 250 mm gewinnen. Beim Ablesen oder Einstellen des Läufers kann man sich um etwa 0,1 mm nach der einen oder anderen Seite irren, was beim Logarithmus einen Fehler von

$$\delta y = \frac{0{,}1}{250} = 0{,}0004$$

verursacht. Hieraus ergibt sich nach unserer Formel

$$\frac{\delta x}{x} = \frac{0{,}0004}{0{,}4343} = 0{,}00092\ldots \approx 0{,}001.$$

Der relative Fehler beim Ablesen ist also für alle Teile der Skala der gleiche.

3. Bei der Bestimmung eines Winkels φ mit Hilfe einer logarithmisch-trigonometrischen Tafel entsteht die Frage, ob man vorteilhafter die Sinus- oder die Tangenstafel benutzt. Wir setzen

$$y_1 = \log \sin \varphi \quad \text{und} \quad y_2 = \log \tan \varphi$$

und nehmen gleiche maximale Fehler δy_1 und δy_2 an (etwa jeweils gleich der Hälfte der letzten Stelle der Mantisse). Bezeichnet man die entsprechenden maximalen Fehler des Winkels φ mit $\delta_1\varphi$ und $\delta_2\varphi$, so erhält man wie oben

$$\delta y_1 = \frac{M}{\sin \varphi} \cdot \cos \varphi \cdot \delta_1 \varphi, \qquad \delta y_2 = \frac{M}{\tan \varphi} \cdot \sec^2 \varphi \cdot \delta_2 \varphi,$$

also

$$\delta_2 \varphi = \delta_1 \varphi \cdot \cos^2 \varphi < \delta_1 \varphi.$$

Somit ergibt sich, daß die Tangenstafel bei gleichem Fehler im Logarithmus einen kleineren Fehler für den Winkel erzeugt als die Sinustafel, also vorteilhafter ist.[1])

4. Als letztes Beispiel wollen wir das Problem der Genauigkeit der Messung eines unbekannten Widerstandes y mit Hilfe einer Wheatstoneschen Brücke betrachten (Abb. 47). Hierbei wird ein beweglicher Kontakt D über einen graduierten Schieber AC so lange verschoben, bis das Galvanometer G keinen Strom mehr anzeigt. Der Widerstand y wird durch die Formel

$$y = \frac{Rx}{a - x} \tag{13}$$

gegeben, wobei $a = AC$, $x = AD$ und R der bekannte Widerstand des Zweiges BC ist.

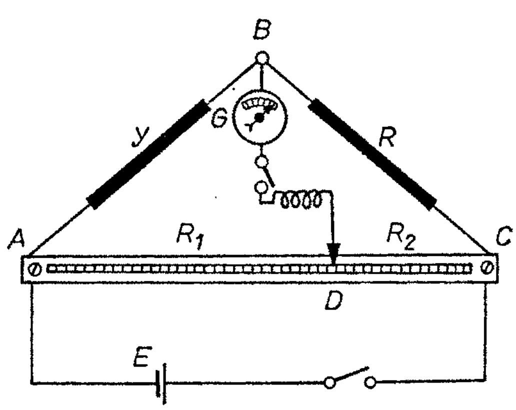

Abb. 47

Nach Formel (12) erhält man

$$\delta y = \left(\frac{Rx}{a - x}\right)_x \cdot \delta x = \frac{aR}{(a - x)^2} \cdot \delta x.$$

Wird diese Gleichung durch (13) dividiert, so erhält man für den (maximalen) relativen Fehler von y den Ausdruck

$$\frac{\delta y}{y} = \frac{a \cdot \delta x}{x(a - x)}.$$

[1]) Bei diesen Berechnungen nahmen wir die Winkel im Bogenmaß an; die Ergebnisse bleiben natürlich gültig, unabhängig davon, in welcher Einheit die Winkel gemessen wurden.

Da der Nenner $x(a - x)$ seinen größten Wert für $x = \frac{a}{2}$ annimmt[1]) und der Fehler δx bei der Längenmessung als von x unabhängig angesehen werden kann, ergibt sich der kleinste Wert für den relativen Fehler bei $x = \frac{a}{2}$. Um ein möglichst genaues Ergebnis zu erhalten, wird daher gewöhnlich der Widerstand R so gewählt, daß der Strom gerade dann verschwindet, wenn sich der Kontakt D etwa in der Mitte der Strecke AC befindet.

§ 3. Grundlegende Sätze der Differentialrechnung

109. Der Satz von Fermat. Die Kenntnis der Ableitung $f'(x)$ einer Funktion $f(x)$ ermöglicht oft Schlußfolgerungen über das Verhalten der Funktion $f(x)$ selbst. In diesem und den folgenden Paragraphen befassen wir uns im wesentlichen mit derartigen Problemen.

Zur Vorbereitung beweisen wir das folgende Lemma.

Lemma. *Die Funktion $f(x)$ habe im Punkt x_0 eine endliche Ableitung. Ist $f'(x_0) > 0$ [bzw. $f'(x_0) < 0$], so gilt für alle rechts von x_0 und in hinreichender Nähe von x_0 gelegenen Punkte x die Beziehung $f(x) > f(x_0)$ [bzw. $f(x) < f(x_0)$] und für alle links von x_0 gelegenen und zu x_0 hinreichend benachbarten Werte von x die Beziehung $f(x) < f(x_0)$ [bzw. $f(x) > f(x_0)$].*

Dafür sagt man auch, die Funktion $f(x)$ *wachse* oder *steige* (bzw. *falle*) *im Punkt* x_0. Handelt es sich um eine einseitige Ableitung, z. B. eine rechtsseitige, so bleibt nur die Aussage für die x-Werte, die rechts von x_0 liegen, gültig.

Beweis. Nach Definition der Ableitung gilt

$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}.$$

Ist $f'(x_0) > 0$ (auf diesen Fall wollen wir uns beschränken), so läßt sich nach Nr. 55, I, Satz 2, eine solche Umgebung $(x_0 - \delta, x_0 + \delta)$ des Punktes x_0 angeben, in der (für $x \neq x_0$)

$$\frac{f(x) - f(x_0)}{x - x_0} > 0 \tag{*}$$

ist. Es sei zunächst $x_0 < x < x_0 + \delta$, also $x - x_0 > 0$. Aus der Ungleichung (*) folgt dann $f(x) - f(x_0) > 0$, d. h. $f(x) > f(x_0)$. Ist aber $x_0 - \delta < x < x_0$, also $x - x_0 < 0$, so ist offenbar auch $f(x) - f(x_0) < 0$, also $f(x) < f(x_0)$. Damit ist das Lemma bewiesen.

Satz von Fermat. *Die Funktion $f(x)$ sei in einem Intervall $\mathcal{X}$ definiert und nehme in einem inneren Punkt c dieses Intervalls einen größten (kleinsten) Wert an. Stimmen die*

---

[1]) Aus der trivialen Ungleichung $x^2 - ax + \frac{a^2}{4} = \left(x - \frac{a}{2}\right)^2 \geqq 0$ erhalten wir unmittelbar

$$x(a - x) \leqq \frac{a^2}{4},$$

womit unsere Behauptung bewiesen ist.

beiderseitigen endlichen Ableitungen in diesem Punkt überein, so ist notwendigerweise $f'(c) = 0$.[1])

Beweis. Die Funktion nehme etwa im Punkt c einen größten Wert an. Die Annahme $f'(c) \neq 0$ führt auf einen Widerspruch. Entweder wäre nämlich $f'(c) > 0$, und dann würde nach dem Lemma für in hinreichender Nähe von c liegende $x > c$ die Beziehung $f(x) > c$ gelten, oder aber es wäre $f'(c) < 0$, und dann würde $f(x) > f(c)$ für alle zu c hinreichend benachbarten $x < c$ sein. In beiden Fällen wäre $f(c)$ kein größter Wert von $f(x)$ im Intervall $\mathcal{X}$. Aus diesem Widerspruch folgt die Richtigkeit des Satzes.

Wir erinnern (Nr. 91, 92) an die geometrische Deutung der Ableitung $y' = f'(x)$ als Richtungskoeffizient der Tangente an die Kurve $y = f(x)$. Wird die Ableitung $f'(c)$ gleich 0, so bedeutet dies geometrisch, daß in dem entsprechenden Punkt dieser Kurve die Tangente parallel zur x-Achse verläuft. Abb. 48 veranschaulicht diesen Sachverhalt.

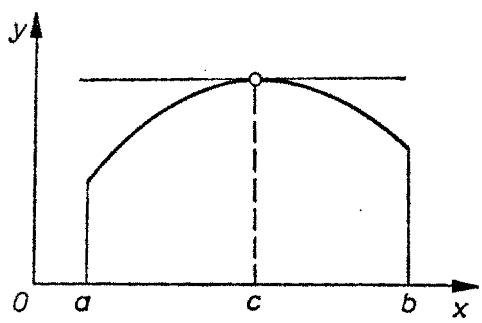

Abb. 48

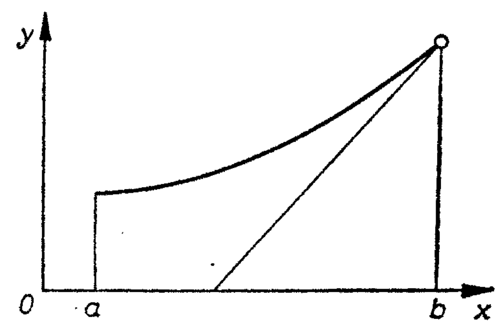

Abb. 49

Beim Beweis wurde die Annahme wesentlich benutzt, daß c ein innerer Punkt des Intervalls ist, da wir sowohl Punkte x rechts von c als auch Punkte links von c in die Betrachtung einbeziehen mußten. Ohne diese Voraussetzung wäre der Satz falsch: *Ist die Funktion* $f(x)$ *in einem abgeschlossenen Intervall definiert und nimmt sie ihren größten* (*kleinsten*) *Wert in einem der Endpunkte dieses Intervalls an, so braucht die Ableitung* $f'(x)$ *in diesem Endpunkt* (*falls sie existiert*) *nicht* 0 *zu sein*. Wir überlassen es dem Leser, ein entsprechendes Beispiel zu finden. Geometrisch wird dieser Sachverhalt durch Abb. 49 veranschaulicht.

Als Anwendung des Satzes von Fermat beweisen wir einen interessanten Satz über die Ableitung einer Funktion.

110. Der Satz von Darboux[2]). *Besitzt eine Funktion* $f(x)$ *im Intervall* $[a, b]$[3]) *eine endliche Ableitung* $f'(x)$, *so nimmt* $f'(x)$ *jeden Wert im Intervall zwischen* $f'(a)$ *und* $f'(b)$ *an.*

Beweis. Zunächst wollen wir voraussetzen, daß $f'(a)$ und $f'(b)$ verschiedene Vorzeichen haben, etwa $f'(a) > 0$ und $f'(b) < 0$, und die Existenz eines Punktes c zwischen a und b beweisen, in welchem die Ableitung 0 ist.

[1]) Diese Aussage gibt im Grunde nur das Wesen des Verfahrens wieder, das Pierre de Fermat (1601–1665, französischer Jurist und Mathematiker) zum Aufsuchen der größen und der kleinsten Funktionswerte benutzte (Fermat kannte den Begriff der Ableitung noch nicht).

[2]) Gaston Darboux, 1842–1917, französischer Mathematiker.

[3]) Wir verstehen darunter, daß im Punkt a die *rechtsseitige* und im Punkt b die *linksseitige* Ableitung existiert. Diese Ableitungen werden im folgenden einfach mit $f'(a)$ und $f'(b)$ bezeichnet.

Aus der Existenz der endlichen Ableitung $f'(x)$ in $[a, b]$ folgt die Stetigkeit der Funktion $f(x)$ in diesem Intervall (Nr. 96, Satz 2); somit nimmt $f(x)$ nach dem zweiten Weierstraßschen Satz (Nr. 85) in einem Punkt c seinen größten Wert an. Dieser Punkt c kann weder mit a noch mit b übereinstimmen, da $f(x)$ nach dem Lemma in allen dem Punkt a rechts benachbarten Punkten größer als $f(a)$ und in allen dem Punkt b links benachbarten Punkten größer als $f(b)$ ist. Somit gilt $a < c < b$. Dann ist aber nach dem Satz von FERMAT $f'(c) = 0$.

Um zum allgemeinen Fall überzugehen, wählen wir eine beliebige Zahl C zwischen $f'(a)$ und $f'(b)$. Es sei etwa $f'(a) > C > f'(b)$. Wir betrachten die Hilfsfunktion $\varphi(x) = f(x) - Cx$. Sie ist stetig und besitzt im Intervall $[a, b]$ die Ableitung $\varphi'(x) = f'(x) - C$. Wegen $\varphi'(a) = f'(a) - C > 0$ und $\varphi'(b) = f'(b) - C < 0$ existiert nach dem bereits Bewiesenen ein Punkt c $(a < c < b)$, in dem

$$\varphi'(c) = f'(c) - C = 0, \quad \text{also} \quad f'(c) = C$$

gilt.

Der soeben bewiesene Satz hat große Ähnlichkeit mit dem zweiten Satz von CAUCHY (Nr. 82), nach dem jede stetige Funktion alle Zwischenwerte annimmt. Der Satz von DARBOUX ist jedoch keinesfalls eine Folgerung aus dem Cauchyschen Satz; denn die Ableitung $f'(x)$ einer stetigen Funktion $f(x)$ braucht nicht stetig zu sein [vgl. beispielsweise in Nr. 102 die Funktion $f(0) = 0$, $f(x) = x^2 \sin \frac{1}{x}$ für $x \neq 0$].

111. Der Satz von Rolle. Vielen Sätzen und Formeln der Differentialrechnung und ihren Anwendungen liegt der folgende einfache, aber wichtige Satz zugrunde, der nach ROLLE benannt wird.[1])

Satz von ROLLE. *Die Funktion $f(x)$ sei im abgeschlossenen Intervall $[a, b]$ definiert und stetig; sie besitze wenigstens im offenen Intervall (a, b) eine endliche Ableitung $f'(x)$, und es seien die Werte der Funktion in den Endpunkten des Intervalls einander gleich, d. h., es sei $f(a) = f(b)$. Dann gibt es zwischen a und b einen Punkt c $(a < c < b)$, in dem $f'(c) = 0$ ist.*

Beweis. Die Funktion $f(x)$ ist stetig im abgeschlossenen Intervall $[a, b]$; daher nimmt sie nach dem zweiten Weierstraßschen Satz (Nr. 85) in diesem Intervall sowohl ihren größten Wert M als auch ihren kleinsten Wert m an.

Wir betrachten die beiden Fälle:

1. $M = m$. Dann ist $f(x)$ im Intervall $[a, b]$ konstant; denn die Ungleichung $m \leqq f(x) \leqq M$ liefert $f(x) = M$ für alle x. Daher gilt $f'(x) = 0$ im ganzen Intervall, so daß man für c jeden Punkt aus (a, b) nehmen kann.

2. $M > m$. Wir wissen, daß diese beiden Werte von der Funktion angenommen werden. Wegen $f(a) = f(b)$ wird aber wenigstens einer von ihnen in einem Punkt c zwischen a und b angenommen. In diesem Fall folgt aus dem Satz von FERMAT, daß die Ableitung $f'(x)$ in diesem Punkt gleich 0 ist. Damit ist der Satz bewiesen.

Geometrisch besagt der Satz von ROLLE folgendes: Wenn die äußeren Ordinaten der Kurve $y = f(x)$ in den Endpunkten einander gleich sind, dann läßt sich (mindestens) ein Punkt der Kurve angeben, in dem die Tangente parallel zur x-Achse verläuft (Abb. 50).

---

[1]) Tatsächlich hat MICHEL ROLLE (1652—1719, französischer Mathematiker) diese Behauptung nur für Polynome formuliert.

Wir machen darauf aufmerksam, daß die Stetigkeit der Funktion $f(x)$ im abgeschlossenen Intervall $[a, b]$ und die Existenz der Ableitung im ganzen offenen Intervall (a, b) für die Gültigkeit des Satzes wesentlich sind.

Die Funktion $f(x) = x - [x]$ erfüllt z. B. im Intervall $[0, 1]$ alle Voraussetzungen des Satzes, mit der Ausnahme, daß sie in $x = 1$ (am rechten Ende) eine Unstetigkeit besitzt. Jedoch ist im ganzen Intervall $(0, 1)$ die Ableitung $f'(x)$ von 0 verschieden, nämlich gleich 1.

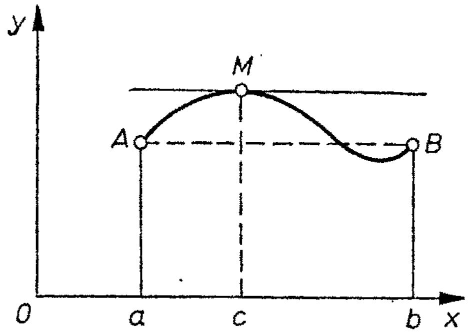

Abb. 50

Die durch $f(x) = x$ für $0 \leqq x \leqq \frac{1}{2}$ und $f(x) = 1 - x$ für $\frac{1}{2} \leqq x \leqq 1$ definierte Funktion erfüllt in demselben Intervall ebenfalls alle Voraussetzungen, mit der einzigen Ausnahme, daß in $x = \frac{1}{2}$ keine (beiderseitige) Ableitung existiert; in der linken Hälfte des Intervalls ist $f'(x) = 1$, in der rechten ist $f'(x) = -1$, d. h., es existiert auch hier kein Punkt c, in dem $f'(c) = 0$ ist. Ebenso wesentlich ist die dritte Voraussetzung des Satzes. Die Funktion $f(x) = x$ erfüllt im Intervall $[0, 1]$ alle Bedingungen des Satzes, außer der Bedingung $f(a) = f(b)$, und ihre Ableitung $f'(x)$ ist überall gleich 1.

Die Zeichnungen möge der Leser selbst anfertigen.

112. Der erste Mittelwertsatz der Differentialrechnung. Wir wenden uns jetzt einigen unmittelbaren Folgerungen aus dem Satz von Rolle zu.

Erster Mittelwertsatz.[1]) *Die Funktion $f(x)$ sei im abgeschlossenen Intervall $[a, b]$ definiert und stetig, und sie besitze wenigstens im offenen Intervall (a, b) eine endliche Ableitung $f'(x)$.*[2]) *Dann gibt es zwischen a und b einen Punkt c $(a < c < b)$ derart, daß die Gleichung*

$$\frac{f(b) - f(a)}{b - a} = f'(c) \tag{1}$$

erfüllt ist.

Beweis. Wir führen eine Hilfsfunktion $F(x)$ ein, die in $[a, b]$ durch folgende Gleichung definiert ist:

$$F(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a).$$

[1]) Auch *Satz von* Lagrange genannt (Joseph Louis Lagrange, 1736–1813, französischer Mathematiker).

[2]) Natürlich folgt die Stetigkeit von $f(x)$ in (a, b), die wir besonders vorausgesetzt haben, aus der zweiten Voraussetzung; aber weder hier noch im folgenden legen wir Wert darauf, die Voraussetzungen in voneinander unabhängigen Teilen zu formulieren.

Diese Funktion genügt allen Voraussetzungen des Satzes von Rolle. Sie ist als Differenz einer stetigen Funktion $f(x)$ und einer linearen Funktion stetig in $[a, b]$; im Intervall (a, b) besitzt sie eine wohlbestimmte endliche Ableitung

$$F'(x) = f'(x) - \frac{f(b) - f(a)}{b - a}.$$

Schließlich überzeugen wir uns durch direktes Einsetzen davon, daß $F(a) = F(b) = 0$ ist, d. h., daß $F(x)$ in den Endpunkten des Intervalls gleiche Werte annimmt. Folglich läßt sich auf die Funktion $F(x)$ der Satz von Rolle anwenden und somit auf die Existenz eines Punktes c in (a, b) schließen, für den $F'(c) = 0$ gilt. Somit ist

$$f'(c) - \frac{f(b) - f(a)}{b - a} = 0, \quad \text{also} \quad f'(c) = \frac{f(b) - f(a)}{b - a},$$

was zu beweisen war.

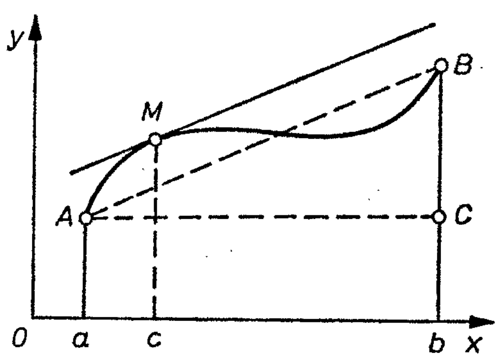

Abb. 51

Der Satz von Rolle ist ein Spezialfall des Mittelwertsatzes. Die obigen Bemerkungen über die erste und die zweite Voraussetzung des Satzes behalten auch hier ihre Gültigkeit. Zur geometrischen Deutung des Mittelwertsatzes (Abb. 51) bemerken wir, daß der Quotient

$$\frac{f(b) - f(a)}{b - a} = \frac{\overline{CB}}{\overline{AC}}$$

den Richtungskoeffizienten der Sekante AB darstellt, während $f'(c)$ der Richtungskoeffizient der Tangente an die Kurve $y = f(x)$ im Punkt mit der Abszisse $x = c$ ist. Somit ist die Aussage des Mittelwertsatzes gleichwertig mit der folgenden: *Auf dem Bogen $\overset{\frown}{AB}$ gibt es stets wenigstens einen Punkt M, in welchem die Tangente parallel zur Sehne $\overline{AB}$ verläuft.* Die Formel

$$\frac{f(b) - f(a)}{b - a} = f'(c) \quad \text{oder} \quad f(b) - f(a) = f'(c) \cdot (b - a)$$

heißt *Lagrangesche Formel*[1]). Sie gilt offenbar auch für den Fall $a > b$.

Wir wählen einen beliebigen Wert x_0 im Intervall $[a, b]$ und erteilen ihm den Zuwachs $\Delta x \gtrless 0$, der nicht über die Grenzen des Intervalls hinausführt. Wir wollen nun die Lagrangesche Formel auf das Intervall $[x_0, x_0 + \Delta x]$ für $\Delta x > 0$ oder auf das

[1]) Auch *Formel der endlichen Zuwächse* genannt.

Intervall $[x_0 + \Delta x, x_0]$ für $\Delta x < 0$ anwenden. Die Zahl c, die zwischen x_0 und $x_0 + \Delta x$ liegt, können wir für beide Fälle in der Form

$$c = x_0 + \theta \cdot \Delta x \quad (0 < \theta < 1)$$

darstellen.[1]) Dann erhält die Formel von LAGRANGE die Gestalt

$$\frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} = f'(x_0 + \theta \Delta x) \tag{1a}$$

oder

$$\Delta f(x_0) = f(x_0 - \Delta x) - f(x_0) = f'(x_0 + \theta \Delta x) \cdot \Delta x \quad (0 < \theta < 1). \tag{2}$$

Diese Gleichung, die den genauen Ausdruck für den Zuwachs der Funktion bei beliebigem endlichem Zuwachs Δx des Arguments liefert, stellt man naturgemäß der Näherungsgleichung [Nr. 107, Formel (3a)]

$$\Delta f(x_0) = f(x_0 + \Delta x) - f(x_0) \approx f'(x_0) \cdot \Delta x$$

gegenüber, deren relativer Fehler gegen 0 strebt, sobald Δx unendlich klein ist. Daher stammt die Benennung „Formel der endlichen Zuwächse".

Die Formel von LAGRANGE hat den Nachteil, daß darin die unbekannte Zahl θ (oder c) auftritt.[2]) Das macht sich jedoch bei vielen Anwendungen dieser Formel in der Analysis nicht störend bemerkbar.

113. Der Grenzwert der Ableitung. Ein nützliches Beispiel einer solchen Anwendung liefert die folgende Bemerkung. Wir setzen voraus, die Funktion $f(x)$ sei im Intervall $[x_0, x_0 + H]$, $H > 0$, stetig und besitze für $x > x_0$ eine endliche Ableitung $f'(x)$. *Existiert der (endliche oder unendliche) Grenzwert*

$$\lim_{x \to x_0 + 0} f'(x) = K,$$

so ist er die rechtsseitige Ableitung im Punkt x_0.

Für $0 < \Delta x < H$ gilt die Beziehung (1a). Für $\Delta x \to 0$ strebt, da θ beschränkt ist, das Argument der Ableitung, nämlich $x_0 + \theta \Delta x$, gegen x_0, so daß die rechte Seite dieser Gleichung, und mit ihr auch die linke, gegen den Grenzwert K strebt, was zu beweisen war.

Die analoge Aussage läßt sich natürlich unter entsprechenden Voraussetzungen auch für die linke Umgebung des Punktes x_0 herleiten.

Als Beispiel betrachten wir die Funktion

$$f(x) = x \arcsin x + \sqrt{1 - x^2}$$

im Intervall $[-1, 1]$. Ist $-1 < x < 1$, so ergibt sich nach den gewöhnlichen Regeln der Differentialrechnung leicht

$$f'(x) = \arcsin x + \frac{x}{\sqrt{1 - x^2}} - \frac{x}{\sqrt{1 - x^2}} = \arcsin x.$$

---

[1]) Manchmal sagt man auch, θ sei ein „echter Bruch"; damit ist dann aber eventuell ein unendlicher nichtperiodischer Dezimalbruch zwischen 0 und 1 gemeint, da θ keine rationale Zahl zu sein braucht.

[2]) In einfachen Fällen läßt sich diese Zahl bestimmen; beispielsweise ist für die quadratische Funktion $f(x) = ax^2 + bx + c$, wie man leicht verifiziert, $\theta = \frac{1}{2}$. Selbstverständlich ist im allgemeinen die Zahl θ von x_0 und Δx abhängig.

Für $x \to 1 - 0$ (bzw. $x \to -1 + 0$) strebt diese Ableitung offenbar gegen den Grenzwert $\frac{\pi}{2}$ (bzw. $-\frac{\pi}{2}$). Somit existieren für $x = \pm 1$ die (einseitigen) Ableitungen

$$f'(+1) = \frac{\pi}{2}, \qquad f'(-1) = -\frac{\pi}{2}.$$

Die obige Bemerkung wird oft bei folgenden Sachverhalten angewandt: Aus der Tatsache, daß der Ausdruck für die Ableitung gegen ∞ (bzw. $-\infty$) strebt, wenn sich x von der einen oder anderen Seite her dem Wert x_0 nähert, schließt man, daß im Punkt x_0 selbst die entsprechende einseitige Ableitung gleich ∞ (bzw. $-\infty$) ist.

Als Beispiel betrachten wir nochmals die Funktionen $f_1(x) = x^{1/3}$ und $f_2(x) = x^{2/3}$, die wir schon in Nr. 101 behandelt hatten. Für sie gilt (für $x \gtrless 0$, wobei für $x < 0$ unter $x^{1/3}$ die Funktion $-|x|^{1/3}$ verstanden wird)

$$f_1'(x) = \frac{1}{3x^{2/3}}, \quad f_2'(x) = \frac{2}{3x^{1/3}}.$$

Da der erste dieser Ausdrücke für $x \to \pm 0$ gegen ∞ strebt und der zweite für $x \to +0$ bzw. $x \to -0$ die Grenzwerte ∞ bzw. $-\infty$ besitzt, können wir schließen, daß für $f_1(x)$ im Punkt $x = 0$ die beiderseitige Ableitung existiert und gleich ∞ ist, während $f_2(x)$ in diesem Punkt nur einseitige Ableitungen besitzt, nämlich ∞ als rechte und $-\infty$ als linke Ableitung.

Hieraus ergibt sich ferner folgende Tatsache: Existiert eine endliche Ableitung $f'(x)$ in einem bestimmten Intervall, so ist sie eine Funktion, die dort weder hebbare Unstetigkeiten noch Sprungstellen haben kann. Sie ist in jedem Punkt entweder stetig oder besitzt eine Unstetigkeit zweiter Art (vgl. Nr. 102, Punkt 2).

114. Die Cauchysche Formel (zweiter Mittelwertsatz)[1]. Der erste Mittelwertsatz läßt sich in folgender Weise verallgemeinern.

Satz von Cauchy. *Die Funktionen $f(x)$ und $g(x)$ seien im abgeschlossenen Intervall $[a, b]$ stetig; wenigstens im offenen Intervall (a, b) seien die endlichen Ableitungen $f'(x)$ und $g'(x)$ vorhanden; im ganzen Intervall (a, b) sei $g'(x) \neq 0$. Dann existiert zwischen a und b ein Punkt c $(a < c < b)$ derart, daß*

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(c)}{g'(c)} \tag{3}$$

gilt.

Diese Formel nennt man *Cauchysche Formel* oder *zweiten Mittelwertsatz der Differentialrechnung.*

Beweis. Zunächst zeigen wir, daß der Nenner auf der linken Seite unserer Gleichung nicht gleich 0 ist, da sonst der Ausdruck keinen Sinn hätte. Wäre $g(b) = g(a)$, so müßte nach dem Satz von Rolle in einem bestimmten Intervallpunkt $g'(x) = 0$ sein, entgegen der Voraussetzung $g'(x) \neq 0$. Also gilt $g(b) \neq g(a)$.

Wir betrachten jetzt die Hilfsfunktion

$$F(x) = f(x) - f(a) - \frac{f(b) - f(a)}{g(b) - g(a)} [g(x) - g(a)].$$

[1]) auch *erweiterter Mittelwertsatz* genannt.

Sie genügt allen Bedingungen des Satzes von ROLLE. Erstens ist sie stetig in $[a, b]$, da $f(x)$ und $g(x)$ stetig sind. Zweitens existiert in (a, b) die Ableitung $F'(x)$, nämlich

$$F'(x) = f'(x) - \frac{f(b) - f(a)}{g(b) - g(a)} \cdot g'(x).$$

Schließlich überzeugen wir uns durch direktes Einsetzen davon, daß $F(a) = F(b) = 0$ gilt. Infolgedessen existiert im Intervall (a, b) ein Punkt c derart, daß $F'(c) = 0$ ist. Mit anderen Worten, es gilt

$$f'(c) - \frac{f(b) - f(a)}{g(b) - g(a)} \cdot g'(c) = 0, \quad \text{also} \quad f'(c) = \frac{f(b) - f(a)}{g(b) - g(a)} \cdot g'(c).$$

Wenn wir durch $g'(c)$ dividieren [das ist wegen $g'(c) \neq 0$ möglich], so erhalten wir die gesuchte Gleichung.

Offenbar ist der erste Mittelwertsatz ein Spezialfall des zweiten. Er ergibt sich aus diesem, wenn man $g(x) = x$ setzt. Geometrisch ist der Satz von CAUCHY ebenso zu deuten wie der Satz von LAGRANGE. Um dies dem Leser besser verständlich zu machen, gehen wir zu einer anderen Bezeichnung über. Wir ersetzen x durch t und bezeichnen die Funktionen mit $\varphi(t)$ und $\psi(t)$. Variiert t im Intervall $[\alpha, \beta]$, so können wir die Cauchysche Formel folgendermaßen schreiben:

$$\frac{\psi(\beta) - \psi(\alpha)}{\varphi(\beta) - \varphi(\alpha)} = \frac{\psi'(\gamma)}{\varphi'(\gamma)} \quad (\alpha < \gamma < \beta). \tag{4}$$

Wir betrachten jetzt eine Kurve, die durch die Parametergleichungen

$$x = \varphi(t), \quad y = \psi(t) \quad (\alpha \leqq t \leqq \beta) \tag{5}$$

gegeben ist. Dann ist die linke Seite der Formel (4) auch hier der Richtungskoeffizient der Sehne, die die Enden des Kurvenbogens verbindet, und die rechte der Richtungskoeffizient der Tangente in einem inneren Punkt des Bogens, der dem Wert $t = \gamma$ entspricht [Nr. 106, Formel (11)].

Bemerkung. Diese Überlegungen legen den Gedanken nahe, die Cauchysche Formel aus der Formel von LAGRANGE zu folgern, wie wir es taten. Das Wesen dieser Herleitung besteht darin, daß man aus der Parameterdarstellung (5) die unmittelbare Abhängigkeit $y = f(x)$ ermittelt, und dann erweisen sich Formel (4) und Formel (1) als gleichbedeutend.

§ 4. Ableitungen und Differentiale höherer Ordnung

115. Definition der Ableitungen höherer Ordnung. Hat eine Funktion $y = f(x)$ in einem bestimmten Intervall $\mathscr{X}$ eine endliche Ableitung $y' = f'(x)$, so daß diese wieder eine Funktion von x ist, so kann es vorkommen, daß diese Funktion in einem Punkt x_0 aus $\mathscr{X}$ ihrerseits eine Ableitung besitzt, die endlich oder auch unendlich sein kann. Diese Ableitung nennt man die *Ableitung zweiter Ordnung* oder einfach die *zweite Ableitung* der Funktion $y = f(x)$ in dem betreffenden Punkt und bezeichnet sie mit einem der Symbole

$$\frac{d^2y}{dx^2}, y'', D^2y; \quad \frac{d^2f(x_0)}{dx^2}, f''(x_0), D^2f(x_0).$$

So haben wir z. B. in Nr. 92 gesehen, daß die Geschwindigkeit v der Bewegung eines Punktes gleich der Ableitung des von dem Punkt durchlaufenen Weges nach der Zeit t ist, $v = \frac{\mathrm{d}s}{\mathrm{d}t}$, die Beschleunigung a aber die Ableitung der Geschwindigkeit v nach der Zeit, also $a = \frac{\mathrm{d}v}{\mathrm{d}t}$. Also ist die Beschleunigung die zweite Ableitung des Weges nach der Zeit, $a = \frac{\mathrm{d}^2 s}{\mathrm{d}t^2}$.

Hat $y = f(x)$ im ganzen Intervall $\mathcal{X}$ (d. h. in jedem Punkt dieses Intervalls) eine endliche zweite Ableitung, so nennt man deren (endliche oder unendliche) Ableitung in einem Punkt x_0 aus $\mathcal{X}$, sofern sie existiert, die *Ableitung dritter Ordnung* oder einfach die *dritte Ableitung* der Funktion $y = f(x)$ in diesem Punkt und bezeichnet sie folgendermaßen:

$$\frac{\mathrm{d}^3 y}{\mathrm{d}x^3},\ y''',\ \mathrm{D}^3 y; \quad \frac{\mathrm{d}^3 f(x_0)}{\mathrm{d}x^3},\ f'''(x_0),\ \mathrm{D}^3 f(x_0).$$

Entsprechend gehen wir von der dritten zur vierten Ableitung über usw. Ist der Begriff der $(n-1)$-ten Ableitung definiert und ist diese Ableitung im Intervall $\mathcal{X}$ vorhanden und endlich, so nennt man deren Ableitung in einem Punkt x_0 dieses Intervalls die *Ableitung n-ter Ordnung* oder die *n-te Ableitung* der Ausgangsfunktion $y = f(x)$. Zur Bezeichnung benutzt man die Symbole

$$\frac{\mathrm{d}^n y}{\mathrm{d}x^n},\ y^{(n)},\ \mathrm{D}^n y; \quad \frac{\mathrm{d}^n f(x_0)}{\mathrm{d}x^n},\ f^{(n)}(x_0),\ \mathrm{D}^n f(x_0).$$

Wenn man die Lagrangesche oder Cauchysche Symbolik benutzt, kann es sich als notwendig erweisen, die Veränderliche anzugeben, nach der differenziert wurde. Dann setzt man diese Veränderliche als unteren Index (und läßt diese Striche oben fort):

$$y_{x^2},\quad \mathrm{D}_{x^3} y,\quad f_{x^n}(x_0) \quad \text{usw.}$$

bzw.

$$y_{xx},\quad \mathrm{D}_{xxx} y,\quad f_{\underbrace{x\ldots x}_{n\text{-mal}}}(x_0) \quad \text{usw.}$$

Beispielsweise schreibt man $a = s_{t^2}$ oder $a = s_{tt}$. (Natürlich hat man auch hier die Symbole

$$\frac{\mathrm{d}^n f}{\mathrm{d}x^n},\ f^{(n)} \quad \text{bzw.} \quad f_{x^n},\ \mathrm{D}^n f \quad \text{bzw.} \quad \mathrm{D}_{x^n} f \quad \text{bzw.} \quad f_{\underbrace{x\ldots x}_{n\text{-mal}}} \quad \text{usw.}$$

als Funktionszeichen anzusehen.)

Somit haben wir den Begriff der n-ten Ableitung induktiv definiert, indem wir von der ersten Ableitung sukzessive zu den folgenden übergingen.

Die Beziehung

$$y^{(n)} = (y^{(n-1)})',$$

welche die n-te Ableitung definiert, nennt man eine *Rekursionsformel*, weil sie die n-te auf die $(n-1)$-te Ableitung „zurückführt".

Die Ableitungen n-ter Ordnung selbst lassen sich (bei gegebenem n) mit Hilfe der dem Leser bereits bekannten Regeln und Formeln berechnen.

Ist etwa

$$y = \frac{1}{2} x^4 - \frac{1}{6} x^3 + 2x^2 + \frac{4}{3} x - \frac{1}{2},$$

so erhält man

$$y' = 2x^3 - \frac{1}{2} x^2 + 4x + \frac{4}{3}, \quad y'' = 6x^2 - x + 4, \quad y''' = 12x - 1, \quad y^{(4)} = 12,$$

so daß alle folgenden Ableitungen identisch 0 sind. Für

$$y = \ln\left(x + \sqrt{x^2 + 1}\right)$$

ergibt sich beispielsweise

$$y' = \frac{1}{\sqrt{x^2+1}}, \quad y'' = -\frac{x}{(x^2+1)^{3/2}}, \quad y''' = \frac{2x^2-1}{(x^2+1)^{5/2}} \quad \text{usw.}$$

Auch bei Ableitungen höherer Ordnung kann der Begriff der einseitigen Ableitung (Nr. 100) induktiv eingeführt werden. Wenn eine Funktion $y = f(x)$ nur in einem bestimmten Intervall $\mathfrak{X}$ definiert ist, ist unter einer Ableitung beliebiger Ordnung in den Endpunkten des Intervalls stets die einseitige Ableitung zu verstehen.

116. Allgemeine Formeln für Ableitungen höherer Ordnung. Um die n-te Ableitung einer Funktion zu bestimmen, muß man im allgemeinen vorher die Ableitungen aller vorangehenden Ordnungen berechnen. In einigen Fällen gelingt es jedoch, einen allgemeinen Ausdruck für die n-te Ableitung aufzustellen, der unmittelbar von n abhängt und keine der vorhergehenden Ableitungen enthält.

Zur Herleitung solcher allgemeinen Ausdrücke sind die Formeln

$$(c \cdot u)^{(n)} = c \cdot u^{(n)}, \quad (u \pm v)^{(n)} = u^{(n)} \pm v^{(n)}$$

dienlich, welche die dem Leser bereits bekannten Regeln I und II aus Nr. 97 auf den Fall höherer Ableitungen verallgemeinern. Sie ergeben sich einfach durch sukzessive Anwendung dieser Regeln.

1. Zunächst betrachten wir die Potenzfunktion $y = x^\mu$, wobei μ eine beliebige reelle Zahl ist. Falls $\mu > 0$ keine ganze Zahl ist, sei $x \geqq 0$; falls $\mu < 0$ ist, sei $x \neq 0$. Wir erhalten sukzessive

$$y' = \mu x^{\mu-1}, \quad y'' = \mu(\mu - 1)\, x^{\mu-2}, \quad y''' = \mu(\mu - 1)(\mu - 2)\, x^{\mu-3}, \ldots$$

Hieraus läßt sich leicht folgende allgemeine Regel gewinnen:

$$y^{(n)} = \mu(\mu - 1) \cdots (\mu - n + 1)\, x^{\mu-n},$$

die jedoch strenggenommen noch eines Beweises bedarf. Hierzu benutzen wir vollständige Induktion. Unter der Annahme, diese Regel gelte bereits für einen bestimmten Wert von n, differenzieren wir sie noch einmal und gelangen zu

$$\begin{aligned}[y^{(n)}]' = y^{(n+1)} &= \mu(\mu - 1) \cdots (\mu - n + 1)\, [x^{\mu-n}]' \\ &= \mu(\mu - 1) \cdots (\mu - n - 1)(\mu - n)\, x^{\mu-(n+1)},\end{aligned}$$

so daß unsere Formel sich auch für die $(n + 1)$-te Ableitung als gültig erweist. Hieraus wiederum folgt die Richtigkeit der Formel für alle natürlichen n.

Nehmen wir etwa $\mu = -1$, so erhalten wir

$$\left(\frac{1}{x}\right)^{(n)} = (-1)\,(-2)\,\cdots\,(-n)\,x^{-1-n} = \frac{(-1)^n \cdot n!}{x^{n+1}},$$

und für $\mu = -\frac{1}{2}$ ist

$$\left(\frac{1}{\sqrt{x}}\right)^{(n)} = \left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\cdots\left(-\frac{2n-1}{2}\right)x^{-1/2-n} = \frac{(-1)^n\,(2n-1)!!}{(2x)^n\,\sqrt{x}}$$

(hier ist natürlich $x > 0$) vorausgesetzt) usw.[1])

Ist μ selbst eine natürliche Zahl m, so ist die m-te Ableitung von x^m bereits die Konstante $m!$, und alle folgenden Ableitungen sind 0. Hieraus geht hervor, daß das auch für Polynome m-ten Grades der Fall ist.

2. Für den etwas allgemeineren Ausdruck

$$y = (a + bx)^\mu \quad (a, b = \text{const})$$

erhalten wir auf ebenso einfache Weise

$$y^{(n)} = \mu(\mu - 1)\cdots(\mu - n + 1)\cdot b^n \cdot (a + bx)^{\mu-n}.$$

Insbesondere ergibt sich wie oben $\left(x \neq -\frac{a}{b}\right)$

$$\left(\frac{1}{a+bx}\right)^{(n)} = \frac{(-1)^n\,n!\,b^n}{(a+bx)^{n+1}},$$

$$\left(\frac{1}{\sqrt{a+bx}}\right)^{(n)} = \frac{(-1)^n\,(2n-1)!!\,b^n}{2^n(a+bx)^n\,\sqrt{a+bx}} \qquad \left(x > -\frac{a}{b}\right).$$

3. Es sei jetzt $y = \ln x$. Zunächst gilt für $x > 0$

$$y' = (\ln x)' = \frac{1}{x}.$$

Die $(n-1)$-te Ableitung von y' entnehmen wir der entsprechenden Formel aus Beispiel 1, indem wir in dieser Formel n durch $n - 1$ ersetzen. Wir erhalten dann

$$y^{(n)} = (y')^{(n-1)} = \left(\frac{1}{x}\right)^{(n-1)} = \frac{(-1)^{n-1}\,(n-1)!}{x^n}.$$

4. Ist $y = a^x$ $(a > 0)$, so gilt

$$y' = a^x \cdot \ln a, \quad y'' = a^x \cdot (\ln a)^2, \ldots$$

Die allgemeine Formel

$$y^{(n)} = a^x(\ln a)^n$$

---

[1]) Das Zeichen $n!!$ bedeutet bei geradem n das Produkt der geraden natürlichen Zahlen von 2 bis n und bei ungeradem n das Produkt der ungeraden natürlichen Zahlen von 1 bis n. So ist z. B.

$$7!! = 1 \cdot 3 \cdot 5 \cdot 7 \quad \text{und} \quad 10!! = 2 \cdot 4 \cdot 6 \cdot 8 \cdot 10.$$

läßt sich leicht durch Induktion beweisen. Insbesondere gilt offenbar

$$(e^x)^{(n)} = e^x.$$

5. Wir setzen jetzt $y = \sin x$. Dann ist

$$y' = \cos x, \quad y'' = -\sin x, \quad y''' = -\cos x,$$
$$y^{(4)} = \sin x, \quad y^{(5)} = \cos x, \ldots$$

Auf diesem Wege ist der gesuchte allgemeine Ausdruck für die n-te Ableitung etwas unübersichtlich. Das wird jedoch sofort einfacher, wenn man die erste Ableitung in der Form $y' = \sin\left(x + \frac{\pi}{2}\right)$ schreibt. Offenbar muß man bei jeder Differentiation das Argument um $\frac{\pi}{2}$ vergrößern, so daß sich

$$(\sin x)^{(n)} = \sin\left(x + n \cdot \frac{\pi}{2}\right)$$

ergibt. Analog erhält man auch die Formel

$$(\cos x)^{(n)} = \cos\left(x + n \cdot \frac{\pi}{2}\right).$$

6. Wir betrachten die Funktion $y = \frac{1}{x^2 - a^2}$. Wenn wir sie auf die Form

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

bringen, können wir Beispiel 2 (und die zu Anfang erwähnten allgemeinen Regeln) benutzen. So erhalten wir

$$\left(\frac{1}{x^2 - a^2}\right)^{(n)} = \frac{(-1)^{(n)}\, n!}{2a}\left[\frac{1}{(x-a)^{n+1}} - \frac{1}{(x+a)^{n+1}}\right].$$

7. Im Fall der Funktion $y = e^{ax} \sin bx$ benutzen wir einen Kunstgriff. Wir erhalten nämlich

$$y' = a e^{ax} \sin bx + b e^{ax} \cos bx;$$

wenn wir durch

$$\sin\varphi = \frac{b}{\sqrt{a^2 + b^2}}, \quad \cos\varphi = \frac{a}{\sqrt{a^2 + b^2}}$$

einen Hilfswinkel φ einführen, so kann die erste Ableitung in der Form

$$y' = \sqrt{a^2 + b^2} \cdot e^{ax} \cdot (\sin bx \cdot \cos\varphi + \cos bx \cdot \sin\varphi) = \sqrt{a^2 + b^2} \cdot e^{ax} \cdot \sin(bx + \varphi)$$

geschrieben werden. Wenn wir jetzt weiter differenzieren, gelangen wir durch Induktion leicht zu der allgemeinen Regel

$$y^{(n)} = (a^2 + b^2)^{n/2} \cdot e^{ax} \sin(bx + n\varphi).$$

8. Jetzt befassen wir uns noch einmal mit der Funktion $y = \arctan x$. Wir stellen uns zunächst die Aufgabe, $y^{(n)}$ durch y auszudrücken. Wegen $x = \tan y$ gilt

$$y' = \frac{1}{1 + x^2} = \cos^2 y = \cos y \cdot \sin\left(y + \frac{\pi}{2}\right).$$

Wenn wir nochmals nach x differenzieren (und beachten, daß y eine Funktion von x ist), so erhalten wir

$$y'' = \left[-\sin y \cdot \sin\left(y + \frac{\pi}{2}\right) + \cos y \cdot \cos\left(y + \frac{\pi}{2}\right)\right] \cdot y'$$
$$= \cos^2 y \cdot \cos\left(2y + \frac{\pi}{2}\right) = \cos^2 y \cdot \sin 2\left(y + \frac{\pi}{2}\right).$$

Die weitere Differentiation liefert

$$y''' = \left[-2\sin y \cdot \cos y \cdot \sin 2\left(y + \frac{\pi}{2}\right) + 2\cos^2 y \cdot \cos 2\left(y + \frac{\pi}{2}\right)\right] \cdot y'$$
$$= 2\cos^3 y \cdot \cos\left(3y + 2 \cdot \frac{\pi}{2}\right) = 2\cos^3 y \cdot \sin 3\left(y + \frac{\pi}{2}\right).$$

Die allgemeine Formel

$$y^{(n)} = (n-1)!\cos^n y \cdot \sin n\left(y + \frac{\pi}{2}\right)$$

läßt sich durch Induktion beweisen.

Führen wir (für $x > 0$) den Winkel

$$z = \arctan\frac{1}{x} = \frac{\pi}{2} - y$$

ein, so läßt sich diese Formel folgendermaßen schreiben:

$$y^{(n)} = (n-1)!\,\frac{1}{(1+x^2)^{n/2}} \cdot \sin n(\pi - z)$$

oder schließlich in der Form

$$y^{(n)} = (-1)^{n-1}(n-1)!\,\frac{1}{(1+x^2)^{n/2}} \cdot \sin n \arctan\frac{1}{x}.$$

9. Zum Schluß wollen wir zur Übung die Formel

$$D^n(x^{n-1}e^{1/x}) = (-1)^n\frac{e^{1/x}}{x^{n+1}} \qquad (n = 1, 2, 3, \ldots)$$

herleiten. Ihre Richtigkeit für $n = 1$ und $n = 2$ prüft man unmittelbar nach. Wir nehmen jetzt an, sie gelte für alle Werte n bis zu einem bestimmten $n \geqq 2$, und beweisen, daß sie dann auch für $n + 1$ richtig ist.[1]))Zu diesem Zweck betrachten wir den Ausdruck

$$D^{n+1}(x^n e^{1/x} = D^n[D(x^n e^{1/x})] = D^n[nx^{n-1}e^{1/x} - x^{n-2}e^{1/x}]$$
$$= n \cdot D^n(x^{n-1}e^{1/x}) - D[D^{n-1}(x^{n-2}e^{1/x})].$$

Nach Induktionsannahme läßt er sich in der Form

$$D^{n+1}(x^n e^{1/x}) = n \cdot (-1)^n\frac{e^{1/x}}{x^{n+1}} - D\left[(-1)^{n-1}\frac{e^{1/x}}{x^n}\right] = (-1)^{n+1}\frac{e^{1/x}}{x^{n+2}}$$

schreiben, was zu beweisen war. Also gilt die Formel für alle natürlichen n.

117. Die Leibnizsche Formel. Wie wir zu Beginn von Nr. 116 bemerkten, lassen sich die Regeln I und II aus Nr. 97 unmittelbar auf Ableitungen beliebiger Ordnung übertragen. Komplizierter ist es bei der Regel III, die sich auf die Differentiation eines Produktes bezieht.

---

[1]) Wir weisen den Leser ausdrücklich auf diese eigenartige Form der Induktion hin. Wir benutzen hier (vgl. den Text) nämlich die Richtigkeit unserer Formel für n und $n - 1$.

Wir nehmen an, daß jede der Funktionen $u(x)$ und $v(x)$ Ableitungen bis zur n-ten Ordnung einschließlich hat, und beweisen, daß dann auch ihr Produkt $y = u \cdot v$ eine n-te Ableitung besitzt, und geben einen Ausdruck dafür an.

Unter Anwendung von Regel III differenzieren wir dieses Produkt sukzessive. Dabei erhalten wir

$$\begin{aligned} y' &= u'v + uv', \\ y'' &= u''v + 2u'v' + uv'', \\ y''' &= u'''v + 3u''v' + 3u'v'' + uv''', \ldots \end{aligned}$$

Das Gesetz, nach dem alle diese Formeln aufgebaut sind, läßt sich unschwer ermitteln: Die rechten Seiten erinnern an die Entwicklung der Potenzen des Binoms $u + v$, $(u + v)^2$, $(u + v)^3, \ldots$, nur daß an Stelle der Potenzen von u und v die Ableitungen der entsprechenden Ordnung treten. Die Ähnlichkeit tritt noch stärker hervor, wenn wir in den Formeln an Stelle von u und v die Symbole $u^{(0)}$ und $v^{(0)}$ schreiben. Wenn wir diese Regel auf beliebiges n ausdehnen, kommen wir zu der allgemeinen Formel

$$\begin{aligned} y^{(n)} = (uv)^{(n)} &= \sum_{i=0}^{n} \binom{n}{i} u^{(n-i)} v^{(i)} \\ &= u^{(n)}v + nu^{(n-1)}v' + \frac{n(n-1)}{1 \cdot 2} u^{(n-2)}v'' + \cdots \\ &\quad + \frac{n(n-1)\cdots(n-i+1)}{1 \cdot 2 \cdots i} u^{(n-1)}v^{(i)} + \cdots + uv^{(n)}. \end{aligned} \tag{1}$$

Zum Beweis dieser Formel verwenden wir wieder Induktion. Wir nehmen an, die Formel gelte für einen bestimmten Wert von n. Wenn für die Funktionen u, v auch $(n + 1)$-te Ableitungen existieren, so ist eine weitere Differentiation nach x möglich. Dann erhalten wir

$$y^{(n+1)} = \sum_{i=0}^{n} \binom{n}{i} [u^{(n-i)}v^{(i)}]' = \sum_{i=0}^{n} \binom{n}{i} u^{(n-i+1)}v^{(i)} + \sum_{i=0}^{n} \binom{n}{i} u^{(n-i)}v^{(i+1)}.$$

Jetzt vereinigen wir die Summanden der letzten beiden Summen, die dieselben Produkte der Ableitungen der Funktionen u und v enthalten (die Summe der Ordnungen der Ableitungen ist bei diesen Produkten, wie man sieht, stets gleich $n + 1$). Das Produkt $u^{(n+1)}v^{(0)}$ tritt nur in der ersten Summe auf (für $i = 0$); der Koeffizient ist $\binom{n}{0} = 1$. Ebenso ist $u^{(0)}v^{(n+1)}$ nur in der zweiten Summe (als Summand mit dem Index $i = n$) mit dem Koeffizienten $\binom{n}{n} = 1$ enthalten. Alle übrigen Produkte, die in dieser Summe auftreten, haben die Form $u^{(n+1-k)}v^{(k)}$, wobei $1 \leqq k \leqq n$ gilt. Jedes dieser Produkte kommt sowohl in der ersten Summe (als Summand mit dem Index $i = k$) als auch in der zweiten Summe (als Summand mit dem Index $i = k - 1$) vor. Die Summe der entsprechenden Koeffizienten ist $\binom{n}{k} + \binom{n}{k-1}$. Bekanntlich gilt aber

$$\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}.$$

Somit gelangen wir wegen $\binom{n+1}{0} = \binom{n+1}{n+1} = 1$ schließlich zu

$$y^{(n+1)} = u^{(n+1)}v^{(0)} + \sum_{k=1}^{n} \binom{n+1}{k} u^{[(n+1)-k]}v^{(k)} + u^{(0)}v^{(n+1)}$$

$$= \sum_{k=0}^{n+1} \binom{n+1}{k} u^{[(n+1)-k]}v^{(k)}.$$

Für $y^{(n+1)}$ erhalten wir also einen Ausdruck, der zu (1) völlig analog ist (nur ist die Zahl n durch die Zahl $n+1$ ersetzt). Damit ist die Gültigkeit der Formel (1) für alle natürlichen n bewiesen.

Diese Formel heißt *Leibnizsche Formel.* Bei der Herleitung allgemeiner Ausdrücke für n-te Ableitungen ist sie oft von Nutzen.

Eine ähnliche Formel kann auch für die n-te Ableitung eines Produktes mehrerer Faktoren $y = u \cdot v \cdots t$ aufgestellt werden. Sie ist der Entwicklung des Polynoms $(u + v + \cdots + t)^n$ ähnlich.

118. Beispiele.

1. Wir wollen mit Hilfe der Leibnizschen Formel (1) die Ableitung

$$(x^2 \cdot \cos ax)^{(50)}$$

bestimmen. Dazu setzen wir $v = x^2$, $u = \cos ax$. Dann ist

$$u^{(k)} = a^k \cdot \cos\left(ax + k \cdot \frac{\pi}{2}\right), \qquad v' = 2x,\ v'' = 2,\ v''' = v^{(4)} = \cdots = 0.$$

Somit sind in Formel (1) alle Summanden mit Ausnahme der ersten drei gleich 0, und wir erhalten

$$(uv)^{(50)} = x^2 \cdot a^{50} \cdot \cos\left(ax + 50 \cdot \frac{\pi}{2}\right) + \frac{50}{1} \cdot 2x \cdot a^{49} \cdot \cos\left(ax + 49 \cdot \frac{\pi}{2}\right)$$

$$+ \frac{50 \cdot 49}{1 \cdot 2} \cdot 2 \cdot a^{48} \cdot \cos\left(ax + 48 \cdot \frac{\pi}{2}\right)$$

$$= a^{48}[(2450 - a^2x^2)\cos ax - 100ax \cdot \sin ax].$$

2. Wir kommen auf das Beispiel 7 aus Nr. 116 zurück; jetzt können wir den allgemeinen Ausdruck für die n-te Ableitung der Funktion

$$y = e^{ax} \sin bx$$

mit Hilfe der Leibnizschen Formel direkt angeben:

$$y^{(n)} = e^{ax}\left[\sin bx \left(a^n - \frac{n(n-1)}{1 \cdot 2} a^{n-2}b^2 + \cdots\right)\right.$$

$$\left. + \cos bx \left(na^{n-1}b - \frac{n(n-1)(n-2)}{1 \cdot 2 \cdot 3} a^{n-3}b^3 + \cdots\right)\right].$$

3. Wir bestimmen die $(n+1)$-te Ableitung der Funktion $y = \arcsin x$. Zunächst ist

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

so daß nach der Leibnizschen Formel

$$y^{(n+1)} = \left(\frac{1}{\sqrt{1+x}}\,\frac{1}{\sqrt{1-x}}\right)^{(n)}$$

$$= \left(\frac{1}{\sqrt{1+x}}\right)^{(n)} \frac{1}{\sqrt{1-x}} + n\left(\frac{1}{\sqrt{1+x}}\right)^{(n-1)}\left(\frac{1}{\sqrt{1-x}}\right)'$$

$$+ \frac{n(n-1)}{1\cdot 2}\left(\frac{1}{\sqrt{1+x}}\right)^{(n-2)}\left(\frac{1}{\sqrt{1-x}}\right)''$$

$$+ \frac{n(n-1)(n-2)}{1\cdot 2\cdot 3}\left(\frac{1}{\sqrt{1+x}}\right)^{(n-3)}\left(\frac{1}{\sqrt{1-x}}\right)''' + \cdots$$

gilt. Wenden wir nun zur Berechnung der Ableitungen von $\frac{1}{\sqrt{1+x}}$ und $\frac{1}{\sqrt{1-x}}$ die in Nr. 116, Beispiel 2, gewonnenen Formeln an, so erhalten wir schließlich

$$y^{(n+1)} = \frac{1}{2^n\sqrt{1-x^2}}\left\{\frac{(2n-1)!!}{(1+x)^n} - n\,\frac{(2n-3)!!\,1!!}{(1+x)^{n-1}(1-x)} + \frac{n(n-1)}{1\cdot 2}\,\frac{(2n-5)!!\,3!!}{(1+x)^{n-2}(1-x)^2} + \cdots\right\}.$$

4. Man bestimme den Wert aller aufeinanderfolgenden Ableitungen der Funktion $y = \arctan x$ für $x = 0$.

Wegen $y' = \frac{1}{1+x^2}$ gilt $y'\cdot(1+x^2) = 1$. Wir bilden die n-te Ableitung beider Seiten dieser Gleichung (unter Benutzung der Leibnizschen Formel):

$$(1+x^2)\,y^{(n+1)} + 2nx\cdot y^{(n)} + n(n-1)\cdot y^{(n-1)} = 0.$$

Nun setzen wir $x = 0$. Bezeichnen wir den Wert der Ableitung für $x = 0$ mit dem unteren Index 0, so erhalten wir

$$y_0^{(n+1)} = -n(n-1)\cdot y_0^{(n-1)}. \tag{*}$$

Ferner nimmt $y'' = -\frac{2x}{(1+x^2)^2}$ für $x = 0$ den Wert 0 an: $y_0'' = 0$. Aus der Relation (*) geht hervor, daß stets $y_0^{(2m)} = 0$ ist. Für die Ableitungen ungerader Ordnung gilt die Rekursionsformel $y_0^{(2m+1)} = -(2m-1)\,2m y_0^{(2m-1)}$. Beachten wir, daß $y_0' = 1$ ist, so erhalten wir hieraus

$$y_0^{(2m+1)} = (-1)^m\,(2m)!.$$

Dasselbe Resultat kann man auch aus der allgemeinen Formel des Beispiels 8 in Nr. 116 herleiten.

5. Für die Funktion $y = \arcsin x$ bestimme man ebenfalls den Wert aller Ableitungen für $x = 0$.

Hinweis. Die Leibnizsche Formel ist diesmal auf die Beziehung

$$(1-x^2)\,y'' - xy' = 0$$

anzuwenden.

Lösung.

$$y_0^{(2m)} = 0, \quad y_0^{(2m-1)} = 1^2\cdot 3^2\cdots(2m-1)^2 = [(2m-1)!!]^2.$$

Mit Hilfe der allgemeinen Ausdrücke in Beispiel 3 gelangt man nicht so einfach zu diesem Resultat.

6. *Die Legendreschen Polynome.* Zum Schluß wollen wir uns noch mit den wichtigen Legendreschen Polynomen¹) befassen. Sie werden durch

$$X_n(x) = c_n \frac{d^n(x^2-1)^n}{\cdot dx} \qquad (n = 1, 2, 3, \ldots)$$

definiert, wobei die konstanten Koeffizienten c_n je nach dem Zweck verschieden gewählt werden können.

Zunächst überzeugen wir uns davon, daß das Polynom $X_n(x)$ (vom Grad n) genau n verschiedene reelle Nullstellen besitzt, die sämtlich zwischen -1 und 1 liegen.

Der Einfachheit halber setzen wir vorerst $c_n = 1$.

Es ist leicht zu sehen, daß das Polynom

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

und seine $n-1$ aufeinanderfolgenden Ableitungen für $x = \pm 1$ den Wert 0 annehmen. Dann besitzt seine erste Ableitung nach dem Satz von Rolle (Nr. 111) eine Nullstelle zwischen -1 und 1. Nach dem gleichen Satz besitzt die zweite Ableitung zwei Nullstellen zwischen -1 und 1 usw. bis zur $(n-1)$-ten Ableitung, die außer den Nullstellen -1 und 1 noch $n-1$ Nullstellen dazwischen besitzt. Wenden wir auf die $(n-1)$-te Ableitung nochmals den Satz von Rolle an, so erhalten wir das gewünschte Ergebnis.

Wir behalten den Koeffizienten $c_n = 1$ bei und bestimmen nun den Wert des Polynoms $X_n(x)$ für $x = \pm 1$. Wenn wir die Potenz $(x^2-1)^n$ als Produkt aus $(x+1)^n$ und $(x-1)^n$ ansehen, können wir nach der Leibnizschen Formel

$$X_n(x) = (x+1)^n \cdot \frac{d^n(x-1)^n}{dx^n} + \binom{n}{1} \frac{d(x+1)^n}{dx} \cdot \frac{d^{n-1}(x-1)^n}{dx^{n-1}} + \cdots + \frac{d^n(x+1)^n}{dx^n} \cdot (x-1)^n$$

schreiben. Da alle Summanden, vom zweiten angefangen, den Faktor $x-1$ enthalten, für $x = 1$ also 0 werden, ist offenbar $X_n(1) = 2^n n!$.

Analog erhalten wir $X_n(-1) = (-1)^n \cdot 2^n \cdot n!$.

Setzt man in der Definitionsgleichung des Legendreschen Polynoms $X_n(x)$ insbesondere

$$c_n = \frac{1}{2^n \cdot n!},$$

so erhält man ein Polynom, das sehr häufig vorkommt. Wir werden es künftig stets mit $P_n(x)$ bezeichnen. Es ist dadurch charakterisiert, daß es in den Punkten $x = 1$ und $x = -1$ die Werte

$$P_n(1) = 1, \qquad P_n(-1) = (-1)^n$$

annimmt.

Mit Hilfe der Leibnizschen Formel kann man ferner leicht zeigen, daß die Legendreschen Polynome der Beziehung

$$(x^2-1) X_n'' + 2x X_n' - n(n+1) X_n = 0$$

genügen, die in der Theorie dieser Polynome eine wichtige Rolle spielt. Es handelt sich um die sogenannte *Differentialgleichung der Legendreschen Polynome.*

Setzen wir nämlich $y = (x^2-1)^n$, so ist $y' = 2nx(x^2-1)^{n-1}$, also $(x^2-1)\,y' = 2nx \cdot y$. Wir bilden jetzt die $(n+1)$-te Ableitung beider Seiten dei letzten Gleichung. Nach der Leibnizschen Formel ist

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

¹) Adrien Marie Legendre, 1752—1833, französischer Mathematiker.

Hieraus folgt

$$(x^2 - 1)\, y^{(n+2)} + 2xy^{(n+1)} - n(n+1)\, y^{(n)} = 0,$$

und nach Multiplikation mit c_n ergibt sich die zu beweisende Relation.

119. Differentiale höherer Ordnung. Wir wollen uns jetzt den Differentialen höherer Ordnung zuwenden. Auch sie werden induktiv definiert.

Unter dem *Differential zweiter Ordnung* oder dem *zweiten Differential* einer Funktion $y = f(x)$ in einem bestimmten Punkt versteht man das Differential des (ersten) Differentials der Funktion in diesem Punkt, in Zeichen

$$\mathrm{d}^2 y = \mathrm{d}(\mathrm{d}y).$$

Das Differential dritter Ordnung oder das dritte Differential ist das Differential des zweiten Differentials:

$$\mathrm{d}^3 y = \mathrm{d}(\mathrm{d}^2 y).$$

Allgemein versteht man unter dem Differential n-ter Ordnung oder dem n-ten Differential der Funktion $y = f(x)$ das Differential ihres $(n-1)$-ten Differentials,

$$\mathrm{d}^n y = \mathrm{d}(\mathrm{d}^{n-1} y).$$

Man schreibt diese Differentiale für den speziellen Punkt $x = x_0$, in welchem die Differentiale gebildet werden, auch folgendermaßen:

$$\mathrm{d}^2 f(x_0),\quad \mathrm{d}^3 f(x_0), \ldots, \mathrm{d}^n f(x_0), \ldots$$

Bei der Berechnung von Differentialen höherer Ordnung ist es sehr wichtig, zu beachten, daß $\mathrm{d}x$ eine willkürliche und von x unabhängige Zahl ist, die bei der Bildung des Differentials nach x als konstanter Faktor angesehen werden muß. Dann erhalten wir (wobei wir stets die Existenz der entsprechenden Ableitungen voraussetzen)

$$\mathrm{d}^2 y = \mathrm{d}(\mathrm{d}y) = \mathrm{d}(y' \cdot \mathrm{d}x) + \mathrm{d}y' \cdot \mathrm{d}x = (y'' \cdot \mathrm{d}x)\, \mathrm{d}x = y'' \cdot \mathrm{d}x^2,$$

$$\mathrm{d}^3 y = \mathrm{d}(\mathrm{d}^2 y) = \mathrm{d}(y'' \cdot \mathrm{d}x^2) = \mathrm{d}y'' \cdot \mathrm{d}x^2 = (y''' \cdot \mathrm{d}x)\, \mathrm{d}x^2 = y''' \cdot \mathrm{d}x^3$$

usw.[1]) Das naheliegende allgemeine Gesetz

$$\mathrm{d}^n y = y^{(n)}\, \mathrm{d}x^n \tag{2}$$

läßt sich wiederum durch Induktion beweisen. Daraus folgt

$$y^{(n)} = \frac{\mathrm{d}^n y}{\mathrm{d}x^n},$$

so daß jetzt auch dieses Symbol als Bruch aufgefaßt werden kann.

Unter Benutzung von Gleichung (2) läßt sich jetzt die Leibnizsche Formel leicht auf Differentiale umformen. Es genügt, beide Seiten dieser Formel mit $\mathrm{d}x^n$ zu multiplizieren, um

$$\mathrm{d}^n(uv) = \sum_{i=0}^{n} \binom{n}{i}\, \mathrm{d}^{n-i} u \cdot \mathrm{d}^i v \qquad (\mathrm{d}^0 u = u,\ \mathrm{d}^0 v = v)$$

zu erhalten. Leibniz selbst hatte die nach ihm benannte Formel für Differentiale aufgestellt.

[1]) Unter $\mathrm{d}x^2, \mathrm{d}x^3, \ldots$ ist stets die Potenz des Differentials zu verstehen, also $(\mathrm{d}x)^2, (\mathrm{d}x)^3, \ldots$ Das Differential der Potenzfunktion bezeichnet man dagegen mit $\mathrm{d}(x^2), \mathrm{d}(x^3), \ldots$

120. Nichtinvarianz der Differentiale höherer Ordnung. Bekanntlich ist das (erste) Differential einer Funktion gegenüber einer Variablentransformation invariant. Dadurch erhebt sich naturgemäß die Frage, ob die Differentiale höherer Ordnung dieselbe Eigenschaft besitzen. Wir werden am Beispiel des zweiten Differentials zeigen, daß dies nicht der Fall ist.

Es sei also $y = f(x)$ und $x = \varphi(t)$, so daß y als mittelbare Funktion von t betrachtet werden kann, $y = f(\varphi(t))$. Das (erste) Differential bezüglich t kann in der Form $\mathrm{d}y = y_x\,\mathrm{d}x$ geschrieben werden, wobei $\mathrm{d}x = x_t\,\mathrm{d}t$ eine Funktion von t ist. Wir berechnen das zweite Differential bezüglich t und finden

$$\mathrm{d}^2y = \mathrm{d}(y_x\,\mathrm{d}x) = \mathrm{d}y_x\,\mathrm{d}x + y_x\,\mathrm{d}(\mathrm{d}x).$$

Das Differential $\mathrm{d}y_x$ kann unter erneuter Benutzung der Invarianz des (ersten) Differentials in der Form $\mathrm{d}y_x = y_{xx}\,\mathrm{d}x$ geschrieben werden, so daß sich schließlich

$$\mathrm{d}^2y = y_{xx}\cdot \mathrm{d}x^2 + y_x\,\mathrm{d}^2x \tag{3}$$

ergibt, während für die unabhängige Veränderliche x das zweite Differential die Gestalt $\mathrm{d}^2y = y_{xx}\cdot\mathrm{d}x^2$ hatte. Der Ausdruck (3) für d^2y ist natürlich der allgemeinere; ist insbesondere x die unabhängige Veränderliche, so ist $\mathrm{d}^2x = 0$, und es bleibt nur das erste Glied übrig.

Betrachten wir ein Beispiel. Es sei $y = x^2$, also, wenn x vorerst die unabhängige Veränderliche ist,

$$\mathrm{d}y = 2x\,\mathrm{d}x, \qquad \mathrm{d}^2y = 2\mathrm{d}x^2.$$

Wir setzen jetzt $x = t^2$; dann ist $y = t^4$ und

$$\mathrm{d}y = 4t^3\,\mathrm{d}t, \qquad \mathrm{d}^2y = 12t^2\,\mathrm{d}t^2.$$

Der neue Ausdruck für $\mathrm{d}y$ kann aus dem alten gewonnen werden, wenn man dort $x = t^2$, $\mathrm{d}x = 2t\,\mathrm{d}t$ setzt. Anders verhält es sich mit d^2y. Würden wir dort ebenso substituieren, so erhielten wir $8t^2\,\mathrm{d}t^2$ anstatt $12t^2\,\mathrm{d}t^2$.

Wenn wir aber die Gleichung $\mathrm{d}y = 2x\,\mathrm{d}x$ nach t differenzieren, wobei wir x als Funktion von t betrachten, so kommen wir, entsprechend (3), auf die Formel

$$\mathrm{d}^2y = 2\mathrm{d}x^2 + 2x\,\mathrm{d}^2x.$$

Setzen wir hier $x = t^2$, $\mathrm{d}x = 2t\,\mathrm{d}t$, $\mathrm{d}^2x = 2\mathrm{d}t^2$, so erhalten wir das richtige Ergebnis, nämlich $12t^2\,\mathrm{d}t^2$.

Ist also x nicht die unabhängige Veränderliche, dann hat man das Differential zweiter Ordnung d^2y durch das Differential von x gemäß der zweigliedrigen Formel (3) auszudrücken. Für Differentiale dritter und höherer Ordnung steigt beim Übergang zu einer neuen unabhängigen Veränderlichen die Zahl der Zusatzglieder weiter an. Wenn man also die höheren Ableitungen $y_{xx}, y_{xxx}, \ldots$ durch Differentiale ausdrücken will,

$$y_{xx} = y_{x^2} = \frac{\mathrm{d}^2y}{\mathrm{d}x^2}, \quad y_{xxx} = y_{x^3} = \frac{\mathrm{d}^3y}{\mathrm{d}x^3}, \ldots, \tag{4}$$

so darf man dementsprechend die Differentiale nicht nach einer beliebigen Veränderlichen bilden, sondern nur nach der Veränderlichen x.

121. Differentiation nach dem Parameter. Man kann übrigens die Ausdrücke für die Ableitungen nach x auch mit Hilfe von Differentialen schreiben, die nach einer beliebigen Veränderlichen t gebildet werden, jedoch werden diese Ausdrücke dann ziemlich kompliziert. Sehen wir alle nachstehend aufgeführten Differentiale als nach t gebildet an, so haben wir nacheinander

$$y_x = \frac{dy}{dx}, \quad y_{xx} = \left(\frac{dy}{dx}\right)_x = \frac{d\left(\frac{dy}{dx}\right)}{dx} = \frac{\frac{dx \cdot d^2y - d^2x \cdot dy}{dx^2}}{dx},$$

also

$$y_{xx} = y_{x^2} = \frac{dx \cdot d^2y - d^2x \cdot dy}{dx^3}; \tag{5}$$

$$y_{xxx} = y_{x^3} = \left(\frac{dx \cdot d^2y - d^2x \cdot dy}{dx^3}\right)_x = \frac{d\left(\frac{dx \cdot d^2y - d^2x \cdot dy}{dx^3}\right)}{dx}$$

$$= \frac{\frac{dx^3(dx \cdot d^3y - d^3x \cdot dy) - 3dx^2\, d^2x(dx \cdot d^2y - d^2x \cdot dy)}{dx^6}}{dx},$$

also schließlich

$$y_{xxx} = y_{x^3} = \frac{dx(dx \cdot d^3y - d^3x \cdot dy) - 3d^2x(dx \cdot d^2y - d^2x \cdot dy)}{dx^5}, \tag{6}$$

usw. Die Formeln (5) und (6) sowie die nach diesem Schema gebildeten weiteren Formeln sind die allgemeinsten. Nehmen wir in ihnen x als unabhängige Veränderliche an, so werden d^2x, d^3x und die folgenden Differentiale gleich 0, und wir kommen auf Formel (4) zurück.

Diese Formeln für die Ableitungen von y nach x liefern die sogenannte *Differentiation nach dem Parameter.*

Sind x und y als Funktionen eines Parameters t gegeben,

$$x = \varphi(t), \quad y = \psi(t),$$

so wird, wie wir in Nr. 106 gesehen haben, unter gewissen Bedingungen dadurch auch y als Funktion von x definiert: $y = f(x)$. Falls die Ableitungen von x und y nach t existieren, so existieren auch die entsprechenden Ableitungen von y nach x und lassen sich durch die oben angegebenen Formeln ausdrücken.

Manchmal ist es zweckmäßiger, die Ableitungen von y nach x durch die Ableitungen (und nicht durch die Differentiale) von x und y nach t auszudrücken. Man erhält sie leicht aus den Differentialausdrücken, indem man Zähler und Nenner durch dt, dt^3, $dt^5, \ldots$ dividiert. So kommt man zu den Formeln

$$y_x = \frac{\frac{dy}{dt}}{\frac{dx}{dt}} = \frac{y_t}{x_t},$$

$$y_{xx} = y_{x^2} = \frac{\frac{dx}{dt} \cdot \frac{d^2y}{dt^2} - \frac{d^2x}{dt^2} \cdot \frac{dy}{dt}}{\left(\frac{dx}{dt}\right)^3} = \frac{x_t y_{t^2} - x_{t^2} y_t}{(x_t)^3};$$

und analog

$$y_{xxx} = y_{x^3} = \frac{x_t(x_t y_{t^3} - x_{t^3} y_t) - 3x_{t^2}(x_t y_{t^2} - x_{t^2} y_t)}{(x_t)^5}$$

usw.

122. Endliche Differenzen. Die Funktion $f(x)$ sei in einem Intervall $\mathcal{X}$ definiert; alle Werte von x, von denen hier die Rede ist, mögen diesem Intervall angehören. Wir legen einen bestimmten Zuwachs Δx der Veränderlichen fest (dabei setzen wir $\Delta x > 0$ voraus, obgleich wir ebensogut $\Delta x < 0$ annehmen könnten) und setzen

$$\Delta f(x) = f(x + \Delta x) - f(x);$$

diesen Ausdruck nennen wir die *erste Differenz* der Funktion $f(x)$. Unter der *zweiten Differenz* verstehen wir die erste Differenz dieser ersten Differenz,

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

Höhere Differenzen werden induktiv definiert:

$$\Delta^n f(x) = \Delta(\Delta^{n-1} f(x)).$$

Übrigens kann die n-te Differenz durch eine geschlossene Formel angegeben werden, nämlich durch

$$\begin{aligned}\Delta^n f(x) &= \sum_{i=0}^{n} (-1)^i \binom{n}{i} f(x + (n - i)\,\Delta x) \\ &= f(x + n\Delta x) - \frac{n}{1} \cdot f(x + (n-1)\,\Delta x) \\ &\quad + \frac{n(n-1)}{1 \cdot 2} f(x + (n-2)\,\Delta x) - \cdots + (-1)^n f(x),\end{aligned}$$

die diese Differenz unmittelbar durch die Werte der Funktion $f(x)$ in den äquidistanten Punkten

$$x,\ x + \Delta x,\ x + 2\Delta x,\ \ldots,\ x + n\Delta x$$

ausdrückt. Diese Formel läßt sich durch Induktion leicht beweisen, was dem Leser überlassen sei. Wir stellen jetzt diese endlichen Differenzen den Ableitungen und Differentialen gegenüber.

Zu diesem Zweck setzen wir voraus, die Funktion $f(x)$ besitze im abgeschlossenen Intervall $[x_0, x_0 + n\Delta x]$ die $n - 1$ stetigen Ableitungen

$$f'(x), f''(x), \ldots, f^{(n-1)}(x)$$

und wenigstens im offenen Intervall $(x_0, x_0 + n\Delta x)$ die endliche Ableitung $f^{(n)}(x)$. Dann gilt

$$\Delta^n f(x_0) = f^{(n)}(\xi_n) \cdot \Delta x^n \qquad (x_0 < \xi_n < x_0 + n\Delta x). \tag{7}$$

Für $n = 1$ reduziert sich (7) auf den Mittelwertsatz, den einfachsten Spezialfall. Um diese Aussage durch Induktion zu beweisen, setzen wir die Richtigkeit der Formel (7) für $\Delta^{n-1} f(x_0)$ voraus (natürlich bei entsprechender Änderung der Voraussetzungen) und beweisen, daß (7) unter den ursprünglichen Voraussetzungen gilt. Aus diesen Voraussetzungen folgt, daß für die Funktion $\Delta f(x) = f(x + \Delta x) - f(x)$ im Intervall $[x_0, x_0 + (n - 1)\,\Delta x]$ auf jeden Fall die geänderte Formel (7) angewandt werden kann. Daher können wir schreiben:

$$\Delta^{n-1}[\Delta f(x_0)] = \Delta^n f(x_0) = [f^{(n-1)}(\xi_{n-1} + \Delta x) - f^{(n-1)}(\xi_{n-1})] \cdot \Delta x^{n-1}, \tag{8}$$

wobei $x_0 < \xi_{n-1} < x_0 + (n - 1)\,\Delta x$ ist. Wenden wir auf die rechte Seite dieser Gleichung den Mittelwertsatz an,[1]) so erhalten wir unmittelbar Formel (7), wobei

$$x_0 < \xi_{n-1} < \xi_n < \xi_{n-1} + \Delta x < x_0 + n\Delta x$$

gilt.

Wenn $f^{(n)}(x)$ auch im Punkt x_0 existiert und dort stetig ist, so folgt aus (7) für $\Delta x \to 0$ (d. h. $\xi_n \to x_0$)

$$f^{(n)}(x_0) = \lim_{\Delta x \to 0} \frac{\Delta^n f(x_0)}{\Delta x^n}. \tag{9}$$

Wir bemerken, daß diese interessante Formel, die die n-te Ableitung mit Hilfe eines einzigen Grenzübergangs zu erhalten gestattet, unter der einzigen Voraussetzung gilt, daß diese Ableitung im Punkt x_0 existiert. Das bedeutet, daß in einer Umgebung des Punktes x_0 dann auch

[1]) Das dürfen wir, da die Funktion $f^{(n-1)}(x)$ im Intervall $[\xi_{n-1}, \xi_{n-1} + \Delta x]$ stetig ist und im Innern des Intervalls eine endliche Ableitung $f^{(n)}(x)$ besitzt.

die Ableitungen

$$f'(x), f''(x), \ldots, f^{(n-1)}(x)$$

existieren, also für hinreichend kleines Δx die Formel (8) angewandt werden kann.

Da die Ableitung $f^{(n)}(x_0)$ existiert und Formel (2) aus Nr. 96 anwendbar ist, können wir

$$f^{(n-1)}(\xi_{n-1}) - f^{(n-1)}(x_0) = f^{(n)}(x_0) \cdot (\xi_{n-1} - x_0) + \alpha \cdot (\xi_{n-1} - x_0)$$

und

$$f^{(n-1)}(\xi_{n-1} + \Delta x) - f^{(n-1)}(x_0) = f^{(n)}(x_0) \cdot (\xi_{n-1} + \Delta x - x_0) + \beta \cdot (\xi_{n-1} + \Delta x - x_0)$$

schreiben, wobei α und β von Δx abhängen und mit Δx gegen 0 streben. Hieraus und aus (8) folgt[1])

$$\Delta^n f(x_0) = [f^{(n)}(x_0) + \gamma] \cdot \Delta x^n,$$

wobei γ eine neue unendlich kleine Größe ist. Schließlich kommen wir zur Formel (9), indem wir diese Gleichung durch Δx^n dividieren und den Grenzübergang für $\Delta x \to 0$ durchführen.

Wir betonen noch einmal, daß Formel (9) nur unter der Voraussetzung gültig ist, daß die Ableitung $f^{(n)}(x_0)$ existiert. Der Grenzwert auf der rechten Seite kann auch dann existieren, wenn diese Ableitung nicht existiert.[2])

Wir betrachten als Beispiel die Funktion

$$f(0) = 0, \qquad f(x) = x^3 \cdot \sin\frac{1}{x} \qquad (x \neq 0)$$

an der Stelle $x_0 = 0$. Für sie existiert die erste Ableitung

$$f'(0) = 0, \qquad f'(x) = 3x^2 \cdot \sin\frac{1}{x} - x \cdot \cos\frac{1}{x} \qquad (x \neq 0),$$

aber im Punkt 0 keine zweite Ableitung, weil der Quotient

$$\frac{f'(0 + \Delta x) - f'(0)}{\Delta x} = \frac{3\Delta x^2 \cdot \sin\frac{1}{\Delta x} - \Delta x \cdot \cos\frac{1}{\Delta x}}{\Delta x} = 3\Delta x \cdot \sin\frac{1}{\Delta x} - \cos\frac{1}{\Delta x}$$

für $\Delta x \to 0$ keinen Grenzwert besitzt. Es gilt jedoch

$$\frac{\Delta^2 f(0)}{\Delta x^2} = \frac{f(0 + 2\Delta x) - 2f(0 + \Delta x) + f(0)}{\Delta x^2} = \frac{8\Delta x^3 \cdot \sin\frac{1}{2\Delta x} - 2\Delta x^3 \cdot \sin\frac{1}{\Delta x}}{\Delta x^2}$$

$$= 8\Delta x \cdot \sin\frac{1}{2\Delta x} - 2\Delta x \cdot \sin\frac{1}{\Delta x} \to 0.$$

§ 5. Die Taylorsche Formel

123. Die Taylorsche Formel für ein Polynom. Ist $p(x)$ ein Polynom n-ten Grades,

$$p(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots + a_n x^n, \tag{1}$$

so finden wir durch n-maliges Differenzieren

$$\begin{aligned}
p'(x) &= a_1 + 2 \cdot a_2 x + 3 \cdot a_3 x^2 + \cdots + n \cdot a_n x^{n-1},\\
p''(x) &= 1 \cdot 2 \cdot a_2 + 2 \cdot 3 a_3 x + \cdots + (n-1)\, n a_n x^{n-2},\\
p'''(x) &= 1 \cdot 2 \cdot 3 \cdot a_3 + \cdots + (n-2)(n-1)\, n a_n x^{n-3},\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
p^{(n)}(x) &= 1 \cdot 2 \cdot 3 \cdots n \cdot a_n.
\end{aligned}$$

[1]) Es gilt nämlich $0 < \xi_{n-1} - x_0 < (n-1)\,\Delta x$ (für $\Delta x > 0$).

[2]) Die Formel (9) gibt also keineswegs eine neue, der früheren Definition äquivalente Definition des Begriffs der n-ten Ableitung an!

Setzen wir in allen diesen Formeln $x = 0$, so erhalten wir Ausdrücke für die Koeffizienten des Polynoms, und zwar durch die Werte des Polynoms und seiner Ableitungen für $x = 0$,

$$a_0 = p(0), \quad a_1 = \frac{p'(0)}{1!}, \quad a_2 = \frac{p''(0)}{2!}, \ldots, a_n = \frac{p^{(n)}(0)}{n!}.$$

Wir setzen diese Werte für die Koeffizienten in (1) ein:

$$p(x) = p(0) + \frac{p'(0)}{1!}x + \frac{p''(0)}{2!}x^2 + \cdots + \frac{p^{(n)}(0)}{n!}x^n. \tag{2}$$

Diese Formel unterscheidet sich von (1) dadurch, daß die Koeffizienten durch $p(0)$, $p'(0), \ldots$ ausgedrückt sind.

Statt ein Polynom nach Potenzen von x zu entwickeln, kann man es auch nach Potenzen von $x - x_0$ entwickeln, wobei x_0 irgendein fester x-Wert ist:

$$p(x) = A_0 + A_1(x - x_0) + A_2(x - x_0)^2 + \cdots + A_n(x - x_0)^n. \tag{3}$$

Setzen wir $x - x_0 = \xi$, $p(x) = p_0(x_0 + \xi) = P(\xi)$, so erhalten wir für die Koeffizienten des Polynoms

$$P(\xi) = A_0 + A_1\xi + A_2\xi^2 + A_3\xi^3 + \cdots + A_n\xi^n$$

nach dem bereits Bewiesenen die Ausdrücke

$$A_0 = P(0), \quad A_1 = \frac{P'(0)}{1!}, \quad A_2 = \frac{P''(0)}{2!}, \quad \ldots, \quad A_n = \frac{P^{(n)}(0)}{n!}.$$

Nun ist aber

$$P(\xi) = p(x_0 + \xi), \quad P'(\xi) = p'(x_0 + \xi), \quad P''(\xi) = p''(x_0 + \xi), \ldots,$$

also $P(0) = p(x_0)$, $P'(0) = p'(x_0)$, $P''(0) = p''(x_0), \ldots$ und

$$A_0 = p(x_0), A_1 = \frac{p'(x_0)}{1!}, \quad A_2 = \frac{p''(x_0)}{2!}, \ldots, \quad A_n = \frac{p^{(n)}(x_0)}{n!}, \tag{4}$$

d. h., die Koeffizienten der Entwicklung (3) sind durch die Werte des Polynoms und seiner Ableitungen für $x = x_0$ ausgedrückt.

Die Ausdrücke (4) setzen wir in (3) ein:

$$p(x) = p(x_0) + \frac{p'(x_0)}{1!}(x - x_0) + \frac{p''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{p^{(n)}(x_0)}{n!}(x - x_0)^n. \tag{5}$$

Die Formel (5) wird ebenso wie ihr Spezialfall (2) (für $x_0 = 0$) *Taylorsche Formel*[1]) genannt. Bekanntlich besitzt sie wichtige Anwendungen in der Algebra.

[1]) Nach Brook Taylor, 1685–1731, englischer Mathematiker. Übrigens wird die Formel (2) oft auch Maclaurinsche Formel genannt (nach Colin Maclaurin, 1698–1746, schottischer Mathematiker).

Bemerkung. Wir weisen noch auf folgendes hin (was uns im weiteren von Nutzen sein wird). Ist ein Polynom in der Form

$$p(x) = c_0 + \frac{c_1}{1!}(x - x_0) + \frac{c_2}{2!}(x - x_0)^2 + \cdots + \frac{c_n}{n!}(x - x_0)^n$$

vorgegeben, so gilt offenbar notwendigerweise

$$p(x_0) = c_0, \quad p'(x_0) = c_1, \quad p''(x_0) = c_2, \ldots, p^{(n)}(x_0) = c_n.$$

124. Entwicklung einer beliebigen Funktion. Das Restglied in der Peanoschen Form. Wir wenden uns jetzt der Untersuchung einer beliebigen Funktion $f(x)$ zu, die im allgemeinen natürlich kein Polynom ist. Dabei setzen wir voraus, diese Funktion besitze in einem bestimmten Punkt x_0 Ableitungen aller Ordnungen bis zur n-ten einschließlich. Das bedeutet, genauer gesagt, daß die Funktion in einem den Punkt x_0 enthaltenden Intervall $[a, b]$ definiert ist und dort Ableitungen aller Ordnungen bis zur $(n - 1)$-ten einschließlich,

$$f'(x), f''(x), f'''(x), \ldots, f^{(n-1)}(x),$$

besitzt und daß außerdem die n-te Ableitung $f^{(n)}(x)$ im Punkt x_0 existiert.[1]) Dann kann auch für die Funktion $f(x)$ nach dem Vorbild von (5) das Polynom

$$p(x) = f(x_0) + \frac{f'(x_0)}{1!}(x - x_0)$$
$$+ \frac{f''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n \tag{6}$$

gebildet werden.

Nach der Bemerkung am Schluß von Nr. 123 haben dieses Polynom und seine Ableitungen (bis zur n-ten einschließlich) im Punkt x_0 dieselben Werte wie die Funktion $f(x)$ und ihre Ableitungen.

Ist aber die Funktion $f(x)$ kein Polynom n-ten Grades, so gilt die Gleichung $f(x) = p(x)$ nicht mehr. Das Polynom $p(x)$ ist, wie sich zeigen wird, nur eine Näherungsformel für die Funktion $f(x)$. Daher ist die Untersuchung der Differenz

$$r(x) = f(x) - p(x) \tag{7}$$

von besonderem Interesse.

Zunächst stellen wir fest, daß *diese Differenz für* $x \to x_0$ (*im Vergleich zu* $x - x_0$) *von höherer als* n*-ter Ordnung klein wird:*

$$r(x) = o\big((x - x_0)^n\big). \tag{8}$$

Auf Grund der Eigenschaft von $p(x)$ gilt für $r(x)$ offenbar

$$r(x_0) = r'(x_0) = r''(x_0) = \cdots = r^{(n)}(x_0) = 0. \tag{9}$$

Nun können wir folgende allgemeinere Aussage beweisen: *Ist für irgendeine Funktion* $r(x)$, *die im Punkt* x_0 *Ableitungen bis zur* n*-ten Ordnung einschließlich besitzt, die Bedingung* (9) *erfüllt, so gilt für sie die Beziehung* (8).

[1]) Ist x_0 einer der Intervallendpunkte, so ist unter der Ableitung in diesem Punkt die entsprechende *einseitige* Ableitung zu verstehen.

Den Beweis führen wir durch Induktion. Für $n = 1$ lautet die Behauptung folgendermaßen: Genügt eine Funktion $r(x)$, die im Punkt x_0 eine Ableitung (erster Ordnung) besitzt, der Bedingung

$$r(x_0) = r'(x_0) = 0,$$

so gilt

$$r(x) = o(x - x_0).$$

Das läßt sich unmittelbar verifizieren:

$$\lim_{x \to x_0} \frac{r(x)}{x - x_0} = \lim_{x \to x_0} \frac{r(x) - r(x_0)}{x - x_0} = r'(x_0) = 0.$$

Wir setzen jetzt voraus, die oben formulierte Behauptung sei für beliebiges $n \geqq 1$ richtig, und beweisen, daß sie gültig bleibt, wenn wir n durch $n + 1$ ersetzen. Das heißt aber: Wenn für irgendeine Funktion $r(x)$, die im Punkt x_0 Ableitungen bis zur $(n + 1)$-ten Ordnung einschließlich besitzt, die Bedingungen

$$r(x_0) = r'(x_0) = r''(x_0) = \cdots = r^{(n)}(x_0) = r^{(n+1)}(x_0) = 0 \tag{9*}$$

erfüllt sind, so gilt

$$r(x) = o\big((x - x_0)^{n+1}\big). \tag{8*}$$

Den Bedingungen (9*) können wir entnehmen, daß die Funktion $r'(x)$ einer Bedingung vom Typ (9) genügt, so daß nach Induktionsvoraussetzung

$$r'(x) = o\big((x - x_0)^n\big)$$

gilt. Nach dem Mittelwertsatz (Nr. 112) ist aber

$$r(x) = r(x) - r(x_0) = r'(c)\,(x - x_0),$$

wobei c zwischen x und x_0 liegt. Da $|c - x_0| < |x - x_0|$ ist, gilt

$$r'(c) = o\big((c - x_0)^n\big) = o\big((x - x_0)^n\big),$$

so daß wir schließlich zu (8*) gekommen sind, was zu beweisen war. Somit gilt unsere Behauptung für jedes natürliche n, und für die Differenz (7) ist tatsächlich die Beziehung (8) erfüllt. Beachten wir (6), so erhalten wir die Formel

$$\begin{aligned} f(x) = f(x_0) &+ \frac{f'(x_0)}{1!}(x - x_0) \\ &+ \frac{f''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + o\big((x - x_0)^n\big), \end{aligned} \tag{10}$$

die sich von (5) durch das Restglied (8) unterscheidet. In dieser Form wurde das Restglied von PEANO[1]) angegeben. Formel (10) nennt man daher auch *Taylorsche Formel mit Peanoschem Restglied.*

Die soeben bewiesene Formel ist eine natürliche Verallgemeinerung der Formel (3) aus Nr. 96, die man folgendermaßen schreiben kann:

$$f(x) = f(x_0) + f'(x_0)\,(x - x_0) + o(x - x_0).$$

---

[1]) GUISEPPE PEANO, 1858—1932, italienischer Mathematiker.

Sie entspricht dem Fall $n = 1$. Dort war die Funktion $f(x)$ bis auf eine unendlich kleine Größe von höherer als erster Ordnung als lineare Funktion dargestellt worden, hier dagegen stellen wir sie als Polynom n-ten Grades dar, und zwar bis auf eine unendlich kleine Größe von höherer als n-ter Ordnung genau.

Es läßt sich leicht zeigen, daß diese Darstellung der Funktion $f(x)$ eindeutig ist; d. h., wenn in der Umgebung von x_0 gleichzeitig

$$f(x) = A_0 + A_1(x - x_0) + A_2(x - x_0)^2 + \cdots + A_n(x - x_0)^n + o\big((x - x_0)^n\big)$$

und

$$f(x) = B_0 + B_1(x - x_0) + B_2(x - x_0)^2 + \cdots + B_n(x - x_0)^n + o\big((x - x_0)^n\big)$$

gilt, dann ist notwendigerweise

$$A_0 = B_0, A_1 = B_1, \ldots, A_n = B_n.$$

Aus der Identität

$$\begin{aligned} &A_0 + A_1(x - x_0) + A_2(x - x_0)^2 + \cdots + A_n(x - x_0)^n + o\big((x - x_0)^n\big) \\ &= B_0 + B_1(x - x_0) + B_2(x - x_0)^2 + \cdots + B_n(x - x_0)^n + o\big((x - x_0)^n\big) \end{aligned}$$

ergibt sich nämlich für $x \to x_0$ unmittelbar $A_0 = B_0$. Wenn wir diese Glieder weglassen und durch $x - x_0$ dividieren ($x \neq x_0$), so erhalten wir

$$\begin{aligned} &A_1 + A_2(x - x_0) + \cdots + A_n(x - x_0)^{n-1} + o\big((x - x_0)^{n-1}\big) \\ &= B_1 + B_2(x - x_0) + \cdots + B_n(x - x_0)^{n-1} + o\big((x - x_0)^{n-1}\big); \end{aligned}$$

hieraus schließen wir analog (für $x \to x_0$) $A_1 = B_1$ usw.

Manchmal ist es zweckmäßig, Formel (10) in anderer Gestalt zu verwenden. Das Restglied läßt sich auch in der Form

$$r(x) = \frac{\alpha}{n!}(x - x_0)^n$$

darstellen, wobei α von x abhängt und mit $x - x_0$ gegen 0 strebt. Setzen wir diesen Ausdruck ein, so erhalten wir

$$\begin{aligned} f(x) = f(x_0) &+ \frac{f'(x_0)}{1!}(x - x_0) \\ &+ \frac{f''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0) + \alpha}{n!}(x - x_0)^n. \end{aligned} \tag{10a}$$

Bringen wir in Formel (10) $f(x_0)$ nach links und setzen $x - x_0 = \Delta x$, so nimmt (10) die Gestalt

$$\Delta f(x_0) = f'(x_0)\,\Delta x + \frac{1}{2!}f''(x_0)\,\Delta x^2 + \cdots + \frac{1}{n!}f^{(n)}(x_0)\,\Delta x^n + o(\Delta x^n) \tag{10b}$$

an. In dieser Gestalt ähnelt sie der Formel (3) aus Nr. 96;

$$\Delta f(x_0) = f'(x_0) \cdot \Delta x + o(\Delta x).$$

Diese sondert nur einen einzigen Hauptteil aus dem unendlich kleinen Zuwachs $\Delta f(x_0)$ der Funktion aus, wobei wie immer angenommen wird, daß Δx die unendlich kleine Grundgröße ist, während in (10b) die Glieder aller Ordnungen bis zur n-ten einschließlich aufgeschrieben sind, die sämtlich einfachste unendlich kleine Größen im Sinne von Nr. 63 sind. Bis auf das Restglied genau ist damit der Zuwachs der Funktion nach Potenzen des Zuwachses der unabhängigen Veränderlichen entwickelt. Betrachten wir schließlich noch

$$f'(x_0)\cdot\Delta x = \mathrm{d}f(x_0),\quad f''(x_0)\cdot\Delta x^2 = \mathrm{d}^2 f(x_0),\ldots, f^{(n)}(x_0)\cdot\Delta x^n = \mathrm{d}^n f(x_0),$$

so können wir (10b) in der Form

$$\Delta f(x_0) = \mathrm{d}f(x_0) + \frac{1}{2!}\mathrm{d}^2 f(x_0) + \cdots + \frac{1}{n!}\mathrm{d}^n f(x_0) + o(\Delta x^n)$$

schreiben. Hieraus ersieht man, daß (für $\Delta x \to 0$) die sukzessiven Differentiale (bis auf die Fakultäten im Nenner) mit den einfachsten unendlich kleinen Gliedern der entsprechenden Ordnung in der Entwicklung des unendlich kleinen Zuwachses der Funktion übereinstimmen.

125. Beispiele. Am einfachsten sieht die Taylorsche Formel im Fall $x_0 = 0$ aus[1]):

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \cdots + \frac{f^{(n)}(0)}{n!}x^n + o(x^n). \tag{11}$$

Auf diesen Spezialfall läßt sich das Problem immer zurückführen, wenn man $x - x_0$ als neue unabhängige Veränderliche einführt.

Als Beispiele betrachten wir die Entwicklung einiger elementarer Funktionen nach dieser Formel.

1. Es sei $f(x) = \mathrm{e}^x$. Dann ist $f^{(k)}(x) = \mathrm{e}^x$ für jedes $k = 1, 2, 3, \ldots$ Da in diesem Fall $f(0) = 1$ und $f^{(k)}(0) = 1$ ist, gilt nach Formel (11)

$$\mathrm{e}^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + o(x^n).$$

2. Ist $f(x) = \sin x$, so gilt $f^{(k)}(x) = \sin\left(x + k\frac{\pi}{2}\right)$, also

$$f(0) = 0, \qquad f^{(2m)}(0) = \sin m\pi = 0,$$

$$f^{(2m-1)}(0) = \sin\left(m\pi - \frac{\pi}{2}\right) = (-1)^{m-1} \qquad (m = 1, 2, 3, \ldots).$$

Wenn wir in Formel (11) nun $n = 2m$ setzen, so folgt

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots + (-1)^{m-1}\frac{x^{2m-1}}{(2m-1)!} + o(x^{2m}).$$

3. Analog erhalten wir für $f(x) = \cos x$

$$f^{(k)}(x) = \cos\left(x + k\frac{\pi}{2}\right), \qquad f(0) = 1, \qquad f^{(2m)}(0) = (-1)^m,$$

$$f^{(2m-1)}(0) = 0 \qquad (m = 1, 2, 3, \ldots).$$

Somit ist (mit $n = 2m + 1$)

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots + (-1)^m\frac{x^{2m}}{(2m)!} + o(x^{2m+1}).$$

---

[1]) Auch diese Formel wird nach Maclaurin benannt (vgl. die Fußnote auf S. 232).

4. Wir betrachten jetzt die Potenzfunktion x^m (für $x > 0$), wobei m weder eine natürliche Zahl noch 0 sein soll. In diesem Fall werden für $x \to 0$ entweder die Funktion selbst (wenn $m < 0$) oder ihre Ableitungen (von einer bestimmten Ordnung an, etwa für $n > m$) gleich ∞. Folglich darf man hier schon nicht mehr $x_0 = 0$ wählen.

Wir wollen $x_0 = 1$ annehmen, d. h. x^m nach Potenzen von $x - 1$ entwickeln. Übrigens können wir, wie bereits erwähnt, $x - 1$ als neue Veränderliche einführen. Wir wollen sie wieder mit x bezeichnen, also die Funktion $(1 + x)^m$ nach Potenzen von x entwickeln, wobei jetzt $1 + x > 0$ sei. Wir wissen aus Nr. 116, Beispiel 2, daß

$$f^{(k)}(x) = m \cdot (m-1) \cdots (m-k+1)\,(1+x)^{m-k},$$

also

$$f(0) = 1, \qquad f^{(k)}(0) = m(m-1)\cdots(m-k+1)$$

ist. Die Entwicklung lautet demnach

$$(1+x)^m = 1 + mx + \frac{m(m-1)}{1 \cdot 2}\,x^2 + \cdots + \frac{m(m-1)\cdots(m-n+1)}{1 \cdot 2 \cdots n}\,x^n + o(x^n).$$

Insbesondere erhalten wir beispielsweise für $n = 2$ und $m = -1, \frac{1}{2}, -\frac{1}{2}$

$$\frac{1}{1+x} = 1 - x + x^2 + o(x^2),$$

$$\sqrt{1+x} = 1 + \frac{1}{2}\,x - \frac{1}{8}\,x^2 + o(x^2),$$

$$\frac{1}{\sqrt{1+x}} = 1 - \frac{1}{2}\,x + \frac{3}{8}\,x^2 + o(x^2).$$

Die erste dieser Entwicklungen erhält man sehr leicht auf elementarem Wege; das Restglied ist hier einfach gleich $\frac{x^3}{1+x}$. Die zweite und dritte würden jedoch längere Berechnungen erfordern (vgl. Nr. 63).

5. Wenn wir zur Logarithmusfunktion $\ln x$ übergehen, die für $x \to +0$ gegen $-\infty$ konvergiert, so ist es (wie im vorigen Beispiel) vorteilhaft, die Funktion $f(x) = \ln(1 + x)$ zu untersuchen und sie nach Potenzen von x zu entwickeln (wieder sei $1 + x > 0$). Dann gilt (Nr. 116, Beispiel 3)

$$f^{(k)}(x) = \frac{(-1)^k\,(k-1)!}{(1+x)^k},$$ [1]

$$f(0) = 0, \qquad f^{(k)}(0) = (-1)^{k-1}\,(k-1)!.$$

Hieraus folgt

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + (-1)^{n-1}\,\frac{x^n}{n} = o(x^n).$$

6. Es sei jetzt $f(x) = \arctan x$. Wir hatten in Nr. 118, Beispiel 4, bereits die Ableitungen dieser Funktion für $x = 0$ angegeben:

$$f^{(2m)}(0) = 0, \qquad f^{(2m-1)}(0) = (-1)^{m-1}\,(2m-2)!;$$

ihre Entwicklung lautet also

$$\arctan x = x - \frac{x^3}{3} + \frac{x^5}{5} - \cdots + (-1)^{m-1}\,\frac{x^{2m-1}}{2m-1} + o(x^{2m}).$$

[1]) Unter 0! verstehen wir, wie immer, die Zahl 1.

7. Für die Funktion $f(x) = \tan x$ ist die Bildung der Koeffizienten nach der Taylorschen Formel recht kompliziert. Trotzdem ist es nicht schwierig, die ersten Glieder zu ermitteln. Da beispielsweise

$$f'(x) = \frac{1}{\cos^2 x}, \quad f''(x) = 2\,\frac{\sin x}{\cos^3 x}, \quad f'''(x) = 2\,\frac{1 + 2\sin^2 x}{\cos^4 x},$$

$$f^{(4)}(x) = 8 \sin x\,\frac{2 + \sin^2 x}{\cos^5 x}$$

ist, gilt

$$f(0) = 0, \quad f'(0) = 1, \quad f''(0) = 0, \quad f'''(0) = 2, \quad f^{(4)}(0) = 0,$$

also

$$\tan x = x + \frac{x^3}{3} + o(x^4).$$

Indem man bekannte Entwicklungen benutzt, kann man ohne Berechnung der Ableitungen die Entwicklung auch für kompliziertere Funktionen direkt hinschreiben. Beispielsweise kann die Formel für $\tan x$ auch aus den Entwicklungen von $\sin x$ und $\cos x$ gewonnen werden. Wir wollen noch einige weitere Beispiele anführen und dabei die Koeffizienten aller Potenzen von x bis zu einer bestimmten Potenz einschließlich berechnen. Die höheren Potenzen werden wir jedoch (ohne sie aufzuschreiben) gleich in das Restglied aufnehmen.

8. Man entwickle die Funktion $e^{\sin x}$ bis x^3. Nach Beispiel 1 gilt

$$e^{\sin x} = 1 + \sin x + \frac{1}{2}\sin^2 x + \frac{1}{6}\sin^3 x + o(x^3),^{1)}$$

nach Beispiel 2

$$\sin x = x - \frac{1}{6}x^3 + o(x^4),$$

also

$$e^{\sin x} = 1 + \left(x - \frac{1}{6}x^3\right) + \frac{1}{2}x^2 + \frac{1}{6}x^3 + o(x^3) = 1 + x + \frac{1}{2}x^2 + o(x^3),$$

da sich die Glieder mit x^3 wegheben.

Analog ergibt sich

$$e^{\tan x} = 1 + x + \frac{1}{2}x^2 + \frac{1}{2}x^3 + o(x^3).$$

9. Man bestimme die Entwicklung von $\ln \cos x$ bis zum Glied x^6.

Nach Beispiel 5 gilt

$$\begin{aligned}\ln \cos x &= \ln\,[1 + (\cos x - 1)] \\ &= (\cos x - 1) - \frac{1}{2}(\cos x - 1)^2 + \frac{1}{3}(\cos x - 1)^3 + o(x^6).^{2)}\end{aligned}$$

Nach Beispiel 3 ist aber $\cos x - 1 = -\frac{1}{2}x^2 + \frac{1}{24}x^4 - \frac{1}{720}x^6 + o(x^7)$ und damit

$$\ln \cos x = \left(-\frac{1}{2}x^2 + \frac{1}{24}x^4 - \frac{1}{720}x^6\right) - \frac{1}{2}\left(\frac{1}{4}x^4 - \frac{1}{24}x^6\right) + \frac{1}{3}\left(-\frac{1}{8}x^6\right) + o(x^6)$$

oder, zusammengefaßt,

$$\ln \cos x = -\frac{1}{2}x^2 - \frac{1}{12}x^4 - \frac{1}{45}x^6 + o(x^6).$$

---

1) Eigentlich müßte hier $o(\sin^3 x)$ stehen; da jedoch die unendlich kleinen Größen x und $\sin x$ äquivalent sind, können wir auch $o(x^3)$ schreiben.

2) Da $1 - \cos x$ die gleiche Ordnung wie x^2 besitzt, kann man statt $o\left((\cos x - 1)^3\right)$ auch $o(x^6)$ schreiben (vgl. Nr. 61).

Analog erhält man

$$\ln(x + \sqrt{1 + x^2}) = x - \frac{1}{6}x^3 + \frac{3}{40}x^5 + o(x^5)$$

und

$$\ln\frac{\sin x}{x} = -\frac{1}{6}x^2 - \frac{1}{180}x^4 - \frac{1}{2835}x^6 + o(x^6).$$

(Für $x = 0$ setzen wir $\ln\frac{\sin x}{x}$ gleich 0, heben also die Unstetigkeit auf.)

Alle diese Entwicklungen, die wir ohne unmittelbare Benutzung der Taylorschen Formel erhalten haben, können natürlich auch nach dieser Formel gewonnen werden; dabei würden wir genau dieselben Koeffizienten bekommen, da die Taylor-Entwicklung einer Funktion eindeutig ist.

Bemerkung. Da die hier betrachteten Funktionen in der Umgebung des Punktes $x = 0$ Ableitungen aller Ordnungen besaßen, brauchten wir uns bei der Wahl der Zahl n in Formel (11) keinerlei Einschränkungen aufzulegen, d. h., wir konnten die Entwicklung dieser Funktionen bis zu einer beliebigen Potenz von x fortsetzen.

126. Andere Formen des Restgliedes. Die Taylorsche Formel mit dem Peanoschen Restglied kann vielfach angewendet werden (vgl. das folgende Kapitel); diese Anwendungen sind jedoch sozusagen „lokaler" Natur, sie beziehen sich auf den gleichen Punkt x_0. Soweit andere x-Werte in Betracht gezogen werden, liegen sie nach Voraussetzung „genügend nahe" bei x, wobei über die „Nähe" nichts Genaues bekannt ist.

Natürlich versucht man, die Funktion $f(x)$ mit Hilfe des Polynoms $p(x)$ mit vorgegebener Genauigkeit zu approximieren.

Damit das Polynom $p(x)$ diesem Zweck dienen kann, muß man die Differenz (7) für gegebenes x abschätzen können. In diesem Fall kann die Peanosche Form, die nur die Konvergenz von $r(x)$ gegen 0 bei $x \to 0$ charakterisiert, nicht benutzt werden; ihr ist nicht zu entnehmen, für welche x-Werte das Polynom $p(x)$ die Funktion $f(x)$ mit vorgegebener Genauigkeit wiedergibt; sie sagt auch nichts darüber aus, wie man bei gegebenem x die Größe des Restgliedes $r(x) = r_n(x)$ durch Vergrößerung von n beeinflussen könnte, usw.[1]) Wir wollen daher eine andere Form des Restgliedes $r_n(x)$ herleiten.

Der Bestimmtheit halber betrachten wir das Intervall $[x_0, x_0 + H]$, $H > 0$, rechts vom Punkt x_0 und setzen die Funktion $f(x)$ als in diesem Intervall definiert voraus. (Der Fall, daß $f(x)$ im Intervall $[x_0 - H, x_0]$ gegeben ist, läßt sich analog erledigen.) Wir setzen diesmal mehr voraus, nämlich, daß im ganzen Intervall $[x_0, x_0 + H]$ die ersten n Ableitungen

$$f'(x), f''(x), f'''(x), \ldots, f^{(n)}(x)$$

existieren und stetig sind und daß überdies im offenen Intervall $(x_0, x_0 + H)$ die $(n + 1)$-te Ableitung $f^{(n+1)}(x)$ existiert und endlich ist.

Nach (6) und (7) gilt

$$r_n(x) = f(x) - f(x_0) - \frac{f'(x_0)}{1!}(x - x_0) - \frac{f''(x_0)}{2!}(x - x_0)^2 - \cdots - \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n. \tag{12}$$

[1]) Man beachte, daß das Restglied $r(x)$ im allgemeinen von n abhängt. Um dies hervorzuheben, werden wir es künftig mit $r_n(x)$ bezeichnen.

Jetzt greifen wir einen beliebigen Wert x aus dem Intervall $(x_0, x_0 + H]$ heraus und bilden entsprechend der rechten Seite in Formel (12), indem wir dort die Konstante x_0 durch die Veränderliche z ersetzen, die Hilfsfunktion

$$\varphi(z) = f(x) - f(z) - \frac{f'(z)}{1!}(x - z)$$
$$- \frac{f''(z)}{2!}(x - z)^2 - \cdots - \frac{f^{(n)}(z)}{n!}(x - z)^n,$$

wobei wir die unabhängige Veränderliche z als im Intervall $[x_0, x]$ variierend voraussetzen. In diesem Intervall ist $\varphi(z)$ stetig und nimmt an den Endpunkten die Werte [vgl. (12])

$$\varphi(x_0) = r_n(x), \qquad \varphi(x) = 0$$

an. Außerdem existiert im Intervall (x_0, x) die Ableitung

$$\varphi'(x) = -f'(z) - \left[\frac{f''(z)}{1!}(x - z) - f'(z)\right] - \left[\frac{f'''(z)}{2!}(x - z)^2 - \frac{f''(z)}{1!}(x - z)\right]$$
$$- \left[\frac{f^{(4)}(z)}{3!}(x - z)^3 - \frac{f'''(z)}{2!}(x - z)^2\right] - \cdots$$
$$- \left[\frac{f^{(n+1)}(z)}{n!}(x - z)^n - \frac{f^{(n)}(z)}{(n - 1)!}(x - z)^{n-1}\right]$$

oder, vereinfacht,

$$\varphi'(z) = -\frac{f^{(n+1)}(z)}{n!}(x - z)^n.$$

Wir nehmen jetzt eine beliebige Funktion $\psi(z)$, die im Intervall $[x_0, x]$ stetig ist und deren Ableitung $\psi'(z)$ wenigstens im offenen Intervall (x_0, x) nicht verschwindet. Auf die Funktionen $\varphi(z)$ und $\psi(z)$ wenden wir die Cauchysche Formel aus Nr. 114 an,

$$\frac{\varphi(x) - \varphi(x_0)}{\psi(x) - \psi(x_0)} = \frac{\varphi'(c)}{\psi'(c)},$$

wobei

$$x_0 < c < x \quad \text{oder} \quad c = x_0 + \theta(x - x_0) \quad (0 < \theta < 1)$$

gilt. Wegen $\varphi(x) = 0$, $\varphi(x_0) = r_n(x)$, $\varphi'(x) = -\dfrac{f^{(n+1)}(c)}{n!}(x - c)^n$ ist

$$r_n(x) = \frac{\psi(x) - \psi(x_0)}{\psi'(c)} \cdot \frac{f^{(n+1)}(c)}{n!}(x - c)^n.$$

Setzen wir jetzt an Stelle von $\psi(z)$ irgendwelche den Voraussetzungen genügende Funktionen ein, so erhalten wir verschiedene Formen des Restgliedes $r_n(x)$.

Setzen wir beispielsweise $\psi(z) = (x - z)^p$, wobei $p > 0$ ist, so gilt

$$\psi'(z) = -p(x - z)^{p-1} \quad (x_0 < z < x).$$

Offenbar genügt diese Funktion den Voraussetzungen. Daher gilt

$$r_n(x) = \frac{-(x-x_0)^p}{-p(x-c)^{p-1}} \cdot \frac{f^{(n+1)}(c)}{n!}(x-c)^n = \frac{f^{(n+1)}(c)}{n!p}(x-c)^{n+1-p}(x-x_0)^p.$$

Wegen $c = x_0 + \theta(x - x_0)$ ist

$$x - c = x - x_0 - \theta(x - x_0) = (1-\theta)(x - x_0)$$

und schließlich

$$r_n(x) = \frac{f^{(n+1)}\big(x_0 + \theta(x-x_0)\big)}{n!p}(1-\theta)^{n+1-p}(x-x_0)^{n+1} \quad (0 < \theta < 1).$$

Dieser Ausdruck heißt *Restglied nach* Schlömilch[1])-Roche[2]).

Hieraus lassen sich, wenn man für p einen konkreten Wert einsetzt, speziellere Formen des Restgliedes gewinnen. Setzt man etwa $p = n + 1$, so erhält man das *Restglied in der Lagrangeschen Form,*

$$r_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-x_0)^{n+1} \quad (x_0 \lesseqgtr c \lesseqgtr x),$$

die besonders einfach ist. Dieses Restglied erinnert an das nächstfolgende Glied der Taylorschen Formel, nur daß die $(n + 1)$-te Ableitung nicht im Punkt x_0, sondern in einem Wert c (zwischen x und x_0) zu nehmen ist.

Die Taylorsche Formel mit dem Lagrangeschen Restglied hat somit die Gestalt

$$\begin{aligned} f(x) = f(x_0) &+ \frac{f'(x_0)}{1!}(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \cdots \\ &+ \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n + \frac{f^{(n+1)}(c)}{(n+1)!}(x-x_0)^{n+1} \quad (x_0 \lesseqgtr c \lesseqgtr x). \end{aligned} \tag{13}$$

Bringt man hier das Glied $f(x_0)$ nach links, so kann man leicht die direkte Verallgemeinerung der Formel des Mittelwertsatzes (Nr. 112) erkennen, die man folgendermaßen schreiben kann:

$$f(x) - f(x_0) = f'(c) \cdot (x - x_0).$$

Obgleich das Restglied in der Lagrangeschen Form angesichts seiner Einfachheit tatsächlich am meisten benutzt wird, ist diese Form in einzelnen Fällen ungeeignet zur Abschätzung des Restgliedes, und man muß sich anderer Formen bedienen, die weniger einfach sind. Wir erwähnen hier nur das *Restglied in der Cauchyschen Form,* das man aus der allgemeinen Schlömilch-Rocheschen Form für $p = 1$ erhält:

$$r_n(x) = \frac{f^{(n+1)}\big(x_0 + \theta(x-x_0)\big)}{n!}(1-\theta)^n(x-x_0)^{n+1}.$$

127. Näherungsformeln. Wir setzen der Einfachheit halber in Formel (13) nun $x_0 = 0$ und schreiben an Stelle von c jetzt θx, wobei $0 < \theta < 1$ ist:

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \cdots + \frac{f^{(n)}(0)}{n!}x^n + \frac{f^{(n+1)}(\theta x)}{(n+1)!}x^{n+1}. \tag{14}$$

[1]) Oskar Schlömilch, 1823—1901, deutscher Mathematiker.

[2]) Edouard Albert Roche, 1820—1883, französischer Mathematiker.

Lassen wir hier das Restglied weg, so erhalten wir die Näherungsformel

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

die eine Funktion komplizierter Natur durch ein Polynom ersetzt. Jedoch können wir jetzt den Fehler dieser Formel abschätzen, da er (dem Absolutbetrag nach) gerade gleich dem weggelassenen Restglied ist. Ist z. B. die $(n+1)$-te Ableitung (wenigstens für ein zwischen 0 und x variierendes Argument) dem Absolutbetrag nach durch eine Zahl M beschränkt, so ist

$$|r_n(x)| \leqq \frac{Mx^{n+1}}{(n-1)!}.$$

Als Beispiele betrachten wir die elementaren Funktionen. Es erübrigt sich, die Berechnungen aus Nr. 125 zu wiederholen; wir schreiben lediglich das Restglied in der neuen Form.

1. Wir setzen $f(x) = \mathrm{e}^x$. Die Näherungsformel lautet

$$\mathrm{e}^x \approx 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!};$$

da

$$r_n(x) = \frac{\mathrm{e}^{\theta x}}{(n+1)!}x^{n+1}$$

das Restglied ist, können wir beispielsweise den Fehler für $x > 0$ folgendermaßen abschätzen:

$$|r_n(x)| < \mathrm{e}^x \frac{x^{n+1}}{(n+1)!}.$$

Ist insbesondere $x = 1$, also

$$\mathrm{e} \approx 1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!},$$

so ist

$$|r_n(1)| < \frac{3}{(n-1)!}.$$

Eine ähnliche Formel hatten wir schon in Nr. 37 zur angenäherten Berechnung der Zahl e benutzt, jedoch war die Abschätzung des Restgliedes, die wir dort auf anderem Wege erhielten, genauer.

2. Setzen wir $f(x) = \sin x$, so erhalten wir

$$\sin x \approx x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots + (-1)^{m-1}\frac{x^{2m-1}}{(2m-1)!}.$$

In diesem Fall ergibt sich für das Restglied

$$r_{2m}(x) = \frac{\sin\left(\theta x + (2m+1)\frac{\pi}{2}\right)}{(2m+1)!}x^{2m+1} = (-1)^m \cos\theta x \cdot \frac{x^{2m+1}}{(2m+1)!},$$

und der Fehler läßt sich leicht abschätzen:

$$|r_{2m}(x)| \leqq \frac{|x|^{2m+1}}{(2m+1)!}.$$

Begnügen wir uns insbesondere mit einem einzigen Glied und setzen

$$\sin x \approx x,$$

so genügt es, um den Fehler in vorgeschriebener Weise klein zu machen (etwa kleiner als 0,001), daß wir (für $x > 0$)

$$\frac{x^3}{6} < 0{,}001 \quad \text{oder} \quad x < 0{,}1817$$

wählen, was etwa 10° ausmacht. Bei Benutzung der zweigliedrigen Formel

$$\sin x \approx x - \frac{x^3}{6}$$

genügt es, um die gleiche Genauigkeit zu erhalten,

$$\frac{x^5}{120} < 0{,}001 \quad \text{oder} \quad x < 0{,}6544 \quad \text{(das entspricht etwa } 37°\,30')$$

zu wählen; ist aber $x < 0{,}4129$ (das entspricht etwa 23° 30′), so ist der Fehler sogar $< 0{,}0001$, usw.

Wir sehen, daß in diesem Beispiel mit wachsender Gliederzahl das Taylorsche Polynom die Ausgangsfunktion immer genauer und in einem stets größeren Bereich wiedergibt. Das zeigt Abb. 52a (vgl. S. 242), in der neben der Kurve $y = \sin x$ auch die Polynome

$$y = x, \quad y = x - \frac{x^3}{6}, \quad y = x - \frac{x^3}{6} + \frac{x^5}{120} \quad \text{usw.}$$

angegeben sind.

3. Analog haben wir für die Funktion $f(x) = \cos x$

$$\cos x \approx 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots + (-1)^m \frac{x^{2m}}{2m!},$$

wobei

$$r_{2m+1}(x) = (-1)^{m+1} \cos \theta x \cdot \frac{x^{2m+2}}{(2m+2)!}$$

ist, so daß

$$|r_{2m+1}(x)| \leqq \frac{|x|^{2m+2}}{(2m+2)!}$$

ist. Beispielsweise gilt für $\cos x \approx 1 - \frac{x^2}{2}$

$$|r_3(x)| \leqq \frac{x^4}{24},$$

also ist der Fehler $< 0{,}0001$ für $x < 0{,}2213$ (das entspricht etwa 13°) usw.

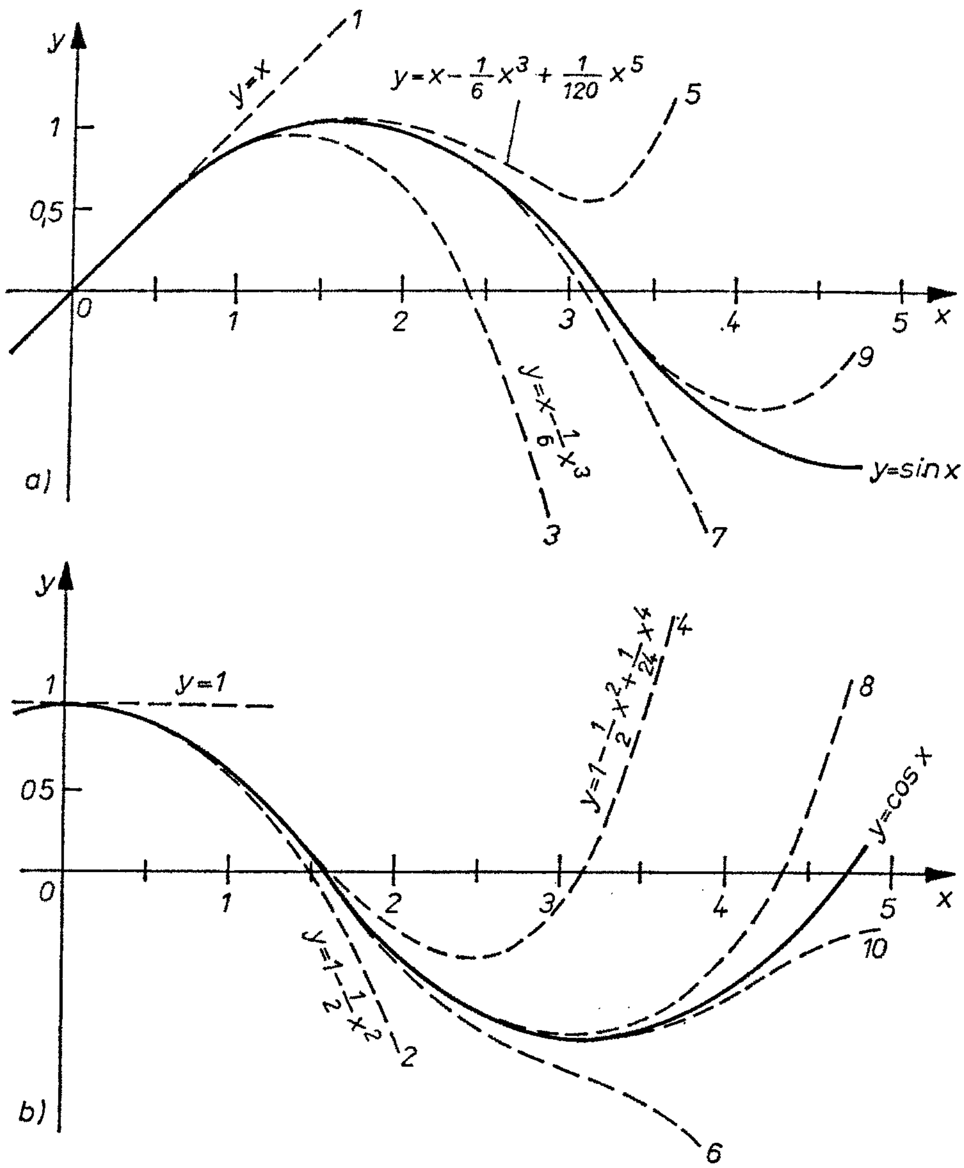

Abb. 52

In Abb. 52b sind zum Vergleich die Kurven von $y = \cos x$ und der Polynome

$$x = 1, \quad y = 1 - \frac{x^2}{2}, \quad y = 1 - \frac{x^2}{2} + \frac{x^4}{24} \quad \text{usw.}$$

angegeben.

Wir machen den Leser auf den wesentlichen Fortschritt gegenüber den Formeln in Nr. 62, 63, 107 aufmerksam. Jetzt können wir die Fehlergrenzen angeben und durch die Formeln jede beliebige Genauigkeit erreichen (da wir $\lim\limits_{n\to\infty} \frac{|x|^n}{n!} = 0$ in Nr. 35, Beispiel 1, bewiesen haben).

Wir zeigen noch, daß die Taylorsche Formel auch die Grundlage für Näherungsformeln einer ganz anderen Art ist.

4. Als erstes Beispiel gehen wir auf die *Huygenssche*[1]) *Formel* für die angenäherte Rektifikation eines Kreisbogens ein, der klein im Vergleich zum Radius ist.

Es sei s die Länge des Bogens, d die entsprechende Sehne und δ die Sehne, die dem halben Bogen entspricht (Abb. 53a). Wir stellen uns die Aufgabe, den Bogen s möglichst genau durch

[1]) Christian Huygens, 1629—1695, niederländischer Physiker.

eine Näherungsformel

$$s \approx Ad + B\delta$$

darzustellen, wobei A und B noch zu bestimmende Koeffizienten sind. Ist r der Kreisradius und $2x$ der zum Bogen s gehörende Zentriwinkel, so ist

$$d = 2r \cdot \sin x = 2r\left(x - \frac{1}{6}x^3 + \frac{\theta'}{120}x^5\right) \qquad (0 < \theta' < 1)$$

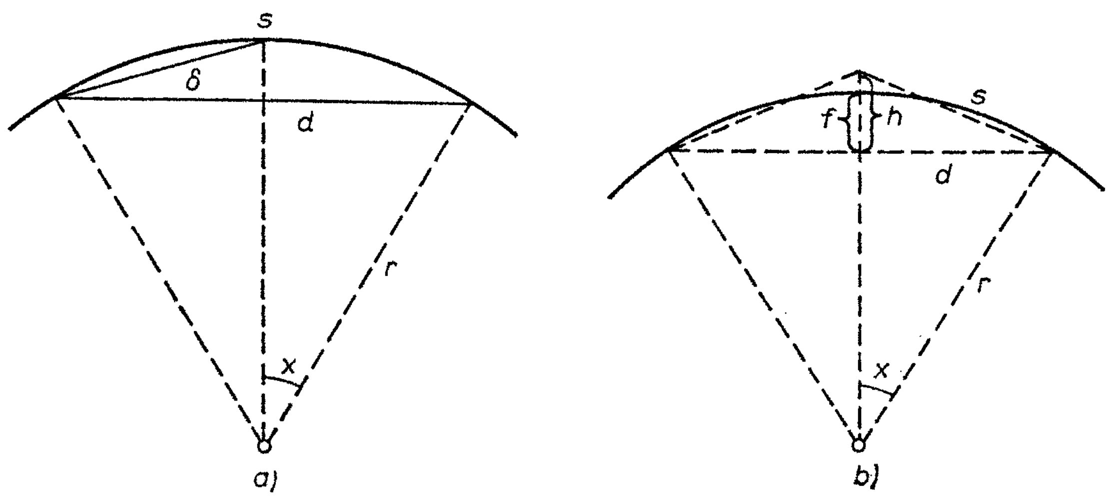

Abb. 53

und analog, wenn wir x durch $\frac{x}{2}$ ersetzen,

$$\delta = 2r\sin\frac{x}{2} = 2r\left(\frac{x}{2} - \frac{1}{48}x^3 + \frac{\theta''}{3840}x^5\right) \qquad (0 < \theta'' < 1).$$

Hieraus folgt

$$Ad + B\delta = 2r\left[\left(A + \frac{1}{2}B\right)\cdot x - \left(\frac{1}{6}A + \frac{1}{48}B\right)\cdot x^3 + \left(\frac{\theta'}{120}A + \frac{\theta''}{3840}B\right)x^5\right],$$

während $s = 2rx$ ist. Naturgemäß wählt man A und B so, daß

$$A + \frac{1}{2}B = 1, \qquad \frac{1}{6}A + \frac{1}{48}B = 0$$

ist; denn dann besteht die Differenz zwischen der linken und der rechten Seite der Formel nur aus Gliedern, die x^5 enthalten. Für die Koeffizienten A und B erhalten wir die Werte $A = -\frac{1}{3}$, $B = \frac{8}{3}$, und die Formel erhält die Gestalt

$$s \approx \frac{8\delta - d}{3} = 2\delta + \frac{2\delta - d}{3}.$$

Ihr Fehler Δ läßt sich offenbar folgendermaßen abschätzen:

$$|\Delta| < r\cdot\frac{x^5}{180}.$$

Beispielsweise gilt bei einem Zentriwinkel von 30°, d. h. $x = \frac{\pi}{12}$, gemäß dieser Abschätzung $|\Delta| < r \cdot 0{,}000007$; in Wirklichkeit ist $s = r \cdot 0{,}523\,599\ldots$, und nach der Huygensschen Formel ergibt sich $s = r \cdot 0{,}523\,593$, so daß die Differenz die angegebene Grenze nicht überschreitet.

• 5. Zum gleichen Zweck gab P. L. Tschebyscheff[1]) folgende Regel an: Der Bogen ist annähernd gleich der Summe der Schenkel des gleichschenkligen Dreiecks, das über der Sehne errichtet ist und dessen Höhe h das $\frac{2}{\sqrt{3}}$-fache der Pfeilhöhe f beträgt (Abb. 53b).

Wir setzen zunächst $h = \gamma f$. Später ergibt sich, daß wir für $\gamma = \sqrt{\frac{4}{3}}$ eine (im bestimmten Sinne) beste Annäherung erhalten.

Wie wir eben gesehen haben, ist

$$\frac{1}{2} d = r \cdot \sin x = r \left(x - \frac{1}{2} x^3 + \frac{\theta_1}{120} x^5 \right) \quad (0 < \theta_1 < 1);$$

analog ergibt sich

$$h = \gamma f = \gamma r (1 - \cos x) = \gamma r \left(\frac{1}{2} x^2 - \frac{\theta_2}{24} x^4 \right) \quad (0 < \theta_2 < 1).$$

Bezeichnen wir die Summe der Seiten des Tschebyscheffschen gleichschenkligen Dreiecks mit s^*, so ist

$$s^* = 2 \sqrt{\left(\frac{1}{2} d\right)^2 + h^2} = 2rx \sqrt{\left(1 - \frac{1}{6} x^2 + \frac{\theta_1}{120} x^4\right)^2 + \gamma^2 \left(\frac{1}{2} x - \frac{\theta_2}{24} x^3\right)^2}$$

$$= 2rx \sqrt{1 + \left(\frac{\gamma^2}{4} - \frac{1}{3}\right) x^2 + a x^4 + b x^6 + c x^8}.$$

Jetzt setzen wir, um das Glied mit x^2 unter der Wurzel zu eliminieren, den Koeffizienten von x^2 gleich 0, d. h.

$$\gamma = \frac{2}{\sqrt{3}}.$$

Um den Fehler abzuschätzen, schreiben wir den Ausdruck für s^* in der Form

$$s^* = 2rx \sqrt{1 + A x^4}, \tag{15}$$

wobei sich für A ein Ausdruck ergibt, der die zweite und vierte Potenz von x enthält.

Für $x < \frac{\pi}{2}$ erhalten wir $x^2 < 2{,}5$, $x^4 < 6{,}5$, woraus sich für A die Abschätzung $|A| < 0{,}06$, also $|A|\, x^4 < 0{,}4$ ergibt. Bezeichnen wir aus Zweckmäßigkeitsgründen Ax^4 mit y, so ist nach dem Mittelwertsatz (Nr. 112)

$$\sqrt{1 + A x^4} = \sqrt{1 + y} = 1 + \frac{y}{2\sqrt{1 + \theta y}} \quad (0 < \theta < 1).$$

Den letzten Bruch schätzt man folgendermaßen ab:

$$\left| \frac{y}{2\sqrt{1 + \theta y}} \right| < \frac{|y|}{2\sqrt{1 - |y|}} = \frac{|A|\, x^4}{2\sqrt{1 - |A|\, x^4}} < \frac{0{,}06 x^4}{2\sqrt{0{,}6}} < \frac{1}{2} \cdot 0{,}1 x^4.$$

Vergleichen wir den Ausdruck (15) für s^* mit den obigen Resultaten, so ergibt sich

$$s^* = s + \varrho \quad \text{mit} \quad |\varrho| < 0{,}1 r x^5.$$

Der Fehler ist von der gleichen Ordnung wie in der Huygensschen Formel.

---

[1]) Pafnuti Lwowitsch Tschebyscheff, 1821—1894, russischer Mathematiker.

Wir kommen auf die Taylorsche Formel mit Restglied in Kapitel XI in Band II zurück, das sich mit unendlichen Reihen befaßt. Dort wird diese Formel eine äußerst wichtige Rolle spielen.

§ 6. Interpolation

128. Die Grundaufgabe der Interpolation. Die Formel von Lagrange. Wir nehmen an, für eine im Intervall $[a, b]$ definierte Funktion $f(x)$ seien $m + 1$ ihrer Werte

$$f(x_0), f(x_1), \ldots, f(x_m) \tag{1}$$

in den Punkten $x_0, x_1, \ldots, x_m$ des Intervalls bekannt, und es sei gefordert, den Wert von $f(x)$ für irgendeinen weiteren x-Wert mit Hilfe dieser Werte zu ermitteln.

Hierin besteht die *Grundaufgabe der Interpolation.* Selbstverständlich sind in dieser Fragestellung viele Unbestimmtheiten enthalten. Gewöhnlich wird diese Aufgabe folgendermaßen verstanden: Gesucht wird ein Polynom $L(x)$ niedrigsten Grades, das in den gegebenen Punkten x_i $(i = 0, 1, \ldots, m)$, die man *Interpolationspunkte* nennt, dieselben Werte $f(x_i)$ annimmt wie die Funktion $f(x)$; für beliebiges x aus $[a, b]$ setzt man dann näherungsweise

$$f(x) \approx L(x). \tag{2}$$

Eine derartige Näherungsbeziehung nennt man *Interpolationsformel.*

Somit muß zunächst eine Interpolationsformel gefunden und danach, unter bestimmten Voraussetzungen bezüglich der Funktion $f(x)$, der Fehler der Näherungsformel (2) abgeschätzt werden.

Zur Bestimmung eines Polynoms $L(x)$, das den Bedingungen

$$L(x_i) = f(x_i) \quad (i = 0, 1, \ldots, m) \tag{3}$$

genügt, ist es zweckmäßig, die Polynome m-ten Grades

$$l_k(x) = \frac{(x - x_0)\cdots(x - x_{k-1})\,(x - x_{k+1})\cdots(x - x_m)}{(x_k - x_0)\cdots(x_k - x_{k-1})\,(x_k - x_{k+1})\cdots(x_k - x_m)} \quad (k = 0, 1, \ldots, m)$$

einzuführen, die für $x = x_k$ den Wert 1 annehmen und für $x = x_i$ (bei $i \neq k$) den Wert 0 haben. Offenbar erfüllt nun das Polynom

$$L(x) = \sum_{k=0}^{m} f(x_k)\, l_k(x) \tag{4}$$

alle Bedingungen (3). Der Grad dieses Polynoms ist nicht höher als m und durch die Bedingungen (3) eindeutig bestimmt. Man nennt es das *Lagrangesche Interpolationspolynom* und die Näherungsgleichung (2) die *Lagrangesche Interpolationsformel.*

Das Polynom $l_k(x)$ läßt sich kürzer schreiben, wenn man den Ausdruck

$$\omega(x) = (x - x_0)\,(x - x_1)\cdots(x - x_m)$$

einführt, der in den Interpolationspunkten $x_0, x_1, \ldots, x_m$ verschwindet. Offenbar ist nämlich

$$(x - x_0)\cdots(x - x_{k-1})\,(x - x_{k+1})\cdots(x - x_m) = \frac{\omega(x)}{x - x_k} \quad (x \quad x_k)$$

und

$$(x_k - x_0)\cdots(x_k - x_{k-1})\,(x_k - x_{k+1})\cdots(x_k - x_m)$$
$$= \lim_{x\to x_k} \frac{\omega(x)}{x - x_k} = \lim_{x\to x_k} \frac{\omega(x) - \omega(x_k)}{x - x_k} = \omega'(x_k).$$

Somit gilt

$$l_k(x) = \frac{\omega(x)}{\omega'(x_k)\,(x - x_k)} \quad \text{und} \quad L(x) = \sum_{k=0}^{m} \frac{\omega(x)}{\omega'(x_k)\,(x - x_k)} \cdot f(x_k).$$

129. Das Restglied der Lagrangeschen Formel. Wir wenden uns jetzt der Abschätzung der Differenz $f(x) - L(x)$ zu, wobei x ein beliebiger fester Wert im Intervall $[a, b]$, aber kein Interpolationspunkt ist. Wir setzen voraus, die Funktion $f(z)$ habe in diesem Intervall Ableitungen aller Ordnungen bis zur $(m + 1)$-ten einschließlich.

Für jede Konstante K besitzt die Funktion

$$\varphi(z) = f(z) - L(z) - K \cdot \omega(z)$$

ebenfalls $m + 1$ Ableitungen und verschwindet in den Interpolationspunkten x_i $(i = 0, 1, \ldots, m)$. Wir wählen jetzt K so, daß auch für $z = x$ die Funktion $\varphi(x)$ verschwindet, d. h., wir setzen

$$K = \frac{f(x) - L(x)}{\omega(x)} \tag{5}$$

[wegen $x \neq x_i$ ist $\omega(x) \neq 0$]. Nach dem Satz von Rolle (Nr. 111) gibt es in den $m + 1$ Intervallen zwischen den $m + 2$ Nullstellen $x, x_0, x_1, \ldots, x_m$ der Funktion $\varphi(z)$ mindestens $m + 1$ verschiedene Nullstellen ihrer Ableitung $\varphi'(z)$. Wenden wir den Satz von Rolle jetzt auf die Funktion $\varphi'(z)$ und auf die m Intervalle zwischen ihren $m + 1$ Nullstellen an, so folgt die Existenz von m verschiedenen Nullstellen der zweiten Ableitung $\varphi''(z)$, usw. Setzen wir diese Überlegung fort, so kommen wir mit dem $(m + 1)$-ten Schritt zur Existenz einer Nullstelle ξ der $(m + 1)$-ten Ableitung $\varphi^{(m+1)}(z)$, so daß

$$\varphi^{(m+1)}(\xi) = 0 \quad (a < \xi < b) \tag{6}$$

gilt. Nun ist aber $L^{(m+1)}(z) \equiv 0$, weil der Grad des Polynoms $L(z)$ nicht höher als m ist, und $\omega^{(m+1)}(z) = (m + 1)!$. Nach Definition der Hilfsfunktion $\varphi(z)$ ist also

$$\varphi^{(m+1)}(z) = f^{(m+1)}(z) - K \cdot (m + 1)!,$$

so daß aus (6)

$$K = \frac{f^{(m+1)}(\xi)}{(m + 1)!}$$

folgt. Schließlich finden wir unter Berücksichtigung von (5)

$$f(x) = L(x) + \frac{f^{(m+1)}(\xi)}{(m + 1)!}\,\omega(x) \quad (a < \xi < b). \tag{7}$$

Dies ist die *Lagrangesche Interpolationsformel mit Restglied.* Sie ist im Gegensatz zu (2) keine Näherungsgleichung, sondern sie gilt streng.

Bemerkung. Gilt im Intervall $[a, b]$

$$\max |f^{(m+1)}(z)| = M_{m+1} < \infty,$$

so erhalten wir, da dort $|\omega(z)| \leqq (b-a)^{m+1}$ ist, für den Fehler der Formel (2) die Abschätzung

$$|f(x) - L(x)| \leqq \frac{M_{m+1}}{(m+1)!} (b-a)^{m+1}.$$

Die rechte Seite strebt für $m \to \infty$ nur für eine sehr kleine Klasse von Funktionen $f(x)$ gegen 0, beispielsweise für solche Funktionen, die in $[a, b]$ beliebig oft differenzierbar sind und deren Ableitungen sämtlich durch dieselbe Konstante M beschränkt sind. In diesem Fall strebt der Fehler der Formel (2) mit zunehmender Zahl der Interpolationspunkte, unabhängig davon, in welcher Weise diese Punkte ausgewählt wurden, gleichmäßig gegen 0. Wie MARCINKIEWICZ[1]) bewiesen hat, läßt sich das für jede stetige Funktion durch geeignete, der Funktion angepaßte Wahl der aufeinanderfolgenden Systeme von Interpolationspunkten erreichen. Nach dem Satz von FABER[2]) gibt es jedoch keine Vorschrift, nach welcher man die Interpolationspunkte so wählen könnte, daß dies für alle stetigen Funktionen gleichzeitig der Fall ist. Auf diese und ähnliche Probleme können wir hier jedoch nicht näher eingehen.

130. Interpolation mit mehrfachen Punkten. Die Hermitesche Formel. Man kann auch eine *allgemeinere Interpolationsaufgabe* stellen: Man gibt $m+1$ Punkte $x_0, x_1, \ldots, x_m$ und neben den Werten der Funktion $f(x)$ in diesen Punkten noch die Werte der Ableitungen vor:

$$\left.\begin{array}{l} f(x_0), f'(x_0), \ldots, f^{(n_0)}(x_0), \\ f(x_1), f'(x_1), \ldots, f^{(n_1)}(x_1), \\ \cdots\cdots\cdots\cdots\cdots \\ f(x_m), f'(x_m), \ldots, f^{(n_m)}(x_m), \end{array}\right\} \tag{8}$$

dabei sind $n_0, n_1, \ldots, n_m$ nichtnegative ganze Zahlen. Die Anzahl dieser Bedingungen ist gleich

$$(n_0 + 1) + (n_1 + 1) + \cdots + (n_m + 1) = N.$$

Die Aufgabe, die Werte der Funktion $f(x)$ für jeden von den Interpolationspunkten verschiedenen Wert x aus $[a, b]$ unter Benutzung der unter (8) angegebenen Voraussetzungen zu berechnen, verstehen wir im einfachsten Fall folgendermaßen: Gesucht ist ein Polynom $H(x)$ niedrigsten Grades, das in jedem Interpolationspunkt x_i nebst seinen Ableitungen bis zur Ordnung n_i einschließlich dieselben Werte annimmt wie die Funktion $f(x)$ und ihre entsprechenden Ableitungen. Dann setzt man näherungsweise

$$f(x) \approx H(x). \tag{9}$$

Die Punkte x_i nennt man *Interpolationspunkte der Vielfachheit* $n_i + 1$.

---

[1]) JOSEF MARCINKIEWICZ, 1915?—?, polnischer Mathematiker.

[2] EORG FABER, 1877—1966, deutscher Mathematiker.

Man kann die Existenz und die Eindeutigkeit eines Polynoms $H(x)$ höchstens $(N-1)$-ten Grades beweisen, das allen genannten Bedingungen genügt. Man nennt es das *Hermitesche Interpolationspolynom*[1]) und die Formel (9) die *Hermitesche Interpolationsformel.*

Setzen wir alle n_i gleich 0, so kommen wir auf die Lagrangesche Formel (2) zurück. Andere Spezialfälle der Hermiteschen Formel haben wir schon kennengelernt: Wählt man nur einen einzigen Interpolationspunkt x_0, aber der Vielfachheit $n+1$, d. h., fordert man von einem Polynom $T(x)$ höchstens n-ten Grades, daß in x_0 sein Wert und die Werte von n seiner Ableitungen mit den entsprechenden Werten der Funktion $f(x)$ und ihrer Ableitungen übereinstimmen, so genügt das Taylorsche Polynom

$$T(x) = f(x_0) + \frac{f'(x)}{1!}(x-x_0) + \cdots + \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n$$

diesen Forderungen [Nr. 124, Formel (6)]. Somit ist die Näherungsformel

$$f(x) \approx T(x)$$

(vgl. Nr. 127) ebenfalls ein Spezialfall der Hermiteschen Interpolationsformel.

Das Restglied der Formel (9), das sie zu einer echten Gleichung macht, läßt sich durch ähnliche Überlegungen herleiten, wie wir sie in Nr. 129 angestellt haben. Wir betrachten das Polynom N-ten Grades

$$\Omega(z) = (z-x_0)^{n_0+1}(z-x_1)^{n_1+1}\cdots(z-x_m)^{n_m+1}$$

und setzen für $a \leqq z \leqq b$

$$\Phi(z) = f(z) - H(z) - K\cdot\Omega(z), \quad K = \text{const.}$$

Nehmen wir an, die Funktion $f(z)$ habe im Intervall $[a, b]$ Ableitungen bis zur Ordnung N, so gilt das auch für $\Phi(z)$. Wir wählen einen von den Interpolationspunkten verschiedenen Wert $z = x$ und die Konstante K so, daß

$$K = \frac{f(x) - H(x)}{\Omega(x)} \quad [\Omega(x) \neq 0] \tag{10}$$

gilt. Dann verschwindet $\Phi(z)$ auch für $z = x$. Ferner hat $\Phi(z)$ sicher $N+1$ Nullstellen, wenn jede so oft gezählt wird, wie ihre Vielfachheit angibt.[2]) Wenden wir nacheinander den Satz von Rolle wie oben an [mit der zusätzlichen Vereinbarung, daß jede mehrfache Nullstelle von $\Phi(z)$ bei den einzelnen Schritten auch als Nullstelle der entsprechenden Ableitung zu zählen ist], so gelangen wir schließlich zu der Aussage, daß auch die N-te Ableitung $\Phi^{(N)}(z)$ in einem gewissen Punkt ξ verschwindet. Hieraus folgt

$$K = \frac{f^{(N)}(\xi)}{N!}$$

---

[1]) Charles Hermite, 1822—1901, französischer Mathematiker.

[2]) Wir erweitern hier den Begriff der *Vielfachheit* einer Nullstelle, der dem Leser für Polynome bekannt ist, auf eine beliebige Funktion $\Phi(z)$. Die Zahl α heißt *p-fache Nullstelle* von $\Phi(z)$, wenn zugleich mit $\Phi(z)$ auch die Ableitungen einschließlich der $(p-1)$-ten in diesem Punkt verschwinden.

und nach (10)

$$f(x) = H(x) + \frac{f^{(N)}(\xi)}{N!}\,\Omega(x). \tag{11}$$

Das ist die *Hermitesche Interpolationsformel mit Restglied.*

Die Lagrangesche Formel mit dem Restglied (7) ist ein Spezialfall der Hermiteschen Formel. Nehmen wir einen einzigen Interpolationspunkt x_0 der Vielfachheit $n + 1$, so erhalten wir als Spezialfall von (11) die Taylorsche Formel mit dem Lagrangeschen Restglied [Nr. 126, Formel (13)].

IV. Untersuchung von Funktionen mit Hilfe der Ableitungen

§ 1. Studium des Funktionsverlaufs

131. Eine Bedingung dafür, daß eine Funktion eine Konstante ist. Wenn man untersuchen will, wie sich eine Funktion ändert, wird man zunächst fragen, unter welchen Bedingungen sie in einem gegebenen Intervall konstant bleibt bzw. sich monoton ändert (Nr. 57).

Satz 1. *Die Funktion $f(x)$ sei in einem Intervall $\mathcal{X}$ definiert*[1]) *und stetig und besitze im Innern des Intervalls eine endliche Ableitung. Dann ist $f(x)$ in $\mathcal{X}$ genau dann konstant, wenn für alle x im Innern von $\mathcal{X}$*

$$f'(x) = 0$$

ist.

Beweis. Die Notwendigkeit der Bedingung liegt auf der Hand: Aus $f(x) = \text{const}$ folgt $f'(x) = 0$. Wir beweisen nun, daß sie auch hinreichend ist. Es sei also $f'(x) = 0$ im Innern von $\mathcal{X}$. Wir wählen irgendeinen inneren Punkt x_0 und einen beliebigen anderen Punkt x aus $\mathcal{X}$. Für das Intervall $[x_0, x]$ bzw. $[x, x_0]$ sind alle Voraussetzungen des Mittelwertsatzes (Nr. 112) erfüllt. Folglich können wir

$$f(x) - f(x_0) = f'(x)\,(x - x_0)$$

schreiben, wobei c zwischen x_0 und x, also im Innern von $\mathcal{X}$ liegt. Nach Voraussetzung ist aber $f'(c) = 0$, so daß für alle x aus $\mathcal{X}$ die Beziehung

$$f(x) = f(x_0) = \text{const}$$

gilt. Damit ist unsere Behauptung bewiesen.

Die sich hieraus ergebende einfache Folgerung ist in der Integralrechnung von großer Bedeutung.

Folgerung. *Sind in einem Intervall $\mathcal{X}$ zwei Funktionen $f(x)$ und $g(x)$ definiert und stetig, existiert im Innern dieses Intervalls die endlichen Ableitungen $f'(x)$ und $g'(x)$ und gilt für jedes x im Innern von $\mathcal{X}$*

$$f'(x) = g'(x),$$

so unterscheiden sich diese Funktionen im gesamten Intervall nur um eine Konstante:

$$f(x) = g(x) + C \quad (C = \text{const}).$$

Zum Beweis braucht man nur den vorigen Satz auf die Differenz $f(x) - g(x)$ an-

---

[1]) Das Intervall $\mathcal{X}$ kann abgeschlossen, offen oder halboffen, endlich oder unendlich sein.

zuwenden. Da ihre Ableitung $(f(x) - g(x))' = f'(x) - g'(x)$ im Innern von $\mathfrak{X}$ gleich 0 ist, ist die Differenz selbst eine Konstante.

Die Besonderheiten der Anwendung dieses Satzes sollen an Beispielen erläutert werden.

1. Wir betrachten die beiden Funktionen

$$\arctan x \quad \text{und} \quad \arcsin \frac{x}{\sqrt{1+x^2}} \qquad (-\infty < x < \infty).$$

Da für die Ableitung der zweiten Funktion

$$D \arcsin \frac{x}{\sqrt{1+x^2}} = \frac{1}{\left(1 - \dfrac{x^2}{1+x^2}\right)^{1/2}} \cdot \frac{\sqrt{1+x^2} - \dfrac{x^2}{\sqrt{1+x^2}}}{1+x^2} = \frac{1}{1+x^2}$$

gilt und die Ableitung der ersten Funktion damit übereinstimmt, unterscheiden sich diese Funktionen im gesamten Intervall $(-\infty, \infty)$ um eine Konstante. Es ist also

$$\arctan x = \arcsin \frac{x}{\sqrt{1+x^2}} + C.$$

Um den Wert dieser Konstanten zu bestimmen, kann man hier beispielsweise $x = 0$ setzen. Da dann $\arctan x$ und $\arcsin x$ beide 0 werden, muß auch C gleich 0 sein. Somit haben wir die Identität

$$\arctan x = \arcsin \frac{x}{\sqrt{1+x^2}} \qquad (-\infty < x < \infty)$$

bewiesen, die übrigens in Nr. 50 mit Hilfe elementarer Überlegungen hergeleitet worden war.

2. Analog beweise man die Beziehung

$$\arcsin x = \arctan \frac{x}{\sqrt{1-x^2}} \qquad (-1 < x < 1).$$

3. Wir betrachten jetzt die Funktionen $\arctan x$ und $\dfrac{1}{2} \arctan \dfrac{2x}{1-x^2}$. Man verifiziert leicht, daß ihre Ableitungen in allen Punkten x übereinstimmen, mit Ausnahme von $x = \pm 1$ (für diese Punkte ist die zweite Funktion nicht definiert). Daher läßt sich die Identität

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

nur für jedes der Intervalle $(-1, 1)$, $(-\infty, -1)$, $(1, \infty)$ *einzeln* herleiten. Interessant hierbei ist, daß auch die Werte der Konstanten C in diesen Intervallen verschieden sind. Für das erste Intervall ist $C = 0$ (man setze $x = 0$); für die beiden anderen Intervalle gilt $C = \dfrac{\pi}{2}$ bzw. $C = -\dfrac{\pi}{2}$ (das ergibt sich leicht, wenn man die Fälle $x \to -\infty$ bzw. $x \to \infty$ betrachtet).

Alle diese Beziehungen lassen sich auch mit elementaren Methoden beweisen.

Bemerkung. Die Bedeutung des Satzes 1 zeigt sich beispielsweise dann, wenn eine Funktion gegeben ist, aus deren Definition nicht unmittelbar ersichtlich ist, daß sie konstant ist. Derartige Fälle werden wir im folgenden noch öfter antreffen.

132. Eine Bedingung für die Monotonie einer Funktion. Wir wollen jetzt klären, wie man mit Hilfe der Ableitung einer Funktion auf das Wachsen (bzw. Fallen) der Funktion selbst in einem gegebenen Intervall schließen kann.

Zunächst beschäftigen wir uns mit Funktionen, die monoton wachsend im weiteren Sinne, d. h. nicht fallend (bzw. monoton fallend im weiteren Sinne, d. h. nicht wachsend) sind (Nr. 57).

Satz 2. *Die Funktion $f(x)$ sei in einem Intervall $\mathcal{X}$ definiert und stetig und besitze im Innern dieses Intervalls eine endliche Ableitung $f'(x)$. Dann ist $f(x)$ in $\mathcal{X}$ monoton wachsend (bzw. fallend) im weiteren Sinne genau dann, wenn die Bedingung*

$$f'(x) \geqq 0 \quad (\textit{bzw. } \leqq 0) \textit{ im Innern von } \mathcal{X}$$

erfüllt ist.[1])

Beweis. Wir zeigen zunächst, daß die Bedingung notwendig ist. Ist $f(x)$ monoton nicht fallend, so erhalten wir, wenn wir einen Wert x im Innern von $\mathcal{X}$ wählen und ihm einen Zuwachs $\Delta x > 0$ erteilen,

$$f(x + \Delta x) \geqq f(x), \qquad \frac{f(x + \Delta x) - f(x)}{\Delta x} \geqq 0;$$

und in der Grenze $\Delta x \to 0$ ergibt sich $f'(x) \geqq 0$.

Jetzt zeigen wir, daß die Bedingung hinreichend ist. Es sei $f'(x) \geqq 0$ im Innern von $\mathcal{X}$. Wir wählen zwei Werte x' und x'' $(x' < x'')$ aus dem Intervall $\mathcal{X}$ und wenden auf $f(x)$ im Intervall $[x', x'']$ den Mittelwertsatz an:

$$f(x'') - f(x') = f'(c)\,(x'' - x') \quad (x' < c < x'');$$

wegen $f'(c) \geqq 0$ ist

$$f(x'') \geqq f(x'),$$

und $f(x)$ ist tatsächlich monoton nicht fallend.

Bisher haben wir den Fall nicht ausgeschlossen, daß $f(x)$ in bestimmten Intervallen konstant und dementsprechend die Ableitung in diesen Intervallen identisch 0 ist. Wenn wir diese Möglichkeit ausschließen, erhalten wir Aussagen über die Monotonie der Funktion im engeren Sinne.

Satz 3. *Unter denselben Voraussetzungen hinsichtlich der Stetigkeit der Funktion $f(x)$ und der Existenz ihrer Ableitung $f'(x)$ sind folgende Bedingungen zusammen notwendig und hinreichend dafür, daß $f(x)$ im engeren Sinn monoton wächst (bzw. fällt):*

a) $f'(x) \geqq 0$ $(\leqq 0)$ *für alle x im Innern von $\mathcal{X}$;*

b) $f'(x)$ *ist in keinem Teilintervall von $\mathcal{X}$ identisch* 0.

Beweis. Die Bedingung ist notwendig: Wächst nämlich $f(x)$ in $\mathcal{X}$, so gilt nach Satz 2

$$f'(x) \geqq 0,$$

so daß Bedingung a) erfüllt ist. Bedingung b) ist ebenfalls erfüllt. Wäre $f'(x)$ in einem Teilintervall ständig gleich 0, so wäre nach Satz 1 die Funktion $f(x)$ dort konstant, im Widerspruch zur Voraussetzung.

Die Bedingung ist hinreichend: Die Bedingungen a) und b) des Satzes seien erfüllt. Dann ist $f(x)$ nach Satz 2 auf jeden Fall nicht fallend. Wählen wir in $\mathcal{X}$ zwei Werte x'

---

[1]) Zwar formulieren wir die Sätze gleichzeitig für monoton wachsende und monoton fallende Funktionen, beschränken uns jedoch bei den Beweisen auf den Fall wachsender Funktionen.

und x'' $(x' < x'')$, so gilt nicht nur

$$f(x') \leqq f(x''), \tag{1}$$

sondern auch

$$f(x') \leqq f(x) \leqq f(x'') \quad \text{für alle } x \text{ aus } [x', x'']. \tag{2}$$

Wir zeigen nun, daß das Gleichheitszeichen in (1) tatsächlich nicht auftreten kann. Wäre nämlich $f(x') = f(x'')$, so würden wir auf Grund von (2)

$$f(x') = f(x) = f(x'')$$

für alle x aus $[x', x'']$ erhalten, $f(x)$ wäre im Intervall $[x', x'']$ konstant, und es wäre in diesem Intervall ständig $f'(x) = 0$, entgegen der Bedingung b). Also ist

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

für $x' < x''$, d. h., die Funktion $f(x)$ ist im engeren Sinne wachsend. Damit ist der Satz bewiesen.

Der hier nachgewiesene Zusammenhang zwischen dem Vorzeichen der Ableitung und der Art der Änderung der Funktion wird geometrisch einleuchtend, wenn man daran denkt (Nr. 91, 92), daß die Ableitung geometrisch den Richtungskoeffizienten der Tangente an die Kurve der Funktion darstellt. Das Vorzeichen dieses Richtungskoeffizienten bringt zum Ausdruck, daß die Tangente nach oben bzw. nach unten geneigt ist, d. h., daß die Kurve selbst ansteigt bzw. fällt (Abb. 54).

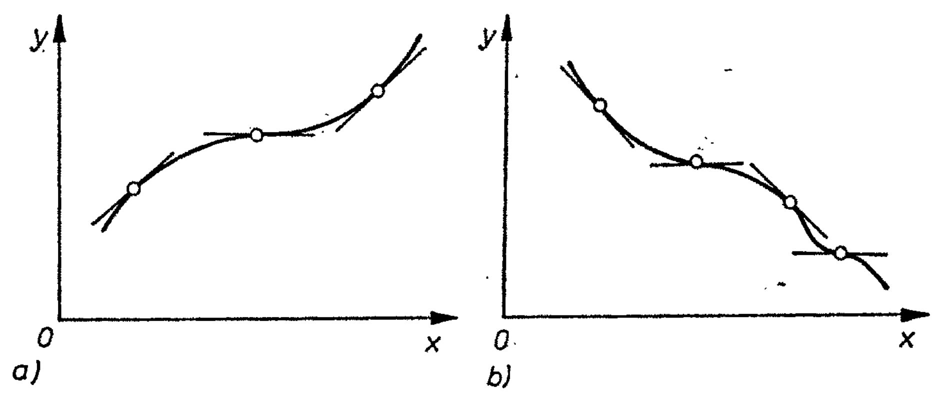

Abb. 54

Jedoch kann hierbei die Tangente in einzelnen Punkten auch horizontal verlaufen, d. h., die Ableitung einer (sogar im engeren Sinne) wachsenden (bzw. fallenden) Funktion kann für einzelne Werte von x gleich 0 sein.

Beispiele.

1. Das einfachste Beispiel für diesen Sachverhalt liefert die Funktion $f(x) = x^3$. Sie wächst zwar überall im engeren Sinne, jedoch wird ihre Ableitung $f'(x) = 3x^2$ für $x = 0$ gleich 0.

2. Analog liegt der Fall bei der Funktion

$$f(x) = x - \sin x.$$

Auch sie wächst, aber ihre nichtnegative Ableitung $f'(x) = 1 - \cos x$ wird für die Werte $x = 2k\pi$ $(k = 0, \pm 1, \pm 2, \ldots)$ gleich 0.

3. Um schließlich zu zeigen, daß die Ableitung einer wachsenden Funktion in einem endlichen Intervall sogar unendlich oft 0 werden kann, betrachten wir die Funktion

$$f(x) = e^{\sin\frac{1}{x} - \frac{1}{x}} \quad \text{für} \quad x > 0, \qquad f(0) = 0.$$

Offenbar gilt $\lim\limits_{x\to+0} f(x) = 0$, so daß unsere Funktion auch für $x = 0$ stetig ist. Für $x > 0$ gilt

$$f'(x) = e^{\sin\frac{1}{x} - \frac{1}{x}} \cdot \left(1 - \cos\frac{1}{x}\right)\frac{1}{x^2} \geqq 0,$$

wobei das Gleichheitszeichen für $x = \dfrac{1}{2k\pi}$ $(k = 1, 2, 3, \ldots)$ gilt.

Wir bemerken, daß

$$0 \leqq f'(x) < 2e \cdot \frac{\frac{1}{x^2}}{e^{1/x}} \to 0 \quad \text{für} \quad x \to +0$$

gilt, und hieraus folgt nach Nr. 113, daß auch $f'(0) = 0$ ist.

Es lassen sich Beispiele wachsender (bzw. fallender) Funktionen angeben, bei welchen die Punkte, in denen die Ableitung 0 wird, noch komplizierter verteilt sind. Solche Fälle sind jedoch selten, und in der Praxis benutzt man oft das folgende hinreichende Kriterium: *Gilt mit eventueller Ausnahme endlich vieler Punkte überall $f'(x) > 0$ (bzw. < 0), so ist $f(x)$ wachsend (bzw. fallend).*

Dieses Kriterium ist für die Anwendungen sehr bequem.

Als Beispiel betrachten wir die Funktion $f(x) = \left(1 + \dfrac{1}{x}\right)^x$ für $x > 0$ und beweisen, daß sie wächst.

Es genügt zu zeigen, daß ihr Logarithmus

$$g(x) = \ln f(x) = x[\ln(x + 1) - \ln x]$$

wächst; offenbar gilt

$$g'(x) = [\ln(x + 1) - \ln x] - \frac{1}{x + 1}.$$

Da nach dem Mittelwertsatz (Nr. 112)

$$\ln(x + 1) - \ln x = \frac{1}{\xi} \qquad (x < \xi < x + 1)$$

gilt, ist $g'(x) > 0$. Also wächst $g(x)$, was zu beweisen war.

133. Beweis von Ungleichungen. Das angegebene einfache Kriterium für die Monotonie läßt sich zum Beweis von Ungleichungen erfolgreich anwenden.

1. Wir beweisen, daß für $0 < x < \dfrac{\pi}{2}$ die Beziehung

$$\sin x > \frac{2}{\pi} x$$

gilt. Es sei $f(x) = \dfrac{\sin x}{x}$ $\left(0 < x \leqq \dfrac{\pi}{2}\right)$. Die Ableitung

$$f'(x) = \frac{\cos x(x - \tan x)}{x^2} \qquad \left(0 < x < \frac{\pi}{2}\right)$$

ist negativ; denn es ist $x < \tan x$. Also fällt $f(x)$, und es ist $f(x) > f\left(\dfrac{\pi}{2}\right) = \dfrac{2}{\pi}$ für $0 < x < \dfrac{\pi}{2}$.

2. Die Funktion $f(x) = \cos x - 1 + \dfrac{1}{2}x^2$ wird für $x = 0$ gleich 0. Ihre Ableitung

$$f'(x) = -\sin x + x$$

ist für $x > 0$ wegen $\sin x < x$ positiv. Also ist die Funktion $f(x)$ für $x \geqq 0$ wachsend, und für $x > 0$ gilt $f(x) > f(0) = 0$; somit ist

$$\cos x > 1 - \frac{1}{2} x^2.$$

Hieraus erhalten wir analog für $x > 0$ die Beziehung

$$\sin x > x - \frac{1}{6} x^3,$$

usw.

3. Man beweise für $0 < x < \frac{\pi}{2}$ die Beziehung

$$\tan x > x + \frac{1}{3} x^3.$$

Dazu genügt es zu zeigen, daß für die angegebenen x-Werte die Ableitung der Funktion $\tan x - x - \frac{1}{3} x^3$, die gleich $\sec^2 x - 1 - x^2$ ist, positiv ist, d. h., daß $\tan^2 x - x^2 > 0$ ist. Das führt aber auf die aus Nr. 54, Formel (9), bekannte Ungleichung $\tan x > x$.

4. Da die Funktion $f(x) = \ln x - x$ $(x > 0)$ die Ableitung

$$f'(x) = \frac{1}{x} - 1 \begin{cases} > 0 & \text{für} \quad 0 < x < 1, \\ < 0 & \text{für} \quad x > 1 \end{cases}$$

besitzt, ist sie wachsend, solange x im Intervall $(0, 1]$ variiert, und fallen im Intervall $[1, \infty)$. Daher ist klar, daß $f(1) = -1$ der größte Wert der Funktion ist, so daß für $x > 0$ die Beziehung

$$\ln x \leqq x - 1$$

gilt.

5. Wir betrachten noch die Funktion $f(x) = x^\alpha - \alpha x$ für $x \geqq 0$ (wobei wir $0 < \alpha < 1$ annehmen wollen). Für die Ableitung gilt

$$f'(x) = \alpha(x^{\alpha-1} - 1) \begin{cases} > 0 & \text{für} \quad 0 < x < 1, \\ < 0 & \text{für} \quad x > 1, \end{cases}$$

und wir können wie in Beispiel 4 schließen, daß für $x > 0$

$$x^\alpha - \alpha x \leqq 1 - \alpha \tag{3}$$

ist. Diese einfache Ungleichung ist der Ausgangspunkt zur Herleitung einer Reihe klassischer Ungleichungen. Im Zusammenhang damit ist es zweckmäßig, sie noch in anderer Gestalt zu formulieren.

Setzt man $x = \frac{a}{b}$, wobei a und b beliebige positive Zahlen sind, und bezeichnet man $1 - \alpha$ mit β, so geht (3) über in

$$a^\alpha b^\beta \leqq \alpha a + \beta b \qquad (a, b, \alpha, \beta > 0; \alpha + \beta = 1). \tag{3a}$$

Manchmal führt man die Zahlen $k = \frac{1}{\alpha} > 1$ und $k' = \frac{1}{\beta} > 1$ ein, so daß $k' = \frac{k}{k-1}$ gilt. Ersetzt man in (3a) a und b durch a^k bzw. $b^{k'}$, so erhält man

$$ab \leqq \frac{1}{k} a^k + \frac{1}{k'} b^{k'} \qquad \left(a, b > 0; k, k' > 1; \frac{1}{k} + \frac{1}{k'} = 1\right). \tag{3b}$$

6. Zunächst kann man die Ungleichung (3a) auf den Fall beliebig vieler miteinander multiplizierter Potenzen ausdehnen. Dabei vollzieht sich der Übergang von zwei zu drei Faktoren

folgendermaßen [unter zweimaliger Anwendung von (3a)]:

$$a^\alpha b^\beta c^\gamma = a^\alpha \cdot \left(b^{\frac{\beta}{\beta+\gamma}} c^{\frac{\gamma}{\beta+\gamma}}\right)^{\beta+\gamma} \leqq \alpha a + (\beta+\gamma) \cdot b^{\frac{\beta}{\beta+\gamma}} c^{\frac{\gamma}{\beta+\gamma}}$$

$$\leqq \alpha a + (\beta+\gamma)\left(\frac{\beta}{\beta+\gamma} b + \frac{\gamma}{\beta+\gamma} c\right) = \alpha a + \beta b + \gamma c;$$

also ist

$$a^\alpha b^\beta c^\gamma \leqq \alpha a + \beta b + \gamma c \qquad (a, b, c, \alpha, \beta, \gamma > 0; \alpha + \beta + \gamma = 1).$$

Analog vollzieht sich der Übergang von n auf $n+1$ Faktoren; durch Induktion erhält man die allgemeinere Ungleichung (wobei wir die Bezeichnungen geändert haben):

$$a_1^{q_1} a_2^{q_2} \cdots a_n^{q_n} \leqq q_1 a_1 + q_2 a_2 + \cdots + q_n a_n$$

$$(a_1, \ldots, a_n, q_1, \ldots, q_n) > 0; q_1 + \cdots + q_n = 1).$$

An Stelle der q_i kann man auch beliebige Zahlen $p_i > 0$ einführen, indem man $q_i = \frac{p_i}{\sum_j p_j}$ setzt, so daß $\sum_i q_i = 1$ ist. Dann läßt sich die Ungleichung folgendermaßen schreiben:

$$\left(a_1^{p_1} a_2^{p_2} \cdots a_n^{p_n}\right)^{1/\sum_j p_j} \leqq \frac{p_1 a_1 + p_2 a_2 + \cdots + p_n a_n}{p_1 + p_2 + \cdots + p_n} \tag{4}$$

$$(a_1, \ldots, a_n, p_1, \ldots, p_n > 0).$$

Für $p_1 = p_2 = \cdots = p_n = 1$ erhalten wir die bekannte Ungleichung

$$\sqrt[n]{a_1 a_2 \cdots a_n} \leqq \frac{a_1 + a_2 + \cdots + a_n}{n}, \tag{4a}$$

welche besagt, daß *das geometrische Mittel positiver Zahlen höchstens gleich ihrem arithmetischen Mittel ist.* Somit ist Ungleichung (4) eine naturgemäße Verallgemeinerung dieser klassischen Aussage.

7. Wir wollen nun die sogenannte *Cauchy-Höldersche Ungleichung*

$$\sum_{i=1}^n a_i b_i \leqq \left\{\sum_{i=1}^n a_i^k\right\}^{1/k} \cdot \left\{\sum_{i=1}^n b_i^{k'}\right\}^{1/k'} \qquad \left(a_i, b_i > 0; k, k' > 1; \frac{1}{k} + \frac{1}{k'} = 1\right) \tag{5}$$

beweisen.[1]) CAUCHY hatte diese Ungleichung für den Spezialfall $k = k' = 2$ bewiesen:

$$\sum_{i=1}^n a_i b_i \leqq \sqrt{\sum_{i=1}^n a_i^2} \cdot \sqrt{\sum_{i=1}^n b_i^2}. \tag{5a}$$

Wir setzen zunächst

$$\sum_{i=1}^n a_i^k = \sum_{i=1}^n b_i^{k'} = 1 \tag{6}$$

voraus, so daß die zu beweisende Ungleichung die Form

$$\sum_{i=1}^n a_i b_i \leqq 1$$

erhält. In (3b) setzen wir nacheinander $a = a_i$, $b = b_i$ ($i = 1, 2, \ldots, n$) und addieren alle so erhaltenen Ungleichungen. Wegen (6) gelangen wir zu dem gesuchten Ergebnis.

Der allgemeine Fall läßt sich auf den eben betrachteten Spezialfall zurückführen, wenn man

---

[1]) OTTO HÖLDER, 1859–1937, deutscher Mathematiker.

an Stelle der Zahlen, a_i, b_i die Zahlen

$$a_i' = \frac{a_i}{\left\{\sum\limits_{j=1}^{n} a_j^k\right\}^{1/k}}, \quad b_i' = \frac{b_i}{\left\{\sum\limits_{j=1}^{n} b_j^{k'}\right\}^{1/k'}}$$

einführt, für welche die Bedingung (6) erfüllt ist. Nach dem bereits Bewiesenen gilt

$$\sum_{i=1}^{n} a_i' b_i' \leqq 1,$$

was mit (5) gleichwertig ist.

8. Aus der Cauchy-Hölderschen Ungleichung erhält man unmittelbar eine weitere wichtige Ungleichung, die *Minkowskische Ungleichung*[1])

$$\left\{\sum_{i=1}^{n} (a_i + b_i)^k\right\}^{1/k} \leqq \left\{\sum_{i=1}^{n} a_i^k\right\}^{1/k} + \left\{\sum_{i=1}^{n} b_i^k\right\}^{1/k} \qquad (a_i, b_i > 0;\ k > 1). \tag{7}$$

Offenbar gilt

$$\sum_{i=1}^{n} (a_i + b_i)^k = \sum_{i=1}^{n} a_i(a_i + b_i)^{k-1} + \sum_{i=1}^{n} b_i(a_i - b_i)^{k-1}.$$

wendet man auf die letzten beiden Summen die Cauchy-Höldersche Ungleichung (5) an, so erhält man (wegen $\frac{1}{k} + \frac{1}{k'} = 1$) die Beziehung

$$\begin{aligned} \sum_{i=1}^{n} (a_i + b_i)^k &\leqq \left\{\sum_{i=1}^{n} a_i^k\right\}^{1/k} \cdot \left\{\sum_{i=1}^{n} (a_i + b_i)^{(k-1)k'}\right\}^{1/k'} \\ &\quad + \left\{\sum_{i=1}^{n} b_i^k\right\}^{1/k} \cdot \left\{\sum_{i=1}^{n} (a_i + b_i)^{(k-1)k'}\right\}^{1/k'} \\ &= \left[\left\{\sum_{i=1}^{n} a_i^k\right\}^{1/k} + \left\{\sum_{i=1}^{n} b_i^k\right\}^{1/k}\right] \cdot \left\{\sum_{i=1}^{n} (a_i + b_i)^k\right\}^{1/k'} \end{aligned}$$

und gelangt schließlich zu (7), indem man durch den letzten Faktor kürzt.

134. Maxima und Minima. Notwendige Bedingungen. Ist eine Funktion $f(x)$ in einem Intervall $[a, b]$ definiert und stetig und ist sie in diesem Intervall nicht monoton, so läßt sich ein Teilintervall $[\alpha, \beta]$ von $[a, b]$ angeben, in dem sie in einem inneren, d. h. zwischen α und β liegenden Punkt einen größten oder kleinsten Wert annimmt.

In der graphischen Darstellung der Funktion (Abb. 55) entsprechen solchen Intervallen charakteristische „Berge" oder „Täler". Man sagt, die Funktion $f(x)$ besitze im Punkt x_0 ein *Maximum* (bzw. *Minimum*), wenn es eine Umgebung $(x_0 - \delta, x_0 + \delta)$

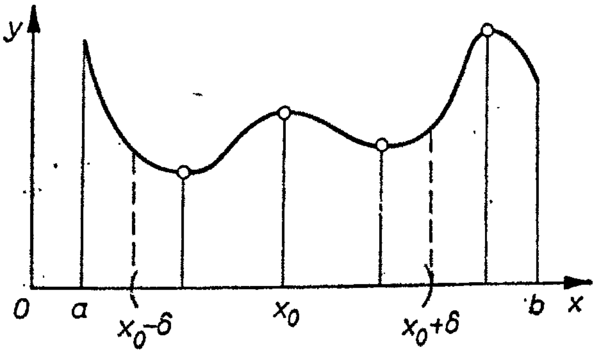

Abb. 55

[1]) Hermann Minkowski, 1864—1909, deutcher Mathematiker.

dieses Punktes im Definitionsintervall der Funktion gibt derart, daß für alle Punkte x dieser Umgebung die Ungleichung

$$f(x) \leqq f(x_0) \quad [\text{bzw.} \quad f(x) \geqq f(x_0)]$$

erfüllt ist.

Mit anderen Worten: Eine Funktion $f(x)$ besitzt im Punkt x_0 ein Maximum (bzw. Minimum), wenn der Wert $f(x_0)$ nicht kleiner (bzw. nicht größer) ist als die Werte der Funktion in allen Punkten einer Umgebung dieses Punktes (die sehr klein sein kann!). Wir betonen, daß bei dieser Definition des Maximums (Minimums) verlangt wird, daß die Funktion auf beiden Seiten von Punkt x_0 definiert ist. Da hierbei nur auf Punkte einer gewissen (evtl. sehr kleinen) Umgebung von x_0 Bezug genommen wird, spricht man auch vielfach von einem *relativen Maximum* (*Minimum*).[1]) Existiert eine solche Umgebung, in deren Grenzen (für $x \neq x_0$) die echte Ungleichung

$$f(x) < f(x_0) \quad [\text{bzw.}\ f(x) > f(x_0)]$$

erfüllt ist, so sagt man auch, die Funktion besitze im Punkt x_0 ein *Maximum* (bzw. *Minimum*) *im engeren Sinne*, anderenfalls spricht man von einem *Maximum* (bzw. *Minimum*) *im weiteren Sinne*.

Besitzt eine stetige Funktion in den Punkten x_0 und x_1 Maxima, so folgt aus dem zweiten Weierstraßschen Satz für das Intervall $[x_0, x_1]$, daß sie einen kleinsten Wert in einem bestimmten Punkt x_2 zwischen x_0 und x_1 in diesem Intervall annimmt, dort also ein Minimum besitzt. Analog liegt zwischen zwei Minima auf jeden Fall ein Maximum. In diesem einfachsten (aber in der Praxis wichtigsten) Fall, daß eine Funktion überhaupt nur endlich viele Maxima und Minima besitzt, folgen diese abwechselnd aufeinander.

Für Maxima und Minima wird die gemeinsame Bezeichnung *Extrema* benutzt.[2])

Wir stellen uns nun die Aufgabe, alle Argumentwerte zu ermitteln, an denen die Funktion einen Extremwert besitzt. Bei der Lösung dieses Problems wird die Ableitung eine wesentliche Rolle spielen.

Zunächst setzen wir voraus, die Funktion $f(x)$ besitze im Intervall (a, b) eine endliche Ableitung. Hat $f(x)$ im Punkt x_0 einen Extremwert, so können wir mit Hilfe des Satzes von FERMAT (Nr. 109), den wir auf das Intervall $(x_0 - \delta, x_0 + \delta)$ anwenden, schließen, daß $f'(x_0) = 0$ ist. Wir haben damit eine *notwendige Bedingung* für das Vorliegen eines Extremwertes. Extremwerte können somit nur in solchen Punkten vorliegen, in denen die Ableitung der Funktion gleich 0 ist. Solche Punkte heißen *stationär*.[3])

Man darf jedoch nicht denken, jedem stationären Punkt entspreche ein Extremwert der Funktion. Die genannte Bedingung ist nämlich nicht hinreichend. Wir haben beispielsweise in Nr. 132, Beispiel 1, gesehen, daß die Ableitung $y' = 3x^2$ der Funktion $y = x^3$ in $x = 0$ den Wert 0 hat. In diesem Punkt besitzt jedoch die Funktion keinen Extremwert, sie ist vielmehr ständig wachsend.

Erweitert man die Klasse der zu untersuchenden Funktionen und läßt zu, daß in endlich vielen Punkten, zu denen z. B. die Endpunkte gehören, keine (endliche) Ableitung existiert, so ist es sogar nicht ausgeschlossen, daß ein Extremwert in einen

---

[1]) Zum Begriff des *absoluten* Maximums (Minimums) vgl. S. 269.

[2]) Diese Ausdrücke stammen aus dem Lateinischen; maximum = das Größte, minimum = das Kleinste, extremum = das Äußerste.

[3]) In diesen Punkten kommt die Änderung der Funktion sozusagen zum Stillstand. Die Geschwindigkeit ihrer Änderung (Nr. 92) wird 0.

dieser Punkte fällt; der Satz von FERMAT behauptet $f'(x) = 0$ nur unter der Voraussetzung, daß eine endliche Ableitung vorhanden ist.

Beispielsweise besitzt die Funktion $x^{2/3}$, falls sie für $x < 0$ als $|x|^{2/3}$ erklärt ist, offenbar ein Minimum für $x = 0$, während ihre linksseitige Ableitung in diesem Punkt gleich $-\infty$ und ihre rechtsseitige Ableitung gleich ∞ ist (Nr. 101). Auch die Funktion $y = |x|$ hat im Punkt $x = 0$ ein Minimum, obgleich in diesem Punkt keine beiderseitige Ableitung vorhanden ist (Nr. 100). Daher können auch in Punkten, in denen keine beiderseitige endliche Ableitung existiert, Extremwerte der Funktion angenommen werden. Natürlich müssen keineswegs in solchen Punkten Extremwerte der Funktion liegen. Als Beispiele können die Funktionen $y = x^{1/3}$ $(= -|x|^{1/3}$ für $x < 0)$ und $y = x \cdot \sin \frac{1}{x}$ [für $x \neq 0$, $y(0) = 0$] dienen. Die erste besitzt im Punkt $x = 0$ eine unendliche Ableitung (Nr. 101), die zweite hat in diesem Punkt überhaupt keine Ableitung (Nr. 102, 1); der Punkt $x = 0$ ist aber weder für die erste noch für die zweite dieser Funktionen ein Extremwert (weil in jeder Umgebung dieses Punktes beide Funktionen sowohl positive als auch negative Werte annehmen).

135. Hinreichende Bedingungen. Erste Regel. Wenn also ein Punkt x_0 für eine Funktion $f(x)$ ein stationärer Punkt ist oder in diesem Punkt keine beiderseitige endliche Ableitung existiert, so kann man ihn sozusagen nur als „extremwertverdächtig" auffassen und muß weitere Untersuchungen anstellen.

Diese Untersuchung besteht darin, daß man nachprüft, ob eine hinreichende Bedingung für das Auftreten eines Extremwertes erfüllt ist. Solche Bedingungen wollen wir nun herleiten. Wir setzen voraus, in einer Umgebung $(x_0 - \delta, x_0 + \delta)$ des Punktes x_0 (jedenfalls für $x \neq x_0$) existiere eine endliche Ableitung $f'(x)$, und diese habe in dieser Umgebung links von x_0 und ebenso auch rechts von x_0 ständig dasselbe Vorzeichen. Dann sind folgende drei Fälle möglich:

I. Es ist $f'(x) > 0$ für $x < x_0$ und $f'(x) < 0$ für $x > x_0$, d. h., *die Ableitung $f'(x)$ ändert beim Durchgang durch den Punkt x_0 ihr Vorzeichen (geht von positiven zu negativen Werten über)*. Dann ist $f(x)$ im Intervall $[x_0 - \delta, x_0]$ wachsend, und im Intervall $[x_0, x_0 + \delta]$ fallend (Nr. 132), so daß der Wert $f(x_0)$ der größte im Intervall $[x_0 - \delta, x_0 + \delta]$ ist, d. h., die Funktion $f(x)$ hat im Punkt x_0 ein *Maximum im engeren Sinne.*

II. Es gilt $f'(x) < 0$ für $x < x_0$ und $f'(x) > 0$ für $x > x_0$, d. h., *die Ableitung $f'(x)$ ändert ihr Vorzeichen beim Durchgang durch den Punkt x_0 (geht von negativen zu positiven Werten über)*. Dann hat $f(x)$ im Punkt x_0 ein *Minimum im engeren Sinne.*

III. Es gilt $f'(x) > 0$ sowohl für $x < x_0$ als auch für $x > x_0$ oder aber $f'(x) < 0$ sowohl links als auch rechts von x_0, d. h., *$f'(x)$ ändert das Vorzeichen im Punkt x_0 nicht.* Dann ist die Funktion entweder ständig wachsend oder ständig fallend. In beliebiger Nähe von x_0 gibt es auf einer Seite Punkte x, in welchen $f(x) < f(x_0)$ gilt, und auf der anderen Seite Punkte x, für die $f(x) > f(x_0)$ ist, so daß im Punkt x_0 *kein Extremwert* vorliegt.

Die verschiedenen Möglichkeiten sind in Abb. 56 graphisch dargestellt.

Somit haben wir eine *erste Regel* zur Untersuchung eines „extremwertverdächtigen" Wertes x_0 erhalten: *Wir setzen in der Ableitung $f'(x)$ zunächst $x < x_0$ und danach $x > x_0$ und bestimmen das Vorzeichen von $f'(x)$ in der linksseitigen und der rechtsseitigen Umgebung des Punktes x_0. Ändert hierbei $f'(x)$ das Vorzeichen von Plus in Minus, so liegt ein Maximum vor; wechselt das Vorzeichen von Minus in Plus, so handelt es sich um ein Minimum. Ändert sich das Vorzeichen nicht, so liegt in diesem Punkt kein Extremwert vor.*

Diese Regel löst das Problem vollständig für den Fall, daß das Intervall (a, b), wie es oft vorkommt, nur endlich viele stationäre Punkte enthält bzw. nur endlich viele Punkte, in denen keine endliche Ableitung vorhanden ist:

$$a < x_1 < x_2 < \cdots < x_k < x_{k+1} < \cdots < x_n < b. \tag{4}$$

Dann existiert nämlich in jedem Intervall

$$(a, x_1), (x_1, x_2), \ldots, (x_k, x_{k+1}), \ldots, (x_n, b)$$

eine endliche Ableitung $f'(x)$, und außerdem behält $f'(x)$ in jedem solchen Intervall, wie sich sogleich zeigen wird, dasselbe Vorzeichen bei.

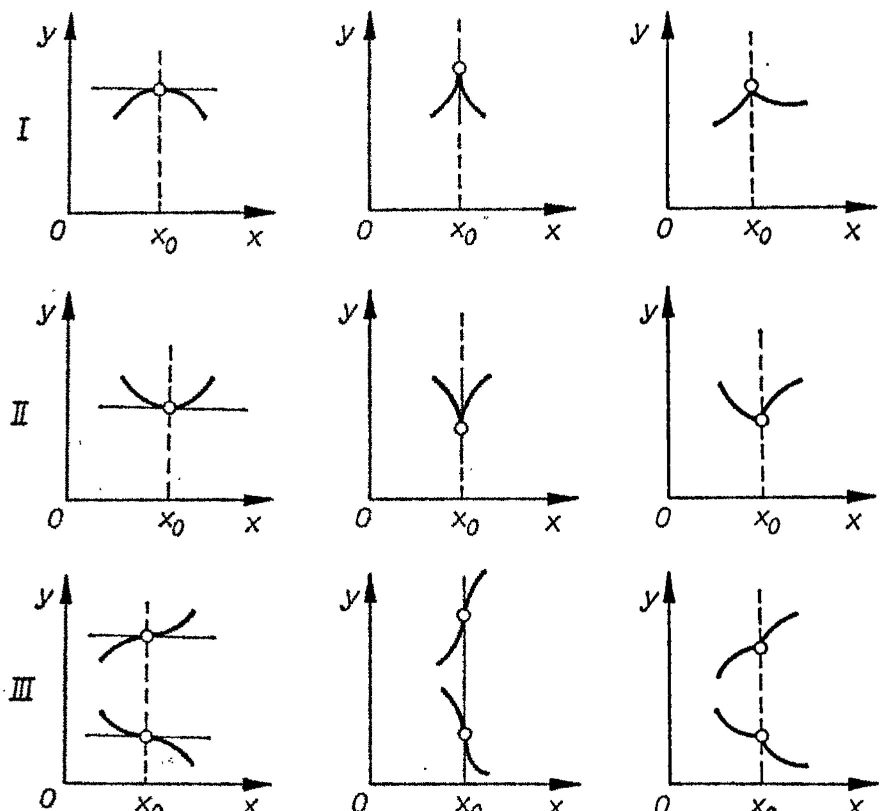

Abb. 56

Würde nämlich $f'(x)$ beispielsweise im Intervall (x_k, x_{k+1}) das Vorzeichen wechseln, so müßte $f'(x)$ nach dem Satz von Darboux (Nr. 110) in einem bestimmten Punkt zwischen x_k und x_{k+1} den Wert 0 annehmen. Das ist aber nicht möglich, da alle Nullstellen der Ableitung bereits unter den Punkten (4) enthalten sind. Diese Bemerkung ist verschiedentlich bei praktischen Anwendungen von Nutzen. Das Vorzeichen der Ableitung $f'(x)$ ist im gesamten Intervall (x_k, x_{k+1}) bestimmt, wenn man ihren Wert (oder auch nur das Vorzeichen) in einem einzigen Punkt dieses Intervalls berechnet hat.

136. Beispiele.

1. Man bestimme die Extremwerte der Funktion $f(x) = (x + 2)^2 (x - 1)^3$.
Ihre Ableitung existiert überall und ist endlich:

$$f'(x) = 2(x + 2)(x - 1)^3 + 3(x + 2)^2 (x - 1)^2 = (x + 2)(x - 1)^2 (5x + 4).$$

Die Nullstellen der Ableitung (die stationären Punkte) sind

$$x_1 = -2, \qquad x_2 = -\frac{4}{5} = -0{,}8, \qquad x_3 = 1.$$

Durch diese Werte wird das Intervall $(-\infty, \infty)$ in folgende Teilintervalle zerlegt:

$$(-\infty, -2), \quad (-2, -0{,}8), \quad (-0{,}8, 1), \quad (1, \infty).$$

Um das Vorzeichen der Ableitung in diesen Intervallen zu bestimmen, braucht man es unter Benutzung der obigen Bemerkung nur für konkrete Werte, etwa -3, -1, 0 und 2 zu ermitteln. Wir bestimmen das Vorzeichen der verschiedenen Faktoren und erhalten für die Ableitung folgende Vorzeichen:

$$\begin{array}{lll} \text{im Intervall} & (-\infty, -2) & (-)(+)(-) = + \\ & (-2, -0{,}8) & (+)(+)(-) = - \\ & (-0{,}8, 1) & (+)(+)(+) = + \\ & (1, \infty) & (+)(+)(+) = + \end{array}$$

Hieraus geht hervor, daß $f(x)$ in $x = -2$ ein Maximum, in $x = -0{,}8$ ein Minimum besitzt; dagegen existiert in $x = 1$ kein Extremwert.

Gewöhnlich geht man jedoch anders vor und setzt keine konkreten Werte in die Ableitung ein. Wir beginnen mit $x = -2$. Das Produkt der letzten beiden Faktoren der Ableitung ist in $x = -2$ negativ, also auch (auf Grund der Stetigkeit) in der Nähe dieses Punktes (sowohl in der linksseitigen als auch in der rechtsseitigen Umgebung). Der Faktor $x + 2$ jedoch wechselt, wenn x wachsend den Wert -2 durchläuft, das Vorzeichen von Minus in Plus, so daß die Ableitung ihr Vorzeichen von Plus in Minus ändert, also die Funktion in $x = -2$ ein Maximum besitzt. Für $x = -\frac{4}{5}$ (und in der Nähe dieses Wertes) sind die ersten beiden Faktoren der Ableitung positiv, der letzte Faktor $5x + 4$ jedoch (und mit ihm die gesamte Ableitung) ändert beim Durchlaufen dieses Wertes das Vorzeichen von Minus in Plus; die Funktion hat also hier ein Minimum. Schließlich behalten beim Durchlaufen des Wertes $x = 1$ nicht nur der erste und der dritte Faktor das Vorzeichen bei, sondern auch der zweite, da das Quadrat stets positiv ist. Hier liegt also kein Extremwert vor.

Nachdem man die Punkte x kennt, in denen unsere Funktion Extremwerte besitzt, lassen sich diese Werte selbst leicht berechnen: Für das Maximum erhält man $f(-2) = 0$ und für das Minimum $f(-0{,}8) \approx -8{,}40$. In Abb. 57 ist der Verlauf dieser Funktion graphisch veranschaulicht.[1])

2. Man bestimme die Extremwerte der Funktion

$$f(x) = \sin^3 x + \cos^3 x.$$

Da die Funktion die Periode 2π besitzt, genügt es, sich auf die x-Werte im Intervall $[0, 2\pi]$ zu beschränken. Die Ableitung dieser Funktion existiert überall:

$$f'(x) = 3\sin^2 x \cdot \cos x - 3\cos^2 x \cdot \sin x = 3\sin x \cdot \cos x \cdot (\sin x - \cos x).$$

Die Nullstellen der Ableitung (die stationären Punkte) sind

$$0, \frac{\pi}{4}, \frac{\pi}{2}, \pi, \frac{5\pi}{4}, \frac{3\pi}{2}, 2\pi.$$

Beim Durchgang (unter Beachtung der Periodizität) durch $x = 0$ ändert $\sin x$ sein Vorzeichen von Minus in Plus; die Ableitung insgesamt ändert das Vorzeichen von Plus in Minus, da die letzten beiden Faktoren in der Nähe des Punktes $x = 0$ negativ bleiben. Es liegt also ein Maximum vor. Der Faktor $\sin x - \cos x$ wird für $x = \frac{\pi}{4}$ gleich 0 und ändert in diesem Punkt sein Vorzeichen von Minus in Plus. Ebenso verhält sich die Ableitung, da die ersten beiden Faktoren dort positiv sind. Folglich liegt hier ein Minimum vor.

Analog untersucht man auch die übrigen stationären Punkte. Sie alle sind Punkte, in denen $f(x)$ ein Maximum bzw. Minimum besitzt; diese Extremwerte wechseln einander ab.

---

[1]) Hier und in den folgenden Beispielen illustrieren wir den Verlauf durch Kurven. Das Problem, wie man diese Kurven zweckmäßig konstruiert, wird in § 3, insbesondere in Nr. 149, Beispiel 3, ausführlich behandelt werden.

Wenn wir diese x-Werte in die Funktion einsetzen, erhalten wir die Maximal- und Minimalwerte selbst:

$$\text{Maxima:}\quad f(0) = f(2\pi) = 1,\quad f\left(\frac{\pi}{2}\right) = 1,\quad f\left(\frac{5\pi}{4}\right) = -\frac{\sqrt{2}}{2} \approx -0{,}71,$$

$$\text{Minima:}\quad f\left(\frac{\pi}{4}\right) = \frac{\sqrt{2}}{2} \approx 0{,}71,\quad f(\pi) = -1,\quad f\left(\frac{3\pi}{2}\right) = -1.$$

Abb. 58 zeigt den Funktionsverlauf (vgl. Nr. 147, Beispiel 1).

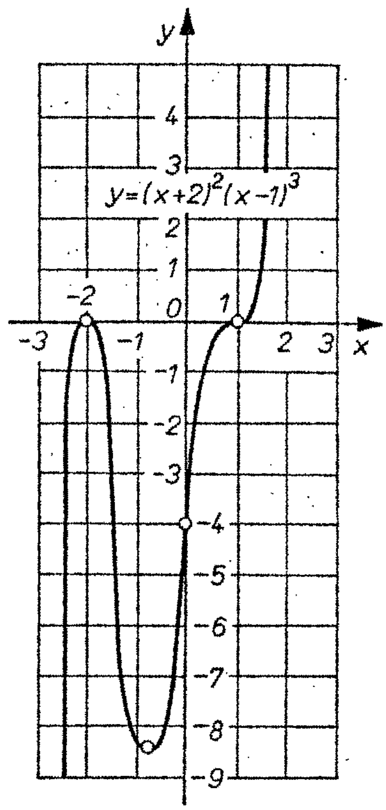

Abb. 57

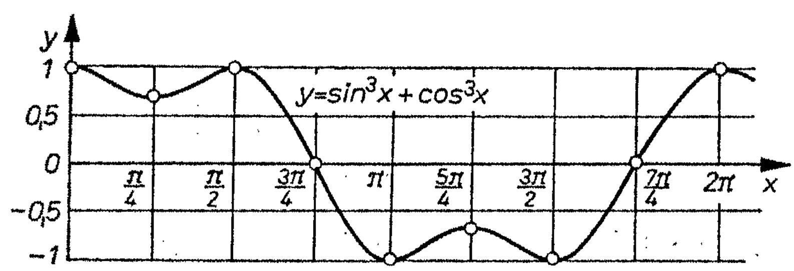

Abb. 58

3. Man bestimme die Extremwerte der (für negatives Argument wie üblich definierten) Funktion

$$f(x) = x^{2/3} - (x^2 - 1)^{1/3}.$$

Diesmal existiert die endliche Ableitung

$$f'(x) = \frac{2}{3}x^{-1/3} - \frac{1}{3}(x^2 - 1)^{-2/3} \cdot 2x = \frac{2}{3} \cdot \frac{(x^2 - 1)^{2/3} - x^{4/3}}{x^{1/3} \cdot (x^2 - 1)^{2/3}}$$

überall mit Ausnahme der Punkte $x = 0$ und $x = \pm 1$. Wenn sich x (von beiden Seiten) diesen Werten nähert, strebt die Ableitung gegen $\pm\infty$.

Die Nullstellen der Ableitung ermitteln wir, indem wir den Zähler gleich 0 setzen; wir finden $x = \pm\frac{1}{\sqrt{2}}$. Also sind „extremwertverdächtig" die Punkte

$$-1,\ -\frac{1}{\sqrt{2}},\ 0,\ \frac{1}{\sqrt{2}},\ +1.$$

Für $x = 0$ (und in der Nähe dieses Punktes) sind der Zähler und der zweite Faktor des Nenners positiv. Der Faktor $x^{1/3}$ im Nenner aber ändert sein Vorzeichen von Minus in Plus, die Ableitung also ebenfalls. Somit liegt hier ein Minimum vor. Für $x = \frac{1}{\sqrt{2}}$ (und in der Umgebung) bleibt der Nenner positiv. Den Zähler aber schreiben wir (da es sich um Werte von x in der Umgebung von $\frac{1}{\sqrt{2}}$ handelt) in der Form $(1 - x^2)^{2/3} - x^{4/3}$. Er wird für $x = \frac{1}{\sqrt{2}}$ gleich 0, mit abnehmendem x wächst er, und mit wachsendem x wird er kleiner, so daß sich das Vorzeichen von Plus in Minus ändert. Also liegt ein Maximum vor. Dasselbe gilt für $x = -\frac{1}{\sqrt{2}}$. Beim Durchgang durch $x = 1$ ändert der Faktor $(x^2 - 1)^{2/3}$ im Zähler, der in diesem Punkt gleich 0 ist, sein Vorzeichen nicht. Dies trifft auch für die Ableitung zu, so daß für $x = 1$ kein Extremwert vorliegt. Dasselbe gilt für $x = -1$.

Für die Maxima erhalten wir $f\left(\pm\frac{1}{\sqrt{2}}\right) = \sqrt[3]{4} \approx 1{,}59$, für das Minimum $f(0) = 1$. Den Funktionsverlauf zeigt Abb. 49 (vgl. Nr. 149, Beispiel 4).

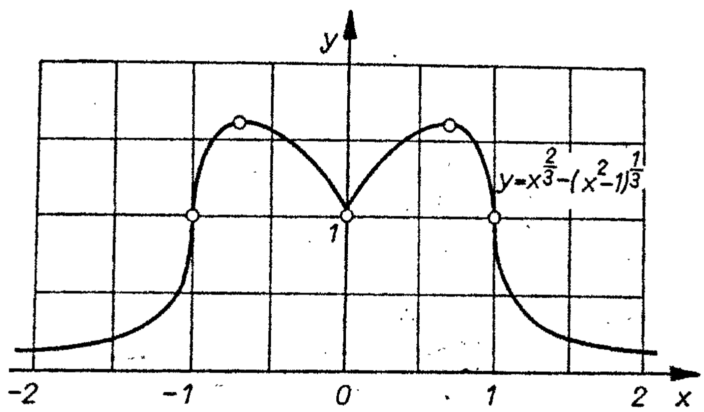

Abb. 59

4. *Gedämpfte Schwingungen.* Die Bewegung eines Punktes verlaufe nach dem Gesetz

$$s = Ae^{-kt}\sin\omega t,$$

wobei s der (von einer bestimmten Ausgangslage an gerechnete) zurückgelegte Weg und t die (vom Beginn der Bewegung an gerechnete) Zeit ist. Alle konstanten Größen, wie A, k, ω, sowie die Veränderliche t sehen wir als positiv an. Wir wollen uns die graphische Darstellung dieser Abhängigkeit klarmachen und sie mit der uns bereits bekannten Sinusschwingung $s = A\sin\omega t$ vergleichen.

Da $e^{-kt} > 0$ ist, schneiden offenbar beide Kurven die x-Achse in denselben Punkten $t = n\frac{\pi}{\omega}$ $(n = 1, 2, 3, \ldots)$. Wir bemerken, daß die Funktion $s = A\sin\omega t$ in den Punkten $t = \left(n + \frac{1}{2}\right)\frac{\pi}{\omega}$, in denen ihre Ableitung $s' = A\omega\cos\omega t$ gleich 0 wird, abwechselnd Maxima und Minima besitzt. Für die Ableitung der Funktion (vgl. Nr. 99, Beispiel 30) finden wir

$$\begin{aligned} s' &= Ae^{-kt}(\omega\cos\omega t - k\sin\omega t) \\ &= A\sqrt{\omega^2 + k^2}\,e^{-kt}\left(\frac{\omega}{\sqrt{\omega^2 + k^2}}\cos\omega t - \frac{k}{\sqrt{\omega^2 + k^2}}\sin\omega t\right). \end{aligned}$$

Durch

$$\frac{\omega}{\sqrt{\omega^2 + k^2}} = \cos\varphi, \qquad \frac{k}{\sqrt{\omega^2 + k^2}} = \sin\varphi$$

führen wir einen Hilfswinkel φ ein und können dann diese Ableitung folgendermaßen umformen:

$$s' = A\sqrt{\omega^2 + k^2}\,e^{-kt}\cos(\omega t + \varphi).$$

Sie wird in den Punkten

$$t = \left(n + \frac{1}{2}\right)\frac{\pi}{\omega} - \frac{\varphi}{\omega}$$

gleich 0; da der Kosinus beim Durchlaufen des Wertes 0 sein Vorzeichen ändert, ist leicht einzusehen, daß die Funktion in diesen Punkten tatsächlich Extrema hat, und zwar für gerades n Maxima und für ungerades n Minima. Im Vergleich zur Sinusschwingung ist eine *Verschiebung der Extrema um* $\frac{\varphi}{\omega}$ *nach links* eingetreten.

Man prüft leicht nach, daß alle Maxima positiv und alle Minima negativ sind. Bezeichnet man den n-ten Extremwert mit A_n, so gilt

$$\left|\frac{A_n}{A_{n+1}}\right| = \mathrm{e}^{\frac{k\pi}{\omega}},$$

so daß die Amplituden in geometrischer Progression abnehmen.

Die graphische Darstellung (für den einfachsten Spezialfall) zeigt Abb. 60. Eine Bewegung dieser Art heißt *gedämpfte Schwingung.*

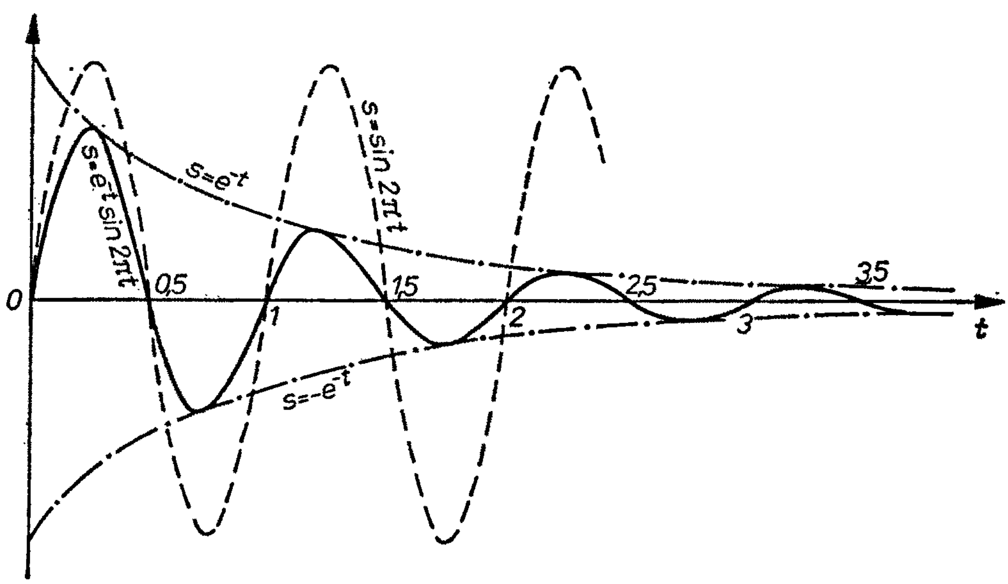

Abb. 60

Bemerkung. Für die meisten der in der Praxis vorkommenden Fälle erweist sich die in Nr. 135 aufgestellte Regel zur Bestimmung der Extrema als völlig ausreichend. Man muß sich jedoch darüber klar sein, daß es Fälle gibt, in denen sie nicht anwendbar ist, und zwar dann, wenn in beliebiger Nähe eines zu untersuchenden Punktes unendlich viele andere extremwertverdächtige Punkte liegen und die Ableitung auf der einen oder anderen Seite dieses Punktes kein festes Vorzeichen beibehält. Als Beispiel betrachten wir die durch

$$f(x) = x^2 \cdot \sin\frac{1}{x} \quad \text{für} \quad x \neq 0, \quad f(0) = 0$$

definierte Funktion. Wir wissen bereits, daß sie für $x = 0$ die Ableitung $f'(0) = 0$ besitzt (Nr. 102, 2). In beliebiger Nähe des stationären Punktes $x = 0$ ändert die Ableitung

$$f'(x) = 2x \cdot \sin\frac{1}{x} - \cos\frac{1}{x}$$

sowohl links auch als rechts unendlich oft ihr Vorzeichen. Hier ist im Punkt $x = 0$ kein Extremwert vorhanden. Die Funktion

$$f(x) = x^2\left(1 + \sin\frac{1}{x}\right) \quad \text{für} \quad x \neq 0, \quad f(0) = 0$$

besitzt zwar die gleiche Besonderheit, hat jedoch in $x = 0$ offenbar ein Minimum. In beiden Fällen ist die Regel nicht anwendbar.

137. Zweite Regel. Zur Bestimmung der Extremwerte kann die Untersuchung des Vorzeichens der Ableitung in der Nähe des „verdächtigen" Punktes durch die Untersuchung des Vorzeichens der zweiten Ableitung in diesem Punkt selbst ersetzt werden. Wir wollen dies beweisen.

Die Funktion $f(x)$ besitze also nicht nur in der Umgebung des Punktes x_0 eine Ableitung $f'(x_0)$, sondern in demselben Punkt x_0 auch eine zweite Ableitung $f''(x_0)$. Der Punkt x_0 sei stationär, d. h., es sei $f'(x_0) = 0$. Ist $f''(x_0) > 0$, so ist nach dem Lemma in Nr. 109 die Funktion $f'(x)$ im Punkt $x = x_0$ wachsend, d. h., links von x_0 gilt $f'(x) < f'(x_0) = 0$, rechts von x_0 gilt $f'(x) > f'(x_0) = 0$. Somit ändert die Ableitung $f'(x)$ ihr Vorzeichen von Minus in Plus; also hat $f(x)$ im Punkt $x = x_0$ ein Minimum. Ist $f''(x_0) < 0$, so fällt $f'(x)$ im Punkt $x = x_0$ und ändert das Vorzeichen von Plus in Minus, so daß also ein Maximum vorliegt.

Somit können wir eine *zweite Regel* für die Untersuchung eines „extremwertverdächtigen" Punktes x_0 formulieren: *Wir setzen x_0 in die zweite Ableitung $f''(x)$ ein. Ist $f''(x_0) > 0$, so besitzt $f(x)$ in x_0 ein Minimum, ist aber $f''(x_0) < 0$, so besitzt $f(x)$ in x_0 ein Maximum.*

Diese Regel hat eigentlich einen engeren Anwendungsbereich. Sie ist z. B. auf solche Punkte, für die keine endliche erste Ableitung existiert (denn dort kann erst recht keine zweite Ableitung existieren), offenbar nicht anwendbar. In denjenigen Fällen, in denen die zweite Ableitung gleich 0 ist, versagt die Regel ebenfalls. Die Lösung des Problems hängt dann vom Verhalten der höheren Ableitungen ab (vgl. Nr. 138).

Will man diese Regel auf das Beispiel 2 anwenden, so muß man die zweite Ableitung errechnen,

$$f''(x) = 6 \sin x \cos x\,(\cos x + \sin x) - 3(\sin^3 x + \cos^3 x).$$

Für $x = 0\ (2\pi), \frac{\pi}{2}, \pi, \frac{3\pi}{2}$ wird der erste Summand gleich 0, und das Vorzeichen von $f''(x)$ ist dem von $f(x) = \sin^3 x + \cos^3 x$ entgegengesetzt, und zwar negativ für $x = 0\ (2\pi), \frac{\pi}{2}$ (hier liegen Maxima vor) und positiv für $x = \pi$ und $x = \frac{3\pi}{2}$ (Minima). Für $x = \frac{\pi}{4}$ und $x = \frac{5\pi}{4}$ reduziert sich $f''(x)$ wegen $\sin x = \cos x$ auf den Ausdruck $6 \sin^3 x$, so daß für den ersten dieser Punkte die zweite Ableitung positiv (also liegt ein Minimum vor) und für den zweiten negativ (Maximum) ist.

Hierzu ein anderes Beispiel: Man bestimme die Extremwerte der Funktion

$$f(x) = \frac{x^2 - 5x + 6}{x^2 + 1}.$$

Die Ableitung $f'(x) = 5\,\frac{x^2 - 2x - 1}{(x^2 + 1)^2}$ wird zugleich mit dem Zähler 0. Ihre Nullstellen sind $x_1 = 1 - \sqrt{2} \approx -0{,}41$ und $x_2 = 1 + \sqrt{2} \approx 2{,}41$. Wir differenzieren die erste Ableitung noch einmal als Produkt,

$$\frac{5}{(x^2 + 1)^2} \cdot (x^2 - 2x - 1),$$

und erhalten

$$f''(x) = \frac{5}{(x^2 + 1)^2}\,(2x - 2) + \cdots,$$

wobei die Punkte ein Glied ersetzen, das den Faktor $x^2 - 2x - 1$ enthält und für uns unwesentlich ist, da es für die Werte x, die wir einsetzen, ohnehin gleich 0 ist. Man sieht leicht, daß $f''(z_1) < 0$ und $f''(x_2) > 0$ gilt; folglich liefert der Wert x_1 ein Maximum, $f(x_1) \approx 7{,}04$, während bei x_2 ein Minimum vorliegt, $f(x_2) \approx -0{,}03$.

Die Funktion ist in Abb. 61 graphisch dargestellt (vgl. Nr. 149, Beispiel 5).

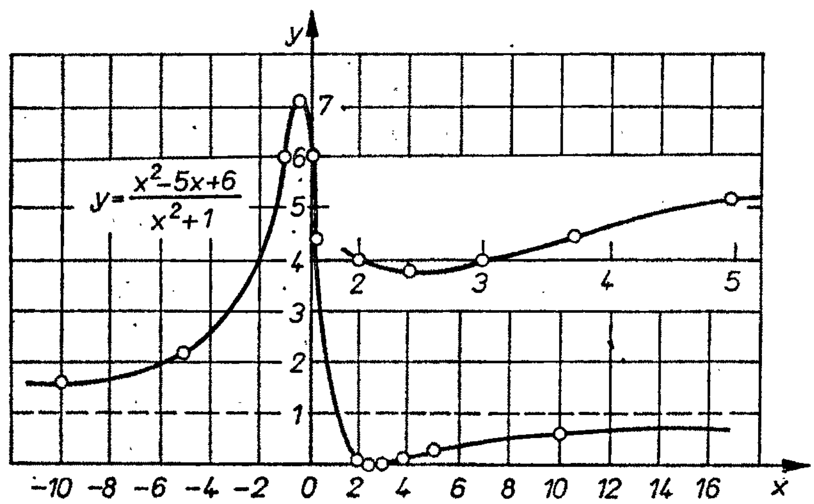

Abb. 61

Schließlich wollen wir noch ein geometrisches Problem betrachten: Man bestimme auf einer in der Ebene durch die Gleichung $y = f(x)$ gegebenen Kurve (K) denjenigen Punkt $M(x, y)$, für den der Abstand $r = \overline{PM}$ von einem gegebenen Punkt $P(\xi, \eta)$ ein Extremum ist (Abb. 62).

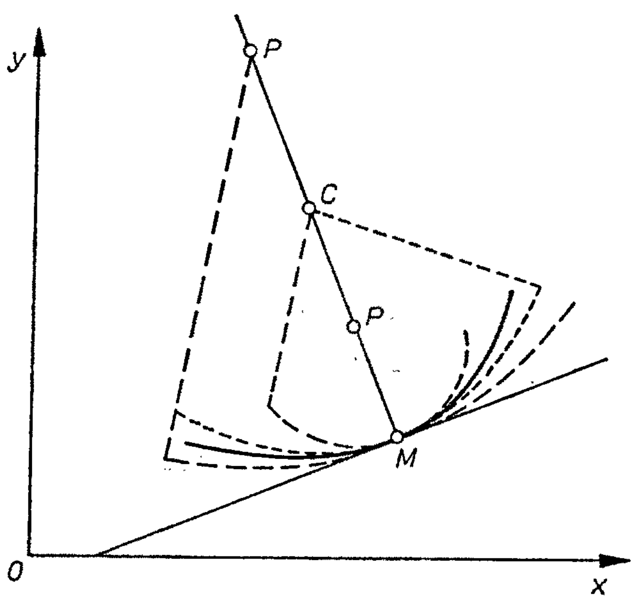

Abb. 62

An Stelle von r können wir die Funktion

$$u = \frac{1}{2} r^2 = \frac{1}{2} [(x - \xi)^2 + (y - \eta)^2]$$

betrachten, wobei $y = f(x)$ ist. Setzen wir die Ableitung

$$u_x = x - \xi + (y - \eta) \cdot y_x$$

gleich 0, so sehen wir, daß ein Punkt $M(x, y)$ auf der Kurve (K) den Abstand r nur dann zu einem Extremum macht, wenn die Bedingung

$$\xi - x + y_x(\eta - y) = 0$$

erfüllt ist. Mit anderen Worten: Der Punkt $P(\xi, \eta)$ muß auf der Geraden

$$X - x + y_x(Y - y) = 0$$

liegen, die durch den (gesuchten) Punkt $M(x, y)$ der Kurve senkrecht zur Tangente verläuft.[1]) Man nennt sie die *Normale* der Kurve.

Wir nehmen nun an, der Punkt $P(\xi, \eta)$ liege wirklich auf der durch den Punkt $M(x, y)$ verlaufenden Normalen der Kurve (K), und stellen die Frage, ob der Abstand $\overline{PM}$ ein Extremwert ist. Die Beantwortung hängt vom Vorzeichen der zweiten Ableitung

$$u_{xx} = 1 + y_x^2 + (y - \eta) \cdot y_{xx}$$

ab. Dieser Ausdruck wird nur im Punkt C mit den Koordinaten

$$\xi = x - y_x \cdot \frac{1 + y_x^2}{y_{xx}}, \quad \eta = y + \frac{1 + y_x^2}{y_{xx}}$$

gleich 0 ($y_{xx} \neq 0$ vorausgesetzt). Das Problem bleibt also für diesen Punkt noch offen. Der Punkt C trennt auf der Normalen diejenigen Punkte P, für die $u'' < 0$ und der Abstand $\overline{PM}$ ein Maximum ist, von den Punkten P, für welche $u'' = 0$ und der Abstand $\overline{PM}$ ein Minimum ist.

Später, in Nr. 243 und Nr. 253, werden wir sehen, daß dieser Grenzpunkt C auf der Normalen in vieler Hinsicht bemerkenswert ist.

138. Die Benutzung höherer Ableitungen. Wie wir sahen, besitzt die Funktion $f(x)$ im Punkt x_0 ein Minimum, wenn $f'(x_0) = 0$ und $f''(x_0) > 0$ ist; dagegen besitzt sie ein Maximum, wenn $f'(x_0) = 0$ und $f''(x_0) < 0$ ist. Der Fall, daß sowohl $f'(x_0) = 0$ als auch $f''(x_0) = 0$ sind, blieb unberücksichtigt.

Wir setzen nun voraus, die Funktion $f(x)$ besitze im Punkt $x = x_0$ Ableitungen bis zur Ordnung n, die bis zur $(n - 1)$-ten einschließlich dort gleich 0 sind,

$$f'(x_0) = f''(x_0) = \cdots = f^{(n-1)}(x_0) = 0,$$

während $f^{(n)}(x_0) \neq 0$ ist. Wir entwickeln den Zuwachs $f(x) - f(x_0)$ der Funktion nach der Taylorschen Formel nach Potenzen der Differenz $x - x_0$ und setzen das Restglied in der Peanoschen Form an [Nr. 124, Formel (10a)]. Da alle k-ten Ableitungen für $k < n$ im Punkt x_0 gleich 0 sind, erhalten wir

$$f(x) - f(x_0) = \frac{f^{(n)}(x_0) + \alpha}{n!}(x - x_0)^n.$$

Da $\alpha \to 0$ für $x \to x_0$ gilt, stimmt für genügend nahe bei x_0 gelegenes x das Vorzeichen von $f^{(n)}(x_0) + \alpha$ mit dem Vorzeichen von $f^{(n)}(x_0)$ überein, und zwar sowohl für $x < x_0$ als auch für $x > x_0$. Wir betrachten zwei Fälle:

1. Die Zahl n sei ungerade: $n = 2k + 1$. Beim Übergang von Werten, die kleiner als x_0 sind, zu solchen, die größer als x_0 sind, kehrt $(x - x_0)^n$ sein Vorzeichen um. Da sich aber das Vorzeichen des ersten Faktors hierbei nicht ändert, wechselt das Vorzeichen der Differenz $f(x) - f(x_0)$. Somit kann $f(x)$ im Punkt x_0 keinen Extremwert besitzen, da in der Nähe dieses Punktes sowohl Werte, die kleiner als $f(x_0)$ sind, als auch solche, die größer als $f(x_0)$ sind, angenommen werden.

2. Die Zahl n sei gerade; $n = 2\,k$. In diesem Fall ändert $f(x) - f(x_0)$ beim Übergang von x-Werten, die kleiner als x_0 sind, zu x-Werten, die größer als x_0 sind, ihr

---

[1]) Ihr Richtungskoeffizient $-\frac{1}{y_x}$ ist der mit dem entgegengesetzten Vorzeichen versehene reziproke Wert des Richtungskoeffizienten y_x der Tagente.

Vorzeichen nicht; denn dann ist $(x - x_0)^n > 0$ für alle x. Offenbar stimmt sowohl in der linksseitigen als auch in der rechtsseitigen Umgebung des Punktes x_0 das Vorzeichen von $f(x) - f(x_0)$ mit dem Vorzeichen von $f^{(n)}(x_0)$ überein. Wenn also $f^{(n)}(x_0) > 0$ gilt, dann ist $f(x) > f(x_0)$ in der Umgebung von x_0, und $f(x)$ besitzt in x_0 ein Minimum (im engeren Sinne). Gilt aber $f^{(n)}(x_0) < 0$, so besitzt die Funktion ein Maximum (im engeren Sinne).

Hieraus erhalten wir folgende *Regel*:

Ist die erste der im Punkt x_0 nicht verschwindenden Ableitungen von ungerader Ordnung, so besitzt die Funktion im Punkt x_0 weder ein Maximum noch ein Minimum. Ist diese Ableitung aber von gerader Ordnung, so hat die Funktion im Punkt x_0 ein Maximum bzw. Minimum, je nachdem, ob diese Ableitung negativ bzw. positiv ist.

Beispielsweise ist der Punkt $x = 0$ für die Funktion $f(x) = e^x + e^{-x} + 2\cos x$ stationär, da in diesem Punkt

$$f'(x) = e^x - e^{-x} - 2\sin x$$

gleich 0 ist.

Ferner ist

$$\begin{aligned} f''(x) &= e^x + e^{-x} - 2\cos x, & f''(0) &= 0; \\ f'''(x) &= e^x - e^{-x} + 2\sin x, & f'''(0) &= 0; \\ f^{(4)}(x) &= e^x + e^{-x} + 2\cos x, & f^{(4)}(0) &= 4. \end{aligned}$$

Da die erste Ableitung, welche nicht verschwindet, von gerader Ordnung ist, liegt ein Extremwert vor, und zwar ein Minimum; denn es gilt $f^{(4)}(0) > 0$.

Bemerkung. Obgleich das oben hergeleitete Kriterium das Problem der Extremwerte in einer sehr umfassenden Klasse von Fällen löst, ist es — wenigstens theoretisch — nicht allumfassend. Eine Funktion, die nicht identisch konstant ist, kann nämlich in einer Umgebung des zu untersuchenden Punktes Ableitungen beliebiger Ordnungen besitzen, die in diesem Punkt sämtlich verschwinden.

Als Beispiel betrachten wir (nach Cauchy) die Funktion

$$f(x) = e^{-1/x^2} \quad \text{für} \quad x \neq 0, \qquad f(0) = 0.$$

Für $x \neq 0$ besitzt sie Ableitungen aller Ordnungen:

$$f'(x) = \frac{2}{x^3} e^{-1/x^2}, \qquad f''(x) = \left(-\frac{6}{x^4} + \frac{4}{x^6}\right) e^{-1/x^2}, \ldots$$

und allgemein

$$f^{(n)}(x) = P_n\left(\frac{1}{x}\right) \cdot e^{-1/x^2} \qquad (n = 1, 2, 3, \ldots), \tag{5}$$

wobei $P_n(z)$ ein Polynom $3n$-ten Grades ist. Das kann man leicht induktiv nachweisen.

Wir wollen nun zeigen, daß die Funktion auch im Punkt $x = 0$ Ableitungen aller Ordnungen besitzt und daß diese sämtlich gleich 0 sind. In der Tat ist zunächst

$$\frac{f(x) - f(0)}{x} = \frac{\frac{1}{x}}{e^{1/x^2}} \to 0 \quad \text{für} \quad x \to 0,$$

also $f'(0) = 0$. Das folgt daraus, daß e^z für $z \to \infty$ von höherer Ordnung unendlich groß wird als jede Potenz z^k, d. h., daß

$$\lim_{z \to \infty} \frac{z^k}{e^z} = 0$$

ist (vgl. Nr. 65). Hier ist $z = \frac{1}{x^2}$ (für $x \to 0$).

Wir nehmen nun an, die zu beweisende Behauptung sei für alle Ableitungen bis zur n-ten einschließlich richtig. Dann gilt [vgl. (5)]

$$\frac{f^{(n)}(x) - f^{(n)}(0)}{x} = \frac{\frac{1}{x} P_n\left(\frac{1}{x}\right)}{e^{1/x^2}} \to 0 \quad \text{für} \quad x \to 0,$$

da der Zähler selbst eine Summe von Gliedern der Form $\frac{c}{x^m}$ ist. Somit ist auch $f^{(n+1)}(0) = 0$. Damit ist die Behauptung durch Induktion bewiesen.

Obgleich unmittelbar einleuchtend ist, daß die gegebene Funktion für $x = 0$ ein Minimum besitzt, kann man diese Tatsache nicht mit Hilfe der obigen Regel finden.

139. Das Aufsuchen größter und kleinster Werte. Die Funktion $f(x)$ sei in einem endlichen abgeschlossenen Intervall $[a, b]$ definiert und stetig. Bisher interessierten wir uns nur für ihre (relativen) Maxima und Minima, jetzt aber stellen wir die Frage, wie man ihren (absolut) größten und ihren (absolut) kleinsten Wert findet, den sie in diesem Intervall annimmt.[1] Nach dem zweiten Satz von WEIERSTRASS (Nr. 85) existieren solche größten bzw. kleinsten Werte. Wir beschränken uns auf die Betrachtung des größten Wertes. Wird er in einem Punkt zwischen a und b angenommen, so ist er gleichzeitig eines der Maxima (offenbar das größte). Der größte Wert kann jedoch auch in einem der Endpunkte des Intervalls, in a oder in b, angenommen werden (Abb. 63). Daher muß man alle Maxima der Funktion $f(x)$ und ihre Werte in den Endpunkten des Intervalls, $f(a)$ und $f(b)$, zum Vergleich heranziehen. Die größte dieser Zahlen ist dann der größte Wert von $f(x)$ in $[a, b]$.

Analog findet man auch den kleinsten Wert von $f(x)$.

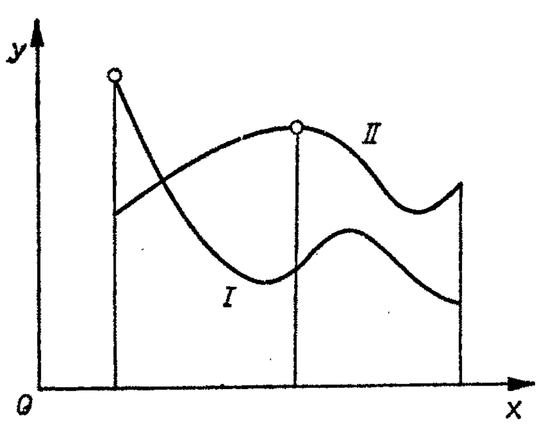

Abb. 63

Es seien z. B. der größte und der kleinste Wert der Funktion $f(x) = \sin^3 x + \cos^3 x$ im Intervall $\left[-\frac{\pi}{4}, \frac{3\pi}{4}\right]$ zu bestimmen. Die beiden Maxima, die gleich 1 sind, sind größer als die Werte der Funktion an den Intervallenden; denn es ist $f\left(-\frac{\pi}{4}\right) = f\left(\frac{3\pi}{4}\right) = 0$. Folglich ist 1 der größte Wert der Funktion in diesem Intervall. Das Minimum, das gleich 0,7 ... ist, ist größer als die Funktionswerte in den Intervallenden, so daß der kleinste Wert 0 ist.

[1]) Wir unterscheiden also zwischen Maxima bzw. Minima einerseits und größten bzw. kleinsten Werten andererseits. Maximum und Minimum sind größter bzw. kleinster Wert in einer gewissen Umgebung (daher sprechen wir auch von *relativen* Maxima bzw. Minima), werden also im „lokalen" Sinne gebraucht. Zieht man alle Werte, die die Funktion im Intervall annimmt, zum Vergleich heran, so spricht man vom größten bzw. kleinsten Wert, gelegentlich auch vom *absoluten* Maximum bzw. Minimum.

Für das Intervall $\left[\frac{\pi}{4}, \frac{3\pi}{2}\right]$ hätte man als größten Wert den größeren der beiden Maxima 1 und $-0{,}7\ldots$ zu wählen, die bei $x = \frac{\pi}{2}$ und $x = \frac{5\pi}{4}$ angenommen werden, da an den Intervallenden die Werte $f\left(\frac{\pi}{4}\right) = 0{,}7\ldots$ und $f\left(\frac{3\pi}{2}\right) = -1$ angenommen werden, die kleiner als 1 sind.

Der kleinste Wert liegt im rechten Endpunkt des Intervalls und stimmt mit dem Minimum in $x = \pi$ überein.

Will man auf die Untersuchung der Maxima und Minima verzichten und nur die absoluten Extrema finden, so kann man anders vorgehen. Man braucht nur die Funktionswerte in allen „extremwertverdächtigen" Punkten zu berechnen und mit den Funktionswerten $f(a)$ und $f(b)$ in den Intervallenden zu vergleichen. Die größte bzw. die kleinste dieser Zahlen ist offenbar genau der größte bzw. der kleinste aller Funktionswerte.

Beispielsweise vergleichen wir für das Intervall $\left[-\frac{\pi}{4}, \frac{3\pi}{4}\right]$ die Werte $f(0) = 1$, $f\left(\frac{\pi}{4}\right) = 0{,}7\ldots$, $f\left(\frac{\pi}{2}\right) = 1$ mit den Werten in den Intervallenden $f\left(-\frac{\pi}{4}\right) = f\left(\frac{3\pi}{4}\right) = 0$, und für das Intervall $\left[\frac{\pi}{4}, \frac{3\pi}{2}\right]$ vergleichen wir die Zahlen $f\left(\frac{\pi}{2}\right) = 1$, $f(\pi) = -1$, $f\left(\frac{5\pi}{4}\right) = -0{,}7\ldots$ mit den Werten in den Intervallenden (den „Randwerten") $f\left(\frac{\pi}{4}\right) = 0{,}7\ldots$ und $f\left(\frac{3\pi}{2}\right) = -1$.

Bemerkung. Bei angewandten Aufgaben tritt sehr oft der einfache Fall ein, daß zwischen a und b nur ein einziger „extremwertverdächtiger" Punkt x_0 liegt. Besitzt die Funktion in diesem Punkt ein Maximum (bzw. Minimum), so ist ohne Vergleich mit den Randwerten klar, daß dies auch der größte (bzw. kleinste) Wert der Funktion im Intervall ist (vgl. Abb. 64). In solchen Fällen ist es oft einfacher festzustellen, ob es sich um ein Maximum oder Minimum handelt, als die speziellen Werte der Funktion zu berechnen und zu vergleichen (insbesondere, wenn in dem Funktionsausdruck willkürliche Konstanten auftreten).

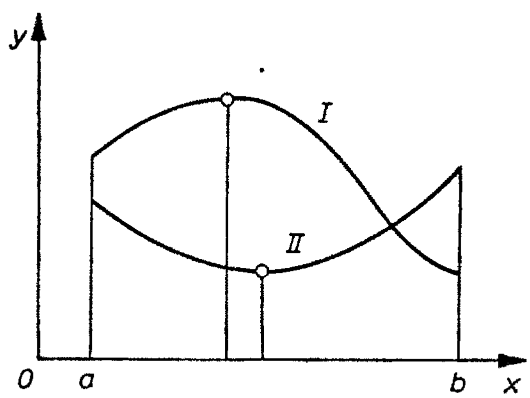

Abb. 64

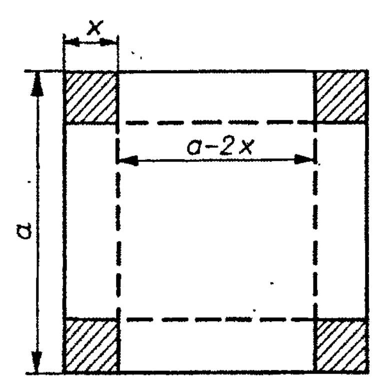

Abb. 65

Es sei noch betont, daß diese Ausführungen ohne Einschränkung sowohl für offene Intervalle (a, b) als auch für unendliche Intervalle gültig bleiben (an den Endpunkten hat man dann die obere bzw. untere Grenze der Funktionswerte zu betrachten).

140. Aufgaben. Wir wollen nun an Hand von Beispielen eine Reihe von Aufgaben aus verschiedenen Gebieten studieren, deren Lösung sich auf die Ermittlung des größten oder kleinsten Wertes einer Funktion zurückführen läßt. Im übrigen sind häufig diese Werte an sich nicht so interessant wie die Punkte (die Werte des Arguments), in denen sie angenommen werden.

1. Aus einem quadratischen Stück Blech mit der Seite a wird eine offene rechteckige Schachtel hergestellt, indem an den Ecken gleiche Quadrate herausgeschnitten werden und der Rand umgebogen wird (Abb. 65). Wie erhält man eine Schachtel mit größtem Rauminhalt?

Bezeichnen wir die Seite des herauszuschneidenden Quadrates mit x, so ergibt sich das

Volumen y der Schachtel zu $y = x(a - 2x)^2$, wobei x im Intervall $\left[0, \frac{a}{2}\right]$ variieren kann. Das Problem ist damit auf die Bestimmung des größten Wertes der Funktion y in diesem Intervall zurückgeführt.

Da die Ableitung $y' = (a - 2x)(a - 6x)$ zwischen 0 und $\frac{a}{2}$ eine einzige Nullstelle $x = \frac{a}{6}$ besitzt, ist klar, daß dieser Wert, wenn er ein Maximum der Funktion liefert, zugleich auch den größten Wert liefert. Anders ausgedrückt: Für $x = \frac{a}{6}$ erhalten wir $y = \frac{2a^3}{27}$, während die Randwerte der Funktion 0 sind. Folglich erhält man tatsächlich für $x = \frac{a}{6}$ den größten Wert von y.

2. Gegeben sei ein Baumstamm von kreisförmigem Querschnitt mit dem Durchmesser d. Er soll so bearbeitet werden, daß man einen Balken größter Tragfähigkeit mit rechteckigem Querschnitt erhält.

Hinweis. In der Festigkeitslehre wird gezeigt, daß die Tragfähigkeit eines rechteckigen Balkens dem Produkt bh^2 proportional ist, wobei b die Grundseite und h die Höhe des Querschnittes ist.

Wegen $h^2 = d^2 - b^2$ handelt es sich also um die Bestimmung des größten Wertes des Ausdrucks $y = bh^2 = b(d^2 - b^2)$, wobei die „unabhängige Veränderliche" b im Intervall $(0, d)$ variiert.

Die Ableitung $y' = d^2 - 3b^2$ wird nur einmal innerhalb dieses Intervalls gleich 0, nämlich im Punkt $b = \frac{d}{\sqrt{3}}$. Die zweite Ableitung $y'' = -6b$ ist negativ; folglich liegt in diesem Punkt ein Maximum vor, das zugleich auch der größte Wert ist.

Für $b = \frac{d}{\sqrt{3}}$ ist $h = d\sqrt{\frac{2}{3}}$, so daß $d:h:b = \sqrt{3}:\sqrt{2}:1$ gilt. Aus Abb. 66 ist ersichtlich, wie man das gesuchte Rechteck konstruiert (der Durchmesser wird in drei gleiche Teile geteilt, und in den Teilungspunkten werden die Senkrechten errichtet). Im Bauwesen wird gewöhnlich die Beziehung $h:b = 7:5 = 1{,}4$ vorgeschrieben. Das ist ein Näherungswert für $\sqrt{2}$.

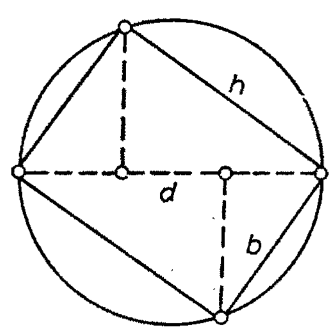

Abb. 66

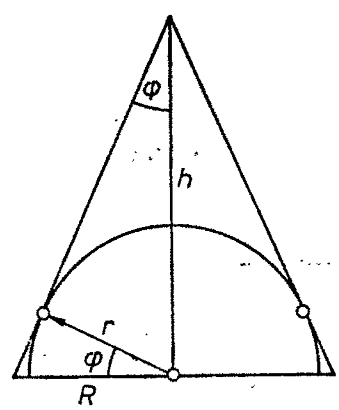

Abb. 67

3. Um eine Halbkugel mit dem Radius r ist ein gerader Kreiskegel kleinsten Volumens zu beschreiben. Dabei wird angenommen, daß die Grundflächen von Halbkugel und Kegel konzentrisch in einer Ebene liegen (Abb. 67). Es empfiehlt sich, eine zweckmäßige unabhängige Veränderliche zu wählen. Als solche nimmt man den Winkel φ an der Spitze des Kegels. In den Bezeichnungen der Abb. 67 ist $R = \frac{r}{\cos\varphi}$, $h = \frac{r}{\sin\varphi}$, so daß für das Volumen die Beziehung

$$v = \frac{1}{3}\pi R^2 h = \frac{\frac{1}{3}\pi r^3}{\cos^2\varphi \cdot \sin\varphi}$$

gilt. Damit das Volumen v seinen kleinsten Wert annimmt, ist offenbar notwendig, daß der Ausdruck $y = \cos^2\varphi \sin\varphi$ im Nenner seinen größten Wert erhält (da der Zähler nicht von φ abhängt), wenn φ im Intervall $\left(0, \frac{\pi}{2}\right)$ variiert. Nun gilt

$$y_\varphi = -2\cos\varphi \cdot \sin^2\varphi + \cos^3\varphi = 2\cos^3\varphi\left(\frac{1}{2} - \tan^2\varphi\right);$$

zwischen 0 und $\frac{\pi}{2}$ wird dieser Ausdruck nur für $\tan\varphi = \frac{1}{\sqrt{2}}$ gleich 0, also für $\varphi = \arctan\frac{1}{\sqrt{2}}$ (d. h. $\varphi = 35°\,15'\,52''$). Beim Durchgang durch diesen Wert ändert sich das Vorzeichen von Plus in Minus. Dieser Winkel liefert für y den größten und für das Volumen v den kleinsten Wert.

4. Eine Last vom Gewicht G, die auf einer horizontalen Ebene liegt, soll durch Einwirkung einer Kraft verschoben werden (Abb. 68). Unter welchem Winkel zur Horizontalen muß (unter Berücksichtigung der Reibung) diese Kraft F angreifen, wenn sie möglichst klein sein soll? Der Reibungskoeffizient μ ist vorgegeben.

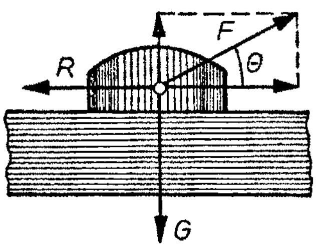

Abb. 68

Hinweis. Die Reibung R wird proportional zu dem Druck angenommen, den der Körper auf die Ebene ausübt (Coulombsches Gesetz), und ist der Bewegung entgegengerichtet. Der Proportionalitätsfaktor μ ist genau der *Reibungskoeffizient*.

Wir bestimmen die Kraft F, die einem gegebenen Winkel θ entspricht. Zerlegen wir sie in eine Horizontal- und eine Vertikalkomponente, so erhalten wir für diese Komponenten die Größen $F \cdot \cos\theta$ und $F \cdot \sin\theta$. Der von dem Körper auf die Ebene ausgeübte Druck ist $G - F \cdot \sin\theta$, so daß nach dem Coulombschen Gesetz $R = \mu(G - F \cdot \sin\theta)$ gilt.

Die Horizontalkomponente $F \cdot \cos\theta$ der Schubkraft F muß diese Reibung gerade kompensieren,

$$F \cdot \cos\theta = \mu \cdot (G - F \cdot \sin\theta),$$

woraus sich

$$F = \frac{\mu G}{\cos\theta + \mu\sin\theta}$$

ergibt. Es handelt sich also um die Bestimmung des kleinsten Wertes dieser Funktion oder, da der Zähler konstant ist, des größten Wertes der Funktion $y = \cos\theta + \mu\sin\theta$, wobei θ das Intervall $\left[0, \frac{\pi}{2}\right]$ durchläuft. Die Ableitung $y_\theta = \mu \cdot \cos\theta - \sin\theta$ wird gleich 0, wenn $\tan\theta = \mu$ oder $\theta = \arctan\mu$ ist. Dieser Winkel θ heißt *Reibungswinkel*. Da $y_{\theta\theta} = -\mu\sin\theta - \cos\theta < 0$ gilt, ist es am zweckmäßigsten, die Kraft unter dem Reibungswinkel angreifen zu lassen. Ist z. B. ein Stein auf einem Holzbrett zu verschieben, so ist $\mu = 0{,}4$, und es ergibt sich $\theta \approx 22°$.

5. Die Kosten für eine Schiffsreise auf einem Fluß pro Stunde errechnen sich nach der empirischen Formel $a + bv^3$, wobei a und b für das betreffende Schiff festgesetzte Konstanten sind und v die Geschwindigkeit des Schiffes in Knoten (1 Knoten = 1,85 km/h) ist.[1] Bei welcher

[1]) In dieser Formel bezieht sich der konstante Teil a des Aufwandes auf die Amortisation und die Kosten für die Besatzung, das zweite Glied $b\,v^3$ auf den Preis für den Kraftstoff.

(„ökonomischen“) Geschwindigkeit legt das Schiff eine beliebige Entfernung mit geringstem Aufwand zurück?

Für 1 km werden $\frac{1}{1{,}85v}$ Stunden benötigt. Der entsprechende Aufwand wird durch die Formel

$$\frac{1}{1{,}85v}(a + bv^3) = \frac{1}{1{,}85}\left(bv^2 + \frac{a}{v}\right)$$

ausgedrückt.

Setzen wir die Ableitung des Ausdrucks $y = bv^2 + \frac{a}{v}$ gleich 0, so erhalten wir $y_v = 2bv - \frac{a}{v^2} = 0$, also $v = \sqrt[3]{\frac{a}{2b}}$. Da $y_{vv} = 2b + \frac{2a}{v^3}$ positiv ist, liefert dieser Wert der Geschwindigkeit den geringsten Aufwand.

Zahlenbeispiel: $a = 40$, $b = 0{,}01$, $v = \sqrt[3]{2000} \approx 12{,}6$ (Knoten).

6. Eine Lichtquelle möge in vertikaler Richtung auf der Geraden OB verschiebbar sein (Abb. 69). Welchen Abstand von der Horizontalen OA muß man wählen, damit im Punkt A die Beleuchtungsstärke maximal ist?

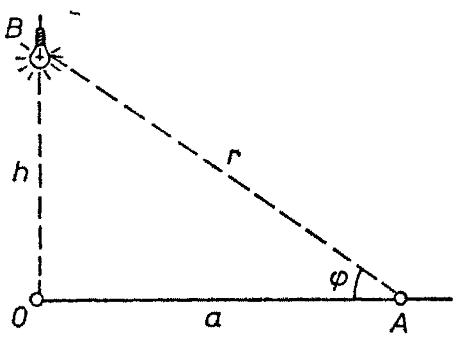

Abb. 69

Hinweis. Die Beleuchtungsstärke J ist proportional $\sin\varphi$ und umgekehrt proportional dem Quadrat der Entfernung $r = \overline{AB}$, d. h., es gilt

$$J = c \cdot \frac{\sin\varphi}{r^2},$$

wobei c von der Lichtstärke der Lichtquelle abhängt.

Wählen wir $h = \overline{OB}$ als unabhängige Veränderliche, so erhalten wir

$$\sin\varphi = \frac{h}{r}, \qquad r = \sqrt{h^2 + a^2}$$

und

$$J = c \cdot \frac{h}{(h^2 + a^2)^{3/2}} \qquad (0 < h < \infty).$$

Ferner verschwindet die Ableitung

$$J_h = c \cdot \frac{a^2 - 2h^2}{(h^2 + a^2)^{5/2}}$$

für $h = \frac{a}{\sqrt{2}} \approx 0{,}7a$ und ändert ihr Vorzeichen beim Durchlaufen dieses Wertes von Plus in Minus. Dieser Wert von h ist also auch die günstigste Entfernung.

Als unabhängige Veränderliche kann man auch den Winkel φ wählen. Dann gilt

$$r = \frac{a}{\cos\varphi}, \qquad J = \frac{c}{a^2} \cdot \cos^2\varphi \sin\varphi,$$

und das Problem läuft auf die Bestimmung des größten Wertes der Funktion $y = \cos^2 \varphi \sin \varphi$ im Intervall $\left(0, \frac{\pi}{2}\right)$ hinaus. Wir wissen aber bereits (vgl. Aufgabe 3), daß dieser größte Wert für einen Winkel φ_0 angenommen wird, für den $\tan \varphi_0 = \frac{1}{\sqrt{2}}$ gilt. Für den Abstand h erhalten wir wieder den Wert $a \cdot \tan \varphi_0 = \frac{a}{\sqrt{2}}$.

7. Von einem Punkt A an der Hauptbahnlinie AB (Abb. 70) sollen dauernd Frachten zum Punkt C gebracht werden, der sich in einer Entfernung $CB = l$ von der Eisenbahnlinie befindet. Die Kosten der Beförderung einer Gewichtseinheit pro Wegeinheit seien α bei Beförderung durch Bahn und β bei Beförderung auf Straßen. Von welchem Punkt M muß die Straße MC abgezweigt werden, wenn die Kosten für den Transport von A über M nach C möglichst niedrig sein sollen?

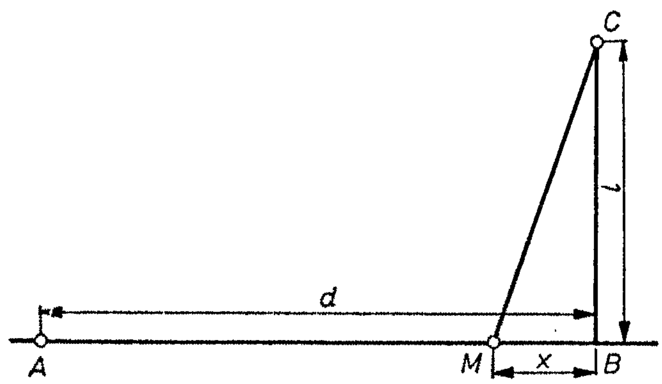

Abb. 70

In den Bezeichnungen der Abb. 70 ergeben sich die Kosten pro Gewichtseinheit bei willkürlicher Festlegung des Punktes M zu

$$y = \alpha(d - x) + \beta \sqrt{x^2 + l^2} \qquad (0 \leqq x \leqq d).$$

Hieraus folgt

$$y_x = \frac{\beta x}{\sqrt{x^2 + l^2}} - \alpha = \beta \left(\frac{x}{\sqrt{x^2 + l^2}} - k\right) \quad \left(k = \frac{\alpha}{\beta}\right).$$

Für $k \geqq 1$ ($\alpha \geqq \beta$) bleibt dieser Ausdruck negativ. Die Funktion y fällt dann mit wachsendem x von 0 bis d und nimmt ihren kleinsten Wert offenbar für $x = d$ an. In diesem Fall ist es natürlich am günstigsten, die Straße direkt im Punkt A beginnen zu lassen.

Dasselbe trifft auch für den Fall $k < 1$ zu, wenn gleichzeitig

$$\frac{kl}{\sqrt{1 - k^2}} \geqq d$$

gilt. Für $k < 1$ hat nämlich der Ausdruck $\frac{x}{\sqrt{x^2 + l^2}} - k$ die einzige Nullstelle

$$x = \frac{kl}{\sqrt{1 - k^2}}.$$

Diese Nullstelle liegt jedoch außerhalb des Variationsintervalls von x (oder in einem Endpunkt), so daß im Innern die Ableitung y_x negativ ist.

Lediglich dann, wenn die erwähnte Nullstelle $< d$ ist, bestimmt dieser x-Wert die Lage eines Punktes M zwischen A und B, für den die Transportkosten am geringsten sind.

Bemerkung. Wir nehmen diesen Fall zum Anlaß, den Leser auf folgenden Sachverhalt aufmerksam zu machen. Bei der Bestimmung des größten bzw. des kleinsten Wertes einer Funktion in einem bestimmten Intervall kann es leicht vorkommen, daß im Innern dieses Intervalls überhaupt keine Nullstellen der Ableitung (oder andere „extremwertverdächtige" Werte) liegen. Das beruht darauf, daß die Funktion in dem betrachteten Intervall monoton wachsend oder fallend ist, also sowohl ihren größten als auch ihren kleinsten Wert in den Endpunkten des Intervalls annimmt.

In der letzten Aufgabe liegt bei bestimmten Verhältnissen zwischen den Ausgangsgrößen gerade dieser Sachverhalt vor.

§ 2. Konvexe (und konkave) Funktionen

141. Definition einer konvexen (konkaven) Funktion. Neben der Klasse der monotonen (wachsenden bzw. fallenden) Funktionen ist die Klasse der sogenannten konvexen bzw. konkaven Funktionen von Bedeutung. Eine in einem (abgeschlossenen, halboffenen, offenen, endlichen oder unendlichen) Intervall $\mathscr{X}$ definierte und stetige Funktion $f(x)$ heißt *konvex* (*von unten konvex*), wenn für beliebige Punkte x_1 und x_2 aus $\mathscr{X}(x_1 \lessgtr x_2)$ die Ungleichung

$$f(q_1x_1 + q_2x_2) \leqq q_1 \cdot f(x_1) + q_2 \cdot f(x_2) \tag{1}$$

erfüllt ist, wie die positiven Zahlen q_1 und $q_2 = 1 - q_1$ auch gewählt sein mögen. Eine Funktion heißt *konkav* (*von oben konvex*), wenn an Stelle von (1) die Beziehung

$$f(q_1x_1 + q_2x_2) \geqq q_1 \cdot f(x_1) + q_2 \cdot f(x_2) \tag{1a}$$

gilt.[1])

Ist $f(x)$ konvex (bzw. konkav), so ist offenbar die Funktion $-f(x)$ konkav (bzw. konvex) und umgekehrt. Diese einfache Bemerkung ermöglicht uns in vielen Fällen, die Untersuchung auf die konvexen Funktionen zu beschränken.

Die obige Definition der Konvexität einer Funktion hat eine einfache geometrische Bedeutung. Zunächst bemerken wir, daß

$$x = q_1x_1 + q_2x_2 \quad (x_1 < x_2) \tag{2}$$

unter unseren Bedingungen für q_1 und q_2 zwischen x_1 und x_2 liegt; umgekehrt kann jede Zahl x zwischen x_1 und x_2 eindeutig in dieser Form dargestellt werden. Dabei gilt für die Koeffizienten

$$q_1 = \frac{x_2 - x}{x_2 - x_1}, \quad q_2 = \frac{x - x_1}{x_2 - x_1}. \tag{2a}$$

Betrachten wir (vgl. Abb. 71) die Kurve der Funktion $f(x)$ und das Bogenstück zwischen den Punkten $A_1(x_1, y_1)$ und $A_2(x_2, y_2)$, wobei $y_1 = f(x_1)$, $y_2 = f(x_2)$ ist, so finden wir auf der linken Seite der Ungleichung (1) mit den Koeffizienten (2a) die Ordinate des Punktes A auf dem Bogen A_1A_2 mit der Abszisse x. Auf ihrer rechten Seite dagegen

---

[1]) Der Begriff der konvexen (bzw. konkaven) Funktion wurde von J. L. W. V. Jensen (1859 bis 1925, dänischer Ingenieur und Mathematiker) eingeführt, der jedoch von einer spezielleren Beziehung als (1) bzw. (1a) ausging, nämlich von

$$f\left(\frac{x_1 + x_2}{2}\right) \underset{(\geqq)}{\leqq} \frac{f(x_1) + f(x_2)}{2}.$$

Sie entspricht den Werten $q_1 = q_2 = \frac{1}{2}$. Im Fall stetiger Funktionen, auf den wir uns beschränken, ist diese Definition der obigen gleichwertig.

steht die Ordinate des Punktes B der Sehne A_1A_2,

$$y = \frac{x_2 - x}{x_2 - x_1} y_1 + \frac{x - x_1}{x_2 - x_1} y_2, \tag{3}$$

mit der gleichen Abszisse.

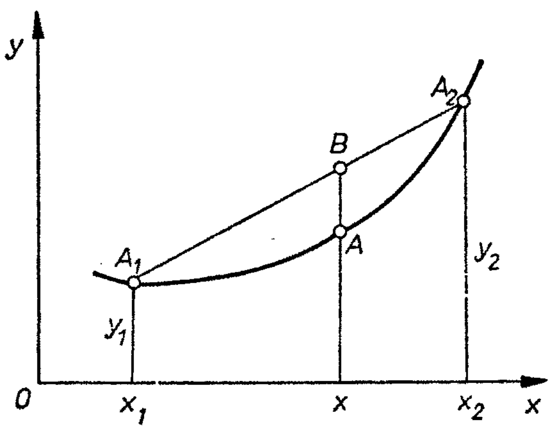

Abb. 71

Somit ist eine konvexe Funktion dadurch charakterisiert, daß alle Punkte jedes Bogens ihrer Kurve unterhalb der zugehörigen Sehne liegen oder auf dieser. (Im Fall einer konkaven Funktion wäre das Wort „unterhalb" durch „oberhalb" zu ersetzen.) Mit der Funktion selbst heißt auch ihre Kurve $y = f(x)$ konvex (bzw. konkav).

Ein triviales Beispiel einer konvexen (und zugleich konkaven) Funktion ist die lineare Funktion $f(x) = ax + b$. Für sie ist die Beziehung (1) immer erfüllt, und zwar mit dem Gleichheitszeichen. Auch die Funktion $f(x) = x^2$ ist konvex, wie man an Hand der Definition leicht unmittelbar nachprüft:

$$(q_1x_1 + q_2x_2)^2 = q_1x_1^2 + q_2x_2^2 - q_1q_2(x_1 - x_2)^2 < q_1x_1^2 + q_2x_2^2,$$

wenn $q_1, q_2 > 0$, $q_1 + q_2 = 1$ ist. Andere Beispiele konvexer Funktionen findet der Leser im weiteren Verlauf unserer Darlegungen.

142. Einfachste Sätze über konvexe Funktionen.

1. *Das Produkt einer konvexen Funktion und einer positiven Konstanten ist eine konvexe Funktion.*

2. *Die Summen zweier oder mehrerer konvexer Funktionen sind ebenfalls konvex.*

In beiden Fällen ergibt sich der Beweis sofort aus der Definition.

Bemerkung. Das Produkt zweier konvexer Funktionen braucht keine konvexe Funktion zu sein. Ein Beispiel hierfür wird später gegeben werden (in der Fußnote auf S. 280).

3. *Ist $\varphi(u)$ konvex und monoton wachsend und ist $u = f(x)$ ebenfalls konvex, dann ist auch die mittelbare Funktion $\varphi(f(x))$ konvex.*

Da f konvex [vgl. (1)] und φ monoton wachsend ist, gilt

$$\varphi\big(f(q_1x_1 + q_2x_2)\big) \leqq \varphi\big(q_1 \cdot f(x_1) + q_2 \cdot f(x_2)\big),$$

und da φ konvex ist, ist die rechte Seite höchstens gleich

$$q_1 \cdot \varphi\big(f(x_1)\big) + q_2 \cdot \varphi\big(f(x_2)\big),$$

so daß sich schließlich die Ungleichung

$$\varphi(f(q_1x_1 + q_2x_2)) \leqq q_1 \cdot \varphi(f(x_1)) + q_2 \cdot \varphi(f(x_2))$$

ergibt, die ihrerseits eine Beziehung vom Typ (1) für die Funktion $\varphi(f(x))$ darstellt.

Wir empfehlen dem Leser, die in der folgenden Tabelle enthaltenen analogen Behauptungen zu beweisen:

$\varphi(u)$	$u = f(x)$	$\varphi(f(x))$
konvex, fallend	*konkav*	*konvex*
konkav, wachsend	*konkav*	*konkav*
konkav, fallend	*konvex*	*konkav*

4. *Sind $y = f(x)$ und $x = g(y)$ eindeutige, zueinander inverse Funktionen (in entsprechenden Intervallen), so gilt gleichzeitig*

$f(x)$	$g(y)$
konvex, wachsend	*konkav, wachsend*
konvex, fallend	*konvex, fallend*
konkav, fallend	*konkav, fallend*

Wollen wir beispielsweise die erste Zeile von links nach rechts beweisen, so setzen wir $f(x_1) = y_1$, $f(x_2) = y_2$, so daß $x_1 = g(y_1)$, $x_2 = g(y_2)$ gilt. Auf Grund von (1) ist

$$f(q_1x_1 + q_2x_2) \leqq q_1 \cdot f(x_1) + q_2 \cdot f(x_2) = q_1y_1 + q_2y_2.$$

Da nach dem Satz über die Umkehrfunktion (Nr. 83) die Funktion $g(y)$ ebenfalls wächst, gilt

$$g(q_1y_1 + q_2y_2) \geqq g(f(q_1x_1 + q_2x_2)) = q_1 \cdot g(y_1) + q_2 \cdot g(y_2),$$

womit bewiesen ist, daß $g(y)$ konkav ist [vgl. (1a)].[1]

5. *Eine im Intervall $\mathcal{X}$ konvexe Funktion $f(x)$, die keine Konstante ist, kann ihren größten Wert nicht im Innern dieses Intervalls annehmen.*

Beweis (indirekt). Wir nehmen an, die Funktion nehme ihren größten Wert in einem inneren Punkt x_0 des Intervalls an. Da $f(x)$ keine Konstante ist, kann dieser Punkt in ein Teilintervall (x_1, x_2) von $\mathcal{X}$ eingebettet werden, $x_1 < x_0 < x_2$, derart, daß wenigstens in einem der Endpunkte des Intervalls der Funktionswert (echt) kleiner als im Punkt x_0 ist. Es sei etwa

$$f(x_1) < f(x_0), \quad f(x_2) \leqq f(x_0).$$

Wir setzen $x_0 = q_1x_1 + q_2x_2$, multiplizieren die erste Ungleichung mit q_1, die zweite mit q_2 und addieren. Wir erhalten

$$q_1 \cdot f(x_1) + q_2 \cdot f(x_2) < f(x_0) = f(q_1x_1 + q_2x_2),$$

was der Konvexität von f widerspricht. Damit ist die Behauptung bewiesen.

6. *Ist das Intervall $[x_1, x_2]$, wobei $x_1 < x_2$ ist, in einem Intervall $\mathcal{X}$ enthalten, in dem die Funktion $f(x)$ konvex ist, so gilt die Beziehung* (1) *entweder stets mit dem Gleichheitszeichen oder stets mit dem Ungleichheitszeichen.*

---

[1]) Alle in der Tabelle formulierten Behauptungen sind aus Abb. 71 unmittelbar abzulesen.

In den Bezeichnungen der Abb. 71 läßt sich dieser Sachverhalt geometrisch folgendermaßen ausdrücken: *Der Bogen A_1A_2 stimmt entweder mit der Sehne A_1A_2 überein, oder aber er liegt vollständig (mit Ausnahme der Endpunkte) unterhalb der Sehne.*

Zum Beweis betrachten wir die lineare Funktion (3), die in den Punkten x_1 und x_2 die gleichen Werte annimmt wie $f(x)$. Zur Abkürzung bezeichnen wir diese lineare Funktion mit $l(x)$.

Da f und $-l$ konvex sind, ist die Differenz

$$\varphi(x) = f(x) - l(x) = f(x) + [-l(x)]$$

ebenfalls konvex (vgl. Satz 2). Dann gilt entweder $\varphi(x) \equiv 0$ im Intervall $[x_1, x_2]$ oder $\varphi \not\equiv 0$. Aus $\varphi(x) \equiv 0$ folgt $f(x) \equiv l(x)$ für dieses Intervall, d. h., der Bogen stimmt mit der Sehne überein, und die Beziehung (1) gilt stets mit dem Gleichheitszeichen. Im zweiten Fall muß im ganzen Intervall (x_1, x_2) aber $\varphi(x) < 0$ gelten. Würde nämlich φ in diesem Intervall auch nichtnegative Werte annehmen, so würde ihr größter Wert im Intervall $[x_1, x_2]$ im Innern dieses Intervalls liegen, was für eine von einer Konstanten verschiedene konvexe Funktion nicht möglich ist (Satz 5). Daher ist im Innern des Intervalls $f(x) < l(x)$, die Kurve liegt unterhalb der Sehne, und in der Beziehung (1) steht stets das Ungleichheitszeichen.

Gilt für jedes in $\mathcal{X}$ enthaltene Intervall $[x_1, x_2]$ mit $x_1 < x_2$ die Beziehung (1) mit dem Ungleichheitszeichen, so nennen wir die Funktion $f(x)$ *streng konvex*. Analog wird der Begriff einer *streng konkaven* Funktion definiert. Diese Terminologie wird auch für die Kurve $y = f(x)$ verwendet.

143. Bedingungen für die Konvexität einer Funktion. Auf Grund von (2) und (2a) läßt sich die grundlegende Ungleichung (1) in der Form

$$f(x) \leqq \frac{x_2 - x}{x_2 - x_1} f(x_1) + \frac{x - x_1}{x_2 - x_1} f(x_2)$$

schreiben oder, in symmetrischer Gestalt,

$$(x_2 - x) f(x_1) + (x_1 - x_2) f(x) + (x - x_1) f(x_2) \geqq 0. \tag{4}$$

Schließlich läßt sich diese Bedingung auch in Determinantenform schreiben:

$$\begin{vmatrix} 1 & x_1 & f(x_1) \\ 1 & x & f(x) \\ 1 & x_2 & f(x_2) \end{vmatrix} \geqq 0. \tag{5}$$

In allen Fällen wird angenommen, daß x zwischen x_1 und x_2 liegt. Wir wollen künftig stets $x_1 < x_2$ voraussetzen.

Wir bemerken beiläufig, daß die Bedingung für die Konvexität einer Funktion in der Form (5) eine unmittelbare geometrische Deutung erhält, wenn wir beachten, daß diese Determinante den doppelten Flächeninhalt des Dreiecks A_1AA_2 (Abb. 72) aus-

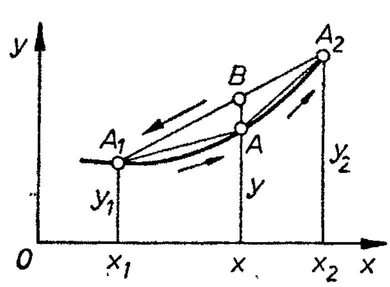

Abb. 72

drückt, und zwar positiv, wenn das Dreieck positiv orientiert ist, d. h., wenn sein Umfang entgegen dem Uhrzeigersinn umlaufen wird.

Wir weisen besonders darauf hin, daß *in allen diesen Bedingungen das Gleichheitszeichen ausgeschlossen ist, wenn es sich um strenge Konvexität handelt.*

Zum Nachweis der Konvexität einer Funktion ergeben sich zweckmäßige Bedingungen, wenn man die Ableitungen der Funktion heranzieht.

Satz 1. *Die Funktion $f(x)$ sei im Intervall $\mathcal{X}$ definiert und stetig und besitze dort eine endliche Ableitung $f'(x)$. Dann ist $f(x)$ in $\mathcal{X}$ genau dann konvex, wenn ihre Ableitung $f'(x)$ (im weiteren Sinn) wachsend ist.*

Die Bedingung ist notwendig; $f(x)$ sei also konvex. Unter der Annahme $x_1 < x < x_2$ können wir (4) in der Form

$$\frac{f(x) - f(x_1)}{x - x_1} \leqq \frac{f(x_2) - f(x)}{x_2 - x} \tag{6}$$

schreiben. Strebt jetzt x gegen x_1 bzw. x_2, so erhalten wir in der Grenze

$$f'(x_1) \leqq \frac{f(x_2) - f(x_1)}{x_2 - x_1} \tag{7a}$$

bzw.

$$f'(x_2) \geqq \frac{f(x_2) - f(x_1)}{x_2 - x_1}, \tag{7b}$$

woraus $f'(x_1) \leqq f'(x_2)$ folgt, so daß $f'(x)$ tatsächlich (im weiteren Sinne) wachsend ist.[1])

Die Bedingung ist hinreichend. Wir setzen nun voraus, daß $f'(x)$ (im weiteren Sinne) wächst. Um die Ungleichung (6) zu beweisen, wenden wir auf ihre beiden Seiten den Mittelwertsatz (Nr. 112) an:

$$\frac{f(x) - f(x_1)}{x - x_1} = f'(\xi_1), \qquad \frac{f(x_2) - f(x)}{x_2 - x} = f'(\xi_2),$$

wobei $x_1 < \xi_1 < x < \xi_2 < x_2$ gilt. Da nach Voraussetzung $f'(\xi_1) \leqq f'(\xi_2)$ gilt, ist die Beziehung (6) tatsächlich erfüllt; hieraus läßt sich die Beziehung (4) gewinnen, welche die Konvexität der Funktion $f(x)$ sichert.

Satz 2. *Die Punktion $f(x)$ sei im Intervall $\mathcal{X}$ definiert und dort nebst ihrer Ableitung $f'(x)$ stetig; ferner besitze sie im Innern von $\mathcal{X}$ eine endliche zweite Ableitung $f''(x)$. Dann ist $f(x)$ in $\mathcal{X}$ genau dann konvex, wenn im Innern von $\mathcal{X}$ die Beziehung*

$$f''(x) \geqq 0 \tag{8}$$

gilt.

In Verbindung mit dem vorhergehenden Satz genügt es, auf die Funktion $f'(x)$ den Satz 2 aus Nr. 132 anzuwenden.

Für die Konkavität einer Funktion erhält man analog die Bedingung

$$f''(x) \leqq 0. \tag{8*}$$

[1]) Für das Folgende ist es wichtig zu bemerken, daß zum Beweis der Ungleichungen (7a) und (7b) nur die Existenz der Ableitung in den Punkten x_1 bzw. x_2 benutzt wurde.

Somit wird durch die Forderung

$$f''(x) > 0 \quad (< 0) \tag{9}$$

offenbar die strenge Konvexität (Konkavität) gewährleistet. Sie schließt nämlich aus, daß $f(x)$ in irgendeinem Intervall linear ist (Nr. 142, Satz 6).

Jetzt können wir viele weitere Beispiele für konvexe oder konkave Funktionen angeben:

1. Die Funktion a^x $(a > 0,\ a \neq 1)$ ist im Intervall $(-\infty, \infty)$ konvex; es ist nämlich $(a^x)'' = a^x \cdot (\ln a)^2 > 0$.

2. Die Funktion $\ln x$ ist im Intervall $(0, \infty)$ konkav; denn es ist $(\ln x)'' = -\frac{1}{x^2} < 0$ (vgl. Nr. 142, Satz 4).

3. Für die Funktion $x \cdot \ln x$ ist (im gleichen Intervall) die zweite Ableitung $\frac{1}{x}$ positiv, also die Funktion selbst konvex.

4. Für die Funktion x^r ist (im gleichen Intervall) die zweite Ableitung gleich $r(r-1)\,x^{r-2}$; hieraus ist ersichtlich, daß für $r > 1$ und für $r < 0$ die Funktion konvex ist, für $0 < r < 1$ jedoch konkav[1]), usw.

Bei allen diesen Beispielen liegt faktisch strenge Konvexität bzw. Konkavität vor.

Zum Abschluß wollen wir noch ein wichtiges geometrisches Charakteristikum für die Konvexität einer Funktion $f(x)$ untersuchen. Zu diesem Zweck ziehen wir an Stelle der in Nr. 141 betrachteten Sehne der Kurve von $f(x)$ jetzt die Tangente in einem beliebigen Kurvenpunkt (Abb. 73) heran.

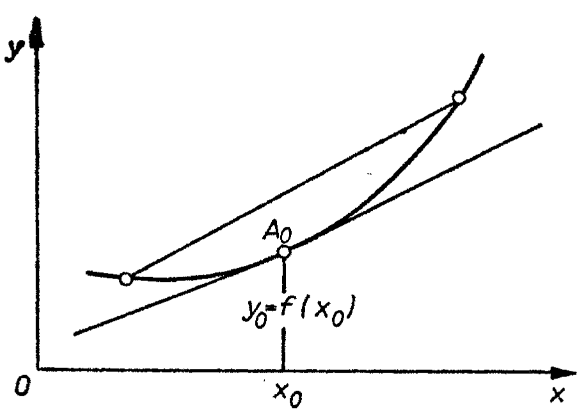

Abb. 73

Satz 3. *Die Funktion $f(x)$ sei im Intervall $\mathcal{X}$ definiert und stetig und besitze dort eine endliche Ableitung $f'(x)$. Dann ist $f(x)$ genau dann konvex, wenn alle Kurvenpunkte oberhalb jeder Tangente (oder auf dieser Tangente) liegen.*

Die Bedingung ist notwendig. Die im Punkt $A_0(x_0, f(x_0))$ an die Kurve $y = f(x)$ gelegte Tangente besitzt den Richtungskoeffizienten $f'(x_0)$. Die Tangentengleichung lautet

$$y = f(x_0) + f'(x_0)\,(x - x_0).$$

Es ist zu zeigen, daß aus der Konvexität von $f(x)$ für beliebige Punkte x_0 und x aus dem Intervall $\mathcal{X}$ die Ungleichung

$$f(x) \geqq f(x_0) + f'(x_0)\,(x - x_0) \tag{10}$$

[1]) Dieses Beispiel liefert nebenbei den Nachweis, daß das Produkt zweier konvexer Funktionen keine konvexe Funktion zu sein braucht. So ist z. B. die Funktion $-x^{1/3}$ konvex, während ihr Quadrat, d. h. die Funktion $x^{2/3}$, konkav ist.

folgt. Sie ist gleichwertig mit den beiden Ungleichungen

$$f'(x_0) \leqq \frac{f(x) - f(x_0)}{x - x_0} \quad \text{für} \quad x > x_0 \tag{11a}$$

und

$$f'(x_0) \geqq \frac{f(x) - f(x_0)}{x - x_0} \quad \text{für} \quad x < x_0; \tag{11b}$$

diese Ungleichungen stimmen jedoch mit den Ungleichungen (7a) bzw. (7b) überein, die wir beim Beweis von Satz 1 erhalten hatten (gerade unter der Voraussetzung, daß die Funktion konvex ist), wenn wir in der ersten $x_2 = x$ und $x_1 = x_0$ und in der zweiten $x_2 = x_0$ und $x_1 = x$ setzen.

Die Bedingung ist hinreichend. Wir setzen jetzt voraus, daß die Ungleichung (10) erfüllt ist oder, was dasselbe ist, daß die Ungleichungen (11a) und (11b) gelten. Dann kann man aus ihnen die Ungleichungen (7a) und (7b) aufstellen, aus denen $f'(x_1) \leqq f'(x_2)$ folgt, so daß also $f'(x)$ eine wachsende Funktion ist. Daraus folgt aber, wie wir wissen (Satz 1), die Konvexität der Funktion $f(x)$.

Bemerkung Wir machen den Leser darauf aufmerksam, daß die Notwendigkeit der Ungleichung (10) (vgl. die Fußnote auf S. 279) für gegebenes x_0 und beliebiges $x \neq x_0$ tatsächlich unter der alleinigen Voraussetzung bewiesen wurde, daß die Ableitung $f'(x_0)$ in eben diesem Punkt x_0 existiert.

144. Die Jensensche Ungleichung und ihre Anwendung. Nach Definition einer konvexen Funktion [vgl. (1)] gilt

$$f(q_1x_1 + q_2x_2) \leqq q_1 \cdot f(x_1) + q_2 \cdot f(x_2) \qquad (q_1, q_2 > 0;\, q_1 + q_2 = 1).$$

Man kann beweisen, daß für konvexe Funktionen eine allgemeinere Ungleichung (die nach Jensen benannt wird) gilt, nämlich

$$f(q_1x_1 + q_2x_2 + \cdots + q_nx_n) \leqq q_1 \cdot f(x_1) + q_2 \cdot f(x_2) + \cdots + q_n \cdot f(x_n) \tag{12}$$
$$(q_1, \ldots, q_n > 0;\, q_1 + \cdots + q_n = 1),$$

wie die Werte $x_1, x_2, \ldots, x_n$ aus dem Grundintervall $\mathcal{X}$ auch gewählt sein mögen. Für $n = 2$ ist sie, wie wir wissen, richtig. Wir nehmen nun an, sie sei für irgendeine natürliche Zahl $n \geqq 2$ gültig, und beweisen, daß sie dann auch für $n + 1$ gilt. Mit anderen Worten, wir zeigen, daß wir, wenn wir $n + 1$ Werte $x_1, \ldots, x_n, x_{n+1}$ aus $\mathcal{X}$ und $n + 1$ positive Zahlen $q_1, \ldots, q_n, q_{n+1}$ wählen, deren Summe gleich 1 ist, folgende Beziehung erhalten:

$$f(q_1x_1 + \cdots + q_nx_n + q_{n+1}x_{n+1}) \leqq q_1 \cdot f(x_1) + \cdots + q_n \cdot f(x_n) + q_{n+1} \cdot f(x_{n+1}). \tag{13}$$

Zu diesem Zweck ersetzen wir links die Summe $q_nx_n + q_{n+1}x_{n+1}$ der letzten beiden Summanden durch den einen Summanden

$$(q_n + q_{n+1}) \left(\frac{q_n}{q_n + q_{n+1}} x_n + \frac{q_{n+1}}{q_n + q_{n+1}} x_{n+1}\right).$$

Dadurch können wir die Ungleichung (12) benutzen und nachweisen, daß der Ausdruck auf der linken Seite von (13) die Summe

$$q_1 \cdot f(x_1) + \cdots + (q_n + q_{n+1}) \cdot f\left(\frac{q_n}{q_n + q_{n+1}} x_n + \frac{q_{n+1}}{q_n + q_{n+1}} x_{n+1}\right)$$

nicht übertrifft. Nun brauchen wir nur noch auf den Wert der Funktion im letzten Summanden die grundlegende Ungleichung (1) anzuwenden, um zu (13) zu gelangen. Somit ist (12) durch Induktion vollständig bewiesen.

Gewöhnlich werden an Stelle der Faktoren q_i, deren Summe gleich 1 ist, beliebige positive Zahlen p_i eingeführt. Setzen wir in (12)

$$q_i = \frac{p_i}{p_1 + \cdots + p_n},$$

so erhalten wir

$$f\left(\frac{\Sigma\, p_i x_i}{\Sigma\, p_i}\right) \leqq \frac{\Sigma\, p_i \cdot f(x_i)}{\Sigma\, p_i}. \tag{12*}$$

Im Fall einer konkaven Funktion f muß das Ungleichheitszeichen offenbar umgekehrt werden. Wenn wir für $f(x)$ verschiedene Funktionen einsetzen, so können wir wichtige konkrete Ungleichungen gewinnen, und zwar alle als Spezialfälle von (12)!

Wir bringen Beispiele:

1. Es sei $f(x) = x^k$ mit $x > 0$, $k > 1$ [$f(x)$ ist eine konvexe Funktion]. Es gilt

$$\left(\frac{\Sigma\, p_i x_i}{\Sigma\, p_i}\right)^k \leqq \frac{\Sigma\, p_i x_i^k}{\Sigma\, p_i} \quad \text{oder} \quad (\Sigma\, p_i x_i)^k \leqq (\Sigma\, p_i)^{k-1} \cdot \Sigma\, p_i x_i^k.$$

Ersetzen wir in der letzten Ungleichung p_i durch $b_i^{\frac{k}{k-1}}$ und x_i durch $\dfrac{a_i}{b_i^{\frac{1}{k-1}}}$, so gelangen wir zu der uns bereits bekannten Ungleichung von Cauchy-Hölder

$$\Sigma\, a_i b_i \leqq \{\Sigma\, a_i^k\}^{\frac{1}{k}} \cdot \left\{\Sigma\, b_i^{\frac{k}{k-1}}\right\}^{\frac{k-1}{k}}$$

[vgl. Nr. 133, Formel (5)].

2. Setzen wir $f(x) = \ln x$ mit $x > 0$ [$f(x)$ ist eine konkave Funktion], so erhalten wir

$$\frac{\Sigma\, p_i \cdot \ln x_i}{\Sigma\, p_i} \leqq \ln \frac{\Sigma\, p_i x_i}{\Sigma\, p_i}.$$

Durch Entlogarithmieren kommen wir zu der uns ebenfalls bereits bekannten Ungleichung

$$\{\Pi\, x_i^{p_i}\}^{1/\Sigma p_i} \leqq \frac{\Sigma\, p_i x_i}{\Sigma\, p_i}$$

[vgl. Nr. 133, Formel (4)].[1])

3. Schließlich setzen wir $f(x) = x \ln x$ mit $x > 0$ [($f(x)$ ist eine konvexe Funktion]. Dann ergibt sich

$$\frac{\Sigma\, p_i x_i}{\Sigma\, p_i} \ln \frac{\Sigma\, p_i x_i}{\Sigma\, p_i} \leqq \frac{\Sigma\, p_i x_i \ln x_i}{\Sigma\, p_i}.$$

Wir multiplizieren mit $\Sigma p_i x_i$ und gehen von den Logarithmen zu den Zahlen selbst über; so erhalten wir die Ungleichung

$$\frac{\Sigma\, p_i x_i}{\Sigma\, p_i} \leqq \{\Pi\, x_i^{p_i x_i}\}^{\frac{1}{\Sigma p_i x_i}}.$$

Setzen wir insbesondere $p_i = \dfrac{1}{x_i}$, so folgt

$$\frac{n}{\Sigma \dfrac{1}{x_i}} \leqq \sqrt[n]{\Pi\, x_i}.$$

[1]) Ebenso wie Σ eine Summe bezeichnet, so bedeutet Π ein Produkt.

Dehnt man den Begriff des harmonischen Mittels[1]) auf den Fall mehrerer Zahlen aus, so läßt sich diese Ungleichung auch folgendermaßen formulieren: *Das harmonische Mittel positiver Zahlen ist höchstens gleich ihrem geometrischen Mittel.*

145. Wendepunkte. Beim Zeichnen der Kurve einer Funktion $f(x)$ (damit wird sich der folgende Paragraph befassen) sind die sogenannten Wendepunkte von $y = f(x)$ von besonderem Interesse.

Ein Punkt $M(x_0, f(x_0))$ einer Kurve heißt *Wendepunkt*, wenn er ein Stück der Kurve, auf dem $f(x)$ konvex (von unten konvex) ist, von einem Stück trennt, auf dem diese Funktion konkav (von oben konvex) ist (Abb. 74).

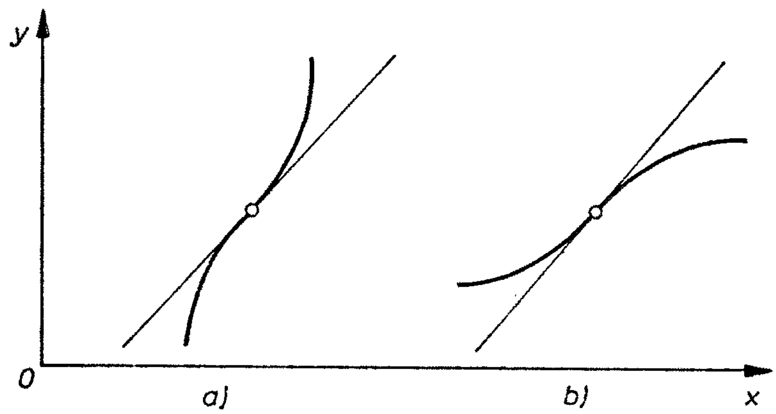

Abb. 74

Angenommen, $f(x)$ besitze in dem betrachteten Intervall eine endliche Ableitung, so ist diese nach Satz 2 aus Nr. 143 in einer bestimmten Umgebung $[x_0 - \delta, x_0]$ links von x_0 wachsend und in der Umgebung $[x_0, x_0 + \delta]$ rechts von x_0 fallend oder umgekehrt links fallend und rechts wachsend. Im ersten Fall besitzt $f'(x)$ für $x = x_0$ ein Maximum, im zweiten ein Minimum. Setzen wir noch die Existenz einer endlichen zweiten Ableitung $f''(x)$ mindestens für $x = x_0$ voraus, so ist notwendigerweise $f''(x_0) = 0$ vgl. (Nr. 134).

Die Bedingung $f''(x_0) = 0$ spielt dieselbe Rolle bei der Ermittlung der Wendepunkte von $f(x)$ wie die Bedingung $f'(x_0) = 0$ bei der Ermittlung der Extremwerte von $f(x)$: Sie ist notwendig, aber nicht hinreichend. Davon überzeugt man sich am einfachsten an Hand eines Beispiels. Für die Funktion $f(x) = x^4$ ist $f''(x) = 12x^2 \geqq 0$ im Intervall $(-\infty, \infty)$, so daß nach Satz 2 aus Nr. 143 die Funktion $f(x)$ in diesem gesamten Intervall konvex ist, obgleich $f''(x)$ im Punkt $x = 0$ gleich 0 ist.

Existiert die zweite Ableitung $f''(x)$ überall im Innern des betrachteten Intervalls, so sind die Abszissen der Wendepunkte unter ihren Nullstellen zu suchen. Dabei muß jeder solcher Punkt x_0 untersucht werden.

In bestimmten Umgebungen $[x_0 - \delta, x_0)$ und $(x_0, x_0 + \delta]$ links bzw. rechts von x_0 möge $f''(x)$ ein bestimmtes Vorzeichen beibehalten. Dann läßt sich zur Ermittlung eines Wendepunktes folgende Regel angeben:

Ändert $f''(x)$ beim Durchlaufen von $x = x_0$ das Vorzeichen, so liegt ein Wendepunkt vor, anderenfalls jedoch nicht (vgl. Nr. 135).

Wir weisen noch auf folgendes hin: Auf den Kurvenstücken, die durch einen Wendepunkt $(x_0, f(x_0))$ getrennt werden, ist die Kurve streng konvex bzw. streng konkav.

Wir betrachten als Beispiel die Funktion $f(x) = \sin x$. Die Ableitung $f''(x) = -\sin x$ ist in den Punkten $x = k\pi$ (k ganz) gleich 0 und ändert beim Durchlaufen dieser Werte das Vorzeichen.

---

[1]) Vgl. die Fußnote 2 auf S. 73.

Folglich sind alle Punkte der Sinuskurve, die auf der x-Achse liegen, Wendepunkte. Man sieht leicht, daß die Sinuskurve in den Intervallen $([2m-1]\,\pi, 2m\pi)$ konvex (von unten konvex) und in den Intervallen $(2m\pi, [2m+1]\,\pi)$ konkav (von oben konvex) ist.

Man könnte auch, wie wir es in Nr. 138 bei der Bestimmung der Extremwerte einer Funktion taten, in einem zu untersuchenden Punkt x_0, für den $f''(x_0) = 0$ gilt, die höheren Ableitungen heranziehen. Auf diesem Wege erhält man die folgende Regel: *Ist die erste Ableitung (von höherer als zweiter Ordnung), die in einem Punkt x_0 nicht gleich 0 ist, von ungerader Ordnung, so liegt ein Wendepunkt vor; ist diese Ableitung dagegen von gerader Ordnung, so liegt kein Wendepunkt vor.*

Zum Abschluß weisen wir noch auf eine bemerkenswerte Eigenschaft der in einem Wendepunkt an eine Kurve $y = f(x)$ gelegten Tangente hin (wenn eine solche Tangente existiert). Die Kurve geht in diesem Punkt von einer Seite der Tangente auf die andere über, d. h., Kurve und Tangente durchsetzen einander (vgl. Abb. 74).

Das ist klar, wenn die Tangente vertikal verläuft (vgl. Abb. 43a, b, S. 195). Wir wollen jetzt den Fall untersuchen, daß die Tangente schräg oder horizontal verläuft; dabei setzen wir die Existenz einer endlichen Ableitung $f'(x_0)$ voraus. Wir nehmen etwa an, links vom Wendepunkt, für $x_0 - \delta \leqq x < x_0$, sei die Kurve konvex, rechts von ihm, für $x_0 < x \leqq x_0 + \delta$, konkav (das entspricht Abb. 74b). In diesem Fall verläuft die Kurve für $x < x_0$ oberhalb der Tangente (oder auf ihr) und für $x > x_0$ unterhalb der Tangente (oder auf ihr), d. h., es gilt

$$
\begin{aligned}
f(x) &\geqq f(x_0) + f'(x_0)\,(x - x_0) \quad \text{für} \quad x < x_0,\\
f(x) &\leqq f(x_0) + f'(x_0)\,(x - x_0) \quad \text{für} \quad x > x_0.
\end{aligned}
$$

Die erste dieser Ungleichungen stimmt aber mit der Ungleichung (10) aus Nr. 143 überein (man beachte die Bemerkung dort). Die zweite ist das Analogon der Ungleichung (10) für eine konkave Funktion.

Bemerkung. Oft wird gerade diese Eigenschaft einer Kurve als Definition des Wendepunktes benutzt. Sie ist jedoch der obigen Definition keineswegs gleichwertig. Eine Kurve braucht in einem Wendepunkt keine Tangente zu besitzen, so daß dann die zweite Definition nicht anwendbar ist. Es kann auch der umgekehrte Fall eintreten: Die Kurve kann die Tangente in einem Punkt durchsetzen, der nicht einen konvexen Teil von einem konkaven trennt; dann ist die erste Definition nicht anwendbar. Solche Kurven zeigen Abb. 43c und d (S. 195). Interessanter ist jedoch die Kurve

$$
y = x^5\left(1 + \sin^2\frac{1}{x}\right) \quad \text{für} \quad x \neq 0\,, \; y = 0 \quad \text{für} \quad x = 0,
$$

die im Koordinatenursprung die x-Achse berührt und sie durchsetzt. Hier existiert sogar eine stetige zweite Ableitung, sie ändert jedoch in der Nähe von $x = 0$, und zwar sowohl links als auch rechts, unendlich oft ihr Vorzeichen. Davon möge sich der Leser durch Ausrechnen überzeugen.

§ 3. Das Zeichnen von Kurven

146. Aufgabenstellung. Nachdem wir die Methoden der Differentialrechnung kennengelernt haben, kommen wir auf das Problem des *Kurvenzeichnens* (vgl. Nr. 47) zurück. Zunächst möge es sich darum handeln, die Kurve einer in einem endlichen Intervall

$[a, b]$ stetigen Funktion $y = f(x)$ graphisch darzustellen. Dabei besteht jetzt unser Hauptziel darin, den *Verlauf* der Funktion möglichst genau zu charakterisieren. Die Genauigkeit der einzelnen Ordinaten interessiert uns in geringerem Maße.

Das gewöhnlich angewandte Verfahren der „punktweisen" Konstruktion (Nr. 47), wobei die Punkte mehr oder weniger dicht, jedoch zufällig und ohne Beziehung zu den (zunächst noch unbekannten) Besonderheiten der Kurve gewählt werden, ist unzureichend. Es erfordert vor allem die Berechnung einer großen Anzahl von Koordinaten, was in der Praxis unzweckmäßig ist. Der Hauptmangel liegt jedoch in etwas anderem: Die punktweise Konstruktion ist prinzipiell deshalb ungeeignet, weil sie gerade angesichts der Zufälligkeiten in der Auswahl der zu berechnenden Ordinaten nicht gewährleistet, daß das gestellte Ziel, nämlich eine Übersicht über den wesentlichen Funktionsverlauf zu gewinnen, erreicht wird.

Wir nehmen an, die Funktion $y = f(x)$ habe eine endliche Ableitung $y' = f'(x)$, mit Ausnahme höchstens endlich vieler Punkte, in denen die Ableitung ∞ ist. Links und rechts von diesen Punkten nehme $f'(x)$ den Wert ∞ mit ein und demselben oder aber mit verschiedenen Vorzeichen an. Dann ermöglichen es die Methoden der Differentialrechnung, gewisse „Stütz"-Punkte anzugeben, die für die vorliegende Kurve charakteristisch sind und mit deren Hilfe die Kurve mit der notwendigen Genauigkeit gezeichnet werden kann.

Zunächst meinen wir hier die Punkte, in denen sich die Richtung der Kurve ändert, die höchsten Punkte der „Berge" bzw. die tiefsten Punkte der „Täler", also die Extrema der Funktion (Nr. 134 bis 138). Man muß weiter noch alle jene Punkte hinzunehmen, in denen die Tangente horizontal oder vertikal verläuft, selbst wenn es sich dabei nicht um Extremwerte der Funktion handelt. Natürlich müssen auch die Endpunkte der Kurve beachtet werden.

Wenn diese Punkte auf der Zeichnung vermerkt sind (gewöhnlich sind es nicht sehr viele), kann man eigentlich die Kurve schon zeichnen.

Die so erhaltene Kurve gibt den Funktionsverlauf schon ziemlich vollständig wieder; sie zeigt die Abschnitte, auf denen die Funktion wächst bzw. fällt, und liefert auch die Punkte, in denen die Geschwindigkeit der Änderung der Funktion auf 0 heruntergeht ($y' = 0$) oder bis ins Unendliche wächst ($y' = \pm\infty$).

Die Darstellung läßt sich weiter verbessern, wenn man die Stücke berücksichtigt, auf denen die Kurve konvex (von unten konvex) bzw. konkav (von oben konvex) ist, und die diese Kurvenstücke trennenden Wendepunkte berechnet (Nr. 143, 145).

147. Das Schema zur Konstruktion einer Kurve. Beispiele. Die Funktion $y = f(x)$ sei in dem betrachteten Intervall $[a, b]$ zweimal differenzierbar, eventuell mit Ausnahme einzelner Punkte, in denen die Ableitung $y' = f'(x)$ unendliche Werte annimmt, deren Vorzeichen zu beiden Seiten eines solchen Punktes gleich oder auch verschieden sein können.

Dann verfährt man beim Zeichnen der Kurve der Funktion $y = f(x)$ folgendermaßen:

1. Man bestimmt die x-Werte, für welche die Ableitung $y' = f'(x)$ gleich 0 oder gleich ∞ ist, und untersucht, inwieweit es sich dabei um Extremwerte handelt.

2. Man bestimmt diejenigen x-Werte, für welche die zweite Ableitung $y'' = f''(x)$ gleich 0 ist, und untersucht, welche davon Wendepunkten entsprechen.

3. Man berechnet die Werte von $y = f(x)$ in allen diesen x-Werten sowie in den Endpunkten a und b des betrachteten Intervalls.

Die Ergebnisse faßt man am besten in einer Tabelle zusammen (vgl. die Beispiele weiter unten), wobei die Besonderheiten der berechneten Punkte vermerkt werden, z. B. Maximum, Minimum, $y' = 0$, $y' = \infty$ oder $y' = -\infty$, $y' = \pm\infty$ oder $y' = \mp\infty$ Wendepunkt.[1])

Manchmal nimmt man nach Belieben noch weitere Punkte hinzu, etwa die Schnittpunkte der Kurve mit den Achsen.

Nachdem diese Werte auf der Zeichnung eingetragen sind, legt man durch diese Punkte eine Kurve, wobei man alle erwähnten Besonderheiten zu berücksichtigen hat. Dabei haben wir uns natürlich auf den in der Praxis üblicherweise vorkommenden Fall beschränkt, daß es insgesamt nur endlich viele Punkte gibt, in denen die erste Ableitung gleich 0 oder gleich $\pm\infty$ ist und die zweite Ableitung verschwindet.

Die Kurve verläuft dann in den Intervallen zwischen ihnen stets ansteigend oder fallend und ist von unten oder von oben konvex.

Berechnen und Zeichnen einer Kurve vereinfachen sich, wenn die Funktion bei einer Vorzeichenänderung von x ihre Werte nicht ändert (eine gerade Funktion ist), so daß die Kurve symmetrisch zur vertikalen Achse verläuft. Einen entsprechenden Vorteil bietet die Symmetrie bezüglich des Koordinatenursprungs; das findet analytisch seinen Ausdruck darin, daß die Funktion bei einem Vorzeichenwechsel von x ebenfalls nur ihr Vorzeichen ändert (eine ungerade Funktion ist).

Beispiele.

1. In Nr. 136, Beispiel 2, untersuchten wir bereits das Verhalten der Funktion

$$y = \sin^3 x + \cos^3 x.$$

Mit Hilfe ihrer Ableitung fanden wir die x-Werte, in denen sie Extremwerte besitzt, und berechneten diese Extremwerte der Funktion. Dabei konnten wir uns auf Grund der Periodizität auf das Intervall $[0, 2\pi]$ beschränken. Die Zeichnung braucht ebenfalls nur für dieses Intervall angefertigt zu werden.

Jetzt müssen wir die Nullstellen der zweiten Ableitung bestimmen. Wir schreiben diese in der Form

$$y'' = \frac{9}{2}(\sin x + \cos x)\cdot\left(\sin 2x - \frac{2}{3}\right)$$

und stellen fest, daß die erste Klammer für $x = \frac{3\pi}{4} \approx 2{,}36$ und $x = \frac{7\pi}{4} \approx 5{,}50$ und die zweite für $x \approx 0{,}36$ (21°), 1,21 (69°), 3,51 (201°) und 4,35 (249)° gleich 0 wird. In allen diesen Punkten ändert y'' das Vorzeichen, so daß es sich um Wendepunkte handelt.

Wir stellen nun folgende Tabelle auf:

$x = 0$	0,36	0,78	1,21	1,57	2,36	3,14	3,51
$y = 1$	0,86	0,71	0,86	1	0	-1	$-0{,}86$
$y' = 0$ Maximum	Wendepunkt	$y' = 0$ Minimum	Wendepunkt	$y' = 0$ Maximum	Wendepunkt	$y' = 0$ Minimum	Wendepunkt

[1]) Die Beziehung $y' = \pm\infty$ soll andeuten, daß die Ableitung links von dem betreffenden Punkt $+\infty$ und rechts $-\infty$ wird; entsprechend bei $y' = \mp\infty$.

$x = 3{,}94$	4,35	4,71	5,50	6,28
$y = -0{,}71$	$-0{,}86$	-1	0	1
$y' = 0$ Maximum	Wendepunkt	$y' = 0$ Minimum	Wendepunkt	$y' = 0$ Maximum

An Hand dieser Tabelle wurde die Kurve in Abb. 58 (S. 262) gezeichnet.

Bemerkung. Der Leser möge berücksichtigen, daß die in diesem Buch wiedergegebenen Zeichnungen in Anbetracht des kleinen Maßstabs die Zahlenwerte nicht so genau wiedergeben können, wie man sie durch Berechnung erhält. Er möge diese Zeichnungen unter Zugrundelegung eines größeren Maßstabs selbst nochmals anfertigen.

2. Wir betrachten die Funktion

$$y = \sin x + \sin 2x.$$

Sie ist nicht nur periodisch, sondern auch ungerade. Dadurch kann man das x-Intervall weiter verkürzen, nämlich auf $[0, \pi]$.

In diesem Intervall wird die Ableitung

$$y' = \cos x + 2 \cos 2x = 4 \cos^2 x + \cos x - 2$$

gleich 0 für $\cos x = \dfrac{-1 \pm \sqrt{33}}{8}$, d. h. für $x \approx 0{,}94$ (54°) und 2,57 (147°). Da die zweite Ableitung

$$y'' = -\sin x - 4 \sin 2x = -\sin x(1 + 8 \cos x)$$

für den ersten dieser Werte offenbar negativ ist, besitzt die Funktion an dieser Stelle ein Maximum. Analog ergibt sich für den zweiten Wert ein Minimum.

Die zweite Ableitung verschwindet für $x = 0$ und $x = \pi \approx 3{,}14$, da dort $\sin x$ gleich 0 ist, und für $x \approx 1{,}70$ (97°), da dort der in Klammern stehende Faktor gleich 0 ist, wobei sich in allen Fällen beim Durchlaufen des Wertes das Vorzeichen ändert (also Wendepunkte vorliegen).

Man bekommt also die folgende Tabelle:

$x = 0$	0,94	1,70	2,09	2,57	3,14
$y = 0$	1,76	0,74	0	$-0{,}37$	0
Wendepunkt	$y' = 0$ Maximum	Wendepunkt		$y' = 0$ Minimum	Wendepunkt

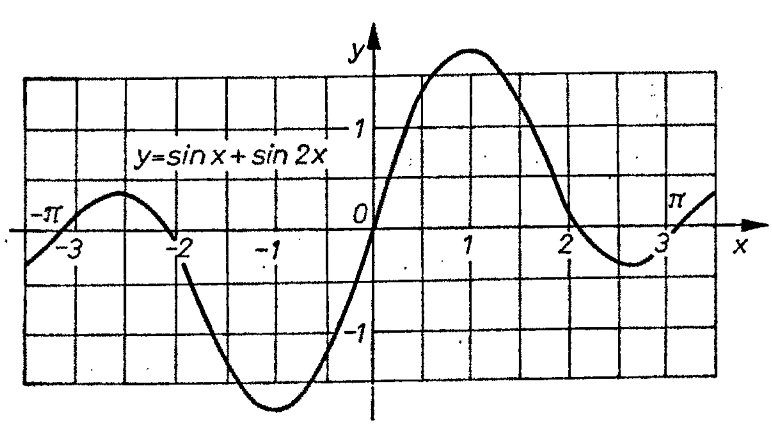

Abb. 75

Zu den obigen x-Werten haben wir noch den Wert $x = \frac{2}{3}\pi \approx 2{,}09\,(120^\circ)$ hinzugefügt, für den $y = 0$ ist (die Kurve die x-Achse schneidet). Die an Hand dieser Punkte gezeichnete Kurve zeigt Abb. 75. Für das Intervall $[-\pi, 0]$ erhält man sie durch zweifache Spiegelung des Kurvenbogens, und zwar einmal an der y-Achse und danach an der x-Achse.

148. Unendlichkeitsstellen. Unendliche Intervalle. Asymptoten. Es ist zweckmäßig, die Klasse der zu untersuchenden Funktionen in zwei Richtungen zu erweitern. Erstens lassen wir jetzt zu, daß die Funktion $y = f(x)$ für einzelne x-Werte unendlich große Werte annimmt. Das bedeutet also folgendes: Ist x_0 ein solcher Wert, so strebt $f(x)$ gegen ∞ oder $-\infty$, wenn sich x von der einen oder anderen Seite her dem Wert x_0 nähert. Zweitens wollen wir uns für das Verhalten einer Funktion in einem unendlichen Intervall interessieren.

Da die Ausmaße einer Zeichnung natürlich endlich sind, muß man sich in beiden Fällen mit einem Teil der Kurve begnügen. Dabei wollen wir in der Zeichnung auf solche Teile der Kurve verzichten, von denen wir uns auf Grund des gezeichneten Teiles eine Vorstellung machen können.

Wir wollen uns zunächst mit dem Fall beschäftigen, daß eine Funktion etwa in $x = x_0$ eine Unstetigkeitsstelle hat. Strebt x von einer Seite her gegen x_0, so strebt die Funktion monoton gegen ∞ oder $-\infty$, wenn in jedem endlichen Teil des x-Intervalls die Ableitung $y' = f'(x)$ nur endlich oft ihr Vorzeichen wechselt.

Auf den verschiedenen Seiten von x_0 (falls x_0 nicht in einen Endpunkt des Intervalls fällt) kann die Funktion „unendliche Grenzwerte" verschiedener Vorzeichen besitzen. In jedem Fall nähert sich die Kurve, die ins Unendliche verläuft, unbegrenzt der vertikalen Geraden $x = x_0$, und zwar, je nach dem Vorzeichen des unendlichen Grenzwertes, oben oder unten. Diese Gerade ermöglicht eine anschauliche Vorstellung des Kurvenverlaufs auch über die Grenzen der Zeichnung hinaus (Abb. 76).

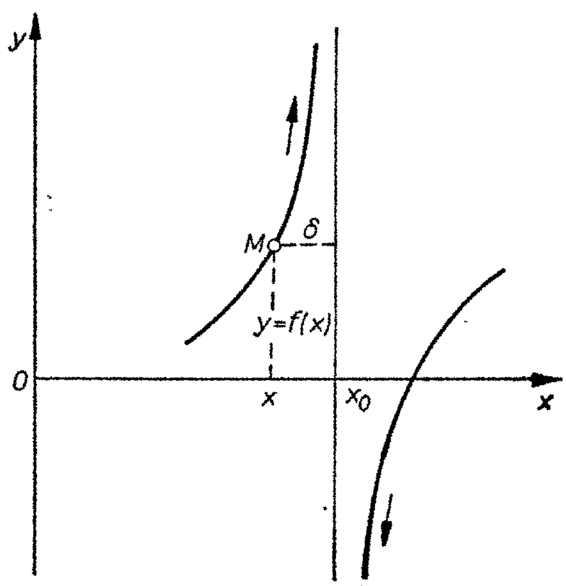

Abb. 76

Als Beispiele können die uns bereits bekannten graphischen Darstellungen von $y = \frac{a}{x}$ für $x = 0$ (Abb. 10, S. 100), $y = \tan x$ für $x = (2k+1)\frac{\pi}{2}$ (Abb. 16, S. 102), $y = \log_a x$ für $x = 0$ (Abb. 14, S. 101) dienen.

Bei einem (einseitig oder beiderseitig) unendlichen Intervall leistet uns manchmal eine horizontale oder schräge Gerade, die sich der Kurve unbegrenzt nähert, denselben Dienst. Im Zusammenhang damit geben wir folgende allgemeine Definition.

Es sei eine Kurve gegeben, von der ein Zweig in einer bestimmten Richtung ins Unendliche verläuft. Strebt der Abstand δ eines Kurvenpunktes von einer bestimmten Geraden gegen 0, wenn der Punkt selbst ins Unendliche rückt, so nennt man diese Gerade eine *Asymptote* der Kurve.

In unserem obigen Beispiel ($f(x) \to \pm\infty$ für $x \to x_0$) hatten wir es mit einer vertikalen Asymptote zu tun; jetzt untersuchen wir horizontale und schräge Asymptoten der durch $y = f(x)$ gegebenen Kurve.

Beispielen für horizontale Asymptoten sind wir bereits begegnet: für die Kurve $y = \frac{a}{x}$ ist es die Gerade $y = 0$ bei $x \to \pm\infty$ (Abb. 10), für die Kurve $y = \arctan x$ sind es die Geraden $y = \frac{\pi}{2}$ und $y = -\frac{\pi}{2}$ bei $x \to \infty$ bzw. $x \to -\infty$ (Abb. 21, S. 107), für die Kurve $y = a^x$ ist es die Gerade $y = 0$ bei $x \to -\infty$, wenn $a > 1$, und bei $x \to \infty$, wenn $a < 1$ ist (Abb. 13, S. 101).

Für $x \to \infty$ ist die Gerade $Y = b$ genau dann Asymptote der Kurve $y = f(x)$ (Abb. 77), wenn

$$\lim_{x\to\infty} \delta = \lim_{x\to\infty} |y - b| = 0 \quad \text{bzw.} \quad \lim_{x\to\infty} y = \lim_{x\to\infty} f(x) = b$$

gilt. Somit läuft die Frage nach einer horizontalen Asymptote einfach auf die Frage nach diesem Grenzwert bzw. nach dem entsprechenden Grenzwert für $x \to -\infty$ hinaus. Es kann (wie bei der Kurve $y = \arctan x$) vorkommen, daß es zwei horizontale Asymptoten gibt.

Beispiele für schräge Asymptoten sind die aus der analytischen Geometrie bekannten Asymptoten $y = \pm\frac{b}{a}x$ der Hyperbel

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \quad \text{bzw.} \quad y = \pm\frac{b}{a}\sqrt{x^2 - a^2} \tag{1}$$

(vgl. auch Abb. 7, S. 98).

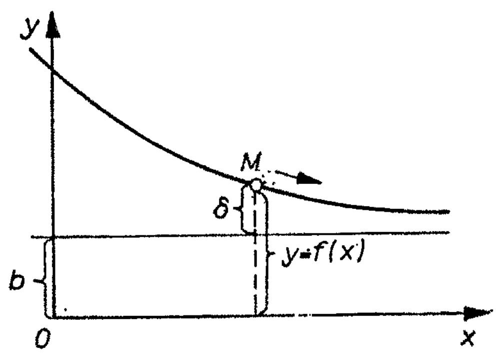

Abb. 77

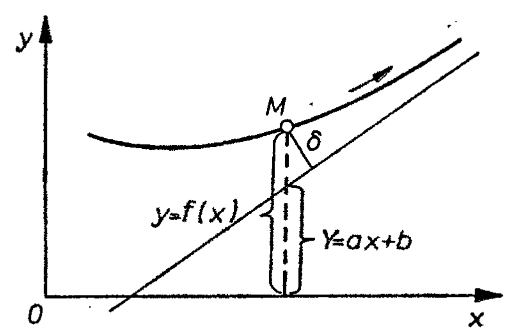

Abb. 78

Wir nehmen jetzt an, die Kurve $y = f(x)$ besitze die beliebige Asymptote

$$Y = ax + b \tag{2}$$

(Abb. 78), und betrachten sie für $x \to \infty$. Da sich die Ordinatendifferenz $|y - Y|$ nur um einen konstanten Faktor (der gleich dem Kosinus des Winkels zwischen der Asymptote und der x-Achse ist) vom Abstand δ unterscheidet, muß für $x \to \infty$ zugleich mit δ auch diese Differenz gegen 0 streben:

$$\lim_{x\to\infty} (y - ax - b) = 0. \tag{3}$$

Nach Division durch x erhalten wir hieraus

$$\lim_{x\to\infty} \frac{y}{x} = a. \tag{4}$$

Außerdem liefert Gleichung (3) unmittelbar

$$\lim_{x\to\infty} (y - ax) = b. \tag{5}$$

Wenn die Gerade (2) Asymptote einer gegebenen Kurve sein soll, müssen die Bedingungen (4) und (5) erfüllt sein. Diese sind also notwendig. Sie sind auch hinreichend, wie man durch Umkehrung der Schlüsse leicht nachweist. Es handelt sich dabei um die sukzessive Ermittlung der Grenzwerte (4) und (5), durch welche die Koeffizienten der Geradengleichung (2) bestimmt werden.

Natürlich muß für $x \to -\infty$ eine analoge Untersuchung durchgeführt werden.

Für $x \to \infty$ erhalten wir im Fall der Hyperbel (1)

$$\frac{y}{x} = \pm\frac{b}{a}\cdot\frac{\sqrt{x^2-a^2}}{x} = \pm\frac{b}{a}\cdot\sqrt{1-\frac{a^2}{x^2}} \to \pm\frac{b}{a}$$

und dann

$$y \mp \frac{b}{a}x = \pm\frac{b}{a}\left(\sqrt{x^2-a^2} - x\right) = \mp\frac{ab}{x+\sqrt{x^2-a^2}} \to 0,$$

und wir gelangen zu den uns bereits bekannten Asymptoten

$$y = \pm\frac{b}{a}x.$$

Jetzt können wir unser Schema zur Konstruktion der Kurve einer Funktion durch die folgenden Punkte ergänzen:

4. Man bestimme die x-Werte, für welche $y = f(x)$ den Wert ∞ annimmt, wobei das Vorzeichen zu berücksichtigen ist, und zeichne die entsprechenden vertikalen Asymptoten.

5. Man bestimme die horizontalen oder die gegen die x-Achse geneigten Asymptoten der Kurve (und zwar einzeln für $x \to \infty$ und $x \to -\infty$, falls das Intervall nach beiden Seiten unendlich ist).

Wir bringen wieder einige Beispiele.

149. Beispiele.

3. Wir kommen auf die Funktion

$$y = (x+2)^2(x-1)^3$$

zurück, für die wir bereits die Extremwerte in Nr. 136, Beispiel 1, ermittelten.

Diese Funktion ist für $-\infty < x < \infty$ stetig. Für $x \to \pm\infty$ strebt nicht nur y, sondern auch $\frac{y}{x}$ gegen ∞, so daß keine Asymptoten vorhanden sind.

Wir betrachten ergänzend dazu die zweite Ableitung

$$y'' = 2(x-1)(10x^2 + 16x + 1).$$

Sie verschwindet für $x = 1$, $x = -0{,}07$ und $x = -1{,}53$, wobei sie ihr Vorzeichen ändert (Wendepunkte).

Wir stellen folgende Tabelle auf:

$x = -2$	$-1{,}53$	$-0{,}8$	$-0{,}07$	0	1
$y = 0$	$-3{,}58$	$-8{,}40$	$-4{,}56$	-4	0
$y' = 0$ Maximum	Wendepunkt	$y' = 0$ Minimum	Wendepunkt		$y' = 0$ Wendepunkt

Die graphische Darstellung zeigt Abb. 57, S. 262.

4. Es sei (wieder mit $x^{1/3} = -|x|^{1/3}$ für $x < 0$)

$$y = x^{2/3} - (x^2 - 1)^{1/3}$$

(vgl. Nr. 136, Beispiel 3). Die Funktion bleibt im Intervall $(-\infty, \infty)$ stetig. Schreibt man sie in der Form

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

so stellt man leicht fest, daß $y \to 0$ für $x \to \pm\infty$ gilt, so daß die Kurve unserer Funktion die x-Achse zur Asymptote hat (sowohl links als auch rechts). Die zweite Ableitung y'' besitzt keine Nullstelle. Wendepunkte liegen nur an den Punkten ($x \neq 0$) vor, an denen die Ableitung y' den Wert ∞ annimmt. Da die Funktion gerade ist, verläuft sie symmetrisch zur y-Achse. Die Tabelle lautet:

$x = -\infty$	-1	$-0{,}71$	0	$0{,}71$	1	∞
$y = 0$	1	$1{,}59$	1	$1{,}59$	1	1
	$y' = \infty$	$y' = 0$ Maximum	$y' = \pm\infty$ Minimum	$y' = 0$ Maximum	$y' = -\infty$	

Die Funktion ist in Abb. 59, S. 263, graphisch dargestellt.

5. $y = \dfrac{x^2 - 5x + 6}{x^2 + 1}$ (vgl. Nr. 137). Die Funktion ist stetig im Intervall $(-\infty, \infty)$. Für $x \to \pm\infty$ gilt offenbar $\lim y = 1$, d. h., es handelt sich um eine horizontale Asymptote. Die zweite Ableitung

$$y'' = -10\,\frac{(x + 1)\,(x^2 - 4x + 1)}{(x^2 + 1)^3}$$

wird für $x = -1$, $x = 2 + \sqrt{3} \approx 3{,}73$ und $x = 2 - \sqrt{3} \approx 0{,}27$ gleich 0 und ändert dabei ihr Vorzeichen (Wendepunkte). Die Tabelle lautet:

$x = -\infty$	-10	-5	-1	$-0{,}41$	0	$0{,}27$	2	$2{,}41$	3	$3{,}73$	5	10	∞
$y = 1$	$1{,}55$	$2{,}15$	6	$7{,}04$	6	$4{,}40$	0	$-0{,}03$	0	$0{,}08$	$0{,}23$	$0{,}55$	1
			Wendepunkt	$y' = 0$ Maximum		Wendepunkt		$y' = 0$ Minimum		Wendepunkt			

Die Kurve ist in Abb. 61, S. 266, gezeichnet. Hier wirkt sich der kleine Maßstab störend auf die Übersichtlichkeit der Zeichnung aus, besonders im x-Intervall zwischen 2 und 5. Dieser Teil der Kurve ist deshalb in vergrößertem Maßstab noch einmal dargestellt.

Wir betrachten noch eine Reihe neuer Beispiele.

6. $y = \dfrac{(x-1)^3}{(x+1)^2}$.

Die Funktion wird $-\infty$ für $x = -1$. Da für $x \to \pm\infty$

$$\frac{y}{x} \to 1, \quad y - x = \frac{-5x^2 + 2x - 1}{(x+1)^2} \to -5$$

gilt, besitzt die Kurve die Asymptote $Y = x - 5$. Wir bilden die Ableitungen

$$y' = \frac{(x-1)^2\,(x+5)}{(x+1)^3}, \quad y'' = \frac{24(x-1)}{(x+1)^4}.$$

Die erste Ableitung verschwindet für $x = 1$ (Wendepunkt) und für $x = -5$ (Maximum). Andere Wendepunkte sind nicht vorhanden. Mit Hilfe der Tabelle

$x = -10$	-5	-3	-1	0	1	5	10
$y = -16{,}4$	$-13{,}5$	-16	$-\infty$	-1	0	$1{,}78$	$6{,}05$
	$y' = 0$ Maximum				$y' = 0$ Wendepunkt		

stellen wir die Funktion unter Berücksichtigung der Asymptote graphisch dar (Abb. 79).

7. $y = \sqrt{\dfrac{x^3}{x-a}} \quad (a > 0)$.

Hiernach hat die Funktion nur dann reelle Werte, wenn $x \leqq 0$ oder $x > a$ gilt. Für $x = a$

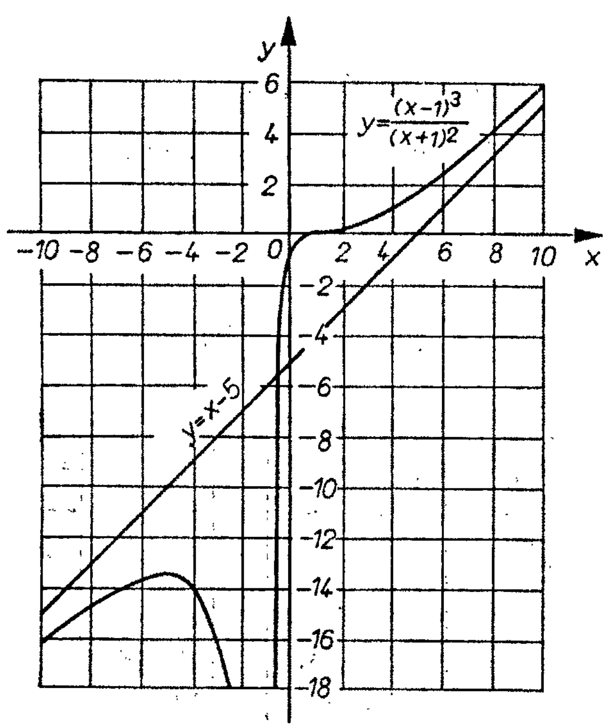

Abb. 79

wird sie ∞. Setzen wir $x > a$ voraus, so erhalten wir für $x \to \infty$

$$\frac{y}{x} = \sqrt{\frac{x}{x-a}} \to 1,$$

$$y - x = \frac{x}{\sqrt{x-a}} \cdot \frac{a}{\sqrt{x} + \sqrt{x-a}} \to \frac{a}{2},$$

so daß sich die Kurve für positive x der Asymptote $Y = x + \frac{a}{2}$ nähert.

Analog erhält man für negative x-Werte eine andere Asymptote, nämlich $Y = -x - \frac{a}{2}$. Die Ableitung

$$y' = \frac{1}{y} \cdot \frac{x^2\left(x - \frac{3}{2}a\right)}{(x-a)^2} = \left(x - \frac{3}{2}a\right)\sqrt{\frac{x}{(x-a)^3}}$$

wird für $x = \frac{3}{2}a$ gleich 0, wobei sie ihr Vorzeichen von Minus in Plus ändert (Minimum). Sie verschwindet auch für $x = 0$; dies ist aber ein Endpunkt des Intervalls $(-\infty, 0]$, in dem wir die Funktion betrachten, und von einem (relativen) Extremwert kann hier natürlich nicht die Rede sein. Die zweite Ableitung lautet

$$y'' = \frac{1}{y} \cdot \frac{\frac{3}{4}a^2 x}{(x-a)^3}.$$

Sie ist positiv sowohl für $x < 0$ als auch für $x > a$, so daß die Kurve stets (von unten) konvex ist.

Berechnen wir noch die Ordinate $y = 2{,}60a$, die sich für $x = \frac{3}{2}a$ ergibt, so reichen die Angaben bereits zur graphischen Darstellung aus (Abb. 80).

8. $y = \sqrt{\dfrac{a^3 - x^3}{3x}} \qquad (a > 0).$

Die Veränderliche x besitzt nur den Wertebereich $(0, a]$; für $x = 0$ wird die Funktion ∞.

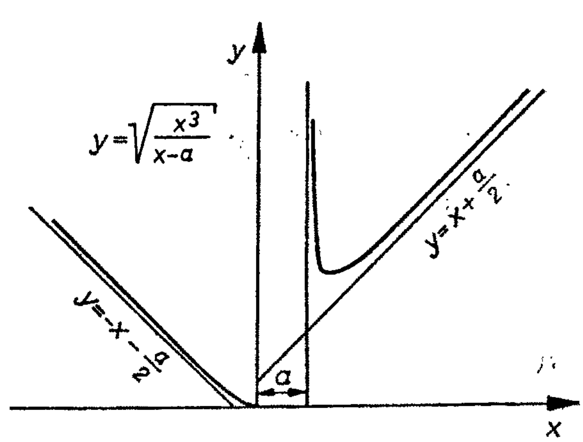

Abb. 80

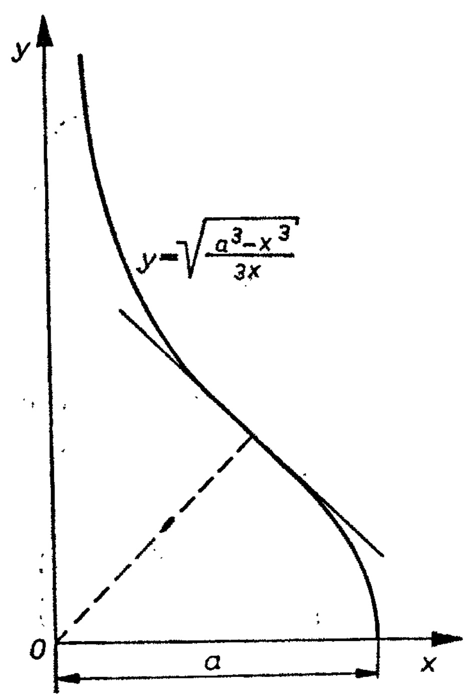

Abb. 81

Die Ableitung

$$y' = -\frac{a^3 + 2x^3}{6x^2 y} = -\frac{1}{2}\left(\frac{x}{y} + \frac{y}{x}\right)$$

ist stets negativ, so daß die Funktion fallend ist. An der Stelle $x = a$ gilt $y' = -\infty$.

Die zweite Ableitung

$$y'' = \frac{1}{2}(y - xy')\left(\frac{1}{x^2} - \frac{1}{y^2}\right)$$

verschwindet nur für $y = x = \frac{a}{\sqrt[3]{4}} \approx 0{,}63a$, wobei sie ihr Vorzeichen ändert (Wendepunkt). Hierbei gilt offenbar $y' = -1$. Abb. 81 zeigt die Kurve.

§ 4. Auswertung unbestimmter Ausdrücke

150. Der unbestimmte Ausdruck der Form $\frac{0}{0}$. Wir wenden jetzt den Begriff der Ableitung und die in § 3 und § 5 des vorhergehenden Kapitels bewiesenen Sätze zur Auswertung unbestimmter Ausdrücke an.

Die folgenden Sätze 1 bis 4 stammen im wesentlichen von G. F. A. de l'Hospital[1]) und Joh. Bernoulli. Die darin zum Ausdruck kommenden Regeln werden gewöhnlich nach de l'Hospital benannt.

Zunächst beschäftigen wir uns mit dem grundlegenden Fall des unbestimmten Ausdrucks der Form $\frac{0}{0}$, d. h., wir untersuchen den Grenzwert des Quotienten zweier gegen 0 strebender Funktionen $f(x)$ und $g(x)$ (für gegen a strebendes x). Wir beginnen mit einem einfachen Satz, der unmittelbar den Begriff der Ableitung benutzt.

Satz 1. *Die Funktionen $f(x)$ und $g(x)$ seien im Intervall $[a, b]$ definiert, es sei $\lim\limits_{x\to a} f(x) = 0$, $\lim\limits_{x\to a} g(x) = 0$; es seien ferner die endlichen (rechtsseitigen) Ableitungen $f'(a)$ und $g'(a)$ vorhanden, und es sei $g'(a) \neq 0$. Dann gilt*

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{f'(a)}{g'(a)}.$$

Beweis. Die Existenz der endlichen Ableitungen $f'(a)$, $g'(a)$ gewährleistet die rechtsseitige Stetigkeit der Funktionen $f(x)$ und $g(x)$ im Punkt a. Somit gilt $f(a) = \lim\limits_{x\to a} f(x) = 0$ und $g(a) = \lim\limits_{x\to a} g(x) = 0$. Da $g'(a) \neq 0$ ist, gilt nach dem Lemma aus Nr. 109 auch für solche x-Werte, die hinreichend nahe bei a liegen, $g(x) \neq 0$. Auf diese Werte beschränken wir uns, so daß der Ausdruck $\frac{f(x)}{g(x)}$ sinnvoll ist.

Nunmehr können wir diesen Quotienten in der Form

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(a)}{g(x) - g(a)} = \frac{\dfrac{f(x) - f(a)}{x - a}}{\dfrac{g(x) - g(a)}{x - a}}$$

[1]) Guillaume François Antoine de l'Hospital, 1661—1704, französischer Mathematiker.

schreiben. Gehen wir hier für $x \to a$ zur Grenze über, so erhalten wir das Gewünschte.

Beispiele.

1. Man bestimme den Grenzwert

$$\lim_{x\to 0} \frac{e^x - e^{-x}}{\ln(e - x) + x - 1}.$$

Nach Satz 1 ist er gleich dem Quotienten der Ableitungen im Punkt $x = 0$;

$$\left.\frac{e^x + e^{-x}}{1 - \dfrac{1}{e - x}}\right|_{x=0} = \frac{2}{1 - \dfrac{1}{e}} = \frac{2e}{e - 1}.$$

2. Man bestimme

$$\lim_{x\to 1} \frac{\sqrt{2x - x^4} - \sqrt[3]{x}}{1 - \sqrt[4]{x^3}}.$$

Es ergibt sich

$$\left.\frac{\dfrac{1 - 2x^3}{\sqrt{2x - x^4}} - \dfrac{1}{3\sqrt[3]{x^2}}}{-\dfrac{3}{4\sqrt[4]{x}}}\right|_{x=1} = \frac{16}{9}.$$

Ist gleichzeitig auch noch $f'(a) = 0$ und $g'(a) = 0$, so kann man folgende Verallgemeinerung des Satzes 1 benutzen, in welcher Ableitungen höherer Ordnung auftreten.

Satz 2. *Die Funktionen $f(x)$ und $g(s)$ seien im Intervall $[a, b]$ definiert, es sei $\lim\limits_{x\to a} f(x) = 0$, $\lim\limits_{x\to a} g(x) = 0$, und es mögen im Intervall $[a, b]$ endliche Ableitungen aller Ordnungen bis zur $(n - 1)$-ten einschließlich existieren,*

$$f'(x), f''(x), \ldots, f^{(n-1)}(x), g'(x), g''(x), \ldots, g^{(n-1)}(x),$$

und für $x = a$ seien alle diese Ableitungen gleich 0. Ferner seien die endlichen Ableitungen $f^{(n)}(a)$ und $g^{(n)}(a)$ vorhanden, und es gelte $g^{(n)}(a) \neq 0$. Dann ist

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{f^{(n)}(a)}{g^{(n)}(a)}.$$

Beweis. Wir wenden auf jede der Funktionen $f(x)$ und $g(x)$ im Intervall $[a, x]$, $a < x \leqq b$, die Taylorsche Formel mit dem Peanoschen Restglied an [vgl. Nr. 124, Formel (10a)]. Dann erhalten wir

$$f(x) = \frac{f^{(n)}(a) + \alpha}{n!}(x - a)^n, \quad g(x) = \frac{g^{(n)}(a) + \beta}{n!}(x - a)^n,$$

wobei α und β für $x \to a$ gegen 0 streben.

Wegen $g^{(n)}(a) \neq 0$ geht aus der zweiten dieser Gleichungen zunächst hervor, daß $g(x)$ wenigstens für solche x-Werte von 0 verschieden ist, die hinreichend nahe bei a liegen. Beschränken wir uns auf diese Werte, so ist der Quotient $\dfrac{f(x)}{g(x)}$ sinnvoll.

Dann erhalten wir aus den obigen Gleichungen unmittelbar das gewünschte Ergebnis:

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f^{(n)}(a) + \alpha}{g^{(n)}(a) + \beta} = \frac{f^{(n)}(a)}{g^{(n)}(a)}.$$

Beispiel.

3. Man bestimme

$$\lim_{x \to a} \frac{e^x - e^{-x} - 2x}{x - \sin x}.$$

Hier gilt

$$\begin{array}{llll} f(x) = e^x - e^{-x} - 2x, & f(0) = 0; & g(x) = x - \sin x, & g(0) = 0; \\ f'(x) = e^x + e^{-x} - 2, & f'(0) = 0; & g'(x) = 1 - \cos x, & g'(0) = 0; \\ f''(x) = e^x - e^{-x}, & f''(0) = 0; & g''(x) = \sin x, & g''(0) = 0; \\ f'''(x) = e^x - e^{-x}, & f'''(0) = 2; & g'''(x) = \cos x, & g'''(0) = 1. \end{array}$$

Folglich ist der gesuchte Grenzwert gleich 2.

Obgleich für die meisten Fälle die bewiesenen Sätze zur Auswertung unbestimmter Ausdrücke der Form $\frac{0}{0}$ ausreichen, ist in der Praxis häufig der folgende Satz zweckmäßiger.

Satz 3. *Die Funktionen $f(x)$ und $g(x)$ seien im Intervall $(a, b]$ definiert, es sei $\lim_{x \to a} f(x) = 0$, $\lim_{x \to a} g(x) = 0$; im Intervall $(a, b]$ mögen die endlichen Ableitungen $f'(x)$ und $g'(x)$ vorhanden und $g'(x) \neq 0$ sein; schließlich sei der (endliche oder nicht endliche) Grenzwert*

$$\lim_{x \to a} \frac{f'(x)}{g'(x)} = K$$

vorhanden. Dann gilt auch

$$\lim_{x \to a} \frac{f(x)}{g(x)} = K.$$

Beweis. Wir ergänzen die Definitionen von $f(x)$ und $g(x)$ dadurch, daß wir diese Funktionen für $x = a$ gleich 0 setzen: $f(a) = g(a) = 0$.[1]) Dann sind diese Funktionen im ganzen abgeschlossenen Intervall $[a, b]$ stetig: Ihre Werte im Punkt a stimmen mit den Grenzwerten für $x \to a$ überein, in den anderen Punkten ist die Stetigkeit durch die Existenz der endlichen Ableitungen gewährleistet. Durch Anwendung des zweiten Mittelwertsatzes (Nr. 114) erhalten wir

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(a)}{g(x) - g(a)} = \frac{f'(c)}{g'(c)},$$

wobei $a < c < x$ ist. Daß $g(x) \neq 0$, also $g(x) \neq g(a)$ ist, folgt aus der Voraussetzung $g'(x) \neq 0$ und war damals beim Beweis in Nr. 114 gezeigt worden.

[1]) Natürlich könnte man die Funktionen für $x = a$ als definiert und stetig voraussetzen; bei Anwendungen ist jedoch die oben gegebene Formulierung der Voraussetzungen des Satzes oft geeigneter (vgl. etwa Satz 3*).

Wenn $x \to a$ gilt, strebt offenbar auch c gegen a, so daß sich

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{c \to a} \frac{f'(c)}{g'(c)} = K$$

ergibt, was zu beweisen war.

Somit führt dieser Satz den Grenzwert des Quotienten von Funktionen auf den Grenzwert des Quotienten der Ableitungen zurück, falls dieser Grenzwert existiert.

Oft ist es leichter, den Grenzwert des Quotienten der Ableitungen zu bestimmen, vielfach sogar mit elementaren Mitteln.

Beispiel.

4. Man bestimme

$$\lim_{x \to 0} \frac{\tan x - x}{x - \sin x}.$$

Für den Quotienten der Ableitungen ergibt sich

$$\frac{\dfrac{1}{\cos^2 x} - 1}{1 - \cos x} = \frac{1}{\cos^2 x} \cdot \frac{1 - \cos^2 x}{1 - \cos x} = \frac{1 + \cos x}{\cos^2 x};$$

für $x \to 0$ strebt dieser Quotient offenbar gegen 2. Dies ist nach dem Satz der gesuchte Grenzwert.

Der Satz 1 ist in diesem Fall nicht anwendbar, da für $x = 0$ sowohl die Ableitung des Zählers als auch die des Nenners gleich 0 ist. Nun könnte man zwar das Problem mit Hilfe des Satzes 2 lösen, doch müßte man dabei drei aufeinanderfolgende Ableitungen der gegebenen Funktionen berechnen.

Wir machen den Leser darauf aufmerksam, daß hier auch der Quotient der Ableitungen wieder einen unbestimmten Ausdruck der Form $\frac{0}{0}$ ergab. Dieser ließ sich jedoch mit elementaren Umformungen auswerten. In anderen Fällen muß man unter Umständen den Satz 3 wiederholt anwenden. (Auf Grund der Voraussetzungen ist das etwas anderes als die Anwendung von Satz 2.) Es sei betont, daß dabei jede Vereinfachung der erhaltenen Ausdrücke, wie Kürzen gemeinsamer Faktoren, Benutzung bereits bekannter Grenzwerte u. dgl., zulässig ist (bei Satz 2 ist das nicht der Fall!).

Im folgenden Beispiel wird Satz 3 dreimal nacheinander angewandt. Nach der ersten Anwendung kürzen wir durch e^x, nach der zweiten lassen wir den Faktor e^x im Nenner weg (da er gegen 1 strebt). Dadurch vereinfachen sich die Rechnungen wesentlich.

Beispiele.

5. $$\lim_{x \to 0} \frac{xe^{2x} + xe^x - 2e^{2x} + 2e^x}{(e^x - 1)^3} = \lim_{x \to 0} \frac{2xe^{2x} + e^{2x} + xe^x + e^x - 4e^{2x} + 2e^x}{3(e^x - 1)^2 \cdot e^x}$$
$$= \lim_{x \to 0} \frac{2xe^x - 3e^x + 3 + x}{3(e^x - 1)^2}$$
$$= \frac{1}{3} \lim_{x \to 0} \frac{2xe^x + 2e^x - 3e^x + 1}{(2e^x - 1)\, e^x}$$
$$= \frac{1}{6} \lim_{x \to 0} \frac{-e^x + 2xe^x + 1}{e^x - 1} = \frac{1}{6} \lim_{x \to 0} \frac{2xe^x + e^x}{e^x} = \frac{1}{6}.$$

6. $$\lim_{x \to 0} \frac{(1 + x)^{1/x} - e}{x} = \lim_{x \to 0} (1 + x)^{1/x} \frac{x - (1 + x) \ln (1 + x)}{x^2(1 + x)}.$$

Da wir schon wissen, daß der erste Faktor auf der rechten Seite gegen e strebt, genügt es, den zweiten Faktor zu untersuchen. Wir wenden Satz 3 zweimal an und finden den Grenzwert $-\frac{1}{2}$. Der gesuchte Grenzwert ist also $-\frac{e}{2}$.

Der Satz 3 läßt sich leicht auf den Fall erweitern, daß x gegen einen unendlichen Grenzwert strebt, $a \to \pm\infty$ (was natürlich für die Sätze 1 und 2 nicht geht). Es gilt nämlich

Satz 3*. *Die Funktionen $f(x)$ und $g(x)$ seien im Intervall $[c, \infty)$ definiert, wobei $c > 0$ ist; es sei $\lim\limits_{x\to\infty} f(x) = 0$, $\lim\limits_{x\to\infty} g(x) = 0$, und es mögen im Intervall $[c, \infty)$ endliche Ableitungen $f'(x)$ und $g'(x)$ existieren mit $g'(x) \neq 0$. Schließlich existiere der (endliche oder nicht endliche) Grenzwert*

$$\lim_{x\to\infty} \frac{f'(x)}{g'(x)} = K.$$

Dann gilt auch

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = K.$$

Beweis. Wir setzen $x = \frac{1}{t}$, $t = \frac{1}{x}$. Für $x \to \infty$ gilt dann $t \to +0$ und umgekehrt. Nach Voraussetzung gilt

$$\lim_{t\to+0} f\left(\frac{1}{t}\right) = 0, \quad \lim_{t\to+0} g\left(\frac{1}{t}\right) = 0$$

sowie

$$\lim_{t\to+0} \frac{f'\left(\frac{1}{t}\right)}{g'\left(\frac{1}{t}\right)} = K.$$

Auf die Funktionen $f\left(\frac{1}{t}\right)$ und $g\left(\frac{1}{t}\right)$ der neuen Veränderlichen t läßt sich Satz 3 anwenden, was

$$\lim_{t\to+0} \frac{f\left(\frac{1}{t}\right)}{g\left(\frac{1}{t}\right)} = \lim_{t\to+0} \frac{f'\left(\frac{1}{t}\right)\cdot\left(-\frac{1}{t^2}\right)}{g'\left(\frac{1}{t}\right)\cdot\left(-\frac{1}{t^2}\right)} = \lim_{t\to+0} \frac{f'\left(\frac{1}{t}\right)}{g'\left(\frac{1}{t}\right)} = K$$

liefert;[1]) somit ist auch

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = K,$$

was zu beweisen war.

[1]) Wir differenzieren die Funktionen $f\left(\frac{1}{t}\right)$ und $g\left(\frac{1}{t}\right)$ als mittelbare Funktion nach t.

Bemerkung. Manchmal läßt sich bei der Auswertung unbestimmter Ausdrücke der betrachteten Gestalt formal die Anwendung der obigen Sätze umgehen, indem man die Taylorsche Reihenentwicklung (Nr. 124 und 125) benutzt. Es gelte also $x \to 0$ (auf diesen Fall läßt sich die Sache immer zurückführen).

Gelingt es, mit Hilfe bekannter Entwicklungen in Zähler und Nenner die Hauptglieder abzuspalten,

$$f(x) = ax^n + o(x^n), \quad g(x) = bx^m + o(x^m),$$

so läßt sich sofort der Grenzwert des Bruches $\dfrac{f(x)}{g(x)}$ angeben. Er ist gleich 0, $\dfrac{a}{b}$ oder $\pm\infty$, je nachdem, ob n größer, gleich oder kleiner als m ist;[1] vgl. Nr. 62 und 63.

So erhalten wir in Beispiel 1, wenn wir die Funktionen e^x, e^{-x} und $\ln(e - x) - 1 = \ln\left(1 - \dfrac{x}{e}\right)$ durch die ersten Glieder ihrer Entwicklung ersetzen,

$$\lim_{x\to 0} \frac{(1 + x + \cdots) - (1 - x + \cdots)}{\left(-\dfrac{x}{e} + \cdots\right) + x} = \lim_{x\to 0} \frac{2x + \cdots}{\left(1 - \dfrac{1}{e}\right)x + \cdots} = \frac{2e}{e - 1}.$$

Analog ergibt sich in Beispiel 4

$$\lim_{x\to 0} \frac{\left(x + \dfrac{x^3}{3} + \cdots\right) - x}{x - \left(x - \dfrac{x^3}{6} + \cdots\right)} = \lim_{x\to 0} \frac{\dfrac{x^3}{3} + \cdots}{\dfrac{x^3}{6} + \cdots} = 2.$$

Der Leser möge die Beispiele 3 und 5 als Übungsaufgabe nach der gleichen Methode behandeln.

151. Der unbestimmte Ausdruck der Form $\dfrac{\infty}{\infty}$. Wir wollen jetzt Ausdrücke der Form $\dfrac{\infty}{\infty}$ untersuchen, d. h. den Grenzwert des Quotienten zweier Funktionen $f(x)$ und $g(x)$, die beide gegen ∞ (für $x \to a$) streben.

Wir zeigen, daß auch in diesem Fall die l'Hospitalsche Regel anwendbar ist. Der folgende Satz ist eine einfache Umformulierung von Satz 3.

Satz 4. *Die Funktionen $f(x)$ und $g(x)$ seien im Intervall $(a, b]$ definiert, und es sei $\lim\limits_{x\to a} f(x) = \infty$, $\lim\limits_{x\to a} g(x) = \infty$. Die endlichen Ableitungen $f'(x)$ und $g'(x)$ seien im Intervall $(a, b]$ vorhanden, und es gelte $g'(x) \neq 0$. Schließlich sei der (endliche oder nicht endliche) Grenzwert*

$$\lim_{x\to a} \frac{f'(x)}{g'(x)} = K$$

vorhanden. Dann gilt auch

$$\lim_{x\to a} \frac{f(x)}{g(x)} = K.$$

[1]) Im letzten Fall ist das Vorzeichen auf Grund der Vorzeichen von a und b leicht zu ermitteln, es ist aber auch (bei *ungeradem* $m - n$) das Vorzeichen von x zu berücksichtigen.

Beweis. Wir betrachten zunächst den Fall, daß K endlich ist. Da $g'(x)$ nicht gleich 0 wird, behält diese Ableitung nach dem Satz von DARBOUX (Nr. 110) ihr Vorzeichen bei, also ändert sich $g(x)$ monoton (Nr. 132). Dann ist klar, daß $g'(x) < 0$ ist und $g(x)$ mit fallendem x monoton wachsend gegen ∞ strebt. Wir können daher annehmen, daß stets $g(x) > 0$ gilt.

Zu einem vorgegebenen $\varepsilon > 0$ läßt sich nach Voraussetzung eine Zahl $\eta > 0$ angeben, derart, daß für $a < x < a + \eta$ die Beziehung

$$\left|\frac{f'(x)}{g'(x)} - K\right| < \frac{\varepsilon}{2}$$

gilt. Wir setzen zur Abkürzung $a + \eta = x_0$ und wählen x zwischen a und x_0. Auf das Intervall $[x, x_0]$ wenden wir den zweiten Mittelwertsatz an:[1])

$$\frac{f(x) - f(x_0)}{g(x) - g(x_0)} = \frac{f'(c)}{g'(c)},$$

wobei $x < c < x_0$ ist. Also gilt

$$\left|\frac{f(x) - f(x_0)}{g(x) - g(x_0)} - K\right| < \frac{\varepsilon}{2}. \tag{1}$$

Wir schreiben nun die leicht zu verifizierende Identität

$$\frac{f(x)}{g(x)} - K = \frac{f(x_0) - K \cdot g(x_0)}{g(x)} + \left[1 - \frac{g(x_0)}{g(x)}\right]\left[\frac{f(x) - f(x_0)}{g(x) - g(x_0)} - K\right]$$

auf, aus der

$$\left|\frac{f(x)}{g(x)} - K\right| \leqq \left|\frac{f(x_0) - K \cdot g(x_0)}{g(x)}\right| + \left|\frac{f(x) - f(x_0)}{g(x) - g(x_0)} - K\right|$$

folgt. Für $x < x_0 = a + \eta$ wird der zweite Summand auf der rechten Seite auf Grund von (1) kleiner als $\frac{\varepsilon}{2}$. Da $g(x) \to \infty$ für $x \to a$ gilt, strebt der erste Summand hierbei gegen 0, und es läßt sich ein $\delta > 0$ angeben (man kann $\delta < \eta$ annehmen) derart, daß für $a < x < a + \delta$ auch der erste Summand kleiner als $\frac{\varepsilon}{2}$ wird. Für diese x-Werte erhalten wir dann

$$\left|\frac{f(x)}{g(x)} - K\right| < \varepsilon,$$

womit die Behauptung bewiesen ist.[2])

Wenn $K = \infty$ ist [und $f'(x) \neq 0$ wenigstens in der Umgebung von a ist], gilt, wenn wir die Rollen von f und g vertauschen,

$$\lim_{x \to a} \frac{g'(x)}{f'(x)} = 0, \quad \text{also auch} \quad \lim_{x \to a} \frac{g(x)}{f(x)} = 0,$$

[1]) Hierin besteht der wesentliche Unterschied gegenüber dem Beweis von Satz 3. Hier kann man den Mittelwertsatz nicht auf das Intervall $[a, x]$ anwenden. Denn wie man die Funktionen $f(x)$ und $g(x)$ im Punkt a auch definiert, stets erhält man Funktionen, die in a nicht stetig sind.

[2]) Wir betonen, daß wir bei unseren Überlegungen die Voraussetzung $\lim f(x) = \infty$ nicht benutzt haben (vgl. den Beweis des Stolzschen Satzes in Nr. 33).

woraus schließlich

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \infty$$

folgt, da (wenigstens in der Umgebung von a) offenbar auch $f(x) > 0$ und $g(x) > 0$ gilt.[1])

Der Beweis läßt sich ohne wesentliche Änderungen auch auf den Fall $a = -\infty$ übertragen. Ebenso könnte der Satz auch für das Intervall $[b, a)$, $b < a$, bewiesen werden, und zwar sowohl für endliches a als auch für $a = \infty$. Somit kann auch Satz 4 automatisch auf den Fall eines unendlichen Grenzwerts des Arguments übertragen werden.

Als Beispiel erhält man leicht die uns bereits bekannten Grenzwerte

$$7.\ \lim_{x\to\infty} \frac{\ln x}{x^\mu} = \lim_{x\to\infty} \frac{\frac{1}{x}}{\mu x^{\mu-1}} = \lim_{x\to\infty} \frac{1}{\mu x^\mu} = 0 \qquad (\text{für } \mu > 0).$$

$$8.\ \lim_{x\to\infty} \frac{x^\mu}{a^x} = \lim_{x\to\infty} \frac{\mu x^{\mu-1}}{a^x \cdot \ln a} \qquad (a > 1, \mu > 0).$$

Ist $\mu > 1$, so erhalten wir auf der rechten Seite wieder einen unbestimmten Ausdruck vom gleichen Typ $\frac{\infty}{\infty}$. Setzen wir aber diesen Prozeß fort und wenden Satz 4 wiederholt an, so erhalten wir schließlich im Zähler eine Potenz mit negativem (oder verschwindendem) Exponenten. Daher gilt in jedem Fall

$$\lim_{x\to\infty} \frac{x^\mu}{a^x} = 0.$$

Wir schließen noch eine allgemeine Bemerkung bezüglich der Sätze 3 (3*) und 4 an. In ihnen wird der Grenzwert eines Quotienten von Funktionen unter der Voraussetzung hergeleitet, daß der Grenzwert des Quotienten der Ableitungen existiert. Die Umkehrung dieser Sätze ist jedoch nicht richtig. Es ist durchaus möglich, daß der Grenzwert des Quotienten der Funktionen existiert, der Grenzwert des Quotienten der Ableitungen aber nicht.

Beispielsweise existiert der Grenzwert

$$\lim_{x\to\infty} \frac{x + \sin x}{x} = \lim_{x\to\infty} \left(1 + \frac{\sin x}{x}\right) = 1,$$

obgleich der Quotient der Ableitungen, der gleich $1 + \cos x$ ist, für $x \to \infty$ gar keinen Grenzwert besitzt.

152. Andere Formen unbestimmter Ausdrücke. Die bisher bewiesenen Sätze bezogen sich auf unbestimmte Ausdrücke der Form $\frac{0}{0}$ und $\frac{\infty}{\infty}$.

Hat man einen unbestimmten Ausdruck der Form $0 \cdot \infty$, so kann man ihn auf die Form $\frac{0}{0}$ oder $\frac{\infty}{\infty}$ zurückführen und dann die l'Hospitalsche Regel anwenden.

Es sei

$$\lim_{x\to a} f(x) = 0, \quad \lim_{x\to a} g(x) = \infty.$$

[1]) Der Fall $K = -\infty$ kann unter den Voraussetzungen des Satzes nicht eintreten.

Dann ist

$$f(x) \cdot g(x) = \frac{f(x)}{\frac{1}{g(x)}} = \frac{g(x)}{\frac{1}{f(x)}}.$$

Der erste dieser Quotienten ist für $x \to a$ ein unbestimmter Ausdruck der Form $\frac{0}{0}$, der zweite ein Ausdruck der Form $\frac{\infty}{\infty}$.

Beispiel.

9. $\lim\limits_{x\to+0} (x^{\mu} \cdot \ln x) = \lim\limits_{x\to+0} \frac{\ln x}{x^{-\mu}} = \lim\limits_{x\to+0} \frac{\frac{1}{x}}{-\mu x^{-\mu-1}} = \lim\limits_{x\to+0} \frac{x^{\mu}}{-\mu} = 0$
(wir setzen $\mu > 0$ voraus).

Auf die Form $\frac{0}{0}$ bzw. $\frac{\infty}{\infty}$ lassen sich auch stets unbestimmte Ausdrücke der Form $\infty - \infty$ zurückführen. Ist also der Ausdruck $f(x) - g(x)$ mit

$$\lim_{x\to a} f(x) = \infty, \quad \lim_{x\to a} g(x) = \infty$$

gegeben, so kann man beispielsweise folgende Umformung vornehmen, durch welche dieser Ausdruck auf die Form $\frac{0}{0}$ gebracht wird:

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

Übrigens läßt sich der Grenzwert oft auch einfacher gewinnen.

Beispiel.

10. Es gilt

$$\lim_{x\to0} \left(\cot^2 x - \frac{1}{x^2}\right) = \lim_{x\to0} \frac{x^2 \cdot \cos^2 x - \sin^2 x}{x^2 \cdot \sin^2 x},$$

es ist aber

$$\frac{x^2 \cdot \cos^2 x - \sin^2 x}{x^2 \cdot \sin^2 x} = \frac{x \cdot \cos x + \sin x}{x} \cdot \frac{x \cdot \cos x - \sin x}{x \cdot \sin^2 x}.$$

Der Grenzwert des ersten Faktors läßt sich elementar bestimmen:

$$\lim_{x\to0} \frac{x \cdot \cos x + \sin x}{x} = \lim_{x\to0} \left(\cos x + \frac{\sin x}{x}\right) = 2.$$

Auf den zweiten Faktor wenden wir Satz 3 an:

$$\lim_{x\to0} \frac{x \cdot \cos x - \sin x}{x \cdot \sin^2 x} = \lim_{x\to0} \frac{-x \cdot \sin x}{\sin^2 x + 2x \cdot \sin x \cdot \cos x} = \lim_{x\to0} \frac{-1}{\frac{\sin x}{x} + 2 \cos x} = -\frac{1}{3}.$$

Somit ist der gesuchte Grenzwert gleich $-\frac{2}{3}$.

Im Fall unbestimmter Ausdrücke der Form 1^∞, 0^0, ∞^0 empfiehlt es sich, diese Ausdrücke vorher zu logarithmieren.

Es sei $y = [f(x)]^{g(x)}$; dann ist $\ln y = g(x) \cdot \ln f(x)$. Der Grenzwert von $\ln y$ ist ein unbestimmter Ausdruck der bereits behandelten Form $0 \cdot \infty$. Nehmen wir an, es sei durch eine der oben angeführten Methoden gelungen, $\lim\limits_{x\to 0} \ln y$ zu bestimmen; er sei gleich einer endlichen Zahl k, ∞ bzw. $-\infty$. Dann ergibt sich für $\lim\limits_{x\to a} y$ der Wert e^k, ∞ bzw. 0.

Beispiele.

11. Es sei

$$y = \left(\frac{\sin x}{x}\right)^{1/(1-\cos x)}$$

Man bestimme $\lim y$ für $x \to 0$ (unbestimmter Ausdruck der Form 1^∞).

Setzen wir $x > 0$ voraus (darauf können wir uns, da y gerade ist, beschränken), so gilt

$$\ln y = \frac{\ln \sin x - \ln x}{1 - \cos x}.$$

Durch Anwendung von Satz 3 (und Benutzung der bereits in den vorigen Beispielen erhaltenen Ergebnisse) finden wir

$$\lim_{x\to 0} \ln y = \lim_{x\to 0} \frac{\dfrac{\cos x}{\sin x} - \dfrac{1}{x}}{\sin x} = \lim_{x\to 0} \frac{x \cdot \cos x - \sin x}{x \cdot \sin^2 x} = -\frac{1}{3}$$

und hieraus

$$\lim_{x\to 0} y = e^{-1/3} = \frac{1}{\sqrt[3]{e}}.$$

12. Es sei

$$y = \left(\frac{\pi}{2} - \arctan x\right)^{1/\ln x}$$

Für $x \to \infty$ wird hieraus ein unbestimmter Ausdruck der Form 0^0. Es gilt

$$\ln y = \frac{\ln\left(\dfrac{\pi}{2} - \arctan x\right)}{\ln x} \qquad \left(\text{Fall } \frac{\infty}{\infty}\right).$$

Nach der l'Hospitalschen Regel erhalten wir

$$\lim_{x\to\infty} \ln y = \lim_{x\to\infty} \frac{-\dfrac{1}{\dfrac{\pi}{2} - \arctan x} \cdot \dfrac{1}{1 + x^2}}{\dfrac{1}{x}}$$

$$= \lim_{x\to\infty} \frac{\dfrac{x}{1 + x^2}}{\arctan x - \dfrac{\pi}{2}} = \lim_{x\to\infty} \frac{\dfrac{1 - x^2}{(1 + x^2)^2}}{\dfrac{1}{1 + x^2}} = \lim_{x\to\infty} \frac{1 - x^2}{1 + x^2} = -1,$$

also $\lim\limits_{x\to\infty} y = \dfrac{1}{e}$.

§ 5. Die angenäherte Lösung von Gleichungen

153. Einführende Bemerkungen. Wir stellen uns jetzt die Aufgabe, die Nullstellen einer vorgegebenen Funktion $f(x)$, d. h. die Wurzeln (Lösungen) der Gleichung

$$f(x) = 0, \tag{1}$$

zu ermitteln. Wir werden übrigens dieses Problem unter der Voraussetzung behandeln, daß die uns interessierende Nullstelle ξ isoliert, d. h. in einem Intervall $[a, b]$,

$$a < \xi < b,$$

enthalten ist, in dem keine anderen Nullstellen liegen.

Wenn überdies $f(x)$ in den Endpunkten des Intervalls Werte $f(a)$ und $f(b)$ verschiedenen Vorzeichens besitzt, so ist es, wie in Nr. 81 im Zusammenhang mit der Anwendung des ersten Zwischenwertsatzes dargelegt wurde, durch sukzessive Zerlegung des die Nullstelle enthaltenden Intervalls und Bestimmung des Vorzeichens von $f(x)$ in den Teilungspunkten möglich, dieses Intervall beliebig zu verkleinern und in dieser Weise die Nullstelle angenähert zu berechnen.

Dieses Verfahren ist jedoch trotz seiner prinzipiellen Einfachheit in der Praxis wenig brauchbar, weil es zu viele Rechnungen erfordert. In diesem Paragraphen wird der Leser wesentlich einfachere Verfahren zur angenäherten Berechnung von (isolierten) Lösungen der Gleichung (1) kennenlernen, die systematischer und schneller zum Ziel führen. Hierbei werden wir wieder die grundlegenden Begriffe und Methoden der Differentialrechnung benutzen.

Wir werden dabei stets folgende Bedingungen als erfüllt voraussetzen:

1. Die Funktion $f(x)$ ist nebst ihren Ableitungen $f'(x)$ und $f''(x)$ im Intervall $[a, b]$ stetig.

2. Die Funktionswerte $f(a)$ und $f(b)$ an den Endpunkten des Intervalls haben verschiedene Vorzeichen: $f(a) \cdot f(b) < 0$.

3. Die Ableitungen $f'(x)$ und $f''(x)$ sind von 0 verschieden und behalten beide im gesamten Intervall $[a, b]$ ein bestimmtes Vorzeichen bei.

Aus der Stetigkeit von $f(x)$ und der Bedingung 2 folgt, daß zwischen a und b eine Lösung ξ der Gleichung (1) liegt (Nr. 80). Da $f'(x)$ das Vorzeichen beibehält (Bedingung 3), wächst oder fällt $f(x)$ im Intervall $[a, b]$ monoton und kann somit nur ein einziges Mal gleich 0 werden, d. h., die Lösung ist isoliert. Die Bedingung 3 besagt geometrisch, daß die Kurve von $y = f(x)$ nicht nur in einer Richtung verläuft [$f(x)$ ständig wächst oder ständig fällt], je nach dem Vorzeichen von $f'(x)$ (vgl. Nr. 132), sondern daß $f(x)$ überdies (streng) konvex von unten bzw. von oben ist, je nach dem Vorzeichen von $f''(x)$ (vgl. Nr. 143). Abb. 82 zeigt die vier möglichen Fälle, die den verschiedenen Kombinationen der Vorzeichen von $f'(x)$ und $f''(x)$ entsprechen. In der Algebra wird gezeigt, daß bei der Berechnung (reeller) Wurzeln algebraischer Gleichungen immer eine solche Sachlage geschaffen werden kann, daß die Bedingungen 1 bis 3 erfüllt sind. Sie schränken also in diesem Fall die Anwendbarkeit der zu entwickelnden Verfahren prinzipiell nicht ein. Dies kann man jedoch hinsichtlich transzendenter (d. h. nicht algebraischer) Gleichungen nicht sagen. In der Praxis sind jedoch unsere Einschränkungen wenig störend, da die oben formulierten Bedingungen in den meisten Fällen erfüllt sind.

154. Die Regula falsi (Sehnenmethode). Ist das Intervall $[a, b]$ hinreichend klein, so können wir mit einer gewissen Näherung annehmen, daß der Zuwachs von $f(x)$ dem Zuwachs des Arguments proportional ist (Mittelwertsatz), wenn x in diesen Grenzen variiert. Bezeichnen wir die Nullstelle der Funktion mit ξ, so gilt also insbesondere näherungsweise

$$\frac{f(\xi) - f(a)}{f(b) - f(a)} = \frac{\xi - a}{b - a},$$

woraus wegen $f(\xi) = 0$

$$\xi \approx a - \frac{(b - a) \cdot f(a)}{f(b) - f(a)}$$

folgt.

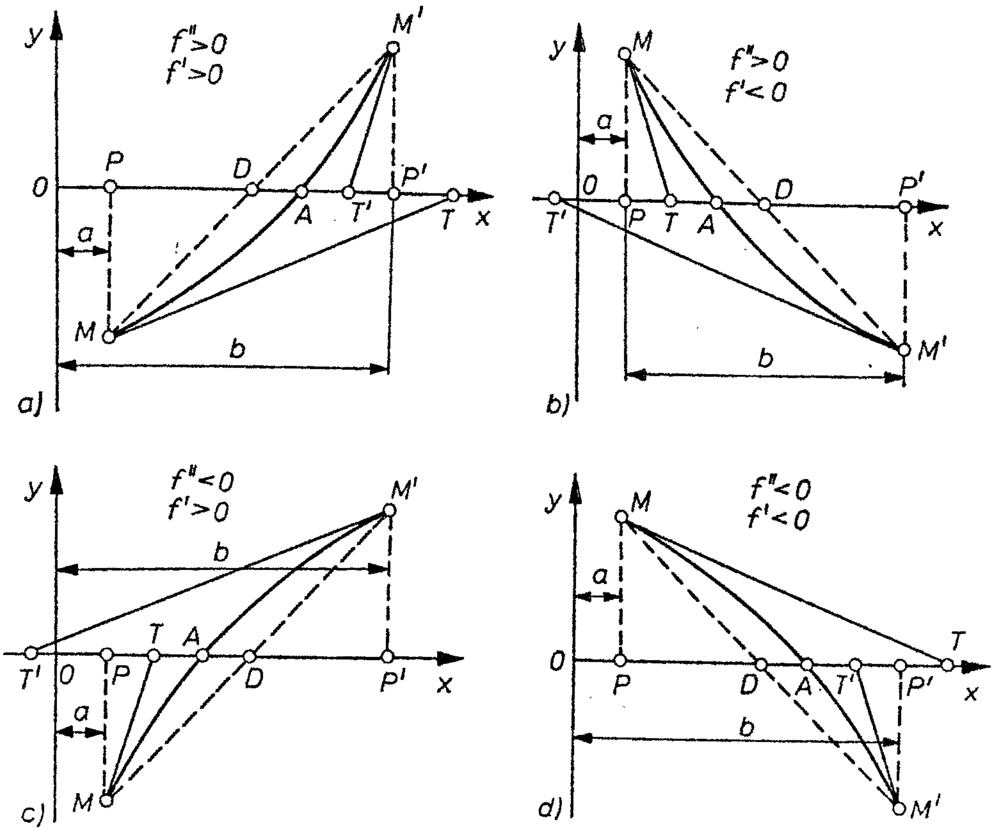

Abb. 82

Somit kann als Näherungswert der Nullstelle hier die Zahl

$$x_1 = a - \frac{(b-a)\cdot f(a)}{f(b)-f(a)} \tag{2}$$

angenommen werden. Diesen Ausdruck kann man offenbar auch in der Gestalt

$$x_1 = b - \frac{(b-a)\cdot f(b)}{f(b)-f(a)} \tag{2*}$$

schreiben. Diese Regel zur Ermittlung eines Näherungswertes einer Nullstelle nennt man *Regula falsi*[1]). Sie gestattet eine einfache geometrische Deutung. Wir ersetzen den Bogen $\widehat{MM'}$ der Kurve (Abb. 82) durch die Sehne MM'. Deren Gleichung kann z. B. in der Form

$$y - f(a) = \frac{f(b)-f(a)}{b-a}(x-a) \tag{3}$$

geschrieben werden. Unsere Regel beruht im Grunde darauf, daß an Stelle des Schnittpunktes A der Kurve mit der x-Achse der Schnittpunkt D dieser Sehne mit der x-Achse bestimmt wird. Tatsächlich erhält man, wenn man in (3) $y = 0$ setzt, für die Abszisse x_1 des Punktes D gerade den Ausdruck (2). Dementsprechend heißt dieses Näherungsverfahren auch *Sehnenmethode* (oder *Sekantenmethode*).

Wir wollen jetzt die Lage des Punktes x_1 in bezug auf die Nullstelle ξ untersuchen. Es ist unmittelbar klar, daß x_1 zwischen a und b liegt; es bleibt zunächst noch offen, auf welcher Seite von ξ sich x_1 befindet.

Da wir es in den Fällen a) und b) [bzw. c) und d)] mit einer von unten (von oben) konvexen Funktion zu tun haben, liegt die Kurve MM' unterhalb (oberhalb) der Sehne MM', d. h., es gilt

$$f(x) \underset{(>)}{<} f(a) + \frac{f(b)-f(a)}{b-a}(x-a) \qquad (a < x < b). \tag{4}$$

---

[1]) Diese Bezeichnung stammt daher, daß die Regel auf einer „falschen" (d. h. nicht den Tatsachen entsprechenden) Annahme beruht, nämlich der Annahme, daß die Zuwächse von Funktion und Argument einander proportional seien.

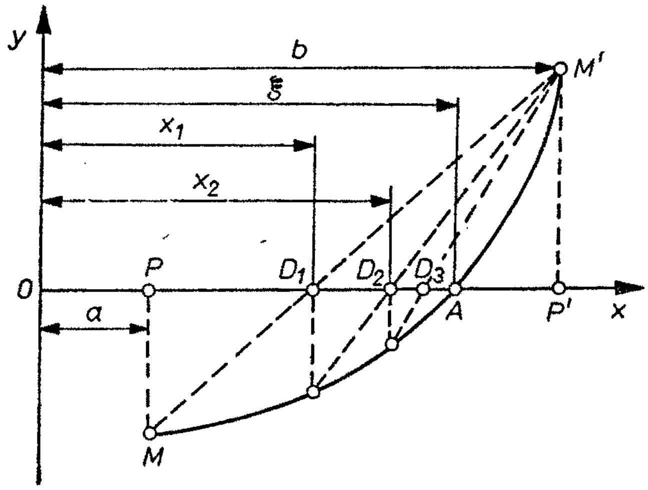

Abb. 83

Setzen wir hier $x = x_1$, so erhalten wir unmittelbar

$$f(x_1) \underset{(>)}{<} 0,$$

so daß $f(x_1)$ stets das dem Vorzeichen von $f''(x)$ entgegengesetzte Vorzeichen hat. Hieraus können wir schließen, daß der Wert x_1 in den Fällen a) und b) zwischen a und ξ, in den Fällen b) und c) dagegen zwischen ξ und b liegt.

Wir beschränken uns auf die Fälle a) und d) und wenden unsere Regel nochmals an, und zwar jetzt auf das Intervall $[x_1, b]$. Ersetzen wir in (2) den Wert a durch x_1, so erhalten wir einen neuen Näherungswert für die Nullstelle ξ,

$$x_2 = x_1 - \frac{(b - x_1) \cdot f(x_1)}{f(b) - f(x_1)},$$

der nach dem soeben Bewiesenen zwischen x_1 und ξ liegt. Dieses Verfahren kann beliebig oft wiederholt werden und liefert eine wachsende Folge von Näherungswerten:

$$a < x_1 < x_2 < \cdots < x_n < x_{n+1} < \cdots < \xi.$$

Dabei sind je zwei aufeinanderfolgende Werte x_n und x_{n+1} durch die zu (2) analoge Formel

$$x_{n+1} = x_n - \frac{(b - x_n) \cdot f(x_n)}{f(b) - f(x_n)} \tag{5}$$

verknüpft.

Wir zeigen, daß mit wachsendem n der Wert x_n gegen ξ strebt. Da nämlich die Veränderliche x_n monoton wachsend und beschränkt ist (z. B. durch die Zahl ξ), muß sie gegen einen bestimmten endlichen Grenzwert $\alpha \leqq \xi$ konvergieren. Gehen wir in Gleichung (5) zur Grenze über und benutzen dabei die Stetigkeit von $f(x)$, so erhalten wir

$$\frac{(b - \alpha) \cdot f(\alpha)}{f(b) - f(\alpha)} = 0,$$

also $f(\alpha) = 0$. Da es in $[a, b]$ außer ξ keine anderen Lösungen der Gleichung (1) gibt, ist $\alpha = \xi$.[1])

Abb. 83 veranschaulicht die stufenweise Annäherung der Schnittpunkte D_1, D_2, ... der aufeinanderfolgenden Sehnen mit der x-Achse an den gesuchten Punkt A.

Es ist leicht einzusehen, daß in den Fällen b) und c) die wiederholte Anwendung der Regel zu einer abnehmenden Folge von Näherungswerten

$$b > x_1 > x_2 > \cdots > x_n > x_{n+1} > \cdots > \xi$$

führt, die von rechts gegen ξ streben.

---

[1]) Die Konvergenz des Verfahrens läßt sich auch ohne die Voraussetzung über die zweite Ableitung beweisen; dann ist es aber nicht ausgeschlossen, daß die Punkte x_n von der einen Seite von ξ auf die andere überwechseln.

Somit kann man in allen Fällen die Nullstelle mit beliebiger Genauigkeit berechnen, wenn man die Regel hinreichend oft anwendet. Hierbei bleibt übrigens die Frage offen, wie man die Genauigkeit der berechneten Näherungswerte x_n abschätzt. Um sie zu beantworten, wenden wir auf die Differenz $f(x_n) - f(\xi)$ den Mittelwertsatz (Nr. 112) an:

$$f(x_n) - f(\xi) = (x_n - \xi) \cdot f'(c) \qquad (\xi \lesseqgtr c \lesseqgtr x_n).$$

Hieraus folgt

$$x_n - \xi = \frac{f(x_n)}{f'(c)}.$$

Bezeichnet man den kleinsten Wert von $|f'(x)|$ im betrachteten Intervall (den man ein für allemal im voraus berechnen kann) mit m, so erhält man

$$|x_n - \xi| \leqq \frac{|f(x_n)|}{m}. \tag{6}$$

Somit liefert der Wert von $f(x_n)$ selbst eine Genauigkeitsschranke für die Güte der Approximation.

Nun betrachten wir ein Beispiel. Die Gleichung

$$x^3 - 2x^2 - 4x - 7 = 0$$

besitzt eine Wurzel zwischen 3 und 4. Bezeichnen wir nämlich ihre linke Seite mit $f(x)$, so erhalten wir

$$f(3) = -10 < 0, \qquad f(4) = 9 > 0.$$

Wir wollen diese Wurzel bis auf 0,01 genau berechnen. Im Intervall [3, 4] behalten beide Ableitungen

$$f'(x) = 3x^2 - 4x - 4, \qquad f''(x) = 6x - 4$$

ihr positives Vorzeichen bei (Fall a); der kleinste Wert der ersten ist $m = 11$.

Wir haben also

$$x_1 = 3 - \frac{f(3)}{f(4) - f(3)} = 3 + \frac{10}{19} = 3 + 0{,}52\ldots,$$

abgerundet $x_1 = 3{,}52$. Wegen $f(3{,}52) = -2{,}246592$ ist nach Ungleichung (6) die geforderte Genauigkeit noch nicht erreicht.

Wir setzen die Berechnung fort:

$$x_2 = 3{,}52 - \frac{0{,}48 \cdot f(3{,}52)}{f(4) - f(3{,}52)} = 3{,}52 + \frac{1{,}078364\,16}{11{,}246592} = 3{,}52 + 0{,}09\ldots,$$

abgerundet $x_2 = 3{,}61$. Nun ist $f(3{,}61) = -0{,}458319$; benutzen wir Ungleichung (6), so stellen wir fest, daß das Ziel immer noch nicht erreicht ist. Schließlich ist

$$x_3 = 3{,}61 - \frac{0{,}39 \cdot f(3{,}61)}{f(4) - f(3{,}61)} = 3{,}61 + \frac{0{,}178744\,41}{9{,}458319} = 3{,}61 + 0{,}0188\ldots$$

Aufgerundet ergibt dies $x_3 = 3{,}63$. Da wir „auf der Seite der Nullstelle" aufgerundet haben, könnten wir damit die Nullstelle bereits überschritten haben. Daß dies nicht der Fall ist, zeigt $f(3{,}63) = -0{,}041653$. Diesmal liefert uns Ungleichung (6) jedoch

$$|x_3 - \xi| = \xi - x_3 < \frac{0{,}041\ldots}{11} < 0{,}004.$$

Somit ist

$$3{,}630 < \xi < 3{,}634, \qquad \xi = 3{,}63(+0{,}004).$$

Auf dieses eine Beispiel wollen wir uns beschränken, da diese Sehnenmethode noch wenig zweckmäßig ist. Man zieht die Tangentenmethode, zu der wir jetzt übergehen, vor.

155. Die Newtonsche Regel (Tangentenmethode). Wir benutzen wieder die obigen Voraussetzungen (Nr. 153) bezüglich der Funktion $f(x)$. Die gesuchte Nullstelle ξ dieser Funktion sei isoliert im Intervall $[a, b]$: $a < \xi < b$. Wir gehen von einem der Endpunkte dieses Intervalls, beispielsweise von b, aus und schreiben die Taylorsche Formel mit dem Lagrangeschen Restglied an:

$$0 = f(\xi) = f(b) + f'(b) \cdot (\xi - b) + \frac{1}{2} f''(c) \cdot (\xi - b)^2 \qquad (\xi < c < b). \tag{7}$$

Lassen wir das Restglied unberücksichtigt, so können wir näherungsweise

$$f(b) + f'(b) \cdot (\xi - b) \approx 0$$

setzen, woraus

$$\xi \approx b - \frac{f(b)}{f'(b)}$$

folgt. Auf diesem Wege gelangen wir zu einem Näherungswert der Nullstelle ξ,

$$x_1' = b - \frac{f(b)}{f'(b)}. \tag{8}$$

Dieser Ausdruck läßt sich ebenfalls geometrisch veranschaulichen. Wir betrachten die im Punkt M' mit der Abszisse b an die Kurve $y = f(x)$ gelegte Tangente. Ihre Gleichung hat die Form

$$y - f(b) = f'(b) \cdot (x - b).$$

Setzen wir hier $y = 0$, so finden wir die Abszisse des Schnittpunktes T' der Tangente mit der x-Achse, also genau (8). Das Wesen der Sache besteht also darin, daß der Kurvenbogen $\widehat{MM'}$ näherungsweise durch die in einem der Endpunkte des Intervalls an die Kurve gelegte Tangente (vgl. Abb. 82) ersetzt wird. Dieses nach NEWTON[1]) benannte Verfahren wird daher auch *Tangentenmethode* genannt.

Es erhebt sich jedoch die Frage, wo der nach Formel (8) erhaltene Wert x_1' liegt.

Aus Abb. 82 ist ersichtlich, daß der Schnittpunkt der Tangente mit der x-Achse sogar außerhalb des betrachteten Intervalls liegen kann. Wir beweisen nun folgendes: *Hat der Wert $f(b)$ das gleiche Vorzeichen wie $f''(x)$* [d. h. in den Fällen a) und d)], *dann liegt x_1' zwischen ξ und b.*

Da $f(b)$ und $f'(b)$ gleiches Vorzeichen haben, folgt aus (8) unmittelbar, daß $x_1' < b$ ist. Andererseits folgt aus (7) und (8)

$$\xi - x_1' = \xi - b + \frac{f(b)}{f'(b)} = -\frac{1}{2} \cdot \frac{f''(c)}{f'(b)} (\xi - b)^2. \tag{9}$$

In den betrachteten Fällen besitzt $f''(x)$ das gleiche Vorzeichen wie $f'(x)$, folglich ist $\xi < x_1'$. Daher gilt schließlich $\xi < x_1' < b$.

Analog erhalten wir an Stelle von (8) den Näherungswert

$$x_1' = a - \frac{f(a)}{f'(a)}, \tag{8*}$$

wenn wir vom Punkt a ausgehen und die Tangente im Endpunkt M (mit der Abszisse a) an die Kurve legen. Hinsichtlich des nach Formel (8*) errechneten Wertes kann man wie oben folgendes feststellen: *Hat der Wert $f(a)$ das gleiche Vorzeichen wie $f''(x)$* [d. h. in den Fällen b) und c)], *so liegt x_1' zwischen a und ξ.*

Somit ist für jeden der vier möglichen Fälle gezeigt, von welchem Endpunkt aus eine erfolgreiche Approximation der Nullstelle nach der Newtonschen Regel gewährleistet ist.

Die wiederholte Anwendung dieser Regel liefert in den Fällen a) und d) eine fallende Folge von Werten:

$$b > x_1' > x_2' > \cdots > x_n' > x_{n+1}' > \cdots > \xi,$$

---

[1]) ISAAC NEWTON, 1643—1727, englischer Mathematiker und Physiker.

in den Fällen b) und c) dagegen eine wachsende Folge von Werten:

$$a < x'_1 < x'_2 < \cdots < x'_n < x'_{n+1} < \cdots < \xi,$$

wobei jeder Wert aus dem vorhergehenden nach der Formel

$$x'_{n+1} = x'_n - \frac{f(x'_n)}{f'(x'_n)} \tag{10}$$

berechnet wird.

Auch hier läßt sich leicht beweisen, daß x'_n gegen ξ strebt. Die monotone und beschränkte Veränderliche x'_n besitzt einen endlichen Grenzwert β. Fehen wir in (10) zur Grenze über und berücksichtigen dabei die Stetigkeit der beiden Funktionen $f(x)$ und $f'(x)$, so finden wir $\frac{f(\beta)}{f'(\beta)} = 0$, also $f(\beta) = 0$ und $\beta = \xi$.

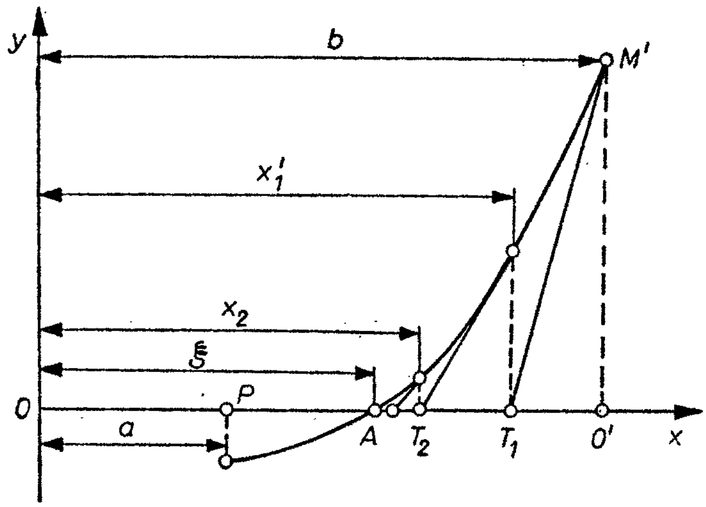

Abb. 84

Abb. 84 veranschaulicht die Annäherung der Schnittpunkte $T_1, T_2, \ldots$ der aufeinanderfolgenden Tangenten mit der x-Achse an den Punkt A.

Somit gestattet auch die Newtonsche Regel durch wiederholte Anwendung, die Nullstelle ξ mit beliebiger Genauigkeit zu berechnen. Dabei läßt sich die Genauigkeit der bereits berechneten Näherungswerte wie oben nach Formel (6) abschätzen.

Um die Geschwindigkeit zu charakterisieren, mit der die Differenzen $x_n - \xi$ abnehmen, kehren wir zur Formel (9) zurück. Darin ersetzen wir b und x'_n und x'_1 durch x'_{n+1}:

$$x'_{n+1} - \xi = \frac{1}{2} \cdot \frac{f''(c)}{f'(x'_n)} (x' - \xi)^2.$$

Bezeichnet man den größten Wert von $|f''(x)|$ in dem gegebenen Intervall $[a, b]$ mit M [und mit m wieder den kleinsten Wert von $|f'(x)|$], so erhält man hieraus

$$|x'_{n+1} - \xi| \leqq \frac{M}{2m} \cdot |x'_n - \xi|^2. \tag{11}$$

Da auf der rechten Seite ein Quadrat steht, ist eine äußerst schnelle Annäherung von x'_n an ξ (wenigstens von einer bestimmten Stelle an) gewährleistet, wodurch die Tangentenmethode eines der wirkungsvollsten Verfahren zur näherungsweisen Berechnung von Nullstellen wird.

Die Ungleichung (11) leistet noch etwas mehr. Wenn die Genauigkeit eines berechneten Wertes x'_n bereits abgeschätzt ist, z. B. mit Hilfe der Ungleichung (6), so gestattet die Ungleichung (11), die Genauigkeit des noch nicht berechneten Wertes x'_{n+1} abzuschätzen. Dies erweist sich als nützlich, wenn man wissen will, ob man zweckmäßigerweise abrundet oder aufrundet.

Wir wenden uns jetzt Beispielen zu. Ihre Lösung setzt die Anwendung aller zur Verfügung stehenden Rechenhilfsmittel voraus.

156. Beispiele und Übungen. Hier werden wir ausschließlich die Tangentenmethode benutzen.

1. Die Wurzel der Gleichung

$$x^3 - 2x^2 - 4x - 7 = 0$$

ist bis auf 0,01 genau zu berechnen. Als bekannt wird vorausgesetzt, daß sie im Intervall (3, 4) enthalten ist (vgl. Nr. 154). Es ist

$$f(x) = x^3 - 2x^2 - 4x - 7, \qquad f(3) = -10 < 0, \qquad f(4) = +9 > 0,$$

$$f'(x) = 3x^2 - 4x - 4 > 0, \qquad f''(x) = 6x - 4 > 0 \qquad (\text{für } 3 \leqq x \leqq 4)$$

[Fall a)]. Der kleinste Wert von $|f'(x)|$ ist $m = 11$.

Wir gehen von dem Endpunkt des Intervalls $b = 4$ aus, für den das Vorzeichen von $f(x)$ mit dem von $f''(x)$ übereinstimmt. Nach (8) ist

$$x_1' = 4 - \frac{f(4)}{f'(4)} = 4 - \frac{9}{28} = 4 - 0{,}32\ldots,$$

aufgerundet $x_1' = 4 - 0{,}3 = 3{,}7$. Da $f(x_1') = f(3{,}7) = 1{,}473$ ist, gilt nach Ungleichung (6) $x_1' - \xi < \frac{1{,}473}{11} < 0{,}14$, d. h., die erreichte Genauigkeit genügt noch nicht.

Weiter erhalten wir

$$x_2' = 3{,}7 - \frac{f(3{,}7)}{f'(3{,}7)} = 3{,}7 - \frac{1{,}473}{22{,}27} = 3{,}7 - 0{,}066\ldots;$$

wir setzen $x_2' = 3{,}7 - 0{,}066 = 3{,}634$. Jetzt ist $f(x_2') = f(3{,}634) = 0{,}042\ldots$, also nach (6) $x_2' - \xi < \frac{0{,}042}{11} < 0{,}004$. Daher ist $3{,}630 < \xi < 3{,}634$ und $\xi = 3{,}63$ mit der geforderten Genauigkeit. (Dieses Resultat haben wir in Nr. 154 nach der Regula falsi erst in drei Schritten erhalten.)

2. Als zweites Beispiel wollen wir die Gleichung

$$x \cdot \log x = 1$$

lösen.

Wir benutzen dieses Beispiel, um dem Leser zu erläutern, wie man die graphische Darstellung einer Funktion zur vorläufigen Orientierung über die Lage der Wurzeln der Gleichung benutzen kann.

Der Wert x, der die Gleichung

$$\log x = \frac{1}{x}$$

erfüllt, ist offenbar die Abszisse des Schnittpunktes der Kurven

$$y = \log x \quad \text{und} \quad y = \frac{1}{x}.$$

Schon die recht grobe Darstellung dieser Kurven (Abb. 85) läßt erkennen, daß die gesuchte Wurzel zwischen 2 und 3 liegt. Das folgt leicht, wenn wir $f(x) = x \cdot \log x - 1$ setzen; dann ist

$$f(2) = -0{,}39793\ldots < 0, \qquad f(3) = 0{,}43136\ldots > 0.$$

Wir berechnen diese Wurzel auf 0,0001 genau.

Offenbar ist für $2 \leqq x \leqq 3$

$$f'(x) = \log x + \log e > 0, \qquad f''(x) = \frac{\log e}{x} > 0$$

[Fall a)]; wir können $m = 0{,}7$ setzen.

Da $f(3)$ das gleiche Vorzeichen wie $f''(x)$ besitzt, ergibt sich nach Formel (8)

$$x_1' = 3 - \frac{f(3)}{f'(3)} = 3 - \frac{0{,}43136\ldots}{0{,}91141\ldots} = 3 - 0{,}473\ldots$$

Wir setzen $x_1' = 3 - 0{,}47 = 2{,}53$. Dann gilt $f(x_1') = f(2{,}53) = 0{,}019894$, so daß wir $x_1' - \xi \leqq \frac{0{,}0199}{0{,}7} < 0{,}03$ erhalten. Weiter ist

$$x_2' = 2{,}53 - \frac{f(2{,}53)}{f'(2{,}53)} = 2{,}53 - \frac{0{,}019894\ldots}{0{,}83741\ldots} = 2{,}53 - 0{,}02375\ldots$$

Wir wählen $x_2' = 2{,}53 - 0{,}0237 = 2{,}5063$. Gemäß Ungleichung (6) schätzen wir den Fehler ab:

$$f(2{,}5063) = 0{,}000096\ldots,$$

$$x_2' - \xi < \frac{0{,}000096\ldots}{0{,}7} < 0{,}0002,$$

also $2{,}5061 < \xi < 2{,}5063$. In diesem Fall haben wir bereits mit der geforderten Genauigkeit

$$\xi = 2{,}5062 \pm 0{,}0001.$$

In Wirklichkeit übertrifft der Näherungswert 2,5062 den Wert für ξ etwas, da $f(2{,}5062) > 0$ ist.

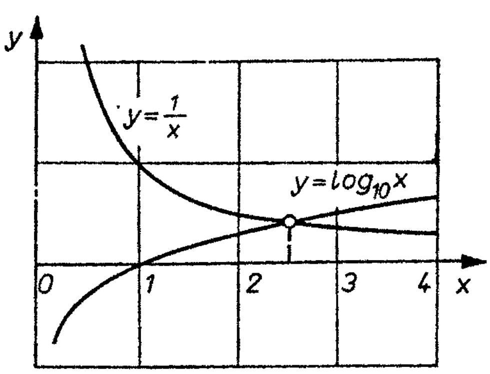

Abb. 85

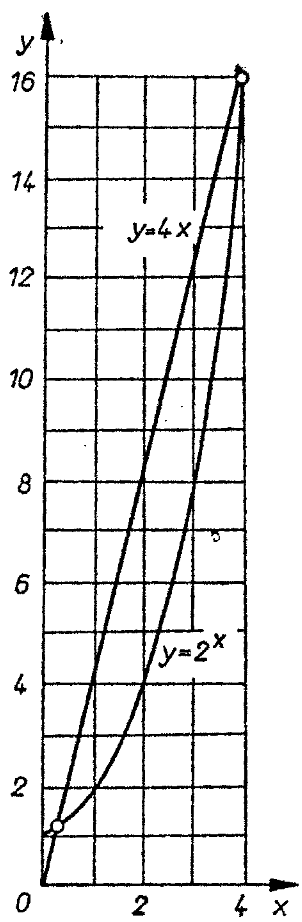

Abb. 86

3. Wir kehren zu der Gleichung

$$2^x = 4x$$

zurück, von der bereits in Nr. 81 die Rede war. Wir haben dort gesehen, daß eine Wurzel dieser Gleichung zwischen 0 und 0,5 liegt. Das läßt sich auch mit Hilfe der graphischen Darstellungen der Funktionen $y = 2^x$ und $y = 4x$ leicht ermitteln; aus Abb. 86 ist klar ersichtlich, daß sich diese Kurven außer im Punkt mit der Abszisse 4 noch in einem Punkt mit der Abszisse ξ zwischen 0 und 0,5 schneiden. Wir wollen diese Wurzel bis auf 0,00001 genau berechnen.

Für $0 \leqq x \leqq 0{,}5$ erhalten wir

$$f(x) = 2^x - 4x, \qquad f'(x) = 2^x \cdot \ln 2 - 4 < 0, \qquad f''(x) = 2^x \cdot \ln^2 2 > 0$$

[Fall b)]. Hier ist $m = 4 - \sqrt{2}\ln 2 > 3$, $M = \sqrt{2}\ln^2 2 < 0{,}7$, $\frac{M}{2m} < 0{,}12$. Da $f(0) = 1$ das gleiche Vorzeichen wie $f''(x)$ besitzt, beginnen wir mit $a = 0$; nach (6) ist der Fehler dieses Näherungswertes kleiner als $\frac{1}{3}$; dann läßt sich aber auf Grund von (11) der Fehler im voraus folgendermaßen abschätzen:

$$\xi - x_1' < 0{,}12 \cdot \frac{1}{9} < 0{,}014.$$

Daher runden wir den nach (8*) berechneten Wert

$$x_1' = \frac{1}{\ln 2 - 4} = \frac{1}{3{,}306852\ldots} = 0{,}30\ldots$$

auf zwei Dezimalstellen ab: $x_1' = 0{,}30$. Wir benutzen den Wert $f(0{,}30) = 0{,}031144$ und schätzen nach Ungleichung (6) den Fehler genauer ab:

$$\xi - x_1' < \frac{0{,}031144\ldots}{3} < 0{,}011;$$

dann ergibt sich nach (11)

$$\xi - x_2' < 0{,}12 \cdot 0{,}000121 < 0{,}000015,$$

so daß wir uns der geforderten Genauigkeit nähern.

Die folgende Näherung

$$x_2' = 0{,}30 - \frac{0{,}031144\ldots}{0{,}8533643\ldots - 4} = 0{,}30 + \frac{0{,}031144\ldots}{3{,}1466356\ldots} = 0{,}309897\ldots$$

runden wir „auf der Seite der Wurzel“ auf fünf Stellen auf: $x_2' = 0{,}30990$. Wegen $f(0{,}30990) = 0{,}000021\ldots > 0$ ist dieser Wert trotzdem kleiner als die Wurzel. Sein Fehler beträgt aber auf Grund von (6) in Wirklichkeit

$$\xi - x_2' < \frac{0{,}000022}{3} < 0{,}00001,$$

so daß wir schließlich

$$\xi = 0{,}30990(+0{,}00001)$$

erhalten.

4. Die Gleichung

$$\tan x = x$$

hat unendlich viele Wurzeln. Dies läßt sich sofort der Abb. 87 entnehmen, da unendlich viele Schnittpunkte der Tangenskurve $y = \tan x$ mit der Geraden $y = x$ vorkommen. Wir stellen uns die Aufgabe, die kleinste positive Wurzel dieser Gleichung zu berechnen; sie liegt zwischen $\frac{5\pi}{4}$ und $\frac{3\pi}{2}$.

Da der Tangens für $x = \frac{3\pi}{2}$ den Wert ∞ annimmt, schreiben wir die vorgegebene Gleichung in eine geeignete Form um:

$$f(x) = \sin x - x \cdot \cos x = 0.$$

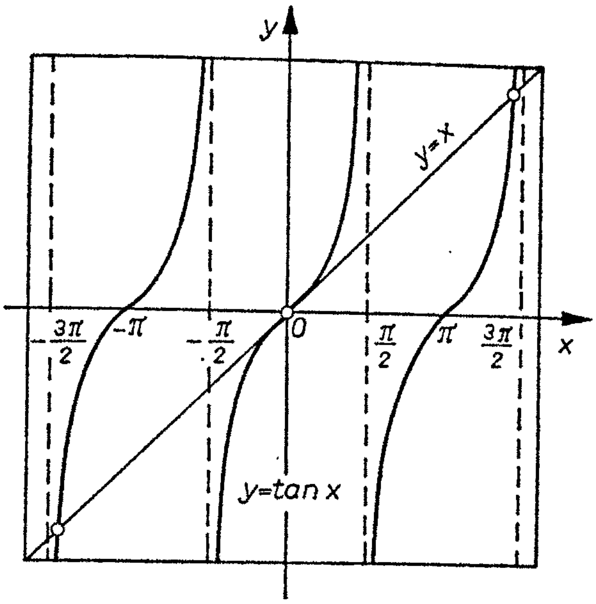

Abb. 87

Wir haben dann

$$f\left(\frac{5\pi}{4}\right) = -\frac{\sqrt{2}}{2}\left(1 - \frac{5\pi}{4}\right) > 0, \qquad f\left(\frac{3\pi}{2}\right) = -1 < 0;$$

$$f'(x) = x \cdot \sin x < 0, \qquad m > 2{,}7; \qquad f''(x) = \sin x + x \cdot \cos x < 0$$

[Fall d)].

Wir beginnen mit $b = \frac{3\pi}{2} = 4{,}7123889\ldots$ und erhalten

$$x_1' = \frac{3\pi}{2} - \frac{2}{3\pi} = 4{,}7123889\ldots - 0{,}2122066\ldots$$

Hier zeigt sich folgendes: In den (älteren) Tafeln der trigonometrischen Größen (und ihrer Logarithmen) sind die Winkel in Graden, Minuten und Sekunden angegeben. Daher müssen wir die Rundung des Korrekturgliedes 0,2122066 auf diese Einheit umrechnen. Wir nehmen dafür 12° 10′, was der etwas größeren Zahl 0,21223484 ... entspricht (Rundung auf der Seite der Wurzel), so daß wir

$$x_1' = 4{,}5000406\ldots \quad (257°\ 50')$$

erhalten. Ferner ist

$$f(x_1') = -\cos 12°\ 10' + 4{,}5000406\ldots \cdot \sin 12°\ 10' = -0{,}0291274\ldots,$$

$$f'(x_1') = -4{,}398962\ldots; \qquad x_1' - \xi < \frac{0{,}03}{2{,}7} < 0{,}012.$$

Im zweiten Schritt ergibt sich

$$x_2' = 4{,}5000406\ldots - \frac{0{,}0291274\ldots}{4{,}398962\ldots} = 4{,}5000406\ldots - 0{,}0066215\ldots;$$

wir runden das Korrekturglied auf: 0,0066177 ... (22′ 45″) und nehmen

$$x_2' = 4{,}4934229\ldots \quad (257°\ 27'\ 15'').$$

Da $f(x_2') = -0{,}000059\ldots$ ist, gilt

$$x_2' - \xi < \frac{0{,}00006}{2{,}7} < 0{,}0000223.$$

Somit ist $4{,}4934006\ldots < \xi < 4{,}4934229\ldots$, und wir können $\xi = 4{,}4934(+0{,}00003)$ setzen.

5. Die Leistungsfähigkeit der Newtonschen Methode zeigt sich besonders dann, wenn das Intervall, das die Wurzel enthält, hinreichend klein ist. Zum Abschluß berechnen wir die Wurzel der Gleichung $x^2 - 2x - 5 = 0$ mit großer Genauigkeit, und zwar bis auf $\frac{1}{10^{10}}$ genau, wobei wir vom Intervall $(2; 2{,}1)$ ausgehen, in dem diese Wurzel liegt.

Hier ist

$$f(x) = x^3 + 2x - 5, \quad f(2) = -1 < 0, \quad f(2{,}1) = 0{,}061 > 0,$$

$$f'(x) = 3x^2 - 2 > 0, \quad f''(x) = 6x > 0 \quad (\text{für } 2 \leqq x \leqq 2{,}1)$$

[Fall a)]. Man findet leicht $m = 10$, $M < 12{,}6$, so daß $\frac{M}{2m} < 0{,}63$ gilt.

Wir beginnen mit $b = 2{,}1$. Nach Formel (6) erhalten wir $b - \xi < \frac{0{,}061}{10} = 0{,}0061$. Unter Benutzung von Ungleichung (11) können wir nun im voraus ermitteln, welche Genauigkeit wir von x_1' erwarten dürfen: $x_1' - \xi < 0{,}63 \cdot 0{,}0061^2 < 0{,}000024$. Daher runden wir die Zahl

$$x_1' = 2{,}1 - \frac{f(2{,}1)}{f'(2{,}1)} = 2{,}1 - \frac{0{,}061}{11{,}23} = 2{,}1 - 0{,}00543\ldots$$

„auf der Seite der Wurzel" auf fünf Stellen: $x_1' = 2{,}1 - 0{,}00544 = 2{,}09456$. Da

$$f(x_1') = f(2{,}09456) = 0{,}000095078690816$$

ist, läßt sich jetzt nach Formel (6) der Fehler genauer abschätzen:

$$x_1' - \xi < \frac{0{,}000095\ldots}{10} < 0{,}00001.$$

Wir gehen zu x_2' über und ziehen (11) nochmals heran; wir berechnen im voraus

$$x_2' - \xi < 0{,}63 \cdot 0{,}00001^2 = 0{,}000000000063.$$

Daher unterscheidet sich die Zahl

$$x_2' = 2{,}09456 - \frac{0{,}000095078690816}{11{,}16154447808} = 2{,}09456 - 0{,}000008518416\ldots,$$

die auf elf Stellen gerundet ist und somit $x_2' = 2{,}09456 - 0{,}00000851841 = 2{,}09455148159$ beträgt, von der gesuchten Wurzel um weniger als $0{,}00000000007$. Also gilt

$$2{,}09455148152 < \xi < 2{,}09455148159,$$

d. h. $\xi = 2{,}0945514815 \left(+\frac{1}{10^{10}}\right)$.

157. Die kombinierte Methode. Diese Methode besteht in der gleichzeitigen Anwendung der Tangentenmethode und der Sehnenmethode.

Der Einfachheit halber betrachten wir Fall a). Die Näherungswerte x_1 und x_1' berechnen wir wie oben mit Hilfe der Formeln (2) und (8):

$$x_1 = a - \frac{(b-a) \cdot f(a)}{f(b) - f(a)}, \qquad x_1' = b - \frac{f(b)}{f'(b)}.$$

Dann ist nach dem Bewiesenen

$$a < x_1 < \xi < x_1' < b.$$

Für den folgenden Schritt jedoch ersetzen wir jetzt in diesen Formeln a und b durch x_1 und x_1':

$$x_2 = x_1 - \frac{(x_1' - x_1) \cdot f(x_1)}{f(x_1') - f(x_1)}, \qquad x_2' = x_1' - \frac{f(x_1')}{f'(x_1')}.$$

Dieses Verfahren läßt sich unbegrenzt fortsetzen. Haben wir Näherungswerte x_n und x_n' erhalten, zwischen denen die Wurzel ξ liegt, so gehen wir zu dem folgenden Paar von Näherungswerten über, und zwar nach den Formeln

$$x_{n+1} = x_n - \frac{(x_n' - x_n) \cdot f(x_n)}{f(x_n') - f(x_n)}, \qquad x_{n+1}' = x_n' - \frac{f(x_n)}{f'(x_n')}.$$

Die zweite ist identisch mit (10); die erste dagegen unterscheidet sich dadurch wesentlich von (5), daß *hier der Punkt b durch den Punkt x_n' ersetzt ist, der immer näher an ξ heranrückt.*

Schreiben wir die Ungleichung (4) für unseren Fall in der Form

$$\frac{x-a}{f(x)-g(a)} > \frac{b-a}{f(b)-f(a)}$$

und setzen darin $a = x_n$ und $x = x_n'$, so ist leicht zusehen, daß das erwähnte Ersetzen von b durch x_n' nur zu einer *schnelleren Approximation von x_n an die gesuchte Wurzel* führt (geometrisch ist dies offensichtlich!).

Somit approximieren wir bei der kombinierten Methode die Wurzel gleichzeitig von unten und von oben; die Näherungswerte streben von verschiedenen Seiten gegen diese. In den Fällen a) und d) strebt x_n von links gegen ξ und x_n' von rechts; in den Fällen b) und c) dagegen ist es offenbar umgekehrt. Die Größe $x_n' - x_n$ gibt unmittelbar eine Abschätzung für die Güte der erreichten Näherung, und darin liegt der Vorteil der kombinierten Methode. Ihre Anwendung erläutern wir durch Beispiele.

158. Beispiele und Übungen. Hier soll nur die kombinierte Methode benutzt werden.

1. Man bestimme die drei reellen Wurzeln der Gleichung

$$f(x) = 2x^3 - x^2 - 7x + 5 = 0,$$

und zwar bis auf 0,001 genau.

Durch eine einfache Zeichnung der Kurve von $y = f(x)$ finden wir die Intervalle, in denen diese Wurzeln liegen:

$$-2 < \xi_1 < -1, \qquad 0 < \xi_2 < 1, \qquad 1 < \xi_3 < 2.$$

Dies überprüft man leicht an Hand der Vorzeichenwechsel der Funktion.

(a) Im Intervall $[-2, -1]$ gilt

$$f'(x) = 6x^2 - 2x - 7 > 0, \qquad f''(x) = 12x - 2 < 0$$

[Fall c)]. Wegen $f(-2) = -1 < 0$, $f(-1) = 9 > 0$ muß man die Newtonsche Regel auf den linken Endpunkt des Intervalls anwenden. Es ist $f'(-2) = 21$ und

$$x_1' = -2 - \frac{-1}{21} = -1{,}952\ldots, \qquad x_1 = -1 - \frac{9}{9-(-1)} = -1{,}9.$$

Runden wir den Wert x_1' ab, so erhalten wir $-1{,}96 < \xi_1$. Runden wir ihn jedoch (auf der Seite der Wurzel) auf, so erhalten wir $-1{,}95$. Nun ist aber $f(-1{,}95) = 0{,}01775 > 0$, d. h., in diesem Fall haben wir die Wurzel überschritten. Das ist vorteilhaft, da man das Intervall verkleinern kann, in dem die Wurzel liegt. Wir lassen den obigen Wert x_1 unberücksichtigt und setzen $x_1' = -1{,}96$, $x_1 = -1{,}95$. Ferner ist

$$f(-1{,}96) = -0{,}180672, \qquad f'(-1{,}96) = 19{,}9696,$$

$$x_2' = -1{,}96 + \frac{0{,}180672}{19{,}9696} = -1{,}96 + 0{,}00904\ldots = -1{,}95095\ldots,$$

$$x_2 = -1{,}95 - \frac{0{,}01 \cdot 0{,}01775}{0{,}01775 + 0{,}180672} = -1{,}95 - 0{,}00089\ldots = -1{,}95089\ldots$$

Da ξ_1 zwischen diesen Grenzen liegen muß, gilt

$$\xi_1 = -1{,}9509 (\pm 0{,}0001)$$

(so daß die geforderte Genauigkeit sogar mehr als erreicht ist).

(b) Im Intervall [0, 1] bleibt die erste Ableitung $f'(x)$ negativ, während die zweite Ableitung im Punkt $x = \frac{1}{6}$, wo sie gleich 0 wird, ihr Vorzeichen ändert. Das zwingt uns, das Intervall weiter zu verkleinern.

Wenn wir es mit dem Wert $x = 0{,}5$ versuchen, erhalten wir $f(0{,}5) = 1{,}5 < 0$. Da $f(1) = -1 < 0$ ist, muß ξ_2 im Innern des Intervalls [0,5; 1] liegen, wo $f''(x)$ positiv bleibt [Fall b)]. Auch hier wenden wir die Newtonsche Regel auf den linken Endpunkt an. Es folgt

$$x_1' = 0{,}5 + \frac{1{,}5}{6{,}5} = 0{,}7307 \approx 0{,}74, \qquad x_1 = 1 - \frac{0{,}5}{2{,}5} = 0{,}80.$$

Rundet man x_1' auf der Seite der Wurzel, so wird die Wurzel nicht überschritten, da $f(0{,}74) = 0{,}082848 > 0$ ist.

Schließlich ist

$$x_2' = 0{,}74 + \frac{0{,}082848}{5{,}1944} = 0{,}755\ldots, \qquad x_2 = 0{,}80 - \frac{0{,}01296}{0{,}298848} = 0{,}756\ldots,$$

so daß $0{,}755\ldots < \xi_2 < 0{,}756\ldots$ gilt. Demnach ist

$$\xi_2 = 0{,}756(\pm 0{,}001).$$

(c) Im Intervall [1, 2] bleibt die zweite Ableitung positiv, während die erste Ableitung ihr Vorzeichen in

$$x = \frac{1 + \sqrt{43}}{6} \approx 1{,}26,$$

wo sie gleich 0 wird, ändert.

Wir versuchen es mit dem Wert 1,5: Es ist $f(1{,}5) = -1$, während $f(2) = 3$ ist, so daß $1{,}5 < \xi_3 < 2$ gilt. Die Ableitung $f'(x)$ bleibt in diesem Intervall positiv [Fall a)]. Es gilt

$$x_1 = 1{,}5 + \frac{1}{8} \approx 1{,}6, \qquad x_1' = 2 - \frac{3}{13} \approx 1{,}7.$$

Die Wurzel wird auch hier nicht übersprungen, da $f(1{,}7) = 0{,}036$ ist.

Somit ist

$$x_2 = 1{,}6 + \frac{0{,}0568}{0{,}604} = 1{,}6 + 0{,}094\ldots = 1{,}694\ldots,$$

$$x_2' = 1{,}7 - \frac{0{,}036}{6{,}94} = 1{,}7 - 0{,}005\ldots = 1{,}694\ldots,$$

also $\xi_3 = 1{,}694(+0{,}001)$.

Bemerkung. Da nach dem Vietaschen Wurzelsatz die Summe der Wurzeln gleich 0,5 sein muß, kann dies zur Nachprüfung benutzt werden.

2. Die Gleichung $f(x) = x^4 - 3x^2 + 75x - 10000 = 0$ besitzt zwei reelle Wurzeln: eine zwischen -11 und -10, die andere zwischen 9 und 10. Man berechne sie auf 0,00001 genau.

(a) Im Intervall $[-11, -10]$ gilt

$$f'(x) = 4x^3 - 6x + 75 < 0, \qquad f''(x) = 12x^2 - 6 > 0$$

[Fall b)]. Wir erhalten

$$x_1' = -11 + \frac{3453}{5183} = -10{,}33\ldots \approx -10{,}3,$$

$$x_1 = -10 - \frac{1050}{4503} = -10{,}23\ldots \approx -10{,}2.$$

Im ersten Fall rundeten wir auf der Seite der Wurzel, haben sie jedoch nicht übersprungen.

Ferner ist

$$x_2' = -10{,}3 + \frac{164{,}3181}{4234{,}108} = -10{,}262\ldots \approx -10{,}262,$$

$$x_2 = -10{,}2 - \frac{25{,}27984}{417{,}1165} = -10{,}260\ldots \approx -10{,}260$$

(hier trifft dasselbe zu). Und schließlich ist

$$x_3' = -10{,}262 + \frac{4{,}334569118736}{4186{,}137218912} = -10{,}262 + 0{,}0010354\ldots$$
$$= -10{,}2609645\ldots,$$

$$x_3 = -10{,}260 - \frac{0{,}008070038048}{8{,}369759358736} = -10{,}260 - 0{,}0009642\ldots$$
$$= -10{,}2609642\ldots,$$

also $\xi_1 = -10{,}260964(-0{,}000001)$ (sogar mit größerer Genauigkeit als gefordert).

(b) Im Intervall $[9, 10]$ gilt $f'(x) > 0$ und $f''(x) > 0$ [Fall a)]. Hier ist

$$x_1 = 9 + \frac{3007}{3457} = 9 + 0{,}869\ldots \approx 9{,}87$$

(auf der Seite der Wurzel),

$$x_1' = 10 - \frac{450}{4015} = 10 - 0{,}112\ldots \approx 9{,}89;$$

$$x_2 = 9{,}87 + \frac{1{,}2389658878}{77{,}4689008} = 9{,}87 + 0{,}01599\ldots = 9{,}88599\ldots,$$

$$x_2' = 9{,}89 - \frac{15{,}52060641}{3885{,}106676} = 9{,}89 - 0{,}003993\ldots = 9{,}886006\ldots,$$

also $\xi_2 = 9{,}88600(\pm 0{,}00001)$.

3. Wir betrachten die Gleichung

$$f(x) = x \sin x - 0{,}5 = 0.$$

Stellen wir die Funktionen $y = \sin x$ und $y = \dfrac{0{,}5}{x}$ graphisch dar (Abb. 88), so sehen wir, daß sie sich in endlich vielen Punkten schneiden, so daß unsere Gleichung unendlich viele Wurzeln besitzt. Aus der Zeichnung ist aber auch ersichtlich, daß die kleinste positive Wurzel ξ in der Nähe von 0,7 liegt. Wir wollen sie auf 0,000001 genau berechnen.

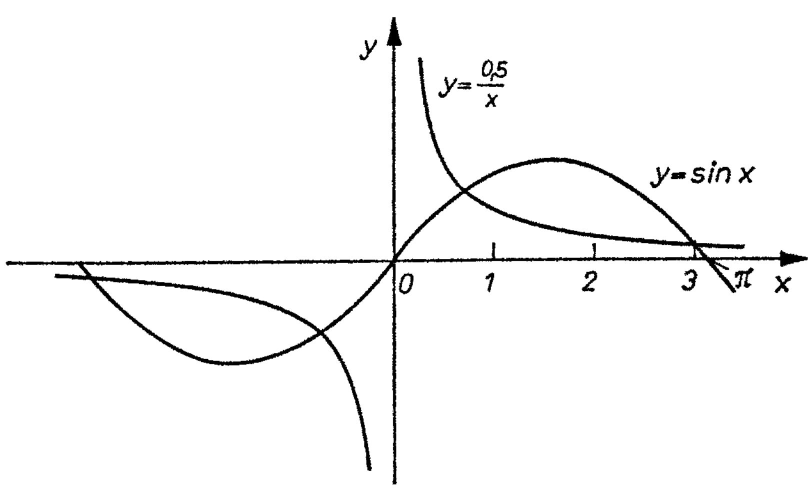

Abb. 88

(Hier ist die Bemerkung zu Aufgabe 4 in Nr. 156 über das Runden in Teilen eines Grades zu berücksichtigen.)

Setzen wir die Werte $a = 0{,}6981317\ldots(40^\circ)$ und $b = 0{,}7853982\ldots(45^\circ)$ in die Funktionsgleichung ein, so erhalten wir im ersten Fall einen negativen Wert, im zweiten einen positiven. Also gilt $a < \xi < b$. Beide Ableitungen $f'(x)$ und $f''(x)$ sind in diesem Intervall positiv [Fall a)].

Das Schema der Berechnung lautet

$$x_1 = 0{,}6981317\ldots + 0{,}0419512\ldots,$$

$$x_1' = 0{,}7853982\ldots - 0{,}0438510\ldots;$$

die erste Korrektur „runden“ wir auf $0{,}0418879\ldots$ $(2^\circ\,24')$ und die zweite auf $0{,}0439231\ldots$ $(2^\circ\,31')$, so daß wir schließlich

$$x_1 = 0{,}7400196\ldots\ (42^\circ\,24'),\quad x_1' = 0{,}7414741\ldots\ (42^\circ\,29')$$

erhalten. Ferner ist

$$x_2 = 0{,}7400196\ldots + 0{,}0008211\ldots = 0{,}7408407\ldots,$$

$$x_2' = 0{,}7414741\ldots - 0{,}0006329\ldots = 0{,}7408412\ldots,$$

und hieraus erhalten wir mit der geforderten Genauigkeit $\xi = 0{,}740841(\pm 0{,}0000005)$.

4. Zum Schluß untersuchen wir die Gleichung

$$f(x) = x^4 - x - 1 = 0.$$

Wir haben in Nr. 81 gesehen, daß sie eine Wurzel zwischen $a = 1{,}22$ und $b = 1{,}23$ hat. Man untersuche, welche Genauigkeit bei der Bestimmung dieser Wurzel in zwei Schritten der kombinierten Methode erreicht wird. Das Rechenschema lautet:

$$x_1 = 1{,}22 + \frac{0{,}0000466544}{0{,}06353115} = 1{,}22073\ldots \approx 1{,}2207,$$

$$x_1' = 1{,}23 - \frac{0{,}05886641}{6{,}443468} = 1{,}22086\ldots \approx 1{,}2209;$$

$$x_2 = 1{,}2207 + \frac{0{,}00000005533760598398}{0{,}001255538012096} = 1{,}22074407\ldots,$$

$$x_2' = 1{,}2209 - \frac{0{,}0009788499821761}{6{,}279478581316} = 1{,}2207441\ldots,$$

und somit ist $\xi = 1{,}2207441(\pm 0{,}0000001)$.

V. Funktionen mehrerer Veränderlicher

§ 1. Grundbegriffe

159. Funktionale Abhängigkeit zwischen Veränderlichen. Beispiele. Bisher haben wir die gemeinsame Veränderung zweier Veränderlicher betrachtet, von denen die eine von der anderen abhing: Durch den Wert der unabhängigen Veränderlichen war der Wert der abhängigen Veränderlichen oder der Funktion bestimmt. In der Wissenschaft und im Leben kommen jedoch viel öfter Fälle vor, in denen mehrere unabhängige Veränderliche auftreten und es zur Bestimmung des Wertes der Funktion unerläßlich ist, vorher die Werte anzugeben, die alle diese unabhängigen Veränderlichen annehmen.

1. So ist z. B. das Volumen V eines Kreiszylinders eine Funktion des Radius R seiner Grundfläche und seiner Höhe H; die Abhängigkeit zwischen diesen Veränderlichen wird durch die Formel

$$V = \pi R^2 H$$

ausgedrückt, aus der wir, wenn wir die Werte der unabhängigen Veränderlichen R und H kennen, den entsprechenden Wert von V bestimmen können.

Das Volumen V eines Kegelstumpfs ist offenbar eine Funktion von drei unabhängigen Veränderlichen, und zwar eine Funktion der Radien R und r der Grund- bzw. Deckfläche und der Höhe H; diese Abhängigkeit wird durch die Formel

$$V = \frac{\pi H}{3}(R^2 + Rr + r^2)$$

ausgedrückt.

2. Nach dem Ohmschen Gesetz hängt die Spannung U in einem elektrischen Stromkreis mit dem Widerstand R und der Stromstärke I durch $U = R \cdot I$ zusammen. Wenn U und R gegeben sind, so kann man I als Funktion von U und R bestimmen:

$$I = \frac{U}{R}.$$

3. Die Temperatur einer Gasmenge, die sich in einem Kolbenzylinder befindet, sei nicht konstant; dann hängen das Volumen V und der Druck p eines Gasmols mit seiner (absoluten) Temperatur T nach der Clapeyronschen Formel

$$pV = RT \quad (R = \text{const})$$

zusammen. Nimmt man V und T als unabhängige Veränderliche, so kann man hieraus p durch

$$p = \frac{RT}{V}$$

ausdrücken.

4. Untersucht man den physikalischen Zustand eines Körpers, so findet man oft, daß sich seine Eigenschaften von Punkt zu Punkt ändern, beispielsweise die Dichte, die Temperatur, das elektrische Potential usw. Alle diese Größen sind „Punktfunktionen" oder, wenn man will, Funktionen der Koordinaten x, y, z des Punktes. Ändert sich der physikalische Zustand des Körpers mit der Zeit, so kommt zu diesen unabhängigen Veränderlichen auch noch die Zeit t hinzu. In diesem Fall haben wir es mit einer Funktion von vier unabhängigen Veränderlichen zu tun.

Die Zahl dieser Beispiele kann der Leser selbst beliebig vergrößern.

Um den Begriff der Funktion im Fall mehrerer unabhängiger Veränderlicher zu präzisieren, beginnen wir mit dem einfachsten Fall, dem zweier Veränderlicher.

160. Funktionen zweier Veränderlicher und ihr Definitionsbereich. Ist von der Änderung *zweier* unabhängiger Veränderlicher x und y die Rede, so muß stets angegeben werden, welche Paare (x, y) von Werten sie annehmen können; die Menge $\mathcal{M}$ dieser Paare wird der *Wertebereich* der Veränderlichen x, y genannt.

Der Begriff der Funktion wird in denselben Worten definiert wie im Fall einer einzigen unabhängigen Veränderlichen:

Die Veränderliche z (mit dem Wertebereich oder Wertevorrat $\mathcal{Z}$) wird *Funktion der unabhängigen Veränderlichen x und y in* (auch *auf*) *der Menge $\mathcal{Z}$* genannt, wenn jedem Wertepaar (x, y) aus $\mathcal{M}$ nach einer bestimmten Vorschrift ein bestimmter Wert z (aus $\mathcal{Z}$) zugeordnet ist.

Hier handelt es sich um eine *eindeutige* Funktion; die Definition läßt sich jedoch leicht auf den Fall mehrdeutiger Funktionen verallgemeinern.

Die Menge $\mathcal{M}$ ist der *Definitionsbereich* der Funktion. Die Veränderlichen x und y werden die *Argumente* der Funktion z genannt. Die funktionale Abhängigkeit zwischen z einerseits und x und y andererseits wird wie im Fall einer einzigen unabhängigen Veränderlichen bezeichnet:

$$z = f(x, y), \quad z = \varphi(x, y), \quad z = z(x, y) \quad \text{usw.}$$

Gehört das Paar (x_0, y_0) zu $\mathcal{M}$, so bezeichnet $f(x_0, y_0)$ den speziellen (Zahlen-) Wert, den die Funktion $f(x, y)$ für $x = x_0$ und $y = y_0$ annimmt.

Wir führen einige Beispiele für Funktionen an, die analytisch, d. h. durch Formeln, gegeben sind, wobei wir auch ihren Definitionsbereich angeben. Die Formeln

$$z = xy \tag{a}$$

und

$$z = x^2 + y^2 \tag{b}$$

definieren die Funktionen für alle Paare (x, y) ohne Ausnahme. Die Formeln

$$z = \sqrt{1 - x^2 - y^2} \tag{c}$$

und

$$z = \frac{1}{\sqrt{1 - x^2 - y^2}} \tag{d}$$

dagegen gelten (wenn wir uns auf endliche reelle z-Werte beschränken) nur für diejenigen Paare (x, y), die der Ungleichung

$$x^2 + y^2 \leqq 1 \quad \text{bzw.} \quad x^2 + y^2 < 1$$

genügen.

Durch die Formel

$$z = \arcsin \frac{x}{a} + \arcsin \frac{y}{b} \tag{e}$$

wird die Funktion z für alle diejenigen Werte von x und y definiert, die einzeln den Ungleichungen

$$-a \leqq x \leqq a, \qquad -b \leqq y \leqq b$$

genügen.

In allen diesen Fällen gaben wir den umfassendsten (natürlichen, Nr. 46, Punkt 2) Gültigkeitsbereich der Formeln an.

Wir betrachten jetzt ein Beispiel.

Die Seiten eines Dreiecks mögen sich beliebig ändern, jedoch soll der Umfang konstant gleich $2p$ sein. Bezeichnet man die Länge von zwei der Seiten mit x bzw. y, so hat die dritte Seite die Länge $2p - x - y$; das Dreieck ist durch die Seiten x und y vollständig bestimmt. Wie hängt nun der Flächeninhalt z des Dreiecks von ihnen ab?

Nach der Heronschen Formel gilt für diesen Flächeninhalt

$$z = \sqrt{p(p-x)(p-y)(x+y-p)}. \tag{f}$$

Was den Definitionsbereich $\mathcal{M}$ dieser Funktion betrifft, so wird er dieses Mal durch das konkrete Problem, das auf die Funktion führte, eingeengt. Da die Länge jeder Dreieckseite positiv und kleiner als der halbe Umfang ist, müssen die Ungleichungen

$$0 < x < p, \qquad 0 < y < p, \qquad x + y > p$$

gelten; damit ist das Gebiet $\mathcal{M}$ bestimmt.[1])

Während also für Funktionen einer Veränderlichen der übliche Wertebereich des Arguments ein (endliches oder unendliches) Intervall ist, kommen bei Funktionen zweier Veränderlicher sehr verschiedenartige und komplizierte (natürliche) Wertebereiche der Argumente vor.

Die Untersuchung solcher Bereiche wird durch ihre geometrische Interpretation beträchtlich erleichtert. Wählen wir in der Ebene zwei zueinander senkrechte Achsen und tragen auf ihnen wie üblich die Werte x und y ab, so ist bekanntlich jedem Paar (x, y) eindeutig ein Punkt in der Ebene zugeordnet, der diese Werte als Koordinaten hat, und umgekehrt.

Daher ist es zur Charakterisierung der Menge der Paare (x, y), für die die Funktion definiert ist, am einfachsten, anzugeben, welche Figur in der x, y-Ebene von den entsprechenden Punkten ausgefüllt ist.

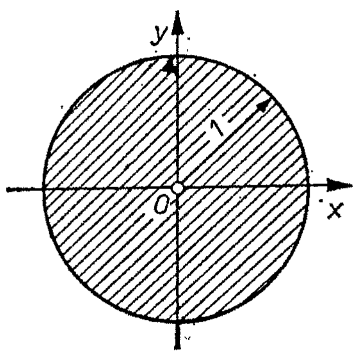

Abb. 89

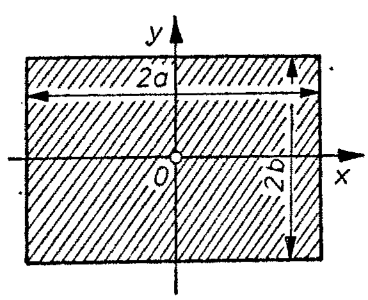

Abb. 90

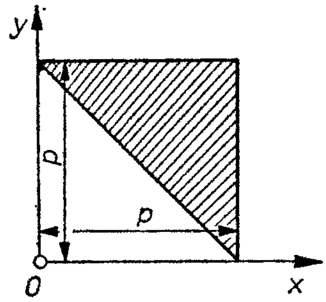

Abb. 91

Man sagt, die Funktionen (a) und (b) seien *in der ganzen Ebene*, die Funktionen (c) und (d) in einem abgeschlossenen Kreis (da der Rand dazugehört) bzw. in einem offenen Kreis (ohne Rand) (Abb. 89) definiert; die Funktion (e) ist in einem Rechteck (Abb. 90), die Funktion (f) schließlich in einem offenen Dreieck (Abb. 91) definiert.

[1]) Unabhängig davon, daß der Ausdruck für die Funktion auch in einem größeren Gebiet sinnvoll ist, z. B. in dem Bereich, für den gleichzeitig $x > p$, $y > p$ ist.

Diese geometrische Interpretation ist so bequem, daß man meist die Zahlenpaare (x, y) selbst als „Punkte" und die Menge dieser Punkte, die eine bestimmte geometrische Figur bildet, mit dem Namen dieser Figur bezeichnet. So ist die Menge der „Punkte" oder Paare (x, y), die den Ungleichungen

$$a \leqq x \leqq b, \quad c \leqq y \leqq d$$

genügen, ein „Rechteck" (Rechteckfläche mit Rand), dessen Ausmaße gleich $b - a$ bzw. $d - c$ sind. Wir bezeichnen es mit $[a, b; c, d]$, ähnlich wie ein Intervall. Die Menge der „Punkte" oder Paare (x, y), die der Ungleichung

$$(x - \alpha)^2 + (y - \beta)^2 \leqq r^2$$

genügen, ist ein Kreis (Kreisfläche einschließlich der Peripherie) mit dem Radius r und dem Mittelpunkt (α, β), usw.

Ähnlich wie eine Funktion $y = f(x)$ durch eine Kurve (Nr. 47) geometrisch dargestellt wird, läßt sich auch die Gleichung $z = f(x, y)$ geometrisch deuten. Wir nehmen im Raum ein rechtwinkliges x, y, z-Koordinatensystem an; in der x, y-Ebene stellen wir den Wertebereich $\mathcal{M}$ der Veränderlichen x und y dar, und schließlich errichten wir in jedem Punkt $M(x, y)$ dieses Bereichs eine Senkrechte auf der x, y-Ebene und tragen hierauf den Wert $z = f(x, y)$ ab. Der geometrische Ort dieser so erhaltenen Punkte ist die *räumliche Darstellung* der Funktion. Das ist im allgemeinen eine Fläche (im üblichen anschaulichen Sinne); die Beziehung $z = f(x, y)$ wird dann die *Gleichung der Fläche* genannt.

Als Beispiele sind in Abb. 92, 93 und 94 die Funktionen

$$z = xy, \qquad z = x^2 + y^2, \qquad z = \sqrt{1 - x^2 - y^2}$$

bildlich dargestellt. Die erste ist ein *hyperbolisches Paraboloid,* die zweite ein *Rotationsparaboloid* und die dritte eine *Halbkugel.*

Zum Schluß weisen wir noch darauf hin, daß man auch eine Veränderliche $x_{m,n}$ betrachten kann, deren Werte mit zwei natürlichen Indizes m und n numeriert sind (jeder von ihnen durchläuft unabhängig vom anderen die Folge der natürlichen Zah-

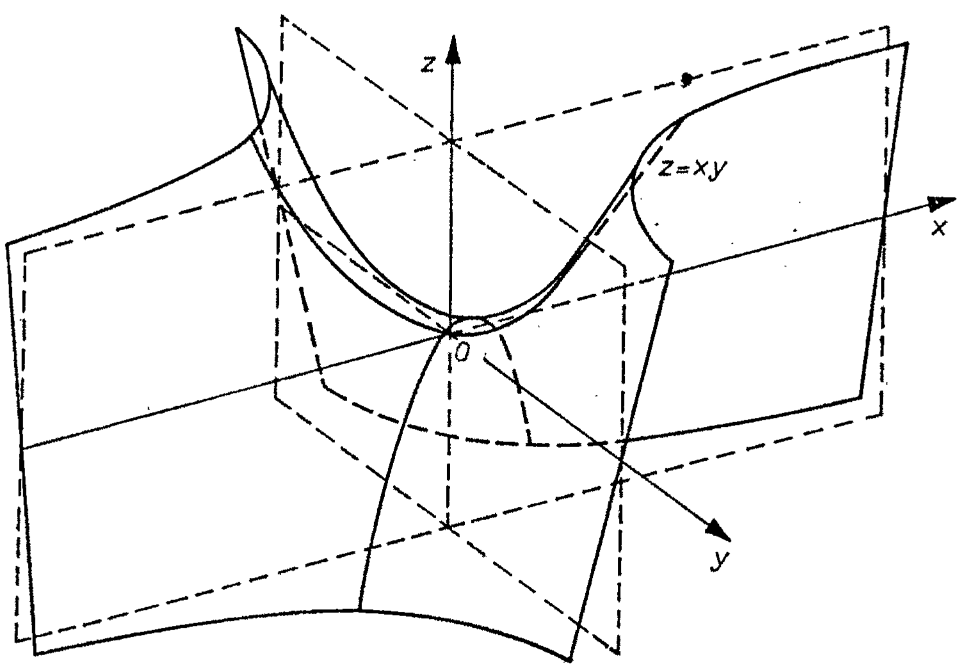

Abb. 92

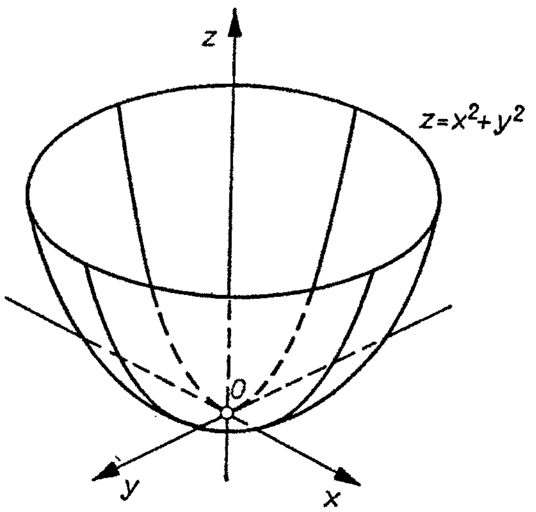

Abb. 93

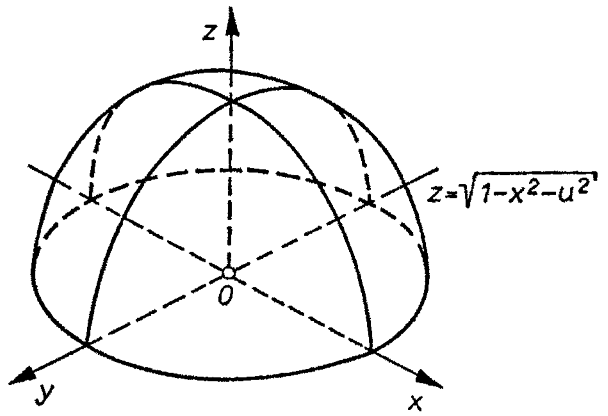

Abb. 94

len). Solche Veränderliche sind in gewissem Sinn eine Verallgemeinerung einer diskreten Veränderlichen.

Als Beispiele mögen

$$x_{m,n} = \frac{(m+n)!}{m!\,n!}, \quad x_{m,n} = \frac{1}{m^2+n^2}, \quad x_{m,n} = \frac{(m+1)\cdot n}{m\cdot(n+1)}$$

dienen.

Im Grunde müßte man hier die Indizes m und n als unabhängige Veränderliche ansehen und die Veränderliche $x_{m,n}$ als Funktion von ihnen. In diesem Fall wird der Wertebereich der unabhängigen Veränderlichen m, n, also der Definitionsbereich der „Funktion" $x_{m,n}$, geometrisch durch die „Gitterpunkte" im ersten Quadranten dargestellt.

161. Der arithmetische n-dimensionale Raum. Wir gehen nun zu Funktionen von n unabhängigen Veränderlichen ($n \geqq 3$) über und wollen uns zunächst mit den Systemen der Werte dieser Veränderlichen befassen.

Im Fall $n = 3$ kann ein solches System (x, y, z) dreier Zahlen, wie ohne weiteres klar ist, noch anschaulich-geometrisch gedeutet werden, und zwar als Punkt des Raumes, und die Menge dieser Tripel als Teil des Raumes oder geometrischer Körper. Für $n > 3$ besteht diese Möglichkeit der unmittelbaren anschaulich-geometrischen Interpretation nicht mehr, weil wir uns einen Raum von mehr als drei Dimensionen nicht mehr vorstellen können.

Wenn man also die geometrischen Vorstellungen (die sich für Funktionen zweier und dreier Veränderlicher als fruchtbar erwiesen haben) auch auf die Theorie der Funktionen von mehr als drei Veränderlichen ausdehnen will, muß man den Begriff des n-dimensionalen „Raumes" für $n > 3$ in die Analysis einführen.

Unter einem *(n-dimensionalen) „Punkt" M* versteht man ein System von n reellen Zahlen $x_1, x_2, \ldots, x_n$; die Zahlen $x_1, x_2, \ldots, x_n$ selbst sind die Koordinaten dieses „Punktes".[1]) Die Menge aller möglichen n-dimensionalen „Punkte" bildet den *n-di-*

[1]) Da wir es hier nicht mit einer festen Zahl von Veränderlichen zu tun haben, ist es zweckmäßig, sie nicht mit verschiedenen Buchstaben, sondern mit einem einzigen Buchstaben mit verschiedenen Indizes zu bezeichnen. Somit bedeutet (abweichend von der bisherigen Praxis) x_i nicht den i-ten Wert irgendeiner Veränderlichen, sondern die i-te Veränderliche selbst, die die Werte ihres Wertebereichs durchläuft.

mensionalen „Raum" (der manchmal auch *arithmetischer Raum* oder *Zahlenraum* genannt wird).

Es ist zweckmäßig, den Begriff des *„Abstandes"* $\overline{MM'}$ zwischen zwei (n-dimensionalen) „Punkten"

$$M(x_1, x_2, \ldots, x_n) \quad \text{und} \quad M'(x_1', x_2', \ldots, x_n')$$

einzuführen, und zwar setzt man in Analogie zum zwei- und dreidimensionalen Fall, der in der analytischen Geometrie behandelt wird,

$$\begin{aligned} \overline{MM'} = \overline{M'M} &= \sqrt{\sum_{i=1}^{n} (x_i' - x_i)^2} \\ &= \sqrt{(x_1' - x_1)^2 + (x_2' - x_2)^2 + \cdots + (x_n' - x_n)^2}; \end{aligned} \tag{1}$$

für $n = 2$ und $n = 3$ stimmt dieser „Abstand" mit dem gewöhnlichen Abstand zwischen den entsprechenden geometrischen Punkten (der Ebene bzw. des Raumes) überein.

Wenn wir noch einen weiteren „Punkt"

$$M''(x_1'', x_2'', \ldots, x_n'')$$

hinzunehmen, so gilt für die „Abstände" $\overline{MM'}$, $\overline{M'M''}$ und $\overline{MM''}$ die Ungleichung

$$\overline{MM''} \leqq \overline{MM'} + \overline{M'M''}, \tag{2}$$

die an den folgenden bekannten Satz aus der Geometrie erinnert: *Die Länge einer Seite eines Dreiecks ist nicht größer als die Summe der Längen der beiden anderen Seiten.* (Daher nennt man diese Ungleichung auch *Dreiecksungleichung.*)

Beweis. Für jedes System reller Zahlen $a_1, a_2, \ldots, a_n$ und $b_1, b_2, \ldots, b_n$ gilt die Ungleichung[1])

$$\sqrt{\sum_{i=1}^{n} (a_i + b_i)^2} \leqq \sqrt{\sum_{i=1}^{n} a_i^2} + \sqrt{\sum_{i=1}^{n} b_i^2}.$$

Setzt man hier

$$a_i = x_i' - x_i, \quad b_i = x_i'' - x_i', \quad \text{also} \quad a_i + b_i = x_i'' - x_i \ (i = 1, 2, \ldots, n),$$

[1]) Diese Ungleichung ist der Spezialfall der Minkowskischen Ungleichung (7) aus Nr. 133 für $k = 2$. Erhebt man beide Seiten ins Quadrat und streicht auf beiden Seiten die einander gleichen Glieder, so geht sie in die Cauchysche Ungleichung [Nr. 133, Formel (5a)] über. Sie läßt sich aber auch direkt völlig elementar beweisen, und damit ist auch die obige Ungleichung bewiesen.

Das quadratische Trinom $\sum_{i=1}^{n} a_i^2 x^2 + 2 \sum_{i=1}^{n} a_i b_i x + \sum_{i=1}^{n} b_i^2 = \sum_{i=1}^{n} (a_i x + b_i)^2$ nimmt offenbar keine negativen Werte an; es kann also nicht zwei verschiedene reelle Nullstellen x haben. Daher ist der Ausdruck

$$\sum_{i=1}^{n} a_i^2 \cdot \sum_{i=1}^{n} b_i^2 - \left\{\sum_{i=1}^{n} a_i b_i\right\}^2$$

nicht negativ, und diese Aussage ist der Cauchyschen Ungleichung äquivalent.

so erhält man

$$\sqrt{\sum_{i=1}^{n}(x_i''-x_i)^2} \leqq \sqrt{\sum_{i=1}^{n}(x_i'-x_i)^2} + \sqrt{\sum_{i=1}^{n}(x_i''-x_i')^2},$$

was (2) äquivalent ist. Somit gilt diese wesentliche Eigenschaft des Abstandes auch in unserem „Raum".

Im n-dimensionalen „Raum" können wir auch stetige „Kurven" betrachten. Bekanntlich (Nr. 106) liefern die Gleichungen $x = \varphi(t)$, $y = \psi(t)$, wobei $\varphi(t)$ und $\psi(t)$ in einem Intervall $[t', t'']$ stetige Funktionen eines *Parameters* t sind, in der x, y-Ebene eine stetige Kurve. Analog können wir mit drei stetigen Funktionen

$$x = \varphi(t), \quad y = \psi(t), \quad z = \chi(t) \quad (t' \leqq t \leqq t'')$$

eine stetige Kurve im (gewöhnlichen dreidimensionalen) Raum darstellen. Wir verallgemeinern das und betrachten jetzt n stetige Funktionen von t,

$$x_1 = \varphi_1(t), \quad x_2 = \varphi_2(t), \quad \ldots, \quad x_n = \varphi_n(t) \quad (t' \leqq t \leqq t'').$$

Dann stellt die Menge der „Punkte"

$$\big(\varphi_1(t), \varphi_2(t), \ldots, \varphi_n(t)\big),$$

die wir für die verschiedenen Werte des Parameters t erhalten, eine *stetige „Kurve"* im n-dimensionalen „Raum" dar. Setzt man

$$x_1' = \varphi_1(t'), \ldots, x_n' = \varphi_n(t'); \quad x_1'' = \varphi_1(t''), \ldots, x_n'' = \varphi_n(t''),$$

so kann man sagen, daß diese „Kurve" den „Punkt" $M'(x_1', \ldots, x_n')$ mit dem „Punkt" $M''(x_1'', \ldots, x_n'')$ verbindet.

Wenn alle Funktionen $\varphi_1, \ldots, \varphi_n$ linear sind, geht die „Kurve" in eine *„Gerade"* über:

$$x_1 = \alpha_1 t + \beta_1, \ldots, x_n = \alpha_n t + \beta_n;$$

hier werden die Koeffizienten $\alpha_1, \ldots, \alpha_n$ als nicht gleichzeitig verschwindend ($|\alpha_1| + |\alpha_2| + \cdots + |\alpha_n| > 0$) und der Parameter t von $-\infty$ bis ∞ variierend angesehen. Die „Punkte" fassen wir als nach wachsenden Parameterwerten geordnet auf; für $t' < t < t''$ liegt von den entsprechenden „Punkten" M', M, M'' der „Punkt" M zwischen den beiden anderen, da er hinter M' und vor M'' kommt. Unter diesen Bedingungen kann man leicht zeigen, daß für die Abstände zwischen den Punkten die Beziehung $\overline{M'M''} = \overline{M'M} + \overline{MM''}$ gilt, was für eine Gerade im gewöhnlichen Raum charakteristisch ist.

Die Gleichung der *„Geraden"*, die durch zwei gegebene „Punkte" $M'(x_1', \ldots, x_n')$ und $M''(x_1'', \ldots, x_n'')$ geht, kann offenbar in der Gestalt

$$x_1 = x_1' + t(x_1'' - x_1'), \ldots, x_n = x_n' + t(x_n'' - x_n') \quad (-\infty < t < \infty)$$

geschrieben werden, wobei die „Punkte" M' und M'' den Werten $t = 0$ bzw. $t = 1$ entsprechen. Variiert t nur zwischen 0 und 1, so erhält man die „geradlinige Strecke", die diese „Punkte" verbindet.

Eine „Kurve", die aus endlich vielen „geradlinigen Strecken" besteht, wird *Polygonzug* oder *Streckenzug* genannt.

162. Beispiele für Bereiche im n-dimensionalen Raum. Wir betrachten jetzt einige Beispiele von „*Körpern*" oder „*Bereichen*" im n-dimensionalen „Raum".

1. Die Menge der Punkte $M(x_1, \ldots, x_n)$, deren Koordinaten unabhängig voneinander den Ungleichungen

$$a_1 \leqq x_1 \leqq b_1, a_2 \leqq x_2 \leqq b_2, \ldots, a_n \leqq x_n \leqq b_n$$

genügen, heißt (*n-dimensionales*) „*rechtwinkliges Parallelepiped*" und wird mit

$$[a_1, b_1; a_2, b_2; \ldots; a_n, b_n]$$

bezeichnet. Für $n = 2$ erhält man hieraus ein „*Rechteck*", wovon schon in Nr. 160, Beispiel (e), die Rede war; dem dreidimensionalen „Parallelepiped" entspricht im Raum das gewöhnliche rechtwinklige Parallelepiped (auch *Quader* oder *Rechtflach* genannt).

Wenn in diesen Beziehungen die Gleichheit ausgeschlossen wird, also

$$a_1 < x_1 < b_1, a_2 < x_2 < b_2, \ldots, a_n < x_n < b_n,$$

so ist dadurch das *offene* „rechtwinklige Parallelepiped"

$$(a_1, b_1; a_2, b_2; \ldots; a_n, b_n)$$

definiert, im Unterschied zu dem oben betrachteten abgeschlossenen[1]) „Parallelepiped". Die Differenzen $b_1 - a_1, b_2 - a_2, \ldots, b_n - a_n$ nennt man in beiden Fällen die „*Abmessungen*" („*Kanten*") des „Parallelepipeds", der „Punkt"

$$\left(\frac{a_1 + b_1}{2}, \frac{a_2 + b_2}{2}, \ldots, \frac{a_n + b_n}{2}\right)$$

wird sein *Mittelpunkt* genannt.

Unter einer *Umgebung des „Punktes"* $M_0(x_1^0, x_2^0, \ldots, x_n^0)$ versteht man ein beliebiges offenes „Parallelepiped"

$$(x_1^0 - \delta_1, x_1^0 + \delta_1; x_2^0 - \delta_2, x_2^0 + \delta_2; \ldots; x_n^0 - \delta_n, x_n^0 + \delta_n) \tag{3}$$

$(\delta_1, \delta_2, \ldots, \delta_n > 0)$ mit dem „Mittelpunkt" M_0; meist nimmt man den „Würfel"

$$(x_1^0 - \delta, x_1^0 + \delta; x_2^0 - \delta, x_2^0 + \delta; \ldots; x_n^0 - \delta, x_n^0 + \delta)$$

$(\delta > 0)$, dessen sämtliche Kanten gleich 2δ sind.

2. Wir betrachten die Menge der „Punkte" $M(x_1, x_2, \ldots, x_n)$, deren Koordinaten den Ungleichungen

$$x_1 \geqq 0, x_2 \geqq 0, \ldots, x_n \geqq 0, \quad x_1 + x_2 + \cdots + x_n \leqq h \quad (h > 0)$$

genügen. Für $n = 2$ ist das dieser Menge entsprechende geometrische Bild ein *gleichschenkliges rechtwinkliges Dreieck*, für $n = 3$ ein *Tetraeder* (Abb. 95). Im allgemeinen Fall nennt man es (abgeschlossenes) *Simplex*[2]) (im Unterschied zu einem offenen, das man erhält, wenn in diesen Beziehungen die Gleichheit ausgeschlossen wird).

---

[1]) Man kann auch ein *unendliches* Parallelepiped betrachten, bei dem eines (oder auch mehrere) der definierenden Intervalle unendlich ist (sind). Wenn von einem n-dimensionalen Parallelepiped die Rede ist, ohne daß etwas dazu gesagt wird, meinen wir stets ein endliches.

[2]) Von dem lateinischen Wort simplex = einfach. In der Tat ist ein Simplex der einfachste mehrflächige Körper mit der in dem betreffenden Raum kleinstmöglichen Anzahl Seitenflächen.

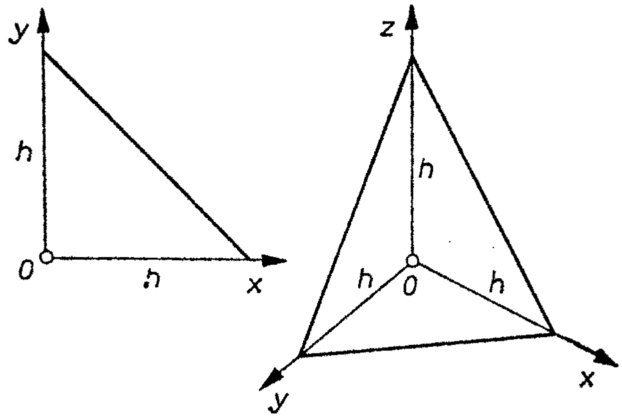

Abb. 95

3. Schließlich bildet die Menge der „Punkte“ $M(x_1, x_2, \ldots, x_n)$, die durch die Ungleichung

$$(x_1 - x_1^0)^2 + (x_2 - x_2^0)^2 + \cdots + (x_n - x_n^0)^2 \leqq r^2 \quad (\text{bzw. } < r^2)$$

definiert ist, wenn $M_0(x_1^0, x_2^0, \ldots, x_n^0)$ ein konstanter „Punkt“ und r eine konstante positive Zahl ist, eine abgeschlossene (bzw. offene) *n-dimensionale „Kugel“* mit dem Radius r und dem „Mittelpunkt“ M_0. Mit anderen Worten, eine „Kugel“ ist die Menge der „Punkte“ M, deren „Abstand“ von einem konstanten „Punkt“ M_0 nicht größer (bzw. kleiner) als r ist. Offenbar ist diese „Kugel“ für $n = 2$ ein Kreis (vgl. Nr. 160) und für $n = 3$ die gewöhnliche Kugel.

Die offene „Kugel“ mit beliebigem Radius $r > 0$ und mit dem „Mittelpunkt“ $M_0(x_1^0, \ldots, x_n^0)$ kann ebenfalls als „Umgebung“ dieses Punktes betrachtet werden; zum Unterschied von den früher betrachteten (parallelepipedischen) Umgebungen spricht man hier von einer „sphärischen“ („Kugel“-)Umgebung.

Es ist zweckmäßig, sich darüber klar zu werden, daß in jeder Kugelumgebung eines Punktes eine Würfelumgebung und in jeder Würfelumgebung eine Kugelumgebung enthalten ist.

Es sei zunächst ein „Paralelepiped“ (3) mit dem „Mittelpunkt“ M_0 gegeben. Es genügt, eine offene „Kugel“ mit demselben Mittelpunkt und einem Radius r, der kleiner als alle δ_i $(i = 1, 2, \ldots, n)$ ist, zu nehmen, damit diese „Kugel“ in dem gegebenen „Parallelepiped“ enthalten ist. In der Tat gilt für jeden „Punkt“ $M(x_1, x_2, \ldots, x_n)$ dieser „Kugel“ (für jedes $i = 1, 2, \ldots, n$)

$$|x_i - x_i^0| \leqq \sqrt{\sum_{k=1}^{n} (x_k - x_k^0)^2} = \overline{MM_0} < r < \delta_i$$

oder

$$x_i^0 - \delta_i < x_i < x_i^0 + \delta_i,$$

so daß dieser „Punkt“ zu dem gegebenen „Parallelepiped“ gehört.

Wenn umgekehrt ursprünglich eine „Kugel“ mit dem Radius r und dem „Mittelpunkt“ M_0 gegeben ist, so ist ein „Parallelepiped“ (3), z. B. für $\delta_1 = \delta_2 = \cdots = \delta_n = \dfrac{r}{\sqrt{n}}$, in ihr enthalten. Das folgt daraus, daß jeder „Punkt“ $M(x_1, x_2, \ldots, x_n)$ dieses „Parallelepipeds“ von dem „Punkt“ M_0 um

$$\overline{MM_0} = \sqrt{\sum_{k=1}^{n} (x_k - x_k^0)^2} < \sqrt{\sum_{k=1}^{n} \delta_k^2} = r$$

entfernt ist, also zur gegebenen „Kugel“ gehört.

163. Allgemeine Definition des offenen und des abgeschlossenen Bereichs. Wir nennen einen „Punkt" $M'(x_1', x_2', \ldots, x_n')$ einen *inneren* „Punkt" einer Menge $\mathcal{M}$ (im n-dimensionalen „Raum"), wenn er nebst einer hinreichend kleinen Umgebung zur Menge $\mathcal{M}$ gehört.

Aus der am Schluß von Nr. 162 bewiesenen Aussage folgt offenbar, daß es gleichgültig ist, um welchen Typ einer Umgebung es sich handelt, um eine „parallelepipedische" (Würfel-) oder eine „sphärische" (Kugel-) Umgebung.

Für ein *offenes* „rechtwinkliges Parallelepiped"

$$(a_1, b_1; \ldots; a_n, b_n) \tag{4}$$

ist jeder „Punkt" ein innerer Punkt. In der Tat, ist

$$a_1 < x_1' < b_1, \ldots, a_n < x_n' < b_n,$$

so kann man leicht ein solches $\delta > 0$ finden, daß

$$a_1 < x_1' - \delta < x_1' + \delta < b_1, \ldots, a_n < x_n' - \delta < x_n' + \delta < b_n$$

gilt.

Analog ist bei einer offenen „Kugel" mit dem Radius r und dem „Mittelpunkt" M_0 jeder ihrer „Punkte" M' ein innerer Punkt. Wenn wir ϱ so wählen, daß

$$0 < \varrho < r - \overline{M'M_0}$$

ist und um M' eine „Kugel" mit dem Radius ϱ beschreiben, so ist diese ganz in der ursprünglichen „Kugel" enthalten: Sobald $\overline{MM'} < \varrho$ ist, gilt (vgl. Nr. 161, Formel (2))

$$\overline{MM_0} \leqq \overline{MM'} + \overline{M'M_0} < \varrho + \overline{M'M_0} < r;$$

also gehört der „Punkt" M zur ursprünglichen „Kugel".

Dieselbe Schlußfolgerung kann man auch für ein offenes Simplex

$$x_1 > 0, \ldots, x_n > 0, \qquad x_1 + \cdots + x_n < h \quad (h > 0)$$

ziehen.

Mengen dieser Art, die ganz aus inneren „Punkten" bestehen, werden „*offene Bereiche*" genannt.

Somit sind das offene „rechtwinklige Parallelepiped", die offene „Kugel", das offene Simplex Beispiele offener Bereiche.

Wir verallgemeinern jetzt den Begriff des Häufungspunktes (Nr. 52) auf den Fall einer Menge $\mathcal{M}$ im n-dimensionalen „Raum". Ein „Punkt" M_0 heißt „*Häufungspunkt*" der Menge $\mathcal{M}$, wenn in jeder Umgebung (gleichgültig welchen Typs) von M_0 ein von M_0 verschiedener „Punkt" der Menge $\mathcal{M}$ enthalten ist.

Ein Häufungspunkt eines offenen Bereichs, der nicht zu ihm gehört, heißt „*Randpunkt*" dieses Bereichs. Die Gesamtheit der Randpunkte bildet den „*Rand des Bereichs*". Ein offener „Bereich" zusammen mit dem „Rand" heißt „*abgeschlossener Bereich*".

Man sieht leicht, daß für ein offenes „Parallelepided" (4) die Punkte $M(x_1, \ldots, x_n)$, für die

$$a_1 \leqq x_1 \leqq b_1, \ldots, a_n \leqq x_n \leqq b_n$$

ist, Randpunkte sind, wenn dabei mindestens in einem Fall Gleichheit gilt.

Genauso sind für die oben betrachteten offenen „Kugeln" alle „Punkte" M, für die $\overline{MM_0} = r$ ist, Randpunkte.

Schließlich sind für ein offenes Simplex (5) die Punkte $M(x_1, \ldots, x_n)$, die den Beziehungen

$$x_1 \geqq 0, \ldots, x_n \geqq 0, \quad x_1 + \cdots + x_n \leqq h$$

genügen, Randpunkte, wenn auch nur in einem einzigen Fall Gleichheit gilt.

Somit sind das abgeschlossene „rechtwinklige Parallelepiped", die abgeschlossene „Kugel" und das abgeschlossene Simplex Beispiele für abgeschlossene „Bereiche".

Wenn wir im folgenden von (offenen oder abgeschlossenen) „Bereichen" sprechen, meinen wir immer „Bereiche" in diesem speziellen Sinne (also „Kugeln", „Parallelepipede", Simplexe).

Wir wollen jetzt zeigen, daß ein abgeschlossener „Bereich" alle seine Häufungspunkte enthält.

Es seien ein abgeschlossener „Bereich" $\overline{\mathfrak{D}}$ und außerhalb von ihm ein „Punkt" M_0 gegeben. Wir beweisen, daß dann M_0 kein Häufungspunkt $\overline{\mathfrak{D}}$ sein kann.

Den abgeschlossenen Bereich $\overline{\mathfrak{D}}$ erhält man aus einem offenen Bereich $\mathfrak{D}$, indem man zu ihm noch seinen Rand $\mathfrak{E}$ hinzunimmt. Offenbar ist M_0 kein Häufungspunkt für $\mathfrak{D}$; folglich kann M_0 so von einer offenen „Kugel" umgeben werden, daß in ihr keine „Punkte" aus $\mathfrak{D}$ enthalten sind. Dann können aber auch keine „Punkte" aus $\mathfrak{E}$ in ihr enthalten sein. Wäre nämlich irgendein „Punkt" M' aus $\mathfrak{E}$ darin enthalten, so wäre darin auch eine ganze Umgebung des „Punktes" M' enthalten, und in dieser Umgebung gäbe es keinen „Punkt" aus $\mathfrak{D}$, im Gegensatz zur Definition des Häufungspunktes und der Menge $\mathfrak{E}$ als Rand. Somit gibt es in der erwähnten „Kugel" keine „Punkte" aus $\overline{\mathfrak{D}}$, und damit ist unsere Behauptung bewiesen.

Allgemein wird eine Menge von Punkten, die ihre sämtlichen Häufungspunkte enthält, *abgeschlossen* genannt. Somit ist ein abgeschlossener Bereich ein Spezialfall einer abgeschlossenen Menge.

Wir führen nun noch eine Reihe von Bezeichnungen ein. Eine „Punktmenge" $\mathcal{M}$ heißt *beschränkt*, wenn sie ganz in einem „rechtwinkligen Parallelepiped" enthalten ist.

Ein „Bereich" heißt *zusammenhängend*, wenn je zwei seiner Punkte durch einen „Polygonzug", dessen sämtliche Punkte im „Bereich" liegen, verbunden werden können. In Abb. 96 sind einige ebene zusammenhängende Bereiche dargestellt.

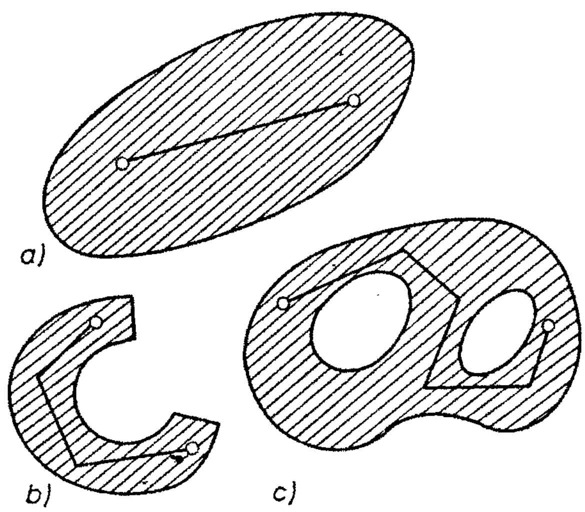

Abb. 96

Ein (offener oder abgeschlossener) beschränkter und zusammenhängender Bereich in einem n-dimensionalen Raum ist in gewissem Sinn einem endlichen (offenen bzw. abgeschlossenen) Intervall analog. Dem Leser dürfte aber klar sein, wie sich das Bild beim Übergang vom zweidimensionalen zum n-dimensionalen ($n \geqq 2$) Fall kompliziert. Dem einfachen und typischen Intervall, dessen Rand ausschließlich aus zwei Punkten besteht, steht eine Vielzahl von Bereichen mit komplizierten Begrenzungen (Rändern) gegenüber.

Alle Ausführungen der letzten Nummern sind als bloße geometrische Analogien zu betrachten, für die es (im Fall $n > 3$) keine geometrische Darstellung im gewöhnlichen Raum gibt. Es sei jedoch darauf hingewiesen, daß der n-dimensionale (arithmetische) Raum tatsächlich nur der erste Schritt zu einer höchst fruchtbaren Verallgemeinerung des Raumbegriffs ist, die vielen Teilgebieten der modernen Analysis zugrunde liegt.[1])

164. Funktionen von n Veränderlichen. Es seien n Veränderliche $x_1, x_2, \ldots, x_n$ gegeben, deren Werte einer Punktmenge $\mathcal{M}$ des n-dimensionalen Raumes entnommen seien. Diese Veränderlichen nennt man *unabhängige Veränderliche*. Der Begriff der Funktion und alles, was in diesem Zusammenhang im Fall zweier unabhängiger Veränderlicher (Nr. 160) gesagt wurde, läßt sich unmittelbar auch auf unseren allgemeinen Fall übertragen, so daß wir nicht näher darauf einzugehen brauchen. Bezeichnet man den Punkt $(x_1, x_2, \ldots, x_n)$ mit M, so nennt man die Funktion $u = f(x_1, x_2, \ldots, x_n)$ dieser Veränderlichen manchmal auch *Funktion des Punktes M* und schreibt $u = f(M)$.

Wir wollen nun annehmen, auf einer Punktmenge $\mathcal{P}$ des n-dimensionalen Raumes seien n Funktionen von m Veränderlichen $t_1, t_2, \ldots, t_m$ (wobei m unabhängig von n ist) gegeben:

$$x_1 = \varphi_1(t_1, t_2, \ldots, t_m), \quad \ldots, \quad x_n = \varphi_n(t_1, t_2, \ldots, t_m) \tag{5}$$

oder kürzer

$$x_1 = \varphi_1(P), \ldots, x_n = \varphi_n(P), \tag{5a}$$

wobei P den Punkt $(t_1, t_2, \ldots, t_m)$ des m-dimensionalen Raumes bezeichnet. Wir nehmen überdies an, daß sich, wenn $P(t_1, t_2, \ldots, t_m)$ die Menge $\mathcal{P}$ durchläuft, der ihm entsprechende n-dimensionale Punkt M mit den Koordinaten (5) oder (5a) nur innerhalb der Grenzen der n-dimensionalen Menge $\mathcal{M}$ variiert, in der die Funktion $u = f(x_1, x_2, \ldots, x_n) = f(M)$ definiert ist.

Dann kann man die Veränderliche u als mittelbare Funktion der unabhängigen Veränderlichen $t_1, t_2, \ldots, t_m$ (in der Menge $\mathcal{P}$) ansehen, und zwar durch Vermittlung der Veränderlichen $x_1, x_2, \ldots, x_n$:

$$u = f\big(\varphi_1(t_1, t_2, \ldots, t_m), \ldots, \varphi_n(t_1, t_2, \ldots, t_m)\big);$$

u ist eine Funktion der Funktionen $\varphi_1, \ldots, \varphi_n$ (vgl. Nr. 51).

Diese Art der Definition einer mittelbaren Funktion durch die Funktionen $\varphi_1, \ldots, \varphi_n$ und die Funktion f nennt man (wie auch im einfachsten Fall der Funktion einer Veränderlichen) *Superposition* oder *Verkettung*.

Die Klasse der Funktionen mehrerer Veränderlicher, mit der wir es zunächst zu tun haben werden, ist nicht sehr groß. Im wesentlichen ergibt sie sich durch Verkettung aus elementaren Funktionen einer Veränderlichen (Nr. 48, 50) und folgenden Funktionen zweier Veränderlicher:

$$z = x \pm y, \quad z = xy, \quad z = \frac{x}{y} \quad \text{und} \quad z = x^y,$$

d. h. den vier Grundrechenarten und der Potenzfunktion.

---

[1]) Wir setzten bisher meist geometrische Termini, die in einer vom üblichen Sinne verschiedenen Bedeutung benutzt wurden, in Anführungszeichen, etwa „Punkt“, „Abstand“, „Bereich“ usw. Von jetzt ab werden wir generell darauf verzichten.

Die Grundrechenarten, wiederholt auf die unabhängigen Veränderlichen $x_1, x_2, \ldots, x_n$ und Konstante angewandt, führen auf Polynome:

$$P(x_1, x_2, \ldots, x_n) = \sum_{\nu_1, \nu_2, \ldots, \nu_n} C_{\nu_1, \nu_2, \ldots, \nu_n} x_1^{\nu_1} x_2^{\nu_2} \cdots x_n^{\nu_n}$$

(*ganze rationale Funktionen*)[1] und auf Quotienten zweier solcher Polynome

$$Q(x_1, x_2, \ldots, x_n) = \frac{\sum C_{\nu_1, \nu_2, \ldots, \nu_n} x_1^{\nu_1} x_2^{\nu_2} \cdots x_n^{\nu_n}}{\sum C'_{\mu_1, \mu_2, \ldots, \mu_n} x_1^{\mu_1} x_2^{\mu_2} \cdots x_n^{\mu_n}}$$

(*gebrochene rationale Funktionen*).

Zieht man die elementaren Funktionen einer Veränderlichen heran, so kommt man auf solche Funktionen wie

$$f(x, y, z) = \frac{\ln (x + y + z)}{\sqrt{x^2 + y^2 + z^2}},$$

$$\varphi(x, y, z, t) = \sin xy + \sin yz + \sin zt + \sin tx,$$

usw.

Die Bemerkungen, die wir in Nr. 46 über die Definition von Funktionen durch analytische Ausdrücke gemacht haben, lassen sich auf unseren Fall übertragen.

165. Grenzwert von Funktionen mehrerer Veränderlicher. Wir nehmen an, die Funktion $f(x_1, \ldots, x_n)$ sei auf einer Punktmenge $\mathcal{M}$ mit dem Häufungspunkt $M_0(a_1, a_2, \ldots, a_n)$ definiert.

Analog wie im Fall von Funktionen einer Veränderlichen sagt man, die *Funktion* $f(x_1, \ldots, x_n)$ *habe,* wenn die Veränderlichen $x_1, \ldots, x_n$ gegen $a_1, \ldots, a_n$ streben, *den Grenzwert* A, wenn zu jeder Zahl $\varepsilon > 0$ eine Zahl $\delta > 0$ existiert derart, daß

$$|f(x_1, \ldots, x_n) - A| < \varepsilon \tag{*}$$

gilt, sobald

$$|x_1 - a_1| < \delta, \ldots, |x_n - a_n| < \delta$$

ist.

Dabei wird der Punkt $(x_1, \ldots, x_n)$ als zu $\mathcal{M}$ gehörig und von dem Punkt $(a_1, \ldots, a_n)$ verschieden vorausgesetzt. Die Ungleichung (*) für die Funktion muß also für alle Punkte der Menge $\mathcal{M}$, die in einer hinreichend kleinen Umgebung

$$(a_1 - \delta, a_1 + \delta; \ldots; a_n - \delta, a_n + \delta)$$

von M_0 liegen, erfüllt sein, mit Ausnahme des Punktes M_0 (wenn er zu $\mathcal{M}$ gehört).

Den (*n-fachen*) Grenzwert bezeichnet man mit

$$A = \lim_{\substack{x_1 \to a_1 \\ \cdots\cdots \\ x_n \to a_n}} f(x_1, \ldots, x_n). \tag{6}$$

Führt man für die Punkte $(x_1, \ldots, x_n)$ und $(a_1, \ldots, a_n)$ die Bezeichnungen M bzw. M_0 ein, so kann man diese Definition in geometrischer Terminologie auch folgendermaßen formulieren: Eine Zahl A heißt *Grenzwert* der Funktion $f(M)$ für gegen M_0 konvergierende Punkte M (oder Grenzwert im Punkt M_0), wenn zu jedem $\varepsilon > 0$ eine Zahl

[1]) Wir erinnern an die Bedeutung des Summenzeichens. Hier handelt es sich um mehrfache Summen, d. h., die Summanden hängen von mehreren Summationsindizes ab.

$r > 0$ existiert derart, daß

$$|f(M) - A| < \varepsilon$$

gilt, sobald der Abstand $\overline{M_0 M}$ kleiner als r ist.

Wie oben wird dabei M als zu $\mathcal{M}$ gehörig und von M_0 verschieden vorausgesetzt. Somit muß die Ungleichung für die Funktion $f(M)$ in allen Punkten der Menge $\mathcal{M}$, die in einer hinreichend kleinen sphärischen Umgebung des Punktes M_0 liegen, mit Ausnahme des Punktes selbst, erfüllt sein.

Dementsprechend kann man den Grenzwert einer Funktion auch in der Form

$$A = \lim_{M \to M_0} f(M) \tag{6*}$$

schreiben.

Aus der Bemerkung in Nr. 161 über die Äquivalenz der verschiedenen Arten von Umgebungen folgt unmittelbar die Gleichwertigkeit dieser beiden Definitionen.

Analog kann man den Begriff eines unendlichen Grenzwertes einer Funktion definieren. Im Fall $A = \infty$ bzw. $-\infty$ ist die Ungleichung

$$|f(x_1, \ldots, x_n) - A| < \varepsilon$$

durch die Ungleichung

$$f(x_1, \ldots, x_n) > E \quad \text{bzw.} \quad f(x_1, \ldots, x_n) < -E$$

zu ersetzen, wobei E eine beliebig vorgegebene positive Zahl ist.

Wir erwähnen zum Schluß noch den Fall, daß einige der unabhängigen Veränderlichen $x_1, \ldots, x_n$ gegen einen unendlichen Grenzwert streben.

Man könnte den Begriff des Häufungspunktes $M_0(a_1, \ldots, a_n)$ eines Bereichs $\mathcal{M}$ auch auf den Fall erweitern, daß alle Koordinaten dieses Punktes (oder einige von ihnen) ∞ sind.[1])

Beispielsweise ist der Punkt $(\infty, \ldots, \infty)$ Häufungspunkt für $\mathcal{M}$, wenn in diesem Bereich ein Punkt mit beliebig großen (positiven) Koordinaten existiert.

Unter dieser Voraussetzung sagt man, die Funktion $f(x_1, \ldots, x_n)$ habe die Zahl A als Grenzwert (wenn alle Veränderlichen $x_1, \ldots, x_n$ gegen ∞ streben), falls zu jedem $\varepsilon > 0$ eine Zahl $\Delta > 0$ existiert derart, daß

$$|f(x_1, x_2, \ldots, x_n) - A| < \varepsilon$$

gilt, sobald

$$x_1 > \Delta, \quad x_2 > \Delta, \ldots, x_n > \Delta$$

ist. Man schreibt

$$A = \lim_{\substack{x_1 \to \infty \\ \cdots\cdots \\ x_n \to \infty}} f(x_1, \ldots, x_n).$$

Wir kommen nochmals auf die Veränderliche $x_{m,n}$ aus Nr. 160 zurück. Man sagt, diese Veränderliche habe bei unbeschränktem Wachsen beider Indizes m und n den Grenzwert A, wenn zu jedem $\varepsilon > 0$ eine Zahl N existiert derart, daß

$$|x_{m,n} - A| < \varepsilon \quad \text{für} \quad m > N \quad \text{und} \quad n > N$$

---

[1]) Dann heißt M_0 *uneigentlicher Punkt.*

ist, in Zeichen

$$A = \lim_{\substack{m\to\infty \\ n\to\infty}} x_{m,n} \quad \text{oder einfach} \quad A = \lim x_{m,n}.$$

Man sieht leicht, wie der Fall $A = \infty$ oder $-\infty$ zu behandeln ist.

166. Reduktion auf Folgen. Wir betrachten im n-dimensionalen Raum eine Punktfolge[1])

$$\{M_k(x_1^{(k)}, \ldots, x_n^{(k)})\} \quad (k = 1, 2, \ldots).$$

Wir sagen, daß diese Folge *gegen den Häufungspunkt* $M_0(a_1, \ldots, a_n)$ *konvergiert*, wenn für $k \to \infty$

$$\overline{M_0 M_k} \to 0 \tag{7}$$

gilt.

Statt dessen könnte man auch fordern, daß die Koordinaten des Punktes M_k einzeln gegen die entsprechenden Koordinaten des Punktes M_0 streben, d. h., daß

$$x_1^{(k)} \to a_1, \ldots, x_n^{(k)} \to a_n \tag{8}$$

gilt.

Die Äquivalenz der beiden Definitionen folgt aus der in Nr. 161 bewiesenen Aussage über die Äquivalenz der Umgebung der beiden Typen.

In der Tat bedeutet die Bedingung (7), daß für jede beliebig gewählte Zahl $r > 0$ der Punkt M_k für hinreichend große k der Ungleichung $\overline{M_0 M_k} < r$ genügt, d. h. in der (offenen) Kugel mit dem Radius r um M_0 liegt; die Forderung (8) besagt, daß bei jeder beliebig gewählten Zahl $\delta > 0$ der betreffende Punkt (wiederum für hinreichend großes k) den Ungleichungen

$$|x_1^{(k)} - a_1| < \delta, \ldots, |x_n^{(k)} - a_n| < \delta$$

genügt, d. h. in dem (offenen) Parallelepiped

$$(a_1 - \delta, a_1 + \delta; \ldots; a_n - \delta, a_n + \delta)$$

mit dem Mittelpunkt in demselben Punkt enthalten ist.

Der Punkt $M_0(a_1, \ldots, a_n)$ sei Häufungspunkt einer Menge $\mathcal{M}$ im n-dimensionalen Raum. Dann kann man aus $\mathcal{M}$ immer eine Folge $\{M_k\}$ von M_0 verschiedener Punkte auswählen, die gegen M_0 konvergiert. Um das zu beweisen, gehen wir von einer positiven diskreten Veränderlichen $r_k \to 0$ aus. Nach Definition des Häufungspunktes (Nr. 163) gibt es in jeder sphärischen Umgebung von M_0 mit dem Radius r_k einen von M_0 verschiedenen Punkt M_k der Menge $\mathcal{M}$. Die Folge $\{M_k\}$ ist dann offenbar die gesuchte.

Jetzt kann man eine Bedingung formulieren, die notwendig und hinreichend für die Existenz des Grenzwertes (6) bzw. (6*) ist:

Dieser Grenzwert existiert genau dann, wenn für jede gegen M_0 strebende Folge $\{M_k\}$ von Punkten aus $\mathcal{M}$, die von M_0 verschieden sind, die Zahlenfolge $\{f(M_k)\}$ der entsprechenden Funktionswerte gegen A konvergiert.

Die Bedingung ist notwendig. Es gelte (6*), und zu einem gegebenen $\varepsilon > 0$ möge ein $r > 0$ in Einklang mit der Definition aus Nr. 165 gefunden sein. Wenn die Folge

[1]) Die Übertragung der Definition vom ebenen auf den allgemeinen Fall $n > 2$ überlassen wir dem Leser.

der Punkte M_k gegen M_0 konvergiert, so gilt $\overline{M_0 M_k} < r$ für hinreichend großes k; daraus folgt die Ungleichung

$$|f(M_k) - A| < \varepsilon,$$

die gerade besagt, daß $f(M_k)$ gegen A strebt.

Die Bedingung ist hinreichend. Wir nehmen jetzt an, die obige Bedingung sei erfüllt. Um zu beweisen, daß in Übereinstimmung mit der Definition aus Nr. 165 die Gleichung (6*) besteht, nehmen wir das Gegenteil der Aussage dieser Definition an. Dann existiert zu einer Zahl $\varepsilon > 0$ kein entsprechendes r, d. h., wie immer auch die Zahl $r > 0$ gewählt wird, stets läßt sich in $\mathcal{M}$ ein (von M_0 verschiedener) Punkt M' finden derart, daß gleichzeitig

$$\overline{M_0 M'} < r, \quad \text{aber} \quad |f(M') - A| \geqq \varepsilon$$

gilt.

Wir wählen eine positive Nullfolge $\{r_k\}$ und nehmen für r der Reihe nach die Zahlen r_k; dann läßt sich nach dem Gesagten zu jedem r_k ein (von M_0 verschiedener) Punkt M_k angeben, für den

$$\overline{M_0 M_k} < r_k, \quad \text{aber} \quad |f(M_k) - A| \geqq \varepsilon$$

gilt. Die so konstruierte Punktfolge $\{M_k\}$ konvergiert gegen M_0, während die Zahlenfolge $\{f(M_k)\}$ nicht den Grenzwert A haben kann, entgegen der Voraussetzung. Mit diesem Widerspruch ist aber unsere Behauptung bewiesen.

Offenbar liefert die eben bewiesene Bedingung eine andere Definition des Grenzwertes einer Funktion (in der „Sprache der Folgen").

Somit läßt sich auch für Funktionen mehrerer Veränderlicher das Problem des Grenzwertes einer Funktion auf den Grenzwert von Folgen (Nr. 53) zurückführen. Das Ergebnis kann leicht auch auf den Fall ausgedehnt werden, daß die Zahlen A, $a_1, \ldots, a_n$ oder einige von ihnen ∞ sind.

Dieser Sachverhalt gestattet es, alle grundlegenden Begriffe und Sätze der in Kapitel I entwickelten Theorie der Grenzwerte in der gleichen Weise auch auf den neuen Typus des Grenzwertes auszudehnen, wie wir es in Nr. 55 für den Grenzwert von Funktionen einer unabhängigen Veränderlichen getan haben.

167. Beispiele.

1. Wenn man den Satz über den Grenzwert eines Produktes benutzt, so läßt sich zeigen, daß

$$\lim_{\substack{x_1 \to a_1 \\ \cdots\cdots \\ x_n \to a_n}} C x_1^{\nu_1} \cdots x_n^{\nu_n} = C a_1^{\nu_1} \cdots a_n^{\nu_n}$$

gilt, wobei $C, a_1, \ldots, a_n$ beliebige reelle und $\nu_1, \ldots, \nu_n$ nichtnegative ganze Zahlen sind. Bezeichnet man mit $P(x_1, \ldots, x_n)$ eine ganze rationale Funktion (Nr. 164),

$$P(x_1, \ldots, x_n) = \sum_{\nu_1, \ldots, \nu_n} C_{\nu_1, \ldots, \nu_n} x_1^{\nu_1} \cdots x_n^{\nu_n},$$

so erhält man nach dem Satz über den Grenzwert einer Summe

$$\lim_{\substack{x_1 \to a_1 \\ \cdots\cdots \\ x_n \to a_n}} P(x_1, \ldots, x_n) = P(a_1, \ldots, a_n).$$

Analog erhält man für eine gebrochene rationale Funktion (Nr. 164)

$$Q(x_1, \ldots, x_n) = \frac{\sum C_{\nu_1, \ldots, \nu_n} x_1^{\nu_1} \cdots x_n^{\nu_n}}{\sum C'_{\mu_1, \ldots, \mu_n} x_1^{\mu_1} \cdots x_n^{\mu_n}}$$

nach dem Satz über den Grenzwert eines Quotienten

$$\lim_{\substack{x_1 \to a_1 \\ \cdots\cdots \\ x_n \to a_n}} Q(x_1, \ldots, x_n) = Q(a_1, \ldots, a_n),$$

natürlich unter der Voraussetzung, daß der Nenner im Punkt $(a_1, \ldots, a_n)$ nicht 0 wird.

2. Wir betrachten die Potenzfunktion x^y für $x > 0$ und beliebiges y. Sind $a > 0$ und b beliebige reelle Zahlen, so gilt

$$\lim_{\substack{x \to a \\ y \to b}} x^y = a^b.$$

In der Tat gilt für beliebige Folgen $x_n \to a$ und $y_n \to b$ nach Nr. 78

$$x_n^{y_n} = e^{y_n \cdot \ln x_n} \to e^{b \cdot \ln a} = a^b;$$

das liefert in der „Sprache der Folgen" das geforderte Ergebnis.

3. Für die Folgen $\{x_n\}$ und $\{y_n\}$ sei bekannt, daß sie die Grenzwerte a bzw. b haben. Wir fragen nach dem Grenzwert der Ausdrücke

$$x_n \pm y_n, \quad x_n \cdot y_n, \quad \frac{x_n}{y_n}, \quad x_n^{y_n}$$

bei den sogenannten unbestimmten Ausdrücken, die durch die Symbole

$$\infty - \infty, \; 0 \cdot \infty, \; \frac{0}{0}, \; \frac{\infty}{\infty}, \; 1^\infty, \; 0^0, \; \infty^0$$

charakterisiert werden. Wie wir wissen (Nr. 31 und 78), braucht der Grenzwert überhaupt nicht zu existieren, und wenn er existiert, so kann er für dieselben a und b verschiedene Werte haben, je nach dem speziellen Bildungsgesetz der Folgen $\{x_n\}$ und $\{y_n\}$.

Wenn wir uns an die Definition des Grenzwertes von Funktionen zweier unabhängiger Veränderlicher in der „Sprache der Folgen" erinnern, so wird klar, daß die erwähnte „Unbestimmtheit" damit zusammenhängt, daß die Grenzwerte

$$\lim_{\substack{x \to \infty \\ y \to \infty}} (x - y), \quad \lim_{\substack{x \to 0 \\ y \to \pm\infty}} x \cdot y, \quad \lim_{\substack{x \to 0 \\ y \to 0}} \frac{x}{y}, \quad \lim_{\substack{x \to \pm\infty \\ y \to \pm\infty}} \frac{x}{y},$$

$$\lim_{\substack{x \to 1 \\ y \to \pm\infty}} x^y, \quad \lim_{\substack{x \to 0 \\ y \to 0}} x^y, \quad \lim_{\substack{x \to \infty \\ y \to 0}} x^y$$

nicht existieren.

4. Wir fragen nach dem Grenzwert $\lim\limits_{\substack{x \to 0 \\ y \to 0}} \frac{xy}{x^2 + y^2}$. (Die Funktion ist hier in der ganzen Ebene mit Ausnahme des Punktes $x = 0, y = 0$ definiert.) Wir betrachten die beiden Teilfolgen

$$\left\{M_k\left(\frac{1}{k}, \frac{1}{k}\right)\right\} \quad \text{und} \quad \left\{M'_k\left(\frac{2}{k}, \frac{1}{k}\right)\right\},$$

die gegen den Punkt (0, 0) konvergieren; es zeigt sich, daß für alle k

$$f(M_k) = f\left(\frac{1}{k}, \frac{1}{k}\right) = \frac{1}{2} \quad \text{und} \quad f(M_k') = f\left(\frac{2}{k}, \frac{1}{k}\right) = \frac{2}{5}$$

gilt. Hieraus folgt schon, daß der erwähnte Grenzwert nicht existiert.

Analog kann man sich davon überzeugen, daß der Grenzwert $\lim\limits_{\substack{x\to 0\\ y\to 0}} \frac{x^2 - y^2}{x^2 + y^2}$ nicht existiert.

5. Dagegen existiert der Grenzwert $\lim\limits_{\substack{x\to 0\\ y\to 0}} \frac{x^2 y}{x^2 + y^2}$ und ist gleich 0. Das folgt sofort aus der Ungleichung

$$\left|\frac{x^2 y}{x^2 + y^2}\right| \leqq \frac{1}{2}|x|.$$

Genauso kann man beweisen, daß auch $\lim\limits_{\substack{x\to 0\\ y\to 0}} \frac{x^3 + y^3}{x^2 + y^2} = 0$ gilt.

168. Iterierte Grenzwerte. Außer dem oben betrachteten Grenzwert einer Funktion $f(x_1, x_2, \ldots, x_n)$ für gleichzeitig gegen ihre Grenzwerte strebende Argumente hat man es auch noch mit Grenzwerten anderer Art zu tun, nämlich mit Grenzübergängen, bei denen sukzessive die Argumente einzeln in irgendeiner bestimmten Reihenfolge gegen ihre Grenzwerte streben. Die oben betrachteten Grenzwerte sind n-fache (zweifache, dreifache usw.) Grenzwerte, während die letzteren *iterierte Grenzwerte* genannt werden.

Wir beschränken uns auf den einfachsten Fall einer Funktion $f(x, y)$ zweier Veränderlicher. Wir nehmen dabei an, daß der Variationsbereich der Veränderlichen x, y so beschaffen ist, daß x (unabhängig von y) jeden Wert einer Menge $\mathscr{X}$ annehmen kann, für den a Häufungspunkt ist, aber nicht zu $\mathscr{X}$ gehört; analog möge y (unabhängig von x) in einer Menge $\mathscr{Y}$ variieren, die den nicht zu ihr gehörenden Häufungspunkt b habe. Einen solchen Bereich könnte man symbolisch mit $\mathscr{X} \times \mathscr{Y}$ bezeichnen; beispielsweise

$$(a, a + H; b, b + K) = (a, a + H) \times (b, b + K).$$

Wenn zu jedem festen y aus $\mathscr{Y}$ für die Funktion $f(x, y)$, die eine Funktion von x allein ist, ein Grenzwert für $x \to a$ existiert, so wird dieser Grenzwert im allgemeinen von dem vorgegebenen festen y abhängen:

$$\lim_{x\to a} f(x, y) = \varphi(y).$$

Alsdann kann man die Frage nach dem Grenzwert der Funktion $\varphi(y)$ für $y \to b$ stellen:

$$\lim_{y\to b} \varphi(y) = \lim_{y\to b} \lim_{x\to a} f(x, y).$$

Dies ist einer von zwei iterierten Grenzwerten. Den anderen erhält man, wenn man die Grenzübergänge in der umgekehrten Reihenfolge durchführt:

$$\lim_{x\to a} \lim_{y\to b} f(x, y).$$

Man darf nicht glauben, daß diese iterierten Grenzwerte einander gleich sein müssen. Wenn wir z. B. in dem Gebiet $M(0, \infty; 0, \infty)$

$$f(x, y) = \frac{x - y + x^2 + y^2}{x + y} \tag{a}$$

setzen und $a = b = 0$ annehmen, so erhalten wir

$$\varphi(y) = \lim_{x\to 0} f(x, y) = y - 1, \qquad \lim_{y\to 0}\varphi(y) = \lim_{y\to 0}\lim_{x\to 0} f(x, y) = -1,$$

aber

$$\psi(x) = \lim_{y\to 0} f(x, y) = x + 1, \qquad \lim_{x\to 0}\psi(x) = \lim_{x\to 0}\lim_{y\to 0} f(x, y) = 1.$$

Es kann auch vorkommen, daß der eine der iterierten Grenzwerte existiert und der andere nicht. So existiert z. B. für die Funktionen

$$f(x, y) = \frac{x \cdot \sin\dfrac{1}{x} + y}{x + y} \tag{b}$$

oder

$$f(x, y) = x \cdot \sin\frac{1}{y} \tag{c}$$

in beiden Fällen der iterierte Grenzwert $\lim\limits_{y\to 0}\lim\limits_{x\to 0} f$, nicht aber der iterierte Grenzwert $\lim\limits_{x\to 0}\lim\limits_{y\to 0} f$ (im letzten Beispiel existiert nicht einmal der einfache Grenzwert $\lim\limits_{y\to 0} f$).

Diese einfachen Beispiele zeigen, wieviel *Vorsicht beim Vertauschen zweier Grenzübergänge bei verschiedenen Veränderlichen* nötig ist; viele fehlerhafte Schlußfolgerungen entstehen gerade aus solchen unerlaubten Vertauschungen. Zugleich hängen viele wichtige Fragen der Analysis mit der Vertauschung von Grenzübergängen zusammen; dabei muß natürlich jedes Mal besonders begründet werden, daß das Vertauschen erlaubt ist.

Eine Methode zum Beweis der Vertauschbarkeit liefert der folgende Satz, der zugleich einen Zusammenhang zwischen den zweifachen und den iterierten Grenzwerten vermittelt.

Satz. *Wenn erstens der (endliche oder unendliche) zweifache Grenzwert*

$$A = \lim_{\substack{x\to a\\ y\to b}} f(x, y)$$

existiert und zweitens für jedes y aus $\mathcal{Y}$ der (endliche) einfache Grenzwert bezüglich x,

$$\varphi(y) = \lim_{x\to a} f(x, y),$$

existiert, so existiert auch der iterierte Grenzwert

$$\lim_{y\to b}\varphi(y) = \lim_{y\to b}\lim_{x\to a} f(x, y)$$

und ist gleich dem zweifachen Grenzwert.

Wir beweisen diesen Satz für den Fall endlicher A, a und b. Nach der Definition aus Nr. 165 gibt es zu gegebenem $\varepsilon > 0$ ein $\delta > 0$ derart, daß

$$|f(x, y) - A| < \varepsilon \tag{9}$$

gilt, sobald $|x - a| < \delta$ und $|y - b| < \delta$ ist (wobei x aus $\mathcal{X}$ und y aus $\mathcal{Y}$ ist). Wir wählen jetzt y so, daß die Ungleichung $|y - b| < \delta$ erfüllt ist, und gehen in (9) zur Grenze für $x \to a$ über. Da nach Voraussetzung dabei $f(x, y)$ gegen den Grenzwert $\varphi(y)$ strebt, erhalten wir

$$|\varphi(y) - A| \leqq \varepsilon.$$

Nun erinnern wir uns, daß y hier eine beliebige Zahl aus $\mathcal{Y}$ ist, die der Bedingung $|y - b| < \delta$ genügt; somit kommen wir zu der Schlußfolgerung, daß

$$A = \lim_{y \to b} \varphi(y) = \lim_{y \to b} \lim_{x \to a} f(x, y)$$

gilt, was zu beweisen war.

Wenn außer den beiden Voraussetzungen zu jedem x aus $\mathcal{X}$ auch noch der (endliche) einfache Grenzwert bezüglich y,

$$\psi(x) = \lim_{y \to b} f(x, y),$$

existiert, so existiert, wie aus dem schon Bewiesenen folgt, wenn x und y ihre Rollen vertauschen, auch der zweite iterierte Grenzwert

$$\lim_{x \to a} \psi(x) = \lim_{x \to a} \lim_{y \to b} f(x, y)$$

und ist gleich derselben Zahl A: *In diesem Fall sind die beiden iterierten Grenzwerte gleich.*

Aus diesem Satz ist sofort klar, daß in den Beispielen (a) und (b) der zweifache Grenzwert nicht existiert (warum?). Hiervon kann man sich leicht auch direkt überzeugen.

Im Beispiel (c) existiert dagegen der zweifache Grenzwert: Aus der Ungleichung

$$\left| x \cdot \sin \frac{1}{y} \right| \leqq |x|$$

ersehen wir, daß er gleich 0 ist. Dieses Beispiel zeigt, daß die erste Voraussetzung des Satzes keineswegs die zweite Voraussetzung zur Folge hat.

Man darf jedoch nicht denken, daß die Existenz des zweifachen Grenzwertes für die Gleichheit der iterierten Grenzwerte notwendig ist: Im Beispiel 4 aus Nr. 167 existieren die beiden iterierten Grenzwerte und sind gleich 0, obwohl der zweifache Grenzwert nicht existiert.

§ 2. Stetige Funktionen

169. Stetigkeit und Unstetigkeit von Funktionen mehrerer Veränderlicher. Die Funktion $f(x_1, \ldots, x_n)$ sei in einer Punktmenge $\mathcal{M}$ des n-dimensionalen Raumes definiert, $M'(x_1', \ldots, x_n')$ sei ein Häufungspunkt dieser Menge und gehöre selbst zu $\mathcal{M}$.

Die Funktion $f(x_1, \ldots, x_n)$ heißt im Punkt $M'(x_1', \ldots, x_n')$ *stetig*, wenn die Beziehung

$$\lim_{\substack{x_1 \to x_1' \\ \cdots\cdots \\ x_n \to x_n'}} f(x_1, \ldots, x_n) = f(x_1', \ldots, x_n') \tag{1}$$

gilt; andernfalls heißt die Funktion im Punkt M' *unstetig*.

In der „ε-δ-Sprache“ kann man die Stetigkeit einer Funktion in einem Punkt M' folgendermaßen ausdrücken (Nr. 165): Zu beliebig vorgegebenem $\varepsilon > 0$ muß sich ein $\delta > 0$ finden lassen derart, daß

$$|f(x_1, \ldots, x_n) - f(x_1', \ldots, x_n')| < \varepsilon \tag{2}$$

ist, sobald

$$|x_1 - x_1'| < \delta, \ldots, |x_n - x_n'| < \delta \tag{3}$$

gilt; oder auch anders: Zu $\varepsilon > 0$ muß ein $r > 0$ existieren derart, daß

$$|f(M) - f(M')| < \varepsilon$$

ist, sobald für den Abstand $\overline{MM'} < r$ gilt.

Dabei wird vorausgesetzt, daß der Punkt $M(x_1, \ldots, x_n)$ zur Menge $\mathcal{M}$ gehört; insbesondere kann er auch mit dem Punkt M' zusammenfallen. Mit Rücksicht darauf, daß der Grenzwert der Funktion im Punkt M' gleich ihrem Wert in diesem Punkt ist, wird die übliche Forderung, daß M von M' verschieden sein soll, hier unnötig.

Sieht man die Differenzen $x_1 - x_1', \ldots, x_n - x_n'$ als Zuwächse $\Delta x_1', \ldots, x_n'$ der unabhängigen Veränderlichen an und die Differenz

$$f(x_1, \ldots, x_n) - f(x_1', \ldots, x_n')$$

als Zuwachs der Funktion, so kann man (ebenso wie im Fall von Funktionen einer einzigen Veränderlichen) sagen, die Funktion sei *stetig*, wenn einem unendlich kleinen Zuwachs der unabhängigen Veränderlichen ein unendlich kleiner Zuwachs der Funktion entspricht.

Die oben definierte Stetigkeit einer Funktion in einem Punkt M' is sozusagen eine Stetigkeit in der Gesamtheit der Veränderlichen $x_1, \ldots, x_n$. Wenn sie vorliegt, so ist gleichzeitig auch

$$\lim_{x_1 \to x_1'} f(x_1, x_2', \ldots, x_n') = f(x_1', x_2', \ldots, x_n'),$$

$$\lim_{\substack{x_1 \to x_1' \\ x_2 \to x_2'}} f(x_1, x_2, x_3', \ldots, x_n') = f(x_1', x_2', x_3', \ldots, x_n'),$$

usw.; denn hierbei wird jeweils nur eine bestimmte Art vorgeschrieben, in der sich M dem Punkt M' nähert. Mit anderen Worten, eine stetige Funktion ist insbesondere in jeder einzelnen Veränderlichen x_i stetig, in bezug auf jedes Paar von Veränderlichen x_i, x_j stetig, in bezug auf jedes Tripel, Quadrupel usw.

Mit Beispielen stetiger Funktionen haben wir uns schon befaßt. So war in Nr. 167, Beispiel 1, die Stetigkeit der ganzen und der gebrochenen rationalen Funktionen von n Argumenten in allen Punkten des n-dimensionalen Raumes nachgewiesen worden (die der gebrochenen Funktionen mit Ausnahme derjenigen Punkte, in denen der Nenner 0 wird), ferner in Beispiel 2 die Stetigkeit der Potenzfunktion x^y in allen Punkten der rechten Halbebene ($x > 0$).

Ergänzt man die in der ganzen Ebene mit Ausnahme des Ursprungs definierte Funktion

$$f(x, y) = \frac{xy}{x^2 + y^2} \qquad (\text{für } x^2 + y^2 > 0)$$

durch die Festsetzung $f(0, 0) = 0$, so erhält man ein Beispiel für eine unstetige Funktion. Sie ist im Ursprung unstetig; denn nach Nr. 167, Beispiel 4, hat sie für $x \to 0$, $y \to 0$ keinen Grenzwert. Hier beobachten wir nun ein interessantes Phänomen. Die betrachtete Funktion $f(x, y)$ ist zwar im Punkt $(0, 0)$ in bezug auf beide Veränderlichen nicht stetig, aber in diesem Punkt sowohl in x als auch in y einzeln stetig; das folgt aus $f(x, 0) = f(0, y) = 0$. Im übrigen ist das durchaus verständlich, wenn man bedenkt, daß bei der Stetigkeit in $(0, 0)$ in x bzw. in y einzeln nur die Annäherung an den Punkt $(0, 0)$ längs der x- bzw. der y-Achse in Betracht gezogen wird, während unendlich viele andere Arten der Annäherung beiseite gelassen werden.

Wenn für eine Funktion $f(M)$ für gegen M' strebendes M kein wohlbestimmter endlicher Grenzwert

$$\lim_{M \to M'} f(M)$$

existiert, so sagt man, die Funktion sei im Punkt M' *unstetig*, selbst wenn $f(M)$ im Punkt M' nicht definiert ist (vgl. die Bemerkung in Nr. 66).

Die Unstetigkeitspunkte einer Funktion können isolierte Punkte sein, wie im vorigen Beispiel, können aber auch eine Linie, eine Fläche usw. ausfüllen. So haben z. B. die Funktionen zweier Veränderlicher

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

folgende Unstetigkeitsstellen: die erste die Geraden $y = \pm x$, die zweite den Kreis $x^2 + y^2 = 1$. Für die Funktionen dreier Veränderlicher

$$\frac{x + y + z}{xy - z}, \quad \frac{1}{x^2 + y^2 - z^2}$$

erfüllen die Unstetigkeitsstellen folgende Flächen: das hyperbolische Paraboloid $z = xy$ bzw. den Kegel $z^2 = x^2 + y^2$.

170. Das Rechnen mit stetigen Funktionen. Der Satz über die Stetigkeit der Summe, der Differenz, des Produktes und des Quotienten zweier stetiger Funktionen (vgl. Nr. 67) läßt sich leicht formulieren und beweisen; das möge der Leser selbst tun.

Wir gehen nur auf die Verkettung stetiger Funktionen näher ein. Wie in Nr. 164 nehmen wir an, neben einer in einer Menge $\mathcal{M}$ von n-dimensionalen Punkten $M(x_1, \ldots, x_n)$ gegebenen Funktion $u = f(x_1, \ldots, x_n)$ seien noch n Funktionen

$$x_1 = \varphi_1(t_1, \ldots, t_m), \quad \ldots, \quad x_n = \varphi_n(t_1, \ldots, t_m) \tag{4}$$

in einer Menge $\mathcal{P}$ von m-dimensionalen Punkten $P(t_1, \ldots, t_m)$ definiert, und der Punkt M mit den Koordinaten (4) liege nicht außerhalb der Menge $\mathcal{M}$.

Satz. *Sind die Funktionen $\varphi_i(P)$ $(i = 1, \ldots, n)$ sämtlich im Punkt $P'(t_1', \ldots, t_m')$ aus $\mathcal{P}$ stetig und ist die Funktion $f(M)$ im entsprechenden Punkt $M'(x_1', \ldots, x_n')$ mit den Koordinaten*

$$x_1' = \varphi_1(t_1', \ldots, t_m'), \quad \ldots, \quad x_n' = \varphi_n(t_1', \ldots, t_m')$$

stetig, so ist auch die mittelbare Funktion

$$u = f\big(\varphi_1(t_1, \ldots, t_m), \ \ldots, \ \varphi_n(t_1, \ldots, t_m)\big) = f\big(\varphi_1(P), \ldots, \varphi_n(P)\big)$$

im Punkt P' stetig.

In der Tat existiert zu gegebenem $\varepsilon > 0$ zunächst eine Zahl $\delta > 0$ derart, daß aus (3) die Beziehung (2) folgt (da f stetig ist). Da die Funktionen $\varphi_1, \ldots, \varphi_n$ stetig sind, gibt es zu diesem δ eine Zahl $\eta > 0$ derart, daß aus

$$|t_1 - t_1'| < \eta, \ldots, |t_m - t_m'| < \eta \tag{5}$$

die Ungleichungen

$$\begin{aligned} |x_1 - x_1'| &= |\varphi_1(t_1, \ldots, t_m) - \varphi_1(t_1', \ldots, t_m')| < \delta, \ldots, \\ |x_n - x_n'| &= |\varphi_n(t_1, \ldots, t_m) - \varphi_n(t_1', \ldots, t_m')| < \delta \end{aligned}$$

folgen. Dann ergibt sich aber aus (5) auch

$$\begin{aligned}&|f(x_1,\ldots,x_n)-f(x_1',\ldots,x_n')|\\&=\left|f\big(\varphi_1(t_1,\ldots,t_m),\ldots,\varphi_n(t_1,\ldots,t_m)\big)-f\big(\varphi_1(t_1',\ldots,t_m'),\ldots,\varphi_n(t_1',\ldots,t_m')\big)\right|<\varepsilon;\end{aligned}$$

womit unsere Behauptung bewiesen ist.

171. In einem Bereich stetige Funktionen. Die Sätze von Bolzano-Cauchy. Man sagt, eine Funktion $f(x_1,\ldots,x_n)$ sei *auf einer Punktmenge $\mathcal{M}$ des n-dimensionalen Raumes stetig,* wenn sie in jedem einzelnen Punkt dieser Menge, welcher Häufungspunkt dieser Menge ist, stetig ist. Von jetzt ab werden wir uns auf den Fall beschränken, daß die Menge $\mathcal{M}$ ein offener oder abgeschlossener (zusammenhängender) Bereich (Nr. 163) ist, ähnlich wie wir stetige Funktionen einer einzigen Veränderlichen in einem Intervall betrachtet haben.

Wir wollen jetzt einige Eigenschaften von Funktionen mehrerer Veränderlicher untersuchen, die in einem Bereich des n-dimensionalen Raumes stetig sind. Sie sind den Eigenschaften der Funktionen einer einzigen Veränderlichen, die in einem Intervall stetig sind (Kap. II, § 5), völlig analog.

Bei der Darlegung beschränken wir uns der Kürze halber auf den Fall zweier unabhängiger Veränderlicher. Die Ergebnisse lassen sich ohne weiteres auf den allgemeinen Fall übertragen. Einige Hinweise in dieser Richtung werden nebenbei gegeben werden.

Wir formulieren zunächst einen Satz, der dem ersten Satz von BOLZANO-CAUCHY für Funktionen einer Veränderlichen (Nr. 80) analog ist.

Satz. *Die Funktion $f(x, y)$ sei in einem zusammenhängenden Bereich $\mathcal{D}$ definiert und stetig. Wenn die Funktion in zwei Punkten $M_0(x_0, y_0)$ und $M_1(x_1, y_1)$ dieses Bereichs Werte verschiedenen Vorzeichens annimmt, wenn also etwa*

$$f(x_0, y_0) < 0, \quad f(x_1, y_1) > 0$$

ist, so gibt es in diesem Bereich auch einen Punkt $M'(x', y')$, in dem die Funktion den Wert 0 annimmt: $f(x', y') = 0$.

Den Beweis führen wir auf den Fall einer Funktion einer unabhängigen Veränderlichen zurück.

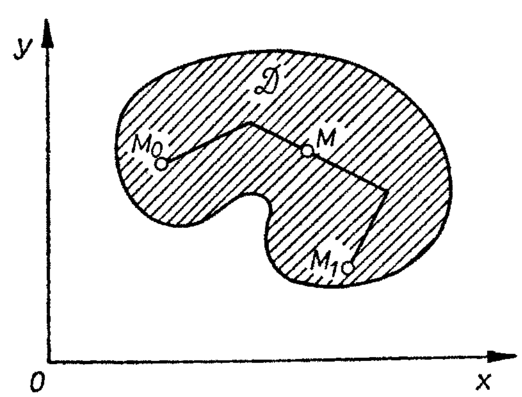

Abb. 97

Da der Bereich $\mathcal{D}$ zusammenhängend ist, kann man die Punkte M_0 und M_1 durch einen Streckenzug verbinden, der ganz in $\mathcal{D}$ liegt (Abb. 97). Durchläuft man den Streckenzug, so nimmt entweder in irgendeinem seiner Eckpunkte die Funktion den Wert 0 an (damit wäre der Satz bewiesen), oder aber dies ist nicht der Fall. Dann muß es aber eine Strecke des Streckenzugs geben, an deren Endpunkten die Funktion Werte verschiedenen Vorzeichens annimmt. Wir ändern die Bezeichnung der Punkte und nehmen an, M_0 und M_1 seien gerade die Endpunkte dieser Strecke. Ihre Gleichungen lauten (vgl. Nr. 161)

$$x = x_0 + t(x_1 - x_0), \quad y = y_0 + t(y_1 - y_0) \quad (0 \leqq t \leqq 1).$$

Wenn der Punkt $M(x, y)$ diese Strecke durchläuft, so geht unsere ursprüngliche Funktion $f(x, y)$ in eine mittelbare Funktion der einen Veränderlichen t über:

$$F(t) = f\big(x_0 + t(x_1 - x_0), y_0 + t(y_1 - y_0)\big) \quad (0 \leqq t \leqq 1).$$

Sie ist (nach dem Satz aus Nr. 170) stetig, da sowohl $f(x, y)$ als auch die linearen Funktionen von t, die an Stelle ihrer Argumente eingesetzt wurden, stetig sind. Für $F(t)$ gilt aber

$$F(0) = f(x_0, y_0) < 0, \quad F(1) = f(x_1, y_1) > 0.$$

Wenn wir auf die von der einzigen Variablen t abhängende Funktion $F(t)$ den in Nr. 80 bewiesenen Satz anwenden, so können wir schließen, daß $F(t') = 0$ ist für einen Wert t' zwischen 0 und 1. Nach Definition der Funktion $F(t)$ gilt somit

$$f\big(x_0 + t'(x_1 - x_0), y_0 + t'(y_1 - y_0)\big) = 0.$$

Der Punkt $M'(x', y')$, wobei $x' = x_0 + t'(x_1 - x_0)$ und $y' = y_0 + t'(y_1 - y_0)$ ist, leistet das Gewünschte.

Hieraus erhalten wir wie in Nr. 82 den zweiten Satz von Bolzano-Cauchy, den man übrigens auch direkt hätte erhalten können.

Der Übergang zum n-dimensionalen Raum ($n > 2$) bereitet keinerlei Schwierigkeiten; denn in einem n-dimensionalen zusammenhängenden Bereich können die Punkte ebenfalls durch einen „Streckenzug" verbunden werden, und das Problem reduziert sich auf die Untersuchung eines Gliedes dieses Streckenzugs, längs dessen die Funktion nur von einem einzigen Parameter abhängt.

172. Der Satz von Bolzano-Weierstrass. Für die weiteren Darlegungen benötigen wir die Verallgemeinerung des Satzes von Bolzano-Weierstrass (Nr. 41) auf den Fall einer Folge von Punkten im Raum beliebiger (endlicher) Dimension; wir beschränken uns auch hierbei wieder auf den ebenen Fall.

Satz. *Aus jeder beschränkten Punktfolge*

$$M_1(x_1, y_1), \quad M_2(x_2, y_2), \quad \ldots, \quad M_n(x_n, y_n), \ldots$$

kann man eine Teilfolge

$$M_{n_1}(x_{n_1}, y_{n_1}), \quad M_{n_2}(x_{n_2}, y_{n_2}), \quad \ldots, \quad M_{n_k}(x_{n_k}, y_{n_k}), \ldots$$
$$(n_1 < n_2 < \cdots < n_k < \cdots; n_k \to \infty)$$

auswählen, die gegen einen Grenzwert konvergiert. Dieser ist natürlich ein Häufungspunkt der Folge.

Erster Beweis. Dabei übertragen wir die in Nr. 41 im Eindimensionalen durchgeführten Überlegungen auf unseren zweidimensionalen Fall. Da die gegebene Punktfolge beschränkt ist, gibt es ein (endliches) Rechteck $[a, b; c, d]$, in dem sie vollständig erhalten ist. Wir halbieren sowohl das Intervall $[a, b]$ der x-Werte als auch das

Intervall $[c, d]$ der y-Werte:

$$\left[a, \frac{a+b}{2}\right], \quad \left[\frac{a+b}{2}, b\right] \quad \text{und} \quad \left[c, \frac{c+d}{2}\right], \quad \left[\frac{c+d}{2}, d\right].$$

Wenn wir jede Hälfte des ersten Intervalls mit jeder Hälfte des zweiten kombinieren, erhalten wir die vier Rechtecke

$$\text{(I)} \left[a, \frac{a+b}{2}; c, \frac{c+d}{2}\right], \qquad \text{(II)} \left[\frac{a+b}{2}, b; c, \frac{c+d}{2}\right],$$

$$\text{(III)} \left[a, \frac{a+b}{2}; \frac{c+d}{2}, d\right], \qquad \text{(IV)} \left[\frac{a+b}{2}, b; \frac{c+d}{2}, d\right],$$

in die das ursprüngliche Rechteck $[a, b; c, d]$ zerlegt wurde (Abb. 98).

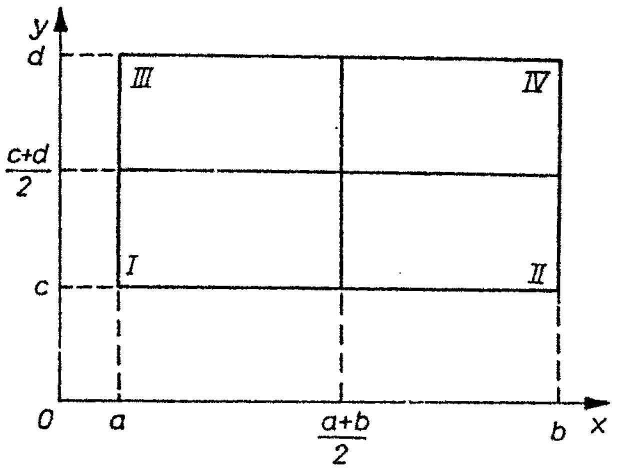

Abb. 98

Mindestens in einem dieser Teile muß eine unendliche Menge von Punkten der gegebenen Folge erhalten sein; denn sonst wären im ganzen Rechteck nur endlich viele dieser Punkte enthalten. Es sei $[a_1, b_1; c_1, d_1]$ dasjenige der Rechtecke (I), (II), (III), (IV) (oder eines davon, wenn es mehrere geben sollte), in dem eine unendliche Menge von Punkten unserer Folge enthalten ist.

Dieses so erhaltene Rechteck teilen wir wieder durch Seitenhalbierung in vier kleinere Rechtecke und greifen von diesen dasjenige (oder eines von mehreren) heraus, das unendlich viele Punkte der gegebenen Folge enthält; wir bezeichnen es mit $[a_2, b_2; c_2, d_2]$.

Diesen Prozeß der sukzessiven Zerlegung der Rechtecke setzen wir unbegrenzt fort. Nach k Schritten haben wir das Rechteck $[a_k, b_k; c_k, d_k]$ erhalten, in welchem eine unendliche Menge von Punkten M_n enthalten ist. Die Abmessungen dieses Rechtecks,

$$b_k - a_k = \frac{b-a}{2^k}, \qquad d_k - c_k = \frac{d-c}{2^k},$$

streben für $k \to \infty$ gegen 0.

Wir wenden jetzt den Intervallschachtelungssatz aus Nr. 38 einzeln auf die Folge der Intervalle $[a_k, b_k]$ der x-Werte und auf die Folge der Intervalle $[c_k, d_k]$ der y-Werte an. Dann folgt, daß die Endpunkte a_k und b_k sowie die c_k und d_k gegen gemeinsame Grenzwerte streben:

$$\lim a_k = \lim b_k = \bar{x} \quad \text{und} \quad \lim c_k = \lim d_k = \bar{y}. \tag{6}$$

Man kann somit sagen, daß sich die Folge der Rechtecke $\{[a_k, b_k; c_k, d_k]\}$ auf den Punkt $\overline{M}(\bar{x}, \bar{y})$ „zusammenzieht".

Jetzt nehmen wir als M_{n_1} einen beliebigen Punkt unserer Folge, der in dem Rechteck $[a_1, b_1; c_1, d_1]$ liegt, ferner sukzessive die Punkte $M_{n_2}, M_{n_3}, \ldots$, wobei wir im allgemeinen Fall als $M_{n_k}(x_{n_k}, y_{n_k})$ einen beliebigen Punkt der Folge wählen, der auf die schon früher ausgewählten Punkte folgt und im k-ten Rechteck $[a_k, b_k; c_k, d_k]$ enthalten ist. Das kann man tatsächlich tun, da jedes der Rechtecke unendlich viele Punkte M_n enthält. Wegen $a_k \leqq x_{n_k} \leqq b_k$ und $c_k \leqq x_{n_k} \leqq d_k$ ist nach (6)

$$\lim_{k\to\infty} x_{n_k} = \bar{x}, \quad \lim_{k\to\infty} y_{n_k} = \bar{y},$$

so daß die ausgewählte Teilfolge $\{M_{n_k}\}$ gegen den Punkt $\overline{M}(\bar{x}, \bar{y})$ als Grenzwert konvergiert (vgl. Nr. 166).

Zweiter Beweis. Einfacher kommt man zum Ziel, wenn man den Satz benutzt, der in Nr. 41 für den eindimensionalen Fall bewiesen wurde. Sind die Punkte unserer Folge in dem endlichen Rechteck $[a, b; c, d]$ enthalten, so ist

$$a \leqq x_n \leqq b, \quad c \leqq y_n \leqq d \quad (\text{für } n = 1, 2, 3, \ldots).$$

Wenden wir den Satz aus Nr. 41 zuerst auf die Folge $\{x_n\}$ an, so erhalten wir eine Teilfolge $\{x_{n_k}\}$, die gegen einen Grenzwert $\bar{x}$ konvergiert. Somit haben für die Teilfolge der Punkte

$$(x_{n_1}, y_{n_1}), (x_{n_2}, y_{n_2}), \ldots, (x_{n_k}, y_{n_k}), \ldots$$

die ersten Koordinaten schon einen Grenzwert. Nun wenden wir denselben Satz auf die Folge $\{y_{n_k}\}$ der zweiten Koordinaten an und können eine Teilfolge $\{y_{n_{k_m}}\}$ auswählen, die gegen einen Grenzwert $\bar{y}$ strebt. Dann strebt offenbar die Teilfolge der Punkte

$$(x_{n_{k_1}}, y_{n_{k_1}}), (x_{n_{k_2}}, y_{n_{k_2}}), \ldots, (x_{n_{k_m}}, y_{n_{k_m}}), \ldots$$

gegen den Grenzwert $(\bar{x}, \bar{y})$.

Beide Überlegungen lassen sich leicht auf den n-dimensionalen Raum ($n > 2$) übertragen. Bei der ersten ändert sich nur die Anzahl der Teile, in die der n-dimensionale Quader zerlegt wird, wenn jedes der ihn definierenden Intervalle halbiert wird; im allgemeinen Fall gibt es n dieser Intervalle, also 2^n Teile.

173. Die Weierstraßschen Sätze. Mit Hilfe des soeben bewiesenen Satzes kann man zunächst für Funktionen zweier Veränderlicher den *Zwischenwertsatz* (ersten Satz von Weierstrass) beweisen.

Satz. *Ist eine Funktion $f(x, y)$ in einem beschränkten abgeschlossenen (nicht notwendig zusammenhängenden) Bereich $\mathfrak{D}$ definiert und stetig, so ist sie beschränkt, d. h., ihre sämtlichen Werte liegen zwischen zwei endlichen Grenzen:*

$$m \leqq f(x, y) \leqq M.$$

Wir beweisen diesen Satz indirekt, durch Überlegungen, die denen aus Nr. 84 völlig analog sind. Die Funktion $f(x, y)$ sei also nicht beschränkt, wenn der Punkt (x, y) in $\mathfrak{D}$ variiert. Dann ließe sich zu jedem n in $\mathfrak{D}$ ein Punkt $M_n(x_n, y_n)$ finden derart, daß

$$|f(x_n, y_n)| > n \tag{7}$$

ist. Nach dem Satz aus Nr. 172 könnte man aus der beschränkten Folge $\{M_n\}$ eine Teilfolge $\{M_{n_k}\}$ auswählen, die gegen einen Grenzpunkt $\overline{M}(\bar{x}, \bar{y})$ konvergiert.

Dieser Punkt $\overline{M}$ müßte im Bereich $\mathfrak{D}$ liegen. Denn sonst wären alle Punkte M_{n_k} von $\overline{M}$ verschieden, und $\overline{M}$ wäre ein Häufungspunkt des Bereichs $\mathfrak{D}$, der nicht zu $\mathfrak{D}$ gehörte, entgegen unserer Voraussetzung über die Abgeschlossenheit von $\mathfrak{D}$ (vgl. Nr. 163).

Da $f(x, y)$ in M stetig ist, gilt

$$f(M_{n_k}) = f(x_{n_k}, y_{n_k}) \to f(\overline{M}) = f(\bar{x}, \bar{y}),$$

im Widerspruch zu (7).

Der zweite Weierstraßsche Satz läßt sich mit Hilfe des vorhergehenden Satzes genau so formulieren und beweisen wie in Nr. 85.

Bemerkung. Beide Weierstraßschen Sätze lassen sich ohne wesentliche Änderung der Überlegungen auch auf den Fall übertragen, daß die Funktion in einer beliebigen beschränkten abgeschlossenen Menge $\mathcal{M}$ (die also kein Bereich zu sein braucht) stetig ist.

Ebenso wie bei Funktionen einer einzigen Veränderlichen heißt die Differenz zwischen der oberen und der unteren Grenze der Werte einer in einer Menge $\mathcal{M}$ definierten und beschränkten Funktion $f(x, y)$ die *Schwankung der Funktion* in (auf) dieser Menge. Ist $\mathcal{M}$ beschränkt und abgeschlossen (insbesondere ein beschränkter abgeschlossener Bereich) und ist f in $\mathcal{M}$ stetig, so ist die Schwankung einfach die Differenz zwischen dem größten und dem kleinsten Wert.

174. Die gleichmäßige Stetigkeit. Wir wissen, daß die Stetigkeit einer Funktion $f(x, y)$ in einem bestimmten Punkt (x_0, y_0) einer Menge $\mathcal{M}$, in der sie gegeben ist, in der „ε-δ-Sprache“ folgendermaßen definiert werden kann: Zu jedem $\varepsilon > 0$ gibt es ein $\delta > 0$ derart, daß die Ungleichung

$$|f(x, y) - f(x_0, y_0)| < \varepsilon$$

für jeden Punkt (x, y) aus $\mathcal{M}$ erfüllt ist, für den

$$|x - x_0| < \delta, \quad |y - y_0| < \delta$$

gilt. Nun sei $f(x, y)$ auf der ganzen Menge $\mathcal{M}$ stetig; dann entsteht die Frage, ob man zu gegebenem $\varepsilon > 0$ ein $\delta > 0$ finden kann, welches für *alle* Punkte (x_0, y_0) aus $\mathcal{M}$ das Gewünschte leistet. Ist das für jedes ε möglich, so sagt man, $f(x, y)$ sei in $\mathcal{M}$ *gleichmäßig stetig.*

Satz von Cantor. *Ist eine Funktion $f(x, y)$ in einem beschränkten abgeschlossenen Bereich $\mathfrak{D}$ stetig, so ist sie dort gleichmäßig stetig.*

Beweis (indirekt). Für ein $\varepsilon > 0$ möge kein $\delta > 0$ existieren, das gleichzeitig für alle Punkte (x_0, y_0) von $\mathfrak{D}$ geeignet ist. Wir betrachten nun eine monoton fallend gegen 0 strebende Folge positiver Zahlen δ_n. Da keine von ihnen gleichzeitig für alle Punkte (x_0, y_0) aus $\mathfrak{D}$ das Gewünschte leisten kann, gäbe es zu jedem δ_n in $\mathfrak{D}$ einen wohlbestimmten Punkt (x_n, y_n), für den δ_n nicht das Verlangte leistet. Das würde aber bedeuten, daß es in $\mathfrak{D}$ einen Punkt (x_n', y_n') gäbe, für welchen zwar

$$|x_n' - x_n| < \delta_n \quad \text{und} \quad |y_n' - y_n| < \delta_n,$$

aber

$$|f(x_n', y_n') - f(x_n, y_n)| \geqq \varepsilon \tag{8}$$

wäre. Da aber die Folge $\{(x_n, y_n)\}$ beschränkt ist, könnte man nach dem Satz von Bolzano-Weierstrass eine Teilfolge $\{(x_{n_k}, y_{n_k})\}$ herausgreifen derart, daß $x_{n_k} \to \bar{x}$,

$y_{n_k} \to \bar{y}$ gelten würde, wobei der Limespunkt $(\bar{x}, \bar{y})$ auf Grund der Abgeschlossenheit von $\mathfrak{D}$ zu $\mathfrak{D}$ gehören würde.

Da ferner $|x'_{n_k} - x_{n_k}| < \delta_{n_k}$ und $|y'_{n_k} - y_{n_k}| < \delta_{n_k}$ wäre und mit wachsendem k natürlich $n_k \to \infty$ und $\delta_{n_k} \to 0$ gilt, würde

$$x'_{n_k} - x_{n_k} \to 0 \quad \text{und} \quad y'_{n_k} - y_{n_k} \to 0,$$

also auch $x'_{n_k} \to \bar{x}$ und $y'_{n_k} \to \bar{y}$ folgen. Da $f(x, y)$ in $(\bar{x}, \bar{y})$ stetig ist und dieser Punkt zu $\mathfrak{D}$ gehört, würde dann sowohl $f(x_{n_k}, y_{n_k}) \to f(\bar{x}, \bar{y})$ als auch $f(x'_{n_k}, y'_{n_k}) \to f(\bar{x}, \bar{y})$ gelten und somit

$$f(x_{n_k}, y_{n_k}) - f(x'_{n_k}, y'_{n_k}) \to 0,$$

was aber der Ungleichung (8) widerspricht. Damit ist der Satz bewiesen.

Zur Formulierung einer Folgerung aus diesem Satz benötigen wir den Begriff des *Durchmessers einer Menge*: Darunter versteht man die obere Grenze der Abstände je zweier beliebiger Punkte der Menge.

Folgerung. *Ist $f(x, y)$ in einem beschränkten abgeschlossenen Bereich $\mathfrak{D}$ stetig, so gibt es zu gegebenem $\varepsilon > 0$ ein $\delta > 0$ derart, daß bei jeder Zerlegung von $\mathfrak{D}$ in abgeschlossene Teilbereiche $\mathfrak{D}_1$ $\ldots, \mathfrak{D}_n$, die höchstens Randpunkte gemeinsam haben und deren Durchmesser kleiner als δ sind, die Schwankung von $f(x, y)$ in jedem einzelnen dieser Teilbereiche kleiner als ε ist.*

Es genügt, für δ die in der Definition der gleichmäßigen Stetigkeit vorkommende Zahl δ zu nehmen. Ist der Durchmesser des Teilbereichs $\mathfrak{D}_i$ kleiner als δ, so ist die Entfernung je zweier seiner Punkte (x, y) und (x_0, y_0) kleiner als δ,

$$\sqrt{(x - x_0)^2 + (y - y_0)^2} < \delta.$$

Hieraus folgt erst recht $|x - x_0| < \delta$ und $|y - y_0| < \delta$, also die Ungleichung $|f(x, y) - f(x_0, y_0)| < \varepsilon$. Wählt man also (x, y) und (x_0, y_0) so, daß $f(x_0, y_0)$ bzw. $f(x, y)$ der größte bzw. der kleinste Funktionswert in $\mathfrak{D}_i$ ist, so hat man die gewünschte Aussage.

Man sieht leicht, daß sich dieser Satz ohne Änderung (ähnlich dem Satz von WEIERSTRASS) auf den Fall einer Funktion übertragen läßt, die auf einer beliebigen beschränkten abgeschlossenen Menge $\mathcal{M}$ stetig ist.

175. Der Überdeckungssatz. Der in Nr. 88 bewiesene, so äußerst nützliche Satz läßt sich ebenfalls auf den mehrdimensionalen Fall verallgemeinern. Wir betrachten den zweidimensionalen Fall.

Es sei Σ ein System offener Bereiche σ; ist jeder Punkt einer Menge $\mathcal{M}$ in einem dieser Bereiche σ enthalten, so sagt man, *Σ überdecke $\mathcal{M}$.*

Überdeckungssatz (BOREL). *Wird eine beschränkte abgeschlossene Menge $\mathcal{M}$ von Punkten der Ebene durch ein unendliches System $\Sigma = \{\sigma\}$ offener Bereiche überdeckt, so gibt es ein endliches Teilsystem*

$$\Sigma^* = \{\sigma_1, \sigma_2, \ldots, \sigma_n\},$$

das ebenfalls schon $\mathcal{M}$ überdeckt.

seiBeweis (indirekt). $\mathcal{M}$ möge also nicht durch endlich viele σ aus Σ überdeckbar n. Da $\mathcal{M}$ beschränkt ist, ist $\mathcal{M}$ in einem Rechteck $[a, b;\ c, d]$ enthalten. Durch Halbierung der beiden Intervalle $[a, b]$ und $[c, d]$ zerlegen wir dieses Rechteck, ebenso wie beim Beweis des Satzes von BOLZANO-WEIERSTRASS (Nr. 172), in vier Rechtecke. Dabei wird auch $\mathcal{M}$ in Teilmengen zerlegt, die in diesen Teilrechtecken enthalten sind. Dabei kann es durchaus weniger als vier solcher Teilmengen geben, beispielsweise, wenn ein Teilrechteck keinen Punkt von $\mathcal{M}$ enthält. Mindestens eine dieser Teilmengen, etwa $\mathcal{M}_1$, wäre jedoch nicht durch endlich viele σ überdeckbar, denn sonst wäre ganz $\mathcal{M}$ durch endlich viele σ überdeckbar. Das Rechteck, das $\mathcal{M}_1$ enthält, bezeichnen wir mit $[a_1, b_1; c_1, d_1]$.

Wir zerlegen es wieder in vier Teilrechtecke, von denen mindestens eines, das wir mit $[a_2, b_2; c_2, d_2]$ bezeichnen, eine Teilmenge $\mathcal{M}_2$ von $\mathcal{M}$ enthalten müßte, die nicht durch enlich viele σ überdeckbar ist.

Wenn wir diesen Prozeß unbeschränkt fortsetzen, kommen wir mit dem k-ten Schritt zu einem Rechteck $[a_k, b_k; c_k, d_k]$, das eine Teilmenge $\mathcal{M}_k$ von $\mathcal{M}$ enthält, welche nicht durch endlich viele σ überdeckbar ist. Wie in Nr. 172 könnten wir schließen, daß sich das Rechteck $[a_k, b_k; c_k, d_k]$ auf einen Punkt $(\bar{x}, \bar{y})$ zusammenzieht, für den

$$\lim a_k = \lim b_k = \bar{x}, \quad \lim c_k = \lim d_k = \bar{y}$$

gilt. Der Punkt $\bar{M}(\bar{x}, \bar{y})$ gehört aber zu $\mathcal{M}$; denn für jede Umgebung $(\bar{x} - \delta, \bar{x} + \delta; \bar{y} - \delta, \bar{y} + \delta)$ von $\bar{M}$ ist für hinreichend große k

$$\bar{x} - \delta < a_k < b_k < \bar{x} + \delta, \quad \bar{y} - \delta < c_k < d_k < \bar{y} + \delta,$$

so daß in dieser Umgebung die Teilmenge $\mathcal{M}_k$ von $\mathcal{M}$, die auf Grund unserer Konstruktion tatsächlich aus unendlich vielen Punkten besteht, enthalten wäre. Daher wäre $\bar{M}$ Häufungspunkt von $\mathcal{M}$, auf Grund der Abgeschlossenheit von $\mathcal{M}$ also ein Punkt von $\mathcal{M}$.

Dann wäre aber $\bar{M}$ in einem der Bereiche σ, etwa in σ_0, enthalten. Da σ_0 offen ist, enthält σ_0 auch eine ganze Umgebung

$$(\bar{x} - \delta, \bar{x} + \delta; \bar{y} - \delta, \bar{y} + \delta)$$

von $\bar{M}$. Wie oben läßt sich leicht zeigen, daß für hinreichend großes k das Rechteck $[a_k, b_k; c_k, d_k]$ in dieser Umgebung läge, also auch die darin enthaltene Teilmenge $\mathcal{M}_k$ von $\mathcal{M}$.

Somit könnte $\mathcal{M}_k$ von einem einzigen Bereich σ_0 überdeckt werden, während $\mathcal{M}_k$ so gewählt sein sollte, daß $\mathcal{M}_k$ nicht durch endlich viele σ überdeckt werden konnte. Mit diesem Widerspruch ist der Satz bewiesen.

In den Anwendungen des Überdeckungssatzes, die der Leser in Nr. 176 und in anderen Teilen dieses Werkes findet, die ist Menge $\mathcal{M}$ meist ein abgeschlossener Bereich. Gelegentlich wenden wir ihn aber auch auf andere abgeschlossene Mengen, beispielsweise stetige Kurven, an.

176. Weitere Beweise der grundlegenden Sätze.

1. *Der erste Zwischenwertsatz.* Die Funktion $f(x, y)$ sei in einem beschränkten abgeschlossenen Bereich $\mathfrak{D}$ stetig. Dann kann jeder Punkt (x', y') dieses Bereichs in eine Umgebung σ' eingebettet werden derart, daß innerhalb ihrer Grenzen bei vorgegebenem $\varepsilon > 0$

$$|f(x, y) - f(x', y')| < \varepsilon$$

oder

$$f(x', y') - \varepsilon < f(x, y) < f(x', y') + \varepsilon$$

gilt. Somit ist $f(x, y)$ im Bereich σ' beschränkt.

Wenden wir den Borelschen Überdeckungssatz auf das System $\Sigma = \{\sigma'\}$ dieser Umgebungen an, so können wir aus Σ endlich viele Umgebungen $\sigma_1, \sigma_2, \ldots, \sigma_n$ auswählen, die zusammen den ganzen Bereich $\mathfrak{D}$ überdecken. Ist

$$m_i \leqq f(x, y) \leqq M_i \quad \text{in} \quad \sigma_i \quad (i = 1, 2, \ldots, n),$$

so gilt in $\mathfrak{D}$, wenn wir die kleinste der Zahlen m_i mit m und die größte der Zahlen M_i mit M bezeichnen,

$$m \leqq f(x, y) \leqq M,$$

was zu beweisen war.

2. *Der Satz von* Cantor. Es sei $\varepsilon > 0$ beliebig gegeben. Zu jedem Punkt (x', y') existiert eine Umgebung

$$\sigma' = (x' - \delta', x' + \delta'; y' - \delta', y' + \delta')$$

derart, daß zu jedem ihrer Punkte (x, y) (aus $\mathfrak{D}$) die Ungleichung

$$|f(x, y) - f(x', y')| < \frac{\varepsilon}{2}$$

gilt. Ist (x_0, y_0) ein anderer Punkt dieser Art, so daß also auch

$$|f(x', y') - f(x_0, y_0)| < \frac{\varepsilon}{2}$$

gilt, so ist

$$|f(x, y) - f(x_0, y_0)| < \varepsilon. \tag{9}$$

Wir ersetzen jedes Rechteck σ' durch ein Rechteck mit demselben Mittelpunkt und einem Viertel der Fläche,

$$\bar{\sigma}' = \left(x' - \frac{\delta'}{2},\ x' + \frac{\delta'}{2};\ y' - \frac{\delta'}{2},\ y' + \frac{\delta^2}{2}\right).$$

Das System $\bar{\Sigma} = \{\bar{\sigma}'\}$ dieser offenen Rechtecke überdeckt den Bereich $\mathfrak{D}$. Nach dem Überdeckungssatz können wir aus $\bar{\Sigma}$ ein endliches System von Rechtecken

$$\bar{\sigma}_i = \left(x_i - \frac{\delta_i}{2},\ x_i + \frac{\delta_i}{2};\ y_i - \frac{\delta_i}{2},\ y_i + \frac{\delta_i}{2}\right)$$

mit derselben Eigenschaft auswählen. Schließlich bezeichnen wir die kleinste aller Zahlen $\frac{\delta_i}{2}$ mit δ.

Es seien (x, y) und (x_0, y_0) zwei beliebige Punkte des Bereichs $\mathfrak{D}$, für welche

$$|x - x_0| < \delta, \qquad |y - y_0| < \delta \tag{10}$$

gilt. Der Punkt (x_0, y_0) gehört einer der Umgebungen $\bar{\sigma}_i$ an, etwa der Umgebung

$$\bar{\sigma}_{i_0} = \left(x_{i_0} - \frac{\delta_{i_0}}{2},\ x_{i_0} + \frac{\delta_{i_0}}{2};\ y_{i_0} - \frac{\delta_{i_0}}{2},\ y_{i_0} + \frac{\delta_{i_0}}{2}\right),$$

so daß

$$|x_0 - x_{i_0}| < \frac{\delta_{i_0}}{2}, \quad |y_0 - y_{i_0}| < \frac{\delta_{i_0}}{2}$$

gilt. Aus (10) folgt wegen $\delta \leqq \frac{\delta_{i_0}}{2}$ offenbar $|x - x_0| < \frac{\delta_{i_0}}{2}$, $|y - y_0| < \frac{\delta_{i_0}}{2}$ und demnach $|x - x_{i_0}| < \delta_{i_0}$, $|y - y_{i_0}| < \delta_{i_0}$. Die Punkte (x, y) und (x_0, y_0) liegen offenbar beide in einer der ursprünglich definierten Umgebungen

$$(x_{i_0} - \delta_{i_0}, x_{i_0} + \delta_{i_0}; y_{i_0} - \delta_{i_0}, y_{i_0} + \delta_{i_0});$$

es gilt also die Ungleichung (9).

Somit kann man zu $\varepsilon > 0$ unabhängig von der Lage des Punktes (x_0, y_0) ein solches $\delta > 0$ auswählen; damit ist bewiesen, daß die Funktion $f(x, y)$ gleichmäßig stetig ist.

§ 3. Ableitungen und Differentiale von Funktionen mehrerer Veränderlicher

177. Partielle Ableitungen und partielle Differentiale. Zur Vereinfachung der Schreibweise und der Darlegung beschränken wir uns hier auf Funktionen dreier Veränderlicher; alles hierfür Gesagte gilt aber auch für Funktionen beliebig vieler Veränderlicher.

Es sei in einem beliebigen (offenen) Bereich $\mathcal{D}$ eine Funktion $u = f(x, y, z)$ gegeben; wir wählen in diesem Bereich den Punkt $M_0(x_0, y_0, z_0)$. Erteilen wir y und z die konstanten Werte y_0 und z_0 und variieren x, so hängt u (in einer Umgebung von x_0) nur von der einen Veränderlichen x ab. Man kann jetzt die Aufgabe stellen, die Ableitung von u im Punkt $x = x_0$ zu berechnen. Addiert man zu diesem Wert x_0 den Zuwachs Δx, so erhält die Funktion den Zuwachs

$$\Delta_x u = \Delta_x f(x_0, y_0, z_0) = f(x_0 + \Delta x, y_0, z_0) - f(x_0, y_0, z_0),$$

den man *partiellen Zuwachs* (bezüglich x) nennen könnte, weil er durch die Änderung des Wertes einer einzigen Veränderlichen hervorgerufen wird. Nach Definition der Ableitung ist diese der Grenzwert (natürlich mit der Einschränkung „sofern er existiert“)

$$\lim_{\Delta x \to 0} \frac{\Delta_x u}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x, y_0, z_0) - f(x_0, y_0, z_0)}{\Delta x}.$$

Diese Ableitung wird *partielle Ableitung der Funktion* $f(x, y, z)$ *nach* x *im Punkt* (x_0, y_0, z_0) genannt.

Wie wir sehen, sind in dieser Definition nicht alle Koordinaten gleichwertig; y_0 und z_0 sind fest, während x variabel ist und gegen x_0 strebt.

Die partielle Ableitung bezeichnet man mit einem der Symbole[1])

$$\frac{\partial u}{\partial x}, \quad \frac{\partial f(x_0, y_0, z_0)}{\partial x}; \quad u_x, \quad f_x(x_0, y_0, z_0); \quad \mathrm{D}_x u, \quad \mathrm{D}_x f(x_0, y_0, z_0).$$

[1]) Der Vorschlag, an Stelle von d dieses runde ∂ zur Bezeichnung der partiellen Ableitung zu benutzen, stammt von Carl Gustav Jacob Jacobi (1804—1851, deutscher Mathematiker).

Der Buchstabe x, der in dieser Beziehung als Index auftritt, weist darauf hin, nach welcher Veränderlichen die Ableitung gebildet werden soll; er hat nichts damit zu tun, in welchem Punkt die Ableitung berechnet wird.[1])

Analog betrachtet man den Grenzwert

$$\lim_{\Delta y \to 0} \frac{\Delta_y u}{\Delta y} = \lim_{\Delta y \to 0} \frac{f(x_0, y_0 + \Delta y, z_0) - f(x_0, y_0, z_0)}{\Delta y},$$

wenn x und z als konstant angesehen werden und y variabel ist. Diesen Grenzwert nennt man die *partielle Ableitung der Funktion* $f(x, y, z)$ *nach* y *im Punkt* (x_0, y_0, z_0); man bezeichnet diese Ableitung mit

$$\frac{\partial u}{\partial y}, \quad \frac{\partial f(x_0, y_0, z_0)}{\partial y}; \quad u_y, \quad f_y(x_0, y_0, z_0); \quad \mathrm{D}_y u, \quad \mathrm{D}_y f(x_0, y_0, z_0).$$

Entsprechend wird die *partielle Ableitung von* $f(x, y, z)$ *nach* z *im Punkt* (x_0, y_0, z_0) definiert.

Die Berechnung der partiellen Ableitungen bietet im Grunde nichts Neues gegenüber der Berechnung der gewöhnlichen Ableitung.

Beispiele.

1. Es sei $u = x^y$ $(x > 0)$; die partiellen Ableitungen dieser Funktion sind

$$\frac{\partial u}{\partial x} = y \cdot x^{y-1}, \quad \frac{\partial u}{\partial y} = x^y \cdot \ln x.$$

Die erste von ihnen berechnet man wie die Ableitung einer Potenz von x (für $y = \text{const}$), die zweite wie die Ableitung der Exponentialfunktion von y (für $x = \text{const}$).

2. Ist $u = \arctan \frac{x}{y}$, so ist

$$\frac{\partial u}{\partial x} = \frac{y}{x^2 + y^2}, \quad \frac{\partial u}{\partial y} = -\frac{x}{x^2 + y^2}.$$

3. Für $u = \frac{x}{x^2 + y^2 + z^2}$ erhalten wir

$$\frac{\partial u}{\partial x} = \frac{y^2 + z^2 - x^2}{(x^2 + y^2 + z^2)^2}, \quad \frac{\partial u}{\partial y} = \frac{-2xy}{(x^2 + y^2 + z^2)^2}, \quad \frac{\partial u}{\partial z} = \frac{-2xz}{(x^2 + y^2 + z^2)^2}.$$

Es sei $z = y \cdot f(x^2 - y^2)$, wo $f(u)$ eine willkürliche Funktion ist (die eine Ableitung besitzt). Man zeige, daß für z immer die Beziehung

$$\frac{1}{x} \cdot \frac{\partial z}{\partial x} + \frac{1}{y} \cdot \frac{\partial z}{\partial y} = \frac{z}{y^2}$$

erfüllt ist.

Nach der Kettenregel (der Strich bezeichnet die Ableitung nach u) gilt nämlich

$$\frac{\partial z}{\partial x} = y \cdot f'(x^2 - y^2) \cdot 2x = 2xy \cdot f'(x^2 - y^2), \quad \frac{\partial z}{\partial y} = f(x^2 - y^2) - 2y^2 \cdot f'(x^2 - y^2);$$

---

[1]) Auch hier sind die Symbole $\frac{\partial f}{\partial x}$, f_x, $\mathrm{D}_x f$ als Ganzes aufzufassen, als Funktionssymbol für die partielle Ableitung nach x. Auf solche Hinweise werden wir aber künftig verzichten.

(Übrigens finden sich in der Literatur auch die Bezeichnungen f_1 oder $\mathrm{D}_1 f$, die andeuten sollen, daß die Funktion nach der ersten der Veränderlichen, von denen sie abhängt, abgeleitet werden soll. — *Anm. d. Red.*)

hieraus folgt

$$\frac{1}{x}\cdot\frac{\partial z}{\partial x}+\frac{1}{y}\cdot\frac{\partial z}{\partial y}=2y\cdot f'(x^2-y^2)+\frac{1}{y}\cdot f(x^2-y^2)-2y\cdot f'(x^2-y^2)=\frac{z}{y^2}.$$

5. Die Seite a eines Dreiecks kann durch die beiden anderen Seiten b und c und den eingeschlossenen Winkel α folgendermaßen ausgedrückt werden:

$$a=\sqrt{b^2+c^2-2bc\cdot\cos\alpha}.$$

Dann ist $\dfrac{\partial a}{\partial b}=\dfrac{b-c\cdot\cos\alpha}{\sqrt{b^2+c^2-2bc\cdot\cos\alpha}}=\dfrac{b-c\cdot\cos\alpha}{a},\quad \dfrac{\partial a}{\partial\alpha}=\dfrac{bc\cdot\sin\alpha}{a}.$

6. Aus der Physik ist die Clapeyronsche Formel $pV=RT$ (mit $R=\text{const}$) bekannt; sie beschreibt den Zusammenhang zwischen dem Volumen V, dem Druck p und der absoluten Temperatur T eines Mols eines idealen Gases. Eine der Größen p, V, T kann also als Funktion der beiden anderen definiert werden. Sind p, V die unabhängigen Veränderlichen und ist T die Funktion, $T=\dfrac{pV}{R}$, so ist

$$\frac{\partial T}{\partial p}=\frac{V}{R},\quad \frac{\partial T}{\partial V}=\frac{p}{R}.$$

Sind p und T die unabhängigen Veränderlichen, so ist V eine Funktion dieser beiden, $V=\dfrac{RT}{p}$, und es gilt

$$\frac{\partial V}{\partial p}=-\frac{RT}{p^2},\quad \frac{\partial V}{\partial T}=\frac{R}{p}.$$

Schließlich seien V und T die unabhängigen Veränderlichen und p die Funktion: $p=\dfrac{RT}{V}$; dann ist

$$\frac{\partial p}{\partial V}=-\frac{RT}{V^2},\quad \frac{\partial p}{\partial T}=\frac{R}{V}.$$

Hieraus erhält man die in der Thermodynamik wichtige Beziehung

$$\frac{\partial p}{\partial V}\cdot\frac{\partial V}{\partial T}\cdot\frac{\partial T}{\partial p}=-\frac{RT}{V^2}\cdot\frac{R}{p}\cdot\frac{V}{R}=-\frac{RT}{pV}=-1.$$

Wir weisen nochmals darauf hin, daß die Jacobische Bezeichnung (mit dem runden ∂) als Funktionszeichen aufzufassen ist, nicht etwa als Quotient oder als Bruch. Die obige Beziehung zeigt mit hinreichender Deutlichkeit den wesentlichen Unterschied in der Art der Bezeichnung der gewöhnlichen und der partiellen Ableitungen: Würde es sich bei den Ableitungen auf der linken Seite um gewöhnliche handeln, so könnte man sie als Quotienten ein und derselben Differentiale auffassen; durch Kürzen erhielte man 1. Hier aber steht auf der rechten Seite die Zahl -1. Eine partielle Ableitung ist also kein Quotient von Differentialen!

Das Produkt der partiellen Ableitung $\dfrac{\partial u}{\partial x}$ mit einem beliebigen Zuwachs Δx nennt man *partielles Differential der Funktion u nach x*; man schreibt

$$\mathrm{d}_x u=\frac{\partial u}{\partial x}\cdot\Delta x.$$

Wenn man auch hier unter dem Differential $\mathrm{d}x$ der unabhängigen Veränderlichen x den Zuwachs Δx versteht, so kann man diese Formel in der Gestalt

$$\mathrm{d}_x u=\frac{\partial u}{\partial x}\cdot\mathrm{d}x$$

schreiben.

Analog ist

$$\mathrm{d}_y u = \frac{\partial u}{\partial y} \cdot \mathrm{d}y, \qquad \mathrm{d}_z u = \frac{\partial u}{\partial z} \cdot \mathrm{d}z.$$

Somit sehen wir, daß man die partiellen Ableitungen ebenfalls in Form von Brüchen darstellen könnte:

$$\frac{\mathrm{d}_x u}{\mathrm{d}x}, \quad \frac{\mathrm{d}_y u}{\mathrm{d}y}, \quad \frac{\mathrm{d}_z u}{\mathrm{d}z};$$

jedoch muß dann unbedingt jeweils angegeben werden, nach welcher Veränderlichen das Differential zu bilden ist.

178. Der vollständige (totale) Zuwachs einer Funktion. Erteilt man, von den Werten $x = x_0$, $y = y_0$, $z = z_0$ der unabhängigen Veränderlichen ausgehend, diesen die Zuwächse Δx, Δy, Δz, so erhält die Funktion $u = f(x, y, z)$ den Zuwachs

$$\Delta u = \Delta f(x_0, y_0, z_0) = f(x_0 + \Delta x, y_0 + \Delta y, z_0 + \Delta z) - f(x_0, y_0, z_0),$$

den man den *vollständigen Zuwachs* der Funktion nennt.

Im Fall einer Funktion $y = f(x)$ einer einzigen Veränderlichen gilt [unter der Voraussetzung, daß in x_0 eine (endliche) Ableitung $f'(x_0)$ existiert] für diesen Zuwachs der Funktion die Formel [vgl. Nr. 96, Formel (2)]

$$\Delta y = \Delta f(x_0) = f'(x_0) \cdot \Delta x + \alpha \cdot \Delta x,$$

wobei α von Δx abhängt und für gegen 0 strebendes Δx gegen 0 strebt.

Wir wollen nun eine analoge Formel für den Zuwachs einer Funktion $u = f(x, y, z)$ herleiten:

$$\begin{aligned} \Delta u &= \Delta f(x_0, y_0, z_0) \\ &= f_x(x_0, y_0, z_0) \cdot \Delta x + f_y(x_0, y_0, z_0) \cdot \Delta y + f_z(x_0, y_0, z_0) \cdot \Delta z \\ &\quad + \alpha \cdot \Delta x + \beta \cdot \Delta y + \gamma \cdot \Delta z, \end{aligned} \tag{1}$$

wobei α, β, γ von Δx, Δy, Δz abhängen und gegen 0 streben, sobald alle diese Zuwächse gegen 0 streben. Damit (1) richtig ist, müssen wir aber der Funktion stärkere Einschränkungen auferlegen.

Satz. *Existieren die partiellen Ableitungen $f_x(x, y, z)$, $f_y(x, y, z)$, $f_z(x, y, z)$ nicht nur im Punkt (x_0, y_0, z_0), sondern auch in einer Umgebung dieses Punktes und sind sie überdies in (x_0, y_0, z_0) als Funktionen von x, y, z stetig, so gilt die Formel* (1).

Zum Beweis schreiben wir den vollständigen Zuwachs Δu in der Gestalt

$$\begin{aligned} \Delta u = {} & [f(x_0 + \Delta x, y_0 + \Delta y, z_0 + \Delta z) - f(x_0, y_0 + \Delta y, z_0 + \Delta z)] \\ & + [f(x_0, y_0 + \Delta y, z_0 + \Delta z) - f(x_0, y_0, z_0 + \Delta z)] \\ & + [f(x_0, y_0, z_0 + \Delta z) - f(x_0, y_0, z_0)]. \end{aligned}$$

Jede dieser Differenzen ist der partielle Zuwachs der Funktion in einer Veränderlichen. Da wir die Existenz der partiellen Ableitungen in einer Umgebung des Punktes (x_0, y_0, z_0) vorausgesetzt haben, können wir für hinreichend kleine Δx, Δy, Δz auf diese

Differenzen einzeln den Mittelwertsatz (Nr. 12) anwenden,[1]) und wir erhalten

$$\Delta u = f_x(x_0 + \theta \Delta x, y_0 + \Delta y, z + \Delta z) \cdot \Delta x$$
$$+ f_y(x_0, y_0 + \theta_1 \Delta y, z_0 + \Delta z) \cdot \Delta y + f_z(x_0, y_0, z_0 + \theta_2 \Delta z) \cdot \Delta z.$$

Dabei sind $\theta, \theta_1, \theta_2$ Zahlen zwischen 0 und 1. Setzt man jetzt

$$f_x(x_0 + \theta \Delta x, y_0 + \Delta y, z_0 + \Delta z) = f_x(x_0, y_0, z_0) + \alpha,$$
$$f_y(x_0, y_0 + \theta_1 \Delta y, z_0 + \Delta z) = f_y(x_0, y_0, z_0) + \beta,$$
$$f_z(x_0, y_0, z_0 + \theta_2 \Delta z) = f_z(x_0, y_0, z_0) + \gamma,$$

so erhält man für Δu einen Ausdruck der gewünschten Gestalt (1). Für $\Delta x \to 0$, $\Delta y \to 0$, $\Delta z \to 0$ streben die Argumente auf den linken Seiten dieser Gleichungen gegen x_0, y_0, z_0 (eben weil $\theta, \theta_1, \theta_2$ zwischen 0 und 1 liegen). Da die partiellen Ableitungen in diesem Bereich als stetig vorausgesetzt wurden, streben sie selbst gegen die Ableitungen auf der rechten Seite, die Größen α, β, γ also gegen 0. Damit ist alles bewiesen.

Dieser Satz erlaubt übrigens den Schluß, daß aus der Existenz und der Stetigkeit der partiellen Ableitungen in einem gegebenen Punkt die Stetigkeit der Funktion in diesem Punkt folgt. (Dagegen ergibt sich — vgl. S. 339 — aus der bloßen Existenz der partiellen Ableitungen noch nicht die Stetigkeit der Funktion.) Für $\Delta x \to 0$, $\Delta y \to 0$, $\Delta z \to 0$ gilt in der Tat offenbar auch $\Delta u \to 0$.

Um Formel (1) etwas kürzer schreiben zu können, führen wir den Abstand

$$\varrho = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}$$

der Punkte (x_0, y_0, z_0) und $(x_0 + \Delta x, y_0 + \Delta y, z_0 + \Delta z)$ ein. Dann können wir schreiben:

$$\alpha \cdot \Delta x + \beta \cdot \Delta y + \gamma \cdot \Delta z = \left(\alpha \cdot \frac{\Delta x}{\varrho} + \beta \cdot \frac{\Delta y}{\varrho} + \gamma \cdot \frac{\Delta z}{\varrho}\right) \cdot \varrho.$$

Bezeichnen wir noch den Ausdruck in der Klammer rechts mit ε, dann wird

$$\alpha \cdot \Delta x + \beta \cdot \Delta y + \gamma \cdot \Delta z = \varepsilon \cdot \varrho,$$

wobei ε von $\Delta x, \Delta y, \Delta z$ abhängt und gegen 0 strebt, wenn diese Zuwächse gegen 0 streben oder, kürzer, wenn der Abstand ϱ gegen 0 strebt. Dann läßt sich (1) in folgender Gestalt schreiben:

$$\Delta u = \Delta f(x_0, y_0, z_0) = f_x(x_0, y_0, z_0) \cdot \Delta x + f_y(x_0, y_0, z_0) \cdot \Delta y$$
$$+ f_z(x_0, y_0, z_0) \cdot \Delta z + \varepsilon \cdot \varrho \tag{2}$$

mit $\varepsilon \to 0$ für $\varrho \to 0$. Die Größe $\varepsilon \cdot \varrho$ kann offenbar in der Form $o(\varrho)$ geschrieben werden (wenn man das in Nr. 60 eingeführte Landau-Symbol auf Funktionen mehrerer Veränderlicher erweitert).

Bei unseren Überlegungen hatten wir den Fall nicht ausgeschlossen, daß $\Delta x, \Delta y, \Delta z$

[1]) Man kann z. B. die erste Differenz als Zuwachs der Funktion $f(x, y_0 + \Delta y, z_0 + \Delta z)$ der einzigen Veränderlichen x zwischen dem Punkt $x = x_0$ und dem Punkt $x = x_0 + \Delta x$ ansehen. Die Ableitung dieser Funktion, also $f_x(x, y_0 + \Delta y, z_0 + \Delta z)$, existiert nach Voraussetzung für alle Werte von x im Intervall $[x_0, x_0 + \Delta y]$, so daß der Mittelwertsatz anwendbar ist. Ebenso schließt man für die übrigen Differenzen.

einzeln oder auch alle drei zugleich 0 sind. Als wir also von den Limesbeziehungen

$$\alpha \to 0, \quad \beta \to 0, \quad \gamma \to 0, \quad \varepsilon \to 0$$

für $\Delta x \to 0$, $\Delta y \to 0$, $\Delta z \to 0$ sprachen, meinten wir den erweiterten Limesbegriff, der nicht ausschließt, daß die Zuwächse selbst den Wert 0 annehmen können, während sie gegen 0 streben. Man vergleiche dazu die analoge Bemerkung in Nr. 96.

Beim Beweis des Satzes haben wir von der Funktion u, die von mehr als einer Veränderlichen abhängt, mehr gefordert als in dem entsprechenden Fall einer Funktion einer einzigen Veränderlichen. Um nun zu zeigen, daß ohne diese Forderungen die Formeln (1) bzw. (2) nicht richtig zu sein brauchen, betrachten wir zum Abschluß nachstehendes Beispiel (wobei wir uns der Einfachheit halber auf eine Funktion zweier unabhängiger Veränderlicher beschränken). Es sei

$$f(x, y) = \frac{x^2 y}{x^2 + y^2} \quad \text{für} \quad x^2 + y^2 > 0, \qquad f(0, 0) = 0.$$

Diese Funktion ist in der ganzen Ebene stetig; für $(0, 0)$ folgt das aus Nr. 167, Beispiel 5. Ferner existieren die partiellen Ableitungen nach x und y ebenfalls in der ganzen Ebene. Für $x^2 + y^2 > 0$ ist offenbar

$$f_x(x, y) = \frac{2xy^3}{(x^2 + y^2)^2}, \quad f_y(x, y) = \frac{x^2(x^2 - y^2)}{(x^2 + y^2)}.$$

Im Ursprung gilt $f_x(0, 0) = f_y(0, 0) = 0$. Das folgt unmittelbar aus der Definition der partiellen Ableitung, denn es ist $f(x, 0) = f(0, y) = 0$. Nun läßt sich leicht zeigen, daß die partiellen Ableitungen im Ursprung nicht mehr stetig sind; für die erste beispielsweise braucht man nur $y = x = \frac{1}{n} \to 0$ zu setzen.

Die Formeln (1) bzw. (2) gelten für unsere Funktion im Punkt $(0, 0)$ nicht mehr. Andernfalls müßte nämlich

$$\Delta f(0,0) = \frac{\Delta x^2 \cdot \Delta y}{\Delta x^2 + \Delta y^2} = \varepsilon \cdot \sqrt{\Delta x^2 + \Delta y^2}$$

sein, wobei ε für $\Delta x \to 0$ und $\Delta y \to 0$ verschwindet. Setzt man insbesondere $\Delta y = \Delta x > 0$, so ist

$$\frac{1}{2} \Delta x = \varepsilon \cdot \sqrt{2} \cdot \Delta x, \quad \text{also} \quad \varepsilon = \frac{1}{2\sqrt{2}},$$

und ε strebt für $\Delta x \to 0$ nicht gegen 0. Auch die Funktion

$$f(x, y) = \sqrt{|xy|}$$

weist in $(0, 0)$ eine analoge Singularität auf, was sich der Leser selbst überlegen möge.

179. Das vollständige Differential. Für eine Funktion $y = f(x)$ hatten wir untersucht, ob und wann ihr Zuwachs $\Delta y = \Delta f(x_0) = f(x_0 + \Delta x) - f(x_0)$ in der Gestalt (vgl. Nr. 103)

$$\Delta f(x_0) = A \cdot \Delta x + o(\Delta x) \quad (A = \text{const}) \tag{3}$$

darstellbar ist. Es zeigte sich (Nr. 104), daß dies genau dann möglich ist, wenn in $x = x_0$ die Ableitung $f'(x_0)$ existiert und endlich ist; dabei gilt für die in (3) vorkommende Größe A die Beziehung $A = f'(x_0)$. Den linearen Anteil des Funktionszuwachses, also

$$A \cdot \Delta x = f'(x_0) \cdot \Delta x = y' \cdot \Delta x,$$

nannten wir Differential dy.

Für eine in einem (etwa offenen) Bereich $\mathfrak{D}$ definierte Funktion $f(x, y, z)$ mehrerer, beispielsweise dreier, Veränderlicher kann man die entsprechende Frage stellen: Wann ist der Zuwachs

$$\Delta u = \Delta f(x_0, y_0, z_0) = f(x_0 + \Delta x, y_0 + \Delta y, z_0 + \Delta z) - f(x_0, y_0, z_0)$$

in der Form

$$\Delta f(x_0, y_0, z_0) = A \cdot \Delta x + B \cdot \Delta y + C \cdot \Delta z + o(\varrho) \tag{4}$$

darstellbar, wobei A, B, C Konstante sind und $\varrho = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}$ ist?

Wie in Nr. 103 läßt sich leicht zeigen: Wenn eine Darstellung (4) möglich ist, dann existieren in (x_0, y_0, z_0) die partiellen Ableitungen nach jeder Veränderlichen, und es ist

$$f_x(x_0, y_0, z_0) = A, \quad f_y(x_0, y_0, z_0) = B, \quad f_z(x_0, y_0, z_0) = C.$$

Setzen wir nämlich in (4) beispielsweise $\Delta y = \Delta z = 0$ und $\Delta x \neq 0$, so erhalten wir

$$\frac{f(x_0 + \Delta x, y_0, z_0) - f(x_0, y_0, z_0)}{\Delta x} = A + \frac{o(|\Delta x|)}{\Delta x},$$

woraus folgt, daß

$$f_x(x_0, y_0, z_0) = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x, y_0, z_0) - f(x_0, y_0, z_0)}{\Delta x} = A$$

existiert. Somit kann eine Darstellung (4) nur in der Gestalt

$$\begin{aligned} &\Delta f(x_0, y_0, z_0) \\ &= f_x(x_0, y_0, z_0) \cdot \Delta x + f_y(x_0, y_0, z_0) \cdot \Delta y + f_z(x_0, y_0, z_0) \cdot \Delta z + o(\varrho) \end{aligned} \tag{5}$$

vorliegen oder, kürzer, in der Gestalt

$$\Delta u = u_x \cdot \Delta x - u_y \cdot \Delta y + u_z \cdot \Delta z + o(\varrho). \tag{5*}$$

Während aber im Fall einer Funktion einer einzigen Veränderlichen die Existenz der Ableitung $y' = f'(x_0)$ für das Bestehen der Identität (3) auch hinreichend war, gewährleistet die bloße Existenz der partiellen Ableitungen

$$u_x = f_x(x_0, y_0, z_0), \quad u_y = f_y(x_0, y_0, z_0), \quad u_z = f_z(x_0, y_0, z_0)$$

jetzt das Vorliegen der Identität (4) nicht mehr. Das haben wir für Funktionen zweier Veränderlicher an den Beispielen in Nr. 178 erkannt. Dort waren auch hinreichende Bedingungen für das Bestehen der Beziehung (4) angegeben worden: Die partiellen Ableitungen sollten in einer Umgebung des Punktes (x_0, y_0, z_0) existieren und in diesem Punkt stetig sein. Nun kann man leicht zeigen, daß diese Bedingungen für die Gültigkeit der Formeln (5) oder (5*) keineswegs notwendig sind. Das folgt im Grunde schon daraus, daß sie für Funktionen einer Veränderlichen (die man nach Belieben auch als Spezialfall von Funktionen mehrerer Veränderlicher ansehen kann) nicht notwendig sind.

Gilt Formel (5) bzw. (5*), so sagt man, die Funktion $f(x, y, z)$ sei im Punkt (x_0, y_0, z_0) *total* (*vollständig*) *differenzierbar* und nennt in diesem Fall den Ausdruck

$$\begin{aligned} &u_x \cdot \Delta x + u \cdot \Delta y + u_z \cdot \Delta z \\ &= f_z(x_0, y_0, z_0) \cdot \Delta x + f_y(x_0, y_0, z_0) \cdot \Delta y + f_z(x_0, y_0, z_0) \cdot \Delta z, \end{aligned}$$

d. h. den linearen Anteil des Funktionszuwachses, das *totale* (*vollständige*) *Differential* der Funktion; man bezeichnet es mit $\mathrm{d}u$ oder mit $\mathrm{d}f(x_0, y_0, z_0)$.

Für Funktionen mehrerer Veränderlicher sind also die Aussagen „die Funktion besitzt in einem Punkt (partielle) Ableitungen nach allen Veränderlichen" und „die Funktion besitzt ein (totales) Differential" keineswegs äquivalent, wie es bei Funktionen einer einzigen Veränderlichen der Fall ist. Die Existenz eines totalen Differentials besagt etwas mehr. Allerdings werden wir meist Existenz und Stetigkeit der partiellen Ableitungen voraussetzen, und dann ist die (totale) Differenzierbarkeit, d. h. die Existenz des totalen Differentials, gewährleistet.

Unter den Differentialen $\mathrm{d}x$, $\mathrm{d}y$, $\mathrm{d}z$ der unabhängigen Veränderlichen versteht man beliebige Zuwächse Δx, Δy, Δz.[1]) Daher kann man schreiben:

$$\mathrm{d}f(x_0, y_0, z_0) = f_x(x_0, y_0, z_0) \cdot \mathrm{d}x + f_y(x_0, y_0, z_0) \cdot \mathrm{d}y + f_z(x_0, y_0, z_0) \cdot \mathrm{d}z$$

oder

$$\mathrm{d}u = u_x \cdot \mathrm{d}x + u_y \cdot \mathrm{d}y + u_z \cdot \mathrm{d}z.$$

Das totale Differential ist also gleich der Summe der partiellen Differentiale (Nr. 177).

180. Geometrische Deutung im Fall einer Funktion zweier Veränderlicher. Wir wollen nun, ähnlich wie in Nr. 91 und Nr. 104 bei der Ableitung und dem Differential einer Funktion einer einzigen Veränderlichen, die obigen Ausführungen geometrisch deuten; wir wenden uns daher dem Begriff der Tangente an eine Kurve $\mathcal{K}$ in einem Punkt M_0 auf $\mathcal{K}$ zu.

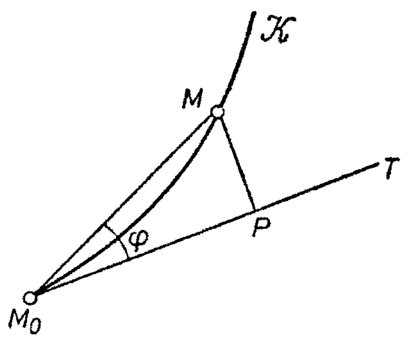

Abb. 99

Wir hatten die Tangente M_0T (Abb. 99) als Grenzlage der Sekante M_0M für gegen 0 strebenden Abstand $\overline{M_0M}$ definiert (Nr. 91).

Offenbar kann man diese Definition (gleichwertig) auch folgendermaßen formulieren:

Eine Gerade M_0T heißt *Tangente an eine Kurve* $\mathcal{K}$ im Kurvenpunkt M_0, wenn der Abstand $\overline{MP}$ eines veränderlichen Punktes M der Kurve $\mathcal{K}$ von der Geraden M_0T für gegen 0 strebenden Abstand $\overline{M_0M}$ von höherer Ordnung unendlich klein wird als $\overline{M_0M}$ (d. h., wenn der Quotient $\dfrac{\overline{MP}}{\overline{M_0M}}$ gegen 0 strebt).[2])

---

[1]) Identifiziert man das Differential der unabhängigen Veränderlichen x mit dem Differential von x als Funktion der unabhängigen Veränderlichen x, y, z, so kann man nach der allgemeinen Formel die Beziehung

$$\mathrm{d}x = x_x \cdot \Delta x + x_y \cdot \Delta y + x_z \cdot \Delta z = 1 \cdot \Delta x + 0 \cdot \Delta y + 0 \cdot \Delta z = \Delta x$$

aufstellen; dann ist die Identität $\mathrm{d}x = \Delta x$ bewiesen.

[2]) Das bedeutet nämlich, daß $\sin\varphi$, also der Winkel φ zwischen der Sekante M_0M und der Geraden M_0T (vgl. Abb. 99) gegen 0 strebt.

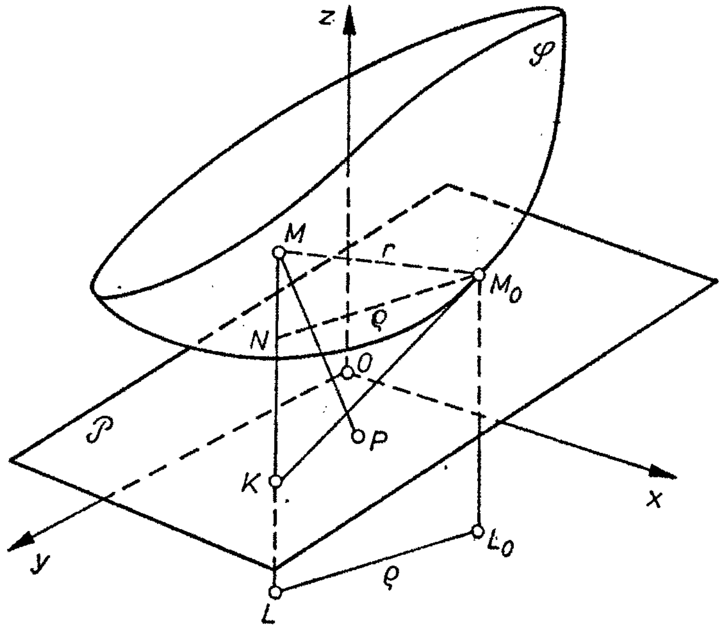

Abb. 100

Wir betrachten jetzt eine Fläche $\mathcal{S}$ und einen Punkt M_0 auf $\mathcal{S}$ (Abb. 100).

Ähnlich wie die Tangente an eine Kurve definieren wir jetzt die Tangentialebene: Eine Ebene M_0K heißt *Tangentialebene* an die Fläche $\mathcal{S}$ im Punkt M_0 von $\mathcal{S}$, wenn der Abstand $\overline{MP}$ eines veränderlichen Flächenpunktes M von dieser Ebene für gegen 0 strebenden Abstand $\overline{M_0M}$ von höherer Ordnung unendlich klein wird als $\overline{M_0M}$ (d. h., wenn der Quotient $\dfrac{\overline{MP}}{\overline{M_0M}}$ gegen 0 strebt).

Die Fläche sei in der Form $z = f(x, y)$ in rechtwinkligen Koordinaten gegeben (vgl. Nr. 160). Wir wählen darauf einen Punkt $M_0(x_0, y_0, z_0)$ [dort ist also $z_0 = f(x_0, y_0)$] und untersuchen, unter welchen Bedingungen eine durch M_0 gehende Ebene $\mathcal{P}$, deren Gleichung

$$Z - z_0 = A(X - x_0) + B(Y - y_0) \tag{6}$$

lautet, die Definition für die Tangentialebene genügt.

Wir ziehen ML parallel der z-Achse (vgl. Abb. 100) und fällen von M_0 auf ML das Lot $\overline{M_0N}$. Da sich $\overline{MK}$ von $\overline{MP}$ durch einen von 0 verschiedenen konstanten Faktor unterscheidet, können wir an Stelle von $\dfrac{\overline{MP}}{\overline{MM_0}}$ den Quotienten $\dfrac{\overline{MK}}{\overline{MM_0}}$ untersuchen. Nun wollen wir beweisen, daß wir dabei $r = \overline{MM_0}$ durch $\varrho = \overline{M_0N}$ ersetzen dürfen, ohne daß sich die Definition der Tangentialebene ändert.

Wenn für $M \to M_0$ der Quotient $\dfrac{\overline{MK}}{\varrho}$ gegen 0 strebt, dann gilt das um so mehr für $\dfrac{\overline{MK}}{r}$, denn es ist $r > \varrho$. Nehmen wir umgekehrt an, $\dfrac{\overline{MK}}{r}$ strebe gegen 0; wir beweisen, daß dann auch $\dfrac{\overline{MK}}{\varrho}$ gegen 0 strebt. Es genügt zu zeigen, daß für $M \to M_0$ der Quotient $\dfrac{r}{\varrho}$ beschränkt bleibt (weil $\dfrac{\overline{MK}}{\varrho} = \dfrac{\overline{MK}}{r} \cdot \dfrac{r}{\varrho}$ ist).

Bis auf das Vorzeichen ist $\overline{MK}$ gleich

$$z - Z = z - z_0 - A(x - x_0) - B(y - y_0)$$

oder, mit $x - x_0 = \Delta x$, $y - y_0 = \Delta y$, $z - z_0 = \Delta z = \Delta f(x_0, y_0)$, gleich

$$\Delta z - (A\Delta x + B\Delta y).$$

Nach unserer Annahme gilt wenigstens für hinreichend nahe bei M_0 gelegene Punkte M

$$|\Delta z - (A\Delta x + B\Delta y)| < \frac{1}{2} r = \frac{1}{2} \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2},$$

also

$$\frac{|\Delta z|}{\varrho} < |A| \cdot \frac{|\Delta x|}{\varrho} + |B| \cdot \frac{|\Delta y|}{\varrho} + \frac{1}{2} \sqrt{1 + \left(\frac{|\Delta z|}{\varrho}\right)^2},$$

um so mehr (Vergrößerung der rechten Seite)

$$\frac{|\Delta z|}{\varrho} < |A| + |B| + \frac{1}{2}\left(1 + \frac{|\Delta z|}{\varrho}\right).$$

Hieraus folgt $\dfrac{|\Delta z|}{\varrho} < 2(|A| + |B|) + 1$ und schließlich

$$\frac{r}{\varrho} = \sqrt{1 + \left(\frac{|\Delta z|}{\varrho}\right)^2} < 2(|A| + |B| + 1),$$

was zu beweisen war.

Somit ist eine Ebene (6) genau dann Tangentialebene an eine Fläche, wenn der Quotient

$$\frac{\Delta z - (A\Delta x + B\Delta y)}{\varrho}$$

mit ϱ gegen 0 strebt, d. h., wenn

$$\Delta z = \Delta f(x_0, y_0) = A \cdot \Delta x + B \cdot \Delta y + o(\varrho)$$

gilt [vgl. (4)].

Wir kommen also schließlich zu folgendem Ergebnis: *Eine Fläche $z = f(x, y)$ hat im Punkt $M_0(x_0, y_0, z_0)$, wobei $z_0 = f(x_0, y_0)$ ist, genau dann eine (nicht zur z-Achse parallele) Tangentialebene, wenn für $x = x_0$, $y = y_0$ die Funktion $f(x, y)$ ein totales Differential hat.*

Da unter dieser Bedingung die Koeffizienten A und B mit den partiellen Ableitungen $f_x(x_0, y_0)$ bzw. $f_y(x_0, y_0)$ übereinstimmen müssen, lautet die Gleichung der Tangentialebene

$$Z - z_0 = f_x(x_0, y_0) \cdot (X - x_0) + f_y(x_0, y_0) \cdot (Y - y_0).$$

Gewöhnlich läßt man den Index weg und schreibt

$$Z - z = f_x(x, y) \cdot (X - x) + f_y(x, y) \cdot (Y - y). \tag{7}$$

Man sieht leicht: *Schneidet man die Fläche und ihre Tangentialebene mit einer zur z-Achse parallelen Ebene durch M_0, so erhält man eine Kurve und eine Tangente an diese.*[1]

---

[1]) In Nr. 234 betrachten wir das allgemeinere Problem der Tangenten an beliebige Kurven, die auf einer Fläche durch einen gegebenen Punkt gehen.

Insbesondere ergeben sich beim Schnitt der Fläche mit den Ebenen $Y = y_0$ und $X = x_0$ Kurven, deren Richtungskoeffizienten gleich

$$f_x(x_0, y_0) \quad \text{bzw.} \quad f_y(x_0, y_0)$$

sind.[1])

In Abb. 101 stellen die Strecken $\overline{K_1M_1}$, $\overline{K_2M_2}$ und $\overline{KM}$ die partiellen Zuwächse und den totalen Zuwachs der Funktion dar, die Strecken $\overline{K_1N_1}$, $\overline{K_2N_2}$ und $\overline{KN}$ die partiellen Differentiale bzw. das totale Differential (vgl. Nr. 104 und Abb. 44).

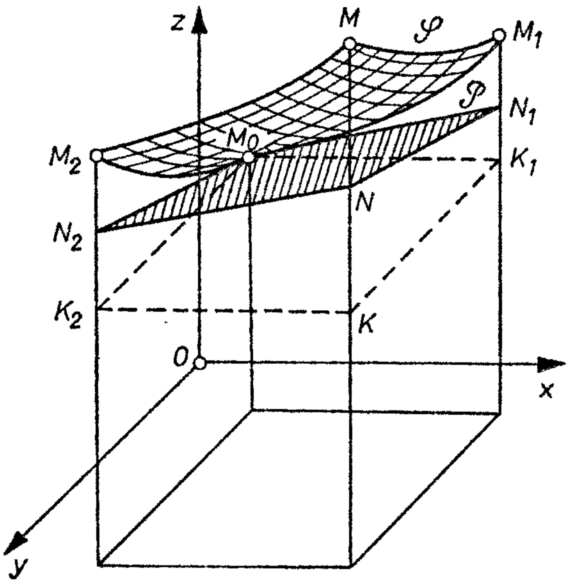

Abb. 101

181. Ableitungen mittelbarer Funktionen. Die Funktion

$$u = f(x, y, z)$$

sei in einem offenen Bereich $\mathfrak{D}$ definiert, und jede der Veränderlichen x, y, z sei ihrerseits eine Funktion einer in einem Intervall definierten Veränderlichen t:

$$x = \varphi(t), \quad y = \psi(t), \quad z = \chi(t).$$

Außerdem möge für die in Frage kommenden t der Punkt (x, y, z) nicht außerhalb $\mathfrak{D}$ liegen. Wir erhalten also eine mittelbare Funktion

$$u = f\bigl(\varphi(t), \psi(t), \chi(t)\bigr).$$

Die Funktion u möge stetige partielle Ableitungen u_x, u_y, u_z besitzen,[2]) und x_t, y_t, z_t mögen existieren. Dann kann man die Existenz des (totalen) Differentials der mittelbaren Funktion beweisen und es berechnen.

Erteilen wir t den Zuwachs Δt, so erhalten x, y, z die Zuwächse Δx, Δy, Δz und u den Zuwachs Δu.

Wenn wir den Zuwachs der Funktion u in der Form (1) darstellen (das können wir auf Grund der vorausgesetzten Stetigkeit der partiellen Ableitungen tun), so erhalten wir

$$\Delta u = u_x \cdot \Delta x + u_y \cdot \Delta y + u_z \cdot \Delta z + \alpha \cdot \Delta x + \beta \cdot \Delta y + \gamma \cdot \Delta z,$$

[1]) Man überlegt sich leicht, in bezug auf welches Koordinatensystem diese Richtungskoeffizienten berechnet werden.

[2]) Es würde genügen, die Existenz des totalen Differentials vorauszusetzen.

wobei α, β, γ für gegen 0 strebende $\Delta x, \Delta y, \Delta z$ selbst gegen 0 streben. Dividieren wir diese Gleichung durch Δt, so folgt

$$\frac{\Delta u}{\Delta t} = u_x \cdot \frac{\Delta x}{\Delta t} + u_y \cdot \frac{\Delta y}{\Delta t} + u_z \cdot \frac{\Delta z}{\Delta t} + \alpha \cdot \frac{\Delta x}{\Delta t} + \beta \cdot \frac{\Delta y}{\Delta t} + \gamma \cdot \frac{\Delta z}{\Delta t}.$$

Nun lassen wir Δt gegen 0 streben. Da x, y, z stetig sind (wir hatten sie sogar als differenzierbar vorausgesetzt), streben auch $\Delta x, \Delta y, \Delta z$ gegen 0. Daher streben auch α, β, γ gegen 0. In der Grenze erhalten wir die *Kettenregel*

$$u_t = u_x \cdot x_t + u_y \cdot y_t + u_z \cdot z_t. \tag{8}$$

Unter unseren Annahmen existiert also die Ableitung der mittelbaren Funktion. Benutzen wir die üblichen Bezeichnungen, so können wir (8) in folgender Gestalt schreiben:

$$\frac{\mathrm{d}u}{\mathrm{d}t} = \frac{\partial u}{\partial x} \cdot \frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial u}{\partial y} \cdot \frac{\mathrm{d}y}{\mathrm{d}t} + \frac{\partial u}{\partial z} \cdot \frac{\mathrm{d}z}{\mathrm{d}t}. \tag{9}$$

Nun sollen x, y, z nicht nur von einer Veränderlichen t, sondern von mehreren Veränderlichen abhängen; beispielsweise sei

$$x = \varphi(t, v), \quad y = \psi(t, v), \quad z = \chi(t, v).$$

Außer der Existenz und der Stetigkeit der partiellen Ableitungen von $f(x, y, z)$ (übrigens würde wieder die Voraussetzung der Existenz eines totalen Differentials von f genügen) setzen wir die Existenz der partiellen Ableitungen von x, y, z nach t und v voraus.

Wenn wir φ, ψ und χ in f einsetzen, so erhalten wir eine Funktion der beiden Veränderlichen t und v; es entsteht die Frage, ob die partiellen Ableitungen u_t und u_v existieren und wie man sie berechnet.

Dieser Fall unterscheidet sich nicht wesentlich von dem schon behandelten; denn bei der Berechnung einer partiellen Ableitung einer Funktion zweier Veränderlicher ist eine der Veränderlichen als fest anzusehen, so daß eine Funktion einer Veränderlichen übrigbleibt. Daher bleibt Formel (8) ungeändert, während die Kettenregel (9) jetzt folgendermaßen lautet:

$$\frac{\partial u}{\partial t} = \frac{\partial u}{\partial x} \cdot \frac{\partial x}{\partial t} + \frac{\partial u}{\partial y} \cdot \frac{\partial y}{\partial t} + \frac{\partial u}{\partial z} \cdot \frac{\partial z}{\partial t}. \tag{9*}$$

182. Beispiele.

1. Wir betrachten

$$u = x^y.$$

Setzt man $x = \varphi(t)$, $y = \psi(t)$ und bildet das totale Differential nach der eben bewiesenen Regel für mittelbare Funktionen, so erhält man die schon bekannte Formel von LEINBIZ und JOH. BERNOULLI:

$$u_t = y \cdot x^{y-1} \cdot x_t + x^y \cdot \ln x \cdot y_t.$$

Wir hatten sie in Nr. 99, Beispiel 23, unter anderen Bezeichnungen mit Hilfe eines Kunstgriffes erhalten.

2. Die Funktion $u = f(x, y, z)$ habe stetige partielle Ableitungen, und es sei

$$x = \eta - \zeta, \quad y = \zeta - \xi, \quad z = \xi - \eta.$$

Dann ist

$$\frac{\partial u}{\partial \xi} = -\frac{\partial u}{\partial y} + \frac{\partial u}{\partial z}, \quad \frac{\partial u}{\partial \eta} = \frac{\partial u}{\partial x} - \frac{\partial u}{\partial z}, \quad \frac{\partial u}{\partial \zeta} = -\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y}.$$

3. Setzt man unter denselben Voraussetzungen bezüglich f

$$y = y(x) \quad \text{und} \quad z = z(x),$$

behält also x als unabhängige Veränderliche bei, und nimmt man an, $y(x)$ und $z(x)$ seien nach x differenzierbar, so erhält man für u als mittelbare Funktion von x die Beziehung

$$\frac{du}{dx} = \frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} \cdot \frac{dy}{dx} + \frac{\partial u}{\partial z} \cdot \frac{dz}{dx}$$

oder

$$\frac{du}{dx} = f_x(x, y(x), z(x)) + f_y(x, y(x), z(x)) \cdot y'(x) + f_z(x, y(x), z(x)) \cdot z'(x).$$

Hier spielt x dieselbe Rolle wie t in Formel (8).

4. Sind x und y unabhängig und setzt man für z

$$z = z(x, y)$$

ein (wobei die partiellen Ableitungen z_x und z_y existieren sollen), so gelten für die mittelbare Funktion $u = f(x, y, z(x, y))$ die Beziehungen

$$\frac{\partial u}{\partial x} = f_x(x, y, z(x, y)) + f_z(x, y, z(x, y)) \cdot z_x(x, y),$$

$$\frac{\partial u}{\partial y} = f_y(x, y, z(x, y)) + f_z(x, y, z(x, y)) \cdot z_y(x, y).$$

5. Als weitere Beispiele für die Anwendung der Formel (9) betrachten wir das Problem, die Determinante

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & a_{1n} \\ a_{21} & a_{22} & a_{2n} \\ \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & a_{nn} \end{vmatrix}$$

partiell abzuleiten, wenn ihre Elemente a_{ik} ($i = 1, 2, \ldots, n$; $k = 1, 2, \ldots, n$) differenzierbare Funktionen eines Parameters t sind. Nach dem Entwicklungssatz ist

$$\Delta = A_{1k} \cdot a_{1k} + A_{2k} \cdot a_{2k} + \cdots + A_{ik} \cdot a_{ik} + \cdots + A_{nk} \cdot a_{nk},$$

wobei die algebraischen Komplemente $A_{1k}, \ldots, A_{nk}$ das Element a_{ik} nicht enthalten. Daher folgt

$$\frac{\partial \Delta}{\partial a_{ik}} = A_{ik}.$$

Somit ist nach Formel (9)

$$\frac{d\Delta}{dt} = \sum_{k=1}^{n} \sum_{i=1}^{n} \frac{\partial \Delta}{\partial a_{ik}} \frac{da_{ik}}{dt} = \sum_{k=1}^{n} \sum_{i=1}^{n} A_{ik} \cdot \frac{da_{ik}}{dt}.$$

Die Summe $\sum_{i=1}^{n} A_{ik} \frac{da_{ik}}{dt}$ liefert die Entwicklung einer Determinante, die sich von der gegebenen nur dadurch unterscheidet, daß die Elemente ihrer k-ten Spalte durch deren Ableitungen nach t ersetzt sind. Hieraus folgt die Regel: *Die Ableitung einer Determinante Δ ist gleich der Summe von n Determinanten, die man aus Δ dadurch erhält, daß man der Reihe nach die Elemente ihrer ersten, zweiten, ..., n-ten Spalte durch ihre Ableitungen ersetzt.*

Formel (8) stimmt mit der Formel $u_t = u_x \cdot x_t$ überein, wenn u nur von der einzigen Veränderlichen x abhängt. Wir weisen hier nochmals auf den Unterschied in den Voraussetzungen hin, unter denen diese Formeln hergeleitet wurden. Hängt u nur von einer Veränderlichen ab, so braucht nur die Existenz der Ableitung u_x vorausgesetzt zu werden; im Fall mehrerer Veränderlicher dagegen mußten wir neben der Existenz noch die Stetigkeit der Ableitungen u_x, u_y, ... voraussetzen. Die folgenden Beispiele zeigen, daß die bloße Existenz der Ableitungen für die Gültigkeit der Formel (8) im allgemeinen nicht ausreicht.

6. Es sei

$$f(x, y) = \frac{x^2 y}{x^2 + y^2} \quad \text{für} \quad x^2 + y^2 > 0, \qquad f(0, 0) = 0.$$

Diese Funktion hat, wie wir gesehen haben, in allen Punkten, einschließlich des Ursprungs, partielle Ableitungen, wobei $f_x(0, 0) = 0$, $f_y(0, 0) = 0$ ist. In diesem Punkt sind diese partiellen Ableitungen unstetig.

Führt man durch $x = t$ und $y = t$ eine neue Veränderliche t ein, so erhält man eine mittelbare Funktion von t. Nach (8) würde für die Ableitung dieser Funktion in $t = 0$

$$u_t = u_x \cdot x_t + u_y \cdot y_t = 0$$

gelten. Setzt man andererseits die Werte $x = t$ und $y = t$ in die Funktion ein, so erhält man

$$u = \frac{t^2 \cdot t}{t^2 + t^2} = \frac{1}{2} t,$$

durch Differentiation also $u_t = \frac{1}{2}$ für jedes t, auch für $t = 0$. Somit ist hier Formel (8) nicht anwendbar.

7. Es sei

$$f(x, y) = \frac{x^{5/3} \cdot y}{x^2 + y^2} \quad \text{für} \quad x^2 + y^2 > 0, \qquad f(0, 0) = 0.$$

Diese Funktion verhält sich im Ursprung (0, 0) ähnlich. Setzt man hier $x = y = t$, so erhält man die zusammengesetzte Funktion $u = \frac{1}{2} t^{2/3}$, die für $t = 0$ unendliche einseitige Ableitungen besitzt. Setzt man aber $x = t$ und

$$y = t^{4/3} \sin \frac{1}{t} \quad \text{für} \quad t \neq 0, \qquad y = 0 \quad \text{für} \quad t = 0,$$

so hat die durch

$$u = \begin{cases} \dfrac{t \cdot \sin \dfrac{1}{t}}{1 + t^{2/3} \cdot \sin^2 \dfrac{1}{t}} & \text{für} \quad t \neq 0, \\ 0 & \text{für} \quad t = 0 \end{cases}$$

definierte Funktion in $t = 0$ überhaupt keine Ableitungen.

183. Der Mittelwertsatz. Die Funktion $f(x, y, z)$ sei in einem abgeschlossenen Bereich $\mathcal{D}$ definiert und stetig und besitze im Innern von $\mathcal{D}$ (d. h. in jedem inneren Punkt) stetige partielle Ableitungen f_x, f_y, f_z. Wir betrachten zwei Punkte

$$M_0(x_0, y_0, z_0) \quad \text{und} \quad M_1(x_0 + \Delta x, y_0 + \Delta y, z_0 + \Delta z)$$

aus $\mathcal{D}$, die man durch eine geradlinige Strecke $\overline{M_0 M_1}$ verbinden kann, welche ganz in $\mathcal{D}$ liegt.

Dann gilt die Formel

$$\begin{aligned} \Delta f(x_0, y_0, z_0) &= f(x_0 + \Delta x, y_0 + \Delta y, z_0 + \Delta z) - f(x_0, y_0, z_0) \\ &= f_x(x_0 + \theta \Delta x, y_0 + \theta \Delta y, z_0 + \theta \Delta z) \cdot \Delta x \\ &\quad + f_y(\ldots) \cdot \Delta y + f_z(\ldots) \cdot \Delta z \qquad (0 < \theta < 1), \end{aligned} \tag{10}$$

die der bekannten Formel des Mittelwertsatzes für eine Funktion einer Veränderlichen [Nr. 112, Formel (2)] völlig analog ist.

Zum Beweis setzen wir in $f(x, y, z)$

$$x = x_0 + t \cdot \Delta x, \quad y = y_0 + t \cdot \Delta y, \quad z = z_0 + t \cdot \Delta z \tag{11}$$

(für $0 \leqq t \leqq 1$), d. h., wir betrachten unsere Funktion in den Punkten der Strecke $\overline{M_0 M_1}$. Die mittelbare Funktion

$$F(t) = f(x_0 + t \cdot \Delta x, y_0 + t \cdot \Delta y, z_0 + t \cdot \Delta z)$$

ist im ganzen t-Intervall [0, 1] stetig (Nr. 170) und hat in seinem Innern eine Ableitung, für die nach Formel (8)

$$\begin{aligned} F'(t) &= f_x(x_0 + t \cdot \Delta x, y_0 + t \cdot \Delta y, z_0 + t \cdot \Delta z) \cdot \Delta x \\ &\quad + f_y(\ldots) \cdot \Delta y + f_z(\ldots) \cdot \Delta z \end{aligned} \tag{*}$$

gilt; denn aus (11) folgt

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \Delta x, \quad \frac{\mathrm{d}y}{\mathrm{d}t} = \Delta y, \quad \frac{\mathrm{d}z}{\mathrm{d}t} = \Delta z.$$

Wenden wir auf die Funktion $F(t)$ im Intervall [0, 1] die Formel (2) aus Nr. 112 an, so ergibt sich

$$F(1) - F(0) = F'(\theta) \quad (0 < \theta < 1).$$

Gemäß der Definition von $F(t)$ ist

$$F(1) - F(0) = f(x_0 + \Delta x, y_0 + \Delta y, z_0 + \Delta z) - f(x_0, y_0, z_0);$$

setzt man an Stelle von $F'(\theta)$ den Wert (*) für $t = \theta$ ein, so kommt man zu Formel (10).

Als einfaches Beispiel für die Anwendung dieser Formel geben wir nachstehende Folgerung an:

Ist eine Funktion $f(x, y, z)$ in einem abgeschlossenen und zusammenhängenden Bereich $\mathfrak{D}$ stetig und sind im Innern von $\mathfrak{D}$ die partiellen Ableitungen f_x, f_y, f_z gleich 0, so ist diese Funktion im ganzen Bereich $\mathfrak{D}$ konstant:

$$f = \text{const.}$$

Es seien $M_0(x_0, y_0, z_0)$ und $M(x, y, z)$ zwei beliebige Punkte aus $\mathfrak{D}$. Da $\mathfrak{D}$ zusammenhängend ist, kann man diese Punkte durch einen Streckenzug verbinden, der ganz in $\mathfrak{D}$ verläuft. Ist $M_1(x_1, y_1, z_1)$ die auf M_0 folgende Ecke, so erhält man, wenn man in (11) $x_0 + \Delta x = x_1$, $y_0 + \Delta y = y_1$, $z_0 + \Delta z = z_1$ setzt,

$$f(x_1, y_1, z_1) = f(x_0, y_0, z_0);$$

geht man so von einer Ecke zur nächsten, so ergibt sich schließlich

$$f(x, y, z) = f(x_0, y_0, z_0),$$

was zu beweisen war.

184. Die Ableitung in einer bestimmten Richtung. Die partiellen Ableitungen einer Funktion $f(M) = f(x, y, z)$ nach x, y, z bringen die Geschwindigkeit zum Ausdruck, mit der sich die Funktion in der Richtung der Koordinatenachsen ändert. Beispielsweise ist f_x die Geschwindigkeit der Änderung der Funktion in der x-Richtung: Man nimmt an, ein Punkt bewege sich nur auf einer Parallelen zur x-Achse. Bei vielen physikalischen Problemen ist es nun interessant, die Geschwindigkeit der Änderung einer Funktion $f(M)$ auch in anderen Richtungen zu untersuchen. Das ist beispielsweise der Fall, wenn ein Temperaturfeld gegeben ist, d. h. die Temperatur $f(M)$ in jedem Punkt M eines zu untersuchenden Körpers. Die Gesetze der Verteilung und Ausbreitung der Wärme hängen wesentlich von der Geschwindigkeit des Absinkens (bzw. Ansteigens) der Temperatur in den verschiedenen Richtungen ab. Wir wollen nun den Begriff der Änderungsgeschwindigkeit bzw. der Ableitung einer Funktion in einer beliebig vorgegebenen Richtung präzisieren. Dabei werden wir wiederum Formel (9) anzuwenden haben.

Die Funktion $f(M)$ sei in einem (offenen) Bereich definiert. Wir betrachten einen beliebigen Punkt $M_0(x_0, y_0, z_0)$ dieses Bereichs und eine beliebig gerichtete Gerade (einen Strahl) l durch diesen Punkt (Abb. 102).

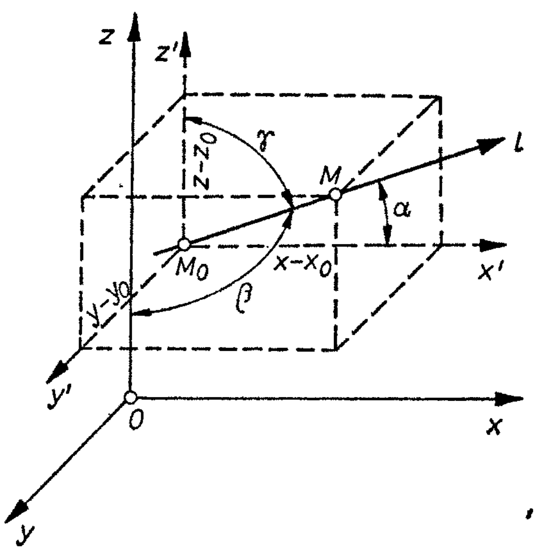

Abb. 102

Es sei $M(x, y, z)$ irgendein anderer Punkt dieses Strahls, $\overline{M_0M}$ die Länge der Strecke zwischen M_0 und M (positiv, wenn die Richtung von M_0M mit der Richtung des Strahles übereinstimmt; negativ, wenn sie ihr entgegengesetzt ist).

Nun nähere sich M unbegrenzt dem Punkt M_0. Den Limes

$$\lim_{M\to M_0} \frac{f(M) - f(M_0)}{\overline{M_0M}}$$

nennt man die *Ableitung von $f(M)$ in der Richtung l* (oder *längs des Strahles l*); man bezeichnet ihn mit

$$\frac{\partial f(M_0)}{\partial l} = \frac{\partial f(x_0, y_0, z_0)}{\partial l}.$$

Diese Ableitung charakterisiert die Geschwindigkeit der Änderung der Funktion im Punkt M_0 in der Richtung l. Insbesondere können, wie schon erwähnt, die gewöhnlichen partiellen Ableitungen $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, $\frac{\partial f}{\partial z}$ als solche „Richtungsableitungen" angesehen werden.

Wir nehmen nun an, $f(x, y, z)$ habe in dem zu untersuchenden Bereich stetige partielle Ableitungen. (Wieder würde es genügen, die Existenz des totalen Differentials vorauszusetzen.) Der Strahl l bilde mit den Achsen die Winkel α, β, γ. Wir zeigen nun, daß *unter unseren Annahmen die Ableitung in der Richtung l existiert und daß*

$$\frac{\partial f(x_0, y_0, z_0)}{\partial l} = \frac{\partial f(x_0, y_0, z_0)}{\partial x} \cdot \cos\alpha + \frac{\partial f(x_0, y_0, z_0)}{\partial y} \cdot \cos\beta + \frac{\partial f(x_0, y_0, z_0)}{\partial z} \cdot \cos\gamma \tag{12}$$

gilt.

Beweis. Setzt man $\overline{M_0M} = t$, so ist

$$x - x_0 = t \cdot \cos\alpha, \quad y - y_0 = t \cdot \cos\beta, \quad z - z_0 = t \cdot \cos\gamma.$$

Somit kann man längs des Strahles l die Koordinaten x, y, z als Funktion von t ansehen,

$$x = x_0 + t \cdot \cos\alpha, \quad y - y_0 + t \cdot \cos\beta, \quad z = z_0 + t \cdot \cos\gamma, \tag{13}$$

und die Funktion $f(M) = f(x, y, z)$ als mittelbare Funktion $\varphi(t)$ von t. Dem Punkt M_0 entspricht dabei der Wert $t = 0$.

Daher gilt

$$\frac{\partial f(M_0)}{\partial l} = \lim_{M \to M_0} \frac{f(M) - f(M_0)}{\overline{M_0M}} = \lim_{t \to 0} \frac{\varphi(t) - \varphi(0)}{t} = \varphi'(0),$$

wenn nur $\varphi'(0)$ existiert. Die Ableitung $\varphi'(t)$ existiert aber unter unseren Voraussetzungen und läßt sich nach Formel (9) folgendermaßen ausdrücken:

$$\varphi'(t) = \frac{\partial f}{\partial x} \cdot \frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial f}{\partial y} \cdot \frac{\mathrm{d}y}{\mathrm{d}t} + \frac{\partial f}{\partial z} \cdot \frac{\mathrm{d}z}{\mathrm{d}t}.$$

Auf Grund von (13) erhalten wir

$$\varphi'(t) = \frac{\partial f}{\partial x} \cdot \cos\alpha + \frac{\partial f}{\partial y} \cdot \cos\beta + \frac{\partial f}{\partial z} \cdot \cos\gamma,$$

und daraus folgt für $t = 0$ unsere Behauptung.

Wir legen uns nun die Frage vor, *in welcher Richtung eine Funktion in einem gegebenen Punkt am schnellsten wächst.* Natürlich hat diese Frage nur dann Sinn, wenn die Ableitungen

$$a = \frac{\partial f(x_0, y_0, z_0)}{\partial x}, \qquad b = \frac{\partial f(x_0, y_0, z_0)}{\partial y}, \qquad c = \frac{\partial f(x_0, y_0, z_0)}{\partial z} \tag{14}$$

nicht gleichzeitig 0 werden (mit anderen Worten, wenn nicht jede Richtungsableitung gleich 0 ist). Unter dieser Annahme formen wir die rechte Seite von (12) folgendermaßen um:

$$\begin{aligned} &a \cdot \cos\alpha + b \cdot \cos\beta + x \cdot \cos\gamma \\ &= \sqrt{a^2 + b^2 + c^2} \cdot \left(\frac{a}{\sqrt{\dots}} \cdot \cos\alpha + \frac{b}{\sqrt{\dots}} \cdot \cos\beta + \frac{c}{\sqrt{\dots}} \cdot \cos\gamma\right). \end{aligned}$$

Die Brüche in der Klammer lassen sich als Richtungskosinus einer Richtung g auffassen:

$$\frac{a}{\sqrt{\cdots}} = \cos\lambda, \quad \frac{b}{\sqrt{\cdots}} = \cos\mu, \quad \frac{c}{\sqrt{\cdots}} = \cos\nu;$$

dann ergibt der in (*) untersuchte Ausdruck

$$\sqrt{a^2+b^2+c^2}\cdot(\cos\lambda\cdot\cos\alpha+\cos\mu\cdot\cos\beta+\cos\nu\cdot\cos\gamma).$$

Bezeichnen wir mit (g, l) den Winkel zwischen den Richtungen g und l, so erhalten wir nach einer aus der analytischen Geometrie bekannten Formel

$$\frac{\partial f}{\partial l} = \sqrt{a^2+b^2+c^2}\cdot\cos(g, l). \tag{15}$$

Jetzt ist klar, daß diese Ableitung ihren größten Wert erreicht, wenn l mit g übereinstimmt, also für

$$\frac{\partial f}{\partial g} = \sqrt{a^2+b^2+c^2} = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2+\left(\frac{\partial f}{\partial y}\right)^2+\left(\frac{\partial f}{\partial z}\right)^2}$$

Der Vektor $\boldsymbol{g}$ mit den Projektionen (14) auf die Koordinatenachsen zeigt die Richtung an, in der die Funktion am schnellsten wächst; seine Länge $|\boldsymbol{g}|$ liefert den Wert der entsprechenden Ableitung. Dieser Vektor wird der *Gradient* der *Funktion* $f(M) = f(x, y, z)$ genannt.

Schreibt man (15) in der Gestalt

$$\frac{\partial f}{\partial l} = |\boldsymbol{g}|\cdot\cos(\boldsymbol{g}, l),$$

so ist leicht zu sehen, daß der Vektor, den man erhält, wenn man in der Richtung l die Strecke $\frac{\partial f}{\partial l}$ aufträgt, einfach die Projektion des Gradienten auf diese Richtung ist.

185. Die Invarianz des (ersten) Differentials. Die Funktion $u = f(x, y, z)$ habe stetige partielle Ableitungen u_x, u_y, u_z, und x, y, z seien ihrerseits Funktionen der neuen Veränderlichen t und v:

$$x = \varphi(t, v), \quad y = \psi(t, v), \quad z = \chi(t, v),$$

die wieder stetige partielle Ableitungen $x_t, x_v, y_t, y_v, z_t, z_v$ haben mögen. Dann existieren nach Nr. 181 nicht nur die Ableitungen der mittelbaren Funktion u nach t und v, sondern diese sind auch in t und v stetig, wie leicht aus (8) zu ersehen ist.

Wären x, y, z unabhängige Veränderliche, so würde, wie wir wissen, für das vollständige Differential von u

$$du = u_x\cdot dx + u_y\cdot dy + u_z\cdot dz$$

gelten. In unserem Fall hängt u über x, y, z von t und v ab. Daher gilt für das Differential bezüglich dieser Variablen

$$du = u_t\cdot dt + u_v\cdot dv.$$

Nun ist aber nach (8)

$$u_t = u_x\cdot x_t + u_y\cdot y_t + u_z\cdot z_t$$

und analog

$$u_v = u_x \cdot x_v + u_y \cdot y_v + u_z \cdot z_v.$$

Setzen wir diese Werte in den Ausdruck für du ein, so ergibt sich

$$du = (u_x \cdot x_t + u_y \cdot y_t + u_z \cdot z_t) \cdot dt + (u_x \cdot x_v + u_y \cdot y_v + u_z \cdot z_v) \cdot dv$$

oder aber

$$du = u_x \cdot (x_t \cdot dt + x_v \cdot dv) + u_y \cdot (y_t \cdot dt + y_v \cdot dv) + u_z \cdot (z_t \cdot dt + z_v \cdot dv).$$

Man sieht leicht, daß die Ausdrücke in den Klammern gerade die Differentiale der Funktionen x, y, z nach t und v sind; daher gilt

$$du = u_x \cdot dx + u_y \cdot dy + u_z \cdot dz.$$

Wir sind damit zu derselben Form des Differentials gekommen wie in dem Fall, daß x, y, z die unabhängigen Veränderlichen sind; allerdings haben die Symbole dx, dy, dz jetzt eine andere Bedeutung.

Für Funktionen mehrerer Veränderlicher ist also ebenfalls das Differential invariant, wie wir schon für Funktionen einer einzigen Veränderlichen erkannt hatten.[1]

Es kann vorkommen, daß jede der Größen x, y, z von anderen Veränderlichen abhängt, beispielsweise $x = \varphi(t)$, $y = \psi(t, w)$, $z = \chi(v, w)$. Dann können wir immer annehmen, es sei

$$x = \varphi_1(t, v, w), \quad y = \psi_1(t, v, w), \quad z = \chi_1(t, v, w),$$

und alle obigen Überlegungen bleiben gültig.

Folgerung. Sind x und y Funktionen einer einzigen Veränderlichen, so gilt bekanntlich (vgl. Nr. 105)

$$d(cx) = c \cdot dx, \quad d(x \pm y) = dx \pm dy, \quad d(xy) = y \cdot dx + x \cdot dy,$$

$$d\left(\frac{x}{y}\right) = \frac{y \cdot dx - x \cdot dy}{y^2}.$$

Diese Formeln bleiben auch dann gültig, wenn x und y Funktionen mehrerer Veränderlicher sind, wenn also

$$x = \varphi(t, v, \ldots), \quad y = \chi(t, v, \ldots)$$

gilt. Das zeigen wir für die letzte Formel.

Zunächst nehmen wir x und y als unabhängige Veränderliche; dann ist

$$d\left(\frac{x}{y}\right) = \frac{1}{y} \cdot dx - \frac{x}{y^2} \cdot dy = \frac{x \cdot dx - x \cdot dy}{y^2}.$$

Unter dieser Annahme hat also das Differential dieselbe Gestalt wie für Funktionen x und y einer einzigen Veränderlichen. Auf Grund der Invarianz des Differentials kann

[1] Übrigens gilt diese Behauptung auch dann, sobald vorausgesetzt wird, daß alle betrachteten Funktionen ein totales Differential haben. Um sich davon zu überzeugen, genügt es zu zeigen, daß eine Verkettung von Funktionen, die ein totales Differential besitzen, wieder eine ebensolche Funktion liefert.

man jetzt sagen, daß diese Formel auch dann gilt, wenn x und y Funktionen mehrerer Veränderlicher sind.

Diese Eigenschaft des totalen Differentials und die Folgerung lassen sich benutzen, um die Berechnung von Differentialen zu vereinfachen; beispielsweise ist

$$1.\ \mathrm{d}\arctan\frac{x}{y}=\frac{1}{1+\left(\frac{x}{y}\right)^2}\cdot\mathrm{d}\left(\frac{x}{y}\right)=\frac{y\cdot\mathrm{d}x-x\cdot\mathrm{d}y}{x^2+y^2},$$

$$\begin{aligned}2.\ \mathrm{d}\frac{x}{x^2+y^2+z^2}&=\frac{(x^2+y^2+z^2)\,\mathrm{d}x-x\cdot\mathrm{d}(x^2+y^2+z^2)}{(x^2+y^2+z^2)^2}\\&=\frac{(y^2+z^2-x^2)\,\mathrm{d}x-2xy\cdot\mathrm{d}y-2xz\cdot\mathrm{d}z}{(x^2+y^2+z^2)^2}.\end{aligned}$$

Da die Koeffizienten der Differentiale der unabhängigen Veränderlichen die entsprechenden partiellen Ableitungen sind, lassen sich deren Werte hieraus leicht berechnen. So erhalten wir für $u=\arctan\frac{x}{y}$ unmittelbar

$$\frac{\partial u}{\partial x}=\frac{y}{x^2+y^2},\qquad\frac{\partial u}{\partial y}=-\frac{x}{x^2+y^2}$$

und für $u=\dfrac{x}{x^2+y^2+z^2}$

$$\frac{\partial u}{\partial x}=\frac{y^2+z^2-x^2}{(x^2+y^2+z^2)^2},\qquad\frac{\partial u}{\partial y}=-\frac{2xy}{(x^2+y^2+z^2)^2},\qquad\frac{\partial u}{\partial z}=-\frac{2xz}{(x^2+y^2+z^2)^2}$$

(vgl. Nr. 177, Beispiel 2 und 3).

186. Anwendung des vollständigen Differentials in der Näherungsrechnung. Analog dem Differential einer Funktion einer einzigen Veränderlichen (Nr. 108) läßt sich das totale Differential einer Funktion mehrerer Veränderlicher mit Erfolg bei angenäherten Rechnungen zur Fehlerabschätzung verwenden. Es sei etwa eine Funktion $u=f(x,y)$ gegeben, deren Argumentwerte x und y wir nur mit einem Fehler Δx bzw. Δy angeben können. Dann ergibt sich auch der Wert von u aus diesen ungenauen Argumentwerten nur mit einem Fehler $\Delta u=f(x+\Delta x,y+\Delta y)-f(x,y)$. Es geht jetzt um die Abschätzung dieses Fehlers, wenn solche Abschätzungen für Δx und Δy bekannt sind.

Ersetzt man (näherungsweise) den Zuwachs der Funktion durch ihr Differential (was im Sinne der Näherungsrechung nur für hinreichend kleine Werte von Δx und Δy erlaubt ist), so erhält man

$$\Delta u=\frac{\partial u}{\partial x}\cdot\Delta x+\frac{\partial u}{\partial y}\cdot\Delta y. \tag{16}$$

Hier können sowohl die Fehlerglieder Δx und Δy als auch ihre Koeffizienten positiv oder negativ sein; ersetzt man alle beide durch ihre absoluten Beträge, so gelangt man zu der Ungleichung

$$|\Delta u|\leqq\left|\frac{\partial u}{\partial x}\right|\cdot|\Delta x|+\left|\frac{\partial u}{\partial y}\right|\cdot|\Delta y|.$$

Bezeichnet man mit δu, δx, δy die maximalen absoluten Fehler (bzw. die Grenzen für diese absoluten Fehler), so ergibt sich offenbar

$$\delta u=\left|\frac{\partial u}{\partial x}\right|\cdot\delta x+\left|\frac{\partial u}{\partial y}\right|\cdot\delta y. \tag{17}$$

Beispiele.

1. Zunächst leitet man mit Hilfe dieser Formeln eine für die Praxis der Näherungsrechnung wichtige Regel ab. Es sei $u = xy$ $(x > 0, y > 0)$, also $du = y\,dx + x\,dy$; ersetzt man die Differentiale durch endliche Zuwächse, so folgt $\Delta u = y\,\Delta x + x\,\Delta y$ [vgl. (16)] oder, wenn man zu den Fehlergrenzen übergeht, $\delta u = y \cdot \delta x + x \cdot \delta y$. Dividiert man durch $u = xy$, so erhält man

$$\frac{\delta u}{u} = \frac{\delta x}{x} + \frac{\delta y}{y}, \tag{18}$$

d. h. die folgende Regel: *Der maximale relative Fehler eines Produktes ist gleich der Summe der maximalen relativen Fehler der Faktoren.*

Wir hätten auch einfacher vorgehen, und zwar zuerst logarithmieren und dann differenzieren können: $u = xy$, also

$$\ln u = \ln x + \ln y, \qquad \frac{du}{u} = \frac{dx}{x} + \frac{dy}{y} \qquad \text{usw.}^{1)}$$

Ist $u = \frac{x}{y}$, so kommen wir nach derselben Methode zu

$$\ln u = \ln x - \ln y, \qquad \frac{du}{u} = \frac{dx}{x} - \frac{dy}{y}.$$

Gehen wir zu den absoluten Beträgen und den maximalen Fehlern über, so kommen wir wieder zur Formel (18). *Somit ist auch der maximale relative Fehler eines Quotienten gleich der Summe der maximalen relativen Fehler von Dividend bzw. Divisor.*

2. Die Fehlerrechnung wird vielfach in der Topographie angewandt, in der Hauptsache bei der Berechnung nicht direkt meßbarer Stücke eines Dreiecks aus meßbaren Stücken. Es folgen einige Beispiele aus diesem Gebiet.

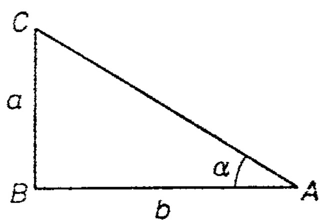

Abb. 103

In einem rechtwinkligen Dreieck ABC (Abb. 103) seien die Kathete $AB = b$ und der anliegende Winkel $BAC = \alpha$ gemessen; die zweite Kathete wird aus $a = b \cdot \tan\alpha$ berechnet. Wie wirken sich Meßfehler bei b und α auf a aus?

Durch Differenzieren ergibt sich

$$da = \tan\alpha \cdot db + \frac{b}{\cos^2\alpha} \cdot d\alpha,$$

also

$$\delta a = \tan\alpha \cdot \delta b + \frac{b}{\cos^2\alpha} \cdot \delta\alpha.$$

Es möge $b = 121{,}56\text{ m} \pm 0{,}05\text{ m}$, $\alpha = 25°\,21'\,40'' \pm 12''$ gemessen worden sein. Für a ergibt sich $a = 57{,}63$ m. Um nach unserer Formel δa zu bestimmen, setzen wir darin

$$\delta b = 0{,}05, \qquad \delta\alpha = \frac{12''}{206\,265''}$$

1) Wir machen hier darauf aufmerksam, daß wir das Differential von $\ln u$ so berechnet haben, als wäre u eine unabhängige Veränderliche, obwohl u in Wirklichkeit eine Funktion von x und y ist (vgl. Nr. 185). Das ist auch später zu beachten.

($\delta\alpha$ muß natürlich im Bogenmaß ausgedrückt werden; eine Bogeneinheit ist aber gleich $\frac{60'' \cdot 60' \cdot 360^\circ}{2\pi} = 206265''$). Wir erhalten

$$\tan\alpha \cdot \delta b = 0{,}0237, \qquad \frac{b}{\cos^2\alpha}\,\delta\alpha = 0{,}0087,$$

also gerundet $\delta a = 0{,}04$. Somit ist $a = 57{,}62\ \text{m} \pm 0{,}04\ \text{m}$.

3. Man bestimme den Fehler, der bei der Berechnung der Seite a eines schiefwinkligen Dreiecks ABC (Abb. 104) nach dem Kosinussatz

$$a = \sqrt{b^2 + c^2 - 2bc \cdot \cos\alpha}$$

auftritt. Benutzen wir die Resultate des Beispiels 5 aus Nr. 177, so können wir nach Formel (17) sofort

$$\delta a = \frac{b - c \cdot \cos\alpha}{a} \cdot \delta b + \frac{c - b \cdot \cos\alpha}{a} \cdot \delta c + \frac{bc \cdot \sin\alpha}{a} \cdot \delta\alpha$$

aufschreiben. Aus Abb. 104 finden wir sofort

$$b - c \cdot \cos\alpha = a \cdot \cos\gamma, \quad c - b \cdot \cos\alpha = a \cdot \cos\beta, \quad bc \cdot \sin\alpha = a \cdot h_a,$$

wobei h_a die von A auf a gefällte Höhe ist. Daher ist

$$\delta a = \cos\gamma \cdot \delta b + \cos\beta \cdot \delta c + h_a \cdot \delta\alpha;$$

hieraus ergibt sich der Einfluß der Fehler δb, δc, $\delta\alpha$ auf δa.

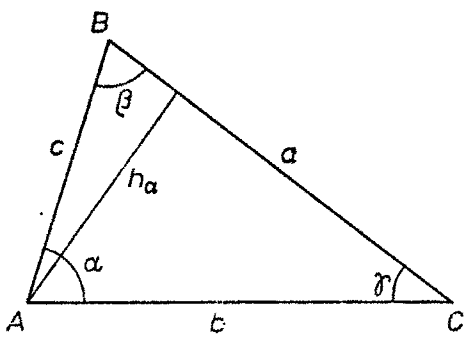

Abb. 104

187. Homogene Funktionen. Bekanntlich wird ein Polynom *homogen* genannt, wenn seine Glieder sämtlich denselben Grad haben. Beispielsweise ist

$$3x^2 - 2xy + 5y^2$$

ein homogenes Polynom zweiten Grades. Multipliziert man x und y mit einem Faktor t, so erhält das Polynom den Faktor t^2. Entsprechend verhalten sich beliebige homogene Polynome.

Jedoch können auch Funktionen komplizierterer Natur diese Eigenschaft besitzen. So multipliziert sich auch

$$x \cdot \frac{\sqrt{x^4 + y^4}}{x - y} \cdot \ln\frac{x}{y}$$

mit t^2, wenn man x und y durch tx und ty ersetzt; diese Funktion verhält sich also ähnlich wie ein homogenes Polynom zweiten Grades.

Naturgemäß nennt man diese Funktion *homogene Funktion zweiten Grades.*

Allgemein definiert man: Eine in einem Bereich $\mathfrak{D}$ definierte Funktion $f(x_1, \ldots, x_n)$ von n Argumenten wird *homogen vom Grad m* genannt, wenn sich bei der Multiplikation der einzelnen Argumente mit dem Faktor t die Funktion mit dem Faktor t^m

multipliziert, d. h., wenn die Gleichung

$$f(tx_1, \ldots, tx_n) = t^m \cdot f(x_1, \ldots, x_n) \tag{19}$$

identisch erfüllt ist.

Der Einfachheit halber nehmen wir an, $x_1, \ldots, x_n$ und t seien positiv. Der Bereich $\mathfrak{D}$, in dem wir die Funktion f untersuchen, möge mit jedem Punkt $M(x_1, \ldots, x_n)$ auch alle Punkte der Form $M_t(tx_1, \ldots, tx_n)$ für $0 < \alpha \leqq t \leqq \beta$ enthalten, d. h. jeden Strahl, der vom Nullpunkt ausgeht, durch M hindurchgeht und ganz in $\mathfrak{D}$ liegt.

Der Homogenitätsgrad m kann eine beliebige reelle Zahl sein; beispielsweise ist die Funktion

$$x^\pi \cdot \sin\frac{y}{x} + y^\pi \cdot \cos\frac{x}{y}$$

eine homogene Funktion von x und y vom Grad π.

Wir wollen jetzt einen *allgemeinen* Ausdruck für homogene Funktionen vom Grad m herleiten.

Es sei zunächst $f(x_1, \ldots, x_n)$ eine homogene Funktion nullten Grades; dann ist

$$f(tx_1, tx_2, \ldots, tx_n) = f(x_1, x_2, \ldots, x_n).$$

Setzen wir $t = \dfrac{1}{x_1}$, so folgt

$$f(x_1, x_2, \ldots, x_n) = f\left(1, \frac{x_2}{x_1}, \ldots, \frac{x_n}{x_1}\right).$$

Führen wir die Funktion

$$\varphi(u_1, \ldots, u_{n-1}) = f(1, u_1, \ldots, u_{n-1})$$

von $n - 1$ Argumenten ein, so zeigt sich, daß

$$f(x_1, x_2, \ldots, x_n) = \varphi\left(\frac{x_2}{x_1}, \ldots, \frac{x_n}{x_1}\right)$$

ist. Somit ist jede homogene Funktion nullten Grades darstellbar als Funktion der Verhältnisse ihrer Argumente zu einem dieser Argumente. Das Umgekehrte ist offenbar ebenfalls richtig, so daß die obige Identität einen allgemeinen Ausdruck einer homogenen Funktion nullten Grades darstellt.

Ist $f(x_1, \ldots, x_n)$ eine homogene Funktion m-ten Grades (und nur dann), so ist $\dfrac{f(x_1, \ldots, x_n)}{x_1^m}$ eine homogene Funktion nullten Grades; also ist in diesem Fall

$$\frac{f(x_1, x_2, \ldots, x_n)}{x_1^m} = \varphi\left(\frac{x_2}{x_1}, \ldots, \frac{x_n}{x_1}\right).$$

Somit ergibt sich die allgemeine Gestalt einer homogenen Funktion m-ten Grades zu

$$f(x_1, x_2, \ldots, x_n) = x_1^m \cdot \varphi\left(\frac{x_2}{x_1}, \ldots, \frac{x_n}{x_1}\right).$$

Beispiel. $x \cdot \dfrac{\sqrt{x^4 + y^4}}{x - y} \cdot \ln\dfrac{x}{y} = x^2 \cdot \dfrac{\sqrt{1 + \left(\dfrac{y}{x}\right)^4}}{\dfrac{y}{x} - 1} \cdot \ln\dfrac{y}{x}.$

188. Die Eulersche Formel. Wir nehmen nun an, eine homogene Funktion $f(x, y, z)$ vom Grad m habe in einem offenen Bereich $\mathfrak{D}$ stetige partielle Ableitungen nach allen Argumenten.[1]) Wir greifen einen beliebigen Punkt (x_0, y_0, z_0) aus $\mathfrak{D}$ heraus; dann ist nach Definition (19) für jedes $t > 0$

$$f(tx_0, ty_0, tz_0) = t^m \cdot f(x_0, y_0, z_0).$$

Diese Identität differenzieren wir nach t: die linke Seite nach der Kettenregel[2]), die rechte einfach als Potenz von t. Wir erhalten

$$f_x(tx_0, ty_0, tz_0) \cdot x_0 + f_y(tx_0, ty_0, tz_0) \cdot y_0 + f_z(tx_0, ty_0, tz_0) \cdot z_0$$
$$= mt^{m-1} \cdot f(x_0, y_0, z_0).$$

Für $t = 1$ folgt

$$f_x(x_0, y_0, z_0) \cdot x_0 + f_y(x_0, y_0, z_0) \cdot y_0 + f_z(x_0, y_0, z_0) \cdot z_0 = m \cdot f(x_0, y_0, z_0).$$

Also gilt für jeden Punkt (x, y, z) die Beziehung

$$f_x(x, y, z) \cdot x + f_y(x, y, z) \cdot y + f_z(x, y, z) \cdot z = m \cdot f(x, y, z). \tag{20}$$

Man nennt sie die *Eulersche Formel.*

Wie wir gesehen haben, genügt jede homogene Funktion m-ten Grades, die stetige partielle Ableitungen besitzt, dieser Gleichung. Wir zeigen jetzt das Umgekehrte: Jede Funktion, die nebst ihren partiellen Ableitungen stetig ist und der Eulerschen Formel (20) genügt, ist eine homogene Funktion m-ten Grades.

Es sei also $f(x, y, z)$ eine solche Funktion. Für festes x_0, y_0, z_0 betrachten wir für $t > 0$ folgende Funktion von t:

$$\varphi(t) = \frac{f(tx_0, ty_0, tz_0)}{t^m}.$$

Sie ist für alle in Frage kommenden $t > 0$ definiert und stetig. Wir berechnen ihre Ableitung $\varphi'(t)$ nach der Regel für die Differentiation eines Bruches und erhalten im Zähler (bis auf den Faktor t^{m-1})

$$[f_x(tx_0, ty_0, tz_0) \cdot x_0 + f_y(tx_0, ty_0, tz_0) \cdot y_0 + f_z(tx_0, ty_0, tz_0) \cdot z_0] \cdot t$$
$$- m \cdot f(tx_0, ty_0, tz_0).$$

Ersetzen wir in (20) x, y, z durch tx_0, ty_0, tz_0, so sehen wir, daß dieser Zähler gleich 0 ist. Daher ist $\varphi'(t) = 0$ und $\varphi(t) = c$ (c = const) für $t > 0$. Um c zu bestimmen, setzen wir $t = 1$ in der Definitionsgleichung für $\varphi(t)$. Dann ergibt sich

$$c = f(x_0, y_0, z_0).$$

Also ist

$$\varphi(t) = \frac{f(tx_0, ty_0, tz_0)}{t^m} = f(x_0, y_0, z_0)$$

oder

$$f(tx_0, ty_0, tz_0) = t^m \cdot f(x_0, y_0, z_0),$$

was zu beweisen war.

Man kann also sagen, daß die Eulersche Formel (20) die homogenen Funktionen vom Grad m ebensogut charakterisiert wie die definierende Gleichung (19).

[1]) Wir beschränken uns nur der Einfachheit halber auf Funktionen dreier Veränderlicher.

[2]) Um diese Regel anwenden zu können, haben wir die Stetigkeit der partiellen Ableitungen vorausgesetzt (vgl. Nr. 181).

§ 4. Ableitungen und Differentiale höherer Ordnung

189. Ableitungen höherer Ordnung. Wenn die Funktion[1] $u = f(x, y, z)$ in dem (offenen) Bereich $\mathfrak{D}$ eine partielle Ableitung nach einer der Veränderlichen hat, so kann diese Ableitung, die selbst eine Funktion von x, y, z ist, ihrerseits in einem Punkt (x_0, y_0, z_0) partielle Ableitungen nach derselben oder nach einer anderen Veränderlichen haben. Für die Ausgangsfunktion $u = f(x, y, z)$ sind diese Ableitungen *partielle Ableitungen zweiter Ordnung* (oder *zweite partielle Ableitungen*).

Ist z. B. die Funktion nach x abgeleitet, so bezeichnet man ihre zweiten Ableitungen nach x, y, z folgendermaßen[2]:

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 f(x_0, y_0, z_0)}{\partial x^2},$$

$$\frac{\partial^2 u}{\partial x\, \partial y} = \frac{\partial^2 f(x_0, y_0, z_0)}{\partial x\, \partial y},$$

$$\frac{\partial^2 u}{\partial x\, \partial z} = \frac{\partial^2 f(x_0, y_0, z_0)}{\partial x\, \partial z}$$

oder

$$u_{x^2} = u_{xx} = f_{xx}(x_0, y_0, z_0),$$

$$u_{xy} = f_{xy}(x_0, y_0, z_0),$$

$$u_{xz} = f_{xz}(x_0, y_0, z_0).$$

Analog definieren wir die Ableitungen *dritter, vierter, … Ordnung* (die dritten, vierten, … Ableitungen). Die partiellen Ableitungen *n-ter Ordnung* können induktiv definiert werden. Die partiellen Ableitungen höherer Ordnung nach verschiedenen Veränderlichen, z. B.

$$\frac{\partial^2 u}{\partial x\, \partial y}, \quad \frac{\partial^2 u}{\partial y\, \partial x}, \quad \frac{\partial^4 u}{\partial x\, \partial y\, \partial z^2}, \quad \ldots,$$

werden *gemischte partielle Ableitungen* genannt.

Beispiele.

1. Es sei $u = x^4 y^3 z^2$; dann ist z. B.

$$\begin{array}{llll} u_x = 4x^3y^3z^2, & u_{xy} = 12x^3y^2z^2, & u_{xyz} = 24x^3y^2z, & u_{xyzx} = 72x^2y^2z, \\ u_y = 3x^4y^2z^2, & u_{yx} = 12x^3y^2z^2, & u_{yxx} = 36x^2y^2z^2, & u_{yxxz} = 72x^2y^2z, \\ u_z = 2x^4y^3z, & u_{zx} = 8x^3y^3z, & u_{zxy} = 24x^3y^2z, & u_{zxyx} = 72x^2y^2z. \end{array}$$

2. Wir hatten schon in Nr. 177 die partiellen Ableitungen für die Funktion $u = \arctan \frac{x}{y}$ gebildet:

$$\frac{\partial u}{\partial x} = \frac{y}{x^2 + y^2}, \quad \frac{\partial u}{\partial y} = -\frac{x}{x^2 + y^2};$$

---

[1]) Wir beschränken uns auch hier auf Funktionen dreier Veränderlicher.

[2]) Diese Bezeichnungen sind wieder als einheitliche Symbole aufzufassen; ∂x^2 im Nenner bedeutet $\partial x\, \partial x$ und besagt, daß *zweimal* nach x differenziert wird. Der im folgenden auftretende Index x^2 bedeutet xx. Das ist stets zu beachten.

wir berechnen jetzt die weiteren Ableitungen:

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{y}{x^2+y^2}\right) = -\frac{2xy}{(x^2+y^2)^2},$$

$$\frac{\partial^2 u}{\partial x\,\partial y} = \frac{\partial}{\partial y}\left(\frac{y}{x^2+y^2}\right) = \frac{x^2-y^2}{(x^2+y^2)^2},$$

$$\frac{\partial^2 u}{\partial y\,\partial x} = \frac{\partial}{\partial x}\left(-\frac{x}{x^2+y^2}\right) = \frac{x^2-y^2}{(x^2+y^2)^2},$$

$$\frac{\partial^2 u}{\partial y^2} = \frac{\partial}{\partial y}\left(-\frac{x}{x^2+y^2}\right) = \frac{2xy}{(x^2+y^2)^2};$$

$$\frac{\partial^3 u}{\partial x^2\,\partial y} = \frac{\partial}{\partial y}\left(-\frac{2xy}{(x^2+y^2)^2}\right) = \frac{6xy^2-2x^3}{(x^2+y^2)^3},$$

$$\frac{\partial^3 u}{\partial y\,\partial x^2} = \frac{\partial}{\partial x}\left(\frac{x^2-y^2}{(x^2+y^2)^2}\right) = \frac{6xy^2-2x^3}{(x^2+y^2)^3}$$

usw.

3. Für die Funktion

$$u = \frac{1}{\sqrt{x^2+y^2+z^2}} = (x^2+y^2+z^2)^{-1/2}$$

gilt.

$$\frac{\partial u}{\partial x} = -x\cdot(x^2+y^2+z^2)^{-3/2},$$

$$\frac{\partial^2 u}{\partial x^2} = 3x^2\cdot(x^2+y^2+z^2)^{-5/2} - (x^2+y^2+z^2)^{-3/2}.$$

Analoge Ausdrücke erhalten wir auch für $\frac{\partial^2 u}{\partial y^2}$, $\frac{\partial^2 u}{\partial z^2}$. Wenn wir diese drei Ausdrücke addieren, so ergibt sich, daß die Funktion u der Gleichung

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0$$

genügt.

4. Es sei $y = f(x+at) + \varphi(x-at)$, wobei $a = \text{const}$ und $f(u)$, $\varphi(u)$ zwei beliebige Funktionen sind, die eine erste und eine zweite Ableitung besitzen. Wir zeigen, daß y der Gleichung

$$\frac{\partial^2 y}{\partial t^2} = a^2\cdot\frac{\partial^2 y}{\partial x^2}$$

genügt, wie auch immer die Funktionen f und φ beschaffen sind.

Wenn wir die Regeln für die Differentiation zusammengesetzter Funktionen benutzen, so finden wir[1])

$$\frac{\partial y}{\partial x} = f'(x+at) + \varphi'(x-at), \qquad \frac{\partial^2 y}{\partial x^2} = f''(x+at) + \varphi''(x-at),$$

$$\frac{\partial y}{\partial t} = f'(x+at)\cdot a + \varphi'(x-at)\cdot(-a),$$

$$\frac{\partial^2 y}{\partial t^2} = f''(x+at)\cdot a^2 + \varphi''(x-at)\cdot(-a)^2 = a^2\cdot\frac{\partial^2 y}{\partial x^2},$$

was zu beweisen war.

[1]) Die Striche an $f, \varphi, \ldots$ bedeuten die Ableitungen nach dem Argument u der Funktion $f(u)$, $\varphi(u)$.

5. Wir beweisen, daß der Ausdruck

$$z = x \cdot \varphi\left(\frac{y}{x}\right) = \psi\left(\frac{y}{x}\right),$$

wobei φ und ψ beliebige Funktionen sind (die eine erste und eine zweite Ableitung besitzen), der Gleichung

$$x^2 \cdot \frac{\partial^2 z}{\partial x^2} + 2xy \cdot \frac{d^2 z}{\partial x\, \partial y} + y^2 \cdot \frac{\partial^2 z}{\partial y^2} = 0$$

genügt.

Es gilt

$$\frac{\partial z}{\partial x} = \varphi\left(\frac{y}{x}\right) - \frac{y}{x} \cdot \varphi'\left(\frac{y}{x}\right) - \frac{y}{x^2} \cdot \psi'\left(\frac{y}{x}\right),$$

$$\frac{\partial z}{\partial y} = \varphi'\left(\frac{y}{x}\right) + \frac{1}{x} \cdot \psi'\left(\frac{y}{x}\right);$$

$$\frac{\partial^2 z}{\partial x^2} = \frac{y^2}{x^3} \cdot \varphi''\left(\frac{y}{x}\right) + \frac{2y}{x^3} \cdot \psi'\left(\frac{y}{x}\right) + \frac{y^2}{x^4} \cdot \psi''\left(\frac{y}{x}\right),$$

$$\frac{\partial^2 z}{\partial x\, \partial y} = -\frac{y}{x^2} \cdot \varphi''\left(\frac{y}{x}\right) - \frac{1}{x^2} \cdot \psi'\left(\frac{y}{x}\right) - \frac{y}{x^3} \cdot \psi''\left(\frac{y}{x}\right),$$

$$\frac{\partial^2 z}{\partial y^2} = \frac{1}{x} \cdot \varphi''\left(\frac{y}{x}\right) + \frac{1}{x^2} \cdot \psi''\left(\frac{y}{x}\right).$$

Werden die letzten drei Ableitungen mit x^2, $2xy$, y^2 multipliziert und dann addiert, so erhalten wir tatsächlich 0.

190. Der Satz über die gemischten Ableitungen. Sieht man sich die Beispiele 1 und 2 näher an, so fällt auf, daß die gemischten Ableitungen, die nach ein und denselben Veränderlichen, aber in verschiedener Reihenfolge gebildet werden, übereinstimmen.

Es sei zunächst bemerkt, daß dies keinesfalls notwendigerweise aus der Definition der gemischten Ableitungen folgt; es gibt Fälle, in denen sie nicht übereinstimmen.

Als Beispiel betrachten wir die Funktion

$$f(x, y) = xy\,\frac{x^2 - y^2}{x^2 + y^2} \quad \text{für} \quad x^2 + y^2 > 0, \qquad f(0, 0) = 0.$$

Dann ist

$$f_x(x, y) = y \cdot \left[\frac{x^2 - y^2}{x^2 + y^2} + \frac{4x^2y^2}{(x^2 + y^2)^2}\right] \quad \text{für} \quad x^2 + y^2 > 0, \qquad f_x(0, 0) = 0.$$

Setzen wir für x den speziellen Wert 0 ein, so erhalten wir für beliebiges y (auch für $y = 0$) die Beziehung $f_x(0, y) = -y$. Wenn wir diese Funktion nach y differenzieren, so erhalten wir $f_{xy}(0, y) = -1$. Hieraus folgt insbesondere, daß im Punkt (0, 0)

$$f_{xy}(0, 0) = -1$$

gilt. Berechnen wir auf dieselbe Weise f_{yx} im Punkt (0, 0), so erhalten wir

$$f_{yx}(0, 0) = 1.$$

Für die betrachtete Funktion ist also

$$f_{xy}(0, 0) \neq f_{yx}(0, 0).$$

Nichtsdestoweniger ist die an Beispielen bemerkte Übereinstimmung gemischter Ableitungen, die sich nur durch die Reihenfolge der Differentiation unterscheiden, nicht zufällig: Tatsächlich stimmen für eine große Klasse von Fällen unter bestimmten Bedingungen die gemischten Ableitungen überein. Wir beginnen mit folgendem einfachen Satz.

Satz. *Wir setzen voraus:*

a) *$f(x, y)$ sei in einem (offenen) Bereich $\mathfrak{D}$ definiert;*

b) *in diesem Bereich mögen die ersten Ableitungen f_x und f_y sowie die gemischten zweiten Ableitungen f_{xy} und f_{yx} existieren;*

c) *f_{xy} und f_{yx} seien als Funktionen von x und y in dem Punkt (x_0, y_0) von $\mathfrak{D}$ stetig.*
Dann gilt in diesem Punkt

$$f_{xy}(x_0, y_0) = f_{yx}(x_0, y_0). \tag{1}$$

Beweis. Wir betrachten den Ausdruck

$$W = \frac{f(x_0 + h, y_0 + k) - f(x_0 + h, y_0) - f(x_0, y_0 + k) + f(x_0, y_0)}{hk},$$

wobei h und k von 0 verschieden, etwa positiv und außerdem so klein sind, daß das ganze Rechteck $[x_0, x_0 + k; y_0, y_0 + k]$ in $\mathfrak{D}$ enthalten ist; das möge bis zum Schluß der Überlegung so bleiben.

Wir führen jetzt eine Hilfsfunktion von x ein:

$$\varphi(x) = \frac{f(x, y_0 + k) - f(x, y_0)}{k},$$

die im Intervall $[x_0, x_0 + h]$ nach b) die Ableitung

$$\varphi'(x) = \frac{f_x(x, y_0 + k) - f_x(x, y_0)}{k}$$

hat und somit stetig ist. Mit Hilfe dieser Funktion kann

$$W = \frac{1}{h}\left[\frac{f(x_0 + h, y_0 + k) - (x_0 + h, y_0)}{k} - \frac{f(x_0, y_0 + k) - f(x_0, y_0)}{k}\right] \tag{2}$$

in der Gestalt

$$W = \frac{\varphi(x_0 + h) - \varphi(x_0)}{h}$$

geschrieben werden.

Weil für die Funktion $\varphi(x)$ im Intervall $[x_0, x_0 + h]$ alle Voraussetzungen des Mittelwertsatzes (Nr. 112) erfüllt sind, können wir den Ausdruck W folgendermaßen umformen:

$$W = \varphi'(x_0 + \theta h) = \frac{f_x(x_0 + \theta h, y_0 + k) - f_x(x_0 + \theta h, y_0)}{k} \quad (0 < \theta < 1).$$

Da die zweite Ableitung $f_{xy}(x, y)$ existiert, können wir den Mittelwertsatz nochmals anwenden, und zwar dieses Mal auf die von y abhängende Funktion $f_x(x_0 + \theta h, y)$ im Intervall $[y_0, y_0 + k]$.[1]) So erhalten wir schließlich

$$W = f_{xy}(x_0 + \theta h, y_0 + \theta_1 k) \quad (0 < \theta < 1; 0 < \theta_1 < 1). \tag{3}$$

[1]) Zwar haben wir die zweidimensionale Stetigkeit von f_x und f_y nicht vorausgesetzt; wegen der vorausgesetzten Existenz von f_{xy} ist aber f_x bei festem x in y stetig. Analog ist f_y in x stetig.

Der Ausdruck W enthält aber x und y einerseits und h und k andererseits in gleicher Weise. Daher können ihre Rollen vertauscht werden; wenn wir die Hilfsfunktion

$$\psi(y) = \frac{f(x_0 + h, y) - f(x_0, y)}{h}$$

einführen, erhalten wir durch analoge Überlegungen das Ergebnis

$$W = f_{yx}(x_0 + \theta_2 h, y_0 + \theta_3 k) \quad (0 < \theta_2 < 1; 0 < \theta_3 < 1). \tag{4}$$

Wenn wir (3) und (4) vergleichen, erhalten wir

$$f_{xy}(x_0 + \theta h, y_0 + \theta_1 k) = f_{yx}(x_0 + \theta_2 h, y_0 + \theta_3 k).$$

Wir lassen jetzt h und k gegen 0 streben und gehen in dieser Gleichung zur Grenze über. Da die Faktoren $\theta, \theta_1, \theta_2, \theta_3$ beschränkt sind, streben sowohl die Argumente auf der rechten Seite als auch die auf der linken gegen x_0 bzw. y_0. Nach c) erhalten wir also schließlich

$$f_{xy}(x_0, y_0) = f_{yx}(x_0, y_0),$$

was zu beweisen war.

Somit stimmen *stetige* gemischte Ableitungen f_{xy} und f_{yx} immer überein.

In dem oben angeführten Beispiel haben diese Ableitungen

$$f_{xy} = f_{yx} = \frac{x^2 - y^2}{x^2 + y^2} \cdot \left\{1 + \frac{8x^2y^2}{(x^2 + y^2)^2}\right\} \qquad (x^2 + y^2 > 0)$$

für $x \to 0$, $y \to 0$ überhaupt keinen Grenzwert; folglich sind sie im Punkt (0, 0) unstetig. In diesem Fall ist der Satz natürlich nicht anwendbar.

Es ist interessant, das Problem der Gleichung (1) dem in Nr. 168 betrachteten Problem der iterierten Grenzwerte gegenüberzustellen. Setzt man die Existenz der ersten Ableitungen voraus und schreibt den Ausdruck W in der Form (2), so sieht man leicht, daß

$$\lim_{k \to 0} W = \frac{f_y(x_0 + h, y_0) - f_y(x_0, y_0)}{h} \qquad (h = \text{const}) \tag{5}$$

und analog

$$\lim_{h \to 0} W = \frac{f_x(x_0, y_0 + k) - f_x(x_0, y_0)}{k} \qquad (k = \text{const}) \tag{5*}$$

ist. Dann gilt nach Definition der Ableitung

$$f_{yx}(x_0, y_0) = \lim_{h \to 0} \frac{f_y(x_0 + h, y_0) - f_y(x_0, y_0)}{h} = \lim_{h \to 0} \lim_{k \to 0} W, \tag{6}$$

$$f_{xy}(x_0, y_0) = \lim_{k \to 0} \frac{f_x(x_0, y_0 + k) - f_x(x_0, y_0)}{k} = \lim_{k \to 0} \lim_{h \to 0} W. \tag{6*}$$

Somit ist das Problem der Existenz und Gleichheit der gemischten Ableitungen identisch mit dem Problem der Existenz und Gleichheit der iterierten Grenzwerte für den (von h und k abhängenden) Ausdruck W.

Auf Grund dieser Bemerkung kann man den soeben bewiesenen Satz verschärfen: *Wir setzen jetzt nur voraus, daß neben den ersten Ableitungen noch eine der gemischten Ableitungen, etwa $f_{xy}(x, y)$, in der Umgebung des Punktes (x_0, y_0) existiert (in diesem Punkt selbst wird die Existenz nicht vorausgesetzt). Ferner möge der endliche Grenzwert*

$$\lim_{\substack{x \to x_0 \\ y \to y_0}} f_{xy}(x, y) = A$$

existieren. Hieraus ergeben sich schon die Existenz der beiden gemischten Ableitungen im Punkt (x_0, y_0) und die Gleichung (1).[1])

Man kann nämlich, ebenso wie oben, auf Grund der Voraussetzungen zu Gleichung (3) kommen und danach, unter Benutzung der Existenz des Grenzwertes der Funktion $f_{xy}(x, y)$ im Punkt (x_0, y_0), die Existenz des zweifachen Grenzwertes für gleichzeitig gegen 0 strebende h und k erschließen:

$$\lim_{\substack{h \to 0 \\ k \to 0}} W = A .$$

Die einfachen Grenzwerte (5) und (5*) existieren aber nach Voraussetzung; nach dem Satz aus Nr. 168 existieren dann aber auch die iterierten Grenzwerte (6) und (6*) und stimmen einzeln mit dem zweifachen Grenzwert überein. Das bedeutet aber gerade, daß die Ableitungen $f_{xy}(x_0, y_0)$ und $f_{yx}(x_0, y_0)$ existieren und einander gleich sind.

191. Verallgemeinerung. Wir kommen jetzt zum Beweis eines allgemeinen Satzes über gemischte Ableitungen:

Satz. *Die Funktion $u = f(x_1, x_2, \ldots, x_n)$ von n Veränderlichen sei in einem (offenen) n-dimensionalen Bereich $\mathfrak{D}$ definiert und besitze dort alle möglichen partiellen Ableitungen bis zur $(k - 1)$-ten Ordnung einschließlich und alle gemischten Ableitungen k-ter Ordnung, und alle diese Ableitungen seien in $\mathfrak{D}$ stetig.*

Unter diesen Voraussetzungen hängt der Wert jeder k-ten gemischten Ableitung nicht von der Reihenfolge ab, in der sukzessive differenziert wird.

Beweis. Für $k = 2$ ist der Satz schon bewiesen, denn es ist beispielsweise

$$\frac{\partial^2 u}{\partial x_i \, \partial x_j} = \frac{\partial^2 u}{\partial x_j \, \partial x_i}.$$

Um nämlich diesen Fall auf den Satz aus Nr. 190 zurückzuführen, genügt es zu bemerken, daß man bei der Berechnung dieser Ableitungen allen übrigen Veränderlichen (außer x_i und x_j) konstante Werte zuschreiben kann; da diese Ableitungen in allen Veränderlichen stetig sind, sind sie es auch in den Veränderlichen x_i und x_j, wenn die übrigen festgehalten werden. Es sei jetzt $k > 2$.

Zunächst beweisen wir den Satz für den Fall, daß bei der Berechnung einer Ableitung k-ter Ordnung nur zwei aufeinanderfolgende Differentiationen vertauscht werden, d. h., wir beweisen

$$\frac{\partial^k u}{\partial x_{i_1} \, \partial x_{i_2} \cdots \partial x_{i_h} \, \partial x_{i_{h+1}} \cdots \partial x_{i_k}} = \frac{\partial^k u}{\partial x_{i_1} \, \partial x_{i_2} \cdots \partial x_{i_{h+1}} \, \partial x_{i_h} \cdots \partial x_{i_k}}. \tag{7}$$

[1]) Dieser Satz stammt von Hermann Amandus Schwarz (1843–1921, deutscher Mathematiker).

(Hier bedeutet $i_1, i_2, \ldots, i_h, i_{h+1}, \ldots, i_k$ ein System von k Indizes aus der Folge $1, 2, \ldots, n$ mit möglichen Wiederholungen). Wenn wir die zur Berechnung dieser Ableitungen notwendigen Differentiationen durchführen, so sehen wir, daß die Ableitungen $(h-1)$-ter Ordnung in beiden Fällen übereinstimmen. Wenden wir hierauf den für $k = 2$ schon bewiesenen Satz an, so finden wir, daß auch die Ableitungen $(h+1)$-ter Ordnung übereinstimmen. Alsdann sind in beiden Fällen die gleichen Operationen durchzuführen, die natürlich zu denselben Resultaten führen.

Somit gilt Beziehung (7), und der Satz ist für diesen Fall bewiesen. Da aber jede Permutation von Elementen durch Hintereinanderausführung einer Reihe von Vertauschungen je zweier aufeinanderfolgender Elemente erhalten werden kann, ist der Satz auch ganz allgemein bewiesen. Sind die entsprechenden Ableitungen stetig, so kann man die Differentiation nach verschiedenen Veränderlichen miteinander vertauschen.

Die Stetigkeit der Ableitungen werden wir von jetzt an immer voraussetzen, so daß die Reihenfolge der sukzessiven Differentiationen gleichgültig ist. Daher können wir von jetzt an bei der Bezeichnung der gemischten Ableitungen die Differentiationen nach ein und derselben Veränderlichen zusammenfassen. Ist u eine Funktion von $x_1, x_2, \ldots, x_n$, so können wir eine solche Ableitung in der Form

$$\frac{\partial^k u}{\partial x_1^{\alpha_1}\, \partial x_2^{\alpha_2} \cdots \partial x_n^{\alpha_n}}$$

schreiben, wobei $\alpha_1 + \alpha_2 + \cdots + \alpha_n = k$ ist; wenn u eine Funktion von $x, y, \ldots, z$ ist, so schreiben wir

$$\frac{\partial^k u}{\partial x^{\alpha}\, \partial y^{\beta} \cdots \partial z^{\gamma}},$$

wobei $\alpha + \beta + \cdots + \gamma = k$ ist. Die einzelnen „Exponenten" $\alpha_1, \alpha_2, \ldots, \alpha_n$ bzw. $\alpha, \beta, \ldots, \gamma$ können auch 0 sein; tritt etwa ∂x^0 auf, so bedeutet das einfach, daß nach der Veränderlichen x nicht differenziert wird.

192. Ableitungen höherer Ordnung für mittelbare Funktionen. Es sei die Funktion

$$u = f(x_1, x_2, \ldots, x_n)$$

gegeben, die $x_1, x_2, \ldots, x_n$ seien ihrerseits Funktionen der Veränderlichen $t_1, t_2, \ldots, t_m$:

$$x_i = \varphi_i(t_1, t_2, \ldots, t_m) \quad (i = 1, 2, \ldots, n).$$

Bezüglich der Funktionen f und φ_i setzen wir voraus, daß sie stetige partielle Ableitungen nach allen Veränderlichen bis zur k-ten Ordnung einschließlich besitzen. Wir betrachten nun u als mittelbare Funktion der Veränderlichen $t_1, t_2, \ldots, t_m$,

$$u = F(t_1, t_2, \ldots, t_m) = f(\varphi_1, \varphi_2, \ldots, \varphi_n),$$

und zeigen, daß auch die zusammengesetzte Funktion alle Ableitungen bis zur k-ten Ordnung einschließlich besitzt und daß diese Ableitungen ebenfalls stetig sind.

Genauer gesagt werden wir folgenden Satz beweisen: *Sämtliche Ableitungen k-ter Ordnung der Funktion F existieren und lassen sich aus Ableitungen der Funktion f (nach den Argumenten $x_1, x_2, \ldots, x_n$) und der Funktionen φ_i (nach den Argumenten $t_1, t_2, \ldots, t_m$), deren Ordnung nicht höher als k ist, mit Hilfe von Multiplikation und Addition bilden.*

Den Beweis führen wir induktiv. Für $k = 1$ ist die Behauptung richtig; das folgt aus der schon früher bewiesenen Formel für die Ableitung einer mittelbaren Funktion (Nr. 181).

Wir setzen nun voraus, der Satz sei für alle Ableitungen, deren Ordnung kleiner als k ist, schon bewiesen, und zeigen, daß er dann auch für die Ableitungen k-ter Ordnung gilt. Jede k-te Ableitung erhält man aus einer $(k-1)$-ten durch Differentiation nach einer der Veränderlichen t_j. Eine Ableitung $(k-1)$-ter Ordnung erhält man nach Induktionsvoraussetzung aus Ableitungen von f und φ_i höchstens $(k-1)$-ter Ordnung nach den Veränderlichen x und t durch Multiplikation und Addition, d. h. als Summe von Produkten solcher Ableitungen. Wenn wir eines dieser Produkte nach t_j differenzieren, so müssen wir der Reihe nach jeden der Faktoren differenzieren. Ist dieser Faktor eine Ableitung höchstens $(k-1)$-ter Ordnung einer der Funktionen φ_i, so erhalten wir im Ergebnis der Differentiation eine Ableitung derselben Funktion, deren Ordnung nicht höher als k ist. Ist er eine Ableitung höchstens $(k-1)$-ter Ordnung der Funktion f, so können wir diese Ableitung als mittelbare Funktion der Veränderlichen t betrachten und nach t_j differenzieren; diese können wir alsdann durch eine Summe von Produkten ersetzen.[1])

Im Ergebnis erhalten wir für die betreffende Ableitung k-ter Ordnung offenbar einen Ausdruck der erwähnten Form; damit ist unsere Behauptung bewiesen.

Die Stetigkeit der Ableitungen der zusammengesetzten Funktion F folgt aus der Art, wie sie aus Ableitungen von f und φ_i gebildet werden, da diese als stetig vorausgesetzt wurden.

193. Differentiale höherer Ordnung. In einem Bereich $\mathfrak{D}$ sei eine Funktion $u = f(x_1, x_2, \ldots, x_n)$ gegeben, die stetige partielle Ableitungen erster Ordnung besitze. Dann versteht man, wie wir wissen, unter dem (vollständigen) Differential $\mathrm{d}u$ den Ausdruck

$$\mathrm{d}u = \frac{\partial u}{\partial x_1}\,\mathrm{d}x_1 + \frac{\partial u}{\partial x_2}\,\mathrm{d}x_2 + \cdots + \frac{\partial u}{\partial x_n}\,\mathrm{d}x_n,$$

wobei $\mathrm{d}x_1, \ldots, \mathrm{d}x_n$ beliebige Zuwächse der unabhängigen Veränderlichen $x_1, \ldots, x_n$ sind.

Offenbar ist $\mathrm{d}u$ wieder eine Funktion von $x_1, \ldots, x_n$. Setzt man die Existenz stetiger partieller Ableitungen zweiter Ordnung von u voraus, so hat $\mathrm{d}u$ stetige partielle Ableitungen erster Ordnung; man kann also von dem (vollständigen) Differential $\mathrm{d}(\mathrm{d}u)$ dieses Differentials $\mathrm{d}u$ sprechen: Man nennt es *Differential zweiter Ordnung* (oder *zweites Differential*) von u und schreibt d^2u.

Es sei besonders hervorgehoben, daß die Zuwächse $\mathrm{d}x_1, \mathrm{d}x_2, \ldots, \mathrm{d}x_n$ hierbei als Konstante angesehen werden und beim Übergang von einem Differential zu dem nächst höheren unverändert bleiben.

Benutzen wir die Differentiationsregeln aus Nr. 185, so erhalten wir

$$\begin{aligned}\mathrm{d}^2u = \mathrm{d}(\mathrm{d}u) &= \mathrm{d}\left(\frac{\partial u}{\partial x_1}\,\mathrm{d}x_1 + \frac{\partial u}{\partial x_2}\,\mathrm{d}x_2 + \cdots + \frac{\partial u}{\partial x_n}\,\mathrm{d}x_n\right)\\ &= \mathrm{d}\left(\frac{\partial u}{\partial x_1}\right)\cdot \mathrm{d}x_1 + \mathrm{d}\left(\frac{\partial u}{\partial x_2}\right)\cdot \mathrm{d}x_2 + \cdots + \mathrm{d}\left(\frac{\partial u}{\partial x_n}\right)\cdot \mathrm{d}x_n\end{aligned}$$

---

[1]) Die vorausgesetzte Stetigkeit aller Ableitungen von f und φ_i ermöglicht die Anwendung der schon bekannten Kettenregel zur Bildung der Ableitung einer mittelbaren Funktion (Nr. 181).

oder, ausgeschrieben,

$$\begin{aligned}
\mathrm{d}^2 u = {} & \left(\frac{\partial^2 u}{\partial x_1^2}\,\mathrm{d}x_1 + \frac{\partial^2 u}{\partial x_1\,\partial x_2}\,\mathrm{d}x_2 + \cdots + \frac{\partial^2 u}{\partial x_1\,\partial x_n}\,\mathrm{d}x_n\right)\cdot \mathrm{d}x_1 \\
& + \left(\frac{\partial^2 u}{\partial x_2\,\partial x_1}\,\mathrm{d}x_1 + \frac{\partial^2 u}{\partial x_2^2}\,\mathrm{d}x_2 + \cdots + \frac{\partial^2 u}{\partial x_2\,\partial x_n}\,\mathrm{d}x_n\right)\cdot \mathrm{d}x_2 + \cdots \\
& + \left(\frac{\partial^2 u}{\partial x_n\,\partial x_1}\,\mathrm{d}x_1 + \frac{\partial^2 u}{\partial x_n\,\partial x_2}\,\mathrm{d}x_2 + \cdots + \frac{\partial^2 u}{\partial x_n^2}\,\mathrm{d}x_n\right)\cdot \mathrm{d}x_n \\
= {} & \frac{\partial^2 u}{\mathrm{d}x_1^2}\,\mathrm{d}x_1^2 + \frac{\partial^2 u}{\partial x_2^2}\,\mathrm{d}x_2^2 + \cdots + \frac{\partial^2 u}{\partial x_n^2}\,\mathrm{d}x_n^2 \\
& + 2\,\frac{\partial^2 u}{\partial x_1\,\partial x_2}\,\mathrm{d}x_1\,\mathrm{d}x_2 + 2\,\frac{\partial^2 u}{\partial x_1\,\partial x_3}\,\mathrm{d}x_1\,\mathrm{d}x_3 + \cdots \\
& + 2\,\frac{\partial^2 u}{\partial x_2\,\partial x_3}\,\mathrm{d}x_2\,\mathrm{d}x_3 + \cdots + 2\,\frac{\partial^2 u}{\partial x_{n-1}\,\partial x_n}\,\mathrm{d}x_{n-1}\,\mathrm{d}x_n .
\end{aligned}$$

Analog definiert man das *Differential dritter Ordnung* $\mathrm{d}^3 u$, usw. Ist allgemein das Differential $(k-1)$-ter Ordnung $\mathrm{d}^{k-1}u$ schon definiert, so wird das *Differential k-ter Ordnung* $\mathrm{d}^k u$ als (vollständiges) Differential des Differentials $(k-1)$-ter Ordnung definiert[1]:

$$\mathrm{d}^k u = \mathrm{d}(\mathrm{d}^{k-1} u).$$

Wenn für eine Funktion u die stetigen partiellen Ableitungen bis zur k-ten Ordnun einschließlich existieren, so ist die Existenz dieses k-ten Differentials gesichert. Die Ausdrücke für die sukzessiven Differentiale werden jedoch immer komplizierter. Zur Vereinfachung der Schreibweise dient folgendes Verfahren.

Zuerst „klammert man“ in dem Ausdruck für das erste Differential den Buchstaben u „aus“ und schreibt *symbolisch*

$$\mathrm{d}u = \left(\frac{\partial}{\partial x_1}\,\mathrm{d}x_1 + \frac{\partial}{\partial x_2}\,\mathrm{d}x_2 + \cdots + \frac{\partial}{\partial x_n}\,\mathrm{d}x_n\right)\cdot u .$$

Klammern wir jetzt in dem Ausdruck für das zweite Differential u aus, so steht in der Klammer offenbar das Quadrat des Ausdrucks

$$\frac{\partial}{\partial x_1}\,\mathrm{d}x_1 + \frac{\partial}{\partial x_2}\,\mathrm{d}x_2 + \cdots + \frac{\partial}{\partial x_n}\,\mathrm{d}x_n ;$$

man kann also das zweite Differential *symbolisch* folgendermaßen schreiben:

$$\mathrm{d}^2 u = \left(\frac{\partial}{\partial x_1}\,\mathrm{d}x_1 + \frac{\partial}{\partial x_2}\,\mathrm{d}x_2 + \cdots + \frac{\partial}{\partial x_n}\,\mathrm{d}x_n\right)^2\cdot u .$$

Analog läßt sich das dritte Differential schreiben, usw. Allgemein kann man diese Regel so aussprechen: Für jedes k gilt in *symbolischer* Schreibweise

$$\mathrm{d}^k u = \left(\frac{\partial}{\partial x_1}\,\mathrm{d}x_1 + \frac{\partial}{\partial x_2}\,\mathrm{d}x_2 + \cdots + \frac{\partial}{\partial x_n}\,\mathrm{d}x_n\right)^k\cdot u ; \tag{8}$$

[1]) Man kann natürlich auch *partielle* Differentiale (vgl. Nr. 177) beliebiger Ordnung definieren; wir wollen aber darauf nicht näher eingehen.

das ist folgendermaßen zu verstehen: Zunächst ist das in der Klammer stehende „Polynom" nach den Regeln der Algebra in die k-te Potenz zu erheben, und dann sind alle erhaltenen Glieder mit u zu „multiplizieren", wobei u im Zähler hinter ∂^k zu schreiben ist. Erst dann werden die Symbole wieder als Ableitungen bzw. Differentiale gedeutet.

Diese Regel gilt offenbar für $k = 1$ und $k = 2$. Daher genügt es, sie unter der Voraussetzung, daß sie für $\mathrm{d}^k u$ gilt, für $\mathrm{d}^{k+1}u$ zu beweisen.

Unter dieser Induktionsvoraussetzung gilt

$$\mathrm{d}^k u = \sum C_{\alpha_1\alpha_2\ldots\alpha_n} \frac{\partial^k u}{\partial x_1^{\alpha_1}\,\partial x_2^{\alpha_2}\cdots\partial x_n^{\alpha_n}}\,\mathrm{d}x_1^{\alpha_1}\,\mathrm{d}x_2^{\alpha_2}\cdots\mathrm{d}x_n^{\alpha_n},$$

wobei sich die Summation über alle möglichen Systeme nichtnegativer ganzer Zahlen $\alpha_1, \alpha_2, \ldots, \alpha_n$ erstreckt, für die $\alpha_1 + \alpha_2 + \cdots + \alpha_n = k$ ist, und

$$C_{\alpha_1\alpha_2\ldots\alpha_n} = \frac{k!}{\alpha_1!\,\alpha_2!\cdots\alpha_n!}$$

die „Polynomialkoeffizienten" sind.

Unter der Voraussetzung, daß die stetigen Ableitungen $(k+1)$-ter Ordnung existieren, differenzieren wir die obige Formel und erhalten

$$\begin{aligned}\mathrm{d}^{k+1}u = \sum C_{\alpha_1\alpha_2\ldots\alpha_n} \Bigg\{ & \frac{\partial^{k+1}u}{\partial x_1^{\alpha_1+1}\,\partial x_2^{\alpha_2}\cdots\partial x_n^{\alpha_n}}\,\mathrm{d}x_1^{\alpha_1+1}\,\mathrm{d}x_2^{\alpha_2}\cdots\mathrm{d}x_n^{\alpha_n}\\ & + \frac{\partial^{k+1}u}{\partial x_1^{\alpha_1}\,\partial x_2^{\alpha_2+1}\cdots\partial x_n^{\alpha_n}}\,\mathrm{d}x_1^{\alpha_1}\,\mathrm{d}x_2^{\alpha_2+1}\cdots\mathrm{d}x_n^{\alpha_n} + \cdots\\ & + \frac{\partial^{k+1}u}{\partial x_1^{\alpha_1}\,\partial x_2^{\alpha_2}\cdots\partial x_n^{\alpha_n+1}}\,\mathrm{d}x_1^{\alpha_1}\,\mathrm{d}x_2^{\alpha_2}\cdots\mathrm{d}x_n^{\alpha_n+1}\Bigg\}.\end{aligned}$$

Offenbar würden wir dasselbe erhalten, wenn wir den symbolischen Ausdruck

$$\begin{aligned}&\sum C_{\alpha_1\alpha_2\ldots\alpha_n} \frac{\partial^k}{\partial x_1^{\alpha}\;\partial x_2^{\alpha_2}\cdots\partial x_n^{\alpha_n}}\,\mathrm{d}x_1^{\alpha_1}\,\mathrm{d}x_2^{\alpha_2}\cdots\mathrm{d}x_n^{\alpha_n}\\ &\quad\times\left(\frac{\partial}{\partial x_1}\mathrm{d}x_1 + \frac{\partial}{\partial x_2}\mathrm{d}x_2 + \cdots + \frac{\partial}{\partial x_n}\mathrm{d}x_n\right)\end{aligned}$$

gebildet und dann u dahintergeschrieben hätten. Dieses „Produkt" ist jedoch nichts anderes als

$$\begin{aligned}&\left(\frac{\partial}{\partial x_1}\mathrm{d}x_1 + \frac{\partial}{\partial x_2}\mathrm{d}x_2 + \cdots + \frac{\partial}{\partial x_n}\mathrm{d}x_n\right)^k \left(\frac{\partial}{\partial x_1}\mathrm{d}x_1 + \frac{\partial}{\partial x_2}\mathrm{d}x_2 + \cdots + \frac{\partial}{\partial x_n}\mathrm{d}x_n\right)\\ &= \left(\frac{\partial}{\partial x_1}\mathrm{d}x_1 + \frac{\partial}{\partial x_2}\mathrm{d}x_2 + \cdots + \frac{\partial}{\partial x_n}\mathrm{d}x_n\right)^{k+1},\end{aligned}$$

so daß in der Tat

$$\mathrm{d}^{k+1}u = \left(\frac{\partial}{\partial x_1}\mathrm{d}x_1 + \frac{\partial}{\partial x_2}\mathrm{d}x_2 + \cdots + \frac{\partial}{\partial x_n}\mathrm{d}x_n\right)^{k+1}\cdot u$$

gilt, was zu beweisen war.

Diese Überlegungen zeigen, daß das k-te Differential ein *homogenes Polynom k-ten Grades* oder, wie man auch sagt, eine *Form k-ten Grades* in den Differentialen der unabhängigen Veränderlichen ist, deren Koeffizienten die mit ganzzahligen Konstanten (den „Polynomialkoeffizienten“) multiplizierten partiellen Ableitungen k-ter Ordnung sind.

Ist beispielsweise $u = f(x, y)$, so ist

$$d^2u = \frac{\partial^2 u}{\partial x^2} dx^2 + 2 \frac{\partial^2 u}{\partial x\, \partial y} dx\, dy + \frac{\partial^2 u}{\partial y^2} dy^2,$$

$$d^3u = \frac{\partial^3 u}{\partial x^3} dx^3 + 3 \frac{\partial^3 u}{\partial x^2\, \partial y} dx^2\, dy + 3 \frac{\partial^3 u}{\partial x\, \partial y^2} dx\, dy^2 + \frac{\partial^3 u}{\partial y^3} dy^3,$$

$$d^4u = \frac{\partial^4 u}{\partial x^4} dx^4 + 4 \frac{\partial^4 u}{\partial x^3\, \partial y} dx^3\, dy + 6 \frac{\partial^4 u}{\partial x^2\, \partial y^2} dx^2\, dy^2 + 4 \frac{\partial^4 u}{\partial x\, \partial y^3} dx\, dy^3 + \frac{\partial^4 u}{\partial y^4} dy^4$$

usw. Nehmen wir beispielsweise $u = \arctan \frac{x}{y}$, so erhalten wir

$$du = \frac{y dx - x dy}{x^2 + y^2},$$

$$d^2u = \frac{2xy(dy^2 - dx^2) + 2(x^2 - y^2)\, dx\, dy}{(x^2 + y^2)^2},$$

$$d^3u = \frac{(6x^2y - 2y^3)\, dx^3 + (18xy^2 - 6x^3)\, dx^2\, dy}{(x^2 + y^2)^3} + \frac{(6y^3 - 18x^2y)\, dx\, dy^2 + (2x^3 - 6xy^2)\, dy^3}{(x^2 + y^2)^3}$$

usw.

Der Ausdruck für das Differential wird mit wachsender Anzahl von Veränderlichen immer komplizierter. Ist $u = f(x, y, z)$, so lautet das dritte Differential d^3u folgendermaßen:

$$d^3u = \left(\frac{\partial}{\partial x} dx + \frac{\partial}{\partial y} dy + \frac{\partial}{\partial z} dz\right)^3 u$$
$$= \frac{\partial^3 u}{\partial x^3} dx^3 + \frac{\partial^3 u}{\partial y^3} dy^3 + \frac{\partial^3 u}{\partial z^3} dz^3 + 3 \frac{\partial^3 u}{\partial x^2\, \partial y} dx^2\, dy + 3 \frac{\partial^3 u}{\partial x\, \partial y^2} dx\, dy^2$$
$$+ 3 \frac{\partial^3 u}{\partial x^2\, \partial z} dx^2\, dz + 3 \frac{\partial^3 u}{\partial x\, \partial z^2} dx\, dz^2 + 3 \frac{\partial^3 u}{\partial y^2\, \partial z} \partial y^2\, \partial z$$
$$+ 3 \frac{\partial^3 u}{\partial y\, \partial z^2} dy\, dz^2 + 6 \frac{\partial^3 u}{\partial x\, \partial y\, \partial z} dx\, dy\, dz.$$

194. Differentiale mittelbarer Funktionen. Es sei jetzt die mittelbare Funktion

$$u = f(x_1, x_2, \ldots, x_n)$$

gegeben, wobei $x_i = \varphi_i(t_1, t_2, \ldots, t_m)$ $(i = 1, 2, \ldots, n)$ ist. In diesem Fall kann das erste Differential in der Gestalt

$$du = \frac{\partial u}{\partial x_1} dx_1 + \frac{\partial u}{\partial x_2} dx_2 + \cdots + \frac{\partial u}{\partial x_n} dx_n$$

dargestellt werden (auf Grund der Invarianz des ersten Differentials; Nr. 185). Hier sind jedoch $dx_1, \ldots, dx_n$ nicht mehr Differentiale von unabhängigen Veränderlichen, sondern von Funktionen, also im Gegensatz zu vorhin selbst Funktionen und keine Konstanten mehr.

Wenn wir jetzt unter Benutzung der Differentiationsregeln aus Nr. 185 das zweite Differential der Funktion berechnen, erhalten wir

$$\begin{aligned} d^2u &= d\left(\frac{\partial u}{\partial x_1}\right)\cdot dx_1 + d\left(\frac{\partial u}{\partial x_2}\right)\cdot dx_2 + \cdots + d\left(\frac{\partial u}{\partial x_n}\right)\cdot dx_n \\ &\quad + \frac{\partial u}{\partial x_1}\cdot d(dx_1) + \frac{\partial u}{\partial x_2}\cdot d(dx_2) + \cdots + \frac{\partial u}{\partial x_n}\cdot d(dx_n) \\ &= \left(\frac{\partial}{\partial x_1}dx_1 + \frac{\partial}{\partial x_2}dx_2 + \cdots + \frac{\partial}{\partial x_n}dx_n\right)^2\cdot u \\ &\quad + \frac{\partial u}{\partial x_1}\cdot d^2x_1 + \frac{\partial u}{\partial x_2}d^2x_2 + \cdots + \frac{\partial u}{\partial x_n}\cdot d^2x_n. \end{aligned}$$

Wir sehen, daß ein Differential höherer als erster Ordnung keineswegs invariant ist.

Wir betrachten jetzt den Spezialfall, daß $x_1, x_2, \ldots, x_n$ lineare Funktionen von $t_1, t_2, \ldots, t_m$ sind, d. h. den Fall, daß

$$x_i = \alpha_i^{(1)}t_1 + \alpha_i^{(2)}t_2 + \cdots + \alpha_i^{(m)}t_m + \beta_i \quad (i = 1, 2, \ldots, n)$$

ist, wobei $\alpha_i^{(j)}$ und β_i Konstante sind.

In diesem Fall ist

$$dx_i = \alpha_i^{(1)}\,dt_1 + \cdots + \alpha_i^{(m)}\,dt_m = \alpha_i^{(1)}\,\Delta t_1 + \cdots + \alpha_i^{(m)}\,\Delta t_m = \Delta x_i.$$

Wir sehen, daß jetzt alle ersten Differentiale der Funktionen $x_1, x_2, \ldots, x_n$ konstant sind, also nicht mehr von $t_1, t_2, \ldots, t_m$ abhängen; folglich lassen sich die Überlegungen aus Nr. 193 unverändert anwenden. Somit können bei der Ersetzung der unabhängigen Veränderlichen $x_1, x_2, \ldots, x_n$ durch lineare Funktionen der neuen Veränderlichen $t_1, t_2, \ldots, t_m$ die ursprünglichen Ausdrücke auch für die Differentiale höherer Ordnung beibehalten bleiben. In diesen stimmen die Differentiale $dx_1, dx_2, \ldots, dx_n$ mit den Zuwächsen $\Delta x_1, \Delta x_2, \ldots, \Delta x_n$ überein; aber diese Zuwächse sind nicht beliebig, sondern durch die Zuwächse $\Delta t_1, \Delta t_2, \ldots, \Delta t_m$ festgelegt.

Diese einfache und wichtige Bemerkung (die von Cauchy stammt) benutzen wir in Nr. 195.

195. Die Taylorsche Formel. Wir wissen [vgl. Nr. 126, Formel (13)], daß eine Funktion $F(t)$, die $n+1$ stetige Ableitungen besitzt, mit Hilfe der Taylorschen Formel folgendermaßen dargestellt werden kann:

$$\begin{aligned} F(t) &= F(t_0) + F'(t_0)\cdot(t - t_0) \\ &\quad + \frac{1}{2!}F''(t_0)\cdot(t-t_0)^2 + \cdots + \frac{1}{n!}F^{(n)}(t_0)\cdot(t-t_0)^n \\ &\quad + \frac{1}{(n+1)!}F^{(n+1)}\big(t_0 + \theta(t-t_0)\big)\cdot(t-t_0)^{n+1} \qquad (0 < \theta < 1) \end{aligned}$$

(Restglied in der Lagrangeschen Form). Setzen wir

$$t - t_0 = \Delta t = dt, \quad F(t) - F(t_0) = \Delta F(t_0),$$

so geht diese Formel in

$$\Delta F(t_0) = \mathrm{d}F(t_0) + \frac{1}{2!}\,\mathrm{d}^2F(t_0) + \cdots + \frac{1}{n!}\,\mathrm{d}^nF(t_0)$$
$$+ \frac{1}{(n+1)!}\,\mathrm{d}^{n+1}F(t_0 + \theta \cdot \Delta t) \quad (0 < \theta < 1) \tag{*}$$

über. Hierbei ist wichtig hervorzuheben, daß die Größe $\mathrm{d}t$, die in verschiedenen Potenzen in den Differentialausdrücken auf der rechten Seite auftritt, genau gleich dem Zuwachs Δt ist, der in dem Zuwachs der Funktion auf der linken Seite auftritt.

Die Taylorsche Formel (*) läßt sich auf den Fall mehrerer Veränderlicher übertragen.

Der Einfachheit halber beschränken wir uns auf Funktionen $f(x, y)$ zweier Veränderlicher.

Wir setzen voraus, diese Funktion besitze in einer Umgebung eines bestimmten Punktes (x_0, y_0) stetige Ableitungen bis zur $(n+1)$-ten Ordnung einschließlich. Die Zuwächse Δx und Δy seien so beschaffen, daß die geradlinige Verbindungsstrecke der Punkte (x_0, y_0) und $(x_0 + \Delta x,\ y_0 + \Delta y)$ innerhalb der betrachteten Umgebung des Punktes (x_0, y_0) verläuft.

Es soll gezeigt werden, daß unter unseren Voraussetzungen die Beziehung

$$\Delta f(x_0, y_0) = f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0)$$
$$= \mathrm{d}f(x_0, y_0) + \frac{1}{2!}\,\mathrm{d}^2f(x_0, y_0) + \cdots + \frac{1}{n!}\,\mathrm{d}^nf(x_0, y_0)$$
$$+ \frac{1}{(n+1)!}\,\mathrm{d}^{n+1}f(x_0 + \theta\,\Delta x, y_0 + \theta\,\Delta y) \qquad (0 < \theta < 1) \tag{9}$$

gilt, wobei die auf der rechten Seite in verschiedenen Potenzen auftretenden Differentiale $\mathrm{d}x$ und $\mathrm{d}y$ gleich denjenigen Zuwächsen Δx und Δy der unabhängigen Veränderlichen sind, durch die der Zuwachs der Funktion auf der linken Seite erzeugt wird.

Zum Beweis führen wir (wie in Nr. 183) durch

$$x = x_0 + t \cdot \Delta x, \quad y = y_0 + t \cdot \Delta y \quad (0 \leqq t \leqq 1) \tag{10}$$

die neue unabhängige Veränderliche t ein. Setzen wir diese Werte für x und y in die Funktion $f(x, y)$ ein, so erhalten wir eine mittelbare Funktion, die nur von der einzigen Veränderlichen t abhängt:

$$F(t) = f(x_0 + t \cdot \Delta x, y_0 + t \cdot \Delta y).$$

Wir wissen schon, daß Formel (10) geometrisch die Verbindungsstrecke der Punkte $M_0(x_0, y_0)$ und $M_1(x_0 + \Delta x, y_0 + \Delta y)$ bedeutet.

Jetzt sehen wir, daß wir statt des Zuwachses

$$\Delta f(x_0, y_0) = f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0)$$

den Zuwachs der Hilfsfunktion

$$\Delta F(0) = F(1) - F(0)$$

betrachten können, weil beide Zuwächse gleich sind. $F(t)$ ist aber eine Funktion einer Veränderlichen, die nach Nr. 192 sicher $n + 1$ stetige Ableitungen besitzt; folglich gilt (nach der schon früher bewiesenen Taylorschen Formel)

$$\begin{aligned}\Delta F(0) &= F(1) - F(0)\\ &= \mathrm{d}F(0) + \frac{1}{2!}\,\mathrm{d}^2F(0) + \cdots + \frac{1}{n!}\,\mathrm{d}^n F(0) + \frac{1}{(n+1)!}\,\mathrm{d}^{n+1} F(\theta)\\ &(0 < \theta < 1); \end{aligned} \tag{11}$$

dabei ist das rechts in verschiedenen Potenzen auftretende Differential $\mathrm{d}t$ gleich $\Delta t = 1 - 0 = 1$.

Da bei einer linearen Variablensubstitution auch die Differentiale höherer Ordnung invariant sind, kann man schreiben:

$$\mathrm{d}F(0) = f_x(x_0, y_0)\cdot \mathrm{d}x + f_y(x_0, y_0)\cdot \mathrm{d}y = \mathrm{d}f(x_0, y_0),$$

$$\begin{aligned}\mathrm{d}^2F(0) &= f_{x^2}(x_0, y_0)\cdot \mathrm{d}x^2 + 2f_{xy}(x_0, y_0)\cdot \mathrm{d}x\,\mathrm{d}y + f_{y^2}(x_0, y_0)\cdot \mathrm{d}y^2\\ &= \mathrm{d}^2f(x_0, y_0),\end{aligned}$$

usw. Schließlich erhalten wir für das $(n+1)$-te Differential

$$\mathrm{d}^{n+1} F(\theta) = \mathrm{d}^{n+1}f(x_0 + \theta\Delta x, y_0 + \theta\,\Delta y).$$

Es ist wichtig, zu bemerken, daß hier die Differentiale $\mathrm{d}x$ und $\mathrm{d}y$ mit den früher betrachteten Zuwächsen Δx und Δy übereinstimmen; in der Tat ist

$$\mathrm{d}x = \Delta x\cdot \mathrm{d}t = \Delta x, \quad \mathrm{d}y = \Delta y\cdot \mathrm{d}t = \Delta y.$$

Wenn wir dies in (11) einsetzen, so erhalten wir die zu beweisende Formel (9).

Man muß sich jedoch vor Augen halten, daß die Taylorsche Formel zwar in der Differentialform für Funktionen mehrerer Veränderlicher dieselbe einfache Gestalt hat wie für Funktionen einer einzigen Veränderlichen, ausgeschrieben aber viel komplizierter ist. Schreiben wir beispielsweise die ersten drei Glieder für eine Funktion von zwei Veränderlichen aus, so erhalten wir

$$\begin{aligned} f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0) &= [f_x(x_0, y_0)\cdot\Delta x + f_y(x_0, y_0)\cdot\Delta y]\\ &+ \frac{1}{2!}\,[f_{x^2}(x_0, y_0)\cdot\Delta x^2 + 2f_{xy}(x_0, y_0)\cdot\Delta x\,\Delta y + f_{y^2}(x_0, y_0)\cdot\Delta y^2]\\ &+ \frac{1}{3!}\,[f_{x^3}(x_0, y_0)\cdot\Delta x^3 + 3f_{x^2y}(x_0, y_0)\cdot\Delta x^2\,\Delta y\\ &+ 3f_{xy^2}(x_0, y_0)\cdot\Delta x\,\Delta y^2 + f_{y^3}(x_0, y_0)\cdot\Delta y^3] + \cdots. \end{aligned}$$

Die Formel (9) gilt auch für $n = 0$; diesen Spezialfall haben wir schon in Nr. 183 betrachtet.

§ 5. Extremwerte. Größe und kleinste Werte

196. Extremwerte einer Funktion mehrerer Veränderlicher. Notwendige Bedingungen. Die Funktion

$$u = f(x_1, x_2, \ldots, x_n)$$

sei in einem Bereich $\mathcal{D}$ definiert, und $(x_1^0, x_2^0, \ldots, x_n^0)$ sei ein innerer Punkt von $\mathcal{D}$.

Man sagt, $f(x_1, x_2, \ldots, x_n)$ habe in $(x_1^0, x_2^0, \ldots, x_n^0)$ ein (*relatives*) *Maximum* bzw. *Minimum*, wenn man diesen Punkt in eine Umgebung

$$(x_1^0 - \delta, x_1^0 + \delta;\, x_2^0 - \delta, x_2^0 + \delta;\, \ldots;\, x_n^0 - \delta, x_n^0 + \delta)$$

einbetten kann derart, daß für alle Punkte dieser Umgebung die Ungleichung

$$f(x_1, x_2, \ldots, x_n) \leqq f(x_1^0, x_2^0, \ldots, x_n^0) \quad (\text{bzw.} \geqq)$$

gilt.

Kann diese Umgebung so klein gewählt werden, daß Gleichgewicht ausgeschlossen ist, d. h., daß in jedem Punkt dieser Umgebung mit Ausnahme von $(x_1^0, x_2^0, \ldots, x_n^0)$ selbst (man sagt auch: in jedem Punkt dieser punktierten Umgebung) die „echte“ Ungleichung

$$f(x_1, x_2, \ldots, x_n) < f(x_1^0, x_2^0, \ldots, x_n^0) \quad (\text{bzw.} >)$$

gilt, so sagt man, im Punkt $(x_1^0, x_2^0, \ldots, x_n^0)$ liege ein *eigentliches Maximum* (bzw. *eigentliches Minimum*) vor. Anderenfalls heißt das Maximum (bzw. Minimum) *uneigentlich.* Zur gemeinsamen Bezeichnung von Maximum und Minimum wird auch hier das Wort *Extremum* benutzt.[1])

Wir nehmen nun an, unsere Funktion habe im Punkt $(x_1^0, x_2^0, \ldots, x_n^0)$ ein Extremum. Es soll gezeigt werden:

Wenn in diesem Punkt endliche partielle Ableitungen existieren,

$$f_{x_1}(x_1^0, \ldots, x_n^0), \quad \ldots, \quad f_{x_n}(x_1^0, \ldots, x_n^0),$$

so sind sie sämtlich gleich 0, *so daß also das Verschwinden der partiellen Ableitungen erster Ordnung eine notwendige Bedingung für das Vorliegen eines Extremums ist.*

Zum Beweis setzen wir $x_2 = x_2^0, \ldots, x_n = x_n^0$ und lassen nur x_1 veränderlich. Dann erhalten wir eine Funktion von x_1 allein:

$$u = f(x_1, x_2^0, \ldots, x_n^0).$$

Da wir angenommen haben, in $(x_1^0, x_2^0, \ldots, x_n^0)$ liege ein Extremum (etwa ein Maximum) vor, folgt insbesondere, daß in einer gewissen Umgebung $(x_1^0 - \delta, x_1^0 + \delta)$ des Punktes $x_1 = x_1^0$ die Ungleichung

$$f(x_1, x_2^0, \ldots, x_n^0) \leqq f(x_1^0, x_2^0, \ldots, x_n^0)$$

gelten muß, so daß die obige Funktion einer Veränderlichen in $x_1 = x_1^0$ ein Maximum hat. Nach dem Fermatschen Satz (Nr. 109) folgt hieraus

$$f_{x_1}(x_1^0, x_2^0, \ldots, x_n^0) = 0.$$

In derselben Weise kann man zeigen, daß in $(x_1^0, x_2^0, \ldots, x_n^0)$ auch die übrigen partiellen Ableitungen gleich 0 sind.

Die „extremwertverdächtigen“ Punkte sind also diejenigen, in denen die partiellen Ableitungen erster Ordnung verschwinden. Man findet ihre Koordinaten, indem man das folgende Gleichungssystem löst[2]):

$$\left.\begin{aligned} f_{x_1}(x_1, x_2, \ldots, x_n) &= 0, \\ f_{x_2}(x_1, x_2, \ldots, x_n) &= 0, \\ \ldots \quad & \quad . \; . \\ f_{x_n}(x_1, x_2, \ldots, x_n) &= 0. \end{aligned}\right\} \tag{1}$$

[1]) Man verwendet auch hier den Ausdruck „relatives Extremum“.

[2]) Für eine Funktion zweier Veränderlicher, $z = f(x, y)$, die ein totales Differential besitzt, hat die Bedingung $f_x(x, y) = 0$, $f_y(x, y) = 0$ eine einfache geometrische Bedeutung: *Die Tangentialebene* [vgl. Nr. 180, Formel (6)] an *die Fläche* $z = f(x, y)$ *in einem Punkt, in dem ein Extremum vorliegt, ist parallel der* x, y*-Ebene.*

Wie bei Funktionen einer einzigen Veränderlichen wollen wir solche Punkte *stationär* nennen.

Bemerkungen.

I. Die notwendige Bedingung für die Existenz eines Extremums läßt sich im Fall, daß die Funktion ein totales Differential besitzt, kurz auch folgendermaßen schreiben:

$$\mathrm{d}f(x_1, x_2, \ldots, x_n) = 0.$$

Ist nämlich $f_{x_1} = f_{x_2} = \cdots = f_{x_n} = 0$, so gilt für jede Wahl von $\mathrm{d}x_1, \mathrm{d}x_2, \ldots, \mathrm{d}x_n$ stets

$$\mathrm{d}f(x_1, x_2, \ldots, x_n) = f_{x_1} \cdot \mathrm{d}x_1 + f_{x_2} \cdot \mathrm{d}x_2 + \cdots + f_{x_n} \cdot \mathrm{d}x_n = 0.$$

Und umgekehrt: Ist in einem gegebenen Punkt diese Bedingung identisch erfüllt, so sind die partiellen Ableitungen einzeln gleich 0, da die $\mathrm{d}x_1, \mathrm{d}x_2, \ldots, \mathrm{d}x_n$ willkürlich sind.

II. Oft hat die zu untersuchende Funktion $f(x_1, x_2, \ldots, x_n)$ (endliche) partielle Ableitungen im ganzen Bereich; man muß dann die Punkte, in denen die Funktion Extremwerte annimmt, unter den stationären Punkten suchen. Es kann aber vorkommen, daß in einzelnen Punkten einige partielle Ableitungen unendliche Werte annehmen oder gar nicht existieren, während die übrigen gleich 0 sind. Solche Punkte müssen neben den stationären genaugenommen ebenfalls zu den „extremwertverdächtigen" Punkten gerechnet werden (vgl. Nr. 201, Aufgabe 6).

197. Hinreichende Bedingungen (für Funktionen zweier Veränderlicher). Wie auch im Fall einer Funktion einer einzigen Veränderlichen braucht in einem stationären Punkt keineswegs ein Extremum vorzuliegen. Für die einfache Funktion $z = xy$ beispielsweise verschwinden $z_x = y$ und $z_y = x$ gleichzeitig nur in $(0, 0)$, wo $z = 0$ ist. Es ist aber unmittelbar klar, daß die Funktion in jeder Umgebung dieses Punktes sowohl positive als auch negative Werte annimmt, also dort kein Extremum hat. Abb. 92, S. 322, zeigt das hyperbolische Paraboloid $z = xy$; in der Nähe des Ursprungs ist diese Fläche *sattelförmig*, in einer vertikalen Ebene nach oben, in der anderen nach unten gebogen.

Somit entsteht die Frage nach hinreichenden Bedingungen für die Existenz (bzw. Nichtexistenz) eines Extremums in einem stationären Punkt, d. h. nach den Untersuchungen, denen ein solcher „extremwertverdächtiger" Punkt zusätzlich unterworfen werden muß.

Wir beschränken uns zunächst auf Funktionen zweier Veränderlicher. Eine solche Funktion $f(x, y)$ sei in der Umgebung eines stationären Punktes (x_0, y_0) definiert und stetig und habe dort stetige partielle Ableitungen erster und zweiter Ordnung. Es ist also

$$f_x(x_0, y_0) = 0, \quad f_y(x_0, y_0) = 0. \tag{1a}$$

Um festzustellen, ob die Funktion in (x_0, y_0) wirklich ein Extremum hat oder nicht, geht man naturgemäß auf die Differenz

$$\Delta = f(x, y) - f(x_0, y_0)$$

zurück. Wir entwickeln sie nach Taylor (Nr. 195), beschränken uns dabei aber auf zwei Glieder. Übrigens fällt, da (x_0, y_0) ein stationärer Punkt sein sollte, das erste Glied weg, und es ist somit

$$\Delta = \frac{1}{2!}\{f_{xx} \cdot \Delta x^2 + 2 f_{xy} \cdot \Delta x\, \Delta y + f_{yy} \cdot \Delta y^2\}. \tag{2}$$

Dabei bedeuten Δx, Δy die Differenzen $x - x_0$, $y - y_0$; sämtliche Ableitungen sind in ein und demselben Punkt

$$(x_0 + \theta\,\Delta x, y_0 + \theta\,\Delta y) \quad (0 < \theta < 1)$$

zu berechnen. Wir setzen nun

$$a_{11} = f_{xx}(x_0, y_0), \quad a_{12} = f_{xy}(x_0, y_0), \quad a_{22} = f_{yy}(x_0, y_0) \tag{3}$$

und

$$f_{xx}(x_0 + \theta\,\Delta x, y_0 + \theta\,\Delta y) = a_{11} + \alpha_{11},$$

$$f_{xy}(\ldots) = a_{12} + \alpha_{12}, \quad f_{yy}(\ldots) = a_{22} + \alpha_{22},$$

so daß auf Grund der Stetigkeit der zweiten Ableitungen für $\Delta x \to 0$, $\Delta y \to 0$ auch $\alpha_{ik} \to 0$ gilt. Wir können nun Δ in der Gestalt

$$\Delta = \frac{1}{2}\{a_{11}\Delta x^2 + 2a_{12}\Delta x\,\Delta y + a_{22}\,\Delta y^2 + \alpha_{11}\,\Delta x^2 + 2\alpha_{12}\,\Delta x\,\Delta y + \alpha_{22}\,\Delta y^2\}$$

schreiben. Wir werden zeigen, daß das Verhalten von Δ wesentlich von dem Vorzeichen des Ausdrucks $a_{11}a_{22} - a_{12}^2$ abhängt.

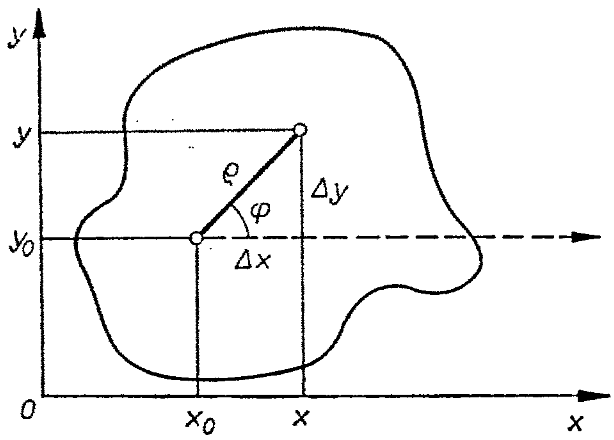

Abb. 105

Zur Vereinfachung der Überlegungen führen wir „Polarkoordinaten" ein, nehmen (x_0, y_0) als Pol und ziehen durch diesen Punkt eine Parallele zur x-Achse als Polarachse (Abb. 105). Es sei $\varrho = \sqrt{\Delta x^2 + \Delta y^2}$ der Abstand zwischen (x_0, y_0) und (x, y), während φ den Winkel bezeichnet, den ϱ mit der Achse bildet. Mit

$$\Delta x = \varrho \cos\varphi, \quad \Delta y = \varrho \sin\varphi \tag{4}$$

läßt sich Δ in der Gestalt

$$\Delta = \frac{\varrho^2}{2}\{a_{11}\cos^2\varphi + 2a_{12}\cos\varphi\cdot\sin\varphi + a_{22}\sin^2\varphi$$
$$+ \alpha_{11}\cos^2\varphi + 2\alpha_{12}\cos\varphi\cdot\sin\varphi + \alpha_{22}\sin^2\varphi\}$$

schreiben.

1. Es sei zunächst $a_{11}a_{22} - a_{12}^2 > 0$.

Dann ist $a_{11}a_{22} > 0$, also $a_{11} \neq 0$, und das erste Trinom in der geschweiften Klammer läßt sich folgendermaßen darstellen:

$$\frac{1}{a_{11}}\cdot[(a_{11}\cos\varphi + a_{12}\sin\varphi)^2 + (a_{11}a_{22} - a_{12}^2)\cdot\sin^2\varphi]. \tag{5}$$

Offenbar ist der Ausdruck in der eckigen Klammer stets positiv, so daß das erwähnte Trinom für alle von 0 verschiedenen Werte von φ das Vorzeichen von a_{11} besitzt. Sein absoluter Betrag nimmt, da er in $[0, 2\pi]$ eine stetige Funktion von φ ist, einen kleinsten (offenbar positiven) Wert m an (vgl. Nr. 85):

$$|a_{11} \cos^2 \varphi + 2a_{12} \cos \varphi \cdot \sin \varphi + a_{22} \sin^2 \varphi| \geqq m > 0 .$$

Für das zweite Trinom in der geschweiften Klammer gilt nach (4)

$$|\alpha_{11} \cos^2 \varphi + 2\alpha_{12} \cos \varphi \cdot \sin \varphi + \alpha_{22} \sin^2 \varphi| \leqq |\alpha_{11}| + 2|\alpha_{12}| + |\alpha_{22}| < m$$

für alle φ, sobald ϱ (und mit ϱ auch Δx und Δy) hinreichend klein ist. Dann hat aber der ganze Ausdruck in der geschweiften Klammer, also auch Δ, dasselbe Vorzeichen wie das erste Trinom, also das Vorzeichen von a_{11}.

Ist also $a_{11} > 0$, so ist auch $\Delta > 0$, d. h., die Funktion hat in dem Punkt (x_0, y_0) ein Minimum; für $a_{11} < 0$ ist $\Delta < 0$, d. h., es liegt ein Maximum vor.

2. Es sei jetzt $a_{11}a_{22} - a_{12}^2 < 0$.

Wir gehen zunächst auf den Fall ein, daß $a_{11} \neq 0$ ist. Dann kann man auch hier die Schreibweise (5) benutzen. Für $\varphi = \varphi_1 = 0$ ist der Ausdruck in der eckigen Klammer positiv, denn er ist dann gleich a_{11}^2. Bestimmt man dagegen $\varphi = \varphi_0$ aus

$$a_{11} \cos \varphi_2 + a_{12} \sin \varphi_2 = 0 \quad (\sin \varphi_2 \neq 0),$$

so wird dieser Ausdruck zu $(a_{11}a_{22} - a_{12}^2) \sin^2 \varphi_2$, also negativ. Für hinreichend kleines ϱ wird das zweite Trinom in der geschweiften Klammer sowohl für $\varphi = \varphi_1$ als auch für $\varphi = \varphi_2$ beliebig klein, so daß das Vorzeichen von Δ gleich dem des ersten Trinoms wird. Somit hat in beliebiger Nähe des betrachteten Punktes (x_0, y_0) — auf Strahlen, die durch die Winkel $\varphi = \varphi_1$ und $\varphi = \varphi_2$ bestimmt werden — die Differenz Δ Werte entgegengesetzten Vorzeichens. Daher kann in diesem Punkt kein Extremum vorliegen.

Ist $a_{11} = 0$ und reduziert sich das erste Trinom in der geschweiften Klammer auf

$$2a_{12} \cos \varphi \cdot \sin \varphi + a_{22} \sin^2 \varphi = \sin \varphi \cdot (2a_{12} \cos \varphi + a_{22} \sin \varphi),$$

so kann man, da sicher $a_{12} \neq 0$ ist, einen Winkel $\varphi_1 \neq 0$ so bestimmen, daß

$$|a_{22}| \cdot |\sin \varphi_1| < 2|a_{12}| \cdot |\cos \varphi_1|$$

ist. Dann besitzt das erwähnte Trinom für $\varphi = \varphi_1$ und $\varphi = \varphi_2 = -\varphi_1$ entgegengesetzte Vorzeichen, und man kann wie oben schließen.

Ist also $a_{11}a_{22} - a_{12}^2 > 0$, so hat in einem stationären Punkt (x_0, y_0) die Funktion $f(x, y)$ ein Extremum, und zwar ein eigentliches Maximum für $a_{11} < 0$ und ein eigentliches Minimum für $a_{11} > 0$. Ist aber $a_{11}a_{22} - a_{12}^2 < 0$, so liegt kein Extremum vor.

Im Fall $a_{11}a_{22} - a_{12}^2 = 0$ muß man zur Entscheidung der Frage, ob ein Extremum vorliegt, höhere Ableitungen heranziehen. Diesen Fall wollen wir hier aber nicht behandeln.

Beispiele.

1. Man untersuche die Funktion

$$z = \frac{x^2}{2p} + \frac{y^2}{2q} \quad (p > 0, q > 0)$$

auf Maxima und Minima.

Offenbar ist

$$z_x = \frac{x}{p}, \quad z_y = \frac{y}{p}.$$

Hieraus folgt, daß der einzige stationäre Punkt der Ursprung ist.

Wir berechnen a_{11}, a_{12}, a_{22} und finden

$$a_{11} = \frac{1}{p}, \quad a_{12} = 0, \quad a_{22} = \frac{1}{q},$$

also $a_{11}a_{22} - a_{12}^2 > 0$. Daher hat in $(0, 0)$ die Funktion z ein Minimum. Übrigens erkennt man das auch unmittelbar.

Das Bild unserer Funktion ist ein elliptisches Paraboloid mit dem Scheitel im Ursprung (vgl. Abb. 93, S. 323).

1. Ist $z = \frac{x^2}{2p} - \frac{z^2}{2q}$ $(p > 0, q > 0)$, so folgt

$$z_x = \frac{x}{p}, \quad z_y = \frac{-y}{q}.$$

Auch hier ist der Punkt $(0, 0)$ stationär.

Wir erhalten

$$a_{11} = \frac{1}{p}, \quad a_{12} = 0, \quad a_{22} = -\frac{1}{q};$$

daher ist $a_{11}a_{22} - a_{12}^2 < 0$, und es liegt kein Extremum vor. Das geometrische Bild dieser Funktion ist das hyperbolische Paraboloid mit dem Scheitel in $(0, 0)$.

3. $z = y^2 + x^4$ bzw. $z = y^2 + x^3$.

In beiden Fällen ist $(0, 0)$ ein stationärer Punkt (und zwar der einzige); dort ist $a_{11}a_{22} - a_{12}^2 = 0$, so daß unser Kriterium versagt. Man sieht aber sofort, daß im ersten Fall ein Minimum, im zweiten überhaupt kein Extremum vorliegt.

Bemerkung. Wir werden in Nr. 236 sehen, daß diese Resultate eng mit dem geometrischen Problem zusammenhängen, wie sich eine Kurve in der Nähe eines „singulären" Punktes verhält.

198. Hinreichende Bedingungen (der allgemeine Fall). Wir wollen jetzt den allgemeinen Fall untersuchen. Die Funktion $f(x_1, x_2, \ldots, x_n)$ sei in der Nähe eines stationären Punktes $(x_1^0, x_2^0, \ldots, x_n^0)$ definiert und stetig und habe dort stetige partielle Ableitungen erster und zweiter Ordnung. Wir betrachten die Differenz

$$\Delta = f(x_1, x_2, \ldots, x_n) - f(x_1^0, x_2^0, \ldots, x_n^0)$$

und entwickeln sie nach der Taylorschen Formel. Wie oben erhalten wir

$$\begin{aligned}\Delta &= \frac{1}{2}\{f_{x_1x_1}\Delta x_1^2 + f_{x_2x_2}\Delta x_2^2 + \cdots + f_{x_nx_n}\Delta x_n^2 \\ &\qquad + 2f_{x_1x_2}\Delta x_1\,\Delta x_2 + 2f_{x_1x_3}\Delta x_1\,\Delta x_3 + \cdots + 2f_{x_{n-1}x_n}\Delta x_{n-1}\,\Delta x_n\} \\ &= \frac{1}{2}\sum_{i,k=1}^{n} f_{x_ix_k}\Delta x_i\,\Delta x_k,\end{aligned}$$

wobei $\Delta x_i = x_i - x_i^0$ ist. Alle Ableitungen sind im Punkt

$$(x_1^0 + \theta\,\Delta x_1, x_2^0 + \theta\,\Delta x_2, \ldots, x_n^0 + \theta\,\Delta x_n) \quad (0 < \theta < 1)$$

zu berechnen.

Wir führen auch hier die Größen

$$f_{x_i x_k}(x_1^0, x_2^0, \ldots, x_n^0) = a_{ik} \quad (i, k = 1, 2, \ldots, n) \tag{6}$$

ein, so daß

$$f_{x_i x_k}(x_1^0 + \theta\,\Delta x_1, \ldots, x_n^0 + \theta\,\Delta x_n) = a_{ik} + \alpha_{ik}$$

und

$$\alpha_{ik} \to 0 \quad \text{für} \quad \Delta x_1 \to 0, \ldots, \Delta x_n \to 0 \tag{7}$$

gilt.[1]) Den kritischen Ausdruck Δ können wir jetzt in der Gestalt

$$\Delta = \frac{1}{2}\left\{\sum_{i,k=1}^{n} a_{ik}\,\Delta x_i\,\Delta x_k + \sum_{i,k=1}^{n} \alpha_{ik}\,\Delta x_i\,\Delta x_k\right\} \tag{8}$$

schreiben.

Die erste Summe ist das zweite Differential von f in dem stationären Punkt; sie ist ein homogenes Polynom zweiten Grades oder, wie man sagt, eine *quadratische Form* in den Veränderlichen $\Delta x_1, \ldots, \Delta x_n$. Die zweite Summe hat eine ähnliche Gestalt, jedoch sind hier auch die Koeffizienten selbst Funktionen derselben Veränderlichen. Die Lösung unseres Problems hängt nun, wie wir alsbald sehen werden, gerade von den Eigenschaften dieser quadratischen Form ab.

In der Algebra nennt man eine quadratische Form

$$\sum_{i,k=1}^{n} a_{ik} y_i y_k \quad (a_{ik} = a_{ki}) \tag{9}$$

in den Veränderlichen $y_1, \ldots, y_n$ *positiv* (bzw. *negativ*) *definit*, wenn sie für alle Werte der Argumente nur positiver (bzw. negativer) Werte fähig ist (natürlich sollen nicht alle Argumente gleichzeitig 0 werden). Beispielsweise ist die Form

$$6y_1^2 + 5y_2^2 + 14y_3^2 + 4y_1y_2 - 8y_1y_3 - 2y_2y_3$$

positiv definit. Das sieht man sofort, wenn man sie in der Gestalt

$$(2y_1 - 3y_3)^2 + 2(y_1 - y_2 + y_3)^2 + 3(y_2 - y_3)^2$$

schreibt.

Wir können auf diese Dinge nicht näher eingehen, sondern hier nur eine von J. J. Sylvester[2]) stammende notwendige und hinreichende Bedingung dafür angeben, daß eine Form (9) positiv definit ist. Sie besteht darin, daß die Ungleichungen

$$a_{11} > 0, \qquad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \qquad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} > 0, \qquad \ldots, \qquad \begin{vmatrix} a_{11} & a_{12} \ldots a_{1n} \\ a_{21} & a_{22} \ldots a_{2n} \\ \cdot & \cdot \\ a_{n1} & a_{n2} \ldots a_{nn} \end{vmatrix} > 0$$

gelten.[3])

[1]) Offenbar ist $a_{ik} = a_{ki}$ und $\alpha_{ik} = \alpha_{ki}$.

[2]) James Joseph Sylvester, 1814—1897, englisch-amerikanischer Mathematiker.

[3]) Wir machen darauf aufmerksam, daß das Glied mit $y_i y_k$ $(i \neq k)$ in der Summe (9) zweimal vorkommt, so daß $a_{ik} = a_{ki}$ die Hälfte des Koeffizienten von $y_i y_k$ ist. Für unser Beispiel läßt sich die Bedingung leicht verifizieren, wenn man berücksichtigt, daß $a_{11} = 6$, $a_{22} = 5$, $a_{33} = 14$, $a_{12} = a_{21} = 2$, $a_{13} = a_{31} = -4$, $a_{23} = a_{32} = -1$ ist.

Wenn man in einer negativ definiten Form die Vorzeichen aller Glieder ändert, so wird sie positiv definit, und umgekehrt. Daher läßt sich leicht ein Kriterium dafür angeben, wann eine Form negativ definit ist: Es müssen die Ungleichungen gelten, die man aus den obigen dadurch erhält, daß man, beginnend mit der ersten, jedes zweite Ungleichheitszeichen umdreht.

Nach diesen Bemerkungen können wir folgende hinreichenden Bedingungen für das Vorliegen eines Extremums formulieren:

Ist das zweite Differential, d. h. die quadratische Form

$$\sum_{i,k=1}^{n} a_{ik}\,\Delta x_i\,\Delta x_k, \tag{10}$$

mit den Werten (6) *der Koeffizienten eine positiv* (*bzw. negativ*) *definite Form, so liegt in dem stationären Punkt* $(x_1^0, \ldots, x_n^0)$ *ein eigentliches Minimum* (*Maximum*) *vor.*

Zum Beweis führen wir den Abstand

$$\varrho = \sqrt{\Delta x_1^2 + \cdots + \Delta x_n^2}$$

zwischen den Punkten $(x_1^0, \ldots, x_n^0)$ und $(x_1, \ldots, x_n)$ ein. Wir ziehen in (8) die Größe ϱ^2 vor die Klammer und setzen $\dfrac{\Delta x_i}{\varrho} = \xi_i$ $(i = 1, 2, \ldots, n)$; dann können wir Δ in der Gestalt

$$\Delta = \frac{\varrho^2}{2}\left\{\sum_{i,k=1}^{n} a_{ik}\xi_i\xi_k + \sum_{i,k=1}^{n} \alpha_{ik}\xi_i\xi_k\right\} \tag{11}$$

schreiben.

Die Zahlen ξ_i werden nicht gleichzeitig 0; daher ist, wenn die Form (10) positiv ist, die erste Summe in (11) stets positiv. Überdies läßt sich wegen

$$\sum_{i=1}^{n} \xi_i^2 = 1 \tag{12}$$

stets eine konstante positive Zahl m finden derart, daß für alle möglichen Werte ξ_i

$$\sum_{i,k=1}^{n} a_{ik}\xi_i\xi_k \geqq m$$

gilt. Diese Summe ist nämlich eine stetige Funktion der Argumente ξ_i im ganzen Raum, insbesondere also auch auf der Menge $\mathcal{M}$ derjenigen Punkte $(\xi_1, \ldots, \xi_n)$, welche der Bedingung (12) genügen (also einer n-dimensionalen Kugelfläche angehören). Diese Menge ist aber offenbar abgeschlossen, d. h., sie enthält ihre Häufungspunkte. Dann nimmt aber die Summe nach dem Weierstraßschen Satz (vgl. die Bemerkung im Anschluß an den Beweis in Nr. 173) in $\mathcal{M}$ ihren kleinsten Wert m an, der, wie alle Werte in $\mathcal{M}$, notwendigerweise positiv ist.

Andererseits wird wegen (7) die zweite Summe in (11) für hinreichend kleine ϱ offenbar dem absoluten Betrag nach kleiner als m, so daß die ganze Klammer positiv ist. In einer hinreichend kleinen Kugel um den Punkt $(x_1^0, \ldots, x_n^0)$ ist also die Differenz Δ positiv, und hieraus folgt, daß in diesem Punkt die Funktion $f(x_1, \ldots, x_n)$ ein eigentliches Minimum hat.

Analog behandelt man den Fall, daß die Form (10) negativ definit ist.

199. Bedingungen dafür, daß kein Extremum vorliegt. Die quadratische Form (9) heißt *indefinit*, wenn sie sowohl positiver als auch negativer Werte fähig ist. Die Form

$$6y_1^2 + y_2^2 + y_3^2 + 8y_1y_2 - 8y_1y_3 - 2y_2y_3$$

beispielsweise ist indefinit; sie ist nämlich für $y_1 = 1$, $y_2 = y_3 = 0$ gleich 6 und für $y_1 = 1$, $y_2 = -1$, $y_3 = 0$ gleich -1.

Jetzt kann man die Aussage aus Nr. 198 folgendermaßen ergänzen:

Ist eine quadratische Form (10) *indefinit, so liegt im Punkt* $(x_1^0, \ldots, x_n^0)$ *bestimmt kein Extremum vor.*

Beweis. Die Form (10) nehme für $\Delta x_i = h_i$ $(i = 1, 2, \ldots, n)$ einen positiven Wert an,

$$\sum_{i,k=1}^{n} a_{ik} h_i h_k > 0, \tag{13}$$

und für $\Delta x_i = \bar{h}_i$ $(i = 1, 2, \ldots, n)$ einen negativen,

$$\sum_{i,k=1}^{n} a_{ik} \bar{h}_i \bar{h}_k < 0.$$

Wir setzen zunächst

$$\Delta x_i = h_i t \quad \text{für} \quad t \neq 0 \quad (i = 1, 2, \ldots, n),$$

was einer Bewegung längs der Geraden entspricht, welche die Punkte $(x_1^0, \ldots, x_n^0)$ und $(x_1^0 + h_1, \ldots, x_n^0 + h_n)$ verbindet. Setzt man in (8) die Größe t^2 vor die Klammer, so erhält man für diesen Fall

$$\Delta = \frac{t^2}{2} \left\{ \sum_{i,k=1}^{n} a_{ik} h_i h_k + \sum_{i,k=1}^{n} \alpha_{ik} h_i h_k \right\}.$$

Die erste Summe in der Klammer ist eine wegen (13) positive Zahl. In der zweiten Summe streben mit $t \to 0$ die Koeffizienten und alle Δx_i gegen 0. Das besagt, daß für hinreichend kleine t der Ausdruck in der geschweiften Klammer (und damit auch die ganze Differenz Δ) positiv wird, d. h., in den Punkten der oben erwähnten Geraden, die hinreichend nahe bei $(x_1^0, \ldots, x_n^0)$ liegen, gilt

$$f(x_1, \ldots, x_n) > f(x_1^0, \ldots, x_n^0).$$

Nimmt man andererseits

$$\Delta x_i = \bar{h}_i t \quad \text{für} \quad t \neq 0 \quad (i = 1, 2, \ldots, n),$$

d. h., bewegt man sich längs einer anderen Geraden, die den Punkt $(x_1^0, \ldots, x_n^0)$ mit $(x_1^0 + \bar{h}_1, \ldots, x_n^0 + \bar{h}_n)$ verbindet, so ist in ihren Punkten, die hinreichend nahe bei $(x_1^0, \ldots, x_n^0)$ liegen (d. h. hinreichend kleinen t entsprechen),

$$f(x_1, \ldots, x_n) < f(x_1^0, \ldots, x_n^0).$$

Damit ist bewiesen, daß in dem betreffenden Punkt weder ein Maximum noch ein Minimum vorliegen kann.

Es kann vorkommen, daß die Form (9) nicht nur dann 0 wird, wenn alle Argumente gleichzeitig 0 werden, sonst aber keine Werte verschiedenen Vorzeichens annimmt. In diesem Fall nennt man sie *semidefinit*. Das gilt beispielsweise für die Form

$$y_1^2 + y_2^2 + y_3^2 + 2y_1y_2 + 2y_1y_3 + 2y_2y_3 = (y_1 + y_2 + y_3)^2;$$

sie nimmt zwar keine negativen Werte an, jedoch den Wert 0 immer dann, wenn

$$y_1 + y_2 + y_3 = 0$$

ist, etwa für $y_1 = y_2 = \frac{1}{2}$ und $y_3 = -1$.

Wenn die Form (10) semidefinit ist, kann man nach unserem Kriterium noch nicht zu einer Entscheidung kommen. Je nach dem Verhalten der höheren Ableitungen kann ein Extremum vorliegen oder auch nicht. Insbesondere muß man die höheren Ableitungen auch heranziehen, wenn alle Ableitungen zweiter Ordnung in dem betreffenden stationären Punkt gleich 0 sind.

Diese Zweifelsfälle werden wir hier jedoch nicht näher untersuchen.

Bemerkung. Für eine Funktion $f(x)$ einer einzigen Veränderlichen reduziert sich die Form (10) auf das einzige Glied

$$f''(x_0) \cdot \Delta x^2,$$

wobei x_0 der stationäre Punkt ist. Diese „Form" ist offenbar definit, und zwar positiv für $f''(x_0) > 0$ und negativ für $f''(x_0) < 0$. Somit ist das Kriterium aus Nr. 137 ein Spezialfall desjenigen aus Nr. 198.

Gehen wir zu einer Funktion $f(x, y)$ zweier Veränderlicher über, so zeigt sich, daß das Ergebnis aus Nr. 197 ebenfalls in dem aus Nr. 198 und Nr. 199 enthalten ist. Man sieht leicht, daß in Nr. 197 beiläufig folgendes bewiesen wurde: Die Form

$$a_{11}\, \Delta x^2 + 2a_{12}\, \Delta x\, \Delta y + a_{22}\, \Delta y^2$$

ist für $a_{11}a_{22} - a_{12}^2 > 0$ definit (und zwar positiv für $a_{11} > 0$ und negativ für $a_{11} < 0$), für $a_{11}a_{22} - a_{12}^2 < 0$ dagegen indefinit.

200. Größte und kleinste Werte einer Funktion. Beispiele. Die Funktion $u = f(x_1, x_2, \ldots, x_n)$ sei in einem beschränkten abgeschlossenen Bereich $\mathcal{D}$ definiert und stetig und habe dort mit eventueller Ausnahme einzelner Punkte endliche partielle Ableitungen. Nach dem Weierstraßschen Satz (Nr. 173) gibt es in diesem Bereich einen Punkt $(x_1^0, x_2^0, \ldots, x_n^0)$, in dem die Funktion ihren größten (bzw. kleinsten) Wert annimmt. Liegt $(x_1^0, x_2^0, \ldots, x_n^0)$ im Innern von $\mathcal{D}$, so hat die Funktion dort offenbar ein Maximum (bzw. Minimum), so daß in diesem Fall der uns interessierende Punkt von vornherein einer der extremwertverdächtigen Punkte ist. Jedoch kann eine Funktion u ihren größten (bzw. kleinsten) Wert auch auf dem Rand des Bereichs annehmen. (Daher stammt die Bezeichnung „relatives Extremum"; hierbei vergleicht man Werte der unmittelbaren Umgebung. Zieht man alle Werte zum Vergleich heran, so spricht man vom „absoluten Extremum".)

Wenn man also den größten (bzw. kleinsten) Wert einer Funktion $u = f(x_1, x_2, \ldots, x_n)$ in einem Bereich $\mathcal{D}$ bestimmen will, muß man alle ihre extremwertverdächtigen Punkte im Innern von $\mathcal{D}$ aufsuchen, dort die Funktionswerte berechnen und sie mit den Werten der Funktion auf dem Rand von $\mathcal{D}$ vergleichen; der größte (bzw. kleinste) Wert unter allen diesen Werten ist dann der größte (kleinste) Wert der Funktion in dem ganzen Bereich $\mathcal{D}$.

Wir wollen das an Beispielen erläutern.

1. Man bestimme den größten Wert der Funktion

$$u = \sin x + \sin y - \sin(x + y)$$

in dem von der x-Achse, der y-Achse und der Geraden $x + y = 2\pi$ begrenzten Dreieck (Abb. 106).

Es ist

$$u_x = \cos x - \cos(x + y), \qquad u_y = \cos y - \cos(x + y).$$

Im Innern des Bereichs verschwinden diese Ableitungen in dem einzigen Punkt $\left(\frac{2\pi}{3}, \frac{2\pi}{3}\right)$, in welchem $u = \frac{3\sqrt{3}}{2}$ ist. Da die Funktion u auf dem Rand, d. h. auf den Geraden $x = 0$, $y = 0$, $x + y = 2\pi$, gleich 0 ist, nimmt sie offenbar in $\left(\frac{2\pi}{3}, \frac{2\pi}{3}\right)$ ihren größten Wert an.

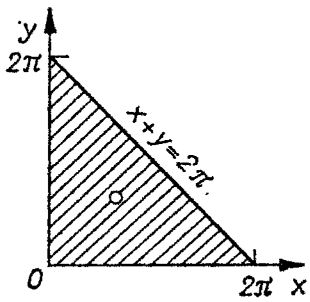

Abb. 106

2. Man bestimme die größten und die kleinsten Werte der Funktion

$$u = a^2x^2 + b^2y^2 + c^2z^2 - (ax^2 + by^2 + cz^2)^2,$$

wenn x, y, z durch die Beziehung $x^2 + y^2 + z^2 = 1$ verknüpft sind (ferner soll $a > b > c > 0$ sein).

Berechnen wir hieraus z^2 und setzen den Ausdruck in u ein, so gelangen wir zu der Funktion

$$u = (a^2 - c^2)\,x^2 + (b^2 - c^2)\,y^2 + c^2 - [(a - c)\,x^2 + (b - c)\,y^2 + c]^2$$

der Veränderlichen x und y im Kreis $x^2 + y^2 \leqq 1$.

Die partiellen Ableitungen

$$u_x = 2x(a - c)\,\{(a + c) - 2[(a - c)\,x^2 + (b - c)\,y^2 + c]\},$$
$$u_y = 2y(b - c)\,\{(b + c) - 2[(a - c)\,x^2 + (b - c)\,y^2 + c]\}$$

verschwinden gleichzeitig in den Punkten

(1) $\quad x = y = 0 \quad (u = 0),$

(2) $\quad x = 0, \quad y = \pm\frac{1}{\sqrt{2}} \quad \left(u = \frac{1}{4}(b - c)^2\right),$

(3) $\quad x = \pm\frac{1}{\sqrt{2}}, \quad y = 0 \quad \left(u = \frac{1}{4}(a - c)^2\right).$

Jetzt müssen die Werte auf dem Rand, d. h. auf dem Kreis $x^2 + y^2 = 1$, untersucht werden. Bestimmen wir hieraus y^2 und setzen den Ausdruck in u ein, so erhalten wir eine Funktion einer einzigen Veränderlichen:

$$u = (a^2 - b^2)\,x^2 + b^2 - [(a - b)\,x^2 + b]^2$$

im Intervall $[-1, 1]$. Im Innern dieses Intervalls verschwindet die Ableitung

$$u_x = 2(a - b)^2\,x(1 - 2x^2)$$

in den Punkten

(4) $\quad x = 0 \quad (u = 0),$

(5) $\quad x = \pm\frac{1}{\sqrt{2}} \quad \left(u = \frac{1}{4}(a - b)^2\right).$

Schließlich untersuchen wir noch die Endpunkte des Intervalls

(6) $\qquad x = \pm 1 \quad (u = 0).$

Es sind also die Werte $u = 0$, $\frac{1}{4}(b-c)^2$, $\frac{1}{4}(a-c)^2$, $\frac{1}{4}(a-b)^2$ zu vergleichen. Der kleinste ist 0, der größte $\frac{1}{4}(a-c)^2$. Das sind auch die gesuchten kleinsten bzw. größten Werte der Funktion, die in den Punkten

$$(0, 0, \pm 1), \quad (0, \pm 1, 0), \quad (\pm 1, 0, 0), \quad \left(\pm \frac{1}{\sqrt{2}}, 0, \pm \frac{1}{\sqrt{2}}\right)$$

angenommen werden.

Im allgemeinen ist bei einer Funktion $f(x, y)$ zweier Veränderlicher der Bereich von einer Kurve (oder von mehreren Kurven) begrenzt; längs dieser Kurve (oder längs jeder dieser Kurven, wenn es mehrere sind) hängen die Veränderlichen x, y entweder voneinander oder beide von einem einzigen Parameter ab, so daß auf dem Rande des Bereichs unsere Funktion $f(x, y)$ nur von einer einzigen Veränderlichen abhängt; daher läßt sich dort ihr größter (bzw. kleinster) Wert nach den Methoden aus Nr. 139 bestimmen. Ist die Kurve etwa in der Parameterdarstellung $x = \varphi(t)$, $y = \psi(t)$ gegeben, wobei t im Intervall $[t_0, T]$ variiert, so ist längs dieser Kurve die Funktion u eine mittelbare Funktion von t,

$$u = f(\varphi(t), \psi(t)),$$

deren größten (bzw. kleinsten) Wert wir bestimmen können.

3. Man bestimme den größten Wert des Produktes $u = xyzt$ der nichtnegativen Zahlen x, y, z, t unter der Bedingung, daß ihre Summe $x + y + z + t = 4c$ konstant ist. Wir zeigen, daß dieses Produkt[1]) am größten ist, wenn alle Faktoren gleich sind, also für $x = y = z = t = c$.

Wir bestimmen t aus der Summe und setzen es in u ein:

$$u = xyz(4c - x - y - z).$$

Dann ist u eine Funktion dreier unabhängiger Veränderlicher x, y, z in dem dreidimensionalen Bereich $x \geqq 0$, $y \geqq 0$, $z \geqq 0$, $x + y + z \leqq 4c$. Dieser Bereich ist das Tetraeder, dessen Seitenflächen die Ebenen $x = 0$, $y = 0$, $z = 0$, $x + y + z = 4c$ sind.

Wir berechnen die partiellen Ableitungen und setzen sie gleich 0:

$$\frac{\partial u}{\partial x} = yz(4c - 2x - y - z) = 0,$$

$$\frac{\partial u}{\partial y} = zx(4c - x - 2y - z) = 0,$$

$$\frac{\partial u}{\partial z} = xy(4c - x - y - 2z) = 0.$$

Im Innern des Bereichs gelten diese Gleichungen nur für den Punkt $x = y = z = c$, in dem $u = c^4$ ist. Da auf dem Rand $u = 0$ ist, nimmt u in diesem Punkt wirklich den größten Wert an. Unsere Behauptung ist damit bewiesen; denn für $x = y = z = c$ ist auch $t = c$.[2])

---

[1]) Nur um etwas Bestimmtes vor Augen zu haben, wählten wir hier vier Faktoren; das Resultat gilt für beliebig viele.

[2]) Hieraus folgt, daß das Produkt $xyzt$ positiver Zahlen, deren Summe gleich $4c$ ist, nicht größer als c^4 wird, so daß

$$\sqrt[4]{xyzt} \leqq c = \frac{x + y + z + t}{4}$$

ist: *Das geometrische Mittel ist höchstens gleich dem arithmetischen Mittel.* Dieses Resultat, das für beliebig viele Zahlen gilt, ist uns bereits aus Nr. 133 [vgl. dort Formel (4a)] bekannt.

Im allgemeinen ist bei einer Funktion $u = f(x, y, z)$ dreier Veränderlicher der Bereich von einer Fläche (oder von mehreren Flächen) begrenzt. Längs dieser Fläche(n) hängen x, y, z nur noch von zwei Parametern ab (als solche können — wie in unserem Beispiel — zwei dieser Veränderlichen genommen werden: $z = 4c - x - y$). Dann hängt auch die Funktion u nur von zwei Parametern ab, so daß die Bestimmung ihres größten (kleinsten) Wertes auf dem Rand schon einfacher ist als das Ausgangsproblem.

Ist eine Funktion $u = f(x_1, x_2, \ldots, x_n)$ nur in einem offenen oder einem unbeschränkten Bereich $\mathfrak{D}$ gegeben, so kann man nicht von vornherein behaupten, daß sie in dem Bereich ihren größten (kleinsten) Wert annimmt. Trotzdem kann in einzelnen Fällen ein solcher Wert existieren. Wir zeigen an einem Beispiel, wie man dann verfährt.

4. Man bestimme den kleinsten Wert der Summe $u = x + y + z + t$ positiver Zahlen x, y, z, t, deren Produkt $xyzt = c^4$ konstant ist.

Wir zeigen, daß sich dieser kleinste Wert für $x = y = z = t = c$ ergibt.[1])

Wir setzen $t = \dfrac{c^4}{xyz}$, also

$$u = x + y + z + \frac{c^4}{xyz}.$$

Wir wollen den kleinsten Wert dieser Funktion dreier Veränderlicher im ersten (offenen und unbeschränkten) Oktanten bestimmen. Wir versuchen die obige Methode anzuwenden: Gibt es im Innern des Bereichs einen Punkt, in dem die Funktion einen kleinsten Wert annimmt, so ist er stationär. Aus

$$u_x = 1 - \frac{c^4}{x^2yz} = 0, \quad u_y = 1 - \frac{c^4}{xy^2z} = 0, \quad u_z = 1 - \frac{c^4}{xyz^2} = 0$$

folgt $x = y = z = c$, also $t = c$ und $u = 4c$.

Wie kann man nun feststellen, daß es sich wirklich um den kleinsten Wert handelt?

Offenbar wächst μ unbegrenzt, wenn man sich den Ebenen $x = 0$, $y = 0$, $z = 0$ nähert bzw. wenn man ins Unendliche fortschreitet. Den vorhin gefundenen Punkt kann man in einen Würfel $[\varepsilon, E; \varepsilon, E; \varepsilon, E]$ einbetten, wobei $\varepsilon > 0$ so klein und $E > 0$ so groß gewählt werde, daß außerhalb dieses Würfels und auf seinen Flächen $u > 4c$ ist. In diesem Würfel, der ein abgeschlossener und beschränkter Bereich ist, hat u einen kleinsten Wert. Jetzt ist aber klar, daß dieser kleinste Wert genau in dem oben gefundenen Punkt angenommen wird und daß dies auch der kleinste Wert in dem ganzen ursprünglichen Bereich ist, was zu beweisen war.

Bemerkung. In den Beispielen 1, 3 und 4 existierte im Innern des zu untersuchenden Bereichs nur ein einziger extremwertverdächtiger Punkt. Man könnte sich davon überzeugen, daß dort ein Maximum (bzw. Minimum) vorliegt. Im Unterschied zu den Verhältnissen bei Funktionen einer einzigen Veränderlichen (vgl. die Bemerkung in Nr. 139) kann man jedoch hieraus noch nicht schließen, daß es sich dabei um den größten (bzw. kleinsten) Wert der Funktion im Bereich handelt. Folgendes einfache Beispiel zeigt, daß ein solcher Schluß tatsächlich zu falschen Resultaten führen kann.

Wir betrachten im Rechteck $[-5, 5; -1, 1]$ die Funktion

$$u = x^3 - 4x^2 + 2xy - y^2.$$

Ihre Ableitungen

$$u_x = 3x^2 - 8x + 2y, \qquad u_y = 2x - 2y$$

verschwinden in diesem Bereich nur im Punkt $(0, 0)$. Mit Hilfe des Kriteriums aus Nr. 197 überzeugt man sich leicht davon, daß in diesem Punkt die Funktion ein Maximum hat (es ist gleich 0). Dieser Wert ist aber nicht der größte im Bereich: Im Punkt $(5, 0)$ gilt ja $u = 25$. Infolgedessen reicht es bei Funktionen mehrerer Veränderlicher, deren größter bzw. kleinster Wert in einem Bereich bestimmt werden soll, nicht aus, nur die (relativen) Maxima bzw. Minima aufzusuchen.

---

[1]) Auch hier ist die Anzahl der Summanden beliebig.

201. Aufgaben. Viele Aufgaben der Mathematik, aber auch aus anderen Gebieten von Wissenschaft und Technik führen auf das Problem, den größten oder den kleinsten Wert einer Funktion zu bestimmen. Die Lösung der Aufgaben 1 bis 4 hängt mit den in Nr. 200 behandelten Beispielen zusammen.

1. Unter allen einem gegebenen Kreis vom Radius R einbeschriebenen Dreiecken bestimme man das Dreieck größten Flächeninhalts (Abb. 107).

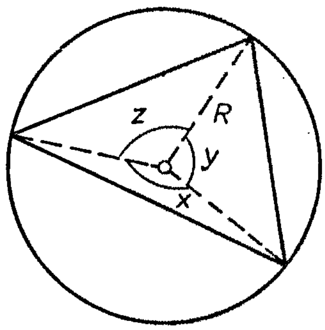

Abb. 107

Bezeichnen x, y, z die Zentriwinkel über den Dreiecksseiten, so gilt $x + y + z = 2\pi$, also $= 2\pi - x - y$. Für den Flächeninhalt P des Dreiecks gilt dann

$$P = \frac{1}{2} R^2 \cdot \sin x + \frac{1}{2} R^2 \cdot \sin y + \frac{1}{2} R^2 \cdot \sin z$$
$$= \frac{1}{2} R^2 \cdot [\sin x + \sin y - \sin(x + y)].$$

Der Variationsbereich von x und y ergibt sich hier zu $x \geqq 0$, $y \geqq 0$, $x + y \leqq 2\pi$. Es sind also diejenigen Werte der Veränderlichen zu bestimmen, für welche die eckige Klammer den größten Wert annimmt.

Nun wissen wir schon (Nr. 200, Beispiel 1), daß das für $x = y = \frac{2\pi}{3}$ der Fall ist. Dann ist auch $z = \frac{2\pi}{3}$. Das gesuchte Dreieck ist also das gleichseitige.

2. Unter allen Dreiecken vom Umfang $2p$ bestimme man dasjenige mit dem größten Flächeninhalt.

Sind x, y, z die Dreiecksseiten, so ist nach der Heronschen Formel

$$P = \sqrt{p(p - x)(p - y)(p - z)}.$$

Man könnte hier $z = 2p - x - y$ setzen und den Flächeninhalt in der Form

$$P = \sqrt{p(p - x)(p - y)(x + y - p)}$$

schreiben und den größten Wert dieser Funktion in dem Dreiecksbereich aus Beispiel (f) von Nr. 160 bestimmen.

Wir gehen anders vor: Das Problem reduziert sich auf die Bestimmung des größten Wertes des Produktes $u = (p - x)(p - y)(p - z)$ positiver Zahlen, deren Summe

$$(p - x) + (p - y) + (p - z) = 3p - 2p = p$$

konstant ist.

Wir wissen aus Nr. 200, Beispiel 3, daß dieser größte Wert angenommen wird, wenn alle Faktoren gleich sind, also wenn $x = y = z = \frac{2p}{3}$ ist. Wieder ergibt sich das gleichseitige Dreieck.

3. Man bestimme unter den dem Ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

einbeschriebenen rechtwinkligen achsenparallelen Parallelepipeden dasjenige größten Volumens.

Sind x, y, z die Koordinaten der Ecke, die im ersten Koordinatenoktanten liegt, so gilt für das Volumen $v = 8xyz$. Statt v kann man auch die Größe

$$u = \frac{v^2}{64a^2b^2c^2} = \frac{x^2}{a^2} \cdot \frac{y^2}{b^2} \cdot \frac{z^2}{c^2}$$

untersuchen, da beide Ausdrücke offenbar für dieselben Werte von x, y, z maximal werden. Damit haben wir aber das Problem auf das Beispiel 3 aus Nr. 200 zurückgeführt. Die Antwort lautet demnach

$$\frac{x^2}{a^2} = \frac{y^2}{b^2} = \frac{z^2}{c^2} = \frac{1}{3}, \quad \text{also} \quad x = \frac{a}{\sqrt{3}}, \quad y = \frac{b}{\sqrt{3}}, \quad z = \frac{c}{\sqrt{3}}.$$

4. Wir nehmen nun an, irgendein Gas (etwa Luft) werde in einem Kompressor vom Druck p_0 auf den Druck $p > p_0$ komprimiert. Für die dabei zur Kompression eines Mols des Gases notwendige Arbeit gilt

$$A = RT_0 \cdot \frac{\gamma}{\gamma - 1}\left[\left(\frac{p}{p_0}\right)^{\frac{\gamma-1}{\gamma}} - 1\right].$$

Hierbei ist R die Gaskonstante, T_0 die absolute Temperatur des Gases vor der Kompression und γ eine Zahl (> 1), die von der Konstruktion des Kompressors abhängt. Die Arbeit A ist offenbar um so geringer, je niedriger die Ausgangstemperatur T_0 ist. Bei höherer Kompression zerlegt man, wenn die Kosten bei der aufzuwendenden Arbeit eine Rolle spielen, den ganzen Prozeß der Kompression in einzelne Stufen, in Etappen, in denen man das komprimierte (und dabei erwärmte) Gas abkühlt.

Es möge etwa ein dreistufiger Kompressor vorliegen mit zwei Kühlkammern, in denen die Temperatur wieder auf T_0 gebracht wird. Bezeichnet man mit p_1 bzw. p_2 den Druck am Ende der ersten bzw. der zweiten Stufe, so gilt für die bei der Kompression zu leistende Arbeit

$$A = RT_0 \cdot \frac{\gamma}{\gamma - 1}\left\{\left[\left(\frac{p_1}{p_0}\right)^{\frac{\gamma-1}{\gamma}} - 1\right] + \left[\left(\frac{p_2}{p_1}\right)^{\frac{\gamma-1}{\gamma}} - 1\right] + \left[\left(\frac{p}{p_2}\right)^{\frac{\gamma-1}{\gamma}} - 1\right]\right\}.$$

Die Frage lautet, wie bei gegebenen p_0, p, T_0 die Drücke p_1 und p_2 zu wählen sind, damit die aufzuwendende Arbeit möglichst klein wird.

Läßt man den konstanten Faktor und konstante Summanden weg, welche die gesuchten Größen p_1 und p_2 nicht beeinflussen, so reduziert sich das Problem auf die Untersuchung des Ausdrucks

$$u = \left(\frac{p_1}{p_0}\right)^{\frac{\gamma-1}{\gamma}} + \left(\frac{p_2}{p_1}\right)^{\frac{\gamma-1}{\gamma}} + \left(\frac{p}{p_2}\right)^{\frac{\gamma-1}{\gamma}}$$

Da das Produkt

$$\left(\frac{p_1}{p_0}\right)^{\frac{\gamma-1}{\gamma}} \cdot \left(\frac{p_2}{p_1}\right)^{\frac{\gamma-1}{\gamma}} \cdot \left(\frac{p}{p_2}\right)^{\frac{\gamma-1}{\gamma}} = \left(\frac{p}{p_0}\right)^{\frac{\gamma-1}{\gamma}}$$

konstant ist, folgt nach Beispiel 4 aus Nr. 200, daß die Summe u dann am kleinsten ist, wenn alle Summanden gleich sind:

$$\left(\frac{p_1}{p_0}\right)^{\frac{\gamma-1}{\gamma}} = \left(\frac{p_2}{p_1}\right)^{\frac{\gamma-1}{\gamma}} = \left(\frac{p}{p_2}\right)^{\frac{\gamma-1}{\gamma}}$$

oder

$$\frac{p_1}{p_0} = \frac{p_2}{p_1} = \frac{p}{p_2},$$

so daß die aufeinanderfolgenden Drücke eine geometrische Progression bilden. Somit gilt

$$p_1 = \sqrt[3]{p_0^2 \cdot p}, \qquad p_2 = \sqrt[3]{p_0 \cdot p^2}.$$

5. In der Ebene sei ein Dreieck mit den Seiten a, b, c gegeben (Abb. 108). Darauf lassen sich unendlich viele Pyramiden gleicher Höhe h errichten. Man soll diejenige bestimmen, die die kleinste Mantelfläche S hat. Das Problem reduziert sich darauf, die Projektion M der Pyramidenspitze zu finden. Ihre Lage ergibt sich aus den drei Loten x, y, z auf die Seiten a, b, c. Jedes

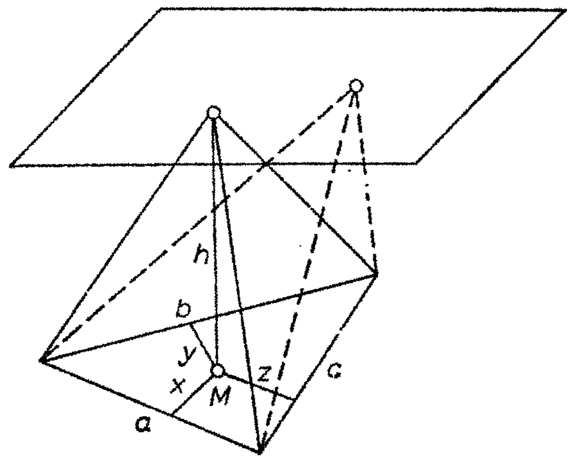

Abb. 108

Lot nehmen wir positiv, wenn M auf derselben Seite wie das Dreieck liegt, sonst negativ; zwischen x, y, z besteht die Beziehung

$$ax + by + cz = 2P, \quad \text{also} \quad z = \frac{2P - ax - by}{c},$$

wobei P der Flächeninhalt des Dreiecks ist.

Für die uns interessierende Mantelfläche S ergibt sich jetzt

$$S = \frac{a}{2}\sqrt{x^2 + h^2} + \frac{b}{2}\sqrt{y^2 + h^2} + \frac{c}{2}\sqrt{z^2 + h^2},$$

wobei für z der oben gefundene Ausdruck einzusetzen ist. Der Variationsbereich der unabhängigen Veränderlichen x, y ist die ganze x, y-Ebene. Nun ist

$$2S_x = \frac{ax}{\sqrt{x^2 + h^2}} - \frac{cz}{\sqrt{z^2 + h^2}} \cdot \frac{a}{c} = 0,$$

$$2S_y = \frac{by}{\sqrt{y^2 + h^2}} - \frac{cz}{\sqrt{z^2 + h^2}} \cdot \frac{b}{c} = 0$$

oder

$$\frac{x}{\sqrt{x^2 + h^2}} = \frac{y}{\sqrt{y^2 + h^2}} = \frac{z}{\sqrt{z^2 + h^2}}, \quad \text{also} \quad x = y = z.$$

Der gesuchte Punkt M ist der Mittelpunkt des dem Dreieck einbeschriebenen Kreises.

Daß diesen Werten von x und y der kleinste Wert für S entspricht, ist ebenso leicht zu beweisen wie im Beispiel 4 aus Nr. 200, indem man benutzt, daß für unbegrenzt wachsende x bzw. y auch S unbegrenzt wächst.

6. Es seien in der Ebene drei Punkte $M_1(a_1, b_1)$, $M_2(a_2, b_2)$, $M_3(a_3, b_3)$ gegeben, die nicht auf einer Geraden liegen. Man bestimme in dieser Ebene denjenigen Punkt, für welchen die Summe der Abstände von den gegebenen Punkten minimal ist.

Für beliebiges $M(x, y)$ setzen wir $\varrho_i = \sqrt{(x - a_i)^2 + (y - b_i)^2}$ $(i = 1, 2, 3)$. Dann ist die Funktion

$$u = \sum_i \varrho_i = \sum_i \sqrt{(x - a_i)^2 + (y - b_i)^2}$$

zu untersuchen. Ihre partiellen Ableitungen

$$\frac{\partial u}{\partial x} = \sum_i \frac{x - a_i}{\varrho_i} = \sum_i \cos\theta_i, \qquad \frac{\partial u}{\partial y} = \sum_i \frac{y - b_i}{\varrho_i} = \sum_i \sin\theta_i,$$

wobei θ_i den Winkel zwischen der Geraden M_iM und der x-Achse bedeutet, existieren mit Ausnahme der gegebenen Punkte überall.

Extremwertverdächtig sind somit zunächst die Punkte M_1, M_2 und M_3, in denen die Ableitungen nicht existieren, sowie ein Punkt M_0 (wir werden sehen, daß er nicht immer existiert), in dem die Ableitungen gleichzeitig verschwinden. Da für unbegrenzt wachsendes x bzw. y auch u unbegrenzt wächst, wird der gesuchte kleinste Wert in einem dieser Punkte angenommen.

Um die stationären Punkte M_0 zu bestimmen, setzen wir beide partielle Ableitungen gleich 0; so erhalten wir

$$\cos\theta_1 + \cos\theta_2 + \cos\theta_3 = 0, \qquad \sin\theta_1 + \sin\theta_2 + \sin\theta_3 = 0.$$

Wir multiplizieren die erste mit $\sin\theta_2$, die zweite mit $\cos\theta_2$ und subtrahieren; es ergibt sich

$$\sin(\theta_1 - \theta_2) = \sin(\theta_2 - \theta_3), \quad \text{also} \quad \theta_1 - \theta_2 = \theta_2 - \theta_3.$$

Analog finden wir

$$\theta_2 - \theta_3 = \theta_3 - \theta_1.$$

Somit müssen die Winkel zwischen je zwei der Geraden M_1M_0, M_2M_0, M_3M_0 sämtlich gleich $\frac{2\pi}{3}$ sein, und M_0 ergibt sich als Schnittpunkt der Bögen, die über den Seiten des Dreiecks $M_1M_2M_3$ konstruiert sind und den Winkel $\frac{2\pi}{3}$ einschließen.

Gibt es in diesem Dreieck keinen Winkel von mindestens $\frac{2\pi}{3}$, so schneiden sich diese Bögen wirklich im Innern des Dreiecks und bestimmen einen Punkt M_0, von dem aus die Seiten unter dem Winkel $\frac{2\pi}{3}$ zu sehen sind (Abb. 109). Als Nächstes muß man die Werte vergleichen, die u

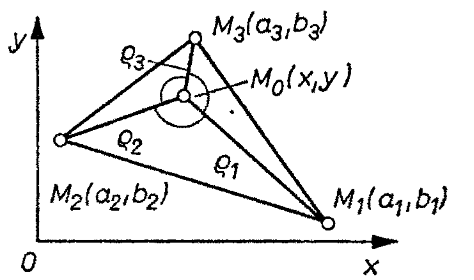

Abb. 109

in diesen vier Punkten annimmt. Wir zeigen, daß der Wert von u in dem stationären Punkt M_0 kleiner ist als in den übrigen (also in der Tat der kleinste ist). Nach dem Kosinussatz ist nämlich

$$\overline{M_1M_2}^2 = \overline{M_0M_1}^2 + \overline{M_0M_2}^2 + \overline{M_0M_1}\cdot\overline{M_0M_2} > \left(\overline{M_0M_2} + \frac{1}{2}\overline{M_0M_1}\right)^2,$$

also

$$\overline{M_1M_2} > \overline{M_0M_2} + \frac{1}{2}\overline{M_0M_1}.$$

Analog ergibt sich

$$\overline{M_1M_3} > \overline{M_0M_3} + \frac{1}{2}\overline{M_0M_1}.$$

Durch Addition folgt $\overline{M_1M_2} + \overline{M_1M_3} > \overline{M_0M_1} + \overline{M_0M_2} + \overline{M_0M_3}$, also

$$u(M_1) > u(M_0).$$

Offenbar kann M_1 hier durch M_2 oder M_3 ersetzt werden, und damit ist der Beweis beendet.

Anders verhält es sich, wenn einer der Dreieckswinkel mindestens $\frac{2\pi}{3}$ ist. Dann existiert kein stationärer Punkt, und u nimmt den kleinsten Wert in einem der Punkte M_1, M_2, M_3 an, und zwar in dem Scheitel des stumpfen Winkels. Eine interessante Besonderheit dieser Aufgabe besteht gerade darin, daß man dabei neben dem stationären Punkt noch solche Punkte zu berücksichtigen hat, in denen die Ableitung nicht existiert (vgl. Nr. 196, Bemerkung II).

7. Wir verallgemeinern die Aufgabe 1: Wir wollen das dem Kreis vom Radius R einbeschriebene $(n+1)$-Eck mit größtem Flächeninhalt P bestimmen.

Wir bezeichnen die Zentriwinkel über den Polygonseiten mit $x_1, x_2, \ldots, x_n, x_{n+1}$. Dann ist $x_1 + x_2 + \cdots + x_n + x_{n+1} = 2\pi$, also

$$x_{n+1} = 2\pi - (x_1 + x_2 + \cdots + x_n).$$

Für P gilt

$$P = \frac{1}{2} R^2 \cdot \sin x_1 + \frac{1}{2} R^2 \cdot \sin x_2 + \cdots + \frac{1}{2} R^2 \cdot \sin x_n + \frac{1}{2} R^2 \cdot \sin x_{n+1}.$$

Setzt man den Wert für x_{n+1} ein, so reduziert sich das Problem auf die Bestimmung des größten Wertes der Funktion

$$u = \sin x_1 + \sin x_2 + \cdots + \sin x_n + \sin [2\pi - (x_1 + x_2 + \cdots + x_n)],$$

wobei der Variationsbereich $\mathfrak{D}$ der Veränderlichen $x_1, x_2, \ldots, x_n$ durch die Ungleichungen

$$x_1 \geqq 0, x_2 \geqq 0, \ldots, x_n \geqq 0, x_1 + x_2 + \cdots + x_n \leqq 2\pi$$

gegeben ist, also ein n-dimensionales Simplex darstellt (Nr. 162).

Nach der allgemeinen Regel berechnen wir die Ableitungen und setzen sie gleich 0:

$$\cos x_1 - \cos (x_1 + x_2 + \cdots + x_n) = 0,$$

$$\cos x_n - \cos (x_1 + x_2 + \cdots + x_n) = 0.$$

Der einzige innere Punkt des Bereichs, in dem diese Bedingungen gelten, ist der Punkt

$$x_1 = x_2 = \cdots = x_n = \frac{2\pi}{n+1} \quad \left(\text{dann ist auch } x_{n+1} = \frac{2\pi}{n+1}\right);$$

ihm entspricht $u = (n+1) \sin \frac{2\pi}{n+1}$.

Um zu zeigen, daß dies wirklich der größte Wert von u ist, benutzen wir Induktion. Für $n = 2$ gilt die Behauptung, wie Beispiel 1 aus Nr. 200 zeigt. Wir nehmen nun an, sie gelte für n Sinusglieder (so daß für ihre Summe der größte Wert gleich $n \cdot \sin \frac{2\pi}{n}$ ist), und zeigen, daß sie dann auch für $n+1$ Sinusglieder gilt.

Auf Grund der allgemeinen Bemerkung muß man den Wert $(n+1) \cdot \sin \frac{2\pi}{n+1}$ mit den Werten vergleichen, welche die Funktion auf dem Rand von $\mathfrak{D}$ annimmt. Wir nehmen etwa die „Simplexfläche" $x_n = 0$; auf ihr ist u eine Funktion von nur $n-1$ Veränderlichen:

$$u = \sin x_1 + \sin x_2 + \cdots + \sin x_{n-1} + \sin [2\pi - (x_1 + x_2 + \cdots + x_{n-1})],$$

deren größter Wert nach Induktionsannahme gleich $n \cdot \sin \frac{2\pi}{n}$ ist. Das kann man genauso für die übrigen Seitenflächen erschließen. Da aber

$$n \cdot \sin \frac{2\pi}{n} < (n+1) \cdot \sin \frac{2\pi}{n+1}$$

gilt, ist unsere Behauptung bewiesen.[1])

---

[1]) Das folgt daraus, daß $\frac{\sin z}{z}$ mit von 0 bis π wachsendem z monoton abnimmt (vgl. Nr. 133, Beispiel 1).

8. Wir betrachten einen elektrischen Stromkreis in Parallelschaltung. Abb. 110 zeigt das Schaltschema; A und B sind die Klemmen der Stromquelle, $P_1, P_2, \ldots, P_n$ Geräte, die den Strom $i_1, i_2, \ldots, i_n$ verbrauchen. Man bestimme bei vorgegebenem Spannungsabfall $2e$ im Stromkreis den Querschnitt der Leiter so, daß auf der ganzen Leitung die geringste Menge Kupfer verbraucht wird.

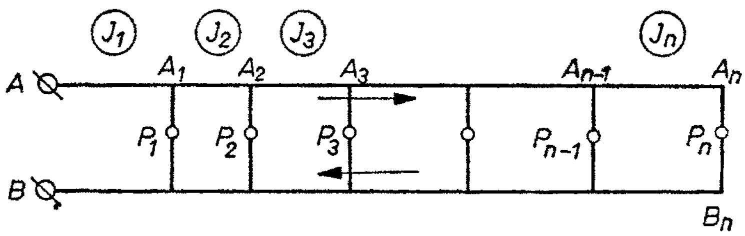

Abb. 110

Offenbar braucht man nur einen der Leiter zu untersuchen, etwa AA_n, da für den anderen dieselben Bedingungen gelten. Mit $l_1, l_2, \ldots, l_n$ bezeichnen wir die Längen der Stücke AA_1, $A_1A_2, \ldots, A_{n-1}A_n$ in m, mit $q_1, q_2, \ldots, q_n$ die Querschnitte in mm². Dann ist

$$u_1 = l_1q_1 + l_2q_2 + \cdots + l_nq_n$$

das Volumen der gebrauchten Kupfermenge in cm³. Wir müssen den kleinsten Wert dieser Größe bestimmen unter der Annahme, daß der Spannungsabfall in AA_n gleich e ist.

Man berechnet leicht, welche Ströme $J_1, J_2, \ldots, J_n$ in den Strecken $AA_1, A_1A_2, \ldots, A_{n-1}A_n$ fließen:

$$J_1 = i_1 + i_2 + \cdots + i_n, \qquad J_2 = i_2 + \cdots + i_n, \ldots, \qquad J_n = i_n.$$

Ist ϱ der Widerstand eines kupfernen Leiters der Länge 1 m vom Querschnitt 1 mm², so gilt für die Widerstände dieser Strecken

$$r_1 = \frac{\varrho l_1}{q_1}, \qquad r_2 = \frac{\varrho l_2}{q_2}, \qquad r_n = \frac{\varrho l_n}{q_n},$$

so daß sich nach dem Ohmschen Gesetz der Spannungsabfall folgendermaßen ausdrücken läßt:

$$e_1 = r_1J_1 = \frac{\varrho l_1 J_1}{q_1}, \qquad e_2 = r_2J_2 = \frac{\varrho l_2 J_2}{q_2}, \qquad e_n = r_nJ_n = \frac{\varrho l_n J_n}{q_n}.$$

Um komplizierte Rechnungen zu vermeiden, führen wir statt der $q_1, q_2, \ldots, q_n$ diese Größen $e_1, e_2, \ldots, e_n$ ein, für welche

$$e_1 + e_2 + \cdots + e_{n-1} + e_n = e, \quad \text{also} \quad e_n = e - e_1 - e_2 - \cdots - e_{n-1}$$

gilt. Dann ist

$$q_1 = \frac{\varrho l_1 J_1}{e_1}, \qquad q_2 = \frac{\varrho l_2 J_2}{e_2}, \qquad q_n = \frac{\varrho l_n J_n}{e_n} = \frac{\varrho l_n J_n}{e - e_1 - e_2 - \cdots - e_{n-1}}$$

und

$$u = \varrho\left[\frac{l_1^2 J_1}{e_1} + \cdots + \frac{l_{n-1}^2 J_{n-1}}{e_{n-1}} + \frac{l_2^n J_n}{e - e_1 - \cdots - e_{n-1}}\right],$$

wobei der Variationsbereich der unabhängigen Veränderlichen $e_1, e_2, \ldots, e_{n-1}$ das offene Simplex

$$e_1 > 0, \quad e_2 > 0, \ldots, e_{n-1} > 0, \quad e_1 + e_2 + \cdots + e_{n-1} < e$$

ist.

Setzen wir die partiellen Ableitungen von u nach allen Veränderlichen gleich 0, so erhalten wir

$$-\frac{l_1^2 J_1}{e_1^2} + \frac{l_n^2 J_n}{(e - e_1 - \cdots - e_{n-1})^2} = 0,$$

$$-\frac{l_2^2 J_2}{e_2^2} + \frac{l_n^2 J_n}{(e - e_1 - \cdots - e_{n-1})^2} = 0,$$

$$-\frac{l_{n-1}^2 J_{n-1}}{e_{n-1}^2} + \frac{l_n^2 J_n}{(e - e_1 - \cdots - e_{n-1})^2} = 0,$$

also (wenn wir e_n wieder einführen)

$$\frac{l_1^2 J_1}{e_1^2} = \frac{l_2^2 J_2}{e_2^2} = \cdots = \frac{l_n^2 J_n}{e_n^2}.$$

Man bezeichnet den gemeinsamen Wert dieser Brüche zweckmäßigerweise mit $\frac{1}{\lambda^2}$ $(\lambda > 0)$, und dann ergibt sich

$$e_1 = \lambda l_1 \sqrt{J_1}, \quad e_2 = \lambda l_2 \sqrt{J_2}, \qquad e_n = \lambda l_n \sqrt{J_n},$$

wobei man λ leicht aus $e_1 + c_2 + \cdots + e_n = e$ erhält:

$$\lambda = \frac{e}{l_1 \sqrt{J_1} + l_2 \sqrt{J_2} + \cdots + l_n \sqrt{J_n}}.$$

Da u unbegrenzt wächst, wenn sich der Punkt $(e_1, e_2, \ldots, e_{n-1})$ dem Rand des Bereichs nähert, nimmt u für diese $e_1, e_2, \ldots, e_{n-1}$ (e_n) tatsächlich den kleinsten Wert an.

Gehen wir wieder auf die $q_1, q_2, \ldots, q_n$ zurück, so ergibt sich

$$q_1 = \frac{\varrho}{\lambda} \sqrt{J_1}, \quad q_2 = \frac{\varrho}{\lambda} \sqrt{J_2}, \qquad q_n = \frac{\varrho}{\lambda} \sqrt{J_n},$$

so daß die am besten geeigneten Querschnitte der Leiter den Quadratwurzeln aus den entsprechenden Stromstärken proportional sind.

9. *Die Methode der kleinsten Quadrate.* Unter diesem Namen ist eine sehr verbreitete Methode zur Auswertung von Meßergebnissen bekannt, deren Wesen in folgendem besteht:

Es seien die Werte dreier[1]) Größen x, y, z zu bestimmen, welche $n > 3$ linearen Gleichungen

$$a_i x + b_i y + c_i z = d_i \qquad (i = 1, 2, \ldots, n)$$

genügen. Hierbei seien einige der Koeffizienten a_i, b_i, c_i, d_i aus Beobachtungen, also nur angenähert, bestimmt. Dabei setzen wir voraus, wenigstens irgendwelche drei dieser Gleichungen hätten eine von 0 verschiedene Determinante, es sei also etwa

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} \neq 0. \tag{14}$$

Die aus den ersten drei Gleichungen berechneten Werte x, y, z werden im allgemeinen die übrigen Gleichungen nicht befriedigen, sei es, weil die Koeffizienten unvermeidliche Fehler enthalten, sei es, weil die Gleichungen selbst nur angenähert gelten. Da wir keinen Grund haben, eine der Gleichungen vor den anderen auszuzeichnen, und mit einem unvermeidlichen Fehler

$$\delta_i = a_i x + b_i x + c_i z - d_i$$

rechnen, welche Werte x, y, z wir auch nehmen, stellen wir uns nur die Aufgabe, die Summe der Quadrate dieser Fehler, also

$$W = \sum_{i=1}^{n} \delta_i^2 = \sum_{i=1}^{n} (a_i x + b_i y + c_i z - d_i)^2,$$

[1]) Auf diesen Fall beschränken wir uns hier nur der Einfachheit halber.

möglichst klein zu machen (daher der Name der Methode).[1]) Mit anderen Worten, wir nehmen an, die Werte von x, y, z, welche am besten mit der Erfahrung übereinstimmen, seien diejenigen, für welche $W = W(x, y, z)$ minimal ist.

Um diese Werte zu bekommen, setzen wir die partiellen Ableitungen von W nach x, y, z gleich 0 und erhalten

$$2 \sum_{i=1}^{n} a_i(a_i x + b_i y + c_i z - d_i) = 0,$$

$$2 \sum_{i=1}^{n} b_i(a_i x + b_i y + c_i z - d_i) = 0,$$

$$2 \sum_{i=1}^{n} c_i(a_i x + b_i y + c_i z - d_i) = 0.$$

Für Summen von Summanden, die sich nur durch verschiedene Indizes unterscheiden, führte C. F. GAUSS eine kurze Bezeichnung ein; er schrieb

$$[aa] \quad \text{statt} \quad \sum_{i=1}^{n} a_i^2, \qquad [ab] \quad \text{statt} \quad \sum_{i=1}^{n} a_i b_i \quad \text{usw.}$$

In der Gaußschen Bezeichnung lautet das Gleichungssystem zur Bestimmung von x, y, z

$$\begin{aligned} &[aa]\,x + [ab]\,y + [ac]\,z = [ad], \\ &[ba]\,x + [bb]\,y + [bc]\,z = [bd], \\ &[ca]\,x + [cb]\,y + [cc]\,z = [cd] \end{aligned}$$

(*Normalgleichungssystem*).

Um sich zu vergewissern, daß x, y, z dadurch eindeutig bestimmt sind, muß man zeigen, daß die Determinante des Systems von 0 verschieden ist. Nach einem bekannten Satz der Algebra gilt für das Quadrat dieser Determinante die Beziehung

$$\begin{vmatrix} [aa] & [ab] & [ac] \\ [ba] & [bb] & [bc] \\ [ca] & [cb] & [cc] \end{vmatrix}^2 = \sum_{(i,j,k)} \begin{vmatrix} a_i & b_i & c_i \\ a_j & b_j & c_j \\ a_k & b_k & c_k \end{vmatrix}^2,$$

wobei über alle möglichen Kombinationen (i, j, k) von n Indizes $1, 2, \ldots, n$ zu je dreien zu summieren ist. Da von allen Determinanten rechts nach Voraussetzung mindestens eine von 0 verschieden ist, folgt daraus, daß auch die Determinante rechts von 0 verschieden ist.

Wir müssen uns noch davon überzeugen, daß die Funktion W für die aus dem Normalgleichungssystem bestimmten Werte der Veränderlichen tatsächlich den kleinsten Wert annimmt. Dazu genügt es zu zeigen, daß W außerhalb einer Kugel von hinreichend großem Radius beliebig groß wird. Zu diesem Zweck betrachten wir die Werte der ersten drei Klammern in dem Ausdruck für W:

$$a_1 x + b_1 y + c_1 z - d_1 = u_1, \quad a_2 x + b_2 y + c_2 z - d_2 = u_2,$$
$$a_3 x + b_3 y + c_3 z - d_3 = u_3.$$

Auf Grund von (14) lassen sich durch diese Werte die Werte von x, y, z mit völlig bestimmten konstanten Koeffizienten linear ausdrücken, so daß, da alle Größen u_1, u_2, u_3 beschränkt sind, auch x, y, z beschränkt sind. Hieraus ist zu ersehen, daß mit unbegrenzt wachsendem $r^2 = x^2 + y^2 + z^2$ auch $u_1^2 + u_2^2 + u_3^2$, also auch W, unbegrenzt wächst.

---

[1]) Vgl. dazu beispielsweise J. W. LINNIK, Methode der kleinsten Quadrate in moderner Darstellung, Berlin 1961 (Übersetzung aus dem Russischen).

VI. Funktionaldeterminanten und ihre Anwendung

§ 1. Formale Eigenschaften der Funktionaldeterminanten

202. Definition der Funktionaldeterminante. In diesem Kapitel und in einigen anderen Teilen dieses Lehrgangs ist eine besondere Art von Determinanten ein wichtiges Hilfsmittel bei unseren Untersuchungen: Determinanten, deren Elemente partielle Ableitungen sind. Zunächst wollen wir einige grundlegende Eigenschaften dieser Determinanten studieren.

Es seien n Funktionen von n Veränderlichen gegeben:

$$\left.\begin{aligned} y_1 &= f_1(x_1, x_2, \ldots, x_n), \\ y_2 &= f_2(x_1, x_2, \ldots, x_n), \\ &\ \\ y_n &= f_n(x_1, x_2, \ldots, x_n). \end{aligned}\right\} \tag{1}$$

Sie seien in einem n-dimensionalen Bereich $\mathcal{D}$ gegeben und mögen dort stetige partielle Ableitungen nach allen Veränderlichen besitzen. Die Determinante

$$\begin{vmatrix} \dfrac{\partial y_1}{\partial x_1} & \dfrac{\partial y_1}{\partial x_2} & \cdots & \dfrac{\partial y_1}{\partial x_n} \\ \dfrac{\partial y_2}{\partial x_1} & \dfrac{\partial y_2}{\partial x_2} & \cdots & \dfrac{\partial y_2}{\partial x_n} \\ & & & \\ \dfrac{\partial y_n}{\partial x_1} & \dfrac{\partial y_n}{\partial x_2} & \cdots & \dfrac{\partial y_n}{\partial x_n} \end{vmatrix}$$

nennt man die *Funktionaldeterminante des* Systems (1). Vielfach wird sie auch nach C. G. J. Jacobi, der als erster ihre Eigenschaften und Anwendungsmöglichkeiten untersuchte, *Jacobian* genannt.[1]) Auch die Abkürzung

$$\frac{D(y_1, y_2, \ldots, y_n)}{D(x_1, x_2, \ldots, x_n)}$$

ist gebräuchlich. Die Funktionaldeterminanten ähneln in einer Reihe von Eigenschaften der gewöhnlichen Ableitung.

---

[1]) Gleichzeitig mit Jacobi führte der russische Mathematiker Michail Wassiljewitsch Ostrogradski (1801—1862) Funktionalderminanten ein.

203. Multiplikation von Funktionaldeterminanten. Neben dem System (1) betrachten wir das System

$$\left.\begin{aligned} x_1 &= \varphi_1(t_1, t_2, \ldots, t_n), \\ x_2 &= \varphi_2(t_1, t_2, \ldots, t_n), \\ &\ \ \vdots \\ x_n &= \varphi_n(t_1, t_2, \ldots, t_n); \end{aligned}\right\} \tag{2}$$

diese Funktionen seien in einem Bereich $\mathcal{P}$ definiert und mögen dort stetige partielle Ableitungen besitzen. Falls der Punkt $(t_1, \ldots, t_n)$ in $\mathcal{P}$ variiert, möge der entsprechende Punkt $(x_1, x_2, \ldots, x_n)$ niemals außerhalb $\mathcal{D}$ liegen, so daß die $y_1, y_2, \ldots, y_n$ als mittelbare Funktionen der $t_1, t_2, \ldots, t_n$ angesehen werden können.

Wir multiplizieren nun die Funktionaldeterminante des Systems (1) mit der Funktionaldeterminante

$$\begin{vmatrix} \frac{\partial x_1}{\partial t_1} & \frac{\partial x_1}{\partial t_2} & \cdots & \frac{\partial x_1}{\partial t_n} \\ \frac{\partial x_2}{\partial t_1} & \frac{\partial x_2}{\partial t_2} & \cdots & \frac{\partial x_2}{\partial t_n} \\ \cdot & \cdot & & \cdot \\ \frac{\partial x_n}{\partial t_1} & \frac{\partial x_n}{\partial t_2} & \cdots & \frac{\partial x_n}{\partial t_n} \end{vmatrix}$$

des Systems (2). Aus der Determinantentheorie setzen wir den Multiplikationssatz als bekannt voraus; in

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \cdot \begin{vmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \cdot & & & \cdot \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{vmatrix} = \begin{vmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \cdot & & & \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{vmatrix}$$

gilt also

$$c_{ik} = a_{i1}b_{1k} + a_{i2}b_{2k} + \cdots + a_{in}b_{nk} \quad (i, k = 1, 2, \ldots, n)$$

(Multiplikation von Zeile mit Spalte). Für die Funktionaldeterminanten gilt somit

$$\begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \cdot & \cdot & & \cdot\ \cdot \\ \frac{\partial y_n}{\partial x_1} & \frac{\partial y_n}{\partial x_2} & \cdots & \frac{\partial y_n}{\partial x_n} \end{vmatrix} \cdot \begin{vmatrix} \frac{\partial x_1}{\partial t_1} & \frac{\partial x_1}{\partial t_2} & \cdots & \frac{\partial x_1}{\partial t_n} \\ \frac{\partial x_2}{\partial t_1} & \frac{\partial x_2}{\partial t_2} & \cdots & \frac{\partial x_2}{\partial t_n} \\ \cdot & \cdot & & \cdot\ \cdot \\ \frac{\partial x_n}{\partial t_1} & \frac{\partial x_n}{\partial t_2} & \cdots & \frac{\partial x_n}{\partial t_n} \end{vmatrix}$$

$$= \begin{vmatrix} \frac{\partial y_1}{\partial x_1}\cdot\frac{\partial x_1}{\partial t_1} + \cdots + \frac{\partial y_1}{\partial x_n}\cdot\frac{\partial x_n}{\partial t_1} & \cdots & \frac{\partial y_1}{\partial x_1}\cdot\frac{\partial x_1}{\partial t_n} + \cdots + \frac{\partial y_1}{\partial x_n}\cdot\frac{\partial x_n}{\partial t_n} \\ \frac{\partial y_2}{\partial x_1}\cdot\frac{\partial x_1}{\partial t_1} + \cdots + \frac{\partial y_2}{\partial x_n}\cdot\frac{\partial x_n}{\partial t_1} & \cdots & \frac{\partial y_2}{\partial x_1}\cdot\frac{\partial x_1}{\partial t_n} + \cdots + \frac{\partial y_2}{\partial x_n}\cdot\frac{\partial x_n}{\partial t_n} \\ \cdot & & \cdot \\ \frac{\partial y_n}{\partial x_1}\cdot\frac{\partial x_1}{\partial t_1} + \cdots + \frac{\partial y_n}{\partial x_n}\cdot\frac{\partial x_n}{\partial t_1} & \cdots & \frac{\partial y_n}{\partial x_1}\cdot\frac{\partial x_1}{\partial t_n} + \cdots + \frac{\partial y_n}{\partial x_n}\cdot\frac{\partial x_n}{\partial t_n} \end{vmatrix}.$$

Da nach der Kettenregel für mittelbare Funktionen für das allgemeine Glied dieser letzten Determinante die Beziehung

$$\frac{\partial y_i}{\partial x_1} \cdot \frac{\partial x_1}{\partial t_k} + \cdots + \frac{\partial y_i}{\partial x_n} \cdot \frac{\partial x_n}{\partial t_k} = \frac{\partial y_i}{\partial t_k} \quad (i, k = 1, 2, \ldots, n)$$

gilt, können wir dafür auch

$$\begin{vmatrix} \dfrac{\partial y_1}{\partial t_1} & \dfrac{\partial y_1}{\partial t_2} & \cdots & \dfrac{\partial y_1}{\partial t_n} \\ \dfrac{\partial y_2}{\partial t_1} & \dfrac{\partial y_2}{\partial t_2} & \cdots & \dfrac{\partial y_2}{\partial t_n} \\ & & & \\ \dfrac{\partial y_n}{\partial t_1} & \dfrac{\partial y_n}{\partial t_2} & \cdots & \dfrac{\partial y_n}{\partial t_n} \end{vmatrix}$$

schreiben. In der abgekürzten Form läßt sich diese Eigenschaft folgendermaßen notieren:

$$\frac{D(y_1, y_2, \ldots, y_n)}{D(x_1, x_2, \ldots, x_n)} \cdot \frac{D(x_1, x_2, \ldots, x_n)}{D(t_1, t_2, \ldots, t_n)} = \frac{D(y_1, y_2, \ldots, y_n)}{D(t_1, t_2, \ldots, t_n)}. \tag{3}$$

Wäre y eine Funktion von x allein und x eine Funktion von t allein, so ergäbe sich die bekannte Kettenregel $\frac{\mathrm{d}y}{\mathrm{d}x} \cdot \frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\mathrm{d}y}{\mathrm{d}t}$. Die oben bewiesene Eigenschaft der Funktionaldeterminanten ist also eine Verallgemeinerung der Kettenregel.

Wir weisen noch besonders auf den Spezialfall hin, daß die Veränderlichen $t_1, t_2, \ldots, t_n$ mit $y_1, y_2, \ldots, y_n$ übereinstimmen, so daß das Funktionensystem (2) die Umkehrung des Systems (1) ist. (Dabei setzen wir die Möglichkeit dieser Invertierung voraus. Näheres darüber im folgenden Paragraphen.) Dann lautet unsere Beziehung

$$\frac{D(y_1, y_2, \ldots, y_n)}{D(x_1, x_2, \ldots, x_n)} \cdot \frac{D(x_1, x_2, \ldots, x_n)}{D(y_1, y_2, \ldots, y_n)} = 1$$

oder

$$\frac{D(y_1, y_2, \ldots, y_n)}{D(x_1, x_2, \ldots, x_n)} = \frac{1}{\dfrac{D(x_1, x_2, \ldots, x_n)}{D(y_1, y_2, \ldots, y_n)}}. \tag{4}$$

Darin erkennen wir die Verallgemeinerung der Differentiationsregel für die Umkehrfunktion.

204. Multiplikation von Funktionalmatrizen (Jacobischen Matrizen). Es seien m Funktionen $y_1, y_2, \ldots, y_m$ von n $(n > m)$ Veränderlichen $x_1, x_2, \ldots, x_n$ gegeben:

$$\begin{aligned} y_1 &= f(x_1, x_2, \ldots, x_n), \\ y_2 &= f_2(x_1, x_2, \ldots, x_n), \\ & \\ y_m &= f_m(x_1, x_2, \ldots, x_n); \end{aligned}$$

dabei seien die $x_1, x_2, \ldots, x_n$ Funktionen von m Veränderlichen $t_1, t_2, \ldots, t_m$:

$$\begin{aligned}
x_1 &= \varphi_1(t_1, t_2, \ldots, t_m),\\
x_2 &= \varphi_2(t_1, t_2, \ldots, t_m),\\
&\\
x_n &= \varphi_n(t_1, t_2, \ldots, t_m).
\end{aligned}$$

Wir nehmen an, daß in beiden Fällen stetige partielle Ableitungen existieren und wollen versuchen, die Funktionaldeterminante der $y_1, y_2, \ldots, y_m$ nach den $t_1, t_2, \ldots, t_m$ zu bestimmen.

In der linearen Algebra wird ein allgemeiner Matrizenmultiplikationssatz bewiesen, von dem der obige Determinantensatz ein Spezialfall ist. Sind die rechteckigen Matrizen

$$\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n}\\ a_{21} & a_{22} & \ldots & a_{2n}\\ \cdot & & & \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} b_{11} & b_{12} & \ldots & b_{1m}\\ b_{21} & b_{22} & \ldots & b_{2m}\\ \cdot & & & \\ b_{n1} & b_{n2} & \ldots & b_{nm} \end{pmatrix} \quad (n > m)$$

gegeben, so versteht man unter ihrem *Produkt* die quadratische Matrix

$$\begin{pmatrix} c_{11} & c_{22} & \ldots & c_{1m}\\ c_{21} & c_{22} & \ldots & c_{2m}\\ & & & \\ c_{m1} & c_{m2} & \ldots & c_{mm} \end{pmatrix},$$

wobei für das allgemeine Glied

$$c_{ik} = a_{i1}b_{1k} + a_{i2}b_{2k} + \cdots + a_{in}b_{nk} \quad (i, k = 1, 2, \ldots, m)$$

gilt. Die Determinante dieser Produktmatrix ist gleich

$$\sum_{(i_1, i_2, \ldots, i_m)} \begin{vmatrix} a_{1i_1} & a_{1i_2} & \ldots & a_{1i_m}\\ a_{2i_1} & a_{2i_2} & \ldots & a_{2i_m}\\ \cdot & & & \\ a_{mi_1} & a_{mi_2} & \ldots & a_{mi_m} \end{vmatrix} \cdot \begin{vmatrix} b_{i_11} & b_{i_12} & \ldots & b_{i_1m}\\ b_{i_21} & b_{i_22} & \ldots & b_{i_2m}\\ \cdot & & & \cdot\\ b_{i_m1} & b_{i_m2} & \ldots & b_{i_mm} \end{vmatrix};$$

dabei erstreckt sich die Summation über alle möglichen Kombinationen $(i_1, i_2, \ldots, i_m)$ der n Indizes $1, 2, \ldots, n$ zu je m. Wenden wir das auf die Funktionalmatrizen

$$\begin{pmatrix} \dfrac{\partial y_1}{\partial x_1} & \dfrac{\partial y_1}{\partial x_2} & \ldots & \dfrac{\partial y_1}{\partial x_n}\\ \dfrac{\partial y_2}{\partial x_1} & \dfrac{\partial y_2}{\partial x_2} & \ldots & \dfrac{\partial y_2}{\partial x_n}\\ \cdot & & & \\ \dfrac{\partial y_m}{\partial x_1} & \dfrac{\partial y_m}{\partial x_2} & \ldots & \dfrac{\partial y_m}{\partial x_n} \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} \dfrac{\partial x_1}{\partial t_1} & \dfrac{\partial x_1}{\partial t_2} & \ldots & \dfrac{\partial x_1}{\partial t_m}\\ \dfrac{\partial x_2}{\partial t_1} & \dfrac{\partial x_2}{\partial t_2} & \ldots & \dfrac{\partial x_2}{\partial t_m}\\ \cdot & & & \\ \dfrac{\partial x_n}{\partial t_1} & \dfrac{\partial x_n}{\partial t_2} & \ldots & \dfrac{\partial x_n}{\partial t_m} \end{pmatrix}$$

an, so erhalten wir

$$\begin{vmatrix} \frac{\partial y_1}{\partial x_1}\frac{\partial x_1}{\partial t_1}+\cdots+\frac{\partial y_1}{\partial x_n}\frac{\partial x_n}{\partial t_1} & \cdots & \frac{\partial y_1}{\partial x_1}\frac{\partial x_1}{\partial t_m}+\cdots+\frac{\partial y_1}{\partial x_n}\frac{\partial x_n}{\partial t_m} \\ \frac{\partial y_2}{\partial x_1}\frac{\partial x_1}{\partial t_1}+\cdots+\frac{\partial y_2}{\partial x_n}\frac{\partial x_n}{\partial t_1} & \cdots & \frac{\partial y_2}{\partial x_1}\frac{\partial x_1}{\partial t_m}+\cdots+\frac{\partial y_2}{\partial x_n}\frac{\partial x_n}{\partial t_m} \\ \cdot & & \\ \frac{\partial y_m}{\partial x_1}\frac{\partial x_1}{\partial t_1}+\cdots+\frac{\partial y_m}{\partial x_n}\frac{\partial x_n}{\partial t_1} & \cdots & \frac{\partial y_m}{\partial x_1}\frac{\partial x_1}{\partial t_m}+\cdots+\frac{\partial y_m}{\partial x_n}\frac{\partial x_n}{\partial t_m} \end{vmatrix}$$

$$= \sum_{(i_1,i_2,\ldots,i_m)} \begin{vmatrix} \frac{\partial y_1}{\partial x_{i_1}} & \frac{\partial y_1}{\partial x_{i_2}} & \cdots & \frac{\partial y_1}{\partial x_{i_m}} \\ \frac{\partial y_2}{\partial x_{i_1}} & \frac{\partial y_2}{\partial x_{i_2}} & \cdots & \frac{\partial y_2}{\partial x_{i_m}} \\ & & & \\ \frac{\partial y_m}{\partial x_{i_1}} & \frac{\partial y_m}{\partial x_{i_2}} & \cdots & \frac{\partial y_m}{\partial x_{i_m}} \end{vmatrix} \cdot \begin{vmatrix} \frac{\partial x_{i_1}}{\partial t_1} & \frac{\partial x_{i_1}}{\partial t_2} & \cdots & \frac{\partial x_{i_1}}{\partial t_m} \\ \frac{\partial x_{i_2}}{\partial t_1} & \frac{\partial x_{i_2}}{\partial t_2} & \cdots & \frac{\partial x_{i_2}}{\partial t_m} \\ & & & \cdot \\ \frac{\partial x_{i_m}}{\partial t_1} & \frac{\partial x_{i_m}}{\partial t_2} & \cdots & \frac{\partial x_{i_m}}{\partial t_m} \end{vmatrix}.$$

Erinnern wir uns der Kettenregel für Funktionen mehrerer Veränderlicher, so können wir die linke Seite in der Form

$$\begin{vmatrix} \frac{\partial y_1}{\partial t_1} & \frac{\partial y_1}{\partial t_2} & \cdots & \frac{\partial y_1}{\partial t_m} \\ \frac{\partial y_2}{\partial t_1} & \frac{\partial y_2}{\partial t_2} & \cdots & \frac{\partial y_2}{\partial t_m} \\ \cdot & & & \\ \frac{\partial y_m}{\partial t_1} & \frac{\partial y_m}{\partial t_2} & \cdots & \frac{\partial y_m}{\partial t_m} \end{vmatrix}$$

schreiben. Verwenden wir wieder die abkürzende Bezeichnung, so finden wir

$$\frac{D(y_1, y_2, \ldots, y_m)}{D(t_1, t_2, \ldots, t_m)} = \sum_{(i_1,i_2,\ldots,i_m)} \frac{D(y_1, y_2, \ldots, y_m)}{D(x_{i_1}, x_{i_2}, \ldots, x_{i_m})} \cdot \frac{D(x_{i_1}, x_{i_2}, \ldots, x_{i_m})}{D(t_1, t_2, \ldots, t_m)}, \tag{5}$$

wobei sich die Summation über alle Kombinationen zu je m aus den n Indizes $1, 2, \ldots, n$ erstreckt. Für $m = 1$ geht diese Formel in die Formel für die Differentiation einer mittelbaren Funktion (bei Verkettung mehrerer Veränderlicher) über:

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \sum_i \frac{\partial y}{\partial x_i} \cdot \frac{\mathrm{d}x_i}{\mathrm{d}t},$$

deren Verallgemeinerung sie also ist.

Wir wollen noch den Spezialfall $n = 3$, $m = 2$ anführen:

$$\frac{D(y_1, y_2)}{D(t_1, t_2)} = \frac{D(y_1, y_2)}{D(x_1, x_2)} \cdot \frac{D(x_1, x_2)}{D(t_1, t_2)} + \frac{D(y_1, y_2)}{D(x_2, x_3)} \cdot \frac{D(x_2, x_3)}{D(t_1, t_2)} + \frac{D(y_1, y_2)}{D(x_3, x_1)} \cdot \frac{D(x_3, y_1)}{D(t_1, t_2)}. \tag{6}$$

Gerade diese Formel wird häufig angewandt.

Wir haben damit für Funktionaldeterminanten eine Reihe formaler Eigenschaften hergeleitet, die denen der gewöhnlichen Ableitungen analog sind. Hierher gehört auch eine Formel, die wir in Nr. 210, Beispiel 8, aufstellen werden. Die weitestgehende Analogie zwischen Ableitungen und Funktionaldeterminanten zeigt sich aber in der Rolle, die sie in der Theorie der impliziten Funktionen (vgl. den nächsten Paragraphen) und insbesondere bei Variablensubstitutionen in Doppelintegralen, dreifachen und allgemein in mehrfachen Integralen spielen. Das werden wir in Band III sehen.

§ 2. Implizite Funktionen

205. Der Begriff der impliziten Funktion einer Veränderlichen. Wir nehmen an, die Werte der beiden Veränderlichen x und y seien durch eine Gleichung verknüpft, die, nachdem alle Glieder auf die linke Seite gebracht worden sind, im allgemeinen Fall die Gestalt

$$F(x, y) = 0 \tag{1}$$

habe. Hier ist $F(x, y)$ eine in einem bestimmten Bereich gegebene Funktion zweier Veränderlicher. Entspricht jedem x (in einem gewissen Bereich) ein (oder auch mehr als ein Wert) y, der zusammen mit x die Gleichung (1) befriedigt, so wird dadurch eine *eindeutige* (bzw. *mehrdeutige*) Funktion $y = f(x)$ definiert, für welche die Gleichung

$$F\big(x, f(x)\big) = 0 \tag{2}$$

identisch in x gilt.

Betrachten wir beispielsweise die Gleichung

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - 1 = 0; \tag{1a}$$

sie definiert offenbar y als zweideutige Funktion von x im Intervall $[-a, a]$,

$$y = \pm\frac{b}{a}\sqrt{a^2 - x^2}.$$

Setzt man diesen Ausdruck an Stelle von y in (1a) ein, so erhält man eine *Identität* in x.

Hier konnten wir für y einen sehr einfachen analytischen Ausdruck in x finden, sogar eine elementare Funktion. Das geht aber keineswegs immer. Nehmen wir die Gleichung

$$y - x - \varepsilon \sin y = 0 \quad (0 < \varepsilon < 1),$$

die wir (in anderen Bezeichnungen) schon in Nr. 83 untersuchten, so wissen wir zwar, daß dadurch y als eindeutige Funktion von x definiert ist, können diese aber in geschlossener Form nicht durch elementare Funktionen ausdrücken.

Eine Funktion $y = f(x)$ wird *implizit* genannt, wenn sie mittels einer nicht nach y aufgelösten Gleichung der Gestalt (1) gegeben ist; man nennt sie *explizit*, wenn man die unmittelbare Abhängigkeit der Veränderlichen y von der Veränderlichen x betrachtet. Selbstverständlich beziehen sich diese Termini nur auf die Art und Weise, in der eine Funktion $y = f(x)$ vorgegeben ist, und nicht etwa auf die Funktion an sich. Strenggenommen kann man die implizite und die explizite Art der Vorgabe einer Funktion nur dann mit voller Klarheit einander gegenüberstellen, wenn man unter der expliziten Vorgabe einen expliziten analytischen Ausdruck versteht; läßt man aber als explizit die Vorgabe mit Hilfe einer beliebigen Vorschrift zu (vgl. Nr. 45), so ist die Definition einer Funktion y von x mit Hilfe einer Gleichung der Gestalt (1) so gut wie jede andere.

Im einfachsten Fall, wenn die Gleichung (1) algebraisch ist, d. h., wenn $F(x, y)$ ein Polynom in x und y ist, heißt die dadurch definierte (eventuell mehrdeutige) implizite Funktion y von x *algebraisch.* Ist der Grad der Gleichung in y nicht höher als 4, so gibt es für diese algebraische Funktion einen expliziten Ausdruck in Radikalen; ist der Grad höher als 4, so gelingt diese Auflösung nur in Ausnahmefällen.

Im folgenden interessieren wir uns nur für das Problem der Existenz und Eindeutigkeit der impliziten Funktion (ebenso wie für gewisse andere Eigenschaften), jedoch unabhängig davon, ob sie durch eine analytische Formel explizit dargestellt werden kann. Übrigens ist diese Fragestellung für uns nicht neu; mit einem Spezialfall haben wir uns schon beschäftigt, als es sich nämlich darum handelte, Existenz und Eigenschaften der Umkehrfunktion zu untersuchen, welche durch die Gleichung

$$y - f(x) = 0$$

als implizite Funktion von y definiert war.

Die geometrische Behandlung dieses Problems ist lehrreich. Die Gleichung (1) liefert unter gewissen Bedingungen eine Kurve in der Ebene [beispielsweise die Gleichung (1a) bekanntlich eine Ellipse; Abb. 111]; man nennt sie dann die *implizite Kurvengleichung.*

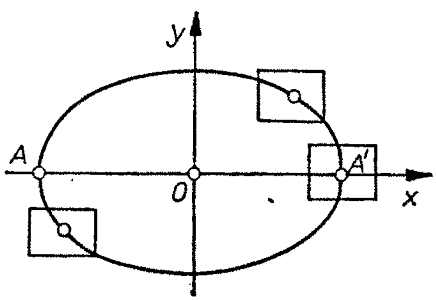

Abb. 111

Das Problem besteht nun darin, festzustellen, ob die Kurve (1) (oder Teile von ihr) durch eine gewöhnliche Funktionsgleichung der Form $y = f(x)$ beschrieben werden kann, wobei $f(x)$ eine eindeutige Funktion ist; das bedeutet geometrisch, daß die Kurve (bzw. Teile von ihr) von Parallelen zur y-Achse nur in einem einzigen Punkt geschnitten wird (werden).

Will man es nur mit eindeutigen Funktionen zu tun haben, so muß man, wie sich am Beispiel der Ellipse zeigt, nicht nur den Variationsbereich von x, sondern auch den von y gewissen Einschränkungen unterwerfen.

Wir wollen zur Abkürzung sagen, im Rechteck $(a, b; c, d)$ definiere (1) die *Veränderliche y als eindeutige Funktion von x,* wenn (1) für jeden x-Wert aus dem Intervall (a, b) eine und nur eine Lösung y im Intervall (c, d) hat.

Meist wird uns ein bestimmter Punkt (x_0, y_0) interessieren, der der Gleichung (1) genügt (auf der Kurve liegt), und als solche Rechtecke werden Umgebungen dieses Punktes auftreten. So kann man sagen, daß im Fall der Ellipse (Abb. 111) die Gleichung (1a) in jeder hinreichend kleinen Umgebung jedes Ellipsenpunktes die Ordinate y als eindeutige Funktion der Abszisse x definiert, mit Ausnahme ihrer Scheitel A und A' auf der großen Achse, wo das nicht der Fall ist.

206. Existenz der impliziten Funktion. Wir stellen nun Bedingungen auf, die die Existenz einer eindeutigen und stetigen impliziten Funktion sichern.

Satz I. *Wir nehmen an:*

1. *die Funktion* $F(x, y)$ *sei in einem rechteckigen Bereich*

$$\mathfrak{D} = [x_0 - \Delta, x_0 + \Delta; y_0 - \Delta', y_0 + \Delta']$$

mit dem Mittelpunkt (x_0, y_0) *definiert und stetig;*

2. *die Funktion* $F(x, y)$ *verschwinde in diesem Punkt,* $F(x_0, y_0) = 0$;

3. *bei konstantem* x *wachse (bzw. falle)* $F(x, y)$ *mit wachsendem* y.

Dann gilt:

a) *In einer Umgebung von* (x_0, y_0) *definiert Gleichung* (1) *die Veränderliche* y *als eindeutige Funktion von* x, $y = f(x)$;

b) *für* $x = x_0$ *nimmt diese Funktion den Wert* y_0 *an:* $f(x_0) = y_0$;

c) $f(x)$ *ist stetig.*

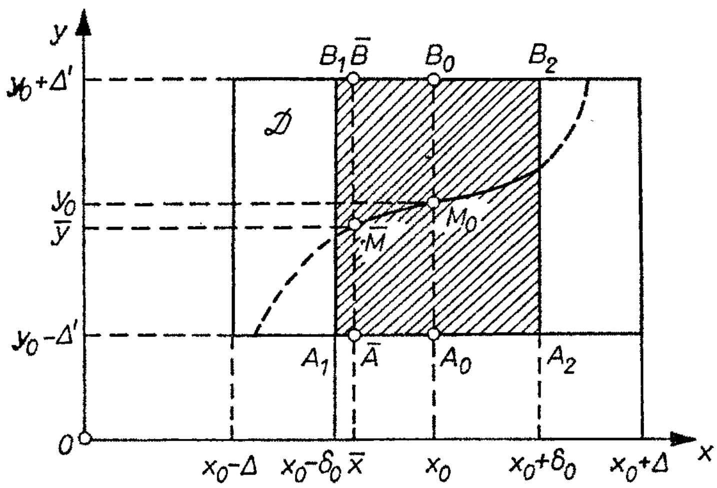

Abb. 112

Beweis. Wir bewegen uns zunächst auf der durch $M_0(x_0, y_0)$ gehenden Vertikalen (Abb. 112), d. h., wir setzen $x = x_0$. Dann wird die zu untersuchende Funktion $F(x, y)$ zu einer Funktion $F(x_0, y)$ von y allein, also eine Funktion einer Veränderlichen. Nach Voraussetzung 2 verschwindet sie für $y = y_0$. Nach Voraussetzung 3 wächst (bzw. fällt, aber hierauf wollen wir wie üblich beim Beweis nicht eingehen) die Funktion $F(x_0, y)$ mit wachsendem y, so daß ihre Werte für $y < y_0$ negativ, für $y > y_0$ dagegen positiv sind. Insbesondere hat sie daher in den Punkten $A_0(x_0, y_0 - \Delta')$ und $B_0(x_0, y_0 + \Delta')$ Werte verschiedenen Vorzeichens, und zwar ist

$$F(A_0) = F(x_0, y_0 - \Delta') < 0, \quad F(B_0) = F(x_0, y_0 + \Delta') > 0.$$

Wir untersuchen jetzt die Horizontalen durch A_0 und B_0, d. h. halten nun $y = y_0 - \Delta'$ und $y = y_0 + \Delta'$ fest. Dann erhalten wir zwei Funktionen der einen Veränderlichen x, nämlich $F(x, y_0 - \Delta')$ und $F(x, y_0 + \Delta')$, von denen, wie wir gesehen haben, die erste in $x = x_0$ negativ, die zweite positiv ist. Nach Voraussetzung 1 sind diese Funktionen stetig;[1]) daher gibt es eine Umgebung $(x_0 - \delta_0, x_0 + \delta_0)$ von x_0 $(0 < \delta_0 \leqq \Delta)$, in der beide Funktionen ihr Vorzeichen beibehalten (vgl. das Lemma in Nr. 80), so daß für das Intervall $x_0 - \delta_0 < x < x_0 + \delta_0$ die Beziehungen

$$F(x, y_0 - \Delta') < 0, \quad F(x, y_0 + \Delta') > 0$$

[1]) Wir hatten $F(x, y)$ als stetig in x, y vorausgesetzt; dann ist diese Funktion aber auch in jeder Variablen einzeln stetig.

gelten. Mit anderen Worten, auf der unteren bzw. der oberen Seite des Ausgangsrechtecks, längs der Strecken $\overline{A_1 A_2}$ und $\overline{B_1 B_2}$, welche die Länge $2\delta_0$ haben und symmetrisch zu A_0 und B_0 liegen, hat $F(x, y)$ negative bzw. positive Werte.

Nun halten wir im Intervall $(x_0 - \delta_0, x_0 + \delta_0)$ einen beliebigen Wert $x = \bar{x}$ fest und betrachten die Strecke, welche die Punkte $\bar{A}(\bar{x}, y_0 - \Delta')$ und $\bar{B}(\bar{x}, y_0 + \Delta')$ verbindet. Längs dieser Strecke ist $F(\bar{x}, y)$ wieder eine Funktion von y allein. Da sie nach Voraussetzung 1 stetig ist[1]) und an den Endpunkten des Intervalls $[y_0 - \Delta', y_0 + \Delta']$ Werte verschiedenen Vorzeichens hat,

$$F(\bar{A}) = F(\bar{x}, y_0 - \Delta') < 0, \quad F(\bar{B}) = F(\bar{x}, y_0 + \Delta') > 0,$$

muß sie nach dem Satz von Bolzano-Cauchy (Nr. 80) für einen zwischen $y_0 - \Delta'$ und $y_0 + \Delta'$ gelegenen Wert $y = \bar{y}$ den Wert 0 annehmen:

$$F(\bar{x}, \bar{y}) = 0.$$

Jetzt folgt aus Voraussetzung 3, daß für $y > \bar{y}$ die Beziehung $F(\bar{x}, y) > 0$ und für $y < \bar{y}$ die Beziehung $F(\bar{x}, y) < 0$ gilt, so daß $\bar{y}$ der einzige Wert in $(y_0 - \Delta', y_0 + \Delta')$ ist, der zusammen mit $x = \bar{x}$ die Gleichung (1) befriedigt. Auf jeder Vertikalen $\bar{A}\bar{B}$ gibt es nur einen einzigen Punkt $\bar{M}$ $(\bar{x}, \bar{y})$, für den die linke Seite 0 wird. Somit definiert in der Umgebung $(x_0 - \delta_0, x_0 + \delta_0; y_0 - \Delta', y_0 + \Delta')$ des Punktes (x_0, y_0) Gleichung (1) tatsächlich y als eindeutige Funktion $y = f(x)$ von x.

Übrigens zeigt diese Überlegung auch auf Grund von Voraussetzung 2, daß $f(x_0) = y_0$ ist. Aus $F(x_0, y_0) = 0$ ersehen wir nämlich, daß y_0 der einzige Wert von y im Intervall $(y_0 - \Delta', y_0 + \Delta')$ ist, der zusammen mit $x = x_0$ die Gleichung (1) befriedigt.

Nun muß noch die Stetigkeit von $y = f(x)$ in $(x_0 - \delta_0, x_0 + \delta_0)$ gezeigt werden. Für $x = x_0$ ergibt sich das unmittelbar aus der obigen Überlegung, die auf jedes kleinere Rechteck mit dem Mittelpunkt $M_0(x_0, y_0)$ ebenfalls anwendbar ist. Hätten wir Δ' durch $\varepsilon < \Delta'$ ersetzt, so wären wir, wie oben, zu einem $\delta \leqq \delta_0$ gekommen derart, daß für jedes x aus $(x_0 - \delta, x_0 + \delta)$ der einzige ihm entsprechende Wert y, der zusammen mit x der Gleichung (1) genügt, gerade zwischen $y_0 - \varepsilon$ und $y_0 + \varepsilon$ gelegen hätte. Somit wäre für $|x - x_0| < \delta$

$$|f(x) - y_0| = |f(x) - f(x_0)| < \varepsilon,$$

womit die Stetigkeit von $f(x)$ in $x = x_0$ bewiesen ist.

Der Beweis für einen beliebigen Punkt $x = \bar{x}$ ist für dem $x = x_0$ analog. Der Punkt $\bar{M}(\bar{x}, \bar{y})$, wobei $\bar{y} = f(\bar{x})$ ist, genügt derselben Bedingung wie der Punkt $M_0(x_0, y_0)$; denn es ist $f(\bar{x}, \bar{y}) = 0$. Daher ist, wie oben, in einer Umgebung von $\bar{M}(\bar{x}, \bar{y})$ durch Gleichung (1) die Veränderliche y als eindeutige Funktion von x definiert, die zudem in $x = \bar{x}$ stetig ist. Auf Grund der Eindeutigkeit stimmt diese Funktion mit $f(x)$ überein; damit ist die Stetigkeit von $f(x)$ in $x = \bar{x}$ bewiesen.

Wir haben somit einen Satz über die Existenz der impliziten Funktion bewiesen, ohne auf die Frage der Berechnung ihrer Werte oder ihrer analytischen Darstellung einzugehen. Damit werden wir uns in Kapitel XII befassen.

Der soeben bewiesene Satz ist offenbar eine Verallgemeinerung des Satzes aus Nr. 83.

---

[1]) Vgl. die Fußnote auf S. 414.

207. Die Differenzierbarkeit der impliziten Funktion. Jetzt verschärfen wir die Annahmen über $F(x, y)$ und können dann auch die Existenz der Ableitung der Funktion $y = f(x)$ beweisen.

Satz II. *Wir setzen voraus:*

1. *$F(x, y)$ sei im Rechteck*

$$\mathfrak{D} = [x_0 - \Delta, x_0 + \Delta; y_0 - \Delta', y_0 + \Delta']$$

mit dem Mittelpunkt (x_0, y_0) definiert und stetig;

2. *die partiellen Ableitungen F_x und F_y existieren in $\mathfrak{D}$ und sind dort stetig;*
3. *$F(x, y)$ verschwindet in (x_0, y_0), d. h. $F(x_0, y_0) = 0$;*
4. *$F_y(x_0, y_0)$ ist von 0 verschieden.*

Dann gelten die Behauptungen a), b), c) *von Satz* I *und außerdem*

d) *$f(x)$ besitzt eine stetige Ableitung.*

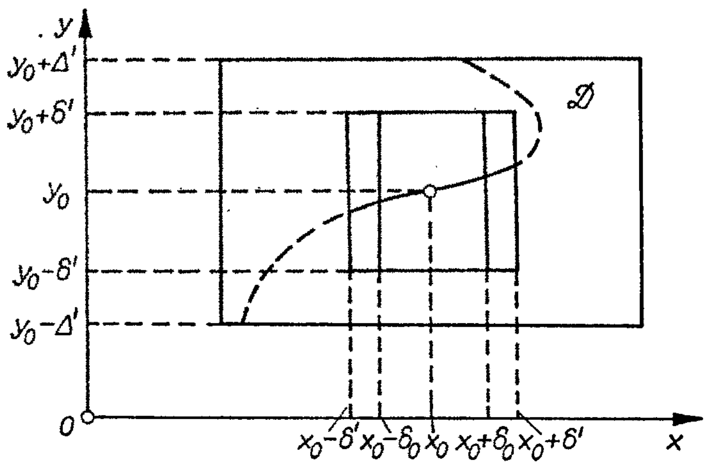

Abb. 113

Beweis (Abb. 113). Es sei etwa $F_y(x_0, y_0) > 0$. Da nach Voraussetzung 2 diese partielle Ableitung stetig ist, gibt es ein Quadrat

$$[x_0 - \delta', x_0 + \delta'; y_0 - \delta', y_0 + \delta'] \quad (\delta' < \Delta, \delta' < \Delta')$$

derart, daß in allen seinen Punkten $F_y(x, y) > 0$ ist.[1]) Dann sind für dieses Quadrat alle Voraussetzungen von Satz I erfüllt: Die Monotonie von $F(x, y)$ bei festem x folgt gerade daraus, daß $F_y > 0$ ist (vgl. Nr. 132). Daher können wir a), b), c) als bewiesen annehmen.

Um auch d) zu beweisen, wollen wir unter y gerade die durch Gleichung (1) definierte und ihr identisch genügende implizite Funktion $y = f(x)$ verstehen. Wir erteilen x den Zuwachs Δx; dem Wert $x + \Delta x$ entspricht ein Wert $y + \Delta y = f(x + \Delta x)$, der zusammen mit ihm die Gleichung (1) erfüllt: $F(x + \Delta x, y + \Delta y) = 0$. Offenbar gilt auch für den Zuwachs ΔF von F die Beziehung

$$\Delta F(x, y) = F(x + \Delta x, y + \Delta y) - F(x, y) = 0.$$

Stellen wir ΔF nach Formel (1) aus Nr. 178 dar, so erhalten wir

$$0 = \Delta F(x, y) = F_x(x, y) \cdot \Delta x + F_y(x, y) \cdot \Delta y + \alpha \, \Delta x + \beta \, \Delta y,$$

[1]) Auch für Funktionen mehrerer Veränderlicher gilt nämlich eine Aussage, die der des Lemmas aus Nr. 80 für Funktionen einer Veränderlichen analog ist.

wobei α und β von Δx und Δy abhängen und gegen 0 streben, wenn Δx und Δy (gleichzeitig) gegen 0 streben. Hieraus folgt

$$\frac{\Delta y}{\Delta x} = -\frac{F_x(x, y) + \alpha}{F_y(x, y) + \beta}.$$

Nun lassen wir Δx gegen 0 streben; auf Grund der schon bewiesenen [vgl. c)] Stetigkeit von $y = f(x)$ strebt dann auch Δy gegen 0, also gilt auch $\alpha \to \beta$, $\beta \to 0$. Wegen $F_y \neq 0$ existiert der Limes der rechten Seite, also auch die Ableitung von y nach x:

$$f'(x) = y_x = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = -\frac{F_x(x, y)}{F_y(x, y)}. \tag{3}$$

Setzen wir $f(x)$ statt y ein, so folgt

$$f'(x) = -\frac{F_x\big(x, f(x)\big)}{F_y\big(x, f(x)\big)};$$

da im Zähler und Nenner stetige Funktionen von stetigen Funktionen stehen und der Nenner nicht verschwindet, folgt hieraus, daß auch $f'(x)$ stetig ist. Damit ist der Satz bewiesen.

Es ist bemerkenswert, daß man aus Eigenschaften der Funktion $F(x, y)$ sofort auf Eigenschaften der Funktion $y = f(x)$ schließen kann, die nicht unmittelbar gegeben ist.

208. Implizite Funktionen mehrerer Veränderlicher. Analog der Gleichung (1) kann man auch eine Gleichung in mehreren Veränderlichen untersuchen:

$$F(x_1, x_2, \ldots, x_n, y) = 0. \tag{4}$$

Unter bestimmten Bedingungen wird hierdurch y als implizite Funktion der n Veränderlichen $x_1, x_2, \ldots, x_n$ definiert, $y = f(x_1, x_2, \ldots, x_n)$, die im allgemeinen mehrdeutig ist. Setzt man sie an Stelle von y ein, so folgt

$$F\big(x_1, x_2, \ldots, x_n, f(x_1, x_2, \ldots, x_n)\big) = 0$$

identisch in $x_1, x_2, \ldots, x_n$. Wir sagen, in dem $(n + 1)$-dimensionalen Parallelepiped

$$(a_1, b_1; a_2, b_2; \ldots; a_n, b_n; c, d)$$

definiere Gleichung (4) die Veränderliche y als *eindeutige Funktion von* $x_1, x_2, \ldots, x_n$, wenn für jeden Punkt $(x_1, x_2, \ldots, x_n)$ in dem n-dimensionalen Parallelepiped

$$(a_1, b_1; a_2, b_2; \ldots; a_n, b_n)$$

die Gleichung (4) eine und nur eine Lösung im Intervall (c, d) hat.

Als solches Parallelepiped wird meist eine *Umgebung* des uns interessierenden Punktes $(x_1^0, x_2^0, \ldots, x_n^0)$ fungieren.

Auf Gleichung (4) bezieht sich nun

Satz III. *Wir setzen voraus:*

1. $F(x_1, x_2, \ldots, x_n, y)$ *sei in einem* $(n + 1)$*-dimensionalen Parallelepiped*

$$\mathscr{D} = [x_1^0 - \Delta_1, x_1^0 + \Delta_1; \ldots; x_n^0 - \Delta_n, x_n^0 + \Delta_n; y_0 - \Delta', y_0 + \Delta']$$

mit dem Mittelpunkt $(x_1^0, x_2^0, \ldots, x_n^0, y_0)$ *definiert und stetig;*

2. *die partiellen Ableitungen* $F_{x_1}, \ldots, F_{x_n}, F_y$ *existieren in* $\mathfrak{D}$ *und sind dort stetig;*
3. $F(x_1^0, \ldots, x_n^0, y_0) = 0$;
4. $F_y(x_1^0, \ldots, x_n^0, y_0) \neq 0$.

Dann gilt:

a) *In einer Umgebung von* $(x_1^0, \ldots, x_n^0, y_0)$ *definiert Gleichung* (4) *die Veränderliche* y *als eindeutige Funktion* $y = f(x_1, \ldots, x_n)$ *von* $x_1, \ldots, x_n$;

b) *für* $x_1 = x_1^0, \ldots, x_n = x_0^n$ *nimmt diese Funktion den Wert* y_0 *an:* $f(x_1^0, \ldots, x_n^0) = y_0$;

c) $f(x_1, \ldots, x_n)$ *ist in bezug auf die Gesamtheit der Argumente stetig;*

d) $f(x_1, \ldots, x_n)$ *hat stetige partielle Ableitungen* $f_{x_1}, f_{x_2}, \ldots, f_{x_n}$.

Der Beweis ist dem Beweis der Sätze I und II völlig analog.

Schließlich kann im allgemeinsten Fall ein System von m Gleichungen mit $n + m$ Veränderlichen gegeben sein:

$$\left.\begin{aligned} F_1(x_1, \ldots, x_n; y_1, \ldots, y_m) &= 0, \\ F_2(x_1, \ldots, x_n; y_1, \ldots, y_m) &= 0, \\ &\;\;\vdots \\ F_m(x_1, \ldots, x_n; y_1, \ldots, y_m) &= 0. \end{aligned}\right\} \tag{5}$$

Hier handelt es sich um die Definition eines Systems von m Veränderlichen $y_1, \ldots, y_m$ als implizite Funktionen der n Veränderlichen $x_1, \ldots, x_n$,

$$y_1 = \varphi_1(x_1, \ldots; x_n), \ldots, y_m = \varphi_m(x_1, \ldots, x_n),$$

so daß beim Einsetzen in (5) die Identitäten

$$\begin{aligned} F_1\big(x_1, \ldots, x_n; \varphi_1(x_1, \ldots, x_n), \ldots, \varphi_m(x_1, \ldots, x_n)\big) &= 0, \\ F_2\big(x_1, \ldots, x_n; \varphi_1(x_1, \ldots, x_n), \ldots, \varphi_m(x_1, \ldots, x_n)\big) &= 0, \\ &\;\;\vdots \\ F_m\big(x_1, \ldots, x_n; \varphi_1(x_1, \ldots, x_n), \ldots, \varphi_m(x_1, \ldots, x_n)\big) &= 0 \end{aligned}$$

entstehen.

Man sagt, im $(n + m)$-dimensionalen Parallelepiped

$$(a_1, b_1; \ldots; a_n, b_n; c_1, d_1; \ldots; c_m, d_m)$$

definiere das System (5) die Veränderlichen $y_1, y_2, \ldots, y_m$ als *eindeutige Funktionen von* $x_1, \ldots, x_n$, wenn für jeden Punkt $(x_1, \ldots, x_n)$ im n-dimensionalen Parallelepiped

$$(a_1, b_1; \ldots; a_n, b_n)$$

das System (5) ein und nur ein System von Lösungen $y_1, \ldots, y_m$ besitzt, die dem m-dimensionalen Parallelepiped

$$(c_1, d_1; \ldots; c_m, d_m)$$

angehören.

Beim Problem der Existenz einer eindeutigen impliziten Funktion, die durch eine einzige Gleichung (1) oder (4) definiert wird, spielt, wie wir gesehen haben, die Forderung, daß in dem betrachteten Punkt, der der Gleichung der impliziten Funktion genügt, die Ableitung F_y (also die Ableitung nach derjenigen Veränderlichen, welche als implizite Funktion definiert werden soll) von 0 verschieden ist, die entscheidende Rolle.

Entsprechend dem über die Funktionaldeterminante Gesagten ist es nicht überraschend, daß es bei unserem allgemeinen Problem auf die Funktionaldeterminante ankommt. Bei dem Problem der Existenz eindeutiger impliziter Funktionen $y_1, \ldots, y_m$, die durch ein Gleichungssystem (5) definiert werden, darf die Funktionaldeterminante der links stehenden Funktionen nach den Veränderlichen $y_1, \ldots, y_m$,

$$J = \frac{D(F_1, \ldots, F_m)}{D(y_1, \ldots, y_m)} = \begin{vmatrix} \frac{\partial F_1}{\partial y_1} & \frac{\partial F_1}{\partial y_2} & \cdots & \frac{\partial F_1}{\partial y_{m-1}} & \frac{\partial F_1}{\partial y_m} \\ \frac{\partial F_2}{\partial y_1} & \frac{\partial F_2}{\partial y_2} & \cdots & \frac{\partial F_2}{\partial y_{m-1}} & \frac{\partial F_2}{\partial y_m} \\ & & & & \\ \frac{\partial F_{m-1}}{\partial y_1} & \frac{\partial F_{m-1}}{\partial y_2} & \cdots & \frac{\partial F_{m-1}}{\partial y_{m-1}} & \frac{\partial F_{m-1}}{x y_m} \\ \frac{\partial F_m}{\partial y_1} & \frac{\partial F_m}{\partial y_2} & \cdots & \frac{\partial F_m}{\partial y_{m-1}} & \frac{\partial F_m}{\partial y_m} \end{vmatrix} \tag{6}$$

im Punkt $(x_1^0, \ldots, x_n^0, y_1^0, \ldots, y_m^0)$ nicht verschwinden.

Satz IV. *Wir setzen voraus:*

1. *Alle Funktionen $F_1, \ldots, F_m$ seien in einem $(n + m)$-dimensionalen rechtwinkligen Parallelepiped*

$$\mathfrak{D} = [x_1^0 - \Delta_1, x_1^0 + \Delta_1; \ldots; x_n^0 - \Delta_n, x_n^0 + \Delta_n;$$
$$y_1^0 - \Delta_1', y_1^0 + \Delta_1'; \ldots; y_m^0 - \Delta_m', y_m^0 + \Delta_m']$$

mit dem Mittelpunkt $(x_1^0, \ldots, x_n^0, y_1^0, \ldots, y_m^0)$ definiert und stetig;

2. *in $\mathfrak{D}$ seien die partiellen Ableitungen dieser Funktionen nach allen Argumenten vorhanden und stetig;*

3. *der Punkt $(x_1^0, \ldots, y_m^0)$ genüge dem System* (5);

4. *die Funktionaldeterminante J sei in diesem Punkt von* 0 *verschieden.*

Dann gilt:

a) *In einer Umgebung des Punktes $(x_1^0, \ldots, y_m^0)$ definiert das System* (5) *die Veränderlichen $y_1, \ldots, y_m$ als eindeutige Funktionen von $x_1, \ldots, x_n$:*

$$y_1 = f_1(x_1, \ldots, x_n), \ldots, y_{m-1} = f_{m-1}(x_1, \ldots, x_n), y_m = f_m(x_1, \ldots, x_n);$$

b) *für die Werte $x_1 = x_0^1, \ldots, x_n = x_n^0$ nehmen diese Funktionen die Werte $y_1^0, \ldots, y_{m-1}^0$, y_m^0 an:*

$$f_1(x_1^0, \ldots, x_n^0) = y_1^0, \ldots, f_{m-1}(x_1^0, \ldots, x_n^0) = y_{m-1}^0, f_m(x_1^0, \ldots, x_n^0) = y_m^0;$$

c) *die Funktionen $f_1, \ldots, f_m$ sind stetig;*

d) *die Funktionen $f_1, \ldots, f_m$ haben stetige partielle Ableitungen nach allen Argumenten.*

Beweis. Den Beweis führen wir durch Induktion. Für $m = 1$, wenn das System aus einer einzigen Gleichung besteht, ist der Satz richtig. Dann handelt es sich einfach um den in Satz III erledigten Fall. Wir nehmen nun an, der Satz sei richtig, wenn das System aus $m-1$ Gleichungen besteht und es darum geht, $m-1$ implizite Funktionen zu definieren; auf Grund dieser Annahme beweisen wir ihn für ein System von m Gleichungen.

Da die Funktionaldeterminante J im Punkt $(x_1^0, \ldots, y_m^0)$ von 0 verschieden ist, gibt es in ihrer letzten Spalte mindestens ein in diesem Punkt ebenfalls von 0 verschiedenes Element; es sei also beispielsweise

$$\frac{\partial F_m(x_1^0, \ldots, y_m^0)}{\partial y_m} \neq 0.$$

Dann definiert nach Satz III die letzte Gleichung des Systems (5) in einer Umgebung $\mathfrak{D}^*$ des Punktes $(x_1^0, \ldots, y_m^0)$ die Veränderliche y_m als eindeutige Funktion der übrigen Argumente,

$$y_m = \varphi(x_1, \ldots, x_n; y_1, \ldots, y_{m-1}), \tag{7}$$

so daß in diesen Argumenten identisch

$$F_m\big(x_1, \ldots, x_n; y_1, \ldots, y_{m-1}, \varphi(x_1, \ldots, y_{m-1})\big) = 0 \tag{8}$$

gilt. Diese Funktion φ ist stetig und hat stetige partielle Ableitungen; außerdem ist

$$\varphi(x_1^0, \ldots, x_n^0; y_1^0, \ldots, y_{m-1}^0) = y_m^0. \tag{9}$$

Es ist wichtig, darauf hinzuweisen, daß die Gleichung

$$F_m(x_1, \ldots, x_n; y_1, \ldots, y_n) = 0$$

der Gleichung (7) äquivalent ist, da wir uns von vornherein auf die Umgebung $\mathfrak{D}^*$ beschränken: In $\mathfrak{D}^*$ genügen ihr ein und dieselben Wertesysteme der Veränderlichen $x_1, \ldots, x_n, y_1, \ldots, y_m$.

Wenn wir die letzte der Gleichungen (5) durch diese Gleichung (7) ersetzen und φ an Stelle von y_m in die übrigen Gleichungen des Systems (5) einsetzen, so erhalten wir ein neues System von $m-1$ Gleichungen mit $n+m-1$ Veränderlichen:

$$\left.\begin{aligned} &\Phi_1(x_1, \ldots, x_n; y_1, \ldots, y_{m-1}) = 0, \\ &\Phi_2(x_1, \ldots, x_n; y_1, \ldots, y_{m-1}) = 0, \\ &\qquad \cdot \\ &\Phi_{m-1}(x_1, \ldots, x_n; y_1, \ldots, y_{m-1}) = 0; \end{aligned}\right\} \tag{10}$$

dabei wurde zur Abkürzung (für $j = 1, 2, \ldots, m-1$)

$$\Phi_j(x_1, \ldots, x_n; y_1, \ldots, y_{m-1}) = F_j\big(x_1, \ldots, x_n; y_1, \ldots, y_{m-1}, \varphi(x_1, \ldots, y_{m-1})\big) \tag{11}$$

gesetzt.

Bleibt man im Bereich $\mathfrak{D}^*$, so ist das System (5) dem System (10) mit der Zusatzgleichung (7) äquivalent.

Wenn wir also noch beweisen können, daß durch das System (10) in einer hinreichend kleinen Umgebung d^* des Punktes $(x_1^0, \ldots, y_{m-1}^0)$ die $m-1$ Veränderlichen $y_1, \ldots, y_{m-1}$ als eindeutige Funktionen von $x_1, \ldots, x_n$ definiert werden,

$$y_1 = f_1(x_1, \ldots, x_n), \ldots, y_{m-1} = f_{m-1}(x_1, \ldots, x_n), \tag{12}$$

so ist nach (7) auch die Veränderliche y_m als ebensolche eindeutige Funktion definiert,

$$\begin{aligned} y_m &= f_m(x_1, \ldots, x_n) \\ &= \varphi\bigl(x_1, \ldots, x_n; f_1(x_1, \ldots, x_n), \ldots, f_{m-1}(x_1, \ldots, x_n)\bigr), \end{aligned} \tag{12a}$$

und die Behauptung a) wäre damit völlig bewiesen.[1])

Wir wenden uns nun wieder dem System (10) zu und beweisen, daß dafür in der Umgebung des Punktes $(x_1^0, \ldots, y_{m-1}^0)$ die Bedingungen erfüllt sind, die den Voraussetzungen 1 bis 4 entsprechen. Für die ersten beiden folgt das unmittelbar aus den Eigenschaften der Funktionen F_j und φ, und zwar auf Grund von (11). Ebenso liefert die Voraussetzung 3 zusammen mit (11) und (9) für $j = 1, \ldots, m-1$ die Beziehungen

$$\begin{aligned} \Phi_j(y_1^0, \ldots, y_{m-1}^0) &= F_j\bigl(x_1^0, \ldots, y_{m-1}^0, \varphi(x_1^0, \ldots, y_{m-1}^0)\bigr) \\ &= F_j(x_1^0, \ldots, y_{m-1}^0, y_m^0) = 0. \end{aligned}$$

Nun müssen wir noch die Funktionaldeterminante

$$J^* = \frac{D(\Phi_1, \ldots, \Phi_{m-1})}{D(y_1, \ldots, y_{m-1})} = \begin{vmatrix} \dfrac{\partial \Phi_1}{\partial y_1} & \dfrac{\partial \Phi_1}{\partial y_2} & \cdots & \dfrac{\partial \Phi_1}{\partial y_{m-1}} \\ \dfrac{\partial \Phi_2}{\partial y_1} & \dfrac{\partial \Phi_2}{\partial y_2} & \cdots & \dfrac{\partial \Phi_2}{\partial y_{m-1}} \\ \cdot & & & \cdot \\ \dfrac{\partial \Phi_{m-1}}{\partial y_1} & \dfrac{\partial \Phi_{m-1}}{\partial y_2} & \cdots & \dfrac{\partial \Phi_{m-1}}{\partial y_{m-1}} \end{vmatrix}$$

(das Analogon von J) betrachten und uns davon überzeugen, daß sie im Punkt $(x_1^0, \ldots, y_{m-1}^0)$ von 0 verschieden ist. Zu diesem Zweck transformieren wir die Determinante J, indem wir zu den Elementen ihrer ersten $m-1$ Spalten die mit $\dfrac{\partial \varphi}{\partial y_1}, \ldots, \dfrac{\partial \varphi}{\partial y_{m-1}}$ multiplizierten Elemente ihrer m-ten Spalte addieren:

$$J = \begin{vmatrix} \dfrac{\partial F_1}{\partial y_1} + \dfrac{\partial F_1}{\partial y_m} \cdot \dfrac{\partial \varphi}{\partial y_1} & \cdots & \dfrac{\partial F_1}{\partial y_{m-1}} + \dfrac{\partial F_1}{\partial y_m} \cdot \dfrac{\partial \varphi}{\partial y_{m-1}} & \dfrac{\partial F_1}{\partial y_m} \\ \dfrac{\partial F_2}{\partial y_1} + \dfrac{\partial F_2}{\partial y_m} \cdot \dfrac{\partial \varphi}{\partial y_1} & \cdots & \dfrac{\partial F_2}{\partial y_{m-1}} + \dfrac{\partial F_2}{\partial y_m} \cdot \dfrac{\partial \varphi}{\partial y_{m-1}} & \dfrac{\partial F_2}{\partial y_m} \\ & & & \\ \dfrac{\partial F_{m-1}}{\partial y_1} + \dfrac{\partial F_{m-1}}{\partial y_m} \cdot \dfrac{\partial \varphi}{\partial y_1} & \cdots & \dfrac{\partial F_{m-1}}{\partial y_{m-1}} + \dfrac{\partial F_{m-1}}{\partial y_m} \cdot \dfrac{\partial \varphi}{\partial y_{m-1}} & \dfrac{\partial F_{m-1}}{\partial y_m} \\ \dfrac{\partial F_m}{\partial y_1} + \dfrac{\partial F_m}{\partial y_m} \cdot \dfrac{\partial \varphi}{\partial y_1} & \cdots & \dfrac{\partial F_m}{\partial y_{m-1}} + \dfrac{\partial F_m}{\partial y_m} \cdot \dfrac{\partial \varphi}{\partial y_{m-1}} & \dfrac{\partial F_m}{\partial y_m} \end{vmatrix}$$

[1]) Wir weisen darauf hin, daß das $(n + m - 1)$-dimensionale (offene) Parallelepiped d^* als so klein angenommen wird, daß die es definierenden Intervalle in den entsprechenden Intervallen enthalten sind, welche das $(n + m)$-dimensionale Parallelepiped $\mathfrak{D}^*$ definieren. Diejenige Umgebung des Punktes $(x_0^1, \ldots, y_m^0)$, von der in der Behauptung a) die Rede ist, wird gerade durch alle Intervalle bestimmt, die mit d^* verknüpft sind, wenn man dazu noch das letzte der Intervalle hinzunimmt, welche mit $\mathfrak{D}^*$ verknüpft sind.

Denkt man sich hierin $y_m = \varphi(x_1, \ldots, x_{m-1})$ gesetzt, so zeigt sich, daß alle Elemente, außer denen der letzten Zeile und denen der letzten Spalte, die partiellen Ableitungen der Funktionen Φ_j nach $y_1, \ldots, y_{m-1}$ sind. Differenzieren wir nämlich gemäß (11) die Funktion Φ_j als mittelbare Funktion nach $y_1, \ldots, y_{m-1}$ (unter Benutzung der Regel aus Nr. 181), so erhalten wir für $j = 1, \ldots, m-1$

$$\frac{\partial \Phi_j}{\partial y_1} = \frac{\partial F_j}{\partial y_1} + \frac{\partial F_j}{\partial y_m} \cdot \frac{\partial \varphi}{\partial y_1}, \ldots, \frac{\partial \Phi_j}{\partial y_{m-1}} = \frac{\partial F_j}{\partial y_{m-1}} + \frac{\partial F_j}{\partial y_m} \cdot \frac{\partial \varphi}{\partial y_{m-1}}.$$

Differenzieren wir andererseits die Identität (8) nach $y_1, \ldots, y_{m-1}$, so ergibt sich[1])

$$\frac{\partial F_m}{\partial y_1} + \frac{\partial F_m}{\partial y_m} \cdot \frac{\partial \varphi}{\partial y_1} = 0, \ldots, \frac{\partial F_m}{\partial y_{m-1}} + \frac{\partial F_m}{\partial y_m} \cdot \frac{\partial \varphi}{\partial y_{m-1}} = 0.$$

Somit sind die Elemente der letzten Zeile (außer dem letzten) sämtlich gleich 0. Es ist also

$$J = \begin{vmatrix} \dfrac{\partial \Phi_1}{\partial y_1} & \cdots & \dfrac{\partial \Phi_1}{\partial y_{m-1}} & \dfrac{\partial F_1}{\partial y_m} \\ \dfrac{\partial \Phi_2}{\partial y_1} & \cdots & \dfrac{\partial \Phi_2}{\partial y_{m-1}} & \dfrac{\partial F_2}{\partial y_m} \\ \cdot & & & \cdot \\ \dfrac{\partial \Phi_{m-1}}{\partial y_1} & \cdots & \dfrac{\partial \Phi_{m-1}}{\partial y_{m-1}} & \dfrac{\partial F_{m-1}}{\partial y_m} \\ 0 & & 0 & \dfrac{\partial F_m}{\partial y_m} \end{vmatrix}.$$

Entwickelt man diese Determinante nach den Elementen der letzten Zeile, so folgt

$$J = J^* \cdot \frac{\partial F_m}{\partial y_m}.$$

Schließlich setzen wir hierin $x_1 = x_1^0, \ldots, y_{m-1} = y_{m-1}^0$. Dann ergibt sich, daß $y_m = \varphi(x_1, \ldots, y_{m-1})$ nach (9) gleich y_m^0 ist. Da dann nach Voraussetzung 4 Determinante J von 0 verschieden ist, gilt das auch für J^*, was zu beweisen war.

Für das System (10), das $m = 1$ Gleichungen enthält, war unser Satz als richtig angenommen worden. Daher definiert dieses System in der Umgebung des Punktes $(x_1^0, \ldots, y_{m-1}^0)$ eindeutige und stetige Funktionen (12), die stetige partielle Ableitungen haben. Außerdem genügen diese Funktionen der Bedingung b):

$$f_1(x_1^0, \ldots, x_n^0) = y_1^0, \ldots, f_{m-1}(x_1^0, \ldots, x_0^1) = y_{m-1}^0. \tag{13}$$

Hieraus folgt, daß die m-te Funktion (12a) ebenfalls stetig ist und stetige Ableitungen besitzt, und schließlich, unter Berücksichtigung von (13) und (9),

$$\begin{aligned} f_m(x_1^0, \ldots, x_n^0) &= \varphi\big(x_1^0, \ldots, x_n^0; f_1(x_1^0, \ldots, x_n^0), \ldots, f_{m-1}(x_1^0, \ldots, x_n^0)\big) \\ &= \varphi(x_1^0, \ldots, x_n^0; y_1^0, \ldots, y_{m-1}^0) = y_m^0. \end{aligned}$$

Damit ist der Satz bewiesen.

[1]) Da nämlich die in (8) links stehende Funktion identisch 0 ist, sind auch ihre Ableitungen nach jedem Argument gleich 0.

Bemerkung. Wir machen den Leser auf den *lokalen* Charakter aller Sätze über die Existenz impliziter Funktionen aufmerksam: *Es handelt sich immer nur um eine gewisse Umgebung des betrachteten Punktes.* Aber auch in dieser Gestalt sind diese Sätze nützlich. Beispielsweise wird sich das in Kapitel VII zeigen, wo es bei der Untersuchung der Eigenschaften des geometrischen Bildes in einem gegebenen Punkt völlig genügt, sich auf seine unmittelbare Umgebung zu beschränken.

209. Berechnung der Ableitungen impliziter Funktionen. Die Überlegungen, mit deren Hilfe wir die Sätze über die Existenz der impliziten Funktionen bewiesen haben, vermittelten im allgemeinen keine Vorstellung davon, wie man die Ableitungen (erster Ordnung) implizierter Funktionen berechnen kann. Von höheren Ableitungen war überhaupt nicht die Rede. Jetzt wollen wir uns mit diesen wichtigen Fragen beschäftigen.

Wir beginnen mit dem einfachsten Fall, daß eine Gleichung (1) gegeben ist. Es seien in der Umgebung eines zu untersuchenden Punktes die Voraussetzungen des Satzes II erfüllt; als besonders wesentlich wird sich die Bedingung $F_y \neq 0$ herausstellen.

Wir geben ein einfaches Verfahren zur Berechnung von y_x an (wenn die Existenz schon bekannt ist). Bekanntlich geht die Gleichung (1), wenn man darin die implizite Funktion $y = f(x)$ einsetzt, in eine Identität über [vgl. Nr. 205, Formel (2)]. Wenn wir also unter y diese Funktion von x verstehen, so ist die linke Seite von (1), also $F(x, y)$, eine mittelbare Funktion von x, die identisch 0 ist. Dann ist auch ihre Ableitung nach x gleich 0. Differenziert man F nach der Regel aus Nr. 181, so ergibt sich[1])

$$F_x(x, y) + F_y(x, y) \cdot y_x = 0, \tag{14}$$

also, da $F_y \neq 0$ ist,

$$y_x = -\frac{F_x(x, y)}{F_y(x, y)}; \tag{15}$$

damit haben wir eine uns schon bekannte Formel erhalten [vgl. Nr. 207, Formel (3)].

Jetzt können wir weiter schließen. Hat $F(x, y)$ stetige Ableitungen zweiter Ordnung, so kann der in (15) auf der rechten Seite stehende Ausdruck nach x differenziert werden; dann existiert also auch die Ableitung von y_x nach x, d. h. die zweite Ableitung y_{xx} der impliziten Funktion y. Führt man die Differentiation aus und setzt jedesmal für y_x den Ausdruck (15) ein, so folgt

$$\begin{aligned} y_{xx} &= \frac{(F_{xy} + F_{yy} \cdot y_x) \cdot F_x - (F_{xx} + F_{xy} \cdot y_x) F_y}{F_y^2} \\ &= \frac{2F_x F_y F_{xy} - F_y^2 F_{xx} - F_x^2 F_{yy}}{F_y^3}. \end{aligned}$$

Die zweite Ableitung ist also wieder eine stetige Funktion von x.

Hat $F(x, y)$ stetige Ableitungen dritter Ordnung, so existiert offenbar auch die dritte Ableitung y_{xxx} von y. Der Ausdruck dafür ergibt sich unmittelbar durch Differentiation von y_{xx}, usw. Durch Induktion läßt sich zeigen, daß *die Existenz stetiger Ableitungen von $F(x, y)$ bis zur Ordnung k ($k > 1$) einschließlich auch die Existenz einer stetigen k-ten Ableitung der impliziten Funktion gewährleistet.*

[1]) Im Grunde haben wir diese Überlegungen schon früher angestellt; vgl. die Fußnote auf S. 422.

Nachdem somit die Existenz der sukzessiven Ableitungen der impliziten Funktion bewiesen ist, lassen sie sich einfach durch wiederholtes Differenzieren der Identität (14) berechnen; dabei ist zu berücksichtigen, daß y eine Funktion von x ist. Beispielsweise liefert die erste Differentiation dieser Identität die Beziehung

$$F_{xx} + F_{xy} \cdot y_x + (F_{xy} + F_{yy} \cdot y_x) \cdot y_x + F_y \cdot y_{xx} = 0, \tag{16}$$

also (wegen $F_y \neq 0$)

$$y_{xx} = -\frac{F_{xx} + 2F_{xy} \cdot y_x + F_{yy} \cdot y_x^2}{F_y}.$$

Setzen wir hier für y_x den Ausdruck (15) ein, so erhalten wir den obigen Ausdruck für y_{xx}, usw.

Analog liegt die Sache bei der Gleichung (4) mit mehreren Veränderlichen. Hier nehmen wir die Voraussetzungen des Satzes III als erfüllt an. Verstehen wir unter y die durch (4) definierte implizite Funktion, so geht (4) in eine Identität über. Halten wir $x_2, \ldots, x_n$ fest und betrachten y als Funktion von x_1 allein, so erhalten wir durch Differentiation der Identität nach x_1 die Beziehung

$$F_{x_1} + F_y \cdot y_{x_1} = 0, \quad \text{also} \quad y_{x_1} = -\frac{F_{x_1}}{F_y};$$

ebenso erhält man

$$y_{x_2} = -\frac{F_{x_2}}{F_y}, \ldots, y_{x_n} = -\frac{F_{x_n}}{F_y}, \quad \text{usw.}$$

Braucht man alle Ableitungen erster, zweiter, ... Ordnung, so ist es einfacher, gleich die Differentiale $\mathrm{d}y, \mathrm{d}^2y, \ldots$ zu berechnen. Bildet man nämlich das vollständige Differential der Identität, d. h., setzt man das vollständige Differential ihrer linken Seite gleich 0 (unter Benutzung der Invarianz des ersten Differentials; Nr. 185), so folgt

$$\frac{\partial F}{\partial x_1}\,\mathrm{d}x_1 + \frac{\partial F}{\partial x_2}\,\mathrm{d}x_2 + \cdots + \frac{\partial F}{\partial x_n}\,\mathrm{d}x_n + \frac{\partial F}{\partial y}\,\mathrm{d}y = 0,$$

also

$$\mathrm{d}y = -\frac{\dfrac{\partial F}{\partial x_1}}{\dfrac{\partial F}{\partial y}}\,\mathrm{d}x_1 - \cdots - \frac{\dfrac{\partial F}{\partial x_n}}{\dfrac{\partial F}{\partial y}}\,\mathrm{d}x_n.$$

Nun ist aber

$$\mathrm{d}y = \frac{\partial y}{\partial x_1}\,\mathrm{d}x_1 + \cdots + \frac{\partial y}{\partial x_n}\,\mathrm{d}x_n.$$

Da die $\mathrm{d}x_1, \ldots, \mathrm{d}x_n$ willkürlich sind, folgt hieraus

$$\frac{\partial y}{\partial x_1} = -\frac{\dfrac{\partial F}{\partial x_1}}{\dfrac{\partial F}{\partial y}}, \ldots, \frac{\partial y}{\partial x_n} = -\frac{\dfrac{\partial F}{\partial x_n}}{\dfrac{\partial F}{\partial y}}$$

in Übereinstimmung mit dem Obigen. Durch weitere Differentiation ergibt sich

$$\left[\frac{\partial^2 F}{\partial x_1^2}\,\mathrm{d}x_1 + \cdots + \frac{\partial^2 F}{\partial x_1\,\partial x_n}\,\mathrm{d}x_n + \frac{\partial^2 F}{\partial x_1\,\partial y}\,\mathrm{d}y\right]\mathrm{d}x_1 + \cdots + \frac{\partial F}{\partial y}\,\mathrm{d}^2 y = 0,$$

woraus sich

$$\frac{\partial^2 y}{\partial x_1^2},\quad \frac{\partial^2 y}{\partial x_1\,\partial x_2},\quad \frac{\partial^2 y}{\partial x_2^2},\ldots$$

usw. berechnen lassen.

Wir sehen, daß bei allen diesen Berechnungen tatsächlich die Bedingung

$$F_y = \frac{\partial F}{\partial y} \neq 0$$

eine entscheidende Rolle spielt.

Wir wollen nun das Gleichungssystem (5) untersuchen. Dabei nehmen wir an, in der Umgebung eines festen Punktes seien die Voraussetzungen von Satz IV erfüllt. Wieder weisen wir auf die Bedeutung der Voraussetzung $J \neq 0$ hin.

Die impliziten Funktionen $y_1, \ldots, y_m$ haben partielle Ableitungen nach $x_1, \ldots, x_n$. Man berechnet sie, indem man die Identität differenziert, welche sich aus (5) ergeben, wenn man unter $y_1, \ldots, y_m$ gerade diese impliziten Funktionen versteht. Beispielsweise liefert die Differentiation nach x_1 die Beziehungen

$$\begin{cases} \dfrac{\partial F_1}{\partial x_1} + \dfrac{\partial F_1}{\partial y_1}\cdot\dfrac{\partial y_1}{\partial x_1} + \cdots + \dfrac{\partial F_1}{\partial y_m}\cdot\dfrac{\partial y_m}{\partial x_1} = 0, \\[3ex] \dfrac{\partial F_m}{\partial x_1} + \dfrac{\partial F_m}{\partial y_1}\cdot\dfrac{\partial y_1}{\partial x_1} + \cdots + \dfrac{\partial F_m}{\partial y_m}\cdot\dfrac{\partial y_m}{\partial y_1} = 0. \end{cases}$$

Das ist ein lineares Gleichungssystem in $\dfrac{\partial y_1}{\partial x_1}, \ldots, \dfrac{\partial y_m}{\partial x_1}$ mit der von 0 verschiedenen Determinante

$$J = \frac{D(F_1, \ldots, F_m)}{D(y_1, \ldots, y_m)}.$$

Hieraus folgt

$$\frac{\partial y_1}{\partial x_1} = -\frac{\dfrac{D(F_1, \ldots, F_m)}{D(x_1, \ldots, y_m)}}{\dfrac{D(F_1, \ldots, F_m)}{D(y_1, \ldots, y_m)}}, \ldots, \frac{\partial y_m}{\partial x_1} = -\frac{\dfrac{D(F_1, \ldots, F_m)}{D(y_1, \ldots, x_1)}}{\dfrac{D(F_1, \ldots, F_m)}{D(y_1, \ldots, y_m)}}.$$

Analoge Ausdrücke erhält man für die Ableitungen von $y_1, \ldots, y_m$ nach $x_2, \ldots, x_n$.

Besitzen die Funktionen $F_1, \ldots, F_m$ stetige partielle Ableitungen zweiter Ordnung, so haben die rechten Seiten aller dieser Formeln (stetige) Ableitungen nach allen Argumenten, also existieren stetige zweite Ableitungen der impliziten Funktionen. Allgemein gilt, wie man leicht induktiv beweist: *Besitzen die Funktionen $F_1, \ldots, F_m$ stetige Ableitungen bis zur Ordnung k einschließlich, so besitzen auch die impliziten Funktionen stetige Ableitungen (nach allen Argumenten) bis zur Ordnung k einschließlich.*

Man berechnet die Ableitungen der impliziten Funktionen auch im allgemeinen Fall entweder durch Differentiation der Identität (5) nach den entsprechenden Veränderlichen oder durch Bildung der totalen Differentiale. Das sich zur Ermittlung der Ableitungen oder der Differentiale ergebende Gleichungssystem ist linear, und seine Determinante ist die als von 0 verschieden vorausgesetzte Funktionaldetermiante J. Diese Bemerkungen werden an den Beispielen deutlich.

210. Beispiele.

1. Die Veränderlichen x und y seien durch die Gleichung

$$\ln \sqrt{x^2 + y^2} = \arctan \frac{y}{x}$$

verknüpft. Wenn wir nach x differenzieren (und dabei y als Funktion von x auffassen), so erhalten wir

$$\frac{x + yy'}{x^2 + y^2} = \frac{xy' - y}{x^2 + y^2} \quad \text{oder} \quad x + yy' = xy' - y;$$

ferner

$$1 + y'^2 + yy'' = xy'';$$

usw. Aus der ersten Gleichung folgt

$$y' = \frac{x + y}{x - y},$$

aus der zweiten (wenn man den Ausdruck für y' einsetzt)

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

usw.

2. Es sei die Gleichung $F(x, y) = x^3 + y^3 - 3axy = 0$ gegeben. Man bestimme die Extrema der dadurch definierten impliziten Funktion $y(x)$.

Hier ist

$$F_x = 3(x^2 - ay), \qquad F_y = 3(y^2 - ax).$$

Nach (15) muß, damit $y_x = 0$ wird, $F_x = 0$ sein. Lösen wir das System der Gleichungen $F = 0$ und $F_x = 0$, so finden wir zwei Wertepaare x, y:

$$x = 0, \qquad y = 0 \quad \text{und} \quad x = a\sqrt[3]{2}, \qquad y = a\sqrt[3]{4}.$$

Im ersten Punkt verschwindet aber auch F_y, so daß wir nicht behaupten können, in seiner Umgebung werde y durch die gegebene Gleichung als eindeutige Funktion von x definiert. Daher lassen wir den Punkt (0, 0) beiseite.

Im zweiten Punkt ist $F_y = 3a^2\sqrt[3]{2} > 0$, also ist Satz II anwendbar. Um uns vom Vorliegen eines Extremums zu überzeugen, berechnen wir y_{xx} für $x = a\sqrt[3]{2}$. Am einfachsten gehen wir von (16) aus, indem wir dort $y_x = 0$ setzen[1]):

$$y_{xx} = -\frac{F_{xx}}{F_y}.$$

Da für $x = a\sqrt[3]{2}$ die Bezeichnung $F_{xx} = 6x > 0$ gilt, ist $y_{xx} < 0$, und es liegt ein Maximum vor.

---

[1]) Das ist nicht der allgemeinste Ausdruck für y_{xx}. Er gilt nur in dem uns interessierenden Punkt $\left(a\sqrt[3]{2}, a\sqrt[3]{4}\right)$.

3. Die implizite Funktion $z(x, y)$ sei durch die Gleichung

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

gegeben. Hier finden wir nacheinander

$$\frac{x\,\mathrm{d}x}{a^2} + \frac{y\,\mathrm{d}y}{b^2} + \frac{z\,\mathrm{d}z}{c^2} = 0, \quad \mathrm{d}z = -\frac{c^2x}{a^2z}\,\mathrm{d}x - \frac{c^2y}{b^2z}\,\mathrm{d}y,$$

also

$$\frac{\partial z}{\partial x} = -\frac{c^2x}{a^2z}, \quad \frac{\partial z}{\partial y} = -\frac{c^2y}{b^2z}.$$

Ferner ist

$$\frac{\mathrm{d}x^2}{a^2} + \frac{\mathrm{d}y^2}{b^2} + \frac{\mathrm{d}z^2}{c^2} + \frac{z\,\mathrm{d}^2z}{c^2} = 0;$$

hieraus folgt, wenn man den Ausdruck für $\mathrm{d}z$ einsetzt,

$$\mathrm{d}^2z = -\frac{c^4}{z^3}\left[\left(\frac{x^2}{a^2} + \frac{z^2}{c^2}\right)\frac{\mathrm{d}x^2}{a^2} + \frac{2xy}{a^2b^2}\,\mathrm{d}x\,\mathrm{d}y + \left(\frac{y^2}{b^2} + \frac{z^2}{c^2}\right)\frac{\mathrm{d}y^2}{b^2}\right],$$

woraus sich

$$\frac{\partial^2 z}{\partial x^2} = -\frac{c^4}{a^2z^3}\left(\frac{x^2}{a^2} + \frac{z^2}{c^2}\right), \quad \frac{\partial^2 z}{\partial x\,\partial y} = -\frac{c^4xy}{a^2b^2z^3}$$

$$\frac{\partial^2 z}{\partial y^2} = -\frac{c^4}{b^2z^3}\left(\frac{y^2}{b^2} + \frac{z^2}{c^2}\right)$$

ergibt, usw.

4. Es sei $z(x, y)$ durch die Gleichung $z = x + y\varphi(z)$ definiert. Man zeige, daß unter der Voraussetzung $1 - y\varphi'(z) \neq 0$

$$\frac{\partial z}{\partial y} = \varphi(z)\,\frac{\partial z}{\partial x}$$

gilt.

Offenbar ist

$$\frac{\partial z}{\partial x} = \frac{1}{1 - y\varphi'(z)}, \quad \frac{\partial z}{\partial y} = \frac{\varphi(z)}{1 - y\varphi'(z)},$$

und daraus ergibt sich das Gewünschte.

5. Durch $y = x\varphi(z) + \psi(z)$ sei z als implizite Funktion von x und y definiert. Unter der Annahme $x\varphi'(z) + \psi'(z) \neq 0$ beweise man, daß z der Differentialgleichung

$$\frac{\partial^2 z}{\partial x^2}\left(\frac{\partial z}{\partial y}\right)^2 - 2\,\frac{\partial z}{\partial x}\,\frac{\partial z}{\partial y}\,\frac{\partial^2 z}{\partial x\,\partial y} + \frac{\partial^2 z}{\partial y^2}\left(\frac{\partial z}{\partial x}\right)^2 = 0$$

oder

$$rq^2 - 2pqs + tp^2 = 0$$

genügt, wobei zur Abkürzung

$$\frac{\partial z}{\partial x} = p, \quad \frac{\partial z}{\partial y} = q, \quad \frac{\partial^2 z}{\partial x^2} = r, \quad \frac{\partial^2 z}{\partial x\,\partial y} = s, \quad \frac{\partial^2 z}{\partial y^2} = t$$

gesetzt ist.

Durch sukzessives Differenzieren nach x und y ergibt sich

$$\varphi(z) + [x\varphi'(z) + \psi'(z)]\,p = 0, \qquad [x\varphi'(z) + \psi'(z)]\,q = 1$$

und

$$2\varphi'(z)\,p + [x\varphi''(z) + \psi''(z)]\,p^2 + [x\varphi'(z) + \psi'(z)]\,r = 0,$$
$$\varphi'(z)\,q + [x\varphi''(z) + \psi''(z)]\,pq + [x\varphi'(z) + \psi'(z)]\,s = 0,$$
$$[x\varphi''(z) + \psi''(z)]\,q^2 + [x\varphi'(z) + \psi'(z)]\,t = 0.$$

Multiplikation dieser Gleichungen mit q^2, $-2pq$ bzw. p^2 und Addition liefert die zu beweisende Beziehung.

6. Es sei das System

$$x + y + z + u = a, \quad x^2 + y^2 + z^2 + u^2 = b^2, \quad x^3 + y^3 + z^3 + u^3 = c^3$$

gegeben, das y, z, u als Funktionen von x definiert. Hier gilt

$$1 + y' + z' + u' = 0, \quad x + yy' = zz' + uu' = 0,$$
$$x^2 + y^2y' + z^2z' + u^2u' = 0.$$

Setzt man die Determinante

$$\begin{vmatrix} 1 & 1 & 1 \\ y & z & u \\ y^2 & z^2 & u^2 \end{vmatrix} = (z - y)\,(u - y)\,(u - z)$$

als von 0 verschieden voraus, so folgt

$$y' = -\frac{(z - x)\,(u - x)}{(z - y)\,(u - y)} \quad \text{usw.}$$

7. Die Veränderlichen x, y, z seien mit r, θ, φ durch die Beziehungen

$$x = r\cos\theta\cos\varphi, \quad y = r\sin\theta\cos\varphi, \quad z = r\sin\varphi$$

verknüpft, wobei

$$0 < r < \infty, \qquad -\frac{\pi}{2} < \theta < \frac{\pi}{2}, \qquad -\frac{\pi}{2} < \varphi < \frac{\pi}{2}$$

sei. Für die Funktionaldeterminante gilt

$$J = \frac{D(x, y, z)}{D(r, \theta, \varphi)} = \begin{vmatrix} \cos\theta\cos\varphi & -r\sin\theta\cos\varphi & -r\cos\theta\sin\varphi \\ \sin\theta\cos\varphi & r\cos\theta\cos\varphi & -r\sin\theta\sin\varphi \\ \sin\varphi & 0 & r\cos\varphi \end{vmatrix} = r^2\cos\varphi > 0.$$

Die obigen Beziehungen definieren r, θ, φ als Funktionen von x, y, z. Zur Berechnung der Ableitungen dieser Funktionen bilden wir die totalen Differentiale

$$\begin{aligned} \cos\theta\cos\varphi\,\mathrm{d}r - r\sin\theta\cos\varphi\,\mathrm{d}\theta - r\cos\theta\sin\varphi\,\mathrm{d}\varphi &= \mathrm{d}x, \\ \sin\theta\cos\varphi\,\mathrm{d}r + r\cos\theta\cos\varphi\,\mathrm{d}\theta - r\sin\theta\sin\varphi\,\mathrm{d}\varphi &= \mathrm{d}y, \\ \sin\varphi\,\mathrm{d}r \qquad\qquad\qquad + r\cos\varphi\,\mathrm{d}\varphi &= \mathrm{d}z. \end{aligned}$$

Hieraus ergibt sich

$$\mathrm{d}r = \frac{r^2\cos\theta\cos^2\varphi}{J}\,\mathrm{d}x + \frac{r^2\sin\theta\cos^2\varphi}{J}\,\mathrm{d}y + \frac{r^2\sin\varphi\cos\varphi}{J}\,\mathrm{d}z,$$

$$\mathrm{d}\theta = -\frac{r\sin\theta}{J}\,\mathrm{d}x + \frac{r\cos\theta}{J}\,\mathrm{d}y,$$

$$\mathrm{d}\varphi = -\frac{r\cos\theta\sin\varphi\cos\varphi}{J}\,\mathrm{d}x - \frac{r\sin\theta\sin\varphi\cos\varphi}{J}\,\mathrm{d}y + \frac{r\cos^2\varphi}{J}\,\mathrm{d}z.$$

Damit haben wir im Grunde die uns interessierenden Ableitungen schon gefunden, da wir ja J kennen:

$$\frac{\partial r}{\partial x} = \cos\theta\cos\varphi, \quad \frac{\partial r}{\partial y} = \sin\theta\cos\varphi, \quad \frac{\partial r}{\partial z} = \sin\varphi,$$

$$\frac{\partial \theta}{\partial x} = -\frac{\sin\theta}{r\cos\varphi}, \quad \frac{\partial \theta}{\partial y} = \frac{\cos\theta}{r\cos\varphi}, \quad \frac{\partial \theta}{\partial z} = 0,$$

$$\frac{\partial \varphi}{\partial x} = -\frac{\cos\theta\sin\varphi}{r}, \quad \frac{\partial \varphi}{\partial y} = -\frac{\sin\theta\sin\varphi}{r}, \quad \frac{\partial \varphi}{\partial z} = \frac{\cos\varphi}{r}.$$

Übrigens lassen sich die Gleichungen leicht nach r, θ, φ auflösen (θ und φ nur zwischen 0 und π eindeutig)

$$r = \sqrt{x^2 + y^2 + z^2}, \quad \theta = \arctan\frac{y}{x}, \quad \varphi = \arctan\frac{z}{\sqrt{x^2 + y^2}}.$$

Hieraus kann man alle diese Ableitungen leicht berechnen und das Ergebnis verifizieren.

8. Als letztes Beispiel der Differentiation impliziter Funktionen führen wir eine Formel an, die noch einmal die Analogie zwischen der Funktionaldeterminante eines Systems von Funktionen und der Ableitung einer einzigen Funktion unterstreicht.

Es sei ein System von n Gleichungen mit $2n$ Veränderlichen gegeben:

$$F_i(x_1, \ldots, x_n; y_1, \ldots, y_n) = 0 \quad (i = 1, \ldots, n).$$

Die Funktionaldeterminante

$$\frac{D(F_1, \ldots, F_n)}{D(y_1, \ldots, y_n)}$$

sei von 0 verschieden. Dann können wir $y_1, \ldots, y_n$ als Funktionen von $x_1, \ldots, x_n$ ansehen, die durch dieses Gleichungssystem definiert sind und die Gleichungen zu Identitäten machen. Wir differenzieren diese Identitäten nach jedem x_j und schreiben die Ergebnisse in der Gestalt

$$-\frac{\partial F_i}{\partial x_j} = \frac{\partial F_i}{\partial y_1}\frac{\partial y_1}{\partial x_j} + \frac{\partial F_i}{\partial y_2}\frac{\partial y_2}{\partial x_j} + \cdots + \frac{\partial F_i}{\partial y_n}\frac{\partial y_n}{\partial x_j} \qquad (i, j = 1, \ldots, n).$$

Die aus den linken Seiten dieser Gleichungen gebildete Determinante ist gleich

$$(-1)^n \frac{D(F_1, \ldots, F_n)}{D(x_1, \ldots, x_n)}.$$

Die aus den rechten Seiten gebildete Determinante dagegen ist offenbar das Produkt der Determinanten

$$\frac{D(F_1, \ldots, F_n)}{D(y_1, \ldots, y_n)} \quad \text{und} \quad \frac{D(y_1, \ldots, y_n)}{D(x_1, \ldots, x_n)}$$

[vgl. Nr. 203, Formel (3)]. Hieraus ergibt sich

$$\frac{D(y_1, \ldots, y_n)}{D(x_1, \ldots, x_n)} = (-1)^n \frac{\dfrac{D(F_1, \ldots, F_n)}{D(x_1, \ldots, x_n)}}{\dfrac{D(F_1, \ldots, F_n)}{D(y_1, \ldots, y_n)}},$$

ein Analogon der Formel (15).

Sind die Gleichungen nach $x_1, \ldots, x_n$ aufgelöst, $x_i = \varphi_i(y_1, \ldots, y_n)$ $(i = 1, \ldots, n)$, so läßt sich dieser Fall dem betrachteten unterordnen, wenn man $F_i = \varphi_i - x_i$ setzt. Da hier $\frac{\partial F_i}{\partial x_j} = -1$ oder 0 ist, je nachdem $i = j$ oder $i \neq j$ ist, reduziert sich der Zähler auf

$$\begin{vmatrix} -1 & 0 \ldots & 0 \\ 0 & -1 \ldots & 0 \\ \cdot & & \cdot \\ 0 & 0 \ldots & -1 \end{vmatrix} = (-1)^n,$$

und die Formel geht über in

$$\frac{D(y_1, \ldots, y_n)}{D(x_1, \ldots, x_n)} = \frac{1}{\dfrac{D(x_1, \ldots, x_n)}{D(y_1, \ldots, y_n)}}.$$

Dieses Resultat ist uns schon bekannt [Nr. 203, Formel (4)].

§ 3. Einige Anwendungen der Theorie der impliziten Funktionen

211. Extrema mit Nebenbedingungen. Wir betrachten das Problem eines Extremums einer Funktion $f(x_1, \ldots, x_{n+m})$ von $n + m$ Veränderlichen unter der Annahme, daß diese Veränderlichen noch m *Nebenbedingungen*

$$\Phi_i(x_1, \ldots, x_n, x_{n+1}, \ldots, x_{n+m}) = 0 \quad (i = 1, \ldots, m) \tag{1}$$

unterworfen sind.

Wir erläutern diesen Begriff des Extremums mit Nebenbedingungen und geben ein Verfahren zu seiner Berechnung an.

Man sagt, in einem Punkt $M_0(x_1^0, \ldots, x_{n+m}^0)$, der einer Nebenbedingung genügt, besitsitze eine Funktion $f(x_1, \ldots, x_{n+m})$ ein *Maximum* (*Minimum*) *mit Nebenbedingungen* (oder ein *bedingtes Maximum* usw.), wenn die Ungleichung

$$f(x_1, \ldots, x_{n+m}) \underset{(\geqq)}{\leqq} f(x_1^0, \ldots, x_{n+m}^0)$$

in einer Umgebung des Punktes M_0 für alle Punkte $(x_1, \ldots, x_{n+m})$ erfüllt ist, welche den Nebenbedingungen genügen.

Wir nehmen an, daß sowohl die Funktion f als auch die Funktion Φ_i in der Umgebung des betrachteten Punktes stetige partielle Ableitungen nach allen Argumenten besitzen. Ferner sei in M_0 mindestens eine der Determinanten m-ter Ordnung, die aus der Matrix

$$\begin{pmatrix} \frac{\partial \Phi_1}{\partial x_1} \ldots \frac{\partial \Phi_1}{\partial x_n} & \frac{\partial \Phi_1}{\partial x_{n+1}} \ldots \frac{\partial \Phi_1}{\partial x_{n+m}} \\ \frac{\partial \Phi_2}{\partial x_1} \ldots \frac{\partial \Phi_2}{\partial x_n} & \frac{\partial \Phi_2}{\partial x_{n+1}} \ldots \frac{\partial \Phi_2}{\partial x_{n+m}} \\ \cdot & \\ \frac{\partial \Phi_m}{\partial x_1} \ldots \frac{\partial \Phi_m}{\partial x_n} & \frac{\partial \Phi_m}{\partial x_{n+1}} \ldots \frac{\partial \Phi_m}{\partial x_{n+m}} \end{pmatrix} \tag{2}$$

der partiellen Ableitungen gebildet werden können, beispielsweise

$$\frac{D(\Phi_1, \ldots, \Phi_m)}{D(x_{n+1}, \ldots, x_{n+m})} = \begin{vmatrix} \frac{\partial \Phi_1}{\partial x_{n+1}} & \cdots & \frac{\partial \Phi_1}{\partial x_{n+m}} \\ \frac{\partial \Phi_2}{\partial x_{n+1}} & \cdots & \frac{\partial \Phi_2}{\partial x_{n+m}} \\ \cdot & & \\ \frac{\partial \Phi_m}{\partial x_{n+1}} & \cdots & \frac{\partial \Phi_m}{\partial x_{n+m}} \end{vmatrix}, \tag{3}$$

von 0 verschieden.[1]) Beschränkt man sich dann auf eine hinreichend kleine Umgebung von M_0, so ist nach Satz IV das System (1) dem System

$$x_{n+1} = \varphi_1(x_1, \ldots, x_n), \ldots, x_{n+m} = \varphi_m(x_1, \ldots, x_n) \tag{4}$$

äquivalent, wobei $\varphi_1, \ldots, \varphi_m$ implizite, durch das System (1) definierte Funktionen sind.

Mit anderen Worten, die Forderung, daß die Werte der Veränderlichen $x_1, \ldots, x_n$, $x_{n+1}, \ldots, x_{n+m}$ den Nebenbedingungen (1) genügen, kann durch die Annahme ersetzt werden, daß die Veränderlichen $x_{n+1}, \ldots, x_{n+m}$ Funktionen (4) von $x_1, \ldots, x_n$ sind. Somit reduziert sich das Problem eines Extremums mit Nebenbedingungen für eine Funktion $f(x_1, \ldots, x_{n+m})$ von $n + m$ Veränderlichen im Punkt $M_0(x_1^0, \ldots, x_n^0, x_{n+1}^0, \ldots, x_{n+m}^0)$ auf das Problem eines gewöhnlichen Extremums für die zusammengesetzte Funktion von n Veränderlichen

$$f\big(x_1, \ldots, x_n; \varphi_1(x_1, \ldots, x_n), \ldots, \varphi_m(x_1, \ldots, x_n)\big) \tag{5}$$

im Punkt $P_0(x_0^1, \ldots, x_n^0)$.

Diese Überlegungen deuten auch den Weg an, auf dem man die Punkte finden kann, in denen $f(x_1, \ldots, x_{n+m})$ ein Extremum mit Nebenbedingungen annimmt: Wenn man die Nebenbedingungen beispielsweise nach den Veränderlichen $x_{n+1}, \ldots, x_{n+m}$ auflösen und explizite Ausdrücke für die Funktionen (4) finden kann, so hat man das Problem auf die Bestimmung eines Extremums der mittelbaren Funktion (5) zurückgeführt. Genaugenommen sind wir in einer Reihe früher behandelter Beispiele (Nr. 200 und 201) so verfahren, etwa als wir den kleinsten Wert der Summe $x + y + z + t$ unter der Bedingung $xyzt = c^4$ bestimmten, und in ähnlichen Fällen.

Wir geben nun einen anderen Weg zur Bestimmung eines solchen Punktes $M_0(x_1^0, \ldots x_{n+m}^0)$ an, und zwar ohne die Annahme, wir hätten explizite Ausdrücke für die (impliziten) Funktionen (4) zur Verfügung. Allerdings benutzen wir auch hierbei die Existenz dieser Funktionen.

In M_0 habe also $f(x_1, \ldots, x_{n+m})$ ein Extremum mit Nebenbedingungen oder, was dasselbe ist, die zusammengesetzte Funktion (5) habe in P_0 ein Extremum. Dann muß nach Bemerkung I aus Nr. 196 ihr Differential in diesem Punkt verschwinden, und zwar identisch in bezug auf die Differentiale der unabhängigen Veränderlichen $\mathrm{d}x_1, \ldots, \mathrm{d}x_n$. Auf Grund der Invarianz des ersten Differentials (Nr. 185) kann man diese Bedingung in der Gestalt

$$\sum_{j=1}^{n+m} \frac{\partial f}{\partial x_j} \, \mathrm{d}x_j = 0 \tag{6}$$

[1]) In diesem Fall sagt man, die Matrix (2) habe im Punkt M_0 den *Rang* m.

schreiben, wobei unter $dx_{n+1}, \ldots, dx_{n+m}$ die Differentiale der Funktionen (4) in P_0 zu verstehen sind, während die partiellen Ableitungen in M_0 zu berechnen sind. Aus Satz IV folgt nämlich

$$\varphi_1(x_1^0, \ldots, x_n^0) = x_{n+1}^0, \ldots, \varphi_m(x_1^0, \ldots, x_n^0) = x_{n+m}^0. \tag{7}$$

Aus (6) darf man natürlich nicht schließen, daß die Koeffizienten der Differentiale verschwinden, da nicht alle diese Differentiale willkürlich sind. Um auf beliebig wählbare Differentiale zu kommen, d. h. auf die Differentiale $dx_1, \ldots, dx_n$ der unabhängigen Veränderlichen, müssen wir hieraus die Differentiale $dx_{n+1}, \ldots dx_{n+m}$ der abhängigen Veränderlichen eliminieren. Das läßt sich leicht machen, wenn man die totalen Differentiale der Nebenbedingungen (1) bildet, wobei man unter $x_{n+1}, \ldots, x_{n+m}$ die Funktionen (4) zu verstehen hat[1]):

$$\sum_{j=1}^{n+m} \frac{\partial \Phi_i}{\partial x_j}\, dx_j = 0 \quad (i = 1, \ldots, m). \tag{8}$$

Hier sind wie oben, nach (7), die partiellen Ableitungen in M_0 zu berechnen. Da nach Voraussetzung die Determinante (3) in diesem Punkt von 0 verschieden ist, können $dx_{n+1}, \ldots, dx_{n+m}$ hieraus linear durch $dx_1, \ldots, dx_n$ ausgedrückt werden. Setzt man diese Ausdrücke in (6) ein, so ergibt sich eine Gleichung der Gestalt

$$A_1\, dx_1 + \cdots + A_n\, dx_n = 0;$$

$A_1, \ldots, A_n$ bedeuten hier n Ausdrücke, die rational bezüglich der partiellen Ableitungen der Φ_j (an der Stelle M_0) sind. Da hier nur die Differentiale $dx_1, \ldots, dx_n$ der unabhängigen Veränderlichen vorkommen, gelten in M_0 die Beziehungen

$$A_1 = 0, \ldots, A_n = 0.$$

Zusammen mit den Nebenbedingungen sind das $n + m$ Gleichungen zur Bestimmung der Unbekannten $x_1, \ldots, x_{n+m}$.

Natürlich haben wir hier nur ein *notwendiges* Kriterium für ein Extremum in $M_0(x_1^0, \ldots, x_{n+m}^0)$ angegeben. Aber auch in dieser Form können die Bedingungen zur Bestimmung eines größten (bzw. kleinsten) Wertes der Funktion f unter den Nebenbedingungen (1) nützlich sein, wenn aus der Art des Problems hervorgeht, daß im Innern des betrachteten Bereichs ein Punkt existieren muß, in dem dieser größte (bzw. kleinste) Wert angenommen wird, oder wenn diese Annahme im Zuge von Gedankengängen gemacht wird, um einen durch andere Überlegungen gefundenen Punkt daraufhin zu untersuchen.

Beispiele werden in Nr. 214 angegeben.

212. Die Lagrangeschen Multiplikatoren (unbestimmte Faktoren). Die oben dargelegte Methode verletzt die Symmetrie in bezug auf die Veränderlichen: ein Teil wird als unabhängig, ein Teil als abhängig angesehen; einige Differentiale werden eliminiert, andere nicht. Das führt manchmal zu Komplikationen bei den Berechnungen. Von Lagrange stammt eine Methode, bei der alle Veränderlichen in gleicher Weise behandelt werden.

Wir multiplizieren die Gleichungen (8) mit vorläufig willkürlichen (unbestimmten) Faktoren λ_i $(i = 1, \ldots, m)$ und addieren die entstehenden Gleichungen zu (6). Dann

[1]) Genauer gesagt, man differenziert die Identität, die man aus (1) erhält, wenn man statt $x_{n+1}, \ldots, x_{n+m}$ die impliziten Funktionen (4) dort einsetzt. Diese Redeweise werden wir auch im folgenden benutzen.

erhalten wir

$$\sum_{j=1}^{n+m} \left(\frac{\partial f}{\partial x_j} + \lambda_1 \frac{\partial \Phi_1}{\partial x_j} + \cdots + \lambda_m \frac{\partial \Phi_m}{\partial x_j} \right) dx_j = 0, \tag{9}$$

wobei wie früher $dx_{n+1}, \ldots, dx_{n+m}$ die Differentiale der impliziten Funktionen (4) bedeuten. In den Überlegungen behalten wir vorläufig die Unterscheidung in den Veränderlichen bei. Die Ableitungen sind in M_0 zu berechnen.

Wir wählen jetzt die Werte der Faktoren $\lambda_i = \lambda_i^0$ $(i = 1, \ldots, m)$ so, daß die Koeffizienten der abhängigen Differentiale $dx_{n+1}, \ldots, dx_{n+m}$ verschwinden:

$$\frac{\partial f}{\partial x_j} + \lambda_1^0 \frac{\partial \Phi_1}{\partial x_j} + \cdots + \lambda_m^0 \frac{\partial \Phi_m}{\partial x_j} = 0 \quad (j = n+1, \ldots, n+m). \tag{10}$$

Das läßt sich machen, da die Determinante (3) des linearen Gleichungssystems zur Bestimmung der $\lambda_1, \lambda_2, \ldots, \lambda_m$ von 0 verschieden ist. Bei diesen Werten der Faktoren nimmt Gleichung (9) die Gestalt

$$\sum_{j=1}^{n} \left(\frac{\partial f}{\partial x_j} + \lambda_1^0 \frac{\partial \Phi_1}{\partial x_j} + \cdots + \lambda_m^0 \frac{\partial \Phi_m}{\partial x_j} \right) dx_j = 0 \tag{9*}$$

an. Hier haben wir es wieder nur mit Differentialen unabhängiger Veränderlicher zu tun; daher müssen ihre Koeffizienten verschwinden. Neben (10) gilt also auch

$$\frac{\partial f}{\partial x_j} + \lambda_1^0 \frac{\partial \Phi_1}{\partial x_j} + \cdots + \lambda_m^0 \frac{\partial \Phi_m}{\partial x_j} = 0 \quad (j = 1, \ldots, n). \tag{10*}$$

Somit haben wir zur Bestimmung der $n + m$ Unbekannten $x_1, \ldots, x_{n+m}$ und der m Faktoren $\lambda_1, \ldots, \lambda_m$ ebenso viele Gleichungen, nämlich die m Nebenbedingungen und die $n + m$ Gleichungen

$$\frac{\partial f}{\partial x_j} + \lambda_1 \frac{\partial \Phi_1}{\partial x_j} + \cdots + \lambda_m \frac{\partial \Phi_m}{\partial x_j} = 0 \quad (j = 1, \ldots, n+m)$$

[vgl. (10) und (10*)].

Um das Hinschreiben dieser Gleichungen zu erleichtern, führt man die Hilfsfunktion

$$F = f + \lambda_1 \Phi_1 + \cdots + \lambda_m \Phi_m$$

ein; dann kann man diese Gleichungen in der Gestalt

$$\frac{\partial F}{\partial x_j} = 0 \quad (j = 1, \ldots, n+m) \tag{11}$$

schreiben. Sie sehen genauso aus wie die Bedingungen für ein gewöhnliches Extremum der Funktion F. Das ist allerdings nur als Hinweis zu betrachten, der die Erinnerung unterstützt.

Auch die Lagrangesche Methode führt nur zu notwendigen Bedingungen. Im übrigen läßt sich das am Schluß von Nr. 211 Gesagte wiederholen.

Bemerkung. In diesen Ausführungen spielte die Voraussetzung über den Rang der Matrix (2), die wir dreimal benutzt haben, eine Rolle. Bei der Lösung von Aufgaben mit Hilfe einer der angegebenen Methoden muß man, um sich davon zu überzeugen, daß man keinen Punkt vergessen hat, in dem die Funktion ein Extremum mit Nebenbedingungen annimmt, vorher nachweisen, daß diese Voraussetzung tatsächlich in allen Punkten des betrachteten Bereichs erfüllt ist, für die die Nebenbedingungen gelten. In einfachen Fällen werden wir das dem Leser überlassen.

213. Hinreichende Bedingungen für ein Extremum mit Nebenbedingungen. Bei diesem Problem beschränken wir uns auf einige Bemerkungen. Wir setzen Existenz und Stetigkeit der zweiten Ableitungen der Funktionen f und Φ_j $(j = 1, \ldots, m)$ voraus. Der Punkt $M_0(x_1^0, \ldots, x_{n+m}^0)$ genüge zusammen mit den Faktoren $\lambda_1^0, \ldots, \lambda_m^0$ den oben angegebenen notwendigen Bedingungen.

Ob in diesem Punkt ein Extremum (mit Nebenbedingungen) vorliegt, hängt, wie in Nr. 198, vom Vorzeichen der Differenz

$$\Delta = f(x_1, \ldots, x_{n+m}) - f(x_1^0, \ldots, x_{n+m}^0)$$

ab, mit dem wesentlichen Vorbehalt, daß auch der Punkt $(x_1, \ldots, x_{n+m})$ der Nebenbedingung (1) oder, was dasselbe ist, der Bedingung (4) genügt. Es ist leicht einzusehen, daß für solche Punkte der Zuwachs der Funktion f durch den Zuwachs von F (wobei alle λ_i gleich λ_i^0 angenommen werden) ersetzt werden kann:

$$\Delta = F(x_1, \ldots, x_{n+m}) - F(x_1^0, \ldots, x_{n+m}^0).$$

Da in M_0 die Bedingungen (11) erfüllt sind (darin besteht gerade der Vorteil des Übergangs zu der Funktion F), kann dieser Zuwachs nach der Taylorschen Formel [vgl. Nr. 198, Formel (8)] folgendermaßen geschrieben werden:

$$\Delta = \frac{1}{2}\left\{\sum_{j,k=1}^{n+m} A_{jk}\,\Delta x_j\,\Delta x_k + \sum_{j,k=1}^{n+m} \alpha_{jk}\,\Delta x_j\,\Delta x_k\right\},$$

wobei

$$\Delta x_j = x_j - x_j^0, \quad A_{jk} = F_{x_j x_k}(x_1^0, \ldots, x_{n+m}^0) \quad (j, k = 1, \ldots, n+m)$$

ist und $\alpha_{jk} \to 0$ gilt, sobald die $\Delta x_1, \ldots, \Delta x_n$ gegen 0 streben [die übrigen Zuwächse $\Delta x_{n+1}, \ldots, \Delta x_{n+m}$ sind auf Grund der Stetigkeit der Funktionen (4) von selbst unendlich klein (im Sinne von Nr. 24)].

Ersetzt man hier alle Zuwächse Δx_j durch die entsprechenden Differentiale dx_j, so ändert sich in bezug auf die unabhängigen Veränderlichen nichts; bei den abhängigen Veränderlichen hat man bei dieser Ersetzung nur an Stelle der Koeffizienten α_{jk} andere unendlich kleine Größen β_{ik} zu setzen:

$$\Delta = \frac{1}{2}\left\{\sum_{j,k=1}^{n+m} A_{jk}\,dx_j\,dx_k + \sum_{j,k=1}^{n+m} \beta_{jk}\,dx_j\,dx_k\right\}.$$

Der Übergang zu den Differentialen ist vorteilhaft, weil die Differentiale der abhängigen und der unabhängigen Veränderlichen durch das lineare System (8) miteinander verknüpft sind. Da die Determinante (3) in M_0 nach Voraussetzung nicht verschwindet, lassen sich hieraus die abhängigen Differentiale linear durch die unabhängigen ausdrücken. Setzt man die Ausdrücke in Δ ein, so erhält man an Stelle der ersten Summe eine quadratische Form in den Differentialen $dx_1, \ldots, dx_n$.

Nun kann man wie in Nr. 198 und 199 folgendes beweisen: *Ist diese Form definit, und zwar positiv (bzw. negativ), so existiert im betreffenden Punkt ein Minimum (bzw. Maximum) mit Nebenbedingungen; ist die Form infinit, so liegt dort kein Extremum mit Nebenbedingungen vor.*

Allerdings ist der praktische Wert dieses Kriteriums nicht groß (vgl. die Bemerkung in Nr. 200).

Wir bringen nun Beispiele und Aufgaben.

214. Beispiele und Aufgaben.

1. Man bestimme das Extremum der Funktion $f = x + y + z + t$ unter der Nebenbedingung $\Phi = xyzt - c^4 = 0$ $(c > 0)$. Der Variationsbereich der Veränderlichen ist durch $x > 0$, $y > 0$, $z > 0$, $t > 0$ gegeben. Dieses Problem haben wir schon in Nr. 200, Beispiel 4, gelöst, indem wir t aus der Nebenbedingung berechneten.

Bilden wir das vollständige Differential, so ergibt sich

$$\frac{dx}{x} + \frac{dy}{y} + \frac{dz}{z} + \frac{dt}{t} = 0, \quad \text{also} \quad dt = -t\left(\frac{dx}{x} + \frac{dy}{y} + \frac{dz}{z}\right).$$

Eliminieren wir dt aus $df = dx + dy + dz + dt = 0$, so erhalten wir die Gleichung

$$\left(1 - \frac{t}{x}\right) dx + \left(1 - \frac{t}{y}\right) dy + \left(1 - \frac{t}{z}\right) dz = 0.$$

Da dx, dy, dz willkürlich sind, ist sie gleichbedeutend mit

$$1 - \frac{t}{x} = 0, \quad 1 - \frac{t}{y} = 0, \quad 1 - \frac{t}{z} = 0;$$

also ist $x = y = z = t = c$.

Wenden wir auf dieselbe Aufgabe die Lagrangesche Methode an, so haben wir die Hilfsfunktion

$$F = x + y + z + t + \lambda xyzt$$

einzuführen[1]) und kommen zu den Bedingungen

$$F_x = 1 + \lambda yzt = 0, \ldots, F_t = 1 + \lambda xyz = 0,$$

woraus sich $yzt = xzt = xyt = xyz$, also $x = y = z = t = c$ ergibt. Um das Resultat aus Nr. 213 zu benutzen, berechnen wir $\lambda = -\frac{1}{c^3}$ und betrachten die Funktion

$$F = x + y + z + t - \frac{xyzt}{c^3}.$$

Für ihr zweites Differential (im Punkt $x = y = z = t = c$) ergibt sich

$$d^2F = -\frac{2}{c}(dx\,dy + dx\,dz + dx\,dt + dy\,dz + dy\,dt + dz\,dt).$$

Durch Bildung des totalen Differentials der Nebenbedingung (in eben diesem Punkt) erhält man

$$dx + dy + dz + dt = 0.$$

Bestimmt man hieraus dt und setzt den Ausdruck in den vorigen ein, so folgt

$$-\frac{2}{c}[dx\,dy + dx\,dz + dy\,dz - (dx + dy + dz)^2]$$
$$= \frac{1}{c}[(dx + dy + dz)^2 + dx^2 + dy^2 + dz^2].$$

Da diese Form offenbar positiv definit ist, liegt in dem Punkt ein Minimum mit Nebenbedingungen vor.

Hieraus darf man jedoch nicht schließen, dieses Minimum sei der kleinste Wert der Funktion $f = x + y + z + t$ unter der angegebenen Nebenbedingung für ihre Argumente; vgl. dazu Nr. 200, Beispiel 4.

[1]) Erinnert man sich der Rolle dieser Funktion, so wird klar, daß der konstante Summand in Φ hier ohne weiteres weggelassen werden kann.

2. Wir wollen noch einmal (vgl. Nr. 200, Beispiel 2) den größten und den kleinsten Wert der Funktion

$$u = a^2x^2 + b^2y^2 + c^2z^2 - (ax^2 + by^2 + cz^2)^2 \qquad (a > b > c > 0)$$

unter der Nebenbedingung

$$x^2 + y^2 + z^2 = 1,$$

d. h. auf der durch diese Gleichung gegebenen Kugel, bestimmen.[1])

Zu diesem Zweck bestimmen wir zunächst nach der Lagrangeschen Methode alle Extrema (mit Nebenbedingungen) der Funktion. Die Hilfsfunktion

$$F = a^2x^2 + b^2y^2 + c^2z^2 - (ax^2 + by^2 + cz^2)^2 + \lambda(x^2 + y^2 + z^2)$$

liefert die Bedingungen

$$x[(a^2 + \lambda) - 2a(ax^2 + by^2 + cz^2)] = 0,$$
$$y[(b^2 + \lambda) - 2b(ax^2 + by^2 + cz^2)] = 0,$$
$$z[(c^2 + \lambda) - 2c(ax^2 + by^2 + cz^2)] = 0,$$

zu denen man noch die Nebenbedingung hinzunehmen muß. Hieraus folgt

(1) $\quad x = 0, \quad y = 0, \quad z = \pm 1 \quad (u = 0);$

(2) $\quad x = 0, \quad y = \pm 1, \quad z = 0 \quad (u = 0);$

(3) $\quad x = \pm 1, \quad y = 0, \quad z = 0 \quad (u = 0);$

(4) $\quad x = 0, \quad y = \pm\frac{1}{\sqrt{2}}, \quad z = \pm\frac{1}{\sqrt{2}} \quad \left(u = \frac{1}{4}(b - c)^2\right);$

(5) $\quad x = \pm\frac{1}{\sqrt{2}}, \quad y = 0, \quad z = \pm\frac{1}{\sqrt{2}} \quad \left(u = \frac{1}{4}(a - c)^2\right);$

(6) $\quad x = \pm\frac{1}{\sqrt{2}}, \quad y = \pm\frac{1}{\sqrt{2}}, \quad z = 0 \quad \left(u = \frac{1}{4}(a - b)^2\right).$

Wählen wir aus den in Klammern angegebenen Werten von u den größten und den kleinsten aus, so kommen wir wieder zu der aus Nr. 200, Beispiel 2, bekannten Lösung der Aufgabe.

3. Wir befassen uns nochmals mit der Aufgabe, den zweckmäßigsten Querschnitt des Leiters in einem elektrischen Stromkreis mit Parallelschaltung zu bestimmen (Nr. 201, Aufgabe 8). In den dortigen Bezeichnungen wollen wir das Extremum der Funktion

$$f(q_1, q_2, \ldots, q_n) = l_1q_1 + l_2q_2 + \cdots + l_nq_n$$

unter der Nebenbedingung

$$\Phi(q_1, q_2, \ldots, q_n) = \frac{\varrho l_1 J_1}{q_1} + \frac{\varrho l_2 J_2}{q_2} + \cdots + \frac{\varrho l_n J_n}{q_n} = e$$

aufsuchen. Dabei brauchen wir übrigens an Stelle der $q_1, q_2, \ldots, q_n$ keine anderen Veränderlichen einzuführen, wie wir das früher getan haben, da sich mit unserer neuen Methode die Aufgabe auch so ganz einfach lösen läßt.

Wenn wir das vollständige Differential von $\Phi = 0$ bilden, erhalten wir für das Differential dq_n den Ausdruck

$$dq_n = -\frac{q_n^2}{l_nJ_n}\left\{\frac{l_1J_1}{q_1^2}\,dq_1 + \cdots + \frac{l_{n-1}J_{n-1}}{q_{n-1}^2}\,dq_{n-1}\right\}.$$

[1]) Da diese Fläche eine abgeschlossene beschränkte Menge ist, folgt die Existenz von Punkten, in denen die Funktion einen größten bzw. kleinsten Wert annimmt, aus dem Weierstraßschen Satz (vgl. die Bemerkung am Schluß von Nr. 173).

Setzen wir diesen in die Gleichung $df = l_1\,dq_1 + \cdots + l_n\,dq_n = 0$ ein, so ergibt sich

$$\left(l_1 + \frac{q_n^2}{J_n}\cdot\frac{l_1 J_1}{q_1^2}\right) dq_1 + \cdots + \left(l_{n-1} - \frac{q_n^2}{J_n}\cdot\frac{l_{n-1}J_{n-1}}{q_{n-1}^2}\right) dq_{n-1} = 0.$$

Da die $dq_1, \ldots, dq_{n-1}$ willkürlich sind, sind ihre Koeffizienten einzeln gleich 0; also ist

$$\frac{q_1^2}{J_1} = \frac{q_2^2}{J_2} = \cdots = \frac{q_{n-1}^2}{J_{n-1}} = \frac{q_n^2}{J_n} = \lambda^2$$

und

$$q_1 = \lambda\sqrt{J_1}, \quad q_2 = \lambda\sqrt{J_2}, \ldots, q_n = \lambda\sqrt{J_n}. \tag{12}$$

Der Proportionalitätsfaktor λ ergibt sich leicht aus der Nebenbedingung:

$$\lambda = \frac{\varrho}{e}\sum_{i=1}^{n} l_i\sqrt{J_i}.$$

Wendet man die Lagrangesche Methode an, so muß man die Hilfsfunktion[1])

$$F(q_1, q_2, \ldots, q_n) = l_1 q_1 + \cdots + l_n q_n + \lambda^2\left(\frac{l_1 J_1}{q_1} + \cdots + \frac{l_n J_n}{q_n}\right)$$

bilden und ihre Ableitungen gleich 0 setzen:

$$\frac{\partial F}{\partial q_1} = l_1 - \frac{\lambda^2 l_1 J_1}{q_1^2} = 0, \ \ldots, \ \frac{\partial F}{\partial q_n} = l_n - \frac{\lambda^2 l_n J_n}{q_n^2} = 0;$$

hieraus erhält man wieder (12) usw.

4. Als komplizierteres Beispiel betrachten wir folgende Aufgabe: Das Ellipsoid $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$ $(a > b > c)$ werde mit der durch seinen Mittelpunkt gehenden Ebene $lx + my + nz = 0$ zum Schnitt gebracht. Man bestimme die Halbachsen der Schnittellipse. Mit anderen Worten, es sind die Extremwerte der Funktion $r^2 = x^2 + y^2 + z^2$ unter den beiden obigen Nebenbedingungen für die Veränderlichen zu bestimmen.

Die Eliminierung der abhängigen Differentiale (Nr. 211) führt hier auf komplizierte Überlegungen. Daher verwenden wir gleich die Lagrangesche Methode.

Um uns davon zu überzeugen, daß der Rang der Matrix

$$\begin{pmatrix} \frac{x}{a^2} & \frac{y}{b^2} & \frac{z}{c^2} \\ l & m & n \end{pmatrix}$$

in allen Punkten des Schnittes von Ellipsoid und Ebene gleich 2 ist,[2]) nehmen wir das Gegenteil an. Aus dem Verschwinden aller Determinanten zweiter Ordnung würde folgen, daß die Elemente der oberen und der unteren Zeile einander proportional sind. Dann würde aber aus $lx + my + nz = 0$ folgen, daß $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 0$, ist, und das ist nicht der Fall.

Wir bilden die Hilfsfunktion

$$F(x, y, z) = x^2 + y^2 + z^2 + \lambda\left(\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2}\right) + 2\mu(lx + my + nz)$$

und setzen ihre Ableitungen gleich 0:

$$x + \lambda\frac{x}{a^2} + \mu l = 0, \quad y + \lambda\frac{y}{b^2} + \mu m = 0, \quad z + \lambda\frac{z}{c^2} + \mu n = 0. \tag{13}$$

---

[1]) Den „unbestimmten Faktor“ setzen wir zweckmäßigerweise in der Form λ^2 an und denken uns die Konstante ϱ darin enthalten.

[2]) Vgl. die Bemerkung in Nr. 212.

Multiplikation dieser Gleichungen mit x, y, z und Addition liefert unter Berücksichtigung der Nebenbedingungen die Beziehung $\lambda = -r^2$.

Nimmt man an, keine der Zahlen l, m, n sei 0, so erkennt man aus (13), daß r von a, b, c verschieden ist. Dann läßt sich (13) umformen:

$$x = -\mu \frac{la^2}{a^2 - r^2}, \quad y = -\mu \frac{mb^2}{b^2 - r^2}, \quad z = -\mu \frac{nc^2}{c^2 - r^2}.$$

Hieraus kann man leicht μ bestimmen, also auch x, y, z. Das läßt sich vermeiden, wenn man die mit l, m, n multiplizierten Gleichungen addiert; dann ergibt sich nämlich

$$\frac{l^2a^2}{a^2 - r^2} + \frac{m^2b^2}{b^2 - r^2} + \frac{n^2c^2}{c^2 - r^2} = 0;$$

hieraus kann man unmittelbar die interessierenden Extremalwerte von r^2 erhalten.

Da die Existenz dieser Extremalwerte von vornherein bekannt ist, hat man damit die Aufgabe vollständig gelöst.

5. Schließlich wollen wir den größten und den kleinsten Wert der quadratischen Form

$$f(x_1, x_2, \ldots, x_n) \equiv \sum_{i,k=1}^{n} a_{ik}x_ix_k \qquad (a_{ik} = a_{ki})$$

unter der Nebenbedingung

$$\varphi(x_1, x_2, \ldots, x_n) \equiv x_1^2 + x_2^2 + \cdots + x_n^2 = 1 \tag{14}$$

bestimmen.[1])

Wir bilden die Lagrangesche Funktion

$$F(x_1, x_2, \ldots, x_n) = f(x_1, x_2, \ldots, x_n) - \lambda(x_1^2 + x_2^2 + \cdots + x_n^2).$$

Eliminieren wir $x_1, x_2, \ldots, x_n$ aus

$$\left.\begin{aligned}
\frac{1}{2}\frac{\partial F}{\partial x_1} &\equiv (a_{11} - \lambda)\cdot x_1 + a_{12}\cdot x_2 + \cdots + a_{1n}\cdot x_n = 0,\\
\frac{1}{2}\frac{\partial F}{\partial x_2} &\equiv a_{21}\cdot x_1 + (a_{22} - \lambda)\cdot x_2 + \cdots + a_{2n}\cdot x_n = 0,\\
&\cdot\quad\cdot\\
\frac{1}{2}\frac{\partial F}{\partial x_n} &\equiv a_{n1}\cdot x_1 + a_{n2}\cdot x_2 + \cdots + (a_{nn} - \lambda)\cdot x_n = 0,
\end{aligned}\right\} \tag{15}$$

so kommen wir zu der Gleichung n-ten Grades

$$\begin{vmatrix} a_{11} - \lambda & a_{12} & a_{1n} \\ a_{21} & a_{22} - \lambda & a_{2n} \\ \cdot & & \\ a_{n1} & a_{n2} & a_{nn} - \lambda \end{vmatrix} = 0 \tag{16}$$

in λ. Ist λ eine ihrer Wurzeln, so läßt sich das lineare Gleichungssystem (15) durch Werte $x_1, x_2, \ldots, x_n$ befriedigen, die nicht sämtlich gleich 0 sind. Durch Multiplikation mit passenden Faktoren kann man dann auch die Bedingung (14) erfüllen. Die Bestimmung dieser Werte ist aber für uns ohne Interesse, da, wie wir sehen werden, sich der kleinste und der größte Wert von f auch ohne sie finden lassen.

Wenn wir nämlich die Gleichungen (15) mit $x_1, x_2, \ldots x_n$ multiplizieren und dann addieren, kommen wir zu

$$f(x_1, x_2, \ldots, x_n) - \lambda(x_1^2 + x_2^2 + \cdots + x_n^2) = 0$$

---

[1]) Hier kann man dieselbe Bemerkung wie in der Fußnote auf S. 436 machen.

oder, wegen (14), zu

$$f(x_1, x_2, \ldots, x_n) = \lambda.$$

Wenn also λ der Gleichung (16) genügt, ist der Wert von f in dem entsprechenden Punkt $(x_1, x_2, \ldots, x_n)$ selbst gleich λ.

Wir sind zu einem eleganten Resultat gelangt: Der gesuchte kleinste bzw. größte Wert der Funktion f unter der Nebenbedingung (14) stimmt mit der kleinsten bzw. größten (reellen) Wurzel der Gleichung (16) überein.[1])

215. Der Begriff der Unabhängigkeit von Funktionen. Wir betrachten das Funktionensystem

$$\left.\begin{aligned} y_1 &= f_1(x_1, x_2, \ldots, x_n), \\ y_2 &= f_2(x_1, x_2, \ldots, x_n), \\ &\ldots \\ y_m &= f_m(x_1, x_2, \ldots, x_n); \end{aligned}\right\} \tag{17}$$

die Funktionen seien nebst ihren partiellen Ableitungen in einem n-dimensionalen offenen Bereich $\mathcal{D}$ definiert und stetig.

Wir untersuchen den Fall, daß der Wert einer dieser Funktionen, etwa y_j, durch das System der Werte $(y_1, \ldots, y_{j-1}, y_{j+1}, \ldots, y_m)$, welche die übrigen Funktionen annehmen, eindeutig bestimmt ist. Genauer, ist $\mathcal{E}_0$ die Menge der $(m-1)$-dimensionalen Punkte, die allen möglichen Punkten $(x_1, \ldots, x_n)$ in $\mathcal{D}$ entsprechen, so gelte in $\mathcal{E}_0$ die funktionale Abhängigkeit

$$y_j = \varphi(y_1, \ldots, y_{j-1}, y_{j+1}, \ldots, y_m), \tag{18}$$

wobei diese Beziehung identisch in den x aus $\mathcal{D}$ gelten soll, wenn an Stelle aller y_i die Funktionen (17) eingesetzt werden.[2]) Dann sagt man, in $\mathcal{D}$ hänge die Funktion y_j von den übrigen ab. Um die Differentialrechnung anwenden zu können, fordern wir zusätzlich, daß die Funktion φ nebst ihren partiellen Ableitungen in einem offenen Bereich $\mathcal{E}$ des $(m-1)$-dimensionalen Raumes, der $\mathcal{E}_0$ enthält, definiert und stetig ist.

Ist insbesondere eine der Funktionen (17), etwa y_j, eine Konstante, dann hängt sie von den übrigen ab: Hier kann man einfach $\varphi = \text{const}$ setzen. Ein System von Funktionen $y_1, \ldots, y_m$ heißt schlechthin abhängig in $\mathcal{D}$, wenn irgendeine Funktion von den übrigen abhängt.

Beispiele.

1. Setzt man

$$y_1 = x_1 + x_2 + \cdots + x_n,$$

$$y_2 = x_1^2 + x_2^2 + \cdots + x_n^2,$$

$$y_3 = x_1x_2 + x_1x_3 + x_2x_3 + \cdots + x_{n-1}x_n,$$

so ergibt sich leicht, daß im ganzen n-dimensionalen Raum die Identität

$$y_2 = y_1^2 - 2y_3$$

erfüllt ist.

---

[1]) Man kann beweisen, daß *alle* Wurzeln dieser Gleichung wegen $a_{ik} = a_{ki}$ reell sind.

[2]) Es ist wesentlich, daß die Funktion φ unter ihren direkten Argumenten keine x enthält.

2. Analog gilt für die Funktionen

$$y_1 = x_1 x_2 - x_3,$$
$$y_2 = x_1 x_3 + x_2,$$
$$y_3 = (x_1^2 + 1)(x_2^2 + x_3^2) - (x_1^2 - 1) x_2 x_3 - x_1(x_2^2 - x_3^2)$$

im dreidimensionalen Raum die Identität

$$y_3 = y_1^2 - y_1 y_2 + y_2^2.$$

Das sind also *abhängige* Funktionen.

Gilt nun weder im Bereich $\mathfrak{D}$ noch in irgendeinem darin enthaltenen Teilbereich eine Identität der Gestalt (18), so heißen die Funktionen $y_1, y_2, \ldots, y_m$ in $\mathfrak{D}$ *unabhängig*.

Das Problem, ob Funktionen unabhängig sind, läßt sich durch Untersuchung der sogenannten *Funktionalmatrix* (*Jacobischen Matrix*)

$$\begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ & & & \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{pmatrix} \tag{19}$$

lösen, die aus den partiellen Ableitungen dieser Funktionen nach allen unabhängigen Veränderlichen besteht.

Für $n \geqq m$ gilt folgender Satz:

Satz I. *Ist wenigstens eine der Determinanten m-ter Ordnung, die aus Elementen der Matrix* (19) *bestehen, in* $\mathfrak{D}$ *von* 0 *verschieden, so sind dort die Funktionen* $y_1, y_2, \ldots, y_n$ *unabhängig.*

Beweis. Es sei

$$\begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_m} \\ & & & \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_m} \end{vmatrix} \neq 0. \tag{20}$$

Wäre nicht diese, sondern irgendeine andere Determinante von 0 verschieden, so könnte man das Problem durch Umnumerieren der Veränderlichen auf den Fall (20) zurückführen. Den Beweis führen wir indirekt. Es sei also etwa y_m durch die übrigen ausdrückbar; es gelte also

$$y_m = \varphi(y_1, y_2, \ldots, y_{m-1}) \tag{21}$$

wenigstens in einem Teilbereich $\mathfrak{D}_0$ von $\mathfrak{D}$. Differenzieren wir diese Identität nach jeder der Veränderlichen x_i $(i = 1, \ldots, m)$, so erhalten wir in $\mathfrak{D}_0$ eine Reihe von Identitäten der Gestalt

$$\frac{\partial y_m}{\partial x_i} = \frac{\partial y_m}{\partial y_1} \cdot \frac{\partial y_1}{\partial x_i} + \frac{\partial y_m}{\partial y_2} \cdot \frac{\partial y_2}{\partial x_i} + \cdots + \frac{\partial y_m}{\partial y_{m-1}} \cdot \frac{\partial y_{m-1}}{\partial x_i} \qquad (i = 1, 2, \ldots, m).$$

Wir sehen, daß die Elemente der letzten Zeile von (20) die Summen der entsprechenden, mit $\frac{\partial y_m}{\partial y_1}, \ldots, \frac{\partial y_m}{\partial y_{m-1}}$ mutliplizierten Elemente der ersten $m-1$ Zeilen sind. Eine solche Determinante ist bekanntlich gleich 0. Das widerspricht aber der Voraussetzung des Satzes. Mit diesem Widerspruch ist gezeigt, daß keine Beziehung der Gestalt (21) bestehen kann.

216. Der Rang einer Funktionalmatrix. Wir gehen nun zum allgemeinen Fall über. Unter dem *Rang* einer Funktionalmatrix (19) in einem Bereich $\mathfrak{D}$ verstehen wir die höchste Ordnung von Determinanten aus Elementen dieser Matrix, die in $\mathfrak{D}$ nicht identisch verschwinden. Es kann natürlich vorkommen, daß alle Elemente einer Matrix (19) identisch verschwinden; dann sagt man, die Matrix habe den Rang 0. Dieser Fall ist aber uninteressant, da hier einfach alle Funktionen $y_1, y_2, \ldots, y_m$ Konstanten sind (vgl. Nr. 183). Ist der Rang der Matrix (19) gleich $\mu \geqq 1$, so existiert mindestens eine Determinante μ-ter Ordnung aus Elementen der Matrix (natürlich $m \geqq \mu$, $n \geqq \mu$ vorausgesetzt), die in $\mathfrak{D}$ nicht identisch verschwindet, während alle Determinanten höherer Ordnung als μ (wenn es solche gibt) identisch 0 sind. Man sagt, der Rang μ werde in einem Punkt des Bereichs angenommen, wenn die betreffende Determinante μ-ter Ordnung gerade in diesem Punkt von 0 verschieden ist.

Satz II. *Der Rang einer Funktionalmatrix in einem Bereich $\mathfrak{D}$ sei $\mu \geqq 1$ und werde im Punkt $M_0(x_1^0, x_2^0, \ldots, x_n^0)$ dieses Bereichs angenommen. Dann sind in einer Umgebung $\mathfrak{D}_0$ dieses Punktes genau μ der m Funktionen (nämlich diejenigen, deren Ableitungen die Elemente der in M_0 nicht verschwindenden Determinante sind) unabhängig, und die übrigen Funktionen hängen von ihnen ab.*

Beweis. Ohne Beschränkung der Allgemeinheit kann man annehmen, in M_0 sei gerade die Determinante

$$\frac{D(y_1, \ldots, y_\mu)}{D(x_1, \ldots, x_\mu)} = \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_\mu} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_\mu} \\ \\ \frac{\partial y_\mu}{\partial x_1} & \frac{\partial y_\mu}{\partial x_2} & \cdots & \frac{\partial y_\mu}{\partial x_\mu} \end{vmatrix} = \begin{vmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_\mu} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_\mu} \\ \cdot\ \cdot \\ \frac{\partial f_\mu}{\partial x_1} & \frac{\partial f_\mu}{\partial x_2} & \cdots & \frac{\partial f_\mu}{\partial x_\mu} \end{vmatrix} \tag{22}$$

von 0 verschieden. Da die partiellen Ableitungen stetig sein sollten, gilt das auch in einer Umgebung von M_0. Daher sind nach Satz I die Funktionen $y_1, y_2, \ldots, y_\mu$ in dieser Umgebung unabhängig.

Wir bezeichnen die Werte dieser Funktionen in M_0 mit $y_1^0, y_2^0, \ldots, y_\mu^0$. Nach Satz IV aus Nr. 208 definieren in einem $(n+\mu)$-dimensionalen Parallelepiped

$$\begin{aligned} \mathcal{M}_0 = (&x_1^0 - \delta_1, x_1^0 + \delta_1; \ldots; x_n^0 - \delta_n, x_n^0 + \delta_n; \\ &y_1^0 - \Delta_1, y_1^0 + \Delta_1; \ldots; y_\mu^0 - \Delta_\mu, y_\mu^0 - \Delta_\mu) \end{aligned} \tag{23}$$

die ersten μ der Gleichungen (17)

$$\left.\begin{aligned} &f_1(x_1, \ldots, x_\mu; x_{\mu+1}, \ldots, x_n) - y_1 = 0, \\ &f_2(x_1, \ldots, x_\mu; x_{\mu+1}, \ldots, x_n) - y_2 = 0, \\ &\cdot\ \cdot \qquad\qquad\qquad\qquad\qquad\qquad\qquad\ \cdot \\ &f_\mu(x_1, \ldots, x_\mu; x_{\mu+1}, \ldots, x_n) - y_\mu = 0 \end{aligned}\right\} \tag{24}$$

die $x_1, x_2, \ldots, x_\mu$ als eindeutige Funktionen der übrigen Veränderlichen $y_1, \ldots, y_\mu$, $x_{\mu+1}, \ldots, x_n$, die in diesen Gleichungen vorkommen:

$$\left.\begin{aligned} x_1 &= \varphi_1(y_1, \ldots, y_\mu; x_{\mu+1}, \ldots, x_n), \\ x_2 &= \varphi_2(y_1, \ldots, y_\mu; x_{\mu+1}, \ldots, x_n), \\ &\ldots \\ x_\mu &= \varphi_\mu(y_1, \ldots, y_\mu; x_{\mu+1}, \ldots, x_n). \end{aligned}\right\} \tag{25}$$

In dem erwähnten Bereich sind die Gleichungssysteme (24) und (25) völlig äquivalent d. h., ihnen genügen ein und dieselben Werte der Veränderlichen $x_1, x_2, \ldots, x_n$ und $y_1, y_2, \ldots, y_n$. Aus demselben Satz, auf den wir uns hier stützen, folgt: Setzt man an Stelle der $x_1, x_2, \ldots, x_\mu$ in (24) die Funktionen (25) ein, so erhält man Identitäten in $y_1, \ldots, y_\mu, x_{\mu+1}, \ldots, x_n$. Für uns ist nun etwas anderes wichtig: Setzt man an Stelle der $y_1, \ldots, y_\mu$ in (25) die Funktionen $f_1, f_2, \ldots, f_\mu$ ein, so ergeben sich Identitäten in $x_1, x_2, \ldots, x_n$ wenigstens in einer Umgebung von $M_0(x_1^0, x_2^0, \ldots, x_n^0)$. Man braucht nämlich diese Umgebung

$$\mathcal{D}_0 = (x_1^0 - \delta_1', x_1^0 + \delta_1'; x_2^0 - \delta_2', x_2^0 + \delta_2'; \ldots; x_n^0 - \delta_n', x_n^0 + \delta_n')$$

nur so zu wählen, daß

$$0 < \delta_1' \leqq \delta_1, \quad 0 < \delta_2' \leqq \delta_2, \quad \ldots, \quad 0 < \delta_n' \leqq \delta_n$$

gilt und außerdem für ihre Produkte sich die aus (24) bestimmten Werte, d. h. die Werte $f_1, f_2, \ldots, f_\mu$, von $y_1^0, y_2^0, \ldots, y_\mu^0$ um weniger als um $\Delta_1, \Delta_2, \ldots, \Delta_\mu$ unterscheiden.[1]) Dann fällt der Punkt $(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_\mu)$ in $\mathcal{M}_0$, und zugleich mit (24) sind auch die Gleichungen (25) erfüllt.

Wir wählen jetzt (falls $m > \mu$ ist) irgendeine der übrigen Funktionen (17), etwa $y_{\mu+1}$, und zeigen, daß sie von den ersten Funktionen $y_1, y_2, \ldots, y_\mu$ abhängt. Setzt man in der Gleichung $y_{\mu+1} = f_{\mu+1}(x_1, \ldots, x_n)$ an Stelle der $x_1, \ldots, x_\mu$ die Funktionen (25) ein, so wird $y_{\mu+1}$ als (mittelbare) Funktion von $y_1, \ldots, y_\mu, x_{\mu+1}, \ldots, x_n$ dargestellt:

$$\begin{aligned} y_{\mu+1} &= \big(f_{\mu+1}\varphi_1(y_1, \ldots, y_\mu; x_{\mu+1}, \ldots, x_n), \ldots, \\ &\qquad \varphi_\mu(y_1, \ldots, y_\mu; x_{\mu+1}, \ldots, x_n); x_{\mu+1}, \ldots, x_n\big) \\ &\equiv F_{\mu+1}(y_1, \ldots, y_\mu; x_{\mu+1}, \ldots, x_n). \end{aligned} \tag{26}$$

Setzt man in dieser Gleichung an Stelle von $y_1, y_2, \ldots, y_\mu, y_{\mu+1}$ die Funktionen $f_1, f_2, \ldots, f_\mu, f_{\mu+1}$ ein, so wird sie auf Grund der obigen Bemerkung identisch von den x in $\mathcal{D}_0$ erfüllt.

Um sich von der Abhängigkeit der Funktion $y_{\mu+1}$ von den Funktionen $y_1, y_2, \ldots, y_\mu$ zu überzeugen, muß man nur noch zeigen, daß die Funktion $F_{\mu+1}$ in (26) in Wirklichkeit die Argumente $x_{\mu+1}, \ldots, x_n$ nicht enthält. Zu diesem Zweck genügt es zu zeigen, daß identisch in $y_1, \ldots, y_\mu, x_{\mu+1}, \ldots, x_n$ die Beziehungen

$$\frac{\partial F_{\mu+1}}{\partial x_{\mu+1}} = 0; \quad \frac{\partial F_{\mu+1}}{\partial x_{\mu+2}} = 0, \quad \ldots, \quad \frac{\partial F_{\mu+1}}{\partial x_n} = 0$$

gelten (vgl. Nr. 183). Wir gehen nur auf die erste Gleichung ein; die übrigen werden analog bewiesen.

[1]) Das geht auf Grund der Stetigkeit der Funktionen $f_1, f_2, \ldots, f_\mu$, die im Punkt M_0 die Werte $y_1^0, y_2^0, \ldots, y_\mu^0$ annehmen.

Wir differenzieren die Beziehungen (24) nach $x_{\mu+1}$, wobei wir $x_1, \ldots, x_\mu$ als Funktionen (25) von $y_1, \ldots, y_\mu, x_{\mu+1}, \ldots, x_n$ ansehen und erhalten

$$\left.\begin{aligned}
\frac{\partial f_1}{\partial x_1}\cdot\frac{\partial \varphi_1}{\partial x_{\mu+1}}+\cdots+\frac{\partial f_1}{\partial x_\mu}\cdot\frac{\partial \varphi_\mu}{\partial x_{\mu+1}}+\frac{\partial f_1}{\partial x_{\mu+1}}&=0,\\
\frac{\partial f_2}{\partial x_1}\cdot\frac{\partial \varphi_1}{\partial x_{\mu+1}}+\cdots+\frac{\partial f_2}{\partial x_\mu}\cdot\frac{\partial \varphi_\mu}{\partial x_{\mu+1}}+\frac{\partial f_2}{\partial x_{\mu+1}}&=0,\\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots&\\
\frac{\partial f_\mu}{\partial x_1}\cdot\frac{\partial \varphi_1}{\partial x_{\mu+1}}+\cdots+\frac{\partial f_\mu}{\partial x_\mu}\cdot\frac{\partial \varphi_\mu}{\partial x_{\mu+1}}+\frac{\partial f_\mu}{\partial x_{\mu+1}}&=0,
\end{aligned}\right\} \tag{27}$$

also lineare Gleichungen für die „Unbekannten" $\frac{\partial \varphi_1}{\partial x_{\mu+1}}, \ldots, \frac{\partial \varphi_\mu}{\partial x_{\mu+1}}$. Aus diesen μ linearen Gleichungen läßt sich die $(\mu+1)$-te lineare Gleichung

$$\frac{\partial f_{\mu+1}}{\partial x_1}\cdot\frac{\partial \varphi_1}{\partial x_{\mu+1}}+\cdots+\frac{\partial f_{\mu+1}}{\partial x_\mu}\cdot\frac{\partial \varphi_\mu}{\partial x_{\mu+1}}+\frac{\partial f_{\mu+1}}{\partial x_{\mu+1}}=0 \tag{27*}$$

herleiten. Daher ist die Determinante $(\mu+1)$-ter Ordnung, deren Elemente die Koeffizienten der „Unbekannten" und die freien Glieder in allen $\mu+1$ Gleichungen (27) und (27*) sind, d. h. die Determinante

$$\begin{vmatrix}
\frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_\mu} & \frac{\partial f_1}{\partial x_{\mu+1}}\\
\frac{\partial f_2}{\partial x_1} & \cdots & \frac{\partial f_2}{\partial x_\mu} & \frac{\partial f_2}{\partial x_{\mu+1}}\\
\cdot & & & \\
\frac{\partial f_\mu}{\partial x_1} & \cdots & \frac{\partial f_\mu}{\partial x_\mu} & \frac{\partial f_\mu}{\partial x_{\mu+1}}\\
\frac{\partial f_{\mu+1}}{\partial x_1} & \cdots & \frac{\partial f_{\mu+1}}{\partial x_\mu} & \frac{\partial f_{\mu+1}}{\partial x_{\mu+1}}
\end{vmatrix},$$

identisch 0 [der Rang der Matrix (19) ist μ]. Die linke Seite von (27*) ist aber nach Definition (26) der Funktion $F_{\mu+1}$ die Ableitung $\frac{\partial F_{\mu+1}}{\partial x_{\mu+1}}$. Nach (27*) ist diese Ableitung somit tatsächlich gleich 0.

In $F_{\mu+1}$ können also die Argumente $x_{\mu+1}, \ldots, x_n$ weggelassen werden: $y_{\mu+1}$ hängt nur von $y_1, \ldots, y_\mu$ ab. Das sollte aber gerade bewiesen werden.

In Beispiel 1 aus Nr. 215 lautet die Funktionalmatrix

$$\begin{pmatrix}
1 & 1 & 1\\
2x_1 & 2x_2 & 2x_n\\
x_2+x_3+\cdots+x_n & x_1+x_3+\cdots+x_n & x_1+x_2+\cdots+x_{n-1}
\end{pmatrix}$$

Wenn man zu den Elementen der dritten Zeile die mit $\frac{1}{2}$ multiplizierten der zweiten Zeile addiert, erhält man eine Zeile, die, wie die erste, aus gleichen Elementen besteht. Hieraus folgt schon, daß alle Determinanten dritter Ordnung verschwinden. Der Rang der Matrix ist gleich 2, und tatsächlich sind je zwei Funktionen unabhängig, die dritte von den anderen beiden abhängig.

Analog lassen sich die Überlegungen auch auf das Beispiel 2 aus Nr. 215 anwenden.

Zum Schluß bemerken wir, daß Fälle möglich sind, in denen in einem Teil eines untersuchten Bereichs eine bestimmte Abhängigkeit von Funktionen besteht, in einem anderen jedoch eine andere Abhängigkeit oder sogar Unabhängigkeit der Funktionen vorliegt.

3. Beispielsweise seien die Funktionen y_1 und y_2 von zwei unabhängigen Veränderlichen x_1 und x_2 in der x_1, x_2-Ebene durch

$$y_1 = \begin{cases} x_1^3 x_2^2 & \text{für} \quad x_1 \geqq 0, \\ 0 & \text{für} \quad x_1 < 0, \end{cases} \qquad y_2 = \begin{cases} x_1^2 x_2^3 & \text{für} \quad x_2 \geqq 0, \\ 0 & \text{für} \quad x_2 < 0 \end{cases}$$

gegeben. Man zeigt leicht, daß diese Funktionen nebst ihren Ableitungen in der ganzen Ebene stetig sind.

Der Rang der Funktionalmatrix ist im ersten Quadranten gleich 2, im zweiten und vierten gleich 1 und im dritten schließlich gleich 0. Nur im ersten Quadranten sind die Funktionen unabhängig.

§ 4. Variablensubstitution

217. Funktionen einer Veränderlichen. Das Ziel dieses Paragraphen besteht darin, eine Vorstellung von dem formalen Prozeß der Variablensubstitution zu vermitteln. Daher wollen wir uns hier nicht ablenken lassen durch eine Untersuchung aller Bedingungen, unter denen die durchzuführenden Operationen zulässig sind (was übrigens ebenfalls nicht schwierig wäre).

Ein beträchtlicher Teil des Inhaltes dieses Paragraphen hätte auch schon früher gebracht werden können; es schien uns jedoch zweckmäßig, alles, was mit Variablensubstitution zusammenhängt, an einer Stelle zu konzentrieren.

Es sei ein Ausdruck

$$W = F(x, y, y_x, y_{xx}, \ldots)$$

gegeben, der die unabhängige Veränderliche x, eine von ihr abhängige Funktion y und ferner Ableitungen dieser Funktion nach x bis zu einer bestimmten Ordnung enthält. Es sei gefordert, zu einem Ausdruck in neuen Veränderlichen t und u überzugehen, die mit x und y durch bestimmte Beziehungen, die sogenannten *Transformationsformeln*, zusammenhängen. Mit anderen Worten, W ist als Funktion von t, u und den entsprechenden Ableitungen von u nach t darzustellen.

Eine solche Variablensubstitution läßt sich gewöhnlich entweder durch das besondere Interesse motivieren, das den Veränderlichen t und u bei dem betrachteten Problem entgegengebracht wird, oder durch die Vereinfachungen, die sich dabei für den Ausdruck W ergeben.

Zunächst behandeln wir den Fall, daß nur die unabhängige Veränderliche transformiert wird, also die Transformationsformel unmittelbar nur x und t verknüpft. Diese Formel sei nach x aufgelöst:

$$x = \varphi(t). \tag{1}$$

Ist y eine Funktion von x, so mittelbar auch von t. Schon in Nr. 121 haben wir die entsprechenden Formeln hergeleitet:

$$\left.\begin{aligned} y_x &= \frac{y_t}{x_t}, \quad y_{xx} = \frac{x_t y_{tt} + x_{tt} y_t}{x_t^3}, \\ y_{xxx} &= \frac{x_t(x_t y_{ttt} - x_{ttt} y_t) - 3x_{tt}(x_t y_{tt} - x_{tt} y_t)}{x_t^5}, \ldots \end{aligned}\right\} \tag{2}$$

Da $x_t, x_{tt}, x_{ttt}, \ldots$ als bekannte Funktionen von t anzusehen sind [man erhält sie aus (1) durch Differentiation], braucht man nur statt $y_x, y_{xx}, \ldots$ die entsprechenden Ausdrücke in $t, y_t, y_{tt}, \ldots$ in W einzusetzen.

Ist die Transformationsformel in der impliziten Form

$$\Phi(x, t) = 0 \tag{3}$$

gegeben, so ist die Aufgabe im Grunde ebenso zu lösen, nur sind $x_t, x_{tt}, \ldots$ nach den Differentiationsregeln für implizite Funktionen zu berechnen.[1])

Nun wollen wir zu dem allgemeinen Fall übergehen, daß sowohl die unabhängige als auch die abhängige Variable transformiert werden. Die Transformationsformeln seien in der expliziten Gestalt

$$x = \varphi(t, u), \qquad y = \psi(t, u) \tag{4}$$

gegeben. Hängt y von x ab, so hiernach u von t, und damit sind auf Grund von (4) sowohl x als auch y mittelbare Funktionen von t. Nach den entsprechenden Differentiationsregeln ist also

$$x_t = \varphi_t + \varphi_u \cdot u_t,$$

$$y_t = \psi_t + \psi_u \cdot u_t;$$

$$x_{tt} = \varphi_{tt} + 2\varphi_{tu} \cdot u_t + \varphi_{uu} \cdot u_t^2 + \varphi_u \cdot u_{tt},$$

$$y_{tt} = \psi_{tt} + \cdots + \psi_u \cdot u_{tt}; \ldots$$

Wir machen den Leser darauf aufmerksam, daß wir mit $x_t, y_t, \ldots$ die „vollständigen" Ableitungen von x und y nach t bezeichnet haben, d. h. unter Berücksichtigung auch der Abhängigkeit der Funktion u von t; dagegen bezeichnen $\varphi_t, \psi_t, \ldots$ die Ableitungen nach t nur insoweit, als t in $\varphi, \psi, \ldots$ als eine von zwei Veränderlichen enthalten ist.

Setzen wir diese Ausdrücke in (2) ein, so finden wir Ausdrücke für die Ableitungen von y nach x in Abhängigkeit von t und u und den Ableitungen von u nach t, usw. Sind die Transformationsformeln nicht nach x und y aufgelöst, sondern in der impliziten Form

$$\Phi(x, y, t, u) = 0, \qquad \Psi(x, y, t, u) = 0 \tag{5}$$

gegeben, so müssen $x_t, y_t, x_{tt}, y_{tt}, \ldots$ hieraus nach den Differentiationsregeln für implizite Funktionen berechnet werden. Beispielsweise erhalten wir aus (5) durch Differentiation nach t (wobei nicht nur x und y, sondern auch u als Funktion von t anzusehen ist) die Beziehungen

$$\Phi_x x_t + \Phi_y y_t + \Phi_t + \Phi_u u_t = 0, \qquad \Psi_x x_t + \Psi_y y_t + \Psi_t + \Psi_u u_t = 0,$$

aus denen x_t, y_t berechnet werden können, usw.

In dem Spezialfall, daß die Transformationsformeln nach den neuen Veränderlichen aufgelöst sind,

$$t = \alpha(x, y), \qquad u = \beta(x, y), \tag{6}$$

kann man zunächst die soeben dargelegte allgemeine Methode benutzen. So erhält man durch Differentiation von (6) nach t (wobei x, y, u als Funktionen von t aufgefaßt werden) die Beziehungen $1 = \alpha_x x_t + \alpha_y y_t$, $u_t = \beta_x x_t + \beta_y y_t$ und hieraus

$$x_t = \frac{\beta_y - \alpha_y u_t}{\alpha_x \beta_y - \alpha_y \beta_x}, \quad y_t = \frac{\alpha_x u_t - \beta_x}{\alpha_x \beta_y - \alpha_y \beta_x},$$

also schließlich

$$y_x = \frac{\alpha_x u_t - \beta_x}{\beta_y - \alpha_y u_t}.$$

In diesem Fall geht man jedoch einfacher anders vor. Man führt die umgekehrte Transformation von t und u zu x und y aus. Durch Differentiation von (6) nach x (wobei y als Funktion von x angesehen wird) ergibt sich $t_x = \alpha_x + \alpha_y y_x$, $u_x = \beta_x + \beta_y y_x$, also

$$u_t = \frac{u_x}{t_x} = \frac{\beta_x + \beta_y y_x}{\alpha_x + \alpha_y y_x}, \tag{7}$$

woraus man für y_x denselben Ausdruck wie oben erhält.

---

[1]) Hängt W noch von x ab, so muß x noch mit Hilfe von (3) eliminiert werden.

Auch hier haben wir t_x, u_x und α_x, β_x zu unterscheiden: Die ersten bezeichnen die „vollständigen" Ableitungen nach x unter Berücksichtigung dessen, daß y von x abhängt; die zweiten bezeichnen die Ableitungen nach x als nach einem der Argumente von α und β.

Der Übergang von den Veränderlichen x, y zu den Veränderlichen t, u nach den Formeln (6) kann geometrisch als Punkttransformation der Ebene (oder eines Teiles der Ebene) gedeutet werden. Werden x, y als Koordinaten eines Punktes M der Ebene angesehen und t, u als Koordinaten eines Punktes P, so führt die Transformation den Punkt M in P über. Wir nehmen dann irgendeine Kurve $\mathcal{K}$ in der Ebene, deren Gleichung $y = f(x)$ ist; dem Funktionalzusammenhang zwischen x und y entspricht eine Abhängigkeit zwischen t und u; also $u = g(t)$, welche in der Ebene eine Kurve $\mathcal{L}$ bestimmt. Bei der betrachteten Transformation geht also die Kurve $\mathcal{K}$ in die Kurve $\mathcal{L}$ über. Zieht man im Punkt M der ersten Kurve die Tangente mit dem Richtungskoeffizienten y_x, so hat die zweite Kurve in P eine Tangente mit dem Richtungskoeffizienten u_t, der nach Formel (7) zu berechnen ist. Somit sind durch die Koordinaten von M auf $\mathcal{K}$ und den Richtungskoeffizienten der Tangente in M sowohl die Koordinaten des entsprechenden Punktes P auf der „transformierten" Kurve $\mathcal{L}$ als auch der Richtungskoeffizient der Tangente in P eindeutig bestimmt. Zieht man also durch einen Punkt P zwei Kurven, die sich in diesem Punkt berühren, so werden sich auch die transformierten Kurven in dem entsprechenden Punkt P berühren. Die betrachtete Punkttransformation der Ebene ist also „berührungstreu" (vgl. Nr. 218, Beispiel 5).

218. Beispiele.

1. Es sei die Gleichung $x^2 y_{xx} + x y_x + y = 0$ gegeben. Wir transformieren sie, indem wir $x = e^t$ setzen.

Nach den Formeln (2) ergibt sich $y_x = e^{-t} \cdot y_t$, $y_{xx} = e^{-2t}(y_{tt} - y_t)$, und die Gleichung geht in die einfachere Gestalt

$$y_{tt} + y = 0$$

über.

2. Man transformiere den Ausdruck

$$W = \frac{y_{xx} - y_x(1 - y_x)^2}{(1 + y_x)^3}$$

mit Hilfe von $x = t - y$.

Dem allgemeinen Schema ordnet sich diese Transformation unter, wenn man $x = t - u$, $y = u$ setzt. Nach (2) ist

$$W = \frac{x_t y_{tt} - x_{tt} y_t - y_t(x_t + y_t)^2}{(x_t + y_t)^3}.$$

Andererseits liefert die Transformationsformel $x_t = 1 - y_t$. Durch Einsetzen finden wir schließlich $W = y_{tt} - y_t$.

3. *Vertauschung der Rollen der Veränderlichen.* Jetzt sollen die unabhängige Veränderliche x und die abhängige Veränderliche y ihre Rollen tauschen; diese Transformation ordnet sich dem allgemeinen Schema unter, wenn man $x = u$, $y = t$ setzt. Wir wollen nun die Ableitung von y nach x durch die von x nach y ausdrücken.

Wir greifen auf die Formel (2) zurück und setzen $t = y$. Wegen $y_y = 1$, $y_{yy} = y_{yyy} = \cdots = 0$ ergibt sich sofort

$$y_x = \frac{1}{x_y}, \quad y_{xx} = -\frac{x_{yy}}{x_y^3}, \quad y_{xxx} = \frac{3x_{yy}^2 - x_y x_{yyy}}{x_y^5}, \ldots$$

Beispielsweise geht $W = y_x y_{xxx} + 3y_{xx}^2$ durch unsere Transformation in $W = -\dfrac{x_{yyy}}{x_y^5}$ über.

4. *Übergang zu Polarkoordinaten.* Sieht man x, y als rechtwinklige Punktkoordinaten an, so stellt $y = f(x)$ eine Kurve dar. Oft ist es nützlich, zu Polarkoordinaten r, θ überzugehen und die Kurve durch die Beziehung $r = g(\theta)$ zu beschreiben. Dann muß man natürlich die zur Be-

schreibung der geometrischen Elemente der Kurve nötigen Größen $x, y, y_x, y_{xx}, \ldots$ durch die entsprechenden Größen $\theta, r, r_\theta, r_{\theta\theta}, \ldots$ ersetzen.

Die Transformationsformeln lauten hier bekanntlich $x = r \cos\theta$, $y = r \sin\theta$. Differenziert man sie nach θ (unter Beachtung der Tatsache, daß r von θ abhängt), so erhält man

$$x_\theta = r_\theta \cos\theta - r \sin\theta,$$

$$y_\theta = r_\theta \sin\theta + r \cos\theta;$$

$$x_{\theta\theta} = r_{\theta\theta} \cos\theta - 2r_\theta \sin\theta - r\cos\theta,$$

$$y_{\theta\theta} = r_{\theta\theta} \sin\theta + 2r_\theta \cos\theta - r \sin\theta, \ldots$$

Hieraus ergibt sich nach (2), wobei θ statt t gesetzt wird,

$$y_x = \frac{r_\theta \sin\theta + r\cos\theta}{r_\theta \cos\theta - r \sin\theta}, \quad y_{xx} = \frac{r^2 + 2r_\theta^2 - rr_{\theta\theta}}{(r_\theta\cos\theta - r\sin\theta)^3}, \ldots$$

Somit erhält man beispielsweise für den Richtungskoeffizienten der Tangente

$$\tan\alpha = y_x = \frac{r_\theta \sin\theta + r\cos\theta}{r_\theta\cos\theta - r\sin\theta};$$

für den Tangens des Winkels ω zwischen Tangente und verlängertem Radiusvektor (Abb. 114),

$$\tan\omega = \tan(\alpha - \theta) = \frac{\tan\alpha - \tan\theta}{1 + \tan\alpha \cdot \tan\theta} = \frac{xy_x - y}{x + yy_x},$$

ergibt sich jetzt die einfache Formel $\tan\omega = \dfrac{r}{r_\theta}$. Eben deswegen beschreibt man die Lage der Tangente an eine in Polarkoordinaten gegebene Kurve durch den Winkel ω.

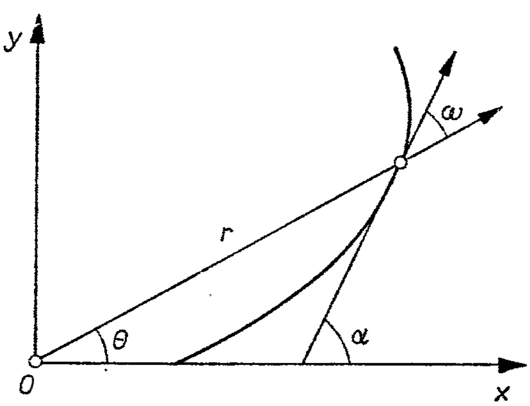

Abb. 114

Wir untersuchen noch den Ausdruck

$$R = \frac{(1 + y_x^2)^{3/2}}{y_{xx}},$$

der, wie wir in Nr. 251 sehen werden, ein wichtiges geometrisches Bestimmungsstück einer Kurve, nämlich den *Krümmungsradius*, liefert. Setzt man hier die oben gefundenen Ausdrücke für y_x und y_{xx} ein, so erhält man nach Vereinfachung

$$R = \frac{(r^2 + r_\theta^2)^{3/2}}{r^2 + 2r_\theta^2 - rr_{\theta\theta}}.$$

5. *Die Legendretransformation.* Die in Nr. 217 gestellte Aufgabe der Variablensubstitution läßt sich verallgemeinern, wenn man zuläßt, daß in den Transformationsformeln selbst schon Ableitungen vorkommen. Wir beschränken uns auf ein einziges Beispiel dieser Art:

$$t = y_x, \qquad u = x \cdot y_x - y,$$

die sogenannte *Legendretransformation.*

Wir differenzieren die zweite Transformationsformel nach x, wobei wir u als mittelbare Funktion von t ansehen; daß und wie t von x abhängt, erkennen wir aus der ersten Transformationsformel. So ergibt sich $u_t \cdot y_{xx} = y_x + x \cdot y_{xx} - y_x = x \cdot y_{xx}$. Hieraus folgt (unter der Annahme $y_{xx} \neq 0$) die Beziehung $u_t = x$. Unter Berücksichtigung beider Transformationsformeln folgt also

$$x = u_t, \qquad y = t \cdot u_t - u.$$

Dadurch wird die Symmetrie der Transformation deutlich: t, u, u_t ergeben sich genauso aus x, y, y_x wie diese Größen aus den erstgenannten.

Differenziert man $u_t = x$ ebenso nach x, so folgt

$$y_{tt} \cdot y_{xx} = 1, \quad \text{slso} \quad y_{xx} = \frac{1}{u_{tt}}.$$

Weiteres Differenzieren liefert $u_{ttt} \cdot y_{xx}^2 + u_{tt} \cdot y_{xxx} = 0$, also $y_{xxx} = -\dfrac{u_{ttt}}{u_t^3}$ usw.

Will man die Legendretransformation geometrisch deuten, so zeigt sich, daß es sich dabei keineswegs um eine Punkttransformation handelt. Zur Bestimmung der Koordinaten t, u von P genügt die Kenntnis der Koordinaten x, y von M nicht; man muß auch den Richtungskoeffizienten y_x der Tangente in diesem Punkt an die Kurve $y = f(x)$ kennen. Trotzdem wird auch hier die Kurve wieder berührungstreu in eine Kurve transformiert.[1])

219. Funktionen mehrerer Veränderlicher. Ersetzung der unabhängigen Veränderlichen. Wir wollen nun die Aufgabe in Angriff nehmen, den Auseruck

$$W = F\left(x, y, \ldots, z, \frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}, \ldots, \frac{\partial^2 z}{\partial x^2}, \frac{\partial^2 z}{\partial x\, \partial y}, \ldots\right),$$

der neben den unabhängigen Veränderlichen x, y, ... und der von ihnen abhängenden Funktion z auch partielle Ableitungen von z nach den Argumenten enthält, zu transformieren.

Aus denselben Gründen wie im oben betrachteten einfachsten Fall kann es auch hier nötig sein, zu neuen Veränderlichen überzugehen, die mit den alten durch Transformationsformeln verknüpft sind. Bezeichnet man die neuen unabhängigen Veränderlichen mit t, u, ... und die von ihnen abhängende Funktion mit v, so besteht die Aufgabe darin, W durch t, u, ..., v sowie durch die Ableitungen von v nach seinen Argumenten auszudrücken. Offenbar genügt es, das in bezug auf die alten Ableitungen $\dfrac{\partial z}{\partial x}, \dfrac{\partial z}{\partial y}, \ldots, \dfrac{\partial^2 z}{\partial x^2}, \dfrac{\partial^2 z}{\partial x\, \partial y}, \ldots$ zu tun. Zur Vereinfachung der Schreibweise wollen wir uns auf je zwei Veränderliche x, y bzw. t, u beschränken.

Auch hier beginnen wir mit dem Fall, daß nur die unabhängigen Variablen zu substituieren sind und die Transformationsformeln die x, y mit den t, u verknüpfen.

Überdies seien sie explizit, d. h. in der Gestalt

$$x = \varphi(t, u), \qquad y = \psi(t, u) \tag{8}$$

gegeben. Da wir z als mittelbare Funktion von t und u (über x, y) auffassen, erhalten wir nach der Kettenregel

$$\frac{\partial z}{\partial t} = \frac{\partial x}{\partial t}\frac{\partial z}{\partial x} + \frac{\partial y}{\partial t}\frac{\partial z}{\partial y}, \qquad \frac{\partial z}{\partial u} = \frac{\partial x}{\partial u}\frac{\partial z}{\partial x} + \frac{\partial y}{\partial u}\frac{\partial z}{\partial y}. \tag{9}$$

Zur Bestimmung der alten Ableitungen $\dfrac{\partial z}{\partial x}$ und $\dfrac{\partial z}{\partial y}$ haben wir somit ein lineares Gleichungs-

---

[1]) Solche berührungstreuen Transformationen spielen in verschiedenen Gebieten von Geometrie und Analysis eine wichtige Rolle. Sie werden auch *Berührungstransformationen* genannt. Die Punkttransformationen und die Legendretransformation sind spezielle Beispiele solcher Berührungstransformationen.

system vorliegen; sie ergeben sich also als Linearkombinationen der neuen:

$$\frac{\partial z}{\partial x} = A\frac{\partial z}{\partial t} + B\frac{\partial z}{\partial u}, \qquad \frac{\partial z}{\partial y} = C\frac{\partial z}{\partial t} + D\frac{\partial z}{\partial u}. \tag{10}$$

Dabei ist wichtig, daß A, B, C, D aus Ableitungen der Funktionen (8) zusammengesetzt sind, aber in keiner Weise von z abhängen. Diese Bemerkung erlaubt, die Formeln (10) auf $\frac{\partial z}{\partial x}$, $\frac{\partial z}{\partial y}$ statt auf z anzuwenden. Auf diesem Wege ergibt sich beispielsweise für $\frac{\partial^2 z}{\partial x^2}$ folgender Ausdruck:

$$\begin{aligned}\frac{\partial^2 z}{\partial x^2} &= \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial x}\right) = A\frac{\partial}{\partial t}\left(\frac{\partial z}{\partial x}\right) + B\frac{\partial}{\partial u}\left(\frac{\partial z}{\partial x}\right)\\ &= A\left(A\frac{\partial^2 z}{\partial t^2} + B\frac{\partial^2 z}{\partial t\,\partial u} + \frac{\partial A}{\partial t}\frac{\partial z}{\partial t} + \frac{\partial B}{\partial t}\frac{\partial z}{\partial u}\right)\\ &\quad + B\left(A\frac{\partial^2 z}{\partial t\,\partial u} + B\frac{\partial^2 z}{\partial u^2} + \frac{\partial A}{\partial u}\frac{\partial z}{\partial t} + \frac{\partial B}{\partial u}\frac{\partial z}{\partial u}\right).\end{aligned}$$

Durch Anwendung von (10) auf die Ableitungen zweiter Ordnung (statt auf z) erhält man Ausdrücke für die Ableitungen dritter Ordnung, usw.

Sind die Transformationsformeln nach den neuen Veränderlichen aufgelöst, $t = \alpha(x, y)$, $u = \beta(x, y)$, so ist es bequemer, nach der umgekehrten Methode vorzugehen, d. h. z als mittelbare Funktion von x, y (über t, u) anzusehen und t nach den alten Veränderlichen zu differenzieren. Dann ergibt sich

$$\frac{\partial z}{\partial x} = \frac{\partial t}{\partial x}\frac{\partial z}{\partial t} + \frac{\partial u}{\partial x}\frac{\partial z}{\partial u}, \qquad \frac{\partial z}{\partial y} = \frac{\partial t}{\partial y}\frac{\partial z}{\partial t} + \frac{\partial u}{\partial y}\frac{\partial z}{\partial u}. \tag{11}$$

Diesmal hängen die Koeffizienten $A = \frac{\partial t}{\partial x}$, $B = \frac{\partial u}{\partial x}$, $C = \frac{\partial t}{\partial y}$, $D = \frac{\partial u}{\partial y}$ von x und y, aber nicht von z ab. Durch wiederholte Anwendung der Formel (11) kann man auch hier die höheren Ableitungen berechnen, beispielsweise[1]):

$$\begin{aligned}\frac{\partial^2 z}{\partial x^2} &= \frac{\partial}{\partial x}\left(A\frac{\partial z}{\partial t} + B\frac{\partial z}{\partial u}\right)\\ &= \frac{\partial A}{\partial x}\frac{\partial z}{\partial t} + \frac{\partial B}{\partial x}\frac{\partial z}{\partial u} + A\frac{\partial}{\partial x}\left(\frac{\partial z}{\partial t}\right) + B\frac{\partial}{\partial x}\left(\frac{\partial z}{\partial u}\right)\\ &= \frac{\partial A}{\partial x}\frac{\partial z}{\partial t} + \frac{\partial B}{\partial x}\frac{\partial z}{\partial u} + A\left(A\frac{\partial^2 z}{\partial t^2} + B\frac{\partial^2 z}{\partial t\,\partial u}\right) + B\left(A\frac{\partial^2 z}{\partial t\,\partial u} + B\frac{\partial^2 z}{\partial u^2}\right).\end{aligned}$$

Liegen schließlich die Transformationsformeln in der impliziten Form

$$\Phi(x, y, t, u) = 0, \qquad \Psi(x, y, t, u) = 0 \tag{12}$$

vor, so kann man nach einer der angegebenen Methoden verfahren und

$$\frac{\partial x}{\partial t}, \quad \frac{\partial x}{\partial u}, \quad \frac{\partial y}{\partial t}, \quad \frac{\partial y}{\partial u} \quad \text{oder} \quad \frac{\partial t}{\partial x}, \quad \frac{\partial t}{\partial y}, \quad \frac{\partial u}{\partial x}, \quad \frac{\partial u}{\partial y}$$

nach den Differentiationsregeln für implizite Funktionen berechnen.

---

[1]) Auch hier ist eine analoge Bemerkung wie auf S. 445 zu machen: Da die Ausdrücke der alten Ableitungen durch die neuen die Veränderlichen x, y enthalten, kann es nach dem Einsetzen in W noch notwendig sein, x und y mit Hilfe der Transformationsformeln zu eliminieren. Der Leser wird auch in späteren Fällen auf diese Besonderheit stoßen.

220. Differentiale. Wir geben jetzt noch eine andere Methode an, um Ausdrücke in alten Ableitungen in solche in neuen zu transformieren, die besonders zweckmäßig ist, wenn in W nicht nur einzelne, sondern alle Ableitungen einer gegebenen Ordnung vorkommen. Es handelt sich um die *Methode zur Berechnung der vollständigen Differentiale*. Sie kann ebenfalls in zwei Formen dargestellt werden, je nachdem, ob t und u oder x und y als unabhängig angesehen werden.

Es seien zunächst t und u unabhängig; wir wollen alle Differentiale nach diesen Veränderlichen bilden (*direkte Methode*). Bilden wir die vollständigen Differentiale der Transformationsformeln (12), so können wir dx und dy linear durch dt und du ausdrücken:

$$\mathrm{d}x = \alpha\,\mathrm{d}t + \beta\,\mathrm{d}u, \qquad \mathrm{d}y = \gamma\,\mathrm{d}t + \delta\,\mathrm{d}u; \tag{13}$$

durch Differentiation dieser Formeln stellen wir d^2x und d^2y als homogene Polynome zweiten Grades in dt und du dar:

$$\mathrm{d}^2x = \varepsilon\,\mathrm{d}t^2 + \zeta\,\mathrm{d}t\,\mathrm{d}u^2 + \eta\,\mathrm{d}u^2, \qquad \mathrm{d}^2y = \theta\,\mathrm{d}t^2 + \iota\,\mathrm{d}t\,\mathrm{d}u + \varkappa\,\mathrm{d}u^2, \tag{14}$$

usw. Die Koeffizienten $\alpha, \beta, \ldots, \iota, \varkappa$ sind bekannte Funktionen von x, y, t, u.

Unter Benutzung der Invarianz des Differentials stellen wir nun dz auf zwei Arten dar:

$$\mathrm{d}z = \frac{\partial z}{\partial x}\mathrm{d}x + \frac{\partial z}{\partial y}\mathrm{d}y = \frac{\partial z}{\partial t}\mathrm{d}t + \frac{\partial z}{\partial u}\mathrm{d}u. \tag{15}$$

Setzt man statt dx und dy die Ausdrücke (13) ein und setzt dann die Koeffizienten von dt und du auf beiden Seiten der Gleichung gleich,[1] so erhält man die linearen Gleichungen

$$\alpha\frac{\partial z}{\partial x} + \gamma\frac{\partial z}{\partial y} = \frac{\partial z}{\partial t} \quad \text{und} \quad \beta\frac{\partial z}{\partial x} + \delta\frac{\partial z}{\partial y} = \frac{\partial z}{\partial u},$$

aus denen man $\dfrac{\partial z}{\partial x}$ und $\dfrac{\partial z}{\partial y}$ berechnen kann.

Analog kann man d^2z auf zwei Arten darstellen (unter Beachtung dessen, daß nicht x und y, sondern t und u die unabhängigen Veränderlichen sind):

$$\begin{aligned}\mathrm{d}^2z &= \frac{\partial^2 z}{\partial x^2}\mathrm{d}x^2 + 2\frac{\partial^2 z}{\partial x\,\partial y}\mathrm{d}x\,\mathrm{d}y + \frac{\partial^2 z}{\partial y^2}\mathrm{d}y^2 + \frac{\partial z}{\partial x}\mathrm{d}^2x + \frac{\partial z}{\partial y}\mathrm{d}^2y\\ &= \frac{\partial^2 z}{\partial t^2}\mathrm{d}t^2 + 2\frac{\partial^2 z}{\partial t\,\partial u}\mathrm{d}t\,\mathrm{d}u + \frac{\partial^2 z}{\partial u^2}\mathrm{d}u^2.\end{aligned} \tag{16}$$

Wir setzen nun statt dx, dy, d^2x, d^2y die Ausdrücke (13) und (14) ein und setzen die Koeffizienten von $\mathrm{d}t^2$, dt du und $\mathrm{d}u^2$ auf beiden Seiten von (16) gleich.[2] Das liefert uns ein System von drei linearen Gleichungen zur Bestimmung von $\dfrac{\partial^2 z}{\partial x^2}, \dfrac{\partial^2 z}{\partial x\,\partial y}, \dfrac{\partial^2 z}{\partial y^2}$ (da $\dfrac{\partial z}{\partial x}, \dfrac{\partial z}{\partial y}$ schon bekannt sind) usw.

Einfacher in der Durchführung ist die *umgekehrte* (*indirekte*) *Methode*, bei der x, y als unabhängig angesehen werden, so daß alle Differentiale nach diesen Veränderlichen zu bilden sind.

Durch sukzessives Differenzieren erhalten wir aus den Transformationsformeln (12)

$$\mathrm{d}t = a\,\mathrm{d}x + b\,\mathrm{d}y, \qquad \mathrm{d}u = c\,\mathrm{d}x + d\,\mathrm{d}y; \tag{17}$$

$$\mathrm{d}^2t = e\,\mathrm{d}x^2 + f\,\mathrm{d}x\,\mathrm{d}y + g\,\mathrm{d}y^2, \qquad \mathrm{d}^2u = h\,\mathrm{d}x^2 + i\,\mathrm{d}x\,\mathrm{d}y + j\,\mathrm{d}y^2, \tag{18}$$

usw. Hier sind die Koeffizienten $a, b, \ldots, i, j$ bekannte Funktionen von x, y, t und u.

Setzt man in (15) statt dt und du die Ausdrücke (17) ein und setzt die Koeffizienten von dx und dy auf beiden Seiten gleich, so folgt unmittelbar

$$\frac{\partial z}{\partial x} = a\frac{\partial z}{\partial t} + c\frac{\partial z}{\partial u}, \qquad \frac{\partial z}{\partial y} = b\frac{\partial z}{\partial t} + d\frac{\partial z}{\partial u}.$$

---

[1] Wir erinnern daran, daß die Gleichung $A\,\mathrm{d}t + B\,\mathrm{d}u = A'\,\mathrm{d}t + B'\,\mathrm{d}u$ für *beliebige* dt und du nur gelten kann, wenn $A = A'$ und $B = B'$ ist.

[2] Die Gleichung $A\,\mathrm{d}t^2 + B\,\mathrm{d}t\,\mathrm{d}u + C\,\mathrm{d}u^2 = A'\,\mathrm{d}t^2 + B'\,\mathrm{d}t\,\mathrm{d}u + C'\,\mathrm{d}u^2$ kann für *beliebige* dt und du nur gelten, wenn $A = A'$, $B = B'$, $C = C'$ ist.

An Stelle von (16) erhalten wir jetzt

$$d^2z = \frac{\partial^2 z}{\partial x^2}\,dx^2 + 2\,\frac{\partial^2 z}{\partial x\,\partial y}\,dx\,dy + \frac{\partial^2 z}{\partial y^2}\,dy^2$$

$$= \frac{\partial^2 z}{\partial t^2}\,dt^2 + 2\,\frac{\partial^2 z}{\partial t\,\partial u}\,dt\,du + \frac{\partial^2 z}{\partial u^2}\,du^2 + \frac{\partial z}{\partial t}\,d^2t + \frac{\partial z}{\partial u}\,d^2u.$$

Einsetzen der Ausdrücke (17), (18) und Koeffizientenvergleich bei dx^2, $dx\,dy$ und dy^2 liefert unmittelbar die Größen $\frac{\partial^2 z}{\partial x^2}$, $\frac{\partial^2 z}{\partial x\,\partial y}$, $\frac{\partial^2 z}{\partial y^2}$ usw.

221. Der allgemeine Fall einer Variablensubstitution. Zum Schluß wollen wir uns dem allgemeinen Fall zuwenden, daß sowohl die Veränderlichen als auch die Funktion zu transformieren sind. Die Transformationsformeln seien nach den alten Variablen aufgelöst:

$$x = \varphi(t, u, v), \qquad y = \psi(t, u, v), \qquad z = \chi(t, u, v). \tag{19}$$

Ist z eine Funktion von x und y, $z = f(x, y)$, so erhält man, wenn man x, y, z durch ihre Ausdrücke in t, u, v ersetzt, eine neue funktionale Abhängigkeit: v wird eine Funktion von t und u.

Wir fassen zunächst t und u als unabhängige Veränderliche auf (direkte Methode) und z durch Vermittlung von x und y als Funktion von t und u. So erhalten wir wie oben die Gleichungen (9) und daraus (10). Hier sind aber unter $\frac{\partial x}{\partial t}, \ldots, \frac{\partial z}{\partial u}$ die vollständigen partiellen Ableitungen von x, y, z nach t bzw. u zu verstehen, die sich aus (19) unter Berücksichtigung der Tatsache ergeben, daß v von t und u abhängt:

$$\frac{\partial x}{\partial t} = \frac{\partial \varphi}{\partial t} + \frac{\partial \varphi}{\partial v}\,\frac{\partial v}{\partial t}, \ldots, \frac{\partial z}{\partial u} = \frac{\partial \chi}{\partial u} + \frac{\partial \chi}{\partial v}\,\frac{\partial v}{\partial u}.$$

Die Koeffizienten A, B, C, D enthalten nicht nur t, u, v, sondern auch die Ableitungen $\frac{\partial v}{\partial t}$, $\frac{\partial v}{\partial u}$, und zwar rational. Die sukzessive Anwendung der Formeln (10) führt auch hier zu Ausdrücken für die zweiten Ableitungen, usw.

Sind die Transformationsformeln nach den neuen Veränderlichen aufgelöst,

$$t = \alpha(x, y, z), \qquad u = \beta(x, y, z), \qquad v = \gamma(x, y, z), \tag{20}$$

so verwendet man gewöhnlich die umgekehrte Methode, d. h., man sieht x und y als die unabhängigen Veränderlichen an. So erhält man

$$\frac{\partial v}{\partial x} = \frac{\partial v}{\partial t}\,\frac{\partial t}{\partial x} + \frac{\partial v}{\partial u}\,\frac{\partial u}{\partial x}, \qquad \frac{\partial v}{\partial y} = \frac{\partial v}{\partial t}\,\frac{\partial t}{\partial y} + \frac{\partial v}{\partial u}\,\frac{\partial u}{\partial y}.$$

An Stelle von $\frac{\partial t}{\partial x}, \ldots, \frac{\partial v}{\partial y}$ muß man hier die Ausdrücke einsetzen, die sich durch Differentiation der Formeln (20) nach x und y ergeben, wobei zu berücksichtigen ist, daß z eine Funktion von x und y ist:

$$\frac{\partial t}{\partial x} = \frac{\partial \alpha}{\partial x} + \frac{\partial \alpha}{\partial z}\,\frac{\partial z}{\partial x}, \ldots, \frac{\partial v}{\partial y} = \frac{\partial \gamma}{\partial y} + \frac{\partial \gamma}{\partial z}\,\frac{\partial z}{\partial y}.$$

So ergeben sich lineare Gleichungen für $\frac{\partial z}{\partial x}$ und $\frac{\partial z}{\partial y}$, aus denen diese Ableitungen leicht durch x, y, z, $\frac{\partial v}{\partial t}$ und $\frac{\partial v}{\partial u}$ ausdrückbar sind.

Die weiteren Ableitungen berechnet man am einfachsten folgendermaßen: Man differenziert die für $\frac{\partial z}{\partial x}$ (bzw. $\frac{\partial z}{\partial y}$) erhaltenen Ausdrücke nochmals nach x (bzw. y), wobei $\frac{\partial v}{\partial t}$ und $\frac{\partial v}{\partial u}$ als mittelbare Funktionen von x und y (über t und u) anzusehen sind, usw.

Haben die Transformationsformeln die allgemeine Gestalt

$$\bar{A}(x, y, z, t, u, v) = 0, \quad \bar{B}(x, y, z, t, u, v) = 0, \quad \bar{C}(x, y, z, t, u, v) = 0, \tag{21}$$

so kann man jede dieser Methoden anwenden. Dabei werden die Differentiationsregeln für implizite Funktionen benutzt.

Man kann auch die Methode der Bildung der vollständigen Differentiale verwenden. Wir beschränken uns hier auf den Fall, daß die alten Veränderlichen x und y als die unabhängigen Veränderlichen angesehen werden (umgekehrte Methode), so daß alle Differentiale nach diesen zu bilden sind.

Aus (21) ergeben sich durch sukzessive Differentiation

$$\left.\begin{aligned} dt &= a_1\,dx + a_2\,dy + a_3\,dz, \quad du = b_1\,dx + b_2\,dy + b_3\,dz,\\ dv &= c_1\,dx + c_2\,dy + c_3\,dz; \end{aligned}\right\} \tag{22}$$

$$\left.\begin{aligned} d^2t &= d_1\,dx^2 + d_2\,dx\,dy + d_3\,dy^2 + d_4\,dx\,dz + d_5\,dy\,dz + d_6\,dz^2 + a_3\,d^2z,\\ d^2u &= e_1\,dx^2 + \cdots + e_6\,dz^2 + b_3\,d^2z,\\ d^2v &= f_1\,dx^2 + \cdots + f_6\,dz^2 + c_3\,d^2z; \ldots \end{aligned}\right\} \tag{23}$$

Setzt man in $dv = \dfrac{\partial v}{\partial t}\,dt + \dfrac{\partial v}{\partial u}\,du$ statt dt, du, dv die Ausdrücke (22) ein, so folgt

$$c_1\,dx + c_2\,dy + c_3\,dz = \frac{\partial v}{\partial t}(a_1\,dx + a_2\,dy + a_3\,dz) + \frac{\partial v}{\partial u}(b_1\,dx + b_2\,dy + b_3\,dz),$$

also

$$dz = A\,dx + B\,dy, \tag{24}$$

wobei A, B die Ableitungen $\dfrac{\partial v}{\partial t}$, $\dfrac{\partial v}{\partial u}$ rational enthalten. Durch Koeffizientenvergleich mit $dz = \dfrac{\partial z}{\partial x}\,dx + \dfrac{\partial z}{dy}\,dy$ ergibt sich

$$\frac{\partial z}{\partial x} = A, \quad \frac{\partial z}{\partial y} = B.$$

Wir betrachten nun (t und u sind keine unabhängigen Veränderlichen) die Gleichung

$$d^2v = \frac{\partial^2 v}{\partial t^2}\,dt^2 + 2\,\frac{\partial^2 v}{\partial t\,\partial u}\,dt\,du + \frac{\partial^2 v}{\partial u^2}\,du^2 + \frac{\partial v}{\partial t}\,d^2t + \frac{\partial v}{\partial u}\,d^2u$$

und substituieren (22) und (23) für dt, du, d^2t, d^2u, d^2v und danach (24) für dz. Aus der sich ergebenden Gleichung läßt sich d^2z bestimmen,

$$d^2z = C\,dx^2 + 2D\,dx\,dy + E\,dy^2,$$

wobei C, D, E die Ableitungen $\dfrac{\partial v}{\partial t}$, $\dfrac{\partial v}{\partial u}$, $\dfrac{\partial^2 v}{\partial t^2}$, $\dfrac{\partial^2 v}{\partial t\,\partial u}$, $\dfrac{\partial^2 v}{\partial u^2}$ rational enthalten. Koeffizientenvergleich mit

$$d^2z = \frac{\partial^2 z}{\partial x^2}\,dx^2 + \frac{\partial^2 z}{\partial x\,\partial y}\,dx\,dy + \frac{\partial^2 z}{\partial y^2}\,dy^2$$

liefert

$$\frac{\partial^2 z}{\partial x^2} = C, \quad \frac{\partial^2 z}{\partial x\,\partial y} = D, \quad \frac{\partial^2 z}{\partial y^2} = E$$

usw.

Auch hier läßt sich die Variablentransformation geometrisch deuten. Sind (x, y, z), (t, u, v) die Koordinaten der Raumpunkte M und P, so ordnen die Transformationsformeln (20) jedem Punkt M einen Punkt P zu, charakterisieren also eine Punkttransformation des Raumes (oder eines Raumteils). Der Abhängigkeit zwischen x, y und z entspricht eine Abhängigkeit zwischen t, u und v, so daß jede Fläche $\mathcal{S}$ dabei in eine Fläche $\mathcal{T}$ übergeht.

Wir haben gesehen, daß die Werte von $x, y, z, \dfrac{\partial z}{\partial x}$ und $\dfrac{\partial z}{\partial y}$ diejenigen von $t, u, v, \dfrac{\partial v}{\partial t}$ und $\dfrac{\partial v}{\partial u}$ eindeutig bestimmen. Aus der bekannten Gleichung der Tangentialebene [vgl. Nr. 180, Formel (6)],

$$Z - z = \frac{\partial z}{\partial x}(X - x) + \frac{\partial z}{\partial y}(Y - y),$$

schließt man leicht, daß zwei sich in M berührenden Flächen $\mathscr{S}_1$ und $\mathscr{S}_2$ bei der betrachteten Transformation zwei Flächen $\mathscr{T}_1$ und $\mathscr{T}_2$ entsprechen, die einander in P berühren. Die Punkttransformation des Raumes ist berührungstreu (vgl. Nr. 222, Beispiel 7).

222. Beispiele.

1. *Übergang zu Polarkoordinaten.* Es sei z eine Punktfunktion in der Ebene: $z = f(M)$. Gewöhnlich wird die Lage eines Punktes durch seine rechtwinkligen Koordinaten x, y bestimmt, so daß z eine Funktion der Veränderlichen x, y ist. Oft ist es jedoch zweckmäßiger, die Lage eines Punktes durch seine Polarkoordinaten r, θ zu charakterisieren; dann muß man auf die neuen Veränderlichen transformieren. Wir vollziehen diesen Übergang mit den verschiedenen Methoden.

Direkte Methode. Es seien r, θ die unabhängigen Veränderlichen. Von den Transformationsformeln

$$x = r\cos\theta, \qquad y = r\sin\theta$$

ausgehend, erhält man nach dem Vorbild von (10)

$$\frac{\partial z}{\partial r} = \cos\theta\,\frac{\partial z}{\partial x} + \sin\theta\,\frac{\partial z}{\partial y}, \qquad \frac{\partial z}{\partial \theta} = -r\sin\theta\,\frac{\partial z}{\partial x} + r\cos\theta\,\frac{\partial z}{\partial y},$$

also

$$\frac{\partial z}{\partial x} = \cos\theta\,\frac{\partial z}{\partial r} - \frac{\sin\theta}{r}\,\frac{\partial z}{\partial \theta}, \qquad \frac{\partial z}{\partial y} = \sin\theta\,\frac{\partial z}{\partial r} + \frac{\cos\theta}{r}\,\frac{\partial z}{\partial \theta}, \tag{25}$$

so daß also $\cos\theta, -\dfrac{\sin\theta}{r}, \sin\theta, \dfrac{\cos\theta}{r}$ die Koeffizienten A, B, C, D sind. Alsdann ist

$$\begin{aligned}\frac{\partial^2 z}{\partial x^2} &= \cos\theta\,\frac{\partial}{\partial r}\left(\cos\theta\,\frac{\partial z}{\partial r} - \frac{\sin\theta}{r}\,\frac{\partial z}{\partial \theta}\right) - \frac{\sin\theta}{r}\,\frac{\partial}{\partial \theta}\left(\cos\theta\,\frac{\partial z}{\partial r} - \frac{\sin\theta}{r}\,\frac{\partial z}{\partial \theta}\right)\\ &= \cos^2\theta\,\frac{\partial^2 z}{\partial r^2} - \frac{2\sin\theta\cos\theta}{r}\,\frac{\partial^2 z}{\partial r\,\partial\theta} + \frac{\sin^2\theta}{r^2}\,\frac{\partial^2 z}{\partial \theta^2} + \frac{2\sin\theta\cos\theta}{r^2}\,\frac{\partial z}{\partial \theta} + \frac{\sin^2\theta}{r}\,\frac{\partial z}{\partial r}.\end{aligned}$$

Analog finden wir

$$\begin{aligned}\frac{\partial^2 z}{\partial y^2} = \sin^2\theta\,\frac{\partial^2 z}{\partial r^2} &+ \frac{2\sin\theta\cos\theta}{r}\,\frac{\partial^2 z}{\partial r\,\partial\theta} + \frac{\cos^2\theta}{r^2}\,\frac{\partial^2 z}{\partial \theta^2}\\ &- \frac{2\sin\theta\cos\theta}{r^2}\,\frac{\partial z}{\partial \theta} + \frac{\cos^2\theta}{r}\,\frac{\partial z}{\partial r}\end{aligned}$$

usw.

Umgekehrte Methode. Es seien x, y die unabhängigen Veränderlichen. Um die Formeln (11) anwenden zu können, müssen wir die Ableitungen $\dfrac{\partial r}{\partial x}, \dfrac{\partial \theta}{\partial x}, \dfrac{\partial r}{\partial y}, \dfrac{\partial \theta}{\partial y}$ kennen. Man könnte sie durch Auflösen der Gleichung, welche die alten mit den neuen Veränderlichen verknüpft, nach diesen Veränderlichen finden. Man kann aber die Differentiation impliziter Funktionen heranziehen, ohne die Gleichungen aufzulösen. Differenziert man die Transformationsformeln nach x und y und sieht dabei r und θ als Funktionen von x und y an, so erhält man

$$1 = \cos\theta \frac{\partial r}{\partial x} - r\sin\theta \frac{\partial \theta}{\partial x}, \quad 0 = \sin\theta \frac{\partial r}{\partial x} + r\cos\theta \frac{\partial \theta}{\partial x}$$

und

$$0 = \cos\theta \frac{\partial r}{\partial y} - r\sin\theta \frac{\partial \theta}{\partial y}, \quad 1 = \sin\theta \frac{\partial r}{\partial y} + r\cos\theta \frac{\partial \theta}{\partial y}.$$

Hieraus folgt

$$\frac{\partial r}{\partial x} = \cos\theta, \quad \frac{\partial \theta}{\partial x} = -\frac{\sin\theta}{r}, \quad \frac{\partial r}{\partial y} = \sin\theta, \quad \frac{\partial \theta}{\partial y} = \frac{\cos\theta}{r}$$

und nach (11) wieder (25) usw.

Berechnung der Differentiale. Wie eben seien x und y die unabhängigen Veränderlichen. Wir bilden die vollständigen Differentiale der Transformationsformeln:

$$\mathrm{d}x = \cos\theta\,\mathrm{d}r - r\sin\theta\,\mathrm{d}\theta, \qquad \mathrm{d}y = \sin\theta\,\mathrm{d}r + r\cos\theta\,\mathrm{d}\theta;$$

hieraus folgt

$$\mathrm{d}r = \cos\theta\,\mathrm{d}x + \sin\theta\,\mathrm{d}y, \quad \mathrm{d}\theta = \frac{-\sin\theta\,\mathrm{d}x + \cos\theta\,\mathrm{d}y}{r},$$

also

$$\mathrm{d}z = \frac{\partial z}{\partial r}\mathrm{d}r + \frac{\partial z}{\partial \theta}\mathrm{d}\theta = \left(\cos\theta \frac{\partial z}{\partial r} - \frac{\sin\theta}{r}\frac{\partial z}{\partial \theta}\right)\mathrm{d}x + \left(\sin\theta \frac{\partial z}{\partial r} + \frac{\cos\theta}{r}\frac{\partial z}{\partial \theta}\right)\mathrm{d}y,$$

was wieder auf (25) führt.

Für die zweiten Differentiale ergibt sich

$$\mathrm{d}^2 r = -\sin\theta\,\mathrm{d}\theta\,\mathrm{d}x + \cos\theta\,\mathrm{d}\theta\,\mathrm{d}y = \frac{\sin^2\theta\,\mathrm{d}x^2 - 2\sin\theta\cos\theta\,\mathrm{d}x\,\mathrm{d}y + \cos^2\theta\,\mathrm{d}y^2}{r},$$

$$\mathrm{d}^2\theta = \frac{-r(\cos\theta\,\mathrm{d}x + \sin\theta\,\mathrm{d}y)\,\mathrm{d}\theta - (\cos\theta\,\mathrm{d}y - \sin\theta\,\mathrm{d}x)\,\mathrm{d}r}{r^2}$$

$$= \frac{2\sin\theta\cos\theta\,\mathrm{d}x^2 - 2(\cos^2\theta - \sin^2\theta)\,\mathrm{d}x\,\mathrm{d}y - 2\sin\theta\cos\theta\,\mathrm{d}y^2}{r^2}.$$

Dann folgt für $\mathrm{d}^2 z$

$$\mathrm{d}^2 z = \frac{\partial^2 z}{\partial r^2}\mathrm{d}r^2 + 2\frac{\partial^2 z}{\partial r\,\partial\theta}\mathrm{d}r\,\mathrm{d}\theta + \frac{\partial^2 z}{\partial\theta^2}\mathrm{d}\theta^2 + \frac{\partial z}{\partial r}\mathrm{d}^2 r + \frac{\partial z}{\partial\theta}\mathrm{d}^2\theta$$

$$= \left(\cos^2\theta \frac{\partial^2 z}{\partial r^2} - \frac{2\sin\theta\cos\theta}{r}\frac{\partial^2 z}{\partial r\,\partial\theta} + \frac{\sin^2\theta}{r^2}\frac{\partial^2 z}{\partial\theta^2} + \frac{\sin^2\theta}{r}\frac{\partial z}{\partial r} + \frac{2\sin\theta\cos\theta}{r^2}\frac{\partial z}{\partial\theta}\right)$$

$$\times\,\mathrm{d}x^2 + 2(\ldots)\,\mathrm{d}x\,\mathrm{d}y + (\ldots)\,\mathrm{d}y^2,$$

woraus sich für die zweiten Ableitungen $\frac{\partial^2 z}{\partial x^2}, \ldots$ dieselben Ausdrücke wie oben ergeben.

Als Beispiel betrachten wir die Ausdrücke

$$W_1 = \left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2, \quad W_2 = \frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2}.$$

Nach unseren Formeln transformieren sie sich in

$$W_1 = \left(\frac{\partial z}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial z}{\partial \theta}\right)^2, \quad W_2 = \frac{\partial^2 z}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2 z}{\partial\theta^2} + \frac{1}{r}\frac{\partial z}{\partial r}.$$

2. *Übergang zu Kugelkoordinaten.* Das Analogon zu den Polarkoordinaten r, θ in der Ebene bilden im Raum die sogenannten *Kugelkoordinaten* ϱ, φ, θ die mit den rechtwinkligen Koordinaten x, y, z durch die Transformationsformeln

$$x = \varrho \sin\varphi \cos\theta, \qquad y = \varrho \sin\varphi \sin\theta, \qquad z = \varrho \cos\varphi$$

verknüpft sind.

Man transformiere die Ausdrücke

$$W_1 = \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial u}{\partial z}\right)^2, \quad W_2 = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}$$

auf Kugelkoordinaten [dabei ist $u = u(x, y, z)$].

Führt man die Transformation in zwei Schritten aus, indem man zunächst $x = r\cos\theta$, $y = r\sin\theta$ setzt und z ungeändert läßt und dann $z = \varrho\cos\varphi$, $r = \varrho\sin\varphi$ (also θ ungeändert) setzt, so kann man die Ergebnisse von Beispiel 1 benutzen.

Beispielsweise gilt für den zweiten Ausdruck

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} + \frac{1}{r}\frac{\partial u}{\partial r},$$

$$W_2 = \left(\frac{\partial^2 u}{\partial z^2} + \frac{\partial^2 u}{\partial r^2}\right) + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} + \frac{1}{r}\frac{\partial u}{\partial r}.$$

Auf Grund von Beispiel 1 läßt sich der Ausdruck in Klammern folgendermaßen schreiben:

$$\frac{\partial^2 u}{\partial \varrho^2} + \frac{1}{\varrho^2}\frac{\partial^2 u}{\partial \varphi^2} + \frac{1}{\varrho}\frac{\partial u}{\partial \varrho};$$

schließlich ist

$$\frac{\partial u}{\partial r} = \sin\varphi\frac{\partial u}{\partial \varrho} + \frac{\cos\varphi}{\varrho}\frac{\partial u}{\partial \varphi}.$$

Also ergibt sich

$$W_2 = \frac{\partial^2 u}{\partial \varrho^2} + \frac{1}{\varrho^2}\frac{\partial^2 u}{\partial \varphi^2} + \frac{1}{\varrho^2\sin^2\varphi}\frac{\partial^2 u}{\partial \theta^2} + \frac{2}{\varrho}\frac{\partial u}{\partial \varrho} + \frac{\cot\varphi}{\varrho^2}\frac{\partial u}{\partial \varphi}.$$

Analog ist

$$W_1 = \left(\frac{\partial u}{\partial \varrho}\right)^2 + \frac{1}{\varrho^2\sin^2\varphi}\left(\frac{\partial u}{\partial \theta}\right)^2 + \frac{1}{\varrho^2}\left(\frac{\partial u}{\partial \varphi}\right)^2$$

3. Man zeige, daß W_1 und W_2 bei einer linearen Transformation der rechtwinkligen Koordinaten in rechtwinklige Koordinaten,

$$x' = a_1 x + b_1 y + c_1 z, \quad y' = a_2 x + b_2 y + c_2 z, \quad z' = a_3 x + b_3 y + c_3 z,$$

bei der die a, b, c den („Orthogonalitäts"-)Bedingungen

$$a_i a_j = b_i b_j + c_i c_j = \begin{cases} 1 & \text{für} \quad i = j, \\ 0 & \text{für} \quad i \neq j \end{cases} \tag{26}$$

genügen, invariant sind.

Wir bilden die Differentiale, wobei x, y, z die unabhängigen Variablen seien:

$$\begin{aligned} \mathrm{d}x' &= a_1\,\mathrm{d}x + b_1\,\mathrm{d}y + c_1\,\mathrm{d}z, & \mathrm{d}^2x' &= 0, \\ \mathrm{d}y' &= a_2\,\mathrm{d}x + b_2\,\mathrm{d}y + c_2\,\mathrm{d}z, & \mathrm{d}^2y' &= 0, \\ \mathrm{d}z' &= a_3\,\mathrm{d}x + b_3\,\mathrm{d}y + c_3\,\mathrm{d}z, & \mathrm{d}^2z' &= 0. \end{aligned}$$

Dann ist

$$du = \frac{\partial u}{\partial x'}(a_1\,dx + b_1\,dy + c_1\,dz) + \frac{\partial u}{\partial y'}(a_2\,dx + b_2\,dy + c_2\,dz)$$

$$+ \frac{\partial u}{\partial z'}(a_3\,dx + b_3\,dy + c_3\,dz),$$

also

$$\frac{\partial u}{\partial x} = a_1\frac{\partial u}{\partial x'} + a_2\frac{\partial u}{\partial y'} + a_3\frac{\partial u}{\partial z'},$$

$$\frac{\partial u}{\partial y} = b_1\frac{\partial u}{\partial x'} + b_2\frac{\partial u}{\partial y'} + b_3\frac{\partial u}{\partial z'},$$

$$\frac{\partial u}{\partial z} = c_1\frac{\partial u}{\partial x'} + c_2\frac{\partial u}{\partial y'} + c_3\frac{\partial u}{\partial z'}.$$

Quadieren und Addieren liefert nach (26)

$$W_1 = \left(\frac{\partial u}{\partial x'}\right)^2 + \left(\frac{\partial u}{\partial y'}\right)^2 + \left(\frac{\partial u}{\partial z'}\right)^2$$

Alsdann ist

$$d^2u = \frac{\partial^2 u}{\partial x'^2}(a_1\,dx + b_1\,dy + c_1\,dz)^2$$

$$+ 2\frac{\partial^2 u}{\partial x'\,\partial y'}(a_1\,dx + b_1\,dy + c_1\,dz)(a_2\,dx + b_2\,dy + c_2\,dz) + \cdots.$$

W_2 ist die Summe der Koeffizienten von dx^2, dy^2, dz^2. Mit Hilfe von (26) ergibt sich leicht

$$W_2 = \frac{\partial^2 u}{\partial x'^2} + \frac{\partial^2 u}{\partial y'^2} + \frac{\partial^2 u}{\partial z'^2}.$$

4. Man transformiere die Gleichung

$$x^2\frac{\partial^2 w}{\partial x^2} + y^2\frac{\partial^2 w}{\partial y^2} + z^2\frac{\partial^2 w}{\partial z^2} + yz\frac{\partial^2 w}{\partial y\,\partial z} + zx\frac{\partial^2 w}{\partial z\,\partial x} + xy\frac{\partial^2 w}{\partial x\,\partial y} = 0$$

auf die neuen Veränderlichen t, u, v, wenn $x = uv$, $y = vt$, $z = tu$ gilt.

Direkte Methode. Es seien t, u, v die unabhängigen Veränderlichen. Dann ist

$$\frac{\partial w}{\partial t} = \frac{\partial w}{\partial y}v + \frac{\partial w}{\partial z}u, \quad \frac{\partial w}{\partial u} = \frac{\partial w}{\partial x}v + \frac{\partial w}{\partial z}t, \quad \frac{\partial w}{\partial v} = \frac{\partial w}{\partial x}u + \frac{\partial w}{\partial y}t.$$

Hieraus folgt

$$x\frac{\partial w}{\partial x} = -\frac{1}{2}t\frac{\partial w}{\partial t} + \frac{1}{2}u\frac{\partial w}{\partial u} + \frac{1}{2}v\frac{\partial w}{\partial v},$$

$$y\frac{\partial w}{\partial y} = \frac{1}{2}t\frac{\partial w}{\partial t} - \frac{1}{2}u\frac{\partial w}{\partial u} + \frac{1}{2}v\frac{\partial w}{\partial v},$$

$$z\frac{\partial w}{\partial z} = \frac{1}{2}t\frac{\partial w}{\partial t} + \frac{1}{2}u\frac{\partial w}{\partial u} - \frac{1}{2}v\frac{\partial w}{\partial v}.$$

Ferner ist

$$x^2 \frac{\partial^2 w}{\partial x^2} = x \frac{\partial}{\partial x}\left(x \frac{\partial w}{\partial x} - w\right)$$

$$= x \frac{\partial}{\partial x}\left(-\frac{1}{2} t \frac{\partial w}{\partial t} + \frac{1}{2} u \frac{\partial w}{\partial u} + \frac{1}{2} v \frac{\partial w}{\partial v} - w\right)$$

$$= \frac{1}{4} t^2 \frac{\partial^2 w}{\partial t^2} + \frac{1}{4} u^2 \frac{\partial^2 w}{\partial u^2} + \frac{1}{4} v^2 \frac{\partial^2 w}{\partial v^2} + \frac{1}{2} uv \frac{\partial^2 w}{\partial u\, \partial v}$$

$$- \frac{1}{2} vt \frac{\partial^2 w}{\partial v\, \partial t} - \frac{1}{2} tu \frac{\partial^2 w}{\partial t\, \partial u} + \frac{3}{4} t \frac{\partial w}{\partial t} - \frac{1}{4} u \frac{\partial w}{\partial u} - \frac{1}{4} v \frac{\partial w}{\partial v},$$

$$yz \frac{\partial^2 w}{\partial y\, \partial z} = z \frac{\partial}{\partial z}\left(y \frac{\partial w}{\partial y}\right) = z \frac{\partial}{\partial z}\left(\frac{1}{2} t \frac{\partial w}{\partial t} - \frac{1}{2} u \frac{\partial w}{\partial u} + \frac{1}{2} v \frac{\partial w}{\partial v}\right)$$

$$= \frac{1}{4} t^2 \frac{\partial^2 w}{\partial t^2} - \frac{1}{4} u^2 \frac{\partial^2 w}{\partial u^2} - \frac{1}{4} v^2 \frac{\partial^2 w}{\partial v^2} + \frac{1}{2} uv \frac{\partial^2 w}{\partial u\, \partial v}$$

$$+ \frac{1}{4} t \frac{\partial w}{\partial t} - \frac{1}{4} u \frac{\partial w}{\partial u} - \frac{1}{4} v \frac{\partial w}{\partial v},$$

usw. Durch Addition aller entsprechenden Ausdrücke und Weglassen konstanter Faktoren erhalten wir die transformierte Gleichung in der Gestalt

$$t^2 \frac{\partial^2 w}{\partial t^2} + u^2 \frac{\partial^2 w}{\partial u^2} + v^2 \frac{\partial^2 w}{\partial v^2} = 0.$$

Bisher haben wir nur die unabhängigen Veränderlichen transformiert. Bei den folgenden Beispielen wird auch die Funktion substituiert.

5. Man transformiere die Gleichung $x^2 \frac{\partial z}{\partial x} + y^2 \frac{\partial z}{\partial y} = z^2$, wenn

$$x = t, \quad y = \frac{t}{1 + tu}, \quad z = \frac{t}{1 + tv}$$

gilt.

Direkte Methode. Es seien t, u die unabhängigen Veränderlichen. Wir differenzieren die dritte Transformationsformel nach t und nach u, wobei wir z und v als Funktionen von t, u (die erste über x, y) ansehen:

$$\frac{\partial z}{\partial x} + \frac{\partial z}{\partial y} \cdot \frac{1}{(1 + tu)^2} = \frac{1 - t^2 \frac{\partial v}{\partial t}}{(1 + tu)^2}, \quad \frac{\partial z}{\partial y} \cdot \frac{-t^2}{(1 + tu)^2} = -\frac{t^2}{(1 + tv)^2} \frac{\partial v}{\partial u}.$$

Hieraus folgt

$$\frac{\partial z}{\partial x} = \frac{1}{(1 + tv)^2}\left(1 - t^2 \frac{\partial v}{\partial t} - \frac{\partial v}{\partial u}\right), \quad \frac{\partial z}{\partial y} = \frac{(1 + tu)^2}{(1 + tv)^2} \frac{\partial v}{\partial u}.$$

Die transformierte Gleichung lautet dann

$$\frac{\partial v}{\partial t} = 0.$$

Nun lösen wir dieselbe Aufgabe anders.

Umgekehrte Methode. Wir lösen die Transformationsformeln nach den neuen Veränderlichen auf,

$$t = x, \quad u = \frac{1}{y} - \frac{1}{x}, \quad v = \frac{1}{z} - \frac{1}{x},$$

und betrachten x und y als unabhängige Veränderliche. Wir differenzieren die dritte Beziehung nach x und nach y (v hängt über t, u von x, y ab):

$$-\frac{1}{z^2}\frac{\partial z}{\partial x}+\frac{1}{x^2}=\frac{\partial v}{\partial t}+\frac{\partial v}{\partial u}\frac{1}{x^2}, \qquad -\frac{1}{z^2}\frac{\partial z}{\partial y}=-\frac{\partial v}{\partial u}\frac{1}{y^2}$$

oder

$$\frac{\partial z}{\partial x}=z^2\left(\frac{1}{x^2}-\frac{\partial v}{\partial t}-\frac{1}{x^2}\frac{\partial v}{\partial u}\right), \qquad \frac{\partial z}{\partial y}=\frac{z^2}{y^2}\frac{\partial v}{\partial u} \qquad \text{usw.}$$

6. Man transformiere

$$W=\frac{\partial^2 z}{\partial x^2}-2\frac{\partial^2 z}{\partial x\,\partial y}+\frac{\partial^2 z}{\partial y^2}$$

auf t, u, v, wenn $t=x+y$, $u=\frac{y}{x}$, $v=\frac{z}{x}$ ist.

Berechnung der Differentiale. Die unabhängigen Veränderlichen seien x, y. Die Differentiale der Transformationsformeln lauten $\mathrm{d}t=\mathrm{d}x+\mathrm{d}y$, $\mathrm{d}u=-\frac{u}{x^2}\,\mathrm{d}x+\frac{1}{x}\,\mathrm{d}y$, $\mathrm{d}v=-\frac{z}{x^2}\,\mathrm{d}x+\frac{1}{x}\,\mathrm{d}z$. Betrachtet man v als (über t, u) mittelbare Funktion von x, y, so gilt für $\mathrm{d}v$

$$\mathrm{d}v=\frac{\partial v}{\partial t}\,\mathrm{d}t+\frac{\partial v}{\partial u}\,\mathrm{d}u=\frac{\partial v}{\partial t}(\mathrm{d}x+\mathrm{d}y)+\frac{\partial v}{\partial u}\left(-\frac{y}{x^2}\,\mathrm{d}x+\frac{1}{x}\,\mathrm{d}y\right).$$

Durch Koeffizientenvergleich folgt

$$\mathrm{d}z=\frac{z}{x}\,\mathrm{d}x+x\frac{\partial v}{\partial t}(\mathrm{d}x+\mathrm{d}y)+\frac{\partial v}{\partial u}\left(-\frac{y}{x}\,\mathrm{d}x+\mathrm{d}y\right). \tag{*}$$

Wir bilden nun die zweiten Differentiale der neuen Veränderlichen:

$$\mathrm{d}^2t=0, \quad \mathrm{d}^2u=\frac{2y}{x^3}\,\mathrm{d}x^2-\frac{2}{x^2}\,\mathrm{d}x\,\mathrm{d}y,$$

$$\mathrm{d}^2v=-\frac{2}{x^2}\,\mathrm{d}x\,\mathrm{d}z+\frac{2z}{x^3}\,\mathrm{d}x^2+\frac{1}{x}\,\mathrm{d}^2z;$$

andererseits ist

$$\begin{aligned}\mathrm{d}^2v&=\frac{\partial^2 v}{\partial t^2}\,\mathrm{d}t^2+2\frac{\partial^2 v}{\partial t\,\partial u}\,\mathrm{d}t\,\mathrm{d}u+\frac{\partial^2 v}{\partial u^2}\,\mathrm{d}u^2+\frac{\partial v}{\partial t}\,\mathrm{d}^2t+\frac{\partial v}{\partial u}\,\mathrm{d}^2u\\&=\frac{\partial^2 v}{\partial t^2}(\mathrm{d}x+\mathrm{d}y)^2+2\frac{\partial^2 v}{\partial t\,\partial u}(\mathrm{d}x+\mathrm{d}y)\left(-\frac{y}{x^2}\,\mathrm{d}x+\frac{1}{x}\,\mathrm{d}y\right)\\&\quad+\frac{\partial^2 v}{\partial u^2}\left(-\frac{y}{x^2}\,\mathrm{d}x+\frac{1}{x}\,\mathrm{d}y\right)^2+\frac{\mathrm{d}v}{\mathrm{d}u}\left(\frac{2y}{x^3}\,\mathrm{d}x^2-\frac{2}{x^2}\,\mathrm{d}x\,\mathrm{d}y\right).\end{aligned}$$

Setzt man die beiden Ausdrücke für d^2v gleich und ersetzt $\mathrm{d}z$ durch (*), so gelangt man zu der folgenden Gleichung für d^2z:

$$\begin{aligned}\mathrm{d}^2z&=2\frac{\mathrm{d}x}{x}\left[\frac{z}{x}\,\mathrm{d}x+x\frac{\partial v}{\partial t}(\mathrm{d}x+\mathrm{d}y)+\frac{\partial v}{\partial u}\left(-\frac{y}{x}\,\mathrm{d}x+\mathrm{d}y\right)\right]-\frac{2z}{x^2}\,\mathrm{d}x^2\\&\quad+x\left[\frac{\partial^2 v}{\partial t^2}(\mathrm{d}x+\mathrm{d}y)^2+2\frac{\partial^2 v}{\partial t\,\partial u}(\mathrm{d}x+\mathrm{d}y)\left(-\frac{y}{x^2}\,\mathrm{d}x+\frac{1}{x}\,\mathrm{d}y\right)\right.\\&\quad\left.+\frac{\partial^2 v}{\partial u^2}\left(-\frac{y}{x^2}\,\mathrm{d}x+\frac{1}{x}\,\mathrm{d}y\right)^2+\frac{\partial v}{\partial u}\left(\frac{2y}{x^3}\,\mathrm{d}x^2-\frac{2}{x^2}\,\mathrm{d}x\,\mathrm{d}y\right)\right].\end{aligned}$$

Hieraus ergeben sich $\frac{\partial^2 z}{\partial x^2}$, $\frac{\partial^2 z}{\partial x\,\partial y}$, $\frac{\partial^2 z}{\partial y^2}$ als Koeffizienten von dx^2, $2dx\,dy$, dy^2. Das gewünschte Resultat erhält man einfacher, wenn man bedenkt, daß d^2z in W übergeht, falls $dx = 1$, $dy = -1$ gesetzt wird. So ergibt sich

$$W = \frac{(x+y)^2}{x^3}\cdot\frac{\partial^2 v}{\partial u^2} = \frac{(1+u)^3}{t}\cdot\frac{\partial^2 v}{\partial u^2}.$$

7. *Die Legendretransformation.* Ähnlich wie in Nr. 218, Beispiel 5, bringen wir auch hier die Legendretransformation als Beispiel einer komplizierteren Substitution, bei der die Transformationsformeln auch Ableitungen enthalten. Wir setzen

$$t = \frac{\partial z}{\partial x}, \quad u = \frac{\partial z}{\partial y}, \quad v = x\frac{\partial z}{\partial x} + y\frac{\partial z}{\partial y} - z.$$

Wir wollen unter $z = f(x, y)$ eine bestimmte Funktion verstehen, die so beschaffen sei, daß

$$J = \frac{D(t, u)}{D(x, y)} = \frac{\partial^2 z}{\partial x^2}\frac{\partial^2 z}{\partial y^2} - \left(\frac{\partial^2 z}{\partial x\,\partial y}\right)^2 \neq 0 \tag{27}$$

gilt.

Wenn wir die dritte Transformationsformel nach x und nach y differenzieren und dabei v als (über t, u) mittelbare Funktion von x, y ansehen, so erhalten wir

$$\frac{\partial v}{\partial t}\frac{\partial^2 z}{\partial x^2} + \frac{\partial v}{\partial u}\frac{\partial^2 z}{\partial x\,\partial y} = x\frac{\partial^2 z}{\partial x^2} + y\frac{\partial^2 z}{\partial x\,\partial y},$$

$$\frac{\partial v}{\partial t}\frac{\partial^2 z}{\partial x\,\partial y} + \frac{\partial v}{\partial u}\frac{\partial^2 z}{\partial y^2} = x\frac{\partial^2 z}{\partial x\,\partial y} + y\frac{\partial^2 z}{\partial y^2}$$

und hieraus

$$x = \frac{\partial v}{\partial t}, \quad y = \frac{\partial v}{\partial u}, \quad \text{also} \quad z = t\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial u} - v, \tag{28}$$

die Transformation ist also symmetrisch.

Differenziert man die erste der Formeln (28) zunächst nach x, dann nach y, so gelangt man zu

$$1 = \frac{\partial^2 v}{\partial t^2}\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 v}{\partial t\,\partial u}\frac{\partial^2 z}{\partial x\,\partial y}, \qquad 0 = \frac{\partial^2 v}{\partial t\,\partial u}\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 v}{\partial u^2}\frac{\partial^2 z}{\partial x\,\partial y}$$

und

$$0 = \frac{\partial^2 v}{\partial t^2}\frac{\partial^2 z}{\partial x\,\partial y} + \frac{\partial^2 v}{\partial t\,\partial u}\frac{\partial^2 z}{\partial y^2}, \qquad 1 = \frac{\partial^2 v}{\partial t\,\partial u}\frac{\partial^2 z}{\partial x\,\partial y} + \frac{\partial^2 v}{\partial u^2}\frac{\partial^2 z}{\partial y^2}.$$

Da [vgl. Nr. 203, Formel (4)]

$$I = \frac{\partial^2 v}{\partial t^2}\frac{\partial^2 v}{\partial u^2} - \left(\frac{\partial^2 v}{\partial t\,\partial u}\right)^2 = \frac{D(x, y)}{D(t, u)} = \frac{1}{J} \neq 0$$

ist, folgt aus diesen Gleichungen

$$\frac{\partial^2 z}{\partial x^2} = \frac{\frac{\partial^2 v}{\partial u^2}}{I}, \quad \frac{\partial^2 z}{\partial x\,\partial y} = -\frac{\frac{\partial^2 v}{\partial t\,\partial u}}{I}, \quad \frac{\partial^2 z}{\partial y^2} = \frac{\frac{\partial^2 v}{\partial t^2}}{I}.$$

Betrachtet man x, y, z und t, u, v als Koordinaten von Raumpunkten, so kann man die Legendretransformation als Transformation des Raumes (aber nicht als Punkttransformation) auffassen. Eine durch eine funktionale Abhängigkeit zwischen z und x, y charakterisierte Fläche geht in eine durch die Abhängigkeit von v und t, u charakterisierte Fläche über. Da t, u, v

$\frac{\partial v}{\partial t}$, $\frac{\partial v}{\partial u}$ nur von $x, y, z, \frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}$ abhängen, ist die Legendretransformation berührungstreu.[1])

8. Man kann die Legendretransformation leicht auf den Fall eines Raumes beliebiger Dimension verallgemeinern. Es sei z eine Funktion von $x_1, x_2, \ldots, x_n$. Wir setzen

$$t_i = \frac{\partial z}{\partial x_i} \quad (i = 1, 2, \ldots, n) \quad \text{und} \quad v = \sum_{i=1}^{n} x_i \frac{\partial z}{\partial x_i} - z;$$

hier ist v die neue Funktion der neuen Veränderlichen $t_1, t_2, \ldots, t_n$. Wir setzen auch hier die Determinante

$$J = \begin{vmatrix} \frac{\partial^2 z}{\partial x_1^2} & \frac{\partial^2 z}{\partial x_1\, \partial x_2} & \frac{\partial^2 z}{\partial x_1\, \partial x_n} \\ \frac{\partial^2 z}{\partial x_2\, \partial x_1} & \frac{\partial^2 z}{\partial x_2^2} & \frac{\partial^2 z}{\partial x_2\, \partial x_n} \\ \\ \frac{\partial^2 z}{\partial x_n\, \partial x_1} & \frac{\partial^2 z}{\partial x_n\, \partial x_2} & \frac{\partial^2 z}{\partial x_n^2} \end{vmatrix}$$

als von 0 verschieden voraus.

Nun differenzieren wir die Definitionsformel für v nach x_k (wobei wir v als mittelbare Funktion von $x_1, \ldots, x_n$ ansehen):

$$\sum_{i=1}^{n} \frac{\partial v}{\partial t_i} \frac{\partial^2 z}{\partial x_i\, \partial x_k} = \sum_{i=1}^{n} x_i \frac{\partial^2 z}{\partial x_i\, \partial x_k} \qquad (k = 1, \ldots, n).$$

Wegen $J \neq 0$ folgt hieraus $\frac{\partial v}{\partial t_i} = x_i \; (i = 1, \ldots, n)$. Somit ist auch $z = \sum_{i=1}^{n} t_i \frac{\partial v}{\partial t_i} - v$, so daß auch diese allgemeine Transformation symmetrisch ist.

9. Abschließend betrachten wir noch ein Beispiel einer besonderen Transformation. Es sei $\varphi(u_1, \ldots, u_n; x_1, \ldots, x_n)$ eine in den $x_1, \ldots, x_n$ homogene Funktion zweiten Grades von $2n$ Veränderlichen. Die Determinante

$$H = \begin{vmatrix} \frac{\partial^2 \varphi}{\partial x_1^2} & \frac{\partial^2 \varphi}{\partial x_1\, \partial x_2} & \frac{\partial^2 \varphi}{\partial x_1\, \partial x_n} \\ \frac{\partial^2 \varphi}{\partial x_2\, \partial x_1} & \frac{\partial^2 \varphi}{\partial x_2^2} & \frac{\partial^2 \varphi}{\partial x_2\, \partial x_n} \\ \cdot & & \cdot \\ \frac{\partial^2 \varphi}{\partial x_n\, \partial x_1} & \frac{\partial^2 \varphi}{\partial x_n\, \partial x_2} & \frac{\partial^2 \varphi}{\partial x_n^2} \end{vmatrix}$$

sei von 0 verschieden. Wir setzen $t_i = \frac{\partial \varphi}{\partial x_i}$ und führen $t_1, \ldots, t_n$ als neue Veränderliche an Stelle der $x_1, \ldots, x_n$ ein. Dann geht φ in eine Funktion $\psi(u_1, \ldots, u_n; t_1, \ldots, t_n)$ über. Man zeige, daß

$$\text{(a)} \; \frac{\partial \psi}{\partial t_i} = x_i, \qquad \text{(b)} \; \frac{\partial \psi}{\partial u_i} = -\frac{\partial \varphi}{\partial u_i} \qquad (i = 1, \ldots, n)$$

gilt.

Wir differenzieren $\varphi = \psi$ nach x_k, wobei wir ψ als (über $t_1, \ldots, t_n$) mittelbare Funktion von $x_1, \ldots, x_n$ betrachten:

$$\frac{\partial \varphi}{\partial x_k} = \sum_{i=1}^{n} \frac{\partial \psi}{\partial t_i} \frac{\partial^2 \varphi}{\partial x_i\, \partial x_k} \qquad (k = 1, \ldots, n).$$

---

[1]) Auch hier beachte man die Fußnote auf S. 449.

Andererseits ist $\frac{\partial \varphi}{\partial x_k}$ eine homogene Funktion ersten Grades in $x_1, \ldots, x_n$. Dann ist nach der Eulerschen Formel (Nr. 188)

$$\frac{\partial \varphi}{\partial x_k} = \sum_{i=1}^{n} \frac{\partial^2 \varphi}{\partial x_k \, \partial x_i} x_i \qquad (k = 1, \ldots, n).$$

Vergleicht man die beiden Ausdrücke für $\frac{\partial \varphi}{\partial x_k}$, so ergibt sich auf Grund von $H \neq 0$ die Beziehung (a). Durch Differentiation nach u_i folgt

$$\frac{\partial \varphi}{\partial u_i} = \frac{\partial \psi}{\partial u_i} + \sum_{k=1}^{n} \frac{\partial \psi}{\partial t_k} \frac{\partial^2 \psi}{\partial x_k \, \partial u_i}.$$

Nun ist aber $\frac{\partial \varphi}{\partial u_i}$ eine homogene Funktion zweiten Grades in $x_1, \ldots, x_n$. Hier liefert die Eulersche Formel $\sum_{k=1}^{n} \frac{\partial}{\partial x_k} \left(\frac{\partial \varphi}{\partial u_i} \right) x_k = 2 \frac{\partial \varphi}{\partial u_i}$, und daraus folgen die Beziehungen (b).

VII. Anwendungen der Differentialrechnung in der Geometrie

§ 1. Analytische Darstellung von Kurven und Flächen

223. Kurven in der Ebene (in rechtwinkligen Koordinaten). In diesem Kapitel bringen wir einige Anwendungen der bisher untersuchten Begriffe, Tatsachen und Methoden der Differentialrechnung auf die Geometrie. Mit einigen haben wir uns ja schon in den Nummern 91, 141, 143, 145, 148 und 180 beschäftigt.

Wir halten es für nützlich, dem Leser einleitend die verschiedenen Arten der analytischen Darstellung von Kurven und Flächen ins Gedächtnis zurückzurufen; das wird in § 1 geschehen. Von vornherein wollen wir vereinbaren, daß die in diesem Kapitel betrachteten Funktionen in der Regel stetig sein und stetige erste Ableitungen nach ihren Argumenten besitzen sollen; nötigenfalls werden wir überdies Existenz und Stetigkeit von Ableitungen höherer Ordnung fordern.

Wir beginnen mit *ebenen* Kurven, und es liege ein rechtwinkliges x, y-Kooordinatensystem zugrunde.

Schon mehrfach haben wir Gleichungen der Gestalt

$$y = f(x) \quad \text{oder} \quad x = g(y) \tag{1}$$

betrachtet und die ihnen entsprechenden Kurven untersucht (Nr. 47, 91 und 146ff.). Ist eine Kurve in dieser Art gegeben, ist also eine der laufenden Koordinaten ihrer Punkte eine (eindeutige) explizite Funktion der anderen, so sagen wir, die Kurve sei in *expliziter Form* gegeben oder dargestellt. Diese Art der Darstellung ist einfach und anschaulich; wie wir sehen werden, kann jede andere Form der Darstellung in gewissem Sinne darauf zurückgeführt werden.

Im Zusammenhang mit der Theorie der impliziten Funktionen behandelten wir auch die sogenannte *implizite* Darstellung einer Kurve, d. h. die Darstellung einer Kurve durch eine Gleichung der Gestalt

$$F(x, y) = 0, \tag{2}$$

die weder nach x noch nach y aufgelöst ist (vgl. Nr. 205ff.). Man spricht dann von einer *impliziten Kurvengleichung.*

Aus den Sätzen über die Existenz einer impliziten Funktion (Nr. 205 und 206) folgt: Ist im Kurvenpunkt (x_0, y_0) bei stetigen F_x und F_y die Bedingung

$$F_x(x_0, y_0) \neq 0 \quad \text{bzw.} \quad F_y(x_0, y_0) \neq 0$$

erfüllt, so läßt sich wenigstens in einer Umgebung dieses Punktes die Kurve durch eine explizite Gleichung (1) der einen oder anderen Form darstellen, und die darin vorkommende Funktion f bzw. g ist nebst ihrer Ableitung stetig. Somit können nur solche Punkte (x_0, y_0) einer Kurve, in denen gleichzeitig

$$F_x(x_0, y_0) = 0, \quad F_y(x_0, y_0) = 0 \tag{3}$$

gilt, eine Singularität in dem Sinne aufweisen, daß in ihrer Umgebung die Kurve nicht durch eine explizite Gleichung (in keiner der beiden Formen) darstellbar ist. Kurvenpunkte, in denen die Beziehungen (3) gelten, heißen *singulär.*

In Nr. 236 beschäftigen wir uns mit dem Verhalten einer Kurve (2) in der Nähe eines singulären Punktes. Im allgemeinen können die singulären Punkte aus der Betrachtung ausgeschlossen werden; wir untersuchen daher zunächst eine Kurve nur in der Umgebung eines nichtsingulären Punktes.

Schließlich hatten wir es vielfach auch mit Gleichungen der Gestalt

$$x = \varphi(t), \quad y = \psi(t) \tag{4}$$

zu tun, die eine Abhängigkeit der laufenden Koordinaten eines Punktes von einem Parameter t zum Ausdruck bringen; diese Gleichungen beschreiben ebenfalls eine ebene Kurve (vgl. etwa Nr. 106). Man spricht hier von *Parametergleichungen* bzw. von einer *Parameterdarstellung einer Kurve.*

Wir betrachten den durch den Parameterwert $t = t_0$ bestimmten Kurvenpunkt, wobei wir annehmen, für $t = t_0$ sei $\varphi'(t_0) \neq 0$. Dann behält in der Umgebung dieses Wertes von t die Ableitung $x_t = \varphi'(t)$ aus Stetigkeitsgründen das Vorzeichen bei; dort ist also $x = \varphi(t)$ monoton (Nr. 132). Unter diesen Bedingungen kann man aber nach Nr. 83 und Nr. 94 auch t als eindeutige Funktion von x ansehen: $t = \theta(x)$. Dabei ist $\theta(x)$ nebst Ableitung stetig. Setzt man diesen Ausdruck für t in den Ausdruck für y ein, so erhält man eine funktionale Abhängigkeit der Veränderlichen y und x,

$$y = \psi\big(\theta(x)\big) = f(x),$$

wobei wiederum f nebst Ableitung stetig ist. Auf diese Weise haben wir (wenigstens in einem den betreffenden Punkt enthaltenden Kurvenstück) eine explizite Kurvengleichung erhalten. Analog kann man schließen, wenn zwar $\varphi'(t_0) = 0$, aber $\psi'(t_0) \neq 0$ ist; nur erhält man jetzt die explizite Kurvengleichung $x = g(y)$.

Wenn dagegen zugleich

$$x_t = \varphi'(t_0) = 0 \quad \text{und} \quad y_t = \psi'(t_0) = 0 \tag{5}$$

ist, kann die Kurve in der Umgebung des betrachteten Punktes nicht durch eine explizite Gleichung dargestellt werden; einen solchen Punkt nennen wir *singulär.*

In Nr. 237 gehen wir kurz auf die Gestalt einer Kurve (4) in der Nähe eines singulären Punktes ein; im allgemeinen werden wir aber hier nur nichtsinguläre Punkte untersuchen.

Es ist wichtig, darauf hinzuweisen, daß alle obigen Ausführungen über einen nichtsingulären Punkt (x_0, y_0), d. h. einen Punkt, für den (5) nicht zutrifft, überdies noch voraussetzen, daß dieser Punkt nur für einen einzigen Parameterwert $t = t_0$ auftritt, mit anderen Worten, ein *einfacher Punkt* ist. Ist dagegen (x_0, y_0) ein *Doppelpunkt* oder *mehrfacher Punkt* und entspricht er beispielsweise zwei verschiedenenen Parameterwerten $t = t_0$ und $t = t_1$, so liegt im allgemeinen dort eine Überschneidung zweier Kurvenstücke vor: Eines davon entspricht den t-Werten in der Nähe von t_0, das andere denen in der Nähe von t_1. In diesem Fall kann die ganze Kurve in der Umgebung des zugegebenen Punktes ebenfalls nicht durch eine explizite Gleichung dargestellt werden. Mehrfachpunkte müssen also ebenfalls zu den singulären Punkten gerechnet werden.[1])

Zusammenfassend können wir sagen: Wir haben nicht versucht, den Begriff einer Kurve geometrisch zu charakterisieren. *Eine Kurve ist für uns der geometrische Ort der Punkte, deren Koordinaten einer analytischen Beziehung der Gestalt* (1), (2) *bzw.* (4) *genügen,* unter der Annahme, daß die darin auftretenden Funktionen nebst ihren (ersten) Ableitungen stetig sind. Zwar können die auf diese verschiedenen Arten bestimmten geometrischen Bilder sich im Großen in ihrer äußeren Gestalt beträchtlich unterscheiden, im Kleinen, in der Umgebung eines nichtsingulären (uns im Fall der Parameterdarstellung überdies einfachen) Punktes ähneln sie alle dem einfachen Bild, das durch Gleichungen der Gestalt (1) geliefert wird.

224. Beispiele. Wir geben hier einen Überblick über besonders häufig vorkommende Kurven, von denen übrigens viele dem Leser aus der analytischen Geometrie bekannt sein dürften.

1. Die *Kettenlinie* (Abb. 41, S. 193). Ihre Gleichung lautet $y = \frac{a}{2}(e^{x/a} + e^{-x/a}) = a \cosh \frac{x}{a}$.
Diese Gestalt nimmt ein an beiden Enden aufgehängter ruhender biegsamer, nicht dehnbarer schwerer Faden (Kette, Draht) an. Daher stammt die Bezeichnung.

Die Kurve ähnelt in der Nähe des Scheitels A (vgl. Abb. 41) einer Parabel, mit wachsendem Abstand vom Scheitel strebt die Kurve aber steiler gegen ∞ als die Parabel. Die Strecke $\overline{OA} = a$ bestimmt ihre Gestalt genauer: Je kleiner a ist, desto steiler ist die Kurve. Die in Abb. 41 dargestellte Lage der Kurve ist keineswegs die einzig mögliche; sie ermöglicht es nur, der Kurvengleichung die einfachste Form zu geben.

2. Die auf die Symmetrieachsen bezogene *Ellipse* hat die Gleichung $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$. Da die Summe der Quadrate von $\frac{x}{a}$ und $\frac{x}{b}$ gleich 1 sein soll, liegt es nahe, diese Größen als Kosinus und Sinus eines Winkels t aufzufassen. Das führt auf die übliche Parameterdarstellung $x = a \cos t$, $y = b \sin t$; variiert t von 0 bis 2π, so wird die Ellipse entgegen dem Uhrzeigersinn umlaufen, beginnend vom Endpunkt $A(a, 0)$ der großen Achse.

Man könnte natürlich auch andere Ausdrücke benutzen, deren Quadratsumme gleich 1 ist, und beispielsweise

$$x = a\,\frac{1-u^2}{1+u^2}, \qquad y = b\,\frac{2u}{1+u^2}$$

setzen, wo u von $-\infty$ bis ∞ variiert. Da für $u \to \pm\infty$ die Beziehungen $x \to -a$, $y \to 0$ gelten, kann man vereinbaren, daß der Punkt $A'(-a, 0)$ für $u = \pm\infty$ angenommen wird.

Analog kann man im Fall der Hyperbel

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

auf Grund der bekannten Beziehung für den hyperbolischen Kosinus und Sinus

$$x = a \cosh t, \qquad y = b \sinh t \qquad (-\infty < t < \infty)$$

---

[1]) Es gibt übrigens einen Fall, daß ein zwei verschiedenen Parameterwerten entsprechender Punkt nicht als Mehrfachpunkt gilt: wenn nämlich dieser Punkt zwei Randwerten des Parameters entspricht und sich in ihm die Kurve schließt. Im Beispiel des Kreises

$$x = a \cos \theta, \qquad y = a \sin \theta \qquad (0 \leqq \theta \leqq 2\pi)$$

ist das der Punkt, der den Parameterwerten $\theta = 0$ und $\theta = 2\pi$ entspricht.

setzen. Eine andere Darstellung (es gibt bei jeder Kurve beliebig viele) ist

$$x = a\,\frac{1+u^2}{1-u^2}, \qquad y = b\,\frac{2u}{1-u^2} \qquad (-\infty < u < \infty;\ u \neq \pm 1).$$

Der Leser möge sich in allen Fällen Gedanken darüber machen, wie der Punkt mit variierendem Parameter die Kurve durchläuft.

3. Die *semikubische* (*Neilsche*) *Parabel* (Abb. 115) $y^2 - cx^3 = 0$ ($c > 0$). Hier ist der Ursprung $(0, 0)$ ein singulärer Punkt. Löst man die Gleichung nach y auf, so erhält man die expliziten Gleichungen für die beiden symmetrischen Zweige der Kurve:

$$y = \pm\sqrt{cx^3} = \pm\sqrt{c}\,x^{3/2}.$$

Da $y' = 0$ für $x = 0$ für beide Zweige gilt, berühren sie beide im Ursprung die x-Achse; es liegt eine Spitze (ein Rückkehrpunkt, vgl. Nr. 236) vor.

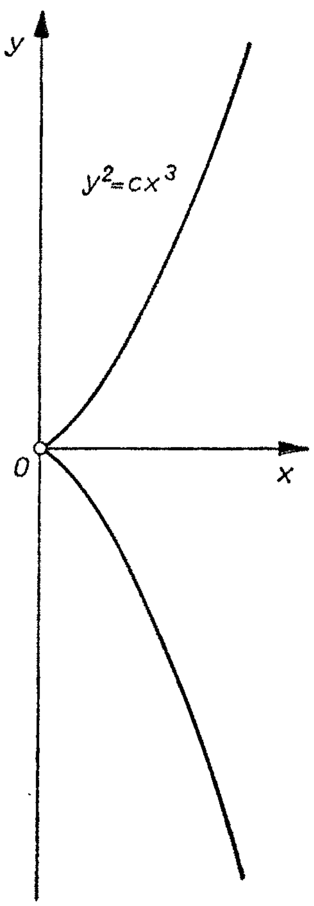

Abb. 115

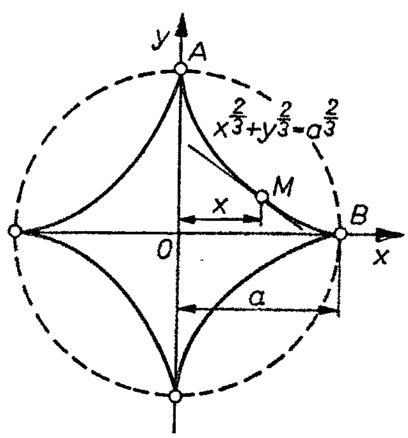

Abb. 116

4. Die *Astroide* (Abb. 116) $x^{2/3} + y^{2/3} = a^{2/3}$ ($a > 0$). Diese Gleichung gehört eigentlich nicht zu den Typen, auf die wir uns beschränken wollten: In jedem der Punkte $(\pm a, 0)$ und $(0, \pm a)$ wird eine der partiellen Ableitungen der linken Seite der Gleichung unendlich (ist also nicht mehr stetig). Übrigens ist es nicht schwer, die Gleichung von den Irrationalitäten zu befreien. Man kann sie leicht auf die Gestalt

$$(x^2 + y^2 - a^2)^3 + 27a^2x^2y^2 = 0$$

bringen. Auch bei dieser Darstellung sind die genannten Punkte singulär.

Aus der Kurvengleichung ist ersichtlich, daß die Kurve innerhalb des Kreises $x^2 + y^2 = a^2$ verläuft und symmetrisch zu beiden Achsen liegt. Bei der Untersuchung können wir uns daher auf den ersten Quadranten beschränken.

Wir lösen die Gleichung nach y auf,

$$y = (a^{2/3} - x^{2/3})^{3/2},$$

und erhalten durch Differentiation

$$y' = -(a^{2/3} - x^{2/3})^{1/2} \cdot x^{-1/3};$$

man sieht, daß für $x = 0$ die Tangente vertikal und für $x = a$ horizontal verläuft. Hieraus folgt, daß alle vier singulären Punkte Spitzen (Rückkehrpunkte) sind.

Um eine Parameterdarstellung der Astroide zu erhalten, benutzen wir die Tatsache, daß nach der Kurvengleichung die Summe der Quadrate von $\left(\frac{x}{a}\right)^{1/3}$ und $\left(\frac{y}{a}\right)^{1/3}$ gleich 1 sein muß. Setzen wir diese Größen gleich $\cos t$ und $\sin t$, so erhalten wir die Parametergleichungen

$$x = a\cos^3 t, \qquad y = a\sin^3 t \qquad (0 \leqq t \leqq 2\pi).$$

Da die Ableitungen

$$x_t = -3a\cos^2 t \sin t, \qquad y_t = 3a\sin^2 t \cos t$$

für $t = 0\ (2\pi), \frac{\pi}{2}, \pi, \frac{3\pi}{2}$ beide verschwinden, entsprechen diesen Parameterwerten die singulären Punkte; es sind natürlich dieselben wie oben.

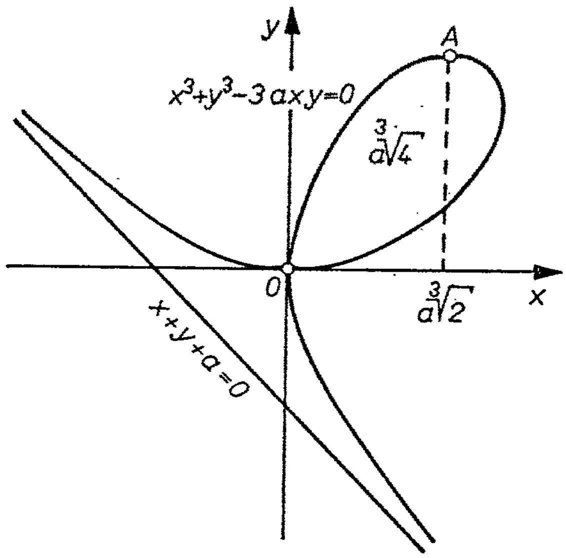

Abb. 117

5. Das *Cartesische Blatt*[1]) (Abb. 117) $x^3 + y^3 - 3axy = 0$ $(a > 0)$. Der Ursprung $(0, 0)$ ist ein singulärer Punkt: dort überschneidet sich die Kurve. Die Kurve hat die Asymptote $x + y + a = 0$, und zwar sowohl für $x \to \infty$ als auch für $x \to -\infty$.

Um das auch rechnerisch zu beweisen, dividieren wir die Gleichung durch x^3:

$$\left(\frac{y}{x}\right)^3 = 3a \cdot \frac{y}{x} \cdot \frac{1}{x} - 1.$$

Hieraus kann man zunächst schließen, daß etwa für $|x| > 3a$ die Größe $\left|\frac{y}{x}\right|$ beschränkt bleibt. Daraus folgt, daß für $x \to \pm\infty$ der Quotient $\frac{y}{x}$ gegen -1 strebt. Andererseits liefert die Gleichung die Beziehung

$$y + x = \frac{3axy}{x^2 - xy + y^2} = \frac{3a \cdot \frac{y}{x}}{1 - \frac{y}{x} + \left(\frac{y}{x}\right)^2},$$

[1]) Vgl. Nr. 210, Beispiel 2. Der Punkt $A\left(a\sqrt[3]{2}, a\sqrt[3]{4}\right)$, der dem Maximum von y als Funktion von x entspricht, ist in Abb. 117 angegeben.

so daß für $x \to \pm\infty$ die Größe $y + x$ gegen $-a$ strebt. Damit ist unsere Behauptung bewiesen (vgl. Nr. 148).

Nimmt man $t = \frac{y}{x}$ als Parameter und setzt in der Kurvengleichung $y = tx$, so erhält man die Parameterdarstellung

$$x = \frac{3at}{t^3+1}, \quad y = \frac{3at^3}{t^3+1} \qquad (-\infty < t < \infty;\, t \neq -1).$$

Für $t \to \pm\infty$ streben beide Koordinaten gegen 0. Man kann also annehmen, daß der Punkt (0, 0) sowohl dem Wert $t = 0$ als auch dem „Wert“ $t = \pm\infty$ entspricht. Läuft t von $-\infty$ bis -1, so läuft der Punkt (x, y), vom Ursprung ausgehend, längs des rechten Zweiges nach ∞. Läuft t von -1 bis 0, so läuft der Kurvenpunkt von ∞ bis in den Ursprung, und zwar längs des linken Zweiges. Wächst t von 0 bis ∞, so durchläuft der Punkt die Schleife (entgegen dem Uhrzeigersinn).

225. Mechanisch erzeugte Kurven. In der Reihe unserer Beispiele betrachten wir jetzt einige Kurven sozusagen mechanischen Ursprungs, die sich durch *Abrollen* einer Kurve auf einer anderen ergeben.

6. Die *Zykloide.* Wir stellen uns vor, auf der x-Achse rolle von links nach rechts ein Kreis vom Radius a mit dem Mittelpunkt A, und zwar ohne Schlupf. Die dabei von einem beliebigen Punkt der Kreisperipherie durchlaufene Kurve nennt man *Zykloide.* Wir wollen beispielsweise die Bahn des Punktes O bei einem Umlauf verfolgen.

Wir betrachten den Kreis nach einer gewissen Drehung in der neuen Lage. Der Kreis berührt die x-Achse jetzt im Punkt N. Der Berührungspunkt ist also auf der Geraden von O nach N gewandert. Unterdessen hat O selbst die Lage M eingenommen, mit dem Kreis also den Weg $\overset{\frown}{NM}$ zurückgelegt. Da der Kreis ohne Schlupf abrollen soll, sind diese Wegstrecken gleich:

$$\overset{\frown}{NM} = \overline{ON}.$$

Wählt man jetzt als Parameter zur Beschreibung der Lage des Punktes den Winkel $t = \sphericalangle NDM$, um den sich der Radius gedreht hat, der bei Beginn der Drehung die vertikale Lage $\overline{AO}$ hatte, so ergeben sich die Koordinaten x und y des Punktes M zu

$$x = \overline{OF} = \overline{ON} - \overline{FN} = \overset{\frown}{NM} - \overline{MG} = at - a\sin t,$$

$$y = \overline{FM} = \overline{NG} = \overline{ND} - \overline{GD} = a - a\cos t.$$

Für die Zykloide gelten also die Parametergleichungen

$$x = a(t - \sin t), \quad y = a(1 - \cos t) \quad (0 \leqq t \leqq 2\pi).$$

Variiert t von $-\infty$ bis ∞, so ergibt sich eine Kurve, die aus unendlich vielen Zweigen besteht, von denen Abb. 118 einen zeigt.

Da die Ableitungen

$$x_t = a(1 - \cos t), \quad y_t = a\sin t$$

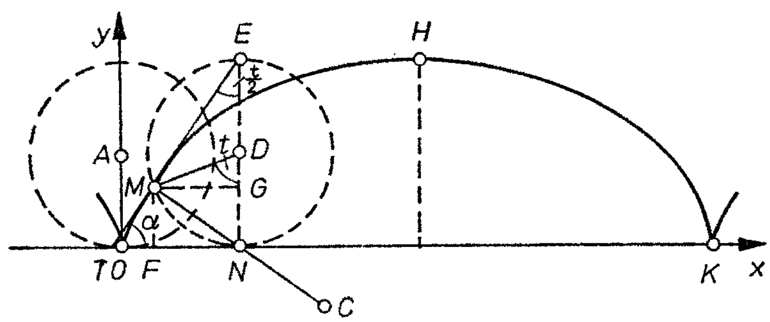

Abb. 118

für $t + 2k\pi$ ($k = 0, \pm 1, \pm 2, \ldots$) beide verschwinden, entsprechen diesen Werten die singulären Punkte der Kurve. Nun ist aber [vgl. Nr. 106, Formel (10)]

$$y_x = \frac{y_t}{x_t} = \frac{\sin t}{1 - \cos t} = \cot \frac{t}{2},$$

so daß beispielsweise für $t \to \pm 0$ (bzw. für $x \to \pm 0$) die Ableitung y_x gegen $\pm\infty$ strebt. Offenbar ist im Ursprung ebenso wie in den anderen singulären Punkten die Tangente vertikal; es liegt eine Spitze (ein Rückkehrpunkt; Nr. 237) vor.

7. *Epizykloiden* und *Hypozykloiden*. Rollt ein Kreis ohne Schlupf außen auf einem anderen ab, so durchläuft ein Punkt des Umfangs des rollenden Kreises eine sogenannte *Epizykloide*. Rollt er dagegen innen ab, so erhält man eine *Hypozykloide*. Wir wollen die Gleichung der Epizykloide aufstellen. Als Ursprung nehmen wir den Mittelpunkt O des ruhenden Kreises, die x-Achse legen wir durch den Berührungspunkt der beiden Kreise im Anfangszustand. Wir wollen die Bahn dieses Berührungspunktes beim Abrollen verfolgen (Abb. 119). Ist der rollende Kreis in die Lage übergegangen, die Abb. 119 zeigt, so ist A in M übergegangen. Der geometrische Ort von M ist zu bestimmen.

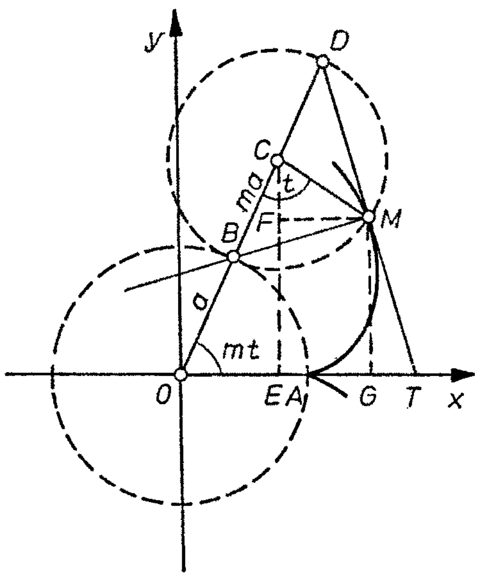

Abb. 119

Es sei a der Radius des ruhenden, ma der des rollenden Kreises. Als Parameter wählen wir den Winkel $t = \sphericalangle MCB$ zwischen dem Radius $\overline{CM}$, der den Mittelpunkt des rollenden Kreises mit dem uns interessierenden Punkt auf der Peripherie verbindet, und dem Radius $\overline{CB}$ durch den Berührungspunkt. Zu Anfang der Bewegung ist dieser Winkel gleich 0.

Zunächst wollen wir untersuchen, worin hier das Rollen ohne Schlupf zum Ausdruck kommt. Der vom Berührungspunkt auf dem ruhenden Kreis zurückgelegte Weg $\widehat{AB}$ muß gleich dem Bogen $\widehat{MB}$ sein, den der Berührungspunkt auf dem rollenden Kreis zurückgelegt hat:

$$a \cdot \sphericalangle AOB = ma \cdot \sphericalangle MCB = mat, \quad \text{also} \quad \sphericalangle AOB = mt.$$

Für die Koordinate x von M ergibt sich

$$x = \overline{OG} = \overline{OE} + \overline{FM} = (a + ma) \cos mt + ma \sin \sphericalangle FCM;$$

es ist aber $\sphericalangle FCM = \sphericalangle BCM - \sphericalangle OCE$ und $\sphericalangle OCE = \frac{\pi}{2} - mt$, also

$$\sphericalangle FCM = (1 + m)\, t - \frac{\pi}{2} \quad \text{und} \quad \sin \sphericalangle FCM = -\cos (1 + m)\, t.$$

Somit ist schließlich

$$x = a[(1 + m) \cos mt + m \cos (1 + m)\, t].$$

In ähnlicher Weise folgt

$$y = a[(1 + m) \sin mt - m \sin (1 + m) t].$$

Diese Gleichungen liefern die *Parameterdarstellung der Epizykloide.*

Wenn der rollende Kreis den ruhenden Kreis wieder in demselben Punkt wie zu Anfang der Bewegung berührt (d. h. für $t = 2\pi$), so hat M einen Kurvenzweig durchlaufen. Bei dem weiteren Abrollen beschreibt er den nächsten Zweig, der dem ersten ähnlich ist, usw. Die Ableitungen

$$x_t = -m(m + 1)\, a[\sin mt - \sin (1 + m) t],$$
$$y_t = m(m + 1)\, a[\cos mt - \cos (1 + m) t]$$

verschwinden gleichzeitig für $t = 2k\pi$ ($k = 0, \pm 1, \pm 2, \ldots$), d. h. jedesmal, wenn der betrachtete Punkt des rollenden Kreises Berührungspunkt ist. Die entsprechenden Punkte der Kurve sind singulär (Rückkehrpunkte).

Im Fall der *Hypozykloide* erhält man in ähnlicher Weise die *Parametergleichungen*

$$y = a[(1 - m) \cos mt + m \cos (1 - m) t],$$
$$y = a[-(1 - m) \sin mt + m \sin (1 - m) t].$$

Auch hier bezeichnet m das Verhältnis des Radius des rollenden zu dem des ruhenden Kreises. Man bemerkt übrigens leicht, daß man diese Gleichungen aus denen der Epizykloide erhält, wenn man m durch $-m$ ersetzt.

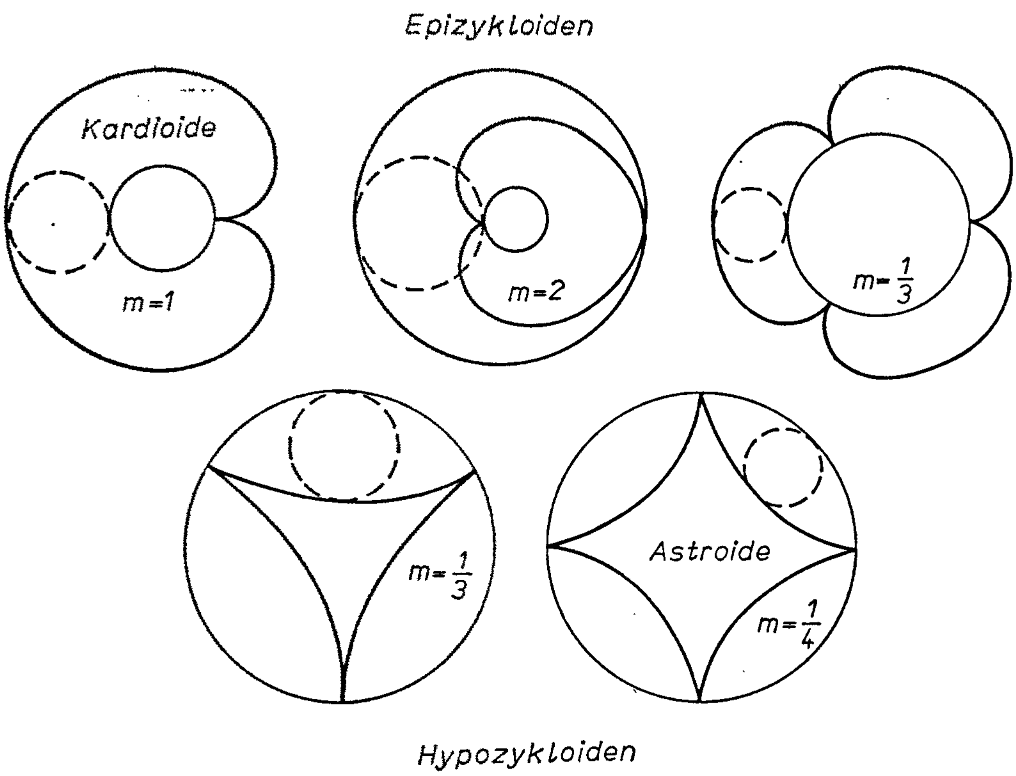

Abb. 120

Abb. 120 zeigt Epizykloiden für $m = 1, 2, \frac{1}{3}$ und Hypozykloiden für $m = \frac{1}{3}$ und $\frac{1}{4}$. In der letzten erkennt man die *Astroide.*[1])

8. Die *Kreisevolvente.* Wir stellen uns vor, um einen Kreis vom Radius a um den Mittelpunkt O sei im Uhrzeigersinn ein Faden gewickelt; sein freies Ende befindet sich im Punkt A. Nun

[1]) Setzt man in den Gleichungen der Hypozykloide $m = \frac{1}{4}$ und ersetzt t durch $-4t$, so erhält man genau die in Beispiel 4 angegebenen Gleichungen.

werde der Faden (entgegen dem Uhrzeigersinn) abgewickelt, und zwar so, daß er immer gespannt bleibt. Die dabei von dem Fadenende beschriebene Kurve heißt *Kreisevolvente* (vgl. Nr. 254 und 156).

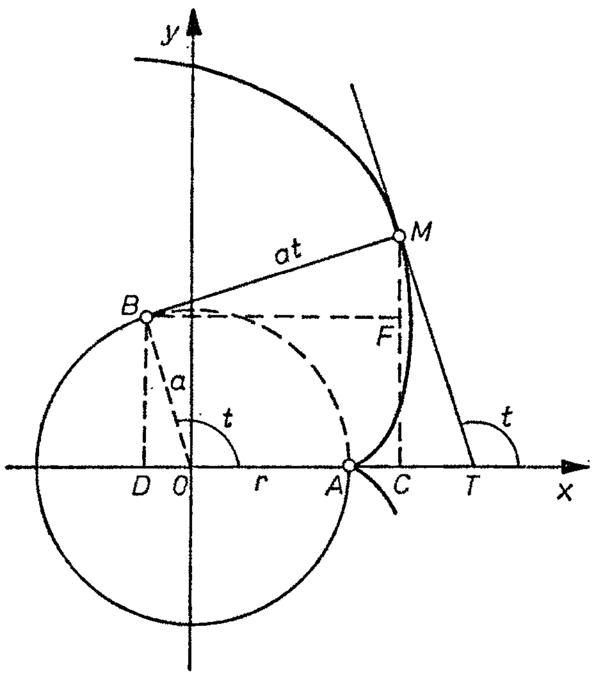

Abb. 121

Wir legen den Ursprung des Koordinatensystems in den Punkt O (Abb. 121) und die x-Achse durch A. Ist der Teil $\widehat{AB}$ des Fadens abgewickelt, so hat er die Lage $\overline{BM}$, die zum Kreis tangential ist, und A ist in den Punkt M übergegangen. Somit ist $\widehat{AB} = \overline{BM}$. Als Parameter nehmen wir $t = \sphericalangle AOB$ zwischen den Radien $\overline{OA}$ und $\overline{OB}$. Die Koordinaten x und y von M ergeben sich folgendermaßen:

$$x = \overline{DC} - \overline{DO} = \overline{BF} - \overline{DO} = \overline{BM} \sin \sphericalangle BMC - \overline{OB} \cos \sphericalangle DOB;$$

nun ist aber $\overline{BM} = \widehat{AB} = at$ und $\sphericalangle BMC = \sphericalangle DOB = \pi - t$, also

$$x = at \sin(\pi - t) - a \cos(\pi - t) = a(t \sin t + \cos t).$$

Ferner ist

$$y = \overline{CM} = \overline{CF} + \overline{FM} = \overline{DB} + \overline{FM} = \overline{OB} \sin \sphericalangle DOB + \overline{BM} \cos \sphericalangle BMC$$
$$= a(\sin t - t \cos t).$$

Somit erhält man folgende Parametergleichungen:

$$x = a(t \sin t + \cos t), \quad y = a(\sin t - t \cos t).$$

Der einzige singuläre Punkt entspricht dem Wert $t = 0$, für den die beiden Ableitungen

$$x_t = at \cos t, \qquad y_t = at \sin t$$

verschwinden.

Wir empfehlen dem Leser, sich davon zu überzeugen, daß man dieselbe Kurve erhält, wenn man eine Gerade ohne Schlupf auf einem Kreis abrollt und die Bahn eines Punktes der Geraden verfolgt.

226. Kurven in der Ebene (in Polarkoordinaten). Beispiele. In vielen Fällen erweist es sich als einfacher, Kurven durch eine Gleichung in Polarkoordinaten darzustellen, nachdem man die gegenseitige Abhängigkeit der laufenden Polarkoordinaten r, θ eines Punktes der Kurve ermittelt hat. Den Polarwinkel θ rechnen wir von der Polarachse an, und zwar positiv in der dem Uhrzeigersinn entgegengesetzten Drehrichtung. Den Radiusvektor r rechnen wir positiv bzw. negativ, je nachdem, ob er in der durch θ bestimmten bzw. in der entgegengesetzten Richtung gemessen wird.

Wie bei rechtwinkligen Koordinaten kann auch hier die funktionale Abhängigkeit zwischen r und θ explizit, implizit oder in Parameterform ausgedrückt werden. Wir beschränken uns in der Hauptsache auf den einfachsten Fall, daß die Kurve durch eine explizite Gleichung der Form $r = f(\theta)$ gegeben ist.

Geht man zu rechtwinkligen Koordinaten über und nimmt wie üblich den Pol als Ursprung, die Polarachse als x-Achse, so liefern die Gleichungen

$$x = r\cos\theta = f(\theta)\cos\theta, \quad y = r\sin\theta = f(\theta)\sin\theta$$

eine Parameterdarstellung der Kurve mit dem Polarwinkel θ als Parameter. Die sich hierbei ergebenden Funktionen von θ sind mit f stetig und besitzen stetige Ableitungen.

Die Formeln

$$x_\theta = r_\theta\cos\theta - r\sin\theta, \quad y_\theta = r_\theta\sin\theta + r\cos\theta$$

zeigen, daß ein singulärer Punkt (im Sinne von Nr. 223) nur für $r = r_\theta = 0$ auftreten kann.

Beispiele.

1. Die *Archimedische Spirale* $r = a\theta$, $a > 0$ (Abb. 122). Die Kurve kann man als Bahn eines Punktes deuten, der sich mit konstanter Geschwindigkeit auf einem Strahl vom Pol aus bewegt, während sich dieser Strahl mit konstanter Winkelgeschwindigkeit um den Pol dreht.

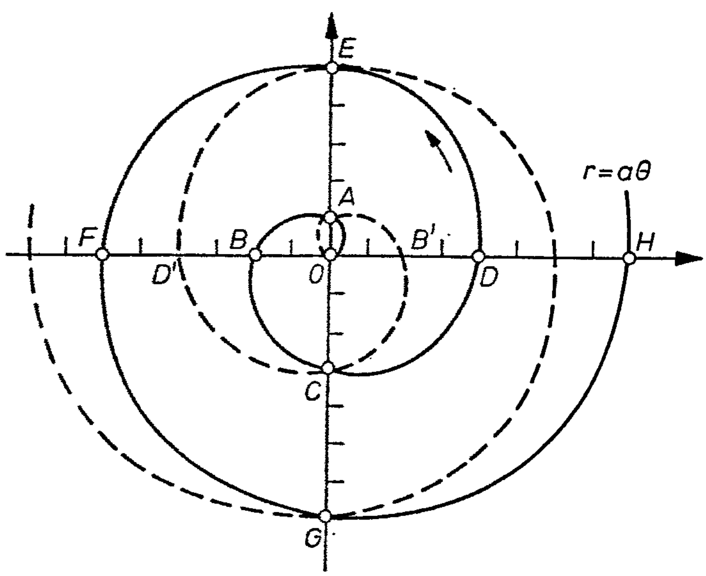

Abb. 122

Zur Konstruktion einer Reihe von Punkten $A, B, C, D, \ldots$ der Kurve tragen wir auf der Vertikalen die Strecke $\overline{OA} = a\frac{\pi}{2}$ ab, dann nehmen wir $\overline{OB} = 2\overline{OA}$, $\overline{OC} = 3\overline{OA}$, $\overline{OD} = 4\overline{OA}$ usw.; denn ihnen entsprechen die Winkel $2\cdot\frac{\pi}{2}$, $3\cdot\frac{\pi}{2}$, $4\cdot\frac{\pi}{2}$ usw. Ändert sich der Winkel θ von 0 bis ∞, so erhalten wir unendlich viele Windungen der Kurve $OABCD$, $DEFGH, \ldots$; der Abstand benachbarter Windungen, auf dem Strahl gerechnet, ist gleich $2\pi a$.

Man kann dem Winkel θ auch negative Werte von 0 bis $-\infty$ erteilen. Dann erhält man den anderen Kurvenzweig $OAB'CD' \ldots$, der gestrichelt gezeichnet ist. Er ist zum ersten symmetrisch.

Übrigens beschreibt die Gleichung $r = a\theta + b$ ebenfalls eine Archimedische Spirale. Dreht man die Polarachse um den Winkel $\alpha = -\frac{b}{a}$, so geht diese Gleichung wieder in $r = a\theta$ über.

2. Die *hyperbolische Spirale* $r = \frac{a}{\theta}$ (Abb. 123). Wächst der Winkel θ gegen ∞, so strebt der Radiusvektor gegen 0. Der Kurvenpunkt bewegt sich auf den Pol zu, ohne jedoch mit ihm zusammenzufallen. Hier ist der Pol ein sogenannter *asymptotischer Punkt* der Kurve. Die Kurve windet sich unendlich oft um den Pol.

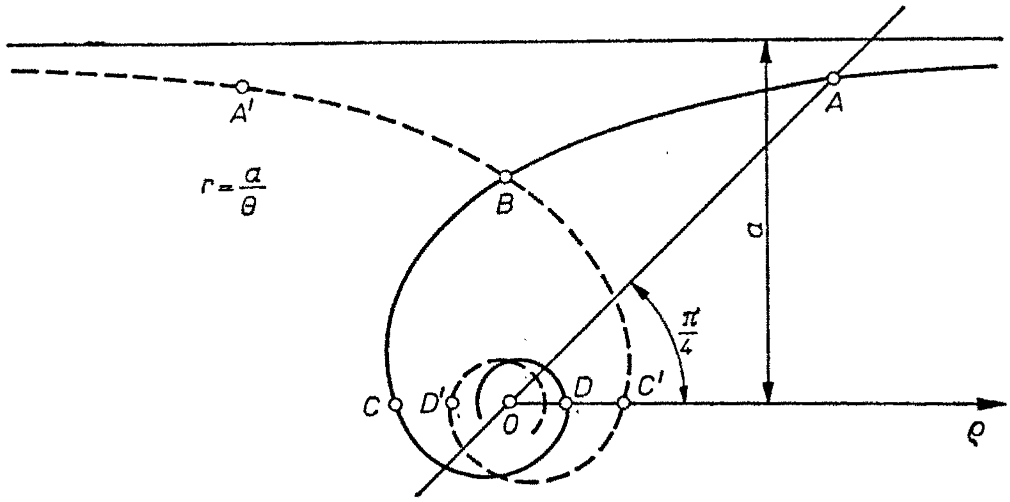

Abb. 123

Zieht man für $\theta = \frac{\pi}{4}$ die entsprechende Gerade, trägt man darauf die Strecke $\overline{OA} = \frac{4c}{\pi}$ ab und setzt dann $\overline{OB} = \frac{1}{2}\overline{OA}$, $\overline{OC} = \frac{1}{2}\overline{OB}$, $\overline{OD} = \frac{1}{2}\overline{OC}, \ldots$, so liegen die Punkte A, B, C, $D, \ldots$ offenbar auf der Kurve.

Auch hier kann man θ negative Werte erteilen. Variiert θ von 0 bis $-\infty$, so ergibt sich wie bei der Archimedischen Spirale der (auch hier getrichelt gezeichnete) zum ersten symmetrische Kurvenzweig $A'BC'D' \ldots$

Um einen Überblick über den Kurvenverlauf im Unendlichen zu erhalten, betrachten wir den vertikalen Abstand eines Kurvenpunktes von der Polarachse: $y = r \sin\theta = a\frac{\sin\theta}{\theta}$. Für $r \to \pm\infty$ oder, was dasselbe ist, für $\theta \to \pm 0$ gilt $\lim y = a$. Somit ist die Parallele zur Polarachse im Abstand a *Asymptote* der Kurve.

3. Die *logarithmische Spirale* $r = ae^{m\theta}$ (Abb. 124). Wächst (bzw. fällt) θ in arithmetischer Progression, so wächst (bzw. fällt) r in geometrischer Progression. Trägt man auf der Polar-

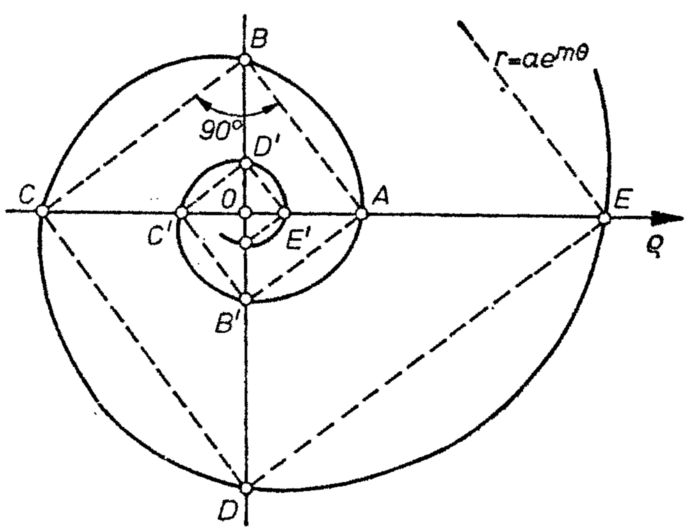

Abb. 124

achse die Strecke $\overline{OA} = a$ ab und auf der dazu senkrechten Geraden die Strecke $\overline{OB} = a e^{m\pi/2}$, so liegen A und B auf der Kurve; konstruiert man jetzt den Streckenzug $ABCDE\ldots$ so ergibt sich aus der Ähnlichkeit der Dreiecke, daß die Strecken $\overline{OA}, \overline{OB}, \overline{OC}, \overline{OD}, \overline{OE}, \ldots$ eine geometrische Folge mit dem Quotienten $e^{m\pi/2}$ bilden. Da die entsprechenden Winkel $0, \frac{\pi}{2}, \frac{2\pi}{2}, \frac{3\pi}{2}$, usw. sind, liegen offenbar alle Punkte $C, D, E, \ldots$ ebenfalls auf der Spirale.

Wächst θ von 0 bis ∞, so beschreibt der Punkt unendlich viele Windungen um den Pol, wobei er sich schnell ins Unendliche entfernt. Die Abstände zwischen den Windungen sind nicht mehr gleich groß. Der Winkel θ kann auch negativ werden. Strebt θ gegen $-\infty$, so strebt r gegen 0. Die Kurve windet sich unendlich oft um den Pol, nähert sich ihm unbegrenzt, ohne ihn zu erreichen (vgl. den Teil $AB'C'D'E'\ldots$ auf der Kurve; Abb. 124). Der Pol ist ein asymptotischer Punkt.

Durch Drehungen der Polarachse um den Pol kann man den Faktor a gleich 1 machen und die Kurvengleichung in ihrer einfachsten Gestalt $r = e^{m\theta}$ erhalten.

4. *Schnecken:* $r = a\cos\theta + b$ (Abb. 125). Man kann sich die Erzeugung dieser Kurven folgendermaßen vorstellen. Ein Kreis vom Durchmesser a sei gegeben. Wählt man den Pol O auf dem Kreis und zieht die Polarachse durch den Mittelpunkt C, so ist für jeden Punkt M des Kreises offenbar $r = a\cos\theta$. Das ist gerade die Polargleichung des Kreises. Ändert sich hier θ von 0 bis 2π, so beschreibt der veränderliche Punkt den Kreis zweimal (entgegen dem Uhrzeigersinn).

Verlängert man jetzt alle Radiusvektoren $\overline{OM'}$ des Kreises um die konstante Strecke $\overline{M'M} = b$ $(b > 0)$, so entsteht aus den so konstruierten Punkten M eine neue Kurve, die man all-

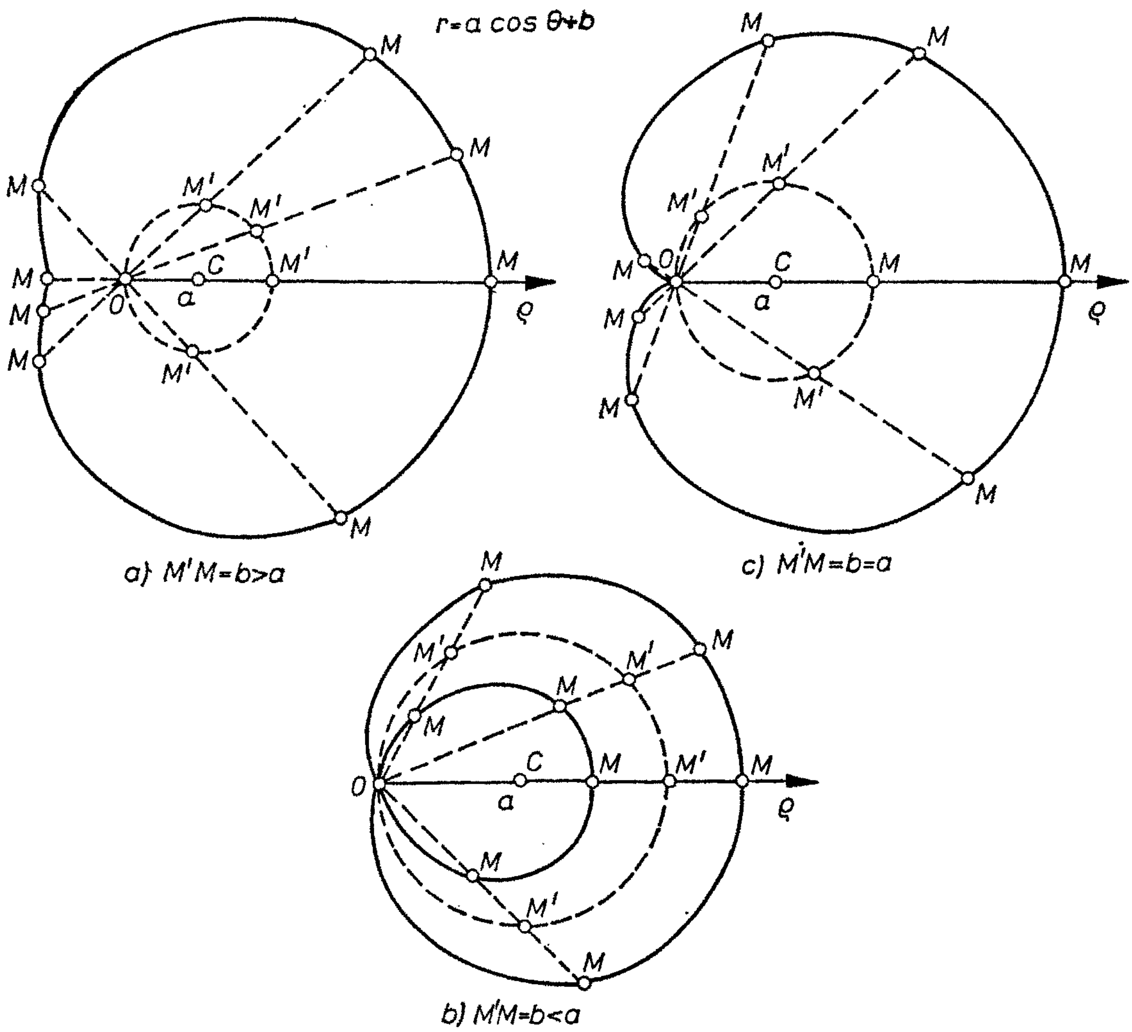

Abb. 125

gemein *Schnecke* nennt. Ihre Gleichung in Polarkoordinaten lautet offenbar

$$r = a \cos \theta + b.$$

Am einfachsten ist der Fall $b > a$; denn dann ist der Radiusvektor immer positiv, und die Kurve umkreist den Pol vollständig (Abb. 125a). Für $b < a$ geht die Kurve durch den Pol, überschneidet sich und bildet eine innere Schleife, wie in Abb. 125b. Um die Winkel θ zu bestimmen, für die der veränderliche Punkt den Pol erreicht, setzen wir in die Kurvengleichung $r = 0$; es folgt $\cos \theta = -\frac{b}{a}$. Diese Gleichung ist wegen $b < a$ lösbar.

Besonders interessant ist der dazwischenliegende Kurventyp, der dem Fall $b = a$ entspricht. Hier liegt der Pol auf der Kurve ($\theta = \pi$), aber es gibt keine Schleife; Abb. 125c zeigt die Kurve. Die Identität dieser Kurve mit der oben betrachteten *Kardioide*, einem Spezialfall einer Epizykloide (Abb. 120), ist offenbar. Wir überlassen es dem Leser, sich davon zu überzeugen.

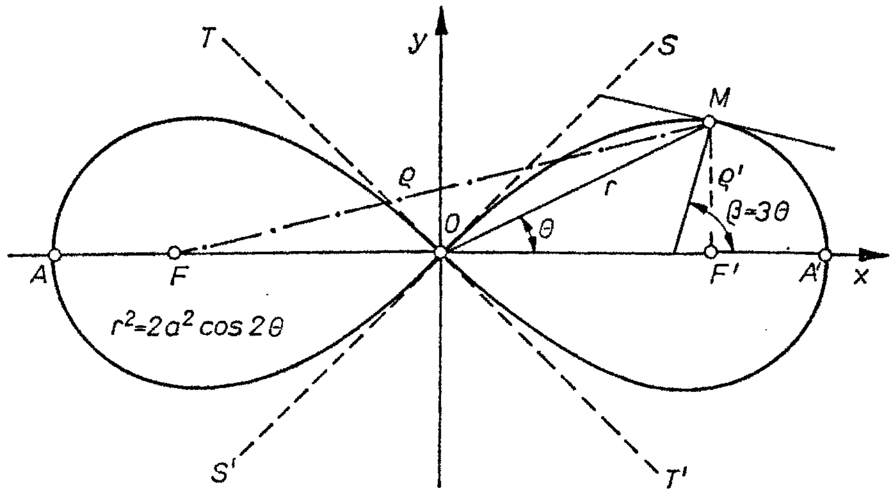

Abb. 126

5. Die *Bernoullische Lemniskate*: $r^2 = 2a^2 \cos 2\theta$ (Abb. 126). Diese Kurve kann man als geometrischen Ort der Punkte M definieren, für die das Produkt ihrer Abstände $\varrho = \overline{FM}$ und $\varrho' = \overline{F'M}$ von zwei gegebenen Punkten F und F', die voneinander die Entfernung $2a$ haben, den konstanten Wert a^2 hat.[1]

In den Bezeichnungen von Abb. 126 gilt für die Dreiecke OMF und OMF' nach dem Kosinussatz $\varrho^2 = r^2 + a^2 + 2ar \cos \theta$, $\varrho'^2 = r^2 + a^2 - 2ar \cos \theta$, also nach Definition

$$\varrho^2 \varrho'^2 = (r^2 + a^2)^2 - 4a^2 r^2 \cos^2 \theta = a^4;$$

hieraus ergibt sich nach elementaren Umformungen

$$r^2 = 2a^2 \cos 2\theta.$$

Das ist die *Gleichung der Lemniskate* in Polarkoordinaten. Da die linke (und damit die rechte) Seite dieser Gleichung keine negativen Werte annehmen kann, kann θ nur in solchen Intervallen variieren, in denen $\cos 2\theta \geqq 0$ ist. Das sind die Intervalle

$$\left(0, \frac{\pi}{4}\right), \quad \left(\frac{3\pi}{4}, \frac{5\pi}{4}\right), \quad \left(\frac{7\pi}{4}, 2\pi\right).$$

Die ganze Kurve liegt in den Winkelräumen zwischen den Geraden SS' und TT', die mit der Polarachse die Winkel $\frac{\pi}{4}$ bzw. $\frac{3\pi}{4}$ einschließen (vgl. Abb. 126). Die Kurve überschneidet sich

[1]) Bei der erwähnten Beziehung zwischen dem Abstand $\overline{FF'}$ und dem konstanten Wert des Produktes $\varrho\varrho'$ gehört offenbar der Mittelpunkt O der Strecke $\overline{FF'}$ der Kurve an ($\varrho = \varrho' = a$). Anders ist es, wenn $\varrho\varrho' = b^2$ ($b \neq a$) ist; dann erhält man die sogenannten *Cassinischen Ovale* (Jean Dominique Cassini, 1625—1712, französischer Astronom).

im Pol, dem die Winkel $\theta = \frac{\pi}{4}, \frac{3\pi}{4}, \frac{5\pi}{4}, \frac{7\pi}{4}$ entsprechen. Geht man in der üblichen Weise zu rechtwinkligen Koordinaten über, so erhält man die (implizite) Gleichung der Lemniskare in der Gestalt

$$(x^2 + y^2)^2 = 2a^2(x^2 - y^2).$$

227. Flächen und Kurven im Raum. Wir wollen hier nicht im einzelnen auf die Anwendung der Differentialrechnung auf die Geometrie im Raum eingehen; das überlassen wir der Differentialgeometrie. Daher beschränken wir uns auf solche räumlichen Probleme, die für die weiteren Teile dieses Werkes notwendig sind.

Wie bisher sollen alle betrachteten Funktionen nebst ihren Ableitungen stetig sein. Ferner sei ein rechtwinkliges Koordinatensystem gegeben.

Wir haben schon gelegentlich gesagt, daß eine Fläche im Raum durch eine Gleichung zwischen den laufenden Koordinaten dargestellt werden kann:

$$z = f(x, y) \tag{6}$$

(vgl. beispielsweise Nr. 160). Eine solche Gleichung heißt ebenso wie die analogen Gleichungen $x = g(y, z)$, $y = h(z, x)$ *explizite Flächengleichung.*

Auf diesen einfachsten Fall lassen sich in gewissem Sinne die anderen Darstellungsformen einer Fläche zurückführen.

Oft ist eine Fläche durch eine Gleichung der Gestalt

$$F(x, y, z) = 0 \tag{7}$$

gegeben, die nach keiner der Variablen aufgelöst ist (also durch eine *implizite* Flächengleichung). Ist in einem ihr genügenden Punkt (x_0, y_0, z_0) wenigstens eine der partiellen Ableitungen F_x, F_y, F_z von 0 verschieden, so ist in einer Umgebung dieses Punktes die Fläche auch durch eine explizite Gleichung darstellbar. Ist etwa $F_z(x_0, y_0, z_0) \neq 0$, so definiert die Gleichung nach Satz III aus Nr. 208 wenigstens in einer Umgebung des betrachteten Punktes die Veränderliche z als eindeutige Funktion von x und y, also $z = f(x, y)$, die überdies nebst ihren Ableitungen in beiden Argumenten stetig ist.

Eine Ausnahme kann also nur in einem singulären Punkt der Fläche auftreten, in dem zugleich die drei Beziehungen $F_x = 0$, $F_y = 0$, $F_z = 0$ gelten.

Die Gleichung

$$F(x, y) = 0, \tag{8}$$

in der eine der Koordinaten überhaupt nicht vorkommt, kann ebenfalls als Gleichung einer Fläche gedeutet werden. In der x, y-Ebene beschreibt sie eine Kurve. Konstruiert man über ihr als Leitkurve eine Zylinderfläche mit zur z-Achse parallelen Erzeugenden, so genügen alle Punkte dieser Fläche und nur diese Punkte der betrachteten Gleichung (da z darin nicht vorkommt und durch nichts eingeschränkt ist).

Analog lassen sich die Gleichungen $G(y, z) = 0$ und $H(z, x) = 0$ deuten.

Wir wenden uns jetzt *Kurven im Raum* zu. Die einfachste Art, eine Kurve im Raum vorzugeben, besteht darin, daß zwei laufende Koordinaten, etwa y und z, als Funktionen der dritten gegeben sind:

$$y = f(x), \quad z = g(x). \tag{9}$$

Das ist das natürliche Analogon einer expliziten Vorgabe einer Kurve in der Ebene. Auch hier könnte man von expliziten Kurvengleichungen sprechen.

Ebenso wie im Fall der Ebene lassen sich auch die übrigen analytischen Darstellungen einer Raumkurve auf die explizite Form bringen.

Jede der Gleichungen (9) kann gedeutet werden als Gleichung der Projektion unserer Kurve auf eine Koordinatenebene (die x, y- bzw. die x, z-Ebene) oder als Gleichung des *projizierenden Zylinders* [vgl. (8)] mit Erzeugenden, die der z-Achse bzw. der y-Achse parallel sind.

Allgemeiner kann man eine Raumkurve als Schnitt zweier Flächen erklären. Werden diese Flächen durch die Gleichungen

$$F(x, y, z) = 0 \quad \text{bzw.} \quad G(x, y, z) = 0 \tag{10}$$

beschrieben, so liefern diese beiden Gleichungen zusammen die analytische Darstellung ihrer Schnittkurve. Die Gleichungen (10) werden implizite Kurvengleichungen genannt.

Aus den partiellen Ableitungen von F und G bilde man die Matrix

$$\begin{pmatrix} F_x & F_y & F_z \\ G_x & G_y & G_z \end{pmatrix}, \tag{11}$$

Es sei eine ihrer Unterdeterminanten, etwa $\begin{vmatrix} F_y & F_z \\ G_y & G_z \end{vmatrix}$, in dem betrachteten Punkt von 0 verschieden. Dann können auf Grund von Satz IV aus Nr. 208 in der Umgebung dieses Punktes die Gleichungen (10) durch Gleichungen der Form (9) ersetzt werden (die darin vorkommenden Funktionen sind nebst ihren Ableitungen stetig).

Somit kann es nur in der Umgebung eines solchen Punktes der Kurve, in dem alle drei Unterdeterminanten der Matrix (11) verschwinden, vorkommen, daß sich die Kurvengleichungen nicht auf die einfachste Form bringen lassen. Wieder nennt man solche Punkte *singulär.*

228. Parameterdarstellung. Wir kommen nun zum Schluß zur Parameterdarstellung von Kurven und Flächen im Raum, wobei wir diesmal mit den Kurven beginnen.

Ebenso wie im Fall einer ebenen Kurve denken wir uns die Koordinaten des veränderlichen Punktes einer Raumkurve als Funktion einer Hilfsveränderlichen (eines *Parameters*) t,

$$x = \varphi(t), \quad y = \psi(t), \quad z = \chi(t), \tag{12}$$

so daß bei sich änderndem t der Punkt, dessen Koordinaten durch diese Gleichungen gegeben sind, die betrachtete Kurve durchläuft [im Fall der Gleichungen (9) spielte x die Rolle dieses Parameters].

Ist für einen Punkt der Kurve wenigstens eine der Ableitungen x_t, y_t, z_t von 0 verschieden, so kann man wie im Fall einer ebenen Kurve in der Umgebung dieses Punktes leicht von der Parameterdarstellung zu einer expliziten Darstellung übergehen. Nur in der Umgebung eines singulären Punktes, in dem alle diese Ableitungen verschwinden, kann es vorkommen, daß dieser Übergang nicht möglich ist.

Ebenso wie im Fall ebener Kurven müssen die sogenannten Mehrfachpunkte, d. h. die Punkte, die zwei oder mehr Parameterwerten entsprechen, zu den singulären Punkten gerechnet werden. Auch hier gilt das in der Fußnote auf S. 464 Gesagte.

Bei der Parameterdarstellung einer Fläche braucht man zur Bestimmung der Lage eines Punktes auf der Fläche offenbar zwei Parameter [in der expliziten Darstellung (6) sind das x und y]. Es seien die Gleichungen

$$x = \varphi(u, v), \quad y = \psi(u, v), \quad z = \chi(u, v) \tag{13}$$

gegeben, wobei (u, v) in einem abgeschlossenen Bereich Δ einer u, v-Ebene variiere. Wir bilden die Matrix

$$\begin{pmatrix} \varphi_u & \psi_u & \chi_u \\ \varphi_v & \psi_v & \chi_v \end{pmatrix} \tag{14}$$

und nehmen an, für $u = u_0$, $v = v_0$ sei wenigstens eine ihrer Unterdeterminanten von 0 verschieden, also etwa

$$\begin{vmatrix} \varphi_u & \psi_u \\ \varphi_v & \psi_v \end{vmatrix} \neq 0.$$

Schreiben wir die ersten beiden Gleichungen (13) in der Gestalt

$$\varphi(u, v) - x = 0, \quad \psi(u, v) - y = 0,$$

so können wir auf Grund des Satzes IV aus Nr. 208 behaupten, daß durch dieses System zweier Gleichungen mit den vier Veränderlichen u, v, x, y (wenigstens in der Umgebung der uns interessierenden Werte) die Veränderlichen u, v als eindeutige Funktionen von x, y definiert sind:

$$u = g(x, y), \quad v = h(x, y);$$

diese Funktionen sind überdies nebst ihren Ableitungen stetig. Setzen wir diese u, v in die dritte Gleichung (13) ein, so gelangen wir zu der üblichen Darstellung der Fläche durch die expliziten Gleichungen

$$z = \chi\big(g(x, y), h(x, y)\big) = f(x, y);$$

die Funktion f ist nebst ihren Ableitungen stetig.

Nur wenn alle drei Unterdeterminanten der Matrix (14) gleichzeitig verschwinden (ein entsprechender Punkt heißt singulär), kann es vorkommen, daß diese Darstellung nicht möglich ist.

Offenbar kann im Zusammenhang mit der Parameterdarstellung einer Fläche ebenfalls der Begriff des *einfachen* bzw. *mehrfachen Flächenpunktes* eingeführt werden: Einen einfachen Punkt erhält man nur für ein einziges Wertesystem (u, v), einen mehrfachen für mindestens zwei Wertesysteme.[1])

Wir erhalten nun in der Parameterdarstellung (13) einen der Parameter fest, setzen also etwa $u = u_0$. Dann erhalten wir offenbar die Gleichungen einer Kurve

$$x = \varphi(u_0, v), \quad y = \psi(u_0, v), \quad z = \chi(u_0, v),$$

deren Punkte sämtlich auf der Fläche liegen. Durch Variation von u_0 erhalten wir eine ganze Schar solcher (u)-Kurven. Setzen wir $v = v_0$, so erhalten wir auf der Fläche die Kurve

$$x = \varphi(u, v_0), \quad y = \psi(u, v_0), \quad z = \chi(u, v_0);$$

die (v)-Kurven bilden ebenfalls eine solche Kurvenschar.

Da man u und v als Koordinaten der Punkte der Fläche ansehen kann, nennt man diese Kurven *Koordinatenlinien der Fläche.* Durch einen einfachen Punkt der Fläche geht offenbar je eine Koordinatenlinie jeder Schar.

---

[1]) Bei einer *geschlossenen* Fläche (d. h. einer Fläche ohne Rand, etwa einer Kugel) können ihre Punkte den Punkten eines Bereichs Δ der u, v-Ebene nicht eineindeutig zugeordnet werden. In diesem Fall sind bei jeder Parameterdarstellung Mehrfachpunkte unvermeidlich.

Vergleicht man die verschiedenen Arten der analytischen Vorgabe von Flächen [vgl. (6), (7) und (13)] und von Raumkurven [(9), (10) und (12)], so läßt sich das am Schluß von Nr. 223 Gesagte wiederholen. In der Umgebung eines nichtsingulären (einfachen) Punktes läßt sich alles auf den anschaulichen Fall der expliziten Darstellung zurückführen.

229. Beispiele.

1. Die *Vivianische Kurve*[1]). So nennt man die Schnittkurve einer Kugelfläche und eines geraden Kreiszylinders über dem Kugelradius als Durchmesser (Abb. 127). Der Kugelradius sei R; liegen die Achsen wie in Abb. 127, so lauten die Gleichungen der beiden Flächen offenbar

$$x^2 + y^2 + z^2 = R^2, \qquad x^2 + y^2 = Rx.$$

Beide zusammen liefern die Gleichung unserer Kurve.

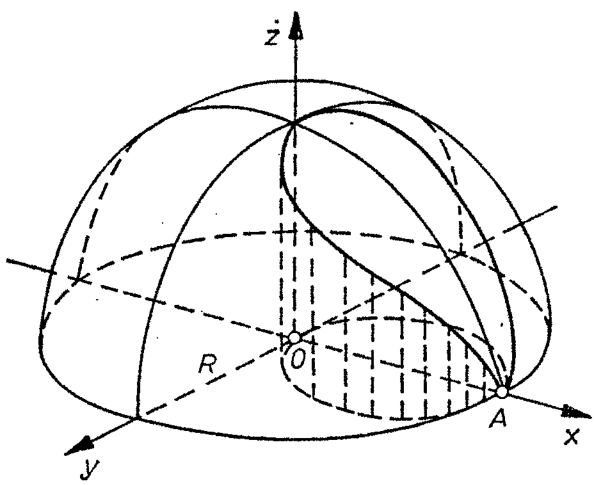

Abb. 127

Die Kurve hat die Gestalt einer gebogenen Acht; sie überschneidet sich im Punkt $(R, 0, 0)$. Dieser Punkt ist also sicher singulär. Tatsächlich hat die Matrix

$$\begin{pmatrix} 2x & 2y & 2z \\ 2x - R & 2y & 0 \end{pmatrix}$$

die Unterdeterminanten

$$\begin{vmatrix} 2y & 2z \\ 2y & 0 \end{vmatrix} = -4yz, \quad \begin{vmatrix} 2z & 2x \\ 0 & 2x - R \end{vmatrix} = 4xz - 2Rz, \quad \begin{vmatrix} 2x & 2y \\ 2x - R & 2y \end{vmatrix} = 2Ry,$$

die in diesem Punkt offenbar sämtlich verschwinden.

Die Parameterdarstellung der Vivianischen Kurve lautet etwa

$$x = R\sin^2 t, \quad y = R\sin t\cos t, \quad z = R\cos t.$$

Das läßt sich leicht verifizieren: Diese Ausdrücke genügen den impliziten Kurvengleichungen für alle t; variiert t von 0 bis 2π, so erhält man die ganze Kurve. Der Punkt $(R, 0, 0)$ ergibt sich für $t = \frac{\pi}{2}$ und für $t = \frac{3\pi}{2}$. Wie zu erwarten war, ist es ein Doppelpunkt.

2. Es gibt Fälle, in denen die Parameterdarstellung ganz zwangsläufig aus der Erzeugungsweise der Kurve folgt. Als Beispiel betrachten wir die Schraubenlinie. Ihre Entstehung kann man sich folgendermaßen vorstellen. Ein ursprünglich in A befindlicher Punkt M (Abb. 128) drehe sich mit gleichbleibender Geschwindigkeit um die z-Achse (etwa im Uhrzeigersinn) und bewege sich gleichzeitig ebenfalls mit gleichförmiger Geschwindigkeit parallel zu dieser (in positiver Richtung). Die Bahn von M heißt *Schraubenlinie*. Als Parameter, der die Lage von M

[1]) Vincenzo Viviani, 1622—1703, italienischer Mathematiker.

bestimmt, kann man den Winkel t nehmen, den die Projektion $\overline{OP}$ der Strecke $\overline{OM}$ mit der x-Achse bildet. Die Koordinaten x und y von M sind dieselben wie von P, also $x = a \cos t$, $y = a \times \sin t$, wobei a der Radius des von P beschriebenen Kreises ist. Die vertikale Veränderliche z wächst proportional mit t (da sowohl die Translation als auch die Drehung gleichförmig erfolgen), also $z = ct$. Somit lautet die Parameterdarstellung der Schraubenlinie

$$x = a \cos t, \qquad y = a \sin t, \qquad z = ct. \tag{15}$$

Hier haben wir eine *Linksschraube* erhalten. Bei einem Rechtssystem wäre dieselbe Kurve als Rechtsschraube zu bezeichnen.

Aus (15) kann man t leicht eliminieren und eine explizite Kurvengleichung erhalten:

$$x = a \cos \frac{z}{c}, \qquad y = a \sin \frac{z}{c}.$$

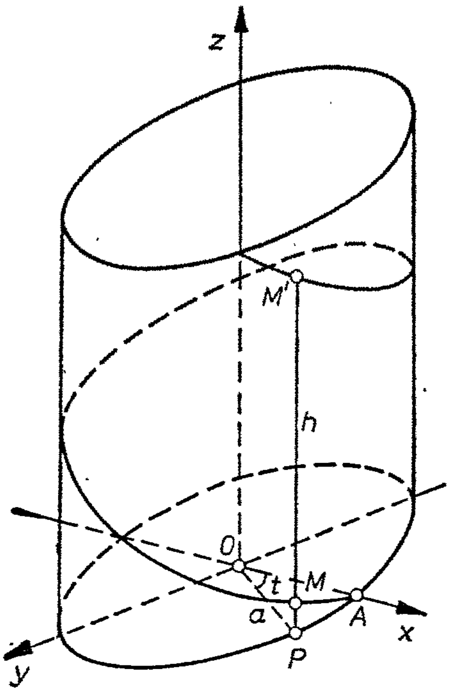

Abb. 128

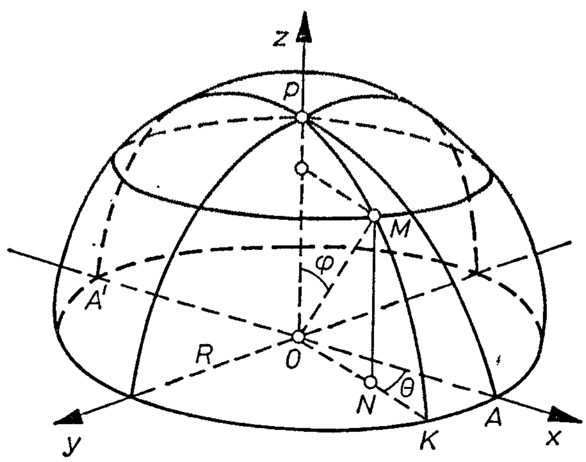

Abb. 129

3. Wir betrachten die *Kugelfläche* vom Radius R um den Ursprung (Abb. 129); ihre implizite Gleichung lautet bekanntlich

$$x^2 + y^2 + z^2 = R^2.$$

Um eine Parameterdarstellung zu erhalten, zeichnen wir die „Äquatorialebene" AKA' und durch die Pole P, P' und den Punkt M den Meridian PMP'. Die Lage von M auf der Kugel wird bestimmt etwa durch die Winkel $\varphi = \sphericalangle\, POM$ und $\theta = \sphericalangle\, AOK$. Ferner ist $z = \overline{NM} = R \cos \varphi$; wegen $\overline{ON} = R \sin \varphi$ ergeben sich für x und y (für M und N dieselben): $x = \overline{ON} \times \cos \theta$, $y = \overline{ON} \sin \theta$. Somit finden wir die Parameterdarstellung der Kugelfläche in der Gestalt

$$x = R \sin \varphi \cos \theta, \qquad y = R \sin \varphi \sin \theta, \qquad z = R \cos \varphi,$$

wobei φ von 0 bis π und θ von 0 bis 2π variiert.

Die Zuordnung zwischen den Punkten der Sphäre und denen des Rechtecks $[0, \pi; 0, 2\pi]$ der φ, θ-Ebene ist jedoch nicht eindeutig[1]: Die Werte $\theta = 0$ und $\theta = 2\pi$ führen zu demselben Punkt der Fläche, und außerdem erhält man für $\varphi = 0$ (bzw. $\varphi = \pi$) bei jedem Wert von θ immer nur einen einzigen Punkt, den Pol P (bzw. P').

[1]) Vgl. die Fußnote auf S. 477.

Ersetzt man φ durch $\lambda = \frac{\pi}{2} - \varphi$ und läßt λ von $-\frac{\pi}{2}$ bis $\frac{\pi}{2}$ variieren und θ zwischen $-\pi$ und π, so erhält man die *geographischen Koordinaten* (*Breite* und *Länge*).

Die Unterdeterminanten $R^2 \sin^2\varphi \cos\theta$, $R^2 \sin^2\varphi \sin\theta$, $R^2 \sin\varphi \cos\varphi$ der Matrix

$$\begin{pmatrix} R\cos\varphi\cos\theta & R\cos\varphi\sin\theta & -R\sin\varphi \\ -R\sin\varphi\sin\theta & R\sin\varphi\cos\theta & 0 \end{pmatrix}$$

verschwinden sämtlich für $\varphi = 0$ und $\varphi = \pi$. Offenbar sind jedoch die beiden Pole nur singulär in bezug auf diese spezielle analytische Darstellung der Sphäre.

Man sieht leicht, daß die *Meridiane* (θ = const) und die *Parallelkreise* (φ = const) je eine Schar von Koordinatenlinien auf der Kugel bilden.

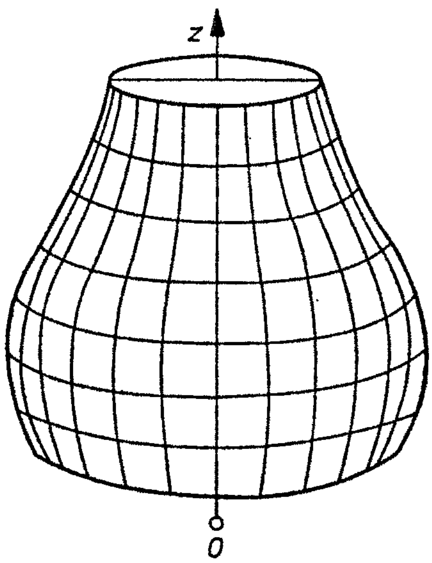

Abb. 130

4. Das vorstehende Beispiel läßt sich folgendermaßen verallgemeinern. In der x, z-Ebene sei eine Erzeugende durch die Parametergleichungen

$$x = \varphi(u), \qquad z = \psi(u) \tag{16}$$

gegeben; es sei $\varphi(u) \geqq 0$. Wir wollen sie wie einen starren Körper um die z-Achse rotieren lassen (Abb. 130). Bezeichnet v den Drehwinkel, so kann man die Gleichungen der so entstehenden Rotationsfläche in der Gestalt

$$x = \varphi(u)\cos v, \qquad y = \varphi(u)\sin n, \qquad z = \psi(u) \qquad (0 \leqq v \leqq 2\pi)$$

schreiben. Nimmt man in der x, z-Ebene den Halbkreis $x = R\sin u$, $z = R\cos u$ und dreht ihn um die z-Achse, so erhält man gerade die Parameterdarstellung der so entstehenden Kugelfläche in der obigen Form (in etwas anderen Bezeichnungen).

Der Leser möge sich vergewissern, daß als singuläre Punkte der Rotationsflächen nur Punkte der Drehachse in Frage kommen oder aber solche Punkte, die bei der Drehung aus singulären Punkten der Erzeugenden entstehen.

Als Koordinatenlinien können auch hier die verschiedenen Lagen der Erzeugenden (Meridiane) und die Parallelkreise dienen.

5. Fügt man zu der Drehbewegung der Kurve (16) noch eine zur Drehachse parallele Translation hinzu, so erhält man (unter der Annahme, daß beide Bewegungen gleichförmig verlaufen) die *allgemeine Schraubenfläche*

$$x = \varphi(u)\cos v, \qquad y = \varphi(u)\sin v, \qquad z = \psi(u) + cv.$$

Nimmt man insbesondere als Erzeugende die positive x-Achse, $x = u$, $z = 0$ ($u \geqq 0$), und führt eine Schraubung aus, so erhält man die *gewöhnliche Schraubenfläche*

$$x = u\cos v, \qquad y = u\sin v, \qquad z = cv.$$

Bei der allgemeinen Schraubenfläche besteht die eine Schar der Koordinatenlinien aus den verschiedenen Lagen der Erzeugenden ($v = \text{const}$) und die andere aus den Schraubenlinien ($u = \text{const}$).

§ 2. Tangente und Tangentialebene

230. Die Tangente an eine ebene Kurve in rechtwinkligen Koordinaten. Der Begriff der *Tangente* ist uns schon einige Male begegnet (vgl. etwa Nr. 91). Eine durch eine explizite Gleichung

$$y = f(x)$$

gegebene Kurve, wobei f eine nebst ihrer Ableitung stetige Funktion ist, hat in jedem Punkt (x, y) eine Tangente, für deren Richtungskoeffizienten die Beziehung $\tan\alpha = y_x = f'(x)$ gilt. Die Gleichung der Tangente lautet also

$$Y - y = y_x(X - x). \tag{1}$$

Hier (wie auch im folgenden) bedeuten X, Y die laufenden Koordinaten und x, y die Koordinaten des Berührungspunktes.

Auch die Gleichung der *Normalen*, d. h. der durch den Berührungspunkt gehenden Senkrechten zur Tangente, ergibt sich leicht:

$$Y - y = -\frac{1}{y_x}(X - x) \quad \text{oder} \quad X - x + y_x(Y - y) = 0. \tag{2}$$

Im Zusammenhang mit Tangente und Normale betrachtet man auch einige Strekken, nämlich $\overline{TM}$ und $\overline{MN}$ und ihre Projektionen $\overline{TP}$ bzw. $\overline{PN}$ auf die x-Achse

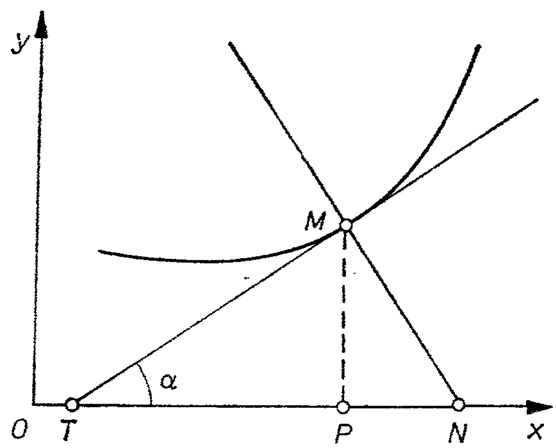

Abb. 131

(Abb. 131). Letztere werden *Subtangente* bzw. *Subnormale* genannt und mit *sbt* bzw. *sbn* bezeichnet. Setzt man in (1) und (2) $Y = 0$, so ergibt sich leicht

$$sbt = \overline{TP} = \left|\frac{y}{y_x}\right|, \quad sbn = \overline{PN} = |y y_x|. \tag{3}$$

Dann kann man aus den Dreiecken MPT und MPN auch die Längen des Tangenten- und des Normalenabschnitts bestimmen:

$$t = \overline{TM} = \left|\frac{y}{y_x}\sqrt{1 + y_x^2}\right|, \quad n = \overline{MN} = \left|y\sqrt{1 + y_x^2}\right|. \tag{4}$$

Ist die Kurve implizit gegeben,

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

so kann man sie sich in der Umgebung eines nichtsingulären Punktes $M(x, y)$ durch eine explizite Gleichung beschrieben vorstellen. Ist in M etwa $F_y(x, y) \neq 0$, so kann man die Kurve in der Form $y = f(x)$ darstellen, wobei f nebst seiner Ableitung stetig ist. Hieraus folgt, daß die Kurve in M eine Tangente hat, deren Gleichung in der Gestalt (1) geschrieben werden kann. Nun wissen wir aber, daß [vgl. Nr. 209, Formel (15)] dann

$$y_x = -\frac{F_x(x, y)}{F_y(x, y)}$$

ist. Setzt man ein, so folgt nach einfachen Umformungen die in x und y symmetrische Tangentengleichung

$$F_x(x, y)\,(X - x) + F_y(x, y)\,(Y - y) = 0. \tag{5}$$

Zu demselben Resultat gelangt man, wenn zwar im Punkt M die Ableitung F_y verschwindet, aber $F_x \neq 0$ ist. Nur in einem singulären Punkt wird diese Gleichung bedeutungslos, und ohne zusätzliche Untersuchungen (vgl. Nr. 236) kann man dann über die Tangente nichts aussagen.

Für die Gleichung der Normalen ergibt sich in unserem Fall offenbar

$$F_y(x, y)\,(X - x) - F_x(x, y)\,(Y - y) = 0.$$

Ist schließlich die Kurve in der Parameterdarstellung

$$x = \varphi(t), \quad y = \psi(t)$$

gegeben, so existiert bekanntlich im Fall $\varphi'(t) \neq 0$ die Tangente an die Kurve, und für den Richtungskoeffizienten gilt

$$\tan \alpha = \frac{y_t}{x_t} \tag{6}$$

[vgl. Nr. 106, Formel (11)]. Dann erhält man die Tangentengleichung in der Gestalt

$$Y - y = \frac{y_t}{x_t}(X - x) \quad \text{oder} \quad \frac{X - x}{x_t} = \frac{Y - y}{y_t}.$$

In dieser Form gilt die Gleichung auch dann, wenn zwar $x_t = 0$, aber $y_t \neq 0$ ist.[1] Nur in einem singulären Punkt, wo $x_t = 0$ und $y_t = 0$ ist, wird die Gleichung bedeutungslos, und das Problem, ob eine Tangente existiert, bleibt offen (vgl. Nr. 237).

Manchmal multipliziert man zweckmäßigerweise beide Nenner mit dt und schreibt die Tangentengleichung in der Gestalt

$$\frac{X - x}{dx} = \frac{Y - y}{dy}. \tag{7}$$

---

[1]) Wie man in der analytischen Geometrie oft vereinbart, bedeutet das folgendes: Ist in der Proportion

$$\frac{X - x}{a} = \frac{Y - y}{b}$$

eines der Hinterglieder 0, so soll auch einfach das entsprechende Vorderglied gleich 0 sein.

231. Beispiele.

1. Die *Parabel* $y^2 = 2px$. Sieht man y als Funktion von x an, so ergibt sich durch Differentiation $yy_x = p$. Somit [vgl. (3)] ist die Subnormale der Parabel eine Konstante. Hieraus folgt eine einfache Konstruktion für die Normale (und damit für die Tangente) an die Parabel.

Nach (4) gilt für den Normalenabschnitt

$$n = \sqrt{y^2 + p^2}.$$

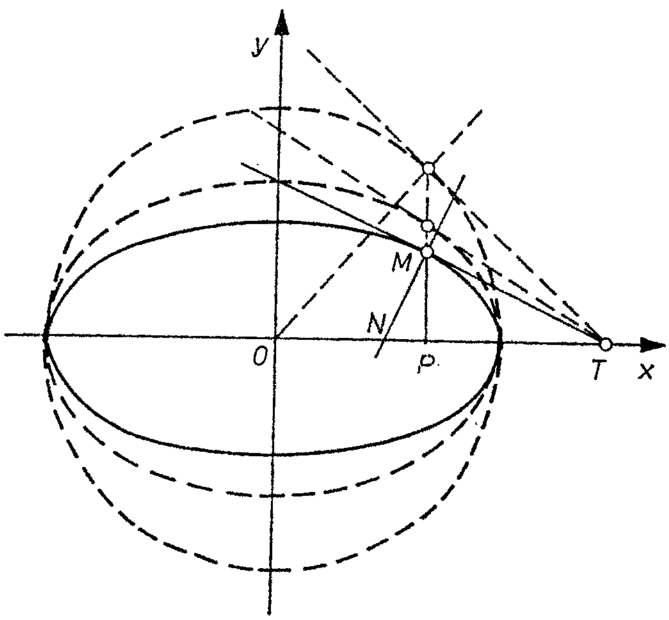

Abb. 132

2. Die *Ellipse* $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ (Abb. 132). Nach (5) lautet die Tangentengleichung

$$\frac{x}{a^2}(X - x) + \frac{y}{b^2}(Y - y) = 0.$$

Man kann sie in der einfacheren Gestalt

$$\frac{xX}{a^2} + \frac{yY}{b^2} = 1$$

schreiben, wenn man die Ellipsengleichung heranzieht. Für $Y = 0$ folgt $X = \frac{a^2}{x}$. Somit hängt der Schnittpunkt T der Tangente mit der x-Achse weder von y noch von b ab. Daher gehen die Tangenten an die Ellipsen, welche verschiedenen Werten von b entsprechen, in Punkten mit der Abszisse x sämtlich durch ein und denselben Punkt T der x-Achse. Da sich für $b = a$ ein Kreis ergibt, an den man die Tangente leicht konstruieren kann, läßt sich T leicht bestimmen, und damit hat man das aus Abb. 132 ersichtliche einfache Konstruktionsverfahren für die Ellipsentangente.[1])

Die Länge des Normalenabschnittes ergibt sich zu

$$n = \frac{\sqrt{b^4x^2 + a^4y^2}}{a^2}.$$

Denselben Ausdruck erhält man auch im Fall der *Hyperbel* $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$.

[1]) Diese Eigenschaft der Ellipsentangente hängt damit zusammen, daß die Ellipse als orthogonale Projektion eines Kreises vom Radius a angesehen werden kann, der in einer zur Ellipsenebene geneigten Ebene liegt.

3. Die *Astroide* $x^{2/3} + y^{2/3} = a^{2/3}$ (Abb. 116, S. 465). Die Tangentengleichung

$$x^{-1/3}(X - x) + y^{-1/3}(Y - y) = 0$$

kann mit Hilfe der Kurvengleichung auf die Gestalt

$$\frac{X}{x^{1/3}} + \frac{Y}{y^{1/3}} = a^{2/3} \quad \text{oder} \quad \frac{X}{a^{2/3}x^{1/3}} + \frac{Y}{a^{2/3}y^{1/3}} = 1$$

gebracht werden. Dies ist eine „Achsenabschnittsgleichung". Daher schneidet die Tangente auf den Achsen die Abschnitte $a^{2/3}x^{1/3}$ und $a^{2/3}y^{1/3}$ ab. Hieraus folgt eine interessante Eigenschaft der Astroide. Bezeichnet man die Länge des Tangentenabschnittes zwischen den Achsen mit τ, so gilt

$$\tau^2 = a^{4/3}x^{2/3} + a^{4/3}y^{2/3} = a^2, \quad \text{also} \quad \tau = a = \text{const}.$$

Somit schneiden die Symmetrieachsen der Astroide auf allen Tangenten gleiche Stücke heraus.

4. Die *Zykloide* $x = a(t - \sin t)$, $y = a(1 - \cos t)$ (Abb. 118, S. 467). Wir hatten in Nr. 225, Beispiel 6, schon die Beziehung $y_x = \cot\frac{t}{2}$, d. h.

$$\tan\alpha = \cot\frac{t}{2} = \tan\left(\frac{\pi}{2} - \frac{t}{2}\right),$$

gefunden und können $\alpha = \frac{\pi}{2} - \frac{t}{2}$ nehmen.

Wir erinnern uns (Abb. 118) der Beziehung $t = \sphericalangle MDN$, also $\sphericalangle MEN = \frac{t}{2}$. Verlängert man $\overline{EM}$ bis zum Schnittpunkt T mit der x-Achse, so folgt $\sphericalangle ETx = \frac{\pi}{2} - \frac{t}{2} = \alpha$. Daher ist die Verbindungsgerade zwischen einem Zykloidenpunkt und dem höchsten Punkt des rollenden Kreises (in entsprechender Lage) gerade die Tangente. Also ist $\overline{MN}$ die Normale.

Im folgenden wird uns ein Ausdruck für die Strecke n der Normalen von Nutzen sein, den man leicht aus dem rechtwinkligen Dreieck MEN erhält, nämlich

$$n = \overline{MN} = 2a\sin\frac{t}{2}.$$

5. Die *Epizykloide*

$$x = a[(1 + m)\cos mt - m\cos(1 + m)t], \quad y = a[(1 + m)\sin mt - m\sin(1 + m)t]$$

(Abb. 119, S. 468). Schreiben wir die Ableitungen in der Gestalt

$$x_t = 2am(1 + m)\sin\frac{t}{2}\cos\left(m + \frac{1}{2}\right)t, \quad y_t = 2am(1 + m)\sin\frac{t}{2}\sin\left(m + \frac{1}{2}\right)t,$$

so gelangen wir zu

$$\tan\alpha = \frac{y_t}{x_t} = \tan\left(m + \frac{1}{2}\right)t, \quad \text{also} \quad \alpha = \left(m + \frac{1}{2}\right)t.$$

Verbindet man (Abb. 119) den Punkt D mit M, so bildet diese Gerade mit der x-Achse den Winkel xTD; für ihn gilt

$$\sphericalangle xTD = \sphericalangle DOT + \sphericalangle ODT = mt + \frac{t}{2},$$

Daher ist $\overline{DT}$ die Tangene in M und $\overline{MB}$ die Normale.

6. Die *Kreisevolvente* $x = a(t\sin t + \cos t)$, $y = a(\sin t - t\cos t)$ (Abb. 121, S. 470). Hier ist

$$\tan\alpha = \frac{y_t}{x_t} = \tan t, \quad \text{also} \quad \alpha = t.$$

Somit ist die Tangente $\overline{MT}$ dem Radius $\overline{OB}$ parallel, und $\overline{BM}$ ist die Normale der Kurve.

Bemerkung. Die Resultate der Beispiele 4 und 6 hätte man auch ohne weitere Herleitungen, von kinematischen Erwägungen ausgehend, erhalten können. Beim Abrollen einer Kurve längs einer anderen ist der Berührungspunkt jeweils momentanes Drehzentrum für die sich bewegende Figur, so daß die Normale der Bahnkurve jedes ihrer Punkte durch diesen Berührungspunkt geht.

232. Die Tangente in Polarkoordinaten. Ist eine Kurve durch eine Gleichung der Form $r = f(\theta)$ gegeben, so erhält man durch den üblichen Übergang zu rechtwinkligen Koordinaten die Parameterdarstellung

$$x = r\cos\theta = f(\theta)\cos\theta, \quad y = r\sin\theta = f(\theta)\sin\theta,$$

wobei θ der Parameter ist. Nach der allgemeinen Formel (6) ist dann

$$\tan\alpha = \frac{y_\theta}{x_\theta} = \frac{r_\theta\sin\theta + r\cos\theta}{r_\theta\cos\theta - r\sin\theta}.$$

Untersucht man eine Kurve in Polarkoordinaten, so bestimmt man jedoch die Lage der Tangente nicht durch den Winkel α mit der Polarachse, sondern mit Hilfe des Winkels ω, den sie mit dem verlängerten Radiusvektor bildet (Abb. 114, S. 447, und Abb. 133). Wir hatten schon die einfache Beziehung

$$\tan\omega = \frac{r}{r_\theta} \tag{8}$$

gefunden (vgl. Nr. 218, Beispiel 4).

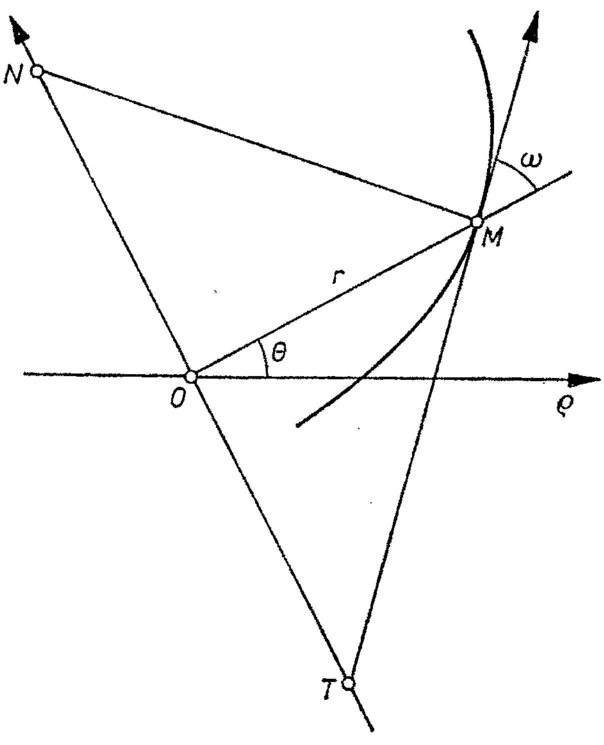

Abb. 133

Ebenso betrachtet man an Stelle der in Nr. 230 behandelten Strecken t, n, sbt, sbn hier andere. Wir ziehen durch den Pol O eine zum Radiusvektor senkrechte Gerade (die sich dreht, wenn der Punkt die Kurve durchläuft) und verlängern Tangente und Normale bis zu den Schnittpunkten T bzw. N mit ihr. Die Strecken $\overline{TM}$ bzw. $\overline{MN}$ heißen *Polarabschnitte von Tangente* bzw. *Normale*, ihre Projektionen $\overline{TO}$ bzw. $\overline{ON}$

auf die Gerade *polare Subtangente* bzw. *polare Subnormale*. Wir notieren das durch den Index p. Aus (8) erhält man leicht

$$sbt_p = \overline{TO} = r \tan \omega = \frac{r^2}{r_\theta}, \quad sbn_p = \overline{ON} = r \cot \omega = r_\theta$$

und hieraus

$$t_p = \overline{TM} = \left| \frac{r}{r_\theta} \sqrt{r^2 + r_\theta^2} \right|, \quad n_p = \overline{MN} = \sqrt{r^2 + r_\theta^2}.$$

233. Beispiele.

1. Die *Archimedische Spirale* $r = a\theta$ (Abb. 122, S. 471). Hier ist $r_\theta = a$, also $sbn_p = a = \text{const.}$ Dadurch erhält man leicht die Lage des Punktes N und damit Normale und Tangente.

Wegen $\tan \omega = \theta$ gilt $\tan \omega \to \infty$ für $\theta \to \infty$; somit strebt ω gegen einen rechten Winkel.

2. Die *hyperbolische Spirale* $r = \frac{a}{\theta}$ (Abb. 123, S. 472). Diesmal ist $r_\theta = -\frac{\alpha}{\theta^2}$, $sbt_p = |-a|$ $= \text{const}$, so daß sich die Tangente ebenfalls leicht konstruieren läßt.

3. Die *logarithmische Spirale* $r = ae^{m\theta}$ (Abb. 134). Es ist $r_\theta = ma\, e^{m\theta}$, also $\tan \omega = \frac{1}{m}$ $= \text{const}$, somit $\omega = \text{const}$. Demnach hat die logarithmische Spirale die bemerkenswerte Eigenschaft, daß der Winkel zwischen Radiusvektor und Tangente konstant bleibt. Anders ausgedrückt: *Die logarithmische Spirale schneidet alle ihre Radiusvektoren unter demselben Winkel.* Durch diese Eigenschaft erinnert sie an den Kreis, bei dem das ebenfalls der Fall ist. Dort ist dieser Winkel ein rechter. Übrigens ergibt sich der Kreis als Spezialfall der logarithmischen Spirale, nämlich für $m = 0$.

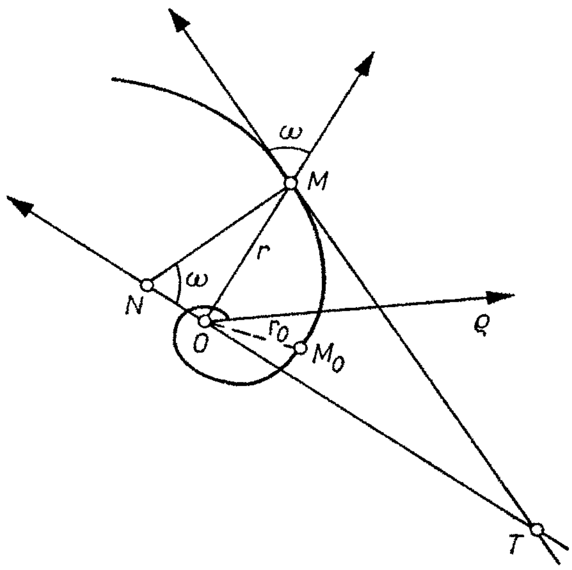

Abb. 134

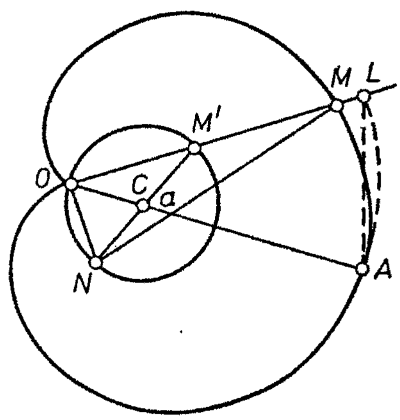

Abb. 135

4. Die *Schnecken* $r = a \cos \theta + b$ (Abb. 135). Hier hängt $sbn_p = |r_\theta| = |-a \sin \theta|$ nicht von b ab. Somit ist für die auf einem Strahl liegenden Punkte verschiedener Schnecken (die verschiedenen Werten von b entsprechen) die polare Subnormale dieselbe, d. h. der Punkt N derselbe. Für $b = 0$ ergibt sich der Kreis, bei dem die Konstruktion der Normalen trivial ist. Dann kann man aber auch für die übrigen Schnecken leicht die Normalen zeichnen. Aus $\triangle\, MON$ erhält man für die polare Normale

$$n_p = \sqrt{a^2 + 2ab \cos \theta + b^2}.$$

Eine besonders einfache Gestalt hat diese Normale für die Kardioide ($b = a$)

$$n_p = 2a \left|\cos \frac{\theta}{2}\right|;$$

diesen Fall zeigt Abb. 135.

5. Die *Lemniskate* $r^2 = 2a^2 \cos 2\theta$ (Abb. 126, S. 474). Durch Differentiation nach θ ergibt sich

$$r r_\theta = -2a^2 \sin 2\theta .$$

Nach (8) folgt durch Division dieser Gleichungen

$$\tan \omega = \frac{r}{r_\theta} = -\cot 2\theta, \quad \text{also} \quad \omega = 2\theta + \frac{\pi}{2}.$$

Sind α und β die Richtungswinkel von Tangente und Normale, so ergibt sich

$$\beta = \alpha - \frac{\pi}{2}, \quad \alpha = \omega + \theta = 3\theta + \frac{\pi}{2},$$

also $\beta = 3\theta$. *Der Neigungswinkel der Normalen der Lemniskate ist gleich dem dreifachen Polarwinkel des Berührungspunktes.* Das liefert ein einfaches Verfahren zur Normalenkonstruktion, also auch zur Tangentenkonstruktion.

234. Die Tangente an eine Raumkurve. Die Tangentialebene an eine Fläche.

1°. Bei einer Raumkurve bleibt die Tangentendefinition wörtlich dieselbe wie bei einer ebenen Kurve (vgl. Nr. 91). Wir beschränken uns auf den Fall, daß die Kurve in der Parameterdarstellung

$$x = \varphi(t), \quad y = \psi(t), \quad z = \chi(t)$$

gegeben ist.

Wir wählen einen bestimmten Wert von t und damit einen bestimmten Punkt M (x, y, z) auf der Kurve; dieser Punkt sei nicht singulär und einfach (vgl. Nr. 223). Nun erteilen wir t den Zuwachs Δt; dem neuen Parameterwert $t + \Delta t$ entspricht dann der Punkt $M_1(x + \Delta x, y + \Delta y, z + \Delta z)$. Die Gleichung der Sekante lautet

$$\frac{X - x}{\Delta x} = \frac{Y - y}{\Delta y} = \frac{Z - z}{\Delta z},$$

wobei X, Y, Z die laufenden Koordinaten sind. Die geometrische Bedeutung dieser Gleichung ändert sich nicht, wenn wir alle Nenner durch Δt dividieren:

$$\frac{X - x}{\dfrac{\Delta x}{\Delta t}} = \frac{Y - y}{\dfrac{\Delta y}{\Delta t}} = \frac{Z - z}{\dfrac{\Delta z}{\Delta t}}.$$

Wenn diese Gleichung beim Grenzübergang $\Delta t \to 0$ sinnvoll bleibt, wird dadurch die Existenz der Grenzlage der Sekante, d. h. der Tangente, gesichert. In der Grenze erhalten wir

$$\frac{X - x}{x_t} = \frac{Y - y}{y_t} = \frac{Z - z}{z_t}; \tag{9}$$

das ist tatsächlich die Gleichung einer Geraden, da nicht alle Nenner verschwinden. Somit existiert in jedem nichtsingulären Punkt die Tangente und wird durch diese Gleichung beschrieben. Für einen singulären Punkt bleibt die Frage offen.

Bemerkung. Wir gingen in den Gleichungen zur Grenze für $\Delta t \to 0$ über; nun zeigen wir, daß dies der Annahme $\overline{MM_1} \to 0$ gleichwertig ist. Da die Funktionen φ, ψ, χ stetig sind, folgt aus $\Delta t \to 0$ auch

$$\overline{MM_1} = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2} \to 0.$$

Um auch das Umgekehrte zu beweisen, geben wir $\varepsilon > 0$ vor. Da $\overline{MM_1}$ eine stetige Funktion von Δt ist, hat für $|\Delta t| \geqq \varepsilon$ diese Funktion einen kleinsten Wert δ, der offenbar positiv ist (da der betrachtete Punkt einfach sein sollte, d. h. sich für keinen von t verschiedenen Parameterwert $\overline{MM_1} = 0$ ergeben sollte). Dann ist aber

$$\text{für} \quad \overline{MM_1} < \delta \quad \text{notwendigerweise} \quad |\Delta t| < \varepsilon,$$

womit alles bewiesen ist.

Manchmal schreibt man (9) in der Gestalt

$$\frac{X - x}{dx} = \frac{Y - y}{dy} = \frac{Z - z}{dz},$$

die sich aus (9) durch Multiplikation aller Nenner mit dt ergibt.

Sind α, β, γ die Winkel zwischen der Tangente und den Koordinatenachsen, so gilt für die Richtungskosinus

$$\cos\alpha = \frac{x_t}{\pm\sqrt{x_t^2 + y_t^2 + z_t^2}}, \qquad \cos\beta = \frac{y_t}{\pm\sqrt{x_t^2 + y_t^2 + z_t^2}},$$
$$\cos\gamma = \frac{z_t}{\pm\sqrt{x_t^2 + y_t^2 + z_t^2}}.$$

Die Wahl eines bestimmten Vorzeichens der Wurzel entspricht der Wahl einer bestimmten Richtung der Tangente.

Das Problem der Tangente an eine durch implizite Gleichungen $F(x, y, z) = 0$ und $G(x, y, z) = 0$ gegebene Kurve untersuchen wir in 3°.

2°. Es sei eine Fläche durch die explizite Gleichung $z = f(x, y)$ gegeben. In Nr. 180 haben wir den Begriff der *Tangentialebene* definiert und unter der Annahme, daß $f(x, y)$ differenzierbar ist[1]), die Gleichung dieser Tangentialebene gefunden [vgl. Nr. 180, Formel (6)]:

$$Z - z = f_x(x, y)\,(X - x) + f_y(x, y)\,(Y - y).$$

Gewöhnlich setzt man

$$\frac{\partial z}{\partial x} = f_x(x, y) = p, \qquad \frac{\partial z}{\partial y} + f_y(x, y) = q$$

und schreibt die Gleichung der Tangentialebene in der Gestalt

$$Z - z = p(X - x) + q(Y - y). \tag{10}$$

[1]) Wir setzen hier Existenz und Stetigkeit der partiellen Ableitungen, also vollständige Differenzierbarkeit, voraus (vgl. Nr. 179).

Sind $\cos\lambda$, $\cos\mu$, $\cos\nu$ die Richtungskosinus der Flächennormalen (d. h. der im Berührungspunkt auf der Tangentialebene errichteten Senkrechten), so gilt

$$\begin{aligned} \cos\lambda &= \frac{-p}{\pm\sqrt{1+p^2+q^2}}, \qquad \cos\mu = \frac{-q}{\pm\sqrt{1+p^2+q^2}}, \\ \cos\nu &= \frac{1}{\pm\sqrt{1-p^2+q^2}}; \end{aligned} \tag{11}$$

die verschiedenen Vorzeichen der Wurzel entsprechen den beiden entgegengesetzten Normalenrichtungen.

Wir legen nun auf der Fläche durch den betrachteten Punkt eine beliebige Kurve $x = \varphi(t)$, $y = \psi(t)$, $z = \chi(t)$, so daß identisch in t

$$\chi(t) = f\big(\varphi(t), \psi(t)\big)$$

gilt. Durch Differenzieren nach t (vgl. Nr. 181) ergibt sich

$$\chi'(t) = p\varphi'(t) + q\psi'(t). \tag{*}$$

Wir betrachten die Tangente an die Kurve in einem nichtsingulären Punkt in der Gestalt (9). Ersetzen wir in (*) die Ableitungen φ', ψ', χ' durch die nach (9) dazu proportionalen Differenzen $X - x$, $Y - y$, $Z - z$, so gelangen wir zu (10). Somit liegt die Tangente (9) mit allen Punkten in der Tangentialebene (10). *Wir können daher die Tangentialebene einer Fläche in einem ihrer Punkte als diejenige Ebene definieren, in der alle Tangenten an alle Kurven liegen, die auf der Fläche durch diesen Punkt gehen.*[1])

Ist die Fläche implizit durch $F(x, y, z) = 0$ gegeben, so kann man unter der Annahme, daß F_z im betrachteten Punkt von 0 verschieden ist, in seiner Umgebung die Fläche explizit durch $z = f(x, y)$ beschreiben, so daß die Existenz der Tangentialebene gewährleistet ist. Da in diesem Fall

$$p = \frac{\partial z}{\partial x} = -\frac{F_x}{F_z}, \qquad q = \frac{\partial z}{\partial y} = -\frac{F_y}{F_z}$$

gilt, erhält man durch Einsetzen dieser Werte von p und q in (10) leicht

$$F_x(x, y, z)\,(X - x) + F_y(x, y, z)\,(Y - y) + F_z(x, y, z)\,(Z - z) = 0. \tag{12}$$

Offenbar läßt sich die Tangentialebene auch dann durch diese Gleichung beschreiben, wenn zwar $F_z = 0$, aber eine der beiden anderen Ableitungen F_x, F_y von 0 verschieden ist. Nur im Fall eines singulären Punktes ist die Gleichung bedeutungslos (und das Problem, ob eine Tangentialebene existiert, bleibt offen).

3°. Jetzt überlegt man sich leicht, wie man die Tangente an eine Kurve findet, die in der impliziten Form

$$F(x, y, z) = 0, \quad G(x, y, z) = 0,$$

d. h. als Schnitt zweier Flächen gegeben ist.

Ist der betrachete Kurvenpunkt nicht singulär, so kann in seiner Umgebung die Kurve auch durch explizite Gleichungen beschrieben werden (vgl. Nr. 227), so daß

---

[1]) Zum Teil haben wir das schon in Nr. 180 erwähnt.

die Existenz der Tangente gewährleistet ist. Diese Tangente bildet die Schnittgerade der Tangentialebenen an die beiden Flächen, ist also durch die Gleichungen

$$\left.\begin{aligned} F_x(X-x)+F_y(Y-y)+F_z(Z-z)&=0,\\ G_x(X-x)+G_y(Y-y)+G_z(Z-z)&=0 \end{aligned}\right\} \tag{13}$$

zu beschreiben. Da in einem nichtsingulären Punkt wenigstens eine der Unterdeterminanten der Koeffizientenmatrix von 0 verschieden ist, wird durch diese Gleichungen tatsächlich eine Gerade bestimmt.

4°. Nun untersuchen wir den Fall, daß eine Fläche in der Parameterdarstellung

$$x=\varphi(u,v),\quad y=\psi(u,v),\quad z=\chi(u,v)$$

gegeben ist.

Wieder beschränken wir uns auf einen nichtsingulären einfachen Punkt. Da in seiner Umgebung nach Nr. 228 die Fläche auch durch eine explizite Gleichung beschrieben werden kann, existiert die Tangentialebene sicher. Ihre Gleichung kann in der Gestalt

$$A(X-x)+B(Y-y)+C(Z-z)=0 \tag{14}$$

geschrieben werden, deren Koeffizienten A, B, C noch zu bestimmen sind.

Setzt man in die Flächengleichungen für v den Wert ein, der einem bestimmten Punkt entspricht, so erhält man die Gleichungen der Koordinatenlinie [der (v)-Kurve], die durch diesen Punkt geht. Die Tangente an diese Kurve in diesem Punkt lautet [vgl. (9)]

$$\frac{X-x}{x_u}=\frac{Y-y}{y_u}=\frac{Z-z}{z_u}.$$

Analog erhält man für festes u die durch den gegebenen Punkt gehende Koordinatenlinie der anderen Schar [(u)-Kurve], deren Tangente die Gleichung

$$\frac{X-x}{x_v}=\frac{Y-y}{y_v}=\frac{Z-z}{z_v}$$

besitzt.

Da diese beiden Tangenten in der Tangentialebene (14) liegen müssen, hat man die Bedingungen

$$Ax_u+By_u+Cz_u=0,\quad Ax_v+By_v+Cz_v=0$$

zu erfüllen. Somit müssen die Koeffizienten A, B, C den Unterdeterminanten der Matrix

$$\begin{pmatrix} x_u & y_u & z_u\\ x_v & y_v & z_v \end{pmatrix}$$

proportional sein.

Gewöhnlich setzt man sie diesen Unterdeterminanten gleich:

$$A=\begin{vmatrix} y_u & z_u\\ y_v & z_v \end{vmatrix},\quad B=\begin{vmatrix} z_u & x_u\\ z_v & x_v \end{vmatrix},\quad C=\begin{vmatrix} x_u & y_u\\ x_v & y_v \end{vmatrix}. \tag{15}$$

Am einfachsten läßt sich jetzt die Gleichung der Tangentialebene als Determinanten-gleichung

$$\begin{vmatrix} X-x & Y-y & Z-z \\ x_u & y_u & z_u \\ x_v & y_v & z_v \end{vmatrix} = 0 \tag{16}$$

schreiben. In einem nichtsingulären Punkt wird dadurch tatsächlich eine Ebene dargestellt.

Für die Richtungskosinus der Normalen hat man

$$\begin{aligned} \cos\lambda &= \frac{A}{\pm\sqrt{A^2+B^2+C^2}}, \qquad \cos\mu = \frac{B}{\pm\sqrt{A^2+B^2+C^2}}, \\ \cos\nu &= \frac{C}{\pm\sqrt{A^2+B^2+C^2}}. \end{aligned} \tag{17}$$

235. Beispiele.

1. Wir betrachten die *Schraubenlinie* (Abb. 128, S. 479) $x = a\cos t$, $y = a\sin t$, $z = ct$. Hier ist

$$x_t = -a\sin t, \qquad y_t = a\cos t, \qquad z_t = c,$$

und die Tangentengleichung lautet

$$\frac{X-x}{-a\sin t} = \frac{Y-y}{a\cos t} = \frac{Z-z}{c}.$$

Die Richtungskosinus der Tangente sind

$$\cos\alpha = -\frac{a\sin t}{\sqrt{a^2+c^2}}, \qquad \cos\beta = \frac{a\cos t}{\sqrt{a^2+c^2}}, \qquad \cos\gamma = \frac{c}{\sqrt{a^2+c^2}}.$$

Offenbar ist $\cos\gamma = \text{const}$, also $\gamma = \text{const}$. Denkt man sich die Schraubenlinie auf einen geraden Kreiszylinder aufgewickelt, so kann man sagen, daß die Schraubenlinie alle Erzeugenden dieses Zylinders unter konstantem Winkel schneidet.[1])

2. Das *Ellipsoid* $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$. Die Gleichung der Tangentialebene ergibt sich aus (12) unter Berücksichtigung der Gleichung des Ellipsoids zu

$$\frac{xX}{a^2} + \frac{yY}{b^2} + \frac{zZ}{c^2} = 1.$$

3. Der *Kegel* (*zweiter Ordnung*) $\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0$. Die Tangentialebene hat die Gleichung

$$\frac{xX}{a^2} + \frac{yY}{b^2} - \frac{zZ}{c^2} = 0.$$

Die Spitze (0, 0, 0) dieses Kegels ist ein singulärer Punkt; dort wird diese Gleichung bedeutungslos, und es gibt keine Tangentialebene in diesem Punkt.

4. Die *Vivianische Kurve* (Abb. 127, S. 478) $x^2 + y^2 + z^2 = R^2$, $x^2 + y^2 = Rx$. Wir erhalten für die Tangente die Gleichungen [vgl. [13)]

$$xX + yY + zZ = R^2, \qquad (2x - R)X + 2yY = Rx.$$

Nur in dem singulären Punkt $(R, 0, 0)$ liefern sie keine Gerade mehr.

---

[1]) Schneidet man den Zylindermantel längs einer Erzeugenden auf und wickelt ihn ab, so geht die Schraubenlinie in eine Gerade über, die natürlich alle Vertikalen unter demselben Winkel schneidet. Diese Überlegung läßt das obige Resultat trivial erscheinen.

5. Die *Schraubenfläche* $x = u \cos v$, $y = \sin v$, $z = cv$. Nach (16) lautet die Gleichung der Tangentialebene

$$\begin{vmatrix} X - x & Y - y & Z - z \\ \cos v & \sin v & 0 \\ -u \sin v & u \cos v & c \end{vmatrix} = 0.$$

Unter Berücksichtigung der Flächengleichung kann man sie auf die Gestalt

$$\sin v \cdot X - \cos v \cdot Y + \frac{u}{c} \cdot Z = uv$$

bringen.

236. Singuläre Punkte ebener Kurven. Wir gehen hier auf das Verhalten einer durch die implizite Gleichung $F(x, y) = 0$ gegebenen Kurve in der Nähe eines *singulären* Punktes (x_0, y_0) näher ein. Wir streben keine Vollständigkeit an, sondern wollen den Leser nur mit den hauptsächlichen Typen singulärer Punkte vertraut machen. Dabei setzen wir F als stetig und zweimal stetig differenzierbar voraus. Ohne Beschränkung der Allgemeinheit sei $x_0 = y_0 = 0$. Sonst hätte man nur den Ursprung in den singulären Punkt zu verschieben. Es sei also

$$F(0, 0) = 0, \quad F_x(0, 0) = 0, \quad F_y(0, 0) = 0.$$

Ferner setzen wir $a_{11} = F_{xx}(0, 0)$, $a_{12} = F_{xy}(0, 0)$, $a_{22} = F_{yy}(0, 0)$. Von den Zahlen a_{11}, a_{12}, a_{22} sei mindestens eine von 0 verschieden. Wir klassifizieren die möglichen Fälle nach dem Vorzeichen von $a_{11}a_{22} - a_{12}^2$. Die folgenden Überlegungen hängen aufs engste mit denen aus Nr. 197 zusammen.

1. $a_{11}a_{22} - a_{12}^2 > 0$.

Dann hat bekanntlich $F(x, y)$ in $(0, 0)$ ein Extremum. In einer hinreichend kleinen (punktierten) Umgebung dieses Punktes ist also $F > 0$ oder $F < 0$. Mit anderen Worten: In dieser Umgebung gibt es außer dem Ursprung überhaupt keinen Punkt der Kurve; der Ursprung ist ein *isolierter Punkt* der Kurve.

Beispiele. $x^2 + y^2 = 0$ bzw. $(x^2 + y^2)(x + y - 1) = 0$. Der Ursprung gehört beiden Kurven an und ist für beide ein isolierter Punkt. Während aber die erste nur aus diesem einzigen Punkt besteht, umfaßt die zweite noch die Gerade $x + y = 1$, die nicht durch den Ursprung geht.

2. $a_{11}a_{22} - a_{12}^2 < 0$.

Wie in Nr. 197 kann man $F(x, y)$ in der Umgebung des Ursprungs in der Gestalt

$$F(x, y) = \frac{1}{2}\{a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + \alpha_{11}x^2 + 2\alpha_{12}xy + \alpha_{22}y^2\}$$

darstellen, wobei für $x \to 0$, $y \to 0$ alle α gegen 0 streben. In Polarkoordinaten ϱ, φ gilt

$$F(x, y) = \frac{\varrho^2}{2}\{a_{11}\cos^2\varphi + 2a_{12}\cos\varphi\sin\varphi + a_{22}\sin^2\varphi$$
$$+ \alpha_{11}\cos^2\varphi + 2\alpha_{12}\cos\varphi\sin\varphi + \alpha_{22}\sin^2\varphi\}.$$

Setzen wir noch $a_{22} \neq 0$ voraus, so hat das Trinom $a_{11} + 2a_{12}t + a_{22}t^2$ zwei verschiedene reelle Nullstellen t_1 und t_2 (es sei $t_1 < t_2$) und läßt sich in der Gestalt $a_{22}(t - t_1)(t - t_2)$ darstellen. Wir setzen $\varphi_1 = \arctan t_1$, $\varphi_2 = \arctan t_2$, also $t_1 = \tan \varphi_1$,

$t_2 = \tan \varphi_2$. Jetzt kann man das erste Trinom in der geschweiften Klammer auf die Gestalt

$$\begin{aligned} &a_{11} \cos^2 \varphi + 2a_{12} \cos \varphi \sin \varphi + a_{22} \sin^2 \varphi \\ &= a_{22} \cos^2 \varphi (\tan \varphi - \tan \varphi_1)(\tan \varphi - \tan \varphi_2) \end{aligned} \tag{18}$$

bringen. Hieraus ersieht man, daß die unter den Winkeln φ_1 und φ_2 zur x-Achse durch den Ursprung gehenden Geraden [der Kürze halber nennen wir sie (φ_1) und (φ_2)] die Ebene in zwei Bereiche teilen, in deren einem das Trinom positiv, in deren anderem es negativ ist (Abb. 136).[1]

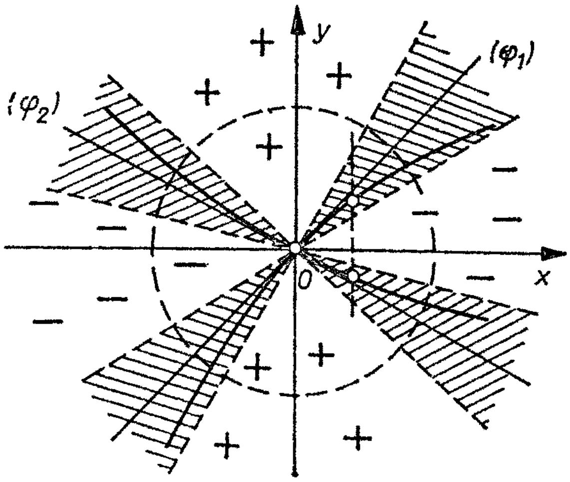

Abb. 136

Wir schließen jetzt die Geraden (φ_1) und (φ_2) in zwei beliebig schmale Winkelbereiche ein, in zwei Paar Scheitelwinkel, die zwischen den Geraden $(\varphi_1 - \varepsilon)$ und $(\varphi_1 + \varepsilon)$ bzw. zwischen $(\varphi_2 - \varepsilon)$ und $(\varphi_2 + \varepsilon)$ enthalten sind (sie sind in Abb. 136 schraffiert). Nimmt man einen Kreis von hinreichend kleinem Radius r_ε um den Ursprung, so kann man behaupten, daß er auf Grund der Wahl der Winkelbereiche in zwei Winkelbereiche zerfällt, von denen jeder so beschaffen ist, daß in ihm die Funktion $F(x, y)$ ein bestimmtes Vorzeichen beibehält, in dem einen das positive, in dem anderen das negative (vgl. Abb. 136). Da nämlich bei einer Änderung des Winkels außerhalb der Intervalle $(\varphi_1 - \varepsilon, \varphi_1 + \varepsilon)$ und $(\varphi_2 - \varepsilon, \varphi_2 + \varepsilon)$ das Trinom (18) nicht verschwindet, bleibt sein absoluter Betrag größer als eine positive Zahl m_ε. Andererseits wird für hinreichend kleines ϱ der absolute Betrag des Ausdrucks

$$\alpha_{11} \cos^2 \varphi + 2\alpha_{12} \cos \varphi \sin \varphi + \alpha_{22} \sin^2 \varphi$$

kleiner als m_ε. Hieraus folgt unsere Behauptung (vgl. die Überlegung in Nr. 197, Fall 1).

Wir betrachten nun zwei schraffierte Scheitelsektoren des Kreises, beispielsweise die durch die Geraden $(\varphi_1 - \varepsilon)$ und $(\varphi_1 + \varepsilon)$ begrenzten. Da auf diesen Geraden die Funktion entgegengesetzte Vorzeichen hat, gibt es auf jeder Vertikalen, die diese Sektoren schneidet, einen Punkt, in dem $F(x, y)$ verschwindet, d. h. einen Punkt unserer Kurve. Das folgt aus der bekannten Eigenschaft stetiger Funktionen (Nr. 80), wenn man sie auf die bei festem x nur von y abhängende Funktion $F(x, y)$ anwendet.[2]

---

[1]) Hier vertiefen wir das in Nr. 197, Fall 2, Gesagte. Dort genügte es uns, die Existenz zweier Geraden zu beweisen, auf denen das Trinom verschiedenes Vorzeichen hat.

[2]) Vgl. den Beweis von Satz I aus Nr. 206 über die Existenz einer impliziten Funktion.

Somit liegt innerhalb jedes Paares der schraffierten Sektoren ein Zweig der Kurve, der durch den Ursprung geht, während es innerhalb des Kreises außerhalb dieser Sektoren keine Kurvenpunkte gibt. Da ε beliebig klein sein kann, ergibt sich hieraus und aus den Stetigkeitsvoraussetzungen, daß im Ursprung diese Zweige die Geraden (φ_1) bzw. (φ_2) berühren.

Zwar blieb noch die Frage offen, ob dieser Punkt der einzige auf dieser Vertikalen ist, in dem $F(x, y) = 0$ ist. Gäbe es zwei, so würde nach dem Satz von ROLLE (Nr. 111) zwischen ihnen auf dieser Vertikalen ein Punkt existieren, in dem $F_y(x, y) = 0$ wäre. Somit ist die Einzigkeit bewiesen, wenn wir noch zeigen, daß dies wenigstens in einer hinreichend kleinen Umgebung des Ursprungs nicht der Fall sein kann.

Diesen Beweis führen wir indirekt. Es sei also $F_y(x_n, y_n) = 0$ für eine Folge von Punkten (x_n, y_n), wobei $x_n \to 0$ und $\dfrac{y_n}{x_n} \to \tan\varphi_1 = t_1$ gilt. Auf $F_y(x, y)$ wenden wir den Mittelwertsatz [Nr. 183, Formel (10)] an:

$$\begin{aligned} 0 &= F_y(x_n, y_n) - F_y(0, 0) \\ &= F_{xy}(\theta_n, x_n, \theta_n y_n) \cdot x_n + F_{yy}(\theta_n x_n, \theta_n y_n) \cdot y_n \quad (0 < \theta_n < 1) \end{aligned}$$

oder

$$F_{xy}(\theta_n x_n, \theta_n y_n) + F_{yy}(\theta_n x_n, \theta_n y_n) \cdot \frac{y_n}{x_n} = 0.$$

Gehen wir hier zur Grenze über, so erhalten wir $a_{12} + a_{22} t_1 = 0$, also $t_1 = -\dfrac{a_{12}}{a_{22}}$, was nicht der Fall ist: Ein solcher Wert von t_1 könnte nur dann existieren, wenn die Nullstellen des Trinoms $a_{11} + 2a_{12}t + a_{22}t^2$ einander gleich wären.

Hieraus folgt beiläufig, daß in einer hinreichend kleinen Umgebung des Ursprungs kein Punkt der beiden Zweige außer dem Ursprung selbst singulär ist.

Analog läßt sich der Fall behandeln, daß $a_{22} = 0$, aber $a_{11} \neq 0$, oder $a_{11} = a_{22} = 0$, aber $a_{12} \neq 0$ ist; im letzten Fall sind die Geraden (φ_1) und (φ_2) die Koordinatenachsen.

Unter der Voraussetzung $a_{11}a_{22} - a_{12}^2 < 0$ ist also der Punkt $(0, 0)$ ein *Doppelpunkt* der Kurve: In ihm schneiden sich zwei Zweige der Kurve, von denen jeder in diesem Punkt eine Tangente hat. Die Richtungskoeffizienten dieser Tangenten erhält man immer aus der Gleichung $a_{11} + 2a_{12}t + a_{22}t^2 = 0$. Nur muß man bei $a_{22} = 0$ annehmen, daß sie neben einer endlichen Wurzel eine unendliche Wurzel hat.

Als Beispiel mögen folgende uns schon bekannten Kurven dienen:

$$(x^2 + y^2)^2 + 2a^2(y^2 - x^2) = 0 \quad \text{(Lemniskate; Abb. 126, S. 474).}$$

$$x^3 + y^3 - 3axy = 0 \quad \text{(Cartesisches Blatt; Abb. 117, S. 466),}$$

für die der Ursprung ein Doppelpunkt ist. Im ersten Fall gilt $a_{11} = -4a^2$, $a_{12} = 0$, $a_{22} = 4a^2$, $t_1 = 1$, $t_2 = -1$, so daß die Winkelhalbierenden dort Tangenten sind. Im zweiten ist $a_{11} = a_{22} = 0$, $a_{12} = -3a$, $t_1 = 0$, $t_2 = \infty$; die Koordinatenachsen sind Tangenten.

3. $a_{11}a_{22} - a_{12}^2 = 0$.

Wir setzen auch hier $a_{22} \neq 0$ voraus. Das quadratische Trinom $a_{11} + 2a_{12}t + a_{22}t^2$ hat in diesem Fall die doppelte Nullstelle $t_1 = -\dfrac{a_{12}}{a_{22}}$. Wir setzen wieder, wie oben, $\varphi_1 = \arctan t_1$ und ziehen durch den Ursprung eine Gerade unter dem Winkel φ_1 zur x-Achse. Wir schließen sie in einen Winkelbereich zwischen den Geraden $(\varphi_1 - \varepsilon)$ und

$(\varphi_1 + \varepsilon)$ ein (in Abb. 137 schraffiert). Mit Hilfe ähnlicher Überlegungen wie oben kann man zeigen, daß außerhalb des schraffierten Bereichs, in hinreichender Nähe des Ursprungs, die Funktion $F(x, y)$ ein bestimmtes Vorzeichen beibehält, und zwar dasselbe auf beiden Seiten: das positive bzw. das negative, je nachdem, ob $a_{22} > 0$ oder $a_{22} < 0$ ist. Jetzt hat die Funktion auf den Geraden $(\varphi_1 \pm \varepsilon)$ dasselbe Vorzeichen, und der Satz von CAUCHY ist nicht anwendbar.

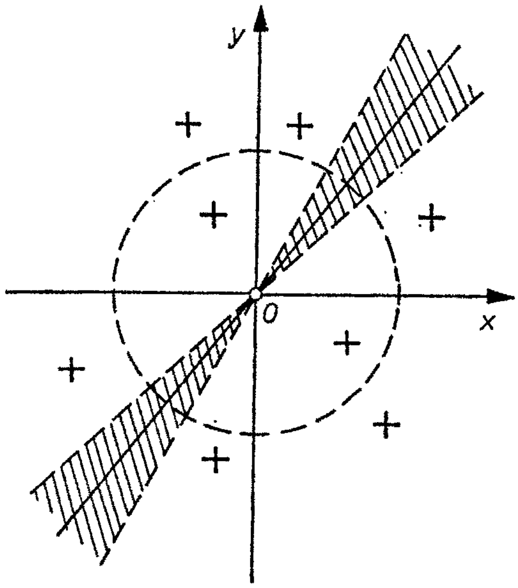

Abb. 137

Wir wollen auf diesen Fall, der tiefergehende Überlegungen erfordert und bei dem höhere Ableitungen herangezogen werden müssen, nicht näher eingehen und beschränken uns auf die Zusammenstellung der wichtigsten Fälle.

a) In der Nachbarschaft des Ursprungs gibt es außer diesem keine Kurvenpunkte, es ist ein *isolierter Punkt* (wie im Fall 1° in Nr. 234).

Beispiele. $x^4 + y^2 = 0$ oder $(x^4 + y^2)(x + y - 1) = 0$. Für beide „Kurven“ ist der Ursprung ein isolierter Punkt.

b) In den beiden schraffierten Winkelbereichen (in hinreichender Nähe des Ursprungs) liegen auf jeder Vertikalen zwei Punkte unserer Kurve; durch den Ursprung gehen zwei Kurvenzweige, die dort die gemeinsame Tangente (φ_1) haben: Der Ursprung ist ein *Doppelpunkt* (wie im Fall 2° in Nr. 234).

Beispiel. $x^4 - y^2 = 0$, d. h. $y = \pm x^2$, zwei Parabeln, die im Ursprung die x-Achse berühren.

c) In einem der schraffierten Bereiche gibt es keine Kurvenpunkte, im anderen zwei Zweige, die sozusagen im Ursprung enden und dort die gemeinsame Tangente (φ_1) haben. Hier haben wir einen neuen Typus eines singulären Punktes vor uns, einen *Rückkehrpunkt* oder eine *Spitze*. Je nachdem, ob die beiden Zweige auf verschiedenen Seiten der gemeinsamen Tangente liegen oder nicht, unterscheidet man Rückkehrpunkte *erster* und *zweiter Art*.

Als Beispiel einer Kurve mit einem Rückkehrpunkt erster Art kann die semikubische Parabel

$$y^2 - x^3 = 0$$

(Abb. 115, S. 465) dienen.

Für den selteneren Fall eines Rückkehrpunktes zweiter Art nennen wir die Kurve

$$x^5 - (y - x^2)^2 = 0 \quad \text{oder} \quad y = x^2 \pm x^2\sqrt{x} \quad (x \geqq 0).$$

Beide Kurvenzweige berühren im Ursprung die x-Achse, liegen aber (wenigstens in hinreichender Nähe) oberhalb derselben (Abb. 138).

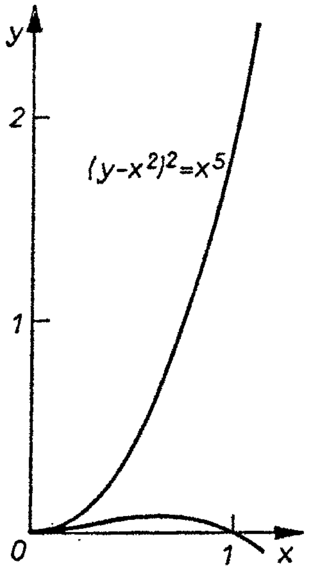

Abb. 138

Ist $a_{11} = a_{12} = a_{22} = 0$, so muß man Ableitungen höherer Ordnung heranziehen. In diesem Fall sind kompliziertere Typen singulärer Punkte möglich (*dreifache* oder allgemeiner *n-fache Punkte* usw.).

237. Parameterdarstellung der Kurve. Wir gehen nun noch auf singuläre Punkte ebener Kurven ein, die in der Parameterdarstellung

$$x = \varphi(t), \quad y = \psi(t)$$

gegeben sind.

Es sei für $t = t_0$

$$x_0' = \varphi'(t_0) = 0 \quad \text{und} \quad y_0' = \psi'(t_0) = 0,$$

aber von den Ableitungen zweiter Ordnung x_0'' und y_0'' sei wenigstens eine, etwa x_0'', von 0 verschieden.

Wir legen die Sekante durch die Kurvenpunkte (x_0, y_0) und (x, y), die den Parameterwerten t_0 und t entsprechen. Ihre Gleichung kann in der Form

$$\frac{X - x_0}{x - x_0} = \frac{Y - y_0}{y - y_0}$$

geschrieben werden. Nach der Taylorschen Formel [mit dem Peanoschen Restglied; vgl. Nr. 124, Formel (10a)] ist wegen $x_0' = y_0' = 0$

$$x - x_0 = \frac{1}{2}(x_0'' + \alpha)(t - t_0)^2, \quad y - y_0 = \frac{1}{2}(y_0'' + \beta)(t - t_0)^2,$$

wobei α und β für $t \to t_0$ gegen 0 streben. Setzen wir diese Ausdrücke für $t \neq t_0$ in die Sekantengleichung ein und dividieren beide Nenner durch $\frac{1}{2}(t - t_0)^2$, so erhalten wir

$$\frac{X - x_0}{x_0'' + \alpha} = \frac{Y - y_0}{y_0'' + \beta}.$$

Hier kann man für $t \to t_0$ zur Grenze übergehen[1]); es ergibt sich die Tangentengleichung

$$\frac{X - x_0}{x_0''} = \frac{Y - y_0}{y_0''} \quad \text{oder} \quad Y - y_0 = \frac{y_0''}{x_0''}(X - x_0). \tag{19}$$

Wir haben $x_0'' \neq 0$ vorausgesetzt; es sei etwa $x_0'' > 0$. Dann hat $x = \varphi(t)$ für $t = t_0$ ein (eigentliches) Minimum (vgl. Nr. 137), d. h., für Werte von t in der Nähe von t_0 (sowohl für $t < t_0$ als auch für $t > t_0$) ist $x > x_0$. Somit laufen in (x_0, y_0) zwei Kurvenzweige zusammen, die $t < t_0$ und $t > t_0$ entsprechen; sie haben eine gemeinsame (horizontale oder schräge) Tangente und liegen beide rechts von der Vertikalen $x = x_0$. Es liegt also ein Rückkehrpunkt vor (Abb. 139). Das ist ein wichtiger Fall eines singulären Punktes einer Kurve in Parameterdarstellung.

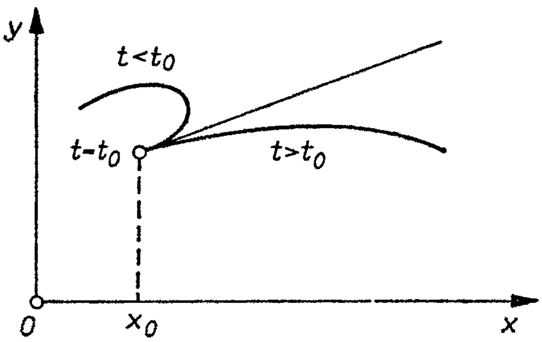

Abb. 139

Um herauszubekommen, welcher Art dieser Rückkehrpunkt ist, treiben wir die Untersuchung etwas weiter. Zu diesem Zweck ziehen wir die dritten Ableitungen heran und schreiben die Zuwächse $x - x_0$ und $y - y_0$ in der Form

$$x - x_0 = \frac{1}{2} x_0''(t - t_0)^2 + \frac{1}{6}(x_0''' + \tilde{\alpha})(t - t_0)^3,$$

$$y - y_0 = \frac{1}{2} y_0''(t - t_0)^2 + \frac{1}{6}(x_0''' + \tilde{\beta})(t - t_0)^3,$$

wobei $\tilde{\alpha}$ und $\tilde{\beta}$ für $t \to t_0$ gegen 0 streben.

Unter Benutzung von Gleichung (19) rechnen wir die Ordinate Y des Punktes der Tangente mit der Abszisse x aus; wir finden

$$Y - y_0 = \frac{y_0''}{x_0''}(x - x_0) = \frac{1}{2} y_0''(t - t_0)^2 + \frac{1}{6} \cdot \frac{y_0''}{x_0''}(x_0''' + \tilde{\alpha})(t - t_0)^3.$$

Nun bilden wir die Differenz der Ordinaten Y und y zu derselben Abszisse x:

$$Y - y = \frac{1}{6}\left(\frac{x_0'''y_0'' - x_0''y_0'''}{x_0''} + \tilde{\gamma}\right)(t - t_0)^3,$$

wobei $\tilde{\gamma}$ für $t \to t_0$ gegen 0 strebt.

Ist jetzt $x_0'''y_0'' - x_0''y_0''' \neq 0$ (was gewöhnlich der Fall sein dürfte), so hat $Y - y$ für $t < t_0$ und für $t > t_0$, d. h. für die beiden Kurvenzweige, die in (x_0, y_0) zusammenstoßen (natürlich unter der Annahme, daß man sich auf zu t_0 hinreichend benachbarte

[1]) Vgl. die Bemerkung in Nr. 234, die auch hier anwendbar ist, wenn der betrachtete Punkt als *einfach* vorausgesetzt wird.

t beschränkt), verschiedenes Vorzeichen. Die Kurvenzweige verlaufen auf verschiedenen Seiten der Tangente; es handelt sich um einen Rückkehrpunkt *erster Art*.

Solche Singularitäten sind uns schon mehrmals begegnet: bei der Zykloide, bei Epi- und Hypozykloiden, bei der Kreisevolvente (vgl. dazu Abb. 118 bis 121); diese Kurven haben Rückkehrpunkte erster Art.

Ist aber $x_0'''y_0'' - x_0''y_0''' = 0$, so beginnt die Entwicklung der Differenz $Y - y$ nach Potenzen von $t - t_0$ mit der vierten oder einer höheren Potenz. Ist diese gerade, so liegt ein Rückkehrpunkt *zweiter Art* vor.

§ 3. Berührung von Kurven

238. Die Einhüllende einer Kurvenschar. Haben zwei Kurven einen Punkt M_0 gemeinsam und dort eine gemeinsame Tangente, so sagt man, die Kurven *berühren* einander in M_0. Dieser Paragraph befaßt sich mit einigen Problemen der Berührung ebener Kurven.

Ehe wir zum Begriff der Einhüllenden einer Kurvenschar übergehen, befassen wir uns mit dem Begriff der Kurvenschar selbst. Wir hatten schon mehrmals mit Kurvengleichungen zu tun, in denen außer den laufenden Koordinaten x und y des veränderlichen Punktes ein (oder mehrere) Parameter vorkam(en). Handelt es sich um einen Parameter, etwa a, so lautet die Gleichung

$$F(x, y, a) = 0. \tag{1}$$

Die linke Seite ist eine Funktion dreier Veränderlicher, von denen wir eine nur deshalb anders nennen, weil sie eine besondere Rolle spielt: Um eine bestimmte Kurve zu erhalten, muß man dem Parameter a einen bestimmten Wert erteilen. Ändert man diesen Wert, etwa innerhalb eines Intervalls, so erhält man im allgemeinen verschiedene Kurven (die hinsichtlich ihrer Gestalt oder Lage voneinander abweichen).

Die Gesamtheit aller dieser Kurven nennt man eine *einparametrige Kurvenschar*, die Gleichung (1) die *Gleichung der Schar*.

Manchmal existiert für eine solche Kurvenschar eine Kurve, die jede Kurve der Schar in einem oder in mehreren Punkten berührt und dabei ganz aus solchen Berührungspunkten besteht (Abb. 140). Eine solche Kurve wird *Einhüllende* der Schar genannt. Wir wollen zeigen, wie man feststellt, ob eine Einhüllende existiert, und wie man sie dann findet.

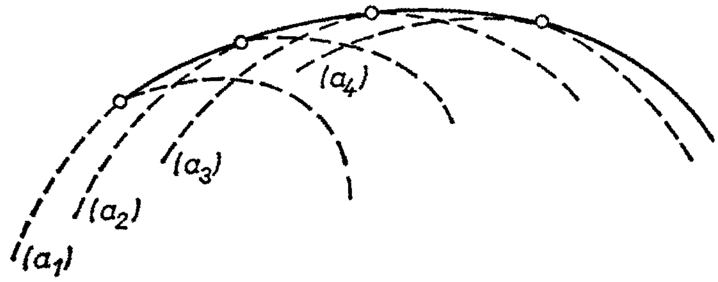

Abb. 140

Zu diesem Zweck nehmen wir zunächst an, es existiere eine Einhüllende. Der Einfachheit halber setzen wir dabei voraus, es handele sich um eine Einhüllende (genauer: um einen Zweig der Einhüllenden), die jede Kurve der Schar in einem einzigen Punkt berührt. Dann sind die Koordinaten dieses Berührungspunktes eindeutig durch die Angabe der Kurve der Schar, d. h. durch den Wert des Parameters a, bestimmt:

$$x = \varphi(a), \quad y = \psi(a). \tag{2}$$

Da die Einhüllende ganz aus Berührungspunkten besteht, liefern diese Beziehungen eine Parameterdarstellung der Einhüllenden.

Existenz und Stetigkeit der partiellen Ableitungen von F nach allen drei Variablen und der Ableitungen von φ und ψ seien vorausgesetzt.

Der Punkt (2) liegt auf der Kurve (1), die durch eben diesen Parameterwert bestimmt wird; daher gilt identisch in a

$$F\big(\varphi(a), \psi(a), a\big) = 0. \tag{3}$$

Durch vollständiges Differenzieren[1]) nach a erhalten wir nach Nr. 181 und 185

$$F_x\,\mathrm{d}x + F_y\,\mathrm{d}y + F_a\,\mathrm{d}a = 0, \tag{4}$$

wobei die Ableitungen für die in (3) angegebenen Argumente zu berechnen sind und $\mathrm{d}x$ und $\mathrm{d}y$ die Differentiale der Funktionen (2) bezeichnen.

Jetzt wollen wir die Tatsache, daß die Einhüllende im Punkt (2) die Kurve (1) berührt, analytisch ausdrücken. Die Tangente an die Kurve (1) — vgl. Nr. 230, Formel (5) —, deren Gleichung

$$F_x(X - x) + F_y(Y - y) = 0 \tag{5}$$

lautet, und die Tangente an die Kurve (2) — vgl. Nr. 230, Formel (7) — mit der Gleichung

$$\frac{X - x}{\mathrm{d}x} = \frac{Y - y}{\mathrm{d}y} \tag{6}$$

müssen zusammenfallen. Diese Bedingung läßt sich in der Form

$$F_x\,\mathrm{d}x + F_y\,\mathrm{d}y = 0 \tag{7}$$

ausdrücken. Dabei verstehen wir wie oben unter x und y die Werte (2) und unter $\mathrm{d}x$ und $\mathrm{d}y$ die Differentiale der Funktionen (2).

Die Gleichungen (5) und (6) beschreiben nur dann Tangenten an Kurven, wenn der betrachtete Punkt nicht singulär ist. Trotzdem gilt die Gleichung (7) sogar dann, wenn dieser Punkt für die eine oder die andere Kurve singulär ist.

Vergleicht man (7) mit (4) und berücksichtigt, daß $\mathrm{d}a$ eine beliebige Zahl ist, so findet man $F_a = 0$ oder, ausgeschrieben,

$$F_a\big(\varphi(a), \psi(a), a\big) = 0. \tag{8}$$

Die Identitäten (3) und (8) zeigen, daß die uns bekannten Funktionen (2) identisch in a dem Gleichungssystem

$$F(x, y, a) = 0, \quad F_a(x, y, a) = 0 \tag{9}$$

genügen müssen.

Wenn also eine Einhüllende existiert, so erhält man ihre Parameterdarstellung durch Auflösen des Systems (9) nach x und y.

---

[1]) Hier benutzen wir übrigens die Stetigkeit der partiellen Ableitungen von F.

Wenn dieses System für veränderliches a keine Lösung als Funktion von a hat, ist die Lage klar; dann gibt es keine Einhüllende. Wir wollen nun annehmen, wir hätten durch Auflösung des Systems (9) Gleichungen (2) erhalten, die eine Kurve ohne singuläre Punkte darstellen.[1]) Ist sie Einhüllende unserer Kurvenschar?

Da die Funktionen (2) den Gleichungen (9) genügen, sind die Identitäten (3) und (8) erfüllt. Differenzieren wir die erste, so erhalten wir (4), vergleichen wir sie mit (8), so kommen wir zu (7). Ist der Punkt (2) für kein a auf der entsprechenden Kurve singulär, so daß (5) tatsächlich die Gleichung der Tangente an die Kurve ist, so ist Gleichung (7) die Bedingung dafür, daß diese Tangente mit der Tangente (6) an die Kurve (2) zusammenfällt. Dann ist also die Kurve (2) tatsächlich Einhüllende der Schar.

Insbesondere ist das der Fall, wenn beispielsweise die Kurven der gegebenen Schar überhaupt keine singulären Punkte besitzen.

Gibt es dagegen solche singulären Punkte und bildet bei Änderung von a ihr geometrischer Ort eine Kurve (2), so genügen die entsprechenden Funktionen φ und ψ dem System (9),[2]) obwohl in diesem Fall die Kurve nicht Einhüllende zu sein braucht.

Beim Vorliegen singulärer Punkte muß also die durch Auflösen des Systems (9) erhaltene Kurve (2) noch überprüft werden: Sie kann Einhüllende sein oder aber der geometrische Ort der singulären Punkte auf den Kurven der Schar oder schließlich zum Teil Einhüllende, zum Teil Ort der singulären Punkte.

Gewöhnlich bleibt man beim Aufsuchen der Einhüllenden nicht beim Gleichungssystem (9) stehen, sondern geht weiter, man eliminiert daraus a. Man erhält dann eine Beziehung

$$\Phi(x, y) = 0, \tag{10}$$

die a nicht mehr enthält und eine notwendige und hinreichende Bedingung dafür darstellt, daß für Paare x, y ein solcher Wert von a gefunden werden kann, für den zusammen mit diesen Paaren beide Gleichungen (9) erfüllt sind.

Alle Punkte der durch Auflösung des Systems (9) erhaltenen Kurve (2) müssen der Gleichung (10) genügen. Wenn daher diese keine Kurve beschreibt, dann ist klar, daß keine Einhüllende existiert. Gibt es aber zu (10) eine Kurve — man nennt sie die *Diskriminantenkurve* der Schar —, so muß man wie oben nachprüfen. Sie kann Einhüllende sein (wenn eine solche existiert), aber auch geometrischer Ort der singulären Punkte (wenn es solche gibt). Außerdem gibt es noch eine weitere Möglichkeit, die durch Nachprüfen ausgeschaltet werden muß. Die Diskriminantenkurve kann auch eine oder mehrere spezielle Kurven der Schar enthalten. Das ist dann der Fall, wenn unendlich vielen Punkten der Diskriminantenkurve ein und derselbe Wert a entspricht, so daß die Gleichungen (9) erfüllt sind.[3])

Diese Ausführungen erläutern wir an Beispielen.

239. Beispiele.

1. Man bestimme die Einhüllende der Schar der Kreise $(x - a)^2 + y^2 = r^2$ $(r = \text{const})$ (Abb. 141).

---

[1]) Falls einzelne singuläre Punkte vorliegen, beschränken wir uns auf solche Parameterintervalle, die keine kritischen Werte enthalten.

[2]) Für sie ist (3), also auch (9) erfüllt. Alsdann gilt (7), wie oben im Text erwähnt; vergleicht man mit (4), so gelangt man zu (8).

[3]) Operiert man direkt mit den Gleichungen (9), so ist diese Möglichkeit ausgeschlossen, weil man die Gleichungen gerade bei veränderlichem a zu lösen versucht.

Durch Differenzieren nach a folgt $-2(x-a)=0$. Eliminiert man a, so ergibt sich $y^2-r^2=0$, also $y=\pm r$. Das sind zwei zur x-Achse parallele Geraden, offenbar Einhüllende.[1])

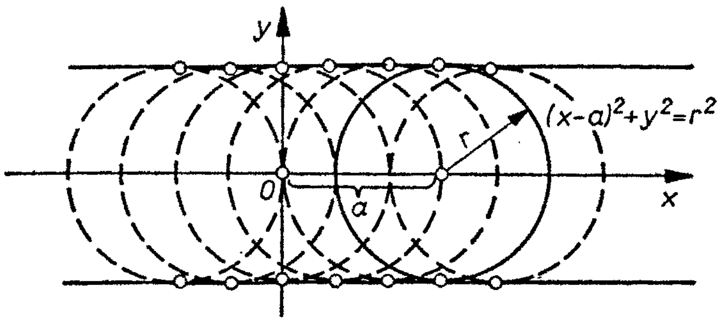

Abb. 141

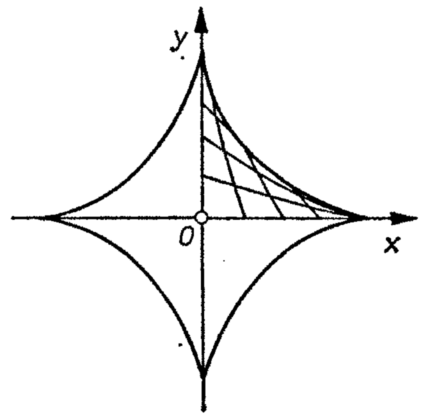

Abb. 142

2. Man bestimme die Einhüllende der verschiedenen Lagen der Geraden, auf welchen von der Koordinatenachse die gleiche Strecke a herausgeschnitten wird (Abb. 142).

Nimmt man (etwa im ersten Quadranten) als Parameter den Winkel θ, den die Senkrechten zu der Geradenschar mit der x-Achse bilden, so lautet die Geradengleichung

$$\frac{x}{\sin\theta}+\frac{y}{\cos\theta}=a.$$

Differenziert man nach θ, so folgt

$$-\frac{x}{\sin^2\theta}\cos\theta+\frac{y}{\cos^2\theta}\sin\theta=0 \qquad \text{oder} \qquad \frac{x}{\sin^3\theta}=\frac{y}{\cos^3\theta}.$$

Daraus folgt

$$\frac{\dfrac{x}{\sin\theta}}{\sin^2\theta}=\frac{\dfrac{y}{\cos\theta}}{\cos^2\theta}$$

oder (nach korrespondierender Addition)

$$\frac{\dfrac{x}{\sin\theta}}{\sin^2\theta}=\frac{\dfrac{y}{\cos\theta}}{\cos^2\theta}=\frac{\dfrac{x}{\sin\theta}+\dfrac{y}{\cos\theta}}{\sin^2\theta+\cos^2\theta}=a;$$

also ist $x=a\sin^3\theta$, $y=a\cos^3\theta$.

Darin erkennt man die Parameterdarstellung der Astroide (vgl. Nr. 224, Beispiel 4, mit $t=\dfrac{\pi}{2}-\theta$), die hier tatsächlich Einhüllende ist.

Übrigens hatten wir schon in Nr. 231, Beispiel 3, auf diese Eigenschaft der Astroide hingewiesen.

---

[1]) Schreibt man die Gleichung der Schar in der Form

$$x-a\pm\sqrt{r^2-y^2}=0,$$

so ergibt sich beim Differenzieren nach a die sinnlose Beziehung $-1=0$. Daraus könnte man schließen, es gäbe keine Einhüllende. Das wäre aber falsch, da unsere Methode Existenz und Stetigkeit der partiellen Ableitungen der linken Seite der Schargleichung voraussetzt, während es hier für $y=\pm r$ keine endliche Ableitung nach y gibt.

3. In vielen Fällen begrenzt also die Einhüllende sozusagen den Teil der Ebene, der von den Kurven der Schar eingenommen wird. Das ist aber nicht immer so, wie das Beispiel $y = (x - a)^3$ (Abb. 143) zeigt. Hier ist die x-Achse Einhüllende; sie schneidet alle Kurven der Schar.

Analog ist es bei dem folgenden, komplizierteren Beispiel:

4. Man bestimme die Einhüllende der Parabelschar $y = a^2(x - a)^2$. Differentiation nach a liefert

$$2a(x - a)^2 - 2a^2(x - a) = 2a(x - a)(x - 2a) = 0,$$

also entweder $x = a$ $(y = 0)$ oder $x = 2a$ $(y = a^4)$. so daß die Diskriminantenkurve aus der Geraden $y = 0$ und der Kurve $16y = x^4$ besteht. Die erste berührt alle Parabeln im Scheitel, die zweite hat mit jeder Parabel drei Punkte gemeinsam: Sie berührt sie bei $x = 2a$ und schneidet sie bei $x = -2a \pm 2a\sqrt{2}$.

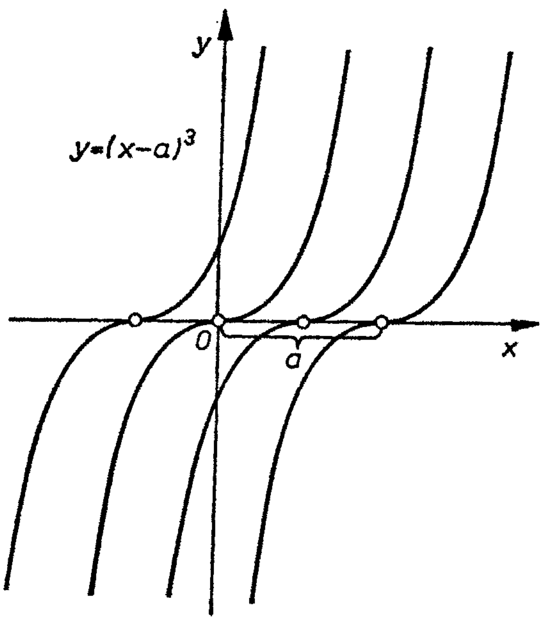

Abb. 143

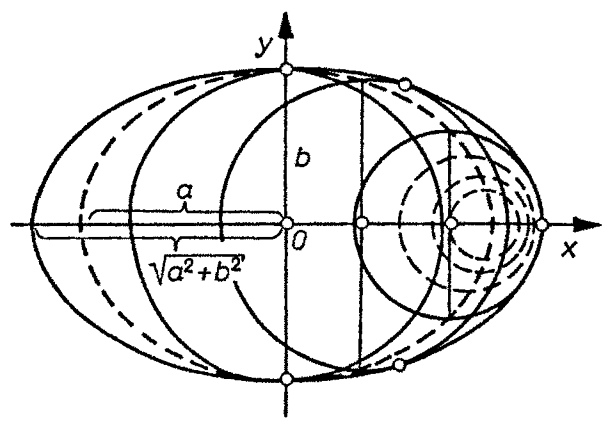

Abb. 144

5. Wir betrachten die Ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$. Gesucht ist die Einhüllende der Schar der Kreise über den der y-Achse parallelen Ellipsensehnen als Durchmesser (Abb. 144).

Als Parameter nehmen wir die Abszisse t des Kreismittelpunktes und schreiben die Gleichung der Kreisschar in der Gestalt

$$F(x, y, t) = (x - t)^2 + y^2 - \frac{b^2}{a^2}(a^2 - t^2) = 0,$$

wobei t im Intervall $[-a, +a]$ variiert. Dann gilt

$$F_t = -2(x - t) + \frac{2b^2}{a^2}t = 0, \quad \text{also} \quad t = \frac{a^2}{a^2 + b^2}x.$$

Setzen wir diese Wert von t in die Beziehung $F = 0$ ein, so erhalten wir die Gleichung der Einhüllenden in der Gestalt

$$\left(x - \frac{a^2x}{a^2 + b^2}\right)^2 + y^2 - \frac{b^2}{a^2}\left(a^2 - \frac{a^4x^2}{(a^2 + b^2)^2}\right) = 0$$

oder, umgeformt,

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

Das ist eine Ellipse mit derselben Achsenlage wie die gegebene.

Es ist interessant, daß diese Ellipse nicht alle Kreise der Schar berührt. Das sieht man leicht ein, wenn man nicht t aus den Gleichungen $F = 0$ und $F_t = 0$ eliminiert, sondern mit ihrer Hilfe x und y durch t ausdrückt:

$$x = \frac{a^2 + b^2}{a^2}\, t, \qquad y = \pm \frac{b}{a^2} \sqrt{a^4 - (a^2 + b^2)\, t^2}.$$

Hieraus erkennt man, daß y nur für $|t| \leqq \dfrac{a^2}{\sqrt{a^2 + b^2}}$ reell sein kann. Demnach existiert nur für denjenigen Teil der Kreisschar, der diesen Werten von t entspricht, die Einhüllende.

Dieses lehrreiche Beispiel zeigt, daß die Parameterdarstellung der Einhüllenden zweckmäßiger sein kann, da man daraus leichter erkennt, für welchen Teil der gegebenen Schar wirklich eine Einhüllende existiert.

6. Für die Schar der konzentrischen Kreise $x^2 + y^2 = a$ $(a \geqq 0)$ gibt es keine Einhüllende (differenziert man nach a, so erhält man links 0, rechts 1).

7. Wir betrachten die zwei Scharen semikubischer Parabeln

$$\text{(a)} \quad (y - a)^2 - x^3 = 0 \quad \text{und} \quad \text{(b)} \quad y^2 - (x - a)^3 = 0$$

(Abb. 145). Die Diskriminantenkurven sind

$$\text{(a)} \quad x = 0, \qquad \text{(b)} \quad y = 0,$$

in beiden Fällen die Träger der singulären Punkte. Im Fall (b) ist die Kurve jedoch auch Einhüllende, während es im Fall (a) keine Einhüllende gibt.

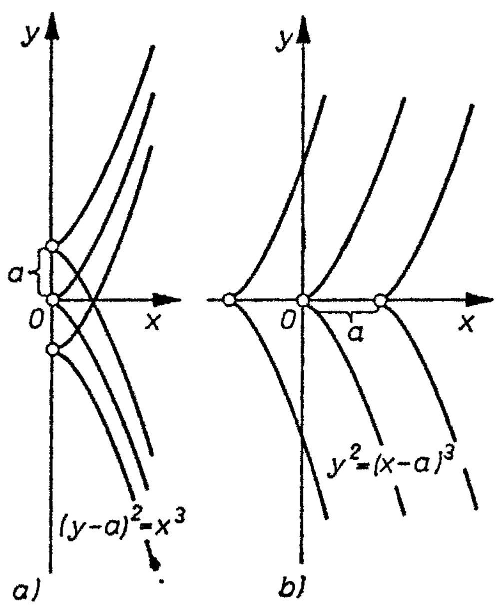

Abb. 145

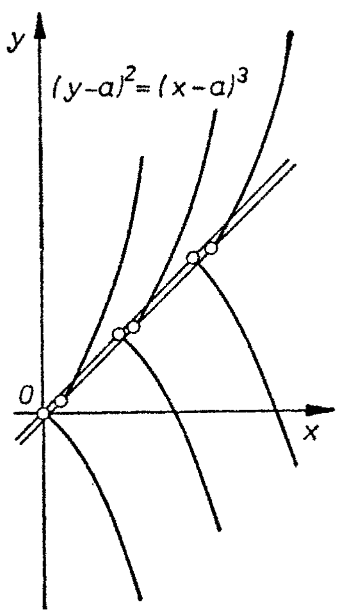

Abb. 146

8. Ein komplizierteres Beispiel desselben Typus liefert die Schar der semikubischen Parabeln $(y - a)^2 - (x - a)^3 = 0$ (Abb. 146). Hier zerfällt die Diskriminantenkurve in die Geraden $y = x$ und $y = x - \dfrac{4}{27}$. Die erste ist nur der geometrische Ort der singulären Punkte, die zweite ist Einhüllende.

9. Zum Schluß betrachten wir die Geradenschar $4(1 + t)\, x = t^2 y$. Differenziert man nach t und eliminiert t aus der Kurvengleichung und der Beziehung $4x = 2ty$, so erhält man

$$x(x + y) = 0.$$

Das sind die beiden Geraden $x = 0$ und $y = -x$, die (für $t = 0$ bzw. $t = -2$) zu unserer Geradenschar gehören. Keine von ihnen ist Einhüllende, keine ist Träger der singulären Punkte. Es gibt also keine Einhüllende.

Dieses Beispiel illustriert unsere früheren Bemerkungen, daß die Gleichung (10) nicht die Einhüllende, sondern eine oder mehrere Kurven der Schar darstellen kann. Hätten wir t nicht eliminiert, sondern versucht, x und y durch t bei veränderlichem t auszudrücken, so hätte sich das als unmöglich herausgestellt.

240. Charakteristische Punkte. Mit dem Begriff der Einhüllenden hängt ein anderer interessanter geometrischer Begriff eng zusammen, der Begriff der charakteristischen Punkte.

Wir betrachten eine Kurve der Schar $F(x, y, a) = 0$, die durch den Wert a des Parameters bestimmt ist. Diesem a erteilen wir den Zuwachs Δa; dem Parameterwert $a + \Delta a$ entspricht die Kurve

$$F(x, y, a + \Delta a) = 0,$$

die der ersten „benachbart" ist.

Es kann vorkommen, daß sich bei hinreichend kleinem Δa die beiden Kurven in einem oder in mehreren Punkten schneiden. Strebt Δa gegen 0, so werden sich diese Schnittpunkte irgendwie längs der ersten Kurve bewegen. Strebt dabei einer von ihnen gegen eine bestimmte Grenzlage, so nennt man diesen Grenzwert einen *charakteristischen Punkt* auf der Ausgangskurve (Abb. 147). Wir machen den Leser darauf

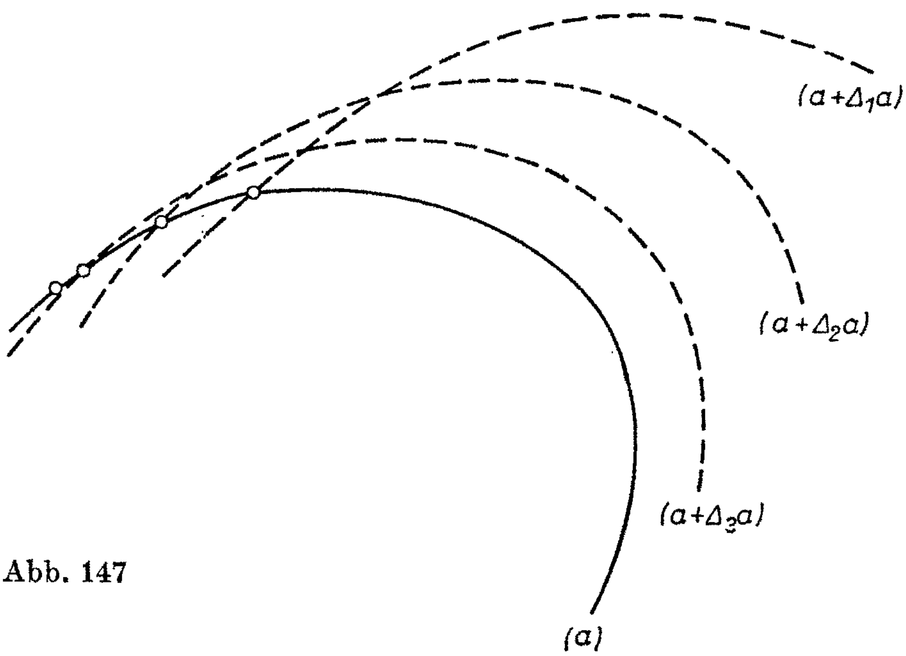

Abb. 147

aufmerksam, daß der Begriff des charakteristischen Punktes nicht nur mit der Kurve, auf der dieser liegt, sondern mit der ganzen Schar zusammenhängt. Von einem charakteristischen Punkt einer einzelnen Kurve zu sprechen, wäre sinnlos.

Ein charakteristischer Punkt der obengenannten Kurven muß dem Gleichungssystem

$$F(x, y, a) = 0, \quad F(x, y, a + \Delta a) = 0$$

genügen oder dem ihm gleichwertigen System

$$F(x, y, a) = 0, \quad \frac{F(x, y, a + \Delta a) - F(x, y, a)}{\Delta a} = 0. \tag{11}$$

Strebt Δa gegen 0, so kommen wir zu dem uns schon bekannten System (9)

$$F(x, y, a) = 0, \quad F_a(x, y, a) = 0,$$

dem somit bei gegebenem a die Koordinaten eines charakteristischen Punktes genügen müssen.

Genauer gesagt: Sind x und y die Koordinaten des Schnittpunktes, so kann man statt (11) nach dem Mittelwertsatz auch

$$F(x, y, a) = 0, \quad F_a(x, y, a + \theta\,\Delta a) = 0 \quad (0 < \theta < 1)$$

schreiben. Streben x, y für $\Delta a \to 0$ gegen $\bar{x}, \bar{y}$, so überzeugt man sich auf Grund der Stetigkeit von F und F_a durch Grenzübergang davon, daß die Koordinaten $\bar{x}, \bar{y}$ des charakteristischen Punktes dem System (9) genügen.

Wir wollen nun annehmen, daß auf jeder Kurve der Schar charakteristische Punkte liegen. Dann kann man die Frage nach dem geometrischen Ort der charakteristischen Punkte stellen. Wird dieser durch eine Kurve der Form (2) dargestellt, so müssen die Funktionen $\varphi(a)$, $\psi(a)$, die in ihren Gleichungen vorkommen, dem System (9) genügen, d. h. in der Menge der Lösungen dieses Systems nach x, y enthalten sein. Ebenso genügen alle Punkte dieses geometrischen Ortes auch der Gleichung (10), d. h., dieser geometrische Ort ist notwendigerweise Bestandteil der Diskriminantenkurve.

Demnach ist der geometrische Ort der charakteristischen Punkte, wenn es solche gibt, entweder (vollständig oder teilweise) Einhüllende oder Träger der singulären Punkte.

Man überzeugt sich leicht davon, daß in den Beispielen 1, 2, 4 und 5 aus Nr. 239 der geometrische Ort der charakteristischen Punkte mit der Einhüllenden übereinstimmt. Das ist im gewissen Sinne der allgemeine Fall. Im Beispiel 7 (a) ist dieser geometrische Ort nur Träger der singulären Punkte, dagegen gibt es in den Beispielen 3 und 7 (b) keine Schnittpunkte von Kurven der Schar (obwohl eine Einhüllende existiert).

241. Ordnung der Berührung zweier Kurven. Wir betrachten zwei Kurven, die sich in einem Punkt M_0 berühren.

Sind die Kurven durch die expliziten Gleichungen $y = f(x)$ und $Y = g(x)$ gegeben und hat M_0 die Abszisse x_0, so gilt auf Grund der Übereinstimmung von Ordinaten und Richtungskoeffizienten

$$f(x_0) = g(x_0), \quad f'(x_0) = g'(x_0).$$

Um zu charakterisieren, daß die betrachteten Kurven in der Nähe von M_0 benachbart sind, wählen wir auf den Kurven je einen Punkt M und m mit der Abszisse

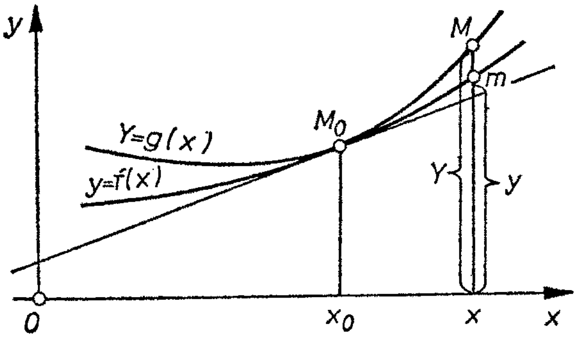

Abb. 148

x (vgl. Abb. 148) und bestimmen die Ordnung des Kleinwerdens der Strecke

$$\overline{mM} = Y - y = g(x) - f(x) = \varphi(x)$$

in bezug auf die Differenz $x - x_0$. Ist diese Ordnung mindestens $n + 1$, so sagen wir, *die Berührung der Kurven in M_0 habe mindestens die Ordnung n* (ist die erstgenannte Ordnung gleich $n + 1$, so sprechen wir von einer Berührung n-ter Ordnung).

Im Fall einer Berührung ist stets

$$\varphi(x_0) = g(x_0) - f(x_0) = 0, \quad \varphi'(x_0) = g'(x_0) - f'(x_0) = 0.$$

Es mögen nun in x_0 alle Ableitungen von $f(x)$ und $g(x)$ bis zur Ordnung $n + 1$ einschließlich existieren, und es sei

$$f''(x_0) = g''(x_0), \ldots, f^{(n)}(x_0) = g^{(n)}(x_0),$$

also

$$\varphi''(x_0) = g''(x_0) - f''(x_0) = 0,$$

$$\varphi^{(n)}(x_0) = g^{(n)}(x_0) - f^{(n)}(x_0) = 0.$$

Über die Größen $f^{(n+1)}(x_0)$ und $g^{(n+1)}(x_0)$ wollen wir zunächst nichts voraussetzen. Wenden wir auf $\varphi(x)$ die Taylorsche Formel mit Peanoschem Restglied [vgl. Nr. 124, Formel (10a)] an, so folgt

$$\overline{mM} = Y - y = \varphi(x) = \frac{\varphi^{(n+1)}(x_0) + \alpha}{(n+1)!}(x - x_0)^{n+1}, \tag{12}$$

und es ist

$$\lim_{x \to x_0} \frac{\overline{mM}}{(x - x_0)^{n+1}} = \frac{\varphi^{(n+1)}(x_0)}{(n+1)!} = \frac{g^{(n+1)}(x_0) - f^{(n+1)}(x_0)}{(n+1)!}.$$

Ist also $f^{(n+1)}(x_0) \neq g^{(n+1)}(x_0)$, so liegt eine Berührung n-ter Ordnung vor; ist aber $f^{(n+1)}(x_0) = g^{(n+1)}(x_0)$, so ist die Ordnung der Berührung höher als n. Hieraus folgt (unter der Annahme, daß die erwähnten Ableitungen existieren):

Die Berührung der Kurven $y = f(x)$ und $Y = g(x)$ im Punkt mit der Abszisse x_0 ist genau dann von der Ordnung n, wenn die Bedingungen

$$f(x_0) = g(x_0), \quad f'(x_0) = g'(x_0), \quad \ldots, \quad f^{(n)}(x_0) = g^{(n)}(x_0), \tag{13}$$

$$f^{(n+1)}(x_0) \neq g^{(n+1)}(x_0) \tag{14}$$

erfüllt sind. Ist (14) noch nicht nachgewiesen, so kann man nur sagen, die Ordnung der Berührung sei mindestens gleich n.

Ist die Ordnung der Berührung genau gleich n, so folgt aus (12) sofort, daß *bei geradem n die Kurven in M_0 einander schneiden (durchsetzen), während das bei ungeradem n nicht der Fall ist.*

Bemerkung. Nach diesen Ergebnissen wenden wir uns nochmals der Definition der Ordnung der Berührung zu. Diese Definition hängt anscheinend von der Wahl des Koordinatensystems ab. In Wirklichkeit ist das jedoch nicht der Fall (die y-Achse soll der gemeinsamen Tangente nicht parallel sein), so daß die Ordnung der Berührung zweier Kurven ein echter geometrischer Begriff ist.

Dreht man nämlich das Koordinatensystem um einen beliebigen Winkel α, so lassen sich die neuen Koordinaten $\bar{x}, \bar{y}$ mit Hilfe der Transformationsformeln

$$\bar{x} = x \cos\alpha + y \sin\alpha, \quad \bar{y} = -x \sin\alpha + y \cos\alpha$$

durch die alten ausdrücken. Im x, y-System sei die Kurve $y = f(x)$ gegeben. Versteht man in diesen Formeln unter y eben diese Funktion, so liefern sie eine Parameterdarstellung der Kurve im neuen System, mit x als Parameter. Offenbar können die Ableitungen

$$\frac{\mathrm{d}\bar{x}}{\mathrm{d}x} = \cos\alpha + \frac{\mathrm{d}y}{\mathrm{d}x}\sin\alpha, \quad \frac{\mathrm{d}\bar{y}}{\mathrm{d}x} = -\sin\alpha + \frac{\mathrm{d}y}{\mathrm{d}x}\cos\alpha$$

nicht gleichzeitig verschwinden, so daß im neuen System kein Punkt singulär ist. Dann ist aber klar, daß die erste dieser Ableitungen in dem uns interessierenden Punkt nicht verschwindet (denn sonst wäre die Tangente an die Kurve in diesem Punkt der $\bar{y}$-Achse parallel). Daher läßt sich in seiner Umgebung die Kurve auch im neuen System durch eine explizite Gleichung $\bar{y} = \bar{f}(\bar{x})$ ausdrücken.

Jetzt sieht man leicht, daß (vgl. Nr. 121)

$$\frac{\mathrm{d}\bar{y}}{\mathrm{d}\bar{x}} = \frac{-\sin\alpha + \dfrac{\mathrm{d}y}{\mathrm{d}x}\cos\alpha}{\cos\alpha + \dfrac{\mathrm{d}y}{\mathrm{d}x}\sin\alpha}, \quad \frac{\mathrm{d}^2\bar{y}}{\mathrm{d}\bar{x}^2} = \frac{\dfrac{\mathrm{d}^2y}{\mathrm{d}x^2}}{\left(\cos\alpha + \dfrac{\mathrm{d}y}{\mathrm{d}x}\sin\alpha\right)^3}$$

und allgemein

$$\frac{\mathrm{d}^k\bar{y}}{\mathrm{d}\bar{x}^k} = R_k\left(\frac{\mathrm{d}y}{\mathrm{d}x}, \frac{\mathrm{d}^2y}{\mathrm{d}x^2}, \ldots, \frac{\mathrm{d}^ky}{\mathrm{d}x^k}\right)$$

gilt, wobei R_k eine rationale Funktion bezeichnet. Hieraus folgt: Sobald für zwei Funktionen y von x die Gleichungen (13) erfüllt sind, so sind für die beiden entsprechenden Funktionen $\bar{y}$ von $\bar{x}$ analoge Beziehungen erfüllt. Ebenso folgt, wenn (13) gilt, aus (14) dieselbe Ungleichung für die neuen Funktionen; denn sonst würde die inverse Transformation von (14) auf eine Gleichheit führen.

Damit ist unsere Behauptung vollständig bewiesen.

242. Eine der Kurven ist implizit gegeben. Die zweite Kurve sei nun durch die implizite Gleichung

$$G(x, y) = 0 \tag{15}$$

gegeben. Der Punkt $M_0(x_0, y_0)$ sei kein singulärer Punkt dieser Kurve, und zwar sei $G_y(x_0, y_0) \neq 0$. Dann definiert (15) in der Umgebung von M_0 eine eindeutig bestimmte Funktion $y = g(x)$, und um die Ordnung der Berührung zu bestimmen, kann man die bekannten Bedingungen (13) und (14) benutzen.

Da uns jedoch der explizite Ausdruck für $g(x)$ nicht vorliegt, ist es zweckmäßig, diese Bedingungen auf eine Form zu bringen, bei der nur die tatsächlich gegebene Funktion G benutzt wird.

Zu diesem Zweck wollen wir annehmen, daß die Werte von $g(x)$ und die der Ableitungen $g'(x), g''(x), \ldots, g^{(n)}(x)$ sukzessive, und zwar eindeutig, durch die Gleichung (15) und diejenigen Gleichungen, die man daraus durch Differentiation nach x erhält, bestimmt werden können, wenn man unter y eben $g(x)$ versteht (vgl. Nr. 209):

$$G(x, \underline{g(x)}) = 0,$$
$$G_x(x, g(x)) + G_y(x, g(x))\, \underline{g'(x)} = 0,$$
$$G_{xx} + 2G_{xy}\, g'(x) + G_{yy}[g'(x)]^2 + G_y \underline{g''(x)} = 0,$$
$$\cdots\cdots\cdots\cdots$$
$$G_{\underbrace{x\ldots x}_{n}} + \cdots + G_y \underline{g^{(n)}(x)} = 0.\text{[1]}$$

Setzen wir für $x = x_0$ in diesen Gleichungen für $g(x_0), g'(x_0), \ldots, g^{(n)}(x_0)$ überall $f(x_0), f'(x_0), \ldots, f^{(n)}(x_0)$ ein, so erhalten wir die den Bedingungen (13) völlig gleichwertigen Bedingungen

$$G(x_0, \underline{f(x_0)}) = 0,$$
$$G_x(x_0, f(x_0)) + G_y(x_0, f(x_0))\, \underline{f'(x_0)} = 0,$$
$$G_{xx} + 2G_{xy}\, f'(x_0) + G_{yy}[f'(x_0)]^2 + G_y \underline{f''(x_0)} = 0,$$
$$\cdots\cdots\cdots\cdots$$
$$G_{\underbrace{x\ldots x}_{n}} + \cdots + G_y \underline{f^{(n)}(x_0)} = 0.\text{[1]}$$

Um sie übersichtlich darstellen zu können, setzen wir

$$\Phi(x) = G(x, f(x)). \tag{16}$$

Dann lauten die obigen Bedingungen

$$\Phi(x_0) = 0, \quad \Phi'(x_0) = 0, \quad \ldots, \quad \Phi^{(n)}(x_0) = 0. \tag{17}$$

Sind also (in einem Punkt mit der Abszisse x_0) die Bedingungen (17) *erfüllt, so ist die Ordnung der Berührung zwischen der Kurve* (15) *und der Kurve $y = f(x)$ mindestens n. Sie ist gleich n, wenn überdies*

$$\Phi^{(n+1)}(x_0) \neq 0 \tag{18}$$

gilt.

243. Schmiegungskurven. Wir nehmen nun an, an Stelle der Kurven (15) sei eine $(n+1)$-parametrige Kurvenschar

$$G(x, y, \underbrace{a, b, \ldots, l}_{n+1}) = 0 \tag{19}$$

gegeben. Dann kann man naturgemäß die Frage stellen, ob man durch geeignete Wahl der Parameter in dieser Schar eine Kurve bestimmen kann, welche die Kurve $y = f(x)$ im Punkt $M_0(x_0, f(x_0))$ von möglichst hoher Ordnung berührt (natürlich hängt die mögliche Ordnung der Berührung von der Schar ab).

[1]) In jeder Gleichung ist diejenige Größe unterstrichen, die sich daraus *eindeutig* ergibt, wenn die vorhergehenden Größen schon bestimmt sind.

Eine solche Kurve heißt *Schmiegungskurve* im Punkt M_0. Genauer müßte man sagen: Schmiegungskurve aus der und der Schar, denn für eine einzelne Kurve (15) ist der Terminus sinnlos.

Zur Bestimmung der Schmiegungskurve führen wir die zu (16) analoge Bezeichnung

$$\Phi(x, a, b, \ldots, l) = G\big(x, f(x), a, b, \ldots, l\big) \tag{*}$$

ein und schreiben eine Reihe von Bedingungen in der Art von (17) auf:

$$\left.\begin{array}{l} \Phi(x_0, a, b, \ldots, l) = 0, \quad \Phi_x(x_0, a, b, \ldots, l) = 0, \ldots, \\ \Phi_{\underbrace{x\ldots x}_{n}}(x_0, a, b, \ldots, l) = 0. \end{array}\right\} \tag{20}$$

Das ist ein System von $n + 1$ Gleichungen mit den $n + 1$ Unbekannten $a, b, \ldots, l$. Vielfach bestimmt dieses System eindeutig ein System von Parameterwerten, und damit hat man eine Schmiegungskurve gefunden, welche mit der gegebenen Kurve eine Berührung mindestens n-ter Ordnung hat.

Ist dabei

$$\Phi_{\underbrace{x\ldots x}_{n+1}}(x_0, a, b, \ldots, l) \neq 0,$$

so hat die Berührung genau die Ordnung n. Diesen Fall sieht man als den Normalfall an.

In den Ausnahmepunkten, in denen überdies

$$\Phi_{\underbrace{x\ldots x}_{n+1}}(x, a, b, \ldots, l) = 0 \tag{21}$$

ist, spricht man von einer *Hyperschmiegung*. Diese Punkte kann man finden, indem man die Gleichungen (20) und (21) zusammen als System von $n + 2$ Gleichungen mit den $n + 2$ Unbekannten $x_0, a, b, \ldots, l$ auffaßt.

Beispiele.

1. Die *Schmiegungsgerade*. Die Gleichung $y = ax + b$ definiert eine zweiparametrige Geradenschar. Daher ist die höchste Ordnung der Berührung, die im allgemeinen Fall zu erwarten ist, gleich 1.

Hier gilt

$$\Phi(x, a, b) = y - ax - b, \quad \Phi_x(x, a, b) = y' - a, \quad \Phi_{xx}(x, a, b) = y'',$$

wenn unter y die Funktion $f(x)$ verstanden wird. Setzen wir zur Bestimmung der Parameter a und b für $x = x_0$ die Größen y, y', y'' gleich 0, so erhalten wir

$$y_0 - ax_0 - b = 0, \qquad y_0' - a = 0.$$

Hieraus folgt $a = y_0'$ und $b = y_0 - y_0' x_0$. Setzen wir dies in die Schargleichung ein, so folgt

$$y = y_0 + y_0'(x - x_0),$$

worin man sofort die Gleichung der Tangente erkennt. *Die Schmiegungsgerade ist also die Tangente.*

Die Berührung ist, wie schon erwähnt, im allgemeinen von erster Ordnung. Sie ist von höherer Ordnung in den einzelnen Punkten, in denen überdies $y_0'' = 0$ ist (etwa in Wendepunkten).

2. Der *Schmiegungskreis* (Kreis hier natürlich immer im Sinne von Kreislinie). Die Gleichung $(x - \xi)^2 + (y - \eta)^2 = R^2$ stellt eine dreiparametrige Kreisschar dar. Die höchste Berührungsordnung ist im allgemeinen 2.

Wie üblich sei $y = f(x)$; wir erhalten

$$\Phi(x, \xi, \eta, R) = (x - \xi)^2 + (y - \eta)^2 - R^2,$$

$$\frac{1}{2} \Phi_x(x, \xi, \eta, R) = x - \xi - (y - \eta)\, y',$$

$$\frac{1}{2} \Phi_{xx}(x, \xi, \eta, R) = 1 + y'^2 + (y - \eta)\, y'';$$

die Parameter ξ, η, R lassen sich aus den Gleichungen

$$(x_0 - \xi)^2 + (y_0 - \eta)^2 = R^2, \quad x_0 - \xi + (y_0 - \eta)\, y_0' = 0, \quad 1 + y_0'^2 + (y_0 - \eta)\, y_0'' = 0$$

bestimmen. Unter der Annahme $y_0'' \neq 0$ erhält man aus den letzten beiden die Koordinaten des Mittelpunktes

$$\xi = x_0 - y_0' \cdot \frac{1 + y_0'^2}{y_0''}, \quad \eta = y_0 + \frac{1 + y_0'^2}{y_0''}, \tag{22}$$

aus der ersten also für den Radius

$$R = \frac{(1 + y_0'^2)^{3/2}}{|y_0''|}. \tag{23}$$

Dadurch ist der Schmiegungskreis bestimmt.

Nach den Ausführungen in Nr. 241 durchsetzt meistens die Tangente die Kurve nicht, dagegen tut es der Schmiegungskreis. Ausnahmen können nur in den Punkten auftreten, in denen die Ordnung der Berührung höher als im Normalfall ist.

244. Ein anderer Zugang zu den Schmiegungskurven. Es seien eine Kurve $y = f(x)$ und eine $(n + 1)$-parametrige Kurvenschar der Gestalt (19) gegeben. Auf der Kurve wählen wir $n + 1$ willkürliche Punkte mit den Abszissen $x_1, x_2, \ldots, x_{n+1}$. Damit eine Kurve der Schar durch diese Punkte hindurchgeht, müssen die $n + 1$ Bedingungen

$$\Phi(x_1, a, b, \ldots, l) = 0, \quad \ldots, \quad \Phi(x_{n+1}, a, b, \ldots, l) = 0$$

erfüllt sein. Vielfach sind dadurch die Parameterwerte eindeutig bestimmt; wir bezeichnen sie mit $\bar{a}, \bar{b}, \ldots, \bar{l}$.

Nun nehmen wir an, daß die Parameterwerte $\bar{a}, \bar{b}, \ldots, \bar{l}$ gegen bestimmte Grenzwerte $a, b, \ldots, l$ streben, wenn diese $n + 1$ Punkte irgendwie gegen einen bestimmten Kurvenpunkt mit der Abszisse x_0 streben. Man kann sagen, daß die durch die genannten Punkte verlaufende Kurve der Schar sich dabei in eine *Grenzkurve* deformiert.

Um diese zu bestimmen, gehen wir folgendermaßen vor. Die Funktion

$$\Phi(x, \bar{a}, \bar{b}, \ldots, \bar{l})$$

verschwindet für $n + 1$ Werte von x, nämlich für $x_1 < x_2 < \cdots < x_{n+1}$. Nach dem Satz von Rolle (vgl. Nr. 111) verschwindet dann die erste Ableitung für n Werte $x_1' < x_2' < \cdots < x_n'$, die zweite für $n - 1$ Werte $x_1'' < x_2'' < \cdots < x_{n-1}'', \ldots,$ die $(n - 1)$-te für zwei Werte $x_1^{(n-1)} < x_2^{(n-1)}$ und schließlich die n-te für einen Wert $x_1^{(n)}$. Alle diese x-Werte liegen zwischen x_1 und x_{n+1}.

Somit gelten die $n + 1$ Gleichungen

$$\Phi(x_1, \bar{a}, \bar{b}, \ldots, \bar{l}) = 0, \quad \Phi_x(x_1', \bar{a}, \bar{b}, \ldots, \bar{l}) = 0,$$

$$\Phi_{xx}(x_1'', \bar{a}, \bar{b}, \ldots, \bar{l}) = 0, \quad \ldots, \quad \Phi_{\underbrace{x\ldots x}_{n}}(x_1^{(n)}, \bar{a}, \bar{b}, \ldots, \bar{l}) = 0.$$

Gilt jetzt gleichzeitig $x_1 \to x_0, x_2 \to x_0, \ldots, x_{n+1} \to x_0$, so gilt $\tilde{a} \to a$, $\tilde{b} \to b, \ldots, \tilde{l} \to l$ und offenbar auch $x_1' \to x_0, x_1'' \to x_0, \ldots, x_1^{(n)} \to x_0$. Durch Grenzübergang in den obigen Gleichungen kommen wir zu dem uns schon bekannten System (20), das die Schmiegungskurve bestimmt.

Existiert also eine Grenzlage für die durch $n+1$ Punkte der gegebenen Kurve verlaufende Kurve der Schar, so ist diese Grenzkurve die Schmiegungskurve.

In diesem Zusammenhang sagt man manchmal (nicht sehr streng, aber anschaulich), die Schmiegungskurve aus einer $(n+1)$-parametrigen Kurvenschar sei „die Kurve, die durch $n+1$ unendlich benachbarte Punkte" der gegebenen Kurve gehe. Insbesondere geht die Tangente durch *zwei* unendlich benachbarte Punkte, der Schmiegungskreis durch *drei*.

§ 4. Die Länge einer ebenen Kurve[1])

245. Hilfssätze. Wir betrachten eine (geschlossene oder nicht geschlossene) ebene Kurve $\widehat{AB}$, die in Parameterdarstellung gegeben sei:

$$x = \varphi(t), \quad y = \psi(t) \quad (t_0 \leqq t \leqq T). \tag{1}$$

Einstweilen setzen wir φ und ψ nur als stetig voraus. Die Kurve besitze keine Mehrfachpunkte, so daß jeder Punkt genau einem Parameterwert t entspricht (mit eventueller Ausnahme der — im Fall einer geschlossenen Kurve — zusammenfallenden Endpunkte).[2]) Unter diesen Annahmen sprechen wir von einer *einfachen stetigen Kurve.*

Um für eine solche Kurve den Begriff der Länge definieren zu können, beginnen wir mit einigen Hilfssätzen. Es sei $t_0 \leqq t' < t'' \leqq T$; den beliebigen Parameterwerten t', t'' mögen die Punkte M', M'' entsprechen.

Lemma 1. *Zu jedem $\delta > 0$ gibt es ein $\eta > 0$ derart, daß im Fall $t'' - t' < \eta$ für die Länge der Sehne die Beziehung $\overline{M'M''} < \delta$ gilt.*

In der Tat gibt es zu δ auf Grund der (gleichmäßigen) Stetigkeit von φ und ψ ein $\eta > 0$ derart, daß für alle Paare t', t'' mit $|t' - t''| < \eta$ zugleich

$$|\varphi(t'') - \varphi(t')| < \frac{\delta}{\sqrt{2}}, \quad |\psi(t'') - \psi(t')| < \frac{\delta}{\sqrt{2}},$$

also

$$\overline{M'M''} = \sqrt{[\varphi(t'') - \varphi(t')]^2 + (\psi(t'') - \psi(t'))]^2} < \delta$$

gilt.

Lemma 2. *Im Fall einer nicht geschlossenen Kurve gibt es zu jedem $\varepsilon > 0$ ein $\delta > 0$ derart, daß sobald $\overline{M'M''} < \delta$ gilt, auch die Differenz der entsprechenden Parameterwerte $t'' - t'$ kleiner als ε wird.*

[1]) Obwohl dieser Problemkreis im Grunde zur Integralrechnung gehört, beginnen wir schon hier mit seiner Behandlung, da wir im nächsten Paragraphen den Begriff der Bogenlänge und ihre Eigenschaften brauchen. Allerdings müssen wir die tatsächliche Berechnung von Bogenlängen auf den Band II verschieben.

[2]) Vgl. die Fußnote auf S. 464.

Den Beweis führen wir indirekt. Wir nehmen an, es gäbe für ein $\varepsilon > 0$ zu jedem $\delta > 0$ zwei Punkte $M'(t')$ und $M''(t'')$ derart, daß $\overline{M'M''} < \delta$, aber $t'' - t' \geqq \varepsilon$ wäre. Es sei $\{\delta_n\}$ eine gegen 0 strebende Folge; für $\delta = \delta_n$ $(n = 1, 2, 3, \ldots)$ gäbe es dann zwei Folgen von Punkten $M'_n(t')$ und $M''_n(t'')$, für die

$$\overline{M'_n M''_n} < \delta_n, \quad \text{aber} \quad t''_n - t'_n \geqq \varepsilon \quad (n = 1, 2, 3, \ldots)$$

wäre. Nach dem Lemma von Bolzano-Weierstrass (vgl. Nr. 41) könnte man ohne Einschränkung der Allgemeinheit annehmen, daß dabei

$$t'_n \to t^*, \quad t''_n \to t^{**}$$

gilt (das läßt sich notfalls durch Übergang zu einer Teilfolge erreichen). Offenbar ist

$$t^{**} - t^* \geqq \varepsilon,$$

also $t^* \neq t^{**}$. Für die entsprechenden Punkte M^* und M^{**} würde aber $\overline{M^*M^{**}} = 0$ gelten; diese Punkte würden also zusammenfallen. Das kann aber nicht sein, da die Kurve weder Mehrfachpunkte hat noch geschlossen ist. Mit diesem Widerspruch ist die Behauptung bewiesen.

Für geschlossene Kurven gilt die Behauptung nicht: Die Sehne $\overline{M'M''}$ kann beliebig klein werden, während t' beliebig nahe bei t_0 und t'' beliebig nahe bei T liegt.

246. Richtung auf einer Kurve. Wir wollen nun annehmen, der Punkt A entspreche dem Parameterwert $t = t_0$, der Punkt B dem Wert $t = T$; den Punkt A nennen wir den *Anfangspunkt*, B den *Endpunkt* der Kurve. Im allgemeinen ordnen wir die Punkte M der Kurve nach wachsenden t, d. h., von zwei von A und B verschiedenen Punkten nennen wir denjenigen den „späteren“ oder „folgenden“ Punkt, der dem größeren Wert des Parameters entspricht. Damit wird auf der Kurve eine *Richtung* oder ein *Richtungssinn* definiert. Formal hängt diese Definition jedoch von der speziellen Parameterdarstellung (1) ab. Wir wollen nun zeigen, daß tatsächlich der Begriff der Richtung nicht von der konkreten Art abhängt, in der die Kurve gegeben ist.

Wir beginnen mit dem einfacheren Fall einer nicht geschlossenen Kurve.

Hat die nicht geschlossene Kurve $\overset{\frown}{AB}$ neben der Darstellung (1) *die* (*ebenfalls doppelpunktfreie*) *Darstellung*

$$x = \varphi^*(u), \quad y = \varphi^*(u) \quad (u_0 \leqq u \leqq U), \tag{1*}$$

wobei φ^* und ψ^* wie oben stetig sind, entspricht ferner dem Wert $u = u_0$ der Punkt A, dem Wert $u = U$ der Punkt B, so bestimmen die Darstellungen (1) *und* (1*) *auf der Kurve ein und dieselbe Richtung.*

Jedem Wert von t entspricht ein Punkt der Kurve, der seinerseits den Wert von u eindeutig bestimmt; umgekehrt entspricht jedem u genau ein bestimmter Wert von t. Somit ist u eine eindeutige Funktion von t, $u = \omega(t)$, welche überdies für zwischen t_0 und T variierendes t jeden ihrer Werte nur einmal annimmt. Insbesondere ist $\omega(t_0) = u_0$ und $\omega(T) = U$.

Nach Lemma 1 entsprechen zwei hinreichend benachbarten Werten von t beliebig benachbarte Kurvenpunkte; dann entsprechen ihnen nach Lemma 2 auch zwei beliebig benachbarte Werte von u; also ist die Funktion $u = \omega(t)$ stetig.

Hieraus kann man schließen, daß diese Funktion monoton wachsend (im engeren Sinne) ist. Würde nämlich für $t_0 < t' < t''$ die Beziehung $u' = \omega(t') > u'' = \omega(t'') > u_0 = \omega(t_0)$ gelten, so gäbe es nach einer bekannten Eigenschaft der stetigen Funktionen (vgl. Nr. 82) zwischen t_0 und t' einen Wert t''', für den $\omega(t''') = u''$ wäre, so daß dieser Wert u'' von der Funktion $u = \omega(t)$ zweimal (für $t = t''$ und $t = t'''$) angenommen würde, entgegen dem oben Bewiesenen.

Damit ist also gezeigt, daß $u = \omega(t)$ mit t wächst; somit ist klar, daß die Anordnung der Punkte nach wachsendem Parameter t der Anordnung dieser Punkte nach wachsendem u völlig gleichwertig ist. Also ist die so definierte Richtung, die man die *Richtung auf der Kurve von A nach B* nennen könnte, ein echter geometrischer Begriff.

Ersetzt man etwa t durch $-t'$ und ordnet die Punkte nach wachsendem t' an, so gelangt man zum Begriff der *Richtung auf der Kurve von B nach A*; sie ergibt sich offenbar auch, wenn man die Punkte nach abnehmenden Parameterwerten t anordnet. Natürlich hängt diese Richtung ebenfalls nicht von der speziellen Wahl des Parameters ab.

Zum Schluß wollen wir noch auf das Problem der Richtung auf einer geschlossenen Kurve eingehen. Auf einer solchen Kurve wählen wir zwei beliebige, von A verschiedene Punkte C und D, denen die Parameterwerte $t = t_1$ bzw. $t = t_2 > t_1$ entsprechen mögen. Somit folgt in der oben mit Hilfe des Parameters t festgelegten Anordnung der Punkt D auf den Punkt C. Man kann zeigen, daß jede Richtung auf der Kurve, die durch eine beliebige, aber diese Anordnung von C und D unverändert lassende Parameterdarstellung festgelegt ist, mit der erstgenannten übereinstimmt.

Entsprechen nämlich den Werten $t = t_0^*$ und $t = T^*$ (mit $t_0 < t_0^* < t_1$ und $t_2 < T^* < T$) die Punkte A^* und B^*, so kann man für den nichtgeschlossenen Bogen A^*B^* ähnlich wie oben schließen. Da aber t_0^* beliebig nahe bei t_0 und T^* beliebig nahe bei T gewählt werden kann, gilt der Schluß für die ganze Kurve.

Somit kann man von der Richtung (oder dem Durchlaufssinn) von A über C und D nach A sprechen, da sie nicht von der Wahl der Parameterdarstellung der Kurve abhängt. Analog läßt sich die Richtung von A über D und C nach A festlegen.

247. Die Länge einer Kurve. Additivität der Bogenlänge. Wir gehen von der Darstellung (1) einer Kurve $\widehat{AB}$ und der durch wachsendes t festgelegten Richtung aus. Auf der Kurve wählen wir eine Reihe von Punkten

$$A = M_0, M_1, M_2, \ldots, M_i, M_{i+1}, \ldots, M_n = B, \tag{2}$$

und zwar so, daß sie in der erwähnten Richtung nacheinander durchlaufen werden. Die entsprechenden Parameterwerte seien

$$t_0 < t_1 < t_2 < \cdots < t_i < t_{i+1} < \cdots < t_n. \tag{3}$$

Verbinden wir je zwei aufeinanderfolgende dieser Punkte durch geradlinige Strecken (Abb. 149), so erhalten wir einen der Kurve $\widehat{AB}$ einbeschriebenen Streckenzug

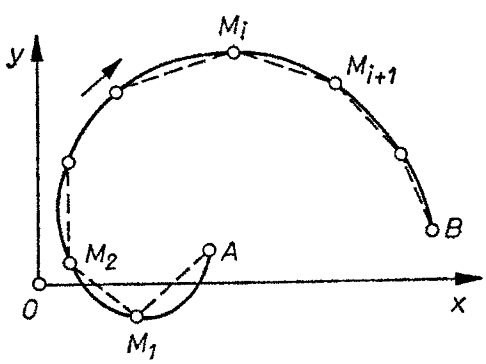

Abb. 149

$M_0M_1 \ldots M_{n-1}M_n$. In Nr. 246 hatten wir die Unabhängigkeit der Richtung von der speziellen Wahl der Parameterdarstellung (1) gezeigt. Das gilt dann natürlich auch für den Streckenzug.

Unter der *Länge der Kurve* $\widehat{AB}$ versteht man nun die obere Grenze S der Menge der Längen p aller möglichen einbeschriebenen Streckenzüge:

$$S = \sup \{p\}.$$

Ist S endlich, so nennt man die Kurve *rektifizierbar*.[1])

Aus dieser Definition folgt, daß *die Länge jedes einer Kurve* $\widehat{AB}$ *einbeschriebenen Streckenzugs höchstens gleich der Länge* S *der Kurve ist*. Insbesondere gilt das auch für die die Endpunkte A und B verbindende Sehne $\overline{AB}$.

Nun wählen wir auf $\widehat{AB}$ einen Punkt C zwischen A und B, der einem Wert $t = \bar{t}$ zwischen t_0 und T entspricht ($t_0 < \bar{t} < T$).

Ist $\widehat{AB}$ *rektifizierbar, so sind auch die Bögen* $\widehat{AC}$ *und* $\widehat{CB}$ *einzeln rektifizierbar*. Umgekehrt folgt aus der Rektifizierbarkeit dieser Bögen die der ganzen Kurve $\widehat{AB}$. Sind S, S', S'' die Längen von $\widehat{AB}, \widehat{AC}, \widehat{CB}$, so gilt

$$S = S' + S''. \tag{4}$$

Zum Beweis setzen wir zunächst die Kurve $\widehat{AB}$ als rektifizierbar voraus und beschreiben ihr beliebige Streckenzüge der Länge p' bzw. p'' ein, die den Bögen $\widehat{AC}$ und $\widehat{CB}$ entsprechen mögen. Diese beiden Streckenzüge bilden einen der Kurve $\widehat{AB}$ einbeschriebenen Streckenzug der Länge $p = p' + p''$. Da $p \leqq S$, also

$$p' + p'' \leqq S \tag{5}$$

gilt, ist offenbar auch $p' \leqq S$ und $p'' \leqq S$. Somit sind die Mengen $\{p'\}$ und $\{p''\}$ durch die (auf Grund der Rektifizierbarkeit von $\widehat{AB}$ endliche) Zahl S beschränkt. Daher haben die Bögen $\widehat{AC}$ und $\widehat{CB}$ die endlichen Längen $S' = \sup \{p'\}$, $S'' = \sup \{p''\}$.

Nach einer Eigenschaft der oberen Grenze (vgl. Nr. 11) gibt es Zahlen p' und p'' in beliebiger Nähe von S' bzw. S''. Aus (5) ergibt sich also durch Grenzübergang

$$S' + S'' \leqq S. \tag{6}$$

Nun seien $\widehat{AC}$ und $\widehat{CB}$ rektifizierbar. Wir beschreiben der Kurve $\widehat{AB}$ einen beliebigen Streckenzug der Länge p ein. Ist C eine Ecke dieses Streckenzugs, so zerfällt dieser unmittelbar in zwei Streckenzüge der Länge p' bzw. p'', die den Bögen $\widehat{AC}$ bzw. $\widehat{CB}$ einbeschrieben sind. Ist C keine Ecke, so nehmen wir diesen Punkt zusätzlich als Ecke hinzu. Dadurch wird die Länge des Streckenzugs nicht kleiner (Abb. 150; Dreiecksungleichung!). Der neue Streckenzug besteht dann wieder aus zwei einzelnen Streckenzügen. In jedem Fall gilt $p \leqq p' + p'' \leqq S' + S''$. Die Menge $\{p\}$ ist nach oben beschränkt (S' und S'', also auch $S' + S''$ sind endlich). Daher ist $\widehat{AB}$ rektifizierbar, und für die Länge gilt

$$S = \sup \{p\} \leqq S' + S''.$$

Vergleicht man diese Ungleichung mit (6), so erhält man die gewünschte Gleichheit (4).

---

[1]) Wir weisen den Leser darauf hin, wie wichtig hier der genaue Begriff der *Richtung* auf der Kurve und auf dem einbeschriebenen Streckenzug ist. Könnte man den Punkt M_i nach Belieben wählen, so wäre S immer gleich ∞.

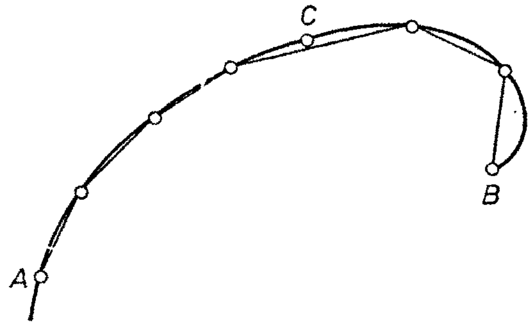

Abb. 150

Die Bogenlänge ist also additiv [vgl. Nr. 21, c)].

Natürlich gilt diese Additivität auch im Fall von mehr als zwei (aber endlich vielen) einzelnen Bögen.

248. Hinreichende Bedingungen für die Rektifizierbarkeit.[1]) **Differential der Bogenlänge.** Bisher untersuchten wir den allgemeinen Fall einer stetigen einfachen Kurve (1). Wir wollen nun hinreichende Bedingungen dafür angeben, daß die Kurve rektifizierbar ist, und insbesondere weitere Eigenschaften der Bogenlänge untersuchen. Zu diesem Zweck wollen wir, wie in diesem Kapitel meist, wieder voraussetzen, daß die Ableitungen $\varphi'(t)$ und $\psi'(t)$ existieren und stetig sind.

Wir zeigen, daß unter diesen Annahmen die Kurve (1) *rektifizierbar ist.*

Beweis. Wir betrachten den Streckenzug mit den Ecken in den Punkten (2), die den Parameterwerten (3) entsprechen. Die Koordinaten von M_i sind

$$x_i = \varphi(t_i), \quad y_i = \psi(t_i) \quad (i = 0, 1, 2, \ldots, n).$$

Dann gilt

$$p = \sum_{i=0}^{n-1} \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}.$$

Nach dem Mittelwertsatz (vgl. Nr. 112) ist aber

$$x_{i+1} - x_i = \varphi(t_{i+1}) - \varphi(t_i) = \varphi'(\tau_i) \cdot (t_{i+1} - t_i),$$
$$y_{i+1} - y_i = \psi(t_{i+1}) - \psi(t_i) = \psi'(\bar{\tau}_i) \cdot (t_{i+1} - t_i)$$
$$(t_i < \tau_i < t_{i+1}; \quad t_i < \bar{\tau}_i < t_{i+1}),$$

also

$$p = \sum_{i=0}^{n-1} \left\{ \sqrt{[\varphi'(\tau_i)]^2 + [\psi'(\bar{\tau}_i)]^2} \cdot (t_{i+1} - t_i) \right\}. \tag{7}$$

Sind L und $\bar{L}$ die größten Werte von $|\varphi'(t)|$ bzw. $|\psi'(t)|$ im Intervall $[t_0, T]$, so ergibt sich aus (7) die Abschätzung

$$p \leqq \sqrt{L^2 + \bar{L}^2} \cdot (T - t_0). \tag{8}$$

Die Menge $\{p\}$ ist nach oben beschränkt; somit hat die Kurve eine endliche Länge S, ist also rektifizierbar.

Da $S = \sup \{p\}$ gilt, folgt aus (8) nebenbei eine Abschätzung für S nach oben:

$$S \leqq \sqrt{L^2 + \bar{L}^2} \cdot (T - t_0), \tag{9}$$

[1]) Allgemeinste (notwendige und hinreichende) Bedingungen für die Rektifizierbarkeit werden in Band III angegeben.

die wir alsbald benutzen. Übrigens brauchen wir auch eine Abschätzung nach unten: Sind l und $\bar{l}$ die kleinsten Werte von $|\varphi'(t)|$ bzw. $|\psi'(t)|$ in $[t_0, T]$, so finden wir aus (7) analog zu (8) die Beziehung $p \geqq \sqrt{l^2 + \bar{l}^2} \cdot (T - t_0)$, um so mehr also

$$S \geqq \sqrt{l^2 + \bar{l}^2} \cdot (T - t_0). \tag{9*}$$

Ändert sich t und damit die Lage des Punktes $M(t)$ auf der Kurve, so wird die Länge des veränderlichen Bogens $\widehat{AM}$ eine Funktion des Parameters t; wir schreiben dafür $S = s(t)$. Nun geben wir t den positiven Zuwachs Δt; dann wandert M längs der Kurve in Richtung auf B in die Lage M' (Abb. 151). Die Größe S erhält einen positiven Zuwachs Δs, der gleich der Länge des Bogens $\widehat{MM'}$ ist (auf Grund der in Nr. 247 bewiesenen Additivität). Die Funktion $s(t)$ ist also wachsend.

Wir betrachten nun an Stelle des Intervalls $[t_0, T]$ das Intervall $[t_0, t_0 + \Delta t]$ und wenden auf den Bogen $\widehat{MM'}$ der Länge Δs die Abschätzungen (9) und (9*) an:

$$\sqrt{l^2 + \bar{l}^2} \cdot \Delta t \leqq \Delta s \leqq \sqrt{L^2 + \bar{L}^2} \cdot \Delta t,$$

wobei unter l und $\bar{l}$ bzw. L und $\bar{L}$ die kleinsten bzw. größten Werte von $|\varphi'(t)|$ und $|\psi'(t)|$ im Intervall $[t, t + \Delta t]$ zu verstehen sind. Hieraus folgt

$$\sqrt{l^2 + \bar{l}^2} \leqq \frac{\Delta s}{\Delta t} \leqq \sqrt{L^2 + \bar{L}^2};$$

da aber auf Grund der Stetigkeit der Ableitungen für $\Delta t \to 0$ die beiden Zahlen l und L gegen $|\varphi'(t)|$, die Zahlen $\bar{l}$ und $\bar{L}$ gegen $|\psi'(t)|$ streben, streben die beiden Wurzeln gegen den gemeinsamen Grenzwert $\sqrt{[\varphi'(t)]^2 + [\psi'(t)]^2}$. Daher strebt $\frac{\Delta s}{\Delta t}$ gegen denselben Grenzwert. Offenbar gilt das auch für $\Delta t < 0$.

Somit ist die veränderliche Bogenlänge $s(t)$ eine differenzierbare Funktion des Parameters t; für die Ableitung nach dem Parameter gilt

$$s'(t) = \lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t} = \sqrt{[\varphi'(t)]^2 + [\psi'(t)]^2}$$

oder, kürzer,

$$s_t = \sqrt{x_t^2 + y_t^2}. \tag{10}$$

Durch Quadrieren und Multiplikation mit dt^2 erhält man die einfache Formel

$$ds^2 = dx^2 + dy^2, \tag{11}$$

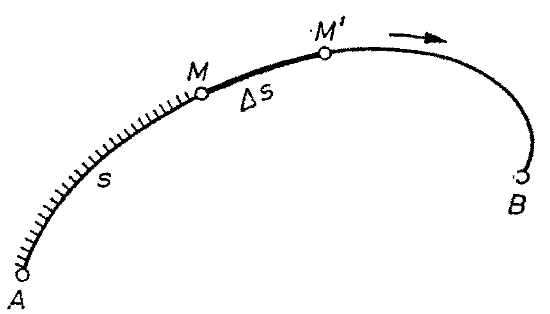

Abb. 151

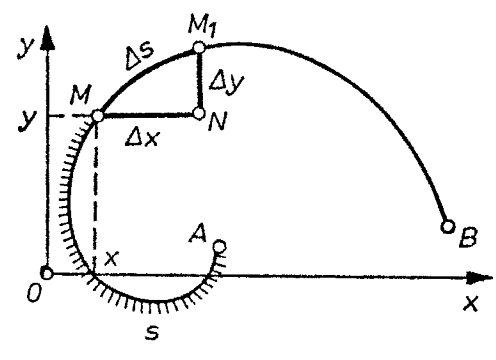

Abb. 152

die überdies eine anschaulich-geometrische Bedeutung hat: Abb. 152 zeigt das (krummlinige) rechtwinklige Dreieck MNM_1; seine Katheten sind die Koordinatenzuwächse von M, also $\overline{MN} = \Delta x$, $\overline{NM_1} = \Delta y$, seine Hypotenuse ist der Bogen $\widehat{MM_1} = \Delta s$, der Zuwachs des Bogens $\widehat{AM} = s$. Es gilt also eine Art pythagoreischer Lehrsatz, zwar nicht für die Zuwächse selbst, aber immerhin für ihre Differentiale.

Es sei noch auf einige Spezialfälle der wichtigen Formel (10) hingewiesen, die verschiedenen Formen der Definitionsgleichung der Kurve entsprechen. Ist diese explizit durch $y = f(x)$ gegeben, so ist x der Parameter; s hängt von x ab, $s = s(x)$, und (10) lautet jetzt

$$s_x = \sqrt{1 + y_x^2}. \tag{10a}$$

Ist die Kurve in Polarkoordinaten $r = g(\theta)$ gegeben, so ist das bekanntlich gleichbedeutend mit der Parameterdarstellung

$$x = r\cos\theta, \quad y = r\sin\theta,$$

wobei θ der Parameter ist. Jetzt ist die Bogenlänge eine Funktion von θ, $s = s(\theta)$. Wegen

$$x_\theta = r_\theta\cos\theta - r\sin\theta, \quad y_\theta = r_\theta\sin\theta + r\cos\theta$$

gilt

$$x_\theta^2 + y_\theta^2 = r_\theta^2 + r^2,$$

so daß (10) die Gestalt

$$s_\theta = \sqrt{r_\theta^2 + r^2} \tag{10b}$$

annimmt.

Oft is es zweckmäßig, als Anfangspunkt A nicht einen der Endpunkte des Bogens, sondern irgendeinen inneren Punkt zu nehmen. Dann muß man natürlich die Bögen, die in Richtung wachsender Parameterwerte verlaufen, positiv, die anderen negativ nehmen. Demgemäß erhält die Bogenlänge ein Vorzeichen. Auch diese vorzeichenbehaftete Bogenlänge s nennen wir einfach Bogenlänge. Die Formeln (10), (11), (10a), (10b) bleiben in allen Fällen gültig.

Wählt man als positive Richtung nicht diejenige nach wachsenden Parameterwerten, wie man das üblicherweise tut, sondern nach abnehmenden, so muß man in den Formeln (10), (10a), (10b) die Wurzel negativ nehmen.

249. Der Bogen als Parameter. Positive Richtung der Tangente. Da der veränderliche Bogen $s = s(t)$ eine stetige monoton wachsende Funktion des Parameters t ist, kann man diesen seinerseits als eindeutige und stetige Funktion von s ansehen: $t = \omega(s)$, wobei s zwischen 0 und der Länge S der ganzen Kurve variiert (vgl. Nr. 83). Setzt man diesen Ausdruck für t in die Gleichung (1) ein, so erhält man die laufenden Koordinaten x und y als Funktionen von s:

$$x = \varphi\big(\omega(s)\big) = \Phi(s), \quad y = \psi\big(\omega(s)\big) = \Psi(s).$$

Zweifellos ist der Bogen s als „krummlinige Abszisse" des Punktes M der natürlichste Parameter zur Bestimmung seiner Lage.

Da man nicht notwendig die Bogenlänge von A aus zu zählen braucht, kann s sowohl positive als auch negative Werte annehmen.

Der Punkt M der Kurve (1) sei nicht singulär, so daß also $s_t = \sqrt{x_t^2 + y_t^2} > 0$ ist; dann existiert nach Nr. 94 für den entsprechenden Wert s (und in seiner Nachbarschaft) die stetige Ableitung

$$t_s = \omega'(s) = \frac{1}{\sqrt{x_t^2 + y_t^2}}.$$

Daher existieren auch die stetigen Ableitungen

$$x_s = \Phi'(s), \quad y_s = \Psi'(s).$$

Nimmt man in (11) alle Differentiale in bezug auf s, so ergibt sich

$$\left(\frac{dx}{ds}\right)^2 + \left(\frac{dy}{ds}\right)^2 = 1. \tag{12}$$

Ist also M ein nichtsingulärer Punkt bei der Parameterdarstellung (1), so ist er auch nichtsingulär beim Übergang zum Parameter s. Ferner liefert Formel (12) folgenden nützlichen Satz:

Es sei M ein nichtsingulärer Punkt einer Kurve. Ist M_1 ein veränderlicher Punkt dieser Kurve und strebt M_1 gegen M, so strebt das Verhältnis der Länge der Sehne $\overline{MM_1}$ und der Länge des Bogens[1]) $\overset{\frown}{MM_1}$ *gegen* 1:

$$\lim_{\overset{\frown}{MM_1} \to 0} \frac{\overline{MM_1}}{\overset{\frown}{MM_1}} = 1. \tag{13}$$

Beweis. Wir nehmen die Bogenlänge als Parameter; der Punkt M entspreche dem Wert s, der Punkt M_1 dem Wert $s + \Delta s$. Die Koordinaten dieser Punkte seien x, y bzw. $x + \Delta x, y + \Delta y$. Dann ist

$$\overset{\frown}{MM_1} = |\Delta s| \quad \text{und} \quad \overline{MM_1} = \sqrt{\Delta x^2 + \Delta y^2},$$

also

$$\frac{\overline{MM_1}}{\overset{\frown}{MM_1}} = \frac{\sqrt{\Delta x^2 + \Delta y^2}}{|\Delta s|} = \sqrt{\left(\frac{\Delta x}{\Delta s}\right)^2 + \left(\frac{\Delta y}{\Delta s}\right)^2}$$

Gehen wir rechts für $\Delta s \to 0$ zur Grenze über, so erhalten wir nach (12) gerade das gewünschte Resultat.

Bisher bestimmten wir die Lage der Tangente an eine Kurve in einem (nichtsingulären) Punkt M durch ihren Richtungskoeffizienten $\tan\alpha$, ohne die beiden entgegengesetzten Tangentenrichtungen zu unterscheiden; für beide hat ja $\tan\alpha$ denselben Wert. Bei einigen Untersuchungen ist es jedoch notwendig, *einen* dieser Richtungssinne festzulegen.

Wir stellen uns vor, auf der Kurve sei ein Anfangspunkt gegeben, von dem aus die Bogenlänge zu rechnen ist, und ebenso die Richtung, in der sie wächst. Als Parameter, der die Lage eines Punktes auf der Kurve bestimmt, nehmen wir die Bogenlänge.

Der Punkt M möge der Bogenlänge s entsprechen. Wir erteilen s den positiven Zuwachs Δs; die Bogenlänge $s + \Delta s$ bestimmt einen Punkt M_1, der von M aus in Rich-

---

[1]) Der Einfachheit halber bedeute in der Formel $\overline{MM_1}$ die Länge der Sehne, $\overset{\frown}{MM_1}$ die Länge des Bogens.

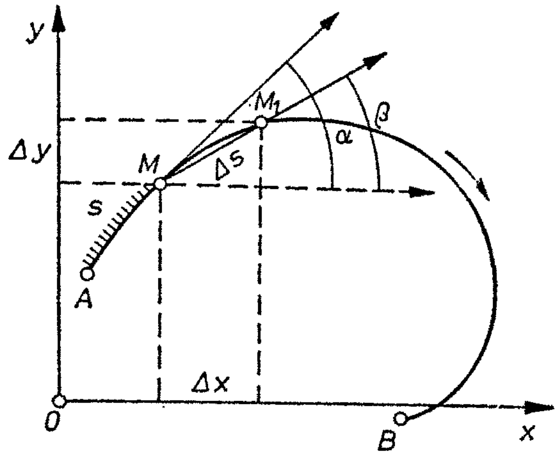

Abb. 153

tung wachsender Bogenlänge liegt. Die Sekante richten wir von M nach M_1, den Winkel, den diese Sekantenrichtung mit der positiven x-Achse bildet, bezeichnen wir mit β. Projiziert man $\overline{MM_1}$ auf die Koordinatenachsen (Abb. 153), so ergibt sich nach bekannten Regeln

$$\mathrm{Proj}_x(\overline{MM_1}) = \Delta x = \overline{MM_1}\cos\beta, \quad \mathrm{Proj}_y(\overline{MM_1}) = \Delta y = \overline{MM_1}\sin\beta,$$

also

$$\cos\beta = \frac{\Delta x}{\overline{MM_1}}, \quad \sin\beta = \frac{\Delta y}{\overline{MM_1}}.$$

Wegen $\widehat{MM_1} = \Delta s$ kann man dafür auch

$$\cos\beta = \frac{\Delta x}{\Delta s}\frac{\widehat{MM_1}}{\overline{MM_1}}, \quad \sin\beta = \frac{\Delta y}{\Delta s}\frac{\widehat{MM_1}}{\overline{MM_1}} \tag{14}$$

schreiben.

Wir nennen diejenige Tangentenrichtung *positiv*, die nach der Seite der wachsenden Bogenlänge zeigt. Genauer: Sie wird als Grenzlage des im Sinne der obigen Ausführungen gerichteten Strahles MM_1 für $\Delta s \to 0$ definiert. Bezeichnet man den Winkel zwischen der positiven Tangentenrichtung und der positiven x-Achse mit α, so erhält man unter Beachtung von (13) aus (14) die Beziehungen

$$\cos\alpha = \frac{\mathrm{d}x}{\mathrm{d}s}, \quad \sin\alpha = \frac{\mathrm{d}y}{\mathrm{d}s}. \tag{15}$$

Diese Formeln bestimmen α bis auf $2k\pi$ (k ganz) genau; sie legen daher tatsächlich eine der beiden Tangentenrichtungen, und zwar die positive, fest.

Bemerkung. Alles in den Nr. 245 bis 249 über ebene Kurven Dargelegte läßt sich ohne wesentliche Änderungen auf den Fall räumlicher Kurven

$$x = \varphi(t), \quad y = \psi(t), \quad z = \chi(t) \quad (t_0 \leqq t \leqq T) \tag{1*}$$

übertragen. Der Begriff der Länge einer Kurve wird ebenso definiert wie in Nr. 247. Haben die Funktionen φ, ψ, χ stetige Ableitungen, so ist die Länge endlich und die Kurve rektifizierbar. Die Länge des veränderlichen Bogens (vom Anfangspunkt bis zum laufenden Punkt, der dem Parameterwert t entspricht), $s = s(t)$, ist nach t differenzierbar, und für die Ableitung nach t gilt

$$s_t = \sqrt{x_t^2 + y_t^2 + z_t^2}. \tag{10*}$$

Hieraus ergibt sich für das Differential der Bogenlänge

$$ds^2 = dx^2 + dy^2 + dz^2. \tag{11*}$$

Falls keine singulären Punkte vorhanden sind (vgl. Nr. 228), kann man zu einer Parameterdarstellung übergehen, bei der die Bogenlänge s der Parameter ist. Schließlich läßt sich der Begriff der positiven Tangentenrichtung einführen, für deren Richtungskosinus die Beziehungen

$$\cos\alpha = \frac{dx}{ds}, \quad \cos\beta = \frac{dy}{ds}, \quad \cos\gamma = \frac{dz}{ds} \tag{15*}$$

gelten.

§ 5. Die Krümmung einer ebenen Kurve

250. Die Krümmung. Es sei eine einfache Kurve

$$x = \varphi(t), \quad y = \psi(t) \quad (t_0 \leqq t \leqq T) \tag{1}$$

gegeben; diesmal werden φ und ψ jedoch als *zweimal* stetig differenzierbar vorausgesetzt. Wir wollen einen Bogen dieser Kurve betrachten, der keine singulären Punkte enthält.

Ziehen wir in jedem Punkt die Tangente (etwa in positiver Richtung), so wird sie sich bei einer Verschiebung des Berührungspunktes drehen, da die Kurve irgendwie „gekrümmt" ist. Dadurch unterscheidet sich eine Kurve wesentlich von einer Geraden, bei der die (mit ihr zusammenfallende) Tangente in allen Punkten dieselbe Richtung hat.

Ein wichtiges Element, das den Verlauf der Kurve charakterisiert, ist nun der „Grad des Gekrümmtseins" oder die „Krümmung", die sie in verschiedenen Punkten aufweist. Diese Krümmung läßt sich durch eine Zahl ausdrücken.

Es sei $\widehat{MM_1}$ (Abb. 154) ein Bogen der Kurve; wir betrachten die (positiv gerichteten) Tangenten MT und M_1T_1 in den Endpunkten dieses Bogens.

Naturgemäß charakterisiert man die Krümmung der Kurve durch den Winkel, um den sich die Tangente dreht, und zwar bezogen auf die Einheit der Bogenlänge, also durch den Quotienten $\frac{\omega}{\sigma}$, wobei ω im Bogenmaß und σ in der gewählten Längen-

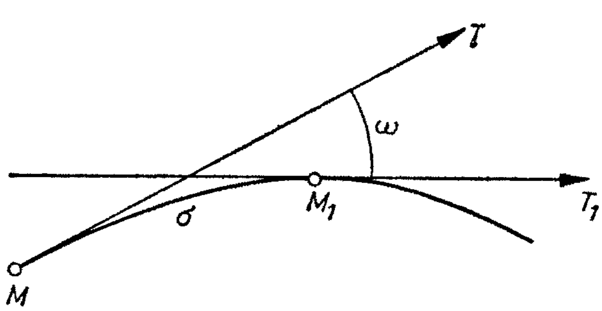

Abb. 154

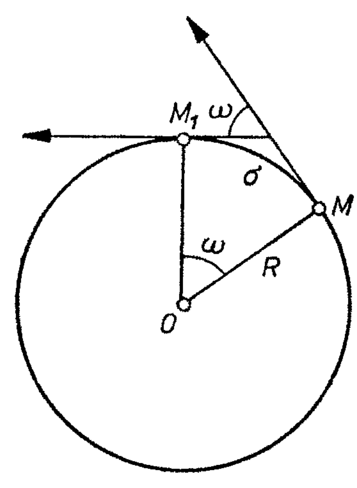

Abb. 155

einheit gemessen wird. Dieses Verhältnis nennt man die *mittlere Krümmung* des Kurvenbogens.

Im allgemeinen wird die mittlere Krümmung auf den einzelnen Kurvenabschnitten verschieden sein. Übrigens existiert eine (einzige) Kurve, für die die mittlere Krümmung überall konstant ist, nämlich der Kreis.[1]) Für den Kreis gilt nämlich (Abb. 155)

$$\frac{\omega}{\sigma} = \frac{\omega}{R\omega} = \frac{1}{R},$$

unabhängig von dem einzelnen Bogenstück des Kreises.

Von der mittleren Krümmung eines Bogens $\widehat{MM_1}$ gehen wir nun zu der Krümmung in einem Punkt über.

Unter der *Krümmung einer Kurve in einem Punkt M* versteht man den Grenzwert, gegen den die mittlere Krümmung des Bogens $\widehat{MM_1}$ strebt, wenn M_1 längs der Kurve gegen M strebt. (Die Existenz dieses Grenzwertes setzen wir einstweilen voraus).

Für die Krümmung k einer Kurve in einem gegebenen Punkt gilt also

$$k = \lim_{\sigma \to 0} \frac{\omega}{\sigma}.$$

Für den Kreis ist offenbar $k = \frac{1}{R}$, d. h., *die Krümmung des Kreises ist das Reziproke seines Radius.*

Bemerkung. Der Begriff der mittleren Krümmung und der der Krümmung in einem gegebenen Punkt sind dem Begriff der mittleren Geschwindigkeit bzw. dem der Geschwindigkeit eines sich bewegenden Punktes in einem gegebenen Zeitpunkt völlig analog. Man kann sagen, die mittlere Krümmung chrakterisiere die mittlere Geschwindigkeit der Änderung der Tangentenrichtung, die Krümmung in einem Punkt dagegen die Momentangeschwindigkeit der Änderung dieser Richtung in dem betreffenden Punkt.

Wir wollen nun einen analytischen Ausdruck für die Krümmung herleiten, auf Grund dessen wir sie berechnen können. Dabei nehmen wir an, die Kurve sei in Parameterdarstellung gegeben.

Zunächst sei die Bogenlänge der Parameter. Wie wir aus Nr. 249 wissen, ist diese Darstellung immer möglich, wenn der betrachtete Kurvenbogen keine singulären Punkte enthält.

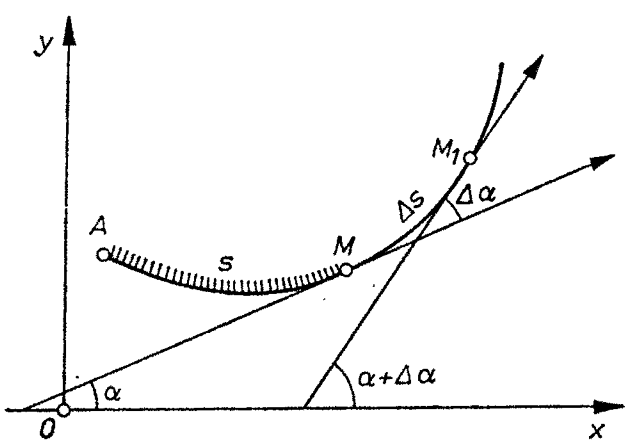

Abb. 156

[1]) abgesehen natürlich von der *Geraden*, wo sie überall 0 ist.

Auf einem solchen Teilstück der Kurve wählen wir den (also sicher nicht singulären) Punkt M, dem der Wert s der Bogenlänge entsprechen möge. Erteilen wir s den positiven Zuwachs Δs, so gelangen wir zum Punkt $M_1(s + \Delta s)$; vgl. Abb. 156. Der Zuwachs $\Delta\alpha$ des Neigungswinkels der Tangente beim Übergang von M zu M_1 liefert den Winkel ω zwischen den beiden Tangenten: $\omega = \Delta\alpha$.

Wegen $\sigma = \Delta s$ ist also die mittlere Krümmung gleich $\dfrac{\Delta\alpha}{\Delta s}$.

Nun strebe $\widehat{MM_1} = \Delta s$ gegen 0. Dann ergibt sich für die Krümmung der Kurve in M die Beziehung

$$k = \lim_{\Delta s \to 0} \frac{\Delta\alpha}{\Delta s} = \frac{\mathrm{d}\alpha}{\mathrm{d}s}. \tag{2}$$

Übrigens ist wichtig, daß diese Formel nur bis aufs Vorzeichen richtig ist, da die Krümmung nach unserer Definition eine nichtnegative Zahl ist. Der Ausdruck rechts kann aber auch negativ werden.

Sowohl $\Delta\alpha$ als auch Δs können negativ sein, so daß man eigentlich $\omega = |\Delta\alpha|$, $\sigma = |\Delta s|$ und schließlich

$$k = \left|\frac{\mathrm{d}\alpha}{\mathrm{d}s}\right|$$

schreiben müßte. Darauf ist auch im folgenden zu achten.

Um Formel (2) in einer Gestalt zu erhalten, die zur unmittelbaren Berechnung geeignet ist (und um die Existenz der Krümmung selbst zu beweisen), gehen wir zu einer beliebigen Parameterdarstellung der Kurve (1) über.

Da $M(t)$ kein singulärer Punkt, somit $x_t^2 + y_t^2 > 0$ ist, können wir ohne Beschränkung der Allgemeinheit $x_t = \varphi'(t) \neq 0$ annehmen.

Nun schreiben wir Formel (2) anders:

$$k = \frac{\mathrm{d}\alpha}{\mathrm{d}s} = \frac{\dfrac{\mathrm{d}\alpha}{\mathrm{d}t}}{\dfrac{\mathrm{d}s}{\mathrm{d}t}} = \frac{\alpha_t}{s_t}. \tag{3}$$

Da $s_t = \sqrt{x_t^2 + y_t^2}$ gilt [vgl. Nr. 148, Formel (10)], braucht man nur noch α_t zu bestimmen. Wegen $\tan\alpha = \dfrac{y_t}{x_t}$ und $\alpha = \arctan\dfrac{y_t}{x_t}$ [vgl. Nr. 106, Formel (11)] ist also

$$\alpha_t = \frac{1}{1 + \left(\dfrac{y_t}{x_t}\right)^2} \frac{x_t y_{tt} - x_{tt} y_t}{x_t^2} = \frac{x_t y_{tt} - x_{tt} y_t}{x_t^2 + y_t^2}. \tag{4}$$

Setzen wir diese Ausdrücke für α_t und s_t in (3) ein, so erhalten wir schließlich

$$k = \frac{x_t y_{tt} - x_{tt} y_t}{(x_t^2 + y_t^2)^{3/2}}. \tag{5}$$

Diese Formel eignet sich vorzüglich zur direkten Berechnung; denn die darin vorkommenden Ableitungen ergeben sich sofort aus der Parameterdarstellung.

Ist die Kurve in der expliziten Form $y = f(x)$ gegeben, so lautet diese Formel

$$k = \frac{y_{xx}}{(1 + y_x^2)^{3/2}}. \tag{5a}$$

Für eine Kurve $r = g(\theta)$ in Polarkoordinaten schließlich geht man wie üblich zur Parameterdarstellung in kartesischen Koordinaten mit θ als Parameter über. Dann erhält man mit Hilfe von (5) die Formel

$$k = \frac{r^2 + 2r_\theta^2 - rr_{\theta\theta}}{(r^2 + r_\theta^2)^{3/2}}. \tag{5b}$$

251. Krümmungskreis und Krümmungsradius. In vielen Untersuchungen ist es zweckmäßig, die Kurve in der Nähe des betrachteten Punktes angenähert durch einen Kreis zu ersetzen, der dieselbe Krümmung hat wie die Kurve in diesem Punkt.

Unter dem *Krümmungskreis* einer Kurve in einem gegebenen Kurvenpunkt M verstehen wir den Kreis, der

a) die Kurve in M berührt,

b) in der Nähe des Punktes nach derselben Seite konvex ist wie die Kurve,

c) dieselbe Krümmung hat wie die Kurve in M (Abb. 157).

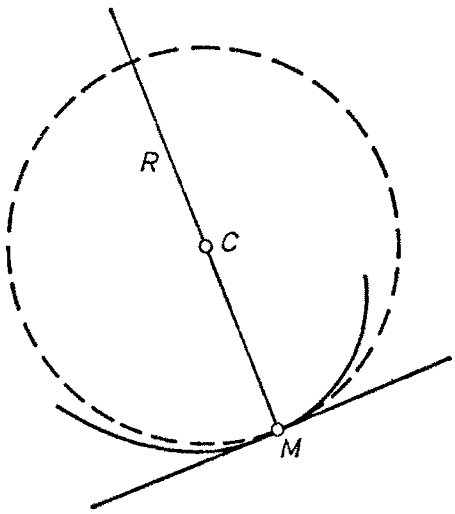

Abb. 157

Der Mittelpunkt des Krümmungskreises wird *Krümmungsmittelpunkt*, sein Radius *Krümmungsradius* (der Kurve in dem gegebenen Punkt) genannt.

Aus der Definition des Krümmungskreises folgt, daß der Krümmungsmittelpunkt immer auf der Kurvennormalen in dem betrachteten Punkt, und zwar nach der konkaven (d. h. auf der zur Konvexitätsrichtung entgegengesetzten) Seite der Kurve liegt. Ist die Krümmung der Kurve in einem gegebenen Punkt gleich k, so gilt für den Krümmungsradius R offenbar $R = \frac{1}{k}$, da nach Nr. 250 für den Kreis $k = \frac{1}{R}$ ist.

Wenn wir die in Nr. 250 für die Krümmung hergeleiteten Ausdrücke benutzen, so erhalten wir für den Krümmungsradius die Formeln

$$R = \frac{ds}{d\alpha}, \tag{6}$$

$$R = \frac{(x_t^2 + y_t^2)^{3/2}}{x_t y_{tt} - x_{tt} y_t}, \tag{7}$$

$$R = \frac{(1 + y_x^2)^{3/2}}{y_{xx}}, \tag{7a}$$

$$R = \frac{(r^2 + r_\theta^2)^{3/2}}{r^2 + 2r_\theta^2 - r r_{\theta\theta}}, \tag{7b}$$

die je nach der Art der Kurvengleichung anzuwenden sind.

Aus allen diesen Formeln ergibt sich der Krümmungsradius mit dem Vorzeichen, wie oben die Krümmung. Hier wollen wir aber dieses Vorzeichen nicht vernachlässigen, sondern seine geometrische Bedeutung erläutern.

Zu diesem Zweck führen wir den Begriff der positiven Normalenrichtung ein. Wir haben schon in Nr. 249 auseinandergesetzt, daß auf der Tangente diejenige Richtung als positiv bezeichnet wird, die nach der Seite der wachsenden Bogenlänge weist. Auf der Normalen wählen wir diejenige Richtung als *positiv*, die in bezug auf die positive Tangentenrichtung ebenso orientiert ist wie die y-Achse zur x-Achse. Beispielsweise muß bei der üblichen Anordnung der Achsen die Normale mit der positiven Tangentenrichtung den Winkel $+\frac{\pi}{2}$ entgegen dem Uhrzeigersinn bilden.

Sieht man jetzt den Krümmungsradius $R = \overline{MC}$ als gerichtete Strecke an, die auf der Normalen liegt, so gibt man ihm naturgemäß das positive Vorzeichen, wenn er in die positive Normalenrichtung weist, anderenfalls das Minuszeichen. In Abb. 158 ist der Krümmungsradius für die Kurve (I) positiv, für die Kurve (II) negativ.

Wir wollen uns nun davon überzeugen, daß das Vorzeichen des Krümmungsradius, das sich nach jeder der obigen Formeln ergibt, tatsächlich dieser Definition entspricht. Dabei ist es jedoch wichtig zu betonen, daß in allen Fällen die positive Richtung dem wachsenden Parameter (t, x bzw. θ) entspricht.

Am einfachsten überzeugt man sich im Fall der explizit gegebenen Kurve. Hier (Abb. 158) ist die Tangente nach rechts gerichtet, daher die Normale nach oben. Ist $y_{xx} > 0$ sowohl im betrachteten Punkt als auch — aus Stetigkeitsgründen — in seiner Umgebung, so ist die Kurve nach unten konvex (vgl. Nr. 143) und R positiv. So ergibt sich R auch aus Formel (7a).

Für $y_{xx} < 0$ ist die Kurve nach oben konvex, R negativ, in völliger Übereinstimmung mit (7a).

Dasselbe kann man bei den übrigen Formeln zeigen.

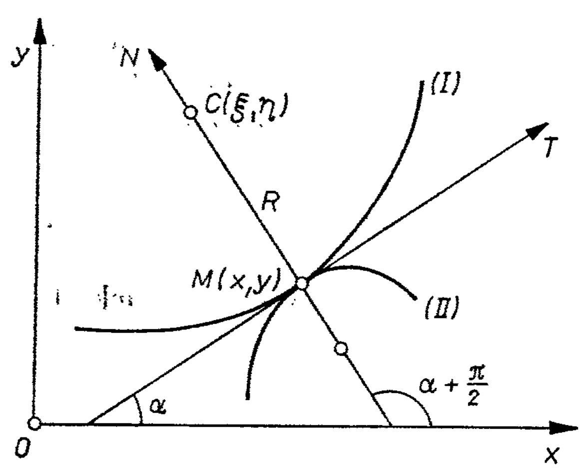

Abb. 158

252. Beispiele.

1. Die *Kettenlinie* $y = a \cosh \frac{x}{a}$ (Abb. 41, S. 193). Hier ist (vgl. Nr. 99, Beispiel 28)

$$\sqrt{1 + y_x^2} = \cosh \frac{x}{a} = \frac{y}{a} \quad \text{und} \quad y_{xx} = \frac{1}{a} \cosh \frac{x}{a} = \frac{y}{a^2}.$$

Daher ist nach (7a)

$$R = \frac{\left(\frac{y}{a}\right)^3}{\frac{y}{a^2}} = \frac{y^2}{a}.$$

Wie man leicht sieht, erhält man denselben Ausdruck für das Normalenstück $n = \overline{MN}$, so daß man den Krümmungsmittelpunkt C folgendermaßen konstruieren kann: Das Normalenstück $\overline{MN}$ (vgl. Abb. 41, S. 193) muß man auf der Normalen um sich selbst in positiver Richtung verschieben.

2. Die *Astroide* $x^{2/3} + y^{2/3} = a^{2/3}$ (Abb. 116, S. 465). Die Ableitungen y_x und y_{xx} lassen sich auch ohne Auflösen der Gleichung durch Differentiation der impliziten Funktionen bestimmen:

$$x^{-1/3} + y^{-1/3}y' = 0 \quad \text{oder} \quad x^{1/3}y' + y^{1/3} = 0, \quad \text{also} \quad y' = -\left(\frac{y}{x}\right)^{1/3}$$

ferner

$$\frac{1}{3} x^{-2/3}y' + \frac{1}{3} y^{-2/3}y' + x^{1/3}y'' = 0, \quad \text{also} \quad y'' = -\frac{a^{2/3}}{3xy^{2/3}}\, y' = \frac{a^{2/3}}{3x^{4/3}y^{1/3}}.$$

Setzt man y' und y'' in Formel (7a) ein, so ergibt sich $R = 3(axy)^{1/3}$.

3. Die *Zykloide* $x = a(t - \sin t)$, $y = a(1 - \cos t)$ (Abb. 118, S. 467). Wegen $\alpha = \frac{\pi}{2} - \frac{t}{2}$ (vgl. Nr. 231, Beispiel 4) ist $d\alpha = -\frac{1}{2}\, dt$; ferner gilt

$$x_t = a(1 - \cos t), \quad y_t = a \sin t, \quad x_t^2 + y_t^2 = 4a^2 \sin^2 \frac{t}{2},$$

also

$$s_t = \sqrt{x_t^2 + y_t^2} = 2a \sin \frac{t}{2} \quad \text{und} \quad ds = 2a \sin \frac{t}{2}\, dt.$$

Hier benutzen wir zur Berechnung von R die Formel (6):

$$R = \frac{ds}{d\alpha} = \frac{2a \sin \frac{t}{2}\, dt}{-\frac{1}{2}\, dt} = -4a \sin \frac{t}{2}.$$

Erinnert man sich des in Nr. 231, Beispiel 4, hergeleiteten Ausdrucks für den Normalenabschnitt n, so ergibt sich $R = -2n$. Hieraus folgt die aus Abb. 118 ersichtliche Konstruktion des Krümmungsmittelpunktes C.

4. Die *Kreisevolvente* $x = a(\cos t + t \sin t)$, $y = a(\sin t - t \cos t)$ (Abb. 121, S. 470). Hier gilt nach Nr. 231, Beispiel 6, offenbar $\alpha = t$, also $d\alpha = dt$. Andererseits ist

$$x_t = at \cos t, \quad y_t = at \sin t, \quad x_t^2 + y_t^2 = a^2t^2,$$

also

$$s_t = at, \quad ds = at \cdot dt.$$

Hiernach ergibt sich einfach

$$R = \frac{ds}{d\alpha} = at = \overline{MB}.$$

Somit ist der Berührungspunkt B (der Punkt, in dem sich der Faden vom Kreis löst) Krümmungsmittelpunkt für die Bahnkurve des Fadenendes M. *Der geometrische Ort der Krümmungsmittelpunkte der Kurve ist der ursprüngliche Kreis.*

Hier stoßen wir auf einen Spezialfall eines Sachverhaltes, den wir in Nr. 255 allgemein untersuchen werden.

5. Die *logarithmische Spirale* $r = a\,e^{m\theta}$ (Abb. 134, S. 486). Hier ist $r_\theta = mr$, $r_{\theta\theta} = m^2 r$. Setzt man das in (7b) ein, so ergibt sich

$$R = \frac{(r^2 + m^2r^2)^{3/2}}{r^2 + 2m^2r^2 - m^2r^2} = r\sqrt{1 + m^2}.$$

Nun ist $m = \cot\omega$ (vgl. Nr. 233, Beispiel 3), so daß R in der Gestalt

$$R = \frac{r}{\sin\omega}$$

geschrieben werden kann. Dann ersieht man unmittelbar aus Abb. 134, daß der Polarenabschnitt der Normalen $n_p = \overline{NM}$ ist. Krümmungsmittelpunkt ist also der Punkt N. Das liefert ein einfaches Verfahren zur Konstruktion des Krümmungsmittelpunktes der logarithmischen Spirale.

6. Die *Kardioide* $r = a\,(1 + \cos\theta)$ (Abb. 135, S. 486). Hier ist $r_\theta = -a\sin\theta$, $r_{\theta\theta} = -a\cos\theta$. Man rechnet leicht

$$r^2 + r_\theta^2 = 4a^2\cos^2\frac{\theta}{2}$$

aus, es bleibt noch $r_\theta^2 - rr_{\theta\theta} = a^2(1 + \cos\theta) = 2a^2\cos^2\frac{\theta}{2}$ zu berechnen, und es ergibt sich sogleich nach (7a)

$$R = \frac{4}{3}a\cos\frac{\theta}{2}.$$

Vergleicht man das mit dem Ausdruck für den Polarenabschnitt der Normalen für die Kardioide (vgl. Nr. 233, Beispiel 4), so folgt

$$R = \frac{2}{3}n_p.$$

7. Die *Lemniskate* $r^2 = 2a^2\cos 2\theta$ (Abb. 126, S. 474). Wir haben in Nr. 233, Beispiel 5, gesehen, daß hier $\alpha = 3\theta + \frac{\pi}{2}$ ist, also $d\alpha = 3d\theta$. Dann ergibt sich nach Formel (6) sofort

$$R = \frac{ds}{d\alpha} = \frac{1}{3}s_\theta = \frac{1}{3}\sqrt{r^2 + r_\theta^2} = \frac{1}{3}n_p = \frac{2a^2}{3r}.$$

Da wir die Normale an die Lemniskate schon konstruieren können, ergibt sich hieraus auch ein Verfahren zur Konstruktion des Krümmungsmittelpunktes.

8. Die *Parabel* $y^2 = 2px$. Wir differenzieren implizit und finden nacheinander

$$yy_x = p, \quad yy_{xx} + y_x^2 = 0, \quad \text{also} \quad y^3y_{xx} = -p^2.$$

Jetzt ergibt sich nach Formel (7a)

$$R = \frac{(1 + y_x^2)^{3/2}}{y_{xx}} = \frac{[y^2 + (yy_x)^2]^{3/2}}{y^3y_{xx}} = \frac{(y^2 + p^2)^{3/2}}{-p^2} \qquad (y > 0).$$

Da nach Nr. 231, Beispiel 1, für den Normalenabschnitt $n = \sqrt{y^2 + p^2}$ gilt, ist

$$R = -\frac{n^3}{p^2}.$$

6. *Ellipse* und *Hyperbel:* $\frac{x^2}{a^2} \pm \frac{y^2}{b^2} = 1$. Durch Differenzieren ergibt sich

$$\frac{x}{a^2} \pm \frac{y y_x}{b^2} = 0, \quad \text{also} \quad y y_x = \mp \frac{b^2 x}{a^2};$$

ferner

$$y y_{xx} = \mp \frac{b^2}{a^2} - y_x^2 \quad \text{oder} \quad y^3 y_{xx} = -\frac{b^4}{a^2}\left(\frac{x^2}{a^2} \pm \frac{y^2}{b^2}\right) = -\frac{b^4}{a^2}.$$

Wie oben folgt hieraus

$$R = -\frac{(b^4 x^2 + a^4 y^2)^{3/2}}{a^4 b^4} \qquad (y > 0).$$

Nach Nr. 231, Beispiel 2, gilt für den Normalenabschnitt $n = \frac{\sqrt{b^4 x^2 + a^4 y^2}}{a^2}$; also ist

$$R = -\frac{a^2}{b^4} n^3.$$

Bekanntlich gilt sowohl bei der Ellipse als auch bei der Hyperbel für den Halbparameter p die Beziehung $p = \frac{b^2}{a}$. Daher ergibt sich auch hier für R derselbe Ausdruck wie bei der Parabel.

Für alle drei Kegelschnitte ist der Krümmungsradius proportional der dritten Potenz des Normalenabschnittes.

10. Zum Schluß bringen wir noch einige Ausführungen über ein praktisches Problem, bei dem gerade die Änderung der Krümmung längs der Kurve wesentlich benutzt wird, nämlich über sogenannte Übergangskurven bei der Anlage von Eisenbahnschienen in Kurven.

In der Mechanik wird gezeigt, daß bei der Bewegung eines Massepunktes längs einer Kurve eine Zentrifugalkraft

$$F = \frac{m v^2}{R}$$

auftritt; dabei ist m die Masse des Punktes, v seine Geschwindigkeit und R der Krümmungsradius in dem betrachteten Punkt.

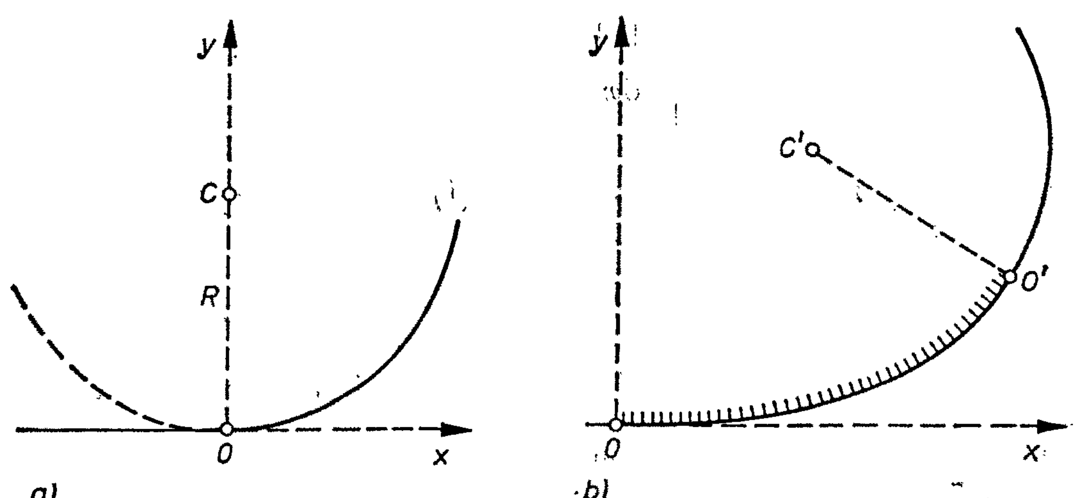

Abb. 159

Würde der geradlinige Teil der Eisenbahnschiene unmittelbar an den krummlinigen, kreisbogenförmigen Teil anschließen (Abb. 159a), so würde beim Übergang zu diesem Teil die Zentrifugalkraft plötzlich auftreten und dabei einen heftigen, starken Stoß verursachen, der für das rollende Material und für den Oberbau schädlich wäre. Um das zu vermeiden, verbindet man den geradlinigen Teil mit dem krummlinigen durch eine sogenannte Übergangskurve (Abb. 159b). Längs dieser Kurve ändert sich der Krümmungsradius allmählich, er nimmt von ∞ (dem Wert, den er am Ende des geradlinigen Stückes hat) bis zum Radius des Kreises (dem Wert zu Beginn des kreisförmigen Stückes) allmählich ab, und dementsprechend wächst auch die Zentrifugalkraft allmählich.

Als Übergangskurve nimmt man meist die kubische Parabel $y = \frac{x^3}{6q}$. Dann ist offenbar

$$y' = \frac{x^2}{2q}, \quad y'' = \frac{x}{q},$$

und für den Krümmungsradius erhält man

$$R = \frac{q}{x}\left(1 + \frac{x^4}{4q^2}\right)^{3/2}$$

Für $x = 0$ gilt $y' = 0$ und $R = \infty$, unsere Kurve berührt im Ursprung die x-Achse und hat die Krümmung 0.[1])

Manchmal wird auch ein Lemniskatenbogen als Übergangskurve gewählt.

253. Die Koordinaten des Krümmungsmittelpunktes. Wir wollen nun Formeln für die Koordinaten des Krümmungsmittelpunktes aufstellen; die Koordinaten des Kurvenpunktes M bezeichnen wir mit x, y, die des entsprechenden Krümmungsmittelpunktes C mit ξ, η.

Der Krümmungsradius $R = \overline{MC}$ (Abb. 158, S. 524) liegt auf einem Strahl der gerichteten Normalen, der mit der x-Achse den Winkel $\alpha + \frac{\pi}{2}$ bildet. Für die Projektionen von $\overline{MC}$ auf die x- bzw. y-Achse gilt

$$\xi - x = R\cos\left(\alpha + \frac{\pi}{2}\right) = -R\sin\alpha, \quad \eta - y = R\sin\left(\alpha + \frac{\pi}{2}\right) = R\cos\alpha.$$

Hieraus ergibt sich

$$\xi = x - R\sin\alpha, \quad \eta = y + R\cos\alpha. \tag{8}$$

Unter Benutzung der in Nr. 251, Formel (6), und Nr. 249, Formel (15), hergeleiteten Beziehungen $R = \frac{\mathrm{d}s}{\mathrm{d}\alpha}$, $\cos\alpha = \frac{\mathrm{d}x}{\mathrm{d}s}$, $\sin\alpha = \frac{\mathrm{d}y}{\mathrm{d}s}$ folgt für die Koordinaten des Krümmungsmittelpunktes schließlich

$$\xi = x - \frac{\mathrm{d}y}{\mathrm{d}\alpha}, \quad \eta = y + \frac{\mathrm{d}x}{\mathrm{d}\alpha}. \tag{9}$$

Ist die Kurve in der Parameterdarstellung (1) gegeben, so ergibt sich aus (9), wenn man den Ausdruck für α_t benutzt,

$$\xi = x - \frac{x_t^2 + y_t^2}{x_t y_{tt} - x_{tt} y_t}\, y_t, \quad \eta = y + \frac{x_t^2 + y_t^2}{x_t y_{tt} - x_{tt} y_t}\, x_t. \tag{10}$$

Hier sind also ξ und η Funktionen desselben Parameters t wie x und y.

[1]) Mit Hilfe der Differentialrechnung (vgl. Nr. 134, 135) läßt sich leicht zeigen, daß R nur bis $x = 0{,}946\sqrt{q}$ abnimmt, wo sich der Minimumwert $R = 1{,}390\sqrt{q}$ ergibt. Nur dieser Teil der Kurve wird in der Praxis benutzt.

Ist die Kurve in der expliziten Form $y = f(x)$ gegeben, so gehen die Formeln (10) über in

$$\xi = x - \frac{1 + y_x^2}{y_{xx}} y_x, \quad \eta = y + \frac{1 + y_x^2}{y_{xx}}. \tag{10a}$$

Die Formeln (10) sind auch anwendbar, wenn die Kurve in Polarkoordinaten $r = g(\theta)$ gegeben ist; dabei ist, wie üblich, θ als Parameter zu nehmen.

Vergleichen wir (10a) mit den Formeln für den Grenzpunkt auf der Normalen, die wir bei der Lösung der Aufgabe in Nr. 137 (Abb. 62, S. 266) erhalten hatten, so finden wir, daß dieser Punkt gerade der Krümmungsmittelpunkt ist.

Ein noch wichtigeres Resultat erhält man durch Vergleich der Formeln (10a) und (7a) mit den Formeln (22) und (23) aus Nr. 243: *Der Krümmungskreis einer Kurve in einem gegebenen Punkt ist nichts anderes als der Schmiegungskreis.* Anders ausgedrückt (Nr. 244), *der Krümmungskreis ist die Grenzlage des Kreises, der durch drei Kurvenpunkte geht, wenn diese drei Punkte mit dem gegebenen Punkt zusammenfallen.*

Dieses Ergebnis ließ sich natürlich voraussehen: Im Fall einer Berührung zweiter Ordnung zwischen gegebener Kurve und Kreis stimmen die Ordinate y und die beiden Ableitungen y_x und y_{xx} in dem betreffenden Punkt für Kreis und Kurve überein, so daß in diesem Punkt auch die Konvexitätsrichtungen und die Krümmungen übereinstimmen, die ja beide nur von diesen Ableitungen abhängen.

254. Evolute und Evolvente. Abwicklung der Evolute. Bewegt sich ein Punkt $M(x, y)$ längs einer gegebenen Kurve, so beschreibt der entsprechende Krümmungsmittelpunkt $C(\xi, \eta)$ im allgemeinen ebenfalls eine Kurve. Den geometrischen Ort der Krümmungsmittelpunkte einer gegebenen Kurve nennt man ihre *Krümmungsmittelpunktskurve* oder *Evolute.* Umgekehrt nennt man die Ausgangskurve in bezug auf ihre Evolute die *Evolvente* dieser Kurve.

Die Formeln (10) bzw. (10a) aus Nr. 253 für die Koordinaten ξ, η des Krümmungsmittelpunktes C, die den Parameter t (bzw. x) enthalten, können also als Parameterdarstellung der Evolute angesehen werden. Manchmal eliminiert man den Parameter und erhält dadurch die Evolutengleichung in der impliziten Form

$$F(\xi, \eta) = 0.$$

Beispiele.

1. Man bestimme die Evolute der Parabel $y^2 = 2px$. Unter Benutzung der Resultate aus Nr. 252, Beispiel 8, ergibt sich

$$y y_x = p, \qquad y^3 y_{xx} = -p^2,$$

so daß nach Formel (10a) für die Koordinaten des Krümmungsmittelpunktes

$$\xi = x - y y_x \frac{y^2 + (y y_x)^2}{y^3 y_{xx}} = x + \frac{y^2 + p^2}{p} = 3x + p = \frac{3y^2}{2p} = p,$$

$$\eta = y + y \frac{y^2 + (y y_x)^2}{y^3 y_{xx}} = y - \frac{y}{p^2}(y^2 + p^2) = -\frac{y^3}{p^2}$$

folgt. Also lautet die Parametergleichung der Evolute der Parabel (y ist Parameter)

$$\xi = \frac{3y^2}{2p} + p, \qquad \eta = -\frac{y^3}{p^2}.$$

Durch Elimination von y erhält man $y^2 = \frac{2p}{3}(\xi - p)$, $y^3 = -p^2\eta$ und hieraus schließlich

$$\eta^2 = \frac{8}{27p}(\xi - p)^3.$$

Die Evolute der Parabel ist also eine semikubische Parabel (Abb. 160).

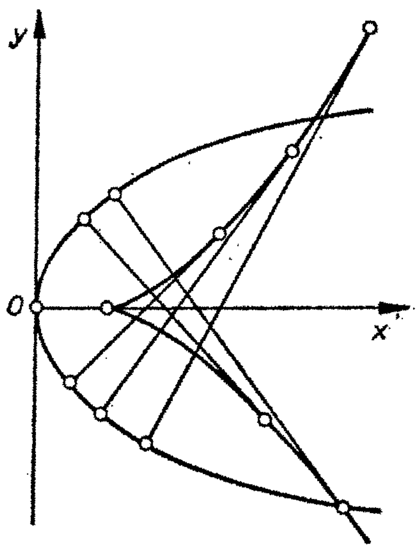

Abb. 160

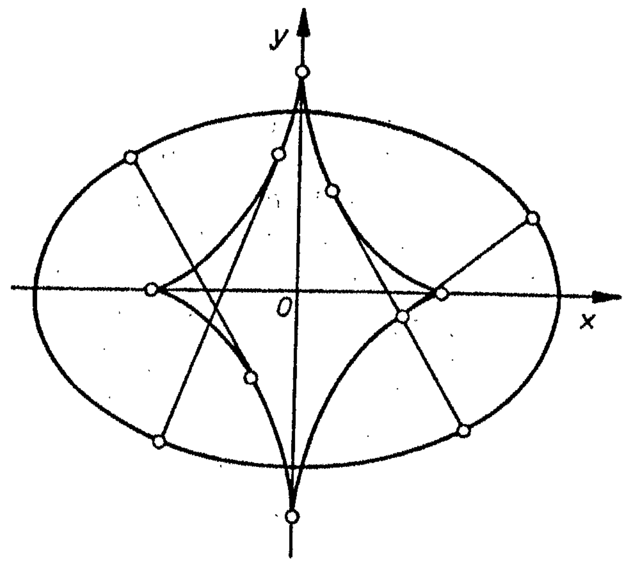

Abb. 161

2. Man bestimme die Evolute der Ellipse $x = a\cos t$, $y = b\sin t$. Hier ist

$$x_t = -a\sin t, \quad x_{tt} = -a\cos t, \quad y_t = b\cos t, \quad y_{tt} = -b\sin t.$$

Durch Einsetzen in Formel (10) folgt

$$\xi = a\cos t - \frac{b\cos t(a^2\sin^2 t + b^2\cos^2 t)}{ab} = \frac{a^2 - b^2}{a}\cos^3 t, \qquad \eta = -\frac{a^2 - b^2}{b}\sin^3 t.$$

Das ist eine Parameterdarstellung der Ellipsenevolute. Eliminiert man t, so erhält man die Gleichung dieser Kurve in der Gestalt

$$(a\xi)^{2/3} + (b\eta)^{2/3} = c^{4/3} \qquad (c^2 = a^2 - b^2).$$

Die Kurve erinnert an die Astroide; sie ergibt sich aus dieser durch Streckung in der vertikalen Richtung (Abb. 161).

Analog ergibt sich (mit Hilfe der hyperbolischen statt der trigonometrischen Funktionen) für die Hyperbel $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$ die Evolute

$$(a\xi)^{2/3} - (b\eta)^{2/3} = c^{4/3} \qquad (c^2 = a^2 + b^2).$$

3. Man bestimme die Evolute der Astroide $x^{2/3} + y^{2/3} = a^{2/3}$. In Nr. 252, Beispiel 2, hatten wir

$$y_x = -\left(\frac{y}{x}\right)^{1/3}, \quad y_{xx} = \frac{1}{3}\left(\frac{a^2}{x^4 y}\right)^{1/3}$$

gefunden. Setzt man das in Formel (10a) ein, so findet man nach Vereinfachung

$$\xi = x + 3x^{1/3}y^{2/3}, \qquad \eta = y + 3x^{2/3}y^{1/3}.$$

Hieraus kann man unter Heranziehung der Astroidengleichung x und y eliminieren:

$$\xi + \eta = (x^{1/3} + y^{1/3})^3, \qquad \xi - \eta = (x^{1/3} - y^{1/3})^3,$$

$$(\xi + \eta)^{2/3} + (\xi - \eta)^{2/3} = 2(x^{2/3} + y^{2/3}) = 2a^{2/3}.$$

Dreht man die Koordinatenachsen um 45°, so erhält man die neuen Koordinaten ξ_1, η_1 aus ξ, η durch $\xi_1 = \frac{\xi+\eta}{\sqrt{2}}$, $\eta_1 = -\frac{\xi-\eta}{\sqrt{2}}$, so daß in diesem neuen Koordinatensystem die Gleichung der gesuchten Evolute die Gestalt

$$\xi_1^{2/3} + \eta_1^{2/3} = (2a)^{2/3}$$

annimmt. Darin erkennen wir wieder eine Astroide. Die Evolute der Astroide ist also ebenfalls eine Astroide, aber mit Achsen, die gegenüber den ursprünglichen um 45° gedreht sind, und mit doppelt so großen Abmessungen (Abb. 162).

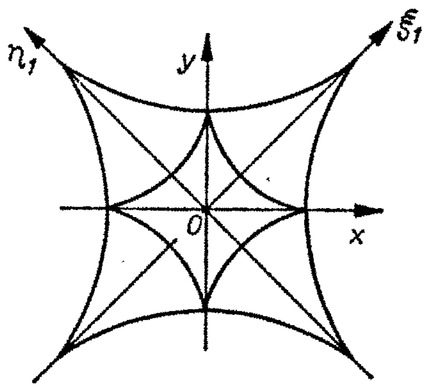

Abb. 162

4. Man bestimme die Evolute der Zykloide $x = a(t - \sin t)$, $y = a(1 - \cos t)$. Nach Nr. 231, Beispiel 4, gilt für die Zykloide

$$\alpha = \frac{\pi}{2} - \frac{t}{2}, \qquad d\alpha = -\frac{1}{2}\,dt;$$

zweckmäßigerweise benutzt man dann die Formeln (9). Setzt man darin diesen Wert für $d\alpha$ ein, so ergibt sich

$$\xi = x + 2y_t, \quad \eta = y - 2x_t \quad \text{oder} \quad \xi = a(t + \sin t), \quad \eta = -a(1 - \cos t).$$

Mit $t = \tau - \pi$ geht diese Parameterdarstellung in $\xi = -\pi a + a(\tau - \sin\tau)$, $\eta = -2a + a \times (1 - \cos\tau)$ über. Hieraus ersieht man, daß die Evolute der Zykloide eine zu dieser kongruente Zykloide ist, die um πa nach links (parallel zur x-Achse in negativer Richtung) und um $2a$ nach unten (parallel zur y-Achse, ebenfalls in negativer Richtung) verschoben ist.

Der Leser möge sich davon überzeugen, daß die Evoluten von Epizykloide und Hypozykloide ebenfalls der ursprünglichen Kurve kongruent sind und sich aus dieser durch Drehung ergeben.

5. Man bestimme die Evolute der logarithmischen Spirale $r = a\,e^{m\theta}$. Die in Nr. 252, Beispiel 5, angegebene geometrische Konstruktion des Krümmungsmittelpunktes ermöglicht es, seine Polarkoordinaten r_1 und θ_1 zu bestimmen. Nach Abb. 134 (S. 486) ist

$$r_1 = n_p = r\cot\omega = mr, \qquad \theta_1 = \theta + \frac{\pi}{2}.$$

Eliminiert man hieraus und aus der Kurvengleichung r und θ, so ergibt sich

$$r_1 = ma\,e^{m\left(\theta_1 - \frac{\pi}{2}\right)} = a_1\,e^{m\theta_1}.$$

Durch Drehung der Polarachse um einen passenden Winkel kann man diese Gleichung mit der ursprünglichen identifizieren. Somit ist die Evolute der logarithmischen Spirale dieselbe, aber durch eine Drehung der Polarachse entstehende Spirale.

Der Konstruktion von Evolventen für eine gegebene Kurve wenden wir uns zu, nachdem wir einige Eigenschaften von Evolute und Evolvente untersucht haben.

255. Eigenschaften von Evolute und Evolvente. Die Parameterdarstellung der Evolute lautete

$$\xi = x - R \sin \alpha, \quad \eta = y + R \cos \alpha, \tag{8}$$

wobei x, y, R, α Funktionen des Parameters sind. Wir wollen nun x und y als dreimal stetig differenzierbare Funktionen des Parameters voraussetzen.[1]) Dann ergibt sich aus (8)

$$d\xi = dx - R \cos \alpha \, d\alpha - dR \sin \alpha, \quad d\eta = dy - R \sin \alpha \, d\alpha + dR \cos \alpha.$$

Unter Beachtung von

$$R \cos \alpha \, d\alpha = \frac{ds}{d\alpha} \frac{dx}{ds} d\alpha = dx, \quad R \sin \alpha \, d\alpha = \frac{ds}{d\alpha} \frac{dy}{ds} d\alpha = dy$$

erhält man schließlich

$$d\xi = -\sin \alpha \, dR, \quad d\eta = \cos \alpha \, dR. \tag{11}$$

Wir beschränken uns nun auf einen solchen Teil der Kurve, für den R weder 0 noch ∞ wird und außerdem dR nicht verschwindet. Dadurch können weder auf der gegebenen Kurve noch auf ihrer Evolute singuläre Punkte auftreten. Wegen $dR \neq 0$ ändert sich der Krümmungsradius monoton; entweder er wächst, oder er fällt.

Durch Division der Formeln (11) finden wir

$$\frac{d\eta}{d\xi} = -\cot \alpha = -\frac{1}{\tan \alpha} = -\frac{1}{\frac{dy}{dx}};$$

also sind die Richtungskoeffizienten der Tangenten an Evolute und Evolvente zueinander reziprok und entgegengesetzten Vorzeichens. Demnach stehen die Tangenten aufeinander senkrecht. Somit gilt:

1. *Die Normale der Evolvente ist im Krümmungsmittelpunkt Tangente der Evolute.*

Wir betrachten die Schar der Normalen der Evolvente; sie hängt von einem einzigen Parameter ab (etwa von demjenigen, durch den die Lage des Punktes auf der gegebenen Kurve bestimmt wird). Aus unseren Ausführungen folgt, daß *die Evolute die Einhüllende dieser Normalenschar ist.*

Zur Übung möge sich der Leser auf einem anderen Wege davon überzeugen: Er möge, ausgehend von der Normalengleichung

$$(X - x) x_t + (Y - y) y_t = 0$$

(wobei der Parameter t in x, y, x_t, y_t steckt), nach den Methoden aus Nr. 238 die Einhüllende bestimmen und zeigen, daß sie mit der Evolute (10) zusammenfällt. Man kann auch zeigen, daß der *Krümmungsmittelpunkt charakteristischer Punkt auf der Normalen ist,* d. h. Grenzlage des Schnittpunktes der gegebenen Normalen mit unendlich benachbarten Normalen.

Wir wollen nun einen Bogen σ der Evolute untersuchen. Wir quadrieren die Gleichungen (11), addieren und finden, unter Beachtung der Formel (11) aus Nr. 248 für das Bogendifferential, $d\sigma^2 = d\xi^2 + d\eta^2 = dR^2$, also

$$d\sigma = \pm dR \tag{12}$$

---

[1]) In R stecken ja schon zweite Ableitungen.

oder (wegen $\mathrm{d}R \neq 0$)

$$\frac{\mathrm{d}\sigma}{\mathrm{d}R} = \pm 1.$$

Da dieser Quotient eine stetige Funktion des Parameters ist, also nicht zwischen -1 und $+1$ schwanken kann (es gibt ja keine Zwischenwerte!), ist er auf dem ganzen Kurvenstück gleich einer dieser Zahlen. Mit anderen Worten, auf der rechten Seite von (12) steht für das ganze Kurvenstück dasselbe Vorzeichen.

Dieses Vorzeichen hängt von der Wahl des Richtungssinns ab, in dem der Bogen auf der Evolute gerechnet wird. Wählt man ihn so, daß der Bogen mit dem Krümmungsradius R wächst, so gilt in (12) das Pluszeichen. Wächst σ in der Richtung, die abnehmendem R entspricht, so gilt das Minuszeichen. Wir wählen die erste Möglichkeit; dann ist

$$\mathrm{d}R = \mathrm{d}\sigma, \quad \text{also} \quad R - \sigma = c = \text{const}, \tag{13}$$

und wir erhalten den Satz:

2. *Der Krümmungsradius unterscheidet sich von der Bogenlänge der Evolute um eine Konstante.*

Somit ist die Differenz der Krümmungsradien in zwei Punkten der Evolvente gleich dem Bogen der Evolute zwischen den entsprechenden Krümmungsmittelpunkten. Hieraus folgt übrigens ein interessantes Verfahren zur Rektifizierung des Evolutenbogens.

Die soeben bewiesene Eigenschaft der Evolute erlaubt nämlich eine elegante mechanische Deutung. Um sie leichter darlegen zu können, nehmen wir an, der Krümmungsradius R, der nicht verschwindet und auf dem ganzen Kurvenstück das Vorzeichen beibehält, sei überall positiv. Das kann man durch passende Wahl der Richtung, in der die Bogenlänge auf der Evolvente gerechnet wird, erreichen. Rechnen wir ferner den Bogen auf der Evolvente von dem Punkt P an, der dem kleinsten Krümmungsradius entspricht, so ist auch $\sigma > 0$. Unter diesen Annahmen ist auch die in Gleichung (13) auftretende Konstante c positiv.

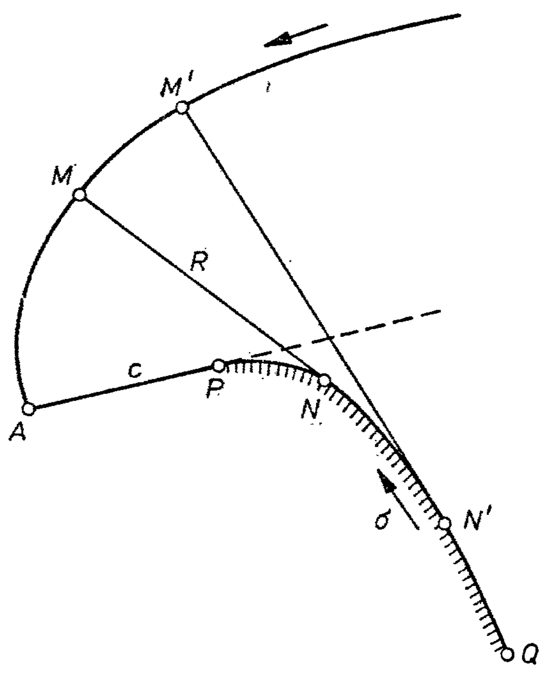

Abb. 163

Wir stellen uns nun vor, auf der Evolute sei ein biegsamer, nicht dehnbarer Faden vom Ende Q (Abb. 163) zum Anfang P aufgewickelt. Er verläßt die Evolute im Anfangspunkt P als Tangente und endet im Abstand c von P im entsprechenden Punkt A der Evolvente. Wir wollen den Faden von der Evolute abwickeln, ihn aber gespannt halten. Es sei QNM eine beliebige Lage des Fadens; da $\overline{NM}$ gerade um die Länge des Bogens $\widehat{PN} = \sigma$ größer ist als $\overline{PA} = c$, ist $\overline{NM}$ der Krümmungsradius R, d. h., M liegt auf der Evolvente.

Die Evolvente kann also als Abwicklung eines Fadens beschrieben werden, der vorher auf der Evolute aufgewickelt war.[1]) Man kann daher auch sagen: Die Evolvente ist die Bahn des Punktes A einer Geraden AP, wenn die Gerade ohne Schlupf längs der Evolute abgerollt wird.

Zum Schluß leiten wir noch eine Formel für den Krümmungsradius ϱ der Evolute her. Es sei β der Winkel zwischen der Tangente der Evolute und der x-Achse; dann ist offenbar

$$\beta = \alpha \pm \frac{\pi}{2}, \quad \text{also} \quad \mathrm{d}\beta = \mathrm{d}\alpha. \tag{14}$$

Daher ist [vgl. (13) und (14)]

$$\varrho = \frac{\mathrm{d}\sigma}{\mathrm{d}\beta} = \frac{\mathrm{d}R}{\mathrm{d}\alpha} = \frac{\mathrm{d}s}{\mathrm{d}\alpha}\,\frac{\mathrm{d}R}{\mathrm{d}s} = R\,\frac{\mathrm{d}R}{\mathrm{d}s}. \tag{15}$$

Es sei daran erinnert, daß hier vorausgesetzt wird, σ wachse mit R. Sonst müßte rechts ein Minuszeichen stehen.

Nimmt man an, σ wachse mit s, so kann man

$$\varrho = R\left|\frac{\mathrm{d}R}{\mathrm{d}s}\right| \tag{16}$$

schreiben, worin der Fall $\frac{\mathrm{d}R}{\mathrm{d}s} > 0$ (R wächst mit s) und der Fall $\frac{\mathrm{d}R}{\mathrm{d}s} < 0$ (R fällt mit wachsendem s) vereinigt sind.

256. Bestimmung der Evolvente. Wir sehen, daß man sich jede Evolvente aus ihrer Evolute durch Abwicklung eines auf die Evolute gewickelten Fadens entstanden denken kann oder, was im Grunde dasselbe ist, durch Rollen einer Geraden längs der Evolute (ohne Schlupf).

Wir wollen nun die umgekehrte Aussage beweisen. *Wenn eine Gerade ohne Schlupf längs einer gegebenen Kurve rollt, so ist die Bahn jedes ihrer Punkte eine Evolute dieser Kurve.* Jede Kurve hat also unendlich viele Evolventen.

Es sei die Kurve $\widehat{PN}$ (Abb. 164) in der Parameterdarstellung $\xi = \varphi(t)$, $\eta = \psi(t)$ gegeben; φ und ψ mögen stetige Ableitungen zweiter Ordnung besitzen. Ferner sollen auf dem betrachteten Kurvenstück weder mehrfache noch singuläre Punkte existieren. Die Bogenlänge σ der Kurve werde von P an gerechnet.

---

[1]) Aus dieser Eigenschaft erklären sich die Bezeichnungen Evolute und Evolvente (lat. evolvere = abwickeln).

Auf der Tangente in P, die nach der Seite wachsender Bogenlänge gerichtet sei, wählen wir einen beliebigen Punkt A, dessen Abstand von P (unter Berücksichtigung des Vorzeichens) wir mit c bezeichnen. Wir verfolgen seine Bahn, wenn die Gerade PA ohne Schlupf längs der gegebenen Kurve gleitet. Bei der neuen Lage der Geraden,

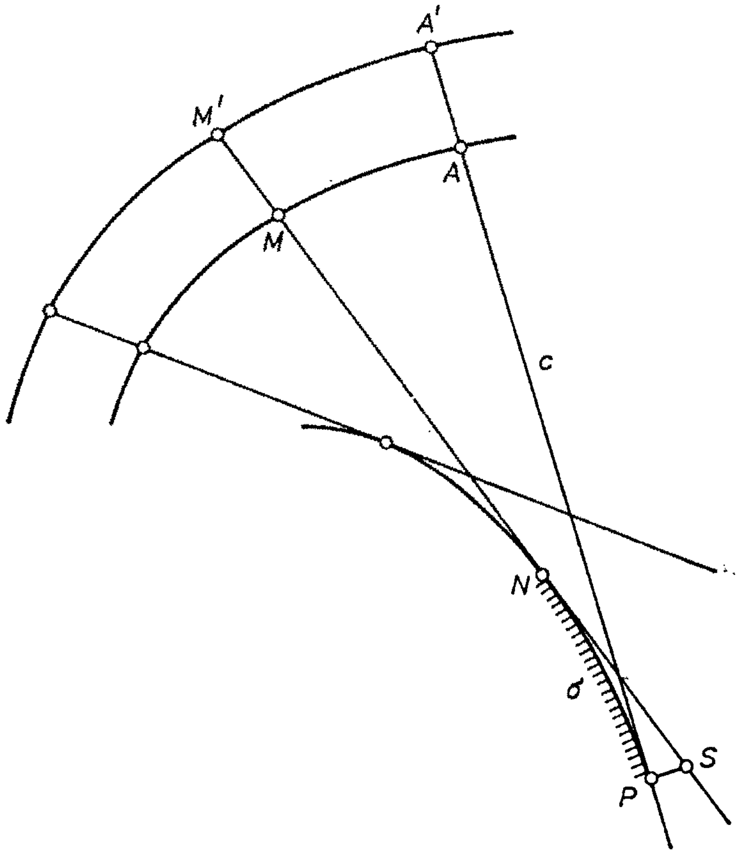

Abb. 164

wenn N Berührungspunkt ist, geht P in S über und A in M. Offenbar ist $\overline{SN} = \widehat{PN} = \sigma$, also $\overline{NM} = c - \sigma$. Die Koordinaten von N und M seien ξ, η bzw. x, y; der Winkel zwischen der Geraden SN und der x-Achse sei β; durch Projektion von NM auf die Achsen erhält man

$$x = \xi + (c - \sigma)\cos\beta, \quad y = \eta + (c - \sigma)\sin\beta. \tag{17}$$

Diese Gleichung ist die Parameterdarstellung der gesuchten Bahnkurve.

Differentiation liefert

$$dx = d\xi - \cos\beta\, d\sigma - (c - \sigma)\sin\beta\, d\beta,$$
$$dy = d\eta - \sin\beta\, d\sigma + (c - \sigma)\cos\beta\, d\beta.$$

Wegen

$$\cos\beta = \frac{d\xi}{d\sigma}, \quad \sin\beta = \frac{d\eta}{d\sigma} \tag{18}$$

[vgl. Nr. 249, Formel (15)] läßt sich das vereinfachen zu

$$dx = -(c - \sigma)\sin\beta\, d\beta, \quad dy = (c - \sigma)\cos\beta\, d\beta.$$

Die Fälle $d\beta = 0$ oder $\sigma = c$ lassen wir beiseite[1]); dann folgt durch Division

$$\tan\alpha = \frac{dy}{dx} = -\cot\beta = -\frac{1}{\frac{d\eta}{d\xi}}.$$

Hieraus ersieht man schon, daß die Tangenten an die beiden Kurven senkrecht aufeinanderstehen, so daß die gegebene Kurve tatsächlich Einhüllende der Normalenschar der konstruierten Kurve ist, d. h. ihre Evolute. Die konstruierte Kurve ist also Evolvente der gegebenen, was bewiesen werden sollte.

Als Beispiel, wie man die Evolvente auf diesem Wege erhält, möge die schon früher behandelte Kreisevolvente dienen; vgl. dazu Nr. 225, Beispiel 8, und Nr. 252, Beispiel 4.

---

[1]) Ihnen entsprechen singuläre Punkte auf der konstruierten Kurve.

Anhang

Das Problem der Erweiterung von Funktionen

257. Bemerkungen zum Funktionsbegriff. In der auf S. 93 gegebenen Definition des Funktionsbegriffs werden solche Begriffe wie „Zuordnung“ und „Zuordnung nach einer Vorschrift oder nach einem Gesetz“ benutzt. Wir wollen ergänzend zeigen, daß man ohne diese Begriffe auskommen bzw. sie auf allgemeinere, bereits definierte mathematische Begriffe zurückführen kann. Das ist möglich, wenn man sich mengentheoretischer Begriffsbildungen bedient. Dabei ist natürlich an dieser Stelle eine systematische Entwicklung dieser Begriffe nicht möglich. (Beipsielsweise definieren wir den Begriff des geordneten Paares anschaulich und nicht durch Rückgriff auf mengentheoretische Begriffsbildungen.)

Zur Präzisierung des Begriffs der Abbildung, durch die den Elementen einer Menge M jeweils gewisse Elemente einer Menge N zugeordnet werden sollen, müssen wir zunächst sagen, was wir unter *geordneten Paaren* von Individuen verstehen wollen. Wir nennen das aus den Elementen a und b gebildete Paar $[a, b]$ geordnet, wenn die Reihenfolge der Elemente innerhalb des Paares nicht beliebig ist. (Der Leser denke z. B. daran, wie die Koordinaten eines Punktes in der Ebene mit Hilfe eines Koordinatensystems angegeben werden.) *Geordnete Paare* $[a, b]$ *und* $[c, d]$ *sind also dann und nur dann gleich, wenn* $a = c$ *und* $b = d$ *ist.*

Aus vorgegebenen Mengen M und N bilden wir jetzt die *Produktmenge* $M \times N$ (oder auch *Kreuzmenge* genannt), indem wir alle geordneten Paare $[a, b]$, in denen $a \in M$ und $b \in N$ ist, zu einer Menge zusammenfassen. Es gilt also

$$[a, b] \in M \times N \quad \text{genau dann, wenn } a \in M \quad \text{und} \quad b \in N.$$

Jede Teilmenge F der Produktmenge $M \times N$ ($F \subseteqq M \times N$) nennen wir eine *Abbildung aus* der Menge M *in* die Menge N. (Zur Bezeichnung von Abbildungen werden im allgemeinen die Buchstaben f, g, F, φ usw. verwendet.) Wenn F eine Abbildung aus M in N und $[x, y] \in F$ ist, so nennen wir y ein *Bild* von x bei der Abbildung F und x ein *Urbild* von y bei der Abbildung F. Mit diesen sehr anschaulichen Begriffsbildungen und Bezeichnungen wird offenbar gerade das präzisiert, was man mit den Bezeichnungen „Abbildung“ oder „Zuordnung“ erfassen will.

Nach der oben gegebenen Definition brauchen nun weder alle Elemente von M noch alle Elemente von N in den geordneten Paaren aufzutreten. Uns interessieren aber im allgemeinen nur die Elemente aus M, die in N wirklich Bilder besitzen, bzw. nur die Elemente aus N, die in M tatsächlich Urbilder besitzen. Zur genauen Abgrenzung nennen wir die Menge aller Elemente $x \in M$, für die es ein $y \in N$ mit $[x, y] \in F$ gibt, *Definitionsbereich* (auch Vorbereich, Urbildbereich oder Argumentbereich) der Abbildung F (aus M in N). Die Menge aller $y \in N$, für die es ein $x \in M$ mit $[x, y] \in F$ gibt, nennen wir den *Wertebereich* (auch Wertevorrat, Nachbereich oder Bildbereich) der Abbildung F (aus M in N).

Ist bei einer Abbildung F der Definitionsbereich die Menge M und der Wertebereich die Menge N, wird also jedem $x \in M$ durch F mindestens ein $y \in N$ zugeordnet, und hat bei dieser Zuordnung auch jedes $y \in N$ mindestens ein Urbild $x \in M$, so spricht man auch von einer *Abbildung* von M *auf* N.

Hat eine Abbildung F die Eigenschaft, daß aus $[x, y] \in F$ und $[x, z] \in F$ stets $y = z$ folgt, so nennen wir diese Abbildung *eindeutig*. Bei einer solchen Abbildung wird also jedem $x \in M$ höchstens ein $y \in N$ zugeordnet.

Hat eine Abbildung F darüber hinaus die Eigenschaft, daß aus $[x, z] \in F$ und $[y, z] \in F$ stets $x = y$ folgt, so heißt diese Abbildung *umkehrbar eindeutig* oder *eineindeutig*. In diesem Fall gibt es also auch zu jedem $y \in N$ höchstens ein Urbild $x \in Y$.

Nach diesen Vorbereitungen können wir nun den Begriff der Funktion als eine spezielle Abbildung folgendermaßen definieren:

Eine *Funktion* f ist eine eindeutige Abbildung von M auf N. Dabei ist M der Definitionsbereich und N der Wertebereich der Funktion f.

Funktionen sind also Mengen geordneter Paare, die in der zweiten Stelle eindeutig sind.[1]) In diesem Sinne wird durch die Funktion f jedem Element x des Definitionsbereichs von f genau ein Element y des Wertebereichs von f zugeordnet. Statt der oben eingeführten Schreibweise $[x, y] \in f$ schreibt man auch $y = f(x)$. Diese Schreibweise besagt also nach der hier gegebenen Definition, daß $y = f(x)$ der Funktionswert der Funktion f an der Stelle x ist. Bei dieser Auffassung können wir *nicht* sagen, daß „die Veränderliche y eine Funktion der Veränderlichen x“ ist. Vielmehr sind die Variablen Zeichen, für welche Zahlen aus dem Definitionsbereich bzw. aus dem Wertebereich eingesetzt werden können. Trotz dieser unterschiedlichen Auffassung bemerkt der Leser jedoch leicht, daß sowohl die in Nr. 45 gegebene Definition des Funktionsbegriffs als auch die hier angegebene Definition denselben Sachverhalt erfassen.

258. Funktionen einer Veränderlichen. Wir betrachten eine Funktion $f(x)$, die in einem (endlichen oder unendlichen) Intervall $\mathfrak{X}$ oder — allgemeiner — auf einem Bereich $\mathfrak{X}$ definiert ist, der aus endlich vielen solchen einzelnen Intervallen besteht. Ist die Funktion $f(x)$ in $\mathfrak{X}$ stetig und besitzt sie dort stetige Ableitungen bis zur n-ten Ordnung einschließlich ($n \geqq 1$), so sagt man, $f(x)$ *gehöre in* $\mathfrak{X}$ *zur Klasse* C^n.

Gehört ein Endpunkt irgendeines dieser Intervalle zu $\mathfrak{X}$, so handelt es sich dort um einseitige Ableitungen.[2])

Die Funktion $f(x)$ gehöre in einem Bereich $\mathfrak{X}$, der nicht die ganze Zahlengerade umfaßt, zur Klasse C^n ($n = 1, 2, 3, \ldots$). In einem den Bereich $\mathfrak{X}$ umfassenden Bereich $\mathfrak{X}^*$ existiere eine Funktion $f^*(x)$ der Klasse C^n, die im Durchschnitt von $\mathfrak{X}$ und $\mathfrak{X}^*$ mit $f(x)$ übereinstimmt. Dann ist f^* eine *Erweiterung* (*Fortsetzung*) *von* f *auf* $\mathfrak{X}^*$ unter Erhaltung der Klasse. Die Funktion f nennt man dann die *Einschränkung* von f^* auf $\mathfrak{X}$.

Ist es immer möglich, eine Funktion in dieser Art auf einen umfassenderen Bereich zu erweitern? Hierauf gibt der folgende Satz Antwort.

Satz. *Jede Funktion $f(x)$ der Klasse C^n ($n = 1, 2, 3, \ldots$) in einem abgeschlossenen Bereich*[3]) $\mathfrak{X}$ *kann auf die ganze Zahlengerade $\mathfrak{X}^* = (-\infty, \infty)$ unter Erhaltung der Klasse erweitert werden.*

---

[1]) Dem Leser sind ja „Wertetabellen“ von Funktionen bekannt. An dieser Art der Darstellung einer Funktion knüpft die hier gegebene Definition im Grunde an.

[2]) Oder, was unter unserer Annahme dasselbe ist, um die Grenzwerte der Ableitungen bei unbegrenzter Annäherung des Punktes an den Endpunkt von der Seite des Intervalls her.

[3]) d. h. in einem aus einem oder mehreren abgeschlossenen Intervallen der Gestalt $[a, b]$, $[a, \infty)$, $(-\infty, b]$ bestehenden Bereich

Wir zeigen, daß hier die Erweiterung einfach durch Polynome erreicht werden kann. Zu diesem Zweck bringen wir zunächst einige vorbereitende Bemerkungen.

Wir haben in Nr. 123 gesehen, daß das Polynom n-ten Grades

$$p(x) = c_0 + \frac{c_1}{1!}(x-\alpha) + \frac{c_2}{2!}(x-\alpha)^2 + \cdots + \frac{c_n}{n!}(x-\alpha)^n \tag{1}$$

und seine Ableitungen im Punkt $x = \alpha$ gerade die Werte $c_0, c_1, c_2, \ldots, c_n$ annehmen.

Es sei ferner gefordert, ein Polynom zu konstruieren, das mit seinen n Ableitungen neben den angegebenen Werten in $x = \alpha$ noch in einem anderen Punkt $x = \beta$ die vorgegebenen Werte $d_0, d_1, d_2, \ldots, d_n$ annimmt.

Wir schreiben das gesuchte Polynom in der Gestalt

$$p(x) + (x-\alpha)^{n+1} q(x), \tag{2}$$

wobei $p(x)$ das Polynom (1) ist, während das Polynom $q(x)$ vom Grad n noch zu bestimmen ist. Für jedes $q(x)$ genügt das Polynom (2) in $x = \alpha$ den gestellten Bedingungen. Nun differenzieren wir das Polynom (2) nacheinander n-mal und setzen jeweils $x = \beta$. Die so entstehenden Ausdrücke setzen wir gleich $d_0, d_1, d_2, \ldots, d_n$ und erhalten so ein System linearer Gleichungen für $q(\beta), q'(\beta), q''(\beta), \ldots, q^{(n)}(\beta)$, aus denen sich diese Werte bestimmen lassen. Aus diesen läßt sich dann nach einer zu (1) analogen Formel $q(x)$ finden (vgl. Nr. 130).

Wir wenden uns jetzt dem Beweis der oben formulierten Aussage zu. Der Bereich $\mathcal{X}$ bestehe aus abgeschlossenen Intervallen $\mathcal{X}_k$ $(k = 1, 2, \ldots, m)$, die von links nach rechts numeriert seien. Wir setzen in diesen Intervallen $f^* = f$ und ergänzen ihre Definition in folgender Weise. Ist der linke Endpunkt a_1 von $\mathcal{X}_1$ eine endliche Zahl, so setzen wir für $x < a_1$ die Funktion f^* gleich einem Polynom der Gestalt (1), für

$$c_0 = f(a_1), \quad c_1 = f'(a_1), \quad \ldots, \quad c_n = f^{(n)}(a_1).$$

Analog läßt sich die Funktion f auch rechts von $\mathcal{X}_m$ fortsetzen, wenn der rechte Endpunkt b_m dieses Intervalls eine endliche Zahl ist. Schließlich identifizieren wir im Intervall (b_k, a_{k+1}) $(k = 1, 2, \ldots, m-1)$, das $\mathcal{X}_k$ von $\mathcal{X}_{k+1}$ trennt, f^* mit dem Polynom, das nebst seinen n Ableitungen in den Punkten $x = b_k$ und $x = a_{k+1}$ dieselben Werte annimmt wie die Funktion f und ihre Ableitungen. Man sieht leicht, daß die so definierte Funktion f^* die geforderte Erweiterung von f auf den ganzen Bereich $\mathcal{X}^* = (-\infty, \infty)$ liefert.

259. Die Problemstellung im zweidimensionalen Fall. Die Sachlage wird beim Übergang zu Funktionen mehrerer Veränderlicher schnell komplizierter. Wir beschränken uns im folgenden auf Funktionen zweier Veränderlicher. Die Resultate, die wir dabei erhalten, lassen sich dann auf den Fall beliebig vieler Veränderlicher übertragen.

Wir betrachten einen Bereich $\mathcal{M}$ im zweidimensionalen Raum, wobei wir darunter entweder einen offenen Bereich oder aber einen offenen Bereich nebst Teilen seines Randes $\mathcal{L}$ oder nebst seinem ganzen Rand verstehen wollen (dann ist der Bereich abgeschlossen).

Wenn wir die Definition der Funktionen der Klasse C^n $(n \geqq 1)$ auf diesen Fall ausdehnen wollen, stoßen wir auf eine eigenartige Schwierigkeit. In einem Punkt des Randes $\mathcal{L}$ von $\mathcal{M}$ kann es vorkommen, daß eine bestimmte partielle Ableitung nicht definiert ist. Ist etwa $\mathcal{M}$ der abgeschlossene Kreis $x^2 + y^2 \leqq 1$, so kann man in den Punkten $(0, \pm 1)$ nicht von einer partiellen Ableitung nach x sprechen, denn für $y = \pm 1$ kann man dem Wert $x = 0$ keinen Zuwachs erteilen, ohne den Bereich zu verlassen, in dem die Funktion definiert ist. Analog kann man die partielle Ableitung nach y in den Punkten $(\pm 1, 0)$ nicht definieren.

Wenn wir von einer in $\mathcal{M}$ stetigen partiellen Ableitung (einer bestimmten Ordnung und eines bestimmten Typs) sprechen, wollen wir in einem Randpunkt M_0 unter dieser Ableitung — für die wir kein neues Zeichen einführen — nur den Grenzwert verstehen, gegen den die in einem inneren Punkt M berechnete gleichnamige Ableitung strebt, wenn M gegen M_0 strebt, unabhängig davon, ob er tatsächlich die Rolle der Ableitung spielt oder nicht.

Aus den weiteren Darlegungen wird nach und nach klar werden, daß der erwähnte Grenzwert — in einer umfassenden Klasse von Fällen — auch tatsächlich die Ableitung ist, sobald die Lage des Punktes M_0 in bezug auf den Bereich es überhaupt erlaubt, von der Ableitung des betreffenden Typs zu sprechen. Übrigens werden wir das für den einfachsten Fall eines rechteckigen Bereichs jetzt beweisen.

Es sei also die Funktion $f(x, y)$ in einem Rechteck $\mathcal{M}$ nebst allen ihren Ableitungen bis zur n-ten Ordnung $(n \geqq 1)$ einschließlich stetig, und der Punkt $M_0(x_0, y_0)$ liege auf der Strecke der Geraden $y = y_0$, die zum Rand des Rechtecks (Abb. 165) und zum Bereich selbst gehört.

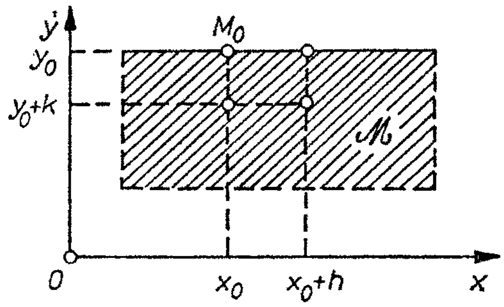

Abb. 165

Wir beginnen mit f_y, wo das Problem einfach zu erledigen ist. Nach dem Mittelwertsatz (Nr. 112) strebt der Differenzenquotient

$$\frac{f(x_0, y_0 + k) - f(x_0, y_0)}{k} = f_y(x_0, y_0 + \theta k) \quad (0 < \theta < 1)$$

für $k \to 0$ gerade gegen den Grenzwert $f_y(x_0, y_0)$, der also auch Ableitung im gewöhnlichen Sinne ist (vgl. Nr. 113). Dagegen kann für f_x der entsprechende Differenzenquotient als Limes

$$\frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h} = \lim_{k \to 0} \frac{f(x_0 + h, y_0 + k) - f(x_0, y_0 + k)}{h}$$

angesehen werden. Diesen Ausdruck formen wir wieder nach dem Mittelwertsatz um:

$$\frac{f(x_0 + h, y_0 + k) - f(x_0, y_0 + k)}{h} = f_x(x_0 + \theta h, y_0 + k) \quad (0 < \theta < 1).$$

Für $h \to 0$, $k \to 0$ strebt er gegen den Grenzwert $f_x(x_0, y_0)$. Nach dem Satz aus Nr. 168 ist dieser Doppellimes auf Grund der Existenz des einfachen Limes für $k \to 0$ zugleich iterierter Limes:

$$\begin{aligned} f_x(x_0, y_0) &= \lim_{h \to 0} \lim_{k \to 0} \frac{f(x_0 + h, y_0 + k) - f(x_0, y_0 + k)}{h} \\ &= \lim_{h \to 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h}, \end{aligned}$$

so daß auch hier die nur als Grenzwert der Ableitung definierte Zahl $f_x(x_0, y_0)$ wirklich die Ableitung ist. Das überträgt sich sukzessive auch auf die höheren Ableitungen.

Somit können wir auf Grund der obigen Vereinbarung jetzt von stetigen Ableitungen in jedem Bereich $\mathcal{M}$ sprechen, wie auch immer seine (zum Bereich gehörigen) Randpunkte in bezug auf diesen Bereich liegen. Eine Funktion $f(x, y)$ gehört in dem zweidimensionalen Bereich $\mathcal{M}$ der Klasse C^n ($n \geqq 1$) an, wenn sie in $\mathcal{M}$ stetig ist und stetige Ableitungen aller Typen und Ordnungen bis zur n-ten einschließlich besitzt. Nun möge $\mathcal{M}$ nicht die ganze Ebene umfassen; wenn in einem $\mathcal{M}$ umfassenden Bereich $\mathcal{M}^*$ eine Funktion f^* existiert, die ebenfalls der Klasse C^n angehört und die im Durchschnitt von $\mathcal{M}$ und $\mathcal{M}^*$ mit f übereinstimmt, so sagen wir, f^* sei eine *Erweiterung* (*Fortsetzung*) *von f auf $\mathcal{M}^*$*, unter Erhaltung der Klasse. Naturgemäß entsteht auch hier die Frage: Ist eine solche Erweiterung auf einen größeren Bereich, insbesondere auf die ganze Ebene, immer möglich? Wie wir zeigen werden, lautet die Antwort für einen abgeschlossenen Bereich $\mathcal{M}$ bejahend, sobald sein Rand bestimmten einfachen Bedingungen genügt. Zur Erleichterung der Darlegung werden wir übrigens $\mathcal{M}$ als beschränkt voraussetzen, obwohl die Aussage auch für unbeschränkte Bereiche gilt.

Die hier behandelten Resultate stammen in der Hauptsache von H. WHITNEY[1]) und M. R. HESTENES[2]).

260. Der Hauptsatz über die Erweiterung. Um den Beweis des Hauptsatzes zu vereinfachen, beweisen wir zunächst einige Hilfssätze.

Lemma I. *Die Funktion $\varphi(u, v)$ gehöre im Bereich $\mathcal{P}: (a < u < b;\ 0 \leqq v < \Delta)$ der Klasse C^n ($n \geqq 1$) an.*[3]) *Dann existiert eine Erweiterung von φ auf das ganze Rechteck*

$$\mathcal{P}^* : (a, b;\ -\Delta, \Delta), \quad \text{d. h. für} \quad (a < u < b,\ -\Delta < v < \Delta),$$

und zwar unter Erhaltung der Klasse.

Wir bestimmen $n + 1$ Zahlen $\lambda_1, \lambda_2, \ldots, \lambda_{n+1}$ aus dem System der $n + 1$ linearen Gleichungen

$$(-1)^k \lambda_1 + \left(-\frac{1}{2}\right)^k \lambda_2 + \cdots + \left(-\frac{1}{n+1}\right)^k \lambda_{n+1} = 1 \tag{3}$$

$$(k = 0, 1, 2, \ldots, n).$$

Das geht, da die Systemdeterminante die sogenannte *Vandermondesche Determinante* für die voneinander verschiedenen Zahlen $-1, -\frac{1}{2}, \ldots, -\frac{1}{n+1}$ ist, die bekanntlich von 0 verschieden ist.

Wir definieren nun in $\mathcal{P}^*$ eine Funktion $\varphi^*(u, v)$, indem wir $\varphi^*(u, v) = \varphi(u, v)$ für $v \geqq 0$ und

$$\varphi^*(u, v) = \lambda_1 \varphi(u, -v) + \lambda_2 \varphi\left(u, -\frac{1}{2} v\right) + \cdots + \lambda_{n+1} \varphi\left(u, -\frac{1}{n+1} v\right) \tag{4}$$

für $v < 0$ setzen. Ist u_0 ein beliebiger Wert von u aus (a, b), so ist zunächst

$$\lim_{\substack{u \to u_0 \\ v \to -0}} \varphi^*(u, v) = (\lambda_1 + \lambda_2 + \cdots + \lambda_{n+1})\, \varphi(u_0, 0) = \varphi(u_0, 0)$$

---

[1]) HASSLER WHITNEY, 20. Jh., amerikanischer Mathematiker.

[2]) MAGNUS R. HESTENES, 20. Jh., amerikanischer Mathematiker.

[3]) Das Intervall (a, b) kann auch unendlich sein, ebenso wie die positive Zahl Δ gleich ∞ sein kann.

auf Grund der ersten Bedingungen (3), die $k = 0$ entspricht. Damit ist die Stetigkeit von φ^* in allen Punkten von $\mathcal{P}^*$ bewiesen, die auf der Geraden $v = 0$ liegen; die Stetigkeit in den übrigen Punkten von $\mathcal{P}^*$ ist trivial. Nun untersuchen wir das Problem der Existenz und Stetigkeit der Ableitungen von φ^* in $\mathcal{P}^*$. Auch hier brauchen nur die Punkte der Geraden $v = 0$ untersucht zu werden. Für alle Ableitungen

$$\frac{\partial^{i+k}\varphi^*(u, v)}{\partial u^i\, \partial v^k} \quad (1 \leqq i + k \leqq n) \tag{5}$$

zeigen wir die Gültigkeit der Beziehung

$$\lim_{\substack{u \to u_0 \\ v \to -0}} \frac{\partial^{i+k}\varphi^*(u, v)}{\partial u^i\, \partial v^k} = \frac{\partial^{i+k}\varphi(u_0, 0)}{\partial u^i\, \partial v^k}. \tag{6}$$

Zu diesem Zweck differenzieren wir (4) i-mal nach u und dann k-mal nach v $(v < 0)$,

$$\frac{\partial^{i+k}\varphi^*(u, v)}{\partial u^i\, \partial v^k} = (-1)^k \lambda_1 \frac{\partial^{i+k}\varphi(u, -v)}{\partial u^i\, \partial v^k} + \left(-\frac{1}{2}\right)^k \lambda_2 \frac{\partial^{i+k}\varphi\left(u, -\frac{1}{2}v\right)}{\partial u^i\, \partial v^k} + \cdots$$
$$+ \left(-\frac{1}{n+1}\right)^k \lambda_{n+1} \frac{\partial^{i+k}\varphi\left(u, -\frac{1}{n+1}v\right)}{\partial u^i\, \partial v^k},$$

und gehen für $u \to u_0$, $v \to -0$ zur Grenze über. Auf Grund von (3) erhalten wir schließlich (6).

Somit ist die Existenz eines (einzigen) Grenzwertes für jede Ableitung (5) sowohl von der Seite $v > 0$ als auch von der Seite $v < 0$ her sicher. Nimmt man als Wert der Ableitung (5) in den Punkten der Geraden $v = 0$ diesen Grenzwert, so erhält man überdies eine in ganz $\mathcal{P}^*$ stetige Funktion. Der Punkt $(u_0, 0)$ ist aber ein innerer Punkt von $\mathcal{P}^*$, wo wir eine Ableitung im eigentlichen Sinne brauchen. Hier können wir uns auf das in Nr. 259 Bewiesene stützen: Dieser Grenzwert ist zugleich die eigentliche Ableitung.

Also ist φ^* tatsächlich die gesuchte Erweiterung von φ auf $\mathcal{P}^*$.

Lemma II. *Die Funktion $f(x, y)$ gehöre in einem beschränkten offenen Bereich $\mathcal{M}$ der Klasse C^n an.*[1] *Kann man jeden Punkt des Randes $\mathcal{L}$ von $\mathcal{M}$ mit einer Umgebung versehen, in der eine Erweiterung von f unter Erhaltung der Klasse möglich ist, so kann man diese Erweiterung auf die ganze Ebene $\mathcal{E}$ erstrecken.*

Für jeden Punkt M des abgeschlossenen Bereichs $\bar{\mathcal{M}} = \mathcal{M} + \mathcal{L}$ gibt es entweder eine Umgebung, in der die Funktion f definiert ist und der Klasse C^n angehört, oder aber eine Umgebung, auf welche f unter Erhaltung der Klasse erweitert werden kann.[2] Diese Umgebung kann man beispielsweise als offenen Kreis $\bar{\sigma} = \mathcal{K}(M, 3r)$ um M vom Radius $3r$ wählen. Somit läßt sich der ganze abgeschlossene Bereich $\bar{\mathcal{M}}$ nicht nur durch ein System $\bar{\Sigma}$ überdecken, das aus diesen Kreisen $\bar{\sigma}$ besteht, sondern auch durch ein System Σ, das aus Kreisen $\sigma = \mathcal{K}(M, r)$ besteht, deren Radius nur ein Drittel davon beträgt.

[1]) Dieser Bereich braucht nicht zusammenhängend zu sein, und einstweilen setzen wir auch nichts über die Gestalt seines Randes voraus.

[2]) je nachdem, ob M zum offenen Bereich $\mathcal{M}$ oder zum Rand $\mathcal{L}$ gehört

Da $\mathcal{M}$ und damit $\bar{\mathcal{M}}$ beschränkt ist, kann man den Borelschen Überdeckungssatz (Nr. 175) anwenden und $\bar{\mathcal{M}}$ durch ein endliches System $\Sigma_m = \{\sigma_1, \sigma_2, \ldots, \sigma_m\}$ überdecken, das aus Σ ausgewählt ist. Hier ist $\sigma_i = \mathcal{K}(M_i, r_i)$ $(i = 1, 2, \ldots, m)$; gleichzeitig betrachten wir auch die Kreise $\sigma_i' = \mathcal{K}(M_i, 2r_i)$, $\sigma_i'' = \mathcal{K}(M_i, 3r_i)$.

Man kann leicht eine Funktion $h_i(M) = h_i(x, y)$ der Klasse C^n in $\mathcal{E}$ konstruieren derart, daß

$$h_i(M) = 0 \quad \text{in} \quad \sigma_i \quad \text{und} \quad h_i(M) = 1 \quad \text{in} \quad \mathcal{E} - \sigma_i' \quad (i = 1, 2, \ldots, m)$$

gilt. So kann man beispielsweise (nach den Methoden aus Nr. 258) eine Funktion $h(t)$ der Klasse C^n auf dem ganzen Intervall $-\infty < t < \infty$ definieren derart, daß $h(t) = 0$ für $t \leqq 1$ und $h(t) = 1$ für $t \geqq 2$ ist und dann

$$h_i(M) = h\left(\frac{\overline{MM_i}}{r_i}\right)$$

setzen. Mit Hilfe der Funktionen h_i bilden wir die Funktionen

$$H_1 = H_1(M) = 1 - h_1, \quad H_i = H_i(M) = h_1 \cdot h_2 \cdots h_{i-1}(1 - h_i)$$
$$(1 < i \leqq m);$$

sie gehören ebenfalls der Klasse C^n in $\mathcal{E}$ an. Offenbar ist

$$H_j = 0 \quad \text{in} \quad \sigma_i \quad (\text{für alle } j > i), \tag{7}$$

$$H_i = 0 \quad \text{in} \quad \mathcal{E} - \sigma_i', \tag{8}$$

denn in σ_i verschwindet der Faktor h_i und in $\mathcal{E} - \sigma_i'$ der Faktor $1 - h_i$. Wegen

$$H_1 + H_2 + \cdots + H_i = (1 - h_1) + h_1(1 - h_2) + \cdots + h_1 h_2 \cdots h_{i-1}(1 - h_i)$$
$$= 1 - h_1 h_2 \cdots h_i$$

gilt

$$H_1 + H_2 + \cdots + H_i = 1 \quad \text{in} \quad \sigma_i, \tag{9}$$

weil dort h_i verschwindet.

Jetzt möge φ_i in σ_i'' mit f oder der obigen Erweiterung von f übereinstimmen; außerhalb σ_i'' sei $\varphi_i = f$ in den Punkten von $\mathcal{M}$ und $\varphi_i = 0$ in den übrigen Punkten. Die Funktion $\varphi_i H_i$ verschwindet in $\mathcal{E} - \sigma_i'$ [vgl. (8)] und gehört offenbar in der ganzen Ebene $\mathcal{E}$ zu C^n. Schließlich setzen wir in allen Punkten von $\mathcal{E}$

$$f^* = \sum_{j=1}^{m} \varphi_j H_j.$$

Damit ist eine Funktion f^* in der ganzen Ebene definiert, die der Klasse C^n angehört.

Nun wählen wir einen beliebigen Punkt M aus $\mathcal{M}$; er gehört zu einem Kreis σ_i. Da alle $\varphi_j(M) = f(M)$ sind und außerdem in diesem Punkt [wegen (9) und (7)]

$$H_1 + H_2 + \cdots + H_i = 1, \quad \text{aber} \quad H_j = 0 \quad \text{für} \quad j > i$$

gilt, ist $f^*(M) = f(M)$. Somit ist f^* die gesuchte Erweiterung.

Jetzt können wir auch für eine Funktion zweier Veränderlicher den Erweiterungssatz beweisen, allerdings mit Einschränkungen über den Rand des Bereichs.

Wir wollen eine einfache Kurve ohne singuläre Punkte, die durch die Gleichungen

$$x = \varphi(t), \quad y = \psi(t) \tag{10}$$

definiert ist, wobei t in einem Intervall $\mathcal{T}$ variiert, eine *glatte Kurve der Klasse* C^n ($n \geqq 1$) nennen, wenn φ und ψ in diesem Intervall der Klasse C^n angehören.

Satz I. *Gehört* $f(x, y)$ *in einem beschränkten abgeschlossenen Bereich* $\mathcal{M}$, *dessen Rand* $\mathcal{L}$ *aus einer oder mehrereren sich nicht überschneidenden glatten Kurven der Klasse* C^n ($n \geqq 1$) *besteht, zur Klasse* C^n, *so kann diese Funktion unter Erhaltung der Klasse auf die ganze Ebene* $\mathcal{E}$ *erweitert werden.*

Es sei $M_0(x_0, y_0)$ ein beliebiger Punkt des Randes $\mathcal{L}$, der Einfachheit halber $x_0 = y_0 = 0$. Dieser Punkt liegt auf einer der Kurven, aus denen $\mathcal{L}$ besteht (und ist einer ihrer nicht singulären Punkte). Dann kann man ohne Beschränkung der Allgemeinheit annehmen, in der Umgebung von M_0 sei die Kurve durch eine explizite Gleichung $y = g(x)$ beschreibbar, wobei g zur Klasse C^n gehört, und der Bereich $\mathcal{M}$ liege oberhalb dieser Kurve, d. h., in der Nähe von M_0 sei $\mathcal{M}$ durch $y \geqq g(x)$ definiert (Abb. 166a).

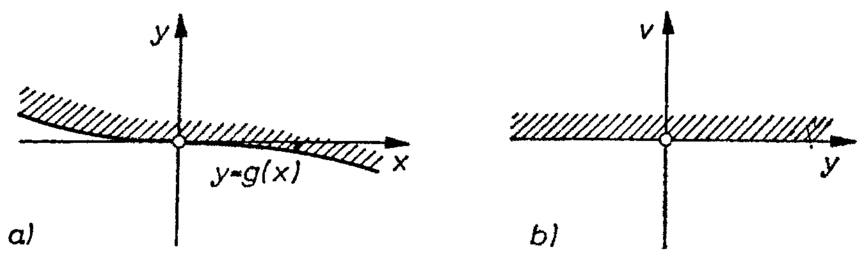

Abb. 166

Substituieren wir $x = u$, $y = g(u) + v$, so geht $f(x, y)$ in die Funktion

$$\varphi(u, v) = f\big(u, g(u) + v\big)$$

über, die in der Nähe des Punktes $u = v = 0$, nämlich für $v \geqq 0$ (Abb. 166b) zur Klasse C^n gehört. Dann kann man nach Lemma I die Funktion φ unter Erhaltung der Klasse auch auf $v < 0$ (unter Beschränkung auf Punkte in hinreichender Nähe des Ursprungs) fortsetzen. Vermittelt $\varphi^*(u, v)$ diese Erweiterung, so erkennt man bei der Rückkehr zu den alten Veränderlichen, daß

$$f^*(x, y) = \varphi^*\big(x, y - g(x)\big)$$

die Fortsetzung von f auf eine Umgebung von M_0 liefert.

Auf Grund von Lemma II können wir jetzt schließen, daß f tatsächlich unter Erhaltung der Klasse auf die ganze Ebene $\mathcal{E}$ fortsetzbar ist.

261. Verallgemeinerung. Dieses Resultat ist jedoch für praktische Bedürfnisse völlig unzureichend, da man es oft mit Bereichen zu tun hat, deren Ränder „Ecken“ haben. Wir wollen von einer *stückweise glatten Kurve* der Klasse C^n sprechen, wenn sie aus endlich vielen glatten Bögen der Klasse C^n besteht, die unter bestimmten, von 0 und π verschiedenen Winkeln aneinanderstoßen.

Satz II. *Satz* I *bleibt gültig, wenn der Rand* $\mathcal{L}$ *des Bereichs* $\mathcal{M}$ *aus einer oder mehreren (endlichen vielen) sich nicht überschneidenden stückweise glatten Kurven der Klasse* C^n *besteht.*

Wir haben schon gesehen, daß jedem Punkt des Randes $\mathcal{L}$, der keine Ecke ist, eine Umgebung zugeordnet werden kann, innerhalb welcher die Funktion f unter Erhaltung der Klasse fortgesetzt werden kann. Jetzt wollen wir beweisen, daß diese auch für einen Eckpunkt $M_0(x_0, y_0)$ gilt.

Auch hier nehmen wir wieder $x_0 = y_0 = 0$; man kann ohne Beschränkung der Allgemeinheit ferner annehmen, daß die im Ursprung zusammenstoßenden Bögen in diesem Punkt Halbtangenten besitzen, von denen eine mit der positiven x-Achse übereinstimmt, während die andere mit ihr einen bestimmten Winkel bildet (Abb. 167). In diesem Fall lassen sich die Bögen in hinreichender Nähe des Ursprungs durch die Gleichungen $y = g(x)$ und $x = h(y)$ ausdrücken; dabei ist $g'(0) = 0$, und g und h gehören beide der Klasse C^n an.

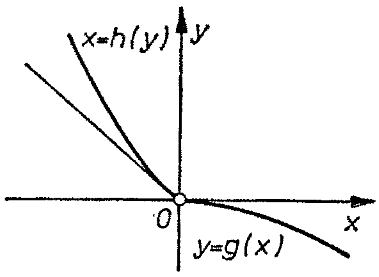

Abb. 167

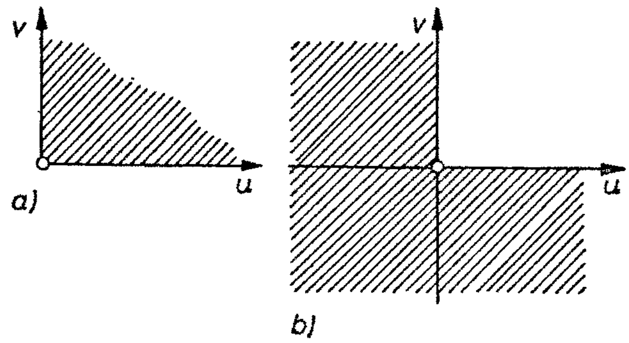

Abb. 168

Wir substituieren

$$x = u + h(v), \quad y = g(u) + v. \tag{11}$$

Da die Funktionaldeterminante

$$J = \begin{vmatrix} 1 & h'(v) \\ g'(u) & 1 \end{vmatrix} = 1 - g'(u)\,h'(v)$$

dieser Funktionen im Punkt $u = v = 0$ gleich 1 wird, hat das System (11) in der Umgebung der Nullstellen aller Argumente eine eindeutige Umkehrung

$$u = \lambda(x, y), \quad v = \mu(x, y), \tag{12}$$

wobei λ und μ ebenfalls der Klasse C^n angehören (Nr. 209).

Für $v = 0$ und $u \geqq 0$ folgen aus (11) die Beziehungen $y = g(x)$ und $x \geqq 0$, so daß dem positiven Teil der u-Achse der erste der genannten Bögen entspricht; analog überzeugt man sich davon, daß der positiven v-Achse der zweite der Bögen entspricht.

Offenbar entsprechen bei dieser Transformation den zwei Winkelbereichen, in die durch diese Bögen die Umgebung des Ursprungs in der x, y-Ebene geteilt wird, diejenigen beiden rechten Winkel (ein konvexer und ein einspringender), in welche durch die positiven Teile der u- bzw. v-Achse in der u, v-Ebene die Umgebung des Ursprungs geteilt wird (Abb. 168a, b).

Setzen wir (11) in f ein, so erhalten wir die transformierte Funktion

$$\varphi(u, v) = f\bigl(u + h(v), g(u) + v\bigr),$$

die in dem einen oder dem anderen der rechtwinkligen Bereiche definiert ist und der Klasse C^n angehört.

Handelt es sich um den konvexen Winkel (Abb. 168a), so läßt sich nach Lemma I die Funktion φ zunächst auf den vierten Quadranten fortsetzen und dann die so erhaltene Funktion (unter Vertauschung der Rollen von u und v) auf den zweiten und dritten Quadranten, d. h. auf die volle Umgebung des Ursprungs.

Komplizierter ist es, wenn es sich um die „einspringende" Ecke (Abb. 168b) handelt. Dann geht man folgendermaßen vor. Zunächst setzt man nach Lemma I (unter

Umkehrung des Vorzeichens von u) die Funktion φ von der linken Halbebene auf die rechte fort (es geht dabei immer nur um Punkte in der Nähe des Ursprungs) und erhält so eine Funktion φ_1 in der vollen Umgebung des Ursprungs. Dann betrachtet man die Funktion $\psi = \varphi - \varphi_1$ in der unteren Halbebene und setzt sie unter Benutzung der beim Beweis von Lemma I angewendeten Methode auf die obere Halbebene fort, was eine Funktion ψ_1 in der vollen Umgebung des Ursprungs liefert. Im dritten Quadranten ist aber $\psi_1 = \psi = \varphi - \varphi_1 = 0$, und daraus folgt nach der erwähnten Methode, daß auch im zweiten Quadranten $\psi_1 = 0$ ist. Setzt man jetzt in der Umgebung des Ursprungs $\varphi^* = \psi_1 + \varphi_1$, so ist im zweiten und dritten Quadranten $\psi_1 = 0$ und $\varphi_1 = \varphi$, so daß auch $\varphi^* = \varphi$ ist, im vierten Quadranten ist $\psi_1 = \psi = \varphi - \varphi_1$ und wieder $\varphi^* = (\varphi - \varphi_1) + \varphi_1 = \varphi$. Somit ist die so konstruierte Funktion φ^* die Erweiterung von φ auf die volle Umgebung des Ursprungs.

Mit Hilfe der umgekehrten Transformation (12) erhält man die Fortsetzung

$$f^*(x, y) = \varphi^*\big(\lambda(x, y), \mu(x, y)\big)$$

der Funktion f. Der Beweis wird wie bei Satz I mit Hilfe von Lemma II zu Ende geführt.

262. Schlußbemerkungen. Der Satz über die Erweiterung von Funktionen findet vielfach Anwendung. Wir beschränken uns hier darauf, die Verallgemeinerung einer Reihe von lokalen (d. h. mit der Umgebung eines bestimmten Punktes zusammenhängenden) Sätzen und Formeln der Analysis anzugeben, die mit seiner Hilfe gewonnen werden können, und zwar auf den Fall, daß der erwähnte Punkt auf dem Rand des betrachteten Bereichs liegt und nicht im Inneren, wie gewöhnlich angenommen wird.

Beispielsweise sei in einem von einem Rand $\mathscr{L}$ des oben betrachteten Typs begrenzten abgeschlossenen Bereich $\mathscr{M}$ eine nebst ihren partiellen Ableitungen f_x und f_y stetige Funktion $z = f(x, y)$ gegeben. Dann gilt für einen inneren Punkt (x_0, y_0) von $\mathscr{M}$ nach Nr. 178 für den vollständigen Zuwachs der Funktion

$$\begin{aligned} \Delta z &= f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0) \\ &= f_x(x_0, y_0)\,\Delta x + f_y(x_0, y_0)\,\Delta y + \alpha\Delta x + \beta\Delta y, \end{aligned} \tag{13}$$

wobei α und β mit Δx und Δy gegen 0 streben. Die beim Beweis dieser Formel angewandte Methode versagt im allgemeinen, wenn (x_0, y_0) auf dem Rand liegt. Dabei gilt aber die Formel auch in diesem Fall, wenn nur Δx und Δy so gewählt werden, daß der Punkt $(x_0 + \Delta x,\ y_0 + \Delta y)$ nicht außerhalb $\mathscr{M}$ zu liegen kommt. Davon überzeugt man sich leicht, wenn man zunächst die Formel für die Fortsetzung f^* von f auf die ganze Ebene anschreibt und diese Formel dann für die Punkte von $\mathscr{M}$ auf die Ausgangsfunktion f überträgt.

In allen Fällen, in denen den Schlußfolgerungen die Formel (13) zugrunde lag, erhalten wir jetzt eine wesentliche Ausdehnung des früheren Resultates.

So existiert unter den Voraussetzungen über f das vollständige Differential (Nr. 179) nicht nur in den inneren Punkten von $\mathscr{M}$, sondern auch auf dem Rand. Für Flächen $z = f(x, y)$ beispielsweise können wir jetzt auch in Punkten ihres Randes von einer Tangentialebene (Nr. 180) sprechen.

Auf der oben angegebenen Formel beruht bekanntlich auch die Differentiationsregel für mittelbare Funktionen (Nr. 181). Besitzen die Funktionen

$$x = \varphi(t), \quad y = \psi(t) \quad (t_0 \leqq t \leqq T) \tag{14}$$

Ableitungen und liegen die Punkte $\big(\varphi(t), \psi(t)\big)$ alle im Inneren von $\mathscr{M}$, so gilt für die zusammengesetzte Funktion $z = f\big(\varphi(t), \psi(t)\big)$ die Formel $t_t = f_x x_t + f_y y_t$. Jetzt läßt sie sich auch auf den Fall ausdehnen, daß die „Kurve" (14) dem Rand von $\mathcal{M}$ beliebig nahe kommt usw.

Ohne auf Einzelheiten einzugehen, bringen wir noch ein wichtiges Beispiel. Es seien die Funktionen

$$x = \varphi(u, v), \quad y = \psi(u, v) \tag{15}$$

nebst ihren Ableitungen in einem abgeschlossenen Bereich $\mathcal{P}$ der u, v-Ebene (mit dem Rand $\mathcal{K}$) stetig, und in einem Punkt (u_0, v_0) von $\mathcal{P}$ sei die Funktionaldeterminante

$$J = \frac{D(\varphi, \psi)}{D(u, v)}$$

von 0 verschieden. Liegt (u_0, v_0) im Innern von $\mathcal{P}$, so ist das System (15) nach Satz IV aus Nr. 208 umkehrbar, so daß in der Umgebung von (x_0, y_0), wo $x_0 = \varphi(u_0, v_0)$, $y_0 = \psi(u_0, v_0)$ gilt, u und v eindeutige Funktionen von x und y sind:

$$u = \lambda(x, y), \quad v = \mu(x, y), \tag{15*}$$

die nebst ihren Ableitungen in dieser Umgebung stetig sind. Beschränkt man sich also auf Werte in der Nähe von u_0, v_0, x_0, y_0, so kann man sagen, die Beziehungen (15) und (15*) seien völlig gleichwertig. Das haben wir beispielsweise beim Beweis der Aussage benutzt, daß die Fläche

$$x = \varphi(u, v), \quad y = \psi(u, v), \quad z = \chi(u, v),$$

wobei (u, v) in $\mathcal{P}$ variiert, in der Nähe eines ihrer nicht singulären Punkte M_0 (der $u = u_0$, $v = v_0$ entspricht) durch eine explizite Gleichung beschrieben werden kann (vgl. Nr. 228). Auf Randpunkte unserer Fläche waren aber unsere Überlegungen nicht anwendbar, da in der u, v-Ebene der Punkt (u_0, v_0) nicht auf dem Rand $\mathcal{K}$ von $\mathcal{P}$ liegen durfte.

Jetzt können wir jedoch durch Benutzung der Fortsetzungen φ^* und ψ^* von φ und ψ das Resultat über die Umkehrung eines Funktionensystems auch auf den Fall ausdehnen, daß der Punkt (u_0, v_0) auf dem Rand $\mathcal{K}$ von $\mathcal{P}$ liegt. Dem an (u_0, v_0) anschließenden Teil des Bereichs $\mathcal{P}$ entspricht in der x, y-Ebene ein an (x_0, y_0) anschließender Bereich, in dem die Umkehrung ebenfalls möglich ist. Entsprechend läßt sich auch das erwähnte geometrische Ergebnis ergänzen.

Die angegebenen Beispiele reichen aus, um dem Leser die Wichtigkeit der bewiesenen Sätze sowohl für die Analysis selbst als für ihre Anwendungen vor Augen zu führen. Weitere Beispiele für die Erweiterung von Funktionen wird er in den folgenden Bänden finden.

Namen- und Sachverzeichnis