

Werk

Titel: Mathematische Annalen

Jahr: 1928

Kollektion: Mathematica

Digitalisiert: Niedersächsische Staats- und Universitätsbibliothek Göttingen

Werk Id: PPN235181684_0100

PURL: http://resolver.sub.uni-goettingen.de/purl?PPN235181684_0100

Übergeordnetes Werk

Werk Id: PPN235181684

PURL: <http://resolver.sub.uni-goettingen.de/purl?PPN235181684>

OPAC: <http://opac.sub.uni-goettingen.de/DB=1/PPN?PPN=235181684>

Terms and Conditions

The Goettingen State and University Library provides access to digitized documents strictly for noncommercial educational, research and private purposes and makes no warranty with regard to their use for other purposes. Some of our collections are protected by copyright. Publication and/or broadcast in any form (including electronic) requires prior written permission from the Goettingen State- and University Library.

Each copy of any part of this document must contain these Terms and Conditions. With the usage of the library's online system to access or download a digitized document you accept the Terms and Conditions.

Reproductions of material on the web site may not be made for or donated to other repositories, nor may be further reproduced without written permission from the Goettingen State- and University Library.

For reproduction requests and permissions, please contact us. If citing materials, please give proper attribution of the source.

Contact

Niedersächsische Staats- und Universitätsbibliothek Göttingen
Georg-August-Universität Göttingen
Platz der Göttinger Sieben 1
37073 Göttingen
Germany
Email: gdz@sub.uni-goettingen.de

MATHEMATISCHE ANNALEN

BEGRÜNDET 1868 DURCH

ALFRED CLEBSCH UND CARL NEUMANN

FORTGEFÜHRT DURCH

FELIX KLEIN

UNTER MITWIRKUNG

VON

LUDWIG BIEBERBACH, HARALD BOHR, L. E. J. BROUWER,
RICHARD COURANT, WALTHER V. DYCK, OTTO HÖLDER,
THEODOR V. KÁRMÁN, ARNOLD SOMMERFELD

GEGENWÄRTIG HERAUSGEGEBEN

VON

DAVID HILBERT

IN GÖTTINGEN

ALBERT EINSTEIN

IN BERLIN

OTTO BLUMENTHAL

IN AACHEN

CONSTANTIN CARATHÉODORY

IN MÜNCHEN

100. BAND



BERLIN

VERLAG VON JULIUS SPRINGER

1928

Inhalt des einhundertsten Bandes.

(In alphabetischer Ordnung.)

	Seite
Ackermann, W. , in Göttingen. Über die Erfüllbarkeit gewisser Zähl- ausdrücke	638
v. Alexits, G. , in Budapest. Über die Annäherung einer stetigen Funktion durch die Cesàroschen Mittel ihrer Fourierreihe	264
Baldus, R. , in Karlsruhe. Zur Axiomatik der Geometrie. I. Über Hilberts Vollständigkeitsaxiom	321
v. Behr, J. , in Köln. Mendelismus	722
Bergmann, St. , in Berlin. Zwei Sätze über Funktionen von zwei komplexen Veränderlichen	399
Bohr, H. , in Kopenhagen. Über die Verallgemeinerungen fastperiodischer Funk- tionen	357
Brunn, H. , in München. Fundamentalsatz von den Stützen eines Eigebietes .	634
Courant, R., K. Friedrichs und H. Lewy in Göttingen. Über die partiellen Differenzgleichungen der mathematischen Physik	32
Csillag, P. , in Budapest. Untersuchungen über die Borelschen Verallgemeine- rungen des Picardschen Satzes	367
Franklin, Ph. , in Cambridge (U. S. A.). A Set of Continuous Orthogonal Functions	522
Friedrichs, K., R. Courant und H. Lewy in Göttingen. Über die partiellen Differenzgleichungen der mathematischen Physik	32
Götzky, Fr. , in Bieber. Über eine zahlentheoretische Anwendung von Modul- funktionen zweier Veränderlicher	411
Größ, G. , in Berlin. Über Gewebe auf Flächen in dreidimensionalen allgemeinen metrischen Räumen	1
Haar, A. , in Szeged. Über adjungierte Variationsprobleme und adjungierte Extremalflächen	481
Hartogs, F. , in München und A. Rosenthal in Heidelberg. Über Folgen analytischer Funktionen	212
Hölder, O. , in Leipzig. Der zweite Mittelwertssatz der Integralrechnung für komplexe Größen	438
Hopf, H. , in Berlin. Zur Topologie der Abbildungen von Mannigfaltigkeiten. Erster Teil. Neue Darstellung der Theorie des Abbildungsgrades für topo- logische Mannigfaltigkeiten	579
Hurewicz, W. , in Amsterdam und K. Menger in Wien. Dimension und Zusammenhangsstufe	618
Jacob, M. , in Wien. Über den Eindeutigkeitssatz in der Theorie der verall- gemeinerten trigonometrischen Integrale	278
Jarník, V. , in Göttingen. Über Gitterpunkte in mehrdimensionalen Ellipsoiden	699
Kneser, H. , in Greifswald. Glättung von Flächenabbildungen	609
Krull, W. , in Freiburg i. Br. Galoissche Theorie der unendlichen algebraischen Erweiterungen	687
Lewy, H., R. Courant und K. Friedrichs in Göttingen. Über die partiellen Differenzgleichungen der mathematischen Physik	32

	Seite
Mahler, K. , in Göttingen. Über einen Satz von Mellin	384
Menger, K. , in Wien. Untersuchungen über allgemeine Metrik	75
Menger, K. , in Wien und W. Hurewicz in Amsterdam. Dimension und Zusammenhangsstufe	618
v. Sz. Nagy, J. , in Klausenburg. Über die charakteristischen Zahlen einer Kurve vom Maximal-Klassenindex	164
v. Sz. Nagy, J. , in Klausenburg. Über die Züge der ebenen Kurven vom Maximal-Klassenindex	179
v. Neumann, J. , in Göttingen. Zur Theorie der Gesellschaftsspiele	295
Ore, Ö. , in New Haven (U. S. A.). Abriss einer arithmetischen Theorie der Galois- schen Körper. (Erste Mitteilung)	650
Radó, T. , in Szeged (Ungarn). Über das Flächenmaß rektifizierbarer Flächen	445
Rosenthal, A. , in Heidelberg und F. Hartogs in München. Über Folgen analytischer Funktionen	212
Schmidt, R. , in Kiel. Die trigonometrische Approximation für eine Klasse von verallgemeinerten fastperiodischen Funktionen	334
Segre, B. , in Rom. Sui moduli delle curve poligonali, e sopra un complemento al teorema di esistenza di Riemann	537
Shoda, K. , in Göttingen. Über die Automorphismen einer endlichen Abelschen Gruppe	674
Szász, O. , in Frankfurt a. M. Über die Fourierschen Reihen gewisser Funk- tionenklassen	530
Szegő, G. , in Königsberg. Zur Theorie der schlichten Abbildungen	188
Trefftz, E. , in Dresden. Konvergenz und Fehlerschätzung beim Ritzschen Ver- fahren	503
Wilson, W. , in Amsterdam. Representation of Manifolds	552
Berichtigung zu den Arbeiten: J. G. van der Corput, „Zum Teilerproblem“, Math. Annalen, 98, S. 697—716, und L. W. Nieland, „Zum Kreisproblem“, in demselben Band S. 717—736	480
Berichtigung zu der Arbeit: B. L. van der Waerden, „Eine Verallgemeinerung des Bézoutschen Theorems“, Math. Annalen 99, S. 497—541	752

Über Gewebe auf Flächen in dreidimensionalen allgemeinen metrischen Räumen.

Von

Gerhard Grüß in Berlin.

Die Theorie der Kurven und Flächen in allgemeinen metrischen Räumen ist in den Grundzügen ausführlich von Herrn Finsler¹⁾ und auf anderem Wege von Herrn Berwald²⁾ entwickelt worden, während die ersten Ansätze dazu sich schon bei Bliss, Landsberg und Synge³⁾ finden. Die Untersuchungen von Herrn Berwald und Finsler sind indessen so allgemeiner Natur, daß die Anwendung der dort entwickelten Methoden auf ein spezielles differentialgeometrisches Problem als Ergänzung der allgemeinen Theorie Interesse zu verdienen scheint. Einen solchen Beitrag liefert die vorliegende Arbeit, und zwar beschäftigt sie sich mit den „Gewebe“ oder „Kurvennetzen ohne Umwege“ auf Flächen im dreidimensionalen allgemeinen metrischen Raum. Ihre Untersuchung eignet sich deshalb besonders als ein erstes Beispiel zur allgemeinen Theorie, weil man zunächst nur die elementarsten Begriffe, Bogenlänge und Winkel, benutzt

¹⁾ Paul Finsler, Über Kurven und Flächen in allgemeinen Räumen, Inaugural-Dissertation, Göttingen 1918.

²⁾ L. Berwald, Über Parallelübertragung in Räumen mit allgemeiner Maßbestimmung, Jahresber. d. Deutsch. Math. Ver. **34**, Heft 9—12, S. 213 ff. Untersuchung der Krümmung allgemeiner metrischer Räume auf Grund des in ihnen herrschenden Parallelismus, Math. Zeitschrift **25** (1926), S. 40 ff. Über zweidimensionale allgemeine metrische Räume, I. Teil, Journ. f. d. r. u. ang. Math. **156**, Heft 3, II. Teil ebenda Heft 4.

³⁾ G. A. Bliss, A generalization of the notion of angle, Trans. of the Amer. Math. Soc. **7** (1906). G. Landsberg, Über die Totalkrümmung, Jahresber. d. Deutsch. Math. Ver. **16** (1907), S. 36 und: Krümmungstheorie und Variationsrechnung, Math. Annalen **65** (1908), S. 313 ff. J. L. Synge, A generalization of the Riemannian line-element, Trans. of the Amer. Math. Soc. **27** (1925), S. 61 ff.

Vgl. auch L. Berwald, Über die erste Krümmung der Kurven bei allgemeiner Maßbestimmung, Lotos, Prag, **67/68**, S. 52 ff. P. Funk und L. Berwald: Flächeninhalt und Winkel in der Variationsrechnung, ebenda S. 45 ff.

1-5.

und so die Wirksamkeit der Theorie an einer Grundaufgabe prüft, andererseits aber auch kompliziertere Fragen anknüpfen kann. Die Arbeit bedient sich der Finslerschen Theorie, die sich für das vorliegende Problem als wirksam erweist: es zeigt sich, daß in ihrem Rahmen die Sätze über Gewebe auf Flächen im Euklidischen Raum beim Übergang zu einer allgemeinen Maßbestimmung im großen und ganzen erhalten bleiben, soweit sie nichts über die Krümmung der Gewebekurven aussagen.

Ob sich dasselbe mit Hilfe der Berwaldschen Theorie erreichen läßt, möge einstweilen dahingestellt bleiben; gewisse Gründe scheinen mir dagegen zu sprechen, daß eine Verallgemeinerung der Gewebetheorie in so einfacher Weise möglich ist wie im Anschluß an die Theorie von Herrn Finsler.

Dagegen scheinen sich die Krümmungseigenschaften der Gewebe nicht ohne weiteres übertragen zu lassen.

Die Einordnung der Sätze über „Gewebe“ bei Riemannscher Maßbestimmung in den allgemeinen Rahmen, der hier gegeben wird, deckt einige vielleicht nicht uninteressante Zusammenhänge auf.

I.

Im Raum $\{x, y, z\}$ wird durch das Grundintegral eines einfachsten regulären Variationsproblems eine Maßbestimmung eingeführt: das Bogenelement der Kurve $\{x(t), y(t), z(t)\}$ ist

$$(1) \quad ds = F\left(x(t), y(t), z(t), \frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}\right) dt;$$

die Bogenlänge einer Kurve $\{u(t), v(t)\}$ auf der Fläche

$$(2) \quad x = x(u, v), \quad y = y(u, v), \quad z = z(u, v)$$

wird durch das Integral

$$(3) \quad s = \int_{t_1}^{t_2} F\left(x(u, v), \dots, \frac{\partial x}{\partial u} \frac{du}{dt} + \frac{\partial x}{\partial v} \frac{dv}{dt}, \dots\right) dt \\ = \int_{t_1}^{t_2} f\left(u, v, \frac{du}{dt}, \frac{dv}{dt}\right) dt$$

gemessen.

Beim Übergang von den Parametern u, v zu anderen Parametern \bar{u}, \bar{v} durch die in einem Bereich $\mathfrak{B}(u, v)$ reguläre Transformation

$$(4) \quad \begin{cases} u = u(\bar{u}, \bar{v}) \\ v = v(\bar{u}, \bar{v}) \end{cases}, \quad \frac{\partial(u, v)}{\partial(\bar{u}, \bar{v})} \neq 0$$

geht das Bogenelement der Flächenkurve über in

$$ds = \bar{f}\left(\bar{u}(\tau), \bar{v}(\tau), \frac{d\bar{u}}{d\tau}, \frac{d\bar{v}}{d\tau}\right) d\tau.$$

Die Funktion $f\left(u, v, \frac{du}{dt}, \frac{dv}{dt}\right)$ und alle aus ihr durch eine reguläre Substitution (4) hervorgehenden Funktionen sind hinsichtlich der Längenmessung auf der Fläche gleichwertig; da auch die anderen geometrischen Größen (Winkel, Krümmung usw.) invariant gegenüber der Substitution (4) definiert werden, kann man $f\left(u, v, \frac{du}{dt}, \frac{dv}{dt}\right)$ die „Maßbestimmung“ der Fläche nennen.

Die Funktion $f\left(u, v, \frac{du}{dt}, \frac{dv}{dt}\right)$ möge folgende Voraussetzungen erfüllen:

Sie sei definiert für alle u, v eines Bereiches, auf dessen Untersuchung wir uns beschränken, und für alle $u' = \frac{du}{dt}$, $v' = \frac{dv}{dt}$ mit Ausnahme des Wertepaares $u' = 0$, $v' = 0$.

Sie sei definit und speziell $\neq 0$ in dem Bereich.

Sie sei positiv homogen erster Ordnung in bezug auf u', v' .

Sie besitze stetige partielle Ableitungen genügend hoher Ordnung nach allen Variablen.

Der dritten Voraussetzung zufolge gelten die Gleichungen

$$(5) \quad \left. \begin{aligned} f(u, v, ku', kv') &= kf(u, v, u', v') \\ \frac{\partial f}{\partial \alpha}(u, v, ku', kv') &= \frac{\partial f}{\partial \alpha}(u, v, u', v') \quad (\alpha = u', v') \\ \frac{\partial^2 f}{\partial \alpha \partial \beta}(u, v, ku', kv') &= \frac{1}{k} \frac{\partial^2 f}{\partial \alpha \partial \beta}(u, v, u', v') \quad (\alpha, \beta = u', v') \\ u' \frac{\partial f}{\partial u'} + v' \frac{\partial f}{\partial v'} &= f; \quad \frac{1}{v'^2} \frac{\partial^2 f}{\partial u'^2} = -\frac{1}{u'v'} \frac{\partial^2 f}{\partial u' \partial v'} = \frac{1}{u'^2} \frac{\partial^2 f}{\partial v'^2} = f_1(u, v, u', v'). \end{aligned} \right\} k > 0$$

Sämtliche über $f(u, v, u', v')$ gemachten Voraussetzungen sind auch für die aus f durch Anwendung von (4) erzeugten Funktionen erfüllt.

Durch die Gleichung

$$(6) \quad d\bar{u} \frac{\partial f}{\partial u'}(u, v, u', v') + d\bar{v} \frac{\partial f}{\partial v'}(u, v, u', v') = 0$$

wird dem gerichteten Linienelement $\{u, v, u'dt, v'dt\}$ das transversale Linienelement zugeordnet: $\{\bar{u} = u, \bar{v} = v, d\bar{u}, d\bar{v}\}$. Sind u' und v' durch die Differentialgleichungen $u' = u'(u, v)$, $v' = v'(u, v)$ einer einparametrischen Kurvenschar als Funktionen des Ortes auf der Fläche gegeben, dann stellt Gleichung (6) die Differentialgleichung der Transversalen der ersten Kurvenschar dar; ist diese speziell eine Schar von Extremalen des Variationsproblems

$$\delta \int_{t_1}^{t_2} f(u, v, u', v') dt = 0,$$

also eine Schar von ∞^1 Integralkurven der Euler-Weierstraßschen Differentialgleichung

$$(7) \quad G(u, v, u', v', u'', v'') \equiv (u'v'' - v'u'')f_1 + \frac{\partial^2 f}{\partial v' \partial u} - \frac{\partial^2 f}{\partial u' \partial v} = 0,$$

dann schneiden die Integralkurven der Gleichung (6) auf den Extremalen Kurvenstücke ab, für die das Integral

$$\int f(u, v, u', v') dt$$

denselben Wert hat. Mit anderen Worten: die Transversalen einer Schar geodätischer Linien (im Sinne der Maßbestimmung f) sind geodätische Äquidistante⁴).

Den Winkelbegriff hat Herr Finsler⁵) der allgemeinen Maßbestimmung folgendermaßen angepaßt: Der Cosinus des Schnittwinkels φ der Kurven C und Γ ist gleich dem Grenzwert des Verhältnisses der transversalen Projektion auf C einer auf Γ abgetragenen Bogenlänge zu dieser Bogenlänge selbst, für den Fall, daß die projizierte Bogenlänge gegen Null konvergiert. „Transversale Projektion“ soll heißen, daß die projizierende Kurve Transversale von C ist. Aus $\cos \varphi$ folgt φ durch die Bestimmung, daß $|\varphi|$ möglichst klein und $\text{sign} \varphi$ willkürlich ist. Herr Finsler zeigt, daß $\cos \widehat{C\Gamma} = \cos \widehat{\Gamma C}$ nur für den Fall der Riemannschen Maßbestimmung $f = \sqrt{Eu'^2 + 2Fu'v' + Gv'^2}$ gilt, oder wenn sich die Kurven C und Γ berühren: dann ist $\cos \widehat{C\Gamma} = \cos \widehat{\Gamma C} = 1$. Da wir es hier mit einer allgemeinen Maßbestimmung zu tun haben, und auch von Null verschiedene Winkel auftreten werden, muß eine Verabredung getroffen werden, aus der hervorgeht, welcher von den beiden Schenkeln eines Winkels jeweils zu bevorzugen ist. In den folgenden Untersuchungen spielen meist nur Paare von Winkeln eine Rolle, die einen Schenkel gemeinsam haben. Demgemäß verabreden wir, daß bei solchen benachbarten Winkeln der gemeinsame Schenkel bevorzugt sein soll: seine Transversale soll die projizierende Kurve sein, auf den anderen Schenkeln werden die zu projizierenden Längen abgetragen. Wir wollen dafür kurz sagen: „Die Winkel werden zum gemeinsamen Schenkel hin gemessen.“ Wenn ausnahmsweise auch einzelne Winkel auftreten, wird die Abhängigkeit von der Reihenfolge der Schenkel beachtet werden.

Sind die Gleichungen der sich schneidenden Kurven

$$(C): \quad u = u_0(t), \quad v = v_0(t)$$

⁴) Vgl. z. B. A. Kneser, Lehrbuch der Variationsrechnung, 2. Aufl. 1925, S. 138ff., sowie C. Carathéodory, Acta Math. 47 (1925), Heft 3.

⁵) Finsler, a. a. O. S. 38 ff.

und

$$(\Gamma): \quad u = u_1(\tau), \quad v = v_1(\tau),$$

und ist φ der von der Kurve (Γ) zur Kurve (C) hin gemessene Winkel, dann ist nach der Finslerschen Definition ⁶⁾

$$(8) \quad \cos \varphi = \frac{u'_1 f_{u'}(u, v, u'_0, v'_0) + v'_1 f_{v'}(u, v, u'_0, v'_0)}{f(u, v, u'_1, v'_1)} \quad \left(f_{u'} = \frac{\partial f}{\partial u'}, f_{v'} = \frac{\partial f}{\partial v'} \right).$$

Man bemerke, daß der Winkel zwischen zwei Richtungen derselbe ist, wenn man ihn im Sinne der Maßbestimmung f und in dem der Maßbestimmung $P(u, v) \cdot f$ ($P \neq 0$) mißt. Im folgenden wird verschiedentlich ein solcher „Multiplikator“ P benutzt werden, der eine Funktion allein des Ortes auf der Fläche ist. Er soll durch die Forderung bestimmt werden, daß eine beliebig vorgegebene Schar von ∞^1 Kurven eine Schar geodätischer Linien ist im Sinne der Maßbestimmung $P(u, v) \cdot f(u, v, u', v')$. Die Frage nach diesem Multiplikator $P(u, v)$ ist also ein spezieller Fall des von Darboux ⁷⁾ allgemein gelösten „inversen Problems der Variationsrechnung“. Sei die Kurvenschar durch ihre Differentialgleichungen

⁶⁾ Die Größe $\cos \varphi$ läßt sich an der Eichkurve der f -Geometrie, d. h. der Indikatrix des zugehörigen Variationsproblems, geometrisch deuten. Ihre Gleichung ist bei $u' = \lambda \xi$, $v' = \lambda \eta$ in kartesischen Koordinaten $f(u, v, \xi, \eta) = 1$, wobei $\lambda > 0$ aus dieser Gleichung folgt, oder in Polarkoordinaten ($\varrho \cos \vartheta = \xi$, $\varrho \sin \vartheta = \eta$) $f(u, v, \cos \vartheta, \sin \vartheta) = 1$, ϱ ; u und v sind Parameter. Der Cosinus des von der Richtung (u', v') zur Richtung (u'_0, v'_0) gemessenen Winkels φ ist bei (s. Figur)

$$u' = \lambda \xi = \lambda \varrho \cos \vartheta, \quad v' = \lambda \eta = \lambda \varrho \sin \vartheta, \\ u'_0 = \lambda_0 \xi_0 = \lambda_0 \varrho_0 \cos \vartheta_0, \quad v'_0 = \lambda_0 \eta_0 = \lambda_0 \varrho_0 \sin \vartheta_0,$$

$$\cos \varphi = \xi \cdot f_{\xi}(\xi_0, \eta_0) + \eta \cdot f_{\eta}(\xi_0, \eta_0),$$

$$\cos \varphi = \sqrt{f_{\xi}^2(\xi_0, \eta_0) + f_{\eta}^2(\xi_0, \eta_0)} \cdot \varrho(\xi, \eta) \cos \alpha.$$

Wegen $\xi \cdot f_{\xi} + \eta \cdot f_{\eta} = 1$ ist $\sqrt{f_{\xi}^2(\xi_0, \eta_0) + f_{\eta}^2(\xi_0, \eta_0)} = 1$, $\varrho(\xi_0, \eta_0) \cos \varepsilon$ und endlich $\cos \varphi = QO : Q_0O$. Wenn die Indikatrix überall konvex ist, folgt aus $\varphi = 0$ $Q = Q_0$ und $P = P_0$: Zwei Linienelemente sind identisch, wenn sie den Winkel Null bilden. Ist die Indikatrix nicht durchweg konvex, dann gilt dies nur bei Beschränkung auf eine genügend kleine Umgebung der einen Richtung. Herr Finsler macht diese Einschränkung nicht. Die Konstruktion der Transversalenrichtung ist eindeutig; die Konstruktion zweier Richtungen, die mit einer gegebenen Richtung einen vorgeschriebenen Winkel bilden, nur dann, wenn die Indikatrix konvex ist; Entsprechendes gilt für die Konstruktion der Winkelhalbierenden zweier Richtungen.

⁷⁾ Darboux, *Théorie des surfaces* 3, Nr. 604, 605. O. Bolza, *Vorlesungen über die Variationsrechnung*, S. 37.

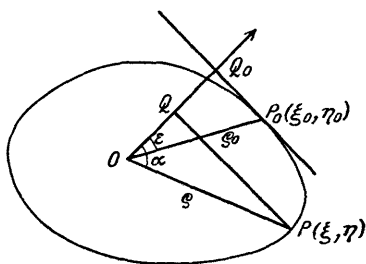


Fig. 1.

$$(9) \quad \begin{cases} \frac{du}{dt} = u'(u, v) \\ \frac{dv}{dt} = v'(u, v) \end{cases}$$

gegeben, wobei t einen beliebigen Parameter auf den Kurven bedeute. Die Integralkurven dieser Gleichung sind dann und nur dann geodätische Linien im Sinne der Maßbestimmung $P \cdot f$, wenn ihre Gleichungen die Euler-Weierstraßsche Differentialgleichung (7) $G = 0$ befriedigen, in der f durch $P \cdot f$ zu ersetzen ist. Sie stellt folgende partielle Differentialgleichung für P dar:

$$P(u, v) \cdot \left\{ f_1 \left[u'^2 \frac{\partial v'}{\partial u} + u' v' \left(\frac{\partial v'}{\partial v} - \frac{\partial u'}{\partial u} \right) - v'^2 \frac{\partial u'}{\partial v} \right] + \frac{\partial^2 f}{\partial v' \partial u} - \frac{\partial^2 f}{\partial u' \partial v} \right\} \\ + \frac{\partial P}{\partial u} \cdot \frac{\partial f}{\partial v'} - \frac{\partial P}{\partial v} \cdot \frac{\partial f}{\partial u'} = 0$$

oder

$$(10) \quad G \cdot P(u, v) + \frac{\partial P}{\partial u} \cdot f_v' - \frac{\partial P}{\partial v} \cdot f_u' = 0,$$

wobei in G die Funktion f einzusetzen ist. Sind die Integralkurven der Gleichung (9) Extremalen im Sinne der Maßbestimmung f , dann folgt aus (10) und (6)

$$\frac{\partial P}{\partial u} d\bar{u} + \frac{\partial P}{\partial v} d\bar{v} = dP = 0$$

längs der Transversalen der Kurvenschar, d. h. die Kurven $P = \text{konst.}$ sind Äquidistanten im f -Sinne, was unmittelbar einleuchtet.

II.

Entsprechend der Definition der Kurvennetze ohne Umwege auf Flächen im Euklidischen Raum⁸⁾ wird der Begriff der Gewebe beim Übergang zur allgemeinen Maßbestimmung (3) folgendermaßen gefaßt: Ein Kurvennetz heißt „ohne Umwege im Sinne der Maßbestimmung f “ oder kurz ein f -Gewebe, wenn die im f -Sinne gemessene Bogenlänge eines beliebigen geschlossenen nur aus Netzkurven bestehenden Weges null ist, oder, was offenbar dasselbe ist, wenn die auf Kurven des Netzes gemessene Entfernung irgend zweier Punkte unabhängig ist von der speziellen Wahl des

⁸⁾ Ausführliche Literatur vgl. G. Scheffers, Einführung in die Theorie der Kurven in der Ebene und im Raum, 2. Aufl., 1 (1921), S. 181. Der Name „Kurvennetz ohne Umwege“ zuerst bei G. Scheffers, Ebene Kurvennetze ohne Umwege, Berichte über die Verh. d. Kgl. Sächs. Ges. d. Wissensch. zu Leipzig, math.-phys. Kl., 57 (1905), S. 353. Die charakteristische Eigenschaft dieser Netze kommt schon vor bei R. Rothe, Bemerkungen über ein spezielles krummliniges Koordinatensystem, Sitzungsber. d. B. M. G. 1 (1902), S. 47.

verbindenden Weges. Voraussetzung ist dabei, daß die Netzkurven in bestimmter Weise orientiert sind: beim Durchlaufen der Kurven im einen oder anderen Sinne ist die Bogenlänge positiv bzw. negativ zu zählen, was stets möglich ist, da $f(u, v, u', v')$ definit ist. Das Netz

$$(11) \quad \begin{cases} u = u(\alpha, \beta) \\ v = v(\alpha, \beta) \end{cases} \quad \frac{\partial(u, v)}{\partial(\alpha, \beta)} \neq 0$$

ist also dann und nur dann ein Gewebe, wenn für jeden geschlossenen Weg aus Kurven $\alpha = \text{konst.}$ bzw. $\beta = \text{konst.}$

$$\int_{\mathcal{C}} ds_{\alpha, \beta} \equiv \int_{\mathcal{C}} \left[f\left(u, v, \frac{\partial u}{\partial \alpha}, \frac{\partial v}{\partial \alpha}\right) d\alpha + f\left(u, v, \frac{\partial u}{\partial \beta}, \frac{\partial v}{\partial \beta}\right) d\beta \right] = 0$$

oder

$$(12) \quad J \equiv \frac{\partial}{\partial \alpha} f\left(u, v, \frac{\partial u}{\partial \beta}, \frac{\partial v}{\partial \beta}\right) - \frac{\partial}{\partial \beta} f\left(u, v, \frac{\partial u}{\partial \alpha}, \frac{\partial v}{\partial \alpha}\right) = 0$$

gilt. Eine leichte Rechnung zeigt, daß diese Bedingung mit der von Herrn Rothe⁹⁾ angegebenen übereinstimmt, wenn

$$f(u, v, u', v') = \sqrt{E u'^2 + 2F u' v' + G v'^2}$$

ist. Die Integrabilitätsbedingung läßt sich durch Einführung der Winkelhalbierenden

$$(13) \quad \begin{cases} u = u_0(t, A) \\ v = v_0(t, A) \end{cases}$$

der Netzkurven auf eine für die Untersuchung der Gewebeeigenschaft brauchbarere Form bringen. Dabei hat man unter den Kurven (13) entsprechend der Finslerschen Winkeldefinition und der Verabredung, die Winkelpaare zum gemeinsamen Schenkel hin zu messen, die Integralkurven folgender Differentialgleichung zu verstehen:

$$(14) \quad \frac{u_\alpha f_{u'}(u, v, u'_0, v'_0) + v_\alpha f_{v'}(u, v, u'_0, v'_0)}{f(u, v, u_\alpha, v_\alpha)} = \frac{u_\beta f_{u'}(u, v, u'_0, v'_0) + v_\beta f_{v'}(u, v, u'_0, v'_0)}{f(u, v, u_\beta, v_\beta)},$$

in der

$$u_\alpha = \frac{\partial u}{\partial \alpha}, \quad v_\alpha = \frac{\partial v}{\partial \alpha}, \quad u_\beta = \frac{\partial u}{\partial \beta}, \quad v_\beta = \frac{\partial v}{\partial \beta}$$

den Gleichungen (11) zufolge Funktionen von u und v sind. Wegen der positiven Homogenität nullter Ordnung von $f_{u'}$, $f_{v'}$ hinsichtlich u' , v' erhält man aus (14)

$$\frac{du_0}{dv_0} = \lambda(u, v),$$

⁹⁾ R. Rothe, Über die Bekleidung einer Fläche mit einem Gewebe („Kurvennetz ohne Umwege“), Sitzungsber. d. B. M. G., VII. Jahrg. (1907), 1. Stück, S. 15.

wo $\lambda(u, v)$ eine i. a. mehrdeutige Funktion des Ortes (u, v) bedeutet; der Grad der Mehrdeutigkeit ist, wenn die Maßbestimmung in den Variablen u' und v' algebraisch ist, die von Landsberg¹⁰⁾ eingeführte charakteristische Zahl des Feldes. In dem vorliegenden Falle genügt es, einen eindeutigen Zweig $\lambda_0(u, v)$ von $\lambda(u, v)$ herauszugreifen und durch Integration der Gleichung

$$\frac{d u_0}{d v_0} = \lambda_0(u, v)$$

in einem Bereich der u - v -Ebene, in dem $\lambda(u, v)$ keinen Verzweigungspunkt besitzt, eine Schar von Winkelhalbierenden des Netzes zu konstruieren, die durch Gleichung (13) dargestellt sein möge. Der Winkel ω zwischen den Kurven (13) und den Netzkurven, der der „halbe Maschenwinkel“ des Netzes heißen soll, ist eine durch die Gleichung (14) wohlbestimmte Funktion von u und v :

$$(15) \quad \cos \omega = \frac{u_\alpha f_{u'}(u, v, u'_0, v'_0) + \dots}{f(u, v, u_\alpha, v_\alpha)} = \frac{u_\beta f_{u'}(u, v, u'_0, v'_0) + \dots}{f(u, v, u_\beta, v_\beta)}.$$

Man führe nun in die Integrabilitätsbedingung (12) aus Gleichung (15)

$$f(u, v, u_\alpha, v_\alpha) = \frac{1}{\cos \omega} [u_\alpha f_{u'}(u, v, u'_0, v'_0) + v_\alpha f_{v'}(u, v, u'_0, v'_0)],$$

$$f(u, v, u_\beta, v_\beta) = \frac{1}{\cos \omega} [u_\beta f_{u'}(u, v, u'_0, v'_0) + v_\beta f_{v'}(u, v, u'_0, v'_0)]$$

ein; so erhält man nach leichter Zwischenrechnung

$$(16) \quad J = \frac{\partial(u, v)}{\partial(\alpha, \beta)} \left[\frac{G(u_0, v_0)}{\cos \omega} + \left(\frac{\partial \omega}{\partial u} f_{v'}(u, v, u'_0, v'_0) - \frac{\partial \omega}{\partial v} f_{u'}(u, v, u'_0, v'_0) \right) \frac{d}{d\omega} \left(\frac{1}{\cos \omega} \right) \right],$$

worin $G(u_0, v_0)$ die linke Seite der Gleichung (7) ist, genommen für die Funktionen $u = u_0(t, A)$, $v = v_0(t, A)$, d. h. für die Winkelhalbierenden. Wegen

$$\frac{\partial(u, v)}{\partial(\alpha, \beta)} \neq 0$$

geht, wenn noch

$$\frac{1}{\cos \omega} = p(u, v)$$

gesetzt wird, die Integrabilitätsbedingung (12) über in

$$(17) \quad p \cdot G(u_0, v_0) + \frac{\partial p}{\partial u} f_{v'}(u, v, u'_0, v'_0) - \frac{\partial p}{\partial v} f_{u'}(u, v, u'_0, v'_0) = 0.$$

Für die eben durchgeführte Umformung der Gleichung (12) ist $\cos \omega \neq 0$

¹⁰⁾ G. Landsberg, Über die Krümmung in der Variationsrechnung, Math. Annalen 65 (1908), S. 321.

in dem ganzen Bereich erforderlich. Dies ist indessen sicherlich erfüllt, da aus $\cos \omega = 0$

$$u_\alpha f_{u'}(\dots u'_0, v'_0) + v_\alpha f_{v'}(\dots u'_0, v'_0) = 0,$$

$$u_\beta f_{u'}(\dots u'_0, v'_0) + v_\beta f_{v'}(\dots u'_0, v'_0) = 0$$

und weiter

$$\frac{\partial(u, v)}{\partial(\alpha, \beta)} = 0$$

folgen würde, weil $f_{u'}$ und $f_{v'}$ wegen $f \neq 0$ nicht zugleich verschwinden können.

Die vorbereitenden Betrachtungen mögen durch die Einführung einiger neuer Bezeichnungen abgeschlossen werden, die eine kürzere Formulierung der Resultate gestatten.

1. In ihrem Schnittpunkt schließen je eine α -Kurve und eine β -Kurve des Netzes (11) den gleichen Winkel ω mit der durch den Schnittpunkt gehenden Kurve der Schar (13) ein. In seiner Nähe sind die α - und β -Kurven also „Spiegelbilder“ voneinander bezüglich der Integralkurve $\{u_0, v_0\}$ von (14). Dafür werde kurz gesagt: Das α - β -Netz ist *spiegelbildlich*¹¹⁾ bezüglich der Schar $\{u_0, v_0\}$ der Winkelhalbierenden.

2. Weiter soll ein Netz (11), das zu der Schar

$$u = u_0(t, A), \quad v = v_0(t, A)$$

spiegelbildlich ist, bezüglich der Schar

$$u = u_1(\tau, B), \quad v = v_1(\tau, B)$$

*gestreift*¹²⁾ heißen, wenn der Winkel ω längs der Kurven $B = \text{konst.}$ konstant ist.

3. Endlich denke man sich zu einer vorgegebenen Schar von ∞^1 Kurven eine zweite Schar so konstruiert, daß ihre Elemente auf denen der ersten Schar gleiche Stücke im Sinne der Maßbestimmung f abschneiden; etwa dadurch, daß man die erste Schar durch eine beliebige Kurve schneidet, so daß der Cosinus des Winkels stets $\neq 0$, $\neq 1$ ist, und von der Schnittkurve aus auf den Kurven der ersten Schar gleiche Längen abträgt und dann entsprechende Punkte verbindet. Dann sollen die Kurven der zweiten Schar schiefwinklige Äquidistante oder kürzer *schiefe Äquidistante* der ersten Schar heißen, wenn sie nicht zugleich Transversalen sind, andernfalls schlechthin Äquidistante. Das Adjektiv „schiefwinklig“ soll darauf

¹¹⁾ Vgl. R. v. Lilienthal, Vorlesungen über Differentialgeometrie 2, S. 245.

¹²⁾ Vgl. R. Rothe, Über die Bekleidung einer Oberfläche mit einem biegsamen unausdehnbaren Netz, Sitzungsber. d. B. M. G. 5 (1906), 1. Stück, S. 9.

hinweisen, daß die Gleichung (6) nicht erfüllt ist für die Linienelemente der ersten und zweiten Schar, d. h. daß $\cos \omega \neq 0$ ist, und ferner, daß $\cos \omega \neq 1$ ist.

Nach diesen Vorbereitungen lassen sich folgende Sätze über gestreifte Kurvennetze ohne Umwege aussprechen.

III.

Satz Ia. *Ein zu einer einparametrischen Extremalenschar*

$$(18) \quad \begin{cases} u = u_0(t, A) \\ v = v_0(t, A) \end{cases}$$

spiegelbildliches Netz

$$(19) \quad \begin{cases} u = u(\alpha, \beta) \\ v = v(\alpha, \beta) \end{cases}$$

ist ein Gewebe, wenn es bezüglich der Transversalen

$$(20) \quad \begin{cases} u = u_1(\tau, B) \\ v = v_1(\tau, B) \end{cases}$$

gestreift ist.

Satz Ib. *Wenn ein Gewebe (19) zur Extremalenschar (18) spiegelbildlich ist, dann ist es bezüglich der zu (18) transversalen Schar (20) gestreift.*

Satz Ic. *Wenn ein Gewebe bezüglich der Transversalen seiner Winkelhalbierenden gestreift ist, dann sind die Winkelhalbierenden Extremalen.*

Die Beweise folgen sofort aus den Gleichungen

$$(6) \quad du_1 f_u'(u, v, u'_0, v'_0) + dv_1 f_v'(u, v, u'_0, v'_0) = 0$$

und

$$(17) \quad J = \frac{G(u_0, v_0)}{\cos \omega} + \frac{\partial}{\partial u} \left(\frac{1}{\cos \omega} \right) f_u'(u, v, u'_0, v'_0) - \frac{\partial}{\partial v} \left(\frac{1}{\cos \omega} \right) f_v'(u, v, u'_0, v'_0) = 0;$$

denn aus

$$(21) \quad G(u_0, v_0) = 0$$

und

$$(22) \quad \frac{\partial}{\partial u} \left(\frac{1}{\cos \omega} \right) du_1 + \frac{\partial}{\partial v} \left(\frac{1}{\cos \omega} \right) dv_1 = 0$$

folgt (17), (Beweis von Satz Ia).

Aus (17) und (21) folgt (22), (Beweis von Satz Ib).

Aus (17) und (22) folgt (21), weil

$$\frac{1}{\cos \omega} = \frac{f(u, v, u_a, v_a)}{u_a f_u'(\dots u'_0, v'_0) + v_a f_v'(\dots u'_0, v'_0)} \neq 0$$

ist (Beweis von Satz Ic).

Der Inhalt des Satzes I kann durch einen geometrischen Beweis unmittelbar veranschaulicht werden, indem man die ursprüngliche Definition des Cosinus als Limes des Verhältnisses zweier (verallgemeinerter) Bogenlängen $\int ds = \int f dt$ benutzt.

Um Satz Ia auf diesem Wege zu beweisen, zeigt man zunächst, daß folgender Hilfssatz gilt:

Die Transversalen einer Extremalenschar schneiden auf den beiden Kurvenscharen eines Netzes, das zur Extremalenschar spiegelbildlich und bezüglich der Transversalenschar gestreift ist, gleiche Längen ab, d. h. sie sind schiefe Äquidistante „des Netzes“.

Beweis. $\{u_1(\tau, B), v_1(\tau, B)\}$ und $\{u_1(\tau, B + dB), v_1(\tau, B + dB)\}$ seien zwei benachbarte Transversalen, B bedeute den auf den Extremalen gemessenen Abstand der ersten Transversale von einer willkürlichen 0-Transversalen. Die beiden Kurven schneiden auf zwei beliebigen Kurven $\beta = \beta_1$ und $\alpha = \alpha_1$ des Netzes $u = u(\alpha, \beta)$, $v = v(\alpha, \beta)$ Bogenlängen $\int f dt$ ab, deren Differentiale sind:

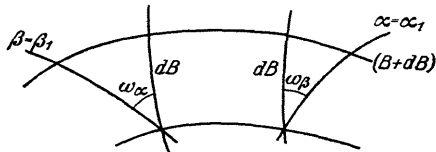


Fig. 2.

$$ds_\alpha = ds_{\beta=\beta_1} = \frac{d}{dB} \left[\int f(u(\alpha, \beta_1), v(\alpha, \beta_1), \frac{\partial u}{\partial \alpha}, \frac{\partial v}{\partial \alpha}) d\alpha \right] dB,$$

$$ds_\beta = ds_{\alpha=\alpha_1} = \frac{d}{dB} \left[\int f(u(\alpha_1, \beta), v(\alpha_1, \beta), \frac{\partial u}{\partial \beta}, \frac{\partial v}{\partial \beta}) d\beta \right] dB,$$

wo die Integrale über die den Schnittpunkten der Transversalen mit den Netzkurven entsprechenden α - bzw. β -Bereiche zu erstrecken sind. Hinsichtlich der Winkel ω_α und ω_β gilt

$$\cos \omega_\alpha = \frac{dB}{ds_\alpha}, \quad \cos \omega_\beta = \frac{dB}{ds_\beta}.$$

Aus der Voraussetzung $\omega_\alpha = \omega_\beta$ folgt

$$\int_{B_1}^{B_2} \frac{d}{dB} \left[\int f(u, v, u_\alpha, v_\alpha) d\alpha \right] dB = \int_{B_1}^{B_2} \frac{d}{dB} \left[\int f(u, v, u_\beta, v_\beta) d\beta \right] dB,$$

w. z. b. w.

Ebenso schließt man, daß die Transversalen auf zwei Kurven $\alpha = \alpha_1$ und $\alpha = \alpha_2$ des Netzes gleiche Längen abschneiden. Daraus folgt offensichtlich sofort, daß das betrachtete Netz ein Gewebe ist, also der Satz Ia. Denn es ist $s_1 + s_2 = s_3 + s_4$ für jedes einfache Netzviereck.

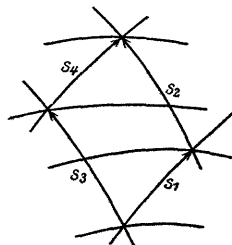


Fig. 3.

Um auch (Ib) und (Ic) geometrisch zu beweisen, scheint man einen Satz benutzen zu müssen, der insofern zu allgemein ist, als er von allgemeinen, nicht notwendig gestreiften Geweben handelt. Es liegt dies daran, daß die Voraussetzungen von (Ib): Das Gewebe ist spiegelbildlich zu einer Extremalenschar, und (Ic): Das Gewebe ist zur Transversalenschar seiner Winkelhalbierenden gestreift — nicht geometrisch-anschaulich interpretiert werden können. Man benutzt also den

Satz II. *Die Transversalen der Winkelhalbierenden eines Gewebes sind schiefe Äquidistante dieses Netzes.*

Beweis. Die Winkelhalbierenden eines Gewebes genügen der Gleichung (17), sie sind also, der Gleichung (10) zufolge, Extremalen des Variationsproblems

$$\int_{t_1}^{t_2} f(u, v, u', v') \cdot p(u, v) dt = \int_{t_1}^{t_2} \frac{f(u, v, u', v')}{\cos \omega} dt = \text{Extr.},$$

und ihre Transversalen sind geodätische Äquidistante im Sinne der Maßbestimmung $\frac{f}{\cos \omega}$; also ist längs irgendeiner Transversalen, die wie oben durch den Parameter B bestimmt sein möge,

$$\frac{d}{dB} \left(\int_{t(B_0)}^{t(B)} \frac{f(u, v, u'_0, v'_0)}{\cos \omega} dt \right) = \frac{f(u, v, u'_0, v'_0)}{\cos \omega} \frac{dt}{dB}$$

konstant, wobei ω der wohlbestimmte halbe Maschenwinkel des gegebenen Netzes ist. Nach seiner Definition ist nun

$$\cos \omega = \frac{f(u_0, v_0, u'_0, v'_0) \frac{dt}{dB}}{f(u, v, u_\alpha, v_\alpha) \frac{d\alpha}{dB}}$$

und auch

$$\cos \omega = \frac{f(u_0, v_0, u'_0, v'_0) \frac{dt}{dB}}{f(u, v, u_\beta, v_\beta) \frac{d\beta}{dB}}.$$

Also folgt, daß $f(u, v, u_\alpha, v_\alpha) \frac{d\alpha}{dB}$ und $f(u, v, u_\beta, v_\beta) \frac{d\beta}{dB}$ längs jeder Transversalen konstant sind, d. h. aber, die Transversalen sind schiefe Äquidistante des Netzes. Sind die Winkelhalbierenden nun Extremalen (Voraussetzung des Satzes Ib), dann gilt außer

$$f(u, v, u_\alpha, v_\alpha) \frac{d\alpha}{dB}, \quad f(u, v, u_\beta, v_\beta) \frac{d\beta}{dB} = \text{konst.}$$

auch

$$f(u_0, v_0, u'_0, v'_0) \frac{dt}{dB} = \text{konst.}$$

längs der Transversalen, also ist ω längs dieser Kurven konstant, d. h. das Gewebe ist gestreift. (I b.)

Endlich Satz I c. Die Transversalen sind geodätische Äquidistante im Sinne der Maßbestimmung $\frac{1}{\cos \omega} \cdot f(u, v, u', v')$, d. h. es ist

$$\frac{f(u_0(t_1, A_1), v_0(t_1, A_1), u'_0, v'_0)}{\cos \omega_1} \frac{dt_1}{dB} = \frac{f(u_0(t_2, A_2), v_0(t_2, A_2), u'_0, v'_0)}{\cos \omega_2} \frac{dt_2}{dB}.$$

Da das Gewebe gestreift ist ($\omega_1 = \omega_2 = \dots$), sind die Transversalen im Sinne der Maßbestimmung f selbst äquidistant und also sind ihre Gefällkurven, nämlich die Winkelhalbierenden des Gewebes, Extremalen¹³).

IV.

Satz I beschäftigte sich nur mit gestreiften Geweben: die Integrabilitätsbedingung zerfiel in zwei Gleichungen. Sieht man von dieser Möglichkeit ab, dann erhält man als Zusammenfassung der Integrabilitätsbedingung (17) und der Gleichung (10) folgendes Merkmal für allgemeine Gewebe:

Ein Netz ist dann und nur dann ein Gewebe, wenn seine Winkelhalbierenden geodätische Linien im Sinne der Maßbestimmung $\frac{1}{\cos \omega} \cdot f(u, v, u', v')$ sind, wobei $\omega(u, v)$ den halben Maschenwinkel des Netzes bedeutet.

Satz II ist die geometrische Interpretation der im vorhergehenden enthaltenen notwendigen Bedingung. Die hinreichende Bedingung läßt sich folgendermaßen als Konstruktionsvorschrift für beliebige Gewebe geometrisch formulieren.

Satz III. *Man konstruiere zu einer beliebig gegebenen einparametrischen Kurvenschar*

$$(23a) \quad \begin{cases} u = U(\alpha, C) \\ v = V(\alpha, C) \end{cases} \quad (C \text{ Scharparameter, } \alpha \text{ Parameter auf den Kurven der Schar)}$$

eine beliebige Schar schiefwinkliger Äquidistanten

$$(24) \quad \begin{cases} u = u_1(\tau, B) \\ v = v_1(\tau, B) \end{cases}$$

bestimme mit Hilfe der Differentialgleichung

$$(6) \quad \frac{du_1}{d\tau}(u, v) f_u(u, v, u'_0, v'_0) + \frac{dv_1}{d\tau}(u, v) f_v(u, v, u'_0, v'_0) = 0$$

¹³ C. Carathéodory, Die Methode der geodätischen Äquidistanten und das Problem von Lagrange, Acta Mathem. 47 (1925), Heft 3.

irgendeine Schar¹⁴⁾

$$(25) \quad \begin{cases} u = u_0(t, A) \\ v = v_0(t, A), \end{cases}$$

deren Transversalen also die Kurven (24) sind, und suche die Kurvenschar

$$(23b) \quad \begin{cases} u = \bar{U}(\beta, D) \\ v = \bar{V}(\beta, D), \end{cases}$$

die mit (23a) zusammen spiegelbildlich zu (25) ist. Dann bildet das Netz (23a, 23b) ein Gewebe.

Beweis. Zunächst bilden die Scharen (23a, 23b) wirklich ein Kurvennetz; denn da die Kurvenscharen (23a) und (23b) bezüglich der Schar (25) spiegelbildlich sind, ist $\frac{\partial \bar{u}}{\partial \beta} : \frac{\partial \bar{v}}{\partial \beta}$ eine von $\frac{\partial U}{\partial \alpha} : \frac{\partial V}{\partial \alpha}$ verschiedene Lösung der Gleichung

$$\frac{U' f_u'(u_0, v_0, u'_0, v'_0) + V' f_v'(u_0 \dots)}{f(u, v, U, V')} = \frac{\frac{\partial U}{\partial \alpha} f_u'(u_0, v_0, u'_0, v'_0) + \frac{\partial V}{\partial \alpha} f_v'(u_0 \dots)}{f\left(u, v, \frac{\partial U}{\partial \alpha}, \frac{\partial V}{\partial \alpha}\right)};$$

eine solche Lösung $U' : V'$ wird es, von speziellen Funktionen f abgesehen, nur dann nicht geben, wenn

$$\frac{\partial U}{\partial \alpha} f_u'(u_0, v_0, u'_0, v'_0) + \frac{\partial V}{\partial \alpha} f_v'(u_0, v_0, u'_0, v'_0) = 0$$

ist, was durch die andere Voraussetzung, daß die Transversalen der Kurven $\{u_0, v_0\}$ schiefe Äquidistante der Schar (23a) sind, ausgeschlossen ist. Also ist

$$\delta = \frac{\partial U}{\partial \alpha} \frac{\partial \bar{V}}{\partial \beta} - \frac{\partial V}{\partial \alpha} \frac{\partial \bar{U}}{\partial \beta} \neq 0,$$

und man kann daher das Kurvennetz (23a, 23b) so darstellen:

$$(26) \quad \begin{cases} u = u(\alpha, \beta) \\ v = v(\alpha, \beta). \end{cases}$$

Um die Gewebeeigenschaft dieses Netzes zu beweisen, geht man so vor. Auf den schiefen Äquidistanten $u = u_1(\tau, B)$, $v = v_1(\tau, B)$ der Kurven $u = u(\alpha, \beta = \text{konst.})$, $v = v(\alpha, \beta = \text{konst.})$ führe man als den die einzelne Kurve charakterisierenden Parameter B den Kurvenparameter α der Kurven $\beta = \text{konst.}$ ein, und als Parameter τ auf der einzelnen $B = \alpha = \text{konst.}$ Kurve den Scharparameter β der anderen Schar. Das geschieht durch Elimination von α aus (26) und τ aus (24), wodurch man $\beta = \beta(u, v)$,

¹⁴⁾ Hinsichtlich der Mehrdeutigkeit gilt dasselbe, was auf S. 8 über die Konstruktion der Halbierenden eines beliebigen Winkels gesagt wurde.

$B = B(u, v)$ erhält. Die Auflösung dieser Gleichungen nach u, v ist wegen

$$\frac{\partial u_1}{\partial \tau} \frac{\partial v}{\partial \alpha} - \frac{\partial v_1}{\partial \tau} \frac{\partial u}{\partial \alpha} \neq 0$$

möglich und ergibt, wenn man noch B durch α ersetzt, die Gleichung

$$(27) \quad \begin{cases} u = U(\alpha, \beta) \\ v = V(\alpha, \beta) \end{cases}$$

für die Schar $\beta = \text{konst.}$ und ihre schiefen Äquidistanten $\alpha = \text{konst.}$

Die Bedingung der schiefen Äquidistanz ist:

$$\int_{\alpha_1}^{\alpha} f\left(U, V, \frac{\partial U}{\partial \alpha}, \frac{\partial V}{\partial \alpha}\right) d\alpha \Big|_{\beta=\beta_1} = \int_{\alpha_1}^{\alpha} f\left(U, V, \frac{\partial U}{\partial \alpha}, \frac{\partial V}{\partial \alpha}\right) d\alpha \Big|_{\beta=\beta_2}$$

für jedes α, β_1 und β_2 ; also gilt

$$(28) \quad \frac{\partial}{\partial \beta} f\left(U(\alpha, \beta), V(\alpha, \beta), \frac{\partial U}{\partial \alpha}(\alpha, \beta), \frac{\partial V}{\partial \alpha}(\alpha, \beta)\right) = 0.$$

Um zu zeigen, daß die Scharen $\alpha = \text{konst.}$ auf den Integralkurven

$$u = u_0(t, A), \quad v = v_0(t, A)$$

der Differentialgleichung

$$\frac{\partial U}{\partial \beta} f_u(u, v, u'_0, v'_0) + \frac{\partial V}{\partial \beta} f_v(u, v, u'_0, v'_0) = 0$$

gleiche Längen $\int \frac{f}{\cos \omega} dt$ abschneiden, muß die Relation

$$\frac{\partial}{\partial \beta} \int_{t_1}^{t_2} \frac{f(u_0, v_0, u'_0, v'_0)}{\cos \omega} dt = 0$$

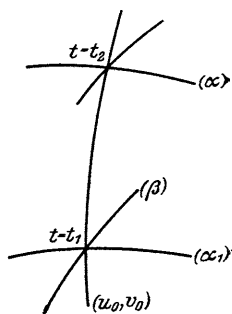


Fig. 4.

bewiesen werden, worin die obere und untere Grenze Funktionen von β, α_1, α sind. Führt man unter dem Integral

$$f(u_0, v_0, u'_0, v'_0) = u'_0 f_u(u_0 \dots) + v'_0 f_v(u_0 \dots)$$

ein und setzt

$$\int_{(t_1)}^{(t_2)} \left\{ \frac{f_u(u_0, v_0, u'_0, v'_0)}{\cos \omega} du_0 + \frac{f_v(u_0, v_0, u'_0, v'_0)}{\cos \omega} dv_0 \right\} = \mathfrak{S}(\beta, \alpha_1, \alpha);$$

dann ist zu zeigen, daß

$$\frac{\partial}{\partial \beta} \mathfrak{S}(\beta, \alpha_1, \alpha) = 0$$

ist für jedes α . Dafür ist

$$(29) \quad \frac{\partial^2}{\partial \alpha \partial \beta} \mathfrak{S} = 0$$

notwendig und hinreichend. Denn daraus folgt

$$\frac{\partial \mathfrak{S}}{\partial \beta} = a(\beta, \alpha_1),$$

und da $\mathfrak{S}(\beta, \alpha_1, \alpha_1) = 0$ ist für jedes β , also auch

$$\frac{\partial}{\partial \beta} \mathfrak{S}(\beta, \alpha_1, \alpha_1) = 0,$$

und da α in $a(\beta, \alpha_1)$ nicht vorkommt, ist $\frac{\partial \mathfrak{S}}{\partial \beta} = a(\beta, \alpha_1) = 0$ für jedes α .

Die Differentiation des Integrals ergibt wegen $u_0(t, A) = U(\alpha, \beta)$, $v_0(t, A) = V(\alpha, \beta)$

$$\frac{\partial \mathfrak{S}}{\partial \alpha} = \frac{f_{u'}(u_0, v_0, u'_0, v'_0)}{\cos \omega} \frac{\partial U}{\partial \alpha} + \frac{f_{v'}(u_0, v_0, u'_0, v'_0)}{\cos \omega} \frac{\partial V}{\partial \alpha},$$

und da

$$\cos \omega = \frac{\frac{\partial U}{\partial \alpha} f_{u'}(u_0, v_0, u'_0, v'_0) + \frac{\partial V}{\partial \alpha} f_{v'}(u_0, v_0, u'_0, v'_0)}{f\left(U, V, \frac{\partial U}{\partial \alpha}, \frac{\partial V}{\partial \alpha}\right)}$$

ist,

$$\frac{\partial \mathfrak{S}}{\partial \alpha} = f\left(U, V, \frac{\partial U}{\partial \alpha}, \frac{\partial V}{\partial \alpha}\right).$$

Wegen Gleichung (28) ist Gleichung (29) erfüllt. Also sind die schiefen Äquidistanten (24) geodätische Äquidistante im Sinne der Maßbestimmung

$\frac{f}{\cos \omega}$. Als Gefällkurven der Schar geodätischer Äquidistanten sind die Kurven

$$(25) \quad \begin{cases} u = u_0(t, A) \\ v = v_0(t, A) \end{cases}$$

geodätische Linien der Maßbestimmung $\frac{f}{\cos \omega}$ und befriedigen daher die Gleichung

$$(10) \quad G \cdot P(u, v) + \frac{\partial P}{\partial u} f_{v'} - \frac{\partial P}{\partial v} f_{u'} = 0$$

für

$$P = \frac{1}{\cos \omega} = p(u, v), \quad u = u_0(t, A), \quad v = v_0(t, A),$$

d. h. die Gleichung (17). Also ist die Integrabilitätsbedingung erfüllt und das Netz ist ein Gewebe.

An die ursprüngliche Integrabilitätsbedingung (12) läßt sich eine Bemerkung über den auf Gewebekurven gemessenen Abstand zweier Punkte

knüpfen. Aus Gleichung (12) folgt, daß eine Funktion $\Psi(\alpha, \beta) = \Phi(u, v)$ existiert, für die

$$\frac{\partial \Psi}{\partial \alpha} = f\left(u, v, \frac{\partial u}{\partial \alpha}, \frac{\partial v}{\partial \alpha}\right), \quad \frac{\partial \Psi}{\partial \beta} = f\left(u, v, \frac{\partial u}{\partial \beta}, \frac{\partial v}{\partial \beta}\right)$$

gilt. Benutzt man die Gleichung (15), so findet man aus

$$\begin{aligned} \frac{\partial u}{\partial \alpha} \left(p f_u'(u_0 \dots) - \frac{\partial \Phi}{\partial u} \right) + \frac{\partial v}{\partial \alpha} \left(p f_v'(u_0 \dots) - \frac{\partial \Phi}{\partial v} \right) &= 0, \\ \frac{\partial u}{\partial \beta} \left(p f_u'(u_0 \dots) - \frac{\partial \Phi}{\partial u} \right) + \frac{\partial v}{\partial \beta} \left(p f_v'(u_0 \dots) - \frac{\partial \Phi}{\partial v} \right) &= 0 \end{aligned}$$

wegen

$$\frac{\partial(u, v)}{\partial(\alpha, \beta)} \neq 0$$

$$(30) \quad \frac{\partial \Phi}{\partial u} = p \cdot f_u'(u_0 \dots), \quad \frac{\partial \Phi}{\partial v} = p \cdot f_v'(u_0 \dots),$$

woraus man durch Integration die Funktion $\Phi(u, v)$ findet, die bis auf eine willkürliche Konstante die auf dem Gewebe gemessene Bogenlänge darstellt. Es ist

$$\frac{\partial^2 \Phi}{\partial u \partial v} - \frac{\partial^2 \Phi}{\partial v \partial u} \equiv p \cdot G + \frac{\partial p}{\partial u} f_v'(u_0 \dots) - \frac{\partial p}{\partial v} f_u'(u_0 \dots).$$

Für ein gestreiftes Gewebe gilt

$$f_u'(u_0 \dots) = \lambda(u, v) \frac{\partial p}{\partial u}, \quad f_v'(u_0 \dots) = \lambda(u, v) \frac{\partial p}{\partial v},$$

und wegen

$$0 = G(u_0, v_0) \equiv \left. \frac{d}{du} f_v'(u_0 \dots) \right|_{v=\text{konst.}} - \left. \frac{d}{dv} f_u'(u_0 \dots) \right|_{u=\text{konst.}}$$

ist weiter

$$\lambda = \mu(p),$$

so daß aus Gleichung (30) $d\Phi = \mu(p) p dp$

$$(31) \quad \Phi(u, v) = M(p) + C$$

folgt bei

$$\frac{dM}{dp} = \frac{p f_u'(u_0 \dots)}{\frac{\partial p}{\partial u}} = \frac{p f_v'(u_0 \dots)}{\frac{\partial p}{\partial v}} = \lambda p.$$

Für gestreifte Gewebe ist also die auf den Netzkurven gemessene Bogenlänge eine Funktion von ω allein. Es gilt auch die Umkehrung: Aus

$$\Phi(u, v) = M(p) + C$$

und Gleichung (30) folgt nämlich

$$\begin{aligned} p f_u'(u_0 \dots) - \frac{dM}{dp} \cdot \frac{\partial p}{\partial u} &= 0, \\ p f_v'(u_0 \dots) - \frac{dM}{dp} \cdot \frac{\partial p}{\partial v} &= 0, \end{aligned}$$

also wegen $p \neq 0$, $\frac{dM}{dp} \neq 0$

$$\frac{\partial p}{\partial v} f_u(u_0 \dots) - \frac{\partial p}{\partial u} f_v(u_0 \dots) = 0,$$

d. h. $p = \text{konst.}$ längs der Transversalen der Winkelhalbierenden.

Aus (31) folgt, daß sämtliche Punkte des gestreiften Gewebes, die auf einer und derselben Äquidistanten liegen, den auf dem Gewebe gemessenen „Abstand“ null haben, weil ja längs dieser Kurven $p = \text{konst.}$ ist. Die Gleichung (31) gilt nicht, wenn p im ganzen Bereich konstant ist. Ein solches gestreiftes Gewebe mit überall konstantem Maschenwinkel $2\omega = 2 \arccos \frac{1}{p}$ bilden z. B. für $f = \sqrt{u'^2 + v'^2}$ in einer Ebene die beiden Scharen logarithmischer Spiralen, deren Gleichungen in Polarkoordinaten $r = e^{\pm a\varphi}$ sind. Die Winkelhalbierenden dieses gestreiften Gewebes sind Extremalen, nämlich die Geraden durch den Nullpunkt.

V.

Zwischen den eben bewiesenen Sätzen und dem Hilbertschen Unabhängigkeitstheorem besteht ein Zusammenhang, der die Wirksamkeit der Finslerschen Erweiterung des Winkelbegriffes erklärlich erscheinen läßt. Zu den Sätzen II und III über allgemeine Gewebe führten zwei wesentliche Schritte: erstens die Umformung der Integrabilitätsbedingung (12) in die Euler-Weierstraßsche Differentialgleichung (17) $G(p \cdot f) = 0$; zweitens der Nachweis, daß schiefe Äquidistante eines Netzes mit dem halben Maschenwinkel ω geodätische Äquidistante im $p \cdot f$ -Sinne darstellen, daß also die Winkelhalbierenden $\frac{f}{\cos \omega}$ -Extremalen sind, wenn ihre Transversalen schiefe Äquidistante des Netzes darstellen. Der erste Schritt wurde hier dem eigentlichen Beweis als eine identische Transformation der Integrabilitätsbedingung vorangeschickt, die sich zwar unschwer durchführen läßt, deren Nutzen aber von vornherein nicht einzusehen ist, so daß sie als ein erst nachträglich gerechtfertigter Kunstgriff erscheint. Während der zweite Teil des Beweises nicht entbehrlich ist, erspart man sich die Aufstellung der Integrabilitätsbedingung und ihre Diskussion durch Anwendung des Hilbertschen Satzes, der für ein Weierstraßsches Variationsproblem

$$(32) \quad \delta \int_{t_0}^{t_1} \varphi(u, v, u', v') dt = 0$$

bekanntlich folgendermaßen lautet:

Sind $\frac{du_0}{dt} = u'_0(u, v)$, $\frac{dv_0}{dt} = v'_0(u, v)$ die Differentialgleichungen der ∞^1

Extremalen eines Feldes, dann ist das Integral

$$\int_{(C)} \left\{ u' \frac{\partial}{\partial u'} \varphi(u, v, u'_0, v'_0) + v' \frac{\partial}{\partial v'} \varphi(u, v, u'_0, v'_0) \right\} dt$$

von dem ganz im Felde liegenden Integrationsweg $C = (u(t), v(t))$ unabhängig, also eine Funktion allein von dem Anfangs- und Endpunkt des Integrationsweges. Dabei muß der Weg $(u(t), v(t))$ stetig sein und bis auf endlich viele Ecken in jedem Punkte eine Tangente besitzen.

Man setze nun

$$\varphi(u, v, u', v') dt = \underset{(\varphi)}{ds}$$

und führe den von der Kurve $(u(t), v(t))$ zu der durch den Punkt gehenden Extremalen $(u_0(t), v_0(t))$ des Problems (32) gemessenen Winkel ω durch

$$\cos \omega \underset{(\varphi)}{=} \frac{u' \varphi_{u'}(u_0, v_0, u'_0, v'_0) + v' \varphi_{v'}(u_0, v_0, u'_0, v'_0)}{\varphi(u, v, u', v')} = \frac{1}{\pi(u, v)}$$

ein; dann geht das Hilbertsche unabhängige Integral, längs eines geschlossenen Weges erstreckt, über in

$$(33) \quad \oint \underset{(\varphi)}{\cos \omega} \underset{(\varphi)}{ds} = 0.$$

Nun sei nach dem Verfahren des Satzes III ein Netz

$$(26) \quad \begin{cases} u = u(\alpha, \beta) \\ v = v(\alpha, \beta) \end{cases}$$

konstruiert, dessen Winkelhalbierende (im f -Sinne)

$$(25) \quad \begin{cases} u = u_0(t, A) \\ v = v_0(t, A) \end{cases}$$

Extremalen im Sinne der Maßbestimmung $\frac{f}{\cos \omega}$ sind, unter ω der im f -Sinne von den Netzkurven zu der Winkelhalbierenden gemessene Winkel verstanden. Dies wird genau wie oben bewiesen. Die Konstruktion des Netzes (26) wird wieder auf einen Bereich beschränkt, in dem die Differentialgleichung der Schar (25) keinen Verzweigungspunkt hat, so daß dieser Bereich ein Feld im Sinne des Variationsproblems (32) darstellt, wenn man

$$\varphi(u, v, u', v') \equiv \frac{f(u, v, u', v')}{\cos \omega} \quad (\cos \omega \neq 0)$$

setzt. Man beachte, daß

$$\underset{(\varphi)}{\cos \omega} = \cos \omega$$

ist, und wende auf einen aus Netzkurven gebildeten geschlossenen Weg den Hilbertschen Unabhängigkeitssatz in der Form der Gleichung (33) an:

$$\int_{\odot} \cos \omega \frac{ds}{(\varphi)} = \int_{\odot} \cos \omega \frac{f}{\cos \omega} dt = \int_{\odot} ds = 0,$$

womit Satz III schon bewiesen ist.

Es ist einleuchtend, daß die Sätze Ib, Ic und II sich nicht mit Hilfe des Hilbertschen Unabhängigkeitssatzes beweisen lassen, dagegen wohl Satz Ia und der Hilfssatz.

Beweis für Satz Ia. Wenn das Netz spiegelbildlich zu Extremalen im f -Sinne und gestreift bezüglich ihrer Äquidistanten ist, kann man außer wie beim Beweis von Satz III so vorgehen: Die Anwendung des Hilbertschen Integralsatzes für $\varphi \equiv f$ auf eine geschlossene Kurve, die aus zwei Äquidistanten und den von ihnen herausgeschnittenen Stücken der Netzkurven besteht, liefert, da längs der Äquidistanten $\cos \omega = 0$ ist, die Gleichung

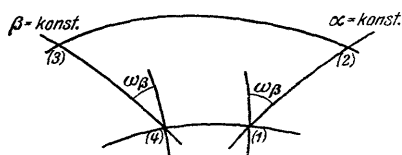


Fig. 5.

$$\int_{(1)}^{(2)} \cos \omega_\alpha ds_\alpha + \int_{(3)}^{(4)} \cos \omega_\beta ds_\beta = 0,$$

in der

$$ds_\alpha = f(u, v, u_\alpha, v_\alpha) d\alpha |_{\beta=\text{konst.}},$$

$$ds_\beta = f(u, v, u_\beta, v_\beta) d\beta |_{\alpha=\text{konst.}}$$

gesetzt ist, oder, wenn die Äquidistanten die Kurven $\beta = \beta_1$ und $\beta = \beta_2$ treffen:

$$\int_{(1)}^{(2)} \cos \omega_{\alpha_1} ds_{\alpha_1} + \int_{(3)}^{(4)} \cos \omega_{\alpha_2} ds_{\alpha_2} = 0$$

bei

$$ds_{\alpha_1} = f(u, v, u_\alpha, v_\alpha) d\alpha |_{\beta=\beta_1}, \quad ds_{\alpha_2} = f(u, v, u_\alpha, v_\alpha) d\alpha |_{\beta=\beta_2}.$$

Da $\omega_\alpha = \omega_\beta$ längs der Äquidistanten konstant und $\cos \omega_\alpha \neq 0$ ist, folgt durch Grenzübergang

$$ds_\alpha + ds_\beta = 0, \quad ds_{\alpha_1} + ds_{\alpha_2} = 0.$$

Integriert man diese Gleichungen über Kurvenstücke des Netzes, die zwischen gleichen Äquidistanten liegen, dann erhält man den Inhalt des Hilfssatzes; durch Integration über ein ganzes Netzviereck dagegen den des Satzes Ia.

Der Hilfssatz läßt sich hier noch etwas allgemeiner aussprechen, als er oben gebraucht wurde:

Die Transversalen einer Extremalenschar im Sinne $p \cdot f = \frac{f}{\cos \omega}$ schneiden auf den Kurven des zur Extremalenschar mit dem Winkel $\omega = \arccos \frac{1}{p}$ spiegelbildlichen Netzes gleiche Längen $\int f dt$ aus.

Beweis. Die Gleichung (33) erhält für $\varphi = \frac{f}{\cos \omega}$ die Form

$$\int_{\sigma} \frac{f}{\cos \omega} \cos \omega' dt = 0,$$

wobei ω' der Winkel zwischen dem Integrationsweg und der Extremalen ist, ω der halbe Maschenwinkel. Der Integrationsweg sei derselbe wie oben.

Längs der Transversalen ist

$$\cos \omega' = 0,$$

längs der Netzkurven

$$\cos \omega' = \cos \omega;$$

also folgt

$$\int_{(1)}^{(2)} f dt + \int_{(3)}^{(4)} f dt = 0.$$

Offenbar enthält dieser Hilfssatz den früheren; denn f -Extremalen sind auch $p \cdot f$ -Extremalen, wenn $p = p(u, v)$ längs der Transversalen konstant ist. Andererseits liefert er mit den Gleichungen (17) und (10) zusammen den Satz II.

VI.

Die im vorhergehenden bewiesenen Sätze sind dank der Finslerschen Terminologie eine fast wörtliche Verallgemeinerung derjenigen, die Herr Rothe¹⁵⁾ für den speziellen Fall der Riemannschen Maßbestimmung bewiesen hat. Nur an einer Stelle müssen die Sätze wesentlich anders formuliert werden, und das läßt hinsichtlich der Gewebe auf Flächen im Euklidischen Raum auf eine Eigenschaft schließen, die, wie sich zeigen wird, zwar nicht auf den Fall der allgemeinen Maßbestimmung übertragbar, aber auch nicht für die Riemannsche Maßbestimmung charakteristisch ist.

Im vorhergehenden spielt mehrmals folgende Konstruktion eine Rolle: Ein Netz ist spiegelbildlich zu der Kurvenschar, deren Transversalen schiefe Äquidistante der Netzkurven sind. An den entsprechenden Stellen der Untersuchungen von Herrn Rothe wird verlangt, das Netz soll zu den schiefen Äquidistanten selbst spiegelbildlich sein, nicht zu der Schar, deren Transversalen die Äquidistanten sind. Für die Riemannsche Maßbestimmung ist dieser Unterschied tatsächlich belanglos, denn auf Flächen im Euklidischen Raum ist ein zu einer Kurvenschar spiegelbildliches Netz auch zur Transversalenschar spiegelbildlich, wenn man nur die eine Schar der Netzkurven entgegengesetzt orientiert.

¹⁵⁾ R. Rothe, Über die Bekleidung einer Fläche mit einem Gewebe, Sitzungsber. d. B. M. G. 7 (1907), 1. Stück, S. 17.

Bei einer beliebigen Maßbestimmung trifft dies durchaus nicht zu; andernfalls müßte ja aus

$$\frac{u_\alpha f_u'(u, v, u'_0, v'_0) + v_\alpha f_v'(u, v, u'_0, v'_0)}{f(u, v, u_\alpha, v_\alpha)} = \frac{u_\beta f_u'(u, v, u'_0, v'_0) + v_\beta f_v'(u, v, u'_0, v'_0)}{f(u, v, u_\beta, v_\beta)}$$

und

$$u'_1 f_u'(u, v, u'_0, v'_0) + v'_1 f_v'(u, v, u'_0, v'_0) = 0$$

stets

$$\frac{u_\alpha f_u'(u, v, u'_1, v'_1) + v_\alpha f_v'(u, v, u'_1, v'_1)}{f(u, v, u_\alpha, v_\alpha)} = - \frac{u_\beta f_u'(u, v, u'_1, v'_1) + v_\beta f_v'(u, v, u'_1, v'_1)}{f(u, v, u_\beta, v_\beta)}$$

folgen, was z. B. für $f = \sqrt[4]{u'^2 + v'^2}$ nicht der Fall ist.

Aus der eben bemerkten Eigenschaft der Netze der Flächen im Euklidischen Raum ergibt sich:

Das Netz

$$(26) \quad \begin{cases} u = u(\alpha, \beta) \\ v = v(\alpha, \beta) \end{cases}$$

ist das Diagonalnetz des Netzes der Winkelhalbierenden (im Sinne der Riemannschen Maßbestimmung)

$$(25) \quad \begin{cases} u = u_0(t, A) \\ v = v_0(t, A) \end{cases}$$

und ihrer Transversalen

$$(24) \quad \begin{cases} u = u_1(\tau, B) \\ v = v_1(\tau, B) \end{cases}$$

und umgekehrt.

Dabei sind Diagonalnetze folgendermaßen definiert: *Zwei Netze sind Diagonalnetze zueinander, wenn gegenüberliegende Ecken eines einfachen Netzvierecks des einen Netzes auf einer und derselben Kurve des anderen Netzes liegen, für den Fall, daß sämtliche Seitenlängen des Netzvierecks gegen Null streben.*

Die notwendige und hinreichende Bedingung für die Diagonaleigenschaft der Netze (26) und (24, 25) ist

$$D \equiv \begin{vmatrix} u_\alpha v'_0 - v_\alpha u'_0 & -(u_\alpha v'_1 - v_\alpha u'_1) \\ u_\beta v'_0 - v_\beta u'_0 & u_\beta v'_1 - v_\beta u'_1 \end{vmatrix} = 0$$

oder

$$\frac{\left(\frac{du}{dv}\right)_\alpha - \left(\frac{du}{dv}\right)_0}{\left(\frac{du}{dv}\right)_\beta - \left(\frac{du}{dv}\right)_0} : \frac{\left(\frac{du}{dv}\right)_\alpha - \left(\frac{du}{dv}\right)_1}{\left(\frac{du}{dv}\right)_\beta - \left(\frac{du}{dv}\right)_1} = -1.$$

Da im Sinne Riemannscher Maßbestimmung die Kurvenscharen (25) und (24) orthogonal sind und mit dem Netz (26) der obigen Bemerkung

zufolge gleiche Winkel einschließen, ist das Doppelverhältnis der vier Richtungen tatsächlich gleich -1 , und die Netze (24, 25) und (26) sind Diagonalnetze, wie behauptet wurde.

Nun ist der Begriff des Diagonalnetzes rein lagegeometrisch, andererseits entspricht der Orthogonalität die Transversalität der allgemeinen Maßbestimmung. Daher möchte man vermuten, daß allgemein ein Kurvennetz (26) Diagonalnetz ist zu dem Netz der Kurven (25) und ihrer Transversalen (24), wenn nur das ursprüngliche Kurvennetz spiegelbildlich ist sowohl zu der Schar (u_0, v_0) , als auch zu der Schar (u_1, v_1) . Diese Vermutung bestätigt sich indessen nur bei Beschränkung auf Maßbestimmungen $f(u, v, u', v')$ mit symmetrischer Transversalitätsbedingung, für die also die Gleichungen

$$(34) \quad \begin{cases} u'_1 f_{u'}(\dots u'_0, v'_0) + v'_1 f_{v'}(\dots u'_0, v'_0) = 0 \\ u'_0 f_{u'}(\dots u'_1, v'_1) + v'_0 f_{v'}(\dots u'_1, v'_1) = 0 \end{cases}$$

zugleich befriedigt sind.

Die Indikatrix des entsprechenden Variationsproblems ist eine Eilinie, deren Halbmesser sich zu Paaren konjugierter zusammenfassen lassen. Herr Radon¹⁶⁾ hat, von diesem kurvengeometrischen Problem ausgehend, sämtliche Funktionen f bestimmt, die den Gleichungen (34) genügen. Da ich von der oben charakterisierten Fragestellung — Übertragbarkeit der Diagonaleigenschaften von Netzen auf den Fall allgemeiner Maßbestimmungen — ausging, gelangte ich zu den Funktionalgleichungen (34) und habe aus ihnen unmittelbar die Funktionen f bestimmt. Man geht etwa so vor: Wenn zunächst $v'_0 > 0$, $v'_1 > 0$ ist, wird

$$\frac{u'_0}{v'_0} = z, \quad \frac{u'_1}{v'_1} = w$$

gesetzt. Dann folgt aus (34) wegen $u' f_{u'} + v' f_{v'} = f$

$$w = z - \frac{f(u, v, z, 1)}{f_{u'}(u, v, z, 1)},$$

$$z = w - \frac{f(u, v, w, 1)}{f_{u'}(u, v, w, 1)}.$$

u und v treten nur als Parameter auf und werden im folgenden fortgelassen; man setzt noch

$$\begin{aligned} f(z, 1) &= f(z), \\ f_{u'}(z, 1) &= f'(z) = \frac{df}{dz} \end{aligned}$$

¹⁶⁾ J. Radon, Über eine besondere Art ebener konvexer Kurven. Berichte d. Kgl. Sächs. Ges. d. Wiss., Math.-Phys. Kl. 68 (1916), S. 123 ff. Vgl. auch: P. Funk, Über den Begriff „extremale Krümmung“ und eine kennzeichnende Eigenschaft der Ellipse, Math. Zeitschr. 3 (1919), S. 87 ff.

und erhält

$$(35) \quad \begin{cases} w = z - \frac{f(z)}{f'(z)} \\ z = w - \frac{f(w)}{f'(w)}. \end{cases}$$

Die Bedingungen $f'(z) \neq 0$, $f'(w) \neq 0$ sind wegen der Voraussetzung

$$v'_0 > 0, \quad v'_1 > 0$$

sicher erfüllt. Aus (35) folgt mit $\frac{f(z)}{f'(z)} = \psi(z)$:

$$\psi(z) + \psi(z - \psi(z)) = 0,$$

und wenn man noch

$$z - \psi(z) = \varphi(z)$$

setzt:

$$z - \varphi(z) + \varphi(z) - \varphi(\varphi(z)) = 0,$$

also

$$(36a) \quad \varphi(\varphi(z)) = z.$$

Man wählt also irgendeine Funktion $\varphi(z)$, die mit ihrer Umkehrfunktion identisch ist; ihr geometrisches Bild in einer z - φ -Ebene ist eine zur Geraden $z = \varphi$ symmetrische Kurve. Durch Integration erhält man:

$$f(z) = C e^{\int \frac{dz}{z - \varphi(z)}},$$

muß also zu (36a) noch die Bedingung

$$(36b) \quad \varphi(z) - z \neq 0$$

hinzufügen. C ist eine willkürliche Funktion von u und v . Aus

$$f(z) = f\left(\frac{u'}{v'}, 1\right)$$

folgt

$$(37a) \quad f(u', v') = v' C e^{\int \frac{dz}{z - \varphi(z)}}, \quad z = \frac{u'}{v'}, \quad v' > 0.$$

Da die Gleichungen (34) für jedes beliebige Verhältnis $\frac{u'_0}{v'_0}$ und das z. B. aus der ersten der Gleichungen berechnete Verhältnis $\frac{u'_1}{v'_1}$ erfüllt sein sollen, bedeutet die oben benutzte Voraussetzung $v'_0 > 0$, $v'_1 > 0$ keine Einschränkung. Die Form (37a) versagt aber für $v' \leq 0$. In diesem Fall führt man eine entsprechende Betrachtung durch; insbesondere ersetzt man, wenn v' verschwindet, da ja nicht zugleich auch $u' = 0$ ist, z durch $\zeta = \frac{1}{z}$. Man erhält dann

$$(37b) \quad f(u', v') = u' C e^{\int \frac{d\zeta}{\zeta - \bar{\varphi}(\zeta)}}, \quad \zeta = \frac{v'}{u'}, \quad \bar{\varphi}(\bar{\varphi}(\zeta)) = \zeta, \quad u' > 0.$$

$\varphi(z)$ und $\bar{\varphi}(\zeta)$ sind nicht unabhängig: für $u' > 0$, $v' > 0$ ist

$$\varphi(z) \cdot \bar{\varphi}(\zeta) = 1.$$

Man verifiziert leicht, daß umgekehrt jede Funktion (37a) oder (37b) zu einer symmetrischen Transversalitätsbedingung führt. Also stellt (37a), (37b) die allgemeine Lösung der Funktionalgleichungen (34) dar. Speziell ist die Riemannsche Maßbestimmung

$$f = \sqrt{E u'^2 + 2 F u' v' + G v'^2}$$

durch die Hyperbel

$$\varphi(z) = -\frac{F \cdot z + G}{E \cdot z + F}$$

charakterisiert. Die Mannigfaltigkeit der eindeutigen Funktionen $\varphi(z)$, die den Bedingungen (36a) und (36b) genügen, ist indessen sehr viel größer, es gibt also außer der Riemannschen noch sehr viel Maßbestimmungen mit symmetrischer Transversalitätsbedingung im Gegensatz zu dem Ergebnis bei dem räumlichen Problem¹⁷⁾.

Um von einer *Maßbestimmung* mit vertauschbarer Transversalitätsbedingung zu reden, muß noch bewiesen werden, daß die genannte Eigenschaft der Funktion $f(u, v, u', v')$ gegenüber der regulären Punkttransformation

$$(38) \quad \begin{cases} u = u(\bar{u}, \bar{v}) & \frac{\partial(u, v)}{\partial(\bar{u}, \bar{v})} \neq 0 \\ v = v(\bar{u}, \bar{v}) \end{cases}$$

invariant ist. Das ist tatsächlich der Fall. Denn wenn durch die Transformation (38) $f(u, v, u', v')$ in

$$g(\bar{u}, \bar{v}, \bar{u}', \bar{v}') \equiv f(u(\bar{u}, \bar{v}), v(\bar{u}, \bar{v}), \frac{\partial u}{\partial \bar{u}} \bar{u}' + \frac{\partial u}{\partial \bar{v}} \bar{v}', \frac{\partial v}{\partial \bar{u}} \bar{u}' + \frac{\partial v}{\partial \bar{v}} \bar{v}')$$

übergeht, ist

$$\begin{aligned} \bar{u}'_1 g_{\bar{u}'}(\bar{u}, \bar{v}, \bar{u}'_0, \bar{v}'_0) + \bar{v}'_1 g_{\bar{v}'}(\bar{u}, \bar{v}, \bar{u}'_0, \bar{v}'_0) &= \bar{u}'_1 f_{u'}(u, v, u'_0, v'_0) + \bar{v}'_1 f_{v'}(u, v, u'_0, v'_0), \\ \bar{u}'_0 g_{\bar{u}'}(\bar{u}, \bar{v}, \bar{u}'_1, \bar{v}'_1) + \bar{v}'_0 g_{\bar{v}'}(\bar{u}, \bar{v}, \bar{u}'_1, \bar{v}'_1) &= \bar{u}'_0 f_{u'}(u, v, u'_1, v'_1) + \bar{v}'_0 f_{v'}(u, v, u'_1, v'_1).^{18)} \end{aligned}$$

Die durch die Gleichungen (37a, b) charakterisierten Maßbestimmungen

¹⁷⁾ Vgl. W. Blaschke, Räumliche Variationsprobleme mit symmetrischer Transversalitätsbedingung, Berichte d. Kgl. Sächs. Ges. d. Wiss., Math.-Phys. Kl., S. 50 ff.

¹⁸⁾ Vgl. z. B. O. Bolza, a. a. O. S. 347.

nehmen insofern eine Mittelstellung zwischen der Riemannschen Maßbestimmung und einer ganz beliebigen ein, als bei der Riemannschen Maßbestimmung (und nur bei ihr) *jeder* Winkel von der Reihenfolge der Schenkel unabhängig ist, bei der eben betrachteten nur der *rechte* Winkel, und bei einer beliebigen Maßbestimmung *gar kein* Winkel. Darauf beruht auch der schon oben angedeutete Zusammenhang mit der Diagonaleigenschaft von Kurvennetzen, der sich so formulieren läßt:

Satz IV. *Die Kurven*

$$(26) \quad \begin{cases} u = u_0(t, A) \\ v = v_0(t, A) \end{cases}$$

mögen die Winkelhalbierenden (im f -Sinne) des Netzes

$$(27) \quad \begin{cases} u = u(\alpha, \beta) \\ v = v(\alpha, \beta) \end{cases}$$

sein, und die Kurven

$$(25) \quad \begin{cases} u = u_1(\tau, B) \\ v = v_1(\tau, B) \end{cases}$$

die Transversalen der Kurven (26).

Dann ist von den drei Tatsachen:

1. Die Transversalitätsbedingung von $f(u, v, u', v')$ ist symmetrisch;
2. Die Kurven (25) halbieren den Außenwinkel des Netzes;
3. Die Kurvennetze (27) und (25, 26) sind Diagonalnetze; eine jede die Folge der beiden anderen.

Beweis. Die allgemeinen Voraussetzungen des Satzes sind:

$$(39) \quad \frac{u_\alpha f_{u'}(u_0, v_0, u'_0, v'_0) + v_\alpha f_{v'}(u_0, v_0, u'_0, v'_0)}{f(u, v, u_\alpha, v_\alpha)} = \cos \omega = \frac{u_\beta f_{u'}(u_0, v_0, u'_0, v'_0) + v_\beta f_{v'}(u_0, v_0, u'_0, v'_0)}{f(u, v, u_\beta, v_\beta)}$$

$$(40) \quad u_1^{\alpha} f_{u'}(u_0, v_0, u'_0, v'_0) + v_1^{\alpha} f_{v'}(u_0, v_0, u'_0, v'_0) = 0.$$

Die erste der drei Zusatzbedingungen ist

$$(41) \quad u_0^{\alpha} f_{u'}(u_1, v_1, u'_1, v'_1) + v_0^{\alpha} f_{v'}(u_1, v_1, u'_1, v'_1) = 0,$$

die zweite

$$(42) \quad \cos \omega'_\alpha = \frac{u_\alpha f_{u'}(u_1, v_1, u'_1, v'_1) + v_\alpha f_{v'}(u_1, v_1, u'_1, v'_1)}{f(u, v, u_\alpha, v_\alpha)} = \cos \omega'$$

$$= - \frac{u_\beta f_{u'}(u_1, v_1, u'_1, v'_1) + v_\beta f_{v'}(u_1, v_1, u'_1, v'_1)}{f(u, v, u_\beta, v_\beta)} = \cos \omega'_\beta,$$

die dritte

$$(43) \quad D = (u_\alpha v'_0 - v_\alpha u'_0)(u_\beta v'_1 - v_\beta u'_1) + (u_\beta v'_0 - v_\beta u'_0)(u_\alpha v'_1 - v_\alpha u'_1) = 0.$$

a) Aus (39), (40), (41), (42) folgt

$$(44) \quad \begin{cases} u'_1 = -\lambda f_{v'}(u_0, v_0, u'_0, v'_0), & v'_1 = \lambda f_{u'}(u_0, v_0, u'_0, v'_0) \quad (\lambda \neq 0) \\ u'_0 = -\kappa f_{v'}(u_1, v_1, u'_1, v'_1), & v'_0 = \kappa f_{u'}(u_1, v_1, u'_1, v'_1) \quad (\kappa \neq 0), \end{cases}$$

$$D = \kappa f(u, v, u_\alpha, v_\alpha) \cos \omega' \lambda f(u, v, u_\beta, v_\beta) \cos \omega$$

$$- \kappa f(u, v, u_\beta, v_\beta) \cos \omega' \lambda f(u, v, u_\alpha, v_\alpha) \cos \omega,$$

also

$$D = 0.$$

b) Aus (39), (40), (43) folgt

$$0 = (u_\alpha v'_0 - v_\alpha u'_0) \lambda \cos \omega f(u, v, u_\beta, v_\beta) + (u_\beta v'_0 - v_\beta u'_0) \lambda \cos \omega f(u, v, u_\alpha, v_\alpha),$$

und wegen $\lambda \cos \omega \neq 0$, $f(u, v, u_\alpha, v_\alpha) \neq 0$, $f(u, v, u_\beta, v_\beta) \neq 0$

$$0 = v'_0 \left(\frac{u_\alpha}{f(u, v, u_\alpha, v_\alpha)} + \frac{u_\beta}{f(u, v, u_\beta, v_\beta)} \right) - u'_0 \left(\frac{v_\alpha}{f(u, v, u_\alpha, v_\alpha)} + \frac{v_\beta}{f(u, v, u_\beta, v_\beta)} \right),$$

dazu die Gleichung (42)

$$0 = f_{v'}(u_1, v_1, u'_1, v'_1) \left(\frac{u_\alpha}{f(u, v, u_\alpha, v_\alpha)} + \frac{u_\beta}{f(u, v, u_\beta, v_\beta)} \right)$$

$$+ f_{v'}(u_1, v_1, u'_1, v'_1) \left(\frac{v_\alpha}{f(u, v, u_\alpha, v_\alpha)} + \frac{v_\beta}{f(u, v, u_\beta, v_\beta)} \right).$$

Also ist

$$v'_0 f_{v'}(u_1, v_1, u'_1, v'_1) + u'_0 f_{u'}(u_1, v_1, u'_1, v'_1) = 0,$$

denn

$$\frac{u_\alpha}{f(u, v, u_\alpha, v_\alpha)} + \frac{u_\beta}{f(u, v, u_\beta, v_\beta)} = 0,$$

$$\frac{v_\alpha}{f(u, v, u_\alpha, v_\alpha)} + \frac{v_\beta}{f(u, v, u_\beta, v_\beta)} = 0$$

ist wegen $\frac{\partial(u, v)}{\partial(\alpha, \beta)} \neq 0$ ausgeschlossen.

c) Endlich folgt durch Elimination von u'_0, v'_0, u'_1, v'_1 aus Gleichung (43) mit Hilfe von (40) und (41)

$$\frac{u_\alpha f_{u'}(u_1, v_1, u'_1, v'_1) + v_\alpha f_{v'}(u_1, v_1, u'_1, v'_1)}{f(u, v, u_\alpha, v_\alpha)} + \frac{u_\beta f_{u'}(u_1, v_1, u'_1, v'_1) + v_\beta f_{v'}(u_1, v_1, u'_1, v'_1)}{f(u, v, u_\beta, v_\beta)} = 0,$$

d. h.

$$\cos \omega'_\alpha = \cos \omega'_\beta.$$

Damit ist der Satz bewiesen.

VII.

Die Gewebe auf Flächen im Euklidischen Raum weisen, obwohl sie durch Merkmale definiert sind, die nur den Begriff der Bogenlänge benutzen, einige charakteristische Krümmungseigenschaften auf, die in der folgenden

von Herrn Rothe¹⁹⁾ bewiesenen Relation ihren Ausdruck finden:

$$(45) \quad (g_\alpha - g_\beta) \frac{\partial \Phi}{\partial \alpha} \frac{\partial \Phi}{\partial \beta} = \frac{\partial(2\omega, \Phi)}{\partial(\alpha, \beta)};$$

darin bedeuten g_α bzw. g_β die geodätische Krümmung der Gewebekurve $\alpha = \text{konst.}$ bzw. $\beta = \text{konst.}$, $\Phi(\alpha, \beta)$ den auf dem Gewebe gemessenen Abstand des Punktes (α, β) von einem festen Punkt, 2ω den *ganzen* Maschenwinkel des Gewebes. Im Gegensatz zu dem oben Untersuchten scheint sich dieser Satz nicht auf den Fall der allgemeinen Maßbestimmung übertragen zu lassen, wenigstens nicht in so einfacher Form. Das soll zum Schluß kurz dargetan werden.

Für eine Kurve im zweidimensionalen Raum hat man als geodätische Krümmung im Sinne der Maßbestimmung $f(u, v, u', v')$ die von Herrn Finsler²⁰⁾ als „erste Krümmung“ bezeichnete Invariante gegen Punkt- und Parametertransformation

$$(46) \quad k = \frac{G(u, v)}{\pm \sqrt{f_1 f^3}} = \frac{1}{\pm \sqrt{f_1 f^3}} \cdot \left[(u'v'' - v'u'')f_1 + \frac{\partial^2 f}{\partial v' \partial u} - \frac{\partial^2 f}{\partial u' \partial v} \right]$$

anzusehen, die schon von Landsberg²¹⁾ als „extremale Krümmung“ in die Variationsrechnung eingeführt wurde.

Um eine der Gleichung (45) analoge Relation zu finden, wird man für die Kurven $\alpha = \text{konst.}$ bzw. $\beta = \text{konst.}$ des Netzes

$$(47) \quad \begin{cases} u = u(\alpha, \beta) \\ v = v(\alpha, \beta) \end{cases}$$

gemäß der Gleichung (46) die Funktionen $k(\alpha)_{\beta=\text{konst.}}$ und $k(\beta)_{\alpha=\text{konst.}}$ bilden und zwischen ihnen eine aus dem Gewebecharakter des Netzes folgende Verbindung herzustellen suchen. Um die Formeln zu vereinfachen, wähle man als Netz der Parameterlinien das Gewebe selbst, eine offenbar unwesentliche Einschränkung, von der auch bei der Rotheschon Gleichung (45) Gebrauch gemacht ist. Durch die Transformation (47) wird

$$f(u, v, u', v') = g(\alpha, \beta, \alpha', \beta'),$$

und wenn man auf den Netzkurven noch α bzw. β selbst als Parameter einführt, wird für

$$(48) \quad \begin{cases} \alpha = \text{konst.}, & \alpha' = 0, \alpha'' = 0, \beta' = 1, \beta'' = 0, & g_\alpha = g(\alpha, \beta, 0, 1) \\ \beta = \text{konst.}, & \alpha' = 1, \alpha'' = 0, \beta' = 0, \beta'' = 0, & g_\beta = g(\alpha, \beta, 1, 0). \end{cases}$$

¹⁹⁾ R. Rothe, Über die Bekleidung einer Fläche mit einem Gewebe („Kurvennetze ohne Umwege“), Sitz.-Ber. d. B. M. G. 7 (1907), 1. Stück, S. 12 ff.

²⁰⁾ P. Finsler, a. a. O. S. 59 ff.

²¹⁾ O. Bolza, a. a. O. S. 346.

Die Gleichung der Winkelhalbierenden des Netzes ist jetzt

$$(49) \quad \begin{cases} \alpha = \alpha_0(t, A) \\ \beta = \beta_0(t, A), \end{cases}$$

und für den halben Maschenwinkel $\omega = \omega(\alpha, \beta)$ des Netzes gilt

$$(50) \quad \cos \omega = \frac{g_{\alpha'}(\alpha, \beta, \alpha'_0, \beta'_0)}{g(\alpha, \beta, 1, 0)} = \frac{g_{\beta'}(\alpha, \beta, \alpha'_0, \beta'_0)}{g(\alpha, \beta, 0, 1)}.$$

Die geodätischen Krümmungen der Netzkurven sind

$$(51) \quad \begin{cases} k(\alpha) = k_\beta = \frac{\frac{\partial^2}{\partial \alpha \partial \beta'} g(\alpha, \beta, 1, 0) - \frac{\partial^2}{\partial \beta \partial \alpha'} g(\alpha, \beta, 1, 0)}{\sqrt{g_1(\alpha, \beta, 1, 0) \cdot g^3(\alpha, \beta, 1, 0)}}, & g_1 = \frac{1}{\alpha'^2} \frac{\partial^2 g}{\partial \beta'^2} = g_{\beta' \beta'}, \\ k(\beta) = k_\alpha = \frac{\frac{\partial^2}{\partial \alpha \partial \beta} g(\alpha, \beta, 0, 1) - \frac{\partial^2}{\partial \beta \partial \alpha'} g(\alpha, \beta, 0, 1)}{\sqrt{g_1(\alpha, \beta, 0, 1) \cdot g^3(\alpha, \beta, 0, 1)}}, & g_1 = \frac{1}{\beta'^2} \frac{\partial^2 g}{\partial \alpha'^2} = g_{\alpha' \alpha'}. \end{cases}$$

Die Homogenitätsbedingung für g liefert die Gleichungen

$$(52) \quad \begin{cases} g(\alpha, \beta, 1, 0) = g_{\alpha'}(\alpha, \beta, 1, 0), \\ g(\alpha, \beta, 0, 1) = g_{\beta'}(\alpha, \beta, 0, 1). \end{cases}$$

Für das Folgende empfiehlt es sich, einen von Herrn Berwald²²⁾ in die Differentialgeometrie allgemeiner Räume eingeführten Begriff zu benutzen. Er definiert den Flächeninhalt ψ des von den Vektoren (ξ_1, η_1) und (ξ_2, η_2) aufgespannten Parallelogrammes in bezug auf das Linienelement $(\alpha, \beta, \alpha', \beta')$ durch die Gleichung

$$\psi(\xi_1, \eta_1; \xi_2, \eta_2; \alpha, \beta, \alpha', \beta') = \sqrt{g_1(\alpha, \beta, \alpha', \beta') \cdot g^3(\alpha, \beta, \alpha', \beta')} (\xi_1 \eta_2 - \xi_2 \eta_1).$$

Danach hat das von den Tangential-Einheitsvektoren $(1, 0)$ und $(0, 1)$ der Kurven $\alpha = \text{konst.}$, $\beta = \text{konst.}$ gebildete Parallelogramm den Inhalt

$$(53) \quad \psi(\alpha) = \sqrt{g_1(\alpha, \beta, 1, 0) \cdot g^3(\alpha, \beta, 1, 0)}$$

bezüglich der Tangente der Kurve $\beta = \text{konst.}$ als Linienelement, und den Inhalt

$$(54) \quad \psi(\beta) = \sqrt{g_1(\alpha, \beta, 0, 1) \cdot g^3(\alpha, \beta, 0, 1)}$$

bezüglich der Tangente der Kurve $\alpha = \text{konst.}$

Nur im Sinne der Riemannschen Metrik ist

$$\frac{\partial}{\partial \alpha'} \sqrt{g_1 \cdot g^3} = 0 = \frac{\partial}{\partial \beta'} \sqrt{g_1 \cdot g^3},$$

wie man leicht zeigt²³⁾.

²²⁾ L. Berwald, Über zweidimensionale allgemeine metrische Räume, I. Teil, Journ. f. d. r. u. ang. Math. 156, Heft 3.

²³⁾ P. Funk und L. Berwald, Flächeninhalt und Winkel in der Variationsrechnung, Lotos, Prag, 67/68 (1919/20), S. 47.

Endlich muß man noch neben dem bisher betrachteten halben Maschenwinkel, der von den Netzkurven zur Winkelhalbierenden gemessen wurde, die Netzwinkel $2\omega_\alpha$ und $2\omega_\beta$ einführen: $2\omega_\alpha$ wird von der Kurve $\beta = \text{konst.}$ zur Kurve $\alpha = \text{konst.}$ gemessen, $2\omega_\beta$ im entgegengesetzten Sinne.

Im allgemeinen ist

$$\begin{aligned} 2\omega_\alpha &\neq 2\omega_\beta, \\ 2\omega_\alpha &\neq 2\omega \neq 2\omega_\beta. \end{aligned}$$

Für die Cosinus dieser Winkel findet man

$$\cos 2\omega_\alpha = \frac{g_{\alpha'}(\alpha, \beta, 0, 1)}{g(\alpha, \beta, 1, 0)}, \quad \cos 2\omega_\beta = \frac{g_{\beta'}(\alpha, \beta, 1, 0)}{g(\alpha, \beta, 0, 1)},$$

und wegen Gleichung (52)

$$(55) \quad \begin{cases} g_{\alpha'}(\alpha, \beta, 1, 0) \cdot \cos 2\omega_\alpha = g_{\alpha'}(\alpha, \beta, 0, 1) \\ g_{\beta'}(\alpha, \beta, 0, 1) \cdot \cos 2\omega_\beta = g_{\beta'}(\alpha, \beta, 1, 0). \end{cases}$$

Bei Anwendung der Gleichungen (53), (54) und (55) folgt aus (51)

$$(56) \quad \begin{cases} k(\alpha) \cdot \psi(\alpha) = \frac{\partial}{\partial \alpha} [g_{\beta'}(\alpha, \beta, 0, 1) \cos 2\omega_\beta] - \frac{\partial}{\partial \beta} g_{\alpha'}(\alpha, \beta, 1, 0) \\ k(\beta) \cdot \psi(\beta) = -\frac{\partial}{\partial \beta} [g_{\alpha'}(\alpha, \beta, 1, 0) \cos 2\omega_\alpha] + \frac{\partial}{\partial \alpha} g_{\beta'}(\alpha, \beta, 0, 1). \end{cases}$$

Jetzt wird die Voraussetzung benutzt, daß die Kurven $\alpha = \text{konst.}$, $\beta = \text{konst.}$ ein Gewebe bilden: es existiert eine Funktion $\Phi(\alpha, \beta)$, für die

$$(57) \quad \begin{cases} g(\alpha, \beta, 1, 0) = \frac{\partial \Phi}{\partial \alpha} \\ g(\alpha, \beta, 0, 1) = \frac{\partial \Phi}{\partial \beta} \end{cases}$$

ist. Aus (57) und (52) folgt durch Addition der beiden Gleichungen (56)

$$(58) \quad \begin{aligned} k(\alpha) \cdot \psi(\alpha) + k(\beta) \cdot \psi(\beta) &= \frac{\partial}{\partial \alpha} \left[\frac{\partial \Phi}{\partial \beta} \cos 2\omega_\beta \right] - \frac{\partial}{\partial \beta} \left[\frac{\partial \Phi}{\partial \alpha} \cos 2\omega_\alpha \right], \\ &= \frac{\partial^2 \Phi}{\partial \alpha \partial \beta} (\cos 2\omega_\beta - \cos 2\omega_\alpha) + \frac{\partial \Phi}{\partial \beta} \frac{\partial}{\partial \alpha} \cos 2\omega_\beta - \frac{\partial \Phi}{\partial \alpha} \frac{\partial}{\partial \beta} \cos 2\omega_\alpha. \end{aligned}$$

Diese Gleichung stellt die Verallgemeinerung der Rothereschen Relation (45) dar, einen Zusammenhang zwischen Krümmung, Bogenlänge, Winkel und Fläche des von den Tangential-Einheitsvektoren gebildeten Parallelogramms eines Gewebes. Im allgemeinen scheint diese Beziehung sich nicht vereinfachen zu lassen. Setzt man

$$2\omega_\alpha = 2\omega = 2\omega_\beta$$

voraus, so beschränkt man sich schon auf die Riemannsche Maßbestimmung

$$g = \sqrt{E\alpha'^2 + 2F\alpha'\beta' + G\beta'^2},$$

ebenso durch die Annahme

$$\psi(\alpha, \beta, 1, 0) = \psi(\alpha, \beta, 0, 1) = \psi(\alpha, \beta).$$

Dann erhält man

$$\psi(\alpha, \beta) \cdot [k(\alpha) + k(\beta)] = -\sin 2\omega \cdot \frac{\partial(2\omega, \Phi)}{\partial(\alpha, \beta)},$$

worin nach Gleichungen (53), (54), (55) und (57)

$$\psi(\alpha, \beta) = +\sqrt{EG - F^2}, \quad \cos 2\omega = \frac{F}{+\sqrt{EG}}, \quad \sqrt{E} = \frac{\partial\Phi}{\partial\alpha}, \quad \sqrt{G} = \frac{\partial\Phi}{\partial\beta}$$

ist; also

$$-(k(\alpha) + k(\beta)) \cdot \frac{\partial\Phi}{\partial\alpha} \cdot \frac{\partial\Phi}{\partial\beta} = \frac{\partial(2\omega, \Phi)}{\partial(\alpha, \beta)}.$$

Dies ist tatsächlich die Formel (45), da

$$k(\alpha) = k_\beta = g_{\beta=\text{konst.}}$$

$$k(\beta) = k_\alpha = -g_{\alpha=\text{konst.}}$$

zu setzen ist, um über das Vorzeichen der geodätischen Krümmung, das Herr Finsler offen läßt, der klassischen Differentialgeometrie gemäß zu verfügen²⁴⁾.

²⁴⁾ Vgl. z. B. J. Knoblauch, Einleitung in die allgemeine Theorie der krummen Flächen, S. 248.

(Eingegangen am 20. 11. 1927.)

Über die partiellen Differenzgleichungen der mathematischen Physik.

Von

R. Courant, K. Friedrichs und H. Lewy in Göttingen.

Ersetzt man bei den klassischen linearen Differentialgleichungsproblemen der mathematischen Physik die Differentialquotienten durch Differenzenquotienten in einem — etwa rechtwinklig angenommenen — Gitter, so gelangt man zu algebraischen Problemen von sehr durchsichtiger Struktur. Die vorliegende Arbeit untersucht nach einer elementaren Diskussion dieser algebraischen Probleme vor allem die Frage, wie sich die Lösungen verhalten, wenn man die Maschen des Gitters gegen Null streben läßt. Dabei beschränken wir uns vielfach auf die einfachsten, aber typischen Fälle, die wir derart behandeln, daß die Anwendbarkeit der Methoden auf allgemeinere Differenzgleichungen und solche mit beliebig vielen unabhängigen Veränderlichen deutlich wird.

Entsprechend den für Differentialgleichungen geläufigen Fragestellungen behandeln wir Randwert- und Eigenwertprobleme für elliptische Differenzgleichungen und das Anfangswertproblem für hyperbolische bzw. parabolische Differenzgleichungen. Wir werden an einigen typischen Beispielen beweisen, daß der Grenzübergang stets möglich ist, nämlich daß die Lösungen der Differenzgleichungen gegen die Lösungen der entsprechenden Differentialgleichungsprobleme konvergieren; ja wir werden sogar erkennen, daß bei elliptischen Gleichungen i. a. die Differenzenquotienten beliebig hoher Ordnung gegen die entsprechenden Differentialquotienten streben. Die Lösbarkeit der Differentialgleichungsprobleme setzen wir nirgends voraus; vielmehr erhalten wir durch den Grenzübergang hierfür einen einfachen Beweis¹⁾. Während aber beim elliptischen

¹⁾ Unsere Beweismethode läßt sich ohne Schwierigkeit so erweitern, daß sie bei beliebigen linearen elliptischen Differentialgleichungen das Rand- und Eigenwertproblem und bei beliebigen linearen hyperbolischen Differentialgleichungen das Anfangswertproblem zu lösen gestattet.

Fälle einfache und weitgehend von der Wahl des Gitters unabhängige Konvergenzverhältnisse herrschen, werden wir bei dem Anfangswertproblem hyperbolischer Gleichungen erkennen, daß die Konvergenz allgemein nur dann vorhanden ist, wenn die Verhältnisse der Gittermaschen in verschiedenen Richtungen gewissen Ungleichungen genügen, die durch die Lage der Charakteristiken zum Gitter bestimmt werden.

Das typische Beispiel ist für uns im elliptischen Falle das Randwertproblem der Potentialtheorie. Seine Lösung von der Lösung des entsprechenden Differenzgleichungsproblems her ist übrigens in den letzten Jahren mehrfach behandelt worden²⁾. Allerdings werden dabei im Gegensatz zu der vorliegenden Arbeit meist spezielle Eigenschaften der Potentialgleichung benutzt, so daß die Anwendbarkeit der Methode auf andere Probleme nicht ohne weiteres zu übersehen ist.

Abgesehen von dem gekennzeichneten Hauptziel der Arbeit werden wir im Anschluß an die elementare algebraische Diskussion des Randwertproblems elliptischer Gleichungen dessen Zusammenhang mit dem aus der Statistik bekannten Probleme der Irrwege erörtern.

I. Der elliptische Fall.

§ 1.

Vorbemerkungen.

1. Definitionen.

Wir betrachten zunächst in der Ebene mit den rechtwinkligen Koordinaten x, y ein quadratisches Punktgitter der Maschenweite $h > 0$, etwa alle Punkte mit den Koordinaten $x = nh, y = mh$,

$$m, n = 0, \pm 1, \pm 2, \dots$$

²⁾ J. le Roux, Sur le problème de Dirichlet, Journ. de mathém. pur. et appl. (6) 10 (1914), p. 189. R. G. D. Richardson, A new method in boundary problems for differential equations, Transactions of the Americ. Mathem. Soc. 18 (1917), p. 489 ff. H. B. Philips and N. Wiener, Nets and the Dirichlet Problem, Publ. of the Mass. Institute of Technology (1925).

Leider waren diese Abhandlungen dem ersten der drei Verfasser bei der Abfassung seiner Note „Zur Theorie der partiellen Differenzgleichungen“, Gött. Nachr. 23. X. 1925, an welche die vorliegende Arbeit anschließt, entgangen.

Vgl. ferner: L. Lusternik, Über einige Anwendungen der direkten Methoden in der Variationsrechnung, Recueil de la Société Mathém. de Moscou, 1926. G. Bouligand, Sur le problème de Dirichlet, Ann. de la soc. polon. de mathém. 4, Krakau 1926.

Über die Bedeutung des Differenzenansatzes und über weitere sie verwendende Arbeiten vgl. R. Courant, Über direkte Methoden in der Variationsrechnung, Math. Annalen 97, S. 711 und die dort angegebene Literatur.

Es sei G ein Gebiet der Ebene, begrenzt von einer stetigen, doppel-punktfreien geschlossenen Kurve. Dann soll das zugehörige — bei ge-nügend kleiner Maschenweite eindeutig bestimmte — Gittergebiet G_h aus allen denjenigen Gitterpunkten bestehen, welche in G liegen und sich von einem festen vorgegebenen Gitterpunkt aus G durch eine zusammen-hängende Kette von Gitterpunkten verbinden lassen. Wir nennen zusammen-hängende Kette von Gitterpunkten eine Folge solcher Punkte, bei der jeder Punkt einer der vier Nachbarpunkte des folgenden ist. Als Rand-punkt von G_h bezeichnen wir einen solchen, dessen vier Nachbarpunkte nicht alle zu G_h gehören. Alle anderen Punkte von G_h nennen wir innere Punkte.

Wir betrachten Funktionen u, v, \dots des Ortes im Gitter, d. h. Funk-tionen, welche nur für die Gitterpunkte definiert sind. Wir bezeichnen sie auch mit $u(x, y), v(x, y), \dots$. Für ihre vorderen und hinteren Differenzenquotienten verwenden wir die folgenden Abkürzungen:

$$\begin{aligned} \frac{1}{h}(u(x+h, y) - u(x, y)) &= u_x, & \frac{1}{h}(u(x, y+h) - u(x, y)) &= u_y, \\ \frac{1}{h}(u(x, y) - u(x-h, y)) &= u_{\bar{x}}, & \frac{1}{h}(u(x, y) - u(x, y-h)) &= u_{\bar{y}}. \end{aligned}$$

Entsprechend bilden wir Differenzenquotienten höherer Ordnung, z. B.

$$(u_x)_{\bar{x}} = u_{x\bar{x}} = u_{\bar{x}x} = \frac{1}{h^2}(u(x+h, y) - 2u(x, y) + u(x-h, y))$$

usw.

2. Differenzenausdrücke und Greensche Umformungen.

Zu der einfachsten allgemeinen Übersicht über lineare Differen-zenausdrücke zweiter Ordnung gelangen wir nach dem Muster der Theorie der partiellen Differentialgleichungen, indem wir aus zwei Funktionen u und v und ihren vorderen Differenzenquotienten einen bilinearen Ausdruck

$$\begin{aligned} B(u, v) &= a u_x v_x + b u_x v_y + c u_y v_x + d u_y v_y + \alpha u_x v + \beta u_y v + \gamma u v_x \\ &\quad + \delta u v_y + g u v \end{aligned}$$

bilden, wobei

$$a = a(x, y), \dots, \quad \alpha = \alpha(x, y), \dots, \quad g = g(x, y)$$

Funktionen im Gitter sind.

Aus dem Bilinear Ausdruck erster Ordnung leiten wir einen Differen-zenausdruck zweiter Ordnung in folgender Weise ab: Wir bilden die Summe

$$h^2 \sum_{G_h} \sum B(u, v)$$

über alle Punkte eines Gebietes G_h im Gitter, wobei in $B(u, v)$ für die Differenzenquotienten zwischen einem Randpunkte und einem nicht zu G_h

gehörigen Punkte Null zu setzen ist. Die Summe formen wir nun durch partielle Summation um (d. h. wir ordnen nach v), und zerspalten sie in eine Summe über die Menge der inneren Punkte G'_h und eine Summe über die Menge der Randpunkte Γ_h . Wir erhalten so:

$$(1) \quad h^2 \sum_{G_h} \sum B(u, v) = -h^2 \sum_{G'_h} \sum v L(u) - h \sum_{\Gamma_h} v \mathfrak{R}(u).$$

$L(u)$ ist der für alle inneren Punkte von G_h definierte lineare „Differenzenausdruck zweiter Ordnung“:

$$L(u) = (a u_x)_{\bar{x}} + (b u_x)_{\bar{y}} + (c u_y)_{\bar{x}} + (d u_y)_{\bar{y}} \\ - a u_x - \beta u_y + (\gamma u)_{\bar{x}} + (\delta u)_{\bar{y}} - g u.$$

$\mathfrak{R}(u)$ ist für jeden Randpunkt ein linearer Differenzenausdruck, dessen genaue Gestalt wir hier nicht angeben.

Ordnet man $\sum_{G_h} \sum B(u, v)$ nach u , so erhält man

$$(2) \quad h^2 \sum_{G_h} \sum B(u, v) = -h^2 \sum_{G'_h} \sum u M(v) - h \sum_{\Gamma_h} u \mathfrak{S}(v).$$

$M(v)$ heißt der zu $L(u)$ adjungierte Differenzenausdruck; er lautet:

$$M(v) = (a v_x)_{\bar{x}} + (b v_y)_{\bar{x}} + (c v_x)_{\bar{y}} + (d v_y)_{\bar{y}} \\ + (\alpha v)_{\bar{x}} + (\beta v)_{\bar{y}} - \gamma v_x - \delta v_y - g v,$$

während $\mathfrak{S}(v)$ ein $\mathfrak{R}(u)$ entsprechender Differenzenausdruck für den Rand ist.

Die Formeln (1), (2) und die aus ihnen folgende Formel

$$(3) \quad h^2 \sum_{G_h} \sum (v L(u) - u M(v)) + h \sum_{\Gamma_h} (v \mathfrak{R}(u) - u \mathfrak{S}(v)) = 0$$

nennen wir die Greenschen Formeln.

Der einfachste und wichtigste Fall ergibt sich, wenn die Bilinearform symmetrisch ist, d. h. wenn die Gleichungen $b = c$, $\alpha = \gamma$, $\beta = \delta$ bestehen. In diesem Falle stimmt der Ausdruck $L(u)$ mit seinem adjungierten $M(u)$ überein; wir nennen ihn deshalb selbstadjungiert, und er ist schon aus dem quadratischen Ausdruck

$$B(u, u) = a u_x^2 + 2b u_x u_y + d u_y^2 + 2\alpha u_x u + 2\beta u_y u + g u^2$$

ableitbar.

Wir beschränken uns im folgenden meist auf Ausdrücke $L(u)$, die sich selbst adjungiert sind. Der Charakter des Differenzenausdruckes $L(u)$ hängt vor allem von der Natur derjenigen Glieder aus der quadratischen Form $B(u, u)$ ab, die in den ersten Differenzenquotienten quadratisch sind. Wir nennen diesen Teil von $B(u, u)$ die charakteristische Form:

$$P(u, u) = a u_x^2 + 2b u_x u_y + d u_y^2.$$

Je nachdem nun $P(u, u)$ in den Differenzenquotienten (positiv) definit oder indefinit ist, nennen wir den zugehörigen Differenzenausdruck $L(u)$ elliptisch oder hyperbolisch.

Der Differenzenausdruck

$$\Delta u = u_{x\bar{x}} + u_{y\bar{y}},$$

mit dem wir uns vorzugsweise in den folgenden Paragraphen beschäftigen werden, ist elliptisch. Er entsteht nämlich aus dem quadratischen Ausdruck

$$B(u, u) = u_x^2 + u_y^2 \quad \text{bzw.} \quad u_{\bar{x}}^2 + u_{\bar{y}}^2.$$

Die zugehörigen Greenschen Formeln lauten also:

$$(4) \quad h^2 \sum_{G_h} \sum_{G_h} (u_x^2 + u_y^2) = -h^2 \sum_{G_h} \sum_{G_h} u \Delta u - h \sum_{\Gamma_h} u \mathfrak{R}(u),^3$$

$$(5) \quad h^2 \sum_{G_h} \sum_{G_h} (v \Delta u - u \Delta v) + h \sum_{\Gamma_h} (v \mathfrak{R}(u) - u \mathfrak{R}(v)) = 0.$$

Der Differenzenausdruck $\Delta u = u_{x\bar{x}} + u_{y\bar{y}}$ ist offenbar das Analogon des Differentialausdruckes $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}$ für eine Funktion $u(x, y)$ der kontinuierlichen Variablen x und y . Ausführlich geschrieben lautet der Differenzenausdruck

$$\Delta u = \frac{1}{h^2} \{u(x+h, y) + u(x, y+h) + u(x-h, y) + u(x, y-h) - 4u(x, y)\}.$$

Es ist also $\frac{h^2}{4} \Delta u$ der Überschuß des arithmetischen Mittels der Funktionswerte in den vier Nachbarpunkten über den Funktionswert in dem betreffenden Punkt.

Ganz ähnliche Überlegungen führen zu linearen Differenzenausdrücken vierter Ordnung und entsprechenden Greenschen Formeln, wenn wir von bilinearen Differenzenausdrücken ausgehen, welche aus Differenzenquotienten zweiter Ordnung gebildet sind. Wir begnügen uns mit dem Beispiel des Differenzenausdruckes

$$\Delta \Delta u = u_{xx\bar{x}\bar{x}} + 2u_{xx\bar{x}\bar{y}} + u_{yy\bar{y}\bar{y}}.$$

Er entspringt aus dem quadratischen Ausdruck

$$B(u, u) = (u_{x\bar{x}} + u_{y\bar{y}})^2 = (\Delta u)^2,$$

wenn wir die Summe

$$h^2 \sum_{G_h} \sum_{G_h} \Delta u \Delta v$$

³ Der Randausdruck $\mathfrak{R}(u)$ läßt sich hier so beschreiben: Sind u_0, u_1, \dots, u_ν die Funktionswerte in dem betreffenden Randpunkte und seinen ν Nachbarpunkten ($\nu \leq 3$), so ist

$$\mathfrak{R}(u) = \frac{1}{h} (u_1 + \dots + u_\nu - \nu u_0).$$

nach v ordnen, etwa indem wir in der Formel (5) an Stelle von u den Ausdruck Δu setzen. Wir müssen dabei beachten, daß in dem Ausdruck $\Delta \Delta u$ der Funktionswert an einer Stelle mit den Funktionswerten in seinen Nachbarpunkten und deren Nachbarpunkten verknüpft ist und daher nur für solche Punkte des Gebietes G_h definiert ist, die innere Punkte auch von G'_h sind (vgl. 5) und deren Gesamtheit wir mit G''_h bezeichnen wollen. Wir erhalten dann die Greensche Formel

$$(6) \quad \hbar^2 \sum_{G_h} \sum \Delta u \cdot \Delta v = \hbar^2 \sum_{G''_h} \sum v \cdot \Delta \Delta u + \hbar \sum_{\Gamma_h + \Gamma'_h} v \cdot \mathfrak{K}(u),$$

wo $\mathfrak{K}(u)$ ein für jeden Punkt des Randstreifens $\Gamma_h + \Gamma'_h$ definierbarer linearer Differenzenausdruck ist, den wir nicht näher angeben. Γ'_h bedeutet dabei die Menge der Randpunkte von G'_h .

§ 2.

Randwertprobleme und Eigenwertprobleme.

1. Die Theorie des Randwertproblems.

Die Randwertaufgabe für lineare elliptische homogene Differenzengleichungen zweiter Ordnung, welche der klassischen Randwertaufgabe für partielle Differentialgleichungen entspricht, formulieren wir folgendermaßen:

In einem Gittergebiete G_h sei ein selbstadjungierter elliptischer linearer Differenzenausdruck zweiter Ordnung $L(u)$ gegeben. Er möge aus einem quadratischen Ausdruck $B(u, u)$ entspringen, der positiv-definit ist in dem Sinne, daß er nicht verschwinden kann, wenn nicht u_x und u_y selbst verschwinden.

Man bestimme nun in G_h eine solche der Differenzengleichung

$$L(u) = 0$$

genügende Funktion u , welche in den Randpunkten dieses Gittergebietes mit vorgegebenen Werten übereinstimmt.

Unsere Forderung wird dargestellt durch ebenso viele lineare Gleichungen wie es innere Gitterpunkte des Gittergebietes, also zu bestimmende Funktionswerte u gibt⁴⁾. Einige dieser Gleichungen, nämlich soweit sie zu Gitterpunkten gehören, welche mit ihren vier Nachbarn im Innern liegen, sind homogen; andere, bei welchen Randpunkte des Gittergebietes mit eingehen, sind inhomogen. Setzen wir die rechten Seiten dieses in-

⁴⁾ Bildet man zu einer beliebigen Differenzengleichung zweiter Ordnung $L(u) = 0$, indem man sie als ein lineares Gleichungssystem auffaßt, das transponierte Gleichungssystem, so wird dieses durch die adjungierte Differenzengleichung $M(v) = 0$ dargestellt. Die oben betrachtete selbstadjungierte Differenzengleichung stellt also ein lineares Gleichungssystem mit symmetrischem Koeffizientenschema dar.

homogenen Gleichungssystems, d. h. die Randwerte von u gleich Null, so folgt aus der Greenschen Formel (1), wenn wir dort $u = v$ setzen, sofort das Verschwinden von $B(u, u)$ und wegen des Definitheitscharakters von $B(u, u)$ das Verschwinden von u_x, u_y und damit auch von u . Die Differenzgleichung hat also die Lösung $u = 0$, wenn die Randwerte verschwinden, oder mit anderen Worten, die Lösung ist durch die Randwerte, wenn überhaupt, eindeutig bestimmt, da die Differenz zweier Lösungen mit denselben Randwerten verschwinden muß. Wenn aber ein lineares Gleichungssystem mit ebenso vielen Unbekannten wie Gleichungen die Eigenschaft besitzt, daß bei verschwindenden rechten Seiten auch die Unbekannten sämtlich verschwinden müssen, so besagt der Fundamentalsatz der Gleichungstheorie, daß bei beliebig vorgegebenen rechten Seiten genau eine Lösung vorhanden sein muß. In unserem Falle folgt somit die Existenz einer Lösung der Randwertaufgabe.

Wir sehen also, daß bei unseren elliptischen Differenzgleichungen die eindeutige Bestimmtheit und die Existenz der Lösung der Randwertaufgabe durch den Fundamentalsatz der Theorie der linearen Gleichungen miteinander zusammenhängen, während in der Theorie der partiellen Differentialgleichungen bekanntlich beide Tatsachen mit ganz verschiedenen Methoden bewiesen werden müssen. Der Grund für diese Schwierigkeit ist darin zu erblicken, daß Differentialgleichungen nicht mehr mit endlich vielen Gleichungen äquivalent sind; und man sich daher nicht mehr auf die Gleichheit der Anzahl von Unbekannten und Gleichungen berufen kann.

Da die Differenzgleichung

$$\Delta u = 0$$

aus dem positiv-definiten quadratischen Ausdruck

$$h^2 \sum_{G_h} \sum (u_x^2 + u_y^2)$$

entspringt, ist also das Randwertproblem dieser Differenzgleichung stets eindeutig lösbar.

Ganz entsprechend wie für die Differenzgleichungen zweiter Ordnung entwickelt sich die Theorie für Differenzgleichungen höherer, z. B. vierter Ordnung, wofür das Beispiel der Differenzgleichung

$$\Delta \Delta u = 0$$

genügen möge. Hier müssen die Werte der Funktion u in dem Randstreifen $\Gamma_h + \Gamma'_h$ vorgegeben werden. Offenbar liefert auch die Differenzgleichung $\Delta \Delta u = 0$ ebensoviel lineare Gleichungen wie unbekannte Funktionswerte in den Punkten von G_h'' . Um die eindeutige Lösbarkeit der Randwertaufgabe nachzuweisen, brauchen wir wieder nur zu zeigen, daß

eine Lösung, deren Werte im Randstreifen $\Gamma_h + \Gamma'_h$ Null sind, notwendig identisch verschwindet. Zu dem Zweck bemerken wir, daß die Summe über den zugehörigen quadratischen Ausdruck:

$$(7) \quad h^2 \sum_{G'_h} \sum (\Delta u)^2$$

für eine solche Funktion verschwindet, wie wir sofort erkennen, wenn wir diese Summe nach der Greenschen Formel (6) umformen. Das Verschwinden der Summe (7) zieht aber das Verschwinden von Δu in allen Punkten von G'_h nach sich, und das kann bei verschwindenden Randwerten nach dem oben Bewiesenen nur stattfinden, wenn die Funktion u überall den Wert Null annimmt. Damit ist aber unsere Behauptung bewiesen und die eindeutige Lösbarkeit der Randwertaufgabe des Differenzenausdruckes sichergestellt⁵⁾.

2. Beziehungen zu Minimumproblemen.

Die obige Randwertaufgabe steht in Zusammenhang mit dem folgenden Minimumproblem: Unter allen im Gittergebiet G_h definierten Funktionen $\varphi(x, y)$, welche in den Randpunkten vorgeschriebene Werte annehmen, ist eine solche $\varphi = u(x, y)$ zu suchen, für welche die über das Gittergebiet erstreckte Summe

$$h^2 \sum_{G_h} \sum B(\varphi, \varphi)$$

einen möglichst kleinen Wert annimmt. Dabei setzen wir voraus, daß der quadratische Differenzenausdruck erster Ordnung $B(u, u)$ in dem oben (vgl. S. 36) genannten Sinne positiv-definit ist. Daß sich aus dieser Minimumforderung als Bedingung für die Lösung $\varphi = u(x, y)$ die Differenzengleichung $L(\varphi) = 0$ ergibt, wo $L(\varphi)$ der in der obigen (vgl. S. 35 (1)) Weise aus $B(\varphi, \varphi)$ abgeleitete Differenzenausdruck zweiter Ordnung ist, erkennt man entweder nach den Regeln der Differentialrechnung, indem man die Summe $h^2 \sum_{G_h} \sum B(\varphi, \varphi)$ als Funktion der endlichen vielen Werte von φ in den Gitterpunkten ansieht oder analog dem üblichen Verfahren in der Variationsrechnung.

Beispielsweise ist das Randwertproblem, eine Lösung der Gleichung $\Delta \varphi = 0$ zu finden, die vorgegebene Randwerte annimmt, mit der Aufgabe

⁵⁾ Vergleiche für die wirkliche Durchführung der Lösung unserer Randwertprobleme durch iterierende Verfahren u. a. die Abhandlung: Über Randwertaufgaben bei partiellen Differenzengleichungen von R. Courant, Zeitschr. f. angew. Mathematik u. Mechanik 6 (1925), S. 322–325. Im übrigen sei verwiesen auf den Bericht von H. Henky, Zeitschr. f. angew. Math. u. Mech. 2 (1922), S. 58 ff.

gleichwertig, die Summe $h^2 \sum_{G_h} \sum (\varphi_x^2 + \varphi_y^2)$ unter allen Funktionen, die die Randwerte annehmen, zum Minimum zu machen.

Ganz Ähnliches gilt für Differenzengleichungen vierter Ordnung, wobei wir uns wiederum auf das Beispiel von $\Delta \Delta \varphi = 0$ beschränken. Die zu dieser Differenzengleichung gehörige Randwertaufgabe ist mit dem Problem gleichwertig, die Summe $h^2 \sum_{G_h} \sum (\Delta \varphi)^2$ unter allen Funktionen $\varphi(x, y)$ zum Minimum zu machen, deren Werte in dem Randstreifen Γ'_h vorgegeben sind. Außer dieser Summe führen auch noch andere in den zweiten Ableitungen quadratische Ausdrücke durch die Forderung, sie zum Minimum zu machen, auf die Gleichung $\Delta \Delta u = 0$, so z. B. die Summe:

$$h^2 \sum_{G'_h} \sum (u_{xx}^2 + 2u_{xy}^2 + u_{yy}^2),$$

in der sämtliche in G_h auftretenden zweiten Differenzenquotienten vorkommen sollen.

Daß die gestellten Minimumprobleme immer eine Lösung besitzen, folgt aus dem Satz, daß eine stetige Funktion von endlichen vielen Veränderlichen (den Funktionswerten von φ in den Gitterpunkten) stets ein Minimum besitzen muß, wenn diese Funktion nach unten beschränkt ist und wenn sie gegen Unendlich strebt, sobald mindestens eine der unabhängigen Veränderlichen es tut⁶⁾.

3. Die Greensche Funktion.

Ähnlich wie die Randwertaufgabe der homogenen Gleichung $L(u) = 0$ kann man auch die Randwertaufgabe der unhomogenen Gleichung $L(u) = -f$ behandeln. Es genügt, bei der unhomogenen Gleichung sich auf den Fall zu beschränken, daß die Randwerte von u überall verschwinden, da wir für andere Randwerte die Lösung durch Addition einer geeigneten Lösung der homogenen Gleichung erhalten. Um das lineare Gleichungssystem, welches durch die Randwertaufgabe von $L(u) = -f$ repräsentiert ist, zu lösen, wählen wir zunächst die Funktion $f(x, y)$ so, daß sie in einem Gitterpunkte mit den Koordinaten $x = \xi$, $y = \eta$ den Wert $-\frac{1}{h^2}$, in allen andern Gitterpunkten den Wert Null annimmt. Ist $K(x, y; \xi, \eta)$ die am Rande verschwindende Lösung der so entstehenden speziellen noch vom Parameterpunkt (ξ, η) abhängigen Differenzengleichung, so wird die zu

⁶⁾ Daß die Voraussetzungen für die Anwendungen dieses Satzes gegeben sind, ist sehr leicht einzusehen.

einer beliebigen Funktion gehörige Lösung durch die Summe

$$u(x, y) = h^2 \sum_{(\xi, \eta) \text{ in } G_h} K(x, y; \xi, \eta) f(\xi, \eta)$$

dargestellt.

Die Funktion $K(x, y; \xi, \eta)$ in ihrer Abhängigkeit von den Punkten (x, y) und (ξ, η) nennen wir die Greensche Funktion des Differentialausdruckes $L(u)$. Bezeichnen wir mit $\bar{K}(x, y; \xi, \eta)$ die Greensche Funktion des adjungierten Ausdruckes $M(v)$, so gilt die Relation

$$K(\bar{\xi}, \bar{\eta}; \xi, \eta) = \bar{K}(\xi, \eta; \bar{\xi}, \bar{\eta}),$$

die man auch unmittelbar aus der Greenschen Formel (5) folgert, wenn man dort $u = K(x, y; \xi, \eta)$ und $v = \bar{K}(x, y; \bar{\xi}, \bar{\eta})$ setzt. Für einen selbstadjungierten Differenzenausdruck ergibt sich aus der obigen Beziehung die Symmetrierelation:

$$K(\bar{\xi}, \bar{\eta}; \xi, \eta) = K(\xi, \eta; \bar{\xi}, \bar{\eta}).$$

4. Eigenwertprobleme.

Selbstadjungierte Differenzenausdrücke $L(u)$ geben Anlaß zu Eigenwertproblemen von folgendem Typ: Es sind die Werte eines Parameters λ — die Eigenwerte — zu suchen, für die die Differenzengleichung

$$L(u) + \lambda u = 0$$

in G_h eine auf dem Rande Γ_h verschwindende Lösung — die Eigenfunktion — besitzt.

Das Eigenwertproblem ist äquivalent dem Hauptachsenproblem der quadratischen Form $B(u, u)$. Es gibt ebenso viele Eigenwerte $\lambda^{(1)}, \dots, \lambda^{(N)}$ wie innere Gitterpunkte im Gebiet G_h und ebenso viele zugehörige Eigenfunktionen $u^{(1)}, \dots, u^{(N)}$. Das System der Eigenfunktionen und Eigenwerte und ihre Existenz ergibt sich aus dem Minimumproblem:

Unter allen am Rande verschwindenden Funktionen $\varphi(x, y)$, die den $m - 1$ Orthogonalitätsbedingungen

$$h^2 \sum_{G_h} \sum \varphi u^{(v)} = 0 \quad (v = 1, \dots, m - 1)$$

und der Normierungsbedingung

$$h^2 \sum_{G_h} \sum \varphi^2 = 1$$

genügen, ist diejenige $\varphi = u$ gesucht, für die die Summe

$$h^2 \sum_{G_h} \sum B(\varphi, \varphi)$$

den kleinsten Wert annimmt. Der Wert dieses Minimums ist der m -te Eigenwert und die Funktion, für die es angenommen wird, ist die m -te Eigenfunktion⁷⁾.

§ 3.⁸⁾

Zusammenhänge mit dem Problem der Irrwege.

Unser Thema steht in Beziehung zu einer Frage der Wahrscheinlichkeitsrechnung, nämlich dem Problem der Irrwege in einem begrenzten Gebiet⁹⁾. Man stelle sich in einem Gittergebiet G_h die Gitterstrecken als Wege vor, längs deren ein Partikel von einem Gitterpunkt zu einem Nachbarpunkt wandern kann. In diesem Straßennetz möge nun unser Partikel ziellos herumirren, indem es an jeder Straßenecke unter den vier verfügbaren Richtungen eine nach dem Zufall auswählt — alle vier seien gleich wahrscheinlich —. Die Irrfahrt endet, sobald ein Randpunkt von G_h erreicht ist, wo unsere Partikel absorbiert werden mögen.

Wir fragen:

1. Welches ist die Wahrscheinlichkeit $w(P; R)$ daß man bei der Irrfahrt von einem Punkte P ausgehend irgend einmal in dem Randpunkte R ankommt?

2. Welches ist die mathematische Hoffnung $v(P; Q)$, daß man bei einer solchen von P ausgehenden Irrfahrt, ohne den Rand zu treffen, einen Punkt Q von G_h berührt?

⁷⁾ Wegen der Orthogonalität $h^2 \sum_{G_h} u^{(\nu)} u^{(\mu)} = 0$ ($\nu \neq \mu$) der Eigenfunktionen läßt sich jede am Rande verschwindende Funktion $g(x, y)$ des Gitters nach den Eigenfunktionen in der Form

$$y = \sum_{\nu=1}^N c^{(\nu)} u^{(\nu)}$$

entwickeln, wo die Koeffizienten $c^{(\nu)}$ durch die Gleichung

$$c^{(\nu)} = \sum_{G_h} g u^{(\nu)}$$

bestimmt sind.

Auf diese Weise erhalten wir insbesondere die folgende Darstellung der Greenschen Funktion:

$$K(x, y; \xi, \eta) = -\frac{1}{h^2} \sum_{\nu=1}^N \frac{u^{(\nu)}(x, y) \cdot u^{(\nu)}(\xi, \eta)}{\lambda^{(\nu)}}.$$

⁸⁾ Für die Durchführung des Grenzüberganges in § 4 ist § 3 entbehrlich.

⁹⁾ Gerade in der Art wie hier die Grenzen des Gebietes hineinspielen, liegt ein wesentlicher Unterschied der folgenden Betrachtung gegenüber bekannten Überlegungen, die z. B. im Zusammenhange mit der Brownschen Molekularbewegung durchgeführt worden sind.

Diese Wahrscheinlichkeit bzw. mathematische Hoffnung wollen wir durch folgenden Prozeß genauer erklären. Wir denken uns im Punkte P die Einheit irgendeiner Substanzmenge vorhanden. Die Substanz möge sich in unserem Straßennetz mit einer konstanten Geschwindigkeit ausbreiten, etwa in der Zeiteinheit eine Gitterstrecke zurücklegen. In jedem Gitterpunkte soll nach jeder der vier Richtungen genau ein Viertel der dort ankommenden Substanz weiterströmen. Die Substanzmenge, die in einem Randpunkte ankommt, soll dort festgehalten werden. Ist der Ausgangspunkt P ein Randpunkt, so soll die Substanzmenge überhaupt dort bleiben.

Unter der Wahrscheinlichkeit $w(P; R)$ überhaupt bei einer von P ausgehenden Irrfahrt an den Randpunkt R zu gelangen, ohne vorher den Rand berührt zu haben, verstehen wir die Substanzmenge, die sich nach unendlicher Zeit in diesem Randpunkte angesammelt hat.

Unter der Wahrscheinlichkeit $E_n(P; Q)$ in genau n Schritten vom Punkte P zum Punkte Q zu gelangen, ohne den Rand zu berühren, verstehen wir die im Punkte Q nach n Zeiteinheiten befindliche Substanzmenge, falls P und Q innere Punkte sind. Ist P oder Q ein Randpunkt, so setzen wir sie gleich Null.

Die Größe $E_n(P; Q)$ ist gerade die Anzahl der von P nach Q führenden den Rand nicht treffenden Wege von n Schritten, durch 4^n dividiert; es ist also $E_n(P; Q) = E_n(Q; P)$.

Unter der mathematischen Hoffnung $v(P; Q)$ bei einem oben gekennzeichneten Irrwege überhaupt einmal von P aus zum Punkte Q zu gelangen, verstehen wir die unendliche Summe aller dieser Wahrscheinlichkeiten

$$v(P; Q) = \sum_{\nu=0}^{\infty} E_{\nu}(P; Q),^{10)}$$

also für innere Punkte P und Q die Summe aller Substanzmengen, die in den verschiedenen Zeitmomenten den Punkt Q durchlaufen haben. Es wird also dem Erreichen des Punktes Q der Erwartungswert 1 zugeschrieben. Für Randpunkte ist diese Hoffnung gleich Null.

Bezeichnen wir die im Randpunkte R mit genau n Schritten ankommende Menge mit $F_n(P; R)$, so ist die Wahrscheinlichkeit $w(P; R)$ durch die unendliche Reihe

$$w(P; R) = \sum_{\nu=0}^{\infty} F_{\nu}(P; R)$$

dargestellt, deren sämtliche Glieder positiv sind, und deren Teilsummen nie größer als Eins sein können, weil die am Rande ankommende Substanz

¹⁰⁾ Ihre Konvergenz werden wir sogleich beweisen.

nur einen Teil der ursprünglichen Substanzmenge ausmacht. Damit ist aber die Konvergenz dieser Reihe gesichert.

Man kann nun leicht einsehen, daß die Wahrscheinlichkeiten $E_n(P; Q)$, d. h. die nach genau n Schritten in einem Punkte Q anlangende Substanzmenge mit wachsendem n gegen Null strebt. Ist nämlich in irgendeinem Punkte Q , von dem aus ein Randpunkt R in m Schritten zu erreichen sei, $E_n(P; Q) > \alpha > 0$, so wird nach m Schritten in diesem Randpunkt R mindestens die Substanzmenge $\frac{\alpha}{4^m}$ ankommen; da aber wegen der Konvergenz der Summe $\sum_{v=0}^{\infty} F_v(P; R)$ die an den Randpunkt R ankommende Substanzmenge mit der Zeit gegen Null strebt, so müssen auch die Größen $E_n(P; Q)$ selber mit wachsendem n gegen Null streben; d. h. die Wahrscheinlichkeit bei einem unendlich langen Wege im Innern zu bleiben, ist Null.

Hieraus ergibt sich, daß die gesamte Substanzmenge schließlich an den Rand ankommen muß; mit anderen Worten, daß die über alle Randpunkte R erstreckte Summe

$$\sum_R w(P; R) = 1$$

ist.

Wir haben noch die Konvergenz der unendlichen Reihe für die mathematische Hoffnung $v(P; Q)$

$$v(P; Q) = \sum_{v=0}^{\infty} E_v(P; Q)$$

zu beweisen.

Zu dem Zweck bemerken wir, daß die Größen $E_n(P; Q)$ der folgenden Relation genügen

$$E_{n+1}(P; Q) = \frac{1}{4} \{E_n(P; Q_1) + E_n(P; Q_2) + E_n(P; Q_3) + E_n(P; Q_4)\} \\ [n \geq 1],$$

wo Q_1 bis Q_4 die vier Nachbarpunkte des inneren Punktes Q sind. D. h. die nach $n+1$ Schritten im Punkte Q ankommende Substanzmenge besteht aus dem vierten Teil der nach n Schritten in den vier Nachbarpunkten von Q ankommenden Substanzmenge. Ist einer der Nachbarpunkte von Q z. B. $Q_1 = R$ Randpunkt, so kommt die Tatsache, daß zum Punkte Q von diesem Randpunkte aus keine Substanzmenge weiter fließt, dadurch zum Ausdruck, daß wir $E_n(P; R)$ gleich Null gesetzt haben. Ferner ist für einen inneren Punkt $E_0(P; P) = 1$ und sonst $E_0(P; Q) = 0$.

Aus diesen Relationen ergeben sich für die Teilsummen

$$v_n(P; Q) = \sum_{v=0}^n E_v(P; Q)$$

die Gleichungen

$$v_{n+1}(P; Q) = \frac{1}{4} \{v_n(P; Q_1) + v_n(P; Q_2) + v_n(P; Q_3) + v_n(P; Q_4)\},$$

wenn P nicht mit Q zusammenfällt; andernfalls ist

$$v_{n+1}(P; P) = 1 + \frac{1}{4} \{v_n(P; P_1) + v_n(P; P_2) + v_n(P; P_3) + v_n(P; P_4)\},$$

d. h. die Hoffnung, von einem Punkte zu sich selbst zurückzukommen, setzt sich zusammen aus der Hoffnung, auf einem nicht verschwindenden Wege den Punkt P wieder zu erreichen, nämlich $\frac{1}{4} \{v_n(P; P_1) + v_n(P; P_2) + v_n(P; P_3) + v_n(P; P_4)\}$ und aus der Hoffnung Eins, die ausdrückt, daß ursprünglich die gesamte Substanz in diesem Punkte vorhanden war.

Es genügen also die Größen $v_n(P; Q)$ der folgenden Differenzengleichung¹¹⁾

$$\Delta v_n(P; Q) = \frac{4}{h^2} E_n(P; Q), \quad \text{wenn } P \neq Q \text{ ist,}$$

$$\Delta v_n(P; Q) = \frac{4}{h^2} (E_n(P; Q) - 1), \quad \text{wenn } P = Q \text{ ist.}$$

$v_n(P; Q)$ ist gleich Null, wenn Q ein Randpunkt ist.

Die Lösung dieser Randwertaufgabe ist, wie schon früher auseinandergesetzt, für irgendwelche rechten Seiten eindeutig bestimmt (vgl. S. 38); sie hängt stetig von den rechten Seiten ab. Da nun die Größen $E_n(P; Q)$ gegen Null streben, so konvergieren die Lösungen $v_n(P; Q)$ gegen die Lösungen $v(P; Q)$ der Differenzgleichung

$$\Delta v(P; Q) = 0, \quad \text{wenn } P \neq Q \text{ ist,}$$

$$\Delta v(P; Q) = -\frac{4}{h^2}, \quad \text{wenn } P = Q \text{ ist,}$$

mit den Randwerten $v(P; R) = 0$.

¹¹⁾ Dabei bezieht sich die Δ -Operation auf den variablen Punkt Q .

Diese Gleichung läßt sich als eine Gleichung vom Wärmeleitungstypus auffassen. Betrachtet man nämlich die Funktion $v_n(P; Q)$ anstatt als Funktion des Index n unserer oben zugrunde gelegten Vorstellung gemäß als Funktion der Zeit t , die zu n proportional ist, indem man $t = n\tau$ und $v_n(P; Q) = v(P; Q; t) = v(t)$ setzt, so können wir die obigen Gleichungen in der folgenden Form schreiben;

$$\Delta v(t) = \frac{4\tau}{h^2} \cdot \frac{v(t+\tau) - v(t)}{\tau} \quad \text{für } P \neq Q,$$

$$\Delta v(t) = \frac{4\tau}{h^2} \left(\frac{v(t+\tau) - v(t)}{\tau} - 1 \right) \quad \text{für } P = Q.$$

Über den Grenzübergang von einer ähnlichen Differenzgleichung zu einer parabolischen Differentialgleichung vgl. Teil II, § 6, S. 67.

Wir sehen also, daß die mathematische Hoffnung $v(P; Q)$ existiert und nichts anderes ist als die zur Differenzgleichung $\Delta u = 0$ zugehörige Greensche Funktion $K(P; Q)$ noch mit dem Faktor 4 versehen. Die Symmetrie der Greenschen Funktion $K(P; Q) = K(Q; P)$ ist eine unmittelbare Folge der Symmetrie der Größen $E_n(P; Q)$, mit deren Hilfe sie definiert wurde.

Die Wahrscheinlichkeit $w(P; R)$ genügt hinsichtlich P der Relation

$$w(P; R) = \frac{1}{4} \{w(P_1; R) + w(P_2; R) + w(P_3; R) + w(P_4; R)\},$$

also der Differenzgleichung

$$\Delta w = 0.$$

Sind nämlich P_1, P_2, P_3, P_4 die vier Nachbarpunkte des inneren Punktes P , so muß jeder Weg von P nach R über einen dieser vier Wege führen, und jede der vier Wegrichtungen ist gleich wahrscheinlich. Ferner ist die Wahrscheinlichkeit, von einem Randpunkt R zu einem andern R' zu gelangen, $w(R, R') = 0$, außer wenn die beiden Punkte R und R' zusammenfallen, wo $w(R, R) = 1$ gilt. Es ist also $w(P; R)$ die Lösung der Randwertaufgabe $\Delta w = 0$, wobei im Randpunkte R der Randwert 1 in allen anderen Punkten der Randwert 0 vorgeschrieben ist. Die Lösung der Randwertaufgabe bei beliebig vorgegebenen Randwerten $u(R)$ hat dann einfach die Gestalt $u(P) = \sum_R w(P; R) u(R)$, wobei über alle Randpunkte R zu summieren ist¹²⁾. Setzen wir hierin für u die Funktion $u \equiv 1$ ein, so erhalten wir wieder die Relation $1 = \sum_R w(P; R)$.

Die hier gegebene Auffassung der Greenschen Funktion als Hoffnung läßt unmittelbar weitere Eigenschaften erkennen. Wir erwähnen nur die Tatsache, daß die Greensche Funktion wächst, wenn man von dem Gebiete G zu einem in G als Teilgebiet enthaltenen Teilgebiete \bar{G} übergeht; es wächst dann nämlich für jedes n die Anzahl der möglichen Gitterwege, von einem Punkte P zu einem anderen Q zu gelangen, ohne den Rand zu berühren.

Natürlich herrschen für mehr als zwei unabhängige Veränderliche entsprechende Beziehungen. Wir begnügen uns mit dem Hinweis, daß auch andere elliptische Differenzgleichungen eine ähnliche Wahrscheinlichkeitsauffassung zulassen.

Führt man den Grenzübergang zu verschwindender Maschenweite durch, was sich mit den Methoden des folgenden Paragraphen einfach ausführen

¹²⁾ Man erkennt übrigens leicht, daß die Wahrscheinlichkeit $w(P; R)$, an den Rand zu gelangen, der von der Greenschen Funktion $K(P; Q)$ hinsichtlich Q gebildete Randausdruck $\Re(K(P, Q))$ ist, indem man in der Greenschen Formel (5) $u(x, y)$ mit $w(P, Q)$, $v(x, y)$ mit $v(P, Q)$ identifiziert.

läßt, so geht die Greensche Funktion im Gitter bis auf einen Zahlenfaktor in die Greensche Funktion der Potentialgleichung über; eine ähnliche Beziehung besteht zwischen dem Ausdruck $\frac{w(P; R)}{h}$ und der normalen Ableitung der Greenschen Funktion am Rande des Gebietes. Auf diese Weise ließe sich z. B. die Greensche Funktion der Potentialgleichung als die spezifische mathematische Hoffnung deuten, von einem Punkte zu einem anderen zu gelangen¹³⁾, ohne den Rand zu berühren.

Nach dem Grenzübergang vom Gitter zum Kontinuum ist der Einfluß der bei den Irrwegen vorgeschriebenen Gitterrichtungen verschwunden. Dieser Tatsache Rechnung zu tragen, indem man den Grenzübergang mit einem allgemeineren Irrfahrtenproblem ohne Richtungsbeschränkung vornimmt, ist eine prinzipiell interessante Aufgabe, welche jedoch über den Rahmen dieser Abhandlung hinaus führt und auf die wir bei anderer Gelegenheit zurückzukommen hoffen.

§ 4.

Grenzübergang zur Lösung der Differentialgleichung.

1. Die Randwertaufgabe der Potentialtheorie.

Bei der Durchführung des Grenzüberganges von der Lösung der Differenzengleichungsprobleme zu der Lösung der entsprechenden Differentialgleichungen wollen wir hinsichtlich des Randes und der Randwerte auf die größtmögliche Allgemeinheit in der Formulierung verzichten, um das für unsere Methoden Charakteristische klarer hervortreten zu lassen¹⁴⁾. Wir setzen demgemäß voraus, daß in der Ebene ein einfach zusammenhängendes Gebiet G vorgegeben ist, dessen Berandung aus endlich vielen mit stetiger Tangente versehenen Kurvenbögen gebildet wird. In einem G im Innern enthaltenden Gebiete sei eine stetige und mit stetigen partiellen Ableitungen erster und zweiter Ordnung versehene Funktion $f(x, y)$ gegeben. Für das zu der Maschenweite h und zum Gebiete G gehörige Gittergebiet G_h sei die Randwertaufgabe der Differenzengleichung $\Delta u = 0$ mit denjenigen Randwerten, welche von der Funktion $f(x, y)$ in den Randpunkten von G_h angenommen werden, gelöst; die Lösung heiße $u_h(x, y)$. Wir wollen beweisen, daß die Gitterfunktion u_h mit verschwindender Maschenweite h gegen die Lösung u der Randwertaufgabe der partiellen

¹³⁾ Dabei ist dem Erreichen eines Flächenstücks als Erwartungswert sein Flächeninhalt zugeschrieben.

¹⁴⁾ Es sei jedoch bemerkt, daß die Ausdehnung unserer Methoden auf allgemeinere Ränder und Randwerte keinerlei prinzipielle Schwierigkeiten bereitet.

Differentialgleichung $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$ für das Gebiet G konvergiert, wobei die Randwerte für das Gebiet G wiederum durch diejenigen Werte geliefert werden, welche die Funktion $f(x, y)$ auf dem Rande von G annimmt. Weiter werden wir zeigen, daß für jedes ganz im Innern von G liegende Gebiet die Differenzenquotienten beliebiger Ordnung von u_h gleichmäßig gegen die entsprechenden partiellen Differentialquotienten der Grenzfunktion $u(x, y)$ streben.

Bei der Durchführung des Konvergenzbeweises ist es bequem, die Forderung, daß $u(x, y)$ die Randwerte annimmt, durch die folgende schwächere Forderung zu ersetzen: Ist S_r derjenige Randstreifen des Gebietes G , dessen Punkte vom Rande eine Entfernung kleiner als r besitzen, so strebt das Integral

$$\frac{1}{r} \iint_{S_r} (u - f)^2 dx dy$$

mit abnehmendem r gegen Null¹⁵⁾.

Unser Konvergenzbeweis beruht auf der Tatsache, daß für jedes ganz im Innern des Gebietes G liegende Teilgebiet G^* die Funktion $u_h(x, y)$ und jeder Differenzenquotient bei abnehmendem h beschränkt bleibt und „gleichartig stetig“ ist in folgendem Sinne: Es gibt für jede dieser Funktionen $w_h(x, y)$ eine nur von dem Gebiete und nicht von h abhängige Größe $\delta(\varepsilon)$ derart, daß

$$|w_h(P) - w_h(P_1)| < \varepsilon$$

ist, sobald die beiden Gitterpunkte P und P_1 des Gittergebietes G_h in dem gegebenen Teilgebiet liegen und voneinander einen kleineren Abstand als $\delta(\varepsilon)$ besitzen.

¹⁵⁾ Daß tatsächlich unsere schwächere Randwertforderung zur *eindeutigen* Kennzeichnung der Lösung genügt, folgt aus dem leicht zu beweisenden Satze: Wenn für eine im Innern von G der Differentialgleichung $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$ genügende Funktion die obige Form der Randbedingung mit $f(x, \hat{y}) = 0$ erfüllt ist und $\iint_G \left(\left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial u}{\partial y} \right)^2 \right) dx dy$ existiert, so ist $u(x, y)$ identisch Null. (Vgl. Courant,

„Über die Lösungen der Diff.-Gl. der Physik“, Math. Annalen 85, insbesondere S. 296 ff.)

Im Falle von zwei unabhängigen Veränderlichen läßt sich aus unserer schwächeren Forderung die tatsächliche Annahme der Randwerte folgern; im Falle von mehr Variablen darf man das Entsprechende schon deswegen nicht allgemein erwarten, weil es dort bekanntlich Ausnahmepunkte am Rande geben kann, in denen die Randwerte nicht mehr angenommen zu werden brauchen, während jedoch für die schwächere Forderung stets eine Lösung existiert.

Haben wir einmal die behauptete gleichartige Stetigkeit bewiesen, so können wir bekanntlich eine Teilfolge unserer Funktionen u_n so auswählen, daß sie mit ihren Differenzenquotienten jeder Ordnung in jedem Teilgebiet G^* gleichmäßig gegen eine Grenzfunktion $u(x, y)$ bzw. deren Differentialquotienten strebt. Die Grenzfunktion besitzt dementsprechend Ableitungen beliebig hoher Ordnung in jedem inneren Teilgebiet G^* von G und genügt dort der partiellen Differentialgleichung $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$. Wenn wir dann noch zeigen, daß sie die Randbedingung befriedigt, so erkennen wir in ihr die Lösung unseres Randwertproblems für das Gebiet G . Da diese Lösung eindeutig bestimmt ist, so zeigt sich nachträglich, daß nicht nur eine Teilfolge der Funktionen u_n , sondern diese Funktionenfolge selbst die ausgesprochene Konvergenzeigenschaft besitzt.

Die gleichartige Stetigkeit unserer Größen wird sich durch den Nachweis folgender Tatsachen ergeben:

1. Bei abnehmendem h bleiben die über das Gittergebiet G_h erstreckten Summen

$$h^2 \sum_{G_h} \sum u^2 \quad \text{und} \quad h^2 \sum_{G_h} \sum (u_x^2 + u_y^2)$$

beschränkt¹⁶⁾.

2. Genügt $w = w_h$ in einem Gitterpunkt G_h der Differenzengleichung $\Delta w = 0$ und bleibt bei abnehmendem h die Summe

$$h^2 \sum_{G_h^*} \sum w^2,$$

erstreckt über ein zu einem Teilgebiet G^* von G gehöriges Gittergebiet G_h^* , beschränkt, so bleibt für jedes feste ganz im Innern von G^* liegende Teilgebiet G^{**} auch die über das zugehörige Gittergebiet G_h^{**} erstreckte Summe

$$h^2 \sum_{G_h^{**}} \sum (w_x^2 + w_y^2)$$

bei abnehmendem h beschränkt.

Zusammen mit 1. folgt hieraus, da sämtliche Differenzenquotienten w der Funktion u_h wieder der Differenzengleichung $\Delta w = 0$ genügen, daß jede der Summen

$$h^2 \sum_{G_h^*} \sum w^2$$

beschränkt ist.

3. Aus der Beschränktheit dieser Summen folgt schließlich die Beschränktheit und gleichartige Stetigkeit aller Differenzenquotienten selbst.

¹⁶⁾ Hier und gelegentlich im folgenden lassen wir bei Gitterfunktionen den Index h fort.

2. Beweis der Hilfssätze.

Der Beweis der Tatsache 1 folgt daraus, daß die Funktionswerte u_n selbst beschränkt sind. Denn der größte und der kleinste Wert der Funktion wird am Rande angenommen¹⁷⁾, strebt also gegen vorgegebene endliche Werte. Die Beschränktheit der Summe $h^2 \sum_{G_h} \sum_{G_h} (u_x^2 + u_y^2)$ ist eine unmittelbare Folge der im § 2, 2. formulierten Minimumeigenschaft unserer Gitterfunktion, wonach sicherlich

$$h^2 \sum_{G_h} \sum_{G_h} (u_x^2 + u_y^2) \leq h^2 \sum_{G_h} \sum_{G_h} (f_x^2 + f_y^2)$$

gilt. Die Summe rechts strebt aber mit abnehmender Maschenweite gegen das Integral $\int_G \int_G \left[\left(\frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial f}{\partial y} \right)^2 \right] dx dy$, welches nach unseren Voraussetzungen existiert.

Um den unter 2. formulierten Hilfssatz zu beweisen, betrachten wir die Quadratsumme

$$h^2 \sum_{Q_1} \sum_{Q_1} (w_x^2 + w_x^2 + w_y^2 + w_y^2),$$

wobei die Summation sich auf alle inneren Punkte eines Quadrates Q_1 bezieht (vgl. Fig. 1). Die Funktionswerte auf den äußeren Seiten S_1 des Quadrates Q_1 bezeichnen wir mit w_1 , die auf der zweiten Randreihe S_0 mit w_0 . Dann liefert die Greensche Formel

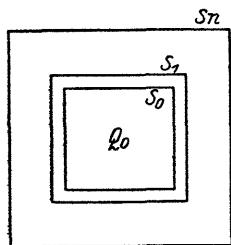


Fig. 1.

$$(8) \quad \begin{aligned} h^2 \sum_{Q_1} \sum_{Q_1} (w_x^2 + w_x^2 + w_y^2 + w_y^2) \\ = \sum_S (w_1^2 - w_0^2) \leq \sum_{S_1} w^2 - \sum_{S_0} w^2, \end{aligned}$$

wobei die Summation rechts über die beiden äußeren Randreihen S_1 und S_0 zu erstrecken ist, und wo w_1 und w_0 sich auf benachbarte Punkte beziehen. Wir betrachten nun eine Reihe von konzentrischen Quadraten $Q_0, Q_1, Q_2, \dots, Q_N$ mit den Rändern S_0, S_1, \dots, S_N , von denen jedes aus dem vorangehenden dadurch entsteht, daß der Kranz der nächsten Nachbarpunkte hinzukommen wird (vgl. Fig. 1). Auf jedes dieser Quadrate wenden wir die Abschätzung (8) an und beachten, daß stets

$$2 h^2 \sum_{Q_0} \sum_{Q_0} (w_x^2 + w_y^2) \leq h^2 \sum_{Q_k} \sum_{Q_k} (w_x^2 + w_x^2 + w_y^2 + w_y^2)$$

¹⁷⁾ Ausdrücklich bemerken wir im Hinblick auf die Übertragung der Methode auf andere Differentialgleichungen, daß wir uns von dieser Eigenschaft unabhängig machen können. Dazu brauchen wir nur die Ungleichung (15) heranzuziehen oder die Schlußweise der Alternative anzuwenden (vgl. S. 55).

für $k \geq 1$ ist. Addieren wir der Reihe nach die n Ungleichungen

$$2h^2 \sum_{Q_0} \sum (w_x^2 + w_y^2) \leq \sum_{S_{k+1}} w^2 - \sum_{S_k} w^2 \quad (0 \leq k < n),$$

so erhalten wir

$$2nh^2 \sum_{Q_0} \sum (w_x^2 + w_y^2) \leq \sum_{S_n} w^2 - \sum_{S_0} w^2 \leq \sum_{S_n} w^2.$$

Diese Ungleichung summieren wir von $n=1$ bis $n=N$. So ergibt sich

$$N^2 h^2 \sum_{Q_0} \sum (w_x^2 + w_y^2) \leq \sum \sum w^2,$$

wobei wir die Summe rechts nur vergrößern, wenn wir sie über das ganze Quadrat Q_N erstrecken.

Lassen wir nun bei Verkleinerung der Maschenweite die Quadrate Q_0 und Q_N gegen zwei feste im Innern von G liegende konzentrische Quadrate mit dem Abstände a streben, so konvergiert Nh gegen a , und wir finden, daß unabhängig von der Maschenweite

$$(9) \quad h^2 \sum_{Q_0} \sum (w_x^2 + w_y^2) \leq \frac{1}{a^2} h^2 \sum_{Q_N} \sum w^2$$

bleibt.

Diese Ungleichung gilt — bei hinreichend kleiner Maschenweite — natürlich nicht nur für zwei Quadrate Q_0 und Q_N , sondern mit einer anderen Konstanten a für irgend zwei Teilgebiete von G , von denen das eine ganz im Innern des anderen liegt. Damit ist die Behauptung von 2. bewiesen¹⁸⁾.

Um nun drittens nachzuweisen, daß in jedem inneren Teilgebiet die Funktion w_h und ihre sämtlichen Differenzenquotienten w_h die Verfeinerung der Maschenweite beschränkt und gleichartig stetig bleiben, betrachten wir ein Rechteck R mit den Eckpunkten P_0, Q_0, P, Q (vgl. Fig. 2), dessen Seiten $P_0 Q_0$ und PQ der x -Achse parallel sind und die Länge a haben.

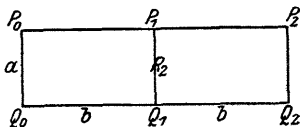


Fig. 2.

Wir gehen aus von der Darstellung

$$w(Q_0) - w(P_0) = h \sum_{PQ} w_x + h^2 \sum_R \sum w_{xy}$$

¹⁸⁾ Wenn wir nicht annehmen, daß $\Delta w = 0$ ist, so erhalten wir an Stelle der Ungleichung (9)

$$(10) \quad h^2 \sum_{G^*} \sum (w_x^2 + w_y^2) \leq c_1 h^2 \sum_{G^*} w^2 + c_2 h^2 \sum_{G^*} (\Delta w)^2$$

bei geeigneten von h unabhängigen Konstanten c_1, c_2 , wobei G^{**} ganz im Innern des Gebietes G^* liegt, das seinerseits im Innern von G enthalten ist.

und der aus ihr folgenden Ungleichung

$$(11) \quad |w(Q_0) - w(P_0)| \leq h \sum_{PQ} |w_x| + h^2 \sum_R \sum |w_{xy}|.$$

Wir lassen nun die Rechtecksseite PQ zwischen einer Anfangslage $P_1 Q_1$ im Abstände b von $P_0 Q_0$ und einer Endlage $P_2 Q_2$ im Abstände $2b$ von $P_0 Q_0$ laufen und summieren die $\frac{b}{h} + 1$ zugehörigen Ungleichungen (11). Wir erhalten so die Abschätzung

$$|w(P_0) - w(Q_0)| \leq \frac{1}{b+h} h^2 \sum_{R_2} \sum |w_x| + h^2 \sum_{R_2} \sum |w_{xy}|,$$

indem wir die Summationsgebiete auf das ganze Rechteck $R_2 = P_0 Q_0 P_2 Q_2$ ausdehnen. Nach der Schwarzischen Ungleichung folgt daraus:

$$(12) \quad |w(P_0) - w(Q_0)| \leq \frac{1}{b} \sqrt{2ab} \sqrt{h^2 \sum_{R_2} \sum w_x^2} + \sqrt{2ab} \sqrt{h^2 \sum_{R_2} \sum w_{xy}^2}.$$

Da die hier auftretenden mit h^2 multiplizierten Summen nach Annahme beschränkt bleiben, so folgt, daß die Differenz $|w(P_0) - w(Q_0)|$ zugleich mit ihrem Abstände a gegen Null strebt und zwar unabhängig von der Maschenweite, da wir für jedes Teilgebiet G^* von G die Größe b festhalten können. Damit ist die gleichartige Stetigkeit von $w = w_h$ in der x -Richtung bewiesen. Entsprechend ergibt sie sich für die y -Richtung und damit für jedes innere Teilgebiet G^* von G . Die Beschränktheit der Funktion w_h in G^* folgt schließlich aus ihrer gleichartigen Stetigkeit und der Beschränktheit von $h^2 \sum_{G^*} \sum w_h^2$.

Mit diesem Nachweis ist die Existenz einer Teilfolge von Funktionen u gesichert, welche gegen eine Grenzfunktion $u(x, y)$ konvergiert und zwar mit sämtlichen Differenzenquotienten in dem oben gekennzeichneten Sinne gleichmäßig für jedes innere Teilgebiet von G . Diese Grenzfunktion $u(x, y)$ besitzt also in G überall stetige partielle Differentialquotienten beliebiger Ordnung und genügt der partiellen Differentialgleichung des Potentials

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0.$$

3. Die Randbedingung.

Um zu beweisen, daß die Lösung die oben formulierte Randbedingung erfüllt, zeigen wir zunächst, daß für jede Gitterfunktion v die Ungleichung

$$(13) \quad h^2 \sum_{S_r, h} v^2 \leq A r^2 h^2 \sum_{S_r, h} (v_x^2 + v_y^2) + B r h \sum_{\Gamma_h} v^2$$

besteht, wo $S_{r,h}$ derjenige Teil des Gittergebietes G ist, der innerhalb des Randstreifens S_r liegt. Dieser Randstreifen S_r war (vgl. S. 48) aus allen Punkten von G gebildet, deren Abstand vom Rande kleiner als r ist; er wird außer von Γ noch von einer Kurve Γ_r begrenzt. Ferner bedeuten A und B nur vom Gebiet und nicht von der Funktion v oder von der Maschenweite h abhängige Konstanten.

Um die obige Ungleichung nachzuweisen, zerlegen wir den Rand Γ von G in eine endliche Anzahl von Stücken, für welche der Winkel der Tangente entweder mit der x -Achse oder mit der y -Achse oberhalb einer positiven Schranke (etwa 30°) bleibt. Es sei z. B. γ ein solches zur

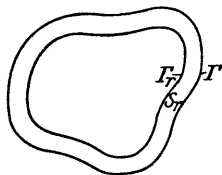


Fig. 3.

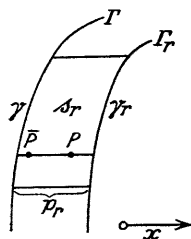


Fig. 4.

x -Achse hinreichend steil geneigtes Stück von Γ (vgl. Fig. 4). Die Parallelen zur x -Achse durch die Endpunkte des Stückes γ schneiden aus der Näherungskurve Γ_r ein Stück γ_r aus und begrenzen zusammen mit γ und γ_r ein Stück s_r des Randstreifens S_r . Der in dem Streifen s_r enthaltene Teil des Gittergebietes G_h heiße $s_{r,h}$ und der zugehörige Teil des Randes Γ_h heiße γ_h .

Wir denken uns durch einen Gitterpunkt P_h von $s_{r,h}$ die Parallele zur x -Achse gezogen. Sie trifft den Rand γ_h in einem Punkte \bar{P}_h . Dasjenige Stück dieser Parallelen, das in $s_{r,h}$ liegt, bezeichnen wir mit $p_{r,h}$. Seine Länge ist sicher kleiner als cr , da r der größte senkrechte Abstand eines Punktes aus S_r von Γ ist. Dabei hängt die Konstante c nur von dem kleinsten Neigungswinkel einer Tangente von γ mit der x -Achse ab.

Nun besteht zwischen dem Wert von v im Punkte P_h und ihrem Werte in \bar{P}_h die Beziehung

$$v(P_h) = v(\bar{P}_h) \pm h \sum_{P_h \bar{P}_h} v_x,$$

woraus sich durch Quadrieren und Anwendung der Schwarzschen Ungleichung

$$v(P_h)^2 \leq 2v(\bar{P}_h)^2 + 2cr \cdot h \sum_{P_h \bar{P}_h} v_x^2$$

ergibt. Summieren wir hinsichtlich P_h in der x -Richtung, so erhalten wir

$$h \sum_{p_r} v^2 \leq 2cr v(\bar{P}_h)^2 + 2c^2 r^2 h \sum_{p_r} v_x^2.$$

Summieren wir noch einmal in der y -Richtung, so entsteht die Relation

$$(14) \quad h \sum_{S_r} \sum_{\Gamma_h} v^2 \leq 2cr \sum_{\Gamma_h} v(\bar{P}_h) + 2c^2 r^2 h \sum_{S_r} \sum_{\Gamma_h} v_x^2,$$

die wir nur noch für die anderen Stücke γ von Γ entsprechend aufzustellen und dann zu addieren haben, um leicht die gewünschte Ungleichung (13) zu erhalten¹⁹⁾.

Wir setzen nun

$$v_h = u_h - f_h,$$

sodaß v_h am Rande Γ_h verschwindet. Da dann $h^2 \sum_{G_h} \sum_{\Gamma_h} (v_x^2 + v_y^2)$ bei abnehmendem h beschränkt bleibt, so erhalten wir aus (13)

$$(16) \quad \frac{h^2}{r} \sum_{S_{r,h}} \sum_{\Gamma_h} v^2 \leq \kappa r,$$

wo κ eine nicht von der Funktion v oder der Maschenweite abhängige Konstante ist. Erstrecken wir die Summe links nicht über den ganzen Randstreifen $S_{r,h}$, sondern nur über die Differenz von zwei solchen; $S_{v,h} - S_{e,h}$, so bleibt die Ungleichung (16) mit der selben Konstanten κ gültig, und wir können den Grenzübergang zu verschwindender Maschenweite vollziehen. Aus der Ungleichung (16) entsteht dann

$$\frac{1}{r} \iint_{S_r - S_e} v^2 dx dy \leq \kappa r, \quad v = u - f.$$

Lassen wir nun den kleineren Randstreifen S_e dem Rande zustreben, so erhalten wir die Ungleichung

$$\frac{1}{r} \iint_{S_r} v^2 dx dy = \frac{1}{r} \iint_{S_r} (u - f)^2 dx dy \leq \kappa r,$$

die gerade ausdrückt, daß die Grenzfunktion u die von uns geforderte Randbedingung erfüllt.

¹⁹⁾ Durch dieselbe Betrachtungsweise, die zum Nachweis der Ungleichung (13) führt, läßt sich auch die Ungleichung

$$(15) \quad h^2 \sum_{G_h} \sum_{\Gamma_h} v^2 \leq c_1 h \sum_{\Gamma_h} v^2 + c_2 h^2 \sum_{G_h} \sum_{\Gamma_h} (v_x^2 + v_y^2)$$

ableiten, in der die Konstanten c_1, c_2 nur vom Gebiet G , aber nicht von der Maschen-einteilung abhängen.

4. Anwendbarkeit der Methode auf andere Probleme.

Unsere Methode stützt sich wesentlich auf die in dem obigen Hilfsatz ausgesprochene Ungleichheitsbeziehung (10)²⁰⁾, weil aus ihr die beiden letzten auf S. 49 genannten Hauptpunkte des Beweises folgen; sie macht keinerlei Gebrauch von speziellen Grundlösungen oder sonstigen speziellen Eigenschaften unserer Differenzenausdrücke und läßt sich daher unmittelbar sowohl auf den Fall von beliebig vielen unabhängigen Variablen als auf das Eigenwertproblem der Differentialgleichung $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \lambda u = 0$ übertragen und liefert dabei hinsichtlich der Konvergenzverhältnisse genau dieselben Resultate wie oben²¹⁾. Auch eine Übertragung auf lineare Differentialgleichungen anderer Art, insbesondere solche mit nicht konstanten Koeffizienten, erfordert nur einige naheliegende Modifikationen. Der wesentliche Unterschied besteht immer nur im Nachweis der Beschränktheit von $h^2 \sum \sum u_n^2$, die allerdings nicht bei einem beliebigen solchen linearen Probleme vorliegt. Aber im Falle der Unbeschränktheit dieser Summe läßt sich zeigen, daß das allgemeine Randwertproblem der betreffenden Differentialgleichung auch wirklich keine Lösung besitzt, daß aber dafür in diesem Falle nicht verschwindende Lösungen des zugehörigen *homogenen* Problems, d. h. Eigenfunktionen, existieren²²⁾.

5. Das Randwertproblem von $\Delta \Delta u = 0$.

Um zu zeigen, daß sich die Methode auch auf den Fall von Differentialgleichungen höherer Ordnung übertragen läßt, behandeln wir im folgenden kurz das Randwertproblem der Differentialgleichung

$$\frac{\partial^4 u}{\partial x^2} + 2 \frac{\partial^4 u}{\partial x^2 \partial y^2} + \frac{\partial^4 u}{\partial y^4} = 0.$$

Wir suchen eine Lösung dieser partiellen Differentialgleichung in unserem Gebiete G , für welche die Funktionswerte und ihre ersten Ableitungen am Rande vorgegeben sind, und zwar durch diejenigen Werte, welche von einer vorgegebenen Funktion $f(x, y)$ am Rande definiert werden.

²⁰⁾ Hinsichtlich der Anwendung entsprechender Integralungleichungen vgl. K. Friedrichs, Die Rand- und Eigenwertprobleme aus der Theorie der elastischen Platten, Math. Annalen 98, S. 222.

²¹⁾ Es ist dann zugleich bewiesen, daß jede Lösung eines solchen Differentialgleichungsproblems Ableitungen jeder Ordnung besitzt.

²²⁾ Vgl. Courant-Hilbert, Methoden der mathematischen Physik, 1, Kap. III, § 3, wo mit Hilfe einer entsprechenden Alternative die Theorie der Integralgleichungen behandelt wird. Vgl. auch die demnächst erscheinende Göttinger Dissertation von W. v. Koppenfels.

Dabei setzen wir wie oben (S. 47) voraus, daß $f(x, y)$ in einem das Gebiet G enthaltenden Gebiete der Ebene mit den ersten und zweiten Ableitungen stetig ist.

Wir ersetzen unser Differentialgleichungsproblem durch die Aufgabe, die Differenzgleichung $\Delta \Delta u = 0$ für das Gittergebiet G zu lösen, wobei in den Punkten des Randstreifens $\Gamma_h + \Gamma'_h$ die Funktion u dieselben Werte wie die vorgegebene Funktion $f(x, y)$ annehmen soll. Nach § 2 wissen wir, daß diese Randwertaufgabe für G_h auf eine und nur eine Weise lösbar ist. Wir werden zeigen, daß bei Verfeinerung der Maschenweite h diese Lösung in jedem inneren Teilgebiet von G mit allen Differenzenquotienten gegen die Lösung unserer Differentialgleichung bzw. gegen die entsprechenden Differentialquotienten konvergiert.

Zu diesem Zwecke bemerken wir erstens, daß für die Lösung $u = u_h$ unseres Differenzenproblems die Summe

$$h^2 \sum_{G_h} \sum (u_{xx}^2 + 2u_{xy}^2 + u_{yy}^2)$$

bei abnehmender Maschenweite beschränkt bleibt. Wegen der Minimumeigenschaft der Lösung unseres Differenzenproblems (vgl. S. 39) ist nämlich diese Summe nicht größer als die entsprechende Summe

$$h^2 \sum_{G_h} \sum (f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2)$$

und diese konvergiert bei Verfeinerung der Maschenweite gegen das Integral

$$\iint_G \left(\frac{\partial^2 f}{\partial x^2} + 2 \frac{\partial^2 f}{\partial x \partial y} + \frac{\partial^2 f}{\partial y^2} \right) dx dy,$$

welches nach unseren Voraussetzungen existiert.

Aus der Beschränktheit der Summe

$$h^2 \sum_{G_h} \sum (u_{xx}^2 + 2u_{xy}^2 + u_{yy}^2)$$

folgt unmittelbar die Beschränktheit von $h^2 \sum_{G_h} \sum (\Delta u)^2$, weiterhin auch die von

$$h^2 \sum_{G_h} \sum (u_x^2 + u_y^2) \quad \text{und} \quad h^2 \sum_{G_h} \sum u^2.$$

Es besteht nämlich für beliebige w die Ungleichung

$$(15) \quad h^2 \sum_{G_h} \sum w^2 \leq c h^2 \sum_{G_h} \sum (w_x^2 + w_y^2) + c h \sum_{\Gamma_h} w^2$$

(vgl. (15), S. 54). Indem man in dieser Ungleichung die Funktion w durch die ersten Differenzenquotienten von w ersetzt und auf diejenigen Teilgebiete von G_h anwendet, für welche diese Differenzenquotienten de-

finit sind, ergibt sich die weitere Ungleichung

$$h^2 \sum_{G_h} \sum (w_x^2 + w_y^2) \leq ch^2 \sum_{G_h} \sum (w_{xx}^2 + 2w_{xy}^2 + w_{yy}^2) + ch \sum_{\Gamma_h + \Gamma'_h} (w_x^2 + w_y^2),$$

wo die Konstanten c wieder nicht von der Funktion und der Maschenweite abhängen. Wir wenden diese Ungleichungen auf $w = u_h$ an und beachten dabei die Beschränktheit der Summen über $\Gamma_h + \Gamma'_h$ auf der rechten Seite — diese Randsummen konvergieren ja definitionsgemäß gegen die entsprechenden mit $f(x, y)$ gebildeten Integrale —. Somit folgt aus der Beschränktheit von

$$h^2 \sum_{G_h} \sum (u_{xx}^2 + 2u_{xy}^2 + u_{yy}^2)$$

die Beschränktheit von

$$h^2 \sum_{G_h} \sum (u_x^2 + u_y^2) \quad \text{und} \quad h^2 \sum_{G_h} \sum u^2.$$

Drittens setzen wir in der Ungleichung

$$(10) \quad h^2 \sum_{G^{**}} \sum (w_x^2 + w_y^2) \leq ch^2 \sum_{G^*} \sum w^2 + ch^2 \sum_{G^*} \sum (\Delta w)^2$$

(vgl. S. 51), wo G^* ein G^{**} im Innern enthaltendes Teilgebiet von G ist, für w nacheinander die Ausdrücke Δu , Δu_x , Δu_y , Δu_{xx} , ... ein, die ja alle der Gleichung $\Delta w = 0$ genügen. Es folgt dann sukzessive, daß für alle inneren Teilgebiete G^* von G die Summen

$$h^2 \sum_{G^*} \sum (w_x^2 + w_y^2),$$

d. h.

$$h^2 \sum_{G^*} \sum (\Delta u_x^2 + \Delta u_y^2), \quad h^2 \sum_{G^*} \sum (\Delta u_{xx}^2 + \Delta u_{xy}^2), \dots$$

zugleich mit den schon als beschränkt bekannten Summen

$$h^2 \sum_{G_h} \sum u^2, \quad h^2 \sum_{G_h} \sum (u_x^2 + u_y^2)$$

und

$$h^2 \sum_{G_h} \sum (\Delta u)^2$$

beschränkt bleiben.

Schließlich setzen wir für w in die Ungleichung (10) der Reihe nach die Funktionen u_{xx} , u_{xy} , u_{yy} , u_{xxx} , ... ein, für die nach dem eben Bewiesenen

$$h^2 \sum_{G_h} \sum (\Delta w)^2, \quad \text{d. h.} \quad h^2 \sum_{G_h} \sum (\Delta u_{xx})^2, \dots$$

beschränkt bleibt. Wir erkennen dann, daß für alle Teilgebiete auch die Summen

$$h^2 \sum_{G_h} \sum (u_{xxx}^2 + u_{xxy}^2), \quad h^2 \sum_{G_h} \sum (u_{xyx}^2 + u_{xyy}^2), \dots$$

beschränkt bleiben.

Aus dieser Tatsache können wir nunmehr wie auf S. 51 ff. schließen, daß sich aus unserer Folge von Gitterfunktionen eine Teilfolge auswählen läßt, die in jedem inneren Teilgebiet von G mit sämtlichen Differenzenquotienten gleichmäßig gegen eine im Innern von G stetige Grenzfunktion bzw. gegen deren Differentialquotienten konvergiert.

Wir haben noch zu zeigen, daß diese Grenzfunktion, die offenbar der Differentialgleichung $\Delta \Delta u = 0$ genügt, auch noch die vorgeschriebenen Randbedingungen erfüllt. Dabei begnügen wir uns analog wie oben damit, diese Randbedingungen in der Form

$$\iint_{S_r} (u - f)^2 dx dy \leq cr^2, \quad \iint_{S_r} \left[\left(\frac{\partial u}{\partial x} - \frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial u}{\partial y} - \frac{\partial f}{\partial y} \right)^2 \right] dx dy \leq cr^2$$

auszusprechen²³⁾. Daß die Grenzfunktion diese Bedingungen erfüllt, ergibt sich aber, indem wir das Schlußverfahren von S. 53 wörtlich auf die Funktion u und ihre ersten Differenzenquotienten anwenden.

Wegen der eindeutigen Bestimmtheit der Lösung unserer Randwertaufgabe erkennt man jetzt nachträglich, daß nicht nur eine ausgewählte Teilfolge, sondern die Funktionenfolge u selbst die angegebenen Konvergenzeigenschaften besitzt.

II. Der hyperbolische Fall.

§ 1.

Die Gleichung der schwingenden Saite.

Im zweiten Teil dieser Arbeit beschäftigen wir uns mit Anfangswertproblemen von hyperbolischen linearen Differentialgleichungen und werden beweisen, daß unter gewissen Voraussetzungen die Lösungen entsprechender Differenzgleichungen bei Verfeinerung der Maschenweite des zugrunde gelegten Gitters gegen die Lösung der Differentialgleichung konvergieren.

Wir können die hier auftretenden Verhältnisse am einfachsten an dem naheliegenden Beispiel der Schwingungsgleichung

$$(1) \quad \frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} = 0$$

darlegen. Dabei beschränken wir uns auf dasjenige Anfangswertproblem, in dem auf der Geraden $t=0$ die Werte der Lösung u und ihrer Ableitungen gegeben sind.

²³⁾ Daß die Randwerte für Funktion und Ableitungen tatsächlich angenommen werden, läßt sich unschwer zeigen. Vgl. die entsprechenden Betrachtungen bei K. Friedrichs loc. cit.

Um die entsprechende Differenzgleichung anzugeben, legen wir in der x, t -Ebene ein quadratisches achsenparalleles Gitter der Maschenweite h . Wir ersetzen die Differentialgleichung (1) durch die Differenzgleichung

$$u_{i\bar{i}} - u_{x\bar{x}} = 0$$

in den Bezeichnungen von S. 34. Greifen wir einen Gitterpunkt P_0 heraus, so verbindet die zugehörige Differenzgleichung den Wert der Funktion u in diesem Punkte mit den Werten in den vier Nachbarpunkten. Kennzeichnen wir wieder die vier Nachbarwerte durch die vier Indizes 1, 2, 3, 4 (s. Fig. 5), so nimmt die Differenzgleichung die einfache Gestalt

$$(2) \quad u_1 + u_3 - u_2 - u_4 = 0$$

an. Hierbei geht der Wert der Funktion u im Punkte P selbst nicht in die Gleichung ein.

Wir denken uns das Gitter in zwei verschiedene Teilgitter zerlegt, wie in der Fig. 5 durch Kreise und Kreuze angedeutet ist. Die Differenzen-

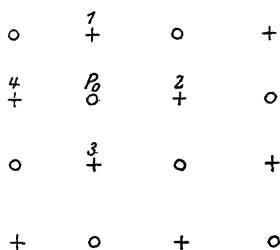


Fig. 5.

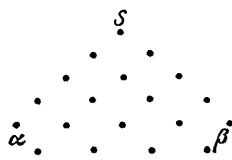


Fig. 6.

gleichung verbindet dann nur die Werte der Funktion in jedem der Teilgitter untereinander. Wir wollen uns daher auf eines der beiden Teilgitter beschränken. Als Anfangsbedingungen haben wir hier die Werte der Funktion u auf den beiden Gitterreihen $t = 0$ und $t = h$ vorzugeben. Wir geben zunächst die Lösung dieses Anfangswertproblems explizite an; d. h. wir drücken den Wert der Lösung in irgendeinem Punkte S durch die vorgegebenen Werte auf den beiden Anfangsreihen aus. Man erkennt sofort, daß der Wert in einem Punkte der Reihe $t = 2h$ eindeutig bestimmt ist lediglich durch die mit ihm verknüpften drei Werte auf den beiden ersten Reihen. Der Wert in einem Punkte auf der vierten Reihe ist eindeutig bestimmt durch die Werte der Lösung in gewissen drei Punkten der zweiten und dritten Reihe, also auch durch gewisse Werte auf den beiden ersten Reihen. Allgemein wird zu einem Punkte S ein gewisser Abhängigkeitsbereich auf den beiden ersten Reihen gehören; man erhält ihn, wenn man durch den Punkt S die Linien $x + t = \text{konst.}$ und $x - t = \text{konst.}$ zieht, bis sie die zweite Reihe in den Punkten α und β treffen (vgl. Fig. 6).

Das Dreieck $S\alpha\beta$ nennen wir dann das Bestimmtheitsdreieck, weil in ihm sämtliche u -Werte sich nicht ändern, sobald sie auf den ersten beiden Reihen festgehalten werden. Die Seitenlinien des Dreiecks nennen wir Bestimmtheitslinien.

Bezeichnet man nun die Differenzen von u in Richtung der Bestimmtheitslinien durch u' und u'' oder genauer

$$\begin{aligned} u'_1 &= u_1 - u_4, & u''_1 &= u_1 - u_2, \\ u'_2 &= u_2 - u_3, & u''_4 &= u_4 - u_3, \end{aligned}$$

so nimmt die Differenzgleichung etwa die Form

$$u'_1 = u'_2$$

an. D. h. auf einer Bestimmtheitslinie sind die Differenzen nach der anderen Bestimmtheitsrichtung konstant, also gleich einer der *vorgegebenen* Differenzen zwischen zwei Punkten der ersten beiden Reihen. Andererseits ist die Differenz $u_S - u_\alpha$ eine Summe über die Differenzen u' längs der Bestimmtheitslinie $S\alpha$, so daß wir unter Benutzung der eben gemachten Bemerkung als Schlußformel (in leicht verständlicher Bezeichnung)

$$(3) \quad u_S = u_\alpha + \sum_{\alpha_1}^{\beta_1} u'$$

erhalten.

Wir lassen nun die Maschenweite h gegen Null streben, wobei die vorgegebenen Werte auf der zweiten oder ersten Reihe gegen eine zweimal stetig differenzierbare Funktion $f(x)$ und die Differenzenquotienten $\frac{u'}{h\sqrt{2}}$ gegen eine stetig differenzierbare Funktion $g(x)$ gleichmäßig konvergieren mögen. Offenbar geht dabei die rechte Seite von (3) gleichmäßig in den Ausdruck

$$(4) \quad f(x-t) + \frac{1}{\sqrt{2}} \int_{x-t}^{x+t} g(\xi) d\xi$$

über, wenn S gegen den Punkt (t, x) konvergiert. Dies ist der bekannte Ausdruck der Lösung der Schwingungsgleichung (1) mit den Anfangswerten $u(x, 0) = f(x)$ und $\frac{\partial u}{\partial t}(x, 0) = \sqrt{2}g(x)$. Damit ist gezeigt, daß die Lösungen unseres Differenzgleichungsproblems bei abnehmender Maschenweite gegen die Lösung des Differentialgleichungsproblems konvergieren, wenn wir die Anfangswerte (in der oben angegebenen Weise) konvergieren lassen.

§ 2.

Über den Einfluß der Wahl des Gitters.

Die Abhängigkeitsgebiete bei Differenzen- und Differentialgleichung.

Die im § 1 betrachteten Verhältnisse legen folgende Überlegungen nahe.

Ebenso wie für die Lösung einer linearen hyperbolischen Differentialgleichung im Punkte S nur ein gewisser Teil der Anfangswerte maßgebend ist, nämlich das von den Charakteristiken durch S ausgeschnittene „Abhängigkeitsgebiet“, besitzt auch die Lösung einer Differenzgleichung im Punkte S ein gewisses Abhängigkeitsgebiet, das man erhält, indem man die Bestimmtheitslinien vom Punkte S aus zieht. In § 1 fielen nun die Richtungen der Bestimmtheitslinien der Differenzgleichung mit den charakteristischen Richtungen der Differentialgleichung zusammen, wodurch auch die Abhängigkeitsgebiete in der Grenze übereinstimmten. Diese Tatsache hing aber wesentlich von der Orientierung des Gitters in der (x, t) -Ebene ab und beruhte ferner darauf, daß wir das Gitter quadratisch gewählt hatten. Wir legen jetzt allgemeiner ein rechteckiges achsenparalleles Gitter zugrunde, dessen Maschenweite in der t -Richtung (Zeitmasche) gleich h und diejenige der x -Richtung (Raummasche) gleich κh mit konstantem κ ist. Das Abhängigkeitsgebiet der Differenzgleichung $u_{ti} - u_{x\bar{x}} = 0$ für dieses Gitter wird ganz im Innern des Abhängigkeitsgebietes der Differentialgleichung $\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} = 0$ liegen oder wird es selbst in seinem Innern enthalten, je nachdem ob $\kappa < 1$ oder $\kappa > 1$ ist.

Hieraus ergibt sich eine merkwürdige Tatsache: Läßt man im Falle $\kappa < 1$ die Maschenweite h gegen Null abnehmen, so kann die Lösung der Differenzgleichung im allgemeinen *nicht* gegen die Lösung der Differentialgleichung konvergieren. Ändert man nämlich etwa bei der Schwingungsgleichung (1) die Anfangswerte der Lösung der Differentialgleichung in der Umgebung der Endpunkte α und β des Abhängigkeitsgebietes (vgl. Fig. 7), so zeigt die Formel (4), daß sich auch die Lösung selbst im Punkte (x, t) ändert. Für die Lösungen der Differenzgleichungen im Punkte S sind aber die Vorgaben in den Punkten α und β irrelevant, da diese außerhalb des Abhängigkeitsgebietes der Differenzgleichungen liegen. — Daß im Falle $\kappa > 1$ Konvergenz statthat, werden wir im § 3 beweisen. Vgl. hierzu Fig. 9, S. 62.

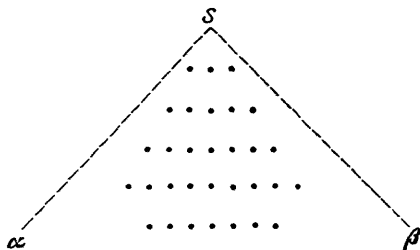


Fig. 7.

Betrachtet man dagegen z. B. die Differentialgleichung

$$(5) \quad 2 \frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2} = 0$$

in den beiden räumlichen Variablen x, y und der zeitlichen t , und ersetzt sie durch entsprechende Differenzgleichungen in geradlinigen Gittern, so ist es im Gegensatz zum Falle von nur zwei unabhängigen Variablen unmöglich, die Mascheneinteilung so zu wählen, daß die Abhängigkeitsgebiete der Differenzen- und Differentialgleichung zusammenfallen; denn das Abhängigkeitsgebiet der Differenzgleichungen wird ein Viereck, während das der Differentialgleichung ein Kreis ist. Wir werden später (vgl. § 4) die Mascheneinteilung so wählen, daß das Bestimmtheitsgebiet der Differenzgleichung das Bestimmtheitsgebiet der Differentialgleichung im Innern enthält, und zeigen, daß wieder Konvergenz stattfindet.

Überhaupt wird ein wesentliches Ergebnis dieses Teils sein, daß man bei jeder linearen hyperbolischen homogenen Differentialgleichung zweiter Ordnung das Gitter so wählen kann, daß die Lösung der Differenzgleichung gegen die Lösung der Differentialgleichung konvergiert, wenn man die Maschenweiten gegen Null streben läßt (vgl. hierzu §§ 3, 4, 7, 8).

§ 3.

Grenzübergang bei beliebigen rechteckigen Gittern.

Wir betrachten zunächst wieder die Schwingungsgleichung

$$(1) \quad \frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} = 0,$$

legen aber nunmehr ein rechteckiges achsenparalleles Gitter zugrunde, dessen zeitliche Maschenweite h , dessen Raummasche κh ist. Die zugehörige Differenzgleichung lautet:

$$(6) \quad L(u) = \frac{1}{h^2}(u_1 - 2u_0 + u_3) - \frac{1}{\kappa^2 h^2}(u_2 - 2u_0 + u_4) = 0,$$

wobei sich die Indizes auf den Mittelpunkt P_0 und die Ecken P_1, P_2, P_3, P_4 eines „Elementarrhombus“ (vgl. Fig. 8) beziehen. Vermöge der Gleichung

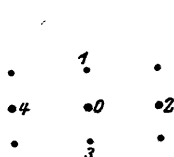


Fig. 8.

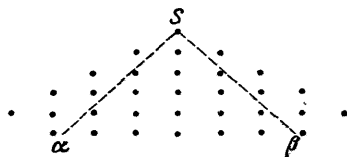


Fig. 9.

$L(u) = 0$ können wir den Wert der Funktion u in einem Punkte S durch ihre Werte auf demjenigen Stück der beiden Anfangsreihen $t = 0$ und $t = h$ darstellen, das man erhält, wenn man vom Punkte S aus (vgl. Fig. 6, S. 59)

die zu den Seiten eines Elementarrhombus parallelen „Bestimmtheitslinien“ zieht. Wir denken uns die Anfangswerte so vorgegeben, daß sie und die zwischen ihnen gebildeten ersten Differenzenquotienten bei abnehmender Maschenweite und bei festem κ gleichmäßig gegen stetige vorgegebene Funktionen auf der Geraden $t=0$ konvergieren. Für die Lösungen der Differenzengleichungen läßt sich wohl eine explizite Darstellung durch ihre Anfangswerte aufstellen (entsprechend (3) in § 1); sie ist aber nicht so einfach, daß man unmittelbar den Grenzübergang zu verschwindender Maschenweite ausführen kann. Wir schlagen daher einen anderen Weg ein, der uns die Behandlung auch des allgemeinen Problem ermöglichen wird²⁴⁾.

Wir multiplizieren den Differenzenausdruck $L(u)$ mit $(u_1 - u_3)$ und formen das Produkt unter Beachtung der folgenden Identitäten um:

$$(7) \quad (u_1 - u_3)(u_1 - 2u_0 + u_3) = (u_1 - u_0)^2 - (u_0 - u_3)^2,$$

$$(8) \quad (u_1 - u_3)(u_2 - 2u_0 + u_4) = (u_1 - u_0)^2 - (u_0 - u_3)^2 \\ - \frac{1}{2} [(u_1 - u_2)^2 + (u_1 - u_4)^2 - (u_2 - u_3)^2 - (u_4 - u_3)^2].$$

Wir erhalten so

$$(9) \quad 2(u_1 - u_3)L(u) = \frac{2}{h^2} \left(1 - \frac{1}{\kappa^2}\right) [(u_1 - u_0)^2 - (u_0 - u_3)^2] \\ + \frac{1}{h^2 \kappa^2} [(u_1 - u_2)^2 + (u_1 - u_4)^2 - (u_2 - u_3)^2 - (u_4 - u_3)^2].$$

Wir summieren nun das Produkt (9) über alle Elementarrhomben eines Bestimmtheitsdreiecks $S\alpha\beta$. Die auf der rechten Seite von (9) auftretenden Differenzenquadrate kommen stets in zwei benachbarten Elementarrhomben vor, mit verschiedenen Vorzeichen versehen. Sie heben sich bei der Summation fort, sobald beide Elementarrhomben zum Dreieck $S\alpha\beta$ gehören; es bleibt also nur eine vom Rande des Dreiecks berührende Summe von Differenzenquadraten übrig. Wir erhalten so die Relation:

$$(10) \quad h^2 \sum_{S\alpha\beta} \sum 2 \frac{u_1 - u_3}{h} L(u) \\ = h \sum_{S\alpha} \left[2 \left(1 - \frac{1}{\kappa^2}\right) \left(\frac{\dot{u}}{h}\right)^2 + \frac{1}{\kappa^2} \left(\frac{u'}{h}\right)^2 \right] \\ + h \sum_{S\beta} \left[2 \left(1 - \frac{1}{\kappa^2}\right) \left(\frac{\dot{u}}{h}\right)^2 + \frac{1}{\kappa^2} \left(\frac{u'}{h}\right)^2 \right] \\ - h \sum_{I\text{II}} \left[2 \left(1 - \frac{1}{\kappa^2}\right) \left(\frac{\dot{u}}{h}\right)^2 + \frac{1}{\kappa^2} \left(\frac{u'}{h}\right)^2 + \frac{1}{\kappa^2} \left(\frac{u'}{h}\right)^2 \right].$$

²⁴⁾ Zum folgenden vgl. K. Friedrichs und H. Lewy, Über die Eindeutigkeit usw., Math. Annalen (98 1928), S. 192 ff., wo analoge Umformungen für Integrale benutzt werden.

Hier bedeuten u' und u'' Differenzen in Bestimmtheitsrichtungen wie in § 1, während \dot{u} die Differenz der Funktionswerte in zwei Nachbarpunkten bezeichnet, deren Verbindungslinie zur t -Achse parallel ist. Die Summen sind über alle aus zwei Parallelreihen bestehenden Randstreifen zu erstrecken, so daß sämtliche vorkommenden Differenzen u' , u'' , \dot{u} einmal und nur einmal auftreten.

Für eine Lösung von $L(u) = 0$ verschwindet also die rechte Seite von (10). Die dort auftretende Summe über die Anfangsreihen I und II bleibt beschränkt, wenn wir die Maschenweite h (bei festgehaltenem \varkappa) zu Null abnehmen lassen; sie geht nämlich in ein Integral über vorgegebene Funktionen auf der Anfangslinie über. Infolgedessen bleiben auch die in (10) über S_α und S_β erstreckten Summen beschränkt. Ist nun, wie wir fordern müssen (vgl. S. 61), $\varkappa \geq 1$, also $1 - \frac{1}{\varkappa^2}$ nicht negativ, so folgt weiter die Beschränktheit der einzelnen Summen

$$h \sum_{S_\alpha} \left(\frac{u'}{h}\right)^2, \quad h \sum_{S_\beta} \left(\frac{u''}{h}\right)^2,$$

die wir über irgendwelche Bestimmtheitslinien erstreckt denken können.

Hieraus können wir die „gleichartige Stetigkeit“ (vgl. 1. Teil § 4) der Folge der Gitterfunktionen in allen Richtungen der Ebene ableiten²⁵⁾; da die Werte von u auf der Anfangslinie beschränkt sind, folgt die Existenz einer gleichmäßig gegen eine Grenzfunktion $u(x, t)$ konvergierenden Teilfolge.

Zugleich mit der Funktion u genügen auch ihre ersten und zweiten Differenzenquotienten der Differenzgleichung $L(u) = 0$. Die Anfangswerte dieser Differenzenquotienten lassen sich vermittelt der Gleichung $L(u) = 0$ durch solche erste, zweite und dritte Differenzenquotienten von u ausdrücken, in denen nur Punkte der beiden Anfangsreihen I und II auftreten. Wir verlangen von ihnen, daß sie gegen stetige Grenzfunktionen streben, d. h. daß etwa die vorgegebenen Anfangswerte $u(x, 0)$, $u_t(x, 0)$ drei- bzw. zweimal stetig nach x differenzierbar sind.

Danach können wir unsere oben angestellte Konvergenzbetrachtung anstatt auf u auch auf seine ersten und zweiten Differenzenquotienten anwenden, also eine Teilfolge auswählen, so daß diese Differenzenquotienten gleichmäßig gegen Funktionen streben, die dann die ersten bzw. zweiten Ableitungen der Grenzfunktion $u(x, t)$ sein müssen. Die Grenzfunktion u

²⁵⁾ Sind S_1 und S_2 zwei Punkte im Abstand δ , so verbinde man sie durch einen Streckenzug aus zwei Strecken $S_1 S$ und $S S_2$, von denen die erste der einen, die zweite der anderen Bestimmtheitsrichtung parallel ist. Es gilt dann die Abschätzung

$$|u_{S_1} - u_{S_2}| \leq |u_{S_1} - u_S| + |u_S - u_{S_2}| \leq \sqrt{\delta} \sqrt{h \sum_{S_1 S} \left(\frac{u'}{h}\right)^2} + \sqrt{\delta} \sqrt{h \sum_{S S_2} \left(\frac{u''}{h}\right)^2}.$$

genügt infolgedessen der Differentialgleichung $\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} = 0$, die in der Grenze aus der Differenzgleichung $L(u) = 0$ entsteht; sie stellt also die Lösung des Anfangswertproblems dar. Da diese Lösung eindeutig bestimmt ist, konvergiert jede Teilfolge der Gitterfunktionen und damit die Folge selbst gegen die Grenzfunktion.

§ 4.

Die Schwingungsgleichung in drei Variablen.

Wir behandeln nun die Schwingungsgleichung

$$(11) \quad 2 \frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2} = 0$$

und knüpfen an die im § 2 gemachten Bemerkungen über die Beziehung der Abhängigkeitsbereiche an. Das Abhängigkeitsgebiet der Differentialgleichung (11) ist der Kreiskegel mit einer zur t -Richtung parallelen Achse und dem Öffnungswinkel α , mit $\operatorname{tg} \alpha = \frac{1}{\sqrt{2}}$. In irgendeinem rechtwinkligen achsenparallelen Gitter setzen wir entsprechend die Differenzgleichung

$$(12) \quad 2 u_{ii} - u_{xx} - u_{yy} = 0$$

an. Durch sie werden die Funktionswerte u in den Punkten eines „Elementaroktaeders“ miteinander verknüpft. Sie gestattet, den Funktionswert in einem Punkte S eindeutig durch die Funktionswerte in gewissen Punkten der beiden Anfangsebenen $t = 0$ und $t = h$ auszudrücken. Wir erhalten zu jedem Punkte S eine Pyramide der Bestimmtheit, die aus den beiden Grundlinien als Abhängigkeitsgebiet zwei Rhomben ausschneidet.

Lassen wir die Maschenweiten etwa unter Festhaltung ihrer Verhältnisse gegen Null streben, so können wir eine Konvergenz der Folge der Gitterfunktionen gegen die Lösung der Differentialgleichung höchstens dann erwarten, wenn die Bestimmtheitspyramide den Bestimmtheitskegel der Differentialgleichung im Inneren enthält. Das einfachste Gitter dieser Eigenschaft wird dasjenige sein, das so liegt, daß die Bestimmungs- pyramide den Bestimmungskegel von außen berührt. Unsere Differentialgleichung ist gerade so gewählt, daß dies für ein *kubisches* achsenparalleles Gitter eintritt.

Die Differenzgleichung (12) nimmt in diesem Gitter in den Bezeichnungen der Figur 10 die Gestalt an:

$$(13) \quad L(u) = \frac{2}{h^2}(u_{\alpha'} - 2u_0 + u_{\alpha}) - \frac{1}{h^2}(u_1 - 2u_0 - u_3) - \frac{1}{h^2}(u_2 - 2u_0 - u_4),$$

in die der Funktionswert u_0 im Mittelpunkt P übrigens nicht mehr eingeht. Die Werte der Lösung auf den beiden Anfangsebenen seien die Werte einer viermal stetig nach x_1, y_2, t differenzierbaren Funktion.

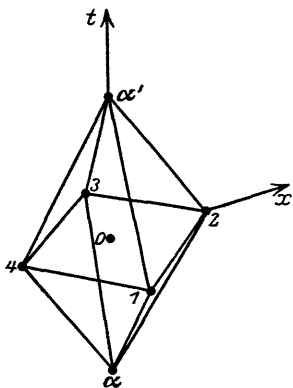


Fig. 10.

Wir benutzen zum Konvergenzbeweise wieder die im § 3 entwickelte Methode, indem wir für die Lösung unserer Differenzgleichung die dreifache Summe

$$h^3 \sum \sum \sum 2 \frac{u_{\alpha'} - u_{\alpha}}{h} L(u) = 0$$

bilden, die über alle Elementaroktaeder der vom Punkte S ausstrahlenden Bestimmtheitspyramide zu erstrecken ist. Auf Grund der fast wörtlich zu übernehmenden Schlußweise erkennen wir, daß die Werte der Funktion u in inneren Punkten der Bestimmtheitspyramide herausfallen und daß

nur noch Flächensummen über die vier Seitendoppelflächen F und die beiden Grundflächen I II der Pyramide übrigbleiben.

Bezeichnen wir mit u' die Differenz der Funktionswerte in zwei Punkten, die durch eine Seitenlinie eines Elementaroktaeders verbunden wird, so lautet die entstehende Formel

$$(14) \quad \sum_F \sum (u')^2 - \sum_I \sum_{II} (u')^2 = 0,$$

in der über sämtliche auf diesen Flächen enthaltenen Differenzen u' zu summieren ist, so daß jede solche Differenz nur einmal auftritt²⁶⁾. Da die Doppelsumme über die beiden Anfangsflächen beschränkt bleibt, weil sie ja in ein Integral über Anfangswerte übergeht, so bleibt auch die Summe über die „Bestimmtheitsflächen“ F beschränkt.

Wir wenden unsere Betrachtung anstatt auf u selbst wieder auf ihre ersten, zweiten und dritten Differenzenquotienten, die ja selbst der Differenzgleichung (13) genügen und deren Anfangswerte mittels (13) sich durch erste bis vierte Differenzenquotienten allein aus Werten auf den ersten beiden Anfangsebenen ausdrücken lassen. Ist $w = w_h$ einer der Differenzenquotienten bis zur dritten Ordnung, so wissen wir, daß die über eine Bestimmtheitsfläche F erstreckte Summe $h^2 \sum \sum \left(\frac{w'}{h}\right)^2$ beschränkt bleibt.

Hieraus ergibt sich aber durch genau denselben Schluß, den wir im ersten Teil, § 4 angewandt haben, daß die Funktion u mit ihren ersten und

²⁶⁾ Es ist das Gitter gerade so gewählt, daß die Differenzen von u zwischen den beiden Flächen F nicht mehr auftreten.

zweiten Differenzenquotienten gleichartig stetig ist. Es gibt also eine gegen Null abnehmende Folge von Maschenweiten, so daß diese Größen, die ja am Anfang beschränkt sind, gegen stetige Grenzfunktionen konvergieren, und zwar offenbar gegen die Lösung der Differentialgleichung einschließlich ihrer ersten und zweiten Ableitungen, was genau wie früher (§ 3) folgt.

Anhang.

Ergänzungen und Verallgemeinerungen.

§ 5.

Beispiel einer Differentialgleichung erster Ordnung.

Wir haben im § 2 gesehen, daß unter Umständen das Abhängigkeitsgebiet der Differentialgleichung nur einen Teil des Abhängigkeitsgebietes der Differenzengleichung ausmacht, und daß also der Einfluß des Restgebietes in der Grenze herausfällt. Dies Phänomen können wir an dem Beispiel der Differentialgleichung erster Ordnung $\frac{\partial u}{\partial t} = 0$ explizite verfolgen, wenn wir sie durch die Differenzengleichung

$$(15) \quad 2u_t - u_x + u_{\bar{x}} = 0$$

ersetzen. Sie lautet ausgeschrieben in den Bezeichnungen der Fig. 5 (S. 59)

$$(16) \quad u_1 = \frac{u_3 + u_4}{2}.$$

Die Differenzengleichung verbindet wieder nur Punkte eines Teilgitters untereinander. Das Anfangswertproblem besteht darin, daß man auf der Reihe $t=0$ der Funktion u in den Punkten $x=2ih$ diejenigen Werte f_{2i} vorschreibt, die dort eine stetige Funktion $f(x)$ annimmt.

Wir betrachten etwa den Punkt S auf der t -Achse im Abstände $2nh$. Man verifiziert leicht die Darstellung der Lösung u in S :

$$(17) \quad u_S = \sum_{i=-n}^n \frac{1}{2^{2n}} \binom{2n}{n+i} f_{2i}.$$

Die Summe auf der rechten Seite strebt bei Verfeinerung der Maschenweite, d. h. bei $n \rightarrow \infty$ einfach gegen den Wert f_0 . Man entnimmt das aus der Stetigkeit von $f(x)$ und dem Verhalten der Binomialkoeffizienten bei wachsendem n . (Vgl. den nächsten Paragraphen.)

§ 6.

Die Wärmeleitungsgleichung.

Die Differenzengleichung (16) des § 5 läßt sich auch als Analogon einer ganz anderen Differentialgleichung auffassen, nämlich der Wärme-

Leitungsgleichung

$$(18) \quad 2 \frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = 0.$$

In irgendeinem rechteckigen achsenparallelen Gitter lautet die entsprechende Differenzgleichung

$$(19) \quad 2 \left(\frac{u_1 - u_0}{l} \right) = \left(\frac{u_2 + u_4 - 2u_0}{h^2} \right).$$

wo l die Zeit, h die Raummasche ist. Beim Grenzübergang zu verschwindender Maschenweite behält die Differenzgleichung nur dann ihre Form, wenn l proportional mit h^2 abnimmt. Setzen wir insbesondere $l = h^2$, so fällt der Wert u_0 aus der Gleichung heraus und es entsteht die Differenzgleichung

$$(16) \quad u_1 = \frac{u_2 + u_4}{2},$$

deren Auflösung durch die Formel (17) gegeben wird:

$$(17) \quad u(0, t) = \sum_{i=-n}^n \frac{1}{2^{2n}} \binom{2n}{n+i} f_{2i}.$$

Ein Punkt ξ der x -Achse ist bei abnehmender Maschenweite immer durch den Index

$$(20) \quad 2i = \frac{\xi}{h}$$

gekennzeichnet. Die Maschenweite h ist mit der Ordinate t des Aufpunktes S durch die Gleichung

$$(21) \quad 2n h^2 = t$$

festgelegt.

Wir wollen untersuchen, was aus der Formel (17) entsteht, wenn h gegen Null, d. h. n gegen unendlich strebt. Unter Benutzung der Formel (21) schreiben wir die Gleichung (17) in die Gestalt

$$(22) \quad u(0, t) = \sum_{i=-n}^n \frac{\sqrt{2n}}{2 \cdot 2^{2n} \sqrt{t}} \binom{2n}{n+i} f_{2i} \cdot 2h.$$

Für den Koeffizienten von $2h f_{2i} = 2h f(\xi)$ verwenden wir die Abkürzung

$$\frac{1}{2\sqrt{t}} g_{2n}(\xi) = \frac{\sqrt{2n}}{2 \cdot 2^{2n} \sqrt{t}} \binom{2n}{n + \frac{\xi}{\sqrt{2t}} \sqrt{n}}.$$

Den Grenzwert dieses Koeffizienten, den man gewöhnlich mit Hilfe der Stirlingschen Formel bestimmt, wollen wir hier berechnen, indem wir die

Funktion $g_{2n}(\xi)$ als Lösung einer gewöhnlichen Differenzgleichung auffassen und den Grenzübergang zu verschwindender Maschenweite h und damit zur Differentialgleichung ausführen. Als diese Differenzgleichung findet man

$$\frac{1}{2h}(g_n(\xi + 2h) - g_n(\xi)) = -\frac{1}{2h}g_n(\xi)\frac{2i+1}{n+i+1}$$

(indem wir $g_n(\xi)$ anstatt $g_{2n}(\xi)$ schreiben). Oder

$$\frac{1}{2h}(g_h(\xi + 2h) - g_h(\xi)) = -g_h(\xi)\frac{\xi+h}{t+h\xi+2h^2}.$$

$g_h(\xi)$ genügt außerdem der Normierungsbedingung

$$\sum_{i=-n}^n g_h(\xi) \cdot 2h = 2\sqrt{t}.$$

Diese Summe ist über das Abhängigkeitsgebiet der Differenzgleichung zu erstrecken, das, wenn $h \rightarrow 0$ strebt, in der Grenze die ganze x -Achse erfüllt.

Durch einfache Überlegungen erkennt man, daß $g_h(\xi)$ gleichmäßig gegen die Lösung $g(x)$ der Differentialgleichung

$$g'(x) = -g(x)\frac{x}{t}$$

mit der Nebenbedingung

$$\int_{-\infty}^{+\infty} g(x) dx = 2\sqrt{t}$$

konvergiert. Aus der Formel (23) entsteht dann durch Grenzübergang

$$u(0, t) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi t}} e^{-\frac{\xi^2}{2t}} f(\xi) d\xi,$$

die bekannte Lösung der Wärmeleitungsgleichung.

Die Betrachtungen dieses Paragraphen übertragen sich ohne weiteres auf den Fall von Differentialgleichungen

$$4 \frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2} = 0$$

usw. bei mehr unabhängigen Veränderlichen.

§ 7.

Die allgemeine lineare homogene Differentialgleichung zweiter Ordnung in der Ebene.

Wir behandeln die Differentialgleichung

$$(23) \quad \frac{\partial^2 u}{\partial t^2} - k^2 \frac{\partial^2 u}{\partial x^2} + \alpha \frac{\partial u}{\partial t} + \beta \frac{\partial u}{\partial x} + \gamma u = 0.$$

Die Koeffizienten sind zweimal stetig nach x , t differenzierbar, während die Anfangswerte auf der Geraden $t=0$ dreimalig stetig nach x differenzierbar sind. Wir ersetzen die Differentialgleichung in einem Gitter mit der Zeitmaschenweite h und der Raummaschenweite κh , so daß in einer Umgebung des zu betrachtenden Stückes der Anfangsgeraden $1 - \frac{k^2}{\kappa^2} > \varepsilon > 0$ für unser konstantes κ gilt, durch die Differenzengleichung

$$(24) \quad L(u) = u_{i\bar{i}}(x, t) - k^2 u_{x\bar{x}}(x, t) + \alpha u_t + \beta u_x + \gamma u = 0$$

und wählen die Anfangswerte wie in § 3 (vgl. S. 63).

Zum Konvergenzbeweis formen wir wieder die Summe

$$h^2 \sum_{S\alpha\beta} \sum 2 \frac{u_1 - u_3}{h} L(u)$$

unter Verwendung der Identitäten (7), (8) um. Außer einer Summe (vgl. (10)) über den Rand des Dreiecks $S\alpha\beta$ (vgl. Fig. 6) tritt dann noch eine über das ganze Dreieck $S\alpha\beta$ erstreckte Summe auf, deren absoluter Betrag sich nach oben mit Hilfe der Schwarzischen Ungleichung durch

$$Ch^2 \sum_{S\alpha\beta} \sum \left[\left(\frac{u'}{h} \right)^2 + \left(\frac{u'}{h} \right)^2 + \left(\frac{\dot{u}}{h} \right)^2 + u^2 \right]$$

abschätzen läßt, wo die Konstante C nicht von der Funktion u , der Maschenweite h und in einer gewissen Umgebung der Anfangslinee auch nicht vom Punkte S abhängt.

Hier können wir weiter $h^2 \sum_{S\alpha\beta} \sum u^2$ nach oben durch

$$C_1 h^2 \sum_{S\alpha\beta} \sum \left(\frac{\dot{u}}{h} \right)^2 + C_2 h \sum_{\text{I II}} u^2$$

abschätzen²⁷⁾, wo für die Konstanten C_1 , C_2 dasselbe gilt, was für C gesagt wurde.

Wir erhalten so eine Ungleichung von der Form

$$(25) \quad \begin{aligned} & h \sum_{S\alpha} \left[2 \left(1 - \frac{k^2}{\kappa^2} \right) \left(\frac{\dot{u}}{h} \right)^2 + \frac{k^2}{\kappa^2} \left(\frac{u'}{h} \right)^2 \right] \\ & + h \sum_{S\beta} \left[2 \left(1 - \frac{k^2}{\kappa^2} \right) \left(\frac{\dot{u}}{h} \right)^2 + \frac{k^2}{\kappa^2} \left(\frac{u'}{h} \right)^2 \right] \\ & \leq C_3 h^2 \sum_{S\alpha\beta} \sum \left[\left(\frac{\dot{u}}{h} \right)^2 + \left(\frac{u'}{h} \right)^2 + \left(\frac{u'}{h} \right)^2 \right] + D, \end{aligned}$$

wo D eine für alle Punkte S und Maschenweiten h feste Schranke für die auftretenden Summen über die Anfangsgerade ist.

²⁷⁾ Vgl. zum Beweise die verwandte Ungleichung S. 64 unten.

Als Spitzen S unserer Dreiecke wählen wir nun von der Anfangsgeraden ausgehend der Reihe nach die Punkte $S_0, S_1, \dots, S_n = S$ auf einer Parallelen zur t -Achse. Durch Summation der zugehörigen Ungleichungen (25) können wir die Ungleichung

$$(26) \quad \begin{aligned} & h^2 \sum_{S S_0 \alpha} \sum \left[2 \left(1 - \frac{k^2}{\varkappa^2} \right) \left(\frac{\dot{u}}{h} \right)^2 + \frac{k^2}{\varkappa^2} \left(\frac{u'}{h} \right)^2 \right] \\ & + h^2 \sum_{S S_0 \beta} \sum \left[2 \left(1 - \frac{k^2}{\varkappa^2} \right) \left(\frac{\dot{u}}{h} \right)^2 + \frac{k^2}{\varkappa^2} \left(\frac{u'}{h} \right)^2 \right] \\ & \leq n h C_3 \sum_{S \alpha \beta} \sum \left[\left(\frac{\dot{u}}{h} \right)^2 + \left(\frac{u'}{h} \right)^2 + \left(\frac{u'}{h} \right)^2 \right] + n h D \end{aligned}$$

erhalten. Beachten wir nun, daß man eine Differenz u' bzw. u' durch zwei Differenzen \dot{u} und eine Differenz u' bzw. u' ausdrücken kann, so ergibt sich, daß wir die linke Seite von (26) höchstens verkleinern, wenn wir sie durch

$$C_4 h^2 \sum_{S \alpha \beta} \sum \left[\left(\frac{\dot{u}}{h} \right)^2 + \left(\frac{u'}{h} \right)^2 + \left(\frac{u'}{h} \right)^2 \right]$$

ersetzen. Beschränken wir uns nun auf eine solche Umgebung $t \leq n h$ der Anfangsgeraden, in der

$$C_4 - n h C_3 = C_5 > 0$$

ist, so erhalten wir aus (26)

$$(27) \quad C_5 h^2 \sum_{S \alpha \beta} \sum \left[\left(\frac{\dot{u}}{h} \right)^2 + \left(\frac{u'}{h} \right)^2 + \left(\frac{u'}{h} \right)^2 \right] \leq \frac{C_4}{C_5} D.$$

Aus der in (27) ausgedrückten Beschränktheit der linken Seite ergibt sich nach (25) die Beschränktheit von

$$h \sum_{S \alpha} \left(\frac{u'}{h} \right)^2 + h \sum_{S \beta} \left(\frac{u'}{h} \right)^2,$$

woraus sich wie in § 3 die gleichartige Stetigkeit von u ergibt.

Wir wenden die Ungleichung (25) anstatt auf die Funktion u selbst auf deren erste und zweite Differenzenquotienten w an, die auch Differenzengleichungen genügen, deren Glieder zweiter Ordnung wie in (24) lauten. In den Zusatzgliedern können zwar noch Ableitungen von u auftreten, die sich nicht durch w ausdrücken lassen, aber deren mit h^2 multiplizierte Flächenquadratsumme schon als beschränkt angenommen werden kann. Das aber genügt, um auf diese Differenzengleichung für w denselben Schluß anzuwenden, den wir oben auf u angewandt haben. Wir können somit die gleichartige Stetigkeit und Beschränktheit der Funktionen u und ihrer ersten und zweiten Ableitungen folgern, die infolgedessen eine Teilfolge besitzen, die gleichmäßig gegen die Lösung des

Anfangswertproblems der Differentialgleichung konvergiert. Aus deren Eindeutigkeit folgt wieder, daß die Funktionenfolge selber konvergiert.

Wir müssen dabei allerdings voraussetzen, daß die Differenzenquotienten bis zur dritten Ordnung auf und zwischen den beiden Anfangsreihen gleichmäßig gegen stetige Grenzfunktionen konvergieren²⁸⁾.

§ 8.

Das Anfangswertproblem einer beliebigen hyperbolischen linearen Differentialgleichung zweiter Ordnung.

Wir wollen nun zeigen, daß die bisher entwickelten Methoden dazu ausreichen, das Anfangswertproblem einer beliebigen linearen homogenen hyperbolischen Differentialgleichung zweiter Ordnung zu lösen. Es genügt dabei, sich auf den Fall von drei Variablen zu beschränken. Der Gedankengang läßt sich unmittelbar auf mehr Variable übertragen. Man kann leicht einsehen, daß das allgemeinste derartige Problem durch eine Variablentransformation auf folgendes zurückgeführt werden kann: Eine Funktion $u(x, y, t)$ zu finden, die der Differentialgleichung

$$(28) \quad u_{tt} - (au_{xx} + 2bu_{xy} + cu_{yy}) + \alpha u_t + \beta u_x + \gamma u_y + \delta u = 0$$

genügt und die mit ihren ersten Ableitungen auf der Fläche $t = 0$ vorgegebene Werte annimmt. Dabei sollen die Koeffizienten Funktionen der Variablen x, y, t sein und den Bedingungen

$$a > 0, \quad c > 0, \quad ac - b^2 > 0$$

genügen.

Die Koeffizienten setzen wir dabei als dreimal nach x, y, t , und die Anfangswerte u als viermal bzw. u_t als dreimal nach x, y stetig differenzierbar voraus.

Wir denken uns ferner die Koordinaten x und y um einen Punkt der Anfangsebene so gedreht, daß in ihm $b = 0$ ist. Dann ist in einer gewissen Umgebung G dieses Punktes sicher die Bedingung

$$a - |b| > 0, \quad c - |b| > 0$$

erfüllt. Auf diese Umgebung beschränken wir unsere Betrachtungen. Wir können dann eine dreimal stetig differenzierbare Funktion $d > 0$ so wählen, daß

$$(29) \quad \left. \begin{array}{l} a - d \\ c - d \\ d - |b| \end{array} \right\} > \varepsilon > 0$$

²⁸⁾ Diese Voraussetzung und die über die Differenzierbarkeit der Koeffizienten der Differentialgleichung und ferner die Beschränkung auf eine genügend kleine Umgebung der Anfangsgeraden lassen sich in Sonderfällen mildern.

mit konstantem ε gilt. Dann setzen wir die Differentialgleichung in die Form

$$(30) \quad u_{tt} - (a-d)u_{xx} - (c-d)u_{yy} - \frac{1}{2}(d+b)(u_{xx} + 2u_{xy} + u_{yy}) \\ - \frac{1}{2}(d-b)(u_{xx} - 2u_{xy} + u_{yy}) + \alpha u_t + \beta u_x + \gamma u_y + \delta u = 0.$$

Wir legen nun in den Raum das Gitter der Punkte

$$t = lh, \quad x + y = m\kappa h, \quad x - y = n\kappa h \quad (l, m, n = \dots - 1, 0, 1, 2, \dots)$$

und ersetzen die Gleichung (30) durch eine Differenzgleichung $L(u) = 0$ in diesem Gitter.

Wir ordnen zu dem Zweck jedem Gitterpunkt P_0 folgende Nachbarpunkte zu: Die Punkte P'_α bzw. P_α , die aus P_0 durch Verschiebung um h bzw. $-h$ in Richtung der t -Achse entstehen; ferner die Punkte P_1, \dots, P_8 , die mit P_0 in derselben Parallelebene zur (x, y) -Ebene liegen; vgl. Fig. 11. Diese Punkte bilden ein „Elementaroktaeder“ mit den Eckpunkten $P'_\alpha, P_\alpha, P_1, P_2, P_3, P_4$.

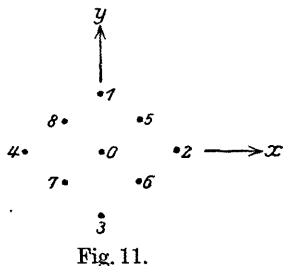


Fig. 11.

Für jeden Gitterpunkt P_0 , der innerhalb von G liegt, ersetzen wir die in (30) auftretenden zweiten Differentialquotienten folgendermaßen durch Differenzenquotienten aus dem zu P_0 gehörigen Elementaroktaeder.

Wir ersetzen

$$u_{tt} \quad \text{durch} \quad \frac{1}{h^2} (u'_\alpha - 2u_0 + u_\alpha) \\ u_{xx} \quad \text{durch} \quad \frac{1}{\kappa^2 h^2} (u_3 - 2u_0 + u_4) \\ u_{yy} \quad \text{durch} \quad \frac{1}{\kappa^2 h^2} (u_1 - 2u_0 + u_3) \\ u_{xx} + 2u_{xy} + u_{yy} \quad \text{durch} \quad \frac{4}{\kappa^2 h^2} (u_6 - 2u_0 + u_8) \\ u_{xx} - 2u_{xy} + u_{yy} \quad \text{durch} \quad \frac{4}{\kappa^2 h^2} (u_5 - 2u_0 + u_7).$$

Die in (30) auftretenden ersten Differentialquotienten ersetzen wir durch irgendwelche entsprechende Differenzenquotienten in dem Elementaroktaeder. Den Koeffizienten in der Differenzgleichung geben wir die Werte, die die Koeffizienten der Differentialgleichung im Punkte P_0 annehmen.

Auf den ersten beiden Anfangsebenen $t = 0$ und $t = h$ denken wir uns die Funktionswerte so vorgegeben, daß sie bei Verfeinerung der Maschenweiten unter Festhaltung des Verhältnisses κ der Raum- zur Zeit-

maschenweite gegen die vorgegebenen Anfangswerte auf $t=0$ streben, wobei die zwischen den beiden Ebenen $t=0$ und $t=h$ gebildeten Differenzenquotienten bis zur vierten Ordnung gegen die entsprechenden vorgegebenen Differentialquotienten gleichmäßig konvergieren sollen.

Die Lösung der Differenzgleichung $L(u)=0$ in einem Punkte ist eindeutig durch die Werte auf den beiden Grundflächen der durch ihn gehenden Bestimmtheitspyramide bestimmt.

Für den Konvergenzbeweis bilden wir die über alle Elementaroktaeder einer Bestimmtheitspyramide erstreckte Summe

$$h^3 \sum \sum \sum 2 \frac{u_{a'} - u_a}{h} L(u)$$

und formen sie mittels der Identitäten (7), (8) um. Dadurch entsteht einmal eine mit h^3 multiplizierte Raumsumme, die in den ersten Differenzenquotienten quadratisch ist, und ferner eine mit h^2 multiplizierte Summe über die Seitendoppelflächen, in denen die Quadrate *aller* auf und zwischen den Doppelflächen vorkommenden Differenzenquotienten vom Typus $u_a - u_0, u_a - u_1, \dots, u_a - u_8$ auftreten, wobei ihre Koeffizienten wegen (29) größer als eine feste positive Konstante sind, wenn wir überdies noch das Verhältnis $\frac{1}{\varkappa}$ von Zeit- zu Raummaschenweite genügend klein wählen.

Von hier aus können wir ganz ebenso vorgehen wie in den §§ 7, 4 und nachweisen, daß die Lösungen unserer Differenzgleichung gegen die Lösung der Differentialgleichung konvergieren.

(Eingegangen am 1. 9. 1927.)

Untersuchungen über allgemeine Metrik.

Von

Karl Menger in Wien.

Inhaltsübersicht.

	Seite
Einleitung	75
Erste Untersuchung. Theorie der Konvexität	76
Zweite Untersuchung. Die euklidische Metrik	113
Dritte Untersuchung. Entwurf einer Theorie der n -dimensionalen Metrik	142

Einleitung.

Von Metrik ist in verschiedenen Teilen der Geometrie die Rede, in der Elementargeometrie, in der Differentialgeometrie, in der Maßtheorie, in der Punktmengenlehre. Zwischen den verschiedenen Gesichtspunkten, unter denen das Messen untersucht wird, einen Zusammenhang auf allgemeinste (mengentheoretischer) Grundlage herzustellen, ist eine der Aufgaben dieser Untersuchungen.

Wir knüpfen an den Begriff des metrischen Raumes der Punktmengenlehre. Zu diesem Begriff gelangte Fréchet, indem er wichtige Eigenschaften der euklidischen Abstandsmetrik (jene nämlich, welche für die Lehre von den reellen Funktionen und zur topologischen Untersuchung der Teilmengen euklidischer Räume ausreichend sind) für abstrakte Mengen postulierte, hingegen von der Tatsache abstrahierte, daß im n -dimensionalen euklidischen Raum der Punkt als n -Tupel von reellen Zahlen gegeben ist und der Abstand zweier Punkte durch eine gewisse Funktion von den Koordinatendifferenzen der betreffenden Punkte dargestellt wird. Diese Metrik ist ein sehr bequemes (wenn auch nicht ganz naturgemäßes) Hilfsmittel für topologische Untersuchungen, und als solches *Hilfsmittel* ist sie bisher verwendet worden. Daß die Metrik *an sich* in ihrer allgemeinsten Form etwas vernachlässigt wurde und hierdurch einige Problemkreise der

mententheoretischen Geometrie unbearbeitet blieben, die sowohl an sich als auch zur systematischen und allgemeinen Fundierung von klassischen geometrischen Disziplinen von Interesse sind, — dies geht schon daraus hervor, daß nicht eine einzige spezifisch-metrische gestaltliche Eigenschaft metrischer Räume überhaupt definiert worden ist, daß eine Charakterisierung der wichtigsten Räume, speziell der euklidischen, unter den metrischen Räumen nicht angegeben worden ist, und daß von einer allgemeinen n -dimensionalen Metrik oder vom abstrakten Kern des Begriffes des Riemannschen Raumes gar nicht die Rede war.

Diese und ähnliche Fragen sollen in einer Reihe von Untersuchungen, von der die drei ersten Aufsätze nun erscheinen, behandelt werden. Ein gemeinsames Band, welches unsere Untersuchungen über allgemeine Metrik verknüpft, bildet die bekannte *Dreiecksungleichung* (die Tatsache, daß in einem Dreieck keine Seite größer als die Summe der beiden anderen Seiten sein kann) — und die mannigfachsten Verschärfungen, Abschwächungen und höherdimensionalen Analoga dieser elementaren Beziehung, deren Auswirkungen sich auf einen großen Teil der gesamten Mathematik erstrecken.

Die Abhandlungen dieses Zyklus sind so gehalten, daß sie vom Leser weder besondere Vorkenntnisse noch eigene Ergänzungen der Beweise erfordern.

Erste Untersuchung: Theorie der Konvexität.

1. Der metrische Raum.
2. Zwischenpunkte in metrischen Räumen.
3. Der Begriff der Konvexität.
4. Mittelpunkte in kompakten konvexen Räumen.
5. Mittelpunkte in vollständigen konvexen Räumen.
6. Existenz von geodätischen Bögen in vollständigen konvexen Räumen.
7. Einige Eigenschaften geodätischer Bögen.
8. Einige Eigenschaften konvexer Räume und Mengen.
9. Über Konvexifizierbarkeit.
10. Über metrisch-singuläre Punkte und ihre Verteilung.
11. Räume mit Strecken.
12. Räume mit Geraden.

1. Der metrische Raum.

Als *metrischen Raum* bezeichnet man eine Menge (deren Elemente *Punkte* genannt werden), wenn für jedes Paar a, b von Elementen ein *Abstand* ab definiert ist, d. i. eine reelle Zahl $ab = ba > 0$ für $a \neq b$ und $ab = 0$ für $a = b$, so daß für jedes Punktetripel a, b, c die sogen. *Dreiecksungleichung* $ab + bc \geq ac$ besteht.

Auf Grund der Abstandsdefinition (ohne Rücksicht darauf, ob der Abstand die Dreiecksungleichung erfüllt oder nicht) läßt sich im Raum auf eine natürliche Weise ein *Limes* definieren, indem man die Punktefolge $\{p_n\}$ ($n = 1, 2, \dots$ ad inf.) konvergent und den Punkt p den *Limespunkt* dieser Punktefolge nennt, falls $\lim_{n=\infty} p_n p = 0$ gilt und nur dann.

Von den Konsequenzen der Dreiecksungleichung heben wir hier bloß eine hervor, auf die wir im folgenden öfters unter dem Namen „*Stetigkeit der Metrik*“ hinweisen werden: Sind $\{a_n\}$ und $\{b_n\}$ ($n = 1, 2, \dots$ ad inf.) zwei Punktefolgen, so daß $\lim_{n=\infty} a_n = a$ und $\lim_{n=\infty} b_n = b$ gilt, dann ist $\lim_{n=\infty} a_n b_n = a b$. Wäre diese Behauptung unrichtig, dann gäbe es unter den Zahlen $a_n b_n$ unendlich viele, etwa die Zahlen der Teilfolge $\{a_{i_n} b_{i_n}\}$ ($n = 1, 2, \dots$ ad inf.), die sich von $a b$ um mehr als eine gewisse positive Zahl ε unterscheiden. Es wäre also für alle n

$$(\dagger) \quad |a_{i_n} b_{i_n} - a b| > \varepsilon > 0.$$

Da die Folgen $\{a_{i_n}\}$ bzw. $\{b_{i_n}\}$ als Teilfolgen von $\{a_n\}$ bzw. $\{b_n\}$ gegen a bzw. b konvergieren, so gilt für eine gewisse natürliche Zahl \bar{n} :

$$a_{i_{\bar{n}}} a < \frac{\varepsilon}{4} \quad \text{und} \quad b_{i_{\bar{n}}} b < \frac{\varepsilon}{4}.$$

Daher ergibt die Dreiecksungleichung, angewendet auf das Punkttripel $a_{i_{\bar{n}}}, a, b_{i_{\bar{n}}}$

$$|a b_{i_{\bar{n}}} - a_{i_{\bar{n}}} b_{i_{\bar{n}}}| < \frac{\varepsilon}{4}$$

und angewendet auf das Punkttripel $a, b, b_{i_{\bar{n}}}$

$$|a b - a b_{i_{\bar{n}}}| < \frac{\varepsilon}{4};$$

also gilt

$$|a b - a_{i_{\bar{n}}} b_{i_{\bar{n}}}| < \frac{\varepsilon}{2}$$

im Widerspruch zu (\dagger) . Damit ist die Annahme von der Unrichtigkeit unserer Behauptung ad absurdum geführt und die Stetigkeit der Metrik bewiesen.

2. Zwischenpunkte in metrischen Räumen.

Nach der Dreiecksungleichung gilt für je drei Punkte a, b, c des Raumes $a b + b c \geq a c$. Es ist naheliegend, jene Punkttripel zu betrachten, für welche in der Dreiecksungleichung das Gleichheitszeichen gilt. Wir wollen, wenn die Beziehung $a b + b c = a c$ besteht, den Punkt b , falls er sowohl von a als auch von c verschieden ist, einen *Zwischenpunkt von a*

und b nennen oder statt dessen auch sagen, b liege zwischen a und c . Eine einfache Untersuchung zeigt, daß derartige Punkttripel in ihren Verhältnissen zueinander zwar nicht alle Relationen von Pasch¹⁾ erfüllen, aber doch einige wichtige dieser Beziehungen²⁾.

Mit Rücksicht auf die Kommutativität der Addition der reellen Abstandszahlen und auf die Symmetrie der Abstandsfunktion ($ab = ba$) gilt offenbar:

1. Wenn b zwischen a und c liegt, dann liegt b auch zwischen c und a . Es gilt ferner:

2. Wenn b zwischen a und c liegt, dann liegt weder a zwischen b und c , noch c zwischen a und b . Denn würden etwa die Gleichungen $ab + bc = ac$ und $ba + ac = bc$ zusammenbestehen, so würde $ab = 0$, also $a = b$ folgen und das widerspräche der Voraussetzung, daß b zwischen a und c liegt.

3. Liegt b zwischen a und c und liegt zugleich c zwischen a und d , dann liegt b zwischen a und d und es liegt c zwischen b und d . Indem wir für drei Punkte a, b, c , von denen b zwischen a und c liegt, einfach abc schreiben, können wir die eben ausgesprochene Behauptung auch so symbolisieren:

$$\left. \begin{array}{l} abc \\ acd \end{array} \right\} \rightarrow \left\{ \begin{array}{l} abd \\ bcd \end{array} \right.$$

Da die Dreiecksungleichung für die Punkttripel a, b, d und b, c, d sicher erfüllt ist, gibt es zwei nicht-negative Zahlen v und w , so daß gilt

$$(\dagger) \quad ab + bd = ad + v, \quad bc + cd = bd + w.$$

Substituieren wir den Wert von bd , der sich aus der zweiten Gleichung ergibt, in die erste, so erhalten wir $ab + bc + cd - w = ad + v$. Falls nun b zwischen a und c liegt, so ist die linke Seite der letzten Gleichung

¹⁾ Vgl. Pasch, Vorlesungen über neuere Geometrie, 1. Aufl. 1882, 2. Aufl. bei Springer 1926. Vgl. insbes. § 1.

²⁾ Nicht nur die Definition von *gestaltlichen Eigenschaften* (vgl. meinen Bericht über die Dimensionstheorie, Jahresbericht d. deutsch. Mathem. Ver. 35, insbes. S. 119), sondern auch die Definition der *elementarsten Beziehungen*, welche, wie z. B. die Zwischenbeziehung, dadurch definiert werden, daß sie gewissen Beziehungen noch einfacherer Natur genügen, — auch diese Definitionen sind ein *in gewissem Maße willkürliches* Umgrenzen der zu definierenden Beziehung gegenüber anderen, welches nur durch *Angleichung an den natürlichen Sprachgebrauch* und gewisse *Zweckmäßigkeitsgründe* bestimmt wird, und demnach nicht nur formal, sondern auch inhaltlich *in verschiedener Weise* erfolgen kann. Beispielsweise weicht unser Zwischenbegriff vom elementargeometrischen etwas ab (siehe unten). Wo es sich darum handelt, die beiden auseinanderzuhalten, wäre etwa unser „zwischen“ als „metrisch zwischen“ zu bezeichnen.

$= ac + cd - w$; falls c zwischen a und d liegt, ist dies weiter gleich $ad - w$; dies soll $= ad + v$ sein, woraus wegen der Nichtnegativität von v und w folgt, daß $v = w = 0$ ist. Aus (†) ergibt sich also $ab + bd = ad$, $bc + cd = bd$. Zum Beweise, daß b zwischen a und d und daß c zwischen b und d liegt, haben wir noch zu zeigen, daß b von a und von d und daß c von b und von d verschieden ist. Es ergibt sich $a \neq b$, $b \neq c$, $c \neq d$ daraus, daß b zwischen a und c und daß c zwischen a und d liegt. Wegen $bc > 0$ und $cd > 0$ ist $bd = bc + cd > 0$; also ist auch $b \neq d$. Damit ist unsere Behauptung bewiesen.

Setzt man in der bewiesenen Formel d statt a , c statt b , b statt c , a statt d und berücksichtigt man die Eigenschaft 1 der Zwischenrelation, so sieht man, daß umgekehrt

$$\left. \begin{array}{l} abd \\ bcd \end{array} \right\} \rightarrow \left\{ \begin{array}{l} abc \\ acd \end{array} \right.$$

gilt, also

$$\left. \begin{array}{l} abc \\ acd \end{array} \right\} \equiv \left\{ \begin{array}{l} abd \\ bcd \end{array} \right.$$

Eine systematische Untersuchung der von Pasch begründeten Axiomatik des *elementargeometrischen* Zwischenbegriffes wurde von Huntington und Kline³⁾ durchgeführt. Die beiden Autoren stellen unter anderem folgendes System von unabhängigen Postulaten für den Zwischenbegriff auf:

Forderungen A und C , welche mit den von uns bewiesenen Eigenschaften 1 und 2 übereinstimmen.

Eine Forderung D , welche besagt, daß, wenn b zwischen a und c liegt, je zwei von den drei Punkten a, b, c verschieden sind. Wir hatten in die Definition unseres Zwischenbegriffes aufgenommen, daß b von a und von c verschieden ist. Die Punkte a und c sind dann wegen $ac = ab + bc > 0$ eo ipso verschieden.

Eine Bedingung II , welche in der obigen Symbolik geschrieben werden kann

$$\left. \begin{array}{l} abc \\ acd \end{array} \right\} \rightarrow bcd.$$

Sie ist in der von uns bewiesenen Eigenschaft 3 enthalten.

Ein Postulat B , welches besagt, daß von je drei verschiedenen Punkten einer Zwischenpunkt der beiden anderen ist. Diese Forderung scheint mir indes eigentlich eine solche *an den Raum* und nicht eine solche an den Zwischenbegriff zu sein. Sie ist in unsere Zwischendefinition nicht aufgenommen und wird von dieser im allgemeinen auch nicht erfüllt.

Eine Bedingung I , welche in der obigen Symbolik geschrieben werden kann:

$$\left. \begin{array}{l} abc \\ bcd \end{array} \right\} \rightarrow abd.$$

³⁾ Trans. Am. Math. Soc. 18 (1917), S. 301.

Diese Bedingung erfüllt unser Zwischenbegriff im allgemeinen *nicht*. Man sieht dies, wenn man für die vier Punkte a, b, c, d als Abstände etwa festsetzt:

$$ab = bc = ad = cd = 1, \quad ac = bd = 2$$

oder:

$$ab = bc = 3, \quad ac = 6, \quad cd = 4, \quad bd = 7, \quad ad = 5.$$

Unsere Zwischenbeziehung erfüllt also die Huntington-Klinesche Bedingung II, nicht aber die Bedingung I.

Daß wir für die folgenden Untersuchungen mit der Eigenschaft 3 der Zwischenbeziehung auskommen und die Relation II von Huntington und Kline nicht benötigen, erscheint bemerkenswert, wenn man sich vergegenwärtigt, welche Konsequenzen die Tatsache hat, daß unsere Zwischenbeziehung diese Relation II nicht erfüllt. Es hängt hiermit beispielsweise zusammen, daß ein Punkt, welcher zwischen zwei Zwischenpunkten von a und b liegt, selbst nicht notwendig Zwischenpunkt von a und b ist. Es seien nämlich die Abstände der Punkte a, b, c, d, e etwa die folgenden: $ac = bd = 2$, $ec = ed = 3$, $ea = eb = 5$, $ad = bc = cd = 6$, $ab = 8$. Man verifiziert ohne weiteres, daß für je drei dieser Punkte die Dreiecksungleichung erfüllt ist, daß sowohl c als auch d Zwischenpunkt von a und b und daß e Zwischenpunkt von c und d ist, daß aber e nicht Zwischenpunkt von a und b ist⁴).

Anschaulich läßt sich die von unserer Zwischenbeziehung erfüllte Eigenschaft 3 folgendermaßen symbolisieren:

Aus



folgt



und umgekehrt.

Die unserer Zwischenbeziehung *nicht* zukommende Eigenschaft der elementargeometrischen Zwischenbeziehung läßt sich entsprechend symbolisieren:

⁴) Unser (vom elementargeometrischen abweichende) Zwischenbegriff ist, wie noch erwähnt werden möge, auch einer Anwendung auf die Erfahrungswelt fähig. Betrachten wir etwa das Netz der europäischen Schnellzuglinien. Es entspricht der Ausdrucksweise von Reisenden, als *Abstand* zweier Städte die (etwa in Stunden gemessene) Dauer der schnellsten Fahrt zwischen den beiden Orten zu bezeichnen. Sehen wir dann ab von den praktisch meist zu vernachlässigenden Unterschieden zwischen der Fahrdauer von A nach B und der Fahrdauer von B nach A , so bilden vermöge dieser Abstandsdefinition die Stationen offenbar einen die Dreiecksungleichung erfüllenden metrischen Raum. Es entspricht wieder einer gebräuchlichen Ausdrucksweise, von der Stadt B zu sagen, sie liege *zwischen* den Städten A und C , wenn man von A ohne Umweg via B nach C gelangen kann, d. i. aber, präzis gesprochen, wenn für die Abstände $AB + BC = AC$ gilt. Es liegt dann beispielsweise zwischen Wien und Amsterdam sowohl Frankfurt a. M. als auch Leipzig, obwohl Frankfurt weder zwischen Wien und Leipzig noch zwischen Leipzig und Amsterdam und obwohl Leipzig weder zwischen Wien und Frankfurt noch zwischen Frankfurt und Amsterdam liegt. Und es liegt Erfurt zwischen Frankfurt a. M. und Leipzig, also zwischen zwei Städten, die zwischen Wien und Amsterdam liegen, ohne selbst zwischen Wien und Amsterdam zu liegen, weil eine schnelle Verbindung von Amsterdam und Wien via Erfurt nicht existiert.

Aus $\left\{ \begin{array}{cccc} \bullet & \bullet & \bullet & \\ & \bullet & \bullet & \bullet \end{array} \right.$
 folgt $\left\{ \begin{array}{cccc} \bullet & \bullet & & \bullet \\ \bullet & & \bullet & \bullet \end{array} \right.$

Wir leiten nun noch eine weitere Eigenschaft unserer Zwischenbeziehung her:

4. Für jedes Punktepaar a, b des Raumes ist die Menge bestehend aus a, b und allen Zwischenpunkten von a und b abgeschlossen. Sei nämlich der Punkt c ein von a und von b verschiedener Häufungspunkt von Punkten, die zwischen a und b liegen. Es existiert dann eine Folge $\{c_n\}$ ($n=1, 2, \dots$ ad inf.) von Punkten zwischen a und b , so daß $c = \lim_{n=\infty} c_n$ ist. Für jedes n gilt $ac_n + c_nb = ab$. Wegen der Stetigkeit der Metrik folgen aus $c = \lim_{n=\infty} c_n$ die Beziehungen $ac = \lim_{n=\infty} ac_n$ und $bc = \lim_{n=\infty} bc_n$. Also ist $ac + cb = ab$, d. h. c ist Zwischenpunkt von a und b . Damit ist die Behauptung bewiesen. Die Menge aller Zwischenpunkte von a und b werden wir bisweilen mit $Z(a, b)$, die (nach dem Bewiesenen abgeschlossene) Menge bestehend aus a, b und $Z(a, b)$ mit $\bar{Z}(a, b)$ bezeichnen.

3. Der Begriff der Konvexität.

Die klassische Definition, welche insbesondere den Minkowskischen Untersuchungen zugrunde liegt, lautet: Eine abgeschlossene Menge A des R_n , des n -dimensionalen euklidischen Raumes, heißt konvex, wenn sie zu je zweien ihrer Punkte die durch die beiden Punkte bestimmte Strecke als Teil enthält. Diese Definition des Konvexitätsbegriffes ist prinzipiell auf Teilmengen des R_n bzw. auf Teilmengen von Räumen beschränkt, in denen Strecken definiert sind, und je zwei Punkte eine Strecke bestimmen. In einem allgemeinen metrischen Raum ist dies keineswegs der Fall. Wollen wir daher für allgemeine metrische Räume und ihre Teilmengen einen Konvexitätsbegriff definieren, so müssen wir uns nach einem anderen Prinzip für die Definition umsehen. Ein solches liefert uns der (wie wir sahen, für allgemeine metrische Räume definierbare) Begriff des *Zwischenpunktes*. Denken wir uns nämlich ein konvexes Gebilde eines euklidischen Raumes mit einer lichtundurchlässigen Substanz ausgefüllt, dann können wir, wenn wir in irgendeinem beliebigen Punkt des Gebildes eine Lichtquelle anbringen, und in irgendeinen beliebigen davon verschiedenen Punkt unser Auge begeben, niemals die Lichtquelle erblicken, weil stets zwischen der Lichtquelle und unserem Auge ein (lichtundurchlässiger) Punkt des

Gebildes liegt. Diese Eigenschaft der euklidischen konvexen Körper läßt sich aber für allgemeine metrische Räume aussprechen, indem wir sagen:

Ein metrischer Raum heißt konvex, wenn zu je zweien seiner Punkte ein Zwischenpunkt existiert.

Um die *Berechtigung* der Bezeichnung *konvex* für den eben definierten Begriff nachzuweisen, d. h. um zu zeigen, daß die hervorgehobene Eigenschaft der konvexen Gebilde für dieselben wirklich *charakteristisch* ist, werden wir im Verlaufe dieser Untersuchung zu beweisen haben, daß einerseits jede konvexe Teilmenge eines euklidischen Raumes, wofern sie in sich betrachtet und als metrischer Raum aufgefaßt wird, im Sinn unserer Definition konvex ist, und daß andererseits jede nicht-konvexe euklidische Menge, in sich betrachtet, nicht unter die in unserem Sinne konvexen Räume fällt.

Der erste Teil dieser Behauptung liegt auf der Hand, denn jeder Punkt der Strecke, welche durch die Punkte a und b eines euklidischen Raumes bestimmt wird, ist Zwischenpunkt von a und b . Der R_n und jeder konvexe Körper eines euklidischen Raumes liefern also Beispiele von konvexen metrischen Räumen im Sinn unserer Definition. Eine Kreislinie und eine Kugeloberfläche sind, wofern als Abstand je zweier Punkte die Länge der durch sie bestimmten Strecke festgesetzt wird, offenbar *nicht* konvex. Sie sind dagegen konvex, wofern als Abstand je zweier Punkte die Länge des kürzesten die beiden Punkte verbindenden Bogens festgesetzt wird.

Die Kugel mit Sehnenmaß und die Kugel mit Bogenmaß sind homöomorphe metrische Räume, von denen der zweite konvex, der erste nicht konvex ist. Daraus geht hervor, daß die *Konvexität eine spezifisch-metrische Eigenschaft* eines metrischen Raumes⁵⁾ ist, die homöomorphen, aber verschiedenen metrisierten Räumen nicht notwendig gleichzeitig zukommen muß; eine Eigenschaft m. a. W., die durch topologische Abbildungen zerstört werden kann.

Es sei nun eine Teilmenge M eines metrischen Raumes vorgegeben. Wir nennen M eine *konvexe Menge*, wenn die Menge M zu je zweien ihrer Punkte einen Zwischenpunkt enthält. Eine Teilmenge eines Raumes heißt also dann und nur dann konvex, wenn sie, in sich betrachtet, ein konvexer Raum ist. Jede im klassischen Sinn konvexe Teilmenge eines euklidischen Raumes ist daher auch in unserem Sinn eine konvexe Teilmenge.

⁵⁾ Durch die angegebene Konvexitätsdefinition wird (vgl. Einleitung S. 76) scheinbar zum erstenmal eine spezifisch metrische gestaltliche Eigenschaft metrischer Räume erfaßt.

4. Die Existenz von Mittelpunkten in kompakten konvexen Räumen.

Für die folgenden Untersuchungen ist wesentlich, daß in gewissen konvexen Räumen zu je zwei Punkten a und b nicht nur ein Zwischenpunkt, sondern ein *Mittelpunkt* existiert, d. h. ein Punkt, dessen Abstand sowohl von a als auch von b gleich dem halben Abstand ab ist. Aus dem Beispiel einer Kreislinie oder einer Kugeloberfläche, falls als Abstand je zweier Punkte die Länge des kürzesten Bogens zwischen ihnen festgesetzt wird, geht hervor, daß in einem kompakten metrischen Raum zu zwei Punkten unter Umständen *mehrere* Mittelpunkte existieren können. Im Kreis gibt es bei Bogenmaß zu je zwei gegenüberliegenden Punkten zwei Mittelpunkte; auf der Kugeloberfläche mit Bogenmaß bilden die Mittelpunkte von je zwei gegenüberliegenden Punkten (Polen) eine (äquatoriale) Kreislinie. — Wir setzen nun über den konvexen Raum zunächst voraus, daß er *kompakt* sei, d. h. daß jede unendliche Teilmenge des Raumes mindestens einen Häufungspunkt im Raum besitzt, und wir beweisen allgemein:

Zu je zwei Punkten a und b eines kompakten konvexen Raumes und zu jeder positiven Zahl λ existiert ein Punkt c , der zwischen a und b liegt und die Beziehung erfüllt $\frac{ac}{bc} = \lambda$. Der Abstand dieses Punktes c von a ist offenbar $\frac{\lambda}{\lambda+1} ab$. Für $\lambda = 1$ ist also c ein Mittelpunkt von a und b .

Wir erinnern zunächst an eine Konsequenz der vorausgesetzten Kompaktheit des Raumes: (*) Ist A eine abgeschlossene Teilmenge eines kompakten Raumes und ist a ein gegebener Punkt des Raumes, dann enthält A einen Punkt a' , dessen Abstand vom Punkte a durch keinen Punkt von A übertroffen wird, und einen Punkt a'' , dessen Abstand von a durch keinen Punkt von A unterschritten wird. Diese Bemerkung ist eine unmittelbare Konsequenz des bekannten Satzes, daß jede auf einer abgeschlossenen Teilmenge eines kompakten metrischen Raumes definierte stetige Funktion in mindestens einem Punkt ihres Definitionsbereiches ihren Maximal- und ihren Minimalwert annimmt.

Es seien nun a und b zwei Punkte eines kompakten Raumes und es sei λ eine gegebene positive Zahl. Wir betrachten die Menge A bestehend aus allen Punkten, welche zwischen a und b liegen und von a einen Abstand $\leq \frac{\lambda}{\lambda+1} ab$ haben. Die Menge A ist Durchschnitt der abgeschlossenen Menge $\bar{Z}(a, b)$ und der abgeschlossenen Kugelumgebung $\bar{U}\left(a; \frac{\lambda}{\lambda+1} ab\right)$ ⁶⁾. Die Menge A ist also abgeschlossen. Zufolge der Bemerkung (*) existiert

⁶⁾ Mit $\bar{U}(p; r)$ bezeichnen wir die (abgeschlossene) Menge aller Punkte des Raumes, die vom Punkt p einen Abstand $\leq r$ haben.

ein Punkt a' von A , dessen Abstand vom Punkt a durch keinen Punkt von A übertroffen wird. Wir zeigen nun:

1. *Der Punkt a' ist Häufungspunkt von Punkten, die zwischen a' und b liegen.* Die Menge $\bar{Z}(a', b)$ ist abgeschlossen. Wäre die Behauptung 1 unrichtig, dann wäre a' nicht Häufungspunkt von $\bar{Z}(a', b)$, dann wäre also auch die Menge $\bar{Z}(a', b) - (a')$ abgeschlossen. Da diese Menge den Punkt a' nicht enthält, so hätte sie von ihm einen gewissen positiven Abstand und enthielte nach der Bemerkung (*) einen Punkt, er heiße b' , dessen Abstand vom Punkt a' durch keinen Punkt von $\bar{Z}(a', b) - (a')$ unterschritten wird. Wegen der Konvexität des Raumes gäbe es einen Punkt c , der zwischen a' und b' liegt. Da b' zwischen a' und b liegt und c zwischen a' und b' , so liegt nach Eigenschaft 3 der Zwischenbeziehung c auch zwischen a' und b , ist also Punkt der Menge $\bar{Z}(a', b) - (a')$. Da c zwischen a' und b' liegt, ist c von b' verschieden, es gilt also $b'c > 0$. Da c zwischen a' und b' liegt, ist $a'c + cb' = a'b'$. Also ist c ein Punkt der Menge $\bar{Z}(a', b) - (a')$, dessen Abstand von a' $a'c < b'a'$ ist im Widerspruch zur Definition des Punktes b' . Damit ist die Annahme von der Ungültigkeit der Behauptung 1 ad absurdum geführt. Wir zeigen

2. *Jeder Punkt, der zwischen a' und b liegt, hat einen Abstand $> \frac{\lambda}{\lambda+1} ab$ von a .* Sei nämlich c ein Punkt zwischen a' und b . Nach Voraussetzung liegt a' zwischen a und b . Aus diesen beiden Tatsachen folgt nach der Eigenschaft 3 der Zwischenbeziehung:

(†) c ist Zwischenpunkt von a und b ;

(††) a' ist Zwischenpunkt von a und c .

Der Punkt c ist nach (†) ein Zwischenpunkt von a und b , dessen Abstand von a' nach (††) den Abstand ac übertrifft. Nach Definition des Punktes a' kann also c nicht in der Menge A liegen. Da A alle Zwischenpunkte von a und b , die von a einen Abstand $\leq \frac{\lambda}{\lambda+1} ab$ haben, enthält, so muß also $a'c > \frac{\lambda}{\lambda+1} ab$ sein, w. z. b. w.

Wir betrachten nun den Punkt a' . Als Punkt der Menge A hat a' einen Abstand $\leq \frac{\lambda}{\lambda+1} ab$ von a . Nach 1. ist a' Häufungspunkt von Punkten zwischen a' und b ; also ist nach 2. a' Häufungspunkt von Punkten, deren Abstand von $a > \frac{\lambda}{\lambda+1} ab$ ist. Wegen der Stetigkeit der Metrik ist dies nur möglich, wenn $a'c = \frac{\lambda}{\lambda+1} ab$ ist. Damit ist die Existenz eines Zwischenpunktes c von a und b im Abstand $\frac{\lambda}{\lambda+1} ab$ von a erwiesen.

5. Die Existenz von Mittelpunkten in vollständigen konvexen Räumen.

Es sei $\{p_n\}$ ($n = 1, 2, \dots$ ad inf.) eine Punktefolge eines metrischen Raumes. Die Folge heißt eine *Cauchysche Folge*, wenn zu jeder Zahl $\varepsilon > 0$ eine natürliche Zahl n existiert derart, daß für je zwei natürliche Zahlen $n' > n$ und $n'' > n$ die Beziehung gilt $|p_n p_{n'} - p_n p_{n''}| < \varepsilon$. Ein metrischer Raum heißt *vollständig*, wenn jede Cauchysche Punktefolge des Raumes einen Limespunkt im Raum besitzt. Ein kompakter Raum ist a fortiori vollständig.

Um die Existenz von Mittelpunkten in vollständigen konvexen Räumen nachzuweisen, müssen wir, da die Bemerkung (*) für abgeschlossene Teilmengen vollständiger Räume im allgemeinen nicht zutrifft, etwas anders vorgehen, als in kompakten Räumen. Wir führen zunächst einen Hilfsbegriff ein: Sind zwei Punkte a und b eines metrischen Raumes gegeben und ist r eine positive Zahl kleiner als der Abstand ab , dann sagen wir, der Punkt c sei ein *letzter Zwischenpunkt von a und b in $\bar{U}(a; r)$* , falls c ein in $\bar{U}(a; r)$ gelegener Zwischenpunkt von a und b ist und die Menge $\bar{U}(a; r)$ keinen Zwischenpunkt d von a und b enthält, so daß c Zwischenpunkt von a und d ist. Wenn $\bar{U}(a; r)$ keinen Zwischenpunkt von a und b enthält, dann nennen wir a selbst den letzten Zwischenpunkt von a und b in $\bar{U}(a; r)$. Wir zeigen nun:

In einem vollständigen Raum existiert zu je zwei Punkten a und b und zu jeder positiven Zahl r , die kleiner als der Abstand ab ist, ein letzter Zwischenpunkt von a und b in $\bar{U}(a; r)$. Wir nehmen an, die Behauptung treffe für ein vorgelegtes Punktepaar a, b und eine vorgelegte Zahl $0 < r < ab$ nicht zu, und leiten aus dieser Annahme einen Widerspruch her durch Konstruktion einer gewissen transfiniten Punktefolge.

Wir setzen $a = c_0$ und nehmen an, es sei α eine Ordinalzahl der ersten oder zweiten Zahlenklasse, so daß für jede Ordinalzahl β , die > 0 und $< \alpha$ ist, bereits ein Punkt c_β zwischen a und b bestimmt sei gemäß folgenden Bedingungen:

$$(\dagger) \quad 0 \leq a c_\gamma < a c_\beta < r \quad (\gamma < \beta < \alpha);$$

$$(\dagger\dagger) \quad c_\delta c_\gamma + c_\gamma c_\beta = c_\delta c_\beta \quad (\delta < \gamma < \beta < \alpha).$$

Unter diesen Voraussetzungen wählen wir einen Punkt c_α gemäß folgenden Bedingungen:

1. Ist α eine *Grenzzahl*, dann betrachten wir eine gegen α konvergente Folge von Ordinalzahlen $\{\alpha_n\}$ ($n = 1, 2, \dots$ ad inf.). Für die Abstände $a c_{\alpha_n}$ gilt wegen (\dagger) $0 < a c_{\alpha_{n-1}} < a c_{\alpha_n} < r$ ($n = 1, 2, \dots$ ad inf.). Die Abstände $a c_{\alpha_n}$ bilden also eine beschränkte monoton wachsende und daher konvergente Zahlenfolge. Da ferner wegen $(\dagger\dagger)$ für je zwei Punkte $c_{\alpha_k}, c_{\alpha_l}$ ($k < l$) die Beziehung gilt $a c_{\alpha_k} + c_{\alpha_k} c_{\alpha_l} = a c_{\alpha_l}$, so bilden die Punkte $\{c_{\alpha_n}\}$ ($n = 1, 2, \dots$ ad inf.) eine Cauchysche Folge. Also existiert wegen der Vollständigkeit des Raumes ein Limespunkt dieser Punktefolge und

diesen Limespunkt wählen wir als c_α . Offenbar ist für jedes $\beta < \alpha$ $ac_\beta < ac_\alpha$. Ferner ist $ac_\alpha < r$. Wäre nämlich $ac_\alpha = r$, so könnte die Menge $\bar{U}(a; r)$ keinen Zwischenpunkt d von a und b enthalten, so daß c_α Zwischenpunkt von a und d ist, d. h. es wäre c_α ein letzter Zwischenpunkt von a und b in $\bar{U}(a; r)$, während doch ein solcher laut Annahme nicht existiert. Es gilt also

$$(\dagger') \quad 0 < ac_\gamma < ac_\beta < r \quad (\gamma < \beta < \alpha + 1)$$

und wegen der Stetigkeit der Metrik für jedes β, γ , ($\beta < \gamma < \alpha$)

$$c_\beta c_\gamma + c_\gamma c_\alpha = c_\beta c_\alpha,$$

also wegen (\dagger)

$$(\dagger\dagger') \quad c_\delta c_\gamma + c_\gamma c_\beta = c_\delta c_\beta \quad (\delta < \gamma < \beta < \alpha + 1).$$

2. Ist α eine *isolierte* Zahl, dann wählen wir als c_α einen der Punkte von $\bar{U}(a; r)$, die zwischen a und b liegen und für welche, wenn $\alpha > 1$ ist, $c_{\alpha-1}$ zwischen a und c_α liegt. Ein solcher Punkt existiert sicher, weil $c_{\alpha-1}$ laut Annahme nicht letzter Zwischenpunkt von a und b in $\bar{U}(a; r)$ ist. Auch durch diese Wahl werden die Relationen (\dagger') und ($\dagger\dagger'$) offenbar erfüllt.

Nach dem Prinzip der transfiniten Induktion ergibt sich also aus der Annahme von der Nichtexistenz eines letzten Zwischenpunktes von a und b in $\bar{U}(a; r)$, daß für jede Ordinalzahl α der ersten und zweiten Zahlenklasse ein Punkt c_α existiert, so daß für $\beta < \alpha$ $ac_\beta < ac_\alpha < r$ ist. Das ist aber unmöglich, da bekanntlich jede beständig wachsende Folge reeller Zahlen abzählbar ist. Die Annahme der Nichtexistenz eines letzten Zwischenpunktes von a und b in $\bar{U}(a; r)$ führt also zu einem Widerspruch, womit unsere Behauptung bewiesen ist.

Es seien nun in einem vollständigen Raum zwei Punkte a und b und eine positive Zahl λ vorgelegt. Es sei c ein nach dem Bewiesenen existierender letzter Zwischenpunkt von a und b in $\bar{U}\left(a; \frac{\lambda}{\lambda+1} ab\right)$ und es sei d ein letzter Zwischenpunkt von b und c in $\bar{U}\left(b; \frac{1}{\lambda+1} ab\right)$. Wir behaupten: Wenn der Raum konvex ist, so ist $c = d$. Um dies nachzuweisen, machen wir die Annahme $c \neq d$ und leiten aus ihr einen Widerspruch her.

Wegen der Konvexität des Raumes folgt aus der Annahme $c \neq d$ zunächst die Existenz eines Punktes e zwischen c und d . Aus der Eigenschaft 3 der Zwischenrelation ergibt sich, da der Punkt e zwischen c und d und der Punkt d entweder $= b$ ist oder zwischen c und b liegt, daß e zwischen c und b liegt. Da überdies c entweder $= a$ ist oder zwischen a und b liegt, so ist e Zwischenpunkt von a und b .

Da e zwischen c und d liegt und d zwischen c und b oder $= b$ ist, so liegt d , falls $d \neq b$ ist, zwischen e und b und es liegt e zwischen c und b .

Ferner sieht man: da c , wenn $c \neq a$, zwischen a und b liegt und d , wenn $d \neq b$, zwischen c und b , so liegt c , wenn $c \neq a$, zwischen a und d . Da überdies e zwischen c und d liegt, so liegt c , wenn $c \neq a$, zwischen a und e .

Nun bestehen zwei Möglichkeiten: Entweder könnte gelten $ea \leq \frac{\lambda}{\lambda+1} ab$; dann läge also e in $\bar{U}\left(a; \frac{\lambda}{\lambda+1} ab\right)$. Da e Zwischenpunkt von a und b ist und c entweder $= a$ oder Zwischenpunkt von a und e , so widerspräche dies der Voraussetzung, daß c letzter Zwischenpunkt von a und b in $\bar{U}\left(a; \frac{\lambda}{\lambda+1} ab\right)$ ist. Oder es müßte $ea > \frac{\lambda}{\lambda+1} ab$ sein. Dann wäre, da e Zwischenpunkt von a und b ist, $eb < \frac{1}{\lambda+1} ab$, also wäre e Punkt von $\bar{U}\left(b; \frac{1}{\lambda+1} ab\right)$. Da e Zwischenpunkt von b und c ist und d entweder $= b$ oder Zwischenpunkt von b und e , so widerspräche dies der Voraussetzung, daß d letzter Zwischenpunkt von b und c in $\bar{U}\left(b; \frac{1}{\lambda+1} ab\right)$ ist. Die Annahme $c \neq d$ führt also in beiden Fällen zu einem Widerspruch und es muß daher $c = d$ sein. Daraus folgt aber $ac = ad = \frac{\lambda}{\lambda+1} ab$, $bc = bd = \frac{1}{\lambda+1} ab$, womit die Existenz eines Zwischenpunktes c von a und b , so daß $\frac{ac}{bc} = \lambda$ gilt, erwiesen ist.

Auf Grund der Stetigkeit der Metrik zeigt man leicht, daß für je zwei Punkte a und b eines metrischen Raumes und für jede positive Zahl λ die Menge aller Punkte c , für die $\frac{ac}{bc} = \lambda$ gilt, abgeschlossen ist. Aus dem Vorangehenden wissen wir, daß diese Menge, falls der metrische Raum vollständig und konvex ist, nicht leer sein kann.

6. Geodätische Bögen in vollständigen Räumen.

Es sei nun ein vollständiger Raum gegeben und eine abgeschlossene Teilmenge M des Raumes, welche zu je zweien ihrer Punkte einen Mittelpunkt enthält. Es seien a_0 und a_1 zwei Punkte von M . Wir wählen einen der in M sicherlich existierenden Mittelpunkte von a_0 und a_1 und bezeichnen ihn mit $a\left(\frac{1}{2}\right)$. Sein Abstand sowohl von a_0 als auch von a_1 ist $\frac{1}{2}a_0a_1$. Wir nehmen an, es seien bereits für $p = 1, 2, \dots, 2^n$ Punkte $a\left(\frac{p}{2^n}\right)$ definiert, so daß der Abstand

$$(\dagger) \quad a\left(\frac{p}{2^n}\right)a\left(\frac{q}{2^n}\right) = \frac{|p-q|}{2^n} a_0a_1.$$

Wir wählen dann einen der in M enthaltenen Mittelpunkte von $a\left(\frac{p}{2^n}\right)$ und $a\left(\frac{p+1}{2^n}\right)$ und bezeichnen ihn mit $a\left(\frac{2p+1}{2^{n+1}}\right)$. Wir setzen ferner $a\left(\frac{2p}{2^{n+1}}\right) = a\left(\frac{p}{2^n}\right)$. Dadurch ist für $p = 1, 2, \dots, 2^{n+1}$ ein Punkt $a\left(\frac{p}{2^{n+1}}\right)$

definiert. Betrachten wir den Abstand von $a\left(\frac{p}{2^{n+1}}\right)$ und $a\left(\frac{q}{2^{n+1}}\right)$. Sind p und q beide gerade Zahlen, so ergibt sich der Abstand aus (†) $= \frac{|p-q|}{2^{n+1}} a_0 a_1$. Es sei eine der beiden Zahlen ungerade, etwa $p = 2p'$, $q = 2q' - 1$, wo p' und q' ganze Zahlen sind, und es sei etwa $p < q$. Der Punkt $a\left(\frac{q}{2^{n+1}}\right)$ ist Zwischenpunkt von $a\left(\frac{q'-1}{2^n}\right)$ und $a\left(\frac{q'}{2^n}\right)$ und für $p' < q' - 1$ ist $a\left(\frac{q'-1}{2^n}\right)$ Zwischenpunkt von $a\left(\frac{q'}{2^n}\right)$ und $a\left(\frac{p'}{2^n}\right)$. Also ist, wofern nicht $q' - 1 = p'$ gilt, $a\left(\frac{q}{2^{n+1}}\right)$ Zwischenpunkt von $a\left(\frac{p'}{2^n}\right)$ und $a\left(\frac{q'}{2^n}\right)$. Der Abstand $a\left(\frac{q}{2^{n+1}}\right) a\left(\frac{q'-1}{2^n}\right) = \frac{1}{2^{n+1}} a_0 a_1$; nach (†) ist der Abstand $a\left(\frac{q'-1}{2^n}\right) a\left(\frac{p'}{2^n}\right) = \frac{p'-q'+1}{2^n} a_0 a_1$, also ist $a\left(\frac{q}{2^{n+1}}\right) a\left(\frac{p'}{2^n}\right) = \frac{|p-q|}{2^{n+1}} a_0 a_1$. Analog beweist man diese Formel für den Fall $p > q$, und für den Fall, daß p und q zwei ungerade Zahlen sind. Dadurch ist die Formel (†) für $n + 1$ bewiesen. Wir können demnach durch eine abzählbare Folge von Auswahlen (nämlich von Wahlen der sukzessiven Mittelpunkte) sukzessive für jede natürliche Zahl n und für jedes $p \leq 2^n$ einen Punkt $a\left(\frac{p}{2^n}\right)$ definieren, so daß für je zwei dieser Punkte die Formel (†) gilt. Das heißt, wir können für jede dyadisch rationale Zahl r einen Punkt $a(r)$ definieren, so daß für je zwei dyadisch rationale Zahlen r und r' die Beziehung gilt

$$(††) \quad a(r) a(r') = |r - r'| \cdot a_0 a_1.$$

Wir bezeichnen die abzählbare Menge der durch eine bestimmte Folge von Wahlen erhaltenen Punkte $a(r)$, wo r eine dyadisch rationale Zahl bedeutet, mit A und nennen \bar{A} die abgeschlossene Hülle von A . Wir zeigen, daß für jede Zahl r , die der Beziehung $0 \leq r \leq a_0 a_1$ genügt, genau ein Punkt $a(r)$ von \bar{A} existiert, der von a_0 den Abstand r hat.

Einerseits können *nicht mehrere* Punkte von \bar{A} existieren, welche von a_0 den Abstand r haben. Da jeder Punkt von A und mithin jeder Punkt von \bar{A} Häufungspunkt der Menge A ist, genügt es zu zeigen, daß nicht zwei Häufungspunkte von A im Abstand r von a_0 existieren können. Angenommen, es existierten zwei verschiedene im Abstand r von a_0 gelegene Häufungspunkte von A , etwa p und q , so daß $pq = d > 0$. Dann existierten zwei Punkte p' und q' von A , so daß $p'p < \frac{d}{8}$, $q'q < \frac{d}{8}$ gilt. Wegen der Stetigkeit der Metrik könnten dabei p' und q' so gewählt werden, daß

$|a_0 p' - r| \leq \frac{d}{8}$ und $|a_0 q' - r| \leq \frac{d}{8}$, also $p' q' \leq |a_0 p' - a_0 q'| \leq \frac{d}{4}$ gilt. Andererseits wären p' und q' zwei Punkte, die um weniger als je $\frac{d}{8}$ von zwei Punkten (nämlich p und q), die voneinander den Abstand d haben, entfernt wären; also müßte nach der Dreiecksungleichung $p' q' \geq \frac{3d}{4}$ sein. Das ist ein Widerspruch, also existiert höchstens ein Punkt von \bar{A} im Abstand r von a_0 .

Andererseits existiert aber für jedes r , das der Beziehung $0 \leq r \leq a_0 a_1$ genügt, *mindestens ein* Punkt von \bar{A} im Abstand r von a_0 . Für $r = 0$ und $r = a_0 a_1$ ist dies klar. Für jedes r zwischen diesen Grenzen können wir eine Folge $\{d_n\}$ ($n = 1, 2, \dots$ ad inf.) von dyadisch rationalen Zahlen bestimmen, so daß $\lim a_0 a(d_n) = r$ ist. Da die Punktefolge $\{a(d_n)\}$ eine Cauchysche Folge ist, existiert wegen der Vollständigkeit des Raumes ein Limespunkt der Folge, dessen Abstand von a_0 wegen der Stetigkeit der Metrik $= r$ sein muß.

Es existiert also für jede Zahl r mit $0 \leq r \leq a_0 a_1$ *genau ein* Punkt von \bar{A} im Abstand r von a_0 . Wir bezeichnen diesen Punkt mit $a(r)$. Wegen der Stetigkeit der Metrik gilt die Formel ($\dagger\dagger$) für je zwei dieser Punkte $a(r)$, wenn r irgendeine reelle Zahl zwischen 0 und $a_0 a_1$ bezeichnet.

Die Menge \bar{A} ist also eindeutiges und, mit Rücksicht auf die Formel ($\dagger\dagger$) abstandstreues, also beiderseits stetiges und umkehrbar eindeutiges, d. h. topologisches Bild von der Menge aller reellen Zahlen zwischen 0 und $a_0 a_1$, demnach ein einfacher stetiger Bogen zwischen a_0 und a_1 . Da nach Konstruktion A Teil von M und nach Voraussetzung M abgeschlossen ist, so ist der Bogen \bar{A} Teil der Menge M . Wir haben also bewiesen den

Satz von der Existenz geodätischer Bögen. *Es sei M eine abgeschlossene Menge eines vollständigen Raumes, die zu je zweien ihrer Punkte mindestens einen Mittelpunkt enthält. Dann enthält die Menge M zu je zweien ihrer Punkte a_0 und a_1 einen einfachen Bogen, der mit einer Strecke der Länge $a_0 a_1$ kongruent ist, d. h. dessen Punkte sich auf die reellen Zahlen zwischen 0 und $a_0 a_1$ derart abbilden lassen, daß der Abstand je zweier Punkte gleich ist dem absoluten Betrag von der Differenz der zwei entsprechenden reellen Zahlen. Für eine Menge von der vorausgesetzten Art nennen wir einen Bogen von dieser Eigenschaft einen geodätischen Bogen zwischen a_0 und a_1 .*

Ein vollständiger konvexer Raum enthält, wie wir im vorigen Abschnitt festgestellt haben, zu je zweien seiner Punkte einen Mittelpunkt. Im eben bewiesenen Existenzsatz ist daher enthalten das

Korollar. *Ein vollständiger konvexer Raum enthält zu je zweien seiner Punkte einen sie verbindenden geodätischen Bogen.*

7. Einige Eigenschaften geodätischer Bögen.

Wir beweisen zunächst:

In einem vollständigen konvexen Raum sind die geodätischen Bögen zwischen den Punkten a_0 und a_1 unter den aus Zwischenpunkten von a_0 und a_1 bestehenden abgeschlossenen Mengen dadurch gekennzeichnet, daß sie zu je zweien ihrer Punkte genau einen Mittelpunkt enthalten.

Erstens enthält jeder geodätische Bogen zwischen a_0 und a_1 , wie man aus der Formel ($\dagger\dagger$), bzw. aus der abstandstreuen Abbildbarkeit des Bogens auf die reellen Zahlen zwischen 0 und $a_0 a_1$ erkennt, zu je zweien seiner Punkte genau einen Mittelpunkt.

Sei *zweitens* M eine aus Zwischenpunkten von a_0 und a_1 bestehende abgeschlossene Menge eines vollständigen Raumes, die zu je zweien ihrer Punkte genau einen Mittelpunkt enthält. Nach dem Satz von der Existenz geodätischer Bögen enthält M jedenfalls einen geodätischen Bogen B zwischen a_0 und a_1 . Wir machen nun die Annahme, daß M einen außerhalb von B gelegenen Punkt p enthält, und führen dieselbe zu einem Widerspruch. Als Punkt von M ist p nach Voraussetzung Zwischenpunkt von a_0 und a_1 . Wir denken die Numerierung von a_0 und a_1 so gewählt, daß $p a_1 \leq p a_0$ gilt. Der Bogen B enthält dann einen Punkt p' im Abstand $p a_1$ von a_1 und einen (eventuell mit a_0 zusammenfallenden) Punkt p'' im Abstand $2 p a_1$ von a_1 . Nach dem Existenzsatz enthält ferner die Menge M einen geodätischen Bogen B' zwischen p und a_0 . Dieser Bogen besteht aus Zwischenpunkten von p und a_0 . Läuft man von p längs B' nach a_0 , so existiert ein (eventuell mit a_0 zusammenfallender) erster Punkt von B , den man auf diesem Wege trifft. Wir nennen diesen Punkt q . Als Punkt von B' ist q entweder mit a_0 identisch oder Zwischenpunkt von p und a_0 . Es sind nun a priori zwei Fälle denkbar:

Entweder ist $a_0 p'' \leq a_0 q$. Da p'' und q Punkte von B sind, ist in diesem Fall p'' entweder mit q oder mit a_0 identisch oder Zwischenpunkt von a_0 und q . Da ferner q entweder mit a_0 identisch oder Zwischenpunkt von a_0 und p ist, so ist p'' entweder mit a_0 identisch oder Zwischenpunkt von a_0 und p . Da überdies p Zwischenpunkt von a_0 und a_1 ist, so liegt p zwischen p'' und a_1 . Es gilt also $p'' p + p a_1 = p'' a_1 = 2 p a_1$, also $p'' p = p a_1$. D. h. p ist Mittelpunkt von p'' und a_1 . Nun existiert aber auf B ein von p verschiedener Mittelpunkt von p'' und a_1 , nämlich der Punkt p' . Mithin enthält M zwei verschiedene Mittelpunkte von p'' und a_1 , was der Voraussetzung über M widerspricht.

Oder es ist $a_0 p'' > a_0 q$. Dann ist $a_1 p'' < a_1 q$. Da q entweder mit a_0 identisch oder Zwischenpunkt von a_0 und p ist und da p Zwischenpunkt von a_0 und a_1 ist, ist p Zwischenpunkt von q und a_1 , also gilt

$a_1 p + pq = a_1 q$. Da p' Zwischenpunkt von q und a_1 ist, gilt weiter $p q + p a_1 = p' q + p' a_1$. Nun ist nach Voraussetzung $p a_1 = p' a_1$, mithin $p q = p' q$. Da p'' Zwischenpunkt von p' und q ist, so ist $p' q > p' p''$. Wegen der vorausgesetzten Beziehung $p' p'' = p a_1$ gilt daher $p' q > a_1 p$ und mithin $p' q > \frac{1}{2} a_1 q$. Es existiert daher ein Punkt r auf dem Bogen B' zwischen p' und q im Abstand $\frac{1}{2} a_1 q$ von q . Da auch auf dem Bogen B ein von r verschiedener Mittelpunkt von q und a_1 existiert, widerspricht dies der Voraussetzung, daß M genau einen Mittelpunkt von q und a_1 enthält.

Die Annahme, daß die Menge M einen Punkt außerhalb von B enthält, ist also für jeden Fall zu einem Widerspruch geführt, womit der Beweis der behaupteten Kennzeichnung der geodätischen Bögen abgeschlossen ist.

Unter allen Bögen zwischen den Punkten a_0 und a_1 eines vollständigen konvexen Raumes sind die geodätischen Bögen dadurch charakterisiert, daß ihre Länge den kleinstmöglichen Wert hat, nämlich dem Abstand $a_0 a_1$ gleich ist. Wir betrachten *erstens* irgendeine endliche, etwa n Punkte $p_i = a(r_i)$ ($i = 1, \dots, n$) enthaltende Teilmenge unseres oben konstruierten geodätischen Bogens, wobei $0 \leq r_i < r_{i+1} \leq a_0 a_1$ ($i = 1, \dots, n-1$) gelten möge. Es gilt nach der Formel (††) für jede solche Menge $\sum_{i=1}^{n-1} p_i p_{i+1} = a_0 a_1$.

Also ist die Länge unseres Bogens $= a_0 a_1$. Sei *zweitens* irgendein Bogen B zwischen den Punkten a_0 und a_1 vorgegeben. Kleiner als $a_0 a_1$ kann die Länge des Bogens wegen der Dreiecksungleichung nicht sein. Wir nehmen an, B sei kein geodätischer Bogen, enthalte also zwei Punkte b_0 und b_1 , von welchen kein Mittelpunkt dem Bogen B angehört. Dabei denken wir die Bezeichnung so gewählt, daß man von a_0 längs B nach a_1 laufend erst auf b_0 und dann auf b_1 stößt. Der zwischen den Punkten b_0 und b_1 gelegene Teilbogen ist zusammenhängend, enthält also einen Punkt c , der von b_0 und b_1 gleich weit entfernt ist. Da aber c nach Voraussetzung kein Mittelpunkt von b_0 und b_1 ist, so gilt nach der Dreiecksungleichung $b_0 c + c b > b_0 b_1$; also ist die Länge des zwischen b_0 und b_1 gelegenen Teilbogens von $B > b_0 b_1$. Die Länge des zwischen a_0 und b_0 gelegenen Teilbogens ist $\geq a_0 b_0$, die Länge des Teilbogens zwischen b_1 und a_1 ist $\geq b_1 a_1$. Also ist die Länge von $B > a_0 b_0 + b_0 b_1 + b_1 a_1$ und folglich $> a_0 a_1$. Damit ist die behauptete Charakterisierung der geodätischen Bögen bewiesen.

Unter allen Bögen zwischen a und b sind die geodätischen Bögen dadurch gekennzeichnet, daß für je zwei von a und b verschiedene Punkte p und q eines geodätischen Bogens p entweder zwischen a und q oder zwischen q und b liegt oder mit q identisch ist. Daß die Bedingung für

einen geodätischen Bogen erfüllt ist, folgt aus der in die Definition eines geodätischen Bogens aufgenommenen abstandstreuen Abbildbarkeit des Bogens auf eine Menge reeller Zahlen. Nehmen wir andererseits an, es sei B ein Bogen, für welchen die Bedingung nicht erfüllt ist, für welchen also zwei Punkte p und q existieren, so daß $ap + pq > aq$, $qp + pb > qb$, $pq > 0$ gilt. Die Reihenfolge, in welcher man von a längs B nach b laufend die vier betrachteten Punkte trifft, ist entweder a, p, q, b oder a, q, p, b . Im ersten Fall betrachten wir den Ausdruck $ap + pq + pb$. Nach Voraussetzung ist die Summe der beiden ersten Terme $> aq$; nach der Dreiecksungleichung ist ferner $aq + qb \geq ab$; mithin ist der betrachtete Ausdruck $> ab$. Im zweiten Fall betrachten wir den Ausdruck $aq + qp + pb$. Nach Voraussetzung ist die Summe der beiden letzten Terme $> qb$; nach der Dreiecksungleichung ist $aq + qb \geq ab$; mithin ist der betrachtete Ausdruck $> ab$. Die (entweder endliche oder unendliche) Länge des Bogens B ist also jedenfalls $> ab$, also ist B kein geodätischer Bogen.

Wir beweisen nun noch eine gewisse Abgeschlossenheitseigenschaft des Systems aller geodätischen Bögen eines konvexen vollständigen Raumes. Eine Folge $\{M_n\}$ ($n = 1, 2, \dots$ ad inf.) von Mengen heißt im topologischen Sinn *konvergent*, wenn für jeden Punkt des Raumes, von dem sämtliche Umgebungen Punkte mit unendlich vielen Mengen M_n gemein haben, sämtliche Umgebungen Punkte sogar von fast allen Mengen M_n enthalten. Die Menge M aller Punkte des Raumes, deren sämtliche Umgebungen Punkte mit unendlich vielen (und daher mit fast allen) Mengen einer konvergenten Mengenfølge gemein haben, wird mit $\lim_{n \rightarrow \infty} M_n$ bezeichnet und heißt der *Limes* oder die *Näherungsgrenze* der Mengen M_n . Es gilt nun:

Wenn in einem kompakten Raum B_n ein geodätischer Bogen zwischen den Punkten a_n und b_n ist und $\lim a_n = a$ und $\lim b_n = b$ gilt, dann ist, falls die Folge B_n ($n = 1, 2, \dots$ ad inf.) konvergiert, ihr Limes ein geodätischer Bogen zwischen a und b , der sich im Falle $a = b$ auf den Punkt a reduziert.

Wir nehmen zunächst an, es sei $a \neq b$. Die Menge B ist als Limes einer Folge von abgeschlossenen Mengen abgeschlossen. Sie enthält ferner zu je zweien ihrer Punkte mindestens einen Mittelpunkt. Denn sind etwa c und d zwei Punkte von B , dann existieren zwei Punktfolgen $\{c_n\}$ und $\{d_n\}$ ($n = 1, 2, \dots$ ad inf.), so daß c_n und d_n in B_n liegen und so daß gilt: $c = \lim_{n \rightarrow \infty} c_n$, $d = \lim_{n \rightarrow \infty} d_n$. Bezeichnet r_n den in B_n enthaltenen Mittelpunkt von c_n und d_n , so besitzt die Menge dieser Punkte r_n wegen der Kompaktheit des Raumes einen Häufungspunkt r . Dieser Punkt r muß in B liegen und wegen der Stetigkeit der Metrik Mittelpunkt von c und d sein. Wir leiten nun einen Widerspruch her aus der Annahme, daß die

Menge B zu ihren zwei Punkten c und d zwei verschiedene Mittelpunkte s und t enthält. Da s und t Punkte der Menge B sind und $B = \lim B_n$ ist, so existieren zwei Punktfolgen $\{s_n\}$ und $\{t_n\}$ ($n = 1, 2, \dots$ ad inf.), so daß s_n und t_n auf B_n liegen und $s = \lim_{n=\infty} s_n$ und $t = \lim_{n=\infty} t_n$ gilt. Da s und t Mittelpunkte von c und d sind, muß wegen der Stetigkeit der Metrik $\lim_{n=\infty} \frac{s_n c_n}{s_n d_n} = \lim_{n=\infty} \frac{t_n c_n}{t_n d_n} = 1$ gelten. Bezeichnet r_n den auf B_n gelegenen Mittelpunkt von c_n und d_n , so ist demnach wegen der Stetigkeit der Metrik $\lim_{n=\infty} s_n r_n = \lim_{n=\infty} t_n r_n = 0$, und dies ist wegen $s \neq t$ ein Widerspruch. — Endlich sind die Punkte von B durchwegs Zwischenpunkte von a und b . Als abgeschlossene, aus Zwischenpunkten von a und b bestehende Menge, die zu je zweien ihrer Punkte genau einen Mittelpunkt enthält, ist die Menge B ein geodätischer Bogen zwischen a und b , wie behauptet. Ist $a = b$, so ist $\lim a_n b_n = 0$, und die Menge B besteht aus dem Punkte a .

Der Begriff des geodätischen Bogens ist keineswegs nur für konvexe metrische Räume definierbar. In jedem metrischen Raum kann man einen Bogen zwischen den Punkten a und b , dessen Länge von der Länge keines, die Punkte a und b verbindenden Bogens unterschritten wird, als *geodätischen Bogen* bezeichnen. Allerdings existieren nicht in jedem metrischen Raum geodätische Bögen, da ja selbst ein kompakter zusammenhängender Raum nicht einmal Bögen, geschweige denn Bögen von endlicher Länge enthalten muß. In einer der folgenden Untersuchungen über allgemeine Metrik, welche von der Geodäsie metrischer Räume handeln wird, werden wir gelegentlich der Untersuchung dieser geodätischen Linien auch gewisse Bögen von *unendlicher Länge*, nämlich solche, die Limes von geodätischen Bögen endlicher Länge sind, als geodätisch bezeichnen. Die konvexen Räume sind unter den metrischen Räumen dadurch charakterisiert, daß in ihnen zu je zwei Punkten ein dieselben verbindender geodätischer Bogen existiert, dessen Länge gleich dem Abstand seiner Endpunkte ist.

Wir wollen nun die allgemeinen geodätischen Bögen noch kurz einordnen unter die *Extremalen hinsichtlich gewisser Variationsprobleme in allgemeinen metrischen Räumen*. Es sei in einem metrischen Raum eine nicht-negative Funktion definiert, welche jedem Punkt p des Raumes eine nicht-negative Zahl $f(p)$ zuordnet. Es sei B ein Teilbogen des Raumes und M eine endliche Teilmenge von B . Es seien a_0, a_1, \dots, a_n die Punkte von M (unter denen auch die Endpunkte von B vorkommen können), und zwar so geordnet, wie sie bei einer Durchlaufung des Bogens (etwa in der Richtung von a nach b) angetroffen werden. Ist p_i irgendein Punkt des durch a_{i-1} und a_i bestimmten Teilbogens von B , so setzen wir die obere Schranke aller Zahlen $\sum_{i=1}^n f(p_i) \cdot a_{i-1} a_i$, welche durch verschiedene Wahlen der Punkte p_i erhalten werden können, $= J(f, M_B)$. Die obere Schranke aller Zahlen $J(f, M_B)$, wenn M die verschiedenen endlichen Teilmengen von B durchläuft, nennen wir *das Kurvenintegral von f längs dem Bogen B* und bezeichnen diese Zahl mit $J(f, B)$. Ist B ein a und b verbindender Bogen, so daß $J(f, B)$ endlich ist und für jeden a und b verbindenden Bogen B' die Beziehung gilt $J(f, B) \leq J(f, B')$, dann nennen wir B *einen Minimalbogen hinsichtlich der Funktion f* . Die geodätischen Bögen eines metrischen Raumes

sind dann offenbar die Minimalbögen des Raumes hinsichtlich der Funktion $f = 1$ (oder, was dasselbe ist, hinsichtlich der Funktion $f = \text{konst.}$).

Der Satz von der Abgeschlossenheit des Systems aller geodätischen Bögen ordnet sich dann in folgenden Satz ein, den wir hier ohne seinen einfachen Beweis anführen:

Voraussetzung. *Es sei A ein metrischer Raum von folgender Art: Zu jedem $\varepsilon > 0$ existiert ein $\delta > 0$, so daß je zwei Punkte des Raumes, deren Abstand $< \delta$ ist, durch einen Teilbogen von A , dessen Länge $< \varepsilon$ ist, verbindbar sind. Es sei in A eine beschränkte gleichmäßig stetige Funktion f definiert. — **Behauptung.** Sind die Bögen B_n zwischen den Punkten a_n und b_n Minimalbögen hinsichtlich f , und ist der Bogen B zwischen $a = \lim a_n$ und $b = \lim b_n$ gleich dem Limes der Bogenfolge $\{B_n\}$, dann ist auch B Minimalbogen hinsichtlich f .*

8. Einige Eigenschaften konvexer Räume und Mengen.

In einem euklidischen Raum fällt für je zwei verschiedene Punkte a und b die Menge $Z(a, b)$ aller Zwischenpunkte von a und b mit der durch a und b bestimmten Strecke zusammen. Denn für jeden Punkt c dieser Strecke gilt $ac + cb = ab$, und für jeden Punkt außerhalb der Strecke gilt $ac + cb > ab$. Eine abgeschlossene Menge A eines euklidischen Raumes ist, in sich betrachtet, ein vollständiger metrischer Raum. Ist eine abgeschlossene Teilmenge A eines euklidischen Raumes in sich betrachtet ein konvexer Raum, so enthält sie zu je zweien ihrer Punkte einen sie verbindenden Bogen, bestehend aus Zwischenpunkten von a und b , also enthält sie die durch a und b bestimmte Strecke, da ein anderer aus Zwischenpunkten von a und b bestehender Bogen nicht existiert. Die Menge A ist also eine konvexe Menge im klassischen Sinn. Daß umgekehrt jede im klassischen Sinn konvexe abgeschlossene Teilmenge eines euklidischen Raumes, in sich betrachtet, ein vollständiger, im Sinn unserer Definition konvexer Raum ist, wurde bereits im Abschnitt 3 erwähnt. Wir sehen also: *Unter den abgeschlossenen Teilmengen euklidischer Räume sind die im klassischen Sinn konvexen Mengen und nur diese, in sich betrachtet, konvexe Räume im Sinne unserer Definition.*

Aus den bewiesenen Sätzen über geodätische Bögen in konvexen Räumen ergeben sich auch unmittelbar einige *topologische* Eigenschaften konvexer metrischer Räume. Zu je zweien seiner Punkte enthält ein vollständiger konvexer Raum einen sie verbindenden Bogen. Eine Menge, die zu je zwei Punkten einen sie verbindenden Teilbogen enthält, und daher insbesondere *zusammenhängend* ist, heißt ein *Semikontinuum*. Ein konvexer vollständiger Raum enthält zu je zweien seiner Punkte a und b einen Bogen, der eine Länge und folglich auch einen Durchmesser gleich dem Abstand ab hat. Hinreichend benachbarte Punkte des Raumes sind also durch beliebig kurze Bogen verbindbar, d. h. der Raum ist *zusammenhängend im kleinen*, und mithin, falls er kompakt ist, nach dem Theorem von Hahn und Mazurkiewicz *stetig durchlaufbar*.

Sind a und b zwei Punkte eines vollständigen konvexen Raumes und p und q zwei Zwischenpunkte von a und b , so enthält die Menge $\bar{Z}(a, b)$ aller Zwischenpunkte von a und b einen Teilbogen zwischen p und a und einen Teilbogen zwischen q und a , also einen Teilbogen zwischen p und q . Die Menge $\bar{Z}(a, b)$ ist also ein *Semikontinuum*, demnach, da sie abgeschlossen ist (Eigenschaft 4 der Zwischenbeziehung), falls der Raum kompakt ist, ein *Kontinuum*, d. h. eine kompakte zusammenhängende Menge.

Dieses Kontinuum ist nicht notwendig zusammenhängend im kleinen, wie man an folgendem einfachen Beispiel sieht. Man betrachte eine Halbkugel S vom Durchmesser 1, begrenzt durch den Kreis K . Es seien a und b zwei gegenüberliegende Punkte von K . Es sei ferner $\{H_n\}$ ($n = 1, 2, \dots$ ad inf.) eine Folge von Halbkreisen unserer Halbkugel, welche durch die Punkte a und b gehen und gegen den einen der beiden durch a und b bestimmten Halbkreise von K — derselbe heiße H — konvergieren. Man wähle nun auf H eine abzählbare, in H überall dicht liegende Menge von Punkten und ordne dieselbe in eine Folge $\{p_k\}$ ($k = 1, 2, \dots$ ad inf.). Wir betrachten für jedes k den auf S gelegenen, zu K senkrechten Halbkreis durch den Punkt p_k ; und wir nennen F_k den Abschnitt dieses Halbkreises, welcher den Punkt p_k enthält und eine Bogenlänge $= \frac{1}{k}$ hat. Wir legen nun das (im kleinen zusammenhängende) Kontinuum

$$R = H + \sum_{n=1}^{\infty} H_n + \sum_{k=1}^{\infty} F_k$$

als Raum zugrunde und setzen als Abstand je zweier Punkte p und q von R die in Bogenmaß gemessene Länge des kürzesten p und q verbindenden Teilbogens von R fest. Man bestätigt ohne weiteres, daß diese Definition den Forderungen an den Abstandsbegriff genügt, daß insbesondere die Dreiecksungleichung erfüllt ist. Weiters ist der so definierte metrische Raum R konvex; denn je zwei Punkte p und q sind durch einen Teilbogen des Raumes von kürzester Länge verbunden und jeder Punkt eines solchen geodätischen Bogens zwischen p und q ist Zwischenpunkt von p und q . — Betrachten wir nun die bei der Konstruktion von R erwähnten Punkte a und b und die Menge $\bar{Z}(a, b)$, bestehend aus a , b und allen Zwischenpunkten von a und b . Offenbar ist H und jeder der Bögen H_n Teil von $\bar{Z}(a, b)$. Sei nun p ein Punkt des Raumes, welcher weder auf H noch auf einem der Bögen H_n liegt. Der Punkt p liegt auf einem der Bögen F_k etwa auf jenem Abschnitt von F_k , der zwischen H_n und H_{n+1} liegt. Es sei etwa der Schnittpunkt q von F_k mit H_n der von p am wenigsten entfernte oder evtl. einer der beiden von p am wenigsten entfernten Schnittpunkte von F_k mit der Menge $\sum_{n=1}^{\infty} H_n$. Der kürzeste Bogen zwischen p und a besteht offenbar aus dem zwischen p und q gelegenen

Abschnitt von F_k und dem zwischen q und a gelegenen Abschnitt von H_n . Der kürzeste Bogen zwischen p und b besteht aus dem zwischen p und q gelegenen Abschnitt von F_k und dem zwischen q und b gelegenen Abschnitt von H_n . Es ist also $ap + pb$ gleich der doppelten Länge des zwischen p und q gelegenen Abschnittes von F_k plus der Länge von H_n ; mithin ist $ap + pb > ab$, da ab gleich der Länge von H_n ist. Liegt der Punkt p weder in H noch in einer der Mengen H_n , so ist also p kein Zwischenpunkt von a und b . Mithin ist $\bar{Z}(a, b) = H + \sum_{n=1}^{\infty} H_n$. Diese Menge ist aber nicht zusammenhängend im kleinen.

Dagegen ist in einem konvexen vollständigen Raum eine Menge $\bar{Z}(a, b)$ offenbar stets in den Punkten a und b zusammenhängend im kleinen. Zusammenfassend können wir also sagen:

Ein vollständiger konvexer Raum ist ein im kleinen zusammenhängendes Semikontinuum. Für je zwei Punkte a und b eines vollständigen konvexen Raumes ist die Menge $\bar{Z}(a, b)$, bestehend aus a, b und allen Zwischenpunkten von a und b , ein Semikontinuum, welches in den Punkten a und b aber sonst nicht notwendig zusammenhängend im kleinen ist.

9. Über Konvexifizierbarkeit.

L sei eine vorgegebene Limesklasse, d. h. eine Menge, für deren Elemente ein den gewöhnlichen Limeseigenschaften genügender Limes definiert sei. Wir nennen L konvexifizierbar, wenn es möglich ist, je zwei Punkten a und b von L einen Abstand \overline{ab} zuzuordnen, so daß 1. $\overline{ab} = \overline{ba} > 0$ für $a \neq b$ und $\overline{ab} = 0$ für $a = b$ ist, daß 2. für je drei Punkte a, b, c die Dreiecksungleichung $\overline{ab} + \overline{bc} \geq \overline{ac}$ erfüllt ist, daß 3. für die Punktefolge $\{a_n\}$ ($n = 1, 2, \dots$ ad inf.) und den Punkt a $\lim_{n=\infty} \overline{a_n a} = 0$ dann und nur dann gilt, wenn $\lim_{n=\infty} a_n = a$ gilt, und daß 4. der aus L so entstehende metrische Raum konvex ist. Wir beweisen nun:

Es sei R ein kompakter metrischer Raum, in dem je zwei Punkte durch einen Bogen von endlicher Länge verbunden sind, und von der Art, daß zu jedem $\varepsilon > 0$ ein $\delta > 0$ existiert, so daß je zwei Punkte, deren Abstand $< \delta$ ist, durch einen Bogen von einer Länge $< \varepsilon$ verbindbar sind. Dann ist der Raum R konvexifizierbar.

Der Raum R ist als metrischer Raum eine Limesklasse. Wir ordnen je zwei Punkten a und b des Raumes die untere Schranke von den Längen aller a und b verbindenden Bögen zu und bezeichnen diese Zahl mit \overline{ab} . Wir zeigen nun, daß R durch diese Zahlen \overline{ab} konvexifiziert wird.

1. Es ist offenbar $\overline{ab} = \overline{ba} > 0$ für $a \neq b$ und $\overline{ab} = 0$ für $a = b$.
2. Für je drei Punkte a, b, c des Raumes gilt $\overline{ab} + \overline{bc} \geq \overline{ac}$. Angenommen nämlich, es wäre für drei Punkte $\overline{ab} + \overline{bc} = \overline{ac} - r$ ($r > 0$).

Dann existierte ein Bogen B' zwischen a und b und ein Bogen B'' zwischen b und c , so daß die Summe der Längen von B' und $B'' < \overline{ac}$ wäre. Die Summe der Bögen B' und B'' enthielte aber einen a und b verbindenden Bogen. Seine Länge wäre a fortiori $< \overline{ac}$. Dies widerspräche aber der Definition der Zahl \overline{ac} . Damit ist die Annahme von der Ungültigkeit der Dreiecksungleichung ad absurdum geführt.

3. Ist $\{a_n\}$ ($n = 1, 2, \dots$ ad inf.) eine Punktfolge von R , welche gegen den Punkt a konvergiert, für welche also $\lim_{n=\infty} a_n a = 0$ gilt, dann gilt auch $\lim_{n=\infty} \overline{a_n a} = 0$ und umgekehrt. Wir setzen *erstens* voraus, es sei $\lim_{n=\infty} a_n a = 0$. Wir geben eine Zahl $\varepsilon > 0$ vor und haben zu zeigen, daß für fast alle n $\overline{a_n a} < \varepsilon$ gilt. Nach der Voraussetzung über den Raum existiert ein $\delta > 0$, so daß für $a_n a < \delta$ $\overline{a_n a} < \varepsilon$ gilt. Wegen $\lim_{n=\infty} a_n a = 0$ ist für fast alle n $a_n a < \delta$. Mithin sind für jedes vorgegebene $\varepsilon > 0$ fast alle $\overline{a_n a} < \varepsilon$, und es gilt also $\lim_{n=\infty} \overline{a_n a} = 0$. Ist *zweitens* $\lim_{n=\infty} \overline{a_n a} = 0$, so gilt a fortiori $\lim_{n=\infty} a_n a = 0$, denn es ist $0 \leq a_n a \leq \overline{a_n a}$.

4. Der durch die Zahlen \overline{ab} definierte metrische Raum ist konvex. Wir beweisen dies, indem wir zwei beliebige Punkte a und b des Raumes vorgeben und die Existenz eines Punktes c nachweisen, so daß $\overline{ac} = \overline{bc} = \frac{1}{2} \overline{ab}$ ist. Wenn R einen Bogen zwischen a und b enthält, dessen Länge $= \overline{ab}$ ist, dann erfüllt der Punkt c , welcher auf diesem Bogen in der Mitte zwischen a und b liegt, offenbar diese Bedingung. Man kann nun zeigen, daß in einem unsere Bedingungen erfüllenden metrischen Raum R zu je zwei Punkten a und b ein Bogen von der Länge \overline{ab} (d. i. ein Bogen, dessen Länge von der Länge keines a und b verbindenden Bogens unterschritten wird, also ein in diesem Sinn geodätischer Bogen) tatsächlich existiert. Wir können unsere Behauptung von der Existenz eines Punktes c aber auch direkt nachweisen. Jedenfalls existiert ja zu den Punkten a und b ein Folge $\{B_n\}$ von Bögen, deren Längen gegen \overline{ab} konvergieren. Nennen wir r_n jenen Punkt von B_n , für welchen die Länge des Teilbogens zwischen a und r_n gleich der Länge des Teilbogens zwischen r_n und b ist. Wegen der Kompaktizität des Raumes existiert ein Häufungspunkt r dieser Punkte r_n . Wir zeigen nun, daß $\overline{ar} = \overline{br} = \frac{1}{2} \overline{ab}$ gilt. Wir beweisen die Beziehung $\overline{ar} = \frac{1}{2} \overline{ab}$. Angenommen *erstens*, es sei $\overline{ar} = \frac{1}{2} \overline{ab} + d$ ($d > 0$). Nach unserer Voraussetzung über den Raum liegen fast alle Punkte r_n so nahe an dem Punkt r , daß $\overline{r_n r} < \frac{d}{4}$ ist. Wir wählen unter diesen Punkten r_n einen, welcher überdies so nahe an r liegt, daß die Länge des zwischen a und r_n gelegenen Teilbogens von $B_n < \frac{1}{2} \overline{ab} + \frac{d}{4}$ ist. Wegen $\overline{r_n r} < \frac{d}{4}$

existiert ein Bogen B' zwischen r_n und r , dessen Länge $< \frac{d}{4}$ ist. Die Summe von B' und dem zwischen a und r_n gelegenen Teilbogen von B_n enthält einen Teilbogen zwischen a und r , dessen Länge $< \frac{1}{2}\overline{ab} + \frac{d}{2}$ ist, was der Voraussetzung $\overline{ar} = \frac{1}{2}\overline{ab} + d$ widersprechen würde. Damit ist die Beziehung $\overline{ar} > \frac{1}{2}\overline{ab}$ ad absurdum geführt. — Angenommen *zweitens*, es sei $\overline{ar} = \frac{1}{2}\overline{ab} - d$ ($d > 0$). Für fast alle n wäre $\overline{r_n r} < \frac{d}{4}$ und die Länge von $B_n < \overline{ab} + \frac{d}{2}$. Für hinreichend große n existierte also ein Bogen B'_n zwischen r und r_n von einer Länge $< \frac{d}{4}$ und wäre die Länge des zwischen r_n und b gelegenen Teilbogens von $B_n \leq \frac{1}{2}\overline{ab} + \frac{d}{4}$. Wäre $\overline{ar} = \frac{1}{2}\overline{ab} - d$, so existierte ein Bogen \overline{B} zwischen a und r von einer Länge $< \frac{1}{2}\overline{ab} - \frac{3}{4}d$. Für hinreichend großes n würde die Summe der Bögen \overline{B} , B'_n , B_n einen Bogen zwischen a und b enthalten, von einer Länge $< \frac{1}{2}\overline{ab} - \frac{3}{4}d + \frac{d}{4} + \frac{1}{2}\overline{ab} + \frac{d}{4} \leq \overline{ab} - \frac{d}{4}$. Dies widerspräche der Definition von \overline{ab} . Damit ist auch die Beziehung $\overline{ar} < \frac{1}{2}\overline{ab}$ ad absurdum geführt. Es ist also $\overline{ar} = \frac{1}{2}\overline{ab}$ und ebenso beweist man $\overline{br} = \frac{1}{2}\overline{ab}$. Der Punkt r ist mithin ein Mittelpunkt von a und b und der Raum R ist konvex. Damit ist unser Satz bewiesen.

Der eben bewiesene Satz lehrt, daß für die Konvexifizierung eines kompakten Raumes R hinreichend sei, daß je zwei Punkte von R durch einen Bogen von endlicher Länge und hinreichend benachbarte Punkte durch Bögen von beliebig kleiner Länge verbunden sind. Diese Eigenschaft ist sehr verwandt dem Zusammenhang im kleinen. Ein kompakter Raum heißt ja nach Hahn zusammenhängend im kleinen, wenn hinreichend benachbarte Punkte des Raumes durch Bögen von beliebig kleinem Durchmesser verbunden sind. Es ergibt sich die Frage, ob nicht bereits der Zusammenhang im kleinen, welcher wesentlich schwächer als die von uns benützte Eigenschaft ist, für die Konvexifizierbarkeit eines kompakten Raumes hinreichend sei. (Dieses Problem ist auf Grund des Bewiesenen offenbar äquivalent mit der Frage, ob sich jeder im kleinen zusammenhängende metrische Raum so ummetrisieren läßt, daß je zwei hinreichend benachbarte Punkte des Raumes durch einen Bogen verbunden sind, von dem nicht nur der Durchmesser, sondern die Länge unter einem vorgeschriebenen Wert liegt.) Das Problem ist aus folgendem Grund von Interesse. Wir haben oben nachgewiesen, daß jeder kompakte konvexe Raum zusammenhängend im kleinen ist. Es ist also auch jeder kompakte konvexifizierbare Raum zusammenhängend im kleinen. Wenn die Antwort auf das angegebene Problem positiv lautet, so wäre also Zusammenhang im kleinen eines kompakten Raumes (oder was nach Hahn und Mazurkiewicz

gleichbedeutend ist, stetige Durchlaufbarkeit) für die Konvexifizierbarkeit notwendig und hinreichend. *Es wären Konvexifizierbarkeit und Zusammenhang im kleinen identisch* und es wäre damit wohl einer der merkwürdigsten Zusammenhänge der mengentheoretischen Geometrie hergestellt. Wenn aber die Antwort auf die angeführte Frage negativ lauten sollte, wenn es also im kleinen zusammenhängende, aber nicht konvexifizierbare kompakte Räume geben sollte, dann ist zu vermuten, daß die Konvexifizierbarkeit eine Verschärfung des Begriffes vom Zusammenhang im kleinen darstellt, die auch topologisch von Interesse ist.

10. Über metrisch-singuläre Punkte und ihre Verteilung.

In einem metrischen Raum, in dem geodätische Bögen definiert sind, ist es möglich, die metrischen Eigenschaften der Punkte, insbesondere jener Punkte, welche in metrischer Hinsicht gewisse Singularitäten aufweisen, näher zu beschreiben. Liegt ein Punkt p und ein in p endender geodätischer Bogen B vor, dann kann es sein, daß sich B über p hinaus *geodätisch fortsetzen* läßt, in dem Sinn, daß ein in p endender geodätischer Bogen B' existiert, so daß die Summe von B' und einem in p endenden Segment von B ein zwischen seinen Endpunkten geodätischer Bogen ist. In konvexen Räumen hängen die diesbezüglichen Verhältnisse, wie ja die gesamte Geodäsie, mit den gegenseitigen Verhältnissen der Zwischenpunkte zusammen. Beispielsweise läßt sich ein im Punkt p eines konvexen Raumes endender geodätischer Bogen B dann und nur dann im angeführten Sinn über p hinaus geodätisch fortsetzen, wenn ein Punkt q existiert, so daß p Zwischenpunkt von q und einem Punkt von B ist. Wir beweisen zunächst einen Satz über die Verteilung von gewissen Zwischensingularitäten in beliebigen metrischen Räumen. Aus diesem Satz wird sich die Verteilung der metrischen Singularitäten konvexer Räume ergeben.

Es sei R ein metrischer Raum. Zu je zwei Punkten a und b von R ist

1. *die Menge $\bar{Z}(a, b)$, bestehend aus a, b und allen Zwischenpunkten von a und b abgeschlossen;*

2. *die Menge $\bar{Z}[b; a]$, bestehend aus b und allen Punkten c , so daß b zwischen a und c liegt und ebenso die entsprechende Menge $\bar{Z}[a; b]$ abgeschlossen;*

3. *die durch a und b bestimmte „Geradenmenge“ $G(a, b)$, bestehend aus a, b und allen Punkten c , so daß einer der drei Punkte a, b, c Zwischenpunkt der beiden anderen ist, abgeschlossen.*

Ist R ein kompakter metrischer Raum, so ist

4. *die Menge aller Punkte, die Zwischenpunkte von irgend zwei Punkten des Raumes sind, ein F_σ , d. h. Summe von abzählbar vielen abgeschlossenen*

Mengen; und es ist die Menge aller Punkte, die Zwischenpunkte von keinem Punktepaar des Raumes sind, ein G_δ , d. h. Produkt von abzählbar vielen offenen Mengen;

5. zu jedem gegebenen Punkt p des Raumes die Menge aller Punkte q , die zwischen p und irgendeinem Punkt des Raumes liegen und ebenso die Menge aller Punkte q , so daß p zwischen q und irgendeinem Punkt des Raumes liegt, ein F_σ .

Die erste Behauptung wurde im Abschnitt 2 als Eigenschaft 4 der Zwischenbeziehung bewiesen. In analoger Weise ergeben sich aus der Stetigkeit der Metrik die Behauptungen 2. Die in der Behauptung 3 definierte Geradenmenge $G(a, b)$ ist Summe $\bar{Z}(a, b)$, $\bar{Z}[a; b]$ und $\bar{Z}[b; a]$, also abgeschlossen.

Es sei nun R ein kompakter Raum. Wir betrachten für eine natürliche Zahl n die Menge M_n aller Punkte p von folgender Eigenschaft: Es existieren zwei Punkte a_p und b_p , deren Abstand $\geq \frac{1}{n}$ ist und von denen p Zwischenpunkt ist. Wir behaupten, die Menge M_n sei abgeschlossen. Sei nämlich p ein Häufungspunkt von M_n . Es existiert also eine gegen p konvergente Folge $\{p_k\}$ ($k = 1, 2, \dots$ ad inf.) von Punkten der Menge M_n . Es seien a_k und b_k Punkte in Abständen $\geq \frac{1}{n}$ voneinander, so daß p_k Zwischenpunkt von a_k und b_k ist. Wegen der vorausgesetzten Kompaktheit des Raumes existiert ein Punktepaar a, b , so daß für eine Teilfolge $\{a_{i_k}\}$ bzw. $\{b_{i_k}\}$ von $\{a_k\}$ bzw. $\{b_k\}$ $a = \lim_{k \rightarrow \infty} a_{i_k}$ und $b = \lim_{k \rightarrow \infty} b_{i_k}$ gilt. Wegen der Stetigkeit der Metrik folgt aus der Voraussetzung $a_{i_k} b_{i_k} \geq \frac{1}{n}$ ($k = 1, 2, \dots$ ad inf.), daß $ab \geq \frac{1}{n}$ gilt. Wegen der Stetigkeit der Metrik folgt weiter aus der Gültigkeit der Beziehungen $a_{i_k} p_{i_k} + p_{i_k} b_{i_k} = a_{i_k} b_{i_k}$ ($k = 1, 2, \dots$ ad inf.) die Beziehung $ap + pb = ab$. Es ist also p Zwischenpunkt von a und b und mithin Punkt der Menge M_n . Für jede natürliche Zahl n ist also die Menge M_n abgeschlossen. Jeder Punkt, welcher Zwischenpunkt von zwei Punkten des Raumes ist, ist Zwischenpunkt von zwei Punkten, die irgendeinen nicht-verschwindenden Abstand voneinander haben, ist also für irgendeine natürliche Zahl n Punkt der Menge M_n . Die Menge aller Punkte, welche Zwischenpunkte von irgend zwei Punkten des Raumes sind, ist also identisch mit der Summe der abzählbar vielen Mengen M_n , deren jede abgeschlossen ist, ist also ein F_σ . Die Menge aller Punkte, welche Zwischenpunkte von keinem Punktepaar des Raumes sind, ist das Komplement der Menge $\sum_{n=1}^{\infty} M_n$, also Produkt von abzählbar vielen offenen Mengen. Damit ist die Behauptung 4 bewiesen.

Analog beweist man die Behauptung 5. Es sei ein Punkt p und eine natürliche Zahl n vorgegeben. Die Menge $M_n(p)$ aller Punkte q des Raumes, für die ein Punkt r in einem Abstand $\geq \frac{1}{n}$ von p existiert, so daß q zwischen r und p liegt, ist wegen der Kompaktizität des Raumes und der Stetigkeit der Metrik abgeschlossen, woraus so wie beim Beweis der Behauptung 4 folgt, daß die Menge aller Punkte q , die zwischen p und irgendeinem Punkt des Raumes liegen, ein F_σ ist.

Damit ist unser Satz bewiesen. Wir wenden ihn nun auf die konvexen Räume an.

Je zwei Punkte a und b eines vollständigen konvexen Raumes bestimmen eine Geradenmenge $G(a, b)$, welche ein Semikontinuum ist. Die Abgeschlossenheit der Menge $G(a, b)$ gilt, wie wir sahen, für beliebige metrische Räume. Daß die Menge $\bar{Z}(a, b)$ in einem vollständigen konvexen Raum auch ein Semikontinuum ist, wurde oben bereits bewiesen. Es ist noch zu zeigen, daß für je zwei Punkte a und b eines vollständigen konvexen Raumes die Menge $\bar{Z}[a; b]$, bestehend aus a und allen Punkten c , so daß a zwischen b und c liegt, ein Semikontinuum ist. Es sei c ein von a verschiedener Punkt der Menge $\bar{Z}[a; b]$. Da der Raum vollständig und konvex ist, enthält er einen c und a verbindenden Teilbogen, bestehend aus Zwischenpunkten von a und c . Ist d Zwischenpunkt von a und c , so ist, da c Punkt von $\bar{Z}[a; b]$ ist und daher a zwischen b und c liegt, a auch Zwischenpunkt von b und d . Der aus Zwischenpunkten von a und c bestehende Bogen zwischen a und c enthält also nur Punkte der Menge $\bar{Z}[a; b]$. Die Menge $\bar{Z}[a; b]$ enthält demnach zu jedem ihrer Punkte einen ihn mit a verbindenden Teilbogen und daher zu je zweien ihrer Punkte einen sie verbindenden Bogen und ist also ein Semikontinuum. Die Abgeschlossenheit der Menge $\bar{Z}[a; b]$ wurde für beliebige metrische Räume nachgewiesen. Es ist also in einem vollständigen konvexen Raum die Menge $\bar{Z}[a; b]$ und ebenso natürlich die Menge $\bar{Z}[b; a]$ ein Semikontinuum. Die Menge $\bar{Z}(a; b)$ hat mit der Menge $\bar{Z}[a; b]$ den Punkt a , mit der Menge $\bar{Z}[b; a]$ den Punkt b gemein. Also ist die Menge $G(a, b)$ als Summe dieser drei Semikontinua ein Semikontinuum, wie behauptet.

Ist in einem vollständigen konvexen Raum der Punkt p Zwischenpunkt der Punkte a und b , dann und nur dann läßt sich der p und a verbindende geodätische Bogen über p hinaus nach b fortsetzen. In einem kompakten Raum existiert zu jedem vom Punkt a ausgehenden Bogen B ein extremer Punkt $e(B, a)$, über den hinaus sich der Bogen B von a aus nicht fortsetzen läßt. Dieser Punkt $e(B, a)$ ist offenbar extremer Punkt für alle geodätischen Bögen, welche ihn mit a verbinden.

Betrachten wir einen Punkt p und alle von p ausgehenden geodätischen Bögen! Auf jedem der Bögen existiert ein hinsichtlich p extremer Punkt. Die Menge aller dieser hinsichtlich p extremen Punkte bezeichnen wir als *die zum Punkt p gehörige Schale des Raumes*. Diese Menge ist offenbar identisch mit der Menge aller Punkte q , zu denen kein Punkt r existiert, so daß q zwischen p und r liegt. Nun haben wir oben als Behauptung 5 über die Verteilung metrischer Singularitäten nachgewiesen, daß für jeden Punkt p die Menge aller Punkte, die zwischen p und irgendeinem Punkt des Raumes liegen, ein F_σ ist. Die zum Punkt p gehörige Schale des Raumes ist offenbar das Komplement dieser Menge und daher ein G_δ . *Es ist also für jeden Punkt eines kompakten konvexen Raumes die zu ihm gehörige Schale ein G_δ , welches übrigens, wie wir nun an einem einfachen Beispiele zeigen wollen, nicht abgeschlossen sein muß.*

Im R_3 sei K eine den Nullpunkt enthaltende und in der XY -Ebene gelegene Kreislinie. Wir bezeichnen als Raum die Summe der beiden schiefen Kegel, welche durch K und die beiden Punkte mit den Koordinaten $x = y = 0, z = \pm 1$ bestimmt werden. Als Abstand je zweier Punkte dieses Raumes setzen wir den euklidischen Abstand der betreffenden Punkte fest. Die zum Punkt $x = y = 0, z = 1$ gehörige Schale des Raumes besteht offenbar aus dem Mantel des den Punkt nicht enthaltenden Kegels, ausgenommen die auf der Z -Achse gelegenen Punkte dieses Mantels, aber inklusive den Punkt $x = y = 0, z = -1$. Diese Menge ist offenbar nicht abgeschlossen. In dem betrachteten Raum ist übrigens auch die Menge aller Punkte, die auf der zu irgendeinem Punkt des Raumes gehörigen Schale liegen (die Menge aller Punkte, die Extrempunkte irgendeines in ihnen endenden Bogens sind), nicht abgeschlossen. Denn diese Menge besteht aus den Mänteln der beiden schiefen Kegel, ausgenommen die auf der Z -Achse gelegenen Zwischenpunkte der beiden Kegelspitzen.

Ist $e(B, a)$ Extrempunkt des Bogens B hinsichtlich des Punktes a , so kann der Fall eintreten, daß ein in $e(B, a)$ endendes Segment des Bogens B sich über den Punkt $e(B, a)$ hinaus geodätisch fortsetzen läßt. Betrachten wir beispielsweise eine Kreislinie mit Bogenmaß und einen Punkt p der Kreislinie, so läßt sich jeder Bogen zwischen p und q über q hinaus bis zu dem p gegenüberliegenden Punkt p' , aber nicht weiter, geodätisch fortsetzen. Hingegen läßt sich jedes in q endende Segment dieses Bogens über p' hinaus fortsetzen. Wenn der Punkt $e(B, a)$ extremer Punkt hinsichtlich *aller* Punkte des Bogens B ist, dann nennen wir ihn einen *Extrempunkt von B schlechthin*.

Jene Punkte des Raumes, welche für alle in ihnen endende Bögen Extrempunkte sind, nennen wir die *Fluchtpunkte* des Raumes. Die Fluchtpunkte eines konvexen Raumes sind offenbar identisch mit jenen Punkten,

die von keinem Punktepaar des Raumes Zwischenpunkte sind. Die Punkte des Raumes, welche nicht Fluchtpunkte sind, welche also zwischen zwei Punkten des Raumes und daher im Innern eines geodätischen Bogens liegen, nennen wir *Durchgangspunkte*. Nach der Behauptung 4 über die Verteilung von metrischen Singularitäten *ist die Menge der Fluchtpunkte eines kompakten konvexen Raumes ein G_δ , die Menge der Durchgangspunkte ein F_σ .*

In einem vollständigen konvexen Raum ist die Menge aller Durchgangspunkte ein im kleinen zusammenhängendes, konvexes und im Raume überall dicht liegendes Semikontinuum. Es seien p und q zwei Durchgangspunkte eines vollständigen konvexen Raumes. Der Raum enthält einen geodätischen Bogen zwischen p und q , dessen Länge gleich dem Abstand pq ist. Jeder innere Punkt dieses Bogens ist offenbar ein Durchgangspunkt. Die Menge aller Durchgangspunkte enthält also zu je zweien ihrer Punkte einen sie verbindenden Bogen, ist mithin ein Semikontinuum. Sie enthält zu hinreichend benachbarten Punkten beliebig kurze Bögen, ist also zusammenhängend im kleinen; sie ist konvex, weil sie zu je zweien ihrer Punkte einen sie verbindenden geodätischen Bogen enthält. Ist p irgendein Punkt des Raumes, so gibt es sicher einen in p endenden geodätischen Bogen, dessen sämtliche innere Punkte Durchgangspunkte sind. Jeder Punkt des Raumes ist also Häufungspunkt von Durchgangspunkten, d. h. die Menge aller Durchgangspunkte liegt im Raume dicht. Damit ist unser Satz bewiesen.

Die Menge der Fluchtpunkte des Raumes muß übrigens weder abgeschlossen noch ein F_σ sein und kann in einem kompakten konvexen Raum unter Umständen ebenfalls dicht liegen. Betrachten wir das Beispiel 8 meiner Grundzüge einer Theorie der Kurven (Math. Annalen 95, S. 285f.). Es beschreibt eine kompakte Kurve, in welcher die Endpunkte dicht liegen (wobei ich als Endpunkt eines Kontinuums einen Punkt bezeichne, zu dem beliebig kleine Umgebungen existieren, deren Begrenzungen genau einen Punkt enthalten). Setzen wir nun in der a. a. O. beschriebenen Kurve als Abstand je zweier Punkte die euklidisch gemessene Länge des kürzesten die beiden Punkte verbindenden Teilbogens dieser Kurve fest, dann wird die erwähnte Kurve ein konvexer kompakter Raum. Jeder Endpunkt eines kompakten konvexen Raumes ist Fluchtpunkt. Andererseits sieht man sofort, daß in unserem Raum die Endpunkte die einzigen Fluchtpunkte sind. Sie bilden eine im Raum überall dicht liegende Menge, welche ein G_δ , aber kein F_σ ist. Es sei hier auf die Analogie hingewiesen, die überhaupt in metrischen Räumen zwischen Endpunkten und Fluchtpunkten und zwischen den Durchgangspunkten und den Punkten höherer Ordnung in vieler Hinsicht besteht.

11. Über Räume mit Strecken.

In euklidischen Räumen gilt bekanntlich der Satz, daß der Durchschnitt zweier (und beliebig vieler) abgeschlossener konvexer Mengen konvex ist. Wir untersuchen die diesbezüglichen Verhältnisse in allgemeinen metrischen Räumen.

Damit in einem vollständigen metrischen Raum der Durchschnitt von je zwei abgeschlossenen konvexen Mengen konvex sei, ist notwendig und hinreichend, daß der Raum zu je zweien seiner Punkte höchstens einen sie verbindenden geodätischen Bogen enthält. Es seien erstens K_1 und K_2 zwei abgeschlossene konvexe Mengen eines Raumes, der zu je zweien seiner Punkte höchstens einen sie verbindenden geodätischen Bogen enthält. Es seien p und q irgend zwei Punkte des Durchschnittes $K_1 K_2$. Da p und q in K_1 liegen und K_1 konvex und abgeschlossen ist, so enthält K_1 wegen der Vollständigkeit des Raumes nach dem Existenzsatz von Abschnitt 6 einen p und q verbindenden geodätischen Bogen. Ebenso enthält K_2 einen p und q verbindenden geodätischen Bogen, und da der Raum nach Voraussetzung mehr als einen geodätischen Bogen zwischen p und q nicht enthält, so müssen diese zwei geodätischen Bögen zwischen p und q identisch sein, also im Durchschnitt $K_1 K_2$ liegen. Je zwei Punkte dieses Durchschnittes sind also durch einen geodätischen Teilbogen des Durchschnittes verbunden, womit die Konvexität des Durchschnittes bewiesen ist. — Sei zweitens R ein Raum, der zu zweien seiner Punkte, etwa zu den Punkten p und q zwei nicht identische, sie verbindende geodätische Bögen B und B' enthält. Da B und B' nicht identisch sind, so enthält einer der beiden Bögen, etwa B , einen Punkt b , welcher nicht in B' liegt. Da die Endpunkte von B und B' den beiden Bögen gemeinsam sind, ist b innerer Punkt von B . Bewegt man sich von b längs B nach rechts, so trifft man einen ersten Punkt, welcher auch zu B' gehört — derselbe heiße f —, und ebenso trifft man von b längs B nach links laufend einen ersten Punkt, welcher zu B' gehört, derselbe heiße g . Es sei \bar{B} der abgesehen von seinen Endpunkten zu B' fremde Abschnitt von B zwischen f und g ; und es sei \bar{B}' der abgesehen von seinen Endpunkten zu \bar{B} fremde Abschnitt von B' zwischen f und g . Die beiden Bögen \bar{B} und \bar{B}' sind zwischen f und g geodätisch, also konvexe Mengen; ihr Durchschnitt ist die aus den Punkten f und g bestehende Menge, also nicht konvex. Damit ist unser Satz bewiesen.

In einem Raum, in dem zu je zwei Punkten genau ein sie verbindender geodätischer Bogen existiert, ist der Durchschnitt von einem beliebigen System S von abgeschlossenen konvexen Mengen konvex. Wenn jede Menge des Systems S zu je zwei Punkten des Durchschnitts aller Mengen aus S den einzigen die beiden Punkte verbindenden geodätischen Bogen enthält, so enthält dieser Durchschnitt selbst zu je zweien seiner Punkte den sie verbindenden geodätischen Bogen und ist daher konvex.

In einem konvexen vollständigen Raum, in dem zu je zwei Punkten genau ein sie verbindender geodätischer Bogen existiert, gibt es zu jeder Menge M genau eine konvexe Hülle, d. h. eine kleinste M enthaltende abgeschlossene konvexe Menge. Der Durchschnitt aller M enthaltenden ab-

geschlossenen konvexen Mengen ist nämlich nach dem Bewiesenen eine abgeschlossene konvexe Menge, welche M enthält, und offenbar existiert kein echter abgeschlossener M enthaltender konvexer Teil dieser Menge^{7a)}.

Existiert in einem konvexen Raum zu je zwei Punkten genau ein sie verbindender geodätischer Bogen, dann nennen wir denselben die durch die betreffenden Punkte bestimmte *Strecke*. Wir geben nun einige Kennzeichnungen jener konvexen Räume an, in denen je zwei Punkte eine Strecke bestimmen.

Damit in einem vollständigen konvexen Raum je zwei Punkte eine Strecke bestimmen, ist notwendig und hinreichend, daß für je zwei Punkte a und b des Raumes die Menge $\bar{Z}(a, b)$ ein einfacher Bogen sei, oder, was gleichbedeutend ist, daß zu je zwei Punkten genau ein Mittelpunkt existiert.

Wenn *erstens* für zwei Punkte a und b eines vollständigen konvexen Raumes die Menge $\bar{Z}(a, b)$ ein einfacher Bogen ist, dann ist dieser Bogen offenbar der einzige geodätische Bogen zwischen den Punkten a und b , da ja ein geodätischer Bogen aus Zwischenpunkten seiner Endpunkte besteht. Dann besitzen ferner die Punkte a und b nur einen Mittelpunkt, nämlich den auf den Bogen $\bar{Z}(a, b)$ gelegenen Mittelpunkt, da es ja außerhalb dieses Bogens überhaupt keine Zwischenpunkte von a und b gibt. Wir nehmen *zweitens* an, für zwei Punkte a und b eines vollständigen konvexen Raumes sei die Menge $\bar{Z}(a, b)$ kein einfacher Bogen. Jedenfalls existiert ein geodätischer Bogen B zwischen a und b . Aus der Annahme folgt, daß ein nicht in B enthaltener Zwischenpunkt c von a und b existiert. Der Punkt c ist sowohl mit a als auch mit b durch einen geodätischen Bogen verbunden und die Summe dieser beiden geodätischen Bögen ist ein geodätischer Bogen B' zwischen a und b , welcher von B verschieden ist, da er den Punkt c enthält. Aus der Annahme folgt also die Existenz von mindestens zwei verschiedenen geodätischen Bögen zwischen a und b . — Laufen wir auf dem Bogen B' von c aus nach rechts, so treffen wir auf einen ersten Punkt, der auch zu B gehört, derselbe heiße f . Desgleichen existiert ein erster Punkt g von B auf dem Bogen B' , wenn man diesen von c aus nach links durchläuft. Die Punkte f und g sind durch zwei bis auf die Endpunkte fremde geodätische Bögen verbunden, besitzen also zwei verschiedene Mittelpunkte. Damit ist unsere Behauptung in allen Teilen bewiesen.

Damit in einem vollständigen konvexen Raum je zwei Punkte eine Strecke bestimmen, ist notwendig und hinreichend, daß für je zwei Punkte a und b des Raumes und für je zwei verschiedene Zwischenpunkte p und q von a und b der Punkt p entweder zwischen a und q oder zwischen

^{7a)} [Zusatz bei der Korrektur:] Allgemeinere diesbezügliche Sätze habe ich in der Notiz „Über konvexe Hüllen“ (Wiener akad. Anz. 1928, Nr. 11) bewiesen.

b und q liegt. Die Bedingung ist notwendig. Denn wenn in einem vollständigen konvexen Raum je zwei Punkte eine Strecke bestimmen, so ist für je zwei Punkte a und b die Menge $\bar{Z}(a, b)$ ein geodätischer Bogen zwischen a und b und für je zwei verschiedene innere Punkte p und q eines geodätischen Bogens zwischen a und b gilt eine der zwei angeführten Relationen. — Die Bedingung ist hinreichend. Dazu haben wir zu zeigen: Sind a und b zwei Punkte eines vollständigen konvexen Raumes, so daß für je zwei verschiedene Zwischenpunkte p und q von a und b eine der beiden angeführten Relationen besteht, dann ist die Menge $\bar{Z}(a, b)$ ein Bogen (und mithin der einzige geodätische Bogen zwischen a und b). Jedenfalls existiert, da der Raum vollständig und konvex ist, ein geodätischer Bogen B zwischen a und b . Angenommen, es existiere ein nicht in B enthaltener Zwischenpunkt p von a und b . Es gibt auf B einen einzigen Punkt q , so daß $\frac{aq}{bq} = \frac{ap}{aq}$ gilt. Die Punkte p und q wären verschiedene Zwischenpunkte von a und b und es wäre p weder Zwischenpunkt von a und q , noch von b und q im Widerspruch gegen die Voraussetzung. Damit ist die Behauptung bewiesen.

12. Über Räume mit Geraden.

Der metrische Zwischenbegriff ermöglicht nicht nur eine Beschreibung metrischer Singularitäten, sondern gibt auch Mittel an die Hand, um sukzessive von den allgemeinen metrischen Räumen zu den Räumen der Elementargeometrie zu gelangen. Insbesondere das elementargeometrische Axiom, daß je zwei verschiedene Punkte eine Gerade bestimmen, erfährt von unseren Gesichtspunkten aus weitgehende Aufspaltungen. Zur Durchführung der diesbezüglichen Analyse definieren wir zunächst einen Hilfsbegriff.

Wenn von drei nicht durchwegs identischen Punkten a, b, c einer entweder mit einem anderen identisch ist oder Zwischenpunkt von den beiden anderen ist, dann wollen wir sagen, *die Punkte a, b, c liegen auf einer Geraden* und schreiben hierfür $\underline{a, b, c}$. (In euklidischen Räumen liegen drei Punkte im angeführten Sinn auf einer Geraden offenbar dann und nur dann, wenn sie im gewöhnlichen Sinn auf einer Geraden liegen.) Unter Zugrundelegung der Symbolik von Abschnitt 2 ist die Relation $\underline{a, b, c}$ äquivalent damit, daß nicht $a = b = c$ gilt und eine der sechs Relationen $a = b, b = c, a = c, abc, bca, cab$ erfüllt ist. Die Relation $\underline{a, b, c}$ genügt daher folgenden Bedingungen: 1. Gilt für irgend drei Punkte a_1, a_2, a_3 $\underline{a_1, a_2, a_3}$, dann gilt auch $\underline{a_{i_1}, a_{i_2}, a_{i_3}}$, wenn i_1, i_2, i_3 irgendeine Permutation der Ziffern 1, 2, 3 darstellt. 2. Für je zwei verschiedene Punkte a und b des Raumes gilt $\underline{a, b, b}$.

Die euklidischen Räume haben folgende für ihre Metrik fundamentale Eigenschaft: Wenn von irgend vier paarweise verschiedenen Punkten des Raumes zwei Tripel auf einer Geraden liegen, dann liegen auch die beiden anderen Tripel auf einer Geraden. Die Konsequenzen dieser Eigenschaft werden wir am Ende dieses Abschnittes betrachten. Zunächst untersuchen wir metrische Räume, welche folgende schwächere (den euklidischen Räumen demnach gleichfalls zukommende) Eigenschaft besitzen:

Dreitripleigenschaft. Wenn von vier paarweise verschiedenen Punkten des Raumes drei Tripel auf einer Geraden liegen, dann liegt auch das vierte Tripel auf einer Geraden.

Gehen wir auf die Definition des Begriffes „drei Punkte liegen auf einer Geraden“ zurück, so besagt die Dreitripleigenschaft: Sind a, b, c, d irgend vier paarweise verschiedene Punkte des Raumes, für welche

eine der Relationen $I_1 abc, I_2 acb, I_3 bac,$

eine der Relationen $II_1 abd, II_2 adb, II_3 bad,$

eine der Relationen $III_1 acd, III_2 adc, III_3 cad$

gilt, dann gilt

eine der Relationen $IV_1 bcd, IV_2 bdc, IV_3 cbd.$

Nun verifiziert man sofort, daß in einem beliebigen metrischen Raum auf Grund der Eigenschaft 3 der Zwischenbeziehung gilt:

Aus dem Zusammenbestehen von	I_1, II_1, III_1	folgt	IV_1
” ” ”	I_1, II_1, III_2	”	IV_2
” ” ”	I_1, II_2, III_2	”	IV_3
” ” ”	I_1, II_3, III_3	”	IV_3
” ” ”	I_2, II_1, III_1	”	IV_3
” ” ”	I_2, II_2, III_1	”	IV_2
” ” ”	I_2, II_2, III_2	”	IV_1
” ” ”	I_2, II_3, III_3	”	IV_1
” ” ”	I_3, II_1, III_3	”	IV_3
” ” ”	I_3, II_2, III_3	”	IV_2
” ” ”	I_3, II_3, III_1	”	IV_1
” ” ”	I_3, II_3, III_2	”	IV_2

Ebenso ergibt sich auf Grund der Eigenschaft 3 der Zwischenbeziehung in beliebigen metrischen Räumen als

widerspruchsvoll das Zusammenbestehen von	I_1, II_1, III_3
" " " "	I_1, II_2, III_1
" " " "	I_1, II_3, III_1
" " " "	I_2, II_1, III_2
" " " "	I_2, II_1, III_3
" " " "	I_2, II_3, III_1
" " " "	I_2, II_3, III_2
" " " "	I_3, II_1, III_2
" " " "	I_3, II_2, III_1
" " " "	I_3, II_2, III_2

Von den 27 Kombinationen des Zusammenbestehens von einer der Relationen I_i , einer der Relationen II_k und einer der Relationen III_l bleiben also nur vier, in denen die für den Raum postulierte Dreitripleeigenschaft eine tatsächliche Forderung ausspricht, nämlich die Fälle, in welchen zusammenbesteht

das Relationentripel α :	I_1, II_3, III_2
" " β :	I_2, II_2, III_3
" " γ :	I_3, II_1, III_1
" " δ :	I_3, II_3, III_3 .

Wir betrachten zunächst die Tripel α, β, γ . Man sieht, daß auf Grund der Eigenschaft 3 der Zwischenbeziehung mit dem Relationentripel α die Relation IV_2 unverträglich ist. Denn aus IV_2 und II_3 würde folgen bac im Widerspruch zu I_1 . Aus analogen Gründen ist IV_3 mit dem Relationentripel α unverträglich und man zeigt ebenso: In beliebigen metrischen Räumen ist von den drei Relationen IV_i

mit dem Relationentripel α höchstens IV_1 verträglich,
" " " β " IV_3 "
" " " γ " IV_2 "

Die Dreitripleeigenschaft des Raumes besagt also, angewendet auf vier Punkte, zwischen denen die Relationentripel α bzw. β bzw. γ erfüllt sind, daß für sie IV_1 bzw. IV_3 bzw. IV_2 erfüllt ist. Da die Relationentripel α, β, γ bei Permutation der Bezeichnung der Punkte a, b, c, d identisch sind, besagt also die Dreitripleeigenschaft: Für je vier Punkte a, b, c, d des Raumes folgt aus abd, acd, bac die Relation bdc .

Wir zeigen nun: *Ein konvexer Raum, welcher vier Punkte a, b, c, d enthält, für welche die Relationen abd, acd, bac, bdc zusammenbestehen,*

enthält auch vier Punkte a', b', c', d' , für welche bloß die Relationen $a'c'd', b'a'c', b'd'c'$, aber nicht die Relation $a'b'd'$ gilt. Zunächst folgt aus den vorausgesetzten Beziehungen

$$ab + bd = ad, \quad ac + cd = ad, \quad ab + ac = bc, \quad bd + cd = bc$$

offenbar:

$$ad = bc, \quad ac = bd, \quad ab = cd.$$

Da ferner der Raum als konvex vorausgesetzt ist, können wir einen Zwischenpunkt von c und d wählen. Wir bezeichnen denselben mit d' . Überdies setzen wir $a' = a, b' = b, c' = c$. Es gilt dann offenbar $b'a'c'$. Ferner ist d' Zwischenpunkt von c und d , also gilt $a'c'd'$ und aus analogen Gründen gilt $b'd'c'$. Hingegen gilt nicht $a'b'd'$. Es gilt ja $b'd' = bd + dd', a'd' = ad' = ac + cd' = bd + cd'$. Daher ist

$$a'b' + b'd' = ab + bd + dd' > a'd' = bd + cd',$$

denn es ist $ab = cd > cd'$. Es gilt also nicht $a'b'd'$, womit unsere Behauptung bewiesen ist. Aus ihr folgt:

In einem konvexen Raum, welcher die Dreitripleigenschaft besitzt, existieren keine Punktequadrupel, für welche die Relationen abd, acd, bac zusammenbestehen. In der Tat, wenn in einem konvexen Raum ein Quadrupel existiert, für welches diese drei Relationen zusammenbestehen, dann existiert ja im Raum auch ein Quadrupel, für welches diese Relationen zusammenbestehen, ohne daß abd erfüllt wäre, was der Dreitriple-relation widersprechen würde. —

Wir betrachten nun vier Punkte, a, b, c, d , für welche das Relationentriplel δ erfüllt ist. Die Dreitripleigenschaft des Raumes fordert, daß für die vier Punkte eine der drei Relationen IV_1, IV_2, IV_3 gilt. Nun würde aber aus dem Zusammenbestehen von IV_1 und I_3 nach der Eigenschaft 3 der Zwischenbeziehung die Relation acd folgen im Widerspruch zu III_3 . Es ist IV_1 mit dem Zusammenbestehen von I_3, II_3, III_3 unverträglich und in derselben Weise zeigt man, daß die Relationen IV_2 und IV_3 mit dem Relationentriplel δ unverträglich sind. In einem Raum, welcher die Dreitripleigenschaft besitzt, kann also ein Punktequadrupel, welches die Relationen des Tripels δ erfüllt, nicht existieren. Das bisher Bewiesene zusammenfassend, können wir sagen: *Unter den konvexen Räumen sind jene, welche die Dreitripleigenschaft besitzen, dadurch gekennzeichnet, daß sie keine vier Punkte a, b, c, d enthalten, für welche, sei es die Relationen abd, acd, bac , sei es die Relationen bac, bad, cad zusammenbestehen, für die m. a. W. eines der Relationentriplel $\alpha, \beta, \gamma, \delta$ gilt.*

Es sei nun ein vollständiger konvexer Raum vorgelegt, der vier Punkte a, b, c, d enthält, für welche die Relationen abd, acd, bac zu-

sammenbestehen. Wir betrachten geodätische Bögen $B_{ab}, B_{ac}, B_{bd}, B_{cd}$ zwischen bzw. a und b , a und c , b und d , c und d . Da sowohl b als auch c Zwischenpunkt von a und d ist, ist sowohl $B_{ab} + B_{bd}$ als auch $B_{ac} + B_{cd}$ ein geodätischer Bogen zwischen a und d . Wir nennen B den ersteren, B' den letzteren dieser beiden geodätischen Bögen zwischen a und d und zeigen, daß die beiden nicht identisch sind. Dazu genügt es, nachzuweisen, daß der in B' liegende Punkt c nicht in B enthalten ist. Aus bac folgt, daß $c \neq b$ und daß c nicht Punkt von B_{ab} sein kann. Der Punkt c kann auch nicht in B_{bd} liegen, denn aus bcd und abd würde folgen abc , während doch bac vorausgesetzt ist. Es ist also c nicht Punkt von B und mithin $B \neq B'$. Der Raum enthält also zwei nicht identische geodätische Bögen zwischen a und d . Daraus folgt: *Wenn in einem vollständigen konvexen Raum je zwei Punkte eine Strecke bestimmen, dann enthält der Raum keine vier Punkte a, b, c, d , für welche die Relationen abd, acd, bac zusammenbestehen, und mithin kein Punktequadrupel, für welches eines der Relationentripel α, β, γ gilt.*

Es sei nun ein vollständiger konvexer Raum vorgelegt, der vier Punkte a, b, c, d enthält, für welche die Relationen bac, bad, cad zusammenbestehen, vier Punkte also, unter denen einer zwischen je zweien von den übrigen drei liegt. Wir betrachten dann drei geodätische Bögen B_{ab}, B_{ac}, B_{ad} zwischen a und bzw. den Punkten b, c, d . Da a Zwischenpunkt von b und c ist, ist $B_{ab} + B_{ac}$ ein geodätischer Bogen zwischen b und c . Ebenso ist $B_{ab} + B_{ad}$ ein geodätischer Bogen zwischen b und d und ist $B_{ac} + B_{ad}$ ein geodätischer Bogen zwischen c und d . Die Menge $D = B_{ab} + B_{ac} + B_{ad}$ enthält daher offenbar zu je zweien ihrer Punkte einen sie verbindenden geodätischen Bogen, ist also konvex. Die Menge D ist Summe von drei Bögen (nämlich von B_{ab}, B_{ac}, B_{ad}), die paarweise nur einen allen dreien gemeinsamen Endpunkt (nämlich a) gemein haben. Eine derartige Menge wollen wir ein *Dreibein* nennen (speziell unsere Menge D ein Dreibein zwischen b, c, d mit dem Scheitel a). Enthält umgekehrt ein Raum ein konvexes Dreibein, etwa zwischen den Punkten b, c, d mit dem Scheitel a , dann liegt der Punkt a zwischen je zweien von den Punkten b, c, d . *Ein vollständiger konvexer Raum enthält also vier Punkte, für welche das Relationentripel δ erfüllt ist, dann und nur dann, wenn er ein konvexes Dreibein enthält.*

Eine Zusammenfassung der bisherigen Resultate dieses Abschnittes ergibt den Satz:

Unter den vollständigen konvexen Räumen, in denen je zwei Punkte eine Strecke bestimmen, sind diejenigen, welche die Dreitripel-eigenschaft besitzen, dadurch gekennzeichnet, daß sie kein konvexes Dreibein enthalten.

In der Tat, ein konvexer vollständiger Raum mit Strecken, der die Dreitripleigenschaft besitzt, enthält keine vier Punkte, welche dem Relationentripel δ genügen, und daher kein konvexes Dreibein. Umgekehrt enthält ein Raum mit Strecken keine vier Punkte, welche dem Relationentripel α, β, γ genügen, und, wofern er kein konvexes Dreibein enthält und konvex und vollständig ist, auch keine vier Punkte, für welche das Relationentripel δ erfüllt ist. Er besitzt demnach die Dreitripleigenschaft.

Wir können diesem Satz noch eine andere Form geben. Ein konvexes Dreibein ist nämlich eine konvexe Menge, die kein geodätischer Bogen, aber Summe von zwei geodätischen Bögen ist. Ein konvexes Dreibein zwischen den Punkten b, c, d mit dem Scheitel a ist beispielsweise Summe eines geodätischen Bogens zwischen b und c und eines geodätischen Bogens zwischen a und d . Enthält umgekehrt ein Raum, in dem je zwei Punkte eine Strecke bestimmen, eine konvexe Menge, welche keine Strecke, aber Summe von zwei Strecken ist, dann enthält D , wie wir hier ohne Beweis erwähnen wollen, ein konvexes Dreibein als Teil. *Unter den vollständigen konvexen Streckenräumen sind daher jene, welche die Dreitripleigenschaft besitzen, auch dadurch gekennzeichnet, daß in ihnen die Strecken die einzigen konvexen Mengen sind, welche sich als Summe zweier Strecken darstellen lassen.*

Wenn in einem vollständigen konvexen Raum je zwei Punkte eine Strecke bestimmen und der Raum sogar überdies die Dreitripleigenschaft besitzt, so ist hierdurch doch nicht gewährleistet, daß jede Strecke, welche über einen ihrer Endpunkte hinaus geodätisch fortsetzbar ist, über diesen Punkt hinaus *eindeutig* geodätisch fortsetzbar ist. Nehmen wir beispielsweise als Raum die Menge aller Punkte der Ebene, für die sei es $x \geq 0, y \geq 0$, sei es $x = 0$ gilt (also die Y -Achse und ein Viertel der Ebene). Als Abstand setzen wir fest für je zwei Punkte des Ebenenviertel und für je zwei Punkte der Y -Achse den euklidischen Abstand, für je zwei Punkte x, y und x', y' , für die $y > 0, y' < 0$ gilt, den Wert $-y' + \sqrt{x^2 + y^2}$. Der so definierte metrische Raum ist vollständig und konvex. Je zwei seiner Punkte bestimmen eine Strecke und er enthält kein konvexes Dreibein. Hingegen ist jede Strecke durch den Nullpunkt, welche in dem zum Raum gehörigen Ebenenviertel liegt, geodätische Fortsetzung von der Strecke zwischen den Punkten $x = 0, y = -1$ und $x = y = 0$, über den Nullpunkt hinaus. Wir beweisen nun:

• *Damit in einem vollständigen konvexen Raum je zwei verschiedene Punkte eine Geradenmenge bestimmen, die durch je zwei verschiedene ihrer Punkte bestimmt ist, ist notwendig und hinreichend, daß der Raum die folgende Zweitripleigenschaft besitzt: Wenn von irgend vier paarweise verschiedenen Punkten zwei Tripel auf einer Geraden liegen, dann liegen auch die beiden anderen Tripel auf einer Geraden.*

Wenn für je zwei verschiedene Punkte a und b eines vollständigen

konvexen Raumes die Geradenmenge $G(a, b)$ durch je zwei verschiedene ihrer Punkte bestimmt ist, dann nennen wir $G(a, b)$ die durch a und b bestimmte Gerade und jede durch zwei Punkte des Raumes bestimmte Geradenmenge eine Gerade. *Erstens* ist klar, daß ein Raum mit Geraden die Zweitripleigenschaft besitzt. Es seien nämlich a, b, c, d vier verschiedene Punkte des Raumes, so daß $\overline{a, b, c}$ und $\overline{a, b, d}$ gilt. Wir haben nachzuweisen, daß $\overline{a, c, d}$ und $\overline{b, c, d}$ gilt. Da c und d zwei verschiedene Punkte von $G(a, b)$ sind, so bestimmen c und d nach Voraussetzung diese Gerade, d. h. es ist $G(c, d) = G(a, b)$. Es enthält also $G(c, d)$ sowohl a als auch b , d. h. es gilt $\overline{a, c, d}$ und $\overline{b, c, d}$. *Zweitens* ist die Zweitripleigenschaft für die Existenz von Geraden hinreichend. Seien nämlich a und b zwei verschiedene Punkte und seien c und d zwei verschiedene Punkte von $G(a, b)$. Wir haben nachzuweisen, daß aus der Zweitripleigenschaft folgt $G(a, b) = G(c, d)$. Dazu haben wir zu zeigen: Wenn p ein Punkt von $G(a, b)$ ist, dann liegt p auch in $G(c, d)$ und umgekehrt. Es sei *erstens* p ein Punkt von $G(a, b)$, es gelte also $\overline{a, b, p}$. Nach Voraussetzung gilt $\overline{a, b, c}$ und $\overline{a, b, d}$. Wegen der Zweitripleigenschaft des Raumes folgt aus $\overline{a, b, p}$ und $\overline{a, b, c}$ die Beziehung $\overline{a, c, p}$ und ebenso folgt $\overline{a, d, p}$ aus $\overline{a, b, p}$ und $\overline{a, b, d}$. Aus $\overline{a, c, p}$ und $\overline{a, d, p}$ folgt $\overline{p, c, d}$, d. h. p ist Punkt von $G(c, d)$. Sei *zweitens* p Punkt von $G(c, d)$; es gilt also $\overline{p, c, d}$. Da nach Voraussetzung $\overline{a, b, c}$ und $\overline{a, b, d}$ gilt, gilt $\overline{a, b, d}$ und $\overline{b, c, d}$, woraus man ganz wie im ersten Fall $\overline{a, b, p}$ herleitet. Damit ist unsere Behauptung bewiesen.

Analysieren wir unter Zugrundelegung der obigen Bezeichnungen die Zweitripleigenschaft, so sehen wir, daß sie folgendes bedeutet: Für je vier Punkte a, b, c, d des Raumes folgt aus dem Zusammenbestehen von einer der Relationen I_i und II_j die Gültigkeit von einer Relation III_k und IV_l . Nun ergibt sich auf Grund der Eigenschaft 3 der Zwischenbeziehung für jeden metrischen Raum:

Aus dem Zusammenbestehen von	I_1	und	II_2	folgen	III_2	und	IV_3 .
" " "	"	"	I_2	"	II_1	"	III_1 " IV_3 .
" " "	"	"	I_3	"	II_3	"	III_3 " IV_1 .
" " "	"	"	I_3	"	II_2	"	III_3 " IV_2 .

Berücksichtigt man die bewiesene Unmöglichkeit des Zusammenbestehens gewisser Relationentripel I_i, II_j, III_k und bedenkt man, daß die Zweitripleigenschaft die Dreitripleigenschaft impliziert und daß demnach auch in einem konvexen Raum mit der Zweitripleigenschaft keine vier Punkte existieren können, für welche eines der Relationentripel $\alpha, \beta, \gamma, \delta$ erfüllt ist, so sehen wir, daß die Zweitripleigenschaft fordert:

- Aus I_1 und II_1 folgt entweder III_1 und IV_1 oder III_2 und IV_2 .
 „ I_2 „ II_3 „ III_3 und IV_1 .
 „ I_2 „ II_2 „ entweder III_1 und IV_2 oder III_2 und IV_1 .
 „ I_3 „ II_1 „ III_3 und IV_1 .
 „ I_3 „ II_3 „ entweder III_1 und IV_1 oder III_2 und IV_2 .

Da die erste und die letzte und ebenso die zweite und die vierte dieser Forderungen durch Wechsel in der Bezeichnung der Punkte a, b, c, d aufeinander zurückführbar sind, fordert also die Zweitripleigenschaft für je vier Punkte des Raumes:

1. Aus acb und adb folgt stets entweder adc und dcb oder acd und cdb . Diese Forderung ist, wie wir wissen, in vollständigen konvexen Räumen gleichbedeutend mit der Forderung, daß je zwei Punkte des Raumes eine Strecke bestimmen.

2. Aus abc und abd folgt stets entweder acd und bcd oder adc und bdc . Diese Forderung ist eine der Forderung 1 analoge Bedingung, welche offenbar mit der *Eindeutigkeit* der geodätischen Fortsetzbarkeit von geodätisch fortsetzbaren geodätischen Bögen äquivalent ist.

3. Aus abc und bad folgt dac und dbc . Es ist dies eine der elementargeometrischen Zwischenrelationen (vgl. oben S. 79), welche unserem Zwischenbegriff im allgemeinen nicht zukommt.

Auf die Beziehungen der dritten Forderung zu den zwei ersten wollen wir nicht mehr eingehen. Wir erwähnen hier nur noch, daß unsere Überlegungen insbesondere eine *Charakterisierung der sogenannten linearen Räume* unter den metrischen ermöglichen, wie gelegentlich an anderer Stelle ausgeführt werden soll.

Zweite Untersuchung: Die euklidische Metrik.

Vorbemerkung.

1. Problemstellung.
2. Zurückführung des Problems auf die Einbettung von $n + 3$ Punkten in den R_n .
3. Über merkwürdige Systeme von $n + 3$ Punkten.
4. Eine notwendige Bedingung für die Einbettbarkeit von $n + 2$ Punkten in den R_n .
5. Über Systeme von $n + 3$ Punkten, von denen je $n + 2$ in den R_n abstandstreu einbettbar sind. — Die pseudo-euklidischen Systeme.
6. Über Systeme, welche mehr als $n + 3$ Punkte enthalten.
7. Ein Kriterium für die Einbettbarkeit von $n + 3$ Punkten in den R_n .
8. Zurückführung des Problems auf die Einbettung von $n + 1$ Punkten in den R_n .

9. Eine notwendige und hinreichende Bedingung für die Einbettbarkeit von $n + 1$ Punkten in den R_n .
10. Notwendige und hinreichende Bedingungen für die Einbettbarkeit eines Raumes in den R_n .
11. Charakterisierung des R_n , des Halb- R_n und der euklidischen konvexen Körper.
12. Ausblicke.

Vorbemerkung.

Herr M. Biedermann in Amsterdam hat folgenden Satz gefunden: *Wenn von je drei Punkten eines zusammenhängenden metrischen Raumes einer zwischen den beiden anderen liegt, dann ist der Raum homöomorph entweder mit der Strecke oder mit der Halbgeraden oder mit der Geraden.* Dabei sagt Herr Biedermann im Anschluß an meine Arbeit „Über geodätische Linien in allgemeinen metrischen Räumen“ (Proc. Ac. Amsterdam, 29, 1926, S. 166), der Punkt b liege zwischen den Punkten a und c , wenn $ab + bc = ac$ gilt, wobei xy den Abstand der Punkte x und y bezeichnet. Der kurze Beweis von Herrn Biedermann lautet so: Ist R ein metrischer Raum, in dem von je drei Punkten einer zwischen den beiden anderen liegt, dann existieren zu jedem Punkt a von R und zu jeder positiven Zahl r höchstens zwei Punkte von R , die von a den Abstand r haben. Angenommen nämlich, es gäbe drei verschiedene Punkte b, c, d , die vom Punkt a den Abstand r haben! Dann betrachten wir die Punkte a, b, c . Nach Voraussetzung liegt einer von ihnen zwischen den beiden anderen. Es kann nicht b zwischen a und c liegen, denn wegen $ab = ac$ würde aus der Relation $ba + bc = ac$ folgen, daß $bc = 0$ ist, während doch die Punkte b und c als verschieden vorausgesetzt wurden. Ebenso kann nicht c zwischen a und b liegen. Es muß also a zwischen b und c liegen. Also gilt $bc = ab + bc = 2r$. Dieselbe Überlegung kann man mit den Punktetripeln a, b, d und a, c, d anstellen und erhält so $bd = 2r$ und $cd = 2r$. Wegen $bc = bd = cd > 0$ sind aber b, c, d drei Punkte, von denen keiner zwischen den beiden anderen liegt. Die Annahme der Existenz von b, c, d führt also zu einem Widerspruch gegen die Voraussetzung. Es existieren zu jedem Punkt a von R und zu jeder positiven Zahl r höchstens zwei Punkte in einem Abstand r von a , m. a. W. die Begrenzung jeder Kugelumgebung in R enthält höchstens zwei Punkte. Zu jedem Punkt von R existieren beliebig kleine Umgebungen, deren Begrenzungen höchstens zwei Punkte enthalten, also ist R nach einem kurventheoretischen Satz (vgl. Math. Annalen 95, S. 303) homöomorph entweder mit einem Kreis oder mit einer zusammenhängenden Teilmenge der Geraden. Da sich in einem metrischen Raum, der mit einem Kreis homöomorph ist, stets drei Punkte angeben lassen, von denen keiner zwischen den beiden anderen liegt, so bleibt also nur die zweite Möglichkeit, womit der Biedermannsche Satz bewiesen ist. Zugleich warf Herr Biedermann die Frage auf, ob sich nicht durch Entfernungsrelationen zwischen mehreren Punkten auch die höherdimensionalen euklidischen Räume topologisch charakterisieren ließen.

Es schien mir zunächst, daß ein beliebiger metrischer Raum durch die Eigenschaft, daß von je drei seiner Punkte einer zwischen den beiden anderen liegt, nicht nur, wie Herr Biedermann zeigte, topologisch, sondern metrisch als Teilmenge der Gerade charakterisiert sei, d. h. daß ein Raum von dieser Eigenschaft nicht nur eindeutig und umkehrbar stetig, sondern eindeutig und abstandstreu auf eine Teilmenge der Gerade abgebildet werden könne. Dies trifft jedoch, wie ich bald an einem

einfachen Beispiel feststellte, ohne Einschränkungen nicht immer zu. Besteht nämlich der Raum R aus vier Punkten p_1, p_2, p_3, p_4 , welche folgende Abstände haben: $p_1 p_2 = p_3 p_4 = p_1 p_3 = p_2 p_4 = 1$, $p_1 p_4 = p_2 p_3 = 2$, — dann ist R nicht abstandstreu in die Gerade einbettbar, obwohl je drei Punkte von R abstandstreu in die Gerade eingebettet werden können. Eine genauere Untersuchung der diesbezüglichen Verhältnisse, zu welcher ich durch die Frage von Herrn Biedermann veranlaßt worden bin, führte mich nun, allerdings erst nach Ausarbeitung neuartiger geometrischer Methoden, tatsächlich auf eine metrische Charakterisierung der Teilmengen euklidischer Räume und des R_n selbst.

1. Problemstellung.

Es liege ein metrischer Raum R vor, d. h. eine Menge von Elementen (die wir Punkte nennen), so daß je zwei Elementen x und y von R eine reelle Zahl $xy = yx > 0$, für $x \neq y$, und $xy = 0$ für $x = y$ zugeordnet ist und so, daß für je drei Punkte x, y und z von R die Dreiecksungleichung $xy + yz \geq z$ gilt. Wir sagen, ein metrischer Raum R sei *abstandstreu* in den metrischen Raum A *einbettbar*, wenn R eindeutig auf eine Teilmenge R' von A abgebildet werden kann, so daß der Abstand je zweier Punkte von R' gleich ist dem Abstand ihrer Urbilder in R . Eine solche abstandstreue Abbildung ist offenbar auch eindeutig und beiderseits stetig. Eine Menge, z. B. R' , auf die R abstandstreu abbildbar ist, nennen wir auch mit R *abstandsgleich*.

Wir fragen nun nach notwendigen und hinreichenden Bedingungen dafür, daß ein metrischer Raum abstandstreu in den R_n , in den n -dimensionalen euklidischen Raum, einbettbar ist, und insbesondere nach den Bedingungen, unter denen ein Raum mit dem R_n selbst abstandsgleich ist. Bei unseren diesbezüglichen Untersuchungen werden wir niemals Gebrauch von der für einen metrischen Raum vorausgesetzten Dreiecksungleichung machen. Wir stellen daher unsere Frage gleich allgemeiner. Wir bezeichnen eine Menge von Elementen („Punkten“), so daß je zwei Punkten x und y eine reelle Zahl $xy = yx > 0$ für $x \neq y$ und $xy = 0$ für $x = y$ definiert ist, als *halbmetrischen Raum* und stellen nun allgemein die Frage:

Welche Bedingungen sind notwendig und hinreichend dafür, daß ein halbmetrischer Raum abstandstreu in den R_n einbettbar sei? Und wann ist ein halbmetrischer Raum insbesondere mit dem R_n abstandsgleich?

2. Zurückführung des Problems auf die Einbettung von $n + 3$ Punkten in den R_n .

Wir beweisen zunächst einen Satz, durch welchen das Problem zurückgeführt wird auf die Frage nach den Bedingungen, unter denen ein System von $n + 3$ Punkten abstandstreu in den R_n einbettbar ist.

Satz 1. *Ein halbmétrischer Raum, von dem je $n + 3$ Punkte abstandstreu in den R_n einbettbar sind, kann abstandstreu in den R_n eingebettet werden.*

Es sei R der vorgelegte halbmétrische Raum. Es sind zwei Fälle möglich: Entweder R enthält $n + 1$ Punkte, die nicht abstandstreu in den R_{n-1} einbettbar sind, oder je $n + 1$ Punkte von R sind abstandstreu in den R_{n-1} einbettbar.

A. Wir beweisen die Einbettbarkeit von R in R_n erstens für den Fall, daß R $n + 1$ Punkte $a_1, a_2, \dots, a_n, a_{n+1}$ enthält, die nicht abstandstreu in den R_{n-1} eingebettet werden können. Nach Voraussetzung sind je $n + 3$ Punkte von R , also erst recht die $n + 1$ Punkte a_1, \dots, a_{n+1} abstandstreu in den R_n einbettbar. Es sei a'_1, \dots, a'_{n+1} ein System von Punkten des R_n , so daß $a'_i a'_k = a_i a_k$ ($i, k = 1, 2, \dots, n + 1$) gilt. Da die $n + 1$ Punkte a_1, \dots, a_{n+1} nicht abstandstreu in den R_{n-1} einbettbar sein sollen, so liegen die Punkte a'_1, \dots, a'_{n+1} nicht in einer Hyperebene des R_n). Sind $n + 1$ positive Zahlen r_1, \dots, r_{n+1} vorgegeben, so existiert daher höchstens ein Punkt des R_n , welcher für $i = 1, 2, \dots, n + 1$ den Abstand r_i von a'_i hat (Bemerkung (*)).

Es sei nun p ein beliebig gewählter und sodann festgehaltener Punkt von R . Da nach Voraussetzung je $n + 3$ Punkte von R , also erst recht je $n + 2$ Punkte von R , abstandstreu in den R_n einbettbar sind, so können insbesondere auch die $n + 2$ Punkte a_1, \dots, a_{n+1}, p in den R_n abstandstreu eingebettet werden. Es existiert daher ein Punkt, also nach der Bemerkung (*) genau ein Punkt — wir wollen ihn p' nennen —, so daß gilt $p' a'_i = p a_i$ ($i = 1, \dots, n + 1$). Wir können demnach auf diese Weise jedem Punkt p von R genau einen Punkt p' des R_n zuordnen, mit anderen Worten: wir können R durch unsere Zuordnung eindeutig auf eine Teilmenge des R_n abbilden.

Wir wollen nachweisen, daß diese Abbildung abstandstreu ist. Es seien p und q irgend zwei Punkte von R und seien p' und q' ihre Bilder im R_n vermöge der angegebenen Abbildung. Wir haben zu zeigen, daß $p' q' = p q$ gilt. Dabei können wir offenbar annehmen, daß sowohl p als auch q von sämtlichen a_i verschieden sind. p' und q' sind die einzigen Punkte im R_n , welche den Bedingungen $p' a'_i = p a_i$, $q' a'_i = q a_i$ ($i = 1, \dots, n + 1$) genügen. Wäre $p' q' \neq p q$, so existierte daher im R_n kein abstandstreu Bild der Punkte $a_1, \dots, a_{n+1}, p, q$, während doch nach Voraussetzung je $n + 3$ Punkte von R abstandstreu in den R_n einbettbar sind. Also muß $p' q' = p q$ sein und es ist gezeigt, daß durch die angegebene Abbildung R abstandstreu in den R_n eingebettet wird.

^{*)} Als Hyperebene des R_n bezeichnen wir einen Teil— R_{n-1} des R_n , also einen linearen Teilraum, dessen Dimension um 1 geringer ist als die des Gesamtraumes.

B. Wir wenden uns zum Beweise des Satzes im zweiten Fall. Wir setzen also voraus, es liege ein Raum R vor, von dem je $n + 3$ Punkte in den R_n und je $n + 1$ Punkte in den R_{n-1} abstandstreu einbettbar sind. Wir werden zeigen, daß der Raum R nicht nur in den R_n , sondern sogar in den R_{n-1} abstandstreu einbettbar ist. Wir beweisen sogar: *Ein Raum, von dem je $n + 2$ Punkte in den R_n und je $n + 1$ Punkte in den R_{n-1} abstandstreu einbettbar sind, kann in den R_{n-1} abstandstreu eingebettet werden.*

Diese Behauptung ist trivial für $n = 1$. Denn jeder Raum, von dem je zwei Punkte abstandstreu in den R_0 einbettbar, d. h.⁸⁾ identisch sind, ist von der Mächtigkeit 0 oder 1, also abstandstreu in den R_0 einbettbar. Wir machen die Annahme (\dagger), es sei für alle Zahlen $m \leq n - 1$ bewiesen, daß ein Raum, von dem je $m + 2$ Punkte in den R_m und je $m + 1$ Punkte in den R_{m-1} abstandstreu einbettbar sind, abstandstreu in den R_{m-1} eingebettet werden kann.

Es liege dann ein Raum R vor, von dem je $n + 2$ Punkte in den R_n und je $n + 1$ Punkte in den R_{n-1} abstandstreu einbettbar sind. Wieder sind zwei Fälle möglich:

B $_{\alpha}$. Entweder sind je n Punkte von R in den R_{n-2} abstandstreu einbettbar. Da überdies nach Voraussetzung je $n + 1$ Punkte von R abstandstreu in den R_{n-1} einbettbar sind, so ist auf Grund der Annahme (\dagger) R abstandstreu in den R_{n-2} , also erst recht in den R_{n-1} , einbettbar.

B $_{\beta}$. Oder R enthält n Punkte, welche nicht in den R_{n-2} abstandstreu einbettbar sind.

Wir bemerken zunächst: Wenn $n + 2$ Punkte p_1, \dots, p_{n+2} abstandstreu in den R_n einbettbar sind und je $n + 1$ von diesen Punkten abstandstreu in den R_{n-1} eingebettet werden können, dann sind die $n + 2$ Punkte abstandstreu in den R_{n-1} einbettbar.

Betrachten wir nämlich irgend $n + 1$ von den Punkten p_1, \dots, p_{n+2} , etwa $p_1, \dots, p_{i-1}, p_{i+1}, \dots, p_{n+2}$. Nach Voraussetzung sind diese $n + 1$ Punkte in den R_{n-1} abstandstreu einbettbar. Es existieren also in einer Hyperebene des R_n $n + 1$ Punkte $\bar{p}_1, \dots, \bar{p}_{i-1}, \bar{p}_{i+1}, \dots, \bar{p}_{n+2}$, so daß $\bar{p}_j \bar{p}_k = p_j p_k$ ($j, k = 1, \dots, i - 1, i + 1, \dots, n + 2$) gilt. Das n -dimensionale Volumen des im R_n durch diese $n + 1$ Punkte \bar{p} bestimmten Simplexes ist, da die Endpunkte in einer Hyperebene des R_n liegen, gleich Null. Nach Voraussetzung existieren im R_n $n + 2$ Punkte p'_1, \dots, p'_{n+2} , so daß $p'_i p'_k = p_i p_k$ ($i, k = 1, \dots, n + 2$) gilt. Unsere Behauptung geht dahin, daß p_1, \dots, p_{n+2} abstandstreu in den R_{n-1} einbettbar sind. Wir haben also zu zeigen, daß die $n + 2$ Punkte p'_1, \dots, p'_{n+2} in einer Hyperebene

⁸⁾ Als R_0 , als nulldimensionalen euklidischen Raum, bezeichnen wir einen aus einem einzigen Punkt bestehenden Raum.

des R_n liegen. Dazu genügt es zu zeigen, daß für jedes i die $n + 1$ Punkte $p'_1, \dots, p'_{i-1}, p'_{i+1}, \dots, p'_{n+2}$ in einer Hyperebene des R_n liegen. Wäre dies aber für irgend ein i nicht der Fall, so bestimmten die betreffenden $n + 1$ Punkte im R_n ein n -dimensionales Simplex mit positivem Volumen. Das ist aber unmöglich, da, wie wir gesehen haben, ein kantengleiches Simplex mit dem Volumen Null existiert und das Volumen eines n -dimensionalen Simplexes eindeutig durch dessen Kantenlängen bestimmt ist.

Auf den Fall B_β zurückkommend sehen wir nun: Nach Voraussetzung sind je $n + 2$ Punkte von R in den R_n , je $n + 1$ Punkte von R in den R_{n-1} abstandstreu einbettbar. Nach der eben bewiesenen Bemerkung sind also je $n + 2$ Punkte von R in den R_{n-1} abstandstreu einbettbar. Überdies existieren nach Voraussetzung $n + 1$ Punkte von R , welche nicht in den R_{n-2} abstandstreu einbettbar sind. Dann ist aber R auf Grund des bereits erledigten Falles A abstandstreu in den R_{n-1} einbettbar. Damit ist der Beweis von Satz 1 in allen Fällen durchgeführt⁹⁾.

3. Über merkwürdige Systeme von $n + 3$ Punkten.

Auf Grund von Satz 1 handelt es sich nun um die Bedingungen, unter denen $n + 3$ Punkte in den R_n abstandstreu einbettbar sind. Daß die Einbettbarkeit von je $n + 2$ Punkten in den R_n hierfür nicht immer hinreichend ist, geht für $n = 1$ schon aus dem in der Vorbemerkung erwähnten einfachen Beispiel hervor. Wir beweisen nun

Satz 2. Für jedes $n > 0$ existieren Räume, welche $n + 3$ Punkte enthalten und nicht in den R_n abstandstreu einbettbar sind, obwohl je $n + 2$ von ihren Punkten abstandstreu in den R_n einbettbar sind.

Wir bezeichnen die Punkte mit a_1, \dots, a_{n+3} und treffen zur Konstruktion eines den Satz 2 beweisenden Beispiels folgende Abstandsfestsetzungen:

$$\begin{aligned} a_i a_i &= 0 & (i = 1, \dots, n + 3), \\ a_i a_n &= 1 & (i \neq k; \quad i, k = 1, \dots, n + 1), \\ a_i a_{n+2} &= r_n & (i = 1, \dots, n + 1), \end{aligned}$$

wobei r_n den Abstand des Schwerpunktes von den Eckpunkten in einem n -dimensionalen Simplex, dessen sämtliche Kanten die Länge 1 haben, bezeichnet.

Die bisherigen Festsetzungen lassen sich im R_n realisieren, indem wir mit a'_i das Bild von a_i bezeichnen und a'_1, \dots, a'_{n+1} als Eckpunkte eines n -dimensionalen Simplexes, dessen sämtliche Kanten die Länge 1 haben, und a'_{n+2} als den Schwerpunkt dieses Simplexes wählen. Wir setzen

⁹⁾ [Zusatz bei der Korrektur:] Einen kurzen Beweis eines allgemeineren Theorems gab ich in den „Bemerkungen zur zweiten Untersuchung über allgemeine Metrik“ (Proc. Ac. Amsterdam, 30, S. 710).

ferner fest:

$$a_{n+3} a_i = r_n \quad (i = 1, \dots, n + 1).$$

Da a'_{n+2} der einzige Punkt des R_n ist, welcher den Abstand r_n von allen Punkten a'_i ($i = 1, \dots, n + 1$) hat, so lassen sich die Abstandsverhältnisse der Punkte a_1, \dots, a_{n+3} dann und nur dann im R_n realisieren, wenn $a_{n+2} a_{n+3} = 0$ und demnach a'_{n+2} mit a'_{n+3} identisch ist. Wir setzen statt dessen fest:

$a_{n+2} a_{n+3} = d$, wo d den doppelten Abstand zwischen dem Schwerpunkt und einer $(n - 1)$ -dimensionalen Seite eines n -dimensionalen Simplexes, dessen sämtliche Kanten die Länge 1 haben, bezeichnet.

Dann sind also die $n + 3$ Punkte a_1, \dots, a_{n+3} nicht abstandstreu in den R_n einbettbar. Man sieht aber leicht, daß je $n + 2$ von diesen Punkten abstandstreu in den R_n eingebettet werden können. Für die $(n + 2)$ -Tupel $a_1, \dots, a_{n+1}, a_{n+2}$ und $a_1, \dots, a_{n+1}, a_{n+3}$ ist dies bereits nachgewiesen. Die übrigen $(n + 2)$ -Tupel sind untereinander abstandsgleich; es genügt daher eines von ihnen, etwa $a_2, \dots, a_{n+1}, a_{n+2}, a_{n+3}$, zu betrachten. Die Abstandsverhältnisse dieser $n + 2$ Punkte lassen sich aber im R_n realisieren: Man wähle a'_2, \dots, a'_{n+1} als Eckpunkte eines $(n - 1)$ -dimensionalen Simplexes S , dessen sämtliche Kanten die Länge 1 haben; sodann bestimme man die Gerade senkrecht zu der Hyperebene durch die Punkte a'_2, \dots, a'_{n+1} , welche den Schwerpunkt S enthält; man hat nun a'_{n+2} und a'_{n+3} als jene beiden Punkte dieser Geraden zu wählen, welche vom Schwerpunkt von S den Abstand $\frac{1}{2} d$ besitzen. — Damit ist Satz 2 bewiesen.

4. Eine notwendige Bedingung für die Einbettbarkeit von $n + 2$ Punkten in den R_n .

Wir führen nun zunächst eine im folgenden vielverwendete Bedingung an, welcher die Abstände von $n + 2$ Punkten, die abstandstreu in den R_n einbettbar sind, genügen müssen. Sind $n + 1$ Punkte des R_n gegeben, p_1, \dots, p_{n+1} , und bezeichnet $r_{i,k}$ den Abstand der Punkte p_i und p_k , dann ist bekanntlich¹⁰⁾ das n -dimensionale Volumen $v(p_1, \dots, p_{n+1})$ des durch die $n + 1$ Punkte bestimmten Simplexes gegeben durch den Ausdruck

$$v^2(p_1, \dots, p_{n+1}) = \frac{(-1)^{n+1}}{2^n (n!)^2} \cdot \begin{vmatrix} 0 & 1 & 1 & \dots & 1 \\ 1 & 0 & r_{1,2}^2 & \dots & r_{1,n+1}^2 \\ 1 & r_{1,2}^2 & 0 & \dots & r_{2,n+1}^2 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & r_{n+1,1}^2 & r_{n+1,2}^2 & \dots & 0 \end{vmatrix} \quad (\text{Volumensformel})$$

¹⁰⁾ Vgl. z. B. R. H. Schouten, Mehrdimensionale Geometrie 2, (Samml. Schubert 36), S. 123.

Indem wir für $r_{i,k}$ die Abstandszahl $p_i p_k$ setzen und berücksichtigen, daß $p_i p_k = p_k p_i$ und $p_i p_i = 0$ ist, können wir die auf der rechten Seite der Volumensformel auftretende Determinante in einer nicht mißzuverstehenden Abkürzung auch schreiben als

$$\begin{vmatrix} 0 & 1 \\ 1 & (p_i p_k)^2 \end{vmatrix}_{i,k=1,\dots,n+1}$$

Wir wollen nun allgemein, wenn $n+1$ Punkte und ihre gegenseitigen Abstände gegeben sind (auch wenn die $n+1$ Punkte nicht speziell als Punkte eines R_n gegeben sind), die Bezeichnung einführen:

$$(*) \quad D(p_1, \dots, p_{n+1}) = \begin{vmatrix} 0 & 1 \\ 1 & (p_i p_k)^2 \end{vmatrix}_{i,k=1,\dots,n+1}.$$

Sind die $n+1$ Punkte p_1, \dots, p_{n+1} in den R_{n-1} abstandstreu einbettbar, dann existieren in einer Hyperebene des R_n $n+1$ Punkte p'_1, \dots, p'_{n+1} , so daß $p'_i p'_k = p_i p_k$ ($i, k = 1, \dots, n+1$) gilt. Das durch die $n+1$ Punkte p'_1, \dots, p'_{n+1} bestimmte n -dimensionale Simplex hat das Volumen Null. Es gilt also der Volumensformel zufolge $D(p'_1, \dots, p'_{n+1}) = 0$ und mithin wegen $p_i p_k = p'_i p'_k$ auch $D(p_1, \dots, p_{n+1}) = 0$. Wir haben also bewiesen:

Satz 3. Wenn $n+2$ Punkte p_1, \dots, p_{n+2} abstandstreu in den R_n einbettbar sind, so gilt

$$D(p_1, \dots, p_{n+2}) = \begin{vmatrix} 0 & 1 \\ 1 & (p_i p_k)^2 \end{vmatrix}_{i,k=1,\dots,n+2} = 0.$$

5. Über Systeme von $n+3$ Punkten, von denen je $n+2$ in den R_n abstandstreu einbettbar sind. — Die pseudoeuklidischen Systeme.

Wir wissen aus Satz 2, daß ein System von $n+3$ Punkten, von denen je $n+2$ abstandstreu in den R_n eingebettet werden können, nicht notwendig abstandstreu in den R_n einbettbar sein muß. Wir wollen nunmehr über die $n+3$ Punkte p_1, \dots, p_{n+3} voraussetzen, daß je $n+2$ von ihnen abstandstreu in den R_n einbettbar sind, und wollen untersuchen, was aus dieser Voraussetzung für das System der $n+3$ Punkte folgt. Es sind zwei Fälle möglich:

A. Entweder: je $n+1$ von den $n+3$ Punkten sind abstandstreu in den R_{n-1} einbettbar. Dann sind von den $n+3$ Punkten auf Grund des Satzes, den wir im Falle B_β von Satz 1 bewiesen haben, je $n+2$ abstandstreu in den R_{n-1} einbettbar und daher sind nach Satz 1 die $n+3$ Punkte abstandstreu in den R_{n-1} einbettbar.

B. Oder es gibt unter den $n+3$ Punkten $n+1$ Punkte, die nicht abstandstreu in den R_{n-1} eingebettet werden können. Wir denken uns

die Bezeichnung so gewählt, daß die Punkte p_1, \dots, p_{n+1} nicht in den R_{n-1} eingebettet werden können.

Nach Voraussetzung sind je $n + 2$ Punkte, insbesondere also die Punkte p_1, \dots, p_{n+2} , abstandstreu in den R_n einbettbar. Es existiert daher im R_n ein System von Punkten p'_1, \dots, p'_{n+2} , so daß $p'_i p'_k = p_i p_k$ ($i, k = 1, \dots, n + 2$) gilt. Da die Punkte p_1, \dots, p_{n+1} nicht in den R_{n-1} abstandstreu einbettbar sind, liegen die Punkte p'_1, \dots, p'_{n+1} nicht in einer Hyperebene des R_n .

Nach Voraussetzung sind ferner die Punkte $p_1, \dots, p_{n+1}, p_{n+3}$ abstandstreu in den R_n einbettbar. Es existiert daher ein Punkt p'_{n+3} (und zwar, da die Punkte p'_1, \dots, p'_{n+1} nicht in einer Hyperebene liegen, ein *einzig*er solcher Punkt), so daß $p'_{n+3} p'_i = p_{n+3} p_i$ ($i = 1, \dots, n + 1$) gilt.

Nach Voraussetzung sind endlich für jedes i ($i = 1, \dots, n + 1$) die Punkte $p_1, \dots, p_{i-1}, p_{i+1}, \dots, p_{n+2}, p_{n+3}$ abstandstreu in den R_n einbettbar; mithin gilt nach Satz 3 für jedes i ($1 \leq i \leq n + 1$) die Beziehung

$$(*) \quad D(p_1, \dots, p_{i-1}, p_{i+1}, \dots, p_{n+2}, p_{n+3}) = 0.$$

Für jedes $i \leq n + 1$ tritt in dieser Beziehung sowohl der Punkt p_{n+2} als auch der Punkt p_{n+3} wirklich auf. Wir ersetzen nun in der i -ten dieser Beziehungen den Term $p_{n+2} p_{n+3}$ durch x . Wir erhalten auf diese Weise eine quadratische Gleichung in x^2 , welche wir mit $D_i(x^2) = 0$ bezeichnen wollen. Gehen wir auf die Definition (*) des Ausdruckes $D(\dots)$ zurück, so sehen wir, daß man schreiben kann:

$$D_i(x^2) \equiv \begin{vmatrix} 0 & 1 & 1 & 1 \\ 1 & (p_k p_l)^2 & (p_k p_{n+2})^2 & (p_k p_{n+3})^2 \\ 1 & (p_{n+2} p_l)^2 & 0 & x^2 \\ 1 & (p_{n+3} p_l)^2 & x^2 & 0 \end{vmatrix} \quad (k, l = 1, \dots, i-1, i+1, \dots, n+1)$$

Berücksichtigen wir, daß, falls nicht $k = n + 2, l = n + 3$ oder $k = n + 3, l = n + 2$ ist, $p'_k p'_l = p_k p_l$ ($k, l = 1, \dots, n + 3$) gilt, so können wir auch schreiben:

$$(**) \quad D_i(x^2) \equiv \begin{vmatrix} 0 & 1 & 1 & 1 \\ 1 & (p'_k p'_l)^2 & (p'_k p'_{n+2})^2 & (p'_k p'_{n+3})^2 \\ 1 & (p'_{n+2} p'_l)^2 & 0 & x^2 \\ 1 & (p'_{n+3} p'_l)^2 & x^2 & 0 \end{vmatrix} \quad (k, l = 1, \dots, i-1, i+1, \dots, n+1)$$

Die Gleichung $D_i(x^2) = 0$ kann nicht identisch verschwinden. Wir wollen nämlich zeigen, daß der Koeffizient von x^4 in dieser Gleichung $\neq 0$ ist. In der Tat, dieser Koeffizient ergibt sich, wenn wir die Determinante (***) nach Unterdeterminanten der beiden letzten Spalten entwickeln,

$$= - \begin{vmatrix} 0 & 1 \\ 1 & (p'_k p'_l)^2 \end{vmatrix} \quad (k, l = 1, \dots, i-1, i+1, \dots, n+1)$$

also $= -D(p'_1, \dots, p'_{i-1}, p'_{i+1}, \dots, p'_{n+1})$. Diese Größe ist aber $\neq 0$, denn wäre sie $= 0$, so bestimmten nach der Volumensformel die Punkte $p'_1, \dots, p'_{i-1}, p'_{i+1}, \dots, p'_{n+1}$ ein Simplex mit dem Volumen 0, lägen also in einer $(n-2)$ -dimensionalen Ebene des R_n . Dies ist aber unmöglich, weil in diesem Fall die $n+1$ Punkte p'_1, \dots, p'_{n+1} in einer Hyperebene des R_n lägen, im Widerspruch zur Voraussetzung.

Jede der Gleichungen $D_i(x^2) = 0$ hat also zwei Lösungen. Beide sind positiv und von den positiven Quadratwurzeln aus ihnen kann man sofort eine geometrische Bedeutung angeben: Die *eine* ist für jedes i der Abstand $p'_{n+2} p'_{n+3}$, die *andere* ist der Abstand des Punktes p'_{n+2} von jenem Punkte — wir wollen ihn p''_{n+3} nennen — welcher zu p'_{n+3} symmetrisch liegt hinsichtlich der Hyperebene H_i , die durch die Punkte $p'_1, \dots, p'_{i-1}, p'_{i+1}, \dots, p'_{n+1}$ bestimmt wird; (oder, was dasselbe ist, der Abstand von p'_{n+3} und dem Punkt p''_{n+2} , welcher zu p_{n+2} hinsichtlich H_i symmetrisch liegt). In der Tat, die beiden $(n+2)$ -Tupel von Punkten $p'_1, \dots, p'_{i-1}, p'_{i+1}, \dots, p'_{n+2}, p'_{n+3}$ und $p'_1, \dots, p'_{i-1}, p'_{i+1}, \dots, p'_{n+2}, p''_{n+3}$ liegen ja im R_n , erfüllen also nach Satz 3 die Beziehungen $D(p'_1, \dots, p'_{i-1}, p'_{i+1}, \dots, p'_{n+2}, p'_{n+3}) = 0$ und $D(p'_1, \dots, p'_{i-1}, p'_{i+1}, \dots, p'_{n+2}, p''_{n+3}) = 0$. Jede der beiden Größen $(p'_{n+2} p'_{n+3})^2$ und $(p'_{n+2} p''_{n+3})^2$ genügt also der Gleichung $D_i(x^2) = 0$ und weitere Lösungen kann diese Gleichung nicht besitzen.

Zufolge (**) genügt $(p_{n+2} p_{n+3})^2$ für jedes i der Gleichung $D_i(x^2) = 0$. Es ist also $(p_{n+2} p_{n+3})^2$ eine gemeinsame Lösung der $n+1$ quadratischen Gleichungen $D_i(x^2) = 0$, ($i = 1, 2, \dots, n+1$). Demnach kommen für $p_{n+2} p_{n+3}$ nur folgende Möglichkeiten in Betracht:

a) Es kann $(p_{n+2} p_{n+3})^2 = (p'_{n+2} p'_{n+3})^2$ und mithin $p_{n+2} p_{n+3} = p'_{n+2} p'_{n+3}$ sein. In diesem Fall sind die $n+3$ Punkte p_1, \dots, p_{n+3} *abstandstreu* in R_n einbettbar. Diese Lösung $(p'_{n+2} p'_{n+3})^2$, die **euklidische Lösung**, haben die $n+1$ quadratischen Gleichungen $D_i(x^2) = 0$, wie wir sahen, immer gemeinsam.

b) Im Falle daß $p'_{n+2} p'_{n+3} = p'_{n+2} p''_{n+3} = \dots = p'_{n+2} p^{n+1}_{n+3}$ gilt — geometrisch gesprochen: im Falle, daß p'_{n+2} gleich weit entfernt ist von den Spiegelbildern des Punktes p'_{n+3} hinsichtlich der Hyperebenen H_i — und *nur* in diesem Falle besitzen die $n+1$ Gleichungen noch eine weitere gemeinsame Lösung, nämlich $p_{n+2} p''_{n+3}$. Es sind dann zwei Fälle möglich, die wir unterscheiden müssen:

b') *Entweder* liegt der Punkt p'_{n+3} in mindestens einer der Hyperebenen H_i , etwa in H_i . Dann fällt also der Punkt p'_{n+3} mit seinem Spiegelbild p''_{n+3} zusammen und es ist $p'_{n+2} p'_{n+3} = p'_{n+2} p''_{n+3}$ ($i = 1, \dots, n+1$). Jede der Gleichungen $D_i(x^2) = 0$ besitzt dann also die euklidische Lösung $(p'_{n+2} p'_{n+3})^2$ zur Doppelwurzel; eine andere Lösung als die euklidische kann

also nicht existieren und wir sehen: *Liegt p'_{n+3} in einer Hyperebene H_i , dann sind die $n+3$ Punkte p_1, \dots, p_{n+3} abstandstreu in den R_n einbettbar.*

b") Oder p'_{n+3} liegt in keiner Hyperebene H_i ($i=1, 2, \dots, n+1$). Dann ist also p'_{n+3} von seinen Spiegelbildern p_{n+3}^i verschieden und diese Spiegelbilder sind, wegen der Verschiedenheiten der Hyperebenen H_i , untereinander paarweise verschieden. Es sind dann wieder zwei Fälle zu unterscheiden.

b̄") Entweder die $n+1$ Punkte p_{n+3}^i liegen in einer Hyperebene des R_n . (Auf diesen Ausnahmefall bin ich durch Herrn D. van Dantzig aufmerksam gemacht worden.) Wir denken im R_n rechtwinklige Cartesische Koordinaten eingeführt und es seien ${}_k x_1, \dots, {}_k x_n$ die Koordinaten des Punktes p'_k ($k=1, \dots, n+3$). Die Koordinaten des Punktes p_{n+3}^i bezeichnen wir mit ${}_{n+3}^i x_1, \dots, {}_{n+3}^i x_n$ ($i=1, \dots, n+1$). Wenn wir die Punkte p'_1, \dots, p'_{n+1} festhalten, dann sind die Koordinaten ${}_{n+3}^i x_k$ ($i=1, \dots, n+1, k=1, \dots, n$) lineare Funktionen der Koordinaten ${}_{n+3} x_1, \dots, {}_{n+3} x_n$. Die $n+1$ Punkte p_{n+3}^i liegen in einer Hyperebene dann und nur dann, wenn das Volumen des durch die Punkte p_{n+3}^i bestimmten Simplexes gleich Null ist, d. h. wenn

$$\left| \begin{array}{c} {}_{n+3}^i x_k, 1 \\ (i=1, \dots, n+1) \\ (k=1, \dots, n) \end{array} \right| \equiv \left| \begin{array}{cccc} {}_{n+3}^1 x_1, & {}_{n+3}^1 x_2, & \dots, & {}_{n+3}^1 x_n, & 1 \\ {}_{n+3}^2 x_1, & {}_{n+3}^2 x_2, & \dots, & {}_{n+3}^2 x_n, & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ {}_{n+3}^{n+1} x_1, & {}_{n+3}^{n+1} x_2, & \dots, & {}_{n+3}^{n+1} x_n, & 1 \end{array} \right| = 0$$

gilt. Substituieren wir in dieser Beziehung für die Größen ${}_{n+3}^i x_k$ die linearen Ausdrücke in den ${}_{n+3} x_i$, so erhalten wir eine Gleichung $f({}_{n+3} x_1, \dots, {}_{n+3} x_n) = 0$ von n -tem Grad in den ${}_{n+3} x_i$. Die Punkte des R_n , welche dieser Gleichung genügen, sind also identisch mit jenen Punkten, für welche die Spiegelpunkte hinsichtlich der Hyperebenen H_i in eine Hyperebene fallen, für welche mit anderen Worten das durch die Spiegelpunkte an den H_i bestimmte Simplex degeneriert. Die Menge aller dieser Punkte bildet eine Hyperfläche n -ten Grades im R_n , welche wir die zu den Punkten p'_1, \dots, p'_{n+1} gehörige *Ausnahmshyperfläche* nennen und mit $F(p'_1, \dots, p'_{n+1})$ bezeichnen wollen. (Es läßt sich zeigen, daß in homogenen Koordinaten mit den Punkten p'_1, \dots, p'_{n+1} als Fundamentalphunkten die Ausnahmsfläche $F(p'_1, \dots, p'_{n+1})$ bei geeigneter Wahl eines Einheitspunktes

die Gleichung hat $\sum_{i=1}^{n+1} \frac{1}{x_i} = 0$). Der Fall, daß die $n+1$ Punkte p_{n+3}^i in einer Hyperebene liegen, ereignet sich dann und nur dann, wenn p'_{n+3} Punkt von $F(p'_1, \dots, p'_{n+1})$ ist. Wenn die $n+1$ Punkte p_{n+3}^i in einer Hyperebene liegen, dann existiert kein Punkt des R_n , der von ihnen allen gleichweit entfernt wäre. Die $n+1$ Gleichungen $D_i(x^2) = 0$ haben also in diesem Fall außer $(p'_{n+2} p'_{n+3})^2$ keine gemeinsame Lösung, es muß mithin

$p_{n+2}p_{n+3} = p'_{n+2}p'_{n+3}$ sein, d. h. die $n+3$ Punkte p_1, \dots, p_{n+3} sind abstandstreu in den R_n einbettbar.

$\bar{b}'')$ Oder die $n+1$ Punkte p'_{n+3} liegen in keiner Hyperebene des R_n . Dies ist nach dem Vorangehenden dann und nur dann der Fall, wenn p'_{n+3} nicht in $F(p'_1, \dots, p'_{n+1})$ liegt. In diesem Fall bestimmen die Punkte p'_{n+3} , da sie paarweise verschieden sind, eindeutig eine $(n-1)$ -dimensionale Sphäre von endlichem Radius, welche wir die Spiegelpunktsphäre von p'_{n+3} hinsichtlich der Punkte p'_1, \dots, p'_{n+1} nennen und mit $S(p'_{n+3}; p'_1, \dots, p'_{n+1})$ bezeichnen wollen. Der Punkt p'_{n+2} , der nach Voraussetzung von den Punkten p'_{n+3} gleichweit entfernt ist, liegt also im Zentrum von $S(p'_{n+3}; p'_1, \dots, p'_{n+1})$.

Es gilt dann $p'_{n+2}p'_{n+3} \neq p'_{n+2}p'_{n+3}$. Wäre dies nämlich unrichtig, so müßte p'_{n+2} von p'_{n+3} denselben Abstand haben, wie von den Punkten p'_{n+3} . Nun liegen aber alle Punkte, welche von p'_{n+3} und p'_{n+3} den gleichen Abstand haben, in H_i , und da die Punkte p_1, \dots, p_{n+1} in keiner Hyperebene liegen, so existiert kein Punkt, welcher allen $n+1$ H_i gemein wäre. In diesem Fall besitzen also die $n+1$ Gleichungen $D_i(x^2) = 0$ auch eine von der euklidischen Lösung verschiedene gemeinsame Lösung, welche wir die **pseudo-euklidische Lösung** nennen wollen. Der Abstand $p_{n+2}p_{n+3}$ ist dann *entweder* $= p'_{n+2}p'_{n+3}$, in welchem Fall die $n+3$ Punkte p_1, \dots, p_{n+3} abstandstreu in den R_n einbettbar sind, *oder* $= p'_{n+2}p'_{n+3}$, in welchem Fall die $n+3$ Punkte p_1, \dots, p_{n+3} , obwohl je $n+2$ von ihnen in den R_n einbettbar sind, *nicht* abstandstreu in den R_n eingebettet werden können und wir von einem *pseudoeuklidischen* $(n+3)$ -Tupel sprechen wollen.

Betrachten wir unser pseudoeuklidisches $(n+3)$ -Tupel p_1, \dots, p_{n+3} . Es enthält, wie wir gesehen haben, $n+1$ Punkte, die nicht abstandstreu in den R_{n-1} eingebettet werden können. Wir hatten die Bezeichnung so gewählt, daß die Punkte p'_1, \dots, p'_{n+1} nicht in einer Hyperebene des R_n liegen. Wir haben ferner gesehen, daß der Punkt p'_{n+3} in keiner der $n+1$ Hyperebenen H_i liegt; da unsere diesbezüglichen Überlegungen in p'_{n+2} und p'_{n+3} symmetrisch waren, so folgt, daß auch p'_{n+2} in keiner der Hyperebenen H_i liegt. Es sind also auch die $2(n+1)(n+1)$ -Tupel $p_1, \dots, p_{i-1}, p_{i+1}, \dots, p_{n+1}, p_{n+2}$ und $p_1, \dots, p_{i-1}, p_{i+1}, \dots, p_{n+1}, p_{n+3}$ ($i = 1, 2, \dots, n+1$) nicht in den R_{n-1} abstandstreu einbettbar. Wir wollen nun zeigen, daß auch keines der $\frac{n(n+1)}{2}$ übrigen $(n+1)$ -Tupel von Punkten unseres pseudoeuklidischen $(n+3)$ -Tupels sich abstandstreu in den R_{n-1} einbetten läßt. Wären nämlich etwa die Punkte $p_1, \dots, p_{i-1}, p_{i+1}, \dots, p_{j-1}, p_{j+1}, \dots, p_{n+1}, p_{n+2}, p_{n+3}$ ($1 \leq i < j \leq n+1$) abstandstreu in den R_{n-1} einbettbar, so lägen die $n+1$ Punkte

$$p'_1, \dots, p'_{i-1}, p'_{i+1}, \dots, p'_{j-1}, p'_{j+1}, \dots, p'_{n+1}, p'_{n+2}, p'_{n+3},$$

und ebenso die $n + 1$ Punkte

$$p'_1, \dots, p'_{i-1}, p'_{i+1}, \dots, p'_{j-1}, p'_{j+1}, \dots, p'_{n+1}, p'_{n+2}, p'_{n+3},$$

oder, wie wir statt dessen kurz sagen wollen, die $(n + 1)$ -Tupel (i) und (j) , in einer Hyperebene des R_n . Das ist aber unmöglich. Denn *erstens* liegen die n Punkte

$$p'_1, \dots, p'_{i-1}, p'_{i+1}, \dots, p'_{j-1}, p'_{j+1}, \dots, p'_{n+1}, p'_{n+2}$$

nicht in einer $(n - 2)$ -dimensionalen Ebene des R_n , weil sonst die Punkte $p'_1, \dots, p'_{i-1}, p'_{i+1}, \dots, p'_{n+1}, p'_{n+2}$ in einer Hyperebene lägen, was nach dem Bewiesenen nicht der Fall ist. *Zweitens* ist $p'_{n+3} p'_m = p''_{n+3} p'_m$ ($m = 1, \dots, i - 1, i + 1, \dots, j - 1, j + 1, \dots, n + 1, n + 2$). Lägen also die $(n + 1)$ -Tupel (i) und (j) in einer Hyperebene des R_n , so wäre $p'_{n+3} = p''_{n+3}$, im Widerspruch gegen die Annahme, daß je zwei von den $n + 1$ Punkten p'_{n+3} verschieden sind. Wir sehen also, daß *keine* $n + 1$ Punkte unseres pseudoeuklidischen $(n + 3)$ -Tupels abstandstreu in den R_{n-1} einbettbar sind.

Wir hatten, um die vorangehenden Überlegungen durchführen zu können, die Punkte p_1, \dots, p_{n+1} unter den $n + 3$ Punkten ausgewählt, weil sie nicht abstandstreu in den R_{n-1} einbettbar sind. Da dies, wie wir nun festgestellt haben, für je $n + 1$ Punkte eines pseudoeuklidischen $(n + 3)$ -Tupels zutrifft, so können, wie immer die $n + 3$ Punkte numeriert werden, die vorstehenden Überlegungen durchgeführt werden. Betrachten wir unter den $n + 3$ Punkten $\bar{p}_1, \dots, \bar{p}_{n+3}$ eines pseudoeuklidischen $(n + 3)$ -Tupels irgendeinen Punkt, etwa \bar{p}_i , und irgend $n + 1$ von ihm verschiedene Punkte, etwa $\bar{p}_1, \dots, \bar{p}_{i-1}, \bar{p}_{i+1}, \dots, \bar{p}_{j-1}, \bar{p}_{j+1}, \dots, \bar{p}_{n+3}$. Wir numerieren die Punkte so, daß wir $\bar{p}_i = p_{n+3}$ und $\bar{p}_j = p_{n+2}$ setzen und die $n + 1$ übrigen Punkte \bar{p}_k irgendwie mit p_1, \dots, p_{n+1} bezeichnen. Bilden wir ein mit den Punkten p_1, \dots, p_{n+1} abstandsgleiches System von Punkten im R_n p'_1, \dots, p'_{n+1} , so liegen diese Punkte sicher nicht in einer Hyperebene des R_n und p'_{n+3} liegt nicht in der durch die Punkte p'_1, \dots, p'_{n+1} bestimmten Ausnahmshyperfläche $F(p'_1, \dots, p'_{n+1})$. Da wir \bar{p}_i und die von ihm verschiedenen $n + 1$ Punkte willkürlich unter den $n + 3$ Punkten $\bar{p}_1, \dots, \bar{p}_{n+3}$ gewählt hatten, so sehen wir: *Von $n + 3$ Punkten in pseudo-euklidischer Lage lassen sich keine $n + 2$ abstandstreu in die durch irgend $n + 1$ von ihnen bestimmte Ausnahmshyperfläche des R_n einbetten.*

Die Resultate dieses Abschnittes zusammenfassend können wir aussprechen:

Satz 4. *Wenn von $n + 3$ Punkten je $n + 2$ abstandstreu in den R_n einbettbar sind, dann sind die $n + 3$ Punkte entweder euklidisch, d. h. in den R_n abstandstreu einbettbar, oder pseudoeuklidisch, wobei pseudo-euklidische $(n + 3)$ -Tupel folgende Eigenschaften haben:*

1. Von einem pseudoeuclidischen $(n+3)$ -Tupel sind keine $n+1$ Punkte in den R_{n-1} und keine $n+2$ Punkte in eine durch $n+1$ von ihnen bestimmte Ausnahmshyperfläche des R_n abstandstreu einbettbar.

2. Liegen $n+1$ Punkte p_1, \dots, p_{n+1} vor, welche in den R_n , aber nicht in den R_{n-1} abstandstreu einbettbar sind,

dann kann,

zu jedem $(n+1)$ -Tupel von Abstandszahlen $p_i p_{n+2}$ ($i=1, \dots, n+1$), welches den Bedingungen genügt, daß die $n+2$ Punkte p_1, \dots, p_{n+2} in den R_n , aber nicht in die durch $n+1$ von ihnen bestimmte Ausnahmshyperfläche des R_n und daß keine $n+1$ von ihnen in den R_{n-1} abstandstreu einbettbar sind,

und nur zu einem $(n+1)$ -Tupel von Zahlen, welches diesen Bedingungen genügt,

auf genau eine Weise ein $(n+1)$ -Tupel von Abstandszahlen $p_i p_{n+3}$ ($i=1, \dots, n+1$) bestimmt werden, so daß die $n+3$ Punkte p_1, \dots, p_{n+3} ein pseudoeuclidisches $(n+3)$ -Tupel bilden.

Dabei ist auch der pseudoeuclidische Abstand $p_{n+2} p_{n+3}$ aus den Zahlen $p_i p_k$ ($i, k=1, \dots, n+2$) eindeutig berechenbar. — Entspricht vermöge der Punkte p_1, \dots, p_{n+1} dem $(n+1)$ -Tupel von Abstandszahlen $p_i p_{n+2}$ ($i=1, \dots, n+1$) das $(n+1)$ -Tupel $p_i p_{n+3}$ ($i=1, \dots, n+1$), dann entspricht umgekehrt diesem letzteren $(n+1)$ -Tupel das erstere.

3. Ist unter den Voraussetzungen von 2. p'_1, \dots, p'_{n+1} ein zu den Punkten p_1, \dots, p_{n+1} abstandsgleiches Punktesystem im R_n , und sind die Zahlen $p_i p_{n+2}$ ($i=1, \dots, n+1$) gemäß den Bedingungen von 2. vorgegeben (so daß also genau ein p'_{n+2} im R_n existiert, welcher weder in einer der Hyperebenen H_i durch die Punkte $p'_1, \dots, p'_{i-1}, p'_{i+1}, \dots, p'_{n+1}$, noch in der Ausnahmshyperfläche $F(p'_1, \dots, p'_{n+1})$ liegt, und von den Punkten p'_i die Abstände $p_i p_{n+2}$ hat), — so bilden die $n+3$ Punkte p_1, \dots, p_{n+3} ein pseudoeuclidisches $(n+3)$ -Tupel dann und nur dann, wenn die Zahl $p_{n+3} p_i$ gleich ist dem Abstand des Zentrums der Spiegelpunktsphäre $S(p'_{n+2}; p'_1, \dots, p'_{n+1})$ von p'_i ($i=1, \dots, n+1$) und wenn $p_{n+2} p_{n+3}$ gleich ist dem Radius dieser Sphäre.

Sehr einfach liegen die Verhältnisse für pseudoeuclidische *Quadrupel*, d. h. für $n=1$. Seien etwa p_1, p_2, p_3, p_4 vier nicht in die Gerade einbettbare Punkte, von denen je drei in die Gerade eingebettet werden können. Es seien p'_1, p'_2, p'_3 drei mit p_1, p_2, p_3 abstandsgleiche Punkte der Geraden. Indem wir p_1, p_2, p_3 eventuell umnummerieren, können wir erreichen, daß $p'_1 p'_2 + p'_2 p'_3 = p'_1 p'_3$ gilt. Wenn das Quadrupel p_1, p_2, p_3, p_4 pseudoeuclidisch ist, so liegt nach Satz 4 der Punkt p'_4 gleich weit entfernt von den Spiegelbildern p_3^1 und p_3^2 des Punktes p_3 hinsichtlich der Punkte p'_2 und p'_1 . Daraus entnimmt man leicht, daß gilt:

$$p'_4 p_3^1 = p'_4 p_3^2 = p'_1 p'_2, \quad p'_4 p'_1 = p'_2 p'_3, \quad p'_4 p'_2 = p'_3 p'_1.$$

Alle pseudoeuklidischen Quadrupel lassen sich also bei geeigneter Numerierung geben durch die Festsetzungen:

$$p_1 p_2 = p_3 p_4 = a,$$

$$p_3 p_3 = p_1 p_4 = b,$$

$$p_1 p_3 = p_2 p_4 = a + b.$$

Jedes pseudoeuklidische Quadrupel läßt sich daher abstandstreu einbetten in eine Kreislinie, in der als Abstand je zweier Punkte die Länge des kürzeren Bogens zwischen den betreffenden Punkten festgesetzt ist.

Wir diskutieren nun noch den Fall $n = 2$, d. i. den Fall der pseudoeuklidischen Quintupel. Betrachten wir drei nicht in einer Geraden liegende Punkte a_1, a_2, a_3 der Ebene; der geometrische Ort aller Punkte, für welche die drei Spiegelpunkte an den Geraden durch a_1 und a_2 , durch a_1 und a_3 , durch a_2 und a_3 in einer Geraden liegen, ist bekanntlich der den Punkten a_1, a_2, a_3 umgeschriebene Kreis. Die Ausnahmskurve, welche durch drei nicht in einer Geraden liegenden Punkte der Ebene bestimmt wird, ist also der die drei Punkte enthaltende Kreis. Satz 4 lehrt daher, daß von den Punkten eines antieuklidischen Quintupels keine drei in eine Gerade und keine vier in einen Kreis abstandstreu einbettbar sind.

Werden in der Ebene drei Punkte p_1, p_2, p_3 festgelegt, so heißt die Zuordnung, vermöge welcher jedem Punkt p der Ebene der Mittelpunkt des Kreises durch die drei Spiegelpunkte von p an den Geraden $p_1 p_2, p_1 p_3, p_2 p_3$ entspricht, die hinsichtlich des Dreiecks p_1, p_2, p_3 *isogonale Transformation* der Ebene. Diese Zuordnung, (über welche im Enzyklopädieartikel von Berkhan-Meyer III A B 10, S. 1207 ausführlich referiert wird), ist eine *quadratische involutorische Cremonatransformation*, vermöge welcher jedem Punkt einer Geraden des Fundamentaldreiecks p_1, p_2, p_3 der gegenüberliegende Eckpunkt des Fundamentaldreiecks entspricht, und jeder Punkt des dem Fundamentaldreieck umgeschriebenen Kreises den Punkten der unendlich fernen Geraden entspricht. Abgesehen von diesen singulären Kurven ist die Transformation *ein-eindeutig*. Zwei einander vermöge dieser involutorischen Transformation entsprechende Punkte heißen *isogonale Gegenpunkte* hinsichtlich p_1, p_2, p_3 . Die Winkelhalbierenden des Fundamentaldreiecks werden durch die Transformation in sich übergeführt; der Mittelpunkt des dem Fundamentaldreieck eingeschriebenen Kreises und die Mittelpunkte der drei dem Fundamentaldreieck angeschriebenen Kreise sind die Fixpunkte der Transformation. Berücksichtigen wir diese Terminologie, so ergibt sich aus Satz 4 für $n = 2$:

Von den Punkten eines pseudoeuklidischen Quintupels sind keine drei in eine Gerade und keine vier in einen Kreis abstandstreu einbettbar. Bilden wir zu drei Punkten, welche in die Ebene, aber nicht in die Gerade abstandstreu einbettbar sind, ein abstandsgleiches System von Punkten p'_1, p'_2, p'_3 der Ebene, so ist es zu jedem Punkt p_4 , der von den Punkten p_1, p_2, p_3 dieselben Abstände hat, welche ein Punkt p'_4 der Ebene, der weder mit den drei Punkten p'_1, p'_2, p'_3 auf einem Kreis noch mit zweien dieser Punkte auf einer Geraden liegt, von den Punkten p'_1, p'_2, p'_3 hat (und nur zu einem so gelegenen Punkt), möglich, eindeutig Abstände eines Punktes p_5 von p_1, p_2, p_3, p_4 zu definieren, so daß die Punkte p_1, p_2, p_3, p_4, p_5 ein pseudoeuklidisches Quintupel bilden: nämlich, indem man $p_5 p_i$ ($i = 1, 2, 3$) gleich dem Abstand des Punktes p'_i und des isogonalen Gegenpunktes von p'_4 hinsichtlich des Dreiecks p'_1, p'_2, p'_3 festsetzt und indem man $p_4 p_5$ gleich dem Radius des Kreises durch die Spiegelpunkte von p'_4 hinsichtlich der drei Geraden durch p'_1, p'_2, p'_3 setzt.

Ein pseudoeuklidisches Quintupel ist im allgemeinen nicht abstandstreu in eine Kugeloberfläche mit Bogenmaß einbettbar, wie man schon an dem Quintupel ersieht, welches sich aus dem Beispiel von Satz 2 für $n = 2$ ergibt.

Für $n > 2$ ist die Abbildung des R_n auf sich selbst, welche jedem Punkt den Mittelpunkt des Spiegelsimplexes hinsichtlich eines vorgegebenen Fundamentalsimplexes zuordnet, eine involutorische Cremonatransformation n -ten Grades, welche abgesehen von den Seitenebenen des Fundamentalsimplexes und abgesehen von der Ausnahmshyperfläche des R_n hinsichtlich der Fundamentalpunkte eineindeutig ist. Die Ausnahmshyperfläche ist von n -tem Grad und reduziert sich für $n > 2$ im allgemeinen nicht auf die dem Fundamentalsimplex umgeschriebene Sphäre.

6. Über Systeme, welche mehr als $n + 3$ Punkte enthalten.

Während für Systeme, die aus $n + 3$ Punkten bestehen, aus der Einbettbarkeit von je $n + 2$ Punkten in den R_n mit Rücksicht auf den Ausnahmefall pseudo-euklidischer $(n + 3)$ -Tupel die Einbettbarkeit des ganzen Systems in den R_n nicht gefolgert werden kann, beweisen wir nunmehr:

Satz 5. Enthält der Raum R mehr als $n + 3$ Punkte und sind je $n + 2$ von ihnen abstandstreu in den R_n einbettbar, dann kann R abstandstreu in den R_n eingebettet werden.

Wir setzen voraus, daß je $n + 2$ von den mindestens $n + 4$ Punkten von R abstandstreu in den R_n einbettbar sind. Auf Grund von Satz 1 genügt es zum Beweise unserer Behauptung, wenn wir zeigen, daß je $n + 3$ Punkte von R in den R_n abstandstreu eingebettet werden können, oder was gleichbedeutend ist (da je $n + 2$ Punkte von R in den R_n abstandstreu einbettbar sind), daß R keine $n + 3$ Punkte in pseudo-euklidischer Lage enthält. Um dies nachzuweisen, leiten wir einen Widerspruch her aus dem Zusammenbestehen der beiden folgenden Annahmen:

Annahme 1. R enthält $n + 3$ Punkte p_1, \dots, p_{n+3} in pseudo-euklidischer Lage.

Annahme 2. R enthält einen Punkt p_{n+4} , welcher von allen $n + 3$ erwähnten Punkten verschieden ist.

Wir bestimmen zunächst zu den Punkten p_1, \dots, p_{n+3} im R_n ein System von Punkten p'_1, \dots, p'_{n+3} , so daß gilt:

Annahme 3. $p'_i p'_k = p_i p_k$ ($i, k = 1, \dots, n + 3$, ausgenommen $i = n + 2, k = n + 3$ und $i = n + 3, k = n + 2$).

Nach Satz 4 liegt p'_{n+3} im Mittelpunkt der Sphäre S , welche bestimmt wird durch die $n + 1$ Spiegelbilder des Punktes p'_{n+2} an den Hyperebenen H_i , wobei H_i die durch die Punkte $p'_1, \dots, p'_{i-1}, p'_{i+1}, \dots, p'_{n+1}$ bestimmte Hyperebene bezeichnet. Der Abstand $p_{n+2} p_{n+3}$ ist gleich dem Radius von S . Wir bezeichnen mit p'_{n+4} den einzigen Punkt des R_n , welcher für $i = 1, \dots, n + 1$ von p'_i den Abstand $p_{n+4} p_i$ hat, der also genügt der

Annahme 4. $p'_{n+4} p'_i = p_{n+4} p_i$ ($i = 1, \dots, n + 1$).

Je n von den $n + 1$ Hyperebenen H_i haben genau einen Punkt gemein, und zwar einen der Punkte p'_i . Der Punkt p'_{n+4} ist nach Annahme 2 von

den $n + 1$ Punkten p'_i verschieden, liegt also in höchstens $n - 1$ von den Hyperebenen H_i . Es gibt also mindestens zwei unter den H_i , welche den Punkt p'_{n+4} nicht enthalten. Wir denken die Numerierung der Punkte p_1, \dots, p_{n+1} so gewählt, daß p'_{n+4} weder in H_1 noch in H_{n+1} liegt. Es ist dann also erfüllt:

Annahme 5. Keines der beiden $(n + 1)$ -Tupel $p'_2, \dots, p'_{n+1}, p'_{n+4}$ und $p'_1, \dots, p'_n, p'_{n+4}$ liegt in einer Hyperebene.

Es sind nun zwei Fälle zu unterscheiden.

Fall I. Es gilt $p'_{n+4} \neq p'_{n+2}$ und $p'_{n+4} \neq p'_{n+3}$.

Nach Satz 4 existiert zu p'_{n+3} außer p'_{n+2} kein Punkt des R_n , dessen Abstand von p'_{n+3} verschieden wäre von dem Abstand, den sein Urbild in R von p_{n+3} hat. Da nach Annahme 1 p'_{n+4} nicht mit p'_{n+2} zusammenfällt, ist also $p'_{n+2} p'_{n+4} = p_{n+2} p_{n+4}$ und aus analogen Gründen gilt $p'_{n+3} p'_{n+4} = p_{n+3} p_{n+4}$. Statt der Annahme 4 können wir also im Fall I zugrunde legen die weitergehende

Annahme 6. $p'_{n+4} p'_i = p_{n+4} p_i$ ($i = 1, \dots, n + 3$).

Wir behaupten nun zunächst: Die $n + 3$ Punkte p_2, \dots, p_{n+4} bilden ein pseudoeuklidisches $(n + 3)$ -Tupel. Da nach Annahme 1 je $n + 2$ Punkte von R abstandstreu in den R_n einbettbar sind, haben wir nur nachzuweisen, daß die $n + 3$ Punkte p_2, \dots, p_{n+4} nicht abstandstreu in den R_n eingebettet werden können. Betrachten wir zu diesem Zweck die Punkte $p'_2, \dots, p'_{n+1}, p'_{n+4}$. Da dieselben nach Annahme 5 nicht in einer Hyperebene des R_n liegen, so existiert im R_n höchstens ein Punkt, welcher für $i = 2, \dots, n + 1, n + 4$ von p'_i den Abstand $p_{n+2} p_i$ hat. Dieser einzige Punkt ist p'_{n+2} , denn er erfüllt, wie man aus den Annahmen 3 und 6 ablesen kann, die geforderten Bedingungen. Aus analogen Gründen ist p'_{n+3} der einzige Punkt des R_n , welcher den Bedingungen genügt, $p'_{n+3} p'_i = p_{n+3} p_i$ ($i = 2, \dots, n + 1, n + 4$). Nun ist aber $p'_{n+2} p'_{n+3} \neq p_{n+2} p_{n+3}$. Also sind die $n + 3$ Punkte p_2, \dots, p_{n+4} nicht abstandstreu in den R_n einbettbar und die Punkte p_2, \dots, p_{n+4} bilden ein pseudoeuklidisches $(n + 3)$ -Tupel, wie behauptet.

Fall II. Es gilt entweder $p'_{n+4} = p'_{n+2}$ oder $p'_{n+4} = p'_{n+3}$.

Dann gilt offenbar auch $p'_{n+2} = p'_{n+3}$. Die Punkte $p'_2, p'_3, \dots, p'_{n+1}$ liegen nicht in einer $(n - 2)$ -dimensionalen Ebene. Also gibt es außer den Punkten p'_{n+2} und p'_{n+3} keinen Punkt, der für $i = 2, \dots, n + 1$ von p'_i den Abstand $p_i p_{n+4} = p'_i p'_{n+2}$ hätte. Die $n + 3$ Punkte p_2, \dots, p'_{n+4} sind mithin nicht in den R_n einbettbar und bilden also auch im Fall II ein pseudoeuklidisches $(n + 3)$ -Tupel.

Wir führen nun das Zusammenbestehen 1 und 2 ad absurdum, indem wir einen Widerspruch herleiten gegen das Zusammenbestehen der Annahme 1, daß die Punkte p_1, \dots, p_{n+3} ein pseudoeuklidisches $(n + 3)$ -Tupel

bilden, und der eben bewiesenen Tatsache, daß die Punkte p_2, \dots, p_{n+4} ein pseudoeuklidisches $(n+3)$ -Tupel bilden. Wir nennen zu diesem Zweck \bar{H}_i ($i=2, \dots, n+1$) die Hyperebene durch die Punkte $p'_2, \dots, p'_{i-1}, p'_{i+1}, \dots, p'_{n+1}, p'_{n+4}$ und \bar{H}_{n+4} die Hyperebene durch die Punkte p'_2, \dots, p'_{n+1} . Wir bezeichnen ferner mit \bar{p}_{n+3}^i ($i=2, \dots, n+1, n+4$) den zu p'_{n+3} hinsichtlich \bar{H}_i symmetrischen Punkt. Da die Punkte p_2, \dots, p_{n+4} ein pseudoeuklidisches $(n+3)$ -Tupel bilden, so muß der Punkt p'_{n+3} gleich weit von den $n+1$ Punkten \bar{p}_{n+3}^i entfernt sein. Offenbar ist $\bar{H}_{n+4} = H_1$ und mithin $\bar{p}_{n+3}^{n+4} = p_{n+3}^1$. Es müssen also die $n+1$ Punkte \bar{p}_{n+3}^i auf der Sphäre S liegen, welche auch die $n+1$ Punkte p_{n+3}^i trägt.

Damit sind wir bei einem Widerspruch angelangt. Nach Annahme 5 enthält weder H_1 noch H_{n+1} den Punkt p'_{n+4} . Die Ebene \bar{H}_{n+1} , welche durch die Punkte $p'_2, \dots, p'_n, p'_{n+4}$ bestimmt wird, ist daher sowohl von H_1 , als auch von H_{n+1} verschieden. Mithin ist der Punkt \bar{p}_{n+3}^{n+1} sowohl von p_{n+3}^1 als auch von p_{n+3}^{n+1} verschieden. Betrachten wir nun das $(n-1)$ -dimensionale Simplex, welches durch die Punkte $p'_2, \dots, p'_n, p'_{n+3}$ bestimmt wird, und lassen wir dasselbe um die $(n-2)$ -dimensionale Ebene durch die Punkte p'_2, \dots, p'_n als Achse rotieren! Die Kreislinienbahn, welche der Punkt p'_{n+3} bei dieser Bewegung beschreibt, schneidet die Sphäre S in zwei Punkten. Diese beiden Punkte sind die einzigen Punkte von S , welche Spiegelbilder von p'_{n+3} hinsichtlich einer Hyperebene durch die Punkte p'_2, \dots, p'_n sein können. Mehr als zwei solcher Punkte kann S also nicht enthalten. Die Punkte p_{n+3}^1 und p_{n+3}^{n+1} sind zwei solche auf S gelegene Spiegelbilder von p'_{n+3} , nämlich hinsichtlich der Hyperebenen H_1 und H_{n+1} und diese beiden Punkte sind untereinander verschieden. Auch \bar{p}_{n+3}^{n+1} ist ein auf S gelegenes Spiegelbild von p'_{n+3} , nämlich hinsichtlich der Hyperebene \bar{H}_{n+1} , und muß daher entweder mit p_{n+3}^1 oder mit p_{n+3}^{n+1} zusammenfallen. Das ist aber ein Widerspruch, denn wir haben eben aus Annahme 5 hergeleitet, daß \bar{p}_{n+3}^{n+1} sowohl von p_{n+3}^1 , als auch von p_{n+3}^{n+1} verschieden ist. Damit ist Satz 5 bewiesen.

Wir können einen Teil unserer bisherigen Ergebnisse zusammenfassen, indem wir folgende Bezeichnungweise einführen: Ein halbmetrischer Raum R heiße **pseudoeuklidisch vom Grade n** , wenn er nicht in den R_n abstandstreu einbettbar ist, obwohl je $n+2$ Punkte von R in den R_n abstandstreu einbettbar sind.

Es gilt dann auf Grund unserer bisherigen Resultate folgendes eigenartige

Theorem I. *Jeder vom Grade n pseudoeuklidische Raum enthält genau $n+3$ Punkte. Für jedes $n > 0$ existieren vom Grad n pseudoeuklidische Räume, welche genau $n+3$ Punkte enthalten.*

7. Ein Kriterium für die Einbettbarkeit von $n + 3$ Punkten in den R_n .

Wir leiten nun eine Bedingung her, die von euklidischen $(n + 3)$ -Tupeln erfüllt und von pseudo-euklidischen $(n + 3)$ -Tupeln nicht erfüllt wird. Sind $n + 3$ Punkte p_1, \dots, p_{n+3} vorgelegt, so zeigen wir, daß $D(p_1, \dots, p_{n+3}) = 0$ eine Bedingung der gewünschten Art ist¹¹⁾.

1. Sind die $n + 3$ Punkte in den R_n abstandstreu einbettbar, so auch in eine n -dimensionale Ebene des R_{n+2} , also auch in eine Hyper-ebene des R_{n+2} , und bestimmen daselbst ein $(n + 2)$ -dimensionales Simplex mit dem Volumen 0. Also gilt nach der Volumensformel $D(p_1, \dots, p_{n+3}) = 0$.

2. Wir wollen zeigen, daß, falls die Punkte p_1, \dots, p_{n+3} ein pseudo-euklidisches $(n + 3)$ -Tripel bilden, $D(p_1, \dots, p_{n+3}) \neq 0$ gilt. Wir ersetzen zu diesem Zweck in der Beziehung $D(p_1, \dots, p_{n+3}) = 0$ den Term $p_{n+2} p_{n+3}$ durch x und bezeichnen die so entstehende Gleichung in x^2 mit $D(x^2) = 0$. Die Gleichung $D(x^2) = 0$ kann nicht identisch verschwinden, denn der Koeffizient von x^4 ist $-D(p_1, \dots, p_{n+1})$, und diese Größe ist $\neq 0$. Da nämlich die Punkte p_1, \dots, p_{n+3} nach Annahme ein pseudo-euklidisches $(n + 3)$ -Tupel bilden, so sind die Punkte p_1, \dots, p_{n+1} in den R_n , aber nicht in den R_{n-1} abstandstreu einbettbar, und ihre Bildpunkte im R_n bestimmen ein n -dimensionales Simplex, dessen Volumen $\neq 0$ ist, so daß nach der Volumensformel $D(p_1, \dots, p_{n+1}) \neq 0$ ist. Die Gleichung $D(x^2) = 0$ hat also zwei Lösungen, die wir mit x^2 und \bar{x}^2 bezeichnen wollen. Unsere Behauptung ist bewiesen, wenn wir zeigen, daß der pseudo-euklidische Abstand $p_{n+2} p_{n+3}$ die Gleichung $D(x^2) = 0$ nicht befriedigt.

Um dies zu zeigen, betrachten wir in einer n -dimensionalen Ebene J des R_{n+2} ein System von Punkten p'_1, \dots, p'_{n+1} , so daß $p'_i p'_k = p_i p_k$ ($i, k = 1, \dots, n + 1$) gilt. Ferner betrachten wir eine Hyperebene H des R_{n+2} , welche J enthält. In H , aber außerhalb von J , wählen wir zwei Punktfolgen $\{p_{n+2}^k\}$ und $\{p_{n+3}^k\}$ ($k = 1, 2, \dots$ ad inf.), so daß $\lim_{k=\infty} p_{n+2}^k = p'_{n+2}$ und $\lim_{k=\infty} p_{n+3}^k = p'_{n+3}$ ist. Es gilt dann offenbar:

$$\begin{aligned}
 (\dagger) \quad & \lim_{k=\infty} p_{n+2}^k p'_i = p'_{n+2} p'_i, & \lim_{k=\infty} p_{n+3}^k p'_i = p'_{n+3} p'_i, \\
 & \lim_{k=\infty} p_{n+2}^k p_{n+3}^k = p'_{n+2} p'_{n+3}.
 \end{aligned}$$

Da für jedes k die Punkte $p'_1, \dots, p'_{n+1}, p_{n+2}^k, p_{n+3}^k$ in H liegen, so gilt nach der Volumensformel für jedes k die Beziehung $D(p'_1, \dots, p'_{n+1}, p_{n+2}^k, p_{n+3}^k) = 0$.

¹¹⁾ Übrigens ließe sich diese Bedingung, obwohl ihre Gültigkeit für euklidische $(n + 3)$ -Tupel nicht nur hinreichend, sondern auch *notwendig* ist, vermutlich durch eine wesentlich schwächere Bedingung ersetzen, die bereits hinreichend wäre.

Wir ersetzen in der k -ten dieser Beziehungen den Term $p_{n+2}^k p_{n+3}^k$ durch x und bezeichnen die so entstehende Gleichung in x^2 mit $D_k(x^2) = 0$. Sie hat zwei Lösungen, nämlich $x_k^2 = (p_{n+2}^k p_{n+3}^k)^2$ und $\bar{x}_k^2 = (p_{n+2}^k \bar{p}_{n+3}^k)^2$, wobei \bar{p}_{n+3}^k das Spiegelbild des Punktes p_{n+3}^k hinsichtlich J in H bezeichnet. Es ist

$$\lim_{k=\infty} |p_{n+2}^k p_{n+3}^k - p_{n+2}^k \bar{p}_{n+3}^k| = \lim_{k=\infty} |x_k - \bar{x}_k| = 0.$$

Die Diskriminante der Gleichung $D_k(x^2) = 0$, d. h. das Quadrat von $|x_k^2 - \bar{x}_k^2|$ hängt stetig von den Größen $p_i' p_k'$ ab. Also gilt wegen (†) für die beiden Lösungen x^2 und \bar{x}^2 der Gleichung $D(x^2) = 0$ $|x^2 - \bar{x}^2| = \lim_{k=\infty} |x_k^2 - \bar{x}_k^2|$, und daher ist $|x^2 - \bar{x}^2| = 0$. Die Gleichung $D(x^2) = 0$ hat also eine Doppelwurzel. Eine Lösung von $D(x^2) = 0$ ist aber sicher $p_{n+2}' p_{n+3}'$. Dies ist also die Doppelwurzel der Gleichung, und eine davon verschiedene Lösung besitzt $D(x^2) = 0$ nicht. Da $p_{n+2} p_{n+3} \neq p_{n+2}' p_{n+3}'$ gilt, ist also $D((p_{n+2}' p_{n+3}')^2) = D(p_1, \dots, p_{n+3}) \neq 0$, wie behauptet. Wir können mithin sagen:

Satz 6. *Ein System von $n + 3$ Punkten p_1, \dots, p_{n+3} ist abstandstreu in den R_n dann und nur dann einbettbar, wenn je $n + 2$ Punkte abstandstreu in den R_n eingebettet werden können und die Beziehung gilt: $D(p_1, \dots, p_{n+3}) = 0$.*

8. Zurückführung des Problems auf die Einbettung von $n + 1$ Punkten in den R_n .

Durch die Sätze 5 und 6 ist unser Problem zurückgeführt auf die Frage nach den Bedingungen, unter denen $n + 2$ Punkte abstandstreu in den R_n eingebettet werden können. Wir beweisen diesbezüglich einen Satz, welcher das Problem weiter reduziert, nämlich auf die Frage nach den Bedingungen für die Einbettbarkeit von $n + 1$ Punkten.

Satz 7. *Die $n + 2$ Punkte p_1, \dots, p_{n+2} sind abstandstreu in den R_n dann und nur dann einbettbar, wenn je $n + 1$ Punkte abstandstreu in den R_n eingebettet werden können und die Beziehung gilt: $D(p_1, \dots, p_{n+2}) = 0$.*

Die Notwendigkeit der Bedingung bildet den Inhalt von Satz 3. Daß die Bedingung auch hinreichend ist, zeigen wir nun mit Hilfe einer bereits im vorangehenden verwendeten Methode.

Wir nehmen an, es liegen $n + 2$ Punkte p_1, \dots, p_{n+2} vor, von denen je $n + 1$ in den R_n abstandstreu eingebettet werden können und die der Bedingung $D(p_1, \dots, p_{n+2}) = 0$ genügen. Es sind zwei Fälle möglich:

A. *Entweder* sind je $n + 1$ von den $n + 2$ Punkten abstandstreu in den R_{n-1} einbettbar. Dann können alle $n + 2$ Punkte abstandstreu in den R_{n-1} eingebettet werden. Denn da je $n + 1$ von ihnen in den R_{n-1}

eingebettet werden können, so müßten die $n + 2$ Punkte, falls sie nicht abstandstreu in den R_{n-1} eingebettet werden könnten, ein pseudoeuklidisches $(n + 2)$ -Tupel bilden. Dann wäre aber die Bedingung $D(p_1, \dots, p_{n+2}) = 0$ nach Satz 6 nicht erfüllt.

B. Oder es existieren $n + 1$ unter den $n + 2$ Punkten, etwa p_1, \dots, p_{n+1} , die nicht abstandstreu in den R_{n-1} eingebettet werden können. Dann bestimmen wir im R_n ein System von Punkten p'_1, \dots, p'_{n+1} , so daß $p'_i p'_k = p_i p_k$ gilt ($i, k = 1, \dots, n + 1$). Die Punkte p'_1, \dots, p'_{n+1} liegen nicht in einer Hyperebene des R_n . Wir ersetzen in der Beziehung $D(p_1, \dots, p_{n+2}) = 0$ den Term $p_{n+1} p_{n+2}$ durch x und bezeichnen die so entstehende quadratische Gleichung in x^2 mit $D(x^2) = 0$. Diese Gleichung kann nicht identisch verschwinden, denn der Koeffizient von x^4 ist $-D(p_1, \dots, p_{n-1}) = -D(p'_1, \dots, p'_{n-1})$, und diese Größe ist $\neq 0$. Denn wäre sie $= 0$, so lägen die Punkte p'_1, \dots, p'_{n-1} in einer $(n - 2)$ -dimensionalen Ebene des R_n , was unmöglich ist, da sonst die Punkte p'_1, \dots, p'_n in einer Hyperebene lägen, im Widerspruch zur Voraussetzung. Die Gleichung $D(x^2) = 0$ hat also höchstens zwei Lösungen. Die Gleichung bleibt unverändert, wenn man in ihr $p_i p_k$ durch $p'_i p'_k$ ($i, k = 1, \dots, n + 2$) ersetzt. Bezeichnet \bar{p}'_{n+2} das Spiegelbild von p'_{n+2} hinsichtlich der Hyperebene durch die Punkte p'_1, \dots, p'_n , so gilt $D(p'_1, \dots, p'_{n+1}, p'_{n+2}) = 0$ und $D(p'_1, \dots, p'_{n+1}, \bar{p}'_{n+2}) = 0$. Es sind also $(p'_{n+1} p'_{n+2})^2$ und $(p'_{n+1} \bar{p}'_{n+2})^2$ die Lösungen der Gleichung $D(x^2) = 0$. Einem dieser beiden Werte muß $(p_{n+1} p_{n+2})^2$ gleich sein. In beiden Fällen aber existiert ein zu den Punkten p_1, \dots, p_{n+2} abstandsgleiches System von Punkten im R_n , nämlich $p'_1, \dots, p'_{n+1} p'_{n+2}$ bzw. $p'_1, \dots, p'_{n+1}, \bar{p}'_{n+2}$. Damit ist Satz 7 bewiesen.

9. Eine notwendige und hinreichende Bedingung für die Einbettbarkeit von $n + 1$ Punkten in den R_n .

Durch Satz 7 ist unser Problem zurückgeführt auf die Frage nach den Bedingungen für die Einbettbarkeit von $n + 1$ Punkten in den R_n . Wir führen folgende Bezeichnungen ein: Ist die reelle Zahl a positiv oder Null, also nicht negativ, so schreiben wir $\text{sign } a = \text{sign}(-1)$; ist a nicht positiv, so schreiben wir $\text{sign } a = \text{sign}(+1)$. Wir beweisen nun:

Satz 8. *Notwendig und hinreichend, damit die $n + 1$ Punkte p_1, \dots, p_{n+1} abstandstreu in den R_n einbettbar seien, ist, daß je n Punkte abstandstreu in den R_{n-1} eingebettet werden können und daß $\text{sign } D(p_1, \dots, p_{n+1}) = \text{sign}(-1)^n$ gilt.*

Die Notwendigkeit der Bedingung folgt aus der Volumsformel (S. 119). In der Tat, wenn $n + 1$ Punkte p_1, \dots, p_{n+1} abstandstreu in den R_n einbettbar sind, so bestimmen ihre Bildpunkte im R_n , p'_1, \dots, p'_{n+1} , ein

n -dimensionales Simplex mit dem Volumen $v(p'_1, \dots, p'_{n+1})$. Der Volumensformel zufolge ist das Produkt von $D(p'_1, \dots, p'_{n+1}) = D(p_1, \dots, p_{n+1})$ mit $(-1)^{n+1}$ und einem positiven Faktor gleich der positiven Zahl $v^2(p'_1, \dots, p'_{n+1})$. Es kann mithin $\text{sign } D(p_1, \dots, p_{n+1})$ nicht $= \text{sign } (-1)^n$ sein.

Um zu zeigen, daß die Bedingung auch hinreichend ist, betrachten wir $n+1$ Punkte p_1, \dots, p_{n+1} , von denen je n abstandstreu in den R_{n-1} einbettbar sind. Es sind zwei Fälle möglich:

A. Entweder sind je n von den Punkten abstandstreu in den R_{n-2} einbettbar. Dann bilden die $n+1$ Punkte entweder ein euklidisches oder ein pseudo-euklidisches $(n+1)$ -Tupel. Im ersteren Fall sind sie abstandstreu in den R_{n-2} , also a fortiori in den R_n einbettbar. Der zweite Fall kann, wie wir zeigen wollen, nicht eintreten, falls die Bedingung $\text{sign } D(p_1, \dots, p_{n+1}) \neq \text{sign } (-1)^n$ erfüllt ist. Mit anderen Worten: wir wollen zeigen: Falls die Punkte p_1, \dots, p_{n+1} ein pseudo-euklidisches $(n+1)$ -Tupel bilden, so ist $\text{sign } D(p_1, \dots, p_{n+1}) = \text{sign } (-1)^n$ ¹²⁾.

Wir ersetzen im Ausdruck $D(p_1, \dots, p_{n+1})$ den Term $p_n p_{n+1}$ durch x und bezeichnen die so entstehende quadratische Funktion von x^2 mit $D(x^2)$. Aus Satz 6 und seinem Beweis geht hervor, daß die Gleichung $D(x^2) = 0$ eine von $(p_n p_{n+1})^2$ verschiedene Zahl als Doppelwurzel besitzt. Das Vorzeichen des Funktionswertes $D(x^2)$ ist also für alle Werte von x^2 , ausgenommen die eine Nullstelle, gleich. Für sehr große Werte von x^2 stimmt das Vorzeichen des Funktionswertes mit dem Vorzeichen des Koeffizienten von x^4 überein. Dieser Koeffizient ist $-D(p_1, \dots, p_{n-1})$. Da die Punkte p_1, \dots, p_{n+1} ein pseudo-euklidisches $(n+1)$ -Tupel bilden, so sind die Punkte p_1, \dots, p_{n-1} in den R_{n-2} , aber nicht in den R_{n-3} abstandstreu einbettbar. Wir können mithin im R_{n-2} ein zu den Punkten p_1, \dots, p_{n-1} abstandsgleiches System von Punkten p'_1, \dots, p'_{n-1} angeben, die nicht in einer Hyperebene des R_{n-2} liegen und mithin ein $(n-2)$ -dimensionales Simplex mit nicht-verschwindendem Volumen bestimmen. Wegen der bereits bewiesenen Notwendigkeit der Bedingung von Satz 8 ist also $\text{sign } D(p'_1, \dots, p'_{n-1}) = \text{sign } D(p_1, \dots, p_{n-1}) = \text{sign } (-1)^{n-1}$. Das Vorzeichen des Koeffizienten von x^4 und mithin das Vorzeichen von $D(x^2)$ für hinreichend große x^2 ist folglich $= \text{sign } (-1)^n$. Demnach ist für alle x^2 , ausgenommen die Nullstelle von $D(x^2)$, $\text{sign } D(x^2) = (-1)^n$. Wenn also $\text{sign } D(p_1, \dots, p_{n+1}) \neq \text{sign } (-1)^n$ vorausgesetzt wird, so kann der Fall eines pseudo-euklidischen $(n+1)$ -Tupels nicht vorliegen und die $n+1$ Punkte sind im Fall A abstandstreu in den R_n einbettbar.

¹²⁾ [Zusatz bei der Korrektur]: Daraus (und übrigens schon aus den Überlegungen von S. 131) geht zugleich hervor, daß ein pseudo-euklidisches Punktsystem in keinem euklidischen Raum irgend einer Dimension abstandstreu einbettbar ist.

B. Oder es sind n von den $n + 1$ Punkten, etwa die Punkte p_1, \dots, p_n , nicht abstandstreu in den R_{n-2} einbettbar. Wir betrachten im R_n eine Hyperebene H und bestimmen in ihr ein zu den Punkten p_1, \dots, p_n abstandsgleiches System von Punkten p'_1, \dots, p'_n . Da die Punkte p'_1, \dots, p'_n nicht in einer $(n - 2)$ -dimensionalen Ebene des R_n liegen, so liegen die Punkte p'_1, \dots, p'_{n-1} nicht in einer $(n - 3)$ -dimensionalen Ebene. Die Punkte $p_1, \dots, p_{n-1}, p_{n+1}$ sind abstandstreu in den R_{n-1} einbettbar. Wir können also in irgendeiner die Punkte p'_1, \dots, p'_{n-1} enthaltenden Hyperebene einen Punkt p' bestimmen, der für $i = 1, \dots, n - 1$ von p'_i den Abstand $p_{n+1} p_i$ hat. Wir lassen nun das $(n - 1)$ -dimensionale Simplex, welches durch die Punkte $p'_1, \dots, p'_{n-1}, p'$ bestimmt wird, um die $(n - 2)$ -dimensionale Ebene durch die Punkte p'_1, \dots, p'_{n-1} als Achse rotieren. Die Kreislinie K , welche der Punkt p' bei dieser Bewegung beschreibt, ist die Menge aller Punkte, die für $i = 1, \dots, n - 1$ von p'_i den Abstand $p_{n+1} p_i$ haben. K hat mit H zwei Punkte gemein, die wir \bar{p} und $\bar{\bar{p}}$ nennen wollen. (Die beiden fallen dann und nur dann zusammen, wenn die Punkte $p_1, \dots, p_{n-1}, p_{n+1}$ in den R_{n-2} abstandstreu einbettbar sind.) Da K auf H senkrecht steht, so sind \bar{p} und $\bar{\bar{p}}$ diejenigen unter den Punkten von K , welche vom Punkt p'_n den kleinsten und den größten Abstand haben. Ist etwa $p'_n \bar{p} \leq p'_n \bar{\bar{p}}$, so existiert für jeden Wert x , welcher den Ungleichungen $p'_n \bar{p} \leq x \leq p'_n \bar{\bar{p}}$ genügt, ein Punkt p'_{n+1} von K , der den Abstand x von p'_n hat, und es existieren dann also im R_n $n + 1$ Punkte p'_1, \dots, p'_{n+1} , so daß $p'_i p'_k = p_i p_k$ ($i, k = 1, \dots, n + 1$) gilt.

Wir betrachten nun für unsere $n + 1$ Punkte den Ausdruck $D(p_1, \dots, p_{n+1})$. Wir ersetzen in ihm den Term $p_n p_{n+1}$ durch x und bezeichnen die so entstehende quadratische Funktion von x^2 mit $D(x^2)$. Nach der Volumensformel ist $D(x^2)$ dann und nur dann $= 0$, wenn die Punkte p_1, \dots, p_{n+1} abstandstreu in den R_{n-1} einbettbar sind. Durchläuft p' die Menge K , so verschwindet $D(p'_1, \dots, p'_n, p')$ für $p' = \bar{p}$ und $p' = \bar{\bar{p}}$, und nur für diese Werte. Berücksichtigen wir, daß $p'_i p'_k = p_i p_k$ ($i, k = 1, \dots, n + 1$, ausgenommen $i = n, k = n + 1$ und $i = n + 1, k = n$) gilt, so sehen wir, daß $D(x^2)$ als Nullstellen die beiden Werte $(p'_n \bar{p})^2$ und $(p'_n \bar{\bar{p}})^2$ hat. Der Koeffizient von x^4 in $D(x^2) = 0$ ist $-D(p_1, \dots, p_{n-1})$, also ist $D(x^2) = -D(p_1, \dots, p_{n-1}) \cdot [x^2 - (p'_n \bar{p})^2] \cdot [x^2 - (p'_n \bar{\bar{p}})^2]$. Es gilt $\text{sign } D(p_1, \dots, p_{n-1}) = \text{sign } (-1)^{n-1}$. Wenn also $\text{sign } D(x^2) \neq (-1)^n$ ist, dann muß $\text{sign} [(x^2 - (p'_n \bar{p})^2)(x^2 - (p'_n \bar{\bar{p}})^2)] \neq \text{sign } (+1)$ sein, also $(p'_n \bar{p})^2 \leq x^2 \leq (p'_n \bar{\bar{p}})^2$ gelten. Falls für unsere $n + 1$ Punkte die Voraussetzung gilt: $\text{sign } D(p_1, \dots, p_{n+1}) \neq \text{sign } (-1)^n$, dann ist also $(p'_n \bar{p})^2 \leq (p_n p_{n+1})^2 \leq (p'_n \bar{\bar{p}})^2$, und dann sind nach dem Bewiesenen die

Punkte p_1, \dots, p_{n+1} abstandstreu in den R_n einbettbar. Damit ist Satz 8 bewiesen.

10. Notwendige und hinreichende Bedingungen für die Einbettbarkeit eines Raumes in den R_n .

Die bisher bewiesenen Sätze zusammengenommen ergeben das

Theorem II. *Damit ein halbmétrischer Raum R abstandstreu in den R_n ($n \geq 0$) einbettbar sei, ist notwendig und hinreichend die Gültigkeit folgender Bedingungen:*

(B_{n+2}^n) : Für je $n+2$ Punkte von R gilt $D(p_1, \dots, p_{n+2}) = 0$.

(B_k) : Für jedes $k \leq n+2$ und für je k Punkte von R gilt $\text{sign } D(p_1, \dots, p_k) \neq \text{sign } (-1)^{k-1}$.

Enthält R genau $n+3$ Punkte, dann wird die Einbettbarkeit in den R_n durch die Bedingungen B_{n+2}^n und B_k ($k \leq n+2$) nur gewährleistet, wenn noch die zugleich auch notwendige Bedingung erfüllt ist:

(B_{n+3}^n) : $D(p_1, \dots, p_{n+3}) = 0$.

Damit ein in den R_n abstandstreu einbettbarer Raum nicht in den R_{n-1} abstandstreu einbettbar sei, ist notwendig und hinreichend die Bedingung

(\bar{B}^{n-1}) : R enthält $n+1$ Punkte, für die $D(p_1, \dots, p_{n+1}) \neq 0$ ist.

Man bemerkt, daß keinerlei Voraussetzungen über Separabilität oder auch nur über die Mächtigkeit von R notwendig sind.

Eine Betrachtung der Bedingungen unseres Theorems zeigt auch, warum die Aussage nicht nur für métrische, sondern für beliebige halbmétrische Räume gilt, mit anderen Worten: warum die Gültigkeit der Dreiecksungleichung nicht eigens vorausgesetzt werden muß. Für einen Raum, der mit einer Teilmenge eines R_n abstandsgleich ist, gilt die Dreiecksungleichung von selbst. Andererseits wird sie auch durch die Bedingungen unseres Theorems gewährleistet. Denn angenommen, es sei für je drei Punkte von R_n $\text{sign } D(p_1, p_2, p_3) \neq \text{sign } (-1)^2$, also

$$(\dagger) \quad D(p_1, p_2, p_3) \leq 0.$$

Es ist

$$(\dagger\dagger) \quad D(p_1, p_2, p_3) = - (p_1 p_2 + p_1 p_3 + p_2 p_3) (p_1 p_2 - p_1 p_3 + p_2 p_3) \\ (p_1 p_2 - p_1 p_3 + p_2 p_3) (-p_1 p_2 + p_1 p_3 + p_2 p_3),$$

und man bestätigt sofort, daß aus (\dagger) und $(\dagger\dagger)$ folgt, daß jeder der drei letzten Faktoren auf der rechten Seite von $(\dagger\dagger)$ nicht-negativ ist. Die Bedingung B_3 ist also mit der Dreiecksungleichung äquivalent, was übrigens

a priori klar ist, da ja die Dreiecksungleichung nichts anderes ist als die Bedingung für die abstandstreue Einbettbarkeit dreier Punkte des Raumes in die Ebene.

Sogar wenn wir über den Raum R bloß voraussetzen, daß je zweien seiner Punkte x und y eine nicht-negative Zahl $xy = yx$ zugeordnet ist, auch dann ist R auf Grund der Bedingungen unseres Theorems eo ipso ein halbmétrischer und mithin métrischer Raum. Denn B_2^0 und \bar{B}^1 besagen, daß zwei Punkte p_1 und p_2 dann und nur dann abstandstreu in den R_0 einbettbar, d. h. identisch sind, wenn $D(p_1, p_2) = 0$ ist. Nun ist aber $D(p_1, p_2) = 2 \cdot (p_1 p_2)^2$. Es ist also $p_1 p_2 = 0$ für $p_1 = p_2$ und $p_1 p_2 > 0$ für $p_1 \neq p_2$. Die beiden Bedingungen, die man in die Definition des métrischen Raumes neben die Symmetrie und Nichtnegativität der Abstandsfunktion aufnimmt, nämlich die Nullbedingung und die Dreiecksungleichung, stellen sich also als die Bedingungen B_2^0 , \bar{B}^1 und B_3 dar.

Im Fall $n = 1$, d. h. für die abstandstreue Einbettbarkeit eines Raumes R in die Gerade, ist, falls R mehr als 4 Punkte enthält, notwendig und hinreichend die Bedingung B_3^1 , d. i. $D(p_1, p_2, p_3) = 0$. Nach ($\dagger\dagger$) bedeutet dies, daß einer der drei letzten Faktoren auf der rechten Seite von ($\dagger\dagger$) verschwindet. Unter Zugrundelegung der „Zwischen“-Bezeichnung besagt dies, daß von je drei Punkten von R einer zwischen den beiden anderen liegt, also die Biedermannsche Bedingung (siehe Vorbemerkung). Enthält R 4 Punkte, dann muß noch eine Bedingung, etwa B_4^1 , d. i. $D(p_1, p_2, p_3, p_4) = 0$, hinzutreten, um die pseudoeuklidischen Quadrupel auszuschließen.

11. Charakterisierung des R_n , des Halb- R_n und der euklidischen konvexen Körper.

Es sei $\{p_n\}$ ($n = 1, 2, \dots$ ad inf.) eine Folge von Punkten des halbmétrischen Raumes R , so daß für jede Zahl $\varepsilon > 0$ eine natürliche Zahl n existiert mit der Eigenschaft, daß für je zwei natürliche Zahlen $n' > n$ und $n'' > n$ $|p_n p_{n'} - p_n p_{n''}| < \varepsilon$ gilt. Dann heißt die Punktfolge eine *Cauchysche Folge*. Ein halbmétrischer Raum R heißt *vollständig*, wenn R zu jeder Cauchyschen Punktfolge einen Häufungspunkt enthält. In einem euklidischen Raum sind die abgeschlossenen Mengen, und nur diese, als métrische Räume aufgefaßt, vollständig. Notwendig und hinreichend, damit ein in den R_n abstandstreu einbettbarer Raum R abstandsgleich mit einer *abgeschlossenen* Teilmenge des R_n sei, ist demnach, daß R *vollständig* ist.

Ein halbmétrischer Raum R heißt *konvex* (vgl. die erste Untersuchung über allgemeine Metrik), wenn R zu je zweien seiner Punkte a und b einen Punkt c *zwischen* ihnen enthält, so daß $ac + cb = ab$ gilt. Ein vollständig konvexer Raum enthält zu je zweien seiner Punkte (a. a. O. S. 87) eine zwischen den beiden Punkten zusammenhängende Menge von Zwischen-

punkten. Im R_n ist die Menge aller Punkte zwischen den Punkten a und b identisch mit der durch a und b bestimmten Strecke.

Ist also ein vollständiger konvexer Raum in den R_n , aber nicht in den R_{n-1} abstandstreu einbettbar, dann ist die mit R abstandsgleiche Teilmenge des R_n abgeschlossen, sie liegt nicht in einer Hyperebene des R_n und sie enthält zu je zweien ihrer Punkte die durch dieselben bestimmte Strecke. Sie ist also ein konvexer n -dimensionaler Körper im Sinn der klassischen Definition, und zwar ein beschränkter oder unbeschränkter, je nachdem der Raum R kompakt ist (d. h. zu jeder seiner unendlichen Teilmengen ein Häufungselement enthält) oder nicht. Damit ein in den R_n , aber nicht in den R_{n-1} abstandstreu einbettbarer Raum R mit einem konvexen n -dimensionalen Körper abstandsgleich sei, ist also notwendig und hinreichend Vollständigkeit und Konvexität (im Sinne der Konvexitätstheorie) von R .

Satz (†). Enthält die abgeschlossene Menge A des R_n zu je zweien ihrer Punkte p_1 und p_2 1. einen Zwischenpunkt von p_1 und p_2 , 2. einen Punkt p^1 , so daß p_1 zwischen p^1 und p_2 liegt, 3. einen Punkt p^2 , so daß p_2 zwischen p^2 und p_1 liegt, dann enthält die Menge A zu je zweien ihrer Punkte die durch sie bestimmte Gerade als Teilmenge. Enthält die abgeschlossene Menge A des R_n zu je zwei Punkten p_1 und p_2 1. einen Zwischenpunkt von p_1 und p_2 , 2. entweder einen Punkt p^1 , so daß p_1 zwischen p^1 und p_2 liegt, oder einen Punkt p^2 , so daß p_2 zwischen p^2 und p_1 liegt, — dann enthält die Menge A zu je zweien ihrer Punkte mindestens eine der beiden durch sie bestimmten Halbgeraden.

Wir beweisen etwa den zweiten Teil des Satzes (der erste Teil kann ganz analog bewiesen werden). Wäre die Behauptung unrichtig, so enthielte die Menge A zwei Punkte und bloß Punkte eines begrenzten Stückes von der durch sie begrenzten Geraden g , also wegen der Konvexität von A ein abgeschlossenes Intervall von g . Zu den Endpunkten p_1 und p_2 des betreffenden Intervalls enthielte A nach Voraussetzung entweder einen Punkt p^1 oder einen Punkt p^2 , also jedenfalls einen Punkt von g außerhalb des Intervalls. Das ist ein Widerspruch.

Satz (††). Enthält eine abgeschlossene konvexe Menge A des R_n $n+1$ Punkte, welche nicht in einer Hyperebene liegen, und die $\frac{n(n+1)}{2}$ Geraden durch je zwei von diesen Punkten, dann ist A mit dem R_n identisch.

Der Beweis von Satz (††) liegt auf der Hand. Unter Berücksichtigung von Satz (†) folgt aus ihm:

Damit ein in den R_n , aber nicht in den R_{n-1} abstandstreu einbettbarer Raum mit dem R_n abstandsgleich sei, ist notwendig und hinreichend, daß R vollständig ist und daß R zu je zweien seiner Punkte p_1 und p_2

enthält: 1. einen Zwischenpunkt; 2. einen Punkt p^1 , so daß p_1 zwischen p^1 und p_2 liegt; 3. einen Punkt p^2 , so daß p_2 zwischen p^2 und p_1 liegt. Damit ist die Charakterisierung des R_n erledigt.

Satz (†††). Voraussetzung: Es sei A eine abgeschlossene Menge des R_n , welche nicht in einer Hyperebene des R_n liegt, und welche zu je zweien ihrer Punkte p_1 und p_2 enthält: 1. einen Zwischenpunkt; 2. *entweder* einen Punkt p^1 , so daß p_1 zwischen p^1 und p_2 liegt, *oder* einen Punkt p^2 , so daß p_2 zwischen p^2 und p_1 liegt. Behauptung: A ist *entweder* identisch mit dem R_n oder identisch mit einem Halb- R_n .

Die Behauptung ist für $n=1$ eine unmittelbare Folge von Satz (†). Wir nehmen an, sie sei bereits bewiesen für $n-1$. Es sei dann A eine Menge, welche die Voraussetzungen von Satz (†††) erfüllt. A enthält $n+1$ Punkte, p_1, \dots, p_{n+1} , welche nicht in einer Hyperebene liegen. Ist A mit dem R_n identisch, so sind wir am Ziel. Andernfalls ist nach Satz (††) mindestens eine von den $\frac{n(n+1)}{2}$ Geraden durch die Punkte p_i und p_k ($i, k = 1, \dots, n+1$) nicht zur Gänze in A enthalten. Wir denken die Bezeichnung so gewählt, daß die Gerade g durch p_n und p_{n+1} nicht ganz in A enthalten ist.

Es sei \bar{g}_n jene in A enthaltene Halbgerade von g , zu der keine umfassendere in A enthaltene Halbgerade existiert. Es sei e_n der Endpunkt von \bar{g}_n . Mit g_i ($i=1, 2, \dots, n-1$) bezeichnen wir die zu g parallele Gerade durch den Punkt p_i . Auch jede einzelne dieser $n-1$ Geraden g_i kann nicht zur Gänze in A liegen. Denn wäre dies für g_i der Fall, so könnten wir eine in g_i dichte Folge $\{p_k\}$ ($k=1, 2, \dots$ ad inf.) von Punkten bestimmen und alle Punkte p_k gehörten zu A ; dann läge auch für jedes k die Strecke durch p_k und e_n in A und es wäre, wegen der Abgeschlossenheit von A die abgeschlossene Hülle von der Summe aller dieser Strecken Teil von A ; diese enthielte aber $g - \bar{g}_n$, während die letztere Menge zu A fremd vorausgesetzt ist.

Wir können also für $i=1, \dots, n-1$ auf der Geraden g_i einen Punkt e_i bestimmen, so daß die durch e_i begrenzte zu \bar{g}_n gleichgerichtete Halbgerade \bar{g}_i von g_i Teil von A und die andere Halbgerade zu A fremd ist. Die n Punkte e_1, \dots, e_n bestimmen eine Hyperebene; denn lägen sie in einer $(n-2)$ -dimensionalen Ebene, so lägen wegen der Parallelität der Geraden g_i die $n+1$ Punkte p_1, \dots, p_{n+1} in einer Hyperebene, was nach Voraussetzung nicht der Fall ist. Nach dem für $n-1$ als bewiesen angenommenen Satz (†††) ist mindestens eine Hälfte von H , etwa die Halb-Hyperebene H_1 , Teil von A . Jede Halbgerade durch einen Punkt p von H_1 , die zu \bar{g}_n parallel und gleichgerichtet ist, ist daher Teil von A . Denn wählen wir auf \bar{g}_n eine Punktfolge q_k ($k=1, 2, \dots$ ad inf.), so daß

$\lim_{k=\infty} p q_k = \infty$ ist, so ist die abgeschlossene Hülle von der Summe aller Strecken durch p und q_k Teil von A ; diese Menge enthält aber die in Rede stehende Halbgerade. Also ist der eine der beiden an H_1 grenzenden Viertelräume, er heiße V_1 , Teil von A .

Wir zeigen nun, daß das Innere des an H grenzenden zu V_1 fremden Halb- R_n , er heiße $V_{3,4}$, zu A fremd ist. Sei nämlich p ein innerer Punkt von $V_{3,4}$. Es sei q ein innerer Punkt von H_1 derart, daß die Gerade g_n und die Strecke durch p und q in einer Ebene liegen. Würde nun p in A liegen, dann wäre auch die Strecke zwischen p und q Teil von A ; dann enthielte aber A eine \bar{g}_n als echten Teil enthaltende Halbgerade von g_n als Teil und dies widerspricht der Voraussetzung. Also ist p nicht Punkt von A und das Innere von $V_{3,4}$ zu A fremd.

Betrachten wir nun einen Punkt a der $(n-2)$ -dimensionalen Randebene der Halbhyperebene H_1 und die zu \bar{g}_n parallele und gleichgerichtete Halbgerade \bar{g}_a durch a . Auf \bar{g}_a wählen wir eine gegen a konvergente Punktfolge $\{a_k\}$ ($k=1, 2, \dots$ ad inf.). Wir verbinden für jedes k den Punkt a_k mit einem inneren Punkt h von H_1 . Da von jeder Geraden, welche mit A zwei Punkte gemein hat, nach Satz (†) mindestens eine Halbgerade Teil von A ist, und da $V_{3,4}$ zu A fremd ist, so muß für jedes k die Halbgerade von h über a_k ins Unendliche Teil von A sein. Auch die abgeschlossene Hülle von der Summe aller dieser Halbgeraden ist in A enthalten. Diese Menge enthält aber eine Halbgerade der Halbhyperebene $H_2 = H - H_1$, nämlich die Halbgerade von h über a ins Unendliche. Indem wir a die ganze $(n-2)$ -dimensionale Randebene der beiden Halbhyperebenen H_1 und H_2 durchlaufen lassen, erkennen wir, daß ganz H_2 , also ganz H , Teil von A ist. Mithin ist der eine an H grenzende Halb- R_n Teil von A , während das Innere des anderen, nämlich $V_{3,4}$, zu A fremd ist. Falls also A nicht mit dem R_n identisch ist, ist A ein Halb- R_n , womit Satz (†††) bewiesen ist.

Fassen wir die Resultate dieses Abschnittes zusammen, so können wir sagen:

Theorem III. *Damit ein halbmétrischer Raum abstandsgleich sei*
 a) *mit dem R_n , b) mit dem Halb- R_n , c) mit einem n -dimensionalen konvexen Körper des R_n — ist notwendig und hinreichend, daß R vollständig sei, daß R die Bedingungen B_{n+2}^n , B_k ($k \leq n+1$) und \bar{B}^{n-1} von Theorem I erfüllt und daß R zu je zweien seiner Punkte p_1 und p_2 enthält:*

a) *einen Zwischenpunkt, einen Punkt p^1 , so daß p_1 zwischen p^1 und p_2 liegt, und einen Punkt p^2 , so daß p_2 zwischen p^2 und p_1 liegt;*

b) *einen Zwischenpunkt und entweder einen Punkt p^1 oder einen Punkt p^2 von den erwähnten Eigenschaften, aber nicht für je zwei Punkte p_1 und p_2 sowohl einen Punkt p^1 als auch einen Punkt p^2 ;*

c) *einen Zwischenpunkt.*

Die Bedingung B_{n+3}^n muß nicht vorausgesetzt werden, weil ein konvexer Raum sicher unendlich viele Punkte enthält. Aus Stetigkeitsgründen würde es natürlich auch genügen, die Bedingungen B_{n+2}^n und B_k ($k < n+1$) in Theorem II bloß für die Punkte einer abzählbaren in R dichten Menge vorzusetzen.

12. Ausblicke.

Wir sind im vorangehenden geradewegs auf die Lösung des eingangs formulierten Problems losgeschritten und haben manche für dieses Ziel unerhebliche Verschärfung, welche wir den Zwischenresultaten hätten geben können, vernachlässigt. Auch sind wir auf diverse Konsequenzen unserer Ergebnisse für Determinanten und speziell für geränderte symmetrische Determinanten mit Diagonalgliedern Null nicht eingegangen.

Von *geometrischen* Problemen, welche durch die vorangehenden Untersuchungen nahegelegt werden, seien die folgenden erwähnt: Die euklidischen Räume und ihre Teilmengen sind gewiß nicht die einzigen, deren Metrik sich durch Relationen zwischen ihren endlichen Punktesystemen charakterisieren läßt. *Welches sind die anderen Räume, für welche dies möglich ist?* Insbesondere sollte eine solche Charakterisierung für die *Räume von konstanter Krümmung* und deren Teilmengen geliefert werden und eventuell durch Heranziehung von Relationen zwischen abzählbaren Punktesystemen, für den *Hilbertschen Raum*¹³⁾. Ferner geben unsere Überlegungen Mittel zur *systematischen Aufzählung gewisser nicht-euklidischer Metriken* an die Hand, insbesondere auch für den Fall nicht-reeller Abstände. Da weiters die Untersuchung jener Punktetripel eines allgemeinen metrischen Raumes, für welche $D(p_1, p_2, p_3) = 0$ ist, große Fruchtbarkeit gezeigt hat, indem wir auf sie eine Theorie der Konvexität begründen konnten, so dürfte auch eine *Untersuchung jener n -Tupel von Punkten eines metrischen Raumes, für welche $D(p_1, \dots, p_n) = 0$ ist*, zur Erkenntnis gewisser metrischer Eigenschaften beitragen. Endlich lassen sich, wie in einem anderen Aufsatz ausgeführt werden soll, durch diese und ähnliche Überlegungen *gewisse differentialgeometrische Eigenschaften*, wie z. B. die Krümmung von Kurven und vielleicht auch die Krümmung von Flächen erfassen. Sie lassen sich, auch wofern sie nicht im ganzen Raum konstant sind, als *Limesausdrücke in allgemeinen metrischen Räumen definieren und koordinatenlos behandeln.*

¹³⁾ [Zusatz bei der Korrektur]: Vgl. über diese und verwandte Probleme meine inzwischen erschienenen Bemerkungen, Proc. Ac. Amsterdam 30, S. 710, sowie Wiener akad. Anz. 1928, Nr. 12.

Dritte Untersuchung: Entwurf einer Theorie der n -dimensionalen Metrik.

- I. Analyse der Abstandsmetrik.
 1. Stetigkeit der Metrik und Dreiecksungleichung.
 2. Bogenlänge und Dreiecksungleichung.
 3. Vollständigkeit der Dreiecksungleichung.
- II. Die n -dimensionale Metrik.
 1. Die euklidische n -dimensionale Metrik.
 2. Stetigkeit der Metrik und Simplexungleichung.
 3. Allgemeine n -dimensionale Metrik.
 4. Volumslimes und Abstandslimes.
 5. Das Verschwinden der Volumsfunktion.
 6. Das Nichtverschwinden der Volumsfunktion.
 7. Die Vollständigkeit der Simplexungleichung.
 8. Flächenmetrik in Räumen mit Dreitripel- und Zweitripel-eigenschaft.

Die Gültigkeit der Dreiecksungleichung wird für den metrischen Raum stets ohne irgendwelche Interpretation postuliert. Eine notwendige und hinreichende Bedingung für irgendwelche Vernünftigkeitseigenschaften der Metrik scheint die Dreiecksungleichung allerdings nicht zu sein. Doch versuche ich durch einige Bemerkungen, welche den Inhalt des ersten Teiles dieser Untersuchung bilden, sie auf Grund gewisser naturgemäßer Forderungen an die Metrik wenigstens in gewissem Umfang plausibel zu machen.

Im zweiten Teil gebe ich die Grundbegriffe und -tatsachen einer Theorie der n -dimensionalen Metrik an, in welche sich die in der mengentheoretischen Geometrie bisher ausschließlich betrachtete Metrik als ein-dimensionaler Fall, als Abstandsmetrik, einordnet. Die Ausführungen, in welchen das Problem der tatsächlichen Einführung einer den allgemeinen Bedingungen genügenden n -dimensionalen Metrik (das Problem der n -dimensionalen Metrisation) offen gelassen wird, sind als Vorstudien gedacht zu einer umfassenderen Behandlung des großen Problems der Verknüpfung von Maß- und Dimensionsbegriff auf der Basis der mengentheoretischen Geometrie.

I. Analyse der Abstandsmetrik.

1. Stetigkeit der Metrik und Dreiecksungleichung.

Es sei eine Menge gegeben, die wir als Raum und deren Elemente wir als Punkte bezeichnen. Eine ganz primitive (gleichsam bloß qualitative) Abstandsmetrik besteht darin, daß je zwei Paare von Punkten a, b

und c, d miteinander vergleichbar sind in dem Sinn, daß man sagen kann, a und b seien voneinander gleich weit, mehr oder weniger entfernt, wie c und d . Hierbei müssen die Relationen „gleich weit entfernt“, bzw. „mehr entfernt“ und „weniger entfernt“ die Transitivitäts-, und Symmetrie- bzw. Asymmetrie-Eigenschaften der Gleichheits- bzw. Größer- und Kleiner-Relation besitzen. Von sich selbst sollen je zwei Punkte gleich weit entfernt sein (d. h. a und a sind gleich weit entfernt, wie b und b). Dabei soll endlich jeder Punkt von sich selbst am wenigsten entfernt sein, d. h. für jeden Punkt a und jeden von a verschiedenen Punkt $b \neq a$ sind a und b mehr entfernt wie a und a . Durch eine derartige Metrik wird dem Raum eine *Limesdefinition* aufgeprägt. Man kann nämlich festsetzen, die Punktefolge $\{p_n\}$ ($n = 1, 2, \dots$) konvergiere gegen den Punkt p , wenn für jeden Punkt $q \neq p$ fast alle Punkte p_n von p weniger entfernt sind, als q und p . Betrachtet man einen festen Punkt p und einen von p verschiedenen Punkt q , so kann man die Menge aller Punkte, welche von p weniger entfernt sind wie q , eine *Kugel* mit dem Zentrum p nennen und kann die Menge aller Punkte, die von p gleich weit entfernt sind wie q , als die *Begrenzung* der erwähnten Kugel bezeichnen.

Die beschriebene Vergleichbarkeit der Punktepaare entspricht einer *Abbildung der Menge aller Punktepaare auf eine geordnete Menge* von irgend welchen Symbolen. Besonders weittragend ist eine Abbildung der Menge aller Punktepaare auf eine geordnete Menge von Symbolen, für welche eine *Subtraktion* definiert ist, also speziell auf eine Menge reeller Zahlen mit Null als kleinstem Element, m. a. W. *die Definition einer nicht-negativen reellen Funktion auf der Menge aller Punktepaare*, durch welche je zwei Punkten a und b eine Zahl zugeordnet wird, die wir den *Abstand* von a und b nennen und mit $ab = ba$ bezeichnen, wobei identische Punkte und nur solche den Abstand Null haben. Eine Menge mit einer derartigen Abstandsfunktion nennen wir einen *halbmetrischen Raum*.

Die Abbildung der Menge aller Punktepaare auf eine geordnete Menge von Symbolen, für welche eine Subtraktion definiert ist, hat darum eine besondere Wichtigkeit, weil sie gestattet, an die Metrik eine wichtige naturgemäße und anschauliche Forderung zu stellen, nämlich, *daß sie gewissen Stetigkeitsbedingungen genüge*. Wir heben hier bloß die folgenden Definitionen hervor: Die Metrik heißt *stetig im Punktepaar a, b* , falls die Abstände von jedem mit a, b hinlänglich benachbarten Punktepaar sich vom Abstand ab beliebig wenig unterscheiden; streng gesprochen, indem wir als Abstand der geordneten Punktepaare a, b und c, d die Zahl $(ac + bd)$ festsetzen: Die Metrik heißt im Punktepaar a, b stetig, wenn zu jedem $\varepsilon > 0$ ein $\delta > 0$ existiert, so daß für jedes Punktepaar aus $(ac + bd) < \delta$ die Beziehung folgt: $|ab - cd| < \varepsilon$. In entsprechender Weise kann man

z. B. *Halbstetigkeit* der Metrik (in gewissen Punktepaaren und schlecht-hin) definieren. *Man kann die Erfüllung von Stetigkeitseigenschaften zur Klassifikation der Metriken verwenden.*

Ein Beispiel möge dies erläutern. Auf der Menge aller Paare von identischen Punkten ist jede Metrik in trivialer Weise unterhalb stetig. Wie Herr Vietoris bemerkt hat, ist die Forderung, daß die *Metrik eines vorgelegten halbmetrischen Raumes auf der Menge aller Paare von identischen Punkten* im Sinne der eben ausgeführten Gedanken *gleichmäßig stetig* (also auch oberhalb stetig) sei, äquivalent mit der Forderung, daß im gegebenen halbmetrischen Raum eine *Voisinage* im Sinne von Fréchet definiert sei, d. h. daß eine Funktion $f(\varrho)$ existiere mit $\lim_{\varrho=0} f(\varrho) = 0$, so daß für je

drei Punkte a, b, c aus $ac < \varrho$ und $bc < \varrho$ die Beziehung $ab < f(\varrho)$ folgt. In der Tat, die gleichmäßige Stetigkeit der Metrik auf der Menge der Paare von identischen Punkten besagt: Zu jedem $\varepsilon > 0$ existiert eine Zahl $\varphi(\varepsilon) > 0$, so daß für je zwei Punkte a und b , welche von einem Punkte c (also von einem Punktepaar c, c , welches aus zwei identischen Punkten besteht) einen Abstand $< \varphi(\varepsilon)$ haben (d. h. für je zwei Punkte a, b , für welche $(ac + bc) < \varphi(\varepsilon)$ gilt), die Beziehung besteht $ab < \varepsilon$. Offenbar erfüllt $\varphi(\varepsilon)$ die Bedingung $\lim_{\varepsilon=0} \varphi(\varepsilon) = 0$ und kann als monoton abnehmende

und sogar als beständig abnehmende Funktion von ε angenommen werden. Es existiert dann eine zu $\varphi(\varepsilon)$ inverse Funktion f , d. h. eine Funktion, so daß $f[\varphi(\varepsilon)] = \varepsilon$ gilt. Dieselbe erfüllt offenbar die Bedingung $\lim_{\varepsilon=0} f(\varphi(\varepsilon)) = 0$ und nehmen wir irgendeine

Zahl $\varrho = \frac{\varphi(\varepsilon)}{2}$, so folgt für je drei Punkte a, b, c aus $ac < \varrho$ und $bc < \varrho$ die Gültigkeit der Beziehung $(ac + bc) = 2\varrho = \varphi(\varepsilon)$ und mithin die Beziehung $ab < f[\varphi(\varepsilon)] = \varepsilon$. In analoger Weise zeigt man, daß die Metrik eines halbmetrischen Raumes, in dem eine Voisinage definiert ist, auf der Menge der Paare von identischen Punkten gleichmäßig stetig ist.

Es möge nun die Metrik eines vorgelegten halbmetrischen Raumes der schärfsten Stetigkeitsforderung genügen, nämlich *gleichmäßig stetig* sein. Die Forderung besagt, daß es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt, so daß für alle Punktequadrupel a_1, a_2, b_1, b_2 , welche der Bedingung $a_1 b_1 + a_2 b_2 < \delta$ genügen, stets auch die Beziehung $|a_1 a_2 - b_1 b_2| < \varepsilon$ gilt. Mit anderen Worten: Bezeichnet $f(\delta)$ die obere Schranke der Werte der Funktion $|a_1 a_2 - b_1 b_2|$ auf der Menge aller Punktequadrupel, die der Bedingung $a_1 b_1 + a_2 b_2 = \delta$ genügen, dann gilt $\lim_{\delta=0} f(\delta) = 0$. Ist nun

die mit x monoton wachsende (gleichsam *stetigkeitserzeugende*) Funktion $f(x)$ die in gewissem Sinn einfachste Funktion, welche der vorgeschriebenen Bedingung $\lim_{x=0} f(x) = 0$ genügt, nämlich $f(x) = x$, so gilt für je vier Punkte a_1, a_2, b_1, b_2 die Beziehung:

$$(\dagger) \quad |a_1 a_2 - b_1 b_2| \leq a_1 b_1 + a_2 b_2.$$

Dies ist aber gleichbedeutend damit, daß für je drei Punkte a, b, c die sogenannte *Dreiecksungleichung*

$$ab \leq ac + bc$$

gilt. Denn setzt man *erstens* in (†) $a = a_1, b = a_2, c = b_1 = b_2$, so erhält man die Dreiecksungleichung. Setzen wir *zweitens* die Gültigkeit der Dreiecksungleichung für je drei Punkte voraus und seien vier Punkte a_1, a_2, b_1, b_2 gegeben. Zuzufolge der Dreiecksungleichung gelten die Beziehungen

$$a_1 a_2 \leq a_1 b_1 + a_2 b_1, \quad a_2 b_1 \leq a_2 b_2 + b_1 b_2,$$

also ist

$$a_1 a_2 \leq a_1 b_1 + a_2 b_2 + b_1 b_2$$

und mithin gilt die Beziehung (†).

Die Dreiecksungleichung ergibt sich also als die in gewissem Sinn einfachste Realisierung der Forderung nach gleichmäßiger Stetigkeit der Metrik.

2. Bogenlänge und Dreiecksungleichung.

Es sei in einem halbmetrischen Raum ein einfacher Bogen B (d. h. ein topologisches Bild einer Strecke) gegeben. Es sei M eine endliche, etwa $n + 1$ Punkte enthaltende Teilmenge von B , welche die beiden Endpunkte a und b von B enthält. Wir durchlaufen B in einer der beiden möglichen Richtungen, etwa von a nach b , und bezeichnen die $n + 1$ Punkte der Menge M in der Reihenfolge, in der sie bei dieser Durchlaufung des Bogens angetroffen werden, mit $a_0 = a, a_1, \dots, a_{n-1}, a_n = b$.

Wir setzen $\lambda_B(M) = \sum_{k=1}^n a_{k-1} a_k$ und bezeichnen mit $L(B)$ bzw. mit $l(B)$ die obere bzw. die untere Schranke aller Zahlen $\lambda_B(M)$, wenn M alle endlichen Teilmengen von B durchläuft. Wir nennen $L(B)$ bzw. $l(B)$ auch *die obere bzw. die untere Länge des Bogens B* . Offenbar ist für jeden a und b verbindenden Bogen B , wenn ab den Abstand der Punkte a und b bezeichnet, $L(B) \geq ab$. Ferner gilt für jeden Bogen B' , welcher Teilmenge des Bogens B ist, $L(B') \leq L(B)$ und $l(B') \leq l(B)$.

Herr L. Vietoris hat mich darauf hingewiesen, daß die Menge aller reellen Zahlen zwischen 0 und 1 ($0 \leq x \leq 1$), wenn man als Abstand der Punkte x und y die Zahl $(x - y)^2$ festsetzt, einen Bogen darstellt, welcher im angeführten Sinn die untere Länge Null hat. In Verfolgung des Gedankens von Herrn Vietoris sieht man: *Zu jeder reellen Zahl $\gamma < 1$ gibt es einen halbmetrischen Raum mit gleichmäßig stetiger Metrik, in welchem für je drei Punkte a, b, c die Beziehung*

$$(†) \quad ab + bc \geq \gamma \cdot ac$$

erfüllt ist und welcher Bögen von der unteren Länge Null enthält. Sei eine Zahl $\gamma < 1$ vorgegeben. Um die Richtigkeit der Behauptung einzusehen, betrachte man die Menge aller reellen Zahlen x , die der Ungleichung $0 \leq x \leq 1$ genügen und setzen als Abstand $r(x, y)$ der Punkte x und y die Zahl $r(x, y) = |x - y|^\alpha$ fest, wo α eine reelle Zahl bedeutet,

deren nähere Bestimmung wir uns noch vorbehalten. Die Menge aller reellen Zahlen zwischen 0 und 1 wird durch die Festsetzung zu einem halbmetrischen Raum, dessen Metrik offenbar gleichmäßig stetig ist. Es seien nun x, y, z drei Punkte unseres Raumes (reelle Zahlen zwischen 0 und 1). Eine elementare Rechnung zeigt, daß der Ausdruck

$$r(x, z) - r(x, y) - r(y, z) = |x - z|^{\alpha} - |x - y|^{\alpha} - |y - z|^{\alpha},$$

als Funktion von y aufgefaßt (d. h. wenn x und z festgehalten werden und y zwischen 0 und 1 variiert), sein Maximum annimmt für den Fall $|x - y| = |y - z| = \frac{1}{2}|x - z|$. Halten wir $r(x, z)$ fest, so ist also, wenn y zwischen 0 und 1 variiert, $r(x, y) + r(y, z) \geq 2 \cdot \frac{|x - z|^{\alpha}}{2^{\alpha}} = 2^{1-\alpha} r(x, z)$.

Setzen wir nun $\alpha = 1 - \frac{\log \gamma}{\log 2}$, wo γ die gegebene Zahl < 1 ist, so ist $2^{1-\alpha} = \gamma$ und der halbmetrische Raum erfüllt also für je drei seiner Punkte die Beziehung (\dagger). Es bleibt noch zu zeigen, daß der Raum Bögen mit der unteren Länge Null enthält. Es ist der Raum selbst und daher jeder seiner Teilbögen ein Bogen von der Länge Null. Betrachten wir, um dies zu beweisen, für eine natürliche Zahl n die endliche Menge M_n , welche aus den $1 + 2^n$ Punkten $x_k = \frac{k}{2^n}$ ($k = 0, 1, \dots, 2^n$) besteht. Es ist $r(x_i, x_{i+1}) = \left(\frac{1}{2^n}\right)^{\alpha}$ und mithin, wenn wir den Raum als Bogen B bezeichnen, $\lambda_B(M_n) = 2^n \left(\frac{1}{2^n}\right)^{\alpha} = 2^{n(1-\alpha)} = \gamma^n$. Da nun $\gamma < 1$ ist, gilt $\lim_{n \rightarrow \infty} \lambda_B(M_n) = \lim_{n \rightarrow \infty} \gamma^n = 0$ und mithin ist die untere Schranke aller Zahlen $\lambda_B(M)$, wenn M alle endlichen Teilmengen von B durchläuft, $= 0$. D. h. aber, die untere Länge des Bogens B und mithin die untere Länge von allen Teilbögen des Raumes ist Null.

Wir sehen daher, daß die Dreiecksungleichung unter allen linearen Dreipunkterelationen eines halbmetrischen Raumes die schwächste ist, welche, für alle halbmetrischen Räume gemeinsam, die Abwesenheit von Bögen mit der unteren Länge Null gewährleistet. In der Tat, in einem metrischen Raum [d. h. in einem halbmetrischen Raum, welcher der Dreiecksungleichung genügt, d. i. der Ungleichung (\dagger) für $\gamma = 1$] ist offenbar die untere Länge jedes Bogens mindestens gleich dem Abstand der Endpunkte des Bogens, also für jeden Bogen von Null verschieden. Wird jedoch in der postulierten linearen Dreipunkterelation (\dagger) der Koeffizient γ irgendwie verkleinert, so gibt es halbmetrische Räume mit gleichmäßig stetiger Metrik, welche der abgeschwächten Forderung genügen, und dabei Bögen von der unteren Länge Null enthalten, ja sogar derartige Räume, deren sämtliche Teilbögen die untere Länge Null besitzen.

3. Vollständigkeit der Dreiecksungleichung.

Wir haben im vorangehenden die Dreiecksungleichung als die in gewissem Sinn einfachste, nämlich lineare, Realisierung der Forderung nach gleichmäßiger Stetigkeit und zugleich als die schwächste lineare Dreipunkte-Relation kennengelernt, welche, für alle halbmetrischen Räume gemeinsam, die Abwesenheit von Bögen mit der unteren Länge Null gewährleistet. Zum Abschluß sei bemerkt, daß die Dreiecksungleichung in gewissem Sinn auch eine vollständige, nicht verschärfbare Forderung an die Abstandsmetrik darstellt: Zu je drei gegebenen nicht negativen Zahlen, welche in jeder Anordnung der Dreiecksungleichung genügen, existieren ja in jedem mehr als eindimensionalen euklidischen Raum drei Punkte, deren paarweise Abstände gleich den gegebenen Zahlen sind. Will man also die wichtigsten Räume, nämlich die euklidischen, unter die metrischen Räume einordnen, so darf in die Definition des metrischen Raumes keine zur Dreiecksungleichung hinzutretende Relation zwischen den paarweisen Abständen je dreier Punkte hinzugefügt werden.

II. Die n -dimensionale Metrik.

Die Definition des Begriffes „metrischer Raum“ ist eine Präzisierung der (für die Theorie der reellen Funktionen und für die Topologie wichtigsten) Eigenschaften der *Abstandsmetrik euklidischer Räume* unter Abstraktion von der speziellen Natur der Punkte euklidischer Räume, nämlich unter Abstraktion von der Koordinatendarstellung der euklidischen Punkte. Diese Abstandsmetrik handelt von den *Punktepaaren* der Räume. Zur Begründung einer Theorie der *n -dimensionalen Metrik* in allgemeinen Mengen kann man ausgehen von der Betrachtung der $(n+1)$ -*Tupel* von Punkten der Mengen. Auch hier ist es naturgemäß, unter Abstraktion von der Koordinatendarstellung der Punkte, an die n -dimensionale Metrik *euklidischer Räume* anzuknüpfen.

1. Die euklidische n -dimensionale Metrik.

Sind

$$p^1 = (x_1^1, x_2^1, \dots, x_k^1), \quad p^2 = (x_1^2, x_2^2, \dots, x_k^2), \dots, \quad p^{n+1} = (x_1^{n+1}, x_2^{n+1}, \dots, x_k^{n+1})$$

$n+1$ Punkte des R_k , des k -dimensionalen euklidischen Raumes, so ist ihnen eine nicht-negative Zahl, das *Volumen des durch die $n+1$ Punkte bestimmten n -dimensionalen Simplexes*, zugeordnet, eine Zahl, welche wir kurz mit $p^1 p^2 \dots p^{n+1}$ bezeichnen und welche gegeben ist durch den Ausdruck

$$p^1 p^2 \dots p^{n+1} = + \sqrt{\sum_{i_1, i_2, \dots, i_n=1}^k \begin{vmatrix} x_{i_1}^1 & x_{i_2}^1 & \dots & x_{i_n}^1 & 1 \\ x_{i_1}^2 & x_{i_2}^2 & \dots & x_{i_n}^2 & 1 \\ \dots & \dots & \dots & \dots & \dots \\ x_{i_1}^{n+1} & x_{i_2}^{n+1} & \dots & x_{i_n}^{n+1} & 1 \end{vmatrix}^2},$$

wobei die Summation über $\binom{k}{n}$ Summanden zu erstrecken ist.

Diese für die Menge aller $(n+1)$ -Tupel von Punkten definierte n -dimensionale Volumsfunktion besitzt insbesondere die folgenden Eigenschaften:

S_n (Symmetriebedingung): Für je $n+1$ Punkte p_1, p_2, \dots, p_{n+1} gilt, wenn i_1, i_2, \dots, i_{n+1} irgendeine Permutation der Zahlen $1, 2, \dots, n+1$ bezeichnet, $p_{i_1} p_{i_2} \dots p_{i_{n+1}} = p_1 p_2 \dots p_{n+1}$.

O_n^e (Euklidische Nullbedingung): Für je $n+1$ Punkte p_1, p_2, \dots, p_{n+1} , welche in einer $(n-1)$ -dimensionalen Ebene liegen, gilt $p_1 p_2 \dots p_{n+1} = 0$.

P_n^e (Euklidische Positivitätsbedingung): Für je $n+1$ Punkte p_1, p_2, \dots, p_{n+1} , welche nicht in einer $(n-1)$ -dimensionalen Ebene liegen, gilt $p_1 p_2 \dots p_{n+1} > 0$.

U_n (Simplexungleichung): Von den $n+2$ n -dimensionalen Simplexen, welche durch irgend $n+2$ Punkte gegeben sind, hat keines ein größeres Volumen als die Summe der $n+1$ übrigen; mit anderen Worten: für je $n+2$ Punkte p_1, p_2, \dots, p_{n+2} des Raumes ist $p_1 p_2 \dots p_{n+1} \leq \sum_{i=1}^{n+1} p_1 \dots p_{i-1} p_{i+1} \dots p_{n+2}$.

U_1 ist die Dreiecksungleichung. U_2 besagt, daß in keinem Tetraeder eine Seitenfläche einen größeren Inhalt hat, als die drei übrigen Flächen zusammen („Tetraederungleichung“). Bezeichnet man ferner als nulldimensionale Ebene eine genau einen Punkt enthaltende Menge, so liefern O_1 und P_1 die in die Definition der allgemeinen Abstandsmetrik aufgenommene Forderung, daß der Abstand zweier Punkte dann und nur dann Null ist, wenn die beiden Punkte identisch sind.

Die Gültigkeit der Simplexungleichung in euklidischen Räumen ist geometrisch evident. Sie kann rechnerisch aus der Definition des n -dimensionalen Volumens hergeleitet werden, so wie die Gültigkeit der Dreiecksungleichung (in üblicher Weise unter Benützung der Schwarzschen Ungleichung) für die euklidische Abstandsdefinition hergeleitet werden kann. Wir gehen hier auf diese im allgemeinen Fall etwas langwierige Rechnung (bei welcher man auf höherdimensionale Analoga der Schwarzschen Ungleichung stößt) nicht ein, da sie mit den geometrischen Problemen der allgemeinen Metrik nichts zu tun hat.

Wir übergehen ferner die Frage, inwieweit die genannten Forderungen voneinander abhängig sind, bzw. sich durch formale Abänderungen unabhängig machen lassen. Es sei diesbezüglich auf eine Bemerkung von Lindenbaum (Fundamenta Math. 8, S. 211) hingewiesen, daß hinsichtlich der gewöhnlichen Abstandsmetrik durch geeignete Formu-

lierung der Dreiecksungleichung U_1 (wofern dieselbe nämlich in der Form $xy + xz \geq yz$ angenommen wird) das Symmetriepostulat S_1 beweisbar wird, also nicht eigens vorausgesetzt werden muß, und daß das Positivitätspostulat P_1 dahin abgeschwächt werden kann, daß je zwei verschiedene Punkte einen von Null verschiedenen Abstand haben (während seine Positivität sodann bewiesen werden kann).

2. Stetigkeit der Metrik und Simplexungleichung.

Betrachten wir nun eine abstrakte Menge, die wir als Raum und deren Elemente wir als Punkte bezeichnen. Eine primitive n -dimensionale Metrik besteht in der *Vergleichbarkeit der $(n+1)$ -Tupel von Punkten* in dem Sinn, daß man sagen kann, die Punkte a_1, a_2, \dots, a_{n+1} bestimmen ein Simplex mit größerem, kleinerem oder gleichem Volumen wie die Punkte b_1, b_2, \dots, b_{n+1} . Besonders weittragend ist im allgemeinen Fall, so wie im Fall $n=1$, eine Vergleichung der $(n+1)$ -Tupel von Punkten auf Grund der Definition einer *nicht-negativen* reellen *Funktion* auf der Menge aller $(n+1)$ -Tupel.

Über das Verschwinden der n -dimensionalen Volumsfunktion postulieren wir dabei einstweilen nur die folgende

O_n^s (Schwache Nullbedingung): *Für jedes „unechte“ $(n+1)$ -Tupel von Punkten, d. h. für je $n+1$ Punkte p_1, p_2, \dots, p_{n+1} , unter denen identische Punkte vorkommen, gilt $p_1 p_2 \dots p_{n+1} = 0$.*

In euklidischen Räumen ergibt sich diese schwache Nullbedingung (welche übrigens für $n=1$ mit O_1 identisch ist) aus der Nullbedingung O_n auf Grund der Tatsache, daß je m Punkte, wenn $m < n+1$ ist, in einer $(n-1)$ -dimensionalen Ebene liegen.

Ist in einem Raum eine solche *n -dimensionale Volumsfunktion* definiert, dann *prägt sie der Menge aller geordneten $(n+1)$ -Tupel von Punkten eine Abstands- und eine Limesdefinition auf*, ganz unabhängig davon, ob im Raume auch eine Abstandsfunktion, d. h. eine Funktion der Punktepaare, definiert ist oder nicht. Wir geben die Definition bloß für den Fall $n=2$, also für die Flächenmetrik, explizit an, da hieraus bereits klar wird, wie die analogen (bloß wesentlich längeren) Formeln für höhere Dimensionen zu lauten haben.

Es seien in einem Raum, in dem eine Flächenmetrik vorliegt (d. h. eine nicht-negative Funktion der Punktetripel definiert ist, welche den Bedingungen S_2, U_2, O_2^s genügt), zwei geordnete Tripel von Punkten a_1, a_2, a_3 , und b_1, b_2, b_3 gegeben. Als ihren Abstand im Sinne der Flächenmetrik bezeichnen wir folgenden in den beiden Tripeln symmetrischen Ausdruck, der als Summe von Dreiecksinhalten ausschließlich auf Grund der Flächenfunktion berechenbar ist und den wir $r \begin{pmatrix} a_1, a_2, a_3 \\ b_1, b_2, b_3 \end{pmatrix}$ nennen:

$$(\dagger) \quad r \begin{pmatrix} a_1, a_2, a_3 \\ b_1, b_2, b_3 \end{pmatrix} = \frac{1}{2} \left(\begin{array}{l} a_1 b_1 b_2 + a_2 b_1 b_2 + a_1 a_2 b_1 + a_1 a_2 b_2 + \\ + a_1 b_1 b_3 + a_3 b_1 b_3 + a_1 a_3 b_1 + a_1 a_3 b_3 + \\ + a_2 b_2 b_3 + a_3 b_2 b_3 + a_2 a_3 b_2 + a_2 a_3 b_3 \end{array} \right).$$

Fordern wir, daß die Flächenfunktion hinsichtlich dieser aus ihr abgeleiteten Abstandsfunktion eine *gleichmäßig stetige Funktion* der Punkttripel sei, so bedeutet dies: Zu jedem $\varepsilon > 0$ existiert ein $\delta > 0$, so daß für je zwei geordnete Punkttripel a_1, a_2, a_3 und b_1, b_2, b_3 , welche einen Abstand $r \begin{pmatrix} a_1, a_2, a_3 \\ b_1, b_2, b_3 \end{pmatrix} < \delta$ besitzen, die Beziehung $|a_1 a_2 a_3 - b_1 b_2 b_3| < \varepsilon$ gilt. Mit anderen Worten: Bezeichnet $f(\delta)$ die obere Schranke der Werte der Funktion $|a_1 a_2 a_3 - b_1 b_2 b_3|$ auf der Menge aller Punkttripel, die der Beziehung $r \begin{pmatrix} a_1, a_2, a_3 \\ b_1, b_2, b_3 \end{pmatrix} = \delta$ genügen, dann gilt $\lim_{\delta=0} f(\delta) = 0$. Ist nun diese mit x monoton wachsende stetigkeitserzeugende Funktion $f(x)$ die in gewissem Sinn einfachste Funktion, welche der vorgeschriebenen Bedingung $\lim_{x=0} f(x) = 0$ genügt, nämlich $f(x) = x$, so gilt für je sechs Punkte $a_1, a_2, a_3, b_1, b_2, b_3$ die Beziehung

$$(\dagger\dagger) \quad |a_1 a_2 a_3 - b_1 b_2 b_3| \leq r \begin{pmatrix} a_1, a_2, a_3 \\ b_1, b_2, b_3 \end{pmatrix},$$

wo für $r \begin{pmatrix} a_1, a_2, a_3 \\ b_1, b_2, b_3 \end{pmatrix}$ der Ausdruck (\dagger) zu substituieren ist. Dies ist aber gleichbedeutend damit, daß für je vier Punkte a, b, c, d die Tetraederungleichung

$$(U_2) \quad abc \leq abd + acd + bcd$$

gilt. Setzen wir nämlich *erstens* in $(\dagger\dagger)$ und (\dagger)

$$a = a_1, \quad b = a_2, \quad c = a_3, \quad d = b_1 = b_2 = b_3,$$

dann erhält man unter Berücksichtigung der vorausgesetzten schwachen Nullbedingung O_2^s die Tetraederungleichung U_2 . — Setzen wir *zweitens* die Gültigkeit von U_2 für je vier Punkte voraus und seien sechs Punkte $a_1, a_2, a_3, b_1, b_2, b_3$ gegeben. Zufolge der Tetraederungleichung gelten die Beziehungen:

$$a_2 a_3 b_1 \leq a_1 a_2 a_3 + a_1 a_2 b_1 + a_1 a_3 b_1,$$

$$a_3 b_1 b_2 \leq a_2 a_3 b_1 + a_2 a_3 b_2 + a_2 b_1 b_2,$$

$$b_1 b_2 b_3 \leq a_3 b_1 b_2 + a_3 b_1 b_3 + a_3 b_2 b_3,$$

also, wenn man diese Ungleichungen addiert:

$$(\dagger) \quad |b_1 b_2 b_3 - a_1 a_2 a_3| \leq a_1 a_2 b_1 + a_1 a_3 b_1 + a_2 a_3 b_2 + a_2 b_1 b_2 + a_3 b_1 b_3 + a_3 b_2 b_3;$$

ferner ist auf Grund der Tetraederungleichung:

$$a_1 a_3 b_2 \leq a_1 a_2 a_3 + a_1 a_2 b_2 + a_2 a_3 b_2,$$

$$a_1 b_2 b_3 \leq a_1 a_3 b_2 + a_1 a_3 b_3 + a_3 b_2 b_3,$$

$$b_1 b_2 b_3 \leq a_1 b_2 b_3 + a_1 b_1 b_3 + a_1 b_1 b_2,$$

also, wenn man diese Ungleichungen addiert:

$$(\sim \sim) \quad | b_1 b_2 b_3 - a_1 a_2 a_3 | \leq a_1 a_2 b_2 + a_2 a_3 b_2 + a_1 a_3 b_3 + a_3 b_2 b_3 + a_1 b_1 b_3 + a_1 b_1 b_2.$$

Addiert man die Ungleichungen (\dagger) und $(\dagger\dagger)$ und dividiert durch 2, so erhält man die Beziehung $(\dagger\dagger)$, wenn man den Ausdruck (\dagger) daselbst substituiert. Damit ist die Äquivalenz von $(\dagger\dagger)$ und U_2 erwiesen. — Diese Betrachtung, deren Analogon hinsichtlich der Abstandsmetrik oben (S. 144 f.) durchgeführt worden ist, läßt sich offenbar für beliebige Dimensionen n anstellen und ergibt den Satz:

Die Simplexungleichung U_n ist die in gewissem Sinn einfachste, nämlich lineare Realisierung der Forderung nach gleichmäßiger Stetigkeit der n -dimensionalen Volumsfunktion auf der Menge der Punkte- $(n+1)$ -Tupel, wofern für die $(n+1)$ -Tupel lediglich auf Grund der n -dimensionalen Volumsfunktion ein geeigneter Abstand definiert wird.

3. Allgemeine n -dimensionale Metrik.

Zur näheren Festlegung der Gesetze der n -dimensionalen Metrik bieten sich nun zwei Wege dar: Der *erste* besteht darin, daß man in einer abstrakten Menge eine n -dimensionale Metrik *unabhängig von eventuell in der Menge definierten niedrigerdimensionalen Metriken* einführt als nicht-negative Funktion der $(n+1)$ -Tupel von Punkten, die den Gesetzen S_n, U_n genügt. Diese Gesetze, sowie die Stetigkeitseigenschaften der n -dimensionalen Metrik können, wie wir sahen, independent von niedrigerdimensionalen Metriken untersucht werden. Ferner kann die lediglich den Identitätsbegriff verwendende schwache Nullbedingung O_n^s postuliert werden. Weitere Anhaltspunkte für das Verschwinden der n -dimensionalen Volumsfunktion liefert folgende aus U_n sich ergebende Bemerkung: Wenn für $n+2$ Punkte $n+1$ unter den $n+2$ Simplexen, welche sich aus den Punkten bilden lassen, das Volumen Null haben, dann haben alle $n+2$ Simplexe das Volumen Null. Niedrigerdimensionale Metriken kann man durch Grenzübergang in gewissem Umfang einzuführen suchen.

Der *zweite* Weg besteht in der Einführung einer n -dimensionalen Metrik in Räume, für welche niedrigerdimensionale Metriken bereits definiert sind. In diesem Fall ist die n -dimensionale Metrik mit den vorliegenden niedrigerdimensionalen Metriken, insbesondere mit der Abstandsmetrik des Raumes zu verknüpfen.

4. Volumslimes und Abstandslimes.

Wir verfolgen zunächst den zweiten Weg zur Betrachtung der *Stetigkeitseigenschaften* der Metrik. Wenn in einem Raum eine Abstandsmetrik vorliegt, dann ist durch dieselbe auch eine Limesdefinition für die Menge aller geordneten $(n+1)$ -Tupel von Punkten des Raumes gegeben. Man kann nämlich sagen, die geordneten $(n+1)$ -Tupel $\{p_1^k, p_2^k, \dots, p_{n+1}^k\}$ ($k=1, 2, \dots$ ad inf.) *konvergieren im Sinne des Abstandslimes* gegen das $(n+1)$ -Tupel p_1, p_2, \dots, p_{n+1} , falls für die Abstände $\lim_{k=\infty} p_i^k p_i = 0$ ($i=1, 2, \dots, n+1$) gilt. Nun haben wir oben bereits auf Grund der n -dimensionalen Volumsfunktion einen Abstand und daher einen Limes für die Menge der geordneten $(n+1)$ -Tupel definiert. Z. B. sagen wir für $n=2$ von einer Folge von Punkttripeln $\{p_1^k, p_2^k, p_3^k\}$ ($k=1, 2, \dots$ ad inf.), sie konvergiere *im Sinne des Flächenlimes* gegen das Punkttripel p_1, p_2, p_3 , wenn $\lim_{k=\infty} r(p_1^k, p_2^k, p_3^k) = 0$ ist. Jede dieser beiden Limesdefinitionen ergibt eine Stetigkeitsdefinition, und das Verhältnis dieser beiden Definitionen wollen wir nun untersuchen. Dazu setzen wir für die n -dimensionale Metrik die Bedingungen S_n, U_n und O_n^s voraus und wollen überdies die n -dimensionale Metrik mit der Abstandsmetrik verknüpfen durch folgende Forderung: Die n -dimensionale Metrik soll im Sinne des Abstandslimes auf der Menge der unechten $(n+1)$ -Tupel, also auf der Menge jener $(n+1)$ -Tupel, welche auf Grund der schwachen Nullbedingung O_n^s das Volumen Null bestimmen, stetig sein; mit anderen Worten wir machen die Forderung:

C_n^s (Schwache Stetigkeitsforderung): *Ist p_1, p_2, \dots, p_{n+1} ein unechtes $(n+1)$ -Tupel von Punkten und $p_1^k, p_2^k, \dots, p_{n+1}^k$ ($k=1, 2, \dots$ ad inf.) eine Folge von $(n+1)$ -Tupeln von Punkten, welche gegen p_1, p_2, \dots, p_{n+1} im Sinne des Abstandslimes konvergiert, d. h. so daß $\lim_{k=\infty} p_i^k p_i = 0$ ($i=1, 2, \dots, n+1$) gilt, dann ist $\lim_{k=\infty} p_1^k p_2^k \dots p_{n+1}^k = 0$.*

Wir beweisen sodann: *Jede Folge $\{p_1^k, p_2^k, \dots, p_{n+1}^k\}$ ($k=1, 2, \dots$ ad inf.), welche gegen ein $(n+1)$ -Tupel p_1, p_2, \dots, p_{n+1} im Sinne des Abstandslimes konvergiert, konvergiert gegen dieses $(n+1)$ -Tupel auch im Sinne des Volumslimes.* Jede auf der Menge der $(n+1)$ -Tupel im Sinne des Volumslimes stetige Funktion ist daher auf der Menge der $(n+1)$ -Tupel auch im Sinne des Abstandslimes stetig. *Insbesondere ist also die n -dimensionale Volumsfunktion selbst, von welcher durch C_n^s die Stetigkeit im Sinne des Abstandslimes bloß auf der Menge aller unechten $(n+1)$ -Tupel gefordert wird, da sie im Sinne des Volumslimes, wie oben*

gezeigt wurde, stetig ist, auch *im Sinne des Abstandslimes auf der Menge aller $(n + 1)$ -Tupel stetig.*

Wir führen den Beweis der Einfachheit halber für $n = 2$. Er verläuft für höhere Dimensionen völlig analog. Wir haben zu beweisen, daß stets aus

$$(*) \quad \lim_{k=\infty} p_i^k p_i = 0 \quad (i = 1, 2, 3),$$

wenn wir $r_k = r \left(\begin{matrix} p_1^k, p_2^k, p_3^k \\ p_1, p_2, p_3 \end{matrix} \right)$ setzen, $\lim_{k=\infty} r_k = 0$ folgt. Gehen wir auf die

Definition (\dagger) von $r \left(\begin{matrix} a_1, a_2, a_3 \\ b_1, b_2, b_3 \end{matrix} \right)$ zurück, so sehen wir, daß r_k Summe von 12 Dreiecksinhalten ist, deren jeder (abgesehen von Permutationen der Ecken) die Gestalt hat: $p_i p_j p_i^k$ oder $p_i p_i^k p_j^k$. Auf Grund der schwachen Stetigkeitsforderung C_n^s und der Voraussetzung (*) konvergiert also mit wachsendem k jeder der 12 Summanden und mithin r_k gegen Null, w. z. b. w.

Die Umkehrung des bewiesenen Satzes, die Behauptung nämlich, daß jede im Sinne des Volumslimes konvergente Folge von $(n + 1)$ -Tupeln auch im Sinne des Abstandslimes konvergiert, ist ohne eine gewisse Einschränkung schon in euklidischen Räumen unrichtig. Allerdings läßt sich die Einschränkung für euklidische Räume leicht formulieren: Konvergiert eine Folge von $(n + 1)$ -Tupeln im Sinne des Volumslimes gegen ein $(n + 1)$ -Tupel *mit nicht-verschwindendem Volumen*, dann konvergiert sie auch im Sinne des Abstandslimes und zwar gegen dasselbe Grenz- $(n + 1)$ -Tupel, während dies, wenn es sich um Grenz- $(n + 1)$ -Tupel mit verschwindendem Volumen handelt, offenbar nicht der Fall zu sein braucht. In allgemeinen metrischen Räumen gilt die Umkehrung der oben bewiesenen Behauptung auch nicht unter Beschränkung auf Grenz- $(n + 1)$ -Tupel mit positivem Volumen. Betrachten wir beispielsweise den in der ersten Untersuchung über allgemeine Metrik (S. 111) erwähnten Raum, bestehend aus den Punkten der Ebene mit den Koordinaten $x = 0$ und mit den Koordinaten $x \geq 0$ und $y \geq 0$ (also die Menge bestehend aus der y -Achse und einem Viertel der Ebene). Als Abstand definieren wir (wie a. a. O.) für je zwei Punkte die Länge des kürzesten sie verbindenden Bogens der Ebene (dieser kürzeste Bogen ist eine Strecke, wenn es sich um zwei Punkte des Ebenenviertels oder um zwei Punkte der y -Achse handelt, und ist Summe zweier Strecken, wenn es sich um einen Punkt des Ebenenviertels und einen Punkt mit negativer Ordinate handelt). Für je drei Punkte a, b, c definieren wir als Flächeninhalt: Falls a, b, c im Ebenenviertel liegen, den euklidischen Inhalt des Dreiecks; falls a und b im Ebenenviertel liegen und c eine negative Ordinate hat, den euklidischen Inhalt des durch a, b und den

Punkt $(0, 0)$ bestimmten Dreiecks; falls a und b eine negative Ordinate haben, Null. Setzen wir dann

$$a_n = (0, 1) = a, \quad b_n = (1, 0) = b, \quad c_n = \left(0, \frac{1-n}{n}\right) \quad (n = 1, 2, \dots \text{ ad inf.}),$$

dann konvergiert offenbar, wenn wir den Punkt $(0, 0)$ mit c bezeichnen, die Folge von Punktetripeln $\{a_n, b_n, c_n\}$ gegen a, b, c wohl im Sinne des Volumslimes, nicht aber im Sinne des Abstandslimes.

5. Das Verschwinden der Volumsfunktion.

Das Verschwinden der euklidischen n -dimensionalen Volumsfunktion wird durch die Bedingungen O_n und P_n vollständig beschrieben. Diese Bedingungen stützen sich wesentlich auf die Definition von $(n-1)$ -dimensionalen Ebenen im R_k . In allgemeinen metrischen Räumen liegt etwas Entsprechendes zunächst nicht vor. Hier steht, wenn wir die n -dimensionale Metrik unabhängig von eventuell vorliegenden niedrigerdimensionalen Metriken einführen, wie wir erwähnt haben, zunächst nur die schwache Nullbedingung zur Verfügung, welche eine hinreichende, aber im allgemeinen natürlich keineswegs notwendige Bedingung für das Verschwinden der Volumsfunktion liefert, — wozu noch die erwähnten Konsequenzen der Simplexungleichung U_n für das Verschwinden der Volumsfunktion, oder, genauer gesagt, für die Gesamtheit aller $(n+1)$ -Tupel mit verschwindendem Volumen hinzutreten.

Wird die n -dimensionale Metrik in Verbindung mit einer im Raum vorliegenden $(n-1)$ -dimensionalen Metrik eingeführt, so ergibt sich noch eine wichtige, über die schwache Nullbedingung wesentlich hinausgehende Bedingung für das Verschwinden der n -dimensionalen Volumsfunktion. Wir können dann nämlich fordern, daß $n+1$ Punkte, für welche in einer der möglichen Anordnungen in der $(n-1)$ -dimensionalen Simplexungleichung das Gleichheitszeichen gilt, ein n -dimensionales Simplex mit dem Volumen Null bestimmen. Es handelt sich dabei also um $(n+1)$ -Tupel von Punkten, für welche das $(n-1)$ -dimensionale Volumen von einem der $n+1$ aus ihnen bildbaren n -Tupel gleich ist der Summe der $(n-1)$ -dimensionalen Volumina der n übrigen n -Tupel. Wir können dies formulieren als

O_n^m (Mittlere Nullbedingung): Gilt für $n+1$ Punkte p_1, p_2, \dots, p_{n+1} ,
 $p_1 p_2 \dots p_n = \sum_{i=1}^n p_1 \dots p_{i-1} p_{i+1} \dots p_n p_{n+1}$, dann ist $p_1 p_2 \dots p_n p_{n+1} = 0$.

Diese mittlere Nullbedingung enthält offenbar die schwache Nullbedingung als Teilaussage. Angenommen nämlich, es liege ein unechtes $(n+1)$ -Tupel von Punkten vor, etwa bestehend aus den Punkten

$p_1, p_2, \dots, p_n, p_{n+1}$, unter welchen die Punkte p_n und p_{n+1} identisch sind. Dann ist auf Grund von O_{n-1}^s jede der Zahlen $p_1 \dots p_{i-1} p_{i+1} \dots p_n p_{n+1} = 0$ ($i = 1, 2, \dots, n - 1$), da jede dieser Zahlen das Volumen eines $(n - 1)$ -dimensionalen Simplexes mit zwei identischen Ecken, nämlich p_n und p_{n+1} , darstellt. Also reduziert sich der Ausdruck $\sum_{i=1}^n p_1 \dots p_{i-1} p_{i+1} \dots p_n p_{n+1}$ auf p_1, p_2, \dots, p_n , so daß die durch O_n^m geforderte Beziehung von selbst erfüllt ist.

Wir sehen ferner, daß in euklidischen Räumen die Bedingung O_n^m auf Grund der dort giltigen starken Nullbedingung O_n erfüllt ist. Denn wenn für $n + 1$ Punkte p_1, p_2, \dots, p_{n+1} die in O_n^m geforderte Bedingung erfüllt ist, so liegt p_{n+1} in der durch die Punkte p_1, \dots, p_n bestimmten Ebene und zwar innerhalb oder auf dem Rande des durch diese Punkte bestimmten $(n - 1)$ -dimensionalen Simplexes. Jedenfalls bestimmen dann also die Punkte p_1, p_2, \dots, p_{n+1} ein Simplex mit dem n -dimensionalen Volumen Null.

Betrachten wir insbesondere den Fall $n = 2$. Die Bedingung O_2^m besagt, daß für je drei Punkte, für welche bei einer der drei möglichen Anordnungen in der Dreiecksungleichung das Gleichheitszeichen gilt, ein Simplex mit dem zweidimensionalen Volumen Null bestimmen; d. h. aber in der Ausdrucksweise der ersten Untersuchung über allgemeine Metrik, daß je drei Punkte, von welchen einer *Zwischenpunkt* der beiden anderen ist, oder, wie wir dies in allgemeinen Räumen auch ausgedrückt haben (a. a. O., S. 106), daß je drei Punkte, welche auf einer Geraden liegen, ein Dreieck vom Flächeninhalt Null bestimmen.

Es legt dieser Sachverhalt die Einführung eines *n-dimensionalen Zwischenbegriffes* nahe, welcher durch die n -dimensionale Metrik erklärbar wird. Wir sagen beispielsweise, ein Punkt a eines Raumes mit zweidimensionaler Metrik liegt zwischen drei Punkten b, c, d , falls $abc + abd + acd = bcd$ gilt. In dieser Ausdrucksweise können wir die mittlere Nullbedingung O_n^m auch folgendermaßen aussprechen:

O_n^m (Mittlere Nullbedingung): *Je $n + 1$ Punkte, von welchen einer zwischen den n anderen liegt, bestimmen ein Simplex mit dem n -dimensionalen Volumen Null.*

Allerdings, eine zugleich *notwendige* Bedingung für das Verschwinden der n -dimensionalen Volumsfunktion wird durch die mittlere Nullbedingung für $n \geq 3$ selbst in euklidischen Räumen nicht geliefert. Betrachten wir nämlich vier Punkte, welche in einer Ebene liegen und daselbst die Ecken eines Quadrates bilden. Die vier Punkte bestimmen, da sie in einer Ebene liegen, ein dreidimensionales Simplex mit dem dreidimensionalen Inhalt Null, obwohl keiner der vier Punkte zwischen den drei anderen (d. h. weder

im Innern noch auf dem Rande des durch die drei anderen gebildeten Dreieckes) liegt.

Sehen wir aber näher zu, so zeigt sich, daß die Festlegung einer *notwendigen* Bedingung für das Verschwinden der n -dimensionalen Volumsfunktion mit Hilfe der $(n - 1)$ -dimensionalen Volumsfunktion für $n \geq 3$ selbst in euklidischen Räumen *unmöglich* ist. Betrachten wir nämlich in der Ebene vier Punkte a, b, c, d mit den Abständen $ab = ac = bc = ad = cd = 1$, $bd = \sqrt{3}$. Dann bestimmen diese vier Punkte, da sie in einer Ebene liegen, ein dreidimensionales Volumen Null. Je drei dieser Punkte bestimmen ein Dreieck mit dem Flächeninhalt $\frac{\sqrt{3}}{4}$. Es existiert aber auch ein dreidimensionales Simplex mit nicht-verschwindendem Volumen, dessen Seitenflächen die Inhalte $\frac{\sqrt{3}}{4}$ haben, nämlich das Tetraeder, dessen sechs Kanten die Länge 1 besitzen.

Zur Festlegung der Nullstellen der n -dimensionalen Volumsfunktion müssen wir also, solange wir den Raum keiner Einschränkung unterwerfen, entweder auch auf die $(n - 2)$ -dimensionalen und noch niedrigerdimensionalen Metriken zurückgreifen (was sehr kompliziert ist), oder wir müssen uns damit begnügen, die Gesamtheit der $(n + 1)$ -Tupel, für welche das n -dimensionale Volumen verschwindet, implizit festzulegen durch die Bedingungen:

a) *Je $n + 1$ Punkte, von denen einer zwischen den n anderen liegt, bestimmen ein Simplex mit dem n -dimensionalen Volumen Null.*

b) *Bestimmen für $n + 2$ Punkte $n + 1$ von den $n + 2$ aus ihnen bildbaren $(n + 1)$ -Tupeln das n -dimensionale Volumen Null, dann bestimmen alle aus ihnen bildbaren $(n + 1)$ -Tupel das n -dimensionale Volumen Null.*

c) *Bestimmt jedes der $(n + 1)$ -Tupel einer im Sinne des (Abstandslimes oder allgemeiner des) Volumslimes konvergenten Folge von $(n + 1)$ -Tupeln das n -dimensionale Volumen Null, dann bestimmt auch das Grenzwert $(n + 1)$ -Tupel das n -dimensionale Volumen Null.*

Die Forderung a) ist die mittlere Nullbedingung, die Forderung b) ist eine unmittelbare Folge der Simplexungleichung U_n , die Forderung c) eine Folge der aus U_n sich ergebenden Stetigkeit der Metrik.

6. Das Nichtverschwinden der Volumsfunktion.

Wir können auf Grund der impliziten Aussagen b) und c) aus der Forderung a) explizite Aussagen über das Verschwinden der Volumsfunktion herleiten. Wir wollen sagen, $n + 1$ Punkte bestimmen *von nullter Ordnung*

das n -dimensionale Volumen Null, wenn sie ein unechtes $(n + 1)$ -Tupel bilden, d. h. identische Punkte enthalten und daher schon auf Grund der schwachen Nullbedingung das n -dimensionale Volumen Null bestimmen. Wir sagen, $n + 1$ Punkte bestimmen *von erster Ordnung* das Volumen Null, wenn einer von ihnen zwischen den n übrigen liegt, so daß die $n + 1$ Punkte auf Grund der mittleren Nullbedingung das Volumen Null bestimmen. Wir sagen, $n + 1$ Punkte bestimmen *von höchstens zweiter Ordnung* das Volumen Null, wenn ein $n + 2$ -ter Punkt existiert, welcher zusammen mit je n von den gegebenen $n + 1$ Punkten das Volumen Null bestimmt, so daß also die $n + 1$ Punkte auf Grund von einmaliger Anwendung der Simplexungleichung das Volumen Null bestimmen. Schon in euklidischen Räumen gibt es, wie wir gesehen haben, Punkte, welche von genau zweiter (d. h. von höchstens zweiter, aber nicht schon von erster) Ordnung das Volumen bestimmen (z. B. bestimmen die vier Eckpunkte eines Quadrates von genau zweiter Ordnung das dreidimensionale Volumen Null). Eine $(n - 1)$ -dimensionale Abhängigkeit von höherer als zweiter Ordnung in diesem Sinn (d. h. die Bestimmung des n -dimensionalen Volumens Null auf Grund der Forderungen a), b), c) von anderer als nullter, erster oder zweiter Ordnung) ist in euklidischen Räumen unmöglich. In allgemeinen Räumen hingegen müssen wir weitergehen und höhere Ordnungen für das Verschwinden der Volumensfunktion herleiten. Um dies einzusehen, betrachten wir die Flächenmetrik von folgendem einfachen metrischen Raum. Der Raum enthalte sieben Punkte $p_1, p_2, p_3, p_4, p_5, p_6, p_7$, für welche folgende Abstände festgesetzt sind:

$$p_1 p_6 = p_1 p_7 = p_2 p_5 = p_2 p_7 = p_3 p_5 = p_3 p_6 = p_4 p_5 = p_4 p_6 = p_4 p_7 = 1, \\ p_1 p_2 = p_2 p_3 = p_3 p_1 = p_4 p_1 = p_4 p_2 = p_4 p_3 = 2.$$

Die sechs übrigen Abstände werden wir hernach festsetzen. Auf Grund der bisherigen Festsetzungen liegt:

$$p_5 \text{ zwischen } p_2 \text{ und } p_3, p_3 \text{ und } p_4, p_4 \text{ und } p_2, \\ p_6 \text{ zwischen } p_1 \text{ und } p_3, p_3 \text{ und } p_4, p_4 \text{ und } p_1, \\ p_7 \text{ zwischen } p_1 \text{ und } p_2, p_2 \text{ und } p_4, p_4 \text{ und } p_1.$$

Also sind, wenn wir zu dieser Abstandsmetrik eine den oben auseinandergesetzten Bedingungen genügende Flächenfunktion einführen, die Inhalte

$$p_2 p_3 p_4, p_1 p_3 p_4, p_1 p_2 p_4$$

gleich Null und zwar auf Grund der für die Punkte p_5, p_6, p_7 hergeleiteten Zwischenbeziehungen von höchstens zweiter Ordnung, aber auch sicher genau von zweiter Ordnung, weil in den drei Tripeln (p_2, p_3, p_4) , (p_1, p_3, p_4) , (p_1, p_2, p_4) kein Punkt Zwischenpunkt der beiden anderen ist.

Da jedes dieser drei Tripel den Inhalt Null bestimmt, ist nach der Tetraederungleichung auch $p_1 p_2 p_3 = 0$. Wir wollen nun die noch zu bestimmenden Abstände so festsetzen, daß p_1, p_2, p_3 von höherer als zweiter Ordnung, d. h. nicht von höchstens zweiter Ordnung den Inhalt Null bestimmen. Wir setzen zu diesem Zweck

$$p_5 p_6 = p_6 p_7 = p_7 p_5 = 2, \quad p_1 p_5 = p_2 p_6 = p_3 p_7 = \frac{5}{2}.$$

Um zu zeigen, daß die Punkte p_1, p_2, p_3 nicht von höchstens zweiter Ordnung den Inhalt Null bestimmen, haben wir bloß zu verifizieren, daß weder p_5 , noch p_6 , noch p_7 , sowohl mit p_1 und p_2 , als auch mit p_2 und p_3 , als auch mit p_3 und p_1 von erster Ordnung das Volumen Null bestimmen (von p_4 wissen wir dies schon). Aus Symmetriegründen genügt es, den Punkt p_5 zu betrachten. Derselbe liegt zwar zwischen p_2 und p_3 , aber es ist $p_1 p_5 = \frac{5}{2}$, $p_2 p_5 = 1$, $p_1 p_2 = 2$, also ist von den drei Punkten p_1, p_2, p_5 keiner Zwischenpunkt der beiden anderen, d. h. diese drei Punkte bestimmen nicht von erster Ordnung den Inhalt Null.

Es bestimmen also in unserem Beispiel die Punkte p_1, p_2, p_3 auf Grund der Forderungen a), b), c) den Inhalt Null, aber nicht von höchstens zweiter Ordnung. Wir können in unserem Beispiel sagen, sie bestimmen den Inhalt Null von dritter Ordnung, denn es existiert ein Punkt, nämlich p_4 , welcher mit je zweien von ihnen von zweiter Ordnung den Inhalt Null bestimmt.

Es ist klar, daß wir in dieser Weise mit der Definition von Ordnungen fortfahren können, und in allgemeinen metrischen Räumen auf Grund der Forderungen b) und c) fortfahren müssen. Es zeigt sich, was hier ohne Beweis erwähnt sei, daß man in kompakten Räumen nicht über die erste und zweite Klasse von Ordinalzahlen hinauszugehen braucht. Für eine Zahl α der ersten oder zweiten Zahlenklasse aber kann man definieren:

Wenn α eine isolierte Zahl ist, so bestimmen drei Punkte den Inhalt Null von höchstens α -ter Ordnung, falls ein vierter Punkt existiert, welcher mit je zweien der drei gegebenen Punkte den Inhalt Null von höchstens $(\alpha - 1)$ -ter Ordnung bestimmt.

Wenn α eine Grenzzahl ist, so bestimmen drei Punkte den Inhalt Null von höchstens α -ter Ordnung, wenn sie Limestripel einer Folge von Punkttripeln sind, welche durchwegs von einer Ordnung $< \alpha$ den Inhalt Null bestimmen.

Diese beiden Forderungen sind die den Forderungen b) und c) entsprechenden expliziten Aussagen für den Fall $n = 2$ und ganz analog liegen offenbar die Verhältnisse für höhere Dimensionen.

Die Abgeschlossenheit der Menge aller Punkttripel, welche für ein gegebenes α von höchstens α -ter Ordnung den Inhalt Null bestimmen, ist leicht zu zeigen.

Nun ist a priori natürlich der Fall möglich, daß die n -dimensionale Volumensfunktion auch für $(n + 1)$ -Tupel, welche *nicht* durch Induktion bzw. durch transfiniten Induktion auf diesen Wegen erhalten werden können, verschwindet. Es ist ja beispielsweise möglich, daß *allen* Punktetripeln eines metrischen Raumes der Inhalt Null zugeordnet wird, obwohl nur *gewisse* Punktetripel auf Grund der Abstandsmetrik von irgendeiner Ordnung α im definierten Sinn den Inhalt Null bestimmen. Diese Möglichkeit wollen wir nun aber durch eine Forderung ausschließen und diese Forderung stellt offenbar das Analogon zu der in euklidischen Räumen gültigen Positivitätsbedingung dar.

P_n (Allgemeine Positivitätsbedingung): *Je $n + 1$ Punkte, welche nicht von irgendeiner Ordnung α das Volumen Null bestimmen, bestimmen ein positives Volumen.*

Es ist klar, daß in euklidischen Räumen diese Bedingung mit der oben erwähnten Positivitätsbedingung der euklidischen Metrik identisch ist. Denn je $n + 1$ Punkte, welche in einer $(n - 1)$ -dimensionalen Ebene liegen, also je $n + 1$ Punkte, welche in einem euklidischen Raum das n -dimensionale Volumen Null bestimmen, bestimmen, wie wir sahen, von endlicher, ja sogar von höchstens zweiter Ordnung das Volumen Null.

Betrachten wir insbesondere unsere allgemeine Positivitätsbedingung für $n = 1$, d. h. für den Fall der Abstandsmetrik. Von nullter Ordnung bestimmen den Abstand Null die unechten Punktepaare, d. h. die Paare identischer Punkte. Von erster Ordnung bestimmen den Abstand Null zwei Punkte a und b dann, wenn ein dritter Punkt c existiert, der sowohl mit a als auch mit b von nullter Ordnung den Abstand Null bestimmt, d. h. ein Punkt c , der sowohl mit a als auch mit b identisch ist usw. Mit Rücksicht auf die Transitivitätseigenschaft der Identitätsrelation führt aber diese sukzessive Ordnungsbildung über die Identität nicht hinaus, so daß unsere allgemeine Positivitätsforderung für den Fall der Abstandsmetrik besagt, daß je zwei *nicht identische Punkte einen positiven Abstand besitzen* sollen. Das ist aber die Positivitätsforderung in der üblichen Definition der Abstandsmetrik. (Zwei Punkte haben den Abstand Null *nur dann*, wenn sie identisch sind.)

7. Die Vollständigkeit der Simplexungleichung.

Wir haben oben festgestellt, daß die Dreiecksungleichung eine nicht verschärfbare und durch keine weitere Dreipunktrelation ergänzbare Dreipunktebeziehung ist, sofern wir die euklidischen Räume unter die allgemeinen Räume mit Abstandsmetrik einordnen wollen. Wie liegen die Verhältnisse für höhere Dimensionen? Ist die Simplexungleichung eine in diesem Sinn vollständige $(n + 2)$ -Tupelrelation?

Es gilt diesbezüglich folgender elementargeometrischer Satz, den ich hier ohne seinen einfachen Beweis anführe:

α) Sind $n + 1$ positive (durchwegs von Null verschiedene) Zahlen vorgegeben, so ist für die Existenz eines n -dimensionalen Simplexes, dessen Seitenflächen die gegebenen Zahlen zu Volumina haben, notwendig und hinreichend, daß keine der $n + 1$ gegebenen Zahlen größer als die Summe der n übrigen ist.

β) Sind $n + 1$ nicht-negative Zahlen a_1, a_2, \dots, a_{n+1} gegeben, worunter sich mindestens eine Null befindet, dann ist für die Existenz eines n -dimensionalen Simplexes, dessen Seitenflächen die gegebenen Zahlen zu Volumina haben, notwendig und hinreichend, daß sich $n + 1$ Zahlen $\varepsilon_i = \pm 1$ ($i = 1, 2, \dots, n + 1$) bestimmen lassen, so daß die Relation gilt

$$\sum_{i=1}^{n+1} \varepsilon_i \cdot a_i = 0.$$

Im Falle $n = 2$ ist die Zusatzbedingung β überflüssig. Denn wenn drei Zahlen a_1, a_2, a_3 gegeben sind, worunter etwa $a_3 = 0$ ist, so ist schon auf Grund der Bedingung α), d. h. auf Grund der Dreiecksungleichung, gewährleistet, daß die beiden anderen Zahlen gleich sein müssen, also $\varepsilon_1 = +1, \varepsilon_2 = -1, \varepsilon_3 = \pm 1$ gewählt werden können. Für $n = 3$ besagt die Bedingung β), daß, wenn für vier Punkte eines der aus ihnen bildbaren Tripel den Inhalt Null hat, der Inhalt von einem der drei übrigen Tripel gleich der Summe der zwei anderen ist. Für beliebiges n ist der Fall, daß eine der Zahlen $\varepsilon_i = +1$, alle übrigen $= -1$ gewählt werden können, charakteristisch dafür, daß einer der $n + 1$ Punkte des den Zahlen a_1, a_2, \dots, a_{n+1} entsprechenden degenerierenden Simplexes zwischen den übrigen n Punkten liegt. Allgemein entsprechen die $2^n - 1$ Vorzeichenkombinationen für die ε_i den $2^n - 1$ Teilen, in welche der R_n durch $n + 1$ Hyperebenen zerlegt wird.

Satz α ergibt nun, daß die Simplexungleichung hinsichtlich der Simplexe mit nicht-verschwindenden Volumen vollständig ist. Würde hinsichtlich der Simplexe mit nicht-verschwindendem Volumen zur Simplexungleichung noch eine über sie hinausgehende $(n + 2)$ -Punktebeziehung hinzugenommen, so ließen sich die euklidischen Räume nicht unter die allgemeinen Räume mit n -dimensionaler Metrik einordnen.

Anders liegen in gewissem Maße die Verhältnisse hinsichtlich der Simplexe mit verschwindendem Volumen. Es wäre nämlich nach β) mit den Eigenschaften der euklidischen Metrik verträglich, wenn unter die Forderungen an die allgemeine Metrik auch die folgende aufgenommen würde, welche wir hier der Einfachheit halber bloß für $n = 2$ aussprechen:

Wenn für vier Punkte eines der aus ihnen bildbaren Tripel den Inhalt Null bestimmt, dann ist von den Inhalten der drei übrigen Tripel einer gleich der Summe der beiden anderen.

Es ist jedoch fraglich, ob diese Forderung, wenn der Raum keinerlei Einschränkungen unterworfen wird und Tripel enthalten kann, die von höherer als erster Ordnung den Inhalt Null bestimmen, mit den übrigen Forderungen an die Flächenmetrik verträglich ist. Die Forderung ist mit der Tetraederungleichung und der Positivitätsbedingung im allgemeinen sicher unverträglich, wenn sie durch die folgende, in euklidischen Räumen gleichfalls gültige Zusatzforderung ergänzt wird:

Liegt von vier Punkten a, b, c, d etwa b zwischen a und c (d. h. gilt $ab + bc = ac$), dann ist $abd + bcd = acd$.

Betrachten wir nämlich etwa einen Raum, welcher fünf Punkte a, b, c, d, e enthält, zwischen denen folgende Abstände festgesetzt sind:

$$ab = ac = ad = ae = 1, \quad bc = bd = cd = 2, \quad ed = ce = de = \sqrt{2}.$$

Wollen wir in diesem Raum eine den auseinandergesetzten Bedingungen genügende Flächenmetrik einführen, so haben wir zunächst zu beachten, daß a Zwischenpunkt von b und c , von c und d und von d und b ist. Es muß also jedenfalls festgesetzt werden:

$$abc = abd = acd = bcd = 0.$$

(Die Punkte b, c, d bestimmen den Inhalt Null von zweiter Ordnung.) Weitere Tripel, welche den Inhalt Null bestimmen, können, wie man auf Grund der Positivitätsbedingung leicht verifiziert, nicht existieren. Nun müßte auf Grund der angeführten Zusatzforderung gelten:

$$abe + ace = bce, \quad abe + ade = bde, \quad ace + ade = cde.$$

Betrachten wir das Punktequadrupel b, c, d, e . Es ist $bcd = 0$. Trotzdem ist unmöglich, daß eine der drei Zahlen bce, bde, cde gleich der Summe der beiden anderen ist. Denn beispielsweise aus $bce = bde + cde$ würde wegen $bce = abe + ace$ und $bde + cde = abe + ace + 2 \cdot ade$ folgen: $ade = 0$, im Widerspruch zur Positivitätsbedingung.

8. Flächenmetrik in Räumen mit Dreitripel- und Zweitripel-eigenschaft.

Die Schwierigkeit der Probleme vermindert sich und der Bereich der möglichen Aussagen vergrößert sich erheblich, wenn wir über den Raum, in welchem eine n -dimensionale Metrik definiert ist, beschränkende Voraussetzungen machen. Dies sei hier zum Abschluß für den Fall der Flächenmetrik kurz angedeutet.

In euklidischen Räumen gilt der Satz, daß drei Punkte den Flächeninhalt Null dann und nur dann bestimmen, wenn sie auf einer Geraden liegen, und daß drei Punkte dann und nur dann auf einer Geraden liegen, wenn einer von ihnen Zwischenpunkt der beiden anderen ist. Aus diesem

Grunde können drei Punkte eines euklidischen Raumes den Flächeninhalt Null nur entweder von nullter oder von erster Ordnung bestimmen (starke Nullbedingung). In allgemeinen metrischen Räumen können, wie wir festgestellt haben, auch höhere Ordnungen auftreten.

Besonders weittragend ist nun aber die Einführung der Flächenmetrik in Räume, welche die erwähnte Eigenschaft mit den euklidischen Räumen gemein haben: in Räume also, in welchen Punktetripel, die von höherer als erster Ordnung den Flächeninhalt Null bestimmen, nicht auftreten; ausführlich gesprochen, in Räume, in welchen zu keinem Punktetripel a, b, c , das nicht auf einer Geraden liegt, ein vierter Punkt d existiert, so daß in jedem der drei Tripel (a, b, d) , (a, c, d) , (b, c, d) einer der drei Punkte zwischen den beiden anderen liegt. Auf diese Räume sind wir aber in der ersten Untersuchung über allgemeine Metrik (S. 107) bereits von einem anderen Gesichtspunkt aus geführt worden. Wir haben diese Räume als *Räume mit der Dreitripleigenschaft* bezeichnet und gezeigt, daß sie unter den vollständigen konvexen Räumen charakterisiert sind durch die Abwesenheit konvexer Dreibeine.

Ist also in einem Raum mit der Dreitripleigenschaft eine den allgemeinen Forderungen genügende Flächenmetrik definiert, so ist der Inhalt eines Punktetripels dann und nur dann Null, wenn einer der drei Punkte zwischen den anderen liegt.

Als wichtigste Unterklasse dieser Klasse von Räumen haben wir *Räume mit der Zweitripleigenschaft* betrachtet (S. 111) und diese letzteren unter den vollständigen konvexen Räumen dadurch gekennzeichnet, daß in ihnen je zwei verschiedene Punkte eine Gerade bestimmen, die durch je zwei verschiedene ihrer Punkte bestimmt ist. Wir wollen nun zeigen, daß diese Räume auch vom Standpunkt der Flächenmetrik eine wichtige Eigenschaft besitzen:

Ist in einem Raum mit der Zweitripleigenschaft eine den allgemeinen Bedingungen genügende Flächenfunktion definiert, so ist für zwei geordnete Punktetripel der mit Hilfe der Flächenfunktion definierte Abstand dann und nur dann gleich Null, wenn die beiden Punktetripel identisch sind. Für identische Punktetripel ist der Abstand auf Grund der schwachen Nullbedingung in jedem Raum mit Flächenmetrik gleich Null. Es seien nun in einem Raum mit der Dreitripleigenschaft zwei nicht-identische Punktetripel a, b, c und a', b', c' gegeben, von denen etwa das erste einen Flächeninhalt $abc > 0$ bestimmt. Da die beiden geordneten Tripel als nicht identisch vorausgesetzt wurden, sind mindestens zwei entsprechende Punkte verschieden. Es ist also etwa $a \neq a'$. Wir betrachten nun die beiden Flächeninhalte $aa'b$ und $aa'c$ und leiten zunächst einen Widerspruch her aus der Annahme, daß diese

beiden Zahlen $= 0$ seien. Da der Raum nach Voraussetzung die Zweitripel-, und daher erst recht die Dreitripel-eigenschaft besitzt, haben nur jene Tripel den Flächeninhalt Null, in denen ein Punkt Zwischenpunkt der beiden anderen ist. Aus der Annahme würde also folgen, daß in jedem der beiden Tripel a, a', b und a, a', c einer der Punkte zwischen den beiden anderen liegt. Da der Raum die Zweitripel-eigenschaft besitzt, müßte also auch von den drei Punkten a, b, c einer zwischen den beiden anderen liegen, im Widerspruch zur Annahme, daß $abc > 0$ ist. Es ist also von den beiden Zahlen $aa'b$ und $aa'c$ mindestens eine > 0 . Da diese beiden Zahlen unter den zwölf durchwegs nicht-negativen Summanden des Abstandes $r \begin{pmatrix} a, b, c \\ a', b', c' \end{pmatrix}$ vorkommen, folgt daraus, daß die Tripel a, b, c und a', b', c' einen positiven Abstand haben.

Wir beweisen nun die Umkehrung dieser Behauptung: *Ist in einem Raum, welcher die Dreitripel-, aber nicht die Zweitripel-eigenschaft besitzt, eine den allgemeinen Bedingungen genügende Flächenmetrik definiert, dann enthält der Raum zwei nicht identische Punkttripel mit positivem Flächeninhalt, deren Abstand im Sinne der Flächenmetrik gleich Null ist.* In der Tat, der Raum besitzt nicht die Zweitripel-eigenschaft, enthält also ein Punktequadrupel a, b, c, d , so daß für zwei aus dem Quadrupel bildbare Tripel etwa für die Tripel a, b, c und a, b, d einer der Punkte zwischen den beiden anderen liegt, daß dies aber für eines der im Quadrupel enthaltenen Tripel, etwa für b, c, d , nicht der Fall ist. Diese vier Punkte a, b, c, d sind offenbar paarweise verschieden. Da der Raum nach Voraussetzung die Dreitripel-eigenschaft besitzt, ist $abc = abd = 0$, $bcd > 0$. Betrachten wir nun die beiden nicht identischen Tripel b, c, d und a, c, d . Es hat bcd sicher einen positiven Inhalt. Berechnen wir ihren Abstand, so ergibt sich $r \begin{pmatrix} b, c, d \\ a, c, d \end{pmatrix} = abc + abd = 0$. Damit ist die Behauptung bewiesen.

(Eingegangen am 10. 5. 1927.)

Über die charakteristischen Zahlen einer Kurve vom Maximal-Klassenindex¹⁾.

Von

Julius v. Sz. Nagy in Klausenburg.

1. Einleitung.

Unter einer *Kurve* verstehen wir eine reelle stetige und geschlossene ebene Kurve, die aus endlichvielen konvexen Bögen besteht und in jedem Punkte eine bestimmte, mit dem Berührungspunkte sich stetig ändernde Tangente hat. Die Klasse bzw. der Klassenindex der Kurve ist die größte bzw. kleinste Anzahl der Tangenten der Kurve, die von einem beliebigen Punkte ihrer Ebene ausgehen.

In einer vorigen Abhandlung²⁾ haben wir für eine Kurve n -ter Klasse vom Maximal-Klassenindex das Geschlecht p mit dem Zusammenhange eines gewissen, der Kurve zugehörigen Gebietes T definiert. Dieses Gebiet wird von den Punkten der Ebene gebildet, aus denen $n - 2$ Tangenten an die Kurve gezogen werden können. Ist das Gebiet T $p + 1$ -fach zusammenhängend, so ist p das Geschlecht der Kurve. Zerfällt das Gebiet T in die zusammenhängenden Gebiete T_1, T_2, \dots, T_k mit dem Zusammenhange $p_1 + 1, p_2 + 1, \dots, p_k + 1$, so ist

$$p = p_1 + p_2 + \dots + p_k - k + 1.$$

In dieser Abhandlung beweisen wir die folgenden Hauptsätze:

I. *Hat eine ebene Kurve n -ter Klasse vom Maximal-Klassenindex r Spitzen erster Art, t Doppeltangenten, w Wendetangenten ($w = 0$ oder 1)*

¹⁾ Diese Abhandlung wurde am 17. Mai 1926 der ungarischen Akademie der Wissenschaften vorgelegt.

²⁾ „Über Kurven vom Maximal-Klassenindex. Über Kurven vom Maximalindex“, *Math. Annalen* 89 (1923), S. 32—75 [Berichtigung *Math. Annalen* 90 (1923), S. 142 bis 143]. Diese Abhandlung wird im folgenden mit. A. zitiert.

und das Geschlecht p , so sind

$$r = n - 2 + 2p \quad \text{und} \quad t + w = \frac{(n-1)(n-2)}{2} - p.$$

II. Stimmen für eine algebraische Kurve vom Maximal-Klassenindex die algebraische und Realitätsklasse überein, so stimmt auch das algebraische und das Realitätsgeschlecht überein und die Kurve hat nur reelle Tangentensingularitäten.

III. Eine einzügige algebraische Kurve von der algebraischen Klasse n und vom Klassenindex $n - 2$ ist entweder rational oder elliptisch.

Auf Grund des Dualitätsprinzips lassen sich entsprechende Sätze für die ebenen Kurven vom Maximalindex aussprechen.

Der zu Satz I duale Satz ist eine Verallgemeinerung einiger Sätze von Herrn C. Juel über die Anzahl der Doppelpunkte und der Wendetangenten für eine ebene Projektion einer Raumkurve vom Maximalindex, die auf einer Regelfläche zweiter Ordnung liegt³⁾.

2. Elementare Deformationen.

Wir beweisen unsere Sätze mittels fünf elementarer Deformationen α , β , γ , δ und ε .

Wir ersetzen einen Elementarbogen \widehat{AB} einer Kurve C vom Maximalindex, die keine Wendetangente hat, durch einen Elementarbogen mit denselben Endpunkten, so daß die Punkte A und B für die neue Kurve C' ebensolche (gewöhnliche Punkte oder Spitzen erster Art) Punkte sind, wie sie für C waren. Liegt kein Punkt von C im Gebiete zwischen den zwei Elementarbögen mit den Endpunkten A und B , so wendet man auf die Kurve C eine *Deformation* α an.

Ersetzt man die Elementarbögen \widehat{AS} und \widehat{SB} der Kurve C , von denen im Punkte S eine Spitze erster Art gebildet wird, durch solche Elementarbögen \widehat{AS}' und $\widehat{S'B}$, von denen im Punkte S' eine Spitze erster Art gebildet wird, sind ferner die Punkte A und B für die neue Kurve C' ebensolche Punkte, wie für die Kurve C , und liegt endlich kein Punkt der Kurve C in dem von den Bögen ASB und $AS'B$ begrenzten Gebiete, so wendet man auf die Kurve C eine *Deformation* β an.

Wir wenden auf die Kurve C die *Deformation* γ an, wenn wir einen ihrer Elementarbögen durch eine Deformation α so lange deformieren, bis er einen anderen Elementarbogen von C berührt. Die so erhaltenen zwei, einander berührenden Elementarbögen liegen an den entgegengesetzten

³⁾ „Die gewundenen Kurven vom Maximalindex auf einer Regelfläche zweiter Ordnung.“ Kgl. Danske Vidensk. Selks. Skrifter, Nat. og Math. Afd., 8. Raekke (1917), S. 279—294.

Seiten der gemeinsamen Tangente. Dies folgt aus der Eigenschaft der Kurve C vom Maximal-Klassenindex, daß man aus einem Punkte der Ebene entweder die konvexen oder die konkaven Seiten der Kurve C erreichen kann, ohne die Kurve zu überschreiten⁴⁾.

Die Kurve C' , in welche die Kurve C durch Deformationen α , β und γ überführt wird, ist wieder vom Maximal-Klassenindex, weil sie keine stationäre und keine Wendetangente hat und weil man aus einem Punkte der Ebene dieselben Seiten ihrer Elementarbögen erreichen kann. Bewegt sich ein Punkt P in der Ebene, so kann sich die Anzahl der Tangenten der Kurve C aus dem Punkte P nur dann ändern, wenn der Punkt P die Kurve C überschreitet. Die Veränderung ist $+2$ bzw. -2 , je nachdem der Punkt P einen Elementarbogen von C von seiner konkaven bzw. von seiner konvexen Seite aus übertritt.

Berühren die Elementarbögen \widehat{AB} und \widehat{CD} der Kurve C (nach der Durchführung einer Deformation γ) einander im Punkte M , so bildet der Elementarbogen \widehat{AM} mit einem der Elementarbögen \widehat{CM} und \widehat{DM} im Punkte M eine Spitze erster Art. Ist \widehat{CM} dieser Elementarbogen, so bilden auch die Elementarbögen \widehat{BM} und \widehat{DM} im Punkte M wieder eine Spitze erster Art. Durch die Deformation δ werden diese zwei Spitzen im Punkte M voneinander getrennt.

Die Deformation ε bringt durch Deformationen β die durch eine Deformation δ entstandenen zwei Spitzen voneinander in endlichen Abstand.

Die Umkehrung einer Elementardeformation α bzw. β ist wieder eine Deformation α bzw. β . Unter den fünf Elementardeformationen ist nur die Deformation δ unstetig.

3. Die Anwendbarkeit der Deformationen γ , δ , ε auf eine im Endlichen liegende einzügige Kurve C n -ter Klasse vom Maximal-Klassenindex und vom Geschlechte eins.

Die Deformationen α und β lassen sich auf jede Kurve vom Maximal-Klassenindex anwenden. Wir wollen jetzt beweisen, daß man auf eine einzügige, im Endlichen liegende Kurve C vom Maximal-Klassenindex, deren Ordnung $n > 3$ und deren Geschlecht eins ist, auch die Deformation γ (und nachher die Deformationen δ und ε) anwenden kann.

Die Kurve C hat das Geschlecht eins, also bilden die Punkte, aus denen $n - 2$ Tangenten an die Kurve gehen, ein zweifach zusammenhängendes Gebiet T_1 . Wegen der Einzigigkeit von C ist demnach das von C begrenzte endliche Gebiet T_0 (aus dessen Punkten an die Kurve

⁴⁾ Vgl. A., S. 50.

n Tangenten gehen) einfach zusammenhängend. Aus einem Punkte des Gebietes T_1 bzw. T_0 ist die konkave bzw. konvexe Seite der Kurve erreichbar, ohne die Kurve zu überschreiten.

Ist P das kleinste geradlinige konvexe Polygon, das sich um die Spitzen der Kurve C spannen läßt, so hat die Kurve keinen Punkt außerhalb des Polygons P .

Würde nämlich die im Endlichen liegende Kurve C von einer Stützgeraden e des Polygons P geschnitten, so gäbe es auf der mit P entgegengesetzten Seite der Geraden e eine mit e parallele Gerade e_1 , von der die Kurve C nicht getroffen wird. Läßt man nun eine Gerade g von e_1 ausgehend parallel der Geraden e sich nähern, so wird die erste der Geraden g , von der die Kurve C getroffen wird, eine Tangente t , weil kein singulärer Punkt der Kurve außerhalb des Polygons P liegen kann.

Auf der Tangente t kann man aber im Berührungspunkte die konvexe Seite der Kurve erreichen. Daraus folgt, daß man aus den Punkten der Geraden e_1 und t die konvexe Seite der Kurve erreichen kann. Dies ist aber unmöglich, weil das einfach zusammenhängende Gebiet T_0 keine ganze Geraden enthalten kann. Daraus folgt die Richtigkeit der Behauptung. Aus der Behauptung geht insbesondere hervor, daß das Polygon P mindestens drei Eckpunkte enthält, da sich die Kurve C andernfalls nach dem Bewiesenen auf einen Punkt bzw. auf eine Strecke reduzieren müßte.

Die Kurve hat also mindestens drei Spitzen. Hat die Kurve C genau drei Spitzen erster Art, so besteht sie aus drei Elementarbögen, deren Endpunkte die Spitzen der Kurve sind. Bezeichnen nämlich A und B zwei Spitzen der Kurve C , so schneidet die Stützgerade AB des Polygons P den singularitätenfreien Bogen \widehat{AB} außerhalb der Punkte A und B in keinem anderen Punkte. Daraus folgt, daß der Bogen \widehat{AB} ein Elementarbogen ist. Eine solche Kurve ist dritter Klasse und hat keine Doppeltangente⁵⁾.

Ist die Klasse n der Kurve C vom Maximal-Klassenindex vom Geschlechte eins größer als drei, so ist auch die Anzahl r der Spitzen größer als drei. Sobald $r > 3$ ist, kann angenommen werden, daß die Kurve auch einen Elementarbogen hat, dessen Endtangente sich im Gebiete T_1 schneiden. Tritt dies nicht ein, so zeigen wir, daß die Kurve C sich durch Deformationen α und β in eine Kurve C' überführen läßt, für welche die Annahme erfüllt ist.

Läuft ein Punkt die Kurve C in positiver Richtung um (so daß das endliche Gebiet T_0 links bleibt), so werde die Reihenfolge der Spitzen bzw. ihrer Tangenten mit A_1, A_2, \dots, A_r bzw. a_1, a_2, \dots, a_r bezeichnet. Aus

⁵⁾ Vgl. A., S. 60.

der Annahme, daß die Kurve C keinen solchen Elementarbogen hat, dessen Endtangente sich im Gebiete T_1 schneiden, folgt, daß die singularitätenfreien Bögen $\widehat{A_1 A_2}$, $\widehat{A_2 A_3}$, ..., $\widehat{A_r A_1}$ alle Elementarbögen, die Winkel $(a_1 a_2)$, $(a_2 a_3)$, ..., $(a_r a_1)$ alle positiv und kleiner als π sind und ihre Summe 2π ausmacht.

Ist nämlich B_k der Schnittpunkt der Geraden a_k und a_{k+1} und durchläuft der Punkt P , dessen Tangente p ist, den Bogen $\widehat{A_k A_{k+1}}$ ($k=1, 2, \dots, r$; $A_{r+1} \equiv A_1$) von A_k aus, so durchläuft der Schnittpunkt Q der Tangenten a_k und p die im Gebiete T_0 liegende endliche Strecke $A_k B_k$ einfach. Der Punkt Q bewegt sich nämlich auf a_k in einem bestimmten Sinne, während der Punkt P den von Singularitäten freien Bogen $\widehat{A_k A_{k+1}}$ durchläuft, und er kann aus dem Gebiete T_0 nicht austreten. Widrigenfalls gäbe es auf dem Bogen $\widehat{A_k A_{k+1}}$ einen Elementarbogen, dessen Tangenten einander im Gebiete T_1 schneiden würden.

Aus dem Gesagten folgt, daß der Bogen $\widehat{A_k A_{k+1}}$ ein Elementarbogen ist, der ganz im Dreiecke $A_k B_k A_{k+1}$ liegt. Ähnlicherweise liegt der Elementarbogen $\widehat{A_{k+1} A_{k+2}}$ im Dreiecke $A_{k+1} B_{k+1} A_{k+2}$, daß außer den gemeinsamen Punkten der Strecken $A_{k+1} B_k$ und $A_{k+1} B_{k+1}$ mit dem Dreiecke $A_k B_k A_{k+1}$ keinen anderen gemeinsamen Punkt haben kann. Die Winkel (a_1, a_2) , (a_2, a_3) , ..., (a_r, a_1) sind also alle positiv und kleiner als π und ihre Summe macht 2π aus.

Es gibt also wenigstens ein Paar der nacheinander folgenden Winkel, deren Summe kleiner als π ist. Angenommen also, daß $(a_1 a_2) + (a_2 a_3) < \pi$ und A'_2 ein Punkt desjenigen Gebietes ist, welches von den Geraden a_1 und a_3 und von den Elementarbögen $\widehat{A_1 A_2}$ und $\widehat{A_2 A_3}$ begrenzt wird, so kann man den Bogen $\widehat{A_1 A_2 A_3}$ durch eine Deformation β in den Bogen $\widehat{A_1 A'_2 A_3}$ überführen, so daß die Spitzentangente im Punkte A'_2 mit a_1 parallel ist. Die Endtangente des

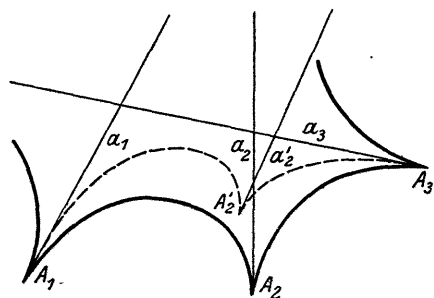


Fig. 1.

Elementar bogens $\widehat{A_1 A'_2}$ schneiden einander also im Gebiete T_1 . (Fig. 1.)

Ist nun \widehat{AB} ein Elementarbogen der Kurve C , dessen Endtangente einander im Gebiete T_1 schneiden, so kann man auf ihn die Deformation γ anwenden, weil es in dem von dem Elementarbogen \widehat{AB} und von seinen Endtangente begrenzten endlichen Gebiete auch solche Elementarbögen

mit den Endpunkten A und B gibt, von denen der übrigbleibende Bogen der Kurve C geschnitten wird. Daraus folgt, daß die Deformationen γ , δ , ε auf eine im Endlichen liegende Kurve $n(> 3)$ -ter Klasse vom Geschlechte eins — nach entsprechenden Deformationen α und β — immer anwendbar sind.

4. Die Einwirkungen der Elementardeformationen auf die Klasse und die Anzahl der Doppeltangenten einer Kurve n -ter Klasse vom Maximal-Klassenindex ohne Wendetangente.

Eine Kurve C n -ter Klasse vom Maximal-Klassenindex, die keine Wendetangente hat, geht durch die Elementardeformationen in eine Kurve C' über, die wieder vom Maximal-Klassenindex ist. Die Kurve C' hat nämlich keinen Doppelpunkt, keine stationäre und Wendetangente und aus einem Punkte der Ebene können entweder die konvexen oder die konkaven Seiten der Elementarbögen der Kurve C' erreicht werden, ohne C' zu überschreiten. Wird also bewiesen, daß aus einem Punkte, von dem dieselbe (konvexe oder konkave) Seite der Kurven C und C' erreicht werden kann, an beide Kurven die gleiche Anzahl von Tangenten gezogen werden kann, so haben die Kurven C und C' dieselbe Klasse.

Bedeutet P den Schnittpunkt der Endtangente des Elementarbogens \widehat{AB} und wendet man auf diesen Elementarbogen die Deformationen α , γ oder δ an, so gehen dieselben Tangenten an beide Kurven C und C' aus dem Punkte P . Deformiert man die Elementarbögen \widehat{AS} und \widehat{SB} , von denen im Punkte S eine Spitze erster Art gebildet wird, durch eine Deformation β oder ε , so geht aus dem Schnittpunkte P der Tangente der Kurve in den Punkten A und B noch eine und nur eine Tangente an jeden der beiden Bögen ASB und $AS'B$. Die Klassen der Kurven C und C' stimmen also überein.

Wendet man auf die Kurve C eine Deformation α oder β an, so bleibt auch die Anzahl der Doppeltangenten unverändert, wie aus den Untersuchungen von Herrn A. Kneser über die Gestalten ebener Kurven folgt⁶⁾.

Herr Kneser hat bewiesen, daß bei irgendeiner stetigen Deformation einer Kurve eine Doppeltangente einfach erhalten bleibt, solange ihre Berührungspunkte getrennt und nicht singular sind. Man kann aber ebenso beweisen, daß eine Doppeltangente auch dann einfach erhalten bleibt, wenn die Berührungspunkte getrennt sind und einer oder beide von ihnen mit einer oder mit zwei Spitzen erster Art zusammenfallen. Daraus folgt, daß

⁶⁾ „Einige allgemeine Sätze über die einfachsten Gestalten ebener Kurven“, *Math. Annalen* **41** (1893), §§ 17—19, S. 369—376.

die Anzahl der Doppeltangenten durch die Deformationen α und β unverändert bleibt.

Durch irgendeine Deformation γ fallen zwei Doppeltangenten in eine zusammen, welche durch die Deformation δ in die gemeinsame Tangente zweier Spitzen verwandelt wird. Diese gemeinsame Spitzentangente ist nur für eine Doppeltangente zu rechnen. Die nacheinandergeführten Deformationen γ , δ und ε vermindern also die Anzahl der Doppeltangenten um eins und vergrößern die Anzahl der Spitzen, die durch die Deformationen α , β und γ nicht verändert wird, um zwei.

5. Im Endlichen liegende einzügige Kurven vom Maximal-Klassenindex und vom Geschlechte eins.

Wir bezeichnen mit $C^{(1)}$ eine im Endlichen liegende Kurve n -ter Klasse vom Maximal-Klassenindex und vom Geschlechte eins, die r Spitzen erster Art und t Doppeltangenten hat. Durch Elementardeformationen, unter denen die Deformationen γ und δ je einmal vorkommen, wird die Kurve $C^{(1)}$ in eine zweizügige Kurve C_1 n -ter Klasse vom Maximal-Klassenindex überführt, die $r + 2$ Spitzen und $t - 1$ Doppeltangenten besitzt. Es ist klar, daß keiner der zwei Züge von C_1 den anderen Zug umgibt.

Wir werden auf die zwei Züge der Kurve C_1 vom Maximal-Klassenindex, die wieder Kurven vom Maximal-Klassenindex und vom Geschlechte eins sind, noch einmal die vorigen Deformationen an und wiederholen dieses Verfahren auf die entstandenen Kurvenzüge, solange die Deformationen γ , δ , ε angewendet werden können, d. h. so lange, bis alle entstandenen Züge dritter Klasse mit drei Spitzen sind. Hat die Kurve $C^{(1)}$ r Spitzen, so geht sie durch $(r - 3)$ -malige Anwendung der Deformationen γ , δ , ε in $r - 2$ Züge dritter Klasse mit je drei Spitzen über, von denen keiner Zug einen anderen umschließen kann.

Die Klasse dieser aus $r - 2$ Zügen dritter Klasse bestehenden Kurve C_{r-3} ist r , weil man aus einem innerhalb eines Zuges liegenden Punkte an diesen Zug drei Tangenten, an die übrigen $r - 3$ Züge je eine Tangente ziehen kann. Die Klasse der Kurve C_{r-3} ist aber (nach § 4) n , also ist $r = n = n + 2p - 2$.

Die Anzahl der Doppeltangenten der Kurve $C^{(1)}$ läßt sich auf Grund unseres folgenden Satzes bestimmen⁷⁾:

IV. Haben die einzügigen ebenen Kurven C_m und C_n keinen gemeinsamen Punkt, keinen Doppelpunkt und keine stationäre und Wendetangente, liegt ferner die Kurve C_m im Endlichen und gehen endlich n bzw. m

⁷⁾ „Über einen v. Staudtschen Satz“, Acta Litt. ac. Sc. Reg. Univ. Franciscose Josephinae, Szeged (Ungarn), Sect. Sc. Math. 2 (1924), S. 67—68.

Tangenten aus einem Punkte der Kurve C_m bzw. C_n an die andere Kurve, so haben die zwei Kurven $m \cdot n$ gemeinsame Tangenten.

Nach diesem Satze haben je zwei Züge dritter Klasse der Kurve C_{r-3} je eine gemeinsame Tangente, die Kurve C_{r-3} hat also $\frac{(n-2)(n-3)}{2}$ Doppeltangenten. Durch $(r-3)$ -fache Anwendung der Deformation δ wurde die Anzahl der Doppeltangenten der Kurve $C^{(1)}$ um $r-3 = n-3$ vermindert, also ist

$$t = \frac{(n-2)(n-3)}{2} + n - 3 = \frac{(n-1)(n-2)}{2} - 1.$$

Damit ist der Satz I für die Kurve $C^{(1)}$ bewiesen.

6. Im Endlichen liegende Kurven vom Maximal-Klassenindex und vom Geschlechte null.

Nimmt man in der Ebene eine dreispitzige Kurve \bar{C} dritter Klasse an, welche die im Endlichen liegende Kurve $C^{(0)}$ vom Maximal-Klassenindex und vom Geschlechte null umschließt, so kann man aus den Punkten des ringförmigen Gebietes zwischen den Kurven $C^{(0)}$ und \bar{C} die konvexen Seiten beider Kurven erreichen. Verkleinert man die Kurve \bar{C} so lange, bis sie die Kurve $C^{(0)}$ berührt, so werden die Kurven $C^{(0)}$ und \bar{C} durch Deformation δ, ε in einem der Berührungspunkte der zwei Kurven in eine einzügige Kurve $C^{(1)}$ vom Maximal-Klassenindex und vom Geschlechte eins vereinigt.

Ist n, r und t bzw. n', r' und t' die Klasse, die Anzahl der Spitzen und diejenige der Doppeltangenten der Kurve $C^{(0)}$ bzw. $C^{(1)}$, so bestehen die Gleichungen

$$r' = n', \quad n' = n + 3, \quad r' = r + 5, \quad t' = \frac{(n'-1)(n'-2)}{2} - 1, \quad t' = t + 3n - 1,$$

woraus $r = n - 2, \quad t = \frac{(n-1)(n-2)}{2}$ ist. Damit ist der Satz I für die Kurve $C^{(0)}$ bewiesen.

Um die Anzahl der Spitzen und Doppeltangenten einer im Endlichen liegenden einzügigen Kurve vom Maximal-Klassenindex bestimmen zu können, ist auch ein anderes Verfahren vorhanden. Es ist nicht schwer einzusehen, daß eine solche Kurve durch Deformationen α und β in ein konvexes oder konkaves Bogenpolygon⁸⁾ überführt werden kann, dessen Spitzen alle oder — im Falle Bogenpolygons — mit der Ausnahme einer Spitze alle auf einer Geraden liegen und ihre Spitzentangenten auf diese Gerade senkrecht fallen. Für ein solches Bogenpolygon kann man die Anzahl der Spitzen und die Anzahl der Doppeltangenten direkt bestimmen.

⁸⁾ Vgl. A., S. 36—41.

7. Im Endlichen liegende irreduzible Kurven vom Maximal-Klassenindex und vom Geschlechte p .

Eine Kurve C n -ter Klasse ist reduzibel und zerfällt in die Kurven C_1, C_2, \dots, C_k n_1 -ter, n_2 -ter, \dots , n_k -ter Klasse, wenn $n = n_1 + n_2 + \dots + n_k$ ist und die Kurven C_1, C_2, \dots, C_k zusammen jeden Zug der Kurve C enthalten, und zwar jeden nur einmal. Tritt das Zerfallen der Kurve bei keinerlei Einteilung ihrer Züge ein, so ist sie irreduzibel.

Eine Kurve n -ter Klasse $C^{(p)}$ vom Maximal-Klassenindex ist dann und nur dann irreduzibel, wenn die Punkte, aus denen an die Kurve $n - 2$ Tangenten gehen, ein zusammenhängendes, und zwar $p + 1$ -fach zusammenhängendes Gebiet T bilden⁹⁾. Man bekommt in der Ebene ein von im Endlichen liegenden Randkurven begrenztes $p + 1$ -fach zusammenhängendes Gebiet, wenn man entweder in der ganzen Ebene oder in einem von einer Randkurve begrenzten einfach zusammenhängenden ebenen Gebiete p Löcher bohrt. Daraus folgt, daß eine im Endlichen liegende irreduzible Kurve $C^{(p)}$ vom Geschlechte p und vom Maximal-Klassenindex aus p Zügen vom Geschlechte eins, von denen keiner einen anderen umschließt, und noch entweder aus einem oder aus keinem Zuge vom Geschlechte null besteht. Der Zug vom Geschlechte null umschließt die übrigen p Züge.

Besteht die Kurve $C^{(p)}$ aus den Zügen C_1, C_2, \dots, C_k ($k = p$ oder $p + 1$) und bezeichnen n, n_1, n_2, \dots, n_k ; r, r_1, r_2, \dots, r_k bzw. t, t_1, t_2, \dots, t_k die Klassen, die Anzahlen der Spitzen bzw. der Doppeltangenten für die Kurven $C^{(p)}, C_1, C_2, \dots, C_k$, so bestehen für $k = p$ die Gleichungen

$$\begin{aligned} n &= n_1 + n_2 + \dots + n_p - 2p + 2, & r &= r_1 + r_2 + \dots + r_p, \\ n_1 &= r_1, & n_2 &= r_2, & \dots, & n_p &= r_p, \\ t &= t_1 + t_2 + \dots + t_p + \sum_{\substack{h \neq j \\ (h, j = 1, 2, \dots, p)}} (n_h - 2)(n_j - 2), \\ t_h &= \frac{(n_h - 1)(n_h - 2)}{2} - 1 & (h &= 1, 2, \dots, p); \end{aligned}$$

für $k = p + 1$ aber die Gleichungen

$$\begin{aligned} n &= n_1 + n_2 + \dots + n_{p+1} - 2p, & r &= r_1 + r_2 + \dots + r_{p+1} \\ n_1 &= r_1, & n_2 &= r_2, & \dots, & n_p &= r_p, & n_{p+1} &= r_{p+1} + 2, \\ t &= t_1 + t_2 + \dots + t_{p+1} + \sum_{\substack{h \neq j \\ (h, j = 1, 2, \dots, p+1)}} (n_h - 2)(n_j - 2), \\ t_h &= \frac{(n_h - 1)(n_h - 2)}{2} - 1 & (h &= 1, 2, \dots, p), & t_{p+1} &= \frac{(n_{p+1} - 1)(n_{p+1} - 2)}{2}, \end{aligned}$$

⁹⁾ Vgl. A., S. 63.

weil die Anzahl der gemeinsamen Tangenten der Kurvenzüge C_h und C_j nach dem Satze IV $(n_h - 2)(n_j - 2)$ ist.

Aus den beiden Gleichungssystemen folgen die Gleichungen

$$r = n - 2 + 2p \quad \text{und} \quad t = \frac{(n-1)(n-2)}{2} - p,$$

wodurch der Satz I für die Kurve $C^{(p)}$ bewiesen ist.

8. Irreduzible und reduzible Kurven vom Maximal-Klassenindex ohne Wendetangente.

Wir bezeichnen mit C eine irreduzible oder reduzible Kurve n -ter Klasse vom Maximal-Klassenindex, die r Spitzen, t Doppeltangenten und das Geschlecht p hat. Die Kurve kann im Endlichen liegen oder sie läßt sich ins Endliche nicht projizieren.

Bilden die Punkte, aus denen $n - 2$ Tangenten an die Kurve C gehen, k in sich zusammenhängende, miteinander nicht zusammenhängende ebene Gebiete T_1, T_2, \dots, T_k mit dem Zusammenhange $p_1 + 1, p_2 + 1, \dots, p_k + 1$, so ist $p = p_1 + p_2 + \dots + p_k - k + 1$. Für eine irreduzible Kurve C ist $k = 1$.

Der Index i der Kurve C ist die niedrigste Anzahl der Punkte, in denen die Kurve von einer Geraden der Ebene getroffen werden kann. Für eine Kurve ohne Wendetangente ist die Ordnung und der Index eine gerade Zahl.

Der Satz I wird für eine Kurve C durch vollständige Induktion bewiesen. Es wird gezeigt, daß der Satz I auch für die Kurven C vom Index i gültig ist, wenn seine Gültigkeit für die Kurven C vom Index $i - 2$ nachgewiesen ist.

Es sei g eine Gerade, von der die Kurve C vom Index i in i Punkten geschnitten wird, wobei $i \geq 2$ ist. Es kann angenommen werden, daß die Gerade g , von der die Kurve C in i Punkten geschnitten wird, von der unendlichfernen Geraden der Ebene abweicht. Es gibt auf der Geraden g immer einen außerhalb der Kurve im Endlichen liegenden Punkt P , aus dem man die konkave Seite der Kurve erreichen kann (weil die Kurve von g in $i \geq 2$ Punkten geschnitten wird). Projiziert man nun eine durch den Punkt P hindurchgehende von g verschiedene Gerade e ins Unendliche und bezeichnet man mit C' bzw. g' die Projektion der Kurve C bzw. der Geraden g , so liegen die Schnittpunkte der Geraden g' mit der Kurve C' alle im Endlichen und man erreicht aus dem unendlichfernen Punkte der Geraden g' (der die Projektion des Punktes P ist) die konkave Seite der Kurve C' . Es kann also angenommen werden, daß alle i Schnittpunkte von g mit C im Endlichen liegen und daß man aus dem unendlichfernen Punkte der Geraden g die konkave Seite der Kurve erreichen kann, ohne die Kurve

zu überschreiten. Bezeichnen A_1, A_2, \dots, A_i diese i Punkte und ihre Aufeinanderfolge auf der Geraden g , so liegt die Strecke $A_1 A_2$ in einem Gebiete G der Ebene, aus dessen Punkten n Tangenten an die Kurve C gehen. Ist M ein im Gebiete G liegendes Oval, von dem die Kurve C in den Punkten A_1 und A_2 berührt wird, so ist die aus den Kurven C und M bestehende Kurve C_M $n + 2$ -ter Klasse und vom Maximal-Klassenindex. Die Kurve C_M bleibt auch dann $n + 2$ -ter Klasse und vom Maximal-Klassenindex, wenn sie durch in den Punkten A_1 und A_2 angewandte Deformationen δ und ε in eine Kurve C'_M überführt wird, die außerhalb der Punkte A_3, A_4, \dots, A_i keine anderen Punkte mit der Geraden g gemeinsam hat.

Nimmt man die Gültigkeit des Satzes I für die Kurven C vom Index $i - 2$ an, so gilt der Satz für die Kurve C'_M , weil ihr Index $i - 2$ ist. Bezeichnen n', r', t', p' die Klasse, die Anzahl der Spitzen, die Anzahl der Doppeltangenten bzw. das Geschlecht der Kurve C'_M , so bestehen die Gleichungen

$$r' = n' - 2 + 2p', \quad t' = \frac{(n' - 1)(n' - 2)}{2} - p',$$

$$n' = n + 2, \quad r' = r + 4, \quad t' = t + 2n - 2.$$

Die Kurve C_M hat nämlich (nach Satz IV) $2n$ gemeinsame Tangenten mit dem Ovale M und deshalb $2n + t$ Doppeltangenten. Diese Anzahl wird durch die in den Punkten A_1 und A_2 durchgeführten Deformationen δ und ε um zwei vermindert.

Aus den vorigen Gleichungen ergeben sich die folgenden:

$$r = n - 2 + 2(p' - 1) \quad \text{und} \quad t = \frac{(n - 1)(n - 2)}{2} - (p' - 1).$$

Wir werden jetzt zeigen, daß $p' = p + 1$ ist.

Wir bezeichnen mit T bzw. T' die Gebiete, aus deren Punkten $n - 2$ bzw. $n' - 2$ Tangenten an die Kurve C bzw. C'_M gehen. Auf Grund der Konstruktion der Kurve C'_M aus der Kurve C entsteht das Gebiet T' aus dem Gebiete T , wenn man die Punkte A_1 und A_2 der Grenze von T durch eine Brücke, d. h. durch eine außerhalb von T liegende Elementarfläche, verbindet. Die Punkte A_1 und A_2 liegen beide an der Grenze eines Gebietes T_h , oder liegen auf den Grenzen zweier der Gebiete T_1, T_2, \dots, T_k . Das Gebiet T entsteht aus dem Gebiete T' , wenn man die Brücke entfernt. Das Entfernen der Brücke ist mit der Anwendung eines Querschnittes topologisch äquivalent, von dem die Punkte A_1 und A_2 der Brücke voneinander getrennt werden.

Verbindet die Brücke zwei Punkte der Grenze des Gebietes T_1 , so vergrößert sie den Zusammenhang des Gebietes um eins, sie läßt aber die

Zusammenhänge der übrigen $k - 1$ Gebiete T_2, T_3, \dots, T_k unverändert. Daraus folgt die Gleichung $p' = (p_1 + 1) + p_2 + p_3 + \dots + p_k - k + 1 = p + 1$.

Werden die Grenzen der Gebiete T_1 und T_2 durch die Brücke miteinander verbunden, so besteht für das Geschlecht p_{12} des aus T_1 und T_2 durch die Brücke entstandenen Gebietes die Gleichung $p_{12} + 1 = (p_1 + 1) + (p_2 + 1) - 1$, d. h. die Gleichung $p_{12} = p_1 + p_2$. Die Zusammenhänge $p_1 + p_2, p_3, p_4, \dots, p_k$ bleiben also durch die Brücke unverändert, die Anzahl der Gebiete wird aber um eins vermindert, so daß auch in diesem Falle $p' = p + 1$ ist.

Damit ist bewiesen, daß der Satz I auch für die Kurven C vom Index i gültig ist, sobald er für die Kurven vom Maximal-Klassenindex und vom Index $i - 2$ gilt. Der Satz I ist also für die Kurven vom Maximal-Klassenindex ohne Wendetangente vollständig bewiesen, sobald seine Gültigkeit für die Kurven vom Maximal-Klassenindex und vom Index Null nachgewiesen ist. Jede solche Kurve läßt sich ins Endliche projizieren. Denn es existiert eine zur Kurve fremde Gerade und diese kann durch eine zentrale Projektion ins unendlich Ferne geworfen werden, wobei die (zur Geraden fremde) Kurve ins Endliche projiziert wird.

Bezeichnet C_0 eine im Endlichen liegende reduzible Kurve vom Maximal-Klassenindex, die in k irreduzible Kurven zerfällt, so werden wir beweisen, daß der Satz I auch für die Kurve C_0 gültig ist, sobald er für die in $k - 1$ irreduzible Kurven zerfallenden und im Endlichen liegenden Kurven vom Maximal-Klassenindex gilt.

Die Punkte der Ebene, aus denen $n - 2$ Tangenten an die Kurve C_0 n -ter Klasse gehen, bilden k zusammenhängende, miteinander nicht zusammenhängende Gebiete T_1, T_2, \dots, T_k . Die k irreduziblen Kurven, in welche die Kurve C_0 zerfällt, bestehen aus den Randkurven der Gebiete T_1, T_2, \dots, T_k .

Ist P_1 bzw. P_2 ein innerer Punkt des Gebietes T_1 bzw. T_2 und bedeuten A_1, A_2, \dots, A_m die aufeinanderfolgenden Schnittpunkte der endlichen Strecke $P_1 P_2$ mit der Kurve C_0 , so liegen die Strecken $A_1 A_2, A_3 A_4, \dots, A_{m-1} A_m$ außerhalb der Gebiete T_h ($h = 1, 2, \dots, k$). Der Punkt A_1 bzw. A_m liegt auf dem Rande des Gebietes T_1 bzw. T_2 . Es gibt also eine solche Strecke $A_j A_{j+1}$, deren Endpunkte auf Ränder verschiedener Gebiete T_{h_1} und T_{h_2} fallen.

Ist M ein im Endlichen außerhalb der Gebiete T_h liegendes Oval von dem die Kurve C_0 in den Punkten A_j und A_{j+1} berührt wird, so gehen die Kurven C_0 und M durch Deformationen δ und ε in den Punkten A_j und A_{j+1} in eine Kurve C'_0 vom Maximal-Klassenindex über, die aus $k - 1$ irreduziblen Kurven besteht. Die Anwendung dieser De-

formationen ist mit der Anwendung einer Brücke äquivalent, von der die Gebiete T_{h_1} und T_{h_2} verbunden werden.

Die Kurve C_0 eine spezielle Kurve C . Die Kurve C'_0 entsteht aus C_0 so, wie die Kurve C' aus C entstanden ist. Wir haben schon bewiesen, daß der Satz I auch für die Kurve C gilt, wenn er für die Kurve C' besteht. Daraus folgt, daß der Satz I auch für die Kurve C_0 gültig bleibt, wenn er für die Kurve C'_0 gültig ist. Besteht also der Satz I für jede Kurve vom Maximal-Klassenindex, die in $k - 1$ im Endlichen liegende irreduzible Kurven zerfällt, so gilt er für jede reduzible Kurve vom Maximal-Klassenindex, die in k irreduzible Kurven zerfällt. Der Satz I wurde aber im vorigen Paragraphen für den Fall irreduzibler im Endlichen liegender Kurven vom Maximal-Klassenindex, d. h. für den Fall C'_0 und $k - 1 = 1$ bewiesen. Der Satz I gilt also erstens für jede Kurve vom Maximal-Klassenindex und vom Index Null und zweitens für jede Kurve C vom Maximal-Klassenindex und vom geraden Index.

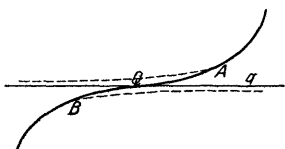
9. Kurven vom Maximal-Klassenindex mit einer Wendetangente.

Eine Kurve \bar{C} vom Maximal-Klassenindex kann höchstens eine Wendetangente q haben, von der die Kurve außerhalb des Wendepunktes Q nicht geschnitten wird¹⁰⁾. Würde nämlich die Kurve von ihrer Wendetangente q außerhalb von Q berührt, so könnte man sie durch passende Deformationen α , von denen die charakteristischen Zahlen der Kurven nicht verändert würden, in eine Kurve \bar{C}' überführen, die von der Wendetangente q nicht getroffen wird.

Man kann einen genug kleinen, den Wendepunkt Q in sich enthaltenden Bogen der Kurve \bar{C} durch einen in der Nähe der Geraden q liegenden hyperbolischen (ins Unendliche sich erstreckenden) Elementarbogen mit den Endpunkten A und B so ersetzen, daß die Punkte A und B gewöhnliche Punkte, die Gerade q eine Assymptote für die aus \bar{C} so erhaltene Kurve C' sind, ferner daß die Klasse, der Klassenindex und die Anzahl der Spitzen für die Kurven \bar{C} und C' übereinstimmen (Fig. 2). Statt der Wendetangente der Kurve \bar{C} hat die Kurve C' um eine Doppeltangente mehr als die Kurve \bar{C} . Es gibt in der Nähe der Geraden q eine Gerade q_1 , von der die Kurve C' nicht getroffen wird. Die Kurve C' läßt sich also ins Endliche projizieren.

Man kann die Möglichkeit der Deformation der Kurve \bar{C} in die Kurve C' an den Dualen Γ und Γ' der Kurven \bar{C} und C' einfacher ein-

Fig. 2.



¹⁰⁾ Vgl. A., S. 48.

sehen. Die Kurve Γ' entsteht aus der Kurve Γ vom Maximalindex, wenn ein genügend kleiner Bogen AB der Kurve Γ , von dem die einzige Spitze S der Kurve enthalten wird, durch eine genügend kleine Schleife so ersetzt wird, daß die Punkte A und B auch für die entstandene Kurve Γ' gewöhnliche Punkte sind. Die Schleife ist eine offene Kurve, die außerhalb eines Doppelpunktes keine anderen Punkt- oder Geradensingularitäten hat.

Aus der Spitze S geht keine Tangente an die Kurve Γ außer der Spitzentangente. Im entgegengesetzten Falle gäbe es nämlich in der Nähe der Tangente t der Kurve Γ aus dem Punkte S eine Gerade, die mit der Kurve wenigstens um vier weniger gemeinsame Punkte hätte, als die Gerade t . Die Kurve Γ wäre also nicht vom Maximalindex.

Schreibt man also um den Punkt S einen Kreis K , von dem ein einziger Bogen γ aus der Kurve Γ ausgeschnitten wird, und hat der Bogen γ außer der Spitze S keine andere Punkt- oder Tangentensingularität, so gibt es einen konzentrischen, den Bogen γ in zwei Punkten A_1 und B_1 schneidenden kleineren Kreis K_1 , der von keiner Tangente, von der die Kurve Γ außerhalb des Kreises K berührt wird, geschnitten wird. Wenn man den im Kreise K_1 liegenden Teilbogen $A_1 B_1$ von γ durch eine im Kreise K_1 liegende Schleife so ersetzt, daß durch diese Ersetzung auch der Bogen γ in eine Schleife überführt wird, so hat die aus Γ so erhaltene Kurve Γ' keine Doppeltangente, weil auch die Kurve Γ keine Doppeltangente hat. Die Kurve Γ' ist also vom Maximalindex. Jede Tangente der Kurve Γ , von der sie außerhalb des Bogens γ berührt wird, berührt auch die Kurve Γ' und trifft beide Kurven in denselben Punkten. Daraus folgt, daß die Ordnungen der Kurven Γ und Γ' übereinstimmen. Die Anzahl der singulären Punkte bzw. die Anzahl der Wendetangenten sind für beide Kurven dieselben.

Auf Grund dessen stimmen also die Werte von n , r und $t + w$, d. h. die Klasse, die Anzahl der Spitzen bzw. der singulären Tangenten für die Kurven \bar{C} und C' überein. Wird also gezeigt, daß auch das Geschlecht für beide Kurven dasselbe ist, so ist der Satz I auch für die Kurve \bar{C} und auf Grund der vorigen für alle Kurven vom Maximal-Klassenindex bewiesen.

Bezeichnet η die Deformation, welche die Kurve \bar{C} in die Kurve C' überführt, so deformiert η nur denjenigen Zug Z der Kurve \bar{C} in einen Zug Z' , der den Wendepunkt Q enthält. Die Deformation η führt die Gebiete T_0 und T'_0 , aus deren Punkten die minimale Anzahl der Tangenten an den Zug Z bzw. Z' gehen, ineinander über. Sind T_1, T_2, \dots, T_k die zur Kurve \bar{C} gehörigen Gebiete, aus deren Punkten die minimale Anzahl der Tangenten an die Kurve \bar{C} gehen, so wird nur dasjenige dieser Gebiete durch die Deformation η deformiert, für welches der Zug Z ein

Rand ist. Daraus folgt, daß die Geschlechter der Kurven \bar{C} und C' dieselben sind, wenn die Geschlechter der Züge Z und Z' übereinstimmen.

Aus den Punkten der Geraden q ist die konvexe Seite des Zuges Z' erreichbar. Das außerhalb von T'_0 liegende Gebiet T'_1 der Ebene kann also ganze Geraden enthalten. Das Gebiet T'_1 ist also zweifach und das Gebiet T'_0 einfach zusammenhängend. Daraus folgt, daß die Kurve Z' vom Geschlecht null ist. Auch der Zusammenhang des Gebietes T_0 ist eins und deshalb ist auch die Kurve Z vom Geschlecht null. Die projektive Ebene wird nämlich von dem einufrigen Rückerschnitte Z nicht zerteilt. Sie wird von Z in ein einfach zusammenhängendes Gebiet verwandelt. Dieses einfach zusammenhängende Gebiet wird vom Querschnitte längs der Wendetangente q in zwei einfach zusammenhängende Gebiete zerteilt. Das eine dieser zwei Gebiete ist T_0 . Die Gebiete T_0 und T'_0 sind also einfach zusammenhängend, die Kurven Z und Z' sind vom Geschlechte null. Damit ist der Satz I für jede Kurve vom Maximal-Klassenindex bewiesen.

9. Algebraische Kurven vom Maximal-Klassenindex.

Wir wollen jetzt den Satz II beweisen.

Bezeichnet p bzw. p^* das Realitäts-Geschlecht bzw. das algebraische Geschlecht der algebraischen Kurve vom Maximal-Klassenindex C^* , für welche die algebraische und Realitäts-Klasse dieselbe ist, so ist $p^* \leq p$, weil die Kurve C^* auch nichtreelle Tangentensingularitäten haben kann. Die Anzahl der nichtreellen Tangentensingularitäten ist wenigstens 2, also besteht auch die Ungleichung $p^* \leq p - 2$, wenn $p^* < p$ ist.

Die Kurve C hat p oder $p + 1$ Züge, sie kann aber nach dem bekannten Harnackschen Satze höchstens $p^* + 1$ Züge haben. Es folgt also, daß $p^* \geq p - 1$ ist. Die Ungleichungen $p^* \leq p - 2$ und $p^* \geq p - 1$ können nicht zugleich bestehen. Damit ist es bewiesen, daß $p^* = p$ ist und die Kurve C keine nichtreelle Tangentensingularität hat.

Aus dem Satz II folgt der Satz III.

Über die Züge der ebenen Kurven vom Maximal-Klassenindex.

Von

Julius v. Sz. Nagy in Klausenburg.

1. Einleitung.

Unter einer *Kurve* verstehen wir eine völlig stetige und geschlossene ebene Kurve. Die Klasse bzw. der Klassenindex der Kurve ist die größte bzw. die kleinste Anzahl der Tangenten der Kurve, die von einem beliebigen Punkte der Ebene ausgehen.

Bilden die Punkte der Ebene, aus denen $n - 2$ Tangenten an die Kurve C n -ter Klasse vom Maximal-Klassenindex gehen, k zusammenhängende Gebiete T_1, T_2, \dots, T_k mit den Zusammenhängen $p_1 + 1, p_2 + 1, \dots, p_k + 1$, so definiert man das Geschlecht p der Kurve C durch die Gleichung

$$p = p_1 + p_2 + \dots + p_k - k + 1.$$

Ist $k = 1$, so ist die Kurve C irreduzibel; ist aber $k > 1$, so zerfällt sie in k irreduzible Kurven C_1, C_2, \dots, C_k . Die Kurve C_h besteht aus den Zügen der Kurve C , von denen das Gebiet T_h begrenzt wird¹⁾.

Bezeichnet t die Anzahl der singulären Tangenten der Kurve C (welche mit der Ausnahme höchstens einer Wendetangente alle Doppeltangenten sind), so besteht die Gleichung

$$p = \frac{(n-1)(n-2)}{2} - t.$$

Das Geschlecht p läßt sich auch durch diese Gleichung definieren²⁾.

Eine einzügige Kurve vom Maximal-Klassenindex ist nach der ersten

¹⁾ J. v. Sz. Nagy, „Über Kurven vom Maximal-Klassenindex. Über Kurven vom Maximalindex“, Math. Annalen 89 (1923), S. 32–75; 90 (1924), S. 152–153. Vgl. § 21, Math. Annalen 89, S. 69–71.

²⁾ J. v. Sz. Nagy, „Über die charakteristischen Zahlen einer Kurve vom Maximal-Klassenindex“, Math. Annalen 100, S. 164.

Definition des Geschlechtes vom Geschlechte 0 oder 1. In dieser Arbeit werden die folgenden wichtigeren Sätze bewiesen.

Eine in k irreduzible Kurven zerfallende Kurve vom Maximal-Klassenindex und vom Geschlechte p hat $p + k - 1$ Züge vom Geschlechte 1 und k oder $k - 1$ Züge vom Geschlechte 0.

Bilden die Kurven C_1 und C_2 vom Klassenindex i_1 bzw. i_2 eine irreduzible Kurve vom Maximal-Klassenindex, so haben sie $i_1 i_2$ gemeinsame Tangenten.

Bilden die Kurven C_1 n_1 -ter und C_2 n_2 -ter Klasse vom Maximal-Klassenindex eine Kurve $n_1 + n_2$ -ter Klasse vom Maximal-Klassenindex, so haben sie $n_1 \cdot n_2$ gemeinsame Tangenten.

2. Über die Geschlechter der Züge der irreduziblen Kurven vom Maximal-Klassenindex.

Für die irreduziblen Kurven vom Maximal-Klassenindex gilt der Satz:

I. *Jede irreduzible Kurve C vom Maximal-Klassenindex und vom Geschlechte p besteht aus p Zügen (vom Maximal-Klassenindex und) vom Geschlechte 1 und noch aus einem oder keinem Zuge vom Geschlechte 0.*

Das ebene Gebiet T , aus dessen Punkten die minimale Anzahl der Tangenten an die Kurve C geht, ist $p + 1$ -fach zusammenhängend. Dieses Gebiet hat also entweder p oder $p + 1$ Ränder.

Jeder Zug der Kurve ist ein Rand des Gebietes T (ausgenommen nur den eventuellen Zug mit Wendetangente, der mit seiner Wendetangente zusammen einen Rand bildet). Bewegt sich nämlich ein Punkt in der Ebene, so kann sich die Anzahl der Tangenten, die an die Kurve gehen, nur dann verändern, wenn der Punkt die Kurve (oder ihre Wendetangente) überschreitet. Überschreitet der Punkt einen Zug von der konvexen Seite, so nimmt die Anzahl der Tangenten um zwei ab. Die Punkte der konkaven Seite von jedem Zuge gehören also dem Gebiete T an. Hat ein Zug eine Wendetangente, so ist er eine unpaare Kurve, weil er nur eine Wendetangente hat. Die Ebene wird von dem Zuge mit Wendetangente in ein einfach zusammenhängendes Gebiet verwandelt. Dieses Gebiet wird von der Wendetangente in zwei einfach zusammenhängende Gebiete zerteilt. Ein einfach zusammenhängendes Gebiet hat nur einen Rand. Daraus folgt, daß der betreffende Zug mit seiner Wendetangente zusammen einen Rand bildet. In unserem Falle hat also die Kurve C entweder p oder $p + 1$ Ränder.

Jeder Zug einer Kurve vom Maximal-Klassenindex ist wieder vom Maximal-Klassenindex. Der Satz ist also bewiesen, wenn wir zeigen, daß die Kurve C höchstens einen Zug vom Geschlechte 0 hat und weder $p + 1$ noch $p - 1$ Züge vom Geschlechte 1 haben kann.

Hat die Kurve C zwei Züge C_1 und C_2 vom Geschlechte 0, so liegen die einfach zusammenhängenden Gebiete T_1 und T_2 , aus deren Punkten die minimale Anzahl der Tangenten an den Zug C_1 bzw. C_2 geht, entweder außerhalb voneinander, oder es liegt das eine Gebiet im Innern des anderen. (Ein dritter Fall kann nicht vorkommen, weil eine Kurve vom Maximal-Klassenindex keinen Doppelpunkt hat und weil sie von der eventuellen einzigen Wendetangente außerhalb des Wendepunktes nicht geschnitten wird). In beiden möglichen Fällen liegen die Punkte der konkaven Seite der Kurven C_1 und C_2 in zwei durch einen Zug der Kurve C voneinander getrennten Gebieten. Zwei Punkte, welche auf der konkaven Seite der Kurven C_1 bzw. C_2 und in genügender Nähe dieser Züge liegen, liegen auch im Gebiete T . Das Gebiet T ist also nicht zusammenhängend und die Kurve C ist — gegen unsere Annahme — reduzibel. Aus diesem Widerspruch folgt, daß die Kurve C höchstens einen Zug vom Geschlechte 0 haben kann.

Das Gebiet G_h , aus dessen Punkten an den Zug C_h vom Geschlechte 1 die maximale Anzahl der Tangenten gezogen werden kann, ist einfach zusammenhängend. Die Punkte auf der konkaven Seite des Zuges C_h liegen außerhalb des Gebietes G_h . Das zusammenhängende Gebiet T hat Punkte außerhalb von G_h , es kann also keinen Punkt im Gebiet G_h haben. Daraus folgt, daß kein Zug von C im Gebiete G_h liegen kann. Hätte die Kurve C $p + 1$ Züge vom Geschlechte 1, so könnte man das Gebiet T aus der projektiven Ebene durch Herausschneiden $p + 1$ einfach zusammenhängender Gebiete herstellen. Das so erhaltene Gebiet ist $p + 2$ -fach zusammenhängend und deshalb wäre die Kurve vom Geschlechte $p + 1$. Hätte die Kurve einen Zug C_0 vom Geschlechte 0 und $p - 1$ Züge vom Geschlechte 1, so könnte man das Gebiet T aus dem einfach zusammenhängenden Gebiete T_0 , aus dessen Punkten an den Zug C_0 die minimale Anzahl der Tangenten gezogen werden kann, durch Herausheben $p - 1$ einfach zusammenhängender Gebiete herstellen. Das Gebiet T wäre dann p -fach zusammenhängend und die Kurve C wäre vom Geschlechte $p - 1$. Damit ist der Satz I bewiesen.

Aus dem Satze I folgt der Satz:

II. *Das Geschlecht einer irreduziblen Kurve vom Maximal-Klassenindex ist gleich der Summe der Geschlechter ihrer Züge.*

3. Über die Geschlechter der Züge der reduziblen Kurven vom Maximal-Klassenindex.

Für reduzible Kurven vom Maximal-Klassenindex gilt der Satz:

III. *Eine in k irreduzible Kurven zerfallende Kurve vom Maximal-Klassenindex hat entweder $k - 1$ oder k Züge vom Geschlecht 0.*

Für den Beweis dieses Satzes zeigen wir erst, daß eine Kurve C vom Maximal-Klassenindex mit den Zügen C_1, C_2, \dots, C_k , welche alle vom Geschlechte 1 sind, nicht reduzibel sein kann.

Jeder Zug C_h vom Geschlechte 1 bestimmt ein einfach zusammenhängendes Gebiet G_h , aus dessen Punkten an den Zug die maximale Anzahl der Tangenten gezogen werden kann. Tritt man aus dem Gebiet G_h heraus, so vermindert sich die Anzahl der Tangenten, die an den Zug C_h (und damit an die Kurve C) gezogen werden können, um zwei. Daraus folgt, daß kein Gebiet G_j im Innern eines anderen Gebietes G_h ($h \neq j$) liegen kann. Widrigenfalls könnte man aus zwei Gebieten von G_1, G_2, \dots, G_k nacheinander austreten, ohne inzwischen in eines der Gebiete G_1, G_2, \dots, G_k einzutreten. Die Kurve C wäre also nicht vom Maximal-Klassenindex. Da die Gebiete G_j außerhalb voneinander liegen, ist das Gebiet T zusammenhängend und mithin die Kurve C irreduzibel.

Damit ist bewiesen, daß eine reduzible Kurve vom Maximal-Klassenindex wenigstens einen Zug vom Geschlechte 0 hat. Zerfällt die reduzible Kurve C n -ter Klasse vom Maximal-Klassenindex in die k irreduziblen Kurven C_1, C_2, \dots, C_k n_1 -ter, n_2 -ter, \dots , n_k -ter Klasse, so bilden je zwei dieser Kurven wieder eine reduzible Kurve vom Maximal-Klassenindex.

Die Kurven C_h und C_j bilden eine reduzible Kurve C_{hj} $n_h + n_j$ -ter Klasse. Die Kurve C_{hj} ist vom Maximal-Klassenindex. Widrigenfalls gäbe es einen Punkt in der Ebene, aus dem man an die Kurve C_{hj} $n_h + n_j - 4 - 2q$ ($q \geq 0$), an die übrigen $k - 2$ Kurven C_m ($m = 1, 2, \dots, k$; $m \neq h, m \neq j$) höchstens n_m , an die Kurve C höchstens $n_1 + n_2 + \dots + n_k - 4 - 2q = n - 4 - 2q$ Tangenten ziehen könnte. Die Kurve C wäre also nicht vom Maximal-Klassenindex.

Daraus folgt der Satz III, denn je zwei von den Kurven C_1, C_2, \dots, C_k haben wenigstens einen Zug vom Geschlechte 0. Die Kurve C hat also mindestens $k - 1$ Züge vom Geschlechte 0, sie hat aber nach dem Satze I höchstens k Züge vom Geschlechte 0.

Aus den Sätzen I, II, III und aus der Definition des Geschlechtes folgen die Sätze:

IV. *Hat eine Kurve vom Maximal-Klassenindex k Züge vom Geschlechte 0, so zerfällt sie entweder in k oder in $k + 1$ irreduzible Kurven.*

V. *Jede in k irreduzible Kurven zerfallende Kurve vom Maximal-Klassenindex und vom Geschlechte p besteht aus $p + k - 1$ Zügen vom Geschlechte 1 und entweder aus k oder $k - 1$ Zügen vom Geschlechte 0.*

VI. *Ist k die Differenz zwischen der Anzahl der Züge vom Geschlechte 1 und dem Geschlechte p einer Kurve vom Maximal-Klassenindex, so zerfällt die Kurve in $k + 1$ irreduzible Kurven.*

VII. Ist die Summe der Geschlechter der Züge von einer Kurve vom Maximal-Klassenindex größer als das Geschlecht der Kurve, so zerfällt die Kurve gewiß.

VIII. Das Geschlecht einer aus m_1 Zügen vom Geschlechte 1 und aus m_0 Zügen vom Geschlechte 0 bestehenden Kurve vom Maximal-Klassenindex ist entweder $m_1 - m_0$ oder $m_1 - m_0 + 1$.

4. Die Anzahl der gemeinsamen Tangenten der Züge von einer irreduziblen Kurve vom Maximal-Klassenindex.

Ist C eine aus den r Zügen C_1, C_2, \dots, C_r bestehende irreduzible Kurve vom Maximal-Klassenindex, so ist das zugehörige Gebiet T zusammenhängend. Aus einem beliebigen Punkte der Ebene geht die maximale Anzahl der Tangenten höchstens an einen Zug der irreduziblen Kurve C . Dasselbe gilt auch für die Kurve C_{hj} , die aus den Zügen C_h und C_j der Kurve C besteht, woraus folgt, daß die Kurve C_{hj} irreduzibel und vom Maximal-Klassenindex ist. Bedeuten i, i_h, i_{hj} bzw. p, p_h, p_{hj} die Klassenindizes bzw. die Geschlechter der Kurven C, C_h, C_{hj} ($h, j = 1, 2, \dots, r; h \neq j$), so ist die Kurve C_{hj} von $n_{hj} = i_h + i_j + 2$ -ter Klasse und vom Geschlechte $p_{hj} = p_h + p_j$ (Satz II).

Sind t_{hj}, t_h bzw. t_j die Anzahlen der singulären Tangenten für die Kurven C_{hj}, C_h bzw. C_j und ist \bar{t}_{hj} die Anzahl der gemeinsamen Tangenten der Züge C_h und C_j , so bestehen (auf Grund der Definition des Geschlechtes einer Kurve vom Maximal-Klassenindex mit gegebener Klasse und gegebener Anzahl der singulären Tangenten) die folgenden Gleichungen:

$$t_{hj} = \frac{(n_{hj} - 1)(n_{hj} - 2)}{2} - p_{hj} = \frac{(i_h + i_j + 1)(i_h + i_j)}{2} - p_h - p_j,$$

$$t_h = \frac{(i_h + 1) \cdot i_h}{2} - p_h, \quad t_j = \frac{(i_j + 1) \cdot i_j}{2} - p_j, \quad \bar{t}_{hj} = t_h + t_j + \bar{t}_{hj}.$$

Aus diesen Gleichungen ergibt sich:

$$\bar{t}_{hj} = t_{hj} - t_h - t_j = \frac{(i_h + i_j + 1)(i_h + i_j)}{2} - \frac{(i_h + 1) i_h}{2} - \frac{(i_j + 1) i_j}{2} = i_h \cdot i_j.$$

Daraus folgt der Satz:

IX. Zwei Züge einer irreduziblen Kurve vom Maximal-Klassenindex haben $i_h \cdot i_j$ gemeinsame Tangenten, wenn i_h und i_j die Klassenindizes der zwei Züge sind.

Der Zug C_1 hat also mit den übrigen $r - 1$ Zügen der Kurve

$$i_1 i_2 + i_1 i_3 + \dots + i_1 i_r = i_1 (i_2 + i_3 + \dots + i_r) = i_1 (i - i_1)$$

gemeinsame Tangenten. Es gilt also der Satz:

X. Ein Zug vom Klassenindex i_h der irreduziblen Kurve C vom Maximal-Klassenindex mit dem Klassenindex i hat mit den übrigen Zügen der Kurve C $i_h(i - i_h)$ gemeinsame Tangenten.

Man kann aus dem Satze IX auch den folgenden Satz ableiten:

XI. Bilden die Kurven C' und C'' vom Klassenindex i' bzw. i'' zusammen eine irreduzible Kurve vom Maximal-Klassenindex, so haben sie $i' \cdot i''$ gemeinsame Tangenten.

Die Kurven C' und C'' bilden eine irreduzible Kurve vom Maximal-Klassenindex, daraus folgt, daß auch die Kurven C' und C'' irreduzible Kurven vom Maximal-Klassenindex sind. Besteht die Kurve C' bzw. C'' aus den Zügen C'_1, C'_2, \dots, C'_r bzw. aus den Zügen $C''_1, C''_2, \dots, C''_s$, deren Klassenindizes i'_1, i'_2, \dots, i'_r bzw. $i''_1, i''_2, \dots, i''_s$ sind, so bestehen die Gleichungen

$$i' = i'_1 + i'_2 + \dots + i'_r \quad \text{und} \quad i'' = i''_1 + i''_2 + \dots + i''_s.$$

Der Zug C'_h hat mit der Kurve C''

$$i'_h i''_1 + i'_h i''_2 + \dots + i'_h i''_s = i'_h \cdot i''$$

gemeinsame Tangenten. Die Kurve C'' hat also mit der Kurve C'

$$i'' i'_1 + i'' i'_2 + \dots + i'' i'_r = i' \cdot i''$$

gemeinsame Tangenten. Damit ist der Satz XI bewiesen.

5. Die Anzahl der gemeinsamen Tangenten der irreduziblen Kurven, in welche eine reduzible Kurve vom Maximal-Klassenindex zerfällt.

Zerfällt die Kurve C vom Maximal-Klassenindex C_1, C_2, \dots, C_k und bedeuten n, n_1, n_2, \dots, n_k bzw. p, p_1, p_2, \dots, p_k die Klassen bzw. die Geschlechter der Kurven C, C_1, C_2, \dots, C_k , so bestehen die Gleichungen

$$n = n_1 + n_2 + \dots + n_k \quad \text{und} \quad p = p_1 + p_2 + \dots + p_k - k + 1.$$

Die Kurven C_h und C_j bilden zusammen eine Kurve C_{hj} $n_{hj} = n_h + n_j$ -ter Klasse und vom Maximal-Klassenindex mit dem Geschlechte $p_{hj} = p_h + p_j - 1$, also bestehen auch die Gleichungen

$$t_{hj} = \frac{(n_{hj} - 1)(n_{hj} - 2)}{2} - p_{hj} = \frac{(n_h + n_j - 1)(n_h + n_j - 2)}{2} - (p_h + p_j - 1),$$

$$t_h = \frac{(n_h - 1)(n_h - 2)}{2} - p_h, \quad t_j = \frac{(n_j - 1)(n_j - 2)}{2} - p_j, \quad t_{hj} = t_h + t_j + \bar{t}_{hj},$$

wo t_{hj}, t_h bzw. t_j die Anzahl der singulären Tangenten der Kurve C_{hj}, C_h bzw. C_j bedeuten, \bar{t}_{hj} die Anzahl der gemeinsamen Tangenten der Kurven C_h und C_j ist

Aus den vorigen Gleichungen folgt die Gleichung

$$\bar{t}_{hj} = t_{hj} - t_h - t_j = n_h \cdot n_j.$$

Es gilt also der folgende Satz:

XII. Zerfällt die Kurve C n -ter Klasse vom Maximal-Klassenindex in die irreduziblen Kurven C_1, C_2, \dots, C_k von den Klassen n_1, n_2, \dots, n_k , so hat die Kurve C_h mit der Kurve C_j $n_h \cdot n_j$, mit den übrigen $k - 1$ Kurven C_l ($l = 1, 2, \dots, k; l \neq h$) $n_h(n - n_h)$ gemeinsame Tangenten.

Durch einen ähnlichen Beweisgang, wie bei dem Beweise des Satzes XI, läßt sich der folgende Satz ableiten:

XIII. Bilden die Kurven C' und C'' n' -ter bzw. n'' -ter Klasse vom Maximal-Klassenindex eine Kurve C $n' + n''$ -ter Klasse vom Maximal-Klassenindex, so haben die Kurven C' und C'' $n' \cdot n''$ gemeinsame Tangenten.

6. Die Anzahl der gemeinsamen Tangenten einer reduziblen Kurve vom Maximal-Klassenindex.

Für jede Kurve vom Maximal-Klassenindex gilt der Satz:

XIV. Sind C_1 und C_2 zwei Züge einer Kurve C vom Maximal-Klassenindex und gehen m_1 und m_2 Tangenten aus einem Punkte des Gebietes G , das von den zwei Zügen (und von der eventuellen einzigen Wendetangente des einen Zuges) begrenzt wird, an den Zug C_1 bzw. C_2 , so haben die Züge C_1 und C_2 $m_1 \cdot m_2$ gemeinsame Tangenten.

Wenn man aus einem Punkte P des Gebietes G die konkaven bzw. die konvexen Seiten beider Züge erreichen kann, so bilden die Züge C_1 und C_2 eine irreduzible bzw. reduzible Kurve vom Maximal-Klassenindex. Für diese zwei Fälle ist der Satz XIV mit den Sätzen IX und XIII schon bewiesen. Es bleibt nur der Fall übrig, wo man aus dem Punkte P die konkave Seite des einen und die konvexe des anderen Zuges erreichen kann.

Wenn man aus dem Punkte P die konvexe Seite des Zuges C_1 n_1 -ter Klasse und vom Klassenindex i_1 und die konkave Seite des Zuges C_2 n_2 -ter Klasse und vom Klassenindex i_2 erreichen kann, so können n_1 bzw. $n_2 - 2 = i_2$ Tangenten aus P an den Zug C_1 bzw. C_2 gezogen werden.

Aus einem Punkte der Ebene kann man entweder die konkaven oder die konvexen Seiten der Elementarbögen einer Kurve C vom Maximal-Klassenindex erreichen, ohne sie zu überschreiten. Es muß also im Gebiete G wenigstens einen Zug C_3 der Kurve C geben, von dem das Gebiet G in zwei Gebiete G_1 und G_2 so zerteilt wird, daß man aus einem Punkte des Gebietes G_1 bzw. G_2 die konvexen Seiten der Züge C_1 und C_2 bzw.

die konkaven Seiten der Züge C_2 und C_3 erreichen kann, ohne einen der Züge C_1 , C_2 und C_3 überschreiten zu müssen.

Ist n_3 bzw. i_3 die Klasse bzw. der Klassenindex der Kurve C_3 , so bilden die Züge C_2 und C_3 eine irreduzible Kurve C_{23} $n_{23} = i_2 + i_3 + 2 = i_2 + n_3$ -ter Klasse vom Maximal-Klassenindex. Die Kurve C_{23} und der Zug C_1 bilden eine reduzible Kurve $n_1 + n_{23}$ -ter Klasse vom Maximal-Klassenindex.

Die aus den Zügen $C_1 C_3$, $C_1 C_2$ bzw. $C_1 C_2 C_3$ bestehenden Kurven sind vom Maximal-Klassenindex, weil die projektive Ebene von jeder dieser Kurven (und von der eventuellen einzigen Wendetangente) in Gebiete so zerteilt wird, daß man aus einem Punkte der Ebene entweder die konkaven oder die konvexen Seiten jeder der drei Kurven erreichen kann, ohne sie zu überschreiten.

Bedeutet t_2, t_3 bzw. t_{23} die Anzahl der gemeinsamen Tangenten des Zuges C_1 mit der Kurve C_2, C_3 bzw. C_{23} , so bestehen die Gleichungen

$$t_{23} = n_1 \cdot n_{23} = n_1 (i_2 + n_3), \quad t_2 = n_1 \cdot n_3 \quad \text{und} \quad t_{23} = t_2 + t_3.$$

Aus diesen Gleichungen folgt die Gleichung $t_2 = n_1 i_2$. Damit ist der Satz XIV bewiesen.

Aus einem Punkte des Gebietes G gehen ebenso viele Tangenten an den Zug C_1 bzw. C_2 , wie aus einem Punkte des Zuges C_2 an C_1 bzw. aus einem Punkte von C_1 an C_2 gehen. Der Satz XIV läßt sich also auf folgende Weise ausdrücken:

XV. Sind C_1 und C_2 Züge einer Kurve vom Maximal-Klassenindex und gehen m_1 Tangenten aus einem Punkte des Zuges C_2 an C_1 und m_2 Tangenten aus einem Punkte des Zuges C_1 an C_2 , so haben die Züge C_1 und C_2 $m_1 \cdot m_2$ gemeinsame Tangenten.

Aus den Sätzen XIV und XV folgen die Sätze:

XVI. Sind C_1 und C_2 zwei Züge einer Kurve vom Maximal-Klassenindex, so haben sie entweder $(n_1 - 2)(n_2 - 2)$, oder $(n_1 - 2) \cdot n_2$, oder $n_1(n_2 - 2)$, oder $n_1 \cdot n_2$ gemeinsame Tangenten, wenn n_1 und n_2 die Klassen der Züge C_1 und C_2 bedeuten.

XVII. Ist die Anzahl der gemeinsamen Tangenten von zwei Zügen einer Kurve C vom Maximal-Klassenindex größer als das Produkt der Klassenindizes der Züge, so ist die Kurve C reduzibel.

7. Kurven vom Maximalindex.

Hat eine Kurve Γ n -ter Ordnung vom Maximalindex d singuläre Punkte (unter denen höchstens eine Spitze vorkommen kann), so kann man das Geschlecht der Kurve durch die Gleichung

$$p = \frac{(n-1)(n-2)}{2} - d$$

definieren. Die aus mehreren Zügen bestehende Kurve Γ zerfällt dann in die Kurven $\Gamma_1, \Gamma_2, \dots, \Gamma_k$ n_1 -, n_2 -, \dots , n_k -ter Ordnung, wenn

$$n = n_1 + n_2 + \dots + n_k$$

ist und die Kurven $\Gamma_1, \Gamma_2, \dots, \Gamma_k$ zusammen jeden Zug der Kurve Γ enthalten und zwar jeden Zug nur einmal. Wenn das Zerfallen der Kurve Γ bei keinerlei Einteilung der Züge eintritt, so ist die Kurve irreduzibel.

Auf Grund der Sätze I—XVII kann man für die Kurven vom Maximalindex die folgenden wichtigeren Sätze aussprechen:

XVIII. *Jede Kurve vom Maximalindex und vom Geschlechte p besteht aus p Zügen vom Geschlechte 1 und aus einem oder keinem Zuge vom Geschlechte 0.*

XIX. *Zerfällt eine Kurve vom Maximalindex und vom Geschlechte p in k irreduzible Kurven, so hat sie $p + k - 1$ Züge vom Geschlechte 1 und k oder $k - 1$ Züge vom Geschlechte 0.*

XX. *Bilden die Kurven Γ_1 und Γ_2 zusammen eine irreduzible Kurve vom Maximalindex, so ist die Anzahl ihrer Schnittpunkte gleich dem Produkt ihrer Indizes.*

XXI. *Bilden die Kurven Γ_1 n_1 -ter und Γ_2 n_2 -ter Ordnung vom Maximalindex eine Kurve $n_1 + n_2$ -ter Ordnung vom Maximalindex, so haben sie $n_1 \cdot n_2$ gemeinsame Punkte.*

XXII. *Sind Γ_1 und Γ_2 zwei Züge einer Kurve vom Maximalindex und wird der Zug Γ_1 bzw. Γ_2 von einer Tangente des anderen Zuges in m_1 bzw. m_2 Punkten geschnitten, so haben die zwei Züge $m_1 m_2$ gemeinsame Punkte.*

(Eingegangen am 12. 4. 1927.)

Zur Theorie der schlichten Abbildungen.

Von

G. Szegö in Königsberg.

Die vorliegende Arbeit enthält einige Bemerkungen über die Abschnitte von schlichten Potenzreihen.

Wir betrachten im folgenden drei Unterklassen der im Einheitskreis $|z| < 1$ regulären Funktionen

$$f(z) = c_0 + c_1 z + c_2 z^2 + \dots + c_n z^n + \dots$$

(S): Das Bild des Kreises $|z| < 1$ bei der Abbildung $w = f(z)$ ist *schlicht*.

(S₀): Das Bild des Kreises $|z| < 1$ bei der Abbildung $w = f(z)$ ist *schlicht* und in bezug auf den Punkt $w = c_0$ *sternförmig*¹⁾.

(E^{*}): $f(z)$ ist *schlicht* und *beschränkt*, und zwar $|f(z)| \leq 1$ für $|z| < 1$.

In Zusammenhang mit (S₀) wird auch die folgende Unterklasse erwähnt:

(K): Das Bild des Kreises $|z| < 1$ bei der Abbildung $w = f(z)$ ist *schlicht* und *konvex*.

In § 2 beweisen wir den folgenden Satz:

I. Die Funktion

$$f(z) = c_0 + c_1 z + c_2 z^2 + \dots + c_n z^n + \dots$$

sei regulär und schlicht im Kreise $|z| < 1$. Dann sind sämtliche Abschnitte

$$c_0 + c_1 z + c_2 z^2 + \dots + c_n z^n \quad (n = 1, 2, 3, \dots)$$

schlicht im Kreise $|z| < \frac{1}{4}$.

¹⁾ Dies bedeutet, daß mit jedem Punkte w auch die geradlinige Verbindungsstrecke $c_0 w$ zum Bildgebiet gehört.

Der Beweis dieses Satzes ist leicht für $n = 2$ und $n \geq 5$, jedoch nicht ganz einfach für $n = 3$ und $n = 4$. Im letzteren Falle benutze ich eine Bemerkung von Herrn W. Rogosinski. Die Konstante $\frac{1}{4}$ kann durch keine größere Zahl ersetzt werden, weil bei der Funktion

$$\frac{z}{(1-z)^2} = z + 2z^2 + 3z^3 + \dots + nz^n + \dots$$

der Abschnitt $z + 2z^2$ eine in $z = -\frac{1}{4}$ verschwindende Ableitung besitzt. Der Beweis liefert zugleich, daß die fraglichen Abschnitte für $n > 2$ sogar in einem (nur von n abhängigen) Kreise $|z| < r_n$, wo $r_n > \frac{1}{4}$ ist, schlicht bleiben. Die genaue Bestimmung des größten Kreises $|z| < r_n$ dieser Art dürfte nicht leicht sein.

Die Fragestellung, welche durch den obigen Satz erledigt wird, erinnert an eine analoge Frage bezüglich der Klasse der beschränkten Funktionen, bzw. der Funktionen mit positivem Realteil, welche von den Herren W. Rogosinski und L. Fejér behandelt worden ist²⁾.

Die entsprechende Aufgabe bei der Klasse (S_0) wird erledigt durch den in § 3 zu beweisenden Satz:

II. Die Funktion

$$f(z) = c_0 + c_1 z + c_2 z^2 + \dots + c_n z^n + \dots$$

sei regulär und schlicht im Kreise $|z| < 1$ und bilde denselben auf ein bezüglich $w = c_0$ sternförmiges Gebiet ab. Dann sind sämtliche Abschnitte

$$c_0 + c_1 z + c_2 z^2 + \dots + c_n z^n \quad (n = 1, 2, 3, \dots)$$

schlicht und „sternförmig“ bezüglich $w = c_0$ im Kreise $|z| < \frac{1}{4}$.

Die Konstante $\frac{1}{4}$ kann auch hier durch keine größere ersetzt werden. Der Beweis dieses Satzes ist leicht für $n = 2$ und $n \geq 4$, dagegen etwas umständlich für $n = 3$.

Mit den Funktionen der Klasse (E^*) hat sich Herr L. Fejér beschäftigt. Er zeigt, daß die Abschnitte derselben der Ungleichung

$$|c_0 + c_1 z + c_2 z^2 + \dots + c_n z^n| \leq 1 + \frac{1}{\sqrt{2}} = 1,7071 \dots \quad (|z| \leq 1)$$

genügen³⁾. Wir bezeichnen bei festem n mit σ_n die obere Grenze von

$$|c_0 + c_1 z + c_2 z^2 + \dots + c_n z^n|$$

²⁾ Vgl. die Literaturzusammenstellung bei L. Fejér, a. a. O. ³⁾, Fußnote ²⁾ auf S. 185.

³⁾ L. Fejér, Über die Koeffizientensumme einer beschränkten und schlichten Potenzreihe, Acta Mathematica 49 (1926), S. 183—190.

für $|z| \leq 1$ und für alle Funktionen von (E^*) ; ferner sei σ die obere Grenze von $\sigma_1, \sigma_2, \dots, \sigma_n, \dots$. Dann ist nach L. Fejér

$$\sigma_1 = 1,25 \leq \sigma \leq 1,7071 \dots$$

In § 4 zeigen wir, daß man die obere Schranke von σ etwas herabdrücken kann. Es gilt

$$\text{III.} \quad \sigma \leq 1,6160 \dots$$

Weiter führt in § 5 eine nicht ganz leichte Rechnung auf

$$\text{IV.} \quad \sigma_2 = \frac{22}{3} - 4\sqrt{\frac{7}{3}} = 1,2232 \dots$$

Es ist bemerkenswert, daß $\sigma_2 < \sigma_1$ ausfällt.

In § 2 wird zum Teil die K. Löwnersche Theorie der Schlitzabbildungen⁴⁾ benutzt.

§ 1.

Hilfssätze.

1. Hilfssatz I. *Es sei $f(z)$ eine Funktion der Klasse (S) . Dann gilt für $|z_1| < 1, |z_2| < 1$ die Ungleichung*

$$(1) \quad \left| \frac{f(z_1) - f(z_2)}{z_1 - z_2} \right| \geq |f'(z_2)| (1 - |z_2|^2) \frac{1 - \bar{z}_2 z_1}{(|z_1 - z_2| + |1 - \bar{z}_2 z_1|)^2}$$

Für den Spezialfall $z_2 = 0$ geht diese Ungleichung in

$$(2) \quad \left| \frac{f(z_1) - f(0)}{z_1} \right| \geq \frac{|f'(0)|}{(1 + |z_1|)^2}$$

über, was einen Teil des Verzerrungssatzes (in der scharfen Fassung) bildet. Auf diesen Spezialfall kann aber (1) durch eine Abbildung des Einheitskreises auf sich leicht zurückgeführt werden⁵⁾.

Wir setzen

$$\frac{z - z_2}{1 - \bar{z}_2 z} = w, \quad \frac{z_1 - z_2}{1 - \bar{z}_2 z_1} = w_1.$$

Es sei ferner $f(z) = g(w)$. Nach (2) gilt

$$(3) \quad \left| \frac{g(w_1) - g(0)}{w_1} \right| \geq \frac{|g'(0)|}{(1 + |w_1|)^2}$$

Nun ist $g(w_1) = f(z_1)$, $g(0) = f(z_2)$ und

$$g'(w) = f'(z) \frac{dz}{dw} = \frac{f'(z)(1 - \bar{z}_2 z)^2}{1 - |z_2|^2}$$

⁴⁾ K. Löwner, Untersuchungen über schlichte konforme Abbildungen des Einheitskreises I, *Math. Annalen* 89 (1923), S. 103—121.

⁵⁾ Den Hinweis auf diesen Kunstgriff verdanke ich Herrn K. Löwner.

folglich

$$g'(0) = f'(z_2)(1 - |z_2|^2),$$

woraus die Behauptung folgt⁶⁾.

Weiter ist nach dem Verzerrungssatze

$$|f'(z_2)| \geq \frac{1 - |z_2|}{(1 + |z_2|)^3} |f'(0)|,$$

so daß wir aus (1)

$$(1') \quad \left| \frac{f(z_1) - f(z_2)}{z_1 - z_2} \right| \geq |f'(0)| \left(\frac{1 - |z_2|}{1 + |z_2|} \right)^2 \frac{1 - \bar{z}_2 z_1}{(|z_1 - z_2| + |1 - \bar{z}_2 z_1|)^2} \\ (|z_1| < 1, |z_2| < 1)$$

erhalten. Es ergibt sich schließlich für $|z_1| < \frac{1}{4}$, $|z_2| < \frac{1}{4}$

$$(4) \quad \left| \frac{f(z_1) - f(z_2)}{z_1 - z_2} \right| > \left(\frac{1 - \frac{1}{4}}{1 + \frac{1}{4}} \right)^2 \frac{1 - \frac{1}{16}}{\left(\frac{1}{2} + 1 + \frac{1}{16} \right)^2} |f'(0)| = \frac{432}{3125} |f'(0)|.$$

2. Hilfssatz II. Es sei

$$f(z) = c_0 + c_1 z + c_2 z^2 + \dots + c_n z^n + \dots$$

eine Funktion der Klasse (S) und $c_1 = 1$. Dann ist für $n \geq 2$

$$(5) \quad |c_n| < en,$$

$$(5') \quad |c_n| < \frac{n^n}{(n-1)^{n-1}},$$

ferner

$$(6) \quad |c_5| < 8.$$

Die Abschätzungen (5) und (5') ((5) ist eine Folge von (5')) stammen von Herrn J. E. Littlewood⁷⁾. Um (6) zu beweisen, verfähre ich in

⁶⁾ Wendet man auf $g(w)$ anstatt (3) die in dem Verzerrungssatze enthaltene obere Abschätzung an, so ergibt sich das folgende Gegenstück zu (1): Es sei $f(z)$ eine Funktion der Klasse (S). Dann gilt für $|z_1| < 1$, $|z_2| < 1$

$$(1^*) \quad \left| \frac{f(z_1) - f(z_2)}{z_1 - z_2} \right| \leq |f'(z_2)| (1 - |z_2|^2) \frac{|1 - \bar{z}_2 z_1|}{(|z_1 - z_2| + |1 - \bar{z}_2 z_1|)^2}.$$

Herr W. Rogosinski hat mich auf die folgenden Ungleichungen aufmerksam gemacht, die aus (1) und (1^{*}) für $z_1 = 0$, $z_2 = z$ folgen:

$$\frac{1 - |z|}{1 + |z|} \leq \left| \frac{z f'(z)}{f(z) - f(0)} \right| \leq \frac{1 + |z|}{1 - |z|}.$$

Durch die obere Abschätzung wird u. a. eine Fragestellung von Herrn J. E. Littlewood (a. a. O. ⁷⁾, S. 508, Theorem 26) beantwortet.

⁷⁾ J. E. Littlewood, On inequalities in the theory of functions, Proc. of the London Math. Soc. (2) 23 (1924), S. 481—519 (Theorem 20).

Anlehnung an eine briefliche Bemerkung von Herrn K. Grandjot folgendermaßen. Für die Funktion

$$g(z) = (f(z^{-2}) - c_0)^{-\frac{1}{2}} = z + \frac{b_1}{z} + \frac{b_3}{z^3} + \dots$$

besagt der „Flächensatz“⁸⁾, daß

$$|b_1|^2 + 3|b_3|^2 + 5|b_5|^2 + \dots \leq 1$$

ist. Aus $f(z) - c_0 = (g(z^{-\frac{1}{2}}))^{-2}$ folgt aber

$$c_5 = -2b_7 + 6b_1b_5 + 3b_3^2 - 12b_1^2b_3 + 5b_1^4,$$

so daß mit $|b_n| = \beta_n$

$$|c_5| \leq 2\beta_7 + 6\beta_1\beta_5 + 3\beta_3^2 + 12\beta_1^2\beta_3 + 5\beta_1^4$$

gilt, wobei

$$\beta_1^2 + 3\beta_3^2 + 5\beta_5^2 + 7\beta_7^2 \leq 1.$$

Nun ist

$$6\beta_1\beta_5 \leq \frac{3}{\sqrt{5}}(\beta_1^2 + 5\beta_5^2) \leq \frac{3}{\sqrt{5}}(1 - 7\beta_7^2),$$

folglich

$$2\beta_7 + 6\beta_1\beta_5 \leq \frac{3}{\sqrt{5}} + 2\beta_7 - \frac{21}{\sqrt{5}}\beta_7^2.$$

Die quadratische Funktion rechts erreicht für $\beta_7 = \frac{\sqrt{5}}{21}$ ihr Maximum, so daß

$$(7) \quad 2\beta_7 + 6\beta_1\beta_5 \leq \frac{3}{\sqrt{5}} + \frac{2\sqrt{5}}{21} - \frac{\sqrt{5}}{21} = \frac{68}{21\sqrt{5}} < \frac{34}{21} < \frac{5}{3}.$$

Andererseits hat man

$$\begin{aligned} 3\beta_3^2 + 12\beta_1^2\beta_3 + 5\beta_1^4 &\leq 3\beta_3^2 + 12(1 - 3\beta_3^2)\beta_3 + 5(1 - 3\beta_3^2)^2 \\ &\leq 45\beta_3^4 - 36\beta_3^3 - 27\beta_3^2 + 12\beta_3 + 5 \\ &\leq \qquad\qquad\qquad -27\beta_3^2 + 12\beta_3 + 5, \end{aligned}$$

weil doch $45\beta_3^4 - 36\beta_3^3 = 45\beta_3^3\left(\beta_3 - \frac{4}{5}\right) \leq 0$ für $0 \leq \beta_3 \leq \frac{1}{3} < \frac{4}{5}$ ist.

Die quadratische Funktion auf der rechten Seite erreicht für $\beta_3 = \frac{2}{9}$ ihr Maximum, so daß

$$(8) \quad 3\beta_3^2 + 12\beta_1^2\beta_3 + 5\beta_1^4 \leq -\frac{4}{3} + \frac{8}{3} + 5 = \frac{19}{3}$$

und also wegen (7) und (8)

$$|c_5| < \frac{5}{3} + \frac{19}{3} = 8$$

ist.

⁸⁾ Vgl. L. Bieberbach, Lehrbuch der Funktionentheorie, Band II, S. 84–85. Leipzig: B. G. Teubner, 1927.

§ 2.

Über die Klasse (S).

Wir dürfen annehmen, daß $c_1 = 1$ ist.

1. Wir beweisen den Satz I zunächst für $n \geq 6$, indem wir zeigen, daß

$$(9) \quad \left| \frac{f(z_1) - f(z_2)}{z_1 - z_2} \right| > \left| \sum_{\nu=n+1}^{\infty} c_{\nu} \frac{z_1^{\nu} - z_2^{\nu}}{z_1 - z_2} \right| \quad \left(|z_1| < \frac{1}{4}, |z_2| < \frac{1}{4}; n \geq 6 \right)$$

ist.

Wegen der Ungleichungen (4) und (5) genügt es

$$(10) \quad e \sum_{\nu=n+1}^{\infty} \frac{\nu^2}{4^{\nu-1}} < \frac{432}{3125}$$

nachzuweisen. Gilt diese Ungleichung für $n = 6$, so gilt sie auch für $n \geq 6$. Nun ist

$$(11) \quad \sum_{\nu=m}^{\infty} \nu^2 x^{\nu} = x^m \frac{m^2(1-x)^2 + (2m+1)x(1-x) + 2x^2}{(1-x)^3} \quad (m = 0, 1, 2, \dots),$$

so daß

$$\sum_{\nu=n+1}^{\infty} \frac{\nu^2}{4^{\nu-1}} = \frac{1}{4^n} \frac{\frac{9}{16}(n+1)^2 + \frac{3}{16}(2n+3) + \frac{1}{8}}{\left(1 - \frac{1}{4}\right)^3} = \frac{9n^2 + 24n + 20}{27 \cdot 4^{n-1}}$$

ist. Wir haben also nur zu zeigen, daß

$$\frac{61}{27 \cdot 128} e < \frac{432}{3125}$$

besteht. Dies ist aber richtig, weil

$$\frac{61}{27 \cdot 128} e < \frac{64 \cdot 3}{27 \cdot 128} = \frac{1}{18} < \frac{432}{3125}.$$

2. Für $n = 5$ versagt die vorige Methode. Sie führt nämlich dann auf die Ungleichung

$$(12) \quad \frac{365}{27 \cdot 256} e < \frac{432}{3125},$$

was wegen

$$\frac{365}{27 \cdot 256} e > \frac{365 \cdot 2,7}{27 \cdot 256} = \frac{73}{512} = \frac{438}{3072} > \frac{432}{3125}$$

falsch ist.

Mit Rücksicht auf (5') genügt es aber anstatt (12)

$$\frac{365}{27 \cdot 256} e - \frac{6^2}{4^5} e + \frac{6 \cdot 6^6}{4^5 \cdot 5^5} < \frac{432}{3125},$$

oder

$$(12') \quad \frac{122}{27} e + 9 \cdot 1,2^5 < \frac{432 \cdot 256}{3125}$$

zu beweisen. Nun ist

$$\begin{aligned} \frac{122}{27} e + 9 \cdot 1,2^5 &< \frac{12,2 \cdot 27,2}{27} + 10,8 \cdot 1,44^2 < 12,2 + 0,1 + 10,8 \cdot 2,1 \\ &= 34,98 < 35 \end{aligned}$$

und

$$\frac{432 \cdot 256}{3125} > \frac{430 \cdot 255}{3125} = \frac{86 \cdot 51}{125} > \frac{86 \cdot 50}{125} + \frac{75}{125} = 34,4 + 0,6 = 35.$$

3. Für $n = 4$ führt die bisher befolgte Methode selbst im Besitze der (bis heute unbewiesenen) Ungleichungen

$$(5'') \quad |c_n| \leq n \quad (n = 4, 5, 6, \dots)$$

nicht zum Ziele. Diese liefern nämlich als obere Schranke für die rechte Seite von (9)

$$\left(\frac{9n^2 + 24n + 20}{27 \cdot 4^{n-1}} \right)_{n=4} = \frac{65}{27 \cdot 16} > \frac{4}{27} > \frac{432}{3125}.$$

Ich benutze daher eine Verschärfung der vorigen Methode, auf welche ich von Herrn W. Rogosinski aufmerksam gemacht worden bin. Es wird dabei die Abschätzung (6) eine Rolle spielen.

Es genügt offenbar die Ungleichung (9) für $z_1 = \frac{1}{4}$, $z_2 = \frac{x}{4}$ mit $|x| = 1$ zu beweisen, was wegen (1') eine Folge von

$$\begin{aligned} \left(\frac{1 - \frac{1}{4}}{1 + \frac{1}{4}} \right)^2 \frac{|1 - \frac{x}{16}|}{\left(\frac{|1-x|}{4} + |1 - \frac{x}{16}| \right)^2} &= \frac{9}{25} \frac{1}{\left| 1 - \frac{x}{16} \right| + \frac{1}{2} |1-x| + \frac{1}{16} \left| \frac{(1-x)^2}{1 - \frac{x}{16}} \right|} \\ &> \sum_{\nu=5}^{\infty} \frac{|c_{\nu}|}{4^{\nu-1}} \left| \frac{1-x^{\nu}}{1-x} \right| \end{aligned}$$

ist. Wir unterscheiden zwei Fälle:

a) $|1-x| \geq 0,4$. Man hat mit Rücksicht auf (6), (5) (vgl. (23))

$$\begin{aligned} \sum_{\nu=5}^{\infty} \frac{|c_{\nu}|}{4^{\nu-1}} \left| \frac{1-x^{\nu}}{1-x} \right| &\leq \frac{8}{|1-x|} \sum_{\nu=5}^{\infty} \frac{|c_{\nu}|}{4^{\nu}} < \frac{8}{|1-x|} \left(\frac{8}{4^5} + e \sum_{\nu=6}^{\infty} \frac{\nu}{4^{\nu}} \right) \\ &< \frac{8}{|1-x|} \left(\frac{8}{4^5} + \frac{19e}{9 \cdot 4^5} \right) = \frac{1}{|1-x|} \frac{72 + 19e}{9 \cdot 128} \\ &\leq \frac{1}{|1-x|} \frac{72 + 19(3 - 0,1)}{9 \cdot 128} \leq \frac{1}{9} \frac{1}{|1-x|}, \end{aligned}$$

so daß wir folgendes zu beweisen haben:

$$\left| \frac{1-x}{1-\frac{x}{16}} \right| + \frac{1}{2} + \frac{1}{16} \left| \frac{1-x}{1-\frac{x}{16}} \right| < \frac{81}{25} = 3,24.$$

Nun ist das erste Glied der linken Seite monoton abnehmend, wenn $\arcc x$ von 0 bis π läuft; es ist ferner $\left| 1 - \frac{x}{16} \right| < 1$ für $|1-x| = 0,4$, $|x| = 1$. Man hat also

$$\left| \frac{1-x}{1-\frac{x}{16}} \right| < \frac{1}{0,4} = 2,5$$

und

$$\frac{1}{16} \left| \frac{1-x}{1-\frac{x}{16}} \right| \leq \frac{1}{16} \frac{2}{1+\frac{1}{16}} = \frac{2}{17} < 0,2.$$

Es ist aber

$$2,5 + 0,5 + 0,2 = 3,2 < 3,24.$$

b) $|1-x| < 0,4$. Man hat mit Rücksicht auf (5') und (12)

$$\begin{aligned} \sum_{v=5}^{\infty} \frac{|c_v|}{4^{v-1}} \left| \frac{1-x^v}{1-x} \right| &< \frac{8,5}{4^4} + e \sum_{v=6}^{\infty} \frac{v^2}{4^{v-1}} < \frac{5}{32} + \frac{365}{27 \cdot 256} e < \frac{5}{32} + \frac{365 \cdot 2,72}{27 \cdot 256} \\ &= \frac{5}{32} + \frac{36,5}{256} + \frac{7,3}{27 \cdot 256} < \frac{76,5}{256} + \frac{0,3}{256} = \frac{76,8}{256} = 0,3, \end{aligned}$$

so daß die zu beweisende Ungleichung folgendermaßen lautet:

$$\left| 1 - \frac{x}{16} \right| + \frac{1}{2} |1-x| + \frac{1}{16} \left| \frac{(1-x)^2}{1-\frac{1}{16}} \right| < \frac{9}{25} \frac{1}{0,3} = 1,2.$$

Nun ist

$$\left| 1 - \frac{x}{16} \right| = \left| \frac{15}{16} + \frac{1-x}{16} \right| < \frac{15,4}{16} < 1 - \frac{0,32}{15} = 1 - 0,02,$$

ferner $\frac{1}{2} |1-x| < 0,2$ und

$$\frac{1}{16} \left| \frac{(1-x)^2}{1-\frac{1}{16}} \right| < \frac{0,16}{16 \left(1-\frac{1}{16}\right)} = \frac{0,16}{15} < \frac{0,30}{15} = 0,02,$$

womit die Behauptung für den Abschnitt $n=4$ gezeigt ist.

4. Für $n=2$ ist die Behauptung leicht zu beweisen. Es ist nämlich⁹⁾ $|c_2| \leq 2$, so daß für $c_2 \neq 0$, $|z_1| < \frac{1}{4}$, $|z_2| < \frac{1}{4}$

$$\left| 1 + c_2(z_1 + z_2) \right| > 1 - \frac{|c_2|}{2} \geq 0$$

gilt.

⁹⁾ Vgl. a. a. O. ⁸⁾, S. 85–86.

Wir betrachten endlich den Fall $n = 3$. Es genügt, den Satz für die von Herrn K. Löwner a. a. O.⁴⁾ betrachteten Schlitzabbildungen zu beweisen¹⁰⁾. Wir setzen dementsprechend

$$(13) \quad c_2 = 2 \int_0^{\infty} \kappa(\tau) e^{-\tau} d\tau, \quad c_3 = 4 \left(\int_0^{\infty} \kappa(\tau) e^{-\tau} d\tau \right)^2 - 2 \int_0^{\infty} \kappa^2(\tau) e^{-2\tau} d\tau,$$

wobei $\kappa(\tau)$ eine stetige Funktion von τ mit $|\kappa(\tau)| = 1$ bezeichnet¹¹⁾.

Die Behauptung lautet dann

$$1 + c_2(z_1 + z_2) + c_3(z_1^2 + z_1 z_2 + z_2^2) \neq 0 \quad \left(|z_1| < \frac{1}{4}, |z_2| < \frac{1}{4} \right).$$

Wir zeigen sogar, daß für $|z_1| = |z_2| = \frac{1}{4}$

$$(14) \quad 1 + \Re \{ c_2(z_1 + z_2) + c_3(z_1^2 + z_1 z_2 + z_2^2) \} > 0$$

ist.

Ersetzt man z_1 und z_2 durch εz_1 bzw. εz_2 und gleichzeitig $\kappa(\tau)$ durch $\bar{\varepsilon} \kappa(\tau)$, $|\varepsilon| = 1$, so bleibt die linke Seite von (14) unverändert. Da ε stets so bestimmt werden kann, daß εz_1 und εz_2 konjugiert komplex sind, so genügt es also (14) für den Fall zu beweisen, daß

$$z_1 = \frac{e^{i\varphi}}{4}, \quad z_2 = \frac{e^{-i\varphi}}{4} \quad (\varphi \text{ reell})$$

ist. Dann lautet die Behauptung:

$$1 + \cos \varphi \cdot \Re \int_0^{\infty} \kappa(\tau) e^{-\tau} d\tau + \frac{2 \cos 2\varphi + 1}{8} \Re \left\{ 2 \left(\int_0^{\infty} \kappa(\tau) e^{-\tau} d\tau \right)^2 - \int_0^{\infty} \kappa^2(\tau) e^{-2\tau} d\tau \right\} > 0.$$

Setzt man

$$\int_0^{\infty} \kappa(\tau) e^{-\tau} d\tau = x_1 + i y_1, \quad 2 \left(\int_0^{\infty} \kappa(\tau) e^{-\tau} d\tau \right)^2 - \int_0^{\infty} \kappa^2(\tau) e^{-2\tau} d\tau = x_2 + i y_2;$$

$$\cos \varphi = \lambda \quad (-1 \leq \lambda \leq 1),$$

so erhält die zu beweisende Ungleichung die folgende Form:

$$1 + \lambda x_1 + \frac{4\lambda^2 - 1}{8} x_2 > 0.$$

Die Ableitung des linksstehenden Ausdruckes nach λ ist $x_1 + \lambda x_2$; sie hat für $|x_1| \geq |x_2|$ konstantes Vorzeichen, wenn $-1 \leq \lambda \leq 1$, so daß in diesem Falle nur die Stellen $\lambda = \pm 1$ zu prüfen sind. Letzteres gilt über-

¹⁰⁾ Vgl. die Bemerkung a. a. O. 4), S. 106.

¹¹⁾ A. a. O. 4). Vgl. die Formeln (51) und (57).

hauptsächlich für $x_2 \leq 0$. Für $|x_1| < x_2$ hingegen muß allein die Stelle $\lambda = -\frac{x_1}{x_2}$ betrachtet werden. Wir haben folglich zu zeigen:

a) $1 - |x_1| + \frac{3}{8}x_2 > 0$, wenn $|x_1| \geq |x_2|$ oder $|x_1| < |x_2|$, $x_2 < 0$ ist.

b) $1 > \frac{x_1^2}{2x_2} + \frac{x_2}{8}$, wenn $|x_1| < |x_2|$, $x_2 > 0$ ist.

Die Ungleichung a) kann, wenn man $x(\tau) = x(\tau) + iy(\tau)$ setzt, folgendermaßen geschrieben werden:

$$|x_1| + \frac{3}{8} \left\{ 2y_1^2 - 2x_1^2 + \int_0^\infty (x^2(\tau) - y^2(\tau)) e^{-2\tau} d\tau \right\} < 1.$$

Wegen $|x_1| \leq 1$ ist aber $|x_1| - \frac{3}{4}x_1^2 \leq \frac{1}{3}$; es genügt also

$$2 \left(\int_0^\infty y(\tau) e^{-\tau} d\tau \right)^2 + \int_0^\infty (1 - 2y^2(\tau)) e^{-2\tau} d\tau < \frac{16}{9}$$

nachzuweisen. Die linke Seite ist hier

$$\leq 2 \int_0^\infty y^2(\tau) (e^{-\tau} - e^{-2\tau}) d\tau + \frac{1}{2} \leq 2 \int_0^\infty (e^{-\tau} - e^{-2\tau}) d\tau + \frac{1}{2} = \frac{3}{2} < \frac{16}{9}.$$

Zum Beweise von b) genügt es

$$1 > \frac{x_2^2}{2x_2} + \frac{x_2}{8} = \frac{5x_2}{8}, \quad \text{d. h. } x_2 < \frac{8}{5}$$

zu zeigen. Bekanntlich¹²⁾ gilt mehr, nämlich

$$|x_2| \leq |x_2 + iy_2| \leq \frac{3}{2} < \frac{8}{5}$$

Damit ist Satz I vollständig bewiesen.

§ 2.

Über die Klasse (S_0) .

Wir dürfen annehmen, daß $c_0 = 0$, $c_1 = 1$ ist.

1. Eine solche Funktion $f(z)$ gehört bekanntlich dann und nur dann zur Klasse (S_0) , wenn

$$(15) \quad \Re z \frac{f'(z)}{f(z)} > 0 \quad \text{für } |z| < 1$$

¹²⁾ A. a. O. 4), S. 120—121. Die Ungleichung $|x_2 + iy_2| \leq \frac{3}{2}$ ist mit $|c_3| \leq 3$ gleichwertig.

gilt¹³⁾. Man hat ferner

$$(16) \quad \frac{1}{2\pi} \int_0^{2\pi} z \frac{f'(z)}{f(z)} d\varphi = 1 \quad (z = r e^{i\varphi}, 0 < r < 1),$$

so daß

$$(17) \quad \Re z \frac{f'(z)}{f(z)} \geq \frac{1-|z|}{1+|z|} \quad (|z| < 1)$$

ist¹⁴⁾.

Außerdem gilt

$$(17') \quad \left| z \frac{f'(z)}{f(z)} \right| \leq \frac{1+|z|}{1-|z|} \quad (|z| < 1).^{15)}$$

Wir setzen nun für $n \geq 2$

$$(18) \quad z + c_3 z^3 + \dots + c_n z^n = s_n(z), \quad f(z) = s_n(z) + \varrho_n(z)$$

und stellen uns die Aufgabe, für $0 < |z| < \frac{1}{4}$

$$(19) \quad s_n(z) \neq 0$$

und

$$(20) \quad \Re z \frac{s_n'(z)}{s_n(z)} = \Re z \frac{f'(z) - \varrho_n'(z)}{f(z) - \varrho_n(z)} = \Re z \frac{f'(z)}{f(z)} + \Re z \frac{\varrho_n(z) \frac{f'(z)}{f(z)} - \varrho_n'(z)}{f(z) - \varrho_n(z)} > 0$$

nachzuweisen. Es genügt zu diesem Zwecke, — behaupte ich —, für $|z| = \frac{1}{4}$

$$(21) \quad |f(z)| > |\varrho_n(z)|$$

und

$$(22) \quad \frac{1 - \frac{1}{4}}{1 + \frac{1}{4}} \geq \frac{1}{4} \frac{\varrho_n(z) \left| \frac{f'(z)}{f(z)} \right| + |\varrho_n'(z)|}{|f(z)| - |\varrho_n(z)|}$$

¹³⁾ Vgl. a. a. O. ⁸⁾, S. 93.

¹⁴⁾ Dies folgt aus dem Harnackschen Lemma: Wenn $u(r, \varphi)$ eine für $r < 1$ reguläre positive Potentialfunktion ist, für welche

$$\frac{1}{2\pi} \int_0^{2\pi} u(r, \varphi) d\varphi = 1$$

gilt, dann hat man $\frac{1-r}{1+r} \leq u(r, \varphi) \leq \frac{1+r}{1-r}$.

¹⁵⁾ Dies folgt aus dem Carathéodoryschen Lemma: Wenn $u(r, \varphi)$ eine für $r < 1$ reguläre positive Potentialfunktion ist, für welche

$$\frac{1}{2\pi} \int_0^{2\pi} u(r, \varphi) d\varphi = 1$$

gilt, wenn ferner $w(z)$ die analytische Funktion bezeichnet, deren Realteil $u(r, \varphi)$ ist, die sich ferner für $z=0$ auf 1 reduziert, so hat man $|w(z)| \leq \frac{1+|z|}{1-|z|}$. — Nach Fußnote ⁶⁾ gilt übrigens (17') für eine beliebige Funktion $f(z)$ der Klasse (S) mit $f(0) = 0$.

zu zeigen. In der Tat ist $\frac{\varrho_n(z)}{f'(z)}$ regulär für $|z| < 1$. Aus der Richtigkeit von (21) für $|z| = \frac{1}{4}$ ergibt sich also dieselbe Ungleichung, also (19) für $0 < |z| < \frac{1}{4}$. Folglich ist $z \frac{s'_n(z)}{s_n(z)}$ regulär für $|z| \leq \frac{1}{4}$ und aus der Richtigkeit von (20) für $|z| = \frac{1}{4}$ folgt dieselbe Ungleichung für $|z| < \frac{1}{4}$. Für $|z| = \frac{1}{4}$ ist aber (20) eine Folge von (22), da doch (17) und (21) gelten.

Wir bemerken zunächst, daß die Koeffizientenabschätzung (5'') für die Klasse (S_0) bekannt ist¹⁶⁾. Zum Beweise von (22) benötigen wir außer der Formel (11) noch die folgende

$$(11') \quad \sum_{\nu=m}^{\infty} \nu x^\nu = x^m \frac{m(1-x) + x}{(1-x)^2} \quad (m = 0, 1, 2, \dots).$$

Es ist also, $|z| = \frac{1}{4}$,

$$(23) \quad |\varrho_n(z)| \leq \sum_{\nu=n+1}^{\infty} \frac{\nu}{4^\nu} = \frac{1}{4^{n+1}} \frac{\frac{3}{4}(n+1) + \frac{1}{4}}{\left(1 - \frac{1}{4}\right)^2} = \frac{3n+4}{9 \cdot 4^n}$$

und nach der in § 1, 1 durchgeführten Rechnung

$$(23') \quad |\varrho'_n(z)| \leq \sum_{\nu=n+1}^{\infty} \frac{\nu^2}{4^{\nu-1}} = \frac{9n^2 + 24n + 20}{27 \cdot 4^{n-1}}.$$

Man hat ferner wegen (17') für $|z| = \frac{1}{4}$

$$\left| \frac{f'(z)}{f(z)} \right| \leq 4 \frac{1 + \frac{1}{4}}{1 - \frac{1}{4}} = \frac{20}{3}.$$

Schließlich ergibt sich aus (2), $|z| = \frac{1}{4}$,

$$|f(z)| \geq \frac{\frac{1}{4}}{\left(1 + \frac{1}{4}\right)^2} = \frac{4}{25}.$$

Für $n = 4$ hat der Ausdruck rechts in (23) den Wert $\frac{1}{144} < \frac{4}{25}$, so daß (21) für $n \geq 4$ gilt. (22) ist nun eine Folge der Ungleichung

$$(24) \quad \frac{\frac{20}{3} \frac{3n+4}{9 \cdot 4^n} + \frac{9n^2 + 24n + 20}{27 \cdot 4^{n-1}}}{\frac{4}{25} - \frac{3n+4}{9 \cdot 4^n}} \leq \frac{12}{5},$$

¹⁶⁾ R. Nevanlinna, Über die konforme Abbildung von Sterngebieten, Översikt av Finska Vetenskaps-Soc. Förh. 63 (A) (1920—21), Nr. 6.

deren Richtigkeit für $n=4$ die für $n \geq 4$ zur Folge hat. Für $n=4$ ist sie aber richtig, weil doch

$$\frac{\frac{5}{108} + \frac{65}{432}}{\frac{4}{25} - \frac{1}{144}} < \frac{\frac{1}{18} + \frac{1}{6}}{\frac{1}{8} - \frac{1}{72}} = \frac{2 \cdot 72}{9 \cdot 8} = 2 < \frac{12}{5}$$

ist.

Bemerkung. Für $n=3$ wird der Ausdruck rechts in (23) gleich $\frac{13}{576} < \frac{4}{25}$. Die Ungleichung (24) ist dagegen in diesem Falle falsch; sie lautet nämlich in diesem Falle

$$\frac{\frac{65}{27 \cdot 16} + \frac{173}{27 \cdot 16}}{\frac{4}{25} - \frac{13}{576}} \leq \frac{12}{5}.$$

Die linke Seite ist aber größer als

$$\frac{\frac{4}{27} + \frac{10}{27}}{\frac{1}{5}} = \frac{70}{27} > \frac{12}{5}.$$

2. Für $n=2$ ist die Behauptung leicht zu beweisen. Es handelt sich dann um die Ungleichung

$$\Re \frac{1 + 2c_2 z}{1 + c_2 z} > 0 \quad (|z| < \frac{1}{4})$$

oder

$$\begin{aligned} \Re(1 + 2c_2 z)(1 + \bar{c}_2 \bar{z}) &= 1 + 3\Re c_2 z + 2|c_2|^2 |z|^2 \\ &= 2 \left| \frac{3}{4} + c_2 z \right|^2 - \frac{1}{8} > 0 \quad (|z| < \frac{1}{4}). \end{aligned}$$

Man hat aber $\left| \frac{3}{4} + c_2 z \right| > \frac{3}{4} - \frac{|c_2|}{4} \geq \frac{1}{4}$, ($c_2 \neq 0$), so daß

$$\Re(1 + 2c_2 z)(1 + \bar{c}_2 \bar{z}) > 2 \cdot \frac{1}{16} - \frac{1}{8} = 0$$

ist.

Wir kommen nun auf den etwas schwierigeren Fall $n=3$. Wir werden beweisen, daß für $|z| \leq \frac{1}{4}$

$$\Re \frac{1 + 2c_2 z + 3c_3 z^2}{1 + c_2 z + c_3 z^2} > 0$$

gilt. Der Nenner wird hier nicht 0, weil $|c_2 z + c_3 z^2| \leq 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{16} = \frac{11}{16} < 1$ ist. Man kann also $|z| = \frac{1}{4}$ annehmen. Es genügt ferner diese Ungleichung für $z = \frac{1}{4}$ zu beweisen. (Man betrachte $\bar{\varepsilon} f(\varepsilon z)$ mit passendem ε , $|\varepsilon| = 1$.) Die Behauptung lautet also:

$$(25) \quad \Re \frac{1 + \frac{c_2}{2} + \frac{3c_3}{16}}{1 + \frac{c_2}{4} + \frac{c_3}{16}} > 0.$$

Die Funktion

$$(26) \quad z \frac{f'(z)}{f(z)} = \frac{1 + 2c_2 z + 3c_3 z^2 + \dots}{1 + c_2 z + c_3 z^2 + \dots} = 1 + 2C_1 z + 2C_2 z^2 + \dots$$

hat für $|z| < 1$ positiven Realteil, so daß nach der Carathéodory-Toeplitzschen Theorie der Funktionen dieser Art¹⁷⁾

$$(27) \quad |C_1| \leq 1, \quad |C_2 - C_1^2| \leq 1 - |C_1|^2$$

gilt. (Umgekehrt gibt es zu vorgeschriebenen C_1, C_2 , welche (27) erfüllen, stets eine Funktion (26) mit positivem Realteil für $|z| < 1$, folglich eine auf die obige Weise normierte Funktion $f(z)$ der Klasse (S_0) .) Nun ist

$$c_2 = 2C_1, \quad c_3 = C_2 + 2C_1^{2m}.$$

Man kann also (25) auch folgendermaßen schreiben:

$$\Re \frac{1 + C_1 + \frac{3}{16}(C_2 + 2C_1^2)}{1 + \frac{C_1}{2} + \frac{1}{16}(C_2 + 2C_1^2)} > 0$$

oder wenn man $C_2 - C_1^2 = \eta(1 - |C_1|^2)$ setzt, wobei $|\eta| \leq 1$ ist,

$$\Re \frac{1 + C_1 + \frac{3}{16}(3C_1^2 + \eta(1 - |C_1|^2))}{1 + \frac{C_1}{2} + \frac{1}{16}(3C_1^2 + \eta(1 - |C_1|^2))} > 0.$$

Wir fassen den letzten Bruch bei festem C_1 und C_2 als eine Funktion von η auf. Sie ist regulär für $|\eta| \leq 1$, weil (vgl. oben)

$$\left| \frac{C_1}{2} + \frac{1}{16}(3C_1^2 + \eta(1 - |C_1|^2)) \right| \leq \frac{|C_1|}{2} + \frac{2|C_1|^2 + 1}{16} \leq \frac{11}{16} < 1$$

ist. Man darf folglich $|\eta| = 1$ annehmen. Wir haben also zu zeigen, daß

$$(25') \quad \Re \left\{ 1 + C_1 + \frac{9}{16}C_1^2 + \frac{3}{16}\eta(1 - |C_1|^2) \right\} \\ \times \left\{ 1 + \frac{\bar{C}_1}{2} + \frac{3}{16}\bar{C}_1^2 + \frac{1}{16}\bar{\eta}(1 - |C_1|^2) \right\} > 0$$

gilt. Wir setzen

$$1 + C_1 + \frac{9}{16}C_1^2 = 3w_1, \quad 1 + \frac{C_1}{2} + \frac{3}{16}C_1^2 = w_2.$$

Dann lautet (25')

$$\Re \left\{ 3w_1 + \frac{3}{16}\eta(1 - |C_1|^2) \right\} \cdot \left\{ \bar{w}_2 + \frac{1}{16}\bar{\eta}(1 - |C_1|^2) \right\} > 0$$

oder

$$\Re w_1 \bar{w}_2 + \frac{(1 - |C_1|^2)^2}{256} + \frac{1 - |C_1|^2}{16} \Re(w_1 + w_2) \bar{\eta} > 0.$$

¹⁷⁾ Vgl. den Enzyklopädieartikel von L. Bieberbach, Neuere Untersuchungen über Funktionen von komplexen Variablen [II C4, S. 379–532], S. 501–504.

Diese Ungleichung ist nun gleichwertig mit

$$\Re w_1 \bar{w}_2 + \frac{(1 - |C_1|^2)^2}{256} > \frac{1 - C_1^2}{16} |w_1 + w_2|,$$

d. h. mit

$$\left(\frac{|w_1 + w_2|}{2} - \frac{1 - |C_1|^2}{16} \right)^2 > \frac{w_1 - w_2^2}{4}$$

oder wegen $|w_1 + w_2| = \left| \frac{4}{3} + \frac{5}{6} C_1 + \frac{3}{8} C_1^2 \right| \geq \frac{4}{3} - \frac{5}{6} - \frac{3}{8} = \frac{1}{8}$ mit

$$|w_1 + w_2| - |w_1 - w_2| > \frac{1 - C_1^2}{8}.$$

Setzt man $C_1 = \zeta$, so lautet diese letzte Ungleichung

$$(28) \quad \left| 8 + 5\zeta + \frac{9}{4}\zeta^2 \right| - |4 + \zeta| > \frac{3}{4}(1 - |\zeta|^2) \quad (|\zeta| \leq 1).$$

3. Wir schreiben

$$\zeta = -4 + \zeta_1 = -4 + r e^{i\varphi};$$

es ist geometrisch klar, daß $3 \leq r \leq 5$ ist, daß ferner bei jedem festen r

$$(29) \quad -\varphi_0(r) \leq \varphi \leq \varphi_0(r)$$

gilt, wobei $\varphi_0(r)$ aus der Gleichung $|-4 + r e^{i\varphi}| = 1$ bestimmt werden kann, $0 < \varphi_0(r) < \frac{\pi}{2}$. Der Punkt $-4 + r e^{i\varphi_0(r)}$ liegt auf dem Einheitskreise. Man erhält

$$(30) \quad \cos \varphi_0(r) = \frac{15 + r^2}{8r}.$$

Wegen $8 + 5\zeta + \frac{9}{4}\zeta^2 = 24 - 13\zeta_1 + \frac{9}{4}\zeta_1^2$ kann nun (28) folgendermaßen geschrieben werden:

$$(28') \quad \left| 24 - 13\zeta_1 + \frac{9}{4}\zeta_1^2 \right| > \frac{3}{4} + Q,$$

wobei $Q = r - \frac{3}{4} | -4 + r e^{i\varphi} |^2 = -12 + r(1 + 6 \cos \varphi) - \frac{3}{4} r^2$ ist. Mit Rücksicht auf (29) und (30) hat man

$$(31) \quad Q \geq -12 + r + \frac{3}{4}(15 + r^2) - \frac{3}{4} r^2 = -\frac{3}{4} + r.$$

Wir schreiben nun (28') in der Form

$$|\alpha - \beta \zeta_1 + \gamma \zeta_1^2| > \frac{3}{4} + Q$$

oder

$$(32) \quad [(\alpha + \gamma r^2) \cos \varphi - \beta r]^2 + (\alpha - \gamma r^2)^2 (1 - \cos^2 \varphi) - \left(\frac{3}{4} + Q \right)^2 > 0,$$

wo übergangsweise $\alpha = 24$, $\beta = 13$, $\gamma = \frac{9}{4}$ gesetzt worden ist. Wir denken uns

$$\cos \varphi = \frac{Q + 12 - r + \frac{3}{4} r^2}{6r}$$

gesetzt und betrachten die linke Seite $f(r, Q)$ von (32) als Funktion von r und Q . Hierbei durchläuft r das Intervall $3 \leq r \leq 5$ und Q ein gewisses, von r abhängendes Intervall $Q_1(r) \leq Q \leq Q_2(r)$. Es ist $Q_1(r) = -\frac{3}{4} + r$. Wir zeigen, daß $f(r, Q)$ bei festem r , $3 \leq r \leq 5$, als Funktion von Q für $Q \geq Q_1(r)$ monoton wächst.

Man hat

$$\begin{aligned} \frac{\partial f(r, Q)}{\partial Q} &= 2 \frac{\alpha + \gamma r^2}{6r} [(\alpha + \gamma r^2) \cos \varphi - \beta r] - 2 \frac{(\alpha - \gamma r^2)^2}{6r} \cos \varphi - 2 \left(\frac{3}{4} + Q \right) \\ &= \frac{4}{3} \alpha \gamma r \cos \varphi - \frac{\beta}{3} (\alpha + \gamma r^2) - \frac{3}{2} - 2Q \\ &= \frac{4}{3} \alpha \gamma \frac{Q + 12 - r + \frac{3}{4} r^2}{6} - \frac{\beta}{3} (\alpha + \gamma r^2) - \frac{3}{2} - 2Q. \end{aligned}$$

Dieser Ausdruck ist positiv für $Q \geq Q_1(r)$, wenn er für $Q = Q_1(r)$ positiv ist, da ja $\frac{\partial^2 f(r, Q)}{\partial Q^2} = \frac{2}{9} \alpha \gamma - 2 > 0$ ausfällt. Nun hat man aber

$$\begin{aligned} \frac{4}{3} \alpha \gamma \frac{15 + r^2}{8} - \frac{\beta}{3} (\alpha + \gamma r^2) - \frac{3}{2} - 2 \left(-\frac{3}{4} + r \right) \\ = \frac{(\alpha - 2\beta) \gamma r^2 - 2r + \alpha(15\gamma - 2\beta)}{6} = -\frac{3}{4} r^2 - 2r + 31 > 0. \end{aligned}$$

Diese Ungleichung ist nämlich für $r = 5$, folglich auch für $3 \leq r \leq 5$ richtig.

Damit ist gezeigt, daß $f(r, Q)$ für $Q = Q_1(r)$ Minimum wird. Dies heißt $\cos \varphi = \cos \varphi_0(r)$ oder $|\zeta| = 1$, $\zeta = e^{i\psi}$. Dann lautet aber die Behauptung

$$(28'') \quad \left| 8e^{-i\psi} + 5 + \frac{9}{4} e^{i\psi} \right| > |4 + e^{i\psi}|,$$

d. h.

$$\left(\frac{41}{4} \cos \psi + 5 \right)^2 + \left(\frac{23}{4} \sin \psi \right)^2 > 17 + 8 \cos \psi$$

oder

$$72 \cos^2 \psi + \frac{189}{2} \cos \psi + 8 + \left(\frac{23}{4} \right)^2 > 0.$$

Die linke Seite wird für $\cos \psi = -\frac{189}{4 \cdot 72}$ Minimum, und zwar beträgt dies

$$-\frac{189^2}{16 \cdot 72} + 8 + \left(\frac{23}{4} \right)^2 > 0,$$

weil $\frac{189^2}{16 \cdot 72} < \frac{(16 \cdot 12)^2}{16 \cdot 72} = 32$ und $\left(\frac{23}{4} \right)^2 > 24$ ist.

Damit ist (28''), also der Satz II bewiesen.

4. D. h. ... Bild des Kreises ... bei der Abbildung

selben Kreises bei der Abbildung $w = zf'(z)$ in bezug auf $w = f(0)$ schlicht und sternförmig ist¹⁸⁾. Hieraus folgt der Satz

II'. Die Funktion

$$f(z) = c_0 + c_1 z + c_2 z^2 + \dots + c_n z^n + \dots$$

sei regulär und schlicht im Kreise $|z| < 1$ und bilde denselben auf ein konvexes Gebiet ab. Dann sind sämtliche Abschnitte

$$c_0 + c_1 z + c_2 z^2 + \dots + c_n z^n \quad (n = 1, 2, 3, \dots)$$

schlicht und „konvex“ im Kreise $|z| < \frac{1}{4}$.

Die Konstante $\frac{1}{4}$ kann auch hier durch keine größere ersetzt werden.

§ 3.

Über die Klasse (E^*): Die obere Abschätzung von σ .

In einer gemeinsamen Arbeit von W. Rogosinski und G. Szegő¹⁹⁾ ist der folgende Satz bewiesen worden:

Es sei $0 < \vartheta \leq \frac{\pi}{2}$. Wenn die Potenzreihe

$$f(z) = c_0 + c_1 z + c_2 z^2 + \dots + c_n z^n + \dots$$

für $|z| < 1$ konvergiert und $|f(z)| \leq 1$ ist, dann gilt für $n = 0, 1, 2, \dots$

$$(33) \quad \left| \sum_{\nu=0}^n \frac{\sin\left(1 - \frac{\nu}{n+1}\right) \vartheta}{\sin \vartheta} c_\nu \right| < \frac{1}{\sin \vartheta} \int_0^\vartheta J_0^2(x) dx,$$

wo $J_0(x)$ die Besselsche Funktion 0-ter Ordnung bedeutet.

Für $\vartheta \rightarrow 0$ geht diese Ungleichung in die folgende über:

$$\left| \sum_{\nu=0}^n \left(1 - \frac{\nu}{n+1}\right) c_\nu \right| \leq 1,$$

welche besagt, daß die arithmetischen Mittel erster Ordnung der Abschnitte von $f(z)$ dem Betrage nach unterhalb der Schranke 1 bleiben.

Herr Fejér hat a. a. O.³⁾, S. 187–188, die letzte Ungleichung benutzt, um $s_n = c_0 + c_1 + c_2 + \dots + c_n$ bei einer Funktion der Klasse (E^*) abzuschätzen. Benutzt man auf ähnliche Weise die Ungleichung (33), so ergibt sich (für $\vartheta = \frac{\pi}{2}$) ein etwas schärferes Resultat.

¹⁸⁾ Vgl. etwa G. Pólya und G. Szegő, Aufgaben und Lehrsätze aus der Analysis 1 (Berlin: Julius Springer 1925), S. 105, Aufgabe 110.

¹⁹⁾ W. Rogosinski und G. Szegő, Über die Abschnitte von Potenzreihen, die in einem Kreise beschränkt bleiben, Math. Zeitschrift 28 (1928), S. 73–94. Vgl. insbesondere § 3.

In der Tat gilt

$$s_n = \sum_{\nu=0}^n \frac{\sin\left(1 - \frac{\nu}{n+1}\right)\vartheta}{\sin\vartheta} c_\nu + \sum_{\nu=1}^n \frac{\sin\vartheta - \sin\left(1 - \frac{\nu}{n+1}\right)\vartheta}{\sin\vartheta} c_\nu,$$

folglich

$$|s_n| < \frac{1}{\sin\vartheta} \int_0^\vartheta J_0^2(x) dx + \sum_{\nu=1}^n \frac{\sin\vartheta - \sin\left(1 - \frac{\nu}{n+1}\right)\vartheta}{\sin\vartheta} |c_\nu|.$$

Wegen $\sum_{\nu=1}^{\infty} \nu |c_\nu|^2 \leq 1$ (vgl. a. a. O.²⁹), S. 188) erhalten wir hieraus

$$|s_n| < \frac{1}{\sin\vartheta} \int_0^\vartheta J_0^2(x) dx + \frac{1}{\sin\vartheta} \left\{ \sum_{\nu=1}^n \frac{[\sin\vartheta - \sin\left(1 - \frac{\nu}{n+1}\right)\vartheta]^2}{\nu} \right\}^{\frac{1}{2}}.$$

Mit der Bezeichnung

$$\frac{[\sin\vartheta - \sin(\vartheta - x)]^2}{x} = h(x) \quad (0 < x \leq \vartheta; h(0) = 0)$$

kann der Ausdruck in der Klammer $\{ \}$ folgendermaßen geschrieben werden:

$$\frac{\vartheta}{n+1} \sum_{\nu=0}^n h\left(\frac{\nu}{n+1}\vartheta\right).$$

Man hat aber für $0 < x \leq \vartheta$

$$x^2 h'(x) = [\sin\vartheta - \sin(\vartheta - x)] [2x \cos(\vartheta - x) - \sin\vartheta + \sin(\vartheta - x)] \geq 0,$$

weil doch

$$\sin\vartheta - \sin(\vartheta - x) \leq x \cos(\vartheta - x)$$

ist. Folglich ist $h(x)$ monoton wachsend und es gilt

$$\frac{\vartheta}{n+1} \sum_{\nu=0}^n h\left(\frac{\nu}{n+1}\vartheta\right) \leq \int_0^\vartheta h(x) dx,$$

d. h.

$$(34) \quad |s_n| < \frac{1}{\sin\vartheta} \int_0^\vartheta J_0^2(x) dx + \frac{1}{\sin\vartheta} \left\{ \int_0^\vartheta \frac{[\sin\vartheta - \sin(\vartheta - x)]^2}{x} dx \right\}^{\frac{1}{2}}.$$

Für $\vartheta = \frac{\pi}{2}$ ist das erste Glied

$$\int_0^{\frac{\pi}{2}} J_0^2(x) dx = 1,0777\dots,^{20)}$$

²⁰⁾ A. a. O.¹⁸⁾, S. 77.

während das zweite

$$= \left\{ \int_0^{\frac{\pi}{2}} \frac{(1 - \cos x)^2}{x} dx \right\}^{\frac{1}{2}} = 2 \left\{ \int_0^{\frac{\pi}{4}} \frac{\sin^4 x}{x} dx \right\}^{\frac{1}{2}} < 2 \left\{ \int_0^{\frac{\pi}{4}} \sin^3 x \frac{2 + \cos x}{3} dx \right\}^{\frac{1}{2}}.$$

Man hat

$$\int_0^{\frac{\pi}{4}} \sin^3 x \frac{2 + \cos x}{3} dx = \left(-\frac{1}{2} \cos x - \frac{1}{24} \cos 2x + \frac{1}{18} \cos 3x + \frac{1}{96} \cos 4x \right)_0^{\frac{\pi}{4}} \\ = \frac{67 - 40\sqrt{2}}{144}$$

und

$$\frac{\sqrt{67 - 40\sqrt{2}}}{6} = 0,5382 \dots,$$

so daß tatsächlich

$$|s_n| < 1,0777 \dots + 0,5382 \dots = 1,6160 \dots$$

gilt.

§ 4.

Über die Klasse (E^*): Die Berechnung von σ_2 .

Es sei

$$f(z) = c_0 + c_1 z + c_2 z^2 + \dots + c_n z^n + \dots$$

eine Funktion der Klasse (E^*), d. h. regulär und schlicht für $|z| < 1$, ferner sei daselbst $|f(z)| \leq 1$. Nach der in der Einleitung gegebenen Definition ist σ_n die obere Grenze von

$$|c_0 + c_1 + c_2 + \dots + c_n|$$

für alle Funktionen der Klasse (E^*). Es ist $\sigma_1 = \frac{5}{4} = 1,25$. Wir stellen uns die Aufgabe, σ_2 zu berechnen.

Es sei

$$(35) \quad \psi(z) = \gamma_1 z + \gamma_2 z^2 + \dots + \gamma_n z^n + \dots$$

eine beliebige Funktion der Klasse (E^*), für die außerdem $\psi(0) = 0$ gilt. Es sei ferner $|\alpha| < 1$. Dann ist

$$(36) \quad f(z) = \frac{\psi(z) - \alpha}{1 - \bar{\alpha}\psi(z)} \quad (|\alpha| < 1)$$

stets eine Funktion der Klasse (E^*). Umgekehrt kann eine beliebige Funktion $f(z)$ der Klasse (E^*) immer in dieser Form dargestellt werden, indem man $\alpha = -f(0)$ setzt.

Wir erhalten durch Koeffizientenvergleichung

$$(37) \quad c_0 = -\alpha, \quad c_1 = \gamma_1(1 - |\alpha|^2), \quad c_2 = (\gamma_2 + \bar{\alpha}\gamma_1^2)(1 - |\alpha|^2), \quad \dots,$$

so daß

$$c_0 + c_1 + c_2 = -\alpha + (\gamma_1 + \gamma_2 + \bar{\alpha}\gamma_1^2)(1 - |\alpha|^2).$$

Es handelt sich um das Maximum des absoluten Betrages dieses Ausdruckes, während α alle Zahlen mit $|\alpha| < 1$ durchläuft und γ_1, γ_2 die Anfangskoeffizienten einer beliebigen Funktion $\psi(z)$ der Klasse (E^*) sind. Man kann hierbei $\gamma_1 = \gamma$ ohne Beschränkung der Allgemeinheit als reell und positiv annehmen.

Es ist offenbar

$$(38) \quad 0 < \gamma \leq 1.$$

Man hat ferner nach einem Satze von G. Pick²¹⁾

$$(39) \quad |\gamma_2| \leq 2\gamma(1 - \gamma).$$

Umgekehrt, wenn zwei Zahlen $\gamma_1 = \gamma, \gamma_2$ diesen beiden Ungleichungen genügen, so existiert stets eine Funktion $\psi(z)$ der Klasse (E^*) mit diesen Anfangskoeffizienten.

Es ist also

$$(40) \quad |c_0 + c_1 + c_2| \leq |-\alpha + (\gamma + \bar{\alpha}\gamma^2)(1 - |\alpha|^2)| + 2\gamma(1 - \gamma)(1 - |\alpha|^2),$$

wo man bei passender Wahl von $f(z)$ das Gleichheitszeichen realisieren kann. Wir wollen das Maximum der rechten Seite von (40) berechnen, wenn α alle Zahlen mit $|\alpha| < 1$ und γ alle Zahlen mit $0 < \gamma \leq 1$ durchläuft.

Wir setzen $\alpha = re^{i\varphi}$ ($0 < r < 1$) und betrachten bei festem r den Ausdruck

$$| -re^{i\varphi} + (\gamma + re^{-i\varphi}\gamma^2)(1 - r^2) | = \sqrt{(Ar \cos \varphi + B)^2 + C^2 r^2 \sin^2 \varphi}.$$

Hierbei haben A, B, C die folgende Bedeutung:

$$A = 1 - \gamma^2(1 - r^2), \quad C = 1 + \gamma^2(1 - r^2), \quad B = -\gamma(1 - r^2).$$

Wir bemerken, daß

$$A^2 - C^2 = 4\gamma B < 0$$

ist. Es handelt sich, wenn man zur Abkürzung $\cos \varphi = \lambda$ ($-1 \leq \lambda \leq 1$) setzt, um die Funktion

$$u(\lambda) = (A^2 - C^2)r^2\lambda^2 + 2ABr\lambda + B^2 + C^2r^2,$$

deren Ableitung nach λ

$$u'(\lambda) = 2(A^2 - C^2)r^2\lambda + 2ABr = 2Br(4\gamma r\lambda + A)$$

ist. Wir unterscheiden zwei Fälle:

²¹⁾ G. Pick, Über die konforme Abbildung eines Kreises auf ein schlichtes und zugleich beschränktes Gebiet, Sitzungsberichte d. math.-naturwiss. Klasse d. Akademie d. Wiss. zu Wien, Zweite Abteilung, 126 (1917), S. 247–263.

$$1. \quad A \geq 4\gamma r.$$

Dann ist $u'(\lambda) < 0$ für $-1 < \lambda \leq 1$, so daß

$$\text{Max}_{-1 \leq \lambda \leq 1} u(\lambda) = u(-1)$$

wird.

$$2. \quad A < 4\gamma r.$$

In diesem Falle ist $u'(\lambda) = 0$ für $\lambda = -\frac{A}{4\gamma r}$, und zwar besitzt hier $u(\lambda)$ ein Maximum, d. h.

$$\text{Max}_{-1 \leq \lambda \leq 1} u(\lambda) = u\left(-\frac{A}{4\gamma r}\right).$$

Die Bedingung 1 besagt, daß

$$(41) \quad \gamma^2(1-r^2) + 4\gamma r \leq 1$$

ist. Bei festem r , $0 < r < 1$, ist die linke Seite dieser Ungleichung monoton wachsend mit γ , und da $1-r^2+4r > 1$ ist, so gibt es einen Wert $g_0(r)$, $0 < g_0(r) < 1$, derart, daß für $0 < \gamma \leq g_0(r)$ die Bedingung 1, für $g_0(r) < \gamma \leq 1$ dagegen 2 erfüllt ist. Es ist übrigens

$$g_0(r) = \frac{-2r + \sqrt{3r^2 + 1}}{1-r^2}.$$

In dem Falle 1 erhalten wir bei festem r und γ als das Maximum der rechten Seite von (40)

$$(42) \quad r + (\gamma - r\gamma^2)(1-r^2) + 2\gamma(1-\gamma)(1-r^2) \\ = r + 3\gamma(1-r^2) - (2+r)(1-r^2)\gamma^2 = P(r, \gamma).$$

In dem Falle 2 setzen wir $\lambda = \cos \varphi = -\frac{A}{4\gamma r}$ und erhalten für dasselbe Maximum

$$\sqrt{\left(B - \frac{A^2}{4\gamma}\right)^2 + C^2\left(r^2 - \frac{A^2}{16\gamma^2}\right)} + 2\gamma(1-\gamma)(1-r^2) = Q(r, \gamma).$$

Nun ist

$$B - \frac{A^2}{4\gamma} = B - \frac{C^2 + 4\gamma B}{4\gamma} = -\frac{C^2}{4\gamma}$$

und

$$r^2 - \frac{A^2}{16\gamma^2} = -\frac{C^2}{16\gamma^2} + \frac{3r^2 + 1}{4},$$

so daß

$$(43) \quad Q(r, \gamma) = \frac{C}{2} \sqrt{3r^2 + 1} + 2\gamma(1-\gamma)(1-r^2) \\ = \frac{1}{2} \sqrt{3r^2 + 1} + 2\gamma(1-r^2) + \gamma^2(1-r^2) \left(\frac{1}{2} \sqrt{3r^2 + 1} - 2\right).$$

Es ist natürlich $P(r, g_0(r)) = Q(r, g_0(r))$. Wir haben zunächst bei festem r , $0 < r < 1$, das Maximum $M(r)$ der stetigen Funktion von γ zu be-

stimmen, welche für $0 < \gamma \leq g_0(r)$ mit $P(r, \gamma)$ und für $g_0(r) < \gamma \leq 1$ mit $Q(r, \gamma)$ übereinstimmt. Es wird dann

$$(44) \quad \sigma_2 = \operatorname{Max}_{0 < r < 1} M(r).^{22)}$$

Nun verschwindet

$$\frac{\partial P(r, \gamma)}{\partial \gamma} = (1 - r^2)[3 - 2(2 + r)\gamma]$$

für den einzigen Wert $\gamma = g_1(r) = \frac{3}{2(2+r)}$. Wir wollen die Bedingung in r dafür ermitteln, daß $g_1(r) \leq g_0(r)$ ist. Dies heißt nach (41)

$$\left(\frac{3}{2(2+r)}\right)^2 (1 - r^2) + \frac{3}{2(2+r)} 4r \leq 1$$

oder

$$(45) \quad 11r^2 + 32r \leq 7.$$

Andersseits verschwindet

$$\frac{\partial Q(r, \gamma)}{\partial \gamma} = (1 - r^2)[2 - (4 - \sqrt{3r^2 + 1})\gamma]$$

nur für $\gamma = g_2(r) = \frac{2}{4 - \sqrt{3r^2 + 1}}$. Berechnen wir die Werte von r , für welche $g_2(r) \leq g_1(r)$ gilt. Diese Ungleichung besagt

$$\frac{4 - \sqrt{3r^2 + 1}}{2} \geq \frac{2(2+r)}{3}$$

oder

$$\frac{4}{3}(1 - r) \geq \sqrt{3r^2 + 1},$$

was, wie man leicht nachrechnet, mit (45) identisch ist.

Es sei also r' die positive Wurzel von $11r^2 + 32r = 7$. Man hat übrigens

$$(46) \quad r' = \frac{3\sqrt{37} - 16}{11}, \quad 0 < r' < 1.$$

Dann gilt

$$(47') \quad g_0(r) \geq g_1(r) \geq g_2(r) \quad \text{für } 0 < r \leq r'$$

und

$$(47'') \quad g_0(r) \leq g_1(r) \leq g_2(r) \quad \text{für } r' \leq r < 1.$$

Das Gleichheitszeichen tritt nur für $r = r'$ ein.

Es sei nun zunächst $0 < r \leq r'$. Dann erreicht $P(r, \gamma)$ im Intervall $0 < \gamma \leq g_0(r)$ für $\gamma = g_1(r)$ sein Maximum. Dagegen wird $Q(r, \gamma)$ im Intervall $g_0(r) \leq \gamma \leq 1$ abnehmen und erreicht für $\gamma = g_0(r)$ sein Maximum. Es ist aber $Q(r, g_0(r)) = P(r, g_0(r)) \leq P(r, g_1(r))$, folglich ist

²²⁾ Wir werden gleich sehen, daß $M(r)$ bis zu einem inneren Punkte des Intervalls $0, 1$ monoton zunimmt und von da ab monoton abnimmt.

das gesuchte Maximum

$$(48) \quad M(r) = P(r, g_1(r)) = r + \frac{9}{2} \frac{1-r^2}{2+r} - \frac{9}{4} \frac{1-r^2}{2+r} \\ = \frac{9+8r-5r^2}{4(2+r)} \quad (0 < r \leq r').$$

Es sei anderseits $r' < r < 1$. Dann ist $P(r, \gamma)$ für $0 < \gamma \leq g_0(r)$ wachsend, erreicht also für $\gamma = g_0(r)$ sein Maximum. Ferner wird $Q(r, \gamma)$ im Intervall $g_0(r) \leq \gamma \leq 1$ an der Stelle $\gamma = g_2(r)$ Maximum. Durch Beachtung von $P(r, g_0(r)) = Q(r, g_0(r)) < Q(r, g_2(r))$ ersieht man also, daß

$$(49) \quad M(r) = Q(r, g_2(r)) = \frac{1}{2} \sqrt{3r^2 + 1} + \frac{2(1-r^2)}{4 - \sqrt{3r^2 + 1}} \quad (r' < r < 1)$$

ist.

Die beiden Ausdrücke (48) und (49) stimmen für $r = r'$ wegen $g_0(r') = g_1(r') = g_2(r')$ überein, so daß $M(r)$ für $r = r'$ stetig ist.

Es sei $0 < r < r'$. Man hat

$$4(2+r)^2 M'(r) = (8-10r)(2+r) - 9 - 8r + 5r^2 \\ = 7 - 20r - 5r^2 > 0,$$

weil

$$5r'^2 + 20r' < 11r'^2 + 32r' = 7$$

ist. Folglich hat man

$$(50) \quad \text{Max}_{0 < r \leq r'} M(r) = M(r').$$

Es sei anderseits $r' < r < 1$. Wir setzen

$$4 - \sqrt{3r^2 + 1} = \varrho, \quad r^2 = \frac{15 - 8\varrho + \varrho^2}{3},$$

so daß

$$M(r) = \frac{4-\varrho}{2} + \frac{2}{\varrho} \left(1 - \frac{15-8\varrho+\varrho^2}{3} \right) = -\frac{8}{\varrho} + \frac{22}{3} - \frac{7}{6}\varrho.$$

Man hat also

$$M'(r) = \left(\frac{8}{\varrho^2} - \frac{7}{6} \right) \frac{d\varrho}{dr}.$$

Wenn r wächst, so nimmt ϱ ab und es gilt

$$2 < \varrho < 4 - \sqrt{3r'^2 + 1}.$$

$M'(r)$ verschwindet für $\varrho = 4\sqrt{\frac{3}{7}}$ und wir behaupten, daß

$$4\sqrt{\frac{3}{7}} < 4 - \sqrt{3r'^2 + 1}$$

gilt. Diese Behauptung kann entweder durch direkte Rechnung oder einfacher folgendermaßen bewiesen werden. Im gegenteiligen Falle müßte $M(r)$ im ganzen Intervalle $0 < r < 1$ monoton wachsen. Dies ist aber

unmöglich, weil doch

$$\lim_{r \rightarrow +0} M(r) = \frac{9}{8}, \quad \lim_{r \rightarrow 1-0} M(r) = 1$$

gilt.

Hiermit wird $M(r)$ für $\rho = 4\sqrt{\frac{3}{7}}$ Maximum und zugleich sieht man, daß dieser Maximalwert größer als $M(r')$ ausfällt. Wir haben also

$$\sigma_2 = \text{Max}_{0 < r < 1} M(r) = -2\sqrt{\frac{7}{3}} + \frac{22}{3} - 2\sqrt{\frac{7}{3}} = \frac{22}{3} - 4\sqrt{\frac{7}{3}} = 1,2232 \dots$$

Bekanntlich wird σ_1 nur bei den Funktionen

$$f(z) = \varepsilon \frac{z + \frac{1}{2}}{1 + \frac{z}{2}} \quad (|\varepsilon| = 1)$$

erreicht. Sie bilden den Einheitskreis $|z| \leq 1$ auf sich selbst ab.

Für welche Funktionen der Klasse (E^*) wird nun σ_2 erreicht, d. h. bei welcher Funktion wird der zweite Abschnitt (bei passendem z , $|z| = 1$) gleich σ_2 ?

Soll σ_2 erreicht werden, so muß zunächst in der Ungleichung (39) das Gleichheitszeichen gelten, d. h. die Funktion $\psi(z)$ muß den Einheitskreis $|z| < 1$ auf den längs eines geradlinigen, von einem Peripheriepunkt senkrecht nach innen gerichteten Schlitzes aufgeschnittenen Einheitskreis abbilden²³⁾. Nach der Abbildung (36) erhalten wir wiederum den aufgeschnittenen Einheitskreis, wobei der Schlitz ein von einem Peripheriepunkt ausgehender, nach innen gerichteter passender Orthogonalbogen ist. Dies ist das Bildgebiet bei der fraglichen Abbildung $w = f(z)$. Man erhält sämtliche Abbildungen dieser Art aus einer einzigen, wenn man das eben erwähnte Bildgebiet einer beliebigen Drehung um den Nullpunkt und einer Spiegelung an der reellen Achse unterwirft.

Königsberg, November 1927.

²³⁾ Vgl. a. a. O. ²¹⁾.

Über Folgen analytischer Funktionen.

Von

F. Hartogs in München und A. Rosenthal in Heidelberg.

Herr W. F. Osgood hat im Jahre 1901 den für Folgen von regulären analytischen Funktionen grundlegenden Satz bewiesen, daß, wenn eine solche in einem Gebiet \mathcal{G} der z -Ebene durchweg konvergiert, stets ein Teilgebiet von \mathcal{G} existieren muß, in welchem die Folge *gleichmäßig* konvergiert¹⁾. Die Teilgebiete \mathcal{G}_ν gleichmäßiger Konvergenz müssen daher in \mathcal{G} überall dicht liegen und ferner infolge eines bekannten Rungeschen Satzes²⁾ einfach zusammenhängend sein. Es liegt nun nahe, sich die Frage vorzulegen, ob die Gebiete \mathcal{G}_ν außer diesen noch anderen Bedingungen unterworfen sind und welches dann der genaue Umfang dieser letzteren ist. Diese Frage wird im ersten Teil der vorliegenden Arbeit behandelt; es zeigt sich, daß in der Tat noch weitere Einschränkungen weniger einfacher Natur hinzutreten, und zwar können die notwendigen und hinreichenden Bedingungen, denen die Stellen gleichmäßiger und ungleichmäßiger Konvergenz genügen, vollständig angegeben werden (Kap. I).

Eine weitere, hiermit zusammenhängende Frage ist die, ob bei einer derartigen Folge die Grenzfunktion, die in jedem \mathcal{G}_ν natürlich regulär ist, für jedes dieser Gebiete (und darüber hinaus auch für die Stellen ungleichmäßiger Konvergenz) willkürlich vorgeschrieben werden kann³⁾. Auch die Beantwortung dieser Frage ist von den Verfassern zu einem gewissen Abschluß gebracht worden, wobei sich gezeigt hat, daß hier bezüglich der

¹⁾ W. F. Osgood, Note on the functions defined by infinite series whose terms are analytic functions, *Ann. of Math.* (2) **3** (1901), p. 25–34.

²⁾ C. Runge, *Acta Math.* **6** (1885), p. 247.

³⁾ Einige diesbezügliche Angaben, welche die einfachsten Fälle betreffen, sowie zugehörige Beispiele (insbesondere von Funktionenfolgen, die in unendlich vielen Gebieten verschiedene analytische Funktionen darstellen) finden sich bereits in der zitierten Osgoodschen Arbeit (§ 2), ferner bei P. Montel, *Ann. Ec. Norm.* (3) **24** (1907), p. 318–320, *Bull. sc. math.* (2) **30** (1906), p. 189–192, sowie *Leçons sur les séries de polynomes* (Paris 1910), p. 110–111, 116–119.

Gebiete \mathfrak{G} , im allgemeinen schärfere Einschränkungen in Kraft treten als bei der obigen ersten Frage. Die hiermit unmittelbar verknüpfte Frage nach der Natur der Grenzfunktion einer beliebigen konvergenten Folge von regulären analytischen Funktionen⁴⁾ findet durch diese Untersuchungen ebenfalls ihre, wenn auch nicht nach allen Richtungen hin völlig abschließende Erledigung (Kap. II).

Der 3. Teil der Arbeit (Kap. III) ist einer genaueren geometrisch-mengentheoretischen Untersuchung der für die Gebiete \mathfrak{G} , gefundenen Bedingungen gewidmet. Mit Hilfe gewisser neuer, auf Punktmengen bezüglicher Begriffsbildungen, die zu einem Analogon der Cantorsche Theorie führen, gelingt es dabei, für jede der beiden Hauptbedingungen durchsichtiger Formulierungen zu finden, in denen nur die mengentheoretischen Eigenschaften der aus den Gebieten \mathfrak{G} , bestehenden Gesamtheit (bzw. ihrer Komplementärmenge) eine Rolle spielen. Die einfachsten Bedingungsformen sind die in § 10 aufgestellten; doch dürften sich für die wirklichen Anwendungen die in §§ 8—9 gegebenen wohl als brauchbarer erweisen.

Eine leichte Verallgemeinerung der in diesen letzteren Untersuchungen enthaltenen Gedankengänge führt (§ 11) noch zur Beantwortung einer allgemeinen topologischen Frage, nämlich unter welchen Bedingungen eine ebene Punktmenge sich aus abzählbar vielen, die Ebene nicht zerlegenden, abgeschlossenen Mengen zusammensetzen läßt (und entsprechend für den n -dimensionalen Raum)⁵⁾.

Kapitel I.

Verteilung der Stellen gleichmäßiger und ungleichmäßiger Konvergenz.

§ 1.

Einfachste notwendige Bedingungen. Hilfssätze.

Ein beliebiges, in der Ebene der komplexen Veränderlichen z gelegenes,

⁴⁾ Auf diese (die vorher erwähnten mit umfassende) Frage hat als noch unbeantwortet insbesondere auch P. Montel mehrfach hingewiesen; s. Ann. Ec. Norm. (3), 24 (1907), p. 332—334; Leçons sur les séries de polynomes (Paris 1910), p. 122.

⁵⁾ Unmittelbar vor Abschluß der vorliegenden Arbeit wurde den Verfassern eine in den Par. C. R. 184 (1927), S. 1634/5 erschienene Note des Herrn M. Lavrentieff bekannt, in der im wesentlichen dieselben Probleme behandelt werden. Die Frage, ob die dort in aller Kürze angegebenen Resultate, die jedenfalls in einer völlig anderen Form erscheinen, zu denen unserer Arbeit in irgendeiner engeren Beziehung stehen, würde einer besonderen Untersuchung bedürfen. Die Frage nach den Eigenschaften der Grenzfunktion wird dort übrigens nur für Folgen *harmonischer* Funktionen behandelt, während die entsprechende Frage für Folgen *analytischer* Funktionen ausdrücklich als eine offenbleibende bezeichnet wird.

beschränktes, einfach zusammenhängendes Gebiet⁶⁾ werde mit \mathfrak{G} , seine Begrenzung mit C bezeichnet. Ist nun eine unendliche Folge von analytischen Funktionen $f_n(z)$ ($n = 1, 2, \dots$) gegeben, welche in \mathfrak{G} sämtlich eindeutig definiert und regulär sind, und ist die Folge im ganzen Gebiet \mathfrak{G} konvergent, so kann man bezüglich jedes einzelnen Punktes von \mathfrak{G} zwei Fälle unterscheiden, je nachdem die Folge in einer gewissen (wenn auch noch so kleinen) Umgebung des betrachteten Punktes *gleichmäßig* konvergiert („Stelle gleichmäßiger Konvergenz“) oder nicht („Stelle ungleichmäßiger Konvergenz“⁷⁾). \mathfrak{G} ist hiernach im allgemeinen in zwei Punkt-mengen zerspalten und es ist nun die Aufgabe, die notwendigen und hinreichenden Bedingungen aufzustellen, denen die Menge der Stellen gleichmäßiger (bzw. der Stellen ungleichmäßiger) Konvergenz genügen muß. Die aus den Stellen ungleichmäßiger Konvergenz bestehende Punktmenge soll im folgenden stets mit Γ , die aus Γ und C zusammengesetzte Punktmenge mit \mathfrak{E} bezeichnet werden.

Eine erste notwendige Bedingung besteht darin, daß \mathfrak{E} stets ein *Kontinuum* (d. h. abgeschlossen und zusammenhängend) sein muß, oder, was hiermit gleichbedeutend, daß die Stellen gleichmäßiger Konvergenz sich zu endlich oder unendlich vielen einfach zusammenhängenden Gebieten $\mathfrak{G}_1, \mathfrak{G}_2, \dots$ zusammensetzen müssen. Denn C ist stets abgeschlossen und ein Häufungspunkt von Γ muß, sofern er nicht auf C liegt, offenbar wiederum zu Γ gehören. Wäre ferner \mathfrak{E} nicht zusammenhängend, sondern in zwei abgeschlossene Punkt-mengen ohne gemeinsamen Punkt zerlegbar, so müßte die eine derselben C enthalten, die andere Γ_0 also in \mathfrak{G} liegen. Ist dann d der Abstand der beiden Punkt-mengen und schließt man Γ_0 in einen oder mehrere einfach geschlossene Streckenzüge S ein, deren jeder Punkt von Γ_0 höchstens den Abstand $\frac{1}{2}d$ besitzt, so liegt auch jedes S noch in \mathfrak{G} , und da die Folge der $f_n(z)$ in der Umgebung jedes Punktes von S gleichmäßig konvergiert, so müßte sie nach dem erwähnten Runge'schen Satze⁸⁾ auch im ganzen Innern von S gleichmäßig konvergieren.

⁶⁾ Unter einem „Gebiet“ \mathfrak{G} wird im folgenden stets eine zusammenhängende offene Punktmenge der Ebene verstanden, d. h. eine Punktmenge, deren jeder Punkt ein innerer Punkt ist und bei der je zwei Punkte derselben durch einen ihr angehörigen endlichen Streckenzug verbunden werden können. Unter einem „Bereich“ ist die aus einem Gebiet \mathfrak{G} durch Hinzufügung der Randpunkte entstehende Menge $\bar{\mathfrak{G}}$ zu verstehen.

⁷⁾ Bei Verwendung des Begriffs der „punktweise gleichmäßigen Konvergenz“ (s. insbes. Pringsheim, Münch. Ber. 1919, S. 419) an Stelle des im Text benutzten würde sich an der Gesamtheit der bezeichneten Stellen nichts ändern (s. Pringsheim, a. a. O. S. 428—429). Wie ferner aus dem Vitalischen Satze (vgl. ¹⁰⁾) hervorgeht, sind die Stellen gleichmäßiger Konvergenz identisch mit denjenigen, in deren Umgebung die $f_n(z)$ gemeinsam beschränkt sind.

⁸⁾ C. Runge, Acta Math. 6 (1885), S. 247.

Eine zweite notwendige Bedingung ergibt sich aus dem bereits in der Einleitung erwähnten, von Herrn Osgood herrührenden Satze, wonach es in \mathfrak{G} (und somit auch in jedem Teilgebiet von \mathfrak{G}) stets Bereiche geben muß, in denen die gegebene Folge *gleichmäßig* konvergiert.

Die beiden genannten Bedingungen zusammen besagen also, daß *die Gebiete gleichmäßiger Konvergenz \mathfrak{G}_ν ($\nu = 1, 2, 3, \dots$) einfach zusammenhängend und in \mathfrak{G} überall dicht sein müssen*, oder, was damit gleichbedeutend ist, daß \mathfrak{C} *stets ein linienhaftes* (d. h. keine Kreisfläche enthaltendes, also nirgends dichtes) *Kontinuum sein muß*^{8a}).

In dem speziellen Fall, daß die Gebiete \mathfrak{G}_ν nur in *endlicher* Zahl vorhanden sind, erweisen sich, wie sich in § 2 ergeben wird, die bisher aufgestellten notwendigen Bedingungen zugleich als hinreichend; d. h. zerfällt das Gebiet \mathfrak{G} nach Wegnahme einer nirgends dichten Teilmenge Γ in eine *endliche* Anzahl von einfach zusammenhängenden Gebieten \mathfrak{G}_ν , so gibt es stets eine Folge von analytischen, in \mathfrak{G} regulären Funktionen, welche in jedem der Gebiete \mathfrak{G}_ν gleichmäßig⁹), hingegen in der Umgebung jedes Punktes von Γ ungleichmäßig konvergiert.

Sind hingegen die Gebiete \mathfrak{G}_ν in *unendlicher* Menge vorhanden, so sind die Bedingungen im allgemeinen nicht mehr hinreichend, wie zunächst an dem folgenden Beispiel gezeigt werden soll.

\mathfrak{G} sei ein Quadrat und werde in 9 kongruente Teilquadrate zerlegt. Das mittlere derselben (exkl. Rand) sei \mathfrak{G}_1 . Jedes der übrigen 8 Teilquadrate werde wieder in je 9 kleinere Quadrate zerlegt, jeweils das mittlere derselben (exkl. Rand) mit \mathfrak{G}_ν ($\nu = 2, 3, \dots, 9$) bezeichnet und dies Verfahren ohne Ende fortgesetzt. Die so bestimmten Gebiete \mathfrak{G}_ν erfüllen alsdann die obigen Bedingungen und \mathfrak{C} besteht daher aus den Rändern der \mathfrak{G}_ν und deren Häufungsstellen.

Dasjenige zu \mathfrak{G}_ν ($\nu = 1, 2, \dots$) konzentrische Quadrat, dessen Seiten dreimal so groß sind, werde (einschl. Rand) mit \mathfrak{S}_ν bezeichnet. Unter diesen Quadraten \mathfrak{S}_ν werde irgend eines, etwa \mathfrak{S}_a , ausgewählt, welches (nebst Rand) ganz im Innern von \mathfrak{G} liegt, und die Fläche des dem zugehörigen Quadrat \mathfrak{G}_a umschriebenen Kreises mit \mathfrak{k} bezeichnet.

Würde nun eine Folge $f_n(z)$ der verlangten Art existieren, so könnten die Funktionen $f_n(z)$ in \mathfrak{k} nicht gleichmäßig beschränkt sein, da sonst nach

^{8a}) Vgl. auch P. Montel, Ann. Ec. Norm. (3) 24 (1907), S. 320–323; Leçons³), S. 118–119.

⁹) Eine Folge „konvergiert in einem Gebiet \mathfrak{G} gleichmäßig“ soll stets bedeuten, daß jeder Punkt von \mathfrak{G} für die Folge eine „Stelle gleichmäßiger Konvergenz“ sei (oder, was damit gleichbedeutend, daß die Folge in jedem zu \mathfrak{G} gehörigen „Bereich“ gleichmäßig konvergiere).

dem Vitalischen Satz¹⁰⁾ die Folge im Innern von \mathfrak{f} gleichmäßig konvergieren würde. Es müßte also für eine gewisse Zahl $n = n_1$ und einen gewissen Punkt $z = z_1$ der Kreisfläche \mathfrak{f} $|f_{n_1}(z_1)| > 1$ gelten und somit (auf Grund des nachfolgenden Hilfssatzes I) einen in \mathfrak{S}_α gelegenen Streifen¹¹⁾ \mathfrak{S}_1 geben, welcher z_1 mit dem Rand R von \mathfrak{S}_α verbindet und in welchem durchweg $|f_{n_1}(z)| > 1$ ist. Da nun kein Gebiet \mathfrak{G}_ν existiert, welches \mathfrak{f} mit R verbindet, so müssen innerhalb \mathfrak{S}_1 notwendig Punkte von Γ liegen, und da jede (beliebig kleine) Umgebung eines Punktes von Γ stets Gebiete \mathfrak{S}_ν enthält, so muß es auch (unendlich viele) Gebiete \mathfrak{S}_ν geben, welche in \mathfrak{S}_1 liegen. Wird eines derselben mit \mathfrak{S}_β bezeichnet, so führt die Wiederholung der obigen Schlußweise zu einer Beziehung $|f_{n_2}(z)| > 2$ ($n_2 > n_1$), welche für alle z eines gewissen in \mathfrak{S}_β gelegenen Streifens \mathfrak{S}_2 gilt. Man gelangt, indem man dies Verfahren ohne Ende fortsetzt, zu einer unendlichen Folge von Streifen $\mathfrak{S}_1, \mathfrak{S}_2, \dots$, so zwar, daß für alle z des k -ten Streifens die Ungleichung $|f_{n_k}(z)| > k$ ($n_k > n_{k-1}$) gilt. Da nun jedes der Quadrate $\mathfrak{S}_\alpha, \mathfrak{S}_\beta, \dots$ in dem vorhergehenden enthalten ist, so existiert ein Punkt $z = z_0$ von \mathfrak{G} , welcher jedem dieser Quadrate und daher auch jedem der Streifen $\mathfrak{S}_1, \mathfrak{S}_2, \dots$ angehört; in diesem Punkt könnte aber wegen $|f_{n_k}(z_0)| > k$ ($k = 1, 2, \dots$) unmöglich Konvergenz stattfinden.

Der hierbei benutzte Hilfssatz, welcher auch bei den nachfolgenden Betrachtungen eine wichtige Rolle spielt, lautet:

Hilfssatz I. Ist die analytische Funktion $f(z)$ im Gebiet \mathfrak{G} eindeutig definiert und regulär und ist für einen gewissen Punkt $z = z_0$ von \mathfrak{G} $|f(z_0)| > \sigma$, so gibt es ein Teilgebiet g von \mathfrak{G} , welches z_0 enthält, sich bis an den Rand von \mathfrak{G} erstreckt (so daß \mathfrak{G} und g mindestens einen gemeinsamen Randpunkt besitzen) und in dem durchweg $|f(z)| > \sigma$ gilt.

Beweis. Diejenigen Punkte von \mathfrak{G} , für welche $|f(z)| > \sigma$ ist, bilden eine offene Menge. Von den Gebieten, aus denen diese letztere besteht, enthält eines, g , den Punkt z_0 . Würde nun g sich nicht bis an den Rand von \mathfrak{G} erstrecken, sondern der Rand c von g ganz im Innern von \mathfrak{G} liegen, so müßte längs c und daher nach einem bekannten Satze auch im Innern von c durchweg die Ungleichung $|f(z)| \leq \sigma$ bestehen.

Angefügt sei noch der folgende

Hilfssatz II. Ist $f_n(z)$ ($n = 1, 2, \dots$) eine Folge von analytischen, in \mathfrak{G} regulären Funktionen, welche in \mathfrak{G} durchweg konvergiert, so existiert eine Folge von ganzen rationalen Funktionen $g_n(z)$, welche in \mathfrak{G} gegen dieselbe Grenzfunktion konvergiert, und zwar an genau den-

¹⁰⁾ Näheres sowie Literatur siehe Encykl. d. math. Wiss. II C 4 (Bieberbach), Nr. 58.

¹¹⁾ Vgl. Fußnote ¹²⁾.

selben Stellen von \mathcal{G} gleichmäßig bzw. ungleichmäßig konvergiert wie die Folge der $f_n(z)$.

Beweis. Es seien g_1, g_2, \dots Polygonflächen, welche sämtlich zu \mathcal{G} gehören, von denen jede im Innern der folgenden liegt und welche so beschaffen sind, daß jeder Punkt von \mathcal{G} einer dieser Polygonflächen (und somit auch allen folgenden) angehört. Nach dem Rungeschen Approximationssatz¹²⁾ kann man die ganze rationale Funktion $g_n(z)$ so bestimmen, daß in g_n $|f_n(z) - g_n(z)| < \frac{1}{n}$ gilt. Die Folge der $g_n(z)$ besitzt dann die behaupteten Eigenschaften; denn beide Folgen haben in \mathcal{G} dieselbe Grenzfunktion und in jedem zu \mathcal{G} gehörigen Bereich, in welchem die eine Folge gleichmäßig konvergiert, tut es auch die andere.

§ 2.

Notwendige und hinreichende Bedingung.

Es bedeute wieder \mathcal{G} ein in der z -Ebene gelegenes, beschränktes, einfach zusammenhängendes Gebiet, $\mathcal{G}_1, \mathcal{G}_2, \dots$ einfach zusammenhängende Teilgebiete von \mathcal{G} , welche in \mathcal{G} überall dicht liegen; die Gesamtheit der keinem \mathcal{G}_n angehörenden Punkte von \mathcal{G} heiße Γ . Irgendein Kreis, der \mathcal{G} (einschließlich Rand) in seinem Innern enthält, werde mit K , irgendeine Folge von Punkten, welche auf Γ überall dicht liegen, mit ζ_1, ζ_2, \dots bezeichnet.

Die *notwendige und hinreichende Bedingung* dafür, daß eine Folge von analytischen, in \mathcal{G} regulären Funktionen $f_n(z)$ existiert, welche in jedem der Gebiete \mathcal{G}_n gleichmäßig⁹⁾, hingegen in der Umgebung jedes Punktes von Γ ungleichmäßig konvergiert, kann alsdann in der folgenden Weise ausgesprochen werden:

Es muß ein System von Streifen $\mathcal{E}_1, \mathcal{E}_2, \dots$ existieren, welche aus dem Gebiet \mathcal{G} herausführen¹³⁾ und ferner so beschaffen sind, daß jede beliebig kleine Umgebung eines Punktes von Γ stets von unendlich vielen dieser Streifen getroffen, daß hingegen jeder einzelne Punkt von Γ selbst höchstens von einer endlichen Anzahl derselben überdeckt wird. („Bedingung A“.)

Hierbei kann die geforderte Eigenschaft des Streifensystems, daß

¹²⁾ C. Runge, Acta Math. 6 (1885), S. 229. — S. a. Encykl. d. math. Wiss. II B 1 (Osgood), S. 89.

¹³⁾ „Streifen“ soll stets ein von einem einfachen Polygon begrenztes Gebiet bezeichnen. Ein Streifen \mathcal{E} „führt aus \mathcal{G} heraus“ soll bedeuten, daß \mathcal{E} sowohl mit \mathcal{G} als auch mit der Kreislinie K Punkte gemein hat (und zwar, wie der Einfachheit halber stets angenommen werden kann, mit K einen einzigen Bogen).

„jede beliebig kleine Umgebung eines Punktes von Γ stets von unendlich vielen der Streifen getroffen werde“, auch durch die ersetzt werden, daß

„der Abstand d_n ¹⁴⁾ des Streifens \mathfrak{S}_n vom Punkte ζ_n mit wachsendem n gegen null konvergiere“.

Denn in einem Streifensystem, welches der ersten Formulierung entspricht, gibt es sicher ein Teilsystem, das der letzteren entspricht; umgekehrt ist klar, daß jedes der zweiten Formulierung entsprechende Streifensystem stets auch der ersten entspricht.

Zusatz. Der Ausspruch behält (wie weiter unten nachgewiesen wird) seine Gültigkeit unverändert bei, wenn gefordert wird, daß *jeder Punkt von \mathfrak{G}* (statt von Γ) höchstens von einer endlichen Anzahl jener Streifen überdeckt werde.

Sind die Gebiete \mathfrak{G}_ν nur in endlicher Anzahl vorhanden, so ist die Bedingung A immer erfüllt¹⁵⁾. Bei dem in § 1 behandelten Beispiel hingegen ist sie nicht erfüllt, wovon man sich auch auf direktem Wege mit Hilfe von Betrachtungen, die mit den in § 1 an dieses Beispiel geknüpften nahezu übereinstimmen, leicht überzeugen kann.

Um zunächst die *Notwendigkeit* der Bedingung A nachzuweisen, nehmen wir an, es existiere eine Funktionenfolge $f_\nu(z)$, welche die verlangten Konvergenzeigenschaften besitzt, wobei wegen des Hilfssatzes II die $f_\nu(z)$ als ganze rationale Funktionen angenommen werden können. Um jeden der Punkte ζ_n werde ein Kreis \mathfrak{k}_n beschrieben, der noch in \mathfrak{G} liegt, und zwar so, daß die Radien r_n dieser Kreise mit wachsendem n gegen 0 konvergieren.

Da die Folge der $f_\nu(z)$ in der Kreisfläche \mathfrak{k}_n nicht gleichmäßig konvergiert, so können nach dem Vitalischen Satz die $f_\nu(z)$ in \mathfrak{k}_n nicht gemeinsam beschränkt sein. Es muß also für eine gewisse Zahl $\nu = \nu_n$ und einen gewissen Punkt $z = z_n$ von \mathfrak{k}_n $|f_{\nu_n}(z_n)| > n$ gelten, wobei noch $\nu_n > \nu_{n-1}$ angenommen werden kann. Nach dem Hilfssatz I gilt dann die Beziehung $|f_{\nu_n}(z)| > n$ auch für alle Punkte eines gewissen Streifens \mathfrak{S}_n , welcher sowohl z_n als auch einen Bogen¹⁶⁾ der Kreislinie K enthält. Der Abstand d_n des Punktes ζ_n vom Streifen \mathfrak{S}_n ist dann kleiner als r_n und

¹⁴⁾ Liegt ζ_n im Innern oder auf dem Rande von \mathfrak{S}_n , so ist $d_n = 0$ zu setzen.

¹⁵⁾ Dies folgt am einfachsten aus dem Umstand, daß die „Bedingung B“ (siehe S. 221–222) in diesem Falle selbstverständlich erfüllt ist (da die Anzahl der Streifen eine endliche ist) und daß (siehe S. 223) das Erfülltsein der Bedingung B das der Bedingung A nach sich zieht. (Um sich *direkt* von der Richtigkeit zu überzeugen, kann man ähnlich wie in Fußnote ²¹⁾ verfahren.)

¹⁶⁾ Da es ja freisteht, bei der Benutzung des Hilfssatzes I statt K einen K einschließenden Kreis K' zugrunde zu legen.

somit $\lim d_n = 0$. Andererseits kann jeder Punkt von \mathfrak{G} nur einer endlichen Anzahl von Streifen \mathfrak{S}_n angehören, da ja andernfalls die Folge der $f_n(z)$ für diesen Punkt divergieren würde.

Bei dem nunmehr zu führenden Nachweise, daß die Bedingung A auch *hinreichend* ist, werde von einem Streifensystem $\mathfrak{S}_1, \mathfrak{S}_2, \dots$ ausgegangen, welches die in der Bedingung A (zweite Formulierung) ausgesprochenen Eigenschaften besitzt. Der Abstand des Punktes ζ_n vom Streifen \mathfrak{S}_n werde wie bisher mit d_n bezeichnet.

Das Streifensystem $\mathfrak{S}_1, \mathfrak{S}_2, \dots$ soll nun zunächst (vgl. obigen „Zusatz“) durch ein anderes, $\mathfrak{T}_1, \mathfrak{T}_2, \dots$ der gleichen Art ersetzt werden, welches jedoch die weitergehende Eigenschaft besitzt, daß (nicht nur jeder Punkt von Γ , sondern) jeder Punkt von \mathfrak{G} und überdies jeder Teilbereich irgendeines \mathfrak{G}_ν höchstens von einer endlichen Anzahl der Streifen getroffen wird. Zu diesem Zwecke mögen zunächst die Gebiete \mathfrak{G}_ν von innen durch Polygonflächen $g_{\nu 1}, g_{\nu 2}, \dots$ approximiert werden (wie beim Beweis des Hilfssatzes II). Falls nun \mathfrak{S}_n keines der Gebiete $g_{1n}, g_{2n}, \dots, g_{nn}$ trifft, möge \mathfrak{T}_n mit \mathfrak{S}_n identisch sein. Trifft hingegen \mathfrak{S}_n eines oder mehrere dieser Gebiete, etwa $g_{\alpha n}, g_{\beta n}, \dots$, so möge \mathfrak{T}_n aus \mathfrak{S}_n dadurch entstehen, daß man von dem durch Vereinigung von \mathfrak{S}_n mit $g_{\alpha, n+1}, g_{\beta, n+1}, \dots$ entstehenden Gebiet die Flächen $g_{\alpha n}, g_{\beta n}, \dots$ wieder fortnimmt (wodurch der Zusammenhang dieses Gebietes keinesfalls unterbrochen wird) und sodann von dem so entstandenen mehrfach zusammenhängenden Gebiet eine einfach zusammenhängende Polygonfläche \mathfrak{T}_n beibehält, deren Abstand von ζ_n wieder höchstens gleich d_n ist und die mit der Kreislinie K einen einzigen Bogen gemein hat. Das so entstehende neue Streifensystem \mathfrak{T}_n besitzt dann offenbar alle behaupteten Eigenschaften.

Wird nun dasjenige Flächenstück, welches die Kreisfläche K mit \mathfrak{T}_n gemein hat, aus K herausgenommen, so bleibt ein einfach zusammenhängender Bereich \mathfrak{R}_n übrig. Irgendein innerer Punkt von \mathfrak{T}_n , dessen Abstand von ζ_n kleiner als $2d_n$ ist, sei ζ'_n . Man kann alsdann nach dem Rungeschen Satz über die gleichzeitige Approximation zweier (oder mehrerer) zu getrennten einfach zusammenhängenden Bereichen gehöriger analytischer Funktionen durch Polynome¹⁷⁾ ein Polynom $f_n(z)$ derart bestimmen, daß

$$1. \quad |f_n(z)| < \frac{1}{n} \quad \text{in } \mathfrak{R}_n,$$

$$2. \quad |f_n(z) - 1| < \frac{1}{2} \quad \text{für } z = \zeta'_n.$$

¹⁷⁾ C. Runge, Acta Math. 6 (1885), S. 229–244. S. a. Encykl. d. math. Wiss. II C 4 (Bieberbach), S. 491.

Die Folge der so bestimmten Funktionen $f_n(z)$ besitzt dann die sämtlichen verlangten Eigenschaften (und zwar konvergiert sie in \mathfrak{G} durchweg gegen 0).

Ist nämlich $z = z_\nu$ irgendein Punkt von \mathfrak{G}_ν , und daher auch von $\mathfrak{g}_{\nu n}$ etwa für $n \geq p$ (wobei die ganze Zahl p überdies größer als ν angenommen werden möge) und beschreibt man um z_ν einen noch in $\mathfrak{g}_{\nu p}$ gelegenen Kreis \mathfrak{k} , so wird \mathfrak{k} von keinem der Streifen $\mathfrak{X}_p, \mathfrak{X}_{p+1}, \dots$ getroffen; die Beziehung $|f_n(z)| < \frac{1}{n}$ gilt daher in \mathfrak{k} für alle $n \geq p$ und die Folge konvergiert somit in der Umgebung von $z = z_\nu$ gleichmäßig gegen null.

Ist andererseits $z = z_0$ irgendein Punkt von Γ , so gibt es höchstens eine endliche Anzahl von Streifen \mathfrak{X}_ν , welche z_0 überdecken, und man hat somit wiederum $\lim f_n(z_0) = 0$. Die Konvergenz ist jedoch in der Umgebung der Stelle z_0 keine gleichmäßige; denn beschreibt man um z_0 einen Kreis mit beliebig kleinem Radius, so liegen in diesem Kreis unendlich viele Punkte ζ_ν , sowie auch unendlich viele Punkte ζ'_ν . Da nun für alle n $|f_n(\zeta'_\nu)| > \frac{1}{2}$ ist, so kann die Konvergenz in diesem Kreise keine gleichmäßige sein.

Kapitel II.

Grenzfunktion vorgeschrieben.

§ 3.

Aufstellung einer hinreichenden Bedingung.

Es bedeute wieder \mathfrak{G} ein beschränktes, einfach zusammenhängendes Gebiet der z -Ebene, K einen dasselbe umschließenden Kreis und $\mathfrak{G}_1, \mathfrak{G}_2, \dots$ endlich oder unendlich viele Teilgebiete von \mathfrak{G} , welche die sämtlichen in §§ 1 und 2 aufgestellten Bedingungen erfüllen. Es existiert dann eine Folge von analytischen, in \mathfrak{G} regulären Funktionen, welche in jedem der Gebiete \mathfrak{G}_ν gleichmäßig, in der Umgebung jedes anderen Punktes von \mathfrak{G} ungleichmäßig konvergiert. Wird nun die Grenzfunktion dieser Folge mit $\varphi(z)$ bezeichnet, so entsteht die Frage, in welchem Umfange und unter welchen Bedingungen (neben der gleichmäßigen und ungleichmäßigen Konvergenz der Folge) auch noch diese Grenzfunktion $\varphi(z)$ willkürlich vorgeschrieben werden kann.

Wird der Bequemlichkeit halber die Grenzfunktion der Folge im Gebiet \mathfrak{G}_n auch mit $\varphi_n(z)$ ($n = 1, 2, \dots$), andererseits für die Punkte von Γ (d. h. für alle übrigen Punkte von \mathfrak{G}) mit $\varphi_0(z)$ bezeichnet, so ist klar, daß für jeden Wert von n $\varphi_n(z)$ eine in \mathfrak{G}_n reguläre analytische Funktion sein muß. Aber auch $\varphi_0(z)$ kann nicht eine Funktion ganz willkürlicher Art sein, sondern muß jedenfalls als Grenz-

funktion einer Folge von Polynomen darstellbar sein. Wenn daher bei den folgenden Betrachtungen von der „vorgeschriebenen Grenzfunktion $\varphi(z)$ “ die Rede ist, so soll dies immer so zu verstehen sein, daß für jedes der Gebiete \mathfrak{G}_n eine daselbst reguläre Funktion $\varphi_n(z)$ vorgeschrieben und ferner eine Funktion $\varphi_0(z)$ als Grenzfunktion einer (mindestens) für alle Punkte von Γ konvergenten Folge von Polynomen $P_\nu(z)$ gegeben sei.

Es wird sich im folgenden zeigen, daß, auch wenn alle Bedingungen des vorigen Kapitels erfüllt sind, die Grenzfunktion $\varphi(z)$ doch nicht (in dem eben angegebenen Sinne) beliebig vorgeschrieben werden kann.

Zunächst seien zwei spezielle Fälle ins Auge gefaßt.

Wenn erstens $\varphi_0(z) = 0$, $\varphi_n(z) = 0$ ($n = 1, 2, \dots$) vorgeschrieben wird, so geht aus § 2 hervor, daß (wenn die bisherigen Bedingungen erfüllt sind) stets eine Funktionenfolge der verlangten Art existiert. Bei dieser speziellen Wahl der vorgeschriebenen Grenzfunktion erweisen sich also die bisherigen Bedingungen zugleich als hinreichend.

Zweitens sei $\varphi_0(z) = 0$, $\varphi_n(z) = 1$ ($n = 1, 2, \dots$) vorgeschrieben. Dann ergibt sich durch folgende Überlegung, daß zu den bisherigen Bedingungen noch eine weitere hinzukommt.

Sei $f_\nu(z)$ ($\nu = 1, 2, \dots$) eine Funktionenfolge der verlangten Art, wobei nach dem Hilfssatz II die $f_\nu(z)$ als Polynome angenommen werden können. Ist z_n ein beliebig gewählter Punkt von \mathfrak{G}_n , so muß es eine Zahl ν_n ($> \nu_{n-1}$) geben, so daß $|f_{\nu_n}(z)| > \frac{1}{2}$ ist und infolgedessen nach Hilfssatz I einen Streifen \mathfrak{S}_n , welcher das Innere von \mathfrak{G}_n mit dem Außenkreis K verbindet und in welchem durchweg $|f_{\nu_n}(z)| > \frac{1}{2}$ gilt. Es kann dann aber keinen Punkt von Γ geben, der von unendlich vielen dieser Streifen überdeckt wird, da ja andernfalls die Funktionenfolge in diesem Punkt nicht gegen 0 konvergieren könnte.

Als neue¹⁸⁾ Bedingung hat sich demnach ergeben, daß hier ein System von Streifen $\mathfrak{S}_1, \mathfrak{S}_2, \dots$ existieren muß, so beschaffen, daß \mathfrak{S}_n das Innere von \mathfrak{G}_n mit der Kreislinie K verbindet, daß aber jeder einzelne Punkt von Γ höchstens von einer endlichen Anzahl der Streifen überdeckt wird.

Diese Bedingung ist aber, wie nunmehr gezeigt werden soll, zugleich eine *hinreichende*, und zwar gilt *letzteres* nicht nur bei der soeben besprochenen speziellen, sondern auch bei jeder beliebigen Wahl der vorgeschriebenen Grenzfunktion $\varphi(z)$; m. a. W. es gilt der folgende Satz:

Liegen die einfach zusammenhängenden Teilgebiete \mathfrak{G}_n im Gebiet \mathfrak{G} überall dicht und gibt es ein System von Streifen $\mathfrak{S}_1, \mathfrak{S}_2, \dots$, so beschaffen, daß \mathfrak{S}_n das Innere von \mathfrak{G}_n mit dem Außenkreis K verbindet,

¹⁸⁾ Daß diese Bedingung nicht etwa schon in der Bedingung A des § 2 enthalten ist, wohl aber umgekehrt diese umfaßt, ergibt sich aus den weiteren Betrachtungen.

daß aber jeder Punkt von Γ höchstens von einer endlichen Anzahl dieser Streifen überdeckt wird („Bedingung B“), so existiert stets eine Folge von Polynomen, welche in \mathfrak{G} gegen eine (im angegebenen Sinne) willkürlich vorgeschriebene Grenzfunktion $\varphi(z)$ konvergiert, und zwar in jedem der Gebiete \mathfrak{G}_n gleichmäßig, in der Umgebung jedes Punktes von Γ ungleichmäßig.

Bemerkung. In ganz ähnlicher Weise wie bei der Bedingung A können auch bei der Bedingung B, ohne deren Tragweite abzuändern, die Worte „jeder Punkt von Γ höchstens ...“ ersetzt werden durch „jeder Punkt von \mathfrak{G} höchstens ...“

Ist die Anzahl der Gebiete \mathfrak{G}_n endlich, so ist die Bedingung B selbstverständlich stets erfüllt.

Beweis. Auch hier kann wieder angenommen werden, daß jeder Streifen \mathfrak{S}_n mit K einen einzigen Bogen gemein hat. Wie auf S. 219 mögen nun zunächst die Gebiete \mathfrak{G}_v von innen durch Polygonflächen $g_{v,1}, g_{v,2}, \dots$ approximiert werden. Dabei kann angenommen werden, daß $g_{v,1}$ (und daher auch $g_{v,2}, \dots$) Punkte mit \mathfrak{S}_v gemein habe; denn sollte dies nicht von vornherein der Fall sein, so kann es durch Fortlassung einer endlichen Anzahl von Polygonflächen (oder auch durch Fortsetzung des Streifens \mathfrak{S}_v nach innen) erreicht werden.

Der Streifen \mathfrak{S}_v werde durch einen Längsschnitt in zwei Streifen $\mathfrak{f}_{v,1}$ und \mathfrak{S}'_v zerlegt (welche ebenfalls das Innere von $g_{v,1}$ mit K verbinden), hierauf \mathfrak{S}'_v in zwei Streifen $\mathfrak{f}_{v,2}$ und \mathfrak{S}''_v von der nämlichen Art usf., so daß \mathfrak{S}_v schließlich in unendlich viele Streifen $\mathfrak{f}_{v,1}, \mathfrak{f}_{v,2}, \dots$ derselben Art zerlegt erscheint.

Nimmt man von der Kreisfläche K die n Polygonflächen $g_{1,n}, g_{2,n}, \dots, g_{n,n}$ nebst den ihnen anhängenden Streifen $\mathfrak{f}_{1,n}, \mathfrak{f}_{2,n}, \dots, \mathfrak{f}_{n,n}$ fort, so ist¹⁹⁾ die übrigbleibende Fläche \mathfrak{H}_n entweder selbst ein einfach zusammenhängender Bereich oder besteht aus einer endlichen Anzahl von solchen, die im allgemeinen voneinander getrennt liegen²⁰⁾. Wiederum getrennt von \mathfrak{H}_n und voneinander liegen die Polygonflächen $g_{1,n-1}, g_{2,n-1}, \dots, g_{n,n-1}$. Auf Grund des obigen Rungeschen Satzes (S. 219) kann man daher ein Polynom $f_n(z)$ bilden, das sich in jedem der Bereiche $g_{v,n-1}$ ($v = 1, 2, \dots, n$) von $\varphi_v(z)$ und überdies in \mathfrak{H}_n von $P_n(z)$ um weniger als $\frac{1}{n}$ unterscheidet (wobei $\varphi_v(z)$ und $P_n(z)$ die auf S. 220—221 angegebene Bedeutung haben).

¹⁹⁾ Die fortgenommenen Flächenstücke einschließlich ihrer Ränder bilden zusammen mit der Kreislinie K ein Kontinuum. Ein solches bestimmt aber stets lauter einfach zusammenhängende Komplementärgebiete.

²⁰⁾ Sollte letzteres nicht der Fall sein, so läßt es sich durch geeignete Einschnürung eines oder mehrerer der Streifen $\mathfrak{f}_{1,n}, \dots, \mathfrak{f}_{n,n}$ leicht erreichen.

Die Folge der $f_n(z)$ ($n = 2, 3, \dots$) besitzt alsdann (abgesehen von einer noch zu erörternden Einschränkung) die behaupteten Eigenschaften.

Denn zunächst hat man in einem beliebigen Gebiet $g_{\nu, \varrho}$:

$$|f_n(z) - \varphi_{\nu}(z)| < \frac{1}{n} \quad \text{für } n \begin{cases} > \varrho \\ \geq \nu \end{cases}$$

und daher konvergiert die Folge in $g_{\nu, \varrho}$ und somit schließlich auch in \mathfrak{G}_{ν} gleichmäßig gegen $\varphi_{\nu}(z)$.

Ist andererseits z_0 ein beliebiger Punkt von Γ , so gibt es nur eine endliche Anzahl von Streifen \mathfrak{S}_{ν} und daher auch von Streifen $\{_{\nu, \varrho}$ ($\nu, \varrho = 1, 2, \dots$), welche z_0 überdecken; infolgedessen gehört z_0 jeder der Flächen \mathfrak{R}_n an, wenn nur n hinreichend groß, etwa $n > n_0$ gewählt wird. Es ist somit

$$|f_n(z_0) - P_n(z_0)| < \frac{1}{n} \quad \text{für } n > n_0$$

und daher $\lim f_n(z_0) = \lim P_n(z_0) = \varphi_0(z_0)$, wie es verlangt war.

Hingegen besteht zunächst noch keinerlei Sicherheit darüber, ob die Konvergenz für die Umgebung jedes Punktes von Γ eine ungleichmäßige ist (was jedoch für einen Punkt von Γ selbstverständlich nur dann zweifelhaft sein kann, wenn die vorgeschriebene Grenzfunktion $\varphi(z)$ in der ganzen Umgebung dieses Punktes regulär sein sollte). Sollte dies noch nicht (oder nicht durchweg) der Fall sein, so kann es stets durch eine nachträgliche Abänderung der Funktionenfolge $f_n(z)$ in folgender Weise erreicht werden.

Nimmt man zunächst für einen Augenblick an, daß die Vorschrift $\varphi_0(z) = 0$, $\varphi_n(z) = 1$ ($n = 1, 2, \dots$) laute, so ist das Gewünschte sicher schon erreicht. Daraus geht aber gemäß § 2 hervor, daß die Gebiete \mathfrak{G}_{ν} auch der Bedingung A genügen, m. a. W. daß *die Bedingung A in der Bedingung B enthalten ist*²¹⁾.

Nachdem dies festgestellt ist, können wir gemäß § 2 eine Folge von Polynomen $g_n(z)$ konstruieren, welche in \mathfrak{G} durchweg gegen 0 konvergiert und zwar in jedem der Gebiete \mathfrak{G}_{ν} gleichmäßig, für die Umgebung jedes Punktes von Γ gleichmäßig.

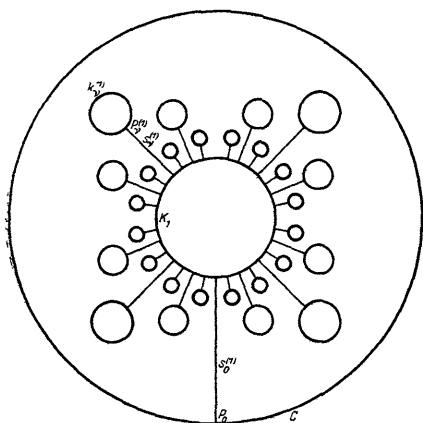
²¹⁾ Hiervon kann man sich auch auf rein geometrischem Wege überzeugen. Seien nämlich ζ_1, ζ_2, \dots Punkte, welche auf den Rändern der sämtlichen Gebiete \mathfrak{G}_{ν} (und daher auch auf Γ) überall dicht liegen. Diejenigen Punkte ζ_i , welche dem Rande von \mathfrak{G}_{ν} angehören, seien mit $\zeta_{\nu 1}, \zeta_{\nu 2}, \dots$ bezeichnet (wobei jedoch Punkte ζ_i , welche den Rändern mehrerer \mathfrak{G}_{ν} angehören, nur bei *einem* [etwa dem ersten] dieser Gebiete aufgezählt werden mögen). Erfüllt nun das Streifensystem \mathfrak{S}_n die Bedingung B, so werde wie im Text jeder der Streifen \mathfrak{S}_{ν} in unendlich viele Teilstreifen $\{_{\nu 1}, \{_{\nu 2}, \dots$ zerlegt und sodann jeder der Streifen $\{_{\nu \alpha}$ im Innern von \mathfrak{G}_{ν} auf beliebigem Wege bis in die vorgeschriebene Nähe des Punktes $\zeta_{\nu \alpha}$ fortgesetzt. Das so entstehende Streifensystem $\{\{_{\nu \alpha}^*\}$ erfüllt dann die Bedingung A (im Sinne der zweiten Formulierung).

Die Funktionenfolge $f_n(z) + \gamma g_n(z)$ konvergiert dann, gleichgültig wie die Konstante γ gewählt wird, ebenfalls gegen die ursprünglich vorgeschriebene Grenzfunktion $\varphi(z)$, und zwar gleichmäßig in jedem \mathfrak{G}_ν . Die Konstante γ läßt sich aber stets so wählen, daß die Konvergenz für die Umgebung jedes Punktes von Γ eine ungleichmäßige wird.

Sind nämlich ζ_i ($i = 1, 2, \dots$) Punkte, die auf Γ überall dicht liegen, so gibt es offenbar höchstens eine Zahl γ_i derart, daß die Folge $f_n(z) + \gamma_i g_n(z)$ ($n = 1, 2, \dots$) in der Umgebung des Punktes ζ_i gleichmäßig konvergiert. Wählt man also $\gamma = \gamma_0$ so, daß es von jeder derartigen Zahl γ_i verschieden ist, so ist die Konvergenz der Folge $f_n(z) + \gamma_0 g_n(z)$ in der Umgebung jeder Stelle ζ_i und daher auch jedes beliebigen Punktes von Γ eine ungleichmäßige.

Zum Nachweise dafür, daß die Bedingungen A und B nicht etwa äquivalent sind, möge noch ein Beispiel angefügt werden, bei welchem die Bedingung A erfüllt, die Bedingung B hingegen nicht erfüllt ist.

Unser Gebiet \mathfrak{G} sei von einem Kreis C begrenzt. Ins Innere von \mathfrak{G} legen wir einen (konzentrischen) Kreis K_1 [„Zentral-Kreis“] und, um ihn herum, eine unendliche Folge von immer kleiner werdenden Kreisen k_ν



Figur 51.

($\nu = 1, 2, 3, \dots$) [„Planeten-Kreise“, die sich gegen sämtliche Punkte von K_1 häufen²²⁾ [vgl. Figur]. K_1 sei mit C und ferner jeder Kreis k_ν ($\nu = 1, 2, 3, \dots$) mit K_1 durch je eine (die übrigen Teile der Figur nicht treffende) Strecke $s_\nu^{(1)}$ bzw. $s_\nu^{(1)}$ ($\nu = 1, 2, 3, \dots$) [„Stiel“] verbunden. Die so entstandene, innerhalb \mathfrak{G} gelegene Figur sei mit \mathfrak{F}_1 bezeichnet. Das von K_1 umschlossene Gebiet soll nun selbst schon eines der Gebiete \mathfrak{G}_ν darstellen und bleibt dementsprechend von der folgenden Konstruktion unberührt; dagegen werden

die „Planeten-Kreise“ noch weiter behandelt: Es werde nämlich jeder Planeten-Kreis durch eine zu \mathfrak{F}_1 ähnliche Figur ersetzt. (D. h. mit jedem $k_\nu^{(1)}$ werde so verfahren wie vorhin mit C und zwar soll dabei der $k_\nu^{(1)}$ und $s_\nu^{(1)}$ gemeinsame Punkt $P_\nu^{(1)}$ dem zu C und $s_0^{(1)}$ gehörenden Punkt P_0

²²⁾ Z. B. denke man K_1 mit einer unendlichen Folge von konzentrischen, gegen K_1 konvergierenden Ringen umgeben und setze in den n -ten dieser Ringe — gleichmäßig verteilt — μ_n „Planeten-Kreise“, so daß μ_n gleichzeitig mit n ins Unendliche wächst.

entsprechen; nach Herstellung der zu \mathfrak{F}_1 ähnlichen Figur werde schließlich die Kreislinie $k_v^{(1)}$ [abgesehen vom Punkt $P_v^{(1)}$] weggelassen. Die so durch das Ersetzen aller Planeten-Kreise aus \mathfrak{F}_1 entstandene Figur sei mit \mathfrak{F}_2 bezeichnet. Wieder beranden die neu entstandenen Zentral-Kreise bereits endgültige Gebiete \mathfrak{G}_v , während wieder die neu entstandenen Planeten-Kreise von \mathfrak{F}_2 durch zu \mathfrak{F}_1 ähnliche Figuren ersetzt werden: Figur \mathfrak{F}_3 . Dieses Verfahren wird unendlich oft iteriert. Die dabei jeweils entstehenden Zentral-Kreis-Flächen stellen insgesamt die Gebiete $\mathfrak{G}_2, \mathfrak{G}_3, \dots, \mathfrak{G}_n, \dots$ dar; die Zentral-Kreis-Linien, die „Stiele“ $s_v^{(2)}$ und die Häufungspunkte von beiden sollen Γ bilden. Dazu kommt dann noch das von Γ und C begrenzte (einfach zusammenhängende) Restgebiet von \mathfrak{G} , welches mit \mathfrak{G}_1 bezeichnet werde²³⁾.

Daß die so entstandene Figur \mathfrak{F}_* der Bedingung A genügt, ist klar; denn man kann an Γ von außen her ein die Bedingung A befriedigendes, ausschließlich in \mathfrak{G}_1 verlaufendes Streifensystem legen, ohne daß überhaupt ein Punkt von Γ überdeckt, geschweige denn unendlich oft überdeckt wird. — Andererseits kann die Bedingung B nicht erfüllt sein. Denn wäre sie befriedigt, so müßte ins Innere von \mathfrak{G}_2 ein Streifen \mathfrak{S}_2 eindringen. Da aber gegen jeden Randpunkt von \mathfrak{G}_2 unendlich viele, beliebig kleine Gebiete \mathfrak{G}_v sich häufen, so müßte \mathfrak{S}_2 ein Gebiet \mathfrak{G}_{n_2} ganz enthalten. Ins Innere von \mathfrak{G}_{n_2} müßte dann ein Streifen \mathfrak{S}_{n_2} eindringen und aus gleichem Grunde innerhalb des Durchschnitts von $(\mathfrak{S}_2, \mathfrak{S}_{n_2})$ ein Gebiet \mathfrak{G}_{n_4} enthalten sein; in dessen Inneres müßte wieder ein Streifen \mathfrak{S}_{n_4} eindringen, usw. Auf diese Weise kämen wir zu einem Punkt von Γ , der von unendlich vielen Streifen $\mathfrak{S}_2, \mathfrak{S}_{n_2}, \mathfrak{S}_{n_4} \dots$ überdeckt würde.

Wir bemerken noch: Bei diesem Beispiel hört offenbar auch die Bedingung A auf, erfüllt zu sein, wenn man in jedes \mathfrak{G}_v ($v \geq 2$) einen (vom Rande nach innen gehenden) Stachel — z. B. einen Radius — hineinsetzt; oder auch, wenn man jedes \mathfrak{G}_v ($v \geq 2$) durch zwei (mittels eines Durchmessers getrennte) halbkreisförmige Teilgebiete ersetzt.

§ 4.

Notwendige und hinreichende Bedingung.

Ist (neben der gleichmäßigen und ungleichmäßigen Konvergenz) die Grenzfunktion $\varphi(z)$ der Funktionenfolge vorgeschrieben, so ist, wie aus den bisherigen Untersuchungen hervorgeht, für die Existenz einer Funktionenfolge

²³⁾ Bezeichnet man also in der Figur \mathfrak{F}_1 das durch Fortnahme des Kreisbereichs K_1 und der Kreisgebiete $k_v^{(1)}$ sowie der zugehörigen „Stiele“ aus \mathfrak{G} entstehende Flächenstück mit \mathfrak{g}_1 , so entsteht \mathfrak{G}_1 durch Vereinigung von \mathfrak{g}_1 mit den sämtlichen zu \mathfrak{g}_1 ähnlichen Flächenstücken der Figuren $\mathfrak{F}_2, \mathfrak{F}_3, \dots$

der verlangten Art die Bedingung A unter allen Umständen eine notwendige und die Bedingung B unter allen Umständen eine hinreichende. Es gibt ferner, wie festgestellt wurde, spezielle Fälle (z. B. $\varphi(z) \equiv 0$), in denen die Bedingung A nicht nur notwendig, sondern auch hinreichend und daher B nicht notwendig ist; andererseits solche (z. B. $\varphi_0(z) = 0$, $\varphi_1(z) = \varphi_2(z) = \dots = 1$), in denen die Bedingung B nicht nur hinreichend, sondern auch notwendig und daher A nicht hinreichend ist. Die Aufgabe ist hiernach, auch für den allgemeinsten Fall eine Bedingung aufzustellen, welche sowohl notwendig als auch hinreichend ist.

Um diese Bedingung in möglichst einfacher Form auszusprechen, sei noch folgende Ausdrucksweise eingeführt. Existiert eine Folge von Polynomen, welche in $\overline{\mathfrak{G}}_n$ (d. h. \mathfrak{G}_n nebst Rand) gemeinsam beschränkt sind und daselbst gegen $\varphi(z)$ konvergieren, so soll dafür kurz gesagt werden, in \mathfrak{G}_n finde „Vitalischer Anschluß“ (nämlich der Funktionswerte $\varphi_n(z)$ an die Randwerte $\varphi_0(z)$) statt. Ist die Grenzfunktion $\varphi(z)$ vorgeschrieben, so steht damit für jedes Gebiet \mathfrak{G}_n fest, ob dort Vitalischer Anschluß stattfindet oder nicht.

Die *notwendige und hinreichende Bedingung* lautet alsdann folgendermaßen:

I. Die Bedingung A muß erfüllt sein.

II. Diejenigen Gebiete \mathfrak{G}_n , bei denen kein Vitalischer Anschluß stattfindet, müssen der Bedingung B genügen²⁴⁾.

Zusatz. Wird die Forderung, daß in der Umgebung jedes Punktes von Γ ungleichmäßige Konvergenz stattfinden solle, beiseite gelassen²⁵⁾, so fällt die Bedingung I fort. Wird hingegen die Forderung bezüglich der gleichmäßigen Konvergenz fortgelassen, so hat dies auf die genannten Bedingungen keinen Einfluß.

Beweis. α) Die genannten Bedingungen sind *notwendig*. Ist nämlich $f_n(z)$ ($n = 1, 2, \dots$) eine Folge von Polynomen, welche in \mathfrak{G} gegen $\varphi(z)$ konvergiert (wobei zunächst über gleichmäßige oder ungleichmäßige

²⁴⁾ D. h. es muß diejenige Bedingung erfüllt sein, die sich aus der Bedingung B ergibt, indem man darin die \mathfrak{G}_n durch die Gesamtheit der Gebiete ohne Vitalischen Anschluß ersetzt (Γ jedoch ungeändert läßt). — Sind die Gebiete, in denen kein Vitalischer Anschluß stattfindet, nur in endlicher Anzahl vorhanden, so ist die Bedingung II stets erfüllt. Beispiel: Wird für die auf S. 224–225 konstruierte Gebiets-einteilung $\varphi_0(z) = 1$, $\varphi_1(z) = 0$, $\varphi_2(z) = \varphi_3(z) = \dots = 1$ vorgeschrieben, so sind die Bedingungen erfüllt und es existiert daher eine Folge von Polynomen der verlangten Art.

²⁵⁾ Diese Forderung ist jedoch, falls die $\varphi_n(z)$ lauter *verschiedene* analytische Funktionen sind und in der Umgebung jedes Punktes von Γ Punkte mindestens *zweier* Gebiete \mathfrak{G}_n liegen, offenbar implizite in den übrigen enthalten; daraus folgt dann, daß in diesem Falle die Bedingung I in der Bedingung II enthalten ist.

Konvergenz nichts angenommen werde²⁶⁾), so sei unterschieden zwischen solchen Gebieten \mathfrak{G}_v , in denen die $f_n(z)$ gemeinsam beschränkt sind, und solchen, bei denen dies nicht der Fall ist. Bei den Gebieten \mathfrak{G}_v der ersteren Art findet Vitalischer Anschluß statt. Die Gebiete \mathfrak{G}_v der zweiten Art (unter denen sich ebenfalls noch solche mit Vitalischem Anschluß befinden können) müssen aber der Bedingung B genügen, da sich (ähnlich wie auf S. 221) Streifen \mathfrak{S}_n konstruieren lassen, welche das Innere je eines solchen Gebiets mit dem Außenkreis K verbinden und in welchen durchweg $|f_n(z)| > n$ gilt. Hieraus folgt aber, daß die Bedingung II erfüllt sein muß.

Wird nun weiter angenommen, daß für die Folge der $f_n(z)$ (mindestens) jeder Punkt von Γ eine Stelle ungleichmäßiger Konvergenz sei, so folgt aus § 2, daß auch die Bedingung A zunächst für die Gesamtheit der Stellen ungleichmäßiger Konvergenz und somit auch für die Teilgesamtheit Γ erfüllt sein muß.

β) Die genannten Bedingungen sind *hinreichend*.

Zunächst werde nur von der Bedingung II vorausgesetzt, daß sie erfüllt sei. Ist dann $\varphi_0(z) = \lim P_n(z)$ die für Γ vorgeschriebene Grenzfunktion, so werde unterschieden zwischen solchen Gebieten \mathfrak{G}_v , in denen die $P_n(z)$ gemeinsam beschränkt sind, und solchen, bei denen dies nicht der Fall ist.

In denjenigen Gebieten \mathfrak{G}_v der ersteren Art, bei denen überdies Vitalischer Anschluß stattzufinden hat, konvergiert die Folge der $P_n(z)$ selbst schon gegen die vorgeschriebene Grenzfunktion $\varphi_v(z)$, und zwar *gleichmäßig*²⁷⁾. Denn ist etwa \mathfrak{G}_α ein Gebiet dieser Art, so ist $\varphi(z)$ in $\overline{\mathfrak{G}}_\alpha$ (d. h. \mathfrak{G}_α mit Rand) durch eine beschränkte Folge von Polynomen $Q_n(z)$ darstellbar. Es stellt also $P_n(z) - Q_n(z)$ ($n = 1, 2, \dots$) eine in $\overline{\mathfrak{G}}_\alpha$ beschränkte Folge von Polynomen dar, welche längs des Randes von $\overline{\mathfrak{G}}_\alpha$ gegen 0 konvergiert. Dann muß sie aber auch im ganzen Innern gegen 0 konvergieren²⁸⁾.

Was zweitens diejenigen Gebiete \mathfrak{G}_v betrifft, für welche die $P_n(z)$ gemeinsam beschränkt sind, jedoch Vitalischer Anschluß nicht stattfindet, so genügen diese nach Voraussetzung der Bedingung B.

²⁶⁾ Doch sollen die Teilgebiete \mathfrak{G}_v nach wie vor als gegeben gelten, und zwar (wie immer) einfach zusammenhängend, in \mathfrak{G} überall dicht und so, daß $\varphi(z)$ mindestens für alle z jedes \mathfrak{G}_v regulär.

²⁷⁾ Letzteres nach dem Vitalischen Satze.

²⁸⁾ Näheres hierüber siehe F. Hartogs, Über die Grenzfunktionen beschränkter Folgen von analytischen Funktionen, Math. Annalen 98 (1927), S. 168. Hiernach ist die Schlußweise des Textes nur dann sichergestellt, wenn der Rand von \mathfrak{G}_α einen Jordanschen Bogen l als Bestandteil enthält und mindestens zwei Punkte von l (als Randpunkte von \mathfrak{G}_α) von innen erreichbar sind.

Es bleiben noch drittens diejenigen Gebiete \mathcal{G}_ν zu betrachten, in denen die $P_n(z)$ nicht gemeinsam beschränkt sind. Auch diese genügen der Bedingung B, wie sich folgendermaßen ergibt. Für jedes einzelne (z. B. das n -te) dieser Gebiete kann man (nach Hilfssatz I) einen Streifen konstruieren, der das Innere dieses Gebietes mit dem Außenkreis K verbindet und in welchem durchweg $|P_{\nu_n}(z)| > n$ ist ($\nu_n > \nu_{n-1}$); jeder Punkt von Γ kann aber dann offenbar höchstens von einer endlichen Anzahl dieser Streifen überdeckt werden.

Damit ist zunächst festgestellt, daß die Gesamtheit derjenigen Gebiete \mathcal{G}_ν , in denen die Folge der $P_n(z)$ nicht selbst schon gegen die vorgeschriebene Grenzfunktion konvergiert (und zwar gleichmäßig), der Bedingung B genügt. Diese Gebiete mögen nunmehr mit $\mathfrak{S}_1, \mathfrak{S}_2, \dots$ bezeichnet werden, die ihnen (im Sinne der Bedingung B) zugeordneten Streifen mit $\mathfrak{S}_1, \mathfrak{S}_2, \dots$.

Die Konstruktion einer Folge von Polynomen $f_n(z)$, welche gegen die vorgeschriebene Grenzfunktion $\varphi(z)$ konvergieren, geschieht nunmehr ganz ähnlich wie beim Beweise in § 3. Die in derselben Weise wie dort durch Unterteilung (Zerfaserung) der Streifen $\mathfrak{S}_1, \mathfrak{S}_2, \dots$ sich ergebenden Streifen $\mathfrak{I}_{\nu, \rho}$ ($\nu, \rho = 1, 2, \dots$) mögen jedoch zunächst (ohne ihre bisherigen Eigenschaften einzubüßen) so abgeändert werden, daß jeder Teilbereich eines Gebietes \mathcal{G}_ν höchstens von einer endlichen Anzahl derselben getroffen wird; dies kann leicht dadurch geschehen, daß man die Streifen $\mathfrak{I}_{\nu, \rho}$ zuerst in eine einfach unendliche Folge $\mathfrak{s}_1, \mathfrak{s}_2, \dots$ umordnet und mit diesen letzteren ebenso verfährt wie auf S. 219 mit den Streifen $\mathfrak{S}_1, \mathfrak{S}_2, \dots$ ²⁹⁾. Konstruiert man dann eine Funktionenfolge $f_n(z)$ auf dieselbe Weise wie beim Beweise in § 3, nur mit dem Unterschied, daß an Stelle der dortigen Gebiete \mathcal{G}_ν hier die Gebiete \mathfrak{S}_ν treten, so konvergiert die so entstehende Funktionenfolge nicht nur in jedem Gebiete \mathfrak{S}_ν , sondern auch in jedem der übrigen Gebiete \mathcal{G}_ν (z. B. in \mathcal{G}_α) gleichmäßig gegen $\varphi(z)$, da ja in einem beliebigen Teilbereich von \mathcal{G}_α für „fast alle“ Werte von n

$$|f_n(z) - P_n(z)| < \frac{1}{n}$$

gilt. Für jeden Punkt von Γ konvergiert die Folge gegen $\varphi_0(z)$.

Wird des weiteren die Bedingung I als erfüllt vorausgesetzt, so kann man genau wie in § 3 aus der soeben gebildeten Funktionenfolge $f_n(z)$ eine neue, $f_n(z) + \gamma_0 g_n(z)$ herstellen, welche überdies in der Umgebung jedes Punktes von Γ ungleichmäßig konvergiert. Damit sind aber die sämtlichen Behauptungen bewiesen.

²⁹⁾ Natürlich tritt dabei an Stelle der dortigen Bedingung, daß \mathfrak{I}_n von ζ_n einen hinreichend kleinen Abstand haben solle, hier die, daß der abgeänderte Streifen \mathfrak{I}_n wieder das Innere des betreffenden Gebietes \mathcal{G}_ν mit dem Kreise K verbinden müsse.

Hinzugefügt sei, daß die Bedingung II noch auf eine etwas andere Form gebracht werden kann, wenn, was jetzt geschehen soll, die Ränder C_ν der Gebiete \mathfrak{G}_ν als *rektifizierbar* vorausgesetzt werden. Gilt alsdann für alle z eines Gebietes \mathfrak{G}_n die Formel:

$$\varphi_n(z) = \frac{1}{2\pi i} \int_{C_n} \frac{\varphi_0(\zeta) d\zeta}{\zeta - z},$$

so werde dafür kurz gesagt, in \mathfrak{G}_n finde „*Cauchyscher Anschluß*“ statt.

Findet in \mathfrak{G}_n Vitalischer Anschluß statt, so findet dort auch Cauchyscher Anschluß statt³⁰⁾. Findet umgekehrt in \mathfrak{G}_n Cauchyscher Anschluß statt, so folgt daraus im allgemeinen nicht, daß dort auch Vitalischer Anschluß stattfinden müsse, wohl aber dann, wenn $\varphi_0(z)$ auf C_n durch eine Folge gemeinsam beschränkter Polynome dargestellt werden kann³⁰⁾.

Dies vorausgeschickt, behalten die an die obigen Bedingungen I und II geknüpften Aussagen *ihre Gültigkeit unverändert bei, wenn in der Bedingung II die Worte „Vitalischer Anschluß“ ersetzt werden durch „Cauchyscher Anschluß“*.

Ist nämlich \mathfrak{G}_a ein Gebiet, in welchem zwar Cauchyscher, aber nicht Vitalischer Anschluß stattfindet, so können die Polynome $P_n(z)$ auf dem Rand C_a dieses Gebiets nicht gemeinsam beschränkt sein; die Gesamtheit der Gebiete \mathfrak{G}_a dieser Art genügt also nach S. 228 (oben) der Bedingung B. Hieraus ergibt sich aber unmittelbar die Behauptung.

§ 5.

Charakteristische Eigenschaften der Grenzfunktion.

Werden die Gebiete gleichmäßiger Konvergenz wie bisher mit \mathfrak{G}_ν bezeichnet, so ist die Grenzfunktion $\varphi(z)$ mindestens für alle z jedes dieser Gebiete regulär. Diejenigen Gebiete \mathfrak{F}_μ ($\mu = 1, 2, \dots$), in denen $\varphi(z)$ regulär ist, müssen also (soweit sie nicht mit je einem der \mathfrak{G}_ν übereinstimmen) durch Vereinigung je einer endlichen oder unendlichen Menge von Gebieten \mathfrak{G}_ν unter Hinzufügung gewisser Teile von Γ entstehen; im Gegensatz zu den \mathfrak{G}_ν können die \mathfrak{F}_μ daher mehrfach (auch unendlich-fach) zusammenhängend sein. Die Gesamtheit der zu keinem \mathfrak{F}_μ gehörigen Punkte von \mathfrak{G} werde mit Γ_0 bezeichnet, so daß Γ_0 eine Teilmenge von Γ darstellt.

Auf Grund der bisherigen Betrachtungen kann alsdann Folgendes nachgewiesen werden:

³⁰⁾ Siehe z. B. P. Montel, *Leçons sur les séries de polynomes à une variable complexe*, Paris 1910, p. 19. Vgl. auch F. Hartogs²⁸⁾, insbes. S. 170 u. 172.

1. Die Gebiete \mathfrak{F}_μ liegen in \mathfrak{G} überall dicht.
2. $\varphi(z)$ ist für Γ_0 durch eine (mindestens für alle Punkte von Γ_0 konvergierende) Folge von Polynomen $P_n(z)$ darstellbar.
3. Γ_0 erfüllt die Bedingung A³¹⁾.
4. Diejenigen Gebiete \mathfrak{F}_μ , bei denen kein Vitalischer Anschluß stattfindet, erfüllen die Bedingung B³¹⁾;

und umgekehrt gilt, daß jede Funktion $\varphi(z)$, welche die Eigenschaften 2. und 4. besitzt, in \mathfrak{G} als Grenzfunktion einer Folge von regulären analytischen Funktionen (oder auch von Polynomen) dargestellt werden kann.

Hieraus geht zugleich hervor, daß das Stattfinden der Eigenschaften 1. und 3. eine unmittelbare Folge des Stattfindens der beiden übrigen ist (und zwar ist, wie ohne weiteres ersichtlich, 1. schon eine Folge von 2. allein).

Nachweis für den ersten Teil der Behauptung. Daß die \mathfrak{F}_μ in \mathfrak{G} überall dicht liegen, folgt unmittelbar aus der entsprechenden Eigenschaft der Gebiete \mathfrak{G}_ν ; ebenso folgt 3. unmittelbar aus der entsprechenden Eigenschaft von Γ . 2. ist selbstverständlich, da ja $\varphi(z)$ auch in \mathfrak{G} durch eine Folge von Polynomen darstellbar ist. Der Nachweis für 4. ist wörtlich wie der Beweis α) auf S. 226—227 zu führen, wobei nur \mathfrak{G}_ν überall durch \mathfrak{F}_μ zu ersetzen und der zweite Absatz ganz fortzulassen ist.

Der Beweis für die Umkehrung stimmt im wesentlichen mit dem Beweis β) (S. 227) überein. Zunächst wird genau wie dort (wobei bloß wieder \mathfrak{G}_ν durchweg durch \mathfrak{F}_μ zu ersetzen ist) nachgewiesen, daß die Gesamtheit derjenigen Gebiete \mathfrak{F}_μ , in denen die Folge der $P_n(z)$ nicht selbst schon gegen $\varphi_n(z)$ konvergiert, der Bedingung B genügt. Diese Gebiete mögen wieder mit $\mathfrak{H}_1, \mathfrak{H}_2, \dots$ bezeichnet werden, die ihnen (im Sinne der Bedingung B) zugeordneten Streifen mit $\mathfrak{S}_1, \mathfrak{S}_2, \dots$, die durch Unterteilung der letzteren entstehenden wieder mit $\mathfrak{I}_{\nu\varrho}$ ($\nu, \varrho = 1, 2, \dots$)³²⁾.

³¹⁾ „Bedingung A“ und „Bedingung B“ sind im folgenden stets so zu verstehen, daß gegenüber der bisherigen Formulierung derselben \mathfrak{G}_ν durch \mathfrak{F}_ν und Γ durch Γ_0 zu ersetzen ist.

³²⁾ Eine Ablenkung der Streifen $\mathfrak{I}_{\nu\varrho}$ ist hier (im Gegensatz zu S. 228) nicht erforderlich, weil hier gleichmäßige Konvergenz für die \mathfrak{F}_μ weder vorgeschrieben ist noch durchweg erzielt werden könnte. Hingegen mögen die Streifen \mathfrak{S}_ν von vornherein so gewählt sein, daß (nicht nur jeder Punkt von Γ_0 , sondern) jeder Punkt von \mathfrak{G} höchstens von einer endlichen Anzahl derselben überdeckt wird. Daß letzteres stets möglich ist (d. h. daß auch dann, wenn die \mathfrak{F}_ν nicht einfach zusammenhängend sind, beide Formen der Bedingung B äquivalent sind) sieht man folgendermaßen: Die \mathfrak{F}_ν mögen durch (endlichfach zusammenhängende) Polygonflächen $f_{\nu\alpha}$ von innen approximiert und letztere durch Kanäle in einfach zusammenhängende, $f_{\nu\mu}^*$, verwandelt werden, jedoch so, daß jeder Punkt von \mathfrak{F}_ν höchstens endlich vielen Kanälen angehört. Der Streifen \mathfrak{S}_ν werde alsdann im Innern von $\mathfrak{F}_1, \mathfrak{F}_2, \dots, \mathfrak{F}_\nu$ (soweit letztere überhaupt getroffen werden) so abgelenkt, daß er $f_{1\nu}^*, f_{2\nu}^*, \dots, f_{\nu\nu}^*$ nicht trifft.

Jedes der Gebiete \mathfrak{G}_ν möge nun von innen durch (ein- oder mehrfach, jedenfalls nur endlichfach zusammenhängende) Polygonflächen $\mathfrak{h}_{\nu,1}, \mathfrak{h}_{\nu,2}, \dots$ approximiert werden, von denen jede ganz im Innern der folgenden liegt. Diejenigen der Flächen $\mathfrak{h}_{\nu,1}, \mathfrak{h}_{\nu,2}, \dots$, welche mehrfach zusammenhängend sind, mögen durch Fortnehmen einer endlichen Anzahl von Kanälen³³⁾ in einfach zusammenhängende $\mathfrak{h}_{\nu,1}^*, \mathfrak{h}_{\nu,2}^*, \dots$ übergeführt werden, jedoch so, daß jeder Punkt von \mathfrak{G}_ν höchstens einer endlichen Anzahl dieser Kanäle angehört. Nimmt man alsdann von der Kreisfläche K die n (ein- oder mehrfach zusammenhängenden) Polygonflächen $\mathfrak{h}_{1,n}, \mathfrak{h}_{2,n}, \dots, \mathfrak{h}_{n,n}$ nebst den ihnen anhängenden Streifen $\mathfrak{f}_{1,n}, \mathfrak{f}_{2,n}, \dots, \mathfrak{f}_{n,n}$ hinweg, so ist¹⁹⁾ die übrigbleibende Fläche \mathfrak{R}_n entweder selbst ein einfach zusammenhängender Bereich oder besteht aus einer endlichen Anzahl von solchen, die im allgemeinen voneinander getrennt liegen²⁰⁾. Wiederum getrennt von \mathfrak{R}_n und voneinander liegen die einfach zusammenhängenden Flächen $\mathfrak{h}_{1,n-1}^*, \mathfrak{h}_{2,n-1}^*, \dots, \mathfrak{h}_{n,n-1}^*$. Infolgedessen kann man ein Polynom $f_n(z)$ herstellen, das sich in jedem der Bereiche $\mathfrak{h}_{\nu,n-1}^*$ ($\nu = 1, 2, \dots, n$) von $\varphi(z)$ und überdies in \mathfrak{R}_n von $P_n(z)$ um weniger als $\frac{1}{n}$ unterscheidet.

Die Folge der $f_n(z)$ konvergiert alsdann in \mathfrak{G} überall gegen $\varphi(z)$.

Ist nämlich $z = z_\alpha$ ein Punkt von \mathfrak{G}_α , und daher auch von $\mathfrak{h}_{\alpha\beta}^*$ etwa für alle $\beta \geq \beta_0$, so gilt

$$|f_n(z_\alpha) - \varphi(z_\alpha)| < \frac{1}{n} \quad \text{für } n \begin{cases} \geq \alpha \\ > \beta_0 \end{cases}$$

und daher ist $\lim f_n(z_\alpha) = \varphi(z_\alpha)$.

Ist andererseits $z = z'$ ein beliebiger Punkt eines der übrigen Gebiete \mathfrak{G}_μ oder auch von Γ_0 , so gibt es höchstens eine endliche Anzahl von Streifen \mathfrak{S}_ν und daher auch von Streifen $\mathfrak{f}_{\nu\rho}$ ($\nu, \rho = 1, 2, \dots$), welche z' überdecken. Infolgedessen gehört z' „fast“ jeder der Flächen \mathfrak{R}_n an, woraus dann

$$\lim f_n(z') = \lim P_n(z') = \varphi(z')$$

folgt.

§ 6.

Eigenschaften der Grenzfunktion auf Γ_0 .

Es bleibt nun noch übrig, die Tragweite der „Bedingung 2.“ des vorigen Paragraphen näher zu untersuchen, d. h. die Frage, welche Eigenschaften eine (zum mindesten für alle Punkte z von Γ_0 definierte) Funktion $\varphi(z)$ besitzen muß, damit sie auf Γ_0 durch eine Folge von Polynomen von z darstellbar sei.

³³⁾ „Kanal“ soll (ebenso wie „Streifen“) ein von einem einfachen Polygon begrenztes Gebiet bezeichnen.

Wird hierbei zunächst angenommen, daß Γ_0 als Vereinigungsmenge von (höchstens) abzählbar unendlich vielen Jordanschen Kurvenbögen darstellbar sei, so läßt sich die gestellte Frage dahin beantworten, daß $\varphi(z)$ auf Γ_0 *lediglich der* (selbstverständlichen) *Einschränkung genügen müsse daselbst eine Funktion der Klasse 0 oder 1 im Baireschen Sinne*³⁴⁾ (also in Γ_0 auf jeder perfekten Menge höchstens punktweise unstetig) *zu sein*. Der Nachweis hierfür ist in den folgenden allgemeineren Betrachtungen enthalten.

Um zu umfassenderen hierauf bezüglichen Aussagen zu gelangen, seien folgende Benennungen eingeführt.

Eine beschränkte, abgeschlossene Punktmenge der z -Ebene, auf welcher x und daher auch³⁵⁾ jede stetige Funktion von x und y durch Polynome von $z (= x + iy)$ *gleichmäßig* approximiert werden kann, werde als „ α -Menge“ bezeichnet. Eine solche kann, wie leicht zu sehen³⁶⁾, niemals die Ebene zerlegen. Jede abgeschlossene Teilmenge einer α -Menge ist offenbar wieder eine α -Menge.

Es läßt sich alsdann nachweisen, daß *jede ungeschlossene Jordansche Kurve eine α -Menge darstellt*³⁷⁾.

³⁴⁾ Vgl. Encykl. d. math. Wiss. II C 9 c (Fréchet-Rosenthal), Nr. 53 u. 58.

³⁵⁾ Ist x auf der beschränkten und abgeschlossenen Punktmenge M durch Polynome von z gleichmäßig approximierbar, so ist es auch $y (= ix - iz)$, in folgedessen auch jede ganze rationale Funktion von x und y und daher schließlich auch jede auf M stetige Funktion von x und y , da eine solche durch ganze rationale Funktionen von x und y gleichmäßig approximiert werden kann.

³⁶⁾ Existierte ein beschränktes Gebiet g , dessen Rand R aus Punkten einer α -Menge A besteht, so würde eine auf A gleichmäßig gegen x konvergierende Folge von Polynomen von z auch in $(g + R)$ durchweg gleichmäßig konvergieren und daher eine in g reguläre Funktion $\psi(z)$ darstellen, welche sich an die Randwerte x stetig anschließt. Infolgedessen müßte der imaginäre Teil von $\psi(z)$ identisch Null, also $\psi(z)$ gleich einer reellen Konstanten und daher x selbst auf R konstant sein.

³⁷⁾ Siehe F. Hartogs²⁸⁾, § 1. — Ein weiteres Beispiel für ein linienhaftes Kontinuum, das eine α -Menge darstellt, ist folgendes.

Werden in das Rechteck $ABCD$ abzählbar unendlich viele, voneinander getrennte Rechtecke \mathfrak{R}_ν ($\nu = 1, 2, \dots$) einbeschrieben, von denen je ein Paar von Gegenseiten auf AD und BC liegen, deren Ecken von A, B, C, D verschieden sind und welche $ABCD$ überall dicht erfüllen, so bilden die (zu AB parallelen) Längsseiten, die Haupt- (d. h. von links oben nach rechts unten verlaufenden) diagonalen d_ν und die Häufungslinien dieser Rechtecke zusammen ein linienhaftes Kontinuum M , welches, wie nun gezeigt werden soll, bei geeigneter Wahl jener Rechtecke eine α -Menge darstellt.

Die $k-1$ ersten Rechtecke seien beliebig gewählt (k beliebige positive Zahl). Zeichnet man die $k-1$ zugehörigen Hauptdiagonalen und fügt die Hauptdiagonale jeder der k Lücken hinzu, so bilden die Längsseiten der $k-1$ Rechtecke zusammen mit den $k-1$ Hauptdiagonalen derselben, den k Lückendiagonalen und den Randlinien

Allgemeiner heie eine beliebige Punktmenge der z -Ebene, auf welcher jede Funktion von x und y der Klasse 0 oder 1 durch Polynome von z (gleich- oder ungleichmig) approximiert werden kann, eine „ β -Menge“.

Jede α -Menge ist zugleich eine β -Menge, da ja jede Funktion von x und y der Klasse 0 oder 1 definitionsgem durch eine Folge von stetigen Funktionen von x und y dargestellt werden kann. Ein Beispiel fur eine β -Menge, die nicht α -Menge ist, liefert jede geschlossene Jordansche Kurve³⁸⁾. Jede abgeschlossene Teilmenge einer β -Menge ist wieder eine β -Menge³⁹⁾.

Es gelten alsdann die beiden folgenden Stze:

1. *Jede beschrnkte, abgeschlossene Menge, deren smtliche „Stucke“⁴⁰⁾ α -Mengen sind, ist selbst wiederum eine α -Menge⁴¹⁾.*

AB und CD einen zickzackformigen Streckenzug s_1 und es gibt daher ein Polynom $P_1(z)$, derart, da $|x - P_1(z)| < \varepsilon_1$ auf s_1 und infolgedessen auch in einem gewissen, s_1 enthaltenden Gebiet g_1 ist.

Man kann nun in jede der Lucken je ein neues Rechteck ($\mathfrak{R}_k, \mathfrak{R}_{k+1}, \dots, \mathfrak{R}_{2k-1}$) hineinsetzen, derart, da die Lngsseiten derselben, ihre Hauptdiagonalen und die $2k$ noch verbleibenden Lucken ganz in g_1 liegen. (Die Hauptdiagonalen der neuen Rechtecke fallen dabei nherungsweise mit je einer der urspringlichen Luckendiagonalen zusammen.) Alsdann liegt (unabhngig von der Wahl aller spteren Rechtecke $\mathfrak{R}_{2k}, \mathfrak{R}_{2k+1}, \dots$) M vollstndig in g_1 und somit gilt die obige Ungleichung fur die ganze Menge M .

Behufs Bildung von $P_2(z)$ verfhrt man nunmehr mit den bisherigen $2k-1$ Rechtecken ebenso wie urspringlich mit den $k-1$ ersten Rechtecken, d. h. bildet aus ihnen einen gewissen zickzackformigen Streckenzug s_2 und sodann ein Polynom $P_2(z)$, derart, da $|x - P_2(z)| < \varepsilon_2$ fur alle Punkte von s_2 ist; bei geeigneter Wahl der $2k$ folgenden Rechtecke gilt dann diese Ungleichung wieder fur alle Punkte von M . In dieser Weise wird fortgefahren, wobei $\lim \varepsilon_n = 0$ angenommen wird.

³⁸⁾ Da eine geschlossene Jordansche Kurve j eine β -Menge ist, geht aus dem nachfolgenden Satz 2 hervor. Da andererseits j die Ebene zerlegt, so kann sie keine α -Menge sein.

³⁹⁾ Die Richtigkeit ergibt sich sofort auf Grund des Satzes, da jede auf einer abgeschlossenen Menge definierte Funktion der Klasse 0 oder 1 zu einer in der ganzen Ebene definierten Funktion der Klasse 0 oder 1 erweitert werden kann. (S. Encykl. d. math. Wiss. II C 90 (Frchet-Rosenthal), S. 1176.)

⁴⁰⁾ Unter einem „Stuck“ einer abgeschlossenen Menge M ist ein Teilkontinuum (oder einzelner Punkt) von M zu verstehen, das nicht in einem anderen Teilkontinuum von M enthalten ist. (Vgl. Encykl. d. math. Wiss. II C 9a (Zoretti-Rosenthal), S. 902.) Selbstverstndlich kann M aus mehr als abzhlbar unendlich vielen „Stucken“ bestehen.

⁴¹⁾ Insbesondere gilt dies also von jeder beschrnkten abgeschlossenen Menge, deren smtliche Stucke ungeschlossene Jordansche Kurven oder Punkte sind. — Aus dem Satze geht noch hervor, da eine α -Menge positiven Inhalt haben kann, und zwar, wenn sie z. B. in einem Rechteck liegt, einen solchen, der dem des Rechtecks beliebig nahekommt.

⁴²⁾ Zunchst kann man nmlich leicht endlich viele, getrennt liegende, einfache Polygonflchen konstruieren, die zusammen M in ihrem Innern enthalten und deren

(Fortsetzung der Funote ⁴²⁾ auf nchster Seite.)

Beweis. Sei M eine solche Menge, σ ein Stück derselben und ε eine beliebig kleine positive Zahl. Da σ eine α -Menge ist, so existiert ein Polynom $P_\sigma(z)$, derart, daß auf σ

$$(1) \quad |x - P_\sigma(z)| < \varepsilon,$$

und diese Ungleichung bleibt dann auch noch für alle z eines gewissen, σ enthaltenden Gebietes g gültig.

Man kann nun weiter, da σ die Ebene nicht zerlegt, eine in g gelegene und σ enthaltende einfach zusammenhängende Polygonfläche p_σ konstruieren, deren Rand M nicht trifft⁴²⁾; die vorige Ungleichung gilt dann insbesondere für alle z dieser Polygonfläche.

Denkt man sich in dieser Weise jedem Stück σ der Menge M eine Polygonfläche p_σ und eine Funktion $P_\sigma(z)$ zugeordnet, so läßt sich nach dem Heine-Borelschen Überdeckungssatz unter den Polygonflächen p_σ eine endliche Anzahl $\mathfrak{P}_1, \mathfrak{P}_2, \dots, \mathfrak{P}_n$ derart auswählen, daß jeder Punkt von M mindestens einer derselben angehört und daß von diesen n Polygonflächen keine vollständig in einer anderen liegt. Da die Ränder dieser Polygone M nicht treffen, so kann man dieses System von n Polygonen aber noch durch ein anderes von den gleichen Eigenschaften ersetzen, bei dem die einzelnen Polygone überdies sämtlich getrennt liegen. Die n Polygonränder bestimmen nämlich in der Ebene endlich viele beschränkte, einfach zusammenhängende Gebiete (Zellen). Wird nun innerhalb jeder Zelle, welche einen oder mehr Punkte von M enthält, ein Polygon gezeichnet, das alle zu M gehörigen Punkte dieser Zelle einschließt, so liegen die Polygonflächen $\mathfrak{Q}_1, \mathfrak{Q}_2, \dots, \mathfrak{Q}_r$ dieser letzteren Art alle getrennt, und da jede derselben in (mindestens) einer der Polygonflächen \mathfrak{P}_ν ($\nu = 1, 2, \dots, n$) vollständig enthalten ist, dieser letzteren aber eine gewisse Funktion $P_\sigma(z)$ zugeordnet war, so gilt für jede der Polygonflächen \mathfrak{Q}_ρ ($\rho = 1, 2, \dots, r$) je eine Ungleichung von der Form (1).

Man kann nun nach dem Rungeschen Approximationssatz ein Polynom $P(z)$ konstruieren, das sich in jeder der Flächen $\mathfrak{Q}_1, \mathfrak{Q}_2, \dots, \mathfrak{Q}_r$ von der zugeordneten Funktion $P_\sigma(z)$ um weniger als ε unterscheidet. Es gilt dann für alle Punkte z eines jeden \mathfrak{Q}_ρ und somit schließlich für

sämtliche Punkte von M einen Abstand kleiner als eine gegebene Zahl δ_n besitzen. (Vgl. z. B. B. v. Kerékjártó, Vorlesungen über Topologie I, Berlin 1923, S. 49—51.) Die σ enthaltende Polygonfläche sei p_n . Ist nun $\lim \delta_n = 0$, so kann angenommen werden, daß die zugehörigen p_n monoton abnehmen; dieselben konvergieren alsdann gleichmäßig gegen σ . Wäre nämlich $\lim \lambda_n = \lambda > 0$, wo λ_n den Maximalabstand der p_n -Randpunkte von σ bedeutet, so müßte der Durchschnitt aller p_n einen Punkt P von M enthalten, der nicht zu σ gehört. Das P enthaltende „Stück“ von M kann nun aber für hinreichend kleines ε nicht durch eine „ ε -Kette“ mit σ verbunden werden (vgl. a. a. O., S. 46) und infolgedessen kann für hinreichend großes n P nicht in p_n liegen.

sämtliche Punkte von M die Ungleichung

$$|x - P(z)| < 2\varepsilon.$$

2. Die Vereinigungsmenge M von endlich oder abzählbar unendlich vielen abgeschlossenen (α - oder) β -Mengen ist stets eine β -Menge⁴³).

Dabei sei zunächst von jeder der gegebenen β -Mengen noch vorausgesetzt, daß sie entweder selbst die Ebene nicht zerlege oder doch als Vereinigungsmenge von (höchstens) abzählbar unendlich vielen abgeschlossenen, die Ebene nicht zerlegenden Mengen darstellbar sei. In § 10 wird jedoch gezeigt werden, daß für jede β -Menge diese Voraussetzung in Wahrheit von selbst erfüllt ist.

Beweis. Auf Grund der letzteren Voraussetzung kann ohne Beschränkung der Allgemeinheit angenommen werden, daß jede der gegebenen β -Mengen B_1, B_2, \dots die Ebene nicht zerlegt, des weiteren, daß jede derselben beschränkt sei, da ja eine nicht beschränkte abgeschlossene Menge als Vereinigungsmenge von abzählbar unendlich vielen beschränkten solchen aufgefaßt werden kann. Bedeutet dann $\varphi(z)$ eine auf M definierte Funktion der Klasse 0 oder 1, so ist $\varphi(z)$ auf B_ν ($\nu = 1, 2, \dots$) als Grenzfunktion von Polynomen $P_{\nu,1}(z), P_{\nu,2}(z), \dots$ darstellbar.

Wird nun zunächst angenommen, daß je zwei der Mengen B_ν keinen Punkt gemein haben, so kann man, da jedes einzelne B_ν die Ebene nicht zerlegt, die n ersten derselben in je ein⁴⁴) einfach geschlossenes Polygon einschließen, derart, daß diese Polygonflächen $\mathfrak{P}_{1n}, \mathfrak{P}_{2n}, \dots, \mathfrak{P}_{nn}$ alle getrennt liegen. Konstruiert man alsdann mit Hilfe des Rungeschen Approximationssatzes ein Polynom $P_n(z)$, das sich in $\mathfrak{P}_{\nu n}$ ($\nu = 1, 2, \dots, n$) von $P_{\nu, n}(z)$ um weniger als $\frac{1}{n}$ unterscheidet, so konvergiert die so entstehende Folge der $P_n(z)$ für jeden Punkt z von M gegen $\varphi(z)$.

Gibt es hingegen unter den Mengen B_ν auch solche, welche Punkte gemein haben, so sei $\mathfrak{P}_\nu, \mathfrak{P}'_\nu, \mathfrak{P}''_\nu, \dots$ eine Folge von B_ν einschließenden, einfachen Polygonbereichen⁴⁵), von denen jeder im Innern des vorangehenden liegt und welche gegen B_ν konvergieren; hierbei treten jedoch, falls B_ν nicht zusammenhängend ist, an Stelle jeder einzelnen Polygonfläche $\mathfrak{P}_\nu^{(n)}$

⁴³) Mit Berücksichtigung der Fußnote ⁴¹) geht aus diesem Satze hervor, daß es β -Mengen gibt, welche ein gegebenes Gebiet (z. B. Rechteck) bis auf die Punkte einer Nullmenge vollständig erfüllen. (Daß eine solche das Gebiet nicht vollständig erfüllen, mit andern Worten keine inneren Punkte haben kann, geht aus dem in der Einleitung erwähnten Osgoodschen Satze hervor.)

⁴⁴) Es würde auch ausreichen, jede der n Mengen in je ein oder endlich viele, voneinander getrennte einfache Polygone einzuschließen.

⁴⁵) „Einfacher Polygonbereich“ = von einem einzigen einfach geschlossenen Polygon begrenzte Fläche.

je eine endliche Anzahl von getrennt liegenden solchen, welche zusammengekommen ebenfalls mit $\mathfrak{F}_\nu^{(n)}$ bezeichnet seien. In jedem Falle möge $\mathfrak{F}_\nu^{(n)}$ überdies so gewählt werden, daß dasselbe keinen der Bereiche $\mathfrak{F}_1^{(n-1)}, \dots, \mathfrak{F}_{\nu-1}^{(n-1)}$ (bzw. keinen Einzelbestandteil eines solchen) vollständig enthält, und ferner so, daß der Rand von $\mathfrak{F}_\nu^{(n)}$ den Rand je eines der Bereiche $\mathfrak{F}_\varrho^{(n-1)}, \mathfrak{F}_\varrho^{(n)}$ ($\varrho = 1, 2, \dots, \nu - 1$) nur an solchen Stellen trifft, an denen er denselben gleichzeitig durchkreuzt.

Werden von $\mathfrak{F}_\nu^{(n)}$ ($n \geq 2$) diejenigen Teile fortgelassen, welche bereits einem oder mehreren der Bereiche $\mathfrak{F}_1^{(n-1)}, \dots, \mathfrak{F}_{\nu-1}^{(n-1)}$ angehören, so ist die übrig bleibende⁴⁶⁾ Fläche $\mathfrak{F}_{\nu n}$ infolge obiger Festsetzungen entweder selbst eine einfache Polygonfläche oder besteht aus einer endlichen Anzahl solcher, die voneinander getrennt liegen. ($\mathfrak{F}_{1n} = \mathfrak{F}_1^{(n)}$). Da die Bereiche $\mathfrak{F}_{1n}, \mathfrak{F}_{2n}, \dots, \mathfrak{F}_{nn}$ (bzw. die Einzelbestandteile derselben) ebenfalls sämtlich voneinander getrennt liegen, so kann man ein Polynom $P_n(z)$ herstellen, das sich in $\mathfrak{F}_{\nu n}$ ($\nu = 1, 2, \dots, n$) von $P_{\nu n}(z)$ um weniger als $\frac{1}{n}$ unterscheidet.

Die Folge der $P_n(z)$ konvergiert alsdann, wie leicht zu sehen, für jeden Punkt z von M gegen $\varphi(z)$. Denn ist $z = z_0$ ein Punkt von B_ν , und zwar ν der kleinste Index, für den dies der Fall ist, so gehört z_0 für hinreichend großes n jedem der Bereiche $\mathfrak{F}_{\nu n}$ an.

Obwohl nun, wie aus diesen Sätzen hervorgeht, der Begriff der β -Menge ein sehr umfassender ist^{46a)}, so ist die Frage, ob ein beliebiges linienhaftes Kontinuum oder auch die mit Γ_0 bezeichnete Menge stets eine β -Menge sei, doch zu verneinen, selbst dann, wenn (für die zur gegebenen Menge komplementären Gebiete) die Bedingung A erfüllt ist. Es gilt nämlich:

Γ_0 kann nur dann eine β -Menge sein, wenn (für die Menge der zu $(\Gamma_0 + C)$ komplementären Gebiete \mathfrak{F}_ν) die Bedingung B erfüllt ist⁴⁷⁾.

⁴⁶⁾ Sollte hierbei vom Bereich $\mathfrak{F}_\nu^{(n)}$ nichts übrigbleiben (also $\mathfrak{F}_{\nu n} = 0$ sein), so ist $\mathfrak{F}_{\nu n}$ im folgenden außer Betracht zu lassen.

^{46a)} Man vergleiche auch Fußnote ⁴³⁾.

⁴⁷⁾ Bei dem Beispiel S. 224 ist also Γ keine β -Menge, also auch nicht als Vereinigungsmenge von abzählbar unendlich vielen α -Mengen (z. B. Jordanschen Kurven) darstellbar (wobei offenbar ausschlaggebend ist, daß Γ neben den „Zentralkreisen“ und „Stielen“ auch noch die Häufungspunkte beider enthält). — Daß hier Γ nicht als Vereinigungsmenge von abzählbar vielen α -Mengen darstellbar ist, läßt sich auch auf direktem Wege nachweisen: Diejenige Teilmenge von Γ , welche aus den Zentralkreisen und deren Häufungsstellen besteht, heiße Γ^* . Wäre Γ und daher auch Γ^* als Vereinigungsmenge von abzählbar vielen α -Mengen darstellbar, etwa $\Gamma^* = \alpha_1 \dot{+} \alpha_2 \dot{+} \dots$, so müßte es unter den Flächen der sämtlichen „Planetenkreise“ $k_\mu^{(\nu)}$ ($\mu, \nu = 1, 2, \dots$) mindestens eine, etwa k^1 geben, die keinen Punkt von α_1 enthält; andernfalls nämlich müßte, da α_1 abgeschlossen ist und jeder Punkt von Γ^* ein Häufungspunkt der $k_\mu^{(\nu)}$ ist, jeder

Ist nämlich die Bedingung B für die Gebiete \mathfrak{F}_v , nicht erfüllt, so müssen, wenn $\varphi_0(z)$ auf Γ_0 als Grenzfunktion einer Folge von Polynomen darstellbar sein soll, *notwendig unendlich viele Gebiete \mathfrak{F}_v , vorhanden sein, bei denen $\varphi_0(z)$ an je eine im Innern des betreffenden Gebietes \mathfrak{F}_v , reguläre Funktion $\varphi_v(z)$ Vitalischen Anschluß besitzt⁴⁸⁾, und zwar muß die Gesamtheit derjenigen Gebiete \mathfrak{F}_v , bei denen dies nicht der Fall ist, der Bedingung B genügen.*

Denn gibt es eine Folge von Polynomen $P_n(z)$ der verlangten Art und unterscheidet man ähnlich wie früher zwischen solchen Gebieten \mathfrak{F}_v , für welche die $P_n(z)$ gemeinsam beschränkt sind, und den übrigen, so ergibt sich die Richtigkeit der letzten Behauptung in ganz analoger Weise wie beim Beweis α) auf S. 226—227.

Daraus folgt aber dann weiter, daß Γ_0 in diesem Falle tatsächlich keine β -Menge sein kann, mit andern Worten: daß $\varphi_0(z)$ nicht als beliebige Funktion von x und y der Klasse 0 oder 1 gewählt werden kann. Denn findet z. B. für das Gebiet \mathfrak{F}_α Vitalischer Anschluß statt, so können die Werte von $\varphi_0(z)$ längs zweier beliebig kleiner Teilkontinua des Randes von \mathfrak{F}_α nicht unabhängig voneinander vorgeschrieben werden⁴⁹⁾.

Es geht aus diesen Betrachtungen hervor, daß wenn die Bedingung B nicht erfüllt ist, die Funktion $\varphi_0(z)$ weit stärkeren Einschränkungen⁵⁰⁾ unterworfen ist, als wenn Γ_0 eine β -Menge darstellt, so zum Beispiel nicht einmal als beliebige ganze rationale Funktion von x und y gewählt werden kann⁵¹⁾.

Punkt von Γ^* zu α_1 gehören, also Γ^* selbst eine α -Menge sein, was wegen der zu Γ^* gehörenden geschlossenen Kurven unmöglich ist. Analog schließt man, daß es im Innern von k^1 einen gewissen Planetenkreis k^2 geben muß, dessen Fläche auch keinen Punkt von α_2 enthält usf. Der (ebenfalls zu Γ^* gehörende) Grenzpunkt von k^1, k^2, \dots wäre dann in keiner der Mengen $\alpha_1, \alpha_2, \dots$ enthalten.

⁴⁸⁾ Oder, was dasselbe bedeutet, unendlich viele Gebiete \mathfrak{F}_v , für deren Rand $\varphi_0(z)$ durch eine gleichmäßig beschränkte Folge von Polynomen darstellbar ist.

⁴⁹⁾ Näheres siehe F. Hartogs²⁸⁾, S. 170.

⁵⁰⁾ Weiteres über den Charakter dieser Einschränkungen siehe F. Hartogs²⁸⁾, S. 170—172, wo insbesondere nachgewiesen ist, daß, falls in \mathfrak{F}_α Vitalischer Anschluß stattfindet, $\varphi_0(z)$ sich längs des Randes von \mathfrak{F}_α „fast überall“ stetig an die im Innern von \mathfrak{F}_α reguläre Funktion $\varphi_\alpha(z)$ anschließen muß.

⁵¹⁾ Beispielsweise kann, wenn die Bedingung B nicht erfüllt ist, $\varphi_0(z)$ nicht gleich einer (nichtkonstanten) ganzen rationalen Funktion $g(x, y)$ mit reellen Koeffizienten sein. Denn findet etwa in \mathfrak{F}_α Vitalischer Anschluß statt, so gibt das über eine in \mathfrak{F}_α gelegene geschlossene Kurve k erstreckte Integral $\frac{1}{2\pi i} \int d \log [\varphi_\alpha(z) - c]$ an, wie oft $\varphi_\alpha(z)$ im Innern von k den Wert c annimmt; dies bleibt aber infolge des in der vorigen Fußnote erwähnten Stetigkeitssatzes auch dann noch richtig, wenn man an die Stelle von k den (als rektifizierbar vorausgesetzten) Rand von \mathfrak{F}_α treten läßt, wo-

(Fortsetzung der Fußnote⁵¹⁾ auf nächster Seite.)

Schließlich ergibt sich als Folgerung noch die bemerkenswerte Tatsache, daß eine auf dem Rand eines einfach zusammenhängenden Gebietes stetige Funktion von x und y nicht unter allen Umständen als Grenzfunktion einer (gleichmäßig oder ungleichmäßig konvergierenden) Folge von Polynomen von z darstellbar ist (daß also der Rand eines solchen Gebietes keine β -Menge zu sein braucht). Da nämlich bei dem auf S. 224 behandelten Beispiel Γ dem Rand des (einfach zusammenhängenden) Gebietes \mathfrak{G}_1 vollständig angehört und die Bedingung B nicht erfüllt ist, so kann nach obigem nicht jede auf dem Rand von \mathfrak{G}_1 stetige Funktion von x und y als Grenzfunktion einer Folge von Polynomen von z dargestellt werden (beispielsweise niemals ein nicht-konstantes Polynom von x und y mit reellen Koeffizienten).

Kapitel III.

Geometrische Umformung der Bedingungen A und B.

§ 7.

Die Hauptpunkte.

M sei eine beliebige abgeschlossene Punktmenge der Ebene⁵²). Jedes Gebiet der Ebene, dessen Rand ausschließlich aus Punkten von M besteht, werde ein „ H -Gebiet“ von M genannt⁵³). Wenn jede Umgebung eines Punktes P (von M) mindestens ein und dann also von selbst unendlich viele H -Gebiete vollständig enthält⁵⁴), dann werde P als „Hauptpunkt“ oder kurz „ H -Punkt“ von M bezeichnet⁵⁵).

bei dann das Integral in $\int d \log [g(x, y) - c]$ übergeht und daher, falls c nicht reell ist, gleich null wird. $\varphi_a(z)$ könnte demnach in \mathfrak{F}_a nur reelle Werte annehmen und müßte also gleich einer reellen Konstanten a , somit $g(x, y)$ infolge des obigen Stetigkeitssatzes längs des Randes von \mathfrak{F}_a ebenfalls gleich a sein und daher in \mathfrak{F}_a mindestens ein Extremum (eventuell eine Extremlinie) besitzen. Da dies für unendlich viele Gebiete \mathfrak{F}_a gilt, so müßte $g(x, y)$ in \mathfrak{G} unendlich viele getrennt liegende Extreme (bzw. Extremlinien) haben.

⁵²) Die folgenden Definitionen und Betrachtungen gelten analog auch für mehrdimensionale oder allgemeinere Räume (topologische Räume mit 2. Abzählbarkeitsaxiom).

⁵³) Man könnte (in der Ebene) dem „ H -Gebiet“ noch die Bedingung auferlegen, beschränkt und einfach-zusammenhängend zu sein; der Begriff des H -Punktes würde dadurch keine Änderung erleiden.

⁵⁴) Oder was dasselbe ist: Wenn eine unendliche Folge von H -Gebieten gegen P konvergiert.

⁵⁵) Ist M außerdem nirgends dicht (auf diesen Fall wird es später ganz besonders ankommen), so könnte man, ohne den Begriff des H -Punktes zu ändern, die „ H -Gebiete“ durch die von M bestimmten Komplementärgebiete ersetzen.

Jeder H -Punkt von M ist natürlich auch Häufungspunkt von M . Die Begriffsbildung und die folgende Untersuchung der H -Punkte ist der Definition und der Cantorsche Theorie der Häufungspunkte und Ableitungen nachgebildet.

Die Menge aller H -Punkte von M bezeichnen wir als die (erste) „ H -Ableitung“ $M^{[1]}$ von M . Daß die H -Ableitung eine abgeschlossene Menge ist, folgt unmittelbar aus der Definition. Die H -Ableitung von $M^{[1]}$ soll „zweite H -Ableitung“ $M^{[2]}$ von M genannt werden und die Wiederholung dieses H -Ableitungs-Prozesses führt zur 3., 4., ..., n -ten H -Ableitung, die wir mit $M^{[3]}$, $M^{[4]}$, ..., $M^{[n]}$ bezeichnen. Ist für jedes ganze positive n die H -Ableitung $M^{[n]}$ nicht leer, so nennen wir den Durchschnitt $\mathfrak{D}(M^{[1]}, M^{[2]}, \dots, M^{[n]}, \dots)$ die „ ω -te H -Ableitung“ $M^{[\omega]}$. Als Durchschnitt einer absteigenden Folge abgeschlossener Mengen ist $M^{[\omega]}$ ebenfalls abgeschlossen und, wenn M oder irgendein $M^{[n]}$ beschränkt ist, nicht leer. Die H -Ableitung von $M^{[\omega]}$ ist $M^{[\omega+1]}$; usw. Durch die beiden Erzeugungsprinzipien „ H -Ableitung“ und „Durchschnittbildung“ wird so für jede Zahl α der 1. oder 2. Zahlenklasse eine „ α -te H -Ableitung“ $M^{[\alpha]}$ definiert.

Wird M nicht als abgeschlossen vorausgesetzt, so gehe man erst zur „abgeschlossenen Hülle“ $\bar{M} = M \dot{+} M'$ von M über, die auch als „0-te H -Ableitung“ $M^{[0]}$ bezeichnet werde; und man verstehe dann unter einem „ H -Punkt von M “ einen H -Punkt von $M^{[0]}$ und unter der „ α -ten H -Ableitung von M “ die α -te H -Ableitung von $M^{[0]}$.

Ferner sagen wir, ein Punkt P (von $M^{[0]}$) sei bezüglich M von der „ H -Ordnung α “ oder kurz von der „Ordnung α “, wenn P der α -ten H -Ableitung von M , aber nicht mehr der $(\alpha + 1)$ -ten H -Ableitung angehört. (Auch für $\alpha = 0$)⁵⁶⁾. Den nicht zu $M^{[0]}$ gehörenden Punkten der Ebene werde die Ordnung -1 zuerteilt. Ein Häufungspunkt von Punkten mindestens α -ter Ordnung ist ebenfalls mindestens von α -ter Ordnung.

Um durch Beispiele zu zeigen, daß man bei gegebenem α Mengen M konstruieren kann, für die alle H -Ableitungen bis zur α -ten einschließlich nicht leer (und voneinander verschieden) sind, während die $(\alpha + 1)$ -te H -Ableitung verschwindet, gehe man von den früheren Figuren $\mathfrak{F}_1, \mathfrak{F}_2, \mathfrak{F}_3, \dots, \mathfrak{F}_n, \dots$ aus, die in der Grenze das Beispiel \mathfrak{F}_* geliefert haben (vgl. S. 225). \mathfrak{F}_n ist damals dadurch entstanden, daß die „Planeten-Kreise“ von \mathfrak{F}_{n-1} durch zu \mathfrak{F}_1 ähnliche Figuren ersetzt wurden. Man kann übrigens (was in dem jetzigen Zusammenhang vorzuziehen ist) \mathfrak{F}_n auch dadurch erzeugen, daß man jeden „Planeten-Kreis“ von \mathfrak{F}_1 durch eine zu \mathfrak{F}_{n-1} ähnliche Figur ersetzt. Die n -te H -Ableitung von \mathfrak{F}_n ⁵⁷⁾ ist nicht leer und besteht aus dem „Zentral-Kreis“ von \mathfrak{F}_1 , so daß $(\mathfrak{F}_n)^{[n+1]}$ verschwindet. Wenn wir die „Planeten-Kreise“ von \mathfrak{F}_1 in eine ein-

⁵⁶⁾ Für einen allen H -Ableitungen angehörenden Punkt könnte die Ordnung = Ω gesetzt werden.

⁵⁷⁾ \mathfrak{F}_n bedeutet hier und im folgenden die Gesamtheit der der Figur \mathfrak{F}_n angehörenden Linien.

fache Folge anordnen und den 1., 2., ..., ν -ten, ... „Planeten-Kreis“ von \mathfrak{F}_1 beziehungsweise durch zu $\mathfrak{F}_1, \mathfrak{F}_2, \dots, \mathfrak{F}_\nu, \dots$ ähnliche Figuren ersetzen, so wird die so entstehende Figur \mathfrak{F}_ω den „Zentral-Kreis“ von \mathfrak{F}_1 zur ω -ten H -Ableitung besitzen; usw. Zur Konstruktion der Beispiele \mathfrak{F}_α , für welche $(\mathfrak{F}_\alpha)^{[\alpha]} \neq 0, (\mathfrak{F}_\alpha)^{[\alpha+1]} = 0$ ist, hat man also die folgenden Erzeugungsprinzipien (die den oben bei der Definition der H -Ableitungen verwendeten entsprechen): 1. Ersetzen jedes „Planeten-Kreises“ von \mathfrak{F}_1 durch eine zu \mathfrak{F}_β ähnliche Figur ($\beta < \alpha$) liefert $\mathfrak{F}_{\beta+1}$; 2. Wenn $\beta_1 < \beta_2 < \dots < \beta_\nu < \dots$ eine Fundamentalfolge ist, die β_ω bestimmt: Ersetzen des ν -ten „Planeten-Kreises“ von \mathfrak{F}_1 durch eine zu \mathfrak{F}_{β_ν} ähnliche Figur für alle ν liefert $\mathfrak{F}_{\beta_\omega}$. — Dagegen sind bei dem früheren Beispiel \mathfrak{F}_* (S. 225) alle folgenden H -Ableitungen mit der ersten (nicht leeren) H -Ableitung identisch.

Eine (nicht leere) Menge, die mit ihrer H -Ableitung⁵⁸⁾ identisch ist, wollen wir als „ H -perfekt“ bezeichnen. Sie ist gleichzeitig „ H -abgeschlossen“, d. h. sie enthält ihre H -Ableitung, und „ H -in sich dicht“, d. h. ist in ihrer H -Ableitung enthalten (oder anders gesagt: jeder ihrer Punkte ist ein H -Punkt von ihr). Wir nennen ferner eine Menge „ H -reduzibel“, wenn irgendeine ihrer H -Ableitungen verschwindet. Ist M eine H -reduzible Menge und N eine beschränkte abgeschlossene Menge, so gibt es, wie man sich leicht überzeugt, unter den bezüglich M definierten Ordnungen aller Punkte von N eine größte; diese bezeichnen wir als die „Maximalordnung“ von N hinsichtlich M . Diese Maximalordnung ist auch dann definiert und $= -1$, wenn $\mathfrak{D}(N, M^{[0]})$ leer ist. Ist N eine beliebige Menge, so gibt es unter den bezüglich M definierten Ordnungen aller Punkte von N (wegen der Wohlordnung der Ordnungszahlen) eine kleinste γ_0 und diese bezeichnen wir als die „Minimalordnung“ von N hinsichtlich M .

Es gilt nun hier das Analogon zum Cantor-Bendixsonschen Satz:

Für jede beliebige Menge M gibt es eine kleinste Zahl α der 1. oder 2. Zahlenklasse, für welche $M^{[\alpha]} = M^{[\alpha+1]}$ ist; wobei dann also $M^{[\alpha]}$ entweder leer oder H -perfekt ist.

Denn: Die H -Ableitungen bilden eine absteigende, wohlgeordnete Reihe von ineinander geschachtelten, abgeschlossenen Mengen; eine solche Reihe besteht aber nach einem bekannten Satz⁵⁹⁾ aus höchstens abzählbar vielen voneinander verschiedenen Gliedern, d. h. es gibt eine kleinste Zahl α der 1. oder 2. Zahlenklasse, so daß von da ab alle Mengen identisch (eventuell leer) sind.

Daraus folgt:

Jede abgeschlossene Menge M ist entweder H -reduzibel oder enthält einen H -perfekten Bestandteil.

⁵⁸⁾ Und also auch mit allen folgenden H -Ableitungen. (Einfachstes Beispiel: Kreisfläche einschl. Rand).

⁵⁹⁾ Vgl. z. B. Encyklopädie d. Math. Wiss. II C 9 a (Zoretti-Rosenthal), S. 871.

Diese beiden Möglichkeiten schließen sich wirklich gegenseitig aus: Eine Menge, die einen H -perfekten Bestandteil enthält, kann niemals H -reduzibel sein; durch den H -Ableitungs-Prozeß nämlich können aus einer H -in sich dichten Teilmenge niemals Punkte verloren gehen.

Wir brauchen später noch den folgenden

Hilfsatz: Ist M eine H -in sich dichte Menge und \mathfrak{A} ein Gebiet, das mindestens einen Punkt von M im Innern enthält, dann ist auch der Durchschnitt $\mathfrak{D}(M, \mathfrak{A}) = D$ H -in sich dicht.

Denn ist P ein beliebiger Punkt von D und $\mathfrak{U}(P)$ eine in \mathfrak{A} gelegene Umgebung von P , so enthält $\mathfrak{U}(P)$ unendlich viele von $M^{(0)}$ und also auch von $D^{(0)}$ begrenzte Gebiete; also ist P ein H -Punkt von D .

Analogien zu anderen Teilen der Cantorsche Theorie werden im § 10 gegeben werden; sie werden erst dort Verwendung finden.

Neben dem oben definierten Begriff des H -Punktes wird im folgenden noch ein modifizierter Begriff benutzt werden, der sich durch Bezugnahme auf eine feste (beliebige) Menge L ergibt. M sei wieder eine abgeschlossene Menge der Ebene⁵²). Jedes Gebiet der Ebene, dessen Rand ausschließlich aus Punkten von M besteht und das in seinem Innern mindestens einen Punkt von L enthält, werde ein „ H -Gebiet von M relativ zu L “ genannt⁵³). Wenn jede Umgebung eines Punktes P (von M) mindestens ein und dann von selbst unendlich viele „ H -Gebiete von M relativ zu L “ vollständig enthält, dann werde P als „ H -Punkt von M relativ zu L “ bezeichnet. Alle im vorstehenden Text für den absoluten Begriff der H -Punkte gegebenen Definitionen und Überlegungen gelten unverändert auch für die H -Punkte relativ zu einer festen Menge L , wenn nur überall der Zusatz „relativ zu L “ beigefügt wird. Der Deutlichkeit halber werde die „ α -te Ableitung von M relativ zu L “ mit $M_L^{(\alpha)}$ bezeichnet. Wird speziell unter L die ganze Ebene oder auch ein \bar{M} enthaltendes Gebiet verstanden, so fällt der Begriff des „ H -Punktes von M relativ zu L “ mit dem absoluten Begriff des H -Punktes von M zusammen⁶⁰).

§ 8.

Umformung der Bedingung B.

Wir verwenden wieder die in § 1 eingeführten Bezeichnungen. Es bedeute also wieder \mathfrak{G} ein ebenes, beschränktes, einfach-zusammenhängendes Gebiet; C seinen Rand; $\mathfrak{G}_1, \mathfrak{G}_2, \dots, \mathfrak{G}_n, \dots$ einfach zusammenhängende⁶¹)

⁶⁰) Ein H -Punkt von M relativ zu einer beliebigen festen Menge L ist natürlich gleichzeitig ein absoluter H -Punkt von M ; aber i. a. nicht umgekehrt.

⁶¹) Die folgenden Betrachtungen bleiben auch gültig, wenn diese Teilgebiete \mathfrak{G}_n nicht einfach zusammenhängend sind (was in §§ 10 und 11 verwendet wird).

Teilgebiete von \mathfrak{G} , die in \mathfrak{G} überall dicht liegen; Γ die Menge der keinem \mathfrak{G}_ν angehörenden Punkte von \mathfrak{G} ; \mathfrak{C} die Vereinigungsmenge ($C + \Gamma$). Außerdem bedeute $\bar{\Gamma}$ die abgeschlossene Hülle von Γ . Ferner zeichnen wir wieder wie früher einen Kreis K , der $(\mathfrak{G} + C)$ im Innern enthält.

Satz. *Mit der Bedingung B ist die folgende Bedingung völlig gleichwertig:*

Γ darf keine Teilmenge Γ^* enthalten, die aus lauter eigenen H -Punkten besteht⁶²⁾ Anders gesagt: Γ darf keine H -in sich dichte Teilmenge Γ^* enthalten.

Ist Γ^* eine H -in sich dichte Teilmenge von Γ und bezeichnet Γ_0^* den Durchschnitt von Γ^* mit einem samt seinem Rand im Innern von \mathfrak{G} gelegenen Gebiet (das wirklich Punkte von Γ^* enthält), so ist die abgeschlossene Hülle $\bar{\Gamma}_0^*$ von Γ_0^* in \mathfrak{C} und also auch in Γ enthalten. Da (Hilfssatz von § 7) Γ_0^* auch H -in sich dicht ist, so ist $\bar{\Gamma}_0^*$ H -perfekt. Wir können also, wenn wir nun statt $\bar{\Gamma}_0^*$ kürzer Γ_* schreiben, die vorstehende Bedingung auch so formulieren:

Γ darf keine H -perfekte Teilmenge Γ_* enthalten⁶³⁾.

Diesen Satz werden wir nachher beweisen. Zuvor aber wollen wir besonders bemerken, daß mit dieser Bedingung keineswegs die Forderung: „ Γ soll H -reduzibel sein“ gleichwertig ist. Da Γ nicht abgeschlossen ist, läßt sich dies jedenfalls nicht unmittelbar aus unserer Bedingung folgern; aber noch mehr: Man kann Beispiele angeben, bei denen Γ keine H -perfekte Teilmenge enthält (und daher [wie auch direkt zu sehen] die Bedingung B erfüllt ist), bei denen aber trotzdem $\bar{\Gamma}$ eine H -perfekte Teilmenge enthält und deshalb Γ nicht H -reduzibel ist⁶⁴⁾.

⁶²⁾ „ Γ^* besteht aus lauter eigenen H -Punkten“ soll bedeuten: Jeder Punkt von Γ^* ist ein H -Punkt von Γ^* .

⁶³⁾ Es muß hervorgehoben werden, daß \mathfrak{C} (nämlich sogar C) sehr wohl einen H -perfekten Bestandteil enthalten kann, auch wenn Bedingung B erfüllt ist. Ein Beispiel hierfür erhält man, indem man als Gebiet \mathfrak{G} das Gebiet \mathfrak{G}_1 unserer früheren Figur \mathfrak{F}_* [vgl. S. 225] wählt, ohne dieses weiter einzuteilen (so daß nur ein Gebiet \mathfrak{G}_ν vorhanden und $\Gamma = 0$ ist). Vgl. auch das Beispiel von ⁶⁴⁾.

Es kann also im obigen Satz nicht Γ durch \mathfrak{C} ersetzt werden.

⁶⁴⁾ Ein derartiges Beispiel erhält man folgendermaßen: In unserer früheren Figur \mathfrak{F}_1 [vgl. S. 224] unterscheide man unter den in eine einfache Folge geordneten „Planeten-Kreisen“ die von gerader und ungerader Ordnung. Jeden Kreis ungerader Ordnung ersetze man, wie früher, durch eine zu \mathfrak{F}_1 ähnliche Figur; während die Kreise gerader Ordnung leer bleiben und beibehalten werden. Jeder Kreis ungerader Ordnung dieser zu \mathfrak{F}_1 ähnlichen Teilfiguren werde dann wieder durch eine zu \mathfrak{F}_1 ähnliche Figur ersetzt, während wieder die Kreise gerader Ordnung dieser Teilfiguren leer bleiben und beibehalten werden. Dieser Prozeß wird unendlich oft iteriert; die in der Grenze entstehende, abgeschlossene Figur sei mit \mathfrak{F}^* bezeichnet. Wir denken uns aus \mathfrak{F}^* die

(Fortsetzung der Fußnote ⁶⁴⁾ auf nächster Seite.)

Will man dennoch unsere Bedingung mittels des Begriffes „ H -reduzibel“ formulieren, so hat man die (am Schluß von § 7 definierten) Relativbegriffe zu verwenden. Deren Benutzung wird ohnedies für den nachherigen Beweis von Bedeutung sein. Wir können dann unseren obigen Satz auch so aussprechen:

*Mit der Bedingung B ist folgende Bedingung gleichwertig:
 Γ darf keine relativ zu \mathfrak{G} H -perfekte Teilmenge enthalten.*

Oder, was jetzt wirklich dasselbe bedeutet:

Γ muß relativ zu \mathfrak{G} H -reduzibel sein⁶⁵⁾.

In diesen beiden Formulierungen kann nunmehr [vgl. hingegen⁶³⁾ und⁶⁴⁾] Γ auch durch $\bar{\Gamma}$ oder durch \mathfrak{C} ersetzt werden.

Daß unsere obige Bedingung in der eben angegebenen Weise umgeformt werden kann, ergibt sich aus dem folgenden

Hilfssatz. Wenn \mathfrak{C} (oder $\bar{\Gamma}$ oder auch Γ) eine relativ zu \mathfrak{G} H -perfekte Teilmenge \mathfrak{C}_* enthält, so enthält Γ eine (absolut) H -perfekte Teilmenge Γ_* ; und umgekehrt.

Beweis. α) Enthält \mathfrak{C} eine relativ zu \mathfrak{G} H -perfekte Teilmenge \mathfrak{C}_* , so kann nicht \mathfrak{C}_* Teil von C sein; C für sich allein besitzt ja überhaupt keinen H -Punkt relativ zu \mathfrak{G} . Also muß \mathfrak{C}_* einen nicht zu C gehörenden Punkt P enthalten. Wegen der Abgeschlossenheit von C ist dann auch eine Umgebung $\mathfrak{U}(P)$ samt ihrem Rand zu C elementenfremd; deshalb ist der Durchschnitt $\mathfrak{D}[\mathfrak{U}(P), \mathfrak{C}_*]$ in Γ enthalten. Wegen des Hilfssatzes von § 7 ist diese letztere Menge H -in sich dicht (relativ zu \mathfrak{G} und daher auch absolut); ihre abgeschlossene Hülle ist also eine H -perfekte Teilmenge von Γ .

β) Umgekehrt: Enthält Γ eine (absolut) H -in sich dichte Teilmenge Γ_* , so auch (wieder wegen des Hilfssatzes von § 7) einen H -perfekten Bestandteil Γ_{**} , der einem samt seinem Rand ganz im Innern von \mathfrak{G}

Flächen aller „Zentralkreise“ herausgeschnitten und alle die „Zentralkreise“ miteinander verknüpfenden Verbindungsstrecken sowie $s_0^{(1)}$ zum Rande C des dann übrigbleibenden (nunmehr einfach-zusammenhängenden) Gebietes \mathfrak{G} gerechnet. Die so entstehende Figur \mathfrak{F}^{**} ist das gewünschte Beispiel.

Bei diesem Beispiel besteht Γ aus allen vorkommenden Kreisen gerader Ordnung und den zugehörigen „Stielen“. Man kann nämlich leicht zeigen, daß alle Häufungspunkte dieser Kreise und ihrer „Stiele“ zu C gehören. (Die Gebiete \mathfrak{G}_* sind die Kreisflächen gerader Ordnung und ferner das von Γ und C begrenzte Restgebiet von \mathfrak{G} .) Γ enthält also überhaupt keinen H -Punkt (weder von Γ noch von \mathfrak{C}); während $\bar{\Gamma}$ einen (ganz zu C gehörenden) H -perfekten Bestandteil, nämlich die Menge der „Zentralkreise“ und ihrer Häufungspunkte, enthält.

⁶⁵⁾ Im Beispiel von ⁶⁴⁾ ist Γ relativ zu \mathfrak{G} H -reduzibel, aber nicht absolut H -reduzibel.

gelegenen Teilgebiet \mathfrak{S} angehört. Γ_{**} ist dann auch relativ zu \mathfrak{G} H -perfekt.

Daß man bei den oben formulierten Bedingungen dann weiterhin zur Forderung „relativ zu \mathfrak{G} H -reduzibel“ übergehen kann, ist für \mathfrak{G} und $\bar{\Gamma}$ wegen der Abgeschlossenheit selbstverständlich; dieser Übergang ist hier nun auch für Γ möglich, eben weil die Aussage „ Γ ist (relativ zu \mathfrak{G}) H -reduzibel oder nicht“ definitionsgemäß dasselbe bedeutet wie die entsprechende Aussage für $\bar{\Gamma}$.

Nach diesen Umformungen des oben angegebenen Satzes gehen wir nunmehr zum Beweis desselben über. Wir beweisen die beiden Bestandteile unseres Satzes nacheinander.

1. Teilsatz. *Enthält Γ eine H -perfekte Teilmenge Γ_* , so kann die Bedingung B nicht erfüllt sein.*

Beweis. Es sei ein beliebiges Streifensystem $\{\mathfrak{S}_\nu\}$ vorgelegt, so daß \mathfrak{S}_ν das Innere von \mathfrak{G}_ν mit K verbindet. \mathfrak{S}_1 sei ein beschränktes H -Gebiet von Γ_* ; dann ist \mathfrak{S}_1 entweder mit einem unserer Gebiete \mathfrak{G}_ν identisch oder \mathfrak{S}_1 enthält ein oder mehrere \mathfrak{G}_ν . Es sei \mathfrak{G}_{n_1} in \mathfrak{S}_1 enthalten. Dann wird der Streifen \mathfrak{S}_{n_1} ins Innere von \mathfrak{G}_{n_1} , also ins Innere von \mathfrak{S}_1 eindringen. Deshalb liegt ein Punkt P_1 des Randes von \mathfrak{S}_1 , d. h. von Γ_* , und zugleich eine Umgebung $\mathfrak{U}_1(P_1)$ dieses Punktes innerhalb \mathfrak{S}_{n_1} . In $\mathfrak{U}_1(P_1)$, also in \mathfrak{S}_{n_1} , muß dann ein H -Gebiet \mathfrak{S}_2 von Γ_* (nebst Rand) liegen, das \mathfrak{G}_{n_1} nicht enthält; \mathfrak{G}_{n_2} sei in \mathfrak{S}_2 enthalten. \mathfrak{S}_{n_2} dringt in das Innere von \mathfrak{G}_{n_2} , also von \mathfrak{S}_2 ein. Deshalb liegt innerhalb \mathfrak{S}_{n_2} ein Punkt P_2 des Randes von \mathfrak{S}_2 , d. h. von Γ_* , und eine (samt Begrenzung) ganz im Gebiet $\mathfrak{U}_1(P_1)$ enthaltene Umgebung $\mathfrak{U}_2(P_2)$. Usw. für eine unendliche Folge von verschiedenen Indizes n_k ($k = 1, 2, 3, \dots$). Ist nun P_0 ein Häufungspunkt der $\{P_k\}$, also ebenfalls zu Γ gehörend, so liegt P_0 in jedem der Gebiete $\mathfrak{U}_k(P_k)$ und somit auch innerhalb jedes der Streifen \mathfrak{S}_{n_k} ($k = 1, 2, 3, \dots$). Das Streifensystem $\{\mathfrak{S}_\nu\}$ kann also die Bedingung B nicht erfüllen.

Bevor wir uns dem wesentlich mehr Mühe bereitenden Beweis des 2. Teilsatzes zuwenden, schicken wir erst die Definition eines bei diesem Beweis benötigten Begriffes voraus:

\mathfrak{A} sei ein beschränkter, abgeschlossener Bereich und Γ (wie beim nachfolgenden Satz) eine relativ zu \mathfrak{G} H -reduzible Menge. μ bezeichne die Maximalordnung von \mathfrak{A} hinsichtlich Γ relativ zu \mathfrak{G} (s. § 7). Wir definieren nun für \mathfrak{A} folgendermaßen einen „Index“ α (hinsichtlich Γ relativ zu \mathfrak{G}): Ist $\mu = -1$ (d. h. enthält \mathfrak{A} keinen Punkt von $\bar{\Gamma}$), so sei auch $\alpha = -1$. Ist hingegen $\mu \geq 0$ (d. h. enthält \mathfrak{A} mindestens einen Punkt von $\bar{\Gamma}$), so sei $\alpha = \mu + 1$

oder $= \mu$, je nachdem (wenigstens) ein H -Gebiet von $\Gamma_{\mathfrak{G}}^{[\mu]}$ relativ zu \mathfrak{G} vollständig in \mathfrak{A} liegt oder nicht.

Der Index eines Teilbereichs von \mathfrak{A} ist offenbar stets \leq dem Index von \mathfrak{A} selbst.

Nun beweisen wir den

2. Teilsatz. *Ist Γ relativ zu \mathfrak{G} H -reduzibel, so ist die Bedingung B erfüllt⁶⁶⁾.*

Beweis. 1. Da in diesem Beweis alle H -Begriffe relativ zu \mathfrak{G} aufzufassen sind, kann der sich immer wiederholende Zusatz „relativ zu \mathfrak{G} “ durchgehend weggelassen werden.

Unter den H -Ableitungen der H -reduziblen Menge Γ gibt es eine letzte, $\Gamma_{\mathfrak{G}}^{[\alpha_0]}$, welche nicht leer ist⁶⁷⁾. Die H -Gebiete von $\Gamma_{\mathfrak{G}}^{[\alpha_0]}$ relativ zu \mathfrak{G} mögen kurz mit \mathfrak{D}_{α_0} bezeichnet werden. Die untere Grenze der Durchmesser aller beschränkten \mathfrak{D}_{α_0} (falls solche vorhanden sind) ist eine Zahl $\delta > 0$; denn gäbe es \mathfrak{D}_{α_0} von beliebig kleinem Durchmesser, so müßten Punkte von $(\alpha_0 + 1)$ -ter Ordnung existieren. (Ist kein beschränktes \mathfrak{D}_{α_0} vorhanden, so kann unter δ eine beliebige positive Zahl verstanden werden.)

2. Wir wollen nun ein (der Bedingung B genügendes) Streifensystem konstruieren, das unsere Gebiete $\mathfrak{G}_1, \mathfrak{G}_2, \dots$ mit K verbindet. Zunächst können wir das Innere von \mathfrak{G}_1 durch einen noch provisorischen Streifen \mathfrak{S}_1^* mit K verbinden, und zwar derart, daß kein Gebiet \mathfrak{D}_{α_0} ganz in \mathfrak{S}_1^* liegt; dies geschieht mittels einer Methode, die wir als die „Flaschenhalsmethode“ (mit Bezug auf α_0)“ bezeichnen wollen:

Wir beginnen den Aufbau von \mathfrak{S}_1^* mit einem Rechteck \mathfrak{R}_1 von folgender Eigenschaft: der Durchmesser von \mathfrak{R}_1 sei $\leq \frac{\delta}{2}$ (so daß also \mathfrak{R}_1 kein \mathfrak{D}_{α_0} vollständig enthält); eine Seite s_0 von \mathfrak{R}_1 soll ganz im Innern von \mathfrak{G}_1 liegen, während die gegenüberliegende Seite s_1 keine Punkte mit \mathfrak{G}_1 gemeinsam haben soll; zugleich kann s_1 (eventuell durch Parallelverschiebung) so gewählt werden, daß s_1 nicht vollständig aus Punkten von $\Gamma_{\mathfrak{G}}^{[\alpha_0]}$ (das ja nirgends dicht ist) besteht. P_1 sei ein nicht zu $\Gamma_{\mathfrak{G}}^{[\alpha_0]}$ gehörender Punkt von s_1 ; es ist dann sogar eine volle Umgebung $U(P_1)$ zu $\Gamma_{\mathfrak{G}}^{[\alpha_0]}$ elementenfremd. Nun werde als Streifenfortsetzung ein an s_1 anschließendes (beliebig kleines) Rechteck \mathfrak{R}_1^* , das einschließlich Rand ganz innerhalb $U(P_1)$ liegt, benutzt (so daß also \mathfrak{R}_1^* keinen Randpunkt irgendeines

⁶⁶⁾ Dieser Teilsatz und der folgende Beweis bleiben auch richtig, wenn man überall den Zusatz „relativ zu \mathfrak{G} “ unterdrückt und also alle H -Begriffe absolut nimmt. Aber bei Weglassung dieses Zusatzes würde, wie die vorhergehenden Bemerkungen zeigen [vgl. insbesondere ⁶⁴⁾ und ⁶⁵⁾], dieser 2. Teilsatz weniger aussagen und deshalb, mit dem 1. Teilsatz zusammengenommen, noch nicht unsern vollen Satz ergeben.

⁶⁷⁾ α_0 existiert auf Grund des § 7 als Maximalordnung von $\bar{\Gamma}$ hinsichtlich Γ (relativ zu \mathfrak{G}).

Ω_{α_0} enthält) [„Flaschenhals“]. An die zu s_1 entgegengesetzte Seite von \mathfrak{R}_1^* werde dann ein neues Rechteck \mathfrak{R}_2 angesetzt, dessen Durchmesser $\leq \frac{\delta}{2}$ ist. \mathfrak{R}_2 werde nun ebenso behandelt wie \mathfrak{R}_1 . s_2 sei eine freie Seite von \mathfrak{R}_2 , die (eventuell nach Parallelverschiebung) nicht vollständig aus Punkten von $\Gamma_{\mathfrak{G}}^{(\alpha_0)}$ bestehe. Der Streifen werde jetzt mittels eines an s_2 anschließenden „Flaschenhalses“ \mathfrak{R}_2^* und dann mit Rechteck \mathfrak{R}_3 (Durchmesser $\leq \frac{\delta}{2}$) fortgesetzt; usw. Es wird möglich sein, mit endlich vielen Schritten nach K zu gelangen⁶⁸⁾, da ja die in der jeweiligen Fortschreitungsrichtung gemessenen Längen der Rechtecke $\mathfrak{R}_1, \mathfrak{R}_2, \mathfrak{R}_3, \dots$ größer als eine geeignete feste positive Zahl genommen werden können. Die Vereinigung $\mathfrak{R}_1 \dot{+} \mathfrak{R}_1^* \dot{+} \mathfrak{R}_2 \dot{+} \mathfrak{R}_2^* \dot{+} \dots \dot{+} \mathfrak{R}_k$ liefert dann offenbar einen Streifen \mathfrak{S}_1^* der verlangten Art. — Bei der Konstruktion von \mathfrak{S}_1^* kann man eine Fortschreitungsrichtung r festhalten (also s_0, s_1, s_2, \dots alle senkrecht zu r). Wir heben aber gleich hier hervor: Wird nicht eine Fortschreitungsrichtung festgehalten, sondern wird ein Wechsel derselben zugelassen (teils senkrecht, teils parallel zu s_0), so kann man einen derartigen Streifen auf Grund der „Flaschenhalsmethode“ auch im Innern eines vorgegebenen, nach K führenden, polygonal begrenzten Bandes \mathfrak{B} konstruieren.

Den „provisorischen“ Streifen \mathfrak{S}_1^* zerspalteten wir nun in drei von \mathfrak{G}_1 nach K führende Teilstreifen. Von dem mittleren dieser drei Teilstreifen schneiden wir ein an s_0 anschließendes, ganz in \mathfrak{G}_1 enthaltenes Rechteck \mathfrak{R}_0 (Teil von \mathfrak{R}_1) ab und den so verkleinerten mittleren Teilstreifen wollen wir als den *endgültigen Streifen* \mathfrak{S}_1 ⁶⁹⁾ nehmen; während wir die beiden anderen (an die Seitenränder von \mathfrak{S}_1^* anstoßenden) Teilstreifen zusammen mit \mathfrak{R}_0 als das „Streifenbett“ \mathfrak{B}_1 von \mathfrak{S}_1 bezeichnen. Derjenige Teil des Randes von \mathfrak{S}_1 , der an \mathfrak{B}_1 angrenzt, heiße das „Streifenufer“ von \mathfrak{S}_1 . Den von \mathfrak{S}_1^* zu \mathfrak{S}_1 führenden Prozeß nennen wir kurz: „Drittellung“.

3. Bezüglich \mathfrak{G}_2 sind folgende Fälle zu unterscheiden:

a) \mathfrak{G}_2 liegt ganz oder wenigstens teilweise außerhalb \mathfrak{S}_1^* . Dann kann man wieder mittels der „Flaschenhalsmethode“ (mit Bezug auf α_0) \mathfrak{G}_2 durch einen provisorischen Streifen \mathfrak{S}_2^* , der kein Gebiet Ω_{α_0} vollständig enthält, mit K verbinden, und zwar so, daß \mathfrak{S}_2^* von \mathfrak{S}_1^* positive Entfernung hat.

b) \mathfrak{G}_2 liegt ganz in \mathfrak{S}_1^* und zugleich ganz oder teilweise in \mathfrak{B}_1 . Dann verbinde man \mathfrak{G}_2 mit K in beliebiger Weise durch einen Streifen \mathfrak{S}_2^* , der ganz innerhalb \mathfrak{B}_1 verläuft und von den Rändern von \mathfrak{B}_1 positive Entfernung hat.

⁶⁸⁾ D. h. von dem letzten dieser Rechtecke \mathfrak{R}_k soll eine freie Seite s_k bereits ganz außerhalb K liegen.

⁶⁹⁾ Damit werde der betr. Streifen einschließlich Rand bezeichnet.

c) \mathfrak{G}_2 liegt ganz in \mathfrak{S}_1 . Der „Index“ von \mathfrak{S}_1 (hinsichtlich Γ) sei α_1 (so daß $0 \leq \alpha_1 \leq \alpha_0$). Ist dann P irgendein innerer Punkt von \mathfrak{G}_2 , so gehört P einem von $\Gamma_{\mathfrak{G}_2}^{[\alpha_1]}$ bestimmten Komplementärgebiet \mathfrak{D} an. Dieses letztere muß aber über \mathfrak{S}_1 hinausragen (da sonst der Index von \mathfrak{S}_1 mindestens gleich $\alpha_1 + 1$ wäre) und man kann daher eine (ganz in \mathfrak{G}_2 und \mathfrak{D} enthaltene) Umgebung $\mathfrak{U}(P)$ durch einen polygonal begrenzten Streifen \mathfrak{T}_2 ⁶⁹⁾, der $\Gamma_{\mathfrak{G}_2}^{[\alpha_1]}$ nicht trifft, über das Streifenufer von \mathfrak{S}_1 hinweg mit dem Innern von \mathfrak{B}_1 verbinden. Die „Maximalordnung“ β_2 von \mathfrak{T}_2 hinsichtlich Γ ist $< \alpha_1$. Gibt es unter den H -Gebieten \mathfrak{D}_{β_2} von $\Gamma_{\mathfrak{G}_2}^{[\beta_2]}$ relativ zu \mathfrak{G} solche, die ganz in \mathfrak{T}_2 liegen, so haben wieder deren Durchmesser eine untere Grenze $\delta_2 > 0$. Man kann deshalb wieder mittels der „Flaschenhalsmethode“ (mit Bezug auf β_2) einen Streifen \mathfrak{R}_2^{**} konstruieren, der innerhalb \mathfrak{T}_2 verläuft, kein Gebiet \mathfrak{D}_{β_2} ganz enthält und aus $\mathfrak{U}(P)$ über das Streifenufer von \mathfrak{S}_1 hinweg nach \mathfrak{B}_1 führt⁷⁰⁾. Der Index von \mathfrak{R}_2^{**} ist dann $\leq \beta_2$. \mathfrak{R}_2^{**} möge das Streifenufer von \mathfrak{S}_1 nur in einer einzigen Strecke s_2^{**} treffen. Die Minimalordnung von s_2^{**} (hinsichtlich Γ) sei γ_2 . Wegen der Abgeschlossenheit von $\Gamma_{\mathfrak{G}_2}^{[\gamma_2+1]}$ müssen die Punkte der γ_2 -ten Ordnung auf s_2^{**} (wenn die Endpunkte von s_2^{**} weggelassen werden) eine offene Menge bilden und daher existiert ein nur aus Punkten γ_2 -ter Ordnung bestehendes, abgeschlossenes Teilintervall s_2^* von s_2^{**} . Dann soll \mathfrak{R}_2^{**} zu einem schmäleren (ebenfalls $\mathfrak{U}(P)$ mit \mathfrak{B}_1 verbindenden) Streifen \mathfrak{R}_2^* verengt werden, der mit dem Streifenufer von \mathfrak{S}_1 nur die Strecke s_2^* gemeinsam hat. Der Index von \mathfrak{R}_2^* ist ebenfalls $\leq \beta_2$, also sicher $< \alpha_1$. Die Vereinigung von \mathfrak{R}_2^* mit einem fortsetzenden, nach K führenden Streifen, der ganz in \mathfrak{B}_1 verläuft und dabei von den Rändern von \mathfrak{B}_1 positiven Abstand besitzt, soll den provisorischen Streifen \mathfrak{S}_2^* bilden.

In jedem der drei Fälle a), b), c) soll aus \mathfrak{S}_2^* wieder durch „Drittelung“ der endgültige Streifen \mathfrak{S}_2 und sein Streifenbett \mathfrak{B}_2 entstehen.

4. Es seien bereits $n - 1$ Streifen $\mathfrak{S}_1, \mathfrak{S}_2, \dots, \mathfrak{S}_{n-1}$ gezogen, die $\mathfrak{G}_1, \mathfrak{G}_2, \dots, \mathfrak{G}_{n-1}$ mit K verbinden; die zugehörigen „provisorischen“ Streifen seien mit $\mathfrak{S}_1^*, \mathfrak{S}_2^*, \dots, \mathfrak{S}_{n-1}^*$, die zugehörigen „Streifenbetten“ mit $\mathfrak{B}_1, \mathfrak{B}_2, \dots, \mathfrak{B}_{n-1}$ bezeichnet. Zu jedem Streifen, provisorischen Streifen und Streifenbett soll sein Rand hinzugerechnet werden. Wir wollen angeben, wie dann in entsprechender Weise der Streifen \mathfrak{S}_n konstruiert werden soll, der \mathfrak{G}_n mit K verbindet.

Als Vorbereitung dazu stellen wir noch die folgenden Definitionen auf. Zunächst definieren wir die „reduzierten Streifenbetten“ $\mathfrak{B}_1^{(n-1)}, \mathfrak{B}_2^{(n-1)}, \dots, \mathfrak{B}_{n-1}^{(n-1)}$: Es entstehe $\mathfrak{B}_\nu^{(n-1)}$ ($1 \leq \nu \leq n - 2$) dadurch aus \mathfrak{B}_ν , daß man alle in \mathfrak{B}_ν

⁷⁰⁾ Im Falle $\beta_2 = -1$ unterbleibt die Anwendung der Flaschenhalsmethode (so daß \mathfrak{R}_2^{**} mit \mathfrak{T}_2 zusammenfällt).

enthaltenen Teile von $\mathfrak{S}_{\nu+1}^*$, $\mathfrak{S}_{\nu+2}^*$, ..., \mathfrak{S}_{n-1}^* aus \mathfrak{B}_ν herausschneidet; während $\mathfrak{B}_{n-1}^{(n-1)} = \mathfrak{B}_{n-1}$ sein möge. Auch jedem $\mathfrak{B}_\nu^{(n-1)}$ soll sein Rand hinzugerechnet werden. Der von den $n-1$ ersten „provisorischen“ Streifen nicht überdeckte Teil von \mathfrak{G} , also $[\mathfrak{G} - (\mathfrak{S}_1^* \dagger \mathfrak{S}_2^* \dagger \dots \dagger \mathfrak{S}_{n-1}^*)]$, soll als das „freie Restgebiet“ \mathfrak{G}_{n-1} bezeichnet werden.

Ferner: Ein Streifen \mathfrak{S}_ν bzw. ein provisorischer Streifen \mathfrak{S}_ν^* wird durch die Ränder der vorhergehenden Streifen \mathfrak{S}_μ ($\mu < \nu$) in Teile zerlegt, die wir die „Abschnitte“ von \mathfrak{S}_ν bzw. \mathfrak{S}_ν^* nennen. Jedem Abschnitt soll sein Rand zugerechnet werden.

5. Um nun den \mathfrak{G}_n mit K verbindenden Streifen \mathfrak{S}_n herzustellen, konstruieren wir zuerst einen „provisorischen“ Streifen \mathfrak{S}_n^* , aus dem dann wieder durch „Drittellung“ der endgültige Streifen \mathfrak{S}_n und sein Streifenbett \mathfrak{B}_n entsteht. Wir unterscheiden dabei wieder drei Fälle:

a) \mathfrak{G}_n liegt ganz oder wenigstens teilweise in \mathfrak{G}_{n-1} . Dann kann man wieder mittels der „Flaschenhalsmethode“ (mit Bezug auf α_0) \mathfrak{G}_n durch einen provisorischen Streifen \mathfrak{S}_n^* , der kein Gebiet \mathfrak{D}_{α_0} vollständig enthält, mit K verbinden, und zwar so, daß \mathfrak{S}_n^* von $(\mathfrak{S}_1^* \dagger \dots \dagger \mathfrak{S}_{n-1}^*)$ eine positive Entfernung hat. (Vgl. hierzu die nachfolgende „Bemerkung 2“.)

b) Kein Punkt von \mathfrak{G}_n liegt in \mathfrak{G}_{n-1} , sondern \mathfrak{G}_n liegt ganz oder teilweise in einem oder mehreren $\mathfrak{B}_\nu^{(n-1)}$. Unter diesen habe $\mathfrak{B}_{\nu_0}^{(n-1)}$ den größten Wert von ν . Dann möge \mathfrak{G}_n mit K in beliebiger Weise durch einen provisorischen Streifen \mathfrak{S}_n^* verbunden werden, der ganz innerhalb $\mathfrak{B}_{\nu_0}^{(n-1)}$ verläuft⁷¹⁾.

c) Kein Punkt von \mathfrak{G}_n liegt in \mathfrak{G}_{n-1} oder in einem der $\mathfrak{B}_\nu^{(n-1)}$, so daß also \mathfrak{G}_n ganz in $(\mathfrak{S}_1 \dagger \mathfrak{S}_2 \dagger \dots \dagger \mathfrak{S}_{n-1})$ enthalten ist. \mathfrak{S}_e sei der letzte dieser Streifen, der Punkte von \mathfrak{G}_n enthält. Dann liegt \mathfrak{G}_n *vollständig* in \mathfrak{S}_e , weil in keinem der an \mathfrak{S}_e anstoßenden Bereiche (nämlich weder in $\mathfrak{B}_e^{(n-1)}$ noch in \mathfrak{S}_{e+1}^* , \mathfrak{S}_{e+2}^* , ..., \mathfrak{S}_{n-1}^*) ein Punkt von \mathfrak{G}_n enthalten ist. P sei nun ein Punkt von \mathfrak{G}_n , dessen Umgebung $\mathfrak{U}(P)$ in \mathfrak{G}_n und in einem einzigen Abschnitt \mathfrak{A} von \mathfrak{S}_e liegt; der „Index“ von \mathfrak{A} (hinsichtlich Γ) sei α (≥ 0). $\mathfrak{U}(P)$ gehört einem Komplementärgebiet \mathfrak{D} von $\Gamma_{\mathfrak{G}}^{[\alpha]}$ an; und zwar muß \mathfrak{D} über das Streifenufer von \mathfrak{A} ⁷²⁾ hinausragen, wie durch eine nachfolgende [der Bestimmung von s_2^* in 3 c) entsprechende] Festsetzung noch sichergestellt wird (Näheres in der unten stehenden „Bemerkung 1“). Man kann daher im Innern von \mathfrak{D} einen Streifen \mathfrak{T}_n ⁶⁹⁾ konstruieren, der $\mathfrak{U}(P)$ über das Streifenufer von \mathfrak{S}_e hin-

⁷¹⁾ Damit dies möglich sei, muß feststehen, daß jede Komponente von $\mathfrak{B}_{\nu_0}^{(n-1)}$ bis an K heranreicht. Dies wird erst nachträglich in der nachfolgenden „Bemerkung 2“ gezeigt.

⁷²⁾ D. h. über den zu \mathfrak{A} gehörenden Teil des Streifenufers von \mathfrak{S}_e .

weg mit dem Innern von \mathfrak{B}_e verbindet, ohne in einen andern Abschnitt von \mathfrak{S}_e einzudringen; und zwar kann \mathfrak{Z}_n so gelegt werden, daß er von etwa in \mathfrak{Q} vorhandenen Stücken der Streifen $\mathfrak{S}_{e+\lambda}^*$ ($1 \leq \lambda \leq n-1-\rho$) positiven Abstand hat. (Dies letztere ist deshalb möglich, weil ein Streifen $\mathfrak{S}_{e+\lambda}^*$, sofern er von einem in \mathfrak{A} enthaltenen Teil eines Gebietes $\mathfrak{G}_{e+\lambda}$ ausgeht, bereits genau in derselben Weise konstruiert worden ist, wie es jetzt für \mathfrak{S}_n^* geschehen soll, und weil $\mathfrak{U}(P)$ von $\mathfrak{S}_{e+\lambda}^*$ getrennt liegt; vgl. auch „Bemerkung 2“.) Demnach führt \mathfrak{Z}_n von $\mathfrak{U}(P)$ nicht nur nach \mathfrak{B}_e , sondern genauer nach $\mathfrak{B}_e^{(n-1)}$ (und zwar unmittelbar dorthin). Die Maximalordnung β_n von \mathfrak{Z}_n hinsichtlich Γ ist $< \alpha$. Nach der „Flaschenhalsmethode“ mit Bezug auf β_n kann man nun innerhalb \mathfrak{Z}_n einen Streifen \mathfrak{R}_n^{**} konstruieren, der kein H -Gebiet von $\Gamma_{\mathfrak{G}}^{[\beta_n]}$ (relativ zu \mathfrak{G}) vollständig enthält und der $\mathfrak{U}(P)$ mit $\mathfrak{B}_e^{(n-1)}$ verbindet⁷³). Der „Index“ von \mathfrak{R}_n^{**} (hinsichtlich Γ) ist dann $\leq \beta_n$. \mathfrak{R}_n^{**} möge ferner das Streifenufer von \mathfrak{S}_e nur in einer einzigen Strecke s_n^{**} treffen. Die Minimalordnung von s_n^{**} (hinsichtlich Γ) sei γ_n . Wegen der Abgeschlossenheit von $\Gamma_{\mathfrak{G}}^{[\gamma_n+1]}$ müssen die Punkte der γ_n -ten Ordnung auf s_n^{**} (wenn die Endpunkte von s_n^{**} weggelassen werden) eine offene Menge bilden und daher existiert ein nur aus Punkten γ_n -ter Ordnung bestehendes, abgeschlossenes Teilintervall s_n^* von s_n^{**} . Dann soll \mathfrak{R}_n^{**} zu einem schmäleren (ebenfalls $\mathfrak{U}(P)$ mit $\mathfrak{B}_e^{(n-1)}$ verbindenden) Streifen \mathfrak{R}_n^* verengt werden, der mit dem Streifenufer von \mathfrak{S}_e nur die Strecke s_n^* gemeinsam hat. Der „Index“ von \mathfrak{R}_n^* (hinsichtlich Γ) ist dann ebenfalls $\leq \beta_n$, also sicher kleiner als α . Aus diesem „provisorischen Streifenkopf“ \mathfrak{R}_n^* möge schließlich der ganze „provisorische Streifen“ \mathfrak{S}_n^* hervorgehen, indem man \mathfrak{R}_n^* im Innern von $\mathfrak{B}_e^{(n-1)}$ auf beliebige Weise bis nach K fortsetzt (immer positiven Abstand von den Rändern haltend)⁷⁴).

Bemerkung 1. Auf Grund der s_n^* betreffenden Festsetzung ist in der Tat sichergestellt, daß \mathfrak{Q} über das Streifenufer von \mathfrak{A} hinausragen muß. Denn: \mathfrak{Q} kann nicht ganz in \mathfrak{A} enthalten sein (da \mathfrak{A} sonst den Index $\alpha+1$ hätte), kann aber andererseits nicht ausschließlich über die \mathfrak{A} von seinen Nachbarabschnitten trennenden Seitenstrecken s hinausführen. Eine solche Seitenstrecke s besteht nämlich [nach unseren Festsetzungen⁷⁵] nur aus Punkten einer Ordnung γ_s . Ist $\alpha > \gamma_s$, so kann $\Gamma_{\mathfrak{G}}^{[\alpha]}$ (und somit

⁷³) Im Falle $\beta_n = -1$ unterbleibt die Anwendung der Flaschenhalsmethode (so daß \mathfrak{R}_n^{**} mit \mathfrak{Z}_n zusammenfällt).

⁷⁴) \mathfrak{S}_n^* liegt hiernach völlig getrennt von $\mathfrak{S}_{e+1}^*, \dots, \mathfrak{S}_{n-1}^*$.

⁷⁵) Man beachte, daß eine solche „Seitenstrecke“ ausschließlich bei der Konstruktion eines neuen „Streifenkopfes“ \mathfrak{R}_n^* entsteht und dann für den betr. neuen Streifen \mathfrak{S}_n , den provisorischen Streifen \mathfrak{S}_n^* und für die im zugehörigen Streifenbett \mathfrak{B}_n eventuell verlaufenden späteren Streifen und provisorischen Streifen eine Trennungslinie je zweier Abschnitte bedeutet.

der Rand von \mathfrak{D}) keinen Punkt mit s gemeinsam haben; ist $\alpha = \gamma_s$, so muß der Teil von s , an den \mathfrak{D} heranreicht, zum Rand von \mathfrak{D} gehören.

Bemerkung 2. Um die Möglichkeit, stets mit den Streifen bis nach K zu gelangen, nachzuweisen, sei folgendes hinzugefügt. Ad a): Da \mathfrak{S}_ν^* entweder von \mathfrak{S}_μ^* ($\mu < \nu$) getrennt liegt oder ganz in \mathfrak{S}_μ^* enthalten ist, so besteht $(\mathfrak{S}_1^* \dot{+} \mathfrak{S}_2^* \dot{+} \dots \dot{+} \mathfrak{S}_{n-1}^*)$ aus endlich vielen ($m \leq n-1$) getrennten, aus \mathfrak{G} nach K führenden Streifen, durch welche die von K umschlossene Kreisfläche nicht zerlegt wird. — Ad c): Es sei \mathfrak{D}_ρ diejenige Komponente von $\mathfrak{D}(\mathfrak{D}, \mathfrak{A})$, welche $\mathfrak{U}(P)$ enthält. Aus dem gleichen Grunde, wie soeben, besteht $(\mathfrak{S}_{\rho+1}^* \dot{+} \mathfrak{S}_{\rho+2}^* \dot{+} \dots \dot{+} \mathfrak{S}_{n-1}^*)$ aus endlich vielen ($\mu \leq n-1-\rho$) getrennten Streifen; von diesen hat jeder, sofern er überhaupt \mathfrak{D}_ρ trifft (also dort beginnt), nach obiger Konstruktionsvorschrift mit dem Rande von \mathfrak{D}_ρ (nämlich mit dem Streifenufer von \mathfrak{A}) genau eine (von $\Gamma_{\mathfrak{G}}^{(a)}$ getrennt liegende) Strecke gemeinsam. Deshalb wird \mathfrak{D}_ρ nach Herausschneiden der etwa in \mathfrak{D}_ρ enthaltenen Abschnitte dieser $\mathfrak{S}_{\rho+\lambda}^*$ ($1 \leq \lambda \leq n-1-\rho$) nicht zerfallen und der Rand des so entstehenden, reduzierten Bereichs wird noch Teile des Streifenufers von \mathfrak{A} enthalten. — Ad b) und c): Dagegen kann $\mathfrak{B}_\nu^{(n-1)}$ aus mehreren getrennten Bereichen bestehen⁷⁶⁾. Aber jeder dieser Teilbereiche muß Bögen von K enthalten. Denn: Der Rand von \mathfrak{B}_ν zerfällt nach Weglassung der beiden außerhalb K [vgl. 68)] gelegenen Stücke in zwei Teile: einen „inneren Rand“ (der dem Streifenufer von \mathfrak{S}_ν angehört) und einen „äußeren Rand“ B_ν . Gemäß unserer Konstruktion wird B_ν von keinem $\mathfrak{S}_{r+\lambda}^*$ ($\lambda > 0$) getroffen⁷⁷⁾. $(\mathfrak{S}_{\nu+1}^* \dot{+} \mathfrak{S}_{\nu+2}^* \dot{+} \dots \dot{+} \mathfrak{S}_{n-1}^*)$ besteht — aus dem gleichen Grunde wie vorhin — aus endlich vielen, getrennten Streifen, deren erste Abschnitte entweder in \mathfrak{S}_ν liegen oder die überhaupt zu \mathfrak{S}_ν fremd sind. Deshalb bestimmt $\mathfrak{S}_\nu \dot{+} (\mathfrak{S}_{\nu+1}^* \dot{+} \dots \dot{+} \mathfrak{S}_{n-1}^*) + B_\nu$ ⁷⁸⁾ kein beschränktes Komplementärgebiet und demnach muß jede Komponente von $\mathfrak{B}_\nu^{(n-1)}$ in der Tat Bögen von K enthalten.

6. Wir beweisen nun, daß das System der nach den vorstehenden Vorschriften konstruierten Streifen $\mathfrak{S}_1, \mathfrak{S}_2, \mathfrak{S}_3, \dots, \mathfrak{S}_n, \dots$ tatsächlich der Bedingung B genügt.

Es sei bereits für $\nu = 2, 3, \dots, n-1$ bewiesen, daß, wenn ein Abschnitt des Streifens \mathfrak{S}_ν oder des provisorischen Streifens \mathfrak{S}_ν^* auf einem vorhergehenden Streifen \mathfrak{S}_μ ($\mu < \nu$) gelegen ist, der erstere einen niedrigeren „Index“ hinsichtlich Γ hat als der betreffende Abschnitt von \mathfrak{S}_μ . Dann

⁷⁶⁾ Dies tritt z. B. dann ein, wenn $\mathfrak{S}_{\nu+1}$ vollständig in \mathfrak{S}_ν liegt.

⁷⁷⁾ Weil $\mathfrak{S}_{r+\lambda}^*$ entweder ganz innerhalb oder ganz außerhalb \mathfrak{S}_ν^* liegt.

⁷⁸⁾ Diese Menge besteht ja aus endlich vielen getrennt liegenden Polygonbereichen und einem (von diesen getrennt liegenden) einfachen endlichen Streckenzug B_ν .

gilt das Gleiche auf Grund unserer oben für \mathfrak{S}_n^* aufgestellten Konstruktionsvorschriften auch für $\nu = n$, und daher (weil für $\nu = 2$ bewiesen) allgemein.

Wenn also ein Punkt $z = z_0$ von den Streifen $\mathfrak{S}_{n_1}, \mathfrak{S}_{n_2}, \dots$ ($n_1 < n_2 < \dots$) überdeckt wird, so bilden die zugehörigen „Index“-Zahlen der z_0 überdeckenden Abschnitte dieser Streifen eine absteigende Folge: $\alpha_{n_1} > \alpha_{n_2} > \dots$ und diese muß (wegen der Wohlordnung der Ordnungszahlen $\leq \alpha_0$) endlich sein.

§ 9.

Umformung der Bedingung A.

Satz. Mit der Bedingung A ist die folgende Bedingung völlig gleichwertig:

*Γ darf keine Teilmenge Γ^{**} enthalten, die aus lauter eigenen H -Punkten⁶²⁾ relativ zu Γ besteht. Oder anders gesagt: Γ darf keine relativ zu Γ H -in sich dichte Teilmenge Γ^{**} enthalten.*

Aus dem gleichen Grunde, wie an der entsprechenden Stelle des § 8, kann man die vorstehende Bedingung auch so formulieren:

*Γ darf keine relativ zu Γ H -perfekte Teilmenge Γ_{**} enthalten.*

Oder, was hier trotz der Nicht-Abgeschlossenheit von Γ dasselbe ist [wie weiter unten gezeigt wird]:

Γ muß relativ zu sich selbst H -reduzibel sein.

Wir können nun auch hier von Γ zu \mathfrak{C} (oder auch zu $\bar{\Gamma}$) übergehen. Zunächst bemerken wir:

Es kann kein H -Punkt des Randes C relativ zu \mathfrak{C} existieren.

Denn: Wäre P ein H -Punkt von C relativ zu \mathfrak{C} , so müßte in einer beliebig kleinen Umgebung $\mathfrak{U}(P)$ ein von Teilen von C begrenztes Gebiet \mathfrak{S} enthalten sein, das in seinem Innern Punkte von \mathfrak{C} , also auch innere Punkte von \mathfrak{G} enthält. Das einzige Gebiet dieser Art ist aber \mathfrak{G} selbst, wohingegen $\mathfrak{U}(P)$ beliebig klein gewählt werden konnte.

Wir beweisen nun den folgenden

*Hilfssatz. Wenn \mathfrak{C} eine relativ zu \mathfrak{C} ⁷⁹⁾ H -perfekte Teilmenge \mathfrak{C}_{**} enthält, dann enthält auch Γ eine relativ zu Γ H -perfekte Teilmenge Γ_{**} .⁸⁰⁾*

Woraus unmittelbar folgt: *Wenn Γ relativ zu sich selbst H -reduzibel ist, dann ist dasselbe auch für \mathfrak{C} der Fall⁸⁰⁾.*

Beweis. \mathfrak{C}_{**} kann (wegen des eben Bemerkten) nicht in C enthalten sein; also muß \mathfrak{C}_{**} einen nicht zu C gehörenden Punkt P enthalten; auch eine hinreichend kleine Umgebung $\mathfrak{U}(P)$ wird samt ihrer abgeschlossenen

⁷⁹⁾ Man darf hier selbstverständlich „relativ zu \mathfrak{C} “ durch „relativ zu Γ “ ersetzen.

⁸⁰⁾ Die Umkehrung ist selbstverständlich.

Hülle keine Punkte mit C gemeinsam haben. $\mathfrak{D}(\mathfrak{U}(P), \mathfrak{C}_{**}) = \Gamma^{**}$ ist H -in sich dicht relativ zu \mathfrak{C} ; also auch relativ zu Γ , weil $\mathfrak{D}(\mathfrak{U}(P), \mathfrak{C})$ in Γ enthalten ist. Auch Γ^{**} ist ein Teil von Γ . Also enthält Γ eine relativ zu Γ H -perfekte Teilmenge $\Gamma_{**} = \overline{\Gamma^{**}}$.

Wegen dieses Hilfssatzes können wir unsere obige Bedingung auch so formulieren:

\mathfrak{C} darf keine relativ zu \mathfrak{C} ⁷⁹⁾ H -perfekte Teilmenge \mathfrak{C}_{**} enthalten.

Oder, was (wegen der Abgeschlossenheit von \mathfrak{C}) dasselbe ist:

\mathfrak{C} muß relativ zu sich selbst ⁷⁹⁾ H -reduzibel sein ⁸¹⁾.

Wegen der Folgerung aus dem Hilfssatz ergibt sich, daß mit dieser Form der Bedingung gleichwertig ist: „ Γ muß relativ zu sich selbst H -reduzibel sein.“ Diese schon oben angegebene Formulierung ist damit gerechtfertigt. —

Zusammenfassend erkennen wir den Unterschied zwischen den neuen Formen der Bedingungen B und A darin, daß in der Bedingung B die relativ zu \mathfrak{G} gebildeten, in der Bedingung A hingegen die relativ zu Γ (oder \mathfrak{C}) gebildeten H -Begriffe auftreten ⁸²⁾.

Wir beweisen nun wieder die beiden Bestandteile unseres Satzes nacheinander.

1. Teilsatz. Enthält Γ eine relativ zu Γ H -perfekte Teilmenge Γ_{**} , so kann die Bedingung A nicht erfüllt sein.

Beweis. Es sei ein beliebiges System von aus \mathfrak{G} nach K führenden Streifen $\{\mathfrak{S}_\mu\}$ vorgelegt, so beschaffen, daß jede beliebig kleine Umgebung jedes Punktes von Γ von unendlich vielen dieser Streifen getroffen wird. \mathfrak{S}_1 sei ein beschränktes H -Gebiet von Γ_{**} relativ zu Γ und Q_1 ein innerhalb \mathfrak{S}_1 gelegener Punkt von Γ . Dann muß es einen aus beliebiger Nähe von Q_1 nach K führenden Streifen \mathfrak{S}_{n_1} geben, der also den Rand von \mathfrak{S}_1 trifft und demnach in seinem Innern Punkte dieses Randes, d. h. von Γ_{**} enthält. Es sei P_1 ein solcher Punkt und $\mathfrak{U}_1(P_1)$ eine innerhalb \mathfrak{S}_{n_1} gelegene Umgebung von P_1 . In $\mathfrak{U}_1(P_1)$, also innerhalb \mathfrak{S}_{n_1} liegt dann ein H -Gebiet \mathfrak{S}_2 von Γ_{**} relativ zu Γ (samt seinem Rand); im Innern von \mathfrak{S}_2 liege Punkt Q_2 von Γ . In beliebige Nähe von Q_2 muß ein (von \mathfrak{S}_{n_1} verschiedener) Streifen \mathfrak{S}_{n_2} gelangen, der also einen Punkt P_2 des \mathfrak{S}_2 -Randes, d. h. von Γ_{**} , sowie eine ganz im Innern von $\mathfrak{U}_1(P_1)$ gelegene Umgebung $\mathfrak{U}_2(P_2)$ desselben enthält. Wieder in $\mathfrak{U}_2(P_2)$, also auch inner-

⁸¹⁾ In den vorstehenden Formulierungen (und zugehörigen Erläuterungen) kann man auch \mathfrak{C} durchweg durch $\overline{\Gamma}$ ersetzen.

⁸²⁾ Dies steht, wie unmittelbar ersichtlich, in Übereinstimmung damit, daß die Bedingung B mehr fordert als die Bedingung A.

halb $\mathfrak{D}(\mathfrak{S}_{n_1}, \mathfrak{S}_{n_2})$ liegt ein H -Gebiet \mathfrak{S}_3 von Γ_{**} relativ zu Γ (samt Rand) und innerhalb \mathfrak{S}_3 ein Punkt Q_3 von Γ . Usw. für eine unendliche Folge von verschiedenen Indizes n_k ($k=1, 2, 3, \dots$). Ist nun P_0 ein Häufungspunkt der $\{P_k\}$, also ebenfalls zu Γ gehörig, so liegt P_0 in jedem der Gebiete $\mathfrak{U}_k(P_k)$ und somit auch in jedem der Streifen \mathfrak{S}_{n_k} ($k=1, 2, 3, \dots$). D. h. unser Streifensystem $\{\mathfrak{S}_\mu\}$ kann die Bedingung A nicht erfüllen.

2. Teilsatz. *Ist Γ relativ zu sich selbst H -reduzibel, so ist die Bedingung A erfüllt.*

Beweis. Wir betrachten eine beliebige Menge von abzählbar vielen Punkten

$$\zeta_1, \zeta_2, \zeta_3, \dots, \zeta_n, \dots,$$

welche auf Γ überall dicht liegen. Jeder Punkt ζ_n sei Mittelpunkt eines (offenen) Kreisgebietes \mathfrak{f}_n vom Radius $\frac{1}{n}$. Dann ist unser Ziel, diese \mathfrak{f}_n durch geeignete Streifen \mathfrak{S}_n mit K zu verbinden, ohne daß irgendein Punkt von unendlich vielen Streifen \mathfrak{S}_n überdeckt wird.

Das Konstruktionsverfahren verläuft im wesentlichen wörtlich ebenso wie bei dem entsprechenden Beweis des 2. Teilsatzes von § 8. Nur sind alle „ H -Begriffe“ (natürlich auch bei der „Index“-Definition) jetzt „relativ zu Γ “ statt — wie früher — „relativ zu \mathfrak{G} “ zu nehmen. Ferner sind die Gebiete \mathfrak{G}_n durch die \mathfrak{f}_n zu ersetzen. Außerdem wollen wir stets dafür sorgen, daß kein Streifenufer irgendeines Streifens \mathfrak{S}_n durch irgendeinen der (ja nur abzählbar vielen) ausgezeichneten Punkte ζ_ν hindurchläuft.

Die Änderung eines Details des Beweises wird nur an einer Stelle in 3c) und 5c) nötig: In 5c) [und entsprechend für $n=2$ an der analogen Stelle von 3c)] ist S. 248, Z. 8 v. u. bis Z. 5 v. u. „ P sei nun ein Punkt von \mathfrak{G}_n, \dots Komplementärgebiet \mathfrak{Q} von $\Gamma_{\mathfrak{G}}^{[a]}$ an;“ zu ersetzen durch: „ ζ_n gehört dem Innern eines Abschnitts \mathfrak{A} von \mathfrak{S}_ρ an; ein in \mathfrak{f}_n enthaltenes, hinreichend kleineres, konzentrisches Kreisgebiet \mathfrak{f}_n^* liegt ebenfalls ganz innerhalb \mathfrak{A} . Der ‚Index‘ von \mathfrak{A} hinsichtlich Γ (relativ zu Γ) sei α (≥ 0). Nun gehört \mathfrak{f}_n^* einem Komplementärgebiet \mathfrak{Q} von $(\Gamma_{\Gamma}^{[a]} - \mathfrak{D}(\mathfrak{f}_n^*, \Gamma_{\Gamma}^{[a]}))$ an.“ Und im darauffolgenden Teil von 5c) ist $\mathfrak{U}(P)$ durch \mathfrak{f}_n zu ersetzen; (nur auf S. 250, Z. 9 ersetze man $\mathfrak{U}(P)$ durch „den Anfang von \mathfrak{T}_n “); ferner soll \mathfrak{f}_n^* von \mathfrak{T}_n nicht getroffen werden.

§ 10.

Neue Umformung der Bedingungen A und B.

Bedeutet M eine beliebige abgeschlossene Menge der Ebene, so wurde in § 7 (S. 238) jedes ebene Gebiet, dessen Rand ausschließlich aus Punkten von M besteht, als ein „ H -Gebiet von M “ bezeichnet. Eine

beschränkte Menge, deren abgeschlossene Hülle überhaupt kein beschränktes H -Gebiet bestimmt⁸³⁾ (m. a. W. keinen Punkt der Ebene einschließt), möge nun „ H -elementar“ heißen. Jede Teilmenge einer H -elementaren Menge ist wieder H -elementar.

Des weiteren werde eine beschränkte Menge M , die entweder selbst keinen H -Punkt bestimmt (deren H -Ableitung also leer ist) oder als Vereinigungsmenge von endlich vielen Mengen dieser Art dargestellt werden kann, als „ H -endlich“, und jede Menge M , die als Vereinigungsmenge von (höchstens) abzählbar vielen H -endlichen Mengen darstellbar ist, als „ H -abzählbar“ bezeichnet.

Die H -endlichen Mengen sind alsdann identisch mit den Vereinigungsmengen von endlich vielen, die H -abzählbaren Mengen mit den Vereinigungsmengen von (höchstens) abzählbar vielen H -elementaren Mengen.

Dies folgt sofort daraus, daß einerseits jede H -elementare Menge zugleich H -endlich ist, andererseits eine beschränkte Menge M , die keinen H -Punkt besitzt, stets als Vereinigungsmenge von endlich vielen H -elementaren Mengen dargestellt werden kann. Letzteres ist folgendermaßen ersichtlich: Bestimmt die beschränkte Menge M und daher auch ihre abgeschlossene Hülle \bar{M} keinen H -Punkt, so müssen die Durchmesser aller H -Gebiete von \bar{M} größer als eine gewisse positive Zahl δ sein. Überdeckt man also die Ebene mit einem quadratischen Netz von der Seitenlänge $\frac{\delta}{2}$, so ist derjenige Teil von \bar{M} , der einem dieser Quadrate (einschließlich Rand) angehört, abgeschlossen und H -elementar.

Wir bemerken noch: Die *abgeschlossenen* H -endlichen bzw. H -abzählbaren Mengen sind gleichzeitig auch Vereinigungsmengen von endlich bzw. (höchstens) abzählbar vielen *abgeschlossenen* H -elementaren Mengen. (Man kann nämlich nunmehr im vorigen Satz jede H -elementare Teilmenge durch ihre abgeschlossene Hülle ersetzen.)

Bezeichnet man nun ferner als „ H -isoliert“ eine Menge M , deren sämtliche Punkte von der H -Ordnung 0 sind (welche also mit $M^{[1]}$ keinen Punkt gemeinsam hat), so gelten die beiden folgenden Sätze (die zu bekannten Sätzen der Cantorsche Theorie analog sind):

Jede H -isolierte Menge ist H -abzählbar.

Denn: Ist M eine H -isolierte Menge und ist ihre H -Ableitung $M^{[1]}$ nicht leer, so hat jeder einzelne Punkt der Menge M einen positiven Abstand von $M^{[1]}$ und man kann daher setzen:

$$M = M_1 \dot{+} M_2 \dot{+} \dots \dot{+} M_n \dot{+} \dots,$$

⁸³⁾ Nicht-beschränkte H -Gebiete sind selbstverständlich stets vorhanden; dieselben können jedoch (vgl. Fußnote ⁵³⁾) bei der Bestimmung der H -Punkte ohnehin außer Betracht bleiben.

wobei M_ν die Gesamtheit derjenigen Punkte von M bedeute, für welche dieser Abstand $\geq \frac{1}{\nu}$ ist. M_ν ist dann eine Menge ohne H -Punkt (also H -endlich, falls beschränkt, andernfalls H -abzählbar); denn ein H -Punkt von M_ν wäre auch H -Punkt von M , also zu $M^{[1]}$ gehörig, während doch M_ν von $M^{[1]}$ einen positiven Abstand ($\geq \frac{1}{\nu}$) besitzt.

Eine beliebige Menge M ist dann und nur dann H -reduzibel, wenn ihre abgeschlossene Hülle \bar{M} H -abzählbar ist⁸⁴. (Jede H -reduzible Menge ist also H -abzählbar.)

Beweis. 1. Ist M H -reduzibel und ist $M^{[\alpha]}$ die erste H -Ableitung, die leer ist, so stellt für jede einzelne der Bedingung $0 \leq \rho < \alpha$ genügende (endliche oder transfinit) Zahl ρ die Gesamtheit derjenigen Punkte von \bar{M} , welche die H -Ordnung ρ besitzen, eine H -isolierte und daher auch H -abzählbare Menge dar, so daß \bar{M} ebenfalls H -abzählbar sein muß.

2. Ist M nicht H -reduzibel, enthält also \bar{M} einen H -perfekten Bestandteil N und wäre \bar{M} und daher auch N H -abzählbar, etwa

$$N = N_1 \dot{+} N_2 \dot{+} \dots \dot{+} N_\nu \dot{+} \dots,$$

wobei die N_ν H -elementare Mengen bedeuten, die als abgeschlossen angenommen werden können, so müßte es einen Punkt P_1 von N geben, der nicht zu N_1 gehört, und daher auch eine von Punkten von N_1 freie Umgebung $\mathfrak{U}_1(P_1)$ desselben. Da der in $\mathfrak{U}_1(P_1)$ gelegene Teil von N jedenfalls den H -Punkt P_1 besitzt, N_3 hingegen keinen H -Punkt, so müßte des weiteren im Innern von $\mathfrak{U}_1(P_1)$ ein Punkt P_2 von N existieren, der nicht zu N_2 gehört, also auch eine Umgebung $\mathfrak{U}_2(P_2)$, die samt ihrem Rand im Innern von $\mathfrak{U}_1(P_1)$ liegt und keinen Punkt von N_2 enthält; sodann eine Umgebung $\mathfrak{U}_3(P_3)$, die samt ihrem Rand im Innern von $\mathfrak{U}_2(P_2)$ liegt und auch keinen Punkt von N_3 enthält; usw. Bedeutet dann P_0 einen Häufungspunkt von P_1, P_2, \dots , so würde P_0 der Menge N angehören, hingegen, da in jedem $\mathfrak{U}_\nu(P_\nu)$ gelegen, keiner der Mengen N_ν ($\nu = 1, 2, 3, \dots$).

Für abgeschlossene Mengen kann der vorige Satz offenbar auch in der folgenden Form ausgesprochen werden:

Eine beliebige abgeschlossene Menge M ist dann und nur dann H -reduzibel, wenn sie als Vereinigungsmenge von (höchstens) abzählbar vielen abgeschlossenen H -elementaren Mengen darstellbar ist.

⁸⁴) Beispiel, bei welchem M H -abzählbar ist, \bar{M} dagegen nicht (also M nicht H -reduzibel): $M =$ Vereinigungsmenge von abzählbar vielen parallelen Strecken, welche in einem Rechteck überall dicht liegen.

Die vorstehenden Überlegungen und Sätze bleiben sämtlich gültig, wenn man an Stelle der absoluten H -Begriffe überall die *relativen H -Begriffe* — relativ zu einer beliebigen Menge L — treten läßt.^{84a)}

Da nun gemäß § 8 die Bedingung B in der Form ausgesprochen werden konnte: „ $\bar{\Gamma}$ (oder \mathfrak{C}) muß relativ zu \mathfrak{G} H -reduzibel sein“ und gemäß § 9 die Bedingung A in der Form: „ $\bar{\Gamma}$ (oder \mathfrak{C}) muß relativ zu sich selbst H -reduzibel sein“, so folgt nunmehr:

Mit der Bedingung B ist folgende Bedingung gleichwertig: $\bar{\Gamma}$ (oder \mathfrak{C}) muß relativ zu \mathfrak{G} H -abzählbar sein; oder auch: als Vereinigungsmenge von (höchstens) abzählbar vielen abgeschlossenen Mengen darstellbar sein, deren jede relativ zu \mathfrak{G} H -elementar ist (m. a. W. keinen Punkt von \mathfrak{G} einschließt).

Mit der Bedingung A ist folgende Bedingung gleichwertig: $\bar{\Gamma}$ (oder \mathfrak{C}) muß relativ zu sich selbst H -abzählbar sein; oder auch: als Vereinigungsmenge von (höchstens) abzählbar vielen abgeschlossenen Mengen darstellbar sein, deren jede relativ zu $\bar{\Gamma}$ (bzw. zu \mathfrak{C}) H -elementar ist (m. a. W. keinen Punkt von $\bar{\Gamma}$ bzw. von \mathfrak{C} einschließt)⁸⁵⁾.

Beide Aussprüche behalten, wie noch gezeigt werden soll, ihre Gültigkeit auch dann noch unverändert bei, wenn darin $\bar{\Gamma}$ überall durch Γ selbst ersetzt wird. Des weiteren kann in der neuen Form der Bedingung B die Forderung, daß die einzelnen Teilmengen „keinen Punkt von \mathfrak{G} einschließen“ sollen, auch ersetzt werden durch die, daß dieselben \mathfrak{G} bzw. die Ebene nicht zerlegen sollen. Um dies nachzuweisen, seien einige Hilfsbetrachtungen vorausgeschickt (und zwar im Hinblick auf ihre Anwendung in § 11 sogleich in einer etwas allgemeineren Form).

Bezeichnet man (mit F. Hausdorff⁸⁶⁾) die Vereinigungsmenge von (höchstens) abzählbar vielen abgeschlossenen Mengen kurz als „ F_σ -Menge“, so gilt:

Die H -abzählbaren F_σ -Mengen sind identisch mit den Vereinigungsmengen von (höchstens) abzählbar vielen abgeschlossenen H -elementaren Mengen.

^{84a)} Die Aussage, daß eine beschränkte Menge M relativ zu L H -elementar sei, ist offenbar gleichbedeutend mit der, daß \bar{M} keinen Punkt von L einschließt.

⁸⁵⁾ Bei unserem Beispiel S. 224—225 liegt der allereinfachste Fall vor: Dort ist $\bar{\Gamma}$ selbst schon relativ zu $\bar{\Gamma}$ H -elementar (während es relativ zu \mathfrak{G} einen H -perfekten Bestandteil enthält). Denn in unserm Beispiel kann die Umgebung jedes beliebigen Punktes P_0 von $\bar{\Gamma}$ mit dem äußeren Kreis C durch einen $\bar{\Gamma}$ nicht treffenden Streifen verbunden werden. Es kann also keinen Punkt P_0 von $\bar{\Gamma}$ geben, der im Innern eines von $\bar{\Gamma}$ begrenzten Gebietes liegt.

⁸⁶⁾ Grundzüge der Mengenlehre, Leipzig 1914, S. 305.

Denn für eine H -abzählbare F_σ -Menge M hat man einerseits:

$$M = A_1 \dot{+} A_2 \dot{+} \dots \dot{+} A_\nu \dot{+} \dots,$$

wobei die A_ν abgeschlossen sind; andererseits:

$$M = M_1 \dot{+} M_2 \dot{+} \dots \dot{+} M_\mu \dot{+} \dots,$$

wobei die M_μ H -elementar sind. Nun ist⁸⁷⁾

$$A_\nu = (A_\nu \cdot M_1) \dot{+} (A_\nu \cdot M_2) \dot{+} \dots \dot{+} (A_\nu \cdot M_\mu) \dot{+} \dots,$$

wobei die rechtsstehenden Summanden H -elementar sind; ersetzt man diese durch ihre abgeschlossenen Hüllen:

$$A_\nu = \overline{(A_\nu \cdot M_1)} \dot{+} \overline{(A_\nu \cdot M_2)} \dot{+} \dots \dot{+} \overline{(A_\nu \cdot M_\mu)} \dot{+} \dots,$$

so ergibt sich:

$$M = \sum_{\nu, \mu} \overline{(A_\nu \cdot M_\mu)}.$$

Hingegen kann der oben bewiesene Satz, daß eine abgeschlossene Menge dann und nur dann H -reduzibel ist, wenn sie als Vereinigungsmenge von abzählbar vielen abgeschlossenen H -elementaren Mengen darstellbar ist, in dieser Form nicht auf F_σ -Mengen ausgedehnt werden; [denn die Vereinigung von abzählbar vielen abgeschlossenen H -elementaren Mengen braucht nicht H -reduzibel zu sein⁸⁸⁾]. Wohl aber dann, wenn man, was ja gestattet ist, in dem Ausspruch des Satzes die Eigenschaft von M , H -reduzibel zu sein, ersetzt durch die, keinen H -perfekten Bestandteil zu enthalten. Es gilt nämlich allgemein:

Eine beliebige F_σ -Menge M enthält dann und nur dann keinen H -perfekten Bestandteil, wenn sie als Vereinigungsmenge von (höchstens) abzählbar vielen abgeschlossenen H -elementaren Mengen darstellbar ist (oder, was nach obigem dasselbe bedeutet, wenn sie H -abzählbar ist)^{88a)}.

Denn: Es sei $M = A_1 \dot{+} A_2 \dot{+} \dots \dot{+} A_\nu \dots$, wobei die A_ν abgeschlossen sind. 1. Wenn M keinen H -perfekten Bestandteil enthält, so enthält auch kein A_ν einen H -perfekten Bestandteil; dann ist also nach dem Vorstehenden jedes A_ν Vereinigungsmenge von (höchstens) abzählbar vielen ab-

⁸⁷⁾ Wir schreiben hier und im folgenden — wie vielfach üblich — den Durchschnitt $\mathfrak{D}(A, B)$ zweier Mengen einfach als Produkt $A \cdot B$.

⁸⁸⁾ Siehe das Beispiel in Fußnote ⁸⁴⁾.

^{88a)} Dagegen braucht eine Menge *beliebiger Art*, welche keinen H -perfekten Bestandteil besitzt, nicht notwendig H -abzählbar zu sein. Zerlegt man z. B. eine H -perfekte (und somit auch perfekte) Menge M in zwei Teilmengen M_1, M_2 , welche keinen perfekten (und daher auch keinen H -perfekten) Bestandteil besitzen, so können M_1 und M_2 trotzdem nicht beide H -abzählbar sein, da es sonst auch M selbst wäre, letzteres im Widerspruch mit dem Satze auf S. 255.

geschlossenen H -elementaren Mengen; daraus folgt auch für M selbst eine Darstellung als derartige Vereinigungsmenge. — 2. Wenn M einen H -perfekten Bestandteil N enthält, so kann nicht $M = M_1 \dot{+} M_2 \dot{+} \dots \dot{+} M_\nu \dot{+} \dots$ sein, wobei die M_ν abgeschlossen und H -elementar sind; denn sonst wäre auch

$$N = (N \cdot M_1) \dot{+} (N \cdot M_2) \dot{+} \dots \dot{+} (N \cdot M_\nu) \dot{+} \dots,$$

wobei die rechtsstehenden Summanden abgeschlossen und H -elementar wären, was nach dem früheren Satz (S. 255) nicht möglich ist.

Auch die letzten beiden Sätze bleiben gültig, wenn die darin vorkommenden H -Begriffe durchweg durch die relativen H -Begriffe ersetzt werden.

Ist M eine nirgends dichte abgeschlossene Menge, so ist die Aussage, daß M H -elementar sei, wie unmittelbar ersichtlich, gleichbedeutend mit der, daß M beschränkt sei und die Ebene nicht zerlege⁸⁹⁾. Aus dem vorigen Satz folgt daher:

Eine nirgends dichte F_σ -Menge enthält dann und nur dann keinen H -perfekten Bestandteil, wenn sie als Vereinigungsmenge von (höchstens) abzählbar vielen beschränkten, abgeschlossenen Mengen dargestellt werden kann, welche die Ebene nicht zerlegen.

Da nun gemäß § 8 die Bedingung B in der Form „ Γ darf keine H -perfekte Teilmenge enthalten“ ausgesprochen werden kann, so ergibt sich mit Berücksichtigung des Umstandes, daß Γ eine F_σ -Menge ist⁹⁰⁾, aus den beiden vorangehenden Sätzen nunmehr folgendes:

Mit der Bedingung B ist jede der folgenden Bedingungen gleichwertig: Γ muß H -abzählbar oder auch als Vereinigungsmenge von (höchstens) abzählbar vielen abgeschlossenen H -elementaren Mengen darstellbar sein oder endlich als Vereinigungsmenge von (höchstens) abzählbar vielen abgeschlossenen Mengen, welche die Ebene nicht zerlegen.

Aus dem vorigen Satz über nirgends dichte F_σ -Mengen geht in Verbindung mit unseren früheren Beispielen⁹¹⁾ noch hervor, daß es abgeschlossene Mengen und sogar solche, die ein einfach zusammenhängendes

⁸⁹⁾ Wir sagen: Eine abgeschlossene Menge M „zerlegt die Ebene“, wenn ihre Komplementärmenge in mindestens zwei Komponenten zerfällt. („Komponente“ einer offenen Menge = Teilgebiet, das in keinem anderen enthalten ist.)

⁹⁰⁾ Nämlich $\Gamma = \gamma_1 \dot{+} \gamma_2 \dot{+} \dots$, wo γ_n die Menge derjenigen Punkte von Γ bedeutet, deren Abstand vom Rand C des Gebietes \mathcal{G} nicht kleiner als $\frac{1}{n}$ ist.

⁹¹⁾ Bei dem Beispiel von S. 224–225 ist der Rand des Gebietes \mathcal{G}_1 mit \mathcal{G} identisch; da nun die Bedingung B nicht erfüllt ist, so besitzt nach § 8 Γ und somit auch \mathcal{G} einen H -perfekten Bestandteil.

Gebiet begrenzen, gibt, welche nicht als Vereinigungsmengen von abzählbar vielen abgeschlossenen, die Ebene nicht zerlegenden Mengen darstellbar sind.

Der vorangehende, auf nirgends dichte F_σ -Mengen bezügliche Satz kann ebenfalls auf eine Relativform gebracht und zwar *relativ zu einem beliebigen Gebiet \mathfrak{G}* formuliert werden. Er behält nämlich seine Gültigkeit unverändert bei, wenn darin die Worte „ H -perfekt“ durch „*relativ zu \mathfrak{G} H -perfekt*“ und gleichzeitig „welche die Ebene nicht zerlegen“ durch „*welche \mathfrak{G} nicht zerlegen*“ ersetzt werden. Um dies festzustellen, genügt es offenbar, folgende Aussage zu beweisen:

Eine nirgends dichte F_σ -Menge ist dann und nur dann relativ zu \mathfrak{G} H -abzählbar, wenn sie als Vereinigungsmenge von (höchstens) abzählbar vielen, beschränkten, abgeschlossenen Mengen darstellbar ist, welche \mathfrak{G} nicht zerlegen⁹²⁾.

Denn: Einerseits läßt sich jede abgeschlossene, relativ zu \mathfrak{G} H -elementare Menge M sofort folgendermaßen als Vereinigungsmenge von abzählbar vielen abgeschlossenen, \mathfrak{G} nicht zerlegenden Mengen darstellen:

$$M = [M - (\mathfrak{G} \cdot M)] + \sum_{\nu} (\mathfrak{P}_\nu \cdot M),$$

wobei $\{\mathfrak{P}_\nu\}$ ($\nu = 1, 2, 3, \dots$) eine Folge von abgeschlossenen Polygonflächen bezeichnet, welche \mathfrak{G} von innen her approximieren. — Andererseits ist jede nirgends dichte, beschränkte, abgeschlossene, \mathfrak{G} nicht zerlegende Menge M als Vereinigung von endlich vielen abgeschlossenen, relativ zu \mathfrak{G} H -elementaren Mengen darstellbar; ist nämlich g eine \mathfrak{G} treffende Gerade und bezeichnet \bar{g}_1 und \bar{g}_2 die beiden von g bestimmten, abgeschlossenen Halbebenen, so ist

$$M = [M - (\mathfrak{G} \cdot M)] \cdot \bar{g}_1 + [M - (\mathfrak{G} \cdot M)] \cdot \bar{g}_2 + (\bar{\mathfrak{G}} \cdot M) \cdot \bar{g}_1 + (\bar{\mathfrak{G}} \cdot M) \cdot \bar{g}_2.$$

Für die Bedingung B, welche in § 8 in der Form „ Γ (oder \mathfrak{C} oder auch $\bar{\Gamma}$) darf keine relativ zu \mathfrak{G} H -perfekte Teilmenge enthalten“ ausgesprochen wurde, ergibt sich somit (wieder mit Berücksichtigung des Umstandes, daß Γ eine F_σ -Menge ist) schließlich noch folgende, gegenüber der obigen (S. 256) vervollständigte Formulierung:

Mit der Bedingung B ist jede der folgenden Bedingungen gleichwertig: $\bar{\Gamma}$ [oder \mathfrak{C} oder auch Γ] muß relativ zu \mathfrak{G} H -abzählbar oder auch als Vereinigungsmenge von (höchstens) abzählbar vielen abgeschlossenen, relativ zu \mathfrak{G} H -elementaren Mengen darstellbar sein oder

⁹²⁾ „Die abgeschlossene Menge M zerlegt das Gebiet \mathfrak{G} “ bedeute dabei, daß \mathfrak{G} nach Weglassung der in \mathfrak{G} enthaltenen Punkte von M in mindestens zwei Komponenten zerfällt.

auch als Vereinigungsmenge von (höchstens) abzählbar vielen abgeschlossenen Mengen, welche \mathfrak{G} nicht zerlegen.

Auch bei der oben (S. 256) ausgesprochenen Form der Bedingung A darf $\bar{\Gamma}$ nunmehr durch Γ selbst ersetzt werden.

Als spezielle Folgerung aus diesem (die Bedingung B betreffenden) Satze ergibt sich noch, daß jede abgeschlossene β -Menge (siehe § 6) als Vereinigungsmenge von höchstens abzählbar vielen abgeschlossenen Mengen darstellbar ist, welche die Ebene nicht zerlegen⁹³). Denn: Eine abgeschlossene β -Menge kann zunächst als Vereinigungsmenge von höchstens abzählbar vielen beschränkten abgeschlossenen β -Mengen dargestellt werden. Ist M eine Menge dieser letzteren Art (also nirgends dicht) und ist C eine Kreislinie, welche M umschließt, so kann man auf die aus M und C bestehende Figur den Satz von S. 236 anwenden, indem man das Innere der Kreislinie C die Rolle des Gebietes \mathfrak{G} und M die Rolle von Γ_0 spielen⁹⁴) läßt. Mithin ist für diese Figur die Bedingung B erfüllt und daher nach dem voranstehenden Satze⁹⁵) M als Vereinigungsmenge von (höchstens) abzählbar vielen abgeschlossenen Mengen darstellbar, welche unser \mathfrak{G} und also auch die Ebene nicht zerlegen.

§ 11.

Verallgemeinerung auf beliebige abgeschlossene Mengen, die die Ebene nicht zerlegen.

Im vorigen Paragraphen haben wir gezeigt: In der Ebene enthält eine nirgends dichte F_σ -Menge M dann und nur dann keinen H -perfekten Bestandteil, wenn M als Vereinigungsmenge von (höchstens) abzählbar vielen beschränkten, abgeschlossenen Mengen darstellbar ist, welche die Ebene nicht zerlegen. Wir wollen hier eine analoge Aussage für beliebige F_σ -Mengen (die nicht als nirgends dicht vorausgesetzt werden) gewinnen. Dies wird zwar nicht mehr auf unsere funktionentheoretischen Untersuchungen Bezug haben; aber es wird uns die Beantwortung der folgenden topologischen Frage liefern: Unter welchen Bedingungen läßt sich eine Menge M der Ebene als Vereinigungsmenge von höchstens abzählbar vielen abgeschlossenen Mengen darstellen, welche die Ebene nicht zerlegen?

⁹³) Was beim Beweis des Satzes 2 von § 6 eine Rolle spielte (und dort vorgehend erwähnt wurde).

⁹⁴) Dies ist zulässig, da Γ_0 definitionsgemäß (§ 5) keiner anderen Einschränkung unterliegt als nirgends dicht und zusammen mit C abgeschlossen zu sein.

⁹⁵) Dieser ist nämlich auch dann noch gültig, wenn an Stelle der Zerlegung von \mathfrak{G} durch Γ in die Gebiete \mathfrak{G}_ν eine Zerlegung von \mathfrak{G} durch Γ_0 in (eventuell mehrfach zusammenhängende) Gebiete \mathfrak{G}_ν (im Sinn von § 5) zugrunde gelegt wird; vgl. auch ⁶¹).

Ist M irgendeine Menge der Ebene, so bezeichne $K(M)$ die Komplementärmenge von M . Für eine abgeschlossene Menge N ist die Aussage, N sei relativ zu $K(N)$ H -elementar, offenbar gleichbedeutend mit der Aussage, daß N beschränkt sei und die Ebene nicht zerlege.

Ist ferner N eine abgeschlossene Teilmenge einer beliebigen Menge M , so ist die Aussage, N sei relativ zu $K(M)$ H -elementar, gleichbedeutend mit der Aussage, daß N beschränkt sei und daß eine N enthaltende, beschränkte, abgeschlossene Teilmenge N^* von M existiere, welche die Ebene nicht zerlegt. Denn ist N relativ zu $K(M)$ H -elementar, so hat entweder N selbst schon die Eigenschaft, die Ebene nicht zu zerlegen, oder nicht. Im letzteren Falle mögen die beschränkten, durch N bestimmten Komplementärgebiete (deren mindestens eins vorhanden ist) mit g_ν ($\nu = 1, 2, \dots$) bezeichnet werden. Jedes g_ν muß dann vollständig zu M gehören, da dasselbe sonst ein H -Gebiet von N relativ zu $K(M)$ wäre. Die beschränkte, abgeschlossene Menge $N^* = N + \sum_\nu g_\nu$ besitzt dann kein beschränktes Komplementärgebiet, hat also die behauptete Eigenschaft. Gibt es umgekehrt eine N enthaltende, beschränkte, abgeschlossene Teilmenge N^* von M , welche die Ebene nicht zerlegt, so ist N^* und daher auch N relativ zu $K(N^*)$ H -elementar, also erst recht relativ zu $K(M)$.

Enthält nun eine F_σ -Menge M keine relativ zu $K(M)$ H -perfekte Teilmenge, so gibt es nach dem (relativ zu nehmenden) Satz von S. 257 (§ 10) eine Darstellung

$$M = N_1 \dot{+} N_2 \dot{+} \dots \dot{+} N_\nu \dot{+} \dots,$$

wobei die N_ν abgeschlossene, relativ zu $K(M)$ H -elementare Mengen bedeuten. Ersetzt man jede derselben durch die ihr im obigen Sinn zugeordnete Menge N_ν^* , so ist M als Vereinigungsmenge von (höchstens) abzählbar vielen beschränkten, abgeschlossenen Mengen dargestellt, welche die Ebene nicht zerlegen. — Ist umgekehrt eine beliebige Menge M in der letztgenannten Weise dargestellt, etwa

$$M = M_1 \dot{+} M_2 \dot{+} \dots \dot{+} M_\nu \dot{+} \dots,$$

so ist M_ν relativ zu $K(M_\nu)$ und daher auch relativ zu $K(M)$ H -elementar, also enthält (wieder nach dem soeben zitierten Satz von § 10) M keine relativ zu $K(M)$ H -perfekte Teilmenge. — Wir haben somit den Satz erhalten:

Eine beliebige Menge M der Ebene ist dann und nur dann als Vereinigungsmenge von (höchstens) abzählbar vielen abgeschlossenen Mengen darstellbar, welche die Ebene nicht zerlegen, wenn M eine F_σ -Menge ist und keine relativ zu $K(M)$ H -perfekte Teilmenge enthält.

Schließlich wollen wir noch unsere *Bedingung B* mit dem Vorstehenden in Zusammenhang bringen. Es sei M eine beschränkte, abgeschlossene Menge der Ebene. Durch die Worte „die Bedingung B bezieht sich auf die Menge der von M bestimmten Komplementärgebiete“ oder, was dasselbe ist, „die Bedingung B bezieht sich auf die Komponenten von $K(M)$ “ soll im folgenden ausgedrückt werden, daß in dem Wortlaut der Streifenbedingung B auf S. 221–222 die $\{\mathfrak{G},\}$ durch die von M bestimmten Komplementärgebiete [also durch die Komponenten von $K(M)$] und gleichzeitig Γ durch M zu ersetzen ist. Unserem Hauptsatz von § 8, wonach die Bedingung B gleichwertig ist mit der Forderung, daß Γ [oder \mathfrak{G}] keine relativ zu \mathfrak{G} H -perfekte Teilmenge enthält, entspricht dann der folgende analoge Satz⁹⁶):

Für eine beschränkte⁹⁷, abgeschlossene Menge M ist die auf die Komponenten von $K(M)$ sich beziehende Bedingung B gleichwertig mit der Forderung, daß M keinen relativ zu $K(M)$ H -perfekten Bestandteil enthält.

Der Beweis von § 8 läßt sich nämlich sinngemäß vollständig auf den vorstehenden Satz übertragen⁹⁸). Nur bei der „Flaschenhalsmethode“ wird eine größere Änderung nötig, weil früher Γ nirgends dicht war, während jetzt über M diese Voraussetzung nicht gemacht wird. An Stelle des früheren, nirgends dichten $\Gamma_{\mathfrak{G}}^{[\alpha_0]}$ tritt jetzt $M_{K(M)}^{[\alpha_0]}$. Für $\alpha_0 \geq 1$ ist nun auch $M_{K(M)}^{[\alpha_0]}$ stets nirgends dicht. Denn ein H -Punkt von M relativ zu $K(M)$ ist gleichzeitig Häufungspunkt von M und von $K(M)$, gehört also der Begrenzung von $K(M)$ an; deshalb ist schon $M_{K(M)}^{[1]}$ und erst recht jede höhere Ableitung nirgends dicht.

Für $\alpha_0 = 0$ hingegen ist eine Abänderung unserer Flaschenhalsmethode nötig (weil $M_{K(M)}^{[0]} = \bar{M} = M$ keineswegs nirgends dicht zu sein braucht): Der Streifen sei bereits bis zum Rechteck \mathfrak{R}_s (vgl. S. 246) einschließlich konstruiert; und s_s sei diejenige freie Seite von \mathfrak{R}_s , von der aus der Streifen fortgesetzt werden soll. Man kann wie früher, an s_s anschließend, verfahren, wenn s_s selbst oder eine in beliebig kleinem Abstand außerhalb \mathfrak{R}_s gelegene, parallele Strecke Punkte von $K(M)$ enthält. Andernfalls

⁹⁶) Derselbe umfaßt den früheren als Spezialfall, wenn man in dem Ausspruch dieses letzteren die Worte „relativ zu \mathfrak{G} “ durch „relativ zu $\Sigma \mathfrak{G}$ “ ersetzt, was offenbar gestattet ist.

⁹⁷) Diesen Zusatz „beschränkt“ kann man (durch Anwendung einer geeigneten Inversion, wobei dann natürlich alle Streifen dem Inversionszentrum ferableiben müssen) beseitigen. Dann sind in der Bedingung B die Streifen nach einem Kreis K zu führen, der [nicht, wie früher, M einschließt, sondern] entweder samt seinem Innern oder samt seinem Außern ganz in $K(M)$ enthalten ist.

⁹⁸) Dabei sind \mathfrak{G} , Γ bzw. \mathfrak{G} , durch $K(M)$, M bzw. die Komponenten von $K(M)$ zu ersetzen.

schließt sich im Äußern von \mathfrak{R}_v an s_v ein ganz aus Punkten von M bestehendes Rechteck an; das größte Rechteck dieser Art sei $\mathfrak{R}_v^{(1)}$ ⁹⁹⁾. Man verlängere nun den Streifen, indem man zunächst an \mathfrak{R}_v dieses $\mathfrak{R}_v^{(1)}$ und sodann (an die s_v gegenüberliegende freie Seite $s_v^{(1)}$ von $\mathfrak{R}_v^{(1)}$ anschließend) noch ein zweites Rechteck $\mathfrak{R}_v^{(2)}$ vom Durchmesser $\leq \frac{\delta}{2}$ ansetzt, dessen freie ($s_v^{(1)}$ gegenüberliegende) Seite $s_v^{(2)}$ Punkte von $K(M)$ enthält. Man kann dann in der früher bei der Flaschenhalsmethode üblichen Weise durch Ansetzen eines „Flaschenhalses“ den Streifen von hier aus weiterführen. — In dem durch Anfügen von $\mathfrak{R}_v^{(1)} + \mathfrak{R}_v^{(2)}$ zu \mathfrak{R}_v entstehenden Gebiet kann kein H -Gebiet von M relativ zu $K(M)$ liegen; denn wäre \mathfrak{S} ein solches, so enthielte bereits auch \mathfrak{R}_v oder $\mathfrak{R}_v^{(2)}$ ein solches H -Gebiet, nämlich eine Komponente von $(\mathfrak{S} - \mathfrak{R}_v^{(1)} \cdot \mathfrak{S})$, was nach der Voraussetzung über δ nicht der Fall ist. —

Fassen wir unsere letzten Resultate zusammen, so ergibt sich für irgendeine *beschränkte*⁹⁷⁾, *abgeschlossene Menge* M der Ebene die Gleichwertigkeit der folgenden drei Aussagen:

1. M ist als Vereinigungsmenge von (höchstens) abzählbar vielen abgeschlossenen, die Ebene nicht zerlegenden Mengen darstellbar.

2. M enthält keine relativ zu $K(M)$ H -perfekte Teilmenge oder, was hier dasselbe besagt: M ist relativ zu $K(M)$ H -reduzibel.

3. Für die Menge der durch M bestimmten Komplementärgebiete ist die Bedingung B erfüllt.

Unsere obigen Betrachtungen haben sich — entsprechend unserem funktionentheoretischen Ausgangspunkt — nur auf die Ebene bezogen. Es lassen sich aber alle Überlegungen und Sätze der §§ 7—11, und insbesondere die im letzten Paragraphen enthaltene Untersuchung der angegebenen topologischen Frage, sofort sinngemäß auf n -dimensionale Euklidische Räume übertragen.

⁹⁹⁾ Wird der Streifen im Innern eines polygonalen Bandes \mathfrak{B} konstruiert, so tritt an Stelle von $\mathfrak{R}_v^{(1)}$ eine aus endlich vielen Rechtecken zusammengesetzte, ganz aus Punkten von M bestehende Treppenpolygonfläche, die in ihrer letzten Fortschreitungsrichtung nicht verlängert werden kann, ohne in $K(M)$ einzudringen.

Über die Annäherung einer stetigen Funktion durch die Cesàroschen Mittel ihrer Fourierreihe.

Von

Georg v. Alexits in Budapest.

Einleitung.

Bezeichne $[0, 2\pi]$ das abgeschlossene Intervall $0 \leq x \leq 2\pi$ und sei die nach 2π periodische Funktion

$$f(x) \sim \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx)$$

im Intervalle $[0, 2\pi]$ im Lebesgueschen Sinne integrierbar. Sei außerdem $S_n^{(\delta)}(x)$ die n -te Partialsumme der Cesàroschen Mittel δ -ter Ordnung ihrer Fourierreihe, also im allgemeinen

$$S_n^{(\delta)}(x) = \frac{a_0}{2} + \frac{1}{\binom{n+\delta}{n}} \sum_{k=1}^n \binom{n-k+\delta}{n-k} (a_k \cos kx + b_k \sin kx).$$

Speziell ist dann $S_n^{(0)}(x)$ der n -te Abschnitt der Fourierreihe selbst:

$$S_n^{(0)}(x) = \frac{a_0}{2} + \sum_{k=1}^n (a_k \cos kx + b_k \sin kx)$$

und $S_n^{(1)}(x)$ das gewöhnliche n -te arithmetische Mittel:

$$S_n^{(1)}(x) = \frac{S_0^{(0)}(x) + S_1^{(0)}(x) + \dots + S_n^{(0)}(x)}{n+1}.$$

Über die arithmetischen Mittel $S_n^{(1)}(x)$ hat Herr S. Bernstein¹⁾ die folgenden beiden Sätze aufgestellt:

¹⁾ S. Bernstein, *Mém. Acad. Roy. Belg.* 4 (1912), S. 1—104.

Genügt die Funktion $f(x)$ im Intervalle $[0, 2\pi]$ überall einer Lipschitzbedingung α -ter Ordnung

$$\left| \frac{f(x') - f(x)}{(x' - x)^\alpha} \right| < M \quad (M = \text{Konst.}),$$

so erhält man in $[0, 2\pi]$ gleichmäßig den Annäherungsgrad

$$S_n^{(1)}(x) - f(x) = O\left(\frac{1}{n^\alpha}\right),$$

wenn nur $0 < \alpha < 1$ ist. Im Falle $\alpha = 1$ aber ist

$$S_n^{(1)}(x) - f(x) = O\left(\frac{\log n}{n}\right).$$

Diese Sätze ergeben sich recht einfach, doch lassen sie manche wichtige Frage offen. Wir werden hier eine Verschärfung und Verallgemeinerung der Bernsteinschen Sätze geben, welche zugleich ein neues Licht auf die gleichmäßige Konvergenz der Fourierreihe selbst wirft.

In neuester Zeit hat Verfasser den zweiten Bernsteinschen Satz für die Cesàroschen Mittel irgendeiner positiven Ordnung δ verallgemeinert und zugleich die Lipschitzbedingung durch eine wesentlich weitergehende ersetzt²⁾:

Wenn an einer Stelle x die Bedingung

$$\frac{1}{h} \int_0^h \frac{|f(x+2t) + f(x-2t) - 2f(x)|}{t^\alpha} dt < K \quad (K = \text{Konst.})$$

erfüllt ist, so besteht in diesem Punkte für $\delta \geq \alpha$

$$S_n^{(\delta)}(x) - f(x) = O\left(\frac{\log n}{n^\alpha}\right).$$

Im § 1 der vorliegenden Arbeit werden wir die Bedingung durch eine sehr allgemeine ersetzen, welche beide Bernsteinsche Sätze auf die Mittel positiver Ordnung auszudehnen erlaubt und außerdem hervortreten läßt, daß diese nur einen sehr enggefaßten Spezialfall eines allgemeineren Approximationssatzes bilden. Darüber hinaus zeigen wir im § 3, daß die Fourierreihe einer stetigen Funktion in diesem sehr allgemeinen Falle gleichmäßig konvergiert und einen ebenso guten Annäherungsgrad besitzt, wie es bisher nur im Falle einer differenzierbaren Funktion bekannt³⁾ war.

Im § 2 wird ein Satz über gleichmäßige Annäherungen hergeleitet, welcher an sich nicht ohne Interesse zu sein scheint, da er den Zusammenhang zwischen den Approximationen in diskreten Punkten und zwischen den gleichmäßigen Annäherungen in Intervallen herstellt.

²⁾ G. Alexits, Acta Univ. Szeged 3 (1927), S. 32—37.

³⁾ H. Lebesgue, Bull. Soc. math. France 38 (1910), S. 184—210.

Aus diesem Satze entnehmen wir dann im § 4 einen Beweis, der, sich auf einen mit unseren Betrachtungen aufs engste zusammenhängenden Satz des Herrn Szász⁴⁾ stützend, zur folgenden Verallgemeinerung des zweiten Bernsteinschen Satzes führt:

Ist $f(x)$ in $[0, 2\pi]$ stetig und von beschränkter Variation, außerdem die Derivierte $f'(x)$ in den Punkten einer in ihrer Stetigkeitsmenge dichten Menge gleichmäßig stetig auf $[0, 2\pi]$, so besteht

$$S_n^{(1)}(x) - f(x) = o\left(\frac{\log n}{n}\right).$$

in $[0, 2\pi]$ gleichmäßig.

Hierbei ist bemerkenswert, daß nicht nur der von Herrn Bernstein angegebene Annäherungsgrad $O\left(\frac{\log n}{n}\right)$ durch den besseren $o\left(\frac{\log n}{n}\right)$ ersetzt wird, sondern dieser bessere Annäherungsgrad auch im Falle einer weitergehenden Bedingung, welche die von Herrn Bernstein angegebene enthält, richtig bleibt.

§ 1.

Der Annäherungsgrad der Cesàroschen Mittel im allgemeinen Falle.

1. Zunächst wollen wir $S_n^{(\delta)}(x)$ auf eine wohlbekannte, allgemein übliche Form bringen:

$$S_n^{(\delta)}(x) = \frac{a_0}{2} + \frac{1}{\binom{n+\delta}{n}} \sum_{k=1}^n \binom{n-k-1+\delta}{n-k} S_k^{(0)}(x).$$

Diese Darstellung ergibt infolge der bekannten Beziehung

$$S_k^{(0)}(x) = \frac{1}{\pi} \int_0^\pi f(x+2t) \frac{\sin(2k+1)t}{\sin t} dt$$

die folgende oft benützte Gleichung:

$$(1) \quad S_n^{(\delta)}(x) - f(x) = \frac{1}{\binom{n+\delta}{n}} \int_0^\pi \varphi(t) K_n(t) dt,$$

wobei

$$\varphi(t) = f(x+2t) + f(x-2t) - 2f(x)$$

und

$$K_n(t) = \sum_{k=1}^n \binom{n-k-1+\delta}{n-k} \frac{\sin(2k+1)t}{\sin t}$$

gesetzt wurde.

⁴⁾ O. Szász, Acta math. 48 (1926), S. 353—362.

2. Wir wollen nun annehmen, daß in der Umgebung des Punktes x die folgende Bedingung erfüllt ist:

$$(2) \quad \frac{1}{h^{1+\alpha}} \int_0^h |\varphi(t)| dt = \frac{\Phi(h)}{h^{1+\alpha}} < G(x),$$

wobei $G(x)$ eine Konstante bezeichnet, deren Wert vom Punkte x , nicht aber von n abhängt, und beweisen dann den

Satz I⁵). *Ist die Bedingung (2) im Punkte x erfüllt, so besteht*

$$S_n^{(\delta)}(x) - f(x) = O\left(\frac{1}{n^\alpha}\right)$$

für alle $\delta > \alpha$, im Falle $\delta = \alpha$ erhalten wir aber

$$S_n^{(\delta)}(x) - f(x) = O\left(\frac{\log n}{n^\alpha}\right).$$

Zum Beweise führen wir die Integration in drei Schritten aus:

$$\int_0^\pi = \int_0^{\frac{1}{2n+1}} + \int_{\frac{1}{2n+1}}^{\frac{1}{2}} + \int_{\frac{1}{2}}^\pi.$$

Das erste dieser Integrale läßt sich leicht abschätzen, wenn wir die Ungleichung

$$\left| \frac{\sin(2k+1)t}{\sin t} \right| \leq 2k+1$$

beachten, denn es ist dann

$$\begin{aligned} & \frac{1}{\binom{n+\delta}{n} \pi} \int_0^{\frac{1}{2n+1}} |\varphi(t)| |K_n(t)| dt \\ &= \frac{1}{\binom{n+\delta}{n} \pi} \int_0^{\frac{1}{2n+1}} |\varphi(t)| \left| \sum_{k=1}^n \binom{n-k-1+\delta}{n-k} \frac{\sin(2k+1)t}{\sin t} \right| dt \\ &\leq \frac{1}{\binom{n+\delta}{n} \pi} \int_0^{\frac{1}{2n+1}} |\varphi(t)| \sum_{k=1}^n \binom{n-k-1+\delta}{n-k} \frac{|\sin(2k+1)t|}{\sin t} dt \end{aligned}$$

⁵ In neuester Zeit hat Herr O. Szász, Acta Univ. Szeged 3 (1927), S. 38—49, einige Sätze bewiesen, welche fast dasselbe besagen, was unser Satz I. Die Sätze des Herrn Szász sind jedoch der Form und dem Inhalt nach mehr zur Darstellung des Sprunges der Funktion $f(x)$, als der Abschätzung des Annäherungsgrades zugeschnitten.

$$\leq \frac{\sum_{k=1}^n \binom{n-k-1+\delta}{n-k} (2k+1)}{\binom{n+\delta}{n}^\pi} \int_0^{\frac{1}{2n+1}} |\varphi(t)| dt$$

$$< (2n+1) \frac{\sum_{k=1}^n \binom{n-k-1+\delta}{n-k}}{\binom{n+\delta}{n}^\pi} \int_0^{\frac{1}{2n+1}} |\varphi(t)| dt.$$

Vermöge dieser Ungleichungen folgt aus (2) unmittelbar

$$\frac{1}{\binom{n+\delta}{n}^\pi} \int_0^{\frac{1}{2n+1}} |\varphi(t)| |K_n(t)| dt < \frac{\sum_{k=1}^n \binom{n-k-1+\delta}{n-k}}{\binom{n+\delta}{n} (2n+1)^\alpha \pi} (2n+1)^{1+\alpha} \Phi\left(\frac{1}{2n+1}\right)$$

$$< \frac{\sum_{k=1}^n \binom{n-k-1+\delta}{n-k}}{\binom{n+\delta}{n} (2n+1)^\alpha \pi} G(x).$$

Es ist aber bekannt, daß

$$\binom{n-k+\delta}{n-k} = \binom{n-k-1+\delta}{n-k} + \binom{n-k-1+\delta}{n-k-1}$$

ist, wenn also diese Gleichung auf $k=0, 1, 2, \dots, n$ wiederholt angewendet wird, so folgt:

$$\sum_{k=1}^n \binom{n-k-1+\delta}{n-k} = \binom{n-1+\delta}{n-1}.$$

Die vorangehende Ungleichung ergibt daher die folgende Abschätzung:

$$(3) \quad \frac{1}{\binom{n+\delta}{n}^\pi} \int_0^{\frac{1}{2n+1}} |\varphi(t)| |K_n(t)| dt \leq \frac{G_1(x)}{(2n+1)^\alpha},$$

wobei $G_1(x)$ und im folgenden $G_2(x), G_3(x), \dots$ solche Konstanten bedeuten, welche nur von der Stelle x , nicht aber von n abhängen.

Um das dritte Integral abzuschätzen, beachten wir die folgende Beziehung⁶⁾:

⁶⁾ S. Chapman, London Proc. (2) 9 (1911), S. 369—409 und M. Riesz, Acta Univ. Szeged 1 (1923), S. 104—113.

$$(4) \quad \frac{|K_n(t)|}{\binom{n+\delta}{n}} < \frac{H}{n^\delta t^{1+\delta}} \quad (H = \text{Konst.})$$

und erhalten dann

$$\begin{aligned} \frac{1}{\binom{n+\delta}{n} x} \int_h^\pi |\varphi(t)| |K_n(t)| dt &< \frac{H}{n^\delta x} \int_h^\pi |\varphi(t)| \frac{dt}{t^{1+\delta}} \\ &\leq \frac{H h^{1+\delta}}{n^\delta x} \int_h^\pi |\varphi(t)| dt < \frac{\text{Konst.}}{n^\delta}. \end{aligned}$$

Es ergibt sich also, je nachdem $\alpha < \delta$ oder $= \delta$ ist:

$$(5) \quad \begin{aligned} \frac{1}{\binom{n+\delta}{n} x} \int_h^\pi |\varphi(t)| |K_n(t)| dt &= o\left(\frac{1}{n^\alpha}\right) & (\alpha < \delta); \\ &= O\left(\frac{1}{n^\alpha}\right) & (\alpha = \delta). \end{aligned}$$

Wir bemerken noch, daß (5) von x unabhängig in $[0, 2\pi]$ gleichmäßig erfüllt ist.

3. Zur Berechnung des zweiten Integrals nehmen wir zunächst $\alpha < \delta$ an. Durch partielle Integration und durch die Anwendung der Ungleichung (4) folgt dann:

$$\begin{aligned} \frac{1}{\binom{n+\delta}{n} \frac{1}{2n+1}} \int_1^h |\varphi(t)| |K_n(t)| dt &< \frac{H}{n^\delta \pi} \int_1^h \frac{|\varphi(t)|}{t^{1+\delta}} dt \\ &= \frac{H}{n^\delta \pi} \left[\frac{\Phi(t)}{t^{1+\delta}} \right]_1^h + \frac{H(1+\delta)}{n^\delta \pi} \int_1^h \frac{\Phi(t)}{t^{2+\delta}} dt \\ &= \frac{H}{n^\delta \pi} \left[\frac{\Phi(t)}{t^{1+\alpha}} \cdot \frac{t^{\alpha-\delta}}{t^\delta} \right]_1^h + \frac{H(1+\delta)}{n^\delta \pi} \cdot \frac{\Phi(\tau)}{\tau^{1+\alpha}} \int_1^h \frac{dt}{t^{1+\delta-\alpha}}. \end{aligned}$$

Nach (2) ist aber

$$\frac{H}{n^\delta \pi} \left[\frac{\Phi(t)}{t^{1+\alpha}} \cdot \frac{t^{\alpha-\delta}}{t^\delta} \right]_1^h < \frac{HG(x)}{n^\delta \pi} \left(\frac{h^\alpha}{h^\delta} + \frac{(2n+1)^\delta}{(2n+1)^\alpha} \right),$$

wie auch

$$\frac{\Phi(\tau)}{\tau^{1+\alpha}} < G(x),$$

da τ ein Punkt aus dem Intervalle $\left[\frac{1}{2n+1}, h\right]$ ist, daher folgt:

$$\frac{1}{\binom{n+\delta}{n} \pi} \int_{\frac{1}{2n+1}}^h |\varphi(t)| |K_n(t)| dt < G(x) \left[\left(\frac{k_1}{n^\delta} + \frac{k_2}{(2n+1)^\alpha} \right) + \left(\frac{k_3}{n^\delta} + \frac{k_4}{(2n+1)^\alpha} \right) \right].$$

Wenn wir also unser Übereinkommen, daß $\alpha < \delta$ ist, beachten, so ergibt sich endlich:

$$(6) \quad \frac{1}{\binom{n+\delta}{n} \pi} \int_{\frac{1}{2n+1}}^h |\varphi(t)| |K_n(t)| dt < \frac{G_2(x)}{n^\alpha}.$$

Die Abschätzungen (1), (3), (5) und (6) haben nun

$$(7) \quad |S_n^{(\delta)}(x) - f(x)| < \frac{G_2(x)}{n^\alpha}$$

zur Folge, womit die Richtigkeit des Annäherungsgrades $O\left(\frac{1}{n^\alpha}\right)$ für die Cesàroschen Mittel der Ordnung $\delta > \alpha$ bewiesen ist.

4. Es bleibt noch zu zeigen, daß im Falle $\alpha = \delta$ der Annäherungsgrad $O\left(\frac{\log n}{n^\alpha}\right)$ ebenfalls erhalten wird. Dabei bleiben die Abschätzungen (3) und (5) unverändert, da (5) auch für den Fall $\alpha = \delta$ berechnet wurde, bei der Abschätzung (3) haben wir aber von der Beziehung der Zahlen α und δ überhaupt keinen Gebrauch gemacht. Wir haben also nur noch einmal das von $\frac{1}{2n+1}$ bis h erstreckte Integral unter der Annahme $\alpha = \delta$ zu betrachten. Aus dieser Annahme folgt aber infolge (2) und (4) genau wie vorher:

$$\begin{aligned} \frac{1}{\binom{n+\delta}{n} \pi} \int_{\frac{1}{2n+1}}^h |\varphi(t)| |K_n(t)| dt &< \frac{H}{n^\alpha \pi} \int_{\frac{1}{2n+1}}^h |\varphi(t)| \frac{dt}{t^{1+\alpha}} \\ &= \frac{H}{n^\alpha \pi} \left[\frac{\Phi(t)}{t^{1+\alpha}} \right]_{\frac{1}{2n+1}}^h + \frac{H(1+\alpha)}{n^\alpha \pi} \cdot \frac{\Phi(\tau)}{\tau^{1+\alpha}} \int_{\frac{1}{2n+1}}^h \frac{dt}{t} \\ &< \frac{G_4(x)}{n^\alpha} + \frac{G_5(x)}{n^\alpha} (\log h + \log(2n+1)). \end{aligned}$$

Diese Ungleichungen vereint mit (1), (3) und (5) behaupten nun das Bestehen von

$$(8) \quad |S_n^{(\delta)}(x) - f(x)| < G_6(x) \frac{\log n}{n^\alpha},$$

womit unser Satz I in allen Teilen bewiesen ist.

5. Die folgende Bemerkung dürfte für das Spätere nicht überflüssig sein: Ist

$$\frac{1}{h^{1+\alpha}} \int_0^h |\varphi(t)| dt < G$$

von x unabhängig erfüllt, so hängt der Wert der Konstanten $G_1(x), \dots, G_6(x)$ in den Beziehungen (3), (5), (6), (7) und (8) weder von der Stelle x noch von n ab, daher bestehen (7) und (8) auf jeder Menge \mathfrak{A} gleichmäßig, auf welcher (2) gleichmäßig besteht. Diese Tatsache drückt aus

Satz II. *Ist die Bedingung (2) auf der Menge \mathfrak{A} aus $[0, 2\pi]$ gleichmäßig erfüllt, so bestehen die Beziehungen*

$$S_n^{(\delta)}(x) - f(x) = O\left(\frac{1}{n^\alpha}\right) \quad (\delta > \alpha)$$

und

$$S_n^{(\delta)}(x) - f(x) = O\left(\frac{\log n}{n^\alpha}\right) \quad (\delta = \alpha)$$

gleichmäßig auf \mathfrak{A} .

Schon in diesem Satze sind die Bernsteinschen Theoreme natürlich enthalten. Denn wir haben nur $\alpha \leq 1$ und $\delta = 1$ zu setzen und dann folgt aus der Lipschitzbedingung für alle Punkte x des Intervalles $[0, 2\pi]$

$$\frac{1}{h^{1+\alpha}} \int_0^h |\varphi(t)| dt \leq \frac{1}{h} \int_0^h \frac{|\varphi(t)|}{t^\alpha} dt < \text{Konst.},$$

was aber nach Satz II die Bernsteinschen Abschätzungen ergibt. Doch enthalten unsere Sätze bedeutend allgemeinere Fälle, als die einer Lipschitzbedingung.

§ 2.

Ein Satz über gleichmäßige Konvergenz.

6. Die Partialsummen $S_n^{(\delta)}(x)$ sind durch absolut konvergente singuläre Integrale dargestellt, deren spezielle Natur einen Zusammenhang zwischen dem Annäherungsgrad in den Punkten einer dichten Menge und zwischen der gleichmäßigen Annäherung in Intervallen in sich birgt. Wir wollen diesen Zusammenhang etwas näher betrachten.

Satz III. *Ist $\delta > 0$ und konvergieren die Mittel $S_n^{(\delta)}(x)$ auf einer in $[0, 2\pi]$ dichten Menge \mathfrak{A} gleichmäßig gegen die in $[0, 2\pi]$ stetige Funktion $f(x)$ mit dem Annäherungsgrade $O(\lambda_n)$, so konvergiert $S_n^{(\delta)}(x)$ im ganzen abgeschlossenen Intervalle $[0, 2\pi]$ gleichmäßig gegen $f(x)$ und zwar besteht dann in $[0, 2\pi]$ gleichmäßig*

$$(9) \quad S_n^{(\delta)}(x) - f(x) = O(\lambda_n).$$

Sei in der Tat x' ein Punkt aus dem Komplement $[0, 2\pi] - \mathfrak{A}$ der Menge \mathfrak{A} . Da \mathfrak{A} in $[0, 2\pi]$ dicht ist, gibt es zu x' eine Folge $\{x_{\nu}\}$ von Punkten aus \mathfrak{A} mit

$$\lim_{\nu \rightarrow \infty} x_{\nu} = x'.$$

Nun ist

$$(10) \quad |S_n^{(\delta)}(x') - f(x')| \\ \leq |S_n^{(\delta)}(x') - S_n^{(\delta)}(x_{\nu_n})| + |S_n^{(\delta)}(x_{\nu_n}) - f(x_{\nu_n})| + |f(x_{\nu_n}) - f(x')|.$$

Da aber die Funktion $f(x)$ in dem abgeschlossenen Intervalle $[0, 2\pi]$ stetig ist, ist sie auch gleichmäßig stetig; man kann also zu dem beliebigen x' und zu der beliebigen Folge $\{x_{\nu}\}$ ein einziges n_0 finden, so daß

$$(11) \quad |f(x') - f(x_{\nu_n})| < \lambda_n \quad (n \geq n_0),$$

wobei wir über die Zahlen λ_n stillschweigend die evidente Voraussetzung gemacht haben, daß $\lambda_n \rightarrow 0$, wenn $n \rightarrow \infty$. Die Beziehung (11) ist also zufolge der soeben angeführten Gründe für alle x' aus $[0, 2\pi] - \mathfrak{A}$ gleichmäßig erfüllt.

Es wurde weiter angenommen, daß $S_n^{(\delta)}(x)$ den Annäherungsgrad $O(\lambda_n)$ auf \mathfrak{A} gleichmäßig erreicht, also ist, da alle x_{ν_n} zu \mathfrak{A} gehören:

$$(12) \quad S_n^{(\delta)}(x_{\nu_n}) - f(x_{\nu_n}) = O(\lambda_n)$$

gleichmäßig erfüllt.

Bezeichne nun für ein Moment

$$\varphi_{\nu_n}(t) = f(x_{\nu_n} + 2t) + f(x_{\nu_n} - 2t) - 2f(x_{\nu_n}),$$

so ist

$$(13) \quad |S_n^{(\delta)}(x') - S_n^{(\delta)}(x_{\nu_n})| \leq \frac{1}{\binom{n+\delta}{n} x_0^{\delta}} \int_0^{\pi} |\varphi(t) - \varphi_{\nu_n}(t)| |K_n(t)| dt,$$

wobei $\varphi(t)$ für den Punkt x' zu bilden ist. Man findet aber

$$|\varphi(t) - \varphi_{\nu_n}(t)| \leq |f(x' + 2t) - f(x_{\nu_n} + 2t)| + |f(x' - 2t) - f(x_{\nu_n} - 2t)| \\ + 2|f(x') - f(x_{\nu_n})|.$$

Der gleichmäßigen Stetigkeit zufolge ist es belanglos, ob auf x' die lineare Transformation $x' \pm 2t$ ausgeübt wird oder nicht, aus (11) folgt demnach

$$|\varphi(t) - \varphi_{\nu_n}(t)| < 4\lambda_n.$$

Nehmen wir noch zu dieser Ungleichung die bekannte Tatsache hinzu, daß

$$\frac{1}{\binom{n+\delta}{n} x_0^{\delta}} \int_0^{\pi} |K_n(t)| dt < C \quad (C = \text{Konst.}),$$

so folgt aus (13):

$$(14) \quad |S_n^{(\delta)}(x') - S_n^{(\delta)}(x_{v_n})| < 4C\lambda_n.$$

Aus (10), (11), (12) und (14) ergibt sich daher, daß

$$S_n^{(\delta)}(x') - f(x') = O(\lambda_n)$$

für alle x' aus $[0, 2\pi] - \mathfrak{A}$ gleichmäßig erfüllt ist. Da diese Relation für die Punkte der Menge \mathfrak{A} postuliert wurde, ist damit gezeigt, daß (9) tatsächlich für jedes x aus $[0, 2\pi]$ gleichmäßig erfüllt ist. Das war aber unsere Behauptung.

§ 3.

Anwendung auf einige Spezialfälle.

7. Wir möchten zuerst einen klassischen, über eine Lipschitzbedingung hinausgehenden Spezialfall hervorheben, wenn nämlich $f(x)$ in $[0, 2\pi]$ von beschränkter Variation ist. Dann besitzt $f(x)$ bekanntlich fast überall eine integrierbare Derivierte $f'(x)$. Daraus folgt, daß die Funktion

$$\psi(t) = \frac{f(t) - f(x)}{t - x}$$

und damit offenbar auch die Funktion $\frac{\varphi(t)}{t}$ in einer genug kleinen Umgebung des Punktes x integrierbar ist; wir erhalten also:

$$\frac{1}{h^2} \int_0^h |\varphi(t)| dt \leq \frac{1}{h} \int_0^h \frac{|\varphi(t)|}{t} dt < G(x),$$

woraus sich nach Satz I

$$S_n^{(1)}(x) - f(x) = O\left(\frac{\log n}{n}\right)$$

ergibt. Die Sätze I und II sind aber wohl bedeutend tiefergehend als dieser Spezialfall. Dies ist aus der folgenden Fassung des Satzes II ersichtlich:

Satz IV. *Erfüllt die Funktion*

$$\Psi(t) = \frac{1}{t-x} \int_x^t |f(u) - f(x)| du$$

in den Punkten x der Menge \mathfrak{A} eine Lipschitzbedingung α -ter Ordnung, so besteht auf der Menge $\overline{\mathfrak{A}}$ aus \mathfrak{A} gleichmäßig

$$\begin{aligned} &= O\left(\frac{1}{n^\alpha}\right) && (\alpha < \delta); \\ S_n^{(\delta)}(x) - f(x) &= O\left(\frac{\log n}{n^\alpha}\right) && (\alpha = \delta), \end{aligned}$$

wenn nur auf $\bar{\mathfrak{A}}$ die Bedingung

$$(15) \quad \Psi(x) = \lim_{t \rightarrow x} \frac{1}{t-x} \int_x^t |f(u) - f(x)| du = 0$$

überall erfüllt ist.

In der Tat ist wegen der angenommenen Lipschitzbedingung nach (15):

$$\frac{\Psi(t)}{(t-x)^\alpha} = \frac{\Psi(t) - \Psi(x)}{(t-x)^\alpha} < M.$$

Daraus folgt für alle x aus $\bar{\mathfrak{A}}$

$$\begin{aligned} \frac{1}{h^{1+\alpha}} \int_0^h |\varphi(t)| dt &\leq \frac{1}{h^{1+\alpha}} \int_0^h |f(x+2t) - f(x)| dt + \frac{1}{h^{1+\alpha}} \int_0^h |f(x-2t) - f(x)| dt \\ &= \frac{1}{(t-x)^{1+\alpha}} \int_x^t |f(u) - f(x)| du = \frac{\Psi(t)}{(t-x)^\alpha} < M, \end{aligned}$$

woraus sich infolge des Satzes II unmittelbar die Behauptung ergibt.

8. Eine besondere Beachtung verdient das folgende Korollar des Satzes IV, welches sich auf die Konvergenz der Fourierreihe bezieht:

Satz V. Ist die Funktion $\Psi(t)$ in $[0, 2\pi]$ von beschränkter Variation, so besteht für ein stetiges $f(x)$:

$$(16) \quad S_n^{(0)}(x) - f(x) = o\left(\frac{1}{n^{1-\varepsilon}}\right)$$

gleichmäßig auf $[0, 2\pi]$, wie klein auch $\varepsilon > 0$ sein mag.

In der Tat besitzt $\Psi(t)$ als Funktion von beschränkter Variation in $[0, 2\pi]$ fast überall eine endliche Ableitung, also erfüllt sie fast überall eine Lipschitzbedingung α -ter Ordnung mit $\alpha < 1$, wobei aber α beliebig nahe an 1 rücken kann. Nach Satz IV ist dann auf einem maßgleichen Kerne $\bar{\mathfrak{A}}$ des Intervalles $[0, 2\pi]$

$$(17) \quad S_n^{(1)}(x) - f(x) = O\left(\frac{1}{n^\alpha}\right)$$

gleichmäßig erfüllt.

Wir wollen nun zeigen, daß diese Menge $\bar{\mathfrak{A}}$ in $[0, 2\pi]$ dicht ist. Wäre dies nicht der Fall, so gäbe es ein Teilintervall J aus $[0, 2\pi]$ der Länge l , in welchem $\bar{\mathfrak{A}}$ nirgends dicht wäre. Dann ist auch die abgeschlossene Hülle $(\bar{\mathfrak{A}} \cdot J)^0$ des Durchschnittes $\bar{\mathfrak{A}} \cdot J$ in J nirgends dicht. Das Komplement der Menge $(\bar{\mathfrak{A}} \cdot J)^0$ ist aber Vereinigung von abzählbar vielen Intervallen, also eine Menge von positivem Maße. Das Maß der Menge $\bar{\mathfrak{A}} \cdot J$ ist daher kleiner als l :

$$m(\bar{\mathfrak{A}} \cdot J) < m(J) = l.$$

Hieraus folgt:

$$m(\bar{\mathfrak{M}}) \leq m(\bar{\mathfrak{M}} \cdot J) + m([0, 2\pi] \cdot \bar{\mathfrak{M}} - J) \leq m(\bar{\mathfrak{M}} \cdot J) + 2\pi - l < 2\pi.$$

Das aber steht zu der Annahme, daß $\bar{\mathfrak{M}}$ und $[0, 2\pi]$ maßgleich sind, in Widerspruch. Die Menge $\bar{\mathfrak{M}}$ ist daher in $[0, 2\pi]$ dicht.

Die Beziehung (17) besteht demnach auf einer in $[0, 2\pi]$ dichten Menge gleichmäßig. Nach Satz III folgt daraus, daß (17) in $[0, 2\pi]$ gleichmäßig erfüllt ist. Aus einem Lebesgueschen Satze⁷⁾ ergibt sich also das gleichmäßige Erfülltsein der Relation

$$S_n^{(0)}(x) - f(x) = O\left(\frac{\log n}{n^\alpha}\right).$$

Nun ist aber α beliebig wenig von 1 verschieden, es gibt daher ein $\beta < \alpha$, für welches

$$S_n^{(0)}(x) - f(x) = o\left(\frac{1}{n^\beta}\right)$$

besteht und β von 1 sich beliebig wenig unterscheidet. Setzen wir noch $\beta = 1 - \varepsilon$, so ist (16) und damit Satz V bewiesen.

9. Das Überraschende in diesem Satze ist, daß wir nicht mehr angenommen haben als die beschränkte Variation der Funktion $\Psi(t)$ und die Stetigkeit von $f(x)$. Denn diese Bedingungen stehen in engster Beziehung zu der de la Vallée-Poussinschen Konvergenzbedingung⁸⁾, welche in nichts anderem besteht, als daß die Funktion

$$X(t) = \frac{1}{t-x} \int_x^t [f(u) - f(x)] dx$$

in $[0, 2\pi]$ von beschränkter Variation sei. Es läßt sich leicht einsehen, daß unsere Bedingung das gleiche leistet wie die de la Vallée-Poussinsche. Wir haben also im Satze V in einem besonders weitgehenden Falle nachgewiesen, daß die Fourierreihe auch bei diesen allgemeinen Bedingungen mit einem sehr guten Annäherungsgrad gegen $f(x)$ konvergiert.

Im allgemeinen läßt sich über die Konvergenz der Fourierreihe ziemlich viel aussagen, wenn wir die folgende Spezialisierung des Satzes IV beachten:

Satz VI. *Ist die Funktion $f(x)$ stetig und erfüllt $\Psi(t)$ in einer im Intervalle $[0, 2\pi]$ dichten Menge eine Lipschitzbedingung α -ter Ordnung, so bestehen in $[0, 2\pi]$ gleichmäßig die Beziehungen*

$$S_n^{(0)}(x) - f(x) = O\left(\frac{\log n}{n^\alpha}\right),$$

⁷⁾ H. Lebesgue, Leçons sur les séries trigonométriques, S. 114–117.

⁸⁾ Ch. J. de la Vallée-Poussin, Palermo Rend. 31 (1911), S. 296–299.

wenn $\alpha < 1$ und

$$S_n^{(0)}(x) - f(x) = O\left(\frac{(\log n)^2}{n}\right),$$

wenn $\alpha = 1$ ist.

In der Tat ist, da $f(x)$ in $[0, 2\pi]$ stetig:

$$\lim_{t \rightarrow x} \frac{1}{t-x} \int_x^t |f(u) - f(x)| du = 0$$

für alle x aus $[0, 2\pi]$, was das Erfülltsein von (15) bedeutet. Es gibt außerdem nach Annahme eine in $[0, 2\pi]$ dichte Menge \mathfrak{A} , auf welcher $\Psi(t)$ eine Lipschitzbedingung α -ter Ordnung erfüllt. Nach Satz IV besteht daher

$$(18) \quad \begin{aligned} S_n^{(1)}(x) - f(x) &= O\left(\frac{1}{n^\alpha}\right) && (\alpha < 1) \\ &= O\left(\frac{\log n}{n}\right) && (\alpha = 1) \end{aligned}$$

auf \mathfrak{A} gleichmäßig. Beachten wir nun, daß die Menge \mathfrak{A} in $[0, 2\pi]$ dicht ist, so besteht (18) nach Satz III in $[0, 2\pi]$ gleichmäßig. Auf Grund des Lebesgueschen Satzes erhalten wir dann die im Satze VI behauptete Beziehung.

§ 4.

Der Annäherungsgrad der arithmetischen Mittel, wenn $f(x)$ stetig und von beschränkter Variation ist.

10. Die Betrachtung des Annäherungsgrades im Falle, daß $f(x)$ eine Lipschitzbedingung der Form

$$\left| \frac{f(t) - f(x)}{t - x} \right| < M$$

erfüllt, verdient vielleicht eine besondere Beachtung, denn nach den vorangehenden allgemeinen Sätzen dürfte man erwarten, daß in diesem Spezialfalle eine bessere Annäherung stattfinden könnte, als in den bedeutend weitergehenden allgemeinen Fällen. Nach dem zweiten Bernsteinschen Satz ist aber der bisher bekannte Annäherungsgrad der arithmetischen Mittel $O\left(\frac{\log n}{n}\right)$, also derselbe, wie wir es in bedeutend allgemeineren Fällen gefunden haben. Wir wollen nun zeigen, daß dieser Bernsteinsche Annäherungsgrad auch in einem etwas allgemeineren Falle, als dem einer Lipschitzbedingung, verbessert werden kann. Beachten wir zu diesem Zwecke, daß eine Funktion, welche in $[0, 2\pi]$ überall eine Lipschitzbedingung erfüllt, ersichtlicherweise stetig und von beschränkter Variation ist, während das Umgekehrte durchaus nicht immer der Fall ist. Wir beweisen nun den folgenden

Satz VII. Ist $f(x)$ in $[0, 2\pi]$ stetig und von beschränkter Variation und ist die Derivierte $f'(x)$ in den Punkten einer in ihrer Stetigkeitsmenge dichten Menge \mathfrak{B} gleichmäßig stetig auf $[0, 2\pi]$, so besteht gleichmäßig der Annäherungsgrad

$$S_n^{(1)}(x) - f(x) = o\left(\frac{\log n}{n}\right)$$

im ganzen abgeschlossenen Intervalle $[0, 2\pi]$.

In der Tat hat Herr Szász⁹⁾ bewiesen, daß

$$(19) \quad \lim_{n \rightarrow \infty} \frac{n}{\log n} (S_n^{(1)}(x) - f(x)) = f'(x+0) - f'(x-0)$$

in allen Punkten gleichmäßig erfüllt ist, in welchen die Funktion $f(x)$ eine Ableitung $f'(x)$ besitzt und in welchen die Gleichung

$$(20) \quad \lim_{h \rightarrow 0} \frac{1}{h} \int_0^h |f'(x+t) - f'(x-t)| dt = 0$$

gleichmäßig besteht.

Da aber die Funktion $f(x)$ von beschränkter Variation ist, besitzt sie in den Punkten der Menge \mathfrak{A} vom Maße 2π eine Ableitung und da jede Ableitung in $[0, 2\pi]$ höchstens punktweise unstetig ist, ist $f'(x)$ auf einer in \mathfrak{A} dichten Menge $\overline{\mathfrak{A}}$ stetig, außerdem gibt es nach Annahme ein in der Menge $\overline{\mathfrak{A}}$ wieder dichter Teil \mathfrak{B} , in deren Punkten $f'(x)$ auf dem Intervalle $[0, 2\pi]$ gleichmäßig stetig ist. Auf \mathfrak{B} ist daher (20) und damit auch (19) gleichmäßig erfüllt. Da aber in diesen Punkten wegen der Stetigkeit der Ableitung

$$f'(x+0) - f'(x-0) = 0$$

ist, folgt aus (19) das gleichmäßige Erfülltsein von

$$(21) \quad S_n^{(1)}(x) - f(x) = o\left(\frac{\log n}{n}\right)$$

auf der Menge \mathfrak{B} . Die Menge \mathfrak{B} ist dicht in $\overline{\mathfrak{A}}$, $\overline{\mathfrak{A}}$ in \mathfrak{A} und \mathfrak{A} dicht in $[0, 2\pi]$, denn es hat das Maß 2π . Die Menge \mathfrak{B} ist also in $[0, 2\pi]$ dicht. Nach dem Satze III folgt daher zufolge (21), daß der Annäherungsgrad $o\left(\frac{\log n}{n}\right)$ in $[0, 2\pi]$ gleichmäßig erreicht wird.

⁹⁾ O. Szász, a. a. O. 4).

Über den Eindeutigkeitsatz in der Theorie der verallgemeinerten trigonometrischen Integrale¹⁾.

Von

M. Jacob in Wien.

Der Ausgangspunkt der folgenden Ausführungen ist eine gemeinsame Definition der Fourierschen Reihe und des Fourierschen Integrales, welche durch die Anwendung des Stieltjesschen Integralbegriffes gewonnen wird. Daraus ergibt sich die Möglichkeit, wichtige Punkte aus der Theorie der trigonometrischen Reihen und Integrale vom einheitlichen Standpunkte zu behandeln.

In der vorliegenden Arbeit wollen wir uns hauptsächlich mit der Eindeutigkeitsfrage beschäftigen. Im § 1 befassen wir uns mit der Weiterführung einiger Sätze von H. Hahn²⁾ und J. C. Burkill³⁾ aus der Theorie der verallgemeinerten Fourierintegrale, während § 2 dem eigentlichen Thema dieser Arbeit gewidmet ist. Zum Schlusse wollen wir im § 3 eine Formulierung des Parsevalschen Theorems angeben, wodurch dieses auf eine erweiterte Funktionenklasse ausgedehnt wird.

§ 1.

Vorbereitende Betrachtungen aus der Theorie der verallgemeinerten Fourierintegrale.

In der Theorie der verallgemeinerten Fourierintegrale handelt es sich hauptsächlich darum, die klassische Fouriersche Integralformel so zu erweitern, daß dieselbe für eine Funktionenklasse gilt, welche durch ein all-

¹⁾ Diese Arbeit ist eine Weiterführung der Resultate meiner Arbeit 1: „Über den Eindeutigkeitsatz in der Theorie der trigonometrischen Integrale“, Math. Annalen 97 (1927), S. 663—674.

²⁾ H. Hahn 1, 2.

³⁾ J. C. Burkill 1, 2.

gemeineres Verhalten der Funktion im Unendlichen gekennzeichnet ist. Jedoch ist die Charakterisierung des Verhaltens der Funktion im Unendlichen keine einheitliche, und man ist gezwungen, dafür verschiedene Bedingungen aufzustellen. Wir wollen zwei dieser Bedingungen besonders hervorheben:

- a) $\left| \frac{f(\xi)}{\xi} \right|$ ist im Unendlichen integrierbar⁴⁾,
 b) $f(\xi)$ ist im Unendlichen periodisch⁵⁾.

Ist a) oder b) im Unendlichen erfüllt, so wollen wir im folgenden sagen, daß die Eigenschaft (U) im Unendlichen erfüllt ist.

Es sei also $f(\xi)$ eine in $[-\infty, \infty]$ meßbare Funktion, welche in jedem endlichen Intervalle integrierbar ist und die im Unendlichen die Eigenschaft (U) besitzt. Dann bilden wir nach Hahn⁶⁾

$$(1) \quad \begin{aligned} \Phi_1(\mu) &= \frac{1}{\pi} \int_{-\infty}^{\infty} f(\xi) \frac{\sin \mu \xi}{\xi} d\xi, \\ \Psi_1(\mu) &= \frac{1}{\pi} \int_{-\infty}^{\infty} f(\xi) \frac{1 - \cos \mu \xi}{\xi} d\xi \end{aligned}$$

und ordnen der Funktion $f(\xi)$ formal den folgenden Ausdruck zu:

$$(2) \quad f(x) \sim \int_0^{\infty} [d\Phi_1(\mu) \cos \mu x + d\Psi_1(\mu) \sin \mu x].$$

Die Zuordnung (2) reduziert sich auf die Fouriersche Reihe von $f(\xi)$, falls $f(\xi)$ rein periodisch ist bzw. auf das klassische Fouriersche Integral, wenn $f(\xi)$ noch schärferen Bedingungen im Unendlichen genügt. Ferner ist bekannt:

A.⁷⁾ In jedem Punkte $\xi = x$, in welchem eine der bekannten Konvergenzbedingungen erfüllt ist, gilt:

$$(3) \quad f(x) = \lim_{\mu=\infty} \int_0^{\mu} [d\Phi_1(\mu) \cos \mu x + d\Psi_1(\mu) \sin \mu x].$$

B.⁸⁾ In jedem Punkte $\xi = x$, in welchem der absolute Integralmittelwert

⁴⁾ Zu dieser Klasse gehören im allgemeinen diejenigen Funktionen, welche sich zur Darstellung durch die klassische Fouriersche Integralformel eignen.

⁵⁾ Es gehören also alle periodischen integrierbaren Funktionen dieser Klasse an; somit umfassen die verallgemeinerten Fourierintegrale auch die Fourierschen Reihen.

⁶⁾ H. Hahn 1, S. 304.

⁷⁾ H. Hahn 1, S. 306.

⁸⁾ H. Hahn 2, S. 464.

$$\lim_{h=0} \frac{1}{h} \int_0^h |f(x+t) + f(x-t) - 2f(x)| dt = 0$$

existiert, gilt der Fejér-Lebesguesche Satz, d. h. es ist

$$(4) \quad f(x) = \lim_{\mu=\infty} \frac{1}{\mu} \int_0^\mu d\lambda \int_0^\lambda [d\Phi_1(\tau) \cos \tau x + d\Psi_1(\tau) \sin \tau x].$$

Burkill⁹⁾ hat unabhängig von Hahn die Eigenschaft (U) in bezug auf a) verallgemeinert, und zwar in folgender Weise: Er geht aus von einer Funktion $F(x)$ (bzw. $G(x)$), welche den nachstehenden Bedingungen genügt:

1. Sie ist in jedem endlichen Intervalle von beschränkter Variation,
2. $F(0) = 0$,
3. $F(x) = \frac{F(x+0) + F(x-0)}{2}$,

4. im Unendlichen existiert das Integral $\int_0^\infty \frac{dF(x)}{x}$.

Dann bildet er

$$(5) \quad \Phi_3(\mu) = \frac{2}{\pi} \int_0^\infty dF(x) \frac{\sin \mu x}{x}$$

bzw.

$$(6) \quad \Psi_3(\mu) = \frac{2}{\pi} \int_0^\infty dG(x) \frac{1 - \cos \mu x}{x}$$

und beweist die folgende Umkehrung:

$$(5.1) \quad F(x) = \int_0^\infty d\Phi_3(\mu) \frac{\sin \mu x}{\mu}$$

bzw.

$$(6.1) \quad G(x) = \int_0^\infty d\Psi_3(\mu) \frac{1 - \cos \mu x}{\mu}.$$

Wie man leicht einsehen kann, bekommt man im Falle, daß $F(x)$ (bzw. $G(x)$) ein unbestimmtes Integral ist, aus den Formeln von Burkill eine direkte Integration der Formeln von Hahn.

Hahn hat nun in der Arbeit 2. den folgenden Satz bewiesen:

Ist $f(\xi)$ eine in $[-\infty, \infty]$ meßbare, in jedem endlichen Intervalle integrierbare Funktion, die im Unendlichen beschränkt ist, so gilt in jedem

⁹⁾ J. E. Burkill 1., S. 515.

Stetigkeitspunkte $\xi = x$ die Formel:

$$(7) \quad f(x) = \lim_{\mu \rightarrow \infty} \frac{1}{\mu} \int_0^{\mu} \left\{ \int_0^{\lambda} \frac{d^2 \Phi_2(\tau)}{d\tau} \cos \tau x + \frac{d^2 \Psi_2(\tau)}{d\tau} \sin \tau x \right\} d\lambda.$$

Dabei bedeutet in (7):

$$(8) \quad \begin{aligned} \Phi_2(\tau) &= \frac{2}{\pi} \int_{-\infty}^{\infty} f(\xi) \left(\frac{\sin \frac{\mu \xi}{2}}{\xi} \right)^2 d\xi, \\ \Psi_2(\tau) &= \frac{1}{\pi} \int_{-1}^1 f(\xi) \frac{\mu \xi - \sin \mu \xi}{\xi^2} d\xi - \frac{1}{\pi} \left(\int_{-\infty}^{-1} f(\xi) \frac{\sin \mu \xi}{\xi^2} d\xi + \int_1^{\infty} f(\xi) \frac{\sin \mu \xi}{\xi^2} d\xi \right) \end{aligned}$$

und das Integral $\left\{ \int_0^{\lambda} \frac{d^2 \Phi_2(\tau)}{d\tau} \cos \tau x + \frac{d^2 \Psi_2(\tau)}{d\tau} \sin \tau x \right\}$ ist in sinngemäßer Art definiert¹⁰⁾.

¹⁰⁾ Das in (7) definierte Integral hängt mit der Verallgemeinerung des Stieltjesschen Integralbegriffes durch Hahn (2, S. 450 und 3, S. 77) eng zusammen. Hahn beweist:

A. Es sei $\Phi(x)$ im Intervalle $[a, b]$ definiert und habe dort überall einseitige Grenzwerte; für $x \leq a$ sei $\Phi(x) = \Phi(a)$, für $x \geq b$ sei $\Phi(x) = \Phi(b)$; ferner sei $g(x)$ in $[a, b]$ von endlicher Variation. Dann hat das verallgemeinerte Stieltjessche Integral

$$\int_a^b d\Phi(x) \cdot g(x)$$

einen eindeutig bestimmten, endlichen Wert.

B. Es sei $g(x)$ samt der Ableitung $g'(x)$ in $[a, b]$ stetig und von endlicher Variation; ferner sei $\Phi(x)$ in $[a, b]$ stetig, und es existieren im Punkte $x = a$ die rechtsseitige Ableitung $\Phi'_+(a)$ und im Punkte $x = b$ die linksseitige Ableitung $\Phi'_-(b)$. Ist nun Z eine Zerlegung

$$a = x_0 < x_1 < x_2 \dots < x_{n-1} < x_n = b$$

des Intervalles $[a, b]$, so bildet man die Summe

$$S(Z) = \sum_{i=0}^{n-1} g(x_i) \left[\frac{\Phi(x_{i+1}) - \Phi(x_i)}{x_{i+1} - x_i} - \frac{\Phi(x_i) - \Phi(x_{i-1})}{x_i - x_{i-1}} \right]$$

und weist nach, daß unter den gemachten Voraussetzungen ein endlicher Grenzwert existiert, der von der Wahl der Zerlegungsfolge unabhängig ist. Dieser Grenzwert wird mit

$$\int_a^b g(x) \frac{d^2 \Phi(x)}{dx}$$

bezeichnet, und es ist

$$\int_a^b g(x) \frac{d^2 \Phi(x)}{dx} = g(b) \Phi'_-(b) - g(a) \Phi'_+(a) - \int_a^b g'(x) d\Phi(x).$$

Es sei an dieser Stelle sogleich bemerkt, daß der zuletzt erwähnte Satz, unter verschärften Voraussetzungen, zum Beweise des Eindeutigkeitsatzes in der Theorie der trigonometrischen Integrale führt¹¹⁾. Es erweist sich nun, daß beim Beweise des Eindeutigkeitsatzes in der Theorie der verallgemeinerten trigonometrischen Integrale eine Erweiterung des Satzes von Hahn im Sinne einer Integration von Burkill eine wesentliche Rolle spielen wird. Diese Erweiterung, welche auch ein selbständiges Interesse verdient, bildet den wesentlichen Inhalt dieses Abschnittes.

Es sei $F(x)$ (bzw. $G(x)$) eine in $[0, \infty]$ definierte Funktion, welche den folgenden Bedingungen genügt:

1. Sie ist in jedem endlichen Intervalle von endlicher Variation,

2. $F(0) = 0$,

$$(9) \quad 3. \quad F(x) = \frac{F(x+0) + F(x-0)}{2},$$

4. es existiert im Unendlichen das Integral $\int \frac{dF(x)}{x^2}$.

Bildet man nun:

$$(10) \quad \begin{aligned} \Phi_4(\mu) &= \frac{1}{\pi} \int_0^\infty dF(x) \left(\frac{\sin \frac{\mu x}{2}}{\frac{x}{2}} \right)^2, \\ \Phi_{4,n}(\mu) &= \frac{1}{\pi] \int_0^n dF(x) \left(\frac{\sin \frac{\mu x}{2}}{\frac{x}{2}} \right)^2, \end{aligned}$$

C. Besitzt ferner $\Phi(x)$ eine erste Ableitung $\Phi'(x)$ und eine zweite $\Phi''(x)$ und gilt in jedem Teilintervalle $[\alpha, \beta]$ von $[a, b]$

$$\Phi'(\beta) - \Phi'(\alpha) = \int_\alpha^\beta \Phi''(x) dx,$$

so ist

$$\int_a^b g(x) \frac{d^2 \Phi(x)}{dx^2} = \int_a^b g(x) \Phi''(x) dx.$$

D. Genügt $g(x)$ und $\Phi_n(x)$ den unter B. gemachten Voraussetzungen, konvergiert $\Phi_n(x)$ in $[a, b]$ gleichmäßig gegen $\Phi(x)$ und ist

$$\Phi'_+(a) = \lim_{n=\infty} \Phi'_{n+}(a), \quad \Phi'_-(b) = \lim_{n=\infty} \Phi'_{n-}(b),$$

so gilt

$$\int_a^b g(x) \frac{d^2 \Phi(x)}{dx^2} = \lim_{n=\infty} \int_a^b g(x) \frac{d^2 \Phi_n(x)}{dx^2}.$$

¹¹⁾ M. Jacob, loc. cit. I (Nachtrag S. 671–674).

bzw.

$$\Psi_4(\mu) = \frac{2}{\pi} \int_0^1 dG(x) \frac{\mu x - \sin \mu x}{x^2} - \frac{2}{\pi} \int_1^\infty dG(x) \frac{\sin \mu x}{x^2},$$

(11)

$$\Psi_{4,n}(\mu) = \frac{2}{\pi} \int_0^1 dG(x) \frac{\mu x - \sin \mu x}{x^2} - \frac{2}{\pi} \int_1^n dG(x) \frac{\sin \mu x}{x^2},$$

so folgt aus den Voraussetzungen, daß in jedem endlichen Intervalle $\Phi_{4,n}(\mu)$ (bzw. $\Psi_{4,n}(\mu)$) gleichmäßig gegen die stetige Grenzfunktion $\Phi_4(\mu)$ (bzw. $\Psi_4(\mu)$) konvergiert. Verlangt man ferner, daß in jedem endlichen Intervalle $\Phi'_4(\mu)$ bzw. $\Psi'_4(\mu)$ integrierbar sein soll, wobei eine Teilfolge $\Phi'_{4,n_i}(\mu)$ bzw. $\Psi'_{4,n_i}(\mu)$ existiere, die fast überall gegen $\Phi'_4(\mu)$ bzw. $\Psi'_4(\mu)$ konvergiert, so kann man die folgende Umkehrung von (10) bzw. (11) beweisen:

$$(10.1) \quad F(x) = \lim_{\mu=\infty} \frac{1}{\mu} \int_0^\mu \left\{ \int_0^\lambda \frac{d^2 \Phi_4(\tau) \sin \tau x}{d\tau \tau} \right\} d\lambda$$

bzw.

$$(11.1) \quad G(x) = \lim_{\mu=\infty} \frac{1}{\mu} \int_0^\mu \left\{ \int_0^\lambda \frac{d^2 \Psi_4(\tau) (1 - \cos \tau x)}{d\tau \tau} \right\} d\lambda.$$

Bei dem nachstehenden Beweise der Umkehrformeln stützen wir uns zunächst auf die Arbeit 2. von Hahn und dann folgen wir den Überlegungen von Burkil. Ohne Beschränkung der Allgemeinheit setzen wir im folgenden $F(x)$ als monoton wachsend voraus. Nach Hahn ergibt sich¹²⁾, wenn wir die Formel (10.1) betrachten:

$$\begin{aligned} (12) \quad J(\mu, x) &= \frac{1}{\mu} \int_0^\mu \left\{ \int_0^\lambda \frac{d^2 \Phi_4(\tau) \sin \tau x}{d\tau \tau} \right\} d\lambda \\ &= \lim_{i=\infty} \frac{1}{\mu} \int_0^\mu \left\{ \int_0^\lambda \frac{d^2 \Phi_{4,n_i}(\tau) \sin \tau x}{d\tau \tau} \right\} d\lambda \\ &= \lim_{i=\infty} \frac{1}{\mu} \int_0^\mu \int_0^\lambda \Phi''_{4,n_i}(\tau) \frac{\sin \tau x}{\tau} d\tau d\lambda \\ &= \lim_{i=\infty} \frac{1}{2\mu x} \int_0^\mu \int_0^\lambda \int_0^{n_i} dF(v) \cos v\tau \frac{\sin \tau x}{\tau} d\tau d\lambda \end{aligned}$$

¹²⁾ H. Hahn 2, S. 458—461.

$$\begin{aligned}
&= \lim_{i=\infty} \int_0^{ni} dF(v) \left[\frac{1}{2\mu\pi} \int_0^\mu \int_0^\lambda \frac{\cos v\tau \sin \tau x}{\tau} d\tau d\lambda \right]^{13)} \\
&= \int_0^\infty dF(v) \varphi_1(\mu, v), \quad 14)
\end{aligned}$$

wobei zur Abkürzung:

$$(13) \quad \varphi_1(\mu, v) = \frac{1}{2\mu\pi} \int_0^\mu \int_0^\lambda \frac{\cos v\tau \sin \tau x}{\tau} d\tau d\lambda$$

gesetzt wird. Eine leichte Umformung ergibt:

$$(13.1) \quad \varphi_1(\mu, v) = \frac{1}{\pi} \int_{\mu \frac{(v-x)}{2}}^{\mu \frac{(v+x)}{2}} \left(\frac{\sin u}{u} \right)^2 du$$

und daraus erhalten wir:

$$(14.1) \quad |\varphi_1(\mu, v)| \leq 1 \quad \text{für jedes } x, \mu, v,$$

¹³⁾ Bei der vorgenommenen Vertauschung der Integrationsfolge stützen wir uns auf den folgenden bekannten Satz: Wenn für $a \leq x \leq b$ und $c \leq y \leq d$ $f(xy)$ eine stetige Funktion von x, y ist und wenn $\varphi(x)$ und $\psi(y)$ von endlicher Variation sind, dann ist:

$$\int_a^b d\varphi(x) \int_c^d d\psi(y) f(xy) = \int_c^d d\psi(y) \int_a^b d\varphi(x) f(xy).$$

¹⁴⁾ Hiermit sind wir zum singulären Integrale gelangt, dessen Kern:

$$\varphi_1(\mu, v) = \frac{1}{\pi} \int_{\mu \frac{(v-x)}{2}}^{\mu \frac{(v+x)}{2}} \left(\frac{\sin u}{u} \right)^2 du$$

zu dem Burkillschen:

$$\varphi_0(\mu, v) = \frac{1}{\pi} \int_{\mu \frac{(v-x)}{2}}^{\mu \frac{(v+x)}{2}} \left(\frac{\sin u}{u} \right) du$$

in demselben Verhältnisse steht, wie der Fejérsche Kern zum Dirichletschen in der Theorie der Fourierschen Reihen und Integrale. Während der Burkillsche Kern im Unendlichen von der Ordnung $O\left(\frac{1}{v}\right)$ ist, so ist der von uns betrachtete Kern von der Ordnung $O\left(\frac{1}{v^2}\right)$ und diese Eigenschaft werden wir hier ganz auswerten. Eine andere Eigenschaft, daß $\varphi_1(\mu, v)$ für jedes μ von endlicher Variation ist in $[-\infty, \infty]$ nützen wir jetzt nicht aus, und das läßt die Möglichkeit offen, die Umkehrsätze zu verschärfen, indem nur das Vorhandensein einseitiger Grenzwerte von $F(x)$ im Endlichen gefordert wird.

(14.2) für $v > x$ ist:

$$|\varphi_1(\mu, v)| \leq \frac{8}{\pi} \cdot \frac{1}{\mu(v-x)^2},$$

(14.3) bei festem x konvergiert mit wachsendem μ der Kern $\varphi_1(\mu, v)$ gleichmäßig in jedem Intervalle $a \leq v \leq b$, sobald $|v - x| \geq \delta$ (δ fest).

Wir zerlegen nun das singuläre Integral $J(\mu, x)$ in drei Teile:

$$(15) \quad J(\mu, x) = \int_0^{\infty} dF(v) \varphi_1(\mu, v) = \int_0^{x-\delta} + \int_{x-\delta}^{x+\delta} + \int_{x+\delta}^{\infty} = J_1 + J_2 + J_3.$$

Zunächst betrachten wir J_3 ; bei festem $\delta > 0$ existiert nach (9.4) das Integral:

$$\int_{\delta}^{\infty} \frac{|dF(v)|}{v^2}$$

und daher ist nach (14.2):

$$(16) \quad J_3 = \int_{x+\delta}^{\infty} dF(v) \varphi_1(\mu, v) = O\left(\frac{1}{\mu} \int_{x+\delta}^{\infty} \frac{|dF(v)|}{v^2}\right) = O\left(\frac{1}{\mu}\right).$$

Nun kann man, bei vorgegebenem $\varepsilon > 0$, 1. ein $\delta = \delta(\varepsilon)$ so bestimmen, daß:

$$(17) \quad \begin{aligned} |F(\xi) - F(x+0)| &< \varepsilon \quad \text{für } x < \xi \leq x + \delta, \\ |F(x-0) - F(\xi)| &< \varepsilon \quad \text{für } x - \delta \leq \xi < x \end{aligned}$$

und 2. kann erreicht werden, daß für alle $\mu \geq \mu_0(\varepsilon, \delta)$ folgende Beziehungen erfüllt sind:

$$(18.1) \quad |J_3| < \varepsilon,$$

$$(18.2) \quad |\varphi_1(\mu, v) - 1| < \varepsilon \quad \text{für } 0 \leq v \leq x - \delta,$$

$$(18.3) \quad \left| \varphi_1(\mu, x) - \frac{1}{2} \right| < \varepsilon.$$

Auf Grund von (17) und (18.3) erhalten wir für J_2 :

$$J_2 = \int_{x-\delta}^{x+\delta} dF(v) \varphi_1(\mu, v) = \frac{F(x+0) - F(x-0)}{2} + O(\varepsilon).$$

Zufolge der Voraussetzung, daß $F(0) = 0$, und wegen (17) und (18.2) bekommt man:

$$J_1 = \int_0^{x-\delta} dF(v) \varphi_1(\mu, v) = F(x-\delta) + O(\varepsilon) = F(x-0) + O(\varepsilon).$$

Es ist also:

$$J(\mu, x) = \frac{F(x+0) + F(x-0)}{2} + O(\varepsilon),$$

und somit ist die Umkehrformel (10.1) bewiesen. Ganz analog geht man beim Beweise der Umkehrformel (11.1) vor.

§ 2.

Beweis des Eindeutigkeitsatzes.

Es sei das nachstehende verallgemeinerte trigonometrische Integral:

$$(19) \quad \int_0^{\infty} [d\Phi(\mu) \cos \mu x + d\Psi(\mu) \sin \mu x]$$

vorgelegt, wobei vorausgesetzt wird: Es seien $\Phi(\mu)$ und $\Psi(\mu)$ in $[0, \infty]$ definierte Funktionen, welche in jedem endlichen Intervalle von endlicher Variation sind; ferner sei $\Phi(0) = \Psi(0) = 0$ und:

$$(19.1) \quad \lim_{\mu=\infty} \int_{\mu}^{\mu+1} |d\Phi(\mu)| = \lim_{\mu=\infty} \int_{\mu}^{\mu+1} |d\Psi(\mu)| = 0.$$

Frage. Wann ist das vorgelegte Integral (19) ein verallgemeinertes Fourierintegral, d. h. wann gibt es eine bis auf Nullmengen eindeutig bestimmte Funktion $f^*(\xi)$ in $[-\infty, \infty]$, so daß $\Phi(\mu)$ und $\Psi(\mu)$ zu ihr als Fouriersche Transformierte im Sinne der Bildung (1) gehören?

Wir behaupten nun den Satz:

Satz I. *Sind die Unbestimmtheitsgrenzen des verallgemeinerten trigonometrischen Integrals (19):*

$$(20) \quad \begin{aligned} G(x) &= \overline{\lim}_{\mu=\infty} J(\mu, x), \\ g(x) &= \underline{\lim}_{\mu=\infty} J(\mu, x), \end{aligned} \quad J(\mu, x) = \int_0^{\mu} [d\Phi(\mu) \cos \mu x + d\Psi(\mu) \sin \mu x]$$

in $[-\infty, \infty]$ *beinahe überall*¹⁵⁾ *endliche, in jedem unendlichen Intervalle integrierbare Funktionen, die im Unendlichen die Eigenschaft (U) besitzen, dann ist (19) ein verallgemeinertes Fourierintegral.*

In dem obigen Satze ist insbesondere die sinngemäße Verallgemeinerung des Cantorschen Satzes enthalten, welche folgendermaßen lautet:

Satz I'. *Konvergiert das verallgemeinerte trigonometrische Integral (19) beinahe überall gegen Null, so verschwinden die Funktionen $\Phi(\mu)$ und $\Psi(\mu)$ in $[0, \infty]$.*

¹⁵⁾ *Beinahe überall* bedeutet überall mit Ausnahme einer Menge, die keinen perfekten Teil enthält; diese Mengen sind aber als Borelsche Mengen nach Hausdorff (Math. Annalen 77 (1916), S. 430—436) und nach Alexandroff (C. R. 162 (1916), S. 323—325) abzählbar. — Es wäre interessant zu untersuchen, inwiefern die Gültigkeit der obigen Sätze erhalten bleibt, wenn man nicht abzählbare Mengen als Ausnahmемengen zuläßt. Es liegt zufolge einer in den Math. Annalen erscheinenden Arbeit von Zygmund: „Über die Beziehung der Eindeutigkeitsfragen in den Theorien der trigonometrischen Reihen und Integrale“ die Vermutung nahe, daß die Eindeutigkeitsmengen in der Theorie der verallgemeinerten trigonometrischen Integrale und in der Theorie der trigonometrischen Reihen dieselben sind. [Die Arbeit Zygmund ist in den Math. Annalen 99 erschienen. Anm. d. Red.]

Beim Beweise des Satzes I spielen einige bekannte Hilfssätze eine wichtige Rolle, die wir sogleich anführen wollen.

Hilfssatz I¹⁶⁾. *Es sei $f(x)$ eine im Intervalle $[a, b]$ beinahe überall endliche, integrierbare Funktion und $F(x)$ eine in $[a, b]$ stetige Funktion, so daß beinahe überall die Ungleichung besteht:*

$$(21) \quad \lim_{h=0} \frac{F(x+h) + F(x-h) - 2F(x)}{h^2} \leq f(x) \leq \overline{\lim}_{h=0} \frac{F(x+h) + F(x-h) - 2F(x)}{h^2},$$

ferner erfülle $F(x)$ im Intervalle $[a, b]$ für jedes x die Bedingung:

$$(21.1) \quad \lim_{h=0} \frac{F(x+h) + F(x-h) - 2F(x)}{h} = 0,$$

dann ist:

$$F(x) = \int_a^x \int_a^y f(t) dt dy + Ax + B.$$

Hilfssatz II¹⁷⁾. *Es sei $V(\tau)$ eine in $[0, \infty]$ definierte Funktion, die in jedem endlichen Intervalle von beschränkter Variation ist; ferner sei $u(\tau)$ eine in $[0, \infty]$ beschränkte, stetige Funktion. Existiert nun der Grenzwert:*

$$\lim_{\mu=\infty} J(\mu) = s, \quad J(\mu) = \int_0^\mu u(\tau) dV(\tau),$$

so ist:

$$\lim_{h=0} \int_0^\infty \left(\frac{\sin \tau h}{\tau h} \right)^2 u(\tau) dV(\tau) = s.$$

Hilfssatz III. (Verallgemeinerung des Hilfssatzes II.) *Es sei $V(\tau)$ und $u(\tau)$ wie vorher definiert. Existieren die endlichen Limites:*

$$\overline{\lim}_{\mu=\infty} J(\mu) = G, \quad \underline{\lim}_{\mu=\infty} J(\mu) = g,$$

so ist:

$$\begin{aligned} \frac{G+g}{2} - M \frac{G-g}{2} &\leq \lim_{h=0} \int_0^\infty \left(\frac{\sin \tau h}{\tau h} \right)^2 u(\tau) dV(\tau) \leq \overline{\lim}_{h=0} \int_0^\infty \left(\frac{\sin \tau h}{\tau h} \right)^2 u(\tau) dV(\tau) \\ &\leq \frac{G+g}{2} + M \frac{G-g}{2} \end{aligned}$$

(M bedeutet eine von $u(\tau)$ und $V(\tau)$ unabhängige Konstante).

Hilfssatz IV¹⁸⁾. *Es sei $V(\tau)$ und $u(\tau)$ wie vorher definiert. Ist ferner:*

$$\lim_{\mu=\infty} \int_\mu^{\mu+1} |dV(\tau)| = 0,$$

¹⁶⁾ Ch. J. de la Vallée-Poussin 1, S. 711.

¹⁷⁾ B. Riemann 1, S. 246—247.

¹⁸⁾ B. Riemann 1, S. 248.

so ist:

$$\lim_{h=0} h \int_0^{\infty} \left(\frac{\sin \tau h}{\tau h} \right)^2 u(\tau) dV(\tau) = 0.$$

Wir wenden uns nun dem Beweise des Eindeutigkeitsatzes zu und bilden:

$$(22) \quad F_1(x) = \frac{2}{\pi} \int_0^{\infty} d\Phi(\mu) \left(\frac{\sin \frac{\mu x}{2}}{\frac{\mu}{2}} \right)^2$$

$$(23) \quad F_2(x) = \frac{2}{\pi} \int_0^1 d\Psi(\mu) \frac{\mu x - \sin \mu x}{\mu^2} - \frac{2}{\pi} \int_1^{\infty} d\Psi(\mu) \frac{\sin \mu x}{\mu^2}$$

und setzen:

$$(24) \quad \underline{F(x) = F_1(x) + F_2(x)}.$$

Die so definierte Funktion $F(x)$ ist eine direkte Übertragung der Riemannschen Funktion $F(x)^{19)}$ aus der Theorie der trigonometrischen Reihen in die der verallgemeinerten trigonometrischen Integrale. Aus den Voraussetzungen über die Funktionen $\Phi(\mu)$ und $\Psi(\mu)$ folgt, nach (22) und (23), daß die Funktionen $F_1(x)$ und $F_2(x)$ in jedem endlichen Intervalle stetig sind und somit auch $F(x)$.

Bildet man weiter:

$$(25) \quad Q_h^2 F_1(x) = \frac{F_1(x+2h) + F_1(x-2h) - 2F_2(x)}{4h^2},$$

so ergibt sich auf Grund von (22):

$$(25.1) \quad Q_h^2 F_1(x) = \int_0^{\infty} \left(\frac{\sin \tau h}{\tau h} \right)^2 d\Phi(\tau) \cos \tau x.$$

Ebenso bekommt man:

$$(25.2) \quad Q_h^2 F_2(x) = \int_0^{\infty} \left(\frac{\sin \tau h}{\tau h} \right)^2 d\Psi(\tau) \sin \tau x.$$

Setzen wir nun:

$$(26) \quad \overline{\lim}_{h=0} Q_h^2 F_i(x) = \overline{D}^2 F_i(x), \quad \underline{\lim}_{h=0} Q_h^2 F_i(x) = \underline{D}^2 F_i(x), \quad (i = 1, 2)$$

¹⁹⁾ $F(x) = \frac{a_0 x^2}{2} - \sum_{\nu=1}^{\infty} \frac{a_{\nu} \cos \nu x + b_{\nu} \sin \nu x}{\nu^2}.$

so erhalten wir auf Grund des Hilfssatzes III beinahe überall die folgende Ungleichung²⁰⁾:

$$(27) \quad \frac{G_i(x) + g_i(x)}{2} - M \frac{G_i(x) - g_i(x)}{2} \leq \underline{D}^2 F_i(x) \leq \overline{D}^2 F_i(x) \\ \leq \frac{G_i(x) + g_i(x)}{2} + M \frac{G_i(x) - g_i(x)}{2}.$$

Nach (27) sind also die zweiten mittleren Ableitungen von $F_i(x)$ eingeschlossen zwischen Funktionenpaaren, die in $[-\infty, \infty]$ laut Voraussetzung beinahe überall endlich sind, in jedem endlichen Intervalle integrierbar und im Unendlichen die Eigenschaft (U) besitzen. Andererseits ergibt sich aus dem Hilfssatze IV für alle x :

$$(27.1) \quad \lim_{h=0} 2h Q_h^2 F_i(x) = \lim_{h=0} \frac{F_i(x+2h) + F_i(x-2h) - 2F_i(x)}{2h} = 0,$$

und daher sind die Voraussetzungen des Hilfssatzes I erfüllt, welcher besagt, daß die mittleren Derivierten $\overline{D}^2 F_i(x)$ und $\underline{D}^2 F_i(x)$ fast überall übereinstimmen müssen.

Bildet man nun:

$$(28) \quad \underline{D}^2 F_1(x) \leq f_1(x) \leq \overline{D}^2 F_1(x), \quad \underline{D}^2 F_2(x) \leq f_2(x) \leq \overline{D}^2 F_2(x)$$

und setzt:

$$(28.1) \quad f^*(x) = \frac{f_1(x) + f_2(x)}{2},$$

so lautet unsere Behauptung:

Die durch (28.1) definierte Funktion $f^(x)$ ist die gesuchte Funktion, deren verallgemeinertes Fourierintegral das vorgelegte Integral (19) ist.*

Denn wie aus der Bildung von $f^*(x)$ hervorgeht, gehört zu $f^*(x)$ ein verallgemeinertes Fourierintegral im Sinne der Zuordnung (2). Dieses sei:

$$(29) \quad f^*(x) \sim \int_0^\infty [d\Phi^*(\mu) \cos \mu x + d\Psi^*(\mu) \sin \mu x],$$

²⁰⁾ Es bedeutet in (27):

$$G_1(x) = \overline{\lim}_{\mu=\infty} \int_0^\mu d\Phi(\tau) \cos \tau x, \quad G_2(x) = \overline{\lim}_{\mu=\infty} \int_0^\mu d\Psi(\tau) \sin \tau x,$$

$$G_1(x) = \underline{\lim}_{\mu=\infty} \int_0^\mu d\Phi(\tau) \cos \tau x, \quad G_2(x) = \underline{\lim}_{\mu=\infty} \int_0^\mu d\Psi(\tau) \sin \tau x.$$

Aus den Forderungen, die wir an die Funktionen $G(x)$ und $g(x)$ gestellt haben, folgt auch, daß $G_i(x)$ und $g_i(x)$ ($i = 1, 2$) denselben genügen müssen.

wobei die Funktionen $\Phi^*(\mu)$ und $\Psi^*(\mu)$ im Sinne von (1)' gebildet werden. Wir haben jetzt nur nachzuweisen, daß:

$$(30) \quad \Phi^*(\mu) = \Phi(\mu), \quad \Psi^*(\mu) = \Psi(\mu).$$

Und nun werden wir die im § 1 bewiesenen Sätze anwenden; zunächst entspricht die Bildung (22) und (23) formal der Bildung (10) und (11). Unsere Voraussetzungen gestatten uns überdies die Umkehrformel von (22) bzw. (23) zu verfeinern. Es ergibt sich nämlich aus dem Hilfssatze I, daß:

$$F_i(x) = \int_a^x \int_a^y f_i(t) dt dy + A_i x + B_i \quad (i=1, 2)$$

und daraus folgt wegen (19.1) auf Grund des Tauberschen Satzes für Integrale, daß

$$F_1'(x) = \int_0^\infty d\Phi(\mu) \frac{\sin \mu x}{\mu}; \quad F_2'(x) = \int_0^1 d\Psi(\mu) \frac{1 - \cos \mu x}{\mu} - \int_1^\infty d\Psi(\mu) \frac{\cos \mu x}{\mu}.$$

Somit sind bei (22) die Bedingungen, die wir im § 1 an (10) gestellt haben, in verschärftem Maße erfüllt und daher bekommen in Analogie zu (10.1) für (22) die folgende Umkehrformel:

$$(31) \quad \Phi(\mu) = \lim_{T=\infty} \frac{2}{\pi T} \int_0^T \left\{ \int_0^\lambda \frac{d^2 F_1(t)}{dt^2} \cdot \frac{\sin xt}{t} \right\} d\lambda$$

und daraus wegen (28) und der Bemerkung loc. cit. 10. C):

$$(31.1) \quad \Phi(\mu) = \lim_{T=\infty} \frac{2}{\pi T} \int_0^T \int_0^\lambda f_1(t) \frac{\sin xt}{t} dt d\lambda.$$

Berücksichtigt man noch, daß $F_1(x) = F_1(-x)$ und $F_2(x) = -F_2(-x)$, so bekommt man aus (31.1):

$$(32) \quad \begin{aligned} \Phi(\mu) &= \lim_{T=\infty} \frac{1}{\pi T} \int_0^T \int_{-\lambda}^\lambda f_1(t) \frac{\sin xt}{t} dt d\lambda \\ &= \lim_{T=\infty} \frac{1}{\pi T} \int_0^T \int_{-\lambda}^\lambda (f_1(t) + f_2(t)) \frac{\sin xt}{t} dt d\lambda \\ &= \lim_{T=\infty} \frac{1}{\pi T} \int_0^T \int_{-\lambda}^\lambda f^*(t) \frac{\sin xt}{t} dt d\lambda. \end{aligned}$$

Ganz analog ergibt sich:

$$(33) \quad \Psi(\mu) = \lim_{T=\infty} \frac{1}{\pi T} \int_0^T \int_{-\lambda}^{\lambda} f^*(t) \frac{1 - \cos \mu t}{t} dt d\lambda.$$

In den Formeln (32) und (33) haben wir vorläufig Analoga zu den Pollard-schen²¹⁾ Formeln gewonnen und haben keineswegs die Tatsache benutzt, daß laut Voraussetzung $f^*(t)$ im Unendlichen die Eigenschaft (U) besitzt. Es existieren aber zufolge der Eigenschaft (U), wie schon vorher bemerkt wurde, die zu $f^*(t)$ gehörenden Transformierten:

$$(32.1) \quad \Phi^*(\mu) = \lim_{\lambda=\infty} \frac{1}{\pi} \int_{-\lambda}^{\lambda} f^*(t) \frac{\sin \mu t}{t} dt,$$

$$(33.1) \quad \Psi^*(\mu) = \lim_{\lambda=\infty} \frac{1}{\pi} \int_{-\lambda}^{\lambda} f^*(t) \frac{1 - \cos \mu t}{t} dt.$$

Daher ergibt sich nach dem Permanenzsatze der Cesàroschen Summierbarkeit, daß die Grenzwerte (32) bzw. (33) mit den Grenzwerten (32.1) bzw. (33.1) übereinstimmen müssen, und daraus folgt, daß:

$$\Phi^*(\mu) = \Phi(\mu), \quad \Psi^*(\mu) = \Psi(\mu),$$

w. z. b. w.

§ 3.

Über die Parsevalsche Formel.

Im Anschluß an die vorherigen Ausführungen wollen wir für eine Klasse der von uns betrachteten Funktionen eine Formulierung der Parsevalschen Formel angeben, deren Idee auf N. Wiener²²⁾ zurückgeht. Formal wird unsere Fassung der Parseval-Formel eine solche sein, daß sie sowohl die Parseval-Formel für die Fourierschen Reihen ergibt, wenn die betrachtete Funktion periodisch und quadratisch integrierbar ist, sowie auch die Parseval-Formel für die Fourierschen Integrale, falls $f(x)$ in $[-\infty, \infty]$ quadratisch integrierbar ist und außerdem noch geeigneten Bedingungen im Unendlichen genügt; überdies wird aber unsere Parseval-Formel für eine weitere Funktionenklasse gelten, für die eine solche bisher nicht aufgestellt wurde.

Die in diesem Paragraphen betrachtete Funktionenklasse sei folgendermaßen gekennzeichnet: $f(x)$ ist eine in $[-\infty, \infty]$ meßbare, in jedem

²¹⁾ S. Pollard 1, S. 458.

²²⁾ N. Wiener 1, S. 590–593 (Theorem III).

endlichen Intervalle quadratisch integrierbare Funktion, die im Unendlichen die Eigenschaft (U) besitzt und für die außerdem der folgende Grenzwert existiert:

$$(34) \quad \lim_{T \rightarrow \infty} \frac{1}{T^{1-\alpha}} \int_{-T}^T (f(x))^2 dx = G_\alpha \quad (0 \leq \alpha \leq 1).$$

Aus der Voraussetzung (34) folgt nach einer leichten Rechnung, daß die folgenden Integrale:

$$\int_{-\infty}^{\infty} \frac{(f(x))^2}{1+|x|^{2-\alpha}}$$

existieren und daher erst recht das Integral:

$$\int_{-\infty}^{\infty} \frac{(f(x))^2}{1+x^2}.$$

Bilden wir nun in Hinsicht auf (1) die Funktionen:

$$(35.1) \quad \varphi_h(\mu) = \Phi(\mu+h) - \Phi(\mu-h) = \frac{2}{\pi} \int_{-\infty}^{\infty} f(x) \frac{\sin hx}{x} \cos \mu x dx,$$

$$(35.2) \quad \psi_h(\mu) = \Psi(\mu+h) - \Psi(\mu-h) = \frac{2}{\pi} \int_{-\infty}^{\infty} f(x) \frac{\sin hx}{x} \sin \mu x dx,$$

so bekommt man durch die Anwendung der Parsevalschen Formel für die klassischen Fourierintegrale²³⁾:

$$(36) \quad A(f, h) = \frac{4}{\pi} \int_{-\infty}^{\infty} (f(x))^2 \left(\frac{\sin hx}{x} \right)^2 dx = \int_{-\infty}^{\infty} [\varphi_h^2(\mu) + \psi_h^2(\mu)] d\mu.$$

Nun kann man einen Satz von N. Wiener verallgemeinern, den wir in folgender Weise aussprechen wollen:

Hilfssatz V²⁴⁾. *Es sei $g(x)$ eine integrierbare Funktion, und es sei:*

$$a) \quad \frac{1}{T^{1-\alpha}} \int_{-T}^T |g(x)| dx < B \quad (0 \leq \alpha \leq 1),$$

$$b) \quad \frac{1}{T^{1-\alpha}} \int_{-T}^T g(x) dx \rightarrow G_\alpha \quad (T \rightarrow \infty),$$

²³⁾ M. Plancherel 1, W. Dorn 1, S. Pollard 2, F. Riesz 1.

²⁴⁾ M. Jacob 2. Für $\alpha = 0$ vgl. S. Bochner und G. H. Hardy 1, N. Wiener 2.

dann ist:

$$\lim_{h=0} \frac{h^{1-\alpha}}{\gamma_\alpha} \int_{-\infty}^{\infty} g(x) \left(\frac{\sin xh}{xh} \right)^2 dx = G_\alpha, \quad \gamma_\alpha = \frac{2^{\alpha-1} \pi}{\Gamma(\alpha+2)} \cdot \frac{1-\alpha}{\sin(1-\alpha) \frac{\pi}{2}}.$$

Setzt man $g(x) = (f(x))^2$, so ist der letzte Hilfssatz auf die rechte Seite von (36) anwendbar, und man bekommt die angekündigte Parsevalsche Formel:

$$(37) \quad \lim_{h=0} \frac{\Delta(f, h)}{\gamma_\alpha^* \cdot h^{1+\alpha}} = G_\alpha, \quad \gamma_\alpha^* = \frac{2^{1+\alpha}}{\Gamma(\alpha+2)} \cdot \frac{1-\alpha}{\sin(1-\alpha) \frac{\pi}{2}}.$$

Für $\alpha = 0$ bekommen wir als Spezialfall die Parseval-Formel für die Fourierschen Reihen, für $\alpha = 1$ die Parseval-Formel für die Fourierschen Integrale. Für $0 < \alpha < 1$ scheint diese Formel neu zu sein. Diese Funktionenklasse ist sicher nicht leer, man setze z. B.

$$(38) \quad f_\alpha(x) = \frac{1}{1 + |x|^{\alpha/2}} \quad (0 < \alpha < 1).$$

Die Funktionen (38) haben nach A. Pringsheim²⁵⁾ im klassischen Sinne eine Fouriersche Integraldarstellung, sie sind aber nicht quadratisch integrierbar in $[-\infty, \infty]$. Sie besitzen aber im Sinne der Bildungsweise (1) eindeutig bestimmte Transformierte $\Phi(\mu)$ und $\Psi(\mu)$, und daher kann ihnen in dem oben definierten Sinne eine Parsevalsche Formel zugeordnet werden.

Literaturverzeichnis.

- S. Bochner und G. H. Hardy: 1. Note on two theorems of Norbert Wiener, *Journal Lond. Math. Soc.* **1** (1926), S. 240–242.
- J. C. Burkill: 1. The expression in Stieltjes integrals of the inversion formulae of Fourier and Hankel, *Proceed. Lond. Math. Soc.* **25** (1926), S. 513–524.
2. On Mellin's inversion formula, *Proceed. Camb. Phil. Soc.* **23** (1926), S. 356–360.
- W. Dorn: 1. Fouriersche Integrale als Grenzwerte Fourierscher Reihen, *Sitzungsberichte der Wiener Akad. der Wissensch.* 1926, S. 127–147.
- H. Hahn: 1. Über die Verallgemeinerung der Fourierschen Integralformel, *Acta mathem.* **49** (1926), S. 301–353.
2. Über die Methode der arithmetischen Mittel in der Theorie der verallgemeinerten Fourierintegrale, *Sitzungsberichte der Wiener Akad. der Wissensch.* (1925), S. 449–470.
3. Über Folgen linearer Operationen, *Monatsh. für Math. u. Phys.* **32** (1922), S. 1–88.

²⁵⁾ A. Pringsheim 1.

- M. Jacob: 1. Über den Eindeutigkeitssatz in der Theorie der trigonometrischen Integrale, *Math. Annalen* **97** (1927), S. 663–674.
2. Über ein Theorem von Bochner-Hardy-Wiener, erscheint im *Journ. Lond. Math. Soc.* **3** (1928).
- B. Riemann: 1. Über die Darstellbarkeit einer Funktion durch eine trigonometrische Reihe, *Gesammelte Werke*, 2. Aufl., S. 227–271.
- M. Plancherel: 1. Contribution à l'étude de la représentation d'une fonction arbitraire par des integrales définies, *Rendiconti Palermo* **30** (1910).
- S. Pollard: 1. Identification of the coefficients in a trigonometrical integral, *Proceed. Lond. Math. Soc.* **25** (1926), S. 451–468.
2. On Fourier's integral, *Proc. Lond. Math. Soc.* (2) **26** (1926), S. 12–24.
- A. Pringsheim: 1. Über neue Gültigkeitsbedingungen der Fourierschen Integralformel, *Math. Annalen* **68** (1910).
- F. Riesz: Sur la formule d'inversion de Fourier, *Acta litt. ac. scient. Univ. Hung.* **3** (1927), S. 235–241.
- Ch. J. de la Vallée-Poussin: 1. Sur l'unicité du développement trigonométrique, *Bullet. de l'Acad. de Belgique* 1912, S. 702–718.
- N. Wiener: 1. On the representation of functions by trigonometrical integrals, *Math. Zeitschr.* **24** (1925), S. 575–616.
2. On a theorem of Bochner and Hardy, *Journ. Lond. Math. Soc.* **2** (1927), S. 118–123.

(Eingegangen am 30. 10. 1927.)

Zur Theorie der Gesellschaftsspiele¹⁾.

Von

J. v. Neumann in Berlin.

Einleitung.

1. Die Frage, deren Beantwortung die vorliegende Arbeit anstrebt, ist die folgende:

n Spieler, S_1, S_2, \dots, S_n , spielen ein gegebenes Gesellschaftsspiel \mathcal{G} . Wie muß einer dieser Spieler, S_m , spielen, um dabei ein möglichst günstiges Resultat zu erzielen?

Die Fragestellung ist allgemein bekannt, und es gibt wohl kaum eine Frage des täglichen Lebens, in die dieses Problem nicht hineinspielte; trotzdem ist der Sinn dieser Frage kein eindeutig klarer. Denn sobald $n > 1$ ist (d. h. ein eigentliches Spiel vorliegt), hängt das Schicksal eines jeden Spielers außer von seinen eigenen Handlungen auch noch von denen seiner Mitspieler ab; und deren Benehmen ist von genau denselben egoistischen Motiven beherrscht, die wir beim ersten Spieler bestimmen möchten. Man fühlt, daß ein gewisser Zirkel im Wesen der Sache liegt.

Wir müssen also versuchen, zu einer klaren Fragestellung zu kommen. Was ist zunächst ein Gesellschaftsspiel? Es fallen unter diesen Begriff sehr viele, recht verschiedenartige Dinge: von der Roulette bis zum Schach, vom Bakkarat bis zum Bridge liegen ganz verschiedene Varianten des Sammelbegriffes „Gesellschaftsspiel“ vor. Und letzten Endes kann auch irgendein Ereignis, mit gegebenen äußeren Bedingungen und gegebenen Handelnden (den absolut freien Willen der letzteren vorausgesetzt), als Gesellschaftsspiel angesehen werden, wenn man seine Rückwirkungen auf die in ihm handelnden Personen betrachtet²⁾. Was ist nun das gemeinsame Merkmal aller dieser Dinge?

¹⁾ Der Inhalt dieser Arbeit ist (mit einigen Kürzungen) am 7. XII. 1926 der Göttinger Math. Ges. vorgetragen worden.

²⁾ Es ist das Hauptproblem der klassischen Nationalökonomie: was wird, unter gegebenen äußeren Umständen, der absolut egoistische „homo oeconomicus“ tun?

Man darf wohl annehmen, daß es dieses ist:

Ein Gesellschaftsspiel besteht aus einer bestimmten Reihe von Ereignissen, deren jedes auf endlich viele verschiedene Arten ausfallen kann. Bei gewissen unter diesen Ereignissen hängt der Ausfall vom Zufall ab, d. h.: es ist bekannt, mit welchen Wahrscheinlichkeiten die einzelnen möglichen Resultate eintreten werden, aber niemand vermag sie zu beeinflussen. Die übrigen Ereignisse aber hängen vom Willen der einzelnen Spieler S_1, S_2, \dots, S_n ab. D. h.: es ist bei jedem dieser Ereignisse bekannt, welcher Spieler S_m seinen Ausfall bestimmt, und von den Resultaten welcher anderer („früherer“) Ereignisse er im Moment seiner Entscheidung bereits Kenntnis hat. Nachdem der Ausfall aller Ereignisse bereits bekannt ist, kann nach einer festen Regel berechnet werden, welche Zahlungen die Spieler S_1, S_2, \dots, S_n aneinander zu leisten haben.

Es ist leicht, diese mehr qualitative Erklärung in die Form einer exakten Definition zu bringen. Diese Definition des Gesellschaftsspieles würde so lauten:

Um ein Gesellschaftsspiel \mathcal{G} vollständig zu beschreiben, sind die folgenden Angaben notwendig, die zusammen die „Spielregel“ ergeben:

a) Es muß angegeben werden, wie viele vom Zufall abhängige Ereignisse oder „Ziehungen“ und wieviel vom Willen der einzelnen Spieler abhängige Ereignisse oder „Schritte“ erfolgen. Diese Anzahlen seien z bzw. s , die „Ziehungen“ bezeichnen wir mit E_1, E_2, \dots, E_z , die „Schritte“ mit F_1, F_2, \dots, F_s .

β) Es muß angegeben werden, auf wie viele Arten jede „Ziehung“ E_μ und jeder „Schritt“ F_ν ausfallen kann. Diese Anzahlen seien M_μ bzw. N_ν ($\mu = 1, 2, \dots, z$, $\nu = 1, 2, \dots, s$). Wir bezeichnen die betreffenden Resultate kurz mit ihren Nummern $1, 2, \dots, M_\mu$ bzw. $1, 2, \dots, N_\nu$.

γ) Bei jeder „Ziehung“ E_μ müssen die Wahrscheinlichkeiten $\alpha_\mu^{(1)}, \alpha_\mu^{(2)}, \dots, \alpha_\mu^{(M_\mu)}$ der einzelnen Resultate $1, 2, \dots, M_\mu$ gegeben sein. Natürlich ist

$$\alpha_\mu^{(1)} \geq 0, \alpha_\mu^{(2)} \geq 0, \dots, \alpha_\mu^{(M_\mu)} \geq 0,$$

$$\alpha_\mu^{(1)} + \alpha_\mu^{(2)} + \dots + \alpha_\mu^{(M_\mu)} = 1.$$

δ) Bei jedem „Schritt“ F_ν muß erstens derjenige Spieler S_m angegeben sein, der den Ausfall dieses „Schrittes“ bestimmt („dessen Schritt“ F_ν ist): $S_{(F_\nu)}$. Ferner müssen die Nummern aller „Ziehungen“ und „Schritte“ angegeben sein, über deren Ausfall er im Momente seiner Entscheidung über F_ν Kenntnis hat. (Diese „Ziehungen“ und „Schritte“ nennen wir „früher“ als F_ν .)

Damit die ganze Sache möglich ist und zeitlich-kausal vorstellbar sei, darf es keine Zyklen $F_{r_1}, F_{r_2}, \dots, F_{r_p}, F_{r_{p+1}} = F_{r_1}$ geben, derart, daß stets F_{r_q} „früher“ ist, als $F_{r_{q+1}}$ ($q = 1, 2, \dots, p$).

ϵ) Schließlich müssen n Funktionen f_1, f_2, \dots, f_n gegeben sein. Jede von ihnen ist abhängig von $z + s$ Variablen, die bzw. die Werte

$$1, 2, \dots, M_1; \quad 1, 2, \dots, M_2; \quad \dots; \quad 1, 2, \dots, M_z;$$

$$1, 2, \dots, N_1; \quad 1, 2, \dots, N_2; \quad \dots; \quad 1, 2, \dots, N_s$$

durchlaufen. Diese Funktionen haben reelle Zahlen als Werte, und es gilt identisch

$$f_1 + f_2 + \dots + f_n \equiv 0.$$

Wenn nun im Laufe einer zu Ende gespielten Partie die Resultate der z „Ziehungen“ und der s „Schritte“ bzw. $x_1, x_2, \dots, x_z, y_1, y_2, \dots, y_s$ ($x_\mu = 1, 2, \dots, M_\mu, y_\nu = 1, 2, \dots, N_\nu; \mu = 1, 2, \dots, z, \nu = 1, 2, \dots, s$) waren, so erhalten die Spieler S_1, S_2, \dots, S_n voneinander³⁾ die Summen

$$f_1(x_1, \dots, x_z, y_1, \dots, y_s), \quad f_2(x_1, \dots, x_z, y_1, \dots, y_s), \quad \dots,$$

$$f_n(x_1, \dots, x_z, y_1, \dots, y_s).$$

(Trotz der etwas langatmigen Beschreibung handelt es sich hier, wenn man genau zusieht, um recht einfache und klare Dinge. Übrigens hätten wir die Definition in mehreren Beziehungen etwas allgemeiner fassen können: so hätten wir z. B. zulassen können, daß die M_μ, N_ν und $\alpha_1^{(\mu)}, \alpha_2^{(\mu)}, \dots, \alpha_{M_\mu}^{(\mu)}$ von den Resultaten der „früheren“ „Ziehungen“ und „Schritte“ abhängen, u. ä.; indessen überzeugt man sich leicht davon, daß dabei nichts wesentlich Neues herauskommt.)

2. Mit dieser Definition ist der Begriff des Gesellschaftsspieles genau umschrieben. Es tritt aber auch ganz klar in Erscheinung, was wir bereits am Anfang von 1. berührten, daß nämlich die Ausdrucksweise: „ S_m sucht ein möglichst günstiges Resultat zu erzielen“ eine recht unklare ist. Ein für den Spieler S_m möglichst günstiges Resultat ist offenbar ein möglichst großer Wert von f_m , aber wie soll überhaupt irgendein Wert von f_m durch S_m „erzielt“ werden? S_m ist ja allein gar nicht in der Lage, den Wert von f_m festzulegen! f_m hängt von den Variablen $x_1, \dots, x_z, y_1, \dots, y_s$ ab, und von diesen wird nur ein Teil durch den Willen von S_m bestimmt (nämlich diejenigen y_ν , für die S_m den „Schritt“ F_ν hat, d. h. $S_{(F_\nu)} = S_m$ ist); die übrigen Variablen hängen vom Willen der Mitspieler (nämlich alle übrigen y_ν) oder vom Zufall (nämlich alle x_μ) ab.

In unserem Falle ist der „unvoraussehbare“ Zufall noch der leichter zu beherrschende Faktor. In der Tat: nehmen wir an, ein f_m hinge außer von jenen y_ν , die S_m bestimmt ($S_{(F_\nu)} = S_m$), nur von den x_μ (die vom

³⁾ Die Identität

$$f_1 + f_2 + \dots + f_n \equiv 0$$

drückt aus, daß die Spieler nur aneinander Zahlungen leisten, die Gesamtheit aber weder gewinnt noch verliert.

Zufall bestimmt werden) ab. Dann wird S_m jedenfalls das Folgende voraussehen können: Wenn ich auf eine bestimmte Weise spiele, so habe ich die und die Resultate (d. i. Werte von f_m) mit den und den Wahrscheinlichkeiten zu erwarten (die Wahrscheinlichkeiten $\alpha_1^{(\mu)}, \alpha_2^{(\mu)}, \dots, \alpha_{M\mu}^{(\mu)}$ sind ja gegeben) — unabhängig davon, nach welchen Prinzipien die übrigen Spieler handeln! Wenn wir nun annehmen, daß unter „günstigstem Resultat“ ein möglichst hoher Erwartungswert zu verstehen ist (und diese oder eine ähnliche Annahme muß gemacht werden, um die Methoden der Wahrscheinlichkeitsrechnung anwenden zu können⁴⁾), so ist die Aufgabe prinzipiell gelöst. Denn es handelt sich um ein einfaches Maximumproblem: S_m muß die Werte der von ihm zu bestimmenden unter den Variablen y , so wählen, daß der (allein von diesen abhängige) Erwartungswert von f_m möglichst groß wird.

Dieser Typus von Gesellschaftsspielen ist es, der in der Wahrscheinlichkeitsrechnung, in der sog. „Theorie der Glücksspiele“ behandelt wird. Ein charakteristisches Beispiel hierfür ist die Roulette: es sind $k+1$ Spieler da (S_1, \dots, S_k sind die „Pointeurs“, S_{k+1} der „Bankier“), S_{k+1} hat überhaupt keinen Einfluß auf das Spiel⁵⁾ und das Resultat von S_l, f_l , ($l = 1, 2, \dots, k$) ist allein vom Zufall und von seinen eigenen Handlungen abhängig⁶⁾.

Schon der Name „Glücksspiele“ zeigt, daß das Hauptgewicht auf die vom Zufall abhängigen Variablen x_μ , und nicht auf die vom Willen der Spieler abhängigen Variablen y , gelegt wird. Aber gerade das ist es, was uns hier beschäftigt wird. Es soll versucht werden, die Rückwirkungen der Spieler aufeinander zu untersuchen, die Konsequenzen des (für alles soziale Geschehen so charakteristischen!) Umstandes, daß jeder Spieler auf die Resultate aller anderen einen Einfluß hat und dabei nur am eigenen interessiert ist.

I. Allgemeine Vereinfachungen.

1. Die in der Einleitung gegebene Definition des Gesellschaftsspieles ist ziemlich kompliziert, was angesichts des Umstandes, daß es beliebige verwickelte Gesellschaftsspiele geben kann, motiviert erscheinen mag. Trotz-

⁴⁾ Die bekannten Einwände gegen den Erwartungswert (die seine Ersetzung durch die sog. moralische Hoffnung u. ä. erstreben), wollen wir unberücksichtigt lassen: es sind andere Schwierigkeiten, die den Gegenstand unserer Betrachtungen bilden.

⁵⁾ Er hat es auch nicht nötig, denn auf Grund der Spielregeln gewinnt er pro Partie 2.70% nach dem Umsatz.

⁶⁾ Wie man auf Grund der vorhergehenden Fußnote vermuten wird, ist das in diesem Falle eindeutig zu erzielende Resultat für das Verhalten der Pointeurs ein recht triviales: sie müssen möglichst den Umsatz 0 haben, je näher sie ihm kommen, desto besser!

dem lassen sich alle in dieser Definition enthaltenen Gesellschaftsspiele auf eine viel einfachere Normalform bringen; sozusagen auf die einfachstdenkbare Form überhaupt. Wir behaupten nämlich:

Es genügt, Gesellschaftsspiele folgender Art zu betrachten:

Es ist $Z = 1$ (d. h. es findet nur eine „Ziehung“ statt).

Es ist $s = n$, und zwar ist der ν -te „Schritt“ der des Spielers S_ν , ($S_{(F_\nu)} = S_\nu$).

Die Relation „früher“ besteht nie (d. h. jeder Spieler muß seine Dispositionen treffen, ohne etwas über die anderen oder die „Ziehung“ zu wissen).

Das Spiel verläuft also so: Jeder Spieler S_m ($m = 1, 2, \dots, n$) wählt eine Zahl $1, 2, \dots, N_m$ aus, ohne die Wahlen der übrigen zu kennen; und dann findet eine Ziehung statt, bei der die Zahlen $1, 2, \dots, M$ mit den Wahrscheinlichkeiten $\alpha_1, \alpha_2, \dots, \alpha_M$ herauskommen können. Die Resultate der Spieler sind (wenn die „Ziehung“ und die n „Schritte“ x, y_1, y_2, \dots, y_n ergaben):

$$f_1(x, y_1, \dots, y_n), f_2(x, y_1, \dots, y_n), \dots, f_n(x, y_1, \dots, y_n).$$

Daß diese scheinbar sehr weitgehende Einschränkung der Möglichkeiten in Wahrheit keine wesentliche ist, können wir so einsehen:

Die „Schritte“ des Spielers S_m (die mit $S_{(F_\nu)} = S_m$) seien diejenigen mit den Nummern $\nu_1^{(m)}, \nu_2^{(m)}, \dots, \nu_{\sigma_m}^{(m)}$. Es ist klar, daß die Annahme, S_m könnte schon vor Beginn des Spieles sagen, welche Wahlen er bei diesen Schritten treffen wird, eine unstatthafte ist; d. h. eine Beschränkung seiner Willensfreiheit und eine Änderung (Verschlechterung) seiner Chancen mit sich bringt. Denn der Entschluß von S_m bei jedem dieser „Schritte“ wird ja im allgemeinen wesentlich dadurch beeinflußt werden, wie die Resultate derjenigen „Ziehungen“ und „Schritte“ waren, von denen er im Momente des Entschlusses Kenntnis hat.

Demgegenüber darf wohl angenommen werden, daß er bereits vor Anfang des Spieles auf die folgende Frage zu antworten weiß: Wie wird der $\nu_k^{(m)}$ -te „Schritt“ ausfallen ($k = 1, 2, \dots, \sigma_m$), wenn die Resultate aller „Ziehungen“ und „Schritte“ vorliegen, die „früher“ als $\nu_k^{(m)}$ sind? D. h. daß der Spieler von vornherein weiß, wie er in einer genau umschriebenen Situation handeln wird; daß er mit einer fertigen Theorie ins Spiel geht. Selbst wenn dies bei einem Spieler nicht der Fall ist, leuchtet es wohl ein, daß eine derartige Annahme keineswegs seine Chancen verschlechtert.

Demgemäß definieren wir die „Spielmethode“ des Spielers S_m folgendermaßen:

Um die „Spielmethode“ eines Spielers S_m ($m = 1, 2, \dots, n$) vollständig zu beschreiben, sind die folgenden Angaben notwendig:

S_m habe, wie oben, die „Schritte“ mit den Nummern $v_1^{(m)}, v_2^{(m)}, \dots, v_{\sigma_m}^{(m)}$. Bei der Entschlußfassung zum $v_k^{(m)}$ -ten „Schritte“ ($k = 1, 2, \dots, \sigma_m$) vorliegend, d. h. „früher“ als dieser, seien die „Ziehungen“ und „Schritte“ mit den bzw. Nummern $\bar{\mu}_1^{(m,k)}, \bar{\mu}_2^{(m,k)}, \dots, \bar{\mu}_{\alpha_{m,k}}^{(m,k)}$ und $\bar{v}_1^{(m,k)}, \bar{v}_2^{(m,k)}, \dots, \bar{v}_{\beta_{m,k}}^{(m,k)}$.

Es muß dann für jede mögliche Kombination von Resultaten der genannten „Ziehungen“ und „Schritte“ (es sind ihrer offenbar nur endlich viele möglich) angegeben werden, wie der Entschluß von S_m über den $v_k^{(m)}$ -ten „Schritt“ lauten (d. h. wie dieser Schritt ausfallen) wird.

Man sieht sofort, daß es für S_m nur endlich viele Spielmethoden gibt, wir bezeichnen diese mit $\mathfrak{S}_1^{(m)}, \mathfrak{S}_2^{(m)}, \dots, \mathfrak{S}_{\Sigma_m}^{(m)}$.

Nun zeigt man offenbar ganz leicht (und dabei kommt die Annahme von Einleitung 1., Definition des Gesellschaftsspieles δ , über das Fehlen von Zyklen zur Anwendung), daß der Verlauf des Spieles auf eine mögliche und eindeutige Weise beschrieben ist, wenn angegeben wird:

1. Welcher „Spielmethoden“ $\mathfrak{S}^{(1)}, \mathfrak{S}^{(2)}, \dots, \mathfrak{S}^{(n)}$ sich bzw. die Spieler S_1, S_2, \dots, S_n bedienen.
2. Was die Resultate der „Ziehungen“ E_1, E_2, \dots, E_z sind.

Man beachte dabei die zwei folgenden Umstände: Erstens liegt es im Wesen des Begriffes der „Spielmethode“, daß alles, was ein Spieler über die Handlungen seiner Mitspieler und dem Ausfall von „Ziehungen“ erfahren oder folgern kann, bereits innerhalb der „Spielmethode“ Berücksichtigung findet. Folglich muß die Wahl der Spielmethode selbst bei jedem Spieler in absoluter Unkenntnis der Wahlen der übrigen Spieler und der Resultate der „Ziehungen“ erfolgen.

Zweitens ist hierdurch das getrennte Erfolgen der Ziehungen E_1, E_2, \dots, E_z (wobei bei E_μ , $\mu = 1, 2, \dots, z$, die Zahlen $1, 2, \dots, M_\mu$ mit den bzw. Wahrscheinlichkeiten $\alpha_\mu^{(1)}, \alpha_\mu^{(2)}, \dots, \alpha_\mu^{(M_\mu)}$ herauskommen können) ganz gleichgültig geworden: die Spieler müssen ja unabhängig davon handeln, d. h. ihre „Spielmethoden“ wählen. Dann hindert uns aber nichts, diese z Ziehungen zu einer einzigen Ziehung H zusammenzuziehen, wobei die Zahlenkomplexe

$$x_1, x_2, \dots, x_z \quad (x_\mu = 1, 2, \dots, M_\mu, \quad \mu = 1, 2, \dots, z)$$

mit den bzw. Wahrscheinlichkeiten $\alpha_1^{(x_1)} \cdot \alpha_2^{(x_2)} \cdot \dots \cdot \alpha_z^{(x_z)}$ herauskommen können; oder was dasselbe ist: die Zahlen $1, 2, \dots, M$ ($M = M_1 \cdot M_2 \cdot \dots \cdot M_z$) mit den entsprechenden Wahrscheinlichkeiten, die wir $\beta_1, \beta_2, \dots, \beta_M$ nennen wollen.

Wir können also 2. folgendermaßen modifizieren:

- 2'. Es muß angegeben werden, was das Resultat der „Ziehung“ H ist. (H kann die Resultate $1, 2, \dots, M$ mit den bzw. Wahrscheinlichkeiten $\beta_1, \beta_2, \dots, \beta_M$ haben.)

Nun sind die in 1. angegebenen Wahlen der Spieler S_1, S_2, \dots, S_n (als „Schritte“ aufgefaßt) und die in 2'. angegebene „Ziehung“ (unter Berücksichtigung des Umstandes, daß jeder „Schritt“ in absoluter Unkenntnis der übrigen Umstände erfolgt) dem ursprünglichen Gesellschaftsspiele \mathcal{G} vollkommen äquivalent; und sie bilden ihrerseits offenbar ein Gesellschaftsspiel \mathcal{G}' , das in der Tat von der am Anfang dieses Paragraphen erwähnten einfachen Form ist.

2. Als letztes, von unserem Gesichtspunkt aus unwesentliches Element soll jetzt auch noch die „Ziehung“ aus dem Spiele eliminiert werden; das geschieht dadurch, daß wir an Stelle der tatsächlichen Ergebnisse für die einzelnen Spieler ihre Erwartungswerte betrachten. Genauer:

Wenn die Spieler S_1, S_2, \dots, S_n die „Spielmethoden“

$$\mathfrak{E}_{u_1}^{(1)}, \mathfrak{E}_{u_2}^{(2)}, \dots, \mathfrak{E}_{u_n}^{(n)} \quad (u_m = 1, 2, \dots, \Sigma_m, \quad m = 1, 2, \dots, n)$$

wählten (wir können übrigens bereits davon absehen, daß es sich um „Spielmethoden“ und nicht um eigentliche „Schritte“ handelt, und einfach von den Wahlen u_1, u_2, \dots, u_n sprechen), und bei der „Ziehung“ H die Zahl $v (= 1, 2, \dots, M)$ herauskam, so seien die Ergebnisse für die Spieler S_1, S_2, \dots, S_n bzw.

$$f_1(v, u_1, \dots, u_n), \quad f_2(v, u_1, \dots, u_n), \quad \dots, \quad f_n(v, u_1, \dots, u_n).$$

Wenn nun nur die Wahlen u_1, u_2, \dots, u_n bekannt sind, die „Ziehung“ v aber noch nicht, so sind die Erwartungswerte der f_1, f_2, \dots, f_n diese:

$$g_m(u_1, \dots, u_n) = \sum_{v=1}^M \beta_v f_m(v, u_1, \dots, u_n) \quad (m = 1, 2, \dots, n)$$

(Aus $f_1 + f_2 + \dots + f_n \equiv 0$ folgt $g_1 + g_2 + \dots + g_n \equiv 0$.) Es ist ganz im Geiste der Methode der Wahrscheinlichkeitsrechnung, wenn wir von der „Ziehung“ überhaupt absehen und so tun, als ob es nur auf die Erwartungswerte g_1, g_2, \dots, g_n ankäme. Damit gewinnen wir aber den folgenden, noch weiter schematisierten und vereinfachten Grundtypus des Gesellschaftsspieles.

Jeder der Spieler S_1, S_2, \dots, S_n wählt eine Zahl, und zwar S_m eine der Zahlen $1, 2, \dots, \Sigma_m$ ⁷⁾ ($m = 1, 2, \dots, n$). Jeder hat seinen Entschluß zu fassen, ohne über die Resultate der Wahlen seiner Mitspieler Kenntnis zu haben. Wenn sie die Wahlen x_1, x_2, \dots, x_n getroffen haben

⁷⁾ Wir könnten auch noch alle Σ_m gleich machen, indem wir ein Σ annehmen, das nicht kleiner ist als irgend ein Σ_m , und jedesmal den Σ_m -ten Fall in $\Sigma - \Sigma_m + 1$ Unterfälle weiter teilen, von denen jeder genau dieselbe Wirkung hat, wie der ursprüngliche. Diese Vereinfachung ist aber unwesentlich.

($x_m = 1, 2, \dots, \Sigma_m, m = 1, 2, \dots, n$), so erhalten sie bzw. die folgenden Summen:

$$g_1(x_1, \dots, x_n), \quad g_2(x_1, \dots, x_n), \quad \dots, \quad g_n(x_1, \dots, x_n).$$

(Dabei ist identisch $g_1 + g_2 + \dots + g_n \equiv 0$.)

Damit ist diejenige Form der Spielregel erreicht, die (trotzdem sie, wie wir soeben zeigten, im wesentlichen nichts an Allgemeinheit verloren hat), nur noch die für uns wesentlichen Merkmale des Gesellschaftsspieles zeigt. Vom „Glücksspiel“ ist nichts mehr da: die Handlungen aller Spieler bestimmen das Resultat restlos (weil ja so operiert wird, als ob es ein jeder von ihnen nur auf den Erwartungswert abgesehen hätte). Aber dafür tritt das am Ende der Einleitung hervorgehende Prinzip in voller Schärfe in Erscheinung: jedes g_m hängt von allen x_1, x_2, \dots, x_n ab.

Der aus der Wahrscheinlichkeitsrechnung bekannte Fall: g_m hängt nur von x_m ab (was natürlich nicht für alle m eintreten kann), erscheint nun als vollkommen trivial.

II. Der Fall $n = 2$.

1. Weiter können wir zunächst in der bisherigen Allgemeinheit nicht kommen, es erweist sich vielmehr als zweckmäßig, jetzt den einfachsten Fall für n zu betrachten. Der Fall $n = 0$ ist sinnlos, der Fall $n = 1$ (wegen $g_1 + \dots + g_n \equiv 0$) ebenfalls, beidemal ist kein eigentliches Gesellschaftsspiel vorhanden. Es ist also der Fall $n = 2$, der nun in Frage kommt.

Da $g_1 + g_2 \equiv 0$ ist, kann $g_1 = g, g_2 = -g$ gesetzt werden. Dann lautet die Beschreibung des allgemeinen 2-Personen-Spieles so:

Die Spieler S_1, S_2 wählen irgendwelche der Zahlen $1, 2, \dots, \Sigma_1$ bzw. $1, 2, \dots, \Sigma_2$ und zwar jeder ohne die Wahl des anderen zu kennen. Wenn sie die Zahlen x bzw. y gewählt haben, so erhalten sie die Summen $g(x, y)$ bzw. $-g(x, y)$.

Dabei kann nun $g(x, y)$ jede beliebige Funktion (definiert für $x = 1, 2, \dots, \Sigma_1, y = 1, 2, \dots, \Sigma_2!$) sein.

Es ist leicht, sich ein Bild von den Tendenzen zu machen, die in einem solchen 2-Personen-Spiele miteinander kämpfen: Es wird von zwei Seiten am Werte von $g(x, y)$ hin und her gezerrt, nämlich durch S_1 , der ihn möglichst groß, und durch S_2 , der ihn möglichst klein machen will. S_1 gebietet über die Variable x , und S_2 über die Variable y . Was wird geschehen?

2. Wenn S_1 die Zahl x ($x = 1, 2, \dots, \Sigma_1$) gewählt hat, so hängt sein Resultat $g(x, y)$ auch noch von der Wahl y des S_2 ab, ist aber

jedenfalls $\geq \text{Min}_y g(x, y)$. Und diese untere Grenze kann durch geeignete Wahl von x gleich $\text{Max}_x \text{Min}_y g(x, y)$ (und nicht größer!) gemacht werden. D. h. wenn S_1 es will, so kann er $g(x, y)$ (unabhängig von S_2 !) jedenfalls

$$\geq \text{Max}_x \text{Min}_y g(x, y)$$

machen. Ebenso zeigt man: wenn S_2 es will, so kann er $g(x, y)$ (unabhängig von S_1 !) jedenfalls

$$\leq \text{Min}_y \text{Max}_x g(x, y)$$

machen.

Wenn nun

$$\text{Max}_x \text{Min}_y g(x, y) = \text{Min}_y \text{Max}_x g(x, y) = M$$

ist, so folgt aus dem Obigen, sowie daraus, daß S_1 das $g(x, y)$ möglichst groß und S_2 es möglichst klein machen will, daß $g(x, y)$ den Wert M haben wird. Denn S_1 hat das Interesse, es groß zu machen, und kann verhindern, daß es kleiner als M wird; S_2 hingegen hat das Interesse, es klein zu machen und kann verhindern, daß es größer als M wird. Folglich wird es den Wert M haben.

Nun ist zwar allgemein

$$\text{Max}_x \text{Min}_y g(x, y) \leq \text{Min}_y \text{Max}_x g(x, y),$$

aber es besteht keineswegs stets das $=$ -Zeichen. Es ist vielmehr leicht, solche $g(x, y)$ anzugeben, bei denen das $<$ -Zeichen gilt, wo also diese Überlegung versagt. Das einfachste derartige Beispiel ist das folgende:

$$\begin{aligned} \Sigma_1 = \Sigma_2 = 2, \quad g(1, 1) = 1, \quad g(1, 2) = -1, \\ g(2, 1) = -1, \quad g(2, 2) = 1. \end{aligned}$$

(Es ist offenbar $\text{Max Min} = -1$ und $\text{Min Max} = 1$.)

Ein anderes Beispiel ist die sog. „Morra“⁸⁾:

$$\begin{aligned} \Sigma_1 = \Sigma_2 = 3, \quad g(1, 1) = 0, \quad g(1, 2) = 1, \quad g(1, 3) = -1, \\ g(2, 1) = -1, \quad g(2, 2) = 0, \quad g(2, 3) = 1, \\ g(3, 1) = 1, \quad g(3, 2) = -1, \quad g(3, 3) = 0. \end{aligned}$$

(Auch hier ist $\text{Max Min} = -1$ und $\text{Min Max} = 1$.)

Daß diese Schwierigkeit auftritt, kann man sich auch so klarmachen:

$\text{Max}_x \text{Min}_y g(x, y)$ ist das beste Resultat, das S_1 erzielen kann, wenn ihn S_2 vollkommen durchschaut: wenn S_2 , sooft S_1 x spielt, ein solches y

⁸⁾ Auch „Verbrecher-Bakkarat“ oder „Knobeln“ genannt. In der üblichen Formulierung heißen 1, 2, 3 „Papier“, „Stein“, „Schere“ („Papier verdeckt den Stein, Stein schleift die Schere, Schere schneidet das Papier“).

spielt, das $g(x, y) = \text{Min}_y g(x, y)$ wird. (Auf Grund der Spielregeln durfte S_2 nicht wissen, was S_1 spielen wird, er mußte also aus anderen Gründen wissen, wie S_1 spielt, das ist es, was wir mit „durchschauen“ andeuten wollen.) Ebenso ist $\text{Min}_y \text{Max}_x g(x, y)$ das beste Resultat, das S_2 erzielen kann, wenn ihn S_1 durchschaut hat. Wenn die beiden Zahlen gleich sind, so bedeutet dies: es ist gleichgültig, welcher von den beiden Spielern der feinere Psychologe ist, das Spiel ist so unempfindlich, daß immer dasselbe herauskommt. Es ist klar, daß dies bei den beiden angeführten Spielen nicht der Fall ist: hier kommt alles darauf an, den Gegner zu durchschauen, zu erraten, ob er 1 oder 2 (bzw. 1, 2 oder 3) wählen wird.

Die Verschiedenheit der zwei Größen Max Min und Min Max bedeutet eben, daß von den zwei Spielern S_1 und S_2 nicht jeder gleichzeitig der klügere sein kann.

3. Es gelingt aber trotzdem mittels eines Kunstgriffes, die Gleichheit der zwei oben erwähnten Ausdrücke zu erzwingen.

Zu diesem Zwecke werden die Verhaltungsmöglichkeiten der Spieler S_1, S_2 folgendermaßen erweitert: Es wird von S_1 nicht verlangt, daß er sich am Anfang des Spieles für irgendeine der Zahlen $1, 2, \dots, \Sigma_1$ entscheide. Er soll nur Σ_1 Wahrscheinlichkeiten

$$\xi_1, \xi_2, \dots, \xi_{\Sigma_1} \quad (\xi_1 \geq 0, \xi_2 \geq 0, \dots, \xi_{\Sigma_1} \geq 0, \xi_1 + \xi_2 + \dots + \xi_{\Sigma_1} = 1)$$

angeben und sodann die Zahlen $1, 2, \dots, \Sigma_1$ aus einer Urne mit den Wahrscheinlichkeiten $\xi_1, \xi_2, \dots, \xi_{\Sigma_1}$ ziehen. Er wählt dann die gezogene Zahl. Dies scheint zwar eine Beeinträchtigung seines freien Entschlusses zu sein: denn nicht er bestimmt x ; ist es aber nicht: denn will er unbedingt ein bestimmtes x haben, so kann er $\xi_x = 1, \xi_u = 0$ (für $u \neq x$) festsetzen. Demgegenüber schützt er sich gegen das „Durchschaut-werden“: denn wenn etwa $\xi_1 = \xi_2 = \frac{1}{2}$ ist, so vermag niemand (selbst er nicht!) vorauszusagen, ob er 1 oder 2 wählen wird!

Ebenso soll S_2 verfahren: auch er wählt nur Σ_2 Wahrscheinlichkeiten $\eta_1, \eta_2, \dots, \eta_{\Sigma_2}$, und verfährt entsprechend.

Die Gesamtheit der $\xi_1, \xi_2, \dots, \xi_{\Sigma_1}$ wollen wir mit ξ und die Gesamtheit der $\eta_1, \eta_2, \dots, \eta_{\Sigma_2}$ mit η bezeichnen. Wenn S_1 ξ und S_2 η wählt, so hat S_1 den Erwartungswert

$$h(\xi, \eta) = \sum_{p=1}^{\Sigma_1} \sum_{q=1}^{\Sigma_2} g(p, q) \xi_p \eta_q$$

und S_2 den Erwartungswert $-h(\xi, \eta)$. Die neue Funktion $h(\xi, \eta)$ umfaßt die alte $g(x, y)$ offenbar im folgenden Sinne: wenn $\xi_x = \eta_y = 1$ und $\xi_u = \eta_v = 0$ (für $u \neq x, v \neq y$) ist, so ist $h(\xi, \eta) = g(x, y)$.

Nun können wir für $h(\xi, \eta)$ genau dieselben Überlegungen anstellen, wie vorhin für $g(x, y)$. Wenn S_1 ξ gewählt hat, so ist sein Erwartungswert mindestens $\text{Min}_\eta h(\xi, \eta)$; er kann also den minimalen Erwartungswert $\text{Max}_\xi \text{Min}_\eta h(\xi, \eta)$ (unabhängig von S_2 !) erzwingen. Ebenso kann S_2 verhindern, daß der Erwartungswert von S_1 den maximalen Wert $\text{Min}_\eta \text{Max}_\xi h(\xi, \eta)$ übersteigt. Wieder ist

$$\text{Max}_\xi \text{Min}_\eta h(\xi, \eta) \leq \text{Min}_\eta \text{Max}_\xi h(\xi, \eta),$$

und es fragt sich, ob stets das =-Zeichen gilt.

Man beachte, daß wir diesmal bessere Aussichten haben als bei $g(x, y)$: denn $g(x, y)$ konnte irgendeine Funktion sein, während $h(\xi, \eta)$ eine Bilinearform ist! Trotzdem also $h(\xi, \eta)$ eigentlich eine Verallgemeinerung von $g(x, y)$ ist, ist es als Funktion von viel einfacherem Typus als dieses. In der Tat werden wir im Abschnitt 3 beweisen, daß die Relation

$$\text{Max}_\xi \text{Min}_\eta h(\xi, \eta) = \text{Min}_\eta \text{Max}_\xi h(\xi, \eta)$$

(Max_ξ erstreckt über alle ξ mit $\xi_1 \geq 0, \dots, \xi_{s_1} \geq 0, \xi_1 + \dots + \xi_{s_1} = 1$, Min_η erstreckt über alle η mit $\eta_1 \geq 0, \dots, \eta_{s_2} \geq 0, \eta_1 + \dots + \eta_{s_2} = 1$) für alle Bilinearformen $h(\xi, \eta)$ besteht.

4. Wir setzen (unter Vorwegnahme des Resultates)

$$\text{Max}_\xi \text{Min}_\eta h(\xi, \eta) = \text{Min}_\eta \text{Max}_\xi h(\xi, \eta) = M.$$

Die Menge derjenigen ξ , für die $\text{Min}_\eta h(\xi, \eta)$ seinen Maximalwert M annimmt, sei \mathfrak{A} ; die Menge derjenigen η , für die $\text{Max}_\xi h(\xi, \eta)$ seinen Minimalwert M annimmt, sei \mathfrak{B} . Aus diesen Definitionen folgen dann die folgenden Relationen ohne weiteres:

1. Wenn ξ zu \mathfrak{A} gehört, so ist stets $h(\xi, \eta) \geq M$.
2. Wenn η zu \mathfrak{B} gehört, so ist stets $h(\xi, \eta) \leq M$.
3. Wenn ξ nicht zu \mathfrak{A} gehört, so gibt es ein η mit $h(\xi, \eta) < M$.
4. Wenn η nicht zu \mathfrak{B} gehört, so gibt es ein ξ mit $h(\xi, \eta) > M$.
5. Wenn ξ zu \mathfrak{A} und η zu \mathfrak{B} gehört, so ist $h(\xi, \eta) = M$.

Auf Grund dieser Relationen 1. bis 5. ist man wohl berechtigt zu erklären:

S_1 bzw. S_2 muß jedenfalls einen zu \mathfrak{A} gehörigen Komplex ξ bzw. einen zu \mathfrak{B} gehörigen Komplex η wählen, einerlei welchen. Eine Partie hat für S_1 bzw. S_2 den Wert M bzw. $-M$.

Ein 2-Personen-Spiel ist offenbar als „gerecht“ zu bezeichnen, wenn $M=0$ ist; und als „symmetrisch“, wenn die Spieler S_1, S_2 dieselben Rollen haben. D. h. wenn bei Vertauschung von ξ und η (dies setzt

natürlich $\Sigma_1 = \Sigma_2$ voraus) sich auch $h(\xi, \eta)$ und $-h(\xi, \eta)$ vertauschen, also wenn

$$h(\xi, \eta) = -h(\eta, \xi)$$

oder, was dasselbe ist,

$$g(x, y) = -g(y, x)$$

ist. Also: wenn die Bilinearform $h(\xi, \eta)$, oder auch die Matrix $g(x, y)$, schiefsymmetrisch ist. In diesem Falle ist es natürlich auch „gerecht“, was man so einsehen kann:

$$\begin{aligned} -\text{Max}_{\xi} \text{Min}_{\eta} h(\xi, \eta) &= \text{Min}_{\xi} \text{Max}_{\eta} -h(\xi, \eta) = \text{Min}_{\xi} \text{Max}_{\eta} h(\eta, \xi) \\ &= \text{Min}_{\eta} \text{Max}_{\xi} h(\xi, \eta), \end{aligned}$$

d. h.

$$-M = M, \quad M = 0 \text{ } ^{\circ}$$

Man überzeugt sich leicht, daß in unseren zwei Beispielen (in 2.) $M = 0$ ist, und zwar umfaßt \mathfrak{A} nur $\xi_1 = \xi_2 = \frac{1}{2}$ bzw. $\xi_1 = \xi_2 = \xi_3 = \frac{1}{3}$, und \mathfrak{B} $\eta_1 = \eta_2 = \frac{1}{2}$ bzw. $\eta_1 = \eta_2 = \eta_3 = \frac{1}{3}$. D. h.: beide Spiele sind gerecht (die „Morra“ ist sogar symmetrisch), und in beiden muß jeder Spieler alle Zahlen durcheinander wählen, und zwar alle mit der gleichen Wahrscheinlichkeit.

Es ist vielleicht nicht uninteressant, noch auf den folgenden Umstand mit Nachdruck hinzuweisen. Trotzdem im Abschnitte 1. der Zufall (durch die Einführung der Erwartungswerte und Streichung der „Ziehungen“) aus den zu betrachtenden Gesellschaftsspielen eliminiert wurde, ist er hier wieder von selbst aufgetreten: selbst wenn die Spielregel keinerlei „hazarde“ Elemente enthält (d. h. Ziehungen aus Urnen) — wie etwa die beiden Beispiele aus 2. —, ist es doch unumgänglich notwendig, das „hazarde“ Element, bei der Angabe der Verhaltensmaßregeln für die Spieler, wieder in Betracht zu ziehen. Das Zufallsabhängige („hazarde“, „statistische“) liegt so tief im Wesen des Spieles (wenn nicht im Wesen der Welt) begründet, daß es gar nicht erforderlich ist, es durch die Spielregel künstlich einzuführen: auch wenn in der formalen Spielregel davon keine Spur ist, bricht es sich von selbst die Bahn.

^o) Dabei ist $\text{Max Min} = \text{Min Max}$ benützt worden, d. h. unser relativ tiefer Satz über Bilinearformen. Trivial, d. h. aus $\text{Max Min} \leq \text{Min Max}$, folgt hier offenbar nur

$$\text{Max Min} \leq 0, \quad \text{Min Max} \geq 0.$$

Während der endgültigen Abfassung dieser Arbeit wurde mir die Note von Herrn E. Borel in den Comptes rendus vom 10. Jan. 1927 (Sur les systèmes de formes linéaires ... et la théorie du jeu, S. 52–55) bekannt. Borel formuliert die auf Bilinearformen bezügliche Frage für ein symmetrisches 2-Personen-Spiel und stellt fest, daß keine Beispiele für $\text{Max Min} < \text{Min Max}$ bekannt sind.

Unser vorstehendes Resultat beantwortet seine Fragestellung.

III. Beweis des Satzes $\text{Max Min} = \text{Min Max}$.

1. Wir ändern etwas unsere Bezeichnungen ab, indem wir für Σ_1, Σ_2 bzw. $M+1, N+1$ schreiben, und für $g(p, q)$ α_{pq} . Wir haben dann:

$$h(\xi, \eta) = \sum_{p=1}^{M+1} \sum_{q=1}^{N+1} \alpha_{pq} \xi_p \eta_q.$$

Infolge der Bedingungen

$$\xi_1 + \dots + \xi_M + \xi_{M+1} = 1, \quad \eta_1 + \dots + \eta_N + \eta_{N+1} = 1$$

ist der Komplex ξ bereits durch ξ_1, \dots, ξ_M bestimmt, und ebenso der Komplex η bereits durch η_1, \dots, η_N . Es ist dann (wir haben keinen Anlaß, die Koeffizienten zu bestimmen):

$$h(\xi, \eta) = \sum_{p=1}^M \sum_{q=1}^N u_{pq} \xi_p \eta_q + \sum_{p=1}^M v_p \xi_p + \sum_{q=1}^N w_q \eta_q + r.$$

Wir werden auch hiervon nur einen Teil benutzen, indem wir stetige Funktionen zweier Variablenreihen $f(\xi, \eta)$ mit der folgenden Eigenschaft untersuchen werden:

(K). Wenn $f(\xi', \eta) \geq A$, $f(\xi'', \eta) \geq A$ ist, so ist auch für jedes $0 \leq \vartheta \leq 1$, $\xi = \vartheta \xi' + (1 - \vartheta) \xi''$ (d. h. $\xi_p = \vartheta \xi'_p + (1 - \vartheta) \xi''_p$, $p=1, 2, \dots, M$) $f(\xi, \eta) \geq A$. Wenn $f(\xi, \eta') \leq A$, $f(\xi, \eta'') \leq A$ ist, so ist auch für jedes $0 \leq \vartheta \leq 1$, $\eta = \vartheta \eta' + (1 - \vartheta) \eta''$ (d. h. $\eta_q = \vartheta \eta'_q + (1 - \vartheta) \eta''_q$, $q=1, 2, \dots, N$) $f(\xi, \eta) \leq A$.

(Daß das sowohl in den ξ wie in den η lineare $h(\xi, \eta)$ diese Eigenschaft (K) hat, ist klar.) Für diese Funktionen $f(\xi, \eta)$ werden wir beweisen:

$$\text{Max}_{\xi} \text{Min}_{\eta} f(\xi, \eta) = \text{Min}_{\eta} \text{Max}_{\xi} f(\xi, \eta),$$

wobei Max_{ξ} über $\xi_1 \geq 0, \dots, \xi_M \geq 0, \xi_1 + \dots + \xi_M \leq 1$ und Min_{η} über $\eta_1 \geq 0, \dots, \eta_N \geq 0, \eta_1 + \dots + \eta_N \leq 1$ zu erstrecken ist. Dies können wir auch so schreiben:

$$\begin{aligned} & \text{Max}_{\xi_1 \geq 0} \text{Max}_{\xi_2 \geq 0} \dots \text{Max}_{\xi_M \geq 0} \text{Min}_{\eta_1 \geq 0} \text{Min}_{\eta_2 \geq 0} \dots \text{Min}_{\eta_N \geq 0} f(\xi, \eta) \\ & \xi_1 \leq 1 \quad \xi_1 + \xi_2 \leq 1 \quad \xi_1 + \dots + \xi_M \leq 1 \quad \eta_1 \leq 1 \quad \eta_1 + \eta_2 \leq 1 \quad \eta_1 + \dots + \eta_N \leq 1 \\ & = \text{Min}_{\eta_1 \geq 0} \text{Min}_{\eta_2 \geq 0} \dots \text{Min}_{\eta_N \geq 0} \text{Max}_{\xi_1 \geq 0} \text{Max}_{\xi_2 \geq 0} \dots \text{Max}_{\xi_M \geq 0} f(\xi, \eta). \\ & \eta_1 \leq 1 \quad \eta_1 + \eta_2 \leq 1 \quad \eta_1 + \dots + \eta_N \leq 1 \quad \xi_1 \leq 1 \quad \xi_1 + \xi_2 \leq 1 \quad \xi_1 + \dots + \xi_M \leq 1 \end{aligned}$$

2. Wir führen die folgenden Bezeichnungen ein:

$$M^{\xi_r} f(\xi_1, \dots, \xi_r, \eta_1, \dots, \eta_s) = \text{Max}_{\xi_r \geq 0} f(\xi_1, \dots, \xi_r, \eta_1, \dots, \eta_s),$$

$$\xi_1 + \dots + \xi_r \leq 1$$

$$M^{\eta_s} f(\xi_1, \dots, \xi_r, \eta_1, \dots, \eta_s) = \text{Min}_{\eta_s \geq 0} f(\xi_1, \dots, \xi_r, \eta_1, \dots, \eta_s).$$

$$\eta_1 + \dots + \eta_s \leq 1$$

Wie man sieht, hebt M^{ξ_r} bzw. M^{η_s} die Abhängigkeit des f von ξ_r bzw. η_s auf. Wir wollen beweisen: wenn f der Bedingung (K) (in 1.) genügt, so ist

$$\begin{aligned} & M^{\xi_1} M^{\xi_2} \dots M^{\xi_p} M^{\eta_1} M^{\eta_2} \dots M^{\eta_q} f \\ &= M^{\eta_1} M^{\eta_2} \dots M^{\eta_q} M^{\xi_1} M^{\xi_2} \dots M^{\xi_p} f. \end{aligned}$$

Der Beweis ist offenbar geführt, wenn die zwei folgenden Behauptungen bewiesen sind:

α) Wenn $f = f(\xi_1, \dots, \xi_r, \eta_1, \dots, \eta_s)$ stetig ist und die Eigenschaft (K) hat, so gilt dasselbe von $M^{\xi_r} f$ und $M^{\eta_s} f$.

β) Wenn $f = f(\xi_1, \dots, \xi_r, \eta_1, \dots, \eta_s)$ stetig ist und die Eigenschaft (K) hat, so ist

$$M^{\xi_r} M^{\eta_s} f = M^{\eta_s} M^{\xi_r} f.$$

Zuerst beweisen wir α). Es genügt aber $M^{\xi_r} f$ zu betrachten: für $M^{\eta_s} f$ verlaufen die Überlegungen ebenso.

Es ist

$$\begin{aligned} M^{\xi_r} f(\xi_1, \dots, \xi_r, \eta_1, \dots, \eta_s) &= f^*(\xi_1, \dots, \xi_{r-1}, \eta, \dots, \eta_s) \\ &= \text{Max}_{\substack{\xi_r \geq 0 \\ \xi_1 + \dots + \xi_r \leq 1}} f(\xi_1, \dots, \xi_r, \eta_1, \dots, \eta_s). \end{aligned}$$

Daß aus der Stetigkeit von f die von f^* folgt, ist klar. Es müssen noch die zwei Eigenschaften in (K) untersucht werden.

Erstens sei

$$f^*(\xi'_1, \dots, \xi'_{r-1}, \eta_1, \dots, \eta_s) \geq A, \quad f^*(\xi''_1, \dots, \xi''_{r-1}, \eta_1, \dots, \eta_s) \geq A,$$

Die f^* entsprechen Maximalwerten von f in endlichen Intervallen, die, da f stetig ist, angenommen werden; etwa für ξ'_r bzw. ξ''_r . Dann ist

$$f(\xi'_1, \dots, \xi'_r, \eta_1, \dots, \eta_s) \geq A, \quad f(\xi''_1, \dots, \xi''_r, \eta_1, \dots, \eta_s) \geq A,$$

und weil f dem (K) genügt (wir setzen $\xi_1 = \vartheta \xi'_1 + (1 - \vartheta) \xi''_1, \dots, \xi_{r-1} = \vartheta \xi'_{r-1} + (1 - \vartheta) \xi''_{r-1}$ und $\xi_r = \vartheta \xi'_r + (1 - \vartheta) \xi''_r$)

$$f(\xi_1, \dots, \xi_r, \eta_1, \dots, \eta_s) \geq A.$$

Dabei folgt aus

$$\xi'_r \geq 0, \quad \xi'_1 + \dots + \xi'_r \leq 1, \quad \xi''_r \geq 0, \quad \xi''_1 + \dots + \xi''_r \leq 1$$

sofort

$$\xi_r \geq 0, \quad \xi_1 + \dots + \xi_r \leq 1.$$

Also ist für das Maximum f^* um so mehr

$$f^*(\xi_1, \dots, \xi_{r-1}, \eta_1, \dots, \eta_s) \geq A.$$

Zweitens sei

$$f^*(\xi_1, \dots, \xi_{r-1}, \eta'_1, \dots, \eta'_s) \leq A, \quad f^*(\xi_1, \dots, \xi_{r-1}, \eta''_1, \dots, \eta''_s) \leq A.$$

Wegen der Maximaleigenschaft von f^* gilt für alle ξ_r mit

$$\xi_r \geq 0, \quad \xi_1 + \dots + \xi_r \leq 1$$

dann

$$f^*(\xi_1, \dots, \xi_r, \eta'_1, \dots, \eta'_s) \leq A, \quad f^*(\xi_1, \dots, \xi_r, \eta''_1, \dots, \eta''_s) \leq A.$$

Da f dem (K) genügt, hat dies wieder

$$f(\xi_1, \dots, \xi_r, \eta_1, \dots, \eta_s) \leq A$$

($\eta_1 = \vartheta \eta'_1 + (1 - \vartheta) \eta''_1, \dots, \eta_s = \vartheta \eta'_s + (1 - \vartheta) \eta''_s$) zur Folge; und da das für alle oben genannte ξ_r gilt,

$$f^*(\xi_1, \dots, \xi_{r-1}, \eta_1, \dots, \eta_s) \leq A.$$

Damit ist unsere Behauptung α) restlos bewiesen.

3. Weiter soll nun gezeigt werden, daß stets (d. h. für alle $\xi_1, \dots, \xi_{r-1}, \eta_1, \dots, \eta_{s-1}$) $M^{\xi_r} M^{\eta_s} f = M^{\eta_s} M^{\xi_r} f$ ist. Wenn wir in $f(\xi_1, \dots, \xi_r, \eta_1, \dots, \eta_s)$ die Variablen $\xi_1, \dots, \xi_{r-1}, \eta_1, \dots, \eta_{s-1}$ festhalten, so genügt es, als Funktion von ξ_r, η_s allein, offenbar auch noch der Bedingung (K). Es bleibt also übrig zu beweisen (wir schreiben ξ, η für ξ_r, η_s):

Wenn $f(\xi, \eta)$ eine stetige Funktion ist, und wenn aus $f(\xi', \eta) \geq A, f(\xi'', \eta) \geq A$ für $\xi' \leq \xi \leq \xi''$ $f(\xi, \eta) \geq A$ folgt, und aus $f(\xi, \eta') \leq A, f(\xi, \eta'') \leq A$ für $\eta' \leq \eta \leq \eta''$ $f(\xi, \eta) \leq A$ folgt, so ist

$$\text{Max}_{\xi} \text{Min}_{\eta} f(\xi, \eta) = \text{Min}_{\eta} \text{Max}_{\xi} f(\xi, \eta).$$

$0 \leq \xi \leq a \quad 0 \leq \eta \leq b \quad 0 \leq \eta \leq b \quad 0 \leq \xi \leq a$

(Wir schreiben a und b für $1 - \xi_1 - \dots - \xi_{r-1}$ bzw. $1 - \eta_1 - \dots - \eta_{s-1}$.)

Die zu beweisende Behauptung kann auch so formuliert werden: Es gibt einen „Sattelpunkt“ ξ_0, η_0 ($0 \leq \xi_0 \leq a, 0 \leq \eta_0 \leq b$), d. h. $f(\xi_0, \eta)$ nimmt in $0 \leq \eta \leq b$ sein Minimum für $\eta = \eta_0$ an, und $f(\xi, \eta_0)$ nimmt in $0 \leq \xi \leq a$ sein Maximum für $\xi = \xi_0$ an.

In der Tat ist erstens jedenfalls

$$\text{Max}_{\xi} \text{Min}_{\eta} f(\xi, \eta) \leq \text{Min}_{\eta} \text{Max}_{\xi} f(\xi, \eta),$$

und zweitens folgt aus der soeben formulierten Behauptung

$$\text{Max}_{\xi} \text{Min}_{\eta} f(\xi, \eta) \geq \text{Min}_{\eta} f(\xi_0, \eta) = f(\xi_0, \eta_0)$$

$$\text{Min}_{\eta} \text{Max}_{\xi} f(\xi, \eta) \leq \text{Max}_{\xi} f(\xi, \eta_0) = f(\xi_0, \eta_0),$$

also

$$\text{Max}_{\xi} \text{Min}_{\eta} f(\xi, \eta) = \text{Min}_{\eta} \text{Max}_{\xi} f(\xi, \eta) = f(\xi_0, \eta_0).$$

Es gilt also, zwei ξ_0, η_0 von der genannten Beschaffenheit zu finden.

ξ sei fest gegeben, für welche Werte von η in $0 \leq \eta \leq b$ nimmt $f(\xi, \eta)$ sein Minimum an? Die Antwort ist leicht: Wegen der Stetigkeit von f ist diese Menge abgeschlossen und wegen der zweiten Voraussetzung über f (aus $f(\xi, \eta') \leq A$, $f(\xi, \eta'') \leq A$ folgt $f(\xi, \eta) \leq A$ für alle $\eta' \leq \eta \leq \eta''$) ist sie konvex; die einzigen abgeschlossenen und konvexen Zahlenmengen sind aber die Intervalle (mit Endpunkten). Diese Menge wird also ein Teilintervall des Intervalles $0, b$ sein; wir nennen es $K'(\xi)$, $K''(\xi)$.

Wenn η fest gegeben ist, so sieht man ebenso ein, daß diejenigen ξ in $0 \leq \xi \leq a$, für die $f(\xi, \eta)$ sein Maximum annimmt, ein Teilintervall (mit Endpunkten) von $0, a$ bilden; wir nennen es $L'(\eta)$, $L''(\eta)$.

Offenbar ist stets $K'(\xi) \leq K''(\xi)$, $L'(\eta) \leq L''(\eta)$. Ferner folgt aus der Stetigkeit von $f(\xi, \eta)$, daß $K'(\xi)$, $L'(\eta)$ nach unten, und $K''(\xi)$, $L''(\eta)$ nach oben halbstetige Funktionen sind¹⁰⁾.

Nun sei wieder ξ^* fest gegeben. Wir bilden die Menge aller ξ^{**} mit der folgenden Eigenschaft: Es gibt ein η^* , so daß $f(\xi^*, \eta)$ seinen Minimalwert (in $0 \leq \eta \leq b$) in $\eta = \eta^*$ annimmt, und $f(\xi, \eta^*)$ seinen Maximalwert (in $0 \leq \xi \leq a$) in $\xi = \xi^{**}$ annimmt. D. h.: die Vereinigungsmenge aller Intervalle $L'(\eta^*) \leq \xi^{**} \leq L''(\eta^*)$, wenn η^* das ganze Intervall $K'(\xi^*) \leq \eta^* \leq K''(\xi^*)$ durchläuft.

Im Intervalle $K'(\xi^*) \leq \eta^* \leq K''(\xi^*)$ nimmt die nach unten halbstetige Funktion $L'(\eta^*)$ ihr Minimum und die nach oben halbstetige Funktion $L''(\eta^*)$ ihr Maximum an; also hat die Menge der ξ^{**} sowohl ein kleinstes als auch ein größtes Element. Sie enthält aber auch jedes dazwischen liegende ξ' , was man sich so klarmachen kann: Wäre das nicht der Fall, so läge jedes Intervall $L'(\eta^*)$, $L''(\eta^*)$ ganz vor oder ganz nach ξ' , und es gäbe solche von jeder Sorte (die zum kleinsten bzw. größten ξ^{**} gehören). Da η^* ein Intervall durchläuft, hätten die beiden Sorten von η^* einen gemeinsamen Häufungspunkt η' . Da in beliebiger Nähe von η' also sowohl $L'(\eta^*) \leq \xi'$ als auch $L''(\eta^*) \geq \xi'$ vorkommt (und L' , L'' nach unten bzw. oben halbstetig ist), muß $L'(\eta') \leq \xi'$, $L''(\eta') \geq \xi'$ sein; d. h. ξ' gehört doch zu einem der Intervalle: zu $L'(\eta')$, $L''(\eta')$.

¹⁰⁾ Wir wollen den Beweis für $K'(\xi)$ skizzieren, für die drei anderen Funktionen geht er ebenso.

Wenn $K'(\xi) = 0$ ist, ist die Behauptung trivial, da stets $K'(\xi) \geq 0$ ist; es sei also $K'(\xi) > 0$. Für $0 \leq \eta \leq K'(\xi) - \varepsilon$ ($\varepsilon > 0$) ist stets $f(\xi, \eta) \neq \text{Min}_\eta f(\xi, \eta)$, und da $f(\xi, \eta)$ stetig ist, $f(\xi, \eta) \leq \text{Min}_\eta f(\xi, \eta) - \delta$ (für ein geeignetes $\delta > 0$). Wenn also ζ genügend nahe bei ξ liegt, so ist noch immer $f(\zeta, \eta) \leq \text{Min}_\eta f(\zeta, \eta) - \frac{1}{2}\delta$ (weil sowohl $f(\zeta, \eta)$ als auch $\text{Min}_\eta f(\zeta, \eta)$ stetig ist); d. h. $f(\zeta, \eta)$ nimmt sein Minimum (in η , für $0 \leq \eta \leq b$) in $0 \leq \eta \leq K'(\xi) - \varepsilon$ nirgends an. Also muß $K(\zeta) \geq K(\xi) - \varepsilon$. Das ist aber gerade die behauptete Halbstetigkeit nach unten.

Unsere ξ^{**} bilden also ein Teilintervall (mit Endpunkten) von $0, a$, wir nennen es $H'(\xi^*), H''(\xi^*)$. $H'(\xi^*)$ ist das Minimum der $L'(\eta^*)$, $H''(\xi^*)$ das Maximum der $L''(\eta^*)$, für $K'(\xi^*) \leq \eta^* \leq K''(\xi^*)$. Man sieht leicht ein, daß wieder $H'(\xi^*)$ nach unten und $H''(\xi^*)$ nach oben halbstetig ist (dies folgt aus den entsprechenden Eigenschaften von $K'(\xi^*)$, $K''(\xi^*)$ und $L'(\eta^*)$, $L''(\eta^*)$).

Wir sind offenbar am Ziele, wenn wir ein ξ^* ($0 \leq \xi^* \leq a$) ausfindig machen können, welches gleichzeitig ein ξ^{**} ist, d. h. eines mit $H'(\xi^*) \leq \xi^* \leq H''(\xi^*)$.

Gäbe es kein solches ξ^* , so läge jedes Intervall $H'(\xi^*), H''(\xi^*)$ ganz vor oder ganz nach ξ^* , und es gäbe solche von jeder Sorte (nämlich $\xi^* = a$ bzw. $\xi^* = 0$). Da ξ^* ein Intervall durchläuft, hätten die beiden Sorten von ξ^* einen gemeinsamen Häufungspunkt ξ' . Da in beliebiger Nähe von ξ' also sowohl $H'(\xi^*) \leq \xi^*$ also auch $H''(\xi^*) \geq \xi^*$ vorkommt (und H', H'' nach unten bzw. nach oben halbstetig ist), muß $H'(\xi') \leq \xi', H''(\xi') \geq \xi'$ sein; d. h. ξ' gehört doch zum Intervalle $H'(\xi'), H''(\xi')$.

Damit ist aber die letzte Behauptung (und somit auch die Behauptung β) bewiesen. Wir haben also unseren Satz restlos bewiesen.

IV. Der Fall $n = 3$.

Nachdem wir in den Abschnitten II, III den Fall $n = 2$ erledigt haben, wenden wir uns dem nächst komplizierten Falle $n = 3$ zu.

Es liege also ein 3-Personen-Spiel vor, das im Sinne der Beschreibung am Ende des Abschnittes I durch drei Funktionen g_1, g_2, g_3 von drei Variablen x, y, z ($x = 1, 2, \dots, \Sigma_1, y = 1, 2, \dots, \Sigma_2, z = 1, 2, \dots, \Sigma_3$) charakterisiert ist; dabei gilt identisch

$$g_1 + g_2 + g_3 = 0.$$

Es war im Falle $n = 2$ möglich, den Wert einer Partie für jeden Spieler S_1, S_2 zwingend zu bestimmen, es ergab sich:

$$\text{Wert für } S_1 = \text{Max}_\xi \text{Min}_\eta \sum_{p=1}^{\Sigma_1} \sum_{q=1}^{\Sigma_2} g(p, q) \xi_p \eta_q = \text{Max}_\xi \text{Min}_\eta \sum_{p=1}^{\Sigma_1} \sum_{q=1}^{\Sigma_2} g_1(p, q) \xi_p \eta_q,$$

$$\text{Wert für } S_2 = - \text{Min}_\eta \text{Max}_\xi \sum_{p=1}^{\Sigma_1} \sum_{q=1}^{\Sigma_2} g(p, q) \xi_p \eta_q = \text{Max}_\eta \text{Min}_\xi \sum_{p=1}^{\Sigma_1} \sum_{q=1}^{\Sigma_2} g_2(p, q) \xi_p \eta_q,$$

wobei gilt

$$\text{Wert für } S_1 + \text{Wert für } S_2 = 0.$$

Versuchen wir nun auch im Falle $n = 3$ die Werte einer Partie für die drei Spieler S_1, S_2, S_3 zu berechnen! Nehmen wir etwa an, diese Werte wären bzw. w_1, w_2, w_3 . Dann ist es klar, daß diese Werte, um

allgemein und ohne jede weitere Erörterung befriedigend zu sein, die folgende Eigenschaft haben müßten: Keine zwei Spieler dürfen in der Lage sein, sich durch Koalition beim Spiele einen größeren Erwartungswert verschaffen zu können, als die Summe der ihnen zugeteilten „Werte einer Partie“. Ferner muß $w_1 + w_2 + w_3 = 0$ sein: denn die Spieler leisten ja nur Zahlungen aneinander.

Wenn aber

$$\text{Max}_{\xi} \text{Min}_{\eta} \sum_{p=1}^{\Sigma_1} \sum_{q=1}^{\Sigma_2} \sum_{r=1}^{\Sigma_3} (g_1(pqr) + g_2(pqr)) \xi_{pq} \eta_r = M_{1,2},$$

$$\text{Max}_{\xi} \text{Min}_{\eta} \sum_{p=1}^{\Sigma_1} \sum_{q=1}^{\Sigma_2} \sum_{r=1}^{\Sigma_3} (g_1(pqr) + g_3(pqr)) \xi_{pr} \eta_q = M_{1,3},$$

$$\text{Max}_{\xi} \text{Min}_{\eta} \sum_{p=1}^{\Sigma_1} \sum_{q=1}^{\Sigma_2} \sum_{r=1}^{\Sigma_3} (g_2(pqr) + g_3(pqr)) \xi_{qr} \eta_p = M_{2,3}$$

gesetzt wird (die ξ_{pq} bilden ein System von Wahrscheinlichkeiten, ebenso die η_r ; analog für ξ_{pr} , η_q und ξ_{qr} , η_p), so können S_1 und S_2 , dadurch daß sie koalieren, ein gewöhnliches 2-Personen-Spiel gegen S_3 spielen und sich dabei (nach dem vorhin Gesagten) den Erwartungswert $M_{1,2}$ erzwingen; ebenso S_1 und S_3 den Erwartungswert $M_{1,3}$; und S_2 und S_3 den Erwartungswert $M_{2,3}$. Also muß

$$w_1 + w_2 \geq M_{1,2}, \quad w_1 + w_3 \geq M_{1,3}, \quad w_2 + w_3 \geq M_{2,3}, \\ w_1 + w_2 + w_3 = 0$$

sein. Dies ist offenbar dann und nur dann möglich, wenn

$$M_{1,2} + M_{1,3} + M_{2,3} \leq 0$$

ist.

Nun ist, wie wir in 2. zeigen werden, stets

$$M_{1,2} + M_{1,3} + M_{2,3} \geq 0,$$

und es ist leicht, Beispiele anzugeben, wo das $>$ -Zeichen gilt. Ein solches 3-Personen-Spiel ist z. B. das folgende:

$\Sigma_1 = \Sigma_2 = \Sigma_3 = 3$. Wenn es unter den x_1, x_2, x_3 (d. h. den Wahlen der S_1, S_2, S_3 , wir schreiben dafür bisher auch x, y, z) zwei solche gibt, daß $x_\mu = \nu$, $x_\nu = \mu$ ist, so bilden μ, ν ein „echtes Paar“. Es gibt offenbar entweder kein echtes Paar oder ein einziges.

Wenn es kein „echtes Paar“ gibt, so sei $g_1 = g_2 = g_3 = 0$. Wenn es ein „echtes Paar“ gibt, so sei es μ, ν , und die dritte der Zahlen 1, 2, 3 sei λ . Dann sei $g_\mu = g_\nu = 1$, $g_\lambda = -2$.

Bei diesem Spiele ist offenbar $M_{1,2} = M_{1,3} = M_{2,3} = 2$ (irgend zwei koalierte S_μ, S_ν können, indem sie ν bzw. μ wählen, ein „echtes Paar“ bilden und dadurch dem dritten, S_λ , die Summe 2 abnehmen!), also $M_{1,2} + M_{1,3} + M_{2,3} = 6 > 0$.

Der Sinn des Versagens eines jeden Wertungsversuches bei diesem Spiele ist offenbar der folgende: Um die Summe 2 zu gewinnen, brauchen sich nur irgendwelche der drei Spieler zusammenzutun, sie können dann den dritten ohne weiteres ausplündern, trotzdem die Spielregel absolut symmetrisch, d. h. das Spiel formal gerecht ist¹¹⁾. Aus der Symmetrie würde folgen, daß der Wert für jeden Spieler gleich 0 sein muß, dies ist aber offenbar falsch: Zwei Spieler brauchen nur zu wollen, und sie können sich dann den Gewinn 2 verschaffen! Wie ist dieser Widerspruch aufzulösen?

2. Gehen wir systematisch vor. Es ist stets

$$M_{1,2} + M_{1,3} + M_{2,3} \geq 0. \quad 12)$$

Denn es ist offenbar:

$$M_{1,2} = \text{Max}_{\xi} \text{Min}_{\eta} \sum_{p=1}^{\Sigma_1} \sum_{q=1}^{\Sigma_2} \sum_{r=1}^{\Sigma_3} (g_1(pqr) + g_2(pqr)) \xi_{pq} \eta_r$$

(auf Grund unseres Satzes über 2-Personen-Spiele)

$$= \text{Min}_{\eta} \text{Max}_{\xi} \sum_{p=1}^{\Sigma_1} \sum_{q=1}^{\Sigma_2} \sum_{r=1}^{\Sigma_3} (g_1(pqr) + g_2(pqr)) \xi_{pq} \eta_r$$

$$= - \text{Max}_{\eta} \text{Min}_{\xi} \sum_{p=1}^{\Sigma_1} \sum_{q=1}^{\Sigma_2} \sum_{r=1}^{\Sigma_3} g_3(pqr) \xi_{pq} \eta_r,$$

wir müssen also

$$\text{Max}_{\eta'} \text{Min}_{\xi'} \sum_{p,q,r} g_3(pqr) \xi'_{pq} \eta'_r + \text{Max}_{\eta''} \text{Min}_{\xi''} \sum_{p,q,r} g_2(pqr) \xi''_{pr} \eta''_q$$

$$+ \text{Max}_{\eta'''} \text{Min}_{\xi'''} \sum_{p,q,r} g_1(pqr) \xi'''_{qr} \eta'''_p \leq 0$$

beweisen, d. h. für alle Systeme $\bar{\eta}'$, $\bar{\eta}''$, $\bar{\eta}'''$

$$\text{Min}_{\xi'} \sum_{p,q,r} g_3(pqr) \xi'_{pq} \bar{\eta}'_r + \text{Min}_{\xi''} \sum_{p,q,r} g_2(pqr) \xi''_{pr} \bar{\eta}''_q$$

$$+ \text{Min}_{\xi'''} \sum_{p,q,r} g_1(pqr) \xi'''_{qr} \bar{\eta}'''_p \leq 0.$$

¹¹⁾ Man sieht hieran, daß unser Beispiel alles andere als ein Fall von „Pathologie“ von Spielen ist: es ist vielmehr ein in praxi recht häufiger und charakteristischer Fall. Im Einklang damit werden wir in IV, 3. und V, 1. sehen, daß es sogar der allgemeine Fall des 3-Personen-Spieles ist.

¹²⁾ Inhaltlich ist dies ohne weiteres klar: S_2 und S_3 können in Koalition gegen S_1 bestenfalls $M_{2,3}$ erzwingen, also S_1 für sich allein (gegen alle) bestenfalls $-M_{2,3}$ (wegen unseres Satzes über das 2-Personen-Spiel); ebenso kann S_2 für sich allein bestenfalls $-M_{1,3}$ erzwingen. Koaliert können aber S_1 und S_2 bestenfalls $M_{1,2}$ erzwingen; „l'union fait la force“, d. h.

$$-M_{2,3} - M_{1,3} \leq M_{1,2}, \quad M_{1,2} + M_{1,3} + M_{2,3} \geq 0.$$

Dies ist aber der Fall, es genügt ja

$$\xi'_{p,q} = \bar{\eta}'''_q \bar{\eta}''_q, \quad \xi''_{p,r} = \bar{\eta}'''_r \bar{\eta}'_r, \quad \xi'''_q = \bar{\eta}''_q \bar{\eta}'_r$$

zu setzen, dann wird (wegen $g_1 + g_2 + g_3 = 0$)

$$\sum_{p,q,r} g_3(pqr) \bar{\eta}'_r \bar{\eta}''_q \bar{\eta}'''_p + \sum_{p,q,r} g_2(pqr) \bar{\eta}'_r \bar{\eta}''_q \bar{\eta}'''_p + \sum_{p,q,r} g_1(pqr) \bar{\eta}'_r \bar{\eta}''_p \bar{\eta}'''_q = 0.$$

Daß das Zeichen $>$ wirklich vorkommt, haben wir gesehen, der Fall des Zeichens $=$ ist also als ein ausgearteter Grenzfall anzusehen.

Nehmen wir nun an, der Spieler S_1 erhebt Anspruch auf einen Gewinn von w_1 pro Partie. Wie kann er seinen Anspruch durchsetzen? Offenbar auf zwei Wegen.

Erstens kann er versuchen allein zu spielen. Dann gerät er im wesentlichen in ein 2-Personen-Spiel, in dem er auf der einen Seite steht und S_2, S_3 auf der anderen (koaliert). Der Wert dieses Spieles für ihn ist also $-M_{2,3}$ pro Partie. Diese Lösung kommt also nur für $w_1 \leq -M_{2,3}$ in Frage; nehmen wir daher $w_1 > -M_{2,3}$ an.

Dann bleibt nur die zweite Möglichkeit übrig: er muß versuchen S_2 oder S_3 zum Bundesgenossen zu bekommen. Da er im Bunde mit S_2 oder S_3 pro Partie die Summe $M_{1,2}$ oder $M_{1,3}$ gewinnen kann, aber davon w_1 für sich behalten will, so kann er S_2 die Summe $M_{1,2} - w_1$ pro Partie als Preis des Bündnisses anbieten, und S_3 die Summe $M_{1,3} - w_1$. Es ist jedoch vollkommen ausgeschlossen, daß S_2 oder S_3 dieses Angebot annimmt, wenn sie miteinander verbündet mehr als $(M_{1,2} - w_1) + (M_{1,3} - w_1)$ pro Partie gewinnen können. D. h. wenn

$$(M_{1,2} - w_1) + (M_{1,3} - w_1) < M_{2,3}; \quad w_1 > \frac{1}{2}(M_{1,2} + M_{1,3} - M_{2,3})$$

ist.

Wir können also sagen: S_1 hat gar keine Aussicht, einen Anspruch w_1 durchzusetzen, der

$$> -M_{2,3}, \quad > \frac{1}{2}(M_{1,2} + M_{1,3} - M_{2,3})$$

ist. Die zweite Zahl ist \geq als die erste (wegen $M_{1,2} + M_{1,3} + M_{2,3} \geq 0$), also muß jedenfalls

$$w_1 \leq \frac{1}{2}(M_{1,2} + M_{1,3} - M_{2,3}) = \bar{w}_1$$

sein. Ebenso kann gezeigt werden: es muß

$$w_2 \leq \frac{1}{2}(M_{1,2} + M_{2,3} - M_{1,3}) = \bar{w}_2,$$

$$w_3 \leq \frac{1}{2}(M_{1,3} + M_{2,3} - M_{1,2}) = \bar{w}_3$$

sein.

Nun sind aber diese oberen Grenzen für die Ansprüche der drei Spieler, $\bar{w}_1, \bar{w}_2, \bar{w}_3$, ohne weiteres realisierbar. Denn wenn sich etwa S_1, S_2 verbünden, so können sie (gegen S_3) den Gewinn $M_{1,2} = \bar{w}_1 + \bar{w}_2$ erzielen; und ebenso können sich S_1, S_3 bzw. S_2, S_3 durch ein Bündnis die Gewinne $M_{1,3} = \bar{w}_1 + \bar{w}_3$ bzw. $M_{2,3} = \bar{w}_2 + \bar{w}_3$ pro Partie sichern. Also: die höchstmöglichen und dabei dennoch vollkommen motivierten Ansprüche der drei Spieler S_1, S_2, S_3 sind bzw. $\bar{w}_1, \bar{w}_2, \bar{w}_3$ (als Gewinn pro Partie).

3. Inwiefern ist aber diese Wertung mit der in 1. erkannten Unmöglichkeit einer allgemeinen Wertung vereinbar? Wenn $M_{1,2} + M_{1,3} + M_{2,3} = 0$ ist, so besteht ja keine Schwierigkeit: dann ist

$$\bar{w}_1 = -M_{2,3}, \quad \bar{w}_2 = -M_{1,3}, \quad \bar{w}_3 = -M_{1,2},$$

d. h. jeder Spieler kann seinen Ansprüchen allein, auch ohne Hilfe eines anderen (und einer möglichen Koalition seiner Gegner trotzend), Geltung verschaffen. Es können also alle drei Spieler ihre Ansprüche gleichzeitig durchsetzen, demgemäß ist auch

$$\bar{w}_1 + \bar{w}_2 + \bar{w}_3 = 0.$$

Anders ist es für $M_{1,2} + M_{1,3} + M_{2,3} > 0$. Wegen

$$\bar{w}_1 > -M_{2,3}, \quad \bar{w}_2 > -M_{1,3}, \quad \bar{w}_3 > -M_{1,2}$$

kann dann kein Spieler seinen Anspruch allein durchsetzen, und wegen

$$\bar{w}_1 + \bar{w}_2 + \bar{w}_3 = \frac{1}{2}(M_{1,2} + M_{1,3} + M_{2,3}) > 0$$

können niemals alle drei zugleich befriedigt werden. Aber wegen

$$\bar{w}_1 + \bar{w}_2 = M_{1,2}, \quad \bar{w}_1 + \bar{w}_3 = M_{1,3}, \quad \bar{w}_2 + \bar{w}_3 = M_{2,3}$$

ist jedes Paar von Spielern, welches sich (zum Ausplündern des dritten) verbündet, des Erfolges gewiß: sie können ihre Ansprüche voll befriedigen, der dritte wird freilich pro Partie nur bzw. $-M_{2,3}, -M_{1,3}, -M_{1,2}$ erhalten, und daher um $\frac{1}{2}(M_{1,2} + M_{1,3} + M_{2,3})$ hinter seinem motivierten Anspruch zurückbleiben.

Dies kann auch so formuliert werden: Jeder der drei Spieler S_1, S_2, S_3 muß trachten, sich mit einem anderen Spieler zu verbünden. Wenn ihm das gelingt, so erhält er pro Partie die bzw. Summe

$$\frac{1}{2}(M_{1,2} + M_{1,3} - M_{2,3}), \quad \frac{1}{2}(M_{1,2} + M_{2,3} - M_{1,3}),$$

$$\frac{1}{2}(M_{1,3} + M_{2,3} - M_{1,2}),$$

wenn es ihm aber nicht gelingt (d. h. wenn die zwei anderen koalieren), so erhält er bzw. nur

$$-M_{2,3}, \quad -M_{1,3}, \quad -M_{1,2}.$$

Eine noch etwas variierte Beschreibung des Sachverhaltes, die vielleicht die prägnanteste ist, ist die folgende:

α) Eine Partie hat für die Spieler S_1, S_2, S_3 die bzw. „Grundwerte“

$$v_1 = \frac{1}{3}(M_{1,2} + M_{1,3} - 2M_{2,3}), \quad v_2 = \frac{1}{3}(M_{1,2} + M_{2,3} - 2M_{1,3}),$$

$$v_3 = \frac{1}{3}(M_{1,3} + M_{2,3} - 2M_{1,2}).$$

Das ist eine regelrechte Wertung da $v_1 + v_2 + v_3 = 0$ ist.

β) Aber über die „Grundwerte“ hinaus besteht für irgend zwei Spieler, die sich gegen den dritten verbünden, die Möglichkeit, je $\frac{1}{6}D$ zu gewinnen, während der dritte (gleichfalls über seinen „Grundwert“ hinaus) $\frac{1}{3}D$ verliert. Dabei ist

$$D = M_{1,2} + M_{1,3} + M_{2,3} > 0 \quad {}^{13)}.$$

(Auch der zuerst behandelte Fall $D = M_{1,2} + M_{1,3} + M_{2,3} = 0$ kann in diese Formulierung mit einbezogen werden: *α)* ist das dortige Resultat und *β)* fällt wegen $D = 0$ fort.)

Man sieht an dieser Lösung sofort: das 3-Personen-Spiel ist etwas wesentlich anderes als das von zwei Personen. Die eigentliche Spielmethode der einzelnen Spieler tritt zurück: sie bietet nichts Neues, da die (unbedingt eintretende) Bildung von Koalitionen das Spiel zu einem 2-Personen-Spiele macht. Aber der Wert der Partie für einen Spieler hängt nicht nur von der Spielregel ab, er wird vielmehr ganz entscheidend dadurch beeinflusst (wenigstens, sobald $D > 0$ ist), welche der drei an sich gleichmöglichen Koalitionen $S_1, S_2; S_1, S_3; S_2, S_3$ zustande gekommen ist. Es macht sich geltend, was dem schablonenmäßigen und ganz ausgeglichenen 2-Personen-Spiele noch völlig fremd ist: der Kampf.

V. Ansätze für $n > 3$.

1. Für $n > 3$ ist es bis jetzt nicht gelungen, allgemeingültige Resultate zu erzielen. Der beste Wegweiser, der hier zur Verfügung steht, mag die Analogie zu den bereits erledigten Fällen $n = 2, 3$ sein; diese sollen deshalb hier noch einmal zusammengestellt werden:

¹³⁾ Es ist übrigens

$$v_1 = -M_{2,3} + \frac{1}{3}D, \quad v_2 = -M_{1,3} + \frac{1}{3}D, \quad v_3 = -M_{1,2} + \frac{1}{3}D.$$

$n = 2$. Es wird definiert:

$$M = \text{Max}_{\xi} \text{Min}_{\eta} \sum_{p,q} g_1(pq) \xi_p \eta_q^{14}.$$

Das Spiel hat für die Spieler S_1, S_2 die bzw. Werte $M, -M$ pro Partie.

$n = 3$. Es wird definiert:

$$M_{1,2} = \text{Max}_{\xi} \text{Min}_{\eta} \sum_{p,q,r} (g_1(pqr) + g_2(pqr)) \xi_{pq} \eta_r$$

$$M_{1,3} = \text{Max}_{\xi} \text{Min}_{\eta} \sum_{p,q,r} (g_1(pqr) + g_3(pqr)) \xi_{pr} \eta_q$$

$$M_{2,3} = \text{Max}_{\xi} \text{Min}_{\eta} \sum_{p,q,r} (g_2(pqr) + g_3(pqr)) \xi_{qr} \eta_p^{14}$$

$$D = M_{1,2} + M_{1,3} + M_{2,3}.$$

Es ist $D \geq 0$, und es sind zwei Fälle zu unterscheiden, je nachdem, ob $D = 0$ oder $D > 0$ ist.

$D = 0$. In diesem Falle hat das Spiel für S_1, S_2, S_3 die bzw. Werte $-M_{2,3}, -M_{1,3}, -M_{1,2}$ pro Partie.

$D > 0$. In diesem Falle hat das Spiel für S_1, S_2, S_3 die bzw. „Grundwerte“ $-M_{2,3} + \frac{1}{3}D, -M_{1,3} + \frac{1}{3}D, -M_{1,2} + \frac{1}{3}D$ pro Partie. Zu den „Grundwerten“ ist aber noch ein weiteres Glied zu addieren, um die richtigen Werte zu erhalten; dieses rührt daher, daß irgend zwei Spieler, die sich gegen den dritten verbünden (einerlei welche zwei!), sich über den „Grundwert“ hinaus noch einen Gewinn von je $\frac{1}{3}D$ pro Partie verschaffen können, während der dritte Spieler $\frac{1}{3}D$ pro Partie (über seinen Grundwert hinaus) verliert.

Aus dieser Zusammenstellung sieht man klar: der Fall $n = 2$ und der Fall $n = 3, D = 0$ gehören zum selben Typus. Dagegen repräsentiert der Fall $n = 3, D > 0$ (wie bereits am Schlusse von IV festgestellt wurde) einen neuen Typus. Wir wollen diese zwei Typen als den eindeutigen bzw. den symmetrisch-mehrdeutigen bezeichnen; es ist wohl zu erkennen, was mit diesen Benennungen gemeint ist.

Besteht nun Aussicht, daß sich auch für $n > 3$ alle Gesellschaftsspiele auf diese zwei Typen bringen lassen? Oder hat man mit der Möglichkeit neuer Komplikationen zu rechnen? Es wäre insbesondere das Auftreten von asymmetrisch-mehrdeutigen Typen ins Auge zu fassen, d. h. von solchen, bei denen die entscheidenden Koalitions-Möglichkeiten nicht mehr

¹⁴) Die Max_{ξ} und Min_{η} sind zu erstrecken über alle Systeme von Wahrscheinlichkeiten, d. h. wir verlangen

$$\text{alle } \xi_p \geq 0, \sum_p \xi_p = 1; \quad \text{alle } \xi_{pq} \geq 0, \sum_{pq} \xi_{pq} = 1; \quad \text{usw.}$$

und analog

$$\text{alle } \eta_p \geq 0, \sum_p \eta_p = 1; \quad \text{usw.}$$

symmetrisch über alle Spieler verteilt sind. (Bei $n = 3$ ist das ja nicht der Fall: die eventuellen Asymmetrien der Spielregel gehen völlig in den „Grundwerten“ der drei Spieler auf, zur Koalitionsbildung aber sind alle Spieler gleichfähig: denn alle drei Koalitionen S_1, S_2 ; S_1, S_3 ; S_2, S_3 kommen gleichmäßig in Betracht.) Diese Frage soll im folgenden etwas näher betrachtet werden.

2. Um ein allgemeines n -Personen-Spiel zu charakterisieren, führen wir die folgenden Konstanten ein:

$$M_{\{\mu_1, \mu_2, \dots, \mu_k\}} = \text{Max}_{\xi} \text{Min}_{\eta} \sum_{p_1=1}^{\sum_1} \sum_{p_2=1}^{\sum_2} \dots \sum_{p_n=1}^{\sum_n} (g_{\mu_1}(p_1, \dots, p_n) + \dots + g_{\mu_k}(p_1, \dots, p_n)) \xi_{p_{\mu_1}, \dots, p_{\mu_k}} \eta_{p_{\nu_1}, \dots, p_{\nu_{n-k}}},$$

wobei $\mu_1, \mu_2, \dots, \mu_k$ irgendwelche k verschiedene unter den Zahlen $1, 2, \dots, n$ sind und $\nu_1, \nu_2, \dots, \nu_{n-k}$ die übrigen (Max_{ξ} ist zu nehmen für alle $\xi_{p_{\mu_1}, \dots, p_{\mu_k}} \geq 0$, $\sum \xi_{p_{\mu_1}, \dots, p_{\mu_k}} = 1$, und Min_{η} für alle $\eta_{\nu_1, \dots, \nu_{n-k}} \geq 0$, $\sum \eta_{\nu_1, \dots, \nu_{n-k}} = 1$). $M_{\{\mu_1, \mu_2, \dots, \mu_k\}}$ ist offenbar diejenige Summe, deren Gewinn pro Partie die koalitierten Spieler $S_{\mu_1}, \dots, S_{\mu_k}$ gegen die koalitierten Spieler $S_{\nu_1}, \dots, S_{\nu_{n-k}}$ erzwingen können (das Spiel ist ja nur ein 2-Personen-Spiel).

Offenbar ist $M_{\{\}} = 0$. Ferner folgt aus unserem Satze über 2-Personen-Spiele, daß $M_{\{\mu_1, \dots, \mu_k\}} = -M_{\{\nu_1, \dots, \nu_{n-k}\}}$ ist. Schließlich seien μ_1, \dots, μ_k ; ν_1, \dots, ν_l ; $\varrho_1, \dots, \varrho_{n-k-l}$ drei zueinander komplementäre Teilmengen von $1, 2, \dots, n$. Wenn die Spieler $S_{\mu_1}, \dots, S_{\mu_k}$, ferner $S_{\nu_1}, \dots, S_{\nu_l}$ und $S_{\varrho_1}, \dots, S_{\varrho_{n-k-l}}$ fest koalitiert sind, so ist dies ein 3-Personen-Spiel, und es ist (wir versehen die auf dieses letztere Spiel bezüglichen Größen mit einem Akzent)

$$M'_{1,2} = M_{\{\mu_1, \dots, \mu_k, \nu_1, \dots, \nu_l\}}$$

$$M'_{2,3} = M_{\{\mu_1, \dots, \mu_k, \varrho_1, \dots, \varrho_{n-k-l}\}} = -M_{\{\nu_1, \dots, \nu_l\}},$$

$$M'_{1,3} = M_{\{\nu_1, \dots, \nu_l, \varrho_1, \dots, \varrho_{n-k-l}\}} = -M_{\{\mu_1, \dots, \mu_k\}}.$$

Nach unseren Resultaten über 3-Personen-Spiele ist aber

$$M'_{1,2} + M'_{1,3} + M'_{2,3} \geq 0,$$

d. h.

$$M_{\{\mu_1, \dots, \mu_k, \nu_1, \dots, \nu_l\}} \geq M_{\{\mu_1, \dots, \mu_k\}} + M_{\{\nu_1, \dots, \nu_l\}}.$$

Zusammenfassend kann also gesagt werden:

Ein gegebenes n -Personen-Spiel ordnet jeder Teilmenge $\mu_1, \mu_2, \dots, \mu_k$ von $1, 2, \dots, n$ eine Konstante $M_{\{\mu_1, \dots, \mu_k\}}$ zu (nämlich diejenige Summe, deren Gewinn pro Partie die Koalition der Spieler $S_{\mu_1}, \dots, S_{\mu_k}$ gegen die Koalition der übrigen erzwingen kann). Das System der Konstanten $M_{\{\mu_1, \dots, \mu_k\}}$ erfüllt stets die folgenden drei Bedingungen:

$$1. M_{\{ \}} = 0.$$

2. $M_{\{\mu_1, \dots, \mu_k\}} + M_{\{\nu_1, \dots, \nu_{n-k}\}} = 0$, wenn μ_1, \dots, μ_k und ν_1, \dots, ν_{n-k} komplementäre Teilmengen von $1, 2, \dots, n$ sind.

3. $M_{\{\mu_1, \dots, \mu_k, \nu_1, \dots, \nu_l\}} \geq M_{\{\mu_1, \dots, \mu_k\}} + M_{\{\nu_1, \dots, \nu_l\}}$, wenn μ_1, \dots, μ_k und ν_1, \dots, ν_l elementfremde Teilmengen von $1, 2, \dots, n$ sind¹⁵⁾.

Es ist nicht schwer, die Umkehrung zu beweisen, d. h. zu jedem System von Zahlen $M_{\{\mathfrak{M}\}}$ (\mathfrak{M} durchläuft alle 2^n Teilmengen von $1, 2, \dots, n$), das den Bedingungen 1.—3. genügt, ein Gesellschaftsspiel anzugeben, bei dem die genannten Konstanten eben diese Werte $M_{\{\mathfrak{M}\}}$ haben. Wir sehen davon ab, hier ein solches Beispiel — das keineswegs tieflegend ist — durchzudiskutieren.

3. Ich glaube die Vermutung aussprechen zu dürfen, daß die Wert- und Koalitionsverhältnisse bei einem Gesellschaftsspiele durch diese 2^n Konstanten allein bestimmt sind. Für $n = 2, 3$ ist das, wie wir sahen, der Fall¹⁶⁾, für $n > 3$ steht der allgemeine Beweis noch aus. Denn während bei $n = 2$ überhaupt nicht koalitiert werden kann und bei $n = 3$ nur auf eine Art (nämlich „zwei gegen einen“), wachsen die Möglichkeiten für $n = 3$ rasch an: schon bei $n = 4$ muß man entscheiden, ob Koalitionen „drei gegen einen“ oder „zwei gegen zwei“ sich bilden werden, d. h. bei welchen Bündnissen die daran beteiligten Spieler die besten Chancen haben werden. Bei $n = 4$ gelingt noch die Diskussion der Hauptfälle (allein auf Grund der $M_{\{\mathfrak{M}\}}$!), aber eine befriedigende allgemeine Theorie fehlt zur Zeit noch.

Wenn unsere Vermutung richtig ist, so haben wir damit alle Gesellschaftsspiele auf eine letzte natürliche Normalform gebracht: jedes System von 2^n Konstanten $M_{\{\mathfrak{M}\}}$, die den Bedingungen 1.—3. genügen, stellt eine Klasse „taktisch äquivalenter“ Gesellschaftsspiele vor¹⁷⁾.

¹⁵⁾ Inhaltlich ist diese Behauptung ebenso klar, wie die in Fußnote ¹²⁾ S. 313 betrachtete.

¹⁶⁾ Es ist für $n = 2$

$$M_{\{ \}} = 0, \quad M_{\{1\}} = M, \quad M_{\{2\}} = -M, \quad M_{\{1,2\}} = 0;$$

und für $n = 3$

$$M_{\{ \}} = 0, \quad M_{\{1\}} = -M_{2,3}, \quad M_{\{2\}} = -M_{1,3}, \quad M_{\{3\}} = -M_{1,2}, \quad M_{\{1,2\}} = M_{1,2}, \\ M_{\{1,3\}} = M_{1,3}, \quad M_{\{2,3\}} = M_{2,3}, \quad M_{\{1,2,3\}} = 0.$$

¹⁷⁾ Eine gewisse Normierungsmöglichkeit für die $M_{\{\mathfrak{M}\}}$ besteht noch darin, daß man, in Analogie zu den „Werten“ (einer Partie) für $n = 2$ und den „Grundwerten“ für $n = 3$, „Grundwerte“ v_1, v_2, \dots, v_n für die Spieler S_1, S_2, \dots, S_n einführt. Für den darüber hinausgehenden Teil des Spieles erhält man dann natürlich die neuen Konstanten

$$M_{\{\mathfrak{M}\}}^* = M_{\{\mathfrak{M}\}} - \sum_{p \text{ in } \mathfrak{G}} v_p.$$

Zum Schlusse möchte ich noch bemerken, daß in einem demnächst erscheinenden Nachtrage eine numerische Berechnung von einigen bekannten 2-Personen-Spielen erfolgen soll (Poker, allerdings mit gewissen schematisierenden Vereinfachungen, Bakkarat). Die Übereinstimmung der dabei herauskommenden Resultate mit den bekannten Faustregeln der Spieler (so z. B. der Beweis der Notwendigkeit des „Bluffens“ beim Poker) kann als eine empirische Bestätigung der Resultate unserer Theorie angesehen werden.

Man wählt die v_p zweckmäßigerweise so, daß

$$M_{\{1\}}^* = M_{\{2\}}^* = \dots = M_{\{n\}}^*, \quad v_1 + v_2 + \dots + v_n = 0$$

ist, d. h. jeder Spieler für sich allein gleichstark ist, und nur in den Koalitionsmöglichkeiten Unterschiede bestehen.

(Aus 1.—3. folgert man leicht, daß der gemeinsame Wert der

$$M_{\{1\}}^*, M_{\{2\}}^*, \dots, M_{\{n\}}^* \leq 0$$

ist; wenn er = 0 ist, so sind alle $M_{\{p\}}^* = 0$, d. h. das Spiel — nach Auszahlung der „Grundwerte“ — eindeutig. Er gibt somit eine Art Maß für die Mehrdeutigkeit des Spieles, d. h. die taktischen Möglichkeiten, die es bietet).

(Eingegangen am 24. 7. 1927.)

Zur Axiomatik der Geometrie. I.

Über Hilberts Vollständigkeitsaxiom¹⁾.

Von

Richard Baldus in Karlsruhe.

1. Das am eingehendsten untersuchte und am weitesten bekannte Axiomensystem der Euklidischen Geometrie ist das von D. Hilbert in seinen berühmten „Grundlagen der Geometrie“ entwickelte. Es bestand in der ursprünglichen Fassung, in der Festschrift zur Enthüllung des Göttinger Gauß-Weber-Denkmal (1899), aus folgenden fünf Axiomgruppen: I. Axiome der Verknüpfung, II. Axiome der Anordnung, III. Parallelenaxiom, IV. Axiome der Kongruenz, V. Archimedisches Axiom. Von der 2. Auflage (1903) an wurde das Parallelenaxiom hinter die Kongruenzaxiome geschoben, weiterhin wurde in der letzten Gruppe dem Archimedischen Axiom das Vollständigkeitsaxiom hinzugefügt, das nach Hilbert nun „den Schlußstein des ganzen Axiomensystems bildet“²⁾.

2. Der Wortlaut des Vollständigkeitsaxioms ist³⁾: V2 (Axiom der Vollständigkeit). Die Elemente (Punkte, Geraden, Ebenen) der Geometrie bilden ein System von Dingen, welches bei Aufrechterhaltung sämtlicher genannten Axiome keiner Erweiterung mehr fähig ist, d. h.: zu dem System der Punkte, Geraden, Ebenen ist es nicht möglich, ein anderes System von Dingen hinzuzufügen, so daß in dem durch Zusammensetzung entstehenden System sämtliche aufgeführten Axiome I bis IV, V1 erfüllt sind.“ Gemeint

¹⁾ Bis auf Nr. 8 und 9 im wesentlichen, mit Ausnahme der Anmerkungen, vorgetragen auf der Kissinger Mathematikertagung, September 1927, vgl. Jahresber. d. Deutsch. Mathem. Vergg. 37 (1928), 2. Abt., S. 3. Auf Anregung von Herrn P. Bernays hin oder veranlaßt durch die neuesten Arbeiten der Herren P. Finsler und B. Baer sind zur ursprünglichen Fassung der vorliegenden Arbeit aus dem September 1927 Nr. 8 und 9, sowie die Anmerkungen ²³⁾, ²⁴⁾, ²⁶⁾ hinzugekommen.

²⁾ D. Hilbert, „Grundlagen der Geometrie“ (weiterhin kurz als „Grundlagen“ zitiert), 4. bis 6. Auflage, S. 23.

³⁾ „Grundlagen“, S. 22.

ist dabei nach Hilbert eine Erweiterung, bei der die früheren Axiome in der früheren Weise gültig bleiben sollen, d. h. ein Punkt, der vor der Erweiterung zwischen zwei Punkten liegt, soll dies auch nach der Erweiterung tun, kongruente Strecken und Winkel sollen kongruent bleiben usw.

Den Axiomen der 1. Auflage mußte das Vollständigkeitsaxiom hinzugefügt werden, weil dadurch erst jede Deutung des Axiomensystems notwendig zur Cartesischen analytischen Geometrie führt; man erspart damit die Einführung eines ausdrücklichen Stetigkeitsaxioms, z. B. des noch später zu nennenden Cantorsche Axioms.

Das Vollständigkeitsaxiom stellt die originellste Leistung Hilberts in der Axiomatik dar und ist außer in der Geometrie auch in den Axiomensystemen anderer mathematischer Gebiete aufgenommen worden⁴). Um so merkwürdiger ist es, daß es in der Literatur, wenigstens soweit ich sie kenne, keine Stelle gibt, die über das von Hilbert selbst Gesagte mathematisch-axiomatisch wesentlich hinausgehend sich mit diesem merkwürdigen Axiom befaßt; man findet nur rein referierende Erwähnungen, ohne eigene Stellungnahme⁵). Und doch läßt sich, wie wir erkennen werden, noch manches darüber sagen.

3. Das Hilbertsche Vollständigkeitsaxiom fordert die Vollständigkeit der drei Systeme von gedachten Dingen, der Punkte, Geraden und Ebenen. $\{A\}$ sei nun eine Deutung des Axiomensystems, die zwar in den Punkten vollständig, aber in den übrigen Elementen (Geraden, Ebenen) erweiterungsfähig sei. $\{B\}$ sei eine solche aus $\{A\}$ erweiterte Deutung, g eine ihrer nicht in $\{A\}$ enthaltenen Geraden. Dann müßte g zufolge Axiom I 3 wenigstens zwei Punkte von $\{B\}$ und damit auch von $\{A\}$ enthalten, diese würden nach I 1 in $\{A\}$ eine Gerade bestimmen, die auch Gerade von $\{B\}$ wäre, und als solche nach I 2 mit g identisch sein müßte. Daher ist $\{A\}$ in den Geraden und, wie man entsprechend erkennt, auch in den Ebenen nicht erweiterungsfähig⁶). Die Systeme der Geraden und

⁴) So von Hilbert selbst schon vor der Geometrie, 1890, in der Arithmetik, vgl. „Grundlagen“ S. 240; weiterhin auch in der Mengenlehre, vgl. P. Finsler, „Über die Grundlegung der Mengenlehre. I. Teil“, Math. Zeitschr. 25 (1926), S. 683—713, insbesondere S. 691.

⁵) Eine Ausnahme macht M. Geiger, „Systematische Axiomatik der Euklidischen Geometrie“, Augsburg 1924, S. 271, in dem Anhang S. 265—269, doch beziehen sich seine Bemerkungen mehr auf die Ästhetik der Axiomatik, während es uns hier auf das wirkliche, man könnte sagen praktische, Arbeiten mit den Axiomen ankommt.

⁶) Derselbe Beweis liefert den schärferen Satz: Enthalten zwei Deutungen der Hilbertschen Axiome der Verknüpfung dieselben Punkte und sind die Geraden (Ebenen) der einen von beiden Deutungen auch Gerade (Ebenen) in der anderen, dann enthalten die beiden Deutungen auch dieselben Geraden (Ebenen). Die Anwendung dieses Satzes auf die Geraden und Ebenen zeigt, daß die Deutungen identisch sind.

Ebenen sind demnach vollständig, wenn das System der Punkte vollständig ist, d. h.:

Man kann Hilberts Vollständigkeitsaxiom dadurch verschärfen, daß man nur die Vollständigkeit der Punkte (nicht auch die der Geraden und Ebenen) fordert.

4. Die axiomatische Frage, ob in der Euklidischen Geometrie das Parallelenaxiom aus den übrigen Axiomen beweisbar sei, hat in bekannter Weise zur Nichteuklidischen Geometrie im engsten Sinne, zur hyperbolischen Geometrie geführt, in welcher ein dem Euklidischen widersprechendes Parallelenaxiom gilt⁷⁾, während alle übrigen Axiome der Euklidischen Geometrie beibehalten werden. Sucht man in möglichst weitgehender Anlehnung an Hilberts Axiomensystem der Euklidischen Geometrie ein Axiomensystem der hyperbolischen Geometrie, dann wird man naturgemäß das Parallelenaxiom noch weiter zurückschieben als Hilbert, nämlich an das Ende des Axiomensystems, muß dann aber das Vollständigkeitsaxiom als Schlußstein weglassen und (vor dem Parallelenaxiom) ein Stetigkeitsaxiom verwenden, etwa das Cantorsche, das, kurz gesprochen, lautet: „Enthält eine Gerade eine unendliche Folge von Strecken, von denen jede innerhalb der vorhergehenden liegt, und konvergieren die Längen dieser Strecken gegen Null, dann gibt es einen Punkt, der innerhalb aller dieser Strecken liegt“⁸⁾.

Man hat demnach nun folgende Reihenfolge der Axiomgruppen für die Euklidische Geometrie: I. Verknüpfung, II. Anordnung, III. Kongruenz, IV. Stetigkeit (Archimedisches und Cantorsches Axiom), V. Euklidisches Parallelenaxiom⁹⁾. Setzt man an die Stelle von V ein hyperbolisches Parallelenaxiom, dann erhält man statt der Euklidischen die hyperbolische Geometrie. *Diese Axiomgruppen I bis IV liefern daher den diesen beiden Geometrien gemeinsamen Teil, das ist in der Bezeichnung J. Bolyais die absolute Geometrie.*

5. In der absoluten Geometrie kann man nun in bekannter Weise (rechtwinklige) Punktkoordinaten einführen, indem man zunächst in einer Geraden, der X -Achse, einen Koordinatenanfangspunkt O und einen von ihm verschiedenen Einheitspunkt E annimmt und irgendeinem Punkt R

⁷⁾ Bei R. Baldus, „Nichteuklidische Geometrie (Hyperbolische Geometrie der Ebene)“, Sammlung Göschen, Berlin und Leipzig 1927, 152 S. (weiterhin zitiert als „N.G.“) findet man S. 70 eine enge Fassung eines Nichteuklidischen Parallelenaxioms.

⁸⁾ „N.G.“, S. 44—45 enthält eine engere Fassung des Cantorsche Axioms, die, wie in einer späteren Note gezeigt werden wird, noch verschärft werden kann. Hier, wo es sich um die Betrachtung des Vollständigkeitsaxioms handelt, möge davon abgesehen werden.

⁹⁾ Enge Fassung „N.G.“, S. 55.

der X -Achse als x -Koordinate die mit der Einheit OE gemessene Länge der Strecke OR zuordnet, und zwar mit dem positiven oder negativen Vorzeichen, je nachdem R auf der Halbgeraden $O \rightarrow E$ liegt oder nicht. Nach dem Archimedischen Axiom gehört zu jedem Punkte der X -Achse ein reeller Koordinatenwert, nach dem Cantorsche Axiom erschöpfen diese Koordinatenwerte die reellen Zahlen (d. h. die unendlichen Dezimalbrüche). Die Beziehung zwischen den Punkten der X -Achse und den reellen Zahlen (einschließlich der Null) ist daher *ausnahmslos umkehrbar-eindeutig*¹⁰⁾. Nun wählt man eine Ebene durch die X -Achse als (X, Y) -Ebene und versieht ihre beiden durch die X -Achse getrennten Halbebenen mit entgegengesetzten Vorzeichen. Jetzt findet man zu irgendeinem Punkte Q der (X, Y) -Ebene ein reelles Zahlenpaar (x, y) von Koordinaten, indem man das Lot aus Q auf die X -Achse fällt, dessen — immer existierender — Fußpunkt R sei; die x -Koordinate von R ist auch die von Q , und die mit der Einheit OE gemessene Länge der Strecke QR ist die y -Koordinate von Q mit dem Vorzeichen der Halbebene, in der Q liegt. Damit sind die Punkte der (X, Y) -Ebene den Paaren reeller Zahlen (x, y) ein-eindeutig zugeordnet. Den durch die (X, Y) -Ebene getrennten Halbräumen gibt man entgegengesetzte Vorzeichen und kann nun zu irgendeinem Punkte P des Raumes ein reelles Zahlentripel (x, y, z) von Koordinaten bestimmen, indem man das — wieder immer existierende — Lot PQ auf die (X, Y) -Ebene fällt; dieses liefert die z -Koordinate, während die (x, y) -Koordinaten von Q gleichzeitig zu P gehören. Daraus folgt:

In der absoluten Geometrie lassen sich die Punkte des Raumes den Tripeln (x, y, z) reeller Zahlen ein-eindeutig zuordnen.

6. Wegen der Ein-eindeutigkeit dieser Zuordnung bilden die Punkte der absoluten Geometrie ein vollständiges System, da ja jeder neu hinzudenkbare Punkt ein Zahlentripel beanspruchen würde, das schon vergeben ist. Dann folgt aber aus der Schlußweise von Nr. 3 ohne weiteres, daß auch die Geraden und Ebenen der absoluten Geometrie vollständige Systeme bilden. Demnach gilt der dem Vollständigkeitsaxiom nachgebildete

*Vollständigkeitssatz*¹¹⁾. *Es ist nicht möglich, zu einer Deutung der Axiome der absoluten Geometrie (Nr. 4, I—IV) irgendwelche Dinge („Punkte“, „Gerade“, „Ebenen“) so hinzuzudenken, daß in der erweiterten Deutung wieder die Axiome der absoluten Geometrie gelten und daß die axiomatischen Beziehungen der Elemente (Punkte, Gerade, Ebenen) der ursprünglichen Deutung bei der Erweiterung erhalten bleiben.*

¹⁰⁾ Weiterhin kurz als „ein-eindeutig“ bezeichnet.

¹¹⁾ „N. G.“, S. 51 ist der Vollständigkeitsatz für die absolute Geometrie der Ebene abgeleitet.

7. Wir hatten die absolute Geometrie gewonnen, indem wir das Parallelenaxiom und das Vollständigkeitsaxiom aus dem Hilbertschen System der Euklidischen Geometrie wegließen und das Cantorsche Axiom hinzufügten. Denkt man sich den soeben bewiesenen Vollständigkeitsatz der absoluten Geometrie wieder als Axiom eingeführt, dann folgt daraus in bekannter Weise die Aussage des Cantorschen Axioms und es ergibt sich damit, daß man die absolute Geometrie auch gewinnen kann, indem man aus dem Hilbertschen Axiomensystem nur das Parallelenaxiom wegläßt und, ohne Einführung des Cantorschen Axioms, das Vollständigkeitsaxiom beibehält. Da man aus der absoluten Geometrie, wie gesagt, die Euklidische durch Anfügung des Parallelenaxioms erhält, ergibt sich daraus:

Das Vollständigkeitsaxiom hat in Hilberts Axiomensystem der Euklidischen Geometrie nichts mit dem Parallelenaxiom zu tun, es braucht nicht den Schlußstein des Axiomensystems zu bilden, sondern kann vor das Parallelenaxiom und hinter die übrigen Axiome gesetzt werden¹²⁾.

8. Dieses Verschieben des Vollständigkeitsaxioms ist gleichbedeutend mit einer Verschärfung der Hilbertschen Fassung, da aus dem Vordersatze das Parallelenaxiom weggelassen wird.

Im Anschluß an die hier vorliegenden Ausführungen macht mich Herr P. Bernays auf Grund einer Vorlesungsbemerkung von Herrn D. Hilbert darauf aufmerksam, daß man in dieser Verschärfung noch weiter gehen kann, *da man im Vordersatze des Vollständigkeitsaxioms nur die Axiome der Gruppen I, II, dazu die linearen Kongruenzaxiome III 1—3 und das Archimedische Axiom aufzunehmen braucht.* Es genügt somit die Forderung der *linearen Vollständigkeit*, und das Vollständigkeitsaxiom hat auch mit der Winkelgeometrie nichts zu tun. Dies kann man z. B. in folgender Weise erkennen:

Zunächst müssen bei jeder Erweiterung einer Deutung der Axiome I nach Nr. 3 neue *Punkte* auftreten. Nun liege eine Erweiterung einer Deutung der Axiomgruppen I, II vor. *Dann kann man stets eine alte Gerade mit einem neuen Punkte finden:* Denn sind A, B, C, D die vier alten, nicht in einer Ebene liegenden Punkte des Axioms I 8 und ist N ein neuer Punkt, dann liegt er entweder auf einer der sechs Verbindungsgeraden von zweien dieser vier Punkte, womit die gesuchte Gerade gefunden ist, oder er tut dies nicht. In diesem letzten Falle kann man sofort eine alte Ebene mit einem neuen Punkte finden, indem man die Ebene ABN betrachtet. In dieser liegt, wenn sie alte Ebene ist, die gesuchte Ebene vor. Ist sie aber neue Ebene, dann hat sie mit der alten

¹²⁾ Dasselbe gilt natürlich für die hyperbolische Geometrie, wenn man statt des Euklidischen ein hyperbolisches Parallelenaxiom einführt.

Ebene BCD nach I 7 noch einen weiteren — nach I 8 nicht auf der Geraden AB liegenden — Punkt gemeinsam und zwar einen neuen Punkt, da sonst ABN nach I 4 nicht neue Ebene wäre. Daher ist nun in BCD eine alte Ebene mit einem neuen Punkte gefunden.

Um nun eine alte Gerade mit einem neuen Punkte zu finden, betrachten wir die alte Ebene mit dem neuen Punkte. Nach I 3 enthält sie drei alte, nicht in einer Geraden liegende Punkte, sie mögen P, Q, R heißen, X sei ihr neuer Punkt. Liegt X auf einer der drei Geraden durch zwei alte Punkte, dann ist dies die gesuchte Gerade, andernfalls gibt es nach II 2 einen alten Punkt S zwischen Q und R und die Gerade SX ist entweder die gesuchte alte Gerade mit dem neuen Punkt oder sie ist eine neue Gerade. Im letzteren Falle kann sie P nicht enthalten und trifft daher nach dem Axiom von Pasch II 4 eine der alten Geraden PQ oder PR und zwar in einem neuen Punkte, da sonst SX nicht eine neue Gerade wäre. Damit ist in jedem Fall eine alte Gerade mit einem neuen Punkte gefunden.

Nun schließt man analog Nr. 5 und 6: Liegt irgendeine Deutung der Axiomgruppen I, II und der Axiome III 1—3 sowie des Archimedischen und des Cantorsche Axioms vor, dann kann diese nicht erweiterungsfähig sein, denn in einer solchen Erweiterung müßte nach dem soeben Gesagten eine alte Gerade mit einem neuen Punkt auftreten. In dieser könnte man, wie in Nr. 5 in der X -Achse, eine Metrik einführen und die alten Punkte dieser Geraden würden schon die Zahlenreihe erschöpfen.

Daraus ergibt sich zunächst folgender, gegenüber der Fassung von Nr. 6

Verschärfter Vollständigkeitssatz: Es ist nicht möglich, zu einer Deutung der Axiome der Verknüpfung, der Anordnung, der linearen Kongruenz sowie des Archimedischen und des Cantorsche Axioms irgendwelche Dinge („Punkte“, „Gerade“, „Ebenen“) so hinzuzudenken, daß in der erweiterten Deutung wieder die genannten Axiome gelten und daß die durch sie in der ursprünglichen Deutung bestimmten Beziehungen bei der Erweiterung erhalten bleiben.

Weiterhin folgt nun aus einem zu Nr. 7 analogen Schlusse, daß es genügt, im Vordersatze des Vollständigkeitsaxioms nach den Axiomgruppen I und II nur die Axiome III 1—3 und das Archimedische Axiom aufzuführen.

9. Die kongruenten Übertragungen, die durch die Kongruenzaxiome III gewährleistet werden, gestatten vielfach aus dem Verhalten bestimmter Elemente auf das Verhalten aller Elemente zu schließen, wodurch sich, wie der Verfasser gezeigt hat, Verschärfungen Hilbertscher Axiome der

späteren Gruppen ergeben: so genügt es, das Euklidische (wie auch das hyperbolische) Parallelenaxiom für einen bestimmten Punkt und eine bestimmte Gerade auszusprechen (statt für jeden Punkt und jede Gerade¹³) oder das Archimedische Axiom nur für eine bestimmte Halbgerade¹⁴). Ähnlich kann man hier schließen: Tritt in einer Erweiterung einer Deutung auf einer alten Geraden g ein neuer Punkt N auf, ist A ein alter Punkt von g und ist h eine andere alte Gerade mit dem alten Punkte B , dann erhält man durch Abtragung der Strecke AN auf h von B aus — Axiom III 1 — auch auf h einen neuen Punkt. Nach Nr. 8 genügt es daher, durch das Vollständigkeitsaxiom das Auftreten eines neuen Punktes auf einer einzigen alten Geraden auszuschließen.

Das bisher Gesagte gibt zusammengefaßt folgendes gegenüber der Hilbertschen Fassung (Nr. 2)

Verschärftes Vollständigkeitsaxiom: \mathfrak{A} bezeichne die Gesamtheit der Axiome I, II, III 1–4, V 1, g eine (irgendwie) bestimmte Gerade. Die Menge der Punkte von g ist bei Erfüllung der Axiome \mathfrak{A} nicht erweiterungsfähig, d. h. man kann zu einem System von Punkten, Geraden — darunter g —, Ebenen, welche die Axiome \mathfrak{A} erfüllen, nicht irgendwelche Dinge als Punkte — darunter einen neuen Punkt von g —, Gerade, Ebenen derart hinzudenken, daß die durch die Axiome \mathfrak{A} festgelegten Beziehungen im ursprünglichen System bei der Erweiterung erhalten bleiben und daß im erweiterten System wieder die Axiome \mathfrak{A} gelten.

Würde man, was wir vermeiden wollten, Hilberts Reihenfolge der Axiome durch Zerreißen der Axiomgruppen abändern, dann könnte man nach dem soeben Gesagten den Axiomgruppen I, II die linearen Kongruenzaxiome und das Archimedische Axiom folgen lassen und dann das soeben ausgesprochene Vollständigkeitsaxiom anschließen.

10. Aus dem Vollständigkeitsatz von Nr. 6 folgt nun eine bemerkenswerte Tatsache: Jede Deutung des Axiomensystems der absoluten Geometrie ist vollständig im Sinne des Vollständigkeitsaxioms. Führt man in zwei verschiedenen Deutungen der absoluten Geometrie rechtwinkelige Punktkoordinaten nach Nr. 5 ein, dann sind die Punkte dieser Deutungen durch die Zahlentripel ein-eindeutig aufeinander bezogen. Und doch gibt es axiomatisch wesentlich verschiedene Deutungen, wie man sofort erkennt, indem man das Axiomensystem der absoluten Geometrie einmal im Euklidischen, ein anderes Mal im hyperbolischen Raume deutet. Diese beiden Deutungen sind zwar in den Punkten, aber nicht mit Erhaltung der

¹³) „N.G.“, S. 55 (und 70).

¹⁴) R. Baldus, „Über das Archimedische Axiom“, Math. Zeitschr. 26 (1927), S. 757–761.

axiomatisch festgelegten Beziehungen in den Geraden und Ebenen ein-eindeutig aufeinander beziehbar, was schon aus dem Verhalten der Parallelen folgt¹⁵⁾. Nicht alle Deutungen des Axiomensystems der absoluten Geometrie sind daher (holoëdrisch) isomorph. D. h.:

Ein Axiomensystem kann im Sinne des Vollständigkeitsaxioms vollständig sein, ohne daß dessen sämtliche Deutungen isomorph sind, d. h. unter Erhaltung der axiomatisch festgelegten Begriffe und Beziehungen elementweise ein-eindeutig aufeinander beziehbar¹⁶⁾.

11. Erst nach Fortsetzung des Axiomensystems der absoluten Geometrie in einem Parallelenaxiom gelangt man in der Euklidischen wie in der hyperbolischen Geometrie zu den Gleichungen der Geraden und der Ebene, zur Trigonometrie und zum analytischen Erfassen aller axiomatischen Beziehungen sowie zu lauter isomorphen Deutungen.

Bezeichnet man ein Axiomensystem als *monomorph*¹⁷⁾, wenn es nur

¹⁵⁾ Ausführlich spricht sich dies mit alleiniger Verwendung axiomatisch festgelegter Begriffe und Beziehungen folgendermaßen aus: Sind zwei Deutungen $\{\mathfrak{D}\}$ und $\{\mathfrak{D}'\}$ der absoluten Geometrie aufeinander isomorph abbildbar, dann müssen den Punkten einer Ebene α die Punkte einer Ebene α' entsprechen, einem Punkt P von α entspricht ein Punkt P' von α' , einer Geraden g von α , die P nicht enthält, eine Gerade g' von α' , die P' nicht enthält, einer Geraden h in α durch P eine Gerade h' in α' durch P' , einem Schnittpunkte von h mit g ein Schnittpunkt von h' mit g' . Ist $\{\mathfrak{D}\}$ Euklidisch und wäre $\{\mathfrak{D}'\}$ hyperbolisch, dann gibt es genau eine Gerade durch P in α , welche mit g keinen Punkt gemeinsam hat, während in α' mehrere Gerade durch P' mit g' keinen Punkt gemeinsam hätten, daher können diese beiden Deutungen nicht isomorph sein.

¹⁶⁾ Sind Deutungen eines Axiomensystems vollständig im Sinne Hilberts, d. h. nicht erweiterungsfähig, dann brauchen sie nach dem soeben Gesagten nicht isomorph zu sein; nachdem die Bezeichnung „Vollständigkeitsaxiom“ eingeführt ist, kann man daher die Wahl der Bezeichnung „vollständig“ z. B. bei H. Weyl, „Philosophie der Mathematik und Naturwissenschaft“ (im Handbuch der Philosophie), München und Berlin 1927, S. 162 nicht als sprachlich glücklichen Griff bezeichnen, wenn man auf S. 22 den Satz findet, „Ein Axiomensystem ist vollständig, wenn irgend zwei inhaltliche Interpretationen desselben notwendig isomorph sind“.

¹⁷⁾ Hierfür finden sich in der Literatur auch noch andere Bezeichnungen, z. B. „kategorisch“, „vollständig“, vgl. Anm. ¹⁰⁾.

Ein weiteres Beispiel eines vollständigen, polymorphen Systems mit einer unendlichen Menge von Elementen gab in der Kissinger Diskussion Fräulein E. Noether. Sie stellt mir folgende Ausführungen darüber zur Verfügung: Die algebraisch abgeschlossenen, absolut algebraischen Körper bilden ohne Angabe der Charakteristik ein Beispiel für vollständige, nicht isomorphe Bereiche. Dabei heißt ein Körper absolut algebraisch, wenn jedes seiner Elemente algebraisch ist in bezug auf den darin enthaltenen Primkörper, d. h. den aus dem Einheitselement abgeleiteten Körper. Ein Körper \mathfrak{K} heißt algebraisch abgeschlossen, wenn jedes Element eines beliebigen Erweiterungskörpers, das algebraisch in bezug auf \mathfrak{K} — und folglich in bezug auf den Primkörper — ist, zu \mathfrak{K} gehört; anders ausgedrückt: wenn jedes Polynom mit Ko-

noch isomorphe Deutungen zuläßt¹⁸⁾ (Satz von Nr. 8), dann ist damit gezeigt:

Das Axiomensystem der absoluten Geometrie ist vollständig im Sinne des Vollständigkeitsaxioms, aber nicht monomorph. Erst durch Hinzufügung eines (Euklidischen oder hyperbolischen) Parallelenaxioms wird das Axiomensystem monomorph.

Das ist die wirkliche Bedeutung des Parallelenaxioms.

Während andere Axiome der Euklidischen Geometrie innerhalb einzelner möglicher Deutungen selektiv wirken, den Geltungsbereich der vorangehenden Axiome einschränken, indem sie aus den weitesten Deutungen der vorhergehenden Axiome Teile als unbrauchbar ausscheiden, Deutungen verengern, schränkt das Parallelenaxiom keine Deutungen ein, sondern greift aus ihnen isomorphe heraus und verwirft die übrigen¹⁹⁾.

effizienten aus \mathfrak{K} in Linearfaktoren mit Koeffizienten aus \mathfrak{K} zerfällt. Die oben genannten Körper lassen sich also definieren: Jedes Element ist algebraisch in bezug auf den darin enthaltenen Primkörper und der Körper genügt dem Vollständigkeitsaxiom bezüglich dieser Forderung. Die Existenz dieser Körper ist durch Steinitz bewiesen — bei beliebigem Primkörper. Da der Primkörper aber entweder dem Körper der rationalen Zahlen isomorph sein kann (Charakteristik Null) oder dem Restklassensystem nach einer Primzahl (Charakteristik p), so gibt es unendlich viele nicht isomorphe Interpretationen. Dagegen folgt aus einem weiteren Steinitzschen Satz (Fundamentalsatz der algebraisch abgeschlossenen Erweiterungen), daß bei Vorgabe der Charakteristik die Interpretation bis auf Isomorphie eindeutig ist. (Steinitz, Algebraische Theorie der Körper, Crelle 137.)

¹⁸⁾ Isomorph heißt immer isomorph im Sinne des betreffenden Axiomensystems. So sind z. B. zwei affin aufeinander bezogene Ebenen für den Euklidischen Geometer, der nur die Axiomgruppen I, II eingeführt hat, aufeinander isomorph abgebildet, nach Einführung der Axiome III nicht mehr.

Hier ein Beispiel eines Axiomensystems, das eine beliebige Zahl von Deutungen zuläßt, deren jede endlich viele Elemente enthält, jede ist vollständig im Sinne des Vollständigkeitsaxioms, dabei sind keine zwei dieser Deutungen isomorph: $A_1, A_2, \dots, A_n \dots$ seien die Elemente des Systems. Es gibt eine Beziehung „ A_q folgt auf A_p “ zwischen zwei Elementen, die folgenden Axiomen genügt: I. Folgt A_q auf A_p , dann folgt nicht auch A_p auf A_q . II. Jedes Element folgt genau einem und jedes Element hat genau ein ihm folgendes. Nimmt man einen orientierten Kreis, dann ist eine Gruppe von drei Punkten auf ihm eine mögliche Deutung, ebenso eine Gruppe von vier Punkten, von fünf usw. Man kann keine dieser Deutungen unter Erhaltung aller axiomatischen Beziehungen des „Folgens“ erweitern, weiterhin lassen sich keine zwei der genannten Deutungen so ein-eindeutig aufeinander abbilden, daß bei der Abbildung alle Beziehungen des „Folgens“ erhalten bleiben.

¹⁹⁾ Man kann das Euklidische und das hyperbolische Parallelenaxiom so fassen, daß die beiden Axiome eine vollständige Disjunktion über die Parallelen ausdrücken, vgl. „N. G.“ S. 55 und 70, sowie S. 72–73. Damit zerfällt die Gesamtheit der Deutungen der absoluten Geometrie in die beiden Klassen der Euklidischen und der hyperbolischen Deutungen.

12. Polymorphe und monomorphe Axiomensysteme können sich in wesentlichen Punkten unterscheiden, z. B. das absolute und das Euklidische oder hyperbolische Axiomensystem in folgender Weise in der *Entscheidbarkeit*:

Die absolute Geometrie ist polymorph, sie läßt axiomatisch wesentlich verschiedene Deutungen zu. Es gibt in ihr Fragen, die *sicher nicht entscheidbar sind*, z. B. die nach der Zahl der Parallelen in einer Ebene zu einer Geraden durch einen Punkt, die nach der Winkelsumme des Dreiecks, überhaupt jede einem Parallelenaxiom äquivalente Frage. Die Axiomensysteme der Euklidischen oder hyperbolischen Geometrie dagegen sind monomorph, jede ihrer Deutungen kann außerdem — was für die nun folgende Bemerkung wesentlich ist — in analytischer Form gefaßt werden: daher *fällt hier die grundsätzliche geometrische Entscheidbarkeit mit der analytischen Entscheidbarkeit zusammen* und man wird die eine gleichzeitig mit der anderen bejahen oder bestreiten, je nachdem man Formalist oder Intuitionist ist²⁰).

Ein zweiter Unterschied bezieht sich auf das *Operieren mit speziellen Deutungen*: Bei einem monomorphen System sind alle Deutungen isomorph. Solange es sich nur um Begriffe und Beziehungen handelt, die in dem betreffenden Axiomensystem gefaßt sind, kann man daher aus einer speziellen Deutung *auch mit Mitteln, die aus anderen Axiomensystemen stammen*, schließen, wie man es z. B. bei der Deutung der hyperbolischen Geometrie innerhalb der Euklidischen Einheitskugel tut, wenn man projektive oder Euklidische Sätze dabei in die hyperbolische Sprache übersetzt²¹). *In einer polymorphen Geometrie wie der absoluten wäre das unzulässig*, in ihr dürfte man beispielsweise aus der Deutung im Euklidischen Raume nicht auf die Winkelsumme im Dreiecke schließen.

Hiermit wollen wir die Betrachtungen über monomorphe und polymorphe Axiomensysteme verlassen und zum Vollständigkeitsaxiom zurückkehren.

13. Wer das Hilbertsche Vollständigkeitsaxiom kennenlernt, fühlt sich angesichts der Allgemeinheit seiner Aussage unsicher, hat die Empfindung, daß es sich dem logischen Zugriff entzieht, in seiner Tragweite

²⁰) Populär dargestellt in des Verfassers Schrift „Formalismus und Intuitionismus in der Mathematik“, Karlsruhe 1924, 45 S.

Die soeben durchgeführte Betrachtung soll selbstverständlich keine allgemeine Aussage über polymorphe und monomorphe Axiomensysteme liefern, sondern bezieht sich auf die speziellen Verhältnisse der absoluten und der Euklidischen Geometrie; diese sind zwar stark durch den polymorphen oder monomorphen Charakter, aber nicht allein durch ihn bestimmt.

²¹) Vgl. „N. G.“, S. 144.

nicht übersehen läßt. Hieraus entspringt auch die in Nr. 2 erwähnte Zurückhaltung in seiner Behandlung. Geht man den logischen Gründen für dieses Gefühl nach, dann entdeckt man, daß sich von zwei Seiten her ernstliche axiomatische Einwände gegen das Vollständigkeitsaxiom erheben lassen:

Zunächst haben nach den „Grundlagen“ S. 2 die Axiome der Geometrie die Aufgabe, die Beziehungen wie „zwischen“, „parallel“, „kongruent“ usw. zwischen den gedachten Dingen, den Punkten, Geraden und Ebenen zu beschreiben. Das ist doch wohl so zu verstehen, daß die Axiome von den gedachten Dingen *und nur von diesen* handeln, so daß man z. B. bei der Untersuchung der Brauchbarkeit einer Deutung lediglich die bei der Deutung gedachten Dinge und nicht auch andere Dinge auf ihre formallogische Verträglichkeit mit den Axiomen hin zu prüfen hat. Demgegenüber fordert das Vollständigkeitsaxiom, daß man zu den gedachten Dingen keine anderen Dinge mehr hinzudenken kann, *es enthält demnach eine Aussage nicht nur über die gedachten Dinge, sondern über alle überhaupt denkbaren Dinge*. Um das Vollständigkeitsaxiom als Axiom erhalten zu können, müßte man daher *in den Axiomen auch Aussagen über andere als die in der betreffenden Deutung des Axiomensystems gedachten Dinge zulassen*, was den Axiomenbegriff in der Geometrie in bedenklicher und auch überflüssiger Weise erweitern würde²²⁾.

14. Dazu kommt noch ein anderer, systematisch-axiomatischer Nachteil: Die Aussage des Vollständigkeitsaxioms „Wenn die Axiome A, B, C, \dots, N erfüllt sind, dann ...“ setzt die Geltung aller dieser genannten Axiome A, B, C, \dots, N voraus. Wollte man nun prüfen, ob eines dieser Axiome, etwa B , von den übrigen Axiomen einschließlich des Vollständigkeitsaxioms unabhängig, das soll heißen, aus ihnen nicht beweisbar ist, dann müßte man in bekannter Weise das Axiomensystem $A, \text{non } B, C, \dots, N$ untersuchen, dazu das Vollständigkeitsaxiom „Wenn die Axiome $A, \text{non } B, C, \dots, N$ erfüllt sind, dann ...“. Das ist aber nun nicht mehr das frühere Vollständigkeitsaxiom, weil ja sein Vordersatz anders lautet, er enthält „non B “ statt „ B “. Daher kann man in der bisher einzig bekannten Weise nur zeigen, daß im Vordersatze des Vollständigkeitsaxioms nicht genannte

²²⁾ In Kissingen hat Herr A. Fraenkel in einer Besprechung darauf hingewiesen, daß die Mengenlehre nicht anders monomorph gemacht werden kann, als durch ein „Postulat“ im Sinne Geigers, nämlich durch ein Beschränktheitsaxiom, gegen das ähnliche Einwände vorgebracht werden können, wie gegen das Vollständigkeitsaxiom. Solange dies nicht vermieden werden kann, ähnlich wie es in der Geometrie durch Hinzunahme des Cantorsche Axioms geschieht, ist die Geometrie in gewissem Sinn axiomatisch einfacher als die Mengenlehre, da ihre Axiome nur von den an der Spitze des Axiomensystems eingeführten Dingen zu handeln brauchen.

Axiome von ihm unabhängig sind, z. B. nach unserer Nr. 7 das Parallelenaxiom, *dagegen kann man von keinem im Vollständigkeitsaxiom genannten Axiom zeigen, daß es nicht aus ihm und den übrigen Axiomen folgt*²³⁾.

15. Daß hier eine wirkliche Gefahr vorliegt, möge an zwei Beispielen gezeigt werden: Das Axiom I 8 lautet: „Es gibt wenigstens vier nicht in einer Ebene gelegene Punkte“. Angenommen, es erfüllte eine Deutung der Euklidischen Geometrie dieses Axiom nicht, dann könnte man die dabei gefundene ebene Geometrie durch Hinzudenken des übrigen Euklidischen Raumes erweitern, die Deutung wäre daher nicht vollständig, d. h. *das Axiom I 8 folgt aus dem Vollständigkeitsaxiom*²⁴⁾.

Auch der Zusammenhang zwischen dem Archimedischen und dem Vollständigkeitsaxiom ist beachtenswert: In den „Grundlagen“ S. 22/23 findet sich der Satz: „Die Erfüllbarkeit des Vollständigkeitsaxioms ist wesentlich durch die Voranstellung des Archimedischen Axioms bedingt; in der Tat läßt sich zeigen, daß zu einem System von Punkten, Geraden und Ebenen, welche die Axiome I—IV²⁵⁾ erfüllen, stets noch auf mannigfache Weise solche Elemente hinzugefügt werden können, daß in dem durch Zusammensetzung entstehenden System die Axiome I—IV ebenfalls sämtlich gültig sind; d. h. das Vollständigkeitsaxiom würde einen Widerspruch einschließen, wenn man den Axiomen I—IV nicht noch das Archimedische

²³⁾ Herr Finsler weist — und dagegen ist logisch nichts einzuwenden — a. a. O. S. 699 eine enger gefaßte Unabhängigkeit seiner Axiome nach, nämlich die Unbeweisbarkeit jedes Axioms aus den vorhergehenden. Der Gefahr der Einführung überzähliger Axiome ist damit aber nicht begegnet, da ja auch ein vorhergehendes Axiom aus einem späteren beweisbar sein kann.

²⁴⁾ Ähnliche Überlegungen gelten für andere Existenzialaxiome in Verbindung mit einem Vollständigkeitsaxiom, z. B. für das Hilbertsche Axiom I 3.

Während so einerseits I 8 aus dem Vollständigkeitsaxiom folgt, kann man andererseits, wie Herr P. Bernays bemerkt, ohne in Widerspruch mit dem Vollständigkeitsaxiom zu geraten, an Stelle von I 8 ein gegenteiliges Axiom einführen. „Es gibt keine vier Punkte, die nicht komplanar sind.“

Auch auf einen andern Umstand sei in diesem Zusammenhange hingewiesen: Die Hinzufügung eines Vollständigkeitsaxioms zu einem Axiomensystem hat nur dann einen Sinn, wenn es unter allen möglichen Deutungen der bisherigen Axiome nicht mehr erweiterungsfähige gibt. Würde man z. B. das Axiom I 7, welches die Dreidimensionalität des Raumes ausspricht, weglassen, alle übrigen Axiome beibehalten und als neue Forderung hinzufügen, das System der Punkte solle endlich viele Dimensionen haben (ohne Angabe einer bestimmten Dimensionszahl), dann wäre jede Deutung dieses Axiomensystems durch Einbettung in einen höheren Raum erweiterungsfähig und es gäbe daher keine Deutung für das aus diesen Axiomen und dem Vollständigkeitsaxiom gebildete Axiomensystem. Auf einen analogen Fall bezieht sich der Einwand gegen Herrn Finslers a. a. O. aufgestelltes Axiomensystem bei R. Baer, „Über ein Vollständigkeitsaxiom in der Mengenlehre“, Math. Zeitschr. 27 (1928), S. 536—539.

²⁵⁾ Verknüpfung, Anordnung, Kongruenz, Euklidisches Parallelenaxiom.

Axiom hinzufügt.“ Das scheint zunächst nichts anderes zu bedeuten, als daß durch einen indirekten Beweis gezeigt wird, daß aus dem Vollständigkeitsaxiom und den übrigen Axiomen das Archimedische Axiom folgt. Tatsächlich ist aber noch folgende Auslegung möglich: Fügt man zu den Axiomen I—IV kein Axiom über die Vergleichbarkeit von Strecken hinzu, dann ist jede Deutung erweiterungsfähig. Fügt man das Archimedische Axiom hinzu, dann gibt es vollständige Deutungen. Dabei bleibt noch die Möglichkeit bestehen, daß auch bei Fortsetzung der Axiome I—IV in einem Nichtarchimedischen Axiom vollständige Deutungen denkbar sind²⁶). Hier wäre die (nach Nr. 14 unmögliche) Feststellung der Unabhängigkeit des Archimedischen Axioms von allen anderen Axiomen einschließlich des Vollständigkeitsaxioms besonders wertvoll.

16. Gegen das Vollständigkeitsaxiom als Axiom, dies ist unser Endergebnis, bestehen ernste Bedenken. Aber seine in genialer Allgemeinheit formulierte Aussage liefert in der absoluten und damit in der Euklidischen wie der hyperbolischen Geometrie den Vollständigkeitsatz von Nr. 6, sowie dessen Verschärfung von Nr. 8. Dieser Satz stellt ein Unikum der Beweistechnik dar, denn bei seinem Beweise wird nur von den an der Spitze des Axiomensystems eingeführten Dingen ausgegangen, werden ausschließlich Axiome benützt, die lediglich von diesen gedachten Dingen handeln, *und doch gelangt man dabei zu einer Aussage über alle überhaupt denkbaren Dinge*. So verdankt man Hilbert im umfassendsten Satze der Geometrie ein einzigartiges Beispiel mathematischer Beweiskraft.

Karlsruhe i. B., im September 1927.

²⁶) Eine andere Auffassung über den logischen Inhalt des hier zitierten Satzes aus den „Grundlagen“ findet man bei P. Finsler, „Erwiderung auf die vorstehende Note des Herrn Baer“, Math. Zeitschr. 27 (1928), S. 540—542 auf S. 542. Ihr tritt Herr R. Baer in seiner unmittelbar darauffolgenden Note „Bemerkungen zu der Erwiderung von Herrn P. Finsler“ entgegen.

(Eingegangen am 27. 9. 1927.)

Die trigonometrische Approximation für eine Klasse von verallgemeinerten fastperiodischen Funktionen.

Von

Robert Schmidt in Kiel.

Einleitung und Definitionen.

Auf der Düsseldorfer Tagung habe ich über eine Klasse von Funktionen vorgetragen, die folgendermaßen erklärt sind:

Definition¹). Eine Funktion $f(x)$, die samt dem Betrage ihres Quadrats in jedem endlichen Intervall im Lebesgueschen Sinne integrierbar ist, soll Wellenfunktion heißen, wenn folgendes zutrifft:

1. Zu jedem Paar $\varepsilon > 0$, $\vartheta > 0$ gibt es ein $L(\varepsilon, \vartheta) > 0$, ein $l(\varepsilon, \vartheta) > 0$ und reelle Zahlen $\tau(\varepsilon, \vartheta)$ derart, daß

$$|f(\xi + \tau) - f(\xi)| \leq \varepsilon$$

wesentlich überall erfüllt ist, d. h. in jedem Intervall $x < \xi < x + L$ der Länge L höchstens bis auf eine Menge $\mathfrak{P}(\varepsilon, \vartheta, \tau, x, L)$ vom Maß

$$|\mathfrak{P}| \leq \vartheta L;$$

und zwar sind die Zahlen τ in jedem Intervall der Länge l mindestens einmal vertreten.

2. Zu jedem $\varepsilon > 0$ gibt es ein $\vartheta(\varepsilon) > 0$ und ein $L(\varepsilon) > 0$, so daß, wenn \mathfrak{Q} eine beliebige Teilmenge eines beliebigen Intervalls $x < \xi < x + L$ vom Maß

$$|\mathfrak{Q}| \leq \vartheta L$$

ist, stets

$$\frac{1}{L} \int_{\mathfrak{Q}} |f(\xi)|^2 d\xi \leq \varepsilon^2$$

erfüllt ist.

¹) Vgl. Jahresbericht der D. M. V. 36, Heft 1/4, S. 2—3. Die dort vorausgesetzte Stetigkeit von $f(x)$ ist im Rahmen der vorliegenden Untersuchung belanglos.

Unter den Wellenfunktionen sind die fastperiodischen Funktionen von Herrn H. Bohr²⁾, ferner die von den Herren W. Stepanoff³⁾ und N. Wiener⁴⁾ angegebenen und untersuchten Funktionenklassen enthalten⁵⁾.

In der vorliegenden Abhandlung werde ich den Problemkreis des Fischer-Rieszschen Satzes, aus dem heraus ich auf die Wellenfunktionen geführt wurde, ganz beiseite lassen. Das Ziel ist vielmehr die Aufstellung und der Beweis eines *Analogons* des Satzes von Herrn Bohr über die *gleichmäßige Approximierbarkeit der fastperiodischen Funktionen durch trigonometrische Polynome*⁶⁾. Mit diesem Analogon ist die Analyse der Wellenfunktionen in genau dem Umfange geleistet, wie es in der Theorie der fastperiodischen Funktionen durch den Bohrschen Approximationssatz der Fall ist.

Was die Formulierung eines solchen Analogons anbetrifft, so wird es sich zunächst darum handeln, dem Begriff der gleichmäßigen Konvergenz einer Funktionenfolge in passender, d. h. der Natur der Wellenfunktionen angemessener Weise einen neuen Begriff zur Seite zu stellen. Die „im Mittel scharfe Konvergenz“ einer Funktionenfolge erweist sich als geeignet:

²⁾ H. Bohr, Zur Theorie der fastperiodischen Funktionen I, II, III, *Acta Mathematica* **45**, S. 29—121; **46**, S. 101—214; **47**, S. 237—281 (1924—26), und eine Reihe weiterer Abhandlungen.

³⁾ W. Stepanoff, Über einige Verallgemeinerungen der fastperiodischen Funktionen, *Math. Annalen* **95** (1926), S. 473—498.

⁴⁾ N. Wiener, On the representation of functions by trigonometrical integrals, *Math. Zeitschr.* **24** (1925), S. 575—616.

⁵⁾ Über den Bereich der Wellenfunktionen hinaus gehen unter Umständen diejenigen Funktionen, die Herr O. Toeplitz: Ein Beispiel zur Theorie der fastperiodischen Funktionen, *Math. Annalen* **98** (1927), S. 281—295, als „Belegungen vom Typus A“ bezeichnet. Diese erfüllen unmittelbar die Forderung 1 der Wellenfunktionen, die Forderung 2 jedoch dann und nur dann, wenn die dort auftretenden Integrale

$\int_0^1 |g_n(s)|^2 ds$ beschränkt sind, während Herr Toeplitz allgemeiner — hierin ist eine Bemerkung, die Herr Toeplitz auf Grund einer mündlichen Mitteilung von mir am Schlusse seiner Arbeit angefügt hat, zu berichtigen — solche Belegungen vom Typus A

betrachtet, für die nur $\sum_{n=1}^{\infty} \frac{1}{2^n} \int_0^1 |g_n(s)|^2 ds$ konvergiert. Dies ergibt sich daraus, daß für jede Wellenfunktion der Grenzwert $\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{x-T}^{x+T} |f(\xi)|^2 d\xi$ gleichmäßig in x existiert

(§ 4), während Herr Toeplitz von seinen Belegungen nur das Vorhandensein von

$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} |f(\xi)|^2 d\xi$ fordert.

⁶⁾ *Acta Mathematica* **46**, S. 184.

Definition. Die Funktionen $f_1(x), f_2(x), \dots$ seien samt den Beträgen ihrer Quadrate in jedem endlichen Intervall im Lebesgueschen Sinne integrierbar. Eine solche Funktionenfolge soll im Mittel scharf konvergent heißen, wenn es zu jedem $\varepsilon > 0$ ein $L(\varepsilon) > 0$ und einen Index $n(\varepsilon)$ gibt derart, daß

$$\frac{1}{L} \int_x^{x+L} |f_q(\xi) - f_p(\xi)|^2 d\xi \leq \varepsilon^2$$

erfüllt ist für alle Indexpaare $p \geq n, q \geq n$ und alle x .

Es ist klar, was unter „scharfer Konvergenz im Mittel gegen eine Grenzfunktion“ zu verstehen ist. — Eine Funktionenfolge, die im Mittel scharf gegen eine Grenzfunktion konvergiert, ist offenbar im Mittel scharf konvergent.

Der Begriff der scharfen Konvergenz im Mittel ist *nicht weiter* als der in der Theorie der fastperiodischen Funktionen gebräuchliche Begriff der Konvergenz im Mittel, bei dem nur

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} |f_q(\xi) - f_p(\xi)|^2 d\xi \leq \varepsilon^2$$

für alle Indexpaare $p \geq n, q \geq n$ gefordert wird; daß jener Begriff wirklich *enger* ist als dieser, lehrt die Funktionenfolge

$$f_n(x) = \begin{cases} 1 & \text{für } -n \leq x \leq +n \\ 0 & \text{sonst,} \end{cases}$$

oder die Folge von fastperiodischen (stetigen reinperiodischen) Funktionen

$$f_n(x) = \begin{cases} \sin \frac{\pi x}{n} & \text{für } kn^2 \leq x \leq kn^2 + n \\ 0 & \text{sonst} \end{cases}$$

($k = 0, \pm 1, \pm 2, \dots$).

Beide Funktionenfolgen konvergieren im Mittel gegen $f(x) \equiv 0$, sie sind aber nicht im Mittel scharf konvergent.

Wenn andererseits die Funktionen $f_1(x), f_2(x), \dots$ reinperiodisch mit gemeinsamer Periode ω und im Mittel konvergent sind, dann sind sie auch im Mittel scharf konvergent ($L = \omega$). Die scharfe Konvergenz im Mittel ist also eine Verallgemeinerung der Konvergenz im Mittel für ein endliches Intervall, jedoch eine nicht so weitgehende wie die Konvergenz im Mittel im Sinne der Bohrschen Theorie.

Trotz ihres Charakters als *Verallgemeinerung der Konvergenz im Mittel* für ein endliches Intervall ist, wie sich zeigen wird, die scharfe Konver-

genz im Mittel als *Analogon der gleichmäßigen Konvergenz* anzusehen; es gilt nämlich — und das ist das Ziel dieser Untersuchung — der folgende *Approximationssatz*:

Zu jeder Wellenfunktion

$$f(x) \sim \sum_{\nu=1}^{\infty} a_{\nu} e^{i\lambda_{\nu}x}$$

gibt es eine Folge von trigonometrischen Polynomen der Form

$$Q_n(x) = \sum_{\nu=1}^{k_n} b_{\nu}^{(n)} e^{i\lambda_{\nu}x},$$

die im Mittel scharf gegen $f(x)$ konvergiert. D. h. also, zu jedem $\varepsilon > 0$ gibt es ein $L(\varepsilon) > 0$ und einen Index $n(\varepsilon)$ derart, daß

$$\frac{1}{L} \int_x^{x+L} |Q_{\nu}(\xi) - f(\xi)|^2 d\xi \leq \varepsilon$$

für alle $\nu \geq n$ und alle x erfüllt ist.

Der erste Abschnitt bringt die Existenz der erforderlichen Integralmittelwerte, insbesondere die Existenz der Fourierkoeffizienten der Wellenfunktionen. Hierbei stellt sich die Identität der Wellenfunktionen mit einer Funktionenklasse heraus, die sich aus der wichtigen Untersuchung von Herrn H. Weyl⁷⁾ zur Theorie der fastperiodischen Funktionen als die allgemeinste ergibt, auf die sich die Methode von Weyl direkt anwenden läßt. Diese Feststellung gestattet es, den Beweis der Parsevalschen Gleichung für Wellenfunktionen durch den Hinweis auf die Entwicklungen von Weyl zu erledigen.

Im zweiten Abschnitt beweise ich den oben formulierten Approximationssatz. — Es darf wohl schon hier kurz beleuchtet werden, warum sich der Beweis des Approximationssatzes wesentlich von den bekannten Beweisen analoger Sätze unterscheidet. Der Grund liegt darin, daß die Definition der scharfen Konvergenz im Mittel nicht an einen Entfernungsbegriff im Funktionenraume anknüpft. Während man zum Beweise von Approximationssätzen hinsichtlich gleichmäßiger Konvergenz oder Konvergenz im Mittel nur *eine* approximierende Funktion (Polynom) nachzuweisen hat, deren Entfernung von der zugrunde gelegten Funktion vorgegeben klein ist, kann eine solche Vereinfachung bei Approximationssätzen hinsichtlich scharfer Konvergenz im Mittel der Natur der Sache nach nicht eintreten.

⁷⁾ H. Weyl, Integralgleichungen und fastperiodische Funktionen, *Math. Annalen* 97 (1927), S. 338—356.

I. Abschnitt.

§ 1.

Vorbereitende Betrachtungen über die Definition der Wellenfunktionen.

Der Zweck der Forderung 2 in der Definition der Wellenfunktionen wird im folgenden der sein, an solchen Stellen ξ abschätzend einzugreifen, wo die Forderung 1 keine Ansprüche an die Differenzen $f(\xi + \tau) - f(\xi)$ stellt, also in den Mengen \mathfrak{P} . Daß die Forderung 2 diesen Zweck zu erfüllen wirklich geeignet ist, ist nicht ohne weiteres erkennbar, da das $L(\varepsilon)$ aus 2 zunächst in keiner Beziehung zum $L(\varepsilon, \vartheta)$ aus 1 steht. Tatsächlich darf jedoch angenommen werden, daß die Längen $L(\varepsilon)$ und $L(\varepsilon, \vartheta)$ übereinstimmen; genauer:

Zu jedem $\varepsilon > 0$ gibt es ein $\vartheta(\varepsilon) > 0$ und ein $L(\varepsilon) > 0$ derart, daß L sowohl als $L(\varepsilon)$ in 2 als auch — mit diesem ϑ — als $L(\varepsilon, \vartheta)$ in 1 geeignet ist.

Das ergibt sich sofort daraus, daß man die Forderungen 1 und 2 ohne sachliche Änderung durch die folgenden Forderungen 1a und 2a ersetzen kann:

1a. *Zu jedem Paar $\varepsilon > 0$, $\vartheta > 0$ gibt es ein $L_1(\varepsilon, \vartheta) > 0$, ein $l(\varepsilon, \vartheta) > 0$ und reelle Zahlen $\tau(\varepsilon, \vartheta)$ derart, daß*

$$|f(\xi + \tau) - f(\xi)| \leq \varepsilon$$

wesentlich überall erfüllt ist, d. h. in jedem Intervall $x < \xi < x + L$ einer beliebigen Länge $L \geq L_1$ höchstens bis auf eine Menge $\mathfrak{P}(\varepsilon, \vartheta, \tau, x, L)$ vom Maß

$$|\mathfrak{P}| \leq \vartheta L;$$

und zwar sind die Zahlen τ in jedem Intervall der Länge l mindestens einmal vertreten.

2a. *Zu jedem $\varepsilon > 0$ gibt es ein $\vartheta(\varepsilon) > 0$ und ein $L_2(\varepsilon) > 0$, so daß, wenn \mathfrak{D} eine beliebige Teilmenge eines Intervalls $x < \xi < x + L$ der beliebigen Länge $L \geq L_2$ vom Maß*

$$|\mathfrak{D}| \leq \vartheta L$$

ist, stets

$$\frac{1}{L} \int_{\mathfrak{D}} |f(\xi)|^2 d\xi \leq \varepsilon^2$$

erfüllt ist.

Beweis für die Gleichwertigkeit von 1 mit 1a und 2 mit 2a. Wenn die Forderungen 1a und 2a erfüllt sind, dann sind es offenbar auch 1 und 2. — Das Umgekehrte ist, soweit es die Forderungen 1 und 1a

betrifft, fast selbstverständlich. Nimmt man nämlich 1 als erfüllt an, und schreibt dort für den Augenblick $L_0(\varepsilon, \vartheta)$ und $l_0(\varepsilon, \vartheta)$ statt $L(\varepsilon, \vartheta)$ und $l(\varepsilon, \vartheta)$, so sind, behaupte ich,

$$L_1(\varepsilon, \vartheta) = L_0\left(\varepsilon, \frac{\vartheta}{2}\right) \quad \text{und} \quad l(\varepsilon, \vartheta) = l_0\left(\varepsilon, \frac{\vartheta}{2}\right)$$

als L_1 und l in 1a geeignet. Denn ist $L \geq L_1$ beliebig, und die natürliche Zahl n gemäß $nL_0 \leq L < (n+1)L_0$ bestimmt, so gilt für ein beliebiges, als $\tau\left(\varepsilon, \frac{\vartheta}{2}\right)$ in 1 brauchbares τ : Die Ungleichung $|f(\xi + \tau) - f(\xi)| \leq \varepsilon$ ist erfüllt in jedem Intervall $x < \xi < x + (n+1)L_0$ höchstens bis auf eine Menge vom Maß $(n+1)\frac{\vartheta}{2}L_0 = \frac{n+1}{n} \cdot \frac{\vartheta}{2} \cdot nL_0 \leq \vartheta L$, und das um so mehr in jedem Intervall $x < \xi < x + L$.

Es bleibt noch zu zeigen: Wenn die Forderung 2 erfüllt ist, dann ist es auch 2a. Zunächst ergibt sich aus 2, wenn man etwa $\varepsilon = 1$ wählt, für $L = L(1)$ die Beschränktheit von

$$\frac{1}{L} \int_x^{x+L} |f(\xi)|^2 d\xi,$$

daus die Beschränktheit von

$$\int_x^{x+1} |f(\xi)|^2 d\xi,$$

und weiter bei jedem festen $A_0 > 0$ die Existenz einer Zahl $K > 0$ derart, daß

$$(1) \quad \frac{1}{A} \int_x^{x+A} |f(\xi)|^2 d\xi \leq K^2$$

ist für alle $A \geq A_0$ und alle x . Endlich ergibt die Schwarzsche Ungleichung

$$(2) \quad \frac{1}{A} \int_x^{x+A} |f(\xi)| d\xi \leq K$$

für alle $A \geq A_0$ und alle x .

In 2 werde für den Augenblick $\vartheta_0(\varepsilon)$ und $L_0(\varepsilon)$ statt $\vartheta(\varepsilon)$ und $L(\varepsilon)$ geschrieben, und K sei so bestimmt, daß (1) für $A_0 = L_0$ erfüllt ist. Ich behaupte nun: Wenn die natürliche Zahl m_0 so gewählt wird, daß

$$\frac{K^2}{m_0} \leq \frac{2\varepsilon^2}{3}$$

ist, dann sind $\vartheta(\varepsilon) = \vartheta_0\left(\frac{\varepsilon}{3}\right)$ und $L_2(\varepsilon) = m_0 L_0\left(\frac{\varepsilon}{3}\right)$ als ϑ und L_2 in 2a geeignet. — Zum Beweise sei das Intervall $J = (x, x + L)$ mit $L \geq L_2$ be-

liebig gewählt, und Ω bezeichne eine beliebige Teilmenge dieses Intervalls vom Maß $|\Omega| \leq \vartheta L$. Es sei ferner

$$mL_0 \leq L < (m+1)L_0 \quad \left(L_0 = L_0 \left(\frac{\varepsilon}{3} \right) \right),$$

so daß also $m \geq m_0$ ist. Ich betrachte das Intervall

$$J_0 = (x, x + mL_0)$$

und seine Teilintervalle

$$J_1 = (x, x + L_0), J_2 = (x + L_0, x + 2L_0), \dots, J_m = (x + (m-1)L_0, x + mL_0).$$

Es bezeichne Ω_0 den Durchschnitt von J_0 und Ω ; es ist

$$|\Omega_0| \leq |\Omega| \leq \vartheta L \leq 2\vartheta mL_0.$$

Ferner bezeichne Ω_μ den Durchschnitt von J_μ und Ω ($\mu = 1, 2, \dots, m$). Die Maße $|\Omega_\mu|$ sollen in der Form

$$(3) \quad |\Omega_\mu| = \vartheta_\mu \cdot L_0 \quad (\mu = 1, 2, \dots, m)$$

geschrieben werden. Es bedeute allgemein n_ν die Anzahl derjenigen Intervalle J_μ ($\mu = 1, \dots, m$), für die

$$(4) \quad \nu \vartheta \leq \vartheta_\mu < (\nu + 1) \vartheta \quad (\nu = 0, 1, 2, \dots)$$

ist. Von einem gewissen ν , etwa $\nu = k + 1$ an sind alle diese Anzahlen $= 0$, und es ist dann

$$n_0 + n_1 + \dots + n_k = m.$$

Ferner ist wegen (3) und (4)

$$n_1 \cdot \vartheta L_0 + n_2 \cdot 2\vartheta L_0 + \dots + n_k \cdot k\vartheta L_0 \leq |\Omega_1| + \dots + |\Omega_m| = |\Omega_0| \leq 2\vartheta mL_0,$$

also

$$\sum_{\nu=1}^k \nu n_\nu \leq 2m.$$

Andererseits kann man, wenn ein bestimmtes μ ins Auge gefaßt wird und

$$\varkappa \vartheta \leq \vartheta_\mu < (\varkappa + 1) \vartheta$$

ist, Ω_μ in $\varkappa + 1$ Teilmengen

$$\Omega_\mu = \Omega_{\mu 1} + \Omega_{\mu 2} + \dots + \Omega_{\mu, \varkappa+1}$$

zerlegen derart, daß

$$|\Omega_{\mu 1}| \leq \vartheta L_0, \dots, |\Omega_{\mu, \varkappa+1}| \leq \vartheta L_0$$

erfüllt ist. Dann wird, da jede der Mengen $\Omega_{\mu 1}, \dots, \Omega_{\mu, \varkappa+1}$ im Intervall J_μ als Menge Ω in 2 geeignet ist,

$$\frac{1}{L_0} \int_{\Omega_\mu} |f(\xi)|^2 d\xi \leq \frac{1}{L_0} \int_{\Omega_{\mu 1}} + \dots + \frac{1}{L_0} \int_{\Omega_{\mu, \varkappa+1}} \leq (\varkappa + 1) \frac{\varepsilon^2}{9}.$$

Nunmehr erhält man

$$\begin{aligned} \frac{1}{mL_0} \int_{\xi_0} |f(\xi)|^2 d\xi &= \frac{1}{m} \left(\frac{1}{L_0} \int_{\xi_1} + \dots + \frac{1}{L_0} \int_{\xi_m} \right) \leq \frac{1}{m} \sum_{\nu=0}^k n_\nu (\nu+1) \frac{\varepsilon^2}{9} \\ &= \frac{1}{m} \cdot \frac{\varepsilon^2}{9} \left\{ \sum_{\nu=1}^k \nu n_\nu + \sum_{\nu=0}^k n_\nu \right\} \leq \frac{1}{m} \cdot \frac{\varepsilon^2}{9} (2m + m) = \frac{\varepsilon^2}{3}; \end{aligned}$$

schließlich

$$\frac{1}{L} \int_{\xi} |f(\xi)|^2 d\xi \leq \frac{mL_0}{L} \cdot \frac{1}{mL_0} \int_{\xi_0} + \frac{L_0}{L} \cdot \frac{1}{L_0} \int_{x+mL_0}^{x+(m+1)L_0},$$

daraus wegen $\frac{mL_0}{L} \leq 1$, $\frac{L_0}{L} \leq \frac{L_0}{L_1} = \frac{1}{m_0}$:

$$\frac{1}{L} \int_{\xi} |f(\xi)|^2 d\xi \leq \frac{\varepsilon^2}{3} + \frac{K^2}{m_0} \leq \varepsilon^2.$$

§ 2.

Übereinstimmung der Wellenfunktionen mit einer Funktionenklasse von Weyl.

Nach diesen Vorbereitungen beweise ich nunmehr die Identität der Klasse der Wellenfunktionen mit der Funktionenklasse, die von Herrn Weyl in der zitierten Arbeit eingeführt wurde, und zwar zuerst in der Form der Definition, die Herr Weyl mit (V) bezeichnet⁸⁾:

Satz 1. *Eine Funktion $f(x)$, die samt dem Betrage ihres Quadrats in jedem endlichen Intervall integrierbar ist, ist dann und nur dann eine Wellenfunktion, wenn es zu jedem $\varepsilon > 0$ ein $L_0(\varepsilon) > 0$, ein $l(\varepsilon) > 0$ und reelle Zahlen $\tau(\varepsilon)$ gibt derart, daß*

$$(5) \quad \frac{1}{L} \int_x^{x+L} |f(\xi + \tau) - f(\xi)|^2 d\xi \leq \varepsilon^2$$

für alle $L \geq L_0$ und alle x erfüllt ist; und zwar sind die Zahlen τ in jedem Intervall der Länge l mindestens einmal vertreten.

1. $f(x)$ sei eine Wellenfunktion. Zu gegebenem $\varepsilon > 0$ können dann $L_0 > 0$ und $\vartheta > 0$ so bestimmt werden, daß

$$(6) \quad \frac{1}{L} \int_{\xi} |f(\xi)|^2 d\xi \leq \frac{\varepsilon^2}{9}$$

⁸⁾ Math. Annalen 97 (1927), S. 354.

im Sinne der Forderung 2a erfüllt ist. Nachdem insbesondere ϑ so fixiert ist, lassen sich l und L_1 so wählen, daß

$$(7) \quad |f(\xi + \tau) - f(\xi)| \leq \frac{\varepsilon}{3}$$

im Sinne der Forderung 1a erfüllt ist. Wenn L_0 die größere der Zahlen L_1 und L_2 ist, so treffen (6) und (7) im angegebenen Sinne jedenfalls für alle $L \geq L_0$ zu. Wegen (7) ist

$$\frac{1}{L} \int_x^{x+L} |f(\xi + \tau) - f(\xi)|^2 d\xi \leq \frac{\varepsilon^2}{9} + \frac{1}{L} \int_{\mathfrak{B}} |f(\xi + \tau) - f(\xi)|^2 d\xi,$$

wo $|\mathfrak{B}| \leq \vartheta L$, so daß also \mathfrak{B} als \mathfrak{Q} in (6) genommen werden kann:

$$\frac{1}{L} \int_{\mathfrak{B}} |f(\xi)|^2 d\xi \leq \frac{\varepsilon^2}{9},$$

ferner

$$\frac{1}{L} \int_{\mathfrak{B}} |f(\xi + \tau)|^2 d\xi \leq \frac{\varepsilon^2}{9},$$

daher

$$\frac{1}{L} \int_x^{x+L} |f(\xi + \tau) - f(\xi)|^2 d\xi \leq \frac{\varepsilon^2}{9} + 2 \frac{\varepsilon^2}{9} + 2 \frac{\varepsilon^2}{9} < \varepsilon^2.$$

2. $f(x)$ möge die Eigenschaft (5) haben. Zu gegebenem Paar $\varepsilon > 0$, $\vartheta > 0$ kann man dann L und l so bestimmen, daß

$$(8) \quad \frac{1}{L} \int_x^{x+L} |f(\xi + \tau) - f(\xi)|^2 d\xi \leq \varepsilon^2 \cdot \vartheta$$

erfüllt ist für alle x , und zwar für Zahlen τ , die in jedem Intervall der Länge l mindestens einmal vertreten sind. Mit diesem $L = L(\varepsilon, \vartheta)$ und diesem $l = l(\varepsilon, \vartheta)$ genügt $f(x)$ der Forderung 1. Denn wenn \mathfrak{B} die Teilmenge des Intervalls $(x, x+L)$ ist, auf der

$$|f(\xi + \tau) - f(\xi)| > \varepsilon$$

ist, so folgt aus (8):

$$\frac{1}{L} \varepsilon^2 |\mathfrak{B}| \leq \frac{1}{L} \int_{\mathfrak{B}} |f(\xi + \tau) - f(\xi)|^2 d\xi \leq \frac{1}{L} \int_x^{x+L} \leq \varepsilon^2 \vartheta,$$

$$|\mathfrak{B}| \leq \vartheta L.$$

Daß $f(x)$ auch der Forderung 2 genügt, folgt so: Zunächst bestimme man L und l so, daß

$$\frac{1}{L} \int_x^{x+L} |f(\xi + \tau) - f(\xi)|^2 d\xi \leq \frac{\varepsilon^2}{4}$$

für alle x und geeignete τ zutrifft. Nachdem x beliebig gewählt ist, nehme man ein τ aus dem Intervall $(-x, -x + l)$, so daß also für $y = x + \tau$ das Intervall $(y, y + L)$ ganz dem von x unabhängigen Intervall $(0, L + l)$ angehört. Es bedeute nun Ω eine vorläufig beliebige meßbare Teilmenge von $(x, x + L)$, und Ω' die zu Ω kongruente Teilmenge von $(y, y + L)$. Dann ist

$$\begin{aligned} \frac{1}{L} \int_{\Omega} |f(\xi)|^2 d\xi &\leq \frac{2}{L} \int_{\Omega} |f(\xi + \tau)|^2 d\xi + \frac{2}{L} \int_x^{x+L} |f(\xi + \tau) - f(\xi)|^2 d\xi \\ &\leq \frac{2}{L} \int_{\Omega'} |f(\eta)|^2 d\eta + \frac{\varepsilon^2}{2}. \end{aligned}$$

Das zuletzt auftretende Integral liegt unterhalb $\frac{\varepsilon^2 L}{4}$, wenn nur $|\Omega|$ unter einer genügend kleinen, von x unabhängigen positiven Schranke gelegen ist. Zu gegebenem $\varepsilon > 0$ läßt sich also in der Tat ein $\vartheta > 0$ so bestimmen, wie es behauptet wird.

Man kann jetzt nach dem Vorgange von Herrn Weyl weiterschließen, also von der Definition (V) zurückgehen auf die ursprüngliche Definition der Weylschen Funktionenklasse (dort S. 16, Nr. 3), dann den Anschluß an die Eigenwerttheorie der linearen Integralgleichungen herstellen und so die Parsevalsche Gleichung gewinnen (dort § 4 und § 6 Schluß). Bei der Durchführung dieses Weges in allen Einzelheiten, insbesondere beim Nachweis der Existenz der Fourierkoeffizienten, stellen sich jedoch Wiederholungen in den Beweisen ein. Aus diesem Grunde erscheint es mir zweckmäßig, von vornherein die Existenz aller in Betracht kommenden Mittelwerte sicherzustellen; dabei wird auch der eben erwähnte direkte Übergang von der einen zur anderen Definition der Weylschen Funktionenklasse eingespart. — Dann ändert sich jedoch die Anordnung des Stoffes bis zur Parsevalschen Gleichung für Wellenfunktionen gegenüber der von Herrn Weyl skizzierten so erheblich, daß es wohl — mindestens aus Gründen der Bequemlichkeit für den Leser — angebracht ist, die Sätze und Beweise mit einiger Ausführlichkeit zu behandeln. Das geschieht in den nächsten beiden Paragraphen. Der entscheidende Teil der Untersuchung von Herrn Weyl, der Beweis der Parsevalschen Gleichung, bleibt natürlich unberührt.

§ 3.

Der Mittelwert $\mathfrak{M}\{f(x)\}$.

Satz 2. Wenn $f(x)$ eine Wellenfunktion ist, dann gibt es zu jedem $\varepsilon > 0$ ein $L_3(\varepsilon) > 0$ derart, daß

$$\left| \frac{1}{L} \int_{x'}^{x'+L} f(\xi) d\xi - \frac{1}{L} \int_{x'}^{x'+L} f(\xi) d\xi \right| \leq \varepsilon$$

erfüllt ist für alle $L \geq L_3$ und alle Paare x', x'' .

Zum Beweise möge das ε des Satzes 1 durch $\frac{\varepsilon}{4}$ ersetzt werden, und es kann und soll dort $l \geq L_0$ angenommen werden. Es folgt: Wenn τ eine beliebige oberhalb l gelegene Zahl τ des Satzes 1 ist, $\tau \geq l$, dann ist

$$(9) \quad \left| \frac{1}{\tau} \int_{x''}^{x''+\tau} f(\xi) d\xi - \frac{1}{\tau} \int_{x'}^{x'+\tau} f(\xi) d\xi \right| \leq \frac{\varepsilon}{2}$$

für alle Paare x', x'' erfüllt. Nachdem nämlich x beliebig gewählt ist, werde unter den Zahlen τ des Satzes 1 eine solche, etwa τ' , gewählt, daß $y = x - \tau'$ in das Intervall $(0, l)$ fällt,

$$0 < y < l.$$

Dann ist

$$\left| \frac{1}{\tau} \int_x^{x+\tau} f(\xi) d\xi - \frac{1}{\tau} \int_0^\tau f(\xi) d\xi \right| \leq \left| \frac{1}{\tau} \int_x^{x+\tau} f(\xi) d\xi - \frac{1}{\tau} \int_y^{y+\tau} f(\xi) d\xi \right| + \left| \frac{1}{\tau} \int_y^{y+\tau} f(\xi) d\xi - \frac{1}{\tau} \int_0^\tau f(\xi) d\xi \right|;$$

$$\begin{aligned} \left| \frac{1}{\tau} \int_x^{x+\tau} f(\xi) d\xi - \frac{1}{\tau} \int_y^{y+\tau} f(\xi) d\xi \right|^2 &= \left| \frac{1}{\tau} \int_y^{y+\tau} f(\xi + \tau') d\xi - \frac{1}{\tau} \int_y^{y+\tau} f(\xi) d\xi \right|^2 \\ &\leq \frac{1}{\tau} \int_y^{y+\tau} |f(\xi + \tau') - f(\xi)|^2 d\xi = J_1, \end{aligned}$$

$$\begin{aligned} \left| \frac{1}{\tau} \int_y^{y+\tau} f(\xi) d\xi - \frac{1}{\tau} \int_0^\tau f(\xi) d\xi \right|^2 &= \left| \frac{1}{\tau} \int_y^{y+\tau} f(\xi) d\xi - \frac{1}{\tau} \int_0^y f(\xi) d\xi \right|^2 \leq \frac{1}{\tau} \int_0^y |f(\xi + \tau) - f(\xi)|^2 d\xi \\ &\leq \frac{1}{l} \int_0^l |f(\xi + \tau) - f(\xi)|^2 d\xi = J_2. \end{aligned}$$

Wegen $\tau \geq l \geq L_0$ kann sowohl auf J_1 als auch auf J_2 die Abschätzung (5) angewendet werden,

$$J_1 \leq \frac{\varepsilon^2}{16} \quad \text{und} \quad J_2 \leq \frac{\varepsilon^2}{16}.$$

Es ist daher

$$(10) \quad \left| \frac{1}{\tau} \int_{x''}^{x'+\tau} - \frac{1}{\tau} \int_{x'}^{x'+\tau} \right| \leq \left| \frac{1}{\tau} \int_{x''}^{x'+\tau} - \frac{1}{\tau} \int_0^{\tau} \right| + \left| \frac{1}{\tau} \int_{x'}^{x'+\tau} - \frac{1}{\tau} \int_0^{\tau} \right| \leq \frac{\varepsilon}{2}.$$

Um den Beweis des Satzes 2 zu Ende zu führen, werde nunmehr das obige $\tau \geq l$ fest gewählt; die Schranke K genüge den Anforderungen von (2) mit $A_0 = L_0$, und es sei $n_0 \geq \frac{8K}{\varepsilon}$. Dann ist

$$L_3 = n_0 \cdot \tau$$

ein L_3 , dessen Existenz zu beweisen ist. Es sei nämlich $L \geq L_3$ beliebig, und

$$n\tau \leq L < (n+1)\tau,$$

also $n \geq n_0$. Man erhält mit Rücksicht auf $\tau \geq l \geq L_0$:

$$\begin{aligned} & \left| \frac{1}{L} \int_x^{x-L} f(\xi) d\xi - \frac{1}{n\tau} \int_x^{x+n\tau} f(\xi) d\xi \right| \\ & \leq \left| \frac{1}{L} - \frac{1}{n\tau} \right| \cdot \int_x^{x+L} |f(\xi)| d\xi + \left| \frac{1}{n\tau} \int_x^{x+L} f(\xi) d\xi - \frac{1}{n\tau} \int_x^{x+n\tau} f(\xi) d\xi \right| \\ & \leq \frac{\tau}{n\tau} \cdot \frac{1}{L} \int_x^{x+L} |f(\xi)| d\xi + \frac{1}{n} \cdot \frac{1}{\tau} \int_{x+\tau n}^{x+n\tau+\tau} |f(\xi)| d\xi \leq \frac{1}{n} K + \frac{1}{n} K \leq \frac{2}{n_0} K \leq \frac{\varepsilon}{4}. \end{aligned}$$

Weiter ist

$$\begin{aligned} & \left| \frac{1}{L} \int_{x''}^{x'+L} f(\xi) d\xi - \frac{1}{L} \int_{x'}^{x'+L} f(\xi) d\xi \right| \\ & \leq \left| \frac{1}{L} \int_{x''}^{x'+L} - \frac{1}{n\tau} \int_{x'}^{x'+n\tau} \right| + \left| \frac{1}{L} \int_{x'}^{x'+L} - \frac{1}{n\tau} \int_{x'}^{x'+n\tau} \right| + \left| \frac{1}{n\tau} \int_{x'}^{x'+n\tau} - \frac{1}{n\tau} \int_{x'}^{x'+n\tau} \right|. \end{aligned}$$

Von diesen drei Summanden sind die ersten beiden unter $\frac{\varepsilon}{4}$ gelegen. Setzt man (10) für die Paare

$$x', x''; x' + \tau, x'' + \tau; \dots; x' + (n-1)\tau, x'' + (n-1)\tau$$

an und bildet das arithmetische Mittel, so erkennt man, daß der dritte Summand unter $\frac{\varepsilon}{2}$, die ganze Summe also unter ε gelegen ist. Damit ist der Satz 2 bewiesen.

Der Satz 2 involviert offenbar die Existenz des Mittelwertes

$$M\{f(x)\} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} f(\xi) d\xi$$

für jede Wellenfunktion, ist jedoch damit nicht ausgeschöpft. — Wenn eine Funktion $F(x)$ diese Eigenschaft hat: Zu jedem $\varepsilon > 0$ gibt es ein $L_3(\varepsilon) > 0$ derart, daß

$$\left| \frac{1}{L} \int_{x''}^{x''+L} F(\xi) d\xi - \frac{1}{L} \int_{x'}^{x'+L} F(\xi) d\xi \right| \leq \varepsilon$$

erfüllt ist für alle $L \geq L_3$ und alle Paare x', x'' , dann will ich sagen,

$$\mathfrak{M}\{F(x)\} \text{ existiert};$$

zugleich soll $\mathfrak{M}\{F(x)\}$ zur Bezeichnung des dann vorhandenen Mittelwertes $M\{F(x)\}$ dienen. — Mit $\mathfrak{M}\{F_1(x)\}, \dots, \mathfrak{M}\{F_n(x)\}$ existiert auch

$$\mathfrak{M}\{c_1 F_1(x) + \dots + c_n F_n(x)\} = c_1 \mathfrak{M}\{F_1(x)\} + \dots + c_n \mathfrak{M}\{F_n(x)\}.$$

Der genaue Inhalt des Satzes 2 läßt sich nun wiedergeben durch den Satz 2a. *Für jede Wellenfunktion $f(x)$ existiert $\mathfrak{M}\{f(x)\}$.*

Wenn $\mathfrak{M}\{F(x)\}$ existiert, und $F(x)$ in der Weise von Parametern abhängt, daß sich die Zahlen $L_3(\varepsilon)$ unabhängig von den Parametern bestimmen lassen, so möge das durch „ $\mathfrak{M}\{F(x)\}$ existiert gleichmäßig in den Parametern“ zum Ausdruck gebracht werden. — Insbesondere existiert dann $\mathfrak{M}\{F(x+s)\}$ gleichmäßig in s .

§ 4.

Der Mittelwert $\mathfrak{M}\{|f(x)|^2\}$ und die Fourierkoeffizienten.

Man überzeugt sich ohne Mühe, daß sich der Beweis für die Existenz von $\mathfrak{M}\{f(x)\}$ ohne wesentliche Änderungen durchführen läßt, wenn von $f(x)$ nur folgendes vorausgesetzt wird:

α) $f(x)$ und $|f(x)|$ sind in jedem endlichen Intervall integrierbar.

β) $f(x)$ genügt der Forderung 1 in der Definition der Wellenfunktionen.

γ) Zu jedem $\varepsilon > 0$ gibt es ein $\vartheta > 0$ und ein $L > 0$, so daß, wenn \mathfrak{D} eine beliebige Teilmenge eines beliebigen Intervalls $x < \xi < x + L$ vom Maß $|\mathfrak{D}| \leq \vartheta L$ ist, stets

$$\frac{1}{L} \int_{\mathfrak{D}} |f(\xi)| d\xi \leq \varepsilon$$

erfüllt ist.

Wenn $f(x)$ eine Wellenfunktion ist, dann haben, behaupte ich, die Funktionen $f(x+s) \cdot \bar{f}(x)$ und $f(x+s) \cdot f(x)$ die Eigenschaften α) bis γ). Beim Beweise kann ich mich auf $f(x+s) \cdot \bar{f}(x)$ beschränken; für $f(x+s) \cdot f(x)$ ergibt sich das Entsprechende, wenn man überall den Querstrich fortläßt.

Die Eigenschaft α) kommt der Funktion $f(x+s)\bar{f}(x)$ offenbar zu. Hinsichtlich γ) folgt das gleiche aus der Schwarzischen Ungleichung:

$$\left(\frac{1}{L} \int_{\Omega} |f(\xi+s)\bar{f}(\xi)| d\xi\right)^2 \leq \frac{1}{L} \int_{\Omega} |f(\xi+s)|^2 d\xi \cdot \frac{1}{L} \int_{\Omega} |f(\xi)|^2 d\xi \leq \varepsilon^4,$$

unabhängig von s .

Endlich möge K so gewählt sein, daß

$$\frac{1}{A} \int_x^{x+A} |f(\xi)|^2 d\xi \leq K^2$$

für alle $A \geq 1$ und alle x zutrifft (1). Zu gegebenem Paar $\varepsilon > 0$, $\vartheta > 0$ kann nunmehr nach Satz 1 ein $L \geq 1$ und ein $l > 0$ so bestimmt werden, daß

$$\frac{1}{L} \int_x^{x+L} |f(\xi+\tau) - f(\xi)|^2 d\xi \leq \frac{\varepsilon^2 \vartheta^2}{4K^2}$$

im Sinne jenes Satzes erfüllt ist. Es ist

$$\begin{aligned} & f(x+s+\tau)\bar{f}(x+\tau) - f(x+s)\bar{f}(x) \\ &= f(x+s+\tau)[\bar{f}(x+\tau) - \bar{f}(x)] + \bar{f}(x)[f(x+s+\tau) - f(x+s)], \end{aligned}$$

und daher mit Hilfe der Schwarzischen Ungleichung

$$\begin{aligned} & \left(\frac{1}{L} \int_x^{x+L} |f(\xi+s+\tau)\bar{f}(\xi+\tau) - f(\xi+s)\bar{f}(\xi)| d\xi\right)^2 \\ & \leq 2 \cdot \frac{1}{L} \int_x^{x+L} |f(\xi+s+\tau)|^2 d\xi \cdot \frac{1}{L} \int_x^{x+L} |f(\xi+\tau) - f(\xi)|^2 d\xi \\ & \quad + 2 \cdot \frac{1}{L} \int_x^{x+L} |f(\xi)|^2 d\xi \cdot \frac{1}{L} \int_x^{x+L} |f(\xi+s+\tau) - f(\xi+s)|^2 d\xi \\ & \leq 4K^2 \frac{\varepsilon^2 \vartheta^2}{4K^2} = \varepsilon^2 \vartheta^2, \end{aligned}$$

$$\frac{1}{L} \int_x^{x+L} |f(\xi+s+\tau)\bar{f}(\xi+\tau) - f(\xi+s)\bar{f}(\xi)| d\xi \leq \varepsilon \vartheta$$

für alle x , und zwar für Werte τ , die in jedem Intervall der Länge l mindestens einmal vertreten sind; außerdem sind L und l unabhängig von s .

Wenn nun \mathfrak{B} die Menge der Stellen ξ des Intervalls $(x, x+L)$ ist, für die

$$|f(\xi+s+\tau)\bar{f}(\xi+\tau) - f(\xi+s)\bar{f}(\xi)| > \varepsilon$$

ist, so erhält man

$$\frac{1}{L} \varepsilon |\mathfrak{B}| \leq \varepsilon \vartheta,$$

$$|\mathfrak{B}| \leq \vartheta L.$$

Damit ist gezeigt, daß $f(x+s)\bar{f}(x)$ für jeden Wert s die Eigenschaft β) hat; und zwar sind die Zahlen $L(\varepsilon, \vartheta)$ und $l(\varepsilon, \vartheta)$ unabhängig von s .

Auf Grund der Bemerkungen zu Beginn dieses Paragraphen ist demnach bewiesen:

Satz 3. Für jede Wellenfunktion $f(x)$ existieren

$$\mathfrak{M}\{f(x+s)\bar{f}(x)\} \quad \text{und} \quad \mathfrak{M}\{f(x+s)f(x)\}$$

gleichmäßig in s .

Für jede Wellenfunktion existiert insbesondere

$$\mathfrak{M}\{|f(x)|^2\}.$$

Ferner existiert

$$m(s) = \mathfrak{M}\{|f(x+s) - f(x)|^2\}$$

gleichmäßig in s , und nach Satz 1 folgt für jedes $\varepsilon > 0$ die Existenz einer Zahl $l(\varepsilon) > 0$ derart, daß

$$(11) \quad m(s) \leq \varepsilon^2$$

in jedem Intervall der Länge l für mindestens ein s erfüllt ist. Die Zahlen $s = \tau = \tau(\varepsilon)$, die der Ungleichung (11) genügen, sind den „zu ε gehörigen Verschiebungszahlen“ der fastperiodischen Funktionen genau analog, und kommen mit diesen überein, wenn $f(x)$ fastperiodisch ist. Weyl nennt diese Zahlen τ „Fastperioden vom Annäherungsgrad ε “. Es steht nichts im Wege, diese Bezeichnungen auf die durch (11) erklärten Zahlen τ für Wellenfunktionen zu übertragen. Die Verschiebungszahlen oder Fastperioden der Wellenfunktionen sind jedoch, weil zu ihrer Definition die Existenz gewisser Mittelwerte vorausgesetzt werden muß oder das doch nur mühsam zu umgehen ist, nicht geeignet, die primäre Rolle zu spielen wie die Verschiebungszahlen in der Bohrschen Theorie. — Wenn τ_1, \dots, τ_n Verschiebungszahlen der Wellenfunktion $f(x)$ sind, die zu ε gehören, so ist $\tau_1 + \dots + \tau_n$ eine Verschiebungszahl von $f(x)$, die zu $n\varepsilon$ gehört, denn man hat

$$m(\tau_1 + \dots + \tau_n) \leq n(m(\tau_1) + \dots + m(\tau_n)) \leq (n\varepsilon)^2.$$

Herr Weyl hat bemerkt, daß für jede Funktion $f(x)$ seiner Klasse — also für jede Wellenfunktion — die zugeordnete Funktion

$$g(s) = \mathfrak{M}\{f(x+s)\bar{f}(x)\}$$

fastperiodisch ist; in der Tat erhält man bei beliebigem t mit Hilfe der Schwarzschen Ungleichung

$$\begin{aligned} |g(s+t) - g(s)|^2 &= |\mathfrak{M}\{[f(x+s+t) - f(x+s)] \cdot \bar{f}(x)\}|^2 \\ &\leq \mathfrak{M}\{|f(x+s+t) - f(x+s)|^2\} \cdot \mathfrak{M}\{|f(x)|^2\} = m(t) \cdot \mathfrak{M}\{|f(x)|^2\}, \end{aligned}$$

und (11) liefert jetzt für alle s und für genügend dicht gelegene τ

$$|g(s+\tau) - g(s)| \leq \varepsilon.$$

Die Stetigkeit von $g(s)$ folgt aus der Stetigkeit von $m(t)$ an der Stelle $t=0$, die ihrerseits aus

$$\lim_{L \rightarrow \infty} \frac{1}{L} \int_0^L |f(\xi+t) - f(\xi)|^2 d\xi = \mathfrak{M}\{|f(x+t) - f(x)|^2\}$$

gleichmäßig in t , und bei festem L nach Lebesgue

$$\lim_{t \rightarrow 0} \int_0^L |f(\xi+t) - f(\xi)|^2 d\xi = 0$$

abzulesen ist.

Wegen

$$\begin{aligned} m(s) &= \mathfrak{M}\{f(x+s)\bar{f}(x+s)\} - \mathfrak{M}\{f(x+s)\bar{f}(x)\} - \mathfrak{M}\{\bar{f}(x+s)f(x)\} \\ &\quad + \mathfrak{M}\{f(x)\bar{f}(x)\} = 2g(0) - g(s) - \bar{g}(s) \end{aligned}$$

ist auch $m(s)$ fastperiodisch im Bohrschen Sinne.

Die Existenz des Fourierkoeffizienten der Wellenfunktionen hängt nun wesentlich von der Gültigkeit des Summensatzes für Wellenfunktionen ab:

Satz 4. *Wenn $f_1(x)$ und $f_2(x)$ Wellenfunktionen sind, dann ist auch die Summe $f(x) = f_1(x) + f_2(x)$ eine Wellenfunktion.*

Denn weil die Funktionen $m_1(s)$ und $m_2(s)$, die den Wellenfunktionen $f_1(x)$ und $f_2(x)$ zugeordnet sind, fastperiodisch sind, so ist es auch ihre Summe $m_1(s) + m_2(s)$. Zu jedem $\varepsilon > 0$ gibt es also ein $l > 0$, und nunmehr in jedem Intervall der Länge l mindestens ein τ derart, daß

$$|m_1(s+\tau) + m_2(s+\tau) - m_1(s) - m_2(s)| \leq \frac{\varepsilon^2}{6}$$

für alle s erfüllt ist. Daraus für $s=0$:

$$m_1(\tau) + m_2(\tau) \leq \frac{\varepsilon^2}{6}.$$

Andererseits gibt es ein L_0 , so daß für alle $L \geq L_0$, alle x und alle s

$$\left| \frac{1}{L} \int_x^{x+L} |f_\nu(\xi+s) - f_\nu(\xi)|^2 d\xi - m_\nu(s) \right| \leq \frac{\varepsilon^2}{6} \quad (\nu = 1, 2)$$

zutritt, daher

$$\frac{1}{L} \int_x^{x+L} |f(\xi + \tau) - f(\xi)|^2 d\xi \leq 2 \left| \frac{1}{L} \int_x^{x+L} |f_1(\xi + \tau) - f_1(\xi)|^2 d\xi - m_1(\tau) \right|$$

$$+ 2 \left| \frac{1}{L} \int_x^{x+L} |f_2(\xi + \tau) - f_2(\xi)|^2 d\xi - m_2(\tau) \right| + 2(m_1(\tau) + m_2(\tau)) \leq \varepsilon^2.$$

Nach Satz 1 ist also $f(x)$ eine Wellenfunktion.

Aus den Sätzen 3 und 4 folgt, wenn $f_1(x)$ und $f_2(x)$ Wellenfunktionen sind, die Existenz von $\mathfrak{M}\{(f_1(x) + f_2(x))^2\}$ und $\mathfrak{M}\{(f_1(x) - f_2(x))^2\}$, daraus weiter wegen $4fg = (f+g)^2 - (f-g)^2$ der

Satz 5. Wenn $f_1(x)$ und $f_2(x)$ Wellenfunktionen sind, dann existieren

$$\mathfrak{M}\{f_1(x) \cdot f_2(x)\} \quad \text{und} \quad \mathfrak{M}\{|f_1(x) \cdot f_2(x)|\}.$$

Für jede Wellenfunktion $f(x)$ existiert insbesondere

$$a(\lambda) = \mathfrak{M}\{f(x) e^{-i\lambda x}\} \quad (\lambda \text{ reell}).$$

Aus der für beliebige b_1, \dots, b_n und reelle paarweise verschiedene μ_1, \dots, μ_n gültigen Identität

$$\mathfrak{M}\left\{ \left| f(x) - \sum_{\nu=1}^n b_\nu e^{i\mu_\nu x} \right|^2 \right\} = \mathfrak{M}\{|f(x)|^2\} - \sum_{\nu=1}^n |a(\mu_\nu)|^2 + \sum_{\nu=1}^n |b_\nu - a(\mu_\nu)|^2$$

erschließt man in bekannter Weise die Abzählbarkeit der Werte λ , für die $a(\lambda) \neq 0$ ausfällt, und nennt diese Werte $\lambda_1, \lambda_2, \dots$ die *Fourierexponenten*, die Konstanten $a_1 = a(\lambda_1), a_2 = a(\lambda_2), \dots$ die *Fourierkoeffizienten*, die Reihe $\sum_{\nu=1}^{\infty} a_\nu e^{i\lambda_\nu x}$ die *Fourierreihe* von $f(x)$,

$$f(x) \sim \sum_{\nu=1}^{\infty} a_\nu e^{i\lambda_\nu x}.$$

Aus der obigen Identität folgt ferner die Konvergenz der Reihe $\sum_{\nu=1}^{\infty} |a_\nu|^2$, und die sog. *Besselsche Ungleichung*

$$\sum_{\nu=1}^{\infty} |a_\nu|^2 \leq \mathfrak{M}\{|f(x)|^2\}.$$

Damit ist alles zusammengestellt, was erforderlich ist, um daran den Beweis von Herrn Weyl für die *Parsevalsche Gleichung* anzuknüpfen:

Satz 6. Für jede Wellenfunktion

$$f(x) \sim \sum_{\nu=1}^{\infty} a_\nu e^{i\lambda_\nu x}$$

ist

$$\sum_{\nu=1}^{\infty} |a_{\nu}|^2 = \mathfrak{M} \{ |f(x)|^2 \}$$

und — damit gleichbedeutend —

$$\lim_{n \rightarrow \infty} \mathfrak{M} \left\{ \left| f(x) - \sum_{\nu=1}^n a_{\nu} e^{i i_{\nu} x} \right|^2 \right\} = 0.$$

II. Abschnitt.

§ 5.

Vorbereitungen.

Hilfssatz. Wenn $f(x)$ eine Wellenfunktion ist, so gibt es zu jedem $\varepsilon > 0$ ein $L_0(\varepsilon) > 0$, und weiterhin zu jedem Paar $\varrho > 0$, $\sigma > 0$ ein $\delta = \delta(\varepsilon, \varrho, \sigma)$ von folgender Beschaffenheit: Aus jeder Wellenfunktion $F(x)$, die $f(x)$ im Mittel mit der Genauigkeit δ approximiert,

$$\mathfrak{M} \{ |F(x) - f(x)|^2 \} \leq \delta^2,$$

läßt sich mit Hilfe von Verschiebungszahlen $\tau_0, \dots, \tau_{m-1}$ der Funktion $f(x)$, die zu ϱ gehören, eine Wellenfunktion $G(x)$ von der Form

$$G(x) = \frac{F(x + \tau_0) + \dots + F(x + \tau_{m-1})}{m}$$

gewinnen, für die

$$\mathfrak{M} \{ |G(x) - f(x)|^2 \} \leq \sigma^2$$

und zugleich

$$\frac{1}{L_0} \int_x^{x+L_0} |G(\xi) - f(\xi)|^2 d\xi \leq \varepsilon^2$$

für alle x zutrifft. Außerdem ist

$$\frac{1}{L_0} \int_x^{x+L_0} |f(\xi + \tau) - f(\xi)|^2 d\xi \leq 2\varepsilon^2$$

für alle x und alle zu ε gehörigen Verschiebungszahlen von $f(x)$ erfüllt.

Beweis. Für jedes $\varepsilon > 0$ möge $l(\varepsilon)$ so beschaffen sein, daß in jedem Intervall der Länge l mindestens eine zu ε gehörige Verschiebungszahl von $f(x)$ gelegen ist. Wegen der gleichmäßigen Existenz von $\mathfrak{M} \{ |f(x+s) - f(x)|^2 \}$ kann man $L_0 = L_0(\varepsilon)$ so wählen, daß

$$\frac{1}{L} \int_x^{x+L} |f(\xi + \tau) - f(\xi)|^2 d\xi \leq 2\varepsilon^2$$

für alle x , alle $\Lambda \geq \Lambda_0$ und alle zu ε gehörigen Verschiebungszahlen von $f(x)$ erfüllt ist. — Nunmehr sei ein $\varepsilon > 0$ gegeben. Die Zahl

$$L_0 = L_0(\varepsilon) = \text{Max} \left(\Lambda_0(\varepsilon), \Lambda_0 \left(\frac{\varepsilon}{5} \right) \right)$$

wird sich als geeignet erweisen.

Zunächst ist die letzte Behauptung des Hilfssatzes offenbar erfüllt. — Es seien ϱ und σ zwei weitere positive Zahlen. Ohne Verlust an Allgemeinheit darf ϱ unterhalb einer vorgegebenen Schranke angenommen werden; es sei

$$\varrho \leq \frac{\sigma}{2} \quad \text{und} \quad \varrho \leq \frac{\varepsilon}{5}.$$

Die natürliche Zahl $k = k(\varepsilon, \varrho, \sigma)$ werde so gewählt, daß

$$k L_0 \geq l(\varrho)$$

ausfällt. Endlich sei

$$\delta = \delta(\varepsilon, \varrho, \sigma) = \text{Min} \left(\frac{\varepsilon}{5k}, \frac{\sigma}{2} \right),$$

und $F(x)$ eine Wellenfunktion, die $f(x)$ mit der Genauigkeit δ approximiert. Für ein passendes L , das oberhalb $2kL_0$ angenommen werden kann, und alle x ist dann

$$\frac{1}{L} \int_x^{x+L} |F(\xi) - f(\xi)|^2 d\xi \leq 2\delta^2.$$

Die natürliche Zahl m werde gemäß

$$2mkL_0 \leq L < 2(m+1)kL_0$$

bestimmt, und nunmehr seien $\tau_0, \dots, \tau_{m-1}$ zu ϱ , also gewiß auch zu $\frac{\varepsilon}{5}$ gehörige Verschiebungszahlen von $f(x)$ aus den Intervallen

$$2\mu kL_0 \leq \tau_\mu < 2\mu kL_0 + l(\varrho) \quad (\mu = 0, \dots, m-1),$$

insbesondere $\tau_0 = 0$, und es werde $\tau_m = L$ gesetzt⁹⁾. Es ist dann

$$\tau_{\mu+1} - \tau_\mu > L_0 \quad (\mu = 0, \dots, m-1)$$

und

$$4mkL_0 > L.$$

Ich behaupte: Die Wellenfunktion

$$G(x) = \frac{1}{m} \sum_{\mu=0}^{m-1} F(x + \tau_\mu)$$

leistet das Verlangte. In der Tat erhält man

⁹⁾ Natürlich braucht τ_m keine Verschiebungszahl zu sein.

$$\begin{aligned}
& \frac{1}{L_0} \int_x^{x+L_0} |G(\xi) - f(\xi)|^2 d\xi \\
\leq & \frac{2}{m} \sum_{\mu=0}^{m-1} \frac{1}{L_0} \int_x^{x+L_0} |f(\xi + \tau_\mu) - f(\xi)|^2 d\xi + \frac{2}{m} \sum_{\mu=0}^{m-1} \frac{1}{L_0} \int_x^{x+L_0} |F(\xi + \tau_\mu) - f(\xi + \tau_\mu)|^2 d\xi \\
\leq & 4 \left(\frac{\varepsilon}{5}\right)^2 + \frac{8k}{4mkL_0} \sum_{\mu=0}^{m-1} \int_x^{x+\tau_{\mu+1}-\tau_\mu} |F(\xi + \tau_\mu) - f(\xi + \tau_\mu)|^2 d\xi \\
\leq & 4 \left(\frac{\varepsilon}{5}\right)^2 + \frac{8k}{L} \sum_{\mu=0}^{m-1} \int_{x+\tau_\mu}^{x+\tau_{\mu+1}} |F(\xi) - f(\xi)|^2 d\xi \\
= & 4 \left(\frac{\varepsilon}{5}\right)^2 + \frac{8k}{L} \int_x^{x+L} |F(\xi) - f(\xi)|^2 d\xi \leq 4 \left(\frac{\varepsilon}{5}\right)^2 + 16k\delta^2 < \varepsilon^2.
\end{aligned}$$

Ferner ist

$$\begin{aligned}
\mathfrak{M} \{ |G(x) - f(x)|^2 \} &= \mathfrak{M} \left\{ \left| \frac{1}{m} \sum_{\mu=0}^{m-1} (F(x + \tau_\mu) - f(x)) \right|^2 \right\} \\
\leq & \mathfrak{M} \left\{ \frac{1}{m} \sum_{\mu=0}^{m-1} |F(x + \tau_\mu) - f(x)|^2 \right\} \\
\leq & \frac{2}{m} \sum_{\mu=0}^{m-1} \mathfrak{M} \{ |F(x + \tau_\mu) - f(x + \tau_\mu)|^2 \} + \frac{2}{m} \sum_{\mu=0}^{m-1} \mathfrak{M} \{ |f(x + \tau_\mu) - f(x)|^2 \} \\
\leq & 2\delta^2 + 2\rho^2 \leq \frac{\sigma^2}{2} + \frac{\sigma^2}{2} = \sigma^2.
\end{aligned}$$

Damit ist der Hilfssatz bewiesen. Der Approximationssatz des nächsten Paragraphen wird sich nun als Folge aus dem Satze 7 ergeben.

Satz 7. *Wenn die Folge von Wellenfunktionen $f_1(x), f_2(x), \dots$ im Mittel gegen die Wellenfunktion $f(x)$ konvergiert, dann kann man daraus eine Folge von Wellenfunktionen $g_1(x), g_2(x), \dots$ gewinnen, die im Mittel scharf gegen $f(x)$ konvergiert. Dies gelingt mit Funktionen $g_\nu(x)$ von der Form*

$$g_\nu(x) = \frac{1}{m_\nu} \sum_{\mu=1}^{m_\nu} f_{n_\nu}(x + c_\mu^{(\nu)}).$$

Zum Beweise seien allgemein für jedes Tripel $\varepsilon > 0, \rho > 0, \sigma > 0$ die Zahlen $L_0 = L_0(\varepsilon)$ und $\delta = \delta(\varepsilon, \rho, \sigma)$ so bestimmt, daß die Aussage des Hilfssatzes für $f(x)$ gültig ist. Nachdem $\varepsilon_1, \varepsilon_2, \dots$ als positive monotone Nullfolge gewählt ist,

$$\varepsilon_1 > \varepsilon_2 > \dots \rightarrow 0,$$

seien L_1, L_2, \dots erklärt durch

$$L_1 = L_0 \left(\frac{\varepsilon_1}{4}\right), \quad L_2 = L_0 \left(\frac{\varepsilon_2}{4}\right), \quad \dots, \quad L_\nu = L_0 \left(\frac{\varepsilon_\nu}{4}\right), \quad \dots$$

Für jedes ν werde ferner der Reihe nach gesetzt

$$(12) \quad \begin{aligned} \delta_1^{(\nu)} &= \delta \left(\frac{\varepsilon_1}{4}, \frac{\varepsilon_\nu}{4\nu}, 1 \right) \\ \delta_2^{(\nu)} &= \delta \left(\frac{\varepsilon_2}{4}, \frac{\varepsilon_\nu}{4\nu}, \delta_1^{(\nu)} \right) \\ &\dots \dots \dots \dots \dots \dots \dots \dots \dots \\ \delta_{\nu-1}^{(\nu)} &= \delta \left(\frac{\varepsilon_{\nu-1}}{4}, \frac{\varepsilon_\nu}{4\nu}, \delta_{\nu-2}^{(\nu)} \right) \\ \delta_\nu^{(\nu)} &= \delta \left(\frac{\varepsilon_\nu}{4}, \frac{\varepsilon_\nu}{4\nu}, \delta_{\nu-1}^{(\nu)} \right). \end{aligned}$$

Ein beliebiges ν werde jetzt festgehalten. Die oberen Indizes ν in (12) können dann vorläufig fortgelassen werden. Für einen passenden Index n ist voraussetzungsgemäß

$$\mathfrak{M} \left\{ |f_n(x) - f(x)|^2 \right\} \leq \delta_\nu^2.$$

Anwendung des Hilfssatzes auf $F(x) = f_n(x)$ liefert eine Wellenfunktion $G(x) = f_{n\nu}(x)$, die den Bedingungen

$$\frac{1}{L_\nu} \int_x^{x+L_\nu} |f_{n\nu}(\xi) - f(\xi)|^2 d\xi \leq \left(\frac{\varepsilon_\nu}{4}\right)^2$$

für alle x , und zugleich

$$\mathfrak{M} \left\{ |f_{n\nu}(x) - f(x)|^2 \right\} \leq \delta_{\nu-1}^2$$

genügt. Auf $F(x) = f_{n\nu}(x)$ läßt sich der Hilfssatz erneut anwenden. Es entsteht $G(x) = f_{n,\nu-1}(x)$ mit

$$\frac{1}{L_{\nu-1}} \int_x^{x+L_{\nu-1}} |f_{n,\nu-1}(\xi) - f(\xi)|^2 d\xi \leq \left(\frac{\varepsilon_{\nu-1}}{4}\right)^2$$

und zugleich

$$\mathfrak{M} \left\{ |f_{n,\nu-1}(x) - f(x)|^2 \right\} \leq \delta_{\nu-2}^2.$$

Und so fort. Beim vorletzten Schritt hat man $f_{n2}(x)$ mit

$$\frac{1}{L_2} \int_x^{x+L_2} |f_{n2}(\xi) - f(\xi)|^2 d\xi \leq \left(\frac{\varepsilon_2}{4}\right)^2$$

und zugleich

$$\mathfrak{M} \left\{ |f_{n2}(x) - f(x)|^2 \right\} \leq \delta_1^2;$$

schließlich $f_{n1}(x)$ mit

$$\frac{1}{L_1} \int_x^{x+L_1} |f_{n1}(\xi) - f(\xi)|^2 d\xi \leq \left(\frac{\varepsilon_1}{4}\right)^2.$$

Die Wellenfunktion

$$g(x) = f_{n_1}(x)$$

hat folgende Eigenschaft: *Es ist*

$$(13) \quad \frac{1}{L_\kappa} \int_x^{x+L_\kappa} |g(\xi) - f(\xi)|^2 d\xi \leq \varepsilon_\kappa^2$$

für alle x und für jeden Index $\kappa = 1, \dots, \nu$ erfüllt. Denn für $\kappa = 1$ stimmt das mit dem Ergebnis des letzten Schrittes überein. Für $\kappa = 2$ kann man so schließen: $g(x)$ ist von der Form

$$g(x) = \frac{1}{m} \sum_{\mu=0}^{m-1} f_{n_2}(x + \tau_\mu),$$

wo $\tau_0, \dots, \tau_{m-1}$ Verschiebungszahlen von $f(x)$ sind, die zu $\frac{\varepsilon_\nu}{4\nu}$, also wegen $\frac{\varepsilon_\nu}{4\nu} < \frac{\varepsilon_2}{4}$ auch zu $\frac{\varepsilon_2}{4}$ gehören. Unter Benutzung der letzten Aussage des Hilfssatzes erhält man

$$\begin{aligned} \frac{1}{L_2} \int_x^{x+L_2} |g(\xi) - f(\xi)|^2 d\xi &\leq \frac{1}{m} \sum_{\mu=0}^{m-1} \frac{1}{L_2} \int_x^{x+L_2} |f_{n_2}(\xi + \tau_\mu) - f(\xi)|^2 d\xi \\ &\leq \frac{2}{m} \sum_{\mu=0}^{m-1} \frac{1}{L_2} \int_x^{x+L_2} |f(\xi + \tau_\mu) - f(\xi)|^2 d\xi + \frac{2}{m} \sum_{\mu=0}^{m-1} \frac{1}{L_2} \int_x^{x+L_2} |f_{n_2}(\xi + \tau_\mu) - f(\xi + \tau_\mu)|^2 d\xi \\ &\leq 4 \left(\frac{\varepsilon_2}{4}\right)^2 + 2 \left(\frac{\varepsilon_2}{4}\right)^2 < \varepsilon_2^2. \end{aligned}$$

Eine ähnliche Erwägung wie im Falle $\kappa = 2$ führt auch allgemein ($2 \leq \kappa \leq \nu$) zum Ziele: $g(x)$ ist von der Form

$$g(x) = \frac{1}{m} \sum_{\mu=0}^{m-1} f_{n_\kappa}(x + \tau_\mu),$$

wo jede der Zahlen $\tau_0, \dots, \tau_{m-1}$ eine Summe von Verschiebungszahlen von $f(x)$ ist, die zu $\frac{\varepsilon_\nu}{4\nu}$ gehören¹⁰); die Anzahl der Summanden ist $\kappa - 1$, die Zahlen $\tau_0, \dots, \tau_{m-1}$ sind also Verschiebungszahlen von $f(x)$, die zu $(\kappa - 1) \frac{\varepsilon_\nu}{4\nu}$, also wegen $(\kappa - 1) \frac{\varepsilon_\nu}{4\nu} < \frac{\varepsilon_\kappa}{4}$ auch zu $\frac{\varepsilon_\kappa}{4}$ gehören. Alles weitere verläuft genau analog zum Falle $\kappa = 2$.

¹⁰) Z. B. erhält man für $\kappa = 3$:

$$g(x) = \frac{1}{m_1} \sum_{\mu=0}^{m_1-1} f_{n_2}(x + \tau_\mu^{(1)}) = \frac{1}{m_1 \cdot m_2} \sum_{\mu=0}^{m_1-1} \sum_{\nu=0}^{m_2-1} f_{n_3}(x + \tau_\mu^{(1)} + \tau_\nu^{(2)}).$$

Ferner hat $g(x)$, wie sich nach dem letzten Schritt, $\kappa = \nu$, ergibt, die Form

$$g(x) = \frac{1}{m} \sum_{\mu=1}^m f_n(x + c_\mu).$$

Für jedes $\nu = 1, 2, \dots$ erhält man eine solche Funktion

$$g_\nu(x) = \frac{1}{m_\nu} \sum_{\mu=1}^{m_\nu} f_{n_\nu}(x + c_\mu^{(\nu)}).$$

Ich behaupte: Die so entstehende Folge von Wellenfunktionen $g_1(x), g_2(x), \dots$ konvergiert im Mittel scharf gegen $f(x)$. Bestimmt man nämlich zu gegebenem $\varepsilon > 0$ den Index n so, daß $\varepsilon_n \leq \varepsilon$ ist, so sind L_n und dieses n solche Werte von $L(\varepsilon)$ und $n(\varepsilon)$, wie sie in der Definition der scharfen Konvergenz im Mittel gegen eine Wellenfunktion gefordert werden. Denn in bezug auf jeden Index $\nu \geq n$ spielt n selbst die Rolle eines κ in (13).

§ 6.

Die trigonometrische Approximation für Wellenfunktionen.

Satz 8. Zu jeder Wellenfunktion

$$f(x) \sim \sum_{\nu=1}^{\infty} a_\nu e^{i\lambda_\nu x}$$

gibt es eine Folge von trigonometrischen Polynomen der Form

$$Q_n(x) = \sum_{\nu=1}^{k_n} b_\nu^{(n)} e^{i\lambda_\nu x} \quad (n = 1, 2, \dots),$$

die im Mittel scharf gegen $f(x)$ konvergiert.

Dieser Satz folgt unmittelbar aus den Sätzen 6 und 7. Bedeuten nämlich $P_k(x)$ die Abschnitte der Fourierreihe von $f(x)$,

$$P_k(x) = \sum_{\nu=1}^k a_\nu e^{i\lambda_\nu x} \quad (k = 1, 2, \dots),$$

so konvergiert die Folge von Wellenfunktionen $f_1(x) = P_1(x)$, $f_2(x) = P_2(x)$, ... im Mittel gegen $f(x)$ (Satz 6), und der Satz 7 liefert die Funktionenfolge

$$\begin{aligned} g_n(x) = Q_n(x) &= \frac{1}{m_n} \sum_{\mu=1}^{m_n} P_{k_n}(x + c_\mu^{(n)}) = \sum_{\nu=1}^{k_n} \left(\frac{1}{m_n} \sum_{\mu=1}^{m_n} a_\nu e^{i\lambda_\nu c_\mu^{(n)}} \right) e^{i\lambda_\nu x} \\ &= \sum_{\nu=1}^{k_n} b_\nu^{(n)} e^{i\lambda_\nu x} \quad (n = 1, 2, \dots), \end{aligned}$$

die im Mittel scharf gegen $f(x)$ konvergiert.

Über die Verallgemeinerungen fastperiodischer Funktionen.

Von

Harald Bohr in Kopenhagen.

An die ersten Arbeiten über fastperiodische Funktionen (Acta math. Bd. 45, 46, 47) hat sich eine Reihe von weiteren Arbeiten angeschlossen. Unter anderem sind mehrere *Verallgemeinerungen* dieser Funktionenklasse angegeben¹⁾. Die ersten solchen Verallgemeinerungen rühren von Stepanoff her (Math. Ann. Bd. 95); unabhängig von Stepanoff wurde auch Wiener (Math. Zeitschr. Bd. 24) bei seinen bedeutsamen Untersuchungen über Fourierintegrale auf eine der Stepanoffschen Funktionenklassen geführt. Eine sehr weitgehende Verallgemeinerung wurde nachher von Besicovitch (Proc. London Math. Soc. Bd. 25) angegeben. Schließlich hat Weyl am Ende seiner sehr interessanten Arbeit „Integralgleichungen und fastperiodische Funktionen“ (Math Ann. Bd. 97), in welcher eine neue Begründung der Hauptsätze der Theorie der eigentlichen fastperiodischen Funktionen gegeben wurde, eine von den oben erwähnten verschiedene Verallgemeinerung des Begriffes der Fastperiodizität aufgestellt und die so verallgemeinerte Funktionenklasse mit Hilfe seiner Methoden untersucht. An diese Weylsche Abhandlung schließt sich die der vorliegenden kleinen Abhandlung vorangehende Arbeit von R. Schmidt; eine kurze Voranzeige seiner Resultate hatte Schmidt schon in einer Note in den Jahresberichten der deutschen Mathematikervereinigung (Bd. 36) veröffentlicht.

Diese verschiedenen Verallgemeinerungen und andere damit zusammenhängende sind neuerdings von Herrn Besicovitch und mir einer eingehenden Untersuchung unterworfen worden. Die allgemeinen Gesichtspunkte, die wir dabei zugrunde legen, sowie eine Anwendung derselben auf die

¹⁾ Ich beschäftige mich im folgenden nur mit Funktionen *einer reellen Variablen* und lasse analytische Funktionen einer komplexen Variablen sowie Funktionen mehrerer Variablen außer Betracht.

Stepanoffschen Funktionen, haben wir bereits in einer Abhandlung in den Mitteilungen der dänischen Akademie (Bd. 8) dargestellt. Eine umfassende Untersuchung des ganzen Fragenkomplexes werden wir demnächst in einer weiteren gemeinsamen Arbeit in den Acta mathematica veröffentlichen. Da so viele der einschlägigen Arbeiten gerade in den Mathematischen Annalen erschienen sind, mag es vielleicht gestattet sein, auch an dieser Stelle auf die Gesichtspunkte von Besicovitch und mir kurz einzugehen — und zwar indem ich im wesentlichen einen Vortrag wiedergebe, den ich im Januar 1928 in der Göttinger Mathematischen Gesellschaft gehalten habe — und darüber hinaus beispielsweise zu berichten, wie die Resultate von Weyl und Schmidt sich in den Rahmen unserer allgemeinen Untersuchungen einordnen.

1. Zunächst ein paar Worte über die Theorie der eigentlichen fast-periodischen Funktionen (kurz f. p. Funktionen).

Unter der abgeschlossenen Hülle einer Menge Φ von Funktionen $F(x)$ ($-\infty < x < \infty$) wollen wir diejenige Funktionenmenge $H(\Phi)$ verstehen, welche aus der gegebenen Menge Φ entsteht, wenn diese durch Hinzufügung aller solcher Funktionen erweitert wird, die *gleichmäßig für alle x* durch Funktionen der Menge Φ approximiert werden können.

Es bezeichne überall im folgenden E die Menge aller Exponentialpolynome (endlichen Summen von „reinen Schwingungen“)

$$s(x) = \sum_{\nu=1}^n a_{\nu} e^{i \lambda_{\nu} x} \quad (\lambda_{\nu} \text{ reell}).$$

Ferner bezeichne F die Menge der f. p. Funktionen, d. h. die Menge aller stetigen Funktionen $f(x)$ ($-\infty < x < \infty$) mit der folgenden Eigenschaft: Zu jedem $\varepsilon > 0$ gibt es eine „relativ dicht“ liegende Menge von Zahlen (Verschiebungszahlen) $\tau = \tau(\varepsilon)$ mit

$$|f(x + \tau) - f(x)| < \varepsilon \quad (-\infty < x < \infty);$$

hierbei sollen die Worte „relativ dicht“ bedeuten, daß in jedem Intervall einer gewissen Länge $l = l(\varepsilon)$ mindestens eine der Zahlen τ vorhanden ist.

Ein Hauptsatz der Theorie der f. p. Funktionen (der Approximationsatz) besagt, daß die Menge F gerade die abgeschlossene Hülle von E ist, also

$$(1) \quad \boxed{F = H(E)}.$$

Durch diese Gleichung wird unsere Funktionenklasse von zwei verschiedenen Gesichtspunkten aus charakterisiert: einerseits durch *Schwingungseigenschaften* als abgeschlossene Hülle von endlichen Summen reiner Schwingungen, und andererseits durch *Verschiebungseigenschaften* (Periodizitätseigenschaften) als f. p. Funktionen.

Zu jeder f. p. Funktion gehört eine Fourierreihe $\sum A_n e^{iA_n x}$, welche ihrerseits die Funktion $f(x)$ eindeutig bestimmt. Der Summationssatz von Bochner (Math. Ann., Bd. 96), der den berühmten Fejérschen Summationssatz für Fourierreihen reinperiodischer Funktionen auf f. p. Funktionen überträgt, liefert einen *Algorithmus*, welcher von der Fourierreihe aus zu gleichmäßig approximierenden Exponentialpolynomen führt.

2. Bei der Aufgabe, die Theorie der f. p. Funktionen zu *verallgemeinern*, wird es sich vor allem darum handeln, den Approximationssatz, d. h. die obige Gleichung (1) zu verallgemeinern. Dabei kann man in zwei verschiedenen Weisen vorgehen.

Entweder kann man von der linken Seite der Gleichung ausgehen, d. h. direkt versuchen, die Definition der Fastperiodizität zu erweitern, vor allem dadurch, daß man die Forderung der Stetigkeit aufgibt und zu Funktionen übergeht, die nur solchen Bedingungen unterworfen sind wie etwa der der Meßbarkeit oder Integrierbarkeit im Lebesgueschen Sinne. So ist z. B. Stepanoff, der als erster und mit großem Erfolg das Problem der Verallgemeinerung in Angriff genommen hat, vorgegangen. Die Aufgabe wird dann sein, die Schwingungseigenschaften der so definierten verallgemeinerten f. p. Funktionen zu studieren.

Oder man kann umgekehrt von der rechten Seite der Gleichung (1) (also von den Schwingungseigenschaften) ausgehen, indem man den Begriff der abgeschlossenen Hülle der Klasse E verallgemeinert, d. h. einen anderen Limesbegriff als den der überall gleichmäßigen Konvergenz zugrunde legt. Man wird hier vor die umgekehrte Aufgabe gestellt, nämlich, für die so definierte abgeschlossene Hülle der Menge aller Exponentialpolynome die zugehörigen verallgemeinerten fastperiodischen Eigenschaften zu erforschen. In dieser Weise ist vor allem Besicovitch vorgegangen.

Für eine systematische Behandlung der Theorie der Verallgemeinerungen f. p. Funktionen bietet sich, unserer Ansicht nach, der letzte Gesichtspunkt als der natürlichere dar. Die Aufgabe ist hier eine ganz klare und eindeutige; man betrachte nach und nach verschiedene Grenzprozesse G , bilde jedesmal die entsprechende abgeschlossene Hülle $H_G(E)$ und suche danach diese Funktionenklasse durch Periodizitätseigenschaften zu charakterisieren.

3. Im folgenden werde ich mich, der Kürze halber, auf Funktionen der Klasse L^2 beschränken, d. h. der Klasse aller Funktionen $f(x)$ ($-\infty < x < \infty$), die samt dem Betrag ihres Quadrats in jedem endlichen Intervall im Lebesgueschen Sinne integrierbar sind; nur diese Klasse L^2 kommt nämlich für die spätere Diskussion der Resultate von Weyl und Schmidt in Betracht. Besicovitch und ich haben übrigens alle Klassen L^2 ,

wo p eine beliebige Zahl ≥ 1 ist, untersucht; hierbei zeigt sich die Klasse L^1 (und nicht etwa die Klasse L^2) insofern als die natürliche Grundlage, als man von dieser ersten Klasse heraus leicht zu den höheren Klassen aufsteigen kann.

In der Theorie der *reinperiodischen* Funktionen der Klasse L^2 , wo es sich um ein *endliches* Intervall (a, b) handelt, wird bekanntlich der folgende Limesbegriff zugrunde gelegt: Es soll $\lim f_n(x) = f(x)$ bedeuten, daß der Mittelwert

$$\frac{1}{b-a} \int_a^b |f(x) - f_n(x)|^2 dx$$

für $n \rightarrow \infty$ gegen 0 strebt. Oder anders ausgedrückt, indem wir in der bekannten Weise einen Entfernungsbegriff zweier Funktionen einführen: Es wird die Entfernung (Distanz) von $f(x)$ und $g(x)$ durch

$$D[f(x), g(x)] = \sqrt{\frac{1}{b-a} \int_a^b |f(x) - g(x)|^2 dx}$$

definiert, und danach der Limesbegriff $\lim f_n(x) = f(x)$ einfach durch

$$D[f(x), f_n(x)] \rightarrow 0 \quad \text{für } n \rightarrow \infty$$

erklärt.

In der Theorie der *fastperiodischen* Funktionen handelt es sich prinzipiell um ein *unendliches* Intervall $-\infty < x < \infty$. Wenn man den obigen Limesbegriff, oder besser den obigen Entfernungsbegriff, von einem endlichen auf ein unendliches Intervall zu übertragen wünscht, wird man vor die Wahl mehrerer verschiedener Möglichkeiten gestellt, von denen jede ihre besonderen Eigentümlichkeiten und ihr besonderes Interesse darbietet. Wir führen drei solcher Entfernungsbegriffe ein, die wir mit $D_S[f(x), g(x)]$, $D_B[f(x), g(x)]$ und $D_W[f(x), g(x)]$ bezeichnen, weil sie auf das engste mit den Verallgemeinerungen der f. p. Funktionen von Stepanoff, Besicovitch und Weyl verknüpft sind.

Der *Stepanoffsche Entfernungsbegriff* ist durch

$$D_S[f(x), g(x)] = \sqrt{\text{Ob. Gr. } \frac{1}{L} \int_x^{x+L} |f(\xi) - g(\xi)|^2 d\xi}$$

gegeben. Hierbei ist L eine feste positive Zahl; ihr Wert ist gleichgültig (sie kann z. B. gleich 1 angenommen werden), weil der Grenzbegriff, welcher durch diesen Entfernungsbegriff festgelegt wird, wie leicht zu sehen, von dem speziellen Werte von L nicht abhängt.

Bei dem *Besicovitchschen Entfernungsbegriff* $D_B[f(x), g(x)]$ wird der Mittelwert sogleich über das ganze Intervall $-\infty < x < \infty$ erstreckt:

$$D_B[f(x), g(x)] = \sqrt{\limsup_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T |f(x) - g(x)|^2 dx}.$$

Der *Weylsche Entfernungsbegriff* ist schließlich ein „Mittelding“ zwischen den beiden obigen; es wird hier zunächst, wie bei Stepanoff, eine feste Länge L betrachtet, die man aber danach über alle Grenzen wachsen läßt:

$$D_W[f(x), g(x)] = \sqrt{\lim_{L \rightarrow \infty} \text{Ob. Gr.} \frac{1}{L} \int_x^{x+L} |f(\xi) - g(\xi)|^2 d\xi}.$$

Es ist mit anderen Worten bei dem Weylschen Entfernungsbegriff die Entfernung zweier Funktionen $f(x)$ und $g(x)$ klein, wenn nur für irgendein (vielleicht ungeheuer großes) L der Mittelwert

$$\frac{1}{L} \int_x^{x+L} |f(\xi) - g(\xi)|^2 d\xi$$

für alle x klein ausfällt.

Bei jedem der drei Entfernungsbegriffe wird der entsprechende Limesbegriff

$$S\text{-lim } f_n(x) = f(x), \quad B\text{-lim } f_n(x) = f(x), \quad W\text{-lim } f_n(x) = f(x)$$

durch die Relation

$$D_S[f(x), f_n(x)] \rightarrow 0, \quad D_B[f(x), f_n(x)] \rightarrow 0, \quad D_W[f(x), f_n(x)] \rightarrow 0$$

erklärt. Und jedesmal wird mit Hilfe des zugrunde gelegten Limesbegriffes die abgeschlossene Hülle der Menge \mathbf{E} gebildet, also beziehungsweise

$$H_S(\mathbf{E}), \quad H_B(\mathbf{E}), \quad H_W(\mathbf{E}).$$

4. Bevor ich zur Besprechung der allgemeinen Methode übergehe, durch welche Besicovitch und ich die — zum Teil schon durch die frühere Literatur bekannten — Verschiebungseigenschaften aufsuchen, welche diese drei Funktionenklassen $H_S(\mathbf{E})$, $H_B(\mathbf{E})$, $H_W(\mathbf{E})$ charakterisieren, seien die folgenden Bemerkungen eingeschaltet über die eigentümlichen Eigenschaften, welche jeder dieser drei Klassen ihr besonderes Interesse verleihen. Die Stepanoffsche Klasse ist die engste und daher die einfachste; in der Tat hat sie sehr viele Eigenschaften mit der (natürlich darin enthaltenen) Klasse der eigentlichen f. p. Funktionen gemeinsam; vor allem bleibt hier — im Gegensatz zu den beiden anderen Klassen — der Satz von der *eindeutigen Bestimmtheit* einer Funktion durch ihre Fourierreihe einigermaßen gut

erhalten; er lautet, daß eine Stepanoffsche Funktion durch ihre Fourierreihe *bis auf eine Menge vom Maße Null* bestimmt ist. Das Interesse der Besicovitschen Klasse liegt nicht nur darin, daß sie bei weitem die umfassendste ist; ihre besondere Bedeutung gewinnt sie dadurch, daß erst hier das *Analogon des Riesz-Fischerschen Satzes* gilt; es besagt, daß die notwendige und hinreichende Bedingung dafür, daß eine unendliche Reihe $\sum A_n e^{iA_n x}$ die Fourierreihe einer Besicovitschen Funktion ist, einfach darin besteht, daß $\sum |A_n|^2$ konvergiert. Schließlich bringt die Weylsche Klasse ihre Stellung als natürliche *Mittelstufe* zwischen der Stepanoffschen und der Besicovitschen Klasse durch mehrere interessante und charakteristische Eigenschaften zum Ausdruck.

5. Indem wir nun den Faden der Entwicklung wieder aufnehmen, stehen wir also vor der Aufgabe, die drei Funktionenklassen $H_S(\mathbf{E})$, $H_B(\mathbf{E})$, $H_W(\mathbf{E})$ durch Periodizitätseigenschaften zu charakterisieren, also vor genau derselben Aufgabe, welche für die mit Hilfe gleichmäßiger Konvergenz abgeschlossene Hülle $\mathbf{H}(\mathbf{E})$ durch die Gleichung

$$(1) \quad \mathbf{H}(\mathbf{E}) = \mathbf{F}$$

gelöst wurde.

Durch eine äußerst einfache, aber prinzipiell wichtige Betrachtung führen Besicovitch und ich dieses Problem auf das entsprechende, durch die Gleichung (1) gelöste Problem in der ursprünglichen Theorie zurück. Indem wir die Tatsache berücksichtigen, daß die gleichmäßige Konvergenz der engste aller betrachteten Grenzprozesse ist — d. h. *daß die mit Hilfe gleichmäßiger Konvergenz abgeschlossene Hülle $\mathbf{H}(\mathbf{E})$ die engste aller Hüllen $H(\mathbf{E})$ ist* — können wir nämlich den Approximationssatz (1) der ursprünglichen Theorie beim Aufbau der verallgemeinerten Theorien direkt als fertiges Resultat benutzen, ohne (wie es die früheren Verfasser getan haben) auf den Beweis dieses Satzes zurückzukommen. Aus der Ungleichung

$$\mathbf{E} < \mathbf{H}(\mathbf{E}) < H_G(\mathbf{E}),$$

wo G irgendeinen der drei obigen Grenzprozesse S -, B - oder W -lim bedeutet, folgt nämlich unmittelbar

$$H_G(\mathbf{E}) = H_G(\mathbf{H}(\mathbf{E})),$$

also, wegen $\mathbf{H}(\mathbf{E}) = \mathbf{F}$,

$$\boxed{H_G(\mathbf{E}) = H_G(\mathbf{F})}.$$

Hiermit ist aber gezeigt, daß das Problem der *fastperiodischen Charakterisierung der Funktionenklasse $H_G(\mathbf{E})$ einfach damit gleichbedeutend ist, die Wirkung zu untersuchen, welche der betrachtete Grenzprozeß G hervor-*

bringt, wenn er auf die eigentlichen f. p. Funktionen der ursprünglichen Klasse F ausgeübt wird. Wir brauchen uns also gar nicht mehr mit der Klasse E , d. h. mit der Theorie der Schwingungen zu befassen; die Verbindung zwischen Schwingungen und Periodizitätseigenschaften ist ein für allemal durch die ursprüngliche Gleichung (1) zustande gebracht; was zu untersuchen übrig bleibt, hat nur noch mit reinen Verschiebungseigenschaften zu tun.

Was die Behandlung der Stepanoffschen Funktionen von diesem Gesichtspunkte aus betrifft, vergleiche man die schon oben zitierte Abhandlung von Besicovitch und mir in den Mitteilungen der dänischen Akademie, und was die entsprechende — aber viel schwierigere — Untersuchung der Besicovitchschen Klasse anbelangt, verweise ich auf unsere gemeinsame, bald erscheinende Arbeit in den Acta mathematica. Hier soll nur mit einigen Worten auf die Weylschen Funktionen eingegangen werden, die ebenfalls in unserer Acta-Arbeit ausführlich behandelt werden.

6. Die Definition der von Weyl in seiner Arbeit über Integralgleichungen und f. p. Funktionen eingeführten verallgemeinerten f. p. Funktionen (wir wollen sie als W. f. p. Funktionen bezeichnen) lautet:

Eine Funktion der Klasse L^2 soll eine W. f. p. Funktion genannt werden, falls es zu jedem $\varepsilon > 0$ eine relativ dicht liegende Menge von Zahlen τ sowie ein zugehöriges L_0 derart gibt, daß für jede dieser Zahlen τ die Ungleichung

$$\text{Ob. Gr. } \frac{1}{L} \int_x^{x+L} |f(\xi + \tau) - f(\xi)|^2 d\xi < \varepsilon$$

für $L > L_0$ besteht.

Das Hauptresultat von Weyl, insofern von Verallgemeinerungen die Rede ist, besagt (obwohl in etwas anderer Formulierung), daß die Klasse F_W der W. f. p. Funktionen gerade mit der oben eingeführten Hülle $H_W(E)$ übereinstimmt, also

$$(2) \quad H_W(E) = F_W.$$

Weyl bewies diesen schönen Satz, indem er zeigte, daß die auf der Theorie der Integralgleichungen beruhende Methode, welche er zum Beweise der Hauptgleichung $H(E) = F$ der ursprünglichen Theorie ausgebildet hatte, auf die Klasse der W. f. p. Funktionen übertragbar war. In unserer Acta-Arbeit werden Besicovitch und ich eine neue Begründung dieser Weylschen Gleichung (2) geben, und zwar in der oben angedeuteten Weise mit Hilfe der Relation $H_W(E) = H_W(F)$. Da wir somit auf den Beweis der Hauptgleichung (1) gar nicht zurückkommen, sondern sie in fertiger Form benutzen, ist es für unsere Herleitung des Weylschen Satzes ganz gleich-

gültig, in welcher Weise diese Gleichung (1) abgeleitet wird — ob mit der ursprünglichen mehr elementareren Methode, ob mit der von Wiener angegebenen Methode, welche die Hilfsmittel der modernen Theorie der reellen Funktionen heranzieht, oder mit der oben erwähnten Weylschen Methode. Und haben wir erst die Gleichung $F_W = H_W(E)$ bewiesen, können wir nachher alle übrigen Eigenschaften der W. f. p. Funktionen: daß die Summe zweier W. f. p. Funktionen wieder eine W. f. p. Funktion ist, daß der Mittelwert einer W. f. p. Funktion existiert, daß sie eine Fourierreihe besitzt usw. sofort ableiten, indem wir von der Charakterisierung $H_W(E)$ (statt von F_W) ausgehen.

Zu leichterem Verständnis des unten Folgenden werde ich den Weylschen Approximationssatz $F_W = H_W(E)$ auch in der folgenden Formulierung aussprechen, wo explizite zum Ausdruck gebracht ist, was es eigentlich bedeutet, daß eine Funktion $f(x)$ der abgeschlossenen Hülle $H_W(E)$ angehört.

Damit eine Funktion $f(x)$ der Klasse L^2 eine W. f. p. Funktion sei, ist notwendig und hinreichend, daß sie in der folgenden Weise durch Summen von Schwingungen approximiert werden kann: Es soll eine Folge von Exponentialpolynomen

$$s_1(x), s_2(x), \dots, s_n(x), \dots$$

sowie zwei dazu gehörige Zahlenfolgen

$$\varepsilon_1 > \varepsilon_2 > \dots > \varepsilon_n \dots \quad (\varepsilon_n \rightarrow 0)$$

und

$$L_1, L_2, \dots, L_n, \dots$$

derart existieren, daß für jedes n die Ungleichung

$$\text{Ob. Gr. } \frac{1}{L_n} \int_x^{x+L_n} |f(\xi) - s_n(\xi)|^2 d\xi < \varepsilon_n$$

besteht.

Hierbei können übrigens, wie aus der Parsevalschen Gleichung folgt, als Approximationssummen $s_n(x)$ einfach die Abschnitte $\sum_1^n A_\nu e^{iA_\nu x}$ der Fourierreihe von $f(x)$ verwendet werden. Es sei jedoch bemerkt, daß dies nur für die Klasse L^2 , und nicht etwa auch für die Klasse L^1 , zutrifft. Doch können auch bei den W. f. p. Funktionen der allgemeinen Klasse L^p ($p \geq 1$) Approximationssummen in einfacher Weise von der Fourierreihe der Funktion aus gebildet werden; in der Tat haben Besicovitch und ich gezeigt, daß die Fejér-Bochnersche Summationsmethode — die innerhalb der Theorie der eigentlichen f. p. Funktionen zu gleichmäßig approximierenden Summen führt — auch in der Theorie der W. f. p. Funktionen der

allgemeinen Klasse L^p Exponentialpolynome der erwünschten Art liefert, d. h. solche, die gegen $f(x)$ im Sinne des hier zugrunde gelegten Entfernungsbegriffes konvergieren.

7. Schließlich werde ich noch kurz die Resultate von Schmidt besprechen. Die Schmidtsche Abhandlung beginnt mit der Definition einer Klasse von verallgemeinerten f. p. Funktionen, die Schmidt als „Wellenfunktionen“ bezeichnet. Diese Definition hat große Ähnlichkeit mit derjenigen der W. f. p. Funktionen, und, wie Schmidt durch einfache Überlegungen zeigt, laufen die beiden Definitionen auch tatsächlich auf genau dasselbe hinaus. Bei der weiteren Behandlung seiner Funktionen benutzt Schmidt ihre Identität mit den W. f. p. Funktionen, indem er sich auf die oben erwähnte Übertragbarkeit der Weylschen Methoden von den eigentlichen f. p. Funktionen auf die W. f. p. Funktionen beruft. Das wesentlich neue, was die Schmidtsche Arbeit über die Weylsche hinaus bringt, besteht darin, daß gezeigt wird, daß die Folge von approximierenden Exponentialpolynomen $s_n(x)$, die in dem oben angegebenen Sinne zu einer beliebigen W. f. p. Funktion existiert, derart gewählt werden kann, daß sie eine gewisse besondere „Glattheit“ (Monotonieeigenschaft) besitzt. In genauer Formulierung besagt das Resultat von Schmidt, daß es möglich ist, die im Weylschen Sinne approximierende Folge $s_n(x)$ so zu wählen, daß, wenn wir ein festes $\varepsilon > 0$ betrachten (und nicht wie oben eine Folge von zu Null abnehmenden Größen $\varepsilon_1, \varepsilon_2, \dots$), wir dann auch für alle hinreichend großen n ein festes $L = L(\varepsilon)$ verwenden können. Es sind mit anderen Worten die Schmidtschen Approximationssummen $s_n(x)$ so beschaffen, daß es zu jedem ε ein $N = N(\varepsilon)$ und ein $L = L(\varepsilon)$ derart gibt, daß die Ungleichung

$$(3) \quad \text{Ob. Gr. } \frac{1}{L} \int_x^{x+L} |f(\xi) - s_n(\xi)|^2 d\xi < \varepsilon$$

für $n > N$ besteht. (Lassen wir ε gegen Null stehen, wird natürlich im allgemeinen L über alle Grenzen wachsen.)

Schmidt beweist diesen Satz, indem er von einer Folge $s_n(x)$ ausgeht, für welche $W\text{-}\lim s_n(x) = f(x)$ ist, und dann diese Folge einem gewissen scharfsinnig ausgedachten Glättungsverfahren (unter Heranziehung der Verschiebungszahlen) unterwirft. Von einer Folge $s_n(x)$, die im Sinne von (3) konvergiert, sagt Schmidt, daß sie „im Mittel scharf“ gegen $f(x)$ konvergiert. Ich wollte jedoch das Resultat von Schmidt lieber so interpretieren, daß es sich nicht so sehr um einen selbständigen Konvergenzprozeß handelt, als vielmehr um eine interessante — bei der betrachteten speziellen Approximationsaufgabe erfüllbare — Nebenbedingung, die man an eine im Weylschen Sinne konvergente Funktionenfolge stellen kann. Verglichen

z. B. mit den drei obigen Konvergenzbegriffen $S\text{-lim } f_n(x) = f(x)$, $B\text{-lim } f_n(x) = f(x)$, $W\text{-lim } f_n(x) = f(x)$, die ja alle drei verschiedene Varianten von Konvergenz im Mittel über ein unendliches Intervall darstellen, scheint nämlich das Schmidtsche Konvergenzverfahren nicht ein Begriff von derselben allgemeinen Bedeutung zu sein, weil es nicht — wie Schmidt selbst betont — mit einem Entfernungsbegriff im Funktionenraume verknüpft werden kann.

In den Untersuchungen von Besicovitch und mir über die Fourierreihen von W. f. p. Funktionen gewinnen wir das Schmidtsche Resultat — das uns von seiner Note in den Jahresberichten bekannt war — in anderer Weise, und zwar so, daß wir zugleich einen einfachen *Algorithmus* angeben, der von der Fourierreihe der Funktion aus zu einer Folge von Exponentialpolynomen der gewünschten Art führt. In der Tat ergibt sich, daß die oben erwähnte Fejér-Bochnersche Summationsmethode, auf die Fourierreihe einer W. f. p. Funktion (sogar der beliebigen Klasse L^p) ausgeübt, Approximationssummen liefert, die nicht nur im Weylschen Sinne konvergieren, sondern darüber hinaus gerade der von Schmidt gestellten Monotonieforderung genügen. Es ist interessant, zu sehen, daß nicht nur in der ursprünglichen Theorie, sondern auch in den verallgemeinerten Theorien der f. p. Funktionen, die Bochnersche Summationsmethode sich von ähnlicher Tragweite zeigt wie ihr Spezialfall, die Cesàro-Fejérsche Methode, in der Theorie der reinperiodischen Funktionen.

(Eingegangen am 7. 7. 1928.)

Untersuchungen über die Borelschen Verallgemeinerungen des Picardschen Satzes.

Von

Paul Csillag in Budapest.

Einleitung.

E. Borel betrachtet in einer bekannten Arbeit¹⁾ Identitäten von der Form $\sum_{i=1}^n a_i(z) e^{g_i(z)} = a_0(z)$, wo sämtliche Funktionen ganze transzendente Funktionen bedeuten und die $a_i(z)$ überdies „schwächer“ wachsen wie alle Quotienten $e^{g_i(z)} / e^{g_k(z)}$.

Sein — in strenger Weise durch O. Blumenthal²⁾ durchgeführter — Beweis zu dem Satze, daß unter diesen Voraussetzungen sämtliche $a_i(z)$ identisch verschwinden, stützt sich wesentlich auf Abschätzungen der Minima ganzer Funktionen auf gewissen Kreisen durch die Modulmaxima eben dieser Funktionen³⁾.

Ich habe gefunden, daß es nicht unzweckmäßig ist, eine andere Tatsache zum Beweise ähnlicher Sätze zu benutzen. Diese Tatsache betrifft die Abschätzung des Quotienten zweier im Einheitskreise regulären analytischen Funktionen auf jedem Kreise, die daselbst auch in bezug auf Vielfachheit dieselben Nullstellen besitzen und im Nullpunkt den Wert Eins haben.

Diese Abschätzung lautet:

$$\text{Max} \left| \frac{f_1(z)}{f_2(z)} \right|_{|z|=r_1} < \left[\text{Max} |f_1(z)|_{|z|=r_2} \text{Max} |f_2(z)|_{|z|=r_2} \right]^{\frac{2r_1}{r_2 - r_1}}$$
$$0 < r_1 < r_2 < 1.$$

1) Sur les zéros des fonctions entières, Acta mathematica 20.

2) Principes de la théorie des fonctions entières d'ordre infini. Paris 1910. Vergleiche außerdem: R. Nevanlinna, Einige Eindeutigkeitssätze usw., Acta mathematica 48. Hier wird ein spezieller Fall des Borelschen Satzes bewiesen. Ein Teil des Beweises ist mit den Rechnungen unseres § 2 nahe verwandt.

3) Vergleiche dazu: Valiron, Recherches sur le théorème de Picard etc., Annales de l'école normale sup. 39.

Der Beweis dieses Hilfssatzes (3.) gelingt leicht mit Hilfe einer Funktion, die die Herren C. Carathéodory und L. Fejér zum Beweise der Jensenschen Ungleichung benutzt haben⁴⁾.

Mit Hilfe dieses Hilfssatzes werden wir unter anderem folgende Verallgemeinerung des Borelschen Satzes beweisen:

Wir betrachten die ganzen transzendenten Funktionen der komplexen Variablen z :

$$g_1(z), g_2(z), \dots, g_n(z), \\ a_0(z), a_1(z), a_2(z), \dots, a_n(z);$$

und die Identität zwischen ihnen

$$a_1(z) e^{g_1(z)} + a_2(z) e^{g_2(z)} + \dots + a_n(z) e^{g_n(z)} = a_0(z);$$

ferner die reelle, monoton ins Unendliche wachsende Funktion der reellen Variablen $\varrho = |z|$

$$g^+(\varrho) = g_1^+(\varrho) + g_2^+(\varrho) + \dots + g_n^+(\varrho),$$

wo $g_i^+(\varrho)$ das Maximum des Realteiles von $g_i(z)$ auf dem Kreise $|z| = \varrho$ bedeutet.

Wir setzen voraus, daß die Funktionen $a_i(z)$ ($i = 1, 2, \dots, n$) durch $g^+(\varrho)$ abschätzbar sind⁵⁾, d. h. es gäbe eine positive Zahl $\delta < 1$, so daß für jedes positive ϱ und $\varepsilon \leq \varrho$

$$\log \sum_{i=0}^n \text{Max} |a_i(z)|_{|z|=\varrho} < \left[\frac{\varrho g^+(\varrho + \varepsilon)}{\varepsilon} \right]^\delta.$$

Die Behauptung unseres Hauptresultats ist, daß unter diesen Voraussetzungen konstante Zahlen c_1, c_2, \dots, c_n existieren, die nicht alle verschwinden, und für die in z identisch

$$\sum_{i=1}^n c_i a_i(z) e^{g_i(z)} = 0.$$

Wir verallgemeinern diesen Satz für Funktionen, die eine isolierte wesentlich singuläre Stelle haben (7.) und für Funktionen, die im Einheitskreis regulär sind und daselbst genügend schnell wachsen (8.). Wir deduzieren aus unserem Satze als Korollar den Satz des Herrn E. Borel (5.) und Verallgemeinerungen gewisser Sätze von Herrn G. Pólya (6.)⁶⁾. Wir

⁴⁾ Comptes Rendus 145 (1907).

⁵⁾ Dieser in (1.) näher ausgeführte Begriff der Abschätzbarkeit ist im wesentlichen mit den Borel-Blumenthalschen Begriffen identisch. Er ist aber vielleicht den aus den Cauchyschen Integralsätzen stammenden Abschätzungen leichter anzupassen.

⁶⁾ W. Saxer, „Über die Picardschen Ausnahmewerte sukzessiver Derivierten“. Ausführung Pólyascher Beweisideen. Math. Zeitschr. 17. Außerdem: W. Saxer, „Sur les valeurs exceptionnelles des dérivées successives des fonctions méromorphes“, Compt. Rend. 182 (1926).

hoffen binnen kurzem auf andere Anwendungen unserer Methode, insbesondere des Hilfssatzes (3.), an anderer Stelle zurückzukommen.

1. Funktionenfamilien, die zu einer monoton wachsenden Funktion gehören.

Wir untersuchen in dieser Arbeit hauptsächlich ganze transzendente Funktionen, und außerdem solche analytische Funktionen, die im Einheitskreise regulär sind.

Um beide Fälle zusammenzufassen, bezeichnen wir mit r die Einheit oder $+\infty$.

Wir bezeichnen außerdem mit

$$\varrho < r, \quad \varepsilon \leq \varrho, \quad \delta < 1$$

positive Zahlen, mit z eine komplexe Variable, mit $f(z)$ eine für $|z| < r$ regulär-analytische Funktion, mit $M(\varrho)$ eine für $0 < \varrho < r$ positive monoton wachsende Funktion, für die erstens

$$(1.1) \quad \lim_{\varrho=r} M(\varrho) = +\infty,$$

zweitens

$$(1.2) \quad \liminf_{\varrho=r} \frac{M(\varrho)}{\varrho} > 0$$

ist. [Die Bedingung (1.2) besagt (1.1) gegenüber nur dann etwas Neues, wenn $r = \infty$ ist.]

Wir sagen, daß $f(z)$ zu $M(\varrho)$ gehört oder durch $M(\varrho)$ abschätzbar ist, wenn für feste $\delta < 1$, $r > \varrho_0 > 0$ und für jedes $\varrho > \varrho_0$, $\varepsilon \leq \varrho$, $\varrho + \varepsilon < r$

$$(1.3) \quad \log \operatorname{Max}_{|z|=\varrho} |f(z)| \leq \left[\frac{\varrho M(\varrho + \varepsilon)}{\varepsilon} \right]^\delta.$$

Daß wir „ δ “ fest nennen, besagt genauer folgendes: δ ist eine nur von der Natur der Funktionen $f(z)$ und $M(\varrho)$ abhängender echter Bruch. Eine Aussage: aus der Abschätzbarkeit von $f(z)$ folgt die von $g(z)$, bedeutet also nur, daß aus der Ungleichheit (1.3) folgt für $\varrho > \varrho'_0$, $\varepsilon < \varrho$ bei einem $\delta' < 1$ (wo δ' vielleicht größer ist als δ)

$$\log \operatorname{Max}_{|z|=\varrho} |g(z)| \leq \left[\frac{\varrho M(\varrho + \varepsilon)}{\varepsilon} \right]^{\delta'}.$$

Dieser Begriff von Abschätzbarkeit hat folgende Eigenschaften:

- a) Eine Konstante gehört zu einer beliebigen $M(\varrho)$.
- b) Wenn $f_1(z)$ und $f_2(z)$ zu $M(\varrho)$ gehören, so gehören $f_1(z) + f_2(z)$, $f_1(z)f_2(z)$ ebenfalls zu $M(\varrho)$.

c) Aus a) und b) folgt: Wenn $f_1(z), f_2(z), \dots, f_n(z)$ zu $M(\varrho)$ gehören, so gehört jede ganze rationale Funktion von diesen Funktionen ebenfalls zu $M(\varrho)$.

d) Wenn $f(z)$ zu $M(\varrho)$ gehört, so gehört auch $f'(z)$ zu $M(\varrho)$.

Beweis.

$$\begin{aligned} \text{Max } |f'(z)|_{|z|=e} &= \frac{1}{2\pi} \text{Max} \left| \int_{|t|=e+\frac{\varepsilon}{2}} \frac{f(t) dt}{(t-z)^2} \right|_{|z|=e} \\ &\leq \frac{4\left(\varrho + \frac{\varepsilon}{2}\right)}{\varepsilon^2} \text{Max } |f(z)|_{|z|=e+\frac{\varepsilon}{2}} \end{aligned}$$

und weiter nach (1.3), wenn wir statt ϱ und ε $\varrho + \frac{\varepsilon}{2}$ und $\frac{\varepsilon}{2}$ setzen,

$$\text{Max } |f'(z)|_{|z|=e} \leq \frac{4\left(\varrho + \frac{\varepsilon}{2}\right)}{\varepsilon^2} e^{\left[\frac{2\left(\varrho + \frac{\varepsilon}{2}\right) M(\varrho + \frac{\varepsilon}{2})}{\varepsilon} \right]^\delta}$$

folglich wegen $\varrho > \frac{1}{2}$ und $\frac{\varrho}{\varepsilon} > 1$ für ein gewisses positives \varkappa

$$< 6\left(\frac{\varrho}{\varepsilon}\right)^2 e^{\left[\frac{3\varrho M(\varrho + \varepsilon)}{\varepsilon} \right]^\delta} < e^{\varkappa \left(\frac{\varrho}{\varepsilon}\right)^\delta + 3\delta \left[\frac{\varrho M(\varrho + \varepsilon)}{\varepsilon} \right]^\delta}$$

und weiter wegen (1.1) für ein geeignetes positives \varkappa_1

$$< e^{\varkappa_1 \left[\frac{\varrho M(\varrho + \varepsilon)}{\varepsilon} \right]^\delta}$$

für genügend große ϱ wegen (1.1)

$$< e^{\left[\frac{\varrho M(\varrho + \varepsilon)}{\varepsilon} \right]^\delta}, \quad \delta_1 = \delta + (1 - \delta)\delta < 1.$$

e) Wenn $f(z)$ zu $M(\varrho)$ gehört, so gehören nach d) alle Derivierten von $f(z)$ zu $M(\varrho)$.

f) Aus c) und e) folgt: Wenn die Funktionen $f_1(z), f_2(z), \dots, f_n(z)$ zu $M(\varrho)$ gehören, so ist jedes Polynom, welches aus diesen Funktionen und aus ihren Derivierten gebildet ist, durch $M(\varrho)$ abschätzbar.

g) Wenn $f^+(\varrho)$ das Maximum des Realteiles von $f(z)$ auf dem Kreise $|z| = \varrho$ bezeichnet und $f(z)$ von solcher Beschaffenheit ist, daß $\lim_{\varrho \rightarrow r} f^+(\varrho) = \infty$, so ist $f(z)$ durch $M(\varrho) = f^+(\varrho)$ abschätzbar, und zwar für ein beliebig kleines δ .

h) Aus den Beweisen erhellt, daß die Eigenschaften a) ... g) der Abschätzbarkeit bestehen bleiben, wenn wir den Begriff dahin modifizieren,

daß wir (1.3) nicht für eine feste Zahl $\delta < 1$, sondern für beliebig kleine δ postulieren.

i) Wenn $M_1(\varrho) < M_2(\varrho)$ für $0 < \varrho < r$, so ist jede Funktion, die durch $M_1(\varrho)$ abschätzbar ist, auch durch $M_2(\varrho)$ abschätzbar.

2. Die linearen Identitäten von der Form:

$$a_1 e^{g_1} + a_2 e^{g_2} + \dots + a_n e^{g_n} = a.$$

Wir betrachten die für $|z| < r$ regulär-analytischen Funktionen

$$(2.1) \quad a_0(z), a_1(z), a_2(z), \dots, a_n(z),$$

$$(2.2) \quad g_1(z), g_2(z), \dots, g_n(z).$$

Wir betrachten außerdem die reelle Funktion

$$(2.3) \quad g(\varrho) = g_1^+(\varrho) + g_2^+(\varrho) + \dots + g_n^+(\varrho).$$

$g_i^+(\varrho)$ bedeutet wie in (1.) das Maximum des Realteiles von $g_i(z)$.

Wir setzen voraus, daß die Funktionen $a_i(z)$ zu $g(\varrho)$ gehören, und daß die Funktionen (2.1), (2.2) der Identität

$$(2.4) \quad a_1(z) e^{g_1(z)} + a_2(z) e^{g_2(z)} + \dots + a_n(z) e^{g_n(z)} = a_0(z)$$

genügen.

Differenzieren wir (2.4) $n - 1$ -mal, so bekommen wir $n - 1$ weitere Identitäten

$$(2.5) \quad \sum_{i=1}^n a_i^{(i)}(z) e^{g_i(z)} = a_0^{(i)}(z) \quad (i = 1, 2, \dots, n - 1),$$

wo die $a_i^{(i)}(z)$ Polynome sind, in denen als Variable die Funktionen (2.1), (2.2) und ihre Derivierten figurieren.

Da die Funktionen (2.1) laut Voraussetzung, die Funktionen (2.2) laut (1.) Eigenschaft g) zu $g(\varrho)$ gehören, so gehören die Funktionen $a_i^{(i)}(z)$ laut (1.) Eigenschaft f) ebenfalls zu $g(\varrho)$.

Eliminieren wir aus den n Identitäten (2.4), (2.5) $n - 1$ der Funktionen $g_i(z)$, so bekommen wir weitere n Identitäten

$$(2.6) \quad A_0(z) e^{g_1(z)} = A_1(z),$$

wo die Funktionen $A_0(z), A_1(z)$ aus den Elementen $a_i^{(i)}(z_1)$ gebildete Determinanten bedeuten. Sie gehören also ebenfalls zu $g(\varrho)$.

Im speziellen ist

$$A_0(z) = |a_i^{(i)}(z)| = \left| \frac{d^i [a_1(z) e^{g_1(z)}]}{dz^i} e^{-g_1(z)} \right| = e^{-g_1(z)} \dots e^{-g_n(z)} \left| \frac{d^i [a_1(z) e^{g_1(z)}]}{dz^i} \right|.$$

Daraus folgt, daß, wenn $A_0(z)$ identisch verschwindet, so verschwindet auch identisch die Wronskische Determinante

$$\left| \frac{d^i [a_i(z) e^{g_i(z)}]}{dz^i} \right|.$$

In diesem Falle existiert die lineare identische Relation von der Form

$$(2.7) \quad c_1 a_1(z) e^{g_1(z)} + \dots + c_n a_n(z) e^{g_n(z)} = 0,$$

wo die komplexen Konstanten c_i nicht alle Null sind, was wir auch so ausdrücken können:

$$|c_1| + |c_2| + \dots + |c_n| > 0.$$

Wir beweisen als Hauptsatz das Bestehen einer Relation von der Form (2.7). Zu diesem Zwecke setzen wir immer voraus, daß $A_0(z)$ nicht identisch verschwindet und leiten daraus einen Widerspruch her.

3. Hilfssatz über die Abschätzung der Quotienten zweier Funktionen, die dieselben Nullstellen besitzen.

Hilfssatz. Wir betrachten zwei im Kreise $|z| < r$ regulär analytische Funktionen mit den Potenzreihenentwicklungen um den Nullpunkt

$$(3.1) \quad f_1(z_1) = z^k + \dots$$

$$(3.2) \quad f_2(z) = z^k + \dots$$

Die beiden Funktionen sollen im Kreise $|z| < r$ auch in bezug auf Vielfachheit dieselben Nullstellen besitzen. Wenn

$$(3.3) \quad \text{Max} |f_1(z)|_{|z|=\varrho} = M_1(\varrho)$$

$$(3.4) \quad \text{Max} |f_2(z)|_{|z|=\varrho} = M_2(\varrho)$$

und $\varrho + \varepsilon < r$, so gilt die Abschätzung:

$$(3.5) \quad \text{Max} \left| \frac{f_1(z_1)}{f_2(z)} \right|_{|z|=\varrho} < \left[\frac{M_1(\varrho + \varepsilon) M_2(\varrho + \varepsilon)}{(\varrho + \varepsilon)^{2k}} \right]^{\frac{2\varrho}{\varepsilon}}$$

Beweis. Wir bezeichnen die gemeinsamen von Null verschiedenen Nullstellen von $f_1(z), f_2(z)$, die absolut kleiner sind als $\varrho + \varepsilon$, mit

$$z_1, z_2, \dots, z_l.$$

Wir setzen

$$z_i^* = \frac{(\varrho + \varepsilon)^2}{z_i} \quad (i = 1, 2, \dots, l)$$

$$(3.6) \quad \lambda(z) = \frac{z_1^* z_2^* \dots z_l^*}{z_1 z_2 \dots z_l} \frac{z - z_1}{z - z_1^*} \frac{z - z_2}{z - z_2^*} \dots \frac{z - z_l}{z - z_l^*}$$

$\lambda(z)$ ist regulär für $|z| < \varrho + \varepsilon$, außerdem ist

$$(3.7) \quad \lambda(0) = 1$$

$$(3.8) \quad |\lambda(z)|_{|z|=\varrho+\varepsilon} = \text{konstant} \geq 1.$$

Die Funktionen $\frac{f_1(z)}{z^k \lambda(z)}$, $\frac{f_2(z)}{z^k \lambda(z)}$ haben im Kreise $|z| < \rho + \varepsilon$ keine Nullstellen, so daß wir für $|z| < \rho + \varepsilon$ setzen können

$$(3.9) \quad \frac{f_1(z)}{z^k \lambda(z)} = e^{\varphi_1(z)}, \quad \frac{f_2(z)}{z^k \lambda(z)} = e^{\varphi_2(z)},$$

wo $\varphi_1(z)$ und $\varphi_2(z)$ für $|z| < \rho + \varepsilon$ regulär sind. Aus (3.8) und (3.9) folgt, daß

$$(3.10) \quad e^{\varphi_1^+(\rho+\varepsilon)} \leq \frac{M_1(\rho+\varepsilon)}{(\rho+\varepsilon)^k}, \quad e^{\varphi_2^+(\rho+\varepsilon)} \leq \frac{M_2(\rho+\varepsilon)}{(\rho+\varepsilon)^k}.$$

Aus 3.9 und aus einer wohlbekannten Hadamard-Borel-Carathéodoryschen Ungleichung folgt weiter, daß

$$(3.11) \quad \begin{aligned} \text{Max} \left| \frac{f_1(z)}{f_2(z)} \right|_{|z|=\rho} &= \text{Max}_{|z|=\rho} |e^{\varphi_1(z) - \varphi_2(z)}| = \text{Max}_{|z|=\rho} [e^{\Re[\varphi_1(z) - \varphi_2(z)]}] \\ &\leq e^{\text{Max}_{|z|=\rho} \Re \varphi_1(z)} + \text{Max}_{|z|=\rho} \Re \varphi_2(z) \leq e^{2 \frac{\rho}{\varepsilon} [\varphi_1^+(\rho+\varepsilon) + \varphi_2^+(\rho+\varepsilon)]}. \end{aligned}$$

Wir verbinden (3.10) und (3.11) zu

$$\text{Max} \left| \frac{f_1(z)}{f_2(z)} \right|_{|z|=\rho} < \left[\frac{M_1(\rho+\varepsilon) M_2(\rho+\varepsilon)}{(\rho+\varepsilon)^{2k}} \right]^{\frac{2\rho}{\varepsilon}},$$

w. z. b. w.

4. Der Hauptsatz über ganze transzendente Funktionen.

Satz I. *Wir bilden aus den ganzen transzendenten Funktionen*

$$g_1(z), g_2(z), \dots, g_n(z)$$

die reelle monotone Funktion

$$g(\rho) = g_1^+(\rho) + g_2^+(\rho) + \dots + g_n^+(\rho).$$

Wenn die ganzen transzendenten Funktionen

$$a_0(z), a_1(z), \dots, a_n(z)$$

durch $g(\rho)$ abschätzbar sind, so folgt aus der Identität

$$a_1(z) e^{g_1(z)} + a_2(z) e^{g_2(z)} + \dots + a_n(z) e^{g_n(z)} = a_0(z)$$

die weitere Identität

$$c_1 a_1(z) e^{g_1(z)} + c_2 a_2(z) e^{g_2(z)} + \dots + c_n a_n(z) e^{g_n(z)} = 0,$$

wo die Konstanten c_1, c_2, \dots, c_n nicht alle verschwinden.

Um Satz I zu beweisen, haben wir nach 2. zu zeigen, daß $A_0(z)$ (in Formel (2.6)) identisch verschwindet.

Die Funktion $g(\varrho)$ entspricht den Bedingungen (1. 1), (1. 2), also können wir nach 1. und 2. behaupten, daß für ein gewisses $\delta < 1$

$$(4. 1) \quad |A_i(z)|_{|z|=\varrho} < e^{\left[\frac{\varrho g(\varrho+\varepsilon)}{\varepsilon}\right]^\delta} \quad (i = 0, 1, 2, \dots, n)$$

Aus (2. 6) folgt weiter, daß, wenn $A_0(z)$ nicht identisch verschwindet, sämtliche Funktionen $A_i(z)$, auch in bezug auf Vielfachheit, dieselben Nullstellen besitzen.

Nach dem Hilfssatz in 3. folgt die Formeln (2. 6), (3. 5), (4. 1) in Betracht gezogen, daß für genügend große ϱ

$$(4. 2) \quad e^{g_i^+(\varrho)} = \text{Max} |e^{g_i(z)}|_{|z|=\varrho} = \text{Max} \left| \frac{A_i(z)}{A_0(z)} \right|_{|z|=\varrho} \\ < \frac{|A_i(0)|}{|A_0(0)|} \left[\frac{\text{Max} |A_0(z)|_{|z|=\varrho+\varepsilon} \text{Max} |A_i(z)|_{|z|=\varrho+\varepsilon}}{|A_0(0)| |A_i(0)| (\varrho+\varepsilon)^{2k}} \right]^{\frac{2\varrho}{\varepsilon}} < e^{12 \frac{\varrho+\varepsilon}{\varepsilon} \left[\frac{(\varrho+\varepsilon)g(\varrho+2\varepsilon)}{\varepsilon} \right]^\delta} \\ (i = 1, 2, \dots, n).$$

Wenn $A_0(0)$ und damit $A_i(0)$ gleich Null sind, so verstehen wir unter $A_0(0)$ bzw. $A_i(0)$ die ersten nicht verschwindenden Koeffizienten der Potenzreihe für die Funktionen $A_0(z)$, $A_i(z)$.

Wir schreiben Anfang und Ende von 4. 2 zusammenfassend

$$(4. 3) \quad g_i^+(\varrho) < 12 \frac{\varrho+\varepsilon}{\varepsilon} \left[\frac{(\varrho+\varepsilon)g(\varrho+2\varepsilon)}{\varepsilon} \right]^\delta \quad (i = 1, 2, \dots, n).$$

Jetzt unterscheiden wir zwei Fälle:

Fall A. Alle Funktionen $g_i(z)$ sind Polynome. h bedeute den Höchstgrad dieser Polynome. Aus (4. 3) folgt für $\varepsilon = \varrho$, daß für eine gewisse Konstante K

$$g_i^+(\varrho) < K \varrho^{h\delta} \quad (i = 1, 2, \dots, n).$$

Das ist ein Widerspruch, da wenigstens eines der Polynome $g_i(z)$ vom Grade h ist. $A_0(z)$ verschwindet also identisch.

Fall B. Wenigstens eine der ganzen Funktionen $g_i(z)$ ist eine ganze transzendente Funktion und kein Polynom. In diesem Falle wächst $g(\varrho)$ schneller als irgendeine Potenz von ϱ .

Nach einem bekannten Borelschen Satz⁷⁾ existieren beliebig große Zahlen ϱ , so daß für ein beliebig kleines ω

$$(4. 4) \quad g \left[\varrho + \frac{1}{\log g(\varrho)} \right] < g(\varrho)^{1+\omega}.$$

⁷⁾ Siehe Borel und Blumenthal, loc. cit.

Setzen wir in (4.3) $2\varepsilon = \frac{1}{\log g(\varrho)}$, so folgt daß

$$(4.5) \quad g_i^+(\varrho) < 24 \cdot 2^{1+\delta} \varrho^{1+\delta} g(\varrho)^{(1+\omega)\delta} [\log g(\varrho)]^{1+\delta} \quad (i = 1, 2, \dots, n)$$

und, weil für genügend große ϱ

$$24 \cdot 2^{1+\delta} \varrho^{1+\delta} [\log g(\varrho)]^{1+\delta} \cdot n < g(\varrho)^{\frac{1-\delta}{2}},$$

folgt weiter aus (4.5) für $\omega = \frac{1-\delta}{2\delta}$ und genügend große ϱ , daß

$$(4.7) \quad g_i^+(\varrho) < \frac{g(\varrho)^{(1+\omega)\delta} g(\varrho)^{\frac{1-\delta}{2}}}{n} = \frac{g(\varrho)^{\frac{1+\delta}{2} + \frac{1-\delta}{2}}}{n} = \frac{g(\varrho)}{n} \quad (i = 1, 2, \dots, n).$$

Wir addieren alle n Ungleichungen (4.7) und erinnern uns an die Definition von $g(\varrho)$. Es folgt in dieser Weise, daß

$$g(\varrho) = \sum_{i=1}^n g_i^+(\varrho) < g(\varrho).$$

Dies ist ein Widerspruch. $A_0(z)$ verschwindet also auch in diesem Falle identisch.

5. Der Borelsche Satz.

Aus Satz I (4.) folgern wir einige Sätze als Korollar. Zuerst wollen wir den Spezialfall von Satz I für $n = 1$ als Satz II in einer etwas veränderten Form noch einmal aussprechen.

Satz II. *Wenn die ganzen Funktionen $a(z), a_0(z)$ durch $g^+(\varrho)$ abschätzbar sind, wo $g^+(\varrho)$ das Maximum des Realteiles von einer ganzen Funktion $g(z)$ bedeutet, so folgt aus der Identität*

$$a(z) e^{g(z)} = a_0(z),$$

daß $a_0(z)$ und $a(z)$ identisch verschwinden.

Beweis. Aus Satz I folgt nämlich, daß für eine von Null verschiedene Konstante c in z identisch

$$c a(z) e^{g(z)} = 0,$$

d. h. es ist identisch

$$a(z) e^{g(z)} = 0,$$

w. z. b. w.

Satz III. *Aus der Identität zwischen ganzen Funktionen*

$$(5.1) \quad a_1(z) e^{g_1(z)} + a_2(z) e^{g_2(z)} + \dots + a_n(z) e^{g_n(z)} = a_0(z)$$

folgt, daß $a_0(z)$ identisch verschwindet unter der Bedingung, daß die ganzen Funktionen

$$(5.2) \quad a_0(z), a_1(z), \dots, a_n(z)$$

durch jede der Funktionen

$$(5.3) \quad g_1^+(\varrho), g_2^+(\varrho), \dots, g_n^+(\varrho)$$

abschätzbar sind.

Beweis. Für $n = 1$ ist der Satz mit Satz I gleichbedeutend. Allgemein beweisen wir den Satz durch vollständige Induktion.

Aus der Annahme folgt nach 1. Eigenschaft i), wie

$$g_i^+(\varrho) \leq \sum_{i=1}^n g_i^+(\varrho) = g(\varrho) \quad (i = 1, 2, \dots, n),$$

daß alle Funktionen (5.2) durch $g(\varrho)$ abschätzbar sind. Nach Satz I existieren also Konstante c_i , von denen z. B. c_1 von Null verschieden ist und für die

$$(5.4) \quad a_1(z) e^{g_1(z)} + \frac{c_2}{c_1} a_2(z) e^{g_2(z)} + \dots + \frac{c_n}{c_1} a_n(z) e^{g_n(z)} = 0.$$

Subtrahieren wir (5.4) von (5.1), so haben wir die weitere Identität

$$\left(1 - \frac{c_2}{c_1}\right) a_2(z) e^{g_2(z)} + \dots + \left(1 - \frac{c_n}{c_1}\right) a_n(z) e^{g_n(z)} = a_0(z)$$

von der Form (5.1), in der nur $n-1$ Funktionen figurieren und die Abschätzbarkeitsannahmen auch erfüllt sind.

Satz IV. Wenn außer den Voraussetzungen des Satzes III wir die weiteren Annahmen machen, daß die Funktionen (5.2) auch durch alle reellen Funktionen

$$[g_i(z) - g_l(z)]_{(z)=\varrho}^+ \quad (i, l = 1, 2, \dots, n; i > l)$$

abschätzbar sind, so folgt aus der Identität (5.1), daß alle Funktionen (5.2) identisch verschwinden.

Beweis. Nach Satz III verschwindet $a_0(z)$ identisch.

Also es ist identisch (nach (5.1))

$$a_2(z) e^{g_2(z) - g_1(z)} + \dots + a_n(z) e^{g_n(z) - g_1(z)} = -a_1(z),$$

woraus nach Satz III und nach unseren Voraussetzungen folgt, daß $a_1(z)$ identisch verschwindet. Daraus erhellt unser Satz durch vollständige Induktion.

Satz IV ist in einer wenig veränderten Fassung der Satz von E. Borel.

Satz V. Wenn $n = 2$, so sind die im Satz I figurierenden Konstanten c_1, c_2 beide von Null verschieden oder eine von den Funktionen $a_1(z), a_2(z)$ verschwindet identisch.

Der Beweis ergibt aus der nach Satz I folgenden Identität

$$c_1 a_1(z) e^{g_1(z)} = -c_2 a_2(z) e^{g_2(z)}.$$

6. Anwendungen der Sätze auf den Picardschen Satz und auf gewisse Pólya-Saxersche Sätze.

Wir beweisen jetzt aus unserem Satze den Satz von E. Picard in seiner einfachsten Form:

Eine Identität zwischen nichtkonstanten ganzen Funktionen $g_1(z)$, $g_2(z)$ wie

$$e^{g_1(z)} + e^{g_2(z)} = 1$$

ist unmöglich.

Zum Beweise setzen wir $a_1(z) = 1$, $a_2(z) = 1$ und folgern aus Satz III (da 1 durch eine beliebige $g(z)$ abschätzbar ist) daß $1 = 0$., Das ist offenbar unmöglich.

Wir beweisen jetzt als eine Verallgemeinerung eines Satzes von Pólya-Saxer⁸⁾ den

Satz VI. *Wenn die ganze transzendente Funktion $f(z)$ keine Nullstellen besitzt, so nehmen alle Derivierten von $f(z) \dots f^n(z)$ außer eventuell 0 sämtliche Werte an.*

Laut Voraussetzung schreiben wir

$$(6.1) \quad f(z) = e^{g(z)},$$

wo $g(z)$ eine ganze Funktion bedeutet.

Daraus folgt durch fortgesetztes Differenzieren

$$f^n(z) = P(g'(z), g''(z), \dots, g^n(z)) e^{g(z)} = P(z) e^{g(z)},$$

wo $P(g'(z), g''(z), \dots, g^n(z))$ ein Polynom von g', g'', \dots, g^n bedeutet. $P(z)$ ist also ganz und durch $g^+(\rho)$ abschätzbar (1. Eigenschaft f).

Hat $f^n(z)$ einen Ausnahmewert d , so ist ($g_1(z)$ ganz)

$$(6.2) \quad f^n(z) = P(z) e^{g(z)} = e^{g_1(z)} + d.$$

Wenn $P(z)$ nicht identisch verschwindet, so existieren nach Satz I und VI zwei von Null verschiedene Konstanten c_1, c_2 , so daß

$$(6.3) \quad c_1 P(z) e^{g(z)} + c_2 e^{g_1(z)} = 0.$$

Aus (6.2) und (6.3) folgt, daß

$$\left(1 + \frac{c_1}{c_2}\right) P(z) e^{g(z)} = d$$

und daraus nach Satz II, daß $d = 0$.

$P(z)$ aber kann nicht identisch verschwinden, wenn nur $f(z)$ transzendent ist.

⁸⁾ W. Saxer, loc. cit.

Satz VII. (Von Pólya-Saxer.)⁸⁾ Wenn die aus der ganzen Funktion $f(z)$ gebildete ganze Funktion $f(z)f'(z)f''(z)$ keine Nullstellen besitzt, so ist $f(z)$ von der Form

$$f(z) = e^{\mu z + \nu},$$

wo μ und ν zwei Konstanten bedeuten.

Beweis. Laut Voraussetzung haben $f(z)$, $f'(z)$, $f''(z)$ keine Nullstellen. Es gibt also ganze Funktionen $g(z)$, $h(z)$, $k(z)$, für die

$$(6.4) \quad \begin{aligned} f(z) &= e^{g(z)}, \\ f'(z) &= g'(z) e^{g(z)} = e^{h(z)+g(z)}, \quad g'(z) = e^{h(z)}, \\ f''(z) &= [h'(z) + g'(z)] e^{g(z)+h(z)} = e^{g(z)+h(z)+k(z)}, \\ h'(z) + e^{h(z)} &= e^{k(z)}. \end{aligned}$$

Aus der letzten Identität folgt nach den Sätzen I und V für gewisse von Null verschiedene Konstanten c_1 , c_2

$$(6.5) \quad c_1 e^{h(z)} + c_2 e^{k(z)} = 0$$

und aus (6.4) und (6.5) folgt weiter, daß

$$e^{h(z)} \left(1 + \frac{c_1}{c_2} \right) = -h'(z).$$

Nach Satz II verschwindet $h'(z)$ identisch, $h(z)$, damit auch $g'(z)$, ist also eine Konstante $g(z)$ eine lineare Funktion von z und $f(z) = e^{g(z)} = e^{\mu z + \nu}$, w. z. b. w.

Wir haben an speziellen Fällen erkannt, aber leider nicht allgemein beweisen können, daß aus $f(z)f^{(i)}(z)f^{(j)}(z) \neq 0$ folgt $f(z) = e^{\mu z + \nu}$ für beliebige ganze i, j .

7. Funktionen, die eine isolierte wesentlich singuläre Stelle haben.

Wir betrachten jetzt die Identitäten von der Form (2.4) zwischen Funktionen, die im Unendlichen eine wesentlich singuläre Stelle haben.

Die Eigenschaften der Abschätzbarkeit in 1. sind auf diese Klasse von Funktionen leicht zu übertragen.

Beim Beweise von d) müssen wir die Integration auf zwei Kreisen vornehmen, von denen der eine von ρ unabhängig ist und der andere den Radius $\rho + \frac{\varepsilon}{z}$ besitzt.

Wir brauchen im folgenden noch zwei Eigenschaften der Abschätzbarkeit, und zwar:

j) Sei K eine Konstante und gehöre $f(z)$ zu $M(\rho)$, so gehört auch $f(z)$ zu $M(\rho) \pm K$.

k) Wenn $f_1(z)$ im Unendlichen eine wesentlich singuläre Stelle hat und zu $M(\rho)$ gehört, so gehört auch $f_1(z)$, $f_2(z)$ zu $M(\rho)$, wenn nur $f_2(z)$

im Unendlichen sich bestimmt verhält, also entweder dort regulär ist oder aber einen Pol hat.

Dies ist klar aus den folgenden Ungleichungen

$$|f_1(z)f_2(z)| < e^{\left[\frac{\rho M(\rho+\varepsilon)}{\varepsilon}\right]^\delta} \rho^k < e^{\left[\frac{\rho M(\rho+\varepsilon)}{\varepsilon}\right]^\delta} (\rho + \varepsilon)^k < e^{2\left[\frac{\rho M(\rho+\varepsilon)}{\varepsilon}\right]^\delta}.$$

Um Satz I jetzt auf unsere neue Funktionsklasse auszudehnen, benutzen wir zwei bekannte Hilfssätze, Korollare des Laurentschen Satzes über Potenzreihenentwicklung einer analytischen Funktion im Äußeren eines Kreises.

Hilfssatz A. Wenn $f(z)$ im Unendlichen einen wesentlich singulären Punkt besitzt, so ist

$$f(z) = f_1(z) + f_2(z),$$

wo $f_1(z)$ eine ganze transzendente Funktion bedeutet und $f_2(z)$ im Unendlichen regulär ist.

Hilfssatz B. Wenn $f(z)$ im Unendlichen einen isolierten wesentlich singulären Punkt besitzt, so ist

$$f(z) = f_1(z) f_2(z),$$

wo $f_1(z)$ eine ganze transzendente Funktion bedeutet und $f_2(z)$ im Unendlichen regulär ist.

Als Anwendung dieser Hilfssätze auf die in (2.6) auftretenden Funktionen ergibt sich

$$(7.1) \quad \begin{cases} g_i(z) = g_{i,1}(z) + g_{i,2}(z) & (i = 1, 2, \dots, n), \\ A_i(z) = A_{i,1}(z) A_{i,2}(z) & (i = 0, 1, 2, \dots, n). \end{cases}$$

Die jetzt im Äußeren eines gewissen Kreises gültigen Identitäten (2.6)

$$(7.2) \quad A_0(z) e^{g_i(z)} = A_i(z) \quad (i = 1, 2, \dots, n)$$

schreiben sich jetzt als

$$(7.3) \quad e^{g_{i,1}(z)} = \frac{A_{i,1}(z)}{A_{0,1}(z)} \frac{A_{i,2}(z)}{A_{0,2}(z)} e^{-g_{i,2}(z)} = \frac{A_{i,1}(z)}{A_{0,1}(z)} B_i(z) \quad (i = 1, 2, \dots, n).$$

Die Funktionen $B_i(z)$ verhalten sich im Unendlichen bestimmt.

Die Funktionen $A_{i,1}(z)$ haben, wenn sie nicht identisch verschwinden nach (7.2) im Äußeren eines festen Kreises R , dieselben Nullstellen. Wenn also $p_i(z)$ geeignete Polynome sind, so können wir setzen

$$A_{i,1}(z) = A_{i,1}^*(z) p_i(z) \quad (i = 1, 2, \dots, n),$$

wo die ganzen Funktionen $A_{i,1}^*(z)$ gemeinsame Nullstellen haben; und statt (7.3) schreiben

$$(7.4) \quad e^{g_{i,1}(z)} = \frac{A_{i,1}^*(z)}{A_{0,1}^*(z)} \frac{p_i(z)}{p_0(z)} B_i(z) = \frac{A_{i,1}^*(z)}{A_{0,1}^*(z)} B_i^*(z),$$

wo die Funktionen $B_i^*(z)$ sich im Unendlichen bestimmt verhalten.

Die ganzen Funktionen $A_{i,1}^*(z)$ gehören nach den am Anfang dieses Paragraphen bewiesenen Eigenschaften der Abschätzbarkeit i) und k) zu $g(\varrho)$ und, weil für eine gewisse Konstante K gewiß

$$g_{i,1}^+(\varrho) < g_i^+(\varrho) + \frac{K}{n},$$

$$g^*(\varrho) = \sum_{i=1}^n g_{i,1}(\varrho) < \sum_{i=1}^n g_i^+(\varrho) + K < g(\varrho) + K$$

zu $g^*(\varrho)$.

Ebenso wie wir aus (2.6) (4.3) gefolgert haben, folgern wir aus (7.4) für eine gewisse positive k , daß

$$(7.5) \quad e^{g_{i,1}^+(\varrho)} < \varrho^k e^{12 \frac{\varrho+\varepsilon}{\varepsilon} \left[\frac{(\varrho+\varepsilon)g^*(\varrho+\varepsilon)}{\varepsilon} \right]^\delta}.$$

Für hinreichend große ϱ ist

$$\varrho^k < e^{[g^*(\varrho+\varepsilon)]^\delta} < e^{\frac{\varrho+\varepsilon}{\varepsilon} \left[\frac{(\varrho+\varepsilon)g^*(\varrho+\varepsilon)}{\varepsilon} \right]^\delta}.$$

Daraus folgt

$$(7.6) \quad e^{g_{i,1}^+(\varrho)} < e^{13 \frac{\varrho+\varepsilon}{\varepsilon} \left[\frac{(\varrho+\varepsilon)g^*(\varrho+\varepsilon)}{\varepsilon} \right]^\delta}.$$

Das ist eine Ungleichung von der Form (4.3), nur der Koeffizient 12 ist mit 13 vertauscht. Wir folgern also aus (7.6) ebenso wie aus (4.3), daß $A_{0,1}^*(z)$ das heißt, das $A_0(z)$ identisch verschwindet.

Damit haben wir den folgenden Satz bewiesen.

Satz I*. *Satz I bleibt gültig, wenn die $2n+1$ Funktionen $a_i(z), g_i(z)$ keine ganzen Funktionen bedeuten, sondern nur solche, die außerhalb eines Kreises regulär sind.*

Natürlich bleiben alle Folgerungen aus Satz I, insbesondere die Sätze II bis VII, für diese Funktionenklasse gültig.

8. Funktionen, die im Einheitskreise regulär sind.

Wir können jetzt unsere Ergebnisse verallgemeinern auf Funktionen, die in einem Kreisringe regulär sind. Wir beschränken uns aber auf Funktionen, die im Einheitskreise regulär sind und dort genügend schnell wachsen.

Wir haben jetzt die Theorie von 1. mit $r=1$ und betrachten die Identität (2.4) mit der Bedingung, daß die Funktionen (2.1), (2.2), (2.3) alle im Einheitskreise regulär sind.

Aus den Identitäten (2.6) gewinnen wir wie in 4. (4.3)

$$e^{g_i^+(\varrho)} \leq e^{12 \frac{\varrho+\varepsilon}{\varepsilon} \left[\frac{(\varrho+\varepsilon)g(\varrho+\varepsilon)}{\varepsilon} \right]^\delta}.$$

Wenn wir für ϱ und ε nur Werte in Betracht ziehen, für welche $\varrho > \frac{1}{2}$ $\varepsilon \leq \varrho$, so folgen für $\delta' = \delta(2 - \delta) = \delta + \delta(1 - \delta) < 1$ die weiteren Ungleichheiten

$$(8.1) \quad g_i^+(\varrho) < \left[\frac{g(\varrho + \varepsilon)}{\varepsilon} \right]^{\delta'} \frac{1}{\varepsilon}.$$

Dies gilt natürlich nur unter der Bedingung, daß $A_0(z)$ nicht identisch verschwindet und $\lim_{\varrho=1} g(\varrho) = +\infty$.

Dies vorausgesetzt präzisieren wir die Worte „ $g(\varrho)$ nächst genügend schnell“ auf zweierlei Weise und erhalten zwei Sätze als Verallgemeinerungen von Satz I.

Im ersten Falle verlangen wir von den $g_i(z)$ weniger und von den $a_i(z)$ mehr, im zweiten umgekehrt von $a_i(z)$ weniger und von $g_i(z)$ mehr.

Zuerst formulieren wir einen bekannten Hilfssatz nach Borel⁹⁾, den wir beim Beweise benutzen:

Hilfssatz A. Sei $U(\varrho)$ monoton wachsend für $\varrho < 1$ und $\lim_{\varrho=1} U(\varrho) = +\infty$. Wenn $\varrho' < 1$ und $\varrho' + \frac{3}{U(\varrho')} < 1$, so gibt es ein ϱ , für welche erstens

$$U\left(\varrho + \frac{1}{U(\varrho)}\right) < 2U(\varrho),$$

zweitens

$$\varrho' < \varrho < \varrho' + \frac{3}{U(\varrho')} < 1.$$

Daraus folgern wir Hilfssatz B und C.

Hilfssatz B. Sei $g(\varrho)$ monoton wachsend für $\varrho < 1$ und für ein beliebig großes Ω $\lim_{\varrho=1} g(\varrho)(1 - \varrho)^\Omega = +\infty$. Unter diesen Voraussetzungen können wir zu einem beliebigen $\varrho_0 < 1$ und $\omega > 0$ ein $1 > \varrho > \varrho_0$ wählen, so daß erstens

$$g\left[\varrho + \frac{1}{g(\varrho)^\omega}\right] < 2^{\frac{1}{\omega}} g(\varrho),$$

zweitens

$$\varrho + \frac{1}{g(\varrho)^\omega} < 1.$$

In Hilfssatz A setzen wir $U(\varrho) = g(\varrho)^\omega$ und schreiben

$$(8.2) \quad g\left(\varrho + \frac{1}{g(\varrho)^\omega}\right) < 2^{\frac{1}{\omega}} g(\varrho), \quad \varrho' < \varrho < \varrho' + \frac{3}{g(\varrho')^\omega}.$$

⁹⁾ Dies ist nur eine andere Form des in 4. benutzten Borelschen Satzes. Siehe Blumenthal, loc. cit. S. 16, Fußnote.

Nach Voraussetzung ist $\overline{\lim}_{\varrho=1} g(\varrho) (1-\varrho)^\omega = +\infty$, also für ein gewisses $\varrho' > \varrho_0$

$$\frac{1}{g(\varrho')} < \frac{(1-\varrho')^\omega}{2^\omega}$$

oder

$$(8.3) \quad \frac{3}{g(\varrho')^\omega} < \frac{1-\varrho'}{2}.$$

Nachdem ϱ' so gewählt ist, wählen wir ϱ nach (8.2). Nach (8.2) und (8.3) folgt, daß

$$(8.4) \quad \varrho + \frac{1}{g(\varrho)^\omega} < \varrho' + \frac{3}{g(\varrho')^\omega} + \frac{1}{g(\varrho)^\omega} < \varrho' + \frac{4}{g(\varrho')^\omega} < \varrho' + \frac{4}{6} (1-\varrho') < 1.$$

Die Ungleichungen (8.2) und (8.4) bilden den Inhalt des Hilfssatzes B.

Hilfssatz C. Für $\varrho < 1$ sei $g(\varrho)$ monoton wachsend und für ein fixes aber beliebiges $\omega < 1$ $\overline{\lim}_{\varrho=1} g(\varrho) (1-\varrho)^\omega > 2$. Dann gibt es zu einem jeden $\varrho_0 < 1$ ein $\varrho > \varrho_0$, $\varrho < 1$, für das erstens

$$g\left[\varrho + \frac{1}{g(\varrho)^\omega}\right] < 2^\omega g(\varrho),$$

zweitens

$$\varrho + \frac{1}{g(\varrho)^\omega} < 1.$$

Der Beweis ist ähnlich wie beim Hilfssatz B.

Jetzt formulieren wir unsere Sätze samt Beweis.

Satz VIII. Seien die analytischen Funktionen

$$g_i(z) \quad (i = 1, 2, \dots, n), \quad a_i(z) \quad (i = 0, 1, 2, \dots, n)$$

regulär für $|z| < 1$, außerdem für jedes positive Ω

$$\overline{\lim}_{\varrho=1} g(\varrho) (1-\varrho)^\omega = +\infty, \quad g(\varrho) = g_1^+(\varrho) + g_2^+(\varrho) + \dots + g_n^+(\varrho).$$

Wenn außerdem die $a_i(z)$ durch $g(\varrho)$ bei einem festen δ abschätzbar sind, so folgt aus der Identität

$$a_1(z) e^{\varrho_1(z)} + \dots + a_n(z) e^{\varrho_n(z)} = a_0(z)$$

die weitere Identität

$$c_1 a_1(z) e^{\varrho_1(z)} + \dots + c_n a_n(z) e^{\varrho_n(z)} = 0, \quad (|c_1| + |c_2| + \dots + |c_n| \neq 0).$$

Zum Beweise leiten wir aus (8.1) einen Widerspruch ab. Wir setzen in Hilfssatz B $\omega = \frac{1-\delta'}{2(1+\delta')}$ und in (8.1) $\varepsilon = \frac{1}{g(\varrho)^\omega}$.

Es folgt aus (8.1) und (8.2), daß für gewisse ϱ

$$g_i^+(\varrho) < \frac{g(\varrho + \varepsilon)^{\delta'}}{\varepsilon^{1+\delta'}} = g\left(\varrho + \frac{1}{g(\varrho)^\omega}\right)^{\delta'} g(\varrho)^{\omega(1+\delta')}$$

$$< 2^{\frac{\delta'}{\omega}} g(\varrho)^{\delta'} g(\varrho)^{\frac{1-\delta'}{2}} = 2^{\frac{\delta'}{\omega}} g(\varrho)^{\frac{1+\delta'}{2}}$$

$$(\delta = 1, 2, \dots, n).$$

Wenn wir diese Ungleichungen addieren, so bekommen wir

$$\sum_{i=1}^n g_i^+(\varrho) = g(\varrho) < n 2^{\frac{\delta'}{\omega}} g(\varrho)^{\frac{1+\delta'}{2}},$$

das heißt, wenn nur ϱ_0 genug nahe der Einheit ist, so folgt $g(\varrho) < g(\varrho)$.
Und das ist der verlangte Widerspruch.

Ähnlich beweisen wir aus Hilfssatz C den

Satz IX. *Seien*

$$a_i(z) \quad (i = 0, 1, 2, \dots, n), \quad g_i(z) \quad (i = 1, 2, \dots, n)$$

regulär für $|z| < 1$, $g(\varrho) = \sum_{i=1}^n g_i^+(\varrho)$, $\lim_{\varrho=1} \overline{\lim} g(\varrho) (1-\varrho)^\Omega > 0$ für ein fixes $\Omega > 1$ und die $a_i(z)$ durch $g(\varrho)$ bei einer beliebig kleinen δ abschätzbar.

Dann folgt aus der Identität $\sum_{i=1}^n a_i(z) e^{g_i(z)} = a_0(z)$ die weitere Identität

$$\sum_{i=1}^n c_i a_i(z) e^{g_i(z)} = 0, \quad \sum_{i=1}^n |c_i| \neq 0.$$

Budapest, Oktober 1927.

(Eingegangen am 18. 10. 1927.)

Über einen Satz von Mellin.

Von

Kurt Mahler in Göttingen.

In seiner Arbeit¹⁾: „Eine Formel für den Logarithmus transzendenter Funktionen von endlichem Geschlecht“ bewies Hj. Mellin, daß die Reihe

$$f(s) = \sum_1^{\infty} \dots \sum_1^{\infty} f(n_1 \dots n_p)^{-s},$$

wo $f(x_1 \dots x_p)$ ein Polynom in p Veränderlichen mit Koeffizienten von positivem Realteil ist, sich in die ganze s -Ebene als eine meromorphe Funktion fortsetzen lasse, die allein Pole auf der reellen Achse besitzt. Er gab ferner eine Abschätzung von $|f(s)|$, wenn s parallel der imaginären Achse ins Unendliche rückt. Diese Abschätzung und zwar in besserer Form, ferner die Fortsetzbarkeit bewies E. Landau auf ganz andere Art²⁾ im Falle $p = 1$ in der Arbeit: „Über einen Mellinschen Satz“. In der vorliegenden Arbeit wird der Mellinsche Satz für den allgemeinen Fall, daß weder $f(x_1 \dots x_p)$, noch der höchste homogene Bestandteil hiervon im ersten „Oktanten“:

$$x_1 > 0, \quad x_2 > 0, \quad \dots, \quad x_p > 0$$

verschwindet, aufs neue und zwar mit ganz elementaren Mitteln, besonders der Eulerschen Summenformel, bewiesen. Je nach den Realitätsannahmen, die über $f(x_1 \dots x_p)$ gemacht sind, werden andere Abschätzungen von $f(s)$ in Streifen parallel der imaginären Achsen erhalten, Abschätzungen, die die Mellinsche und Landausche als Sonderfall enthalten.

Kapitel I.

I.

1. Mit x_1, x_2, \dots, x_p seien p nichtnegative reelle Variable bezeichnet:

$$x_l \geq 0 \quad (l = 1, 2, \dots, p);$$

¹⁾ Acta soc. scient. fennicae 29 (1900), Nr. 4.

²⁾ Archiv der Math. und Physik (3) 24 (1915), 2. Heft.

Weiter seien

$$f(x_1 x_2 \dots x_p) = f(x)$$

$$g(x_1 x_2 \dots x_p) = g(x)$$

zwei Polynome mit beliebigen komplexen Koeffizienten vom Grade m und n . Sie zerfallen folgendermaßen in homogene Bestandteile:

$$f(x) = f_m(x) + f_{m-1}(x) + \dots + f_0(x)$$

$$g(x) = g_n(x) + g_{n-1}(x) + \dots + g_0(x).$$

Der Index gibt dabei die Dimension an.

Sei zur Abkürzung

$$v = \left| \sqrt{x_1^2 + x_2^2 + \dots + x_p^2} \right|.$$

Das Polynom $f(x)$ gehöre zu einer der folgenden drei Klassen, von denen jede eine Unterklasse der vorherigen ist:

$$\text{A: } |f(x)| > 0 \text{ für alle } x_i \geq 0; \quad |f_m(x)| > 0 \text{ für } v > 0$$

$$\text{B: } R(f(x)) > 0 \text{ für alle } x_i \geq 0; \quad R(f_m(x)) > 0 \text{ für } v > 0$$

$$\text{C: } f(x) > 0 \text{ für alle } x_i \geq 0; \quad f_m(x) > 0 \text{ für } v > 0.$$

2. $f_m(x)$ ist in dem beschränkten Bereich $v = 1$ stetig und von Null verschieden; daher existieren zwei positive Konstante c_1, c_2 , so daß gleichmäßig für alle x :

$$2c_1 \leq \left| f_m\left(\frac{x}{v}\right) \right| \leq \frac{c_2}{2}$$

oder

$$2c_1 v^m \leq |f_m(x)| \leq \frac{c_2}{2} v^m$$

ist. Folglich ist von einer Grenze $v_0 \geq 0$ an:

$$c_1 v^m \leq |f(x)| \leq c_2 v^m \text{ für } v \geq v_0.$$

Weiter gibt es eine positive Konstante c_3 , so daß

$$|g(x)| \leq c_3 v^n \text{ für } v \geq v_0$$

ist.

In der Zerlegung

$$f(x) = f_m(x)(1 + r(x)); \quad r(x) = \frac{f_{m-1}(x) + f_{m-2}(x) + \dots + f_0(x)}{f_m(x)}$$

befriedigt $r(x)$ die Ungleichung

$$|vr(x)| \geq c_4 \text{ für } v \geq v_0,$$

wenn c_4 eine hinreichend große positive Konstante bedeutet. Ist also δ ein beliebiger positiver echter Bruch, so gibt es hierzu eine Zahl $v_1 = v_1(\delta) \geq v_0$, so daß zugleich

$$|r(x)| \leq \delta \quad |\text{arc}(1 + r(x))| \leq \delta \quad \text{für } v \geq v_1(\delta)$$

ist und erst recht

$$|\tau r(x)| \leq \delta \quad |\operatorname{arc}(1 + \tau r(x))| \leq \delta \quad \text{für } 0 \leq \tau \leq 1, v \geq v_1(\delta).$$

3. Da $f_m(x)$ in dem endlichen und einfach zusammenhängenden Bereich $v = 1$ stetig und von Null verschieden ist, so stellt hier $\log f_m(x)$ eine eindeutige und stetige Funktion dar, wenn noch der Zweig des Logarithmus in einem beliebigen Punkt beliebig ausgewählt wird. Wir definieren allgemein:

$$\log f_m(x) = \log f_m\left(\frac{x}{v}\right) + m \log v \quad \text{für } v > 0.$$

Diese Funktion ist eindeutig und stetig; es gibt eine positive Konstante M , so daß

$$|J(\log f_m(x))| \leq M \quad \text{für } v > 0$$

ist.

Vermöge der vorigen Ungleichungen für $r(x)$ wird ferner durch die Forderung

$$|J(\log(1 + \tau r(x)))| \leq \frac{\pi}{2} \quad \text{für } v \geq v_1(\delta)$$

eine in den x_i und τ stetige Funktion $\log(1 + \tau r(x))$ definiert.

Schließlich wird durch

$$\log f(x) = \log f_m(x) + \log(1 + r(x)) \quad \text{für } v \geq v_1(\delta)$$

eine eindeutige und stetige Funktion erhalten. $f(x)$ ist noch in dem endlichen einfach zusammenhängenden Bereich $0 \leq v \leq v_1$ stetig und von Null verschieden; $\log f(x)$ läßt sich also auch hierher als eindeutige und stetige Funktion fortsetzen, so daß damit $\log f(x)$ als eindeutige und stetige Funktion im ganzen Bereich der Veränderlichen definiert ist. Vermöge dieser Definition existiert eine positive Konstante N , so daß für alle $x_i \geq 0$

$$|J(\log(f(x)))| \leq N$$

ist.

Gehört $f(x)$ zur Klasse A, so können die Zahlen M und N beliebig groß werden; dagegen darf man ohne Einschränkung fordern:

wenn $f(x)$ zur Klasse B gehört:

$$0 \leq M < \frac{\pi}{2}, \quad 0 \leq N < \frac{\pi}{2};$$

wenn $f(x)$ zur Klasse C gehört:

$$M = N = 0.$$

4. Sei $s = \sigma + ti$ eine komplexe Veränderliche und

$$f_m(x)^{-s} = e^{-s \log f_m(x)}$$

$$f(x)^{-s} = e^{-s \log f(x)}$$

$$(1 + \tau r(x))^{-s} = e^{-s \log(1 + \tau r(x))},$$

mit den früher definierten Zweigen des Logarithmus. Verstehen wir unter c_5 die Konstante

$$c_5 = \max\left(c_1, c_2, \frac{1}{c_1}, \frac{1}{c_2}\right),$$

so ist nach den vorherigen Abschätzungen:

$$\left. \begin{aligned} |f_m(x)^{-s}| &\leq c_5^{|\sigma|} e^{M|t|} v^{-m\sigma} && \text{für } v \geq 0 \\ |f(x)^{-s}| &\leq c_5^{|\sigma|} e^{N|t|} v^{-m\sigma} \\ |g(x)f_m(x)^{-s}| &\leq c_3 c_5^{|\sigma|} e^{M|t|} v^{n-m\sigma} \\ |g(x)f(x)^{-s}| &\leq c_3 c_5^{|\sigma|} e^{N|t|} v^{n-m\sigma} \end{aligned} \right\} \text{für } v \geq v_0.$$

Weiter gibt es einen positiven echten Bruch δ_0 , so daß gleichmäßig

$$|g(x)f(x)^{-s}| \leq c_3 c_5^{|\sigma|} e^{N|t|} v^{n-m\sigma} \quad \text{für } v \leq w, w \geq v_1(\delta)$$

ist, wenn

$$0 < \delta \leq \delta_0.$$

Schließlich besteht die Ungleichung:

$$|(1 + \tau r(x))^{-s}| \leq (1 - \delta)^{-|\sigma|} e^{\delta|t|} \quad \text{für } 0 \leq \tau \leq 1, v \geq v_1(\delta).$$

5. Die Zahlen

$$u_l = \frac{x_l}{v} \geq 0 \quad (l = 1, 2, \dots, p)$$

befriedigen die Identität

$$u_1^2 + u_2^2 + \dots + u_p^2 = 1.$$

Deutet man die x_l als rechtwinklige Koordinaten in einem p -dimensionalen Raum, so liegt der Punkt u_l auf dem ersten „Oktanten“ O der Einheitskugel. Das Oberflächendifferential $d\omega$ dieser Einheitskugel befriedigt die identische Gleichung:

$$dx_1 dx_2 \dots dx_p = v^{p-1} dv d\omega.$$

II.

6. Das Integral

$$J(s) = \int_0^\infty \dots \int_0^\infty g(x)f(x)^{-s} dx_1 dx_2 \dots dx_p$$

geht vermöge der letzten Formel über in

$$J(s) = \int_0^\infty \int_0^\infty v^{p-1} g(uv)f(uv)^{-s} dv d\omega.$$

Dies Integral zerlegen wir in

$$J(s) = \int_0^w \int_0^\infty + \int_0^\infty \int_w^\infty = J_1(s, w) + J_2(s, w) \quad (w \geq v_1(\delta))$$

7. Das Integral $J_1(s, w)$ wird über einen endlichen Bereich erstreckt, der Integrand ist in s regulär, in u, v stetig. Daher stellt $J_1(s, w)$ eine ganze Funktion von s dar. Nach den Abschätzungen in 4. ist

$$|J_1(s, w)| \leq c_3 c_6 c_5^{|\sigma|} e^{N|t|} w^{p+n-m\sigma} \quad c_6 = \int_0^1 d\omega.$$

Wird σ auf das endliche Intervall

$$\sigma_1 \leq \sigma \leq \sigma_2$$

beschränkt, so gilt folglich gleichmäßig in σ :

$$J_1(s, w) = O(e^{N|t|}).$$

Wenn $f(x)$ zur Klasse C gehört, so ist $J_1(s, w)$ in diesem Streifen sogar beschränkt.

8. Im Integral

$$J_2(s, w) = \int_0^1 \int_w^\infty v^{p-1-ms} g(uv) f_m(u)^{-s} \{1 + r(uv)\}^{-s} dv d\omega$$

gestattet der letzte Faktor folgende abbrechende Taylorentwicklung:

$$\begin{aligned} \{1 + r(uv)\}^{-s} &= \sum_0^{k-1} \binom{-s}{l} r(uv)^l + \\ &+ k \binom{-s}{k} r(uv)^k \int_0^1 \{1 + \tau r(uv)\}^{-s-k} (1 - \tau)^{k-1} d\tau; \end{aligned}$$

dabei bedeutet k eine beliebige natürliche Zahl. $J_2(s, w)$ kann daher zerlegt werden:

$$J_2(s, w) = \sum_0^{k-1} \binom{-s}{k} M_k(s, w) + N_k(s, w),$$

und hier ist

$$M_k(s, w) = \int_0^1 \int_w^\infty v^{p-1-ms} g(uv) f_m(u)^{-s} r(uv)^k dv d\omega,$$

$$N_k(s, w) =$$

$$= \int_0^1 \int_w^\infty \int_0^1 k \binom{-s}{k} v^{p-1-ms} g(uv) f_m(u)^{-s} r(uv)^k \{1 + \tau r(uv)\}^{-s-k} (1 - \tau)^{k-1} d\tau dv d\omega$$

9. Im Integral $M_k(s, w)$ ist

$$g(uv) = v^n \{g_n(u) + v^{-1} g_{n-1}(u) + \dots + v^{-n} g_0(u)\},$$

$$r(uv) = v^{-1} \left\{ \frac{f_{m-1}(u)}{f_m(u)} + v^{-1} \frac{f_{m-2}(u)}{f_m(u)} + \dots + v^{-(m-1)} \frac{f_0(u)}{f_m(u)} \right\},$$

folglich:

$$g(uv) r(uv)^l = v^{n-l} \sum_{h=0}^{n+(m-1)l} A_h^{(l)}(u) v^{-h}; \quad A_0^{(0)} = g_n(u).$$

Dabei sind die $A_h^{(l)}(u)$ rationale Funktionen und auf O stetig.

$M_l(s, w)$ nimmt also die Form an:

$$M_l(s, w) = \sum_{h=0}^{n+(m-1)l} \int_0^\infty \int_w^\infty v^{n+p-ms-l-h-1} A_h^{(l)}(u) f_m(u)^{-s} dv d\omega.$$

Wenn

$$\sigma > \frac{n+p}{m}$$

ist, darf die Integration nach v ohne weiteres ausgeführt werden. Man erhält dann

$$M_l(s, w) = \sum_0^{n+(m-1)l} \frac{C_h^{(l)}(s, w)}{ms+l+h-n-p};$$

$$C_h^{(l)}(s, w) = w^{n+p-ms-h-l} \int_0^\infty A_h^{(l)}(u) f_m(u)^{-s} d\omega.$$

Analog wie bei $J_1(s, w)$ folgt, daß die $C_h^{(l)}(s, w)$ ganze Funktionen von s sind, in jedem endlichen Intervall

$$\sigma_1 \leq \sigma \leq \sigma_2$$

befriedigen sie gleichmäßig die Abschätzung

$$C_h^{(l)}(s, w) = O(e^{M|t|});$$

sie sind also beschränkt, wenn $f(x)$ zur Klasse C gehört.

$M_l(s, w)$ ist somit eine meromorphe Funktion mit einfachen Polen höchstens bei

$$s = \frac{n+p-l-h}{m} \quad (h = 0, 1, \dots, n + (m-1)l)$$

mit dem Residuum

$$\frac{1}{m} C_h^{(l)}\left(\frac{n+p-l-h}{m}, w\right);$$

natürlich hängt diese Zahl nur scheinbar von w ab. Wenn sie verschwindet, so ist $M_l(s, w)$ an den betreffenden Stellen regulär.

Insbesondere $M_0(s, w)$ und keine der anderen Funktionen $M_l(s, w)$ besitzt bei

$$s = \frac{n+p}{m}$$

einen einfachen Pol, mit dem Residuum

$$\lambda = \frac{1}{m} \int_0^\infty g_n(u) f_m(u)^{-\frac{n+p}{m}} d\omega.$$

Die Funktionen $M_l(s, w)$ befriedigen gleichmäßig in jedem endlichen Streifen

$$\sigma_1 \leq \sigma \leq \sigma_2$$

für große $|t|$ die Abschätzung

$$M_l(s, w) = O(e^{M|t|}).$$

Wenn $f(x)$ zur Klasse C gehört, so ist $M_l(s, w)$ in diesem Streifen sogar beschränkt.

10. Der Integrand von

$$N_k(s, w) =$$

$$= \int_0^\infty \int_w^1 \int_0^1 k \binom{-s}{k} v^{p-1-ms} g(uv) f_m(u)^{-s} r(uv)^k \{1 + \tau r(uv)\}^{-s-k} (1-\tau)^{k-1} d\tau dv dw$$

besitzt nach den früheren Abschätzungen die Majorante:

$$k \left| \binom{-s}{k} \right| \cdot v^{p-1-ms} \cdot c_3 c_5^{|\sigma|} e^{M|t|} v^n \cdot c_4^k v^{-k} \cdot (1-\delta)^{-|\sigma|-k} \cdot e^{\delta|t|}$$

oder anders geschrieben:

$$k \left| \binom{-s}{k} \right| \cdot \{c_3 c_4^k c_5^{|\sigma|} (1-\delta)^{-|\sigma|-k}\} \cdot e^{(M+\delta)|t|} \cdot v^{n+p-1-k-m\sigma}.$$

Wird s auf den endlichen Bereich

$$\sigma \geq \frac{n+p-k+\varepsilon}{m}, \quad |s| \leq R,$$

wo ε und R^{-1} zwei sehr kleine positive Konstanten sind, beschränkt, so konvergiert folglich das Integral für $N_k(s, w)$ absolut und gleichmäßig; also ist $N_k(s, w)$ in der Halbebene

$$\sigma > \frac{n+p-k}{m}$$

eine reguläre Funktion. In einem endlichen Streifen

$$\frac{n+p-k}{m} < \sigma_1 \leq \sigma \leq \sigma_2$$

ist gleichmäßig in σ für große $|t|$:

$$N_k(s, w) = O(|t|^k e^{(M+\delta)|t|}).$$

Sei jetzt ϑ eine beliebig kleine positive Konstante, $\delta = \frac{\vartheta}{2}$ gewählt, also $w \geq v_1 \left(\frac{\vartheta}{2}\right)$ hinreichend groß; wegen

$$|t|^k = O\left(e^{\frac{\vartheta}{2}|t|}\right)$$

ist alsdann

$$N_k(s, w) = O(e^{(M+\vartheta)|t|}),$$

wie klein auch ϑ sei. Gehört $f(x)$ zur Klasse B, ist also $M < \frac{\pi}{2}$, so kann ϑ so klein gewählt werden, daß auch

$$M + \vartheta < \frac{\pi}{2}$$

ist. Wenn dagegen $f(x)$ zur Klasse C gehört, so kann eine viel bessere Abschätzung im Streifen

$$\frac{n+p-k}{k} < \sigma_1 \leq \sigma \leq \sigma_2$$

erhalten werden. Im Integranden von $N_k(s, w)$ bleibt jetzt $f_m(u)$ und $1 + \tau r(uv)$ immer positiv; der Integrand besitzt folglich die Majorante

$$k \left| \binom{-s}{k} \right| \{c_3 c_4^k c_5\}^{\sigma} (1-\delta)^{-|\sigma|-k} v^{n+p-1-k-m\sigma}$$

und daher ist jetzt gleichmäßig für große $|t|$:

$$N_k(s, w) = O\left(\left|\binom{-s}{k}\right|\right) = O(|t|^k).$$

11. Wir fassen die letzten Ergebnisse zusammen. In der Zerlegung

$$J(s) = J_1(s, w) + \sum_0^{k-1} \binom{-s}{l} M_l(s, w) + N_k(s, w)$$

ist hiernach:

$J_1(s, w)$ eine ganze Funktion;

$M_l(s, w)$ eine meromorphe Funktion mit einfachen Polen höchstens bei

$$s = \frac{n+p-l-h}{m} \quad (h = 0, 1, 2, \dots, n + (m-1)l)$$

mit dem Residuum

$$\frac{1}{m} C_h^{(l)} \left(\frac{n+p-l-h}{m}, w \right);$$

$N_k(s, w)$ eine reguläre Funktion für

$$\sigma > \frac{n+p-k}{k}.$$

Folglich ist auch $J(s)$ in dieser Halbebene analytisch, mit den einfachen Polen höchstens bei

$$s = \frac{n+p-h}{m} \quad (h = 0, 1, \dots, k-1);$$

hiervon scheiden jedoch die Stellen

$$0, -1, -2, \dots$$

gewiß als regulär aus, da in dem Summanden

$$\binom{-s}{l} M_l(s, w)$$

der erste Faktor bei

$$s = 0, -1, \dots, -(l-1)$$

verschwindet, die einfachen Pole des zweiten Faktors aber alle größer oder gleich

$$\frac{n+p-l-[n+(m-1)l]}{m} > -l$$

sind.

Im Streifen

$$\frac{n+p-k}{m} < \sigma_1 \leq \sigma \leq \sigma_2$$

ist

$$\binom{-s}{l} = O(|t|^k) = O(e^{\vartheta|t|}),$$

also nach den früheren Abschätzungen im Fall A und B:

$$J(s) = O(e^{N|t|}) + \sum_0^{k-1} O(e^{(M+\vartheta)|t|}) + O(e^{(M+\vartheta)|t|}) = O(e^{S|t|}),$$

wobei

$$S = \max(M + \vartheta, N).$$

Liegt der Fall B vor, so ist $M < \frac{\pi}{2}$, $N < \frac{\pi}{2}$; da ϑ beliebig klein sein darf, also auch $S < \frac{\pi}{2}$.

Wenn hingegen $f(x)$ zur Klasse C gehört, so gilt die schärfere Abschätzung

$$J(s) = O(1) + \sum_0^{k-1} O(|t|^k) + O(|t|^k) = O(|t|^k).$$

Bei den vorigen Überlegungen durfte k jede natürliche Zahl bedeuten. Wählt man $k=1$, so ergibt sich, daß $J(s)$ in der Halbebene

$$\sigma > \frac{n+p}{m}$$

konvergiert. Nimmt man dagegen für k eine sehr große Zahl, so ergibt sich die Fortsetzbarkeit von $J(s)$ als meromorphe Funktion in die ganze Ebene.

12. Zusammengefaßt lautet das Ergebnis dieses Kapitels somit:

Satz I. *Das Integral*

$$J(s) = \int_0^{\infty} \dots \int_0^{\infty} g(x) f(x)^{-s} dx_1 \dots dx_p$$

konvergiert in der Halbebene

$$\sigma > \frac{n+p}{m}$$

und stellt hier eine reguläre Funktion dar. $J(s)$ läßt sich in die ganze Ebene als meromorphe Funktion fortsetzen, mit einfachen Polen höchstens an den Stellen

$$s = \frac{n+p-h}{m} \quad \left(\begin{array}{l} h = 0, 1, 2, \dots \\ s \neq 0, -1, -2, \dots \end{array} \right).$$

Insbesondere besitzt $J(s)$ im rechtesten Pol:

$$s = \frac{n+p}{m}$$

das Residuum

$$\lambda = \frac{1}{m} \int_0^1 g_n(u) f_m(u)^{-\frac{n+p}{m}} d\omega.$$

In jedem endlichen Streifen

$$\frac{n+p-k}{m} < \sigma_1 \leq \sigma \leq \sigma_2 \quad (k = 0, 1, 2, \dots)$$

besteht die Abschätzung

$$J(s) = O(e^{S|t|})$$

für große $|t|$ gleichmäßig in σ . Wenn $f(x)$ zur Klasse B gehört und für $\log f(x)$ der Hauptwert genommen wird, so wird

$$0 \leq S < \frac{\pi}{2};$$

ist endlich $f(x)$ zur Klasse C gehörig, so besteht sogar die schärfere Abschätzung

$$J(s) = O(|t|^k)$$

für große $|t|$ gleichmäßig in σ , wenn man für $\log f(x)$ den reellen Wert nimmt.

Kapitel II.

13. Die bekannte Eulersche Summenformel

$$\sum_1^{\infty} \mathfrak{F}(w) = \int_0^{\infty} \mathfrak{F}(x) dx + \sum_{l=1}^{k-1} \frac{(-1)^{l+1} B_l}{l!} \mathfrak{F}^{(l-1)}(0) + \int_0^{\infty} \frac{(-1)^{k-1}}{(k-1)!} \sigma_k(x) \mathfrak{F}^{(k)}(x) dx,$$

in der k eine natürliche Zahl, B_l die l -te Bernoullische Zahl und $\sigma_k(x)$ das k -te Bernoullische Polynom mit dem Argument $x - [x]$ bedeutet, darf z. B. auf solche komplexe Funktionen $\mathfrak{F}(x)$ der reellen, nichtnegativen Veränderlichen x angewandt werden, die k -mal stetig differenzierbar sind und für welche in der obigen Gleichung die Summe links und die beiden Integrale rechts absolut konvergieren und die Gleichungen

$$\lim_{x \rightarrow +\infty} \mathfrak{F}^{(l)}(x) = 0 \quad (l = 1, 2, \dots, k)$$

erfüllt sind.

Mit den Abkürzungen

$$h_l^{(k)}(x) = (-1)^l \frac{B_l}{l!} \quad \text{für } l = 0, 1, \dots, k-1$$

$$h_k^{(k)}(x) = \frac{(-1)^{k-1}}{(k-1)!} \sigma_k(x)$$

nimmt die Eulersche Formel die Gestalt an:

$$\sum_1^{\infty} \mathfrak{F}(w) = \sum_{i=0}^k \int_0^{\infty} h_i^{(k)}(x) \mathfrak{F}^{(i)}(x) dx.$$

Offenbar bleiben in dieser Formel die Funktionen $h_i^{(k)}(x)$ unterhalb einer Konstanten, die von l nicht abhängt:

$$|h_i^{(k)}(x)| \leq \varrho_k \quad (l = 0, 1, 2, \dots, k).$$

14. Seien jetzt x_1, x_2, \dots, x_p wieder p nichtnegative reelle Veränderliche und k_1, k_2, \dots, k_p p natürliche Zahlen, ferner

$$\mathfrak{F}(x_1 x_2 \dots x_p) = \mathfrak{F}(x)$$

eine Funktion mit stetigen Ableitungen:

$$\mathfrak{F}_{l_1 l_2 \dots l_p}(x) = \frac{\partial^{l_1+l_2+\dots+l_p} \mathfrak{F}(x)}{\partial x_1^{l_1} \partial x_2^{l_2} \dots \partial x_p^{l_p}} \quad \left(\begin{array}{l} l_\nu = 0, 1, \dots, k_\nu \\ \nu = 1, 2, \dots, p \end{array} \right).$$

Eine p -malige Anwendung der Eulerschen Summenformel ergibt die allgemeine Gleichung:

$$\sum_{w_1=1}^{\infty} \dots \sum_{w_p=1}^{\infty} \mathfrak{F}(w_1 \dots w_p) = \sum_{l_1=0}^{k_1} \dots \sum_{l_p=0}^{k_p} \int \dots \int_0^{\infty} h_{l_1}^{(k_1)}(x_1) \dots h_{l_p}^{(k_p)}(x_p) \mathfrak{F}_{l_1 \dots l_p}(x) dx_1 \dots dx_p.$$

Diese Formel darf z. B. angewandt werden, wenn die Summe links und die

$$(k_1 + 1)(k_2 + 1) \dots (k_p + 1)$$

Integrale rechts absolut konvergieren.

15. Sei

$$k_1 = k_2 = \dots = k_p = k,$$

$$\mathfrak{F}(x) = g(x) f(x)^{-s}$$

mit den Polynomen $f(x)$ und $g(x)$ des vorigen Kapitels. In den Integralen

$$J_{l_1 \dots l_p}(s) = \int \dots \int_0^{\infty} h_{l_1}^{(k)}(x_1) \dots h_{l_p}^{(k)}(x_p) \mathfrak{F}_{l_1 \dots l_p}(x) dx_1 \dots dx_p$$

ergibt sich für $\mathfrak{F}_{l_1 \dots l_p}(x)$ durch Differenzieren

$$\mathfrak{F}_{l_1 \dots l_p}(x) = \sum_{h=0}^l \binom{-s}{h} \varphi_{l_1 \dots l_p}^{(h)}(x) f(x)^{-s-h}, \quad l = \sum_{\nu=1}^p l_\nu,$$

wobei $\varphi_{l_1 \dots l_p}^{(h)}(x)$ ein von s unabhängiges Polynom in den Veränderlichen x vom Grade

$$n + h m - l$$

ist; ist diese Zahl negativ, so verschwindet das Polynom identisch.

16. Im Falle

$$0 \leq l_\nu \leq k - 1 \quad (\nu = 1, 2, \dots, p)$$

kann $J_{l_1 \dots l_p}(s)$ zerlegt werden in der Form:

$$J_{l_1 \dots l_p}(s) = c(l_1 l_2 \dots l_p) \sum_{h=0}^l \binom{-s}{h} \int \dots \int_0^\infty \varphi_{l_1 \dots l_p}^{(h)}(x) f(x)^{-s-h} dx_1 \dots dx_p,$$

wobei

$$c(l_1 l_2 \dots l_p) = h_{l_1}^{(k)}(x_1) \dots h_{l_p}^{(k)}(x_p)$$

eine absolute Konstante ist.

Das Teilintegral

$$J_{l_1 \dots l_p}^{(h)}(s) = \int \dots \int_0^\infty \varphi_{l_1 \dots l_p}^{(h)}(x) f(x)^{-s-h} dx_1 \dots dx_p$$

ist gerade von der Gestalt der Funktion $J(s)$, auf die sich Satz I bezieht. Aus diesem Satz ergibt sich, daß $J_{l_1 \dots l_p}^{(h)}(s)$ in der Halbebene

$$\sigma = \frac{n+p-l}{m}$$

absolut konvergiert; $J_{l_1 \dots l_p}^{(h)}(s)$ besitzt einfache Pole höchstens bei

$$s = \frac{n+p-l-j}{m} \quad \left(\begin{array}{l} j = 0, 1, 2, \dots \\ s \neq 0, -1, -2, \dots \end{array} \right).$$

Speziell ist

$$J_{0,0,\dots,0}(s) = J_{0,0,\dots,0}^{(0)}(s) = J(s);$$

allein diese Funktion besitzt bei

$$s = \frac{n+p}{m}$$

einen einfachen Pol, mit dem Residuum λ .

In dem Streifen

$$\frac{n+p-k}{m} < \sigma_1 \leq \sigma \leq \sigma_2$$

besteht gleichmäßig in σ für große $|t|$ die Abschätzung

$$J_{l_1 l_2 \dots l_p}^{(h)}(s) = O(e^{S|t|});$$

gehört $f(x)$ zur Klasse C, so läßt sich diese Abschätzung verbessern zu

$$J_{l_1 l_2 \dots l_p}^{(h)}(s) = O(|t|^g) \quad (g = \max(k-l, 0)).$$

Sei nun T eine Zahl, die größer als S ist:

$$T > S;$$

im Falle, daß $f(x)$ zur Klasse B gehört, also $S < \frac{\pi}{2}$ ist, nehmen wir auch

$$T < \frac{\pi}{2}$$

an. Es ist dann im Streifen

$$\frac{n+p-k}{m} < \sigma_1 \leq \sigma \leq \sigma_2$$

für große $|t|$:

$$\binom{-s}{h} = O(|t|^h) = O(e^{(T-s)|t|})$$

und daher nach den obigen Abschätzungen:

$$J_{l_1 l_2 \dots l_p}(s) = O(e^{T|t|}),$$

und wenn $f(x)$ zur Klasse C gehört, genauer:

für $l \leq k$:

$$J_{l_1 l_2 \dots l_p}(s) = O(|t|^{\rho+l}) = O(|t|^k),$$

für $l \geq k$:

$$J_{l_1 l_2 \dots l_p}(s) = O(|t|^{l'}).$$

Da nun auf jeden Fall

$$l \leq k \cdot p$$

ist, so besteht die Abschätzung unabhängig von l :

$$J_{l_1 l_2 \dots l_p}(s) = O(|t|^{pk}).$$

17. Sei weiter mindestens eine der Zahlen

$$l_1, l_2, \dots, l_p$$

gleich k . Wegen der Darstellung

$$J_{l_1 \dots l_p}(s) = \sum_{h=0}^l \binom{-s}{h} J_{l_1 \dots l_p}^{(h)}(s),$$

$$J_{l_1 \dots l_p}^{(h)}(s) = \int \dots \int_0^{\infty} h_{l_1}^{(k)}(x_1) \dots h_{l_p}^{(k)}(x_p) \varphi_{l_1 \dots l_p}^{(h)}(x) f(x)^{-s-h} dx_1 \dots dx_p,$$

ergibt sich die absolute und gleichmäßige Konvergenz von $J_{l_1 \dots l_p}^{(h)}(s)$ und damit von $J_{l_1 \dots l_p}(s)$ in dem Bereich

$$\sigma \geq \frac{n+p-k+\varepsilon}{m}, \quad |s| \leq R.$$

Denn es ist mit einer positiven Konstanten $c_{l_1 \dots l_p}^{(h)}$:

$$|\varphi_{l_1 \dots l_p}^{(h)}(x)| \leq v^{hm+n-l} c_{l_1 \dots l_p}^{(h)} \quad \text{für } v \geq v_0;$$

hieraus ergibt sich, ähnlich wie im ersten Kapitel, die Behauptung, da außerdem $l \geq k$ und

$$|h_{l_1}^{(k)}(x_1) \dots h_{l_p}^{(k)}(x_p)| \leq \varrho_k^p$$

ist.

$J_{l_1 \dots l_p}(s)$ ist somit in den Halbebenen

$$\sigma > \frac{n+p-k}{m}$$

regulär. Man findet wie im ersten Kapitel, daß im Streifen

$$\frac{n+p-k}{m} < \sigma_1 \leq \sigma \leq \sigma_2$$

gleichmäßig in σ für große $|t|$

$$J_{l_1 \dots l_p}^{(h)}(s) = O(e^{S|t|}),$$

und folglich

$$J_{l_1 \dots l_p}(s) = O(e^{T|t|})$$

ist. Wenn aber $f(x)$ zur Klasse C gehört, so bleibt $J_{l_1 \dots l_p}^{(h)}(s)$ in jenem Streifen sogar beschränkt und folglich ist für große $|t|$

$$J_{l_1 \dots l_p}(s) = O(|t|^l) = O(|t|^{kp}).$$

18. Das Integral

$$J(s) = \int_0^\infty \dots \int_0^\infty g(x) f(x)^{-s} dx_1 \dots dx_p$$

konvergiert nach früher in der Halbebene

$$\sigma > \frac{n+p}{m}$$

absolut. Nach einem bekannten Cauchyschen Konvergenzsatz konvergiert daher auch

$$Z(s) = \sum_{w_1=1}^\infty \dots \sum_{w_p=1}^\infty g(w) f(w)^{-s}$$

in dieser Halbebene absolut.

19. Nach den letzten vier Abschnitten sind daher für

$$\sigma > \frac{n+p}{m}$$

die Voraussetzungen der Eulerschen Summenformel erfüllt, so daß von ihr Gebrauch gemacht werden darf. Aus den früheren Ergebnissen folgt dann, da k jede natürliche Zahl sein kann:

Satz II. Die Summe

$$Z(s) = \sum_{w_1=1}^\infty \dots \sum_{w_p=1}^\infty g(w) f(w)^{-s}$$

konvergiert in der Halbebene

$$\sigma > \frac{n+p}{m}$$

und stellt hier eine reguläre Funktion dar. $Z(s)$ läßt sich in die ganze Ebene als meromorphe Funktion fortsetzen, mit einfachen Polen höchstens bei

$$s = \frac{n+p-h}{m} \quad \left(\begin{array}{l} h = 0, 1, 2, 3, \dots \\ s \neq 0, -1, -2, -3, \dots \end{array} \right).$$

Insbesondere besitzt $Z(s)$ im rechtesten Pol

$$s = \frac{n+p}{m}$$

das Residuum

$$\lambda = \frac{1}{m} \int_0^1 g_n(u) f_m(u)^{-\frac{n+p}{m}} d\omega.$$

In jedem endlichen Streifen

$$\frac{n+p-k}{m} < \sigma_1 \leq \sigma \leq \sigma_2 \quad (k = 0, 1, 2, \dots)$$

gilt gleichmäßig in σ für große $|t|$ die Abschätzung

$$Z(s) = O(e^{T|t|});$$

gehört $f(x)$ zur Klasse B und nimmt man für $\log f(x)$ den Hauptwert, so kann

$$0 \leq T < \frac{\pi}{2}$$

angenommen werden; gehört $f(x)$ zur Klasse C, so gilt endlich die schärfere Abschätzung

$$Z(s) = O(|t|^{kp}),$$

wenn man für $\log f(x)$ den reellen Wert nimmt.

Krefeld, den 29. März 1927.

Zwei Sätze über Funktionen von zwei komplexen Veränderlichen.

Von

Stefan Bergmann in Berlin.

Die Eigenschaften der Funktionen einer komplexen Veränderlichen, die durch ihr Funktionselement um den Mittelpunkt gegeben ist, sind eingehend untersucht.

Bedeutend weniger ist man auf die entsprechenden Aufgaben aus der Theorie der Funktionen zweier komplexen Veränderlichen eingegangen. Diese zweite Fragestellung ist um so interessanter, da eine unmittelbare Übertragung der Methoden der Funktionentheorie einer komplexen Veränderlichen auf die Theorie von mehreren ernste Schwierigkeiten bietet.

In der vorliegenden Arbeit sollen zwei Sätze aus dem obengenannten Problemkreis über Funktionen von zwei komplexen Veränderlichen mit Hilfe der Algebra von unendlich vielen Veränderlichen abgeleitet werden.

An dieser Stelle möchte ich noch Herrn Hammerstein für seine wertvollen Ratschläge bei der Abfassung der vorliegenden Arbeit danken.

§ 1.

Es sollen in der vorliegenden Arbeit folgende Bezeichnungen gelten:

ξ, η, ζ, τ bedeuten die reellen kartesischen Koordinaten des vierdimensionalen Raumes; $X \equiv \xi + i\eta, Z \equiv \zeta + i\tau$ die komplexen Veränderlichen.

\mathfrak{B} bedeutet einen einfach zusammenhängenden Bereich des vierdimensionalen Raumes, der den Koordinatenanfangspunkt in seinem Innern enthält. Die Berandung von \mathfrak{B} besteht aus stückweise regulären dreidimensionalen Mannigfaltigkeiten.

\mathfrak{K} bezeichne einen Kreisbereich um den Nullpunkt, der ganz im Innern von \mathfrak{B} liegt. $a_{\nu\mu}$ sei eine Abkürzung für die Zahlenfolge

$\frac{1}{\sqrt{\iiint_{\mathfrak{B}} |X^{\nu} Z^{\mu}|^2 d\omega}}$ ¹⁾, wobei $d\omega \equiv d\xi d\eta d\zeta d\tau$ das (vierdimensionale)

Volumenelement bedeutet. $G(X, Z)$ sei schließlich eine im abgeschlossenen \mathfrak{B} reguläre Funktion zweier komplexen Veränderlichen X, Z , dessen Funktionselement

$$(2) \quad \sum_{\substack{(\nu \mu)=\infty \\ (\nu \mu)=1}}^{(\nu \mu)=\infty} b_{\nu \mu} X^{\nu} Z^{\mu}$$

lautet. Es gilt der

Satz. *Zu jedem Bereiche \mathfrak{B} läßt sich eine Folge von Hermiteschen Formen von unendlich vielen Veränderlichen*

$$(3) \quad f_{(n, m)}(\alpha_{\nu \mu}, \bar{\alpha}_{\nu' \mu'}) \equiv \sum_{\substack{(\nu' \mu')=\infty \\ (\nu \mu)=1 \\ (\nu' \mu')=1}}^{(\nu' \mu')=\infty} A_{\nu \mu, \nu' \mu'}^{(n, m)} \alpha_{\nu \mu} \bar{\alpha}_{\nu' \mu'}$$

herstellen, die folgende Eigenschaft besitzt:

ist $G(X, Z)$ im Innern von \mathfrak{B} regulär und existiert

$$\iiint_{\mathfrak{B}} G(X, Z) \overline{G(X, Z)} d\omega, \quad ^2)$$

so existiert

$$(4) \quad \lim_{(n, m) \rightarrow \infty} f_{(n, m)} \left(\frac{b_{\nu \mu}}{a_{\nu \mu}}, \frac{\bar{b}_{\nu' \mu'}}{a_{\nu' \mu'}} \right);$$

existiert umgekehrt der Limes (4), so läßt sich die durch das Funktionenelement (2) in \mathfrak{B} gegebene Funktion $G(X, Z)$ überall im Innern von \mathfrak{B} fortsetzen, und es gilt in \mathfrak{B} die Abschätzung

¹⁾ Da in der vorliegenden Arbeit mit Funktionen zweier komplexen Veränderlichen operiert wird, werden Doppelindizes häufig benutzt. Es wird im folgenden stets die Reihenfolge

$$(1) \quad (0, 0), (1, 0), (0, 1), (2, 0), (1, 1), (0, 2), \dots$$

festgehalten. Unter $(\nu \mu) = k$ werden die Doppelindizes $\nu \mu$ des k -ten Gliedes der Reihenfolge (1) verstanden. Mit $(\nu^0 \mu^0)$ wird das dem (ν, μ) unmittelbar vorhergehende Indizespaar bezeichnet.

Eine Potenzreihe von zwei komplexen Veränderlichen, deren Glieder die Anordnung (1) haben, wird als Diagonalphotenzreihe bezeichnet.

Der Kreisbereich wird nach K. Reinhardt („Über Abbildungen durch analytische Funktionen zweier Veränderlichen“, Math. Annalen 83, 1921, S. 211–255, § 5) definiert als ein konvexer Bereich, dessen Berandung durch die Gleichung: $\sqrt{\xi^2 + \eta^2} = g(\sqrt{\xi^2 + \tau^2})$ gegeben ist, wo g eine reelle stetige Funktion bedeutet.

Die entsprechenden Sätze für den Fall einer komplexen Veränderlichen wurden in der in der Fußnote ²⁾ (S. 410) angegebenen Arbeit § 5 formuliert.

²⁾ Unter $\iiint_{\mathfrak{B}} |v(X, Z)|^2 d\omega$ soll in üblicher Weise $\lim_{\mathfrak{B}_n \rightarrow \mathfrak{B}} \iiint_{\mathfrak{B}_n} |v(X, Z)|^2 d\omega$ verstanden werden, wobei \mathfrak{B}_n eine Folge von ganz im Innern von \mathfrak{B} gelegenen Teilbereichen von \mathfrak{B} bedeutet.

$$(5) \quad |G(X, Z)| \leq \frac{\sqrt{2}}{\pi \gamma^2(X, Z)} \cdot \sqrt{\lim_{(n, m) \rightarrow \infty} f_{(n, m)} \left(\frac{b_{\nu \mu}}{a_{\nu \mu}}, \frac{\overline{b_{\nu' \mu'}}}{\overline{a_{\nu' \mu'}}} \right)},$$

wobei $\gamma(X, Z)$ der Abstand des Punktes (X, Z) von der Berandung von \mathfrak{B} ist.

Beweis. Als erster Punkt des Beweises soll die Existenz eines vollständigen doppelorthogonalen Systems $\varphi_{\nu \mu}(X, Z)$ gezeigt werden. Unter doppelorthogonal wird verstanden, daß die Orthogonalrelationen sowohl bezüglich des Bereiches \mathfrak{B}

$$(6a) \quad \left[\text{d. h. } \iiint_{\mathfrak{B}} \varphi_{\nu \mu} \overline{\varphi_{\nu' \mu'}} d\omega = k_{\nu \mu} \quad (\nu' \mu') = (\nu \mu) \right]$$

$$(6b) \quad \left[\text{d. h. } \iiint_{\mathfrak{B}} \varphi_{\nu \mu} \overline{\varphi_{\nu' \mu'}} d\omega = 0 \quad (\nu' \mu') \neq (\nu \mu) \right],$$

als auch bezüglich des Kreisbereiches \mathfrak{K}

$$(7a) \quad \left[\text{d. h. } \iiint_{\mathfrak{K}} \varphi_{\nu \mu} \overline{\varphi_{\nu' \mu'}} d\omega = 1 \quad (\nu' \mu') = (\nu \mu) \right]$$

$$(7b) \quad \left[\text{d. h. } \iiint_{\mathfrak{K}} \varphi_{\nu \mu} \overline{\varphi_{\nu' \mu'}} d\omega = 0 \quad (\nu' \mu') \neq (\nu \mu) \right]$$

bestehen. Die Existenz des Systems $\varphi_{\nu \mu}$ wird induktiv gezeigt, indem man sich (nachdem die Existenz aller vorhergehenden Lösungen $\varphi_{00}, \varphi_{10}, \varphi_{01}, \dots, \varphi_{\nu^0 \mu^0}$ der entsprechenden Aufgaben bewiesen wurde) folgende Variationsaufgabe stellt:

Unter allen im offenen Bereiche \mathfrak{B} regulären Funktionen ist diejenige zu finden, welche an Stelle von v in das Integral

$$(8) \quad \iiint_{\mathfrak{B}} v(X, Z) \overline{v(X, Z)} d\omega$$

eingesetzt, unter den Nebenbedingungen

$$(9) \quad \iiint_{\mathfrak{K}} v(X, Z) \overline{\varphi_{00}(X, Z)} d\omega = 0, \quad \iiint_{\mathfrak{K}} v(X, Z) \overline{\varphi_{10}(X, Z)} d\omega = 0, \quad \dots,$$

$$\iiint_{\mathfrak{K}} v(X, Z) \overline{\varphi_{\nu^0 \mu^0}(X, Z)} d\omega = 0, \quad \iiint_{\mathfrak{K}} v(X, Z) \overline{v(X, Z)} d\omega = 1$$

(8) zu einem Minimum macht. Der ausführliche Beweis für die Existenz der Lösung dieser Variationsaufgabe soll nur für φ_{00} durchgeführt werden; der Beweis für beliebiges $\varphi_{\nu \mu}$ verläuft Schritt für Schritt analog.

Das Integral (8) besitzt eine endliche untere Grenze k_{00} . (Es ist offenbar: $1 < k_{00} < \frac{\text{Volumen von } \mathfrak{B}}{\text{Volumen von } \mathfrak{K}}$). Es ist deshalb nach einem früher³⁾

³⁾ Vgl. dazu die Arbeiten: „Über Hermite'sche Formen, die zu einem Bereich gehören“ (vorläufige Mitteilung) Sitzungsberichte der Berliner Mathematischen Gesellschaft 26 (1927), S. 178–184 und „Über unendliche Hermite'sche Formen, die zu einem Bereich gehören, nebst Anwendungen auf Fragen der Abbildung durch Funktionen von zwei komplexen Veränderlichen“, erscheint demnächst in der Math. Zeitschrift, Korollar zu Hilfssatz 1.

bewiesenen Satz für die genügend späten v einer Minimalfolge bei gegebenen ε .

$$(10) \quad v(X, Z) \leq \frac{\sqrt{2 \iiint_{\mathfrak{B}} v(X, Z) \overline{v(X, Z)} d\omega}}{\pi \gamma^2(X, Z)} = \frac{\sqrt{2(k_{00} + \varepsilon)}}{\pi \gamma^2(X, Z)}$$

in jedem inneren Punkt von \mathfrak{B} beschränkt.

Man wähle dann eine Folge von Bereichen $\mathfrak{B}_1 < \mathfrak{B}_2 < \dots < \mathfrak{B}_l < \dots < \mathfrak{B}$ mit $\lim_{l \rightarrow \infty} \mathfrak{B}_l = \mathfrak{B}$. In \mathfrak{B}_1 nehme man eine Folge von inneren Punkten.

P_1, P_2, \dots die sich gegen einen inneren Punkt P häufen. Nach dem bekannten Diagonalverfahren kann jetzt aus der Minimalfolge eine solche Teilfolge ausgewählt werden, die auf der Menge P_1, P_2, \dots konvergiert. Der Vitalische Doppelreihensatz lehrt nun die gleichmäßige Konvergenz derselben in jedem \mathfrak{B}_l , da sie dort nach (10) beschränkt ist. Dadurch ist also eine Grenzfunktion φ_{00} erklärt, von der wir behaupten, daß sie die gewünschte Minimaleigenschaft hat, d. h., es ist:

$$(11) \quad \iiint_{\mathfrak{B}} \varphi_{00}(X, Z) \overline{\varphi_{00}(X, Z)} d\omega = k_{00}$$

$$(11^*) \quad \iiint_{\mathfrak{B}} \varphi_{00}(X, Z) \overline{\varphi_{00}(X, Z)} d\omega = 1$$

(11*) ist klar. (11) sieht man so ein: für genügend großes n ist

$$(12) \quad \begin{aligned} & \iiint_{\mathfrak{B}_l} \varphi_{00}(X, Z) \overline{\varphi_{00}(X, Z)} d\omega \\ & \leq \iiint_{\mathfrak{B}_l} v_n(X, Z) \overline{v_n(X, Z)} d\omega + \varepsilon \text{ (Volumen von } \mathfrak{B}_l) \\ & \leq \iiint_{\mathfrak{B}} v_n(X, Z) \overline{v_n(X, Z)} d\omega + \varepsilon \text{ (Volumen von } \mathfrak{B}) \\ & \leq k_{00} + \varepsilon' + \varepsilon \text{ (Volumen von } \mathfrak{B}), \end{aligned}$$

wobei ε und ε' vorgegebene kleine Zahlen sind. Da die Ungleichung (12) für jedes \mathfrak{B}_l , ε und ε' richtig ist, so ist sie auch für \mathfrak{B} richtig. Da andererseits:

$$\iiint_{\mathfrak{B}} \varphi_{00} \overline{\varphi_{00}} d\omega \geq k_{00}$$

folgt die Relation (11).

Der Beweis der Existenz der weiteren $\varphi_{\nu\mu}$ verläuft, wie gesagt, dem soeben durchgeführten für φ_{00} vollkommen analog.

Es ist nunmehr noch zu zeigen, daß das System $\varphi_{\nu\mu}$ die Relation (6b) befriedigt. Um dies einzusehen, bildet man den Ausdruck

$$(13) \quad \psi \equiv \alpha_1 \varphi_{\nu\mu} + \alpha_2 \varphi_{\nu'\mu'} \quad (\nu'\mu') > (\nu\mu),$$

wobei zwischen a_1 und a_2 die Relation

$$(14) \quad |a_1|^2 + |a_2|^2 = 1$$

besteht. ψ befriedigt offenbar die ersten $(\nu^0 \mu^0)$ Orthogonalrelationen [wegen $(\nu^0 \mu^0)$ vgl. Fußnote¹]

$$(15) \quad \iiint_{\mathfrak{B}} \psi \bar{\varphi}_{00} d\omega = 0, \quad \iiint_{\mathfrak{B}} \psi \bar{\varphi}_{10} d\omega = 0, \quad \dots, \quad \iiint_{\mathfrak{B}} \psi \bar{\varphi}_{\nu^0 \mu^0} d\omega = 0,$$

$$(16) \quad \iiint_{\mathfrak{B}} \psi \bar{\psi} d\omega = 1.$$

Es ist deshalb:

$$(17) \quad \iiint_{\mathfrak{B}} \psi \bar{\psi} d\omega \geq k_{\nu\mu}.$$

Andererseits ist

$$(18) \quad \iiint_{\mathfrak{B}} \psi \bar{\psi} d\omega = |a_1|^2 k_{\nu\mu} + |a_2|^2 k_{\nu'\mu'} + a_1 \bar{a}_2 A + a_1 \bar{a}_2 \bar{A} =$$

$$(19) \quad = k_{\nu\mu} + (k_{\nu'\mu'} - k_{\nu\mu}) |a_2|^2 + a_1 \bar{a}_2 A + \bar{a}_1 a_2 \bar{A}$$

(wo $A \equiv A_1 + iA_2 \equiv \iiint_{\mathfrak{B}} \varphi_{\nu\mu} \bar{\varphi}_{\nu'\mu'} d\omega$ ist) gleich. Setzt man nun für a_2 genügend kleine reelle Werte ein, entwickelt man gemäß (14) a_1 in der Umgebung der Stelle $a_2 = 0$ und setzt man die erhaltene Potenzreihe an Stelle von a_1 in (19) ein, so wird

$$(20) \quad \iiint_{\mathfrak{B}} \psi \bar{\psi} d\omega = k_{\nu\mu} + 2a_2 A_1 + a_2^2 [k_{\nu'\mu'} - k_{\nu\mu}] + a_2^3 A_1 + \dots$$

A_1 muß Null sein, sonst hätten wir einen Widerspruch gegen die Ungleichung (17). Ebenso zeigt man, daß A_2 gleich Null ist.

Es sollen nunmehr einige weitere Eigenschaften des Systems $\varphi_{\nu\mu}$ abgeleitet werden.

I. Die Größen $k_{\nu\mu}$ stellen offenbar eine nicht abnehmende Folge dar. Es kann aber geschehen, daß mehrere $k_{\nu\mu}$ gleich sind. Es soll nunmehr gezeigt werden, daß nicht unendlich viele $k_{\nu\mu}$ einander gleich sein können.

Die Funktionen $\frac{\varphi_{\nu\mu}(X, Z)}{\sqrt{k_{\nu\mu}}}$ sind ein normiertes Orthogonalsystem in \mathfrak{B} . In der unter ³) genannten Arbeit (S. 180) wurde gezeigt, daß der Kern eines normierten Orthogonalsystems, nämlich die Reihe

$$(21) \quad \sum_{(\nu\mu)=1}^{(\nu\mu)=\infty} \left| \frac{\varphi_{\nu\mu}(X, Z)}{\sqrt{k_{\nu\mu}}} \right|^2$$

in jedem abgeschlossenen inneren Teilbereich von \mathfrak{B} gleichmäßig konvergiert. Insbesondere ist er im abgeschlossenen \mathfrak{R} beschränkt, und es exi-

stiert das Integral

$$(22) \quad \iiint_{\mathfrak{B}} \sum_{(\nu\mu)=1}^{(\nu\mu)=\infty} \frac{\varphi_{\nu\mu}(X, Z) \overline{\varphi_{\nu\mu}(X, Z)}}{k_{\nu\mu}} d\omega \\ \equiv \sum_{(\nu\mu)=1}^{(\nu\mu)=\infty} \frac{1}{k_{\nu\mu}} \iiint_{\mathfrak{B}} \varphi_{\nu\mu}(X, Z) \overline{\varphi_{\nu\mu}(X, Z)} d\omega \equiv \sum_{(\nu\mu)=1}^{(\nu\mu)=\infty} \frac{1}{k_{\nu\mu}},$$

was aus (7 a) folgt. Daraus schließt man, daß nicht unendlich viele $k_{\nu\mu}$ gleich sein können, w. z. b. w.

II. Das System $\frac{\varphi_{\nu\mu}}{\sqrt{k_{\nu\mu}}}$ ist die Lösungsfolge der folgenden Maximumaufgaben:

Es ist diejenige Funktion $w(X, Z)$ zu bestimmen, die in das Integral

$$(23) \quad \iiint_{\mathfrak{B}} w(X, Z) \overline{w(X, Z)} d\omega$$

eingesetzt, unter den Nebenbedingungen

$$(24) \quad \iiint_{\mathfrak{B}} w(X, Z) \frac{\overline{\varphi_{00}(X, Z)}}{\sqrt{k_{00}}} d\omega = 0, \quad \iiint_{\mathfrak{B}} w(X, Z) \frac{\overline{\varphi_{10}(X, Z)}}{\sqrt{k_{10}}} d\omega = 0, \quad \dots, \\ \iiint_{\mathfrak{B}} w(X, Z) \frac{\overline{\varphi_{\nu^0\mu^0}(X, Z)}}{\sqrt{k_{\nu^0\mu^0}}} d\omega = 0, \quad \iiint_{\mathfrak{B}} w(X, Z) \overline{w(X, Z)} d\omega = 1,$$

(23) zum Maximum macht.

III. Das System $\varphi_{\nu\mu}$ ist im Bereiche \mathfrak{B} vollständig, d. h. jede im offenen Bereiche \mathfrak{B} reguläre Funktion $h(X, Z)$ mit endlichem $\iiint_{\mathfrak{B}} h(X, Z) \overline{h(X, Z)} d\omega$ läßt sich nach $\varphi_{\nu\mu}(X, Z)$ im Innern von \mathfrak{B} entwickeln.

Um dies zu zeigen, sei vorausgesetzt, daß es eine Funktion $H(X, Z)$ mit den angegebenen Eigenschaften gibt, die dies nicht tut.

Bezeichnet man mit $E_{\nu\mu}$:

$$(25) \quad E_{\nu\mu} \equiv \iiint_{\mathfrak{B}} H(X, Z) \overline{H(X, Z)} d\omega - \sum_{(nm)=1}^{(nm)=(\nu\mu)} \frac{\left| \iiint_{\mathfrak{B}} H(X, Z) \overline{\varphi_{nm}(X, Z)} d\omega \right|^2}{k_{nm}},$$

so muß $\lim_{(\nu\mu) \rightarrow \infty} E_{\nu\mu}$ entweder 0, oder größer als Null sein.

1. Ist nun $\lim_{(\nu\mu) \rightarrow \infty} E_{\nu\mu} = 0$, so kann man zu jedem vorgegebenen ε ein endliches $(\nu\mu)$ finden, derart, daß

$$(26) \quad E_{\nu\mu} < \varepsilon$$

ist. Da aber

$$(27) \quad \iiint_{\mathfrak{B}} \left| H(X, Z) - \sum_{(nm)=1}^{(nm)=(\nu\mu)} \frac{\varphi_{nm} \iiint_{\mathfrak{B}} H(X, Z) \overline{\varphi_{nm}(X, Z)} d\omega}{k_{nm}} \right|^2 d\omega$$

$$= \iiint_{\mathfrak{B}} H(X, Z) \overline{H(X, Z)} d\omega - \sum_{(nm)=1}^{(nm)=(\nu\mu)} \frac{\left| \iiint_{\mathfrak{B}} H(X, Z) \overline{\varphi_{nm}(X, Z)} d\omega \right|^2}{k_{nm}} < \varepsilon$$

ist, so folgt nach dem in der Fußnote 3) erwähnten Satz, daß

$$(28) \quad \left| H(X, Z) - \sum_{(nm)=1}^{(nm)=(\nu\mu)} \frac{\varphi_{nm}(X, Z) \iiint_{\mathfrak{B}} H(X, Z) \overline{\varphi_{nm}(X, Z)} d\omega}{k_{nm}} \right| \leq \frac{\varepsilon \sqrt{2}}{\pi \gamma^2(X, Z)}$$

ist, was ein Widerspruch gegen unsere Annahme ist.

2. Ist $\lim_{(\nu\mu) \rightarrow \infty} E_{\nu\mu} > 0$, so ist $\frac{1}{E_{\nu\mu}}$ beschränkt und die Funktion

$$(29) \quad \theta_{\nu\mu}(X, Z) \equiv \frac{1}{\sqrt{E_{\nu\mu}}} \left[H(X, Z) - \sum_{(nm)=1}^{(nm)=(\nu\mu)} \frac{\varphi_{nm}(X, Z) \iiint_{\mathfrak{B}} H(X, Z) \overline{\varphi_{nm}(X, Z)} d\omega}{k_{nm}} \right]$$

genügt den $(\nu\mu)$ ersten Orthogonalrelationen

$$(30) \quad \iiint_{\mathfrak{B}} \theta_{\nu\mu}(X, Z) \frac{\overline{\varphi_{00}(X, Z)}}{\sqrt{k_{00}}} d\omega = 0, \quad \iiint_{\mathfrak{B}} \theta_{\nu\mu}(X, Z) \frac{\overline{\varphi_{10}(X, Z)}}{\sqrt{k_{10}}} d\omega = 0, \quad \dots,$$

$$\iiint_{\mathfrak{B}} \theta_{\nu\mu}(X, Z) \frac{\overline{\varphi_{\nu\rho\mu^{\circ}}(X, Z)}}{\sqrt{k_{\nu\rho\mu^{\circ}}}} d\omega = 0, \quad \iiint_{\mathfrak{B}} \theta_{\nu\mu}(X, Z) \overline{\theta_{\nu\mu}(X, Z)} d\omega = 1.$$

Nach II ist

$$(31) \quad \iiint_{\mathfrak{B}} \theta_{\nu\mu}(X, Z) \overline{\theta_{\nu\mu}(X, Z)} d\omega \leq \frac{1}{k_{\nu\mu}}.$$

Da $\lim_{(\nu\mu) \rightarrow \infty} \frac{1}{k_{\nu\mu}} = 0$ ist, so folgt nach dem oben benutzten Satz, daß

$\lim_{(\nu\mu) \rightarrow \infty} \theta_{\nu\mu}(X, Z) = 0$ ist, so daß auch in diesem Falle ein Widerspruch gegen die ursprüngliche Annahme entsteht.

IV. Es soll gezeigt werden, daß das Funktionselement von $\varphi_{\nu\mu}$ die Gestalt

$$(32) \quad \sum_{(nm)=1}^{(nm)=\infty} O_{\nu\mu, nm} a_{nm} X^n Z^m,$$

hat, wo

$$(33) \quad \parallel O_{\nu\mu, nm} \parallel$$

eine orthogonale Matrix ist. (Wegen der Definition von a_{nm} vgl. S. 399 f.) Nach dem Hartogsschen Satz⁴⁾ läßt sich $\varphi_{\nu\mu}$ in \mathfrak{R} in eine Diagonalpotenzreihe entwickeln. Da sowohl das System $\varphi_{\nu\mu}$ wie das System $a_{\nu\mu} X^\nu Z^\mu$ in \mathfrak{R} orthogonal ist, ist die Matrix (33) orthogonal.

V. Ganz ähnlich erschließt man, daß, wenn sich \mathfrak{B} vermittelt eines Funktionenpaares $V(X, Z)$ und $W(X, Z)$ ein-eindeutig auf einen Kreisbereich des $V-W$ -Raumes abbilden läßt und $|V|$ und $|W|$ in \mathfrak{R} kleiner wie 1 sind, so kann man $\varphi_{\nu\mu}$ ebenfalls in der Gestalt

$$(34) \quad \varphi_{\nu\mu} = \sum_{(nm)=1}^{(nm)=\infty} \Omega_{\nu\mu, nm} V^n W^m \frac{\partial(V, W)}{\partial(X, Z)}$$

darstellen, wo

$$(35) \quad \|\Omega_{\nu\mu, nm}\|$$

wiederum eine orthogonale Matrix bedeutet.

Wir gehen zum Beweis der ersten Behauptung über. Nach III läßt sich jede in einem offenen Bereiche reguläre Funktion $G(X, Z)$ mit endlichen $\iiint_{\mathfrak{B}} G(X, Z) \overline{G(X, Z)} d\omega$ nach dem vollständigen Orthogonalfunktionensystem $\varphi_{\nu\mu}$ entwickeln. Es ist somit in \mathfrak{B}

$$(36) \quad G(X, Z) = \sum_{(\nu\mu)=1}^{(\nu\mu)=\infty} \varphi_{\nu\mu}(X, Z) \frac{1}{k_{\nu\mu}} \iiint_{\mathfrak{B}} G \overline{\varphi_{\nu\mu}} d\omega$$

und ferner in \mathfrak{R}

$$(37) \quad G(X, Z) = \sum_{(\nu\mu)=1}^{(\nu\mu)=\infty} \varphi_{\nu\mu}(X, Z) \iiint_{\mathfrak{R}} G \overline{\varphi_{\nu\mu}} d\omega.$$

Da die Entwicklung nach den Orthogonalfunktionen eindeutig ist, folgt, daß

$$(38) \quad \iiint_{\mathfrak{B}} G \overline{\varphi_{\nu\mu}} d\omega = k_{\nu\mu} \iiint_{\mathfrak{R}} G \overline{\varphi_{\nu\mu}} d\omega$$

ist. Da ferner sowohl G wie φ_{nm} im abgeschlossenen \mathfrak{R} durch eine Diagonalpotenzreihe darstellen läßt, folgt weiter, daß (38) dem Ausdruck

$$(39) \quad k_{\nu\mu} \iiint_{\mathfrak{R}} \left(\sum_{(nm)=1}^{(nm)=\infty} \frac{b_{nm}}{a_{nm}} a_{nm} X^n Z^m \right) \overline{\left(\sum_{(nm)=1}^{(nm)=\infty} O_{\nu\mu, nm} a_{nm} X^n Z^m \right)} d\omega \\ = k_{\nu\mu} \sum_{(nm)=1}^{(nm)=\infty} \frac{b_{nm}}{a_{nm}} \overline{O_{\nu\mu, nm}}$$

⁴⁾ F. Hartogs, „Zur Theorie der analytischen Funktionen mehrerer unabhängiger Veränderlichen, insbesondere über die Darstellung derselben durch Reihen, welche nach Potenzen einer Veränderlichen fortschreiten“, Math. Annalen 62 (1906), S. 1–88 (S. 2).

gleich ist. Aus der Existenz von

$$(40) \quad \iiint_{\mathfrak{B}} G(X, Z) \overline{G(X, Z)} d\omega = \sum_{(\nu\mu)=1}^{(\nu\mu)=\infty} \frac{|\iiint_{\mathfrak{B}} G \overline{\varphi}_{\nu\mu} d\omega|^2}{k_{\nu\mu}}$$

folgt die von

$$(41) \quad \lim_{(NM) \rightarrow \infty} \sum_{(\nu\mu)=1}^{(\nu\mu)=(NM)} \frac{|\iiint_{\mathfrak{B}} G \overline{\varphi}_{\nu\mu} d\omega|^2}{k_{\nu\mu}}.$$

Setzt man in (41) für jedes $\iiint_{\mathfrak{B}} G \overline{\varphi}_{\nu\mu} d\omega$ den in (39) erhaltenen Wert ein, so geht (41) in

$$(42) \quad \lim_{(NM) \rightarrow \infty} \sum_{(\nu\mu)=1}^{(\nu\mu)=(NM)} k_{\nu\mu} \left| \sum_{(\varrho, \sigma)=1}^{(\varrho, \sigma)=\infty} \frac{b_{\varrho\sigma}}{a_{\varrho\sigma}} \overline{O}_{\nu\mu, \varrho\sigma} \right| \\ = \lim_{(NM) \rightarrow \infty} \sum_{\substack{(\varrho, \sigma)=1 \\ (\varrho', \sigma')=1}}^{\substack{(\varrho, \sigma)=\infty \\ (\varrho', \sigma')=\infty}} \frac{b_{\varrho\sigma}}{a_{\varrho\sigma}} \overline{\frac{b_{\varrho'\sigma'}}{a_{\varrho'\sigma'}}} \left(\sum_{(\nu\mu)=1}^{(\nu\mu)=(NM)} k_{\nu\mu} O_{\nu\mu, \varrho\sigma} \overline{O}_{\nu\mu, \varrho'\sigma'} \right)$$

über. Bezeichnet man

$$\sum_{(\nu\mu)=1}^{(\nu\mu)=(NM)} \nu\mu O_{\nu\mu, \varrho\sigma} \overline{O}_{\nu\mu, \varrho'\sigma'} \quad \text{mit} \quad A_{\varrho\sigma, \varrho'\sigma'}^{(NM)}$$

und

$$\sum_{\substack{(\varrho, \sigma)=1 \\ (\varrho', \sigma')=1}}^{\substack{(\varrho, \sigma)=\infty \\ (\varrho', \sigma')=\infty}} A_{\varrho\sigma, \varrho'\sigma'}^{(NM)} \frac{b_{\varrho\sigma}}{a_{\varrho\sigma}} \overline{\frac{b_{\varrho'\sigma'}}{a_{\varrho'\sigma'}}} \quad \text{mit} \quad f_{(NM)} \left(\frac{b_{\varrho\sigma}}{a_{\varrho\sigma}}, \overline{\frac{b_{\varrho'\sigma'}}{a_{\varrho'\sigma'}}} \right)$$

so wird (41)

$$(43) \quad \lim_{(NM) \rightarrow \infty} f_{(NM)} \left(\frac{b_{\nu\mu}}{a_{\nu\mu}}, \overline{\frac{b_{\nu'\mu'}}{a_{\nu'\mu'}}} \right),$$

womit die erste Behauptung des Satzes bewiesen ist.

Wir gehen nunmehr zum Beweis der zweiten Behauptung über.

Ist $G(X, Z)$ in \mathfrak{R} regulär, so läßt es sich dort nach (37) in der Form

$$(44) \quad \sum_{(\nu\mu)=1}^{(\nu\mu)=\infty} \frac{\varphi_{\nu\mu}(X, Z)}{\sqrt{k_{\nu\mu}}} \sqrt{k_{\nu\mu}} \iiint_{\mathfrak{R}} G \overline{\varphi}_{\nu\mu} d\omega$$

darstellen. (44) ist in \mathfrak{B} regulär, denn nach der Schwarzschen Ungleichung ist für alle X, Z in \mathfrak{B}

$$(45) \quad \left| \sum_{(\nu\mu)=1}^{(\nu\mu)=(NM)} \frac{\varphi_{\nu\mu}(X, Z)}{\sqrt{k_{\nu\mu}}} \sqrt{k_{\nu\mu}} \iiint_{\mathfrak{R}} G \overline{\varphi}_{\nu\mu} d\omega \right| \\ \leq \sqrt{\left[\sum_{(\nu\mu)=1}^{(\nu\mu)=(NM)} \frac{|\varphi_{\nu\mu}(X, Z)|^2}{k_{\nu\mu}} \right] \left[\sum_{(\nu\mu)=1}^{(\nu\mu)=(NM)} k_{\nu\mu} \left| \iiint_{\mathfrak{R}} G \overline{\varphi}_{\nu\mu} d\omega \right|^2 \right]}.$$

Es ist nun der erste Faktor des Wurzelausdruckes nach einem in der unter ³⁾ zitierten Arbeit angegebenen Satze in jeden inneren von \mathfrak{B} beschränkt und es gilt für jedes (NM) die Ungleichung

$$(46) \quad \sqrt{\sum_{(\nu\mu)=1}^{(\nu\mu)=(NM)} \frac{|\varphi_{\nu\mu}(X, Z)|^2}{k_{\nu\mu}}} \leq \frac{\sqrt{2}}{\pi\gamma^2(X, Z)}.$$

Der zweite Faktor $\sum_{(\nu\mu)=1}^{(\nu\mu)=(NM)} k_{\nu\mu} \left| \iiint_{\mathfrak{B}} G \bar{\varphi}_{\nu\mu} d\omega \right|^2$ hat nach (39) und (42)

den Wert $f_{(NM)} \left(\frac{b_{\nu\mu}}{a_{\nu\mu}}, \frac{\bar{b}_{\nu\mu}}{a_{\nu\mu}} \right)$, nach Voraussetzung darf der Grenzübergang zu $(NM) \rightarrow \infty$ vollzogen werden.

Daraus erschließt man die gleichmäßige Konvergenz von (44) in jedem abgeschlossenen inneren Teilbereich von \mathfrak{B} , was die zweite Behauptung nach sich zieht.

§ 2.

Sei $V(X, Z)$, $W(X, Z)$ ein normiertes⁵⁾ Funktionenpaar, welches in einem Bereiche \mathfrak{B} regulär ist. Das Funktionenpaar V, W bilde den Bereich \mathfrak{B} umkehrbar eindeutig auf einen Kreisbereich des V - W -Raumes ab. Nach dem Hartogschen Satz lassen sich $V(X, Z)$ und $W(X, Z)$ in dem (in § 1 definierten) Kreisbereiche \mathfrak{R} in eine Diagonalspotenzreihe

$$(47) \quad \begin{aligned} V(X, Z) &\equiv X + \sum_{(\nu\mu)=(20)}^{(\nu\mu)=\infty} c_{\nu\mu} X^\nu Z^\mu \\ W(X, Z) &\equiv Z + \sum_{(\nu\mu)=(20)}^{(\nu\mu)=\infty} e_{\nu\mu} X^\nu Z^\mu \end{aligned}$$

entwickeln. Die Funktionaldeterminante $\frac{\partial(V, W)}{\partial(X, Z)}$ ⁶⁾ hat in \mathfrak{R} die Gestalt:

⁵⁾ Unter einem normierten Funktionenpaar wird ein Paar verstanden, das die Eigenschaft besitzt, daß

$$\begin{aligned} V(0, 0) &= 0, & \left[\frac{\partial V(X, Z)}{\partial X} \right]_{\substack{X=0 \\ Z=0}} &= 1, & \left[\frac{\partial V(X, Z)}{\partial Z} \right]_{\substack{X=0 \\ Z=0}} &= 0, \\ W(0, 0) &= 0, & \left[\frac{\partial W(X, Z)}{\partial X} \right]_{\substack{X=0 \\ Z=0}} &= 0, & \left[\frac{\partial W(X, Z)}{\partial Z} \right]_{\substack{X=0 \\ Z=0}} &= 1 \end{aligned}$$

ist.

⁶⁾ Es wird dabei die Voraussetzung gemacht, daß die Funktion $\frac{\partial(V, W)}{\partial(X, Z)}$ sich gleichmäßig in jedem ganz im Innern von \mathfrak{B} gelegenen Teilbereich durch Polynome approximieren läßt.

$$(48) \quad \frac{\partial(V, W)}{\partial(X, Z)} \equiv 1 + \sum_{(\nu\mu)=(10)}^{(\nu\mu)=\infty} d_{\nu\mu} X^\nu Z^\mu.$$

Es gilt dann der

Satz. Die Koeffizienten $d_{\nu\mu}$ von (48) genügen der Ungleichung

$$|d_{\nu\mu}| \leq \text{konst. } a_{\nu\mu}$$

$$(49) \quad a_{\nu\mu} \equiv \frac{1}{\sqrt{\iiint_{\mathfrak{B}} |X^\nu Z^\mu|^2 d\omega}}$$

Beweis. Um die Formel (49) zu beweisen, betrachten wir die zu dem Bereich \mathfrak{B} gehörige Hermitesche Form

$$(50) \quad \sum_{\substack{(\nu\mu)=1 \\ (\nu'\mu')=1}}^{(\nu\mu)=\infty \\ (\nu'\mu')=\infty} \alpha_{\nu\mu} \bar{\alpha}_{\nu'\mu'} \iiint_{\mathfrak{B}} X^\nu Z^\mu \overline{X^{\nu'} Z^{\mu'}} d\omega.$$

Ohne die Allgemeinheit zu beschränken, können wir voraussetzen, daß \mathfrak{B} sich ganz im Innern der Einheitshyperkugel befindet. (50) ist daher eine *vollstetige, positiv-definite Hermitesche Form*. Nach dem bekannten Hilbertschen Satz läßt sich (50) in der Form:

$$(51) \quad \sum_{(\nu\mu)=1}^{(\nu\mu)=\infty} K_{\nu\mu} \left(\sum_{(nm)=1}^{(nm)=\infty} o_{\nu\mu, nm} \alpha_{nm} \right) \overline{\left(\sum_{(nm)=1}^{(nm)=\infty} o_{\nu\mu, nm} \alpha_{nm} \right)}$$

schreiben. Um eine für den Beweis nötige Abschätzung zu gewinnen, gehen wir wieder von einer (von der in § 1 verschiedenen) Variationsaufgabe aus.

Es soll nämlich das Minimum der Form (50) bei der Nebenbedingung

$$(52) \quad \alpha_{\nu\mu} = 1$$

bei beliebigen, aber fest gewählten $(\nu\mu)$ bestimmt werden. Führt man die Vektoren \mathbf{v}_{nm} (im Hilbertschen Raum) mit den Komponenten

$$(53) \quad \mathbf{v}_{nm} \equiv \left(\frac{o_{00, nm}}{\sqrt{K_{00}}}, \frac{o_{10, nm}}{\sqrt{K_{10}}}, \frac{o_{01, nm}}{\sqrt{K_{01}}}, \dots \right)$$

ein und bezeichnet man mit $*$ das innere Produkt, so läßt sich zeigen: Dasjenige System von Größen $\beta_{nm}^{(\nu\mu)}$, das die Lösung der oben angegebenen Variationsaufgabe darstellt, läßt sich in der Gestalt

$$(54) \quad \beta_{nm}^{(\nu\mu)} = \frac{(\mathbf{v}_{\nu\mu} * \overline{\mathbf{v}_{nm}})}{(\mathbf{v}_{\nu\mu} * \mathbf{v}_{\nu\mu})}$$

schreiben. Das Minimum $\lambda_{\nu\mu}$ von (50) bei der Nebenbedingung (52) wird

$$(55) \quad \lambda_{\nu\mu} = \frac{1}{(\mathbf{v}_{\nu\mu} * \mathbf{v}_{\nu\mu})}$$

gleich. Betreffs der Ableitung Formeln (54) und (55) sei auf die Arbeit „Über die harmonischen Funktionen der Ebene und des Raumes nach Orthogonalfunktionen“⁷⁾ § 2 verwiesen.

Es soll nunmehr für $\lambda_{\nu\mu}$ eine untere Schranke angegeben werden. Nach dem Hilfssatz 2 der unter ³⁾ zitierten Arbeit folgt:

$$(56) \quad \lambda_{\nu\mu} \geq \iiint_{\mathfrak{B}} |X^\nu Z^\mu|^2 d\omega = \frac{1}{a_{\nu\mu}^2}.$$

Wie ferner in der zweiten unter ³⁾ zitierten Arbeit Sätze 3 und 4 gezeigt wurde, besitzt die Funktionaldeterminante $\frac{\partial(V,W)}{\partial(X,Z)}$ eines Funktionenpaares, die den Bereich \mathfrak{B} auf einem Kreisbereich abbildet, folgende Eigenschaft: Setzt man in das Integral

$$(57) \quad \iiint_{\mathfrak{B}} |v(X,Z)|^2 d\omega$$

die Funktionaldeterminante $\frac{\partial(V,W)}{\partial(X,Z)}$ an Stelle von v ein, so wird dieses zum Minimum gemacht, falls zum Vergleiche alle in \mathfrak{B} regulären Funktionen mit der Nebenbedingung

$$(58) \quad v(0,0) = 1$$

herangezogen werden. Da nach der Voraussetzung $\frac{\partial(V,W)}{\partial(X,Z)}$ sich in jedem

ganz im Innern von \mathfrak{B} gelegenen Teilbereiche gleichmäßig durch Polynome approximieren läßt, so sind die Koeffizienten des Funktionselementes von

$\frac{\partial(V,W)}{\partial(X,Z)} \equiv D(X,Z)$ nichts anderes wie $\beta_{nm}^{(00)}$ (d. h. es wird:

$$(59) \quad D(X,Z) \equiv 1 + \sum_{(nm)=(10)}^{(nm)=\infty} d_{nm} X^n Z^m = 1 + \sum_{(nm)=(10)}^{(nm)=\infty} \beta_{nm}^{(00)} X^n Z^m.$$

Nach (54) folgt weiter

$$(60) \quad |d_{nm}| = |\beta_{nm}^{(00)}| = \left| \frac{(v_{00} * \overline{v_{nm}})}{(v_{00} * v_{00})} \right| \leq \frac{\sqrt{(v_{00} * v_{00})(v_{nm} * \overline{v_{nm}})}}{(v_{00} * v_{00})} \\ = \sqrt{\frac{(v_{nm} * \overline{v_{nm}})}{(v_{00} * v_{00})}}.$$

Nach (55) ist nun somit

$$(61) \quad |d_{nm}| \leq \frac{a_{nm}}{\sqrt{(v_{00} * v_{00})}} \leq \frac{a_{nm}}{\sqrt{\text{Volumen des Kreisbereiches in VW Raum}}} \equiv \frac{a_{nm}}{\text{const.}},$$

w. z. b. w.

⁷⁾ Math. Annalen 86 (1922), S. 238—271.

Über eine zahlentheoretische Anwendung von Modulfunktionen zweier Veränderlicher.

Von

Fritz Götzky in Bieber.

Die *Modulfunktionen von mehreren Variablen* hat Herr Hilbert eingeführt. Die ersten Arbeiten darüber sind von Herrn Blumenthal veröffentlicht worden. In letzter Zeit hat sich, und zwar besonders vom zahlentheoretischen Standpunkt aus, Herr Hecke mit diesen Funktionen befaßt, vor allem mit den dem allgemeinen reellen quadratischen Zahlkörper zugeordneten Modulfunktionen. Schon auf diese letzteren scheinen manche fundamentalen Sätze über elliptische Modulfunktionen nicht verallgemeinerungsfähig zu sein. Es ist daher wohl zweckmäßig, zunächst einmal in speziellen quadratischen Zahlkörpern zu arbeiten, in denen diese Funktionen noch besonders einfache Eigenschaften haben, um dann aus den Ergebnissen Schlüsse auf die allgemeineren Modulfunktionen zu ziehen, für deren Untersuchung so auch die arithmetischen Sonderheiten spezieller Zahlkörper nutzbar werden.

In dieser Arbeit wird der Körper $k(\sqrt{5})$, der quadratische Zahlkörper mit der Diskriminante 5, zugrunde gelegt. Das Ziel ist der Beweis des folgenden arithmetischen Satzes mit Hilfe von Modulfunktionen dieses Körpers.

Satz I. *Jede ungerade ganze total positive Zahl aus $k(\sqrt{5})$ ist so oft als Summe von vier ganzen Quadratzahlen desselben Körpers darstellbar, als die achtfache Summe der Normen der Idealeiler der Zahl beträgt. Jede gerade ganze total positive Zahl aus $k(\sqrt{5})$ ist so oft als Summe von vier ganzen Quadratzahlen des Körpers darstellbar, als die 12-fache Summe der Normen der geraden Idealeiler, vermindert um die 24-fache Summe der Normen der ungeraden Idealeiler, beträgt.*

Im ersteren Fall einer ungeraden Zahl, in welchem genaue Analogie zum Jacobischen Satz im rationalen Zahlkörper besteht, ist der Satz be-

reits von Herrn Kirmse¹⁾ auf anderem Wege bewiesen. Man kann leicht einsehen, daß in keinem weiteren quadratischen Zahlkörper jede ganze total positive Zahl als Quadratsumme von vier ganzen Körperzahlen darstellbar ist. Dem Beweis von Satz I liegt der folgende Gedankengang zugrunde.

Satz I, zunächst das Ergebnis empirischer Rechnung, gibt Veranlassung zur Aufstellung der vermutlichen Identität

$$(1) \quad \left(\sum_{\mu} q^{\mu^2} q'^{\mu'^2} \right)^4 = 1 + \sum_{\mu > 0} A_{\mu} q^{\mu} q'^{\mu'}$$

in zwei unabhängigen komplexen Variablen q und q' ($|q| < 1$, $|q'| < 1$), wobei μ links alle ganzen, rechts alle total positiven ganzen Zahlen aus $k(\sqrt{5})$ durchläuft, μ' die Konjugierte zu μ bedeutet und A_{μ} diejenige arithmetische Funktion ist, die nach Satz I die Anzahl der Zerlegungen von μ in vier Quadrate angibt. Auf beiden Seiten dieser Formel treten Reihen vom Typus

$$\sum_{\mu > 0} a_{\mu} q^{\mu} q'^{\mu'}$$

auf, die Herr Hecke als Potenzreihen im Körper $k(\sqrt{5})$ bezeichnet hat. Da jede Funktion auf höchstens eine Art in Form einer solchen Potenzreihe darstellbar ist, so ist die Identität (1) völlig gleichbedeutend mit dem arithmetischen Satz I. Von dem Quotienten

$$\frac{1 + \sum_{\mu > 0} A_{\mu} q^{\mu} q'^{\mu'}}{\left(\sum_{\mu} q^{\mu^2} q'^{\mu'^2} \right)^4},$$

der aus den beiden Seiten der Formel (1) gebildet ist, wird nun nachgewiesen, daß er in seiner Abhängigkeit von τ und τ' , wenn man $q = e^{\frac{\pi i \tau}{\sqrt{5}}}$, $q' = e^{\frac{-\pi i \tau'}{\sqrt{5}}}$ setzt, eine *Modulfunktion* darstellt, d. h. daß er bei Ausübung aller Substitutionen der *Modulgruppe in $k(\sqrt{5})$* nur in endlich viele, und zwar neun weitere Funktionen übergeht. Die so erhaltenen zehn Modulfunktionen erweisen sich sämtlich als *regulär im Fundamentalbereich* dieser Modulgruppe, woraus nach einem allgemeinen Satz ihre Konstanz und damit die Identität (1) folgt.

Die §§ 1 bis 3 handeln von der Theorie der *Modulfunktionen im Körper $k(\sqrt{5})$* . Es wird in § 1 die *Modulgruppe in $k(\sqrt{5})$* gemäß der Hilbertschen Definition²⁾ aufgestellt und ihre Erzeugung durch drei spezielle

¹⁾ Joh. Kirmse, Zur Darstellung total positiver Zahlen als Summen von vier Quadraten, Math. Zeitschr. 21 (1924), S. 195—202.

²⁾ Vgl. Blumenthal, Über Modulfunktionen von mehreren Veränderlichen, Math. Annalen 56 (1903), S. 510.

Modulsubstitutionen durchgeführt. Der in § 2 eingeführte *Fundamentalebereich* \mathfrak{M} , der keinen Spezialfall des von Herrn Blumenthal für die allgemeineren Modulgruppen aufgestellten Fundamentalebereichs darstellt, ist wohl die richtige Verallgemeinerung des bekannten Fundamentalebereichs³⁾ der gewöhnlichen Modulgruppe. § 3 bringt einen allgemeinen Satz über *reguläre Modulfunktionen* des Körpers $k(\sqrt{5})$. In den §§ 4 und 6 werden die Transformationsformeln für die in der zu beweisenden Identität (1) auftretenden Reihen abgeleitet, wobei für die in der rechten Seite enthaltenen Funktionen ein von Herrn Hecke in seiner Arbeit „Analytische Funktionen und algebraische Zahlen. II. Teil⁴⁾“ angewandtes Verfahren benutzt wird. Die in § 5 ausgeführten Abschätzungen der *Thetafunktionen*, die durch die Wahl des Fundamentalebereiches \mathfrak{M} ermöglicht werden, besagen übrigens, daß diese Funktionen im ganzen Definitionsbereich, von trivialen Faktoren abgesehen, absolut genommen oberhalb einer festen positiven Schranke liegen, eine Eigenschaft, die man den elliptischen Thetafunktionen in ihrer Produktdarstellung ohne weiteres ansieht. In § 7 wird der Beweis von Satz I zu Ende geführt, wobei sich ein Zusammenhang zwischen der Dedekindschen η -Funktion aus der Theorie der elliptischen Modulfunktionen und Funktionen des Körpers $k(\sqrt{5})$ ergibt. Erwähnt sei noch, daß zu den Funktionen des § 6, die auf Grund des empirisch gefundenen Satzes I aufgestellt wurden, auch die von Herrn Siegel in seiner Arbeit „Additive Theorie der Zahlkörper. I (Math. Annalen 87 (1922), S. 1—35)“ veröffentlichte approximative Methode führt, die eine Verallgemeinerung der Hardy-Littlewood-Ramanujanschen analytischen Methode auf algebraische Zahlkörper darstellt.

§ 1.

Die Modulgruppe in $k(\sqrt{5})$.

Es sei $k(\sqrt{5})$ der quadratische Zahlkörper mit der Diskriminante 5. Mit $\alpha, \beta, \gamma, \delta$ seien ganze Zahlen dieses Körpers, mit $\alpha', \beta', \gamma', \delta'$ die zugehörigen Konjugierten bezeichnet; τ und τ' sollen zwei unabhängige komplexe Variable sein. Unter einer *Modulsubstitution in $k(\sqrt{5})$* sind dann die beiden simultanen Substitutionen

$$\tau_1 = \frac{\alpha\tau + \beta}{\gamma\tau + \delta}, \quad \tau'_1 = \frac{\alpha'\tau' + \beta'}{\gamma'\tau' + \delta'}$$

³⁾ Der Bereich (bis auf abzurechnende Randpunkte): $-\frac{1}{2} \leq \Re \tau < \frac{1}{2}$, $|\tau_1| \geq 1$, $\Im \tau > 0$ ($\Re \tau = \text{Realteil von } \tau$, $\Im \tau = \text{Imaginärteil von } \tau$).

⁴⁾ Abhandlungen aus dem Math. Seminar der Hamburgischen Universität. III. Bd. (1924), S. 213—236. Leipzig: B. G. Teubner.

mit

$$\alpha \delta - \beta \gamma = 1 \quad ^5)$$

zu verstehen, wofür

$$\tau_1, \tau'_1 = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \tau, \tau'$$

oder einfach

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}.$$

eine abgekürzte Bezeichnung sei. Diese Modulsstitutionen bilden in ihrer Gesamtheit eine Gruppe \mathfrak{G} , die *Modulgruppe in $k(\sqrt{5})$* . Das *Produkt* zweier Modulsstitutionen $G_1 = \begin{pmatrix} \alpha_1 & \beta_1 \\ \gamma_1 & \delta_1 \end{pmatrix}$ und $G_2 = \begin{pmatrix} \alpha_2 & \beta_2 \\ \gamma_2 & \delta_2 \end{pmatrix}$ ist (wenn zuerst G_1 und dann G_2 ausgeführt wird)

$$G_2 G_1 = \begin{pmatrix} \alpha_1 \alpha_2 + \gamma_1 \beta_2 & \beta_1 \alpha_2 + \delta_1 \beta_2 \\ \alpha_1 \gamma_2 + \gamma_1 \delta_2 & \beta_1 \gamma_2 + \delta_1 \delta_2 \end{pmatrix}.$$

Satz 1. Jede Modulsstitution in $k(\sqrt{5})$ läßt sich aus den drei speziellen

$$S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad T = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad U = \begin{pmatrix} \varepsilon & 0 \\ 0 & \varepsilon^{-1} \end{pmatrix}$$

zusammensetzen, wo ε die Grundeinheit $\frac{1+\sqrt{5}}{2}$ ist.

Beweis. Jede ganze Zahl ν aus $k(\sqrt{5})$ ist, da $1, \varepsilon = \frac{1+\sqrt{5}}{2}$ eine Körperbasis bilden, auf genau eine Art in der Gestalt

$$\nu = n_1 + n_2 \varepsilon$$

mit ganzen rationalen n_1, n_2 darstellbar. Also ist jede Substitution

$S_\nu = \begin{pmatrix} 1 & \nu \\ 0 & 1 \end{pmatrix}$ aus $S = S_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ und $S_\varepsilon = \begin{pmatrix} 1 & \varepsilon \\ 0 & 1 \end{pmatrix}$ erzeugbar in der Form

$$(2) \quad S_\nu = S_\varepsilon^{n_2} S^{n_1}.$$

Bei einer Substitution der Gestalt $\begin{pmatrix} \alpha & \beta \\ 0 & \delta \end{pmatrix}$ sind α und δ reziproke Einheiten, da $\alpha \delta = 1$; d. h. α und δ sind darstellbar in der Form $\alpha = \pm \varepsilon^n, \delta = \pm \varepsilon^{-n}$ mit ganzem rationalem n . Folglich ist

$$\begin{pmatrix} \alpha & \beta \\ 0 & \delta \end{pmatrix} = \begin{pmatrix} \varepsilon^n & \pm \beta \\ 0 & \varepsilon^{-n} \end{pmatrix} = \begin{pmatrix} \varepsilon^n & 0 \\ 0 & \varepsilon^{-n} \end{pmatrix} \begin{pmatrix} 1 & \beta_1 \\ 0 & 1 \end{pmatrix}$$

⁵⁾ Läßt man für $\alpha \delta - \beta \gamma$ beliebige total positive Einheiten zu, wie es der Hilbertschen Definition genau entsprechen würde, so gelangt man in diesem Körper, wo die Norm der Grundeinheit -1 ist, zu keinen weiteren Substitutionen.

mit einer ganzen Körperzahl β_1 . Bezeichnet man die Substitution $\begin{pmatrix} \varepsilon & 0 \\ 0 & \varepsilon^{-1} \end{pmatrix}$ mit U , so folgt in Verbindung mit (2), wenn $n_1 + n_2 \varepsilon$ die Basisdarstellung der Zahl β_1 ist,

$$(3) \quad \begin{pmatrix} \alpha & \beta \\ 0 & \delta \end{pmatrix} = U^n S_{\beta_1} = U^n S_\varepsilon^{n_2} S^{n_1}.$$

Nun sei $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ eine Modulsstitution mit $\gamma \neq 0$. Für jedes ν ist

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} \alpha\nu - \beta & \alpha \\ \gamma\nu - \delta & \gamma \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & \nu \\ 0 & 1 \end{pmatrix}.$$

Man kann nun, wie gezeigt werden soll, ν als ganze Körperzahl so bestimmen, daß

$$|N(\gamma\nu - \delta)| < |N(\gamma)|,$$

wenn allgemein $N(\kappa)$ die Norm $\kappa\kappa'$ ⁶⁾ der Körperzahl κ bedeutet. Wegen $N(\gamma\nu - \delta) = N(\gamma)N(\nu - \frac{\delta}{\gamma})$ ist diese Ungleichung gleichbedeutend mit

$$\left| N\left(\nu - \frac{\delta}{\gamma}\right) \right| < 1.$$

Ist $\frac{\delta}{\gamma} = r_1 + r_2\varepsilon$ mit rationalen r_1, r_2 , so wähle man $\nu = n_1 + n_2\varepsilon$ mit solchen ganzen rationalen n_1, n_2 , daß

$$|n_1 - r_1| \leq \frac{1}{2}, \quad |n_2 - r_2| \leq \frac{1}{2}.$$

Dann ist $\left| N\left(\nu - \frac{\delta}{\gamma}\right) \right| = |(n_1 - r_1)^2 + (n_1 - r_1)(n_2 - r_2) - (n_2 - r_2)^2| < 1$.

Es ist also jede Modulsstitution $G = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ mit $\gamma \neq 0$ darstellbar in der Form

$$G = G_1 T S_\nu,$$

wobei $G_1 = \begin{pmatrix} \alpha_1 & \beta_1 \\ \gamma_1 & \delta_1 \end{pmatrix}$ mit $|N(\gamma_1)| < |N(\gamma)|$. Daraus folgt die Darstellbarkeit

$$G = \begin{pmatrix} \alpha_0 & \beta_0 \\ 0 & \delta_0 \end{pmatrix} T S_{\nu_\varepsilon} T S_{\nu_{\varepsilon^{-1}}} \dots T S_{\nu_2} T S_{\nu_1}$$

oder wegen (2) und (3)

$$G = U^n S_\varepsilon^{m_\varepsilon} S^{n_\varepsilon} T S_\varepsilon^{m_{\varepsilon^{-1}}} S^{n_{\varepsilon^{-1}}} T \dots T S_\varepsilon^{m_1} S^{n_1},$$

wobei $n_1, m_1, \dots, n_\varepsilon, m_\varepsilon, n$ ganze rationale Zahlen sind.

⁶⁾ Ist κ irgendeine Zahl aus $k(\sqrt{5})$, so soll κ' stets die Konjugierte zu κ bedeuten.

Da

$$S_e = S^{-1}USU^{-1},$$

so läßt sich also jede Modulsstitution aus S , T und U zusammensetzen. Dabei kann man noch Potenzen von S und U mit negativem Exponenten beseitigen vermöge der Relationen

$$\begin{aligned} (UT)^2 &= 1 \quad \text{oder} \quad U^{-1} = TUT, \\ (ST)^3 &= 1 \quad \text{oder} \quad S^{-1} = TSTST. \end{aligned}$$

§ 2.

Der Fundamentalbereich.

Die Real- und Imaginärteile der komplexen Variablen τ und τ' seien bezeichnet mit $\Re\tau = r$, $\Im\tau = s$, $\Re\tau' = r'$, $\Im\tau' = -s'$, also $\tau = r + is$, $\tau' = r' - is'$). Es soll stets

$$s > 0, \quad s' > 0$$

sein. Die vier Größen r, r', s, s' deute man sich als Koordinaten im vierdimensionalen Raum. Die Modulgruppe \mathfrak{G} transformiert den Bereich

$s > 0, s' > 0$ in sich; denn ist $\tau_1, \tau'_1 = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \tau, \tau'$, so ist, wenn $\bar{\tau}, \bar{\tau}'$ konjugiert komplex zu τ, τ' sind,

$$(4) \quad s_1 = \Im \frac{(\alpha\tau + \beta)(\gamma\bar{\tau} + \delta)}{|\gamma\tau + \delta|^2} = \frac{s}{|\gamma\tau + \delta|^2} > 0, \quad \text{wenn } s > 0,$$

und

$$(5) \quad s'_1 = \Im \frac{-(\alpha'\tau' + \beta')(\gamma'\bar{\tau}' + \delta')}{|\gamma'\tau' + \delta'|^2} = \frac{s'}{|\gamma'\tau' + \delta'|^2} > 0, \quad \text{wenn } s' > 0.$$

Irgend zwei Punkte in diesem Raume sollen *äquivalent in bezug auf eine Gesamtheit von Modulsstitutionen* heißen, wenn sie durch eine dieser Substitutionen ineinander überführbar sind. *Äquivalent* schlechthin soll *äquivalent in bezug auf \mathfrak{G}* bedeuten. Eine Punktmenge heiße ein *Fundamentalbereich für eine Gesamtheit von Modulsstitutionen*, wenn sie zu jedem Punkte τ, τ' genau einen in bezug auf diese Substitutionen äquivalenten enthält.

Hilfssatz 1. Ein ganz im Endlichen gelegener abgeschlossener Bereich \mathfrak{B} innerhalb $s > 0, s' > 0$ kann nur endlich viele einander äquivalente Punkte enthalten.

Beweis. Es seien τ_1, τ'_1 und τ_2, τ'_2 irgend zwei Punkte von \mathfrak{B} , die durch eine Modulsstitution

7) Werden verschiedene Werte τ, τ' durch Indizes ($n = 0, 1, 2, \dots$) unterschieden, so sollen stets r_n, r'_n und $s_n, -s'_n$ die Real- bzw. Imaginärteile von τ_n, τ'_n bedeuten.

$$\tau_1, \tau'_1 = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \tau_2, \tau'_2$$

verknüpft sind. Da für jeden Punkt τ, τ' von \mathfrak{B} $|r|, |r'|$ nach oben, s, s' nach oben und unten positive Schranken besitzen, so liefern die Gleichungen

$$s_1 = \Im \tau_1 = \frac{s_2}{(\gamma r_2 + \delta)^2 + (\gamma s_2)^2},$$

$$\frac{s_1}{|\tau_1|^2} = \Im \left(-\frac{1}{\tau_1} \right) = \frac{s_2}{(\alpha r_2 + \beta)^2 + (\alpha s_2)^2}$$

und die entsprechenden in den gestrichenen Größen Schranken für $|\gamma|, |\gamma'|, |\alpha|, |\alpha'|$ und damit auch für $|\delta|, |\delta'|, |\beta|, |\beta'|$. Da es nun nur endlich viele Möglichkeiten für eine mit ihrer Konjugierten beschränkte ganze Körperzahl gibt, so sind nur endlich viele Werte $\alpha, \beta, \gamma, \delta$ möglich. Es kommen also für die Äquivalenz irgend zweier Punkte in \mathfrak{B} nur beschränkt viele Substitutionen $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ in Betracht. Daher können in \mathfrak{B} zu einem beliebigen Punkt immer nur endlich viele äquivalente existieren.

Im Bereich $s > 0, s' > 0$ seien die folgenden Punktmengen definiert. \mathfrak{U} sei der Bereich

$$\varepsilon'^2 \leq \frac{s'}{s} < \varepsilon^2,$$

der offenbar einen Fundamentalbereich der aus $U = \begin{pmatrix} \varepsilon & 0 \\ 0 & \varepsilon^{-1} \end{pmatrix}$ erzeugten Untergruppe U^n ($n = 0, \pm 1, \pm 2, \dots$) von \mathfrak{G} darstellt.

Mit \mathfrak{Z} sei der Bereich

$$|\tau \tau'| \geq 1$$

bezeichnet, der nach Abrechnung gewisser Randpunkte ein Fundamentalbereich für die Substitution T ist.

Unter $\mathfrak{E}_{s,s'}$ soll die zu dem festen Wertepaar s, s' gehörige Ebene der Variablen r, r' verstanden werden. $\mathfrak{E}_{s,s'}$ sei dann die Punktmenge, die entsteht, wenn man innerhalb $\mathfrak{E}_{s,s'}$ aus jedem System in bezug auf die Substitutionsgruppe

$$r_1 = r + \nu, \quad r'_1 = r' + \nu' \quad (\nu = n_1 + n_2 \varepsilon; n_1, n_2 = 0, \pm 1, \pm 2, \dots)$$

äquivalenter Punkte einen solchen herausgreift, für den

$$(r^2 + s^2)(r'^2 + s'^2) = |\tau \tau'|^2$$

einen möglichst kleinen Wert hat. Die so definierten $\mathfrak{E}_{s,s'}$ bestimmen in ihrer Gesamtheit eine Punktmenge \mathfrak{S} im vierdimensionalen Raum $s > 0, s' > 0$, die auf Grund ihrer Definition ein Fundamentalbereich für die aus S und S_s erzeugbare Gruppe \mathfrak{S}_s von Modulsstitutionen ist.

Hilfssatz 2. Die Punktmenge $(\mathfrak{U}, \mathfrak{S})$ ^{s)} ist ein Fundamentalbereich der Gruppe aller Modulsstitutionen $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ mit $\gamma = 0$.

Beweis. Jeder Punkt τ, τ' ist in bezug auf die Gruppe U^n genau einem Punkt von \mathfrak{U} äquivalent, der vermittelt einer Substitution S_ν , die ja die Imaginärteile ungeändert läßt, auf genau eine Art nach \mathfrak{S} , also nach $(\mathfrak{U}, \mathfrak{S})$, transformierbar ist. Da sich $\begin{pmatrix} \alpha & \beta \\ 0 & \delta \end{pmatrix}$ stets in der Form $S_\nu U^n$ darstellen läßt, so ist Hilfssatz 2 bewiesen.

Die Punktmenge $(\mathfrak{U}, \mathfrak{T}, \mathfrak{S})$ sei mit \mathfrak{M} bezeichnet. Es besteht

Hilfssatz 3. Zu jedem Punkt des Bereiches $s > 0, s' > 0$ existiert in \mathfrak{M} mindestens ein äquivalenter.

Beweis. Nach Hilfssatz 2 ist jeder Punkt von $s > 0, s' > 0$ einem τ_1, τ'_1 von $(\mathfrak{U}, \mathfrak{S})$ äquivalent. Ist $|\tau_1 \tau'_1| \geq 1$, so liegt τ_1, τ'_1 bereits in \mathfrak{M} . Es werde $|\tau_1 \tau'_1| < 1$ angenommen; dann geht τ_1, τ'_1 vermöge T in einen Punkt τ_{11}, τ'_{11} von \mathfrak{T} über mit

$$s_{11} s'_{11} > s_1 s'_1,$$

denn $s_{11} s'_{11} = s_1 s'_1 |\tau_1 \tau'_1|^{-2}$. Durch eine geeignete Substitution $\begin{pmatrix} \alpha & \beta \\ 0 & \delta \end{pmatrix}$ kann man τ_{11}, τ'_{11} wieder in einen Punkt τ_2, τ'_2 von $(\mathfrak{U}, \mathfrak{S})$ transformieren, wobei $s_2 s'_2 = s_{11} s'_{11}$, also

$$s_2 s'_2 > s_1 s'_1.$$

Ist immer noch $|\tau_2 \tau'_2| < 1$, so verfähre man jetzt mit τ_2, τ'_2 wie vorher mit τ_1, τ'_1 , d. h. man transformiere den Punkt zuerst nach \mathfrak{T} und dann nach $(\mathfrak{U}, \mathfrak{S})$ in einen Punkt τ_3, τ'_3 , wobei dann

$$s_3 s'_3 > s_2 s'_2,$$

usw. Nach endlich vielen Schritten muß man zu einem Punkt von \mathfrak{M} gelangen; denn andernfalls erhielte man unendlich viele äquivalente Punkte $\tau_1, \tau'_1, \tau_2, \tau'_2, \tau_3, \tau'_3, \dots$ mit

$$s_1 s'_1 < s_2 s'_2 < s_3 s'_3 < \dots,$$

die im Widerspruch zu Hilfssatz 1 alle in dem endlichen abgeschlossenen Bereich

$$|\tau \tau'| \leq 1, \quad \varepsilon'^2 \leq \frac{s'}{s} \leq \varepsilon^2, \quad s s' \geq s_1 s'_1$$

liegen müßten.

^{s)} Sind $\mathfrak{B}_1, \mathfrak{B}_2, \dots$ irgendwelche Punktmenge, so soll unter $(\mathfrak{B}_1, \mathfrak{B}_2, \dots)$ der Durchschnitt dieser Punktmenge verstanden werden, d. h. diejenige Punktmenge, die aus den gleichzeitig $\mathfrak{B}_1, \mathfrak{B}_2, \dots$ angehörenden Punkten besteht.

Hilfssatz 4. In $(\mathfrak{S}, \mathfrak{T})$ ist $ss' > 0,54$.

Beweis. Es sei

$$f_{s,s'}(r, r') = f(r, r') = (r^2 + s^2)(r'^2 + s'^2).$$

In der Ebene $\mathfrak{S}_{s,s'}$ soll ein Fundamentalbereich der Gruppe

$$r_1 = r + \nu, \quad r'_1 = r' + \nu' \quad (\nu = n_1 + n_2 \varepsilon; n_1, n_2 = 0, \pm 1, \pm 2, \dots)$$

ein $\mathfrak{F}_{s,s'}$ -Bereich heißen. Dann ist auf Grund der Definition von $\mathfrak{S}_{s,s'}$ eine obere Schranke für $f(r, r')$ in irgendeinem $\mathfrak{F}_{s,s'}$ -Bereich stets eine solche in $\mathfrak{S}_{s,s'}$. Das Parallelogramm

$$(6) \quad -\frac{\sqrt{5}}{2} \leq r - r' < +\frac{\sqrt{5}}{2},$$

$$(7) \quad -1 \leq (1+a)r + (1-a)r' < +1$$

auf $\mathfrak{S}_{s,s'}$, wo a eine reelle Zahl aus dem Intervall $-\frac{1}{\sqrt{5}} \leq a \leq +\frac{1}{\sqrt{5}}$ ist,

sei mit \mathfrak{F}_a bezeichnet. \mathfrak{F}_a ist ein $\mathfrak{F}_{s,s'}$ -Bereich; denn zu jedem Punkt r, r' existiert genau ein r_1, r'_1 im Streifen (6), so daß $r_1 = r + n_2 \varepsilon, r'_1 = r' + n_2 \varepsilon'$ mit ganzem rationalem n_2 , und zu r_1, r'_1 gehört in (6) genau ein gleichzeitig in (7) gelegener Punkt r_2, r'_2 mit $r_2 = r_1 + n_1, r'_2 = r'_1 + n_1$, wo n_1 ganz rational ist. Das Parallelogramm \mathfrak{F}_a hat die Ecken

$$(r_1, r'_1), \quad (r_2, r'_2), \quad (-r_1, -r'_1), \quad (-r_2, -r'_2),$$

wo $r_1 = \frac{1}{4}\{2 + (1-a)\sqrt{5}\}, r'_1 = \frac{1}{4}\{2 - (1+a)\sqrt{5}\}, r_2 = \frac{1}{4}\{2 - (1-a)\sqrt{5}\}, r'_2 = \frac{1}{4}\{2 + (1+a)\sqrt{5}\}$.

Es soll nun eine obere Schranke für

$$f(r, r') = r^2 r'^2 + r^2 s'^2 + r'^2 s^2 + s^2 s'^2$$

in \mathfrak{F}_a angegeben werden. Der Ausdruck $r^2 r'^2$ erreicht in \mathfrak{F}_a seinen größten Wert im Punkte $r = -\frac{\sqrt{5}}{4}, r' = \frac{\sqrt{5}}{4}$ der Geraden $r - r' + \frac{\sqrt{5}}{2} = 0$ oder im Punkte $r = -\frac{1}{2(1+a)}, r' = -\frac{1}{2(1-a)}$ der Geraden $(1+a)r + (1-a)r' + 1 = 0$. Im ersteren Fall ist $(r r')^2 = \left(\frac{5}{16}\right)^2$, im zweiten Fall $(r r')^2 = \left(\frac{1}{4(1-a^2)}\right)^2 \leq \left(\frac{5}{16}\right)^2$, da $|a| \leq \frac{1}{\sqrt{5}}$. Das Maximum von $r^2 s'^2 + r'^2 s^2$ liegt in einem Eckpunkt des Parallelogramms. Setzt man zur Abkürzung $ss' = b, \frac{s'}{s} = c$, so ist demnach in \mathfrak{F}_a

$$\text{entweder} \quad f(r, r') \leq \left(\frac{5}{16}\right)^2 + \left(c r_1^2 + \frac{1}{c} r_1'^2\right) b + b^2$$

$$\text{oder} \quad f(r, r') \leq \left(\frac{5}{16}\right)^2 + \left(c r_2^2 + \frac{1}{c} r_2'^2\right) b + b^2.$$

Man wähle nun c so, daß $c r_1^2 + \frac{1}{c} r_1'^2 = c r_2^2 + \frac{1}{c} r_2'^2$, d. h.

$$c^2 = \frac{r_2'^2 - r_1'^2}{r_1^2 - r_2^2} = \frac{1+a}{1-a}.$$

Dann wird

$$c r_1^2 + \frac{1}{c} r_1'^2 = c r_2^2 + \frac{1}{c} r_2'^2 = \frac{9}{8} \cdot \frac{1 - \frac{5}{9} a^2}{\sqrt{1-a^2}} \leq \frac{9}{8},$$

da $a^2 \leq \frac{1}{5}$. Somit gilt in \mathfrak{F}_a

$$(8) \quad f(r, r') \leq \left(\frac{5}{16}\right)^2 + \frac{9}{8} b + b^2,$$

$$\text{falls } \frac{s'}{s} = \sqrt{\frac{1+a}{1-a}}, \quad -\frac{1}{\sqrt{5}} \leq a \leq \frac{1}{\sqrt{5}}.$$

In der Ebene $\mathfrak{E}_{s, s'}$ sei $\mathfrak{F}_a^{(n)}$ das aus \mathfrak{F}_a vermöge der Transformation

$$r_n = \varepsilon^n r, \quad r'_n = \varepsilon'^n r' \quad (n \text{ ganz rational})$$

hervorgehende Parallelogramm, das, wie leicht einzusehen ist, wieder einen $\mathfrak{F}_{s, s'}$ -Bereich darstellt. Wegen

$$f_{s, s'}(r_n, r'_n) = r^2 r'^2 + r^2 \varepsilon^{2n} s'^2 + r'^2 \varepsilon'^{2n} s^2 + s^2 s'^2 = f_{\varepsilon^{-n} s, \varepsilon'^n s'}(r, r')$$

gilt dann Ungleichung (8) in $\mathfrak{F}_a^{(n)}$ und zufolge der am Anfang des Be-

weises gemachten Bemerkung auch in $\mathfrak{E}_{s, s'}$, falls $\frac{\varepsilon^n s'}{\varepsilon^{-n} s} = \sqrt{\frac{1+a}{1-a}}$, also

$\frac{s'}{s} = \varepsilon^{-2n} \sqrt{\frac{1+a}{1-a}}$, $-\frac{1}{\sqrt{5}} \leq a \leq \frac{1}{\sqrt{5}}$, $n = 0, \pm 1, \pm 2, \dots$. Nun ist

$\sqrt{\frac{1+a}{1-a}} = \varepsilon$ bzw. ε^{-1} für $a = \frac{1}{\sqrt{5}}$ bzw. $-\frac{1}{\sqrt{5}}$. Also gilt (8) in $\mathfrak{E}_{s, s'}$ für

alle $\frac{s'}{s} > 0$, d. h. in \mathfrak{E} .

Liegt τ, τ' in $(\mathfrak{E}, \mathfrak{F})$, so muß wegen $f(r, r') = |\tau \tau'|^2 \geq 1$ die Ungleichung

$$\left(\frac{5}{16}\right)^2 + \frac{9}{8} b + b^2 \geq 1$$

gelten. Daraus folgt

$$b \geq \frac{-9 + \sqrt{312}}{16} > 0,54,$$

womit Hilfssatz 4 bewiesen ist.

Aus $s s' > 0,54$ und $\varepsilon'^2 \leq \frac{s'}{s} < \varepsilon^2$ folgt

$$(9) \quad s > 0,45, \quad s' > 0,45 \text{ in } \mathfrak{M}.$$

Ferner ist $s + s' \geq 2\sqrt{s s'} > 2\sqrt{0,54}$ und daher

$$(10) \quad e^{-\frac{\pi}{\sqrt{5}}(s+s')} < \frac{1}{7} \text{ in } \mathfrak{M}.$$

Aus den positiven unteren Schranken für s und s' ergeben sich übrigens leicht auf Grund der Definition von \mathfrak{S} obere Schranken für $|r|$ und $|r'|$ in \mathfrak{M} . In Verbindung mit $\varepsilon'^2 \leq \frac{s'}{s} < \varepsilon^2$ folgt hieraus, daß τ oder τ' in \mathfrak{M} nur dann unendlich werden können, wenn dies für s und s' gleichzeitig der Fall ist. Es hat also \mathfrak{M} nur einen unendlich fernen Punkt.

Satz 2⁹⁾. Die Punktmenge \mathfrak{M} stellt bei Abrechnung gewisser Randpunkte einen Fundamentalbereich der Modulgruppe in $k(\sqrt{5})$ dar.

Beweis. Nach Hilfssatz 3 existiert in \mathfrak{M} zu jedem Punkt des Raumes $s > 0, s' > 0$ ein äquivalenter. Es ist noch zu untersuchen, wann zwei Punkte τ_1, τ'_1 und τ_2, τ'_2 von \mathfrak{M} einander äquivalent sind. Dabei kann man sich wegen Hilfssatz 2 auf den Fall $\gamma \neq 0$, also $|N(\gamma)| \geq 1$, beschränken, wenn

$$(11) \quad \tau_1, \tau'_1 = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \tau_2, \tau'_2.$$

Nun folgt aus (11) nach (4) und (5)

$$(12) \quad s_1 s'_1 = s_2 s'_2 |\gamma \tau_2 + \delta|^{-2} |\gamma' \tau'_2 + \delta'|^{-2} = s_2 s'_2 N |\gamma \tau_2 + \delta|^{-2}. \quad 10)$$

Da

$$(13) \quad N |\gamma \tau_2 + \delta|^2 = N \{(\gamma r_2 + \delta)^2 + \gamma^2 s_2^2\} \geq N(\gamma^2 s_2^2) = N(\gamma)^2 (s_2 s'_2)^2,$$

so folgt

$$N(\gamma)^2 \leq \frac{1}{s_1 s'_1 s_2 s'_2}.$$

Da nun τ_1, τ'_1 und τ_2, τ'_2 in \mathfrak{M} liegen, so ist nach Hilfssatz 4 $s_1 s'_1 > \frac{1}{2}$ und $s_2 s'_2 > \frac{1}{2}$, also

$$N(\gamma)^2 < 4.$$

Somit kommt nur $N(\gamma) = \pm 1$ in Betracht. In diesem Fall hat man wegen (12) und der linken Seite von (13)

$$\frac{s_2 s'_2}{s_1 s'_1} = N \{(\gamma r_2 + \delta)^2 + \gamma^2 s_2^2\} = N \{(r_2 + \nu)^2 + s_2^2\},$$

wo $\nu = \frac{\delta}{\gamma}$ eine ganze Körperzahl ist. Daher ist auf Grund der Definition von \mathfrak{S}

$$N \{(r_2 + \nu)^2 + s_2^2\} \geq N(r_2^2 + s_2^2) = |\tau_2 \tau'_2|^2$$

⁹⁾ Dieser Satz, der der Vollständigkeit halber hier abgeleitet werden soll, findet im folgenden, wo man stets mit Hilfssatz 3 auskommt, keine Verwendung.

¹⁰⁾ Sind $\alpha, \alpha'; \beta, \beta'; \dots$ Paare konjugierter Körperzahlen, u, u', v, v', \dots irgendwelche Variable, so soll unter der „Norm“ Nf , wenn $f = f(\alpha, \beta, \dots, u, v, \dots)$ ein Ausdruck in den $\alpha, \beta, \dots, u, v, \dots$ ist, stets das Produkt $f(\alpha, \beta, \dots, u, v, \dots) \cdot f(\alpha', \beta', \dots, u', v', \dots)$ verstanden werden. Ebenso soll Sf die „Spur“ $f(\alpha, \beta, \dots, u, v, \dots) + f(\alpha', \beta', \dots, u', v', \dots)$ bedeuten.

und somit

$$\frac{s_2 s_2'}{s_1 s_1'} \geq |\tau_2 \tau_2'|^2.$$

Ebenso folgt aber, wenn man von der zu (11) inversen Substitution

$$\tau_2, \tau_2' = \begin{pmatrix} -\delta & \beta \\ \gamma & -\alpha \end{pmatrix} \tau_1, \tau_1' \text{ ausgeht,}$$

$$\frac{s_1 s_1'}{s_2 s_2'} \geq |\tau_1 \tau_1'|^2.$$

Also ist

$$|\tau_1 \tau_1'| \cdot |\tau_2 \tau_2'| \leq 1.$$

Da nun in \mathfrak{M} $|\tau \tau'| \geq 1$, so folgt

$$|\tau_1 \tau_1'| = |\tau_2 \tau_2'| = 1.$$

Zwei verschiedene Punkte von \mathfrak{M} können also nur dann äquivalent sein, wenn sie beide auf dem Randgebiet $|\tau \tau'| = 1$ liegen.

§ 3.

Ein Satz über Modulfunktionen des Körpers $k(\sqrt{5})$.

Ist \mathfrak{B} eine Punktmenge innerhalb $s > 0$, $s' > 0$, so soll eine Funktion von τ und τ' *regulär* in \mathfrak{B} heißen, wenn sie in jedem Punkt von \mathfrak{B} in beiden Variablen regulär analytisch ist.

Satz 3. Eine in \mathfrak{M} reguläre Funktion $\Phi(\tau, \tau')$, die den Transformationsformeln

$$\Phi(\tau + 1, \tau' + 1) = \Phi(\tau, \tau'),$$

$$\Phi(\varepsilon^2 \tau, \varepsilon'^2 \tau') = \Phi(\tau, \tau'),$$

$$\Phi\left(-\frac{1}{\tau}, -\frac{1}{\tau'}\right) = \Phi(\tau, \tau')$$

genügt, ist eine Konstante¹¹⁾.

Beweis. Aus Satz 1 und Hilfssatz 3 folgt, daß $\Phi(\tau, \tau')$ ungeändert bleibt bei Ausübung einer beliebigen Modulusubstitution und im ganzen Bereich $s > 0$, $s' > 0$ regulär ist. Führt man durch die Gleichungen

$$\tau = -\varepsilon' u + v, \quad \tau' = -\varepsilon u + v$$

oder

$$u = \frac{\tau - \tau'}{\sqrt{5}} = S \frac{\tau}{\sqrt{5}}, \quad v = \frac{\varepsilon \tau - \varepsilon' \tau'}{\sqrt{5}} = S \frac{\varepsilon \tau}{\sqrt{5}} \quad (12)$$

¹¹⁾ Dieser Satz weicht offenbar insofern von dem analogen über elliptische Modulfunktionen ab, als bei ihm keine einschränkende Voraussetzung über das Verhalten der Funktion im Unendlichen auftritt.

¹²⁾ Vgl. ¹⁰⁾.

neue Variable u, v ein, so geht $\Phi(\tau, \tau')$ in eine für $\Im(-\varepsilon' u + v) > 0$, $\Im(-\varepsilon u + v) < 0$ reguläre Funktion $\Psi(u, v)$ über, die den Transformationsformeln

$$\Psi(u + 1, v) = \Psi(u, v), \quad \Psi(u, v + 1) = \Psi(u, v)$$

genügt, denn diese Transformationen in u und v bedeuten Modulusubstitutionen für τ und τ' . Es besteht folglich die Fouriersche Entwicklung

$$\Psi(u, v) = \sum_{m, n = -\infty}^{+\infty} a_{mn} e^{2\pi i(mu + nv)}$$

mit

$$a_{mn} = \int_{-\frac{1}{2} + iu_0}^{+\frac{1}{2} + iu_0} \int_{-\frac{1}{2} + iv_0}^{+\frac{1}{2} + iv_0} \Psi(u, v) e^{-2\pi i(mu + nv)} du dv$$

$$\left(u_0 = \frac{s + s'}{\sqrt{5}}, v_0 = \frac{\varepsilon s + \varepsilon' s'}{\sqrt{5}}, s > 0, s' > 0\right).$$

Führt man τ und τ' wieder ein, so wird

$$(14) \quad \Phi(\tau, \tau') = \sum_{\nu} a_{\nu} e^{\frac{2\pi i S \nu \tau}{\sqrt{5}}},$$

$$(15) \quad a_{\nu} = \frac{1}{\sqrt{5}} \iint_{\mathfrak{F}_{s, s'}} \Phi(\tau, \tau') e^{-\frac{2\pi i S \nu \tau}{\sqrt{5}}} d\tau d\tau',$$

wobei $\nu = m + n\varepsilon$ alle ganzen Zahlen aus $k(\sqrt{5})$ durchläuft und $\mathfrak{F}_{s, s'}$ der Bereich $-\frac{1}{2} \leq \frac{r-r'}{\sqrt{5}} \leq \frac{1}{2}$, $-\frac{1}{2} \leq \frac{\varepsilon r - \varepsilon' r'}{\sqrt{5}} \leq \frac{1}{2}$ bei festem s, s' ist, was (bei geeigneter Abrechnung von Randpunkten) ein $\mathfrak{F}_{s, s'}$ -Bereich im Sinne von § 2 (nämlich der Bereich \mathfrak{F}_a für $a = \frac{1}{\sqrt{5}}$) ist. Aus (15) folgt für jedes $s > 0, s' > 0$

$$|a_{\nu}| \leq e^{\frac{2\pi}{\sqrt{5}}(\nu s + \nu' s')} M_{s, s'},$$

wo $M_{s, s'}$ das Maximum von $|\Phi(\tau, \tau')|$ in $\mathfrak{F}_{s, s'}$ und damit in der ganzen Ebene $\mathfrak{C}_{s, s'}$ ist. Setzt man

$$s = \varepsilon^{2n} s_0, \quad s' = \varepsilon'^{2n} s_0 \quad (n = 0, \pm 1, \pm 2, \dots; s_0 > 0)$$

und berücksichtigt, daß $M_{s, s'} = M_{s_0, s_0}$ wegen $\Phi(\varepsilon^{2n} \tau, \varepsilon'^{2n} \tau') = \Phi(\tau, \tau')$, so folgt

$$|a_{\nu}| \leq e^{\frac{2\pi}{\sqrt{5}}(s_0(\nu \varepsilon^{2n} + \nu' \varepsilon'^{2n}))} M_{s_0, s_0}.$$

Für alle ν mit $\nu < 0$ oder $\nu' < 0$ folgt hieraus, wenn man $n \rightarrow +\infty$ bzw. $-\infty$ streben läßt, $a_{\nu} = 0$. In der Reihe (14) treten somit außer

a_0 nur Glieder mit total positivem ν ($\nu > 0$) auf:

$$(16) \quad \Phi(\tau, \tau') = a_0 + \sum_{\nu > 0} a_\nu e^{\frac{2\pi i S \nu \tau}{\sqrt{5}}}.$$

Für $s \rightarrow \infty$, $s' \rightarrow \infty$ strebt folglich $\Phi(\tau, \tau') \rightarrow a_0$. Es ist demnach $\Phi(\tau, \tau')$ auch im unendlich fernen Punkt von \mathfrak{M} stetig, und $|\Phi(\tau, \tau')|$ muß in der zu \mathfrak{M} gehörigen abgeschlossenen Punktmenge \mathfrak{M}' (einschließlich des unendlich fernen Punktes) einen endlichen Maximalwert M annehmen.

Es sei nun $\Phi(\tau, \tau')$ in wenigstens einer Variablen, etwa in τ , nicht konstant, und $|\Phi(\tau, \tau')|$ möge das Maximum M in einem nicht unendlich fernen Punkt τ_1, τ'_1 von \mathfrak{M}' annehmen. Dann umgebe man den Punkt τ_1, τ'_1 in der durch ihn gehenden Ebene der Variablen τ mit einem noch ganz im Regularitätsgebiet gelegenen Kreis. Auf dessen Rand muß es nach einem funktionentheoretischen Satz einen Punkt τ_2, τ'_2 geben, so daß $|\Phi(\tau_2, \tau'_2)| > M$, was unmöglich ist, weil τ_2, τ'_2 nach Hilfssatz 3 einem Punkt von \mathfrak{M}' äquivalent ist. So folgt, daß $|\Phi(\tau, \tau')|$ das Maximum M im unendlich fernen Punkt von \mathfrak{M}' annimmt. Setzt man in (16)

$e^{\frac{2\pi i \tau}{\sqrt{5}}} = q$, $e^{-\frac{2\pi i \tau'}{\sqrt{5}}} = z \cdot q$ ($|q| < 1$, $|zq| < 1$), so muß also der absolute Betrag der im Kreise $|q| \leq c$ ($c < 1$, $c < \frac{1}{|z|}$) regulär analytischen Funktion von q

$$\Phi(\tau, \tau') = a_0 + \sum_{\nu > 0} a_\nu z^{\nu'} q^{\nu + \nu'}$$

das Maximum im Punkte $q = 0$ haben, was auf Grund des bereits oben angewandten Satzes aus der Funktionentheorie nur möglich ist, wenn die Funktion eine Konstante in bezug auf q ist. Setzt man $q = 0$, so folgt

$$\Phi(\tau, \tau') = a_0$$

identisch in τ und τ' , womit Satz 3 bewiesen ist.

Hilfssatz 5. Sind $\Phi_1(\tau, \tau')$, $\Phi_2(\tau, \tau')$, ..., $\Phi_n(\tau, \tau')$ reguläre Funktionen in \mathfrak{M} , die bei Ausübung einer beliebigen Modulsstitution in $k(\sqrt{5})$ lediglich eine Permutation erleiden, so sind sie sämtlich Konstanten.

Der Beweis dieses Satzes ergibt sich ohne weiteres durch Anwendung von Satz 3 auf die n elementar-symmetrischen Funktionen von $\Phi_1(\tau, \tau')$, $\Phi_2(\tau, \tau')$, ..., $\Phi_n(\tau, \tau')$.

Das Ergebnis dieses Paragraphen kann offenbar, wie es auch in dieser Arbeit geschehen soll, verwandt werden, um Identitäten zwischen Funktionen zweier Veränderlicher nachzuweisen, deren Verhalten bei den drei Modulsstitutionen S, T, U bekannt ist.

§ 4.

Thetafunktionen in $k(\sqrt{5})$.

Die Kongruenz

$$(17) \quad \nu x \equiv \mu \pmod{2}$$

hat für alle ganzen Zahlen μ, ν aus $k(\sqrt{5})$ mit $\nu \not\equiv 0 \pmod{2}$ und festem Quotienten $\frac{\mu}{\nu} = \kappa$ ein und dieselbe eindeutig bestimmte Lösung x unter den Zahlen $0, 1, \varepsilon, \varepsilon'$, da dieselben ein vollständiges Restsystem mod 2 bilden. Man nennt dann $\kappa \equiv x \pmod{2}$. Trifft man noch die Vereinbarung, daß die Kongruenz (17) im Falle $\nu \equiv 0 \pmod{2}$, $\mu \not\equiv 0 \pmod{2}$ die Lösung $x = \infty$ hat und nennt dann $\kappa \equiv \infty \pmod{2}$, so ist somit jede Zahl aus $k(\sqrt{5})$ genau einer der fünf Zahlen

$$(18) \quad 0, 1, \varepsilon, \varepsilon', \infty$$

nach dem Modul 2 kongruent. Es seien nun Funktionen $\vartheta_{\kappa\lambda}(\tau, \tau') = \vartheta_{\kappa\lambda}(\tau)$ mit irgendwelchen Zahlen κ, λ aus $k(\sqrt{5})$, die zwei verschiedenen der fünf Zahlen (18) mod 2 kongruent sind ($\kappa + \lambda \not\equiv 0 \pmod{2}$), als Indizes definiert durch die Gleichung

$$(19) \quad \vartheta_{\kappa\lambda}(\tau) = \vartheta_{\lambda\kappa}(\tau) = \sum_{\nu} (-1)^{\frac{\nu(\nu+1)}{2}} \frac{e^{\pi i \nu \tau}}{\sqrt{5}^{\nu}} e^{\pi i \nu \frac{\tau^2}{5}},$$

wo in der Summe ν alle ganzen Zahlen aus $k(\sqrt{5})$ durchläuft, während α und β zwei (eventuell zwei gleiche) der Zahlen $0, 1, \varepsilon, \varepsilon'$ sind, die sich (wegen $\kappa + \lambda \not\equiv 0 \pmod{2}$) aus den Kongruenzen

$$(20) \quad \alpha \equiv \frac{1}{\kappa' + \lambda'} \pmod{2}, \quad \beta \equiv \frac{\kappa' \lambda'}{\kappa' + \lambda'} \pmod{2}$$

eindeutig ergeben. $\vartheta_{\kappa\lambda}(\tau)$ stellt also für alle mod 2 kongruenten Wertepaare κ, λ ein und dieselbe Funktion dar. Formel (19) definiert demnach zehn Funktionen, entsprechend den $\binom{5}{2} = 10$ Kombinationen von zwei verschiedenen der fünf Elemente $0, 1, \varepsilon, \varepsilon', \infty$. Läßt man auch die Zahl ∞ selbst für den Index κ oder λ zu, wobei die Rechenregeln $\frac{1}{\infty} \equiv 0$, $\frac{1}{0} \equiv \infty$, $\infty + \alpha \equiv \infty$, $\alpha \cdot \infty \equiv \infty$ ($\alpha \not\equiv 0 \pmod{2}$), $\frac{\alpha \cdot \infty + \beta}{\gamma \cdot \infty + \delta} \equiv \alpha \cdot \frac{1}{\gamma}$ (α und γ nicht beide $\equiv 0 \pmod{2}$), $\infty' \equiv \infty$, $\infty \equiv \infty \pmod{2}$ gelten sollen, wenn $\alpha, \beta, \gamma, \delta$ Zahlen aus $k(\sqrt{5})$ sind, so besteht

Hilfssatz 6. Für jede Modulsstitution $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ ist

$$(21) \quad \vartheta_{\kappa\lambda}^8\left(\frac{\alpha\tau + \beta}{\gamma\tau + \delta}\right) = N(\gamma\tau + \delta)^4 \vartheta_{\frac{-\delta\kappa + \beta}{\gamma\kappa - \alpha}, \frac{-\delta\lambda + \beta}{\gamma\lambda - \alpha}}^8(\tau). \quad (13)$$

Beweis. Gilt Formel (21) für zwei Modulsstitutionen G_1 und G_2 , so gilt sie auch für das Produkt $G_2 G_1$. Daher genügt es auf Grund von Satz 1 die Formel für die drei speziellen Substitutionen S, U, T zu beweisen. Ersetzt man in (20) κ, λ durch $\kappa - 1, \lambda - 1$ bzw. $\varepsilon'^2 \kappa, \varepsilon'^2 \lambda$, so gehen α, β in α_1, β_1 bzw. α_2, β_2 über, wobei $\alpha_1 \equiv \alpha, \beta_1 \equiv \alpha + \beta - 1, \alpha_2 \equiv \varepsilon'^2 \alpha, \beta_2 \equiv \varepsilon'^2 \beta \pmod{2}$. Berücksichtigt man nun, daß $\vartheta_{\kappa\lambda}^8(\tau)$ ungeändert bleibt, wenn man in der Definitionsformel (19) α oder β durch eine mod 2 kongruente ganze Körperzahl ersetzt, so folgen aus der Darstellung

$$\vartheta_{\kappa\lambda}^8(\tau) = e^{\frac{\pi i}{4} S \frac{\alpha^2 \tau}{\sqrt{5}}} \sum_{\nu} e^{\frac{\pi i S \tau^2 \nu^2 + (\alpha\tau + \beta)\nu}{\sqrt{5}}}$$

die Formeln

$$\vartheta_{\kappa\lambda}^8(\tau + 1) = \vartheta_{\kappa-1, \lambda-1}^8(\tau), \quad \vartheta_{\kappa\lambda}^8(\varepsilon^2 \tau) = \vartheta_{\varepsilon'^2 \kappa, \varepsilon'^2 \lambda}^8(\tau).$$

Die Transformationsformel für die Substitution T möge abgeleitet werden aus der für den Körper $k(\sqrt{5})$ spezialisierten Heckschen Thetaformel¹⁴⁾

$$\sum_{\nu} e^{-\pi\{t_1(\nu+z_1)^2 + t_2(\nu'+z_2)^2\}} = \frac{1}{\sqrt{5}\sqrt{t_1 t_2}} \sum_{\nu} e^{-\frac{\pi}{5}\left(\frac{\nu^2 + \nu'^2}{t_1 + t_2}\right) - \frac{2\pi i}{\sqrt{5}}(\nu z_1 - \nu' z_2)},$$

wobei $\Re t_1 > 0, \Re t_2 > 0, \Re \sqrt{t_1 t_2} > 0$. Setzt man

$$t_1 = -\frac{i\tau}{\sqrt{5}}, \quad t_2 = \frac{i\tau'}{\sqrt{5}}, \quad 2z_1 = \frac{\alpha}{\tau} + \beta, \quad 2z_2 = \frac{\alpha'}{\tau'} + \beta',$$

so folgt nach leichter Umformung

$$e^{\frac{\pi i}{2} S \frac{\alpha\beta}{\sqrt{5}}} e^{\frac{\pi i}{4} S \frac{\beta^2 \tau}{\sqrt{5}}} \sum_{\nu} e^{\frac{\pi i S \tau^2 \nu^2 + (\alpha + \beta\tau)\nu}{\sqrt{5}}} = \frac{1}{\sqrt{\tau\tau'}} e^{-\frac{\pi i}{4} S \frac{\alpha^2}{\tau\sqrt{5}}} \sum_{\nu} e^{-\pi i S \frac{\nu^2 + \left(\frac{\alpha}{\tau} + \beta\right)\nu}{\sqrt{5}}}.$$

Da nun α und β sich vertauschen, wenn man κ, λ durch $-\frac{1}{\kappa}, -\frac{1}{\lambda}$ ersetzt, so besagt diese Formel

$$\vartheta_{\kappa\lambda}^8\left(-\frac{1}{\tau}\right) = (\tau\tau')^4 \vartheta_{-\frac{1}{\kappa}, -\frac{1}{\lambda}}^8(\tau).$$

Somit ist Formel (21) für die Substitutionen S, U, T und damit allgemein bewiesen.

¹³⁾ Ist $\Phi(\tau, \tau') = \Phi(\tau)$ eine Funktion von τ, τ' , so soll unter $\Phi\left(\frac{\alpha\tau + \beta}{\gamma\tau + \delta}\right)$ stets $\Phi\left(\frac{\alpha\tau + \beta}{\gamma\tau + \delta}, \frac{\alpha'\tau' + \beta'}{\gamma'\tau' + \delta'}\right)$ für jede Modulsstitution $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ verstanden werden. Vgl. ferner ¹⁰⁾.

¹⁴⁾ Hecke, Vorlesungen über die Theorie der algebraischen Zahlen. Leipzig 1923, S. 230.

§ 5.

Abschätzung der Thetafunktionen im Fundamentalbereich.

Es sei $e^{\frac{\pi i S \mu \tau}{\sqrt{5}}} = h_\mu$ gesetzt für jede Zahl μ aus $k(\sqrt{5})$. Dann ist nach (19)

$$(22) \quad h_{\alpha^2}^{-\frac{1}{4}} \vartheta_{\kappa\lambda}(\tau) = \sum_{\nu} (-1)^{\frac{S\beta\nu}{\sqrt{5}}} h_{\nu(\nu+\alpha)}.$$

Es sei zunächst $\alpha = 0$ oder 1 .¹⁵⁾ Dann ist, wenn man $\nu = n_1 \varepsilon + n_2 \varepsilon'$ (die Zahlen $\varepsilon, \varepsilon'$ bilden eine Körperbasis) setzt,

$$h_{\alpha^2}^{-\frac{1}{4}} \vartheta_{\kappa\lambda}(\tau) = \sum_{n_1, n_2 = -\infty}^{+\infty} (-1)^{\frac{S\beta(n_1\varepsilon+n_2\varepsilon')}{\sqrt{5}}} h_1^{n_1(n_1-\alpha)} h_\varepsilon^{n_1(n_1+\alpha)} h_{\varepsilon'}^{n_2(n_2+\alpha)}.$$

Addiert man hierzu die Formel mit vertauschten Summationsbuchstaben und greift gleichzeitig die Glieder mit $n_1 = n_2 = 0$, $n_1 = n_2 = -\alpha$ heraus, so folgt

$$\begin{aligned} & |h_{\alpha^2}^{-\frac{1}{4}} \vartheta_{\kappa\lambda}(\tau) - (1 + \alpha)| \\ \leq & -(1 + \alpha) + \frac{1}{2} \sum_{n_1, n_2 = -\infty}^{+\infty} |h_1|^{(n_1-n_2)^2} (|h_\varepsilon|^{n_1(n_1+\alpha)} |h_{\varepsilon'}|^{n_2(n_2+\alpha)} + |h_\varepsilon|^{n_2(n_2+\alpha)} |h_{\varepsilon'}|^{n_1(n_1+\alpha)}). \end{aligned}$$

Nun ist in \mathfrak{M} für jedes reelle Zahlenpaar p, q

$$|h_\varepsilon|^p |h_{\varepsilon'}|^q + |h_\varepsilon|^q |h_{\varepsilon'}|^p \leq |h_1|^p + |h_1|^q,$$

denn $h_\varepsilon h_{\varepsilon'} = h_1$, $|h_\varepsilon| \leq 1$, $|h_{\varepsilon'}| \leq 1$ in \mathfrak{M} und daher

$$|h_1|^p + |h_1|^q - |h_\varepsilon|^p |h_{\varepsilon'}|^q - |h_\varepsilon|^q |h_{\varepsilon'}|^p = (|h_\varepsilon|^p - |h_\varepsilon|^q)(|h_{\varepsilon'}|^p - |h_{\varepsilon'}|^q) \geq 0.$$

Folglich ist in \mathfrak{M}

$$\begin{aligned} |h_{\alpha^2}^{-\frac{1}{4}} \vartheta_{\kappa\lambda}(\tau) - (1 + \alpha)| & \leq -(1 + \alpha) + \frac{1}{2} \sum_{n_1, n_2 = -\infty}^{+\infty} |h_1|^{(n_1-n_2)^2} (|h_1|^{n_1(n_1+\alpha)} + |h_1|^{n_2(n_2+\alpha)}) \\ & \leq -(1 + \alpha) + \sum_{m = -\infty}^{+\infty} |h_1|^{m^2} \sum_{n = -\infty}^{+\infty} |h_1|^{n(n+\alpha)} \end{aligned}$$

und demnach

$$(23) \quad |h_{\alpha^2}^{-\frac{1}{4}} \vartheta_{\kappa\lambda}(\tau)| \geq 2(1 + \alpha) - \sum_{m = +\infty}^{+\infty} |h_1|^{m^2} \sum_{n = -\infty}^{+\infty} |h_1|^{n(n+\alpha)} > \frac{1}{3},$$

da $|h_1| = e^{-\frac{\pi}{\sqrt{5}(s+s')}} < \frac{1}{7}$ in \mathfrak{M} (nach (10)).

¹⁵⁾ Im Falle $\alpha = 0$ ($\kappa + \lambda \equiv \infty \pmod{2}$) ist $\beta = 0, 1, \varepsilon$ oder ε' , im Falle $\alpha = 1$ ($\kappa + \lambda \equiv 1 \pmod{2}$) ist $\beta = 0$ oder 1 .

Jetzt sei $\alpha = \varepsilon$ oder ε' .¹⁶⁾ Dann ist $\beta = b\alpha'$, wo $b = 0$ oder 1 .
Setzt man in (22) $\nu = n_2 - n_1\alpha'$, so folgt

$$(24) \quad h_{\alpha^2}^{-\frac{1}{4}} \vartheta_{\alpha\lambda}(\tau) = \sum_{n_1, n_2 = -\infty}^{+\infty} (-1)^{b(n_1+n_2)} h_1^{n_1(n_1+1)} h_{\alpha}^{n_2(n_2+1)} h_{\alpha'}^{(n_1-n_2)^2}$$

$$= 2 \sum_{n_1, n_2 = 0}^{\infty} (-1)^{b(n_1+n_2)} h_1^{n_1(n_1+1)} h_{\alpha}^{n_2(n_2+1)} (h_{\alpha'}^{(n_1-n_2)^2} + (-1)^b h_{\alpha'}^{(n_1+n_2+1)^2}).$$

Dabei ist

$$h_{\alpha'}^{(n_1-n_2)^2} + (-1)^b h_{\alpha'}^{(n_1+n_2+1)^2} = h_{\alpha'}^{(n_1-n_2)^2} (1 + (-1)^b h_{\alpha'}^{(2n_1+1)(2n_2+1)})$$

$$= h_{\alpha'}^{(n_1-n_2)^2} (1 + (-1)^b h_{\alpha'}^{(2n_1+1)(2n_2+1)-1} \sum_{n=0}^{2n_1+1} (-1)^{(b+1)n} h_{\alpha'}^n).$$

Trägt man dies in (24) ein, dividiert durch $2(1 + (-1)^b h_{\alpha'})$, greift die Glieder mit $n_1^2 + n_2^2 \leq 2$ heraus und berücksichtigt $h_{\alpha} h_{\alpha'} = h_1$, $|h_{\alpha}| \leq 1$, $|h_{\alpha'}| \leq 1$ in \mathfrak{M} , so folgt

$$\left| \frac{\vartheta_{\alpha\lambda}(\tau)}{2 h_{\alpha^2}^{\frac{1}{4}} (1 + (-1)^b h_{\alpha'})} - 1 \right| \leq |h_1| (|h_{\alpha}| + 4|h_{1+\alpha'}|) + 2|h_1|^2 + |h_1|^3 + 7|h_1|^4$$

$$+ \sum_{\substack{n_1, n_2 = 0 \\ n_1^2 + n_2^2 > 2}}^{\infty} (2n_1+1)(2n_2+1) |h_1|^{n_1(n_1+1)} |h_{\alpha}|^{n_2(n_2+1)} |h_{\alpha'}|^{(n_1-n_2)^2}.$$

Nun ist $|h_{\alpha}| + 4|h_{1+\alpha'}| \leq 1 + 4|h_1|^2$ in \mathfrak{M} , weil $1 + 4|h_1|^2 - |h_{\alpha}| - 4|h_{1+\alpha'}| = (1 - |h_{\alpha}|)(1 - 4|h_{1+\alpha'}|) \geq 0$ in \mathfrak{M} , und ferner

$$|h_1^{n_1(n_1+1)} h_{\alpha}^{n_2(n_2+1)} h_{\alpha'}^{(n_1-n_2)^2}| \leq e^{-\frac{\pi}{\sqrt{5}} \{(a'^2 s + \alpha^2 s') n_1^2 - 2(\alpha' s + \alpha s') n_1 n_2 + (s + s') n_2^2\}}$$

$$\leq e^{-\frac{\pi}{\sqrt{5}} s_{\alpha}(1 + \varepsilon'^2)(n_1^2 + n_2^2)} \quad (s_{\varepsilon} = s, s_{\varepsilon'} = s')$$

wegen $n_1^2(\alpha'^2 s + \alpha^2 s') - 2(\alpha' s + \alpha s') n_1 n_2 + (s + s') n_2^2 - s_{\alpha}(1 + \varepsilon'^2)(n_1^2 + n_2^2)$
 $= (\sqrt{\varepsilon^2 s_{\alpha}' - s_{\alpha} \cdot n_1} - \sqrt{s_{\alpha}' - \varepsilon'^2 s_{\alpha} \cdot n_2})^2 \geq 0$ in \mathfrak{M} . Da noch $e^{-\frac{\pi}{\sqrt{5}} s_{\alpha}(1 + \varepsilon'^2)}$
 $= e^{-\frac{\pi}{\varepsilon} s_{\alpha}} \leq 0,42$ wegen (9), so folgt

$$\left| \frac{\vartheta_{\alpha\lambda}(\tau)}{2 h_{\alpha^2}^{\frac{1}{4}} (1 + (-1)^b h_{\alpha'})} - 1 \right| \leq |h_1| + 2|h_1|^2 + 5|h_1|^3 + 7|h_1|^4$$

$$+ \left(\sum_{n=0}^{\infty} (2n+1) p^{n^2} \right)^2 - (1 + 6p + 9p^2) \quad (p = 0,42)$$

¹⁶⁾ Im Falle $\alpha = \varepsilon$ ($\kappa + \lambda \equiv \varepsilon \pmod{2}$) ist $\beta = 0$ oder ε' , im Falle $\alpha = \varepsilon'$ ($\kappa + \lambda \equiv \varepsilon' \pmod{2}$) ist $\beta = 0$ oder ε .

in \mathfrak{M} . Die rechte Seite dieser Ungleichung ist aber $< 1 - \frac{1}{100}$, was aus $|h_1| < \frac{1}{7}$ und $\sum_{n=0}^{\infty} (2n+1)p^{n^2} < 1 + 3p + 5p^4(1 - \frac{7}{5}p^5)^{-1}$ folgt; d. h.

$$(25) \quad \left| \frac{\vartheta_{\kappa\lambda}(\tau)}{2h_{\alpha'}^{\frac{1}{2}}(1+(-1)^b h_{\alpha'})} \right| > \frac{1}{100} \text{ in } \mathfrak{M}.$$

Man bestätigt leicht, daß $b \equiv S \frac{\kappa\alpha'}{\sqrt{5}} \pmod{2}$. Aus (23) und (25) folgt somit

Hilfssatz 7. Die 10 Funktionen

$$\vartheta_{\kappa\lambda}^{-8}(\tau) \quad (\kappa + \lambda \equiv 1 \text{ oder } \infty \pmod{2}),$$

$$\left(1 + (-1)^{S \frac{\kappa\alpha'}{\sqrt{5}} h_{\alpha'}}\right)^8 \vartheta_{\kappa\lambda}^{-8}(\tau) \quad (\kappa + \lambda \equiv \alpha \pmod{2}, \alpha = \varepsilon \text{ oder } \varepsilon')$$

sind regulär in \mathfrak{M} .

§ 6.

Einführung neuer Funktionen und deren Transformation.

Es sei κ eine ganze Zahl aus $k(\sqrt{5})$. Dann soll unter $F_{\kappa}(\tau, \tau') = F_{\kappa}(\tau)$ die Summe

$$(26) \quad F_{\kappa}(\tau) = \sum_{\substack{(\nu, \mu) \\ \nu\mu > 0}} (-1)^{S \frac{\nu\mu\kappa}{\sqrt{5}}} |N(\nu)| e^{\pi i S \frac{\nu\mu\tau}{\sqrt{5}}}$$

verstanden werden. Darin durchläuft (ν) alle ganzen Ideale, μ alle ganzen Zahlen aus $k(\sqrt{5})$, für die $\nu\mu$ total positiv ist. Setzt man $e^{\pi i \frac{\tau}{\sqrt{5}}} = q$, $e^{-\pi i \frac{\tau'}{\sqrt{5}}} = q'$, so läßt sich $F_{\kappa}(\tau)$ in Form einer Potenzreihe in $k(\sqrt{5})$ schreiben:

$$F_{\kappa}(\tau) = \sum_{\mu > 0} a_{\mu} q^{\mu} q'^{\mu'}, \quad a_{\mu} = (-1)^{S \frac{\mu\kappa}{\sqrt{5}}} \sum_{(\nu)|\mu} |N(\nu)|.$$

Die Reihe konvergiert gleichmäßig in jedem Bereich $|q| \leq q_0$, $|q'| \leq q_0$ mit $0 < q_0 < 1$; denn setzt man $q = q'$, so wird

$$F_{\kappa}(\tau) = \sum_{\mu > 0} a_{\mu} q^{\mu+\mu'} = \sum_{n=1}^{\infty} b_n q^n, \quad b_n = \sum_{\mu+\mu'=n} a_{\mu},$$

was eine gewöhnliche Potenzreihe mit dem Konvergenzradius 1 ist, da b_n nicht stärker als eine Potenz von n unendlich wird. Es ist daher $F_{\kappa}(\tau)$

eine im Bereich $s > 0$, $s' > 0$ reguläre Funktion von τ , τ' . Wie ohne weiteres aus (26) folgt, stellt $F_{\kappa}(\tau)$ für alle derselben Restklasse mod 2 zugehörenden κ ein und dieselbe Funktion dar mit

$$(27) \quad F_{\kappa}(\tau + 1) = F_{\kappa-1}(\tau), \quad F_{\kappa}(\varepsilon^2 \tau) = F_{\varepsilon'^2 \kappa}(\tau). \quad 17)$$

Zwecks Ableitung der Transformationsformel für die Substitution T sei die Funktion

$$(28) \quad A(\tau, \tau'; k) = \sum_{(\nu), \mu} \frac{1}{(\nu\tau + \mu)^2 (\nu'\tau' + \mu')^2 |\nu\tau + \mu|^{2k} |\nu'\tau' + \mu'|^{2k}}$$

$$= \sum_{(\nu), \mu} N\{(\nu\tau + \mu) |\nu\tau + \mu|^k\}^{-2}$$

eingeführt, wobei k eine positive Zahl ist und die Summation sich über alle ganzen Ideale (ν) und alle ganzen Körperzahlen μ aus $k(\sqrt{5})$ erstreckt. Zunächst soll die Konvergenz der Reihe nachgewiesen werden.

Zu jedem $s > 0$, $s' > 0$ existiert eine von ν und μ unabhängige Zahl $p > 0$, so daß

$$|\nu\tau + \mu|^2 > p(\nu^2 + \mu^2), \quad |\nu'\tau' + \mu'|^2 > p(\nu'^2 + \mu'^2).$$

Setzt man $\mu = (-1)^a \varepsilon^n \varrho$, läßt a die Zahlen 0 und 1, n alle ganzen rationalen Zahlen, (ϱ) alle ganzen Ideale von $k(\sqrt{5})$ durchlaufen, so durchläuft μ genau alle ganzen Zahlen aus $k(\sqrt{5})$, die Null ausgenommen. Es ist dann

$$N|\nu\tau + \mu|^2 > p^2 N(\nu^2 + \mu^2) = p^2 N(\nu^2 + \varepsilon^{2n} \varrho^2) > p^2 (\varepsilon^{2n} \nu'^2 \varrho^2 + \varepsilon'^{2n} \nu^2 \varrho'^2).$$

Denkt man, was stets möglich ist, ν und ϱ so gewählt, daß $\nu^2 > \frac{1}{2} |N(\nu)|$, $\nu'^2 > \frac{1}{2} |N(\nu)|$, $\varrho^2 > \frac{1}{2} |N(\varrho)|$, $\varrho'^2 > \frac{1}{2} |N(\varrho)|$, so folgt

$$N|\nu\tau + \mu|^2 > \frac{p^2}{4} (\varepsilon^{2n} + \varepsilon'^{2n}) |N(\nu)| \cdot |N(\varrho)|.$$

Daher hat (28) die Majorante

$$\left(\frac{1}{p^2}\right)^{1+k} \sum_{(\nu)} |N(\nu)|^{-2(1+k)} + 2 \left(\frac{4}{p^2}\right)^{1+k} \sum_{n=-\infty}^{+\infty} \sum_{(\nu), (\varrho)} \{(\varepsilon^{2n} + \varepsilon'^{2n}) |N(\nu)| \cdot |N(\varrho)|\}^{-(1+k)}.$$

Hierin ist die erste Summe eine konvergente Zetareihe in $k(\sqrt{5})$ und die letzte Summe ein Produkt von drei einfachen konvergenten Reihen, wenn $k > 0$. Damit ist die absolute Konvergenz der Reihe $A(\tau, \tau'; k)$ im Bereiche $s > 0$, $s' > 0$ nachgewiesen.

¹⁷⁾ Vgl. 13).

Als Teilsumme von (28) konvergiert auch

$$A_x(\tau, \tau'; k) = A_x(\tau; k) = \sum_{\substack{(\nu), \mu \\ \nu \equiv \mu \pmod{2}}} N\{(\nu\tau + \mu) | \nu\tau + \mu |^k\}^{-2},$$

worin x eine ganze Zahl aus $k(\sqrt{5})$ und $k > 0$ ist.

Im Falle $x \equiv 0 \pmod{2}$ ist

$$\begin{aligned} A_x(2\tau; k) &= \frac{1}{16^{1+k}} \sum_{(\nu), \mu} N\{(\nu\tau + \mu) | \nu\tau + \mu |^k\}^{-2} \\ &= \frac{1}{16^{1+k} (\tau\tau')^2 | \tau\tau' |^{2k}} \sum_{(\nu)} |N(\nu)|^{-2(1+k)} + \frac{1}{16^{1+k}} \sum_{\substack{(\nu), \mu \neq 0 \\ \nu \equiv \mu \pmod{2}}} N\{(\nu\tau + \mu) | \nu\tau + \mu |^k\}^{-2}. \end{aligned}$$

Da hierin die letzte Summe sich nur mit $(\tau\tau')^2 | \tau\tau' |^{2k}$ multipliziert, wenn man τ, τ' durch $-\frac{1}{\tau}, -\frac{1}{\tau'}$ ersetzt, so folgt die Funktionalgleichung

$$(29) \quad A_x\left(-\frac{2}{\tau}; k\right) + \frac{\zeta_k(2+2k)}{16^{1+k}} = (\tau\tau')^2 | \tau\tau' |^{2k} \left\{ A_x(2\tau; k) + \frac{\zeta_k(2+2k)}{16^{1+k}} \right\}$$

$(x \equiv 0 \pmod{2}),$

wo $\zeta_k(s) = \sum_{(\nu)} |N(\nu)|^{-s}$.

Im Falle $x \not\equiv 0 \pmod{2}$ ist

$$\begin{aligned} A_x(\tau; k) &= \frac{1}{(\tau\tau')^2 | \tau\tau' |^{2k}} \sum_{(\nu)} |N(2\nu)|^{-2(1+k)} + \sum_{\substack{(\nu), \mu \neq 0 \\ \nu \equiv \mu \pmod{2}}} N\{(\nu\tau + \mu) | \nu\tau + \mu |^k\}^{-2} \\ &= \frac{1}{(\tau\tau')^2 | \tau\tau' |^{2k}} \frac{\zeta_k(2+2k)}{16^{1+k}} + \sum_{\substack{(\mu), \nu \neq 0 \\ \mu \equiv \nu \pmod{2}}} N\{(\nu\tau + \mu) | \nu\tau + \mu |^k\}^{-2}. \end{aligned}$$

Vertauscht man τ, τ' mit $-\frac{1}{\tau}, -\frac{1}{\tau'}$, so folgt

$$(30) \quad A_x\left(-\frac{1}{\tau}; k\right) + \frac{\zeta_k(2+2k)}{16^{1+k}} = (\tau\tau')^2 | \tau\tau' |^{2k} \left\{ A_x(\tau; k) + \frac{\zeta_k(2+2k)}{16^{1+k}} \right\}$$

$(x \not\equiv 0 \pmod{2}).$

Auf die Funktion

$$\begin{aligned} f(m_1, m_2) &= N\{(\nu\tau + \varrho) | \nu\tau + \varrho |^k\}^{-2}, \\ \varrho &= 2(m_1 + m_2 \varepsilon) + \nu x, \quad \varrho' = 2(m_1 + m_2 \varepsilon') + \nu' x', \end{aligned}$$

wende man die Formel

$$\sum_{m_1, m_2 = -\infty}^{+\infty} f(m_1, m_2) = \sum_{n_1, n_2 = -\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{2\pi i(n_1 u_1 + n_2 u_2)} f(u_1, u_2) du_1 du_2$$

an. Dann folgt

$$\begin{aligned}
 & \sum_{\mu} N\{(\nu\tau + 2\mu + \nu\kappa) | \nu\tau + 2\mu + \nu\kappa |^k\}^{-2} \\
 &= \sum_{n_1, n_2 = -\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{e^{2\pi i(n_1 u_1 + n_2 u_2)} du_1 du_2}{N\{(\nu\tau + 2(u_1 + u_2\varepsilon) + \nu\kappa) | \nu\tau + 2(u_1 + u_2\varepsilon) + \nu\kappa |^k\}^{-2}} \\
 &= \frac{1}{4\sqrt{5}} \sum_{n_1, n_2} \int_{-\infty}^{+\infty} \frac{S^{\nu\mu\kappa} \frac{\pi i}{\sqrt{5}} (\mu v - \mu' v')}{(\nu\tau + v)^2 | \nu\tau + v |^{2k} (\nu'\tau' + v')^2 | \nu'\tau' + v' |^{2k}} dv dv' \\
 & \quad (\mu = n_2 - n_1 \varepsilon') \\
 &= \frac{1}{4\sqrt{5}} \sum_{\mu} \frac{(-1)^{\frac{S^{\nu\mu\kappa}}{\sqrt{5}}}}{|N(\nu)|^{1+2k}} \int_{-\infty}^{+\infty} \frac{e^{\frac{\pi i}{\sqrt{5}} \nu \mu v} dv}{(\tau + v)^2 | \tau + v |^{2k}} \int_{-\infty}^{+\infty} \frac{e^{-\frac{\pi i}{\sqrt{5}} \nu' \mu' v'} dv'}{(\tau' + v')^2 | \tau' + v' |^{2k}}.
 \end{aligned}$$

Summiert man noch über alle Ideale (ν) , so folgt

$$A_{\kappa}(\tau; k) = \frac{1}{4\sqrt{5}} \sum_{(\nu), \mu} \frac{(-1)^{\frac{S^{\nu\mu\kappa}}{\sqrt{5}}}}{|N(\nu)|^{1+2k}} \int_{-\infty}^{+\infty} \frac{e^{\frac{\pi i}{\sqrt{5}} \nu \mu v} dv}{(\tau + v)^2 | \tau + v |^{2k}} \int_{-\infty}^{+\infty} \frac{e^{-\frac{\pi i}{\sqrt{5}} \nu' \mu' v'} dv'}{(\tau' + v')^2 | \tau' + v' |^{2k}}.$$

Die so gewonnene Reihe für $A_{\kappa}(\tau; k)$ konvergiert auch noch für $k=0$. Sie hat nämlich, wegen

$$\int_{-\infty}^{+\infty} \frac{e^{\pi i c v}}{(t+v)^2} dv = \begin{cases} 0, & \text{wenn } c \geq 0, \Im t > 0 \text{ oder } c \leq 0, \Im t < 0, \\ 2\pi^2 c e^{-\pi i c t}, & \text{wenn } c < 0, \Im t > 0, \\ -2\pi^2 c e^{-\pi i c t}, & \text{wenn } c > 0, \Im t < 0, \end{cases}$$

an der Stelle $k=0$ den Wert

$$\frac{\pi^4}{5\sqrt{5}} \sum_{\substack{(\nu), \mu \\ \nu\mu > 0}} (-1)^{\frac{S^{\nu\mu\kappa}}{\sqrt{5}}} |N(\mu)| e^{\frac{\pi i S^{\nu\mu\tau}}{\sqrt{5}}} = \frac{\pi^4}{5\sqrt{5}} F_{\kappa}(\tau).$$

Da sie, wie leicht einzusehen, in $k \geq 0$ gleichmäßig konvergiert, so besteht der Grenzwert

$$\lim_{k \rightarrow 0} A_{\kappa}(\tau; k) = \frac{\pi^4}{5\sqrt{5}} F_{\kappa}(\tau),$$

und es folgt durch Grenzübergang ($k \rightarrow 0$) aus (29) und (30)

$$(31) \quad F_{\kappa}\left(-\frac{2}{\tau}\right) + C = (\tau\tau')^2 \{F_{\kappa}(2\tau) + C\} \quad (\kappa \equiv 0 \pmod{2}),$$

$$(32) \quad F_{\kappa}\left(-\frac{1}{\tau}\right) + C = (\tau\tau')^2 \{F_{\kappa}(\tau) + C\} \quad (\kappa \not\equiv 0 \pmod{2}),$$

wobei $C = \frac{5\sqrt{5}}{16\pi^4} \zeta_k(2) = \frac{1}{120}$, da $\zeta_k(2)$, was gleich dem Produkt aus der

gewöhnlichen Zetareihe und einer leicht summierbaren L -Reihe ist, den Wert $\frac{2\sqrt{5}}{375}\pi^4$ hat.

Setzt man allgemein

$$(33) \quad F_{\kappa}(\tau) + \frac{1}{120} = G_{\kappa}(\tau)$$

und außerdem, falls $\kappa \equiv 0 \pmod{2}$,

$$(34) \quad 16 \left\{ F_{\kappa}(4\tau) + \frac{1}{120} \right\} = G_{\infty}(\tau),$$

wobei der Index ∞ die in § 4 festgesetzten Eigenschaften haben möge, definiert man ferner $G_{\kappa}(\tau)$ für jede beliebige Körperzahl κ durch die Forderung $G_{\kappa}(\tau) = G_{\lambda}(\tau)$, wenn die Kongruenz $\kappa \equiv \lambda \pmod{2}$ im Sinne von § 4 besteht, so kann man die Formeln (27), (31) und (32) zusammenfassen in der Gestalt

$$G_{\kappa}(\tau + 1) = G_{\kappa-1}(\tau), \quad G_{\kappa}(\varepsilon^2\tau) = G_{\varepsilon^2\kappa}(\tau), \quad G_{\kappa}\left(-\frac{1}{\tau}\right) = (\tau\tau')^2 G_{-\frac{1}{\kappa}}(\tau).$$

Hieraus folgt (wie beim Beweise von Hilfssatz 6) für eine beliebige

Modulsubstitution $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$

$$(35) \quad G_{\kappa}\left(\frac{\alpha\tau + \beta}{\gamma\tau + \delta}\right) = N(\gamma\tau + \delta)^2 G_{\frac{-\delta\kappa + \beta}{\gamma\kappa - \alpha}}(\tau)$$

und ebenso aus (27), (31), (33)

$$(36) \quad G_{\kappa}\left(2\frac{\alpha\tau + \beta}{\gamma\tau + \delta}\right) = N(\gamma\tau + \delta)^2 G_{\kappa}(2\tau) \quad (\kappa \equiv 0 \pmod{2}).$$

Setzt man für irgend zwei mod 2 inkongruente Indizes κ, λ

$$(37) \quad 8 \{G_{\kappa}(\tau) + G_{\lambda}(\tau)\} - 16 G_0(2\tau) = \Theta_{\kappa\lambda}(\tau, \tau') = \Theta_{\kappa\lambda}(\tau),$$

so erhält man zehn Funktionen $\Theta_{\kappa\lambda}(\tau) = \Theta_{\lambda\kappa}(\tau)$, die wegen (35) und (36) die Transformationseigenschaft

$$(38) \quad \Theta_{\kappa\lambda}\left(\frac{\alpha\tau + \beta}{\gamma\tau + \delta}\right) = N(\gamma\tau + \delta)^2 \Theta_{\frac{-\delta\kappa + \beta}{\gamma\kappa - \alpha}, \frac{-\delta\lambda + \beta}{\gamma\lambda - \alpha}}(\tau)$$

für jede Modulsubstitution $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ besitzen. Daraus folgt in Verbindung mit Hilfssatz 6

Hilfssatz 8. Die zehn Funktionen

$$(39) \quad \Theta_{\kappa\lambda}^2(\tau) \vartheta_{\kappa\lambda}^{-8}(\tau)$$

erleiden bei Ausübung einer beliebigen Modulsubstitution lediglich eine Permutation.

§ 7.

Schluß des Beweises.

Damit auf die zehn Funktionen (39) Hilfssatz 5 anwendbar wird, muß noch ihre Regularität in \mathfrak{M} nachgewiesen werden. Diese folgt nun aus Hilfssatz 7 ohne weiteres in den Fällen $\kappa + \lambda \equiv 1$ oder $\infty \pmod{2}$, während für die restlichen vier Funktionen sich offenbar nur die Regularität

von $(1 + (-1)^{\frac{S \kappa \alpha'}{\sqrt{5}}} h_{\alpha'})^8 \Theta_{\kappa \lambda}^2(\tau) \vartheta_{\kappa \lambda}^{-8}(\tau)$ ($\kappa + \lambda \equiv \alpha \pmod{2}$, $\alpha = \varepsilon$ oder ε') ergibt. In der Tat verschwindet in diesen Fällen, wie aus § 5 hervorgeht,

$\vartheta_{\kappa \lambda}(\tau)$ identisch für $1 + (-1)^{\frac{S \kappa \alpha'}{\sqrt{5}}} h_{\alpha'} = 0$. Zur Beseitigung dieser Schwierig-

keit muß gezeigt werden, wie es in Hilfssatz 10 geschehen soll, daß die vier Funktionen $\Theta_{\kappa \lambda}(\tau)$ ($\kappa + \lambda \equiv \alpha \pmod{2}$, $\alpha = \varepsilon$ oder ε') in ihrer Ab-

hängigkeit von $h_1, h_{\alpha'}$ betrachtet, an der Stelle $h_{\alpha'} = -(-1)^{\frac{S \kappa \alpha'}{\sqrt{5}}}$ von vierter Ordnung identisch in h_1 verschwinden. Der Nachweis hiervon erfolgt mit Hilfe von Modulfunktionen einer Veränderlichen auf Grund des bekannten Satzes:

Hilfssatz 9. Eine zur Modulgruppe \mathfrak{G}_1 (der Modulgruppe im rationalen Zahlkörper) gehörige Modulfunktion $\Phi(u)$, die im Fundamentalebereich \mathfrak{M}_1 (im Bereiche $-\frac{1}{2} \leq \Re u \leq \frac{1}{2}$, $|u| \geq 1$, $\Im u > 0$) regulär und beschränkt ist, ist eine Konstante.

Hilfssatz 10. Die Funktionen

$$(1 + (-1)^{\frac{S \kappa \alpha'}{\sqrt{5}}} h_{\alpha'})^{-4} \Theta_{\kappa \lambda}(\tau) \quad (\kappa + \lambda \equiv \alpha \pmod{2}, \alpha = \varepsilon \text{ oder } \varepsilon')$$

sind regulär in \mathfrak{M} .

Beweis. Die Funktion $\Theta_{1\varepsilon}(\tau)$ geht durch die Substitution

$$\tau = -\varepsilon' u + v, \quad \tau' = -\varepsilon u + v \quad \text{oder} \quad u = S \frac{\tau}{\sqrt{5}}, \quad v = S \frac{\varepsilon \tau}{\sqrt{5}}$$

in eine für $\Im(-\varepsilon' u + v) > 0$, $\Im(-\varepsilon u + v) < 0$ reguläre Funktion $\varphi(u, v)$ über, wobei

$$\varphi(u + 1, v) = \Theta_{1\varepsilon}(\tau - \varepsilon'),$$

$$\varphi\left(-\frac{u+v}{u^2+uv-v^2}, \frac{v}{u^2+uv-v^2}\right) = \Theta_{1\varepsilon}\left(-\frac{1}{\varepsilon^2 \tau}\right).$$

Nun ist nach (38) $\Theta_{1\varepsilon}(\tau - \varepsilon') = \Theta_{1\varepsilon}(\tau)$ und $\Theta_{1\varepsilon}\left(-\frac{1}{\varepsilon^2 \tau}\right) = (\tau \tau')^2 \Theta_{1\varepsilon}(\tau)$, so daß

$$(40) \quad \left\{ \begin{array}{l} \varphi(u + 1, v) = \varphi(u, v), \\ \varphi\left(-\frac{u+v}{u^2+uv-v^2}, \frac{v}{u^2+uv-v^2}\right) = (u^2 + uv - v^2)^2 \varphi(u, v). \end{array} \right.$$

Für $v = 0$ lauten diese Formeln

$$\varphi(u+1, 0) = \varphi(u, 0), \quad \varphi\left(-\frac{1}{u}, 0\right) = u^4 \varphi(u, 0).$$

Dabei ist $\varphi(u, 0)$ eine für $\Im u > 0$ reguläre Funktion.

Die Dedekindsche η -Funktion

$$\eta(u) = e^{\frac{\pi i u}{12}} \prod_{n=1}^{\infty} (1 - e^{2\pi i n u}),$$

eine für $\Im u > 0$ reguläre und nicht verschwindende Funktion, hat die Eigenschaften

$$\eta^{24}(u+1) = \eta^{24}(u), \quad \eta^{24}\left(-\frac{1}{u}\right) = u^{12} \eta^{24}(u); \quad e^{-2\pi i u} \eta^{24}(u) \rightarrow 1,$$

wenn $u \rightarrow \infty$ in \mathfrak{M}_1 . Folglich genügt $\Phi(u) = \varphi^3(u, 0) \eta^{-24}(u)$ den Transformationsformeln

$$\Phi(u+1) = \Phi(u), \quad \Phi\left(-\frac{1}{u}\right) = \Phi(u)$$

und ist somit eine zur Modulgruppe \mathfrak{G}_1 gehörige Modulfunktion. Da $e^{-\pi i u} \varphi(u, 0) \rightarrow c_1$ ¹⁸⁾ für $u \rightarrow \infty$ in \mathfrak{M}_1 , ¹⁹⁾ so strebt

$$(41) \quad \Phi(u) \rightarrow 0, \quad \text{wenn } u \rightarrow \infty \text{ in } \mathfrak{M}_1.$$

Somit ist $\Phi(u)$ regulär und beschränkt in \mathfrak{M}_1 , verschwindet also wegen Hilfssatz 9 und (41) identisch. Daher ist auch

$$(42) \quad \varphi(u, 0) = 0.$$

Aus (40) und (42) folgt für $\varphi'(u, v) = \frac{\partial \varphi(u, v)}{\partial v}$

$$\varphi'(u+1, 0) = \varphi'(u, 0), \quad \varphi'\left(-\frac{1}{u}, 0\right) = u^6 \varphi'(u, 0).$$

Demnach ist $\varphi'^2(u, 0) \eta^{-24}(u)$ eine zur Modulgruppe \mathfrak{G}_1 gehörige Modulfunktion, die wegen $e^{-\pi i u} \varphi'(u, 0) \rightarrow 0$ für $u \rightarrow \infty$ in \mathfrak{M}_1 ¹⁹⁾ nach Hilfssatz 9 eine Konstante, und zwar $= 0$ ist. Folglich ist

$$(43) \quad \varphi'(u, 0) = 0.$$

¹⁸⁾ c_1, c_2, c_3 sollen Konstanten bedeuten.

¹⁹⁾ Es ist nämlich, wie aus (26), (33), (37) wegen $\varphi(u, v) = \theta_{1, \varepsilon}(\tau)$ folgt,

$$\frac{1}{(\pi i)^k} \left\{ \frac{\partial^k}{\partial v^k} \varphi(u, v) \right\}_{v=0} = 8 \sum_{\substack{\nu, \mu \\ \nu \mu > 0, (\mu, 2)=1}} |N(\nu)| \{(-1)^{n_2} + (-1)^{n_1+n_2}\} n_2^k h_1^{n_1},$$

$$h_1 = e^{\pi i u}, \quad \nu \mu = n_1 + n_2 \varepsilon \quad (k = 0, 1, 2, \dots).$$

Da $n_1 \geq 1$ wegen $\nu \mu > 0$ und $(-1)^{n_2} + (-1)^{n_1+n_2} = 0$ für $n_1 = 1$, so enthält diese Potenzreihe in h_1 keine Glieder mit h_1^0 und h_1^1 .

Aus (40), (42), (43) folgt für $\varphi''(u, v) = \frac{\partial^2 \varphi(u, v)}{\partial v^2}$

$$\varphi''(u+1, 0) = \varphi''(u, 0), \quad \varphi''\left(-\frac{1}{u}, 0\right) = u^3 \varphi''(u, 0)$$

und hieraus bei Betrachtung der Funktion $\varphi''^3(u, 0) \eta^{-48}(u)$ wegen $e^{-2\pi i u} \varphi''(u, 0) \rightarrow c_2$ für $u \rightarrow \infty$ in \mathfrak{M}_1^{19}) analog wie vorher

$$(44) \quad \varphi''(u, 0) = 0.$$

Aus (40), (42), (43), (44) folgt für $\varphi'''(u, v) = \frac{\partial^3 \varphi(u, v)}{\partial v^3}$

$$\varphi'''(u+1, 0) = \varphi'''(u, 0), \quad \varphi''' \left(-\frac{1}{u}, 0\right) = u^{10} \varphi'''(u, 0)$$

und hieraus bei Betrachtung von $\varphi'''^6(u, 0) \eta^{-120}(u)$ wegen $e^{-2\pi i u} \varphi'''(u, 0) \rightarrow c_3$ für $u \rightarrow \infty$ in \mathfrak{M}_1^{19})

$$(45) \quad \varphi'''(u, 0) = 0.^{30)}$$

Die Formeln (42), (43), (44), (45) besagen, daß $v^{-4} \varphi(u, v)$ und damit auch $(1 - e^{\pi i v})^{-4} \varphi(u, v)$ eine im ganzen Definitionsbereich $\Im(-\varepsilon' u + v) > 0$, $\Im(-\varepsilon u + v) < 0$ reguläre Funktion von u und v ist. Wegen $e^{\pi i v} = h_s$, $\varphi(u, v) = \Theta_{1_s}(\tau)$ ist also

$$(46) \quad (1 - h_s)^{-4} \Theta_{1_s}(\tau),$$

als Funktion von τ, τ' aufgefaßt, regulär im Bereiche $s > 0$, $s' > 0$. Das gleiche gilt somit von

$$(47) \quad (1 + h_s)^{-4} \Theta_{0_s}(\tau),$$

wozu man gelangt, wenn man in (46) τ, τ' durch $\tau+1, \tau'+1$ ersetzt. Berücksichtigt man nun noch, daß $\Theta_{1_s}(-\tau', -\tau) = \Theta_{1_s}(\tau, \tau')$ und $\Theta_{0_s}(-\tau', -\tau) = \Theta_{0_s}(\tau, \tau')$, was sich unmittelbar aus (26), (33), (37) ergibt, so folgt, wenn man in (46) und (47) τ, τ' mit $-\tau', -\tau$ vertauscht, daß auch $(1 - h_s)^{-4} \Theta_{1_s}(\tau)$ und $(1 + h_s)^{-4} \Theta_{0_s}(\tau)$ regulär im Bereiche $s > 0$, $s' > 0$ sind. Damit ist Hilfssatz 10 bewiesen.

Auf Grund der Hilfssätze 7, 8 und 10 sind nunmehr die Voraussetzungen von Hilfssatz 5 für die zehn Funktionen (39) erfüllt. Folglich sind diese zehn Funktionen Konstanten. Speziell ist also

$$\vartheta_{0\infty}^4(\tau) = c \cdot \Theta_{0\infty}(\tau),$$

³⁰⁾ Setzt man das Verfahren noch einen Schritt fort, so erhält man den folgenden Zusammenhang zwischen $\eta^{24}(u)$ und $\varphi^{(4)}(u, v) = \frac{\partial^4 \varphi(u, v)}{\partial v^4}$:

$$-384 \pi^4 \eta^{24}(u) = \varphi^{(4)}(u, 0).$$

wo c eine Konstante ist. Nun ist nach (19), (20)

$$\vartheta_{0\infty}^4(\tau) = \left(\sum_{\nu} e^{\pi i S \frac{\nu^2 \tau}{\sqrt{5}}} \right)^4 = 1 + \sum_{\nu > 0} a_{\nu} e^{\pi i S \frac{\nu \tau}{\sqrt{5}}},$$

wo a_{ν} gleich der Anzahl der Darstellungen von ν als Summe von vier ganzen Quadratzahlen aus $k(\sqrt{5})$ ist. Ferner folgt aus (26), (33), (34), (37)

$$\begin{aligned} \Theta_{0\infty}(\tau) &= 1 + 8 \sum_{\substack{(\nu), \mu \\ \nu \mu > 0}} |N(\nu)| e^{\pi i S \frac{\nu \mu \tau}{\sqrt{5}}} - 16 \sum_{\substack{(\nu), \mu \\ \nu \mu > 0}} |N(\nu)| e^{2\pi i S \frac{\nu \mu \tau}{\sqrt{5}}} + 128 \sum_{\substack{(\nu), \mu \\ \nu \mu > 0}} |N(\nu)| e^{4\pi i S \frac{\nu \mu \tau}{\sqrt{5}}} \\ &= 1 + \sum_{\mu > 0} \left\{ 8 \sum_{(\nu) | \mu} |N(\nu)| - 4 \sum_{2 | (\nu) | \mu} |N(\nu)| + 8 \sum_{4 | (\nu) | \mu} |N(\nu)| \right\} e^{\pi i S \frac{\mu \tau}{\sqrt{5}}} \end{aligned}$$

Hierbei durchläuft μ alle total positiven ganzen Zahlen aus $k(\sqrt{5})$, während innerhalb der Klammer (ν) in der ersten Summe alle, in der zweiten alle geraden, in der dritten alle durch 4 teilbaren Idealeiler von μ durchläuft. Da die Reihen für $\vartheta_{0\infty}^4(\tau)$ und $\Theta_{0\infty}(\tau)$ mit demselben konstanten Glied beginnen, so muß $c = 1$ sein. Es besteht daher die Identität

$$\vartheta_{0\infty}^4(\tau) = \Theta_{0\infty}(\tau),$$

und durch Koeffizientenvergleich folgt

Satz 4. Jede ganze total positive Zahl μ aus $k(\sqrt{5})$ ist auf genau

$$8 \sum_{(\nu) | \mu} |N(\nu)| - 4 \sum_{2 | (\nu) | \mu} |N(\nu)| + 8 \sum_{4 | (\nu) | \mu} |N(\nu)|$$

Arten als Summe von vier ganzen Quadratzahlen aus $k(\sqrt{5})$ darstellbar.

Betrachtet man ungerade und gerade μ gesondert, so kann man das Ergebnis auch in Form von Satz I der Einleitung aussprechen.

Bieber b. Offenbach, 24. Oktober 1927.

(Eingegangen am 8. 11. 1927.)

Der zweite Mittelwertssatz der Integralrechnung für komplexe Größen.

Von

Otto Hölder in Leipzig.

1.

Formulierung des Satzes.

Nehmen wir zunächst im reellen Intervall $a \dots b$ die reelle Funktion $f(x)$ als monoton und beschränkt, $\varphi(x)$ dagegen als eine nur im Riemannschen Sinne integrierbare reelle Funktion an, so kann das Integral

$$(1) \quad \int_a^b f(x) \varphi(x) dx \quad (a < b)$$

nach dem zweiten Mittelwertssatz in jede der beiden Formen

$$(2) \quad [f(a+0) - f(b-0)] M_1 + f(b-0) \int_a^b \varphi(x) dx$$

und

$$(3) \quad f(a+0) \int_a^b \varphi(x) dx + [f(b-0) - f(a+0)] M_2$$

gesetzt werden. Es bedeutet dabei M_1 einen Mittelwert aus den Werten, die unter der Bedingung $a \leq x \leq b$ von dem Integral

$$(4) \quad \int_a^x \varphi(\alpha) d\alpha$$

angenommen werden, und M_2 einen Mittelwert aus den von dem Integral

$$(5) \quad \int_x^b \varphi(\alpha) d\alpha$$

unter derselben Bedingung angenommenen Werten.

In dem vorliegenden, reellen Fall kann man

$$M_1 = \int_a^{\xi} \varphi(x) dx, \quad M_2 = \int_{\xi'}^b \varphi(x) dx$$

setzen. Führt man diese Ausdrücke in (2) und (3) ein, so erkennt man, daß man $\xi' = \xi$ nehmen und die Formeln (2) und (3) ineinander umrechnen kann. Man gewinnt daraus dann auch unmittelbar die dritte, symmetrische Gestalt

$$f(a+0) \int_a^{\xi} \varphi(x) dx + f(b-0) \int_{\xi}^b \varphi(x) dx$$

der Formel, die am häufigsten erwähnt wird, aber für das Folgende nicht in Betracht kommt.

Es scheint nun noch nicht bemerkt worden zu sein, daß der Mittelwertssatz in den beiden Formen (2) und (3) auch dann gültig ist, wenn zwar $f(x)$ nach wie vor eine monotone und beschränkte reelle Funktion, aber $\varphi(x)$ eine *komplexe*, im Riemannschen Sinne integrierbare¹⁾ Funktion des im *reellen* Intervall $a \dots b$ beweglichen Argumentes bedeutet.

2.

Komplexe Mittelwerte.

Für diese Verallgemeinerung des Satzes kommt es vor allem darauf an, was unter einem „Mittelwert“ von komplexen Größen zu verstehen ist. Denkt man sich eine Anzahl von komplexen Zahlen $\zeta_1, \zeta_2, \dots, \zeta_n$, die zugleich durch Punkte der komplexen Zahlenebene vorgestellt sind, und bildet man aus ihnen mit Hilfe der reell positiven und von Null verschiedenen „Gewichte“ $\alpha_1, \alpha_2, \dots, \alpha_n$ den Ausdruck

$$(6) \quad \frac{\alpha_1 \zeta_1 + \alpha_2 \zeta_2 + \dots + \alpha_n \zeta_n}{\alpha_1 + \alpha_2 + \dots + \alpha_n},$$

so ist dieser durch einen Punkt der komplexen Zahlenebene vorgestellt, der (als innerer oder als Grenzpunkt) dem kleinsten konvexen, um die obigen n Punkte gezogenen Polygon angehört²⁾.

Sind jetzt komplexe Werte

$$(7) \quad \zeta, \zeta', \zeta'', \dots$$

in eventuell unendlicher Zahl gegeben, deren darstellende Punkte wir uns aber in einem ganz im Endlichen gelegenen Teil der komplexen Zahlenebene denken wollen, so kann man auf folgende Weise ein kleinstes, kon-

¹⁾ Dies soll heißen, daß die aus einer Teilung des Intervalls hervorgehende Summe $\sum \Delta_v D_v$ für unendliche Verkleinerung der Intervallteile Δ gegen Null geht, wobei D_v die obere Grenze der Absolutbeträge der Differenzen bedeutet für irgend zwei im Intervall Δ_v genommene Werte der Funktion.

²⁾ Die geometrische Deutung, die Schwarz und Fejér einem Satz von Gauß gegeben haben (Gauß, Werke 3 (1876), S. 112, Anm.; Fejér, Comptes Rendus 145 (1907), S. 460), hat mich hier auf die richtige Auffassung geführt.

vexes und abgeschlossenes Gebiet definieren, dem die Punkte (7) angehören. Man nimmt aus den Werten (7) irgendwelche: $\zeta_1, \zeta_2, \dots, \zeta_n$, in endlicher aber beliebiger Zahl, heraus und bildet aus ihnen mit beliebigen Gewichten den Ausdruck (6). Die sämtlichen komplexen Zahlen, die man auf diese Weise erhält, oder denen man durch solche Zahlen unendlich nahe kommen kann, stellen das in der Zahlenebene zu definierende Gebiet vor.

Man kann auch geometrisch folgendermaßen definieren:

1. Man geht von den Punkten (7) aus;
2. so oft zwei Punkte dem Gebiet angehören, fügt man die Punkte ihrer Verbindungsstrecke dem Gebiet hinzu und fährt mit allen neuen Punkten in derselben Weise fort;
3. außer den Punkten, die so erreicht werden, fügt man noch deren Grenzpunkte hinzu.

Daß sich diese geometrische Definition mit der vorigen deckt, ersieht man durch die Überlegung, daß die für $n=2$ genommene Formel (6):

$$\frac{\alpha_1}{\alpha_1 + \alpha_2} \zeta_1 + \frac{\alpha_2}{\alpha_1 + \alpha_2} \zeta_2$$

bei passender Wahl der Gewichte α_1 und α_2 jeden zwischen ζ_1 und ζ_2 auf der Verbindungsstrecke liegenden Punkt vorstellen kann, und daß die Gleichung

$$\frac{\alpha_1 \zeta_1 + \alpha_2 \zeta_2 + \dots + \alpha_n \zeta_n}{\alpha_1 + \alpha_2 + \dots + \alpha_n} \\ = \frac{\alpha_1 + \alpha_2 + \dots + \alpha_{n-1}}{\alpha_1 + \alpha_2 + \dots + \alpha_n} \cdot \frac{\alpha_1 \zeta_1 + \alpha_2 \zeta_2 + \dots + \alpha_{n-1} \zeta_{n-1}}{\alpha_1 + \alpha_2 + \dots + \alpha_{n-1}} + \frac{\alpha_n}{\alpha_1 + \alpha_2 + \dots + \alpha_n} \cdot \zeta_n$$

eine sukzessive Konstruktion des durch (6) dargestellten Punktes erkennen läßt³⁾, während andererseits jede homogene lineare Funktion der betrachteten Art (mit Koeffizienten, die positiv sind und die Summe 1 haben) von solchen linearen Funktionen wieder eine solche lineare Funktion wird.

Man erkennt ferner, daß auch die Verbindungsstrecke zweier Grenzpunkte stets vollständig zum Gebiet gehört, da jeder Punkt einer solchen Strecke durch solche Punkte, die erreicht werden, angenähert werden kann. Nun erkennt man auch aus der geometrischen Betrachtung, daß die konstruierte Punktmenge zusammenhängend und konvex⁴⁾ und wegen der Hinzurechnung der Grenzpunkte abgeschlossen ist, und daß jede die vor-

³⁾ Entsprechend der sukzessiven Konstruktion des Schwerpunktes bei n einzelnen Massen.

⁴⁾ Ich schließe mich hier der Definition von Minkowski an, wonach die Punktmenge (das Gebiet) konvex ist, wenn die ganze (geradlinige) Verbindungsstrecke zweier Punkte der Menge zur Menge gehört.

gegebenen Punkte (7) enthaltende konvexe und abgeschlossene Punktmenge die jetzt konstruierte Menge enthalten muß. Wir haben also in der Tat das kleinste konvexe und abgeschlossene Gebiet mit den Punkten (7) definiert. Jede in diesem Gebiet oder auf seiner Grenze gelegene komplexe Zahl soll als *Mittelwert* aus den Zahlen (7) in Anspruch genommen werden.

3.

Der Beweis.

Der formale Teil des Beweises ist so ziemlich derselbe wie im reellen Fall⁵⁾. Man betrachtet an Stelle des Integrals (1) die Summe

$$(8) \quad \sum_{\nu=0}^{\nu=n-1} f(\xi_{\nu}) \int_{x_{\nu}}^{x_{\nu+1}} \varphi(x) dx.$$

Dabei sollen die Grenzen a und b des obigen Integrals auch mit x_0 und x_n bezeichnet, und im Intervall von a bis b die Werte x_1, x_2, \dots, x_{n-1} der Reihe nach eingeschaltet sein, während $\xi_0, \xi_1, \xi_2 \dots \xi_{n-1}$ Argumentwerte bedeuten, die irgendwie in den einzelnen Teilintervallen liegen. Es wird nun die Summe (8) durch partielle Summation in

$$\sum_{\nu=0}^{\nu=n-2} \left\{ (f(\xi_{\nu}) - f(\xi_{\nu+1})) \sum_{\varrho=0}^{\varrho=\nu} \int_{x_{\varrho}}^{x_{\varrho+1}} \varphi(x) dx \right\} + f(\xi_{n-1}) \sum_{\varrho=0}^{\varrho=n-1} \int_{x_{\varrho}}^{x_{\varrho+1}} \varphi(x) dx,$$

d. h. in

$$(9) \quad \sum_{\nu=0}^{n-2} \left\{ (f(\xi_{\nu}) - f(\xi_{\nu+1})) \int_a^{x_{\nu+1}} \varphi(x) dx \right\} + f(\xi_{n-1}) \int_a^b \varphi(x) dx$$

übergeführt. Die Differenz zwischen (1) und (8) ist

$$\sum_{\nu=0}^{\nu=n-1} \int_{x_{\nu}}^{x_{\nu+1}} \{ (f(x) - f(\xi_{\nu})) \varphi(x) dx \}.$$

Werden jetzt die Intervallteile unendlich klein gemacht, so wird, wie man aus der Integrabilität der Funktionen erkennt, der Absolutbetrag dieser komplexen Differenz unendlich klein; dasselbe gilt aber auch von der absoluten Differenz zwischen dem zweiten Glied von (9) und der Größe

$$f(b-0) \int_a^b \varphi(x) dx.$$

Man kommt deshalb zu der Gleichung

⁵⁾ M. vgl. z. B. meine Ausführungen in den Göttinger Gelehrten Anzeigen 1894, S. 519—522 oder P. du Bois-Reymond, Journ. f. d. reine u. angew. Math. 79 (1875), S. 42. Übrigens dürfte die Anordnung des Textes sich auch im reellen Fall noch mehr empfehlen.

$$\int_a^b f(x) \varphi(x) dx - f(b-0) \int_a^b \varphi(x) dx$$

$$= [f(\xi_0) - f(\xi_{n-1})] \frac{\sum_{\nu=0}^{n-2} \left\{ (f(\xi_\nu) - f(\xi_{\nu+1})) \int_a^{\xi_{\nu+1}} \varphi(x) dx \right\}}{\sum_{\nu=0}^{n-2} (f(\xi_\nu) - f(\xi_{\nu+1}))} + \varepsilon,$$

wo

$$\lim \varepsilon = 0$$

ist für den Fall, daß die Teilintervalle unendlich klein werden⁶⁾.

Da nun die in eckige Klammer geschlossene Größe dem Wert $f(a+0) - f(b-0)$ unendlich nahe kommt, und der obige Bruch, in dem alle Differenzen, soweit sie von Null verschieden sind, dasselbe Vorzeichen haben, einen Ausdruck von der Form (6) vorstellt, der aus den Größen

$$\zeta_\nu = \int_a^{\xi_{\nu+1}} \varphi(x) dx \quad (\nu = 0, 1, 2, \dots, n-2)$$

gebildet ist, so kann man in der Tat durch Ausdrücke der genannten Form dem Quotienten

$$\frac{\int_a^b f(x) \varphi(x) dx - f(b-0) \int_a^b \varphi(x) dx}{f(a+0) - f(b-0)}$$

unendlich nahe kommen. Es ist also dieser Quotient gleich einem Mittelwert M_1 aus den, jetzt komplexen, Größen (4), womit die Formel (2) bewiesen ist.

Schreibt man die Glieder der Summe (8) in umgekehrter Ordnung

$$\sum_{\nu=n-1}^0 f(\xi_\nu) \int_{\xi_\nu}^{\xi_{\nu+1}} \varphi(x) dx,$$

so ergibt die in gewöhnlicher Weise ausgeführte partielle Summation

$$\sum_{\nu=n-1}^{\nu=1} \left\{ (f(\xi_\nu) - f(\xi_{\nu-1})) \int_{\xi_\nu}^b \varphi(x) dx \right\} + f(\xi_0) \int_a^b \varphi(x) dx,$$

und man erhält durch die entsprechende Überlegung den Mittelwertssatz in der Form (3), in welcher ein Mittelwert der jetzt komplexen Größen (5) auftritt.

4.

Beispiel.

Das folgende Beispiel soll zeigen, daß der eben bewiesene Satz für eine komplexe Funktion $\varphi(x)$ reellen Arguments mehr ergibt als die An-

⁶⁾ Da mit $f(\xi_0) - f(\xi_{n-1})$ multipliziert und dividiert worden ist, ist vorausgesetzt, daß $f(x)$ nicht im ganzen Innern des Intervalles $a \dots b$ konstant ist. Im umgekehrten Fall bedürfte aber die Gleichung (2) keines Beweises.

wendung des reellen Mittelwertsatzes auf das in den reellen und imaginären Teil gespaltene Integral. Ich nehme

$$(10) \quad \int_0^{+\infty} \frac{e^{ix} dx}{1 + \psi(x)},$$

wobei für das reelle, aus der nächsten Nähe von 0 bis $+\infty$ wachsende x die reelle Funktion $\psi(x)$ monoton von 0 bis $+\infty$ wachsen soll. Wenn

$$\frac{1}{1 + \psi(x)} = f(x), \quad e^{ix} = \varphi(x)$$

gesetzt wird, sind die Bedingungen des Mittelwertsatzes erfüllt. Ich führe das Integral zunächst nur über den Intervallausschnitt $\omega \dots \omega'$ ($\omega < \omega'$) und erhalte aus der ersten Formulierung (2) des Mittelwertsatzes die Gleichung

$$(11) \quad \int_{\omega}^{\omega'} \frac{e^{ix} dx}{1 + \psi(x)} = \left[\frac{1}{1 + \psi(\omega + 0)} - \frac{1}{1 + \psi(\omega' - 0)} \right] M_1 + \frac{1}{1 + \psi(\omega' - 0)} \int_{\omega}^{\omega'} e^{ix} dx,$$

wo M_1 einen Mittelwert aus den Größen

$$(12) \quad \int_{\omega}^x e^{i\alpha} d\alpha \quad (\omega \leq x \leq \omega')$$

bedeutet. Da nun

$$\int_{\omega}^x e^{i\alpha} d\alpha = -i(e^{ix} - e^{i\omega})$$

ist, erkennt man zunächst, daß alle die Integrale (12), und also auch alle ihre Mittelwerte absolut kleiner oder gleich 2 sind. Im Hinblick darauf zeigt aber die Formel (11), daß das über den Intervallausschnitt $\omega \dots \omega'$ geführte Integral mit $\omega \rightarrow +\infty$ gegen Null geht, daß somit das Integral (10) konvergent ist.

Jetzt kann man in der Gleichung (11) den Wert ω' gegen $+\infty$ gehen lassen. Man erhält, wenn noch $\omega = 0$ gesetzt wird,

$$(13) \quad \int_0^{+\infty} \frac{e^{ix} dx}{1 + \psi(x)} = \bar{M}_1,$$

wo nun \bar{M}_1 einen Mittelwert aus den Integralen

$$(14) \quad \int_0^x e^{i\alpha} d\alpha = i - i e^{ix} \quad (0 \leq x < +\infty)$$

bedeutet.

Den Ausdruck (14) untersuche ich geometrisch. Er ist in der Zahlenebene durch den Vektor vorgestellt, der von dem Punkt $i e^{ix}$ des Einheitskreises nach dem Punkt $+i$ hinführt. Denkt man sich nun alle die von

$+i$ nach den Punkten des Einheitskreises führenden Vektoren (s. Fig. 1) von $+i$ aus ins Entgegengesetzte verkehrt und dann dieses ganze System, indem es als ein starres Gebilde betrachtet wird, so parallel verschoben, daß der Punkt $+i$ in den Nullpunkt gebracht wird, so ergeben die Endpunkte der betrachteten Vektoren schließlich die Peripherie des in Fig. 2 schraffierten Kreises. In diesem Kreis ist uns also jetzt jenes kleinste konvexe Gebiet gegeben. Der komplexe Mittelwert \bar{M}_1 , d. h. nach Gleichung (13) das Integral (10), muß sich als ein Punkt der schraffierten Kreisscheibe darstellen. Macht nun der nach diesem Punkt vom Nullpunkt aus gezogene Vektor mit der positiven reellen Achse den Winkel γ , so ist $2\vartheta \sin \gamma$ die Länge des Vektors, wobei

$$0 \leq \vartheta \leq 1, \quad 0 \leq \gamma \leq \pi,$$

und es ist unser Integral (10) durch $2\vartheta \sin \gamma (\cos \gamma + i \sin \gamma)$ ausgedrückt.

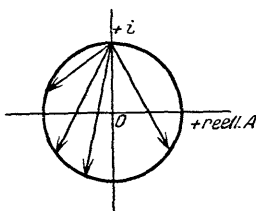


Fig. 1.

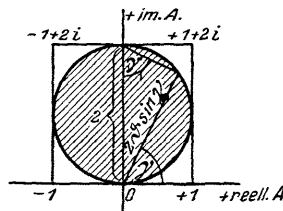


Fig. 2.

Hätte man das Integral (10) in den reellen und den imaginären Teil zerlegt und auf jeden der Teile den reellen Mittelwertssatz angewendet, so hätte sich nur ergeben, daß der reelle Teil dem Intervall $-1 \dots +1$, die imaginäre Koordinate dem Intervall $0 \dots 2$ angehört⁷⁾. Hieraus ergäbe sich also bloß, daß das Integral durch einen Punkt der Quadratafläche (Fig. 2) vorgestellt ist.

⁷⁾ Eine Betrachtung, die der Herleitung der Gleichung (13) über die Gleichung (11) hinweg völlig analog ist, zeigt nämlich, daß die Integrale

$$\int_0^{\infty} \frac{\cos x \, dx}{1 + \varphi(x)}, \quad \int_0^{\infty} \frac{\sin x}{1 + \varphi(x)} \, dx$$

beziehungsweise Mittelwerte sind von

$$\int_0^x \cos \alpha \, d\alpha, \quad \int_0^x \sin \alpha \, d\alpha,$$

wenn x im Intervall $0 \dots +\infty$ variiert.

Über das Flächenmaß rektifizierbarer Flächen¹⁾.

Von

Tibor Radó in Szeged (Ungarn).

Einleitung.

Die vorliegende Arbeit stellt einen Ausschnitt aus Ergebnissen dar, zu welchen ich im Laufe eines systematischen, auf Anregung von Herrn F. Riesz unternommenen Studiums Geöczescher Arbeiten gelangt bin. Um den Leser nicht durch eine ganz allgemein gehaltene und doch unvollständige Theorie zu ermüden, betrachte ich hauptsächlich die im Titel genannte Flächenklasse, die ja für die Anwendungen wohl in erster Linie in Betracht kommt. Auf die Beziehungen der folgenden Entwicklungen zu Geöcze komme ich weiter unten noch zu sprechen und möchte zunächst den Inhalt der vorliegenden Arbeit schildern.

Nach Lebesgue wird unter einer rektifizierbaren Fläche eine solche Fläche verstanden, die einer Parameterdarstellung

$$(1) \quad x = x(u, v), \quad y = y(u, v), \quad z = z(u, v)$$

fähig ist, wobei $x(u, v)$, $y(u, v)$, $z(u, v)$ einer Lipschitzbedingung genügen. Um die Vorstellungen zu fixieren, nehmen wir an, daß die Gleichungen (1) in einem einfachen Jordanbereiche B betrachtet werden²⁾; die Voraussetzung bedeutet dann, daß für irgend zwei Punkte (u_1, v_1) , (u_2, v_2) von B die Ungleichungen

$$(2) \quad \left. \begin{array}{l} |x(u_2, v_2) - x(u_1, v_1)| \\ |y(u_2, v_2) - y(u_1, v_1)| \\ |z(u_2, v_2) - z(u_1, v_1)| \end{array} \right\} \leq K [(u_2 - u_1)^2 + (v_2 - v_1)^2]^{\frac{1}{2}}$$

¹⁾ Ein Auszug dieser Arbeit wurde am 7. 11. 1927 der ungarischen Akademie der Wissenschaften vorgelegt.

²⁾ Unter einem einfachen Jordanbereiche verstehen wir in üblicher Weise die Menge der innerhalb und auf einer Jordankurve liegenden Punkte.

erfüllt sind, wobei K eine endliche Konstante bedeutet. Es sei B_* das aus den inneren Punkten von B bestehende Gebiet. Alsdann existiert das Doppelintegral

$$(3) \quad \iint_{B_*} \left[\left(\frac{\partial(x, y)}{\partial(u, v)} \right)^2 + \left(\frac{\partial(y, z)}{\partial(u, v)} \right)^2 + \left(\frac{\partial(z, x)}{\partial(u, v)} \right)^2 \right]^{\frac{1}{2}} du dv;$$

aus (2) folgt ja bekanntlich, daß der Integrand eine in B_* fast überall erklärte, meßbare, beschränkte Funktion ist. Wir werden (3) als das klassische Doppelintegral bezeichnen.

Aus dem Umstande allein, daß für eine Parameterdarstellung (1) das klassische Doppelintegral existiert, folgt bekanntlich noch keineswegs, daß es das Flächenmaß der betrachteten Fläche liefert. Obwohl die Ansichten darüber, was unter dem Flächenmaß einer allgemeinen stetigen Fläche zu verstehen sei, stark divergieren, sind alle Forscher darin einig, daß eine jede allgemeine Definition des Flächenmaßes, auf rektifizierbare Flächen angewendet, auf das klassische Doppelintegral führen muß. In der vorliegenden Arbeit wird nun unter anderem der Nachweis erbracht, daß gewisse Definitionen des Flächenmaßes dieser Forderung genügen; den auf andern Gebieten tätigen Mathematiker dürfte aber vor allem die Frage interessieren, inwiefern durch derartige Ergebnisse unsere Kenntnisse über das klassische Doppelintegral gefördert werden, und ich möchte daher einiges hervorheben, was in dieser Hinsicht von Interesse sein dürfte.

Die von uns zu betrachtenden Definitionen werden zunächst zwei wichtige Eigenschaften des Integrals (3) in Evidenz setzen, nämlich die *Invarianz* und die *Halbstetigkeit* desselben. — Neben der durch (1) gegebenen rektifizierbaren Fläche, sie heiße F , betrachte man eine zweite, \bar{F} :

$$x = \bar{x}(\bar{u}, \bar{v}), \quad y = \bar{y}(\bar{u}, \bar{v}), \quad z = \bar{z}(\bar{u}, \bar{v}),$$

wobei der Punkt (\bar{u}, \bar{v}) in einem gewissen einfachen Jordanbereiche \bar{B} variiert, in welchem zu (2) analoge Ungleichungen mit einer gewissen Konstanten \bar{K} erfüllt sind. Die auf F bzw. \bar{F} bezüglichen klassischen Doppelintegrale mögen mit I, \bar{I} bezeichnet werden. Ist eine umkehrbar eindeutige stetige Abbildung von B auf \bar{B} vorhanden, so daß für jedes Paar entsprechender Punkte $A = (u, v), \bar{A} = (\bar{u}, \bar{v})$ die Gleichungen

$$x(A) = \bar{x}(\bar{A}), \quad y(A) = \bar{y}(\bar{A}), \quad z(A) = \bar{z}(\bar{A})$$

erfüllt sind, so wird man die Flächen F und \bar{F} nicht als verschieden betrachten; aus unseren Ergebnissen folgt nun unter anderem, daß unter den angeführten Bedingungen tatsächlich $I = \bar{I}$ gilt. Da über die Abbildung von B auf \bar{B} nur Eineindeutigkeit und Stetigkeit vorausgesetzt wurde, so scheint dies aus der allgemeinen Theorie der Transformation

mehrfacher Integrale nicht ohne weiteres zu folgen³⁾. — Es sei, außer F , eine Folge $\{F_n\}$ rektifizierbarer Flächen gegeben; es seien

$$x = x_n(u, v), \quad y = y_n(u, v), \quad z = z_n(u, v)$$

die Gleichungen von F_n , wobei (u, v) in demselben Jordanbereiche B wie bei (1) variiert, und in B zu (2) analoge Ungleichungen mit gewissen Konstanten K_n erfüllt sind (diese Konstanten K_n brauchen in ihrer Gesamtheit nicht beschränkt zu sein). Das auf F_n bezügliche klassische Doppelintegral heiße I_n , das auf F bezügliche I . Wenn nun in B gleichmäßig

$$x_n(u, v) \rightarrow x(u, v), \quad y_n(u, v) \rightarrow y(u, v), \quad z_n(u, v) \rightarrow z(u, v)$$

erfüllt ist, so braucht zwar noch keineswegs $\lim I_n = I$ zu gelten, wohl aber gilt stets $\underline{\lim} I_n \geq I$; diese Beziehung drückt die *Halbstetigkeit von unten des klassischen Doppelintegrals* aus. Bekanntlich spielt in der Variationsrechnung, seit den grundlegenden Untersuchungen von Herrn Tonelli⁴⁾, die Halbstetigkeit gewisser Integrale eine bedeutende Rolle⁵⁾; doch handelt es sich in der einschlägigen Literatur, soweit Doppelintegrale betrachtet werden, um den Fall, wo die Fläche, deren Koordinaten und Ableitungen im Integrale eingehen, in der Form $z = f(x, y)$ gegeben ist⁶⁾. Für die Behandlung von Doppelintegralen, die sich auf in allgemeiner Parameterdarstellung gegebene Flächen beziehen, dürften die in dieser Arbeit entwickelten, das besondere Integral (3) betreffenden Methoden wenigstens eine erste Orientierung liefern⁷⁾.

³⁾ Für den wichtigen Sonderfall, wo die Gleichungen (1) eine umkehrbar eindeutige Zuordnung zwischen den Punkten (u, v) und den Flächenpunkten (x, y, z) bestimmen, folgt der fragliche Invariansatz aus den schönen Untersuchungen von Herrn Schauder über das Janzense Flächenmaß (Theory of surface measure, Fundamenta Mathematicae 8 (1926), p. 1—48).

⁴⁾ Siehe sein Lehrbuch Fondamenti di calcolo delle variazioni, insbesondere Bd. I, S. 26—27, sowie die Berichte von R. Courant, Über direkte Methoden bei Variations- und Randwertproblemen, Jahresbericht der Deutschen Mathematikervereinigung 34 (1926), insbesondere S. 94, und Über direkte Methoden in der Variationsrechnung und über verwandte Fragen, Math. Annalen 97 (1927), S. 711—736.

⁵⁾ Außer dem unter ⁴⁾ zitierten Lehrbuche von Herrn Tonelli siehe vor allem A. Haar, Über das Plateausche Problem, Math. Annalen 97 (1927), S. 124—158.

⁶⁾ Vgl. loc. cit. ⁴⁾ und ⁵⁾, sowie A. Roussel, Sur l'extremum de certains intégrales doubles, Bulletin des sciences mathématiques 51 (1927), p. 268—288.

⁷⁾ Für den Leser, der aus den geometrischen Entwicklungen dieser Arbeit einen direkten Beweis für die Halbstetigkeit des klassischen Doppelintegrals herauspräparieren will, dürfte eine kleine Note, in welcher ich dies für den Sonderfall $\iint (1+p^2+q^2)^{\frac{1}{2}} dx dy$ durchgeführt habe, die nötigen Fingerzeige enthalten (Bemerkung über das Doppelintegral $\iint (1+p^2+q^2)^{\frac{1}{2}} dx dy$, Math. Zeitschr. 26 (1927), S. 408—416).

Die beiden angeführten Sätze über das klassische Doppelintegral (3) fließen aus den geometrischen Deutungen, die wir für dasselbe gewinnen werden. Diesen geometrischen Deutungen liegt die Idee zugrunde, daß trotz des bekannten Schwarzschen Beispiels⁹⁾ der Prozeß der Bestimmung des Flächenmaßes letzten Endes auf den elementargeometrischen Formeln für das Flächenmaß von Polyederflächen beruhen muß. Und zwar werden wir die folgenden drei Prozesse untersuchen (die genauen Begriffsbestimmungen folgen später, hier handelt es sich nur um die anschaulichen Momente).

a) Die Fläche, sie heiße F , wird in endlich viele Flächenstücke F_1, F_2, \dots, F_ν zerlegt, und jedes solche Stück wird auf die drei Koordinatenebenen xy, yz, zx orthogonal projiziert. Sind a_k, b_k, c_k die Flächeninhalte der drei Projektionen von F_k , so betrachtet man die Summe

$$\sum_{k=1}^{\nu} (a_k^2 + b_k^2 + c_k^2)^{\frac{1}{2}}$$

in den meisten Theorien des Flächenmaßes als einen *unteren Näherungswert* für das Flächenmaß von F , indem nämlich in solchen Theorien das Flächenmaß gerade als die obere Grenze dieser Summe (für alle mögliche Zerlegungen von F) definiert wird. Diese obere Grenze heiße $G(F)$.

b) Die Fläche F wird wieder in endlich viele Stücke F_1, F_2, \dots, F_ν zerlegt. Man projiziere F_1 orthogonal auf eine Ebene ε_1 und bezeichne mit φ_1 den Flächeninhalt dieser Projektion, ferner mit λ_1 die obere Grenze von φ_1 für alle Lagen der Ebene ε_1 . Analoge Bedeutung mögen $\lambda_2, \dots, \lambda_\nu$ für F_2, \dots, F_ν haben. Nach Peano hat man dann die Summe $\sum_{k=1}^{\nu} \lambda_k$ als einen *unteren Näherungswert* für das Flächenmaß von F zu betrachten; Peano erklärt das Flächenmaß von F gerade als die obere Grenze dieser Summe für alle möglichen Zerlegungen von F . Diese obere Grenze heiße $P(F)$.

c) Während die beiden bisher geschilderten Prozesse, kurz gesagt, mit *unteren Schranken für das Flächenmaß* operieren, beruht der von Lebesgue eingeführte Prozeß auf *oberen Schranken*. Lebesgue geht von der in elementaren Fällen als richtig erwiesenen Annahme aus, daß das Flächenmaß ein von unten halbstetiges Funktional sei. Nimmt man insbesondere eine gegen F konvergierende Folge von Polyederflächen Π_n , und bezeichnet $E(\Pi_n)$ das im elementargeometrischen Sinne verstandene Flächenmaß von Π_n , so bedeutet die Halbstetigkeit von unten folgendes:

⁹⁾ Vgl. etwa die besonders elegante und elementare Darstellung bei M. Fréchet, Sur l'aire des surfaces polyédrales, Annales de la soc. polonaise de math. 3 (1925), p. 1—3.

es braucht zwar $\lim E(\Pi_n)$ nicht vorhanden, und wenn vorhanden, nicht gleich dem Flächenmaße der Grenzfläche F zu sein — gewiß ist aber $\underline{\lim} E(\Pi_n)$ nicht kleiner als das Flächenmaß von F . Jede gegen F konvergierende Polyederfolge $\{\Pi_n\}$ liefert daher eine obere Schranke (nämlich $\underline{\lim} E(\Pi_n)$) für das Flächenmaß von F . Lebesgue betrachtet gerade die untere Grenze dieser oberen Schranken als das Flächenmaß von F ⁹⁾. Diese untere Grenze heiße $L(F)$.

Die Fläche F sei nun insbesondere rektifizierbar; das klassische Doppelintegral (3) werde der Kürze halber mit $I(F)$ bezeichnet. Im Laufe seiner Untersuchungen gelangte Geöcze insbesondere zum Ergebnis, daß in diesem Falle $G(F) = P(F) = L(F) = I(F)$ gilt¹⁰⁾. Außer den beiden oben angeführten Sätzen folgt hieraus für das klassische Doppelintegral eine geometrische Deutung, auf die ich aufmerksam machen möchte. Faßt man die bei a) und b) verwendeten Größen $\sum (a_k^2 + b_k^2 + c_k^2)^{\frac{1}{2}}$ und $\sum \lambda_k$ zu einem System $\{s\}$, die bei c) verwendeten Größen $\underline{\lim} E(\Pi_n)$ zu einem System $\{S\}$ zusammen, so besteht also $\{s\}$ aus *anschaulichen unteren Schranken*, und das System $\{S\}$ aus *anschaulichen oberen Schranken* für das Flächenmaß. Die Sätze $G(F) = P(F) = L(F) = I(F)$ besagen demnach, daß im Falle rektifizierbarer Flächen zwischen den genannten unteren und oberen Schranken eben nur der durch das klassische Doppelintegral gelieferte Zahlenwert Platz hat; in diesem Falle ergibt sich also das Flächenmaß sozusagen aus einem, auf Grund der geometrischen Anschauung konstruierten, Dedekindschen Schnitte.

Teil II dieser Arbeit enthält eine neue Ableitung der angeführten Sätze. Damit dieselben einen klaren Sinn erhalten, müssen natürlich die in den Prozessen a), b), c) auftretenden Begriffe genau festgelegt werden; bereits in diesem Punkte weicht die vorliegende Darstellung erheblich von Geöcze ab. Die Prozesse a) und b) operieren mit orthogonalen Projektionen. Der einfachste Ansatz wäre, wenn F eine Fläche und \mathfrak{P} ihre Projektion auf einer Ebene bedeutet, das Maß $m\mathfrak{P}$ der Punktmenge \mathfrak{P} als Baustein bei den Prozessen a) und b) zu verwenden; da wir aber in allgemeiner Parameterdarstellung gegebene allgemeine stetige Flächen untersuchen, so erweist sich dieser Ansatz als unbrauchbar. Als *Fundamentalgröße der Theorie* ist eben diejenige Größe zu betrachten, durch welche das unbrauchbare $m\mathfrak{P}$ ersetzt wird, indem es in erster Linie von den Eigenschaften derselben abhängt, ob man über bloße Definitionen hinauskommen wird.

⁹⁾ H. Lebesgue, Intégrale, longueur, aire, Annali di Matematica (3) 7 (1902), p. 231—359.

¹⁰⁾ Z. v. Geöcze: a) Über die rektifizierbare Fläche (ungarisch), Math. és term. tud. értesítő 34 (1916), p. 337—354; b) Über die Peanosche Definition des Flächenmaßes (ungarisch), ibid. 35 (1917), p. 325—360.

Die Erklärung der Geöczeschen Fundamentalgröße beruht auf bewundernswerter geometrischer Intuition, doch ist bereits ihre bloße Definition mit topologischen Schwierigkeiten verbunden¹¹⁾; um diesen und weiteren Komplikationen aus dem Wege zu gehen, führe ich eine neue Fundamentalgröße ein. Ich erkläre zunächst in einfacher Weise eine Teilmenge der Projektion \mathfrak{P} , den *Projektionskern*, und wähle das Lebesguesche Maß des Projektionskernes zur Fundamentalgröße; dies bedeutet das Zurückgehen auf die geometrische Grundidee, welche den schwierigen Geöczeschen Begriffsbildungen implizite zugrunde liegt: daß nämlich die vollständige Projektion \mathfrak{P} deswegen unbrauchbar ist, weil dieselbe unwesentliche und doch störende Bestandteile enthält. *Der Projektionskern soll eben den wesentlichen Teil der Projektion vertreten.* Außer der Anschaulichkeit ihrer Erklärung gewährt die neue Fundamentalgröße den technischen Vorteil, daß ihre Eigenschaften unmittelbar aus der allgemeinen Theorie des Lebesgueschen Maßes ebener Punktengen fließen — eben weil dieselbe selbst das Maß einer solchen Punktmenge darstellt.

Durch die Änderung der Fundamentalgröße erhalten die Prozesse a) und b), sowie die erwähnten Geöczeschen Sätze natürlich eine andere Bedeutung; die oben angeführten Folgerungen für das klassische Doppelintegral bleiben aber bestehen. Bei der Ableitung der fraglichen Sätze dienten mir hauptsächlich die schönen Untersuchungen von Herrn Rademacher über rektifizierbare Flächen und über die Transformation von Doppelintegralen als Vorbild¹²⁾; auch hierdurch ergeben sich gegenüber Geöcze wesentliche Vereinfachungen. Ich bediene mich ferner durchgängig der Fréchet'schen Begriffsbildungen über stetige Flächen¹³⁾; diese Begriffsbildungen, deren Zweckmäßigkeit man bewundern muß, scheinen überhaupt unentbehrlich für die strenge und übersichtliche Durchführung von Untersuchungen zu sein, die sich auf in allgemeiner Parameterdarstellung vorliegende Flächen beziehen, wobei also auf Multiplizität und Anordnung der Flächenpunkte Rücksicht zu nehmen ist.

Die vorliegende Untersuchung bewegt sich durchaus im Lebesgueschen Ideenkreise, wonach als wesentlichste Eigenschaft des Flächenmaßes die

¹¹⁾ Vgl. insbesondere seine Arbeit: *Recherches générales sur la quadrature des surfaces courbes*, Math. und naturwiss. Berichte aus Ungarn 27 (1909), S. 1—21 und S. 131—163.

¹²⁾ H. Rademacher, *Eineindeutige Abbildungen und Meßbarkeit*, Dissertation, Göttingen (1917); *Über partielle und totale Differenzierbarkeit I und II*, Math. Annalen 79 (1919), S. 340—359 und 81 (1920), S. 52—63. — Ich verdanke Herrn F. Riesz die Anregung, die Rademacherschen Arbeiten zwecks Vereinfachung der Geöczeschen Theorie heranzuziehen.

¹³⁾ M. Fréchet, *Sur la distance de deux surfaces*, Annales de la soc. polonaise de math. 3 (1924), p. 4—19.

Halbstetigkeit von unten zu betrachten ist¹⁴⁾. Aus diesem Grunde kommen aus der reichen Literatur über das Flächenmaß, sofern es sich um eine Vergleichung der Methoden und Ergebnisse handelt, unmittelbar nur diejenigen Untersuchungen in Betracht, die sich auf das Lebesguesche Flächenmaß beziehen. Die hierher gehörigen Arbeiten von Geöcze und von Herrn Tonelli, sowie die anschließenden Arbeiten vom Verfasser und von Herrn Saks¹⁵⁾ haben für den wichtigen Sonderfall von Flächen, die durch eine Gleichung $z = f(x, y)$ mit eindeutigem $f(x, y)$ dargestellt werden können, Resultate von überraschender Schärfe und Allgemeinheit zutage gefördert; doch scheinen gerade die zugkräftigsten Methoden wesentlich an die besondere Gleichungsform $z = f(x, y)$ gebunden zu sein. Die Verallgemeinerung der Ergebnisse auf den in dieser Arbeit behandelten Fall von Flächen, die in allgemeiner Parameterdarstellung vorliegen, dürfte noch zu schwierigen Problemen führen; da die bisher erzielten Resultate wesentlich auf der Betrachtung der ebenen Projektionen von Flächenstücken beruhen, so habe ich, mit Rücksicht auf die vielen noch offenen Fragen, die diesbezüglichen Entwicklungen in Teil I etwas ausführlicher und allgemeiner gestaltet, als es für die Untersuchung der rektifizierbaren Flächen unbedingt nötig gewesen wäre.

I. Teil.

Projektionssätze.

§ 1.

Allgemeines über stetige Flächen.

1. Sind $x(u, v)$, $y(u, v)$, $z(u, v)$ in einem einfachen Jordanbereiche B eindeutig und stetig, so stellen die Gleichungen

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

wobei x, y, z rechtwinklige Koordinaten bedeuten, eine stetige Fläche F dar; wenn es sich darum handelt, die Bestimmungsstücke von F vor Augen zu haben, so schreiben wir

$$(4) \quad F: x = x(u, v), \quad y = y(u, v), \quad z = z(u, v), \quad (u, v) \in B.$$

¹⁴⁾ Vgl. M. Fréchet, Sur le prolongement des fonctionnelles semicontinues et sur l'aire des surfaces courbes, *Fundamenta Mathematicae* 7 (1925), p. 210—224. — Einer der leitenden Gesichtspunkte bei der Erklärung unserer Fundamentalgröße bestand eben darin, den durch die Prozesse a) und b) erklärten Größen $G(F)$ und $P(F)$ die Eigenschaft der Halbstetigkeit aufzuprägen.

¹⁵⁾ Wegen einer zusammenfassenden Darstellung der Ergebnisse vgl. die beiden Arbeiten (wo auch die Originalarbeiten angeführt sind): T. Radó, Sur l'aire des surfaces courbes, *Acta litt. ac sc. regiae univ. hung. Franc.-Jos. Szeged* 3 (1927), p. 131—169 und S. Saks, Sur l'aire des surfaces $z = f(x, y)$, *ibid.*, p. 170—176.

Wir setzen nicht voraus, daß zwei verschiedenen Punkten (u, v) zwei verschiedene Punkte (x, y, z) entsprechen. Es ist nun vor allem festzusetzen, unter welchen Umständen die durch zwei solche Gleichungssysteme gegebenen Flächen als identisch gelten sollen. Wir legen dabei die äußerst zweckmäßigen Begriffsbildungen von Herrn Fréchet zugrunde, an welche hier kurz erinnert werden möge¹⁶⁾.

2. Den Ausgangspunkt bildet der Begriff der *Distanz von zwei stetigen Flächen*. Neben der Fläche (4) betrachte man eine zweite

$$(5) \quad \bar{F}: x = \bar{x}(\bar{u}, \bar{v}), \quad y = \bar{y}(\bar{u}, \bar{v}), \quad z = \bar{z}(\bar{u}, \bar{v}), \quad (\bar{u}, \bar{v}) \in \bar{B}.$$

Zwei einfache Jordanbereiche können nun auf unendlich viele Arten topologisch aufeinander abgebildet werden. Es sei T eine topologische Abbildung von B auf \bar{B} und es sei, wenn A ein variabler Punkt von B und \bar{A} der Bildpunkt von A ist, M_T das Maximum von

$$[(x(A) - \bar{x}(\bar{A}))^2 + (y(A) - \bar{y}(\bar{A}))^2 + (z(A) - \bar{z}(\bar{A}))^2]^{\frac{1}{2}}.$$

Die Distanz der beiden Flächen F und \bar{F} wird als die untere Grenze von M_T , bei allen möglichen Abbildungen T , erklärt und soll durch $[F, \bar{F}]$ bezeichnet werden. Man hat

$$[F, \bar{F}] = [\bar{F}, F] \geq 0, \quad [F, F] = 0;$$

ferner gilt, unter F_1, F_2, F_3 irgend drei stetige Flächen verstanden, die Dreiecksungleichung

$$[F_1, F_3] \leq [F_1, F_2] + [F_2, F_3].$$

3. Zwei Flächen F, \bar{F} sollen nun als identisch gelten, wenn ihre Distanz $[F, \bar{F}]$ gleich Null ist. In diesem Falle gelten demnach (4) und (5) als Parameterdarstellungen derselben Fläche. Ein wichtiger Sonderfall ist der, wo eine topologische Abbildung von B auf \bar{B} von der Beschaffenheit vorhanden ist, daß für jedes Paar A, \bar{A} entsprechender Punkte die Gleichungen

$$x(A) = \bar{x}(\bar{A}), \quad y(A) = \bar{y}(\bar{A}), \quad z(A) = \bar{z}(\bar{A})$$

erfüllt sind; in diesem Falle sollen (4) und (5) als *ähnliche Parameterdarstellungen* derselben stetigen Fläche bezeichnet werden. Doch brauchen, wie Herr Fréchet an einem einfachen Beispiele zeigt, zwei Parameterdarstellungen, die zwei im angegebenen Sinne identische Flächen bestimmen, keineswegs ähnlich zu sein¹⁷⁾.

¹⁶⁾ Siehe M. Fréchet, loc. cit. ¹²⁾.

¹⁷⁾ Der Fréchetsche Begriff der Identität von zwei stetigen Flächen wurde von Herrn B. v. Kerékjártó einer tiefgehenden topologischen Analyse unterworfen (Involutions et surfaces continues, Acta litt. ac sc. regiae univ. Franc.-Jos. Szeged **3** (1927). p. 49—67).

4. Wir sagen, eine Folge $\{F_n\}$ stetiger Flächen konvergiert gegen die stetige Fläche F , wenn $[F, F_n] \rightarrow 0$ gilt. Herr Fréchet zeigt, daß für diesen Konvergenzbegriff die von der Theorie der Punktmengen her geläufigen Sätze bestehen bleiben.

Man sieht sogleich: wenn $F_n \rightarrow F$ gilt und F durch (4) gegeben ist, so gestattet F_n eine Darstellung

$$(6) \quad F_n: x = x_n(u, v), y = y_n(u, v), z = z_n(u, v), (u, v) \in B,$$

(wo also für jedes n derselbe Bereich B auftritt), so daß in B gleichmäßig

$$x_n(u, v) \rightarrow x(u, v), \quad y_n(u, v) \rightarrow y(u, v), \quad z_n(u, v) \rightarrow z(u, v)$$

erfüllt ist. Mit Rücksicht auf spätere Anwendung fügen wir noch die evidente Bemerkung hinzu, daß man überdies für jedes n eine besondere Darstellung von F_n vorgeben und verlangen kann, daß (6) dieser besonderen Darstellung ähnlich sei.

5. Eine stetige Fläche F heiße *rektifizierbar*, wenn sich unter ihren Parameterdarstellungen (4) eine solche befindet, bei welcher die Funktionen $x(u, v)$, $y(u, v)$, $z(u, v)$ in B einer Lipschitzbedingung genügen, also in B Ungleichungen der Form (2) erfüllen; jede solche Darstellung einer rektifizierbaren Fläche heiße eine *typische Darstellung* derselben.

6. In der Folge werden wir auf Schritt und Tritt den Nachweis führen müssen, daß eine gewisse, durch einen bestimmten Prozeß der Fläche zugeordnete Größe nur von der Fläche selbst abhängt und nicht etwa von der besonderen Parameterdarstellung, aus welcher die betreffende Größe konstruiert wurde. Wir wollen daher das äußerst einfache Prinzip ausdrücklich formulieren, auf welchem die in Betracht kommenden Invarianzsätze letzten Endes beruhen werden. Dieses Prinzip ergibt sich aus den beiden folgenden evidenten Bemerkungen.

a) Zwei stetige Flächen F_1, F_2 , für welche $[F_1, F_2] = 0$ gilt, bestehen aus denselben Punkten (x, y, z) ; mit anderen Worten: zwei im Sinne von Fréchet identische Flächen sind auch als Punktmengen identisch (die Umkehrung gilt natürlich nicht).

b) Sind zwei Paare F'_1, F''_1 und F'_2, F''_2 von stetigen Flächen gegeben, für welche $[F'_1, F''_1] = 0$, $[F'_2, F''_2] = 0$ gilt, so gilt auch $[F'_1, F'_2] = [F''_1, F''_2]$; mit anderen Worten: die Distanz zweier stetiger Flächen ist von der besonderen Parameterdarstellung derselben unabhängig.

Man hat demnach das folgende *Invarianzprinzip*: wird einer stetigen Fläche eine gewisse Größe durch einen Prozeß zugeordnet, welcher durch alleinige Anwendung der Begriffe Distanz zweier Flächen und Menge der Punkte einer Fläche erklärt werden kann, so hängt die fragliche Größe nur von der Fläche selbst ab.

7. Wir haben noch die Transformationen des rechtwinkligen Koordinatensystems xyz zu betrachten. Stellen die Formeln

$$\begin{aligned}x' &= a_1 x + b_1 y + c_1 z + d_1, \\y' &= a_2 x + b_2 y + c_2 z + d_2, \\z' &= a_3 x + b_3 y + c_3 z + d_3\end{aligned}$$

den Übergang vom System xyz zum neuen, ebenfalls rechtwinkligen, System $x'y'z'$ dar, so setzen wir fest, daß die Gleichungen

$$\left. \begin{aligned}x' &= a_1 x(u, v) + b_1 y(u, v) + c_1 z(u, v) + d_1 \\y' &= a_2 x(u, v) + b_2 y(u, v) + c_2 z(u, v) + d_2 \\z' &= a_3 x(u, v) + b_3 y(u, v) + c_3 z(u, v) + d_3\end{aligned} \right\} (u, v) \in B$$

dieselbe Fläche wie die Gleichungen (4) darstellen. Auf Grund dieser Festsetzung läßt sich das obige Invarianzprinzip offenbar dahin erweitern, daß jeder Prozeß, bei welchem die Koordinatenebenen xy, yz, zx keine ausgezeichnete Rolle spielen und in dessen Erklärung nur die Begriffe Distanz zweier Flächen und Menge der Punkte einer Fläche auftreten, zu solchen Größen führt, welche von der Wahl der Parameterdarstellung und auch von der Wahl des Koordinatensystems xyz unabhängig sind.

§ 2.

Der Projektionskern.

Es liege die stetige Fläche (4) vor. Die Punkte (x, y, z) , welche den Punkten (u, v) von B vermöge (4) entsprechen, bilden im Raume eine beschränkte abgeschlossene Punktmenge E_F . Ist ε eine Ebene, so bezeichne $\mathfrak{P}(F; \varepsilon)$ die orthogonale Projektion von E_F auf der Ebene ε . Die Punktmenge $\mathfrak{P}(F; \varepsilon)$ ist ebenfalls beschränkt und abgeschlossen, und folglich auch meßbar.

Der Kern $\mathfrak{K}(F; \varepsilon)$ der Projektion $\mathfrak{P}(F; \varepsilon)$ wird nun durch folgende Festsetzung erklärt. Ein Punkt P der Ebene ε gehört zu $\mathfrak{K}(F; \varepsilon)$, wenn es eine positive Zahl $\delta = \delta(P) > 0$ gibt derart, daß für jede stetige Fläche F^* mit $[F, F^*] < \delta$ der Punkt P in $\mathfrak{P}(F^*; \varepsilon)$ enthalten ist. Der Kern $\mathfrak{K}(F; \varepsilon)$ kann auch leer sein. Offenbar gilt $\mathfrak{K}(F; \varepsilon) \subset \mathfrak{P}(F; \varepsilon)$.

2. $\mathfrak{K}(F; \varepsilon)$ ist meßbar¹⁸⁾.

Beweis. Es bedeute \mathfrak{K}_n die Menge derjenigen Punkte P der Ebene ε , welche die folgende Eigenschaft besitzen: sobald für eine stetige Fläche F^*

¹⁸⁾ Den folgenden Beweis entnehme ich einer brieflichen Mitteilung von Herrn S. Saks.

die Beziehung $[F, F^*] < \frac{1}{n}$ gilt, liegt P in $\mathfrak{P}(F^*; \varepsilon)$. Offenbar ist

$$\mathfrak{R}(F; \varepsilon) = \sum_{n=1}^{\infty} \mathfrak{R}_n,$$

so daß es genügt, die Meßbarkeit von \mathfrak{R}_n darzutun. Man hat nun, wie unmittelbar ersichtlich

$$\mathfrak{R}_n = \coprod_{[F, F^*] < \frac{1}{n}} \mathfrak{P}(F^*; \varepsilon);^{19)}$$

da jede einzelne Menge $\mathfrak{P}(F^*; \varepsilon)$ abgeschlossen ist, so ist auch ihr Durchschnitt \mathfrak{R}_n abgeschlossen und folglich meßbar.

Das Lebesguesche (zweidimensionale) Maß des Kernes $\mathfrak{R}(F; \varepsilon)$ ist unsere Fundamentalgröße, durch welche wir $m\mathfrak{P}(F; \varepsilon)$ ersetzen. Durch Heranziehung des Invarianzprinzips in I, 1, 7 (d. i. Teil I, § 1, Nr. 7) ergibt sich sogleich, daß $m\mathfrak{R}(F; \varepsilon)$ (und auch $m\mathfrak{P}(F; \varepsilon)$) von der besonderen Parameterdarstellung der Fläche und von der Wahl des Koordinatensystems xyz unabhängig ist.

3. Ist P ein Punkt von $\mathfrak{R}(F; \varepsilon)$, so gibt es nach Definition ein $\delta = \delta(P) > 0$ derart, daß

$$P \in \mathfrak{P}(F^*; \varepsilon) \text{ für } [F, F^*] < \delta$$

gilt. Man sieht nun leicht, daß mit eben diesem δ sogar

$$(7) \quad P \in \mathfrak{R}(F^*; \varepsilon) \text{ für } [F, F^*] < \delta$$

erfüllt ist. (7) bedeutet nämlich: Ist F^* eine feste, der Bedingung $[F, F^*] < \delta$ genügende Fläche, so gibt es ein $\delta^* > 0$ derart, daß

$$P \in \mathfrak{P}(\bar{F}; \varepsilon) \text{ für } [F^*, \bar{F}] < \delta^*$$

gilt. Nun kann, wie sogleich ersichtlich,

$$\delta^* = \delta - [F, F^*]$$

gesetzt werden. Aus $[F^*, \bar{F}] < \delta^*$ folgt dann nämlich zunächst

$$[F, \bar{F}] \leq [F, F^*] + [F^*, \bar{F}] < [F, F^*] + \delta^* = \delta,$$

und daraus, mit Rücksicht auf die Bedeutung von δ , die Beziehung $P \in \mathfrak{P}(\bar{F}; \varepsilon)$.

4. Die Fundamentalgröße $m\mathfrak{R}(F; \varepsilon)$ ist, als Funktion der Fläche F , halbstetig von unten, d. h. es gilt $\liminf m\mathfrak{R}(F_n; \varepsilon) \geq m\mathfrak{R}(F; \varepsilon)$ für $F_n \rightarrow F$.

Bedeutet nämlich \mathfrak{M}_n den Durchschnitt von $\mathfrak{R}(F; \varepsilon)$ und $\mathfrak{R}(F_n; \varepsilon)$, so ist \mathfrak{M}_n , als Durchschnitt zweier meßbarer Mengen, wieder meßbar. Ist P irgendein Punkt von $\mathfrak{R}(F; \varepsilon)$, so gilt nach Nr. 3, wegen $[F, F_n] \rightarrow 0$,

¹⁹⁾ Das Zeichen \coprod bedeutet üblicherweise die Menge der gemeinsamen Punkte (Durchschnitt).

für hinreichend großes n die Beziehung $P \subset \mathfrak{R}(F_n; \varepsilon)$, also auch $P \subset \mathfrak{M}_n$. Nach den ersten Sätzen der Lebesgueschen Maßtheorie folgt hieraus

$$\lim m \mathfrak{M}_n = m \mathfrak{R}(F; \varepsilon).$$

Wegen $\mathfrak{M}_n \subset \mathfrak{R}(F_n; \varepsilon)$ folgt nunmehr

$$\underline{\lim} m \mathfrak{R}(F_n; \varepsilon) \geq \lim m \mathfrak{M}_n = m \mathfrak{R}(F; \varepsilon),$$

w. z. b. w.

5. Die stetige Fläche F sei durch (4) gegeben, und es sei B^* ein in B enthaltener einfacher Jordanbereich. Für die durch

$$F^*: x = x(u, v), \quad y = y(u, v), \quad z = z(u, v), \quad (u, v) \in B^*$$

gegebene stetige Fläche und für irgendeine Ebene ε gilt dann

$$\mathfrak{R}(F^*; \varepsilon) \subset \mathfrak{R}(F; \varepsilon).$$

Beweis. Zu zeigen ist: aus $P \subset \mathfrak{R}(F^*; \varepsilon)$ folgt $P \subset \mathfrak{R}(F; \varepsilon)$. Sei also P ein Punkt von $\mathfrak{R}(F^*; \varepsilon)$; dann gibt es nach Definition ein $\delta^* = \delta^*(P) > 0$ derart, daß

$$(7^*) \quad P \subset \mathfrak{P}(\bar{F}; \varepsilon) \quad \text{für} \quad [F^*, \bar{F}] < \delta^*$$

gilt. Es sei nun F_* irgendeine der Bedingung $[F, F_*] < \delta^*$ genügende Fläche. Aus der Erklärung der Distanz folgt dann, daß F_* eine Darstellung

$$F_*: x = x_*(u, v), \quad y = y_*(u, v), \quad z = z_*(u, v), \quad (u, v) \in B$$

gestattet, wobei für jeden Punkt $A = (u, v)$ von B die Ungleichung

$$(8) \quad [(x(A) - x_*(A))^2 + (y(A) - y_*(A))^2 + (z(A) - z_*(A))^2]^{\frac{1}{2}} < \delta^*$$

statthalt. Für die durch

$$\bar{F}: x = x_*(u, v), \quad y = y_*(u, v), \quad z = z_*(u, v), \quad (u, v) \in B^*$$

gegebene Fläche gilt dann erstens, infolge der Voraussetzung $B^* \subset B$,

$$\mathfrak{P}(\bar{F}; \varepsilon) \subset \mathfrak{P}(F_*; \varepsilon);$$

zweitens, infolge der Ungleichung (8)

$$[\bar{F}, F^*] < \delta^*;$$

und folglich drittens, mit Rücksicht auf (7*),

$$P \subset \mathfrak{P}(\bar{F}; \varepsilon) \subset \mathfrak{P}(F_*; \varepsilon).$$

Damit ist also

$$P \subset \mathfrak{P}(F_*; \varepsilon) \quad \text{für} \quad [F, F_*] < \delta^*$$

erwiesen; mit Rücksicht auf die Erklärung des Projektionskernes stellt dies gerade die zu beweisende Behauptung $P \subset \mathfrak{R}(F; \varepsilon)$ dar.

6. Die stetige Fläche F sei wieder durch (4) gegeben. Es sei $\{B^{(n)}\}$ eine Folge einfacher Jordanbereiche, der folgenden Beschaffenheit.

a) $B^{(1)} \subset B^{(2)} \subset \dots \subset B$.

b) B wird durch die Bereiche $B^{(n)}$ in dem Sinne ausgefüllt, daß jede aus inneren Punkten von B bestehende abgeschlossene Punktmenge für hinreichend großes n in $B^{(n)}$ enthalten ist.

Es sei $F^{(n)}$ die durch

$$F^{(n)}: x = x(u, v), \quad y = y(u, v), \quad z = z(u, v), \quad (u, v) \in B^{(n)}$$

erklärte stetige Fläche. Alsdann gilt

$$m \mathfrak{R}(F^{(n)}; \varepsilon) \rightarrow m \mathfrak{R}(F; \varepsilon).$$

Beweis. Nach Nr. 5 gilt, wegen $B^{(n)} \subset B$, jedenfalls

$$m \mathfrak{R}(F^{(n)}; \varepsilon) \leq m \mathfrak{R}(F; \varepsilon),$$

so daß es genügt,

$$(9) \quad \underline{\lim} m \mathfrak{R}(F^{(n)}; \varepsilon) \geq m \mathfrak{R}(F; \varepsilon)$$

zu zeigen²⁰⁾. Um den Beweis von (9) übersichtlicher zu gestalten, bemerken wir vor allem, daß alles, was in (9) vorkommt, von der besonderen Parameterdarstellung der Flächen $F, F^{(n)}$ unabhängig ist. Wir dürfen und wollen daher durch eine topologische Abbildung von B neue Parameter u, v einführen, und zwar derart, daß B in die Kreisscheibe $u^2 + v^2 \leq 1$ übergeht; um keine neuen Bezeichnungen einführen zu müssen, nehmen wir von vornherein

$$B \equiv u^2 + v^2 \leq 1$$

an. Es sei $\lambda > 1$, und F_λ die durch

$$F_\lambda: x = x(u, v), \quad y = y(u, v), \quad z = z(u, v), \quad u^2 + v^2 \leq \frac{1}{\lambda^2}$$

gegebene stetige Fläche. Dann gilt offenbar $F_\lambda \rightarrow F$ für $\lambda \rightarrow 1$; daraus folgt zunächst, nach Nr. 4,

$$(10) \quad \underline{\lim}_{\lambda \rightarrow 1} m \mathfrak{R}(F_\lambda; \varepsilon) \geq m \mathfrak{R}(F; \varepsilon).$$

Da wegen $\lambda > 1$ die Kreisfläche $u^2 + v^2 \leq \frac{1}{\lambda^2}$ aus inneren Punkten von B besteht, so wird diese Kreisfläche infolge der Eigenschaft b) der Folge $\{B^{(n)}\}$ für hinreichend großes n in $B^{(n)}$ enthalten sein; nach Nr. 5 hat man daher für hinreichend großes (von λ abhängiges) n die Ungleichung

$$m \mathfrak{R}(F^{(n)}; \varepsilon) \geq m \mathfrak{R}(F_\lambda; \varepsilon),$$

²⁰⁾ Man beachte, daß aus den Voraussetzungen noch keineswegs $F^{(n)} \rightarrow F$ folgt, so daß das Ergebnis von Nr. 4 nicht unmittelbar angewendet werden darf.

woraus

$$(11) \quad \lim_{n \rightarrow \infty} m \mathfrak{R}(F^{(n)}; \varepsilon) \geq \lim_{\lambda \rightarrow 1} m \mathfrak{R}(F_\lambda; \varepsilon)$$

folgt. Durch (10) und (11) ist (9) bewiesen.

7. Die stetige Fläche F sei durch (4) gegeben. Der Bereich B werde durch einen einfachen Jordanbogen γ , der zwei verschiedene Randpunkte von B verbindet und sonst aus inneren Punkten von B besteht, in zwei einfache Jordanbereiche B_1, B_2 zerlegt. Wenn in den Gleichungen

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

die Veränderlichkeit von (u, v) der Reihe nach auf B_1, B_2, γ beschränkt wird, so erhält man zwei stetige Flächen F_1, F_2 und einen stetigen Bogen Γ . Es sei Γ_ε die Projektion von Γ auf eine Ebene ε . Alsdann gilt

$$(12) \quad \mathfrak{R}(F; \varepsilon) < \mathfrak{R}(F_1; \varepsilon) + \mathfrak{R}(F_2; \varepsilon) + \Gamma_\varepsilon,$$

und folglich auch

$$(13) \quad m \mathfrak{R}(F; \varepsilon) \leq m \mathfrak{R}(F_1; \varepsilon) + m \mathfrak{R}(F_2; \varepsilon) + m \Gamma_\varepsilon.$$

Beweis. Auch jetzt dürfen wir, zwecks Vereinfachung der Beweisführung, die aus B, B_1, B_2, γ bestehende Figur mit irgendeiner anderen, aus der gegebenen durch topologische Abbildung von B erhältlichen, bequemeren Figur zusammenfallen lassen. Wir dürfen und wollen daher von vornherein annehmen, daß

$$B \equiv 0 \leq u \leq 1, \quad 0 \leq v \leq 1, \quad \gamma \equiv u = \frac{1}{2}, \quad 0 \leq v \leq 1,$$

$$B_1 \equiv 0 \leq u \leq \frac{1}{2}, \quad 0 \leq v \leq 1, \quad B_2 \equiv \frac{1}{2} \leq u \leq 1, \quad 0 \leq v \leq 1$$

ist. (12) ist bewiesen, sobald gezeigt wird: ist P ein Punkt von $\mathfrak{R}(F; \varepsilon)$, der weder in $\mathfrak{R}(F_1; \varepsilon)$ noch in $\mathfrak{R}(F_2; \varepsilon)$ enthalten ist, so ist P ein Punkt von Γ_ε .

Es sei also P ein solcher Punkt. Da P nicht in $\mathfrak{R}(F_1; \varepsilon)$ enthalten ist, so gibt es, auf Grund der Erklärung des Projektionskernes, eine gegen F_1 konvergierende Folge $\{F_n^{(1)}\}$ von stetigen Flächen, so daß P für keinen Wert von n in $\mathfrak{R}(F_n^{(1)}; \varepsilon)$ enthalten ist. Da F_1 durch

$$F_1: x = x(u, v), \quad y = y(u, v), \quad z = z(u, v), \quad (u, v) \in B_1$$

gegeben ist, so gestattet $F_n^{(1)}$ eine Darstellung

$$F_n^{(1)}: x = x_n^{(1)}(u, v), \quad y = y_n^{(1)}(u, v), \quad z = z_n^{(1)}(u, v), \quad (u, v) \in B_1,$$

wobei gleichmäßig in B_1

$$x_n^{(1)}(u, v) \rightarrow x(u, v), \quad y_n^{(1)}(u, v) \rightarrow y(u, v), \quad z_n^{(1)}(u, v) \rightarrow z(u, v)$$

gilt. Da P auch kein Punkt von $\mathfrak{R}(F_2; \varepsilon)$ ist, so gibt es weiter zu

$$F_2: x = x(u, v), \quad y = y(u, v), \quad z = z(u, v), \quad (u, v) \in B_2$$

eine Flächenfolge

$$F_n^{(2)}: x = x_n^{(2)}(u, v), \quad y = y_n^{(2)}(u, v), \quad z = z_n^{(2)}(u, v), \quad (u, v) \in B_2,$$

wobei gleichmäßig in B_2

$$x_n^{(2)}(u, v) \rightarrow x(u, v), \quad y_n^{(2)}(u, v) \rightarrow y(u, v), \quad z_n^{(2)}(u, v) \rightarrow z(u, v)$$

gilt, und P für keinen Wert von n in $\mathfrak{R}(F_n^{(2)}; \varepsilon)$ enthalten ist. Wir führen nunmehr eine stetige Hilfsfläche

$$F_n: x = x_n(u, v), \quad y = y_n(u, v), \quad z = z_n(u, v), \quad (u, v) \in B$$

durch folgende Festsetzung ein. Es sei

$$x_n(u, v) = \begin{cases} x_n^{(1)}(u, v) & \text{für } 0 \leq u \leq \frac{1}{2} - \frac{1}{n}, \quad 0 \leq v \leq 1, \\ x_n^{(2)}(u, v) & \text{für } \frac{1}{2} + \frac{1}{n} \leq u \leq 1, \quad 0 \leq v \leq 1, \end{cases}$$

und für $\frac{1}{2} - \frac{1}{n} \leq u \leq \frac{1}{2} + \frac{1}{n}$, $v = \text{konst.}$ sei $x_n(u, v)$ eine lineare Funktion von u ; überdies sei $x_n(u, v)$ in B stetig. Analog werden $y_n(u, v)$, $z_n(u, v)$ erklärt; F_n ist also gewissermaßen aus $F_n^{(1)}$ und $F_n^{(2)}$ zusammengestückt. Offenbar gilt $F_n \rightarrow F$; wegen $P \in \mathfrak{R}(F; \varepsilon)$ ist daher P , für hinreichend großes n , in $\mathfrak{R}(F_n; \varepsilon)$ enthalten, d. h. es gibt in B einen Punkt $A_n = (u_n, v_n)$, so daß sich der Punkt

$$A_n^* = (x_n(A_n), y_n(A_n), z_n(A_n))$$

auf der Ebene ε in den Punkt P projiziert. Es muß nun

$$\frac{1}{2} - \frac{1}{n} \leq u_n \leq \frac{1}{2} + \frac{1}{n}$$

sein, da sonst, mit Rücksicht auf die Erklärung der Funktionen $x_n(u, v)$, $y_n(u, v)$, $z_n(u, v)$, der Punkt A_n^* entweder auf $F_n^{(1)}$ oder auf $F_n^{(2)}$ und folglich P selbst entweder in $\mathfrak{R}(F_n^{(1)}; \varepsilon)$ oder in $\mathfrak{R}(F_n^{(2)}; \varepsilon)$ liegen würde, was den Annahmen widerspricht. Die Punktfolge A_n hat daher eine Häufungsstelle auf der Strecke γ ; daraus folgt, da $x_n(u, v) \rightarrow x(u, v)$, $y_n(u, v) \rightarrow y(u, v)$, $z_n(u, v) \rightarrow z(u, v)$ gleichmäßig in B gilt, daß die Punkte A_n^* eine Häufungsstelle am Bogen Γ haben. Ist A^* ein am Bogen Γ liegender Häufungspunkt der Punkte A_n^* , so projiziert sich A^* , ebenso wie die Punkte A_n^* , auf der Ebene ε in den Punkt P , womit erkannt ist, daß P in der Projektion Γ_ε von Γ auf ε enthalten ist; dies sollte eben bewiesen werden.

8. Die Bezeichnungen der vorangehenden Nr. 7 mögen beibehalten werden, und es werde noch vorausgesetzt, daß (4) eine in typischer Darstellung vorliegende rektifizierbare Fläche und daß auch der Bogen γ rektifizierbar ist. Aus den Ungleichungen (2) folgt dann, daß auch der Bogen Γ rektifizierbar ist, und daraus folgt bekanntlich weiter, daß die Projektion von Γ auf irgendeiner Ebene das (zweidimensionale) Maß Null hat. Unter den angeführten zusätzlichen Voraussetzungen läßt sich also (13) zu

$$m\mathfrak{R}(F; \varepsilon) \leq m\mathfrak{R}(F_1; \varepsilon) + m\mathfrak{R}(F_2; \varepsilon)$$

verschärfen.

9. Wir wollen schließlich für $m\mathfrak{R}(F; \varepsilon)$ Schranken ableiten. — Aus Bequemlichkeitsgründen führen wir in der uv -Ebene die komplexe Veränderliche $\zeta = u + iv$, und in der ε -Ebene, nach Zugrundelegung eines rechtwinkligen Achsenkreuzes, eine komplexe Veränderliche τ ein. Ordnen wir jedem Punkte des in (4) auftretenden Bereiches B denjenigen τ -Wert zu, welcher in der ε -Ebene durch die Projektion des Punktes $x = x(u, v)$, $y = y(u, v)$, $z = z(u, v)$ dargestellt wird, so entsteht eine in B eindeutige stetige Funktion $\tau = f(\zeta)$; die Menge der Bildpunkte τ , die den Punkten ζ von B vermöge der Gleichung $\tau = f(\zeta)$ entsprechen, ist eben die Projektion $\mathfrak{P}(F; \varepsilon)$.

Sei C die Randkurve von B . Wenn ζ die geschlossene Jordankurve C beschreibt, so beschreibt der Bildpunkt $\tau = f(\zeta)$ in der ε -Ebene eine geschlossene (nicht notwendig einfache) stetige Kurve \mathfrak{C} . Wir erklären in der ε -Ebene eine Punktmenge $\mathfrak{D}(F; \varepsilon)$ durch die Festsetzung: ein Punkt τ^* gehört zu $\mathfrak{D}(F; \varepsilon)$, wenn erstens τ^* nicht auf \mathfrak{C} liegt, und wenn zweitens die Ordnung von τ^* in bezug auf \mathfrak{C} von Null verschieden ist²¹). Um diese Erklärung auf eine geeignetere Form zu bringen, verabreden wir die folgende Bezeichnung. Ist $\varphi(\zeta)$ eine in B eindeutige stetige, am Rande von B von Null verschiedene komplexe Funktion, so bedeute $V_B(\varphi(\zeta))$ die Variation des Argumentes von $\varphi(\zeta)$ bei einem positiven Umlaufe der Randkurve C von B . Wegen $\varphi(\zeta) \neq 0$ am Rande von B ist diese Variation eine wohlbestimmte Zahl von der Form $2k\pi$, unter k eine ganze Zahl ($= 0, \pm 1, \pm 2, \dots$) verstanden. Die Erklärung von $\mathfrak{D}(F; \varepsilon)$ gestattet dann die folgende gleichwertige Fassung: ein Punkt τ^* der Ebene ε gehört zu $\mathfrak{D}(F; \varepsilon)$, wenn

$$(14) \quad \begin{cases} f(\zeta) - \tau^* \neq 0 \text{ am Rande von } B \text{ und} \\ V_B(f(\zeta) - \tau^*) \neq 0 \text{ ist.} \end{cases}$$

²¹) Über den Begriff der Ordnung, sowie über die darauf bezüglichen, im folgenden zu verwendenden Sätze verweisen wir den Leser auf das Buch von Herrn B. v. Kerékjártó, Vorlesungen über Topologie I. (Berlin, Julius Springer, 1923).

Wenn ein Punkt τ^* diese Bedingungen erfüllt, so gilt dasselbe offenbar auch für alle Punkte τ der Ebene ε , die in einer hinreichend kleinen Umgebung von τ^* liegen; die Punktmenge $\mathfrak{D}(F; \varepsilon)$ ist also offen (oder leer) und folglich auch meßbar. Wir beweisen die Beziehung

$$(15) \quad \mathfrak{D}(F; \varepsilon) \subset \mathfrak{K}(F; \varepsilon).$$

Sei τ^* ein Punkt von $\mathfrak{D}(F; \varepsilon)$ und $\{F_n\}$ irgendeine gegen F konvergierende Folge stetiger Flächen; ist F wieder durch (4) gegeben, so gestattet demnach F_n eine Darstellung

$$F_n : x = x_n(u, v), \quad y = y_n(u, v), \quad z = z_n(u, v), \quad (u, v) \in B,$$

wobei $x_n(u, v) \rightarrow x(u, v)$, $y_n(u, v) \rightarrow y(u, v)$, $z_n(u, v) \rightarrow z(u, v)$ gleichmäßig in B gilt. Es habe $f_n(\zeta)$ dieselbe Bedeutung in bezug auf F_n wie $f(\zeta)$ in bezug auf F ; dann gilt offenbar $f_n(\zeta) \rightarrow f(\zeta)$ gleichmäßig in B . Daraus folgt, da für τ^* die Bedingungen (14) erfüllt sind,

$$(16) \quad \begin{cases} f_n(\zeta) - \tau^* \neq 0 \text{ am Rande von } B, \\ V_B(f_n(\zeta) - \tau^*) = V_B(f(\zeta) - \tau^*) \neq 0 \end{cases}$$

für hinreichend großes n , etwa für $n > n_0$. Hieraus folgt weiter, daß für $n > n_0$ die Funktion $f_n(\zeta) - \tau^*$ wenigstens eine Nullstelle in B hat; sonst müßte nämlich, auf Grund des Monodromiesatzes, $V_B(f_n(\zeta) - \tau^*) = 0$ sein²²⁾, was mit (16) im Widerspruche steht. Das Vorhandensein einer Nullstelle von $f_n(\zeta) - \tau^*$ in B bedeutet aber, daß τ^* in $\mathfrak{P}(F_n; \varepsilon)$ enthalten ist. Es gilt also $\tau^* \in \mathfrak{P}(F_n; \varepsilon)$ für hinreichend großes n ; da $\{F_n\}$ eine beliebige gegen F konvergierende Folge war, so folgt daraus mit Rücksicht auf die Erklärung des Projektionskernes sogleich, daß τ^* in $\mathfrak{K}(F; \varepsilon)$ liegt, w. z. b. w. — Aus (15), in Verbindung mit der evidenten Beziehung $\mathfrak{K}(F; \varepsilon) \subset \mathfrak{P}(F; \varepsilon)$, ergibt sich nunmehr die Abschätzung

$$(17) \quad m\mathfrak{D}(F; \varepsilon) \leq m\mathfrak{K}(F; \varepsilon) \leq m\mathfrak{P}(F; \varepsilon).$$

§ 3.

Eine Ableitungsformel.

1. Es liege die stetige Fläche (4) vor. Im Bereiche B werde ein abgeschlossenes, beliebig orientiertes Rechteck R angenommen. Es bedeute F_R die stetige Fläche

$$F_R : x = x(u, v), \quad y = y(u, v), \quad z = z(u, v), \quad (u, v) \in R.$$

Die Größe $m\mathfrak{K}(F_R; \varepsilon)$ hängt dann, wenn die Darstellung (4) und die Ebene ε fixiert sind, nur vom Rechtecke R ab und stellt demnach eine

²²⁾ Siehe loc. cit. ²¹⁾, S. 196, Hilfssatz.

in B erklärte Rechtecksfunktion dar. Es sei (u_0, v_0) ein innerer Punkt von B , und Q ein in B liegendes abgeschlossenes Quadrat, welches den Punkt (u_0, v_0) enthält und sonst beliebig orientiert ist. Wenn der Quotient

$$(18) \quad \frac{m \mathfrak{R}(F_Q; \varepsilon)}{mQ}$$

für $mQ \rightarrow 0$ einen Grenzwert hat, so heißt bekanntlich dieser Grenzwert die Ableitung der Rechtecksfunktion $m \mathfrak{R}(F_R; \varepsilon)$ im Punkte (u_0, v_0) . Wir beweisen den folgenden Satz:

In jedem inneren Punkte von B , in welchem $x(u, v)$, $y(u, v)$, $z(u, v)$ gleichzeitig total differenzierbar sind²³⁾, ist die Rechtecksfunktion $m \mathfrak{R}(F_R; \varepsilon)$ differenzierbar, und zwar ist ihre Ableitung gleich dem absoluten Werte der Determinante

$$\begin{vmatrix} x_u & y_u & z_u \\ x_v & y_v & z_v \\ \lambda & \mu & \nu \end{vmatrix},$$

wobei λ, μ, ν die Richtungskosinuse der Normalen der Ebene ε bedeuten.

2. Da alle im Satze eingehenden Begriffe und Größen von der Wahl des Koordinatensystems xyz unabhängig sind, so dürfen und wollen wir bequemlichkeitshalber die Ebene ε mit der xy -Ebene zusammenfallen lassen.

Wir setzen

$$\begin{aligned} x_u(u_0, v_0) &= a, & x_v(u_0, v_0) &= b, & x(u_0, v_0) &= x_0, \\ y_u(u_0, v_0) &= c, & y_v(u_0, v_0) &= d, & y(u_0, v_0) &= y_0, \\ & & ad - bc &= \Delta; \end{aligned}$$

diese Ableitungen sind nach Voraussetzung vorhanden, und wegen $\lambda = \mu = 0$, $\nu = 1$ handelt es sich um den Beweis der Limesgleichung

$$(19) \quad \frac{m \mathfrak{R}(F_Q; \varepsilon)}{mQ} \rightarrow |\Delta| \quad \text{für } mQ \rightarrow 0,$$

wo also Q ein den Punkt (u_0, v_0) enthaltendes abgeschlossenes Quadrat bedeutet.

Die totale Differenzierbarkeit von $x(u, v)$, $y(u, v)$ im Punkte (u_0, v_0) bedeutet bekanntlich, daß diese Funktionen in der Form

$$\begin{aligned} x(u, v) &= x_0 + a(u - u_0) + b(v - v_0) + [(u - u_0)^2 + (v - v_0)^2]^{\frac{1}{2}} \varphi(u, v), \\ y(u, v) &= y_0 + c(u - u_0) + d(v - v_0) + [(u - u_0)^2 + (v - v_0)^2]^{\frac{1}{2}} \psi(u, v) \end{aligned}$$

²³⁾ Der sehr anschauliche Ausdruck *total differenzierbar* rührt von Rademacher her und bedeutet dasselbe, wie *differenzierbar im Sinne von Stolz*; was darunter zu verstehen ist, wird im Texte alsbald erklärt.

und führen zwei Hilfsquadrate \underline{Q} , \bar{Q} ein, die mit Q konzentrisch und gleichorientiert sind, und zwar soll

$$(26) \quad \begin{cases} l \left(1 - \frac{3\sqrt{2}\mu}{k}\right) \text{ die Seitenlänge von } \underline{Q}, \\ l \left(1 + \frac{3\sqrt{2}\mu}{k}\right) \text{ " " " } \bar{Q} \end{cases}$$

sein²⁶⁾. Dann gilt $\underline{Q} \subset Q \subset \bar{Q}$. Es seien \underline{Q}^* , Q^* , \bar{Q}^* die Bilder dieser Quadrate durch die affine Hilfsabbildung $\tau = f^*(\zeta)$; diese Bilder sind dann bekanntlich Parallelogramme, und man hat

$$(27) \quad \begin{cases} \underline{Q}^* \subset Q^* \subset \bar{Q}^*, \\ m \underline{Q}^* = |\Delta| m \underline{Q} = |\Delta| \left(1 - \frac{3\sqrt{2}\mu}{k}\right)^2 m Q, \\ m \bar{Q}^* = |\Delta| m \bar{Q} = |\Delta| \left(1 + \frac{3\sqrt{2}\mu}{k}\right)^2 m Q. \end{cases}$$

Wir beweisen erstens die Beziehung

$$(28) \quad \mathfrak{B}(F_Q; \varepsilon) \subset \bar{Q}^*,$$

welche mit unseren Bezeichnungen auch wie folgt formuliert werden kann: Ist

$$(29) \quad \tau^* \text{ kein Punkt von } \bar{Q}^* \text{ und gilt } \zeta \subset Q,$$

so gilt

$$(30) \quad f(\zeta) - \tau^* \neq 0.$$

Wir schreiben

$$(31) \quad f(\zeta) - \tau^* = (f^*(\zeta) - \tau^*) + (f(\zeta) - f^*(\zeta)).$$

Es gibt nun in der ζ -Ebene einen und nur einen Punkt ζ^* , so daß $\tau^* = f^*(\zeta^*)$ wird; wegen (29) ist

$$(32) \quad \zeta^* \text{ kein Punkt von } \bar{Q}.$$

Alsdann gilt, mit Rücksicht auf die Erklärung von \bar{Q} und wegen (23), (20), (21),

$$|f^*(\zeta) - \tau^*| = |f^*(\zeta) - f^*(\zeta^*)| \geq k |\zeta - \zeta^*| \geq \frac{3}{2} \sqrt{2} l \mu,$$

$$|f(\zeta) - f^*(\zeta)| = |\zeta - \zeta_0| |\chi(\zeta)| \leq l \sqrt{2} \mu,$$

für $\zeta \subset Q$. Aus (31) folgt daher, für $\zeta \subset Q$,

$$|f(\zeta) - \tau^*| \geq |f^*(\zeta) - \tau^*| - |f(\zeta) - f^*(\zeta)| \geq \frac{1}{2} \sqrt{2} l \mu > 0,$$

womit (30) bewiesen ist.

²⁶⁾ Der Leser wolle sich eine schematische Figur entwerfen.

Wir beweisen zweitens die Beziehung

$$(33) \quad \underline{Q}^* \subset \mathfrak{D}(F_Q; \varepsilon),$$

welche auch wie folgt formuliert werden kann: Ist τ^* ein Punkt der τ -Ebene und gilt

$$(34) \quad \tau^* \in \underline{Q}^*,$$

so gilt auch

$$(35) \quad f(\zeta) - \tau^* \neq 0 \text{ am Rande von } Q, \text{ und}$$

$$(36) \quad V_B(f(\zeta) - \tau^*) \neq 0.$$

Zum Beweise sei wieder ζ^* derjenige Punkt der ζ -Ebene, für welchen $\tau^* = f^*(\zeta^*)$ gilt; wegen (34) gilt dann

$$(37) \quad \zeta^* \in \underline{Q}.$$

Wenn ζ den Rand von Q beschreibt, so beschreibt $\tau = f^*(\zeta)$ den Rand von Q^* ; da, wegen (34), τ^* innerer Punkt von Q^* und da die Abbildung $\tau = f^*(\zeta)$ umkehrbar eindeutig ist, so hat man

$$(38) \quad V_Q(f^*(\zeta) - \tau^*) = \pm 2\pi.$$

Wir vergleichen nun $V_Q(f(\zeta) - \tau^*)$ mit $V_Q(f^*(\zeta) - \tau^*)$. Zu dem Ende schreiben wir wieder

$$(39) \quad f(\zeta) - \tau^* = (f(\zeta) - f^*(\zeta)) + (f^*(\zeta) - \tau^*).$$

Es liege

$$(40) \quad \zeta \text{ am Rande von } Q.$$

Dann gilt, mit Rücksicht auf (23), (37), (40); (20), (21), (40),

$$|f^*(\zeta) - \tau^*| = |f^*(\zeta) - f^*(\zeta^*)| \geq k|\zeta - \zeta^*| \geq \frac{3}{2}\sqrt{2}l\mu,$$

$$|f(\zeta) - f^*(\zeta)| = |\zeta - \zeta_0| |\chi(\zeta)| \leq \sqrt{2}l\mu,$$

also

$$(41) \quad |f(\zeta) - f^*(\zeta)| < |f^*(\zeta) - \tau^*|,$$

sofern (34) und (40) erfüllt sind. Aus (39), (41) folgt erstens unmittelbar (35). Aus (39), (41) folgt ferner, auf Grund des in der komplexen Funktionentheorie oft verwendeten Rouchéschen Satzes²⁷⁾,

$$V_Q(f(\zeta) - \tau^*) = V_Q(f^*(\zeta) - \tau^*),$$

und daraus, wegen (38), auch (36), womit (33) bewiesen ist.

²⁷⁾ Vgl. die rein topologische Fassung bei K. Szilárd, Grundlagen der Funktionentheorie, Math. Zeitschr. 26 (1927), S. 655, Satz 2.

Aus (28), (33) folgt nunmehr, wegen (17),

$$m \underline{Q}^* \leq m \mathfrak{R}(F_Q; \varepsilon) \leq m \bar{Q}^*,$$

d. h. mit Rücksicht auf (25), (26),

$$|\Delta| \left(1 - \frac{3\sqrt{2}\mu}{k}\right)^2 \leq \frac{m \mathfrak{R}(F_Q; \varepsilon)}{m \underline{Q}} \leq |\Delta| \left(1 + \frac{3\sqrt{2}\mu}{k}\right)^2.$$

Da k und Δ von Q nicht abhängen, so folgt hieraus wegen (22) unmittelbar die Ableitungsformel (19).

4. Zweiter Fall: $\Delta = 0$.

In diesem Falle führt die Hilfsabbildung $\tau = f^*(\zeta)$ die ganze Ebene ζ entweder in den einzigen Punkt $\tau_0 = x_0 + i y_0$ oder aber in eine gewisse, durch τ_0 gehende Gerade der τ -Ebene über. Dem Quadrate Q entspricht also dabei entweder der einzige Punkt τ_0 , oder aber eine durch τ_0 gehende Strecke, deren Länge nach (20) höchstens gleich

$$2\sqrt{2}l(a^2 + b^2 + c^2 + d^2)^{\frac{1}{2}}$$

ist, wo wieder l die Seitenlänge von Q bedeutet. Da nach (20), (21)

$$|f(\zeta) - f^*(\zeta)| = |\zeta - \zeta_0| |\chi(\zeta)| \leq \sqrt{2}l\mu \quad \text{für } \zeta \in Q$$

gilt, so kann demnach das Bild von Q durch die Abbildung $\tau = f(\zeta)$, d. i. eben die Projektion $\mathfrak{P}(F_Q; \varepsilon)$, in ein Rechteck mit den Seiten

$$2\sqrt{2}l\mu \quad \text{und} \quad 2\sqrt{2}l\mu + 2\sqrt{2}l(a^2 + b^2 + c^2 + d^2)^{\frac{1}{2}}$$

eingeschlossen werden; also gilt

$$m \mathfrak{P}(F_Q; \varepsilon) \leq 8l^2 \mu (\mu + (a^2 + b^2 + c^2 + d^2)^{\frac{1}{2}}).$$

Wegen $l^2 = m \underline{Q}$ folgt hieraus, mit Rücksicht auf (22),

$$\frac{m \mathfrak{P}(F_Q; \varepsilon)}{m \underline{Q}} \rightarrow 0 \quad \text{für} \quad m \underline{Q} \rightarrow 0.$$

Wegen (17) gilt a fortiori

$$\frac{m \mathfrak{R}(F_Q; \varepsilon)}{m \underline{Q}} \rightarrow 0 \quad \text{für} \quad m \underline{Q} \rightarrow 0,$$

womit die Ableitungsformel (19) auch im Falle $\Delta = 0$ erwiesen ist.

5. Ist die Fläche (4) eine in typischer Darstellung vorliegende rektifizierbare Fläche, d. h. sind in B die Ungleichungen (2) erfüllt, so sind bekanntlich $x(u, v)$, $y(u, v)$, $z(u, v)$ im Innern von B fast überall total differenzierbar²⁸⁾, so, daß also in diesem Falle die Ableitungsformel für die

²⁸⁾ Nach einem allgemeinen, von W. Stepanoff verschärften Satze von Rademacher [siehe H. Rademacher, l. c.²⁴⁾ und W. Stepanoff, Math. Annalen 90 (1923), S. 318 bis 320]. — Der im Texte benötigte Sonderfall findet sich bereits bei Geöcze, loc. cit.¹⁰⁾.

Rechtecksfunktion $m\mathfrak{R}(F_R; \varepsilon)$ fast überall im Innern von B gilt. Mit Rücksicht auf spätere Anwendung werde noch eine einfache Bemerkung hinzugefügt. Ist Q ein in B liegendes abgeschlossenes Quadrat mit der Seitenlänge l , und bedeutet wieder F_Q die Fläche

$$F_Q: x = x(u, v), \quad y = y(u, v), \quad z = z(u, v), \quad (u, v) \subset Q,$$

so folgt aus den Ungleichungen (2) sogleich, daß die Projektion $\mathfrak{P}(F_Q; \varepsilon)$ in eine Kreisscheibe eingeschlossen werden kann, deren Flächeninhalt kleiner als der mit einer endlichen, nur von K abhängigen Konstanten multiplizierte Flächeninhalt von Q ist. Wegen $\mathfrak{R}(F_Q; \varepsilon) \subset \mathfrak{P}(F_Q; \varepsilon)$ folgt hieraus, daß der Quotient

$$\frac{m\mathfrak{R}(F_Q; \varepsilon)}{mQ}$$

unterhalb einer festen, vom Quadrate Q und von der Ebene ε unabhängigen Schranke bleibt (etwa unterhalb der Schranke $6\pi K^2$).

II. Teil.

Anwendungen.

§ 1.

Das Geöczesche Flächenmaß.

1. Bis auf weiteres soll das Koordinatensystem xyz als fest betrachtet werden, da die in diesem Paragraphen zu betrachtenden Größen durch Prozesse erklärt sein werden, bei welchen die Koordinatenebenen xy, yz, zx eine ausgezeichnete Rolle spielen.

In $m\mathfrak{R}(F; \varepsilon)$ lassen wir die Ebene ε der Reihe nach mit den Koordinatenebenen xy, yz, zx zusammenfallen und bezeichnen die erhaltenen Größen mit $a(F), b(F), c(F)$. Wir setzen ferner

$$g(F) = [a(F)^2 + b(F)^2 + c(F)^2]^{\frac{1}{2}}.$$

Aus den Entwicklungen über die Fundamentalgröße $m\mathfrak{R}(F; \varepsilon)$ erhält man unmittelbar eine Reihe von Sätzen über $g(F)$, die wir wohl nicht besonders aufzuzählen brauchen; wenn wir in der Folge gewisse Eigenschaften von $g(F)$ benötigen, so zitieren wir einfach die entsprechenden Entwicklungen über $m\mathfrak{R}(F; \varepsilon)$. Hier werde nur so viel bemerkt, daß $g(F)$, bei festem Koordinatensystem xyz , nur von der Fläche F selbst abhängt.

2. Es liege die stetige Fläche (4) vor. Es seien B_1, B_2, \dots, B_r endlich viele, in B liegende, nicht übereinander greifende einfache Jordanbereiche, und es seien F_1, F_2, \dots, F_r die durch

$$F_k: x = x(u, v), \quad y = y(u, v), \quad z = z(u, v), \quad (u, v) \subset B_k, \\ (k = 1, 2, \dots, r)$$

gegebenen stetigen Flächen. Wir bilden die Summe $g(F_1) + g(F_2) + \dots + g(F_\nu)$ und bezeichnen mit $G(F)$ die obere Grenze (d. i. die kleinste obere Schranke) aller solcher Summen bei festgehaltener Parameterdarstellung, für alle mögliche Systeme B_1, B_2, \dots, B_ν , wobei auch die Anzahl ν dieser Bereiche frei veränderlich ist. Wir nennen $G(F)$ das *Geöczesche Flächenmaß* der Fläche F ²⁹⁾; man hat stets $G(F) \geq 0$, es kann aber $G(F)$ auch $+\infty$ sein. Vor allem wollen wir zeigen, daß $G(F)$ nur von der Fläche F selbst abhängt. Diesen Nachweis gründen wir auf zwei einfache Bemerkungen.

3. Es sei

$$F_n: x = x_n(u, v), \quad y = y_n(u, v), \quad z = z_n(u, v), \quad (u, v) \subset B$$

eine Folge stetiger Flächen, und es gelte gleichmäßig in B

$$x_n(u, v) \rightarrow x(u, v), \quad y_n(u, v) \rightarrow y(u, v), \quad z_n(u, v) \rightarrow z(u, v).$$

Es bedeute ferner G_n das mit Hilfe der angeschriebenen Darstellung gebildete Geöczesche Flächenmaß von F_n , und G das mit Hilfe der Darstellung (4) gebildete Geöczesche Flächenmaß von F . Alsdann gilt

$$(42) \quad \underline{\lim} G_n \geq G.$$

Sei nämlich $\eta > 0$ beliebig vorgegeben. Nach Definition kann dann in B ein System B_1, B_2, \dots, B_ν von nicht übereinander greifenden einfachen Jordanbereichen derart angegeben werden, daß

$$g(F_1) + g(F_2) + \dots + g(F_\nu) > G - \eta$$

gilt, wobei die stetigen Flächen F_1, F_2, \dots, F_ν durch

$$F_k: x = x(u, v), \quad y = y(u, v), \quad z = z(u, v), \quad (u, v) \subset B_k, \\ (k = 1, 2, \dots, \nu)$$

gegeben sind. Wir betrachten die durch

$$F_k^{(n)}: x = x_n(u, v), \quad y = y_n(u, v), \quad z = z_n(u, v), \quad (u, v) \subset B_k, \\ (k = 1, 2, \dots, \nu)$$

gegebenen stetigen Flächen. Alsdann gilt $F_k^{(n)} \rightarrow F_k$ für $n \rightarrow \infty$, und folglich (I, 2, 4)

$$\underline{\lim} g(F_1^{(n)}) \geq g(F_1), \dots, \underline{\lim} g(F_\nu^{(n)}) \geq g(F_\nu).$$

Da nach Definition

$$G_n \geq g(F_1^{(n)}) + \dots + g(F_\nu^{(n)})$$

²⁹⁾ Diese Größe $G(F)$ ist mit der von Geöcze untersuchten nicht identisch, da wir seine Fundamentalgröße durch eine andere ersetzt haben. Eine analoge Bemerkung gilt für die in II, 3, 1 einzuführende Größe $P(F)$.

ist, so ergibt sich weiter

$$\lim G_n \geq \lim g(F_1^{(n)}) + \dots + \lim g(F_r^{(n)}) \geq g(F_1) + \dots + g(F_r) > G - \eta.$$

Da $\eta > 0$ beliebig war, so ist damit (42) bewiesen.

4. Es seien

$$F: \begin{cases} x = x(u, v), & y = y(u, v), & z = z(u, v), & (u, v) \subset B, \\ x = \bar{x}(\bar{u}, \bar{v}), & y = \bar{y}(\bar{u}, \bar{v}), & z = \bar{z}(\bar{u}, \bar{v}), & (\bar{u}, \bar{v}) \subset \bar{B} \end{cases}$$

zwei *ähnliche* Parameterdarstellungen derselben stetigen Fläche F , und G bzw. \bar{G} die mit Hilfe dieser Darstellungen gebildeten Geöczeschen Flächenmaße derselben. Alsdann gilt $G = \bar{G}$.

Beweis. Nach Voraussetzung gibt es eine topologische Abbildung T von B auf \bar{B} , so daß für entsprechende Punkte $A = (u, v)$ und $\bar{A} = (\bar{u}, \bar{v})$ die Gleichungen

$$x(A) = \bar{x}(\bar{A}), \quad y(A) = \bar{y}(\bar{A}), \quad z(A) = \bar{z}(\bar{A})$$

bestehen. Es sei B_1, B_2, \dots, B_r irgendein System von in B liegenden, nicht übereinander greifenden einfachen Jordanbereichen; durch die Abbildung T erhält man daraus in \bar{B} ein System $\bar{B}_1, \bar{B}_2, \dots, \bar{B}_r$ von derselben Beschaffenheit. Die Gleichungen

$$\begin{aligned} x &= x(u, v), & y &= y(u, v), & z &= z(u, v), & (u, v) &\subset B_k, \\ x &= \bar{x}(\bar{u}, \bar{v}), & y &= \bar{y}(\bar{u}, \bar{v}), & z &= \bar{z}(\bar{u}, \bar{v}), & (\bar{u}, \bar{v}) &= \bar{B}_k, \\ & & & & & & (k = 1, 2, \dots, r) \end{aligned}$$

bestimmen dann dieselbe stetige Fläche F_k ; da $g(F_k)$ von der Wahl der Parameterdarstellung unabhängig ist, so ergibt sich demnach dieselbe Summe $g(F_1) + g(F_2) + \dots + g(F_r)$, ob man vom System B_1, B_2, \dots, B_r oder vom System $\bar{B}_1, \bar{B}_2, \dots, \bar{B}_r$ ausgeht. Da diese Summe durch passende Wahl des Systems B_1, B_2, \dots, B_r beliebig nahe an G gebracht werden kann, so folgt hieraus $G \leq \bar{G}$, und durch Vertauschung der Rollen der beiden Darstellungen ergibt sich die inverse Ungleichung $\bar{G} \leq G$.

5. Es seien nunmehr

$$(43) \quad F: \begin{cases} x = x(u, v), & y = y(u, v), & z = z(u, v), & (u, v) \subset B, \end{cases}$$

$$(44) \quad \begin{cases} x = \bar{x}(\bar{u}, \bar{v}), & y = \bar{y}(\bar{u}, \bar{v}), & z = \bar{z}(\bar{u}, \bar{v}), & (\bar{u}, \bar{v}) \subset \bar{B} \end{cases}$$

irgend zwei (nicht notwendig ähnliche) Darstellungen derselben stetigen Fläche F , und G, \bar{G} die mit Hilfe derselben gebildeten Geöczeschen Flächenmaße. Aus der Fréchet'schen Erklärung der Identität zweier Flächen folgt dann vor allem, daß F eine Folge von Darstellungen der Form

$$(45) \quad F: x = x_n(u, v), \quad y = y_n(u, v), \quad z = z_n(u, v), \quad (u, v) \subset B$$

gestattet, wobei

$$x_n(u, v) \rightarrow x(u, v), \quad y_n(u, v) \rightarrow y(u, v), \quad z_n(u, v) \rightarrow z(u, v)$$

gilt gleichmäßig in B , und überdies alle diese Darstellungen der Darstellung (44) *ähnlich* sind. Bedeutet daher G_n das mit Hilfe von (45) gebildete Geöczesche Flächenmaß, so gilt $\lim G_n \geq G$ nach Nr. 3 und $G_n = \bar{G}$ nach Nr. 4, woraus $\bar{G} \geq G$ folgt. Vertauscht man die Rollen von (43) und (44), so ergibt sich ebenso die inverse Ungleichung $G \geq \bar{G}$. Also ist $G = \bar{G}$, d. h. das Geöczesche Flächenmaß $G(F)$ hängt von der Wahl der Parameterdarstellung von F nicht ab.

Es ist dabei, wie oben bemerkt wurde, vorausgesetzt, daß das Koordinatensystem xyz fest ist; ob $G(F)$ von der Wahl dieses Koordinatensystems abhängt oder nicht, muß dahingestellt bleiben.

6. Das Ergebnis von Nr. 3 bedeutet nunmehr, da $G(F)$ nur von der Fläche abhängt, daß

$$\lim G(F_n) \geq G(F) \quad \text{für} \quad F_n \rightarrow F$$

gilt, d. h. daß das Geöczesche Flächenmaß $G(F)$, als Funktion der Fläche F von unten halbstetig ist.

7. Wir wollen nun $G(F)$, im Falle rektifizierbarer Flächen, mit dem klassischen Doppelintegrale identifizieren. Es sei

$$F: x = x(u, v), \quad y = y(u, v), \quad z = z(u, v), \quad (u, v) \subset B$$

eine in typischer Darstellung vorliegende rektifizierbare Fläche. Es bedeute $Q_{ij}^{(n)}$ das Quadrat

$$Q_{ij}^{(n)}: \frac{i}{2^n} \leq u \leq \frac{i+1}{2^n}, \quad \frac{j}{2^n} \leq v \leq \frac{j+1}{2^n} \quad (i, j = 0, \pm 1, \pm 2, \dots; n = 1, 2, \dots)$$

und, falls $Q_{ij}^{(n)}$ in B liegt, $F_{ij}^{(n)}$ die durch

$$F_{ij}^{(n)}: x = x(u, v), \quad y = y(u, v), \quad z = z(u, v), \quad (u, v) \subset Q_{ij}^{(n)}$$

bestimmte stetige Fläche. Es werde schließlich

$$\sigma_n(F) = \sum_{i,j} g(F_{ij}^{(n)})$$

gesetzt (die Summation ist also bei festem n auf diejenigen Quadrate $Q_{ij}^{(n)}$ zu erstrecken, die in B enthalten sind). Wir zerlegen den Beweisgang in drei Schritte³⁰⁾.

³⁰⁾ In der Terminologie der Burkillschen Theorie der Intervallfunktionen [J. C. Burkill, Functions of intervals, Proc. of the London Math. Soc. (2) 22 (1923), p. 275—310] läßt sich das Wesentliche des Beweisganges dahin charakterisieren, daß zunächst $G(F)$ als das Integral einer gewissen Rechtecksfunktion erkannt und dann mit Hilfe der Ableitung dieser Rechtecksfunktion berechnet wird. — In der

a) Unter den angeführten Bedingungen gilt zunächst

$$\underline{\lim} \sigma_n(F) \geq g(F).$$

Beweis. Mit Hilfe eines bekannten, oft verwendeten Verfahrens bilde man eine Folge $\{B^{(n)}\}$ von einfachen Jordanbereichen derart, daß $B^{(n)}$ aus endlich vielen Quadraten $Q_{ij}^{(n)}$ besteht, und daß die Folge $\{B^{(n)}\}$ die in I, 2, 6 beschriebenen Eigenschaften a) und b) besitzt³¹⁾. Es bedeute $F^{(n)}$ die stetige Fläche

$$F^{(n)}: x = x(u, v), \quad y = y(u, v), \quad z = z(u, v), \quad (u, v) \in B^{(n)},$$

und es möge $\sigma_n(F^{(n)})$ dieselbe Bedeutung in bezug auf $F^{(n)}$ haben wie $\sigma_n(F)$ in bezug auf F . Wegen $B^{(n)} \subset B$ gilt dann zunächst

$$\sigma_n(F^{(n)}) \leq \sigma_n(F).$$

Durch wiederholte Anwendung des Ergebnisses von I, 2, 8, indem man nämlich den Bereich $B^{(n)}$ in Quadrate $Q_{ij}^{(n)}$ zerlegt, ergibt sich

$$g(F^{(n)}) \leq \sigma_n(F^{(n)}).$$

Nach I, 2, 6 gilt schließlich $g(F^{(n)}) \rightarrow g(F)$. Aus diesen Beziehungen folgt nunmehr

$$\underline{\lim} \sigma_n(F) \geq \underline{\lim} \sigma_n(F^{(n)}) \geq \underline{\lim} g(F^{(n)}) = g(F),$$

w. z. b. w.

b) Unter denselben Bedingungen gilt sogar $\lim \sigma_n(F) = G(F)$.

Beweis. Da $\sigma_n(F)$ selbst eine derjenigen Summen ist, als deren obere Grenze $G(F)$ erklärt wurde, so gilt jedenfalls $\sigma_n(F) \leq G(F)$, so daß es genügt,

$$(46) \quad \underline{\lim} \sigma_n(F) \geq G(F)$$

zu beweisen. — Es sei nun $\eta > 0$ beliebig vorgegeben. Nach der Erklärung von $G(F)$ gibt es dann im Bereiche B ein System B_1, B_2, \dots, B_r von nicht übereinander greifenden einfachen Jordanbereichen derart, daß

$$g(F_1) + g(F_2) + \dots + g(F_r) > G(F) - \eta$$

gilt, wo allgemein F_k die stetige Fläche

$$F_k: x = x(u, v), \quad y = y(u, v), \quad z = z(u, v), \quad (u, v) \in B_k$$

Young-Burkillschen Theorie des Flächenmaßes [W. H. Young, On the area of surfaces, Proc. Roy. Soc. A 96 (1920), p. 72; J. C. Burkill, The expression of area as an integral, Proc. London Math. Soc. (2) 22 (1923), p. 311–336] wird das Flächenmaß direkt als das Integral einer gewissen Rechtecksfunktion erklärt; es wäre interessant zu untersuchen, in welcher Weise aus dieser Theorie die in der Einleitung angeführten Sätze über das klassische Doppelintegral erschlossen werden können.

³¹⁾ Vgl. Kerékjártó, loc. cit. ³¹⁾, S. 107.

bedeutet. Es möge $\sigma_n(F_k)$ dieselbe Bedeutung in bezug auf F_k haben wie $\sigma_n(F)$ in bezug auf F . Das unter a) gewonnene Ergebnis, auf die Flächen F_1, F_2, \dots, F_r angewendet, liefert dann

$$\lim_{n \rightarrow \infty} \sigma_n(F_1) \geq g(F_1), \dots, \lim_{n \rightarrow \infty} \sigma_n(F_r) \geq g(F_r).$$

Da B_1, B_2, \dots, B_r nicht übereinander greifen und in B liegen, so hat man

$$\sigma_n(F) \geq \sigma_n(F_1) + \dots + \sigma_n(F_r).$$

Aus diesen Beziehungen folgt nunmehr

$$\lim_{n \rightarrow \infty} \sigma_n(F) \geq \lim_{n \rightarrow \infty} \sigma_n(F_1) + \dots + \lim_{n \rightarrow \infty} \sigma_n(F_r) \geq g(F_1) + \dots + g(F_r) > G(F) - \eta.$$

Damit ist, da $\eta > 0$ beliebig war, (46) bewiesen.

c) Wir erklären nunmehr eine Treppenfunktion $\varphi_n(u, v)$ durch folgende Vorschrift. In jedem Quadrate $Q_{ij}^{(n)}$, welches in B enthalten ist³²⁾, habe $\varphi_n(u, v)$ den konstanten Wert

$$\frac{g(F_{ij}^{(n)})}{m Q_{ij}^{(n)}},$$

sonst sei $\varphi_n(u, v) = 0$. Bedeutet B_* das aus den inneren Punkten von B bestehende Gebiet, so ist, nach I, 3, 5, die Funktionenfolge $\{\varphi_n(u, v)\}$ in B_* gleichmäßig beschränkt, und es gilt fast überall in B_*

$$\varphi_n(u, v) \rightarrow \left[\left(\frac{\partial(x, y)}{\partial(u, v)} \right)^2 + \left(\frac{\partial(y, z)}{\partial(u, v)} \right)^2 + \left(\frac{\partial(z, x)}{\partial(u, v)} \right)^2 \right]^{\frac{1}{2}}.$$

Diese Beziehung darf, infolge der gleichmäßigen Beschränktheit der Folge $\varphi_n(u, v)$, gliedweise integriert werden, d. h. das über B_* erstreckte Integral von $\varphi_n(u, v)$ konvergiert gegen das klassische Doppelintegral. Andererseits gilt nach b)

$$\iint_{B_*} \varphi_n(u, v) du dv = \sigma_n(F) \rightarrow G(F),$$

und man hat demnach

$$G(F) = \iint_{B_*} \left[\left(\frac{\partial(x, y)}{\partial(u, v)} \right)^2 + \left(\frac{\partial(y, z)}{\partial(u, v)} \right)^2 + \left(\frac{\partial(z, x)}{\partial(u, v)} \right)^2 \right]^{\frac{1}{2}} du dv.$$

8. Im Falle einer rektifizierbaren Fläche ist also das Geöczesche Flächenmaß $G(F)$ gleich dem aus irgendeiner typischen Darstellung berechneten klassischen Doppelintegral³³⁾. Für $G(F)$ folgt hieraus, im Falle rektifizierbarer Flächen, die Unabhängigkeit von der Wahl des

³²⁾ Bei der Konstruktion von $\varphi_n(u, v)$ ist natürlich bei $Q_{ij}^{(n)}$ der obere Zeiger n fest, und es variieren nur die Zeiger i, j gemäß der Bedingung $Q_{ij}^{(n)} \subset B$.

³³⁾ Geöcze, loc. cit. ¹⁰⁾, b); vgl. indessen ²⁹⁾.

Koordinatensystems xyz , und für das klassische Doppelintegral ergeben sich die beiden in der Einleitung angekündigten Eigenschaften (Invarianz und Halbstetigkeit) unmittelbar aus den entsprechenden Eigenschaften von $G(F)$.

9. Es werde ausdrücklich bemerkt, daß es sehr wohl vorkommen kann, daß bei einer nicht typischen Darstellung einer rektifizierbaren Fläche F das klassische Doppelintegral existiert und einen von $G(F)$ verschiedenen Wert hat. Hingegen liefern nach obigem die aus typischen Darstellungen berechneten klassischen Doppelintegrale stets den Wert $G(F)$; um die Schreibweise zu vereinfachen, dürfen wir daher das auf irgendeine typische Darstellung bezügliche Doppelintegral einfach mit $I(F)$, ohne Angabe der betreffenden typischen Darstellung bezeichnen. In dem hiermit genau erklärten Sinne soll die Bezeichnung $I(F)$ in der Folge verwendet werden.

§ 2.

Das Lebesguesche Flächenmaß.

1. Eine stetige Fläche heiße eine *Polyederfläche* und werde mit Π bezeichnet, falls sie eine Darstellung

$$(47) \quad \Pi: x = x(u, v), \quad y = y(u, v), \quad z = z(u, v), \quad (u, v) \in B$$

der folgenden Beschaffenheit zuläßt. Der Bereich B hat einen einfachen geschlossenen Polygonzug zum Rande, und kann in endlich viele Dreiecke derart zerlegt werden, daß in jedem dieser Dreiecke $x(u, v)$, $y(u, v)$, $z(u, v)$ lineare Funktionen von u, v sind; eine solche Darstellung heiße wieder eine *typische Darstellung der Polyederfläche*. Sind $\delta_1, \delta_2, \dots, \delta_r$ die erwähnten Dreiecke von B , in welchen also $x(u, v)$, $y(u, v)$, $z(u, v)$ linear sind, so entsprechen denselben, vermöge (47), gewisse Dreiecke A_1, A_2, \dots, A_r im Raume xyz ; die Summe der Flächeninhalte von A_1, A_2, \dots, A_r heiße *das elementare Flächenmaß der Polyederfläche Π* und werde mit $E(\Pi)$ bezeichnet. Nach bekannten elementargeometrischen Formeln ist dann $E(\Pi)$ gleich dem klassischen Doppelintegrale; da überdies, bei einer typischen Darstellung von Π , die Funktionen $x(u, v)$, $y(u, v)$, $z(u, v)$ offenbar einer Lipschitzbedingung genügen, so gilt, nach II, 1, 8, $E(\Pi) = G(\Pi) = I(\Pi)$. Daraus ist erstens ersichtlich, daß $E(\Pi)$ von der Wahl der typischen Darstellung unabhängig ist, und zweitens, daß $E(\Pi)$ als Funktion von Π halbstetig von unten ist³⁴⁾. Die Unabhängigkeit von $E(\Pi)$ von der Wahl des Koordinatensystems xyz ist aus elementargeometrischen Gründen evident.

³⁴⁾ Herr Fréchet hat für diese letztere Tatsache einen direkten Beweis geliefert (La semi-continuité en géométrie élémentaire, Nouvelles Annales de Math. 3 (1924).

2. Es sei nunmehr F eine stetige Fläche; für jede gegen F konvergierende Polyederfolge $\{\Pi_n\}$ betrachte man $\underline{\lim} E(\Pi_n)$. Sei $L(F)$ die untere Grenze der Größen $\underline{\lim} E(\Pi_n)$, für alle mögliche gegen F konvergierende Polyederfolgen. Diese durch Lebesgue eingeführte Größe $L(F)$ heiße *das Lebesguesche Flächenmaß* von F ; wie aus der Definition unmittelbar ersichtlich, hängt $L(F)$ nur von der Fläche F selbst ab. — Als Funktion von F ist $L(F)$ von unten halbstetig³⁵⁾. In der Tat, gilt $F_n \rightarrow F$, so gibt es, wie aus der Definition unmittelbar folgt, für jedes n eine Polyederfläche Π_n mit den Eigenschaften

$$(48) \quad [F_n, \Pi_n] < \frac{1}{n}, \quad E(\Pi_n) < L(F_n) + \frac{1}{n}.$$

Wegen $[F, \Pi_n] \leq [F, F_n] + [F_n, \Pi_n]$ gilt dann $\Pi_n \rightarrow F$, und folglich, nach der Erklärung von $L(F)$, auch $\underline{\lim} E(\Pi_n) \geq L(F)$. Andererseits ist, nach (48), $\underline{\lim} L(F_n) \geq \underline{\lim} E(\Pi_n)$. Zusammenfassend hat man also

$$\underline{\lim} L(F_n) \geq \underline{\lim} E(\Pi_n) \geq L(F),$$

w. z. b. w.

3. Für jede stetige Fläche F gilt $G(F) \leq L(F)$.

Beweis. Auf Grund der Erklärung von $L(F)$ gibt es eine Polyederfolge $\{\Pi_n\}$ mit den Eigenschaften $\Pi_n \rightarrow F$, $\lim E(\Pi_n) = L(F)$. Wegen $E(\Pi_n) = G(\Pi_n)$ und II, 1, 6 folgt hieraus

$$L(F) = \lim E(\Pi_n) = \lim G(\Pi_n) \geq G(F).$$

Korollar. Es gilt stets $L(F) \geq m \mathfrak{R}(F; \varepsilon)$, wo ε irgendeine Ebene bedeutet. Da nämlich $L(F)$ und $m \mathfrak{R}(F; \varepsilon)$ von der Wahl des Koordinatensystems xyz unabhängig sind, so darf angenommen werden, daß ε mit der xy -Ebene zusammenfällt; alsdann gilt $m \mathfrak{R}(F; \varepsilon) \leq G(F)$ auf Grund der Erklärung von $G(F)$ und $G(F) \leq L(F)$ auf Grund des soeben bewiesenen Satzes, woraus die Behauptung $m \mathfrak{R}(F; \varepsilon) \leq L(F)$ folgt.

4. Ist F eine in typischer Darstellung vorliegende rektifizierbare Fläche, so gilt $L(F) = I(F)$, d. h. im Falle rektifizierbarer Flächen ist das Lebesguesche Flächenmaß gleich dem aus irgendeiner typischen Darstellung berechneten klassischen Doppelintegrale³⁶⁾.

Beweis. Für jede gegen F konvergierende Polyederfolge $\{\Pi_n\}$ gilt nach Definition $\underline{\lim} E(\Pi_n) \geq L(F)$, und wenn insbesondere $\lim E(\Pi_n)$ vorhanden ist, $\lim E(\Pi_n) \geq L(F)$. Nun aber hat Herr Rademacher im Laufe seiner schönen Untersuchungen über partielle und totale Differenzierbar-

³⁵⁾ Lebesgue, loc. cit. 9).

³⁶⁾ Geöcze, loc. cit. ¹⁰⁾, a). — In diesem Satze ist insbesondere enthalten, daß für jede Polyederfläche $E(\Pi) = L(\Pi)$ gilt; denn nach II, 2, 1 hat man $E(\Pi) = I(\Pi)$.

keit einen Satz bewiesen, welcher in unserer Terminologie besagt, daß zu jeder rektifizierbaren Fläche F Polyederfolgen $\{II_n\}$ mit den Eigenschaften

$$II_n \rightarrow F, \quad \lim E(II_n) = I(F);$$

konstruiert werden können^{36a}). Daraus folgt nach obiger Bemerkung $I(F) \geq L(F)$, während nach II, 1, 7 und II, 2, 3 für jede rektifizierbare Fläche $I(F) = G(F) \leq L(F)$ gilt; also ist $I(F) = L(F)$.

5. Wir besprechen noch einen Satz, der im folgenden Paragraphen Anwendung finden wird, und welcher im wesentlichen besagt, daß $L(F)$ ein inneres Maß ist. Es liege eine stetige Fläche

$$(49) \quad F: x = x(u, v), \quad y = y(u, v), \quad z = z(u, v), \quad (u, v) \subset B$$

vor. Im Bereiche B nehme man ein System B_1, B_2, \dots, B_ν von nicht übereinander greifenden einfachen Jordanbereichen an, und man betrachte die durch

$$F_k: x = x(u, v), \quad y = y(u, v), \quad z = z(u, v), \quad (u, v) \subset B_k \\ (k = 1, 2, \dots, \nu)$$

gegebenen stetigen Flächen. Alsdann gilt

$$(50) \quad L(F_1) + L(F_2) + \dots + L(F_\nu) \leq L(F).$$

Den Beweis führen wir am bequemsten in den folgenden Schritten.

a) F sei eine in typischer Darstellung vorliegende rektifizierbare Fläche. Die in (50) vorkommenden Größen sind dann, nach II, 2, 4, Integrale mit demselben Integranden, und die Behauptung folgt unmittelbar aus der gegenseitigen Lage der Bereiche $B, B_1, B_2, \dots, B_\nu$.

b) F sei wieder eine rektifizierbare Fläche, die Darstellung (49) sei zwar keine typische, doch einer typischen Darstellung ähnlich. Durch eine topologische Abbildung von B kann man dann aus (49) eine typische Darstellung erhalten, und die behauptete Ungleichung folgt aus a) nebst der Tatsache, daß das Lebesguesche Flächenmaß von der Parameterdarstellung nicht abhängt.

c) F sei schließlich eine allgemeine stetige Fläche. Nach der Definition von $L(F)$ gibt es eine Polyederfolge $\{II_n\}$ mit den Eigenschaften

$$II_n \rightarrow F, \quad E(II_n) \rightarrow L(F);$$

man kann dabei für jedes n eine Darstellung von II_n der Form

$$II_n: x = x_n(u, v), \quad y = y_n(u, v), \quad z = z_n(u, v), \quad (u, v) \subset B$$

^{36a}) Siehe H. Rademacher, Partielle und totale Differenzierbarkeit II, Math. Annalen 81 (1920), S. 52–63, insbesondere S. 60, Satz IX.

zugrunde legen derart, daß

$$x_n(u, v) \rightarrow x(u, v), \quad y_n(u, v) \rightarrow y(u, v), \quad z_n(u, v) \rightarrow z(u, v)$$

gilt gleichmäßig in B und daß überdies die angeschriebene Darstellung einer typischen Darstellung von Π_n ähnlich sei (I, 1, 4). Es bedeute dann $F_n^{(k)}$ die stetige Fläche

$$F_n^{(k)}: x = x_n(u, v), \quad y = y_n(u, v), \quad z = z_n(u, v), \quad (u, v) \in B_k, \\ (k = 1, 2, \dots, \nu).$$

Da jede in typischer Darstellung vorliegende Polyederfläche gleichzeitig eine in typischer Darstellung vorliegende rektifizierbare Fläche ist, so kann b) auf die Flächen $\Pi_n, F_n^{(1)}, F_n^{(2)}, \dots, F_n^{(k)}$ angewendet werden; mit Rücksicht auf $L(\Pi_n) = E(\Pi_n)$ erhält man auf diese Weise

$$L(F_n^{(1)}) + \dots + L(F_n^{(\nu)}) \leq E(\Pi_n).$$

Da, wegen $F_n^{(1)} \rightarrow F_1, \dots, F_n^{(\nu)} \rightarrow F_\nu,$

$$\lim_{n \rightarrow \infty} L(F_n^{(1)}) \geq L(F_1), \dots, \lim_{n \rightarrow \infty} L(F_n^{(\nu)}) \geq L(F_\nu)$$

gilt, so ergibt sich daraus

$$L(F) = \lim E(\Pi_n) \geq \lim L(F_n^{(1)}) + \dots + \lim L(F_n^{(\nu)}) \geq L(F_1) + \dots + L(F_\nu),$$

w. z. b. w.

§ 3.

Das Peanosche Flächenmaß.

1. Es liege die stetige Fläche

$$(4) \quad F: x = x(u, v), \quad y = y(u, v), \quad z = z(u, v), \quad (u, v) \in B$$

vor. Es sei B_1, B_2, \dots, B_ν ein System von in B liegenden, nicht übereinander greifenden, einfachen Jordanbereichen, und es seien F_1, F_2, \dots, F_ν die durch

$$F_k: x = x(u, v), \quad y = y(u, v), \quad z = z(u, v), \quad (u, v) \in B_k \\ (k = 1, 2, \dots, \nu)$$

erklärten stetigen Flächen. Man nehme noch ν Ebenen $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_\nu$ beliebig an und bilde die Summe

$$m \mathfrak{S}(F_1; \varepsilon_1) + \dots + m \mathfrak{S}(F_\nu; \varepsilon_\nu).$$

Die obere Grenze aller solcher Summen bei festgehaltener Parameterdarstellung, für alle Werte von ν und für alle Systeme $B_1, B_2, \dots, B_\nu; \varepsilon_1, \varepsilon_2, \dots, \varepsilon_\nu$, heiße das *Peanosche Flächenmaß* und werde mit $P(F)$ bezeichnet³⁷⁾. Der Nachweis dafür, daß $P(F)$ nur von der Fläche selbst

³⁷⁾ Vgl. ²⁹⁾.

abhängt, sowie daß $P(F)$ als Funktion von F von unten halbstetig ist, gestaltet sich analog wie bei $G(F)$; die Einzelheiten können wohl dem Leser überlassen werden. Der wesentliche Beweisgrund ist wieder der Umstand, daß die Fundamentalgröße $m \mathfrak{R}(F; \varepsilon)$ invariant und halbstetig ist. — Da bei der Erklärung von $P(F)$ das Koordinatensystem xyz keine ausgezeichnete Rolle spielt, so ist $P(F)$ auch von der Wahl desselben unabhängig.

2. Es gilt stets $P(F) \leq L(F)$.

Beweis. F sei durch (4) gegeben, und es sei

$$\Sigma = m \mathfrak{R}(F_1; \varepsilon_1) + \dots + m \mathfrak{R}(F_\nu; \varepsilon_\nu)$$

eine derjenigen Summen, als deren obere Grenze $P(F)$ erklärt wurde. Dann gilt, nach dem Korollar zu II, 2, 3,

$$m \mathfrak{R}(F_1; \varepsilon_1) \leq L(F_1), \dots, m \mathfrak{R}(F_\nu; \varepsilon_\nu) \leq L(F_\nu),$$

und folglich, mit Rücksicht auf II, 2, 5,

$$\Sigma \leq L(F_1) + \dots + L(F_\nu) \leq L(F).$$

Jede der fraglichen Summen ist also $\leq L(F)$, und folglich gilt dies auch für ihre obere Grenze $P(F)$.

3. Für jede rektifizierbare Fläche F gilt $P(F) = I(F)$, d. h. im Falle rektifizierbarer Flächen ist das Peanosche Flächenmaß gleich dem aus irgendeiner typischen Darstellung berechneten klassischen Doppelintegrale³⁸⁾.

Beweis. Es sei

$$F: x = x(u, v), \quad y = y(u, v), \quad z = z(u, v), \quad (u, v) \in B$$

eine in typischer Darstellung vorliegende rektifizierbare Fläche. Es bedeute $Q_{ij}^{(n)}$ das Quadrat

$$Q_{ij}^{(n)}: \frac{i}{2^n} \leq u \leq \frac{i+1}{2^n}, \quad \frac{j}{2^n} \leq v \leq \frac{j+1}{2^n} \quad (i, j = 0, \pm 1, \dots; n = 1, 2, \dots),$$

und, falls $Q_{ij}^{(n)}$ in B enthalten ist, $F_{ij}^{(n)}$ die stetige Fläche

$$F_{ij}^{(n)}: x = x(u, v), \quad y = y(u, v), \quad z = z(u, v), \quad (u, v) \in Q_{ij}^{(n)}.$$

Es bedeute schließlich $\lambda_{ij}^{(n)}$ die obere Grenze von $m \mathfrak{R}(F_{ij}^{(n)}; \varepsilon)$ für alle Lagen der Ebene ε . Wir setzen

$$\Sigma_n = \sum_{i,j} \lambda_{ij}^{(n)},$$

wobei, bei festem n , die Summation auf alle in B liegenden Quadrate $Q_{ij}^{(n)}$ zu erstrecken ist. Aus der Erklärung des Peanoschen Flächenmaßes folgt

³⁸⁾ Geöcze, loc. cit.¹⁰⁾, b); vgl. indessen ²⁹⁾.

unmittelbar $\Sigma_n \leq P(F)$; nach II, 3, 2 gilt $P(F) \leq L(F)$; da F rektifizierbar ist, so hat man $L(F) = I(F)$. Zusammenfassend ergibt sich daraus $\Sigma_n \leq P(F) \leq I(F)$; zum Beweise von $P(F) = I(F)$ genügt es also,

$$\lim_{n \rightarrow \infty} \Sigma_n \geq I(F)$$

zu erweisen. Dies kann nun etwa wie folgt geschehen. Es bedeute $\psi_n(u, v)$ die durch folgende Vorschrift erklärte Treppenfunktion: in jedem in B liegenden Quadrate $Q_{ij}^{(n)}$ hat $\psi_n(u, v)$ den konstanten Wert

$$\frac{\lambda_{ij}^{(n)}}{m Q_{ij}^{(n)}},$$

und sonst ist $\psi_n(u, v)$ gleich Null. Man hat dann

$$\Sigma_n = \iint_{B_*} \psi_n(u, v) du dv,$$

unter B_* das aus den inneren Punkten von B bestehende Gebiet verstanden. Wir zeigen nun, daß fast überall in B_*

$$(51) \quad \lim_{n \rightarrow \infty} \psi_n(u, v) \geq \left[\left(\frac{\partial(x, y)}{\partial(u, v)} \right)^2 + \left(\frac{\partial(y, z)}{\partial(u, v)} \right)^2 + \left(\frac{\partial(z, x)}{\partial(u, v)} \right)^2 \right]^{\frac{1}{2}}$$

gilt; auf Grund geläufiger Sätze der Integrationstheorie folgt dann daraus unmittelbar

$$\begin{aligned} \underline{\lim} \Sigma_n &= \lim_{B_*} \iint \psi_n(u, v) du dv \\ &\geq \iint_{B_*} \left[\left(\frac{\partial(x, y)}{\partial(u, v)} \right)^2 + \left(\frac{\partial(y, z)}{\partial(u, v)} \right)^2 + \left(\frac{\partial(z, x)}{\partial(u, v)} \right)^2 \right]^{\frac{1}{2}} du dv = I(F). \end{aligned}$$

Die Richtigkeit von (51) ergibt sich aber wie folgt. Es sei (u_0, v_0) ein solcher Punkt von B_* , in welchem $x(u, v)$, $y(u, v)$, $z(u, v)$ gleichzeitig total differenzierbar sind; dasjenige Quadrat $Q_{ij}^{(n)}$, welches (u_0, v_0) enthält, werde einfach mit Q_n , die entsprechende Fläche $F_{ij}^{(n)}$ mit F_n , und die entsprechende Größe $\lambda_{ij}^{(n)}$ mit λ_n bezeichnet. Bedeutet noch ε irgendeine Ebene, so hat man

$$\psi_n(u_0, v_0) = \frac{\lambda_n}{m Q_n} \geq \frac{m \mathfrak{K}(F_n; \varepsilon)}{m Q_n},$$

da ja λ_n die obere Grenze von $m \mathfrak{K}(F_n; \varepsilon)$ für alle Lagen der Ebene ε bedeutete. Aus dieser Beziehung folgt, mit Rücksicht auf die Ableitungsformel in II, 2, 1, daß $\lim_{n \rightarrow \infty} \psi_n(u_0, v_0)$ wenigstens gleich dem absoluten Betrage der Determinante

$$\begin{vmatrix} x_u & y_u & z_u \\ x_v & y_v & z_v \\ \lambda & \mu & \nu \end{vmatrix}$$

ist, wobei λ, μ, ν die Richtungskosinuse der Ebene ε bedeuten [und natürlich alles im Punkte (u_0, v_0) betrachtet wird]. Da nun durch passende Wahl der Ebene ε der absolute Betrag der angeschriebenen Determinante in die rechte Seite von (51) übergeht, so ist hiermit (51) für alle solche Punkte (u, v) erwiesen, in welchen $x(u, v), y(u, v), z(u, v)$ gleichzeitig total differenzierbar sind; also gilt (51) fast überall in B_* , da die Fläche F in typischer Darstellung vorliegt und daher $x(u, v), y(u, v), z(u, v)$ fast überall in B_* total differenzierbar sind.

4. Durch Zusammenfassung der Ergebnisse in II, 1, 7, II, 2, 4 und II, 3, 3 ergibt sich nunmehr *das Hauptresultat*

$$G(F) = L(F) = P(F) = I(F),$$

d. h. die in der Einleitung hervorgehobene schnittartige Bestimmung des Flächenmaßes rektifizierbarer Flächen.

5. Nach II, 2, 3 und II, 3, 2 gilt für jede stetige Fläche $G(F) \leq L(F)$ und $P(F) \leq L(F)$, d. h. die der Konstruktion von $G(F)$ und $P(F)$ zugrunde liegenden anschaulichen unteren Schranken für das Flächenmaß sind mit den zur Konstruktion von $L(F)$ dienenden anschaulichen oberen Schranken stets verträglich; ob aber zwischen diesen unteren und oberen Schranken auch im allgemeinen Falle nur eine einzige Zahl Platz hat, d. h. ob die für rektifizierbare Flächen erwiesenen Gleichungen $G(F) = L(F)$, $P(F) = L(F)$ auch für jede stetige Fläche erfüllt sind, muß dahingestellt bleiben. Die bejahende Antwort auf diese Frage würde bedeuten, daß auch im Falle einer beliebigen stetigen Fläche das Flächenmaß sozusagen durch einen anschaulichen Dedekindschen Schnitt festgelegt werden kann; mit Rücksicht auf die außerordentliche Mannigfaltigkeit der für das Flächenmaß vorgeschlagenen Definitionen würde dies eine Erkenntnis von prinzipieller Wichtigkeit darstellen³⁹⁾.

Szeged, November 1927.

³⁹⁾ Aus den Ergebnissen meiner unter ¹⁵⁾ zitierten Arbeit folgt leicht, daß die Beziehung $G(F) = L(F)$ für jede in der Form $z = f(x, y)$ darstellbare stetige Fläche gilt. Die schnittartige Bestimmung des Flächenmaßes ist somit a) für jede rektifizierbare Fläche, b) für jede in der Form $z = f(x, y)$ darstellbare stetige Fläche geleistet.

zu den Arbeiten: J. G. van der Corput, „Zum Teilerproblem“, Math. Annalen 98, S. 697—716, und L. W. Nieland, „Zum Kreisproblem“, in demselben Band, S. 717—736.

Wie Herr S. C. van Veen uns brieflich mitteilt, ist in der genannten dem Teilerproblem gewidmeten Arbeit der Übergang von (20) zu (21) nicht erlaubt, und enthält die andere erwähnte Abhandlung zwischen den Beziehungen (22) und (23) einen entsprechenden Fehler. Beim Teilerproblem ist in Formel (20), beim Kreisproblem ist in der Formel, unmittelbar nach (22), in der rechten Seite ein Summationszeichen übersehen worden. Um die Beweise in Ordnung zu bringen, genügen die folgenden Änderungen.

Zum Teilerproblem.

S. 703, Z. 5 v. u. lies $\tau > \frac{\log q}{2 \log 2} + \frac{1}{2}$ statt $\tau > \frac{\log q}{2}$.

S. 705, Z. 1 v. u. und S. 706, Z. 1, 2 und 3 v. o. ersetze man durch

$$\sum_{2^l \Omega < a \leq 2^{l+1} \Omega} a^{-\tau} \cdot a^{-\tau} F(a y, a) < 2^{l+\frac{1}{2}} \Omega \cdot (2^l \Omega)^{-2\tau} \cdot q^{l+1} F(\Omega y, \Omega),$$

so daß nach (20) wegen $2^{2\tau-1} > q$

$$(21) \quad \sum_{a > \Omega} a^{-\tau} \cdot a^{-\tau} F(a y, a) \leq \Omega^{1-2\tau} \cdot 2 q F(\Omega y, \Omega) \sum_{l=0}^{\infty} \left(\frac{q}{2^{2\tau-1}} \right)^l = c_{19} \Omega^{1-2\tau} F(\Omega y, \Omega).$$

S. 706, Z. 11 v. o. lies $\Omega^{-\frac{1}{2}}$ statt $\Omega^{-\frac{3}{2}}$,

Z. 13 v. o. lies $\Omega^{1-2\tau} F(\Omega y, \Omega)$ statt $\Omega^{-2\tau} F(\Omega y, \Omega)$.

Schließlich ändere man noch die folgenden Druckfehler:

S. 708, Z. 6 v. o. lies $\sum_{n > \Omega^2}$ statt $\sum_{n > \omega^2}$.

S. 710, Z. 4 v. o. lies $|z|$ statt z .

Zum Kreisproblem.

S. 724, Z. 8 v. u. lies $\tau > \frac{\log q}{2 \log 2} + \frac{1}{2}$ statt $\tau > \frac{\log q}{2 \log 2}$.

S. 729, Z. 1 v. o. lies $2^{2\tau-1} > q$ statt $2^{2\tau} > q$.

Z. 2 v. o. lies $2 \Omega_1 \sum_{l=0}^{\infty} \left(\frac{q}{2^{2\tau-1}} \right)^l$ statt $\sum_{l=0}^{\infty} \left(\frac{q}{2^{2\tau}} \right)^l$,

Z. 3 v. o. lies $c_{36} \Omega_1^{1-2\tau}$ statt $c_9 q \Omega^{-2\tau}$,

Z. 11 v. o. lies $\Omega^{-\frac{1}{2}}$ statt $\Omega^{-\frac{3}{2}}$,

Z. 13 v. o. lies $\Omega_1^{1-2\tau}$ statt $\Omega_1^{-2\tau}$,

lies $\left(\frac{1}{2} \sqrt{2} \right)^{1-2\tau}$ statt $\left(\frac{1}{2} \sqrt{2} \right)^{-2\tau}$.

Über adjungierte Variationsprobleme und adjungierte Extremalflächen.

Von

Alfred Haar in Szeged.

Das Vorbild für die Entwicklung desjenigen Teiles der Variationsrechnung, in dem die unbekanntenen Funktionen von einer Veränderlichen abhängen, war stets die Theorie der geodätischen Linien auf krummen Flächen, die von Gauß begründet und insbesondere von Darboux ausgebildet wurde. Nicht nur die Untersuchungen von Jacobi und Kneser — an denen man den Einfluß der Theorie der geodätischen Linien klar erkennt —, sondern auch die direkte Methode der Behandlung von Variationsproblemen, die man den berühmten Arbeiten Hilberts verdankt, fußen auf der genannten Theorie¹⁾.

Was nun jenen Teil der Variationsrechnung betrifft, in dem die unbekanntene Funktion von zwei unabhängigen Veränderlichen abhängt, so findet man in der Flächentheorie einen besonderen Fall desselben mit großer Ausführlichkeit behandelt: die Minimalflächen, deren Theorie in einer Reihe von klassischen Arbeiten entwickelt wurde. Dabei denken wir nicht an die mit diesen Flächen zusammenhängenden Existenzfragen (Plateausches Problem), sondern in erster Reihe an Theoreme, die sich mit der Minimalfläche „im kleinen“ beschäftigen, die sich also aus dem Verschwinden der ersten Variation ergeben.

Es liegt nun der Gedanke nahe, diesen Teil der Flächentheorie für die Variationsrechnung nützlich zu machen bzw. die in der Theorie der Minimalflächen erhaltenen Resultate auf allgemeinere Variationsprobleme zu übertragen, in Analogie dessen, was für die Theorie der geodätischen

¹⁾ Hilbert benutzt in erster Reihe das Beispiel der geodätischen Linie, um seine Methode darzulegen (Über das Dirichletsche Prinzip, Jahresbericht d. deutschen Math. Ver. 8, S. 184).

Linien bereits vollzogen wurde. In der Tat gelingt dies für Variationsprobleme von der Form

$$(1) \quad \delta \iint f \left(\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y} \right) dx dy = 0$$

(wo der Integrand nur die Ableitungen der unbekannteten Funktion enthält); dabei bedarf es natürlich einer *neuen Begriffsbildung*, in analoger Weise wie der Transversalitätsbegriff eingeführt werden mußte, um die Theorie der geodätischen Linien auf die Variationsrechnung übertragen zu können. In den folgenden Untersuchungen ist es der Begriff des *adjungierten Variationsproblems und der adjungierten Extremalfläche*, der die Übertragung der Sätze über Minimalflächen auf Variationsprobleme von der obigen Form ermöglicht.

In dem erwähnten Teil der Theorie der Minimalflächen spielt nämlich der von Bonnet eingeführte Begriff der adjungierten Minimalflächen eine fundamentale Rolle. Freilich wird diese Zuordnung, die jeder Minimalfläche eine andere Minimalfläche adjungiert, gewöhnlich in der Weise angegeben und studiert, daß man in der bekannten Weierstraßschen Darstellung der Minimalflächen (die jeder Minimalfläche in bestimmter Weise eine analytische Funktion zuordnet) die daselbst auftretende Funktion einer komplexen Veränderlichen durch das i -fache dieser Funktion ersetzt — eine Behandlungsweise, die zur Verallgemeinerung kaum brauchbar gemacht werden kann.²⁾ Wenn man aber den Begriff der adjungierten Minimalfläche von der Weierstraßschen Darstellung befreit, d. h. wenn man die adjungierte Fläche in allgemeinen Gaußschen Koordinaten untersucht, so gelangt man leicht zu einer solchen Form derselben, die einer Verallgemeinerung auf Variationsprobleme von der oben angegebenen Art fähig ist.

Indem man auf diese Weise zu jeder Extremalfläche des Variationsproblems (1) eine adjungierte Fläche zuordnet, gelangt man zu dem springenden Punkt der folgenden Untersuchungen; diese adjungierte Fläche ist nämlich im allgemeinen Falle keine Extremalfläche des Variationsproblems (1), wohl aber eines anderen Variationsproblems, des adjungierten Variationsproblems, das in einfacher Weise aus dem ursprünglichen ableitbar ist. Man erhält auf diese Weise eine Zuordnung, die jedem Variationsproblem von der Form (1) ein Problem derselben Form zuordnet; diese Beziehung ist involutorisch, und das Problem der Minimalflächen ist eben dadurch ausgezeichnet, daß es ein sich selbst adjungiertes Problem ist (§ 1).

Untersucht man nun die Abbildung einer Extremalfläche von (1) auf ihre adjungierte Fläche (§ 2), so gelangt man zu den Verallgemeinerungen

²⁾ Auch die ursprüngliche Bonnetsche Definition der adjungierten Minimalfläche benutzt ein ganz spezielles Koordinatensystem auf der Fläche (Comptes Rendus 1853, p. 532).

der erwähnten klassischen Sätze der Theorie der Minimalflächen, Von den so erhaltenen Resultaten wollen wir die folgenden hervorheben:

In der Theorie der Minimalflächen gilt der Satz, daß jede Minimalfläche in ihre adjungierte verbiegbare ist. In unserem allgemeineren Ideenkreis gilt der Satz, daß man auf der Extremalfläche und auf ihrer adjungierten eine solche nichteuklidische Maßbestimmung, die nur von den entsprechenden Variationsproblemen abhängt, einführen kann, daß die fragliche Abbildung in diesem Sinne eine Verbiegung ist. Damit gewinnt man das Resultat, daß durch jedes Variationsproblem in ganz bestimmter Weise eine Maßbestimmung auf seinen Extremalflächen festgelegt ist (die nur in dem Falle der Minimalflächen mit der gewöhnlichen Bogenlänge übereinstimmt); das Variationsintegral selbst wird der Flächeninhalt bei Zugrundelegung dieser Definition der Bogenlängen. Wenn man in dieser Maßbestimmung isotherme Parameter einführt, so erhält man ein Differentialgleichungssystem für die Koordinaten der Extremalfläche, das von besonders einfacher Art ist und zum Studium dieser Flächen äußerst geeignet ist. Diese Sätze sind die Verallgemeinerung der bekannten Sätze über die Darstellung der Minimalflächen durch Potentialfunktionen, und diese Differentialgleichungen leisten — wie mir scheint — auch in der weiteren Theorie des Variationsproblems (1) gute Dienste.

§ 1.

Begriff und Haupteigenschaften der adjungierten Extremalfläche und des adjungierten Variationsproblems.

1. Bezeichnungen und Voraussetzungen. Wir legen unseren Untersuchungen ein Variationsproblem von der Form

$$(1) \quad \delta \iint f(z_x, z_y) dx dy = 0$$

zugrunde und betrachten irgendeine Extremalfläche $z = z(x, y)$ desselben; wir setzen zur Abkürzung³⁾

$$z_x = p, \quad z_y = q.$$

Um elegantere Formeln zu gewinnen, schreiben wir die Gleichung dieser Extremalfläche durch Einführung zweier Parameter u, v in der Form

$$x = \xi(u, v), \quad y = \eta(u, v), \quad z = \zeta(u, v)$$

und setzen

$$(2) \quad p_1 = \begin{vmatrix} \eta_u & \zeta_u \\ \eta_v & \zeta_v \end{vmatrix}, \quad p_2 = \begin{vmatrix} \xi_u & \xi_v \\ \zeta_u & \zeta_v \end{vmatrix}, \quad p_3 = \begin{vmatrix} \xi_u & \eta_u \\ \xi_v & \eta_v \end{vmatrix}.$$

³⁾ Wir bezeichnen im folgenden die Ableitungen einer Funktion durch Hinzufügung der entsprechenden Indizes.

Es ist dann

$$\zeta(u, v) = z(\xi(u, v), \eta(u, v)) \quad \text{und} \quad p = -\frac{p_1}{p_3}, \quad q = -\frac{p_2}{p_3},$$

und das vorgelegte Variationsproblem geht — wenn man noch die Bezeichnung

$$(3) \quad F(p_1, p_2, p_3) = p_3 f\left(-\frac{p_1}{p_3}, -\frac{p_2}{p_3}\right)$$

einführt — in das folgende Problem über:

$$(I) \quad \delta \iint F(p_1, p_2, p_3) du dv = 0.$$

Wir nehmen an, daß die Integranden $f(p, q)$ und $F(p_1, p_2, p_3)$ für alle in Betracht kommenden Werte der Argumente hinreichend oft differenzierbare Funktionen sind; die letztere ist außerdem homogen vom ersten Grade. Folglich ist

$$p_1 F_{p_1} + p_2 F_{p_2} + p_3 F_{p_3} = F,$$

und es bestehen — wie man sich leicht überzeugt — die Relationen

$$(3') \quad F_{p_1} = -f_p, \quad F_{p_2} = -f_q, \quad F_{p_3} = f - p f_p - q f_q.$$

Es zeigen diese Formeln, daß die Ableitungen $F_{p_1}, F_{p_2}, F_{p_3}$ homogene Funktionen 0-ten Grades von p_1, p_2, p_3 sind und daher von der Wahl der Parameter u, v unabhängig sind.

2. Definition der adjungierten Extremalfläche. Die Euler-Lagrangeschen Differentialgleichungen des Variationsproblems (I) lauten:

$$\frac{\partial}{\partial u} \frac{\partial F}{\partial \xi_u} + \frac{\partial}{\partial v} \frac{\partial F}{\partial \xi_v} = 0, \quad \frac{\partial}{\partial u} \frac{\partial F}{\partial \eta_u} + \frac{\partial}{\partial v} \frac{\partial F}{\partial \eta_v} = 0, \quad \frac{\partial}{\partial u} \frac{\partial F}{\partial \zeta_u} + \frac{\partial}{\partial v} \frac{\partial F}{\partial \zeta_v} = 0.$$

Mit Rücksicht auf die leicht verifizierbaren Relationen

$$\begin{aligned} \frac{\partial F}{\partial \xi_u} &= \begin{vmatrix} \eta_v & \zeta_v \\ F_{p_2} & F_{p_3} \end{vmatrix}, & \frac{\partial F}{\partial \eta_u} &= \begin{vmatrix} \zeta_v & \xi_v \\ F_{p_2} & F_{p_1} \end{vmatrix}, & \frac{\partial F}{\partial \zeta_u} &= \begin{vmatrix} \xi_v & \eta_v \\ F_{p_1} & F_{p_2} \end{vmatrix}, \\ \frac{\partial F}{\partial \xi_v} &= - \begin{vmatrix} \eta_u & \zeta_u \\ F_{p_2} & F_{p_3} \end{vmatrix}, & \frac{\partial F}{\partial \eta_v} &= - \begin{vmatrix} \zeta_u & \xi_u \\ F_{p_2} & F_{p_1} \end{vmatrix}, & \frac{\partial F}{\partial \zeta_v} &= - \begin{vmatrix} \xi_u & \eta_u \\ F_{p_1} & F_{p_2} \end{vmatrix}, \end{aligned}$$

kann man diese Euler-Lagrangeschen Differentialgleichungen durch Einführung dreier Hilfsfunktionen

$$\bar{\xi}(u, v), \quad \bar{\eta}(u, v), \quad \bar{\zeta}(u, v)$$

(die bis auf einen konstanten Addenden eindeutig bestimmt sind) durch das folgende Differentialgleichungssystem erster Ordnung ersetzen⁴⁾:

⁴⁾ Es ist dies diejenige Form der Euler-Lagrangeschen Differentialgleichung, die man aus dem Verschwinden der ersten Variation ohne die Existenz der zweiten Ableitungen der Extremalfunktion voraussetzen ableiten kann. Vgl. meine Arbeit „Über die Variation der Doppelintegrale“, Journal für die reine u. angew. Math. 149, S. 1–18. Indem ich diese Differentialgleichungen für den Fall der Minimalflächen aufstellte, erkannte ich, daß die auftretenden Hilfsfunktionen mit den Koordinaten der Bonnetschen adjungierten Minimalfläche übereinstimmen, und dieser Umstand führte mich zu den Begriffsbildungen der vorliegenden Untersuchungen.

$$(II) \quad \begin{aligned} \bar{\xi}_u &= \begin{vmatrix} \eta_u & \zeta_u \\ F_{p_2} & F_{p_2} \end{vmatrix}, & \bar{\eta}_u &= \begin{vmatrix} \zeta_u & \xi_u \\ F_{p_2} & F_{p_1} \end{vmatrix}, & \bar{\xi}_u &= \begin{vmatrix} \xi_u & \eta_u \\ F_{p_1} & F_{p_2} \end{vmatrix}, \\ \bar{\xi}_v &= \begin{vmatrix} \eta_v & \zeta_v \\ F_{p_2} & F_{p_2} \end{vmatrix}, & \bar{\eta}_v &= \begin{vmatrix} \zeta_v & \xi_v \\ F_{p_2} & F_{p_1} \end{vmatrix}, & \bar{\xi}_v &= \begin{vmatrix} \xi_v & \eta_v \\ F_{p_1} & F_{p_2} \end{vmatrix}. \end{aligned}$$

Schreibt man diese Gleichungen in der folgenden Form

$$(III) \quad d\bar{\xi} = F_{p_2} d\eta - F_{p_2} d\zeta, \quad d\bar{\eta} = F_{p_1} d\zeta - F_{p_2} d\xi, \quad d\bar{\xi} = F_{p_2} d\xi - F_{p_1} d\eta,$$

so erkennt man unmittelbar — in Anbetracht dessen, daß die Ausdrücke $F_{p_1}, F_{p_2}, F_{p_2}$ von der Wahl der Parameter u, v unabhängig sind —, daß die Fläche

$$x = \bar{\xi}(u, v), \quad y = \bar{\eta}(u, v), \quad z = \bar{\xi}(u, v)$$

bis auf eine Parallelverschiebung eindeutig festgelegt ist, insbesondere von der Wahl der Parameter u, v unabhängig ist. Wir nennen diese Fläche die *in bezug auf das vorgelegte Variationsproblem adjungierte Fläche der betrachteten Extremalfläche*⁵⁾.

Vermöge der Gleichungen (III) ist nicht nur die adjungierte Fläche selbst definiert, sondern es ist auch eine Abbildung dieser Fläche auf die ursprüngliche Extremalfläche festgelegt, indem wir Punkte mit denselben Parameterwerten u, v einander entsprechen lassen.

Wir führen endlich der Symmetrie wegen die folgenden Bezeichnungen ein:

$$(2) \quad \bar{p}_1 = \begin{vmatrix} \bar{\eta}_u & \bar{\xi}_u \\ \bar{\eta}_v & \bar{\xi}_v \end{vmatrix}, \quad \bar{p}_2 = \begin{vmatrix} \bar{\xi}_u & \bar{\xi}_u \\ \bar{\xi}_v & \bar{\xi}_v \end{vmatrix}, \quad \bar{p}_3 = \begin{vmatrix} \bar{\xi}_u & \bar{\eta}_u \\ \bar{\xi}_v & \bar{\eta}_v \end{vmatrix}.$$

Man erhält — mit Rücksicht auf die Gleichungen (II) —

$$\begin{aligned} \bar{p}_1 &= \begin{vmatrix} F_{p_1} \zeta_u - F_{p_2} \xi_u & F_{p_2} \xi_u - F_{p_1} \eta_u \\ F_{p_1} \zeta_v - F_{p_2} \xi_v & F_{p_2} \xi_v - F_{p_1} \eta_v \end{vmatrix} = F_{p_1} \begin{vmatrix} \eta_u & \zeta_u \\ \eta_v & \zeta_v \end{vmatrix} + F_{p_1} F_{p_2} \begin{vmatrix} \zeta_u & \xi_u \\ \zeta_v & \xi_v \end{vmatrix} + \\ &+ F_{p_1} F_{p_2} \begin{vmatrix} \xi_u & \eta_u \\ \xi_v & \eta_v \end{vmatrix} = F_{p_1} (p_1 F_{p_1} + p_2 F_{p_2} + p_3 F_{p_2}) = F F_{p_1}, \end{aligned}$$

⁵⁾ Man erkennt unschwer, daß diese Definition im Falle der Minimalflächen

$$F = \sqrt{p_1^2 + p_2^2 + p_3^2}$$

mit der üblichen Festlegung der adjungierten Minimalfläche übereinstimmt. Führt man nämlich auf der Minimalfläche isotherme Koordinaten ein

$$\xi_u^2 + \eta_u^2 + \zeta_u^2 = \xi_v^2 + \eta_v^2 + \zeta_v^2, \quad \xi_u \xi_v + \eta_u \eta_v + \zeta_u \zeta_v = 0,$$

so gehen die Gleichungen (II) in die folgenden über:

$$\begin{aligned} \bar{\xi}_u &= -\xi_v, & \bar{\eta}_u &= -\eta_v, & \bar{\zeta}_u &= -\zeta_v, \\ \bar{\xi}_v &= \xi_u, & \bar{\eta}_v &= \eta_u, & \bar{\zeta}_v &= \zeta_u, \end{aligned}$$

d. h. ξ und $\bar{\xi}$, η und $\bar{\eta}$, ζ und $\bar{\zeta}$ sind konjugierte Potentialfunktionen; dies ist eben die übliche Definition der adjungierten Minimalfläche.

und in gleicher Weise die Gleichungen

$$(4) \quad \bar{p}_2 = F F_{p_2}, \quad \bar{p}_3 = F F_{p_3},$$

die zu der Proportion

$$(5) \quad \bar{p}_1 : \bar{p}_2 : \bar{p}_3 = F_{p_1} : F_{p_2} : F_{p_3}$$

führen.

3. Definition des adjungierten Variationsproblems. Für das Folgende ist nun die Tatsache von entscheidender Wichtigkeit, daß die soeben definierten *adjungierten Flächen* der Extremalfächen des vorgelegten Variationsproblems die *Extremalfächen eines neuen Variationsproblems sind, das durch das ursprüngliche Variationsproblem völlig bestimmt und von der Wahl der Extremalfäche unabhängig ist*. Wir werden dieses Variationsproblem — das zu dem ursprünglichen *adjungierte Variationsproblem* — einfach charakterisieren und den involutorischen Charakter dieser Beziehung beweisen, indem wir zeigen werden, daß die in bezug auf dieses letztere Variationsproblem adjungierten Flächen der Extremalfächen dieses Variationsproblems die Extremalfächen des ursprünglichen Problems sind.

Um dies klar zu erkennen, wollen wir zunächst x, y als unabhängige Variable nehmen und annehmen, daß die Gleichung der zu der Extremalfäche $z = z(x, y)$ adjungierten Fläche

$$\bar{x} = \bar{\xi}(x, y), \quad \bar{y} = \bar{\eta}(x, y), \quad \bar{z} = \bar{\zeta}(x, y)$$

auf die Form

$$\bar{z} = \bar{z}(\bar{x}, \bar{y})$$

gebracht werden kann. Mit anderen Worten, wir betrachten einen Teil der Extremalfäche, wo die Determinante

$$\bar{p}_3 = \begin{vmatrix} \bar{\xi}_x & \bar{\eta}_x \\ \bar{\xi}_y & \bar{\eta}_y \end{vmatrix} = F F_{p_3} = f(f - p f_p - q f_q)$$

von Null verschieden ist. Alsdann sind — wenn man zur Abkürzung

$$\bar{p} = \frac{\partial \bar{z}}{\partial \bar{x}}, \quad \bar{q} = \frac{\partial \bar{z}}{\partial \bar{y}}$$

einführt — mit Rücksicht auf die Gleichungen (2), (3') und (4)

$$(6) \quad \begin{aligned} \bar{p} &= -\frac{\bar{p}_1}{\bar{p}_3} = -\frac{F_{p_1}}{F_{p_3}} = \frac{f_p}{f - p f_p - q f_q}, \\ \bar{q} &= -\frac{\bar{p}_2}{\bar{p}_3} = -\frac{F_{p_2}}{F_{p_3}} = \frac{f_q}{f - p f_p - q f_q}. \end{aligned}$$

Es zeigen diese Gleichungen die Tatsache (die man auch aus den Gleichungen (5) entnehmen kann), daß die Richtung der Flächennormalen (\bar{p}, \bar{q}) in einem Punkte der adjungierten Fläche nur von der Richtung der Nor-

malen (p, q) im entsprechenden Punkte der Extremalfäche abhängt. Man verifiziert ohne Schwierigkeit die Formel

$$\frac{\partial(\bar{p}, \bar{q})}{\partial(p, q)} = \frac{f}{(f - p f_p - q f_q)^3} (f_{p p} f_{q q} - f_{p q}^2);$$

ist dieser Ausdruck von Null verschieden — was wir von nun an annehmen wollen —, so kann man die Gleichungen (6), die man in die eine Gleichung

$$(6') \quad \bar{p} dp + \bar{q} dq = \frac{df}{f - p f_p - q f_q}$$

vereinigen kann, nach p und q auflösen. Man erhält eine durchsichtige Lösungsformel, die die Symmetrie der beiden Variablenpaare p, q und \bar{p}, \bar{q} klar hervortreten läßt, in folgender Weise: Wir denken uns p, q durch \bar{p}, \bar{q} ausgedrückt und definieren die Funktion $\bar{f}(\bar{p}, \bar{q})$ durch die folgende Gleichung:

$$(7) \quad \bar{f}(\bar{p}, \bar{q}) = \frac{1}{f - p f_p - q f_q},$$

dann ist, — in voller Analogie zu (6), —

$$(8) \quad p = \frac{\bar{f}_{\bar{p}}}{\bar{f} - \bar{p} \bar{f}_{\bar{p}} - \bar{q} \bar{f}_{\bar{q}}}, \quad q = \frac{\bar{f}_{\bar{q}}}{\bar{f} - \bar{p} \bar{f}_{\bar{p}} - \bar{q} \bar{f}_{\bar{q}}}.$$

In der Tat, bringt man die Definitionsgleichung (7) — mit Rücksicht auf (6) — auf die Form

$$(8) \quad f(p, q) \bar{f}(\bar{p}, \bar{q}) = 1 + p \bar{p} + q \bar{q},$$

so erhält man unmittelbar

$$\bar{f} df + f d\bar{f} = (\bar{p} dp + \bar{q} dq) + (p d\bar{p} + q d\bar{q})$$

und wegen (6') und (7)

$$p d\bar{p} + q d\bar{q} = f d\bar{f} = f(\bar{f}_{\bar{p}} d\bar{p} + \bar{f}_{\bar{q}} d\bar{q}),$$

d. h.

$$(9) \quad p = f \bar{f}_{\bar{p}}, \quad q = f \bar{f}_{\bar{q}}.$$

Führt man diese Ausdrücke von p, q in die rechte Seite der Gleichung (8) ein, so ergibt sich

$$(7) \quad f = \frac{1}{\bar{f} - \bar{p} \bar{f}_{\bar{p}} - \bar{q} \bar{f}_{\bar{q}}}.$$

Durch Einsetzen dieses Ausdruckes an Stelle von f in den Gleichungen (9) erhält man die behauptete Gleichung (6).

Wir können unsere Formeln folgendermaßen zusammenfassen:

Unterwirft man die Größen p, q den Bedingungen

$$(10) \quad f \neq 0, \quad f - p f_p - q f_q \neq 0, \quad f_{p p} f_{q q} - f_{p q}^2 \neq 0$$

und führt die Größen \bar{p}, \bar{q} durch die Gleichungen

$$(6) \quad \bar{p} = \frac{f_p}{f - pf_p - qf_q}, \quad \bar{q} = \frac{f_q}{f - pf_p - qf_q},$$

die Funktion $\bar{f}(\bar{p}, \bar{q})$ aber durch die Gleichung

$$(7) \quad \bar{f}(\bar{p}, \bar{q}) = \frac{1}{f - pf_p - qf_q},$$

oder (was damit gleichbedeutend ist)

$$(8) \quad f(p, q) \bar{f}(\bar{p}, \bar{q}) = 1 + p\bar{p} + q\bar{q}$$

ein, so wird

$$(6) \quad p = \frac{\bar{f}_{\bar{p}}}{\bar{f} - \bar{p}\bar{f}_{\bar{p}} - \bar{q}\bar{f}_{\bar{q}}}, \quad q = \frac{\bar{f}_{\bar{q}}}{\bar{f} - \bar{p}\bar{f}_{\bar{p}} - \bar{q}\bar{f}_{\bar{q}}},$$

ferner

$$(7) \quad f = \frac{1}{\bar{f} - \bar{p}\bar{f}_{\bar{p}} - \bar{q}\bar{f}_{\bar{q}}}$$

und

$$(9) \quad \bar{f}_{\bar{p}} = \frac{p}{f}, \quad \bar{f}_{\bar{q}} = \frac{q}{f},$$

$$(9) \quad f_p = \frac{\bar{p}}{\bar{f}}, \quad f_q = \frac{\bar{q}}{\bar{f}}.$$

Es bestehen daher diese Zusammenhänge zwischen den Ableitungen

$\frac{\partial z}{\partial x} = p, \frac{\partial z}{\partial y} = q$ einer Extremalfläche des Variationsproblems

$$(1) \quad \delta \iint f(z_x, z_y) dx dy = 0$$

und den Ableitungen $\bar{z}_{\bar{x}} = \bar{p}, \bar{z}_{\bar{y}} = \bar{q}$ ihrer adjungierten Fläche $\bar{z} = \bar{z}(\bar{x}, \bar{y})$. Wir werden zeigen, daß diese Funktion eine Extremalfunktion des Variationsproblems

$$(1) \quad \delta \iint \bar{f}(\bar{z}_{\bar{x}}, \bar{z}_{\bar{y}}) d\bar{x} d\bar{y} = 0$$

ist, das wir als das zu (1) *adjungierte Variationsproblem* bezeichnen wollen.

4. Fortsetzung. Man kann den Zusammenhang zwischen den Integranden $f(p, q)$ und $\bar{f}(\bar{p}, \bar{q})$ des ursprünglichen und des adjungierten Variationsproblems in eleganter Weise geometrisch charakterisieren. Zu diesem Ende betrachten wir in einem Raum, in dem X, Y, Z die rechtwinkligen Koordinaten bezeichnen, die Fläche zweiter Ordnung

$$Z^2 = X^2 + Y^2 + 1$$

und denken daselbst den Integranden des ursprünglichen Variationsproblems durch die Fläche

$$(11) \quad Z = f(X, Y)$$

realisiert. Aus den elementaren Formeln der analytischen Geometrie entnimmt man unmittelbar, daß der Pol der im Punkte X, Y, Z dieser letzten Fläche gezogenen Tangentialebene in bezug auf die obige Fläche zweiter Ordnung die folgenden Koordinaten besitzt:

$$\bar{X} = \frac{f_X}{f - Xf_X - Yf_Y}, \quad \bar{Y} = \frac{f_Y}{f - Xf_X - Yf_Y}, \quad \bar{Z} = \frac{1}{f - Xf_X - Yf_Y}.$$

Durch Elimination von X, Y erhalten wir aus diesen Gleichungen

$$(12) \quad \bar{Z} = \bar{f}(\bar{X}, \bar{Y}),$$

wobei $\bar{f}(\bar{X}, \bar{Y})$ eben der Integrand des adjungierten Variationsproblems ist. D. h. die durch die Integranden der beiden Variationsprobleme (des ursprünglichen und des adjungierten) festgelegten Flächen (11) und (12) sind polarreziproke Flächen in bezug auf die betrachtete Fläche zweiter Ordnung.

Daraus erkennt man (was übrigens schon aus der völligen Symmetrie der Formeln (6), (7), (6̄), (7̄) folgt) den involutorischen Charakter des Adjungierens; d. h. bildet man das adjungierte Variationsproblem des Variationsproblems (1̄), so gelangt man zu dem ursprünglichen (1) zurück⁶⁾.

Unsere Formeln werden eleganter, wenn man statt der bisher benutzten x, y wieder allgemeine Parameter u, v als unabhängige Veränderliche einführt, was — nach den Bezeichnungen in 1. — darauf hinauskommt, statt $f(p, q)$ die Funktion $F(p_1, p_2, p_3) = p_3 f\left(-\frac{p_1}{p_3}, -\frac{p_2}{p_3}\right)$ als Integranden des Variationsproblems zu nehmen. In entsprechender Weise setzen wir für den Integranden des adjungierten Problems

$$(3) \quad \bar{F}(\bar{p}_1, \bar{p}_2, \bar{p}_3) = \bar{p}_3 \bar{f}\left(-\frac{\bar{p}_1}{\bar{p}_3}, -\frac{\bar{p}_2}{\bar{p}_3}\right).$$

Die Relationen (6), (7) und (6̄), (7̄) erhalten dann die folgende einfache Form (mit Rücksicht auf (3'))

$$(13) \quad \bar{p}_1 = \bar{F} F_{p_1}, \quad \bar{p}_2 = \bar{F} F_{p_2}, \quad \bar{p}_3 = \bar{F} F_{p_3},$$

$$(13) \quad p_1 = F \bar{F}_{\bar{p}_1}, \quad p_2 = F \bar{F}_{\bar{p}_2}, \quad p_3 = F \bar{F}_{\bar{p}_3},$$

die man natürlich auch unabhängig vom Vorangehenden, wie folgt begründen kann: Da nämlich $F_{p_1}, F_{p_2}, F_{p_3}$ homogene Funktionen 0-ter Ordnung sind,

⁶⁾ Auch die Ungleichungen (10) sind symmetrisch aufgebaut; denn man verifiziert unschwer die Relation

$$(f_{pp} f_{qq} - f_{pq}^2) (\bar{f}_{\bar{p}\bar{p}} \bar{f}_{\bar{q}\bar{q}} - \bar{f}_{\bar{p}\bar{q}}^2) = \left(\frac{1}{f \bar{f}}\right)^4,$$

so daß jene Ungleichungen — mit Rücksicht auf (7) — äquivalent mit der Aussage sind, daß $f(p, q)$ und $\bar{f}(\bar{p}, \bar{q})$ endlich und von Null verschieden sind.

so kann man vermöge der Gleichungen der ersten Zeile bzw. aus der Proportion $\bar{p}_1 : \bar{p}_2 : \bar{p}_3 = F_{p_1} : F_{p_2} : F_{p_3}$ die Größen $\frac{p_1}{p_3}$ und $\frac{p_2}{p_3}$ durch $\frac{\bar{p}_1}{\bar{p}_3}$ und $\frac{\bar{p}_2}{\bar{p}_3}$ ausdrücken; alsdann liefert etwa die dritte Gleichung \bar{F} als homogene Funktion erster Ordnung von $\bar{p}_1, \bar{p}_2, \bar{p}_3$.⁷⁾ Komponiert man die Gleichungen (13) — die man auch in der Form

$$\bar{p}_1 dp_1 + \bar{p}_2 dp_2 + \bar{p}_3 dp_3 = \bar{F} dF$$

schreiben kann — bzw. mit p_1, p_2, p_3 , so ergibt sich

$$p_1 \bar{p}_1 + p_2 \bar{p}_2 + p_3 \bar{p}_3 = F \bar{F}$$

und daraus

$$\bar{F} dF + F d\bar{F} = (\bar{p}_1 dp_1 + \bar{p}_2 dp_2 + \bar{p}_3 dp_3) + (p_1 d\bar{p}_1 + p_2 d\bar{p}_2 + p_3 d\bar{p}_3).$$

Folglich ist

$$F d\bar{F} = p_1 d\bar{p}_1 + p_2 d\bar{p}_2 + p_3 d\bar{p}_3,$$

und dies ist der Inhalt der Gleichungen ($\bar{13}$).

Bezeichnen wir daher — wie in 2. — mit p_1, p_2, p_3 bzw. $\bar{p}_1, \bar{p}_2, \bar{p}_3$ die Unterdeterminanten der aus der ursprünglichen Extremalfläche bzw. aus ihrer adjungierten Fläche gebildeten Matrix

$$\begin{pmatrix} \xi_u & \eta_u & \zeta_u \\ \xi_v & \eta_v & \zeta_v \end{pmatrix} \text{ bzw. } \begin{pmatrix} \bar{\xi}_u & \bar{\eta}_u & \bar{\zeta}_u \\ \bar{\xi}_v & \bar{\eta}_v & \bar{\zeta}_v \end{pmatrix},$$

so besteht nach (5) die Proportion $\bar{p}_1 : \bar{p}_2 : \bar{p}_3 = F_{p_1} : F_{p_2} : F_{p_3}$. Führt man daher in der soeben angegebenen Weise die Funktion $\bar{F}(\bar{p}_1, \bar{p}_2, \bar{p}_3)$ ein, so gilt nach ($\bar{13}$) die analoge Proportion

$$(5) \quad p_1 : p_2 : p_3 = \bar{F}_{\bar{p}_1} : \bar{F}_{\bar{p}_2} : \bar{F}_{\bar{p}_3}.$$

Dabei ist natürlich vorausgesetzt, daß man sich auf ein solches Stück der ursprünglichen Extremalfläche beschränkt, daß die Bestimmung von \bar{F} aus den Gleichungen (13) möglich ist, d. h. daß man $\frac{p_1}{p_3}$ und $\frac{p_2}{p_3}$ aus der aus diesen entspringenden Proportion durch $\frac{\bar{p}_1}{\bar{p}_3}$ und $\frac{\bar{p}_2}{\bar{p}_3}$ ausdrücken kann; die Bedingungen hierfür sind in (10) bzw. in der Anmerkung ⁶⁾ angegeben worden.

5. Die adjungierte Fläche als Extremalfläche des adjungierten Variationsproblems. Um den Zusammenhang zwischen adjungierten Flächen und adjungiertem Variationsproblem darzulegen, greifen wir auf die Definitionsgleichung dieser Flächen zurück, die wir in der Form

$$(III) \quad d\bar{\xi} = F_{p_2} d\eta - F_{p_3} d\zeta, \quad d\bar{\eta} = F_{p_1} d\zeta - F_{p_3} d\xi, \quad d\bar{\zeta} = F_{p_2} d\xi - F_{p_1} d\eta$$

⁷⁾ Auf eine andere einfache Berechnung von $\bar{F}(\bar{p}_1, \bar{p}_2, \bar{p}_3)$ wird in 6. hingewiesen.

zugrunde legen, und berechnen die analog aufgebauten Ausdrücke

$$\bar{F}_{\bar{p}_2} d\bar{\eta} - \bar{F}_{\bar{p}_2} d\bar{\zeta}, \quad \bar{F}_{\bar{p}_1} d\bar{\zeta} - \bar{F}_{\bar{p}_3} d\bar{\xi}, \quad \bar{F}_{\bar{p}_2} d\bar{\xi} - \bar{F}_{\bar{p}_1} d\bar{\eta}.$$

Eine einfache Rechnung ergibt mit Rücksicht auf (III) und (13) die Gleichungen

$$\begin{aligned} \bar{F}_{\bar{p}_2} d\bar{\eta} - \bar{F}_{\bar{p}_2} d\bar{\zeta} &= \frac{1}{F} [p_3 (F_{p_1} d\zeta - F_{p_3} d\xi) - p_2 (F_{p_2} d\xi - F_{p_1} d\eta)] \\ &= \frac{1}{F} [-(p_1 F_{p_1} + p_2 F_{p_2} + p_3 F_{p_3}) d\xi + F_{p_1} (p_1 d\xi + p_2 d\eta + p_3 d\zeta)] = -d\xi, \end{aligned}$$

und in gleicher Weise

$$(III) \quad \bar{F}_{\bar{p}_1} d\bar{\zeta} - \bar{F}_{\bar{p}_3} d\bar{\xi} = -d\eta, \quad \bar{F}_{\bar{p}_2} d\bar{\xi} - \bar{F}_{\bar{p}_1} d\bar{\eta} = -d\zeta.$$

Ausführlicher geschrieben lauten diese Gleichungen folgendermaßen:

$$(II) \quad \begin{aligned} -\xi_u &= \begin{vmatrix} \bar{\eta}_u & \bar{\xi}_u \\ \bar{F}_{\bar{p}_2} & \bar{F}_{\bar{p}_3} \end{vmatrix}, & -\eta_u &= \begin{vmatrix} \bar{\xi}_u & \bar{\xi}_u \\ \bar{F}_{\bar{p}_2} & \bar{F}_{\bar{p}_1} \end{vmatrix}, & -\zeta_u &= \begin{vmatrix} \bar{\xi}_u & \bar{\eta}_u \\ \bar{F}_{\bar{p}_1} & \bar{F}_{\bar{p}_2} \end{vmatrix}, \\ -\xi_v &= \begin{vmatrix} \bar{\eta}_v & \bar{\xi}_v \\ \bar{F}_{\bar{p}_2} & \bar{F}_{\bar{p}_3} \end{vmatrix}, & -\eta_v &= \begin{vmatrix} \bar{\xi}_v & \bar{\xi}_v \\ \bar{F}_{\bar{p}_2} & \bar{F}_{\bar{p}_1} \end{vmatrix}, & -\zeta_v &= \begin{vmatrix} \bar{\xi}_v & \bar{\eta}_v \\ \bar{F}_{\bar{p}_1} & \bar{F}_{\bar{p}_2} \end{vmatrix}. \end{aligned}$$

Aus diesen Gleichungen folgen aber durch Elimination von ξ, η, ζ die Gleichungen

$$\begin{aligned} \frac{\partial}{\partial u} \begin{vmatrix} \bar{\eta}_v & \bar{\xi}_v \\ \bar{F}_{\bar{p}_2} & \bar{F}_{\bar{p}_3} \end{vmatrix} - \frac{\partial}{\partial v} \begin{vmatrix} \bar{\eta}_u & \bar{\xi}_u \\ \bar{F}_{\bar{p}_2} & \bar{F}_{\bar{p}_3} \end{vmatrix} &= \frac{\partial}{\partial u} \begin{vmatrix} \bar{\xi}_v & \bar{\xi}_v \\ \bar{F}_{\bar{p}_2} & \bar{F}_{\bar{p}_1} \end{vmatrix} - \frac{\partial}{\partial v} \begin{vmatrix} \bar{\xi}_u & \bar{\xi}_u \\ \bar{F}_{\bar{p}_2} & \bar{F}_{\bar{p}_1} \end{vmatrix} \\ &= \frac{\partial}{\partial u} \begin{vmatrix} \bar{\xi}_v & \bar{\eta}_v \\ \bar{F}_{\bar{p}_1} & \bar{F}_{\bar{p}_2} \end{vmatrix} - \frac{\partial}{\partial v} \begin{vmatrix} \bar{\xi}_u & \bar{\eta}_u \\ \bar{F}_{\bar{p}_1} & \bar{F}_{\bar{p}_2} \end{vmatrix} = 0, \end{aligned}$$

die mit den Euler-Lagrangeschen Differentialgleichungen des adjungierten Variationsproblems

$$(I) \quad \delta \iint \bar{F}(\bar{p}_1, \bar{p}_2, \bar{p}_3) du dv = 0$$

übereinstimmen. Daher sind die adjungierten Flächen tatsächlich Lösungen des adjungierten Variationsproblems.

Die Gleichungen (II) zeigen aber weiter, daß die in bezug auf das soeben angeschriebene Variationsproblem gebildete adjungierte Fläche der Extremalfläche $x = \bar{\xi}(u, v)$, $y = \bar{\eta}(u, v)$, $z = \bar{\zeta}(u, v)$ diejenige Fläche ist, deren Gleichung

$$x = -\xi(u, v), \quad y = -\eta(u, v), \quad z = -\zeta(u, v),$$

d. h. das Spiegelbild der ursprünglichen Extremalfläche in bezug auf den Koordinatenanfangspunkt.

Wir gelangen daher zu dem folgenden Resultat:

Die in bezug auf das Variationsproblem

$$(I) \quad \delta \iint F(p_1, p_2, p_3) du dv = 0$$

gebildete adjungierte Fläche irgendeiner Extremalfläche dieses Problems ist Extremalfläche des adjungierten Variationsproblems

$$(\bar{I}) \quad \delta \iint \bar{F}(\bar{p}_1, \bar{p}_2, \bar{p}_3) du dv = 0;$$

bildet man zu dieser Fläche in bezug auf dieses Variationsproblem die adjungierte Fläche, so erhält man das Spiegelbild der ursprünglichen Extremalfläche in bezug auf den Koordinatenanfangspunkt; diese Fläche ist offenbar ebenfalls Extremalfläche des ursprünglichen Variationsproblems.

§ 2.

Über die Abbildung einer Extremalfläche auf ihre adjungierte Fläche.

6. Invarianz des Variationsintegrals. Wie bereits in 2. bemerkt wurde, es liefern die Gleichungen

$$(III) \quad d\bar{\xi} = \bar{F}_{\bar{p}_3} d\eta - \bar{F}_{\bar{p}_2} d\zeta, \quad d\bar{\eta} = \bar{F}_{\bar{p}_1} d\zeta - \bar{F}_{\bar{p}_3} d\xi, \quad d\bar{\zeta} = \bar{F}_{\bar{p}_2} d\xi - \bar{F}_{\bar{p}_1} d\eta,$$

die zur Definition der adjungierten Fläche dienen, gleichzeitig eine Abbildung zwischen diesen Flächen, indem wir Punkte beider Flächen, die zu denselben Parameterwerten gehören, einander entsprechen lassen; wir beschäftigen uns im Folgenden mit dieser Abbildung.

Aus der obigen Definition der adjungierten Fläche entnehmen wir in 2. die Relationen

$$(4) \quad \bar{p}_1 = F(p_1, p_2, p_3) F_{p_1}, \quad \bar{p}_2 = F(p_1, p_2, p_3) F_{p_2}, \quad \bar{p}_3 = F(p_1, p_2, p_3) F_{p_3};$$

andererseits gilt auf Grund der Definition des adjungierten Variationsproblems die Gleichung

$$(13) \quad \bar{p}_1 = \bar{F}(\bar{p}_1, \bar{p}_2, \bar{p}_3) F_{\bar{p}_1}, \quad \bar{p}_2 = \bar{F}(\bar{p}_1, \bar{p}_2, \bar{p}_3) F_{\bar{p}_2}, \quad \bar{p}_3 = \bar{F}(\bar{p}_1, \bar{p}_2, \bar{p}_3) F_{\bar{p}_3}.$$

Folglich ist

$$(14) \quad F(p_1, p_2, p_3) = \bar{F}(\bar{p}_1, \bar{p}_2, \bar{p}_3)$$

oder (für jedes Gebiet G):

$$\iint_G F(p_1, p_2, p_3) du dv = \iint_G \bar{F}(\bar{p}_1, \bar{p}_2, \bar{p}_3) du dv,$$

d. h. das Integral des ursprünglichen Variationsproblems, erstreckt über einen Teil einer Extremalfläche desselben, ist gleich dem Integral des adjungierten Variationsproblems, erstreckt über den entsprechenden Teil der adjungierten Fläche.

Die Gleichung (14) gibt ein bequemes Mittel zur Herstellung des adjungierten Variationsproblems. Man erhält nämlich den Integranden

$\bar{F}(\bar{p}_1, \bar{p}_2, \bar{p}_3)$, wenn man aus den Gleichungen (4) p_1, p_2, p_3 durch $\bar{p}_1, \bar{p}_2, \bar{p}_3$ ausdrückt und diese Ausdrücke an Stelle der Argumente in die Funktion $F(p_1, p_2, p_3)$ einsetzt.

7. Entsprechende Richtungen. Aus der Definitionsgleichung (III) der adjungierten Fläche folgt unmittelbar die Relation

$$d\xi d\bar{\xi} + d\eta d\bar{\eta} + d\zeta d\bar{\zeta} = 0,$$

d. h. jedem Linienelement auf einer Extremalfläche entspricht auf der adjungierten Fläche ein Linienelement, das auf diesem senkrecht steht.

8. Invarianz einer nichteuklidischen Bogenlänge. Um auf eine wesentlich tiefer liegende Eigenschaft der fraglichen Abbildung zu gelangen, erinnern wir an die bekannte Tatsache, daß jede Minimalfläche in ihre adjungierte Fläche verbiegbar ist. Dies ist natürlich bei allgemeinen Variationsproblemen nicht zu erwarten; wohl kann man aber auf der Extremalfläche bzw. auf der adjungierten Fläche eine nichteuklidische Maßbestimmung, die nur von dem entsprechenden Variationsproblem abhängig ist, derart angeben, daß im Sinne dieser Maßbestimmungen unsere Abbildung eine Verbiegung ist.

Um zu diesen Maßbestimmungen zu gelangen, verfahren wir, wie folgt: Wir bezeichnen mit a, b, c zunächst beliebige, später zu bestimmende Größen und betrachten die Matrix

$$\begin{pmatrix} a d\xi & b d\eta & c d\zeta \\ a F_{p_1} & b F_{p_2} & c F_{p_3} \end{pmatrix};$$

durch Anwendung der bekannten Laplaceschen Identität erhalten wir — mit Rücksicht auf III —

$$\begin{aligned} & (a^2 d\xi^2 + b^2 d\eta^2 + c^2 d\zeta^2)(a^2 F_{p_1}^2 + b^2 F_{p_2}^2 + c^2 F_{p_3}^2) \\ & \quad - (a^2 F_{p_1} d\xi + b^2 F_{p_2} d\eta + c^2 F_{p_3} d\zeta)^2 \\ & = \begin{vmatrix} b d\eta & c d\zeta \\ b F_{p_2} & c F_{p_3} \end{vmatrix}^2 + \begin{vmatrix} c d\zeta & a d\xi \\ c F_{p_3} & a F_{p_1} \end{vmatrix}^2 + \begin{vmatrix} a d\xi & b d\eta \\ a F_{p_1} & b F_{p_2} \end{vmatrix}^2 \\ & = b^2 c^2 d\bar{\xi}^2 + c^2 a^2 d\bar{\eta}^2 + a^2 b^2 d\bar{\zeta}^2. \end{aligned}$$

Wir bringen auf der linken Seite das quadratische Glied zum Verschwinden, indem wir $a^2 F_{p_1}, b^2 F_{p_2}, c^2 F_{p_3}$ bzw. gleich p_1, p_2, p_3 wählen d. h.

$$a^2 = \frac{p_1}{F_{p_1}}, \quad b^2 = \frac{p_2}{F_{p_2}}, \quad c^2 = \frac{p_3}{F_{p_3}}$$

setzen; da bei dieser Wahl von a, b, c

$$a^2 F_{p_1}^2 + b^2 F_{p_2}^2 + c^2 F_{p_3}^2 = p_1 F_{p_1} + p_2 F_{p_2} + p_3 F_{p_3} = F$$

wird, so ergibt die obige Laplacesche Identität

$$F \left(\frac{p_1}{F_{p_1}} d\xi^2 + \frac{p_2}{F_{p_2}} d\eta^2 + \frac{p_3}{F_{p_3}} d\zeta^2 \right) = \frac{p_1 p_2 p_3}{F_{p_1} F_{p_2} F_{p_3}} \left(\frac{F_{p_1}}{p_1} d\bar{\xi}^2 + \frac{F_{p_2}}{p_2} d\bar{\eta}^2 + \frac{F_{p_3}}{p_3} d\bar{\zeta}^2 \right).$$

Nun liefern aber die Relationen (13), (13) in Verbindung mit der Gleichung (14) die Zusammenhänge

$$(13') \quad \frac{p_1}{F_{p_1}} = F^2 \frac{\bar{F}_{\bar{p}_1}}{\bar{p}_1}, \quad \frac{p_2}{F_{p_2}} = F^2 \frac{\bar{F}_{\bar{p}_2}}{\bar{p}_2}, \quad \frac{p_3}{F_{p_3}} = F^2 \frac{\bar{F}_{\bar{p}_3}}{\bar{p}_3}.$$

Wir können daher die rechtsstehende quadratische Form auf die folgende Gestalt bringen

$$F \sqrt{\frac{p_1 p_2 p_3}{F_{p_1} F_{p_2} F_{p_3}}} \sqrt{\frac{\bar{F}_{\bar{p}_1} \bar{F}_{\bar{p}_2} \bar{F}_{\bar{p}_3}}{\bar{p}_1 \bar{p}_2 \bar{p}_3}} \left(\frac{\bar{p}_1}{\bar{F}_{\bar{p}_1}} d\bar{\xi}^2 + \frac{\bar{p}_2}{\bar{F}_{\bar{p}_2}} d\bar{\eta}^2 + \frac{\bar{p}_3}{\bar{F}_{\bar{p}_3}} d\bar{\zeta}^2 \right).$$

Damit gewinnen wir aber die Relation

$$\begin{aligned} \sqrt{\frac{F F_{p_1} F_{p_2} F_{p_3}}{p_1 p_2 p_3}} \left(\frac{p_1}{F_{p_1}} d\xi^2 + \frac{p_2}{F_{p_2}} d\eta^2 + \frac{p_3}{F_{p_3}} d\zeta^2 \right) \\ = \sqrt{\frac{\bar{F} \bar{F}_{\bar{p}_1} \bar{F}_{\bar{p}_2} \bar{F}_{\bar{p}_3}}{\bar{p}_1 \bar{p}_2 \bar{p}_3}} \left(\frac{\bar{p}_1}{\bar{F}_{\bar{p}_1}} d\bar{\xi}^2 + \frac{\bar{p}_2}{\bar{F}_{\bar{p}_2}} d\bar{\eta}^2 + \frac{\bar{p}_3}{\bar{F}_{\bar{p}_3}} d\bar{\zeta}^2 \right), \end{aligned}$$

und diese liefert eben das Gewünschte.

Wir denken uns nämlich auf der ursprünglichen Extremalfläche eine nichteuklidische Maßbestimmung derart eingeführt, daß das Differential der Bogenlänge

$$ds^2 = \sqrt{\frac{F F_{p_1} F_{p_2} F_{p_3}}{p_1 p_2 p_3}} \left(\frac{p_1}{F_{p_1}} d\xi^2 + \frac{p_2}{F_{p_2}} d\eta^2 + \frac{p_3}{F_{p_3}} d\zeta^2 \right)$$

sei und definieren in gleicher Weise auf der adjungierten Fläche als Differential der Bogenlänge

$$d\bar{s}^2 = \sqrt{\frac{\bar{F} \bar{F}_{\bar{p}_1} \bar{F}_{\bar{p}_2} \bar{F}_{\bar{p}_3}}{\bar{p}_1 \bar{p}_2 \bar{p}_3}} \left(\frac{\bar{p}_1}{\bar{F}_{\bar{p}_1}} d\bar{\xi}^2 + \frac{\bar{p}_2}{\bar{F}_{\bar{p}_2}} d\bar{\eta}^2 + \frac{\bar{p}_3}{\bar{F}_{\bar{p}_3}} d\bar{\zeta}^2 \right);$$

diese beiden Definitionen sind völlig *symmetrisch*, indem die Maßbestimmung auf der Extremalfläche in gleicher Weise aus der Gleichung dieser Fläche und aus dem ursprünglichen Variationsproblem aufgebaut ist, wie die Maßbestimmung auf der adjungierten Fläche aus dieser Fläche bzw. aus dem adjungierten Variationsproblem. Wir sind daher zu dem folgenden Satz gelangt:

Bei Zugrundelegung dieser nichteuklidischen Maßbestimmungen ist die Abbildung der Extremalfläche auf die adjungierte Fläche eine Verbiegung.

9. Das Variationsintegral als nichteuklidischer Flächeninhalt. Die beiden Differentialformen ds^2 und $d\bar{s}^2$ sind gleichzeitig reell bzw. rein imaginär. Denn es folgt mit Hilfe der auf Grund der Relationen (13), (13') und (14) leicht verifizierbaren Gleichung

$$\frac{F F_{p_1} F_{p_2} F_{p_3}}{p_1 p_2 p_3} \frac{\bar{F} \bar{F}_{\bar{p}_1} \bar{F}_{\bar{p}_2} \bar{F}_{\bar{p}_3}}{\bar{p}_1 \bar{p}_2 \bar{p}_3} = \frac{1}{F^4} > 0,$$

daß die beiden in ds^2 bzw. $d\bar{s}^2$ auftretenden Wurzeln gleichzeitig reell bzw. rein imaginär sind. Wir wollen im letzteren Falle die obige Definition der Bogenlängen mit i multipliziert als Fundamentalform unserer Maßbestimmungen nehmen, um in jedem Falle reelle Bogenlängen zu erhalten. Dementsprechend setzen wir

$$(15) \quad \begin{aligned} \lambda_1 &= \sqrt{\pm \frac{F F_{p_1} F_{p_2} F_{p_3}}{p_1 p_2 p_3} \frac{p_1}{F_{p_1}}}, & \lambda_2 &= \sqrt{\pm \frac{F F_{p_1} F_{p_2} F_{p_3}}{p_1 p_2 p_3} \frac{p_2}{F_{p_2}}}, & \lambda_3 &= \sqrt{\pm \frac{F F_{p_1} F_{p_2} F_{p_3}}{p_1 p_2 p_3} \frac{p_3}{F_{p_3}}}, \\ \bar{\lambda}_1 &= \sqrt{\pm \frac{\bar{F} \bar{F}_{\bar{p}_1} \bar{F}_{\bar{p}_2} \bar{F}_{\bar{p}_3}}{\bar{p}_1 \bar{p}_2 \bar{p}_3} \frac{\bar{p}_1}{\bar{F}_{\bar{p}_1}}}, & \bar{\lambda}_2 &= \sqrt{\pm \frac{\bar{F} \bar{F}_{\bar{p}_1} \bar{F}_{\bar{p}_2} \bar{F}_{\bar{p}_3}}{\bar{p}_1 \bar{p}_2 \bar{p}_3} \frac{\bar{p}_2}{\bar{F}_{\bar{p}_2}}}, & \bar{\lambda}_3 &= \sqrt{\pm \frac{\bar{F} \bar{F}_{\bar{p}_1} \bar{F}_{\bar{p}_2} \bar{F}_{\bar{p}_3}}{\bar{p}_1 \bar{p}_2 \bar{p}_3} \frac{\bar{p}_3}{\bar{F}_{\bar{p}_3}}}, \end{aligned}$$

wo überall das positive bzw. negative Vorzeichen zu nehmen ist; wir nehmen dasjenige Vorzeichen, das für $\lambda_1, \lambda_2, \lambda_3, \bar{\lambda}_1, \bar{\lambda}_2, \bar{\lambda}_3$ reelle Werte liefert. Dann sind die Größen $\lambda_1, \lambda_2, \lambda_3$ bis auf ihr Vorzeichen eindeutig bestimmt (wir werden darüber später verfügen). Da ferner — wegen (13') —

$$\lambda_1 \bar{\lambda}_1 = \lambda_2 \bar{\lambda}_2 = \lambda_3 \bar{\lambda}_3 = F^2 \sqrt{\frac{1}{F^4}},$$

so wollen wir die Vorzeichen von $\bar{\lambda}_1, \bar{\lambda}_2, \bar{\lambda}_3$ derart bestimmen, daß

$$(16) \quad \lambda_1 \bar{\lambda}_1 = \lambda_2 \bar{\lambda}_2 = \lambda_3 \bar{\lambda}_3 = 1$$

ausfällt.

Schreiben wir nun in Anlehnung an die Flächentheorie

$$(17) \quad \begin{aligned} E^* &= \lambda_1 \xi_u^2 + \lambda_2 \eta_u^2 + \lambda_3 \zeta_u^2, & G^* &= \lambda_1 \xi_v^2 + \lambda_2 \eta_v^2 + \lambda_3 \zeta_v^2, \\ F^* &= \lambda_1 \xi_u \xi_v + \lambda_2 \eta_u \eta_v + \lambda_3 \zeta_u \zeta_v, \end{aligned}$$

und entsprechend für die adjungierte Fläche

$$(17) \quad \begin{aligned} \bar{E}^* &= \bar{\lambda}_1 \bar{\xi}_u^2 + \bar{\lambda}_2 \bar{\eta}_u^2 + \bar{\lambda}_3 \bar{\zeta}_u^2, & \bar{G}^* &= \bar{\lambda}_1 \bar{\xi}_v^2 + \bar{\lambda}_2 \bar{\eta}_v^2 + \bar{\lambda}_3 \bar{\zeta}_v^2, \\ \bar{F}^* &= \bar{\lambda}_1 \bar{\xi}_u \bar{\xi}_v + \bar{\lambda}_2 \bar{\eta}_u \bar{\eta}_v + \bar{\lambda}_3 \bar{\zeta}_u \bar{\zeta}_v, \end{aligned}$$

so kann man die Integrale

$$\iint |\sqrt{E^* G^* - F^{*2}}| du dv \quad \text{bzw.} \quad \iint |\sqrt{\bar{E}^* \bar{G}^* - \bar{F}^{*2}}| du dv$$

bei Zugrundelegung der oben angegebenen Maßbestimmungen als die nichteuklidischen Flächeninhalte der Extremalfläche bzw. der adjungierten

Fläche betrachten. Für die Integranden erhält man einen bemerkenswerten Ausdruck; in der Tat, wendet man die Laplacesche Identität auf die Matrix

$$\begin{pmatrix} \sqrt{\lambda_1} \xi_u & \sqrt{\lambda_2} \eta_u & \sqrt{\lambda_3} \zeta_u \\ \sqrt{\lambda_1} \xi_v & \sqrt{\lambda_2} \eta_v & \sqrt{\lambda_3} \zeta_v \end{pmatrix}$$

an, so ergibt sich

$$|E^* G^* - F^{*2}| = |\lambda_1 \lambda_2 p_3^2 + \lambda_2 \lambda_3 p_1^2 + \lambda_3 \lambda_1 p_2^2| = F^2,$$

und wir erhalten das folgende Resultat

$$\iint |\sqrt{E^* G^* - F^{*2}}| du dv = \iint F du dv,$$

d. h. das vorgelegte Variationsintegral, erstreckt auf einen Teil der Extremalfläche, ist gleich dem nichteuklidischen Flächeninhalt dieses Flächenstückes bei Zugrundelegung der oben angegebenen Maßbestimmung.

Wir bemerken noch, daß man die Größen $\lambda_1, \lambda_2, \lambda_3$ in der folgenden Form schreiben kann

$$\lambda_1 = \sqrt{\pm \frac{p_1}{p_2 p_3} : \frac{\bar{p}_1}{\bar{p}_2 \bar{p}_3}}, \quad \lambda_2 = \sqrt{\pm \frac{p_2}{p_3 p_1} : \frac{\bar{p}_2}{\bar{p}_3 \bar{p}_1}}, \quad \lambda_3 = \sqrt{\pm \frac{p_3}{p_1 p_2} : \frac{\bar{p}_3}{\bar{p}_1 \bar{p}_2}}$$

und entsprechend die Größen $\bar{\lambda}_1, \bar{\lambda}_2, \bar{\lambda}_3$, welche Formeln den symmetrischen Charakter dieser Größen zeigen.

10. Isotherme Parameter. Ein Blick auf die Formeln (15) zeigt, daß die Größen $\lambda_1, \lambda_2, \lambda_3$ homogene Funktionen 0-ten Grades von p_1, p_2, p_3 sind. Dies hat zur Folge — da das Verhältnis der Funktionaldeterminanten p_1, p_2, p_3 bei Einführung neuer Veränderlichen an Stelle der u, v un geändert bleibt — daß $\lambda_1, \lambda_2, \lambda_3$ von der Wahl der Parameter u, v unabhängige Funktionen auf der vorgelegten Extremalfläche bedeuten.

Wir wollen nun statt u, v neue Parameter α, β

$$\alpha = \alpha(u, v), \quad \beta = \beta(u, v)$$

derart einführen, daß die beiden Relationen

$$\begin{aligned} \lambda_1 \xi_\alpha^2 + \lambda_2 \eta_\alpha^2 + \lambda_3 \zeta_\alpha^2 &= \lambda_1 \xi_\beta^2 + \lambda_2 \eta_\beta^2 + \lambda_3 \zeta_\beta^2 \\ \lambda_1 \xi_\alpha \xi_\beta + \lambda_2 \eta_\alpha \eta_\beta + \lambda_3 \zeta_\alpha \zeta_\beta &= 0 \end{aligned}$$

bestehen. Dies ist stets möglich; denn ein klassischer Satz der Theorie der Differentialformen lehrt, daß wenn $\alpha(u, v)$ und $\beta(u, v)$ Lösungen des Differentialgleichungssystems

$$\beta_u = \frac{F^* \alpha_u - E^* \alpha_v}{\sqrt{E^* G^* - F^{*2}}}, \quad \beta_v = \frac{G^* \alpha_u - F^* \alpha_v}{\sqrt{E^* G^* - F^{*2}}}$$

bedeuten, die quadratische Differentialform

$$\lambda_1 d\xi^2 + \lambda_2 d\eta^2 + \lambda_3 d\zeta^2 = E^* du^2 + 2F^* du dv + G^* dv^2$$

in $\theta(\alpha, \beta)(d\alpha^2 + d\beta^2)$

übergeht, wobei $\theta(\alpha, \beta)$ eine Funktion von α, β bedeutet. Führen wir daher solche α, β als neue Parameter ein, so wird

$$E^* = G^*, \quad F^* = 0.$$

Solche Parameter bezeichnen wir als *isotherme Parameter in unserer nichteuklidischen Maßbestimmung* auf der vorgelegten Extremalfäche, die Kurven $\alpha = \text{konst.}$ bzw. $\beta = \text{konst.}$ aber als isotherme Kurven.

In isothermen Parametern (in diesem Sinne) nehmen die Differentialgleichungen unseres Variationsproblems eine besonders einfache Gestalt an, die das Studium dieser Flächen sehr erleichtert. Es handelt sich dabei um das Analogon des bekannten Weierstraßschen Satzes, daß die Gleichungen einer Minimalfläche in isothermen Parametern (im gewöhnlichen Sinne) Potentialfunktionen sind.

Setzen wir zur Abkürzung

$$X = \sqrt{\lambda_1}(\xi_\alpha + i\xi_\beta), \quad Y = \sqrt{\lambda_2}(\eta_\alpha + i\eta_\beta), \quad Z = \sqrt{\lambda_3}(\zeta_\alpha + i\zeta_\beta),$$

so kann man die Gleichungen $E^* = G^*, F^* = 0$, die für die Wahl der Parameter α, β charakteristisch sind, in die eine Gleichung

$$(18) \quad X^2 + Y^2 + Z^2 = 0$$

zusammenfassen. Da ferner — wegen der Invarianz unserer Bogenlänge — gleichzeitig auch $\bar{E}^* = \bar{G}^*, \bar{F}^* = 0$ ist, so gilt, wenn wir die entsprechende Abkürzung

$$\bar{X} = \sqrt{\bar{\lambda}_1}(\bar{\xi}_\alpha + i\bar{\xi}_\beta), \quad \bar{Y} = \sqrt{\bar{\lambda}_2}(\bar{\eta}_\alpha + i\bar{\eta}_\beta), \quad \bar{Z} = \sqrt{\bar{\lambda}_3}(\bar{\zeta}_\alpha + i\bar{\zeta}_\beta)$$

einführen, in gleicher Weise

$$(19) \quad \bar{X}^2 + \bar{Y}^2 + \bar{Z}^2 = 0.$$

Endlich liefert die Gleichung in Nr. 7

$$d\xi d\bar{\xi} + d\eta d\bar{\eta} + d\zeta d\bar{\zeta} = 0$$

in Verbindung mit den Relationen $\lambda_1 \bar{\lambda}_1 = \lambda_2 \bar{\lambda}_2 = \lambda_3 \bar{\lambda}_3 = 1$ die Gleichung:

$$(20) \quad X\bar{X} + Y\bar{Y} + Z\bar{Z} = 0.$$

Aus den Gleichungen (18), (19), (20) folgt aber ohne Schwierigkeit die Proportionalität der Größen X, Y, Z mit den Größen $\bar{X}, \bar{Y}, \bar{Z}$, d. h. es ist — wenn t den Proportionalitätsfaktor bedeutet —

$$\bar{X} = tX, \quad \bar{Y} = tY, \quad \bar{Z} = tZ.$$

Um den Faktor t zu bestimmen, ziehen wir die Gleichungen (II) heran, die wir in der folgenden Form schreiben:

$$\begin{aligned}\frac{\bar{X}}{\sqrt{\lambda_1}} &= t \sqrt{\lambda_1} X = F_{p_2} \frac{Y}{\sqrt{\lambda_2}} - F_{p_2} \frac{Z}{\sqrt{\lambda_3}}, & \frac{\bar{Y}}{\sqrt{\lambda_2}} &= t \sqrt{\lambda_2} Y = F_{p_1} \frac{Z}{\sqrt{\lambda_3}} - F_{p_1} \frac{X}{\sqrt{\lambda_1}}, \\ \frac{\bar{Z}}{\sqrt{\lambda_3}} &= t \sqrt{\lambda_3} Z = F_{p_2} \frac{X}{\sqrt{\lambda_1}} - F_{p_1} \frac{Y}{\sqrt{\lambda_2}}.\end{aligned}$$

Aus diesem homogenen Gleichungssystem in X, Y, Z folgt das Verschwinden der Determinante

$$\begin{vmatrix} -t\sqrt{\lambda_1} & \frac{F_{p_2}}{\sqrt{\lambda_2}} & -\frac{F_{p_2}}{\sqrt{\lambda_3}} \\ -\frac{F_{p_2}}{\sqrt{\lambda_1}} & -t\sqrt{\lambda_2} & \frac{F_{p_1}}{\sqrt{\lambda_3}} \\ \frac{F_{p_2}}{\sqrt{\lambda_1}} & -\frac{F_{p_1}}{\sqrt{\lambda_2}} & -t\sqrt{\lambda_3} \end{vmatrix}$$

$$= -\sqrt{\lambda_1 \lambda_2 \lambda_3} t^3 - \left(\sqrt{\frac{\lambda_1}{\lambda_2 \lambda_3}} F_{p_1}^2 + \sqrt{\frac{\lambda_2}{\lambda_3 \lambda_1}} F_{p_2}^2 + \sqrt{\frac{\lambda_3}{\lambda_1 \lambda_2}} F_{p_3}^2 \right) t.$$

In der soeben erhaltenen Gleichung⁸⁾ für t

$$\lambda_1 \lambda_2 \lambda_3 t^3 + (\lambda_1 F_{p_1}^2 + \lambda_2 F_{p_2}^2 + \lambda_3 F_{p_3}^2) = 0$$

sind aber die beiden Koeffizienten einander gleich, denn man findet ohne Mühe

$$\lambda_1 F_{p_1}^2 + \lambda_2 F_{p_2}^2 + \lambda_3 F_{p_3}^2 = \sqrt{\pm \frac{F_{p_1} F_{p_2} F_{p_3}}{p_1 p_2 p_3}} (p_1 F_{p_1} + p_2 F_{p_2} + p_3 F_{p_3}) = \lambda_1 \lambda_2 \lambda_3.$$

Daher ist

$$t^2 + 1 = 0, \quad \text{d. h. } t = \pm i$$

und

$$\bar{X} = \pm i X, \quad \bar{Y} = \pm i Y, \quad \bar{Z} = \pm i Z.$$

Die wegen (16) damit äquivalenten Gleichungen

$$\begin{aligned}\bar{\xi}_\alpha + i \bar{\xi}_\beta &= \pm i \lambda_1 (\xi_\alpha + i \xi_\beta), & \bar{\eta}_\alpha + i \bar{\eta}_\beta &= \pm i \lambda_2 (\eta_\alpha + i \eta_\beta), \\ \bar{\xi}_\alpha + i \bar{\xi}_\beta &= \pm i \lambda_3 (\zeta_\alpha + i \zeta_\beta)\end{aligned}$$

zeigen zunächst, daß man durch passende Wahl der Vorzeichen von $\lambda_1, \lambda_2, \lambda_3$ erreichen kann — da man über diese noch frei verfügt — daß in diesen Gleichungen das obere Vorzeichen statthabe. Alsdann wird — da $\lambda_1, \lambda_2, \lambda_3$ reell sind —

$$(IV) \quad \begin{aligned}\bar{\xi}_\beta &= \lambda_1 \xi_\alpha, & \bar{\eta}_\beta &= \lambda_2 \eta_\alpha, & \bar{\xi}_\beta &= \lambda_3 \zeta_\alpha, \\ \bar{\xi}_\alpha &= -\lambda_1 \xi_\beta, & \bar{\eta}_\alpha &= -\lambda_2 \eta_\beta, & \bar{\xi}_\alpha &= -\lambda_3 \zeta_\beta.\end{aligned}$$

Wir sind daher zu dem folgenden Analogon des oben genannten Weierstraßschen Satzes gelangt:

⁸⁾ t muß offenbar von Null verschieden sein, da die Annahme $t=0$ auf $\bar{\xi} = \text{konst.}$, $\bar{\eta} = \text{konst.}$, $\bar{\zeta} = \text{konst.}$ führt.

Sind α, β in der zugrunde gelegten nichteuklidischen Maßbestimmung isotherme Parameter, so nehmen die Differentialgleichungen unseres Variationsproblems die einfache Gestalt (IV) an, wobei $\lambda_1, \lambda_2, \lambda_3$ die unter (15) angegebenen Ausdrücke sind.

11. Umkehrung des vorangehenden Satzes. Verstehen wir unter $\lambda_1, \lambda_2, \lambda_3$ die in (15) angegebenen Ausdrücke, wobei $F(p_1, p_2, p_3)$ eine homogene Funktion ersten Grades, p_1, p_2, p_3 aber die Unterdeterminanten der Matrix

$$\begin{pmatrix} \xi_\alpha & \eta_\alpha & \zeta_\alpha \\ \xi_\beta & \eta_\beta & \zeta_\beta \end{pmatrix}$$

bedeuten, und sind diese Funktionen ξ, η, ζ solche Lösungen des Differentialgleichungssystems

$$(IV) \quad \begin{aligned} \bar{\xi}_\beta &= \lambda_1 \xi_\alpha, & \bar{\eta}_\beta &= \lambda_2 \eta_\alpha, & \bar{\zeta}_\beta &= \lambda_3 \zeta_\alpha, \\ \bar{\xi}_\alpha &= -\lambda_1 \xi_\beta, & \bar{\eta}_\alpha &= -\lambda_2 \eta_\beta, & \bar{\zeta}_\alpha &= -\lambda_3 \zeta_\beta, \end{aligned}$$

für die

$$(21) \quad \begin{aligned} \lambda_1 \xi_\alpha^2 + \lambda_2 \eta_\alpha^2 + \lambda_3 \zeta_\alpha^2 &= \lambda_1 \xi_\beta^2 + \lambda_2 \eta_\beta^2 + \lambda_3 \zeta_\beta^2, \\ \lambda_1 \xi_\alpha \xi_\beta + \lambda_2 \eta_\alpha \eta_\beta + \lambda_3 \zeta_\alpha \zeta_\beta &= 0 \end{aligned}$$

ist, so ist $x = \xi(\alpha, \beta)$, $y = \eta(\alpha, \beta)$, $z = \zeta(\alpha, \beta)$ eine Extremalfläche des Variationsproblems $\delta \iint F(p_1, p_2, p_3) d\alpha d\beta = 0$.

Setzt man nämlich für einen Augenblick

$$P_1 = \sqrt{\lambda_2 \lambda_3} \frac{p_1}{F}, \quad P_2 = \sqrt{\lambda_3 \lambda_1} \frac{p_2}{F}, \quad P_3 = \sqrt{\lambda_1 \lambda_2} \frac{p_3}{F},$$

so erkennt man sofort das Bestehen der folgenden Gleichungen

$$\begin{aligned} P_1 \sqrt{\lambda_1} \xi_\alpha + P_2 \sqrt{\lambda_2} \eta_\alpha + P_3 \sqrt{\lambda_3} \zeta_\alpha &= 0, & P_1 \sqrt{\lambda_1} \xi_\beta + P_2 \sqrt{\lambda_2} \eta_\beta + P_3 \sqrt{\lambda_3} \zeta_\beta &= 0, \\ & \pm (P_1^2 + P_2^2 + P_3^2) = 1. \end{aligned}$$

Die vektorielle Bedeutung dieser Gleichungen besteht darin, daß der Vektor vom Betrage Eins ($\varepsilon P_1, \varepsilon P_2, \varepsilon P_3$), wo $\varepsilon = 1$ oder $\varepsilon = i$ ist, senkrecht auf den Vektoren ($\sqrt{\lambda_1} \xi_\alpha, \sqrt{\lambda_2} \eta_\alpha, \sqrt{\lambda_3} \zeta_\alpha$) und ($\sqrt{\lambda_1} \xi_\beta, \sqrt{\lambda_2} \eta_\beta, \sqrt{\lambda_3} \zeta_\beta$) steht, die wegen (21) von gleicher Länge sind und selbst aufeinander senkrecht stehen. Daraus folgt aber, daß das vektorielle Produkt des obigen Einheitsvektors mit jeder dieser letzteren Vektoren bis auf das Vorzeichen mit dem anderen dieser letzteren Vektoren übereinstimmt; d. h.

$$\begin{aligned} \sqrt{\lambda_1} \xi_\alpha &= \pm \varepsilon (P_3 \sqrt{\lambda_2} \eta_\beta - P_2 \sqrt{\lambda_3} \zeta_\beta), & \sqrt{\lambda_1} \xi_\beta &= \mp \varepsilon (P_3 \sqrt{\lambda_2} \eta_\alpha - P_2 \sqrt{\lambda_3} \zeta_\alpha), \\ \sqrt{\lambda_2} \eta_\alpha &= \pm \varepsilon (P_1 \sqrt{\lambda_3} \zeta_\beta - P_3 \sqrt{\lambda_1} \xi_\beta), & \sqrt{\lambda_2} \eta_\beta &= \mp \varepsilon (P_1 \sqrt{\lambda_3} \zeta_\alpha - P_3 \sqrt{\lambda_1} \xi_\alpha), \\ \sqrt{\lambda_3} \zeta_\alpha &= \pm \varepsilon (P_2 \sqrt{\lambda_1} \xi_\beta - P_1 \sqrt{\lambda_2} \eta_\beta), & \sqrt{\lambda_3} \zeta_\beta &= \mp \varepsilon (P_2 \sqrt{\lambda_1} \xi_\alpha - P_1 \sqrt{\lambda_2} \eta_\alpha), \end{aligned}$$

wobei entweder überall das obere oder überall das untere Vorzeichen zu nehmen ist. Beachtet man aber die auf Grund von (15) leicht verifizierbaren Relationen

$$P_1 \sqrt{\lambda_2 \lambda_3} = \pm F_{p_1}, \quad P_2 \sqrt{\lambda_3 \lambda_1} = \pm F_{p_2}, \quad P_3 \sqrt{\lambda_1 \lambda_2} = \pm F_{p_3},$$

so ergeben die letzten Gleichungen durch Multiplikation mit $\sqrt{\lambda_1}$ bzw. $\sqrt{\lambda_2}$, $\sqrt{\lambda_3}$ in Verbindung mit dem zugrunde gelegten Differentialgleichungssystem

$$\begin{aligned} \bar{\xi}_\beta &= \pm \varepsilon (F_{p_2} \eta_\beta - F_{p_2} \zeta_\beta), & \bar{\xi}_\alpha &= \pm \varepsilon (F_{p_2} \eta_\alpha - F_{p_2} \zeta_\alpha), \\ \bar{\eta}_\beta &= \pm \varepsilon (F_{p_1} \zeta_\beta - F_{p_1} \xi_\beta), & \bar{\eta}_\alpha &= \pm \varepsilon (F_{p_1} \zeta_\alpha - F_{p_1} \xi_\alpha), \\ \bar{\zeta}_\beta &= \pm \varepsilon (F_{p_2} \xi_\beta - F_{p_1} \eta_\beta), & \bar{\zeta}_\alpha &= \pm \varepsilon (F_{p_2} \xi_\alpha - F_{p_1} \eta_\alpha). \end{aligned}$$

Diese Gleichungen lehren aber (nach den Ergebnissen in 1.), daß die Funktionen ξ, η, ζ tatsächlich eine Extremalfläche des fraglichen Variationsproblems darstellen. Aus der Realität der auftretenden Funktionen folgt übrigens $\varepsilon = 1$; daher sind $\bar{\xi}, \bar{\eta}, \bar{\zeta}$ die Koordinaten der adjungierten Fläche oder des Spiegelbildes dieser Fläche in bezug auf den Koordinatenanfangspunkt.

12. Analogon eines Darboux'schen Satzes. Bekanntlich bewies Darboux den schönen Satz, daß falls zwei Flächen so ineinander verbiegbar sind, daß die entsprechenden Linienelemente aufeinander senkrecht stehen, so sind diese Flächen adjungierte Minimalflächen. Dieser Satz besitzt in unserem Ideenkreis das folgende Analogon:

Es seien $F(p_1, p_2, p_3)$ und $\bar{F}(\bar{p}_1, \bar{p}_2, \bar{p}_3)$ die Integranden adjungierter Variationsprobleme; wir führen auf den gegebenen Flächen

$$x = \xi(u, v), \quad y = \eta(u, v), \quad z = \zeta(u, v),$$

bzw.

$$x = \bar{\xi}(u, v), \quad y = \bar{\eta}(u, v), \quad z = \bar{\zeta}(u, v)$$

nichteuklidische Maßbestimmungen ein, indem wir die Bogenlängen durch die quadratischen Differentialformen

$$ds^2 = \lambda_1 d\xi^2 + \lambda_2 d\eta^2 + \lambda_3 d\zeta^2$$

bzw.

$$d\bar{s}^2 = \bar{\lambda}_1 d\bar{\xi}^2 + \bar{\lambda}_2 d\bar{\eta}^2 + \bar{\lambda}_3 d\bar{\zeta}^2$$

definieren, wobei $\lambda_1, \lambda_2, \lambda_3$ bzw. $\bar{\lambda}_1, \bar{\lambda}_2, \bar{\lambda}_3$ die in (15) angegebenen Ausdrücke sind ($\lambda_1 \bar{\lambda}_1 = \lambda_2 \bar{\lambda}_2 = \lambda_3 \bar{\lambda}_3 = 1$). Ist die vorgelegte Abbildung zwischen beiden Flächen bei Zugrundelegung dieser Maßbestimmungen eine Verbiegung und sind ferner entsprechende Linienelemente auf beiden Flächen stets orthogonal zueinander, so sind die gegebenen Flächen Extremalflächen der entsprechenden Variationsprobleme, und die eine Fläche ist die adjungierte Fläche der anderen oder deren Spiegelbild in bezug auf den Koordinatenanfangspunkt.

In der Tat, führt man auf der ersten der gegebenen Flächen — wie in 9. — in bezug auf die zugrunde gelegte Maßbestimmung isotherme Parameter α, β ein, so daß

$$E^* = \lambda_1 \xi_\alpha^2 + \lambda_2 \eta_\alpha^2 + \lambda_3 \zeta_\alpha^2 = \lambda_1 \xi_\beta^2 + \lambda_2 \eta_\beta^2 + \lambda_3 \zeta_\beta^2 = G^*,$$

$$F^* = \lambda_1 \xi_\alpha \xi_\beta + \lambda_2 \eta_\alpha \eta_\beta + \lambda_3 \zeta_\alpha \zeta_\beta = 0,$$

so wird infolge unserer Annahmen gleichzeitig auch

$$\bar{E}^* = \bar{\lambda}_1 \bar{\xi}_\alpha^2 + \bar{\lambda}_2 \bar{\eta}_\alpha^2 + \bar{\lambda}_3 \bar{\zeta}_\alpha^2 = \bar{\lambda}_1 \bar{\xi}_\beta^2 + \bar{\lambda}_2 \bar{\eta}_\beta^2 + \bar{\lambda}_3 \bar{\zeta}_\beta^2 = \bar{G}^* = E^* = G^*,$$

$$\bar{F}^* = \bar{\lambda}_1 \bar{\xi}_\alpha \bar{\xi}_\beta + \bar{\lambda}_2 \bar{\eta}_\alpha \bar{\eta}_\beta + \bar{\lambda}_3 \bar{\zeta}_\alpha \bar{\zeta}_\beta = 0.$$

Setzt man wiederum zur Abkürzung

$$\sqrt{\lambda_1}(\xi_\alpha + i\xi_\beta) = X, \quad \sqrt{\lambda_2}(\eta_\alpha + i\eta_\beta) = Y, \quad \sqrt{\lambda_3}(\zeta_\alpha + i\zeta_\beta) = Z,$$

$$\sqrt{\bar{\lambda}_1}(\bar{\xi}_\alpha + i\bar{\xi}_\beta) = \bar{X}, \quad \sqrt{\bar{\lambda}_2}(\bar{\eta}_\alpha + i\bar{\eta}_\beta) = \bar{Y}, \quad \sqrt{\bar{\lambda}_3}(\bar{\zeta}_\alpha + i\bar{\zeta}_\beta) = \bar{Z},$$

so lauten die obigen Gleichungen einfach

$$X^2 + Y^2 + Z^2 = 0, \quad \bar{X}^2 + \bar{Y}^2 + \bar{Z}^2 = 0;$$

außerdem ist — infolge der angenommenen Orthogonalität der entsprechenden Linienelemente —

$$X\bar{X} + Y\bar{Y} + Z\bar{Z} = 0.$$

Daraus folgt wiederum — wie in 9 —

$$(22) \quad \bar{X} = tX, \quad \bar{Y} = tY, \quad \bar{Z} = tZ,$$

wo noch der Proportionalitätsfaktor t zu bestimmen ist. Zunächst folgt aus den letzten Relationen durch Quadrieren der Beträge und Addieren

$$\bar{E}^* + \bar{G}^* = |t|^2 (E^* + G^*).$$

Daher ist

$$|t| = 1, \quad t = \cos \tau + i \sin \tau,$$

wo τ reell ist. Alsdann liefern die Gleichungen (22) mit Rücksicht auf die Realität von $\lambda_1, \lambda_2, \lambda_3$ die Beziehungen

$$\bar{\xi}_\alpha = \lambda_1 (\xi_\alpha \cos \tau - \xi_\beta \sin \tau),$$

$$\bar{\eta}_\alpha = \lambda_2 (\eta_\alpha \cos \tau - \eta_\beta \sin \tau),$$

$$\bar{\zeta}_\alpha = \lambda_3 (\zeta_\alpha \cos \tau - \zeta_\beta \sin \tau).$$

Komponiert man diese Gleichungen mit $\xi_\alpha, \eta_\alpha, \zeta_\alpha$, so ergibt sich links wegen der Orthogonalität der entsprechenden Linienelemente Null, also

$$0 = E^* \cos \tau - F^* \sin \tau = E^* \cos \tau,$$

d. h. es ist $\cos \tau = 0$ also $t = \pm i$, und wir gelangen aus (22) zu dem folgenden Gleichungssystem

$$\bar{\xi}_\beta = \pm \lambda_1 \xi_\alpha, \quad \bar{\eta}_\beta = \pm \lambda_2 \eta_\alpha, \quad \bar{\zeta}_\beta = \pm \lambda_3 \zeta_\alpha,$$

$$\bar{\xi}_\alpha = \mp \lambda_1 \xi_\beta, \quad \bar{\eta}_\alpha = \mp \lambda_2 \eta_\beta, \quad \bar{\zeta}_\alpha = \mp \lambda_3 \zeta_\beta.$$

Da ferner $E^* = G^*$ und $F^* = 0$ ist, so lehrt der in 11. bewiesene Satz, daß die durch die Funktionen $\xi(\alpha, \beta)$, $\eta(\alpha, \beta)$, $\zeta(\alpha, \beta)$ dargestellte Fläche eine Extremalfläche des Variationsproblems

$$\delta \iint F(p_1, p_2, p_3) d\alpha d\beta = 0$$

ist, und die durch $\bar{\xi}$, $\bar{\eta}$, $\bar{\zeta}$ dargestellte Fläche ihre adjungierte Fläche oder deren Spiegelbild in bezug auf den Nullpunkt.

Wir haben im Vorangehenden diejenigen Sätze der Theorie der Minimalflächen auf unser zugrunde gelegtes Variationsproblem übertragen, in denen nur die ersten Ableitungen der Extremalflächen auftreten. Tatsächlich erfordern unsere Sätze nicht die Existenz der zweiten Ableitungen der Extremalfunktion, wie in Anmerkung ⁴⁾ betont wurde. Es wäre aber auch von Interesse, solche Eigenschaften der Minimalflächen zu übertragen, in denen die zweiten Ableitungen auftreten, beispielsweise zu untersuchen, wie die Verallgemeinerung des bekannten Bonnetschen Satzes in unserem Ideenkreis lautet, der die Abbildung der Krümmungskurven der Minimalfläche auf die Asymptotenkurven der adjungierten Fläche aussagt.

(Eingegangen am 14. 12. 1927.)

Konvergenz und Fehlerschätzung beim Ritzschen Verfahren.

Von

E. Treffitz in Dresden.

1. Einleitung.

Wenn man die direkten Methoden der Variationsrechnung in der Form des Ritzschen Verfahrens zur numerischen Lösung von Randwertaufgaben partieller Differentialgleichungen anwendet, so begegnet man dem Übelstande, daß eine befriedigende Fehlerschätzung so gut wie unmöglich ist. Der Ausgangspunkt der folgenden Betrachtungen war der Wunsch, hier einen Fortschritt zu erzielen. Die Betrachtungen sind im wesentlichen Übertragungen von Gedanken, die bei den Minimalaufgaben für endlich viele Parameter, d. h. in der Statik der statisch unbestimmten Fachwerke, seit Maxwell bekannt sind. Den Zusammenhang im einzelnen aufzuweisen, will ich mir der Kürze wegen versagen; ebenso will ich nur an dieser Stelle auf die für die direkten Methoden der Variationsrechnung wichtigen Arbeiten von Courant hinweisen, die ich für die Ergebnisse der letzten Abschnitte benutzt habe. [Über direkte Methoden der Variationsrechnung und über verwandte Fragen, *Math. Annalen* **97** (1927), S. 711; Über ein konvergenzerzeugendes Prinzip in der Variationsrechnung, *Gött. Nachr.* 1922; Über die Eigenwerte bei den Differentialgleichungen der mathematischen Physik, *Math. Zeitschr.* **7** (1920), S. 1.]

Meine Ergebnisse beziehen sich zunächst auf die Lösung der ersten Randwertaufgabe der linearen partiellen Differentialgleichungen vom elliptischen Typus, für welche bei den Gleichungen zweiter Ordnung die Randwerte der gesuchten Funktion, bei denen vierter Ordnung die Randwerte der Funktion und ihrer Normalableitung vorgeschrieben sind. Für eine Reihe praktisch wichtiger Fälle ist auch eine Ausdehnung auf gemischte Randwertaufgaben möglich.

Um die Darstellung zu vereinfachen, rede ich hauptsächlich von den Differentialgleichungen $\Delta u = 0$ und $\Delta \Delta u = 0$; die Übertragung auf Differentialgleichungen von mehr als zwei unabhängigen Variablen und von allgemeinerer Form bedarf keiner Erläuterung.

Das Verfahren von Ritz. Wir betrachten Aufgaben der folgenden Art: Es sei in einer xy -Ebene ein einfach oder mehrfach zusammenhängendes Gebiet Ω gegeben, das von stückweise analytischen doppel-punktslosen Randkurven R begrenzt ist. Gesucht wird die Funktion $u(x, y)$, welche einer linearen partiellen Differentialgleichung $L(u) = 0$ im Innern von Ω genügt, auf dem Rande R vorgeschriebene Randbedingungen erfüllt, und außerdem gewissen Stetigkeits- und Differenzierbarkeitsbedingungen genügt.

Diese Randwertaufgabe sei mit dem Variationsproblem identisch: Unter allen Funktionen einer gegebenen Klasse K diejenige zu suchen, für welche ein „Minimalausdruck“ $D(u)$ ein Minimum wird. Dabei gehört eine Funktion „zur Klasse K “, wenn sie den geforderten Stetigkeits- und Differenzierbarkeitsbedingungen genügt, und gewisse Randbedingungen erfüllt. (NB. Die Randbedingungen für die Differentialgleichung und für das Variationsproblem brauchen nicht notwendig die gleichen zu sein.)

Beim Ritzschen Verfahren konstruiert man Näherungslösungen für dieses Variationsproblem von der Form:

$$(1) \quad v_n(x, y) = q_0(x, y) + \sum_1^n c_\rho q_\rho(x, y).$$

Die Funktionen $q_\rho(x, y)$ werden dabei so gewählt, daß die für das Variationsproblem vorgeschriebenen Randbedingungen bei beliebigen Koeffizienten c_ρ erfüllt sind (z. B. bei gegebenen Randwerten so, daß $q_0(x, y)$ die gegebenen Randwerte annimmt, und daß alle übrigen $q_\rho(x, y)$ ($\rho \geq 1$) auf dem Rande verschwinden.) Außerdem müssen die Funktionen $q_\rho(x, y)$ stetig, hinreichend oft differenzierbar und von solcher Allgemeinheit sein, daß jede konkurrenzfähige Funktion mit den erforderlichen Differentialquotienten beliebig genau approximiert werden kann.

Die Koeffizienten c_ρ in dem Ansatz (1) werden dann so bestimmt, daß der Minimalausdruck $D(v_n)$ möglichst klein wird. Man erhält auf diese Weise eine Folge von Näherungsfunktionen, von der zu zeigen ist, daß sie mit wachsendem n gegen die Lösung $u(x, y)$ des Variationsproblems konvergiert.

Wir beschäftigen uns im folgenden mit der Frage der Konvergenz und der Fehlerschätzung für dieses Verfahren. Der Einfachheit wegen setzen wir die Existenz einer Lösung für das gegebene Variationsproblem voraus; es würde nicht schwierig sein, die Betrachtungen zu einem Exi-

stenzbeweis zu erweitern, es würde aber nichts wesentlich Neues dabei herauskommen, höchstens eine Vereinfachung bekannter Beweise.

Zur Abkürzung will ich noch einige Bezeichnungen einführen. Für die Anwendung des Ritzschen Verfahrens denken wir uns das System der approximierenden Funktionen fest gegeben. Eine Funktion der Form

$q_0(x, y) + \sum_1^n c_e q_e(x, y)$, die mit beliebigen Koeffizienten c_e gebildet ist,

werde eine „Funktion der Klasse K_n “ genannt. Durch die Gleichung

$$(2) \quad \lim_{n \rightarrow \infty} K_n = K$$

wollen wir ausdrücken, daß irgendeine konkurrenzfähige Funktion bei hinreichend hohem n mit beliebiger Genauigkeit angenähert werden kann, insbesondere auch, daß die untere Grenze des Minimalausdruckes für die Funktionen der Klasse K_n mit wachsendem n gegen die untere Grenze des Minimalausdruckes aller konkurrenzfähigen Funktionen konvergiert. Das Ritzsche Verfahren liefert dann für jedes n eine strenge Lösung des Variationsproblems, wenn der Bereich der zulässigen Funktionen auf die Klasse K_n beschränkt wird.

Was nun die Konvergenz des Verfahrens angeht, so steht zunächst nur fest, daß der Minimalausdruck selbst gegen seine untere Grenze konvergiert, wenn $\lim K_n = K$ ist. Um auch die Konvergenz der Funktionswerte selbst zu untersuchen, werde ich jeden einzelnen Wert der Lösung an irgendeiner Stelle ξ, η durch die Minimalwerte (d. h. untere Grenze eines Minimalausdruckes) für gewisse abgeänderte Variationsprobleme — die sogenannten „Nebenprobleme“ — darstellen. Daraus wird sich als Ergebnis des ersten Abschnittes ein Konvergenzkriterium ergeben. Ich zeige dann noch, wie für die gegebenen Variationsprobleme der Fehler des Minimalwertes (= Abweichung des Minimalausdruckes der n -ten Näherungslösung von dem wahren unteren Grenzwert) abgeschätzt werden kann. — Stellt man dann die Funktionswerte durch die Minimalwerte geeigneter „Nebenprobleme“ dar und schätzt die letzteren ab, so ist die Fehlerabschätzung für die Funktionswerte geleistet.

2. Der Grundgedanke des Konvergenzbeweises.

Mechanische Betrachtung. — Wenn man eine am Rande eingespannte Platte durch Kräfte senkrecht zu ihrer Ebene auf Biegung beansprucht, so entspricht jeder Durchbiegung $u(x, y)$ eine bestimmte Verzerrungsenergie A (Formänderungsarbeit). Die Durchbiegung $u(x, y)$ der Gleichgewichtslage ist dadurch ausgezeichnet, daß bei einer Variation (virtuellen Verschiebung), welche die Einspannungsbedingungen nicht verletzt, die

Erhöhung der Verzerrungsenergie gerade von der Arbeit der äußeren Kräfte bestritten wird, so daß in Summa keine Arbeit erforderlich ist, um aus der Gleichgewichtslage heraus eine virtuelle Verschiebung vorzunehmen. Ist die Belastung pro Flächeneinheit der Platte $p(x, y)$, so ist ihre Arbeit bei der Verschiebung δu gleich $\iint p \delta u dx dy$, so daß also für die Gleichgewichtslage bei jedem δu

$$\delta A = \iint p \delta u dx dy$$

oder

$$(3) \quad A - \iint p u dx dy = \text{Minimum}$$

werden muß. Mit den gegebenen Randbedingungen (Einspannung) legt dieses Variationsproblem die Gleichgewichtslage fest. Der zum Minimum zu machende Ausdruck wird die potentielle Energie des Systems genannt. Fragen wir nun nach dem Wert der potentiellen Energie im Gleichgewichtszustande, so bemerken wir (durch partielle Integrationen oder gleichwertige mechanische Betrachtungen), daß im Gleichgewichtsfalle die sogenannte „Arbeit der äußeren Kräfte“ $\iint p u dx dy$ gerade doppelt so groß ist wie die Formänderungsarbeit, so daß der Minimalwert der potentiellen Energie beim Gleichgewicht bis auf das Vorzeichen mit der Formänderungsarbeit übereinstimmt.

Um nun bei einer gegebenen Belastung $p(x, y)$ die Durchbiegung an einer bestimmten Stelle ξ, η zu untersuchen, rufen wir eine besondere virtuelle Verschiebung dadurch hervor, daß wir am Punkte ξ, η eine zusätzliche Einzelkraft von der Größe ε anbringen. Die neue Gleichgewichtslage $v(x, y)$ ist dann durch das Minimum des Ausdrucks

$$(4) \quad A - \iint p v dx dy - \varepsilon v(\xi, \eta)$$

bestimmt; es ist einfach zur Arbeit der äußeren Kräfte der von der Zusatzkraft herrührende Teil $\varepsilon v(\xi, \eta)$ hinzugekommen. Wir können nun den Minimalwert der potentiellen Energie für das dergestalt modifizierte Problem in einer einfachen Weise ausdrücken. Da die Differentialgleichung linear ist, so superponiert sich die von der Zusatzkraft hervorgerufene Zusatzdurchbiegung einfach über die von der Belastung $p(x, y)$ herrührende ursprüngliche Durchbiegung. Die gesamte Formänderungsarbeit können wir dann in folgender Weise berechnen. Wir bringen zunächst die Zusatzkraft ε auf — es kommt auf die Reihenfolge der Belastung ja nicht an —; sie ruft eine Durchbiegung $\varepsilon G(x, y, \xi, \eta)$ hervor, und leistet bei allmählichem Aufbringen die Arbeit $\frac{\varepsilon^2}{2} G(\xi, \eta, \xi, \eta)$. $G(x, y, \xi, \eta)$ ist dabei die Durchbiegung, die eine im Punkte ξ, η wirkende Kraft von der Größe eins hervorrufen würde, oder mathematisch gesprochen, die Greensche Funktion unseres Randwertproblems. Bringen wir jetzt die gegebene Belastung

$p(x, y)$ auf, so ruft sie die eigentlich gesuchte Durchbiegung $u(x, y)$ hervor, und leistet dabei selbst die Arbeit $\frac{1}{2} \iint p u dx dy$, die genau so groß ist, wie wenn die Zusatzkraft nicht vorhanden wäre. Außerdem leistet aber die vorher angebrachte Zusatzkraft ε noch die Arbeit $\varepsilon u(\xi, \eta)$, weil ihr Angriffspunkt sich um das Stück $u(\xi, \eta)$ senkt. Es wird somit die gesamte Formänderungsarbeit

$$(5) \quad A = \frac{\varepsilon^2}{2} G(\xi \eta, \xi \eta) + \varepsilon u(\xi, \eta) + \frac{1}{2} \iint p u dx dy,$$

welche bis auf das Vorzeichen den Minimalwert der potentiellen Energie angibt.

Wir können nun den gesuchten Funktionswert $u(\xi, \eta)$ durch solche Minimalwerte darstellen, wenn wir einmal die Kraft $+\varepsilon$, dann die Kraft $-\varepsilon$ als Zusatzbelastung anbringen, also einmal belasten, und dann entlasten. Die Formänderungsarbeit wird dann

$$\text{bei Belastung} \quad A(+\varepsilon) = \frac{\varepsilon^2}{2} G(\xi \eta, \xi \eta) + \varepsilon u(\xi, \eta) + \frac{1}{2} \iint p u dx dy,$$

$$\text{bei Entlastung} \quad A(-\varepsilon) = \frac{\varepsilon^2}{2} G(\xi \eta, \xi \eta) - \varepsilon u(\xi, \eta) + \frac{1}{2} \iint p u dx dy.$$

Der gesuchte Funktionswert

$$u(\xi, \eta) = \frac{1}{2\varepsilon} \{A(+\varepsilon) - A(-\varepsilon)\}$$

erscheint somit durch die Minimalwerte der „Nebenprobleme“ ausgedrückt, wobei wir unter dem „Nebenproblem“ das Variationsproblem verstehen, welches die Gleichgewichtslage unter Wirkung der ursprünglichen Last $p(x, y)$ und der Zusatzlasten $+\varepsilon$ bzw. $-\varepsilon$ charakterisiert.

3. Mathematische Formulierung des mechanischen Gedankens.

Um die Tragweite des eben entwickelten Gedankens übersehen zu können, will ich ihn in rein mathematischer Form darstellen. Dabei will ich den Gültigkeitsbereich nicht im einzelnen abgrenzen, da ich die Betrachtungen weiterhin vor allem als heuristisches Hilfsmittel benutze. In den hier interessierenden Fällen kann ihre Gültigkeit ohne Schwierigkeit durch unmittelbare Rechnung bestätigt werden.

Es sei die Aufgabe gegeben, aus einer durch Randbedingungen oder dergleichen festgelegten Klasse K von Funktionen die Funktion $u(x, y)$ herauszusuchen, welche einen Ausdruck $D(u)$ zum Minimum macht. Die Existenz der Lösung werde vorausgesetzt. Die klassische Methode der Variationsrechnung besteht dann darin, daß man neben der Lösung u eine inparametrische Schar $v(x, y, \varepsilon)$ von Funktionen der Klasse K betrachtet,

die so beschaffen ist, daß

$$(7) \quad \text{für } \varepsilon = 0 \quad v(x, y, 0) = u(x, y)$$

ist. Vergleicht man nun den Wert des Minimalausdruckes für die Funktionen v mit dem von der wahren Lösung u erreichten Minimalwert $D(u)$, so lautet die Minimalbedingung:

$$(8) \quad \text{für } \varepsilon = 0 \quad \frac{\partial D}{\partial \varepsilon} = 0.$$

Man erhält die zu dem Variationsproblem gehörige Differentialgleichung, wenn man die Funktionen v in der Form

$$v = u + \varepsilon \delta u$$

ansetzt. In den mechanischen Anwendungen, wo u eine Verschiebungsgröße ist, kommt das darauf hinaus, daß man eine Schar von virtuellen Verrückungen betrachtet, die sämtlich einer gegebenen Funktion proportional sind. Hier wollen wir die Vergleichsfunktionen dadurch erzeugen, daß wir an irgendeinem Punkte ξ, η eine Einzellast ε anbringen, durch welche die Gleichgewichtsverschiebung variiert wird. Das stellt sich mathematisch in folgender Form dar:

Wird von der Lösung unseres Variationsproblems der Ausdruck $D(u)$ zum Minimum gemacht, und wollen wir den Funktionswert $u(\xi, \eta)$ an einer bestimmten Stelle ξ, η berechnen, so betrachten wir außer dem gegebenen Variationsproblem das „Nebenproblem“, unter den Funktionen der Klasse K diejenige zu bestimmen, welche den Ausdruck

$$(9) \quad N(v) = D(v) + \varepsilon v(\xi, \eta)$$

zum Minimum macht. Die Lösungen des Nebenproblems $v(x, y, \varepsilon)$, die noch von dem Parameter ε abhängen, seien die Vergleichsfunktionen für das ursprüngliche Variationsproblem. Die Bedingung, daß die Vergleichsfunktion v für den Wert $\varepsilon = 0$ mit u übereinstimmen soll, ist erfüllt. Nun ist

$$(10) \quad \frac{\partial N}{\partial \varepsilon} = \frac{\partial D}{\partial \varepsilon} + \varepsilon \frac{\partial v(\xi, \eta)}{\partial \varepsilon} + v(\xi, \eta).$$

Wegen der Minimaleigenschaft von u wird für $\varepsilon = 0$, $v = u$, $\frac{\partial D}{\partial \varepsilon} = 0$

$$(11) \quad \text{also } \varepsilon = 0, \quad \frac{\partial N}{\partial \varepsilon} = u(\xi, \eta).$$

Wir erhalten also den gesuchten Funktionswert von u , indem wir den Minimalwert des Nebenproblems an der Stelle $\varepsilon = 0$ nach ε differenzieren. Voraussetzung ist natürlich, daß das Nebenproblem auch wirklich eine Lösung hat und daß diese zur Klasse K der für das ursprüngliche Problem zugelassenen Funktionen gehört.

Wir richten unser Augenmerk vor allem auf diejenigen Fälle, in welchen der Minimalausdruck einen quadratischen Differentialausdruck enthält, und demgemäß die Differentialgleichungen linear sind. Für diese setzt sich die Lösung $v(x, y, \varepsilon)$ des Nebenproblems aus der Lösung u des ursprünglichen Problems und der durch einen passenden Faktor normierten Greenschen Funktion $G(xy, \xi \eta)$ in der Form

$$(12) \quad v(x, y, \varepsilon) = u(x, y) - \varepsilon G(xy, \xi \eta)$$

zusammen. Der Minimalausdruck N wird infolgedessen eine quadratische Funktion von ε :

$$(13) \quad N(\varepsilon) = D(u) + \varepsilon u(\xi, \eta) - \frac{\varepsilon^2}{2} G(\xi \eta, \xi \eta),$$

so daß wir den Differentialquotienten $\frac{\partial N}{\partial \varepsilon}$ durch den Differenzenquotienten

$$u(\xi, \eta) = \frac{1}{2\varepsilon} \{N(\varepsilon) - N(-\varepsilon)\}$$

ersetzen können.

Damit erhalten wir zur Bestimmung des Funktionswertes $u(x, y)$ folgende Regel: Wir betrachten außer dem Hauptproblem

$$D(u) = \text{Min.}$$

die beiden Nebenprobleme

$$N(v) = D(v) + v(\xi, \eta) = \text{Min.}$$

und

$$N(v) = D(v) - v(\xi, \eta) = \text{Min.}$$

Sind die Minimalwerte der Nebenprobleme $N(+1)$ und $N(-1)$, so ist

$$(14) \quad u(\xi, \eta) = \frac{1}{2} \{N(+1) - N(-1)\}.$$

Der Vorteil dieser Darstellung, wo der Funktionswert durch die Minimalwerte der Nebenprobleme ausgedrückt erscheint, ist vor allem, daß die Frage nach der Konvergenz der Ritzschen Minimalfolge damit auf die viel einfachere Frage nach der Konvergenz der Minimalwerte selbst zurückgeführt werden kann. An dem Beispiel der Gleichung $\Delta \Delta u = 0$ wollen wir jetzt zeigen, daß die Lösbarkeit des Nebenproblems gesichert ist, wenn die Greensche Funktion im singulären Punkte endlich bleibt, und daß dies auch für die Konvergenz des Ritzschen Verfahrens hinreicht. Der allgemeine Beweis erfolgt nur mit mehr Schreibearbeit so genau analog, daß sich seine besondere Darstellung erübrigt.

4. Variationsprobleme mit lösbarem Nebenproblem. $\Delta \Delta u = 0$.

Als Beispiel einer Randwertaufgabe, für welche das zugehörige Variationsproblem ein lösbares Nebenproblem besitzt, betrachten wir die Rand-

wertaufgabe der Gleichung

$$(15) \quad \Delta \Delta u = 0.$$

Wir suchen eine Funktion, welche im Innern eines gegebenen Gebietes Ω dieser Differentialgleichung genügt und auf der Randkurve die vorgeschriebenen Randwerte

$$(16) \quad u = f(s), \quad \frac{\partial u}{\partial \nu} = g(s)$$

annimmt. Das zugehörige Variationsproblem verlangt, bei den gleichen Randwerten das Integral

$$(17) \quad \frac{1}{2} \iint \Delta u^2 dx dy$$

zum Minimum zu machen. Den Punkt, für den wir den Funktionswert bestimmen wollen, wählen wir zum Koordinatenanfang, und bezeichnen die Funktionswerte an diesem Punkte einfach durch den Index 0. Die zu diesem Punkte als singulärem gehörige Greensche Funktion sei $G(x, y)$; sie genügt als Funktion von x und y der Differentialgleichung $\Delta \Delta G = 0$ und hat im Nullpunkte eine Singularität

$$(18) \quad G = \frac{1}{8\pi} r^2 \ln r + \text{reg. F.}$$

und die Randwerte

$$(19) \quad G = \frac{\partial G}{\partial \nu} = 0.$$

Zu dem Variationsproblem gehört das Nebenproblem: bei den gleichen Randwerten wie oben den Ausdruck

$$(20) \quad N = \frac{1}{2} \iint \Delta v^2 dx dy + \varepsilon v_0$$

zum Minimum zu machen.

Wir zeigen nun zunächst, daß die Lösung des Nebenproblems

$$(21) \quad v(x, y) = u(x, y) - \varepsilon G(x, y)$$

ist, wo $u(x, y)$ die Lösung der ursprünglichen Randwertaufgabe ist. Das rechnen wir einfach aus, indem wir in üblicher Weise den Nullpunkt durch einen Kreis von verschwindendem Radius ausschließen. Betrachten wir außer der Lösung v irgendeine andere Funktion

$$(22) \quad w = v + \delta v,$$

welche die gleichen Randbedingungen erfüllt, so daß am Rande

$$(23) \quad \delta v = \frac{\partial \delta v}{\partial \nu} = 0$$

ist, welche überall stetige Differentialquotienten besitzt, und deren zweite

Differentialquotienten höchstens so unendlich werden, daß das Integral $\iint \{\Delta \delta v\}^2 dx dy$ existiert, so ist:

$$(24) \quad \frac{1}{2} \iint \Delta w^2 dx dy = \frac{1}{2} \iint \Delta v^2 dx dy + \iint \Delta v \Delta \delta v dx dy \\ + \frac{1}{2} \iint \Delta \delta v^2 dx dy.$$

Mit der bekannten partiellen Integration wird das mittlere Integral umgeformt

$$(25) \quad \iint \Delta v \Delta \delta v dx dy = \iint \delta v \Delta \Delta v dx dy + \int \Delta v \frac{\partial \delta v}{\partial \nu} ds - \int \delta v \frac{\partial \Delta v}{\partial \nu} ds.$$

Das Flächenintegral fällt wegen $\Delta \Delta v = 0$ fort. Die Randintegrale liefern wegen (23) keinen Beitrag, soweit sie vom eigentlichen Rande herrühren. Der kleine Kreis um den Nullpunkt liefert wegen der Singularität der Greenschen Funktion den Beitrag

$$- \delta v_0 \int \frac{\partial \Delta v}{\partial \nu} ds = - \varepsilon \delta v_0 = - \varepsilon (w_0 - v_0).$$

Setzen wir dies in (24) ein, so ergibt sich

$$(26) \quad \frac{1}{2} \iint \Delta w^2 dx dy + \varepsilon w_0 = \frac{1}{2} \iint \Delta v^2 dx dy + \varepsilon v_0 + \frac{1}{2} \iint \Delta \delta v^2 dx dy.$$

D. h. für die Funktion w ist der Minimalausdruck

$$N = \frac{1}{2} \iint \Delta w^2 dx dy + \varepsilon w_0$$

tatsächlich größer als für v . Da w eine beliebige konkurrenzfähige Funktion ist, ist also gezeigt, daß v das Nebenproblem tatsächlich löst. Berechnen wir den Wert des Minimalausdrucks für die Lösung v , so erhalten wir durch die gleichen partiellen Integrationen

$$(27) \quad N(\varepsilon) = \frac{1}{2} \iint \Delta u^2 dx dy + \varepsilon u_0 - \frac{\varepsilon^2}{2} G_0.$$

Damit das Nebenproblem wirklich eine Lösung hat, d. h. der Minimalausdruck eine endliche untere Grenze, ist offenbar erforderlich, daß die Greensche Funktion im singulären Punkte endlich bleibt.

In Übereinstimmung mit den Ergebnissen der früheren Abschnitte erhalten wir zur Ermittlung des gesuchten Funktionswertes

$$(28) \quad u(0) = \frac{1}{2} \{N(+1) - N(-1)\},$$

d. h. wir bilden einmal für $\varepsilon = +1$, dann für $\varepsilon = -1$ die Minimalwerte der Nebenprobleme und bekommen den Funktionswert in der halben Differenz.

5. Konvergenz des Ritzschen Verfahrens.

Wir bleiben bei dem Beispiel der Randwertaufgabe für $\Delta u = 0$, mit den gegebenen Randwerten $u = f(s)$, $\frac{\partial u}{\partial \nu} = g(s)$ und dem zugehörigen Minimalproblem $\frac{1}{2} \iint \Delta u^2 dx dy = \text{Min}$. Gegeben sei ein hinreichend allgemeines Funktionensystem $q_0(x, y)$, $q_1(x, y)$, $q_2(x, y)$ usw., wo $q_0(x, y)$ mit seiner Ableitung die vorgeschriebenen Randwerte annimmt und die übrigen $q(x, y)$ mit ihren Normalableitungen auf dem Rande verschwinden, so daß jede Funktion der Klasse K_n d. h. jede Funktion der Form

$$(29) \quad v_n(x, y) = q_0(x, y) + \sum_1^n c_\rho q_\rho(x, y)$$

die Randbedingungen erfüllt, also zur Klasse K der konkurrenzfähigen Funktionen gehört. Das Ritzsche Verfahren konstruiert die n -te Näherungslösung in der Weise, daß innerhalb der Klasse K_n das Variationsproblem streng gelöst wird; d. h. die Koeffizienten werden so bestimmt, daß das Integral $\iint \Delta v_n^2 dx dy$ möglichst klein wird.

Nun setzen die Betrachtungen des dritten Abschnitts nichts über die zulässigen Funktionen voraus, sie sind also unmittelbar auf unseren Fall anwendbar; auch alle Existenzfragen fallen fort, da es sich hier nur um eine Minimumsaufgabe endlich vieler Parameter handelt.

So erhalten wir, wie man auch durch direkte Rechnung ohne weiteres bestätigen kann, das folgende Resultat:

Bezeichnen wir die Minimalwerte des Nebenproblems mit $N_n(+1)$ und $N_n(-1)$, wenn wir nur die Ritzschen Näherungsfunktionen bis zum Index n , also die Klasse K_n zulassen, so wird der n -te Näherungswert

$$(30) \quad v_n(\xi, \eta) = \frac{1}{2} \{N_n(+1) - N_n(-1)\}.$$

Andererseits haben wir im vorigen Abschnitt bewiesen, daß

$$(31) \quad u(\xi, \eta) = \frac{1}{2} \{N(+1) - N(-1)\}$$

ist, wo $N(+1)$ und $N(-1)$ die Minimalwerte der Nebenprobleme sind, wenn alle zulässigen Funktionen zum Vergleich herangezogen werden. Aus diesen beiden Darstellungen folgt die Konvergenz des Ritzschen Verfahrens sofort. Bezüglich der approximierenden Funktionen ist nämlich vorauszusetzen, daß

$$\lim_{n \rightarrow \infty} K_n = K,$$

also

$$(32) \quad \lim_{n \rightarrow \infty} N_n = N \text{ ist.}$$

Haben die Funktionen q diesen Grad der Vollständigkeit, so ist

$$(33) \quad \lim_{n \rightarrow \infty} v_n(\xi, \eta) = u(\xi, \eta).$$

Das Ritzsche Verfahren konvergiert also. Wir können sogar zeigen, daß es gleichmäßig konvergiert. Die Minimalwerte der Nebenprobleme sind nämlich als Funktion von ξ und η stetige Funktionen, die durch die ebenfalls stetigen Funktionen $N_n(+1)$ und $N_n(-1)$ monoton approximiert werden. Nun ist aber nach einem Satze von Dini eine Folge von stetigen Funktionen gleichmäßig konvergent, wenn sie eine stetige Funktion monoton approximiert. (Ich verdanke Herrn Kowalewski den Hinweis darauf, daß sein Beweis dieses Satzes auch für mehrere Veränderliche gültig bleibt.) Damit ist die gleichmäßige Konvergenz des Ritzschen Verfahrens bewiesen.

Ich bemerke noch, daß unsere Betrachtungen auch für gemischte Randwertaufgaben und ihre Variationsprobleme unverändert ihre Gültigkeit behalten. Nur tritt im Minimalausdruck im allgemeinen noch ein Randintegral auf, und mit den Randbedingungen für die gesuchte Funktion ändern sich auch diejenigen für die Greensche Funktion. Es gilt ganz allgemein der folgende Satz:

Das Ritzsche Verfahren konvergiert bei hinreichender Vollständigkeit der approximierenden Funktionen immer dann gleichmäßig, wenn die Greensche Funktion im singulären Punkte $G(\xi, \eta, \xi, \eta)$ endlich bleibt und eine stetige Funktion von ξ und η ist.

6. Variationsprobleme mit nicht lösbarem Nebenproblem. $\Delta u = 0$.

Wenn man die oben für die Gleichung $\Delta \Delta u = 0$ angestellten Betrachtungen auf den klassischen Fall der Gleichung $\Delta u = 0$ übertragen will, so stößt man auf die Schwierigkeit, daß die Greensche Funktion im singulären Punkte logarithmisch unendlich wird, d. h. daß das Nebenproblem keine Lösung mehr hat. Man verdankt Courant einen Vorschlag, der diese Schwierigkeit behebt. Anstatt nämlich in der klassischen Weise bei gegebenen Randwerten das Dirichletsche Integral

$$(34) \quad D(u) = \frac{1}{2} \iint \text{grad}^2 u \, dx \, dy$$

zum Minimum zu machen, sucht man diejenige Funktion, welche bei den gleichen Randbedingungen den Ausdruck

$$(35) \quad D(u) + \lambda \iint \Delta u^2 \, dx \, dy$$

zum Minimum macht. Da die Lösung des ersten Problems $\Delta u = 0$ ergibt, so muß dieselbe auch das erweiterte Problem bei beliebigem λ lösen. In der erweiterten Gestalt entspricht dem Variationsproblem aber eine Diffe-

rentialgleichung vierter Ordnung vom Typus der Gleichung $\Delta \Delta u = 0$, für welche die Greensche Funktion im singulären Punkte endlich bleibt, so daß die Konvergenz des Ritzschen Verfahrens gesichert ist. — Die Methode hat den Nachteil, in der numerischen Rechnung komplizierter zu werden. Es sei deshalb hier auf eine Methode hingewiesen, die im Grunde genommen auf die Courantschen „Glättungsverfahren“ hinausläuft, und die es gestattet, die Funktionswerte auch dann noch zu bestimmen, wenn das Ritzsche Verfahren nicht konvergieren sollte. Ich zeige zu diesem Zwecke, daß beim Ritzschen Verfahren jedenfalls das über ein beliebiges Kurvenstück C genommene Integral $\int_C v_n ds$ konvergiert. Nimmt man einen Kreis, so erhält man den Mittelwert auf demselben, also den Funktionswert im Mittelpunkt. Um die Konvergenz nachzuweisen, betrachtet man — stets bei den gegebenen Randbedingungen — neben dem eigentlichen Variationsproblem:

$$(36) \quad D(u) = \frac{1}{2} \iint \text{grad}^2 u \, dx \, dy = \text{Min.}$$

das Nebenproblem

$$(37) \quad N(v) = D(v) + \varepsilon \int_C v \, ds = \text{Min.}$$

Ist $N(\varepsilon)$ für ein gegebenes ε der Minimalwert des Nebenproblems, so zeigt man genau, wie es im dritten Abschnitt für den Funktionswert gezeigt wurde, daß

$$(38) \quad \int_C u \, ds = \frac{\partial N}{\partial \varepsilon} \quad \text{bei } \varepsilon = 0$$

ist, oder da es sich wieder um lineare Probleme handelt, so daß N eine quadratische Funktion in ε wird,

$$(39) \quad \int_C u \, ds = \frac{1}{2} \{N(+1) - N(-1)\}.$$

Da hier nichts spezielles über die Klasse der zugelassenen Funktionen ausgesagt ist, gilt diese Formel auch für den eingeschränkten Funktionsbereich der Klasse K_n , d. h. es ist für die n -te Ritzsche Näherungslösung

$$(40) \quad \int_C v_n \, ds = \frac{1}{2} \{N_n(+1) - N_n(-1)\}.$$

Nun hat in diesem Falle das Nebenproblem eine Lösung; ich will das nicht beweisen, die Lösung des Nebenproblems liefert bei gegebener Randverschiebung die Deformation einer Membran, welche längs der Kurve C durch eine Last eins pro Längeneinheit belastet ist, und es ist bekannt,

daß das eine endliche Durchbiegung ergibt. Es folgt also genau wie früher

$$(41) \quad \lim_{n \rightarrow \infty} N_n = N$$

und somit

$$(42) \quad \lim_{n \rightarrow \infty} \int_C v_n ds = \int_C u ds.$$

Schließlich sei noch auf eine Methode hingewiesen, die zwar nicht unmittelbar zum Ritzschen Verfahren in Beziehung steht, die aber nützlich wird, wenn man durch irgendein zeichnerisches oder rechnerisches Verfahren eine Näherungslösung kennt und nun einzelne Werte der Funktion oder ihrer Ableitung auf ihre Genauigkeit prüfen will, insbesondere dann, wenn es wesentlich auf einzelne Funktions- oder Ableitungswerte ankommt, wie z. B. in der Potentialtheorie ebener Flüssigkeitsströmungen.

Betrachtet man den Ausdruck (13) für den Minimalwert des Nebenproblems

$$N(\varepsilon) = D(u) + \varepsilon u(\xi, \eta) - \frac{\varepsilon^2}{2} G(\xi\eta, \xi\eta)$$

und die daraus entnommene Formel (14)

$$u(\xi, \eta) = \frac{1}{2} \{N(+1) - N(-1)\},$$

so sieht man, daß sich der unendlich werdende Teil $G(\xi\eta, \xi\eta)$ forthebt, und der Gedanke liegt nahe, diese Singularität von vornherein abzuziehen. Man erhält das folgende Ergebnis, das ich der Kürze wegen auf dem einfachsten Wege ableite.

Es werde wieder die Funktion u gesucht, welche bei gegebenen Randwerten die Gleichung $\Delta u = 0$ befriedigt. Insbesondere soll der Funktionswert an einem bestimmten Punkte ξ, η gefunden werden. Stellen wir diesen Funktionswert in der üblichen Weise durch die Randintegrale dar,

$$(43) \quad 2\pi u(\xi, \eta) = \int u \frac{\partial \ln r}{\partial \nu} ds - \int \ln r \frac{\partial u}{\partial \nu} ds,$$

so kann bei gegebenen Randwerten das erste Integral rechts ausgewertet werden. Nun bezeichnen wir durch $\gamma(x, y)$ diejenige zunächst unbekannte Potentialfunktion, welche am Rande die Werte

$$(44) \quad \gamma = -\ln r$$

annimmt. Dann wird das zweite Integral, wegen $\Delta u = 0$,

$$(45) \quad -\int \ln r \frac{\partial u}{\partial \nu} ds = \int \gamma \frac{\partial u}{\partial \nu} ds = \iint \text{grad } u \text{ grad } \gamma \, dx \, dy \\ = \frac{1}{4} \iint \text{grad}^2 (u + \gamma) \, dx \, dy - \frac{1}{4} \iint \text{grad}^2 (u - \gamma) \, dx \, dy.$$

Diese beiden Integrale lassen sich nun aber durch die Minimalwerte der folgenden Variationsprobleme darstellen, nämlich: den Ausdruck

$$(46) \quad \iint \text{grad}^2 w \, dx \, dy$$

zum Minimum zu machen, wenn die Funktion w am Rande die Werte

$$(47) \quad w = u \pm \ln r$$

annimmt. Es ist damit auch hier der unbekannte Teil des Funktionswertes durch die Minimalwerte geeigneter „Nebenprobleme“ ausgedrückt, und die Abschätzung der Genauigkeit des Funktionswertes auf die Abschätzung von Minimalwerten zurückgeführt. Diese Methode ist weniger einfach als die frühere, dafür läßt sie auch die Bestimmung der Ableitungen zu. Schreiben wir z. B. für

$$(48) \quad 2\pi \frac{\partial u}{\partial \xi} = \frac{\partial}{\partial \xi} \int u \frac{\partial \ln r}{\partial \nu} \, ds - \int \frac{\partial u}{\partial \nu} \frac{\partial \ln r}{\partial \xi} \, ds,$$

so erhalten wir das unbekannte zweite Integral auf die gleiche Weise wie oben, wenn wir jetzt für γ die Potentialfunktion nehmen, welche auf dem Rand die Werte

$$(49) \quad \gamma = - \frac{\partial \ln r}{\partial \xi}$$

annimmt. Die gleiche Betrachtung ist auch auf alle höheren Ableitungen anwendbar.

7. Die Abschätzung der Minimalwerte.

Wir haben in den vorhergehenden Abschnitten dargelegt, in welcher Weise die Funktionswerte für die Lösung eines Minimalproblems durch die Minimalwerte geeigneter Nebenprobleme dargestellt werden können. Der Minimalwert eines Variationsproblems ist dabei die von der Lösung gelieferte untere Grenze für die Werte des Minimalausdrucks. Damit ist, abgesehen von der Vereinfachung der Konvergenzbeweise, die Möglichkeit der Fehlerschätzung zunächst für die erste Randwertaufgabe und im beschränkten Umfang auch für gemischte Randwertaufgaben gegeben. Denn durch die genannte Darstellung ist die Abschätzung der Funktionswerte auf die Abschätzung der Minimalwerte zurückgeführt. Mit dieser wollen wir uns jetzt beschäftigen.

Ich lege den Grundgedanken wieder an dem Beispiel der ersten Randwertaufgabe der Gleichung $\Delta u = 0$ dar. Wenn wir das zugehörige Variationsproblem: Bei den gegebenen Randwerten das Integral $\frac{1}{2} \iint \text{grad}^2 u \, dx \, dy$ zum Minimum zu machen, nach Ritz behandeln, so ist es klar, daß die Minimalausdrücke der Näherungslösungen zu groß ausfallen, denn die Ritzschen Näherungslösungen sind konkurrenzfähige Funktionen, so daß das

Minimalintegral nicht unter den für diese kleinstmöglichen Wert sinken kann. Wenn wir also eine Funktion w finden wollen, welche ein $D(w) < D(u)$ hat, so dürfen wir sie nicht aus der Klasse K der konkurrenzfähigen Funktionen nehmen.

Wir bedienen uns nun eines Gedankens von Courant, wonach der Minimalwert eines Variationsproblems sinkt, wenn durch eine Milderung der den Konkurrenzfunktionen auferlegten Bedingungen der Kreis der zulässigen Funktionen erweitert wird. Wir werden demgemäß das vorliegende Variationsproblem durch Milderung der Randbedingungen abändern, und zwar so, daß das geänderte Problem exakt lösbar wird. Dann kann der errechnete Minimalwert nicht oberhalb von $D(u)$ liegen. Wir ersetzen also für die Näherungslösung w die Randbedingung $u = f(s)$ durch n Bedingungen der Form

$$(50) \quad \int \{w - f(s)\} g_\rho(s) ds = 0.$$

Über die zunächst beliebigen Funktionen $g_\rho(s)$ werden wir dann so verfügen, daß das Problem streng lösbar wird. Soll mit diesen Nebenbedingungen der Ausdruck $\frac{1}{2} \iint \text{grad}^2 w dx dy$ ein Minimum werden, so muß für irgendein δw , wenn wir mit λ_1 bis λ_n die Lagrangeschen Faktoren der Nebenbedingungen bezeichnen,

$$(51) \quad \iint \text{grad} w \text{ grad} \delta w dx dy = \sum_1^n \lambda_\rho \int g_\rho(s) \delta w ds$$

sein, oder mit der üblichen partiellen Integration:

$$(52) \quad - \iint \delta w \Delta w dx dy + \int \delta w \frac{\partial w}{\partial \nu} ds = \int \delta w \left\{ \sum_1^n \lambda_\rho g_\rho(s) \right\} ds.$$

Soll dies für beliebiges δw gelten, so muß

$$(53) \quad \text{im Innern des Gebietes } \Omega \quad \Delta w = 0$$

$$(54) \quad \text{auf dem Rande } R \quad \frac{\partial w}{\partial \nu} = \sum_1^n \lambda_\rho g_\rho(s)$$

sein. Wählen wir nun die $g_\rho(s)$ so, daß sie mit den Normalableitungen eines Systems von Potentialfunktionen übereinstimmen:

$$(55) \quad g_\rho(s) = \frac{\partial p_\rho(x, y)}{\partial \nu},$$

so sind die beiden Minimalbedingungen durch den Ansatz

$$(56) \quad w_n = c_0 + \sum_1^n \lambda_\rho p_\rho(x, y)$$

erfüllt, und die Lagrangeschen Faktoren sind so zu bestimmen, daß den Gleichungen (50) Genüge getan wird. Das sind gerade n Gleichungen für

die n Faktoren. Die Integrationskonstante c_0 , die aus der Minimalbedingung nicht folgt, bestimmt man durch die Forderung, daß der Mittelwert des Randfehlers

$$(57) \quad \int \{w - f(s)\} ds = 0$$

wird. Damit ist das gemilderte Variationsproblem streng gelöst, und nach dem Courantschen Prinzip wird

$$(58) \quad D(w) \leq D(u),$$

was man auch durch direkte Rechnung leicht bestätigen kann.

Diese Methode, die ein Gegenstück zum Ritzschen Verfahren darstellt, ist eine Verallgemeinerung der klassischen Reihenentwicklungen nach Partikularlösungen. In der Tat sind die durch Orthogonalitätseigenschaften ausgezeichneten bekannten Reihenentwicklungen sämtlich als Spezialfälle darin enthalten. Ein Vorzug dieser Methode vor der Ritzschen ist, daß man die Randbedingungen nicht zu erfüllen braucht, auch ist der Konvergenzbeweis besonders einfach, wenn man die Existenz der Lösung voraussetzt und die approximierenden Funktionen so auswählt, daß sie in dem gegebenen Gebiete jede reguläre Potentialfunktion beliebig genau zu approximieren vermögen. (Nach einem Satz von Runge genügen hierzu in einem einfach zusammenhängenden, schlichten Gebiete die rationalen Potentialpolynome.) Betrachten wir nämlich das Integral über das Quadrat des Fehlergradienten

$$(59) \quad F_n = \iint \text{grad}^2(w_n - u) dx dy,$$

so läuft unsere Methode der Koeffizientenbestimmung darauf hinaus, daß unter allen Funktionen der Form (56) dieses Fehlerintegral möglichst klein wird. In der Tat ist unter Anwendung der üblichen partiellen Integrationen

$$(60) \quad \delta F_n = 2 \iint \text{grad}(w_n - u) \text{grad} \delta w dx dy = 2 \int \{w_n - f(s)\} \frac{\partial \delta w}{\partial \nu} ds \\ = 2 \sum_1^n \delta c_e \int \{w_n - f(s)\} \frac{\partial p_e}{\partial \nu} ds,$$

und dies verschwindet, wenn die Koeffizienten c_e nach Gleichung (50) bestimmt werden. Da nun bei der vorausgesetzten Vollständigkeit des Approximationssystems das Fehlerintegral beliebig klein gemacht werden kann, muß es auch in unserem Falle gegen Null gehen. Es ist also

$$(61) \quad \lim_{n \rightarrow \infty} F_n = \lim_{n \rightarrow \infty} \iint \text{grad}^2(w_n - u) dx dy = 0.$$

Haben wir nun ein Gebiet Ω^* , das ganz innerhalb Ω liegt, so können wir jeden Punkt mit einem Kreise vom Radius a umgeben. Für den

Mittelpunkt dieses Kreises ist nach den bekannten Mittelwertsätzen:

$$(62) \quad \text{grad}(w_n - u) = \frac{1}{\pi a^2} \iint \text{grad}(w_n - u) dx dy,$$

und nach der Schwarzschen Ungleichung

$$(63) \quad |\text{grad}(w_n - u)| \leq \frac{1}{a\sqrt{\pi}} \sqrt{\iint \text{grad}^2(w_n - u) dx dy} \leq \frac{1}{a} \sqrt{\frac{F_n}{\pi}},$$

woraus die in Ω^* gleichmäßige Konvergenz der ersten Ableitungen, und damit nach bekannten Sätzen auch aller übrigen Ableitungen folgt.

Analoge Betrachtungen lassen sich für die Differentialgleichungen höherer Ordnung anstellen. Setzt man z. B. für die Gleichung $\Delta \Delta u = 0$ bei gegebenen Randwerten $u = f(s)$, $\frac{\partial u}{\partial \nu} = g(s)$ eine Näherungslösung in der Form $v(x, y) = \sum c_\rho p_\rho(x, y)$ an, wo die $p_\rho(x, y)$ Partikularlösungen der biharmonischen Gleichung sind, so erhält man für den zum Minimum zu machenden Ausdruck $\iint \Delta u^2 dx dy$ sicher einen zu kleinen Wert, wenn man die Koeffizienten aus den Gleichungen

$$(64) \quad \int \{v_n - f(s)\} \frac{\partial \Delta p_\rho}{\partial \nu} ds - \int \left\{ \frac{\partial v_n}{\partial \nu} - g(s) \right\} \Delta p_\rho ds = 0$$

bestimmt, während der zu große Wert durch das Ritzsche Verfahren geliefert wird. —

Die Abschätzung nach unten geschieht immer durch Partikularlösungen, d. h. durch Funktionen, welche der Differentialgleichung genügen, oder mechanisch gedeutet, den Gleichgewichtsbedingungen im Innern. Dazu gehört auch bei den Minimalproblemen der Form

$$\frac{1}{2} \iint \Delta u^2 dx dy + \varepsilon u(\xi, \eta) = \text{Min},$$

daß die Lösung an der Stelle ξ, η die richtige Singularität hat, also in dem hier genannten Falle $-\frac{\varepsilon}{8\pi} r^2 \ln r$.

Eine Erweiterung für die Abschätzung der Minimalwerte habe ich in den Verhandlungen des zweiten internationalen Kongresses für technische Mechanik, Zürich 1926, angegeben. (E. Trefftz, „Ein Gegenstück zum Ritzschen Verfahren“, S. 131.) Man kann nämlich eine Milderung des Variationsproblems auch dadurch erreichen, daß man im Innern des gegebenen Gebietes auf die geforderte Stetigkeit oder Differenzierbarkeit verzichtet. Das ist besonders bei zusammengesetzten Gebieten von praktischer Bedeutung, wie ich dort an einem Beispiele gezeigt habe.

Gemischte Randbedingungen. Die bisherigen Betrachtungen über die Abschätzung der Minimalwerte bezogen sich auf die erste Randwertaufgabe, für welche bei den Differentialgleichungen zweiter Ordnung die

Randwerte der Funktion, bei denen vierter Ordnung die der Funktion und ihrer Normalableitung gegeben sind. Die Schwierigkeiten bei gemischten Randwertaufgaben will ich hier an einem Beispiel erläutern, an dem zugleich deutlich wird, wieweit die angegebenen Methoden auf diesen Fall übertragen werden können. Es sei ein Gebiet Ω in der xy -Ebene gegeben, und es werde die Funktion gesucht, welche der Differentialgleichung $\Delta u = 0$ genügt, und folgende Randbedingungen befriedigt: Der Rand bestehe aus zwei Teilen, Teil I und Teil II. Längs I sollen die Funktionswerte $u = f(s)$, längs II die Werte der Normalableitung $\frac{\partial u}{\partial \nu} = g(s)$ vorgeschrieben sein. Dem entspricht das Variationsproblem, unter allen Funktionen, welche auf I die Werte $f(s)$ annehmen, diejenige zu suchen, welche den Ausdruck

$$(65) \quad \frac{1}{2} \iint \text{grad}^2 u \, dx \, dy - \int_{\text{II}} u g(s) \, ds$$

zum Minimum macht. Wir sehen, daß für die zum Variationsproblem konkurrierenden Funktionen die Erfüllung der Randbedingung II $\frac{\partial u}{\partial \nu} = g(s)$ nicht mehr gefordert zu werden braucht; dafür ist im Minimalausdruck das Randintegral $\int_{\text{II}} u g(s) \, ds$ hinzugetreten. Das Ritzsche Verfahren liefert zu große Werte für den Minimalausdruck. Wenn wir nun genau wie oben eine Funktion konstruieren, die sicher einen zu kleinen Wert ergibt, indem wir nach Milderung der Randbedingungen das Variationsproblem streng lösen, so ist es offenbar, daß eine Milderung der Randbedingungen nur da vorgenommen werden kann, wo dieselben gegeben sind. Längs II, wo für das Variationsproblem keine Randbedingungen gegeben sind, kann sich nichts ändern, infolgedessen muß auch die strenge Lösung des gemilderten Variationsproblems hier die Randbedingung erfüllen, weil diese sich aus dem Variationsproblem selbst ergibt. — Das ist die oben angekündigte Schwierigkeit. Wir müssen zur Unterschreitung des Minimalwertes eine Näherungslösung mit Partikularlösungen ansetzen, welche auf dem Randstück II die Bedingung $\frac{\partial v}{\partial \nu} = g(s)$ erfüllt. Das ist nur in besonderen Fällen auf einfache Weise möglich, wenn etwa der Teil II aus einem geradlinigen oder kreisförmigen Stück der Begrenzung besteht, — immerhin sind darin eine Reihe praktisch wichtiger Fälle enthalten.

Für das angegebene Beispiel würde sich die Rechnung folgendermaßen gestalten. Man macht einen Ansatz

$$(66) \quad v_n(x, y) = c_0 + p_0(x, y) + \sum_1^n c_2 p_2(x, y),$$

in dem die $p_\varrho(x, y)$ Potentialfunktionen sind, von denen $p_0(x, y)$ auf Randteil II die Bedingung $\frac{\partial p_0}{\partial \nu} = g(s)$ erfüllt, während für die übrigen $p_\varrho(x, y)$ dort die Normalableitung verschwindet. Dann ist die Randbedingung hier für alle Werte der Koeffizienten sicher erfüllt. Diese werden so bestimmt, daß

$$(67) \quad \int_I \{v_n - f(s)\} \frac{\partial p_\varrho}{\partial \nu} ds = 0 \quad (0 \leq \varrho \leq n)$$

ist. Geht man von diesen Gleichungen aus, indem man sie als gemilderte Randbedingungen längs des Randstücks I betrachtet, so erkennt man, daß das so gemilderte Variationsproblem durch den Ansatz (66) streng gelöst wird. Also wird der Minimalausdruck sicher zu klein, wie man auch durch direkte Rechnung bestätigen kann.

Damit ist, soweit es mit unseren elementaren Mitteln möglich ist, die Abschätzung der Fehler bei den direkten Methoden der Variationsrechnung ermöglicht. Hoffentlich ist dieser erste Schritt geeignet, auch die praktische Rechnung an einem kritischen Punkte zu fördern.

(Eingegangen am 24. 11. 1927.)

Thus a_n is the n^{th} term of the series

$$0, \frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8}, \frac{1}{16}, \dots$$

and all the a_n are distinct.

Each of these functions v_i is continuous for $0 \leq x \leq 1$, which we take as our fundamental interval. Furthermore, these functions are linearly independent, since the first two obviously are so, and if any v_n ($n \geq 2$) were linearly dependent on a finite number of functions of the set not including v_n , it could not have a discontinuity in its derivative at a_n . Thus we may apply the process of orthogonalization²⁾ to this set and so obtain the normal and orthogonal set f_n , where

$$(2) \quad f_0 = 1, \\ f_{n+1} = \left[v_{n+1} - \sum_{h=1}^n f_h \int_0^1 f_h(t) v_{n+1}(t) dt \right] / \pm \sqrt{\int_0^1 v_{n+1}(t)^2 dt}.$$

We find, for example

$$(3) \quad \begin{aligned} f_0 &= 1, & f_1 &= \sqrt{3}(1-2x), \\ f_2 &= \sqrt{3}(1-4x), & x &\leq \frac{1}{2} \text{ and } \sqrt{3}(4x-3), & x &\geq \frac{1}{2}, \\ f_3 &= \sqrt{3}(10-76x)/19, & x &\leq \frac{1}{4}, \\ & \sqrt{3}(52x-22)/19, & \frac{1}{4} &\leq x \leq \frac{1}{2}, \\ & \sqrt{3}(10-12x)/19, & x &\geq \frac{1}{2}. \end{aligned}$$

The sign of these functions is not determined by (2), we take that sign before the radical which makes the first sign change of each f_n one from plus to minus. This defines in a unique way a set of continuous functions normal and orthogonal on the unit interval.

3. Convergence Properties.

The series expansion of a given function $F(x)$ in terms of the functions f_n is

$$(4) \quad \sum_{i=0}^{\infty} A_i f_i(x), \quad \text{where} \quad A_i = \int_0^1 F(t) f_i(t) dt.$$

As we shall derive the convergence properties from approximation considerations, we recall that, for any linear combination of the first $(n+1)$ f_i ,

$$T_n = \sum_{i=0}^n C_i f_i, \quad \text{we have:}$$

²⁾ cf. e. g. Courant-Hilbert, Methoden der math. Physik, Berlin 1924, p. 34.

$$\begin{aligned}
 (5) \quad E(T_n) &= \int_0^1 (F(x) - T_n(x))^2 dx \\
 &= \int_0^1 F(x)^2 dx - \sum_{i=0}^n A_i^2 + \sum_{i=0}^n (C_i - A_i)^2.
 \end{aligned}$$

This shows that, for n fixed, the mean square error $E(T_n)$ is least when $C_i = A_i$, and T_n is a partial sum of the series (4), $S_n = \sum_{i=1}^n A_i f_i$, and also that, as n increases, $E(S_n)$ decreases.

We shall now show that, if $F(x)$ is continuous, for any positive ε there exists an $N(\varepsilon)$ such that, if $n > N$, $E(S_n) < \varepsilon$. We first define a broken line function of type n , B_n , to be a continuous function, linear in each of the intervals bonded by consecutive points of the set (b_j) ,

$$(6) \quad 0, \frac{1}{2^k}, \frac{2}{2^k}, \dots, a_n, a_n + \frac{1}{2^k}, a_n + \frac{3}{2^k}, \dots, 1.$$

These (except for 1) are simply the a_i , $i \leq n$, arranged in order of magnitude. Every function B_n is clearly a linear combination of the first $(n+1)v_i$, for we may reduce it to zero by subtracting the constant times v_0 which makes it zero at $x=0$, then the constant times v_1 which coincides with the remainder in the first interval, then the constant times v_i ($i = 2^{k-1} + 1$ so that $a_i = \frac{1}{2^k}$) which coincides with the new remainder in the first two intervals, and so on. As it is evident from (2) that each v_i is linearly dependent on the first $(i+1)f_j$, we see that each B_n is a linear combination of the first $(n+1)f_j$, i. e. it is a T_n .

From the continuity of $F(x)$ in the closed interval $0 \leq x \leq 1$, we infer the existence of a $\delta(\varepsilon)$ such that $|F(x_1) - F(x_2)| < \varepsilon$, whenever $|x_1 - x_2| < \delta$. If K is an integer for which

$$(7) \quad \frac{1}{2^K} < \delta(\varepsilon), \quad N = 2^K$$

may be taken as $N(\varepsilon)$. For, as all the points $m/2^K$ ($m = 1, 2, 3, \dots, 2^K - 1$) are points a_i with $i \leq N$, the broken line function which agrees with $F(x)$ at 0, 1 and these points, and is linear in the intervals determined by them is a B_n and hence a T_n for any $n > N$. But, for this function, from (7) we have

$$(8) \quad |F(x) - T_n(x)| < \varepsilon,$$

so that $E(T_n)$ and hence $E(S_n) < \varepsilon$.

We shall next prove that when n exceeds the N of (7), $S_n(x)$ uniformly approximates $F(x)$. A characteristic property of $S_n(x)$ is the fact

that it is the T_n which minimizes $E(T_n)$. To apply this property, we need a lower limit for $E_{12}(T_n)$, the contribution to $E(T_n)$ from an interval $x_1 \leq x \leq x_2$, in which T_n is a linear function of x . We assume that

$$(9) \quad x_2 - x_1 = D < \delta(\varepsilon).$$

Consequently, if $L(x)$ is the linear function for which

$$(10) \quad L(x_1) = F(x_1) \quad \text{and} \quad L(x_2) = F(x_2),$$

we will have throughout the interval $x_1 x_2$

$$(11) \quad |F(x) - L(x)| < \varepsilon \quad (x_1 \leq x \leq x_2).$$

It follows from this that, in this interval

$$(12) \quad |F(x) - T_n(x)| > |L(x) - T_n(x)| - \varepsilon.$$

Recalling that $L(x)$ and $T_n(x)$ are both linear in the interval $x_1 x_2$ and putting

$$(13) \quad H_1 = |F(x_1) - T_n(x_1)|, \quad H_2 = |F(x_2) - T_n(x_2)|,$$

we find for the right member of (12):

$$(14) \quad |L(x) - T_n(x)| - \varepsilon = \frac{1}{D} |(H_2 \pm H_1)x - H_2 x_1 \mp H_1 x_2| - \varepsilon,$$

the upper and lower signs corresponding to the cases in which $T_n(x_1)$ and $T_n(x_2)$ are on opposite sides of, or on the same side of $L(x)$. If we take the upper signs, and integrate the square of this expression over the parts of the interval $x_1 x_2$ where the expression is positive, we find

$$(15) \quad E_{12}(T_n) > \frac{D [(H_1 - \varepsilon)^3 + (H_2 - \varepsilon)^3]}{3(H_1 + H_2)},$$

when H_1 and H_2 are both $\geq \varepsilon$. If either of these is $< \varepsilon$ the corresponding parenthesis does not appear in (15), but as it is negative, it may be inserted without destroying the inequality. If at least one of the H 's is $> 4\varepsilon$, we shall have

$$(16) \quad \frac{1}{H_1 + H_2} > \frac{1}{2[(H_1 - \varepsilon) + (H_2 - \varepsilon)]},$$

and in view of this, (15) becomes:

$$(17) \quad E_{12}(T_n) > \frac{D}{12} [(H_1 - \varepsilon)^2 + (H_2 - \varepsilon)^2], \quad \text{Max}(H_1, H_2) > 4\varepsilon.$$

When the lower signs in (14) are used, we have in place of (15):

$$(18) \quad E_{12}(T_n) > \frac{D [(H_1 - \varepsilon)^3 - (H_2 - \varepsilon)^3 - \varepsilon^2(H_1 - H_2)]}{3(H_1 - H_2)}.$$

When H_1 and H_2 are both $\geq \varepsilon$, the last term in the numerator does not appear. As after division by the denominator it is negative, its presence merely strengthens the inequality. We have inserted it so that (18) will apply to the case in which one of the H 's, say H_1 is $< \varepsilon$. Here the first term does not appear, and may not be supplied by itself since it is a positive multiple of the denominator. However, in view of our second condition, when $H_1 < \varepsilon$, $H_2 > 4\varepsilon$, and

$$(19) \quad \frac{(H_1 - \varepsilon)^3}{H_1 - H_2} - \varepsilon^2 < 0,$$

so that these two terms may be inserted together. As it is easily seen that (18) implies (17), this last inequality gives a lower limit for $E_{12}(T_n)$ which holds in all cases.

Let us now consider S_n , which is linear in the intervals bounded by the b_j of (6). We put

$$(20) \quad H_j = |F(b_j) - S_n(b_j)|.$$

Let M be the greatest of the quantities H_j , and b_g one of the points at which $H_g = M$. Suppose that $M > 4\varepsilon$, and let b_p be the first point to the left of b_g at which $H_p \leq 4\varepsilon$ (or 0, if no such point exists) and b_q (or 1) the first such point to the right. We will compare S_n with a particular T_n , obtained from it by the following process. For values of x outside the interval $b_p \leq x \leq b_q$, $T_n(x)$ coincides with $S_n(x)$. Inside this interval, we put

$$(21) \quad T_n(b_p) = S_n(b_p) \text{ (or } F(0)), \quad T_n(b_q) = S_n(b_q) \text{ (or } F(1)), \\ T_n(b_i) = F(b_i) \quad (p < i < q).$$

In view of the linear character of T_n in the intervals between the b_j , it is defined by these values.

To compare the mean square errors $E(S_n)$ and $E(T_n)$, we need merely compare the contributions from the interval $b_p b_q$. Applying (17), we find

$$(22) \quad E_{pq}(S_n) > \frac{3}{2}(b_q - b_p)\varepsilon^2 + \frac{D}{12}(M - \varepsilon)^2 - 4D\varepsilon^2.$$

Here $D = \frac{1}{2^k}$, so that the elementary b_j intervals are at least D , and at most $2D$ in width. For T_n , we obviously have

$$(23) \quad E_{pq}(T_n) < (b_q - b_p)\varepsilon^2 + 100D\varepsilon^2,$$

since the numerical error is at most ε , cf. (8), except in the end intervals, where it is at most 5ε . From (22) and (23) we find:

$$(24) \quad E_{pq}(S_n) - E_{pq}(T_n) > \frac{D}{12}(M - \varepsilon)^2 - 104D\varepsilon^2 + \frac{1}{2}(b_q - b_p)\varepsilon^2,$$

and since

$$(25) \quad E_{pq}(S_n) - E_{pq}(T_n) = E(S_n) - E(T_n) < 0,$$

we must have³⁾

$$(26) \quad M < 37\varepsilon.$$

Thus all the $H_j < 37\varepsilon$, and in consequence

$$(27) \quad |F(x) - S_n(x)| < 38\varepsilon \quad (n > N(\varepsilon)),$$

which proves our contention about uniform approximation. We may express this result as

Theorem I. *If any function $F(x)$, continuous in the interval $0 \leq x \leq 1$, be expanded in a series of constant multiples of the functions f_j (3), the resulting series (4) will converge to $F(x)$ at all points of the unit interval, and, in fact, uniformly.*

Let us next treat the case in which $F(x)$ is a measurable function with summable square in the unit interval. Such a function can be approximated to an arbitrary degree of exactness, in the sense of least squares, by a continuous function⁴⁾, $C(x)$, so that

$$(28) \quad \int_0^1 [F(x) - C(x)]^2 dx < \eta^2.$$

As we have shown above, (8), that there exists a T_n for which

$$(29) \quad \int_0^1 [C(x) - T_n(x)]^2 dx < \eta^2,$$

for this function $E(T_n) < 4\eta^2$. This implies the same relation for $E(S_n)$, and shows that for the functions now treated the expansion will converge in the mean.

If, at any point of the unit interval, $F(x)$ is continuous, we may show that the expansion converges there in the ordinary sense. Let, then, x_c be a point of continuity of $F(x)$, and $x_1 x_2$ an interval such that

$$(30) \quad |F(x) - F(x_c)| < \varepsilon, \quad x_1 \leq x \leq x_2; \quad x_2 - x_1 = x - x_1.$$

Let us take n so large that $\frac{1}{2^k}$ is small compared with $x_2 - x_1$, and

³⁾ By a longer calculation, which treats the cases more in detail, we may show that, in fact, $M < 17\varepsilon$.

⁴⁾ cf. e. g. E. W. Hobson; *Theory of functions of a Real Variable*, 2nd ed., Cambridge 1921, vol. 1, p. 584.

for S_n consider the H_j for the b_j of the interval $x_1 x_2$ here defined by

$$(31) \quad H_j = |F(x_c) - S_n(b_j)| \quad (x_1 \leq b_j \leq x_2).$$

If more than $1/P$ of these H_j were $\geq 4\varepsilon$, from (17) we would have

$$(32) \quad E(S_n) \geq E_{12}(S_n) > (x_2 - x_1) \varepsilon^2 / P.$$

As $E(S_n)$ approaches zero as n increases, we may take n so large that $1/P < \frac{1}{4}$. When this is done, $\frac{3}{4}$ of the $H_j < 4\varepsilon$, and in particular b_j for which this holds exist on both sides of x_c in the interval $x_1 x_2$. Calling b_p the first such point to the left of x_c , and b_q the first one to the right, we may use the argument given above to establish (27), to show that each of the H_j for b_j adjacent to x_c (or for x_c itself if that is a b_j) $< 37\varepsilon$, and hence

$$(33) \quad |F(x_c) - S_n(x_c)| < 38\varepsilon \quad (n > N_1(\varepsilon)).$$

This proves the convergence of S_n to $F(x)$ at any point of continuity, and an obvious modification of the argument shows that the convergence is uniform for any closed interval in which $F(x)$ is continuous. We have thus established:

Theorem II. *If any measurable function $F(x)$, with summable square in the interval $0 \leq x \leq 1$, be expanded in a series of constant multiples of the functions f_j (3), the resulting series (4) will converge to $F(x)$ at all points of continuity. Further, in any closed interval of continuity the convergence will be uniform. Over the whole interval, the series converges in the mean to $F(x)$.*

The consideration of simple examples shows that if the function $F(x)$ is continuous in each of the intervals $x_1 x_2$ and $x_2 x_3$, but discontinuous at x_2 , the series will in general oscillate between two finite values at x_2 .

4. Other Functions.

We have based the set of function used, f_i (3), on the functions v_i (1) with breaks at a_n , the proper fractions with denominators integral powers of 2. Similarly, we could obtain an orthogonal set of broken line functions from any other set of points p_n . For such a set to have the convergence properties of theorems I and II, it is sufficient that the points p_n be everywhere dense on the unit interval and enumerated in such a way that, of the intervals marked off at any stage by these points, the ratio of the greatest interval to the least remains uniformly bounded for the entire set. With this restriction on the p_n , the theorems can be proved essentially as above.

5. Application to the Haar set.

The set of discontinuous functions used by Haar may be obtained by applying the process of orthogonalization (2) to the derivatives of our functions v_i (1). As the linear combinations of them are step functions, the deductions made from (5) show that the set is complete. As, in a step function, any one step may be altered without disturbing the rest of the function, it is obvious from the minimum property that in any interval between adjacent points of discontinuity b_j , $S_n(x)$ must lie between the greatest and least values of $F(x)$ in the interval. From this remark most of the convergence theorems given by Haar may be deduced immediately without further calculation.

Massachusetts Institute of Technology.

(Eingegangen am 20. 12. 1927.)

Über die Fourierschen Reihen gewisser Funktionenklassen.

Von

Otto Szász in Frankfurt a. M.

§ 1.

$f(x)$ sei eine integrierbare Funktion mit der Fourierschen Reihe

$$(1) \quad f(x) \sim \frac{a_0}{2} + \sum_{\nu=1}^{\infty} (a_{\nu} \cos \nu x + b_{\nu} \sin \nu x);$$

ferner sei

$$s_0 = \frac{a_0}{2}, \quad s_n = \frac{a_0}{2} + \sum_1^n (a_{\nu} \cos \nu x + b_{\nu} \sin \nu x) \quad (n = 1, 2, 3, \dots),$$
$$\dot{\sigma}_n = \frac{s_0 + s_1 + \dots + s_{n-1}}{n} \quad (n = 1, 2, 3, \dots).$$

Vor einigen Jahren hatte ich den Satz bewiesen¹⁾:

Es sei $\lambda > 0$, $0 < \alpha < 1$ und

$$(L) \quad \int_0^{2\pi} [f(x+2t) + f(x-2t) - 2f(x)]^2 dx \leq \lambda t^{2\alpha} \quad \text{für } t > 0,$$

dann ist die Reihe $\sum_1^{\infty} (a_{\nu}^2 + b_{\nu}^2)^{\frac{1}{2}k}$ für $k > \frac{2}{2\alpha+1}$ konvergent; dagegen gibt es der Bedingung (L) genügende Funktionen, für die $\sum (a_{\nu}^2 + b_{\nu}^2)^{\frac{1}{2\alpha+1}}$ divergiert (also $k = \frac{1}{2\alpha+1}$).

¹⁾ Szász, 1, insbes. § 3. (L) ist offenbar eine Verallgemeinerung der Lipschitzbedingung vom Grade α auf quadratische Integralmittelwerte. Vor kurzem erhielt ich die Korrekturbogen einer Arbeit von Hardy und Littlewood (2), worin über solche Funktionenklassen (vgl. Lp) interessante Sätze bewiesen werden. — Die Nummern beziehen sich auf die am Schlusse angeführte Literatur.

Der Beweis beruhte auf der Ungleichung

$$(2) \quad \int_0^{2\pi} [\sigma_n(x) - f(x)]^2 dx < \frac{\lambda \varrho_1}{n^{2\alpha}} \quad (n = 1, 2, 3, \dots)^2)$$

die aus (L) hergeleitet wurde.

Im folgenden verallgemeinere ich dieses Resultat; der Beweisgang bedeutet auch für den eben zitierten Satz eine Vereinfachung. Ich beweise zunächst den

Satz 1. *Es sei $\lambda > 0$, $p > 1$, $-\frac{1}{p} < \alpha < 1$ und*

$$(Lp) \quad \int_0^{2\pi} |f(x+2t) + f(x-2t) - 2f(x)|^p dx \leq \lambda t^{\alpha p}, \quad t > 0;$$

dann ist

$$(2p) \quad \int_0^{2\pi} |\sigma_n(x) - f(x)|^p dx < \varrho_2 \lambda n^{-\alpha p} \quad (n = 1, 2, 3, \dots).$$

Setzt man zur Abkürzung

$$f(x+2t) + f(x-2t) - 2f(x) = \varphi(t, x) = \varphi,$$

so ist bekanntlich

$$\sigma_n(x) - f(x) = \frac{1}{n\pi} \int_0^{\frac{\pi}{2}} \left(\frac{\sin nt}{\sin t} \right)^2 \varphi(t, x) dt;$$

ferner ist³⁾

$$\left(\frac{\sin nt}{\sin t} \right)^2 < \left(\frac{\pi n}{1+nt} \right)^2, \quad 0 < t < \frac{\pi}{2},$$

also

$$|\sigma_n(x) - f(x)| \leq \pi n \int_0^{\frac{\pi}{2}} (1+nt)^{-2} |\varphi(t, x)| dt,$$

oder für irgendein reelles β

$$|\sigma_n(x) - f(x)| \leq \pi n \int_0^{\frac{\pi}{2}} (1+nt)^{\beta-1} |\varphi(t, x)| \cdot (1+nt)^{-\beta-1} dt.$$

²⁾ ϱ_1 hängt nur von α ab; die später auftretenden ϱ_2, \dots hängen von α und einem Parameter p ab.

³⁾ Infolge einer Bemerkung des Herrn Fejér ist nämlich

$$|\sin nt| \leq n \sin t \quad \text{und} \quad nt |\sin nt| \leq nt \leq \frac{\pi}{2} n \sin t,$$

also

$$(1+nt) |\sin nt| \leq \left(1 + \frac{\pi}{2}\right) n \sin t < \pi n \sin t.$$

Wir wenden nun die bekannte Ungleichung an:

$$\left| \int_a^b g(t) h(t) dt \right|^p \leq \int_a^b |g(t)|^p dt \cdot \left(\int_a^b |h(t)|^{\frac{p}{p-1}} dt \right)^{p-1}, \quad p > 1;$$

dann wird

$$|\sigma_n(x) - f(x)|^p \leq (\pi n)^p \int_0^{\frac{\pi}{2}} (1+nt)^{p(\beta-1)} |\varphi|^p dt \cdot \left(\int_0^{\frac{\pi}{2}} (1+nt)^{-\frac{p(\beta+1)}{p-1}} dt \right)^{p-1},$$

und wegen (Lp)

$$\begin{aligned} (3) \quad & \int_0^{2\pi} |\sigma_n(x) - f(x)|^p dx \\ & \leq \lambda (\pi n)^p \int_0^{\frac{\pi}{2}} (1+nt)^{p(\beta-1)} \cdot t^{\alpha p} dt \cdot \left(\int_0^{\frac{\pi}{2}} (1+nt)^{-\frac{p(\beta+1)}{p-1}} dt \right)^{p-1}, \\ & \alpha p > -1. \end{aligned}$$

Nun sei (man beachte, daß $\alpha < 1$)

$$-\frac{1}{p} < \beta < 1 - \alpha - \frac{1}{p}, \quad \text{z. B. } \beta = \frac{1-\alpha}{2} - \frac{1}{p};$$

dann ist

$$\int_0^{\frac{\pi}{2}} (1+nt)^{p(\beta-1)} \cdot t^{\alpha p} dt = n^{-1-\alpha p} \int_0^{\frac{\pi}{2}} (1+\tau)^{p(\beta-1)} \cdot \tau^{\alpha p} d\tau < \varrho_3 n^{-1-\alpha p}$$

und

$$\int_0^{\frac{\pi}{2}} (1+nt)^{-\frac{p(\beta+1)}{p-1}} dt = \frac{1}{n} \int_0^{\frac{\pi}{2}} (1+\tau)^{-\frac{p(\beta+1)}{p-1}} d\tau < \frac{\varrho_4}{n},$$

also nach (3)

$$\int_0^{2\pi} |\sigma_n(x) - f(x)|^p dx < \varrho_2 \lambda n^{-\alpha p}. \quad \text{Qu. e. d.}$$

Je größer p ist, desto schärfer ist die Voraussetzung (Lp), aber auch die Ungleichung (2p); für $p=2$ ist hierin die Ungleichung (16) meiner Arbeit 1. enthalten. Man beachte, daß allgemein

$$\left(\frac{1}{2\pi} \int_0^{2\pi} |F(x)|^\gamma dx \right)^{\frac{1}{\gamma}} \leq \left(\frac{1}{2\pi} \int_0^{2\pi} |F(x)|^\delta dx \right)^{\frac{1}{\delta}}, \quad 0 < \gamma < \delta.$$

§ 2.

Es sei jetzt $1 < p \leq 2$; die Anwendung eines Hausdorffschen Satzes⁴⁾ auf die Fouriersche Reihe

$$f(x) - \sigma_n(x) \sim \sum_{\nu=1}^{n-1} \frac{\nu}{n} (a_\nu \cos \nu x + b_\nu \sin \nu x) + \sum_{\nu=n}^{\infty} (a_\nu \cos \nu x + b_\nu \sin \nu x)$$

ergibt unmittelbar

$$\left(n^{\frac{p}{1-p}} \sum_1^{n-1} \nu^{\frac{p}{p-1}} |c_\nu|^{\frac{p}{p-1}} + \sum_n^\infty |c_\nu|^{\frac{p}{p-1}} \right)^{p-1} < \varrho_5 \int_0^{2\pi} |\sigma_n(x) - f(x)|^p dx,$$

$$c_\nu = a_\nu - i b_\nu;$$

also nach (2p)

$$(4) \quad n^{\frac{p}{1-p}} \sum_1^{n-1} \nu^{\frac{p}{p-1}} |c_\nu|^{\frac{p}{p-1}} + \sum_n^\infty |c_\nu|^{\frac{p}{p-1}} < \varrho_6 \lambda' n^{-\frac{\alpha p}{p-1}}, \quad (\lambda' = \lambda^{\frac{1}{p-1}}),$$

und insbesondere

$$(4') \quad \sum_n^\infty |c_\nu|^{\frac{p}{p-1}} < \varrho_6 \lambda' n^{-\frac{\alpha p}{p-1}}.$$

Ich wende nun die bekannte Ungleichung an

$$\frac{1}{k} \sum_1^k u_\nu \leq \left(\frac{1}{k} \sum_1^k u_\nu^\gamma \right)^{\frac{1}{\gamma}}, \quad \gamma \geq 1, \quad u_\nu \geq 0;$$

dann ist für $n_2 > n_1$, $\gamma = \frac{p}{p-1} \cdot \frac{1}{k}$

$$\frac{1}{n_2 - n_1} \sum_{1+n_1}^{n_2} |c_\nu|^k \leq \left(\frac{1}{n_2 - n_1} \sum_{1+n_1}^{n_2} |c_\nu|^{\frac{p}{p-1}} \right)^{\frac{(p-1)k}{p}}, \quad k \leq \frac{p}{p-1},$$

und nach (4')

$$\frac{1}{n_2 - n_1} \sum_{1+n_1}^{n_2} |c_\nu|^k < \left[\frac{1}{n_2 - n_1} \varrho_6 \lambda' n_1^{-\frac{\alpha p}{p-1}} \right]^{\frac{(p-1)k}{p}}.$$

Es ist also

$$\sum_{1+n_1}^{n_2} |c_\nu|^k < (n_2 - n_1)^{1 - \frac{(p-1)k}{p}} (\varrho_6 \lambda')^{\frac{(p-1)k}{p}} n_1^{-\alpha k};$$

ich setze hier $n_1 = 2^\mu$, $n_2 = 2^{\mu+1}$, dann wird

$$(5) \quad \sum_{1+2^\mu}^{2^{\mu+1}} |c_\nu|^k < 2^{\mu \left(1 - \frac{p-1}{p} k \right) - \alpha \mu k} (\varrho_6 \lambda')^{\frac{p-1}{p} k} \quad (\mu = 1, 2, 3, \dots).$$

⁴⁾ F. Hausdorff, 3.

Nun sei $\alpha > 0$ und $1 - \frac{p-1}{p}k - \alpha k < 0$, das heißt

$$\frac{p}{p-1+\alpha p} < k;$$

dann folgt aus (5) die Konvergenz der Reihe $\sum |c_n|^k$. Somit gilt der Satz 2. *Es sei $1 < p \leq 2$, $0 < \alpha < 1$, und*

$$\int_0^{2\pi} |f(x+2t) + f(x-2t) - 2f(x)|^p dx = O(t^{\alpha p});$$

dann ist die mit den Fourierschen Koeffizienten c_n gebildete Reihe $\sum |c_n|^k$ konvergent für $k > \frac{p}{p(1+\alpha)-1}$.⁵⁾

Für $k = \frac{p}{p(1+\alpha)-1}$ kann die Reihe divergieren; sei z. B.

$$f(x) = |x|^{\alpha - \frac{1}{p}}, \quad -\pi < x < \pi;$$

dann ist offenbar $\alpha p - 1 > -1$, und ⁶⁾

$$\int_{-\pi}^{\pi} |\varphi(t, x)|^p dx = O(t^{\alpha p}),$$

während die Fourierschen Koeffizienten im Falle $\alpha p \neq 1$ die Größenordnung $\nu^{\frac{1}{p} - \alpha - 1}$ haben, so daß $\sum |c_n|^{\frac{p}{(1+\alpha)p-1}}$ divergiert.

Für den Fall $\alpha p = 1$, also $k = 1$ liefert ein Beispiel die Funktion

$$f(x) = 1 \quad \text{für } 0 < x < \pi, \quad f(x) = -1 \quad \text{für } -\pi < x < 0.$$

Jetzt ist offenbar

$$\int_{-\pi}^{\pi} |\varphi(t, x)|^p dx = O(t),$$

und die Fourierschen Koeffizienten haben die Größenordnung $\frac{1}{\nu}$, somit ist $\sum |c_n|$ divergent.

Korollar zu Satz 2. Sei $f(x)$ absolut stetig und $|f'(x)|^p$ integrierbar; dann ist (Lp) mit $\alpha = 1$ erfüllt, so daß $\sum |c_n|^k$ für $k > \frac{p}{2p-1}$ konvergiert.

Es ist nämlich

$$\varphi(t, x) = \int_0^{2t} [f'(x+\tau) - f'(x-\tau)] d\tau,$$

⁵⁾ Unter der weiteren Voraussetzung $\alpha p > 1$ bei Hardy und Littlewood, 2, Theorem 8.

⁶⁾ Vgl. Hardy und Littlewood, 2.

also

$$|\varphi(t, x)|^p \leq (2t)^{p-1} \int_0^{2t} |f'(x + \tau) - f'(x - \tau)|^p d\tau$$

und

$$\int_0^{2\pi} |\varphi(t, x)|^p dx \leq (2t)^{p-1} \int_0^{2t} \left(\int_0^{2\pi} |f'(x + \tau) - f'(x - \tau)|^p dx \right) d\tau = O(t^p)$$

Qu. e. d.

Insbesondere ist $\sum |c_\nu|$ konvergent (wegen $\frac{p}{2p-1} < 1$); dies hat schon Herr Tonelli mit Hilfe des Hausdorffschen Satzes bewiesen⁷⁾; auch das obige Korollar kann auf diese Weise gewonnen werden.

Zusatz. Die Bedingung (L) ist gleichbedeutend mit

$$\sum_{\nu=1}^n \nu^4 |c_\nu|^2 = \lambda O(n^{4-2\alpha}).$$

Zunächst kann man nämlich für (L) schreiben

$$(L') \quad 16\pi \sum |c_\nu|^2 \sin^4 \nu t \leq \lambda t^{2\alpha}, \quad t > 0; \text{ } ^8)$$

hieraus folgt für $t = \frac{\pi}{2n}$, $n > 1$,

$$\sum_1^n |c_\nu|^2 \sin^4 \nu \frac{\pi}{2n} < \lambda \varrho_7 n^{-2\alpha},$$

und wegen $\sin x > \frac{2}{\pi} x$, $0 < x < \frac{\pi}{2}$,

$$(6) \quad \sum_1^n \nu^4 |c_\nu|^2 < \lambda \varrho_8 n^{4-2\alpha}, \quad n > 1.$$

Umgekehrt, wenn (6) gilt, so sei bei gegebenem $0 < t < 1$

$$\frac{1}{t} \leq n < \frac{1}{t} + 1 < \frac{2}{t};$$

dann ist

$$\sum_1^n |c_\nu|^2 \sin^4 \nu t \leq \sum_1^n |c_\nu|^2 \nu^4 t^4 < \lambda \varrho_8 t^4 n^4 \cdot n^{-2\alpha} < \lambda \varrho_9 n^{-2\alpha}$$

und

$$\sum_{\nu=n+1}^{n+s} |c_\nu|^2 \sin^4 \nu t \leq \sum_{\nu=n+1}^{n+s} |c_\nu|^2.$$

Setzt man nun zur Abkürzung

$$\sum_1^n |c_\nu|^2 \cdot \nu^4 = u_n,$$

⁷⁾ Tonelli, 4.

⁸⁾ Vgl. Szász, 1, S. 148.

so ist

$$\begin{aligned} \sum_{n+1}^{n+s} |c_v|^2 &= \sum_{n+1}^{n+s} (u_v - u_{v-1}) v^{-4} \\ &= -\frac{u_n}{(n+1)^4} + \sum_{n+1}^{n+s-1} u_v \left(\frac{1}{v^4} - \frac{1}{(v+1)^4} \right) + \frac{u_{n+s}}{(n+s)^4} < \lambda_{10} n^{-2\alpha}, \end{aligned}$$

also

$$\sum_{n+1}^{n+s} |c_v|^2 < \lambda_{10} t^{2\alpha} \quad (s = 1, 2, 3, \dots).$$

Somit ist mit Rücksicht auf (L') der Zusatz bewiesen.

Literatur.

1. O. Szász, Über den Konvergenzexponenten der Fourierschen Reihen gewisser Funktionenklassen, Sitzungsber. d. Akademie München 1922, S. 135–150.
2. G. H. Hardy und J. E. Littlewood, A convergence criterion for Fourier series, Math. Zeitschr. 28 (1928), S. 612–634.
3. F. Hausdorff, Eine Ausdehnung des Parsevalschen Satzes über Fourierreihen, Math. Zeitschr. 16 (1923), S. 163–169.
4. L. Tonelli, Sulla convergenza assoluta delle serie di Fourier, Atti della Reale Accad. Naz. dei Lincei 1925, serie 6, Rendic. Classe di Scienze fisiche, mat. e nat. 2, p. 145–149.

(Eingegangen am 17. 7. 1928.)

Sui moduli delle curve poligonali, e sopra un complemento al teorema di esistenza di Riemann.

Von

Beniamino Segre in Rom.

1. Si dice che una curva algebrica è ν -gonale, quando contiene (almeno) una g_ν^1 , ma non contiene alcuna serie lineare infinita d'ordine inferiore a ν . Così, ad esempio, le curve 1-gonali sono le curve razionali, e le 2-gonali sono le curve iperellittiche.

Detto p il genere della curva, se l'intero ν uguaglia o supera $\frac{p}{2} + 1$, esiste sempre sulla curva qualche g_ν^1 ¹⁾: dunque, una curva ν -gonale sarà a moduli particolari solo se è $\nu < \frac{p}{2} + 1$. Ci si può chiedere di determinare il numero N dei moduli da cui dipendono le curve ν -gonali di genere $p > 1$, per cui è $\nu < \frac{p}{2} + 1$: tale numero è stato determinato dal prof. F. Severi²⁾, e vale

$$N = 3p - 3 - i,$$

indicando con i l'indice di specialità della serie doppia della g_ν^1 .

Da questa proposizione, Severi ha dedotto, per $\nu = 2$, il numero $2p - 1$ (già noto per altre vie) dei moduli da cui dipendono le curve iperellittiche di genere p , e, per $\nu = 3$, il numero $2p + 1$ dei moduli da cui dipendono le curve trigonali di genere p .³⁾ Però, per ν qualunque, la suddetta proposizione non porge senz'altro il numero N richiesto, espresso solo mediante i valori di p e di ν ; ebbene, basandoci su di essa, noi dimostriamo nella prima parte di questo lavoro che *le curve ν -gonali di genere $p > 1$ (con $\nu < \frac{p}{2} + 1$) dipendono da*

moduli.

$$N = 2p + 2\nu - 5$$

¹⁾ Ciò è corollario immediato di un noto teorema, per cui cfr. F. Severi, *Vorlesungen über algebraische Geometrie* (Leipzig, Teubner, 1921), p. 159.

²⁾ F. Severi, *Sul teorema di esistenza di Riemann*, Rendic. Circ. Mat. di Palermo 46 (1922), p. 105, n.° 5.

³⁾ F. Severi, Nota cit. in ²⁾, n.° 6.

La determinazione di questo numero N vien completata da alcune considerazioni sulle curve canoniche di genere p .

Può forse sembrare che si possa giungere in modo pressocchè immediato a determinare il numero N , col seguente ragionamento. In base al teorema di esistenza di Riemann, si vien a *determinare* un numero finito di enti algebrici di genere p contenenti una g^1 , quando di questa vengano dati gli $\omega = 2v + 2p - 2$ elementi di diramazione: i moduli da cui dipendono le curve v -gonali di genere p son quindi in numero di $\omega - 3$, poichè appunto i gruppi proiettivamente distinti di ω punti di una retta sono $\infty^{\omega-3}$. Però questo ragionamento presuppone in modo essenziale, che una curva di genere p , pel solo fatto di contenere una g^1 , non ne contenga di conseguenza infinite; questa proposizione non è a priori evidente⁴⁾: essa resta dimostrata per $v < \frac{p}{2} + 1$, in virtù della concordanza del suddetto N col numero $\omega - 3$.

La seconda parte di questo lavoro è volta alla determinazione delle curve piane v -gonali normali d'ordine minimo, su cui la g^1 è segata dalle rette di un fascio. Prendendo la questione più in generale (e senza neppure escludere che la curva contenga delle serie lineari di dimensione inferiore a v), noi stabiliamo che

Date genericamente $\omega = 2v + 2p - 2$ rette di un fascio ($v > 2$), affinchè esista (almeno) una curva piana irriducibile di genere p ed ordine $v + n$ (n intero ≥ 0), passante n volte pel centro del fascio, e di cui le ω rette date siano le tangenti (in punti semplici) che appartengono a questo fascio, è necessario e sufficiente che sia

$$n \geq \frac{p-v+2}{2}.$$

È questo un complemento, *proiettivo*, che sembra notevole, del teorema di esistenza (complemento enunciato sotto veste analitica al n. 9). Da esso segue, come caso particolarissimo, il noto teorema di C. Segre che dice una curva algebrica di genere p contenente una g^1 , esser sempre riducibile con una trasformazione birazionale ad una curva piana d'ordine $p + 2$ con punto $(p - v + 2)$ -plo ($v \leq p + 2$).⁵⁾ Si può anzi risol-

⁴⁾ Anzi, essa non è vera per $v > \frac{p}{2} + 1$, mentre che per $v < \frac{p}{2} + 1$ è in contraddizione col noto criterio di Plücker-Clebsch. — Così, se non si conoscesse il teorema di Riemann-Roch, alla stessa stregua si potrebbe argomentare che una curva di genere p , pel fatto di contenere un gruppo di $v < \frac{p}{2} + 1$ punti presentanti solo $v - 1$ condizioni ai gruppi canonici, non ne contenesse infiniti, mentre ciò non è vero.

⁵⁾ C. Segre, *Recherches générales sur les courbes et les surfaces réglées algébriques*, Math. Annalen 30 (1887), p. 203, n.° 19.

V. anche E. Bertini, *La geometria delle serie lineari sopra una curva piana secondo il metodo algebrico*. Ann. di Mat. (2) 22 (1894), p. 1, n.° 43.

vere il problema di minimo cui si è accennato dianzi, e si ha precisamente che:

Una curva algebrica di genere p contenente una g_v^1 , con $v \leq p + 2$, è riducibile ad una curva piana d'ordine $\frac{p+v}{2} + \varepsilon + 1$, con punto $(\frac{p-v}{2} + \varepsilon + 1)$ -plo, ove ε vale 0 od $\frac{1}{2}$, a seconda che $p + v$ è pari o dispari. Se gli elementi di diramazione della g_v^1 formano birapporti generici, tale curva piana soddisfa al problema di minimo proposto, ed è normale.

I modelli piani di curve *trigonal*i dati da questo teorema per $v = 3$, erano già stati trovati da F. Amodeo⁶⁾, senza tuttavia che questi abbia notato la proprietà di *minimo* che li caratterizza.

Si noti poi il caso (già osservato da Severi⁷⁾) della quartica piana di genere 3, di cui le 12 tangenti che passano per un punto generico del piano non son 12 rette *qualunque* di un fascio (come direbbe il conto delle costanti): infatti il modello minimo di curva di genere 3, su cui le rette di un fascio segano una g_4^1 a elementi di diramazione *generici*, è del 5.° ordine. Più in generale risulta che una curva iperspaziale di genere p ed ordine $v < p + 2$, è tale che i $2v + 2p - 2$ iperpiani di un fascio che la toccano, *non* sono $2v + 2p - 2$ iperpiani *generici* di un fascio.

Nell'ultimo enunciato in corsivo, abbiamo trascurato (perchè banale) il caso in cui sia $v > p + 2$. In tale ipotesi la g_v^1 è sempre totalmente contenuta in una serie lineare più volte infinita, ond'essa può venir segata dalle rette di un fascio, su di una curva piana d'ordine v .

I.

2. Seguendo la via indicata da Severi per $v = 3$, consideriamo, per v qualunque purchè maggior di 2, la curva C canonica di genere p . Se C è v -gonale, esiste su essa una g_v^1 *completa*, *priva di punti fissi*; ognuno degli $\infty^1 G_v$ della g_v^1 appartiene ad uno spazio S_{v-2} , e, per determinare l'indice i di specialità del doppio della g_v^1 , occorre esaminare le possibili incidenze di questi S_{v-2} a due a due. Precisamente: se due generici di questi S_{v-2} hanno uno S_σ a comune ($\sigma = -1$ nel caso che non abbiano alcun punto comune), essi appartengono ad uno $S_{2v-\sigma-4}$, onde l'indice i di specialità vale

$$i = p - 2v + \sigma + 3.$$

⁶⁾ F. Amodeo, *Courbes normales trigonales du plan*, Comptes Rendus 130 (1900), p. 1744; e *Curve di gonalità k con punti fissi nella $(k-1)$ esima serie canonica, e curve normali trigonali del piano*. Rendic. Acc. Sc. di Napoli (3) 6 (1900), p. 174, n.° 19 e segt.

⁷⁾ F. Severi, Nota cit. in ²⁾, n.° 3.

In base al citato teorema di Severi, il numero N dei moduli è dato da

$$N = 2p + 2v - \sigma - 6,$$

e quindi la prima proposizione che ci siam proposti di dimostrare *equivalente* a far vedere che è $\sigma = -1$, ossia che quegli S_{v-2} sono a due a due sghembi, e che l'indice i di specialità vale

$$i = p - 2v + 2.$$

Noi dimostreremo ai n.º 3 e 4 quella proposizione, costruendo, dati ad arbitrio gli interi p e v (purchè sia $2 < v < \frac{p}{2} + 1$), una curva di genere p , contenente una g_v^1 completa e priva di punti fissi, tale che la serie doppia abbia proprio l'indice di specialità $i = p - 2v + 2$. E notiamo che non è necessario che la curva costruita sia effettivamente una curva v -gonale, ossia non contenga alcuna serie lineare infinita d'ordine inferiore a v . Basta infatti che questa curva *non sia iperellittica*, talchè si possa considerare in S_{p-1} la relativa curva canonica C di genere p : per ciò che si è detto poc' anzi, gli $\infty^1 S_{v-2}$ di S_{p-1} che si appoggiano a C nei vari G_v della g_v^1 , sono a due a due sghembi, onde lo stesso fatto dovrà presentarsi per ogni curva canonica *generica* di genere p , che contenga una g_v^1 , ossia per ogni curva v -gonale di genere p , che, come tale, sia generica.

Ciò non esclude che, per curve v -gonali particolari, detti S_{v-2} possano presentare particolari incidenze: noi, anzi, determineremo ai n.º 5 e 6 tutte le possibili incidenze, stabilendo in conseguenza il numero dei parametri da cui dipendono le curve v -gonali che offrono tale particolarità, ed inoltre *caratterizzando* dal punto di vista invariante le loro serie lineari g_v^1 .

3. Fissiamo nel piano 4 punti indipendenti O, A, B, C , e consideriamo il sistema lineare Σ , costituito dalle curve d'ordine

$$2v + k \quad \left(2 < v < \frac{p}{2} + 1; \quad k \geq 4 \right)$$

aventi un punto $(v+k)$ -plo in O , e tre punti v -pli in A, B, C . Dico che questo sistema Σ è certo *irriducibile*. Infatti, in caso contrario, la generica curva di Σ dovrebbe, per un noto teorema di Bertini, contenere una parte fissa, oppure comporsi con un certo numero di parti, variabili in uno stesso fascio. Ora l'uno e l'altro caso sono da escludere, poichè fra le curve di Σ vi sono quelle spezzate in k rette arbitrarie del fascio di centro O , ed in v coniche arbitrarie del fascio di punti base O, A, B, C .

La dimensione ed il genere virtuali di Σ valgono rispettivamente:

$$\begin{aligned} \delta_k &= \frac{1}{2}(2v+k)(2v+k+3) - \frac{1}{2}(v+k)(v+k+1) - \frac{3}{2}v(v+1) \\ &= kv + k + v, \end{aligned}$$

$$\Theta_k = \frac{1}{2}(2\nu + k - 1)(2\nu + k - 2) - \frac{1}{2}(\nu + k)(\nu + k - 1) - \frac{3}{2}\nu(\nu - 1) \\ = (k - 1)(\nu - 1).$$

Risulta quindi che è $\delta_k > \Theta_k$, laonde, per un noto teorema di C. Segre, il sistema Σ dev'essere *regolare*.

Dalle precedenti relazioni si ha anzi che:

$$\delta_k - \Theta_k = 2k + 2\nu - 1 > 2d,$$

ove d è un qualunque intero che soddisfa alla doppia limitazione

$$0 \leq d \leq k + \nu - 1.$$

Poniamo:

$$p = \Theta_k - d;$$

in virtù delle ipotesi fatte risulterà:

$$p \geq (k - 2)(\nu - 2) - 2 \geq 0.$$

Un teorema di Castelnuovo⁸⁾ ci assicura allora che le curve di Σ che passano doppiamente per d punti D fissati *ad arbitrio* nel piano, formano un nuovo *sistema lineare irriducibile* Σ' , di genere p e di dimensione $\delta = \delta_k - 3d$; anche Σ' è *regolare*, poichè, in base alle precedenti relazioni risulta $\delta > p$.

Il genere p di Σ' soddisfa alle disequaglianze

$$\theta_k \leq p \leq \Theta_k,$$

ove si è posto:

$$\theta_k = \Theta_k - (k + \nu - 1) = (k - 2)(\nu - 2) - 2;$$

anzi, p può assumere un qualunque valore intero dell'intervallo (θ_k, Θ_k) , bastando scegliere in modo conveniente il numero d dei punti D .

Notiamo ora che (essendo per ipotesi $\nu > 2$) al crescere di k , θ_k e Θ_k crescono indefinitamente, soddisfacendo per ogni k ($k \geq 4$) alle disequaglianze:

$$\theta_k < \Theta_k; \quad \theta_k < \theta_{k+1}; \quad \Theta_k < \Theta_{k+1}; \quad \theta_{k+1} < \Theta_k.$$

Ciò mostra che ciascuno degli intervalli (θ_k, Θ_k) ricopre parzialmente il consecutivo $(\theta_{k+1}, \Theta_{k+1})$, onde questi intervalli comprendono complessivamente tutti i numeri da $\theta_4 = 2\nu - 6$ a $+\infty$. Dunque, dato *comunque* l'intero $\nu > 2$ e l'intero $p > 2\nu - 2$, esiste almeno un $k \geq 4$, tale che p cada nel corrispondente intervallo (θ_k, Θ_k) . Per ciò che precede si potranno scegliere ad arbitrio nel piano i 4 punti O, A, B, C , ed i $d = \Theta_k - p$ punti D , e saremo sicuri che esiste (almeno) una *curva* K ,

⁸⁾ G. Castelnuovo, *Massima dimensione dei sistemi lineari di curve piane di dato genere*, Annali di Mat. (2) 18 (1890), p. 119, n.° 5.

irriducibile di genere p e d'ordine $2\nu + k$, che passa $\nu + k$ volte per O , ν volte per ciascuno dei punti A, B, C , e due volte per ciascuno dei punti D : su tale curva le rette per O segano una g_ν^1 priva di punti fissi.

4. Dico che la curva K che abbiamo costruito alla fine del n.° precedente, ha tutti i requisiti richiesti al n.° 2: per dimostrare ciò, dovremo far vedere che essa non è iperellittica, che la g_ν^1 su essa considerata è completa, e che il doppio di questa serie lineare ha l'indice di specialità $i = p - 2\nu + 2$.

Consideriamo a tal uopo le curve d'ordine $2\nu + k - 3$, aggiunte a $\Sigma' \equiv |K|$: da esse si staccano le tre rette OA, OB, OC , che sono curve fondamentali di Σ' . Le curve del sistema aggiunto puro che contengono un gruppo G_ν della g_ν^1 , si spezzano nella retta per O che contiene tale G_ν , ed in una curva Γ variabile, d'ordine $2\nu + k - 7$, che deve passare $\nu + k - 5$ volte per O , $\nu - 2$ volte per ciascuno dei punti A, B, C , e contenere semplicemente i d punti D . Dico che queste, per le curve d'ordine $2\nu + k - 7$, son condizioni indipendenti.

Potremo intanto limitarci alle condizioni imposte dai punti multipli, poichè i punti *semplici* D , essendo d punti che (n.° 3) posson darsi ad arbitrio nel piano, imporranno certamente d condizioni indipendenti. Per ciò che concerne le condizioni imposte dai punti multipli, basterà far vedere che è possibile di soddisfare a tutte queste condizioni, tranne una arbitraria⁹⁾. Ciò è evidente, quando si osservi che fra le curve d'ordine $2\nu + k - 7$ che soddisfano a tutte quelle condizioni, vi sono quelle spezzate in $k - 3$ rette arbitrarie per O ed in $\nu - 2$ coniche arbitrarie per O, A, B, C .

Dunque, le curve Γ linearmente indipendenti sono in numero di

$$\begin{aligned} & \frac{1}{2}(2\nu + k - 5)(2\nu + k - 6) - \frac{1}{2}(\nu + k - 5)(\nu + k - 4) \\ & \quad - \frac{3}{2}(\nu - 2)(\nu - 1) - d = \\ & = k\nu - k - 2\nu - d + 2 = \Theta_k - d - \nu + 1 = p - \nu + 1. \end{aligned}$$

Questo è pertanto l'indice di specialità della g_ν^1 segata su K dalle rette per O . In base al teorema di Riemann-Roch, da qui si trae che tale g_ν^1 è completa. Si ha inoltre che K non può essere iperellittica, poichè nell'ipotesi contraria, la serie g_ν^1 (che è speciale e priva di punti fissi) dovrebbe comporsi colla g_ν^1 esistente su K , ed il suo indice di specialità varrebbe $p - \frac{\nu}{2}$ invece di $p - \nu + 1$.

⁹⁾ Si potrebbe dimostrare che il sistema delle curve d'ordine $2\nu + k - 7$ che soddisfa a quelle condizioni è *irriducibile e regolare*, imitando il processo seguito al n.° 3 pel sistema Σ : da qui seguirebbe subito che K non è iperellittica.

E veniamo da ultimo alla determinazione dell' *indice i di specialità della serie lineare doppia di g_v^1* . Dovremo vedere quante sono le curve Γ linearmente indipendenti che contengono un gruppo G_v di g_v^1 . Tali curve si spezzano nella retta fissa per O che contiene tale G_v , ed in una parte variabile, i cui caratteri si hanno da quelli delle curve Γ mutandovi k in $k-1$. Per ciò che precede sarà quindi

$$i = \Theta_{k-1} - d - v + 1 = \Theta_k - d - 2v + 2 = p - 2v + 2,$$

che è quanto dovevasi dimostrare.

5. Ritorniamo alla curva C canonica di genere p , e premettiamo alcune considerazioni generali. Avendo su C due serie lineari complete, g_v^e e $g_v^{e'}$, l'una l'altra residua rispetto alla serie canonica (talchè sarà $v + v' = 2p - 2$, $e' = v' - p + e + 1$), i loro gruppi G_v e $G_{v'}$ sono rispettivamente segati su C da spazi S ed S' , di dimensioni $v - e - 1$ e $v' - e' - 1$ ¹⁰⁾. Uno spazio S ed uno spazio S' son congiunti da uno ed un solo iperpiano; e, inversamente, ogni iperpiano per S (per S') contiene uno ed un sol spazio S' (S).

Ciò premesso, supponiamo che *tutti* gli spazi S' abbiano un punto fisso P a comune: preso un qualunque spazio S , *ogni* iperpiano per esso, contenendo uno S' , passerà per P , ond'è necessario che anche S contenga il punto P .

Risulta adunque che, se i vari spazi S' si segano in uno S_σ fisso ($\sigma \geq 0$), questo è altresì spazio comune a tutti gli spazi S ; i quali spazi, poi, non possono segarsi in uno spazio di dimensione superiore a σ , poichè questo dovrebbe pur esser spazio comune agli S' .

Gli iperpiani passanti per S_σ , segano su C una serie lineare di dimensione $p - \sigma - 2$, la quale è *totalmente* contenuta nella serie canonica: infatti S_σ non può appoggiarsi a C , se no g_v^e e $g_v^{e'}$ verrebbero ad avere dei punti fissi. Ed è chiaro che queste serie lineari sono anche l'una l'altra residua rispetto a quella $g_{2p-2}^{p-\sigma-2}$.

Dico che quest'ultima serie lineare è la loro *somma minima*, ossia che non esiste alcuna serie lineare d'ordine $2p - 2$ e dimensione inferiore a $p - \sigma - 2$, che contenga *tutti* i gruppi di $2p - 2$ punti ottenuti aggregando ad un gruppo qualunque di g_v^e un gruppo qualunque di $g_v^{e'}$. Infatti, supposto che esista una serie lineare siffatta, indichiamone con $p - \sigma' - 2$ la dimensione: sarà $\sigma' > \sigma$, e la serie lineare che si considera, deve potersi segare su C cogli iperpiani che contengono un certo $S_{\sigma'}$. Ora, poichè gli iperpiani che congiungono uno spazio S ad un qualunque S' , devono contenere $S_{\sigma'}$, e d'altronde tali iperpiani sono iperpiani *arbitrari*

¹⁰⁾ Cf. F. Severi, op. cit. in ¹⁾, p. 133.

condotti per S , occorre che sempre S passi per S_{σ} ; mentre i vari spazi S hanno *solo* a comune uno S_{σ} .

In base a quest'ultimo ragionamento, ed a ciò che precede, potremo dire che:

Se sulla curva canonica di genere p si hanno due serie lineari complete g_{ν}^{σ} e $g_{\nu'}^{\sigma'}$, l'un l'altra residua rispetto alla serie canonica, i loro gruppi possono venir segati sulla curva da due serie di spazi S ed S' , di dimensione $\nu - \sigma - 1$ e $\nu' - \sigma' - 1$ rispettivamente. Se gli spazi S' passano per uno S_{σ} fisso, anche gli spazi S passano per esso, e la somma minima delle due serie lineari, ha la dimensione $p - \sigma - 2$. Inversamente, da quest'ultimo fatto si deduce l'esistenza di uno S_{σ} fisso, per cui passano tutti gli spazi S ed S' .

6. Supponiamo ora che C sia ν -gonale, e contenga adunque una g_{ν}^1 completa, con $2 < \nu < \frac{p}{2} + 1$. Consideriamo la serie $g_{\nu'}^{\sigma'}$ residua di g_{ν}^1 rispetto alla serie canonica, ed i due sistemi relativi di spazi S ed S' , di cui al numero precedente. Nelle ipotesi attuali è $\sigma = 1$, $\nu' = 2p - \nu - 2$, $\sigma' = p - \nu$, onde gli spazi S (come già s'è detto al n.° 2) hanno la dimensione $\nu - 2$, mentre gli spazi S' hanno la dimensione $p - 3$.

Dico che se due diversi spazi S hanno un punto P a comune, per esso devono passare *tutti* gli spazi S' . Infatti, se esistesse un S' non passante per P , consideriamo i due iperpiani che lo congiungono ai due spazi S che passano per questo punto; essi dovrebbero *coincidere* con l'iperpiano (determinato ed unico) che congiunge S' a P : e ciò è assurdo, poichè questo iperpiano verrebbe a segare C in più di $2p - 2$ punti.

In base al numero precedente, ciò dimostra che se due spazi S hanno un punto a comune, per esso devono passare *tutti* gli ∞^1 spazi S .

Gli spazi S , di dimensione $\nu - 2$, possono, al più, avere uno $S_{\nu-3}$ a comune. Dico che *questo caso va escluso*. Gli ∞^1 spazi S sono ν -secanti la curva C che è d'ordine $2p - 2$; se avessero uno $S_{\nu-3}$ in comune, essi formerebbero un cono d'ordine $\frac{2p-2}{\nu}$ e dimensione $\nu - 1$; poichè questo cono apparterebbe allo $S_{\nu-1}$, dovrebbe aversi:

$$\frac{2p-2}{\nu} \geq p - \nu + 1,$$

ossia:

$$p(\nu - 2) \leq \nu^2 - \nu - 2 < \nu^2 - 4,$$

da cui:

$$p \leq \nu + 1;$$

e ciò contraddice le disequaglianze $2 < \nu < \frac{p}{2} + 1$.

Da quanto precede, tenendo anche presenti i n.° 2 e 3, risulta che:

Gli spazi $S_{\nu-2}$, che determinano su di una curva canonica di genere p , i G_ν di una sua g_ν^1 completa ($2 < \nu < \frac{p}{2} + 1$), non possono a due a due che segarsi in uno S_σ fisso, con $-1 \leq \sigma \leq \nu - 4$. In generale σ vale -1 , ossia quegli $S_{\nu-2}$ sono a due a due sghembi.

Può però accadere che sia $\sigma \geq 0$; affinchè questa particolarità si presenti, è necessario e sufficiente che la somma minima della g_ν^1 considerata e della sua serie residua rispetto alla serie canonica, non sia la serie canonica completa, ma una serie lineare in essa totalmente contenuta, di dimensione $p - \sigma - 2$.

Le curve ν -gonali di genere p , la cui g_ν^1 presenta questa particolarità, dipendono da $2p + 2\nu - \sigma - 6$ moduli.

II.

7. Consideriamo nel piano il sistema lineare Φ , costituito dalle curve d'ordine $n + \nu$ aventi un punto n -plo O . È noto che la dimensione di questo sistema è data da

$$f = \frac{1}{2}\nu(\nu - 1) + n\nu + 2\nu + n,$$

e che la generica curva di Φ è irriducibile.

Suppongasì che fra le curve di Φ ve ne siano di irriducibili il cui genere valga p (≥ 0): una curva siffatta avrà, fuori di O , altre singolarità equivalenti ad un certo numero d di punti doppi, dato da

$$d = \frac{1}{2}\nu(\nu - 1) + n\nu - \nu - n - p + 1.$$

È chiaro che questo numero dovrà risultare maggiore o uguale a zero. Dimosteremo reciprocamente al n.° 8, che basta che sia $d \geq 0$, affinchè si possa asserire che Φ contiene qualche curva irriducibile di genere p : risulterà anzi che la generica curva irriducibile C di Φ il cui genere è p , ha precisamente, fuori del punto O , che è n -plo ordinario, d punti doppi, che verranno indicati colla lettera D .

Dico che i punti D impongono alle curve di Φ , d condizioni distinte. Sia infatti h la dimensione del sistema lineare costituito dalle curve di Φ che passano semplicemente per i punti D . Se a queste curve imponiamo di passare per $\nu + 1$ punti generici di una retta per O , otteniamo un sistema lineare la cui dimensione non può essere inferiore a $h - (\nu + 1)$. D'altronde le curve di quest'ultimo sistema si spezzano in tale retta per O ,

e nelle curve d'ordine $n + \nu - 1$ *aggiunte* a C ; per una nota proposizione si avrà dunque

$$p + 2n + 2\nu - 2 \geq h - (\nu + 1),$$

ossia

$$h \leq f - d.$$

E poichè ovviamente h non può risultare inferiore ad $f - d$, sarà $h = f - d$, il che dimostra l'asserto.

Usufruendo della rappresentazione delle curve di Φ coi punti di uno spazio lineare S_f , si conclude¹¹⁾ che la curva C individua una "falda lineare" di dimensione

$$h = f - d = 3\nu + 2n + p - 1$$

di curve di Φ , con d punti doppi prossimi a quelli di C . Eppertanto C sta in un sistema algebrico irriducibile H , ∞^h , di curve siffatte; la C è una curva *generica* di H .

Tracciamo le $\omega = 2\nu + 2p - 2$ rette per O che toccano C altrove, e consideriamo la totalità delle curve di H che toccano queste rette. Sia K una parte irriducibile di questa totalità, che contenga C , e la cui dimensione verrà indicata colla lettera k . Un teorema di F. Severi, relativo appunto a sistemi continui di curve soggette alla condizione di avere un dato numero di punti doppi e di toccare un dato gruppo di rette di un fascio, ci assicura che *la serie caratteristica di K su C è completa*¹²⁾.

Detta serie caratteristica, fuori dei punti fissi, che cadono nei punti multipli di C e nei punti di contatto delle ω tangenti considerate, ha l'*ordine*

$$(n + \nu)^2 - n^2 - 2d - \omega = 2n + \nu.$$

Detto j il suo *indice di specialità*, la sua dimensione è data, per quanto s'è detto, da

$$2n + \nu - p + j.$$

D'altro canto essa dev'esser inferiore di un'unità, alla dimensione k di K ; sarà dunque

$$k = 2n + \nu - p + j + 1.$$

¹¹⁾ Basta ragionare come nell' Op. cit. in ¹⁾, a p. 314-315.

¹²⁾ V. Nota cit. in ²⁾, n.° 5. — Esso è caso particolare di un noto teorema dovuto a F. Enriques, *Sulla proprietà caratteristica delle superficie algebriche irregolari*, R. Acc. delle Scienze di Bologna 9 (1904), p. 5; però la dimostrazione del teorema generale non è valida, se non quando la serie caratteristica è *non speciale*: ciò è stato osservato da F. Severi, nella Nota V *Sulla teoria degli integrali semplici di 1^a specie appartenenti ad una superficie algebrica*, Rendic. R. Acc. Naz. dei Lincei (5) 30 (1921), p. 296.

Ogni gruppo di ω rette del fascio O , che tocchi una C generica di H , ne tocca quindi, in conseguenza, ∞^k altre; eppertanto i gruppi di ω rette del fascio O tangenti alle curve C , dipendono da

$$h - k = \omega - j$$

parametri. Dunque, se si vuole che detti gruppi di ω rette siano arbitrari nel loro fascio, è necessario e sufficiente che risulti $j = 0$, ossia che la serie caratteristica di K sia non speciale.

Per dimostrare la penultima proposizione enunciata al n.° 1, si deve ora solo più, dati ν e p , determinare n in guisa che detta serie caratteristica risulti non speciale.

8. Prima però, dobbiamo stabilire la proposizione ammessa al n.° precedente, che

dati comunque gl' interi p , n e ν , purchè sia

$$d = \frac{1}{2} \nu(\nu - 1) + n\nu - \nu - n - p + 1 \geq 0,$$

esistono sempre delle curve irriducibili d'ordine $n + \nu$ e genere p , aventi un punto n -plo ordinario O e d punti doppi¹³⁾.

Al sistema lineare Φ considerato al principio del n.° 7, appartengono le curve spezzate in $n + \nu$ rette, di cui n passanti per O . Consideriamo una generica, Γ , di queste curve, spezzata in n rette a_1, a_2, \dots, a_n per O , ed in ν rette ulteriori b_1, b_2, \dots, b_ν : essa, fuori del punto n -plo O , ha

$$\delta = \frac{1}{2} \nu(\nu - 1) + n\nu$$

punti doppi. Dico che questi impongono alle curve di Φ , δ condizioni distinte. All' uopo s' indichi con f' la dimensione del sistema Φ' costituito dalle curve di Φ che passano pei suddetti punti doppi: sarà

$$f' \geq f - \delta,$$

ed occorre far vedere che in questa relazione occorre prendere il segno =. Scegliamo genericamente n punti su a_1, a_2, \dots, a_n rispettivamente; le curve di Φ' che passano per questi punti, si spezzano nelle n rette a_1, a_2, \dots, a_n , ed in una curva d'ordine ν aggiunta alla curva spezzata nelle rette b_1, b_2, \dots, b_ν . Il sistema lineare che così si ottiene ha la dimensione

¹³⁾ La dimostrazione che segue, si avvicina assai a quella data da Severi (ai n.° 1 e 2 della Nota cit. in ²⁾) pel teorema di esistenza; ci permettiamo quindi di sorvolare un po' sui dettagli.

$$\frac{1}{2}\nu(\nu+3) - \frac{1}{2}\nu(\nu-1) = 2\nu,^{14)}$$

onde dovrà aversi

$$2\nu \geq f' - n,$$

il che, per le precedenti relazioni, implica che sia effettivamente

$$f' = f - \delta = n + 2\nu.$$

Ciò premesso, sulla curva Γ si fissino $n + \nu + p - 1$ dei δ punti doppi, da considerarsi come virtualmente inesistenti, in modo che Γ diventi una curva connessa col punto n -plo O , ed altri

$$\delta - (n + \nu + p - 1) = d$$

punti doppi assegnati¹⁵⁾; e si noti che la cosa ha senso, poichè, per ipotesi, risulta $d \geq 0$. Il genere virtuale di Γ varrà

$$\binom{n+\nu-1}{2} - \binom{n}{2} - d = p.$$

Poichè i δ punti doppi assegnati di Γ presentano condizioni indipendenti alle curve di Φ , usufruendo della rappresentazione delle curve di Φ coi punti di uno spazio lineare S_f , si conclude¹⁶⁾ che la curva connessa Γ individua una "falda lineare" di dimensione

$$f - d = 3\nu + 2n + p - 1 > 0$$

di curve di Φ , con d punti doppi prossimi a quelli assegnati di Γ . Eppertanto Γ sta in un sistema algebrico Ω irriducibile ed infinito di curve, la cui curva generica è irriducibile, d'ordine $n + \nu$ e genere p , ed ha un punto n -plo ordinario in O e d punti doppi ulteriori.

9. Ritorniamo ora alla serie caratteristica di K su C . Per ciò che s'è detto al n.° 7, questa serie è *completa*, e vien segata su C dalle curve di Φ che passano pei d punti D doppi di C , e per gli ω punti di contatto delle tangenti per O . Fra queste curve si hanno palesemente quelle spezzate nella *prima polare di O rispetto a C* , ed in *una retta* ulteriore: questo mostra che si ha un gruppo di quella serie lineare, aggiungendo

¹⁴⁾ Ciò si può dire in base ad un teorema dato da Severi (a p. 316 dell' Op. cit. in ¹⁾), il quale dice che *i punti doppi di una qualunque curva, irriducibile o spezzata, d'ordine ν senza singolarità superiori, presentano sempre condizioni indipendenti alle aggiunte d'ordine ν .*

¹⁵⁾ Affinchè Γ risulti connessa, si possono p. es. riguardare come inesistenti gli $n + \nu - 1$ punti ove b_1 sega le altre rette di Γ (con che Γ risulta già connessa e di genere virtuale zero) eppoi p punti arbitrari dei rimanenti punti doppi di Γ .

¹⁶⁾ Basta ragionare come nell' Op. cit. in ¹⁾, a p. 324.

agli $n + \nu$ punti di una sezione rettilinea di C , gli n punti di C che coincidono con O (punti che sono origini di n rami lineari)¹⁷⁾.

Da qui segue facilmente che la serie caratteristica di K è speciale o non speciale, a seconda che esistono o non esistono delle aggiunte d'ordine $n + \nu - 4$ di C , che passano n volte per O .

Il conto delle costanti dimostra subito che queste aggiunte esistono certamente se

$$n \leq \frac{p - \nu + 1}{2}.$$

In base a ciò che precede, per provare la verità della proposizione che abbiamo in vista, basterà far vedere che, preso comunque

$$n \geq \frac{p - \nu + 2}{2}.$$

1° il numero d definito al n.° 7 non è negativo, talchè, giusta il n.° 8, esistono delle curve irriducibili d'ordine $n + \nu$ e genere p , aventi un punto n -plo ordinario O , e d punti doppi;

2° la generica C di queste curve, non ammette nessuna curva aggiunta d'ordine $n + \nu - 4$ che passi n volte per O .

Il primo punto si stabilisce facilmente. Infatti, in virtù delle fatte ipotesi (si ricordi che è $\nu \geq 3$), si ha

$$n(\nu - 1) \geq 2n \geq p - \nu + 2,$$

e pertanto

$$d = \frac{1}{2}\nu(\nu - 3) + n(\nu - 1) - p + 1 \geq \frac{1}{2}(\nu - 2)(\nu - 3) \geq 0.$$

Per dimostrare il secondo punto, ammettiamo (ragionando per assurdo) che la generica di quelle curve C abbia (almeno) un'aggiunta C_1 d'ordine $n + \nu - 4$ che passi n volte per O . Si potrà supporre, come s'è detto, che la curva C stia nel sistema Ω costruito al n.° 8 a partire dalla curva spezzata Γ . Quando C , muovendosi in Ω , tende a Γ , i suoi d punti doppi tendono ai d punti doppi assegnati di Γ . Dunque, al limite C_1 diverrà una

¹⁷⁾ Ciò permette di caratterizzare invariantivamente quella serie caratteristica, in funzione della g_ν^2 segata su C dalle rette per O : essa si ottiene sommando ad un gruppo G_ν di questa g_ν^2 , il doppio di un gruppo G_n di $n \geq 0$ punti, tale che la serie $|G_\nu + G_n|$ sia almeno doppiamente infinita.

In base a questa osservazione, tenuto presente la fine del n.° 7, la proposizione che abbiamo in vista si può anche enunciare sotto forma invariantiva così:

Data su d'un ente algebrico ∞^1 di genere p , una g_ν^2 completa ($\nu > 2$), i cui G_ν di diramazione entro alla ∞^1 razionale dei G_ν , di g_ν^2 non presentino particolarità, affinché esista (almeno) un G_n tale che $|G_\nu + G_n|$ sia più che semplicemente infinita, senza che $|G_\nu + 2G_n|$ risulti speciale, è necessario e sufficiente che si abbia

$$n \geq \frac{p - \nu + 2}{2}.$$

curva Γ_1 d'ordine $n + \nu - 4$, passante semplicemente per questi d punti doppi, ed avente in O un punto n -plo. Dico che una curva Γ_1 siffatta non può esistere.

All' uopo, esaminiamo un pò da vicino come si possono scegliere i d punti doppi assegnati di Γ , o (ciò che fa lo stesso) gli $n + \nu + p - 1$ suoi punti doppi da considerarsi come virtualmente inesistenti. Si ha, per ipotesi,

$$p \leq 2n + \nu - 2,$$

e quindi

$$n + \nu + p - 1 \leq 3n + 2\nu - 3.$$

Ciò premesso, fissiamo tre diverse delle rette b_1, b_2, \dots, b_ν (ciò è possibile essendo $\nu \geq 3$), e siano le b_1, b_2, b_3 . Le rette b_1 e b_2 sono segate dalle ulteriori $n + \nu - 2$ rette di Γ , in $2n + 2\nu - 4$ punti; se aggiungiamo a questi il punto d'intersezione di b_1, b_2 e gli n punti in cui b_3 vien segata dalle rette a_1, a_2, \dots, a_n , otteniamo un totale di $3n + 2\nu - 3$ punti. In base alla precedente relazione, ed a ciò che è stato detto in ¹⁵⁾, i punti doppi di Γ da considerarsi come virtualmente inesistenti, possono venir scelti fra i suddetti $3n + 2\nu - 3$ punti¹⁶⁾.

Da qui segue che, se la curva Γ_1 d'ordine $n + \nu - 4$ considerata poc' anzi esistesse realmente, essa dovrebbe avere con ciascuna delle rette $a_1, a_2, \dots, a_n; b_4, b_5, \dots, b_\nu$, $n + \nu - 3$ intersezioni, e quindi contenerla per intero; e ciò è assurdo, poichè queste rette sono in numero di $n + \nu - 3$, numero che è superiore all'ordine di Γ_1 .

Le considerazioni fin qui svolte dimostrano pienamente la penultima proposizione del n.º 1. Ad essa si può dare veste analitica, enunciando il teorema di esistenza così:

Di una funzione $y(x)$ algebrica irriducibile a $\nu (\geq 3)$ rami, di genere $p (\geq 0)$, si possono scegliere ad arbitrio gli $\omega = 2\nu + 2p - 2$ punti di diramazione (semplici) e le trasposizioni ad essi relative, purchè il gruppo di questi trasposizioni sia transitivo ed il loro prodotto sia l'identità. Una tale funzione è definita a meno di trasformazioni birazionali.

Se gli ω punti di diramazione sono generici, e se è $\nu \leq p + 2$, il legame algebrico fra y ed x , ha, nelle due variabili, un grado che non può discendere al di sotto di

$$\frac{p + \nu + 2}{2};$$

mentre esso può, a piacere, raggiungere un qualsiasi valore intero non inferiore a questo limite.

¹⁵⁾ È ovvio che si dovranno considerare tutti questi $3n + 2\nu - 3$ punti doppi come virtualmente inesistenti, allora e solo allora che nell'ultima relazione valga il segno =.

L'ultima proposizione del n.º 1, è un corollario immediato della precedente. L'unico punto che occorra ancor dimostrare, è l'affermazione fatta, che le curve "minime" che ivi si considerano, sono *normali*. Ora, presa una C di queste curve, se essa potesse ottenersi come proiezione sul piano di una curva sghemba Γ dello stesso ordine, da un punto P esterno a Γ , la retta che congiunge P al centro O del fascio che sega su C la g_v^1 , non potrebbe appoggiarsi a Γ , poichè se no la proiezione di Γ sul piano da uno dei punti d'appoggio, darebbe una curva di genere p ed ordine inferiore a quello di C , su cui le rette per O segherebbero una g_v^1 cogli stessi elementi di diramazione della g_v^1 segata su C .

La molteplicità $\frac{p-v}{2} + \varepsilon + 1$ del punto O per C deve quindi valere zero, il che implica che C sia d'ordine v , e che risulti $v = p + 2$. Poichè C non passa per O (ossia, colle precedenti notazioni, è $n = 0$), per ciò che s'è detto poc' anzi la curva C non può ammettere nessuna aggiunta d'ordine $v - 4$. Ciò significa che la serie lineare segnata su C dalle rette del piano non è speciale, eppertanto (in virtù della $v = p + 2$) essa è *completa*.

Dunque, effettivamente, i modelli minimi considerati sono *normali*.

Representation of Manifolds.

Von

Wilfrid Wilson in Amsterdam¹⁾.

Introduction.

§ 1.

By an n -dimensional simplicial manifold is understood an n -dimensional manifold²⁾ in Brouwer's sense, that is, a connected set of points built up of n -dimensional elements³⁾, any two of which have either no point in common, or a p -dimensional face together with all its q -dimensional faces in common ($p = 0, 1, \dots, n - 1$; $q = 0, 1, \dots, p - 1$), while those elements of common vertex are related in the same way as the simplexes of a simplex star³⁾ of R_n ; if in this definition the restriction that elements of common vertex are related as simplexes of a simplex star of R_n be replaced by the conditions that each $(n - 1)$ -dimensional face belongs exactly to two elements and each p -dimensional face ($0 \leq p \leq n - 2$) to a finite number of faces, while the property of a vertex of being common to several elements is only recognised in so far as it can be derived from the property of $(n - 1)$ -dimensional faces of being common each to two elements, we have an n -dimensional pseudo-manifold of Brouwer⁴⁾. A simplicial manifold or a pseudo-manifold is said to be

¹⁾ Fellow of the International Education Board.

²⁾ Brouwer, „Über Abbildung von Mannigfaltigkeiten“, Math. Annalen 71, p. 97; this work is referred to in the sequel as „A. v. M.“.

³⁾ „A. v. M.“, p. 97. By an n -dimensional element is understood the topological image of an n -dimensional simplex and by a simplex star of R_n a finite number of simplexes dense in a neighbourhood of a common vertex O , no two of which have inner points in common and any two of which have a common p -dimensional face ($0 \leq p \leq n - 1$) but no further common point.

⁴⁾ „Beweis der Invarianz des n -dimensionalen Gebiets“, Math. Annalen 71, p. 305 and 306; the definition there given is here completed in accordance with a verbal communication from Prof. Brouwer.

closed when the number of its elements is finite; otherwise it is said to be open⁵). By the indicatrix⁶) of an element is understood a certain order of its vertices, orders obtained from each other by an even number of permutations being considered equivalent, so that an element has two indicatrices one of which may be arbitrarily assigned as positive. Consider in a simplicial or pseudo-manifold a finite closed chain consecutive elements of which have n -vertices in common and assign a positive indicatrix to one of these elements; from it we determine the negative indicatrix of the following element by writing the newly entering vertex in place of the discarded one⁷); in this way we determine a positive indicatrix for each element of the chain in succession and on returning to the initial element assign a new positive indicatrix to it; if for every closed chain this new positive indicatrix is identical with the original one the manifold is called two-sided, otherwise one-sided⁷). By a locally simplicial n -dimensional manifold is understood a connected topological space⁸) to whose defining system of neighbourhoods there exists a countable equivalent⁹) system each of which is the interior of an n -dimensional element; the manifold is said to be closed or open according as it is or is not a compact topological space¹⁰); the following remarks¹¹) 1, ..., 4 place such manifolds and the problem treated in the following investigation, in a clearer light:

1. In a locally simplicial manifold M it may be assumed that each neighbourhood together with its boundary is an n -dimensional element contained in M (*i. e.*, the topological image of an n -dimensional simplex together with its boundary). For let the neighbourhood U_i be the topological image of the interior of the n -dimensional simplex S_i ; consider in S_i a countable set of n -dimensional simplexes $S_i^{(j)}$ (each with its boundary) such that each inner point is covered by the interior of a sequence of the simplexes $S_i^{(j)}$ converging to this point and let $U_i^{(j)}$ be the image of the interior of $S_i^{(j)}$ in U_i ; $\{U_i^{(j)}\}$ ($i, j = 1, 2, 3, \dots$) is the

⁵) „A. v. M.“, p. 97–98.

⁶) „A. v. M.“, p. 100.

⁷) „A. v. M.“, p. 101.

⁸) Hausdorff, „Grundzüge der Mengenlehre“ (1914), p. 213; this work is referred to in the sequel by “Hausdorff”.

⁹) Hausdorff, p. 260.

¹⁰) Such manifolds have been defined by Weyl, „Die Idee der Riemannschen Fläche“ (1913), p. 17–18; Tibor Radó, „Über den Begriff der Riemannschen Fläche“, *Acta litt. scient. Reg. Univ. Franc. Jos.*, 2, Fasc. II (1925), who proves that a closed, locally simplicial, 2-dimensional manifold is simplicial; and v. Kerékjártó, „Vorlesungen über Topologie I“ (1923), Einleitung, p. 5–6.

¹¹) These remarks have been suggested to me by Prof. Brouwer.

required set of neighbourhoods; assume in the sequel that the boundaries of all neighbourhoods used are contained in M ; then the neighbourhood ω_i together with its boundary being written $\bar{\omega}_i$, $\bar{\omega}_i$ is an n -dimensional element contained in M ; we then have the remark:

2. the manifold M is "condensed" with respect to any sequence of the above neighbourhoods $\omega_1, \omega_2, \dots$ covering it, *i. e.*, if the index $i(P)$ of an arbitrary point P be defined as the least value of ν for which $P \subset \omega_\nu$, then a sequence of points P_1, P_2, \dots has or has not a limit point in M according as the sequence of indices $i(P_1), i(P_2), \dots$ is bounded or monotonely increases respectively. Proof: When $i(P_1), i(P_2), \dots$ is bounded a finite m exists such that $\sum_1^{\infty} P_i \subset \sum_1^m \omega_i$, hence the sequence P_1, P_2, \dots has a limit point $P' \subset \sum_1^m \bar{\omega}_i \subset M$. When the sequence $i(P_1), i(P_2), \dots$ monotonely increases the assumption that P_1, P_2, \dots has a limit point P' leads to a contradiction, for then the subsequence P'_1, P'_2, \dots of P_1, P_2, \dots converging to P' is such that $i(P'_1), i(P'_2), \dots$ is bounded.

We note that two arbitrary points P and Q of M are connected by a finite chain $\omega_{r_1}, \omega_{r_2}, \dots, \omega_{r_r}$ such that $P \subset \omega_{r_1}, \omega_{r_i} \omega_{r_{i+1}} \neq 0$ ($i = 1, 2, \dots, r-1$), $Q \subset \omega_{r_r}$. For the set G of points which can be joined to P in the above manner completely contains each ω_i of which it contains a point and is therefore an open set containing P and having no edge point; hence M being connected, G must contain Q .

3. M is a metrical space by the theorem of Urysohn¹²⁾ since it is a normal space satisfying the 2nd enumerability axiom of Hausdorff.

4. The system of neighbourhoods defining M can be replaced by an equivalent system any two of which having a common point are contained in a single neighbourhood of the original system. Proof. Let $r(P)$ be the greatest value of ρ such that points distant less than ρ from P are contained in one of the neighbourhoods $\omega_1, \omega_2, \dots$ defining M . Define $m_1 \equiv 1$ and $m_{\nu+1}$ as the least value such that $\bar{\omega}_1 + \bar{\omega}_2 + \dots + \bar{\omega}_{m_\nu} \subset \omega_1 + \omega_2 + \dots + \omega_{m_{\nu+1}}$; a sequence of finite numbers m_1, m_2, \dots exists by Borel's theorem; it follows that when P varies in $\bar{\omega}_1 + \dots + \bar{\omega}_{m_\nu}$ there exists a minimum value k_ν of $r(P)$ different from zero. Call a point a $(\sigma+1)$ -point when it is contained in $\omega_1 + \omega_2 + \dots + \omega_{m_{\sigma+1}}$ but not in $\omega_1 + \omega_2 + \dots + \omega_{m_\sigma}$; take, for each σ -point P , a spherical neighbourhood containing P and of radius $\frac{1}{4}k_\sigma$ and in this a neighbourhood

¹²⁾ „Zum Metrisationsproblem“, *Math. Annalen* 94, p. 310, Hauptsatz; extended by Tychonoff to regular spaces, *Math. Annalen* 95.

$\omega_{\nu(P, \sigma)}$ of P of the original system $\omega_1, \omega_2, \dots$. The set of neighbourhoods such as $\omega_{\nu(P, \sigma)}$ satisfies the required condition.

The following investigation, which is a step toward the proof that all locally simplicial manifolds are simplicial, proves that certain locally simplicial manifolds can be covered "once positively" (see Theorem 1, § 2 beneath) by a unique continuous representation of a simplicial manifold, and extends to locally simplicial manifolds the fundamental theorem¹³⁾ and results of Brouwer on the representation of simplicial manifolds.

In the element E' of a locally simplicial manifold consider a segment path¹⁴⁾ of which a segment PQ crosses an $(n-1)$ -dimensional simplex $A_1 A_2 \dots A_n$ in E' of positive indicatrix A_1, A_2, \dots, A_n ; let Q be the end point of PQ for a positive sense of description of the path, that is, the path is regarded as a two-sided, 1-dimensional simplicial manifold in which P, Q is the positive indicatrix of the element PQ ; then the above mentioned crossing is defined to be positive or negative in E' according as A_1, A_2, \dots, A_n, Q is the positive or negative indicatrix of the simplex $A_1 A_2 \dots A_n Q$ in E' respectively¹⁵⁾.

The boundary B of an n -dimensional element E contained in an element E' is a particular case of a Jordan Manifold¹⁶⁾. Let P be a point of E , S_P an $(n-1)$ -dimensional sphere¹⁷⁾ of E' of centre P , B_{ν} ($\nu = 1, 2, \dots$) a sequence of simplicial divisions¹⁸⁾ of B in which $B_{\nu+1}$ is a simplicial division of B_{ν} , and the breadth of the simplexes of B_{ν}

¹³⁾ „A. v. M.“, p. 106, Satz 1; in the above terminology this theorem may be written: If a closed, two-sided, n -dimensional, simplicial manifold μ be uniquely and continuously represented on a simplicial n -dimensional manifold μ' , there exists a finite whole number c , invariant under continuous modification of the representation, with the property that the image of μ covers every region of μ' altogether c times positively; if μ' be one-sided or open c is always zero.

This number c is called the degree of the representation.

¹⁴⁾ As with simplicial manifolds, Brouwer, „A. v. M.“, p. 100, assign to each element of a locally simplicial manifold, a regular Euclidean simplex of fixed length of edge as its „representative simplex“, and let there be a topological correspondence between the element and its representative simplex; then by a segment, segment path, component simplex, $(n-1)$ -dimensional simplex in E' , is understood the image of a segment, segment path, component simplex, $(n-1)$ -dimensional simplex respectively, in the representative simplex of E' .

¹⁵⁾ Brouwer, „Beweis des n -dimensionalen Jordanschen Satzes“, Math. Annalen 71, p. 317, footnote.

¹⁶⁾ Brouwer, „Beweis des n -dimensionalen Jordanschen Satzes“, Math. Annalen 71, p. 314.

¹⁷⁾ By an $(n-1)$ -dimensional sphere of E' is understood the image in E' of an $(n-1)$ -dimensional sphere in the representative simplex of E' .

¹⁸⁾ „A. v. M.“, p. 101.

does not exceed $\frac{1}{2^r}$; assign a positive indicatrix to B and E' , determine the indicatrix of the interior of S_P from that of E' ¹⁹⁾ and that of S_P from that of its interior²⁰⁾, and represent B uniquely and continuously on S_P by projection with centre P ; the degree of this representation is called the order²¹⁾ of P with respect to B in E' ; for sufficiently great r it is identical with the difference between the number of positive and negative crossings of B_r by a segment path from the boundary of E' to P , no vertex or segment of which is in B_r and no segment of which cuts an $(n-2)$ -dimensional face of B_r ; this order has the value ± 1 .²²⁾

Let the system of neighbourhoods of a locally simplicial manifold be replaced by an equivalent⁹⁾ system any two neighbourhoods of which having common points are contained in a single element (see remark 4 above) and thus constitute a two-sided piece of the manifold. Consider two elements E' and E'' such that $E'E'' \neq 0$ and such that a third element $E \subset E'E''$ exists²³⁾. The indicatrix of E' being assigned, that of E is determined so that the order of a point $P \subset E$ with respect to the boundary of E in E' is $+1$ and that of E'' is so assigned that the order of P with respect to the boundary of E in E'' is also $+1$ ²⁴⁾; this determination of the indicatrix of E'' is then independent of the choice of the element $E \subset E'E''$, both E' and E'' being contained in a single element of the manifold. Finally two arbitrary elements E' and E'' of the manifold can be joined by a finite chain any two consecutive elements of which have a third element in common²⁵⁾; if then the indicatrix of the first element E' of this chain be assigned, that of the last element E'' can be determined by successive applications of the above rule; when the indicatrix thus assigned to E'' is independent of the chain of elements joining E' and E'' the locally simplicial manifold is said to be two-sided; otherwise one sided²⁴⁾.

¹⁹⁾ „A. v. M.“, p. 100.

²⁰⁾ „A. v. M.“, p. 108. The indicatrix of the $(n-1)$ -dimensional simplex $A_1 A_2 \dots A_n$ considered as a face of the simplex $A_1 A_2 \dots A_{n+1}$ is defined to be A_1, A_2, \dots, A_n where $A_1, A_2, \dots, A_n, A_{n+1}$ is the indicatrix of $A_1 A_2 \dots A_{n+1}$

²¹⁾ Brouwer, „Über Jordansche Mannigfaltigkeiten“, Math. Annalen 71, p. 323, § 4.

²²⁾ Brouwer, „Über Jordansche Mannigfaltigkeiten“, Math. Annalen 71, p. 323, § 4, Satz 4; we have above used a particular case of this theorem.

²³⁾ Such an element exists when E' and E'' have inner points in common; Hausdorff, Axiom (B).

²⁴⁾ We are here using Brouwer's generalized indicatrix, „Über Jordansche Mannigfaltigkeiten“, p. 324, § 5; for the extension of the indicatrix conception to locally simplicial manifolds I am indebted to Prof. Brouwer personally.

²⁵⁾ See remark 2 above.

If now α be a unique continuous representation of the closed, two-sided, n -dimensional pseudo-manifold or simplicial manifold μ on the closed, two-sided, n -dimensional locally simplicial manifold μ' , we can define and use the simplicial division ζ of μ , the simplicial and modified simplicial representations β and γ respectively corresponding to ζ and approximating to α , the "inner simplex" of each element of μ' , the ordinary points of these inner simplexes and the numbers p and p' of positive and negative image simplexes respectively covering an ordinary point P of μ' , exactly as with simplicial manifolds, obtaining the results²⁶⁾:

(1.1) The number $c = p - p'$ is constant in the ordinary points of an inner simplex of μ' ;

(1.2) The number c has the same value for every simplicial approximation to α .

Again let E' and E'' be neighbourhoods of μ' such that $E' E'' \neq 0$ and let the corresponding inner simplexes J' and J'' be so chosen that $J' J'' \neq 0$; construct a sequence ζ_1, ζ_2, \dots of simplicial divisions of μ for which ζ_{i+1} is a simplicial division of ζ_i and the maximum diameter of any element of ζ_i is $< \frac{1}{2^i}$; if then β'_i and β''_i be the simplicial approximations to α in J' and J'' respectively corresponding to ζ_i , P any ordinary point of J' and J'' with respect to both β'_i and β''_i , and c'_i and c''_i the values of $(p - p')$ at P for β'_i and β''_i respectively, it can be proved that for sufficiently great λ , $c'_i = c''_i$ ²⁷⁾, and hence that —

(1.3) The number c has the same value in the ordinary points of every inner simplex of the locally simplicial manifold μ' .²⁷⁾

As with (1.1) and (1.2), the results (1.4) and (1.5) beneath, follow exactly as for simplicial manifolds²⁸⁾:

(1.4) When the image of μ under α is not everywhere dense in μ' , the degree c is zero;

(1.5) If μ' be one-sided or open c is zero.

With (1.1), (1.2), ..., (1.5), the fundamental theorem²⁹⁾ is established for the case where μ' is locally simplicial and μ a pseudo or a simplicial manifold.

²⁶⁾ „A. v. M.“, p. 101–105.

²⁷⁾ W. Wilson, „Representation of a simplicial manifold on a locally simplicial manifold“, Amsterdam Proceedings 29 (1926), p. 1129 sqq.; for the leading idea of the proof there given the writer was indebted to a remark of Prof. Brouwer.

²⁸⁾ „A. v. M.“, p. 106.

²⁹⁾ „A. v. M.“, p. 106, Satz 1; see footnote ¹³⁾ above.

Part I.

Representation of locally simplicial manifolds on each other.

§ 2.

The object of the following investigation is to establish the fundamental theorem¹³⁾ and define the degree¹³⁾ of representation for the case when both manifolds may be only locally simplicial. It is required to prove the

Theorem 1. *To any closed, two-sided, n -dimensional, locally simplicial manifold M , there can be assigned a closed, two-sided, n -dimensional pseudo-manifold μ and a unique, continuous representation of μ on M of degree $+1$, that is, such that the image of μ covers M once positively.*

Among the countably infinite system of n -dimensional neighbourhoods defining the manifold M , a finite subset E_1, E_2, \dots, E_m exists such that $E_1 + E_2 + \dots + E_m > M$ and E_i contains points of $E_1 + E_2 + \dots + E_{i-1}$, ($i = 2, 3, \dots, m$)³⁰⁾; in each element E_i there can then be assigned m simplexes $J_{i1}, J_{i2}, \dots, J_{im}$ such that J_{i1} is an inner simplex of E_i , J_{i2} an inner simplex of J_{i1} , and so on, such that $J_{1m} + J_{2m} + \dots + J_{mm} > M$ and further such that J_{im} and J_{jm} have points in common when this is the case for E_i and E_j . Consider a simplicial division ζ_1 of E_1 into the set of simplexes

$$M^{(1)} \equiv \{U_i\} + \{V_i\}$$

of which each V_i is contained in E_2 and each U_i contains points not in E_2 ; let this division be of such fineness that when any simplex of $M^{(1)}$ contains a point of J_{i1} it is contained in E_i and when it contains a point of J_{i2} it is contained in J_{i1} , and so on, ($i = 1, 2, \dots, m$). Assign a positive indicatrix in the manifold M and thereby, by the method described in the introduction, to each U_i and V_i so that $\{U_i\} + \{V_i\}$ covers E_1 once positively. The division ζ_1 is such that no U_i can contain a point of J_{21} for otherwise it would be contained in E_2 ; it follows that $\{V_i\}$ covers $E_1 J_{21}$ once positively. Let $V_i^{(2)}$ be the simplicial image of V_i in E_2 ³¹⁾ and assign as its indicatrix the positive indicatrix of V_i in $M^{(1)}$; when P is a point of E_2 it is said to be covered once positively or negatively by $V_i^{(2)}$ according as the indicatrix just determined for $V_i^{(2)}$ is positive or nega-

³⁰⁾ Tibor Radó, loc. cit., Hilfsatz 1; see also remark 2 in the introduction.

³¹⁾ That is, $V_i^{(2)}$ is a simplex of E_2 with the same vertices as V_i ; we recall (footnote ¹⁴⁾ above) that a simplex of E_2 is the image in E_2 of a simplex in the representative simplex of E_2 .

tive respectively in E_2 . The boundary³²⁾ of $\{V_i\}$ consists of two sets of $(n-1)$ -dimensional faces $\{F_u\}$ and $\{F_v\}$, each face F_u being common to a simplex U_i and a simplex V_j while each F_v belongs only to a simplex V_j ; let $F_u^{(2)}$ and $F_v^{(2)}$ be the corresponding faces of $\{V_i^{(2)}\}$. Since U_i contains no point of J_{21} , it follows for a sufficiently fine division ζ_1 that

(a) $\{F_u^{(2)}\}$ contains no point of J_{21} . Again, since $\{U_i\} + \{V_i\}$ covers E_1 once positively and F_v is in the boundary of $\{U_i\} + \{V_i\}$, we have, for a sufficiently fine division ζ_1

(b) $\{F_v^{(2)}\}$ contains no point of J_{11} .

From (a) and (b) we see that the boundary $\{F_u^{(2)}\} + \{F_v^{(2)}\}$ of $\{V_i^{(2)}\}$ contains no point of $J_{11} J_{21}$ so that the number c for the simplicial representation $\{V_i^{(2)}\}$ has the same value in the ordinary points of $J_{11} J_{21}$; both for the identical simplicial representation $\{V_i\}$, c has the value $+1$ in $E_1 J_{21}$, therefore from (1.3) for the case where $c = +1$, the value of c for the representation $\{V_i^{(2)}\}$ is $+1$ for all ordinary points of $J_{11} J_{21}$, i. e.,

(2.1) $\{V_i^{(2)}\}$ covers $J_{11} J_{21}$ once positively.

Consider a simplicial division ζ_2 of E_2 into the set of simplexes

$$\{X_i\} + \{Y_i\} \text{ and other simplexes,}$$

satisfying the following conditions —

(2.2) Each $V_i^{(2)}$ is composed of one or more of the X_i and each X_i is contained in one or more of the $V_i^{(2)}$;

(2.3) $\{X_i\} + \{Y_i\} > J_{21}$, i. e., the boundary of $\{X_i\} + \{Y_i\}$ is contained in $E_2 - J_{21}$;

(2.4) When any simplex X_i or Y_i contains a point of J_{h1} or of $J_{h, k+1}$ it is contained in E_h or $J_{h, k}$ respectively, ($h = 1, 2, \dots, m$; $k = 1, 2, \dots, m-1$)

(2.5) Each Y_i contains points of J_{21} .

This simplicial division is constructed as follows —

Let $\pi_1, \pi_2, \dots, \pi_2$ be the intersections with E_2 of those flat $(n-1)$ -dimensional spaces containing the $(n-1)$ -dimensional faces of the simplexes $V_i^{(2)}$. In each two dimensional face of E_2 containing points of π_1 take a point of π_1 and in those of the remaining two dimensional faces of E_2 which have one or more divided edges take an arbitrary point; from the points thus introduced, by projection on the already constructed

³²⁾ By the boundary of a set of simplexes among which the incidence relations are assigned is understood the set of those $(n-1)$ -dimensional faces which are incident with only one simplex.

simplicial division of the edges, we obtain a simplicial division of certain of the 2-dimensional faces³³); the remaining 2-dimensional faces are undivided. Suppose this process to have been carried out for the p -dimensional faces, $p = 2, 3, \dots, h - 1$ in succession, and proceed thus: in each h -dimensional face of E_2 containing points of π_1 take a point of π_1 and in those of the remaining h -dimensional faces which have one or more divided $(h - 1)$ -dimensional faces take an arbitrary point; from the points thus introduced, by projection on the already constructed simplicial division of the $(h - 1)$ -dimensional faces, we obtain simplicial divisions of certain of the h -dimensional faces; the remaining h -dimensional faces are undivided. Proceeding in this way for $p = h + 1, h + 2, \dots, n$, a simplicial division of E_2 is obtained³³). Proceed with π_2 and each of the component simplexes of E_2 in the same way as with π_1 and E_2 thus obtaining a second subdivision of E_2 ; and so on with π_3, π_4, \dots , and π_l ; the simplexes of the l^{th} division are now further divided and if this further division be sufficiently fine, those of its simplexes each of which is contained in one or more of the $V_i^{(2)}$ are the X_i required, and those of the remaining simplexes which contain points of J_{21} are the Y_i required. Let each X_j be counted once for each $V_i^{(2)}$ containing it and written $V_{ij}^{(2)}$, the indicatrix of $V_{ij}^{(2)}$ being determined from that of $V_i^{(2)}$ by the rule for simplicial manifolds; $V_{ij}^{(2)}$ coincides with X_j in E_2 and is said to cover X_j positively or negatively according as its indicatrix is positive or negative respectively in E_2 . Thus the simplex $V_i^{(2)}$, when composed of the simplexes X_j, X_k, \dots, X_l of E_2 , is divided into the simplexes $V_{ij}^{(2)}, V_{ik}^{(2)}, \dots, V_{il}^{(2)}$; and the simplex X_i , when contained in $V_j^{(2)}, V_k^{(2)}, \dots, V_l^{(2)}$, is covered by $V_{ji}^{(2)}, V_{ki}^{(2)}, \dots, V_{li}^{(2)}$. An h -dimensional face of $V_{ij}^{(2)}$ is said to cover that h -dimensional face of X_j with which it coincides in E_2 . Referring to the definition of the $V_i^{(2)}$ we see that V_i is also the simplicial image of $V_i^{(2)}$ in E_1 ; let V_{ij} be the image of $V_{ij}^{(2)}$ in V_i , divide each U_i into simplexes U_{ij} so that the set of simplexes $\{U_{ij}\} + \{V_{ij}\}$ is a simplicial division of $\{U_i\} + \{V_i\}$ ³⁴) and determine the positive indicatrix of U_{ij} and V_{ij} from that of U_i and V_i respectively; then

$$(2.6) \quad \{U_{ij}\} + \{V_{ij}\} \text{ covers } E_1 \text{ once positively.}$$

The incidence relations among the $V_{ij}^{(2)}$ are defined to be the same as

³³) Use is being here made of the „gemischte Zerlegung“ of Brouwer, „Erweiterung des Definitionsbereichs einer stetigen Funktion“, Math. Annalen 79, p. 210.

³⁴) Again by means of Brouwer's „gemischte Zerlegung“; in this case, for $h = 1, 2, \dots, (n - 1)$ in succession, in each $(h + 1)$ -dimensional simplex of $\{U_i\}$ not yet divided but having one or more divided h -dimensional faces, take an arbitrary point and proceed as in the construction of ζ_2 above.

those among the corresponding V_{ij} , that is, any two simplexes $V_{ij}^{(2)}$ and $V_{kl}^{(2)}$ have an h -dimensional face in common when, and only when, this is the case for V_{ij} and V_{kl} ; for example $V_{ji}^{(2)}$ and $V_{ki}^{(2)}$ may have no common face although both covering the same simplex X_i . Let it be ensured by an arbitrarily small displacement of the vertices of $M^{(1)}$ that no $(n-1)$ -dimensional face of $\{X_i\} + \{Y_i\}$ is covered by more than one boundary face of the set $\{V_{ij}^{(2)}\}$; if then, of the simplexes $V_{ij}^{(2)}$ covering a variable point x of E_2 , p are positive and q negative in E_2 , the number $(p-q)$ remains constant or changes by unity when x crosses a face of $\{X_i\} + \{Y_i\}$ according as this face is not or is covered by a boundary face of $\{V_{ij}^{(2)}\}$.

§ 3.

When the simplex X_i contains points of $J_{11} J_{21}$ the numbers p and q of positive and negative covering simplexes $V_{ji}^{(2)}$ are, by (2.1), such that $(p-q) = +1$. When X_i contains no point of $J_{11} J_{21}$, $(p-q)$ may differ from $+1$; in this case let r be such that

$$(3.1) \quad p - q + r = +1$$

and, c being the absolute value of r , count each simplex X_i which contains points of J_{21} a further c times³⁵⁾ positively or negatively with respect to E_2 according as r is positive or negative respectively, writing the new simplexes $X_{i1}, X_{i2}, \dots, X_{ic}$.

A simplex X_{ij} coincides with X_i in E_2 and in conformity with the previous terminology, is said to cover X_i positively or negatively according as the indicatrix assigned to it is positive or negative respectively in E_2 ; similarly an h -dimensional face of X_{ij} is said to cover that face of X_i with which it coincides in E_2 . Any simplex X_i which contains points of J_{21} is thus covered once positively by the set simplexes, $p+q+c$ in number, $V_{ji}^{(2)}, V_{ki}^{(2)}, \dots, V_{li}^{(2)}, X_{i1}, X_{i2}, \dots, X_{ic}$.

Finally cover each Y_i by a simplex Y_{i1} which is positive in E_2 . The set of simplexes

$$M^{(2)} \equiv \{V_{ij}^{(2)}\} + \{X_{ij}\} + \{Y_{i1}\}$$

then covers J_{21} once positively.

§ 4.

The incidence relations among the simplexes of $M^{(2)}$ must now be so assigned that the boundary of $M^{(2)}$ is contained in $E_2 - J_{21}$. For each

³⁵⁾ This multiplication of elements was suggested by the duplication of elements used by Brouwer. "Transformations of Projective Spaces", Amsterdam Proceedings 29 (1926), No. 6.

$(n - 1)$ -dimensional face F of $\{X_i\} + \{Y_i\}$ we assign the incidence relations among those simplexes of $M^{(2)}$ having $(n - 1)$ -dimensional faces covering F .

The incidence relations among the Y_{i1} are defined to be the same as those among the corresponding Y_i , *i. e.*, when two simplexes Y_i and Y_j have an h -dimensional face in common the simplexes Y_{i1} and Y_{j1} have a common h -dimensional face covering the first mentioned face; thus it is only necessary to consider the case when F is a face of $\{X_i\}$.

(a) Let F be a boundary face of $\{X_i\} + \{Y_i\}$ belonging to a simplex X_j ; the numbers p, q , and c of covering simplexes for X_j are written p_j, q_j , and c_j .

1. Let F be covered by a boundary face Γ of $\{V_{ij}^{(2)}\}$ belonging to the simplex $V_{ij}^{(2)}$ only; then $p_j - q_j = +1$ or -1 ³⁶⁾ according as $V_{ij}^{(2)}$ is positive or negative in E_2 respectively. When $p_j - q_j$ is $+1$, Γ is defined to belong to $V_{ij}^{(2)}$ only and is thus a boundary face of $M^{(2)}$. When $p_j - q_j = -1$ and X_j contains no point of J_{21} there are no simplexes X_{jk} covering X_j and Γ is again defined to belong only to $V_{ij}^{(2)}$ and is again a boundary face of $M^{(2)}$. When $p_j - q_j = -1$ and X_j contains points of J_{21} , then X_j is covered by two positive (in E_2) simplexes X_{j1} and X_{j2} ; in this case Γ is assigned to be a common face of $V_{ij}^{(2)}$ and X_{j1} covering F , while that face of X_{j2} which covers F is assigned to belong to X_{j2} only and thus to be a boundary face of $M^{(2)}$.

2. When F is not covered by a boundary face of $\{V_{ij}^{(2)}\}$, those $(n - 1)$ -dimensional faces of the covering simplexes X_{jk} (when the latter exist) which cover F are defined to be boundary faces of $M^{(2)}$.

The boundary faces of $M^{(2)}$ above defined cover a boundary face F of $\{X_i\} + \{Y_i\}$ and are therefore, by (2.3), contained in $E_2 - J_{21}$.

(b) Let F be a boundary face of $\{X_i\}$, but not of $\{X_i\} + \{Y_i\}$, common to the two simplexes X_j and Y_i .

1. Let F be covered by a boundary face Γ of $\{V_{ij}^{(2)}\}$ belonging to $V_{ij}^{(2)}$ only. When $V_{ij}^{(2)}$ is positive in E_2 , $p_j - q_j = +1$ and thus there are no simplexes X_{jk} covering X_j . Γ is then defined as the common face of $V_{ij}^{(2)}$ and Y_{i1} covering F . When $V_{ij}^{(2)}$ is negative in E_2 , $p_j - q_j = -1$ and when X_j contains points of J_{21} there are two simplexes X_{j1} and X_{j2} covering X_j and positive in E_2 ; Γ is then assigned as a common face of $V_{ij}^{(2)}$ and X_{j1} , while X_{j2} and Y_{i1} are defined to have a common face which covers F ; when X_j contains no point of J_{21} so that there are no covering simplexes X_{j1}, X_{j2}, \dots , then Γ and that face of Y_{i1} which covers

³⁶⁾ Since the values of $p_j - q_j$ on different sides of F differ by unity, and on that side of F not in X_j both numbers p and q are zero.

F are assigned as boundary faces of $M^{(2)}$, *i. e.*, belonging only to $V_{ij}^{(2)}$ and Y_{i1} respectively.

2. When F is not covered by a boundary face of $\{V_{ij}^{(2)}\}$, $p_j - q_j = 0$. When X_j contains no point of J_{21} there are no covering simplexes X_{jk} and that face of Y_i which covers F is defined to belong to Y_{i1} only, and thus to be a boundary face of $M^{(2)}$; when X_j contains points of J_{21} there is one covering simplex X_{j1} which is positive in E_2 , and X_{j1} and Y_{i1} are defined to have a common face covering F . In both 1. and 2. the boundary faces of $M^{(2)}$ cover a face of a simplex X_j containing no point of J_{21} and are therefore contained in $E_2 - J_{21}$.

(c) Let F be the common face of X_i and X_j .

1. When F is covered by a boundary face F of $\{V_{ij}^{(2)}\}$ it is sufficient to consider the case when Γ belongs to a simplex $V_{ki}^{(2)}$ covering X_i . Since $p_i - q_i$ and $p_j - q_j$ differ by unity the covering simplexes $X_{i1}, X_{i2}, \dots, X_{j1}, X_{je}$ when they exist, have all the same sign in E_2 ; when this sign is that of $V_{ki}^{(2)}$ the incidence relations are assigned so that

$V_{ki}^{(2)}$ and X_{j1} have a common face Γ covering F ,

X_{i1} and X_{j2} have a common face covering F ,

and so on, until either the simplexes $V_{ki}^{(2)}, X_{i1}, \dots, X_{ie}$ or $X_{j1}, X_{j2}, \dots, X_{je}$ are exhausted; those faces of the simplexes left over which cover F are then defined to be boundary faces of $M^{(2)}$; when the above mentioned sign differs from that of $V_{ki}^{(2)}$, the incidence relations are assigned that

$V_{ki}^{(2)}$ and X_{i1} have the common face Γ covering F ,

X_{j1} and X_{i2} have a common face covering F ,

and so on, until either the simplexes $V_{ki}^{(2)}, X_{j1}, X_{j2}, \dots, X_{je}$ or $X_{i1}, X_{i2}, \dots, X_{ie}$ are exhausted; those faces of the simplexes left over which cover F are then defined to be boundary faces of $M^{(2)}$; from (3.1) there are only simplexes left over when either X_i or X_j contains no points of J_{21} , *i. e.*, only when F is contained in $E_2 - J_{21}$; hence the boundary faces of $M^{(2)}$ just defined are contained in $E_2 - J_{21}$.

2. When F is not covered by a boundary face of $\{V_{ij}^{(2)}\}$, the incidence relations among the covering simplexes $X_{i1}, X_{i2}, \dots, X_{ie}, X_{j1}, X_{j2}, \dots, X_{je}$ are so assigned that:

X_{i1} and X_{j1} have a common face covering F ,

X_{i2} and X_{j2} have a common face covering F ,

and so on until either the set $X_{i1}, X_{i2}, \dots, X_{ie}$ or the set $X_{j1}, X_{j2}, \dots, X_{je}$ is exhausted, when, of the simplexes left over those faces which cover F are defined to be boundary faces of $M^{(2)}$; again from (3.1), there are

only such simplexes left over when either X_i or Y_j , and thus F , contains no point of J_{21} , hence the boundary faces thereby defined are contained in $E_2 - J_{21}$.

Thus the boundary of $M^{(2)}$ is contained in $E_2 - J_{21}$; further, any two simplexes of $M^{(2)}$ have either no point or an h -dimensional face in common, ($h = 0, 1, \dots, n - 1$).

§ 5.

To each simplex of $M^{(2)}$ there has already been assigned a unique indicatrix, namely such that $M^{(2)}$ covers J_{21} once positively; call this indicatrix the positive indicatrix. Referring to the above assignment of incidence relations we now see that this indicatrix is of opposite or the same sign in E_2 for simplexes of common face according as they cover the same simplex or adjacent simplexes of $\{X_i\} + \{Y_i\}$; but this relation between the signs in E_2 of indicatrices of elements with a common face is that given by the usual rule for simplicial manifolds, hence, given the indicatrix of an element of $M^{(2)}$, this rule determines also a unique indicatrix for each element of $M^{(2)}$ which can be joined to the first element by a chain consecutive elements of which have a common face, *i. e.*, $M^{(2)}$ is two sided. Further the above argument shows that when a positive indicatrix is assigned to one element of $M^{(2)}$ the unique indicatrix thereby determined for every element of $M^{(2)}$ which can be joined to the first mentioned element by a chain of elements of the type above used, is the above defined positive indicatrix.

The vertices of $\{V_{ij}^{(2)}\}$ are now displaced into the corresponding vertices of $\{\bar{V}_{ij}^{(2)}\}$ whereby certain vertices of $\{X_{ij}\} + \{Y_{i1}\}$ are also displaced; let $\bar{V}_{ij}^{(2)}$, \bar{X}_{ij} and \bar{Y}_{i1} be the resulting modified simplicial images in E_2 of $V_{ij}^{(2)}$, X_{ij} and Y_{i1} respectively, the indicatrices of these new simplexes being derived from those of the corresponding old simplexes by writing the new vertices in place of the old; then writing

$$\bar{M}^{(2)} \equiv \{\bar{V}_{ij}^{(2)}\} + \{\bar{X}_{ij}\} + \{\bar{Y}_{i1}\},$$

we have from the above properties of $M^{(2)}$ —

(5.1) $\bar{M}^{(2)}$ is two-sided; any two of its simplexes have either no point or an h -dimensional face in common and for sufficiently fine ζ_1 and ζ_2

(5.2) The boundary of $\bar{M}^{(2)}$ is contained in $E_2 - J_{22}$, *i. e.*, $\bar{M}^{(2)}$ covers J_{22} once positively.

We now prove that any simplex X_{ij} or Y_{i1} of $M^{(2)}$ can be joined to a simplex of $\{V_{ij}^{(2)}\}$ by a chain in $M^{(2)}$ consecutive simplexes of which have an $(n - 1)$ -dimensional face in common. Consider first any simplex X_{ij} and a simplex $V_{hk}^{(2)}$ which contains points of $J_{11} J_{22}$; these two sim-

plexes cover the simplexes X_i and X_k of $\{X_i\} + \{Y_i\}$ respectively, which (by 2.5) can be joined by a chain of simplexes of $\{X_i\} + \{Y_i\}$; let this chain be written

$$X_i, X_1, X_2, \dots, X_\alpha, Y_1, Y_2, \dots, Y_\beta, \dots, \Delta_\gamma, X_k,$$

consecutive simplexes having a common $(n-1)$ -dimensional face and Δ_γ being a simplex X_γ or Y_γ ; further, since $V_{hk}^{(2)}$ contains points of J_{22} , this chain may be so chosen that each face common to two consecutive simplexes contains points of J_{21} and is therefore not covered by a boundary face of $M^{(2)}$. Referring to the incidence relations in $M^{(2)}$ we see that no two simplexes X_{i_m} and X_{i_n} covering the same simplex X_i have a common $(n-1)$ -dimensional face; hence X_{ij} has either a common face with a simplex $V_{\lambda i}^{(2)}$ covering X_i , or with a simplex $V_{\lambda 1}^{(2)}$ covering X_1 , or with a simplex X_{1r} covering X_1 ; we have thus a chain $X_{ij}, V_{\lambda i}^{(2)}$, or a chain $X_{ij}, V_{\lambda 1}^{(2)}$, or a chain X_{ij}, X_{1r} ; in the last case X_{1r} has a common face with some simplex $V_{\lambda 1}^{(2)}$ or $V_{\lambda 2}^{(2)}$ covering X_1 or X_2 respectively, or with a simplex X_{2s} covering X_2 , so that we have one of the chains $X_{ij}, X_{1r}, V_{\lambda 1}^{(2)}$, or $X_{ij}, X_{1r}, V_{\lambda 2}^{(2)}$ or X_{ij}, X_{1r}, X_{2s} ; continuing this process, and noting that $X_{\alpha t}$ has a common face either with some simplex $V_{\lambda \alpha}^{(2)}$ covering X_α or with Y_{11} , that Y_{11} has a common face with Y_{21} and so on, we either obtain a chain containing a simplex of $\{V_{ij}^{(2)}\}$ before reaching X_k or the chain

$$X_{ij}, X_{1r}, X_{2s}, \dots, X_{\alpha t}, Y_{11}, \dots, Y_{\beta 1}, \Delta_{\gamma p}, V_{hk}^{(2)}$$

(where $\Delta_{\gamma p}$ is $X_{\gamma p}$ or $Y_{\gamma 1}$) joining X_{ij} and $V_{hk}^{(2)}$. In the same way, a chain joining any simplex Y_{i1} to a simplex $V_{hk}^{(2)}$ containing points of $J_{11} J_{21}$ can be constructed, hence from the definition of $\bar{M}^{(2)}$ we have.

(5.3) Any simplex of $\bar{M}^{(2)}$ can be joined to a simplex of $\{\bar{V}_{ij}^{(2)}\}$ by a chain in $\bar{M}^{(2)}$ consecutive simplexes of which have a common $(n-1)$ -dimensional face.

By definition, $\bar{V}_{ij}^{(2)}$ is the simplicial image of V_{ij} in E_2 , hence when $\bar{V}_{ij}^{(2)}$ and $\bar{V}_{kl}^{(2)}$ have a common h -dimensional face, V_{ij} and $V_{kl}^{(2)}$ have the $(h+1)$ vertices of this face in common; if therefore we write

$$M^{(12)} \equiv \{U_{ij}\} + \{V_{ij}\} + \{\bar{X}_{ij}\} + \{\bar{Y}_{i1}\},$$

we have from (5.3) and (5.1)

(5.4) Any two simplexes of $M^{(12)}$ have either no point or an h -dimensional face or the $(h+1)$ vertices of such a face in common and are joined in $M^{(12)}$ by a chain, consecutive simplexes of which have either an $(n-1)$ -dimensional face or n vertices in common.

It is required to prove that, when in any chain of simplexes of

$M^{(1,2)}$, consecutive simplexes of which have n vertices in common³⁷), the indicatrices are related by the rule used in simplicial manifolds, the assignment of an indicatrix to one simplex of $M^{(1,2)}$ determines a unique indicatrix for every simplex; when this is the case we shall say that $M^{(1,2)}$ is two-sided.

Indicatrices are already uniquely assigned to each simplex of $\{U_{ij}\} + \{V_{ij}\}$ and $\bar{M}^{(2)} \equiv \{\bar{V}_{ij}^{(2)}\} + \{\bar{X}_{ij}\} + \{\bar{Y}_{i1}\}$, namely such that $\{U_{ij}\} + \{V_{ij}\}$ covers E_1 once positively and $\bar{M}^{(2)}$ covers $J_{2,2}$ once positively; call these the positive indicatrices; then by their definition the positive indicatrices of V_{ij} and $\bar{V}_{ij}^{(2)}$ are identical for each V_{ij} ³⁸).

The two-sidedness of $\bar{M}^{(2)}$ is such that if the positive indicatrix be assigned to one of its simplexes the unique indicatrix determined for every simplex which can be joined to the first by a chain in $\bar{M}^{(2)}$ is also the positive indicatrix. Consider any simplex V_{ij} of $\{V_{ij}\}$ and any simplex Δ of $\{\bar{V}_{ij}^{(2)}\} + \{\bar{X}_{ij}\} + \{\bar{Y}_{i1}\}$ joined to V_{ij} by a chain in $\{V_{ij}\} + \{\bar{X}_{ij}\} + \{\bar{Y}_{i1}\}$; and the corresponding chain in $\bar{M}^{(2)}$ joining $\bar{V}_{ij}^{(2)}$ and the simplex $\bar{\Delta}$ corresponding to Δ . Since elements of the first chain have n vertices in common when the corresponding elements of the second have a common $(n-1)$ -dimensional face, both chains determine the same indicatrix for the simplexes Δ and $\bar{\Delta}$; hence referring to the above mentioned connection of the two-sidedness of $\bar{M}^{(2)}$ with its positive indicatrix and the identity of the positive indicatrices of V_{ij} and $\bar{V}_{ij}^{(2)}$ we have — from the positive indicatrix of V_{ij} the rule for simplicial manifolds determines a unique indicatrix, which is identical with the positive indicatrix, for each simplex Δ which is joined to V_{ij} by a chain in $\{V_{ij}\} + \{\bar{X}_{ij}\} + \{\bar{Y}_{i1}\}$. By definition, the unique indicatrix determined for each simplex of $\{U_{ij}\} + \{V_{ij}\}$ from V_{ij} by a chain wholly in $\{U_{ij}\} + \{V_{ij}\}$ is the positive indicatrix. Consider finally any chain in $M^{(1,2)}$ joining V_{ij} to an arbitrary simplex $\Delta^{(1,2)}$ of $M^{(1,2)}$; let this chain be written $V_{ij}, \Delta_1, \dots, \Delta_k, \dots, \Delta_l, \dots, \Delta_m, \dots, \Delta_n, \Delta^{(1,2)}$ where k, l, m, n, \dots are such that

- none of the simplexes $V_{ij}, \Delta_1, \dots, \Delta_k$ is in $\{U_{ij}\}$,
- each of the simplexes $\Delta_{k+1}, \dots, \Delta_l$ is in $\{U_{ij}\}$,
- none of the simplexes $\Delta_{l+1}, \dots, \Delta_m$ is in $\{U_{ij}\}$,
- each of the simplexes $\Delta_{m+1}, \dots, \Delta_n$ is in $\{U_{ij}\}$.

³⁷) Such a chain of simplexes shall be referred to briefly as a chain of simplexes or merely as a chain.

³⁸) To see this we refer to the definition of $\bar{V}_{ij}^{(2)}$ given above; and to § 2 where V_{ij} and $V_{ij}^{(2)}$ are corresponding component simplexes of V_i and $V_i^{(2)}$, the indicatrices of V_i and $V_i^{(2)}$ being defined as identical and those of V_{ij} and $V_{ij}^{(2)}$ derived from those of V_i and $V_i^{(2)}$, i. e., from the same indicatrix.

and so on; but no U_{ij} has n vertices in common with any simplex of $\{\bar{X}_{ij}\} + \{\bar{Y}_{i1}\}$, hence $\Delta_k, \Delta_{l+1}, \Delta_m, \Delta_{n+1}, \dots$ are all simplexes of $\{V_{ij}\}$ from which it follows that each of the chains $V_{ij}, \dots, \Delta_k; \Delta_k, \dots, \Delta_{l+1}; \Delta_{l+1}, \dots, \Delta_m$ and $\Delta_m, \dots, \Delta_n$ begins with a simplex of $\{V_{ij}\}$ and is wholly contained either in $\{U_{ij}\} + \{V_{ij}\}$ or in $\{V_{ij}\} + \{\bar{X}_{ij}\} + \{\bar{Y}_{i1}\}$; the first of these chains determines a positive indicatrix for Δ_k , the second then determines a positive indicatrix for Δ_{l+1} , the third a positive indicatrix for Δ_m , and so on until finally the indicatrix assigned to $\Delta^{(12)}$ is also the positive indicatrix. Hence

(5.5) $M^{(12)}$ is two-sided and when the indicatrix assigned to a given element of $M^{(12)}$ is the above mentioned positive indicatrix, the indicatrix thereby uniquely determined for each element of $M^{(12)}$, is the positive indicatrix.

Finally from (2.6) and (5.2), for sufficiently fine ζ_1 and ζ_2 ,

(5.6) The simplicial approximation to $M^{(12)}$ in E_i ($i = 1, 2$) covers J_{i2} once positively.

§ 6.

Assume now the construction of a set $M^{(12\dots l)}$ of simplexes in which the incidence relations are such that it has the properties (5.4) and (5.5) of $M^{(12)}$ and the property analogous to (5.6), viz.:

(6.1) The simplicial approximation to $M^{(12\dots l)}$ in E_i covers J_{il} once positively, ($i = 1, 2, \dots, l$).

Let $M^{(12\dots l)} \equiv \{U_i\} + \{V_i\}$, where each simplex V_i is contained in E_{l+1} and each U_i contains points not in E_{l+1} ; assume further the maximum diameters of each U_i and V_i to be such that when one of these simplexes contains a point of J_{j1} it is contained in E_j and that when it contains a point of $J_{j,k+1}$ it is contained in $J_{j,k}$, ($j = 1, 2, \dots, m$; $k = 1, 2, \dots, m - 1$), from which follows that no U_i contains a point of $J_{l+1,1}$. As in the previous paragraphs $U_i^{(k)}$ and $V_i^{(k)}$ represent the simplicial images of U_i and V_i respectively in E_k , while the indicatrices assigned to $U_i^{(k)}$ and $V_i^{(k)}$ are the positive indicatrices of U_i and V_i respectively in $M^{(12\dots l)}$; hence from (6.1) we have:

(6.2) $\{U_i^{(k)}\} + \{V_i^{(k)}\}^{39}$ covers J_{kl} once positively, ($k = 1, 2, \dots, l$); then since U_i contains no point of $J_{l+1,1}$, there follows:

³⁹ By $\{U_i^{(k)}\}$ is understood the set of all simplexes $U_i^{(k)}$, i. e., the suffix i takes all values for which the vertices of U_i are all in E_k ; e. g. in $\{U_i^{(1)}\}$ the summation is extended over a different set of the suffixes i from that in $\{U_i^{(2)}\}$; similarly with $\{V_i^{(k)}\}$.

(6.3) $\{V_i^{(k)}\}$ covers $J_{k,l}J_{l+1,1}$ once positively, ($k = 1, 2, \dots, l$). The incidence relations among the simplexes $V_i^{(l+1)}$ are such that when two simplexes V_i and V_j of $M^{(1, 2, \dots, l)}$ have $(h+1)$ vertices in common, $V_i^{(l+1)}$ and $V_j^{(l+1)}$ have the h -dimensional face of these vertices in common. The boundary of $\{V_i^{(l+1)}\}$ consists of two sets of $(n-1)$ -dimensional faces — faces such as F_u whose n vertices belong to a simplex U_i and faces such as F_v not having this property. Since U_i contains points not in E_{l+1} we have, when the simplexes U_i and V_i are taken sufficiently small,

(a) $\{F_u\}$ contains no point of $J_{l+1,1}$.

If F_v contain a point of $J_{k,l+1}$, ($k \leq l$), then it is contained in J_k ; therefore its simplicial image in E_k is contained in J_k , *i. e.*, a boundary face of $\{U_i^{(k)}\} + \{V_i^{(k)}\}$ lies in J_k ; this contradicts (6.2) hence:

(b) $\{F_v\}$ contains no point of $J_{k,l+1}$, ($k \leq l$).

From (a) and (b) the boundary of $\{V_i^{(l+1)}\}$ contains no point of $J_{k,l+1}J_{l+1,1}$, ($k \leq l$), hence from (1.3) and (6.3) —

(6.4) $\{V_i^{(l+1)}\}$ covers $(J_{1,l+1} + J_{2,l+1} + \dots + J_{l,l+1})J_{l+1,1}$ once positively. Let ζ_{l+1} be a simplicial division of E_{l+1} into simplexes

$\{X_i\}$, $\{Y_i\}$ and other simplexes,

of which the X_i and the Y_i satisfy the conditions analogous to (2.2), (2.3), ..., (2.5), namely

(6.5) Each $V_i^{(l+1)}$ is composed of simplexes X_i and each X_i is contained in one or more of the $V_i^{(l+1)}$;

(6.6) $\{X_i\} + \{Y_i\} > J_{l+1,1}$, *i. e.*, the boundary of $\{X_i\} + \{Y_i\}$ is contained in $E_{l+1} - J_{l+1,1}$;

(6.7) When any X_i or Y_i contains points of J_{k1}, J_{k2}, \dots , or J_{km} it is contained in E_k, J_{k1}, \dots , or $J_{k,m-1}$ respectively, ($k = 1, 2, \dots, m$); and

(6.8) Each Y_i contains points of $J_{l+1,1}$.

As in § 3, each X_i is counted once for each $V_i^{(l+1)}$ containing it and written $V_{ij}^{(l+1)}$, the indicatrix of $V_{ij}^{(l+1)}$ being derived from that of $V_i^{(l+1)}$; then by (6.4), $\{V_{ij}^{(l+1)}\}$ covers $(J_{1,l+1} + J_{2,l+1} + \dots + J_{l,l+1})J_{l+1,1}$ once positively so that when X_i contains points of this domain the numbers p_i and q_i of the covering simplexes $V_{ki}^{(l+1)}$ are such that $p_i - q_i = +1$. When X_i contains no point of $(J_{1,l+1} + J_{2,l+1} + \dots + J_{l,l+1})J_{l+1,1}$, the number $p_i - q_i$ may differ from $+1$; in this case let r_i be such that $p_i - q_i + r_i = +1$, let c_i be the absolute value of r_i and cover each X_i which contains points of $J_{l+1,1}$ by the c_i simplexes $X_{i1}, X_{i2}, \dots, X_{ic_i}$ whose indicatrices have the same sign in E_{l+1} as r_i ; finally each Y_i is covered by one simplex Y_{i1} whose indicatrix is positive in E_{l+1} . Then

the set of simplexes

$$M^{(l+1)} \equiv \{V_{ij}^{(l+1)}\} + \{X_{ij}\} + \{Y_{i1}\}$$

covers $J_{l+1,1}$ once positively.

Let V_{ij} be the image of $V_{ij}^{(l+1)}$ in V_i and divide each U_i into simplexes U_{ij} in such a way that, when $U_{ij}^{(k)}$ and $V_{ij}^{(k)}$ are the images of U_{ij} and V_{ij} in $U_i^{(k)}$ and $V_i^{(k)}$ respectively, we have:

$$\{U_{ij}^{(k)}\} + \{V_{ij}^{(k)}\} \text{ is a simplicial division of } \{U_i^{(k)}\} + \{V_i^{(k)}\} \\ (k = 1, 2, \dots, l).$$

Each vertex of $\{U_{ij}\} + \{V_{ij}\}$ is now displaced into the corresponding vertex of $\{U_{ij}^{(h)}\} + \{V_{ij}^{(h)}\}$, where h is the least integer for which this corresponding vertex exists, and each simplex U_{ij} and V_{ij} is replaced by the simplex \bar{U}_{ij} or \bar{V}_{ij} , of any element of M , whose vertices are the displaced vertices of U_{ij} or V_{ij} respectively. Writing $\bar{M}^{(1,2,\dots,l)} \equiv \{\bar{U}_{ij}\} + \{\bar{V}_{ij}\}$, since $M^{(1,2,\dots,l)}$ satisfies conditions (5.4) and (5.5) we have:

(6.9) Any two simplexes of $\bar{M}^{(1,2,\dots,l)}$ have either no point or an h -dimensional face or the $(h+1)$ vertices of such a face in common and are joined by a chain in $\bar{M}^{(1,2,\dots,l)}$;

(6.10) $\bar{M}^{(1,2,\dots,l)}$ is two sided; and from (6.1) when the simplexes of $M^{(1,2,\dots,l)}$ are sufficiently fine and the division ζ_{l+1} is also sufficiently fine,

(6.11) The simplicial approximation to $\bar{M}^{(1,2,\dots,l)}$ in E_i covers $J_{i,l+1}$ once positively, ($i = 1, 2, \dots, l$).

§ 7.

The incidence relations in $M^{(l+1)}$ can now be assigned. Two simplexes $V_{ij}^{(l+1)}$ and $V_{kh}^{(l+1)}$ are defined to have an h -dimensional face in common when and only when the corresponding simplexes \bar{V}_{ij} and \bar{V}_{kh} of $\bar{M}^{(1,2,\dots,l)}$ have the corresponding $(h+1)$ vertices in common; the incidence relations among the Y_{i1} are defined to be the same as those among the corresponding simplexes Y_i . The remaining incidence relations are assigned by the process of § 4 with E_{l+1} , $J_{l+1,1}$ and $(J_{1,l+1} + J_{2,l+1} + \dots + J_{l,l+1})J_{l+1,1}$ in place of E_2 , $J_{2,1}$ and $J_{1,2}J_{2,1}$ respectively. The boundary of $M^{(l+1)}$ is then contained in $E_{l+1} - J_{l+1,1}$.

The vertices of $\{V_{ij}^{(l+1)}\}$ are now displaced into the corresponding vertices of \bar{V}_{ij} , whereby certain vertices of $\{X_{ij}\} + \{Y_{i1}\}$ are also displaced; let $\bar{V}_{ij}^{(l+1)}$, \bar{X}_{ij} and \bar{Y}_{i1} be the resulting modified simplicial images of $V_{ij}^{(l+1)}$, X_{ij} and Y_{i1} respectively in E_3 , then for a sufficiently fine division ζ_{l+1} and when the simplexes of $M^{(1,2,\dots,l)}$ are sufficiently fine, the boundary of $\bar{M}^{(l+1)} \equiv \{\bar{V}_{ij}^{(l+1)}\} + \{\bar{X}_{ij}\} + \{\bar{Y}_{i1}\}$ is contained in $E_{l+1} - J_{l+1,1}$ and exactly as in § 5, we obtain:

(7.1) Any two simplexes of $\bar{M}^{(l+1)}$ have either no point or an h -dimensional face in common and any simplex of $\bar{M}^{(l+1)}$ can be joined by a chain in $\bar{M}^{(l+1)}$ to a simplex of $\{\bar{V}_{ij}^{(l+1)}\}$;

(7.2) $\bar{M}^{(l+1)}$ is two sided and covers $J_{l+1,2}$ once positively. Writing

$$M^{(1^2 \dots l+1)} \equiv \{\bar{U}_{ij}\} + \{\bar{V}_{ij}\} + \{\bar{X}_{ij}\} + \{\bar{Y}_{i1}\},$$

and noting that when $\bar{V}_{ij}^{(l+1)}$ and $\bar{V}_{kh}^{(l+1)}$ have a common p -dimensional face, \bar{V}_{ij} and \bar{V}_{kh} have the $(p+1)$ vertices of this face in common, (6.9) and (7.1) give the result:

(7.3) Any two simplexes of $M^{(1^2 \dots l+1)}$ have either no point or an h -dimensional face or the $(h+1)$ vertices of such a face in common, ($h = 0, 1, \dots, n-1$), and can be joined by a chain in $M^{(1^2 \dots l+1)}$.

By a method entirely analogous to that used for $M^{(12)}$, from (6.10), (6.11) and (7.2), there results —

(7.4) $M^{(1^2 \dots l+1)}$ is two-sided; and by the assignment of a positive indicatrix to one of its simplexes the unique positive indicatrix thereby determined for every simplex is such that the simplicial approximation to $M^{(1^2 \dots l+1)}$ in E_i covers $J_{i,l+1}$ once positively, ($i = 1, 2, \dots, l+1$).

§ 8.

Beginning with $M^{(12)}$ and repeating the reasoning of the above §§ 6, 7 for $l = 2, 3, \dots, m-1$, we obtain the result —

(8.1) A two-sided set of simplexes $M^{(1^2 \dots m)}$, any two simplexes of which have either no point or $(h+1)$ vertices in common and can be joined by a chain in $M^{(1^2 \dots m)}$, and whose simplicial approximation in E_i covers $J_{i,m}$ once positively, ($i = 1, 2, \dots, m$), can be constructed.

Let A_1, A_2, \dots, A_ν be the vertices of $M^{(1^2 \dots m)}$ and B_1, B_2, \dots, B_ν the vertices of a $(\nu-1)$ -dimensional simplex S of $R_{\nu-1}$; then the set of those n -dimensional faces of S such as $B_\alpha, B_{\alpha_2} \dots B_{\alpha_{n+1}}$, for which the corresponding vertices $A_\alpha, A_{\alpha_2}, \dots, A_{\alpha_{n+1}}$ are vertices of a simplex of $M^{(1^2 \dots m)}$ constitutes a closed, two-sided, n -dimensional pseudo-manifold μ . Consider an infinite sequence $\mu, \mu^{(1)}, \mu^{(2)}, \dots$ of simplicial divisions of μ such that $\mu^{(i+1)}$ is a simplicial division of $\mu^{(i)}$ and further such that the set of all vertices of the sequence is everywhere dense in μ — this is accomplished when in the above used „gemischte Zerlegung“ of Brouwer³⁸) the new vertices a^h in $\mu^{(i)}$ are introduced in each h -dimensional simplex of $\mu^{(i)}$.⁴⁰) Let each vertex a^h be in the centre of the containing face. Let the vertices of $\mu^{(1)}$ be $B_1, B_2, \dots, B_\nu, B_{\nu+1}, \dots, B_{\nu_1}$; and those of $\mu^{(2)}$

⁴⁰) This is the “regular subdivision” of Veblen, Analysis Situs, p. 85–86.

$B_1, B_2, \dots, B_\nu, \dots, B_{\nu_1}, \dots, B_{\nu_2}$; and so on, and represent μ uniquely and continuously on M in the following way:

To each vertex B_i of μ is assigned that point C_i of M which is covered by A_i ;

a vertex B_i of $\mu^{(1)}$ which is not a vertex of μ is the centre of some h -dimensional face $B_{\alpha_1} B_{\alpha_2} \dots B_{\alpha_{h+1}}$ of μ , ($h = 1, 2, \dots, n$), the images $C_{\alpha_1}, C_{\alpha_2}, \dots, C_{\alpha_{h+1}}$ of whose vertices have already been assigned in M — to B_i is then assigned the centre C_i of the simplex $C_{\alpha_1} C_{\alpha_2} \dots C_{\alpha_{h+1}}$ of any element of M containing the $C_{\alpha_1}, C_{\alpha_2}, \dots, C_{\alpha_{h+1}}$; similarly a vertex B_j of $\mu^{(2)}$ not a vertex of $\mu^{(1)}$ is the centre of some h -dimensional face $B_{\beta_1} B_{\beta_2} \dots B_{\beta_{h+1}}$ of $\mu^{(1)}$, ($h = 1, 2, \dots, n$), the images $C_{\beta_1}, C_{\beta_2}, \dots, C_{\beta_{h+1}}$ of whose vertices have already been assigned in M — to B_j is then assigned the centre C_j of the simplex $C_{\beta_1} C_{\beta_2} \dots C_{\beta_{h+1}}$ with respect to any element of M containing $C_{\beta_1}, C_{\beta_2}, \dots, C_{\beta_{h+1}}$; continuing in this way the image of any vertex of the sequence $\mu, \mu^{(1)}, \mu^{(2)}, \dots$ is assigned; this representation is unique and continuous and since the vertices of the sequence $\mu, \mu^{(1)}, \mu^{(2)}, \dots$ are everywhere dense in μ , there is defined a unique, continuous representation of μ on M , whose degree, by (8.1), is ± 1 ; the theorem 1 is therefore established.

Let M be uniquely and continuously represented on a locally simplicial manifold M' ; this representation then determines a unique continuous representation of μ on M' — the degree of the latter representation is called the degree with respect to μ of the representation of M on M' ; if M be simplicial this definition agrees by virtue of the multiplicative property of the degree for simplicial manifolds with the previous one for the case of representation of a simplicial on a locally simplicial manifold; the above mentimed extension of Brouwer's theorem is thus proved.

Part II.

The topological invariance of the degree of representation and of two-sidedness for locally simplicial manifolds.

§ 9.

By the topological invariance of the degree of representation for locally simplicial manifolds is understood its invariance when the system of neighbourhoods of one (or both) of the manifolds is replaced by an equivalent system⁴¹⁾; the object of the following investigation is to establish this invariance, to prove that the degree is multiplicative for successive representations and that it has the value ± 1 for topological representations of a manifold on itself.

⁴¹⁾ In the sense of Hausdorff, "Grundzüge der Mengenlehre", p. 260—261

Let α_1 and α_2 be unique continuous representations of degree c_1 and c_2 of the closed, two-sided, pseudo-manifolds μ_1 and μ_2 respectively on the closed; two-sided, locally simplicial manifold M and let the elements E_i and the associated inner simplexes $J_{i1}, J_{i2}, \dots, J_{im}$ be defined as in the previous chapter; let $\{U_{ij}\}$ be the set of simplexes of μ_1 and $\{V_{ij}\}$ those of μ_2 . By $U_i^{(k)}$ is understood a simplicial or modified simplicial image of U_i in E_k corresponding to α_1 , i. e., the images under α_1 of the vertices of U_i undergo a certain displacement and $U_i^{(k)}$ is that simplex of E_k (when it exists) whose vertices are these displaced images; similarly $V_i^{(k)}$ is a simplicial or modified simplicial image of V_i in E_k corresponding to α_2 ; let the suffix i in $\{U_i^{(k)}\}$ be understood to take all values for which the corresponding $U_i^{(k)}$ exists; then $\{U_i^{(k)}\}$ is the simplicial approximation in E_k to the image of μ_1 and with a similar convention, $\{V_i^{(k)}\}$ is the simplicial approximation in E_k to the image of μ_2 .

(9.1) For sufficiently fine original simplicial divisions of μ_1 and μ_2 , $\{U_i^{(k)}\}$ and $\{V_i^{(k)}\}$ cover J_{k1} positively c_1 and c_2 times respectively.

The set of those simplexes of E_k each of which is covered by one or more simplexes of $\{U_i^{(k)}\}$ is called the base of the approximation $\{U_i^{(k)}\}$ in E_k ; and similarly for $\{V_i^{(k)}\}$.

When $\{U_i^{(k)}\}$ and $\{V_i^{(k)}\}$ have the same base in E_k and if, when any simplex of this common base is covered c' times positively by simplexes of $\{U_i^{(k)}\}$, it is covered c'' times by simplexes of $\{V_i^{(k)}\}$ where $\frac{c'}{c''} = \frac{c_1}{c_2}$, then $\{U_i^{(k)}\}$ and $\{V_i^{(k)}\}$ are said to be similar in E_k . It is required to prove the theorem —

Theorem 2. *Simplicial divisions of μ_1 and μ_2 together with a displacement of the images of their vertices in M can be defined so that the resulting modified simplicial approximations in any element of M are similar.*

Construct a simplicial division of E_1 into simplexes X_j such that each $U_i^{(1)}$ and $V_i^{(1)}$ is composed of one or more of the X_j while each X_j is contained in one or more of the $U_i^{(1)}$ or $V_i^{(1)}$; each $U_i^{(1)}$ and $V_i^{(1)}$ is then divided into simplexes $U_{ij}^{(1)}$ and $V_{ij}^{(1)}$ respectively each of which covers an X_j , and the indicatrices of $U_{ij}^{(1)}$ and $V_{ij}^{(1)}$ are determined from those of $U_i^{(1)}$ and $V_i^{(1)}$ respectively by the usual rule. Then by (9.1)

$$\{U_{ij}^{(1)}\} \text{ and } \{V_{ij}^{(1)}\} \text{ (}^{42}\text{) are similar in } E_1,$$

⁴²⁾ By $\{U_{ij}^{(1)}\}$ is understood the set of all component simplexes of $\{U_i^{(1)}\}$, and, as in previous paragraphs, by $\{U_i^{(1)}\}$ the set of all existing simplexes $U_i^{(1)}$; similarly for $\{V_{ij}^{(1)}\}$.

having the common base $\{X_i\}$ each simplex of which is covered c_1 times by $\{U_{ij}^{(1)}\}$ and c_2 times by $\{V_{ij}^{(1)}\}$.

Let U_{ij} be the image of $U_{ij}^{(1)}$ in U_i and V_{ij} the image of $V_{ij}^{(1)}$ in V_i . For $h=1, 2, \dots, (n-1)$ in succession, in each $(h+1)$ -dimensional simplex of $\{U_i\}$ and $\{V_i\}$ not yet divided but having one or more divided h -dimensional faces, take an arbitrary point a^{h+1} and form the set of all simplexes such as $a^0 a^1 \dots a^{h+1}$ where $a^0 a^1 \dots a^h$ is either one of the undivided h -dimensional faces of the simplex containing a^{h+1} or a component simplex of one of its divided h -dimensional faces³³). Component simplexes of U_i and V_i thus obtained are written U_{ij} and V_{ij} respectively and those U_i and V_i undivided by the above process are re-named U_{i1} and V_{i1} ; then

$\{U_{ij}\} \equiv {}_1\mu_1$ and $\{V_{ij}\} \equiv {}_1\mu_2$ are simplicial divisions of μ_1 and μ_2 respectively.

By A_h is understood any vertex of ${}_1\mu_1$ or ${}_1\mu_2$ and by \bar{A}_h the image of A_h in $\{U_{ij}^{(1)}\}$ or $\{V_{ij}^{(1)}\}$ when such image exists, and when not, the image of A_h under α_1 or α_2 respectively; if then the vertices of any simplex U_{ij} or V_{ij} be $A_{a_1}, A_{a_2}, \dots, A_{a_{n+1}}$, by \bar{U}_{ij}^k or $\bar{V}_{ij}^{(k)}$ is understood the simplex $\bar{A}_{a_1} \bar{A}_{a_2} \dots \bar{A}_{a_{n+1}}$ of E_k when it exists.

Then $\{\bar{U}_{ij}^{(k)}\}$ and $\{\bar{V}_{ij}^{(k)}\}$ are modified simplicial approximations in E_k to the images of ${}_1\mu_1$ and ${}_1\mu_2$ respectively; for sufficiently fine original subdivisions of μ_1 and μ_2 , when $\bar{U}_{ij}^{(1)}$ or $\bar{V}_{ij}^{(1)}$ contains points of J_{11} it is identical with $U_{ij}^{(1)}$ or $V_{ij}^{(1)}$ respectively, so that

the modified simplicial approximations $\{\bar{U}_{ij}^{(1)}\}$ and $\{\bar{V}_{ij}^{(1)}\}$ are similar in J_{11} .

§ 10.

The simplexes $\{U_{ij}\}$ and $\{V_{ij}\}$ of ${}_1\mu_1$ and ${}_1\mu_2$ are now rewritten $\{{}_1U_h\}$ and $\{{}_1V_h\}$ respectively, and when ${}_1U_h$ or ${}_1V_h$ represents the simplex previously written U_{ij} or V_{ij} respectively, then ${}_1U_h^{(k)}$ or ${}_1V_h^{(k)}$ represents the simplex previously written $\bar{U}_{ij}^{(k)}$ or $\bar{V}_{ij}^{(k)}$ respectively, so that the above modified simplicial approximations $\{\bar{U}_{ij}^{(k)}\}$ and $\{\bar{V}_{ij}^{(k)}\}$ are written $\{{}_1U_h^{(k)}\}$ and $\{{}_1V_h^{(k)}\}$; thus

$\{{}_1U_i^{(1)}\}$ and $\{{}_1V_i^{(1)}\}$ are similar in J_{11} ; and for sufficiently fine original simplicial divisions of μ_1 and μ_2 ,

$\{{}_1U_i^{(2)}\}$ and $\{{}_1V_i^{(2)}\}$ cover J_{21} with degrees c_1 and c_2 respectively.

Construct a simplicial division of E_2 into simplexes $X_j^{(2)}$ such that each ${}_1U_i^{(2)}$ and ${}_1V_i^{(2)}$ is composed of one or more of the $X_j^{(2)}$ and each $X_j^{(2)}$ is contained in one or more of the ${}_1U_i^{(2)}$ and ${}_1V_i^{(2)}$; divide each ${}_1U_i^{(2)}$ and ${}_1V_i^{(2)}$ into simplexes ${}_1U_{ij}^{(2)}$ and ${}_1V_{ij}^{(2)}$ respectively each of which covers an

$X_j^{(2)}$ and determine their indicatrices from those of ${}_1U_i^{(2)}$ and ${}_1V_i^{(2)}$ respectively by the rule for simplicial manifolds; then

$$\{{}_1U_{ij}^{(2)}\} \text{ and } \{{}_1V_{ij}^{(2)}\}^{43)} \text{ are similar in } J_{21},$$

having the common base $\{X_j^{(2)}\}$ of which each simplex containing points of J_{21} is covered c_1 times by $\{{}_1U_{ij}^{(2)}\}$ and c_2 times by $\{{}_1V_{ij}^{(2)}\}$.

Those ${}_1U_i^{(1)}$ and ${}_1V_i^{(1)}$ for which corresponding ${}_1U_i^{(2)}$ and ${}_1V_i^{(2)}$ exist are divided into simplexes ${}_1U_{ij}^{(1)}$ and ${}_1V_{ij}^{(1)}$ which are images of ${}_1U_{ij}^{(2)}$ and ${}_1V_{ij}^{(2)}$ in ${}_1U_i^{(1)}$ and ${}_1V_i^{(1)}$ respectively and the corresponding base simplexes are divided into simplexes covered by these ${}_1U_{ij}^{(1)}$ and ${}_1V_{ij}^{(1)}$; the remaining simplexes of the bases of $\{{}_1U_i^{(1)}\}$ and $\{{}_1V_i^{(1)}\}$ in E_1 are simplicially divided as follows — for $h=1, 2, \dots, (n-1)$ in succession, in each $(h+1)$ -dimensional simplex not yet divided but having one or more divided h -dimensional faces take an arbitrary point a^{h+1} and construct the set of all simplexes such as $a^0 a^1 \dots a^{h+1}$ where $a^0 a^1 \dots a^h$ is either one of the undivided h -dimensional faces of the simplex containing a^{h+1} or a component simplex of one of its divided faces; the remaining simplexes of the above bases are either undivided or preserve their previous division⁴³⁾; the remaining simplexes ${}_1U_i^{(1)}$ and ${}_1V_i^{(1)}$ for which the corresponding ${}_1U_i^{(2)}$ and ${}_1V_i^{(2)}$ do not exist are now divided into simplexes ${}_1U_{ij}^{(1)}$ and ${}_1V_{ij}^{(1)}$ respectively each of which covers one of the simplexes $a^0 a^1 \dots a^n$, while those whose base simplex is undivided are rewritten ${}_1U_{i1}^{(1)}$ and ${}_1V_{i1}^{(1)}$ respectively; by this method the ${}_1U_i^{(1)}$ and ${}_1V_i^{(1)}$ have all been simplicially divided in such a way that when any two simplexes ${}_1U_i^{(1)}$ and ${}_1V_j^{(1)}$ have a common base simplex Δ , the corresponding sets of component simplexes have a simplicial division of Δ as common base; hence

$$\{{}_1U_{ij}^{(1)}\} \text{ and } \{{}_1V_{ij}^{(1)}\} \text{ are similar in } J_{11}.$$

${}_1\mu_1$ and ${}_1\mu_2$ are now simplicially divided as follows —

When ${}_1U_i^{(k)}$ exists it is a simplicial image of ${}_1U_i$ in E_k , ($k=1, 2$); let ${}_1U_{ij}$ be the image of ${}_1U_{ij}^{(k)}$ in ${}_1U_i$ and similarly let ${}_1V_{ij}$ be the image of ${}_1V_{ij}^{(k)}$ in ${}_1V_i$; the remaining ${}_1U_i$ and ${}_1V_i$ are divided as follows — for $h=1, 2, \dots, (n-1)$ in succession, in each $(h+1)$ -dimensional simplex not yet divided but having one or more divided h -dimensional faces take an arbitrary point a^{h+1} and proceed as in § 1, writing the component simplexes ${}_1U_{ij}$ and ${}_1V_{ij}$ and rewriting the simplexes undivided by this process ${}_1U_{i1}$ and ${}_1V_{i1}$; then

$$\{{}_1U_{ij}\} \equiv {}_2\mu_1 \text{ and } \{{}_1V_{ij}\} \equiv {}_2\mu_2 \text{ are simplicial divisions of } {}_1\mu_1 \text{ and } {}_1\mu_2$$

⁴³⁾ As in § 9, by $\{{}_1U_{ij}^{(2)}\}$ is understood the set of all component simplexes of $\{{}_1U_i^{(2)}\}$; similarly for $\{{}_1V_{ij}^{(2)}\}$.

respectively, and corresponding modified simplicial approximations are the following —

- (a) $\left\{ \begin{array}{l} \text{Let } A_h \text{ be any vertex of } {}_2\mu_1 \text{ or } {}_2\mu_2; \text{ by } \bar{A}_h \text{ is understood} \\ (1) \text{ the image of } A_h \text{ in } \{ {}_1U_{ij}^{(2)} \} \text{ or } \{ {}_1V_{ij}^{(2)} \} \text{ when such image exists,} \\ \text{or } (2) \text{ the image of } A_h \text{ in } \{ {}_1U_{ij}^{(1)} \} \text{ or } \{ {}_1V_{ij}^{(1)} \} \text{ when such image exists,} \\ \text{but does not exist in case (1),} \\ \text{or } (3) \text{ the image of } A_h \text{ under } \alpha \text{ or } \beta \text{ when (1) and (2) are impossible;} \end{array} \right.$

let the vertices of any simplex ${}_1U_{ij}$ or ${}_1V_{ij}$ be $A_{a_1}, A_{a_2}, \dots, A_{a_{n+1}}$, and let ${}_1\bar{U}_{ij}^{(k)}$ or ${}_1\bar{V}_{ij}^{(k)}$ respectively be the simplex $\bar{A}_{a_1} \bar{A}_{a_2} \dots \bar{A}_{a_{n+1}}$ of E_k when such simplex exists; $\{ {}_1\bar{U}_{ij}^{(k)} \}$ and $\{ {}_1\bar{V}_{ij}^{(k)} \}$ are the required simplicial approximations in E_k to the images of ${}_2\mu_1$ and ${}_2\mu_2$ respectively, ($k = 1, 2, \dots, m$); from their construction it follows that when certain of the simplexes ${}_1U_{ij}^{(1)}$ and ${}_1V_{ij}^{(1)}$ have a common base simplex, the corresponding simplexes ${}_1\bar{U}_{ij}^{(1)}$ and ${}_1\bar{V}_{ij}^{(1)}$ have a common base simplex and therefore, since $\{ {}_1U_{ij}^{(1)} \}$ and $\{ {}_1V_{ij}^{(1)} \}$ are similar in J_{11} , it follows that for sufficiently fine divisions ${}_1\mu_1$ and ${}_1\mu_2$,

- (b) $\{ {}_1\bar{U}_{ij}^{(1)} \}$ and $\{ {}_1\bar{V}_{ij}^{(1)} \}$ are similar in J_{12} ;

again, when ${}_1\mu_1$ and ${}_1\mu_2$ are sufficiently fine, those ${}_1\bar{U}_{ij}^{(2)}$ and ${}_1\bar{V}_{ij}^{(2)}$ which contain points of J_{21} are identical with the corresponding ${}_1U_{ij}^{(2)}$ and ${}_1V_{ij}^{(2)}$ so that

- (c) $\{ {}_1\bar{U}_{ij}^{(2)} \}$ and $\{ {}_1\bar{V}_{ij}^{(2)} \}$ are similar in J_{21} ;

From (a), (b) and (c) we conclude —

Simplicial divisions ${}_2\mu_1$ and ${}_2\mu_2$ of μ_1 and μ_2 respectively, and a displacement of the images under α_1 and α_2 in M of their vertices have been so defined that the resulting modified simplicial approximations in E_k are similar in J_{k2} , ($k = 1, 2$).

§ 11.

Assume now that simplicial divisions ${}_l\mu_1$ and ${}_l\mu_2$ of μ_1 and μ_2 respectively and a displacement of the images under α_1 and α_2 in M of their vertices have been so defined that the resulting modified simplicial approximations in E_k are similar in J_{kl} , ($k = 1, 2, \dots, l$).

Let $\{ {}_lU_i \}$ and $\{ {}_lV_i \}$ be the sets of simplexes of ${}_l\mu_1$ and ${}_l\mu_2$ respectively and $\{ {}_lU_i^{(k)} \}$ and $\{ {}_lV_i^{(k)} \}$ the corresponding modified simplicial approximations which are thus similar in J_{kl} ($k = 1, 2, \dots, l$). When the simplicial divisions $\{ {}_lU_i \}$ and $\{ {}_lV_i \}$ of ${}_l\mu_1$ and ${}_l\mu_2$ are sufficiently fine,

$$\{ {}_lU_i^{(l+1)} \} \text{ and } \{ {}_lV_i^{(l+1)} \} \text{ cover } J_{l+1,1}$$

with degrees c_1 and c_2 respectively.

As in the previous paragraph, construct a simplicial division of E_{i+1} into simplexes $X_j^{(l+1)}$ such that each ${}_iU_i^{(l+1)}$ and ${}_iV_i^{(l+1)}$ is composed of one or more of the $X_j^{(l+1)}$ and each $X_j^{(l+1)}$ is contained in one or more of the ${}_iU_i^{(l+1)}$ or the ${}_iV_i^{(l+1)}$; divide each ${}_iU_i^{(l+1)}$ and ${}_iV_i^{(l+1)}$ into simplexes ${}_iU_{ij}^{(l+1)}$ and ${}_iV_{ij}^{(l+1)}$ respectively each of which covers an $X_j^{(l+1)}$ and determine their indicatrices from those of ${}_iU_i^{(l+1)}$ and ${}_iV_i^{(l+1)}$ respectively; then:

$$\{ {}_iU_{ij}^{(l+1)} \} \text{ and } \{ {}_iV_{ij}^{(l+1)} \} \text{ are similar in } J_{i+1,1},$$

having the common base $\{ X_i^{(l+1)} \}$ of which each simplex containing points of $J_{i+1,1}$ is covered c_1 times by $\{ {}_iU_{ij}^{(l+1)} \}$ and c_2 times by $\{ {}_iV_{ij}^{(l+1)} \}$. Again by a process analogous to that in the previous paragraph, the ${}_iU_i^{(l)}$ and ${}_iV_i^{(l)}$ are divided into simplexes ${}_iU_{ij}^{(l)}$ and ${}_iV_{ij}^{(l)}$ such that:

$$\{ {}_iU_{ij}^{(l)} \} \text{ and } \{ {}_iV_{ij}^{(l)} \} \text{ are similar in } J_{i1}.$$

Assume now that ${}_iU_i^{(k)}$ and ${}_iV_i^{(k)}$ have been divided into ${}_iU_{ij}^{(k)}$ and ${}_iV_{ij}^{(k)}$ in such a way that:

$$\{ {}_iU_{ij}^{(k)} \} \text{ and } \{ {}_iV_{ij}^{(k)} \} \text{ are similar in } J_{ki}, \quad (k = l, l - 1, \dots, h + 1);$$

we require to prove that this last result is valid for $k = h$.

Those ${}_iU_i^{(h)}$ and ${}_iV_i^{(h)}$ for which one of the corresponding simplexes ${}_iU_i^{(k)}$ and ${}_iV_i^{(k)}$, ($k = h + 1, h + 2, \dots, l + 1$) exists are divided into simplexes ${}_iU_{ij}^{(h)}$ and ${}_iV_{ij}^{(h)}$ which are images of ${}_iU_{ij}^{(k)}$ and ${}_iV_{ij}^{(k)}$ in ${}_iU_i^{(h)}$ and ${}_iV_i^{(h)}$ respectively and the corresponding base simplexes are divided into simplexes covered by these ${}_iU_{ij}^{(h)}$ and ${}_iV_{ij}^{(h)}$; the remaining simplexes of the bases of $\{ {}_iU_i^{(h)} \}$ and $\{ {}_iV_i^{(h)} \}$ are divided by the method of the previous paragraph. Those simplexes ${}_iU_i^{(h)}$ and ${}_iV_i^{(h)}$ for which none of the corresponding simplexes ${}_iU_i^{(k)}$ and ${}_iV_i^{(k)}$ ($k = h + 1, h + 2, \dots, l + 1$) exist are also divided into simplexes ${}_iU_{ij}^{(h)}$ and ${}_iV_{ij}^{(h)}$ respectively each of which covers a simplex of the division of the bases of $\{ {}_iU_i^{(h)} \}$ and $\{ {}_iV_i^{(h)} \}$, while those whose base simplex is undivided are rewritten ${}_iU_{i1}^{(h)}$ and ${}_iV_{i1}^{(h)}$ respectively; then:

$$\{ {}_iU_{ij}^{(h)} \} \text{ and } \{ {}_iV_{ij}^{(h)} \} \text{ are similar in } J_{hi}.$$

It follows that each ${}_iU_i^{(k)}$ and ${}_iV_i^{(k)}$ can be divided into simplexes ${}_iU_{ij}^{(k)}$ and ${}_iV_{ij}^{(k)}$ respectively in such a way that:

$$\{ {}_iU_{ij}^{(k)} \} \text{ and } \{ {}_iV_{ij}^{(k)} \} \text{ are similar in } J_{ki} \quad (k = 1, 2, \dots, l).$$

Let ${}_1\mu_1$ and ${}_1\mu_2$ be now simplicially divided by the method analogous to that³³ used for ${}_1\mu_1$ and ${}_1\mu_2$, with ${}_iU_i^{(k)}$, ${}_iV_i^{(k)}$ and $k = 1, 2, \dots, l$ in place of ${}_1U_i^{(k)}$, ${}_1V_i^{(k)}$ and $k = 1, 2$; we thereby obtain simplicial divisions

$$\{ {}_iU_{ij} \} \equiv {}_{i+1}\mu_1 \quad \text{and} \quad \{ {}_iV_{ij} \} \equiv {}_{i+1}\mu_2$$

of ${}_i\mu_1$ and ${}_i\mu_2$ respectively. To these simplicial divisions we construct modified simplicial approximations as follows —

Let A_n be any vertex of ${}_{l+1}\mu_1$ or ${}_{l+1}\mu_2$; by \bar{A}_k is understood the image of A_n in $\{ {}_lU_{ij}^{(k)} \}$ or $\{ {}_lV_{ij}^{(k)} \}$ respectively where k is the greatest integer not exceeding $l+1$ for which such image exists, and when no such image exists, the image of A_n under α_1 or α_2 respectively. Let the vertices of any simplex ${}_lU_{ij}$ or ${}_lV_{ij}$ be $A_{\alpha_1}, A_{\alpha_2}, \dots, A_{\alpha_{n+1}}$ and let ${}_l\bar{U}_{ij}^{(k)}$ or ${}_l\bar{V}_{ij}^{(k)}$ respectively be the simplex $\bar{A}_{\alpha_1} \bar{A}_{\alpha_2} \dots \bar{A}_{\alpha_{n+1}}$ of E_k , ($k = 1, 2, \dots, m$), when such simplex exists; then $\{ {}_l\bar{U}_{ij}^{(k)} \}$ and $\{ {}_l\bar{V}_{ij}^{(k)} \}$ are the required simplicial approximations in E_k , ($k = 1, 2, \dots, m$), to the images of ${}_{l+1}\mu_1$ and ${}_{l+1}\mu_2$ respectively. As with ${}_2\mu_1$ and ${}_2\mu_2$ we then have —

Simplicial divisions ${}_{l+1}\mu_1$ and ${}_{l+1}\mu_2$ of μ_1 and μ_2 respectively and a displacement of the images under α_1 and α_2 of their vertices have been so defined that the resulting modified simplicial approximations in E_k are similar in $J_{k, l+1}$, ($k = 1, 2, \dots, l+1$).

From the result of the previous paragraph, repetition of the above reasoning for $l = 2, 3, \dots, m-1$ establishes the theorem 2.

§ 12.

Let α' be any unique continuous representation of M on another locally simplicial manifold M' (which is first assumed to be closed and two-sided); $\alpha_1 \alpha'$ and $\alpha_2 \alpha'$ are then unique continuous representations of μ_1 and μ_2 respectively on M' ; it follows from theorem 2 that there is a displacement of the images in M' of the vertices of ${}_m\mu_1$ and ${}_m\mu_2$ for which the corresponding modified simplicial representations in any element of M' are similar, hence we have:

Theorem 3. *If α_1 and α_2 be unique continuous representations of degrees c_1 and c_2 respectively of the pseudo-manifolds μ_1 and μ_2 on the closed, two-sided locally simplicial manifold M , α' a unique continuous representation of M on the closed, two-sided locally simplicial manifold M' , c'_1 and c'_2 the degrees of the representations $\alpha_1 \alpha'$ and $\alpha_2 \alpha'$ of μ_1 and μ_2 respectively on M' , then $\frac{c'_1}{c'_2} = \frac{c_1}{c_2}$.*

When $c_1 = c_2 = +1$, this theorem gives $c'_1 = c'_2$; but in this case c'_1 and c'_2 have been defined as the degrees with respect to μ_1 and μ_2 respectively of the representation α' of M on M' ; hence:

Corollary 1. The degree of representation of M on M' is the same with respect to any two pseudo-manifolds which cover M with degree $+1$.

Corollary 2. The degree is multiplicative for successive representations, *i. e.*, if M_1, M_2 and M_3 be closed, two-sided, locally simplicial manifolds, α_{12} and α_{23} representations of degree c_{12} and c_{23} of M_1 on M_2 and M_2 on M_3 respectively, and c_{13} the degree of the product representation $\alpha_{13} = \alpha_{12} \alpha_{23}$ of M_1 on M_3 , then $c_{13} = c_{12} c_{23}$.

For if μ_1 and μ_2 be pseudo-manifolds represented on M_1 and M_2 respectively with degree $+1$, α_{12} determines a representation of μ_1 on M_2 of degree c_{12} , while α_{23} determines representations of μ_1 and μ_2 on M_3 of degrees c_{13} and c_{23} respectively whence from theorem 3,

$$\frac{c_{13}}{c_{23}} = \frac{c_{12}}{+1}.$$

From this result the proofs of Brouwer⁴⁴⁾ give —

Corollary 3. The degree of a topological representation of one closed, two-sided, locally simplicial manifold on another is ± 1 ⁴⁵⁾.

Corollary 4. One-sidedness and two-sidedness are topological invariants of locally simplicial manifolds⁴⁶⁾.

Corollary 5. The degree of representation is a topological invariant for locally simplicial manifolds⁴⁷⁾.

⁴⁴⁾ „Über Jordansche Mannigfaltigkeiten“, Math. Annalen 71, p. 320, §§ 5 and 6.

⁴⁵⁾ Brouwer, „Über Jordansche Mannigfaltigkeiten“, p. 324 and the remarks on p. 598.

⁴⁶⁾ Brouwer, op. cit. p. 324, footnote.

⁴⁷⁾ Brouwer, op. cit. Satz 6.

(Eingegangen am 30. 5. 1927.)

Zur Topologie der Abbildungen von Mannigfaltigkeiten.

Erster Teil.

Neue Darstellung der Theorie des Abbildungsgrades für topologische Mannigfaltigkeiten.

Von

Heinz Hopf in Berlin.

Diese Arbeit bildet den ersten und einleitenden Teil einer größeren Abhandlung, in der die eindeutigen und stetigen Abbildungen einer n -dimensionalen Mannigfaltigkeit auf eine zweite n -dimensionale Mannigfaltigkeit näher untersucht werden. Im Vordergrund steht dabei der von Brouwer eingeführte Begriff der „Abbildungsklasse“, d. h. der Menge aller Abbildungen, die sich aus einer gegebenen Abbildung durch stetige Modifikation herstellen lassen. In der vorliegenden Arbeit wird die von Brouwer bewiesene Tatsache, daß alle Abbildungen einer Klasse denselben „Grad“ besitzen, als Grundlage für den Aufbau der Theorie des Abbildungsgrades in einer erweiterten Gestalt benutzt¹⁾.

Dieser Aufbau wird in zwei Schritten durchgeführt. Der erste besteht in der Begründung der Lehre von dem Grade der Abbildungen eines n -dimensionalen *Elements* auf Punktmengen eines n -dimensionalen euklidischen Raumes (§ 1). Diese Dinge werden jedoch nicht ab ovo dargestellt, vielmehr wird die Theorie des Grades für die $(n - 1)$ -dimensionalen Kugeln

¹⁾ Es werden fortgesetzt Begriffe und Sätze aus den folgenden Arbeiten von Brouwer benutzt:

a) Über Abbildung von Mannigfaltigkeiten, *Math. Annalen* 71 (1911).

b) Über Jordansche Mannigfaltigkeiten, *Math. Annalen* 71 (1911).

c) Beweis der Invarianz der Dimensionenzahl, *Math. Annalen* 70 (1911); (hier wird zwar nicht der Name, aber der Begriff des Abbildungsgrades zuerst eingeführt, und zwar für Abbildungen eines n -dimensionalen Elements, nicht einer geschlossenen Mannigfaltigkeit).

als bekannt vorausgesetzt, und mit ihrer Hilfe werden durch Vermittlung des Begriffs der „Ordnung“²⁾ die Beweise der in Brouwerschen Arbeiten enthaltenen Sätze über das n -dimensionale Element erbracht. Es wäre übrigens leicht, diese Beweisführung zu einem Induktionsschluß von $n - 1$ auf n Dimensionen zu vervollständigen³⁾; jedoch wird hierauf in dieser Arbeit kein Wert gelegt.

Neu dürften in diesem ersten Paragraphen zwei Tatsachen sein: erstens wird für die eineindeutigen Abbildungen, die die Orientierung erhalten, und diejenigen, die die Orientierung umkehren, ein Unterscheidungsmerkmal angegeben, das frei von Vorzeichenbetrachtungen ist und nur auf Stetigkeitsbegriffen beruht. Zweitens wird der Grad durch eine Minimal-eigenschaft charakterisiert: Sein absoluter Betrag in einem Punkte ξ des Bildraumes ist die Minimalzahl der eineindeutigen Bedeckungen einer Umgebung von ξ , welche sich erreichen läßt, wenn man die gegebene Abbildung im Innern, jedoch nicht am Rande, des abgebildeten Elements stetig abändert. Die Übertragung dieses Satzes auf die Abbildungen beliebiger Mannigfaltigkeiten wird einen wesentlichen Punkt des zweiten Teils⁴⁾ dieser Abhandlung bilden.

Der zweite Schritt beim Aufbau der Theorie des Grades ist der folgende: In jeder Klasse von Abbildungen einer n -dimensionalen Mannigfaltigkeit auf eine andere gibt es derart reguläre Abbildungen, daß für sie die Definition des Grades bei Kenntnis der im § 1 hergeleiteten Sätze keinerlei Schwierigkeit macht. Es wird nun gezeigt, daß diese regulären Abbildungen innerhalb der Klasse eine „überall dichte“ und in gewissem Sinne „zusammenhängende“ Menge bilden. Aus der zweiten Eigenschaft folgt leicht, daß sie alle den gleichen Grad besitzen und daß dieser somit eine Invariante der ganzen Klasse ist, womit das angestrebte Ziel im wesentlichen erreicht ist. Es sei noch bemerkt, daß die Eigenschaft, innerhalb der Klasse überall dicht zu liegen, insbesondere auch der Menge derjenigen Abbildungen zukommt, die in einem festen Punkt der Bildmannigfaltigkeit „glatt“ sind, d. h. bei denen die Bildmenge eine Umgebung des betrachteten Punktes endlich oft eineindeutig bedeckt.

Die genannten „regulären“ Abbildungen — im Text werden sie in „halbglatte“, „fastglatte“, „glatte“ abgestuft —, die man durch stetige Abänderung der gegebenen Abbildung erhält, leisten also etwa dasselbe, wie die von Brouwer herangezogenen simplizialen Approximationen.

²⁾ Siehe ¹⁾ b), sowie Hadamard, Note sur quelques applications de l'indice de Kronecker (Tannery, Introduction à la théorie des fonctions II, 2. éd., 1911).

³⁾ Vgl. die unter ²⁾ zitierte Darstellung von Hadamard.

⁴⁾ „Klasseninvarianten von Abbildungen“.

Diese, sowie überhaupt simpliziale Zerlegungen, werden in der neuen Darstellung völlig vermieden; infolgedessen kann man auf die Voraussetzung der „Triangulierbarkeit“ der betrachteten Mannigfaltigkeiten verzichten: diese brauchen nur als „topologische“ Mannigfaltigkeiten erklärt zu sein, d. h. als zusammenhängende topologische Räume, die abzählbare vollständige Systeme von n -dimensionalen euklidischen Umgebungen besitzen, während ihre Zerlegbarkeit in Simplexe offen gelassen werden kann. Ob dies eine tatsächliche Erweiterung des Bereichs der triangulierbaren Mannigfaltigkeiten bedeutet, ist allerdings nicht bekannt und zum mindesten fraglich.

Daß sich der Abbildungsgrad für „topologische“ Mannigfaltigkeiten definieren läßt, ist bereits von W. Wilson⁵⁾ — mit prinzipiell anderen Methoden, als den hier benutzten — gezeigt worden; erst die Kenntnis dieser Untersuchungen und Ergebnisse veranlaßte mich zur Verschärfung und Verallgemeinerung meiner Methoden zwecks neuer Herleitung des Wilsonschen Resultats.

In einer Richtung wird übrigens über die Sätze von Brouwer und Wilson hinausgegangen: der Grad wird auch für offene Mannigfaltigkeiten definiert; dies ist ohne große Komplikationen zu erreichen; man hat sich nur auf gewisse Teilgebiete der Bildmannigfaltigkeit zu beschränken, in deren jedem der Grad konstant ist.

§ 1.

Vorbemerkungen über Abbildungen eines n -dimensionalen Elements auf Punktmengen des n -dimensionalen euklidischen Raumes⁶⁾.

Im n -dimensionalen euklidischen (x^1, x^2, \dots, x^n) -Raum R^n bezeichne p_0 den Koordinatenanfangspunkt, p_ν den Einheitspunkt auf der x^ν -Achse, T_0^n das durch die Punkte p_ν ($\nu = 0, 1, \dots, n$) bestimmte Simplex. Eine Reihenfolge $p_{i_0} p_{i_1} \dots p_{i_n}$ der Ecken nennen wir eine positive oder eine negative Orientierung von T_0^n , je nachdem $[i_0 i_1 \dots i_n]$ eine gerade oder eine ungerade Permutation von $[0 1 \dots n]$ ist oder, was dasselbe bedeutet, je nachdem die durch $p_{i_\nu} = A(p_\nu)$ ($\nu = 0, 1, \dots, n$) eindeutig festgelegte affine Transformation $A(R^n)$ eine positive oder negative Determinante besitzt. Entsprechend legt eine Eckenreihenfolge $q_0 q_1 \dots q_n$ eines beliebigen Simplexes T^n die positive oder negative Orientierung von T^n je nach dem Vorzeichen der Determinante derjenigen affinen

⁵⁾ W. Wilson, Representation of a simplicial manifold on a locally simplicial manifold; Proceedings Amsterdam 29 (1926), S. 1129ff. — Representation of Manifolds; Math. Annalen 100 (1928), S. 552.

⁶⁾ Der Inhalt dieses Paragraphen ist zum Teil nur eine Zusammenstellung und für das folgende zweckmäßige Formulierung bekannter Tatsachen.

Transformationen B fest, die durch $q_\nu = B(p_\nu)$ ($\nu = 0, 1, \dots, n$) bestimmt ist. Als positive bzw. negative Randorientierung eines $(n-1)$ -dimensionalen Randsimplexes T^{n-1} von T^n bezeichnen wir diejenigen Reihenfolgen seiner Ecken, die zusammen mit der nicht zu T^{n-1} gehörigen Ecke von T^n als letztem Punkt die positive bzw. negative Orientierung von T^n definieren. Ist S^{n-1} eine im R^n gelegene $(n-1)$ -dimensionale Kugel, t^{n-1} ein sphärisches Simplex von S^{n-1} , d. h. ein System von n Punkten q_1, q_2, \dots, q_n auf S^{n-1} , die mit dem Kugelmittelpunkt nicht in einer $(n-1)$ -dimensionalen Ebene liegen, so sei die positive Orientierung von t^{n-1} identisch mit der positiven Randorientierung des durch q_1, q_2, \dots, q_n aufgespannten ebenen Simplexes T^{n-1} , wenn man es als Randsimplex des von ihm und dem Kugelmittelpunkt aufgespannten T^n ansieht.

Auf diese Weise werden vermöge der Orientierung des R^n die in diesem liegenden $(n-1)$ -dimensionalen Kugeln zu orientierten Mannigfaltigkeiten.

Sei nun P^n ein zweiter euklidischer n -dimensionaler Raum mit einem $(\xi^1, \xi^2, \dots, \xi^n)$ -Koordinatensystem und seien in ihm die analogen Orientierungsfestsetzungen wie in R^n getroffen. Sind dann S^{n-1} und Σ^{n-1} Kugeln in R^n bzw. P^n , und ist S^{n-1} eindeutig und stetig auf Σ^{n-1} abgebildet, so besitzt diese Abbildung φ einen auch bezüglich des Vorzeichens wohlbestimmten Grad $\gamma(\varphi)$; ist insbesondere φ durch $\xi^\nu = x^\nu$ ($\nu = 1, 2, \dots, n$) gegeben, so ist $\gamma(\varphi) = +1$.

Ist f eine Abbildung der Kugel S^{n-1} auf eine Punktmenge $f(S^{n-1})$ von P^n , ξ ein nicht zu $f(S^{n-1})$ gehöriger Punkt von P^n , Σ^{n-1} eine ξ im Inneren enthaltende Kugel, π die — in allen von ξ verschiedenen Punkten in P^n definierte — Zentralprojektion vom Punkt ξ aus auf Σ^{n-1} , so ist der Grad $\gamma(\pi f)$ der Abbildung $\pi f(S^{n-1})$ unabhängig von der speziellen Wahl von Σ^{n-1} .

Wir schreiben:

$$\gamma(\pi f) = u_\xi(f(S^{n-1}))$$

und nennen diese Zahl die „*Ordnung* des Punktes ξ in bezug auf das Kugelbild $f(S^{n-1})$ “. Da der Grad bei stetiger Änderung der Abbildung konstant bleibt, gilt

Satz I. Ändert man S^{n-1} , f , ξ stetig ab, d. h. definiert man für $0 \leq t \leq 1$ stetige Scharen S_t^{n-1} , f_t , ξ_t von Kugeln bzw. Abbildungen und Punkten, so bleibt die Ordnung $u_{\xi_t}(f_t(S_t^{n-1}))$ ungeändert, falls niemals ξ_t auf $f_t(S_t^{n-1})$ liegt.

Es sei jetzt f nicht nur auf S^{n-1} , sondern in der ganzen von S^{n-1} begrenzten Vollkugel V^n definiert. Dann gilt

Satz II. Ist $\xi \nsubseteq f(V^n)$, so ist $u_\xi(f(S^{n-1})) = 0$.⁷⁾

Denn hat S^{n-1} den Radius r und bezeichnet S_t^{n-1} die Kugel vom Radius t um den Mittelpunkt m von S^{n-1} , S_0^{n-1} also den Punkt m selbst, so ist $\xi \nsubseteq f(S_t^{n-1})$ für $0 \leq t \leq r$, da $f(S_t^{n-1}) \subset f(V^n)$ ist; nach Satz I ist daher $u_\xi(f(S^{n-1})) = u_\xi(f(S_t^{n-1}))$; für hinreichend kleine t ist diese Zahl aber gewiß 0, da die Bildmenge $\pi f(S_t^{n-1})$ auf Σ^{n-1} nur eine Umgebung des Punktes $\pi f(m)$ bedeckt.

Satz II läßt sich verschärfen zu

Satz IIa. Ist $\xi \nsubseteq f(S^{n-1})$, $u_\xi(f(S^{n-1})) \neq 0$, so gibt es eine ξ enthaltende Vollkugel ω^n , die ganz zu $f(V^n)$ gehört.

Denn nimmt man ω^n so klein an, daß man jeden ihrer Punkte α unter Vermeidung von $f(S^{n-1})$ in ξ überführen kann, so ist nach Satz I: $u_\alpha(f(S^{n-1})) \neq 0$, also nach Satz II: $\alpha \subset f(V^n)$.

Hieraus folgt weiter:

Satz IIb. Ist $f(V^n) \subset R^n$ (also $P^n \equiv R^n$), und ist für jeden Punkt x von V^n die Entfernung $\varrho(x, f(x))$ kleiner als der Radius r von S^{n-1} , so gibt es eine den Mittelpunkt m enthaltende Vollkugel v^n , die ganz zu $f(V^n)$ gehört⁸⁾.

Denn die Abbildung $f(V^n)$ läßt sich aus der identischen, d. h. jeden Punkt sich selbst zuordnenden, Abbildung durch stetige Abänderung erzeugen, indem man jeden Punkt x geradlinig mit der Geschwindigkeit 1 nach $f(x)$ laufen läßt; dabei wird wegen $\varrho(x, f(x)) < r$ der Mittelpunkt m niemals von dem Bild eines Randpunktes bedeckt, nach Satz I ist also $u_m(f(S^{n-1})) = u_m(S^{n-1}) = 1$. Hieraus ergibt sich auf Grund von Satz IIa die Richtigkeit der Behauptung IIb.

Wir betrachten jetzt neben R^n und P^n noch einen dritten Raum \mathfrak{R}^n , der auch in der anfangs geschilderten Weise orientiert ist; dann gilt

Satz III. v^n und Ω^n seien Vollkugeln in R^n bzw. P^n mit den Rändern s^{n-1} bzw. Σ^{n-1} , f sei eine eindeutige Abbildung von v^n , Φ eine eineindeutige Abbildung von Ω^n , und es sei $f(v^n) \subset \Omega^n$, $\Phi(\Omega^n) \subset \mathfrak{R}^n$; ferner sei x ein Punkt von v^n , dessen Bild $f(x) = \xi$ weder zu $f(s^{n-1})$ noch zu Σ^{n-1} gehöre; ist weiter $\Phi(\xi) = \zeta$, so gilt die Gleichung

$$u_\xi(f(s^{n-1})) \cdot u_\zeta(\Phi(\Sigma^{n-1})) = u_\zeta(\Phi f(s^{n-1})).$$

(Die Zahlen $u_\xi(\Phi(\Sigma^{n-1}))$ und $u_\zeta(\Phi f(s^{n-1}))$ sind definiert, da infolge der Voraussetzungen über ξ und der Eineindeutigkeit von Φ weder $\xi \subset \Phi(\Sigma^{n-1})$ noch $\zeta \subset \Phi f(s^{n-1})$ ist.)

⁷⁾ Das Zeichen \nsubseteq bedeutet: „nicht enthalten in“.

⁸⁾ Vgl. ¹⁾ c), S. 164, Hilfssatz.

Beweis. π bezeichne, wie früher, die Zentralprojektion von ξ aus auf Σ^{n-1} , analog \wp die Zentralprojektion von ζ aus auf eine Kugel \mathfrak{S}^{n-1} um ζ ; π ist in allen von ξ verschiedenen Punkten von \mathbb{P}^n , \wp in allen von ζ verschiedenen Punkten von \mathfrak{R}^n definiert; πf ist also eine Abbildung von s^{n-1} auf Σ^{n-1} , $\wp \Phi$ eine Abbildung von Σ^{n-1} auf \mathfrak{S}^{n-1} , $\wp \Phi \pi f$ die resultierende Abbildung von s^{n-1} auf \mathfrak{S}^{n-1} ; die Gradzahlen bezeichnen wir wieder mit γ . Dann ist

$$u_{\xi}(f(s^{n-1})) = \gamma(\pi f),$$

$$u_{\zeta}(\Phi(\Sigma^{n-1})) = \gamma(\wp \Phi),$$

also nach der „Produktregel“⁹⁾

$$u_{\xi}(f(s^{n-1})) \cdot u_{\zeta}(\Phi(\Sigma^{n-1})) = \gamma(\wp \Phi \pi f).$$

Es sei nun weiter für jeden von ξ verschiedenen Punkt α von Ω^n derjenige Punkt, der die von α nach $\pi(\alpha)$ gezogene Strecke im Verhältnis $(1-t):t$ teilt, mit $\pi_t(\alpha)$ bezeichnet, so daß also $\pi_0(\alpha) = \pi(\alpha)$, $\pi_1(\alpha) = \alpha$ ist. Dann ist für $0 \leq t \leq 1$ niemals $\xi < \pi_t f(s^{n-1})$, also wegen der Eineindeutigkeit von Φ auch niemals $\zeta < \Phi \pi_t f(s^{n-1})$, die Abbildung $\wp \Phi \pi_t f(s^{n-1})$ ist daher für $0 \leq t \leq 1$ definiert und hängt stetig von t ab. Mithin ist

$$\gamma(\wp \Phi \pi f) = \gamma(\wp \Phi \pi_1 f) = \gamma(\wp \cdot \Phi f).$$

Andererseits ist

$$u_{\zeta}(\Phi f(s^{n-1})) = \gamma(\wp \cdot \Phi f);$$

damit ist Satz III bewiesen.

Nunmehr sind wir in der Lage, den folgenden wichtigen Satz herzuleiten:

Satz IV. *Ist die Abbildung $F(V^n)$ eineindeutig [$F(V^n) \subset \mathbb{P}^n$], so ist entweder für alle nicht auf S^{n-1} gelegenen Punkte a von V^n*

$$u_{F(a)}(F(S^{n-1})) = +1,$$

oder es ist für alle diese Punkte

$$u_{F(a)}(F(S^{n-1})) = -1.$$

Beweis. Die Bildmenge $F(V^n) = M \subset \mathbb{P}^n$ enthält einen inneren Punkt¹⁰⁾; denn andernfalls würde die durch eine beliebige simpliziale Approximation φ_1 der eindeutigen Abbildung F^{-1} gelieferte Bildmenge $\varphi_1(M) = \varphi_1 F(V^n) \subset R^n$ keinen inneren Punkt besitzen; dies stände aber in Widerspruch zu Satz II b, da man φ_1 so wählen kann, daß $\varrho(F^{-1}(\xi), \varphi_1(\xi)) < r$ für jeden Punkt $\xi \in M$, also $\varrho(x, \varphi_1 F(x)) < r$ für jeden Punkt $x \in V^n$ ist.

⁹⁾ Brouwer, ¹⁾ c), S. 326.

¹⁰⁾ Vgl. ¹⁾ c), Satz 1.

In $F(V^n)$ ist also eine Vollkugel Ω^n enthalten; ihr Rand sei Σ^{n-1} . Ist Ω_1^n eine kleinere konzentrische Vollkugel, so ist analog in $F^{-1}(\Omega_1^n)$ eine Vollkugel v^n enthalten. Ist x der Mittelpunkt, s^{n-1} der Rand von v^n , so ist $F(x) \notin F(s^{n-1})$ wegen der Eineindeutigkeit von F , $F(x) \notin \Sigma^{n-1}$ wegen $F(x) \in \Omega_1^n$. Daher sind, wenn wir $F(x) = \xi$, $\Re^n \equiv R^n$ setzen, die Voraussetzungen von Satz III erfüllt; es ist also

$$u_\xi(F(s^{n-1})) \cdot u_x(F^{-1}(\Sigma^{n-1})) = u_x(s^{n-1}) = 1,$$

mithin $u_\xi(F(s^{n-1})) = \pm 1$. Nun kann man s^{n-1} stetig in S^{n-1} überführen, ohne x zu überschreiten; diesem Vorgang entspricht vermöge der eineindeutigen Abbildung F eine Überführung von $F(s^{n-1})$ in $F(S^{n-1})$ unter Vermeidung des Punktes ξ . Nach Satz I ist daher auch $u_\xi(F(S^{n-1})) = \pm 1$.

Da man ferner einen beliebigen Punkt a des Inneren von V^n unter Vermeidung von S^{n-1} in den Punkt x hineinbewegen kann, so folgt ebenso, daß auch $u_{F(a)}(F(S^{n-1})) = u_{F(x)}(F(S^{n-1})) = \pm 1$ ist, womit Satz IV bewiesen ist.

Aus diesem Satz folgt unter Berücksichtigung von Satz II a unmittelbar der Satz von der „Gebietsinvarianz“¹¹⁾.

E^n sei ein „Element“, d. h. das eineindeutige stetige Bild einer Vollkugel V^n in R^n ; infolge der Gebietsinvarianz ist der Rand von E^n mit dem Bild des Randes von V^n identisch.

E^n sei nun der eineindeutigen Abbildung F unterworfen [$F(E^n) \subset P^n$]. Sind v_1^n, v_2^n zwei in E^n gelegene Vollkugeln mit den Rändern s_1^{n-1}, s_2^{n-1} , x_1, x_2 Punkte im Innern von v_1^n bzw. v_2^n , so kann man v_1^n in v_2^n innerhalb E^n stetig und eineindeutig überführen; nach Satz I ist daher $u_{F(x_1)}(F(s_1^{n-1})) = u_{F(x_2)}(F(s_2^{n-1}))$, und diese Zahl ist nach Satz IV entweder $+1$ oder -1 ; sie hängt nicht von der Wahl der Vollkugeln v_i^n und der Punkte x_i^n ab, sondern ist eine Invariante von F ; sie heie der „Grad“ von F , und werde mit $\gamma(F)$ bezeichnet. Je nach dem Vorzeichen von $\gamma(F)$ sagen wir, daß die Bilder der inneren Punkte von E^n „positiv“ oder „negativ“ bedeckt werden.

Aus Satz III folgt die „Produktregel“: $\gamma(F) \cdot \gamma(\Phi) = \gamma(\Phi F)$, wenn Φ eine in $F(E^n)$ definierte eineindeutige Abbildung ist.

Wir wollen nun ein rein topologisches, d. h. von der Verwendung von Koordinatensystemen — wie sie bei der Definition der „Ordnung“ benutzt werden — unabhängiges, Merkmal dafür angeben, wann F den Grad $+1$, wann den Grad -1 hat. Zu diesem Zweck definieren wir: unter einer

¹¹⁾ Dies ist im wesentlichen der zweite Brouwersche Beweis. (Zur Invarianz des n -dimensionalen Gebietes, Math. Annalen 72 (1912)).

„Indikatrix“ verstehen wir den Rand irgendeines Elementes e^n zusammen mit einem inneren Punkt von e^n , dem „Zentrum“ der Indikatrix; eine eindeutige — nicht notwendig eineindeutige — und stetige Abänderung einer Indikatrix in R^n nennen wir eine „Indikatrixdeformation“, falls bei ihr das Bild des Zentrums niemals im Bilde des Randes enthalten ist.

Wir stellen das gesuchte Kriterium nun zunächst für einen Spezialfall her, indem wir den Satz beweisen:

Satz V. Die eineindeutige Abbildung F des im R^n gelegenen Elements E^n auf eine Punktmenge desselben Raumes R^n hat dann und nur dann den Grad $+1$, wenn sich eine in E^n liegende Indikatrix i durch eine Indikatrixdeformation in ihr Bild $F(i)$ überführen läßt.

Beweis. Ein Teilelement e^n von E^n läßt sich als das eineindeutige Bild $e^n = G(v^n)$ einer Vollkugel v^n des R^n auffassen, deren Rand s^{n-1} heiße. Ist nun i eine Indikatrix, die von $G(s^{n-1})$ und dem Bild $G(x)$ eines inneren Punktes x von v^n gebildet wird, und läßt sich i in $F(i)$ vermöge einer Indikatrixdeformation F_t ($0 \leq t \leq 1$) überführen, so ist $u_{F_t G(x)}(F_t G(s^{n-1}))$ unabhängig von t , es ist also $u_{F G(x)}(F G(s^{n-1})) = u_{G(x)}(G(s^{n-1}))$, d. h. $\gamma(FG) = \gamma(G)$; aus der Produktregel folgt daher $\gamma(F) = +1$.

Sei andererseits $\gamma(F) = +1$. Man verschiebe zunächst i so, daß ihr Zentrum in den Mittelpunkt m von v^n fällt; dann halte man das Zentrum fest und lasse die Randpunkte auf den von m ausgehenden Radien auf die Kugel s^{n-1} laufen. Dieser ganze Abänderungsvorgang ist eine Indikatrixdeformation und heiße h_t ($0 \leq t \leq 1$); dabei ist $\gamma(G(v^n)) = u_{G(m)}(G(s^{n-1})) = u_{h_t G(m)}(h_t G(s^{n-1})) = u_m(h_1 G(s^{n-1}))$, wenn man $h_1 G$ als Abbildung von s^{n-1} auf sich auffaßt. Analog definiere man eine Indikatrixdeformation h_t ($3 \geq t \geq 2$) von $i' = F(i)$ so, daß $h_3(i') = i'$ ist und das durch h_2 vermittelte Bild des Zentrums in m , das des Randes auf s^{n-1} liegt; dabei ist, analog dem obigen, $\gamma(FG(v^n)) = \gamma(h_2 FG(s^{n-1}))$, $h_2 FG$ als Abbildung von s^{n-1} auf sich aufgefaßt. Infolge der Voraussetzung $\gamma(F) = +1$ ist nun $\gamma(G) = \gamma(FG)$, also $\gamma(h_1 G) = \gamma(h_2 FG)$. Daher gehören $h_1 G(s^{n-1})$ und $h_2 FG(s^{n-1})$ zu derselben „Klasse“, d. h. das Bild $h_1 G(s^{n-1})$ läßt sich auf s^{n-1} stetig in das Bild $h_2 FG(s^{n-1})$ überführen¹²⁾; dieser Prozeß ist, unter Festhaltung des Zentrums m , eine Indikatrixdeformation von $h_1(i)$ in $h_2(i')$ und heiße h_t ($1 \leq t \leq 2$). h_t stellt also für $0 \leq t \leq 3$ eine Indikatrixdeformation von i in i' dar.

Will man nun den allgemeinen Fall $P^n \cong R^n$ behandeln, so bemerkt man, daß ohne irgendein Koordinatensystem und ohne jegliche willkür-

¹²⁾ H. Hopf, Abbildungsklassen n -dimensionaler Mannigfaltigkeiten, Math. Annalen 96 (1926).

liche Festsetzung das Vorzeichen von $\gamma(F)$ offenbar unbestimmt ist. Man muß die „Orientierungen“ der beiden Räume dadurch in Verbindung miteinander bringen, daß man einer ausgezeichneten eineindeutigen Abbildung $H(e^n)$ eines bestimmten Elements $e^n \in R^n$ auf ein Element $H(e^n) = \varepsilon^n \in P^n$ den Grad $+1$ beilegt; dadurch werden die beiden Räume „relativ orientiert“. Im Fall $R^n \equiv P^n$ muß die „Eichabbildung“ H die Identität sein; bei den am Anfang des Paragraphen gemachten Festsetzungen ist sie die Abbildung $\xi_\nu = x_\nu$ ($\nu = 1, 2, \dots, n$); demnach sind R^n und P^n relativ orientiert, falls jeder der beiden Räume orientiert ist.

Wir setzen nun stets R^n und P^n als relativ orientiert voraus; dann ergibt sich das gewünschte Kriterium für den Wert von $\gamma(F(E^n))$ durch Anwendung der Produktregel: man bilde das Bildelement $F(E^n)$ durch eine eineindeutige Abbildung Δ auf ε^n oder einen Teil von ε^n ab und betrachte außerdem die Abbildung $D(E^n) = H^{-1} \Delta F(E^n)$; dann ist $\gamma(D) = \gamma(\Delta) \cdot \gamma(F)$. Da $\gamma(D)$ und $\gamma(\Delta)$ nach Satz V zu bestimmen sind, ist somit $\gamma(F)$ eineindeutig und topologisch invariant festgelegt. Somit können wir sagen:

Satz Va. Sind R^n und P^n relativ orientiert, so kann man unter den eineindeutigen Abbildungen von Gebieten des R^n auf Gebiete des P^n eine topologisch invariante Unterscheidung machen zwischen solchen, die „positive“ und solchen, die „negative“ Bedeckungen liefern.

Wir betrachten nun beliebige eindeutige, nicht notwendig eindeutig umkehrbare Abbildungen f des Elements E^n mit $f(E^n) \subset P^n$; s^* sei der Rand von E^n . Zwei Abbildungen $f_1(E^n)$ und $f_2(E^n)$ sollen als zur selben „Klasse“ gehörig gelten, wenn sie am Rande übereinstimmen, wenn also $f_1(x) = f_2(x)$ für jeden Punkt $x \in s^*$ ist. Zwei Abbildungen f_1, f_2 derselben Klasse lassen sich innerhalb der Klasse, d. h. ohne Änderung auf s^* , stetig ineinander überführen; man braucht etwa nur alle Punkte $f_2(x)$ geradlinig und gleichförmig in einem festen Zeitintervall in die entsprechenden Punkte $f_1(x)$ laufen zu lassen.

Eine Abbildung $f(E^n)$ heie im Punkt $\xi \in P^n$ „glatt“, wenn es im Innern von E^n endlich viel Gebiete G_α ($\alpha = 1, \dots, k$) gibt, so da f in jedem von ihnen eineindeutig und da $\xi \in f(E^n - \sum_{\alpha=1}^k G_\alpha)$ ist; die Definition habe auch für $k = 0$ Sinn; d. h. f heie in jedem nicht zu $f(E^n)$ gehrigen Punkt glatt.

Satz VIa. f^ sei eine Abbildung des Randes s^* von E^n , ξ ein nicht zu $f^*(s^*)$ gehriger Punkt von P^n . Dann gibt es in der durch f^* bestimmten Klasse \mathfrak{F} Abbildungen, die in ξ glatt sind.*

Beweis. V^n sei eine Vollkugel des R^n mit dem Rand S^{n-1} und durch eine eindeutige Abbildung $G(V^n) = E^n$ auf E^n bezogen; dann ist $G(S^{n-1}) = s^*$, also $\xi \notin f^*G(S^{n-1})$. Es sei $u_\xi(f^*G(S^{n-1})) = c$, p und q seien zwei nicht negative ganze Zahlen mit $p - q = c$; $v_1^n, v_2^n, \dots, v_p^n, v_{p+1}^n, \dots, v_{p+q}^n$ seien zueinander fremde Vollkugeln im Innern von V^n mit den Rändern $s_1^{n-1}, s_2^{n-1}, \dots, s_{p+q}^{n-1}$. Man bilde die v_i^n ($i = 1, 2, \dots, p + q$) durch eindeutige Abbildungen F_i — (z. B. Ähnlichkeitsabbildungen) — so auf Umgebungen von ξ ab, daß $\gamma(F_i) = +1$ für $i \leq p$, $\gamma(F_i) = -1$ für $i > p$ ist. Jedem Punkt x von $S^{n-1}, s_1^{n-1}, s_2^{n-1}, \dots, s_{p+q}^{n-1}$ ordne man denjenigen Vektor $v(x)$ zu, der in ξ beginnt und in $f^*G(x)$ bzw. $F_i(x)$ endet. Diese Vektorfelder besitzen keine Nullstelle. Ordnet man dann dem Punkt x den Schnittpunkt von $v(x)$ mit einer Kugel Σ^{n-1} um ξ zu, so entstehen Abbildungen $\Phi, \varphi_1, \varphi_2, \dots, \varphi_{p+q}$ von $S^{n-1}, s_1^{n-1}, \dots, s_{p+q}^{n-1}$ auf Σ^{n-1} ; dabei ist $\gamma(\Phi) = c$, $\gamma(\varphi_i) = +1$ ($i \leq p$), $\gamma(\varphi_i) = -1$ ($i > p$), also, da $c = p - q$ ist, $\sum_{i=1}^{p+q} \gamma(\varphi_i) = \gamma(\Phi)$. Infolge

dieser Gleichheit ist die „Randwertaufgabe“ lösbar, welche verlangt, in dem durch Herausnahme der Innengebiete der v_i^n aus V^n entstandenen Körper \bar{V}^n eine stetige Zuordnung nirgends verschwindender Vektoren zu definieren, die die bereits vorhandenen Randwerte besitzt¹³). Durch eine derartige Vektorverteilung ist eine Abbildung $F(\bar{V}^n)$ bestimmt, die ξ nicht bedeckt: ist $v(x)$ der dem Punkt x zugeordnete Vektor, so sei $F(x)$ der Endpunkt des in ξ angetragenen Vektors $v(x)$. F schließt sich stetig an f^*G und die F_i an. Der so in V^n definierten Abbildung entspricht vermöge G eine zur Klasse \mathfrak{F} gehörige Abbildung von E^n , die in ξ glatt ist.

Satz VI b. *Ist f eine in ξ glatte Abbildung der Klasse \mathfrak{F} , und bezeichnen p und q die Anzahlen der durch f gelieferten positiven bzw. negativen Bedeckungen des Punktes ξ , so ist die Differenz $p - q$ eine Konstante von \mathfrak{F} und ξ , d. h. lediglich von der Lage des Punktes ξ und von der Randabbildung f^* , aber nicht von der speziellen Wahl von f abhängig.*

Beweis. Da es eindeutige Abbildungen der Vollkugel V^n auf sich vom Grade -1 gibt, z. B. Spiegelungen, gibt es nach der Produktregel unter den eindeutigen Beziehungen zwischen V^n und E^n gewiß solche des Grades $+1$. G sei eine solche: $E^n = G(V^n)$. $F(V^n) = fG(V^n)$ ist in ξ glatt und liefert, da $\gamma(G) = +1$ ist, ebenso wie $f(E^n)$, p positive, q negative Bedeckungen von ξ . F erzeugt in der beim Beweis von Satz VI a geschilderten Weise ein Vektorfeld in V^n , das nur in den $p + q$

¹³) Siehe ¹²), § 5. (Die Lösbarkeit der Randwertaufgabe für \bar{V}^n ist einer Aussage über Abbildungen einer S^{n-1} äquivalent.)

Originalpunkten von ξ Nullstellen besitzt. Ist x_i eine solche Nullstelle und s_i^{n-1} eine x_i , aber keine weitere Nullstelle einschließenden Kugel, so heißt $u_\xi(F(s_i^{n-1}))$ in der üblichen Bezeichnung der „Index“ der Nullstelle, und die Summe aller Indizes ist gleich dem Grade der durch die Vektoren vermittelten Abbildung des Randes S^{n-1} auf eine „Richtungskugel“ von P^n , also gleich $u_\xi(F(S^{n-1}))$ ¹⁴⁾. Da andererseits nach Satz IV $u_\xi(F(s_i^{n-1})) = +1$ oder -1 ist, je nachdem x_i zu einer positiven oder negativen Bedeckung von ξ gehört, ist die Summe der Indizes gleich $p - q$; also ist $p - q = u_\xi(F(S^{n-1}))$; da diese Zahl nur von ξ und der Randabbildung $f^*(s^*)$ abhängt, ist Satz VIb bewiesen.

Die auf Grund der Sätze VIa, VIb eindeutig und topologisch invariant der Klasse \mathfrak{F} und dem Punkt ξ zugeordnete Zahl $p - q$ nennen wir den „Grad der zu \mathfrak{F} gehörigen Abbildungen $f(E^n)$ im Punkte ξ “ und bezeichnen sie durch $\gamma_\xi(f(E^n))$. Der früher für eineindeutige Abbildungen F definierte „Grad“ $\gamma(F)$ ist demnach zu präzisieren als „Grad $\gamma_{F(x)}(\mathfrak{F}(E^n))$ in den Bildern der inneren Punkte x von E^n “.

$\gamma_\xi(f(E^n))$ bleibt nicht nur ungeändert, solange f der Klasse \mathfrak{F} angehört, sondern aus Satz I und der Rolle, die die Zahl $u_\xi(F(S^{n-1}))$ beim Beweis von Satz VIb spielt, folgt der allgemeinere

Satz VIIa. *Der Grad $\gamma_\xi(f(E^n))$ bleibt bei stetiger Abänderung von f konstant, wenn nur bei ihr niemals ξ zu der jeweiligen Bildmenge des Randes s^* gehört,*

und ebenso ergibt sich

Satz VIIb. *Für zwei Punkte ξ_1, ξ_2 , die man ineinander überführen kann, ohne das Bild des Randes zu treffen, ist der Grad derselbe; insbesondere ist er also in einer Umgebung von ξ , die das Bild des Randes ausschließt, konstant.*

Aus den Sätzen VIa und VIb folgt die „Summenregel“.

Satz VIII. *Läßt sich die Menge M derjenigen Punkte von E^n , die durch f auf den Punkt ξ abgebildet werden, in endlich viele, zueinander fremde Teilelemente $e_1^n, e_2^n, \dots, e_k^n$ von E^n einschließen, so ist*

$$\gamma_\xi(f(E^n)) = \sum_{\times} \gamma_\xi(f(e_{\times}^n)).$$

Der Beweis ergibt sich, indem man f in den e_{\times}^n glättet und dann die positiven und negativen Bedeckungen von ξ abzählt.

Aus der Willkür, mit der man p und q im Beweise von Satz VIa wählen kann, und aus $p + q \geq p - q = \gamma_\xi(f(E^n))$ ist ersichtlich:

¹⁴⁾ Bezüglich dieser Addition der Indizes siehe z. B. Feigl, Fixpunktsätze für spezielle n -dimensionale Mannigfaltigkeiten, § 2; Math. Annalen 98 (1927).

Satz IX. Der Grad $\gamma_\xi(f(E^n))$ läßt sich seinem Betrage nach als die Mindestzahl der eindeutigen Bedeckungen einer Umgebung von ξ charakterisieren, die durch — in ξ glatte — Abbildungen der Klasse \mathfrak{F} erreicht wird. Wenn die Mindestzahl von Bedeckungen vorliegt, so sind diese sämtlich gleichsinnig, und zwar positiv oder negativ je nach dem Vorzeichen von $\gamma_\xi(f)$. p und q unterliegen nur den Beschränkungen:

$$p \geq 0, \quad q \geq 0, \quad p - q = \gamma_\xi(f).$$

Hierbei darf $\gamma_\xi(f) = 0$ sein; d. h.

Satz IXa. Ist $\gamma_\xi(f) = 0$, so gibt es in der Klasse \mathfrak{F} Abbildungen, für die ξ nicht zur Bildmenge von E^n gehört.

Es sei besonders darauf hingewiesen, daß es sich bei Satz IX um eindeutige Bedeckungen einer Umgebung von ξ , nicht um solche des Punktes ξ allein handelt. Bezüglich der letzteren gilt im Gegenteil

Satz X. Unter den Abbildungen der Klasse \mathfrak{F} gibt es solche, die einen und nur einen Punkt auf ξ abbilden.

Um eine solche Abbildung herzustellen, betrachten wir, wie beim Beweise von Satz VIa, die Vollkugel V^n , die durch G eindeutig auf E^n bezogen ist. Wir bilden einen inneren Punkt x von V^n auf ξ , und wenn x^* Randpunkt von V^n ist, die Strecke xx^* proportional auf die Strecke von ξ nach $f^*G(x^*)$ ab. Diese in V^n erklärte Abbildung übertragen wir mittels G auf E^n .

Der Gegenstand unserer weiteren Untersuchung wird — nach Erledigung einiger Hilfssätze — die Beantwortung der Frage sein, ob es bei Abbildungen beliebiger n -dimensionaler Mannigfaltigkeiten eine, als „Grad“ zu bezeichnende Zahl gibt, die durch Eigenschaften gekennzeichnet ist, wie sie in den Sätzen VIa, VIb, VIIa, VIIb, VIII ausgesprochen sind. Diese Frage wird bejaht werden.

Im zweiten Teil dieser Abhandlung⁴⁾ werden wir feststellen, ob bzw. mit welchen Abänderungen sich die Sätze IX, IXa, X auf diesen „Grad“ übertragen lassen.

§ 2.

Hilfssätze über die Erweiterung der Definitionsbereiche von Abbildungen.

In diesem Paragraphen bedeutet stets k^l einen l -dimensionalen Kubus, der z. B. in einem euklidischen (x^1, x^2, \dots, x^l) -Raum durch $0 \leq x^\lambda \leq 1$ ($\lambda = 1, 2, \dots, l$) definiert sei, und \mathfrak{K}^l den Rand von k^l ; ferner ist ξ ein fester Punkt in dem mit einem rechtwinkligen $(\xi^1, \xi^2, \dots, \xi^n)$ -Koordinatensystem versehenen Raum P^n . Mit f und F werden stets ein-

deutige und stetige Abbildungen von Kuben k^l oder von Teilen solcher Kuben auf Punktmengen des P^n bezeichnet.

Hilfssatz I. Es sei $l < n$; auf $\mathfrak{R} k^l$ sei eine Abbildung f so definiert, daß $\xi \notin f(\mathfrak{R} k^l)$ ist.

Dann gibt es eine auf $\mathfrak{R} k^l$ mit f identische Abbildung F von k^l , so daß $\xi \notin F(k^l)$ ist.

Beweis¹⁵⁾. k_1^l sei der zu k^l konzentrische seitenparallele Kubus mit der halben Seitenlänge, x_0 der Mittelpunkt der beiden Kuben. Ist $x \in \mathfrak{R} k^l$, so bezeichne $p(x)$ den Punkt auf $\mathfrak{R} k_1^l$, in den x von x_0 aus projiziert wird.

Man konstruiere eine so feine simpliziale Approximationsabbildung $f_1(\mathfrak{R} k^l)$ von $f(\mathfrak{R} k^l)$, daß für jeden Punkt $x \in \mathfrak{R} k^l$ der Fehler $\varrho(f(x), f_1(x))$ kleiner als die Entfernung $\varrho(\xi, f(\mathfrak{R} k^l))$ ist; dann liegt ξ auf keiner der Strecken $\overline{f(x)f_1(x)}$. Wir definieren jetzt: 1) $\overline{F(x)} = f(x)$, $F(p(x)) = f_1(x)$ für $x \in \mathfrak{R} k^l$. 2) F ist auf jeder Strecke $x p(x)$ linear. Damit ist F in $k^l - k_1^l + \mathfrak{R} k_1^l$ in der gewünschten Weise erklärt.

Die Bildmenge $F(\mathfrak{R} k_1^l) = f_1(\mathfrak{R} k^l)$ besteht aus einer endlichen Anzahl $(l - 1)$ -dimensionalen Simplexe; der von ihr und dem Punkt ξ aufgespannte Kegel ist daher l -dimensional, und, da $l < n$ ist, gibt es Geraden durch ξ , die diesem Kegel nicht angehören und mithin keinen Punkt von $F(\mathfrak{R} k_1^l)$ enthalten. α sei eine solche Gerade, μ ein von ξ verschiedener Punkt auf α ; ist $y \in \mathfrak{R} k_1^l$, so liegt also ξ nicht auf der Geraden $\overline{\mu F(y)}$. Wir erklären daher $F(k_1^l)$ in der geforderten Weise, wenn wir festsetzen: 1) $F(x_0) = \mu$; 2) ist $y \in \mathfrak{R} k_1^l$, so ist F auf der Strecke $\overline{\mu F(y)}$ linear.

Zusatz. Sind statt des einen Punktes ξ endlich viele nicht in $f(\mathfrak{R} k^l)$ enthaltene Punkte $\xi_1, \xi_2, \dots, \xi_r$ gegeben, so läßt sich F so konstruieren, daß $\xi_\varrho \notin F(k^l)$ ($\varrho = 1, 2, \dots, r$) ist. Man hat im Beweis von Hilfssatz I nämlich nur die Gerade α so zu wählen, daß sie auf keinem der r von den ξ_ϱ und der simplizialen Menge $f_1(\mathfrak{R} k^l)$ aufgespannten Kegel liegt.

Hilfssatz II. Es sei jetzt $l = n$; auf $\mathfrak{R} k^n$ sei f so definiert, daß $\xi \notin f(\mathfrak{R} k^n)$ ist. x_0 sei ein willkürlicher Punkt im Innern von k^n .

Dann läßt sich $F(k^n)$ so definieren, daß $F(\mathfrak{R} k^n) = f(\mathfrak{R} k^n)$, $F(x_0) = \xi$, $F(x) \neq \xi$ für $x \neq x_0$ ist¹⁶⁾.

¹⁵⁾ Dieser Beweis ist eine Spezialisierung des, mir aus einer unveröffentlichten Bemerkung von Brouwer bekannten, Beweises der Tatsache, daß im R^n eine p -dimensionale und eine q -dimensionale Mannigfaltigkeit nicht „verschlungen“ sein können, wenn $p + q < n - 1$ ist.

¹⁶⁾ Enthalten in Satz X, § 1.

Beweis. Man definiere F als lineare Abbildung aller x_0 mit den Punkten von $\mathfrak{R}(k^n)$ verbindenden Strecken.

Hilfssatz III. Es sei $l = n + 1$; k^{n+1} sei durch $0 \leq x^\nu \leq 1$ im R^{n+1} gegeben. $f(\mathfrak{R} k^{n+1})$ sei so definiert, daß es auf $\mathfrak{R}(k^{n+1})$ höchstens endlich viele ξ -Stellen gibt, d. h. Punkte x , für die $f(x) = \xi$ ist.

Dann läßt sich $F(k^{n+1})$ so definieren, daß $F(\mathfrak{R} k^{n+1}) = f(\mathfrak{R} k^{n+1})$ ist und daß es auf jeder n -dimensionalen Ebene $x^\nu = \text{konst.}$ ($\nu = 1, 2, \dots, n + 1$) höchstens endlich viele ξ -Stellen gibt.

Beweis. x_0 sei ein innerer Punkt von k^{n+1} , der mit keiner der ξ -Stellen auf $\mathfrak{R}(k^{n+1})$ in einer zu einer Koordinatenebene parallelen n -dimensionalen Ebene liegt. Man setze $F(x_0) = \xi$ und definiere F als lineare Abbildung der Strecken, die x_0 mit den Punkten von $\mathfrak{R} k^{n+1}$ verbinden. Dann besteht die Menge der ξ -Stellen von $F k^{n+1}$ aus den Verbindungsstrecken von x_0 mit den ξ -Stellen auf $\mathfrak{R} k^{n+1}$. Da auf keiner dieser Strecken eine x^ν -Koordinate konstant ist, besitzt F die gewünschte Eigenschaft.

Hilfssatz IIa. Q sei eine abgeschlossene Teilmenge von k^n , q die Menge der gemeinsamen Grenzpunkte von Q und $k^n - Q$. In Q sei f so definiert, daß $\xi \notin f(q)$ ist.

Dann gibt es eine in Q mit f identische Abbildung $F(k^n)$, die in $k^n - Q$ höchstens endlich viele ξ -Stellen besitzt.

Beweis. Die Koordinaten $\xi^1, \xi^2, \dots, \xi^n$ der durch f vermittelten Bildpunkte sind stetige Funktionen in Q . Nach einem bekannten Satz¹⁷⁾ läßt sich der Definitionsbereich dieser Funktionen auf ganz k^n erweitern; dadurch wird eine in Q mit f identische Abbildung $F^*(k^n)$ erklärt.

Die Menge N der in der abgeschlossenen Menge $k^n - Q + q$ gelegenen ξ -Stellen von F^* ist abgeschlossen und zu q punktfremd; sie besitzt also eine positive Entfernung a von q . Wir zerlegen k^n in seitenparallele, kongruente Teilkuben, deren Durchmesser kleiner als a sind. Ist dann L die Menge derjenigen Teilkuben, die Punkte von N enthalten, so ist $L \subset k^n - Q$.

In $k^n - L$, also insbesondere in Q , sowie in den gemeinsamen Grenzpunkten von $k^n - L$ und L setzen wir $F = F^*$; dann ist dort $F(x) \neq \xi$ für $x \notin Q$.

¹⁷⁾ Bewiesen von Bohr, Brouwer, Hahn, Hausdorff, Lebesgue, Riesz, Tietze, Urysohn, de la Vallée Poussin; genauere Literaturangabe bei Urysohn, Über die Mächtigkeit der zusammenhängenden Mengen, Math. Annalen 94 (1926), S. 293, und von Kerékjártó, Vorlesungen über Topologie (Berlin 1923), S. 75.

Wir haben F noch in L so zu erklären, daß nur endlich viele ξ -Stellen auftreten. Mit k_i^l ($i = 1, 2, \dots, r_i$; $0 \leq l \leq n - 1$) bezeichnen wir die im Innern von L gelegenen l -dimensionalen Randkuben der Teilkuben k_i^n ($i = 1, 2, \dots, r_n$) von L . Wir erklären nun F zunächst auf allen k_i^0 (d. h. in den Eckpunkten) willkürlich mit der einzigen Einschränkung $\xi \neq F(k_i^0)$; sodann erklären wir F auf Grund von Hilfssatz I für alle k_i^1 , darauf für alle k_i^2 , usw. bis F auf allen k_i^{n-1} , also auf den Rändern aller k_i^n , ξ -stellenfrei definiert ist. Schließlich bestimmen wir auf Grund von Hilfssatz II F in den k_i^n so, daß in jedem genau eine ξ -Stelle x_i liegt.

Zusatz 1. Da sich die Punkte x_i in den k_i^n willkürlich wählen lassen, kann man eine vorgeschriebene, in $k^n - Q$ nirgends dichte Menge frei von ξ -Stellen halten.

Zusatz 2. Liegt die Menge $f(Q)$ ganz in einer ξ enthaltenden Vollkugel Ω^n , so läßt sich F so bestimmen, daß auch $F(k^n) \subset \Omega^n$ ist. Denn man braucht, um das zu erreichen, nur jeden etwa außerhalb Ω^n gelegenen Punkt $F(x)$ durch denjenigen Punkt des Randes von Ω^n zu ersetzen, in welchen $F(x)$ von ξ aus projiziert wird.

Hilfssatz IIIa. k^{n+1} sei durch $0 \leq x^\nu \leq 1$ im R^{n+1} gegeben; Q sei eine abgeschlossene Teilmenge von k^{n+1} , q die Menge der gemeinsamen Grenzpunkte von Q und $k^{n+1} - Q$, $\Re k^{n+1}$ der Rand von k^{n+1} (der auch Punkte von Q enthalten darf). In $Q + \Re k^{n+1}$ sei die Abbildung f definiert; sie habe in q keine, in dem nicht zu Q gehörigen Teil von $\Re k^{n+1}$ höchstens endlich viele ξ -Stellen.

Dann gibt es eine in $Q + \Re k^{n+1}$ mit f identische Abbildung $F(k^{n+1})$, die in jeder Ebene $x^\nu = \text{konst.}$ ($\nu = 1, 2, \dots, n + 1$) höchstens endlich viele nicht zu Q gehörige ξ -Stellen besitzt.

Der Beweis ist dem von Satz IIa völlig analog: F^* und L haben dieselbe Bedeutung wie früher. Man setzt $F = F^*$ im Äußern von L , auf der gemeinsamen Grenze von L und $k^{n+1} - L$, sowie auf $\Re k^{n+1}$. Man hat F noch im Innern von L zu erklären. Dies geschieht ebenso wie früher; nur hat man jetzt auch noch $(n + 1)$ -dimensionale Teilkuben zu behandeln, und man tut dies mit Hilfe von Hilfssatz III.

Zusatz. Der zu Hilfssatz IIa bezüglich $F(k^n)$ gemachte „Zusatz 2“ gilt auch jetzt bezüglich $F(k^{n+1})$.

Hilfssatz IIIb. k^{n+1}, Q, q haben dieselbe Bedeutung wie in Hilfssatz IIIa. $f(Q)$ habe aber jetzt in q endlich viele ξ -Stellen x_1, x_2, \dots, x_a .

Dann gibt es eine in Q mit f identische Abbildung $F(k^{n+1})$ mit folgender Eigenschaft: die Menge der nicht zu Q gehörigen, in irgend-einer Ebene $x^\nu = \text{konst.}$ liegenden ξ -Stellen besitzt höchstens endlich viele Häufungspunkte.

Beweis. Man nehme mit k^{n+1} eine solche Unterteilung in Teilkuben $k_1^{n+1}, k_2^{n+1}, \dots, k_i^{n+1}$ vor, daß jeder von diesen höchstens einen der Punkte x_α enthält und daß kein Punkt x_α auf der gemeinsamen Grenze zweier k_β^{n+1} liegt. Nun definiere man F zunächst mittels Hilfssatz IIa auf den n -dimensionalen Randkuben k_γ^n der k_β^{n+1} so, daß auf jedem k_γ^n höchstens endlich viele ξ -Stellen entstehen. Darauf erkläre man auf Grund von Hilfssatz IIIa F in denjenigen k_β^{n+1} , welche keinen Punkt x_α enthalten derart, daß auf jeder Ebene $x^\nu = \text{konst.}$ höchstens endlich viele neue ξ -Stellen hinzukommen. Schließlich betrachte man die Kuben $k_1^{n+1}, k_2^{n+1}, \dots, k_a^{n+1}$, in welchen die Punkte x_1, x_2, \dots, x_a liegen, einzeln:

In k_a^{n+1} konstruiere man eine Folge seitenparalleler, gegen x_α konvergierender Kuben $k_{a,0}^{n+1} > k_{a,1}^{n+1} > k_{a,2}^{n+1} > \dots$; g_i sei der von $k_{a,i-1}^{n+1}$ und $k_{a,i}^{n+1}$ begrenzte abgeschlossene Bereich, φ_i sei das Maximum der Entfernung $\varrho(\xi, f(x))$ für $x < q \cdot g_i$. Dann definieren wir F zunächst in g_1 , indem wir genau wie oben vorgehen: wir zerlegen g_1 in Teilwürfel und erklären F zunächst auf deren Rändern, darauf in ihrem Innern, so daß auf jeder Ebene $x_\nu = \text{konst.}$ nur endlich viele ξ -Stellen auftreten. Dabei sorgen wir dafür, daß $\varrho(\xi, F(x)) < \varphi_2$ für $x < \Re k_{a,1}^{n+1}$ und $\varrho(\xi, F(x)) < \varphi_1$ für $x < g_1$ wird; mit Rücksicht auf die zu Satz IIa und IIIa gemachten Zusätze können wir dies erreichen. So definieren wir der Reihe nach F in g_2, g_3, \dots derart, daß immer $\varrho(\xi, f(x)) < \varphi_i$ für $x < g_i$ ist. Dann folgt aus $\lim_{x \rightarrow x_\alpha} \varrho(\xi, F(x)) = 0$ und $F(x_\alpha) = \xi$, daß F auch in x_α stetig ist. Da die Abbildung F überdies in keiner Ebene einen von den x_α verschiedenen Häufungspunkt nicht in Q gelegener ξ -Stellen besitzt, erfüllt sie alle Forderungen.

Aus Hilfssatz IIa und IIIb ergeben sich unmittelbar die folgenden Sätze:

Hilfssatz IIa*. Q_0 sei eine abgeschlossene Teilmenge von k_0^{n-1} , q_0 die Menge der gemeinsamen Grenzpunkte von Q_0 und $k^{n-1} - M_0$. In Q_0 sei für $0 \leq t \leq 1$ eine von t stetig abhängende Schar von Abbildungen f_t erklärt, die während des ganzen Intervalls $0 \leq t \leq 1$ in q_0 keine, in Q_0 höchstens endlich viele ξ -Stellen besitzt. f_0 und f_1 seien überdies in dem ganzen k_0^{n-1} definiert und zwar ξ -stellenfrei.

Dann gibt es eine in Q_0 mit f_t identische Schar von Abbildungen $F_t(k_0^{n-1}) [F_0 \equiv f_0, F_1 \equiv f_1]$, für die während des ganzen Intervalls $0 \leq t \leq 1$ nur endlich viele ξ -Stellen auftreten.

Hilfssatz IIIb*. Q_0 sei eine abgeschlossene Teilmenge von k_0^n , q_0 die Menge der gemeinsamen Grenzpunkte von Q_0 und $k_0^n - Q_0$. In $Q_0 + \Re k_0^n$ sei für $0 \leq t \leq 1$ eine von t stetig abhängende Schar von

Abbildungen f_t mit folgenden Eigenschaften gegeben: 1) während des ganzen Intervalls $0 \leq t \leq 1$ treten in $Q_0 + \Re k_0^n$ nur endlich viele ξ -Stellen auf; 2) für jeden einzelnen Wert von t hat die Menge der ξ -Stellen in Q_0 höchstens endlich viele Häufungspunkte. f_0 und f_1 seien überdies in dem ganzen k_0^n definiert, und zwar mit höchstens endlich vielen ξ -Stellen.

Dann gibt es eine in $Q_0 + \Re k_0^n$ mit f_t identische, derartige Schar von Abbildungen $F_t(k_0^n)$ [$F_0 \equiv f_0, F_1 \equiv f_1$], daß für jede einzelne Abbildung $F_t(k_0^n)$ die Menge der ξ -Stellen höchstens endlich viele Häufungspunkte besitzt.

Die Beweise von IIa* und IIIb* sind in IIa und IIIb enthalten: man deute in diesen Sätzen die x^n - bzw. x^{n+1} -Achse als t -Achse, den durch $t=0$ bestimmten Randkubus von k^n bzw. k^{n+1} als k_0^{n-1} bzw. k_0^n ; Q besteht dann aus denjenigen Punkten von k^n bzw. k^{n+1} , deren senkrechte Projektionen auf k_0^{n-1} bzw. k_0^n zu Q_0 gehören, und im Fall des Satzes IIa außerdem aus den beiden durch $t=0$ und $t=1$ ausgezeichneten $(n-1)$ -dimensionalen Randkubus. [In Hilfssatz IIIb sind die Randkuben bereits in der Formulierung des Satzes berücksichtigt.]

Zusatz 1. Aus IIa Zusatz 1 folgt, daß man in IIa* eine in k_0^{n-1} nirgends dichte Menge frei von ξ -Stellen halten kann.

Zusatz 2. Die in „Zusatz 2“ zu IIa gemachte Bemerkung gilt analog auch für IIa* und IIIb*.

§ 3.

Topologische Mannigfaltigkeiten.

Unter einer n -dimensionalen „topologischen Mannigfaltigkeit“, oder kurz *Mannigfaltigkeit* M^n , verstehen wir einen zusammenhängenden topologischen Raum, in dem es ein abzählbares vollständiges System von Umgebungen — im Sinne von Hausdorff — gibt, von welchen eine jede sich eindeutig und stetig auf das Innere einer n -dimensionalen Vollkugel oder, was dasselbe ist, auf den ganzen euklidischen R^n abbilden läßt. Statt von „kompakten“ und „nichtkompakten“ sprechen wir von „geschlossenen“ und „offenen“ Mannigfaltigkeiten. Aus der Definition folgt, daß jeder zusammenhängende offene Teil einer M^n , also z. B. jedes Gebiet im R^n , selbst eine offene, n -dimensionale Mannigfaltigkeit ist.

Eine Punktmenge von M^n , die das eineindeutige und stetige Bild einer abgeschlossenen, n -dimensionalen Vollkugel V^n ist, heißt ein „Element“ E^n von M^n ; nach dem Satz von der Gebietsvarianz entsprechen die im Sinne des Umgebungsbegriffs von M^n definierten „inneren Punkte“ und „Randpunkte“ von E^n den inneren bzw. den Randpunkten von V^n . Ist x ein Punkt von M^n , so gibt es auf Grund der euklidischen Struktur einer Um-

gebung von x Elemente, die x im Innern enthalten. Aus den bekannten Überdeckungssätzen, auf die Innengebiete dieser Elemente angewandt, folgt: Jeder (echte oder unechte) Teil m von M^n läßt sich derart mit einer abzählbaren Menge von Elementen E_1^n, E_2^n, \dots bedecken, daß jeder Punkt von m im Innern mindestens eines E_i^n enthalten ist; ist m kompakt, so leistet bereits eine endliche Menge von Elementen eine derartige Überdeckung.

Wir wollen nun im Fall einer geschlossenen M^n ein endliches, im Fall einer offenen M^n ein abzählbar endliches System E_1^n, E_2^n, \dots von Elementen, das M^n derart überdeckt, eine „Basis“ nennen. Die Elemente einer Basis denken wir uns durch euklidische Vollkugeln V_1^n, V_2^n, \dots dargestellt. Zwischen Teilen verschiedener V_i^n sind vermöge der Koinzidenz der entsprechenden Punkte in M^n eineindeutige und stetige Abbildungen definiert.

V_1^n, V_2^n, \dots sei eine Darstellung einer Basis von M^n . $x'_0, x_1, x'_1, x_2, x'_2, \dots, x_{k-1}, x'_{k-1}, x_k$, seien innere Punkte der V_i^n , derart, daß x'_{i-1} mit x_i stets in derselben V_i^n liegt, während x_i und x'_i entweder identisch oder wenigstens Repräsentanten desselben Punktes in M^n sind. Der Gesamtheit der Strecken $x'_{i-1} x_i$ ($i = 1, 2, \dots, k$) entspricht in M^n ein „Weg“ zwischen den durch x'_0 und x_k dargestellten Punkten a und b von M^n . Man kann je zwei Punkte von M^n durch einen derartigen Weg verbinden; ist nämlich v die Menge der mit a verbindbaren Punkte, so ist v einerseits offen, da, wenn b mit a verbindbar und c ein Punkt einer euklidischen Umgebung von b ist, c mit b , also auch mit a verbunden werden kann; andererseits ist v abgeschlossen, da ein Häufungspunkt c von mit a verbindbaren Punkten b_i in einer seiner Umgebungen mit gewissen b_i , also auch mit a verbunden werden kann. Mithin ist $v = M^n$, da M^n nach Definition zusammenhängend ist.

Eine Mannigfaltigkeit M^n ist „im kleinen kompakt“ und es gilt das „2. Abzählbarkeitsaxiom“; sie ist daher „metrisierbar“¹⁸⁾, d. h. je zwei Punkten x, y läßt sich eine nicht-negative „Entfernung“ $\varrho(xy) = \varrho(yx)$ zuordnen, die die Bedingung $\varrho(xy) + \varrho(yz) \geq \varrho(xz)$ erfüllt, die nur dann verschwindet, wenn $x = y$ ist, und für die dann und nur dann $\lim_{i \rightarrow \infty} \varrho(x_i, x) = 0$ ist, wenn die Punktfolge x_i gegen x konvergiert; die letzte Bedingung besagt, daß der durch die Metrik neu eingeführte Umgebungsbegriff dem ursprünglichen äquivalent ist.

Ein Element E^n von M^n (das kein Element unserer Basis zu sein braucht) heißt „orientiert“, sobald in einer bestimmten, E^n repräsentierenden Vollkugel V^n eine positive Orientierung (s. § 1, Anfang) aus-

¹⁸⁾ Alexandroff, Über die Metrisation der im Kleinen kompakten topologischen Räume, Math. Annalen 92 (1924).

gezeichnet ist. Ist e^n ein in E^n enthaltenes Element, v^n eine e^n repräsentierende Vollkugel, so wird, sobald E^n orientiert ist, in e^n eine Orientierung durch die Forderung „induziert“, daß die durch die Koinzidenz in M^n vermittelte Abbildung von v^n auf einen Teil von V^n den Grad $+1$ habe (vgl. § 1). Umgekehrt induziert eine Orientierung von e^n eine solche von E^n .

Wir nennen nun die Mannigfaltigkeit M^n orientiert, wenn alle ihre Elemente so orientiert sind, daß die Orientierungen je zweier Elemente, von denen das eine das andere enthält, sich gegenseitig induzieren.

Wenn M^n „orientierbar“ ist, d. h. wenn M^n sich auf wenigstens eine Weise orientieren läßt, so gibt es genau zwei verschiedene Orientierungen. Daß es wenigstens zwei Orientierungen gibt, sobald es eine gibt, ist trivial, da man nur die Orientierungen aller Elemente umzukehren braucht, um (im Hinblick auf die Produktregel des § 1) eine neue Orientierung zu erhalten. Hat man andererseits in einem einzigen Elemente E^n eine der beiden möglichen Orientierungen ausgezeichnet, so wird dadurch zunächst in jedem in E^n enthaltenen Element e^n eine Orientierung induziert; diese überträgt sich wieder durch Induktion auf jedes e^n enthaltende Element E_1^n . Somit läßt sich, sobald E^n orientiert ist, jedes E_1^n , das mit E^n ein Gebiet gemeinsam hat, auf höchstens eine Weise vorschrittmäßig orientieren. Da man nun je zwei Punkte durch einen „Weg“ (s. o.) verbinden kann, lassen sich je zwei Elemente durch eine Kette von Elementen verbinden, in der jedes Element in das folgende übergreift. Mithin ist in jedem Element höchstens eine Orientierung möglich, wenn ein Element E^n orientiert ist.

Ist M^n orientierbar, so lassen sich insbesondere die Elemente E_1^n, E_2^n, \dots einer Basis so orientieren, daß jede zwischen Teilen der repräsentierenden V_i^n durch Vermittlung von M^n auftretende eineindeutige Abbildung den Grad $+1$ hat; denn ist z. B. e^n ein im Durchschnitt von E_1^n und E_2^n enthaltenes Element und sind V_1^n, V_2^n, v^n repräsentierende, orientierte Vollkugeln dieser Elemente, so haben die beiden Abbildungen von v^n auf V_1^n bzw. V_2^n den Grad $+1$, also ist dies nach der Produktregel auch der Grad der durch e^n vermittelten Abbildung zwischen einem Teil von V_1^n und einem Teil von V_2^n . Umgekehrt genügt es für die Orientierung von M^n bereits, daß die Elemente E_i^n einer bestimmten Basis derart orientiert seien, daß alle zwischen Teilen der repräsentierenden V_i^n vorkommenden Abbildungen den Grad $+1$ haben. Denn wenn wir mit ε^n diejenigen Elemente bezeichnen, die in mindestens einem E_i^n enthalten sind, so wird die Orientierung in jedem Element ε^n auf Grund der Produktregel *eindeutig* induziert; greifen $\varepsilon_1^n, \varepsilon_2^n$ übereinander, so induzieren sie in einem sie enthaltenden E_i^n *die-selbe* Orientierung, nämlich diejenige eines in ihrem Durchschnitt enthaltenen ε_3^n ; sind $\varepsilon_1^n, \varepsilon_2^n$ zwei beliebige in einem E_i^n enthaltene ε^n , so kann man

sie durch eine Kette von ε^n verbinden, in der jedes Element in das folgende hinübergreift; mithin induzieren alle diese ε_i^n , insbesondere ε_1^n und ε_2^n dieselbe Orientierung in E_i^n . Ein beliebiges Element läßt sich also *eindeutig* mittels Induktion durch die ε^n orientieren. Somit ist gezeigt, daß die Orientierbarkeit von M^n sich unter Zugrundelegung der Darstellung V_1^n, V_2^n, \dots einer Basis folgendermaßen charakterisieren läßt: M^n ist dann und nur dann orientierbar, wenn sich jede V_i^n im elementaren Sinn derart orientieren läßt, daß alle durch M^n vermittelten Abbildungen zwischen Teilen der verschiedenen V_i^n den Grad $+1$ haben. Dieser Satz liefert eine Möglichkeit, für eine vorgelegte M^n die Frage nach der Orientierbarkeit zu untersuchen; ferner ist aus ihm leicht ersichtlich, daß die hier gegebene Definition der Orientierbarkeit im Falle triangulierbarer Mannigfaltigkeiten mit der Brouwerschen Definition ¹⁾ a) übereinstimmt.

§ 4.

Abbildungen und Abbildungsklassen von Mannigfaltigkeiten.

Wir betrachten eindeutige und stetige Abbildungen einer M^n auf eine zweite n -dimensionale Mannigfaltigkeit μ^n , und zwar insbesondere solche Eigenschaften der Abbildungen, die bei deren stetigen Änderungen erhalten bleiben. Dabei verstehen wir unter einer für $0 \leq t \leq 1$ definierten stetigen Abänderung a_t einer Abbildung a_0 eine Schar von Abbildungen mit folgenden zwei Eigenschaften: 1) Für jeden Punkt $x \in M^n$ hängt $a_t(x)$ stetig von t ab; 2) diese Stetigkeit ist gleichmäßig in ganz M^n ; d. h. zu jedem $\delta > 0$ läßt sich ein $\tau > 0$ so angeben, daß aus $|t' - t| < \tau$ für alle Punkte $x \in M^n$ die Beziehung $\varrho(a_{t'}(x), a_t(x)) < \delta$ folgt, wobei ϱ die Entfernungsfunktion einer in μ^n fest eingeführten Metrik ist (vgl. § 3).

(Wenn M^n geschlossen ist, so ist, wie leicht zu sehen, 2) eine Folge von 1). Dasselbe gilt z. B. auch, wenn M^n ein beschränktes Gebiet im R^n ist und die Abbildungen g_t in der abgeschlossenen Hülle von M^n definiert und stetig sind und 1) erfüllen, also insbesondere in dem in § 1 behandelten Fall der Abbildungen eines Elements.)

Wir werden die Abbildungen zunächst in den Umgebungen einzelner Punkte von μ^n untersuchen und beschränken uns dabei auf solche Punkte, in denen die Abbildungen „kompakt“ sind. Hierunter verstehen wir folgendes: *a heißt kompakt in dem Punkte ξ , wenn es eine Umgebung von ξ gibt, deren durch a auf sie abgebildete Originalmenge in M^n kompakt ist.* Für eine geschlossene M^n bedeutet dies keine Einschränkung.

(Die in § 1 behandelten Abbildungen eines Elements sind, als Abbildungen des offenen Innengebiets aufgefaßt, nicht kompakt in den Bildpunkten des Randes, die wir ja auch von der Betrachtung ausschlossen.)

Unter einer „Abbildungsklasse“ \mathfrak{K} verstehen wir eine Menge von Abbildungen mit folgenden Eigenschaften: 1) Je zwei Abbildungen aus \mathfrak{K} lassen sich (in dem oben präzisierten Sinn) stetig ineinander überführen; 2) die Abbildungen aus \mathfrak{K} sind sämtlich in den gleichen Punkten von μ^n kompakt bzw. nicht kompakt. 3) Es gibt keine Menge $\overline{\mathfrak{K}}$, die \mathfrak{K} als echten Teil enthält und auch 1) und 2) erfüllt. — μ^n wird also durch die Punkte, in denen \mathfrak{K} nicht kompakt ist, in eine (endliche oder abzählbar unendliche) Anzahl von Gebieten zerlegt.

(Diese Definition der „Klasse“, angewandt auf den im § 1 behandelten Spezialfall, ist etwas allgemeiner als die dort aufgestellte.)

\mathfrak{K} sei nunmehr eine bestimmte Klasse, ξ ein fester Punkt, in dem \mathfrak{K} kompakt ist. Wir stufen innerhalb \mathfrak{K} die Abbildungen nach der Regularität ihres Verhaltens im Punkte ξ ab:

1. Eine Abbildung $h \in \mathfrak{K}$ heißt „*halbglatt*“ in ξ , wenn die Originalmenge von ξ sich in die Innengebiete von endlich vielen, zueinander fremden Elementen einschließen läßt, deren Bilder einer euklidischen Umgebung von ξ angehören.

2. Eine Abbildung $f \in \mathfrak{K}$ heißt „*fastglatt*“ in ξ , wenn die Originalmenge von ξ endlich ist.

3. Eine Abbildung $g \in \mathfrak{K}$ heißt „*glatt*“ in ξ , wenn es l zueinander fremde Gebiete ($l \geq 0$) von M^n gibt, so daß g in jedem von ihnen eineindeutig ist und daß ξ nicht zu dem Bild der Komplementärmenge dieser Gebiete gehört (vgl. die Definition der „glatten“ Abbildungen in § 1).

Bezeichnen wir die Mengen der in ξ halbglatte, fastglatten und glatten Abbildungen mit \mathfrak{H}_ξ , \mathfrak{F}_ξ bzw. \mathfrak{G}_ξ , so ist $\mathfrak{G}_\xi \subset \mathfrak{F}_\xi \subset \mathfrak{H}_\xi \subset \mathfrak{K}$. Das Ziel dieses Paragraphen ist eine genauere Kenntnis der Verteilung von \mathfrak{G}_ξ , \mathfrak{F}_ξ und \mathfrak{H}_ξ innerhalb \mathfrak{K} .

Zunächst zeigen wir, daß \mathfrak{F}_ξ *überall dicht* in \mathfrak{K} ist, d. h. wir beweisen

Satz I. *Ist a eine Abbildung aus \mathfrak{K} und δ irgendeine positive Zahl, so gibt es eine Abbildung $f \in \mathfrak{F}_\xi$ derart, daß $\rho(a(x), f(x)) < \delta$ für alle Punkte $x \in M^n$ ist.*

Beweis. Φ_ξ sei eine mit einem euklidischen Koordinatensystem versehene so kleine Umgebung von ξ , daß die Entfernung je zweier ihrer Punkte kleiner als δ ist. φ_ξ sei eine in Φ_ξ enthaltene Umgebung von ξ , deren Originalmenge u , die durch a auf φ_ξ abgebildet wird, in M^n kompakt ist. u läßt sich daher in endlich viele Elemente $E_1^n, E_2^n, \dots, E_a^n$ derart einschließen, daß jeder ihrer Punkte im Inneren von wenigstens einem E_a^n liegt, und zwar kann man die Elemente so klein wählen, daß $a(E_a^n) \subset \Phi_\xi$ ist.

Wir setzen nun $f \equiv a$ in der Menge $M^n - \sum_{a=1}^a E_a^n$ und auf dem Rande

dieser Menge. Dieser Rand ist zugleich der Rand von $\sum E_a^n$. Wenn wir f nun noch im Innern von $\sum E_a^n$ so erklären, daß $f(\sum E_a^n) \subset \Phi_\xi$ ist, und daß nur endlich vielen Punkten das Bild ξ zugeordnet wird, so hat die so definierte Abbildung $f(M^n)$ alle geforderten Eigenschaften: sie ist fast-glatt in ξ , da die Gleichung $f(x) = \xi$ in $\sum E_a^n$ endlich viele, in $M^n - \sum E_a^n$ keine Lösungen hat; sie gehört zu \mathfrak{R} , da man a in f dadurch stetig übergehen lassen kann, daß man $a(x)$ festhält, falls x nicht innerer Punkt von $\sum E_a^n$ ist, und für $x \in \sum E_a^n$ den Punkt $a(x)$ im Zeitintervall 1 gleichförmig und geradlinig im Sinne der euklidischen Geometrie von Φ_ξ in den Punkt $f(x)$ laufen läßt, was infolge der Kompaktheit von $\sum E_a^n$ eine gleichmäßige stetige Änderung ist; es ist $\rho(f(x), a(x)) < \delta$ für $x \in \sum E_a^n$, da $f(x) \in \Phi_\xi$, $a(x) \in \Phi_\xi$ ist, und $\rho(f(x), a(x)) = 0$ für alle anderen Punkte.

Wir haben also nur die Abbildung f des Randes \mathfrak{R} von $\sum E_a^n$ auf eine ξ nicht enthaltende Menge des euklidischen Raumes Φ_ξ zu einer Abbildung $f(\sum E_a^n)$ zu erweitern, die höchstens endlich viele ξ -Stellen besitzt. Dies tun wir mittels wiederholter Anwendung des Hilfssatzes II a aus § 2. Nach diesem Hilfssatz können wir f zunächst gewiß in E_1^n so erklären, daß auf dem Rand von E_1^n , der ja in E_1^n nirgends dicht ist, keine ξ -Stelle liegt. Wir definieren nun f durch Schluß von i auf $i+1$.

Sei f in $\sum_{a=1}^i E_a^n$ so erklärt, daß der Rand dieser Menge frei von ξ -Stellen ist; betrachten wir nun E_{i+1}^n : in $E_{i+1}^n \cdot \left(\sum_{a=1}^i E_a^n + \mathfrak{R} \right) = Q$ ist f schon erklärt; die gemeinsame Grenze q dieser Menge mit ihrer Komplementärmenge in E_{i+1}^n gehört teils zu dem Rand von $\sum_{a=1}^i E_a^n$, teils zu \mathfrak{R} , ist also jedenfalls frei von ξ -Stellen; f läßt sich daher auf E_{i+1}^n ausdehnen, und zwar so, daß keinem Randpunkt von E_{i+1}^n ξ als Bild zugeordnet wird. Somit ist f jetzt auch in $\sum_{a=1}^{i+1} E_a^n$ erklärt und hat auf dem Rande dieser Menge keine ξ -Stelle. Mithin ist f überall in der gewünschten Weise zu definieren, und Satz I ist bewiesen.

Aus dem Beweis ist ersichtlich, daß man Satz I folgendermaßen verschärfen kann:

Satz Ia. Die in Satz I genannte Abbildung f läßt sich sogar so bestimmen, daß sie außerhalb eines beliebig vorgeschriebenen, die Originalmenge von ξ enthaltenden, offenen Teils von M^n mit a identisch ist.

Ist $f \in \mathfrak{F}_\xi$, so lassen sich die Originalpunkte x_1, x_2, \dots, x_a von ξ in zueinander fremde Elemente $E_1^n, E_2^n, \dots, E_a^n$ einschließen, deren Bilder in einer beliebig kleinen euklidischen Umgebung Φ_ξ von ξ enthalten sind.

Nach § 1, Satz VI a, kann man um f in jedem Element E_α^n , ohne daß die Bildpunkte Φ_ξ verlassen, so abändern, daß die Abbildung in ξ glatt wird; also ist auch \mathfrak{G}_ξ überall dicht in \mathfrak{R} , d. h. es gilt

Satz II. Jede Abbildung $a \in \mathfrak{R}$ läßt sich durch eine beliebig kleine stetige Abänderung in ξ „glätten“, d. h. in eine Abbildung $g \in \mathfrak{G}_\xi$ überführen.

Da nach Satz I a diese Abänderung die Punkte einer beliebigen, zu der Originalmenge von ξ fremden, abgeschlossenen Punktmenge fest läßt, kann man derartige Abänderungen nacheinander für mehrere Punkte $\xi_1, \xi_2, \dots, \xi_b$ vornehmen; es gilt also

Satz II a. Ist \mathfrak{R} in den Punkten $\xi_1, \xi_2, \dots, \xi_b$ kompakt, so läßt sich die Abbildung $a \in \mathfrak{R}$ durch eine beliebig kleine stetige Abänderung in den Punkten $\xi_1, \xi_2, \dots, \xi_b$ glätten.

Wir werden jetzt zeigen, daß die — nach den vorstehenden Sätzen a fortiori in \mathfrak{R} überall dichte — Menge \mathfrak{S}_ξ zusammenhängend ist; wir werden nämlich folgenden Satz beweisen:

Satz III. Sind h_0, h_1 Abbildungen aus \mathfrak{S}_ξ , so läßt sich der, infolge der Zugehörigkeit von h_0 und h_1 zu \mathfrak{R} mögliche, stetige Überführungsprozeß a_t ($0 \leq t \leq 1$; $a_0 = h_0, a_1 = h_1$) durch einen solchen ersetzen, in dem die die Überführung bewirkenden Abbildungen h_t ($0 \leq t \leq 1$) sämtlich zu \mathfrak{S}_ξ gehören.

Beweis. Satz III läßt sich auf den Spezialfall III* zurückführen, in dem die folgenden beiden zusätzlichen Voraussetzungen erfüllt sind: 1) h_0^* und h_1^* sind nicht nur halb-, sondern sogar fast-glatt¹⁹⁾; 2) es gibt eine h_0^* in h_1^* stetig überführende Abbildungsschar $a_t \in \mathfrak{R}$ ($t' \leq t \leq t''$; $h_0^* = a_{t'}, h_1^* = a_{t''}$), in der $\rho(h_0^*(x), a_t(x)) < \Delta$ für alle $x \in M^n$ und $t' \leq t \leq t''$ ist, wobei Δ eine nur von \mathfrak{R} und ξ , aber nicht von h_0^* und h_1^* abhängende Konstante ist.

Wenn nämlich III* richtig ist, so läßt sich der allgemeine Fall III folgendermaßen auf ihn zurückführen:

h_0 werde in h_1 durch die Schar $a_t \in \mathfrak{R}$ ($0 \leq t \leq 1$; $h_0 = a_0, h_1 = a_1$) übergeführt. c sei eine so große natürliche Zahl, daß $\rho(a_{t_1}(x), a_{t_2}(x)) < \frac{1}{2} \Delta$ für $|t_1 - t_2| \leq \frac{1}{c}$, $x \in M^n$ ist. f_i ($i = 0, 1, \dots, c$) sei irgendeine fast-glatte Abbildung, die sich so wenig von $a_{\frac{i}{c}}$ unterscheidet, daß für alle Punkte x $\rho(f_i(x), a_{\frac{i}{c}}(x)) < \frac{1}{4} \Delta$ ist; die Existenz von f_i ist durch Satz I gesichert. Dann ist $\rho(f_i(x), f_{i+1}(x)) < \Delta$ ($i = 0, 1, \dots, c-1$); nach

¹⁹⁾ h_i^* spielt hier die Rolle von h_i in III.

III* läßt sich daher $\overset{c}{f}_i$ in $\overset{c}{f}_{i+1}$ und mithin f_0 in f_1 durch eine Schar halbglatter Abbildungen überführen, und wir brauchen uns nur noch davon zu überzeugen, daß sich h_0 und h_1 mittels halbglatter Abbildungen in f_0 bzw. f_1 deformieren lassen. Nun dürfen wir aber auf Grund von Satz Ia annehmen, daß h_0 von f_0 nur in einer beliebigen Nähe der Originalmenge von ξ verschieden ist. Diese Menge ist infolge der Halbglattheit von h_0 in endlich vielen, zueinander fremden Elementen enthalten; die Abänderung von h_0 in f_0 geht ganz im Innern dieser Elemente vor sich und ist beliebig klein; sie zerstört daher die Eigenschaft der Halbglattheit in keinem Augenblick. Ebenso verhält es sich mit h_1 und f_1 .

Es genügt also in der Tat, den Spezialfall III* zu beweisen, und wir tun dies, unter Beibehaltung der Bezeichnungen von Satz III, indem wir die Konstante Δ folgendermaßen bestimmen: φ_ξ ist eine euklidische Umgebung von ξ , in deren sämtlichen Punkten \mathfrak{R} kompakt ist; dann sei

$$(1) \quad 0 < \Delta < \frac{1}{3} \varrho(\xi, \mu^n - \varphi_\xi).$$

Es sei nun A die (abgeschlossene) Menge derjenigen Punkte von M^n , für die

$$(2) \quad \varrho(\xi, h_0(x)) \leq \Delta$$

ist; da mithin nach (1) $h_0(A) \subset \varphi_\xi$ ist, ist A kompakt und läßt sich daher ins Innere von endlich vielen Elementen E_β^n ($\beta = 1, 2, \dots, b$) einschließen. Diese seien so klein, daß für zwei Punkte x_β, x'_β desselben Elements E_β^n stets

$$(3) \quad \varrho(h_0(x_\beta), h_0(x'_\beta)) < \Delta$$

ist; ferner lassen sie sich so wählen, daß keine der endlich vielen ξ -Stellen der fastglatten Abbildungen h_0 und h_1 auf dem Rande eines von ihnen liegt, daß also, wenn $\mathfrak{R} E_\beta^n$ den Rand von E_β bezeichnet,

$$(4) \quad \xi \notin h_i(\mathfrak{R} E_\beta^n) \quad (i=0, 1; \beta=1, 2, \dots, b)$$

ist. — Da nach Voraussetzung

$$(5) \quad \varrho(h_0(x), a_t(x)) < \Delta \quad (0 \leq t \leq 1; x \in M^n)$$

ist, so folgt, wenn wir den oben benutzten Punkt x'_β als im Durchschnitt von E_β^n und A gelegen annehmen, aus (2), (3), (5)

$$\varrho(\xi, a_t(x_\beta)) < 3\Delta,$$

also aus (1)

$$(6) \quad a_t(E_\beta^n) \subset \varphi_\xi \quad (0 \leq t \leq 1; \beta=1, 2, \dots, b).$$

Sei $B = \sum_{\beta=1}^b E_\beta^n$, $\mathfrak{R} B$ der Rand von B . Ist dann $y \in M^n - B + \mathfrak{R} B$, so ist $y \notin A$, also nach (2)

$$\varrho(\xi, h_0(y)) > \Delta;$$

hieraus und aus (5) ergibt sich

$$(7) \quad a_t(y) \neq \xi \quad (0 \leq t \leq 1; y \in M^n - B + \mathfrak{R} B).$$

Wir definieren die gesuchte Abbildungsschar h_t nun zunächst in $M^n - B + \mathfrak{R} B$ durch die Festsetzung $h_t = a_t$. ξ -Stellen treten hier infolge (7) nicht auf. Insbesondere sind die Abbildungen h_t also auf $\mathfrak{R}(B)$ definiert, und zwar infolge (6) so, daß

$$h_0(B) \subset \varphi_\xi, \quad h_1(B) \subset \varphi_\xi, \quad h_t(\mathfrak{R} B) \subset \varphi_\xi \quad (0 \leq t \leq 1)$$

ist; da φ_ξ ein euklidisches Koordinatensystem besitzt, ermöglicht dies die Anwendung der Hilfssätze aus § 2 auf die Elemente $E_1^n, E_2^n, \dots, E_b^n$.

Diese seien durch Kuben k_β^n ($\beta = 1, 2, \dots, b$) repräsentiert; k_i^{n-1} seien deren Randkuben ($i = 1, 2, \dots, 2nb$). $q_{i,j}$ sei die gemeinsame Grenze des Durchschnitts $k_i^{n-1} \cdot k_j^{n-1}$ — („Durchschnitt“ im Sinne der Koinzidenz in M^n) — und der in k_i^{n-1} komplementären Menge $k_i^{n-1} - k_i^{n-1} \cdot k_j^{n-1}$; $q_i = \sum_{j \neq i} q_{i,j}$ ist nirgends dicht in k_i^{n-1} und $q = \sum_i q_i$ nirgends dicht in $\sum_i k_i^{n-1}$.

Wir erklären die Schar h_t nun zuerst in k_1^{n-1} so, daß sie $h_0(k_1^{n-1})$ in $h_1(k_1^{n-1})$ überführt, daß $h_t(k_1^{n-1} \cdot \mathfrak{R} B) = a_t(k_1^{n-1} \cdot \mathfrak{R} B)$ ist, daß im ganzen nur endlich oft endlich viele ξ -Stellen auftreten und daß q frei von ξ -Stellen bleibt; das alles kann man auf Grund von Satz IIa*, dessen Voraussetzungen infolge (4) und (7) erfüllt sind, und des „Zusatzes 1“ zu Satz IIa* erreichen. Sei jetzt die Schar h_t in $\sum_{j=1}^i k_j^{n-1}$ so definiert, daß sie die gewünschte Überführung leistet, daß im ganzen nur endlich viele ξ -Stellen auftreten und daß q frei von ξ -Stellen bleibt. Dann sind für k_{i+1}^{n-1} alle Voraussetzungen des Satzes IIa* erfüllt; insbesondere ist die Grenze des Teils von k_{i+1}^{n-1} , in der h_t für $0 < t < 1$ schon festgelegt ist, frei von ξ -Stellen, weil er zu q gehört; man kann also die h_t auch in $\sum_{j=1}^{i+1} k_j$ so definieren, daß es im ganzen nur endlich viele ξ -Stellen gibt und daß diese nicht auf q liegen.

Somit ist eine derartige Definition von h_t in $\sum_{\beta=1}^b \mathfrak{R} k_\beta^n$ möglich, und wir haben nun die Innengebiete der k_β^n zu behandeln. Auf Grund von IIIb* tun wir dies der Reihe nach für $k_1^n, k_2^n, \dots, k_b^n$. Immer läßt sich die Schar $h_t(k_\beta^n)$ so erklären, daß für jedes einzelne t höchstens endlich viele Häufungspunkte von ξ -Stellen vorhanden sind; denn immer sind die Voraussetzungen von IIIb* erfüllt, da die gemeinsame Grenze von $k_\beta^n \cdot \sum_{\lambda=1}^{\beta-1} k_\lambda^n = Q$ und $k_\beta^n - k_\beta^n \cdot \sum_{\lambda=1}^{\beta-1} k_\lambda^n$ zu $\sum_{\beta=1}^b \mathfrak{R} k_\beta^n$ gehört, da also dort im ganzen Intervall

$0 \leq t \leq 1$ nur endlich viele ξ -Stellen auftreten, und da $h_0(k_\beta^n)$ und $h_1(k_\beta^n)$ auch nur endlich viele ξ -Stellen besitzen.

Mithin hat jede Abbildung h_t der Schar, die nunmehr für $0 \leq t \leq 1$ in ganz M^n erklärt ist, die Eigenschaft, daß die Menge der ξ -Stellen höchstens endlich viele Häufungspunkte x_1, x_2, \dots, x_a hat. Eine solche Abbildung ist aber gewiß halbglatte; denn die Punkte x_α ($\alpha = 1, 2, \dots, a$) lassen sich in zueinander fremde Elemente e_α^n einschließen, die auf ihren Rändern keine ξ -Stellen enthalten und die so klein sind, daß die Bilder $h_t(e_\alpha^n) \subset \varphi_\xi$ sind. Außerhalb der e_α^n liegen nur endlich viele, isolierte ξ -Stellen; jede von ihnen läßt sich in analoger Weise in ein kleines Element einschließen. h_t besitzt also die Eigenschaft der „Halbglattheit“; damit ist Satz III bewiesen.

§ 5.

Der Abbildungsgrad.

M^n sei orientiert; im übrigen bleibt der Sinn aller Bezeichnungen aus dem vorigen Paragraphen ungeändert.

Ist h eine in ξ halbglatte Abbildung, so lassen sich die Originalpunkte von ξ in zueinander fremde Elemente $e_1^n, e_2^n, \dots, e_a^n$ einschließen, deren Bilder $h(e_\alpha^n)$ ($\alpha = 1, 2, \dots, a$) in einer Umgebung φ_ξ von ξ liegen, die mit einem euklidischen Koordinatensystem versehen ist. In diesem zeichnen wir willkürlich eine Orientierung als die positive aus. Dann besitzt jede der Abbildungen $h(e_\alpha^n)$ nach § 1 einen wohlbestimmten Grad $\gamma_\xi(h(e_\alpha^n))$; wir behaupten, daß die Summe $\sum_{\alpha=1}^a \gamma_\xi(h(e_\alpha^n))$ unabhängig von der Wahl der Elemente e_α^n , d. h. daß

$$(1) \quad \sum_{\alpha=1}^a \gamma_\xi(h(e_\alpha^n)) = \sum_{\beta=1}^b \gamma_\xi(h(E_\beta^n))$$

ist, falls $E_1^n, E_2^n, \dots, E_b^n$ ebenfalls zueinander fremde Elemente sind, die die Originalpunkte von ξ im Innern enthalten und deren Bilder $h(E_\beta^n)$ in einer euklidischen Umgebung Φ_ξ von ξ liegen; dabei sei in Φ_ξ die positive Orientierung durch φ_ξ induziert (siehe § 3).

Die Behauptung ist gewiß in dem Spezialfall richtig, in dem h nicht nur halb-, sondern sogar fastglatte ist; denn dann lassen sich die Originalpunkte x_1, x_2, \dots, x_d von ξ einzeln in kleine, zueinander fremde Elemente e_δ^n ($\delta = 1, 2, \dots, d$) einschließen, die ganz in je einem e_α^n und einem E_β^n enthalten sind, und deren Bilder $h(e_\delta^n)$ in dem Durchschnitt von φ_ξ und Φ_ξ liegen, und nach § 1, Satz VIII, ist

$$\sum_{\alpha=1}^a \gamma_\xi(h(e_\alpha^n)) = \sum_{\delta=1}^d \gamma_\xi(h(e_\delta^n)) = \sum_{\beta=1}^b \gamma_\xi(h(E_\beta^n)),$$

wobei es wegen der topologischen Invarianz des Grades und unserer Festsetzung der Orientierung gleichgültig ist, welches der Koordinatensysteme von φ_ξ oder Φ_ξ wir bei der Bestimmung von $\gamma_\xi(h(\epsilon_\delta^n))$ im Auge haben.

Die Richtigkeit von (1) für eine beliebige halbglatte Abbildung h ergibt sich aus diesem Spezialfall nun dadurch, daß man — auf Grund von § 4, Satz Ia — h durch eine kleine Abänderung, welche die Ränder aller e_α^n und E_β^n sowie alle Punkte von $M^n - (\sum e_\alpha^n + \sum E_\beta^n)$ festhält, in eine fastglatte Abbildung überführt; hierbei bleibt nach § 1 für jedes einzelne Element e_α^n und E_β^n der Grad konstant.

$\sum_{\alpha=1}^a \gamma_\xi(h(e_\alpha^n))$ ist also unabhängig von der Wahl der e_α^n und mithin eine Funktion $\gamma_\xi(h)$ von h . Wir nennen sie den *Grad* von h .

Sind nun h_0 und h_1 zwei halbglatte Abbildungen aus \mathfrak{R} , so ist

$$(2) \quad \gamma_\xi(h_0) = \gamma_\xi(h_1).$$

Beweis. Nach § 4, Satz III, gibt es für $0 \leq t \leq 1$ eine stetige Schar halbglatte Abbildungen h_t . Aus der Definition von $\gamma_\xi(h_t)$ und aus § 1, Satz VII a, folgt, daß sich jeder Wert t^* in ein t -Intervall einschließen läßt, innerhalb dessen $\gamma_\xi(h_t)$ konstant ist. Mithin ist sowohl die Menge derjenigen t , für die $\gamma_\xi(h_t) = \gamma_\xi(h_0)$ ist, als ihre Komplementärmenge offen. Folglich ist letztere Menge leer, d. h. es ist $\gamma_\xi(h_t) = \gamma_\xi(h_0)$ für $0 \leq t \leq 1$, also gilt insbesondere Gleichung (2) und mithin

Satz I. *Die Grade aller halbglatte Abbildungen aus \mathfrak{R} sind einander gleich. Ihr Wert ist eine Invariante von \mathfrak{R} und ξ . Wir nennen ihn den Grad $\gamma_\xi(\mathfrak{R})$ der Klasse \mathfrak{R} im Punkte ξ .*

Nach § 4, Satz II, gibt es in \mathfrak{R} Abbildungen, die im Punkt ξ glatt sind; für sie folgt aus § 1:

Satz Ia. *Ist g eine in ξ glatte Abbildung aus \mathfrak{R} , und bezeichnen p und q die Anzahlen der durch g gelieferten positiven bzw. negativen Bedeckungen einer Umgebung des Punktes ξ , so ist $p - q = \gamma_\xi(\mathfrak{R})$.*

Daher ist z. B., falls eine Umgebung von ξ bei einer Abbildung aus \mathfrak{R} eineindeutig bedeckt wird, $\gamma_\xi(\mathfrak{R}) = \pm 1$, und, falls ξ gar nicht bedeckt wird, $\gamma_\xi(\mathfrak{R}) = 0$.

Aus den Sätzen des § 1 ist ersichtlich, daß $\gamma_\xi(\mathfrak{R})$ eine topologische Invariante ist, sobald die Elemente von M^n einerseits, eine Umgebung von ξ andererseits „relativ orientiert“ sind; dies ist, da wir M^n als orientiert voraussetzen, der Fall, wenn auch μ^n orientiert ist. Ist aber μ^n nicht orientiert, so ist das Vorzeichen von $\gamma_\xi(\mathfrak{R})$ von der Orientierung der Umgebung von ξ abhängig. Der absolute Betrag von $\gamma_\xi(\mathfrak{R})$ ist in jedem Fall eindeutig bestimmt.

Aus $|\gamma_{\xi}(\mathfrak{R})| = |p - q|$ und aus der Bedeutung der Zahlen p und q ist ersichtlich, daß der absolute Betrag der Grades in allen Punkten einer Umgebung von ξ denselben Wert hat. Er bleibt daher überhaupt konstant, wenn ξ in einem Gebiet Γ variabel ist, in dessen Punkten \mathfrak{R} kompakt, der Grad also erklärt ist. Ist dieses Gebiet, das sich ja als offene Mannigfaltigkeit auffassen läßt (siehe § 3), orientierbar, so ist nach Festsetzung einer positiven Orientierung von Γ auch das Vorzeichen des Grades in Γ eindeutig festgelegt. Ist andererseits \mathfrak{R} in einem Gebiet Γ , über dessen Orientierbarkeit wir nichts voraussetzen, überall kompakt, und ist $|\gamma_{\xi}(\mathfrak{R})| \neq 0$ für $\xi \in \Gamma$, so läßt sich jedes Element von Γ so orientieren, daß in ihm $\gamma_{\xi}(\mathfrak{R}) > 0$ ist; sind E^n und ε^n Elemente von Γ und ist $\varepsilon^n \subset E^n$, so haben die so in ε^n und E^n festgelegten Orientierungen die Eigenschaft, sich gegenseitig zu „induzieren“ (siehe § 3), da eine positive Bedeckung eines Teiles von ε^n auch als positiv zu gelten hat, wenn man ε^n als Teil von E^n auffaßt. Durch unsere Festsetzung $\gamma_{\xi}(\mathfrak{R}) > 0$ ist also Γ orientiert worden. — Es liegt mithin für orientierte M^n folgender Sachverhalt vor:

Satz II. *Die (offene) Menge derjenigen Punkte von μ^n , in denen \mathfrak{R} kompakt ist, wird durch die übrigen Punkte in eine (endliche oder abzählbar unendliche) Anzahl von (zusammenhängenden) Gebieten $\Gamma_1, \Gamma_2, \dots$ zerlegt. In jedem dieser Gebiete Γ_i ist der Betrag des Grades konstant: $|\gamma_{\Gamma_i}(\mathfrak{R})| = |\gamma_{\xi}(\mathfrak{R})|$ ($\xi \in \Gamma_i$); ist Γ_i orientierbar und ist eine Orientierung als positiv ausgezeichnet, — was insbesondere der Fall ist, wenn nicht nur M^n , sondern auch μ^n orientiert ist, — so ist auch das Vorzeichen des Grades $\gamma_{\Gamma_i}(\mathfrak{R})$ bestimmt.*

Satz IIa. *Ist Γ_i nicht orientierbar, so ist $\gamma_{\Gamma_i}(\mathfrak{R}) = 0$.*

Satz IIb. *Es sei \mathfrak{R} überall in μ^n kompakt, (also z. B. M^n geschlossen); ist μ^n orientiert, so ist der Grad konstant: $\gamma_{\xi}(\mathfrak{R}) = \gamma(\mathfrak{R})$ ($\xi \in \mu^n$); ist μ^n nicht orientierbar, so ist $\gamma_{\xi}(\mathfrak{R}) = 0$ ($\xi \in \mu^n$).*

Satz IIc. *Ist M^n geschlossen, μ^n offen, so ist $\gamma(\mathfrak{R}) = 0$;*

denn dann ist die kompakte Bildmenge $k(M^n)$ ($k \subset \mathfrak{R}$) ein echter Teil von μ^n , und es gibt Punkte ξ , die nicht zu $k(M^n)$ gehören, in denen also $\gamma_{\xi}(k) = 0$ ist.

Satz III. *Es sei μ^n einer Abbildung φ auf eine dritte Mannigfaltigkeit \mathfrak{M}^n unterworfen; χ sei ein Punkt von \mathfrak{M}^n , in dem φ kompakt ist. In der Originalmenge Ξ von χ sei \mathfrak{R} kompakt und $\gamma_{\xi}(\mathfrak{R})$ sei konstant für $\xi \in \Xi$. Dann gilt die „Produktregel“*

$$\gamma_{\chi}(\varphi k) = \gamma_{\Xi}(k) \cdot \gamma_{\xi}(\varphi) \quad (k \subset \mathfrak{R}).$$

Zum Beweise „glätte“ (siehe § 4) man φ im Punkte ζ ; bei der so entstandenen, in ζ glatten Abbildung φ^* seien $\xi_1, \xi_2, \dots, \xi_a, \xi_{a+1}, \dots, \xi_{a+b}$ die Originalpunkte von ζ , und zwar werde die Umgebung von ζ durch Umgebungen von $\xi_1, \xi_2, \dots, \xi_a$ positiv, durch Umgebungen von $\xi_{a+1}, \dots, \xi_{a+b}$ negativ bedeckt. Ferner sei k eine in den Punkten $\xi_1, \xi_2, \dots, \xi_{a+b}$ glatte Abbildung aus \mathfrak{R} (§ 4, Satz IIa); bei ihr seien p_i, q_i die positiven bzw. negativen Bedeckungen von ξ_i ($i = 1, 2, \dots, a + b$). Bei der in ζ glatten Abbildung $\varphi^*k(M^n)$ ist dann die Anzahl der positiven Bedeckungen von ζ :

$$\pi = \sum_{\alpha=1}^a p_\alpha + \sum_{\beta=1}^b q_{a+\beta}, \text{ die der negativen Bedeckungen: } \kappa = \sum_{\alpha=1}^a q_\alpha + \sum_{\beta=1}^b p_{a+\beta},$$

folglich ist

$$\begin{aligned} \gamma_\zeta(\varphi^*k) &= \pi - \kappa = \sum_{\alpha=1}^a (p_\alpha - q_\alpha) - \sum_{\beta=1}^b (p_{a+\beta} - q_{a+\beta}) \\ &= \sum_{\alpha=1}^a \gamma_{\xi_\alpha}(\mathfrak{R}) - \sum_{\beta=1}^b \gamma_{\xi_{a+\beta}}(\mathfrak{R}) = (a - b) \cdot \gamma_\zeta(\mathfrak{R}) \\ &= \gamma_\zeta(\varphi) \cdot \gamma_\zeta(\mathfrak{R}). \end{aligned}$$

Der Begriff der „Klasse“ \mathfrak{R} wurde am Anfang des vorigen Paragraphen durch drei Eigenschaften definiert, von denen die zweite in der Identität der Mengen der Punkte bestand, in denen die Abbildungen kompakt sind. Mitunter ist es zweckmäßig, diese Eigenschaft 2) durch die schwächere zu ersetzen, daß die Abbildungen der betrachteten Menge sämtlich in ξ kompakt sind.

Eine solche Menge \mathfrak{R}_ξ heiße eine „Abbildungsklasse in bezug auf ξ .“ Es gilt

Satz IV. Alle Abbildungen aus \mathfrak{R}_ξ haben in ξ den gleichen Grad, oder mit anderen Worten:

Bei stetiger Abänderung einer im Punkte ξ kompakten Abbildung F bleibt der Grad in ξ konstant, wenn nur die Abbildung immer kompakt in ξ bleibt.

Beweis. Hängt die Abbildungsschar F_t für $0 \leq t \leq 1$ in dem am Anfang von § 4 präzisierten Sinne stetig von t ab und sind alle Abbildungen F_t in ξ kompakt, so sind sie in sämtlichen Punkten einer Umgebung φ_ξ von ξ kompakt.

Denn andernfalls gäbe es eine gegen einen Wert t^* konvergierende Folge t_i und eine derartige Folge gegen ξ konvergierender Punkte ξ_i ($i = 1, 2, \dots$), daß F_{t_i} in ξ_i nicht kompakt wäre. Ist aber Φ_ξ eine Umgebung von ξ , in der F_{t^*} kompakt ist, und ε^n ein Element von Φ_ξ , das ξ im Innern enthält, so kann man eine Zahl $\tau > 0$ so angeben, daß aus $|t - t^*| < \tau$, $F_{t^*}(y) \not\subset \Phi_\xi$ stets $F_t(y) \not\subset \varepsilon^n$ folgt. Die Originalmenge

von ε^n bei F_i ist also in der Originalmenge von Φ_ξ bei F_i^* enthalten, mithin kompakt. Im Widerspruch hierzu müßte $\xi_i < \varepsilon^n$ für hinreichend großes i sein.

Alle F_i sind also in einer Umgebung φ_ξ kompakt. H_0 und H_1 seien Abbildungen, die in ξ halbglatt sind, mit F_0 bzw. F_1 zu denselben Klassen gehören, und sich von F_0 bzw. F_1 nur in so kleinen Umgebungen der Originalmengen des Punktes ξ unterscheiden, daß die Originalmengen von φ_ξ ungeändert, also kompakt bleiben. Dann bleibt der Beweis von § 4, Satz III, gültig als Beweis der Tatsache, daß man H_0 in H_1 durch eine Schar von halbglaten Abbildungen H_i überführen kann, die sämtlich in φ_ξ kompakt sind. Hieraus folgt, genau wie beim Beweise von § 5, Satz I, daß $\gamma_\xi(H_0) = \gamma_\xi(H_1)$, also $\gamma_\xi(F_0) = \gamma_\xi(F_1)$ ist.

Wird M^n nicht mehr, wie bisher, als orientiert vorausgesetzt, so sind die am Anfang dieses Paragraphen betrachteten Elemente e_α^n und E_β^n nicht orientiert. Jedoch sind die in (1) auftretenden Grade γ_ξ ihrem absoluten Betrage nach, die Summen in (1) also modulo 2 bestimmt. Die an (1) anschließenden Betrachtungen behalten ihre Gültigkeit für die betreffende Restklasse modulo 2; wir bezeichnen diese als die „Parität“²⁰⁾ der Abbildung. Es ergibt sich ohne weiteres

Satz IV. *Ist M^n nicht orientiert, so bleiben, wenn man „Grad“ durch „Parität“ ersetzt, die Sätze dieses Paragraphen mit folgenden Abänderungen bestehen: In Satz Ia tritt an Stelle von $p - q$ die modulo 2 reduzierte Anzahl der glatten Bedeckungen. Satz IIa und die Teile der Sätze II und IIb, in denen von der Orientierbarkeit von Γ_i oder μ^n die Rede ist, verlieren ihren Sinn.*

²⁰⁾ Brouwer, Aufzählung der Abbildungsklassen endlichfach zusammenhängender Flächen, Math. Annalen 82 (1921), S. 283, Fußnote.

(Eingegangen am 4. 8. 1927.)

Glättung von Flächenabbildungen.

Von

Hellmuth Kneser in Greifswald.

Bekanntlich¹⁾ kommt jeder stetigen Abbildung \mathfrak{A} einer zusammenhängenden orientierten geschlossenen Fläche F auf eine andere, G , in jedem Punkte von G eine bestimmte ganze Zahl n , ihr Grad, zu, und dieser Grad ist derselbe in jedem Punkt von G und bei jeder Abbildung, die sich aus \mathfrak{A} durch stetige Deformation ableiten läßt, d. h. für alle Abbildungen derselben Klasse. Gibt es auf G ein Gebiet, dessen Punkte bei \mathfrak{A} höchstens r -fach überdeckt werden, d. h. Bildpunkte von je höchstens r Punkten von F sind, so ist $|n| \leq r$. Beispiele führen zu der Vermutung:

Jede stetige Abbildung vom Grade n einer geschlossenen orientierten Fläche F auf eine andere G läßt sich stetig deformieren in eine solche, bei der ein Gebiet auf G genau $|n|$ -fach überdeckt wird.

Diese soll im folgenden bewiesen werden. Bei dem entsprechenden Satze über Mannigfaltigkeiten einer von 2 verschiedenen Dimension führt ein Beweis von Herrn H. Hopf²⁾, dem ich die Anregung zu dieser Untersuchung verdanke, zum Ziel.

Die Hilfsmittel sind größtenteils die von Brouwer^{3) 4)} in die Topologie eingeführten und bei der Aufzählung der Abbildungsklassen benutzten, worauf im einzelnen durch Fußnoten verwiesen wird. Des Zusammenhangs halber wurden sie jedoch vollständig dargelegt.

¹⁾ L. E. J. Brouwer, Über Abbildung von Mannigfaltigkeiten, *Math. Annalen* 71 (1911), S. 97–115; B. v. Kerékjártó, *Vorlesungen über Topologie I* (1923), S. 230–237.

²⁾ In einer demnächst in den *Math. Annalen* erscheinenden Arbeit.

³⁾ L. E. J. Brouwer, Over éénéénduidige, continue transformaties van oppervlakken in zichzelf (5de mededeeling). *Kon. Ak. van Wet. Amsterdam, Versl. d. wis. en natuurk. Afd.* 21 (1912), S. 300–309.

⁴⁾ L. E. J. Brouwer, Aufzählung der Abbildungsklassen endlichfach zusammenhängender Flächen, *Math. Annalen* 82 (1921), S. 280–286.

§ 1.

Maßbestimmung und Triangulation.

Auf den Flächen F und G werde in bekannter Weise eine Metrik konstanten Krümmungsmaßes eingeführt, indem die universelle Überlagerungsfläche mit der euklidischen oder nichteuklidischen Ebene, die Fläche selbst mit dem endlichen Fundamentalbereich einer fixpunkte- und umlegungsfreien Bewegungsgruppe \mathcal{G}_F bzw. \mathcal{G}_G — im Falle des Geschlechtes Null mit der Kugelfläche vom Radius Eins — identifiziert werden. Die Abbildung \mathfrak{A} bestimmt eine Abbildung $\overline{\mathfrak{A}}$ der Überlagerungsfläche von F auf die von G . Alle Konstruktionen, die mit Punkten der Überlagerungsflächen vorgenommen werden, sollen gleichzeitig an allen nach \mathcal{G}_F bzw. \mathcal{G}_G äquivalenten Punkten vorgenommen werden, so daß z. B. Abänderungen von $\overline{\mathfrak{A}}$ zugleich auch Abänderungen von \mathfrak{A} bestimmen.

Die Flächen F und G seien trianguliert, und zwar seien die Dreiecke geradlinige Dreiecke im Sinne der Maßbestimmung. Die dabei beteiligten Punkte und Strecken heißen Ecken und Seiten. Zwei verschiedene Seiten sollen höchstens einen Endpunkt, zwei verschiedene Dreiecke entweder eine Ecke oder eine Seite oder keinen Punkt gemeinsam haben. Im Falle des Geschlechtes Null seien die Durchmesser der Dreiecke kleiner als $\pi : 3$.

Werden zwei Dreiecke mit einer gemeinsamen Seite geteilt durch die Verbindungsstrecken der Seitenmitte mit den gegenüberliegenden Ecken, so entsteht eine Triangulation mit denselben Eigenschaften.

§ 2.

Normalabbildungen. Erste Reduktion.

Eine Abbildung von \overline{F} auf \overline{G} heiße halbnormal, wenn sie jedes Dreieck der Triangulation nach einer (ausgearteten oder nicht ausgearteten) Kollineation abbildet, die die Seitenmitten wieder in die Seitenmitten verwandelt. Schreibt man die Lage der drei Ecken vor, so ist dadurch die halbnormale Abbildung des Dreiecks eindeutig bestimmt — im Falle des Geschlechtes Null unter der Voraussetzung, daß die Abstände der drei Punkte kleiner als $\pi : 3$ sind.

Die Abbildung $\overline{\mathfrak{A}}$ werde jetzt in eine halbnormale Abbildung $\overline{\mathfrak{B}}$ deformiert, und zwar in diejenige, die die Ecken von \overline{F} ebenso abbildet wie $\overline{\mathfrak{A}}$. Es wandere z. B. jeder Bildpunkt von seiner ursprünglichen Lage in seine Lage bei $\overline{\mathfrak{B}}$ auf geradem Wege mit konstanter Geschwindigkeit, so daß er in der Zeiteinheit gerade ankommt. Hat G das Geschlecht Null, so ist vorher die Triangulation von F so fein zu wählen, daß das Bild jedes Dreiecks einen Durchmesser kleiner als $\pi : 3$ hat; dann ist die Entfernung

eines Bildpunktes vom Zielpunkt kleiner als $2\pi : 3$, und die Verbindungsstrecke ist als kürzeste eindeutig und in stetiger Abhängigkeit von der Lage des Punktes bestimmt.

Da bei der Deformation von $\bar{\mathfrak{A}}$ in $\bar{\mathfrak{B}}$ die Bilder nach \mathfrak{G}_F äquivalenter Punkte nach \mathfrak{G}_G äquivalente Bahnen beschreiben, ergibt sich eine Deformation der Abbildung \mathfrak{A} in eine halbnormale \mathfrak{B} .

Wird jetzt die Triangulation von F verfeinert und werden für die neu eingeführten Ecken dieselben Bildpunkte vorgeschrieben wie bei \mathfrak{A} , so ist die dadurch bestimmte halbnormale Abbildung im allgemeinen von \mathfrak{B} verschieden; doch läßt sich \mathfrak{B} in sie deformieren nach dem soeben benutzten Verfahren der gleichförmig-geradlinigen Wanderung.

Eine halbnormale Abbildung heiße normal, wenn die Endpunkte jeder Seite von F auf dieselbe Ecke von G oder auf die Endpunkte einer Seite von G abgebildet werden. Nach den Eigenschaften der Triangulation von G wird dann jedes Dreieck von F entweder auf ein Dreieck von G oder auf eine Seite von G (und zwar mit einer Seite auf den einen, mit der gegenüberliegenden Ecke auf den anderen Endpunkt) oder auf eine Ecke von G abgebildet.

Um nun die halbnormale Abbildung \mathfrak{B} in eine normale Abbildung \mathfrak{C} zu deformieren, genügt es, jeder Ecke p von F eine Ecke p' von G zuzuordnen derart, daß den Endpunkten einer Seite von F dieselbe Ecke oder wieder die Endpunkte einer Seite von G zugeordnet sind, und daß, wenn G das Geschlecht Null hat, der Abstand des Bildes von p bei \mathfrak{B} von p' kleiner als $\pi : 3$ ist. Zu diesem Zweck wird die Triangulation von F so verfeinert, daß das Bild der Gesamtheit der an eine Seite von F anstoßenden Dreiecke immer ganz im Innern eines Dreieckssterns von G liegt⁵⁾. Um so mehr liegt dann auch das Bild jedes Dreieckssterns von F in einem Stern von G . Wird nun jeder Ecke p von F eine Ecke p' von G so zugeordnet, daß das Bild des Sternes von p im Stern von p' liegt, so erfüllt diese Zuordnung die gestellte Forderung. In der Tat; das Bild einer Seite pq von F liegt in einem Stern s von G . Die Ecken p und q müssen also zwei dem Stern s angehörenden Ecken p' und q' von G zugeordnet sein, deren Sterne innere Punkte gemeinsam haben, die zugleich

⁵⁾ Als Dreiecksstern bezeichnet man bekanntlich die Gesamtheit der an eine Ecke anstoßenden Dreiecke. Ist d_i der Abstand des Dreiecks A_i von der Gesamtheit der zu ihm punktfremden Dreiecke und d der kleinste der Abstände d_i , ist ferner e kleiner als d und als die kleinste Länge, die die längste Seite eines einem Dreieck von G eingeschriebenen Dreiecks haben kann, so ist jede zusammenhängende Punktmenge vom Durchmesser e in einem Dreiecksstern enthalten. Beziehen sich diese Angaben auf die Fläche G , so braucht die Triangulation von F nur so fein gewählt zu werden, daß die Durchmesser der Bilder der Dreiecke von F sämtlich kleiner sind als $e : 3$.

in s liegen. Das ist nur der Fall, wenn p' und q' zusammenfallen oder Endpunkte einer Seite sind.

In die durch diese Zuordnung bestimmte Normalabbildung \mathcal{C} wird \mathfrak{B} nach dem Verfahren der geradlinig-gleichförmigen Wanderung deformiert.

§ 3.

Falten. Zweite Reduktion.

Die Dreiecke von F zerfallen jetzt in drei Klassen: Ecken-, Seiten- und Volldreiecke, je nachdem ob ihre Ecken auf eine, zwei oder drei verschiedene Ecken von G abgebildet sind. Ebenso zerfallen die Seiten in Eckenseiten, die auf je eine Ecke, und Vollseiten, die auf je eine Seite von G abgebildet sind. Jedes Seitendreieck hat eine Ecken- und zwei Vollseiten. Mit den Vollseiten grenzt es an Voll- oder Seitendreiecke. Jedes Eckendreieck wird von Eckenseiten berandet und grenzt in ihnen an Ecken- oder Seitendreiecke.

Zwei Vollseiten von F heißen verbunden, wenn sie die beiden Vollseiten eines Seitendreiecks sind; sie werden dann auf dieselbe Seite von G abgebildet. Da eine Vollseite höchstens zwei Seitendreiecke berandet und jedes von diesen nur eine weitere Vollseite hat, so ist jede Vollseite mit höchstens zwei Vollseiten verbunden. Daher ordnen sich alle Vollseiten zu endlichen Ketten, in denen benachbarte verbunden sind. (Eine Kette kann auch aus einer einzigen Seite bestehen.) Schließt sich die Kette nicht, so beranden ihr erstes und letztes Glied je ein Volldreieck; besteht sie nur aus einer Seite, so berandet diese zwei Volldreiecke. Jedes dieser beiden Dreiecke wird auf eines der beiden Dreiecke von G abgebildet, die von der Bildseite der Seiten der Kette berandet werden. Die Kette heiße von *erster Art*, wenn beide Dreiecke auf verschiedene Dreiecke abgebildet werden, sonst von *zweiter Art*. Geschlossene Ketten sollen *Ketten dritter Art* heißen.

Ist eine Kette zweiter Art vorhanden, so wird die Abbildung \mathcal{C} so

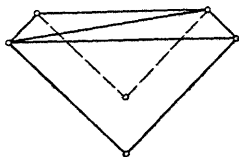


Fig. 1 a.

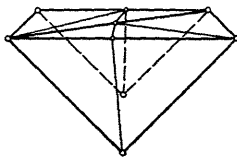


Fig. 1 b.

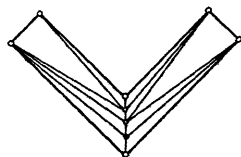


Fig. 1 c.

abgeändert, wie es Fig. 1 a, 1 b, 1 c andeuten⁶⁾. Stellt Fig. 1 a die Abbildung der Seitendreiecke einer Kette zweiter Art und der anstoßenden beiden Volldreiecke dar (Punkte mit denselben Bildpunkten sind nahe beieinander

⁶⁾ Vgl. Brouwer a. a. O. ³⁾, S. 303.

gezeichnet), so wird zuerst die Triangulation von F verfeinert durch die in Fig. 1 b angegebenen Seitenhalbierungen und Dreiecksteilungen (vgl. Schluß von § 1) und dann die Abbildung abgeändert, indem den neu hinzukommenden Ecken die in Fig. 1 c angegebenen Plätze zugewiesen werden. Offenbar entsteht wieder eine Normalabbildung, in die sich die vorliegende durch Deformation verwandeln läßt. Da bei der neuen Abbildung zwei Volldreiecke weniger vorhanden sind, führt dies Verfahren nach einer endlichen Zahl von Wiederholungen zu einer Abbildung \mathfrak{D} , bei der keine Ketten zweiter Art mehr vorhanden sind.

§ 4.

Dritte Reduktion.

Ein Volldreieck von F heiße *positiv*, wenn seine durch die Abbildung \mathfrak{D} auf G übertragene Orientierung mit der von G übereinstimmt, sonst *negativ*.

Sind Δ_1 und Δ_2 die beiden von einer Seite s von G berandeten Dreiecke von G , so geht von jedem auf Δ_1 abgebildeten Volldreieck Δ_1^* von F eine auf s abgebildete Kette erster Art aus, die in einem auf Δ_2 abgebildeten Volldreieck Δ_2^* ihr Ende findet. Diese Zuordnung der auf Δ_1 und Δ_2 abgebildeten Volldreiecke ist eindeutig umkehrbar. Fig. 2 zeigt, daß gepaarte Dreiecke beide positiv oder beide negativ sind. Also werden auf Nachbardreiecke von G und daher auf irgend zwei Dreiecke von G gleichviel positive und negative Volldreiecke abgebildet.

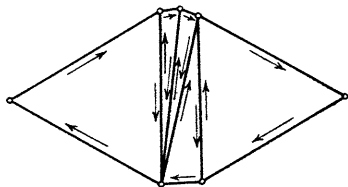


Fig. 2.

Sind nur positive oder nur negative Dreiecke vorhanden, so ist die Anzahl der Bedeckungen, mit dem Plus- bzw. Minuszeichen versehen, der Abbildungsgrad, und der zu beweisende Satz ist bestätigt. Anderenfalls verfährt man wie folgt.

Die Gesamtheit der auf eine bestimmte Ecke von G abgebildeten Zellen (d. h. Ecken, Seiten und Dreiecke) von F zerfällt in eine Anzahl größter zusammenhängender Komplexe, die *Bündel* heißen mögen. Stößt an eine Zelle eines Bündels ein Volldreieck an, so heiße das Bündel ein *Hauptbündel*, sonst ein *Nebenbündel*. Stößt eine Kette erster Art an ein Bündel an, so ist dieses ein Hauptbündel.

Die Abbildung soll so abgeändert werden, daß ein Hauptbündel vorhanden ist, an das sowohl positive wie negative Volldreiecke anstoßen. Ist dies noch nicht der Fall, so können die Hauptbündel in positive und negative eingeteilt werden. Eine Ecke p eines beliebigen positiven Bün-

dels werde mit einer Ecke q eines beliebigen negativen Bündels durch einen Seitenzug verbunden. Dieser Zug werde so gewählt, daß in ihm möglichst wenig Vollseiten auftreten. (Mindestens eine ist vorhanden, sonst würde der ganze Zug demselben Bündel angehören.) Dann enthält der Zug außer je einem Teilzug am Anfang und am Ende keine Ecke (und daher auch keine Seite) eines Hauptbündels. Wäre nämlich r eine solche, so wäre pr oder rq ein Zug mit weniger Vollseiten und mit denselben Eigenschaften (je nachdem, ob r einem negativen oder positiven Hauptbündel angehört). Ein Nebenbündel b kann der Zug nur einmal betreten und verlassen. Hätte er nämlich mit b zwei getrennte Teilzüge gemeinsam, so würde man den dazwischen liegenden Teil, der mehrere Vollseiten enthält, ersetzen durch einen ganz in b verlaufenden Zug, der ja nur Ecken-seiten enthält.

Das Anfangsbündel b_0 des Zuges werde auf die Ecke p von G , die erste Vollseite des Zuges auf die Seite pt , das zweite Bündel b_1 des

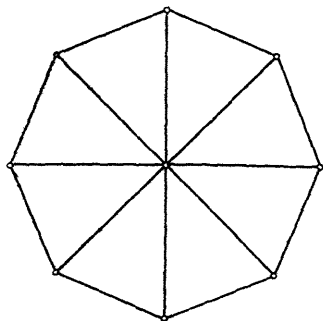


Fig. 3a.

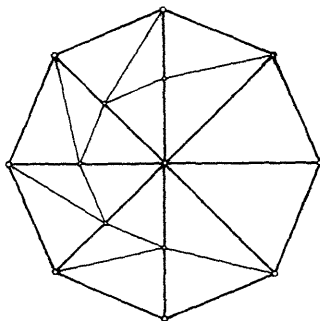


Fig. 3a.

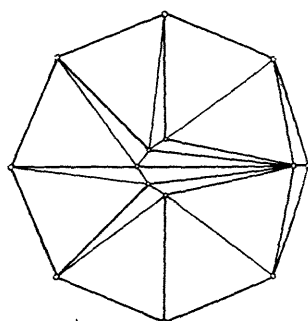


Fig. 3c.

Zuges also auf die Ecke t abgebildet. Fig. 3a stelle den Stern von p dar; die rechts gelegene Ecke sei t . Die an das Bündel b_0 anstoßenden Seiten- und Voldreiecke werden so geteilt, wie es Fig. 3b angibt; wie mit den



Fig. 4a.



Fig. 4b.

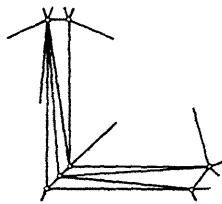


Fig. 4c.

Seitendreiecken zu verfahren ist, zeigt die Teilfigur 4. Dann wird die Abbildung der behandelten Zellen durch stetige Deformation so abgeändert, wie es Fig. 3c zeigt. Die Strecken und Dreiecke aus Fig. 3b erkennt man

an ihrer kombinatorischen Stellung wieder. Jetzt ist offenbar aus b_1 ein positives Hauptbündel geworden. Dagegen ist bei keinem weiteren Bündel des Zuges die Abbildung geändert worden; denn jedes Bündel r , an das Zellen anstoßen, deren Abbildung geändert wurde, wird mit b_0 durch eine einzige Vollseite verbunden, so daß der Zug, wenn er das Bündel r erreichte, in seinem ersten Teil durch einzige Vollseite $b_0 r$ zu ersetzen, also nicht in dem vorher geforderten Sinne der kürzeste wäre. Jetzt hat also der b_1 mit dem Schlußbündel verbindende, um eine Vollseite kürzere Teil des Zuges die Eigenschaft, ein positives Hauptbündel mit einem negativen zu verbinden. Wiederholung dieses Verfahrens führt schließlich zu einer Abbildung, bei der an ein Bündel b^* sowohl positive wie negative Voldreiecke anstoßen.

Das Bündel b^* sei auf die Ecke p^* von G abgebildet. Schreitet man von einem an b^* anstoßenden positiven bzw. negativen Voldreiecke über die eine der beiden an ihm liegenden und an b^* anstoßenden Ketten erster Art fort zu dem nächsten Voldreieck, so wird dies gleichfalls positiv bzw. negativ abgebildet auf das an p^* anstoßende Nachbardreieck von G und stößt gleichfalls an b^* an. Also werden auf jedes an p^* anstoßende Dreieck gleichviele an b^* anstoßende Voldreiecke positiv und gleichviele negativ abgebildet; insbesondere also auf jedes von ihnen mindestens eines positiv und eines negativ.

Nunmehr⁷⁾ werde von jedem an b^* anstoßenden Voll- und Seitendreieck nach Fig. 5 ein Dreieck oder ein aus zwei Dreiecken bestehendes Viereck abgetrennt und die Abbildung so abgeändert, daß alle diese auch auf p^*

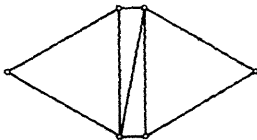


Fig. 5a.

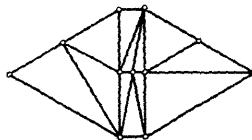


Fig. 5b.

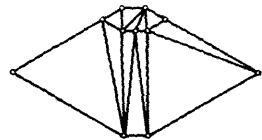


Fig. 5c.

abgebildet werden, wie es Fig. 5 andeutet. Die zu oberst nahe bei einander gelegenen Punkte seien auf p_0 abgebildet; die beiden schmalen Dreiecke seien Seitendreiecke. Dadurch wird zu b^* längs jeder Randlinie ein Kranz von Dreiecken hinzugefügt, und b^* wird zweidimensional zusammenhängend, d. h. man kann irgend zwei Seiten von b^* durch eine ganz zu b^* gehörige Kette von Dreiecken verbinden, in der Nachbardreiecke je eine Seite gemeinsam haben. Außerdem ist dadurch erreicht, daß an jedem Voldreieck ein Seitendreieck liegt, dessen Eckenseite zu b^*

⁷⁾ Vgl. Brouwer a. a. O.³⁾, S. 306.

gehört, und zwar sei dies so geschehen, daß die besagten Seitendreiecke bei zwei Volldreiecken, die auf dasselbe Dreieck von G abgebildet sind, auf dieselbe Seite dies letzteren abgebildet sind. Man wähle zwei Seitendreiecke, die in dieser Weise an zwei Volldreiecken liegen, die, das eine positiv, das andere negativ, auf dasselbe Dreieck Δ von G abgebildet sind. Sind uv bzw. u_1v_1 die Eckenseiten der beiden Seitendreiecke, und zwar u und u_1 die an den beiden Volldreiecken gelegenen Ecken, so lassen sich uv und u_1v_1 durch eine Kette zu b^* gehöriger Dreiecke verbinden. Diese sei zu nächst so gewählt, daß kein Dreieck in ihr zweimal auftritt. Trennt man

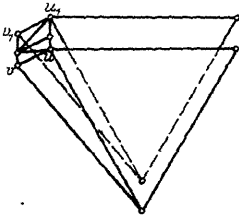


Fig. 6a.

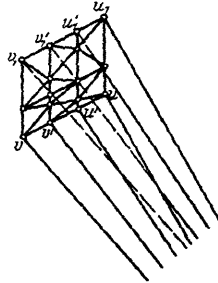


Fig. 6b.

dann nach Fig. 6a und 6b beiderseits einen Streifen von ihr ab, so zeigt die Betrachtung der Orientierung, daß von den beiden in Fig. 6b senkrecht gezeichneten Seitenzügen der eine u' mit u'_1 , der andere v' mit v'_1 verbinden muß. Sonst würde sich nämlich ergeben, daß die beiden Volldreiecke mit gleicher Orientierung auf G abgebildet werden, während doch das eine positiv, das andere negativ ist.

Schließlich⁸⁾ werden die beiden Volldreiecke und die beiden Seitendreiecke nach Fig. 7 geteilt, das Dreieck Δ nach Fig. 8 und die Abbildung

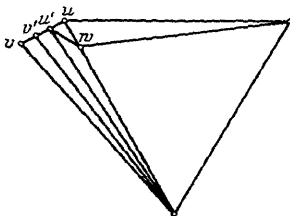


Fig. 7.

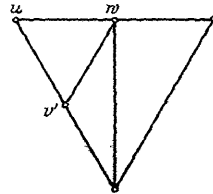


Fig. 8.

dahin abgeändert, daß die Punkte u' und u'_1 so wie der sie verbindende, in Fig. 6b senkrecht gezeichnete Zug in die in Fig. 8 mit w bezeichnete Lage kommen, w und der im anderen Volldreieck entsprechende Punkt w_1

⁸⁾ Vgl. für den Fall der Kugelfläche Brouwer a. a. O.³⁾, S. 308.

ebendorthin, v' , v'_1 und der Streckenzug $v'v'_1$ in den mit v' bezeichneten Punkt von Fig. 8. Alle übrigen Ecken von F werden wie vorher abgebildet. Die so bestimmte halbnormale Abbildung läßt sich auch aus der ursprünglichen durch Deformation ableiten. Bei ihr wird das Dreieck $uv'w$ in Fig. 8 zweimal weniger überdeckt als vorher. Auch diese Abbildung läßt sich in eine normale verwandeln, indem man geeignete Dreiecksteilungen in F vornimmt und dann in den einzelnen Dreiecken von der vorliegenden Abbildung zu der dadurch bestimmten normalen übergeht.

§ 5.

Schlußfolgerung. Bemerkungen.

Die Wirkung der angegebenen Reduktionen ist die folgende. Die erste verwandelt eine beliebige Abbildung in eine normale; die beiden anderen verwandeln eine Normalabbildung wieder in eine solche. Die zweite Reduktion ändert die Triangulation von G nicht, erhöht die Mindestzahl der Überdeckungen bei den einzelnen Dreiecken von G nicht und vermindert die Gesamtzahl der Überdeckungen der Dreiecke. Die dritte ist anwendbar, wenn es die zweite nicht ist, außer wenn der in dem zu beweisenden Satz als Ziel hingestellte Sachverhalt besteht; sie vermindert die Mindestzahl der Überdeckungen der einzelnen Dreiecke von G . Aus diesen Wirkungen folgt, daß die zweite und dritte Reduktion nur endlich oft stattfinden können, daß sich also nach endlich vielen stetigen Deformationen der in der Einleitung behauptete Sachverhalt einstellt. Damit ist das Ziel der Arbeit erreicht, sogar noch mehr; es ist die Abbildung verwandelt in eine solche, bei der alle Punkte von G außerhalb der Seiten einer gewissen Triangulation genau $|n|$ -fach überdeckt sind.

Die in § 1 eingeführten Hilfsmittel, nämlich die Darstellung der Flächen als Fundamentalbereiche euklidischer oder nichteuklidischer Bewegungsgruppen, lassen sich natürlich nicht ohne weiteres auf die entsprechende Fragestellung bei mehrdimensionalen Mannigfaltigkeiten übertragen; doch sind sie nicht so wesentlich, daß sich nicht durch Einführung geeigneter örtlicher Koordinatensysteme Ersatz schaffen ließe. Dagegen lassen sich die zweite und dritte Reduktion ohne weiteres sinngemäß übertragen; nur gestaltet sich die hier gewählte eingehende kombinatorische Darstellung naturgemäß langwieriger.

Dimension und Zusammenhangsstufe.

Von

W. Hurewicz in Amsterdam und K. Menger in Wien.

1. Der Begriff des höherstufigen Zusammenhanges.

In früheren Arbeiten von einem von uns¹⁾ wurde die Möglichkeit einer Beschreibung der Zusammenhangsverhältnisse der Räume auseinandergesetzt, welche die übliche Definition des Zusammenhanges durch Einführung des *Zusammenhanges verschiedener Stufen* erheblich verfeinert. Es wurde dort darauf hingewiesen, daß sowohl die Strecke als auch die Quadratfläche und der Würfelkörper zusammenhängend sind, daß wir aber doch geneigt sind, die Quadratfläche *fester* zusammenhängend als die Strecke und *minder stark* zusammenhängend als den Würfel zu nennen, und daß uns ebenso eine Menge, welche Summe ist von zwei Quadratflächen, die bloß einen Eckpunkt miteinander gemein haben, *schwächer* zusammenhängend erscheint als eine Quadratfläche.

Aus diesen Beispielen konnte auch unschwer eine strenge Präzisierung unserer anschaulichen Vorstellung des festeren Zusammenhanges hergeleitet werden. *Zusammenhängend schlechthin* nannte man bisher einen Raum, welcher nicht Summe ist von zwei fremden nicht-leeren abgeschlossenen Teilmengen. Derartige Räume können, wofern es sich um ihren Vergleich mit fester zusammenhängenden Räumen handelt, auch als *mindestens einstufig zusammenhängend* bezeichnet werden²⁾, wobei es für das Folgende zweckmäßig ist, ihre Definition in der folgenden Form auszusprechen: *Ein Raum heißt mindestens einstufig zusammenhängend, wenn er*

¹⁾ Vgl. Menger, „Bericht über die Dimensionstheorie“, Jahresber. d. deutschen Math. Ver. **35**, S. 144 f. Siehe ferner die Hinweise in den beiden Abhandlungen „Über die Dimension von Punktmengen“, Monatshefte f. Math. u. Phys. **33** (1923), S. 160 und **34** (1924), S. 156 u. S. 161, sowie die Note „Zusammenhangsstufe und Cantorsche Mannigfaltigkeiten“, Proc. Ac. Amsterdam **30**, S. 705.

²⁾ Vgl. Menger, Jahresber. d. deutschen Math. Ver. **35**, S. 144.

erstens mehr als einen Punkt enthält und zweitens nicht Summe ist von zwei abgeschlossenen echten Teilmengen mit leerem Durchschnitt. Das, was der Anschauung entsprechend als festerer Zusammenhang und zwar als nächste Zusammenhangsstufe, als zweistufiger Zusammenhang bezeichnet werden kann, läßt sich nun erklären als Nichtzerlegbarkeit in zwei abgeschlossene echte Teile mit einstufig zusammenhangslosem Durchschnitt, wobei als einstufig zusammenhangslos eine Menge ohne mindestens einstufig zusammenhängenden Teil zu bezeichnen ist. Auf diese Weise erscheint eine Strecke als bloß einstufig zusammenhängend; desgleichen ist die Summe zweier Quadratiflächen, welche nur einen Eckpunkt miteinander gemein haben, bloß einstufig, nicht aber zweistufig zusammenhängend, denn die beiden Quadrate, aus denen diese Menge besteht, sind abgeschlossene echte Teilmengen derselben, deren Durchschnitt aus einem Punkt besteht, also einstufig zusammenhangslos ist.

Daß zur Erklärung des zweistufigen Zusammenhanges Nichtspaltbarkeit in abgeschlossene echte Teilmengen mit einstufig zusammenhangslosem Durchschnitt verlangt werden mußte, erkennt man daraus, daß in zwei abgeschlossene Teilmengen mit einstufig zusammenhangslosem Durchschnitt, von denen die eine nicht-echte Teilmenge ist (d. h. mit der ganzen Menge übereinstimmt), jede zusammenhängende Menge in trivialer Weise zerlegbar ist, auch die Quadratifläche, der wir doch eben eine höhere Zusammenhangsstufe zuzuordnen wünschen. Um beispielsweise das Quadrat als Summe zweier abgeschlossener nicht-echter Teilmengen mit zusammenhangslosem Durchschnitt zu zerlegen, braucht man nur als den einen Summanden irgendeine einstufig zusammenhangslose abgeschlossene Teilmenge des Quadrates und als anderen Summanden das Quadrat selbst zu wählen. Nur Zerlegungen in echte Teilmengen dürfen also zur Definition des höherstufigen Zusammenhanges in Betracht gezogen werden³⁾.

Die angeführte Definition des zweistufigen Zusammenhanges kann rekursiv fortgeführt werden: *Ein Raum heißt*⁴⁾ *mindestens n -stufig zusammenhängend, wenn er nicht Summe ist von zwei abgeschlossenen echten Teilmengen mit $(n - 1)$ -stufig zusammenhangslosem Durchschnitt, wobei eine Menge $(n - 1)$ -stufig zusammenhangslos heißt, wenn sie keinen $(n - 1)$ -stufig zusammenhängenden Teil enthält.* Setzen wir als Ausgangspunkt der Rekursion, analog zur Dimensionstheorie⁵⁾, *die leere Menge und*

³⁾ Daß das Quadrat nicht Summe ist von zwei abgeschlossenen echten Teilmengen mit einstufig zusammenhangslosem Durchschnitt läßt sich in der Tat beweisen, worauf wir hier nicht näher eingehen.

⁴⁾ Vgl. Menger, Proc. Ac. Amsterdam 30, S. 305.

⁵⁾ Über die leere Menge als die (-1) -dimensionale vgl. Menger, Monatshefte f. Math. u. Phys. 33, S. 158 ff. und Urysohn, Fund. Math. 7, S. 66.

nur diese als nullstufig zusammenhangslos fest, so ordnet sich die obige Definition der mindestens einstufig zusammenhängenden Mengen (der mindestens zwei Punkte enthaltenden zusammenhängenden Mengen schlechthin) in die Rekursion ein, wofern hinsichtlich der letzteren Mengen diese rekursive Definition durch die Forderung ergänzt wird, daß eine mindestens einstufig zusammenhängende Menge mehr als einen Punkt enthalte. Ein mindestens n -stufig zusammenhängender, aber nicht mindestens $(n + 1)$ -stufig zusammenhängender Raum kann auch als *genau n -stufig zusammenhängend* oder als *n -stufig zusammenhängend* schlechthin bezeichnet werden.

2. Höherstufiger Zusammenhang und Dimension.

Bei der Aufstellung des Begriffes vom n -stufigen Zusammenhang ergab sich sogleich⁶⁾ die Frage nach den Beziehungen von n -stufigem Zusammenhang und n -Dimensionalität. Ein diesbezügliches Resultat kann durch einen einfachen Induktionsschluß bewiesen werden⁷⁾. *Jeder Raum, welcher eine mindestens n -stufig zusammenhängende Teilmenge enthält, ist mindestens n -dimensional ($n \geq 1$); oder was gleichbedeutend ist: Ein höchstens $(n - 1)$ -dimensionaler Raum ist n -stufig zusammenhangslos ($n \geq 1$). Diese Behauptung ist richtig für $n = 1$. Denn sei ein nulldimensionaler Raum R gegeben. Wir machen die Annahme, R enthielte eine mindestens einstufig zusammenhängende Teilmenge M und leiten aus dieser Annahme einen Widerspruch her. Nach Definition des mindestens einstufigen Zusammenhanges enthielte M sicher zwei verschiedene Punkte p und q . Da R als nulldimensional vorausgesetzt ist, existieren beliebig kleine Umgebungen von p mit leeren Begrenzungen. Insbesondere also existiert eine Umgebung U von p mit leerer Begrenzung, so daß q im Komplement $R - U$ von U läge. Durch die Formel $R = U \uplus (R - U)$ ist R dargestellt als Summe zweier Mengen, die fremd (weil komplementär) und die beide abgeschlossen sind ($R - U$ ist als Komplement der offenen Menge U abgeschlossen, U ist es als offene Menge mit leerer Begrenzung). Die Formel $M = M \cdot U \uplus M \cdot (R - U)$ stellt also eine Zerlegung von M in zwei fremde in M abgeschlossene Mengen dar, von denen jede eine echte Teilmenge von M ist, denn $M \cdot U$ enthält nicht den Punkt q von M , während $M \cdot (R - U)$ nicht den Punkt p von M enthält. Eine solche Zerlegung widerspricht aber der Voraussetzung, M sei mindestens einstufig zusammenhängend. Damit ist die Behauptung für $n = 1$ bewiesen und wir wollen annehmen, es sei bereits bewiesen, daß ein*

⁶⁾ Vgl. Menger, Monatshefte f. Math. u. Phys. 33, S. 160.

⁷⁾ Vgl. Menger, Proc. Ac. Amsterdam 30, S. 706.

höchstens $(n - 2)$ -dimensionaler Raum $(n - 1)$ -stufig zusammenhangslos sei. Es sei dann ein höchstens $(n - 1)$ -dimensionaler Raum R gegeben ($n > 1$). Wir machen die Annahme, R enthalte eine mindestens n -stufig zusammenhängende Teilmenge M und leiten aus dieser Annahme ganz wie im Falle $n = 1$ einen Widerspruch her. Die Menge M enthält zwei verschiedene Punkte p und q . Da R höchstens $(n - 1)$ -dimensional ist, existieren beliebig kleine Umgebungen von p mit höchstens $(n - 2)$ -dimensionalen Begrenzungen. Insbesondere existiert also eine Umgebung U von p mit höchstens $(n - 2)$ -dimensionaler Begrenzung, welche so klein ist, daß der Punkt q im Komplement $R - \bar{U}$ der abgeschlossenen Hülle von U liegt. Die Formel $R = \bar{U} + (R - U)$ stellt eine Zerlegung von R in zwei abgeschlossene Mengen dar, deren Durchschnitt $\bar{U} \cdot (R - U)$, welcher ja mit der Begrenzung von U identisch ist, höchstens $(n - 2)$ -dimensional, also nach der als bewiesen angenommenen Behauptung $(n - 1)$ -stufig zusammenhangslos ist. Durch die Formel $M = M \cdot \bar{U} + M \cdot (R - U)$ wird demnach M dargestellt als Summe zweier in M abgeschlossener Mengen mit $(n - 1)$ -stufig zusammenhangslosem Durchschnitt, wobei jeder der beiden Summanden eine echte Teilmenge von M ist, denn $M \cdot \bar{U}$ enthält nicht den Punkt q , während $M \cdot (R - U)$ nicht den Punkt p enthält. Eine derartige Zerlegung widerspricht aber der Voraussetzung, M sei n -stufig zusammenhängend. Damit ist die Behauptung allgemein bewiesen.

Die *Umkehrung* der Behauptung, der Satz nämlich, daß jeder n -stufig zusammenhangslose Raum höchstens n -dimensional sei, oder, was gleichbedeutend ist, der Satz, daß jeder n -dimensionale Raum einen n -stufig zusammenhängenden Teil enthält, ist in dieser Allgemeinheit *unrichtig*. Bekanntlich gibt es ja nach Sierpiński⁸⁾ einstufig zusammenhangslose eindimensionale Mengen, und Mazurkiewicz hat neuerdings⁹⁾ Mengen jeder beliebigen endlichen Dimensionszahl konstruiert, welche keinen einstufig zusammenhängenden Teil enthalten und in denen überdies sogar je zwei Punkte durch die leere Menge trennbar sind. Beschränkt man sich indessen auf kompakte und halbkompakte¹⁰⁾ Räume, so gilt bekanntlich¹¹⁾ der Satz: *Jeder kompakte oder halbkompakte einstufig zusammenhangslose Raum ist nulldimensional*, oder, was gleichbedeutend ist, *jeder kompakte oder halbkompakte mehr als nulldimensionale Raum enthält mindestens einstufig zusammenhängende Teilmengen*. Es erhob sich daher gleich bei

⁸⁾ Vgl. Sierpiński, Fund. Math. 2, S. 81.

⁹⁾ Vgl. Mazurkiewicz, Fund. Math. 10, S. 311.

¹⁰⁾ Ein Raum heißt (vgl. Menger, Monatshefte f. Math. u. Phys. 34, S. 144) *halbkompakt*, wenn er Summe von abzählbar vielen kompakten Räumen ist.

¹¹⁾ Vgl. Menger, Monatshefte f. Math. 33, S. 159; Urysohn, Fund. Math. 7, S. 75; Mazurkiewicz, Fund. Math. 3, S. 67; Hurewicz, Math. Annalen 96, S. 753.

der Aufstellung der Begriffe von Dimension und höherstufigem Zusammenhang die Frage, ob unter den kompakten und halbkompakten Räumen die Verallgemeinerung der letzterwähnten Behauptung für beliebiges n gilt, und es wurde die Vermutung ausgesprochen¹²⁾, daß tatsächlich jeder kompakte n -dimensionale Raum einen n -stufig zusammenhängenden Teil enthält, daß also, indem wir diese Behauptung mit der oben bewiesenen kombinieren, unter den kompakten Räumen die höchstens n -dimensionalen identisch sind mit den $(n+1)$ -stufig zusammenhangslosen, und die n -dimensionalen mit jenen, welche einen n -stufig zusammenhängenden, aber keinen $(n+1)$ -stufig zusammenhängenden Teil enthalten. Beispielsweise ist die zweidimensionale Menge, bestehend aus zwei Quadratflächen, die bloß einen Eckpunkt miteinander gemein haben, zwar bloß einstufig zusammenhängend, aber sie enthält zweistufig zusammenhängende Teile.

3. Der Begriff der Cantorschen Mannigfaltigkeit.

Eine Begriffsbildung, welche mit jener des höherstufigen Zusammenhanges eine Analogie aufweist, wurde von Urysohn eingeführt¹³⁾. Urysohn bezeichnet als n -dimensionale Cantorsche Mannigfaltigkeit einen kompakten n -dimensionalen Raum, welcher nach Tilgung jeder höchstens $(n-2)$ -dimensionalen abgeschlossenen Teilmenge zusammenhängend bleibt. Urysohn formulierte sodann das folgende Problem, dessen besondere Wichtigkeit und Schwierigkeit er betonte¹⁴⁾: Enthält jeder kompakte n -dimensionale Raum eine n -dimensionale Cantorsche Mannigfaltigkeit?

Wir werden im nächsten Abschnitt diese Frage im positiven Sinn beantworten. Hier reproduzieren wir zunächst den Beweis von einem von uns¹⁵⁾ für die Tatsache, daß eine positive Antwort auf die Urysohnsche Frage äquivalent ist mit der erwähnten Vermutung von der Identität der höchstens n -dimensionalen und der $(n+1)$ -stufig zusammenhangslosen kompakten Räume.

Um diese Äquivalenz einzusehen, bemerken wir vor allem, daß jeder kompakte n -stufig zusammenhängende n -dimensionale Raum eine n -di-

¹²⁾ Vgl. Menger, Monatshefte f. Math. u. Phys. 34, S. 156. Infolge eines Versehens bei der Drucklegung steht a. a. O., die n -dimensionalen Räume seien identisch „mit den $(n+1)$ -stufig zusammenhangslosen, n -stufig zusammenhängenden Räumen“. Es soll selbstverständlich heißen, „mit den $(n+1)$ -stufig zusammenhangslosen Räumen, die einen n -stufig zusammenhängenden Teil enthalten“. Denn wie aus den oben zitierten Beispielen (aus Menger, Jahresber. d. deutschen Math. Ver. 35, S. 144) hervorgeht, ist natürlich nicht jeder n -dimensionale Raum n -stufig zusammenhängend.

¹³⁾ Vgl. Urysohn, Fund. Math. 7, S. 124.

¹⁴⁾ Vgl. Fund. Math. 8, S. 285.

¹⁵⁾ Vgl. Menger, Proc. Ac. Amsterdam 30, S. 707.

mensionale Cantorsche Mannigfaltigkeit ist. Sei nämlich irgendein kompakter n -dimensionaler Raum R gegeben, welcher keine n -dimensionale Mannigfaltigkeit ist, welcher also eine höchstens $(n - 2)$ -dimensionale abgeschlossene Teilmenge A enthält, so daß $R - A$ nicht zusammenhängend ist, d. h. so, daß $R - A$ Summe ist von zwei fremden nicht-leeren relativ abgeschlossenen Teilmengen R' und R'' . Betrachten wir die Menge $R' + A$. Sie ist *erstens* eine echte Teilmenge von R , denn sie ist ja zu der nicht-leeren Teilmenge R'' von R fremd. Die Menge $R' + A$ ist *zweitens* abgeschlossen. Denn A ist nach Voraussetzung abgeschlossen, also ist $R - A$ offen. Die Menge R' ist in $R - A$ abgeschlossen, folglich ist $R'' = (R - A) - R'$ in der offenen Menge $R - A$ offen, also ist R'' offen und mithin ist $R' + A = R - R''$ abgeschlossen. Eine analoge Überlegung zeigt, daß die Menge $R'' + A$ eine abgeschlossene echte Teilmenge von R ist. Durch die Formel $R = (R' + A) + (R'' + A)$ ist R also dargestellt als Summe zweier abgeschlossener echter Teilmengen, deren Durchschnitt (die Menge A) höchstens $(n - 2)$ -dimensional, also, wie wir am Anfang von § 2 zeigten, $(n - 1)$ -stufig zusammenhangslos ist. Demnach ist R höchstens $(n - 1)$ -stufig zusammenhängend.

Machen wir jetzt *erstens* die Annahme I, daß jeder n -dimensionale kompakte Raum eine n -stufig zusammenhängende Teilmenge enthält, dann enthält auf Grund des soeben Bewiesenen jeder kompakte n -dimensionale Raum eine n -dimensionale Cantorsche Mannigfaltigkeit, so daß also aus der Annahme I die Annahme II folgt.

Machen wir *zweitens* die Annahme II, daß jeder kompakte n -dimensionale Raum eine n -dimensionale Cantorsche Mannigfaltigkeit enthält. Wir wollen aus dieser Annahme herleiten, daß jeder $(n + 1)$ -stufig zusammenhangslose kompakte Raum höchstens n -dimensional ist; mit andern Worten, daß jeder kompakte n -dimensionale Raum einen n -stufig zusammenhängenden Teil enthält. Es genügt zu diesem Zweck offenbar, aus der Annahme II herzuleiten, daß jede n -dimensionale Cantorsche Mannigfaltigkeit n -stufig zusammenhängend ist. Diese Behauptung ist richtig für $n = 1$. Denn eine eindimensionale Cantorsche Mannigfaltigkeit, d. h. ein kompakter Raum, der sich durch die (-1) -dimensionale (die leere) Menge nicht in zwei nichtleere abgeschlossene Teile zerlegen läßt, ist offenbar mindestens einstufig zusammenhängend. Wir setzen voraus, es sei auf Grund der Annahme II bereits bewiesen, daß jede $(n - 1)$ -dimensionale Cantorsche Mannigfaltigkeit $(n - 1)$ -stufig zusammenhängend sei. Aus dieser Voraussetzung folgt auf Grund der Annahme II: Jeder $(n - 1)$ -stufig zusammenhangslose kompakte Raum ist höchstens $(n - 2)$ -dimensional. Es sei nun ein n -dimensionaler kompakter Raum R gegeben, welcher nicht n -stufig zusammenhängend ist, welcher also Summe ist von zwei ab-

geschlossenen echten Teilmengen R' und R'' , deren Durchschnitt $(n-1)$ -stufig zusammenhangslos ist. Auf Grund der Annahme II ist, wie soeben bemerkt, der Durchschnitt $R' \cdot R''$ höchstens $(n-2)$ -dimensional; also ist R keine n -dimensionale Cantorsche Mannigfaltigkeit. Mit andern Worten: aus der Annahme II folgt, daß jede n -dimensionale Cantorsche Mannigfaltigkeit n -stufig zusammenhängend ist. Falls also jeder kompakte n -dimensionale Raum eine n -dimensionale Cantorsche Mannigfaltigkeit enthält, dann enthält jeder kompakte n -dimensionale Raum auch einen n -stufig zusammenhängenden Teil, w. z. b. w.

Es sind demnach die beiden folgenden Sätze äquivalent¹⁶⁾:

Jeder kompakte n -dimensionale Raum enthält einen n -stufig zusammenhängenden Teil, und

Jeder kompakte n -dimensionale Raum enthält eine n -dimensionale Cantorsche Mannigfaltigkeit.

Da, wie oben bewiesen wurde, die höchstens n -dimensionalen Räume $(n+1)$ -stufig zusammenhangslos sind, so ist mit den beiden angeführten Sätzen auch der folgende äquivalent^{16a)}:

Unter den kompakten Räumen sind die höchstens n -dimensionalen und die $(n+1)$ -stufig zusammenhangslosen identisch.

Nehmen wir nun an, diese drei untereinander äquivalenten Sätze seien richtig. Wie wir oben sahen, folgt aus dieser Annahme, daß jede n -dimensionale Cantorsche Mannigfaltigkeit n -stufig zusammenhängend ist. Andererseits wurde oben gezeigt, daß ein n -dimensionaler und n -stufig zusammenhängender kompakter Raum eine n -dimensionale Cantorsche Mannigfaltigkeit ist. Wenn also die drei als äquivalent erwiesenen Sätze richtig sind, so ergibt sich¹⁷⁾: *Eine Cantorsche Mannigfaltigkeit ist ein kompakter Raum, für welchen Dimension und Zusammenhangsstufe gleich sind. Die n -dimensionalen Cantorschen Mannigfaltigkeiten sind identisch mit den n -dimensionalen, n -stufig zusammenhängenden kompakten Räumen.*

4. Beweis der Theoreme.

Wir beweisen nunmehr die drei angeführten äquivalenten Theoreme und damit auch die eben erwähnte Behauptung über die Cantorschen Mannigfaltigkeiten, indem wir zeigen, daß jeder kompakte n -dimensionale Raum eine n -dimensionale Cantorsche Mannigfaltigkeit enthält.

Wir erinnern zunächst an folgendes dimensionstheoretische Theorem:

¹⁶⁾ Loc. cit. ¹⁵⁾.

^{16a)} Loc. cit. ¹⁵⁾.

¹⁷⁾ Loc. cit. ¹⁵⁾.

Ist R ein höchstens n -dimensionaler kompakter metrischer Raum, so ist R für jedes $\varepsilon > 0$ Summe von endlich vielen abgeschlossenen Mengen mit Durchmessern $< \varepsilon$, die zu je $n + 2$ fremd sind¹⁸⁾, aber nicht für jedes ε Summe von endlich vielen abgeschlossenen Mengen $< \varepsilon$, welche zu je $n + 1$ fremd sind¹⁹⁾. Dieser Satz rechtfertigt die folgende Begriffsbildung²⁰⁾. Ein kompakter Raum R heiße *n -dimensional vom Grade r* , wenn r die kleinste reelle Zahl ist von der Art, daß es zu jeder reellen Zahl $r' > r$ eine Zerlegung von R gibt in endlich viele abgeschlossene Teilmengen mit Durchmessern $< r'$, die zu je $n + 1$ fremd sind²¹⁾. Es ist in dieser Ausdrucksweise ein kompakter Raum dann und nur dann n -dimensional, wenn er vom Grade Null ($n + 1$)-dimensional, aber von einem positiven Grade n -dimensional ist.

Wir sehen zunächst: *Die Eigenschaft eines kompakten Raumes, n -dimensional von einem Grade $\geq r$ zu sein, ist eine induzible Eigenschaft*, d. h.: Ist n eine natürliche Zahl und r eine reelle Zahl > 0 und ist $\{A_k\}$ ($k = 1, 2, \dots$) eine monoton abnehmende Folge von abgeschlossenen Teilmengen eines kompakten Raumes, von denen jede n -dimensional von einem Grade $\geq r$ ist, dann ist auch der Durchschnitt $A = \bigcap_{k=1}^{\infty} A_k$ n -dimensional von einem Grade $\geq r$.

Andernfalls könnte ja A als Summe von endlich vielen abgeschlossenen Mengen A_1, A_2, \dots, A_m dargestellt werden, deren Durchmesser $< r$ und die zu je $n + 1$ fremd sind. Wir bestimmen m offene Mengen U_1, U_2, \dots, U_m , so daß $U_i \supset A_i$ ($i = 1, 2, \dots, m$) und daß die abgeschlossenen Hüllen \bar{U}_i zu je $n + 1$ fremd sind²²⁾. A ist in der offenen Menge $\bigcup_{i=1}^m U_i$ enthalten und

¹⁸⁾ Vgl. Menger, Monatsh. f. Math. u. Phys. 34, S. 153; Urysohn, Fund. Math. 8, S. 292.

¹⁹⁾ Vgl. Urysohn, Fund. Math. 8, S. 294.

²⁰⁾ Unter dem Namen coefficient d'appaplisement wurde dieselbe von Urysohn, Fund. Math. 8, S. 353, eingeführt.

²¹⁾ Die leere Menge wird als Summand zugelassen, so daß die Zahl r sicher nicht größer als der Durchmesser des Raumes R ist.

²²⁾ Zur Herstellung dieser Mengen U_i kann man folgendermaßen verfahren: Man bilde aus den m gegebenen Mengen A_i alle möglichen Durchschnitte — dies seien etwa die Mengen P_1, P_2, \dots, P_r ($r = 2^m - 1$) —. Man betrachte sodann die Menge A_1 und bezeichne mit S_1 die Summe aller derjenigen Durchschnitte P_i , die zu A_1 fremd sind. Auf Grund der Normalität des Raumes läßt sich eine offene Menge U_1 angeben mit den Eigenschaften $A_1 \subset U_1$; $\bar{U}_1 \cdot S_1 = 0$. Man ersetze jetzt in dem System $\{A_i\}$ die Menge A_1 durch \bar{U}_1 und verfähre mit der Menge A_2 in bezug auf dieses neue System genau in derselben Weise, wie soeben mit der Menge A_1 in bezug auf das ursprüngliche System. So setze man das Verfahren fort; nach m Schritten ist man zu Ende und hat ein System offener Mengen U_1, U_2, \dots, U_m , so daß irgendwelche r ($r < m$) Mengen \bar{U}_i dann und nur dann einen gemeinsamen Punkt haben, wenn dies für die entsprechenden A_i der Fall ist, woraus sich sofort ergibt, daß die U_i den oben formulierten Bedingungen genügen.

infolgedessen ist für hinreichend große Werte von k ²³⁾ $A_n \subset \sum_{i=1}^m U_i$ und somit $A_k = \sum_{i=1}^m A_k \cdot \bar{U}_i$. In dieser Summe ist jeder Summand abgeschlossen und hat einen Durchmesser $< r$; je $n+1$ Summanden sind fremd. Also ist A_k n -dimensional von einem Grad $< r$ im Widerspruch zur Voraussetzung. Damit ist die Behauptung bewiesen ^{23a)}.

Sind A und B zwei Mengen, deren jede n -dimensional von einem Grad $\leq r$ ist, so ist ihre Summe $A+B$ nicht notwendig n -dimensional von einem Grad $\leq r$, auch dann nicht, wenn beide Summanden abgeschlossen sind. Sei beispielsweise A die Summe der abgeschlossenen linearen Intervalle $[0, \frac{1}{4}]$ und $[\frac{1}{2}, \frac{3}{4}]$ und sei B die Summe der abgeschlossenen linearen Intervalle $[\frac{1}{4}, \frac{1}{2}]$ und $[\frac{3}{4}, 1]$. Jede der beiden Mengen A und B ist eindimensional vom Grade $\frac{1}{4}$; ihre Summe, das abgeschlossene Intervall $[0, 1]$, ist eindimensional vom Grade 1.

Es gilt jedoch, wie einer von uns im Anhang zu diesem Aufsatz beweist, der folgende

Additionssatz. Sind A und A' zwei abgeschlossene Mengen eines kompakten Raumes, deren jede n -dimensional von einem Grad $\leq r$ und deren Durchschnitt höchstens $(n-2)$ -dimensional ist, dann ist auch die Summe $A+A'$ n -dimensional von einem Grade $\leq r$.

Eine unmittelbare Folgerung aus dem Additionssatz lautet: *Ist der n -dimensionale kompakte Raum R keine Cantorsche Mannigfaltigkeit, dann enthält er eine echte abgeschlossene Teilmenge, welche von demselben Grade n -dimensional ist, wie der Gesamtraum R .* In der Tat, wenn R keine Cantorsche Mannigfaltigkeit ist, so gilt ja $R = A+B$, wo A und B abgeschlossene echte Teilmengen von R mit höchstens $(n-2)$ -dimensionalem Durchschnitt sind.

Sei nun ein kompakter n -dimensionaler Raum R gegeben. Er ist von einem positiven Grade n -dimensional, etwa vom Grade $r > 0$. Wie oben bewiesen wurde, ist die Eigenschaft eines Raumes, n -dimensional von einem Grade $\geq r$ zu sein, induzibel; also enthält R nach dem Brouwerschen Reduktionstheorem eine abgeschlossene Teilmenge A , welche hinsichtlich der erwähnten Eigenschaft *irreduzibel* ist, d. h. eine abgeschlossene Teilmenge A , welche n -dimensional vom Grade r ist, aber keine echte

²³⁾ Der Durchschnitt der abgeschlossenen Mengen $B_i = A_i = (R - \sum_{k=1}^m U_k)$ ($i=1, 2, \dots$)

(unter R den Raum verstanden) ist nämlich leer, woraus nach dem Cantorschen Durchschnittssatze folgt, daß für hinreichend große Werte von k $B_k = 0$ gilt, wofür auch geschrieben werden kann $A_k \subset U_k$.

^{23a)} Vgl. zu diesem Beweis die sub ¹⁹⁾ zitierte Stelle bei Urysohn.

abgeschlossene Teilmenge enthält, die n -dimensional vom Grade r ist. Nach der erwähnten Folgerung aus dem Additionssatz ist A eine n -dimensionale Cantorsche Mannigfaltigkeit. Damit ist die Behauptung bewiesen: Jeder kompakte n -dimensionale Raum enthält eine n -dimensionale Cantorsche Mannigfaltigkeit.

(Eingegangen am 30. 7. 1927.)

Anhang.

Ein Additionssatz ²⁴⁾.

Von Witold Hurewicz.

Ehe wir an den eigentlichen Beweis des oben angeführten Additionssatzes herangehen, wollen wir einige vorbereitende Sätze beweisen, die auch an sich ein gewisses Interesse zu bieten scheinen.

Satz 1^{24a)}. *Voraussetzung: Es sei in einem separablen Raum R eine n -dimensionale Menge M gegeben und ferner m offene Mengen U_1, U_2, \dots, U_m (die Zahl m ist dabei willkürlich), so daß $M \subset \sum_{k=1}^m U_k$ ist.*

Behauptung: Es gibt m offene Mengen V_1, V_2, \dots, V_m , so daß

$$(1) V_i \subset U_i \quad (i = 1, 2, \dots, m),$$

$$(2) M \subset \sum_{i=1}^m V_i,$$

(3) je $n + 2$ von den Mengen V_i einen leeren Durchschnitt haben.

Betrachten wir zunächst den Fall $n = 0$. Zu jedem Punkt von M existiert dann innerhalb einer der Mengen U_i eine Umgebung mit zu M fremder Begrenzung. Mit Rücksicht auf die Separabilität des Raumes ist M in der Summe von abzählbar vielen solchen Umgebungen, etwa von $W_1, W_2, \dots, W_k, \dots$ enthalten. Wir setzen nun

$$W'_1 = W_1,$$

$$W'_k = W_k - W_k \cdot \sum_{i=1}^{k-1} \overline{W_i} \quad (k = 1, 2, \dots).$$

²⁴⁾ Dieser Satz wurde von mir aufgestellt bei meinen Untersuchungen über stetige Abbildungen von Punktmengen; vgl. Proc. Ac. Amsterdam 30 (1926), S. 164, wo zwar nicht der Additionssatz selbst, jedoch eine unmittelbare Folgerung aus ihm formuliert wurde.

^{24a)} Dieser Satz findet sich (samt Beweis) in meiner Arbeit in Proc. Ac. Amsterdam 30, S. 426.

Die Mengen W'_k sind offen, paarweise fremd und ihre Summe enthält M . Ferner ist jedes W'_k in einer der m Mengen U_i enthalten. Verstehen wir nun unter V_i ($i = 1, 2, \dots, m$) die Summe aller derjenigen W'_k , die in V_i , aber in keiner der Mengen V_1, \dots, V_{i-1} enthalten sind, dann genügen diese m Mengen V_i den Forderungen des Satzes.

Sei nun $n > 0$ ²⁵⁾ und sei die im Satz 1 gegebene Menge M n -dimensional. Wir zerspalten M in $n + 1$ nulldimensionale Mengen M_1, M_2, \dots, M_{n+1} ²⁶⁾. Nach dem für $n = 0$ bewiesenen Satze können wir $n + 1$ Systeme von je m offenen Mengen angeben

$$\begin{array}{ccccccc} V_1^1, & V_2^1, & \dots, & V_m^1, & & & \\ V_1^2, & V_2^2, & \dots, & V_m^2, & & & \\ \dots & \dots & \dots & \dots & \dots & \dots & \\ V_1^{n+1}, & V_2^{n+1}, & \dots, & V_m^{n+1}, & & & \end{array}$$

so daß jede Menge der i -ten Kolonne in U_i liegt, daß M_i in der Summe der m Mengen der i -ten Zeile enthalten ist und daß je zwei Mengen derselben Zeile fremd sind. Es sind dann je $n + 2$ unter den $m \cdot (n + 1)$ Mengen V_i^k fremd, da unter ihnen mindestens zwei Mengen aus einer und derselben Zeile vorkommen. Ist V_i Summe der $n + 1$ Mengen der i -ten Kolonne, so haben offenbar auch je $n + 2$ Mengen V_i einen leeren Durchschnitt und befriedigen daher die Forderungen des Satzes 1.

Eine für das folgende bequemere Formulierung von Satz 1 ist

Satz 1'. *Voraussetzung: Es sei in einem separablen Raum R eine abgeschlossene n -dimensionale Menge M und ein System von m offenen Mengen U_1, U_2, \dots, U_m gegeben, so daß $R = \sum_{i=1}^m U_i$.*

Behauptung: Es gibt ein System von m offenen Mengen W_1, W_2, \dots, W_m , welches den folgenden Bedingungen genügt:

$$(1') \quad W_i \subset U_i \text{ }^{26a)} \quad (i = 1, 2, \dots, m),$$

$$(2') \quad R = \sum_{i=1}^m W_i,$$

(3') jeder Punkt von M gehört höchstens $n + 1$ unter den abgeschlossenen Mengen \overline{W}_i an.

Zunächst definieren wir in bezug auf die Menge M und die m (offenen) Mengen U_i ein System von offenen Mengen V_i , so daß die Forderungen

²⁵⁾ Diesen Teil des Beweises verdanke ich in der hier angegebenen Form im wesentlichen einer mündlichen Mitteilung des Herrn Vietoris.

²⁶⁾ Vgl. Hurewicz, Normalbereiche und Dimensionstheorie, Math. Annalen 96, S. 761; Tumarkin, Zur allgemeinen Dimensionstheorie, Proc. Ac. Amsterdam 28, S. 996.

^{26a)} Wir schreiben mit Menger $W \subset U$ für $\overline{W} \subset U$.

von Satz 1 erfüllt sind. Sodann setzen wir $V_i' = V_i + U_i \cdot (R - M)$. Es ist dann $R = \sum_{i=1}^m V_i'$. Die Mengen W_i bestimmen wir nun so²⁷⁾, daß

1. $W_i \subseteq V_i$ ($i = 1, 2, \dots, m$),
2. $R = \sum_{i=1}^m W_i$.

Dann genügen die Mengen W_i den Forderungen von Satz 1'.

Die Sätze 1 und 1' enthalten den bekannten dimensionstheoretischen Zerlegungssatz von Menger und Urysohn, demzufolge ein kompakter n -dimensionaler Raum für jedes $\varepsilon > 0$ Summe von endlich vielen abgeschlossenen Mengen, deren Durchmesser $< \varepsilon$ und die zu je $n + 2$ fremd sind.

Satz 2. Voraussetzungen. Sei R ein separabler Raum, \mathfrak{S} ein System bestehend aus endlich vielen abgeschlossenen Mengen. Es seien ferner m offene Mengen U_1, U_2, \dots, U_m gegeben, so daß $R = \sum_{k=1}^m U_k$ ist.

Behauptung. Es gibt m abgeschlossene Mengen A_1, A_2, \dots, A_m mit folgenden Eigenschaften:

1. $A_k \subset U_k$ ($k = 1, 2, \dots, m$).
2. $R = \sum_{k=1}^m A_k$.

3. Bezeichnet Π_r den Durchschnitt aus irgend r Mengen A_r und S eine beliebige Menge des Systems \mathfrak{S} , dann gilt

$$\dim S \cdot \Pi_r \leq \dim S - (r - 1),$$

wobei unter Mengen mit einer negativen Dimension stets die leere Menge zu verstehen ist²⁸⁾.

²⁷⁾ Es gilt nämlich allgemein: Ist der metrische Raum R als Summe von endlich vielen offenen Mengen dargestellt: $R = \sum_{i=1}^m U_i$, so gibt es eine Summendarstellung: $R = \sum_{i=1}^m U_i^*$, wo U_i^* wieder offene Mengen sind mit der Eigenschaft: $U_i^* \subseteq U_i$. In der Tat, die abgeschlossene Menge $R - \sum_{i=2}^m U_i$ ist in der offenen Menge U_1 enthalten; folglich gibt es mit Rücksicht auf Normalität von metrischen Räumen eine offene Menge U_1^* , so daß $R - \sum_{i=2}^m U_i \subset U_1^*$ so wie $\subseteq U_1$. Für den ersten Teil dieser Ungleichung kann man auch schreiben: $R = U_1^* + \sum_{i=2}^m U_i$. Wenn wir jetzt das System der offenen Mengen $U_1^*, U_2, U_3, \dots, U_m$ betrachten, auf die Menge U_2 dasselbe Verfahren anzuwenden, wie soeben auf die Menge U_1 usw., so gelangen wir schließlich zu einem System der Mengen U_i^* mit den verlangten Eigenschaften.

²⁸⁾ Für den Spezialfall, daß das System \mathfrak{S} nur aus ein- und nulldimensionalen Mengen besteht, wurde der Satz hinsichtlich kompakter Räume auch von Menger implizite verwendet. Vgl. Proc. Ac. Amst. 29 (1920), S. 481.

Wir verwenden im folgenden übrigens bloß zwei Spezialfälle dieses Satzes: *Erstens* den Fall, daß das System genau eine Menge enthält. In diesem Fall ist hinsichtlich kompakter Räume die Aussage des Satzes äquivalent mit dem *Mengerschen Zerlegungstheorem*, wonach ein kompakter n -dimensionaler Raum als Summe von beliebig kleinen abgeschlossenen Teilen dargestellt werden kann, von denen je m ($m < n + 2$) einen $(n - m + 1)$ -dimensionalen Durchschnitt haben²⁹⁾. *Zweitens* die Behauptung, daß es unter den Voraussetzungen von Satz 2 ein System abgeschlossener Mengen gibt, so daß alle Durchschnitte von der Form $S \cdot \Pi_r$, wo $r \geq \dim s + 2$, leer sind.

Ferner werden wir im folgenden den Satz 2 lediglich für *kompakte* Räume anwenden, doch beweisen wir den Satz für beliebige separable Räume. Die zum Beweise verwendete Methode macht übrigens diese Verallgemeinerung der Behauptung, wie wir sehen werden, notwendig.

Es sei n die größte unter den Dimensionen der Mengen des Systems \mathfrak{S} . Wir können annehmen, daß nur *eine* Menge des Systems \mathfrak{S} die Dimension n besitzt, denn gäbe es mehrere solcher Mengen, dann tilgen wir sie aus dem System \mathfrak{S} und nehmen statt dessen ihre (ebenfalls n -dimensionale) Summe in das System auf. Wenn die Behauptung in bezug auf dieses neue System gilt, dann gilt sie erst recht in bezug auf das ursprüngliche.

Sei M die einzige n -dimensionale Menge des Systems \mathfrak{S} . Wenn $n = 0$ ist, dann besteht das System \mathfrak{S} nur aus der Menge M und die Behauptung des Satzes reduziert sich auf den Spezialfall des Satzes 1' für $n = 0$.

Nehmen wir an, die Behauptung sei bewiesen für $n - 1$. Wir beweisen sie jetzt für n . Wir bestimmen nach Satz 1' ein System offener Mengen V_1, V_2, \dots, V_m gemäß den folgenden Bedingungen:

$$a') V_i \subset U_i \quad (i = 1, 2, \dots, m).$$

$$b') R = \sum_{i=1}^m V_i.$$

c') Jeder Punkt von M gehört höchstens $n + 1$ von den Mengen \bar{V}_i an.

Wir zerlegen nun die Menge M in einen $(n - 1)$ -dimensionalen Teil M' und einen nulldimensionalen Teil M'' .^{29a)} Es sind dabei die folgenden Annahmen gestattet:

$\alpha)$ M'' ist zu den von M verschiedenen Mengen des Systems \mathfrak{S} fremd. Denn sei C die Summe aller von M verschiedenen Mengen aus \mathfrak{S} . Wäre

²⁹⁾ Vgl. Menger, Jahresbericht d. Deutsch. Math. Ver. 36, S. 10 des Anhanges.

^{29a)} Jede n -dimensionale separable Menge ist Summe einer $(n - 1)$ -dimensionalen und einer 0-dimensionalen Menge. Vgl. meine sub 1) zitierte Abhandlung, S. 754, Theorem I.

M'' zu C nicht fremd, dann würden wir die Zerlegung $M = M' + M''$ durch die folgende ersetzen:

$$M = (M' + M \cdot C) + (M'' - M'' \cdot C).$$

Auch hier ist der erste Summand $(n-1)$ -dimensional, als Summe von der $(n-1)$ -dimensionalen Menge M' und der abgeschlossenen höchstens $(n-1)$ -dimensionalen Menge $M \cdot C$.³⁰⁾ Der zweite Summand ist nulldimensional und zu C fremd.

$\beta)$ M'' enthält keinen offenen Teil. Denn wäre der offene Kern K von M'' nicht leer, so würden wir die Zerlegung $M = M' + M''$ ersetzen durch die folgende

$$M = (M' + K) + (M'' - K).$$

Auch hier ist der erste Summand als Summe der $(n-1)$ -dimensionalen Menge M' und der offenen nulldimensionalen Menge K höchstens $(n-1)$ -dimensional³⁰⁾, während der zweite Summand keinen offenen Teil enthält.

Wir machen also die beiden Annahmen $\alpha)$ und $\beta)$ und setzen $R' = R - M''$. Wir betrachten sodann das System \mathfrak{S}' , das aus \mathfrak{S} entsteht, wenn die Menge M durch die Menge M' ersetzt wird. Alle Mengen des Systems \mathfrak{S}' sind in R' abgeschlossen und höchstens $(n-1)$ -dimensional. Wir können also unseren für $n-1$ als bewiesen angenommenen Satz anwenden und ein System von m in R' abgeschlossene Mengen A'_1, A'_2, \dots, A'_m bilden, so daß

$$1'. A'_i \subset V_i \quad (i = 1, 2, \dots, m).$$

$$2'. R' = \sum_{i=0}^m A'_i.$$

3'. Bezeichnet Π'_r den Durchschnitt von irgend r Mengen A'_i und S eine beliebige Menge des Systems \mathfrak{S}' , dann ist

$$\dim S' \cdot \Pi'_r \leq \dim S' - (r-1).$$

Setzen wir nun $A_i = \bar{A}'_i$ ($i = 1, 2, \dots, m$), dann genügen, wie wir nun zeigen wollen, die m abgeschlossenen Mengen A_i den Forderungen des Satzes hinsichtlich des Systems \mathfrak{S} .

1. Es war $A'_i \subset V_i \subset U_i$, folglich ist $A_i = \bar{A}'_i \subset U_i$.

2. Da M'' nach Voraussetzung $\beta)$ keinen offener Teil enthält, ist

$R' = R - M''$ in R dicht. Folglich ist $\sum_{i=1}^m A_i = \overline{\sum_{i=1}^m A'_i} = \bar{R}' = R$.

3. Für jedes i ist A_i Summe von A'_i und einer Teilmenge von M'' . Folglich kann jeder Durchschnitt Π_r in der Form dargestellt werden

$$(*) \quad \Pi_r = \Pi'_r + N,$$

wo Π'_r einen Durchschnitt von r Mengen A'_i und N eine Teilmenge von M'' bedeutet. Ist also S eine von M verschiedene Menge des

³⁰⁾ Vgl. Hurewicz, Math. Annalen 96, S. 761, Satz 37.

Systems \mathfrak{S} , dann gilt mit Rücksicht auf Voraussetzung α)

$$S \cdot \Pi_r = S \cdot \Pi'_r + N \cdot S = S \cdot \Pi'_r.$$

Folglich ist nach Voraussetzung β'

$$\dim S \cdot \Pi_r = \dim S \cdot \Pi'_r \leq \dim S - (r - 1),$$

wie der Satz es fordert.

Wir haben also nur noch Durchschnitte der Form $\Pi_r \cdot M$ zu betrachten.

Wir unterscheiden zwei Fälle:

a) $r \geq n + 2$. In diesem Fall ist Π_r nach Voraussetzung $1'$ im Durchschnitt von $n + 2$ Mengen V_i enthalten. Folglich ist mit Rücksicht auf die Annahme c') der Durchschnitt $\Pi'_r M$ leer, d. h. höchstens $(n - r + 1)$ -dimensional im Einklang mit der Bedingung β des zu beweisenden Satzes.

b) $r < n + 2$. Nach der Formel (*) haben wir (unter Berücksichtigung von $N \subset M''$)

$$(\dagger) \quad \Pi_r M = (\Pi'_r + N) \cdot (M' + M'') = \Pi'_r M' + P,$$

wo P eine Teilmenge von M'' und daher nulldimensional ist. Nach Annahme β' ist die Menge $\Pi'_r M'$ höchstens $(n - 1) - (r - 1) = (n - r)$ -dimensional, also ist $\Pi_r \cdot M$ als Summe einer höchstens $(n - r)$ -dimensionalen und einer nulldimensionalen Menge höchstens $(n - r + 1)$ -dimensional, — wiederum im Einklang mit den Forderungen des Satzes.

Damit ist Satz 2 bewiesen.

Additionssatz. Sind A und A' zwei abgeschlossene Mengen, deren jede von einem Grade $< r$ n -dimensional und deren Durchschnitt höchstens $(n - 2)$ -dimensional ist, dann ist auch $A + A'$ von einem Grade $< r$ n -dimensional.

Nach Voraussetzung und nach Fußnote ²²⁾ kann A mit endlich vielen in A offenen Mengen U_1, U_2, \dots, U_m und A' mit endlich vielen in A' offenen Mengen U'_1, U'_2, \dots, U'_m , überdeckt werden, so daß die Durchmesser der $m + m'$ Mengen U_i und $U'_i < r$ sind und daß je $n + 1$ von den Mengen U_i und je $n + 1$ von den Mengen U'_i fremd sind. Wir wenden nun Satz 2 an, wobei wir als Raum die Menge A und als System \mathfrak{S} das aus der einzigen Menge $A \cdot A'$ bestehende System wählen. Wir bestimmen nach Satz 2 (der, wie oben bemerkt, sich in diesem Falle im wesentlichen auf den Mengerschen Zerlegungssatz reduziert) die abgeschlossenen Mengen F_1, F_2, \dots, F_m so, daß gilt:

$$1. \quad F_i \subset U_i \quad (i = 1, 2, \dots, m).$$

$$2. \quad A = \sum_{i=1}^m F_i.$$

3. Ist Π_r Durchschnitt von irgend r Mengen F_i , so gilt:

$$(*) \quad \begin{aligned} \dim A' \cdot \Pi_r &= \dim A \cdot A' \cdot \Pi_r \leq \dim A \cdot A' - (r - 1) \\ &\leq (n - 2) - (r - 1) = n - r - 1. \end{aligned}$$

Wir betrachten nun das System aller Mengen $A' \cdot \Pi_r$, wo Π_r alle Durchschnitte von irgend r Mengen F_i durchläuft. Dieses System nennen wir \mathfrak{S} und wenden Satz 2 auf die Menge A' als Raum, auf die offenen Mengen U_1, U_2, \dots, U_m und das System \mathfrak{S} an. Wir können dann m' abgeschlossene Mengen F'_1, F'_2, \dots, F'_m bestimmen, so daß gilt:

$$1'. F'_i \subset U_i \quad (i = 1, 2, \dots, m').$$

$$2'. A' = \sum_{i=1}^{m'} F'_i.$$

3'. Ist Π_r' Durchschnitt von irgend r Mengen A'_i und M eine beliebige Menge des Systems \mathfrak{S} , so gilt

$$\dim M \cdot \Pi_r' \leq \dim M - (r - 1).$$

Es ist dann

$$A + A' = \sum_{i=1}^m F_i + \sum_{i=1}^{m'} F'_i,$$

und die Durchmesser der $m + m'$ abgeschlossenen Mengen F_i und F'_i sind durchwegs $< r$. Unsere Behauptung ist also bewiesen, wenn wir zeigen, daß je $n + 1$ von den $m + m'$ Mengen F_i und F'_i einen leeren Durchschnitt haben.

Dies ist mit Rücksicht auf die Wahl der F_i bzw. F'_i als Teilmengen der U_i bzw. U'_i sicher der Fall für je $n + 1$ Mengen, die entweder durchwegs Mengen F_i oder durchwegs Mengen F'_i sind.

Wir haben noch nachzuweisen, daß alle Mengen

$$\Pi_r \cdot \Pi'_s, \quad \text{wo } r \geq 1, s \geq 1, r + s = n + 1 \text{ gilt,}$$

leer sind. Es ist $\Pi_r \cdot \Pi'_s = \Pi_r \cdot A' \cdot \Pi'_s$. Da die Mengen $\Pi_r \cdot A'$ zum System \mathfrak{S} gehören, so ist mit Rücksicht auf die Wahl der Mengen F'_i

$$\dim \Pi_r \cdot \Pi'_s \leq \dim \Pi_r \cdot A' - (s - 1).$$

Kombinieren wir diese Ungleichung mit der Ungleichung (*), so erhalten wir

$$\dim \Pi_r \cdot \Pi'_s \leq (n - r - 1) - (s - 1) = n - (r + s) = -1,$$

d. h. $\Pi_r \cdot \Pi'_s$ ist leer.

Damit ist der Additionssatz bewiesen³¹⁾.

³¹⁾ (Zusatz bei der Korrektur): Es gilt, wie der angeführte Beweis zeigt, noch der folgende allgemeinere Additionssatz, dessen Inhalt im Falle kompakter Räume allerdings nicht über den des im Text angegebenen Satzes hinausgeht:

Es seien in einem separablen Raum zwei abgeschlossene Mengen A und A' mit $(n - 2)$ -dimensionalem Durchschnitt gegeben; es liege ferner ein endliches System offener Mengen U_1, U_2, \dots, U_m vor, und es sei jede der Mengen A und A' als Summe von endlich vielen abgeschlossenen Mengen darstellbar, die zu je $n + 1$ fremd sind und deren jede in einer der Mengen U_i enthalten ist. Behauptet wird, daß eine analoge Summendarstellung auch für die Menge $A + A'$ gilt.

Fundamentalsatz von den Stützen eines Eigebietes.

Von

Hermann Brunn in München.

Daß durch jeden Begrenzungspunkt eines Eigebietes mindestens ein Stützhauptlinear¹⁾ des Raumes geht, in dem das Eigebiet liegt, ist zuerst von Minkowski bewiesen worden²⁾. Ich habe dann einen kürzeren und anschaulichen Beweis zu geben versucht³⁾. Und schließlich hat Carathéodory, eine Bemerkung Minkowskis verwertend, einen durch Kürze und Allgemeinheit sich auszeichnenden analytisch formulierten Beweis erbracht⁴⁾.

Bei der grundlegenden Natur des Satzes ist es vielleicht nicht ohne Interesse, noch einen anderen Beweis kennen zu lernen, der sich bestrebt, so elementar und von Grenzbetrachtungen so unabhängig wie möglich zu sein.

Zunächst darf ich noch einmal auf den Beweis im Archiv der Mathematik zurückkommen, um zu zeigen, daß er von der dort verwendeten, für den Präzisionsmathematiker nicht ganz einwandfreien Hilfsvorstellung der Bewegung (eines unbelegten Strahls in eine gewisse Grenzlage) leicht frei gemacht werden kann.

Wenn r ein Rand-, i ein innerer Punkt des Ovals \mathcal{E} ist, zerfällt die Gerade $ri \equiv V$ durch r in einen „belegten“ Halbstrahl V' (durch i) und einen „unbelegten“ (der außer r keinen \mathcal{E} -Punkt enthält) V'' . Der zu V senkrechte Strahl durch r sei als eine erste Halbiegerade (hier von gestreckten Winkeln $V'V''$) mit H_1 bezeichnet. Sind seine Halbstrahlen H_1', H_1''

¹⁾ Hauptlinear eines Raumes heiße das durch eine lineare Gleichung zwischen den n Koordinaten aus ihm herausgehobene $(n-1)$ fach ausgedehnte Gebilde. Es, wie oft geschieht, allgemein als „Ebene“ seines Raumes zu bezeichnen, will mir ebenso wenig behagen, als man die durch zwei lineare Gleichungen bestimmten $(n-2)$ fach ausgedehnten Gebilde allgemein als „Gerade“ ihres Raumes oder die Gerade als Ebene des zweifach ausgedehnten Raumes wird bezeichnen wollen.

²⁾ Geom. d. Zahlen; der Beweis ist auf den Seiten 13 bis 35 verstreut.

³⁾ Arch. d. Math. u. Phys. (3) 17 (1910), S. 289 bis 292.

⁴⁾ Rend. d. Circ. Mat. di Palermo 32 (1911) im Aufsatz „Über den Variabilitätsbereich der Fourierschen Konstanten usw.“

beide belegt oder beide unbelegt, so erweist sich leicht schon H_1 als die gewünschte Stütze. Ist nur einer belegt, so schließt er mit der belegten Hälfte von V einen von lauter belegten Halbstrahlen ausgefüllten oder „völlig belegten“, der andre mit der unbelegten Hälfte von V einen völlig unbelegten rechten Winkel ein. Die beiden von Halbstrahlen verschiedenen Belegungscharakters eingeschlossenen rechten (Scheitel-)Winkel dagegen bleiben vorderhand noch zweifelhaften Charakters. Wir halbieren sie durch den Strahl H_2 und wiederholen unsere Schlüsse. H_2 erweist sich bereits als die gewünschte Stütze, wenn H_2' und H_2'' gleichen Belegungscharakter zeigen; ist das nicht der Fall, so ist wieder leicht zu sehen, daß der durch die belegten (unbelegten) Hälften von H_1, H_2 eingeschlossene Winkel $\frac{\pi}{4}$ völlig belegt (unbelegt) ist und sich mit dem belegten (unbelegten) rechten Winkel von vorhin zu einem belegten (unbelegten) Winkel von der Größe $\frac{\pi}{2} + \frac{\pi}{4} = \frac{3}{4}\pi$ zusammenschließt, während zwei Scheitelwinkel von der Größe $\frac{\pi}{4}$ zunächst noch unbestimmten Charakters bleiben. So fährt man fort; nach n Halbierungen sind der als völlig belegt nachgewiesene Winkel und sein unbelegter Scheitelwinkel auf

$$\frac{\pi}{2} + \frac{\pi}{2^2} \dots + \frac{\pi}{2^n} = \frac{2^n - 1}{2^n} \pi$$

angewachsen, die sie trennenden unbestimmten Scheitelwinkel auf $\frac{\pi}{2^n}$ herabgesunken, und wenn nicht schon nach einer endlichen Anzahl von Schritten in H_n , so findet sich nach einer unendlichen in dem Strahl H_∞ , dem die H_n als ihrer Grenze sich annähern, offenbar die gewünschte Stütze.

Bei der Verallgemeinerung des Satzes auf drei Dimensionen lassen sich mit dem Begriff der Belegung alle Belegungsschlüsse von den Halbstrahlen, die vom Randpunkt r ausgehen, übertragen auf die Halbebenen, die von einer streifenden Geraden r_1 ausgehen (vgl. den gedruckten Beweis im Archiv) usw.

Im folgenden sei die Oberfläche — d. h. die Gesamtheit der Begrenzungsunkte — eines mit großem deutschen Buchstaben bezeichneten Körpers immer mit dem entsprechenden großen lateinischen Buchstaben bezeichnet. Nun sei zur Darstellung des versprochenen neuen Beweises geschritten. Er wird hier für dreidimensionale Eikörper durchgeführt; seine Verallgemeinerung erhellt von selbst.

Der Beweis stützt sich, wenn wir es in gewöhnlicher Terminologie ausdrücken, auf die äußeren Parallelfächen zur Oberfläche des gegebenen Eikörpers \mathcal{E} . Wir wollen indessen, da man immer praktischer von den Eikörpern als von ihren Oberflächen ausgeht, zunächst nicht von den Parallelfächen, sondern von den von ihnen eingeschlossenen Körpern sprechen.

Wir wollen sie, da die Bezeichnung „Parallelkörper“ etwas Unlogisches haben würde, als Erweiterungskörper, genauer Kugelerweiterungskörper, Erweiterung des gegebenen Körpers durch Kugeln bezeichnen und dadurch auf ihre von uns bevorzugte Definition hinweisen.

Ein Erweiterungskörper \mathfrak{K}_ρ des Eikörpers \mathfrak{E} in unserem Sinne wird gebildet durch die Menge der Punkte, welche in allen Vollkugeln \mathfrak{K}_ρ mit festem Radius ρ enthalten sind, deren Mittelpunkte Punkte von \mathfrak{E} sind. \mathfrak{K}_ρ erweist sich unschwer ebenfalls als Eikörper.

Es ist klar, daß ein Punkt von \mathfrak{K}_ρ mindestens zu einer Erweiterungskugel \mathfrak{K}_ρ von \mathfrak{E} gehört, ist es ein Begrenzungspunkt r von \mathfrak{K}_ρ , so muß er natürlich auf K_ρ liegen, weil er als innerer Punkt von \mathfrak{K}_ρ auch ein innerer von \mathfrak{K}_ρ sein würde. Und r kann, in starkem Gegensatz zu einem i von \mathfrak{K}_ρ , nur zu einem einzigen \mathfrak{K}_ρ — nennen wir es \mathfrak{K}_ρ^r — gehören; denn wenn r zweien Kugeln \mathfrak{K}_ρ und \mathfrak{K}_ρ' angehörte, würde der Verbindungsstreckenkörper zwischen beiden Kugeln, ganz zu \mathfrak{K}_ρ gehörig, r in seinem Innern enthalten. Diese Begründung erhärtet zugleich, daß \mathfrak{K}_ρ^r überhaupt die einzige Kugel vom Radius ρ ist, die r enthält und ganz zu \mathfrak{K}_ρ gehört. Dies bedeutet insofern etwas mehr wie das, was eben vorher behauptet wurde, als für die \mathfrak{K}_ρ ja \mathfrak{E} -Punkte als Mittelpunkte vorausgesetzt werden und es durch das bisher Gesagte noch nicht sichergestellt wurde, ob in \mathfrak{K}_ρ nicht auch Kugeln vom Radius ρ liegen, deren Mittelpunkte nicht zu \mathfrak{E} gehören. Für andere als Eikörper ist das ja in der Tat möglich, z. B. beim Erweiterungskörper einer Hohlkugel, deren Innenraum einen Radius $\sigma < \rho$ hat.

Fürs folgende mag gemerkt werden, daß die Punkte von \mathfrak{K}_ρ eindeutig auf die Mittelpunkte m ihrer \mathfrak{K}_ρ abgebildet sind; umgekehrt gilt das nicht, wie an dem Eckpunkt eines konvexen Polyeders sofort zu sehen ist.

Nun erkennt man leicht, daß unser Stützensatz für jedes \mathfrak{K}_ρ gültig ist. Eine Stütze durch r an \mathfrak{K}_ρ müßte natürlich auch Stütze, somit Tangentialebene T von \mathfrak{K}_ρ^r in r sein. Auf der kugelfreien Seite von T kann nun in der Tat kein \mathfrak{K}_ρ -Punkt p liegen, denn der dann ganz zu \mathfrak{K}_ρ gehörige Verbindungsstreckenkörper zwischen p und \mathfrak{K}_ρ^r würde r in seinem Innern enthalten. T ist also Stütze von \mathfrak{K}_ρ durch r , und die Ebene $T' \parallel T$ durch den Mittelpunkt m von \mathfrak{K}_ρ^r erweist sich hiernach als Stütze von \mathfrak{E} durch m , m als Punkt von E . Läge nämlich ein \mathfrak{E} -Punkt q auf der T -Seite von T' , so würde das \mathfrak{K}_ρ mit Zentrum q auf die Seite von T hinüberreichen, die wir eben als frei von \mathfrak{E} -Punkten erwiesen haben.

Hiernit scheint nun der Beweis schon fertig, da man geneigt ist, die Menge der m mit der Menge der Punkte e auf E zu identifizieren. Aber wir müssen vorderhand noch als möglich annehmen, daß ein e kein m ist und daß \mathfrak{K}_ρ^e , worunter hier das \mathfrak{K}_ρ mit Mittelpunkt e zu verstehen ist, die

Fläche \mathfrak{P}_ϱ nicht erreicht, vielmehr ganz im Innern von \mathfrak{P}_ϱ liegt, wie das bei den Ausweitungskörpern nicht konvex begrenzter Körper in der Tat vorkommt.

Dann würde es eine Kugel \mathfrak{K}_ϱ^e geben mit einem Radius $\varrho' > \varrho$, die vollständig zu \mathfrak{P}_ϱ gehören würde. Wir könnten uns e als Schwerpunkt eines, sagen wir regulären Tetraeders mit den Ecken a_λ ($\lambda = 1, 2, 3, 4$) denken und \mathfrak{K}_ϱ^e in Richtung $e a_\lambda$ verschieben, bis K_ϱ^e zum ersten Mal einen Punkt von P_ϱ enthält; er heiße b_λ , die neue Lage der Kugel \mathfrak{K}_λ und ihr Mittelpunkt e_λ . Evident ist dann $\overline{e e_\lambda} \geq \varrho' - \varrho$, das Tetraeder $e_1 e_2 e_3 e_4$ ein körperliches und e ein Innenpunkt desselben. Da nun aber die \mathfrak{K}_λ , weil sie die b_λ enthalten und noch ganz zu \mathfrak{P}_ϱ gehören nach Früherem mit den \mathfrak{K}^{b_λ} identisch sein müssen, und somit die e_λ sicher \mathfrak{E} -Punkte sind, würde $e_1 e_2 e_3 e_4$ ein ganz zu \mathfrak{E} gehöriges Tetraeder und e ein innerer Punkt von \mathfrak{E} sein, während es doch ein Begrenzungspunkt ist. Somit kann \mathfrak{K}_ϱ^e nicht existieren; \mathfrak{K}_ϱ^e muß P_ϱ in mindestens einem Punkte erreichen, e ist ein m und damit ist für jeden Punkt auf E eine Stütze nachgewiesen.

Noch einen weiteren Beweis unseres Satzes will ich in seiner Anwendung auf dreifach ausgedehnte Eikörper \mathfrak{E} andeuten. Sein Grundgedanke ist ganz besonders einfach. Man kann den von r ausgehenden unbelegten Halbstrahl der Geraden $r i - r$ und i haben die alte Bedeutung — zur Achse eines veränderlichen Kreiskegelmantels C aus Halbstrahlen vom Scheitel r aus wählen und den Kegelwinkel von 0 aus so lang wachsen lassen, bis C in der Lage C_1 einen ersten belegten Halbstrahl H_1 enthält. Dann ist die Tangentialebene T an C_0 längs H_1 eine Stütze durch r an \mathfrak{E} . Denn nähme man auf der Kegelseite von T einen belegten Halbstrahl $H_2 = r e$ an, so würde der Winkel $H_1 H_2 < \pi$ lauter belegte Halbstrahlen enthalten und mit ihnen in das von \mathfrak{E} -Punkten doch freie Innere von C_0 eindringen müssen.

Leider wird eine doppelte Grenzbeachtung nötig, wenn nicht ein erstes belegtes, sondern letztes unbelegtes C , heiße es wieder C_0 , vorhanden ist. Eine der Erzeugenden von C_0 hat dann belegte Halbstrahlen in beliebiger Nähe (tangiert E), und ein genügend nah gewählter von ihnen kann die Rolle des früheren H_1 übernehmen. Das zu zeigen ist nicht schwer, aber bei vollständiger, strenger Darstellung etwas umständlich und wir gehen nicht darauf ein, da wir nach möglichst elementaren Beweisen fahndeten.

Januar 1928.

(Eingegangen am 28. I. 1928.)

Über die Erfüllbarkeit gewisser Zähl ausdrücke.

Von

Wilhelm Ackermann in Göttingen.

Die Herren Bernays und Schönfinkel haben in einer kürzlich erschienenen Abhandlung¹⁾ einige einfache Fälle der Entscheidungsprobleme für die Relationslogik behandelt und erledigt. Es wurden dabei logische Ausdrücke betrachtet, die außer den logischen Grundverknüpfungen „und“, „oder“, „nicht“, „wenn — so“, Prädikaten- und Relationsvariable und ferner All- und Seinszeichen für Individuen enthielten. Derartige Ausdrücke bezeichnet man nach Löwenheim auch als Zähl ausdrücke. Die Lösung des Entscheidungsproblems im Bereiche der Zähl ausdrücke bedeutet die Auf findung eines Verfahrens, durch das man imstande ist, festzustellen, ob ein vorgelegter Zähl ausdruck bei bestimmten Einsetzungen für die variablen Prädikate und Relationen stets eine richtige Aussage darstellt oder nicht. Die so definierte Allgemeingültigkeit eines Zähl ausdrucks kann noch von der Anzahl der Elemente des Individuenbereichs abhängen, für den man sich die Prädikate und Relationen definiert denkt. Die vollständigste Lösung des Entscheidungsproblems ist die, in der die Abhängigkeit der Allgemeingültigkeit von der Individuenanzahl genau angegeben ist.

Man kann sich bei den Zähl ausdrücken auf diejenigen beschränken, die die sogenannte „Normalform“ haben, d. h. bei denen die All- und Seinszeichen alle am Anfang der Formel stehen, und man wird die Zähl ausdrücke zweckmäßig nach der Art der zu ihnen gehörenden Kombinationen von All- und Seinszeichen klassifizieren.

In der Bernays-Schönfinkelschen Arbeit wurde die Frage nach der Allgemeingültigkeit für die folgenden Klassen von Formeln beantwortet:

¹⁾ P. Bernays und M. Schönfinkel, Zum Entscheidungsproblem der math. Logik, *Math. Annalen* 99 (1928).

Zunächst für diejenigen Formeln, bei denen jedes vorkommende Allzeichen jedem vorkommenden Existenzialzeichen vorangeht. Die allgemeinste Gestalt dieser Formeln ist

$$(1) \quad (x_1)(x_2) \dots (x_k)(E y_1) \dots (E y_l) \mathfrak{A}(x_1, \dots, x_k, y_1, \dots, y_l). \text{ } ^2)$$

Die Fälle, daß kein Existenzialzeichen oder kein Allzeichen vorkommt, sind dabei eingeschlossen. Dieser Fall (1) läßt sich sehr einfach behandeln.

Besondere Schwierigkeiten bieten die Fälle, in denen Existenzialzeichen auch vor Allzeichen stehen. Der einfachste Fall ist hier

$$(2) \quad (E x)(y) \mathfrak{A}(x, y).$$

Für diesen wurde in der genannten Arbeit ebenfalls ein Entscheidungsverfahren angegeben.

In der vorliegenden Arbeit will ich zunächst allgemein den Fall behandeln, daß auf ein Existenzialzeichen lauter Allzeichen folgen. Es handelt sich also um den Fall:

$$(3) \quad (E x)(y_1) \dots (y_l) \mathfrak{A}(x, y_1, \dots, y_l).$$

Weiter sollen überhaupt die Ausdrücke betrachtet werden, bei denen nur ein Existenzialzeichen vorkommt. Der allgemeinste Fall ist hier:

$$(4) \quad (x_1) \dots (x_k)(E y)(z_1) \dots (z_l) \mathfrak{A}(x_1, \dots, x_k, y, z_1, \dots, z_l).$$

Übrigens beschränke ich mich bei meinen Überlegungen der Einfachheit halber auf zweigliedrige Relationen. Die Ausdehnung auf mehrgliedrige Relationen macht aber keine sachlichen Schwierigkeiten.

§ 1.

Vorbereitungen und Angabe der Methode.

Statt die Allgemeingültigkeit von Zähl ausdrücken zu betrachten, kann man auch nach ihrer *Erfüllbarkeit* fragen. Ein Zähl ausdrück heißt *erfüllbar*, wenn es überhaupt eine bestimmte Einsetzung für die variablen Prädikate und Relationen gibt, so daß der Zähl ausdrück in eine richtige Behauptung übergeht. Die Allgemeingültigkeit eines Zähl ausdrucks ist gleichbedeutend mit der Nichterfüllbarkeit des entgegengesetzten, d. h. aus dem ursprünglichen durch Negation entstehenden Zähl ausdrucks. Für unsere Überlegungen ist es bequemer, uns mit der Erfüllbarkeit zu beschäftigen. Statt nach der Allgemeingültigkeit von Ausdrücken des Typs (3) und (4), fragen wir daher nach der Erfüllbarkeit von Ausdrücken der folgenden Form:

²⁾ Die Symbolik ist die gleiche wie in: D. Hilbert u. W. Ackermann, Grundzüge der theoretischen Logik (Springer 1928) [zitiert im folgenden als H.-A.].

$$(3') \quad (x)(E y_1) \dots (E y_l) \mathfrak{A}(x, y_1, \dots, y_l),$$

$$(4') \quad (E x_1) \dots (E x_k)(y)(E z_1) \dots (E z_l) \mathfrak{A}(x_1, \dots, x_k, y, z_1, \dots, z_l).$$

Die Erfüllbarkeit oder Nichterfüllbarkeit eines bestimmten Zähl-
ausdrucks ist natürlich ebenso eine Funktion der Anzahl der Elemente
im zugrunde gelegten Individuenbereich wie die Allgemeingültigkeit. Für
einen Bereich mit bestimmter endlicher Anzahl der Elemente ist nun die
Entscheidung über die Erfüllbarkeit eines Ausdrucks ohne weiteres zu
treffen. Man kann hier die Allzeichen und Existenzialzeichen durch end-
liche Konjunktionen bzw. Disjunktionen ersetzen, so daß sich das Problem
auf ein solches des Aussagenkalküls reduziert, das keine Schwierigkeiten
bietet³⁾.

Ich werde nun zeigen, daß für die Ausdrücke der Form (3') und (4')
der folgende Satz gilt:

*Ist ein derartiger Zähl Ausdruck überhaupt erfüllbar, so ist er auch
schon in einem Individuenbereich von bestimmter endlicher Anzahl der
Elemente erfüllbar.*

Mit der Angabe dieser zu einem bestimmten Ausdruck gehörigen
endlichen Anzahl ist das Entscheidungsverfahren gegeben. Diese Anzahl
sei für einen bestimmten Zähl Ausdruck z. B. k . Für einen Individuen-
bereich der endlichen Anzahl k läßt sich, wie eben angeführt, die Ent-
scheidung über die Erfüllbarkeit treffen. Entweder ist nun der betrachtete
Zähl Ausdruck für einen Bereich von k Individuen nicht erfüllbar, dann
gilt dasselbe für jeden Individuenbereich. Oder aber er ist für einen Be-
reich von k Individuen erfüllbar, dann gilt dasselbe für jeden Bereich, der
eine größere Anzahl von Individuen enthält⁴⁾. Man braucht dann nur
noch die Erfüllbarkeit des betreffenden Ausdrucks für Individuenbereiche
der Anzahlen $1, 2, \dots, k - 1$ zu untersuchen, was sich wieder auf ein
Problem des Aussagenkalküls reduziert.

Es kommt also darauf an, für jeden Zähl Ausdruck der Form (3')
oder (4') einen zugehörigen endlichen Individuenbereich anzugeben, für
den der Zähl Ausdruck erfüllbar sein muß, falls er nicht einen logischen
Widerspruch darstellt. Es sei übrigens bemerkt, daß sich die angegebene
Methode nicht für eine allgemeine Lösung des Entscheidungsproblems im
Bereich der Zähl Ausdrücke verwenden läßt, da es Zähl Ausdrücke gibt, die
für keinen endlichen Individuenbereich erfüllbar sind, wohl aber für den
abzählbar unendlichen⁵⁾.

³⁾ Vgl. H.-A., S. 79.

⁴⁾ Vgl. Bernays und Schönfinkel, S. 3.

⁵⁾ Vgl. H.-A., Seite 80.

§ 2.

Die Behandlung der Zählausdrücke von der Form $(x)(Ey)\mathfrak{A}(x, y)$.

Ich behandle in diesem Paragraphen den Bernays-Schönfinkelschen Fall noch einmal, da an ihm das Wesentliche der von mir benutzten Methode auch für die allgemeineren Fälle klarer werden wird. Ich will zunächst noch die weitere Einschränkung machen, daß die betrachteten Zählausdrücke nur eine einzige variable Relation F enthalten. $\mathfrak{A}(x, y)$ ist dann eine Aussagenverbindung, die aus den Grundaussagen $F(x, x)$, $F(x, y)$, $F(y, x)$, $F(y, y)$ zusammengesetzt ist. Um dies anzudeuten, schreibe ich den Zählausdruck als

$$(x)(Ey)\mathfrak{A}(F(x, x), F(x, y), F(y, x), F(y, y)).$$

Es sei dieser Ausdruck in irgendeinem Individuenbereich erfüllbar. $\Phi_0(x, y)$ möge eine zugehörige erfüllende Relation sein. Zu jedem Element x des Individuenbereichs kann man dann ein anderes y finden, so daß

$$\mathfrak{A}(\Phi_0(x, x), \Phi_0(x, y), \Phi_0(y, x), \Phi_0(y, y))$$

richtig ist. Wir denken uns für jedes x eins der zugehörigen y ausgewählt, und bezeichnen es mit $\alpha(x)$. Es gilt dann allgemein, d. h. für jedes x :

$$\mathfrak{A}(\Phi_0(x, x), \Phi_0(x, \alpha(x)), \Phi_0(\alpha(x), x), \Phi_0(\alpha(x), \alpha(x))).$$

Zur Abkürzung bezeichnen wir $\alpha(x)$ mit $\alpha_1(x)$, $\alpha\alpha(x)$ mit $\alpha_2(x)$, $\alpha\alpha\alpha(x)$ mit $\alpha_3(x)$ usw. x_0 möge irgendein willkürlich herausgegriffenes Element des Individuenbereichs bedeuten.

Wir nehmen nun einen neuen Individuenbereich, der aus den Elementen $0, 1, 2, \dots$ besteht. Für diesen Bereich definieren wir eine Relation $\Phi_1(x, y)$ wie folgt:

Für alle i und k soll sein:

$$\Phi_1(i, k) \sim \Phi_0(\alpha_i(x_0), \alpha_k(x_0))$$

($\alpha_0(x_0)$ soll hier x_0 selbst bezeichnen. Das Zeichen \sim bedeutet, daß die links und die rechts stehende Aussage den gleichen Wahrheitswert haben.)

Für den neuen Individuenbereich gilt offenbar:

$$(x)\mathfrak{A}(\Phi_1(x, x), \Phi_1(x, x+1), \Phi_1(x+1, x), \Phi_1(x+1, x+1))$$

Wir beweisen nun den folgenden

Hilfssatz: *Es möge eine ganze Zahl $n \geq 3$ und ein Element a des zuletzt erwähnten Bereiches geben, so daß*

$$\Phi_1(a, a) \sim \Phi_1(a+n, a+n).$$

Dann ist unser Zählausdruck schon in einem Bereiche von n Individuen erfüllbar.

Beweis. Wir nehmen als Elemente des n -zähligen Bereichs $a, a + 1, a + 2, \dots, a + (n - 1)$. Die erfüllende Relation $\Phi_2(x, y)$ definieren wir wie folgt. Für die Paare

$$(a + i, a + i), (a + i, a + i + 1) \text{ und } (a + i + 1, a + i),$$

wo $0 \leq i \leq n - 2$, stimmt Φ_2 mit Φ_1 überein. Ferner ist

$$\Phi_2(a + n - 1, a) \sim \Phi_1(a + n - 1, a + n),$$

$$\Phi_2(a, a + n - 1) \sim \Phi_1(a + n, a + n - 1).$$

Da $n \geq 3$, steht diese letzte Festlegung nicht mit den vorigen in Widerspruch. Im übrigen ist Φ_2 beliebig. Wir ordnen nun jedem Element x des Bereiches $a, a + 1, \dots, a + (n - 1)$ ein anderes $\beta(x)$ zu. Für $x = a, \dots, a + (n - 2)$, ist das $x + 1$, für $a + (n - 1): a$. Aus unserer Definition der Relation Φ_2 ergibt sich dann, daß

$$\mathfrak{A}(\Phi_2(x, x), \Phi_2(x, \beta(x)), \Phi_2(\beta(x), x), \Phi_2(\beta(x), \beta(x)))$$

für alle x eine richtige Aussage darstellt.

Der Zähl Ausdruck

$$(x) (Ey) \mathfrak{A}(F(x, x), F(x, y), F(y, x), F(y, y))$$

ist also in dem Bereich $a, a + 1, \dots, a + (n - 1)$ erfüllbar.

Damit ist der Hilfssatz bewiesen.

Demnach genügt es, wenn wir in dem Bereich $0, 1, 2, \dots$ ein Element a und eine ganze Zahl n von der angegebenen Art finden. Nun muß mindestens einer der folgenden drei Fälle zutreffen:

$$\Phi_1(0, 0) \sim \Phi_1(3, 3),$$

$$\Phi_1(3, 3) \sim \Phi_1(6, 6),$$

$$\Phi_1(0, 0) \sim \Phi_1(6, 6).$$

Daraus ergibt sich, daß der Zähl Ausdruck

$$(x) (Ey) \mathfrak{A}(F(x, x), F(x, y), F(y, x), F(y, y)),$$

wenn überhaupt, schon für einen Bereich mit der Individuenanzahl 6 erfüllbar ist.

Die Lösung des Entscheidungsproblems ist damit für den betrachteten Fall gegeben.

Ich lege übrigens keinen Wert darauf, daß 6 oder die Zahlen, die ich für die komplizierteren Fälle angeben werde, die kleinsten Zahlen dieser Art sind. Es kommt mir nicht darauf an, ein rechnerisch möglichst einfaches Entscheidungsverfahren zu finden, sondern es soll nur die prinzipielle Möglichkeit der Entscheidung mit möglichst einfachen Mitteln dargetan werden.

Die Methode, die bei dem behandelten Falle angewandt wurde, beruhte darauf, daß gewissermaßen das Element $a + n$ mit a identifiziert und die Definition der Funktion Φ_1 entsprechend abgeändert wurde.

Falls der Ausdruck $(x) (Ey) \mathfrak{A}(x, y)$ nicht eine, sondern n variable Relationen $F_1(\dots), \dots, F_n(\dots)$ enthält, so wird ganz ähnlich verfahren. Der Ausdruck $\mathfrak{A}(x, y)$ setzt sich hier durch Verbindung der n Grundaussagen

$$\begin{array}{cccc} F_1(x, x), & F_1(x, y), & F_1(y, x), & F_1(y, y) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ F_n(x, x), & F_n(x, y), & F_n(y, x), & F_n(y, y) \end{array}$$

zusammen. Wir denken uns wieder das zu einem x gehörige y mit $\alpha(x)$ bezeichnet, und können in derselben Weise wie früher zu einem Bereich übergehen, der aus den Elementen $0, 1, 2, \dots$ besteht und für den $\alpha(x)$ gleich $x + 1$ ist. Die n erfüllenden Relationen in diesem Bereich seien

$$\Phi_1(x, y), \Phi_2(x, y), \dots, \Phi_n(x, y).$$

Entsteht $\mathfrak{A}'(x, y)$ aus $\mathfrak{A}(x, y)$ dadurch, daß für F_1, \dots, F_n die Relationen $\Phi_1, \Phi_2, \dots, \Phi_n$ eingesetzt werden, so gilt $(x) \mathfrak{A}'(x, x + 1)$.

Betrachten wir nun irgendein Element a dieses Bereiches. Für die Verteilung von Wahrheit, bzw. Falschheit auf die n Aussagen

$$\Phi_1(a, a), \Phi_2(a, a), \dots, \Phi_n(a, a)$$

gibt es insgesamt 2^n Möglichkeiten; für jedes Element a liegt eine dieser Möglichkeiten vor.

Es gilt nun der folgende

Satz: Falls m eine ganze Zahl ≥ 1 ist, so sind bei der Folge von Elementen $0, 1, \dots, 3(m - 1)$ mindestens m Möglichkeiten eingetreten, oder aber der Zählausdruck ist schon in einem Individuenbereich erfüllbar, der nur aus $3 \cdot (m - 1)$ Elementen besteht.

Wir beweisen die Behauptung durch Induktion nach m . Für $m = 1$ ist sie offenbar richtig. Für ein bestimmtes m sei schon alles gezeigt, und wir gehen jetzt zu $(m + 1)$ über. Der ungünstigste Fall ist der, daß bis zum Gliede $3(m - 1)$ genau m Möglichkeiten vorgekommen sind. Wir nehmen nun die Elemente $3m - 2, 3m - 1, 3m$ hinzu. Kommt hier eine weitere Möglichkeit hinzu, so ist es gut. Andernfalls kommt unter $0, 1, \dots, 3m - 3$ ein Element c vor, so daß

$$\Phi_1(3m, 3m) \sim \Phi_1(c, c),$$

$$\Phi_2(3m, 3m) \sim \Phi_2(c, c),$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$\Phi_n(3m, 3m) \sim \Phi_n(c, c).$$

Wir zeigen dann, daß der Zähl Ausdruck in dem Bereich, der aus den Individuen $0, 1, \dots, 3m - 1$ besteht, erfüllbar ist. Zu diesem Zweck definieren wir in diesem Bereich n Relationen $\Phi'_1(x, y), \dots, \Phi'_n(x, y)$ wie folgt:

Für alle i ($1 \leq i \leq n$) und alle k ($0 \leq k \leq 3m - 2$) soll sein

$$\Phi'_i(k, k + 1) \sim \Phi_i(k, k + 1),$$

$$\Phi'_i(k + 1, k) \sim \Phi_i(k + 1, k),$$

$$\Phi'_i(k, k) \sim \Phi_i(k, k).$$

Ferner sei für alle i

$$\Phi'_i(3m - 1, c) \sim \Phi_i(3m - 1, m),$$

$$\Phi'_i(c, 3m - 1) \sim \Phi_i(3m, 3m - 1).$$

Es werden also gewissermaßen die Elemente c und $3m$ identifiziert. Die beiden letzten Festlegungen stehen nicht mit den vorhergehenden im Widerspruch, da $3m - 1 \neq c + 1$. Im übrigen ist die Definition der Relationen Φ'_1, \dots, Φ'_n beliebig. — Wir bezeichnen nun den Ausdruck, der aus $\mathfrak{A}(x, y)$ dadurch entsteht, daß für F_1, \dots, F_n die bestimmten Relationen Φ'_1, \dots, Φ'_n eingesetzt werden, mit $\mathfrak{A}''(x, y)$. Es bedeute ferner $\beta(x)$ für ein Element x ($0 \leq x \leq 3m - 2$) das Element $x + 1$; $\beta(3m - 1)$ sei gleich c . Es gilt dann für den aus $0, 1, \dots, 3m - 1$ bestehenden Bereich

$$(x)\mathfrak{A}''(x, \beta(x)).$$

Das folgt aus der Gültigkeit von $(x)\mathfrak{A}'(x, x + 1)$ in dem Bereiche $0, 1, 2, \dots$.

Das Bestehen von $(x)\mathfrak{A}''(x, \beta(x))$ besagt aber, daß der Zähl Ausdruck $(x)(Ey)\mathfrak{A}(x, y)$ in einem aus $3m$, oder $3((m + 1) - 1)$ Individuen bestehenden Bereich erfüllt ist. Unser Satz ist damit bewiesen.

Wir wenden diesen Satz jetzt an für $m = 2^n + 1$. Da es nur 2^n Möglichkeiten gibt, so erhalten wir:

Unser Zähl Ausdruck $(x)(Ey)\mathfrak{A}(x, y)$ ist schon in einem Bereiche von $3 \cdot 2^n$ Individuen erfüllbar.

Damit ist die Entscheidungsfrage für die Zähl Ausdrücke der Form $(x)(Ey)\mathfrak{A}(x, y)$ vollständig erledigt.

§ 3.

Die Behandlung der Zähl Ausdrücke der Form

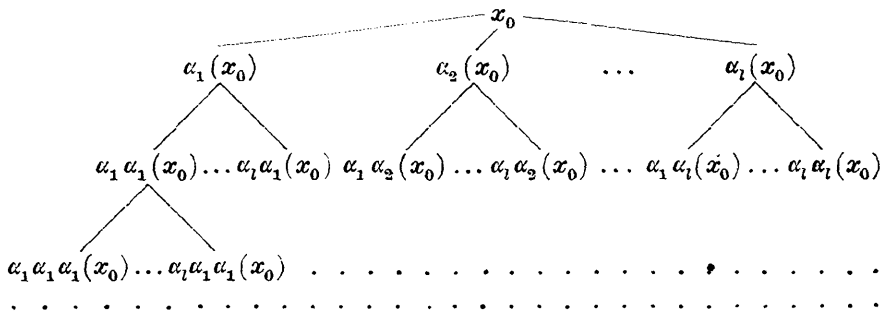
$$(x)(Ey_1)(Ey_2) \dots (Ey_l)\mathfrak{A}(x, y_1, \dots, y_l).$$

Wir betrachten jetzt den allgemeineren Fall, daß hinter dem Allzeichen nicht nur ein, sondern allgemeiner l Existenzialzeichen vorkommen. Die Anzahl der vorkommenden Relationen sei wieder n .

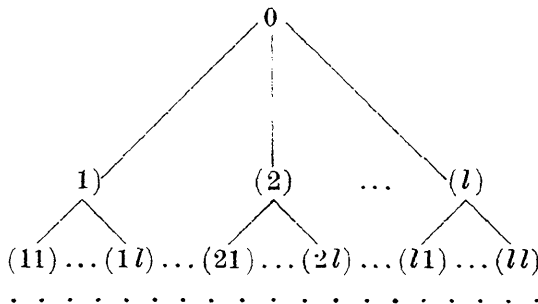
Der Ausdruck sei in irgendeinem Bereich durch die Relationen $\Phi_1(x, y), \dots, \Phi_n(x, y)$ erfüllt. Durch die Einsetzung dieser bestimmten Relationen für die Relationsvariablen gehe $\mathfrak{A}(x, y_1, \dots, y_l)$ in $\mathfrak{A}'(x, y_1, \dots, y_l)$ über. Wir können uns jedem Element x eindeutig l andere Elemente $\alpha_1(x), \alpha_2(x), \dots, \alpha_l(x)$ zugeordnet denken, so daß

$$(x)(\mathfrak{A}'(x, \alpha_1(x), \alpha_2(x), \dots, \alpha_l(x)))$$

eine richtige Aussage ist. x_0 sei ein willkürlich herausgegriffenes Element. Dem Element x_0 sind dann die l Elemente $\alpha_1(x_0), \alpha_2(x_0), \dots, \alpha_l(x_0)$ zugeordnet, einem Element $\alpha_i(x_0)$ die l Elemente $\alpha_1 \alpha_i(x_0), \dots, \alpha_l \alpha_i(x_0)$. Wir können uns diese Zuordnung durch das folgende Schema veranschaulichen.



In dem oben angeschriebenen Schema brauchen natürlich nicht alle Elemente verschieden zu sein. Wir können sie aber immer als verschieden annehmen, und sie mit



bezeichnen. Wir brauchen nämlich für diesen neuen Bereich nur n Relationen Φ'_1, \dots, Φ'_n zu definieren, so daß immer gilt:

$$\Phi'_i[(a_1 \dots a_m), (b_1 \dots b_k)] \sim \Phi_i(\alpha_{a_m} \dots \alpha_{a_1}(x_0), \alpha_{b_k} \dots \alpha_{b_1}(x_0)).$$

Einem beliebigen Element (a_1, \dots, a_m) sind hier die l Elemente:

$$(a_1, \dots, a_m, 1), (a_1, \dots, a_m, 2), \dots, (a_1, \dots, a_m, l)$$

zugeordnet. Entsteht $\mathfrak{U}''(x, y_1, \dots, y_l)$ aus $\mathfrak{U}(x, y_1, \dots, y_l)$ dadurch, daß für F_1, \dots, F_n die Relationen Φ'_1, \dots, Φ'_n eingesetzt werden, so ist für ein beliebiges Element (a_1, \dots, a_m) :

$$\mathfrak{U}''[(a_1, \dots, a_m), (a_1, \dots, a_m 1), \dots, (a_1, \dots, a_m l)]$$

eine richtige Behauptung.

Zu dem „Stammbaum“ eines Elementes $(a_1 \dots a_m)$ in dem zuletzt genannten Schema rechnen wir alle Elemente, deren Bezeichnung mit (a_1, \dots, a_m, \dots) beginnt. Es dürfte klar sein, was man unter den l Elementen zu verstehen hat, die zum 1. Gliede des Stammbaumes eines bestimmten Elementes gehören, unter den l^2 Elementen, die zum 2. Gliede des Stammbaumes gehören, usw.

Wir greifen nun irgend l Elemente b_1, b_2, \dots, b_l heraus, die einem anderen Elemente zugeordnet sind. Für die Verteilung von Wahrheit bzw. Falschheit auf die $n \cdot l^2$ Aussagen

$$\Phi_i(b_p, b_q)$$

gibt es dann $2^{n \cdot l^2}$ Möglichkeiten.

Sei nun m eine ganze Zahl ≥ 1 . Wir betrachten die Möglichkeiten der oben genannten Art, die für zusammengehörige l Elemente eintreten. Es gilt dann der folgende Satz: *In dem Stammbaum eines beliebigen Elementes sind bis zum $(3m - 2)$ -ten Gliede mindestens m verschiedene Möglichkeiten eingetreten, oder aber der Zähl Ausdruck ist schon in einem Individuenbereich erfüllt, der nur aus $\frac{l^{3m-2} - 1}{l - 1}$ Elementen besteht ($l \geq 2$).*

Der Beweis geschieht durch Induktion nach m . Für $m = 1$ stimmt die Behauptung. Für ein bestimmtes m sei schon alles gezeigt. Bis zum $(3m - 2)$ -ten Gliede sind also ungünstigstenfalls genau m Möglichkeiten eingetreten. Wir nehmen nun weitere Elemente bis zum $(3m + 1)$ -ten Gliede. Kommt hier eine neue Möglichkeit vor, so ist es gut. Andernfalls zeigen wir, daß der Zähl Ausdruck schon in dem Bereich erfüllt ist, der aus den Elementen bis zum $3m$ -ten Gliede besteht.

Es gilt nämlich folgendes: Jede Möglichkeit, die bei l zusammengehörigen Elementen des $(3m + 1)$ -ten Gliedes vorkommt, wurde auch schon bis zum $(3m - 2)$ -ten Gliede vertreten, und wir können dort für jede derartige Möglichkeit einen Repräsentanten von l zusammengehörigen Elementen auswählen. Für die Möglichkeit \mathfrak{M}_i seien das $a_{i1}, a_{i2}, \dots, a_{il}$. Liegt nun für l zusammengehörige Elemente des $(3m + 1)$ -ten Gliedes b_1, b_2, \dots, b_l die Möglichkeit \mathfrak{M}_i vor, so wird gewissermaßen jedes b_k mit a_{ik} identifiziert, d. h. in dem neuen Bereiche, der nur aus den Elementen bis zum $3m$ -ten Gliede besteht, wird einem Element a des $3m$ -ten Gliedes, falls ihm vorher b_1, \dots, b_l entsprach, jetzt a_{i1}, \dots, a_{il} zuge-

ordnet. Die neuen erfüllenden Funktionen $\Phi_1'', \dots, \Phi_n''$ werden so festgelegt, daß

$$\Phi_j''(a, a_{ik}) \sim \Phi_j'(a, b_k).$$

Im übrigen bleiben die Relationen unverändert für die Paare, die aus einem Element und seinem zugeordneten bestehen. Da a nicht in dem auf a_{ik} folgenden Glied steht, so ist das möglich. Man erkennt nun leicht, daß auch in dem neuen Bereich der Zähl ausdrück

$$(x)(E y_1) \dots (E y_l) \mathfrak{A}(x, y_1, \dots, y_l)$$

erfüllt ist. Da der neue Bereich aus

$$1 + l + l^2 + \dots + l^{3m} = \frac{l^{3m+1} - 1}{l - 1} \text{ Elementen}$$

besteht, so ist unser Satz bewiesen.

Wir wenden den Satz nun an für $m = 2^{n \cdot l^2} + 1$. Da es nur $2^{n \cdot l^2}$ Möglichkeiten gibt, so erhalten wir das Resultat:

Ein Zähl ausdrück von der Form $(x)(E y_1) \dots (E y_l) \mathfrak{A}(x, y_1, \dots, y_l)$ ist, falls überhaupt, schon in einem Individuenbereich erfüllt, der nur aus $\frac{l^{3 \cdot 2^{n \cdot l^2} + 1} - 1}{l - 1}$ Elementen besteht.

§ 4.

Die Behandlung der Zähl ausdrücke von der Form

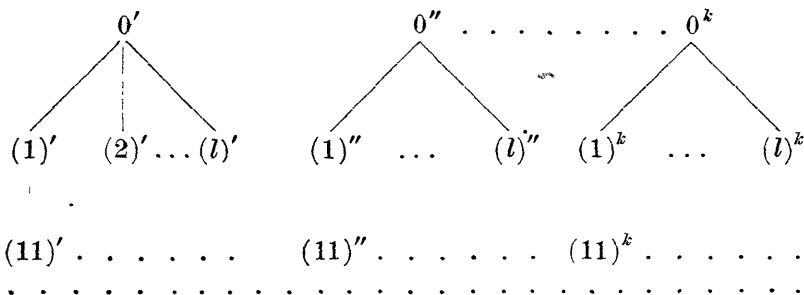
$$(E x_1) \dots (E x_k) (y) (E z_1) \dots (E z_l) \mathfrak{A}(x_1, \dots, x_k, y, z_1, \dots, z_l).$$

Wir erweitern jetzt unsere Betrachtungen auf diejenigen Zähl ausdrücke, bei denen einer Kombination von der Form $(y)(E z_1) \dots (E z_l)$ noch eine beliebige endliche Anzahl k von Existenzialzeichen vorangeht. Die erfüllenden Relationen seien $\Phi_1, \Phi_2, \dots, \Phi_n$.

$\mathfrak{A}(x_1, \dots, x_k, y, z_1, \dots, z_l)$ gehe nach Einsetzung dieser Relationen in $\mathfrak{A}'(x_1, \dots, x_k, y, z_1, \dots, z_l)$ über. Es gibt dann in dem zugrunde gelegten Individuenbereich k Elemente $x_1^0, x_2^0, \dots, x_k^0$, so daß

$$(y)(E z_1) \dots (E z_l) \mathfrak{A}'(x_1^0, \dots, x_k^0, y, z_1, \dots, z_l)$$

eine richtige Aussage darstellt. Die zu einem bestimmten y gehörigen Elemente können wir uns eindeutig bestimmt denken und bezeichnen sie mit $\alpha_1(y), \dots, \alpha_l(y)$. Wir dürfen nun annehmen, daß der Individuenbereich die folgende Gestalt hat: Die $x_1^0, x_2^0, \dots, x_k^0$ sind alle verschieden, wir bezeichnen sie mit $0', 0'', \dots, 0^{(k)}$. Zu jedem dieser k Elemente denken wir uns einen sich immer l -fach verzweigenden Stammbaum, wie wir ihn im vorigen Paragraphen hatten.



Die zu einem Element y gehörigen z_1, z_2, \dots, z_k sind durch das nächste Glied des zu y gehörigen Stammbaumes gegeben. Z. B. gehören zu einem Element $(11)^i$ die l Elemente

$$(111)^i, (112)^i, \dots, (11l)^i.$$

Und entsprechend ist z. B.

$$\mathfrak{A}'(0', 0'', \dots, 0^k, (11)^i, (111)^i, (112)^i, \dots, (11l)^i)$$

eine richtige Aussage. Wir dürfen ferner annehmen, daß die verschiedenen Stammbäume kein Element gemeinsam haben.

Jedem Element des neuen Bereichs ist nämlich ein bestimmtes des alten Bereichs zugeordnet, z. B. ist das für $(123)^i$ das Element $\alpha_3 \alpha_2 \alpha_1(x_i^0)$, und man definiert in dem neuen Bereich n Relationen so, daß ihr Zutreffen für ein Paar von geordneten Elementen davon abhängt, ob die ursprünglichen Relationen für das entsprechende Paar von Elementen zutraten oder nicht.

Betrachten wir ein geordnetes Paar $(0^i, 0^j)$. Für irgendein Φ_p stellt $\Phi_p(0^i, 0^j)$ eine bestimmte Aussage dar, die wahr oder falsch sein kann. Wir wollen diese Aussage mit Γ_{ij}^p bezeichnen. Die Zuordnungen der Elemente denken wir uns durch $\beta_1(x), \dots, \beta_l(x)$ gegeben.

$(y)\mathfrak{A}'(0', \dots, 0^k, y, \beta_1(y), \dots, \beta_l(y))$ ist eine richtige Aussage. Aus der Richtigkeit der Aussage und der besonderen Natur der β folgt, daß auch

$$\begin{aligned} &(Ez_1) \dots (Ez_l) \mathfrak{A}'(0', \dots, 0^k, 0', z_1, \dots, z_l) \\ &\quad \& (Ez_1) \dots (Ez_l) \mathfrak{A}'(0' \dots 0^k, 0'', z_1, \dots, z_l) \\ &\quad \& \dots \& (Ez_1) \dots (Ez_l) \mathfrak{A}'(0', \dots, 0^k, 0^k, z_1, \dots, z_l) \\ &\quad \& (y)(Ez_1) \dots (Ez_l) \mathfrak{A}'(0', \dots, 0^k, y, z_1, \dots, z_k) \end{aligned}$$

eine richtige Aussage ist, und zwar beziehen sich hier sämtliche All- und Existenzialzeichen nur auf den Bereich, der aus dem ursprünglichen entsteht, wenn man die Elemente $0', \dots, 0^k$ fortläßt.

Wir können die letzte Aussage auf die Normalform bringen, indem wir das Allzeichen zuerst setzen und die $(k + 1) \cdot l$ Existenzialzeichen folgen

Abriß einer arithmetischen Theorie der Galoisschen Körper.

(Erste Mitteilung.)

Von

Öystein Ore in New Haven (Conn.) (U. S. A.).

Die vorliegende Abhandlung enthält eine neue arithmetische Theorie der Galoisschen Körper, welche wesentlich mehr als die bekannte Hilbertsche Theorie¹⁾ leistet, und gleichzeitig einen weiteren Einblick in die Struktur der Hilbertschen Gruppen gewährt. Ich habe diese Theorie schon vor längerer Zeit entwickelt, aber da sie eng mit meinen Arbeiten über die Arithmetik der allgemeinen algebraischen Körper²⁾ zusammenhängt, habe ich sie vor dem Erscheinen dieser Abhandlungen nicht veröffentlichen wollen. Die Methode gestattet eine Reihe von Anwendungen auf die arithmetischen Probleme der Galoisschen Körper, gegenseitige Reduktion von Galoisschen Gleichungen, zusammengesetzte Galoissche Körper, Existenz der Einheitswurzeln im Körper; speziell ergeben sich für Abelsche und zyklische Körper mehrere interessante Resultate, worauf ich hier doch nicht eingehe.

Wie in den eben erwähnten Arbeiten über allgemeine Körper ist die Methode, die arithmetischen Eigenschaften des Körpers direkt aus den

¹⁾ D. Hilbert, Grundzüge einer Theorie des Galoisschen Zahlkörpers. Nachrichten d. Königl. Ges. d. Wiss. zu Göttingen 1894, S. 224—236. Man vgl. auch den Bericht von Hilbert im Jahresbericht der D. Mathematikervereinigung 4. Die gleichzeitige Arbeit von Dedekind, Zur Theorie der Ideale, Gött. Nachrichten 1894, S. 271—277 ist hier zu erwähnen.

²⁾ Öystein Ore, Über den Zusammenhang zwischen den definierenden Gleichungen und der Idealtheorie in algebraischen Körpern, erste und zweite Mitteilung. Math. Annalen 96 (1926), S. 313—352 und 97 (1927), S. 569—598. Diese beiden Arbeiten werden im folgenden mit Ore I und Ore II bezeichnet.

Eigenschaften der definierenden Gleichung abzuleiten. Bei Galoisschen Körpern besteht aber der folgende Hauptsatz, der sich als ein Schlüssel zur ganzen Arithmetik dieser Körper erweist: *Wenn \mathfrak{P} ein beliebiges Primideal ist, so kann man das volle Restsystem $(\text{mod } \mathfrak{P}^a)$ (a beliebig groß) durch sukzessive Adjunktionen der Wurzeln von irreduziblen binomischen und trinomischen Normalkongruenzen gewinnen.*

Man zeigt auch leicht, daß dies die einfachst möglichen Normalkongruenzen sind; der trinomische Fall tritt nur dann ein, wenn höhere Verzweigungsgruppen auftreten.

In der vorliegenden ersten Mitteilung werden zunächst in Kapitel 1 die Hilbertschen und Speiserschen Resultate über die Zerlegungsgruppe erwähnt, dann folgen kurz die Eigenschaften vom Zerlegungs-, Trägheits- und ersten Verzweigungskörper, und die Struktur der Faktorgruppe G_Z/G_V wird vollständig abgeleitet, wodurch der reguläre Fall erledigt ist. Zuletzt wird eine später nützliche Anwendung gemacht, indem die notwendige und hinreichende Bedingung für die Existenz einer p -ten Einheitswurzel $(\text{mod } \mathfrak{P}^a)$ angegeben wird.

In Kap. 2 werden dann die Vorbereitungen für den Beweis des Hauptsatzes und für die Behandlung des irregulären Falles gemacht. Die aufgestellten Sätze haben aber, wie es mir scheint, auch an sich Interesse. Für Relativkörper vom Primzahlgrad wird die Existenz der Normalkongruenzen bewiesen, und zuletzt wird für die spätere Anwendung der Fall eines relativ-zyklischen Körpers eingehend studiert.

In der zweiten Mitteilung soll dann der Hauptsatz aufgestellt und zum Studium des irregulären Falles angewandt werden.

Man hätte aus diesen Untersuchungen leicht die Hilbertsche Theorie über den Zusammenhang zwischen den Eigenschaften der Primideale und ihrer zugehörigen Gruppen ableiten können, um aber die Darstellung möglichst kurz zu machen, werden die Hilbertschen Resultate überall als bekannt vorausgesetzt.

Ich bemerke zuletzt, daß es mir eine der wichtigsten Aufgaben der arithmetischen Idealtheorie zu sein scheint, auch für allgemeine Körper eine entsprechende, auf Normalkongruenzen bauende Theorie zu schaffen. Eine Reihe von noch ungelösten Problemen würden dadurch ihre Beantwortung finden. Im regulären Falle leitet man die Lösung leicht aus den Henselschen Resultaten³⁾ ab, in dem schwierigeren irregulären Falle bilden meine Sätze in Kapitel 2 eine Lösung im einfachsten Falle, wo die Irregularitätsordnung gleich 1 ist.

³⁾ Vgl. K. Hensel, Zur Theorie der algebraischen Zahlen I, Kap. 8, § 6. Leipzig 1908.

Kapitel 1.

Zerlegungskörper, Trägheitskörper und Regularkörper.

§ 1.

Übersicht über die Hilbertschen Gruppen.

Im folgenden sollen die Eigenschaften eines allgemeinen Galoisschen Körpers K vom Grade N studiert werden. Es sei G die entsprechende endliche Gruppe von der Ordnung N . Weiter sei p eine beliebige Primzahl und \mathfrak{P} ein Primidealteiler von p . Wenn dann \mathfrak{P} die Ordnungszahl e und Gradzahl f hat, so wird bekanntlich

$$(1) \quad p = (\mathfrak{P}_1 \dots \mathfrak{P}_g)^e, \quad N\mathfrak{P}_i = p^f, \quad N = efg.$$

Dem Primideal \mathfrak{P} entspricht nach Hilbert eine Reihe von ineinander geschachtelten Untergruppen von G , welche in der folgenden Weise definiert sind:

I. Die *Zerlegungsgruppe* G_Z besteht aus der Gesamtheit der Substitutionen Z in G , wofür, wenn ω eine beliebige Zahl in \mathfrak{P} ist, auch immer die Zahl $Z:\omega$ in \mathfrak{P} enthalten ist. Der Zerlegungsgruppe G_Z entspricht ein Unterkörper K_Z von K , der als *Zerlegungskörper* von P bezeichnet wird.

II. Die *Trägheitsgruppe* G_T besteht aus allen Substitutionen T in G , wofür, für jede Zahl ω in K

$$(2) \quad T:\omega \equiv \omega \pmod{\mathfrak{P}}.$$

Die Trägheitsgruppe ist ein Normalteiler der Zerlegungsgruppe, und die Faktorgruppe G_Z/G_T ist zyklisch von der Ordnung f . Die Ordnung von G_T ist somit gleich e .

Der entsprechende Körper K_T heißt der *Trägheitskörper* und ist ein Relativkörper vom Grade f zu K_Z , während K ein Relativkörper vom Grade e zu K_T ist.

III. Die *Verzweigungsgruppen*. Die *erste Verzweigungsgruppe* G_{V_1} besteht aus allen Substitutionen V_1 in G , wofür, für alle ω in K

$$(3) \quad V_1:\omega \equiv \omega \pmod{\mathfrak{P}^2}.$$

G_{V_1} ist ein Normalteiler von G_Z und G_T . Wenn die Ordnungszahl e von \mathfrak{P} die Form $e = e_0 p^s$ hat, wobei $(e_0, p) = 1$, so ist die Faktorgruppe G_T/G_{V_1} zyklisch von der Ordnung e_0 , und die erste Verzweigungsgruppe G_{V_1} hat daher die Ordnung p^s . Der entsprechende *erste Verzweigungskörper* K_{V_1} ist ein Relativkörper vom Grade e_0 zu K_T , während K ein Relativkörper vom Grade p^s in bezug auf K_{V_1} ist.

Der schwierige Fall tritt ein, wenn die Ordnungszahl e von \mathfrak{P} wirklich durch p teilbar ist, so daß G_{V_1} nicht die Einheitsgruppe ist. Die Primzahl p wird in diesem Falle als *irregulär* bezeichnet.

Es sei nun in (3) \mathfrak{P}^{μ_1} die höchste Potenz, wofür die Kongruenz

$$(4) \quad V_1: \omega \equiv \omega \pmod{\mathfrak{P}^{\mu_1}}$$

für alle Substitutionen in G_{V_1} besteht. Diejenigen Substitutionen V_2 in G_{V_1} , wofür noch

$$V_2: \omega \equiv \omega \pmod{\mathfrak{P}^{\mu_1+1}}$$

für alle ω , bilden die *zweite Verzweigungsgruppe* G_{V_2} , woraus man in entsprechender Weise die höheren Verzweigungsgruppen G_{V_i} ableitet.

Man erhält so eine Reihe von eingeschachtelten Untergruppen

$$(5) \quad G_{V_1}, G_{V_2}, \dots, G_{V_k}, G_{V_{k+1}} = 1,$$

wobei allgemein G_{V_i} ein Normalteiler von G_Z, G_T und allen vorangehenden Verzweigungsgruppen ist. Jeder Gruppe G_{V_i} ist eine Zahl μ_i zugeordnet, so daß für eine beliebige Substitution V_i in G_{V_i} die Kongruenz

$$(6) \quad V_i: \omega \equiv \omega \pmod{\mathfrak{P}^{\mu_i}}$$

für alle ω in K besteht, während dieselbe Kongruenz $\pmod{\mathfrak{P}^{\mu_i+1}}$ nicht für alle V_i richtig bleibt. Die Zahl μ_i soll die *Verzweigung* von G_{V_i} heißen, und man hat offenbar

$$(7) \quad \mu_1 < \mu_2 < \dots < \mu_k.$$

Es gibt also keine von der Einheitssubstitution verschiedene Substitution, so daß

$$V: \omega \equiv \omega \pmod{\mathfrak{P}^{\mu_k+1}}$$

für alle ω in K .

Herr A. Speiser⁴⁾ hat zwei interessante Sätze über die Verzweigungen bewiesen: Es ist

$$(8) \quad \mu_k \leq \frac{e}{p-1} + 1$$

und weiter

$$(9) \quad \mu_1 \equiv \mu_2 \equiv \dots \equiv \mu_k \pmod{p}.$$

Die Speisersche Ungleichung ist, wie wir sehen werden, eine andere Form der allgemeinen Dedekind-Henselschen Ungleichung für die Differentenexponenten bei irregulären Primzahlen (vgl. Ore I, Kap. 3, Satz 10); wir werden auch eine obere Grenze für jedes μ_i angeben. Die Kongruenz (9) hat eine besondere Bedeutung für die Form der Normalkongruenzen.

⁴⁾ A. Speiser, Die Zerlegungsgruppe. Journ. f. Math. 149 (1919), S. 174—188.

Die Faktorgruppe $G_{V_i}/G_{V_{i+1}}$ ist Abelsch von der Ordnung p^{s_i} und Typus $\overbrace{p, p, \dots, p}^{s_i}$. Dabei ist natürlich

$$(10) \quad s = s_1 + s_2 + \dots + s_k,$$

und weiter leitet man leicht die einschränkende Bedingung⁵⁾

$$(11) \quad s_i \leq f \quad (i = 1, 2, \dots, k)$$

ab. Die Ordnung einer Gruppe G_{V_i} ist somit $p^{s_1 + \dots + s_k}$.

Aus den hier erwähnten Resultaten folgt sofort der Hauptsatz von Hilbert, daß die Zerlegungsgruppe meta-zyklisch ist.

Der i -ten Verzweigungsgruppe G_{V_i} entspricht weiter der i -te Verzweigungskörper K_{V_i} , der ein Relativkörper vom Grade $p^{s_{i-1}}$ zu $K_{V_{i-1}}$ ist, während K ein Relativkörper vom Grade $p^{s_1 + \dots + s_k}$ zu K_{V_1} ist. Speziell hat man $K_{V_{k+1}} = K$.

§ 2.

Der Zerlegungskörper.

Es sei nun

$$(12) \quad F(x) = 0$$

die irreduzible Galoissche Gleichung N -ten Grades, welche K definiert, und weiter sei ω eine beliebige der Wurzeln dieser Gleichung. Wir werden im folgenden alle auftretenden algebraischen Zahlen als ganz voraussetzen. Aus der Primidealzerlegung (1) folgen dann sofort die Tatsachen⁶⁾:

Das Polynom $F(x)$ in (12) hat die Zerlegung

$$(13) \quad F(x) \equiv \Phi_1(x) \dots \Phi_q(x) \pmod{p^\alpha}$$

in irreduzible Faktoren $\pmod{p^\alpha}$, wobei $\alpha > \delta$ ist, wenn die Diskriminante von $F(x)$ genau durch p^δ teilbar ist. Weiter ist die Zerlegung (13) $\pmod{p^{\alpha-\delta}}$ eindeutig, und die Faktoren $\Phi_i(x)$ sind alle vom selben Grade ef .

Jedem Faktor $\Phi_i(x)$ in (13) entspricht ein Primideal \mathfrak{P}_i in (1) in der Weise, daß

$$(14) \quad \Phi_i(\omega) \equiv 0 \pmod{\mathfrak{P}_i^{e\left(\alpha - \frac{\delta}{2}\right)}},$$

während alle Zahlen $\Phi_j(\omega)$, $j \neq i$ durch feste, von α unabhängige Potenzen von \mathfrak{P}_i teilbar sind.

Es sei nun $\Phi(x)$ ein beliebiger irreduzibler Faktor in (13) und \mathfrak{P} das entsprechende Primideal, so daß nach (14) die Kongruenz

$$(15) \quad \Phi(\omega) \equiv 0 \pmod{\mathfrak{P}^{e\left(\alpha - \frac{\delta}{2}\right)}}$$

⁵⁾ Diese Bedingung kommt schon bei Hilbert vor. Man vgl. die Abhandlung in den Gött. Nachrichten, S. 233.

⁶⁾ Man vgl. Ore II, Einleitung.

besteht. Wenn dann

$$Z_1, Z_2, \dots, Z_{ef}$$

die Substitutionen der Zerlegungsgruppe und

$$\omega = \omega_1, \omega_2, \dots, \omega_{ef}$$

die entsprechenden konjugierten Zahlen zu ω sind, so folgen aus (15) die Kongruenzen

$$\Phi(\omega_i) \equiv 0 \pmod{\mathfrak{P}^{e\left(\alpha - \frac{\delta}{2}\right)}} \quad (i = 1, 2, \dots, ef),$$

woraus man ohne Schwierigkeit den folgenden Satz ableitet⁷⁾:

Satz 1. *Wenn $\Phi(x)$ ein irreduzibler Faktor in der Zerlegung (13) $(\text{mod } p^\alpha)$ und \mathfrak{P} das entsprechende Primideal ist, so hat $\Phi(x)$ die Zerlegung*

$$\Phi(x) \equiv (x - \omega_1) \dots (x - \omega_{ef}) \pmod{\mathfrak{P}^{e(\alpha - \delta)}},$$

wobei $\omega_1 \dots \omega_{ef}$ gewisse der Wurzeln der Gleichung (12) sind. Die Zerlegungsgruppe ist die Gruppe der Vertauschungen dieser Wurzeln.

Es sollen nun kurz die Eigenschaften des Zerlegungskörpers K_Z erwähnt werden. In K_Z zerfällt bekanntlich, wie man auch leicht aus diesen Untersuchungen hätte ableiten können, die Primzahl p in der folgenden Weise:

$$(16) \quad p = \mathfrak{P}_Z \mathfrak{Q}_Z, \quad (\mathfrak{P}_Z, \mathfrak{Q}_Z) = 1, \quad N_Z(\mathfrak{P}_Z) = p,$$

während \mathfrak{P}_Z in K die Zerlegung

$$(17) \quad \mathfrak{P}_Z = \mathfrak{P}^e, \quad n_Z(\mathfrak{P}) = \mathfrak{P}_Z^f$$

hat, wobei n_Z die Relativnorm in bezug auf K_Z bezeichnet.

Da nach (16) \mathfrak{P}_Z ein Primideal ersten Grades ist, so wird jede Zahl in $K_Z \pmod{\mathfrak{P}_Z}$ kongruent einer rationalen Zahl, und da p genau die erste Potenz von \mathfrak{P}_Z enthält, folgt leicht:

Satz 2. *Jede Zahl ω_Z im Zerlegungskörper genügt für ein beliebig hohes α einer Kongruenz*

$$\omega_Z \equiv a \pmod{\mathfrak{P}_Z^\alpha = \text{mod } \mathfrak{P}^{e\alpha}},$$

wobei a rational ist.

Die Zahl a hängt natürlich von α ab. Da wir im folgenden die Zahlen in K nur für Moduln untersuchen werden, welche Potenzen von \mathfrak{P} sind, können wir dabei die Zahlen im Zerlegungskörper als rational voraussetzen.

⁷⁾ Man vgl. die Arbeit von F. Hüttig: Arithmetische Theorie eines Galoisschen Körpers. Diss. Marburg 1907.

Bilden wir nun die Gleichung $\Phi_0(x) = 0$ ef -ten Grades mit Koeffizienten in K_Z , welcher eine Zahl ω in K genügt, so wird offenbar

$$\Phi_0(x) = (x - \omega_1) \dots (x - \omega_{ef}),$$

und daher nach Satz 1 und Satz 2

$$(18) \quad \Phi_0(x) \equiv \Phi(x) \pmod{P_Z^{\alpha-\delta}},$$

wobei wie früher $\Phi(x)$ der rationale, $(\text{mod } p^a)$ irreduzible Faktor in (13) ist, dem \mathfrak{P} entspricht.

Zuletzt soll noch erwähnt werden, daß man die Zahl ω so wählen kann, daß

$$(19) \quad \Phi(x) = \varphi(x)^e + pM(x), \quad M(x) \not\equiv 0 \pmod{p, \varphi(x)}$$

wird, wobei $\varphi(x)$ eine Primfunktion $(\text{mod } p)$ f -ten Grades ist. (Ore I, Kap. 3, Satz 3).

§ 3.

Der Trägheitskörper.

Es sollen weiter kurz die wichtigsten Eigenschaften der Trägheitsgruppe G_T und Trägheitskörper K_T erwähnt werden.

Es sei $\Phi_0(x) = 0$ die Gleichung in K_Z ef -ten Grades, welcher eine beliebige Zahl ω genügt; nach (18) und (19) folgt, daß man ω so wählen kann, daß

$$(20) \quad \Phi_0(x) \equiv (x - \omega)^e (x - \omega^p)^e \dots (x - \omega^{p^{f-1}})^e \pmod{\mathfrak{P}}$$

ist. Da nun G_T aus denjenigen Substitutionen T besteht, wofür immer

$$T: \omega \equiv \omega \pmod{\mathfrak{P}},$$

erhält man sofort aus (20):

Wenn Z eine Substitution in G_Z bezeichnet, welche ω in $\omega^p \pmod{P}$ überführt, so hat G_Z die Form

$$(21) \quad G_Z = Z^i G_T \quad (i = 0, 1, \dots, f-1),$$

wobei

$$(22) \quad Z^f = T_1, \quad T_2 Z = Z T_3;$$

T_1, T_2, T_3 bezeichnen Substitutionen in der Trägheitsgruppe.

Der Trägheitskörper K_T ist ein Relativkörper f -ten Grades zu K_Z , und das Primideal P_Z in (16) hat in K_T die Zerlegung

$$(23) \quad \mathfrak{P}_Z = \mathfrak{P}_T, \quad N_{Z/T}(\mathfrak{P}_T) = \mathfrak{P}_Z^f,$$

und daher ist in K

$$\mathfrak{P}_T = \mathfrak{P}^e, \quad n_T(\mathfrak{P}) = \mathfrak{P}_T.$$

Es folgt hieraus, was wir für die spätere Anwendung bemerken, daß das Primideal \mathfrak{P}_T in jedem Körper zwischen K_T und K in gleiche Primideale ersten Grades in bezug auf K_T zerfallen muß.

Aus der Primidealzerlegung (23) folgt, daß K_T ein *Gradkörper* f -ten Grades in bezug auf K_Z ist, und man leitet daraus die folgenden Tatsachen ab⁵⁾:

Satz 3. *Man kann eine Zahl τ im Trägheitskörper so bestimmen, daß τ eine primitive Wurzel der Kongruenz*

$$(24) \quad x^{p^{f-1}} \equiv 1 \pmod{\mathfrak{P}^{e\alpha}}$$

ist, wobei α beliebig groß gewählt werden kann. Durch die Substitution Z in G_Z geht τ in $\tau^p \pmod{\mathfrak{P}^{e\alpha}}$ über.

Die Bezeichnung primitiv in diesem Satze bedeutet, daß p^{f-1} der kleinste Exponent ist, wofür τ einer Kongruenz (24) genügt.

Für jede nicht durch \mathfrak{P} teilbare Zahl ω in K besteht dann eine Kongruenz

$$\omega \equiv \tau^a \pmod{\mathfrak{P}},$$

und wenn π eine *Primzahl* in K in bezug auf \mathfrak{P} bezeichnet, d. h. eine Zahl, welche genau die erste Potenz von \mathfrak{P} enthält, so läßt sich ω immer in der Form

$$\omega \equiv a_0(\tau) + a_1(\tau)\pi + \dots + a_{e-1}(\tau)\pi^{e-1} \pmod{\mathfrak{P}^{e\alpha}}$$

darstellen, wobei alle $a_i(\tau)$ Zahlen in K_T sind. Die Primzahl π genügt in K_T einer Gleichung $F_T(x) = 0$ e -ten Grades, wo $F_T(x)$ die Form

$$(25) \quad F_T(x) \equiv x^e + p c_1(\tau) x^{e-1} + \dots + p c_e(\tau) \pmod{\mathfrak{P}^{e\alpha}}$$

hat, und wo $c_e(\tau) \not\equiv 0 \pmod{\mathfrak{P}}$.

§ 4

Der reguläre Fall.

Wir werden nun den regulären Fall erledigen, wo die Ordnungszahl e nicht durch p teilbar ist. Nach § 1 folgt, daß in diesem Falle die erste Verzweigungsgruppe G_V , gleich der Einheitsgruppe ist, so daß der erste Verzweigungskörper K_V , mit K identisch ist. Wir hätten dies auch leicht aus den folgenden Untersuchungen ableiten können.

Nach (25) genügt die Primzahl π der Kongruenz

$$\pi^e + p c_e(\tau) \equiv 0 \pmod{\mathfrak{P}^{e+1}},$$

oder wenn

$$c_e(\tau) \equiv \tau^a \pmod{\mathfrak{P}}$$

⁵⁾ Man vgl. Ore II, Kap. I, § 1 oder K. Hensel, loc. cit. Kap. 8, §§ 2—4.

ist, erhält man

$$(26) \quad \pi^e + p\tau^a \equiv 0 \pmod{\mathfrak{P}^{e+1}}.$$

Wenn daher $\mathfrak{P}^{e+\gamma}$ die höchste Potenz von \mathfrak{P} ist, wofür die Kongruenz (26) besteht, kann man

$$(27) \quad \pi^e + p\tau^a \equiv \pi^{e+\gamma} \omega \pmod{\mathfrak{P}^{e+\gamma+1}}$$

setzen. Wir werden nun aber zeigen, daß man eine neue Primzahl π' in K so bestimmen kann, daß die Kongruenz (26) auch mindestens $\pmod{\mathfrak{P}^{e+\gamma+1}}$ besteht. Man braucht nur $\pi' = \pi(1 + \omega_1 \pi^\gamma)$ zu setzen und erhält zur Bestimmung von ω_1

$$\pi'^e + p\tau^a \equiv \pi^e(1 + e\omega_1 \pi^\gamma) + p\tau^a \equiv 0 \pmod{\mathfrak{P}^{e+\gamma+1}}$$

oder nach (27)

$$e\omega_1 + \omega \equiv 0 \pmod{\mathfrak{P}},$$

woraus ω_1 immer bestimmt werden kann. Wir haben daher bewiesen⁹⁾:
Man kann die Primzahl π so wählen, daß sie einer binomischen Kongruenz

$$(28) \quad x^e + p\tau^a \equiv 0 \pmod{\mathfrak{P}^{e\alpha}}$$

für ein beliebig hohes, festes α genügt.

Die Kongruenz (28) hat offenbar e verschiedene Wurzeln, welche man erhält, wenn man auf π die Substitutionen in G_T ausübt. Die Form dieser Wurzeln ist einfach bestimmbar; denn wenn π' eine von π verschiedene Wurzel von (28) ist, so folgt

$$\left(\frac{\pi'}{\pi}\right)^e \equiv 1 \pmod{\mathfrak{P}^{e(\alpha-1)}},$$

und daher hat die Kongruenz

$$(29) \quad x^e \equiv 1 \pmod{\mathfrak{P}^{e\alpha}}$$

für ein beliebig hohes α genau e verschiedene Wurzeln. Dies ist aber bekanntlich nur dann möglich, wenn e ein Teiler von $p^f - 1$ ist, und wir haben daher die Hilbertsche Bedingung abgeleitet:

Wenn $e \not\equiv 0 \pmod{p}$ ist, besteht im Galoisschen Körper für jedes Primideal die Bedingung

$$(30) \quad p^f - 1 \equiv 0 \pmod{e}.$$

Nach (30) können wir $b = \frac{p^f - 1}{e}$ setzen, wo b ganz ist, und die Wurzeln der Kongruenz (29) sind

$$1, \tau^b, \tau^{2b}, \dots, \tau^{p^f - 1 - b},$$

⁹⁾ Man vgl. K. Hensel, loc. cit. Kap. 8, § 6.

und die Kongruenz (28) hat folglich die Wurzeln

$$\pi, \tau^b \pi, \dots, \tau^{p^f - 1 - b} \pi.$$

Wenn daher T eine Substitution in G_T ist, welche π in $\tau^b \pi \pmod{\mathfrak{P}^{ea}}$ überführt, so wird $T^e = 1$, und man hat

$$G_T = T^j \quad (j = 0, 1, \dots, e - 1).$$

Wendet man aber auf π eine Substitution Z in G_T an, wobei Z nach Satz 3 die Zahl τ in τ^p überführt, so sei

$$(31) \quad \pi' = Z: \pi \equiv \tau^\lambda \pi \pmod{\mathfrak{P}^2}$$

gesetzt. Die Zahl π' genügt nach (28) der Kongruenz

$$\pi'^e + p \tau^{pa} \equiv 0 \pmod{\mathfrak{P}^{ea}},$$

woraus nach (31)

$$\pi^e + p \tau^{pa - e\lambda} \equiv 0 \pmod{\mathfrak{P}^{e+1}}$$

folgt, und diese Kongruenz ist nur dann in K lösbar, wenn

$$pa - e\lambda \equiv a \pmod{p^f - 1}$$

oder

$$\lambda = \frac{a(p-1)}{e}$$

eine ganze Zahl ist.

Satz 4. *Im regulären Falle, $e \not\equiv 0 \pmod{p}$, kann man das volle Restsystem $\pmod{\mathfrak{P}^{ea}}$ in K erhalten, wenn man die Wurzel einer binomischen Kongruenz*

$$(32) \quad x^e + p \tau^a \equiv 0 \pmod{\mathfrak{P}^{ea}}$$

zum Trägheitskörper adjungiert. Dabei müssen die Zahlen

$$(33) \quad b = \frac{p^f - 1}{e}, \quad \lambda = \frac{a(p-1)}{e}$$

ganz sein.

Aus der Normalkongruenz (32) leitet man sofort die Konstitution der ganzen Verzweigungsgruppe im regulären Falle ab. Nach (21) folgt sofort

$$G_Z = Z^i T^j \quad (i = 0, 1, \dots, f - 1, j = 0, 1, \dots, e - 1).$$

Nach (22) ist hier Z^f eine Substitution in G_T , die wir einfach bestimmen können. Aus (31) folgt durch Wiederholung

$$(34) \quad Z^i: \pi \equiv \pi \tau^{\frac{p^i - 1}{p - 1} \lambda} \pmod{\mathfrak{P}^2}$$

oder wenn man den Wert (33) für λ , $i = f$ in (34) einführt, erhält man

$$Z^f: \pi \equiv \pi \tau^{ab} \pmod{\mathfrak{P}^2},$$

so daß $Z^f = T^a$ sein muß. Dadurch ist auch die gruppentheoretische Bedeutung der Konstante a in (32) erklärt.

Weiter folgt leicht

$$Z^{-1}: \pi \equiv \pi \tau^{-\lambda p^{f-1}} \pmod{\mathfrak{P}^2},$$

so daß

$$Z^{-1}TZ: \pi \equiv \pi \tau^{pb} \equiv T^p: \pi \pmod{\mathfrak{P}^2}$$

wird. Wir fassen diese Resultate in dem folgenden Satze zusammen:

Satz 5. *Im regulären Falle ist die Zerlegungsgruppe vom Typus*

$$G_Z = Z^i T^j \quad (i = 0, 1, \dots, f-1, j = 0, 1, \dots, e-1),$$

wobei

$$Z^f = T^a, \quad T^e = 1, \quad Z^{-1}TZ = T^p,$$

und a die in Satz 4 vorkommende Konstante ist, und daher $p^f - 1 \equiv a(p-1) \equiv 0 \pmod{e}$.

§ 5.

Der Regulararkörper.

Wir gehen nun zum schwierigeren Falle über, wo e durch p teilbar ist, und setzen

$$e = e_0 p^s, \quad e_0 \not\equiv 0 \pmod{p}.$$

Die Verzweigungsgruppe G_{V_1} hat in diesem Falle die Ordnung p^s , und der erste Verzweigungskörper K_{V_1} ist ein Relativkörper vom Grade e_0 zum Trägheitskörper. Nach einer früheren Bemerkung muß dann das Primideal \mathfrak{P}_T in (23) die Zerlegung

$$(35) \quad \mathfrak{P}_T = \mathfrak{P}_{V_1}^{e_0}, \quad N_{T, V_1}(\mathfrak{P}_{V_1}) = \mathfrak{P}_T$$

haben, und weiter ist

$$\mathfrak{P}_{V_1} = \mathfrak{P}^{p^s}, \quad n_{V_1}(\mathfrak{P}) = \mathfrak{P}_{V_1}.$$

Da nach (35) die Primzahl p in K_{V_1} regulär ist, werden wir im folgenden $K_{V_1} = K_0$ setzen, und K_0 als *Regulararkörper* bezeichnen.

Es sei nun π_0 eine Primzahl in K_0 , d. h. eine Zahl, welche genau die erste Potenz von $\mathfrak{P}_0 = \mathfrak{P}_{V_1}$ enthält. Dann kann man nach § 4 die Zahl π_0 so wählen, daß sie einer Kongruenz

$$\pi_0^{e_0} + p\tau^{a_0} \equiv 0 \pmod{\mathfrak{P}_0^{e_0 a}}$$

genügt, und genau wie in § 4 leitet man den folgenden Satz ab:

Satz 6. *Man kann die Primzahl π_0 im Regulararkörper so bestimmen, daß π_0 der Kongruenz*

$$(36) \quad x^{e_0} + p\tau^{a_0} \equiv 0 \pmod{\mathfrak{P}^{e_0 a}}$$

für ein beliebig hohes a genügt; dabei müssen auch die Zahlen

$$(37) \quad b_0 = \frac{p^f - 1}{e_0}, \quad \lambda = \frac{a_0(p-1)}{e_0}$$

ganz sein.

Wenn $\delta_0 = (p - 1, e_0)$ ist, kann man statt der letzten Bedingung (37)

$$a_0 \equiv 0 \pmod{\frac{e_0}{\delta_0}}$$

schreiben.

Die Wurzeln der Kongruenz (36) sind

$$(38) \quad \pi_0, \tau^{b_0} \pi_0, \dots, \tau^{p^f - 1 - b_0} \pi_0,$$

während π_0 durch die Substitution Z in $\tau^i \pi_0 \pmod{\mathfrak{P}^{e\alpha}}$ übergeht.

Jede Zahl in K kann in der Form

$$\omega \equiv A_0^{(0)} + A_1^{(0)} \pi + \dots + A_{p^s-1}^{(0)} \pi^{p^s-1} \pmod{\mathfrak{P}^{e\alpha}}$$

dargestellt werden, wo die Koeffizienten $A_i^{(0)}$ Zahlen im Regularkörper sind. Die Primzahl π genügt in K_0 einer Gleichung $F_0(x) = 0$ p^s -ten Grades, wo

$$(39) \quad F_0(x) \equiv x^{p^s} + \pi_0 C_1^{(0)} x^{p^s-1} + \dots + \pi_0 C_{p^s}^{(0)} \pmod{\mathfrak{P}^{e\alpha}}$$

und $C_{p^s}^{(0)} \equiv 0 \pmod{\mathfrak{P}_0}$. Bildet man die Gleichung einer anderen Primzahl $\pi' = \pi \tau^\beta$, so kann man offenbar den Exponenten β so wählen, daß

$$(40) \quad C_{p^s}^{(0)} \equiv -1 \pmod{\mathfrak{P}}$$

ist. Wir werden im folgenden immer voraussetzen, daß diese Bedingung erfüllt ist.

Dann wird nach (39)

$$\pi_0 \equiv \pi^{p^s} \pmod{\mathfrak{P}^{p^s+1}},$$

und es folgt sofort, daß, wenn T die Zahl π_0 in $\pi_0 \tau^{b_0}$ überführt, so wird

$$Z : \pi \equiv \pi \tau^{\lambda'}, \quad T : \pi \equiv \pi \tau^{b'} \pmod{\mathfrak{P}^3},$$

wobei

$$\lambda' \equiv \lambda_0 p^{-s}, \quad b' \equiv b_0 p^{-s} \pmod{p^f - 1}.$$

Wie in § 4 beweist man das Analogon zum Satz 5:

Satz 7. Die Zerlegungsgruppe hat den Typus

$$G_Z = Z^i T^j G_{V_1} \quad (i = 0, 1, \dots, f - 1, j = 0, 1, \dots, e - 1),$$

wobei

$$Z^f = T^{a_0} V_1^{(0)}, \quad T^{e_0} = V_2^{(0)}, \quad Z^{-1} T Z = T^p V_3^{(0)}.$$

Die Substitutionen $V_1^{(0)}, V_2^{(0)}, V_3^{(0)}$ sind alle in der ersten Verzweigungsgruppe G_{V_1} enthalten.

Man könnte auch zufügen, daß

$$V_1^{(0)} Z = Z V_2^{(0)}, \quad V_3^{(0)} T = T V_4^{(0)}$$

ist, indem G_{V_1} ein Normalteiler von G_Z und G_T ist.

§ 6.

Die Einheitswurzeln (mod $\mathfrak{P}^{e\alpha}$).

Wir sollen zuletzt, für eine spätere Anwendung, einen Satz über die Einheitswurzeln in K beweisen. Wir sagen, daß eine Zahl ε_0 eine Einheitswurzel (mod $\mathfrak{P}^{e\alpha}$) ist, wenn ε_0 einer Kongruenz

$$(41) \quad x^q - 1 \equiv 0 \pmod{\mathfrak{P}^{e\alpha}}$$

genügt. Für mehrere wichtige Probleme der Arithmetik der algebraischen Körper ist es von Bedeutung, zu wissen, welche Einheitswurzeln (mod $\mathfrak{P}^{e\alpha}$) im Körper vorkommen.

Aus Satz 3 folgt, daß die Kongruenz (41) sicher lösbar ist, wenn q ein Teiler von $p^f - 1$ ist. Nach einer Bemerkung, die man Hensel¹⁰⁾ verdankt, folgt, daß die Kongruenz (41) sonst nur in dem Falle eine Lösung haben kann, wo q eine Potenz von p ist.

Mittels der hier gewonnenen Resultate kann man einfach die notwendige und hinreichende Bedingung dafür angeben, daß K eine p -te Einheitswurzel enthält. Später soll auch die allgemeine Bedingung für eine p^r -te Einheitswurzel (mod $\mathfrak{P}^{e\alpha}$) aufgestellt werden.

Wenn die Kongruenz

$$x^p - 1 \equiv 0 \pmod{\mathfrak{P}^{e\alpha}}$$

eine Lösung $\varepsilon_0 \neq 1$ hat, kann man $\varepsilon_0 = 1 + \eta_0$ setzen, und erhält zur Bestimmung von η_0

$$\eta_0^{p-1} + \binom{p}{1} \eta_0^{p-2} + \dots + p \equiv 0 \pmod{\mathfrak{P}^{e\alpha}}.$$

Wie in § 4 zeigt man, daß diese Kongruenz dann und nur dann lösbar ist, wenn die Kongruenz

$$(42) \quad \eta_0^{p-1} + p \equiv 0 \pmod{\mathfrak{P}^{e+1}}$$

lösbar ist, was die von Wahlin¹¹⁾ und Hensel¹²⁾ aufgestellte Bedingung ist.

Bei den Galoisschen Körpern kann man aber eine noch einfachere Bedingung erhalten. Aus (42) folgt sofort, daß e und somit auch e_0 durch $p - 1$ teilbar sein muß, und man kann daher

$$\eta_0 \equiv \tau^{a'} \pi_0^{\frac{e_0}{p-1}} \pmod{\mathfrak{P}^{\frac{e}{p-1}+1}}$$

¹⁰⁾ K. Hensel, Die multiplikative Darstellung der algebraischen Zahlen für den Bereich eines beliebigen Primteilers. Journ. f. Math. **146** (1916), S. 189—215, § 5.

¹¹⁾ G. E. Wahlin, The Equation $x^e - A = 0 (p)$. Journ. f. Math. **145** (1915), S. 114—138, § 3.

¹²⁾ Man vgl. die eben zitierte Arbeit von Hensel, § 5.

setzen, woraus man leicht nach Satz 6 zur Bestimmung von a'

$$a_0 \equiv -(p-1)a' \pmod{p^f-1}$$

erhält. Dies zeigt, daß a_0 durch $p-1$ teilbar sein muß, und daß

$$a' \equiv -\frac{a_0}{p-1} \pmod{k_0}$$

ist, wobei

$$k_0 = \frac{p^f-1}{p-1}$$

gesetzt wird. Wir haben daher bewiesen:

Satz 8. Die notwendige und hinreichende Bedingung, daß K eine p -te Einheitswurzel $(\text{mod } \mathfrak{P}^{e\alpha})$ enthält, ist

$$e_0 \equiv 0, \quad a_0 \equiv 0 \pmod{p-1},$$

wobei a_0 die in der Normalkongruenz (36) vorkommende Konstante ist.

Wie man leicht sieht, sind die p Einheitswurzeln im Regularkörper enthalten, und weiter

$$\varepsilon_0^{(0)} = 1, \quad \varepsilon_0^{(1)} \equiv 1 + \tau^{-\frac{a_0}{p-1}} \pi_0^{\frac{e_0}{p-1}}, \quad \varepsilon_0^{(2)} \equiv 1 + 2\tau^{-\frac{a_0}{p-1}} \pi_0^{\frac{e_0}{p-1}}, \dots$$

$$\left(\text{mod } \mathfrak{P}^{e(p-1)+1} \right).$$

Kapitel 2.

Hilfssätze über irreguläre Relativkörper vom Primzahlgrade.

§ 1.

Bestimmung der Relativedifferente.

Für die weiteren Untersuchungen sind einige Hilfssätze über irreguläre Körper vom Relativgrade p notwendig, und diese sollen nun hier aufgestellt werden.

Es sei k ein beliebiger Körper, und K ein Relativkörper zu k vom Relativgrade p , wobei wie früher p eine Primzahl bezeichnet. Weiter sei \mathfrak{p} ein beliebiger Primidealteiler von p in k , so daß man

$$(1) \quad p = \mathfrak{p}^e \mathfrak{q}, \quad N(\mathfrak{p}) = p^f$$

hat; die Zahl τ in k kann dann immer so gewählt werden, daß

$$\tau^{p^f} - 1 \equiv 0 \pmod{\mathfrak{p}^{e\alpha}},$$

und jede nicht durch \mathfrak{p} teilbare Zahl in k ist $(\text{mod } \mathfrak{p})$ kongruent einer Potenz von τ . Es soll nun vorausgesetzt werden, daß das Primideal \mathfrak{p} in K eine Primidealzerlegung von der Form

$$(2) \quad \mathfrak{p} = \mathfrak{P}^p, \quad N_k(\mathfrak{P}) = \mathfrak{p}$$

hat, so daß das Primideal \mathfrak{P} die Relativordnung p und Relativgrad 1 hat. Weiter sei π_0 eine Primzahl in bezug auf \mathfrak{p} in k , und π eine Primzahl in bezug auf \mathfrak{P} in K .

Es folgt dann leicht, daß die Zahl π für ein beliebig hohes α einer Kongruenz¹³⁾

$$(3) \quad f(x) = x^p + \pi_0 \alpha_{p-1} x^{p-1} + \dots + \pi_0 \alpha_0 \equiv 0 \pmod{\mathfrak{P}^\alpha}$$

genügt, wobei die Zahlen α_i alle in k enthalten sind, und wo weiter $\alpha_0 \equiv -1 \pmod{\mathfrak{p}}$ vorausgesetzt werden kann. Wenn der Koeffizient α_i genau durch \mathfrak{p}^{a_i} teilbar ist, kann man

$$\alpha_i \equiv \pi_0^{a_i} \beta_i \pmod{\mathfrak{p}^a}$$

setzen, wobei $\beta_i \not\equiv 0 \pmod{\mathfrak{p}}$ ist, und die Kongruenz (3) nimmt dann die Form

$$(4) \quad f(x) \equiv x^p + \pi_0^{a_{p-1}+1} \beta_{p-1} x^{p-1} + \dots + \pi_0^{a_1+1} \beta_1 x + \pi_0 \beta_0 \equiv 0 \pmod{\mathfrak{P}^\alpha}$$

an.

Bei diesen Untersuchungen spielt die Relativdifferente \mathfrak{d} von K in bezug auf k eine wichtige Rolle. Aus einem allgemeinen Satze folgt, daß \mathfrak{d} genau durch \mathfrak{P}^{p-1+e} teilbar ist, wobei ich e die *relative Supplementzahl* von \mathfrak{P} in bezug auf k genannt habe. Aus der Dedekind-Henselschen Ungleichung für Relativkörper folgt dann

$$(5) \quad 1 \leq e \leq ep.$$

Wie wir sehen werden, fallen die folgenden Untersuchungen in zwei Teile, je nachdem e den Maximalwert pe hat, oder $e < pe$ ist.

Die Zahl $f'(\pi)$ enthält genau dieselbe Potenz von \mathfrak{P} wie die Relativdifferente \mathfrak{d} , und man kann daraus die Supplementzahl e bestimmen. Differentiiert man ein Glied $\pi_0^{a_i+1} \beta_i x^i$, in (4) erhält man für $x = \pi$

$$i \pi_0^{a_i+1} \beta_i \pi^{i-1},$$

und wenn $i < p$ ist, wird diese Zahl genau durch $\mathfrak{P}^{p-1+a_i p+i}$ teilbar. Das erste Glied x^p ergibt bei der Differentiation die Zahl $p \pi^{p-1}$, welche genau durch \mathfrak{P}^{p-1+pe} teilbar ist.

Wenn daher für alle i

$$(6) \quad a_i p + i > pe \quad (i = 1, 2, \dots, p-1),$$

so wird e den Maximalwert ep haben. Wenn dagegen die Ungleichung (6) nicht für alle i richtig bleibt, so sei $a_r p + r$ der kleinste Wert von $a_i p + i$. Dann ist offenbar

$$(7) \quad e = a_r p + r.$$

Wir werden nun die beiden Fälle getrennt behandeln.

¹³⁾ Für die folgenden Resultate muß ich auf Ore, I u. II verweisen.

§ 2.

Der binomische Fall: $\varrho = ep$.

Wenn $\varrho = ep$ ist, so soll gezeigt werden, daß man die Primzahl π so wählen kann, daß die Kongruenz (4) sich auf eine einfache Normalform reduziert.

Nach (6) genügt die Zahl π in diesem Falle der Kongruenz

$$(8) \quad \pi^p + \pi_0 \beta_0 \equiv 0 \pmod{\mathfrak{P}^{p(e+1)+1}},$$

und es sei $\mathfrak{P}^{p(e+1)+\gamma}$ die höchste Potenz von \mathfrak{P} , wofür diese Kongruenz besteht, so daß man

$$(9) \quad \pi^p + \pi_0 \beta_0 \equiv \omega_0 \pi^{p(e+1)+\gamma} \pmod{\mathfrak{P}^{p(e+1)+\gamma+1}}$$

setzen kann, wobei $\omega_0 \not\equiv 0 \pmod{\mathfrak{P}}$ als eine Zahl in k vorausgesetzt werden kann.

Wir zeigen nun, daß man eine neue Zahl

$$(10) \quad \pi' = \pi(1 + \omega_1 \pi^\gamma)$$

in K so bestimmen kann, daß π' auch für die höhere Potenz $\mathfrak{P}^{(e+1)p+\gamma+1}$ einer binomischen Kongruenz (8) genügt, wobei aber im allgemeinen die Konstante β_0 einen anderen Wert hat.

Um dies zu beweisen, multiplizieren wir die Kongruenz (9) mit $(1 + \omega_1 \pi^\gamma)^p$ und erhalten nach (10)

$$\pi'^p + \pi_0 \beta_0 (1 + \omega_1 \pi^\gamma)^p \equiv \omega_0 \pi^k \pmod{\mathfrak{P}^{k+1}},$$

wobei der Kürze wegen $k = (e+1)p + \gamma$ geschrieben worden ist. Führt man hier die Potenzierung aus, kommt

$$\pi'^p + \pi_0 \beta_0 + \pi_0 \beta_0 \omega_1^p \pi^{\gamma p} + \omega_1 \pi_0 p \beta_0 \pi^\gamma \equiv \omega_0 \pi^k \pmod{\mathfrak{P}^{k+1}}.$$

Da die beiden Glieder $\omega_0 \pi^k$ und $\omega_1 \pi_0 p \beta_0 \pi^\gamma$ dieselbe Potenz von \mathfrak{P} enthalten, kann man offenbar ω_1 in k so bestimmen, daß

$$\omega_1 \pi_0 \beta_0 p \pi^\gamma \equiv \omega_0 \pi^k \pmod{\mathfrak{P}^{k+1}},$$

und wir erhalten daher

$$\pi'^p + \pi_0 \beta_0 + \pi_0 \beta_0 \omega_1^p \pi^{\gamma p} \equiv 0 \pmod{\mathfrak{P}^{k+1}},$$

oder da nach (9)

$$\pi^p \equiv -\pi_0 \beta_0 \pmod{\mathfrak{P}^k}$$

ist, erhält man sicher

$$\pi_0 \beta_0 \omega_1^p \pi^{\gamma p} \equiv \pm \pi_0^{\gamma+1} \beta_0^{\gamma+1} \omega_1^p \pmod{\mathfrak{P}^{k+1}},$$

wo das rechtsstehende Glied in k enthalten ist, und die gewünschte Kongruenz für π' wird daher

$$\pi'^p + \pi_0 \beta_0 (1 \pm \omega_1^p \pi_0^\gamma \beta_0^\gamma) \equiv 0 \pmod{\mathfrak{P}^{k+1}}.$$

Wie wir sehen, können wir in der Kongruenz für π' dasselbe β_0 wie in (8) anwenden, sobald

$$p + \gamma p > p(e + 1) + \gamma$$

oder

$$\gamma \geq \left[\frac{pe}{p-1} \right] + 1 = \left[\frac{e}{p-1} \right] + 1$$

ist.

Satz 1. *Wenn die Supplementzahl ϱ den Maximalwert pe hat, kann man für ein beliebig hohes α eine solche Primzahl π in K finden, daß π einer binomischen Kongruenz*

$$(11) \quad x^p + \pi_0 \beta_0 \equiv 0 \pmod{\mathfrak{P}^\alpha}$$

genügt.

Man kann dabei auch annehmen, daß β_0 die Form

$$\pi_0 \beta_0 = \pi_0 + \alpha_2 \pi_0^2 + \dots + \alpha_\nu \pi_0^\nu$$

hat, wobei $\nu \leq e + 1 + \frac{1}{p} \left[\frac{e}{p-1} \right]$ ist.

Zuletzt soll auch noch untersucht werden, ob man mehrere Primzahlen π in K bestimmen kann, welche binomische Normalkongruenzen genügen, d. h. ob man eine andere Primzahl π' in K finden kann, so daß π' einer binomischen Kongruenz

$$(12) \quad x^p + \pi_0 \beta'_0 \equiv 0 \pmod{\mathfrak{P}^\alpha}$$

genügt.

Man kann dann

$$(13) \quad \pi' = \pi(\alpha_0 + A\pi^\gamma)$$

setzen, wobei $\alpha_0 \not\equiv 0 \pmod{p}$ eine Zahl in k ist, während $A \not\equiv 0 \pmod{\mathfrak{P}}$ eine Zahl in K ist. Weiter kann man dabei voraussetzen, daß γ nicht durch p teilbar ist. Denn wäre $\gamma = p\gamma_1$, würde man aus der Kongruenz (11) für π

$$\pi^\gamma \equiv \pm \beta_0^{\gamma_1} \pi_0^{\gamma_1} \pmod{\mathfrak{P}^\alpha}$$

erhalten, und wenn dies in (13) eingesetzt würde, könnte man ein neues α_0 so wählen, daß γ einen höheren Wert habe.

Nimmt man nun in (13) die p -te Potenz von π' , erhält man

$$\pi'^p \equiv -\pi_0 \beta'_0 \equiv -\beta_0 \pi_0 (\alpha_0 + A\pi^\gamma)^p \pmod{\mathfrak{P}^\alpha}$$

und folglich ist

$$(14) \quad (\alpha_0 + A\pi^\gamma)^p \equiv \alpha_0^p + A^p \pi^{\gamma p} + p\alpha_0^{p-1} A\pi^\gamma + \dots$$

eine Zahl in $k \pmod{\mathfrak{P}^\alpha}$.

Wenn nun

$$\pi^\gamma A = \gamma_0 \pi^\gamma + \gamma_1 \pi^{\gamma+1} + \dots \pmod{\mathfrak{P}^a}$$

ist, so wird

$$\pi^{\gamma p} A^p \equiv \gamma_0^p \pi^{\gamma p} + \gamma_1^p \pi^{(\gamma+1)p} + \dots \pmod{\mathfrak{P}^{ep+p\gamma+1}}$$

oder wenn man aus (11) den Wert von π^p einführt, erhält man

$$\pi^{\gamma p} A^p \equiv \mu_0 \pmod{\mathfrak{P}^{ep+\gamma p+1}},$$

wobei μ_0 eine Zahl in k ist. Aus (14) erhält man dann, daß die Zahl

$$p\alpha_0^{p-1} A\pi^\gamma + \dots \pmod{\mathfrak{P}^{ep+p\gamma+1}}$$

kongruent einer Zahl in k sein muß, und dies ist sicher nicht möglich, indem γ nach unserer Voraussetzung nicht durch p teilbar ist. Nach (13) erhalten wir folglich $\pi' = \alpha_0 \pi$.

Satz 2. *Es sei $\varrho = ep$ und weiter π eine Primzahl in K , welche der Normalkongruenz (11) genügt. Jede andere Primzahl π' , welche einer binomischen Kongruenz genügt, hat dann die Form $\pi' = \alpha_0 \pi$, wo α_0 eine Zahl in k ist, und die entsprechende Normalkongruenz ist daher*

$$x^p + \alpha_0^p \beta_0 \pi_0 \equiv 0 \pmod{\mathfrak{P}^a}.$$

§ 3.

Der trinomische Fall: $\varrho < pe$.

Wir gehen nun zu dem schwierigeren Falle über, wo

$$1 \leq \varrho < pe$$

ist, und wie wir sehen werden, sind die Verhältnisse dabei etwas verwickelter, indem die Normalkongruenz nicht binomisch, sondern trinomisch wird.

Wenn ϱ die Form (7) hat, so folgt sofort aus (4), wenn $x = \pi$ gesetzt wird

$$\pi^p + \pi_0^{a_{r-1}} \beta_r \pi^r + \beta_0 \pi_0 \equiv 0 \pmod{\mathfrak{P}^{p(a_{r+1})+r+1}},$$

oder wenn

$$\alpha_r = c, \quad cp + r = \varrho, \quad \beta_r \equiv \tau^b \pmod{\mathfrak{P}}$$

gesetzt wird, kann man die Kongruenz in der Form

$$(15) \quad \pi^p + \pi_0^{c+1} \tau^b \pi^r + \beta_0 \pi_0 \equiv 0 \pmod{\mathfrak{P}^{p+e+1}}$$

schreiben.

Es sei nun

$$k = p(c+1) + r + \gamma = p + \varrho + \gamma$$

der größte Exponent, wofür die Kongruenz (15) besteht, so daß man

$$(16) \quad \pi^p + \pi_0^{e+1} \tau^b \pi^r + \beta_0 \pi_0 \equiv \omega_0 \pi^k \pmod{\mathfrak{P}^{k+1}}$$

setzen kann. Es soll dann wie früher gezeigt werden, daß man eine neue Primzahl

$$(17) \quad \pi' = \pi(1 + \omega_1 \pi^\gamma)$$

so bestimmen kann, daß für π' eine trinomische Kongruenz (15) auch für den höheren Modul \mathfrak{P}^{k+1} bestehen wird, wobei aber im allgemeinen die Konstante β_0 einen anderen Wert haben wird. Wenn man nämlich die Kongruenz (16) mit $(1 + \omega_1 \pi^\gamma)^p$ multipliziert, erhält man nach (17)

$$(18) \quad \pi'^p + \pi_0^{e+1} \tau^b \pi'^r (1 + \omega_1 \pi^\gamma)^{p-r} + \pi_0 \beta_0 (1 + \omega_1 \pi^\gamma)^p \equiv \omega_0 \pi^k \pmod{\mathfrak{P}^{k+1}}.$$

Wir werden nun diese Kongruenz wesentlich reduzieren; es ist offenbar

$$(19) \quad \pi_0^{e+1} \tau^b \pi'^r (1 + \omega_1 \pi^\gamma)^{p-r} \equiv \pi_0^{e+1} \tau^b \pi'^r + \pi_0^{e+1} \omega_1 (p-r) \tau^b \pi^{\gamma+r} \pmod{\mathfrak{P}^{k+1}}.$$

und wir werden nun ω_1 dadurch bestimmen, daß

$$(20) \quad \pi_0^{e+1} \omega_1 (p-r) \tau^b \pi^{\gamma+r} \equiv \omega_0 \pi^k \pmod{\mathfrak{P}^{k+1}}$$

gesetzt wird. Dies ist natürlich immer möglich, indem beide Seiten dieselbe Potenz von \mathfrak{P} enthalten. Wird nun (20) in (19) eingeführt, erhält man nach (18) die einfachere Kongruenz für π'

$$(21) \quad \pi'^p + \pi_0^{e+1} \tau^b \pi'^r + \pi_0 \beta_0 (1 + \omega_1 \pi^\gamma)^p \equiv 0 \pmod{\mathfrak{P}^{k+1}}.$$

Nach der Voraussetzung über ϱ ist aber

$$\pi_0 \beta_0 (1 + \omega_1 \pi^\gamma)^p \equiv \pi_0 \beta_0 + \pi_0 \beta_0 \omega_1^p \pi^{\gamma p} \pmod{\mathfrak{P}^{k+1}},$$

und nach (16) folgt leicht

$$\pi_0 \pi^{\gamma p} \equiv \pm \beta_0^\gamma \pi_0^{\gamma+1} \pmod{\mathfrak{P}^{k+1}},$$

so daß

$$\pi_0 \beta_0 (1 + \omega_1 \pi^\gamma)^p \equiv \pi_0 \beta_0 (1 \pm \omega_1^p \beta_0^\gamma \pi_0^\gamma) \pmod{\mathfrak{P}^{k+1}}$$

wird. Nach (21) erhält man daraus die gewünschte trinomische Kongruenz für π'

$$(22) \quad \pi'^p + \pi_0^{e+1} \tau^b \pi'^r + \pi_0 \beta_0 (1 \pm \omega_1^p \beta_0^\gamma \pi_0^\gamma) \equiv 0 \pmod{\mathfrak{P}^{k+1}}.$$

Wie man sieht, kann man in (22) dasselbe β_0 wie in (15) anwenden, sobald

$$\gamma p + p > p + \varrho + \gamma$$

oder

$$\gamma \geq \left[\frac{\varrho}{p-1} \right] + 1$$

ist

Satz 3. Wenn die Supplementzahl die Form

$$\varrho = cp + r < ep, \quad 1 \leq r \leq p - 1$$

hat, kann man für ein beliebig hohes α eine Primzahl π in K bestimmen, welche einer trinomischen Normalkongruenz

$$(23) \quad x^p + \pi_0^{c+1} \tau^b x^r + \pi_0 \beta_0 \equiv 0 \pmod{\mathfrak{P}^\alpha}$$

genügt.

Dabei kann man auch wie früher annehmen, daß $\pi_0 \beta_0$ die Form

$$\pi_0 \beta_0 = \pi_0 + \alpha_2 \pi_0^2 + \dots + \alpha_\nu \pi_0^\nu$$

hat, wobei $\nu \leq 1 + \frac{1}{p} \left[\frac{\varrho p}{p-1} \right]$ ist.

Wir werden nun auch in diesem Falle untersuchen, welche andere Primzahlen in K einer trinomischen Normalkongruenz genügen können, und welche Form die entsprechende Kongruenz hat. Diese Untersuchung ist für die spätere Anwendung auf Galoissche Körper besonders wichtig.

Wir nehmen also an, daß eine Kongruenz

$$(24) \quad x^p + \pi_0^{c+1} \tau^{b'} \gamma'_0 x^r + \beta'_0 \pi_0 \equiv 0 \pmod{\mathfrak{P}^\alpha}$$

für ein beliebig hohes α lösbar sei; dabei kann man auch voraussetzen, daß

$$\beta'_0 \equiv \gamma'_0 \equiv 1 \pmod{\mathfrak{p}}$$

ist. Wie man aber leicht sieht, ist es in diesem Falle möglich, eine Zahl γ_0 in k so zu bestimmen, daß $\gamma_0^{p-r} \equiv \gamma'_0 \pmod{\mathfrak{P}^\alpha}$, und wenn man dann in (24) $\gamma_0 x$ an der Stelle von x schreibt, erhält man eine Kongruenz von der einfacheren Form

$$(25) \quad x^p + \pi_0^{c+1} \tau^{b'} x^r + \beta'_0 \pi_0 \equiv 0 \pmod{\mathfrak{P}^\alpha}.$$

Wir werden also untersuchen, wann diese Kongruenz eine Lösung besitzt, wenn man schon weiß, daß die Kongruenz (23) in K lösbar ist, und die Wurzel π hat.

Es folgt leicht, daß eine Lösung von (25) die Form

$$(26) \quad \pi' = \pi(1 + A\pi^\nu) \pmod{\mathfrak{P}^\alpha}, \quad \nu \geq 1$$

haben muß, wobei $A \not\equiv 0 \pmod{\mathfrak{P}}$ vorausgesetzt wird. Daraus erhält man sofort

$$(27) \quad \pi'^p \equiv \pi^p + A^p \pi^{(\nu+1)p} \pmod{\mathfrak{P}^{e(p+\nu)}}.$$

Wird hier

$$A^p \pi^{\nu+1} \equiv \lambda_0 \pi^{\nu+1} + \lambda_1 \pi^{\nu+2} + \dots \pmod{\mathfrak{P}^\alpha}$$

gesetzt, so leitet man weiter die Kongruenz

$$A^p \pi^{(\gamma+1)p} \equiv \lambda_0^p \pi^{p(\gamma+1)} + \lambda_1^p \pi^{p(\gamma+1)} + \dots \pmod{\mathfrak{P}^{e+p\gamma+1}}$$

ab, und da sicher $p\gamma + 1 \geq p + \gamma$ ist, besteht diese Kongruenz auch $\pmod{\mathfrak{P}^{e+p\gamma+1}}$. Nach der Definition von π ist aber

$$\begin{aligned} \pi^{p(\gamma+1)} &\equiv \pm \pi_0^{\gamma+1} \beta_0^{\gamma+1} \pmod{\mathfrak{P}^{p+e+(\gamma+1)p}} \\ \pi^{p(\gamma+2)} &\equiv \pm \pi_0^{\gamma+2} \beta_0^{\gamma+2} \pmod{\mathfrak{P}^{p+e+(\gamma+1)p}} \\ &\dots \end{aligned}$$

so daß man zuletzt für $A^p \pi^{(\gamma+1)p}$ die Kongruenz

$$(28) \quad A^p \pi^{(\gamma+1)p} \equiv \pi_0^{\gamma+1} \mu_0 \pmod{\mathfrak{P}^{p+e+\gamma+1}}$$

erhält, wobei μ_0 eine Zahl in k ist.

Wird nun (28) in (27) eingesetzt, so folgt

$$\pi'^p \equiv \pi^p + \pi_0^{\gamma+1} \mu_0 \pmod{\mathfrak{P}^{p+e+\gamma+1}},$$

und wenn hier die Werte von π'^p und π^p nach (25) und (23) eingeführt werden, gewinnt man die wichtige Kongruenz

$$(29) \quad \beta'_0 - \beta_0 \equiv \pi_0^c (\tau^b \pi^r - \tau^{b'} \pi'^r) - \pi_0^\gamma \mu_0 \pmod{\mathfrak{P}^{e+\gamma+1}},$$

wodurch die Verbindung zwischen β_0 und β'_0 bestimmt ist.

Die Kongruenz (29) zeigt, daß die Zahl

$$\tau^b \pi^r - \tau^{b'} \pi'^r \pmod{\mathfrak{P}^{r+\gamma+1}}$$

kongruent einer Zahl in k sein muß, und dies ist nach (26) nur dann möglich, wenn $b = b'$ ist. Weiter muß daher

$$\pi^r (1 + A \pi^\gamma)^r - \pi^r \pmod{\mathfrak{P}^{r+\gamma+1}}$$

eine Zahl in k sein, woraus man sofort den Schluß zieht, daß $\pi^{r+\gamma} A \pmod{\mathfrak{P}^{r+\gamma+1}}$ eine Zahl in k ist. Dies ist aber offenbar nur dann möglich, wenn die Bedingung $r + \gamma \equiv 0 \pmod{p}$ erfüllt ist.

Wird nun $b = b'$, $r + \gamma = p\gamma_1$ in (29) eingeführt, so folgt nach einer leichten Umformung

$$\beta'_0 \equiv \beta_0 - \pi_0^\gamma \mu_0 - \pi_0^{e+\gamma_1} \nu_0 \pmod{\mathfrak{P}^{e+\gamma+1}},$$

wobei ν_0 eine Zahl in k ist. Wie man leicht sieht, ist hier $\gamma > c + \gamma_1$ solange $\gamma < \left[\frac{e}{p-1} \right]$ ist, und nach einer früheren Bemerkung kann man immer dies voraussetzen.

Satz 4. Wenn eine trinomische Normalkongruenz

$$(30) \quad x^p + \pi_0^{c+1} \tau^{b'} x^r + \beta'_0 \pi_0 \equiv 0 \pmod{\mathfrak{P}^a}$$

gleichzeitig mit der Kongruenz (23) in K lösbar sein soll, so muß $b' = b$ sein, und wenn

$$\beta'_0 - \beta_0 = \pi_0^\gamma \alpha_0, \quad \alpha_0 \not\equiv 0 \pmod{p}, \quad \gamma < \left[\frac{e}{p-1} \right]$$

ist, so muß die Bedingung $\gamma + r \equiv 0 \pmod{p}$ erfüllt sein. Die Lösung der Kongruenz (30) wird dann die Form $\pi' = \pi(1 + A\pi^r)$ haben.

§ 4.

Anwendung auf relativ-zyklische Körper.

Im folgenden werden wir diese Resultate überhaupt nur in dem Falle anwenden, wo der Körper K ein relativ-zyklischer in bezug auf k ist, und es soll daher dieser Fall mehr eingehend behandelt werden.

Wir betrachten zunächst den binomischen Fall, wo $q = ep$ ist. Wenn dann K in bezug auf k relativ-zyklisch ist, muß die Normalkongruenz

$$(31) \quad x^p + \beta_0 \pi_0 \equiv 0 \pmod{\mathfrak{P}^\alpha}$$

außer π noch $p - 1$ andere Wurzeln in K besitzen. Die Bedingung dafür können wir aber leicht ableiten. Wenn nämlich π' eine von π verschiedene Wurzel von (31) ist, so folgt sofort

$$\left(\frac{\pi'}{\pi} \right)^p \equiv 1 \pmod{\mathfrak{P}^\alpha},$$

und K muß daher die p -ten Einheitswurzeln $(\text{mod } \mathfrak{P}^\alpha)$ enthalten. Wir haben daher bewiesen:

Satz 5. Die notwendige und hinreichende Bedingung, daß die Kongruenz (31) p Wurzeln in K besitzt, ist, daß K eine primitive p -te Einheitswurzel $(\text{mod } \mathfrak{P}^\alpha)$ enthält.

Wenn ε eine Lösung der Kongruenz

$$x^p - 1 \equiv 0 \pmod{\mathfrak{P}^\alpha}$$

ist, so sind die Wurzeln von (31) durch

$$\pi, \varepsilon \pi, \varepsilon^2 \pi, \dots, \varepsilon^{p-1} \pi$$

gegeben. Wie in Kap. 1, § 6 folgt leicht, daß die Zahl e durch $p - 1$ teilbar sein muß, so daß

$$\frac{e}{p-1} = \frac{ep}{p-1}$$

eine ganze Zahl ist. Weiter soll noch bemerkt werden, daß nach Kap. 2, § 2 kann die Zahl $\pi_0 \beta_0$ in (31) eindeutig in der Form

$$(32) \quad \pi_0 \beta_0 = \pi_0 + \tau^{\alpha_1} \pi_0^2 + \dots + \tau^{\frac{\alpha}{p-1} + 1} \pi_0^{\frac{e}{p-1} + 1}$$

geschrieben werden, wobei natürlich einige Glieder fehlen können.

Zuletzt soll noch der trinomische Fall behandelt werden. Wir schreiben im folgenden die Normalkongruenz in einer leicht abgeänderten Form

$$(33) \quad f(x) = x^p - \frac{1}{r} \tau^b \pi_0^{e+1} x^r - \pi_0 \beta_0 \equiv 0 \pmod{\mathfrak{P}^\alpha},$$

und es soll daher untersucht werden, wann diese Kongruenz p Wurzeln in K hat.

Wenn π' eine von π verschiedene Lösung ist, kann man $\pi' = \pi + A$ setzen, und erhält zur Bestimmung von A die Kongruenz

$$(34) \quad A^{p-1} + \frac{f^{(p-1)}(\pi)}{(p-1)!} A^{p-2} + \dots + \frac{f'(\pi)}{1!} \equiv 0 \pmod{\mathfrak{P}^\alpha},$$

und diese Kongruenz muß dann $p-1$ Wurzeln in K besitzen. In (34) ist allgemein ein Koeffizient $f^{(i)}(\pi)$ durch \mathfrak{P}^{p+e-i} oder eine höhere Potenz von \mathfrak{P} teilbar, während $f'(\pi)$ genau durch \mathfrak{P}^{p+e-1} teilbar ist.

Es folgt daher, daß A durch \mathfrak{P} teilbar sein muß, und wenn A genau die Potenz \mathfrak{P}^ν enthält, so zeigt man leicht, daß ν aus der Bedingung

$$\nu(p-1) = p-1 + \varrho$$

bestimmt wird. Daraus folgt, daß die Supplementzahl ϱ durch $p-1$ teilbar ist, so daß man

$$\nu = 1 + \frac{\varrho}{p-1}$$

setzen kann.

Aus (34) erhält man dann

$$(35) \quad A^{p-1} + f'(\pi) \equiv 0 \pmod{\mathfrak{P}^{p+e}},$$

und wie man durch Induktion zeigen kann, wird die Kongruenz (34) für alle α in K lösbar, und hat $p-1$ Wurzeln, sobald die Kongruenz (35) diese Eigenschaft hat.

Es ist aber nach (33)

$$f'(\pi) \equiv -\tau^b \pi_0^{e-1} \pi^{r-1} \equiv -\tau^b \pi^{p+e-1} \pmod{\mathfrak{P}^{p+e}},$$

und wenn man weiter

$$A \equiv \tau^\lambda \pi^\nu \pmod{\mathfrak{P}^{\nu+1}}$$

setzt, nimmt die Kongruenz (35) die einfachere Form

$$\tau^{(p-1)\lambda} \pi^{p+e-1} \equiv \tau^b \pi^{p+e-1} \pmod{\mathfrak{P}^{p+e}}$$

an. Wenn aber diese Kongruenz lösbar sein soll, muß b durch $p-1$ teilbar sein, und die $p-1$ Werte von λ sind durch

$$\lambda \equiv \frac{b}{p-1} \pmod{\frac{p-1}{p-1}}$$

bestimmt.

Satz 6. Die notwendige und hinreichende Bedingung, daß die Kongruenz

$$x^p - \frac{1}{r} \tau^b \pi_0^{c+1} x^r - \pi_0 \beta_0 \equiv (\text{mod } \mathfrak{P}^a)$$

p Wurzeln in K haben soll, ist durch

$$\varrho = c p + r \equiv 0, \quad b \equiv 0 (\text{mod } p - 1)$$

gegeben.

Wie man leicht sieht, haben dann die Wurzeln die Form

$$(36) \quad \pi, \pi + \tau^{\frac{b}{p-1}} \pi^\nu, \pi + 2 \tau^{\frac{b}{p-1}} \pi^\nu, \dots (\text{mod } \mathfrak{P}^{\nu+1}),$$

wo wie früher $\nu = 1 + \frac{\varrho}{p-1}$ gesetzt worden ist. Weiter folgt auch wie früher, daß man in der Normalkongruenz β_0 in der Form (32) darstellen kann.

(Eingegangen am 23. 1. 1928.)

Über die Automorphismen einer endlichen Abelschen Gruppe.

Von

Kenjiro Shoda in Göttingen.

In der folgenden Arbeit soll die Automorphismengruppe einer endlichen Abelschen Gruppe untersucht werden¹⁾. Die Gesamtheit der einstufigen und mehrstufigen isomorphen Abbildungen einer endlichen Abelschen Gruppe \mathfrak{A} in sich (d. h. die der Automorphismen von \mathfrak{A} im üblichen Sinne und der Abbildungen von \mathfrak{A} in ihre homomorphen Untergruppen, die ich auch Automorphismen nenne) bildet bei passender Definition der Verknüpfungen einen Ring (Automorphismenring)²⁾. Ein Automorphismus heißt eigentlich oder uneigentlich, je nachdem die Abbildung einstufig oder mehrstufig ist.

¹⁾ Dieses Thema verdanke ich Fr. E. Noether. Ich danke ihr für eine Reihe von Bemerkungen, die sie mir mitgeteilt hat.

²⁾ Nachträglich bemerke ich, daß der Begriff des Automorphismenrings sich schon in Arbeiten von Herrn A. Chatelet findet: Comptes Rendus 175, S. 85; 177, S. 729. Vgl. seine zusammenfassende Darstellung, Les groupes abéliens finis et les modules de points entiers, Lille 1925, besonders Kap. IV und VII. Dort wird der Automorphismenring eingeführt als eine Menge von ganzzahligen Matrizen X , die modulo einer durch die Abelsche Gruppe eindeutig bestimmten Matrix F betrachtet werden. Dabei durchläuft X alle mit F semi-vertauschbaren Matrizen, d. h. $FX = X'F$, wobei die ganzzahlige Matrix X' nicht notwendig zur Menge der X zu gehören braucht. Bei der in unserer Arbeit zugrundegelegten Elementarteiler-Normalform wird $F = \begin{pmatrix} p^{e_1} & & & \\ & p^{e_2} & & \\ & & \ddots & \\ & & & p^{e_m} \end{pmatrix}$

und die an (1a) und (1b) § 2 anschließende Formel $F_d X = X' F_d$ geht für $d=0$ in die Definition von Herrn A. Chatelet über. Vgl. die entsprechende Definition des Rings, der aus der Gesamtheit der mit einer Matrix vertauschbaren Matrizen besteht. (Frobenius, Über die mit einer Matrix vertauschbaren Matrizen, Sitzungsberichte der Akad. der Wiss. zu Berlin, 1910, § 1.) Dort hat Frobenius die mit der Elementarteiler-Normalform der charakteristischen Matrix semi-vertauschbaren ganzen Matrizen betrachtet. Zum Zusammenhang dieser Definition mit dem Automorphismenring vgl. den Schluß der Einleitung der vorliegenden Arbeit.

Die Gesamtheit der eigentlichen Automorphismen von \mathfrak{A} bildet eine Gruppe (Automorphismengruppe). Wenn man eine Abelsche Gruppe \mathfrak{A} als das direkte Produkt der Abelschen Gruppen \mathfrak{A}_i darstellt, deren Ordnungen Potenzen verschiedener Primzahlen sind, so ist der Automorphismenring \mathfrak{o} bzw. die Automorphismengruppe \mathfrak{G} von \mathfrak{A} die direkte Summe bzw. das direkte Produkt derjenigen der einzelnen Abelschen Gruppen \mathfrak{A}_i . Daher betrachte ich eine Abelsche Gruppe \mathfrak{A} , deren Ordnung eine Primzahlpotenz ist (§ 1). Man kann den Automorphismenring \mathfrak{o} durch zwei den Reihen aus charakteristischen Untergruppen von \mathfrak{A} entsprechende Reihen von zweiseitigen Idealen in \mathfrak{o} genau untersuchen (§ 2). Vermöge dieser Reihen von Idealen kann man durch Strahlbildung (d. h. durch Durchschnittbildung der Automorphismengruppe mit der jeweiligen Restklasse des Einheits-elementes) alle Tatsachen über den Ring in der Theorie der (eigentlichen) Automorphismengruppe \mathfrak{G} übertragen. Dabei entspricht einem Ideale \mathfrak{r} in \mathfrak{o} ein Normalteiler \mathfrak{R} von \mathfrak{G} und dem Restklassenring $\mathfrak{o}/\mathfrak{r}$ die Faktorgruppe $\mathfrak{G}/\mathfrak{R}$. Ist ferner das Produkt zweier Elemente (Klassen) in einem Ideale (Restklassenringe) stets gleich Null (Nullklasse), so ist der entsprechende Normalteiler (die Faktorgruppe) kommutativ und mit einem Unterideale (Unterideal des Restklassenrings) — diese aufgefaßt als Abelsche Gruppe in bezug auf Addition — isomorph. Dadurch erhalte ich zwei Reihen von Normalteilern von \mathfrak{G} , die den Reihen von charakteristischen Untergruppen von \mathfrak{A} entsprechen. Die Struktur von \mathfrak{G} wird durch diejenige der Normalteiler genau erklärt (Satz 7). Dabei erhalte ich den Satz, daß die Auflösbarkeit der Gruppe \mathfrak{G} nur von der von Kongruenzgruppen modulo einer Primzahl abhängt. Herr Speiser³⁾ hat nur eine Reihe von charakteristischen Untergruppen und nur die Automorphismengruppe betrachtet, um einen Satz zu beweisen, der in den weitergehenden Resultaten (Satz 7) enthalten ist, die ich durch Benützung des Automorphismenrings erhalte.

Die Untersuchungen dieser Arbeit sind auf die Krullsche Elementarteilergruppe und daher in die Theorie der mit einer Matrix vertauschbaren Matrizen übertragbar⁴⁾. Dazu braucht man nur die Primzahl durch ein in einem beliebigen Körper irreduzibles Polynom und jeweils die Anzahl der Elemente durch den Rang ersetzen. Diese Übertragbarkeit beruht auf der Tatsache, daß der Polynombereich einer Veränderlichen in einem Körper nur Hauptideale enthält.

³⁾ Speiser, Theorie der Gruppen von endlicher Ordnung, 2. Aufl. § 43. Vgl. Anmerkung ¹⁶⁾.

⁴⁾ Krull, Theorie und Anwendung der verallgemeinerten Abelschen Gruppen, Sitzungsberichte der Heidelberger Akad. der Wiss. 1926. Vgl. meine Arbeit, Über die mit einer Matrix vertauschbaren Matrizen, die in der Math. Zeitschrift erscheinen wird.

§ 1.

Automorphismen. ⁵⁾

Es sei \mathfrak{A} eine endliche Abelsche Gruppe, deren Elemente durch Addition verknüpft werden. \mathfrak{A} sei die direkte Summe $\mathfrak{A}_1 + \mathfrak{A}_2 + \dots + \mathfrak{A}_n$ von n zyklischen Gruppen \mathfrak{A}_κ von den Ordnungen a_κ , $\kappa = 1, 2, \dots, n$. Ist \mathfrak{B} eine (homomorphe) Untergruppe von \mathfrak{A} , so muß \mathfrak{B} die Summe $(\mathfrak{B}_1, \mathfrak{B}_2, \dots, \mathfrak{B}_n)$ sein, wo \mathfrak{B}_κ mit \mathfrak{A}_κ homomorph ist, also die Ordnung von \mathfrak{B}_κ ein Teiler der von \mathfrak{A}_κ ist. Man kann jedes Element von \mathfrak{A} bzw. \mathfrak{B} in der Gestalt

$$x_1 r_1 + x_2 r_2 + \dots + x_n r_n \quad \text{bzw.} \quad y_1 r_1^* + y_2 r_2^* + \dots + y_n r_n^*$$

ausdrücken, wo r_1, r_2, \dots, r_n linear unabhängig sind und x_κ modulo a_κ betrachtet werden soll. Da r_κ^* ein Element von \mathfrak{A} ist, so kann man r_κ^* durch r_1, r_2, \dots, r_n ausdrücken.

$$(1) \quad r_\kappa^* = g_{\kappa 1} r_1 + g_{\kappa 2} r_2 + \dots + g_{\kappa n} r_n,$$

Wegen der Homomorphie von \mathfrak{A}_κ und \mathfrak{B}_κ — $a_\kappa r_\kappa^* = 0$ — muß

$$a_\kappa g_{\kappa 1} r_1 + a_\kappa g_{\kappa 2} r_2 + \dots + a_\kappa g_{\kappa n} r_n = 0$$

sein. Aus der Unabhängigkeit von r_κ folgt $a_\kappa g_{\kappa \lambda} \equiv 0 \pmod{a_\lambda}$. Ist $d_{\kappa \lambda}$ der größte gemeinsame Teiler von a_κ und a_λ , so ist

$$(2) \quad g_{\kappa \lambda} = \frac{a_\lambda}{d_{\kappa \lambda}} h_{\kappa \lambda}$$

Die Abbildung (1) nennt man einen Automorphismus von \mathfrak{A} , wenn sie der Bedingung (2) genügt, und zwar einen eigentlichen Automorphismus, wenn \mathfrak{B} mit \mathfrak{A} übereinstimmt. Die Gesamtheit der Automorphismen bildet einen Ring (*Automorphismenring*) \mathfrak{o} , dessen Addition und Multiplikation durch die der Matrix $(g_{\kappa \lambda})$ (1) mit der Bedingung (2) definiert werden.

Ist speziell a_κ mit a_λ teilerfremd, so ist $g_{\kappa \lambda} r_\lambda$ gleich Null. Wenn man also \mathfrak{A} als die direkte Summe von zyklischen Gruppen darstellt, deren Ordnungen Potenzen von Primzahlen sind, so wird auch der Automorphismenring direkte Summe. Zusammenfassend hat man

Satz 1. *Ist eine Abelsche Gruppe \mathfrak{A} die direkte Summe von Abelschen Gruppen, deren Ordnungen Potenzen von einander verschiedener Primzahlen sind, so ist der Automorphismenring von \mathfrak{A} die direkte Summe der einzelnen Automorphismenringe. Hat \mathfrak{A} den Typus $(p^{e_1}, p^{e_2}, \dots, p^{e_m})$,*

⁵⁾ Vgl. die entsprechende Untersuchung für die Elementarteilergruppe in meiner Arbeit a. a. O. § 1 und § 3.

$e_1 \geq e_2 \geq \dots \geq e_m$, so wird jedes Element des Automorphismenring durch die Matrix

$$(3) \quad P = \begin{pmatrix} h_{11} & h_{12} & h_{13} & \dots \\ p^{e_1 - e_2} h_{21} & h_{22} & h_{23} & \\ p^{e_1 - e_3} h_{31} & p^{e_2 - e_3} h_{32} & h_{33} & \\ \vdots & \vdots & \vdots & \end{pmatrix}$$

dargestellt, wo $h_{\kappa\lambda}$ modulo p^{e_μ} , $\mu = \max(\kappa, \lambda)$, betrachtet werden soll, also $P = 0$ bedeutet, daß jedes $h_{\kappa\lambda} \equiv 0 \pmod{p^{e_\mu}}$.

Die Gesamtheit der eigentlichen Automorphismen von \mathfrak{A} bildet eine Gruppe (*Automorphismengruppe*) \mathfrak{G} . Aus Satz 1 folgt unmittelbar (vgl. Satz 6, § 3), was auch aus der Basisdarstellung klar ist: ist \mathfrak{A} die direkte Summe von Abelschen Gruppen, deren Ordnungen Potenzen von einander verschiedener Primzahlen sind, so ist \mathfrak{G} das direkte Produkt der einzelnen Automorphismengruppen. Daher wollen wir von jetzt an annehmen, daß die vorgegebene Abelsche Gruppe \mathfrak{A} den Typus $(p^{e_1}, p^{e_2}, \dots, p^{e_m})$, $e = e_1 \geq e_2 \geq \dots \geq e_m$, hat.

Satz 2. *Ein Automorphismus P einer Abelschen Gruppe \mathfrak{A} ist dann und nur dann eigentlich, wenn seine Determinante durch p nicht teilbar ist.*

Dann und nur dann existiert der reziproke Automorphismus P^{-1} von P , wenn P eigentlich ist, dann ist $PP^{-1} \equiv E \pmod{p}$. Daher ist die Determinante von P durch p nicht teilbar. Ist umgekehrt die Determinante von P durch p nicht teilbar, so ist die reziproke Matrix von P eine ganze Matrix modulo p^e .

Einfachheit halber wollen wir annehmen, daß unter e_1, e_2, \dots, e_m genau n_κ -mal κ vorkommt, was keine Beschränkung der Allgemeinheit ist, wenn n_κ Null sein darf. Man schreibe P durch Zusammenfassung der Koeffizienten von (3) in Teilmatrizen in der Gestalt

$$(4) \quad P = \begin{pmatrix} P_{11} & P_{12} & \dots & P_{1e} \\ P_{21} & P_{22} & \dots & P_{2e} \\ \vdots & \vdots & \ddots & \vdots \\ P_{e1} & \dots & \dots & P_{ee} \end{pmatrix},$$

wo der Grad von $P_{\kappa\kappa}$ gleich $n_{e-\kappa+1}$ ist. Die Koeffizienten von $P_{\kappa\lambda}$, $\kappa > \lambda$, sind durch p teilbar. Daher kann man den Satz 2 auch so aussprechen:

Satz 2a. *P ist dann und nur dann eigentlich, wenn die Determinante von $P_{\kappa\kappa}$ für jedes κ durch p nicht teilbar ist.*

Die Anzahl der Elemente des Automorphismenrings \mathfrak{o} von \mathfrak{A} ist — wie aus (3) folgt — gleich⁶⁾

$$\chi(\mathfrak{o}) = p^{\eta(0)} = p^{e_1 + 3e_2 + \dots + (2m-1)e_m},$$

die man auch die *Ordnung* von \mathfrak{o} nennt. Ist speziell $e_1 = e_2 = \dots = e_m$, so nennt man \mathfrak{o} einen *Kongruenzring*, der mit $\mathfrak{o}(m, p^e)$ bezeichnet wird. Man kann die Koeffizientenfaktoren $h_{\mu\nu}$ von $P_{\kappa\lambda}$, $\kappa \neq \lambda$, beliebig nehmen, auch wenn P als eigentlich vorausgesetzt wird. Wenn man die Koeffizienten in allen $P_{\kappa\kappa}$ festhält, so erhält man genau $p^{\eta(0) - \sum \kappa n_\kappa^2}$ verschiedene Matrizen in \mathfrak{o} . Die Determinante von $P_{\kappa\kappa}$ muß dann aber durch p nicht teilbar sein, d. h. $P_{e-\kappa+1, e-\kappa+1}$ muß ein Element der *Kongruenzgruppe* des Grades n_κ modulo p^κ sein, die wir mit $\mathfrak{G}(n_\kappa, p^\kappa)$ bezeichnen. Die Ordnung von $\mathfrak{G}(m, p^e)$ ist aber $p^{(e-1)m^2} \psi(m)$, wo $\psi(m)$ die Ordnung von $\mathfrak{G}(m, p^e)$, $\psi(m) = (p^m - 1)(p^m - p) \dots (p^m - p^{m-1})$, bedeutet⁷⁾. Daher ist die Ordnung von \mathfrak{G} gleich

$$\chi(\mathfrak{G}) = p^{\eta(0) - \sum \kappa n_\kappa^2 + \sum (\kappa-1) n_\kappa^2} \prod \psi(n_\kappa) = p^{\eta(0) - \sum_{\kappa=1}^e n_\kappa^2} \prod_{\kappa=1}^e \psi(n_\kappa).$$

§ 2.

Automorphismenring.

Es sei P ein Element von \mathfrak{o} derart, daß

$$(1a) \quad p^d P = 0 \quad \text{bzw.} \quad (1b) \quad F_d P = 0,$$

wo F_d eine diagonale Matrix mit Koeffizienten $p^{e_\kappa - d}$, $p^{e_\kappa - d} = 1$ für $e_\kappa \leq d$, $\kappa = 1, 2, \dots, m$, ist. Ist X ein beliebiges Element von \mathfrak{o} , so ist — wie aus (3) § 1 folgt — jeder Koeffizient der κ -ten Spalte der Matrix $F_d X$ durch $p^{e_\kappa - d}$ teilbar. Daher existiert eine ganze Matrix X' (die im allgemeinen in \mathfrak{o} nicht enthalten ist) derart, daß $F_d X = X' F_d$ ist. Nach dieser Bemerkung beweisen wir: *Die Gesamtheit der Elemente von \mathfrak{o} , die der Bedingung (1a) bzw. (1b) genügen, bildet ein Ideal⁸⁾ in \mathfrak{o} .* Aus $p^d P = 0$, $p^d Q = 0$ folgt nämlich $p^d(P - Q) = 0$. Sind X, Y beliebige Elemente von \mathfrak{o} , so ist auch $p^d X P Y = 0$. Aus $F_d P = F_d Q = 0$ folgt $F_d(P - Q) = 0$ und nach der Bemerkung auch $F_d X P Y = X' F_d P Y = 0$.

⁶⁾ Wenn man p durch ein irreduzibles Polynom des Grades t ersetzt (vgl. Einleitung), so ist der Rang des Automorphismenrings der betreffenden Elementarteilergruppe gleich $t(e_1 + 3e_2 + \dots + (2m-1)e_m)$.

⁷⁾ Speiser, a. a. O. S. 131. Diese Gruppe wird als spezieller Fall von \mathfrak{G} abgeleitet. Siehe Anmerkung ¹⁶⁾.

⁸⁾ Unter einem Ideal soll stets ein zweiseitiges Ideal verstanden werden.

Das durch (1 a) bzw. (1 b) definierte Ideal bezeichnen wir mit $r_d^* = r_1 \hat{\ } r_2 \hat{\ } \dots \hat{\ } r_d$ bzw. $\bar{r}_d^* = \bar{r}_e \hat{\ } \bar{r}_{e-1} \hat{\ } \dots \hat{\ } \bar{r}_{d+1}$.⁹⁾

Diejenigen Elemente a von \mathfrak{A} , die der Bedingung

$$(2 a) \quad p^d a = 0 \quad \text{bzw.} \quad (2 b) \quad a = p^d b$$

genügen, bilden einen Normalteiler

$$\mathfrak{A}_d^* = \mathfrak{A}_1 \hat{\ } \mathfrak{A}_2 \hat{\ } \dots \hat{\ } \mathfrak{A}_d \quad \text{bzw.} \quad \bar{\mathfrak{A}}_d^* = \bar{\mathfrak{A}}_e \hat{\ } \bar{\mathfrak{A}}_{e-1} \hat{\ } \dots \hat{\ } \bar{\mathfrak{A}}_{d+1},$$

die wir nach Herrn Krull die d -te *Loewysche Untergruppe* bzw. *Hauptuntergruppe* von \mathfrak{A} nennen. Die Reihe $\mathfrak{A}_1, \mathfrak{A}_2, \dots, \mathfrak{A}_e$ bzw. $\bar{\mathfrak{A}}_e, \bar{\mathfrak{A}}_{e-1}, \dots, \bar{\mathfrak{A}}_1$ heißt die *vordere* bzw. *hintere Loewysche Kompositionsreihe* von \mathfrak{A} . Wenn man ein Element $b = b_1 r_1 + b_2 r_2 + \dots + b_m r_m$ von \mathfrak{A} durch die Matrix

$$B R_r = (b_1 \ b_2 \ \dots \ b_m) \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{pmatrix}$$

darstellt, so ist die Bedingung (2 b) mit $A F_d R_r = 0$ identisch, da $p^d F_d = 0$ ist. Ferner ist (2 a) mit $a = B F_d R_r$ äquivalent¹⁰⁾.

Transformiert man ein Element a von \mathfrak{A} durch ein Element P von \mathfrak{o} , so bezeichne man das mit $P \times a$. Ist \mathfrak{o}' bzw. \mathfrak{A}' ein System von Elementen von \mathfrak{o} bzw. \mathfrak{A} , so versteht man unter $\mathfrak{o}' \times \mathfrak{A}'$ die Gesamtheit der Elemente von der Gestalt $P \times a$, wo P bzw. a ein Element von \mathfrak{o}' bzw. \mathfrak{A}' ist.

Satz 3. r_d^* und \bar{r}_d^* werden folgendermaßen charakterisiert:

$$(I) \quad r_d^* \times \mathfrak{A} = \mathfrak{A}_d^*, \quad \bar{r}_d^* \times \mathfrak{A} = \bar{\mathfrak{A}}_d^*;$$

$$(II) \quad r_d^* \times \bar{\mathfrak{A}}_d^* = 0, \quad \bar{r}_d^* \times \mathfrak{A}_d^* = 0.$$

⁹⁾ Unter r_d bzw. \bar{r}_{d+1} soll der Restklassenring (die Faktorgruppe) r_d^*/r_{d-1}^* bzw. $\bar{r}_d^*/\bar{r}_{d+1}^*$ verstanden werden, wo r_d^* bzw. \bar{r}_d^* ein Ideal (ein Normalteiler) ist. Die Schreibweise soll das System der Restklassenringe (der Faktorgruppen) hervorheben. Dieses System wird nach Herrn Krull a. a. O. als „Kompositionsreihe“ bezeichnet, wenn es sich um den ganzen Ring (die ganze Gruppe) handelt. Diese sonst in der Literatur nicht übliche Bezeichnung erweist sich als sehr bequem, da die invarianten Bestandteile dadurch hervorgehoben werden; sie wird im folgenden stets angewandt.

¹⁰⁾ Diese Äquivalenz gilt für ein Element von \mathfrak{o} nicht, z. B.

$$p^{e-1} P = \begin{pmatrix} p^{e-1} P_{11} & 0 \dots 0 \\ 0 & 0 \dots 0 \\ \vdots & \\ 0 & 0 \dots 0 \end{pmatrix}, \text{ aber } F_{e-1} P = 0, \text{ wenn } P = \begin{pmatrix} p^{e-1} P_{11} & p^{e-1} P_{12} \dots P_{1e} \\ 0 & 0 \dots 0 \\ \vdots & \vdots \dots \vdots \\ 0 & 0 \dots 0 \end{pmatrix}$$

ist.

\mathfrak{A}_d^* und $\overline{\mathfrak{A}}_d^*$ haben die Eigenschaften¹¹⁾:

$$(III) \quad \mathfrak{o} \times \mathfrak{A}_d^* = \mathfrak{A}_d^*, \quad \mathfrak{o} \times \overline{\mathfrak{A}}_d^* = \overline{\mathfrak{A}}_d^*.$$

Zum Beweis betrachten wir die einspaltige Matrix R_r mit Koeffizienten r_1, r_2, \dots, r_m wie oben. Dann stellt jeder Automorphismus (1) § 1 sich in der Gestalt $R_{r^*} = P R_r$ dar, wo P ein Element von \mathfrak{o} ist. Daraus folgt (I), denn $P R_r$ ist dann und nur dann Null, wenn P Null ist, da r_1, r_2, \dots, r_m linear unabhängig sind. Da die Basis von $\overline{\mathfrak{A}}_d^*$ bzw. \mathfrak{A}_d^* die Gestalt $p^d R_r$ bzw. $F_d R_r$ hat, so wird sie durch P in $p^d P R_r$ bzw. $F_d P R_r$ transformiert (hieraus folgt (III)), welches aber dann und nur dann Null ist, wenn $p^d P$ bzw. $F_d P$ Null ist, also $p^d P = 0$ bzw. $F_d P = 0$ ist. Damit ist (II) bewiesen.

Wir wollen nun die beiden Kompositionsreihen r_1, r_2, \dots, r_e und $\bar{r}_e, \bar{r}_{e-1}, \dots, \bar{r}_1$ genau untersuchen.

A. Es sei

$$(3a) \quad R = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}$$

ein Element von \mathfrak{r}_1 , wo der Grad von R_{22} gleich n_1 ist (also $R_{22} = P_{e_e}$ in (4) § 1). Da $pR = 0$ ist, so sind die Koeffizienten von R_{11} und R_{21} durch p teilbar. Sind R, R^* zwei Elemente von \mathfrak{r}_1 , so ist¹²⁾

$$R R^* = \begin{pmatrix} R_{12} & R_{21}^* & R_{12} & R_{22}^* \\ R_{22} & R_{21}^* & R_{22} & R_{22}^* \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}.$$

(a₁) Wenn $R_{22} = 0$ oder $R_{22}^* = 0$, so ist $S_{22} = 0$. (a₂) Wenn $R_{22} = 0$ bzw. $R_{22}^* = 0$, so ist $S_{21} = 0$ bzw. $S_{12} = 0$. (a₃) Wenn $R_{22} = 0$ und $R_{12} = 0$, so ist $R R^* = 0$. (a₄) Wenn $R_{22}^* = 0$ und $R_{21}^* = 0$, so ist $R R^* = 0$. Aus (a₁) bis (a₄) folgt:

I. Die Gesamtheit der Elemente von \mathfrak{r}_1 derart, daß $R_{22} = 0$, bildet ein Ideal \mathfrak{s}_1 in \mathfrak{r}_1 , dessen Restklassenring $\mathfrak{r}_1/\mathfrak{s}_1$ mit dem Kongruenzring $\mathfrak{o}(n_1, p)$ isomorph ist (vgl. (a₁)).

II. Die Gesamtheit der Elemente von \mathfrak{r}_1 derart, daß $R_{22} = R_{12} = 0$ bzw. $R_{22} = R_{21} = 0$, bildet ein Ideal $\mathfrak{s}_{12}^{(1)}$ bzw. $\mathfrak{s}_{21}^{(1)}$ in \mathfrak{r}_1 . $\mathfrak{s}_{12}^{(1)}$ und $\mathfrak{s}_{21}^{(1)}$ erzeugen das Ideal \mathfrak{s}_1 . Jedes Produkt zweier Elemente von $\mathfrak{s}_{12}^{(1)}$ bzw. $\mathfrak{s}_{21}^{(1)}$ ist gleich Null (vgl. (a₂), (a₃), (a₄)).

¹¹⁾ Diese Eigenschaft besagt, daß \mathfrak{A}_d^* und $\overline{\mathfrak{A}}_d^*$ charakteristische Untergruppen von \mathfrak{A} sind, und zwar handelt es sich um die beiden bekannten Reihen von charakteristischen Untergruppen: (Vgl. etwa Burnside, The theory of groups, 2. Aufl., S. 108; Chatelet, a. a. O. Kap. VII.)

¹²⁾ Denn $\begin{pmatrix} R_{11} & R_{11}^* & R_{11} & R_{12}^* \\ R_{21} & R_{11}^* & R_{21} & R_{12}^* \end{pmatrix} = 0$, was aus der Definition von \mathfrak{r}_1 folgt.

III. Jedes Produkt zweier Elemente von \mathfrak{s}_1 ist im Durchschnitt t_1 der beiden Ideale $\mathfrak{s}_{12}^{(1)}$ und $\mathfrak{s}_{21}^{(1)}$ enthalten (vgl. (a₂)).

IV. Jedes Produkt eines Elementes von t_1 und eines von r_1 ist gleich Null (vgl. (a₃), (a₄)).

V. Das Ideal r_1 bildet eine Abelsche Gruppe in bezug auf Addition, deren Typus (p, p, \dots, p) ist.

VI. Die Ordnungen von $r_1, \mathfrak{s}_1, \mathfrak{s}_{12}^{(1)} \simeq \mathfrak{s}_{21}^{(1)}, t_1$ werden durch

$$\begin{aligned} \chi(r_1) &= p^{(n_1+n_2+\dots+n_e)^2}, & \chi(\mathfrak{s}_1) &= p^{(2n_1+n_2+\dots+n_e)(n_2+\dots+n_e)}, \\ \chi(\mathfrak{s}_{12}^{(1)}) &= p^{(n_1+\dots+n_e)(n_2+\dots+n_e)}, & \chi(t_1) &= p^{(n_2+\dots+n_e)^2} \end{aligned}$$

gegeben.

Nach V. kann nämlich jeder Koeffizient eines Elementes von r_1 genau p verschiedene Werte nehmen. Man kann also die Ordnungen von $r_1, \mathfrak{s}_1, \mathfrak{s}_{12}^{(1)}, t_1$ leicht ausrechnen. Für die Restklassenringe r_2, r_3, \dots, r_e kann man genau wie oben vorgehen.

B. Man schreibe ein Element von \bar{r}_κ^* durch Zusammenfassung der Koeffizienten in Teilmatrizen, wie früher ((4) § 1), in der Gestalt

$$(3b) \quad \begin{pmatrix} p^\kappa P_{11} & p^{\kappa-1} P_{12} & \dots & p P_{1\kappa} & P_{1\kappa+1} & \dots & P_{1e} \\ p^\kappa P_{21} & p^\kappa P_{22} & \dots & p^2 P_{2\kappa} & p P_{2\kappa+1} & \dots & P_{2e} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ p^\kappa P_{e-\kappa 1} & p^\kappa P_{e-\kappa 2} & \dots & & & & P_{e-\kappa e} \\ \vdots & \vdots & & & & & \vdots \\ 0 & 0 & & & & & 0 \\ \vdots & \vdots & & & & & \vdots \\ \vdots & \vdots & & & & & \vdots \\ 0 & 0 & & & & & 0 \end{pmatrix}.$$

Man erkennt diese Gestalt des Elementes, wenn man die Koeffizienten der in der Definition (1b) vorkommenden Matrix F_κ entsprechend in Teilmatrizen zusammenfaßt. Da die Koeffizienten der Teilmatrizen $p^\mu P_{\mu\nu}$ in (3b) für $\mu \geq \nu$ durch p teilbar sind, so ist jedes Produkt zweier Elemente von \bar{r}_κ^* in $\bar{r}_{\kappa+1}^*$ enthalten.

VII. Jedes Produkt zweier Elemente von $\bar{r}_\kappa, \kappa > 1$, ist gleich Null.

VIII. $\bar{r}_\kappa, \kappa \geq 1$, bildet eine Abelsche Gruppe in bezug auf Addition, deren Typus (p, p, \dots, p) ist.

Denn ist P ein Element von $\bar{r}_{\kappa-1}^*$ (d. h. $F_{\kappa-1} P = 0$), so ist pP in \bar{r}_κ^* enthalten, da aus $F_{\kappa-1} P = 0$ auch $F_\kappa pP = 0$ folgt.

IX. Die Ordnung von $\bar{r}_\kappa, \kappa > 1$, ist gleich

$$\begin{aligned} \chi(\bar{r}_\kappa) &= p^{(n_e+n_{e-1}+\dots+n_{e-\kappa})n_e + (n_e+\dots+n_{e-\kappa+1})n_{e-1}+\dots} \\ &= p^{\sum_{\alpha=0}^{e-\kappa-1} n_{e-\alpha}(n_e+n_{e-1}+\dots+n_{e-\alpha})} \end{aligned}$$

Nach VIII. kann nämlich jeder Koeffizient der Teilmatrix $P_{\alpha\beta}$, $\beta - \alpha \leq \kappa$, genau p verschiedene Werte annehmen, wenn man modulo \bar{r}_κ^* betrachtet. Da der Grad von $P_{\kappa\kappa}$ gleich $n_{e-\kappa+1}$ ist, so kann man die Ordnung von \bar{r}_κ ausrechnen. Ist P ein Element von \mathfrak{o} , so ist, da $h_{\kappa\lambda} \pmod{p^\mu}$ zu nehmen ist,

$$(4b) \quad P \equiv \begin{pmatrix} Q_{11} & 0 & 0 \\ Q_{21} & Q_{22} & \\ \vdots & & \\ Q_{e1} & Q_{e2} & \dots & Q_{ee} \end{pmatrix} \pmod{\bar{r}_1^*}.$$

Wenn man ferner alle $Q_{\kappa\kappa-\alpha}$, $\alpha \leq d-1$, $\kappa > \alpha$, Null setzt, so erhält man ein Ideal \bar{s}_d^* in \bar{r}_1 , was aus der Kompositionsregel der Matrizen folgt. Dadurch erhält man eine Kompositionsreihe $\bar{s}_e, \bar{s}_{e-1}, \dots, \bar{s}_3, \bar{1}$, wo $\bar{s}_d^* = \bar{s}_e \hat{\ } \bar{s}_{e-1} \hat{\ } \dots \hat{\ } \bar{s}_{d+1}$ ist. Nun weiß man folgendes:

X. Jedes Produkt zweier Elemente von \bar{s}_κ , $\kappa > 1$, ist gleich Null.

XI. $\bar{1}$ ist mit der direkten Summe der Kongruenzringe $\mathfrak{o}(n_\kappa, p)$, $n_\kappa \neq 0$, $\kappa = 1, 2, \dots, e$, isomorph.

XII. Die Ordnung von \bar{s}_κ , $\kappa > 1$, ist gleich $\chi(\bar{s}_\kappa) = p^{\sum_{\alpha=1}^{e-\kappa} n_{e-\alpha+1} n_{e-\kappa-\alpha+1}}$. Denn der Grad von $Q_{\kappa\kappa}$ ist gleich $n_{e-\kappa+1}$.

§ 3.

Strahlbildung.

Das Hilfsmittel für den Übergang von Ring zu Gruppe bildet der Begriff des *Strahls*. Es sei \mathfrak{o} ein Ring mit Einheitselement E , c ein Unterring von \mathfrak{o} . Die Gesamtheit der Elemente von \mathfrak{o} , die mit dem Einheitselement E modulo c kongruent sind, nennt man den Strahl von \mathfrak{o} in bezug auf c , was mit $\{\mathfrak{o}, c\}$ bezeichnet wird. Ist \mathfrak{r} ein Ideal in c , so versteht man unter $\{\mathfrak{o}, c|\mathfrak{r}\}$ den Strahl $\{\mathfrak{o}, \mathfrak{r}\}$, also $\{\mathfrak{o}, c\} = \{\mathfrak{o}, c|\mathfrak{r}\}$. $\{\mathfrak{o}|\mathfrak{o}|\mathfrak{r}\}$ wird einfach mit $\{\mathfrak{o}|\mathfrak{r}\}$ bezeichnet.

Hilfssatz 1. *Der Strahl $\{\mathfrak{o}, c\}$ ist ein Multiplikatorenbereich für das Ideal \mathfrak{r} in c .*

Denn jedes Element von $\{\mathfrak{o}, c\}$ hat die Gestalt $E + P$, wo P ein Element von c ist. Aus $P\mathfrak{r} \equiv \mathfrak{r}P \equiv 0 \pmod{\mathfrak{r}}$ folgt auch

$$(E + P)\mathfrak{r} \equiv \mathfrak{r}(E + P) \equiv 0.$$

Die Gesamtheit der Einheiten — die Gesamtheit der Elemente mit reziproken Elementen — in \mathfrak{o} bildet eine Gruppe \mathfrak{G} , die wir die *zugehörige Gruppe* von \mathfrak{o} nennen.

Hilfssatz 2. *Der Durchschnitt von \mathfrak{G} und $\{\mathfrak{o}|c\}$ bildet eine Untergruppe¹³⁾ \mathfrak{C} von \mathfrak{G} .*

Aus $P \equiv Q \equiv E \pmod{c}$ folgt nämlich $PQ \equiv Q \pmod{cQ}$. Nach Hilfssatz 1, wobei jetzt r mit c zusammenfällt, ist daher $PQ \equiv Q \equiv E \pmod{c}$. Ferner $P^{-1} \equiv E$, da $PP^{-1} \equiv E$ ist. Aus $P \equiv E$ folgt nämlich $E \equiv P^{-1} \pmod{cP^{-1}}$ und daher $E \equiv P^{-1} \pmod{c}$.

Hilfssatz 3. *Der Durchschnitt von \mathfrak{G} und $\{\mathfrak{o}|c|r\}$ bildet einen Normalteiler \mathfrak{R} von \mathfrak{G} .*

Ist nämlich R ein Element von \mathfrak{G} (d. h. $R \equiv E \pmod{c}$), so folgt nach Hilfssatz 1 aus $P \equiv E \pmod{r}$ auch $RPR^{-1} \equiv E$.

Hilfssatz 4. *Sind zwei Elemente P, Q von \mathfrak{G} modulo r kongruent, so ist PQ^{-1} in \mathfrak{R} enthalten.*

Ist nämlich $P \equiv Q \pmod{r}$, so ist $PQ^{-1} \equiv E$, was aus dem Hilfssatz 1 und der Existenz von Q^{-1} in \mathfrak{G} (Hilfssatz 2) folgt.

Für den Restklassenring von \mathfrak{o} kann man genau wie oben vorgehen. Zusammenfassend hat man

Satz 4. *Ist \mathfrak{o} ein Ring mit Einheitselement, \mathfrak{G} die zugehörige Gruppe von \mathfrak{o} , so entspricht bei der Zuordnung durch Strahlbildung einem durch endlich-oftmalige Ideal- und Restklassenringbildung aus \mathfrak{o} erhaltenen Ring stets eine Gruppe, die man erhält, wenn man jeweils die Worte Ideal durch Normalteiler, Restklassenring durch Faktorgruppe und \mathfrak{o} durch \mathfrak{G} ersetzt.*

Wir beweisen nun

Satz 5. *Es sei c ein Ideal von \mathfrak{o} , r ein Ideal in c . Sind $r + S_1, r + S_2$ zwei Klassen von c/r derart, daß $S_1 S_2 \equiv 0 \pmod{r}$, so sind die Nebengruppen von \mathfrak{R} in \mathfrak{G} miteinander vertauschbar. Die Faktorgruppe $\mathfrak{G}/\mathfrak{R}$ ist mit einem Ideal u des Restklassenringes c/r — dieser aufgefaßt als Abelsche Gruppe in bezug auf Addition — isomorph, wenn die Voraussetzung für jede Klasse von c/r besteht; und es wird $u = c/r$, wenn außerdem der Strahl $\{\mathfrak{o}|c\}$ nur Einheiten enthält.*

Denn sind $\mathfrak{R}P_1 \equiv E + S_1, \mathfrak{R}P_2 \equiv E + S_2 \pmod{r}$, so ist $\mathfrak{R}P_1 P_2 \equiv E + S_1 + S_2$, woraus der erste Teil des Satzes folgt. Wenn der Strahl $\{\mathfrak{o}|c\}$ nur Einheiten enthält, so ist $\mathfrak{G} = \{\mathfrak{o}|c\}$ und $\mathfrak{R} = \{\mathfrak{o}|c|r\}$. Daraus folgt der zweite Teil, da aus $\mathfrak{R}P_1 \equiv \mathfrak{R}$ auch $S_1 \equiv 0$ folgt und umgekehrt.

¹³⁾ Es kann \mathfrak{C} mit \mathfrak{G} übereinstimmen, auch wenn c mit \mathfrak{o} nicht übereinstimmt.

Denn z. B. $\begin{pmatrix} x & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & y \end{pmatrix}$ bildet einen Ring \mathfrak{o} , wo $x, y \pmod{2}$ betrachtet werden, $\begin{pmatrix} x & 0 \\ 0 & 0 \end{pmatrix}$ bildet ein Ideal r_x in \mathfrak{o} . $\{\mathfrak{o}|r_x\}$ enthält aber die einzige Einheit E in \mathfrak{o} .

Zusammenfassend mit Satz 4 weiß man, daß man für den Restklassenring von \mathfrak{o} genau wie oben vorgehen kann.

Satz 6. *Entspricht einem Ideal c von \mathfrak{o} ein Normalteiler \mathfrak{C} von \mathfrak{G} und ist c die direkte Summe von zwei Idealen \mathfrak{v} und \mathfrak{w} in c , so entspricht \mathfrak{v} bzw. \mathfrak{w} ein Normalteiler \mathfrak{B} bzw. \mathfrak{B} von \mathfrak{C} und c das direkte Produkt $\mathfrak{C} = \mathfrak{B} \times \mathfrak{B}$, falls $\mathfrak{B} \neq \mathfrak{C}$, $\mathfrak{B} \neq \mathfrak{C}$ ist¹⁴⁾, also speziell wenn $\mathfrak{C} = \{\mathfrak{o} | c\}$ ist.*

Denn aus $c = \mathfrak{v} + \mathfrak{w}$, $[\mathfrak{v}, \mathfrak{w}] = 0$ (d. h. der Durchschnitt von \mathfrak{v} und \mathfrak{w} ist Null) folgt $\mathfrak{C} = \mathfrak{B} \mathfrak{B}$, da $E + V + W = (E + V)(E + W)$ dann und nur dann eine Einheit ist, wenn $E + V$ und $E + W$ Einheiten sind, wo V bzw. W ein Element von \mathfrak{v} bzw. \mathfrak{w} ist, und jedes Element von \mathfrak{B} ist nach Satz 5 mit jedem Element von \mathfrak{B} vertauschbar, da $\mathfrak{v} \mathfrak{w} = 0$ ist. Ist $[\mathfrak{B}, \mathfrak{B}] = \mathfrak{y}$, so ist $\mathfrak{y} - \mathfrak{C}$ ¹⁵⁾ in $[\mathfrak{v}, \mathfrak{w}]$ enthalten, also $\mathfrak{y} - \mathfrak{C} = 0$ und daher $\mathfrak{y} = \mathfrak{C}$. Ist $\mathfrak{C} = \{\mathfrak{o} | c\}$, so ist $\mathfrak{B} = \{\mathfrak{v} | c | \mathfrak{v}\}$ ($\mathfrak{B} = \{\mathfrak{w} | c | \mathfrak{w}\}$) und dann und nur dann gleich \mathfrak{C} , wenn $\mathfrak{v} = 0$ ($\mathfrak{w} = 0$) ist. Damit ist der Satz bewiesen. Man kann nach Satz 4 diesen Satz auf einen allgemeinen Ring im Sinne des Satzes 4 verallgemeinern.

§ 4.

Automorphismengruppe.

Mit Hilfe der Resultate von § 3 übertragen wir die Ergebnisse von § 2 in die Theorie der Automorphismengruppe.

Satz 7¹⁶⁾. *Die Automorphismengruppe \mathfrak{G} einer endlichen Abelschen Gruppe \mathfrak{A} mit n_x Basiselementen von der Ordnung p^α , $\alpha = 1, 2, \dots, e$, hat zwei Reihen von Normalteilern $\mathfrak{R}_1 \hat{\ } \mathfrak{R}_2 \hat{\ } \dots \hat{\ } \mathfrak{R}_e = \mathfrak{R}_e^*$ und $\overline{\mathfrak{R}}_e \hat{\ } \overline{\mathfrak{R}}_{e-1} \hat{\ } \dots \hat{\ } \overline{\mathfrak{R}}_{e+1} = \mathfrak{R}_d^*$, $d = 1, 2, \dots, e$, die die folgenden Eigenschaften haben:*

(A) *Der Normalteiler \mathfrak{R}_d^* von \mathfrak{G} besteht aus der Gesamtheit der Automorphismen, welche die d -te Hauptuntergruppe $\overline{\mathfrak{A}}_d^*$ elementweise fest-*

¹⁴⁾ \mathfrak{B} und \mathfrak{B} können \mathfrak{C} sein. Der Ring \mathfrak{o} in Anmerkung ¹³⁾ ist die direkte Summe von \mathfrak{r}_x und \mathfrak{r}_y . Die zugehörige Gruppe von \mathfrak{o} ist aber \mathfrak{C} .

¹⁵⁾ Unter $\mathfrak{y} - \mathfrak{C}$ soll die Gesamtheit der Elemente $Y - E$ verstanden werden, wo Y ein Element von \mathfrak{y} ist.

¹⁶⁾ Wir gebrauchen hier die Ausdrucksweise der multiplikativen Verknüpfung der Gruppe \mathfrak{A} , wie es in der Literatur üblich ist. Herr Speiser a. a. O. hat die Reihe $\overline{\mathfrak{A}}_e^*, \overline{\mathfrak{A}}_{e-1}^*, \dots, \overline{\mathfrak{A}}_1^*$ betrachtet und die Reihe $\mathfrak{R}_1^*, \mathfrak{R}_2^*, \dots, \mathfrak{R}_e^*$ (dort $\mathfrak{A}_1, \mathfrak{A}_2, \dots, \mathfrak{A}_r$) erhalten. Weiter hat er die Existenz des Abelschen Normalteilers $\mathfrak{C}_{21}^{(x)}$ von \mathfrak{R}_x (dort $\mathfrak{A}_x / \mathfrak{A}_{x-1}$) bewiesen und dessen Typus, Ordnung und Index gegeben. Ist $e_1 = e_2 = \dots = e_m = e$, so stimmen die beiden Reihen $\mathfrak{R}_1, \mathfrak{R}_2, \dots, \mathfrak{R}_e$ und $\overline{\mathfrak{R}}_e, \overline{\mathfrak{R}}_{e-1}, \dots, \overline{\mathfrak{R}}_1$ überein, d. h. \mathfrak{R}_x ist mit $\overline{\mathfrak{R}}_{e-x+1}$ isomorph. Speziell wird $\mathfrak{R}_e \simeq \mathfrak{R}_1 \simeq \mathfrak{G}(m, p)$ und $\mathfrak{R}_x \simeq \mathfrak{R}_\lambda$, $x \neq \lambda$, $\lambda \neq 1$; letztere sind Abelsche Gruppen von der Ordnung p^{m^2} und vom Typus (p, p, \dots, p) (vgl. Speiser, a. a. O., Satz I13).

lassen, dabei besteht also $\overline{\mathfrak{A}}_d^*$ nach § 2 aus den p^d -ten Potenzen aller Elemente von \mathfrak{A} . Die Faktorgruppe \mathfrak{K}_x enthält einen Normalteiler \mathfrak{S}_x von der Ordnung $p^{(2n_x + n_{x+1} + \dots + n_e)(n_{x+1} + \dots + n_e)}$, dessen Faktorgruppe $\mathfrak{K}_x / \mathfrak{S}_x$ mit der Kongruenzgruppe $\mathfrak{G}(n_x, p)$ isomorph ist. \mathfrak{S}_x wird durch zwei miteinander isomorphe Abelsche Normalteiler $\mathfrak{S}_{12}^{(x)}$ und $\mathfrak{S}_{21}^{(x)}$ von \mathfrak{K}_x erzeugt deren Ordnung und Typus durch $p^{(n_x + \dots + n_e)(n_{x+1} + \dots + n_e)}$ und (p, p, \dots, p) gegeben sind. Der Durchschnitt \mathfrak{T}_x von $\mathfrak{S}_{12}^{(x)}$ und $\mathfrak{S}_{21}^{(x)}$ ist eine Untergruppe des Zentrums von \mathfrak{K}_x von der Ordnung $p^{(n_{x+1} + \dots + n_e)^2}$.

(B) Der Normalteiler $\overline{\mathfrak{R}}_d^*$ von \mathfrak{G} besteht aus der Gesamtheit der Automorphismen, welche die d -te Loewysche Untergruppe \mathfrak{A}_d^* elementweise festlassen, dabei besteht \mathfrak{A}_d^* aus allen Elementen von \mathfrak{A} von der Ordnung $\leq p^d$. Die Faktorgruppe $\overline{\mathfrak{R}}_x, x > 1$, ist eine Abelsche Gruppe von der Ordnung $p^{\sum_{e-\alpha} n_e + \dots + n_{e-x-\alpha}}$ und von dem Typus (p, p, \dots, p) . $\overline{\mathfrak{R}}_1$ hat eine Kompositionsreihe $\overline{\mathfrak{S}}_e, \overline{\mathfrak{S}}_{e-1}, \dots, \overline{\mathfrak{S}}_2, \mathfrak{T}$, wo $\overline{\mathfrak{S}}_x$ eine Abelsche Gruppe von der Ordnung $p^{\sum_{e-\alpha}^{e-x} n_{e-\alpha-1} n_{e-x-\alpha+1}}$ und von dem Typus (p, p, \dots, p) ist. \mathfrak{T} ist mit dem direkten Produkt der Kongruenzgruppe $\mathfrak{G}(n_x, p), x = 1, 2, \dots, e$ isomorph.

Die Gesamtheit der eigentlichen Automorphismen — Einheiten —, die im Strahl $\{\mathfrak{o} | \mathfrak{r}_d^*\}$ bzw. $\{\mathfrak{o} | \overline{\mathfrak{r}}_d^*\}$ des Automorphismenringes \mathfrak{o} liegen, bildet einen Normalteiler $\mathfrak{R}_1 \hat{\ } \mathfrak{R}_2 \hat{\ } \dots \hat{\ } \mathfrak{R}_d = \mathfrak{R}_d^*$ bzw. $\overline{\mathfrak{R}}_e \hat{\ } \overline{\mathfrak{R}}_{e-1} \hat{\ } \dots \hat{\ } \overline{\mathfrak{R}}_{d+1} = \overline{\mathfrak{R}}_d^*$ der Automorphismengruppe \mathfrak{G} (Satz 4), der nach Satz 3 aus der Gesamtheit der Elemente von \mathfrak{G} besteht, die $\overline{\mathfrak{A}}_d^*$ bzw. \mathfrak{A}_d^* elementweise festlassen. Damit ist der erste Teil von (A) bzw. (B) bewiesen.

(A) Wir betrachten den Strahl $\{\mathfrak{o} | \mathfrak{r}_1\}$. Jedes Element des Strahls $\{\mathfrak{o} | \mathfrak{r}_1 | \mathfrak{s}_1\}$ ist *eigentlich*, da in (3a) § 2 die Koeffizienten von R_{11} durch p teilbar und daher die Determinante von $E + R$ durch p nicht teilbar ist. (Vgl. Satz 2, § 1). Bei der Anwendung von Satz 5 handelt es sich also um die schärfere Aussage der Isomorphie des ganzen Ringes. Da jedes Element *eigentlich* ist, bildet der Strahl $\{\mathfrak{o} | \mathfrak{r}_1 | \mathfrak{s}_1\}$ einen Normalteiler von \mathfrak{R}_1 , den wir mit \mathfrak{S}_1 bezeichnen. Die Ordnung von \mathfrak{S}_1 ist gleich $\chi(\mathfrak{s}_1)$. (Vgl. VI, § 2.) Die Faktorgruppe $\mathfrak{R}_1 / \mathfrak{S}_1$ ist mit der Kongruenzgruppe $\mathfrak{G}(n_1, p)$ isomorph. (Vgl. I, § 2 und Satz 2.) Der Strahl $\{\mathfrak{o} | \mathfrak{r}_1 | \mathfrak{s}_{12}^{(1)}\}$ und der Strahl $\{\mathfrak{o} | \mathfrak{r}_1 | \mathfrak{s}_{21}^{(1)}\}$ sind Abelsche Gruppen $\mathfrak{S}_{12}^{(1)}$ und $\mathfrak{S}_{21}^{(1)}$ von der Ordnung $\chi(\mathfrak{s}_{12}^{(1)})$ und von dem Typus (p, p, \dots, p) (vgl. II, V, VI § 2 und Satz 5), und sie erzeugen \mathfrak{S}_1 (vgl. III, § 2). Der Durchschnitt \mathfrak{T}_1 von $\mathfrak{S}_{12}^{(1)}$ und $\mathfrak{S}_{21}^{(1)}$ ist gleich dem Strahl $\{\mathfrak{o} | \mathfrak{r}_1 | \mathfrak{t}_1\}$. \mathfrak{T}_1 ist daher eine Untergruppe des Zentrums von \mathfrak{R}_1 (vgl. IV, § 2 und Satz 5) von der Ordnung $\chi(\mathfrak{t}_1)$ (vgl. III, § 2). Für die Restklassenringe $\mathfrak{r}_2, \mathfrak{r}_3, \dots, \mathfrak{r}_e$ kann man auch genau wie oben vorgehen und die Faktorgruppen $\mathfrak{R}_2, \mathfrak{R}_3, \dots, \mathfrak{R}_e$ untersuchen. Damit ist (A) bewiesen.

(B) Die Koeffizienten der diagonalen Teilmatrizen jedes Elementes von \bar{r}_1^* sind durch p teilbar (vgl. § 2). Jedes Element des Strahls $\{\mathfrak{o} | \bar{r}_1^*\}$ ist daher eigentlich (vgl. Satz 2). $\bar{\mathfrak{H}}_\kappa$, $\kappa > 1$, ist eine Abelsche Gruppe von der Ordnung $\chi(\bar{r}_\kappa)$ und von dem Typus (p, p, \dots, p) (vgl. VII, VIII, IX, § 2 und Satz 5). $\bar{\mathfrak{H}}_1$ hat eine Kompositionsreihe $\bar{\mathfrak{S}}_e, \bar{\mathfrak{S}}_{e-1}, \dots, \bar{\mathfrak{S}}_2, \bar{\mathfrak{T}}$. $\bar{\mathfrak{S}}_\kappa$ ist eine Abelsche Gruppe von der Ordnung $\chi(\bar{s})$ und von dem Typus (p, p, \dots, p) (vgl. VIII, X, XII, § 2 und Satz 5). $\bar{\mathfrak{T}}$ ist mit dem direkten Produkt der Kongruenzgruppen $\mathfrak{G}(n_\kappa, p)$, $n_\kappa \neq 0$, $\kappa = 1, 2, \dots, e$, isomorph (vgl. XI, § 2 und Satz 6), da hier kein Faktor von $\bar{\mathfrak{T}}$ gleich \mathfrak{G} ist. Damit ist (B) bewiesen.

Aus diesem Satz folgt

Satz 8. *Die Automorphismengruppe einer endlichen Abelschen Gruppe mit n_κ Basiselementen von der Ordnung p^κ , $\kappa = 1, 2, \dots, e$, ist dann und nur dann auflösbar, wenn die Kongruenzgruppe des Grades n_κ modulo p , $\kappa = 1, 2, \dots, e$, auflösbar ist.*

Man kann diesen Satz auch direkt so beweisen¹⁷⁾: Die Gesamtheit der Elemente P (4) § 1 von \mathfrak{o} , derart, daß jeder Koeffizient der diagonalen Teilmatrizen $P_{\kappa\kappa}$ durch p teilbar ist, bildet ein nilpotentes Ideal \mathfrak{m} in \mathfrak{o} , da jeder Koeffizient von $P_{\kappa\lambda}$, $\kappa \geq \lambda$, dann durch p teilbar ist. (\mathfrak{m} ist ferner das maximale nilpotente Ideal von \mathfrak{o} , da der Restklassenring $\mathfrak{o}/\mathfrak{m}$ die direkte Summe von $\mathfrak{o}(n_\kappa, p)$ ist, die nach dem MacLagan-Wedderburnschen Satz kein echtes (zweiseitiges) Ideal enthalten.) Der \mathfrak{m} entsprechende Normalteiler \mathfrak{M} von \mathfrak{G} ist gleich dem Strahl $\{\mathfrak{o} | \mathfrak{m}\}$. Denn jeder Koeffizient von $P_{\kappa\kappa}$ ist durch p teilbar, jedes Element des Strahls ist also eigentlich¹⁸⁾. Man bilde die Reihe von Idealen $\mathfrak{m}, \mathfrak{m}^2, \dots, \mathfrak{m}^e = 0$ und die entsprechende Reihe von Normalteilern $\mathfrak{M}, \mathfrak{M}_2, \dots, \mathfrak{M}_e = \mathfrak{G}$ von \mathfrak{G} . Dann erkennt man nach Satz 4, daß dem Restklassenring $\mathfrak{m}^x/\mathfrak{m}^{x+1}$ die Faktorgruppe $\mathfrak{M}_x/\mathfrak{M}_{x+1}$ entspricht. Jedes Produkt zweier Elemente von $\mathfrak{m}^x/\mathfrak{m}^{x+1}$ ist aber Null. Nach Satz 5 ist also $\mathfrak{M}_x/\mathfrak{M}_{x+1}$ mit $\mathfrak{m}^x/\mathfrak{m}^{x+1}$ — dieser aufgefaßt als Abelsche Gruppe gegenüber der Addition — isomorph. Daher ist \mathfrak{M} eine auflösbare Gruppe. Die Faktorgruppe $\mathfrak{G}/\mathfrak{M}$ ist aber mit dem direkten Produkt von $\mathfrak{G}(n_\kappa, p)$ isomorph, woraus Satz 8 folgt.

¹⁷⁾ Dieser Beweis wurde bei der Korrektur (im Juni 1928) hinzugefügt.

¹⁸⁾ Ist allgemein P ein nilpotentes Element eines Rings mit Einheitselement E , so ist $E+P$ eine Einheit des Rings. Es sei nämlich $P^e = 0$, $e \leq 2^\sigma$. Dann ist $(E+P) \cdot (E-P) \cdot (E+P^2) \cdot \dots \cdot (E+P^{2^{\sigma-1}}) = E - P^{2^\sigma} = E$.

Galoissche Theorie der unendlichen algebraischen Erweiterungen.

Von

Wolfgang Krull in Freiburg i. Br.

Der Hauptsatz der Galoisschen Theorie besagt im wesentlichen, daß (bei algebraischen Erweiterungen „erster Art“¹⁾) die Untergruppen der Abbildungsgruppe Γ eines Normalkörpers \mathfrak{N} eindeutig umkehrbar den Unterkörpern von \mathfrak{N} entsprechen — falls \mathfrak{N} über seinem Grundkörper \mathfrak{K}_0 endlich ist, d. h. aus \mathfrak{K}_0 durch Adjunktion von endlich viel Elementen erzeugt werden kann. Ist dagegen \mathfrak{N} über \mathfrak{K}_0 unendlich, verliert, wie aus einem von Dedekind²⁾ angegebenen Beispiel zu ersehen, die eineindeutige Zuordnung zwischen Untergruppen und Unterkörpern ihre Allgemeingültigkeit. Im folgenden wird gezeigt, wie dem hier vorliegenden Übelstand abgeholfen werden kann. Der dabei benutzte Grundgedanke findet sich andeutungsweise schon bei Dedekind, nämlich in der Bemerkung, daß die Gesamtheit der Abbildungen eines (speziellen) unendlichen Normalkörpers „gewissermaßen eine stetige Mannigfaltigkeit“ bilde.

Wir machen die Abbildungsgruppe Γ eines (beliebigen) unendlichen Normalkörpers \mathfrak{N} durch Einführung eines geeigneten Umgebungsbegriffs zu einem topologischen Raum und erhalten dann eineindeutiges Entsprechen zwischen den Unterkörpern von \mathfrak{N} und den im Sinne der Topologie abgeschlossenen Untergruppen von Γ . Insbesondere ergibt sich, daß die Ge-

¹⁾ Die algebraischen Erweiterungen erster Art sind nach E. Steinitz (Algebraische Theorie der Körper, J. f. Math. 130 (1908), S. 167—303) dadurch ausgezeichnet, daß nur Nullstellen irreduzibler Polynome mit lauter verschiedenen Wurzeln adjungiert werden.

²⁾ Dedekind, Über die Permutationen des Körpers aller algebraischen Zahlen; Abhandlungen d. Gesellschaft d. Wissenschaften z. Göttingen 1901 (Festschrift). Bei Dedekind handelt es sich an der betr. Stelle um den Körper aller p^n -ten Einheitswurzeln (p feste Primzahl, $n = 1, 2, \dots$). — In Zukunft wird die Dedekindsche Arbeit kurz mit „D.“ zitiert.

samtheit der Abbildungen aus Γ , die den Invariantenkörper einer beliebigen Untergruppe Δ von Γ elemente weis ungeändert lassen, gerade mit der abgeschlossenen Hülle von Δ identisch ist. Daß es in der Abbildungsgruppe Γ eines beliebigen unendlichen Normalkörpers \mathfrak{K} stets nichtabgeschlossene Untergruppen gibt, folgt im allgemeinen Fall auf Grund des Hauptsatzes im wesentlichen aus körpertheoretischen Überlegungen.

Besonders einfach gestalten sich die Verhältnisse, falls der gegebene Normalkörper \mathfrak{K} über seinem Grundkörper abzählbar ist. Die Abbildungsgruppe Γ ist hier als Raum kompakt und metrisierbar, die Existenz von nichtabgeschlossenen Untergruppen ergibt sich aus elementaren topologischen Betrachtungen, und es besitzt jede abgeschlossene Untergruppe von Γ ein System von höchstens abzählbar unendlich viel Erzeugenden. (Dabei ist allerdings der Begriff des Erzeugendensystems allgemeiner zu fassen als üblich.)

Die Handlichkeit der eingeführten Begriffsbildungen wird schließlich noch an einem Beispiel gezeigt. Nennen wir eine Abbildungsgruppe mit einer einzigen Erzeugenden ideal-zyklisch³⁾, so gilt der Satz, daß der Normalkörper \mathfrak{K} dann und nur dann eine ideal-zyklische Abbildungsgruppe besitzt, wenn jeder endliche Unterkörper von \mathfrak{K} im üblichen Sinne zyklisch ist; ferner zeigt es sich, daß jede ideal-zyklische Gruppe dargestellt werden kann als direktes Produkt von höchstens abzählbar unendlich viel speziellen ideal-zyklischen Gruppen, die isomorph sind entweder zur Additionsgruppe des Restklassensystems nach einer Primzahlpotenz oder zur Additionsgruppe der ganzen p -adischen Zahlen.

§ 1.

Der allgemeine Fall \mathfrak{U} .

Bei unsern Untersuchungen gehen wir von einem Grundkörper \mathfrak{K}_0 aus, der in allen auftretenden Körpern als Unterkörper enthalten sein soll; die Bezeichnung „normal“ bezieht sich stets auf das Verhältnis zu \mathfrak{K}_0 . Ein Körper heißt endlich, abzählbar unendlich, beliebig (algebraisch über \mathfrak{K}_0), wenn er aus \mathfrak{K}_0 durch Adjunktion einer Menge relativ zu \mathfrak{K}_0 algebraischer Elemente von endlicher, abzählbar unendlicher, beliebiger Mächtigkeit entsteht. Endliche Normalkörper bezeichnen wir mit \mathfrak{C} , abzählbar unendliche mit \mathfrak{U} , beliebige mit \mathfrak{U} (oder gelegentlich auch mit \mathfrak{K}). Für die Permutationen (Isomorphismen, Automorphismen) eines Normalkörpers verwenden wir das Wort „Abbildungen“. Abbildungen werden mit kleinen, Abbildungsgruppen und Abbildungssysteme mit großen griechischen Buchstaben bezeichnet.

³⁾ Wegen des Wortes „idealzyklisch“ vgl. Anm. ¹³⁾. „Erzeugende“ ist auch hier in dem am Schlusse von § 1 präzisierten verallgemeinerten Sinn aufzufassen.

Nach Steinitz hat man bei beliebigem Ausgangskörper \mathfrak{K}_0 zwischen algebraischen Erweiterungen erster und zweiter Art zu unterscheiden¹⁾, und es ist bekannt, daß zur Erkenntnis des Aufbaus eines Erweiterungskörpers zweiter Art die Galoissche Theorie allein nicht ausreicht. Wir beschränken uns daher im folgenden ausschließlich auf Erweiterungen erster Art.

Es sei jetzt \mathfrak{U} ein fester Normalkörper über \mathfrak{K}_0 , Γ sei die Galoissche Gruppe seiner Abbildungen. Zu jedem Unterkörper \mathfrak{K} von \mathfrak{U} gibt es eine eindeutig bestimmte Untergruppe $\Delta(\mathfrak{K})$ von Γ , nämlich die größte Untergruppe von Γ , die die Eigenschaft hat, daß ihre sämtlichen Abbildungen den Körper \mathfrak{K} elementeweis ungeändert lassen; entsprechend bestimmt jede Untergruppe Δ von Γ eindeutig einen „Invariantenkörper“ $\mathfrak{K}(\Delta)$, nämlich den größten Unterkörper von \mathfrak{K} , dessen Elemente bei sämtlichen Abbildungen von Δ ungeändert bleiben. Ist nun $\mathfrak{U} = \mathfrak{C}$, so gelten die „reziproken Gleichungen“

$$(1) \quad \underline{\mathfrak{K}(\Delta(\mathfrak{K})) = \mathfrak{K},^4}$$

$$(2) \quad \underline{\Delta(\mathfrak{K}(\Delta)) = \Delta,}$$

und man erhält eine eindeutige umkehrbare Zuordnung zwischen den Unterkörpern von \mathfrak{U} und den Untergruppen von Γ , eine Zuordnung, deren Existenz eben den Hauptsatz der gewöhnlichen Galoisschen Theorie ausmacht. Ist hingegen \mathfrak{U} nicht endlich, so bleibt, wie von Dedekind²⁾ gezeigt wurde, zwar die Gleichung (1) richtig, die Gleichung (2) aber verliert ihre Allgemeingültigkeit, es wird im allgemeinen $\Delta(\mathfrak{K}(\Delta))$ eine echte Obergruppe von Δ darstellen. Es erhebt sich also im Falle eines unendlichen \mathfrak{U} die Frage nach einer unabhängigen Charakterisierung derjenigen Untergruppen Δ von Γ , die der Gleichung $\Delta(\mathfrak{K}(\Delta)) = \Delta$ genügen. Ist \mathfrak{C} irgendein in \mathfrak{U} enthaltener endlicher Normalkörper, so vermittelt jede Abbildung γ von \mathfrak{U} eine Abbildung von \mathfrak{C} , die mit $\gamma_{\mathfrak{C}}$ bezeichnet werden möge. Offenbar gilt die Gleichung $(\gamma \cdot \delta)_{\mathfrak{C}} = \gamma_{\mathfrak{C}} \cdot \delta_{\mathfrak{C}}$, und daraus folgt, daß die Gesamtheit derjenigen Abbildungen von \mathfrak{C} , die durch die Abbildungen einer Untergruppe Δ von Γ vermittelt werden, eine Abbildungsgruppe von \mathfrak{C} darstellt, für die wir zweckmäßig die Schreibweise $\Delta_{\mathfrak{C}}$ benutzen. Ist $\gamma_{\mathfrak{C}}$ irgendeine feste Abbildung von \mathfrak{C} , so verstehen wir unter $\Phi(\gamma_{\mathfrak{C}})$ ⁵⁾ dasjenige Teilsystem von Γ , das aus allen den Abbildungen δ besteht, die der Gleichung $\gamma_{\mathfrak{C}} = \delta_{\mathfrak{C}}$ genügen. *Wir fassen nun die Elemente von Γ als*

⁴⁾ Formel (1) gilt nur für Erweiterungen erster Art. Der Beweis von Formel (1) kann am einfachsten nach dem Vorbild von D. § 6 geführt werden, nur muß man im allgemeinen Falle die Wohlordenbarkeit einer beliebigen algebraischen Erweiterung an Stelle der Abzählbarkeit einer algebraischen Erweiterung eines algebraischen Zahlkörpers benutzen.

⁵⁾ Die Systeme $\Phi(\gamma_{\mathfrak{C}})$ bilden die Nebengruppen bei der Zerlegung von Γ nach $\Delta(\mathfrak{C})$.

Punkte eines topologischen Raumes auf und ordnen dem Punkt γ als Umgebungen die sämtlichen Punktmenge $\Phi(\gamma_{\mathfrak{G}})$ für beliebigen in \mathfrak{U} enthaltenen \mathfrak{G} zu.

Auf diese Weise wird Γ zu einem brauchbaren topologischen Raum, denn es gelten die folgenden Tatsachen⁶⁾:

a) Der Punkt γ ist nach Definition von $\Phi(\gamma_{\mathfrak{G}})$ in jeder seiner Umgebungen enthalten. — *Umgebungsaxiom A.*

b) Der Vereinigungskörper $\mathfrak{G}_1 + \mathfrak{G}_2$ zweier endlicher Normalkörper \mathfrak{G}_1 und \mathfrak{G}_2 stellt selbst einen endlichen Normalkörper dar, und man verifiziert leicht, daß $\Phi(\gamma_{\mathfrak{G}_1 + \mathfrak{G}_2})$ gleich ist dem Durchschnitt von $\Phi(\gamma_{\mathfrak{G}_1})$ und $\Phi(\gamma_{\mathfrak{G}_2})$. Der Durchschnitt zweier Umgebungen ist also stets selbst eine Umgebung. — *Umgebungsaxiom B.*⁷⁾

c) Ist δ ein Punkt der Umgebung $\Phi(\gamma_{\mathfrak{G}})$ von γ , so ist $\delta_{\mathfrak{G}} = \gamma_{\mathfrak{G}}$, $\Phi(\gamma_{\mathfrak{G}})$ ist also auch Umgebung von δ , jede Umgebung stellt eine offene Punktmenge dar. — *Umgebungsaxiom C.*⁷⁾

d) Sind γ und δ verschiedene Punkte von Γ , so gibt es mindestens einen \mathfrak{G} , für den $\gamma_{\mathfrak{G}} \neq \delta_{\mathfrak{G}}$ ausfällt, und dann haben wir in $\Phi(\gamma_{\mathfrak{G}})$ und $\Phi(\delta_{\mathfrak{G}})$ zwei elementefremde Untermengen von Γ , deren eine eine Umgebung von γ , deren andere eine Umgebung von δ ist. — *Trennungsaxiom.*

Auf die Gültigkeit eines Abzählbarkeitsaxioms und auf die Metrisierbarkeit von Γ können wir nur im Falle $\mathfrak{U} = \mathfrak{A}$ rechnen, auf den wir in § 3 zurückkommen. Vorerst brauchen wir bloß die Tatsache, daß mit Hilfe der Umgebungen in Γ die Begriffe des *Häufungspunktes* (Hp.) und der *abgeschlossenen Punktmenge* in üblicher Weise definiert werden können.

Hauptsatz 1. *Jede Untergruppe $\Delta(\mathfrak{R})$ aus Γ ist abgeschlossen, umgekehrt genügt jede abgeschlossene Untergruppe Δ^* von Γ der Gleichung $\Delta(\mathfrak{R}(\Delta^*)) = \Delta^*$. Die Unterkörper von \mathfrak{U} entsprechen also eindeutig umkehrbar den abgeschlossenen Untergruppen von Γ .*⁸⁾

⁶⁾ In der Bezeichnung der Axiome schließe ich mich an F. Hausdorff, Grundzüge der Mengenlehre, 2. Aufl., an.

⁷⁾ Es sei hervorgehoben, daß unter b) und c) mehr bewiesen ist, als die Hausdorffschen Axiome B. und C. fordern:

B. verlangt nur, daß der Durchschnitt zweier Umgebungen von γ eine Umgebung von γ enthält, nicht, daß er selbst eine Umgebung ist, wie für unsern Raum unter b) gezeigt wird.

Nach C. soll jeder Punkt δ einer Umgebung Φ von γ eine Umgebung Ψ besitzen, die ganz in Φ liegt. Unter c) dagegen wird die viel weitergehende Tatsache bewiesen, daß in Γ eine Umgebung von γ auch Umgebung jedes der in ihr enthaltenen Punkte ist.

⁸⁾ Man beachte, daß der Hauptsatz auch für $\mathfrak{U} = \mathfrak{G}$ gilt; denn dann ist Γ endlich und jede Untergruppe abgeschlossen.

a) Es sei δ^* ein Hp. von $\Delta(\mathfrak{K})$. Zu jedem Element a aus \mathfrak{K} gibt es einen a enthaltenden \mathfrak{G} , und nach Definition des Hp. muß es in Δ ein der Gleichung $\delta_{\mathfrak{G}} = \delta_{\mathfrak{G}}^*$ genügendes δ geben. a bleibt daher nach der Definition von $\Delta(\mathfrak{K})$ bei der Abbildung δ^* ungeändert, d. h. δ^* läßt den ganzen Körper \mathfrak{K} in Ruhe und gehört mithin zu $\Delta(\mathfrak{K})$. — $\Delta(\mathfrak{K})$ ist abgeschlossen.

b) Es sei Δ^* abgeschlossen, $\mathfrak{K} = \mathfrak{K}(\Delta^*)$, \mathfrak{G} beliebig, $\mathfrak{K}_{\mathfrak{G}}$ sei der Durchschnitt von \mathfrak{K} und \mathfrak{G} . $\mathfrak{K}_{\mathfrak{G}}$ ist offenbar der Invariantenkörper der durch Δ^* vermittelten Abbildungsgruppe $\Delta_{\mathfrak{G}}^*$ von \mathfrak{G} , d. h. als Unterkörper von \mathfrak{G} genügt $\mathfrak{K}_{\mathfrak{G}}$ der Gleichung $\mathfrak{K}_{\mathfrak{G}} = \mathfrak{K}(\Delta_{\mathfrak{G}}^*)$, und daraus ergibt sich wegen der Endlichkeit von \mathfrak{G} weiter $\Delta_{\mathfrak{G}}^* = \Delta(\mathfrak{K}_{\mathfrak{G}})$.⁹⁾ Ist daher δ^* ein Element aus Γ , das \mathfrak{K} und damit jeden $\mathfrak{K}_{\mathfrak{G}}$ elementeweis ungeändert läßt, so muß $\delta_{\mathfrak{G}}^*$ stets in $\Delta_{\mathfrak{G}}^*$ enthalten sein, d. h. δ^* ist Hp. von Δ^* und gehört wegen der Abgeschlossenheitsvoraussetzung zu dieser Gruppe, die folglich der Gleichung $\Delta^* = \Delta(\mathfrak{K}(\Delta^*))$ genügt.

Durch a) und b) ist der Beweis des Hauptsatzes bereits erledigt.

Gleichzeitig mit δ ist stets auch δ^{-1} und gleichzeitig mit γ und δ stets auch $\gamma \cdot \delta$ Hp. der Gruppe Δ . Die abgeschlossene Hülle Δ^* einer Gruppe Δ ist daher selbst Gruppe, und aus Hauptsatz 1 folgt unmittelbar:

Satz 1. *Die abgeschlossene Hülle Δ^* der Gruppe Δ ist die größte Untergruppe von Γ , die den zu Δ gehörigen Invariantenkörper $\mathfrak{K}(\Delta)$ elementeweis ungeändert läßt.*

Satz 2. *Über abgeschlossene Gruppen gelten die folgenden Tatsachen:*

a) *Jede unendliche Untergruppe von Γ ist in sich dicht, jede abgeschlossene unendliche Untergruppe ist also perfekt.*

b) *Ist Δ kommutativ, so ist auch die abgeschlossene Hülle Δ^* von Δ kommutativ.*

c) *Eine abgeschlossene Gruppe Δ^* ist durch die Gesamtheit der Gruppen $\Delta_{\mathfrak{G}}^*$ eindeutig bestimmt.*

d) *Die abgeschlossene Gruppe Δ^* ist dann und nur dann invariant in Γ , wenn für jedes \mathfrak{G} die Gruppe $\Delta_{\mathfrak{G}}^*$ invariant ist in $\Gamma_{\mathfrak{G}}$.*

a) Es genügt zu zeigen, daß das Einheitselement ε Hp. in Δ ist, d. h. daß es zu jedem \mathfrak{G} in Δ eine Abbildung $\delta \neq \varepsilon$ gibt, die \mathfrak{G} elementeweis ungeändert läßt. Das folgt aber unmittelbar aus dem Umstand, daß die Abbildungsgruppe $\Gamma_{\mathfrak{G}}$ von \mathfrak{G} endlich ist, daß es also unter den unendlich vielen Elementen von Δ sicher zwei verschiedene gibt, die dieselbe Abbildung von \mathfrak{G} vermitteln.

b) Man beachte, daß die Elemente γ und δ dann und nur dann vertauschbar sind, wenn für jeden \mathfrak{G} die Gleichung $\gamma_{\mathfrak{G}} \cdot \delta_{\mathfrak{G}} = \delta_{\mathfrak{G}} \cdot \gamma_{\mathfrak{G}}$ gilt!

⁹⁾ Hierbei ist $\Delta_{\mathfrak{G}}^*$ als Untergruppe der Gesamtabbildungsgruppe $\Gamma_{\mathfrak{G}}$ von \mathfrak{G} aufgefaßt.

c) Die Behauptung folgt aus der Tatsache, daß durch die $\Delta_{\mathfrak{G}}^*$ die isolierten Punkte und die Hp. von Δ^* eindeutig bestimmt sind.

d) Ist jedes $\Delta_{\mathfrak{G}}^*$ invariant in $\Gamma_{\mathfrak{G}}$, so ergibt sich aus der Gleichung $(\beta^{-1} \cdot \alpha \cdot \beta)_{\mathfrak{G}} = \beta_{\mathfrak{G}}^{-1} \cdot \alpha_{\mathfrak{G}} \cdot \beta_{\mathfrak{G}}$, daß für beliebige α aus Δ^* und beliebige β aus Γ das Element $\beta^{-1} \cdot \alpha \cdot \beta$ stets einen Hp. von Δ^* darstellt, also zu Δ^* gehört. Δ^* ist daher invariant in Γ . Der Beweis der Umkehrung ist trivial.

Die Bedeutung des Invarianzkriteriums von Satz 2 d) liegt darin, daß die abgeschlossenen invarianten Untergruppen von Γ eindeutig umkehrbar den Normalunterkörpern von \mathfrak{U} entsprechen.

Wir schließen den Paragraphen mit einer Definition, die vor allem in § 3 und § 4 wichtig sein wird:

Die Elementemenge M aus Γ heißt ein „volles System von Erzeugenden“ der abgeschlossenen Gruppe Δ^ , wenn Δ^* die abgeschlossene Hülle der kleinsten Gruppe Δ ist, die die sämtlichen Elemente von M enthält.*

§ 2.

Existenz von nichtabgeschlossenen Untergruppen.

Es soll jetzt gezeigt werden, daß Γ , falls unendlich, stets nicht-abgeschlossene Untergruppen enthält. Im allgemeinen Fall muß der Beweis dieser Tatsache im wesentlichen auf Grund von körpertheoretischen Überlegungen geführt werden.

Hilfssatz 1. *Ist \mathfrak{U}_1 ein Normalunterkörper von \mathfrak{U} , $\gamma_{\mathfrak{U}_1}$ eine Abbildung von \mathfrak{U}_1 , so gibt es für $\mathfrak{U}_1 \neq \mathfrak{U}$ stets mindestens zwei verschiedene Abbildungen von \mathfrak{U} , die gerade die Abbildung $\gamma_{\mathfrak{U}_1}$ von \mathfrak{U}_1 vermitteln.*

Der Beweis kann im wesentlichen nach Dedekindschem Vorbild geführt werden¹⁰⁾; nur muß man an Stelle der Abzählbarkeit der algebraischen Zahlen die Wohlordenbarkeit von \mathfrak{U} zu Hilfe nehmen.

Hilfssatz 2. *Die Abbildungsgruppe Γ eines unendlichen Normalkörpers ist niemals abzählbar.*

Der Beweis folgt nach bekannten Schlüssen aus Hilfssatz 1 sowie aus der Tatsache, daß zwischen \mathfrak{U} und dem Grundkörper eine abzählbare Kette $\mathfrak{U}_1, \mathfrak{U}_2, \mathfrak{U}_3, \dots$ von Normalkörpern eingeschaltet werden kann, bei der \mathfrak{U}_{i+1} jeweils einen echten Oberkörper von \mathfrak{U}_i darstellt.

Hauptsatz 2. *Die Abbildungsgruppe Γ eines unendlichen Normalkörpers \mathfrak{U} enthält stets nichtabgeschlossene Untergruppen.*

¹⁰⁾ Vgl. D. § 4, Satz IV.

Da Γ unendlich ist, gibt es mindestens eine abzählbare Untergruppe Δ , die nach Hilfssatz 2 sicher von ihrer abgeschlossenen Hülle Δ^* verschieden ist, weil nach Satz 1 die Gruppe Δ^* die Abbildungsgruppe von \mathbb{U} über dem Grundkörper $\mathfrak{K}(\Delta)$ darstellt.

§ 3.

Der Fall \mathfrak{A}^{11} .

Wir untersuchen jetzt den Fall $\mathbb{U} = \mathfrak{A}$ genauer, der z. B. immer dann vorliegt, wenn wir es mit algebraischen Zahlkörpern zu tun haben. In \mathbb{U} gibt es eine, und damit unendlich viele „Körperfundamentalfolgen $\{\mathfrak{E}_i\}$ “ mit den folgenden Eigenschaften: 1. \mathfrak{E}_{i+1} ist Erweiterungskörper von \mathfrak{E}_i . 2. \mathfrak{A} ist Vereinigungskörper der sämtlichen \mathfrak{E}_i ($\mathfrak{A} = \mathfrak{E}_1 + \mathfrak{E}_2 + \mathfrak{E}_3 + \dots$). Für die weiteren Untersuchungen von § 3 legen wir eine feste Fundamentalfolge $\{\mathfrak{E}_i\}$ zugrunde und bezeichnen mit $\gamma_i^{(k)}$ ($\kappa = 1, 2, \dots, r_i$) (statt mit $\gamma_{\mathfrak{E}_i}^{(k)}$) die endlich viel verschiedenen Abbildungen von \mathfrak{E}_i . Das System Σ der abzählbar unendlich viel Umgebungen $\Phi(\gamma_i^{(k)})$ ($i = 1, 2, \dots$) bildet ein „volles Umgebungssystem“ im Hausdorffischen Sinne, d. h. *der Raum Γ genügt dem zweiten Abzählbarkeitsaxiom*. In Zukunft werden außer den $\Phi(\gamma_i^{(k)})$ keine weiteren Umgebungen benutzt.

Eine *Umgebungsfolge* $\{\Phi(\gamma_i^{(k_i)})\}$ ($i = 1, 2, \dots$) heißt „*Fundamentalfolge*“, wenn $\Phi(\gamma_{i+1}^{(k_{i+1})})$ stets in $\Phi(\gamma_i^{(k_i)})$ enthalten ist, d. h. wenn $\gamma_i^{(k_i)}$ jeweils die durch $\gamma_{i+1}^{(k_{i+1})}$ vermittelte Abbildung von \mathfrak{E}_i darstellt.

Satz 3. *Die Punkte γ von Γ entsprechen eindeutig umkehrbar den Umgebungsfundamentalfolgen $\{\Phi(\gamma_i^{(k_i)})\}$.*

Ist einerseits γ aus Γ gegeben, und bedeutet $\gamma_i^{(k_i)}$ die durch γ vermittelte Abbildung von \mathfrak{E}_i , so stellt $\{\Phi(\gamma_i^{(k_i)})\}$ eine durch γ eindeutig bestimmte Umgebungsfundamentalfolge dar; ist andererseits $\{\Phi(\gamma_i^{(k_i)})\}$ vorgelegt, so liefert die Abbildungsfolge $\{\gamma_i^{(k_i)}\}$ der Körper \mathfrak{E}_i eine bestimmte Abbildung γ von \mathfrak{A} , die ihrerseits für jedes i gerade die Abbildung $\gamma_i^{(k_i)}$ von \mathfrak{E}_i vermittelt.

Satz 4. *Der Raum Γ ist kompakt.*

Es sei \mathbb{U} irgendeine unendliche Untermenge von Γ . Da es nur endlich viel verschiedene Abbildungen $\gamma_1^{(k)}$ gibt, muß eine unendliche Untermenge \mathbb{U}_1 von \mathbb{U} existieren, deren Punkte sämtlich in ein und derselben festen Umgebung $\Phi(\gamma_1^{(k_1)})$ liegen. Ebenso muß \mathbb{U}_1 eine unendliche Untermenge \mathbb{U}_2 besitzen, die ganz in einer festen (in $\Phi(\gamma_1^{(k_1)})$ enthaltenen) Um-

¹¹⁾ Ursprünglich hatte ich bei meinen Untersuchungen nur den Fall $\mathbb{U} = \mathfrak{A}$ im Auge. Auf die Möglichkeit der Verallgemeinerung auf beliebige Erweiterungen machte mich J. v. Neumann aufmerksam.

gebung $\Phi(\gamma_2^{(k_2)})$ liegt usw. Man sieht, es gibt eine Umgebungsfundamentalfolge $\{\Phi(\gamma_i^{(k_i)})\}$, deren Glieder sämtlich unendlich viel Elemente von \mathfrak{U} enthalten, und der durch $\{\Phi(\gamma_i^{(k_i)})\}$ nach Satz 3 bestimmte Punkt γ ist Hp. von \mathfrak{U} .

Satz 5. *Jede perfekte Untermenge von Γ ist nicht abzählbar.*

Beweis auf Grund von Satz 3 nach der vom n -dimensionalen Kontinuum her geläufigen Methode. Kombiniert man Satz 5 mit Satz 2a, so ergibt sich sofort, daß jede abzählbare Untergruppe von Γ nicht abgeschlossen ist. *Hauptsatz 2 kann daher für $\mathfrak{U} = \mathfrak{A}$ ohne Benutzung des Wohlordnungssatzes mit ganz einfachen topologischen Überlegungen bewiesen werden.*

Liegen die Punkte γ und γ' in ein und derselben Umgebung $\Phi(\gamma_{i-1}^{(k_{i-1})})$, aber in verschiedenen Umgebungen $\Phi(\gamma_i^{(k_i)})$ und $\Phi(\gamma_i^{(l)})$, so soll i^{-1} als der „Abstand $|\gamma - \gamma'|$ “ von γ und γ' bezeichnet werden¹²⁾.

Man verifiziert leicht, daß bei dieser Abstandsdefinition die Dreiecksungleichung in der Form $|\alpha - \gamma| \leq \max(|\alpha - \beta|, |\beta - \gamma|)$ erfüllt ist, daß also durch unsere Festsetzung Γ zu einem brauchbaren metrischen Raum gemacht wird. Als den *absoluten Betrag* eines Elementes γ aus Γ bezeichnet man zweckmäßig seinen Abstand von der Identität ε .

Nennen wir eine Folge mit einem einzigen Hp. *konvergent*, so gilt:

Satz 6. *Die Folge $\{\gamma^{(k)}\}$ ist dann und nur dann konvergent, wenn das „Cauchysche Konvergenzkriterium“ erfüllt ist, d. h. wenn $|\gamma^{(k)} - \gamma^{(l)}|$ ($k \geq l$, sonst beliebig) mit wachsendem k gegen 0 geht.*

Beweis ohne wesentliche Schwierigkeit nach üblichem Schema. Um zu zeigen, daß das Kriterium notwendig ist, braucht man Satz 4, um zu zeigen, daß es hinreicht, Satz 3.

Ein *unendliches Produkt* $\gamma^{(1)} \cdot \gamma^{(2)} \cdot \gamma^{(3)} \dots$ heiße *konvergent* mit dem Grenzwert γ , wenn die Folge $\{\delta^{(i)} = \gamma^{(1)} \cdot \gamma^{(2)} \dots \gamma^{(i)}\}$ gegen γ konvergiert.

Satz 4. *Das unendliche Produkt $\gamma^{(1)} \cdot \gamma^{(2)} \cdot \gamma^{(3)} \dots$ konvergiert dann und nur dann, wenn die Folge $\{\gamma^{(i)}\}$ gegen die Identität konvergiert.*

Beweis nach bekannten Mustern. Daß die Bedingung hinreicht, folgt aus der Tatsache, daß in Γ der absolute Betrag eines Produktes nie größer ist als der absolute Betrag des absolut größten Faktors.

Unter dem „gewöhnlichen 1. Gruppenpostulat“ wollen wir die Tatsache verstehen, daß in einer Gruppe Δ mit γ und δ stets auch das Produkt $\gamma \cdot \delta$ vorkommen muß, das „*verschärfte 1. Gruppenpostulat*“ dagegen soll identisch sein mit der folgenden Forderung:

¹²⁾ Unter $\Phi(\gamma_0^{(1)})$ versteht man zweckmäßig den Gesamttraum Γ . Der Abstand eines Punktes von sich selbst ist gleich 0 zu setzen.

Sind $\gamma^{(1)}, \gamma^{(2)}, \gamma^{(3)}, \dots$ irgendwelche endlich oder abzählbar unendlich viele Elemente aus Δ , so enthält Δ auch das Produkt $\gamma^{(1)} \cdot \gamma^{(2)} \cdot \gamma^{(3)} \dots$, sofern nur dasselbe bei unendlich viel Komponenten (in einem vorher zu präzisierenden Sinn) konvergiert.

Eine Untergruppe von Γ , die dem verschärften 1. Gruppenpostulat genügt, soll „scharfe Untergruppe“ heißen.

Satz 7. Die scharfen Untergruppen von Γ sind identisch mit den abgeschlossenen Untergruppen.

a) Ist $\gamma^{(1)} \cdot \gamma^{(2)} \cdot \gamma^{(3)} \dots$ ein konvergentes Produkt von Elementen aus Δ mit dem Grenzwert γ , so ist γ Hp. von Δ . — Die abgeschlossenen Untergruppen von Γ sind also scharfe Untergruppen.

b) Ist γ Hp. von Δ , $\{\delta^{(i)}\}$ eine Folge aus Δ mit dem Grenzwert γ , $\gamma^{(i)} = \delta^{(i-1)^{-1}} \cdot \delta^{(i)}$ ($\delta^{(0)} = \varepsilon$), so ist $\gamma = \gamma^{(1)} \cdot \gamma^{(2)} \cdot \gamma^{(3)} \dots$, die scharfen Untergruppen aus Γ sind also abgeschlossen.

Hauptsatz 3. Um den eineindeutigen Zusammenhang zwischen den Unterkörpern von \mathfrak{A} und den Untergruppen von Γ herzustellen, genügt es, bei der Gruppdefinition an Stelle des gewöhnlichen 1. Gruppenpostulates das verschärfte 1. Gruppenpostulat (mit dem oben auseinandergesetzten Konvergenzbegriff) zugrunde zu legen.

Satz 8. Jede scharfe Untergruppe Δ von Γ besitzt (im Sinne von § 1) ein volles System von höchstens abzählbar unendlich viel Erzeugenden.

Es seien $\Phi(\gamma_i^{(l)})$ ($i = 1, 2, \dots; l = l_1, l_2, \dots, l_{m_i}$) diejenigen unter den Umgebungen $\Phi(\gamma_i^{(k)})$, in denen Punkte von Δ vorkommen, $\gamma^{(i,l)}$ sei jeweils ein in $\Phi(\gamma_i^{(l)})$ vorkommendes Element aus Δ . Dann bilden offenbar die abzählbar unendlich viel Elemente $\gamma^{(i,l)}$ ein volles System von Erzeugenden von Δ . — Man erkennt schließlich noch leicht die Richtigkeit der folgenden Tatsache:

Ist \mathfrak{U} ein volles System von Erzeugenden der scharfen Gruppe Δ , so besteht Δ aus allen endlichen und konvergent unendlichen Produkten, die man aus den Elementen von \mathfrak{U} und ihren Reziproken bilden kann.

§ 4.

Ideal-zyklische Abbildungsgruppen¹³⁾.

Die Abbildungsgruppe Z eines Normalkörpers \mathfrak{Z} soll „ideal-zyklisch“ heißen, wenn \mathfrak{Z} (im Sinne von § 3) durch ein einziges Element erzeugt

¹³⁾ Ich spreche von „idealzyklischen“ Gruppen, weil die idealzyklischen Abbildungsgruppen ideale Abelsche Gruppen sind im Sinne von H. Prüfer: „Theorie der Abelschen Gruppen: II. Ideale Gruppen“ (Math. Zeitschr. 22 (1925), S. 222–249). Allgemeiner sind, soweit ich sehe, die sämtlichen Abelschen Abbildungsgruppen samt ihren

wird. Eine ideal-zyklische Gruppe ist nach Satz 2 stets kommutativ. Ferner gilt:

Satz 9. *Besitzt der Normalkörper \mathfrak{Z} eine ideal-zyklische Abbildungsgruppe Z , so ist jeder endliche Normalunterkörper von \mathfrak{Z} , d. h. in diesem Falle jeder endliche Unterkörper von \mathfrak{Z} , im üblichen Sinne zyklisch.*

Es sei γ irgendein erzeugendes Element von Z , H die Gruppe der positiven und negativen Potenzen von γ . Dann muß es in H stets ein Element geben, das eine beliebig vorgeschriebene Abbildung eines beliebig in \mathfrak{Z} enthaltenen \mathfrak{E} vermittelt. Daraus folgt aber, daß die Abbildungsgruppe von \mathfrak{E} zu einer Quotientengruppe von H isomorph und damit zyklisch ist, d. h. \mathfrak{E} stellt einen zyklischen Körper im üblichen Sinne dar. — Es handelt sich jetzt um die Umkehrung von Satz 9. Wir führen zu diesem Zweck die folgenden Abkürzungen und Begriffsbildungen ein:

$\mathfrak{Z} p^r$ ($r = 1, 2, \dots$) möge einen endlichen zyklischen Körper vom Primzahlpotenzgrad p^r bedeuten; unter $\mathfrak{Z} p^\infty$ hingegen möge ein abzählbar unendlicher Normalkörper verstanden werden, zu dem im Sinne von § 3 eine erzeugende Körperfundamentalfolge von der speziellen Gestalt $\{\mathfrak{E}_i = \mathfrak{Z} p^i\}$ gehört. $\mathfrak{Z} p^\infty$ besitzt außer den in der definierenden Fundamentalfolge auftretenden $\mathfrak{Z} p^i$ keine weiteren nichttrivialen Unterkörper. In dem Symbol $\mathfrak{Z} p^r$ kann in Zukunft r sowohl eine endliche positive Zahl als auch ∞ bedeuten.

Wir nennen den Körper \mathfrak{K} das *direkte Produkt* der endlich oder abzählbar unendlich vielen Unterkörper $\mathfrak{K}_1, \mathfrak{K}_2, \mathfrak{K}_3, \dots$ und schreiben $\mathfrak{K} = \mathfrak{K}_1 \cdot \mathfrak{K}_2 \cdot \mathfrak{K}_3 \cdot \dots$, wenn erstens \mathfrak{K} Vereinigungskörper der \mathfrak{K}_i ist, und wenn zweitens \mathfrak{K}_i mit dem Vereinigungskörper der sämtlichen \mathfrak{K}_k ($k \neq i$) nur den Grundkörper als Durchschnitt gemein hat.

Satz 10. *Ist jeder endliche Unterkörper des Normalkörpers \mathfrak{Z} im üblichen Sinne zyklisch, so besitzt \mathfrak{Z} eine (eindeutig bestimmte) direkte Produktdarstellung $\mathfrak{Z} = \mathfrak{Z} p_1^{r_1} \cdot \mathfrak{Z} p_2^{r_2} \cdot \mathfrak{Z} p_3^{r_3} \cdot \dots$, wobei die Primzahlen p_i sämtlich verschieden sind.*

a) Es sei \mathfrak{Z}_p der Vereinigungskörper aller der endlichen Unterkörper von \mathfrak{Z} , deren Grad eine Potenz von p ist. Dann gilt die Gleichung

abgeschlossenen Untergruppen im Prüferschen Sinne „ideal“, während den nicht-abgeschlossenen Untergruppen diese Eigenschaft nicht zukommt. Doch wäre der hier vorliegende Zusammenhang noch genauer zu untersuchen.

Die idealzyklischen Abbildungsgruppen sind sämtlich isomorph zu (echten oder unechten) Quotientengruppen der Additionsgruppe derjenigen „idealen Zahlen“, die H. Prüfer in der Arbeit: „Begründung der algebraischen Zahlentheorie“ (Math. Annalen 94 (1925), S. 198–243) den gewöhnlichen ganzen Zahlen zugeordnet hat. Man könnte die Theorie der idealzyklischen Abbildungsgruppen auf die Herstellung des Zusammenhanges mit der zuletzt erwähnten Prüferschen Arbeit gründen. Doch führt bei unserem Ausgangspunkt der im Text eingeschlagene Weg wohl rascher zum Ziel.

$\mathfrak{B} = \mathfrak{B}_{p_1} \cdot \mathfrak{B}_{p_2} \cdot \mathfrak{B}_{p_3} \cdot \dots$, weil jeder endliche zyklische Körper als direktes Produkt von zyklischen Körpern von Primzahlpotenzgrad dargestellt werden kann. (Das Produkt ist dabei nur über diejenigen Primzahlen p zu erstrecken, für die \mathfrak{B} wirklich einen Unterkörper enthält, dessen Grad eine positive Potenz von p ist.)

b) \mathfrak{B}_p ist ein \mathfrak{B}_{p^n} , denn der Vereinigungskörper zweier endlicher zyklischer Körper $\mathfrak{B}_{p^{n_1}}$ und $\mathfrak{B}_{p^{n_2}}$ ($n_2 \geq n_1$) ist dann und nur dann selbst zyklisch, wenn $\mathfrak{B}_{p^{n_1}}$ in $\mathfrak{B}_{p^{n_2}}$ enthalten ist. Durch a) und b) ist die Existenz der in Satz 10 geforderten Produktzerlegung bewiesen. Auch die Eindeutigkeit läßt sich ohne Schwierigkeit zeigen.

Satz 11. *Satz 9 läßt sich umkehren, d. h. es gilt: Ein Normalkörper \mathfrak{B} besitzt dann und nur dann eine idealzyklische Abbildungsgruppe Z , wenn jeder endliche Unterkörper von \mathfrak{B} im üblichen Sinne zyklisch ist.*

In der Tat, genügt \mathfrak{B} der angegebenen Voraussetzung, so besitzt \mathfrak{B} nach Satz 10 eine direkte Produktdarstellung $\mathfrak{B} = \mathfrak{B}_{p_1^{r_1}} \cdot \mathfrak{B}_{p_2^{r_2}} \cdot \mathfrak{B}_{p_3^{r_3}} \cdot \dots$, und es gibt in Z ein Element γ , das für jedes i eine von der Identität verschiedene Abbildung des in $\mathfrak{B}_{p_i^{r_i}}$ enthaltenen \mathfrak{B}_p vermittelt. Bedeutet H die Gruppe der positiven und negativen Vielfachen von γ , so gilt für $\mathfrak{C} = \mathfrak{B}_{p_1^{e_1}} \cdot \mathfrak{B}_{p_2^{e_2}} \cdot \dots \cdot \mathfrak{B}_{p_r^{e_r}}$ ($e_i \leq r_i$; r endlich) und damit für jeden in \mathfrak{B} enthaltenen \mathfrak{C} die Gleichung $Z_{\mathfrak{C}} = H_{\mathfrak{C}}$;¹⁴⁾ daraus folgt aber, daß Z mit der abgeschlossenen Hülle von H identisch und mithin idealzyklisch ist.

Satz 11 zeigte die Zweckmäßigkeit der folgenden Definition:

Ein Normalkörper \mathfrak{B} soll dann und nur dann idealzyklisch heißen, wenn seine Abbildungsgruppe idealzyklisch ist.

Unter Zp^r ($r = 1, 2, \dots$) wollen wir eine endliche zyklische Gruppe von der Primzahlpotenzordnung p^r verstehen, unter Zp^∞ hingegen eine Gruppe, die zur Additionsgruppe der ganzen p -adischen Zahlen isomorph ist. Ferner möge das ideale direkte Gruppenprodukt folgendermaßen definiert werden:

Die kommutative Gruppe Γ heißt *ideales direktes Produkt*¹⁵⁾ der endlich oder abzählbar unendlich vielen Untergruppen $\Gamma_1, \Gamma_2, \Gamma_3, \dots$, und wir schreiben $\Gamma = \Gamma_1 \cdot \Gamma_2 \cdot \Gamma_3 \cdot \dots$, wenn jedes Element aus Γ dargestellt werden kann als (endliches oder unendliches) Produkt je eines Elementes aus Γ_i , wenn umgekehrt das Produkt je eines Elementes aus Γ_i auch bei

¹⁴⁾ Es wird also hier die Tatsache benutzt, daß das Element γ_e Erzeugendes der Abbildungsgruppe Z_e des endlichen zyklischen Körpers \mathfrak{B}_e ist, wenn bei der Abbildung γ_e keiner der in \mathfrak{B}_e enthaltenen zyklischen Körper vom Primzahlgrad elementeweise un-geändert bleibt.

¹⁵⁾ Man vergleiche die „additive Zerlegung“ der idealen Zahlen in § 7 der in Anm. ¹³⁾ an zweiter Stelle zitierten Prüferschen Arbeit.

unendlich viel Faktoren stets existiert und zu Γ gehört, und wenn schließlich (für jedes i) Γ_i zur Vereinigungsgruppe der sämtlichen Γ_k ($k \neq i$) elementfremd ist.

Satz 12. Jede idealzyklische Abbildungsgruppe Z besitzt eine (eindeutig bestimmte) ideale direkte Punktzerlegung $Z = Z p_1^{r_1} \cdot Z p_2^{r_2} \cdot Z p_3^{r_3} \cdot \dots$, wobei die Primzahlen p_i sämtlich verschieden sind¹⁶⁾.

a) Es sei $\mathfrak{B} = \mathfrak{B} p_1^{r_1} \cdot \mathfrak{B} p_2^{r_2} \cdot \mathfrak{B} p_3^{r_3} \cdot \dots$ die direkte Produktzerlegung des zu Z gehörigen Normalkörpers \mathfrak{B} , Z_{p_i} sei die größte Untergruppe von Z , bei der die sämtlichen Körper $\mathfrak{B} p_k^{r_k}$ ($k \neq i$) elementweis ungeändert bleiben; dann ist Z_{p_i} zur Abbildungsgruppe von $\mathfrak{B} p_i^{r_i}$ isomorph, und es gilt die Gleichung $Z = Z_{p_1} \cdot Z_{p_2} \cdot Z_{p_3} \cdot \dots$.

b) Für endliches r ist offenbar die Abbildungsgruppe von $\mathfrak{B} p^r$ eine $Z p^r$. Wir zeigen noch, daß die Abbildungsgruppe Z^* von $\mathfrak{B} p^\infty$ eine $Z p^\infty$ darstellt. Es sei γ ein erzeugendes Element von Z^* ; dann sieht man zunächst leicht, daß es unter den positiven Potenzen von γ stets ein Element gibt, das eine beliebige vorgeschriebene Abbildung eines beliebigen endlichen Unterkörpers von $\mathfrak{B} p^\infty$ vermittelt, und daß daher jedes Element δ aus Z^* im Sinne von § 3 als Grenzwert einer konvergenten Folge $\{\gamma^{n_i}\}$ ($i = 1, 2, \dots; n_i \geq 0$) dargestellt werden kann. Aus der speziellen Natur von $\mathfrak{B} p^\infty$ folgt weiter: Die Folge $\{\gamma^{n_i}\}$ konvergiert dann und nur dann, wenn $n_i \equiv n_{i+1} \pmod{p^k}$ für $i \geq i_0(k)$ ($k = 1, 2, \dots$); die konvergenten Folgen $\{\gamma^{n_i}\}$ und $\{\gamma^{m_i}\}$ liefern dann und nur dann denselben Grenzwert, wenn $n_i \equiv m_i \pmod{p^k}$ für $i \geq i_0(k)$. Schließlich ist unmittelbar klar, daß $\delta_1 \cdot \delta_2 = \lim \{\gamma^{n_i + m_i}\}$, falls $\delta_1 = \lim \{\gamma^{n_i}\}$, $\delta_2 = \lim \{\gamma^{m_i}\}$. — Die zusammengestellten Tatsachen zeigen: Jedem Element δ aus Z^* kann eindeutig umkehrbar eine ganze p -adische Zahl \tilde{n} zugeordnet werden, mit deren Hilfe man zweckmäßig δ in der symbolischen Form $\delta = \gamma^{\tilde{n}}$ schreibt. Es gilt dann das Multiplikationsgesetz $\gamma^{\tilde{n}} \cdot \gamma^{\tilde{m}} = \gamma^{\tilde{n} + \tilde{m}}$.

Es ist damit die Isomorphie von Z^* zur Additionsgruppe der ganzen p -adischen Zahlen bewiesen, und die Existenz der in Satz 12 geforderten idealen direkten Produktzerlegung endgültig gesichert. Auch die Eindeutigkeit läßt sich leicht zeigen, und zwar entweder unmittelbar oder mit Hilfe der Eindeutigkeit der Zerlegung von Satz 10.

¹⁶⁾ Die Additionsgruppe Z^* der zu den ganzen rationalen Zahlen nach Prüfer gehörigen idealen Zahlen besitzt den Typ: $Z^* = Z 2^\infty \cdot Z 3^\infty \cdot Z 5^\infty \cdot \dots$. Daraus und aus Satz 12 ergibt sich sofort die in Anm. ¹³⁾ aufgestellte Behauptung über die Isomorphie der idealzyklischen Abbildungsgruppen mit den Quotientengruppen von Z^* .

Über Gitterpunkte in mehrdimensionalen Ellipsoiden.

Von

Vojtěch Jarník in Göttingen.

§ 1.

Einleitung.

Es sei $Q(u) = \sum_{\mu, \nu=1}^r a_{\mu, \nu} u_{\mu} u_{\nu}$ ($r \geq 4$) stets eine positiv-definite quadratische Form der u_{μ} . Für $x > 0$ sei $A_Q(x)$ die Anzahl der Gitterpunkte im r -dimensionalen Ellipsoid $Q(u) \leq x$; $J_Q(x)$ sei der Inhalt dieses Ellipsoids,

$$P_Q(x) = A_Q(x) - J_Q(x).$$

Dann gilt

$$(1) \quad P_Q(x) = O\left(x^{\frac{r}{2} - \frac{r}{r+1}}\right)^1, \quad P_Q(x) = O\left(x^{\frac{r-1}{4}}\right)^2.$$

Wenn insbesondere die $a_{\mu, \nu}$ rational sind, so ist³⁾

$$(2) \quad P_Q(x) = O\left(x^{\frac{r}{2}-1}\right) \quad \text{für } r > 4,$$

¹⁾ E. Landau, Zur analytischen Zahlentheorie der definiten quadratischen Formen, Berliner Akademieber. 1915, S. 458—476.

²⁾ E. Landau, Über die Anzahl der Gitterpunkte in gewissen Bereichen, vierte Abhandlung, Göttinger Nachr. 1924, S. 137—150.

³⁾ Vgl. die in der Math. Zeitschr. unter dem gemeinsamen Titel „Über Gitterpunkte in mehrdimensionalen Ellipsoiden“ erschienenen Arbeiten von E. Landau, erste Abhandlung 21 (1924), S. 126—132, zweite Abhandlung 24 (1926), S. 299—310; A. Walfisz, erste Abhandlung 19 (1924), S. 300—307, zweite Abhandlung 26 (1927), S. 106—124; V. Jarník 27 (1927), S. 154—160. Weiter: H. Petersson, Über die Anzahl der Gitterpunkte in mehrdimensionalen Ellipsoiden, Abhandl. aus dem Math. Seminar in Hamburg 5 (1926), S. 116—150. Ch. H. Müntz, Über den Gebrauch willkürlicher Funktionen in der analytischen Zahlentheorie I, Sitzungsber. der Berliner Math. Ges. 24, 2 (1925), S. 81—93 und Zur Gittertheorie n -dimensionaler Ellipsoide, Math. Zeitschr. 25 (1926), S. 150—165.

$$(3) \quad P_Q(x) = O\left(x^{\frac{r}{2}-1} \log^2 x\right) \quad \text{für } r = 4,$$

$$(4) \quad P_Q(x) = \Omega\left(x^{\frac{r}{2}-1}\right) \quad \text{für } r \geq 4.$$

Für $r = 4$ ist noch in einigen Spezialfällen (mit rationalen $a_{\mu,\nu}$) etwas mehr bekannt⁴).

In einer vor einigen Wochen erschienenen Arbeit⁵) hat Herr Walfisz das Problem von Formen mit *irrationalen* Koeffizienten in Angriff genommen und folgende Resultate erhalten:

Er betrachtet nur Formen von der speziellen Gestalt

$$Q(u) = \alpha u_1^2 + \sum_{\mu,\nu=2}^r a_{\mu,\nu} u_\mu u_\nu, \quad \alpha > 0 \text{ irrational, } a_{\mu,\nu} \text{ rational, } r \geq 10$$

und beweist:

$$1. \quad P_Q(x) = o\left(x^{\frac{r}{2}-1}\right).$$

2. Zu jedem $\varphi(x) > 0$ mit $\varphi(x) = o\left(x^{\frac{r}{2}-1}\right)$ läßt sich eine irrationale Zahl $\alpha > 0$ so finden, daß

$$P_Q(x) = \Omega(\varphi(x)).$$

3. Für fast alle $\alpha > 0$ ist trotzdem⁶)

$$P_Q(x) = O\left(x^{\frac{r}{2}-\frac{6}{5}} \log^{\frac{1}{4}} x\right).$$

Ich habe inzwischen das Problem mit einer anderen Methode angegriffen und werde hier folgendes zeigen:

Ich betrachte ausschließlich Formen⁷)

$$Q(u) = \alpha_1 u_1^2 + \alpha_2 u_2^2 + \dots + \alpha_r u_r^2, \quad \alpha_j > 0, \quad r \geq 4.$$

Für diese Formen werde ich beweisen:

$$1. \quad P_Q(x) = O\left(x^{\frac{r}{2}-1} \log x\right) \quad \text{für } r > 4,$$

$$P_Q(x) = O\left(x^{\frac{r}{2}-1} \log^2 x\right) \quad \text{für } r = 4. \quad (\text{Satz 1.})$$

⁴) E. Landau, Über die Anzahl der Gitterpunkte in gewissen Bereichen, Göttinger Nachr. 1912, S. 687—771; H. D. Kloosterman, Über Gitterpunkte in mehrdimensionalen Ellipsoiden, Math. Zeitschr. 24 (1926), S. 519—529; A. Walfisz, Teilerprobleme, Math. Zeitschr. 26 (1927), S. 66—88.

⁵) A. Walfisz, Über Gitterpunkte in mehrdimensionalen Ellipsoiden, dritte Abhandlung, Math. Zeitschr. 27 (1927), S. 245—268.

⁶) „Fast alle“ bedeutet: alle bis auf eine Menge vom Maß Null. „Maß“ bedeutet in dieser Arbeit das Lebesguesche Maß.

⁷) Rationale α_j (einige oder alle) werden nicht ausgeschlossen.

Daß sich dieses Resultat nicht wesentlich verbessern läßt, zeigt die im Falle rationaler α_j gültige Formel (4).

2. Für *fast alle* $\alpha_j > 0$ ist

$$P_Q(x) = O\left(x^{\frac{r}{4} + \varepsilon}\right)$$

bei jedem $\varepsilon > 0$ (Satz 3).

Außerdem werde ich mich mit speziellen Formen der Gestalt

$$(5) \quad Q(u) = \beta_1(u_{1,1}^2 + \dots + u_{r_1,1}^2) + \beta_2(u_{1,2}^2 + \dots + u_{r_2,2}^2) + \dots \\ + \beta_\sigma(u_{1,\sigma}^2 + \dots + u_{r_\sigma,\sigma}^2)$$

mit $\beta_j > 0$, $\sigma \geq 2$, $r_j \geq 4$ beschäftigen, für welche ich folgende, ziemlich scharfe Resultate beweisen werde:

3. Für *fast alle* $\beta_j > 0$ ist

$$P_Q(x) = O\left(x^{\frac{r}{2} - \sigma + \varepsilon}\right)$$

bei jedem $\varepsilon > 0$ (Satz 2).

4. Für *fast alle* $\beta_j > 0$ ist

$$P_Q(x) = \Omega\left(x^{\frac{r}{2} - \sigma} \log^{\frac{\sigma-1}{\sigma+1}} x\right) \quad (\text{Satz 4}).$$

5. Für *alle* $\beta_j > 0$ ist

$$P_Q(x) = \Omega\left(x^{\frac{r}{2} - \sigma}\right) \quad (\text{Satz 5}).$$

Ich könnte zwar auch für Formen (5) *ohne* die Voraussetzung $r_j \geq 4$ einige Resultate bekommen, und zwar unmittelbar aus der Methode des § 5, doch verzichte ich darauf, um nicht unübersichtliche Resultate zu häufen. Ebenso könnte man ohne Mühe das x^ε im Satz 2 und 3 und das

Glied $\log^{\frac{\sigma-1}{\sigma+1}} x$ im Satz 4 noch verschärfen. Endlich ist es zweifellos möglich, die hier erhaltenen Resultate auch auf Formen von einer etwas allgemeineren Gestalt zu übertragen.

Zum Beweis der Sätze 1, 2, 3 benutze ich den bekannten Hardy-Littlewoodschen Ansatz, wie er im Falle von Formen mit rationalen Koeffizienten von Herrn Petersson (loc. cit. ³) konsequent durchgeführt worden ist. Die erzeugende Funktion ist hier freilich keine Potenzreihe mehr, sondern eine Dirichletsche Reihe; statt der Cauchyschen Integralformel wird darum die Perronsche Formel (6) benutzt. Die Unendlichkeit des Integrationsweges bereitet hier einige Schwierigkeiten, die im § 2 überwunden werden; dann läßt sich schon Satz 1 leicht beweisen (§ 3). Dagegen brauche ich zum Beweis der Sätze 2 und 3 noch einen Hilfsatz über die Häufigkeit derjenigen Systeme von ganzen positiven Zahlen

$h_1, h_2, \dots, h_\sigma; k_1, k_2, \dots, k_\sigma$, die bei gegebenem $\gamma_1, \gamma_2, \dots, \gamma_\sigma$ ($\gamma_j > 0$) die $\sigma - 1$ Ausdrücke

$$\left| \frac{h_1}{k_1} \gamma_1 - \frac{h_j}{k_j} \gamma_j \right| \quad (j = 2, 3, \dots, \sigma)$$

„klein“ machen (§ 4).

Zum Beweis der Sätze 4, 5 benutze ich eine ganz elementare (bis auf die Benutzung eines Satzes von Herrn Khintchine bei Satz 4) Methode, die im wesentlichen auf einer Approximation der Formen (5) durch Formen mit rationalen Koeffizienten (oder vielmehr Koeffizientenverhältnissen) beruht.

§ 2.

Vorbereitungen.

Im folgenden sei immer r ganz, $r \geq 4$; $x > 4$; $\alpha_j > 0$ ($j = 1, 2, \dots, r$). Es sei $Q(u)$ eine positiv-definite quadratische Form von der Gestalt

$$Q(u) = \alpha_1 u_1^2 + \alpha_2 u_2^2 + \dots + \alpha_r u_r^2.$$

Wir setzen noch⁶⁾

$$A_Q(x) = \sum_{Q(m) \leq x} 1; \quad J_Q(x) = \frac{\pi^{\frac{r}{2}} x^{\frac{r}{2}}}{\sqrt{\alpha_1 \alpha_2 \dots \alpha_r} \Gamma\left(\frac{r}{2} + 1\right)} = M x^{\frac{r}{2}};$$

$$P_Q(x) = A_Q(x) - J_Q(x).$$

Dabei ist die Summation über alle ganzen m_1, m_2, \dots, m_r zu erstrecken, für welche $Q(m) \leq x$.

Es sei

$$\theta(s) = \sum_{m=-\infty}^{+\infty} e^{-m^2 s} \quad \text{für } \Re(s) > 0;$$

die Dirichletsche Reihe

$$\theta(\alpha_1 s) \theta(\alpha_2 s) \dots \theta(\alpha_r s) = \sum_{n=1}^{\infty} a_n e^{-\lambda_n s} \quad (0 = \lambda_1 < \lambda_2 < \lambda_3 < \dots)$$

ist für $\Re(s) > 0$ absolut konvergent und es ist

$$\sum_{\lambda_n \leq x} a_n = A_Q(x).$$

Für jedes $x > 4$, welches sich nicht in der Form $x = \sum_{j=1}^r m_j^2 \alpha_j$ mit ganzen m_j darstellen läßt, gilt demnach die Formel⁷⁾

$$(6) \quad A_Q(x) = \frac{1}{2\pi i} \int_{\frac{1}{x} - i\infty}^{\frac{1}{x} + i\infty} \theta(\alpha_1 s) \theta(\alpha_2 s) \dots \theta(\alpha_r s) \frac{e^{xs}}{s} ds.$$

⁶⁾ Quadratwurzeln sind mit positivem Realteil zu nehmen.

⁷⁾ Alle benutzten Integrationswege sind geradlinig. Vgl. z. B. Hardy-Riesz, *The General Theory of Dirichlet's Series*, Cambridge Tracts 18 (1915), S. 12.

Wir setzen noch $\Delta(x)$ gleich der kleinsten unter den Zahlen $|x - \sum_{j=1}^r m_j^2 \alpha_j|$, wo die m_j unabhängig voneinander alle ganzen Zahlen durchlaufen.

Hilfssatz 1. Es sei $\gamma > 0$, $\delta \geq 0$. Es sei die Abschätzung

$$P_Q(x) = O(x^\gamma \log^\delta x)$$

für diejenigen x richtig, für welche $\Delta(x) \geq \frac{1}{\frac{r}{x^2}}$; dann ist sie auch für stetig wachsendes x richtig.

Beweis. Es ist¹⁰⁾ $\sum_{\frac{x}{2} \leq Q(m) \leq x} 1 < cx^{\frac{r}{2}}$ für $x > c$; daher gibt es auf der

Strecke $\frac{x}{2} < y < x$ mindestens eine Zahl y mit $\Delta(y) > \frac{x}{cx^{\frac{r}{2}}} > \frac{1}{\frac{r}{y^2}}$. Gesetzt

nun, die Behauptung sei falsch; dann würde es eine Folge von Zahlenpaaren $x_1, D_1, x_2, D_2, \dots$ mit $x_n \rightarrow +\infty, D_n \rightarrow +\infty$ geben, so daß entweder

1. für alle n gelten würde

$$A_Q(x_n) > Mx_n^{\frac{r}{2}} + D_n x_n^\gamma \log^\delta x_n$$

oder

2. für alle n gelten würde

$$A_Q(x_n) < Mx_n^{\frac{r}{2}} - D_n x_n^\gamma \log^\delta x_n.$$

Aus 1. ergibt sich nun ein Widerspruch folgendermaßen: Es sei y_n die untere Grenze der y mit $y \geq x_n, \Delta(y) \geq \frac{1}{\frac{r}{y^2}}$. Dann wäre $\Delta(y_n) \geq \frac{1}{\frac{r}{y_n^2}}$, $x_n \leq y_n < 2x_n$ für $x_n > c$; weiter wäre

$$\Delta(y) \leq \frac{1}{\frac{r}{y^2}} \leq \frac{1}{\frac{r}{x_n^2}} \quad \text{für} \quad x_n < y \leq y_n;$$

also

$$\sum_{x_n < Q(m) \leq y_n} 1 \geq (y_n - x_n) \frac{1}{2} x_n^{\frac{r}{2}} - 1.$$

¹⁰⁾ Mit c bezeichne ich unterschiedslos positive, nur von $r, \alpha_1, \alpha_2, \dots, \alpha_r$ abhängige Zahlen.

Also

$$\begin{aligned} A_Q(y_n) &\geq Mx_n^{\frac{r}{2}} + (y_n - x_n) \frac{1}{2} x_n^{\frac{r}{2}} - 1 + D_n x_n^\gamma \log^\delta x_n \\ &\geq M \left(x_n^{\frac{r}{2}} + \frac{r}{2} (y_n - x_n) y_n^{\frac{r}{2}-1} \right) + \frac{D_n}{2^{1+\gamma}} y_n^\gamma \log^\delta y_n \\ &\geq M y_n^{\frac{r}{2}} + \frac{D_n}{2^{1+\gamma}} y_n^\gamma \log^\delta y_n \quad \text{für } y_n > c(\gamma, \delta), \end{aligned}$$

gegen die Voraussetzung, da $\Delta(y_n) \geq \frac{1}{y_n^{\frac{r}{2}}}$.

Ebenso ergibt sich ein Widerspruch aus 2.: Hier sei y_n die obere Grenze der $y \leq x_n$ mit $\Delta(y) \geq \frac{1}{y^{\frac{r}{2}}}$; also $\Delta(y_n) \geq \frac{1}{y_n^{\frac{r}{2}}}$, $\frac{x_n}{2} < y_n \leq x_n$; $\Delta(y) \leq \frac{1}{y^{\frac{r}{2}}}$ für $y_n \leq y < x_n$, alles für $x_n > c$. Also

$$\begin{aligned} \sum_{y_n < Q(m) \leq x_n} 1 &\geq (x_n - y_n) \frac{1}{2} y_n^{\frac{r}{2}} - 1, \\ A_Q(y_n) &\leq Mx_n^{\frac{r}{2}} + \frac{r}{2} (y_n - x_n) x_n^{\frac{r}{2}-1} + 1 - D_n y_n^\gamma \log^\delta y_n \\ &\leq M y_n^{\frac{r}{2}} - \frac{D_n}{2} y_n^\gamma \log^\delta y_n, \end{aligned}$$

wieder gegen die Voraussetzung.

Hilfssatz 2. Wenn x die Zahlen mit $\Delta(x) \geq \frac{1}{x^{\frac{r}{2}}}$ durchläuft, so ist

$$A_Q(x) = \frac{1}{2\pi i} \int_{\frac{1}{x} - ix^r}^{\frac{1}{x} + ix^r} \theta(\alpha_1 s) \theta(\alpha_2 s) \dots \theta(\alpha_r s) \frac{e^{xs}}{s} ds + O(1).$$

Beweis. Wegen (6) und weil der Integrand für konjugiert komplexe Werte von s konjugiert komplexe Werte annimmt, genügt es, für die betreffenden x zu beweisen, daß

$$(7) \quad \left| \int_{\frac{1}{x} + ix^r}^{\frac{1}{x} + i\xi} \theta(\alpha_1 s) \theta(\alpha_2 s) \dots \theta(\alpha_r s) \frac{e^{xs}}{s} ds \right| < c$$

für alle $\xi > x^r$ und alle $x > c$. Die linke Seite von (7) ist gleich

$$(8) \left\{ \begin{aligned} & \left| \sum_{(m)=-\infty}^{+\infty} \int_{\frac{1}{x} + i\xi}^{\frac{1}{x} + i\bar{\xi}} e^{\frac{(x - \sum_j m_j^2 \alpha_j)(\frac{1}{x} + ti)}{s}} ds \right| \\ & \leq \sum_{(m)=-\infty}^{+\infty} \left| \int_{x^r}^{\xi} \frac{e^{\frac{(x - \sum_j m_j^2 \alpha_j)(\frac{1}{x} + ti)}{t}}}{t} dt \right| + \frac{1}{x} \sum_{(m)=-\infty}^{+\infty} \left| \int_{x^r}^{\xi} \frac{e^{\frac{(x - \sum_j m_j^2 \alpha_j)(\frac{1}{x} + ti)}{t(\frac{1}{x} + ti)}}}{t(\frac{1}{x} + ti)} dt \right| \\ & \leq c \sum_{(m)=-\infty}^{+\infty} e^{-\sum_j m_j^2 \alpha_j \cdot \frac{1}{x}} \left(\frac{1}{x^r |x - \sum_j m_j^2 \alpha_j|} + \frac{1}{x^{r+1}} \right). \end{aligned} \right.$$

Es ist aber erstens $\frac{1}{|x - \sum_j m_j^2 \alpha_j|} \leq x^{\frac{r}{2}}$; zweitens $\sum_{Q(m) \leq 2^n x} 1 < c(2^n x)^{\frac{r}{2}}$
für $n \geq 0$, $x > c$. Also

$$\sum_{(m)=-\infty}^{+\infty} e^{-\sum_j m_j^2 \alpha_j \cdot \frac{1}{x}} < c \sum_{n=1}^{\infty} (2^n x)^{\frac{r}{2}} e^{-2^{n-1}} < cx^{\frac{r}{2}};$$

also ist die rechte Seite in (8) kleiner als

$$cx^{\frac{r}{2}} \left(\frac{1}{x^r} x^{\frac{r}{2}} + \frac{1}{x^{r+1}} \right) < c \text{ für } x > c;$$

w. z. b. w.

Hilfssatz 3¹¹⁾. *Es sei A eine positive Zahl; dann ist*

$$(9) \frac{1}{2\pi i} \int_{\frac{1}{x} - \frac{A}{\sqrt{x}} i}^{\frac{1}{x} + \frac{A}{\sqrt{x}} i} \theta(\alpha_1 s) \theta(\alpha_2 s) \dots \theta(\alpha_r s) \frac{e^{xs}}{s} ds = \frac{\pi^{\frac{r}{2}} x^{\frac{r}{2}}}{\Gamma\left(\frac{r}{2} + 1\right) \sqrt{\alpha_1 \alpha_2 \dots \alpha_r}} + O\left(x^{\frac{r}{4}}\right).$$

Beweis. Nach einer bekannten Transformationsformel ist

$$\theta(\alpha_j s) = \sqrt{\frac{\pi}{\alpha_j s}} (1 + \psi_j(s)),$$

wo

$$\psi_j(s) = 2 \sum_{m=1}^{\infty} e^{-m^2 \frac{\pi^2}{\alpha_j s}};$$

¹¹⁾ Vgl. Petersson, loc. cit. ³⁾, § 2, wo analoge Rechnungen vorkommen.

wir setzen $s = \frac{1}{x} + it$; dann ist auf unserem Integrationsweg¹²⁾

$$\Re\left(\frac{1}{s}\right) = \frac{x}{1+x^2t^2} > c_1 \quad \text{für } x > c_1.$$

Also

$$|\psi_j(s)| \leq 2e^{-\frac{\pi^2 x}{\alpha_j(1+x^2t^2)}} \sum_{m=1}^{\infty} e^{-\frac{\pi^2 c_1}{\alpha_j}(m^2-1)} < c_1 e^{-c_1 \frac{x}{1+x^2t^2}}.$$

Demnach ist die linke Seite von (9) gleich

$$\frac{1}{2\pi} \cdot \frac{\pi^{\frac{r}{2}}}{\sqrt{\alpha_1 \alpha_2 \dots \alpha_r}} \int_{-\frac{A}{\sqrt{x}}}^{\frac{A}{\sqrt{x}}} \frac{e^{x\left(\frac{1}{x}+it\right)}}{\left(\frac{1}{x}+it\right)^{\frac{r}{2}+1}} \left(1 + \chi \cdot e^{-c_1 \frac{x}{1+x^2t^2}}\right) dt,$$

wo $|\chi| < c_1$. Hier ist erstens

$$(10) \quad \int_{-\frac{A}{\sqrt{x}}}^{\frac{A}{\sqrt{x}}} \left| \frac{e^{x\left(\frac{1}{x}+it\right)} \cdot e^{-c_1 \frac{x}{1+x^2t^2}}}{\left(\frac{1}{x}+it\right)^{\frac{r}{2}+1}} \right| dt = O\left(x^{\frac{r}{2}}\right) \int_0^{A\sqrt{x}} \frac{e^{-c_1 \frac{x}{1+t^2}}}{(1+t^2)^{\frac{r}{4}+\frac{1}{2}}} dt;$$

Das Maximum des Integranden für $t > 0$ liegt für $x > c_1$ bei $t^2 + 1 = \frac{c_1 x}{\frac{r}{2} + 1}$ und ist gleich $O\left(x^{-\frac{r}{4}-\frac{1}{2}}\right)$; daher ist die linke Seite von

(10) gleich $O\left(x^{\frac{r}{4}}\right)$. Weiter ist

$$\left| \int_{-\infty}^{\frac{A}{\sqrt{x}}} \frac{e^{x\left(\frac{1}{x}+it\right)}}{\left(\frac{1}{x}+it\right)^{\frac{r}{2}+1}} dt \right| = \left| \int_{\frac{A}{\sqrt{x}}}^{\infty} \frac{e^{x\left(\frac{1}{x}+it\right)}}{\left(\frac{1}{x}+it\right)^{\frac{r}{2}+1}} dt \right| = O \int_{\frac{A}{\sqrt{x}}}^{\infty} \frac{1}{t^{\frac{r}{2}+1}} dt = O\left(x^{\frac{r}{4}}\right).$$

¹²⁾ Mit c_1 bezeichne ich unterschiedslos positive Zahlen, die nur von α_j , r und A ($j = 1, 2, \dots, r$) abhängen.

Endlich ist

$$\int_{-\infty}^{\infty} \frac{e^{x\left(\frac{1}{x}+it\right)}}{\left(\frac{1}{x}+it\right)^{\frac{r}{2}+1}} dt = \frac{2\pi}{\Gamma\left(\frac{r}{2}+1\right)} x^{\frac{r}{2}},$$

womit die Behauptung bewiesen ist.

Hilfssatz 4. Wenn für irgendeine Form

$$Q(u) = \alpha_1 u_1^2 + \dots + \alpha_r u_r^2 \quad (\alpha_j > 0),$$

für irgendein $A > 0$, irgendein $\gamma \geq \frac{r}{4}$ und irgendein $\delta \geq 0$ gilt

$$\int_{\frac{1}{x}+ix^r}^{\frac{1}{x}+i\frac{A}{\sqrt{x}}} |\theta(\alpha_1 s) \dots \theta(\alpha_r s)| \frac{ds}{|s|} = O(x^\gamma \log^\delta x),$$

so ist

$$P_Q(x) = O(x^\gamma \log^\delta x).$$

Beweis. Folgt unmittelbar aus Hilfssatz 1, 2, 3, wenn wir noch bemerken, daß der Integrand für konjugiert komplexe Werte von s konjugiert komplexe Werte annimmt und $|e^{xs}| = e$ ist.

§ 3.

Allgemeine Bemerkungen und Satz 1.

Für das Folgende (§ 3—6) machen wir folgende Verabredungen. Wir legen, bei gegebenem $x > 4$, auf das Intervall $-\infty < t < +\infty$ alle Fareybrüche $\frac{h}{k}$ mit $h \geq 0$, $0 < k \leq \sqrt{x}$, $(h, k) = 1$ und konstruieren noch ihre Medianten, d. h. die Punkte $\frac{h+\bar{h}}{k+\bar{k}}$, wo $\frac{h}{k}$, $\frac{\bar{h}}{\bar{k}}$ zwei benachbarte von unseren Fareybrüchen sind. Mit $\mathfrak{B}_{h,k}$ bezeichnen wir das linksseitig abgeschlossene, rechtsseitig offene Intervall, dessen Endpunkte zwei benachbarte Medianten sind und welches den Punkt $\frac{h}{k}$ enthält. Bekanntlich ist

$$\mathfrak{B}_{h,k} = \left\langle \frac{h}{k} - \frac{\theta}{k\sqrt{x}}, \frac{h}{k} + \frac{\theta'}{k\sqrt{x}} \right\rangle,$$

wo θ, θ' noch von h, k, x abhängen, aber den Ungleichungen $\frac{1}{2} \leq \theta \leq 1$, $\frac{1}{2} \leq \theta' \leq 1$ genügen. Wenn a_1, a_2, a_3 drei reelle Zahlen sind, $a_1 > 0$, so wollen wir, wenn J das Intervall $\langle a_2, a_3 \rangle$ ist, mit $a_1 J$ das Intervall $\langle a_1 a_2, a_1 a_3 \rangle$ bezeichnen.

Es gilt nun bekanntlich¹³⁾ folgendes: Wenn $\binom{lm}{np}$ eine Modulusubsti-

¹³⁾ A. Krazer, Lehrbuch der Thetafunktionen, S. 183—185. Leipzig: B. G. Teubner 1903.

tution ist mit $n > 0$, $\frac{p}{n} = \frac{2h}{k}$ (also $n = k$ oder $n = \frac{k}{2}$), so ist

$$(11) \quad \theta(\alpha_j s) = \frac{\mathfrak{B}}{\sqrt{k(\alpha_j s - 2\pi i \frac{h}{k})}} \bar{\theta}\left(-\pi i \frac{l\alpha_j s - m\pi i}{n\alpha_j s - p\pi i}\right),$$

wo \mathfrak{B} , $\bar{\theta}$ noch von h, k, l, m, n, p abhängen, dabei aber $|\mathfrak{B}|$ kleiner als eine absolute Konstante ist und $\bar{\theta}(s)$ eine der folgenden Gestalten hat:

$$(12) \quad \left\{ \begin{array}{l} \bar{\theta}(s) = \sum_{m=-\infty}^{+\infty} e^{-m^2 s} \quad \text{oder} \quad \bar{\theta}(s) = \sum_{m=-\infty}^{+\infty} (-1)^m e^{-m^2 s} \quad \text{oder} \\ \bar{\theta}(s) = \sum_{m=-\infty}^{+\infty} e^{-\left(\frac{2m-1}{2}\right)^2 s}. \end{array} \right.$$

Wenn nun $s = \frac{1}{x} + ti$ und t im Intervall $\frac{2\pi}{\alpha_j} \mathfrak{B}_{h,k}$ liegt, so ist

$$\Re\left(-\pi i \frac{l\alpha_j s - m\pi i}{n\alpha_j s - p\pi i}\right) = \frac{\pi^2 \alpha_j}{x n^2 \left(\frac{\alpha_j^2}{x^2} + \left(\alpha_j t - \frac{2\pi h}{k}\right)^2\right)} > c;$$

denn es ist

$$\left|\alpha_j t - \frac{2\pi h}{k}\right| < \frac{c}{k\sqrt{x}}, \quad 0 < n \leq k \leq \sqrt{x}.$$

Daher ist für $s = \frac{1}{x} + ti$, t in $\frac{2\pi}{\alpha_j} \mathfrak{B}_{h,k}$ nach (11) und (12)

$$(13) \quad |\theta(\alpha_j s)| < \frac{c}{\sqrt{k} \sqrt[4]{\frac{1}{x^2} + \left(t - \frac{2\pi h}{\alpha_j k}\right)^2}},$$

also

$$(14) \quad |\theta(\alpha_j s)| < \frac{c}{\sqrt{k}} \text{Min}\left(\sqrt{x}, \frac{1}{\sqrt{\left|t - \frac{2\pi h}{\alpha_j k}\right|}}\right) \quad \left(\text{Min}\left(a, \frac{1}{0}\right) = a\right).$$

Satz 1. Es sei $Q(u) = \alpha_1 u_1^2 + \alpha_2 u_2^2 + \dots + \alpha_r u_r^2$, $\alpha_j > 0$, $r \geq 4$; dann ist

$$P_Q(x) = O\left(x^{\frac{r}{2}-1} \log x\right) \quad \text{für} \quad r > 4,$$

$$P_Q(x) = O\left(x^{\frac{r}{2}-1} \log^2 x\right) \quad \text{für} \quad r = 4.$$

Beweis. Ich setze $A = \text{Max} \frac{2\pi}{\alpha_j}$; dann ist

$$(15) \quad \int_{\frac{1}{x} + i\frac{A}{\sqrt{x}}}^{\frac{1}{x} + ix^r} \left| \frac{\theta(\alpha_1 s) \dots \theta(\alpha_r s)}{s} \right| ds \leq \frac{1}{r} \sum_{j=1}^r \int_{\frac{A}{\sqrt{x}}}^{x^r} \left| \frac{\theta^r(\alpha_j s)}{s} \right| dt,$$

wo $s = \frac{1}{x} + it$ (denn $\sqrt[r]{\alpha_1 \alpha_2 \dots \alpha_r} \leq \frac{1}{r}(\alpha_1 + \alpha_2 + \dots + \alpha_r)$ für $\alpha_j \geq 0$).

Aber

$$\int_{\frac{A}{\sqrt{x}}}^{x^r} \left| \frac{\theta^r(\alpha, s)}{s} \right| dt \leq \sum_{h, k} \int_{\frac{2\pi}{\alpha_j} \mathfrak{B}_{h, k}} \left| \frac{\theta^r(\alpha, s)}{s} \right| dt,$$

wo ich nur über diejenigen Paare von h, k summiere, für welche mindestens ein Punkt von $\frac{2\pi}{\alpha_j} \mathfrak{B}_{h, k}$ im Intervall $\left(\frac{A}{\sqrt{x}}, x^r\right)$ liegt. Es ist also sicher $h > 0$, da $\frac{2\pi}{\alpha_j} \mathfrak{B}_{0, 1}$ im Intervall $\left(-\frac{2\pi}{\alpha_j \sqrt{x}}, \frac{2\pi}{\alpha_j \sqrt{x}}\right)$ liegt. Weiter ist für t aus $\frac{2\pi}{\alpha_j} \mathfrak{B}_{h, k}$ (wegen $x > 4$) $|s| > |t| \geq \frac{2\pi}{\alpha_j} \left(\frac{h}{k} - \frac{1}{k\sqrt{x}}\right) > \frac{\pi}{\alpha_j} \frac{h}{k}$.

Nach (13) gilt also

$$\begin{aligned} \int_{\frac{2\pi}{\alpha_j} \mathfrak{B}_{h, k}} \left| \frac{\theta^r(\alpha, s)}{s} \right| dt &< c \int_{\frac{2\pi}{\alpha_j} \mathfrak{B}_{h, k}} \frac{k}{h} \frac{dt}{k^{\frac{r}{2}} \left(\frac{1}{x^2} + \left(t - \frac{2\pi h}{\alpha_j k}\right)^2\right)^{\frac{r}{4}}} \\ &< \frac{c}{k^{\frac{r}{2}-1} h} \int_{-\infty}^{+\infty} \frac{dt}{\left(\frac{1}{x^2} + t^2\right)^{\frac{r}{4}}} = \frac{c x^{\frac{r}{2}-1}}{k^{\frac{r}{2}-1} h}. \end{aligned}$$

Dieser Ausdruck ist über einige $h > 0$ zu summieren, wobei aber stets $h < c x^{r+\frac{1}{2}}$; dann über k mit $0 < k \leq \sqrt{x}$. Das ergibt

$$\int_{\frac{A}{\sqrt{x}}}^{x^r} \left| \frac{\theta^r(\alpha, s)}{s} \right| dt < c x^{\frac{r}{2}-1} \log x \Psi(x),$$

wo $\Psi(x) = 1$ für $r > 4$, $\Psi(x) = \log x$ für $r = 4$. Wegen (15) und Hilfssatz 4 ist damit der Satz 1 bewiesen.

§ 4.

Ein Hilfssatz über diophantische Approximationen.

Wir betrachten nun einen σ -dimensionalen Euklidischen Raum R_σ ; seine Punkte seien durch ihre rechtwinkligen Koordinaten $\gamma_1, \gamma_2, \dots, \gamma_\sigma$ charakterisiert. Es sei $\sigma > 1$. Es seien weiter drei Zahlen C, D, a gegeben mit $0 < C < D, a > 0$; mit W bezeichnen wir den Würfel $C \leq \gamma_j \leq D$ ($j = 1, 2, \dots, \sigma$); mit W_1 bezeichnen wir den Würfel $C \leq \gamma_j \leq D$ ($j = 2, \dots, \sigma$), den wir als Punktmenge eines $(\sigma - 1)$ -dimen-

sionalen Raumes $R_{1, \sigma-1}$ betrachten. Es seien nun ganze nichtnegative Zahlen¹⁴⁾

$$h_1 > 0, h_2 > 0, \dots, h_\sigma > 0, \quad k_1 > 0, k_2 > 0, \dots, k_\sigma > 0, \\ n_2, n_3, \dots, n_\sigma, \varrho$$

und eine Zahl γ_1 mit $C \leq \gamma_1 \leq D$ gegeben. Wir bezeichnen mit $M(h; k; n; \varrho; \gamma_1)$ die Menge derjenigen Punkte $(\gamma_2, \gamma_3, \dots, \gamma_\sigma)$ von W_1 , für welche gilt

$$(16) \quad \left| \frac{h_1}{k_1} \gamma_1 - \frac{h_j}{k_j} \gamma_j \right| < \frac{a}{2^{n_j} k_j 2^\varrho} \quad (j = 2, 3, \dots, \sigma).$$

Das Maß der Menge $M(h; k; n; \varrho; \gamma_1)$ im $(\sigma - 1)$ -dimensionalen Raum $R_{1, \sigma-1}$ ist höchstens

$$\prod_{j=2}^{\sigma} \frac{2a}{2^{n_j} h_j 2^\varrho}.$$

Nun sei $M(h; k; n; \varrho)$ die Menge von allen Punkten $(\gamma_1, \gamma_2, \dots, \gamma_\sigma)$ des σ -dimensionalen Würfels W , in welchen (16) gilt; ihr Maß in R_σ ist höchstens gleich

$$\int_C^D \prod_{j=2}^{\sigma} \frac{2a}{2^{n_j} h_j 2^\varrho} d\gamma_1;$$

also¹⁵⁾

$$\text{m } M(h; k; n; \varrho) \leq \frac{c_2}{2^{n_2+n_3+\dots+n_\sigma} h_2 h_3 \dots h_\sigma 2^{\varrho(\sigma-1)}}.$$

Aus (16) folgt noch

$$\left| \frac{h_1}{k_1} \frac{k_j}{h_j} \gamma_1 - \gamma_j \right| < \frac{a}{2^{n_j} h_j 2^\varrho} \quad (j = 2, 3, \dots, \sigma);$$

für $\varrho > c_2$ ist also, wenn $M(h; k; n; \varrho)$ nicht leer ist,

$$(17) \quad c_2 \frac{h_1}{k_1} < \frac{h_j}{k_j} < c_2 \frac{h_1}{k_1} \quad (j = 2, 3, \dots, \sigma)$$

Wir beschränken uns im folgenden auf solche ϱ .

Es sei nun $l; m_1, m_2, \dots, m_\sigma; n_2, n_3, \dots, n_\sigma; \varrho$ ein System von nichtnegativen ganzen Zahlen. Wir betrachten alle Mengen $M(h; k; n; \varrho)$, für welche gilt

$$(18) \quad 2^l \leq h_1 < 2^{l+1}; \quad 2^{m_j} \leq k_j < 2^{m_j+1} \quad (j = 1, 2, \dots, \sigma).$$

Wenn ein solches $M(h; k; n; \varrho)$ nicht leer sein soll, so muß (17) gelten, also

$$c_2 2^{l+m_j-m_1} < h_j < c_2 2^{l+m_j-m_1} \quad (j = 2, 3, \dots, \sigma).$$

¹⁴⁾ Man übersehe nicht die Unsymmetrie in bezug auf den Index 1!

¹⁵⁾ m M sei das Maß von M in R_σ ; mit c_2 bezeichne ich unterschiedslos positive Zahlen, die nur von σ, C, D, a abhängen.

Die Summe der Maße aller dieser $M(h; k; n; \varrho)$ ist daher höchstens gleich

$$\frac{c_2 2^{(\sigma-1)m_1}}{2^{n_2+n_3+\dots+n_\sigma} 2^{(\sigma-1)l} 2^{m_2+m_3+\dots+m_\sigma}} \cdot \frac{2^{\sigma l} 2^{m_2+m_3+\dots+m_\sigma} 2^{m_1+m_2+\dots+m_\sigma}}{2^{(\sigma-1)m_1}}$$

$$= c_2 \frac{2^l 2^{m_1+m_2+\dots+m_\sigma}}{2^{n_2+n_3+\dots+n_\sigma} 2^{(\sigma-1)l}} = \mathfrak{L}.$$

Wir betrachten nun die Menge $\bar{M}(l; m; n; \varrho)$ derjenigen Punkte des Würfels W , in welchen die Ungleichungen (16) für mehr als

$$\left((\varrho + 1)(l + 1) \prod_{j=1}^{\sigma} (m_j + 1) \prod_{j=2}^{\sigma} (n_j + 1) \right)^2 \frac{2^l 2^{m_1+m_2+\dots+m_\sigma}}{2^{n_2+n_3+\dots+n_\sigma} 2^{(\sigma-1)l}} = \mathfrak{M}$$

Wertsysteme von h_j, k_j mit (18) erfüllt sind; d. h. die Menge der Punkte, die in mehr als \mathfrak{M} Mengen $M(h; k; n; \varrho)$ mit (18) enthalten ist; das Maß dieser Menge ist offenbar höchstens $\frac{\mathfrak{L}}{\mathfrak{M}}$, d. h.

$$(19) \quad m \bar{M}(l; m; n; \varrho) \leq \frac{c_2}{\left((\varrho + 1)(l + 1) \prod_{j=1}^{\sigma} (m_j + 1) \prod_{j=2}^{\sigma} (n_j + 1) \right)^2};$$

alles für $\varrho > c_2$. Der Ausdruck rechts in (19) ist aber das allgemeine Glied einer konvergenten Reihe. Daher bilden diejenigen Punkte von W , die in unendlich vielen Mengen $\bar{M}(l; m; n; \varrho)$ mit $\varrho > c_2$ enthalten sind, eine Punktmenge vom Maß Null, die wir mit $N(C, D, a)$ bezeichnen. Insbesondere haben wir also folgenden Hilfssatz bewiesen:

Hilfssatz 5. *Es sei $0 < C < D$, $a > 0$, σ ganz, $\sigma > 1$. Dann gibt es im Würfel W :*

$$C \leq \gamma_j \leq D \quad (j = 1, 2, \dots, \sigma)$$

des σ -dimensionalen Raumes R_σ eine Punktmenge $N(C, D, a)$ vom Maß Null und von folgender Beschaffenheit:

Zu jedem Punkt $(\gamma_1, \gamma_2, \dots, \gamma_\sigma)$ von $W - N(C, D, a)$ gibt es eine Zahl

$$\varrho_0 = \varrho_0(\gamma_1, \gamma_2, \dots, \gamma_\sigma, a),$$

so daß für jedes System von nichtnegativen ganzen Zahlen

$$l; m_1, m_2, \dots, m_\sigma; n_2, n_3, \dots, n_\sigma; \varrho$$

mit $\varrho > \varrho_0$ die Ungleichungen

$$\left| \frac{h_1}{k_1} \gamma_1 - \frac{h_j}{k_j} \gamma_j \right| < \frac{a}{2^{n_j} k_j 2^\varrho} \quad (j = 2, 3, \dots, \sigma),$$

$$2^l \leq h_1 < 2^{l+1}, \quad 2^{m_1} \leq k_1 < 2^{m_1+1}, \quad 2^{m_2} \leq k_2 < 2^{m_2+1}, \dots,$$

$$2^{m_\sigma} \leq k_\sigma < 2^{m_\sigma+1}$$

höchstens

$$\mathfrak{M} = \left((\varrho + 1)(l + 1) \prod_{j=1}^{\sigma} (m_j + 1) \prod_{j=2}^{\sigma} (n_j + 1) \right)^2 \frac{2^l 2^{m_1 + m_2 + \dots + m_{\sigma}}}{2^{n_2 + n_3 + \dots + n_{\sigma}} 2^{(\sigma-1)l}}$$

Lösungen in ganzen h_j, k_j ($j = 1, 2, \dots, \sigma$) besitzen.

§ 5.

Satz 2.

Hilfssatz 6. Es sei $\sigma \geq 2$; $r_j \geq 4$ ($j = 1, 2, \dots, \sigma$); $\sum_{j=1}^{\sigma} r_j = r$; r_j, σ ganz. Dann gibt es in dem σ -dimensionalen Raum der Punkte $(\beta_1, \beta_2, \dots, \beta_{\sigma})$ eine Punktmenge N vom Maß Null und von folgender Beschaffenheit: Für jeden Punkt $(\beta_1, \beta_2, \dots, \beta_{\sigma})$ mit $\beta_j > 0$ ($j = 1, 2, \dots, \sigma$), welcher nicht zu N gehört, und jedes $\varepsilon > 0$ ist

$$(20) \quad \int_{\frac{A}{\sqrt{x}}}^{x^r} \left| \frac{\theta^{r_1}(\beta_1 s) \theta^{r_2}(\beta_2 s) \dots \theta^{r_{\sigma}}(\beta_{\sigma} s)}{s} \right| dt = O\left(x^{\frac{r}{2} - \sigma + \varepsilon}\right),$$

wo $A = \max \frac{2\pi}{\beta_j}$, $s = \frac{1}{x} + it$.

Aus Hilfssatz 6 folgt dann wegen Hilfssatz 4 sofort

Satz 2. Es sei $\sigma \geq 2$, $r_j \geq 4$ ($j = 1, 2, \dots, \sigma$); $\sum_{j=1}^{\sigma} r_j = r$; r_j, σ ganz. Es sei

$$Q(u) = \beta_1 (u_{1,1}^2 + u_{2,1}^2 + \dots + u_{r_1,1}^2) + \beta_2 (u_{1,2}^2 + u_{2,2}^2 + \dots + u_{r_2,2}^2) \\ + \dots + \beta_{\sigma} (u_{1,\sigma}^2 + u_{2,\sigma}^2 + \dots + u_{r_{\sigma},\sigma}^2).$$

Dann gibt es in dem σ -dimensionalen Raum der Punkte $(\beta_1, \beta_2, \dots, \beta_{\sigma})$ eine Punktmenge N vom Maß Null und von folgender Beschaffenheit:

Für jeden Punkt $(\beta_1, \beta_2, \dots, \beta_{\sigma})$ mit $\beta_j > 0$ ($j = 1, 2, \dots, \sigma$), welcher nicht zu N gehört und jedes $\varepsilon > 0$ ist

$$P_Q(x) = O\left(x^{\frac{r}{2} - \sigma + \varepsilon}\right).$$

Beweis des Hilfssatzes 6. Weil eine Vereinigungsmenge von abzählbar vielen Nullmengen¹⁶⁾ wieder eine Nullmenge ist, genügt es zu zeigen: Zu jedem \bar{C}, \bar{D} mit $0 < \bar{C} < \bar{D}$ gibt es eine Nullmenge $\bar{N}(\bar{C}, \bar{D})$, so daß in jedem Punkt des Würfels \bar{W} : $\bar{C} \leq \beta_j \leq \bar{D}$ ($j = 1, 2, \dots, \sigma$), der nicht zu $\bar{N}(\bar{C}, \bar{D})$ gehört, für jedes $\varepsilon > 0$ die Beziehung (20) erfüllt ist.

¹⁶⁾ Statt „Menge vom Maß Null“ sage ich im folgenden oft „Nullmenge“.

Es sei dauernd $x > 4$ und wir greifen aus \overline{W} irgendeinen Punkt $(\beta_1, \beta_2, \dots, \beta_\sigma)$ heraus. Zu jedem t unseres Integrationsintervalls $(\text{Max} \frac{2\pi}{\beta_j} \frac{1}{\sqrt{x}}, x^r)$ und zu jedem j ($j = 1, 2, \dots, \sigma$) gibt es genau ein Paar h_j, k_j so, daß t im Intervall $\frac{2\pi}{\beta_j} \mathfrak{B}_{h_j, k_j}$ liegt; diese Werte h_j, k_j sind also bestimmte Funktionen von t , die freilich noch von $x, \beta_1, \beta_2, \dots, \beta_\sigma$ abhängen. Wir bilden nun zu jedem System von ganzen Zahlen

$$n_1, n_2, \dots, n_\sigma; \quad h_1, h_2, \dots, h_\sigma; \quad k_1, k_2, \dots, k_\sigma$$

mit

$$h_j \geq 0, \quad 0 < k_j \leq \sqrt{x}, \quad (h_j, k_j) = 1 \quad (j = 1, 2, \dots, \sigma)$$

die Menge derjenigen t unseres Integrationsintervalls, welche im Durchschnitt der σ Intervalle $\frac{2\pi}{\beta_j} \mathfrak{B}_{h_j, k_j}$ ($j = 1, 2, \dots, \sigma$) liegen und die Ungleichungen

$$(21) \quad \frac{1}{2^{n_j+1} k_j \sqrt{x}} < \left| \frac{\beta_j}{2\pi} t - \frac{h_j}{k_j} \right| \leq \frac{1}{2^{n_j} k_j \sqrt{x}} \quad (j = 1, 2, \dots, \sigma)$$

erfüllen. Diese Menge bezeichnen wir mit $Q(h; k; n)$. Weil in $\frac{2\pi}{\beta_j} \mathfrak{B}_{h_j, k_j}$ gilt

$$\left| \frac{\beta_j}{2\pi} t - \frac{h_j}{k_j} \right| \leq \frac{1}{k_j \sqrt{x}},$$

so sind diejenigen Mengen $Q(h; k; n)$ leer, für welche mindestens eine der Zahlen n_j negativ ist; ebenso sind diejenigen Mengen $Q(h; k; n)$ leer, für welche mindestens eine der Zahlen h_j gleich Null ist. Weiter kommen nur die h_j mit $h_j < \bar{c}_2 x^{r+\frac{1}{2}}$ in Betracht¹⁷⁾. Endlich ist jeder Punkt t unseres Integrationsintervalls in genau einer der abzählbar vielen Mengen $Q(h; k; n)$ enthalten, ausgenommen die endlich vielen Punkte $t = \frac{2\pi}{\beta_j} \frac{h_j}{k_j}$.

Es genügt also zu zeigen: für fast alle Punkte $(\beta_1, \beta_2, \dots, \beta_\sigma)$ von \overline{W} gilt

$$\sum_{(h), (k), (n)} \int_{Q(h; k; n)} \left| \frac{\theta^{r_1}(\beta_1 s) \dots \theta^{r_\sigma}(\beta_\sigma s)}{s} \right| dt = O\left(x^{\frac{r}{2} - \sigma + \varepsilon}\right)$$

für jedes $\varepsilon > 0$. Dabei können wir uns nach dem eben Gesagten auf die $Q(h; k; n)$ mit $h_j > 0$, $n_j \geq 0$, $h_j < \bar{c}_2 x^{r+\frac{1}{2}}$ beschränken. Aus Symmetriegründen dürfen und wollen wir uns sogar auf die $Q(h; k; n)$ mit

$$(22) \quad 2^{n_1} k_1 \geq 2^{n_2} k_2 \geq \dots \geq 2^{n_\sigma} k_\sigma$$

beschränken.

¹⁷⁾ \bar{c}_2 bedeuten positive, nur von $r_j, \sigma, \bar{C}, \bar{D}$ abhängige Zahlen.

Es sei nun $l; m_1, m_2, \dots, m_\sigma; n_1, n_2, \dots, n_\sigma$ ein System von ganzen nichtnegativen Zahlen; wir wollen sagen, daß eine Menge $Q(h; k; n)$ mit (22) zur Klasse $[l; m; n]$ gehört, wenn

$$(23) \quad 2^l \leq h_1 < 2^{l+1}, \quad 2^{m_j} \leq k_j < 2^{m_j+1} \quad (j = 1, 2, \dots, \sigma).$$

Jede von unseren Mengen $Q(h; k; n)$ gehört also genau einer Klasse an. Wenn eine Menge $Q(h; k; n)$ der Klasse $[l; m; n]$ nicht leer sein soll, so muß nach (21) (wegen $2^{n_1} k_1 \geq 2^{n_j} k_j$ für $j = 2, 3, \dots, \sigma$) gelten

$$\left| \frac{1}{\beta_1} \frac{h_1}{k_1} - \frac{1}{\beta_j} \frac{h_j}{k_j} \right| < \frac{\bar{c}_3}{2^{n_j} k_j 2^e} \quad (j = 2, 3, \dots, \sigma),$$

wo die ganze Zahl e durch $2^e \leq \sqrt{x} < 2^{e+1}$ definiert ist. Wenn wir nun den Hilfssatz 5 mit $\gamma_j = \frac{1}{\beta_j}$, $C = \frac{1}{D}$, $D = \frac{1}{C}$, $a = \bar{c}_3$ anwenden, finden wir:

Es gibt im Würfel \bar{W} ($\bar{C} \leq \beta_j \leq \bar{D}$) eine Punktmenge $\bar{N}(\bar{C}, \bar{D})$ von folgender Beschaffenheit:

1. Wenn der Punkt $(\beta_1, \beta_2, \dots, \beta_\sigma)$ die Menge $\bar{N}(\bar{C}, \bar{D})$ durchläuft, so durchläuft der Punkt $\left(\frac{1}{\beta_1}, \frac{1}{\beta_2}, \dots, \frac{1}{\beta_\sigma}\right)$ eine Menge vom Maß Null.

2. Zu jedem Punkt $(\beta_1, \beta_2, \dots, \beta_\sigma)$ der Menge $\bar{W} - \bar{N}(\bar{C}, \bar{D})$ gibt es eine Zahl

$$\varrho_0 = \varrho_0(\beta_1, \beta_2, \dots, \beta_\sigma; \bar{C}, \bar{D}, \sigma, r_1, r_2, \dots, r_\sigma),$$

so daß für alle $\varrho \geq \varrho_0$ (also für alle $x \geq 2^{(\varrho_0+1)}$) und für jedes System von ganzen nichtnegativen Zahlen

$$l; m_1, m_2, \dots, m_\sigma; n_1, n_2, \dots, n_\sigma$$

höchstens

$$\bar{M} = 2^{\sigma-1} \left(\log x (l+1) \prod_{j=1}^{\sigma} (m_j+1) \prod_{j=2}^{\sigma} (n_j+1) \right)^2 \frac{2^{l+m_1+m_2+\dots+m_\sigma}}{2^{n_2+\dots+n_\sigma} x^{\frac{\sigma-1}{2}}}$$

Mengen $Q(h; k; n)$ der Klasse $[l; m; n]$ nicht leer sind.

Aus 1. folgt freilich, daß auch

3. die Menge $\bar{N}(\bar{C}, \bar{D})$ das Maß Null hat.

Es sei nun $(\beta_1, \beta_2, \dots, \beta_\sigma)$ im Rest des Beweises ein fester Punkt von $\bar{W} - \bar{N}(\bar{C}, \bar{D})$; und es sei $x > \text{Max}(4, 2^{2\varrho_0+1})$. Wir vereinigen die Klassen $[l; m; n]$ in Oberklassen $\{0\}, \{1\}, \dots, \{\sigma\}$ folgendermaßen: wenn genau τ von den σ Zahlen $2^{m_j+n_j}$ ($j = 1, 2, \dots, \sigma$) größer oder gleich \sqrt{x} sind, so werde die Klasse $[l; m; n]$ zur Oberklasse $\{\tau\}$ gezählt. Wegen (22), (23) ist offenbar

$$2^{n_1+m_1} \geq 2^{n_2+m_2} \geq \dots \geq 2^{n_\sigma+m_\sigma}.$$

Es genügt uns also zu zeigen: für unseren Punkt $(\beta_1, \beta_2, \dots, \beta_\sigma)$, für jedes $\varepsilon > 0$ und für jedes τ ($\tau = 0, 1, 2, \dots, \sigma$) ist

$$(24) \quad \sum_{Q(h; k; n)} \int \left| \frac{\theta^{r_1}(\beta_1 s) \dots \theta^{r_\sigma}(\beta_\sigma s)}{s} \right| dt = O\left(x^{\frac{r}{2} - \sigma + \varepsilon}\right),$$

wenn $Q(h; k; n)$ alle nicht leeren Mengen der Oberklasse $\{\tau\}$ durchläuft. Für ein $Q(h; k; n)$ der Klasse $[l; m; n]$, die zur Oberklasse $\{\tau\}$ gehört, ist nach (14) der Integrand kleiner als¹⁸⁾

$$c_3 \frac{2^{m_1}}{2^l} \frac{x^{\frac{r_1}{2} + \dots + \frac{r_\tau}{2}} \cdot x^{\frac{r_{\tau+1}}{4} + \dots + \frac{r_\sigma}{4}} \cdot 2^{n_{\tau+1} \frac{r_{\tau+1}}{2} + \dots + n_\sigma \frac{r_\sigma}{2}}}{2^{m_1 \frac{r_1}{2} + \dots + m_\tau \frac{r_\tau}{2}}};$$

das Maß von $Q(h; k; n)$ ist wegen (21) kleiner als

$$\frac{c_3}{2^{m_1 + n_1} \sqrt{x}};$$

endlich ist die Anzahl der nicht leeren $Q(h; k; n)$ der Klasse $[l; m; n]$ höchstens $\overline{\mathfrak{M}}$.

Also ist der Beitrag aller $Q(h; k; n)$ der Klasse $[l; m; n]$ zum Integral auf der linken Seite von (24) höchstens gleich

$$(25) \quad \left\{ \begin{aligned} & c_3 x^{\frac{r_1}{2} + \dots + \frac{r_\tau}{2} + \frac{r_{\tau+1}}{4} + \dots + \frac{r_\sigma}{4} - \frac{\sigma}{2}} \prod_{j=1}^{\tau} \frac{1}{2^{n_j + m_j} \left(\frac{r_j}{2} - 1\right)} \cdot \prod_{j=\tau+1}^{\sigma} 2^{n_j \left(\frac{r_j}{2} - 1\right) + m_j} \\ & \times \log^2 x \left((l+1) \prod_{j=1}^{\sigma} (m_j + 1) \prod_{j=2}^{\sigma} (n_j + 1) \right)^2. \end{aligned} \right.$$

Diesen Ausdruck sollen wir nun über die nichtnegativen ganzen Zahlen

$$l; m_1, m_2, \dots, m_\sigma; n_1, n_2, \dots, n_\sigma$$

summieren, wobei zu beachten ist

$$\begin{aligned} 2^{m_j + n_j} &\geq \sqrt{x} \quad \text{für } j \leq \tau; \quad 2^{m_j + n_j} < \sqrt{x} \quad \text{für } j > \tau; \\ 2^{m_j} &\leq \sqrt{x} \quad \text{für } j = 1, 2, \dots, \sigma; \quad 2^l < c_3 x^{r + \frac{1}{2}}. \end{aligned}$$

Es sei nun $1 > \varepsilon > 0$, $\eta = \frac{\varepsilon}{\sigma + 2}$. Mit $c_3(\varepsilon)$ bezeichne ich unterschiedslos positive, nur von $\varepsilon, \beta_j, r_j, \sigma$ ($j = 1, 2, \dots, \sigma$) abhängige Zahlen. Die angekündigte Summation ergibt:

¹⁸⁾ Denn es ist $|s| > t \geq \frac{2\pi}{\beta_1} \left(\frac{h_1}{k_1} - \frac{1}{k_1 \sqrt{x}} \right) > c_3 \frac{h_1}{k_1}$; mit c_3 bezeichne ich unterschiedslos positive, nur von β_j, r_j, σ ($j = 1, 2, \dots, \sigma$) abhängige Zahlen.

$$\sum_{2^l < c_3 x^{r+\frac{1}{2}}} (l+1)^2 < c_3 \log^3 x < c_3(\varepsilon) x^\eta;$$

$$\begin{aligned} \sum_{\substack{2^{m_j+n_j} \geq \sqrt{x} \\ 2^{m_j} \leq \sqrt{x}}} \frac{(m_j+1)^2 (n_j+1)^2}{2^{n_j+m_j} \binom{r_j}{\frac{r_j}{2}-1}} < c_3(\varepsilon) \log^2 x \sum_{2^{m_j} \leq \sqrt{x}} \frac{1}{2^{m_j} \binom{r_j}{\frac{r_j}{2}-1}} \sum_{2^{n_j} \geq \frac{\sqrt{x}}{2^{m_j}}} \frac{1}{2^{n_j(1-\eta)}} \\ < c_3(\varepsilon) \frac{x^\eta}{\sqrt{x}} \sum_{2^{m_j} \leq \sqrt{x}} \frac{1}{2^{m_j \binom{r_j}{\frac{r_j}{2}-2+\eta}}} < c_3(\varepsilon) \frac{x^\eta}{\sqrt{x}}; \end{aligned}$$

$$\begin{aligned} \sum_{2^{m_j+n_j} < \sqrt{x}} (m_j+1)^2 (n_j+1)^2 2^{n_j \binom{r_j}{\frac{r_j}{2}-1}} 2^{m_j} < c_3 \log^4 x \cdot \sqrt{x} \sum_{2^{n_j} < \sqrt{x}} 2^{n_j \binom{r_j}{\frac{r_j}{2}-2}} \\ < c_3(\varepsilon) x^{\frac{r_j}{4} - \frac{1}{2} + \eta}. \end{aligned}$$

(Für $j = 1$ haben wir sogar den überflüssigen Faktor $(n_j + 1)^2$ hinzugefügt; man vergesse bei der Summation nicht die Bedingungen $r_j \geq 4$.)

Aus diesen drei Abschätzungen sieht man aber, daß die durchzuführende Summation des Ausdruckes (25) höchstens

$$c_3(\varepsilon) x^{\frac{r_1}{2} + \dots + \frac{r_\tau}{2} + \frac{r_{\tau+1}}{4} + \dots + \frac{r_\sigma}{4} - \frac{\sigma}{2}} x^{2\eta} \prod_{j=1}^{\tau} \left(\frac{x^\eta}{\sqrt{x}} \right) \prod_{j=\tau+1}^{\sigma} \left(x^{\frac{r_j}{4} - \frac{1}{2} + \eta} \right) = c_3(\varepsilon) x^{\frac{r}{2} - \sigma + \varepsilon}$$

ergibt, w. z. b. w.

§ 6.

Satz 3.

Hilfssatz 7. *Es sei r ganz, $r \geq 4$. Dann gibt es im r -dimensionalen Raum der Punkte $(\alpha_1, \alpha_2, \dots, \alpha_r)$ eine Punktmenge N vom Maß Null und von folgender Beschaffenheit: Für jeden Punkt $(\alpha_1, \alpha_2, \dots, \alpha_r)$ mit $\alpha_j > 0$ ($j = 1, 2, \dots, r$), der nicht zu N gehört, und jedes $\varepsilon > 0$ gilt*

$$\int_{\frac{A}{\sqrt{x}}}^{x^r} \left| \frac{\theta(\alpha_1 s) \theta(\alpha_2 s) \dots \theta(\alpha_r s)}{s} \right| dt = O\left(x^{\frac{r}{4} + \varepsilon}\right);$$

dabei ist $s = \frac{1}{x} + ti$, $A = \text{Max} \frac{2\pi}{\alpha_j}$.

Nach Hilfssatz 4 folgt daraus sofort

Satz 3. *Es sei r ganz, $r \geq 4$. Dann gibt es im r -dimensionalen Raum der Punkte $(\alpha_1, \alpha_2, \dots, \alpha_r)$ eine Punktmenge N vom Maß Null und von folgender Beschaffenheit:*

Es sei $(\alpha_1, \alpha_2, \dots, \alpha_r)$ ein Punkt mit $\alpha_j > 0$ ($j = 1, 2, \dots, r$), der nicht in N liegt; es sei $\varepsilon > 0$; endlich sei

$$Q(u) = \alpha_1 u_1^2 + \alpha_2 u_2^2 + \dots + \alpha_r u_r^2.$$

Dann ist

$$P_Q(x) = O\left(x^{\frac{r}{4} + \varepsilon}\right).$$

Beweis des Hilfssatzes 7. Wir wenden den Hilfssatz 6 mit $r_j = 4$, α_j statt β_j , r statt σ und $4r$ statt r an; N sei die dort erklärte Menge vom Maß Null, $(\alpha_1, \alpha_2, \dots, \alpha_r)$ ein Punkt mit $\alpha_j > 0$ ($j = 1, 2, \dots, r$), der nicht zu N gehört; dann ist nach dem Hilfssatz 6

$$\begin{aligned} \int_{\frac{A}{\sqrt{x}}}^{x^r} \left| \frac{\theta^4(\alpha_1 s) \dots \theta^4(\alpha_r s)}{s} \right| dt &\leq \int_{\frac{A}{\sqrt{x}}}^{x^{4r}} \left| \frac{\theta^4(\alpha_1 s) \dots \theta^4(\alpha_r s)}{s} \right| dt \\ &= O\left(x^{\frac{4r}{2} - r + \varepsilon}\right) = O(x^{r + \varepsilon}). \end{aligned}$$

Nach einer Hölderschen Ungleichung¹⁹⁾ ist aber für $a_v \geq 0$, $x_v \geq 0$, $m > 1$

$$\left(\sum_{v=1}^n a_v x_v\right)^m \leq \sum_{v=1}^n a_v x_v^m \left(\sum_{v=1}^n a_v\right)^{m-1}.$$

Durch Grenzübergang ergibt sich daraus sofort eine analoge Ungleichung für Integrale aus nichtnegativen Funktionen, aus welcher sich ergibt:

$$\begin{aligned} \left(\int_{\frac{A}{\sqrt{x}}}^{x^r} \left| \frac{\theta(\alpha_1 s) \dots \theta(\alpha_r s)}{s} \right| dt\right)^4 &\leq \int_{\frac{A}{\sqrt{x}}}^{x^r} \left| \frac{\theta^4(\alpha_1 s) \dots \theta^4(\alpha_r s)}{s} \right| dt \left(\int_{\frac{A}{\sqrt{x}}}^{x^r} \frac{dt}{s}\right)^3 \\ &= O(x^{r + \varepsilon}) \cdot O(x^\varepsilon) = O(x^{r + 4\varepsilon}), \end{aligned}$$

w. z. b. w.

§ 7.

Satz 4.

Satz 4. Es sei $\sigma \geq 2$, $r_j \geq 1$ ($j = 1, 2, \dots, \sigma$); $\sum_{j=1}^{\sigma} r_j = r$; σ, r_j ganz. Es sei

$$\begin{aligned} Q(u) = \beta_1 (u_{1,1}^2 + u_{2,1}^2 + \dots + u_{r,1}^2) + \beta_2 (u_{1,2}^2 + \dots + u_{r,2}^2) + \dots \\ + \beta_\sigma (u_{1,\sigma}^2 + \dots + u_{r,\sigma}^2). \end{aligned}$$

Dann gibt es eine Punktmenge P des σ -dimensionalen Raumes der Punkte $(\beta_1, \beta_2, \dots, \beta_\sigma)$ vom Maß Null und von folgender Beschaffenheit:

¹⁹⁾ O. Hölder, Über einen Mittelwertsatz, Göttinger Nachrichten 1889, S. 38–47; vgl. besonders S. 44.

Wenn $\beta_j > 0$ ($j = 1, 2, \dots, \sigma$) und der Punkt $(\beta_1, \beta_2, \dots, \beta_\sigma)$ nicht zu P gehört, so ist

$$P_Q(x) = \Omega\left(x^{\frac{r}{2}-\sigma} \log^{\frac{\sigma-1}{\sigma+1}} x\right).$$

Beweis. Herr Khintchine²⁰⁾ hat bewiesen: es gibt im Raum der Punkte $(\beta_1, \beta_2, \dots, \beta_\sigma)$ eine Punktmenge P vom Maß Null und von folgender Beschaffenheit:

Zu jedem Punkt $(\beta_1, \beta_2, \dots, \beta_\sigma)$, der nicht zu P gehört, gibt es eine Folge von Systemen von ganzen Zahlen

$$(26) \quad \left\{ \begin{array}{l} p_{2,n}, p_{3,n}, \dots, p_{\sigma,n}, q_n \quad (n = 1, 2, 3, \dots), \\ \text{mit } q_n \rightarrow +\infty, \quad \left| \frac{\beta_j}{\beta_1} - \frac{p_{j,n}}{q_n} \right| < \frac{1}{q_n^{1+\frac{1}{\sigma-1}} \log^{\frac{1}{\sigma-1}} q_n} \quad (j = 2, 3, \dots, \sigma). \end{array} \right.$$

Wir nehmen ein solches System $(\beta_1, \beta_2, \dots, \beta_\sigma)$ mit $\beta_j > 0$ ($j = 1, 2, \dots, \sigma$) und eine zugehörige Folge (26). Es sei

$$Q_n(u) = \beta_1 \left(u_{1,1}^2 + u_{2,1}^2 + \dots + u_{r,1}^2 + \frac{p_{2,n}}{q_n} (u_{1,2}^2 + \dots + u_{r,2}^2) + \dots + \frac{p_{\sigma,n}}{q_n} (u_{1,\sigma}^2 + \dots + u_{r,\sigma}^2) \right);$$

für ganzzahlige Werte von $u_{i,k}$ (d. h. in jedem Gitterpunkt) nimmt diese Form einen Wert von der Gestalt $\beta_1 \frac{m}{q_n}$ (m ganz) an. Es ist

$$\begin{aligned} |Q_n(u) - Q(u)| &\leq \frac{\beta_1}{q_n^{1+\frac{1}{\sigma-1}} \log^{\frac{1}{\sigma-1}} q_n} (u_{1,2}^2 + \dots + u_{r,\sigma}^2) \\ &\leq \frac{Q(u)}{q_n^{1+\frac{1}{\sigma-1}} \log^{\frac{1}{\sigma-1}} q_n} \quad \text{für } n > c_3. \end{aligned}$$

Es sei M_n ganz, $M_n \leq q_n^{1+\frac{1}{\sigma-1}} \log^{\frac{1}{\sigma-1}} q_n < M_n + 1$. Wenn

$$\beta_1 \frac{M_n + \frac{1}{3}}{q_n} \leq Q(u) \leq \beta_1 \frac{M_n + \frac{2}{3}}{q_n},$$

so ist

$$\beta_1 \frac{M_n + \frac{1}{3}}{q_n} \left(1 - \frac{1}{q_n^{1+\frac{1}{\sigma-1}} \log^{\frac{1}{\sigma-1}} q_n} \right) \leq Q_n(u) \leq \beta_1 \frac{M_n + \frac{2}{3}}{q_n} \left(1 + \frac{1}{q_n^{1+\frac{1}{\sigma-1}} \log^{\frac{1}{\sigma-1}} q_n} \right),$$

²⁰⁾ A. Khintchine, Zur metrischen Theorie der diophantischen Approximationen, Math. Zeitschr. 24 (1926), S. 706–714. Sein Resultat ist noch schärfer und in symmetrischer Form dargestellt.

also (wenn sich das Zeichen o auf wachsendes n bezieht)

$$\beta_1 \frac{M_n + \frac{1}{3} + o(1)}{q_n} \leq Q_n(u) \leq \beta_1 \frac{M_n + \frac{2}{3} + o(1)}{q_n},$$

also

$$\beta_1 \frac{M_n}{q_n} < Q_n(u) < \beta_1 \frac{M_n + 1}{q_n}$$

für $n > c_3$. In dem Gebiet

$$\beta_1 \frac{M_n + \frac{1}{3}}{q_n} \leq Q(u) \leq \beta_1 \frac{M_n + \frac{2}{3}}{q_n}$$

liegen also für $n > c_3$ keine Gitterpunkte; also ist

$$A_Q\left(\beta_1 \frac{M_n + \frac{1}{3}}{q_n}\right) = A_Q\left(\beta_1 \frac{M_n + \frac{2}{3}}{q_n}\right);$$

dagegen ist

$$\begin{aligned} J_Q\left(\beta_1 \frac{M_n + \frac{2}{3}}{q_n}\right) - J_Q\left(\beta_1 \frac{M_n + \frac{1}{3}}{q_n}\right) &= c_3 \frac{\left(M_n + \frac{2}{3}\right)^{\frac{r}{2}} - \left(M_n + \frac{1}{3}\right)^{\frac{r}{2}}}{q_n^{\frac{r}{2}}} > c_3 \frac{M_n^{\frac{r}{2}-1}}{q_n^{\frac{r}{2}}} \\ &> c_3 \left(\frac{M_n}{q_n}\right)^{\frac{r}{2}-\sigma} \left(\log \frac{M_n}{q_n}\right)^{\frac{\sigma-1}{\sigma+1}} \end{aligned}$$

für $n > c_3$. Für $n > c_3$ ist also mindestens eine der beiden Zahlen

$$\left| P_Q\left(\beta_1 \frac{M_n + \frac{1}{3}}{q_n}\right) \right|, \quad \left| P_Q\left(\beta_1 \frac{M_n + \frac{2}{3}}{q_n}\right) \right|$$

größer als

$$c_3 \left(\frac{M_n}{q_n}\right)^{\frac{r}{2}-\sigma} \left(\log \frac{M_n}{q_n}\right)^{\frac{\sigma-1}{\sigma+1}},$$

womit wegen $\frac{M_n}{q_n} \rightarrow +\infty$ der Satz bewiesen ist.

§ 8.

Satz 5.

Satz 5. Es sei $\sigma \geq 2$; $r_j \geq 1$ ($j = 1, 2, \dots, \sigma$); $\sum_{j=1}^{\sigma} r_j = r$; σ, r_j ganz; $\beta_1 > 0, \beta_2 > 0, \dots, \beta_{\sigma} > 0$. Es sei

$$\begin{aligned} Q(u) &= \beta_1(u_{1,1}^2 + u_{2,1}^2 + \dots + u_{r_1,1}^2) + \beta_2(u_{1,2}^2 + \dots + u_{r_2,2}^2) + \dots \\ &\quad + \beta_{\sigma}(u_{1,\sigma}^2 + \dots + u_{r_{\sigma},\sigma}^2). \end{aligned}$$

Behauptung:

$$P_Q(x) = \Omega\left(x^{\frac{\tau}{2}-\sigma}\right).$$

Beweis. Es gibt bekanntlich eine Folge von Systemen von ganzen Zahlen

$$p_{2,n}, p_{3,n}, \dots, p_{\sigma,n}, q_n \quad (n = 1, 2, \dots)$$

mit $q_n \rightarrow +\infty$, $\left| \frac{\beta_j}{\beta_1} - \frac{p_{j,n}}{q_n} \right| < \frac{1}{q_n^{1+\frac{1}{\sigma-1}}}$ ($j = 2, 3, \dots, \sigma$). Es sei

$$Q_n(u) = \beta_1 \left(u_{1,1}^2 + u_{2,1}^2 + \dots + u_{r,1}^2 + \frac{p_{2,n}}{q_n} (u_{1,2}^2 + \dots + u_{r,2}^2) + \dots + \frac{p_{\sigma,n}}{q_n} (u_{1,\sigma}^2 + \dots + u_{r,\sigma}^2) \right);$$

für jedes ganzzahlige System von $u_{i,k}$ hat der Wert von $Q_n(u)$ die Form

$\beta_1 \frac{m}{q_n}$ (m ganz). Es ist

$$|Q_n(u) - Q(u)| \leq \frac{\beta_1}{q_n^{1+\frac{1}{\sigma-1}}} (u_{1,2}^2 + \dots + u_{r,\sigma}^2) \leq \frac{dQ(u)}{q_n^{1+\frac{1}{\sigma-1}}},$$

wo $d = \text{Max}_{j=1,2,\dots,\sigma} \frac{\beta_j}{\beta_1}$ (d ist also ein c_3).

Nun sei \bar{M}_n ganz, $\bar{M}_n \leq \frac{1}{6d} q_n^{1+\frac{1}{\sigma-1}} < \bar{M}_n + 1$. Wenn

$$\beta_1 \frac{\bar{M}_n + \frac{1}{3}}{q_n} \leq Q(u) \leq \beta_1 \frac{\bar{M}_n + \frac{2}{3}}{q_n},$$

so ist

$$\beta_1 \frac{\bar{M}_n + \frac{1}{3}}{q_n} \left(1 - \frac{d}{q_n^{1+\frac{1}{\sigma-1}}} \right) < Q_n(u) < \beta_1 \frac{\bar{M}_n + \frac{2}{3}}{q_n} \left(1 + \frac{d}{q_n^{1+\frac{1}{\sigma-1}}} \right),$$

also für $n > c_3$

$$\beta_1 \frac{\bar{M}_n}{q_n} < Q_n(u) < \beta_1 \frac{\bar{M}_n + 1}{q_n};$$

im Gebiet

$$\beta_1 \frac{\bar{M}_n + \frac{1}{3}}{q_n} \leq Q(u) \leq \beta_1 \frac{\bar{M}_n + \frac{2}{3}}{q_n}$$

liegen also für $n > c_3$ keine Gitterpunkte; also

$$A_Q\left(\beta_1 \frac{\bar{M}_n + \frac{1}{3}}{q_n}\right) = A_Q\left(\beta_1 \frac{\bar{M}_n + \frac{2}{3}}{q_n}\right);$$

andererseits

$$J_Q\left(\beta_1 \frac{\bar{M}_n + \frac{2}{3}}{q_n}\right) - J_Q\left(\beta_1 \frac{\bar{M}_n + \frac{1}{3}}{q_n}\right) > c_3 \frac{\bar{M}_n^{\frac{r}{2}-1}}{q_n^{\frac{r}{2}}} > c_3 \left(\frac{\bar{M}_n}{q_n}\right)^{\frac{r}{2}-\sigma} \quad \text{für } n > c_3.$$

Für $n > c_3$ ist also mindestens eine der beiden Zahlen

$$P_Q\left(\beta_1 \frac{\bar{M}_n + \frac{1}{3}}{q_n}\right), \quad \left| P_Q\left(\beta_1 \frac{\bar{M}_n + \frac{2}{3}}{q_n}\right) \right|$$

größer als $c_3 \left(\frac{\bar{M}_n}{q_n}\right)^{\frac{r}{2}-\sigma}$, womit wegen $\frac{\bar{M}_n}{q_n} \rightarrow \frac{1}{2} \infty$ der Satz bewiesen ist.

Göttingen, den 21. November 1927.

(Eingegangen am 22. 11. 1927.)

Mendelismus.

Von

J. v. Behr in Köln.

§ 1.

Einleitung.

Unsere Problemstellung können wir kurz folgendermaßen skizzieren. Wir betrachten eine Generation von Individuen irgendwelcher Art, mögen sie nun dem Menschen-, Tier- oder Pflanzenreich angehören. Es sei möglich, mit Rücksicht auf n Erbeigenschaften innerhalb dieser Generation eine bestimmte Klasseneinteilung zu unterscheiden, die durch statistische Erhebung zahlenmäßig bekannt sei. Wir fragen jetzt, unter welchen Voraussetzungen wir etwas über die wahrscheinliche Klassenverteilung in der m -ten, auf die Ausgangsgeneration folgenden Generation aussagen können, und wie bei Annahme mendelnder Erbfaktoren, mögen sie nun voneinander unabhängig oder miteinander gekoppelt sein, diese wahrscheinliche Verteilung aussehen wird. Bei dieser Allgemeinheit der Fragestellung geht die vorliegende Arbeit in mehr als einem Punkte über die unten genannten Abhandlungen hinaus. Weinberg¹⁾ und Tietze²⁾ behandeln nur Mendelfaktoren, die voneinander unabhängig wirken. Das Problem der Kopplung ist zum ersten Male in einer 1924 erschienenen Arbeit des Verfassers³⁾ und zwar in allgemeiner Form untersucht worden. In allen eben genannten Arbeiten finden sich Rekursionsformeln, die die Verteilungszahlen einer Generation durch die der vorangehenden ausdrücken. Es fehlt aber die allgemeine Formel, die die Verteilungszahlen der m -ten Generation durch die der ersten ausdrückt, und die zu finden ein Hauptziel dieser Arbeit

¹⁾ Weinberg, Archiv für Rassen- und Gesellschaftsbiologie 7 (1910).

²⁾ Tietze, Zeitschrift für angewandte Mathematik und Mechanik. Oktober 1923.

³⁾ v. Behr, Zur Theorie mendelistischer Massenerscheinungen, Diss. Göttingen 1924

ist. Im § 6 werden statistische Korrelationsmaße eingeführt, die geeignet sind, eine neue, besonders dem Mathematiker und Statistiker verständliche Definition des Mendelismus zu geben. Es wird nämlich bewiesen, daß diese Korrelationsmaße — unter gewissen Voraussetzungen — beim Übergang von einer Generation zur andern in geometrischer Progression klein werden. Da diese Korrelationsmaße — abgesehen von besonderen Komplikationen — der statistischen Messung zugänglich sind, leuchtet ihre Bedeutung für den statistischen Beweis mendelistischer Vererbung ohne weiteres ein. Die in der vorliegenden Arbeit erzielten Resultate sind so allgemeiner Natur, daß die Resultate der früheren Arbeiten sich als einfache Spezialfälle ergeben. Als wichtigster Satz, der auf diese Art einen neuen Beweis erfährt, sei der folgende genannt, der zum ersten Male in der eben zitierten Arbeit des Verfassers bewiesen wurde: Bei mendelistischer Vererbung ist die Klassenverteilung der stationären Endgeneration, die sich aus einer gewissen Anfangsgeneration entwickelt, vom Koppelungsgrad der Erbfaktoren gänzlich unabhängig. Mit andern Worten: Wer Koppelungen zwischen Erbfaktoren statistisch nachweisen will, darf kein stationäres (durchgezüchtetes) Material benutzen; denn dieses liefert auf jeden Fall negative Resultate. Die zahllosen Untersuchungen, die bisher an stationärem Material vorgenommen wurden, erscheinen unter diesem Gesichtspunkte als wertlos.

Diese Arbeit bemüht sich, durch starke Schematisierung und möglichst formale Darstellung mit einem Minimum an biologischen Voraussetzungen auszukommen. Durch Einführung eines symbolischen Rechenmechanismus werden schwierige Rechnungen umgangen und übersichtliche Resultate erzielt.

§ 2.

Ein Erbfaktorenpaar.

Es sei möglich, innerhalb der Generation, die wir betrachten, zweifelsfrei drei Klassen A , B , C zu unterscheiden, denen bzw. α , β , γ Individuen beiderlei Geschlechts zugehören. Hierbei sollen die α , β , γ Relativzahlen sein, so daß

$$(1) \quad \alpha + \beta + \gamma = 1$$

ist. Wann können wir nun etwas über die wahrscheinliche Klassenverteilung der nächsten Generation aussagen? Wir können es jedenfalls, wenn wir etwas wissen 1. über die relative Häufigkeit, mit der geschlechtliche Kreuzungen der verschiedenen Typen ($A \times A$; $A \times B$; $A \times C$; $B \times B$ usw.) zu erwarten sind; wenn wir 2. Wahrscheinlichkeiten

$$(2) \quad W_{\mu, \nu}^{\varrho} \left. \begin{array}{l} \mu \\ \nu \\ \varrho \end{array} \right\} = \alpha, \beta, \gamma,$$

kennen, wo also z. B. $W_{\alpha, \beta}^{\gamma}$ die Wahrscheinlichkeit dafür bedeutet, daß die Kreuzung von A - männlich mit B - weiblich als Abkömmling C -Klasse ergibt. Wir müßten schließlich 3. etwas über die Fruchtbarkeit der einzelnen Elternpaartypen ($A \times A$, $A \times B$ usw.) wissen.

Wir beginnen mit dem letzten Punkt und setzen ad 3. voraus, daß die Fruchtbarkeit der verschiedenen Elternpaartypen gleich groß sei. Ad 1.: Über die relative Häufigkeit, mit der z. B. Ehen zwischen brünetten, dunkeläugigen Männern und blauäugigen Blondinen zu erwarten sind, läßt sich a priori nichts aussagen. Wir wollen daher unsere Betrachtungen auf solche Erbeigenschaften beschränken, die bei der Gattenwahl keine Rolle spielen.

(3) Dann ist eine Verteilung der Elternpaartypen zu erwarten, die einer rein zufälligen Kombination der männlichen mit den weiblichen Individuen entsprechen würde. Um die Häufigkeit der einzelnen Kombinationen zu berechnen, müßte man das Zahlenverhältnis der Geschlechter in den drei Klassen A , B , C kennen.

(4) Indem wir die Betrachtung auf solche Erbeigenschaften beschränken, die vom Geschlechte unabhängig sind, nehmen wir an, daß dieses Zahlenverhältnis in allen drei Klassen dasselbe sei. Auf Grund der bisher eingeführten Voraussetzungen können wir — lediglich mit Hilfe der elementaren Sätze über totale und zusammengesetzte Wahrscheinlichkeit — die folgenden Gleichungen hinschreiben:

$$(5) \quad \alpha' = \sum_{\mu, \nu} \mu \cdot \nu \cdot W_{\mu, \nu}^{\alpha}; \quad \beta' = \sum_{\mu, \nu} \mu \cdot \nu \cdot W_{\mu, \nu}^{\beta}; \quad \gamma' = \sum_{\mu, \nu} \mu \cdot \nu \cdot W_{\mu, \nu}^{\gamma}$$

$$(\mu = \alpha, \beta, \gamma; \nu = \alpha, \beta, \gamma).$$

Statt dessen kann man auch schreiben

$$(5a) \quad \varrho' = \sum_{\mu, \nu} \mu \cdot \nu \cdot W_{\mu, \nu}^{\varrho} \quad (\varrho = \alpha, \beta, \gamma).$$

Hierbei ist α' usw. die Wahrscheinlichkeit dafür, daß sich in der ersten Filialgeneration der Typus A usw. findet. Zur Erläuterung der Gleichungen (5) sei hinzugefügt: Das \sum -Zeichen auf der rechten Seite rührt daher, daß es sich um eine totale Wahrscheinlichkeit handelt. Der einzelne Term, z. B. $\alpha \cdot \beta \cdot W_{\alpha, \beta}^{\gamma}$ bedeutet die Wahrscheinlichkeit dafür, daß eine Kreuzung zwischen A — männlich und B — weiblich zustande kommt [s. (2) und (3)], und daß diese Kreuzung den Typ C hervorbringt.

In den bisherigen Voraussetzungen ist von Mendelismus noch nicht die Rede gewesen. Die Spezialisierung wird erst durch zahlenmäßige Angabe der Größen $W_{\mu,\nu}^e$ vollzogen. Um nun diese W aus den Mendelschen Erbgeln abzuleiten, führen wir vorübergehend folgende Bezeichnung ein:

$$(6) \quad \alpha = (I, I); \quad \beta = (II, II); \quad \gamma = (I, II) = (II, I).$$

Durch diese Bezeichnung soll die erbbiologische Struktur der Individuen, soweit sie für die Klasseneinteilung A, B, C in Betracht kommt, angedeutet werden. I und II ist das Erbfaktorenpaar, auf das sich die Überschrift dieses Paragraphen bezieht. Schon durch formale Verallgemeinerung kommt man auf den Gedanken, neben Gen-(Erbfaktoren-)paaren auch Gentripel, Genquadrupel, allgemein Genreihen I, II, III, IV, \dots zu betrachten. Man hätte dann die Typen $(I, I), (I, II), (I, III)$ usw.; (II, II) usw. zu unterscheiden. Aber auch in diesem Falle wäre das Individuum durch eine *Zweiheit* von Erbfaktoren definiert.

(7) Dies rührt daher, daß das eine Gen von dem männlichen und das andere von dem weiblichen Elter stammt. Sind p, q, r, s irgendwelche Gene aus der Reihe I, II, III, \dots usw., so stellt die folgende Gleichung ein Schema des Erbprozesses im Falle *einer* Genreihe dar (monohybrider Fall):

$$(8) \quad (p, q) \times (r, s) \equiv \left(\frac{p+q}{2}\right) \cdot \left(\frac{r+s}{2}\right) \equiv \frac{1}{4} \cdot (p, r) + \frac{1}{4} \cdot (p, s) + \frac{1}{4} \cdot (q, r) + \frac{1}{4} \cdot (q, s).$$

Das \equiv -Zeichen deutet an, daß es sich um eine symbolische Gleichung handelt.

(8a) Zu lesen ist die Gleichung wie folgt: Bei der geschlechtlichen Kreuzung der Typen (p, q) und (r, s) entstehen — vermöge einer rein zufälligen Kombination — mit der gleichen Wahrscheinlichkeit $\frac{1}{4}$ die Typen $(p, r), (p, s), (q, r), (q, s)$. Schon auf Grund des bisher Gesagten [s. (7) u. (8)] können wir den Fundamentalsatz des Mendelismus verstehen, nach dem der Vererbungsprozeß nichts ist als Kombination (d. h. also Mischung, aber nicht *Vermischung!*) invarianter Erbfaktoren. Wenden wir (8) auf das in diesem Paragraphen zu behandelnde Genpaar I und II an, so erhalten wir nach (6) z. B.

$$(9) \quad \begin{aligned} \gamma \times \gamma &\equiv (I, II) \times (I, II) \\ &\equiv \frac{1}{4}(I, I) + \frac{1}{2}(I, II) + \frac{1}{4}(II, II) \equiv \frac{1}{4} \cdot \alpha + \frac{1}{2} \cdot \gamma + \frac{1}{4} \cdot \beta. \end{aligned}$$

(In dieser symbolischen Gleichung sind α, β, γ natürlich nicht Zahlwerte, sondern Typenbezeichnungen.)

Hieraus folgt nach (8a) und (2):

$$W_{\gamma,\gamma}^{\alpha} = \frac{1}{4}, \quad W_{\gamma,\gamma}^{\beta} = \frac{1}{4}, \quad W_{\gamma,\gamma}^{\gamma} = \frac{1}{2}.$$

In genau derselben Weise kann man die übrigen $W_{\mu,\nu}^{\rho}$ bestimmen, deren Werte in der folgenden Tabelle zusammengestellt sind.

	α	β	γ
α	1, 0, 0	0, 0, 1	$\frac{1}{2}, 0, \frac{1}{2}$
β	0, 0, 1	0, 1, 0	$0, \frac{1}{2}, \frac{1}{2}$
γ	$\frac{1}{2}, 0, \frac{1}{2}$	$0, \frac{1}{2}, \frac{1}{2}$	$\frac{1}{4}, \frac{1}{4}, \frac{1}{2}$

(10)

Gehen wir in dieser Tabelle z. B. in das dritte Kästchen der oberen Reihe, so können wir aus ihm die Werte $W_{\alpha,\gamma}^{\rho}$ ($\rho = \alpha, \beta, \gamma$) entnehmen, und zwar ist $W_{\alpha,\gamma}^{\alpha} = \frac{1}{2}$, $W_{\alpha,\gamma}^{\beta} = 0$ usw.

Setzt man die Werte von (10) in (5) ein, so folgt durch einfache Ausrechnung

$$(11) \quad \alpha' = \left(\alpha + \frac{\gamma}{2}\right)^2, \quad \beta' = \left(\beta + \frac{\gamma}{2}\right)^2, \quad \gamma' = 2\left(\alpha + \frac{\gamma}{2}\right)\left(\beta + \frac{\gamma}{2}\right).$$

Man überzeugt sich noch durch Ausrechnung von der formalen Gültigkeit der Identitäten:

$$(12) \quad \begin{cases} \sum_{\mu,\nu} \mu \cdot \bar{\nu} \cdot W_{\mu,\nu}^{\alpha} = \left(\alpha + \frac{\gamma}{2}\right)\left(\bar{\alpha} + \frac{\bar{\gamma}}{2}\right); & \sum_{\mu,\nu} \mu \cdot \bar{\nu} \cdot W_{\mu,\nu}^{\beta} = \left(\beta + \frac{\gamma}{2}\right)\left(\bar{\beta} + \frac{\bar{\gamma}}{2}\right); \\ \sum_{\mu,\nu} \mu \cdot \bar{\nu} \cdot W_{\mu,\nu}^{\gamma} = \left(\alpha + \frac{\gamma}{2}\right) \cdot \left(\bar{\beta} + \frac{\bar{\gamma}}{2}\right) + \left(\beta + \frac{\gamma}{2}\right) \cdot \left(\bar{\alpha} + \frac{\bar{\gamma}}{2}\right) & (\mu, \nu = \alpha, \beta, \gamma). \end{cases}$$

Hierbei brauchen wir, da es sich um Identitäten handelt, über die überstrichenen Größen einstweilen nicht mehr auszusagen, als die Gleichungen (12) selbst besagen. Die Gleichungen (12) sind die Grundlagen des Folgenden.

§ 3.

Mehrere Erbfaktorenpaare.

Nehmen wir an, daß die eine Erbeigenschaft, von der bisher die Rede war, die Haarfarbe sei mit den Klasseneinteilungen: schwarz, braun, blond. (Hierbei spielt es keine Rolle, ob das Beispiel biologisch einwandfrei ist.) Sei die Augenfarbe mit den Klassen: dunkel, mittel, hell eine zweite Eigenschaft, die sich unabhängig von der ersten vererben möge, dann wird jede der ersten drei Klassen mit Rücksicht auf die zweite Eigenschaft in drei Unterklassen zerfallen. Es ergeben sich so neun Klassen, deren Häufigkeiten bezeichnet werden mit

$$\alpha_1 \alpha_2, \alpha_1 \beta_2, \alpha_1 \gamma_2, \beta_1 \alpha_2, \beta_1 \beta_2, \beta_1 \gamma_2, \gamma_1 \alpha_2 \text{ usw.}$$

Hierbei ist $\alpha_1 \alpha_2$ usw. nicht etwa als Produkt von zwei Zahlen, sondern als *eine* Zahl aufzufassen. In unserem Beispiel wäre z. B. $\alpha_1 \gamma_2$ die relative Häufigkeit der Klasse „schwarz—hell“, und $\beta_1 \beta_2$ würde das entsprechende für die Klasse „braun—mittel“ bedeuten. Im Falle von n Erbeigenschaften hätte man die 3^n Verteilungszahlen $\alpha_1 \alpha_2 \dots \alpha_{n-1} \alpha_n$, $\alpha_1 \alpha_2 \dots \alpha_{n-1} \beta_n$, ... usf. bis $\gamma_1 \gamma_2 \dots \gamma_{n-1} \gamma_n$. Den allgemeinen Typus dieser Verteilungszahlen bezeichnen wir mit $\mu_1 \mu_2 \mu_3 \dots \mu_{n-1} \mu_n$, wo also die $\mu_i = \alpha_i, \beta_i, \gamma_i$ ($i = 1, 2, \dots, n$) sein können. In demselben Sinne wie μ werden wir auch die Buchstaben ν und ϱ gebrauchen. Entsprechend (1) ergibt sich:

$$(13) \quad \sum_{\mu_i} \mu_1 \mu_2 \dots \mu_{n-1} \mu_n = 1 \quad (i = 1, \dots, n)$$

und entsprechend (5a)

$$(14) \quad \varrho'_1 \varrho'_2 \dots \varrho'_{n-1} \varrho'_n = \sum_{\mu_i, \nu_i} \mu_1 \mu_2 \dots \mu_{n-1} \mu_n \cdot \nu_1 \nu_2 \dots \nu_{n-1} \nu_n \cdot W_{\mu_1 \mu_2 \dots \mu_n, \nu_1 \nu_2 \dots \nu_n}^{\varrho_1 \varrho_2 \dots \varrho_{n-1} \varrho_n}$$

Unserer Voraussetzung, daß die n Erbeigenschaften sich unabhängig voneinander vererben, tragen wir durch den Ansatz Rechnung:

$$(15) \quad W_{\mu_1 \mu_2 \dots \mu_{n-1} \mu_n, \nu_1 \nu_2 \dots \nu_{n-1} \nu_n}^{\varrho_1 \varrho_2 \dots \varrho_{n-1} \varrho_n} = W_{\mu_1, \nu_1}^{\varrho_1} \cdot W_{\mu_2, \nu_2}^{\varrho_2} \cdot \dots \cdot W_{\mu_n, \nu_n}^{\varrho_n}$$

wobei die $W_{\mu_i, \nu_i}^{\varrho_i}$ von i unabhängig sein sollen. Also

$$(16) \quad W_{\mu_i, \nu_i}^{\varrho_i} = W_{\mu, \nu}^{\varrho} \quad (i = 1, 2, \dots, n).$$

Die $W_{\mu, \nu}^{\varrho}$ sind die nach (10) bekannten Mendelzahlen.

Jetzt zerlegen wir in (14) $\mu_1 \mu_2 \dots \mu_{n-1} \mu_n$ in das *symbolische* Produkt $\mu_1 \cdot \mu_2 \cdot \mu_3 \cdot \dots \cdot \mu_{n-1} \cdot \mu_n$; desgleichen setzen wir $\nu_1 \nu_2 \dots \nu_{n-1} \nu_n$ gleich $\nu_1 \cdot \nu_2 \cdot \nu_3 \cdot \dots \cdot \nu_{n-1} \cdot \nu_n$. Bevor wir aber in (14) irgend etwas an der Reihenfolge der symbolischen Faktoren ändern, versehen wir die ν_i mit einem Querbalken. Dann geht (14) über in:

$$(17) \quad \varrho'_1 \varrho'_2 \dots \varrho'_{n-1} \varrho'_n = \sum_{\mu_i, \nu_i} \mu_1 \cdot \bar{\nu}_1 \cdot W_{\mu, \nu}^{\varrho_1} \cdot \mu_2 \cdot \bar{\nu}_2 \cdot W_{\mu, \nu}^{\varrho_2} \cdot \dots \cdot \mu_n \cdot \bar{\nu}_n \cdot W_{\mu, \nu}^{\varrho_n}$$

$$(17a) \quad = \sum_{\mu_i, \nu_i} \prod_k^{1, n} \mu_k \cdot \bar{\nu}_k \cdot W_{\mu_k, \nu_k}^{\varrho_k}$$

$$(18) \quad = \varrho'_1 \varrho'_2 \dots \varrho'_{n-1} \varrho'_n = \prod_k \sum_{\mu_k, \nu_k} \mu_k \cdot \bar{\nu}_k \cdot W_{\mu_k, \nu_k}^{\varrho_k}$$

Zerlegt man auch die linke Seite von (18) in symbolische Faktoren, so bleibt diese Gleichung richtig, wenn man setzt:

$$(!19) \quad \varrho'_k = \sum_{\mu_k, \nu_k} \mu_k \cdot \bar{\nu}_k \cdot W_{\mu_k, \nu_k}^{\varrho_k} \quad (k = 1, 2, \dots, n).$$

Das !-Zeichen erinnert daran, daß es sich nur um eine symbolische Gleichung handelt.

Führt man die eben vollzogenen Rechenoperationen rückwärts durch, so gelangt man in eindeutiger Weise von (19) zu (17). Setzt man jetzt die überstrichenen Symbolfaktoren und die nicht überstrichenen je für sich zu wirklichen Zahlen zusammen (die $W_{\mu_k, \nu_k}^{\rho_k}$ sind nach (16) wirkliche Zahlen), so geht (17) in eindeutiger Weise in (14) über. Damit ist (19) als eine erlaubte Schreibweise an Stelle von (14) nachgewiesen. Über das Rechnen mit den Symbolen genügen die folgenden Bemerkungen. Wir werden nur die Operationen der Addition und Multiplikation anwenden.

(20) Die Aufgabe des Querbalkens ist es, eine eindeutige Zusammensetzung der Symbole zu wirklichen Zahlen zu gewährleisten, indem die überstrichenen Symbolfaktoren und die nicht überstrichenen je für sich zu wirklichen Zahlen zusammengefaßt werden. Diese Aufgabe kann er natürlich nur erfüllen, wenn höchstens je *ein* vollständiges Produkt überstrichener Symbole und nicht überstrichener zusammentreffen. Beispiel: Es sei $n = 4$. Dann ist $\bar{\alpha}_2 \gamma_3 \gamma_1 \bar{\beta}_3 \alpha_2 \bar{\beta}_1 \bar{\gamma}_4 \gamma_4$ zu lesen $\gamma_1 \alpha_2 \gamma_3 \gamma_4 \cdot \bar{\beta}_1 \bar{\alpha}_2 \bar{\beta}_3 \bar{\gamma}_4 = \gamma_1 \alpha_2 \gamma_3 \gamma_4 \cdot \beta_1 \alpha_2 \beta_3 \gamma_4$;

(20 a) d. h. wenn der Querbalken seine Aufgabe erfüllt hat, kann er weggelassen werden. Dagegen läßt sich $\alpha_2 \gamma_3 \gamma_1 \beta_3 \alpha_2 \bar{\beta}_1 \bar{\gamma}_4 \gamma_4$ nicht eindeutig zu wirklichen Zahlen zusammensetzen. Der Beweis für das Rechnen mit den Symbolen, den wir im Einzelfalle nachher nicht mehr führen werden, beruht immer auf der — oben erwähnten — Identität der Gleichungen (19) und (14). Bildet man irgendeine *lineare* Funktion der $\rho'_1 \rho'_2 \dots \rho'_n$ einmal nach (19) und dann nach (14), so werden die beiden Resultate, die man so erhält, gewiß in den einzelnen Symbolen identisch sein. Da aber der Querbalken, weil die Funktionen der $\rho'_1 \dots \rho'_n$ *linear* sind, in eindeutiger Weise die Zusammensetzung der Symbole zu wirklichen Zahlen festlegt, müssen die beiden Resultate auch in den wirklichen Zahlen identisch sein. Damit sind die formalen Rechnungen, die im folgenden ausgeführt werden, gerechtfertigt.

Aus (19) folgt wegen (12) [siehe auch (16)]

$$(21) \quad \alpha'_k = \varepsilon_k \cdot \bar{\varepsilon}_k; \quad \beta'_k = \delta_k \cdot \bar{\delta}_k; \quad \gamma'_k = \varepsilon_k \cdot \bar{\delta}_k + \delta_k \cdot \bar{\varepsilon}_k,$$

wo gesetzt ist:

$$(22) \quad \varepsilon_k = \alpha_k + \frac{\gamma_k}{2}, \quad \delta_k = \beta_k + \frac{\gamma_k}{2} \quad (k = 1, 2, \dots, n).$$

Nach (21) wird z. B.

$$(23) \quad \alpha'_1 \alpha'_2 \dots \alpha'_n = \varepsilon_1 \varepsilon_2 \dots \varepsilon_n \cdot \bar{\varepsilon}_1 \bar{\varepsilon}_2 \dots \bar{\varepsilon}_n = (\varepsilon_1 \varepsilon_2 \dots \varepsilon_n)^2 \quad [\text{nach (20 a)}].$$

Nach demselben Verfahren kann man die übrigen $\rho'_1 \rho'_2 \dots \rho'_n$ bilden. Rein mathematisch ist die Situation folgende: Gegeben sind 3^n Zahlen $\rho_1 \rho_2 \dots \rho_{n-1} \rho_n$ ($\rho_k = \alpha_k, \beta_k, \gamma_k$; $k = 1, 2, \dots, n$). Diesen 3^n Zahlen werden

durch die Symboltransformation (21) 3^n neue Zahlen $\varrho'_1 \varrho'_2 \dots \varrho'_{n-1} \varrho'_n$ zugeordnet, diesen wieder durch dieselbe Transformation $\varrho''_1 \varrho''_2 \dots \varrho''_{n-1} \varrho''_n$ usw. Gefragt ist: Wie drücken sich die $\varrho_1^m \varrho_2^m \dots \varrho_{n-1}^m \varrho_n^m$ durch die gegebenen $\varrho_1 \varrho_2 \dots \varrho_{n-1} \varrho_n$ aus? Gemäß (23) erhält man:

$$\alpha''_1 \alpha''_2 \dots \alpha''_{n-1} \alpha''_n = (\varepsilon'_1 \varepsilon'_2 \dots \varepsilon'_{n-1} \varepsilon'_n)^2.$$

Die Größen ε_k bildet man formal nach (22)

$$! \quad \varepsilon'_k = \alpha'_k + \frac{\gamma'_k}{2} = \varepsilon_k \cdot \bar{\varepsilon}_k + \frac{1}{2} (\varepsilon_k \cdot \bar{\delta}_k + \delta_k \cdot \bar{\varepsilon}_k) \quad [\text{nach (21)}],$$

$$(!24) \quad \varepsilon'_k = \frac{\varepsilon_k \cdot \bar{*}_k + *_k \cdot \bar{\varepsilon}_k}{2} \quad (k = 1, 2, \dots, n),$$

wo $*_k = \varepsilon_k + \delta_k$ gesetzt ist.

Entsprechend folgt

$$! \quad \delta'_k = \frac{\delta_k \cdot \bar{*}_k + *_k \cdot \bar{\delta}_k}{2}.$$

Wir führen jetzt folgende — 2^n Zahlen umfassende — Bezeichnung ein

$$\sigma_1 \sigma_2 \dots \sigma_{n-1} \sigma_n \quad (\sigma_k = \varepsilon_k \text{ oder } \delta_k; k = 1, 2, \dots, n).$$

Die $\sigma_1 \sigma_2 \dots \sigma_{n-1} \sigma_n$ sind vermöge (22) lineare Kombinationen der $\varrho_1 \varrho_2 \dots \varrho_{n-1} \varrho_n$, und daher, wenn diese gegeben sind, gleichfalls zahlenmäßig bekannt. Im Gegensatz zu den $\varrho_1 \varrho_2 \dots \varrho_n$, die wir als Individuenzahlen bezeichnen können, wollen wir die $\sigma_1 \sigma_2 \dots \sigma_n$ Gametenzahlen nennen. Die biologische Bedeutung dieser Gametenzahlen ist aus der formalen Darstellung, die wir gewählt haben, nicht ohne weiteres ersichtlich; doch werden hierüber, besonders über das Verhältnis von Individuen- und Gametenzahlen, wohl die Erörterungen des nächsten Paragraphen einige Klärung bringen.

§ 4.

Der mathematische Kern des Problems.

Rein mathematisch ist die Situation jetzt folgende. Gegeben sind 2^n Zahlen $\sigma_1 \sigma_2 \dots \sigma_{n-1} \sigma_n$. Diesen werden 2^n neue Zahlen $\sigma'_1 \sigma'_2 \dots \sigma'_{n-1} \sigma'_n$ durch die Symboltransformation (24) zugeordnet, die wir auch so schreiben können:

$$(!25) \quad \sigma'_i = \frac{\sigma_i \cdot \bar{*}_i + *_i \cdot \bar{\sigma}_i}{2} \quad (\sigma_i = \varepsilon_i \text{ oder } \delta_i; i = 1, 2, \dots, n),$$

$$(!25a) \quad *_i = \delta_i + \varepsilon_i.$$

Durch dieselbe Transformation werden den 2^n Zahlen $\sigma'_1 \sigma'_2 \dots \sigma'_{n-1} \sigma'_n$ 2^n neue Zahlen $\sigma''_1 \sigma''_2 \dots \sigma''_{n-1} \sigma''_n$ zugeordnet usw. Gefragt wird: Wie

drücken sich die Zahlen $\sigma_1^m \sigma_2^m \dots \sigma_n^m$ durch die Zahlen $\sigma_1 \sigma_2 \dots \sigma_n$ aus? Dies ist die Problemstellung, die wir in den Vordergrund schieben möchten, zumal nach (21) die $\varrho'_1 \varrho'_2 \dots \varrho'_n$ quadratische Funktionen der $\sigma_1 \sigma_2 \dots \sigma_n$ werden [siehe z. B. (23)] und daher ebenso die $\varrho_1^{m+1} \varrho_2^{m+1} \dots \varrho_n^{m+1}$ als quadratische Funktionen der $\sigma_1^m \sigma_2^m \dots \sigma_n^m$ berechenbar sind. Solange wir uns darauf beschränken, lineare Funktionen der $\sigma'_1 \sigma'_2 \dots \sigma'_n$ zu betrachten, gewährleistet der Querbalken eine eindeutige Zusammensetzung der Symbole zu wirklichen Zahlen. Dies hört jedoch auf, wenn wir zu den Zahlen $\sigma''_1 \sigma''_2 \dots \sigma''_n$ übergehen, die quadratische Funktionen der $\sigma'_1 \sigma'_2 \dots \sigma'_n$ sind. Wir wollen daher die Rechenregel des Querbalkens durch eine ihr adäquate ersetzen. Der Querbalken hat im wesentlichen nichts anderes getan, als daß er für Symbole mit *gleichem* Index i die Vertauschbarkeit der Faktorenfolge aufhob. Wir können daher, wenn wir diese Funktion des Querbalkens als Rechenregel beibehalten, in (25) den Querstrich einfach fortlassen und schreiben

$$(126) \quad \sigma'_i = \frac{\sigma_i \cdot *i + *i \cdot \sigma_i}{2} \quad (\sigma_i = \delta_i \text{ oder } \varepsilon_i; \quad i = 1, 2, \dots, n).$$

Wir wollen statt dessen nun die folgende Schreibweise benutzen, die einer rein rechnerischen Behandlung besser zugänglich ist.

$$(127) \quad \sigma'_i = \frac{(\sigma_i \cdot *i)_1 + (\sigma_i \cdot *i)_2}{2} \quad (\sigma_i = \delta_i \text{ oder } \varepsilon_i; \quad i = 1, 2, \dots, n).$$

Die für diese Schreibweise geltende Rechenregel heißt folgendermaßen:

(27 a) Unmittelbar bevor die Symbole zu wirklichen Zahlen zusammengesetzt werden, muß das σ_i in der runden Klammer an die Stelle treten, die der Index hinter der Klammer angibt. Dann werden die Symbole zu wirklichen Zahlen zusammengesetzt; jedoch darf unter Symbolen mit gleichem Index i die Reihenfolge der Faktoren hierbei nicht geändert werden. Diese Vorschrift ist der des Querbalkens völlig adäquat. Beispiel: $n = 2$.

$$\begin{aligned} (28) \quad \sigma'_1 \sigma'_2 &= \frac{1}{4} \cdot \{ (\sigma_1 \cdot *1)_1 \cdot (\sigma_2 \cdot *2)_1 + (\sigma_1 \cdot *1)_1 \cdot (\sigma_2 \cdot *2)_2 + (\sigma_1 \cdot *1)_2 \cdot (\sigma_2 \cdot *2)_1 \\ &\quad + (\sigma_1 \cdot *1)_2 \cdot (\sigma_2 \cdot *2)_2 \} \\ &= \sigma'_1 \sigma'_2 = \frac{1}{4} \cdot \{ \sigma_1 \sigma_2 \cdot *1 *2 + \sigma_1 *2 \cdot *1 \sigma_2 + *1 \sigma_2 \cdot \sigma_1 *2 + *1 *2 \cdot \sigma_1 \sigma_2 \} \\ &= \sigma'_1 \sigma'_2 = \frac{1}{2} \cdot \{ \sigma_1 \sigma_2 \cdot *1 *2 + \sigma_1 *2 \cdot *1 \sigma_2 \}. \end{aligned}$$

Diese Gleichung umfaßt natürlich vier Gleichungen, da $\sigma_1 = \delta_1$ oder ε_1 und $\sigma_2 = \delta_2$ oder ε_2 sein kann. Nach (25 a) ist:

$$\sigma_1 *2 = \sigma_1 \delta_2 + \sigma_1 \varepsilon_2, \quad *1 \sigma_2 = \delta_1 \sigma_2 + \varepsilon_1 \sigma_2, \quad *1 *2 = \delta_1 \delta_2 + \delta_1 \varepsilon_2 + \varepsilon_1 \delta_2 + \varepsilon_1 \varepsilon_2.$$

Entsprechend (28) erhalten wir im allgemeinen Falle aus (27)

$$(29) \sigma'_1 \sigma'_2 \dots \sigma'_{n-1} \sigma'_n = \frac{1}{2^n} \cdot \sum_{a_i} (\sigma_1 \cdot *_{1})_{a_1} \cdot (\sigma_2 \cdot *_{2})_{a_2} \cdot (\sigma_3 \cdot *_{3})_{a_3} \cdot \dots \cdot (\sigma_n \cdot *_{n})_{a_n}.$$

Hierbei sind die a_i ($i = 1, 2, \dots, n$) unabhängig voneinander der Werte 1 und 2 fähig. Es ergeben sich also 2^n Summanden. Die Gleichung (29), die keine symbolische Gleichung ist, läßt eine anschauliche Illustration zu. Wir werden im folgenden ein Problem auseinandersetzen, das als ein exaktes Schema für den durch (29) dargestellten Kreuzungs- und Vererbungsprozeß angesehen werden kann und dementsprechend durch (29) seine Lösung erfährt. Auf einem Tisch sei vor uns eine große Anzahl von Papierstreifen ausgebreitet. Auf jedem einzelnen Streifen mögen nebeneinander n Kästchen abgetragen sei, so daß also z. B. für $n = 5$ sich folgendes Bild ergibt:

--	--	--	--	--

. Die n Kästchen eines jeden Streifens werden auf der Rückseite des Papiers durch die mit roter Tinte in natürlicher Anordnung eingetragenen Zahlen $1, 2, \dots, n-1, n$ numeriert. Auf die Vorderseite der Streifen tragen wir in die Kästchen wahllos entweder den Buchstaben δ oder ε ein. (Den Index dürfen wir hier weglassen, da er bereits auf der Rückseite des Kästchens steht.) Bei Ausnutzung aller Möglichkeiten entstehen auf diese Weise 2^n Typen. Entsteht z. B. der Typ $\varepsilon \delta \delta \varepsilon \dots \varepsilon \varepsilon$, so bezeichne $\varepsilon_1 \delta_2 \delta_3 \varepsilon_4 \dots \varepsilon_{n-1} \varepsilon_n$ die relative Häufigkeit, mit der dieser Typ vorkommt. Allgemein werde die relative Häufigkeit der Typen mit $\sigma_1 \sigma_2 \dots \sigma_n$ bezeichnet ($\sigma_i = \varepsilon_i$ oder δ_i).

(30) Nun stellen wir uns einen zweiten Haufen von Papierstreifen her, der mit dem ersten völlig identisch sei, und beginnen, die Papierstreifen paarweise zuzuordnen, indem wir für jedes Paar aus jedem der beiden Haufen je einen Streifen entnehmen. Die also paarweise geordneten Streifen zerschneiden wir, so daß jeder Streifen in seine n Kästchen zerfällt.

(31) Aus den $2n$ Kästchen eines Paares werden in völlig willkürlicher Weise zwei neue Streifen gebildet, jedoch so, daß jeder neue Streifen auf der Rückseite in natürlicher Reihenfolge die mit roter Tinte geschriebenen Zahlen $1, 2, \dots, n$ enthält. Sämtliche neue Streifen werden darauf zu einem Haufen zusammengelegt. Wir fragen nach den relativen Häufigkeiten $\sigma'_1 \sigma'_2 \dots \sigma'_n$, mit denen — wahrscheinlich — *jetzt* die einzelnen Typen vorhanden sind. Auf diese Frage gibt die Gleichung (29) Antwort, was weiter unten in allgemeinerer Form bewiesen werden wird. Um Mißverständnissen vorzubeugen, sei erwähnt, daß dem Streifenpaar in der Biologie das Einzelindividuum entspricht. Das Einzelindividuum ist durch eine Zweierheit von Gameten (Streifen) charakterisiert, nämlich einen, der von dem männlichen, und einen, der von dem weiblichen Elter herrührt. In der bisherigen Pro-

blemstellung sind zwei Spezialisierungen enthalten, von denen wir uns der Reihe nach befreien werden. Die eine ist folgende:

(31a) Nach unserer Darstellung [s. (31)] erscheinen alle möglichen Paare neuer Streifen, die aus einem bestimmten Streifenpaar durch Zerstückelung entstehen können, als „gleichmögliche“ Fälle im Sinne der Wahrscheinlichkeitsrechnung. Diese Voraussetzung werden wir im nächsten Paragraphen fallen lassen und statt (29) die allgemeinere Transformation behandeln:

$$(32) \sigma'_1 \sigma'_2 \dots \sigma'_{n-1} \sigma'_n = \sum_{a_i} (\sigma_1 \cdot *_{1})_{a_1} \cdot (\sigma_2 \cdot *_{2})_{a_2} \dots (\sigma_n \cdot *_{n})_{a_n} \cdot \omega(a_1, a_2, \dots, a_n),$$

wobei die a_i ($i = 1, 2, \dots, n$) unabhängig voneinander der Werte 1 und 2 fähig sind. Das Σ -Zeichen auf der rechten Seite deutet an, daß es sich um eine totale Wahrscheinlichkeit handelt; d. h. es deutet alle möglichen Arten an, auf die ein bestimmter *neuer* Streifen aus alten Streifenpaaren durch Zerstückelung entstehen kann. Die Bedeutung des einzelnen Summanden macht man sich am besten an einem Beispiel klar. Setzen wir $n = 5$ und greifen den Summanden

$$(\sigma_1 \cdot *_{1})_2 \cdot (\sigma_2 \cdot *_{2})_1 \cdot (\sigma_3 \cdot *_{3})_2 \cdot (\sigma_4 \cdot *_{4})_2 \cdot (\sigma_5 \cdot *_{5})_1 \cdot \omega(2, 1, 2, 2, 1)$$

heraus.

$$\begin{aligned} & (\sigma_1 \cdot *_{1})_2 \cdot (\sigma_2 \cdot *_{2})_1 \cdot (\sigma_3 \cdot *_{3})_2 \cdot (\sigma_4 \cdot *_{4})_2 \cdot (\sigma_5 \cdot *_{5})_1 \\ &= *_{1} \cdot \sigma_1 \cdot \sigma_2 \cdot *_{2} \cdot *_{3} \cdot \sigma_3 \cdot *_{4} \cdot \sigma_4 \cdot \sigma_5 \cdot *_{5} = *_{1} \sigma_2 *_{3} *_{4} \sigma_5 \cdot \sigma_1 *_{2} \sigma_3 \sigma_4 *_{5} \end{aligned}$$

[s. (27a)] bedeutet die Wahrscheinlichkeit dafür, daß bei der Paarung der Streifen [s. (30)] aus dem *ersten* Haufen ein Streifen gezogen wird, der im zweiten und fünften Kästchen einen bestimmten vorgeschriebenen Buchstaben σ_2 bzw. σ_5 hat, in den übrigen drei Kästchen aber irgendwie mit ε oder δ besetzt ist, und daß gleichzeitig aus dem *zweiten* Haufen ein Streifen gezogen wird, der im ersten, dritten und vierten Kästchen einen vorgeschriebenen Buchstaben σ_1 bzw. σ_3, σ_4 hat, in den beiden übrigen aber irgendwie mit ε oder δ besetzt ist. $\omega(2, 1, 2, 2, 1)$ bedeutet dann die Wahrscheinlichkeit dafür, daß aus einem solchen Streifenpaar auch wirklich ein *neuer* Streifen entsteht, der im ersten, dritten und vierten Kästchen die eben genannten — aus dem zweiten Haufen stammenden — Buchstaben $\sigma_1, \sigma_3, \sigma_4$ und im zweiten und fünften Kästchen die aus dem ersten Haufen stammenden Buchstaben σ_2 und σ_5 hat.

(33) Nach diesem Beispiel ist es klar, daß $\omega(a_1, a_2, \dots, a_{n-1}, a_n)$ die Wahrscheinlichkeit dafür bedeutet, daß durch Zerstückelung eines vorgegebenen Streifenpaares [s. (30)] ein neuer Streifen entsteht, der im i -ten Kästchen mit einem aus dem a_i -ten Haufen stammenden Buchstaben besetzt ist. ($a_i = 1, 2; i = 1, 2, \dots, n$.)

Diese ω -Größen sollen für alle Streifenpaare der genannten Art dieselben sein; sie sind also von den σ -Größen unabhängig. Nach dem Satz von der totalen Wahrscheinlichkeit wird:

$$(34) \quad \sum_{a_i} \omega(a_1, a_2, \dots, a_n) = 1 \quad (a_i = 1, 2; i = 1, 2, \dots, n).$$

Diese Summe umfaßt 2^n Summanden. In dem Falle, in dem alle ω -Größen einander gleich sind [s. (31) und (31 a)], wird jeder Summand gleich $\frac{1}{2^n}$. Dann geht (32) in Gleichung (29) über, die damit als Lösung des speziellen Streifenproblems nachgewiesen ist. Die zweite spezielle Voraussetzung, von der wir aber bereits in diesem Paragraphen abgehen wollen, ist die, daß sowohl in (27), (29) wie auch in (32) für σ_i nur die Buchstaben ε_i oder δ_i zugelassen waren. Wir hätten statt dieser beiden Buchstaben auch ε_{i1} und ε_{i2} schreiben können.

(35) Es ist aber sofort klar, daß (27), (29) und (32) dieselbe Gestalt behalten, wenn für die σ_i unabhängig voneinander Buchstaben aus der Reihe $\varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{i3}, \varepsilon_{i4}, \dots, \varepsilon_{ik}$ zugelassen werden ($i = 1, 2, \dots, n$). Nur muß man entsprechend (25 a) jetzt setzen:

$$(35 a) \quad *_{i} = \varepsilon_{i1} + \varepsilon_{i2} + \varepsilon_{i3} + \dots + \varepsilon_{ik} \quad (i = 1, \dots, n).$$

Summiert man in (29) und (32) z. B. über σ_3 , indem man σ_3 der Reihe nach die Werte $\varepsilon_{31}, \varepsilon_{32}, \varepsilon_{33}, \dots, \varepsilon_{3k}$ annehmen läßt und dann die k Gleichungen addiert,

(36) so erkennt man, daß sowohl (32) wie (29) und daher auch das mit (29) identische (27) richtig bleiben, wenn man auf beiden Seiten dieser Gleichungen entsprechende σ_i durch $*_{i}$ ersetzt. Da die $\sigma_1 \sigma_2 \sigma_3 \dots \sigma_n$ relative Häufigkeiten bedeuten, ist ihre Summe gleich 1, also

$$(37) \quad \sum_{\sigma_i} \sigma_1 \sigma_2 \sigma_3 \dots \sigma_n = 1 = *_{1} *_{2} *_{3} \dots *_{n} \quad (i = 1, 2, \dots, n).$$

Wir knüpfen jetzt wieder an (29) an, das wir mit Rücksicht auf (36) folgendermaßen schreiben können:

$$(38) \quad v'_1 v'_2 v'_3 \dots v'_n = \frac{1}{2^n} \cdot \sum_{a_i} (v_1 \cdot *_{1})_{a_1} \cdot (v_2 \cdot *_{2})_{a_2} \cdot (v_3 \cdot *_{3})_{a_3} \dots (v_n \cdot *_{n})_{a_n}$$

$$(a_i = 1, 2; i = 1, 2, \dots, n; v_i = \sigma_i \text{ oder } *_{i}; \sigma_i = \varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{ik}).$$

Entsprechend erhalten wir für das mit (29) identische (27)

$$(! 39) \quad v'_i = \frac{(v_i \cdot *_{i})_1 + (v_i \cdot *_{i})_2}{2} \quad (i = 1, \dots, n).$$

Nun bilden wir

$$(40) \quad \sigma''_1 \sigma''_2 \dots \sigma''_n = \frac{1}{2^{1,2}} \cdot \sum_{a_i} (\sigma'_1 \cdot *'_1)_{a_1} \cdot (\sigma'_2 \cdot *'_2)_{a_2} \dots (\sigma'_n \cdot *'_n)_{a_n}.$$

Auf der rechten Seite verfährt man genau, wie in (27a) angegeben ist, und setzt hernach für die $v'_1 v'_2 \dots v'_n$ die Werte aus (38) ein. Man erhält:

$$(41) \quad \sigma''_1 \sigma''_2 \dots \sigma''_n = \frac{1}{2^{2n}} \cdot \sum_{a_i} (\sigma_1 \cdot *_{a_1} \cdot *_{a_1} \cdot *_{a_1} \cdot *_{a_1})_{a_1} \cdot (\sigma_2 \cdot *_{a_2} \cdot *_{a_2} \cdot *_{a_2} \cdot *_{a_2})_{a_2} \cdot \dots \cdot (\sigma_n \cdot *_{a_n} \cdot *_{a_n} \cdot *_{a_n} \cdot *_{a_n})_{a_n}$$

$$(a_i = 1, 2, 3, 4; i = 1, \dots, n).$$

Für diesen Ausdruck gilt genau dieselbe Rechenregel wie in (27a). Man erkennt, daß man dasselbe Resultat erhalten hätte, wenn man in (40) rein formal (39) eingesetzt hätte. Ist m irgendeine ganze Zahl, so folgt entsprechend:

$$(42) \quad \sigma_1^m \sigma_2^m \dots \sigma_n^m = \frac{1}{2^{m \cdot n}} \cdot \sum_{a_i} (\sigma_1 \cdot *_{a_i} \cdot *_{a_i} \cdot \dots \cdot *_{a_i})_{a_i} \cdot \dots \cdot (\sigma_n \cdot *_{a_n} \cdot *_{a_n} \cdot \dots \cdot *_{a_n})_{a_n}$$

$$(a_i = 1, 2, 3, \dots, 2^m; i = 1, 2, \dots, n).$$

(42a) Hierbei besteht die einzelne runde Klammer $(\sigma_i \cdot *_{a_i} \cdot *_{a_i} \cdot \dots \cdot *_{a_i})$ aus 2^m Symbolfaktoren, enthält also $(2^m - 1)$ Sterne. Zu demselben Resultat (42) gelangt man, wenn man die entsprechenden Transformationen $v_i'', v_i''', \dots, v_i^m$ unmittelbar an der Symbolgleichung (39) ausführt, ohne — der Vorschrift entsprechend [s. (26) und (27)] — bei Symbolfaktoren mit gleichem Index i die Reihenfolge der Faktoren zu ändern. Es folgt:

$$(! 43) \quad v_i^m = \frac{1}{2^m} \cdot \sum_{a_i} (v_i \cdot *_{a_i} \cdot *_{a_i} \cdot *_{a_i} \cdot \dots \cdot *_{a_i})_{a_i}$$

$$(a_i = 1, 2, 3, \dots, 2^m; v_i = \sigma_i \text{ oder } *_{a_i}).$$

Auf der rechten Seite steht die Summe aller Permutationen, die zwischen den $(2^m - 1)$ Sternen und dem einen v_i möglich sind. Für $v_i = \sigma_i$ ergibt sich, wenn man das Produkt $\sigma_1^m \sigma_2^m \dots \sigma_n^m$ bildet, genau (42). Die Summe in (42) können wir in n Teilsummen zerlegen.

$$(44) \quad \sigma_1^m \sigma_2^m \dots \sigma_n^m = \frac{1}{2^{m \cdot n}} \cdot \{ \sum^{(1)} + \sum^{(2)} + \sum^{(3)} + \dots + \sum^{(c)} + \dots + \sum^{(n)} \},$$

wobei also z. B. $\sum^{(3)}$ diejenigen Summanden umfaßt, unter deren Indizes a_1, \dots, a_n , die der Werte 1 bis 2^m fähig sind, sich genau drei verschiedene Zahlen aus der Reihe $1, \dots, 2^m$ finden. Wir greifen aus $\sum^{(c)}$ einen beliebigen Summanden heraus und veranschaulichen uns dessen Gestalt an einem Beispiel, indem wir setzen: $n = 6, m = 3, c = 4, a_1 = 5, a_2 = 8, a_3 = 1, a_4 = 5, a_5 = 2, a_6 = 5$. Von diesem Summanden erhalten wir ein anschauliches Bild, wenn wir [s. (42)] die Symbolfaktoren der runden Klammern in der Reihenfolge, die der Index hinter der Klammer angibt, senkrecht untereinandersetzen.

a_1	a_2	a_3	a_4	a_5	a_6
= 5	= 8	= 1	= 5	= 2	= 5
* ₁	* ₂	σ_3	* ₄	* ₅	* ₆
* ₁	* ₂	* ₃	* ₄	σ_5	* ₆
* ₁	* ₂	* ₃	* ₄	* ₅	* ₆
* ₁	* ₂	* ₃	* ₄	* ₅	* ₆
σ_1	* ₂	* ₃	σ_4	* ₅	σ_6
* ₁	* ₂	* ₃	* ₄	* ₅	* ₆
* ₁	* ₂	* ₃	* ₄	* ₅	* ₆
* ₁	σ_2	* ₃	* ₄	* ₅	* ₆

(45) Auf diese Weise setzen sich die Symbole von selbst — gemäß (27a) — zu wirklichen Zahlen zusammen; denn die Horizontalreihen sind wirkliche Zahlen. Wir erhalten so ein Produkt von $2^m = 8$ wirklichen Faktoren. Von diesen 2^m Faktoren enthalten aber nur $c = 4$ Faktoren, nämlich die erste, zweite, fünfte, achte Horizontalreihe, den Buchstaben σ . Die übrigen setzen sich nur aus Sternen zusammen und haben daher nach (37) den Wert 1. Wenn wir nun in (45) die Horizontalreihen permutieren, so werden bei allen Permutationen, bei denen ein Buchstabe σ verschoben wird, auch die in (45) senkrecht über den σ_i stehenden a_i ihre Werte ändern. Wir erhalten also durch diese horizontalen Permutationen neue Summanden aus $\Sigma^{(c)}$, die aber alle, weil zwischen wirklichen Zahlen die Reihenfolge der Faktoren willkürlich ist, einander im Werte gleich sind. Nun lassen sich 2^m Horizontalreihen, von denen $(2^m - c)$ einander gleich sind, auf $2^m! : (2^m - c)!$ verschiedene Arten permutieren.

(46) Wir können also in $\Sigma^{(c)}$ immer $2^m! : (2^m - c)!$ Summanden zusammenfassen, wobei sinnvollerweise $c \leq 2^m$ und $0! = 1$ zu setzen ist. Es fragt sich nun, welchen von diesen einander gleichen Summanden wir als Repräsentanten auswählen. Wir setzen fest, daß alle durch horizontale Permutation ineinander übergehenden Summanden durch denjenigen ersetzt werden sollen, für den gilt:

$$(47) \quad a_i \leq 1 + \text{maximum}(a_1, a_2, a_3, \dots, a_{i-1}),$$

wobei $a_1 = 1$ festgesetzt wird und $\max(a_1, \dots, a_{i-1})$ den größten Zahlwert unter a_1, a_2, \dots, a_{i-1} bedeutet. Man überzeugt sich, daß es stets einen und nur einen Summanden dieser Art gibt. Aus (47) folgt sofort für alle nun übrigbleibenden Summanden, d. h. die Repräsentanten:

$$(48) \quad a_i \leq i \quad (i = 1, 2, \dots, n), \text{ und speziell für die Summanden von } \Sigma^{(c)} \\ a_i \leq c.$$

(44) geht jetzt — wegen (46) bis (48) — über in

$$(49) \quad \sigma_1^m \sigma_2^m \dots \sigma_n^m = \frac{1}{2^{m \cdot n}} \cdot \sum_c^{1, n} \frac{2^m!}{(2^m - c)!} \\ \times \sum_{a_i}^{1, c} (\sigma_1 \cdot *_{1} \cdot *_{1} \cdot \dots \cdot *_{1})_{a_1} \cdot \dots \cdot (\sigma_n \cdot *_{n} \cdot *_{n} \cdot \dots \cdot *_{n})_{a_n} \\ (i = 1, 2, \dots, n; c \leq 2^m).$$

Hierbei sind aber die a_i — im Gegensatz zu (44) und (42) — durch die Ungleichung (47) eingeschränkt, während der Index c in $\sum^{(c)}$ auch für die reduzierte Summandenzahl dieselbe Bedeutung hat wie in (44). Die runde Klammer $(\sigma_i \cdot *_{i} \cdot *_{i} \cdot \dots \cdot *_{i})$ enthält [s. (42a)] immer noch $(2^m - 1)$ Sterne. Da aber in $\sum^{(c)}$ a_i nur der Werte von 1 bis c fähig ist, so werden die Summanden von $\sum^{(c)}$ [s. (45)] nur in den ersten c Horizontalfaktoren den Buchstaben σ enthalten. Die unteren $(2^m - c)$ Horizontalfaktoren enthalten nur Sterne und haben daher nach (37) den Wert 1. Es werden also bei den Summanden von $\sum^{(c)}$ die letzten $(2^m - c)$ Sterne nur zur Bildung solcher belangloser Faktoren beitragen und können daher weggelassen werden. (49) geht dann über in

$$(50) \quad \sigma_1^m \sigma_2^m \dots \sigma_n^m = \frac{1}{2^{n \cdot m}} \cdot \sum_c^{1, n} \frac{2^m!}{(2^m - c)!} \cdot \sum_{a_i}^{1, c} (\sigma_1 \cdot *_{1}^{(c-1)})_{a_1} \cdot (\sigma_2 \cdot *_{2}^{(c-1)})_{a_2} \dots (\sigma_n \cdot *_{n}^{(c-1)})_{a_n},$$

wobei (nach (47)) $a_1 = 1$; $a_i \leq 1 + \max(a_1, \dots, a_{i-1})$; $i = 1, 2, \dots, n$; $c \leq 2^m$. Hierbei ist der obere Index von $*_{i}^{(c-1)}$ nicht mit dem oberen Index m der linken Seite zu verwechseln, sondern es ist z. B.

$$(50a) \quad *_{i}^{(4)} = *_{i} \cdot *_{i} \cdot *_{i} \cdot *_{i}; \quad *_{i}^{(0)} = 1.$$

(50b) Die Bemerkung (36) gilt entsprechend auch für (50).

Es ist zu beachten, daß die $\sum^{(c)}$ von m völlig unabhängig sind. Für $m \rightarrow \infty$ rückt der Koeffizient von $\sum^{(n)}$

$$(51) \quad k_n = \frac{2^m!}{(2^m)^n \cdot (2^m - n)!} \rightarrow 1; \quad \lim_{m \rightarrow \infty} k_n = 1 \quad (\text{bei festem } n).$$

Nun ist die Summe der Zahlkoeffizienten aller Summanden, wie aus dem mit (50) identischen (42) folgt, gleich 1. Schreibt man daher (50) in der Form:

$$(52) \quad \sigma_1^m \sigma_2^m \dots \sigma_n^m = k_n \cdot \sum^{(n)} + R,$$

so wird $R \leq 1 - k_n$, da jeder einzelne Summand von $\sum^{(c)} \leq 1$ ist. Da $0 \leq R$, gilt wegen (51) $\lim_{m \rightarrow \infty} R = 0$. Dann folgt aus (52)

$$(53) \quad \sigma_1^\infty \sigma_2^\infty \dots \sigma_n^\infty = \sum^{(n)} = (\sigma_1 \cdot *_{1}^{(n-1)})_1 \cdot (\sigma_2 \cdot *_{2}^{(n-1)})_2 \cdot (\sigma_3 \cdot *_{3}^{(n-1)})_3 \dots (\sigma_n \cdot *_{n}^{(n-1)})_n.$$

Denn wegen der einschränkenden Voraussetzungen über die a_i enthält $\sum^{(m)}$ nur den einen Summanden, für den gilt: $a_i = i$; ($i = 1, \dots, n$). Setzt man nun die rechte Seite von (53) — entsprechend (27a) oder siehe auch (45) — zu wirklichen Zahlen zusammen, so ergibt sich ein Resultat, das man symbolisch schreiben kann:

$$(!54) \quad \sigma_i^\infty = *_1 *_2 *_3 \dots *_{i-1} \sigma_i *_{i+1} \dots *_{n-1} *_n \quad (i = 1, \dots, n).$$

Diese Symbolgleichung ist insofern merkwürdig, als auf ihrer rechten Seite eine wirkliche Zahl steht. Wir erhalten somit den Satz, daß im stationären Zustande die Symbole σ_i als wirkliche Zahlen aufgefaßt werden dürfen. Die rechte Seite von (54) bedeutet die relative Häufigkeit, mit der — vor Beginn der Streifenmischung — der bestimmte Buchstabe σ sich in einem Kästchen fand, auf dessen Rückseite die Zahl i rot aufgetragen war. Entsprechend (54) folgt:

$$(!54a) \quad *_{i-1}^\infty = *_{i-1} *_{i-2} *_{i-3} \dots *_{i-1} *_{i-1} *_{i+1} \dots *_{n-1} *_{i-1} = 1 \quad [\text{nach (37)}]; \quad i = 1 \dots n.$$

§ 5.

Gekoppelte Genreihen.

Dieser Paragraph knüpft an Gleichung (32) an. Genau so, wie (29) aus (14) abgeleitet wurde, kann man (32) aus (14) durch Modifikation der Gleichung (15) herleiten. Doch sei hiervon Abstand genommen und statt dessen auf die bereits zitierte Arbeit des Verfassers verwiesen, in der eine ähnliche Untersuchung durchgeführt ist. Überträgt man die — im vorigen Paragraphen — an (29) ausgeführten Operationen ganz entsprechend auf (32), so ergibt sich, (50) entsprechend:

$$(55) \quad \begin{aligned} & \sigma_1^m \sigma_2^m \dots \sigma_n^m \\ &= \sum_c^{1, n} \sum_{a_i}^{1, c} (\sigma_1 \cdot *_{i-1}^{(c-1)})_{a_1} \cdot (\sigma_2 \cdot *_{i-2}^{(c-1)})_{a_2} \dots (\sigma_n \cdot *_{i-1}^{(c-1)})_{a_n} \cdot \omega^{(m)}(a_1, a_2, \dots, a_n) \\ & (a_1 = 1; a_i \leq 1 + \max(a_1, a_2, \dots, a_{i-1}); i = 1, 2, \dots, n; c \leq 2^m). \end{aligned}$$

Der Index m bei $\omega^{(m)}$ entspricht genau dem Index m der linken Seite, hat also mit dem Index $(c-1)$ nichts zu tun. Die Größen $\omega^{(m)}$ sind durch die Transformation (32) natürlich eindeutig definiert; sie aber explizit durch die $\omega^{(1)}$ auszudrücken, ist die Aufgabe, die in diesem Paragraphen gelöst werden soll. Die $\omega^{(1)}$ sind übrigens — wegen der Beschränkung des Summationsindex a_i — nicht mit den ω in (32) identisch, sondern setzen sich additiv aus je zweien zusammen, z. B.

$$\omega^{(1)}(1, 1, 2, 2, 1, 2, 2) = \omega(1, 1, 2, 2, 1, 2, 2) + \omega(2, 2, 1, 1, 2, 1, 1).$$

Es gilt daher nach (34)

$$(55a) \quad \sum_{a_i} \omega^{(1)}(a_1, a_2, \dots, a_n) = 1 \quad (a_i = 1, 2; i = 2, 3, \dots, n; a_1 = 1).$$

(55b) (55) bleibt — wegen (36) — richtig, wenn man auf beiden Seiten entsprechende σ_i durch $*_i$ ersetzt. Hierbei schreiben wir kürzer:

$$(56) \quad \sigma_1^m \sigma_2^m \dots \sigma_n^m = \sum_{a_i} \pi(a_1, a_2, \dots, a_n) \cdot \omega^{(m)}(a_1, a_2, \dots, a_n) \\ (a_i = 1; a_i \leq 1 + \max(a_1, \dots, a_{i-1}); i = 1, \dots, n).$$

Hierbei ist gesetzt

$$(57) \quad \pi(a_1, a_2, \dots, a_n) = (\sigma_1 \cdot *_1^{(c-1)})_{a_1} \cdot (\sigma_2 \cdot *_2^{(c-1)})_{a_2} \dots (\sigma_n \cdot *_n^{(c-1)})_{a_n}.$$

Wendet man auf (57) die Transformation (56) — für $m = 1$ — an [s. auch (55b)], so folgt:

$$(58a) \quad \pi'(a_1, \dots, a_n) = (\sigma'_1 \cdot (*'_1)^{(c-1)})_{a_1} \cdot (\sigma'_2 \cdot (*'_2)^{(c-1)})_{a_2} \dots (\sigma'_n \cdot (*'_n)^{(c-1)})_{a_n}$$

$$(58) \quad = \pi'(a_1, \dots, a_n) = \sum_{o_i} \pi(o_1, o_2, \dots, o_n) \cdot f^{(1)} \left\{ \begin{matrix} o_1, o_2, \dots, o_n \\ a_1, a_2, \dots, a_n \end{matrix} \right\} \\ (i = 1, \dots, n; o = o[n]; \text{ wenn } a_i \neq a_k, \text{ so ist } o_i \neq o_k).$$

(58b) $o = o[n]$ ist eine Abkürzung für: $o_i = 1, 2, \dots, i; o_1 = 1; o_i \leq 1 + \max(o_1, \dots, o_{i-1}); i = 1, 2, \dots, n$; gelegentlich werden wir auch schreiben: $o_i = o_i[n]$.

Der Summationsindex o_i ist also hier stärker eingeschränkt als a_i in (56). Die Größen $f^{(1)}$ kann man natürlich explizit angeben, indem man die rechte Seite von (58a) ausrechnet; doch soll dies erst am Schluß dieses Paragraphen [s. (78)] geschehen. Da jedoch die $f^{(1)}$ Produkte gewisser $\omega^{(1)}$ sind [s. (55a)], so folgt:

$$(59) \quad 0 \leq f^{(1)} \left\{ \begin{matrix} o_1, o_2, \dots, o_n \\ a_1, a_2, \dots, a_n \end{matrix} \right\} \leq 1.$$

Diese Beziehung folgt auch, da die $f^{(1)}$ offenkundig ≥ 0 sind, aus der Gleichung

$$(60) \quad 1 = \sum_{o_i} f^{(1)} \left\{ \begin{matrix} o_1, o_2, \dots, o_n \\ a_1, a_2, \dots, a_n \end{matrix} \right\}$$

$$(i = 1, \dots, n; o = o[n], a = a[n]; \text{ wenn } a_i \neq a_k, \text{ so ist } o_i \neq o_k).$$

(60) beweisen wir wie folgt. (55) gilt ganz allgemein, welche Werte wir auch für die v_1, v_2, \dots, v_n der rechten Seite ($v_i = \sigma_i$ oder $*_i; i = 1, \dots, n$) wählen mögen. Wählen wir also die Werte der Gleichungen (54) und (54a), so dürfen wir auf der rechten Seite von (55) die σ als wirkliche Zahlen behandeln und die Sterne gleich 1 setzen. Dann folgt ..

$$(61) \quad \sigma_1^m \sigma_2^m \dots \sigma_n^m = \sigma_1 \sigma_2 \dots \sigma_n \cdot \sum_c \sum_{a_i}^{(c)} \omega^{(m)}(a_1, \dots, a_n)$$

($c = 1, \dots, n$; $i = 1, \dots, n$; $a = a[n]$; $c \leq 2^m$).

Und ebenso folgt aus (57), daß alle $\pi(a_1, a_2, \dots, a_n)$ einander gleich werden, also auch [s. (58)]

$$(62) \quad \pi(o_1, o_2, \dots, o_n) = \sigma_1 \sigma_2 \dots \sigma_n.$$

Aus (61) folgt für $m = 1$

$$\sigma'_1 \sigma'_2 \dots \sigma'_n = \sigma_1 \sigma_2 \dots \sigma_n \cdot \sum_c \sum_{a_i}^{1,2(c)} \omega^{(1)}(a_1 \dots a_n) = \sigma_1 \sigma_2 \dots \sigma_n \quad [\text{nach (55 a)}],$$

und ebenso gilt — wegen (55 b) — $v'_1 v'_2 \dots v'_n = v_1 \dots v_n$ ($v_i = \sigma_i$ oder \ast_i). Jetzt folgt sofort aus (57) und (58 a)

$$(63) \quad \pi'(a_1, \dots, a_n) = \pi(a_1, \dots, a_n) = \sigma_1 \sigma_2 \dots \sigma_n.$$

Definiert man nun Größen $\pi^{(m)}(a_1, a_2, \dots, a_n)$, indem man die Transformation (58) wiederholt auf sich selbst anwendet, so sind durch die folgende Darstellung von $\pi^{(m)}$ wieder Größen $f^{(m)}$ definiert:

$$(64) \quad \pi^{(m)}(a_1, a_2, \dots, a_n) = \sum_{o_i} \pi(o_1, o_2, \dots, o_n) \cdot f^{(m)} \left\{ \begin{matrix} o_1, o_2, \dots, o_n \\ a_1, a_2, \dots, a_n \end{matrix} \right\}$$

($i = 1, \dots, n$; $o = o[n]$; $a = a[n]$; wenn $a_i \neq a_k$, so ist $o_i \neq o_k$).

Nimmt man nun für die π die Werte aus (62), so folgt — entsprechend (63) — der Reihe nach $\pi(a_1, \dots, a_n) = \pi^{(1)}(a_1, \dots, a_n) = \pi^{(2)}(a_1, \dots, a_n) = \dots = \pi^{(m)}(a_1, \dots, a_n) = \sigma_1 \sigma_2 \dots \sigma_n$. Dann liefert (64)

$$(65a) \quad 1 = \sum_{o_i} f^{(m)} \left\{ \begin{matrix} o_1, o_2, \dots, o_n \\ a_1, a_2, \dots, a_n \end{matrix} \right\}$$

($i = 1, \dots, n$; $o = o[n]$; $a = a[n]$; wenn $a_i \neq a_k$, so ist $o_i \neq o_k$).

Damit ist auch (60) bewiesen.

Durch (64) sind die $f^{(m)}$ definiert. Aus der Identität

$$\begin{aligned} & \pi^{(m+1)}(a_1, \dots, a_n) \\ &= \sum_{o_i} \pi'(o_1, \dots, o_n) \cdot f^{(m)} \left\{ \begin{matrix} o_1, \dots, o_n \\ a_1, \dots, a_n \end{matrix} \right\} \equiv \sum_{o_i} \pi(o_1, \dots, o_n) \cdot f^{(m+1)} \left\{ \begin{matrix} o_1, \dots, o_n \\ a_1, \dots, a_n \end{matrix} \right\} \end{aligned}$$

ergeben sich Rekursionsformeln für $f^{(m+1)}$ dadurch, daß man für die π' den sich aus (64) — für $m = 1$ — ergebenden Wert einsetzt.

$$(65) \quad f^{(m+1)} \left\{ \begin{matrix} o_1, \dots, o_n \\ a_1, \dots, a_n \end{matrix} \right\} = \sum_{u_i} f^{(m)} \left\{ \begin{matrix} u_1, \dots, u_n \\ a_1, \dots, a_n \end{matrix} \right\} \cdot f^{(1)} \left\{ \begin{matrix} o_1, \dots, o_n \\ u_1, \dots, u_n \end{matrix} \right\}$$

($i = 1, \dots, n$; $a = a[n]$; $o = o[n]$; $u = u[n]$;
wenn $a_i \neq a_k$, so ist $u_i \neq u_k$; wenn $u_i \neq u_k$, so ist $o_i \neq o_k$).

Wir wollen nun für die Relation

wenn $a_i \neq a_k$, so ist $u_i \neq u_k$

die neue Bezeichnung $a \leq u$ einführen, gelegentlich auch dafür $a_i \leq u_i$ schreiben. Hierbei handelt es sich natürlich nur um eine symbolische Größenbeziehung, denn *ohne* Index haben a_i und u_i sowie a und u gar keinen Zahlwert. Wenn wir nun die $f^{(m)}$ nach (65) durch fortgesetzte Rekursion bestimmen, ergibt sich:

$$(66) \quad \left\{ \begin{array}{l} f^{(m)} \left\{ \begin{array}{l} o_1, o_2, \dots, o_n \\ a_1, a_2, \dots, a_n \end{array} \right\} \\ = \sum_{u_{i_2}} \sum_{u_{i_3}} \dots \sum_{u_{i_m}} f^{(1)} \left\{ \begin{array}{l} u_{1_2}, u_{2_2}, \dots, u_{n_2} \\ a_1, a_2, \dots, a_n \end{array} \right\} \cdot f^{(1)} \left\{ \begin{array}{l} u_{1_3}, u_{2_3}, \dots, u_{n_3} \\ u_{1_2}, u_{2_2}, \dots, u_{n_2} \end{array} \right\} \dots \\ \times f^{(1)} \left\{ \begin{array}{l} u_{1(r+1)}, u_{2(r+1)}, \dots, u_{n(r+1)} \\ u_{1_r}, u_{2_r}, \dots, u_{n_r} \end{array} \right\} \dots f^{(1)} \left\{ \begin{array}{l} o_1, o_2, \dots, o_n \\ u_{1_m}, u_{2_m}, \dots, u_{n_m} \end{array} \right\}. \end{array} \right.$$

$$(66a) \quad u_{i_r} = u_{|r}[n]; \quad r = 2, 3, \dots, m;$$

$$a_i \leq u_1 \leq u_2 \leq u_3 \dots \leq u_r \dots \leq u_m \leq o_i.$$

Hierbei bedeutet z. B. $u_{i_2} \leq u_{i_3}$: Wenn $u_{i_2} \neq u_{k_2}$, so ist $u_{i_3} \neq u_{k_3}$, und entsprechend $u_{i_m} \leq o_i$: Wenn $u_{i_m} \neq u_{k_m}$, so ist $o_i \neq o_k$. Hieraus folgt sofort: Wenn $o_i = o_k$, so ist $u_{i_m} = u_{k_m}$, wofür wir abgekürzt schreiben; $o_i = > u_m$. Diese symbolischen Ungleichungen sind also von rechts nach links und von links nach rechts lesbar. Diese Auffassung erleichtert die Übersicht über den Wertevorrat der u_{i_r} . (66) gilt nur für $m \geq 2$. Jetzt zerlegen wir die Summe auf der rechten Seite von (66) wie folgt:

$$(67) \quad f^{(m)} \left\{ \begin{array}{l} o_1, \dots, o_n \\ a_1, \dots, a_n \end{array} \right\} = \sum_{u_{i_r}}^{(1)} + \sum_{u_{i_r}}^{(2)} + \dots + \sum_{u_{i_r}}^{(s)} + \sum_{u_{i_r}}^{(s+1)} + \dots$$

$$(i = 1, 2, \dots, n; r = 2, 3, \dots, m).$$

$\sum_{u_{i_r}}^{(s)}$ soll diejenigen Summanden der rechten Seite von (66) umfassen, die s Faktoren $f^{(1)}$ enthalten, bei denen die *Zahlwerte* der a_i bzw. u_{i_r} bzw. o_i in der unteren Zeile verschieden sind von den Zahlwerten der genau darüber stehenden a_i bzw. u_{i_r} bzw. o_i ; diese Summanden müssen also außerdem $(m - s)$ Faktoren enthalten, bei denen die Zahlwerte der unteren Zeile mit den genau darüber stehenden völlig übereinstimmen. Wir wollen die Gestalt des Summanden von $\sum^{(s)}$ an einem speziellen Beispiel kennenlernen und setzen zu diesem Zweck: $n = 4$, $m = 6$, $s = 2$, $a_i = 1$, $o_i = i$ ($i = 1, 2, \dots, n$). Aus $\sum^{(2)}$ werde der folgende Summand herausgegriffen:

$$f^{(1)} \left\{ \begin{array}{l} 1, 1, 1, 1 \\ 1, 1, 1, 1 \end{array} \right\} \cdot f^{(1)} \left\{ \begin{array}{l} 1, 1, 1, 1 \\ 1, 1, 1, 1 \end{array} \right\} \cdot f^{(1)} \left\{ \begin{array}{l} 1, 2, 1, 2 \\ 1, 1, 1, 1 \end{array} \right\} \cdot f^{(1)} \left\{ \begin{array}{l} 1, 2, 1, 2 \\ 1, 2, 1, 2 \end{array} \right\} \\ \cdot f^{(1)} \left\{ \begin{array}{l} 1, 2, 3, 4 \\ 1, 2, 1, 2 \end{array} \right\} \cdot f^{(1)} \left\{ \begin{array}{l} 1, 2, 3, 4 \\ 1, 2, 3, 4 \end{array} \right\}.$$

Vorübergehend benutzen wir folgende Abkürzung hierfür:

$$f^{(1)}\left\{\begin{matrix} x \\ x \end{matrix}\right\} \cdot f^{(1)}\left\{\begin{matrix} x \\ x \end{matrix}\right\} \cdot f^{(1)}\left\{\begin{matrix} y \\ x \end{matrix}\right\} \cdot f^{(1)}\left\{\begin{matrix} y \\ y \end{matrix}\right\} \cdot f^{(1)}\left\{\begin{matrix} z \\ y \end{matrix}\right\} \cdot f^{(1)}\left\{\begin{matrix} z \\ z \end{matrix}\right\}.$$

Wir greifen nun einen andern Summanden von $\Sigma^{(2)}$ heraus, der aber auch die beiden Faktoren $f^{(1)}\left\{\begin{matrix} y \\ x \end{matrix}\right\}$ und $f^{(1)}\left\{\begin{matrix} z \\ y \end{matrix}\right\}$ enthält, z. B.

$$f^{(1)}\left\{\begin{matrix} y \\ x \end{matrix}\right\} \cdot f^{(1)}\left\{\begin{matrix} y \\ y \end{matrix}\right\} \cdot f^{(1)}\left\{\begin{matrix} y \\ y \end{matrix}\right\} \cdot f^{(1)}\left\{\begin{matrix} y \\ y \end{matrix}\right\} \cdot f^{(1)}\left\{\begin{matrix} y \\ y \end{matrix}\right\} \cdot f^{(1)}\left\{\begin{matrix} z \\ y \end{matrix}\right\}.$$

Suchen wir so alle Summanden aus $\Sigma^{(2)}$ heraus, die die beiden Faktoren $f^{(1)}\left\{\begin{matrix} y \\ x \end{matrix}\right\}$ und $f^{(1)}\left\{\begin{matrix} z \\ y \end{matrix}\right\}$ enthalten, addieren alle diese Summanden, so folgt:

$$f^{(1)}\left\{\begin{matrix} y \\ x \end{matrix}\right\} \cdot f^{(1)}\left\{\begin{matrix} z \\ y \end{matrix}\right\} \cdot S^{(4)}\left(f^{(1)}\left\{\begin{matrix} x \\ x \end{matrix}\right\}, f^{(1)}\left\{\begin{matrix} y \\ y \end{matrix}\right\}, f^{(1)}\left\{\begin{matrix} z \\ z \end{matrix}\right\}\right).$$

Kürzen wir vorübergehend ab: $f^{(1)}\left\{\begin{matrix} x \\ x \end{matrix}\right\}$ mit η_1 , $f^{(1)}\left\{\begin{matrix} y \\ y \end{matrix}\right\}$ mit η_2 usw., so ist

$$\begin{cases} S^{(4)}(\eta_1, \eta_2, \eta_3) = \eta_1^4 + \eta_1^3(\eta_2 + \eta_3) + \eta_1^2(\eta_2^2 + \eta_2 \cdot \eta_3 + \eta_3^2) \\ \quad + \eta_1(\eta_2^3 + \eta_2^2 \cdot \eta_3 + \eta_2 \cdot \eta_3^2 + \eta_3^3) + \eta_2^4 + \eta_2^3 \cdot \eta_3 + \eta_2^2 \cdot \eta_3^2 + \eta_2 \cdot \eta_3^3 + \eta_3^4. \end{cases}$$

$S^{(4)}$ ist also eine symmetrische Funktion von η_1, η_2, η_3

Als die genau entsprechende symmetrische Funktion ($m-s$)-ten Grades der Größen $\eta_1, \eta_2, \eta_3, \eta_4, \dots$ usw. definieren wir:

$$(68) \quad \begin{cases} S^{(m-s)}(\eta_1, \eta_2, \dots) = \eta_1^{m-s} + \eta_1^{m-s-1} \cdot S^{(1)}(\eta_2, \dots) + \eta_1^{m-s-2} \cdot S^{(2)}(\eta_2, \dots) \\ \quad + \dots + \eta_1 \cdot S^{(m-s-1)}(\eta_2, \dots) + S^{(m-s)}(\eta_2, \dots). \end{cases}$$

$$(68a) \quad S^{(m-s)}(\eta_1) = \eta_1^{m-s}; \quad S^{(0)}(\eta_1) = 1.$$

Führen wir nun für die $f^{(1)}$ in (66) eine kürzere Bezeichnung ein, indem wir z. B. $f^{(1)}\left\{\begin{matrix} u_{14}, u_{24}, \dots, u_{n4} \\ u_{13}, u_{23}, \dots, u_{n3} \end{matrix}\right\}$ abkürzen durch $f^{(1)}\left\{\begin{matrix} u_{14} \\ u_{13} \end{matrix}\right\}$, so geht (67) bzw. (66) über in

$$(69) \quad \begin{aligned} f^{(m)}\left\{\begin{matrix} o_1 \\ u_{,1} \end{matrix}\right\} &= \sum_{s=1}^m \sum_{u_{i,r}}^{(s)} f^{(1)}\left\{\begin{matrix} u_{12} \\ u_{,1} \end{matrix}\right\} \cdot f^{(1)}\left\{\begin{matrix} u_{13} \\ u_{12} \end{matrix}\right\} \dots f^{(1)}\left\{\begin{matrix} o_1 \\ u_{,s} \end{matrix}\right\} \\ &\cdot S^{(m-s)}\left(f^{(1)}\left\{\begin{matrix} u_1 \\ u_1 \end{matrix}\right\}, \dots, f^{(1)}\left\{\begin{matrix} u_r \\ u_r \end{matrix}\right\}, \dots, f^{(1)}\left\{\begin{matrix} u_s \\ u_s \end{matrix}\right\}, f^{(1)}\left\{\begin{matrix} o_1 \\ o_1 \end{matrix}\right\}\right). \end{aligned}$$

Hierbei ist zu beachten, daß a_1 , wie es (66) entsprechen würde, durch u_{11} ersetzt ist. Für die Summationsindizes gelten dieselben Einschränkungen wie in (66a), nur ist außerdem in den symbolischen Ungleichungen noch der Fall der Gleichheit auszuschließen; d. h.

$$(169a) \quad u_1 < \neq u_2 < \neq u_3 \dots < \neq u_s < \neq o_1.$$

Im Gegensatz zu (66) ist jetzt der Fall ausgeschlossen, daß z. B. bei $f^{(1)} \begin{Bmatrix} u_{1(r+1)} \\ u_r \end{Bmatrix} = f^{(1)} \begin{Bmatrix} u_{1(r+1)} \cdots u_{n(r+1)} \\ u_{1r} \cdots u_{nr} \end{Bmatrix}$ die *Zahlwerte sämtlicher* u der unteren Zeile gleich den Zahlwerten der darüber stehenden u seien. Dagegen könnte sehr wohl z. B. $u_{2r} = u_{2(r+1)}$ sein, und nach (66a) ist ja auch stets $u_{1r} = u_{1(r+1)} = 1$ [s. (58b)]. s läuft zwar von 1 bis m ; doch werden von einem bestimmten m ab keine neuen $\sum^{(s)}$ mehr hinzukommen können, da infolge der Beschränkungen der Summationsindizes keine Summanden mehr vorhanden sind. Man sieht leicht ein, daß

$$(70) \quad s \leq (n-1) \quad \text{ist;} \quad (n \geq 2).$$

Wegen der Ungleichungen (69a), aus denen folgt $u_1 < \neq 0$, ist (69) zur Berechnung der Größen $f^{(m)} \begin{Bmatrix} o_1 \\ o_1 \end{Bmatrix} = f^{(m)} \begin{Bmatrix} o_1, o_2, \dots, o_n \\ o_1, o_2, \dots, o_n \end{Bmatrix}$ nicht brauchbar. Doch folgt unmittelbar aus (65), indem man dort a durch o ersetzt,

$$f^{(m+1)} \begin{Bmatrix} o_1 \cdots o_n \\ o_1 \cdots o_n \end{Bmatrix} = f^{(m)} \begin{Bmatrix} o_1 \cdots o_n \\ o_1 \cdots o_n \end{Bmatrix} \cdot f^{(1)} \begin{Bmatrix} o_1 \cdots o_n \\ o_1 \cdots o_n \end{Bmatrix},$$

weil die Summe der rechten Seite in (65) nur diesen einen Summanden enthält. Daher

$$(71) \quad f^{(m)} \begin{Bmatrix} o_1 \cdots o_n \\ o_1 \cdots o_n \end{Bmatrix} = \left(f^{(1)} \begin{Bmatrix} o_1 \cdots o_n \\ o_1 \cdots o_n \end{Bmatrix} \right)^m.$$

Wir wollen nun zur Erläuterung von (69) und (71) alle typischen Gleichungen dieser Art im Falle $n=3$ aufschreiben.

$$f^{(m)} \begin{Bmatrix} 1, 1, 1 \\ 1, 1, 1 \end{Bmatrix} = \left(f^{(1)} \begin{Bmatrix} 1, 1, 1 \\ 1, 1, 1 \end{Bmatrix} \right)^m;$$

$$f^{(m)} \begin{Bmatrix} 1, 1, 2 \\ 1, 1, 1 \end{Bmatrix} = f^{(1)} \begin{Bmatrix} 1, 1, 2 \\ 1, 1, 1 \end{Bmatrix} \cdot S^{(m-1)} \left(f^{(1)} \begin{Bmatrix} 1, 1, 1 \\ 1, 1, 1 \end{Bmatrix}, f^{(1)} \begin{Bmatrix} 1, 1, 2 \\ 1, 1, 2 \end{Bmatrix} \right);$$

entsprechende Gleichungen ergeben sich für $f^{(m)} \begin{Bmatrix} 1, 2, 1 \\ 1, 1, 1 \end{Bmatrix}$ und $f^{(m)} \begin{Bmatrix} 1, 2, 2 \\ 1, 1, 1 \end{Bmatrix}$.

$$f^{(m)} \begin{Bmatrix} 1, 2, 3 \\ 1, 1, 1 \end{Bmatrix} = f^{(1)} \begin{Bmatrix} 1, 2, 3 \\ 1, 1, 1 \end{Bmatrix} \cdot S^{(m-1)} \left(f^{(1)} \begin{Bmatrix} 1, 1, 1 \\ 1, 1, 1 \end{Bmatrix}, f^{(1)} \begin{Bmatrix} 1, 2, 3 \\ 1, 2, 3 \end{Bmatrix} \right)$$

$$+ f^{(1)} \begin{Bmatrix} 1, 1, 2 \\ 1, 1, 1 \end{Bmatrix} \cdot f^{(1)} \begin{Bmatrix} 1, 2, 3 \\ 1, 1, 2 \end{Bmatrix} \cdot S^{(m-2)} \left(f^{(1)} \begin{Bmatrix} 1, 1, 1 \\ 1, 1, 1 \end{Bmatrix}, f^{(1)} \begin{Bmatrix} 1, 1, 2 \\ 1, 1, 2 \end{Bmatrix}, f^{(1)} \begin{Bmatrix} 1, 2, 3 \\ 1, 2, 3 \end{Bmatrix} \right)$$

$$+ f^{(1)} \begin{Bmatrix} 1, 2, 1 \\ 1, 1, 1 \end{Bmatrix} \cdot f^{(1)} \begin{Bmatrix} 1, 2, 3 \\ 1, 2, 1 \end{Bmatrix} \cdot S^{(m-2)} \left(f^{(1)} \begin{Bmatrix} 1, 1, 1 \\ 1, 1, 1 \end{Bmatrix}, f^{(1)} \begin{Bmatrix} 1, 2, 1 \\ 1, 2, 1 \end{Bmatrix}, f^{(1)} \begin{Bmatrix} 1, 2, 3 \\ 1, 2, 3 \end{Bmatrix} \right)$$

$$+ f^{(1)} \begin{Bmatrix} 1, 2, 2 \\ 1, 1, 1 \end{Bmatrix} \cdot f^{(1)} \begin{Bmatrix} 1, 2, 3 \\ 1, 2, 2 \end{Bmatrix} \cdot S^{(m-2)} \left(f^{(1)} \begin{Bmatrix} 1, 1, 1 \\ 1, 1, 1 \end{Bmatrix}, f^{(1)} \begin{Bmatrix} 1, 2, 2 \\ 1, 2, 2 \end{Bmatrix}, f^{(1)} \begin{Bmatrix} 1, 2, 3 \\ 1, 2, 3 \end{Bmatrix} \right).$$

$$f^{(m)} \begin{Bmatrix} 1, 1, 2 \\ 1, 1, 2 \end{Bmatrix} = \left(f^{(1)} \begin{Bmatrix} 1, 1, 2 \\ 1, 1, 2 \end{Bmatrix} \right)^m;$$

$$f^{(m)} \begin{Bmatrix} 1, 2, 3 \\ 1, 1, 2 \end{Bmatrix} = f^{(1)} \begin{Bmatrix} 1, 2, 3 \\ 1, 1, 2 \end{Bmatrix} \cdot S^{(m-1)} \left(f^{(1)} \begin{Bmatrix} 1, 1, 2 \\ 1, 1, 2 \end{Bmatrix}, f^{(1)} \begin{Bmatrix} 1, 2, 3 \\ 1, 2, 3 \end{Bmatrix} \right).$$

Entsprechende Gleichungen ergeben sich, wenn man in den beiden letzten 1, 1, 2 durch 1, 2, 1 oder 1, 2, 2 ersetzt.

$$f^m \left\{ \begin{matrix} 1, 2, 3 \\ 1, 2, 3 \end{matrix} \right\} = \left(f^{(1)} \left\{ \begin{matrix} 1, 2, 3 \\ 1, 2, 3 \end{matrix} \right\} \right)^m.$$

Also im ganzen zwölf Gleichungen.

Eine numerisch brauchbare Rekursionsformel für die S erhalten wir aus (68), indem wir dort η_1 durch ϑ_1 ersetzen und dann bilden:

$$(72) \quad \left\{ \begin{array}{l} \frac{S^{(m-s)}(\eta_1, \eta_2, \dots) - S^{(m-s)}(\vartheta_1, \eta_2, \dots)}{\eta_1 - \vartheta_1} \\ = S^{(m-s-1)}(\eta_1, \vartheta_1) + S^{(m-s-2)}(\eta_1, \vartheta_1) \cdot S^{(1)}(\eta_2, \dots) \\ + S^{(m-s-3)}(\eta_1, \vartheta_1) \cdot S^{(2)}(\eta_2, \dots) + \dots + S^{(0)}(\eta_1, \vartheta_1) \cdot S^{(m-s-1)}(\eta_2, \dots) = \end{array} \right.$$

$$(72) \quad \frac{S^{(m-s)}(\eta_1, \eta_2, \dots) - S^{(m-s)}(\vartheta_1, \eta_2, \dots)}{\eta_1 - \vartheta_1} = S^{(m-s-1)}(\vartheta_1, \eta_1, \eta_2, \dots).$$

Damit ergibt sich folgende Regel zur Berechnung der Größen $S^{(m-s)}$ in (69), die wir der Kürze halber bezeichnen mit $S^{(m-s)}(\eta_1, \dots, \eta_r, \dots, \eta_s, \eta_{s+1})$.

$$\begin{array}{ccccccc} \eta_1^m & \eta_2^m & \eta_3^m & \eta_4^m & \dots & \eta_s^m & \eta_{s+1}^m \\ \frac{\eta_1^m - \eta_2^m}{\eta_1 - \eta_2}, & \frac{\eta_2^m - \eta_3^m}{\eta_2 - \eta_3}, & \frac{\eta_3^m - \eta_4^m}{\eta_3 - \eta_4}, & \dots & & & \\ \frac{\eta_1^m - \eta_2^m}{\eta_1 - \eta_2} \frac{\eta_2^m - \eta_3^m}{\eta_2 - \eta_3}; & \frac{\eta_2^m - \eta_3^m}{\eta_2 - \eta_3} \frac{\eta_3^m - \eta_4^m}{\eta_3 - \eta_4}; & \dots & & & & \\ \text{usw.} & \text{usw.} & \text{usw.} & & & & \end{array}$$

Hieraus folgt, da s nach (70) eine endliche Zahl ist, daß für alle echten Brüche η_r gelten muß

$$(73) \quad \lim_{m \rightarrow \infty} S^{(m-s)}(\eta_1, \dots, \eta_r, \dots, \eta_{s+1}) = 0, \quad \text{wenn } 0 \leq \eta_r < 1 \text{ ist.}$$

(73) bleibt auch richtig, wenn einige der η_r einander gleich sind, auch wenn sie gleich null sein sollten. Von (73) wollen wir sofort eine Anwendung machen, indem wir voraussetzen [s. 59]

$$(74) \quad 0 \leq f^{(1)} \left\{ \begin{matrix} a_1, a_2, \dots, a_n \\ a_1, a_2, \dots, a_n \end{matrix} \right\} < 1, \quad \text{wenn mindestens ein } a_i \neq i; \quad a = a[n].$$

Wenn aber alle $a_i = i$ sind, so folgt aus (60)

$$(74a) \quad 1 = \sum_{o_i} f^{(1)} \left\{ \begin{matrix} o_1, o_2, \dots, o_n \\ 1, 2, \dots, n \end{matrix} \right\} = f^{(1)} \left\{ \begin{matrix} 1, 2, \dots, n \\ 1, 2, \dots, n \end{matrix} \right\},$$

da wegen $a_i \leq o_i$ die Summe \sum nur diesen Summanden enthält. Dann folgt aus (69)

$$(75) \quad \lim_{m \rightarrow \infty} f^{(m)} \left\{ \begin{matrix} o_1 \\ u_{11} \end{matrix} \right\} = f^{(m)} \left\{ \begin{matrix} o_1, o_2, \dots, o_n \\ u_{11}, u_{21}, \dots, u_{n1} \end{matrix} \right\} = 0,$$

wenn mindestens ein $o_i \neq i$ ist, da dann wegen (74) auch (73) gilt und daher die $S^{(m-s)}$ auf der rechten Seite von (69) sämtlich null werden. Nun folgt aus (65a)

$$(75a) \quad \lim_{m \rightarrow \infty} \sum_{o_i} f^{(m)} \left\{ \begin{matrix} o_1, o_2, \dots, o_n \\ a_1, a_2, \dots, a_n \end{matrix} \right\} = 1 = f^\infty \left\{ \begin{matrix} 1, 2, \dots, n \\ a_1, a_2, \dots, a_n \end{matrix} \right\}$$

$$(o = o[n], a = a[n]; a_i < = o_i).$$

Wegen (75) und (75a) folgt aus (64)

$$(75b) \quad \pi^\infty(a_1, a_2, \dots, a_n) = \pi(1, 2, \dots, n),$$

also auch [s. (57)]

$$\pi^\infty(1, 1, 1, \dots, 1) = \pi(1, 2, \dots, n).$$

Diese Gleichung ist aber völlig identisch mit (53) und daher auch mit (54) und (54a). Wir haben damit von neuem (zum ersten Male in der bereits zitierten Arbeit des Verfassers) den Satz bewiesen, daß der Koppelungsgrad der Erbfaktoren auf den Endzustand der Entwicklung überhaupt keinen Einfluß hat. Über die Koppelung [s. (32) bis (33)] wurde nichts anderes vorausgesetzt, als daß ihre Gesetzmäßigkeit die des Zufalls ist und daher mit den Mitteln der Wahrscheinlichkeitsrechnung beschrieben werden kann, denn von der Voraussetzung (74), die unsern Beweis sehr erleichtert hat, werden wir sofort zeigen, daß sie die Allgemeinheit der Problemstellung nicht beeinträchtigt.

(76) Zu diesem Zwecke wollen wir die in (58) eingeführten Größen $f^{(1)} \left\{ \begin{matrix} o_1, o_2, \dots, o_n \\ e_1, e_2, \dots, e_n \end{matrix} \right\}$ (der Buchstabe a ist jetzt durch e ersetzt) explizit durch die $\omega^{(1)}$ [s. (55)] ausdrücken, indem wir die auf der rechten Seite von (58a) enthaltene Substitution durchführen, nachdem wir zuvor dort die Symbole zu wirklichen Zahlen zusammengesetzt haben.

Nach (55) sind wegen $c \leq 2^m$ — für $m=1$ — die a_i in $\omega^{(1)}(a_1, a_2, \dots, a_n)$ ($i=2 \dots n$) der Werte 1 und 2 fähig. Überdies ist $a_1 = 1$. Wir wollen jetzt $\omega^{(1)}(a_1 \dots a_n)$ in das symbolische Produkt

$$(177) \quad \omega^{(1)}(a_1, a_2, \dots, a_n) = \omega_1(a_1) \cdot \omega_2(a_2) \cdot \omega_3(a_3) \dots \omega_n(a_n)$$

zerlegen. (Auf der rechten Seite ist der obere Index (1), da eine Verwechslung nicht möglich ist, fortgelassen.) Wir setzen weiter als Abkürzung

$$(177a) \quad \begin{cases} \omega_i(*) = \omega_i(1) + \omega_i(2); & (i = 2, 3, \dots, n). \\ \omega_1(*) = \omega_1(1), & \text{da } a_1 \text{ nur gleich.} \end{cases} \quad \text{Dagegen ist}$$

$$(177b) \quad \omega_1(*) \cdot \omega_2(*) \cdot \dots \cdot \omega_n(*) = 1 \quad [\text{nach (55a)}].$$

Dann können wir — genau so, wie in (57) die $\pi(a_1 \dots a_n)$ — die Größen $f^{(1)} \left\{ \begin{matrix} o_1 \dots o_n \\ e_1 \dots e_n \end{matrix} \right\}$ als symbolisches Produkt darstellen.

$$(78) \quad f^{(1)} \left\{ \begin{matrix} o_1, o_2, \dots, o_n \\ e_1, e_2, \dots, e_n \end{matrix} \right\} \\ = \sum_{a_i} (\omega_1(a_1) \cdot \omega_1(*)^{(c-1)})_{e_1} \cdot (\omega_2(a_2) \cdot \omega_2(*)^{(c-1)})_{e_2} \cdot \dots \cdot (\omega_n(a_n) \cdot \omega_n(*)^{(c-1)})_{e_n}.$$

$a_i = 1, 2; i = 2, \dots, n; a_1 = 1$; wenn $e_i \neq e_k$, so ist $o_i \neq o_k$; wenn $e_i = e_k$, so ist $a_i =$ oder $\neq a_k$, je nachdem $o_i =$ oder $\neq o_k$ ist. Alle $f^{(1)} \left\{ \begin{matrix} o_1 \dots o_n \\ e_1 \dots e_n \end{matrix} \right\}$, deren Indizes diesen Ungleichungen nicht genügen, sind null. Es ist z. B. $\omega_i(*)^{(4)} = \omega_i(*) \cdot \omega_i(*) \cdot \omega_i(*) \cdot \omega_i(*)$; $\omega_i(*)^{(0)} = 1$.

Im übrigen gilt für die Zusammensetzung der Symbole $\omega_i(a_i)$ und $\omega_i(*)$ zu wirklichen Zahlen dieselbe Rechenregel wie in (50) und (27a). c ist die Anzahl der verschiedenen Zahlwerte, die sich unter e_1, e_2, \dots, e_n finden. Beispiel: $f^{(1)} \left\{ \begin{matrix} 1, 2, 3 \\ 1, 1, 1 \end{matrix} \right\} = 0$; $f^{(1)} \left\{ \begin{matrix} 1, 2, 1 \\ 1, 1, 1 \end{matrix} \right\} = \sum_{a_i} \omega_1(a_1) \cdot \omega_2(a_2) \cdot \omega_3(a_3)$, da $c = 1$ ist. Da $o_3 = o_1$, so ist $a_3 = a_1 = 1$; da $o_1 \neq o_2$, so ist $a_1 \neq a_2$; also muß $a_2 = 2$ sein. $f^{(1)} \left\{ \begin{matrix} 1, 2, 1 \\ 1, 1, 1 \end{matrix} \right\} = \omega^{(1)}(1, 2, 1)$ (nach 77); ebenso $f^{(1)} \left\{ \begin{matrix} 1, 2, 2, 1, 2, 1 \\ 1, 1, 1, 1, 1, 1 \end{matrix} \right\} = \omega^{(1)}(1, 2, 2, 1, 2, 1)$ usw. Es genügt, die Behauptung (76) an einem Beispiel zu illustrieren. Sei etwa — im Gegensatz zu unsrer Voraussetzung (74) —

$$f^{(1)} \left\{ \begin{matrix} 1, 2, 2, 3, 1, 1 \\ 1, 2, 2, 3, 1, 1 \end{matrix} \right\} = 1.$$

Nach (78) wird:

$$f^{(1)} \left\{ \begin{matrix} 1, 2, 2, 3, 1, 1 \\ 1, 2, 2, 3, 1, 1 \end{matrix} \right\} = \omega^{(1)}(1, *, *, *, 1, 1) \cdot [\omega^{(1)}(*, 1, 1, *, *, *) \\ + \omega^{(1)}(*, 2, 2, *, *, *)] \cdot \omega^{(1)}(*, *, *, *, *, *).$$

Da die $\omega^{(1)}$ Wahrscheinlichkeiten sind, die der Gleichung (55a) unterliegen, so folgt, daß jeder Faktor des vorstehenden Produkts gleich 1 sein muß, wenn das ganze Produkt den Wert 1 hat. Es genügt, den Faktor $[\omega^{(1)}(*, 1, 1, *, *, *) + \omega^{(1)}(*, 2, 2, *, *, *)] = 1$ herauszugreifen.

Nach (77) und (77a) können wir schreiben:

$$\sum_{a_i} \omega^{(1)}(a_1, a_2, a_3, \dots, a_6) = 1 \quad (i = 1, 2, \dots, 6), \\ (a_i = 1, 2; i = 2, \dots, 6; a_1 = 1; a_2 = a_3).$$

Ziehen wir diese Gleichung von (55a) ab, so folgt:

$$\sum_{a_i} \omega^{(1)}(a_1, a_2, a_3, \dots, a_6) = 0 \quad (i = 1, \dots, 6), \\ (a_i = 1, 2; i = 2, \dots, 6; a_1 = 1; a_2 \neq a_3).$$

Da nun die $\omega^{(1)}$ Wahrscheinlichkeiten sind, folgt

$$\omega^{(1)}(a_1, a_2, a_3, \dots, a_6) = 0, \quad \text{wenn } a_2 \neq a_3.$$

Jetzt liefert (55) bzw. (56) — für $m = 1$ —

$$\sigma'_1 \sigma'_2 \sigma'_3 \dots \sigma'_6 = \sum_{a_i} \pi(a_1, a_2, a_3, \dots, a_6) \cdot \omega^{(1)}(a_1, a_2, a_3, \dots, a_6) \quad (i=1 \dots n)$$

$$(a_i = 1, 2; \quad i = 2, \dots, n; \quad a_1 = 1; \quad a_2 = a_3).$$

Denken wir nun an die Bedeutung von π [s. (57)], so können wir für diese Gleichung in zweckmäßiger Abänderung der Bezeichnung schreiben:

$$\sigma_1 \overline{\sigma_2 \sigma_3} \sigma'_4 \sigma'_5 \sigma'_6 = \sum_{a_i} \pi(a_1, \overline{a_2 a_3}, a_4, a_5, a_6) \cdot \omega^{(1)}(a_1, \overline{a_2 a_3}, a_4, a_5, a_6)$$

$$(a_i = 1, 2; \quad i = 2, \dots, n; \quad a_1 = 1; \quad a_2 = a_3).$$

Hierbei soll der Querbalken über $\overline{\sigma_2 \sigma_3}$ und $\overline{a_2 a_3}$ andeuten, daß wir jetzt $\overline{\sigma_2 \sigma_3}$ und $\overline{a_2 a_3}$ je als einen Buchstaben auffassen dürfen. Für das oben angeführte Schema der Papierstreifenmischung bedeutet dies, daß das zweite und dritte Kästchen eines jeden Streifens als *ein* Kästchen aufgefaßt werden dürfen, da sie bei der Streifenzerstückelung niemals getrennt werden. Und so wie in diesem Falle gelangt man stets, wenn (74) nicht erfüllt ist, von einem Streifenproblem mit n Kästchen zu einem solchen mit n' Kästchen, wo $n' < n$ ist. Es ist also klar, daß (74) die Allgemeinheit der Problemstellung nicht beeinträchtigt.

§ 6.

Die Transformation der statistischen Korrelationsmaße.

In diesem Paragraphen werden die Buchstaben a_i, e_i, o_i, u_i, z_i als Summationsindizes verwandt. Es gilt immer: $a = a[n]$, $e = e[n]$, $o = o[n]$, $u = u[n]$, $z = z[n]$ [s. (58b)]. Diese Summationsindizes können sich gegenseitig vertreten, so daß also die Auswechslung der Buchstaben lediglich schreibtechnische Bedeutung hat.

Bilden wir nach (64)

$$(79) \quad \pi^{(m)}(1, 1, 1, \dots, 1) = \sum_{o_i} \pi(o_1, o_2, \dots, o_n) \cdot f^{(m)} \left\{ \begin{matrix} o_1, o_2, \dots, o_n \\ 1, 1, \dots, 1 \end{matrix} \right\}$$

$$(i = 1, \dots, n),$$

so ist diese Gleichung — wegen (57) — identisch mit (55) und daher folgt

$$(79a) \quad \omega^{(m)}(o_1, o_2, \dots, o_n) = f^{(m)} \left\{ \begin{matrix} o_1, o_2, \dots, o_n \\ 1, 1, \dots, 1 \end{matrix} \right\} \quad (o_i \leq 2^m \text{ [s. (55)]}).$$

Nach (65a) ist dann

$$(80) \quad 1 = \sum_{o_i} \omega^{(m)}(o_1, o_2, \dots, o_n) \quad (i = 1, \dots, n; \quad o_i \leq 2^m)$$

und nach (65)

$$(80a) \quad \omega^{(m+1)}(o_1, o_2, \dots, o_n) = \sum_{u_i} \omega^{(m)}(u_1, \dots, u_n) \cdot f^{(1)} \left\{ \begin{matrix} o_1, \dots, o_n \\ u_1, \dots, u_n \end{matrix} \right\}$$

$$(i = 1, \dots, n; \quad o_i \leq 2^{m+1}; \quad u_i \leq 2^m; \quad u < = o).$$

Wir wollen nun neue Größen $g^{(m)}(a_1, a_2, \dots, a_n)$ definieren. Es sei

$$(81) \quad g^{(m)}(a_1, a_2, \dots, a_n) = \sum_{o_i} \omega^{(m)}(o_1, \dots, o_n) \quad (i = 1, \dots, n)$$

$$(0_i \leq 2^m; a = > o).$$

Nach (80 a) wird jetzt

$$(82) \quad g^{(m+1)}(a_1, a_2, \dots, a_n) = \sum_{u_i} \omega^{(m)}(u_1, \dots, u_n) \cdot d^{(1)} \left\{ \begin{matrix} a_1, \dots, a_n \\ u_1, \dots, u_n \end{matrix} \right\}$$

$$(i = 1, \dots, n; u < = a).$$

Hierbei ist gesetzt

$$(82a) \quad d^{(1)} \left\{ \begin{matrix} a_1, \dots, a_n \\ u_1, \dots, u_n \end{matrix} \right\} = \sum_{o_i} f^{(1)} \left\{ \begin{matrix} o_1, \dots, o_n \\ u_1, \dots, u_n \end{matrix} \right\}$$

$$(i = 1, \dots, n; a = > o = > u).$$

Läßt man in (81) die a_i alle Werte annehmen, die durch die Bedingung $a = a[n]$ zugelassen sind, so erhält man ein Gleichungssystem, das stets nach den Größen $\omega^{(m)}$ der rechten Seite auflösbar ist, wie man durch Induktionsschluß zeigt; daher sind die Größen $\omega^{(m)}$ als lineare Funktionen gewisser $g^{(m)}$ darstellbar. Führt man nun in (82) und (79) [s. (79a)] statt der $\omega^{(m)}$ diese linearen Funktionen der $g^{(m)}$ ein, so wird aus (79)

$$(83) \quad \pi^{(m)}(1, 1, \dots, 1) = \sum_{e_i} \gamma(e_1, e_2, \dots, e_n) \cdot g^{(m)}(e_1, e_2, \dots, e_n)$$

$$(i = 1, \dots, n).$$

Aus (82) wird entsprechend

$$(84) \quad g^{(m+1)}(a_1, a_2, \dots, a_n) = \sum_{z_i} h^{(1)} \left\{ \begin{matrix} a_1, \dots, a_n \\ z_1, \dots, z_n \end{matrix} \right\} g^{(m)}(z_1, z_2, \dots, z_n)$$

$$(i = 1, \dots, n; a = > z).$$

Um die Größen γ bzw. $h^{(1)}$ zu bestimmen, braucht man nur in (83) bzw. (84) für die $g^{(m)}$ die Werte aus (81) einzusetzen, hiernach nach den $\omega^{(m)}$ zu ordnen und dann (83) bzw. (84) mit (79) [s. (79a)] bzw. (82) zu vergleichen.

$$(85) \quad \pi(o_1, o_2, \dots, o_n) = \sum_{e_i} \gamma(e_1, e_2, \dots, e_n) \quad (i = 1, \dots, n; o < = e)$$

$$(86) \quad d^{(1)} \left\{ \begin{matrix} a_1, a_2, \dots, a_n \\ u_1, u_2, \dots, u_n \end{matrix} \right\} = \sum_{z_i} h^{(1)} \left\{ \begin{matrix} a_1, \dots, a_n \\ z_1, \dots, z_n \end{matrix} \right\}$$

$$(i = 1, \dots, n; u < = z < = a).$$

Indem man die o_i in (85) und die a_i bzw. u_i in (86) alle zugelassenen Werte annehmen läßt, erhält man Gleichungssysteme, die die Berechnung aller möglichen Größen γ und $h^{(1)}$ erlauben. Die durch (85)

definierten Größen γ nennen wir die statistischen Korrelationsmaße der $\sigma_1 \sigma_2 \dots \sigma_n$, bzw. der entsprechenden Typen (Papierstreifen des Schemas). Aus (85) folgt

$$\pi(1, 2, 3, \dots, n) = \gamma(1, 2, 3, \dots, n);$$

daher

$$(87) \quad \pi^\infty(1, 2, \dots, n) = \gamma^\infty(1, 2, \dots, n) = \pi(1, 2, \dots, n) = \gamma(1, 2, \dots, n)$$

[siehe (75 b)].

Also $\gamma^\infty(1, 2, \dots, n) = \gamma(1, 2, \dots, n)$. Dagegen sind alle übrigen

$$(87 a) \quad \gamma^\infty(e_1, e_2, \dots, e_n) = 0,$$

wenn mindestens ein $e_i \neq i$.

Denn aus (85) folgt

$$\pi^\infty(o_1 \dots o_n) = \sum_{e_i} \gamma^\infty(e_1 \dots e_n) = \pi^\infty(1, 2, \dots, n) \quad [\text{siehe (75 b)}].$$

Subtrahiert man hiervon (87):

$$\gamma^\infty(1, 2, \dots, n) = \pi^\infty(1, 2, \dots, n),$$

so folgt

$$(88) \quad \sum'_{e_i} \gamma^\infty(e_1, e_2, \dots, e_n) = 0 \quad (i = 1, \dots, n; e = > o).$$

Hierbei soll der Strich bei \sum' andeuten, daß der Summand $\gamma^\infty(1, 2, \dots, n)$ auszulassen ist. Läßt man in der symbolischen Ungleichung $e = > o$ die untere Grenze o variieren, indem man die o_i alle zugelassenen Werte annehmen läßt ($o = o[n]$), so liefert (88) ein Gleichungssystem, aus dem man sukzessive alle γ^∞ , die in (87 a) genannt sind, als null erweisen kann. Wir haben die Größen γ nicht wegen ihrer Eigenschaft (87 a) angeführt — diese Eigenschaft haben auch andere Größen —, sondern weil sie unter gewissen Voraussetzungen eine besonders einfache Transformation erfahren. Bilden wir nach (83) die Identität

$$\begin{aligned} \pi^{(m+1)}(1, 1, 1, \dots, 1) &= \sum_{e_i} \gamma'(e_1 \dots e_n) \cdot g^{(m)}(e_1 \dots e_n) \\ &\equiv \sum_{e_i} \gamma(e_1 \dots e_n) \cdot g^{(m+1)}(e_1 \dots e_n). \end{aligned}$$

Jetzt setzen wir auf der rechten Seite für $g^{(m+1)}$ den Ausdruck aus (84) ein, ordnen nach $g^{(m)}$ und finden durch Vergleich beider Seiten der vorstehenden Identität

$$(89) \quad \gamma'(z_1 \dots z_n) = \sum_{e_i} \gamma(e_1, e_2, \dots, e_n) \cdot h^{(1)} \left\{ \begin{matrix} e_1 \dots e_n \\ z_1 \dots z_n \end{matrix} \right\}$$

$$(i = 1, \dots, n; z \leq e).$$

Diese Transformation (89) brauchen wir nicht besonders zu untersuchen, da sie — formal — völlig übereinstimmt mit der Transformation

(58). Es gilt also (64) entsprechend:

$$(90) \quad \gamma^{(m)}(z_1 \dots z_n) = \sum_{e_i} \gamma(e_1, e_2, \dots, e_n) \cdot h^{(m)} \left\{ \begin{matrix} e_1 \dots e_n \\ z_1 \dots z_n \end{matrix} \right\}$$

und für $h^{(m)}$ gilt genau die Gleichung (69), nur daß wir den Buchstaben f durch h ersetzen müssen. Damit ist die m -fache Transformation der statistischen Korrelationsmaße γ in allgemeinsten Form behandelt. Wir wenden uns nun einem Spezialfall zu, der die Einführung gerade der Funktionen γ als Maße rechtfertigen wird. Zu diesem Zweck leiten wir aus der Formel (82 a), indem wir für die $f^{(1)}$ auf der rechten Seite die Werte aus (78) einsetzen, folgende Formel ab:

$$(91) \quad d^{(1)} \left\{ \begin{matrix} a_1 \dots a_n \\ u_1 \dots u_n \end{matrix} \right\} = \sum_{r_i} (\omega_1(r_1) \cdot \omega_1(*))^{(c-1)}_{u_1} \dots (\omega_n(r_n) \cdot \omega_n(*))^{(c-1)}_{u_n} \\ (i = 1, \dots, n);$$

$r_i = 1, 2; r_1 = 1; a = > u$; wenn $a_i = a_k$, so ist $r_i = r_k$; c ist die Anzahl der verschiedenen Zahlwerte, die sich unter u_1, u_2, \dots, u_n finden.

Diese Formel wollen wir auf den speziellen Fall anwenden, in dem die $w_i(a_i)$ in (77) nicht symbolische Faktoren, sondern wirkliche Zahlen sind und der Bedingung unterliegen

$$(92) \quad 0 \leq w_i(a_i) \leq 1; \quad \omega_i(*) = 1.$$

Für diese speziellen Werte geht (91) über in

$$(93) \quad d^{(1)} \left\{ \begin{matrix} a_1 \dots a_n \\ u_1 \dots u_n \end{matrix} \right\} = \sum_{r_i} \omega_1(r_1) \cdot \omega_2(r_2) \dots \omega_n(r_n) \quad (i = 1, \dots, n);$$

($a = > u; r_i = 1, 2; r_1 = 1$; wenn $a_i = a_k$, so ist $r_i = r_k$).

An dieser Gleichung ist das Wesentlichste, daß die rechte Seite von u unabhängig ist. Es ist also

$$(94) \quad d^{(1)} \left\{ \begin{matrix} a_1 \dots a_n \\ u_1 \dots u_n \end{matrix} \right\} = d^{(1)} \left\{ \begin{matrix} a_1 \dots a_n \\ a_1 \dots a_n \end{matrix} \right\} \quad (a = > u).$$

Aus (86) folgt

$$(95) \quad d^{(1)} \left\{ \begin{matrix} a_1 \dots a_n \\ a_1 \dots a_n \end{matrix} \right\} = h^{(1)} \left\{ \begin{matrix} a_1 \dots a_n \\ a_1 \dots a_n \end{matrix} \right\}.$$

Ziehen wir diese Gleichung von (86) ab, so folgt wegen (94)

$$(96) \quad 0 = \sum'_{z_i} h^{(1)} \left\{ \begin{matrix} a_1 \dots a_n \\ z_1 \dots z_n \end{matrix} \right\} \quad (i = 1, \dots, n; u < = z < = a).$$

Hierbei soll der Strich bei \sum' andeuten, daß der Summand $h^{(1)} \left\{ \begin{matrix} a_1 \dots a_n \\ a_1 \dots a_n \end{matrix} \right\}$ auszulassen ist. Hält man nun in (96) zunächst die a_i fest und läßt die u_i variieren, jedoch so, daß die symbolische Ungleichung

$u \leq a$ erfüllt bleibt, so erhält man, wenn man hernach auf die a_i variiert, Gleichungssysteme, aus denen man der Reihe nach alle

$$(97) \quad h^{(1)} \left\{ \begin{matrix} a_1 \dots a_n \\ z_1 \dots z_n \end{matrix} \right\}$$

in (96) als null erweisen kann.

Jetzt folgt sofort aus (84)

$$(98) \quad \gamma'(z_1, z_2, \dots, z_n) = \gamma(z_1, z_2, \dots, z_n) \cdot h^{(1)} \left\{ \begin{matrix} z_1 \dots z_n \\ z_1 \dots z_n \end{matrix} \right\}.$$

Wir erkennen also, daß — unter der Voraussetzung (92) — die als statistische Korrelationsmaße bezeichneten Größen γ eine ganz besonders einfache Transformation erfahren, und darum bieten sie sich von selbst als die natürlichen Maße an. Es sei hervorgehoben, daß die Voraussetzung (92) nicht die Unabhängigkeit der Erbfaktoren bedeutet, von der in den §§ 1 bis 4 die Rede war. Auf diesen Fall kommt man erst, wenn man setzt

$$(99) \quad \omega_i(a_i) = \frac{1}{2}; \quad i = 2, \dots, n; \quad \omega_1(a_1) = \omega_1(*) = 1 \quad [\text{siehe (77 a)}].$$

Wegen (97) folgt aus (84)

$$g^{(m+1)}(a_1, a_2, \dots, a_n) = g^{(m)}(a_1, a_2, \dots, a_n) \cdot h^{(1)} \left\{ \begin{matrix} a_1 \dots a_n \\ a_1 \dots a_n \end{matrix} \right\}.$$

Nun folgt aus (81), (93) und (95), daß

$$(100 \text{ a}) \quad g^{(1)}(a_1 \dots a_n) = h^{(1)} \left\{ \begin{matrix} a_1 \dots a_n \\ a_1 \dots a_n \end{matrix} \right\},$$

so daß also

$$(100) \quad g^{(m)}(a_1, a_2, \dots, a_n) = (g^{(1)}(a_1, a_2, \dots, a_n))^m.$$

Es ist klar, daß unter der Wirkung von (92) sich noch andre Vereinfachungen ergeben müssen; so wird z. B. auch $f^{(m)} \left\{ \begin{matrix} o_i \\ u_i \end{matrix} \right\}$ in (69) in einfacher Weise berechenbar, woraus sich dann wieder gewisse Identitäten ergeben. Doch wollen wir auf alle diese Formeln nicht eingehen und zum Schluß nur noch die Frage beantworten: Was wird aus (100), wenn man die Voraussetzung (99) einführt? Nach (81) ist

$$(101) \quad g^{(1)}(a_1, a_2, \dots, a_n) = \sum_{o_i} w^{(1)}(o_1 \dots o_n) \quad (i = 1, \dots, n) \\ (o_1 = 1; o_i = 1, 2; i = 2, \dots, n; a = > o).$$

Der einzelne Summand dieser Summe hat wegen (99) und (77) den Wert 2^{1-n} . Ist α die Anzahl der verschiedenen Zahlwerte, die sich unter den a_1, a_2, \dots, a_n in (101) finden, so ist $2^{\alpha-1}$ die Anzahl der Summanden auf der rechten Seite von (101). Also ist

$$(102) \quad g^{(1)}(a_1 \dots a_n) = 2^{\alpha-n} \quad \text{und nach (100):} \quad g^{(m)}(a_1 \dots a_n) = (2^m)^{\alpha-n}.$$

Setzt man diesen Wert in (83) ein, so folgt — wegen (57) —

$$\sigma_1^m \sigma_2^m \dots \sigma_n^m = \sum_{a_i} \gamma(a_1, a_2, \dots, a_n) \cdot (2^m)^{\alpha-n} \quad (i = 1, \dots, n).$$

Es ist klar, daß diese Gleichung mit der identisch ist, die aus (50) folgt, wenn man dort die rechte Seite nach Potenzen von $\frac{1}{2^m}$ entwickelt. Aus (102) und (100 a) folgt: $h^{(1)} \left\{ \begin{matrix} a_1 \dots a_n \\ a_1 \dots a_n \end{matrix} \right\} = 2^{\alpha-n}$, so daß sich also im Falle der unabhängigen Erbfaktoren [siehe (99)] ergibt:

$$(103) \quad \gamma'(a_1, a_2, \dots, a_n) = \gamma(a_1, a_2, \dots, a_n) \cdot 2^{\alpha-n}.$$

Diese Gleichung, die also besagt, daß die statistischen Korrelationsmaße beim Übergang von einer Generation zur folgenden in geometrischer Progression klein werden, kann geradezu als Definition des mendelistischen Erbganges aufgefaßt werden, wofern es sich um nichtgekoppelte Erbfaktoren handelt. Daß der allgemeineren Gleichung (98), die mit (103) in der Form übereinstimmt, auch eine solche *biologische* Bedeutung zugesprochen werden kann, ist nach den bisher vorliegenden Untersuchungen der Biologen nicht anzunehmen.

(Eingegangen am 20. 10. 1927.)

zu der Arbeit von Bartel L. van der Waerden, „Eine Verallgemeinerung des Bézoutschen Theorems“, Math. Annalen **99** (1928), S. 497–541.

Fräulein Emmy Noether machte mich in liebenswürdiger Weise auf folgenden Fehler im Schlußparagraphen (§ 12) der oben genannten Arbeit aufmerksam. Nachdem der Hauptsatz bewiesen war, sollte im Fall $r=1$ die in § 10 bewiesene Ungleichung $\chi \geq m$ (χ = charakteristische Funktion des Ideals $(\mathfrak{p}, \mathfrak{p}')_Y$; m = Multiplizität des Schnittpunktes Y) zu $\chi = m$ verschärft werden. Die Ungleichung wird zu dem Zweck fälschlich als $\chi \leq m$ zitiert und dann wird richtig $\sum \chi \geq \sum m$ bewiesen, was natürlich ungenügend ist. Um den Beweis zu vervollständigen, hat man $\sum \chi = \sum m$ zu beweisen. Verfolgt man den Beweisgang in § 12, so sieht man, daß an zwei Stellen nur gezeigt wird, daß gewisse Summanden positiv sind, während für den vollständigen Beweis erforderlich ist, daß diese Summanden = 1 sind, und zwar handelt es sich beide Male um Summanden der Gestalt $\chi(\mathfrak{p}, \mathfrak{p}')_Y$, wo das eine Mal \mathfrak{p} das zu einem allgemeinen linearen $(n-1)$ -Raum gehörige und \mathfrak{p}' das zu einer eindimensionalen Mannigfaltigkeit M' gehörige Primideal ist, während das andere Mal umgekehrt \mathfrak{p}' zu einem allgemeinen linearen eindimensionalen Raum und \mathfrak{p} zu einer $(n-1)$ -dimensionalen Mannigfaltigkeit M gehört. Also ist alles zurückzuführen auf die folgende Behauptung: Wenn \mathfrak{p} das zu einer r -dimensionalen Mannigfaltigkeit gehörige homogene Primideal und $\mathfrak{p}' = (l_1, \dots, l_r)$ das zu einem allgemeinen linearen $(n-r)$ -Raum gehörige homogene Primideal ist (l_1, \dots, l_r allgemeine Linearformen), so hat das Ideal $(\mathfrak{p}, \mathfrak{p}')$ als Primärkomponenten lauter Primideale, die zu je einem Schnittpunkt Y gehören. Diese Behauptung nun folgt unmittelbar aus Satz 27, 3 (§ 6).